# Molecular Parataxis: Molecular Locks and Boosters

Arthur Weininger and Susan Weininger


## Abstract

Molecular Locks are molecular assemblies that are engineered to specifically and tightly bind a specific hybrid nucleic acid target Molecular Locks are made from locking components that each home to a part of a target nucleic acid and then lock together when all locking components have been assembled side by side in a process called "Molecular Parataxis" Molecular Locks can be used to bind targets with or without the use of single-stranded hybridizing probes There are a number of advantages to using Molecular Locks to engage targets and these include exacting discrimination of target sequence and geometry. Molecular Locks can be used to signal the presence of a single target molecule and can be used to immobilize small numbers of targets from a crude sample. Molecular Locks have some unique thermal properties that make them suitable for running tests at elevated temperatures in crude environments.


## Contents

## 1. Background

Traditional nucleic acid detection systems have relied upon hybridization of a single-stranded probe nucleic acid to a single-stranded target nucleic acid as the basis for recognition of the target and must be run under conditions tailored to control hybridization of the target nucleic acid. This limits access to the target and often necessitates thermal cycling. Although nucleic acids (e.g., hybridizing probes, RNAi constructs) can bind to specific nucleic acid targets, they can mis-hybridize, are easily degraded and have issues regarding temperature sensitivity, stability and ease of use. The use of Molecular Locks as the basis for target recognition and signal amplification corrects many of the defects of current molecular amplification and detection systems.

Currently, less expensive antibody tests are often used instead of the more expensive DNA and RNA tests to screen biological samples for the presence of a particular disease-causing organism, such as HIV, Hepatitis SARS, or HPV due to cost considerations. Current tests fail to deliver information regarding the genomic constituents responsible for pathogenicity, e.g., whether the pathogenicity clusters that can make H. pylori dangerous are present with H. pylori or without H. pylori (i.e, present in other G.I. residing bacteria that have picked up these genes). Both antibody and nucleic acid tests also currently fail to deliver other genomic information, such as state of integration, which may be critical to clinical assessment and treatment. While there have been new developments in increased detection-throughput (e.g., indicators, beads, and robotics), current practice remains the use of antibody screens backed up by nucleic acid detection by PCR. The cost of PCR in terms of time and material make it unable to be a good economic alternative for antibody tests. In addition, there are often patents, e.g., sequence patents, which block the development of diagnostic tests where a copy of the target must be made and labeled as part of the test.

## 2. Molecular Parataxis Overview

“Parataxis” means “to place side-by-side” from the Greek (“para”, beside + “tassein”, to arrange). A Molecular Lock that specifically binds a nucleic acid target through Molecular Parataxis may be developed by identifying structural signatures within the target, making locking components that bind to these structural signatures, and then adding organizing components that assemble these nucleic acid binding domains in a specific geometry that binds cooperatively to the structural signatures within the target nucleic acid. The assembly of the locking components into the Molecular Lock has another function: molecules with weak but specific binding to a portion of the target can be assembled into a Molecular Lock that has extremely high binding affinity and specificity for the complete target. Molecular Parataxis components are used to detect the target nucleic acid directly without target replication. The detection signal is amplified with a simple nucleic acid polymer. Molecular Locks fix the fidelity of target detection and signal amplification and provide for target immobilization. Molecular Locks allow reaction conditions to be shifted to favor the single-stranded form of the nucleic acid. The binding of a Molecular Lock to a double-stranded target or a target/probe hybrid shifts the equilibrium by selectively stabilizing the double-stranded target. Molecular Locks can directly detect double-stranded nucleic acid targets without a single-stranded nucleic acid probe. Molecular Locks can be designed so that all of the components of detection and signal amplification can be added at the same time; thermal cycling is not required. While some structures of the Molecular Lock

components relax at high temperatures, the lock will assemble when presented with target and lock onto the target and report its presence at elevated temperatures.

## 3. Target Detection by Molecular Parataxis

Most naturally occurring nucleic acid binding domains bind to multiple sites in a given genome in addition to sites within a target. These naturally occurring nucleic acid binding domains can be modified and incorporated into a Molecular Lock that will bind specifically to a target comprised of multiple binding sites by downshifting the binding affinity of these domains and then mounting them onto an oligomerization scaffold to present them to the target as a cooperative assembly. Molecular Lock specificity and high binding affinity for the target is created by assembling multiple, low binding affinity domains into a single cooperative molecular assembly. Binding of the nucleic acid target by a Molecular Lock is accomplished by Molecular Parataxis, the cooperative binding of weaker nucleic acid binding domains that bind side-by-side on targets to create a high-affinity, target-specific binding assembly. Because the binding of the target by a Molecular Lock is accomplished by the cooperative binding of weak binding domains that bind side-by-side on targets to create a high-affinity, targetspecific binding assembly, Molecular Locks bind tightly to the target only when the entire target is present and all binding components of the Molecular Lock are bound.

## 4. Molecular Lock Specificity: Cooperative Binding

The creation of Molecular Lock specificity through cooperative binding of Molecular Locking components is illustrated by the case of a target site containing two binding sites **a** and **b** that are each specifically bound by the DNA binding domains **A** and **B**, respectively. The DNA binding domain **A** is not capable of distinguishing single **a** sites contained in the background genome from the a site contained in the target sequence. Similarly, the DNA binding domain **B** is not capable of distinguishing single **b** sites contained in the background genome from the **b** site contained in the target sequence. If the binding affinity of **A** for **a** and **B** for **b** are on the same order and equal to **n**, we can construct a Molecular Lock that is specific for the target site **a-b** by making some modifications to **A** and **B**.

First, we make mutants of **A** and **B**, called **A´** and **B´**, that bind specifically to **a** and **b**, respectively, but whose affinity for **a** and **b**, respectively, are reduced by *d*. Thus, the affinity of **A´** for **a** is equal to *n-d* and the affinity of **B´** for **b** is equal to *n-d*. Second, we mount **A´** and **B´** on dimerization domains that present the **A´** and **B´** binding domains **(A´- B´)** in the proper geometry to bind to **a-b**. As long as the binding affinity of **A´** for **a** and **B´** for **b** is not downshifted too much, the binding affinity of **A´- B´** for the **a-b** target sequence will be much greater than the affinity of **A** for **a** or **B** for **b**. This is illustrated in Figure 1.

The binding affinity of a molecule that binds cooperatively with two binding domains is on the order of the square of the binding affinity of a molecule binding with only one of the domains. For example, if $n=10^{-7}$M, $d=10^{2}$M, an **A´- B´** dimer (constructed so that the **A´** and **B´** domains bind cooperatively to the target) will bind with an affinity close to $(n-d)^2$ or on the order of $10^{-10}$ M. This means that the **A´- B´** Molecular Lock binds with an affinity that is a thousand times greater for the **a-b** target than the single **a** site. It also means that an **A´- B´** dimer will displace any **A** domains binding to single **a** sites within the target. This is illustrated in Figure 2. **A** and **B** can be used as a wash molecules for **A´- B´** binding to single sites in reagent and diagnostic applications. For a similar reason, a properly designed **A´- B´** Molecular Lock *in vivo*

would not be expected to interfere with normal cellular trafficking of **A** to single **a** binding sites. **A´- B´** Molecular Locks are designed to home to **a-b** target sites and be washed off of single **a** and **b** sites by naturally occurring **A** and **B** domains. It follows that the more domains used in binding the target, the greater is the selectivity of the Molecular Lock for the target. Figure 2 illustrates the case where the target contains two binding sites, **a** and **b,** that are bound by a Molecular Lock comprised of two chaperoned domains **A´** and **B´**.

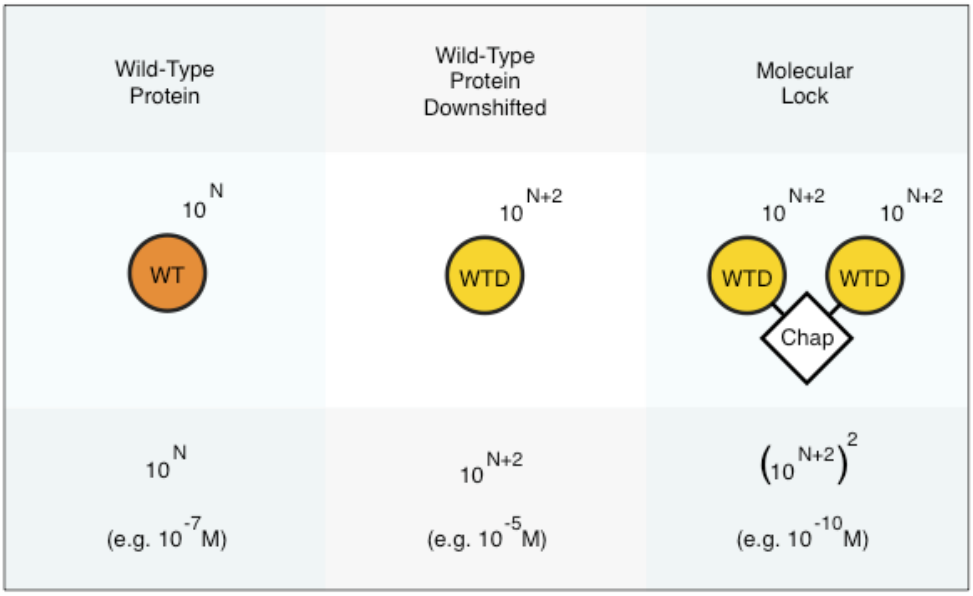



*Figure 1. The binding affinity of a molecular assembly that binds cooperatively with two binding domains is on the order of the square of the binding affinity of a molecule binding with only one of the domains.*

A single binding protein binds wherever its binding site is present and does not provide a basis for discriminating a target sequence from its background (e.g. a viral genome from the host cell background genome). Molecular Locks can be engineered to specifically bind to a target made up of individual binding components that are assembled into a specific spatial pattern (such as a viral promoter or cellular enhancer), even when the individual binding components may be separately found in many places in the background genome. This feature of Molecular Locks distinguishes them from simple DNA binding domains. In addition, each of the Molecular Lock components can be made so that they engage their portion of the target nucleic acid weakly but specifically. The assembled Molecular Lock has high binding affinity for the target nucleic acid but the individual components bind only weakly to portions of the target sequence found elsewhere in the genome. This is an important property of Molecular Locks if used *in situ* or *in vivo*.

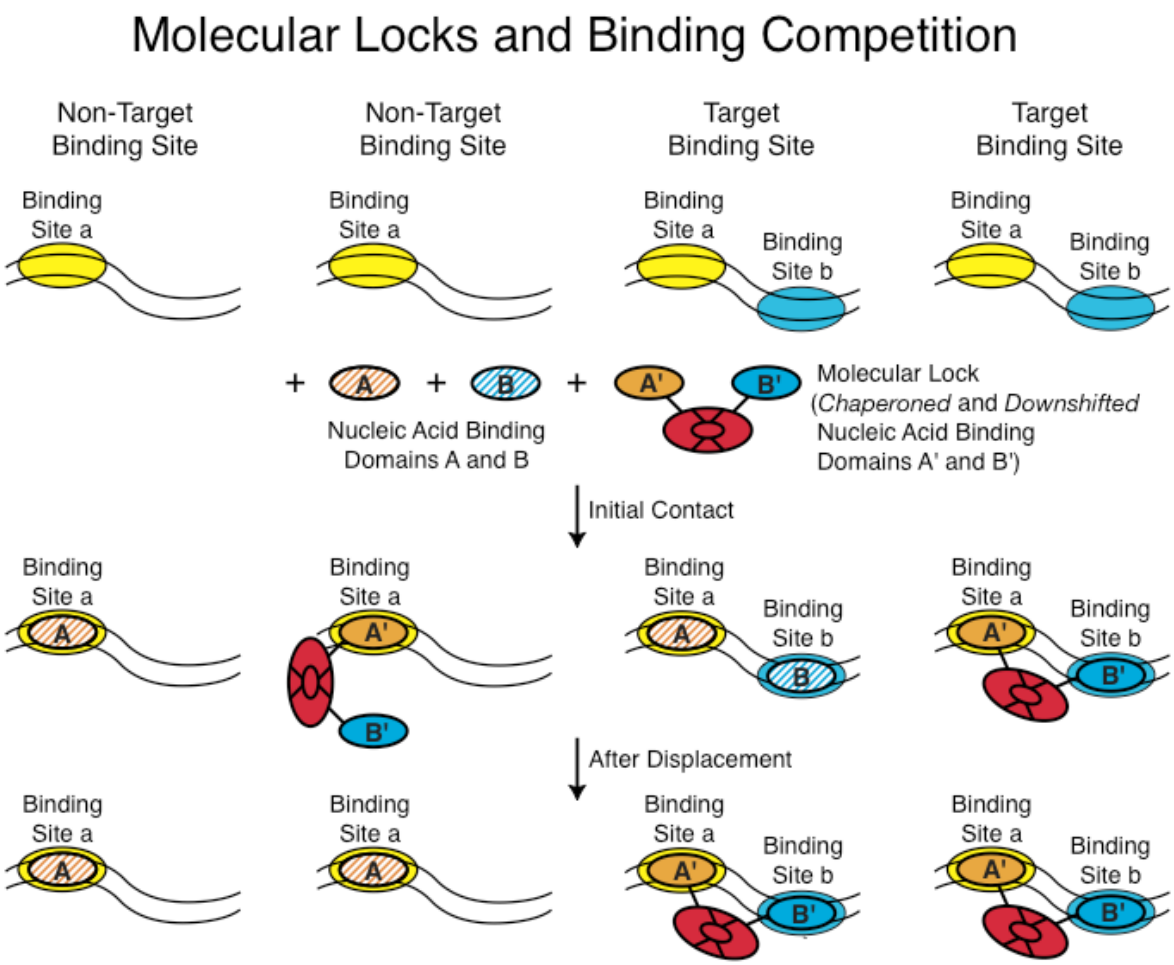



*Figure 2. Downshifting the binding affinity of the binding components of a Molecular Lock provides a method for discriminating a multi-component nucleic acid target from a single component in the background genome.*

## 5. Example of Target Specific Molecular Locks: HIV-Locks™

Human Immunodeficiency Virus (HIV) replication is controlled primarily by cis-acting transcription factors that bind to recognition sites in viral long terminal repeats (LTRs). The structure of the HIV LTR is highly conserved but not invariant. Almost all variations in the HIV LTR to date appear to conserve the register and number of binding sites for certain transcriptional regulators. However, changes such as an addition of a binding site (Sp1) and substitution of a binding site (GABP for NFKB) have been observed (Van Opijnen et al., 2004). Multiple transcription factor binding sites in the HIV-LTR for an HIV-1 genome are highlighted in Figure 3.

```
9361 atttc gtcac atggc ccgag agctg catcc
9391 ggagt actac aaaga ctgct gacat cgagt
9421 tttct acaag ggact ttccg ctggg gactt
9451 tccag gggag gcgtg gcctg ggcgg gactg
9481 gggag tggcg agccc tcaga tgctg catat
9511 aagca gctgc ttttt gcctg tacgg ggtct
9541 ctctg gttag accag atctg agcct gggag
9571 ctctc tggct aacta gggaa cccac tgctt
9601 aagcc tcaat aaagc ttgcc ttgag tgctt
9631 caagt agtgt gtgcc cgtct gttgt gtgac
9661 tctgc tatct agaga tccct cagac cctt
9691 tagtc agtgt ggaaa atctc tagca atata
```

*Figure 3. HIV LTR BASES 9361 to 9720 (Higgins, D et al., 1990). The bases are colored as follows:*
**P50/rel binding sites, P50/p52 binding sites, Sp1 binding sites, TCF-Ialpha binding site, Ap1 binding site, NFAT binding site.**

The Rel/NFKB transcription factor family includes p105, p110, p65, c-Rel and RelB. The DNA-binding portions of these molecules assemble into homodimers and heterodimers, bind to a related set of binding sites, and regulate cellular responses such as the immune response, inflammation, growth and antiapoptosis (Sen and Baltimore, 1986; Ghosh et al. 1990; Huang et al., 2001). As can be seen from Figure 3, they also bind to and regulate HIV. Van Opijnen et al. (2004) catalogue the variations of HIV-1 that include the loss of the first NFKB binding site in the LTR (NFKB-NFKB-Sp1-Sp1-Sp1). There are several molecular assemblies that can be built to specifically recognize HIV. The sequences of two additional p50/p52 binding sites have been characterized by Montanyo et al. (1996) who suggest that these sites may regulate pre-initiation complexes. By way of example, we shall discuss lock components that could be utilized to recognize the two p50/p52 binding sites at positions 9535 and 9566: gggtctctct and gggagctctc.

In order to design a Molecular Lock for these sites, we must first look at the geometry of the target nucleic acid and assemble the components of the Molecular Lock. The target nucleic acid sequence (HIV LTR) contains two binding sites for p50/p52 (see Figure 3 sites highlighted in brown) that are separated by 21 base pairs and have the centers of their binding sites separated by 31 base pairs:

```
HIV TARGET NUCLEIC ACID
                          123456789012345678901
ttttgcctgtacgggtctctctggttagaccagatctgagcctgggagctctctggct
                    123456789012345678901234567890
```

The spacing of the p50/p52 binding sites is such that the binding sites are on the same side of the DNA helix and separated by approximately three turns of the helix (10.4 base pairs per turn x 3 turns = 31.2 base pairs). Figure 4 illustrates one type of Molecular Lock and the locking and anchoring components that will be useful to engage the HIV LTR. Site A (Fig. 4) may be used to release the HIV LTR with the Molecular Lock still engaged. Site B (Fig. 4) may be used to release the Molecular Lock when higher binding affinity wash molecules (protein or nucleic acid) are introduced to compete with the binding of the anchor portion of the Lock to the

anchoring nucleic acid. Site B (Fig. 4) may be used to release the Molecular Lock from the TNA. This is particularly useful if PCR will be used as a follow-up step to target capture.

For this example, we select cro protein binding sites as the anchoring recognition sites. We can construct an anchor for the target side of the HIV-Lock™ and create the cooperativity of binding by building a piece of DNA with embedded cro binding sites whose placement "mirrors" the HIV sites that we have selected. The cro protein binds with high affinity to binding sites contained within the left and right operators of bacteriophage lambda. The binding affinity of cro for OR3 is $10^{-12}$M, for OR2 is $10^{-10}$M, and for OR1 is is $10^{-11}$M (Johnson et al., 1978). The naturally occurring binding site sequences are listed below (Dahlberg and Blattner, 1975):

```
Left Operator Sequences (LETTERS IN RED: OL1, OL2, OL3)
    35581 atgtgctcag tatcaccgcc agtggtattt atgtcaacac cgccagagat aatttatcac
    35641 cgcagatggt tatctgtatg ttttttatat gaatttattt tttgcagggg ggcattgttt

Right Operator Sequences (LETTERS IN RED: OR3, OR2, OR1)
    37921 ggtttctttt ttgtgctcat acgttaaatc tatcaccgca agggataaat atctaacacc
    37981 gtgcgtgttg actattttac ctctggcggt gataatggtt gcatgtacta aggaggttgt
```

In order to bind the HIV LTR using the lock configuration presented in Figure 4, we must engineer a protein that will bridge between the HIV LTR and the anchoring sequences. We choose to construct the anchoring sequences from cro binding sites. Possible mirror sites are as follows (only one strand of the double strand is shown), in order of highest affinity:

```
OR3-OR3 TARGET STRUCTURE MIRROR
                              12345678901234567890123456789012
HIV    ...tttttgcctgtacgggtctctctggttagaccagatctgagcctgggagctctctggct...
MIRROR ...nnnntaaatctatcaccgcaagggataaatccnnnntaaatctatcaccgcaagggataaatcc...
                             012345678901234567890123456789012

OR3-OR1 TARGET STRUCTURE MIRROR
                              12345678901234567890123456789012
HIV    ...tttttgcctgtacgggtctctctggttagaccagatctgagcctgggagctctctggct...
MIRROR ...nnnntaaatctatcaccgcaagggataaatccnnnntatttacctctggcggtgataatggtt...
                             012345678901234567890123456789012

OR2-OR2 TARGET STRUCTURE MIRROR
                              12345678901234567890123456789012
HIV    ...tttttgcctgtacgggtctctctggttagaccagatctgagcctgggagctctctggct...
MIRROR    nnnnaatatctaacaccgtgcgtgttgactatnnnnaatatctaacaccgtgcgtgttgactat...
                             012345678901234567890123456789012
```

The **nnnn** nucleotides, in the sequences above, may be replaced or expanded to produce a restriction site or other feature that allows the release of the target nucleic acid with the Molecular Lock still engaged and the ***nnnn*** nucleotides may be modified similarly to enable the target to be released from the Molecular Lock as described above.

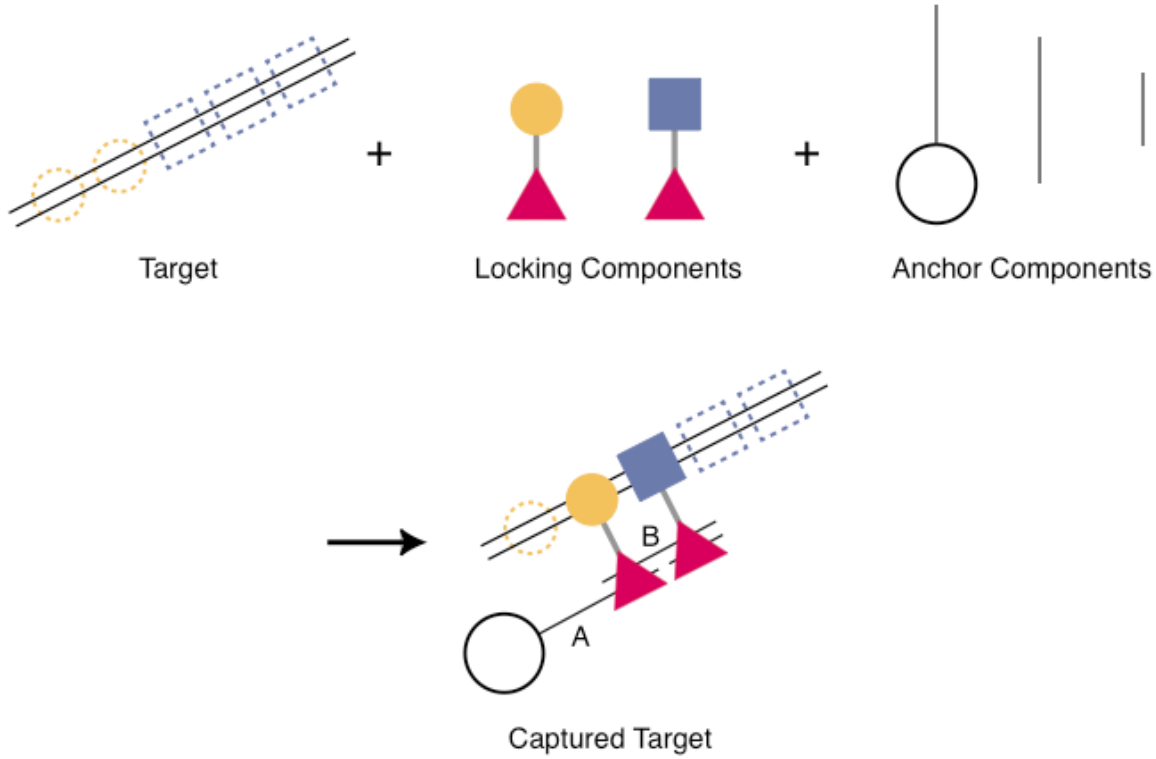



*Figure 4. Locking and anchoring configuration used to capture a nucleic acid target. Sites A and B may incorporate restriction sites or other structures for release of the captured nucleic acid target.*

The protein fragments listed below have the necessary parts for the formation of a Molecular Lock for the selected target sequence (please see Appendices A and B for more information of the structure and assembly of these molecules):

**Fragment 1: P50 (Mus musculus), Residues 1-42 (non-structured N-terminal piece):**

MAEDD PYLGR PEQMF HLDPS LTHTI FNPEV FQPQM ALPTA DG

**Fragment 2: P50 (Mus musculus), Residues 43-359 (DNA-binding,organizing and short linker):**

PYLQI LEQPK QRGFR FRYVC EGPSH GGLPG ASSEK NKKSY PQVKI CNYVG
PAKVI VQLVT NGKNI HLHAH SLVGK HCEDG VCTVT AGPKD MVVGF ANLGI
LHVTK KKVFE TLEAR MTEAC IRGYN PGLLV HSDLA YLQAE GGGDR QLTDR
EKEII RQAAV QQTKE MDLSV VRLMF TAFLP DSTGS FTRRL EPVVS DAIYD
SKAPN ASNLK IVRMD RTAGC VTGGE EIYLL CDKVQ KDDIQ IRFYE EEENG
GVWEG FGDFS PTDVH RQFAI VFKTP KYKDV NITKP ASVFV QLRRK SDLET
SEPKP FLYYP EIKDK EE

**Fragment 3: P50 (Mus musculus), 360-387 (additional linking piece):**

VQRKR QKLMP NFSDS FGGGS GAGAG GGG

**Fragment 4: Cro Protein (Bacteriophage Lambda) 1-66 (DNA binding and organizing):**

MEQRI TLKDY AMRFG QTKTA KDLGV YQSAI NKAIH AGRKI FLTIN ADGSV
YAEEV KPFPS NKKTT A

**Fragment 5: C-rel (Gallus gallus) 1-7 (more structured N-terminal piece):**

MASGA YN

**Fragment 6: C-rel (Gallus gallus) 8-285 (DNA-binding,organizing and short linker):**

PYIEI IEQPR QRGMR FRYKC EGRSA GSIPG EHSTD NNRTY PSIQI MNYYG
KGKVR ITLVT KNDPY KPHPH DLVGK DCRDG YYEAE FGQER RPLFF QNLGI
RCVKK KEVKE AIITR IKAGI NPFNV PEKQL NDIED CDLNV VRLCF QVFLP
DEHGN LTTAL PPVVS NPIYD NRAPN TAELR ICRVN KNCGS VRGGD EIFLL
CDKVQ KDDIE VRFVL NDWEA KGIFS QADVH RQVAI VFKTP PYCKA ITEPV
TVKMQ LRRPS DQEVS ESMDF RYLPD EK

A set of Molecular Lock components can be built that are capable of distinguishing the specific transcription factor binding sites that are found in the HIV-LTR from transcription factor binding sites found elsewhere, including in the background cellular genome. Molecular Parataxis can be used to build an HIV-Lock™ that will specifically bind to and immobilize the HIV LTR and not to sequences in the background genome. From the sequences above, a locking component, A, is built by ligating a Rel binding domain (e.g., p50 or c-Rel) to an organizing domain (e.g., the cro protein from bacteriophage lambda) to create a homodimer that binds a sequence in the HIV LTR and a sequence in the anchoring nucleic acid. The resultant molecule has two DNA binding domains, one that binds to the target (p50) and one that organizes and anchors (cro). When both binding sites in the HIV LTR and both sites in the anchor are bound, the entire complex acts as a Molecular Lock for the HIV LTR.

Several constructs have the correct geometry and specificity to make good locking components including fragments: 1-2-4, 5-2-4, 1-2-3-4, and 5-2-3-4. The structural features of the locking components are shown in Figures 5-7. Figure 5 shows a ribbon depiction of a dimer of Fragment 2-4 (p50-cro). One monomer is yellow and one monomer is purple. The atoms of the ligating residues are shown as spheres. Figure 6 shows a space filling depiction of a dimer of Fragment 2-4 (p50-cro). All atoms are filled. Figure 7 shows the Fragment 2-4 dimer bound to a single recognition site in the target and a single mirror site. As can be seen from Figures 5-7, the geometry of a HIV LTR-p50-cro dimer-anchor binding site complex has a tight, precise geometry and the helices are bound in a parallel fashion. It can also be seen that the assembly and binding of the monomers can proceed independently of one another. The p50-cro dimer can be expected to protect the HIV LTR from degradation and bind to the double operator site when this site is introduced. The geometric constraint can be relaxed by using a Fragment 1-2-3-4 or 5-2-3-4 Molecular Lock.

Other components can function in place of the fragments 1 and 2. Rel has nearly the same geometry as p50 but there are subtle differences between the two proteins that can be harnessed. Rel is capable of binding to the NFKB sites in the LTR but can also interact with sites that are

not bound by NFKB (Sica et al., 1992). Figure 8 shows the residues of p50 in red that c-Rel is missing.

Other Molecular Locks can be constructed that recognize an HIV target sequence even though portions of the LTR targeted (e.g., NFKB and Sp1 binding sites) are found elsewhere. A Molecular Lock can easily discriminate the HIV-LTR (with its two adjacent NFKB binding sites next to three adjacent Sp1 binding sites) from other sites containing identical NFKB binding sites (such as in the Beta-2-microglobulin promoter as well as in the Kappa-light chain promoter in the human genome). Mirror sites and Molecular Locking components for the NFKB, Sp1 and ap1 binding sites in the HIV-LTR, engineered in a similar fashion, can be used to "sandwich" the target genome and link to the Stacked Boosters described in the Signal Amplification section below.

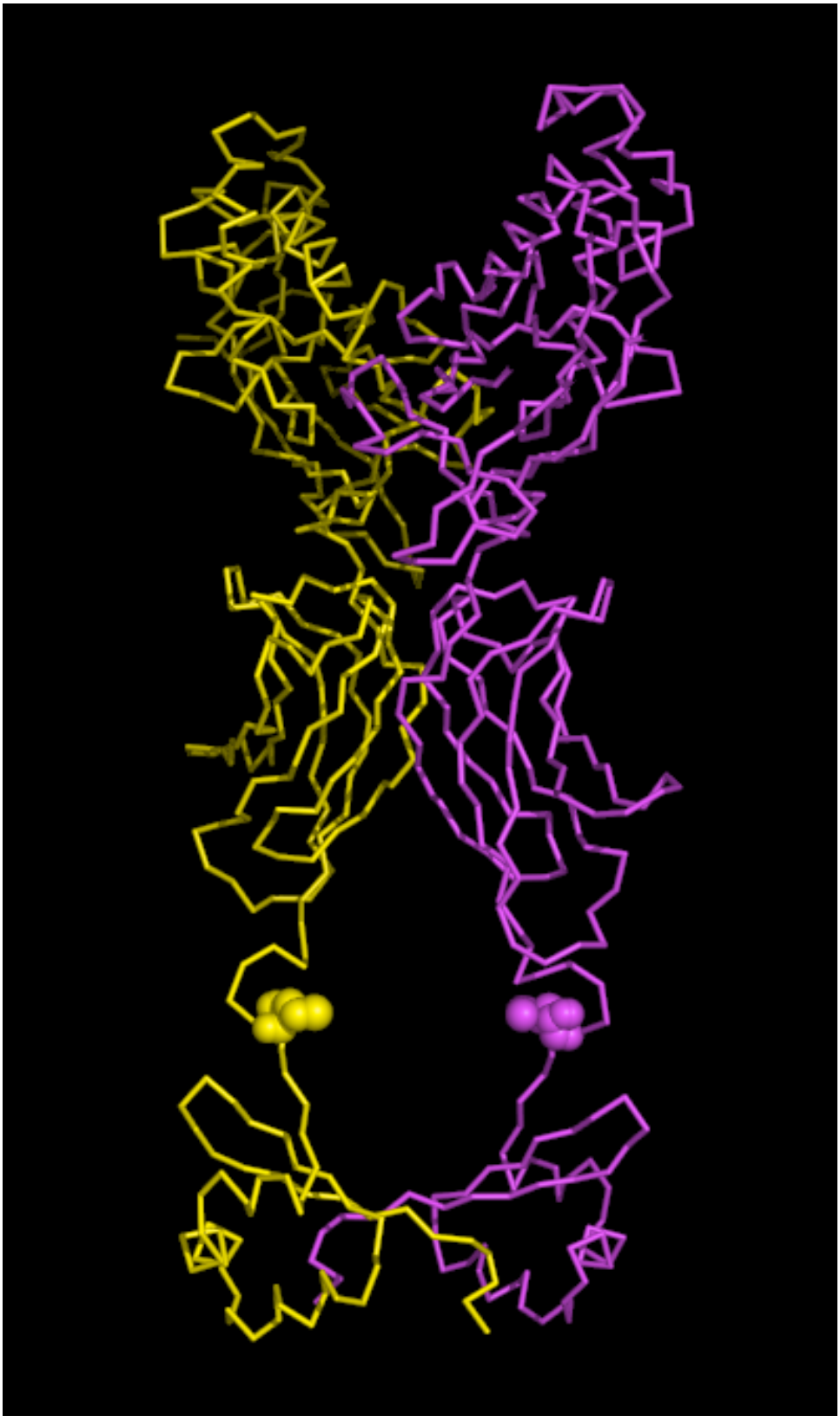


*Figure 5.*

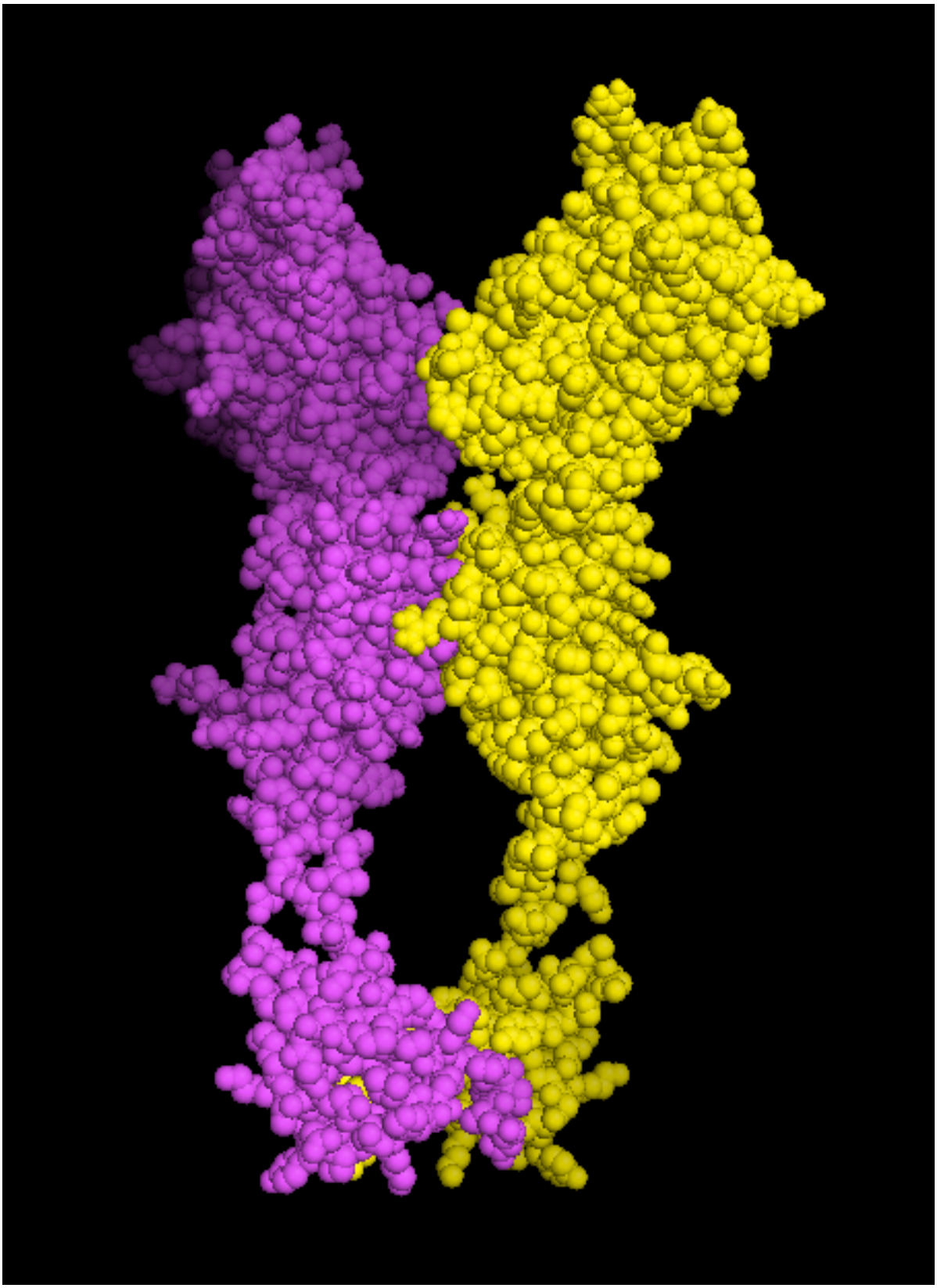


*Figure 6.*

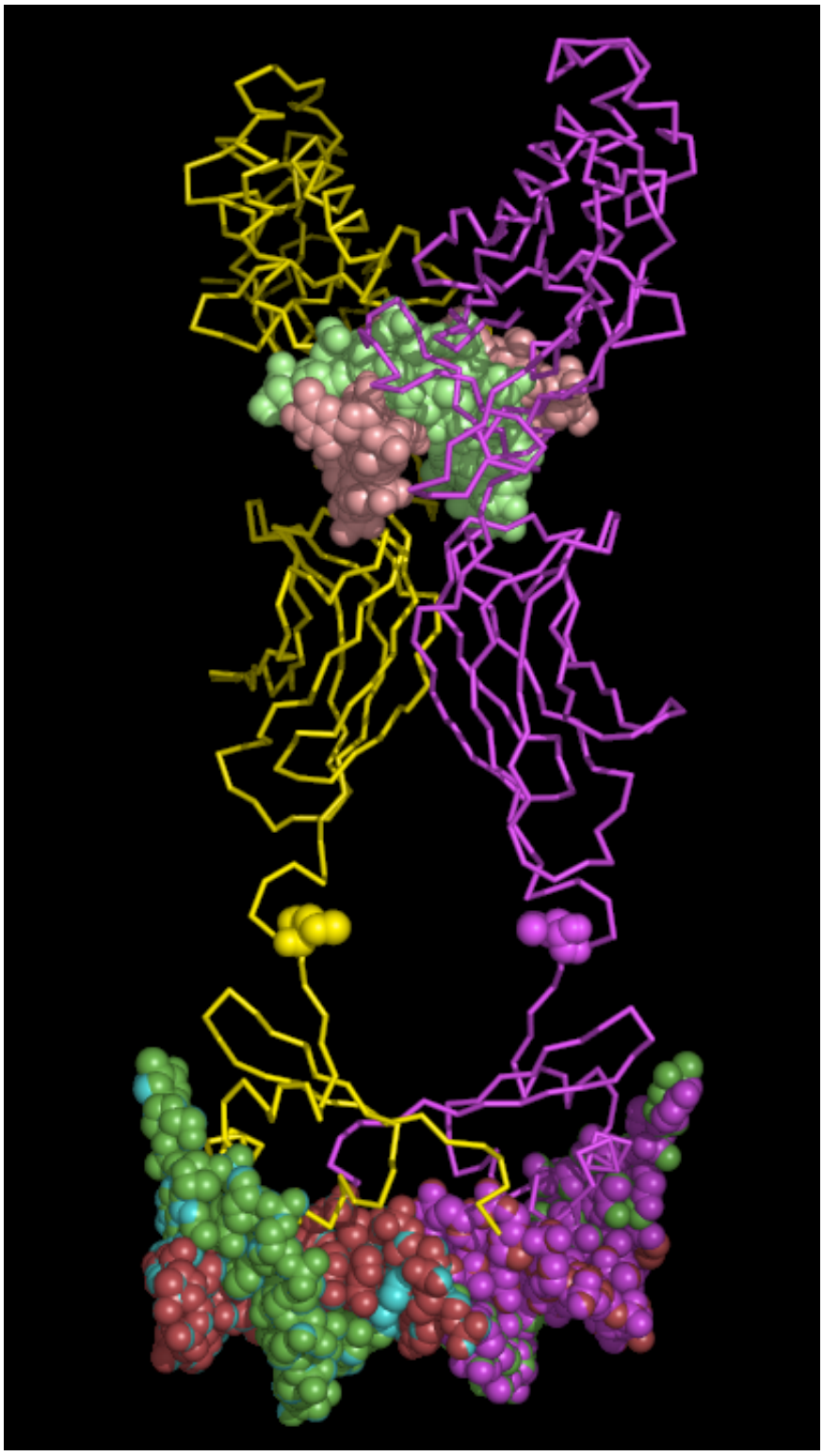


*Figure 7*.

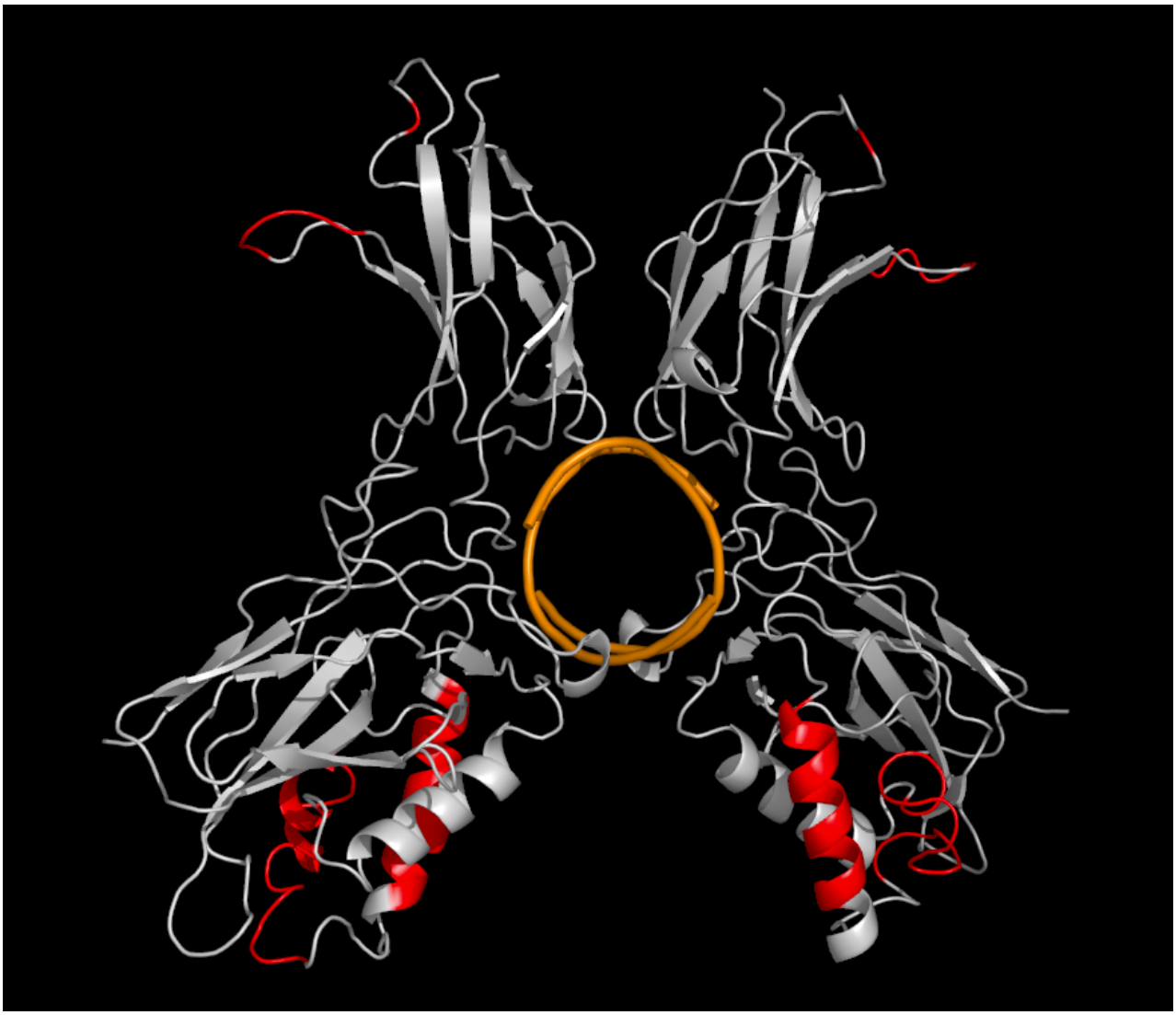


*Figure 8.*

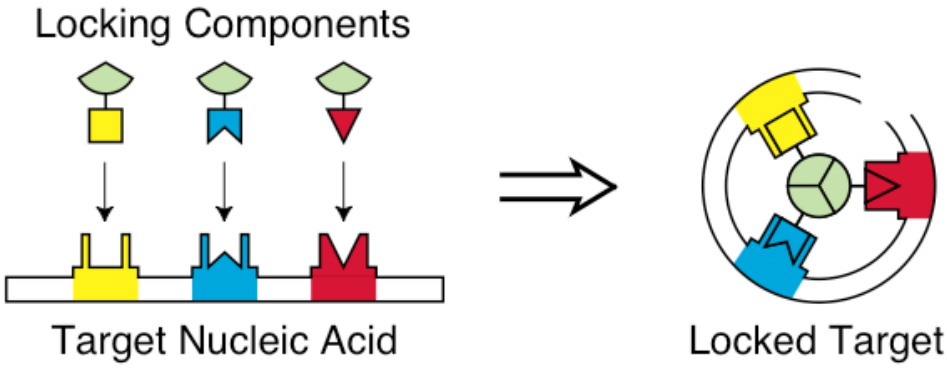



*Figure 9. Molecular Lock geometry for in situ or in vivo use.*

Molecular Locks can be made in different geometries. The geometry illustrated in Figure 9 is particularly useful for Molecular Locks that are to be used *in situ* and *in vivo,* particularly if the Molecular Lock is to be introduced using a viral vector, must accommodate a nuclear localization signal, and must stay engaged except at cell division. In this configuration, there is no geometric mirroring and the lock is target nucleic acid sequence and geometry selective. This Molecular Lock configuration is particularly useful when it is desirable to to target the HIV control region (HIV-LTR) without interfering with normal cellular transcription factor trafficking through cousin sites such as the beta-2 microglobulin and kappa-light chain promoters. In this configuration, a Molecular Lock can bind to a target exclusively in a manner not achievable by single binding domains or proteins because it is composed of multiple binding components that are cooperatively bound together. Each of these binding components is: a) selective for an individual binding region within the target nucleic acid, and b) has the binding affinity of its components down-shifted. Although the individual Molecular Lock components bind weakly, the entire Molecular Lock binds with great strength and specificity as the individual Molecular Lock binding components are cooperatively bound together. This feature of a Molecular Lock allows it to target an individual control region and to not interfere with other cellular sites that contain (even exactly identical) individual binding sites that are present in the targeted control region.

It is clear that any therapeutic molecule based on a single nucleic acid binding protein would interact with both viral and cellular targets and would not be capable of discriminating between the two and can be expected to have no utility *in vivo* as a therapy. This is particularly true for zinc finger DNA binding proteins (e.g. sp1), either naturally occurring or redesigned. Sp1 binds independently to each GC box sequence and physical interaction between adjacent Sp1 molecules is insufficient to give rise to cooperative DNA binding behavior (Narayan, et al. (1997)). Individual zinc fingers alone have no specificity and together they act as a single binding domain for their target sequence. Zinc finger binding sites are present multiple times in most genomes and are always present in both viral and cellular genomes. This is due to the fact that viruses that integrate into the host cell genome generally utilize host cell transcription factors for activation.

Any effective therapy has to address the issue of non-interference with normal cellular trafficking of transcription factors. Effective and useful therapies cannot be made simply by using proteins that could be expected to affect cellular regulation, however slight. Molecular Locking is a method to selectively bind and therefore modulate the activity of a viral target nucleic acid without affecting the host cell genome. By creating a target binding assembly that recognizes both the presence and geometry of the binding sites within the target nucleic acid, target nucleic acids in a sample can be discriminated from the background genome. For this reason, Molecular Locks may become a new class of therapeutic blockers.

## 6. Signal Amplification

Molecular Locks can detect either an existing hybrid nucleic acid target fragment directly from a sample ("probeless") or a single stranded nucleic acid to which a probe nucleic acid is introduced to hybridize with the target nucleic acid. The probe may consist of both a sequence complement to the target as well as an extra strand for attaching a signal booster. This is illustrated in Figure 10.

Boosters are signal amplification polymers that may contain a fixed ratio of labels. This means that the detection of a molecule with a Booster can be reduced to the detection of a ratio of fluorophores. Molecular locking components can be added to nucleic acids of unknown composition and then "read" using a Booster molecule. These Booster molecules can carry multiple labels and amplify a binding event so that it can be detected. Boosters are comprised of simple, linked nucleic acid polymers that may engage the probe and are fixed with their own detectable binding proteins called Booster Locks.

Two configuration choices for Booster polymers are linear and stacking. The Booster Locks prevent mismatch hybridization of Booster nucleic acids, ensuring the link between the signal and the target. The signal amplification factor is controlled by the assembly of Booster polymer nucleic acids and associated binding proteins. For quantitative tests, controlled amplification may be attained by making the Booster Nucleic Acids (BNAs) unique, such that they hybridize to one specific position in the polymer, and thus produce a polymer of a given length. This length produces a given signal strength that can be directly correlated with the target density. For qualitative screening tests, unrestrained amplification is used to achieve the strongest possible signal; Booster Nucleic Acids have been designed by us to form repeating sections that may link together in lengths that are concentration dependent to address the question of whether the target is present in a sample.

The Stacking Booster shown in Figure 10 is a qualitative signal amplification polymer. Stacking Boosters can easily be made with Bacteriophage Lambda Operator sites. The cI repressor protein binds with nanomolar binding affinity to OL1 (.5x), OL2 (.5x), OR1 (1x) and OR2 (2x), where x= 3nm=concentration of repressor required to fill half the OR1 sites (Johnson et al., 1978).

The cI repressor assembles on the OL1, Ol2, OR2 and OR1 sites and causes the DNA to loop back (Dodd et al., 2004). Octamerization of the cI repressors through their c-terminal domains *in vivo* creates a complex that acts as a part of the lytic-lysogenic switch by stabilizing additional cI binding that represses the Lambda $P_{RM}$ promoter (ibid).

The cI repressor has a DNA binding domain for the N-terminus (in green) and a oligomerization domain for the C-terminus (in blue) (Sauer and Anderegg, 1978):

MSTKK KPLTQ EQLED ARRLK AIYEK KKNEL GLSQE SVADK MGMGQ SGVGA
LFNGI NALNA YNAAL LAKIL KVSVE EFSPS IAREI YEMYE AVSMQ PSLRS
EYEYP VFSHV QAGMF SPELR TFTKG DAERW VSTTK KASDS AFWLE VEGNS
MTAPT GSKPS FPDGM LILVD PEQAV EPGDF CIARL GGDEF TFKKL IRDSG
QVFLQ PLNPQ YPMIP CNESC SVVGK VIASQ WPEET FG

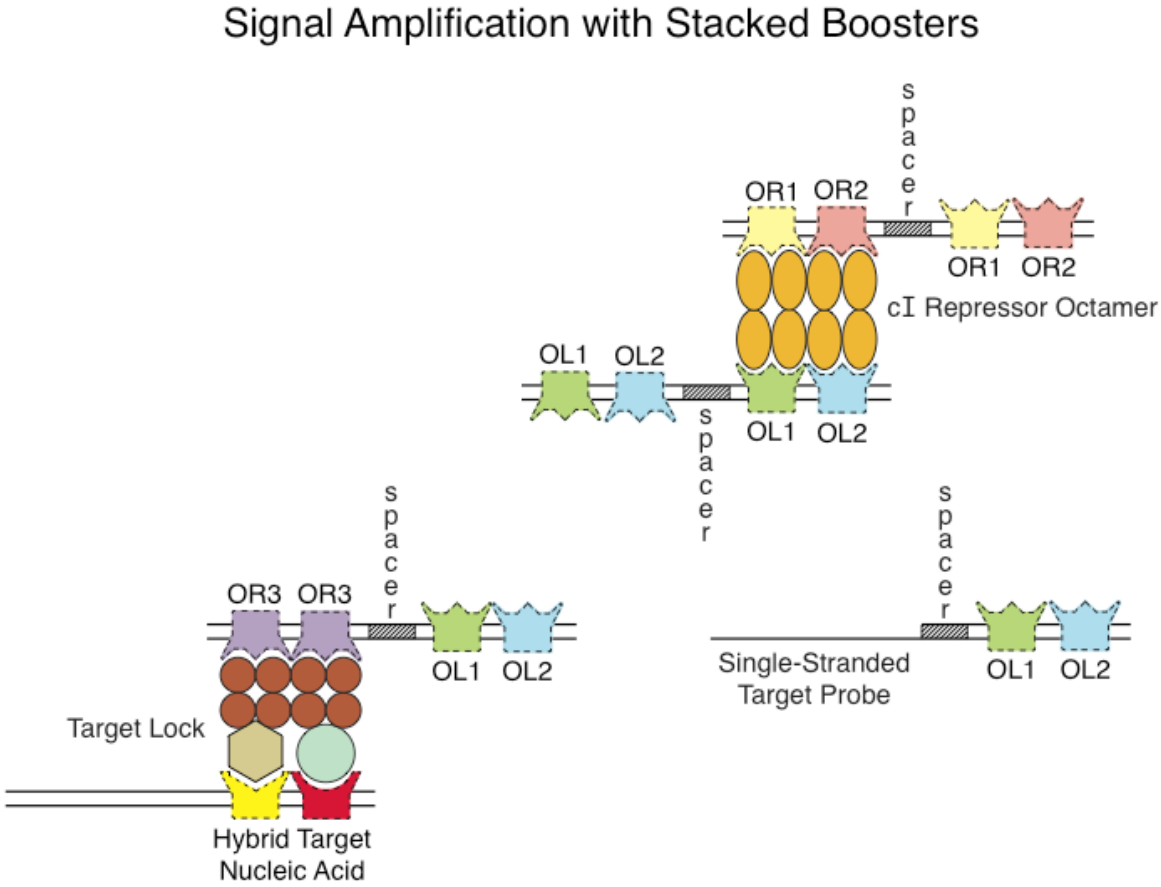



*Figure 10. Stacked Booster components for signal amplification of target recognition and target engagement.*

The natural spacing of the cI binding sites can be seen from the structure of the operators:

**Left Operator Sequences (LETTERS IN RED: OL1, OL2, OL3):**

```
          12345678901234567890123456          123
tatcaccgccagtggtatttatgtcaacaccgccagagataatttatcaccgcagatggtta
                  1234567          12345678901234567890123456
```

**Right Operator Sequences (LETTERS IN RED: OR3, OR2, OR1):**

```
          12345678901234567890123456          1234567
tatcaccgcaagggataaatatctaacaccgtgcgtgttgactatttacctctggcggtga
taa
                  123456          12345678901234567890123456
```

Note that the separation between centers of naturally occurring high affinity cI repressor binding sites is 24 base pairs. This is because the cI repressor protein dimer does not bind pseudosymmetrically about an axis perpendicular to the helix. The bound cI repressor dimer "leans" toward the other bound dimer on the adjacent high affinity binding site. The 31 base pair spacing is optimal for the cro protein because it does bind pseudosymmetrically about an axis perpendicular to the helix. It is possible to build a Booster that is laminated with cI repressor proteins that assemble into octamers with a 24 base pair separation between the central bases of paired sites (7 base pair separation between paired sites). The basic building block for the stacked Booster is shown in Figure 10. As long as the spacers are constructed so that they place the first OL1-OL2 on the other side of the helix from the second OL1-OL2 and the first OR2-OR1 on the other side of the helix from the second OR2-OR1, then the addition of cI repressor will cause the nucleic acid fragments to stack. By manipulating the proteins and protein-binding sites in the Boosters, many different Booster geometries and reporting structures can be created for signal amplification.

There is also considerable flexibility in the type of signal that Boosters can carry. Perhaps the simplest of these is an incorporated fluorophore at the ends of the nucleic acid portion of the signal amplification polymer. In general, fluorophores should be placed 5 bases or more from the end of the Lock binding site in order to put phosphates away from the protein and on the opposite side from the protein. Ten nucleotides plus 3 or 4 bases from the middle of the Lock component binding sequence puts the indicators roughly 120 degrees off of each other. To avoid quenching, fluorophores should be placed on ends that are on opposite sides of the helix. Although this is the easiest and can be done by researchers with use licenses for labeled nucleic acids, there is some question as to which placements of fluorophores in the nucleic acid trigger infringement of the Enzo patents. For this reason, the safest (from our patent standpoint without a complete examination by patent counsel of what positions on the nucleic acids are "free and clear", if any) placement of fluorophores is on proteins that are incorporated in specific geometries into the Booster by binding to sites in the Booster. All incorporation types have geometries that are optimal for FRET or other fluorescence transfer methods. Multiple fluorophores incorporated into a Booster (on the nucleic acid or on proteins bound to the nucleic acid, assemble a fixed ratio of fluorophores to signal the detector. Booster Lock components also can be generated that carry an enzyme that produces a color change (e.g., HRP) and be used to detect nucleic acids in the same manner and formats as antibody-based diagnostics.

It is possible to create a polymer of a fixed size and composition but this is somewhat more laborious as it requires pre-assembly and isolation prior to use. One method uses six single-stranded nucleic acids (listed below as Ia, Ib, Ic, Id, Ie, and If) that are combined in pairs to produce the three initial hybrids (listed below as IIa, IIb, and IIc). Each of these hybrids can combine with the other two remaining hybrids as each hybrid has one single stranded arm extending out on each side (e.g., item IIa in Figure 12 has the extensions "A" and "B'"). Each hybrid cannot polymerize with itself. Each production cycle consists of three nucleic acid hybrids being combined in pairs to produce three new nucleic acid hybrids that are twice as long. The starting and ending hybrids share the same connective property of being able to combine with the other two remaining hybrids in their respective starting or ending set. Any number of cycles may be run, resulting in polymer lengths of the power of 2 of the initial hybrids. (The first two cycles are shown in Figure 12; items IIa, IIb, and IIc producing items IIIa, IIIb, and IIIc, items IIIa, IIIb, and IIIc producing items IVa, IVb, and IVc.) The resultant polymers may be capped with single stranded nucleic acids or hairpins that match the hybrid extensions (as listed below, Va, Vb, and Vc respectively cap IVa, IVb, and IVc).

In the example case, spacers of 6 nucleic acid bases are used to present the protein binding sites approximately two thirds of a rotation around the polymer axis from the previous protein binding site. Spacers may be used but are not necessary for polymer formation. It is also possible to fix the assembly of parts by creating a covalent bond between the Booster Nucleic Acids and the Booster binding proteins.

The fragment name, size and compositions for the fragments used in Figure 11 are listed below (5' to 3'):

```
Fragment Ia  46   ATCTA TCACC gCAAg ggATA AATTg TCAAC ACCgC CAgAg ATAAT T
Fragment Ib  46   AATAA CCATC TgCgg TgATA AATAA TTATC TCTgg CggTg TTgAC A
Fragment Ic  46   ATTTA TCACC gCAgA TggTT ATTTA gTATC ACCgC CAgTg gTATT T
Fragment Id  46   AgTCA ACACg CACgg TgTTA gATAA ATACC ACTgg CggTg ATACT A
Fragment Ie  46   ATCTA ACACC gTgCg TgTTg ACTAA TTATC ACCgC CAgAg gTAAA A
Fragment If  46   ATTTA TCCCT TgCgg TgATA gATTT TTACC TCTgg CggTg ATAAT T
Fragment Va  23   ATTTA TCACC gCAgA TggTT ATT
Fragment Vb  23   ATCTA ACACC gTgCg TgTTg ACT
Fragment Vc  23   ATCTA TCACC gCAAg ggATA AAT
```

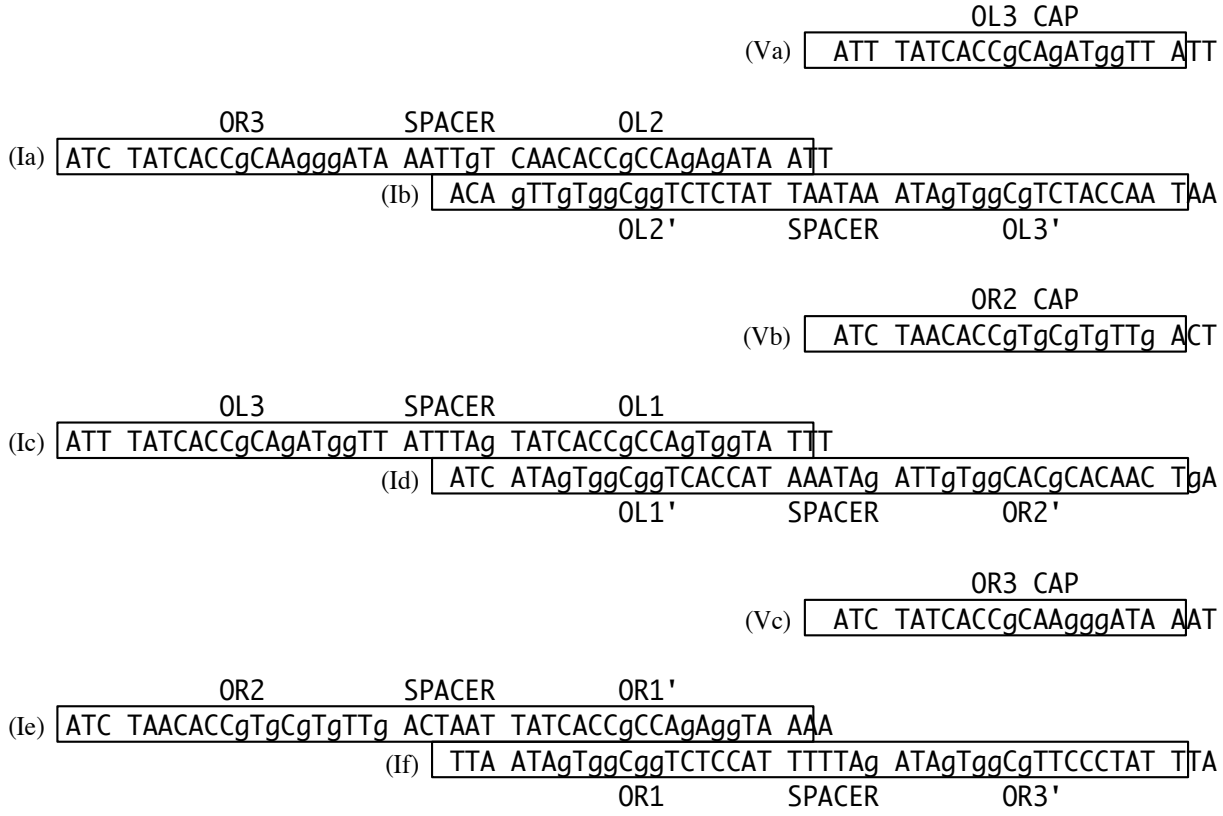



*Figure 11. Components for controlled assembly of Booster.*

A detailed protocol for making fixed length Boosters is available but not included here, however the steps are illustrated graphically in Figure 12. Polymer terminating and hairpin nucleic acids can be incorporated that allow the polymer to be capped and fixed size polymer "rafts" to be made and isolated.

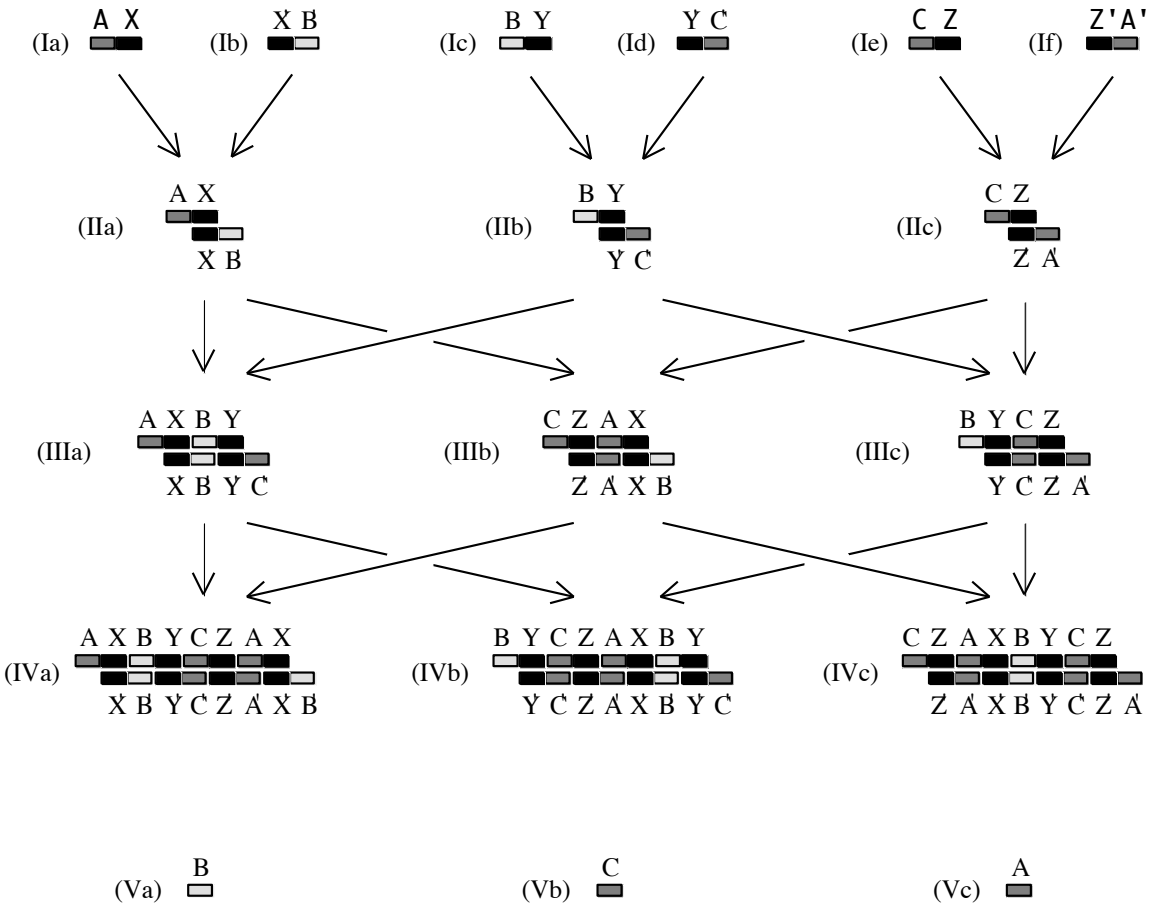



*Figure 12. Controlled assembly of Booster polymer.*

## 7. The Development and Testing of Molecular Lock Components

Molecular Lock components must immobilize the target and create a readable signature that is unique to the target and not found in the background genome(s). Molecular Locks must bind to their target cooperatively. This is tested by running a gel retardation assay with the (1) target present, (2) the target absent (control) and (3) the target present but with portions of the target blocked (site specificity). The target then is doped into a sample and detected.

We will use the example of the HIV long terminal repeat (HIV LTR) to discuss limits of detection. At the beginning of the test, the sample is lysed, fragmented non-specifically and Molecular Lock components (in micromolar concentrations) that bracket and protect a synthesized LTR are added. At a micromolar concentration of a p50-cro lock component described above, all of the Rel binding sites in a sample (viral and cellular) are bound by the p50-cro derived Molecular Lock components shown in Figures 5-7. This result is consistent with values seen for the wild type Rel-binding domains. See Figure 2 in: http://www.jbc.org/cgi/content/full/275/32/24392. This means that all of the Rel binding sites, irrespective of orientation relative to one another can expect to be occupied by Rel binding lock protein. All HIV targets in the sample will be engaged, bracketed and protected at this concentration by Molecular Lock components ready to be picked up by the anchor.

What does it look like when a p50/rel lock component binds to the target? In a gel this looks like a shift in the position of where the target migrates in the gel. Lane 3 in the gel retardation assay in Figure 13 below shows the shift that occurs when a lock component containing a p50/rel-binding domain hops on the target (the HIV LTR). Lane 2 is the control (no target). In Lane 3, there are two bands that show one protein (lower shifted band) and two proteins (higher shifted band) engaging the target (HIV LTR). The gel bands (seen by film exposed to radioactive P32 added (kinased) onto the ends of the target DNA) are simply a means of graphically showing target migration and engagement in the sample.

Lanes 13 and 14 in Figure 13 show a target that has one and two linear booster fragments hybridized to it to show our ability to control the assembly of fixed size boosters that are hybridized to a target through a fragment that has one portion that binds to the target and one part that engages the Booster. Each addition of Booster fragment is precise and sequence specific. Lane 14 shows that each booster fragment added can be assembled in a linear and controlled fashion and add its unique fluorophores to the booster. This creates a molecular bar code. Lane 13 shows that proteins added to the Booster cause a shift in the target engaged Booster in this lane. These Booster proteins can be used to carry the signaling fluorophores in the Booster. In lane 13, small proteins that can carry fluorophores are added can carry its own that has one and two Booster fragment extensions are made (2 fragments). Lane 13 differs from lane 14 in that, in lane 13, a protein has been added to the booster that can carry a fluorescent signal. This is the basis for one of the Booster-assisted PCR products that challenge the Enzo position.

Lane 10 in Figure 13 shows the result when a Booster is hybridized directly to third Sp1 site in the HIV LTR (the third blue region in Figure 3: sequence ggggagtgg). In Figure 13, Lane 10, a Sp1 Lock protein is used to anchor a linear polymer directly to the target by hybridization (as in Figure 13, Lanes 13 and 14). This gel demonstrates some of the versatility of engagement of target sequences by lock proteins. Lock protein components (with great affinity and specificity) can: (1) engage the target directly without hybridization; (2) engage the target by hybridization alone; (3) engage the target by hybridization and stabilization with a Lock protein specific for the target-probe (part probe/part Booster-engaging) nucleic acid; (4) stabilize a Booster hybridized directly to the target; (5) assemble, stabilize and add additional barcoding versatility to the Booster; and (6) anchor the assembly to a surface for counting. Counting copies is as simple as counting the photons coming from fluorophores that are placed in the geometries in the Booster such that they do not quench each other. The Booster fluorophores can either be on the nucleic acid or on the proteins that are bound to the nucleic acid (as in Lanes 13 and Lane 10).

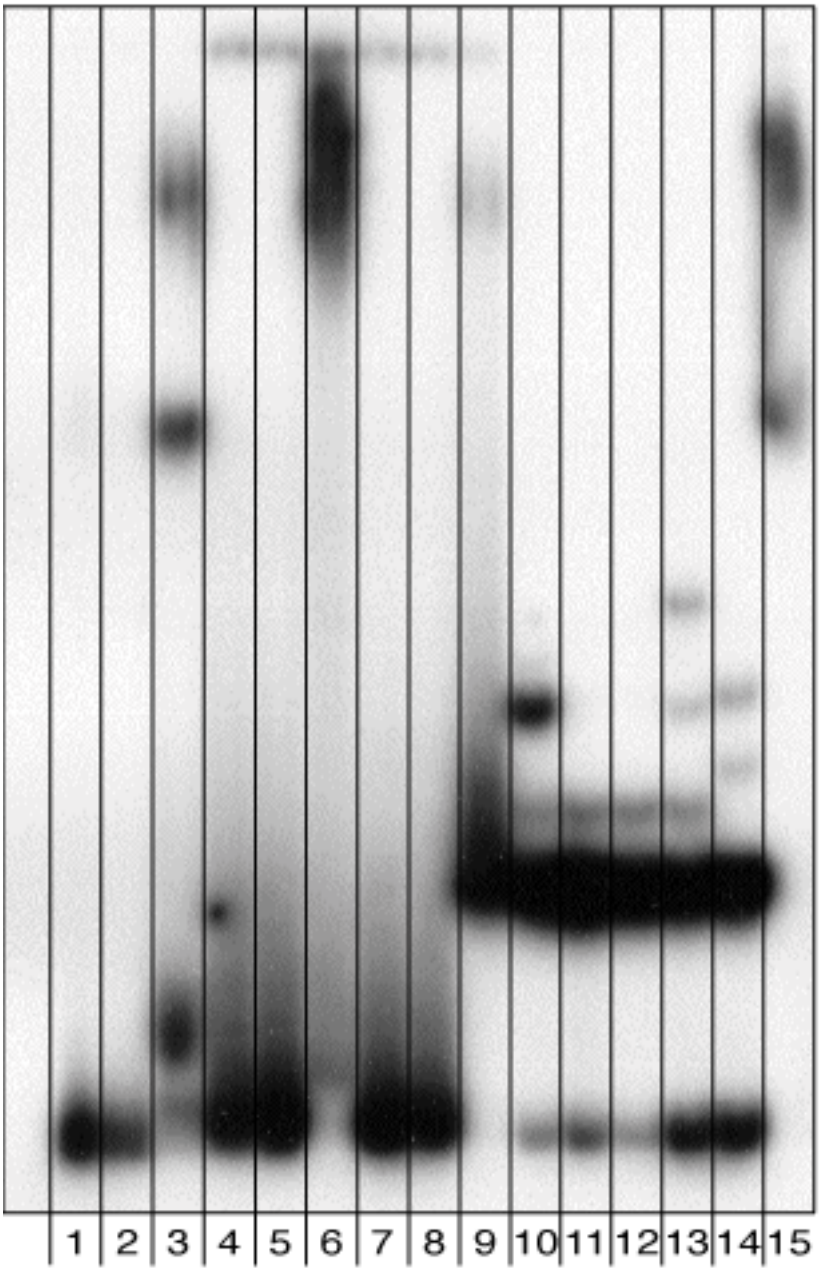



Figure 13. Gel retardation assay using a synthesized $P^{32}$-labelled HIV LTR, Lock components and Booster components.

Figure 14 below is a graphical depiction of the molecular assembly shown in the gel lanes in Figure 13. Although there are other ways to incorporate targets into Molecular Lock assemblies, these three engagements types form the basis for the initial reagent products, including those complementing and replacing PCR.

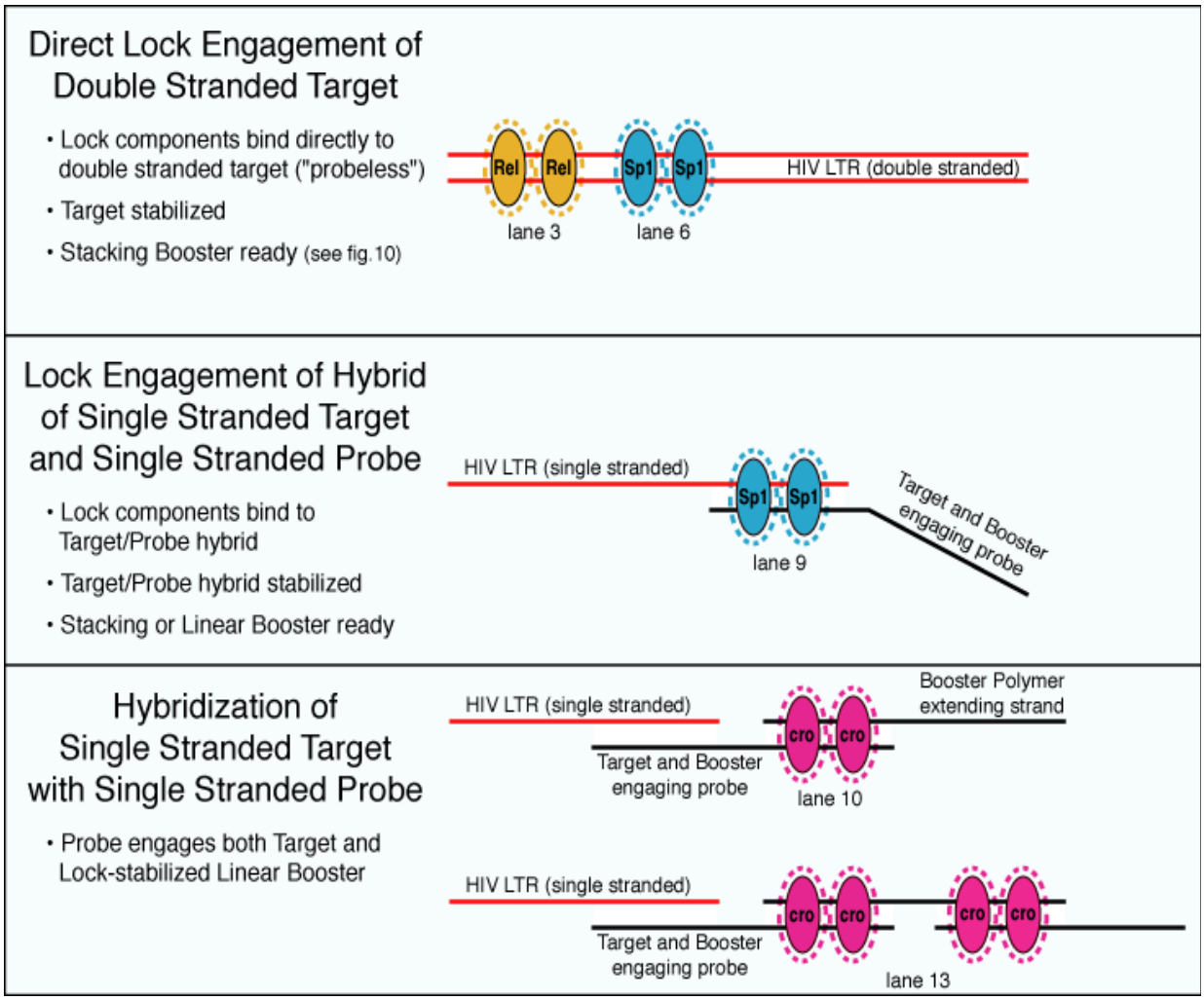



Figure 14. Three examples of how to engage a target (HIV-LTR): a graphical depiction of the molecular assembly of Lock components in the gel in Figure 13.

The target that a Molecular Lock engages can be "locked and dropped" for sequencing. Some Locks are configured to engage specific geometries that are only contained in the target sequence. Other locks are meant to simply engage one feature of the target in order to immobilize it so that other features of the target can be cross-checked and "read." In the first case, only the target is captured. In the second case, only the target is read. The second strategy is useful when there are low titer targets in highly viscous environments where it is desirable to throw a wide net for a target feature but only "read" those targets when one or more additional features are present.


Figure 15. Gel retardation assay using Silver stain.

One of the advantages of the high binding affinity of Lock components is that an "optical event" can be created to signal that the sandwich is present. This can be done with silver stain (see Figure 15), gold particles, etc. This means that it is not necessary to use fluorophores to signal the presence of low copy number targets if the polymerization of the Booster is configured properly to incorporate opaque materials. This feature may enable us to optically "read" low copy number target rafts directly and precisely as digital events.

In the example of detection of the HIV-LTR that we are discussing, the target (HIV LTR) is bound by a Rel binding domain of the Lock component. The Lock component also contains a portion of the cro protein and so can couple the HIV LTR to the anchor using the anchor binding domains of the cro protein. The native unligated cro dimer has an extremely high binding affinity for its target DNA, operator sites in the bacteriophage lambda. This has been measured to be 6.3 x $10^{-14}$M (Jana et al., 1997). Please see Appendix C for a copy of this paper. Normally dimer assembly limits DNA binding. Very few monomers assemble into dimers even at high concentrations. However, in this case the cro dimer is already assembled by virtue of the p50 being bound to the Rel binding sites in the HIV LTR. The binding of the p50 domain to the target shifts the cro monomer-dimer equilibrium to present a high binding affinity dimer to the anchor for binding. The cro protein dimer, when ligated to the Rel binding domain (also assembled as a dimer) to make a Lock component, has a binding affinity of approximately $10^{-11}$M for its target sites in the anchor. This is reduced from native cro because of constraints imposed by the ligation to the Rel binding domain. The cro protein can be expected to anchor any copy of the target HIV that has a lock component bound and is presented with an anchor. The Molecular Locks bound to the target in tandem bind the anchor with the highest binding affinities known in biological systems and have great thermal stability.

There are several ways that temperature affects the system. The most important effects of increased temperature are that the fragmented DNA hybrids become less stable, the Lock components diffuse to their targets faster and the Lock components' domains sample more configurations than at a lower temperature. When two Lock component monomers bind to the target, both the target and the dimer are stabilized. When more than one of these events occur side by side (Molecular Parataxis), the target and the Lock component dimers that stabilize them attain extremely high stability and become insensitive to the effects of temperature. The target and target-bound locks in a bridging configuration, and anchor "box" shown in Figure 10, is an extremely stable configuration. Once the target is locked, increased temperature can be used to destabilize all interactions in the sample that are not specific to target binding and reporting. The "box" can only be disassembled by a chemical change that impacts the nucleic acid backbone and the highly charged atoms that it contains or by cutting the anchor to uncouple the coopertivity.

The main barrier to the anchor picking up the target is the fact that these molecules live in a highly viscous environment with a low Reynolds number. Any bracketed target must diffuse to the target. The immobilization of the target is diffusion limited but the test is configured so that the target is moved past an anchor rich area ensuring the binding of the target to the immobilization area. This is accomplished by creating a gauntlet of anchor binding areas that "filter" the sample. Increases in temperature act to increase the diffusion of Lock components to their targets.

## 8. Limits of Detection

The limit of detection (DL) is an estimate of the concentration of target below which we may not detect target or at which we are fairly certain that we will detect target (usually set at 99% confidence). It is usually determined on an uncomplicated or "clean" sample. For the Molecular Locks this is determined by the concentration of the components. In a dilution series we have been able to reliably see what we believe to be less than ten copies of the target using a booster with fluorophores. The Boosters can be made to carry hundreds of fluorophores in a configuration where there is no significant quenching so single copies are readily counted.

An Instrument Detection Limit (IDL) is the lowest limit that an instrument can detect a signal. Although common photon detectors are highly efficient (http://www.andor.com/biology/?app=67, http://www.andor.com/biology/?app=95 ), the detector in our device does not have to have single photon detection. A single attached booster contains a sufficient number of fluorophores to signal most detectors that an immobilized target is present. Our proprietary device is configured so that the captured target can be interrogated independent of the sample contents in the background.

A Method Detection Limit (MDL) is based on samples which have gone through the entire sample preparation Scheme prior to analysis. We believe this to be about 10%.

The Practical Quantitation Limit (PQL) is normally 3-10 times MDL and is generally understood as the lowest concentration that can be accurately measured (as compared to detected). This means that the PQL is expected to be 30-100 copies.

These values will be higher for systems that do not use the handheld device. See Appendix D and E for information on sensitivity of PCR and bDNA tests, respectively. However, it is currently not established how many actual sample copies are lost when a PCR prep (Boom prep) is done. See Appendix F for the Boom prep description. The values for both PCR and bDNA detection limits are derived from a comparison with doped (not extracted) copies. When we met with the scientists from one major PCR company, we raised this issue of sample copy loss and were told that they are aware of it and it is substantial. We know that as much as 99% of the targets per sample can remain on the silica. Greater than 99% of the targets are commonly degraded (depending on the test) prior to adsorption, especially when the target is RNA. This would mean that less than one in ten thousand copies are actually reaching the test in a Boom preparation in the best of circumstances. The Molecular Lock test is the only technology that can establish this number. If we can have third party contractors document these sample losses, we could establish and safely make the commercial assertion that a Molecular Lock test is the only way to reliably detect extremely low copy number targets.

We believe that the insert to our Molecular Lock products will have a description similar to that of bDNA (see Appendix E):

**Method**

Capture by Molecular Locks and Booster polymers
Booster components hybridized to alkaline phosphatase labeled or fluorescent-labeled proteins
Chemiluminescent or fluorescent detection of Booster molecules
Reportable range: 30 to 500,000 copies/mL
Specificity: All HIV subtypes
Minimum detectable change in viral load: 0.51log10 (across range of assay)
Synonyms: human immunodeficiency virus (HIV) viral load
Booster detection levels are 30-100 copies/mL. Reportable range 30 to 500,000 copies/mL.
Minimum detectable change in viral load .51 log10 (across range of assay).

## 9. References

(Other than those in the appendices.)

Molecular Parataxis methods of manufacture and compositions are patented and patent-pending. For more information, see US patent 5,871,902.


Albright, R. A. and Matthews, B. W. (1998) J. Mol. Biol. 280, 137-51.
Baeuerle, P.A. and Baltimore, D. (1989) Genes Dev. 3, 1689–98.
Brownell,E., et al. (1989) Oncogene 4 (7), 935-42.
Dahlberg,J.E. and Blattner,F.R. (1975) Nucleic Acids Res. 2 (9), 1441-58.
Dodd, I. et al., (2004) Genes and Dev. 18, 344-54.
Ghosh, S., et al. (1990) Cell 62, 1019–29.
Ghosh, G., et al. (1995) Nature 373, 303-10.
Higgins, D., et al. (1990) J. Virol. 64 (8), 4016-20.
Huang, D. B., et al. (2001) Structure 9 (8), 669-78.
Humayun,Z., et al. (1977) J. Mol. Biol. 112 (2), 265-77.
Johnson, A. et al. (1978) PNAS 75(4), 1783-87.
Kieran, M., et al. (1990) Cell 62, 1007–18.
Maniatis, T. and Ptashne M. (1973) Nature 246 (5429), 133-36.
Maniatis, T. et al., (1975) PNAS 72 (3), 1184-88.
Meyer,B.J.,et al. (1980) J. Mol. Biol. 139 (2), 163-94.
Montanyo, M.A. et al. (1996) PNAS 93, 12376-12381.
Narayan, et al. (1997) J. B. C. 272: 7801-9.
Roberts,T.M., et al. (1977) Nature 270 (5634), 274-75.
Sauer, RT and Anderegg, R. (1978) Biochem. 17 (6):1092-100.
Schmid, R.M., et al. (1991) Nature 352, 733–6.
Schwarz,E., et al. (1978) Nature 272 (5652), 410-14.
Sen,R. and Baltimore, D. (1986) Cell 46 (5), 705-17.
Sica, A. et al. (1992) PNAS 89, 1740-44.
Takeda, Y., et al. (1977) J. Mol. Biol. 127, 6177-83.
Takeda, Y., et al. (1986) J. Mol. Biol. 261, 8608-16.
Takeda, Y. et al. (1987) J. Mol. Biol. 196, 149-58.
Toledano,M.B. et al. (1993) Mol. Cell Biol. 13(2), 852-60.
Van Lint C. et al. (1997) J. Virol. 71 (8), 6113-6127
Van Opijnen, T., et al. (2004) J. Virol. 78(7), 3675-3683.

## Appendix A: HIV-Lock™ Structure Background Information

**Cro Repressor Protein Amino Acid Sequence (Bacteriophage Lambda, Recombinant)**
http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=215104

BOLD LETTERS IN RED= HIV-Lock™ FRAGMENT

```
CDS                 38041..38241
                    /note="cro (antirepressor; also tof;66)"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="AAA96582.1"
                    /db_xref="GI:215148"

/translation="MEQRITLKDYAMRFGQTKTAKDLGVYQSAINKAIHAGRKIFLTI
                    NADGSVYAEEVKPFPSNKKTTA"
```

**Cro Repressor Protein DNA Sequence (Bacteriophage Lambda, Recombinant)**
http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=215104

BOLD LETTERS IN RED= HIV-Lock™ FRAGMENT

```
  REFERENCE 16 (bases 38041 to 38241)
  AUTHORS   Roberts,T.M., Shimatake,H., Brady,C. and Rosenberg,M.
  TITLE     Sequence of Cro gene of bacteriophage lambda
  JOURNAL   Nature 270 (5634), 274-275 (1977)

    38041 atggaacaac gcataaccct gaaagattat gcaatgcgct ttgggcaaac caagacagct
    38101 aaagatctcg gcgtatatca aagcgcgatc aacaaggcca ttcatgcagg ccgaaagatt
    38161 tttttaacta taaacgctga tggaagcgtt tatgcggaag aggtaaagcc cttcccgagt
    38221 aacaaaaaaa caacagcata a
```

## Appendix A: HIV-Lock™ Structure Background Information

**Cro Repressor Protein Crystal Structure (Bacteriophage Lambda, Recombinant)**

Cro Bound To A 19 Bp DNA Duplex With Single-Base 5'-Overhang Consensus Operator
Exp. Method: X-ray Diffraction
Classification: Complex (Transcription Regulation/DNA)
Source: Bacteriophage lambda
http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?name=bacteriophage+lambda
Primary Citation: Albright, R. A., Matthews, B. W.: Crystal structure of lambda-Cro bound to a consensus operator at 3.0 A resolution. J Mol Biol 280 pp. 137 (1998).
http://www.rcsb.org/pdb/cgi/explore.cgi?job=summary&pdbId=6CRO&page=40&pid=32441116972417

```
   1 EQRITLKDYA MRFGQTKTAK DLGVYQSAIN KAIHAGRKIF LTINADGSVY AEEVKPFPSN
       EEEHHHHH HHHTHHHHHH HHTS HHHHH HHHHHT  EE EEE TT  EE EEE   SS
```

http://www.rcsb.org/pdb/cgi/explore.cgi?job=chains&pdbId=6CRO&page=40&pid=32441116972417#6CRO:A

## Appendix A: HIV-Lock™ Structure Background Information

**p105 Amino Acid Sequence (Mus musculus, Recombinant)**
http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=6679043

```
LOCUS       NM_008689               3892 bp    mRNA    linear   ROD 23-MAY-
2005
DEFINITION  Mus musculus nuclear factor of kappa light chain gene enhancer in
            B-cells 1, p105 (Nfkb1), mRNA.
ACCESSION   NM_008689
SOURCE      Mus musculus (house mouse)
```

***ITALIC LETTERS = HIV-Lock™ FRAGMENTS***
ITALIC LETTERS IN NON-BOLD RED = HIV-Lock™-1A and B FRAGMENT
ITALIC LETTERS IN BOLD RED = N-terminus of HIV-Lock™-1B
*ITALIC LETTERS IN BOLD BLACK = HIV-Lock™-1*-1 and 2 FRAGMENTS*

/translation="***MADDDPYGTGQMFHLNTALTHSIFNAELYSPEIPLSTDG****PYLQI*
*LEQPKQRGFRFRYVCEGPSHGGLPGASSEKNKKSYPQVKICNYVGPAKVIVQLVTNGK*
*NIHLHAHSLVGKHCEDGVCTVTAGPKDMVVGFANLGILHVTKKKVFETLEARMTEACI*
*RGYNPGLLVHSDLAYLQAEGGGDRQLTDREKEIIRQAAVQQTKEMDLSVVRLMFTAFL*
*PDSTGSFTRRLEPVVSDAIYDSKAPNASNLKIVRMDRTAGCVTGGEEIYLLCDKVQKD*
*DIQIRFYEEEENGGVWEGFGDFSPTDVHRQFAIVFKTPKYKDVNITKPASVFVQLRRK*
*SDLETSEPKPFLYYPEIKDKEE****VQRKRQKLMPNFSDSFGGGSGAGAGGGG***MFGSGGGG
GSTGSPGPGYGYSNYGFPPYGGITFHPGVTKSNAGVTHGTINTKFKNGPKDCAKSDDE
ESLTLPEKETEGEGPSLPMACTKTEPIALASTMEDKEQDMGFQDNLFLEKALQLARRH
ANALFDYAVTGDVKMLLAVQRHLTAVQDENGDSVLHLAIIHLHAQLVRDLLEVTSGLI
SDDIINMRNDLYQTPLHLAVITKQEDVVEDLLRVGADLSLLDRWGNSVLHLAAKEGHD
RILSILLKSRKAAPLIDHPNGEGLNAIHIAVMSNSLPCLLLLVAAGAEVNAQEQKSGR
TPLHLAVEYDNISLAGCLLLEGDAHVDSTTYDGTTPLHIAAGRGSTRLAALLKAAGAD
PLVENFEPLYDLDDSWEKAGEDEGVVPGTTPLDMAANWQVFDILNGKPYEPVFTSDDI
LPQGDMKQLTEDTRLQLCKLLEIPDPDKNWATLAQKLGLGILNNAFRLSPAPSKTLMD
NYEVSGGTIKELMEALQQMGYTEAIEVIQAAFRTPATTASSPVTTAQVHCLPLSSSST
RQHIDELRDSDSVCDSGVETSFRKLSFTESLTGDSPLLSLNKMPHGYGQEGPIEGKI"

## Appendix A: HIV-Lock™ Structure Background Information

**P105 cDNA Sequence (Mus musculus, Recombinant)**
http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=6679043

***ITALIC LETTERS = p105***
***ITALIC LETTERS IN RED**, BOLD RED, BOLD BLACK = **HIV-Lock™ FRAGMENTS***
ITALIC LETTERS IN NON-BOLD RED = HIV-Lock™-1A and B FRAGMENT
ITALIC LETTERS IN BOLD RED = N-terminus of HIV-Lock™-1B
*ITALIC LETTERS IN BOLD BLACK = HIV-Lock™-1*-1 and 2 FRAGMENTS*

```
   1 agcggccgcc gcgggcgcgc tctagcagcg caggccggag ctcagggccc cgcgcgcccg
  61 gcccgccccg cgcttctccg cccgcgccgc agccatggcg cgccgctgag ccgcccgccc
 121 gcccgcccgc gccccgaccc ggctcggctc ccgccggtcc gcgccgctcc gcagcggagc
 181 ccgcaggcga ggagaggccg cgcgcatctc cagggtaccc tcagaggcca gaagagggtg
 241 tcagagccct tgtaactgga gtttgacggt cgtgagctgc gcatcttcac catggcagac
 301 gatgatccct acggaactgg gcaaatgttt catttgaaca ctgctttgac tcactcaata
 361 tttaatgcag aattatattc accagaaata ccactgtcaa cagatggccc ataccttcaa
 421 atattagagc aaccaaaaca gaggggattt cgattccgct atgtgtgtga aggcccatca
 481 cacggagggc ttccgggagc ctctagtgag aagaacaaga aatcctaccc acaggtcaaa
 541 atttgcaact atgtggggcc tgcaaaggtt atcgttcagt tggtcacaaa tggaaaaaac
 601 atccacctgc acgcccacag cctggtgggc aagcactgtg aggacggggt atgcaccgta
 661 acagcaggac ccaaggacat ggtggttggc tttgcaaacc tgggaatact tcatgtgact
 721 aagaaaaagg tatttgaaac actggaagca cggatgacag aggcgtgtat taggggctat
 781 aatcctggac ttctggtgca ttctgacctt gcctatctac aagcagaagg cggaggagac
 841 cggcaactca cagacagaga gaaggagatc atccgccagg cagccgtgca gcagaccaag
 901 gagatggacc tgagcgtggt gcgcctcatg ttcacagcct tcctccctga cagcactggc
 961 agcttcactc ggagactgga gcctgtggtg tcagacgcca tctatgatag caaagccccg
1021 aatgcatcca acctgaaaat cgtgagaatg gacagaacag caggatgtgt gacgggaggg
1081 gaggagattt accttctctg tgacaaggtt cagaaagatg acatccagat tcggttttat
1141 gaagaggaag aaaatggcgg agtttgggaa ggatttgggg acttttcccc cacggatgtt
1201 catagacagt ttgccattgt cttcaaaacg ccaaagtata aggatgtcaa cattacaaag
1261 ccagcttccg tgtttgttca gcttcggagg aaatcagacc tggaaactag tgaaccgaaa
1321 ccctttctct actaccctga aatcaaagac aaagaggaag tgcaaaggaa acgccagaag
1381 cttatgccga acttctcgga cagcttcggc ggcggcagtg gagcgggagc cggtggtgga
1441 ggcatgttcg gtagtggcgg tggcggaggg agtaccggaa gccctggccc agggtatggc
1501 tactcgaact acggatttcc tccctacggt gggattacat tccatcccgg agtcacgaaa
1561 tccaacgcag gggtcaccca tggcaccata aacaccaaat ttaaaaatgg ccctaaagat
1621 tgtgccaaga gtgatgacga ggagagtctg actctccctg agaaggaaac tgaaggtgaa
1681 gggcccagcc tgcccatggc ctgcaccaag acggaaccca tcgccttggc atccaccatg
1741 gaagacaagg agcaggacat gggatttcag gataacctct ttctcgagaa ggctctgcag
1801 ctcgccaggc gacacgccaa cgccctttc gactacgcag tgacgggga tgtgaagatg
1861 ttgctggccg tgcaacgcca tctcaccgcc gtgcaggatg agaatgggga cagtgtctta
1921 cacttagcca tcatccacct ccacgctcag ctcgtgaggg atctgctgga agtcacatct
1981 ggtttgatct ctgatgacat catcaacatg agaaatgacc tgtatcagac acctctgcac
2041 ttggccgtga tcaccaagca ggaagatgta gtagaggatt tgctgagggt tggggctgac
2101 ctgagccttc tggaccgctg gggcaactct gtcctgcacc tagctgccaa agaaggacac
2161 gacagaatcc tcagcatcct gctcaagagc agaaaagcag cgcccttat cgaccacccc
2221 aatggggaag gtctaaatgc catccacata gctgtgatga gcaatagcct gccatgtctg
2281 ctgctgctgg tggctgccgg ggcagaagtc aatgctcagg agcagaagtc tgggcgcacg
2341 ccgctgcacc tggccgtgga gtacgacaac atctccttgg ctggctgcct gcttctggag
2401 ggtgatgccc acgtggacag taccacctat gatgggacta cacctctgca tatagcggcc
2461 ggaagagggt ccaccagact ggcagctctt ctcaaagcag caggagcaga ccccctggtg
2521 gagaactttg agcctctcta tgacctggac gactcttggg agaaggctgg agaagatgag
2581 ggagtggtgc caggtaccac acccctggac atggctgcca actggcaggt atttgacata
2641 ctaaatggga aaccgtatga gcctgtgttc acatctgatg atatactacc acaaggggac
2701 atgaagcagc tgacagaaga cacgaggcta caactctgca aactgctgga aattcctgat
2761 ccagacaaaa actgggccac tctggcacag aagttgggtc tggggatatt gaacaatgcc
```

## Appendix A: HIV-Lock™ Structure Background Information

**P105 cDNA Sequence (Mus musculus, Recombinant), continued**
http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=6679043

*ITALIC LETTERS= p105*

*2821 ttccggctga gtcctgctcc ttctaaaact ctcatggaca actatgaggt ctctgggggt*
*2881 accatcaaag agctgatgga ggccctgcaa cagatgggct acacagaggc cattgaagtg*
*2941 atccaggcag ccttccgcac cccggcaacc acagcctcca gccccgtgac cactgctcag*
*3001 gtccactgtc tgcctctctc gtcttcctcc acgaggcagc acatagatga actccgggat*
*3061 agtgacagcg tctgtgacag tggtgtggag acatccttcc gcaaactcag ctttacagag*
*3121 tctcttactg gagacagccc actgctatct ctgaacaaaa tgccccacgg ttatgggcag*
*3181 gaaggaccta ttgaaggcaa aatt*tagcct gctggccgtt cccccacact gtgtaaacca
3241 aagccctgac agtccattgc atcgtcccaa aggaggaagg caaagcgaat ccaaaggtgc
3301 tggagaatcg ccggcctgca gggtcactcg atttcattca aggccttccg aatttggcgt
3361 ccttcttggt tctgaaatga aatgtagttg ccacgcacag acggtgtcta gcaatcatgg
3421 cgctcgctcg ctcagctgca ctctatggct caggtgcagt gtcttgagct ttctctgctg
3481 ctactggatc acatttgctt tgtgttgtta ctgctgtccc tccgctgggt tcctgctgtc
3541 attaaaaggt gtcgctgtcc ccacccggtg tcctttctag ccatctactg taagttgtgc
3601 attcaaatta agattaagga aaaacatatt tttaaatgag taccttgatg cgcaataaaa
3661 aaaaagacat ttctttttt aatgtggttt atctgtgatt taaaaataaa aaacacatga
3721 acttatcaat atttaaaaca tgctacaatc agtgntgaaa atagtatttt ccccgtttta

## Appendix A: HIV-Lock™ Structure Background Information

**p50 Homodimer Bound to Kb Site Crystal Structure (Mus musculus, Recombinant)**
*

P50 Subunit Bound to Engineered Kb Site, DNA (5'-D(Tgagaattccc)-3'
Exp. Method: X-ray Diffraction
Classification: Complex (Transcription Factor/DNA) Source: Mus musculus
http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?name=Mus+musculus
Primary Citation: Ghosh, G., van Duyne, G., Ghosh, S., Sigler, P. B.:
Structure of NF-kappa B p50 homodimer bound to a kappa B site. Nature 373 pp. 303 (1995).
http://www.rcsb.org/pdb/cgi/explore.cgi?job=summary&pdbId=1NFK&page=20&pid=260641116970571
http://www.rcsb.org/pdb/cgi/explore.cgi?job=chains&pdbId=1NFK&page=20&pid=260641116970571#1NFK:A

```
   1 GPYLQILEQP KQRGFRFRYV CEGPSHGGLP GASSEKNKKS YPQVKICNYV
       EEEESS B  SSS   EEG GG   SS  E ETTTBTTB   EEEEEEES

  51 GPAKVIVQLV TNGKNIHLHA HSLVGKHCED GVCTVTAGPK DMVVGFANLG
     SS EEEEEEE  SSSS  B S SEEESTTEET TEEEEE  SS   EEEE S E

 101 ILHVTKKKVF ETLEARMTEA CIRGYNPGLL VHSDLAYLQA EGGGDRQLTD
     EEE  STTHH HHHHHHHHHH HHTTTTIIII I TT S     SSSS     H

 151 REKEIIRQAA VQQTKEMDLS VVRLMFTAFL PDSTGSFTRR LEPVVSDAIY
     HHHHHHHHHH HHHHHT  SS EEEEEEEEEE E SSS B EE    EE   EE

 201 DSKAPNASNL KIVRMDRTAG CVTGGEEIYL LCDKVQKDDI QIRFYEEEEN
      TTTTTS     B EES S E ESS   EEEE EES   TTTE EEEEEE   S

 251 GGVWEGFGDF SPTDVHRQFA IVFKTPKYKD VNITKPASVF VQLRRKSDLE
     S  EEEE B   GGGBTTTSE EEEE   TTS S   SSEEEE EEEEETTT

 301 TSEPKPFLYY PEIKDKEEME QRITLKDYAM RFGQTKTAKD LGVYQSAINK
     B   EEEEEE             EEEHHHHHH HHTHHHHHHH HTS HHHHHH

 351 AIHAGRKIFL TINADGSVYA EEVKPFPSNK KTTA
     HHHHT  EE EEE TT  EE EEE   SS
```

## Appendix A: HIV-Lock™ Structure Background Information

**C-REL Transcription Factor Amino Acid Sequence**

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=X75042.1

ITALIC LETTERS = HIV-Lock™ FRAGMENTS
ITALIC LETTERS IN NON-BOLD RED = HIV-Lock™-2 FRAGMENT
**ITALIC LETTERS IN BOLD RED = N-terminal extension to p50 fragment in HIV-Lock™-1B**

```
LOCUS       HSRNAREL                2337 bp    mRNA    linear   PRI 18-APR-
2005
DEFINITION  H.sapiens rel proto-oncogene mRNA.
ACCESSION   X75042
VERSION     X75042.1  GI:402648
KEYWORDS    rel oncogene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 2337)
  AUTHORS   Brownell,E., Mittereder,N. and Rice,N.R.
  TITLE     A human rel proto-oncogene cDNA containing an Alu fragment as a
            potential coding exon
  JOURNAL   Oncogene 4 (7), 935-942 (1989)
   PUBMED   2666912
REFERENCE   2
  AUTHORS   Rice,N.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1993) N.R. Rice, Lab of Molec Virology &
            Carcinogenesis, BRI-Basic Research Program, NCI-Frederick Cancer
            Research Facility, PO Box B, Frederick, Maryland 21701, USA
FEATURES             Location/Qualifiers
     source          1..2337
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="1"
                     /cell_line="Daudi Burkitt lymphoma cell line"
     gene            1..2337
                     /gene="c-rel"
     CDS             178..2037
                     /gene="c-rel"
                     /codon_start=1
                     /protein_id="CAA52954.1"

/translation="MASGAYNPYIEIIEQPRQRGMRFRYKCEGRSAGSIPGEHSTDNN
RTYPSIQIMNYYGKGKVRITLVTKNDPYKPHPHDLVGKDCRDGYYEAEFGQERRPLFF
QNLGIRCVKKKEVKEAIITRIKAGINPFNVPEKQLNDIEDCDLNVVRLCFQVFLPDEH
GNLTTALPPVVSNPIYDNRAPNTAELRICRVNKNCGSVRGGDEIFLLCDKVQKDDIEV
RFVLNDWEAKGIFSQADVHRQVAIVFKTPPYCKAITEPVTVKMQLRRPSDQEVSESMD
FRYLPDEKDTYGNKAKKQKTTLLFQKLCQDHVETGFRHVDQDGLELLTSGDPPTLASQ
SAGITVNFPERPRPGLLGSIGEGRYFKKEPNLFSHDAVVREMPTGVSSQAESYYPSPG
PISSGLSHHASMAPLPSSSWSSVAHPTPRSGNTNPLSSFSTRTLPSNSQGIPPFLRIP
VGNDLNASNACIYNNADDIVGMEASSMPSADLYGISDPNMLSNCSVNMMTTSSDSMGE
TDNPRLLSMNLENPSCNSVLDPRDLRQLHQMSSSSMSAGANSNTTVFVSQSDAFEGSD
FSCADNSMINESGPSNSTNPNSHGFVQDSQYSGIGSMQNEQLSDSFPYEFFQV"
```

## Appendix A: HIV-Lock™ Structure Background Information

**C-REL Transcription Factor cDNA Sequences**

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=X75042.1

ITALIC LETTERS = HIV-Lock™ FRAGMENTS
ITALIC LETTERS IN NON-BOLD RED = HIV-Lock™-2 FRAGMENT
**ITALIC LETTERS IN BOLD RED = N-terminal extension to p50 fragment in HIV-Lock™-1B**

```
   1 cggaaggtgt gagccgcaaa cccagcggag ggcgggaaga aggaggaggc ctctagggtg
  61 ntcgggggac tgggggcccc gccggcagag gtccctcggc ctcctgactg actgactgcg
 121 gccgcctccg gccaggacgc tgggagctgc ctgcgggaag gtgcggggag cggagcc atg
 181 gcctccggtg cgtataaccc gtatatagag ataattgaac aacccaggca gaggggaatg
 241 cgttttagat acaaatgtga agggcgatca gcaggcagca ttccagggga gcacagcaca
 301 gacaacaacc gaacataccc ttctatccag attatgaact attatggaaa aggaaaagtg
 361 agaattacat tagtaacaaa gaatgaccca tataaacctc atcctcatga tttagttgga
 421 aaagactgca gagacggcta ctatgaagca gaatttggac aagaacgcag acctttgttt
 481 ttccaaaatt tgggtattcg atgtgtgaag aaaaaagaag taaaagaagc tattattaca
 541 agaataaagg caggaatcaa tccattcaat gtccctgaaa aacagctgaa tgatattgaa
 601 gattgtgacc tcaatgtggt gagactgtgt tttcaagttt ttctccctga tgaacatggt
 661 aatttgacga ctgctcttcc tcctgttgtc tcgaacccaa tttatgacaa ccgtgctcca
 721 aatactgcag aattaaggat ttgtcgtgta aacaagaatt gtggaagtgt cagaggagga
 781 gatgaaatat ttctactttg tgacaaagtt cagaaagatg acatagaagt tcgttttgtg
 841 ttgaacgatt gggaagcaaa aggcatcttt tcacaagctg atgtacaccg tcaagtagcc
 901 attgttttca aaactccacc atattgcaaa gctatcacag aacccgtaac agtaaaaatg
 961 cagttgcgga gaccttctga ccaggaagtt agtgaatcta tggattttag atatctgcca
1021 gatgaaaaag atacttacgg caataaagca aagaaacaaa agacaactct gcttttccag
1081 aaactgtgcc aggatcacgt agaaacaggg tttcgccatg ttgaccagga tggtcttgaa
1141 ctcctgacat caggtgatcc acccaccttg gcctcccaaa gtgctgggat tacagttaat
1201 tttcctgaga gaccaagacc tggtctcctc ggttcaattg gagaaggaag atacttcaaa
1261 aaagaaccaa acttgttttc tcatgatgca gttgtgagag aaatgcctac aggggtttca
1321 agtcaagcag aatcctacta tccctcacct gggcccatct caagtggatt gtcacatcat
1381 gcctcaatgg cacctctgcc ttcttcaagc tggtcatcag tggcccaccc caccccacgc
1441 tcaggcaata caaacccact gagtagtttt tcaacaagga cacttccttc taattcgcaa
1501 ggtatcccac cattcctgag aatacctgtt gggaatgatt taaatgcttc taatgcttgc
1561 atttacaaca atgccgatga catagtcgga atggaagcgt catccatgcc atcagcagat
1621 ttatatggta tttctgatcc caacatgctg tctaattgtt ctgtgaatat gatgacaacc
1681 agcagtgaca gcatgggaga gactgataat ccaagacttc tgagcatgaa tcttgaaaac
1741 ccctcatgta attcagtgtt agacccaaga gacttgagac agctccatca gatgtcctct
1801 tccagtatgt cagcaggcgc caattccaat actactgttt ttgtttcaca atcagatgca
1861 tttgagggat ctgacttcag ttgtgcagat aacagcatga taaatgagtc gggaccatca
1921 aacagtacta atccaaacag tcatggtttt gttcaagata gtcagtattc aggtattggc
1981 agtatgcaaa atgagcaatt gagtgactcc tttccatatg aatttttca agtataactt
2041 gcaagattta aatcctttta aatcttgata ccacctatat agatgcagca ttttgtattt
2101 gtctaactgg ggatataata ctatatttat actgtatata taatactgac tgagaatata
2161 atactgtatt tgagaatata aaaaactttt ttcagggaag aagcatacaa ctttggacat
2221 agcgaataca aaattggaag ctgtcataaa aagacaactc agaggccagg cgcaggngct
2281 cacacctgta atcctagcac tttgggaggc caaggcgggt ggatcacttg agaccag
```

## Appendix A: HIV-Lock™ Structure Background Information

**Crystal Structure Of C-Rel Bound To DNA**

C-Rel Proto-Oncogene Protein Bound to Rel Homology Region
Exp. Method: X-ray Diffraction
Classification: Transcription/DNA
Source: Gallus gallus
http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?name=Gallus+gallus
Primary Citation: Huang, D. B., Chen, Y. Q., Ruetsche, M., Phelps, C. B., Ghosh, G.: X-Ray Crystal Structure of Proto-Oncogene Product C-Rel Bound to the Cd28 Response Element of Il-2 Structure 9 pp. 669 (2001)

```
   1 PYIEIFEQPR QRGMRFRYKC EGRSAGSIPG EHSTDNNKTF PSIQILNYFG
      EEEESB B   SSB  EEGG G S      B SS  SSS     EEEEES  S

  51 KVKIRTTLVT KNEPYKPHPH DLVGKDCRDG YYEAEFGPER RVLSFQNLGI
      EEEEEEEE  SSSS  B SS EEESTTEETT EEEEEE S   S EE    EE

 101 QCVKKKDLKE SISLRISKKI NPFNVPEEQL HNIDEYDLNV VRLCFQAFLP
     EE  TTHHHH HHHHHHTTT   TT   TTTT    TTTTTTE EEEEEEEEEE

 151 DEHGNYTLAL PPLISNPIYD NRAPNTAELR ICRVNKNCGS VKGGDEIFIL
      SSSSB EE    EE   EEE TTTTTTS    B EES SEEE TT   EEEEE

 201 CDKVQKDDIE VRFVLDNWEA KGSFSQADVH RQVAIVFRTP PFLRDITEPI
     ES   GGG E EEEEETTEEE E B  GGGEE TTTEEEEE     SS  SS E

 251 TVKMQLRRPS DQEVSEPMDF RYLPD
     EEEEEEEEGG GTEE   EEE EEE
```

**http://www.rcsb.org/pdb/cgi/explore.cgi?job=summary&pdbId=1GJI&page=0&pid=8818118348057**

## Appendix A: HIV-Lock™ Structure Background Information

**Left and Right Operator DNA Sequences (Bacteriophage Lambda)**
http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=215104

```
LOCUS       LAMCG                   48502 bp    DNA     circular PHG 17-APR-
2002
DEFINITION  Bacteriophage lambda, complete genome.
ACCESSION   J02459 M17233 M24325 V00636 X00906
VERSION     J02459.1

Left Operator Sequences (LETTERS IN RED: OL1, OL2, OL3)
    35581 atgtgctcag tatcaccgcc agtggtattt atgtcaacac cgccagagat aatttatcac
    35641 cgcagatggt tatctgtatg ttttttatat gaatttattt tttgcagggg ggcattgttt

Right Operator Sequences (LETTERS IN RED: OR3, OR2, OR1)
    37921 ggtttctttt ttgtgctcat acgttaaatc tatcaccgca agggataaat atctaacacc
    37981 gtgcgtgttg actattttac ctctggcggt gataatggtt gcatgtacta aggaggttgt
```

## Appendix A: HIV-Lock™ Structure Background Information

**HIV LTR Sequence for Molecular Parataxis™ Paper and Example**
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Nucleotide&list_uids=33359200&dopt=GenBank

```
LOCUS       AY352275               10280 bp    DNA     linear   VRL 05-AUG-2003
DEFINITION  HIV-1 isolate SF33 from USA, complete genome.
ACCESSION   AY352275 M38427
VERSION     AY352275.1  GI:33359200
KEYWORDS    .
SOURCE      Human immunodeficiency virus 1 (HIV-1)
  ORGANISM  Human immunodeficiency virus 1
            Viruses; Retro-transcribing viruses; Retroviridae;
            Orthoretrovirinae; Lentivirus; Primate lentivirus group.
REFERENCE   1  (bases 1 to 10280)
  AUTHORS   York-Higgins,D., Cheng-Mayer,C., Bauer,D., Levy,J.A. and Dina,D.
  TITLE     Human immunodeficiency virus type 1 cellular host range,
            replication, and cytopathicity are linked to the envelope region of
            the viral genome
  JOURNAL   J. Virol. 64 (8), 4016-4020 (1990)
   PUBMED   2370688
REFERENCE   2  (bases 1 to 10280)
  AUTHORS   York-Higgins,D., Cheng-Mayer,C., Bauer,D., Levy,J.A. and Dina,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1993) Medicine, University of California, San
            Francisco, 513 Parnassus, San Francisco, CA 94143-1270, USA
REFERENCE   3  (bases 1 to 10280)
  AUTHORS   York-Higgins,D., Cheng-Mayer,C., Bauer,D., Levy,J.A., Dina,D. and
            Bonneau,K.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2003) Medicine, University of California, San
            Francisco, 513 Parnassus, San Francisco, CA 94143-1270, USA
  REMARK    Nucleotide sequence updated by submitter
COMMENT     On Aug 5, 2003 this sequence version replaced gi:328668.
FEATURES             Location/Qualifiers
     source          1..9714
                     /organism="Human immunodeficiency virus 1"
                     /proviral
                     /mol_type="genomic DNA"
                     /isolate="SF33"
                     /isolation_source="from patient in 1984"
                     /db_xref="taxon:11676"
                     /country="USA: Philadelphia"
                     /focus
                     /note="biologically active upon transfection into human
                     rhabdomyosarcoma (RD-4) and HUT 78 cells; very cytopathic"
     source          9715..10280
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     LTR             1..635
     gene            792..2303
                     /gene="gag"
     CDS             792..2303
                     /gene="gag"
                     /codon_start=1
LTR                9081..9715
```

## Appendix A: HIV-Lock™ Structure Background Information

**HIV LTR Sequence for Molecular Parataxis™ Paper and Example, continued**
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Nucleotide&list_uids=33359200&dopt=GenBank

```
     9361 atttcgtcac atggcccgag agctgcatcc ggagtactac aaagactgct gacatcgagt
     9421 tttctacaag ggactttccg ctgggggactt tccaggggag gcgtggcctg ggcgggactg
     9481 gggagtggcg agccctcaga tgctgcatat aagcagctgc tttttgcctg tacgggtct
     9541 ctctggttag accagatctg agcctgggag ctctctggct aactagggaa cccactgctt
     9601 aagcctcaat aaagcttgcc ttgagtgctt caagtagtgt gtgcccgtct gttgtgtgac
     9661 tctgctatct agagatccct cagacccttt tagtcagtgt ggaaaatctc tagcaatata
     9721 taaatatatc tttgaccttt acagcatatg gtaataactt aaaaattata tgcctaattg
     9781 tgaaaaaaaa aaaagaaaaa agaactcttc ttgccagaat ccaagtccca tgaaagtagc
     9841 caatgctgtc tcattagtta gtaagctaat ggaaatgttg ccagcatttc tttcagtgtc
     9901 tagaaaacag agtgtgcaat gtgccaagtc ttcactgatt tatttttgta agcagcagtg
     9961 taataaaccc aaagaagcca aaaaagcaaa tttttaaaaa ataaatattc atttgctatc
    10021 aagatgggta tgacctttt acccaagcct attactgaca attcagaaag actatgtgaa
    10081 atagtcactc atttatctta attgcatttg caggtactac caccactcaa gttttaaaat
    10141 gtttttaaac actcaagttt gcattccttt agcttttata caagaaacca cattatttta
    10201 catacatatt aattattttc tgacctttca ggaaaaccca ataatataaa tctacaaaat
    10261 gaaataatac tcaagaattc
```

*HIV-LTR in ITALICS*
**P50/rel binding sites in bold red**
**Sp1 binding sites in bold blue**
**TCF-Ialpha binding site in green**
**Ap1 binding site in orange**
**NFAT binding site in magenta**

Selected transcription factor sites designations from:
Van Lint et al. (1997) J.of Virology, 71 (8) 6113-6127 and
Montanyo et al. (1996) PNAS 93,12376-12381.

## Appendix A: HIV-Lock™ Structure Background Information

**BLAST SEQUENCE gggtctctctggttagaccagatctgagcctgggagctctct REPORT**

. Human immunodeficiency virus 1 ---- 83 108 hits [viruses] HIV-1 isolate P18-88 from USA long terminal repeat, partial
. Lentiviral transfer vector pHsCXW . 83 2 hits [other sequences] Lentiviral transfer vector pHsCXW, complete sequence
. HIV-1 vector pNL4-3 ..... 83 2 hits [other sequences] HIV-1 vector pNL4-3, complete sequence
. synthetic construct ..... 83 1 hit [other sequences] Synthetic construct HIV-1 self-inactivating chimeric LTR wi

Organism Report
Human immunodeficiency virus 1 [viruses] taxid 11676Score E
Sequences producing significant alignments: (Bits) Value

| | | | |
|---|---|---|---|
| gb\|AY376266.1\| | HIV-1 isolate P18-88 from USA long terminal re... | 83.8 | 2e-14 |
| gb\|AY376265.1\| | HIV-1 isolate P18-86 from USA long terminal re... | 83.8 | 2e-14 |
| gb\|AY376264.1\| | HIV-1 isolate P18-84 from USA long terminal re... | 83.8 | 2e-14 |
| gb\|AY376259.1\| | HIV-1 isolate P13-95 from USA long terminal re... | 83.8 | 2e-14 |
| gb\|AY376258.1\| | HIV-1 isolate P13-91 from USA long terminal re... | 83.8 | 2e-14 |
| gb\|AY376257.1\| | HIV-1 isolate P13-85 from USA long terminal re... | 83.8 | 2e-14 |
| gb\|AY376256.1\| | HIV-1 isolate P10-96 from USA long terminal re... | 83.8 | 2e-14 |
| gb\|AY376249.1\| | HIV-1 isolate SP13-90 from USA long terminal r... | 83.8 | 2e-14 |
| gb\|AY376248.1\| | HIV-1 isolate SP13-93 from USA long terminal r... | 83.8 | 2e-14 |
| gb\|AY352655.1\| | HIV-1 isolate SE9010 from Sweden, complete genome | 83.8 | 2e-14 |
| gb\|AY468486.1\| | Lentiviral transfer vector pHsCXW, complete seque | 83.8 | 2e-14 |
| gb\|AY423386.1\| | HIV-1 isolate 671-00T103 from Netherlands, comple | 83.8 | 2e-14 |
| gb\|AY423382.1\| | HIV-1 isolate 671-00T93 from Netherlands, complet | 83.8 | 2e-14 |
| gb\|AY423381.1\| | HIV-1 isolate 671-99T12 from Netherlands, complet | 83.8 | 2e-14 |
| gb\|AY352275.1\| | HIV-1 isolate SF33 from USA, complete genome | 83.8 | 2e-14 |
| gb\|AF224507.1\| | HIV-1 strain HIV-1wk from Korea, complete genome | 83.8 | 2e-14 |
| gb\|AY781125.1\| | HIV-1 isolate 01UYTRA1101 from Uruguay, complete | 83.8 | 2e-14 |
| gb\|AY610991.1\| | HIV-1 isolate 03SP123900 from Spain LTR, partial | 83.8 | 2e-14 |
| gb\|AY610979.1\| | HIV-1 isolate 03SP96716 from Spain LTR, partial s | 83.8 | 2e-14 |
| gb\|AY610971.1\| | HIV-1 isolate 03SP97646 from Spain LTR, partial s | 83.8 | 2e-14 |
| gb\|AY819715.1\| | HIV-1 isolate 04RU139095 from Russia, complete ge | 83.8 | 2e-14 |
| gb\|K03455.1\|HIVHXB2CG | Human immunodeficiency virus type 1 (HX... | 83.8 | 2e-14 |
| gb\|AY121475.1\| | HIV-1 isolate 01KMH11 from South Korea nef pro... | 83.8 | 2e-14 |
| gb\|AY121464.1\| | HIV-1 isolate 97PJH8 from South Korea nonfunct... | 83.8 | 2e-14 |
| gb\|AF462793.1\| | HIV-1 isolate 01CMH8 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462792.1\| | HIV-1 isolate 01HSJ4 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462789.1\| | HIV-1 isolate 00YKW1 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462785.1\| | HIV-1 isolate 00KMH11(59) from South Korea nef... | 83.8 | 2e-14 |
| gb\|AF462782.1\| | HIV-1 isolate 99KMH610 from South Korea nef pr... | 83.8 | 2e-14 |
| gb\|AF462781.1\| | HIV-1 isolate 99KMH6 from South Korea nef-like... | 83.8 | 2e-14 |
| gb\|AF462779.1\| | HIV-1 isolate 00JHS11 from South Korea nef pro... | 83.8 | 2e-14 |
| gb\|AF462778.2\| | HIV-1 isolate 00KYR12 from South Korea nef pro... | 83.8 | 2e-14 |
| gb\|AF462776.1\| | HIV-1 isolate 00KBH11 from South Korea nef pro... | 83.8 | 2e-14 |
| gb\|AF462775.1\| | HIV-1 isolate 93KBH7 from South Korea nef-like... | 83.8 | 2e-14 |
| gb\|AF462766.1\| | HIV-1 isolate 00KJK11 from South Korea nef pro... | 83.8 | 2e-14 |
| gb\|AF462757.1\| | HIV-1 isolate 01CWS2 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462755.1\| | HIV-1 isolate 94CWS1 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462753.1\| | HIV-1 isolate 99KMK3 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462748.1\| | HIV-1 isolate 00KGS10 from South Korea nef pro... | 83.8 | 2e-14 |
| gb\|AF462746.1\| | HIV-1 isolate 00JWK12 from South Korea nef pro... | 83.8 | 2e-14 |
| gb\|AF462730.1\| | HIV-1 isolate 91OCH8 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462727.1\| | HIV-1 isolate 00JJS2 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462716.1\| | HIV-1 isolate 99YWS3 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462707.1\| | HIV-1 isolate 00HYH3 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462706.1\| | HIV-1 isolate 00KYJ12 from South Korea nef pro... | 83.8 | 2e-14 |
| gb\|AF462705.2\| | HIV-1 isolate 00KJS3 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AF462704.1\| | HIV-1 isolate 00LJI12 from South Korea nef pro... | 83.8 | 2e-14 |
| gb\|AF462701.1\| | HIV-1 isolate 94LSW12 from South Korea nef pro... | 83.8 | 2e-14 |
| gb\|AF462693.1\| | HIV-1 isolate 99CSR4 from South Korea nef prot... | 83.8 | 2e-14 |
| gb\|AY093617.1\| | HIV-1 isolate GA from USA nef protein (nef) gene, | 83.8 | 2e-14 |
| gb\|AY851682.1\| | HIV-1 isolate SP-LTNP4 from Spain LTR, partial se | 83.8 | 2e-14 |
| gb\|AY851681.1\| | HIV-1 isolate SP-LTNP3 from Spain LTR, partial se | 83.8 | 2e-14 |
| gb\|AY851679.1\| | HIV-1 isolate SP-LTNP2 from Spain LTR, partial se | 83.8 | 2e-14 |
| gb\|AY851678.1\| | HIV-1 isolate SP-LTNP20 from Spain LTR, partial s | 83.8 | 2e-14 |
| gb\|AY851676.1\| | HIV-1 isolate SP-LTNP1 from Spain LTR, partial se | 83.8 | 2e-14 |
| gb\|AY851674.1\| | HIV-1 isolate SP-LTNP6 from Spain LTR, partial se | 83.8 | 2e-14 |

## Appendix A: HIV-Lock™ Structure Background Information

**BLAST SEQUENCE gggtctctctggttagaccagatctgagcctgggagctctct REPORT, continued**

| | | | | |
|---|---|---|---|---|
| gb\|AY851672.1\| | | HIV-1 isolate SP-LTNP18 from Spain LTR, partial s | 83.8 | 2e-14 |
| gb\|AY851671.1\| | | HIV-1 isolate SP-LTNP13 from Spain LTR, partial s | 83.8 | 2e-14 |
| gb\|AY851670.1\| | | HIV-1 isolate SP-LTNP14 from Spain LTR, partial s | 83.8 | 2e-14 |
| gb\|AY851669.1\| | | HIV-1 isolate SP-LTNP11 from Spain LTR, partial s | 83.8 | 2e-14 |
| gb\|AF361871.1\| | AF361871 | HIV-1 isolate 97TZ01 from Tanzania gag... | 83.8 | 2e-14 |
| gb\|AY561239.1\| | | HIV-1 isolate PCM039 from Colombia genomic sequen | 83.8 | 2e-14 |
| gb\|AY561237.1\| | | HIV-1 isolate PCM013 from Colombia gag protein... | 83.8 | 2e-14 |
| gb\|AF075719.1\| | AF075719 | HIV-1 isolate MN clone MNTQ from the USA, | 83.8 | 2e-14 |
| gb\|U21135.1\| | HIV1U21135 | Human immunodeficiency virus type 1 cl... | 83.8 | 2e-14 |
| gb\|AF324493.1\| | AF324493 | HIV-1 vector pNL4-3, complete sequence | 83.8 | 2e-14 |
| gb\|AF070521.1\| | AF070521 | HIV-1 E9 from the USA, complete genome | 83.8 | 2e-14 |
| gb\|U89866.1\| | HIVU89866 | HIV-1 clone L-50 patient SG1 from Italy... | 83.8 | 2e-14 |
| gb\|U89865.1\| | HIVU89865 | HIV-1 clone L-47 patient SG1 from Italy... | 83.8 | 2e-14 |
| gb\|U89863.1\| | HIVU89863 | HIV-1 clone L-29 patient SG1 from Italy... | 83.8 | 2e-14 |
| gb\|U89862.1\| | HIVU89862 | HIV-1 clone L-27 patient SG1 from Italy... | 83.8 | 2e-14 |
| gb\|U89861.1\| | HIVU89861 | HIV-1 clone L-12 patient SG1 from Italy... | 83.8 | 2e-14 |
| gb\|U89860.1\| | HIVU89860 | HIV-1 clone L-8 patient SG1 from Italy,... | 83.8 | 2e-14 |
| gb\|U89859.1\| | HIVU89859 | HIV-1 clone L-4 patient SG1 from Italy,... | 83.8 | 2e-14 |
| gb\|U89858.1\| | HIVU89858 | HIV-1 clone L-3 patient SG1 from Italy,... | 83.8 | 2e-14 |
| gb\|U89857.1\| | HIVU89857 | HIV-1 clone L-2 patient SG1 from Italy,... | 83.8 | 2e-14 |
| gb\|U89856.1\| | HIVU89856 | HIV-1 clone L-1 patient SG1 from Italy,... | 83.8 | 2e-14 |
| gb\|AF019404.1\| | AF019404 | HIV-1 strain TP825 from Spain, 3' long... | 83.8 | 2e-14 |
| gb\|AF019398.1\| | AF019398 | HIV-1 strain NP644 from Spain, 3' long... | 83.8 | 2e-14 |
| gb\|AF019395.1\| | AF019395 | HIV-1 strain NP625 from Spain, 3' long... | 83.8 | 2e-14 |
| gb\|AF019393.1\| | AF019393 | HIV-1 strain NP623 from Spain, 3' long... | 83.8 | 2e-14 |
| gb\|AF019392.1\| | AF019392 | HIV-1 strain NP588 from Spain, 3' long... | 83.8 | 2e-14 |
| gb\|AF019391.1\| | AF019391 | HIV-1 strain NP585 from Spain, 3' long... | 83.8 | 2e-14 |
| gb\|AF286365.1\| | AF286365 | HIV-1 isolate WR27 from USA, complete gen | 83.8 | 2e-14 |
| gb\|AF004394.1\| | AF004394 | HIV-1 strain AD8, complete genome | 83.8 | 2e-14 |
| gb\|AF000542.1\| | AF000542 | HIV-1 strain HP83-A1.ltr from USA, TAR re | 83.8 | 2e-14 |
| gb\|AF000538.1\| | AF000538 | HIV-1 strain DJ93-A1.ltr from USA, TAR re | 83.8 | 2e-14 |
| gb\|AF000537.1\| | AF000537 | HIV-1 strain CW94-B2.ltr from USA, TAR re | 83.8 | 2e-14 |
| gb\|AF000535.1\| | AF000535 | HIV-1 strain BT94-B1.ltr from USA, TAR re | 83.8 | 2e-14 |
| gb\|AF000534.1\| | AF000534 | HIV-1 strain BJ93-A1.ltr from USA, TAR re | 83.8 | 2e-14 |
| gb\|AF000533.1\| | AF000533 | HIV-1 strain AD92-A2.ltr from USA, TAR re | 83.8 | 2e-14 |
| gb\|U26942.1\| | HIV1U26942 | Human immunodeficiency virus clone pNL4-3 | 83.8 | 2e-14 |
| gb\|K02007.1\| | HIVSF2CG | Human immunodeficiency virus type 1, iso... | 83.8 | 2e-14 |
| gb\|AF272008.1\| | AF272008 | HIV-1 isolate 99KMK3 from South Korea ... | 83.8 | 2e-14 |
| gb\|AF272007.1\| | AF272007 | HIV-1 isolate 99OHS12 from South Korea... | 83.8 | 2e-14 |
| gb\|AF272006.1\| | AF272006 | HIV-1 isolate 0HYH3 from South Korea 3... | 83.8 | 2e-14 |
| gb\|U23487.1\| | HIV1U23487 | Human immunodeficiency virus type 1 ki... | 83.8 | 2e-14 |
| gb\|AF237862.1\| | AF237862 | Synthetic construct HIV-1 self-inactiv... | 83.8 | 2e-14 |
| gb\|U81477.1\| | HIVU81477 | HIV-1 patient VE23B from Venezuela, sam... | 83.8 | 2e-14 |
| gb\|U81476.1\| | HIVU81476 | HIV-1 patient VE23A from Venezuela, sam... | 83.8 | 2e-14 |

Alignments

| | | | |
|---|---|---|---|
| Query | 1 | GGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCT | 42 |
| AY376266 | 476 | ..... | 517 |
| AY376265 | 476 | ..... | 517 |
| AY376264 | 476 | ..... | 517 |
| AY376259 | 476 | ..... | 517 |
| AY376258 | 476 | ..... | 517 |
| AY376257 | 476 | ..... | 517 |
| AY376256 | 476 | ..... | 517 |
| AY376249 | 477 | ..... | 518 |
| AY376248 | 477 | ..... | 518 |
| AY352655 | 8976 | ..... | 9017 |
| AY468486 | 454 | ..... | 495 |
| AY468486 | 3566 | ..... | 3607 |
| AY423386 | 68 | ..... | 109 |
| AY423382 | 77 | ..... | 118 |
| AY423381 | 9086 | ..... | 9127 |
| AY423381 | 1 | ..... | 17 |
| AY352275 | 455 | ..... | 496 |
| AY352275 | 9535 | ..... | 9576 |
| AF224507 | 1 | ..... | 42 |
| AF224507 | 9138 | ..... | 9179 |
| AY781125 | 8739 | ..... | 8780 |

## Appendix A: HIV-Lock™ Structure Background Information

**BLAST SEQUENCE gggtctctctggttagaccagatctgagcctgggagctctct REPORT, continued**

| | | | |
|---|---|---|---|
| AY610991 | 399 | ..... | 440 |
| AY610979 | 431 | ..... | 472 |
| AY610971 | 360 | ..... | 401 |
| AY819715 | 8887 | ..... | 8928 |
| K03455 | 454 | ..... | 495 |
| K03455 | 9539 | ..... | 9580 |
| AY121475 | 785 | ..... | 826 |
| AY121464 | 317 | ..... | 358 |
| AF462793 | 745 | ..... | 786 |
| AF462792 | 759 | ..... | 800 |
| AF462789 | 757 | ..... | 798 |
| AF462785 | 683 | ..... | 724 |
| AF462782 | 815 | ..... | 856 |
| AF462781 | 379 | ..... | 420 |
| AF462779 | 745 | ..... | 786 |
| AF462778 | 747 | ..... | 788 |
| AF462776 | 731 | ..... | 772 |
| AF462775 | 731 | ..... | 772 |
| AF462766 | 744 | ..... | 785 |
| AF462757 | 754 | ..... | 795 |
| AF462755 | 745 | ..... | 786 |
| AF462753 | 746 | ..... | 787 |
| AF462748 | 787 | ..... | 828 |
| AF462746 | 765 | ..... | 806 |
| AF462730 | 757 | ..... | 798 |
| AF462727 | 773 | ..... | 814 |
| AF462716 | 769 | ..... | 810 |
| AF462707 | 766 | ..... | 807 |
| AF462706 | 739 | ..... | 780 |
| AF462705 | 773 | ..... | 814 |
| AF462704 | 766 | ..... | 807 |
| AF462701 | 743 | ..... | 784 |
| AF462693 | 734 | ..... | 775 |
| AY093617 | 777 | ..... | 818 |

```
Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental
samples or phase 0, 1 or 2 HTGS sequences)
  Posted date:  Jun 7, 2005  9:47 PM
Number of letters in database: 1312653073
Number of sequences in database:  3181128
Lambda     K      H
    1.37    0.711     1.31
Gapped
Lambda     K      H
    1.37    0.711     1.31
Matrix: blastn matrix:1 -3
Gap Penalties: Existence: 5, Extension: 2
Number of Sequences: 3181128
Number of Hits to DB: 239875019
Number of extensions: 13034202
Number of successful extensions: 23813
Number of sequences better than 1000: 955
Number of HSP's better than 1000 without gapping: 951
Number of HSP's gapped: 23748
Number of HSP's successfully gapped: 1002
Number of extra gapped extensions for HSPs above 1000: 22314
Length of query: 42
Length of database: 14197554961
Length adjustment: 19
Effective length of query: 23
Effective length of database: 14137113529
Effective search space: 325153611167
Effective search space used: 325153611167
A: 0
X1: 11 (21.8 bits)
X2: 15 (29.7 bits)
X3: 25 (49.6 bits)
S1: 12 (24.3 bits)
S2: 15 (30.2 bits)
```

## Appendix A: HIV-Lock™ Structure Background Information

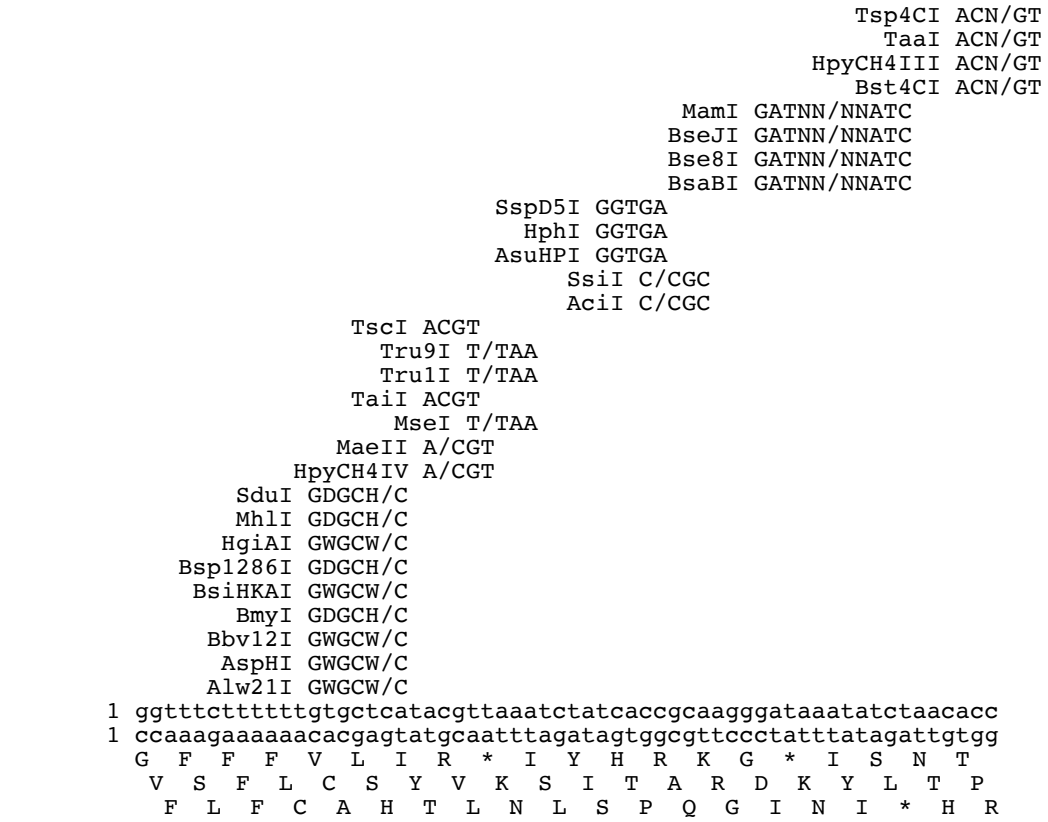
**HIV-Lock™ Literature References**

Albright, R. A. and Matthews (1998) J Mol Biol 280, 137-51
Baeuerle, P.A. and Baltimore, D. (1989) Genes Dev. 3, 1689–98
Brownell, E., et al. (1989) Oncogene 4 (7), 935-942
Dahlberg, J.E. and Blattner,F.R. (1975) Nucleic Acids Res. 2 (9), 1441-58
Ghosh, S., et al. (1990) Cell 62, 1019–29
Ghosh, S., et al. (1995) Nature 373, 303-10
Huang, D. B., et al. (2001) Structure 9 (8), 669-78
Humayun, Z., et al. (1977) J. Mol. Biol. 112 (2), 265-77
Kieran, M., et al. (1990) Cell 62, 1007–18
Meyer, B.J.,et al. (1980) J. Mol. Biol. 139 (2), 163-94
Montanyo, M.A. et al. (1996) PNAS 93,12376-12381.
Roberts, T.M., et al. (1977) Nature 270 (5634), 274-75
Schmid, R.M., et al. (1991) Nature 352, 733–6
Schwarz, E., et al. (1978) Nature 272 (5652), 410-14
Takeda, Y., et al. (1977) J. Mol. Biol. 127, 6177-83
Takeda, Y., et al. (1986) J. Mol. Biol. 261, 8608-16
Van Lint, C. et al. (1997) J.of Virology, 71 (8) 6113-6127
Weininger, S. and Weininger, A. (2005) (Molecular Parataxis paper)

U.S. Pat. Nos. RE 5,460,965.
U.S. Pat. Nos. RE 5,324,818.
U.S. Pat. Nos. RE 5,871,902.

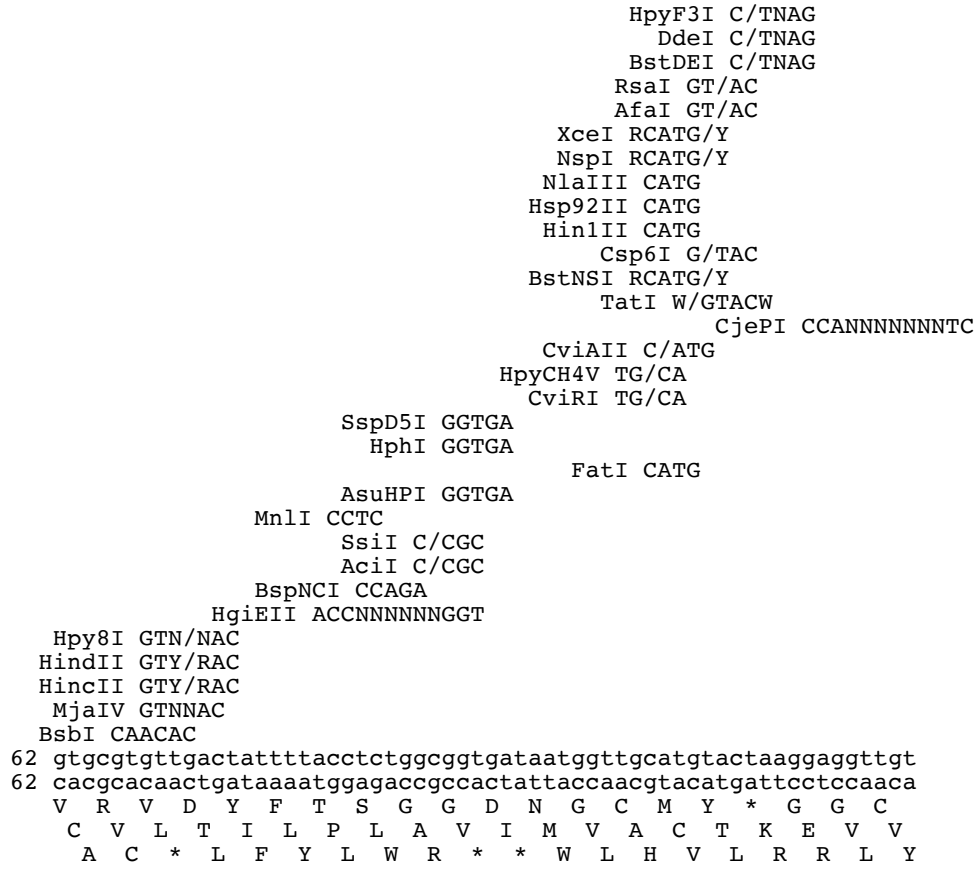
## Appendix A: HIV-Lock™ Structure Background Information

**Cro protein selected restriction digest**
**HIV-Lock™ incorporated residues in bold**



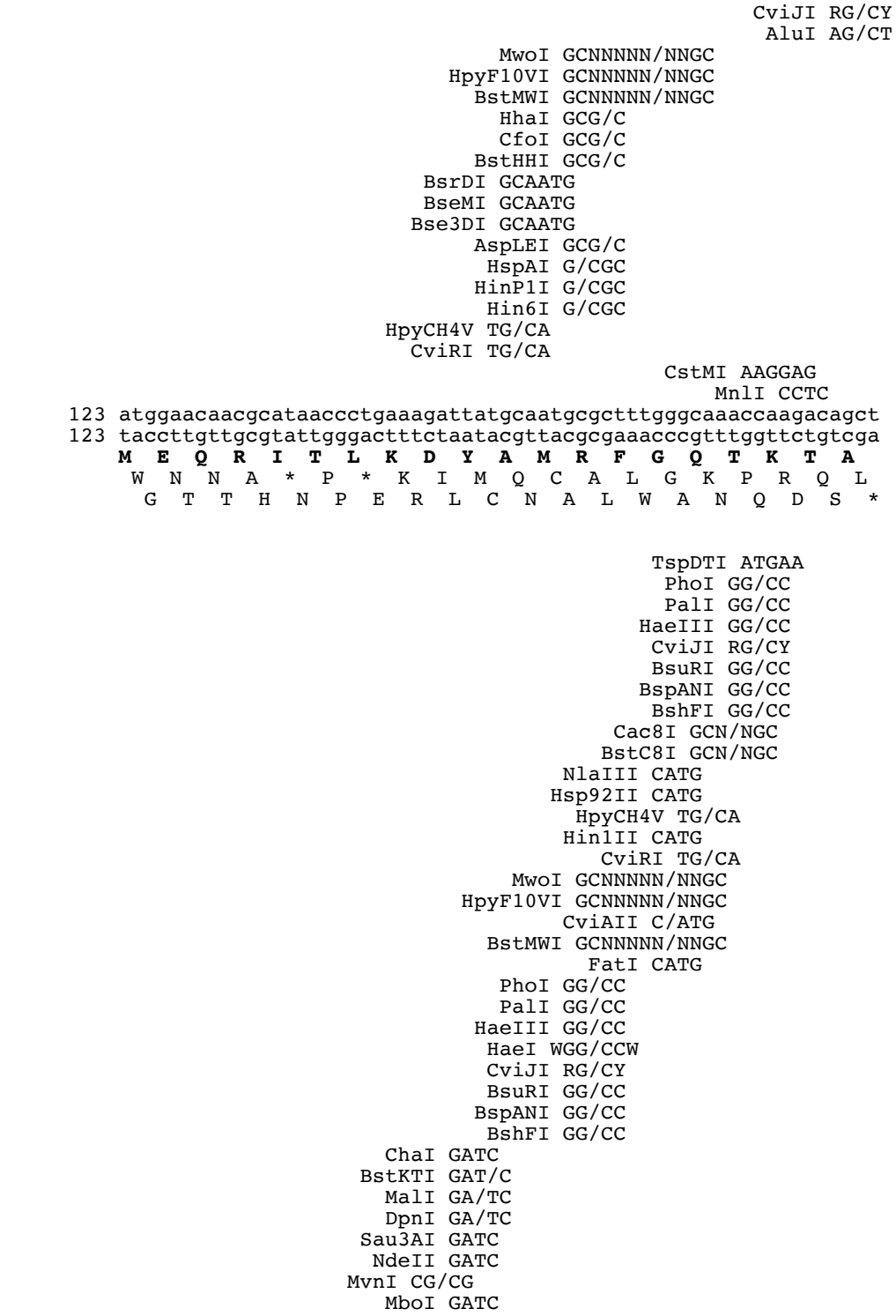
## Appendix A: HIV-Lock™ Structure Background Information

**Cro protein selected restriction digest, continued**
**HIV-Lock™ incorporated residues in bold**



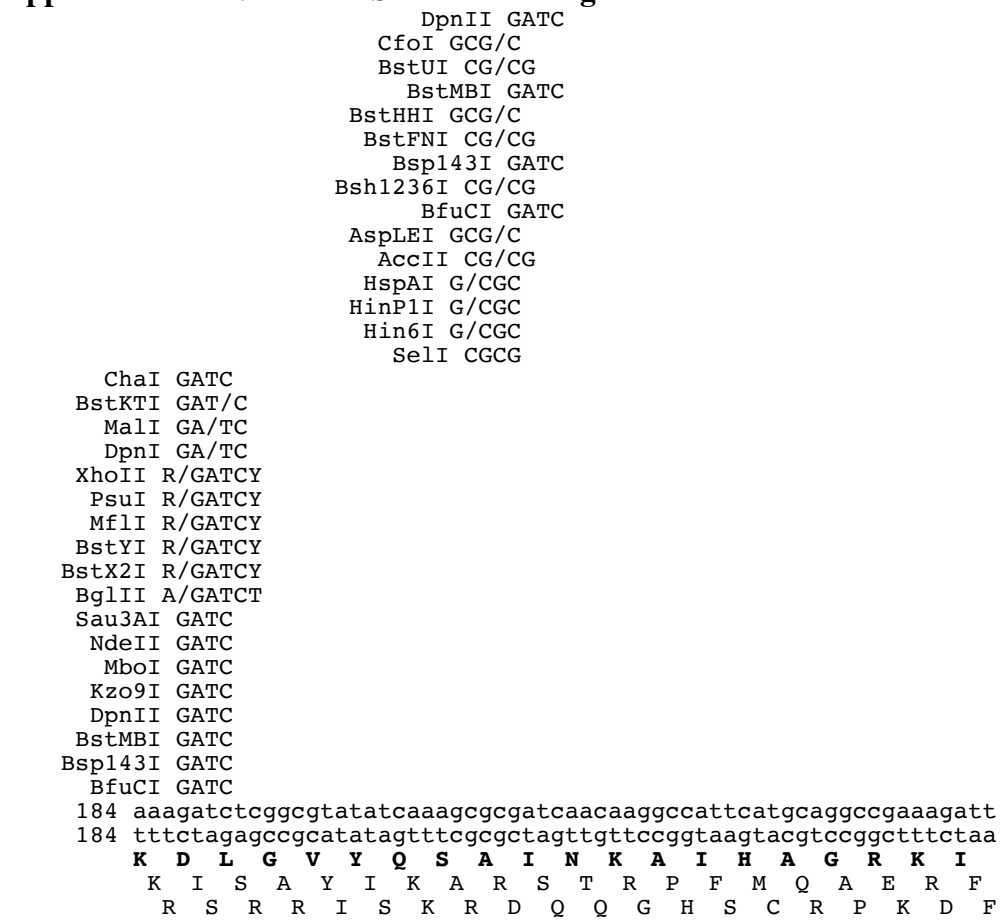
## Appendix A: HIV-Lock™ Structure Background Information

**Cro protein selected restriction digest,continued**
**HIV-Lock™ incorporated residues in bold**



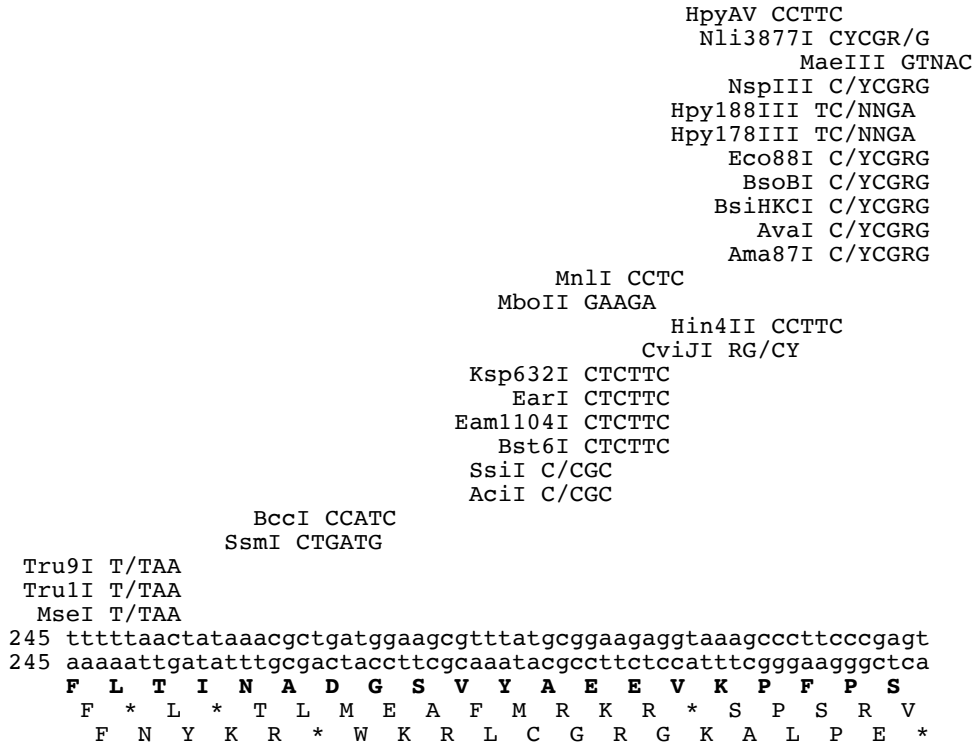

**Cro protein selected restriction digest, continued**
**HIV-Lock™ incorporated residues in bold**

## Appendix A: HIV-Lock™ Structure Background Information



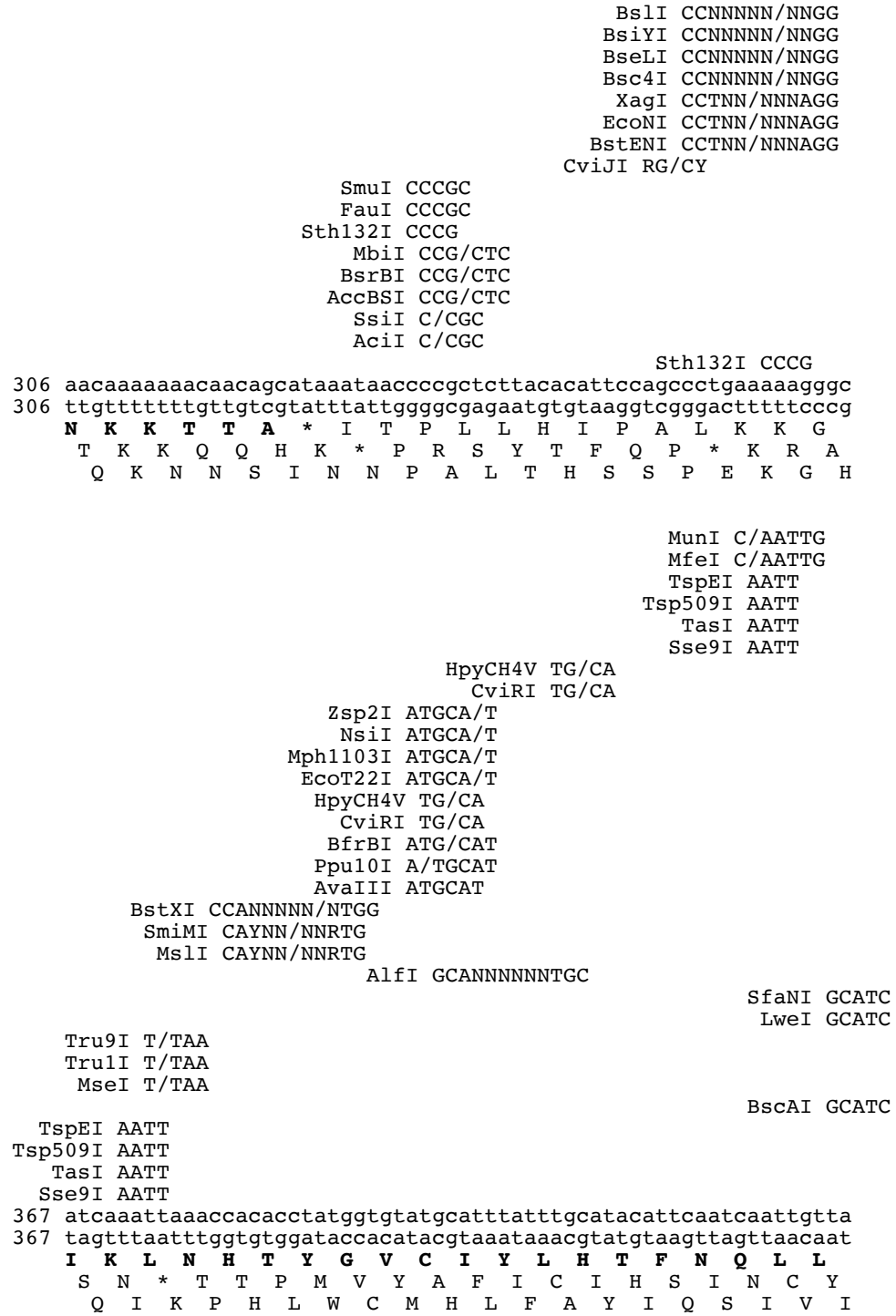
## Appendix A: HIV-Lock™ Structure Background Information

**Cro protein selected restriction digest, continued**
**HIV-Lock™ incorporated residues in bold**



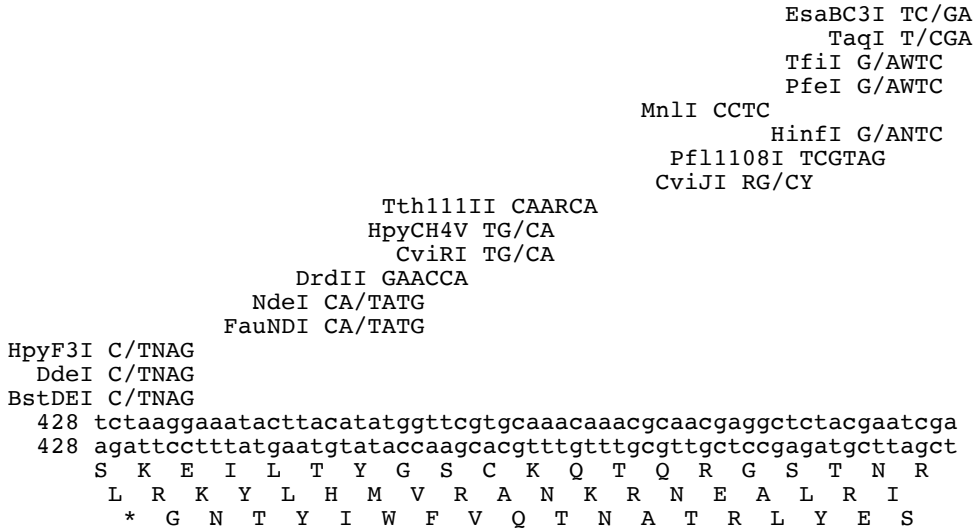
## Appendix A: HIV-Lock™ Structure Background Information

**Cro protein selected restriction digest, continued**
**HIV-Lock™ incorporated residues in bold**



## Appendix A: HIV-Lock™ Structure Background Information

**Cro protein selected restriction digest, continued**
**HIV-Lock™ incorporated residues in bold**



## Appendix A: HIV-Lock™ Structure Background Information

**P50 selected restriction digest**
**HIV-Lock™ incorporated residues in bold**

```
                                                                NciI CC/SGG
                                                               EcoHI CCSGG
                     MaeI C/TAG
                    FspBI C/TAG
   1 agcggccgccgcgggcgcgctctagcagcgcaggccggagctcagggccccgcgcgcccg
   1 tcgccggcggcgcccgcgcgagatcgtcgcgtccggcctcgagtcccggggcgcgcgggc
     S  G  R  R  G  R  A  L  A  A  Q  A  G  A  Q  G  P  A  R  P
      A  A  A  A  G  A  L  *  Q  R  R  P  E  L  R  A  P  R  A  R
       R  P  P  R  A  R  S  S  S  A  G  R  S  S  G  P  R  A  P  G


  62 gcccgccccgcgcttctccgcccgcgccgcagccatggcgcgccgctgagccgcccgccc
  62 cgggcggggcgcgaagaggcgggcgcggcgtcggtaccgcgcggcgactcggcgggcggg
     A  R  P  A  L  L  R  P  R  R  S  H  G  A  P  L  S  R  P  P
      P  A  P  R  F  S  A  R  A  A  A  M  A  R  R  *  A  A  R  P
       P  P  R  A  S  P  P  A  P  Q  P  W  R  A  A  E  P  P  A  R


                NciI CC/SGG
               EcoHI CCSGG
 123 gcccgcccgcgccccgacccggctcggctcccgccggtccgcgccgctccgcagcggagc
 123 cgggcgggcgcggggctgggccgagccgagggcggccaggcgcggcgaggcgtcgcctcg
     A  R  P  R  P  D  P  A  R  L  P  P  V  R  A  A  P  Q  R  S
      P  A  R  A  P  T  R  L  G  S  R  R  S  A  P  L  R  S  G  A
       P  P  A  P  R  P  G  S  A  P  A  G  P  R  R  S  A  A  E  P


 184 ccgcaggcgaggagaggccgcgcgcatctccagggtaccctcagaggccagaagagggtg
 184 ggcgtccgctcctctccggcgcgcgtagaggtcccatgggagtctccggtcttctcccac
     P  Q  A  R  R  G  R  A  H  L  Q  G  T  L  R  G  Q  K  R  V
      R  R  R  G  E  A  A  R  I  S  R  V  P  S  E  A  R  R  G  C
       A  G  E  E  R  P  R  A  S  P  G  Y  P  Q  R  P  E  E  G  V


 245 tcagagcccttgtaactggagtttgacggtcgtgagctgcgcatcttcaccatggcagac
 245 agtctcgggaacattgacctcaaactgccagcactcgacgcgtagaagtggtaccgtctg
     S  E  P  L  *  L  E  F  D  G  R  E  L  R  I  F  T  M  A  D
      Q  S  P  C  N  W  S  L  T  V  V  S  C  A  S  S  P  W  Q  T
       R  A  L  V  T  G  V  *  R  S  *  A  A  H  L  H  H  G  R  R


                                                            SspI AAT/ATT
 306 gatgatccctacggaactgggcaaatgtttcatttgaacactgctttgactcactcaata
 306 ctactagggatgccttgacccgtttacaaagtaaacttgtgacgaaactgagtgagttat
     D  D  P  Y  G  T  G  Q  M  F  H  L  N  T  A  L  T  H  S  I
      M  I  P  T  E  L  G  K  C  F  I  *  T  L  L  *  L  T  Q  Y
       *  S  L  R  N  W  A  N  V  S  F  E  H  C  F  D  S  L  N  I


 367 tttaatgcagaattatattcaccagaaataccactgtcaacagatggcccataccttcaa
 367 aaattacgtcttaatataagtggtctttatggtgacagttgtctaccgggtatggaagtt
     F  N  A  E  L  Y  S  P  E  I  P  L  S  T  D  G  P  Y  L  Q
      L  M  Q  N  Y  I  H  Q  K  Y  H  C  Q  Q  M  A  H  T  F  K
       *  C  R  I  I  F  T  R  N  T  T  V  N  R  W  P  I  P  S  N
```

## Appendix A: HIV-Lock™ Structure Background Information

**P50 selected restriction digest, continued**
**HIV-Lock™ incorporated residues in bold**

```
                          EsaBC3I TC/GA
                             TaqI T/CGA
                                                              SspI AAT/ATT
428 atattagagcaaccaaaacagaggggatttcgattccgctatgtgtgtgaaggcccatca
428 tataatctcgttggttttgtctccctaaagctaaggcgatacacacacttccgggtagt
    I  L  E  Q  P  K  Q  R  G  F  R  F  R  Y  V  C  E  G  P  S
     Y  *  S  N  Q  N  R  G  D  F  D  S  A  M  C  V  K  A  H  H
      I  R  A  T  K  T  E  G  I  S  I  P  L  C  V  *  R  P  I  T


                                AloI GAACNNNNNNTCC
                       MaeI C/TAG
                      FspBI C/TAG
              NciI CC/SGG
             EcoHI CCSGG
489 cacggagggcttccgggagcctctagtgagaagaacaagaaatcctacccacaggtcaaa
489 gtgcctcccgaaggccctcggagatcactcttcttgttctttaggatgggtgtccagttt
    H  G  G  L  P  G  A  S  S  E  K  N  K  K  S  Y  P  Q  V  K
     T  E  G  F  R  E  P  L  V  R  R  T  R  N  P  T  H  R  S  K
      R  R  A  S  G  S  L  *  *  E  E  Q  E  I  L  P  T  G  Q  N


550 atttgcaactatgtggggcctgcaaaggttatcgttcagttggtcacaaatggaaaaaac
550 taaacgttgatacaccccggacgtttccaatagcaagtcaaccagtgtttacctttttg
    I  C  N  Y  V  G  P  A  K  V  I  V  Q  L  V  T  N  G  K  N
     F  A  T  M  W  G  L  Q  R  L  S  F  S  W  S  Q  M  E  K  T
      L  Q  L  C  G  A  C  K  G  Y  R  S  V  G  H  K  W  K  K  H


611 atccacctgcacgcccacagcctggtgggcaagcactgtgaggacggggtatgcaccgta
611 taggtggacgtgcgggtgtcggaccacccgttcgtgacactcctgccccatacgtggcat
    I  H  L  H  A  H  S  L  V  G  K  H  C  E  D  G  V  C  T  V
     S  T  C  T  P  T  A  W  W  A  S  T  V  R  T  G  Y  A  P  *
      P  P  A  R  P  Q  P  G  G  Q  A  L  *  G  R  G  M  H  R  N


672 acagcaggacccaaggacatggtggttggctttgcaaacctgggaatacttcatgtgact
672 tgtcgtcctgggttcctgtaccaccaaccgaaacgtttggacccttatgaagtacactga
    T  A  G  P  K  D  M  V  V  G  F  A  N  L  G  I  L  H  V  T
     Q  Q  D  P  R  T  W  W  L  A  L  Q  T  W  E  Y  F  M  *  L
      S  R  T  Q  G  H  G  G  W  L  C  K  P  G  N  T  S  C  D  *


733 aagaaaaaggtatttgaaacactggaagcacggatgacagaggcgtgtattaggggctat
733 ttctttttccataaactttgtgaccttcgtgcctactgtctccgcacataatccccgata
    K  K  K  V  F  E  T  L  E  A  R  M  T  E  A  C  I  R  G  Y
     R  K  R  Y  L  K  H  W  K  H  G  *  Q  R  R  V  L  G  A  I
      E  K  G  I  *  N  T  G  S  T  D  D  R  G  V  Y  *  G  L  *


794 aatcctggacttctggtgcattctgaccttgcctatctacaagcagaaggcggaggagac
794 ttaggacctgaagaccacgtaagactggaacggatagatgttcgtcttccgcctcctctg
    N  P  G  L  L  V  H  S  D  L  A  Y  L  Q  A  E  G  G  G  D
     I  L  D  F  W  C  I  L  T  L  P  I  Y  K  Q  K  A  E  E  T
      S  W  T  S  G  A  F  *  P  C  L  S  T  S  R  R  R  R  R  P


855 cggcaactcacagacagagagaaggagatcatccgccaggcagccgtgcagcagaccaag
855 gccgttgagtgtctgtctctcttcctctagtaggcggtccgtcggcacgtcgtctggttc
    R  Q  L  T  D  R  E  K  E  I  I  R  Q  A  A  V  Q  Q  T  K
     G  N  S  Q  T  E  R  R  R  S  S  A  R  Q  P  C  S  R  P  R
      A  T  H  R  Q  R  E  G  D  H  P  P  G  S  R  A  A  D  Q  G
```

**P50 selected restriction digest, continued**
**HIV-Lock™ incorporated residues in bold**

```
916 gagatggacctgagcgtggtgcgcctcatgttcacagccttcctccctgacagcactggc
916 ctctacctggactcgcaccacgcggagtacaagtgtcggaaggagggactgtcgtgaccg
```

## Appendix A: HIV-Lock™ Structure Background Information

```
      E  M  D  L  S  V  V  R  L  M  F  T  A  F  L  P  D  S  T  G
       R  W  T  *  A  W  C  A  S  C  S  Q  P  S  S  L  T  A  L  A
        D  G  P  E  R  G  A  P  H  V  H  S  L  P  P  *  Q  H  W  Q


 977 agcttcactcggagactggagcctgtggtgtcagacgccatctatgatagcaaagccccg
 977 tcgaagtgagcctctgacctcggacaccacagtctgcggtagatactatcgtttcggggc
     S  F  T  R  R  L  E  P  V  V  S  D  A  I  Y  D  S  K  A  P
      A  S  L  G  D  W  S  L  W  C  Q  T  P  S  M  I  A  K  P  R
       L  H  S  E  T  G  A  C  G  V  R  R  H  L  *  *  Q  S  P  E


1038 aatgcatccaacctgaaaatcgtgagaatggacagaacagcaggatgtgtgacgggaggg
1038 ttacgtaggttggacttttagcactcttacctgtcttgtcgtcctacacactgccctccc
     N  A  S  N  L  K  I  V  R  M  D  R  T  A  G  C  V  T  G  G
      M  H  P  T  *  K  S  *  E  W  T  E  Q  Q  D  V  *  R  E  G
       C  I  Q  P  E  N  R  E  N  G  Q  N  S  R  M  C  D  G  R  G


1099 gaggagatttaccttctctgtgacaaggttcagaaagatgacatccagattcggttttat
1099 ctcctctaaatggaagagacactgttccaagtctttctactgtaggtctaagccaaaata
     E  E  I  Y  L  L  C  D  K  V  Q  K  D  D  I  Q  I  R  F  Y
      R  R  F  T  F  S  V  T  R  F  R  K  M  T  S  R  F  G  F  M
       G  D  L  P  S  L  *  Q  G  S  E  R  *  H  P  D  S  V  L  *


1160 gaagaggaagaaaatggcggagtttgggaaggatttggggacttttccccccacggatgtt
1160 cttctccttctttttaccgcctcaaacccttcctaaacccctgaaaaggggtgcctacaa
     E  E  E  E  N  G  G  V  W  E  G  F  G  D  F  S  P  T  D  V
      K  R  K  K  M  A  E  F  G  K  D  L  G  T  F  P  P  R  M  F
       R  G  R  K  W  R  S  L  G  R  I  W  G  L  F  P  H  G  C  S


1221 catagacagtttgccattgtcttcaaaacgccaaagtataaggatgtcaacattacaaag
1221 gtatctgtcaaacggtaacagaagttttgcggtttcatattcctacagttgtaatgtttc
     H  R  Q  F  A  I  V  F  K  T  P  K  Y  K  D  V  N  I  T  K
      I  D  S  L  P  L  S  S  K  R  Q  S  I  R  M  S  T  L  Q  S
       *  T  V  C  H  C  L  Q  N  A  K  V  *  G  C  Q  H  Y  K  A


                                            MaeI C/TAG
                                           FspBI C/TAG
                                           BcuI A/CTAGT
                                           AhlI A/CTAGT
1282 ccagcttccgtgtttgttcagcttcggaggaaatcagacctggaaactagtgaaccgaaa
1282 ggtcgaaggcacaaacaagtcgaagcctcctttagtctggacctttgatcacttggcttt
     P  A  S  V  F  V  Q  L  R  R  K  S  D  L  E  T  S  E  P  K
      Q  L  P  C  L  F  S  F  G  G  N  Q  T  W  K  L  V  N  R  N
       S  F  R  V  C  S  A  S  E  E  I  R  P  G  N  *  *  T  E  T


1343 ccctttctctactaccctgaaatcaaagacaaagaggaagtgcaaaggaaacgccagaag
1343 gggaaagagatgatgggactttagtttctgtttctccttcacgtttcctttgcggtcttc
     P  F  L  Y  Y  P  E  I  K  D  K  E  E  V  Q  R  K  R  Q  K
      P  F  S  T  T  L  K  S  K  T  K  R  K  C  K  G  N  A  R  S
       L  S  L  L  P  *  N  Q  R  Q  R  G  S  A  K  E  T  P  E  A
```

## Appendix A: HIV-Lock™ Structure Background Information

**P50 selected restriction digest, continued**
**HIV-Lock™ incorporated residues in bold**

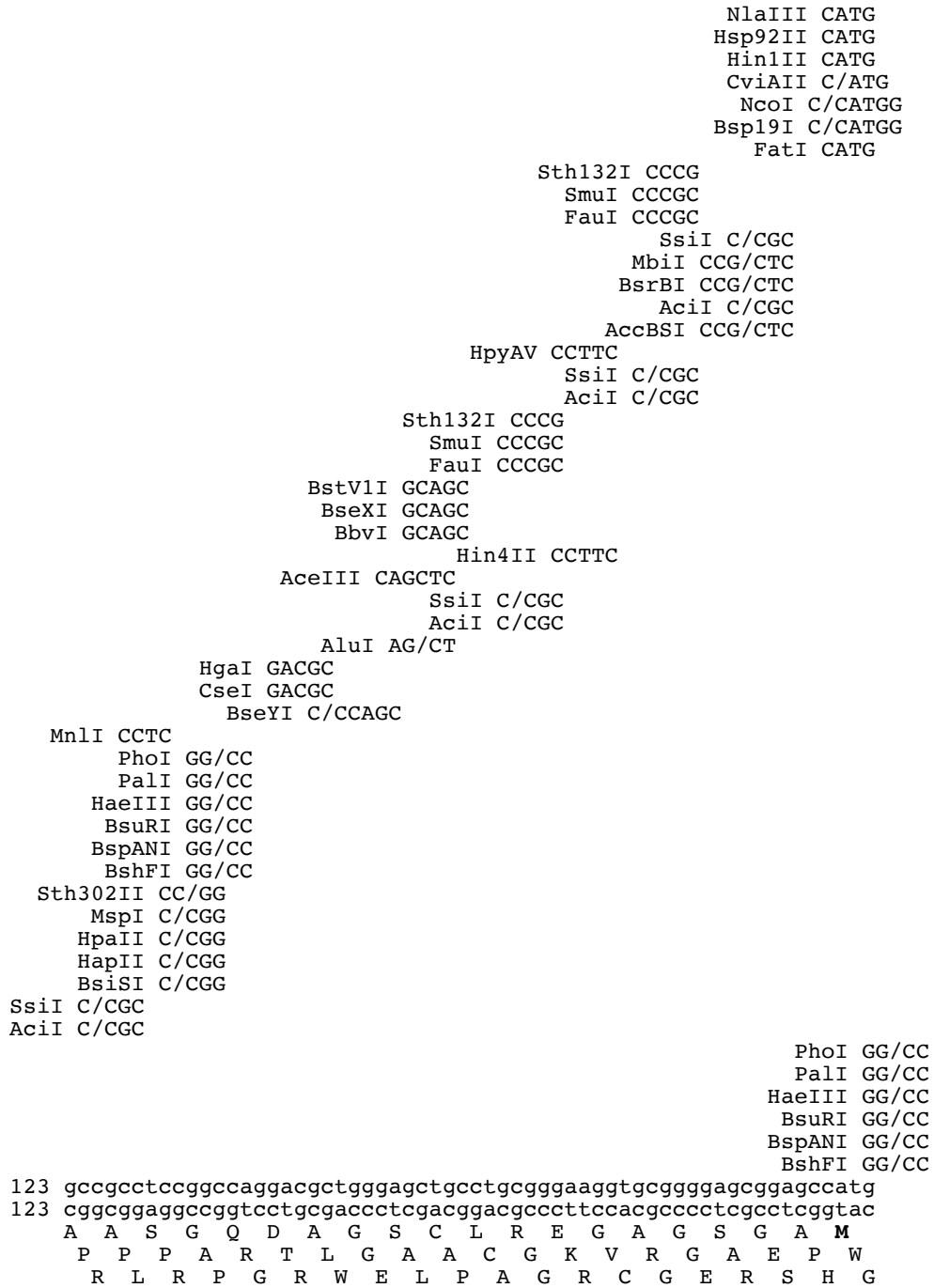
```
    1404 cttatgccgaacttctcggacagcttcggcggcggcagtggagcgggagccggtggtgga
    1404 gaatacggcttgaagagcctgtcgaagccgccgccgtcacctcgccctcggccaccacct
         L  M  P  N  F  S  D  S  F  G  G  G  S  G  A  G  A  G  G  G
          L  C  R  T  S  R  T  A  S  A  A  A  V  E  R  E  P  V  V  E
           Y  A  E  L  L  G  Q  L  R  R  R  Q  W  S  G  S  R  W  W  R


    1465 ggcatgttcggtagtggcggtggcggagggagtaccggaagccctggcccagggtatggc
    1465 ccgtacaagccatcaccgccaccgcctccctcatggccttcgggaccgggtcccataccg
         G  M  F  G  S  G  G  G  G  G  S  T  G  S  P  G  P  G  Y  G
          A  C  S  V  V  A  V  A  E  G  V  P  E  A  L  A  Q  G  M  A
           H  V  R  *  W  R  W  R  R  E  Y  R  K  P  W  P  R  V  W  L


                                                  NciI CC/SGG
                                                 EcoHI CCSGG
    EsaBC3I TC/GA
       TaqI T/CGA
    1526 tactcgaactacggatttcctccctacggtgggattacattccatcccggagtcacgaaa
    1526 atgagcttgatgcctaaaggagggatgccaccctaatgtaaggtagggcctcagtgcttt
         Y  S  N  Y  G  F  P  P  Y  G  G  I  T  F  H  P  G  V  T  K
          T  R  T  T  D  F  L  P  T  V  G  L  H  S  I  P  E  S  R
           L  E  L  R  I  S  S  L  R  W  D  Y  I  P  S  R  S  H  E
```

## Appendix A: HIV-Lock™ Structure Background Information

**c-rel Restriction Digest**
**HIV-Lock™ incorporated residues in BOLD**



## Appendix A: HIV-Lock™ Structure Background Information

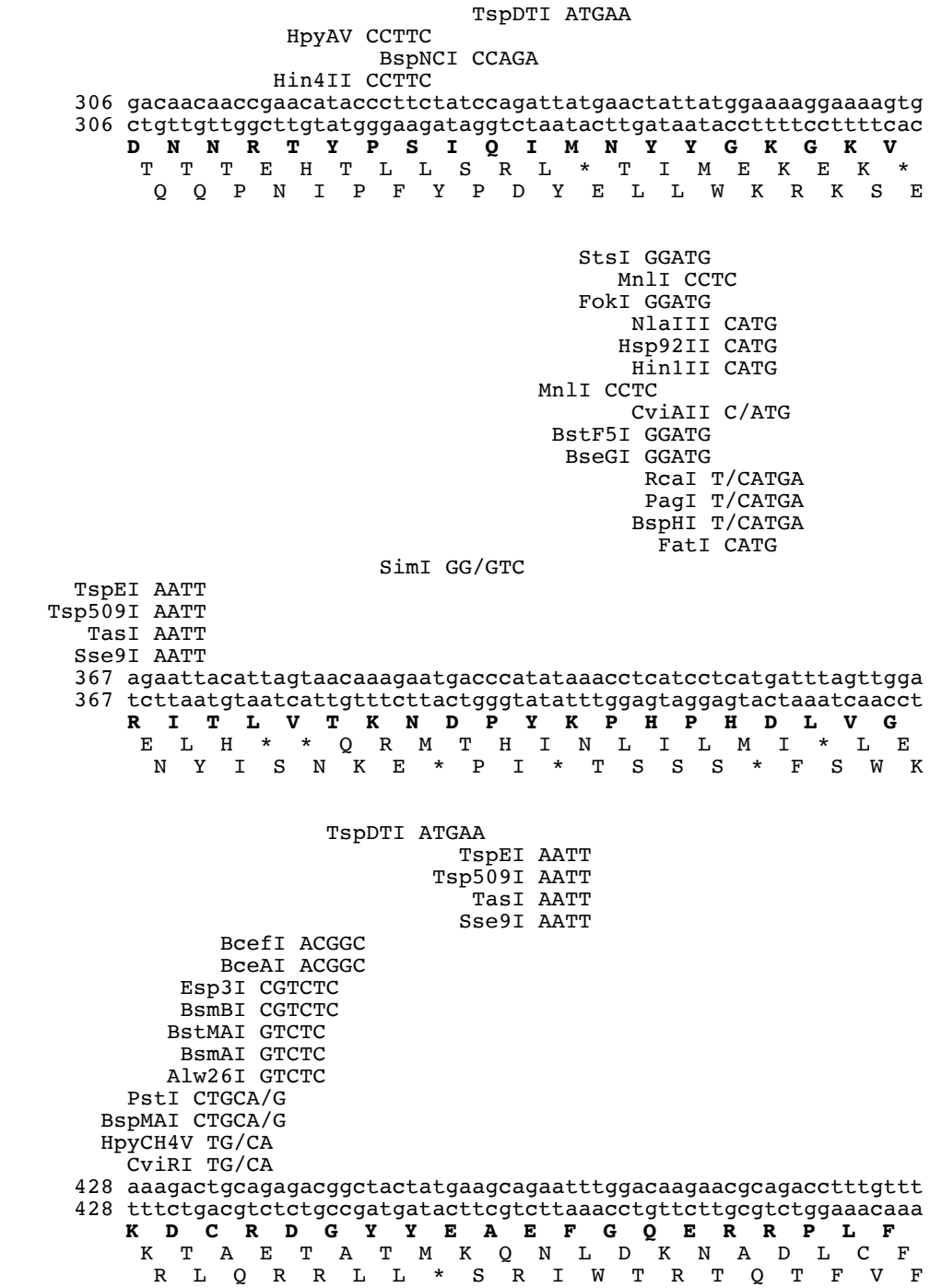
**c-rel Restriction Digest, continued**
**HIV-Lock™ incorporated residues in BOLD**

```
                                TspEI AATT
                               Tsp509I AATT
                                  TasI AATT
                                  Sse9I AATT
                 Sth132I CCCG
                   BscGI CCCGT
     MnlI CCTC
   Sth302II CC/GG
      MspI C/CGG
     HpaII C/CGG
     HapII C/CGG
     BsiSI C/CGG
                                                                PhoI GG/CC
                                                                PalI GG/CC
                                                              HaeIII GG/CC
                                                               BsuRI GG/CC
                                                              BspANI GG/CC
                                                               BshFI GG/CC
 184 gcctccggtgcgtataacccgtatatagagataattgaacaacccaggcagaggggaatg
 184 cggaggccacgcatattgggcatatatctctattaacttgttgggtccgtctcccccttac
     A  S  G  A  Y  N  P  Y  I  E  I  I  E  Q  P  R  Q  R  G  M
      P  P  V  R  I  T  R  I  *  R  *  L  N  N  P  G  R  G  E  C
       L  R  C  V  *  P  V  Y  R  D  N  *  T  T  Q  A  E  G  N  A


                               BstV1I GCAGC
                                BseXI GCAGC
                                 BbvI GCAGC
                                    PctI GAATGC
                               Mva1269I GAATGC
                                   BsmI GAATGC
                                  BsaMI GAATGC
                  HpyAV CCTTC
                          ChaI GATC
                        BstKTI GAT/C
                          MalI GA/TC
                          DpnI GA/TC
                        Sau3AI GATC
                         NdeII GATC
                          MboI GATC
                         Kzo9I GATC
                         DpnII GATC
                        BstMBI GATC
                       Bsp143I GATC
                         BfuCI GATC
                Hin4II CCTTC
                                                        PctI GAATGC
                                                     Mva1269I GAATGC
                                                         BsmI GAATGC
                                                        BsaMI GAATGC
                                                     MnlI CCTC
 245 cgttttagatacaaatgtgaagggcgatcagcaggcagcattccaggggagcacagcaca
 245 gcaaaatctatgtttacacttcccgctagtcgtccgtcgtaaggtcccctcgtgtcgtgt
     R  F  R  Y  K  C  E  G  R  S  A  G  S  I  P  G  E  H  S  T
      V  L  D  T  N  V  K  G  D  Q  Q  A  A  F  Q  G  S  T  A  Q
       F  *  I  Q  M  *  R  A  I  S  R  Q  H  S  R  G  A  Q  H  R
```

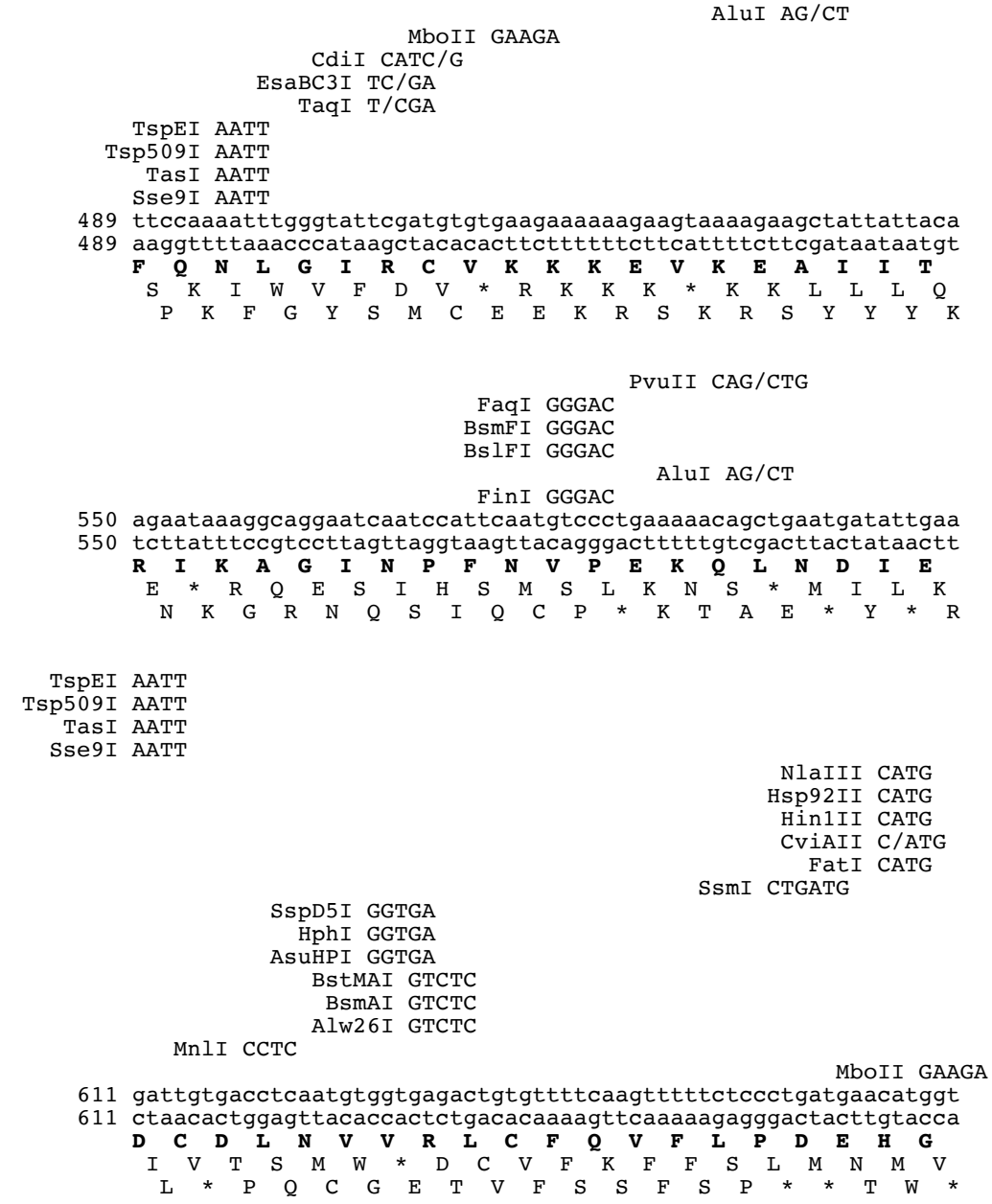
## Appendix A: HIV-Lock™ Structure Background Information

**c-rel Restriction Digest, continued**
**HIV-Lock™ incorporated residues in BOLD**



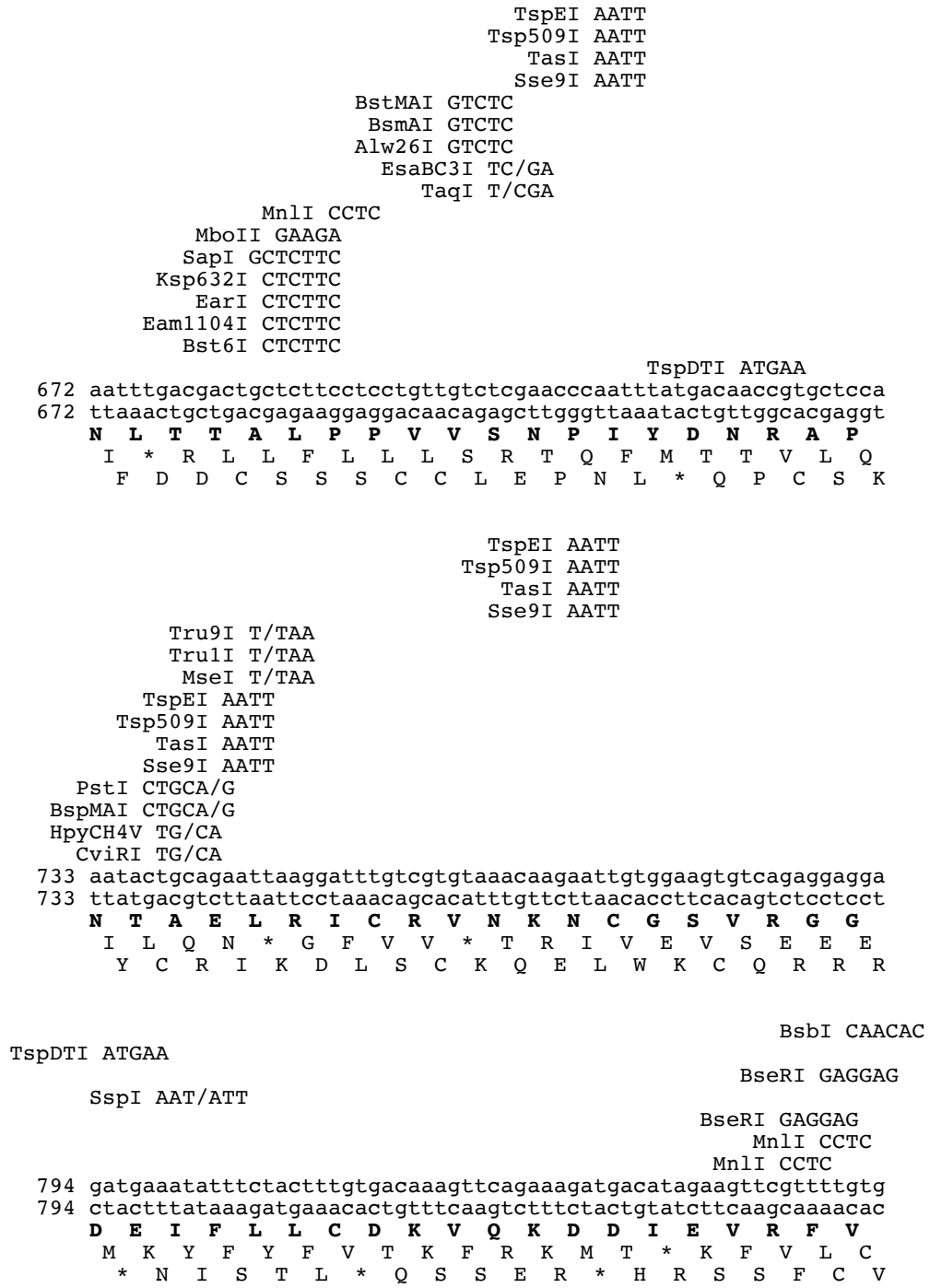
## Appendix A: HIV-Lock™ Structure Background Information

**c-rel Restriction Digest, continued**
**HIV-Lock™ incorporated residues in BOLD**



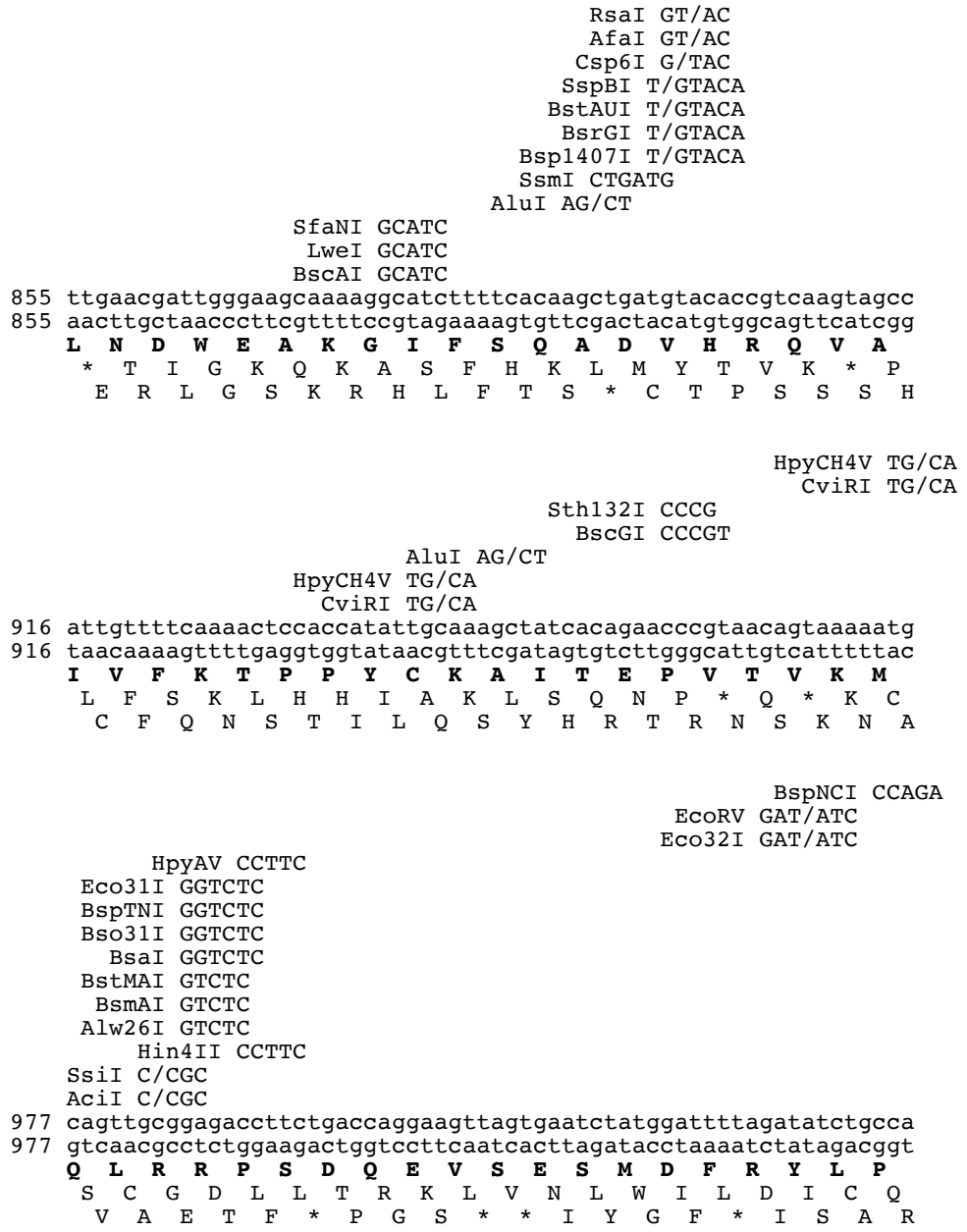
## Appendix A: HIV-Lock™ Structure Background Information

**c-rel Restriction Digest, continued**
**HIV-Lock™ incorporated residues in BOLD**



## Appendix A: HIV-Lock™ Structure Background Information

**c-rel Restriction Digest, continued**
**HIV-Lock™ incorporated residues in BOLD**



## Appendix A: HIV-Lock™ Structure Background Information

**c-rel Restriction Digest, continued**
**HIV-Lock™ incorporated residues in BOLD**

```
                                                          BspNCI CCAGA
                 BcefI ACGGC
                 BceAI ACGGC
   TspDTI ATGAA
    1038 gatgaaaaagatacttacggcaataaagcaaagaaacaaaagacaactctgcttttccag
    1038 ctacttttctatgaatgccgttatttcgtttctttgttttctgttgagacgaaaaggtc
         D  E  K  D  T  Y  G  N  K  A  K  K  Q  K  T  T  L  L  F  Q
          M  K  K  I  L  T  A  I  K  Q  R  N  K  R  Q  L  C  F  S  R
           *  K  R  Y  L  R  Q  *  S  K  E  T  K  D  N  S  A  F  P  E
```

## Appendix B: HIV-Lock™ Sequences

**Comparison of p105 (Human) and HIV-Lock™-1* (A and B; 1 and 2; Mus musculus p50 derived) and HIV-Lock™-2 (1 and 2; c-rel derived) Protein Amino Acid Sequences and Secondary Structure; C-Terminus (Cro protein derived)**

Red amino acids are not exact matches to p105 Mus Musculus
**RED** and **BLACK** bold residues are sequences differences between the C-REL HUMAN and C-REL CHICKEN sequences

```
<
p105           1 MAEDD PYLGR PEQMF HLDPS LTHTI FNPEV FQPQM ALPTA DGPYL QILEQ
HIV-Lock™-1A     MAEDD PYLGR PEQMF HLDPS LTHTI FNPEV FQPQM ALPTA DGPYL QILEQ
2D STRUCTURE                                                    EE EESS

HIV-Lock™-1B                                                  MASGA YNPYL QILEQ
C-REL HUMAN                                                   MASGA YNPYI EIIEQ
C-REL CHICK                                                   MASGA YNPYI EIFEQ
HIV-Lock™-2                                                   MASGA YNPYI EIIEQ
2D STRUCTURE                                                     EE EESB


p105          51 PKQRG FRFRY VCEGP SHGGL PGASS EKNKK SYPQV KICNY VGPAK VIVQL
HIV-Lock™-1*     PKQRG FRFRY VCEGP SHGGL PGASS EKNKK SYPQV KICNY VGPAK VIVQL
2D STRUCTURE     B SSS    EE GGG    SS   EETTT BTTB   EEEE EEES   SS E EEEEE

C-REL HUMAN      PRQRG MRFRY KCEGR SAGSI PGEHS TDNNR TYPSI QIMNY YGKGK VRITL
C-REL CHICK      PRQRG MRFRY KCEGR SAGSI PGEHS TDNNK TFPSI QILNY FGKVK IRTTL
HIV-LOCK™-2      PRQRG MRFRY KCEGR SAGSI PGEHS TDNNR TYPSI QIMNY YGKGK VRITL
2D STRUCTURE     B  SS B  EE GGG S        BSS   SSS     EE EEES   S EE EEEEE


p105         101 VTNGK NIHLH AHSLV GKHCE DGICT VTAGP KDMVV GFANL GILHV TKKKV
HIV-Lock™-1*     VTNGK NIHLH AHSLV GKHCE DGVCT VTAGP KDMVV GFANL GILHV TKKKV
2D STRUCTURE     E SSS S  B  SSEEE STTEE TTEEE EE  S S  EE EE S  EEEE   STTH

C-REL HUMAN      VTKND PYKPH PHDLV GKDCR DGYYE AEFGQ ERRPL FFQNL GIRCV KKKEV
C-REL CHICK      VTKNE PYKPH PHDLV GKDCR DGYYE AEFGP ERRVL SFQNL GIQCV KKKDL
HIV-Lock™-2      VTKND PYKPH PHDLV GKDCR DGYYE AEFGQ ERRPL FFQNL GIRCV KKKEV
2D STRUCTURE     E SSS S  B  SSEEE STTEE TTEEE EEE S   S E E     EEEE   TTHH


p105         151 FETLE ARMTE ACIRG YNPGL LVHPD LAYLQ AEGGG DRQLG DREKE LIRQA
HIV-Lock™-1*     FETLE ARMTE ACIRG YNPGL LVHSD LAYLQ AEGGG DRQLT DREKE IIRQA
2D STRUCTURE     HHHHH HHHHH H-HTT TTIII II TT  S     SSSS       HHHHH HHHHH

C-REL HUMAN      KEAII TRIKA G-INP FN--- ----- ----- ---VP EKQLN DIE-- -----
C-REL CHICK      KESIS LRISK K-INP FN--- ----- ----- ---VP EEQLH NID-- -----
HIV-LOCK™-2      KEAII TRIKA G-INP FN--- ----- ----- ---VP EKQLN DIE-- -----
2D STRUCTURE     HHHHH HHHTT T-  T T --- ----- ----- ---   TTTT   TT-- -----


p105         201 ALQQT KEMDL SVVRL MFTAF LPDST GSFTR RLEPV VSDAI YDSKA PNASN
HIV-Lock™-1*     AVQQT KEMDL SVVRL MFTAF LPDST GSFTR RLEPV VSDAI YDSKA PNASN
2D STRUCTURE     HHHHH HT  S SEEEE EEEEE EE SS S B E E   E E   E E TTT TTS

C-REL HUMAN      ----- -DCDL NVVRL CFQVF LPDEH GNLTT ALPPV VSNPI YDNRA PNTAE
C-REL CHICK      ----- -EYDL NVVRL CFQAF LPDEH GYNTL ALPPL ISNPI YDNRA PNTAE
HIV-Lock™-2      ----- -DCDL NVVRL CFQVF LPDEH GNLTT ALPPV VSNPI YDNRA PNTAE
2D STRUCTURE     ----- -TTTT TEEEE EEEEE EE SS SSB E E   E E   E EETTT TTTS
```

**Comparison of p105 (Human) and HIV-Lock™-1* (A and B; 1 and 2; Mus musculus p50 derived) and HIV-Lock™-2 (1 and 2; c-rel derived) Protein Amino Acid Sequences and Secondary Structure; C-Terminus (Cro protein derived), continued**

## Appendix B: HIV-Lock™ Sequences

Red amino acids are not exact matches to p105 Mus musculus
**RED** and **BLACK** bold residues are sequences differences between the C-REL HUMAN and C-REL CHICK sequences
BLUE residues are cro derived and do not match p50 Mus Musculus

```
p105          251 LKIVR MDRTA GCVTG GEEIY LLCDK VQKDD IQIRF YEEEE NGGVW EGFGD
HIV-Lock™-1*      LKIVR MDRTA GCVTG GEEIY LLCDK VQKDD IQIRF YEEEE NGGVW EGFGD
2D STRUCTURE        B E ES S  EESS    EEE EEES    TTT EEEEE EE    SS  E EEE B

C-REL HUMAN       LRICR VNKNC GSVRG GDEIF LLCDK VQKDD IEVRF VL--- --NDW EAKGI
C-REL CHICK       LRICR VNKNC GSVKG GDEIF ILCDK VQKDD IEVRF VL--- --DNW EAKGS
HIV-Lock™-2       LRICR VNKNC GSVRG GDEIF LLCDK VQKDD IEVRF VL--- --NDW EAKGI
2D STRUCTURE        B E ES SE EETT    EEE EEES    GGG  EEEE EE--- --TTE EEE B


p105          301 FSPTD VHRQF AIVFK TPKYK DINIT KPASV FVQLR RKSDL ETSEP KPFLY
HIV-Lock™-1*      FSPTD VHRQF AIVFK TPKYK DVNIT KPASV FVQLR RKSDL ETSEP KPFLY
2D STRUCTURE        GGG BTTTS EEEEE    TT SS    SSEEE EEEEE ETTT   B    EEEEE

C-REL HUMAN       FSQAD VHRQV AIVFK TPPYC K-AIT EPVTV KMQLR RPSDQ EVSES MDFRY
C-REL CHICK       FSQAD VHRQV AIVFK TPPFL R-DIT EPITV KMQLR RPSDQ EVSEP MDFRY
HIV-Lock™-2       FSQAD VHRQV AIVFK TPPYC K-AIT EPVTV KMQLR RPSDQ EVSES MDFRY
2D STRUCTURE        GGG EETTT EEEEE     S S-  S S EEE EEEEE EGGGT EE    EEEEE


p105          351 YPEIK DKEEV QRKRQ KLMPN FSDSF GGGSG AGAGG GGMFG SGGGG GGTGS
C-REL HUMAN       LPDEK DTYGN KAKKQ KTTLL FQKLC QDHVE TGFRH VDQDG LELLT SGDPP
C-REL CHICK       LPDEK DPYGN KAKRQ RSTLA WQKLI QDCGS AVTER PKATP IPTVN PEGKL
HIV-Lock™-1*-1    YPEIK DKEEV QRKRQ KLMPN FSDSF GGGSG AGAGG GGMEQ RITLK DYAMR
HIV-Lock™-2-1     LPDEK DKEEV QRKRQ KLMPN FSDSF GGGSG AGAGG GGMEQ RITLK DYAMR
2D STRUCTURE      E                                             EEEHH HHHHH
HIV-Lock™-1*-2    YPEIK DKEEM EQRIT LKDYA MRFGQ TKTAK DLGVY QSAIN KAIHA GRKIF
HIV-Lock™-2-2     LPDEK DKEEM EQRIT LKDYA MRFGQ TKTAK DLGVY QSAIN KAIHA GRKIF
2D STRUCTURE      E             EEE HHHHH HHHHH HHHHH HHTS  HHHHH HHHHH T  EE


HIV-Lock™-1*-1 FGQTK TAKDL GVYQS AINKA IHAGR KIFLT INADG SVYAE EVKPF PSNKKTTA
HIV-Lock™-2-1  FGQTK TAKDL GVYQS AINKA IHAGR KIFLT INADG SVYAE EVKPF PSNKKTTA
2D STRUCTURE   HHHHH HHHHH TS HH HHHHH HHH T  EEEE E TT   EEEE E  SS

HIV-Lock™-1*-2 LTINA DGSVY AEEVK PFPSN KKTTA
HIV-Lock™-2-2  LTINA DGSVY AEEVK PFPSN KKTTA
2D STRUCTURE   EEE T T  EE EEE   SS
```

P50/C-REL ALIGNMENT: Ghosh, S. et al. (1990) Cell 62, 1019–29.
PDB STRUCTURES: P50/1NFK.PDB, C-REL/1GJI.PDB, CRO/6CRO.PDB

**Appendices C – F**

**Appendix C:**

Jana R, Hazbun TR, Mollah AKMM, Mossing MC. 1997
A Folded Monomeric Intermediate in the Formation of Lambda Cro Dimer – DNA Complexes
JMB Oct 24; 273(2) 402-416
doi: 10.1006/jmbi.1997.1256
PMID: 9344748

**Appendices D – F: REDACTED**