

Editor-Communicated Paper

A Novel Genetic Algorithm for Designing Mimetic Peptides That Interfere with the Function of a Target Molecule

William Campbell^{1,2}, Laurence Kleiman³, Lajos Baranyi^{1,2}, Zhou Li³, Ahmad Khorchid³, Emiko Fujita¹, Noriko Okada¹, and Hidechika Okada^{*,1}

¹Department of Molecular Biology, Nagoya City University School of Medicine, Nagoya, Aichi 467-8601, Japan, ²Choju Medical Institute, Fukushima Hospital, Toyohashi, Aichi 441-8124, Japan, and ³Lady Davis Institute for Medical Research and McGill AIDS Centre, Jewish General Hospital, Montreal, Quebec, Canada

Communicated by Nobuyuki Kato: Received January 1, 2002. Accepted January 10, 2002

Abstract: We designed a new computer program (MIMETIC), which generates a series of peptides for interaction with a target peptide sequence. The genetic algorithm employed ranks the sequences obtained from one generation to the next by “goodness of fit” to the target. MIMETIC designed recognition peptides to various regions of HIV-1 reverse transcriptase. Among ten peptide candidates synthesized, three inhibited reverse transcription *in vitro*. TLMA2993 and PSTW1594 both targeted the connection domain of reverse transcriptase and ESLA2340 targeted the thumb domain.

Key words: Mimetic peptide, HIV-1, Reverse transcriptase, Target molecule

We designed an evolutionary software program, MIMETIC, which runs on any typical PC computer. MIMETIC generates a series of mimetic peptide sequences with potential to interact with a target peptide by comparing several physico-chemical parameters of each pair of the complementary peptides being analyzed. When the genetic algorithm has evolved through 5,000 iterations, the program ranks the generated peptide sequences by “goodness of fit” to the target. If generated peptides can interfere with the function of a target region in a protein, rational design using these sequences as templates could be a useful approach for drug design. Therefore, we attempted to generate mimetic peptides to recognize biologically active regions of HIV-1 reverse transcriptase (RT) expecting that some of these might interfere with RT function. RT is an intensively studied molecule, and both its sequence and 3 dimensional structure are known (8, 10). RT is a heterodimer composed of p66 and p51 subunits, the latter being a truncated version of p66 and lacking the RNase H domain. The resemblance of the enzyme to a right hand has led to the division of each RT chain into structural subdomains known

as the fingers, palm, thumb, connection and RNase H subdomains (8). The active catalytic site for DNA polymerization is in the palm domain, and the thumb, palm, and finger domains serve to grip the RNA/DNA template. The connection domain serves to connect these domains to the RNase H domain at the carboxy end of p66, and is also involved in the p66/p51 interaction. Based on this information, we selected the palm regions consisting of amino acids 96–115 and 178–191 which contain the polymerase catalytic site amino acids, D110, Y181, and Y188 (3, 11) as well as the YMDD motif which is highly conserved in many DNA polymerases (3). The second target sequence was the thumb domain consisting of amino acids 283–302 which contains two LT pairs believed to interact with the DNA/RNA template (1). The third target region was in the connection domain comprising amino acids 384–413 where mutation affected packaging (12) and genomic placement (6) of tRNA^{Lys3} in the virion.

*Address correspondence to Dr. Hidechika Okada, Department of Molecular Biology, Nagoya City University School of Medicine, Mizuho-cho, Mizuho-ku, Nagoya, Aichi 467-8601, Japan. Fax: 052-842-3460. E-mail: hiokada@med.nagoya-cu.ac.jp

Abbreviations: DCM, dichloromethane; DMF, *N,N*-dimethylformamide; DMSO, dimethyl sulfoxide; EDT, 1,2-ethanedithiol; HBTU, 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HOBt, hydroxybenzotriazole; IC50, 50% inhibitor concentration; NMM, *N*-methylmorpholine; NMP, *N*-methylpyrrolidone; RT, reverse transcriptase; TFA, trifluoroacetic acid.

Materials and Methods

Design of mimetic peptides. We used the evolutionary software program MIMETIC to design mimetic peptide sequences for interaction with target regions in RT. This proprietary software employs a genetic algorithm that generates a series of increasingly optimized peptides to a target by random alteration of amino acids for 5,000 generations. Every peptide sequence generated in this manner is assigned a score based on several physicochemical parameters including: hydrophobic complementarity optimization; average structural similarity optimization; minimization of bulky side chain interference and backbone alignment. Following the final generation, the program re-arranges the peptides into a list according to a scoring method for "goodness of fit" to the target. We synthesized peptides with the highest score and tested their ability to inhibit reverse transcription. We designated each mimetic peptide by the one letter code for the first four amino terminal amino acids followed by its molecular weight.

Peptide synthesis. We synthesized peptides using an AMS 422 Multiple Peptide Synthesizer (ABiMED, Langenfeld, Germany) using standard solid-phase synthesis techniques and 9-fluorenylmethoxycarbonyl (Fmoc) amino acids (Watanabe Chemical Industries Ltd., Hiroshima, Japan). *In situ* activation was by 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) with *N*-methylmorpholine (NMM), and hydroxybenzotriazole (HOBT) as catalytic bases. Amidated peptides (25 μ mol scale) were prepared using Fmoc-PAL-PEG-PS resin (PerSeptive Biosystems, Warrington, U.K.). All residues were double coupled in *N,N*-dimethylformamide (DMF; Peptide synthesizer grade; Watanabe Chem. Ind., Ltd.), *N*-methylpyrrolidone (NMP) and dichloromethane (DCM). Deprotection was achieved using 20% piperidine in DMF. Peptides were then cleaved from the resin, with concomitant removal of side-chain protecting groups by treatment with trifluoroacetic acid (TFA; 80%), thioanisole (12%), 1,2-ethanedithiol (EDT; 6%), and *m*-cresol (2%). After cleavage, the peptides were precipitated with 2 volumes of cold ether for 10 min, collected by centrifugation, washed two times with cold ether and left to dry overnight. Purification was carried out by reversed-phase HPLC (Waters 741 Data Module; Waters 484 Tunable Absorbance Detector; Waters 600E System Controller). Samples of crude peptide were chromatographed on a Waters Delta-PakTM C₁₈ column (40 \times 100 mm, 15 μ m, 100 Å particles) with linear gradient: Milli-Q water/acetonitrile in 0.1% TFA. Peptide mass was confirmed using electrospray mass spectrometry

(Kompact Maldi II, Kratos Analytical, Shimadzu, Japan). Lyophilized peptides were stored desiccated at -30 C.

Virus isolation from infected cells and viral RNA isolation. The monocytic cell line, PLB, chronically infected with the HIV_{III}B strain, was maintained through the addition at 3-4 day intervals of equal numbers of uninfected cells (cell suspension of 2×10^6 cells/ml), in RPMI 1640 medium (GIBCO Products Ltd., Quebec, Canada) supplemented with 10% heat-inactivated fetal calf serum, 2 mM L-glutamine, 250 U/ml of penicillin, and 250 μ g/ml of streptomycin. Virus was isolated from the cell culture medium by first centrifuging the supernatant in a Beckman GS-6R rotor at 3,000 rpm for 30 min, and the virus was then pelleted from the resulting supernatant by centrifuging in a Beckman Ti45 rotor at 35,000 rpm for 1 hr. The viral pellet was then purified by centrifugation at 26,500 rpm for 1 hr through 15% sucrose onto a 65% sucrose cushion, using a Beckman SW41 rotor.

Total viral RNA was extracted from viral pellets using the guanidinium isothiocyanate procedure (2). The pellets were dissolved in 5 mM Tris buffer, pH 7.5.

Analysis of reverse transcription. The unspliced viral genomic RNA in total viral RNA was determined by the ribonuclease protection assay, as previously described (5, 9). Approximately 0.5×10^8 molecules of viral genomic RNA annealed *in vivo* with tRNA^{Lys3} were incubated for 15 min at 37 C in 20 μ l RT buffer (50 mM Tris-HCl pH 7.5, 60 mM KCl, 3 mM MgCl₂, 10 mM DTT) containing 50 ng purified HIV-RT, 10 units RNase, and various dNTP's. The first 6 sequentially incorporated bases are, CTGCTA. tRNA^{Lys3} was extended 1 base by the addition of 5 μ Ci ³²P-dCTP (Du Pont, 3,000 μ Ci/mmol, 10 μ Ci/ml) as the only dNTP added. To extend the tRNA^{Lys3} by 6 bases, the reaction mix contained 0.2 mM dCTP, 0.2 mM dTTP, 5 μ Ci ³²P-dGTP (Du Pont, 3,000 μ Ci/mmol, 10 μ Ci/ml), and 0.05 mM ddATP. The extended primer was ethanol precipitated, resuspended and analyzed on 6% polyacrylamide, 7 M urea, 1 \times TBE. One μ l of peptide inhibitor, in 75% DMSO, was added to the final reaction volume of 20 μ l. We observed a prominent labeled species moving more slowly than 1-base extended tRNA^{Lys3}. This species migrates as a 4-base extended tRNA^{Lys3}, and most likely results from the addition of C to tRNA^{Lys3} already extended in the virion by the first 3 bases (i.e., CTG). We have previously reported that tRNA^{Lys3} in the virus produced from COS cells is present in both unextended and 2-base extended forms (7). In the preparations of PLB-produced HIV-1 used here, apparently 3-base extended forms exist as well. The reverse transcriptase-extended tRNA^{Lys3} products were resolved on 1D PAGE, and the electrophoretic patterns, visualized by autoradiography. The inhibitory effect of the peptide

Table 1. Target peptides in RT and generated mimetic peptides

Target peptide	Subdomain	Mimetic peptide (RT region targeted in brackets)	Abbreviation	Inhibition
96HPAGLKKKKSV TVLDVGDAY115	palm	MWATELILISDSDEVDSIQM (96 115)	MWAT2299	no
178PDIVIQYMD DLYVGSdleI191	palm	RFYPHIHHVGVKSDIEVY (178 191)	RFYP2448	no
283LRGKALTEV IPLTEEALEL302	thumb	ESLALYKSLQQSEMILLELEL (283 302)	ESLA2340	yes
		ESLALYKSLQ (283 292)	ESLA1153	no
		QSEMILLELEL (293 302)	QSEM1205	no
383WGKTPKFKLPIQKETWE TWWTEYWQATWIPE413	connection	TLMALELKGKLLLAGLAPSAFLPLSFPEL (383 413)	TLMA2993	yes
		PSTTPTFLKFQLK (383 395)	PSTT1508	no
		PSTWPTFLKFQLK (383 395)	PSTW1594	yes
		IPARLGHMFMLRRVGL (398 413)	IPAR1900	no
350KTGKYARMRGAHTN363	connection	LSATMAAAAASMS (350 363)	LSAT1256	no

was not affected by the order in which the inhibitor and other reagents were added.

Results

Mimetic Peptides Generated to Targets in RT

The best-fit peptide sequences were generated by MIMETIC and their target regions in RT are listed in Table 1. Each peptide is designated by the one letter code for the first four amino terminal amino acids followed by its molecular weight. MWAT2299 and RFYP2448 were generated to the palm regions (amino acids 96 115) and (amino acids 178 191), respectively. ESLA2340 was generated to the thumb domain (amino acids 283 302) and ESLA1153 and QSEM1205 were generated to subregions of the same domain (amino acids 283 292) and (amino acids 293 302) respectively. TLMA2993 was generated to the connection domain (amino acids 383 314) and PSTT1508, PSTW1594 and IPAR1900 were generated to subregions of the same domain (amino acids 383 395 or 398 413). LSAT1256 was generated to another connection domain (amino acids 350 363).

Inhibition of Reverse Transcription by Mimetic Peptides

We tested the mimetic peptides listed in Table 1. In the control lane (Fig. 1), the reaction is carried out by the addition of the peptide solvent buffer, containing 75% DMSO, and the major, 6-base long extension of the tRNA^{Lys3} can be seen indicating that DMSO does not interfere with the reaction. In the other lanes, the reaction mixture contained 2 µg/20 µl reaction volume of mimetic peptide or control peptide. We found that only

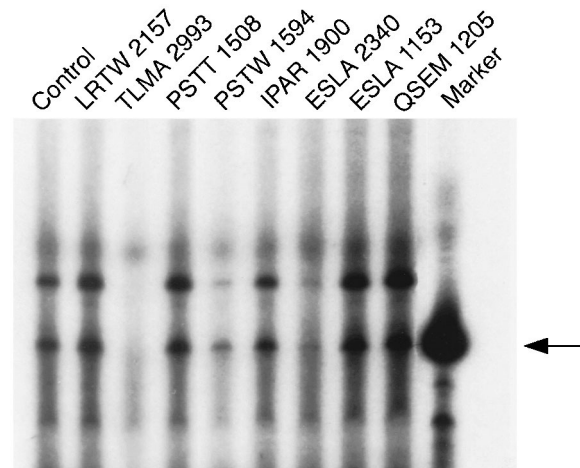


Fig. 1. Inhibition of reverse transcription by mimetic peptides. Total RNA was extracted from PLB-derived virus and included *in vivo* placed tRNA^{Lys3} as the source of primer. Reverse transcription was carried out in the presence of dCTP, dTTP, α -³²P-dGTP, and ddATP. The first 6 bases incorporated are CTGCTA, and the major product is a 6-base extension of the tRNA^{Lys3} due to the presence of ddATP and the absence of dATP. The control is no added peptide. LRTW2157 is an unrelated peptide from complement component C3a. The marker lane is purified tRNA^{Lys3} extended 6 bases in the *in vitro* reverse transcription reaction (without inhibitor present), using this tRNA annealed to the primer binding site in synthetic HIV genomic RNA as the source of primer/template.

TLMA2993, ESLA2340 and PSTW1594 inhibited RT activity (Fig. 1). RFYP2448 and MWAT2299 (targeting the connection domain), and LSAT1256 (targeting the palm domain) did not affect reverse transcription in this assay (data not shown). TLMA2993, a 30 amino acid long mimetic peptide, was the strongest RT inhibitor

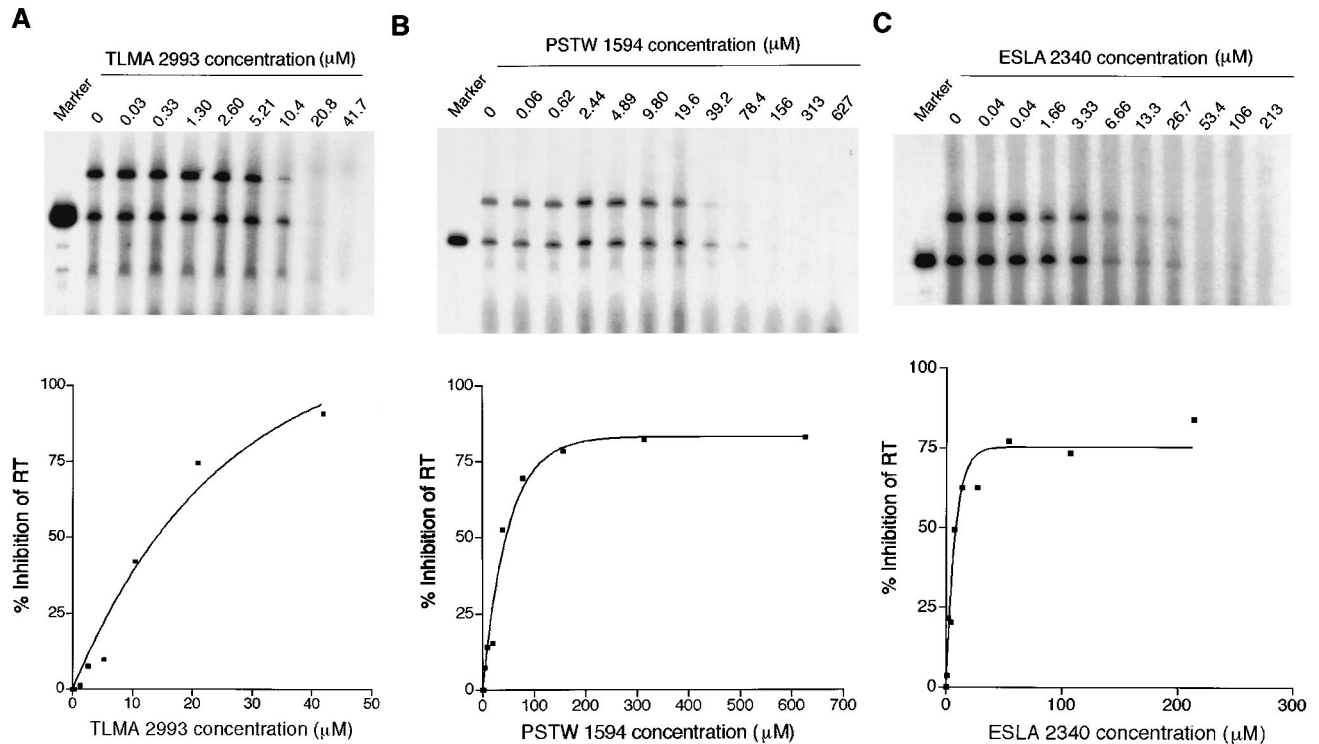


Fig. 2. Inhibition of RT by increasing doses of mimetic peptides. The reaction conditions are identical to those used for the experiments in Fig. 1 with total HIV-1 RNA used as a source of primer/template. Panels A, B and C represent reactions run in the presence of increasing concentrations of TLMA2993, PSTW1594, and ESLA2340, respectively. The upper portion of each panel shows the electrophoretic pattern of the reverse transcription products, and the lower portion of each panel shows the quantitation of these results using phosphor-imaging. The marker lane represents the reverse transcription reaction run under similar conditions, minus inhibitor, and using purified tRNA^{Lys3} annealed to the primer binding site in synthetic HIV genomic RNA as the source of primer/template.

and completely inhibited reverse transcription. This peptide was designed to complement the amino acid sequence in the connection domain starting with amino acid 383 and terminating with amino acid 413. This region includes beta sheet 19, alpha helix L, and beta sheet 20 of RT. Peptides targeting subregions covered by TLMA2993 were tested. IPAR1900 targets the RT sequence from amino acid 398 to amino acid 413, and proved to be inactive as an inhibitor. PSTT1508 and PSTW1594 target the RT sequence connection domain from amino acid 383 to 395, and differ from each other only by the replacement of the fourth amino acid, 387T in PSTT1508 with W in PSTW1594. While PSTT1508 was inactive, PSTW1594 significantly inhibited reverse transcription, although not as efficiently as the larger TLMA2993. Mimetic peptides targeting subregions covered by ESLA2340 were also generated (ESLA1153 and QSEM1205) but these had no effect on RT activity. TLMA2993, PSTW1594, and ESLA2340 inhibited RT activity *in vitro* in a dose dependent manner (Fig. 2). The upper portion of each panel shows the electrophoretic patterns of reverse transcription products with increasing amounts of inhibitor, using total viral RNA as the source of primer/template and the lower portion of each panel

shows the quantitation of these results using phosphor imaging. TLMA2993 shows an IC₅₀ of approximately 12 μM , and the maximum inhibition shown (close to 100%) is obtained at a molar excess of TLMA2993 over p66/p51 RT of approximately 1,000. Inhibition by PSTW1594, which was generated to the first half of the TLMA2993 target sequence, occurs at an IC₅₀ closer to 50 μM , and does not achieve as high an inhibition. The IC₅₀ for ESLA2340 is similar to that of TLMA2993, but the level of maximum inhibition is also lower.

Discussion

We synthesized ten mimetic peptides generated by MIMETIC to target regions within RT presumed to be active from structure-function studies. Of the ten peptides, three were able to inhibit reverse transcription *in vitro*. Close to complete inhibition of reverse transcription by TLMA2993, the strongest of the three inhibitors tested, occurred at a 50% inhibitor concentration (IC₅₀) of approximately 32 μM . For comparison, under similar conditions, the nucleoside analogues AZTTP and ddATP give maximal inhibition at approximately 80 μM , while the non-nucleoside drugs nevirapine and delavirdine

give maximal inhibition at 0.32 and 0.64 μM , respectively (4). Therefore, the capacity of TLMA2993 to inhibit RT *in vitro* is comparable to nucleoside analogs although it is much lower than that of non-nucleoside inhibitors. We have found that 2–3 amino acid insertions into the thumb or connection subdomains of RT result in diminished HIV-1 replication in cells due to blocking of tRNA^{Lys3} and viral precursor protein Pr160 gag-pol incorporation into the virions during their assembly (12). This indicates that the regions targeted by mimetic peptides may not only interfere with enzyme activity, but with the packaging of viral components during assembly, and that a pleiotropic effect of the mimetic peptides may be expected. In summary, our novel genetic algorithm has the potential for generating peptides which interact with a target peptide sequence.

This work was supported by grants from The Organization for Pharmaceutical Safety and Research (OPSR) of Japan and the Medical Research Council of Canada.

References

- 1) Basu, A., Ahluwalia, K.K., Basu, S., and Modak, M.J. 1992. Identification of the primer binding domain in human immunodeficiency virus reverse transcriptase. *Biochemistry* **31**: 616–623.
- 2) Chomczynski, P., and Sacchi, N. 1987. RNA isolation from cultured cells. *Anal. Biochem.* **162**: 156–159.
- 3) Delarue, M., Pock, V., Tordo, N., Moras, D., and Argos, P. 1990. An attempt to unify the structure of polymerases. *Protein Eng.* **3**: 461–467.
- 4) Gu, Z., Quan, Z.Li., Arts, E.J., and Wainberg, M.A. 1995. Effects of non-nucleoside inhibitors of Human Immunodeficiency Virus Type 1 in cell-free recombinant reverse transcriptase assays. *J. Biol. Chem.* **270**: 31046–31051.
- 5) Huang, Y., Kohrchild, A., Gabor, J., Wang, J., Li, X., Darlix, J., Wainberg, M.A., and Kleiman, L. 1998. The role of nucleocapsid and U5 stem/A-rich loop sequences in tRNA^{Lys} genomic placement and initiation of reverse transcription in human immunodeficiency virus type 1. *J. Virol.* **72**: 3907–3915.
- 6) Huang, Y., Kohrchild, A., Wang, J., Parkiak, M.A., Darlix, J., Wainberg, M.A., and Kleiman, L. 1997. Effect of mutations in nucleocapsid protein (NCp7) upon Pr160^{gag-pol} and tRNA^{Lys} incorporation into HIV-1. *Virol.* **71**: 4378–4384.
- 7) Huang, Y., Wang, J., Shalom, A., Li, Z., Kohrchild, A., Wainberg, M.A., and Kleiman, L. 1997. Primer tRNA^{Lys} on the viral genome exists in unextended and two-base extended forms within mature human immunodeficiency virus type 1. *J. Virol.* **71**: 726–728.
- 8) Jacobo-Molina, A., Ding, H., Nanni, R.G., Clark, A.D., Jr., Lu, X., Tantillo, C., Williams, T.L., Kamer, G., Ferris, A.L., Clark, P., Hizi, A., Hughes, S.H., and Arnold, E. 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 6320–6324.
- 9) Kaye, J.F., Richardson, J.H., and Lever, A.M. 1995. Cis-acting sequences involved in human immunodeficiency virus type 1 RNA packaging. *J. Virol.* **69**: 6588–6592.
- 10) Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., and Steitz, T.A. 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**: 1783–1790.
- 11) Larder, B.A., Purifoy, D.J., Powell, K.L., and Darby, G. 1987. Site-specific mutagenesis of AIDS virus reverse transcriptase. *Nature* **327**: 716–717.
- 12) Mak, J., Kohrchild, A., Cao, Q., Huang, Y., Lowy, I., Parkiak, M.A., Parasad, V.R., Wainberg, M.A., and Kleiman, L. 1997. Effects of mutations in Pr160^{gag-pol} upon tRNA^{Lys3} and Pr160^{gag-pol} incorporation into HIV-1. *J. Mol. Biol.* **265**: 419–431.