I. Comparison of translesion bypass of guanine–N2 monoadducts of mitomycin C and guanine-N7 monoadducts of 2,7-diaminomitosene by T7 exo-, Klenow exo-, eta and Klenow exo+ DNA polymerases.

II. Structure-based design, synthesis, structure-conformation and structure-activity relationships studies of D-Phe-Pro-D-Arg-P1'-CONH2 tetrapeptides with inhibitory activity for thrombin.

By

Cristina C. Clement

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2006

UMI Number: 3213273

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Biochemistry in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy.

5-2-2006	**Manfred Philipp, Ph.D.**
Date	**Chair of Examining Committee**
5-2-2006	**Lesley Davenport, Ph.D.**
Date	**Executive Officer**

Gary Quigley, Ph.D._____

Richard Magliozzo, Ph.D._____

Charlotte Russell , Ph.D._____

Horst Schultz, Ph.D. _____

Maria Tomasz,Ph.D._____

Supervisory Committee

The Graduate Center of The City University of New York

ABSTRACT

I. Comparison of translesion bypass of guanine–N2 monoadducts of mitomycin C and guanine-N7 monoadducts of 2,7-diaminomitosene by T7 exo-, Klenow exo-, eta and Klenow exo+ DNA polymerases.

II. Structure-based design, synthesis, structure-conformation and structure-activity relationships studies of D-Phe-Pro-D-Arg-P1'-CONH2 tetrapeptides with inhibitory activity for thrombin.

By

Cristina C. Clement

Adviser Part I: Professor Maria Tomasz

Adviser Part II: Professor Manfred Philipp

The guanine (G)-N^2 DNA monoadduct of mitomycin C (MC), a cytotoxic anticancer drug, inhibits translesion bypass by DNA polymerases. 2,7-Diaminomitosene (2,7-DAM) is the major metabolite of MC in tumor cells, generated by the reduction of MC. 2,7-DAM alkylates DNA in the cell *in situ*, forming an adduct at the N7 position of 2'-

deoxyguanosine (2,7-DAM-dG) and is noncytotoxic. In part I of this study we tested a potential correlation between the lack of cytotoxicity of 2,7-DAM and the relative ease of bypass of this adduct as compared with the MC adduct. 24-mer and 27-mer templates, adducted at a single guanine either with MC or 2,7-DAM were synthesized and submitted to extension of primers by T7 exo-, Klenow exo-, Klenow exo+, and eta DNA polymerases. The G-N7-2,7-DAM adduct was bypassed by all four polymerases, resulting in the production of a fully extended primer. In sharp contrast, the G-N^2-MC monoadduct was not bypassed beyond the adduct position under the same conditions by any of the four polymerases. In parallel experiments in cell free systems, template oligonucleotides containing a single 2,7-DAM-dG-N7 adduct directed selective incorporation of cytosine in the 5'-^{32}P-labeled primer strands opposite the adducted guanine, catalyzed by Klenow (exo-) DNA polymerase. These results showed for the first time that the dG-N7-2, 7-DAM lesion is non-mutagenic in cell-free systems.

In part II of this research structure-based design and molecular docking were employed to design in silico libraries of peptides as potential reversible inhibitors of thrombin. The candidate inhibitors were selected from two original classes of amino acids sequences (1)-D-Phe-Pro-Arg (P1)-D-Pro(P1')-P2'-P3'-$CONH_2$ and (2)-D-Phe-Pro-D-Arg(P1)-P1'-P2'-P3'-$CONH_2$. For the first time in the field of peptides inhibitors for thrombin we showed that the presence of D-Pro at P1' Position and the use of D-Arg instead of L-Arg at P1 Position is responsible for inhibiting hydrolysis of these of peptides by thrombin, causing these sequences to be inhibitors. *In vitro* kinetics of thrombin inhibition showed a specific structure-activity relationship at P1' position in the peptide sequence space (2)-D-Phe-Pro-D-Arg(P1)-P1'-$CONH_2$. The lead peptides (D-

Phe-Pro-D-Arg-D-Ala-$CONH_2$, D-Phe-Pro-D-Arg-D-Thr-$CONH_2$, D-Phe-Pro-D-Arg-D-Cys-$CONH_2$, D-Phe-Pro-D-Arg-D-Ser-$CONH_2$ had competitive or mixed inhibition with respect to thrombin and are characterized by inhibitory constant in the 20-0.8 micromolar range.

ACKNOWLEDGEMENTS

I would like to take the opportunity to express my gratefulness to my primary mentor Professor Manfred Philipp for his guidance, encouragement and support throughout all five years I spent in his lab working on peptides inhibitors of thrombin. Without him, this thesis and specifically the original contribution to the field of reversible peptide inhibitor of thrombin would never have been accomplished.

I am also indebted to my second mentor, Professor Maria Tomasz, who was supporting my work in the field of translesion synthesis of mitomycin C bulky adducts during six years of thesis in her lab. Without her encouragement and support the work on DNA polymerases and translesion synthesis of the mitomycin C adducts would never have been accomplished.

I also like to thank my committee members Professor Horst Schultz, Professor Charlotte Russell, Professor Gary Quigley and Professor Magliozzo for their support and suggestions during the committee meetings.

My thank goes also to Professor Tomasz alumni, Dr. Suresh Kumar and Dr. Arun Das who were supervising my work in her lab during the first two years. I am grateful especially to their training in chemical synthesis of different adducted oligonucleotides with mitomycin C and in developing methods for high pressure liquid chromatography separation of different DNA adducts.

I thank also to Dr. Clifford E. Soll and Dr. Michael Blumenstein for their support on mass and NMR spectroscopy, respectively.

I also very much indebted to Professor Magliozzo and to his lab that allowed me to perform the isothermal titration calorimetry experiments and to Professor Max Diem

and his graduate student Christian Matthaeus who allowed and supervised my experiments on Fourier transformed infrared spectroscopy on peptides-solid films.

Professor Kennelly and his lab at Lehman College are acknowledged for their support with their electrospray ionization mass spectroscopy on peptides.

Professor Dixie Goss lab at Hunter College is acknowledged for the support with their spectropolarimeter which I was allowed to use for the circular dichroism experiments.

I am acknowledging all the financial support from Chemistry Department at Lehman College who was appointed me Adjunct Lecturer in Chemistry through all my thesis work in Professor Philipp lab.

I am especially recognizing the contribution of Julian Gingold to the docking experiments performed on thrombin and different peptides inhibitors. He started the original in silico library of peptides inhibitors for thrombin under the supervision of Professor Philipp and further continued under my supervision to develop the lead compounds using advanced docking softwares.

Many thanks go also to the undergraduate students from Professor Philipp lab who was synthesizing part of peptide libraries, especially to Juan Barquero, Rafael Alcantara, Patricia Chimezie and his graduate student, Janet Gonzalez.

At last, but not the least, I am truly grateful to my parents and my brother for their support and encouragements.

TABLE OF CONTENTS

LIST OF SCHEMES

PART I

PART II

LIST OF TABLES

PART I

PART II

LIST OF FIGURES

PART I

PART II

ABBREVIATIONS

PART I

***A*: absorbance**

bp: base pair;

2,7-DAM: 2,7-diaminomitosene;

dG: deoxyguanosine;

DMS: dimethyl sulfate;

E: enzyme;

ESI-MS (-), electrospray ionization mass spectroscopy (negative mode);

HPLC: high pressure liquid chromatography

LC-MS: liquid chromatography-mass spectroscopy;

MC: mitomycin C;

nt: nucleotide;

PAGE, polyacrylamide gel electrophoresis;

PNK, T4 polynucleotide kinase;

S, substrate;

SAR: structure-activity relationship;

SVD: snake venom phosdiesterase

TE buffer: 1 mM Tris, 1 mM EDTA (pH 7.4); TN buffer, 10 mM Tris, 10 mM NaCl;

TEAA buffer: triethyl ammonium acetate

TLS: translesion synthesis;

UV-VIS spectroscopy-ultraviolet-visible spectroscopy

ABBREVIATIONS

PART II

BSA: bovine serum albumin

CD: circular dichroism

2D-NOESY: 2-D-transferred NOE

DMF: dimethyl formamide

FTIR: fourrier transformed infrared spectroscopy

HPLC: high pressure liquid chromatography

ITC: isothermal titration calorimetry

Ki: inhibitory constant

LPC: ligand-protein-contact (software).

PDB: Protein data bank

PPACK: phenyl-prolyl-argine-chloromethylketone

SAR: structure-activity relationship

S2238: [H-D-Phenylalanyl-L-pipecolyl-L-Arginine-p-nitro-anilide dihydrochloride (H-D-Phe-Pip-Arg-pNA.2HCl)]: substrate for thrombin

PART I

Comparison of translesion bypass of guanine–N2 monoadducts of mitomycin C and guanine-N7 monoadducts of 2,7-diaminomitosene by T7 exo-, Klenow exo-, eta and Klenow exo+ DNA polymerases

By

Cristina C. Clement

ABSTRACT

The guanine (G)-N^2 DNA monoadduct of mitomycin C (MC), a cytotoxic anticancer drug, inhibits translesion bypass by DNA polymerases as shown previously. The non-cytotoxic MC metabolite 2,7-diaminomitosene (2,7-DAM) forms a G-N7 DNA monoadduct in vitro and in vivo. We tested a potential correlation between the relative ease of bypass of this adduct (as compared with the MC adduct) and the lack of cytotoxicity of 2,7-DAM. 24-mer, 27-mer and 36-mer DNA templates, adducted at a single guanine either with MC or 2,7-DAM were synthesized and submitted to extension of primers by T7 exo-, Klenow fragment, eta (η) and Klenow (exo+ polymerases). In the 27-mer template/15-mer primer and 24 mer template/15 mer primer systems the G-N7-2,7-DAM adduct was bypassed by all four polymerases, resulting in the production of a fully extended primer. However, the extension was at a slower rate as compared with the control, non-alkylated template; build-up of abortive products before the lesion was observed suggesting that the polymerases are loosing the processivity as they encounter the bulky adducts. The fully extended primer and the abortive products each represented 15-20 % of expected 100 % extended primer, observed with control template. Each of the four polymerase had different extension efficiency: klenow exo- had the highest efficiency of bypassing the lesion followed by T7 exo- with the lowest efficiency being observed for klenow exo+ and eta polymerases. In all cases substrate: enzyme ratios (S:E) of 1: 1, 2:1 and 3:1 were allowing a higher efficiency for translesion synthesis of both bulky adducts than S:E ratios of 5:1 or higher. In sharp contrast, the G-N^2-MC monoadduct was not bypassed beyond the adduct position under the same condition by none of the four polymerases. These results suggest a positive correlation between

cytotoxicity and inhibition of DNA synthesis in the case of the two mitomycin monoadducts.

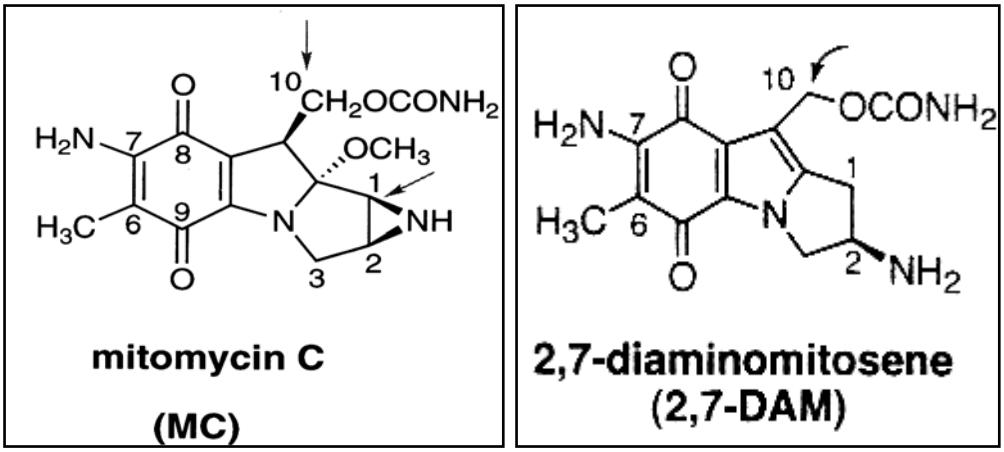
CHAPTER 1

Mitomycin C adducts with DNA: Structure-Activity Relationship

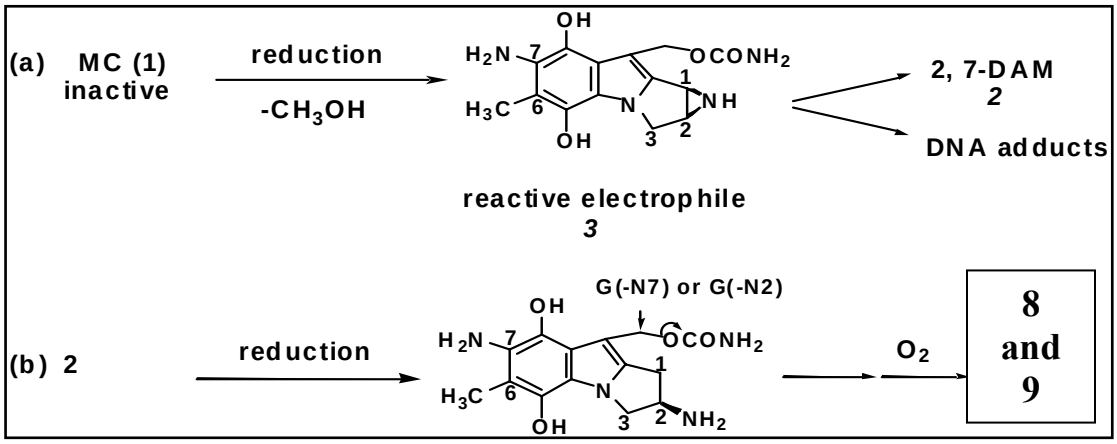
INTRODUCTION

Mitomycin C (MC-**1**, Scheme 1) is an antitumor antibiotic used clinically as a chemotherapeutic agent against several types of cancer (1). 2,7-Diaminomitosene (2,7-DAM; **2**) is a major metabolite found in cells and tissues treated with MC (*2*, *3*). The antitumor activity of MC is believed to be based on its ability to alkylate DNA to form several monofunctional adducts and cross-links in DNA (*4*). MC requires reductive activation for the DNA alkylation process (*5*). However, the activated hydroquinone species **3** is partitioned between the pathway of DNA alkylation and a predominant metabolic pathway, which produces a new quinone, 2,7-DAM (Scheme 2a; *2*, *3*, *6*).

Scheme 1. Structures of mitomycin C (MC) and its metabolite 2,7-diaminomitosene (2,7-DAM).



Scheme 2. (a) Reductive metabolism of MC;
(b) Reductive alkylation of DNA by 2,7-DAM.



In MC-treated tumor cells, six major covalent DNA adducts are formed and the same six adducts can be isolated in reconstituted cell-free systems using flavoreductases or chemical reducing agents for activation of the drug. The structures of the DNA adducts are shown in Figure 1. Four of the six adducts (**4**-**7**) are generated directly from the active form of MC by alkylation of the 2-amino group of guanine in the minor groove of DNA to give both monofunctional (**6**, **7**) and bifunctional (**4**, **5**) DNA adducts (*7*). Adducts **8** and **9** are products of 2,7-DAM. This was shown by direct treatment of the cells or DNA in vitro with purified 2,7-DAM itself, which resulted in the selective formation of adducts **8** and **9**, at high efficiency. 2,7-DAM lacks the aziridine ring and alkylates DNA monofunctionally at the guanine-N7 and -N^2 positions (Scheme 2b). The alkylation, like with MC, requires reductive activation; the relative proportions of the two adducts vary, depending on the conditions of their formation (*7*, *8*).

Earlier studies of guanine-N^2 adducts of mitomycin C site-specificially incorporated in plasmids indicated that the major MC-guanine-N2 monoadduct is cytotoxic in bacteria (10). MC itself is highly cytotoxic when administered to bacterial or

mammalian cells, but the metabolite 2,7-DAM is non-cytotoxic (11). It is proposed that the effect of structurally distinct 2,7-DAM adduct on DNA function may be different from those of MC-DNA adducts. Treatment of mouse mammary tumor cells EMT6 with MC resulted in the isolation of a predominant product, the specific guanine N^7-adduct of 2,7-DAM in the major groove.

This result supports the idea that 2,7-DAM is the major bioreductive metabolite of MC. In contrast with MC, 2,7-DAM-guanine-N^7-DNA adduct has a dramatically lower cytotoxicity under both aerobic and anaerobic conditions. The different effects of MC-DNA and 2,7-DAM adducts supports the concept that specific structural features of the DNA damage may play a critical role in the cytotoxic response to a DNA-targeted chemotherapeutic agent (12).

Another major difference between the reactivity of 2,7-DAM and MC is the sequence specificity requirements for guanine alkylation and also the nitrogen of the purine ring which is chemically modified. It has been shown that both MC and DMC (decarbamoyl mitomycin C) have Pur d(CpG)Pyr . Pur d(CpG)Pyr sequence requirement. This is so because the guanine N^2 atom of GpPyr is more reactive toward the drug than that of GpPur, due to the favorable effect of the negative dipole of the O^2 of the Pyr on the reaction (13).

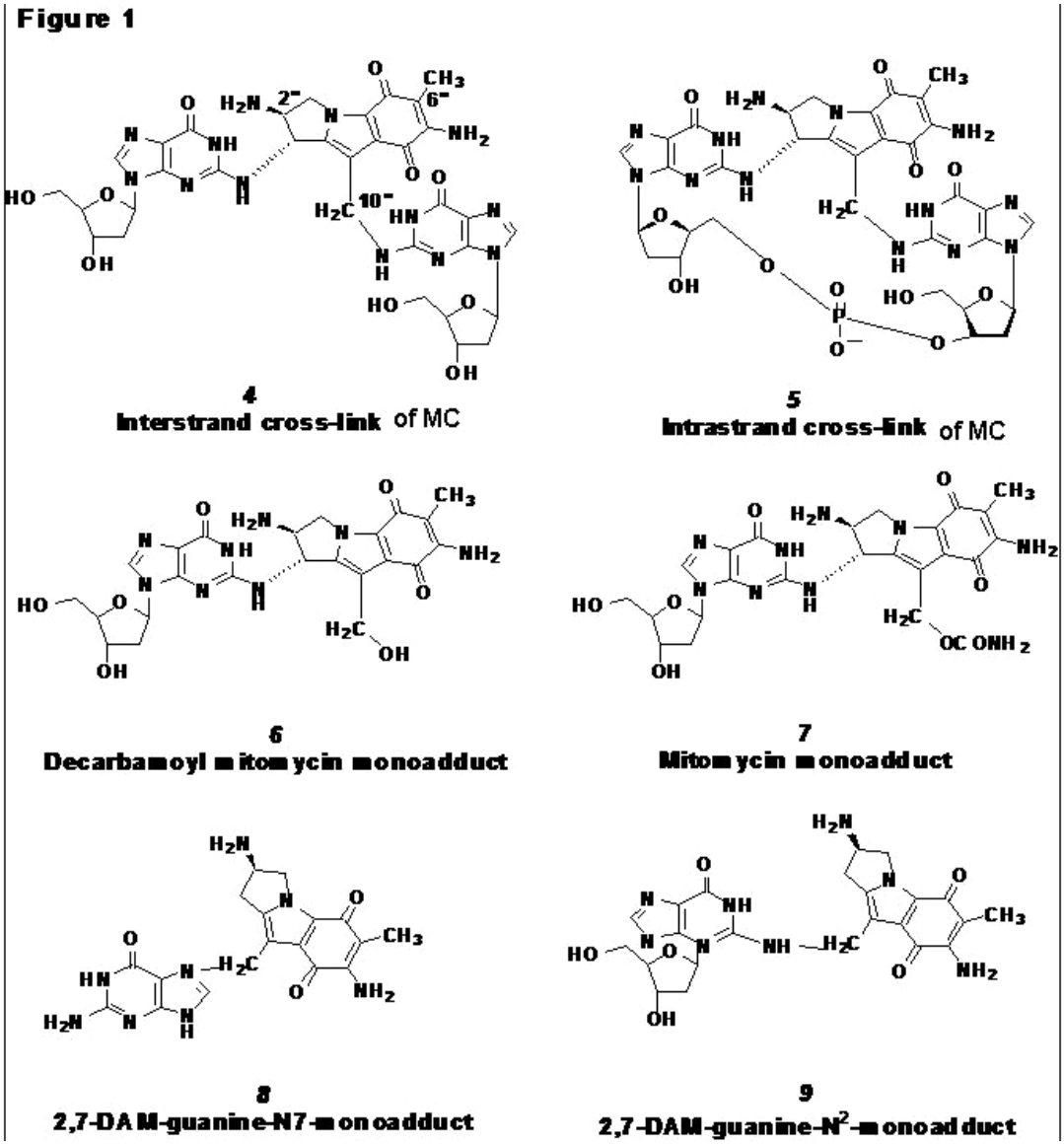


Figure 2: **The major adducts of MC and 2,7-DAM (9).**

The mechanism of selective monoalkylation to N2-guanine was shown previously to be attributed to specific H-bond between the drug and the 2-amino-group of guanine (14). In addition, 5'C is required for H-bond of carbamate to the non-target G-2-NH_2 group in the opposite strand which facilitates the covalent reaction at the target G-NH_2 (15). In contrast, the alkylation of guanines in the major groove of DNA by 2,7-DAM requires the sequence context of longer poly(dG)n . poly(dC)n tracts. The intrinsic S_N2-

type reactivity of 2,7-DAM renders it to alkylate selectively the much more nucleophilic G-N^7. Thus, increasing of (G)n tract length enhances the reactivity of G-N7, since the there is an increase in the negative molecular electrostatic potential minima at N^7 guanines as compared with single guanine residues. The guanine-N^7 atoms in (G)n tracts are the most nucleophilic sites in B-DNA and are subjected to S_N2-type of alkylation (16).

One important biological function related to the cytotoxic effect exerted by MC adducts is the replication of DNA. Monoadducts of mitomycin C can act as strong blocks of the replication fork, thus inhibiting the polymerase reaction. Previously, an in vitro polymerase reaction assay was designed containing site-specifically modified oligodeoxynucleotides at guanine residues with MC as templates and three different polymerases (T7 DNA (Sequenase), AMV reverse transcriptase and E. coli DNA Polymerase I (Klenow fragment) exo^- and exo^+ forms) were used in the polymerase assay (17). The alkylation of oligodeoxynucleotides was performed under conditions which determined the formation of specific N^2-guanine monoadducts. The polymerase reaction was designed such that both the primer extension and single nucleotide incorporation opposite the MC-dG were investigated in order to determine whether the DNA polymerase can bypass the lesion and, in the case of a translesion bypass, whether the nucleotide incorporation was correct (17).

In contrast with other site-specifically located adducts induced by carcinogens or antitumor drugs which exhibit a translesion bypass, in the case of MC monoadduct DNA synthesis was terminated nearly quantitatively at the nucleotide 3' to monoadducts sites, instead of primer extension to full length of the template. There was no translesion bypass when exo^- Klenow fragment was used in presence of Mn^{2+} at high concentration

of deoxynucleotide triphosphates, even though the incorporation of nucleotides opposite the adduct took place (17).

The block of replication induced by the MC monoadducts is proposed to be due to the increased stability of a distorted conformation of the template/primer since the thermal melting of a typical template/primer complex showed higher Tm values for the templates containing MC adduct (ΔTm 0.8-1.0) (17). Additional molecular modeling studies of the modified templates/primer supported the thermal melting results by showing unique hydrogen-bonding between 10”-OC(O)NH2 group of the bound mitosene and bases from the primer/template junction. In addition the 2-“ NH3+ group has three H-Bonds to various electronegative atoms on the template strand (17). The position of 3’-OH was shown not to be altered but the base-pair formation of a nucleotide triphosphate to the MC modified guanine was proposed to be blocked by the large 10”-carbamate group which is stabilized by its H-bonds in an obstructive conformation (17). The replication studies performed in this paper suggested an unexpected role of the mitomycin C monoadducts in the induced-cell cytotoxicity described by earlier studies of Tomasz et al. (11, 12).

The structures of both monoalkylated and cross-linked MC-DNA complexes have been elucidated by NMR spectroscopy, combined with computational molecular modeling methods (18). The monofunctional adducts (figure 1) are located in the minor groove, at the exocyclic amino group of guanine residues, with the selective d(C-G)·d(C-G) sequence specificity. The NMR solution structure of the 2,7-DAM-DNA complex was also determined and showed that the 2,7-DAM molecule is anchored in the major groove of DNA. The presence of 2,7-DAM in the major groove does not alter the overall B-DNA helical structure (19). Alignment in the major groove was proposed to be a novel

feature of the complexation of 2,7-DAM with DNA; other known major groove alkylators such as aflatoxin, possessing aromatic structural elements, form intercalated complexes (20-30). In figure 2 the modeled structures of the two monoadducts of mitomycin C and 2,7-DAM are presented together with the pdb ID code which was used to obtain the atomic coordinates from PDB.

We hypothesized that different structures of the MC and 2,7-DAM monoadducts with DNA may induce different distortions in the templates/primer and thus may affect differently their recognition by the DNA polymerases. These differences in the interaction between the polymerases and the modified DNA templates may play an important role in inducing different cytotoxic effects of MC and 2,7-DAM monoadducts.

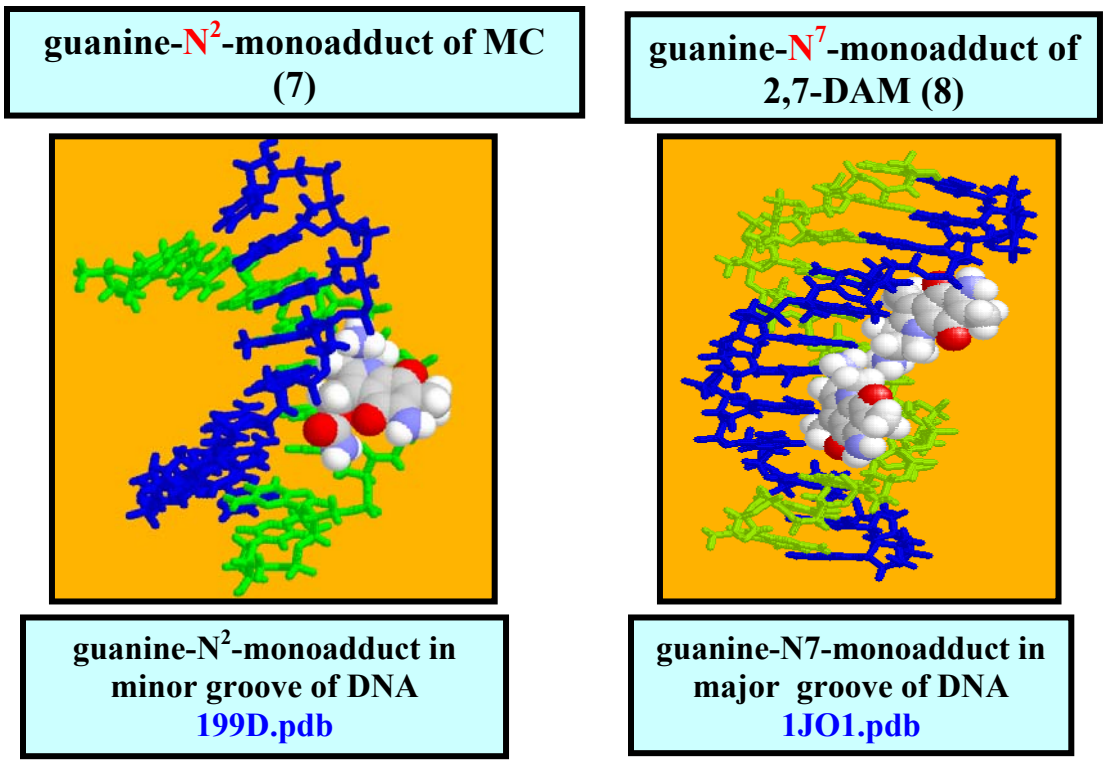


Figure 3: **NMR-solved structures of the monoadducts 7 and 8 (18, 19).** The atomic coordinates for each structure were downloaded from PDB data bank and modeled using the software Rasmol.

PART I: ORIGINAL CONTRIBUTION

The first part of the thesis focused on the relationship between the structure of N7-dG-monoadducts of 2,7-DAM and the N2-dG- monoadducts of MC and their activity related to the inhibition DNA replication in isolated in vitro simple systems composed of primer/template (substrates) and 3 major DNA polymerases, T7 exo-, Klenow exo- and eta (η).

The research presented in Part I of the thesis had 3 major specific aims:

(I) Synthesis of DNA templates with chemically modified guanines at ^{2}N-and N7 position with MC and 2,7-DAM, respectively.

(II) *In vitro* translesion synthesis (TLS) and single nucleotide incorporation kinetics opposite the MC and 2,7-DAM monoadducts performed with Klenowexo-, T7 exo-, eta and Klenow exo+ DNA polymerases.

(III) Physical-chemical investigations *(thermal melting)* of d(G-T-G-G(4)-T-A-T-A-C-C-A-C) (duplex III) self-complementary DNA duplex alkylated at one single guanine (G4) with 2,7-DAM.

The original findings related to the in vitro TLS of these 2 major bulky adducts of MC and its derivative 2,7-DAM contributed both to the knowledge of SAR in the field of MC as an anticancer drug and to the more complex field of the enzymology of different DNA polymerases and their efficiency of performing the DNA replication past structurally-different damaged DNA lesions. A relative new isolated adduct of N2-dG-2,7-DAM (23) was also used to generate templates-substrates for different DNA polymerases and the preliminary experiments related to the translesion synthesis of this adduct are presented in Chapter 2. Of course, we are just at the beginning of understanding the molecular

mechanisms underlying the recognition and the efficient processing of damaged DNA (both at the replication and at the repair levels) and the MC-monoadducts can be used as useful tools in probing more complex questions related to the structural basis of DNA polymerases for recognition of chemically modified templates and their catalytic efficiency of performing the DNA replication with or without error during TLS. Thus, the relationship between the MC-monoadducts structure and their mutagenic potential was further explored in this original thesis project and presented as a starting point for further detailed investigations as described in Chapter 2.

REFERENCES

1. Hata, T., Sano, Y., Sugawara, R., Matsumae, A., Kanamorei, K., Shima, T., and Hoshi, T. (1956) *J. Antibiot., Ser. A 9*, 141-146.

2. Verweij, J., and Pinedo, H. (1990) in *Cancer chemotherapy and biological modifiers, Annual 11* (Pinedo, H. M., Chabner, B. A., and Longo, D. L., Eds.) pp 67-73, Elsevier Science Publishers B.V., Amsterdam.

3. Chirrey, L., Cummings, J., Halbert, G. W., and Smyth, J. F. (1995) *Cancer Chemother. Pharmacol. 35*, 318-322.

4. Lipman, R., and Tomasz, M. (1981) *Biochemistry 20*, 5056-5061.

5. Iyer, V. N., and Szybalski, W. (1963) Iyer, V. N.; Szybalski, W. A. Molecular Mechanism of Mitomycin Action: Linking of Complementary DNA Strands. *Proc. Natl. Acad. Sci. U.S.A.*, *50*, 355-362. *Proc. Natl. Acad. Sci. U.S.A. 60*, 355-362.

6. Tomasz, M. (1994) in *Molecular Aspects of Anticancer Drug-DNA Interactions* (Neidle, S., and Waring, M., Eds.) Vol. 2, pp 312-349, Macmillan, London, and CRC Press, Fort Lauderdale, FL.

7. Li, V., and Kohn, H. (1991) *J. Am. Chem. Soc. 113,* 275-283.

8. Kumar, S., Lipman, R., and Tomasz, M. (1992) *Biochemistry 31*, 1399-1407.

9. Clement, Cristina C.; Utzat, Christopher D.; Ramos, Leilani A.; Das, Arunangshu; Tomasz, Maria; Basu, Ashis K. *Chemical Research in Toxicology* (2005), **18(2)**, 213-223.

10. Ramos, L. A., Lipman, R., Tomasz, M., and Basu, A. K. (1998). *Chem. Res. Toxicol*. *11*, 64-69.

11. Palom, Y., Belcourt, M. F., Tang, L.-Q, Mehta, S. S., Sartorelli, A. C., Pritsos, C. A., Pritsos, K. L., Rockwell, S., and Tomasz, M. (2001). *Biochem. Pharmacol*. *61*, 1517-1529.

12. Palom Y., Belcourt M. F., Kumar G. S., Arai H., Kasai M., Sartorelli A. C., Rockwell S., M. Tomasz. 1998. *Oncology Research*, 10, 509-521.

13. Borowy-Borowski H., Lipman R., M. Tomasz. (1990). *Biochemistry*, 29, 2999-3006.

14. Tomasz M., Das A., Tang K.S., Ford M.G.J., Minnock A., Musser S.M., Waring M.J. (1998). *JACS*, 120, 11581-11593.

15. Kumar G. S., Musser S. M., Cummings J., M. Tomasz. (1996). *JACS*, 118, 9209-9217.

16. Kumar G. S., R. Lipman, Cummings J., M. Tomasz. (1997). *Biochemistry*, 36, 14128-14136.

17. Basu, A. K., Hanrahan, C. J., Malia, S. A., Kumar, S., Bizanek, R., and Tomasz, M. (1993). *Biochemistry 32*, 4708-4718.

18. Sastry M, Fiala R., Lipman R., Tomasz M., Patel D. J. (1995). *J. Mol. Biol*., 247, 338-359.

19. Subramaniam, G., Paz, M. M., Suresh Kumar, G., Das, A., Palom, Y., Clement, C. C., Patel, D. J., and Tomasz, M. (2001). *Biochemistry 40*, 10473-10484.

20. Busby, W. F., Jr., and Wogan, G. N. (1984). In *Chemical Carcinogens* (Searle, C., Ed.) 2nd ed., pp 945-1136, American Chemical Society, Washington, DC.

21. Suzuki, Naomi; Yasui, Manabu; Laxmi, Y. R. Santosh; Ohmori, Haruo; Hanaoka, Fumio; Shibutani, Shinya. (2004), *Biochemistry,* 43(35), 11312-11320.

22. Verweij, J.; den Hartigh, J.; Pinedo, H. M. Anticancer Antibiotics. (1990). In *Cancer Chemotherapy Principles and Practice*; Chabner, B. A., Collins, J. M., Eds.; J. B. Lippincott Company; Philadelphia, PA, pp 382-396.

23. Palom, Y.; Belcourt, M. F.; Musser, S. M.; Sartorelli, A. C.; Rockwell, S.; Tomasz, M. (2000). Structure of Adduct X, the Last Unknown of the Six Major DNA Adducts of Mitomycin C Formed in EMT6 Mouse Mammary Tumor Cells. *Chem. Res. Toxicol.*, *13*, 479-488.

CHAPTER 2

Different translesion bypass of guanine–N^2 monoadducts of mitomycin C and guanine-N7 monoadducts of 2,7-diaminomitosene by eta, Klenow exo-, Klenow exo+ and T7 exo- DNA polymerases

2.1. INTRODUCTION

In the field of eukaryotic translesion synthesis (TLS), DNA polymerases were the enzymatic systems which were investigated at all molecular levels, from structural to more biochemical and cell biological levels. The fidelity, mismatch extension ability, and lesion bypass efficiencies of different polymerases were studied in the context of their structures (1-55). One major conclusion is that, despite the overall similarity of the basic structural features, there is a high degree of specificity of their lesion bypass among the Y-family of DNA polymerases (which are involved primarily in the DNA lesion bypass with high error rates in DNA synthesis) and the replicative DNA polymerases (which synthesize DNA with high degree of accuracy) (11-25). Some DNA polymerases (either from Y-family or from the replicative DNA polymerase family) are able to bypass a particular DNA lesion whereas others are efficient at only the insertion step or the extension step of lesion bypass. Both the structures of the enzymes and the structures of the adducts (i.e. the type of DNA damage) are final determinants of the efficiency and accuracy of lesion bypass (25-66).

In this original research project we took advantage of the structural difference between the well characterized monoadducts of MC and 2,7-DAM (Figure 1, Chapter-1) and synthesized templates containing site-specific alkylated guanines with bulky adducts

of MC or 2,7-DAM and further performed in vitro translesion experiments and single nucleotide insertion kinetics addressing the efficiency of these lesions bypass by four DNA polymerases, Klenow exo-, T7-exo-, eta and Klenow exo+.

Bioorganic and molecular biology methods were used to synthesize 24-mer, 27-mer and 36-mer DNA templates adducted at a single guanine either with MC or 2,7-DAM. The alkylated templates were further annealed with complementary primers which were further used as substrates to conduct the *in vitro* translesion synthesis (TLS) with T7 exo-, Klenow fragment exo-, eta (η) and Klenow (exo+) DNA polymerases.

In the 27-mer template/15-mer primer and 24 mer template/15 mer primer systems the G-N7-2,7-DAM adduct was bypassed by all four polymerases, resulting in the production of a fully extended primer. However, the extension was at a slower rate as compared with the control, non-alkylated template. In sharp contrast, the G-N^2-MC monoadduct was not bypassed beyond the adduct position under the same condition by any of the four polymerases. Single-nucleotide incorporation kinetics using Klenow exo- polymerase showed that dCTP has the highest frequency of incorporation followed by a very low frequency (<1%) of dGTP incorporation opposite the G-N7-2,7-DAM lesion, while no detectable levels of dATP, dTTP incorporation opposite the same lesion were observed. These results suggest a positive correlation between cytotoxicity and inhibition of DNA synthesis in the case of the two mitomycin monoadducts and support the *in vivo* findings that the G-N7-2,7-DAM lesion is not mutagenic (9).

In contrast with polymerase Klenow exo-, polymerase eta was more prone to errors since dGTP was having the highest frequency of incorporation opposite the G-N7-2,7-DAM lesion, a result which is in agreement with other published data related to error-prone replication by polymerase eta (20-30). An interesting outcome was the formation

of a double adduct -5'-(GG)-$(2,7\text{-DAM})_2$- which showed in a similar single nucleotide kinetics assay a higher frequency for the incorporation of dTTP and dGTP opposite the G-N7-2,7-DAM lesion, suggesting that the double adduct is mutagenic in contrast to the single monoadduct. Preliminary results performed with 24 mer templates adducted at one single guanine residue with 2,7-DAM but at the N^2 position, showed a lower efficiency of primer extension directed from templates adducted with G-N^2-2,7-DAM lesion by T7 exo- and Klenow exo- DNA polymerases as compared with the primer extension directed from templates adducted with of G-N7-2,7-DAM lesion. We are still investigating the kinetics of single nucleotide incorporation opposite other lesions of 2,7-DAM, namely -5'-(GG)-$(2,7\text{-DAM})_2$- and G-N^2-2,7-DAM together with the G-N^2-MC lesion, using polymerase eta.

2.2. MATERIALS AND METHODS

2.2.1. MATERIALS

MC was obtained from Dr. Dinesh M. Vyas, Bristol Myers Squibb Co. (Wallingford, CT). 2,7-DAM was obtained by reduction of MC with H_2/PtO_2 as previously described (*3*). [γ-^{32}P]ATP was from DuPont New England Nuclear (Boston, MA) or from Perkin-Elmer Life Sciences (3000 Ci/mmol, 10 mCi/mL). T4 DNA ligase (400 units/μL) and T4 polynucleotide kinase (PNK) (10 000 units/mL) were obtained from New England Biolabs (Beverly, MA). Phosphodiesterase I (snake venom diesterase, *Crotalus adamanteus* venom, EC 3.1.4.1) and alkaline phosphatase (*Escherichia coli*, EC 3.1.3.1) were obtained from Worthington Biochemical Corp. (Freehold, NJ). *E. coli* Klenow (exo-) DNA polymerase (5000 units/mL) was purchased from USB (Cleveland, OH). Exonuclease-free T7 (exo-) DNA polymerase (13 units/μL) was either purchased from

USB or was a gift from Dr. Smita Patel (UMDNJ, Piscataway, NJ) and was active site titrated using the quench-flow instrument in the same lab (>75% active enzyme). DNA polymerase ɳ, the full length clone of the human enzyme containing a C-terminal 6X histidine tag, was a generous gift from Dr. Scott McCulloch and Dr. Thomas A. Kunkel (Laboratories of Molecular Genetics and Structural Biology, NIEHS, NIH, Research Triangle Park, NC). Synthetic oligonucleotides were purchased from the Midland Certified Reagent Company, Inc. (Midland, TX) or were synthesized in the Core Facilities at Hunter College and further purified and desalted through a Sephadex G-25 column. Their purity was confirmed through reversed phase HPLC and negative electrospray ionization mass spectroscopy ESI-MS (-).

2.2.2. General Methods

UV spectra and absorbance readings were recorded in a Cary 3 UV-visible spectrophotometer. All DNA concentrations were determined by using the calculated molar extinction coefficient (E_{260} (strand)) for each oligonucleotide strand and the absorbance readings at 260 nm using the formula: $c = A_{260}/ [E_{260}$ (strand) x L] (where the molar extinction coefficient E is expressed in $M^{-1}cm^{-1}$ and the path length of the cuvette is L = 1.0 cm). Each E_{260} (strand) was calculated using formula: E_{260} (strand) = number of purines x 14,000 + number of pyrimidines x 7,000, where 14,000 and 7,000 are the averaged molar extinction coefficients for purines and for pyrimidines, respectively (14-15-Chapter-1). In the case of the 2,7-DAM alkylated oligonucleotide a correction for E_{260} (strand) was done by adding 5189 (E_{260} of 2,7-DAM) (14-15-Chapter 1). LC-MS was performed with a Hewlett-Packard series 1100 diode array HPLC system connected to a Hewlett-Packard series 1100 MSD mass spectrometer. HPLC was performed in a Beckman System Gold 125 instrument, equipped with a Beckman System Gold 168

diode array detector. The HPLC columns used for analytical purposes were C18, 100 Å (Microsorb MV Rainin, 5 mm x 250 mm) and C4, 300 Å (Microsorb MV Rainin, 5 mm x 250 mm) while the columns for semipreparative and preparative use were C18, 100 Å, Dynamax (5 µm, 1 cm x 25 cm) and C4, 300 Å, Dynamax, (5 µm, 2.14 cm x 25 cm) respectively. 18% polyacrylamide gel electrophoresis (PAGE) with 8 M urea was run in a Sequi-Gen GT (Biorad Inc, CA) apparatus, at 2800 constant voltage for 3.5 h. The gel was exposed to phosporimaging on a Phosphorimager 445 SI (Molecular Dynamics, now Amersham Biosciences) and the quantitative data were produced using the software ImageQuant 5.2.

2.2.3. Specific Methods

2.2.3.1. Synthesis of the site-specifically substituted short oligonucleotide 10 (Scheme 3):

Scheme 3. Alkylated Oligonucleotides (10-13) and Oligonucleotide Templates Constructed by Ligation (14-16).

5'-ACACG(MC)TCAT-3
10 (9-mer)

5'-GTGG(2,7-DAM)TACCAC-3'
11 (10-mer)

5'-CTGG(2,7-DAM)TAATTTAC-3'
12 (12-mer)

5'-CTAGTGG(2,7-DAM)TATCC-3'
13 (12-mer)

5'-ACACG(MC)TCATTAGAGATTGGTAGGG-3'
14 (24-mer)

5'-CTGG(2,7-DAM)TAATTTACTAGAGATTGGTAGGG-3'
15 (27-mer)

5'-CTAGTGG(2,7-DAM)TATCCTAGAGATTGGTA-3'
16 (24-mer)

Synthetic oligonucleotides were incubated with MC or 2,7-DAM under reductive conditions using $Na_2S_2O_4$ or catalytic hydrogenation in aqueous buffer, as described

earlier. Specifically, 9-mer **10** and the 10-mer **11** were synthesized and characterized as reported (14-15-Chapter 1).

2.2.3.1.1. Preparative alkylation of **120 OD_S** total DNA [C1 oligo (**5'-A C A C G T C A T-3'= 9 mer)** and **C2 oligo (3'-T I T G C A I T-5' = 8 mer).**

This reaction was performed using the reductive alkylation procedure of mitomycin C in presence of substoichiometric amounts of sodium dithionate ($Na_2S_2O_4$) which enable the formation of monoaaducts as already was described elsewere (14-15-Chapter 1). The reaction mixture contained 2.4 mM final concentration mononucleotide duplex DNA (12 μmole mononucleotide DNA duplex in 5 ml reaction mix) and 4 fold excess of mitomycin C (9.6 mM MC) in the 5 ml reaction mix. (i.e. a total of 48 μmole total MC). The $Na_2S_2O_4$ was freshly prepared at 1/4 molar concentration of that of mitomycin C (i.e. 2.4 mM final concentration) in anaerobic conditions. Before sodium dithionate was added to the reaction mixture, MC (stock solution in methanol), C1 and C2 oligos (both dissolved into distilled water) were mixed together, rotavaporized to dryness and the residuals were resuspended in 5 ml 0.1 M KH_2PO_4, pH=7.0 (same buffer used for sodium dithionate preparation). The mixture was heated at 95^{o}C for 3 minutes, and the annealing of the two oligos was performed at room temperature for 1 hour, followed by cooling on ice (4^{o}C) for 30 minutes. The DNA duplex formed after annealing is suitable for alkylation with mitomycin C (14-15-Chapter 1):

5'-A C A C G *T C A T-3' 9mer (C1-oligo)

3'-T I T G* C A I T-5' 8 mer (C2-oligo)

The guanine that are alkylated at NH_2 position by MC are indicated with a (*).

The freshly prepared sodium dithionate (under Argon for 20 minutes) was added to the duplex DNA + MC mixture, and reductive alkylation of the DNA duplex was performed

on ice, in the dark, aerobically during 1 hour. The reaction mixture was further applied on a Sephadex G-25 column (56 x 2.5 cm), and the elution of the alkylated DNA together with the nonalkylated DNA was performed using 20 mM ammonium bicarbonate buffer. The first peak eluted from the gel filtration column contains both alkylated and nonalkylated DNA, while the free mitomycin C un-reacted is eluted after the DNA peak. The first peak was further lyophilized and the presence of the alkylated DNA strands each was determined by running the alkylated mixtures of C1 and C2 oligos on a HPLC C4 column, using the reverse-phase buffer described in Materials and Methods.

2.2.3.1.2. Preparative HPLC reverse phase chromatography was employed for purification of C1= 5'-A C A C G T C A T-3'= 9 mer and C2 oligo =3'-T I T G C A I T-5' = 8 mer.

In order to perform the preparative HPLC, a scale-up factor was used (scale-up factor = π r^2 prep / π r^2 anal), in which the ratio between the radius of the preparative column over the radius of the analytical column is used to predict the flow rate that has to be used for the preparative experiment such that the components of interest will eluted with the same retention time as in the analytical run. In the preparative experiment for the purification of C1 and C2 oligos alkylated with mitomycin C, we were using a preparative C4, Dynamax, 300 Å-wide pore, 5 μm, 25 cm. The scale-up factor was 21.6 in this case, predicting that the flow rate would be 20 ml/min. The actual flow rate was 5 ml/min, which is 4 fold less than the flow rate that was actually predicted from the scale-up factor calculations. Thus, the time required the chromatogram was 4 fold higher, i.e. 440 minutes (110 min as for analytical runs x 4). The buffer system and the gradient used in the reverse phase run were the same as those mentioned for the analytical run [Buffer A: 0.1 M triethylammoniumacetate (TEAA), pH=7.0. Buffer B: 30% CH_3CN in

70 % 0.1M TEEA, pH=7.0 with the gradient, 20-60% B in 440 minutes). The individual peaks corresponding to the alkylated oligonucleotides C1 and C2 (as determined by the ratio of Absorption at 260 nm/Absorption at 320 nm which is between 10-14 for alkylated oligos (14)) were collected and lyophilized. The concentration of the corresponding oligos was determined from UV absorption measurements. For each run a 300 μl of mixture of oligos containing around 50 ODs (as absorption units at 260 nm) was injected into the preparative column. The Beckman HPLC system was used for the preparative run as for analytical run.

2.2.3.1.3. HPLC reversed-phase analytical control for alkylation with MC of C1 and C2 oligos.

The HPLC was performed using the silica based C4 analytical column (5 μm, 300 Å) (Rainin) and the following buffer system: Buffer A: 0.1 M triethylammoniumacetate (TEAA), pH=7.0. Buffer B: 30% CH_3CN in 70 % 0.1M TEEA, pH=7.0. The gradient applied was 20%-60% B in 110 minutes. The peaks were monitored by fixing the channel A at 254 nm and channel B at 320 nm. The flow rate for the analytical run was 1ml/min. The % yield of monoalkylation of the corresponding DNA strands was calculated from the integrated areas of the HPLC peaks corresponding to the alkylated oligos relative to the unmodified strands. The % yield of alkylation with mitomycin C was in all cases ~20%.

2.2.3.2. Synthesis of the site-specifically substituted short oligonucleotides 12-13 (Sheme 3).

Preparative alkylation of 168 OD_S total DNA [[5'-CTGGTAATTTAC-3'] and [5'-AATTACCAG-3'] annealed duplex] and of 168 ODs total DNA

[[5'CTAGTGGTATCC-3'] and [5'-GGATACCACT-3'] annealed duplex]].

The reaction of alkylation was performed in 0.1 M potassium phosphate buffer pH 5.5 in the presence of PtO_2 as catalyst, under anaerobic conditions. A total of 168 OD_{260} units of oligonucleotide mixture containing either the unsubstituted oligonucleotide precursor of **12** annealed with 3'-GACCATTAA-5' or the unsubstituted oligonucleotide precursor of **13** annealed with 5'-GGATACCACT-3' was mixed with 4.2 μmole 2,7-DAM and 0.841 mg PtO_2 in 5 mL 0.1 M potassium phosphate buffer pH 5.5, and the mixture was annealed by heating at 95° C for 3 min, then slowly cooling to 25° C, followed by keeping it on ice for 30 min. After annealing, solid ferrous ammonium sulfate (0.084 mg) was added to the mixture then purged with argon for 15 min. The hydrogenation was performed by purging with H_2 for 17 min at room temperature. After purging with argon for 1 h at room temperature the reaction mixture was exposed to air and was applied on Sephadex G-25 fine column. The first UV-absorbing peak containing the unmodified and modified oligonucleotides was collected and further purification of alkylated oligos was performed using a semipreparative C18, column, 100 Å, and the following elution system: eluant A: 30 mM phosphate, pH 5.4, eluant B: 30% CH_3CN in 70% 30 mM phosphate, pH 5.4. The gradient used was 10%-25% B in 10 min, 25-33% B in 60 min and 33-100% B in 20 min, with 3 mL/min flow rate. The absorbance was recorded both at 260 nm and 314 nm and the adducted oligos were identified from their absorption at both wavelengths and collected as described (14-15-Chapter 1).

2.2.3.3. Characterization of adducted oligonucleotides 10-13 (Scheme-3).

Analytical HPLC was used establish homogeneity, using reverse phase columns (Rainin C4 column, 300 Å (Rainin Microsorb MV, 5 x 250 mm)). ESIMS (-) confirmed

the molecular weight of the adducted oligonucleotides. All ESI-MS spectra were acquired using the solvent system: 400 mM hexafluoro-2-propanol (HFIP) (5 ml water, 4.5 ml methanol, 420 μL HFIP, 12 μL triethylamine). The solvent was injected at 50 μL/min flow rate in the ion source. Digestion by snake venom phosphodiesterase and alkaline phosphatase, followed by reverse phase HPLC of the digest confirmed the structure of the nucleoside adducts **7** or **8,** and the nucleoside and drug-nucleoside adduct composition of **10-13** as described before (11-15-Chapter 1). Briefly, 0.3 ODs from the purified C1-MC alkylated oligo were subjected to digestion with snake venom phosphodiesterase (1 unit) and alkaline phosphatase (0.3 units) in 0.1 M Tris-HCl, pH=8.2 containing 2mM $MgCl_2$ final concentration. The reaction was left to proceed for 4 hours, at 45^{o}C. The resulted mixture of nucleosides were tested on an analytical column C18 (Rainin) using the phosphate buffer system (Buffer A: 30 mM phosphate buffer, pH=5.4. Buffer B: 30% CH_3CN in 70% 30 mM phosphate buffer, pH=5.4). The gradient applied was 20-60% B in 60 minutes.

2.2.3.4. Mapping the position of the 2,7-DAM adduct in 11, 12 and 13 (Scheme-3): There are 2 guanines in oligonucleotides 11 and 12 and 3 guanines in oligonucleotide 13 (Scheme 3). The positions of the adducts were determined as follows**.** **11:** The oligonucleotide was 5'-labeled with [^{32}P], heated in hot piperidine and analyzed by PAGE along a Maxam-Gilbert G-lane of labeled unmodified control oligonucleotide. The results showed that **11** was substituted exclusively at G(4) as described in detail earlier (14-15-Chapter 1). Positions in **12** and **13** were determined analogously.

2.2.3.5. Construction of alkylated templates 14-16 (Scheme 3). The short adducted oligonucleotides **10-13** (6 nmol each) were ligated to 5'-phosphorylated

oligonucleotides (6 nmol each) by T4 DNA ligase (400 units/6 nmol duplex DNA) in the presence of a complementary oligomer (6 nmol) which held the ligating oligomers together, using a modified procedure described originally for other mitomycin-adducted constructs (11-15-Chapter-1).

Specific examples of ligation experiments:

2.2.3.5.1. Ligation of the MC modified 9 mer with a 5'-phosphorylated 15 mer in the presence of a complementary 20 mer in order to obtain the 24 mer template. The 9 mer 5'-ACACG*TCAT-3' (3 nmol) with G* alkylated with MC or the unmodified 9 mer (control) was annealed with the 15 mer 5'-P-TAGAGATTGGTAGGG-3' (3 nmol) and the complementary 20 mer 3'-TGTGCAGTAATCTCTAACCA-5' (3nmol) by heating first the three oligonucleotides at 80^{o}C and then cooling slowly to 30^{o}C followed by a incubation at 4^{o}C for 30 minutes. The reaction volume was 10 µl of 50 mM Tris-HCl, pH=7.6, 10 mM $MgCl_2$, and 10 mM dithiothreitol (DTT). The ligation was performed using the T4 ligase (2 units) (ATP dependent, 1mM final concentration), BSA (50 µg/ml) during 16 h at 20 ^{o}C. The reaction was stopped by inactivation of ligase at 65^{o}C during 10 minutes. A small aliquot from the mixture was 5'-end ^{32}P labeled using γ-^{32}P-ATP, and the results of the ligase reaction were analyzed on a 16% polyacrylamide-8M urea gel.

2.2.3.5.2. Ligation of the 2,7-DAM modified 12 mer with a 5'-phosphorylated 15 mer in the presence of a complementary 17 mer in order to obtain the 27 mer template.

The 12 mer **5'- CTGG*TAATTTAC-3'** (3 nmol) with G* alkylated with 2,7-DAM or the unmodified 12 mer (control) was annealed with the 15 mer 5'-p -TAGAGATTGGTAGGG-3' (3 nmol) and the complementary 17 mer 3'-

CATTAAATGATCTCTAA-5' (3nmol) by heating first the three oligonucleotides at 60^{o}C (five minutes) and then cooling slowly to room temperature followed by a incubation at 4^{o}C for 30 minutes. The reaction volume was around 10 µl in 50 mM Tris-HCl, pH=7.6, 10 mM $MgCl_2$, and 10 mM dithiothreitol (DTT) (ligation buffer 1x). The ligation was performed using the T4 DNA ligase (2 units) (ATP dependent, 1mM final concentration), BSA (50 µg/ml) during 16 h at 4^{o}C. The reaction was stopped by inhibition of ligase during 10 minutes heating at 65^{o}C. A small aliquot from the ligation mix was used for ^{32}P labeling and checking of ligation products using 18% polyacrylamide-8M urea sequencing gel. Other specific oligonucleotide components used for the construction of each template were the following (Scheme 3):

24 mer-template (**14**): **10,** 5'-pTAGAGATTGGTAGGG,

5'-ACCAATCTCTAATGACGTGT.

27 mer-template (**15**): **12**, 5'-pTAGAGATTGGTAGGG, 5'-AATCTCAGTAAATTAC.

24 mer-template (**16**): **13**, 5'-pTAGAGATTGGTA, 5'-CCAATCTCTAGGATACCACT.

The templates were detected by UV shadowing, excised from the gel and the gel pieces were soaked overnight at room temperature in 100 mM Tris, 0.5 M NaCl, 5 mM EDTA, pH 8.0. The templates were desalted by passing through a C18-Sep-Pak cartridge (Waters). The elution buffer consisted of 0.1 M triethyl ammonium acetate (TEAA), pH 7.0 : 100% CH_3OH (1:1). The alkylated templates were analyzed for the presence of the MC or 2,7-DAM adduct by digestion with snake venom phosphodiesterase and alkaline phosphatase followed by analysis of the digests by HPLC, as described for characterization of the short oligonucleotides **10-13** above.

The purity of the alkylated templates was confirmed by reverse-phase HPLC, using C4 column (300 Å) and the solvent system based on 0.1 M TEAA buffer, pH 7.0 and

acetonitrile (31). In addition, samples of the alkylated templates were 5'-labeled with ^{32}P and analyzed for purity on 18% PAGE-8 M urea high resolution sequencing gel.

2.2.3.6. 5'-Labeling (phosphorylation) of oligonucleotides with 32P. The primer oligonucleotides were labeled using 18 pmole DNA and 20 μCi [γ-^{32}P] ATP (3,000 Ci/mmol) together with 20 units of T4 polynucleotide kinase (PNK) (10,000 U/ml) in a reaction mixture containing the buffer for PNK provided by the company and BSA (100 μg/mL). The reaction of 5'-phosphorylation was performed at 37^{o}C for 1 h and 50 min and the kinase was inactivated by incubation at 65^{o}C for 15 minutes.

2.2.3.7. 5'-Labeling (phosphorylation) of oligonucleotides with non-radioactive P for ligation: The oligonucleotides required to be 5'-phosphorylated for further ligation experiments were phosphorylated using 6 nanomol DNA and 20 to 30 units of T4 PNK in a buffer used in ligation experiments at 1 mM final ATP concentration under the same conditions as above.

2. 2.3.8. Primer extension assay. All primer extension reactions were performed under steady state conditions. The substrates for the in vitro polymerase assay consisted of unmodified (control) and site-specifically alkylated templates annealed with shorter complementary ^{32}P-labeled primers to form complexes 17, 18 and 19 (Scheme 4). The reaction mixture contained 150-200 nM final duplex DNA concentration, primer:template (1:1.5 molar ratio)) in buffer 1x for different DNA polymerases: 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 5 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT for T7 (exo-) DNA polymerase; 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 7.5 mM DTT for Klenow (exo-) polymerase; 40 mM Tris-HCl, pH 8.0, 10mM $MgCl_2$, 60 mM KCl, 10 mM DTT for eta polymerase. Each buffer was containing in addition 100 μg/ml BSA final concentration.

The enzymes were added to the annealed primer/template complex such that their final concentration in the reaction mix was 15-20 nM (10 fold excess DNA substrate vs. enzyme) except for the eta polymerase where the enzyme was 40 nM and the duplex DNA substrate was 200 nM such that the substrate : enzyme ratio was 5:1. The polymerase reaction was initiated by adding to the preformed [duplex DNA/ polymerase] complex an equimolar mixture of the four dNTPs in a range of 100 – 500 μM concentration of each (20-25 μl final volume of reaction). Equal aliquots from the reaction mix (5-7 μL at each time point) were withdrawn at 1, 3, 5, 10 and 30 or 60 min and quenched with an equal volume (5-7 μL) mixture of 20 mM EDTA, pH 8.0, and sequencing loading gel buffer containing 95% formamide and 0.05% bromphenol blue. Fractionation of all products from the primer extension reaction was performed by PAGE and quantified as described above.

2.2.3.9. Single nucleotide incorporation assay. Four independent reactions, corresponding to the incorporation of dATP, dGTP, dCTP and dTTP opposite the lesions were run at 250-500 μM dNTP concentration using the same procedure and enzymes described for the primer extension assay using the DNA substrate constructs **20**, **21** and **22 (Scheme 4)**.

Scheme 4. Primer/Template Complexes Used in Primer Extension on Adducted Templates.

3'-ATCTCTAACCATCCC-5'-^{32}P (15-mer)
5'-ACACG(MC.)TCATTAGAGATTGGTAGGG-3' (24-mer)
17

3'-AATGATCTCTAACCAT-5'-^{32}P (16-mer)
5'-CTGG(2,7-DAM)TAATTTACTAGAGATTGGTAGGG-3' (27-mer)
18

3'-AGGATCTCTAACCAT-5'-^{32}P (15-mer)
5'-CTAGTGG(2,7-DAM)TATCCTAGAGATTGGTA-3' (24-mer)
19

3'-AGTAATCTCTAACCATCCC-5'-32P (19-mer)
5'-ACACG(MC)TCATTAGAGATTGGTAGGG-3' (24-mer)
20

3'-ATTAAATGATCTCTAACCATCCC-5'-^{32}P (23-mer)
5'-CTGG(2,7-DAM)TAATTTACTAGAGATTGGTAGGG-3' (27-mer)
21

3'-ATAGGATCTCTAACCAT-5'-^{32}P (17-mer)
5'-CTAGTGG(2,7-DAM)TATCCTAGAGATTGGTA-3' (24-mer)
22

3'-AATGATCTCTAACCAT-5'-^{32}P (16-mer)
5'-CTGG(2,7-DAM)$_2$TAATTTACTAGAGATTGGTAGGG-3' (27-mer)
23

3'-ATTAAATGATCTCTAACCATCCC-5'-^{32}P (23-mer)
5'-CTGG(2,7-DAM)$_2$TAATTTACTAGAGATTGGTAGGG-3' (27-mer)
24

2.3.10. Quantitative analysis of primer extension kinetics. Quantitative analysis of the kinetic data were performed using the Phosphorimager software Image Quant. The total counts per each line (corresponding to each time point of the kinetic run) was performed. The DNA polymerase reaction product was calculated by dividing the counts for each individual DNA product (like the counts corresponding to 36 mer, 25 mer, 24 mer, 22 mer, etc.) with the total counts for the specific time point. Thus, the data were expressed as Fraction Product = f (time) or as the %primer remained un-extended as a function of time. The data were thus normalized for the possible errors that

could appear due to differences in the concentration of DNA applied into each wells. %primer remained un-extended was calculated as the fraction between the primer remained unextended at each time point and the primer counted at time '0' multipled with 100.

2.3. RESULTS

2.3.1. Synthesis of the site-specifically substituted short oligonucleotides 10-13 (Scheme 3).

2.3.1.1.Synthesis of MC-alkylated oligonucleotide 5'-ACACG(MC)TCAT-3'-C1-oligo.

The short adducted oligonucleotides **10-13** were designed for ligase-mediated extension to 24- or 27-mer template strands **14-16 (schemes 3 and 4)**. The alkylation reactions of specific duplexes containing the sequence slectivity for MC or 2,7-DAM were performed under the conditions described in Materials and Methods. The alkylated and nonalkylated DNA strands are separated as a single fraction by passing the reaction mix over a Sephadex G-25 column (figure 3).

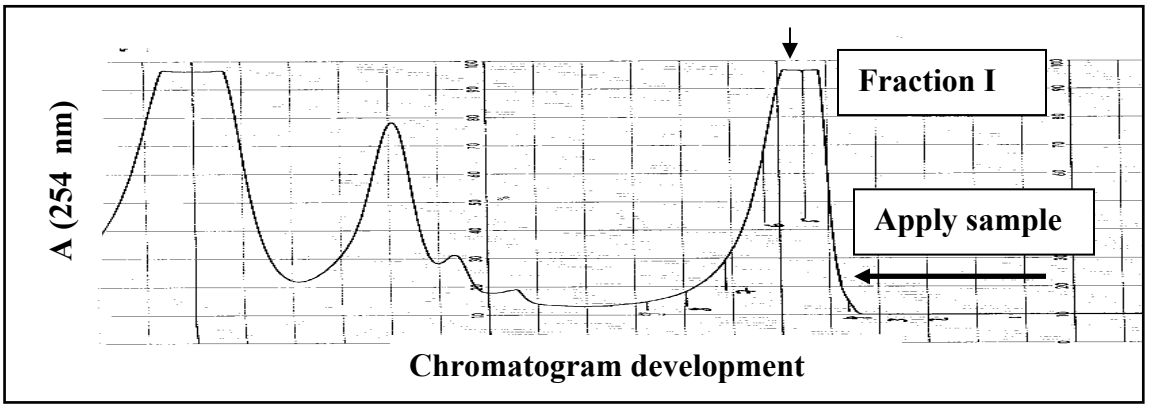


Figure 4: Analysis of oligonucleotides mixture (C1 oligo = 5'-ACACG(MC)TCAT-3' and C2 3'-TITGCAIT-5' = 8 mer oligo) by Sephadex-G 25 gel filtration chromatography after alkylation with mitomycin C. 1st peak (Fraction I) contains the mixture alkylated and non alkylated C1 and C2 oligos.

Fraction I from sephadex G-25 was further used to purify the alkylated oligonucleotides using preparative HPLC as described in the Materials and Methods section. As can be seen from Figure 4 a very powerful separation of alkylated strands from non-alkylated oligonucleotide was achieved using a gradient of 20-60% B in 440 minutes were: Buffer A: 0.1 M triethylammoniumacetate (TEAA), pH=7.0. Buffer B: 30% CH_3CN in 70 % 0.1M TEEA, pH=7.0. the structural characterization of the purified peaks eluted at 67.62 min and 99.15 min was performed using both electrospray ionization ("-" mode) mass spectroscopy ((ESI-)-MS) and nucleosides analysis on analytical C18 columns, after digestion of alkylated oligonucleotides with snake phosphodiesterase and alkaline phosphatase as described in Materials and Methods. The alkylated oligos are having always a ratio of Absorption at 260 nm/Absorption at 320 nm (A(260 nm)/A(320 nm)) between 10-14 (13-14-Chapter 1). As can be seen from HPLC chromatogram in Figure 4 the alkylated DN A strands have higher absorption on channel B (set at 320 nm) than the non-alkylated parent strands which have lower absorption on channel B (at 320 nm). These UV characteristics for alkylated oligos allowed their original purification from the non-alkylated (parent DNA strands) provided the gradied performed enabled their separation as is demonstratd in the preparative HPLC run presented in Figure 4.

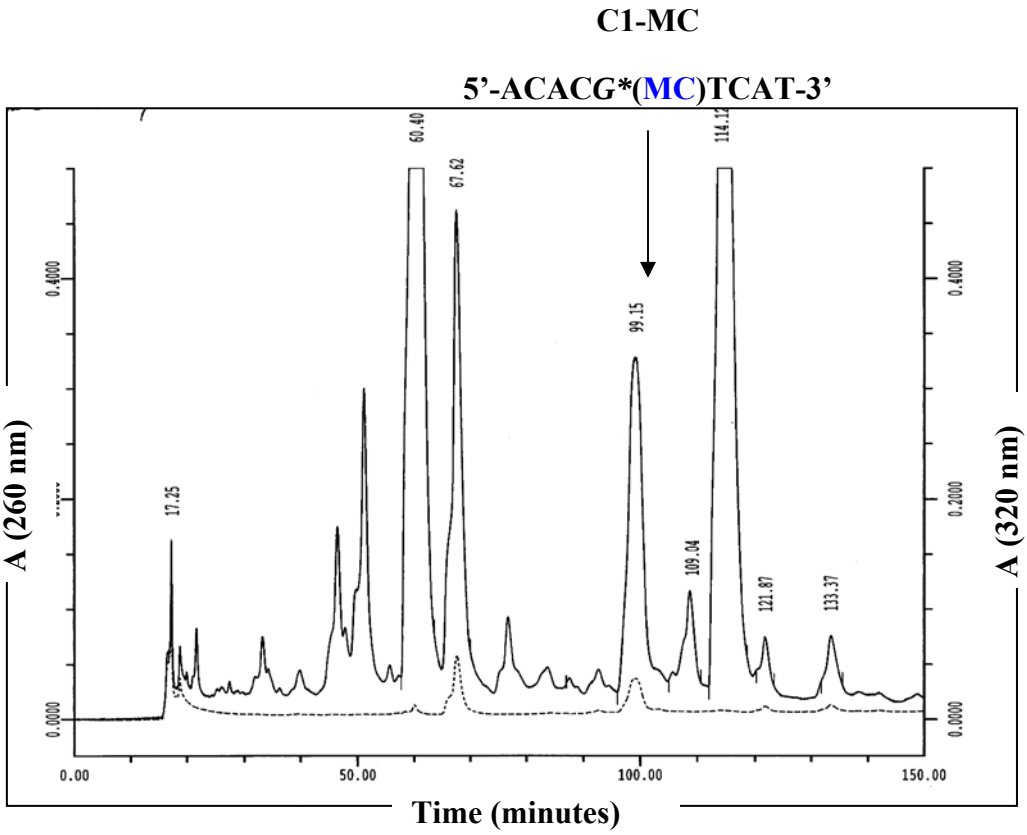


Figure 5. **Analytical Reverse Phase HPLC of the reaction mixture containing 5'-ACACG(MC)TCAT-3' alkylated oligo.** The conditions used for running the HPLC are described in the Materials and Methods section. The peak eluted at 99.15 minutes was identified through ESI-MS (- mode) to be the oligonucleotide C1-MC, alkylated with mitomycin C (MC), i.e. 5'-ACACG*(MC)TCAT-3' using the procedure described in Materials and Methods.

The purity control of the MC alkylated oligo C1 was checked by performing analytical reverse-phase chromatography using C4 column (Rainin), 300 angstoms, Microsorb, and the same TEAA buffer system used for the preparative and semi preparative work as described under Materials and Methods. The purified C1-MC alkylated oligo has the retention time 22.78 minutes and is more than 95 % pure. This C1-MC alkylated oligo was further used in the ligation experiments for obtaining the 24 mer template for polymerases.

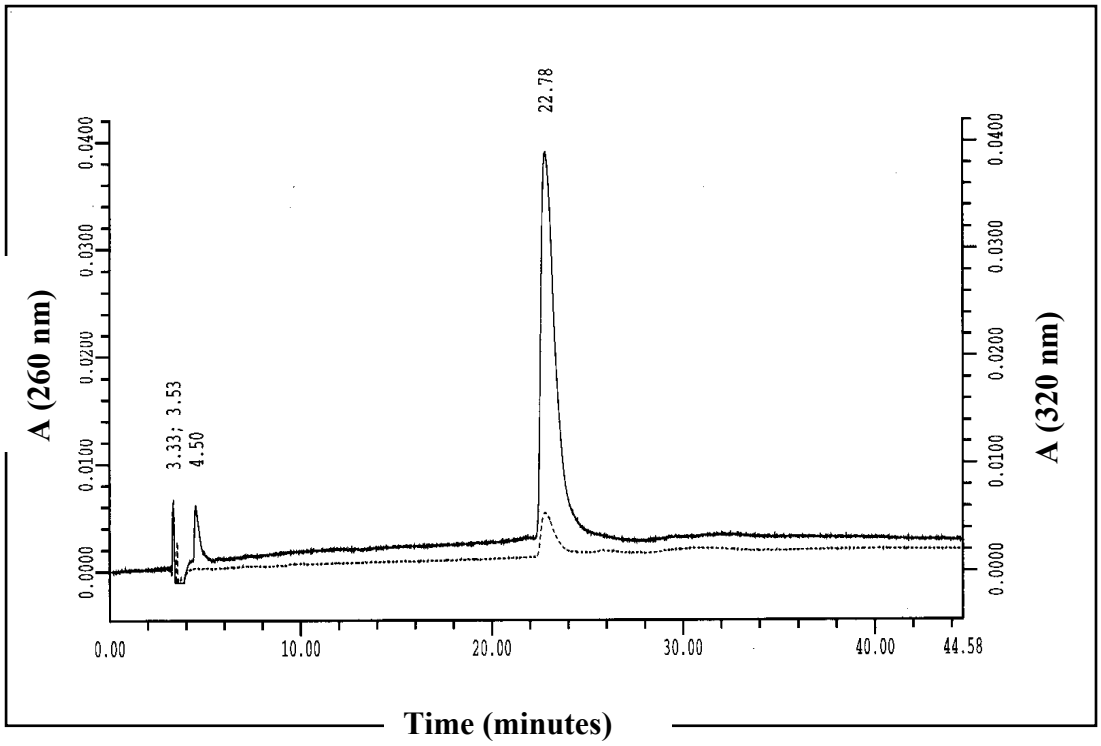


Figure 6. The reversed phase HPLC control of purity for: 5'-A C A C *G* *(MC) T C A T-3' alkylated oligo. The appearance of a double signal (both at 254 nm and at 320 nm) is an indication of the presence of a mitomycin C alkylated guanine as described in Materials and Methods. The peak eleuted at 22.78 minutes was characterized by ESI(-) mass spectroscopy and by analytical analysis of the base composition as being the 5'-A C A C *G* *(MC) T C A T-3' oligo containing one single guanine alkylated with MC (underlined) (see Materials and Methods, Section2.3.1.).

The structural identification of the alkylated guanine at the NH2 position in C1 oligo was performed by running reverse phase HPLC of the corresponding alkylated oligo after digestion with snake venom phosphodiesterase and alkaline phosphatase as described earlier (15-Chapter 1) and under Materials and Methods. In a parallel control run, the non alkylated C1 oligo was digested in the same conditions with the two enzymes (snake venom phosphodiesterase and alkaline phosphatase), to determine the retention time for each individual nucleoside (dA, dT, dC and dG). The chromatogram in

figure 6 (A) support the purity of the alkylated C1 oligo since after digestion with snake venom phosphodiesterase and alkaline phosphatase, there is no free dG eluted (the retention time for dG is around 5 minutes). All the dG is shifted at higher retention time (19.4 minutes) due to the presence of the MC drug covalently bound to NH2 exocyclic position. The structure of the monoadduct of guanine with mitomycin C at NH_2 position was confirmed by taken the scan of the peak at 19.44 minutes between 200 and 400 nm. The presence of the monoadduct is confirmed by the appearance of the UV characteristics at 252 nm and at 315 nm. The ratio 215nm/352nm is 2.3 in the case of monoadducts (15-Chapter 1). The spectra of the monoadduct is presented in figure 6 (B).

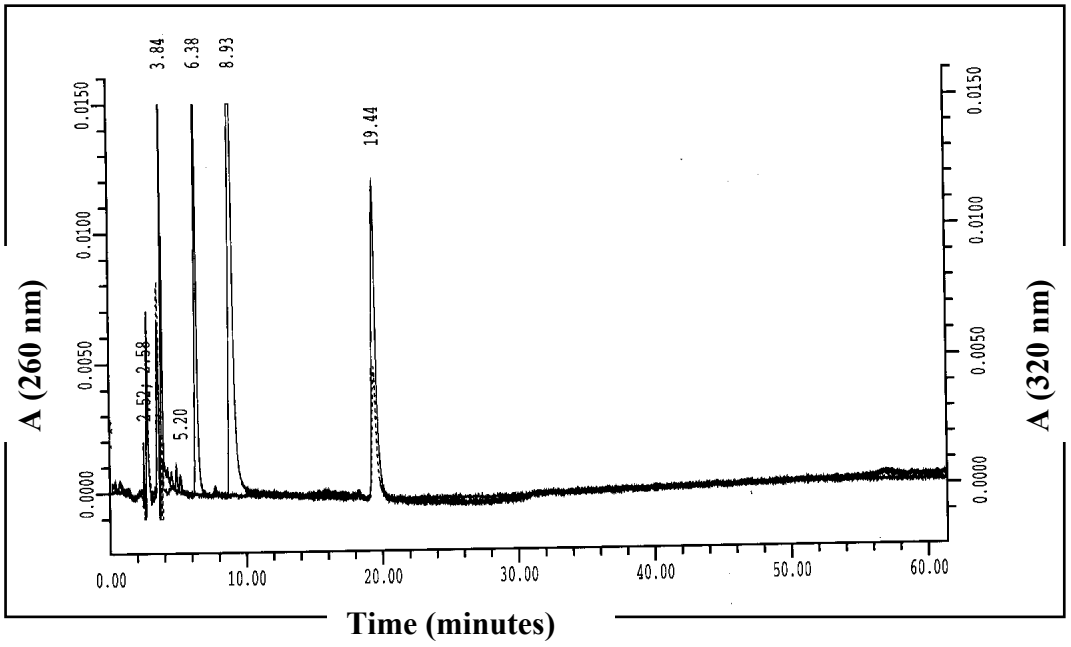


Figure 7 (A-B). The reverse-phase HPLC analytical run of the mixture of nucleosides containing the alkylated guanine from the C1-5'-A C A C G *(MC) T C A T-3' oligo, after snake venom phosphodiesterase and alkaline phosphatase treatment. The monoadduct of guanine with mitomycin C at N2 position (adduct 7 in Figure 1) has the peak shifted at 19.44 minutes as it was identified using authentic standard of this monoadduct (see Materials and Methods Sections 2.3.1-2.3.

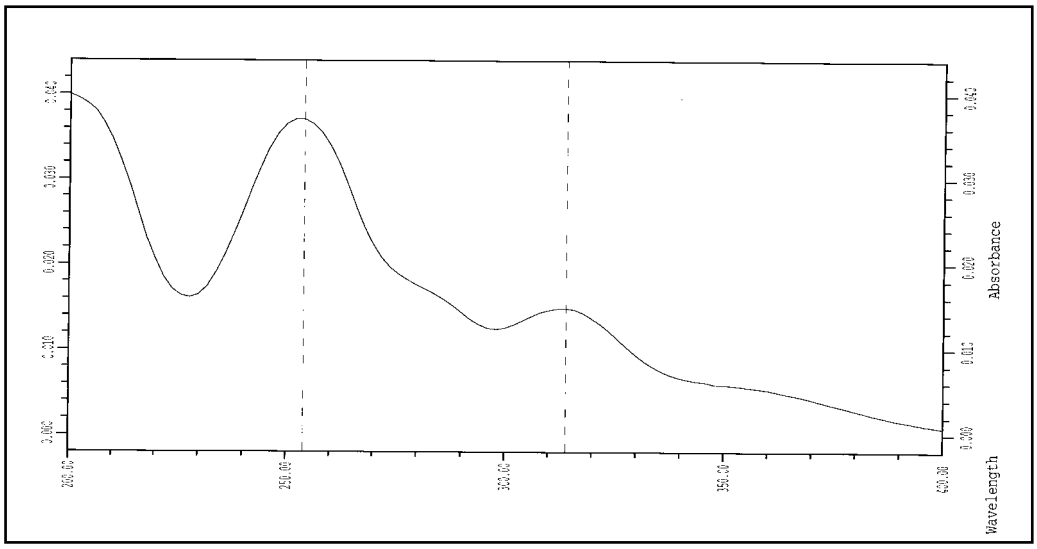


Figure 6 (B). The UV spectrum of guanine N^2-MC-monoadduct 7 (Figure 1).

The spectrum of the alkylated dG-nucleoside eluted at 19.4 min in the reversed-phase HPLC (Figure 6(A)) was recorded using the photodiode array detector of the HPLC system. The UV-spectrum shows the spectral characteristics described already in the literature for this monoadduct (11-15-Chapter 1).

The expected adduct and nt compositions **of C1-5'-A C A C G *(MC) T C A T-3' oligo** were verified by ESI-MS (-) as described in Materials and Methods. For comparison the MS scan of the control non-alkylated oligonucleotide was also recorded (see figures 7 (A) and 7 (B)).

Table 1 summarizes the theoretically predicted molecular weights (MW) and the experimentally determined MW for the oligonucleotides C1 and C1-MC. The MW for the oligonucleotide C1-MC (5'-ACACG(MC)*TCAT-3') was calculated from the MW of the parent oligonucleotide C1 (not alkylated) using the folowing formula: MW of alkylated C1 oligonucleotide = theoretical MW of non-alkylated C1 oligo (2683 Da) to which the MW of the drug mitomycin C was added (MW for MC is 334 Da) after which the MW of the methanol (MW=32.0 Da) was substracted, since when the guanine is

alkylated with MC it looses methanol = 2683Da+334Da-32Da=2985Da predicted MW for the oligonucleotide 5'-ACACG(MC)*TCAT-3'.

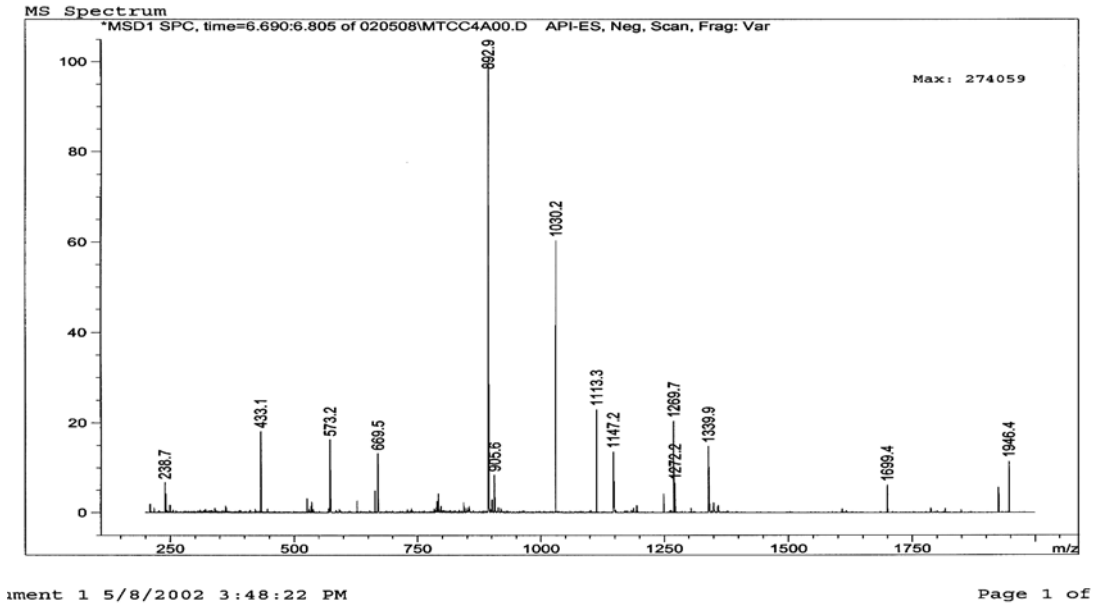


Figure 8-A. ESI-(-) MS spectrum acquired for the control, parent oligonucleotide **5'-ACACGTCAT-3'** (see Materials and Methods). The peaks were deconvoluted and fitted to the molecular ion mass of **5'-ACACGTCAT-3' (Table 1).**

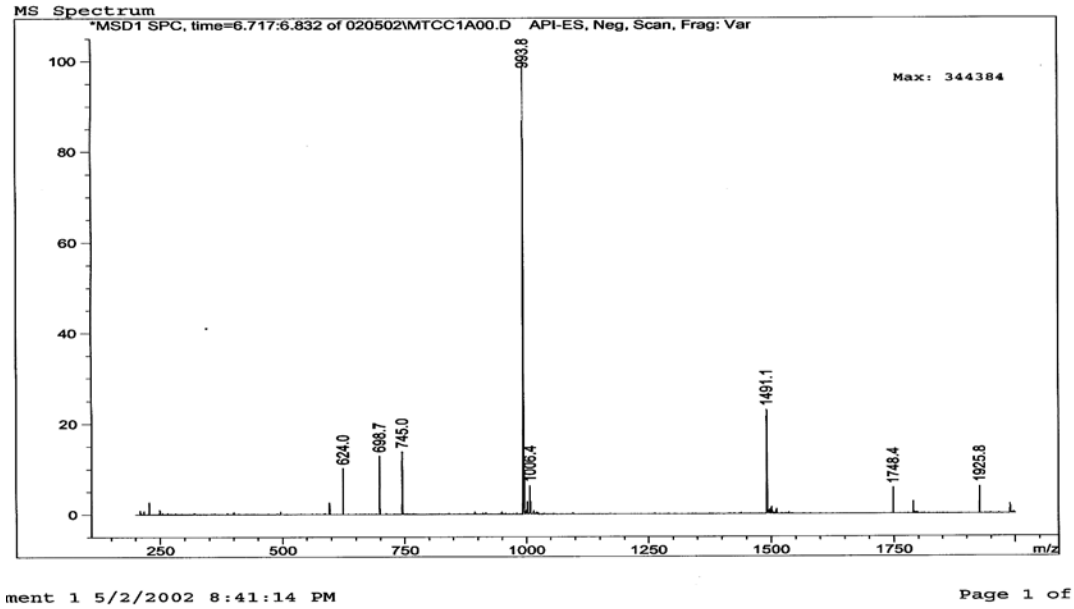


Figure 7 (B). MS spectrum acquired for the alkylated 5'-ACACG(MC)*TCAT-3' oligonucleotide (see Materials and Methods). The peaks were deconvoluted and fitted to the molecular ion mass of **5'-ACACG (MC) TCAT-3' (Table 1).**

Table 1: The molecular masses of the parent and alkylated 5'-ACACGTCAT-3' oligonucleotides as determined from ESI-MS (-) (Materials and Methods).

Oligonucleotide	**Theoretically predicted MW (Da)**	**Experimentally determined MW (Da)**
5'-ACACGTCAT-3'	**2683**	**2682.13**
5'- ACACG(mc)TCAT	**2985**	**2984.48**

In conclusion, the structural investigations (reversed-phase HPLC and ESI_MS) support the fact that the oligonucleotide 5'- ACACG(MC)TCAT-3' (C1-MC) incorporated one single drug molecule at the specific guanine residue.

2.3.1.2. Synthesis and structural characterization of C6 oligonucleotide 5'-CTGG(2,7-DAM)TAATTTAC-3'.

As for the case of MC-alkylated oligonucleotides the alkylated and non-alkylated DNA strands are separated as a single fractiopn by passing the reaction mixture over a Sephadex G-25 column (figure 8). The alkylation of DNA duplex 5'-CTGG*TAATTTAC-3' (C6 oligonucleotide/ 3'-GACC ATTAA-5' (C7 oligonucleotide) was performed nder reductive conditions as described in Materials and Methods.

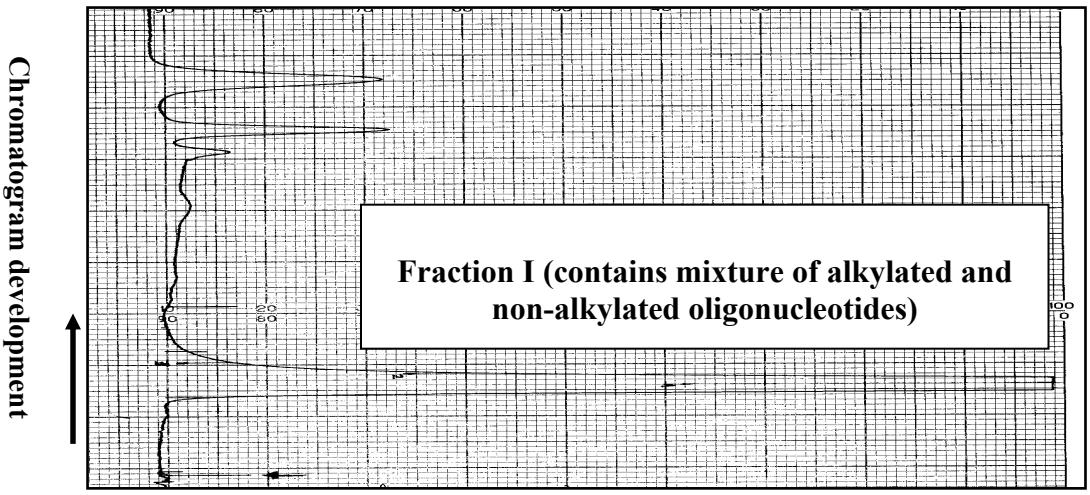


Figure 9: **Separation of the C6 oligonucleotide (5'-CTGG*TAATTTAC-3') and C7 oligonucleotide (3'-GACC ATTAA-5') from the alkylation reaction mixture by gel-filtration chromatography on Sephadex G-25 column.**

Fraction-I peak collected from Sephadex G-25 was further purified for the alkylated oligonucleotide using reversed phase HPLC. In figure 9, the reverse-phase HPLC semi-preparative Fraction-I collected from gel-filtration chromatography is presented. The buffer system used was: Buffer A: 30 mM phosphate buffer, pH=5.4. Buffer B: 30% CH_3CN in 70% 30 mM phosphate buffer, pH=5.4). The gradient: 10%-25%B in 10 minutes, 25-33%B in 60 minutes and 33-100%B in 20 minutes. The column was a semipreparative, Dynamax C18 column. 50 ODs from the reaction mixture (C6-2,7-DAM +parent C6 +parent C7) was applied to the column and run at 3 ml/min. The absorbance recordings were at 260 nm (channel A) and 320 nm (channel B) as indicated on the HPLC chromatogram.

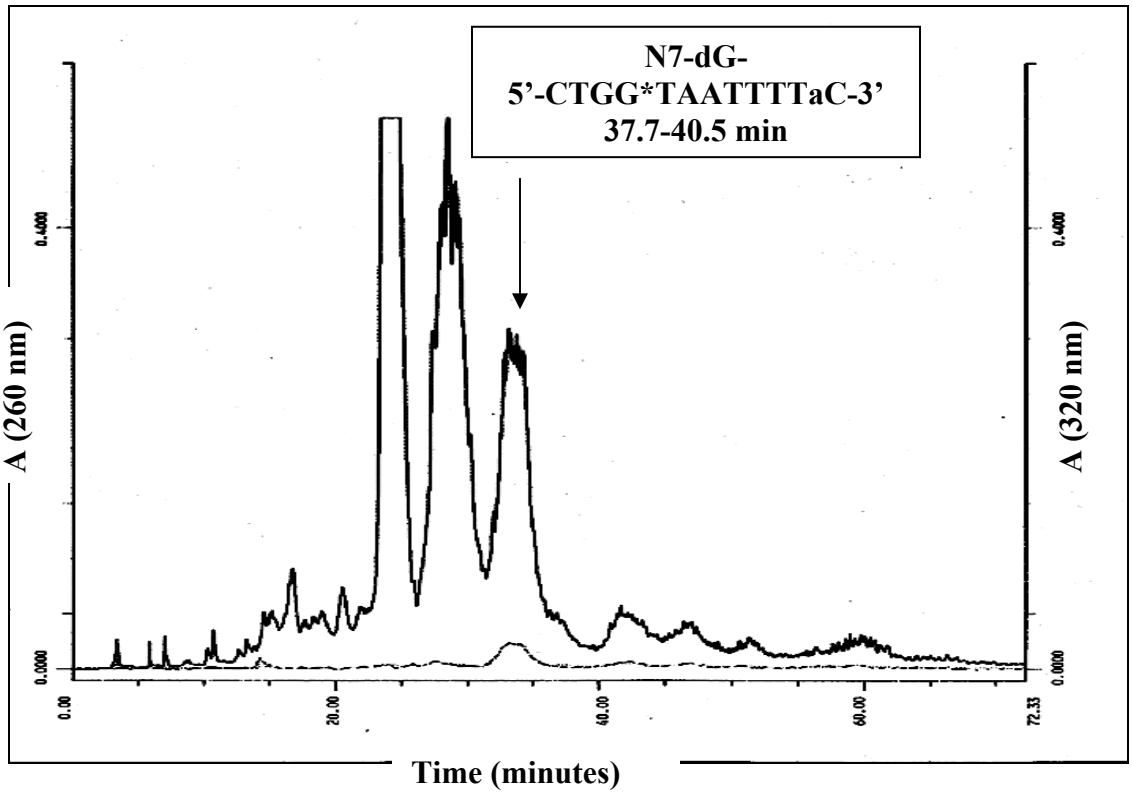


Figure 10. **Reversed-phase HPLC semipreparative purification of C6-5'-CTGG* (2,7-DAM)TAATTTAC-3'alkylated oligo on semipreparative C18 column** (conditions for running the HPLC are described in Materials and Methods). The alkylated oligonucleotides are eluted later than the non-alkylated parent oligonucleotides. The structural identity of the alkylated oligonucleotide (indicated with the arrow) was determined by ESI(-)-MS) as described in materials and methods and presented in figure 12 (B) and table2.

Figure 9 shows that there are many peaks that absorb both at 260 and at 320 nm suggesting that adducts other than N7-2,7-DAM monoadduct were formed. In most cases the 2,7-DAM-G-N7 monoadducted-oligonucleotide eluted immediately after the parent oligonuleotide. As can be seen in figure 9 the C6-oligo alkylated with 2,7-DAM was eluted after parent oligo at 37.7-40.5 minutes. This individual peak was collected, lyophilized and subjected to further purification using semi-preparative HPLC methods described in Materials and Methods. In addition the analytical reversed-phase HPLC was used to test the purity of the 2,7-DAM alkylated oligonucleotide. In figure 10 the HPLC

trace for the analytical control of the purity of the N7-G-2,7-DAM monalkylated oligonucleotide **5'-CTGG* (2,7-DAM)TAATTTAC-3'** is presented.

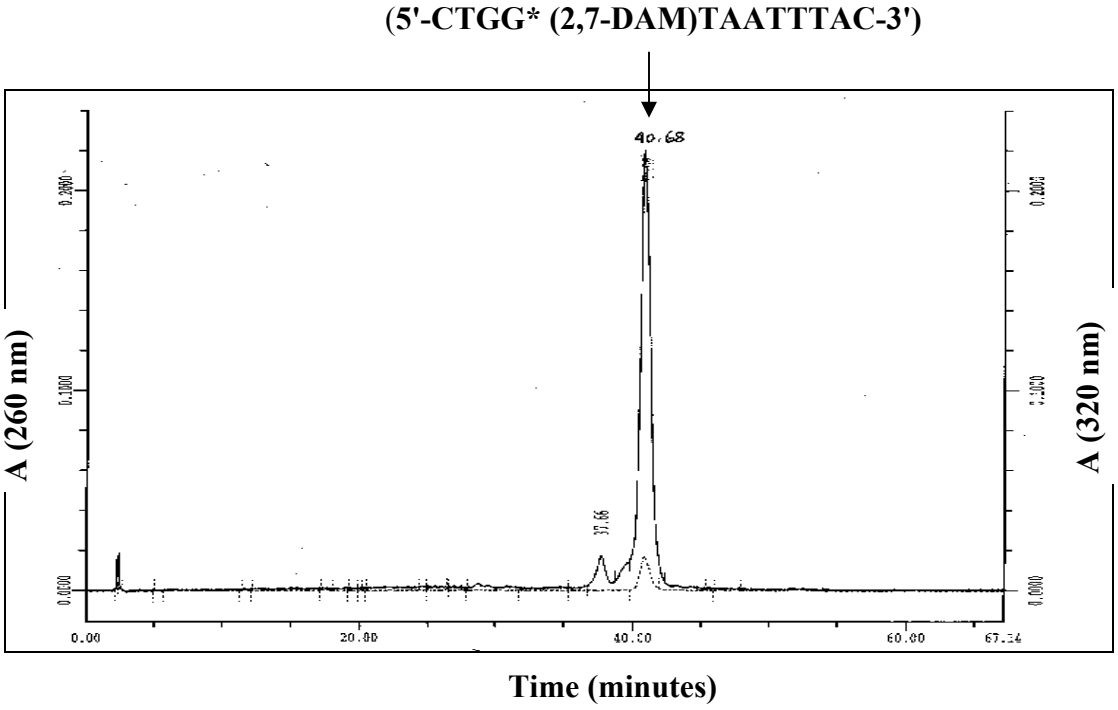


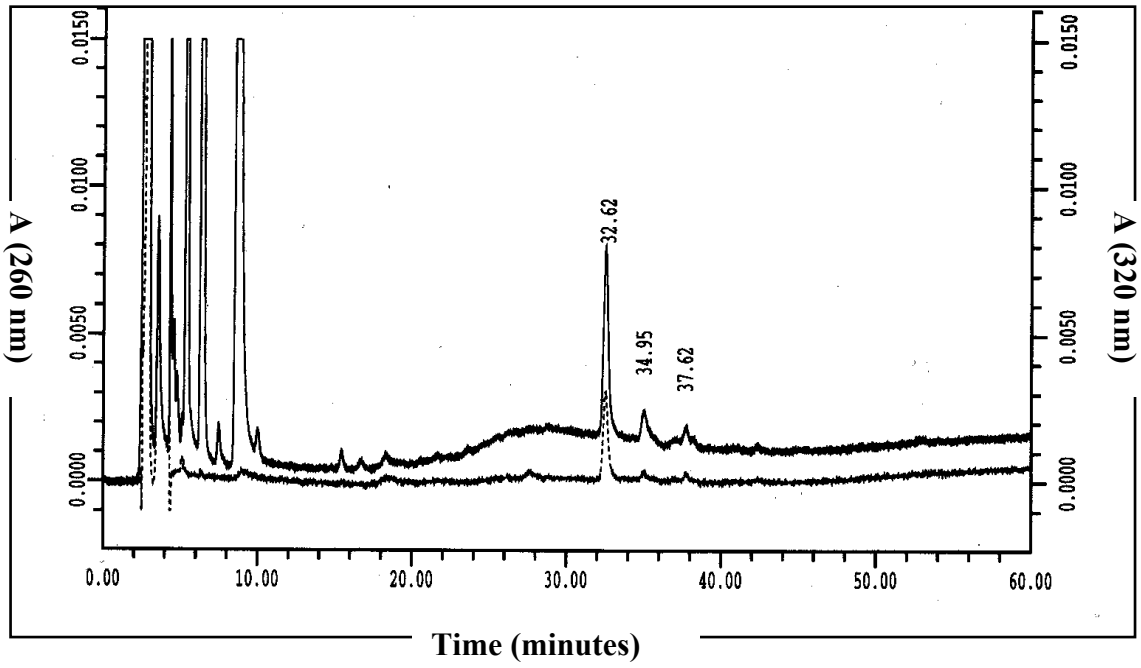
Figure 11. Analytical HPLC reverse phase control of purity of C6-5'-CTGG*(2,7-DAM) TAATTTAC-3' oligonucleotide. The structural identity of the alkylated oligonucleotide was determined by ESI-MS (-) mode as described in Materials and Methods.

As can be seen from the analysis of the chromatogram from figure 10, the C6-2,7-DAM oligo is pure (>95%) and elutes at 40.68 minutes in the HPLC reversed system used: C18 analytical column, run at 1ml/min with the following gradient: 10-25% B in 10 minutes, 25%-33% B in 60 minutes and 33%-100% B in 20 minutes. Buffer A: 30 mM KH_2PO4, pH=5.5; Buffer B: 30% acetonitrile into 70% 30 mM KH_2PO4, pH=5.5.

The structural identity of the G-N7-2,7-DAM monoadduct was determined by subjection to digestion with snake venom phosphodiesterase and alkaline phosphatase of alkylated C6-2,7-DAM oligonucleotide using the same procedure as that described for the digestion of oligos alkylated with mitomycin C (see Materials and Methods).

In the case of an oligonucleotide having **5'----GG----3'** tract, where the second guanine from 5'end is chemically alkylated with 2,7-DAM at N7-position, there is a stable structure containing the two Gs linked together, from which only one guanine carries the 2,7-DAM drug. This structure is resistant to digestion by snake venom phosphodiesterase and alkaline phosphatase (11-15-Chapter 1). In order to obtain just one single N7-G-2,7-DAM it is necessary to heat the digested sample 1 hour at 90^{o}C. Upon depurination the N7-Guanine monoadduct with 2,7-DAM is released (adduct 8, figure 1). Figure 11 (A) shows the analytical HPLC of the digested nonheated samples and figure 11 (B)-the digested heated sample. The structural assignment of the N7-G*-2,7-DAM monoadduct was performed by comparison with an authentic standard described previously (see also Appendix-I-Figure 69) and by inspection of its own UV spectrum (11-15-Chapter 1). The UV-scan of the released G-2,7-DAM monoadduct was recorded during HPLC run using the built-in photodiode array detector. The spectral characteristic of the N7-G monoadduct is the ratio of 1 between the absorbance at 285 nm and the absorbance at 314 nm (figure 12). In Appendix (I) the preparation of an authentic standard for N7-G-2,7-DAM is described together with the HPLC chromatogram used to identify the adduct and its own UV spectrum (Figure 69).

A) Analysis of digested C6-alkylated unheated oligo on C18 analytical column.



B) Analysis of digested C6-alkylated oligo on C18 analytical column after depurination at 90^{o}C for 1 hour as described earlier (15-Chapter 1).

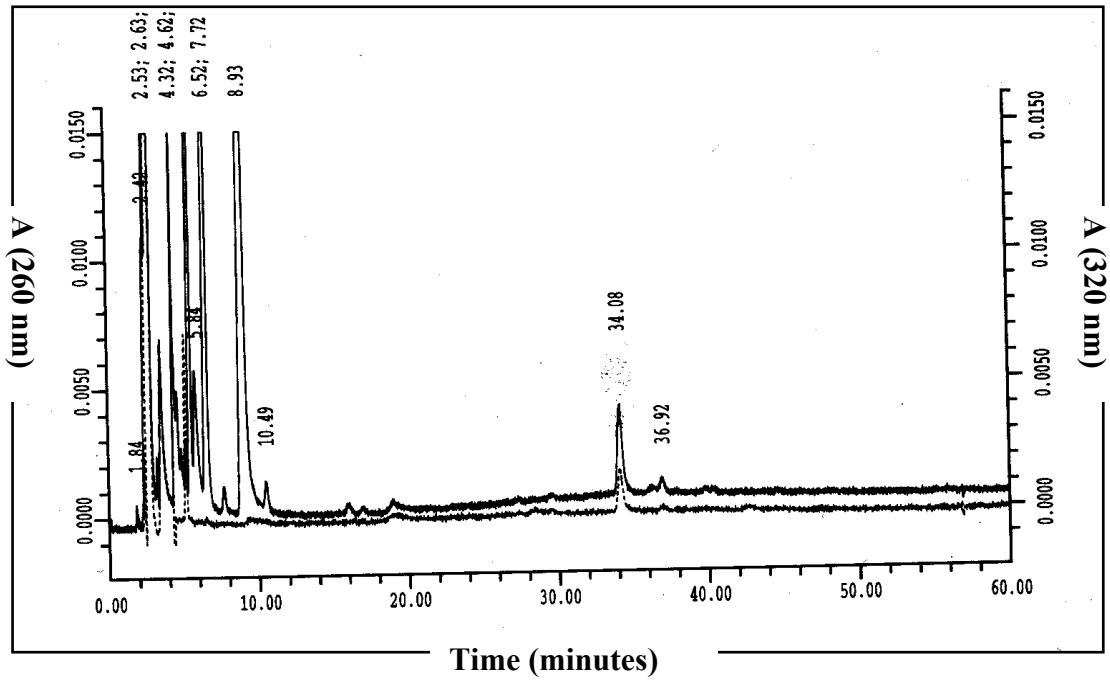


Figure 12 (A-B). HPLC analysis of N7-2,7-DAM monoadduct present in the C6-5'-CTGG*(2,7-DAM) TAATTTAC-3'oligo.

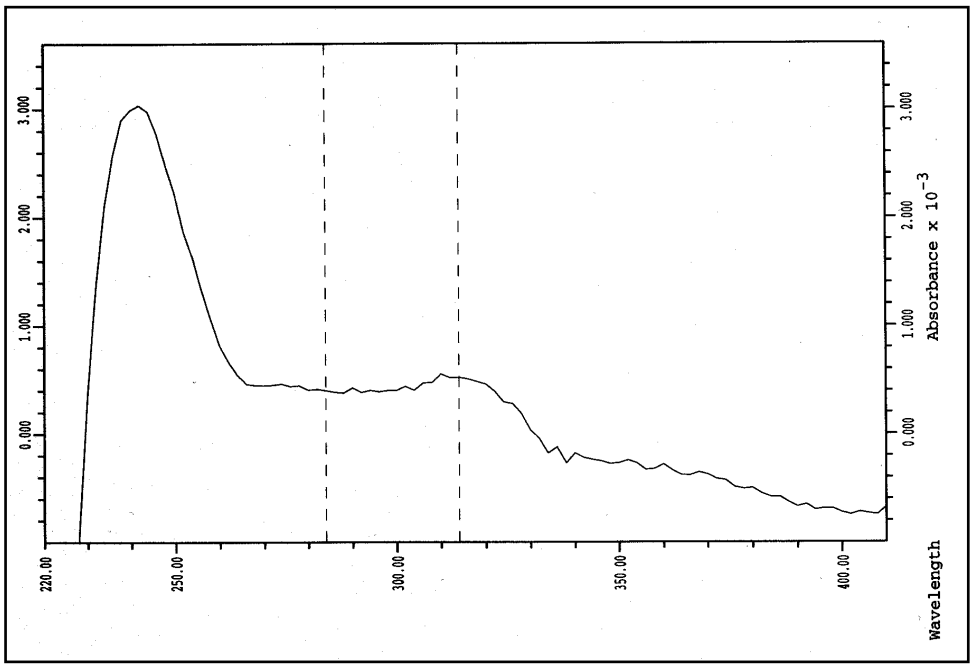


Figure 12: The UV spectrum of guanine N7-2,7-DAM-monoadduct 8 (Figure 1). The UV spectrum is identical with that of the standard-control shown in Figure 69 (B) in Appendix (I).

The elution conditions for both the heated and unheated sample presented in the HPLC chromatograms in Figure 11 (A-B) were: Column: C18, 100 angstroms, 5 μm, Microsorb Rainin; Buffer A: 20 mM NH_4Acetate, pH=5.5; Buffer B: 30% acetonitrile in 70% 20 mM NH_4Acetate, pH=5.5 and the gradient applied: 20%-60% B in 60 minutes, with a flow rate of 1 ml/min. The retention time of the N7-guanine monoadduct 8 (figure 1) is 34.05 minutes (Figure 11) (A-B). The spectral characteristics of the N7-G-2,7-DAM monoadduct are presented in Figure 12.

The expected adduct and nt compositions of **5'-CTGG*(2,7-DAM)TAATTTAC-3' (C6-2,7-DAM)** were verified by ESI-MS (-) as described in Materials and Methods. For comparison the MS scan of the control non-alkylated oligonucleotide was also recorded (figures 13 (A) and 13 (B)).

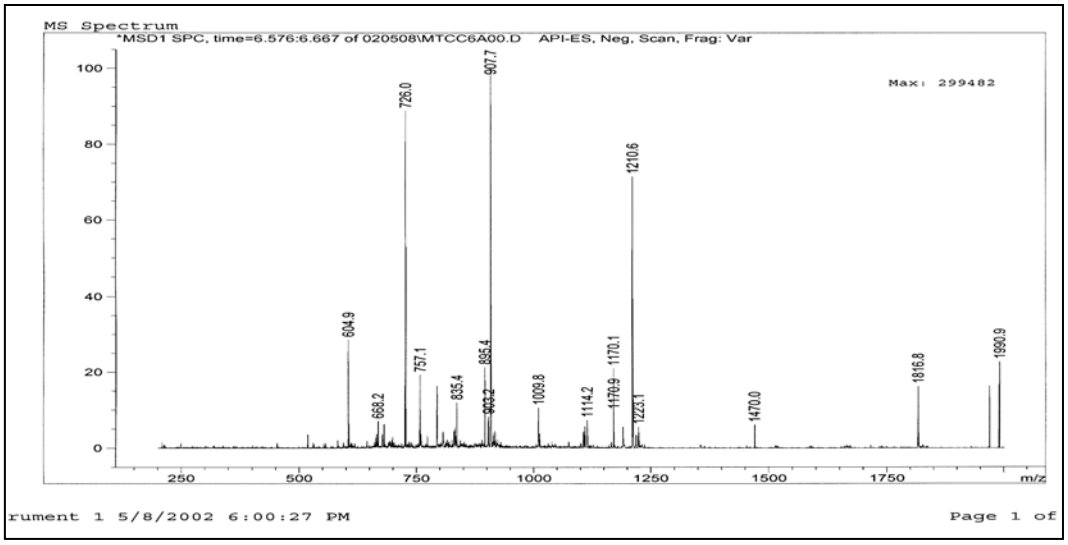


Figure 13 A. **The ESI-MS spectrum for the parent oligonucleotide 5'-CTGGTAATTTAC-3'.** The deconvoluted spectra shows the presence of an oligonucleotide with a MW of 3635.06 Da as expected from theoretical calculations (table 2) (we assumed that the oligo is ionized with (-3), (-4) or (-5) charges corresponding to the m/z ratio of 1210.6, 907.7 and 726.0 respectively).

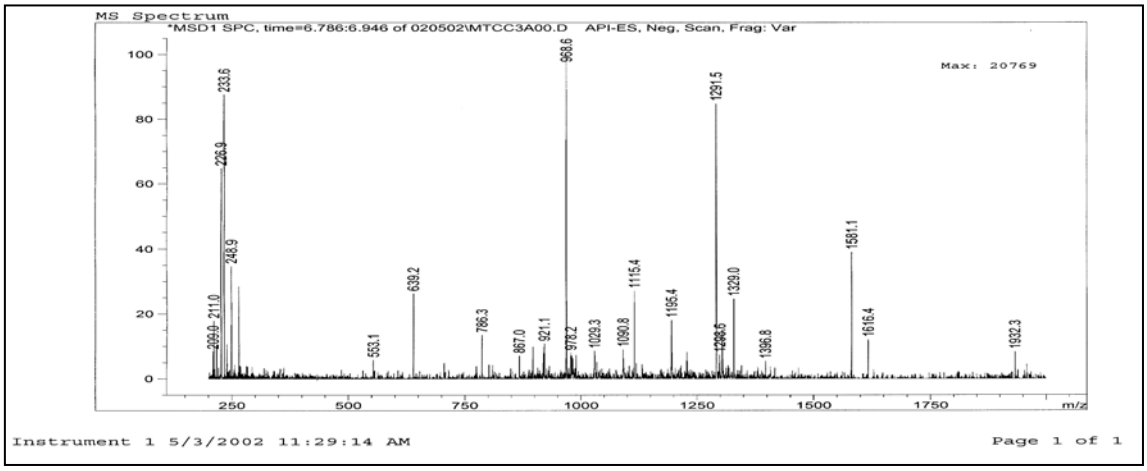


Figure 13 (B). The ESI-MS spectrum for the G-N7-2,7-DAM alkylated oligonucleotide 5'-CTGG*(2,7-DAM) TAATTTAC-3'. The deconvoluted spectra shows the presence of an oligonucleotide with a MW of 3877.92 Da as expected from the theoretical calculations (see Table 2) (we assumed that the oligo is ionized with (-3), (-4) or (-5) charges corresponding to the m/z ratio of 1291.5 and 968.6 respectively).

Table 2 summarizes the theoretically predicted molecular weights and the experimentally determined MW for the oligonucleotides C6 and C6-2,7-DAM. The

for the oligonucleotide C6-2,7-DAM 5'-CTG*G*(2,7-DAM)*TAATTTAC-3' was calculated from the MW of the parent oligonucleotide C6 (not alkylated) using the following formula: MW of alkylated C6 oligonucleotide = theoretical MW of non-alkylated C6 oligo (3634 Da) to which the MW of the drug 2,7-DAM incorporated in the oligonucleotide was added (MW for [2,7-DAM+guanine] is 394 Da and the MW of Guanine is 151.1, such that the MW of the 2,7-DAM drug incorporated into the oligonucleotide is 394-151.1 = 242.9 Da). Thus the MW of the C6-oligo alkylated at one single guanine with 2,7-DAM is expected to be 3634Da + 242.9Da = 3876.9Da.

Table 2: Mass of parent and of alkylated 5'-CTGG*(2,7-DAM)TAATTTAC oligonucleotides determined by ESIMS (-).

Oligonucleotide	Theoretically predicted MW (Da)	Experimentally determined MW (Da)
5'-CTGGTAATTTAC	**3634**	**3634.90**
5'-CTGG*(2,7-DAM)TAATTTAC	**3877**	**3877.92**

***mark the position of alkylated guanine.**

All the structural investigations (reversed-phase HPLC and ESI-MS) support the identity of the oligonucleotide **5'- CTGG*(2,7-DAM)TAATTTAC-3'** (alkylated at guanine with 2,7-diaminomitosene (2,7-DAM)). This oligonucleotide was further used in the ligation experiments to obtain 24 and 36 mers templates for polymerases.

2.3.2. Synthesis, purification and structural characterization of an oligonucleotide containing a single N^2-dG-2,7-DAM monoadduct; *in vitro* conditions that enhanced the selective alkylation of DNA duplexes for favoring N7-G*-2,7-DAM vs. N2-dG-2,7-DAM monoadducts.

Beside the duplex **C6-(5'-CTGG*TAATTTAC-3')** and **C7** oligonucleotide (**3'-GACCATTAA-5'), duplex I,** an independent construct, namely **duplex II,** 5'-CTA**G**T**GG**TATCC-3' (CI) / 3'-TCACCATAGG-5' (CIII) was akylated with 2,7-DAM under reductive conditions which enhanced the production of either N-7-2,7-DAM monoadduct or N^2-dG-2,7-DAM monoadduct. The N^2-dG-2,7-DAM monoaadduct was first discovered and structurally characterized using UV spectroscopy by Palom et al. (23-Chapter-1). It was in our interest to obtain a pure oligonucleotide containing the N^2-dG-2,7-DAM monoadduct which can be further ligated within 24 mer templates and subjected to primer extension in vitro reactions. Thus, an enhanced structure-activity relationship (SAR) of N7-G-2,7-DAM and N^2-dG-2,7-DAM could be performed.

2.3.2.1. Reductive conditions that enhanced the production of N7-dG-2,7-DAM monoadduct using the DNA duplex (I). There was a preference for the alkylation at N7-G-position by 2,7-DAM when the alkylating agent was having higher percentage of decarbamoyl 2,7-DAM (>30%). The conditions for alkylation are described below:

Alkylating agent: 2,7-DAM: 60% 2,7-DAM and 40% decarbamoyl 2,7-DAM (as determined by HPLC).

Hydrogenation time: 25-45 minutes: the longer the hydrogenation time the higher the yield of the N7-G-2,7-DAM monoadduct obtained.

Argon time: 60-90 minutes; after 1 hour the N7-dG-2,7-DAM monoadduct is reaching a plateau (figure 15). Preparative alkylation of oligos **C6 (5'-C T G G*T A A T T T A C-3' = 12 mer)** annealed with **C7 (3'-G A C C A T T A A-5' = 9 mer)** (duplex (I)-DNA) with 2,7-DAM was performed as described earlier in section 1.2.

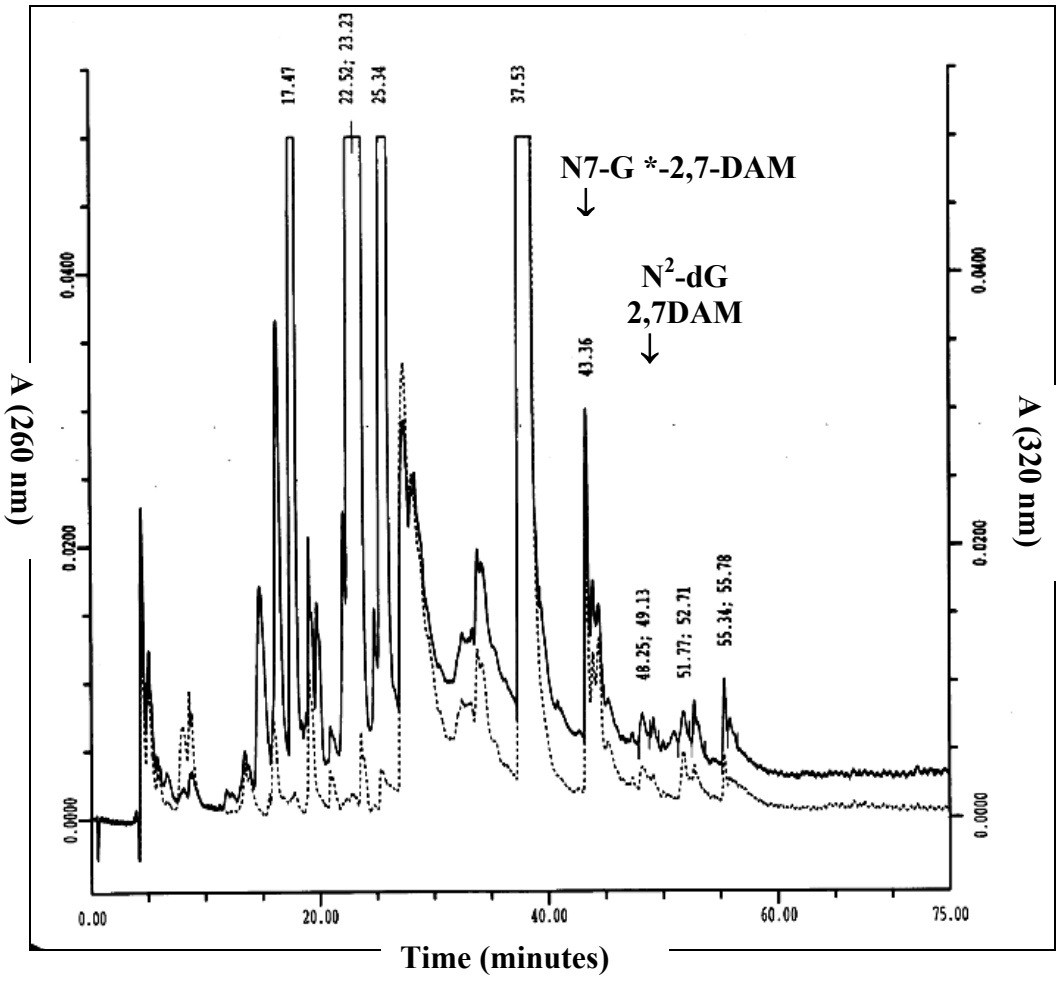


Figure 14: Snake Venom Phosphodiesterase (SVD) treatment of the oligonucleotide mixture at 10 minutes Argon treatment after 40 minutes hydrogenation, under the conditions described at section 1.3.1. It can be seen that the N7-dG-2,7-DAM monoadduct is obtained in the highest yield, as compared with N2-dG-2,7-DAM monoadduct which is obtained in lower yield in these conditions. Both monoadducts were identified based on authentic standards and based on their UV-scans as described in Materials and Methods.

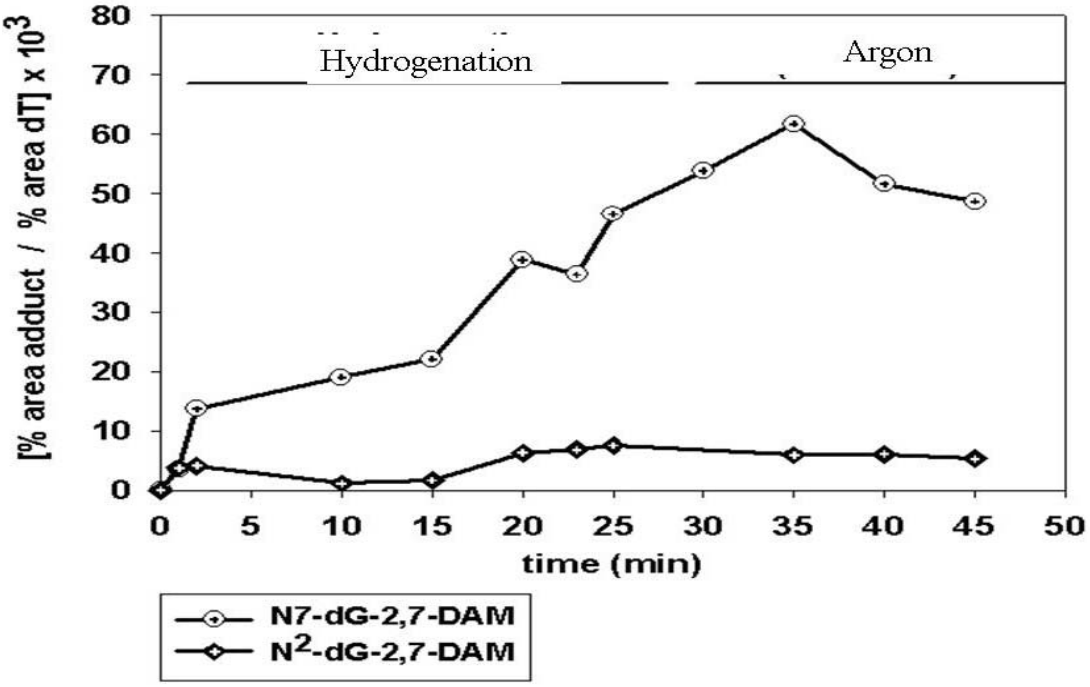


Figure 15: The production of N7-G-2,7-DAM monitored in time by reversed-phase HPLC. The ratio % area of N7-dG-monoadduct and % area of N^2-dG-monoadduct to % area of dT as a function of time of alkylation reaction of duplex (I) (C6-(5'-CTGG*TAATTTAC-3') and C7 oligonucleotide (3'-GACC ATTAA-5')) is plotted. The % area for each monoadduct was calculated using the integration procedure built-in the software for data analysis of the Beckman HPLC system. The hydrogenation time was 25 minutes while the alkylation under argon was during 1 hour, but the graph shows only the first 20 minutes of reaction progress under argon.

The N7-G-2,7-DAM monoadduct is produced at higher yield and with faster rate during both the hydrogenation and the argon time periods. Originally the hydrogenation time was 15-18 minutes (14-15-Chapter 1) but figure 14 supports the fact that increasing hydrogenation time enhances the yield of N7-G-2,7-DAM monoadduct. The N7-G-monoadduct continues to increase in time during argon period when it reaches a plateau before 60 minutes. The chemical mechanism responsible for enhancing the production of

N7-G-monoadduct vs. N^2-dG-monoadduct in the presence of increased amounts of decarbamoyl 2,7-DAM (> 40%) is not known by this time and is still under investigation.

2.3.2.2. Reductive conditions enhancing the production of N7-G-2,7-DAM monoadduct using the DNA duplex (II): 5'-CTAGTGGTATCC-3'-(CI)-3'-TCACCATAGG-5' (CIII). The reductive conditions used for alkylation of duplex (II) with 2,7-DAM are the same as those described in section 2.3.2.1. The oligonucleotide carrying the N7-G-2,7-DAM monoadduct was purified by semipreparative reverse phase HPLC as described in Materials and Methods and presented in figure 16. The purified oligonucleotide was further analyzed on analytical HPLC (nonheated sample-figure 17 (A) and heated sample for depurination-figure 17 (B)).

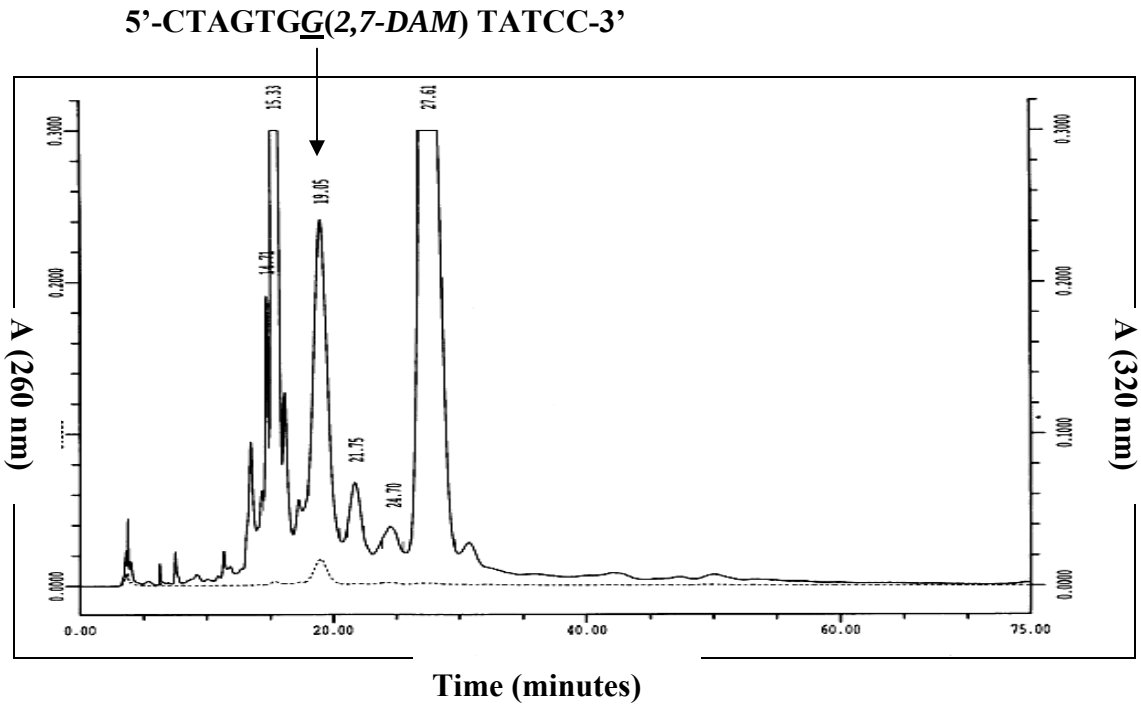


Figure 16: **Reversed-phase HPLC semipreparative purification of C-I-5'-CTAGTGG(2,7-DAM) TATCC-3' alkylated oligo on semipreparative C18 column.** The conditions used to run the semipreparative HPLC column are described in Materials and Methods.

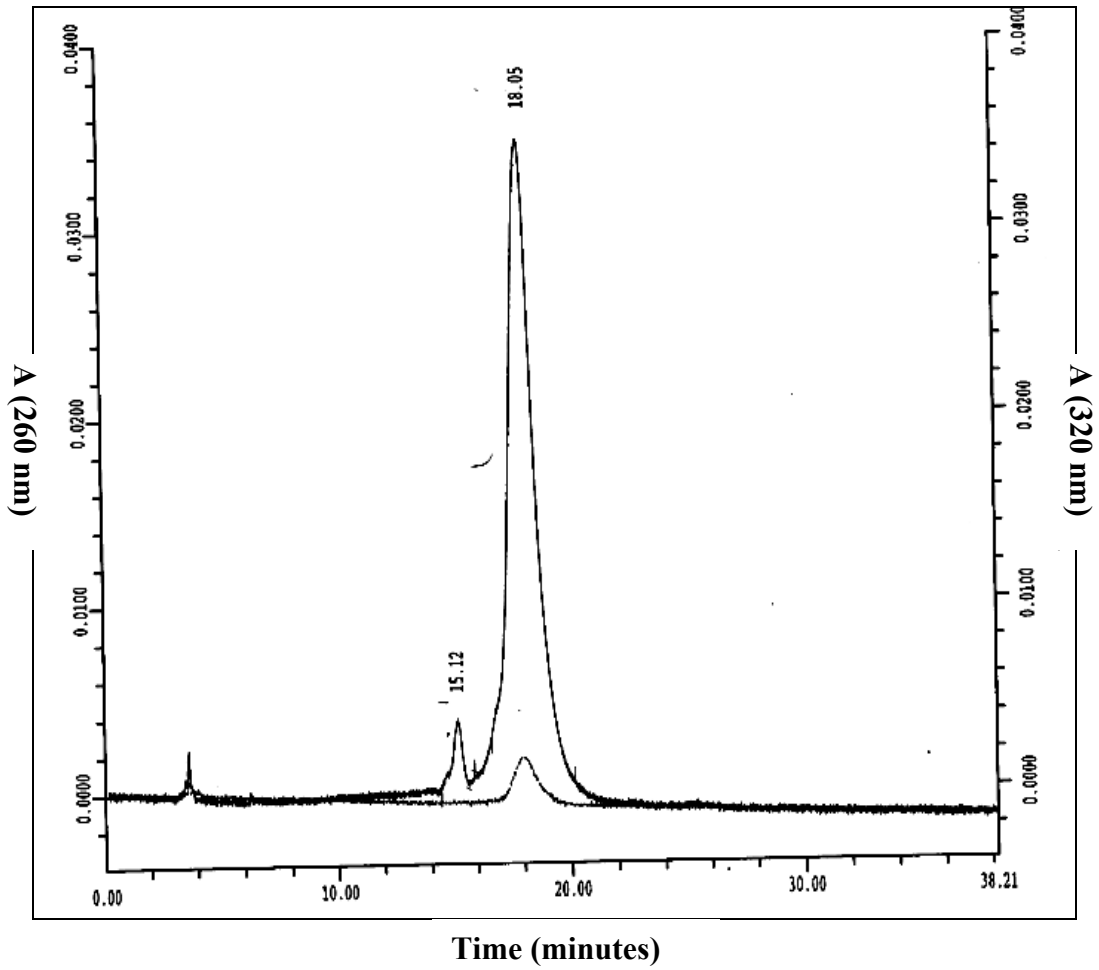


Figure 17 (A): Analytical reversed-phase HPLC control of purity for

C-I-5'-CTAGTG*G*(2,7-DAM) TATCC-3' oligonucleotide- (sample not heated). The alkylated oligo eluted at 18.05 minutes as can be seen from the absorption at both 260 and 320 nm monitored using the photodiode array system of the Beckman HPLC system. The structural analysis of the isolated alkylated oligonucleotide was determined by ESI (-)-MS and by analytical HPLC analysis of nucleosides using alkaline phosphatase and snake venom phosphodiesterase method as described in Materials and Methods.

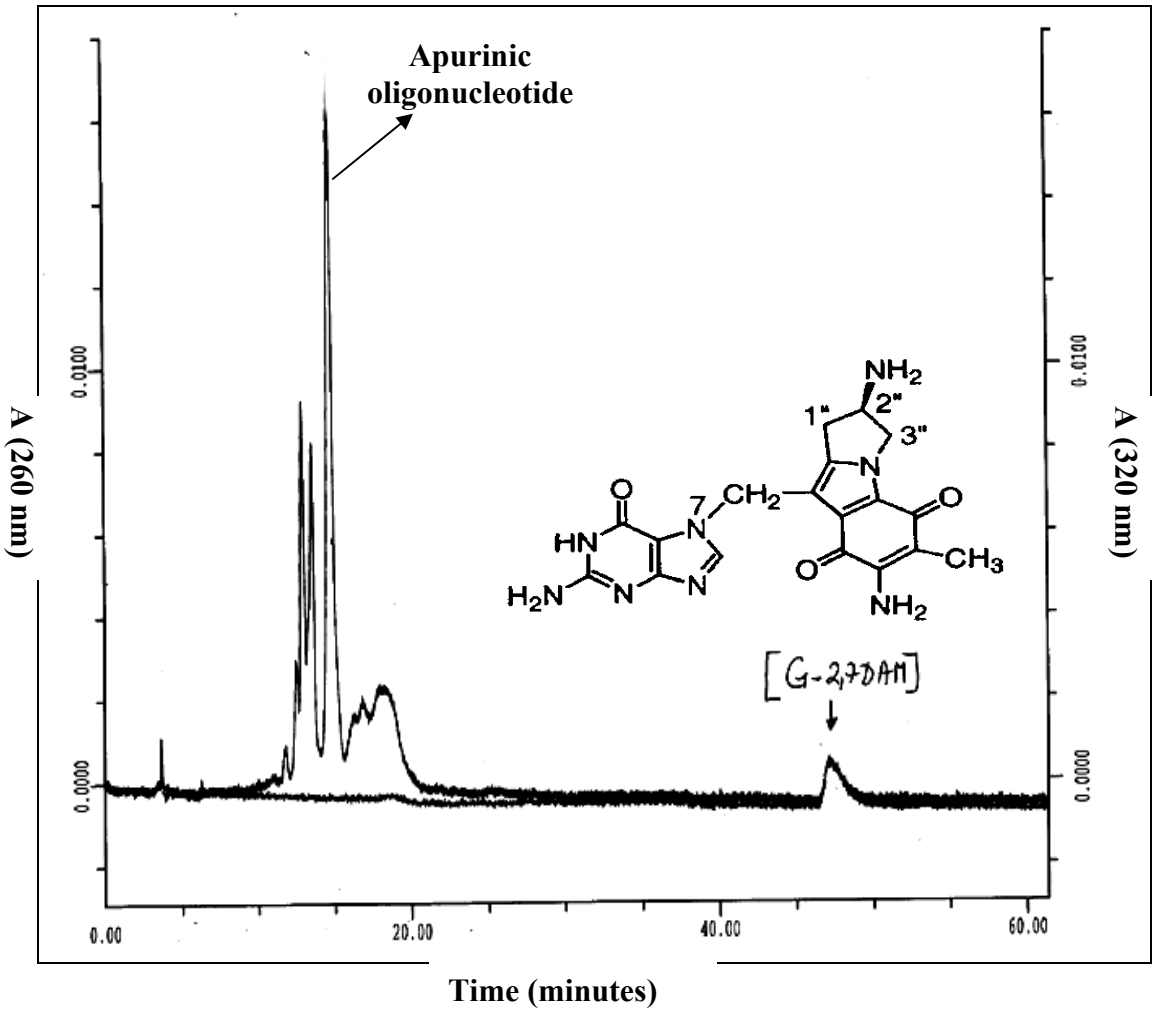


Figure 17 (B): Analytical reversed-phase HPLC control of purity for C-I-5'-CTAGTG*G*(2,7-DAM) TATCC-3' oligonucleotide- (sample heated). Through heating for 1 hour at 90°C the depurination of N7-dG-2,7-DAM oligonucleotide takes place releasing the apurinic oligonucleotide and the G*-N7-2,7-DAM as they are labeled on the HPLC chromatogram. This test shows that the alkylated oligonucleotide was pure and has the N7-G-2,7-DAM monoadduct.

In addition to the depurination of the alkylated oligonucleotide test the pure oligonucleotide was subjected to digestion by SVD and alkaline phosphatase and the N7-monoadduct was identified as described on section 2.3.2.1.

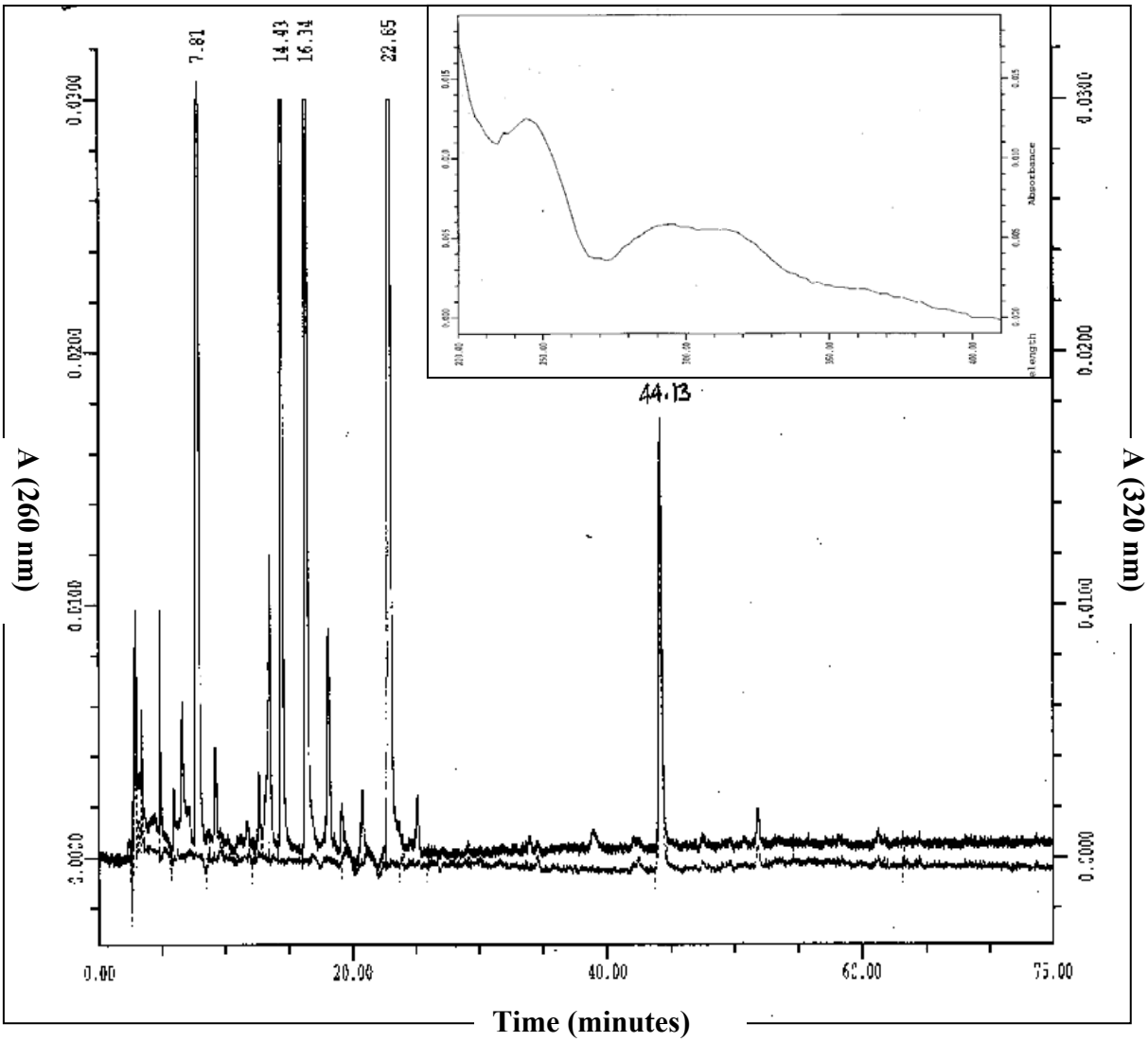


Figure 18: HPLC chromatogram of the SVD and alkaline phosphatase treated C-I-N7-2,7-DAM oligonucleotide. The HPLC chromatogram supports the fact that the oligo CI was alkylated at N7-position with 2,7-DAM by showing the N7-G-2,7-DAM monoadduct peak eluted at 44.13 min in the analytical analysis of the nucleoside pattern on a C18 reversed-phase HPLC. This peak was identified as being the N7-G-2,7-DAM monoadduct by comparison with the authentic standard (Appendix I, Figure 69). The inserted UV-spectrum (inset figure) is characteristic for the N7-2,7-DAM monoadduct as presented earlier (figure 12).

This new 12 mer C-I-5'-CTAGTG*G*(2,7-DAM) TATCC-3' was further used to obtain the 24 mer templates through ligation experiments which were further used for TLS with DNA polymerases.

The ESI-(-)-MS of this oligonucleotide proved that one single drug molecule was incorporated at N7-G position. Table 3 presents the molecular mass determination of C-I-N7-2,7-DAM oligonucleotide and the theoretically calculated mass.

Table 3: Mass of parent and of alkylated 5'-CTAGTG*G*(2,7-DAM)TATCC-3' oligonucleotides determined by ESIMS (-).

Oligonucleotide	Theoretically predicted MW (Da)	Experimentally determined MW (Da)
5'-CTAGTGGTACC	3636.5	3635.65
5'-CTAGTGG*(2,7-DAM)TACC (13)	3878.68	3879.68

2.3.2.3. Reductive conditions enhancing the production of N^2-dG-2,7-DAM monoadduct using the DNA duplex (II): 5'-CTAGTGGTATCC-3'-(CI)-3'-TCACCATAGG-5' (CIII). This section presents experimental evidences that support the view that increasing concentration of 2,7-DAM up to 80% vs. decarbamoyl 2,7-DAM is enhancing the production of N^2-dG-2,7-DAM. The reductive conditions used for alkylation of DNA dduplex (II) are described below:

Alkylating agent: 2,7-DAM: 80% 2,7-DAM and 20% decarbamoyl 2,7-DAM (as determined by HPLC).

Hydrogenation time: 25-45 minutes: the longer the hydrogenation time the higher the yield of the N^2-dG-2,7-DAM monoadduct.

Argon time: 60-90 minutes; after 1 hour the N^2-dG-2,7-DAM monoadduct is reaching a plateau .

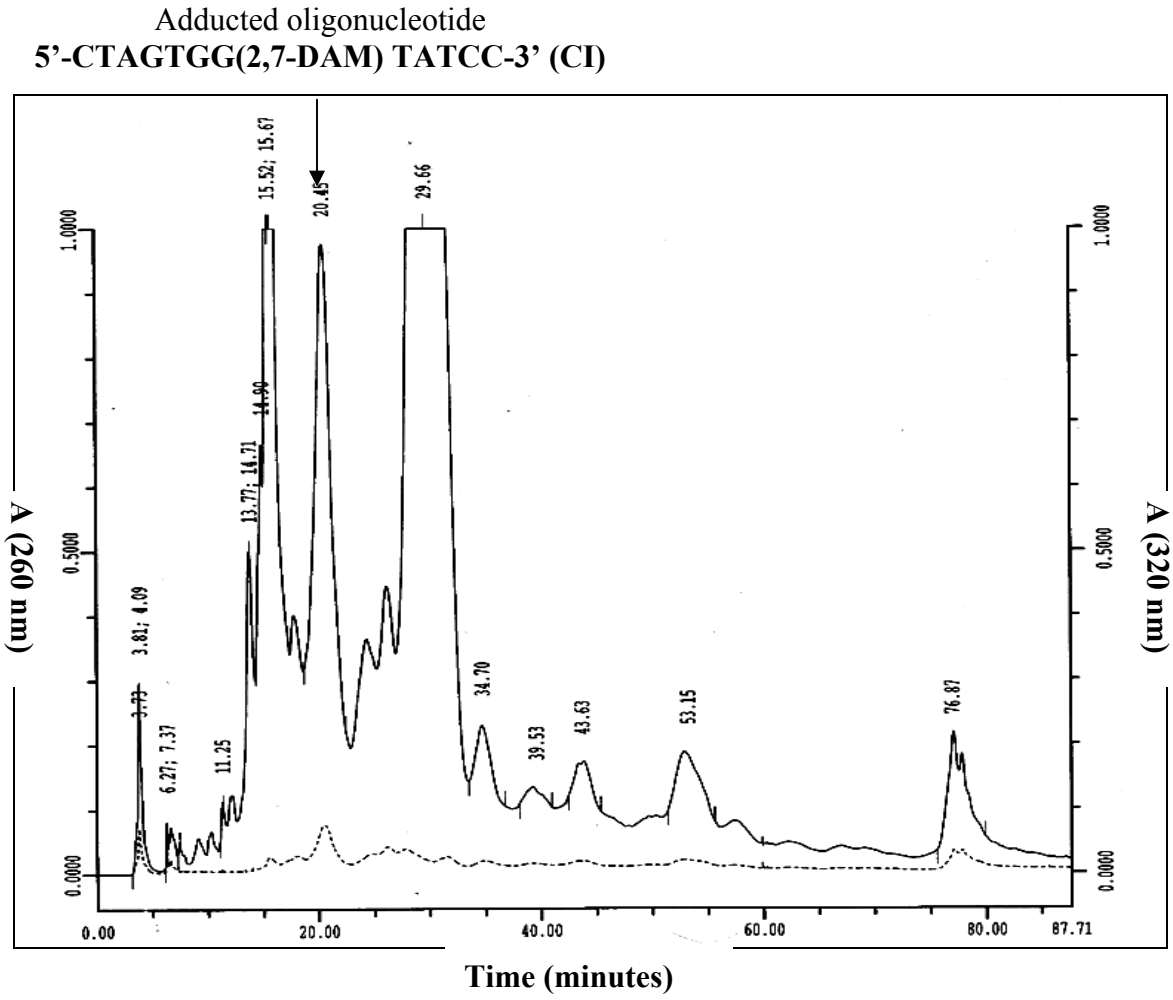
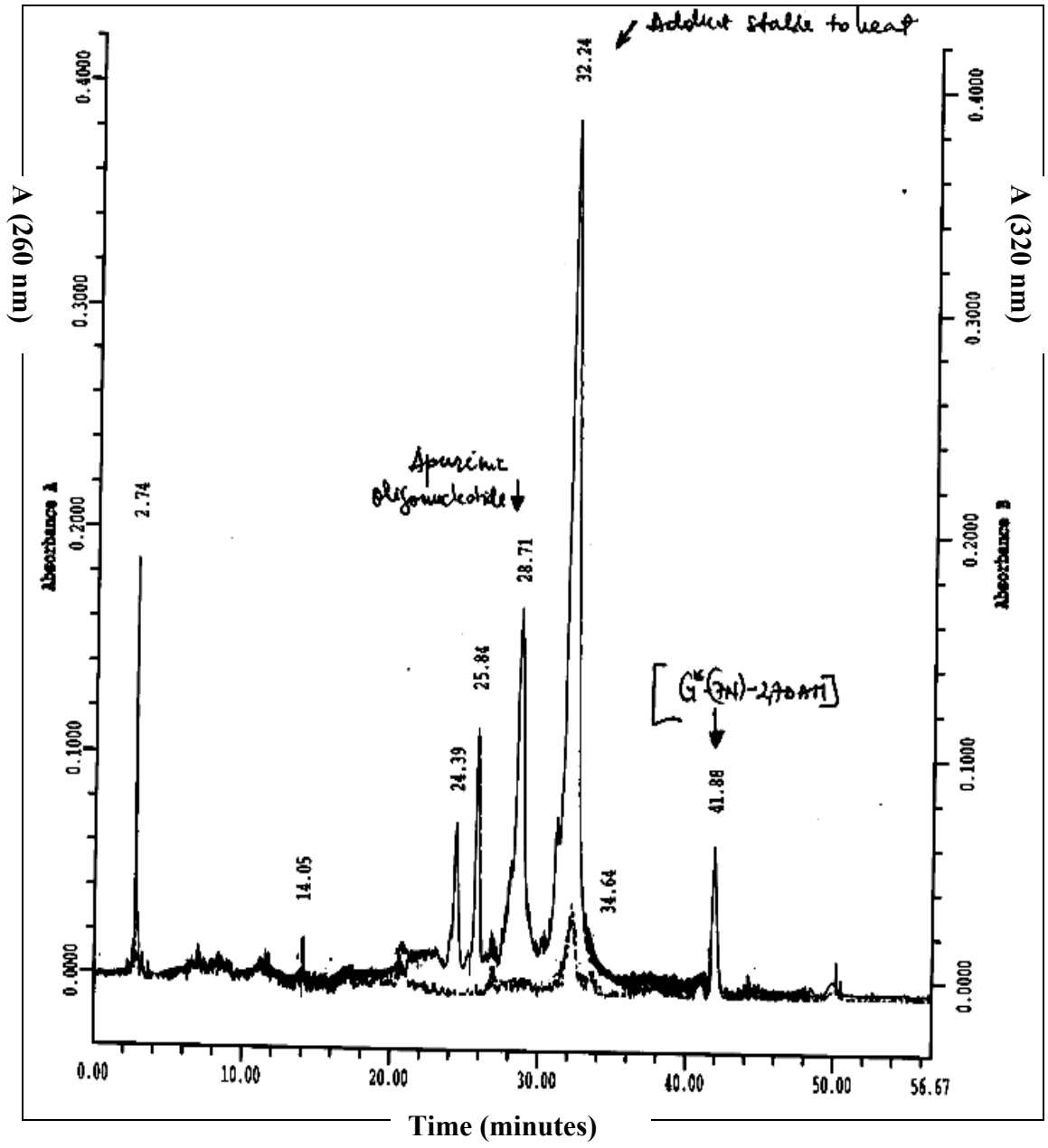


Figure 19. **Semipreparative reversed-phase HPLC of the DNA duplex (II) 5'-CTAGTGGTATCC-3'/(CI)-3'-TCACCATAGG-5' (CIII).** The alkylation with 2,7-DAM was performed under conditions which allowed the synthesis of N^2-dG-2,7-DAM monoadduct. Many adducted oligonucleotides are obtained but the major oligonucleotides adducted at N7 and N2-dG with 2,7-DAM are eluted at around 20 minutes. In most cases the N7-2,7-DAM and N2-2,7-DAM adducted oligonucleotides are not separated during this semipreparative run.

The alkylated oligonucleotides at 20.0 minutes were collected and further purified by HPLC. In order to separate just the N^2-2,7-DAM monoadducted **C-I-5'-CTAGTGGTATCC-3'**oligonucleotide the sample was heated, depurinated to release the

apurinic oligonucleotide, the G-N7-2,7-DAM and the N^2-dG-adducted oligonucleotide which is stable to heat (Figure 20-A-B).

A



B

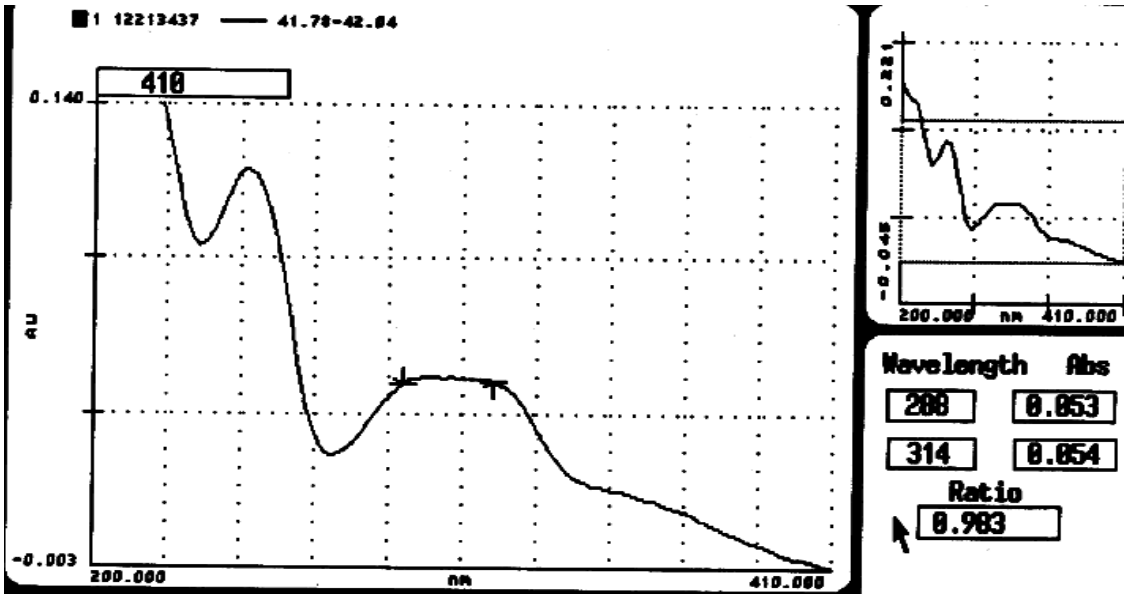
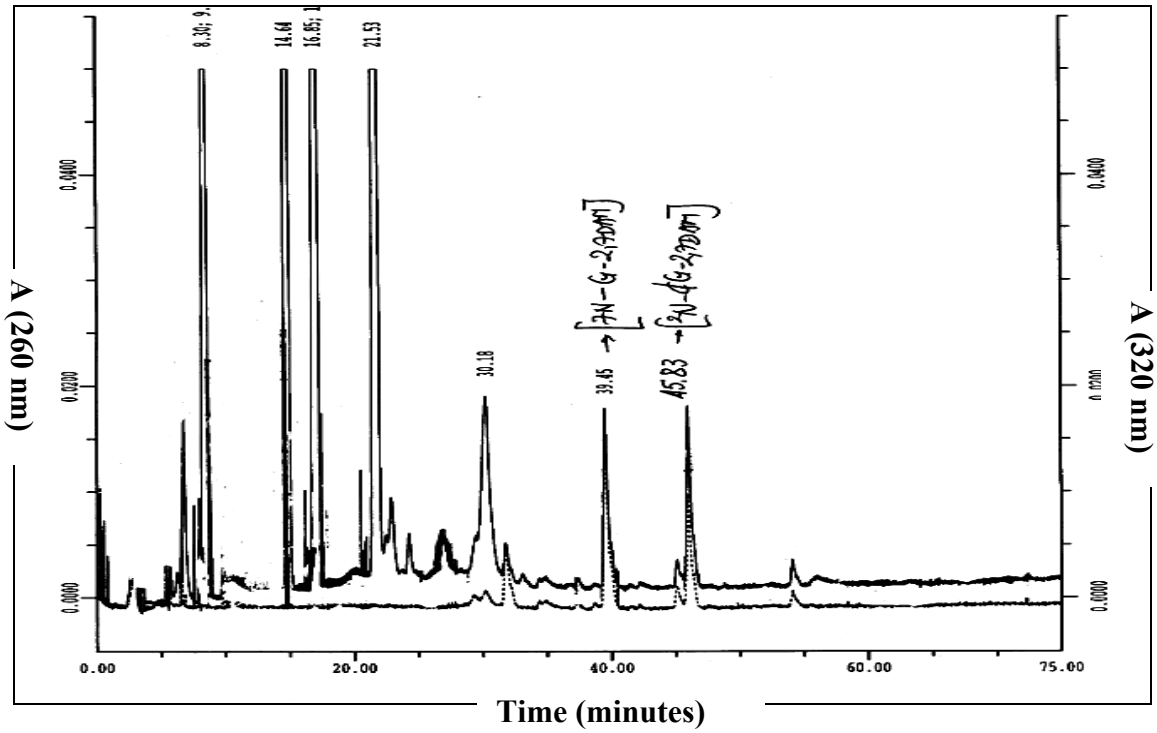


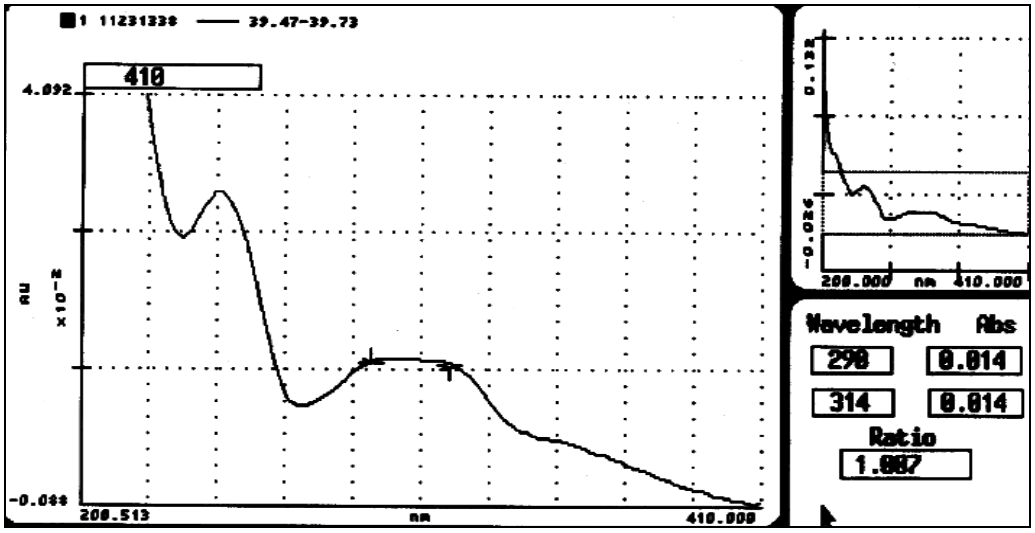
Figure 20. **(A-B) Reversed-phase HPLC of a mixture of N7 and N^2-dG-2,7-DAM monoadducted C-I-5'-CTAGTGGTATCC-3' (depurinated sample).** The G-N7-2,7-DAM monoadducted C-I-oligonucleotide is destroyed by depurination releasing the G-N7-2,7-DAM base and the apurinic oligonucleotides. The N^2-dG-alkylated C-I-oligonucleotide is stable to heat and can be further purified from this mixture. (B) UV-scan of the peak eluted at 41 minutes in Figure 19 (A) shows that the N7-G-2,7-DAM monoadduct was present in the original mixture of oligonucleotides isolated from semipreparative HPLC (Figure 19). The UV scan in the inset on the right side of figure 20 (B) shows that the ratio of absorption at 288 nm and 314 nmis almost 1, a characteristic for N7-G-2,7-DAM monadduct (11-15-Chapter-1).

In addition to the HPLC chromatogram of the depurinated oligonucleotide, the SVD and alkaline phosphatase treatment of purified adducted oligonucleotides presented in Figure 20-A showed that the original adducted oligo was a mixture of 2 different oligonucleotides, one containing the N7-2,7-DAM monoadduct and the other containing the N^2-dG-2,7-DAM monoadduct (Figure 21 (A-C)). In Figure 21 (B) and (C) the two UV-scan of the two known adducts of 2,7-DAM are shown the expected UV characteristics for the 2 monoadducts dG-2,7-DAM.

A



(B) **N7-G-2,7-DAM monoadduct-UV-scan**



(C) N^2-dG-2,7-DAM monoadduct-UV-scan

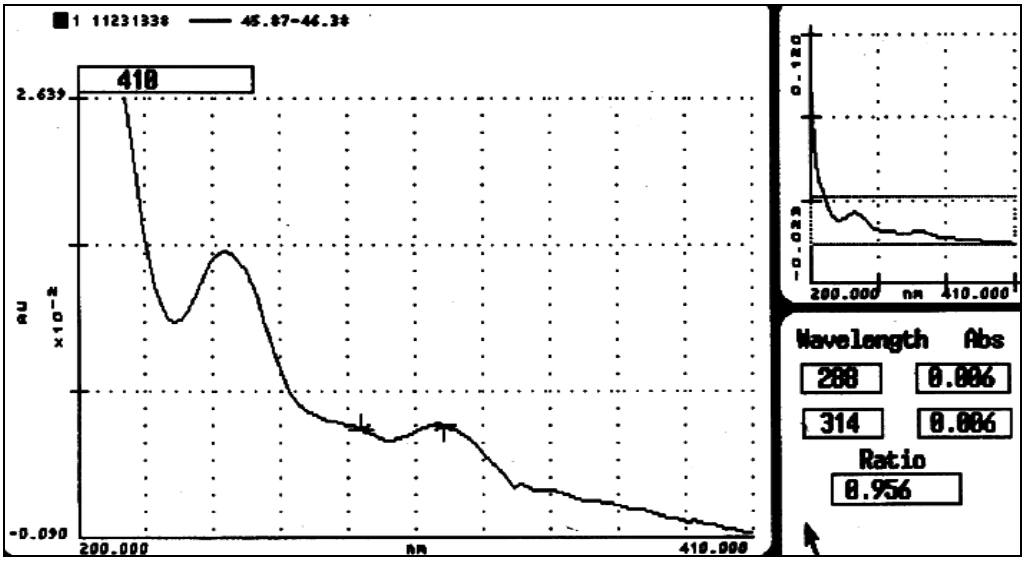
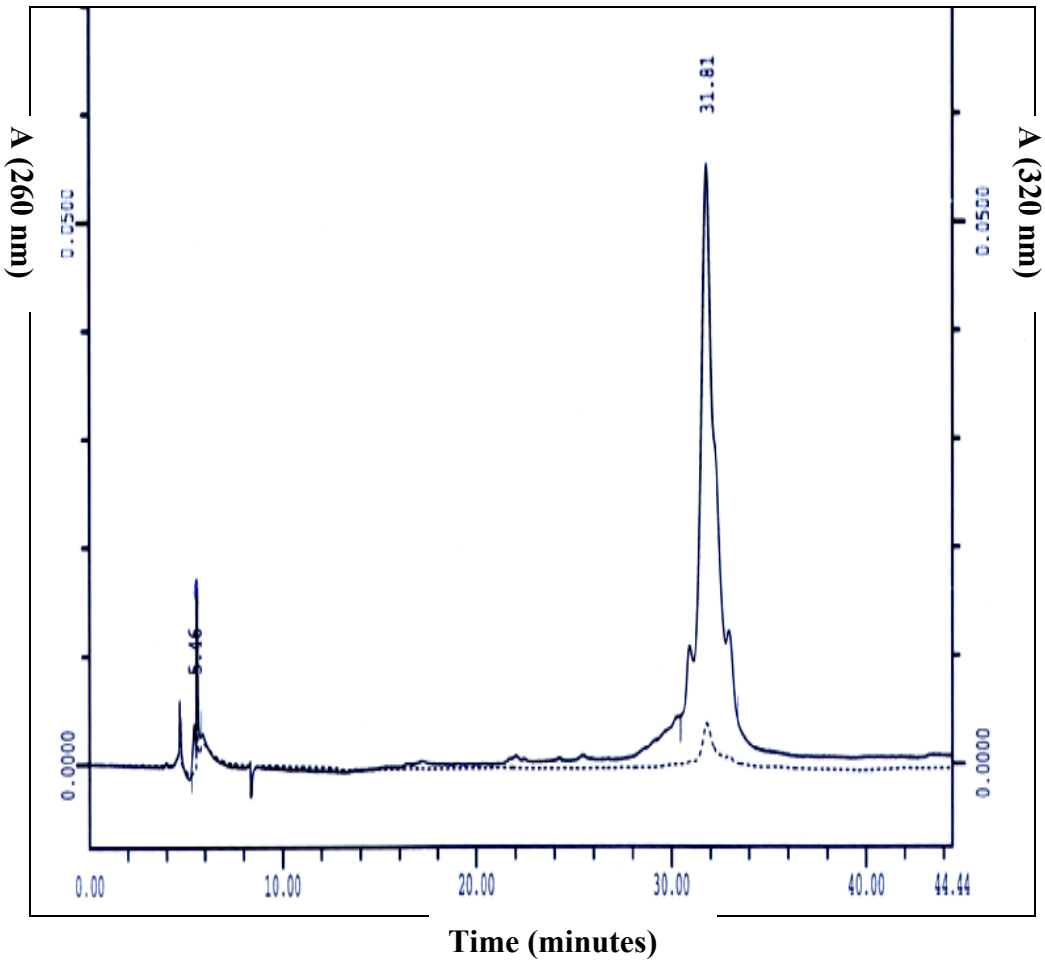


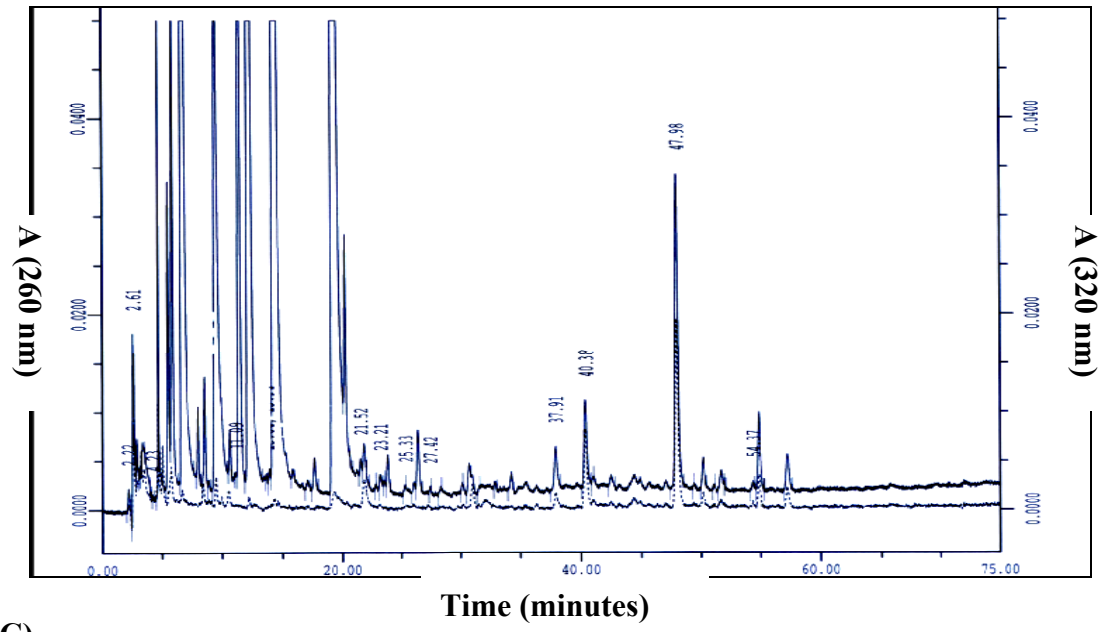
Figure 21 (A-C): The SVD and alkaline phosphatase treatment of the C-I-oligonucleotide-5'-CTAGTGGTATCC-3' containing the N7-G-2,7-DAM and N^2-dG-2,7-DAM monoadducts and UV scans of the isolated adducts. There are 6-7 minutes difference between the elution of N7-G-2,7-DAM base (39.45 min) and the elution of N2-dG-2,7-DAM nucleoside (45.83 min). The two monoadducts were assigned by running in the same conditions (i.e.using the same HPLC machine and the same bufferand gradient system) the authentic standards for N7-G-2,7-DAM and N2-dG-2,7-DAM. (B-C): UV-scans of N7-dG-2,7-DAM and N^2-dG-2,7-DAM monoadducts are showing the expected UV characteristics (12-15-Chapter 1).

The purified N^2-dG-2,7-DAM-C-(I) oligonucleotide was tested on C18 analytical reversed phase HPLC column and further subjected to SVD treatment for nucleoside analysis. The UV characteristics of N^2-dG-2,7-DAM-monoadduct were consistent with the published data and with authentic standards (figure 22 (A), (B) and (C)).

(A)


A (260 nm)
A (320 nm)
0.0500
0.0000
5.46
31.81
0.00
10.00
20.00
30.00
40.00
44.44
Time (minutes)

(B)



(C)

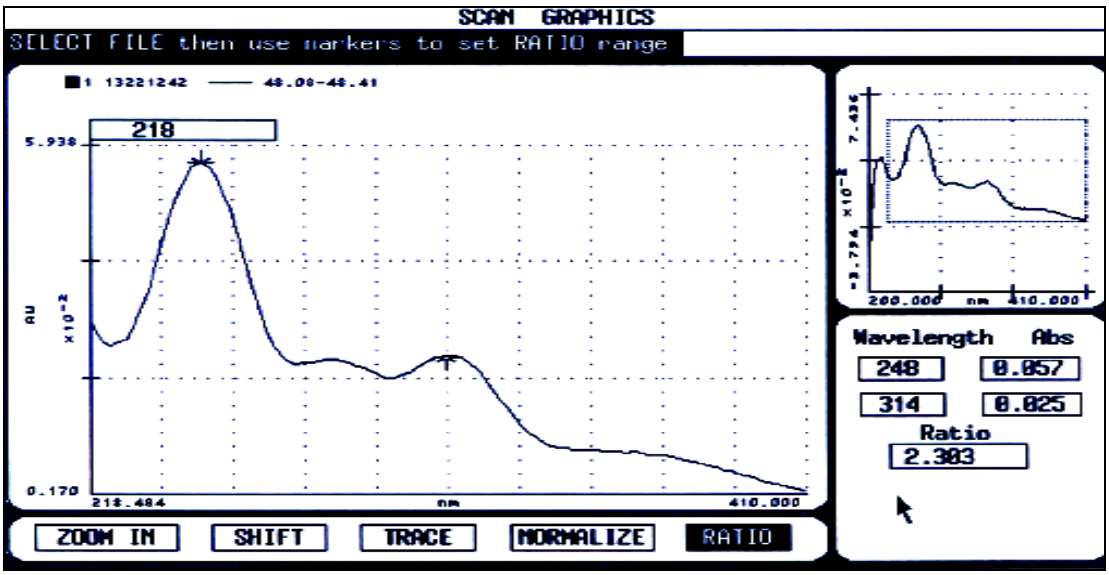


Figure 22 (A-C): Analytical reversed-phase HPLC analysis of purified N^2-dG-2,7-DAM adducted C-I-5'-CTAGTGGTATCC-3'- oligonucleotide (A) and SVD treatment of the purified C-I-5'-CTAGTGGTATCC-3'- oligonucleotide and nucleoside analysis on C18 column (B-C). The UV-scan of the adduct N^2-dG-2,7-DAM monoadduct (C). The N2-dG-2,7-DAM monoadduct eluted at 47 minutes in the reversed-phase system (the conditions for running the nucleosides are described in Materials and Methods).

The ESI-(-)-MS of this oligonucleotide proved that one single drug molecule was incorporated at N7-G position (Figure 23). Table 4 presents the molecular mass determination of C-I-dG-N^2-2,7-DAM oligonucleotide and the theoretically calculated mass.

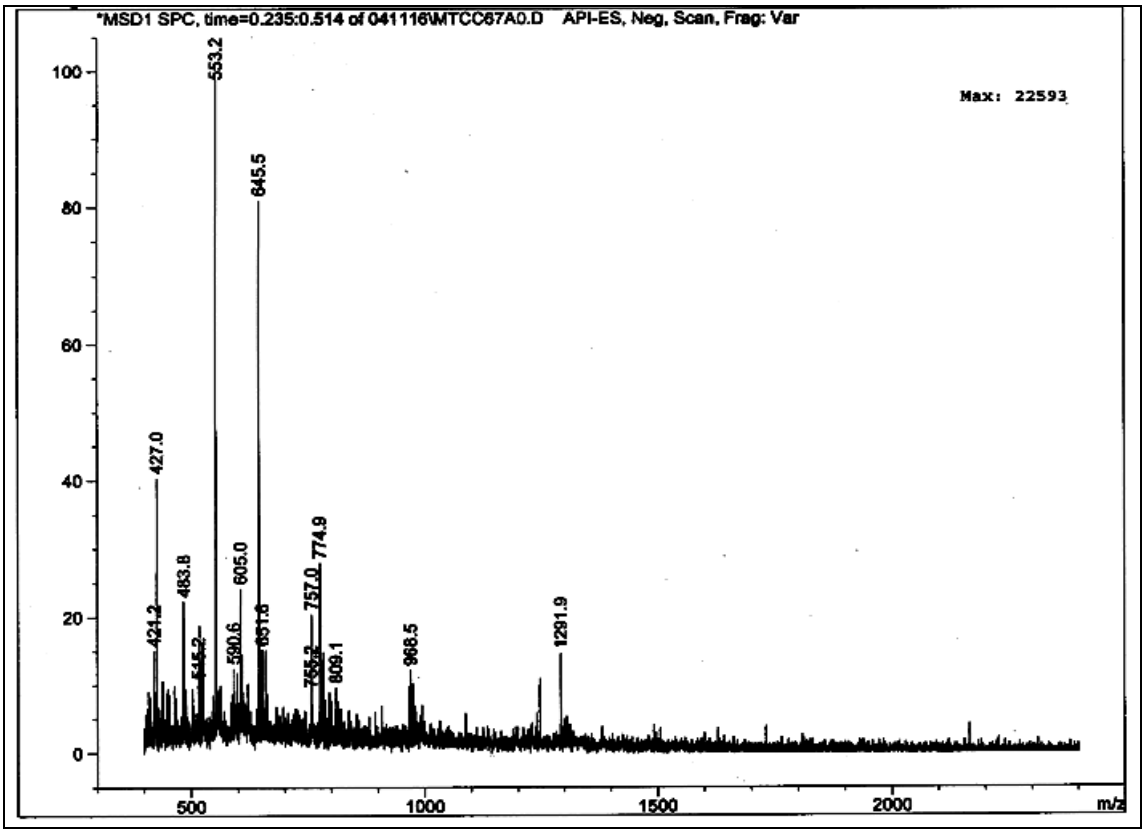


Figure 23: ESI-MS-(-) (negative mode) of purified C-I-5'-CTAGTGGTATCC-3'-oligonucleotide. The MS scan confirms that the oligonucleotide has one single 2,7-DAM drug molecule covalently bound to it (Table 4) (we assumed that the oligo is ionized with (-5) or (-6) charges corresponding to the m/z ratio of 645.5 and 553.2 respectively).

Table 4: Mass of parent and of alkylated 5'-CTAGTGG(2,7-DAM)TATCC-3' oligonucleotides determined by ESIMS (-).

Oligonucleotide	Theoretically predicted MW (Da)	Experimentally determined MW (Da)
5'-CTAGTGGTACC	**3636.5**	**3635.65**
5'-CTAGTGG*(2,7-DAM)TACC (13)	**3878.68**	**3879.68**

The progress of a reaction mix during alkylation of duplex (I) under conditions which favor the production of N^2-dG-2,7-DAM monoadduct (hydrogenation time: 45 minutes and Argon: 60 minutes) is presented in Figure 24 (A-D).

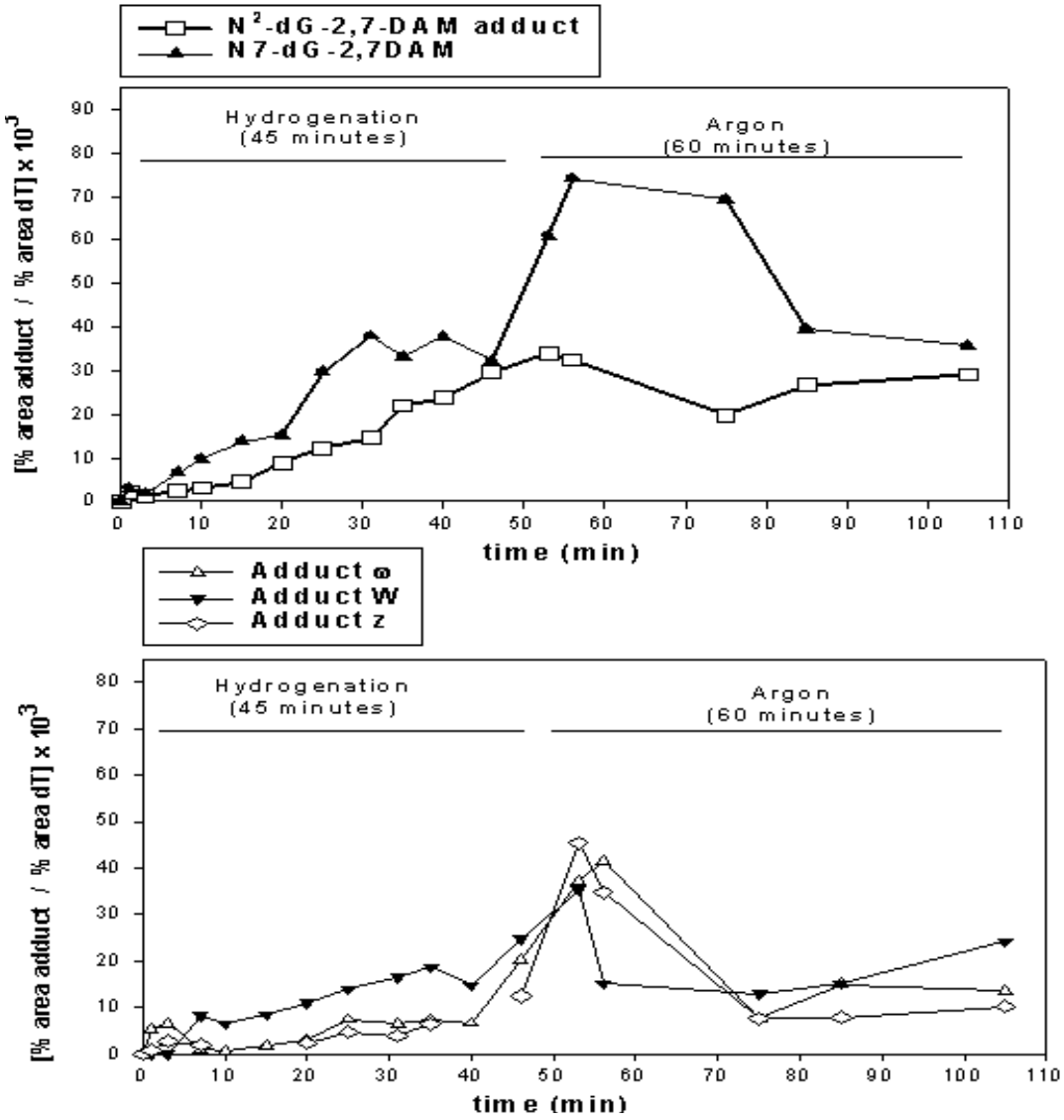
In Figure 24 **(A)** the ratio % area of N7-dG-monoadduct and % area of N^2-dG-monoadduct relative to % area of dT is presented as a function of time of alkylation reaction. In general the N7-G-2,7-DAM monoadduct is unstable supporting the hypothesis that it can be decomposed under anaerobic conditions into a reactive species of 2,7-DAM which can in turn activate the NH2 (N^2) position of guanine giving rise to N^2-dG-2,7-DAM monoadduct. However, an independent pathway for producing the N^2-dG-monoadduct cannot be excluded. In contrast, the N^2-dG-2,7-DAM monoadduct is more stable and is formed at a slower rate than N7-2,7-DAM monoadduct. Also other adducts labeled z, w, W are also formed in parallel with N^2-dG-2,7-DAM adduct, possibly competing with the intermediate reactive species which gives rise also to N^2-dG-2,7-DAM adduct (eventually they are guanine adducts-see Figure 24 (C) and their UV scans). We are still working on additional experiments to test the above hypothesis.

In Figure 24 **(B)** the SVD treatment of a reaction mix at two different data points under Argon is presented as an example for the assignment of the adducts N7-dG-2,7-DAM and N^2-dG-2,7-DAM. In all cases the reaction mix treated with SVD and alkaline

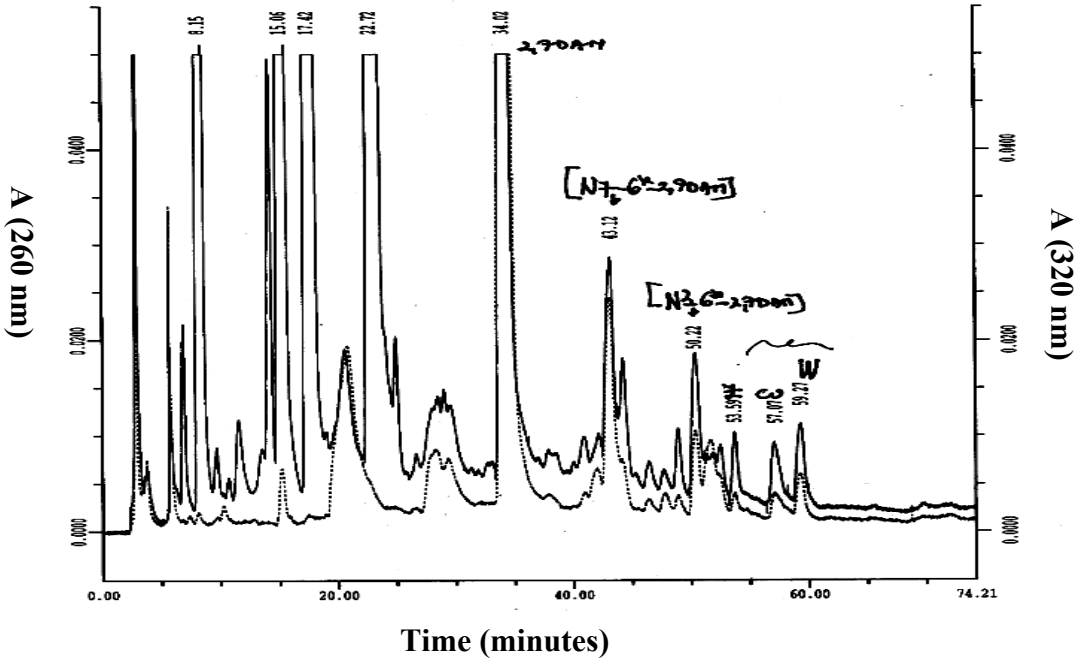
phosphatase is heated to fully release the N7-dG-2,7-DAM. The assignment of the adducts is performed based on authentic standards of N7 and N^2-dG-2,7-DAM and on their UV-characteristics (Figure 24 (C)).

In Figure 24 (D) the ratio % area N7-G*-2,7-DAM / %area N^2-dG-2,7-DAM is plotted as a function of reaction time. This graph suggests indirectly that N7-dG-2,7-DAM is converted into N2-dG-2,7-DAM but there are not direct proofs for this mechanism yet.

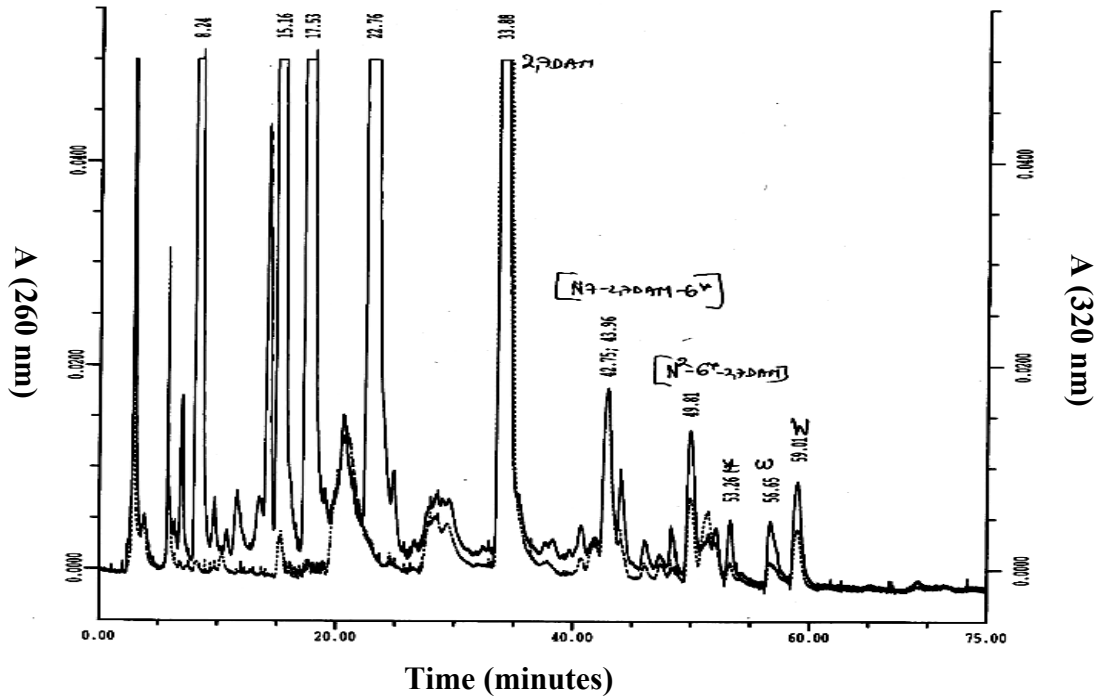
(A)



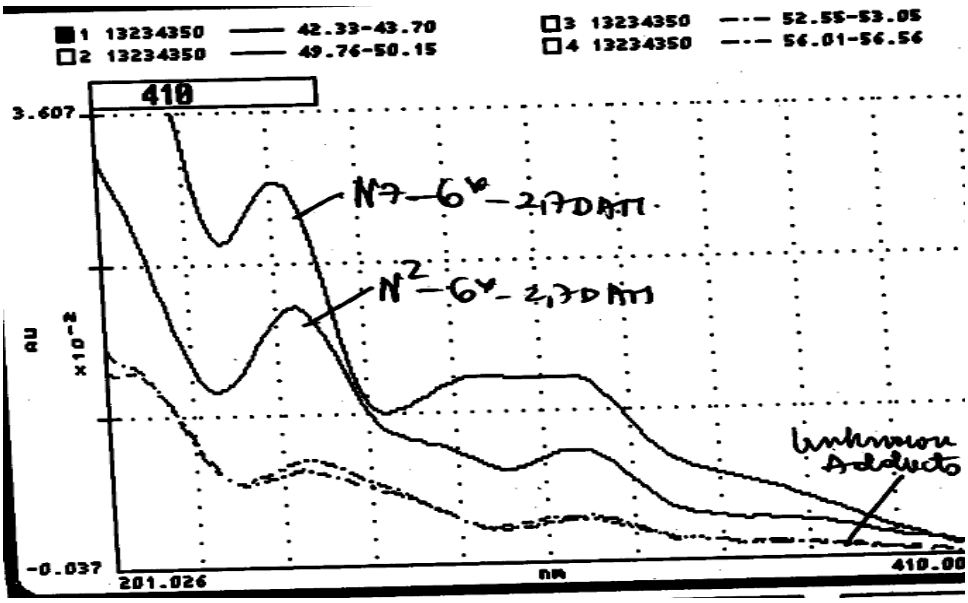
(B) 40 minutes Argon (after 45 minutes hydrogenation)



60 minutes Argon (after 45 minutes hydrogenation)



(C)



(D)

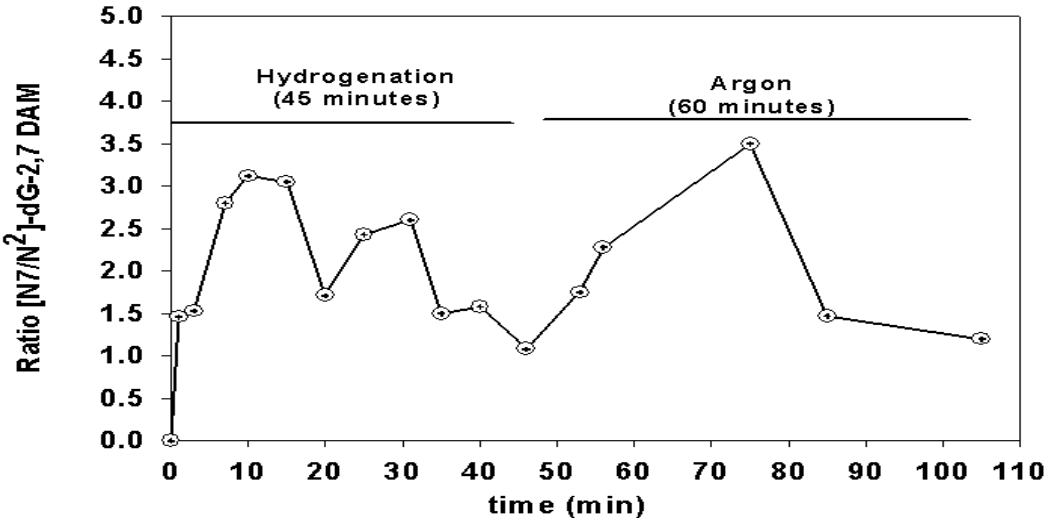


Figure 24 (A-D). Reductive conditions enhancing the production of N^2-dG-2,7-DAM monoadduct using the DNA duplex (I): C6 (5'-C T G GT A A T T T T A C-3' = 12 mer) and C7 (3'-G A C C A T T A A-5' = 9 mer). The reaction conditions together with the results which enhanced the production of the C6-5'-C T G G*T A A T T T T A C-3'-alkylated with 2,7-DAM at N^2-position of dG are presented in APPENDIX (II) (figure 72 (A-C)). The reductive conditions were the same as described in section 1.3.3.

2.3.3. Mapping the position of N7-2,7-DAM adduct within the C6- 5'-CTGG(4)(2,7-DAM)TAATTTAC-3' (construct 12-scheme 3) and within the C-I-5'-CTAGTGG(7)(2,7-DAM)TATCC-3' (construct 13-scheme-3) by Maxam Gilbert.

The positions of the 2,7-DAM-adduct 8 (see figure 1) in oligos **12-13** had to be independently determined, since these oligonucleotides contained more than one alkylatable target guanines, based on the known 5'…-GG-…3' sequence selectivity of formation of adduct **8** upon alkylation of DNA by 2,7-DAM (14-15-Chapter 1). The alkylation is selective in general for G tracts of DNA and for the 3'-G in -NGGN- sequences in duplex oligonucleotides; guanine in -NGN- sequences is essentially unreactive (12-15-Chapter 1). Accordingly, each of the oligonucleotide substrates for 2,7-DAM alkylation was designed to contain the same -TGGTA- target sequence. The oligonucleotide substrate for **13** contained an additional guanine in an -AGT- sequence, which was not expected to be alkylated. The positions of the adduct in **12** and **13** were determined experimentally by PAGE assay of their fragmentation upon treatment with hot piperidine (18) (figure 25).

Assay of the adducted oligonucleotide **12** showed that adduct **8** was located exclusively at nt position 4, as expected (Figure 25a). The same assay of **13**, which has three Gs, indicated that the 2,7-DAM adduct **8** was located in nt position 7 (Figure 25b).

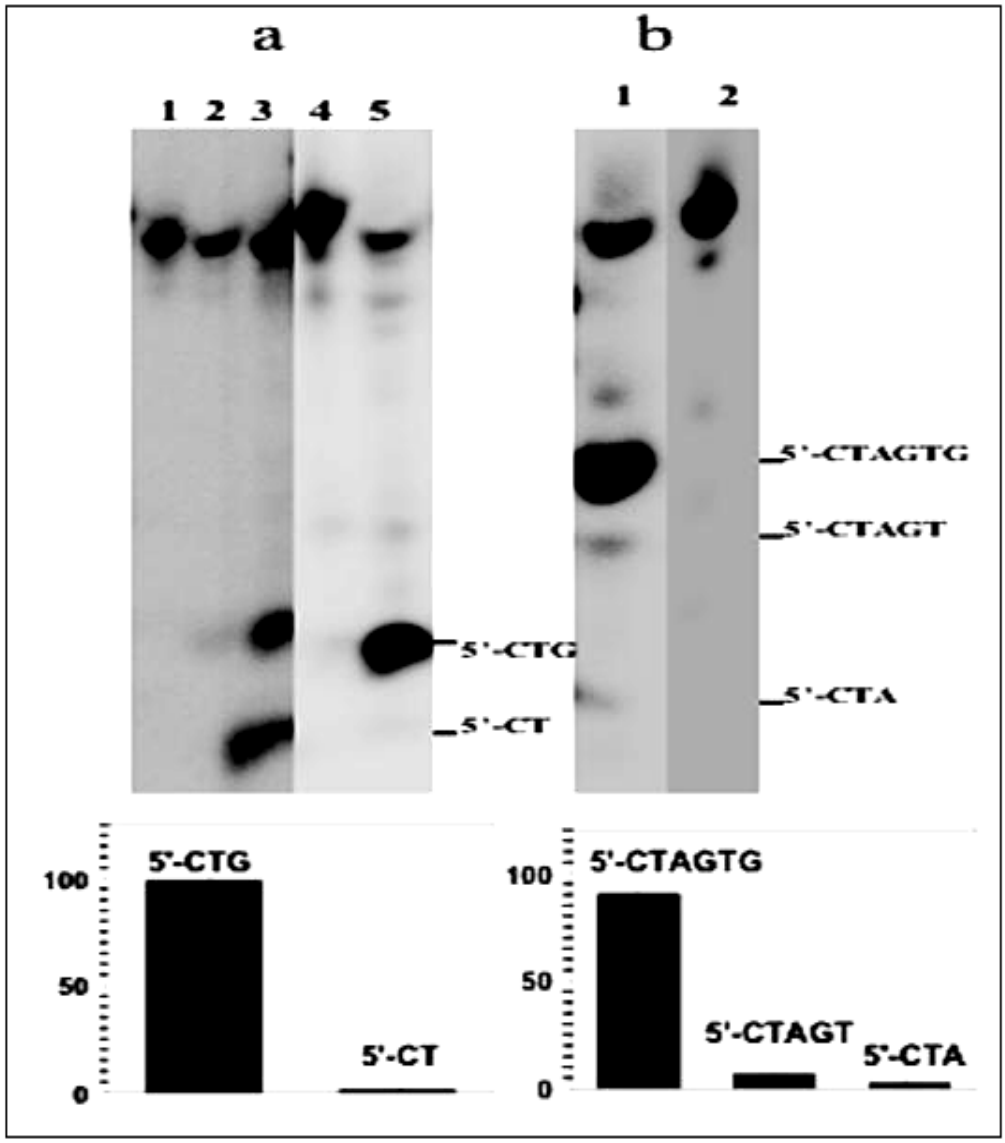


Figure 25: Piperidine cleavage assay of the 2,7-DAM adduct positions in the alkylated oligonucleotides 12 and 13 (Scheme 3).

(a) Assay of 12: PAGE (top), and histogram of % relative fragment intensities in lane 5 (below). PAGE lanes: 1, control; 2, control + piperidine; 3, control + dimethyl sulphate (DMS) + piperidine (*G-lane*); 4, 12; 5, 12 + piperidine.
(b) Assay of 13: PAGE (top), and fragment intensities (below). PAGE lanes: 1, 13 + piperidine; 2, 13.

2.3.4. Synthesis of 24 mer, 27 mer and 36 mer templates substrates for DNA polymerases.

The templates **14-16** were designed to contain the adduct located close to the 5' terminus in order to ensure uninhibited initial binding of the primer-template duplex substrate (**Scheme 4**) to the DNA polymerases. These templates were homogeneous on HPLC and PAGE and showed the expected composition by analysis of nucleosides using SVD treatment of the corresponding oligonuleotides (figures 31-34, Figures 75-78 Appendix IV). A schematical review of the major steps in the synthesis of each template construct is presented in **Scheme 5.**

Scheme 5: Schematical representation of the major steps used to obtain the templates for DNA polymerases.

(I) **Synthesis of 24-mer containing a MC guanine monoadduct**

Synthesis of the 24 mer template with a guanine adduct of MC

Step 1: Synthesis of a 9 mer containing a guanine adduct of MC:

5'-ACACGTCAT-3' **MC** **5'-ACACG*(MC)TCAT-3'**

reductive activation

(H_2, PtO_4)

Step 2: Ligation to make a longer template:

5'-ACACG*(MC)TCAT-3' + 5'-p-TAGAGATTGGTAGGG-3'
3'-TGTGCAGTAATCTCTAACCA-5' complementary strand for annealing)

↓

Product of ligation

5'-ACACG(MC)*TCATTAGAGATTGGTAGGG-3'
24 mer template (MC)

5'-TTTGTT-3-OH + 5'-p-ACACG*(MC)TCAT-3'-OH +
5'-pGGTTGGACGGCTGCGAGGCCG-3'
3'-AAACAATGTGCAGTACCAACCTGCCGACGCTCC-5' (complementary strand for annealing)

↓

Product of ligation:
5'-TTTGTTACACG*(MC)TCATGGTTGGACGGCTGCGAGGCCG-3'
36-mer template (MC)

(II)

Synthesis of the 27 mer template with a guanine 2,7-DAM adduct

Step1: Synthesis of a 12 mer containing one guanine adduct with 2,7-DAM:
5'- CTGGTAATTTAC-3' 2,7-DAM 5'- CTGG*(2,7-DAM)TAATTTAC3'

reductive activation
(H_2, PtO_4)

Step2: Ligation to make a longer template:

5'-CTGG*(2,7-DAM)TAATTTAC-3' + 5'-p-TAGAGATTGGTAGGG-3'
3'-CATTAAATGATCTCTAA-5' (complementary strand for annealing)

↓

Product of ligation:

5'-CTGG*(2,7-DAM)TAATTTACTAGAGATTGGTAGGG-3'
27 mer template (2,7-DAM)

Synthesis of the 24 mer template with a guanine 2,7-DAM adduct

Step1: Synthesis of a 12 mer containing one guanine adduct with 2,7-DAM:

5'-CTAGTGGTATCC-3' 2,7-DAM 5'- CTAGTGG (2,7-DAM)TATCC-3 '

reductive activation
(H_2, PtO_4)

Step2: Ligation to make a longer 24 mer template:

5'-CTAGTGG*(2,7-DAM) TATCC-3' + 5'-p-TAGAGATTGGTA-3'
3'-TCACCATAGGATCTCTAACC-5' (complementary strand for annealing)

↓

Product of ligation:

5'-CTAGTGG*(2,7-DAM) TATCCTAGAGATTGGTA-3'

24 mer template (2,7-DAM)

In order to test if the ligation reactions were successful aliquots from each reaction mix were run on high resolution sequencing gel (20% PAGE containing 8 M urea). The templates products of ligation reactions which contain the alkylated short

oligonucleotides (constructs 10, 12 and 13-Scheme 3) are having slower migration rate than the templates without the alkylated guanine-monoadducts, so that their presence in the reaction mix can be identified (Figures 26-28, and Figures 73 and 74 from APPENDIX III).

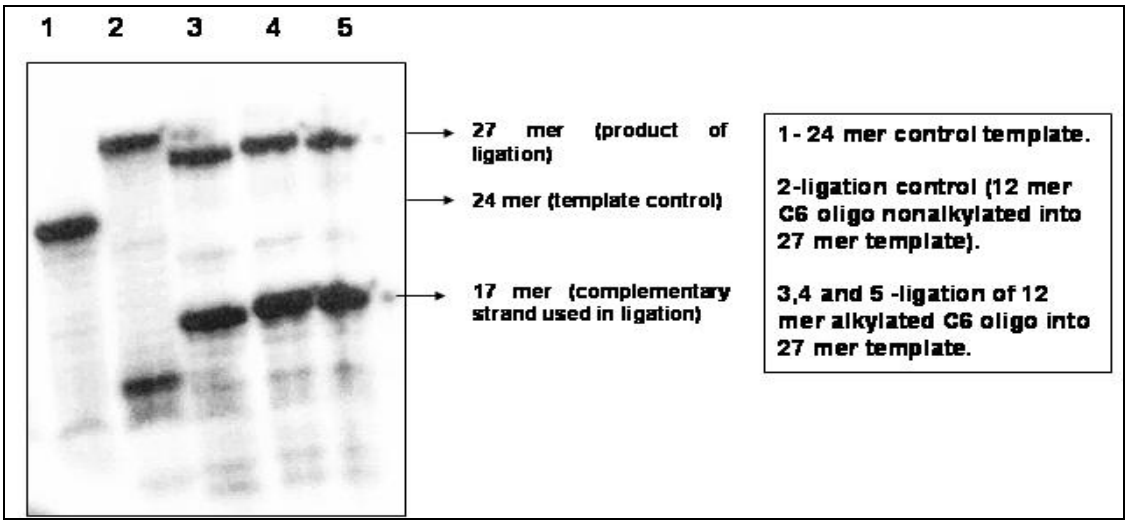


Figure 26: Analysis of ligation reactions involving the ligation of the 12 mer C6-2,7-DAM oligo into the 24 mer template on 8M high resolution PAGE sequencing gel. (the conditions for running the gel are described in Materials and Methods).

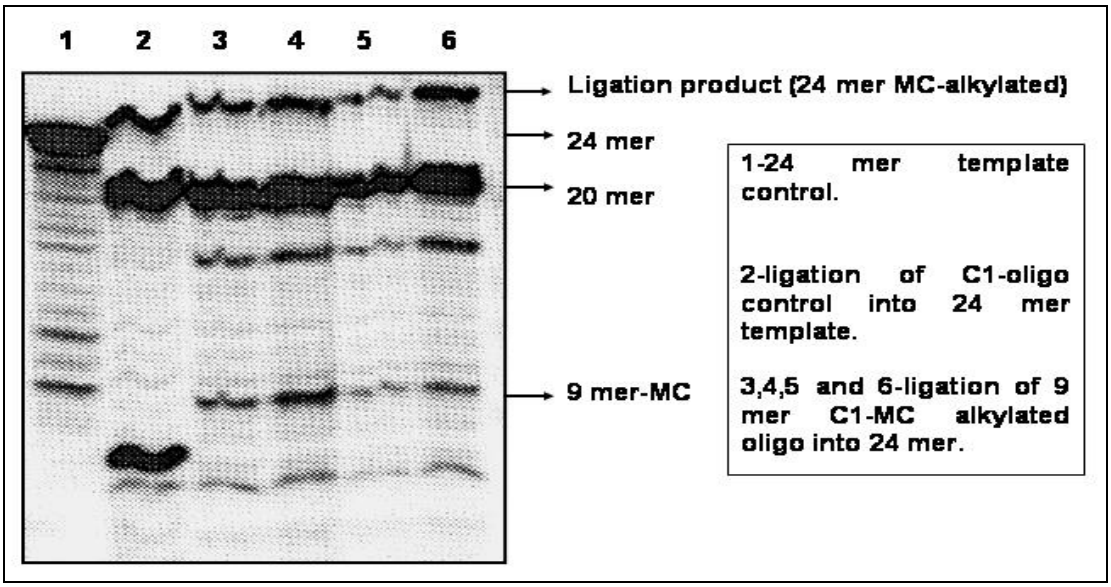


Figure 27: Analysis of ligation reactions involving the ligation of the 9 mer (MC) oligo (C1-MC) into the 24 mer template on 8M high resolution PAGE sequencing gel (the conditions for running the gel are described in Materials and Methods).

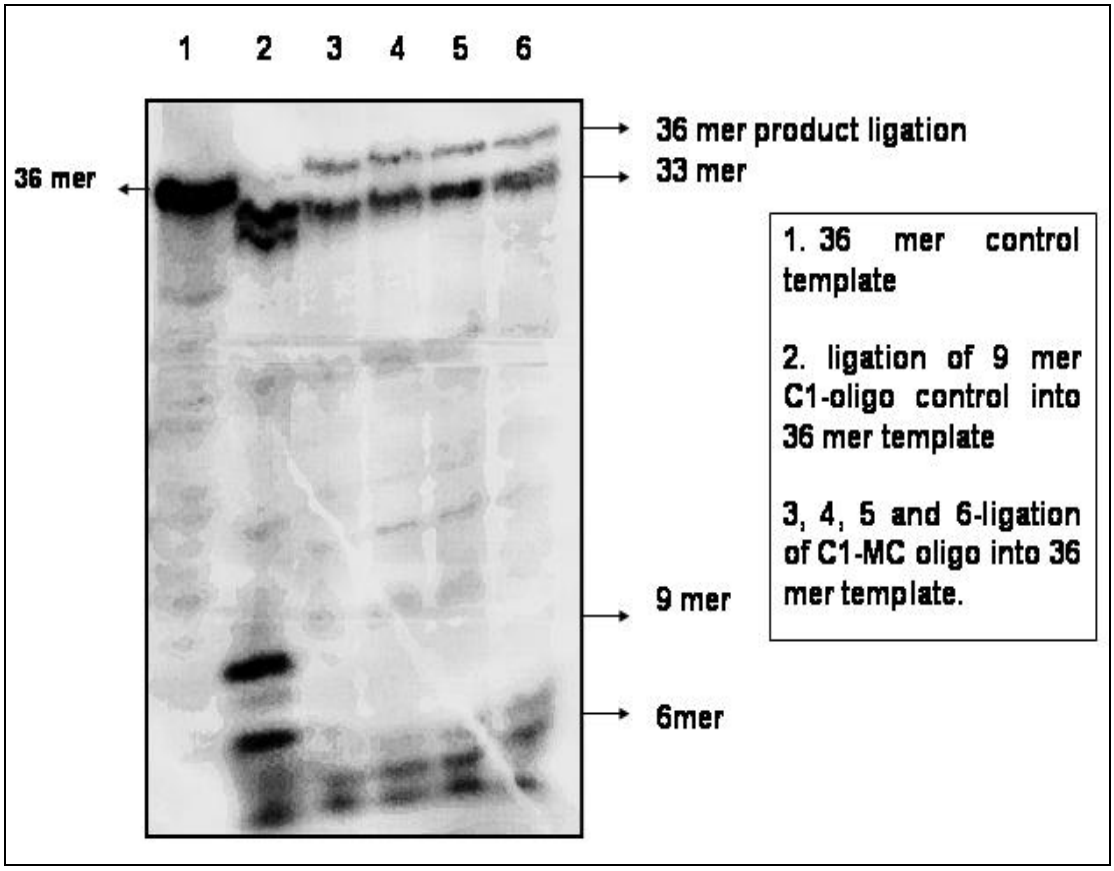


Figure 28: Analysis of ligation reactions involving the ligation of the 9 mer (MC) (C1-MC) oligo into 36 mer template on 8M high resolution PAGE sequencing gel (the conditions for running the gel are described in Materials and Methods).

The ligations experiments were performed successfully for the 27 mer construct containing one guanine alkylated with 2,7-DAM (see Figure 26), also for the 24 and 36 mer constructs containing one guanine alkylated with MC. The bands corresponding to the template 27, 24 or 36 mer alkylated at one guanine with MC or 2,7-DAM have the slowest electrophoretic mobility (as seen from the sequencing gels (Figure 26-28).

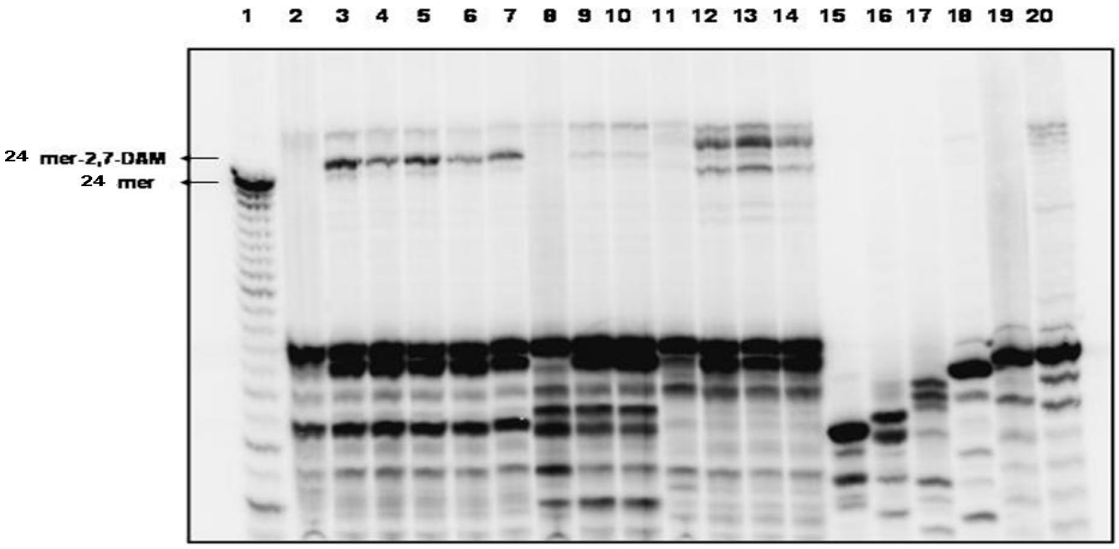


Figure 29: Analysis of ligation reactions involving the ligation of the 12 mer C-I-2,7-DAM oligo [5'-CTAGTG*G*(2,7-DAM) TATCC-3'] into the 24 mer template on 8M high resolution PAGE sequencing gel. (the conditions for running the gel are described in Materials and Methods).

1- 24 mer control template. **2-** ligation reaction mix for ligation of a 12 mer adduct alkylated at one single guanine with 2,7-DAM at (N7) position. **3-** ligation reaction mix for ligation of a 12 mer adduct alkylated at one single guanine with 2,7-DAM at (N7) position. **4-** ligation reaction mix for ligation of a 12 mer adduct alkylated at one single guanine with 2,7-DAM at (N7) position. **5-** ligation reaction mix for ligation of a 12 mer adduct alkylated at one single guanine with 2,7-DAM at (N7) position. **6-** ligation reaction mix for ligation of a 12 mer adduct alkylated at one single guanine with 2,7-DAM at (N7) position. **7-** ligation reaction mix for ligation of a 12 mer adduct alkylated at one single guanine with 2,7-DAM at ^{2}N position. **8-** ligation reaction mix for ligation of a 12 mer adduct alkylated at one single guanine with 2,7-DAM at ^{2}N position. **9-** ligation reaction mix for ligation a 12 mer adduct alkylated at two guanines with 2,7-DAM at (N7) position. **10-**ligation reaction mix for ligation a 12 mer adduct alkylated at two guanines with 2,7-DAM at (N7) position. **11-** ligation reaction mix for ligation a 12 mer adduct alkylated at two guanines with 2,7-DAM at (N7) position. **12-**pool fractions of ligation mix of a 12 mer adduct alkylated at one single guanine with 2,7-DAM at (N7) position (for preparative work). **13-** pool fractions of ligation mix of a 12 mer adduct alkylated at one single guanine with 2,7-DAM at (^{2}N) position (for preparative work). **14-**pool fractions of ligation mix of a 12 mer adduct alkylated at two guanines with 2,7-DAM at (N7) position. **15-** purity control for the 12 mer adduct alkylated at one single guanine with 2,7-DAM at (N7) position. **16-** purity control for the 12 mer adduct alkylated at one single guanine with 2,7-DAM at (^{2}N) position. **17-** purity control for the 12 mer adduct alkylated at two guanines with 2,7-DAM at (N7) position. **18-** 15 mer used for ligation experiments (1). **19-**15 mer used for ligation experiments (2). **20-**17 mer used as complementary strand during ligation experiments.

The results show that the ligation experiments were successful for the 24 mer construct containing one or two guanines alkylated with 2,7-DAM (as is seen on the sequencing gel (Figure 28) the shifted band above the 24 mer control template correspond to the alkylated 24-mer template). The alkylated template has a slower electrophoretic mobility than the non-alkylated control template. It can be seen that part of the new alkylated templates are depurinated, generating an independent band with the same electrophoretic mobility as the control template (Figure 28-lanes 12-14).

2.3.5. Purification and structural characterization of the alkylated 24 mer and 36 mers templates alkylated with MC and 2,7-DAM.

2.3.5.1. Purification of templates from PAGE-urea gels.

The ligation mix was used to purify the alkylated templates (24, 27 mer and 36 mers) from the other oligonucleotides used for ligation using preparative 16% polyacrylamide-8M urea gel. The 24 or 27 mer and the 36 mer templates were excised from the gel and desalted using a Sep-Pak cartridge (Waters) and further purified as described in Materials and Methods (Specific Methods section). The preparative urea gel used for the purification of DNA templates contained as marker lanes ^{32}P-labeled ligations mix such that after autoradiography the positions of the required DNA bands could be determined.

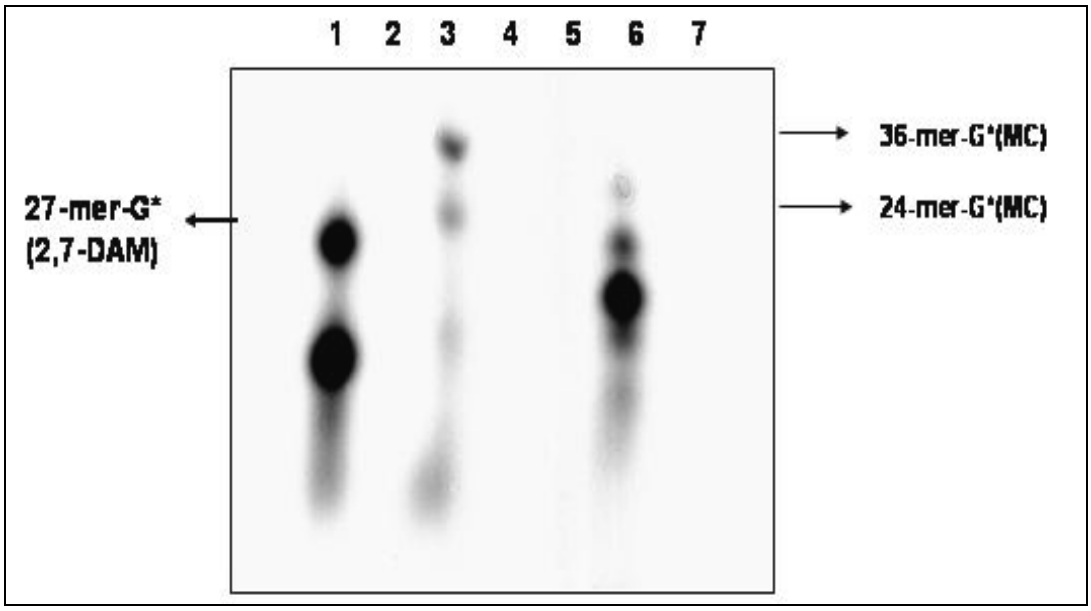


Figure 30: 16% preparative polyacrylamide gel containing 8 M urea for purification of 24 mer-G*(MC), 24 mer-G*(2,7-DAM) and 36 mer-G*(MC).

1. ^{32}P-labeled reaction mix for ligation of C6-2,7-DAM into 27 mer template
2. reaction mix for ligation of C6-2,7-DAM into 27 mer template (for DNA excision from gel)
3. ^{32}P-labeled reaction mix for ligation of C1-MC into 36 mer template
4-5- reaction mix for ligation of C1-MC into 36 mer template template (for DNA excision from gel)
6- ^{32}P-labeled reaction mix for ligation of C1-MC into 24 mer template
7- reaction mix for ligation of C1-MC into 24 mer template (for DNA excision from gel)

2.3.5.2. Structural characterization of the alkylated 24, 27 mer and 36 mer alkylated mitomycin C or with 2,7-DAM.

Before the alkylated templates were further used in the primer extension assays, the structural characterization of the purified DNA templates was performed to determine if the final product of the ligation was having a single guanine alkylated with MC or with 2,7-DAM. The structural characterization of the alkylated templates involved the digestion of the DNA with snake venom phosphodiesterase and alkaline phosphatase, both for the adducts with MC and with 2,7-DAM, followed by the analysis of the

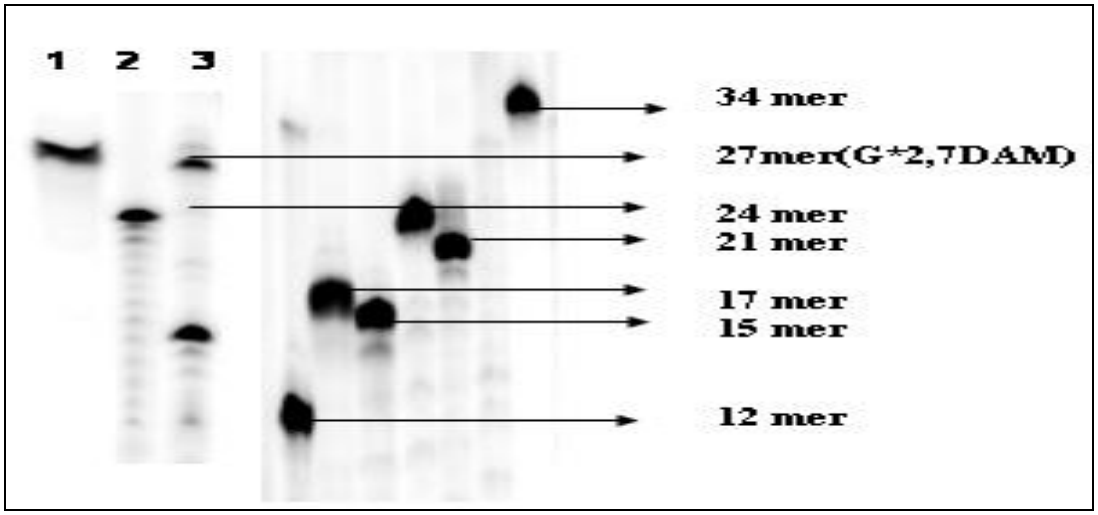
nucleosides using the reverse-phase HPLC (C18 column, 100 Å) and the buffer system as described in the Specfic Methods section and in sections 1.1-1.2 (Figures 75-78, in Appendix IV presents the HPLCs for the digested alkylated templates, as a proof for their structural identity: the alkylated MC or 2,7-DAM nucleoside G is eluted at later retention time than the nonalkylated nucleosides and were identified to be the expected monoadducts based on HPLC analysis of authentic standards and on UV characteristics already described in the literature (11-15-Chapter 1).

2.3.5.3. High resolution sequencing PAGE electrophoresis and reversed-phase HPLC proofs of purity for the MC and 2,7-DAM 24, 27 mer and 36 mer alkylated mitomycin C or with 2,7-DAM.

The purity of adducted templates **14-16 (**Scheme 3) was tested by running ^{32}P-labeled aliquots of each template on a 18% sequencing gel, at 2800 V constant voltage for 3 hours and half. The gel was exposed and processed using a 445 SI PhosphorImager and ImageQuant 5.2 software (as described in detsails in Materials and Methods).

As can be seen from Figures 31-34 and Figures 79-84 in APPENDIX V all the templates are homogeneous pure. In the case of 2,7-DAM alkylated templates we were interested to check if any depurination was taking place, since is known from the literature that the apurinic sites are easily bypassed with incorporation of dATP opposite the apurinic site (34-50). The PAGE test showed that the 2,7-DAM-alkylated template was not depurinated (APPENDIX V-Figure 84).

A



B

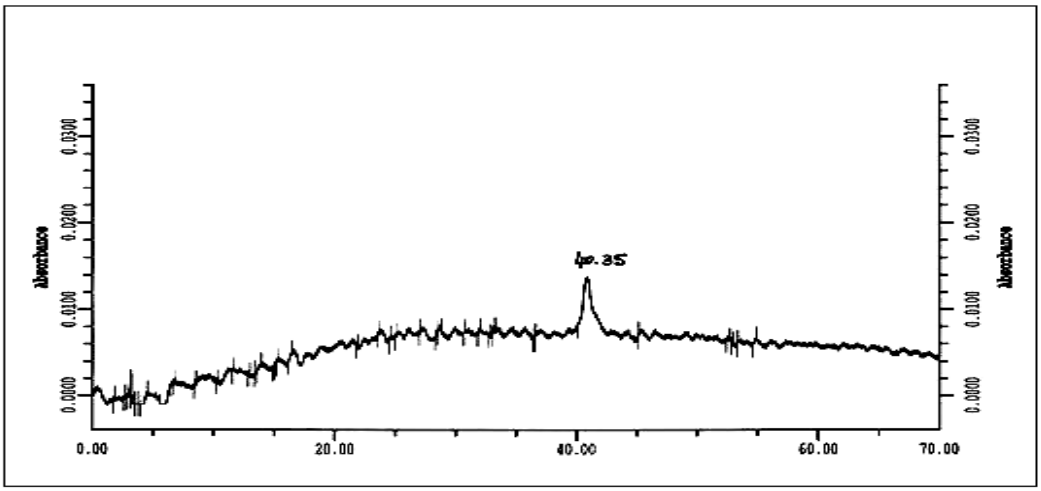
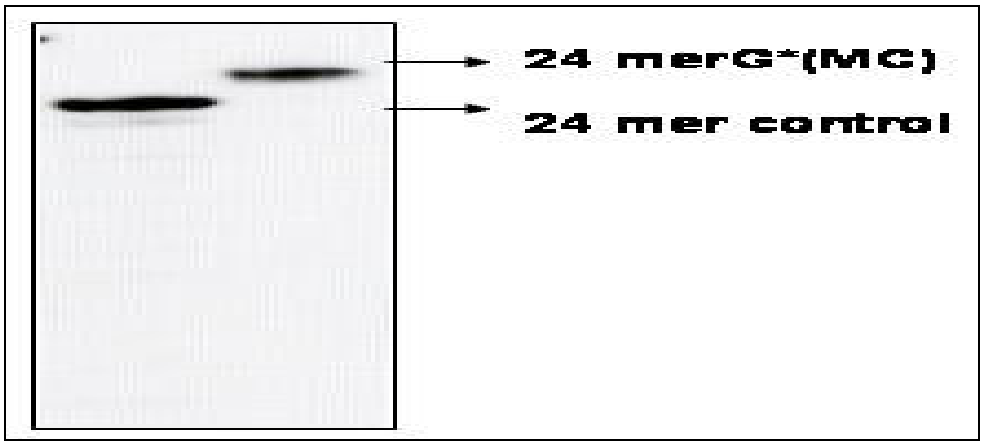


Figure 31. A**.** 18% Sequencing gel analysis of ligation reactions for obtaining the 27-mer-G*-(2,7-DAM) template. 1- **27 mer-G*(2,7-DAM) template purified**. 2- **24 mer template control.** 3- **ligation mix for 27 mer-G*-(2,7-DAM).**

B. Reverse-phase HPLC analysis of the 27-mer-G*-(2,7-DAM) template on a C4 column (300 Å) as described in "Materials and Methods".

A.



B.

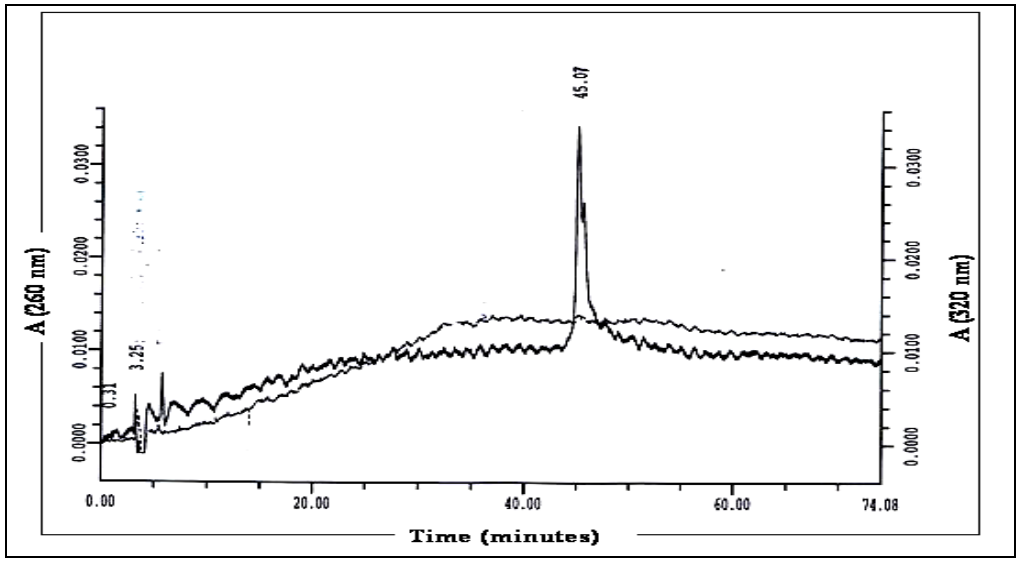
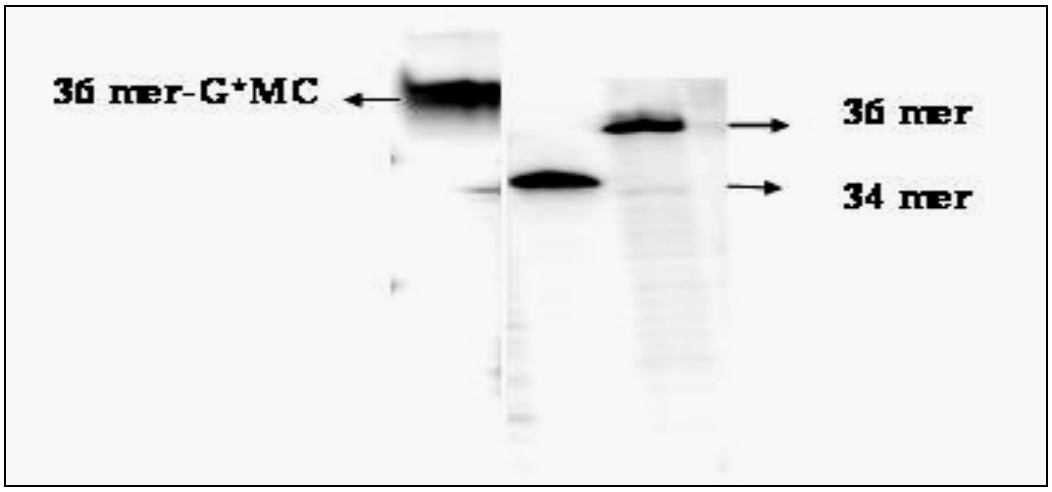


Figure 32: PAGE (A) and HPLC (B) analysis of alkylated 24-mer-MC template. A. 18% Sequencing gel analysis of ligation reactions for obtaining the 24-mer-G*-(MC) template. 1- 24 mer template control. 2-24-mer-G*-(MC) template purified. **B. Reverse-phase HPLC analysis of the 24-mer-G*-(MC) template on a C4 column (300 Å) as described** in "Materials and Methods".

A.



B.

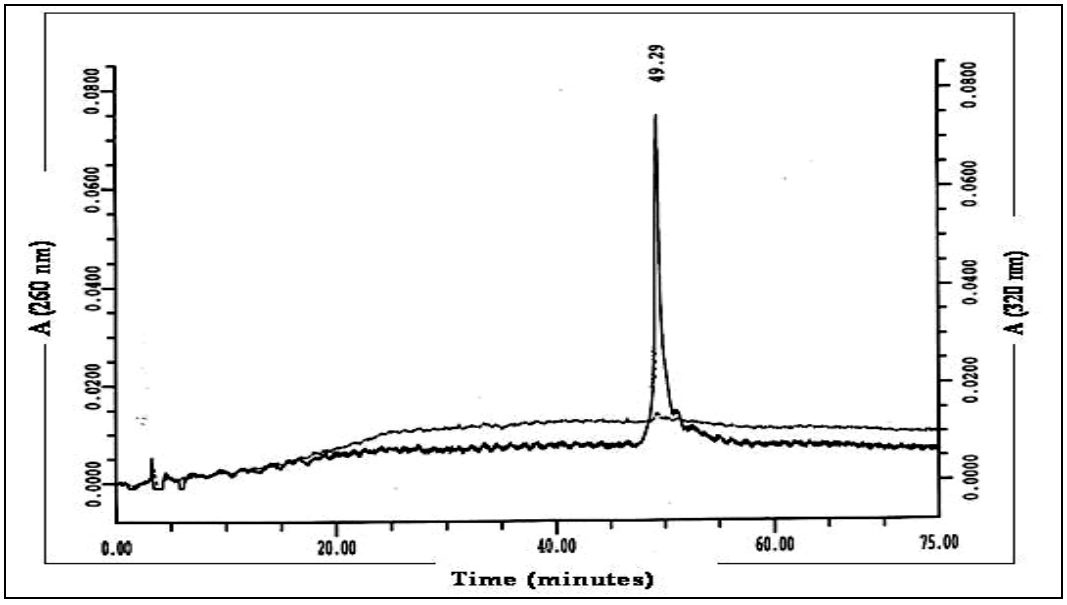


Figure 33: **PAGE (A) and HPLC (B) analysis of alkylated 36-mer-MC template.** A. **18% polyacrylamide high resolution sequencing gel for purity check of 36 mer-G*(MC) template.** B. **Reverse-phase HPLC analysis of 36-mer-G*-(MC) template on a C4 column (300 angstroms)** (details in "Materials and Methods").

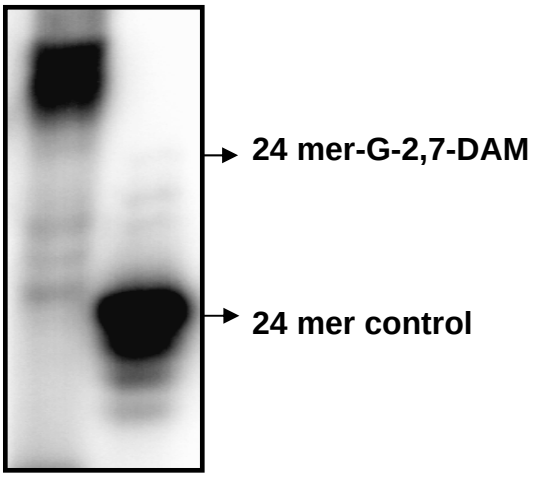


Figure 34: PAGE (A) analysis of 24-mer-2,7-DAM alkylated template (construct 16 from schene 3).

2.3.6. In vitro primer extension and single-nucleotide kinetics assay using Klenow exo-, T7 exo-, eta and Klenow exo+ DNApolymerases.

The *in vitro* polymerase and single-nucleotide kinetics were performed with four different polymerases Klenow fragment (KF) (exo-), T7 (exo-), eta and Klenow exo+ DNApolymerases. We were interested in the following major questions:

1. Is the DNA lesion containing N7-dG-2,7-DAM bypassed during primer extension experiments?

The *purpose* of this experiment was to determine if there is any *translesional bypass* in the case of the template containing one single guanine alkylated with 2,7-DAM or 2 guanines alkylated with 2,7-DAM.

2. Which dNTP has the highest frequency of insertion opposite the G-2,7-DAM lesion? In order to test the mutagenic potential of the N7-2,7-DAM lesion we we performed

single-nucleotide incorporation kinetics, were one single dNTP was incorporated at a time opposite the G*-containing the N7-G*-2,7-DAM lesion.

3. By comparison with N7-G-2,7-DAM lesion is the N^2-dG-MC lesion bypassed or is inhibiting the translesion synthesis (TLS) by Klenow exo-, T7 exo- and eta DNA polymerase?

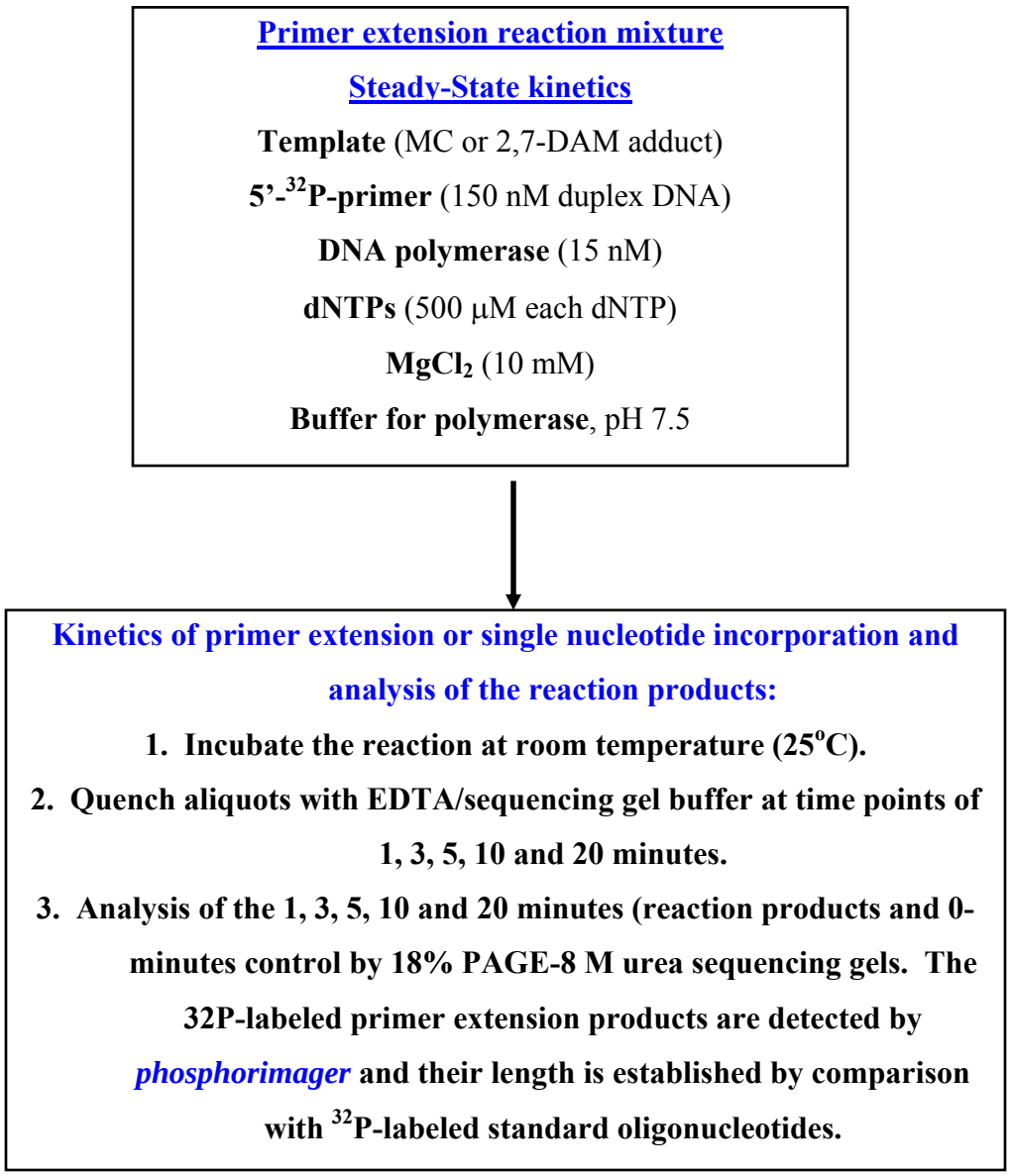
We knew from previous studies that the N^2-dG-MC lesion is inhibiting primer extension by Klenow exo- and T7 exo- DNA polymerases one nucleotide before the lesion (17-Chapter 1). However, it was important to compare the relative rates of primer extension for the two lesions, thus we repeated the translesion synthesis reactions for the N^2-dG-MC lesion. We run at steady state in real time the reactions of TLS not at end point kinetics as were originally run (17), so that we could compare the rate with which the primer is extended in time for the two lesions N^2-dG-MC and N7-G-2,7-DAM.

4. In addition to the questions raised above, how is the TLS performed by DNA polymerase eta in the case of the two lesions N^2-dG-MC and N7-G-2,7-DAM? DNA polymerase eta is known to be error prone translesion synthesis (being part of the Y family of DNA polymerases (3-50). The investigations of TLS using polymerase eta presented in this dissertation research are bringing new evidences for the inhibition or the translesion synthesis (TLS) as a function of the structure of the monoadduct (in this case we compared the minor groove vs. major grove orientations of the two adducts N^2-dG-MC and N7-G-2,7-DAM respectively and their effects upon the TLS).

Footprinting experiments indicate that Klenow Fragment covers 19 and 12 nucleotides on the template and primer strands, respectively (35). Other experiments indicates that templates 19-20 nucleotides and primers 12 nucleotides in length are sufficient to form stable polymerase/DNA complexes (35).

We conduccted the experiments of TLS and single nucleotide incorporation kinetics using primarily 24 mer templates alkylated at one single guanine with MC or 2,7-DAM annealed with 15 mer or 17 mer primers. Our quench flow experiments showed that the polymerization rate for T7 exo- polymerase (kpol) is higher for the 36mer/25 mer (template/primer) constructs than for 24 mer/17 mer (template/primer) (C. Clement –unpublished results-see Figures 85, 86-Appendix VI) suggesting that the DNA polymerases have higher affinity for longer templates/primer constructs. However, most of the *in vitro* TLS assays are conducted with 24 mer-27 mer/15mer-17-mer (templates/primers) constructs as a first test for the abilities of different DNA polymerases to bypass different lesions (36, 37, 38). It was reasonable to construct 24 mers and 27 mers templates primarily adducted at one guanine with MC or 2,7-DAM and conduct the *in vitro* TLS with these substrates. We also succeeded to synthesize a 36 mer template adducted at one single guanine with MC and we run *in vitro* TLS with this template annealed with a 25 mer primer. The template/primer constructs used for TLS and single nucleotide incorporation kinetics are presented in Scheme 4. The conditions used to run all the TLS reactions are described in details in the Specific Methods section and presented schematically in Scheme 6.

Scheme 6: The experimental design used to run primer extension and single nucleotide incorporation kinetics with MC and 2,7-DAM adducted templates:



2.3.6.1. Primer Extension kinetics (TLS) by T7 exo- polymerase.

T7 exo- polymerase was the first enzyme used for the primer extension assay using different substrates constructs (scheme 4). The results of primer extension (TLS) performed by this DNA polymerase are presented in figures 35-41.

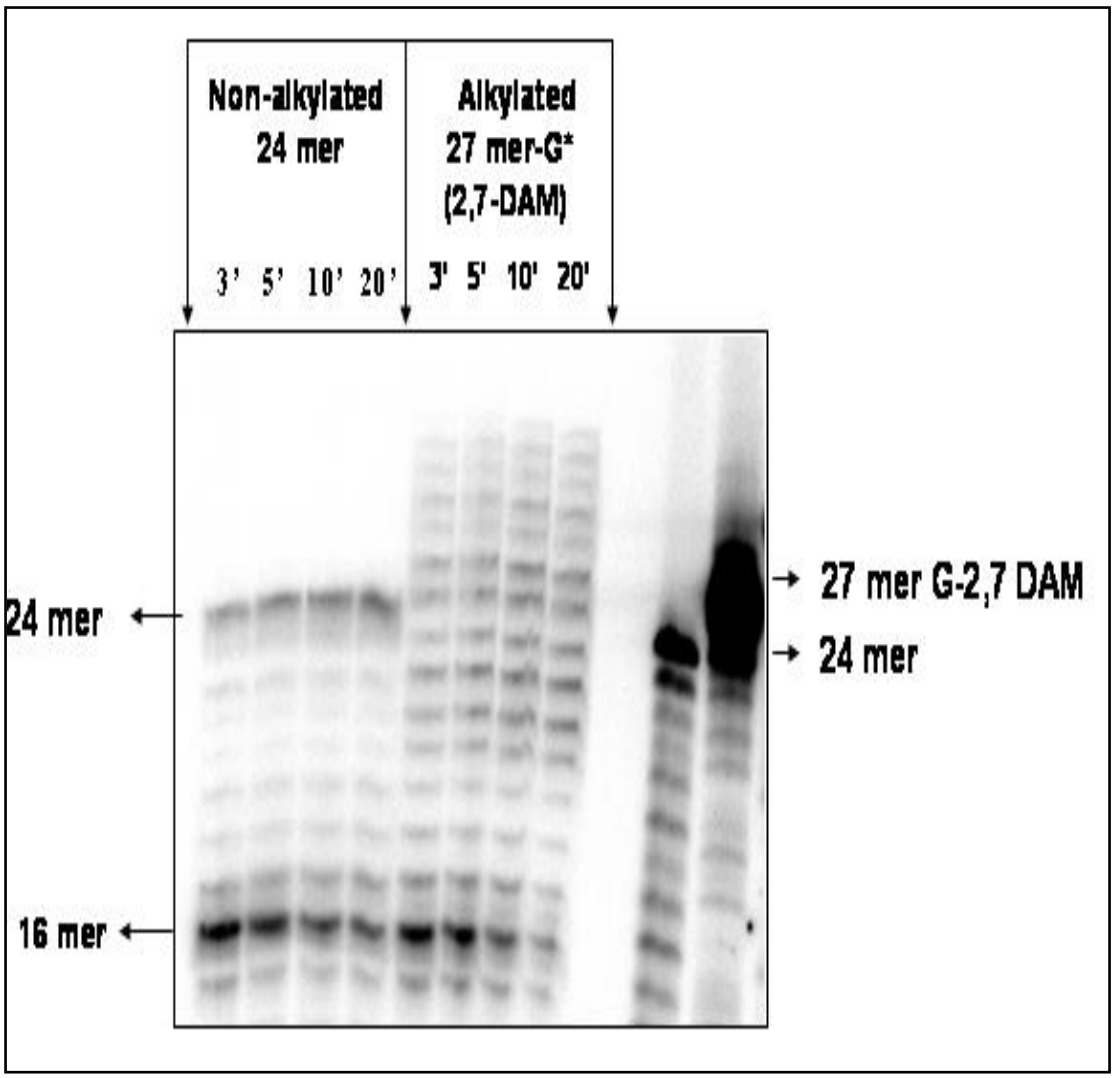
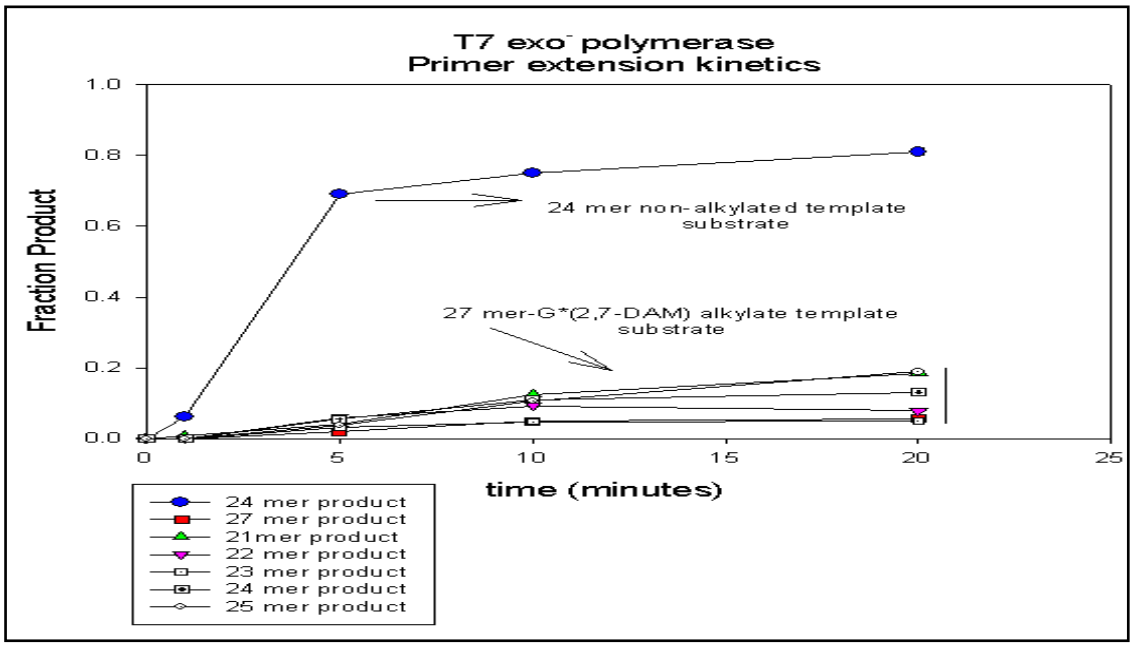


Figure 35: TLS (translesion synthesis) performed with T7 exo- polymerase and 27-mer-2,7-DAM template (substrate 18 in scheme 4) at 250 μM final concentration of each dNTP. The results of TLS showed that the 2,7-DAM lesion was fully bypassed by T7-exo- polymerase resulting in the production of fully extended primers (24 mer products). Extra extension of the alkylated template with 4 more nucleotides maybe due to the primer slippage during the reaction of primer extension.

(A)



(B)

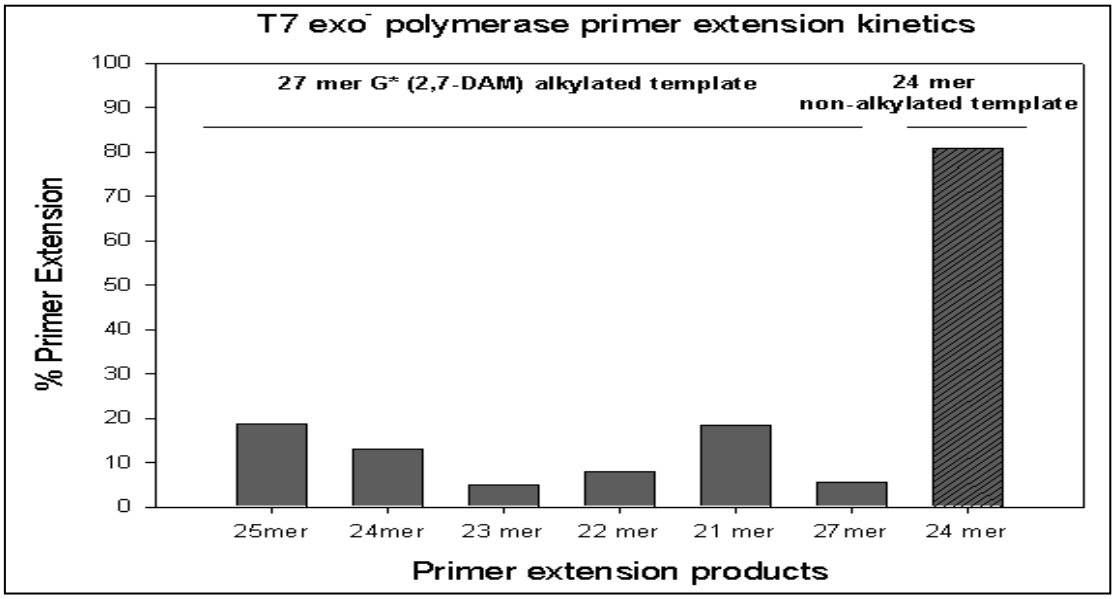


Figure 36: Quantitative measurements of TLS (translesion synthesis) performed with T7 exo- polymerase and 27-mer-2,7-DAM template (substrate 18 in scheme 4) at 250 μM final concentration of each dNTP. The data from figure 35 were quantified as fraction of product (A) or % primer extended (B) using the Phosphorimager and the Image Quant as described in Specific Methods section.

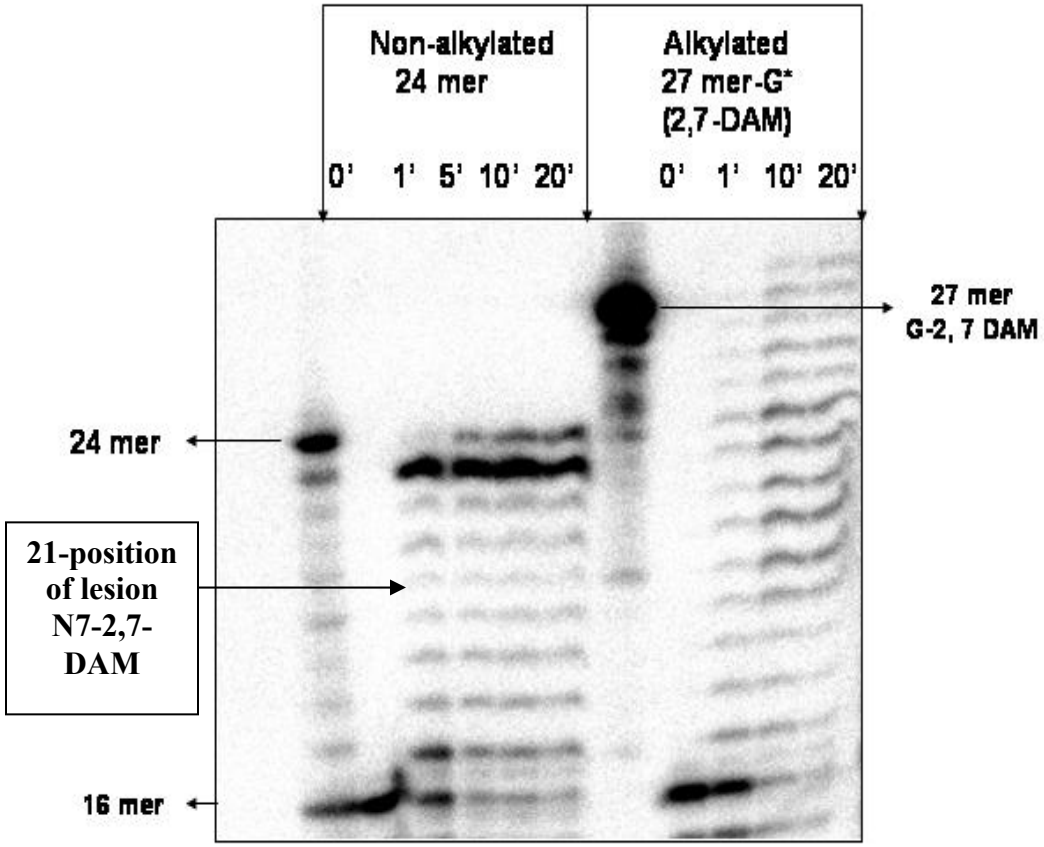
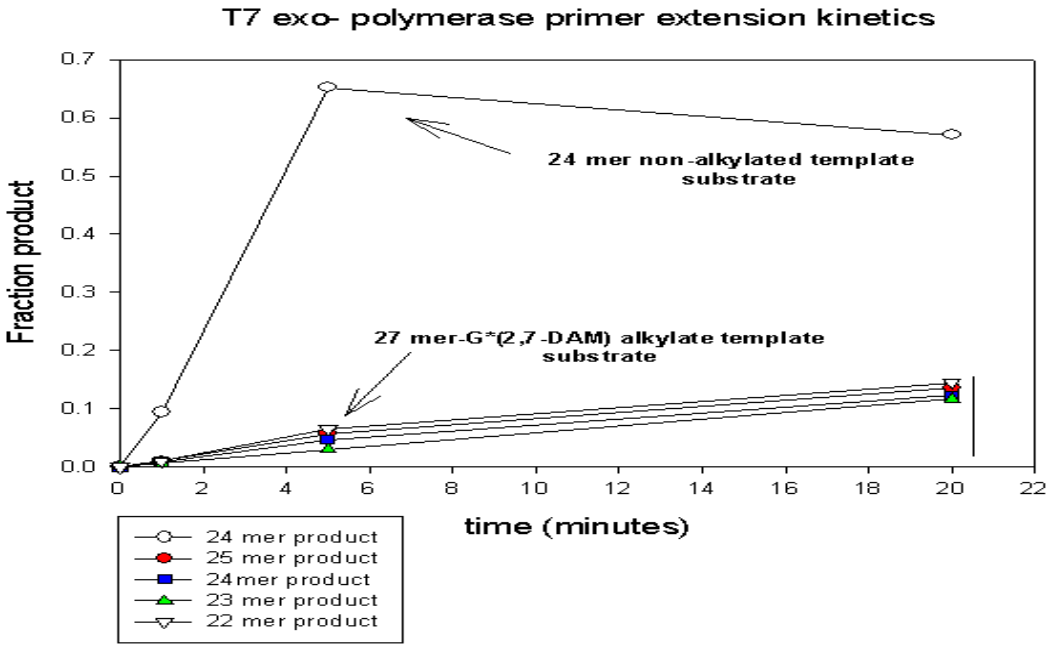


Figure 37: TLS (translesion synthesis) performed with T7 exo- polymerase and 27-mer-2,7-DAM template (substrate 18 in scheme 4) at 300 μM final concentration of each dNTP. The results of TLS showed that the 2,7-DAM lesion was fully bypassed by T7-exo- polymerase resulting in the production of fully extended primers (24 mer products). Extra extension of the alkylated template with 4 more nucleotides may be due to the primer slippage during the reaction of primer extension.

(A)



(B)

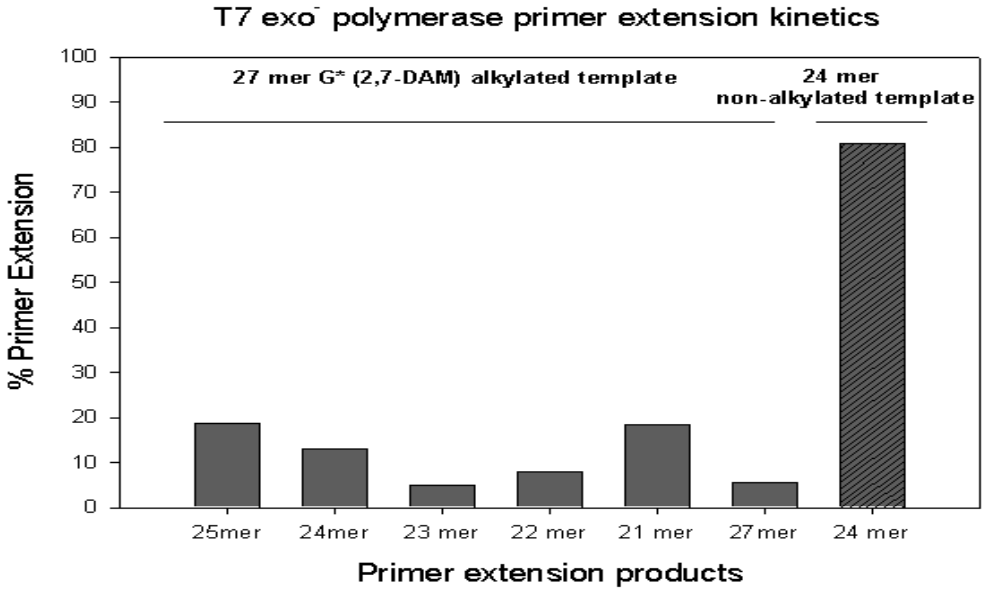


Figure 38: Quantitative measurements of TLS (translesion synthesis) performed with T7 exo- polymerase and 27-mer-2,7-DAM template (substrate 18 in scheme 4) at 250 μM final concentration of each dNTP. The data from figure 37 were quantified as fraction of product or % primer remaine unextended using the Phosphorimager and the Image Quant as described in Specific Methods section.

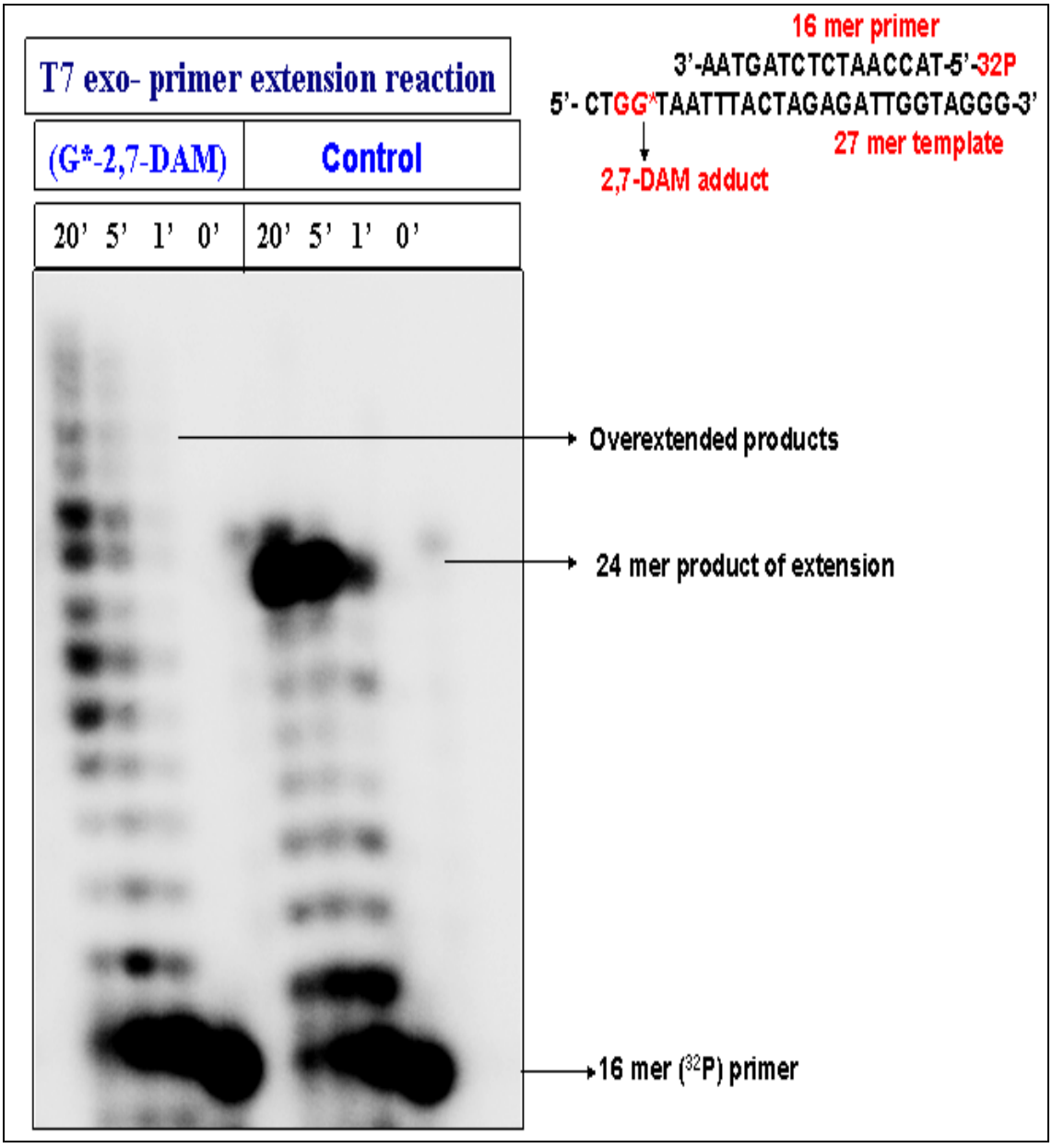


Figure 39: TLS (translesion synthesis) performed with T7 exo- polymerase and 27-mer-2,7-DAM template (substrate 18 in scheme 4) at 500 μM final concentration of each dNTP. The results of TLS showed that the 2,7-DAM lesion was fully bypassed by T7-exo- polymerase resulting in the production of fully extended primers (24 mer products). Extra extension of the alkylated template with 4 more nucleotides maybe due to the primer slippage during the reaction of primer extension.

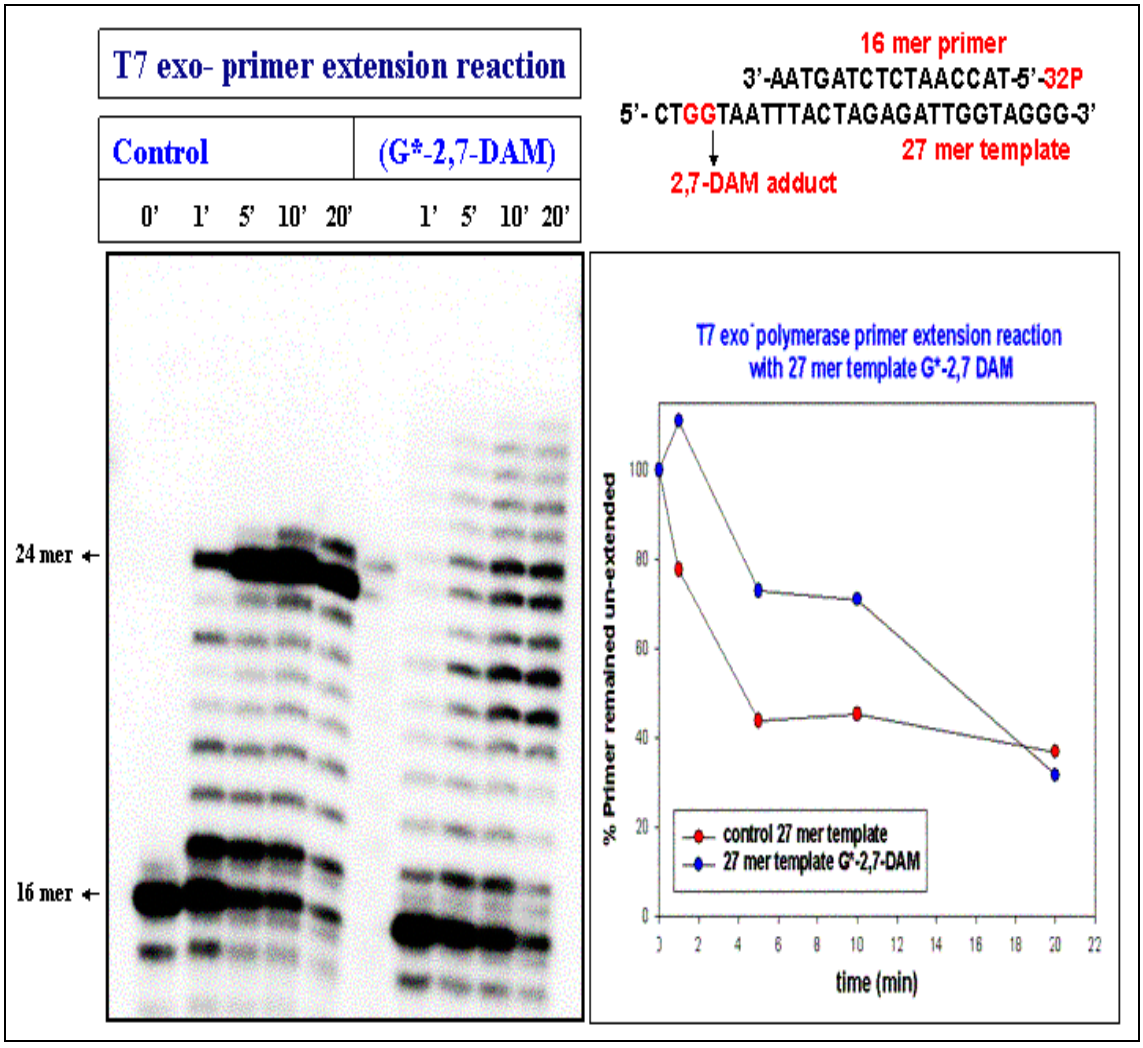


Figure 40: TLS (translesion synthesis) performed with T7 exo- polymerase and 27-mer-2,7-DAM template (substrate 18 in scheme 4) at 500 μM final concentration of each dNTP. The results of TLS showed that the 2,7-DAM lesion was fully bypassed by T7-exo- polymerase resulting in the production of fully extended primers (24 mer products). Extra extension of the alkylated template with 4 more nucleotides may be due to the primer slippage during the reaction of primer extension. The rate of primer utilization is slowere in the case of the alkylated template as compared with control template, suggesting that T7 exo- polymerase has lower affinity for the alkylated substrate as compared with the non-alkylated control template.

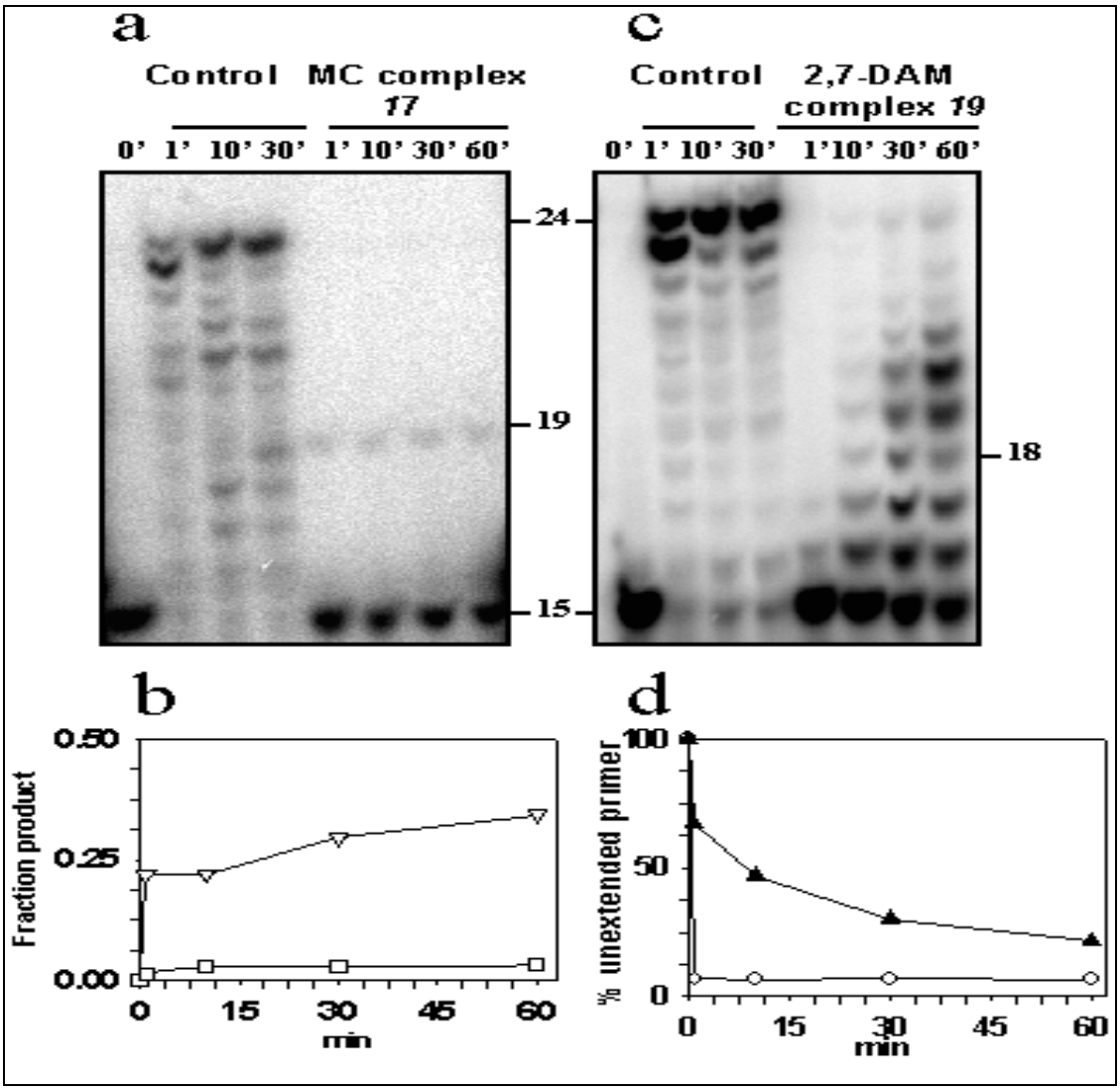


Figure 41: PAGE analysis of primer extension on MC- and 2,7-DAM-adducted templates by T7 (exo-) DNA polymerase. (a) PAGE of primer extension of MC-adducted template/primer complex 17 (scheme 4). Reaction conditions: 250 μM dNTP (each), 200 nM complex 17, 40 nM T7 (exo-) DNA polymerase; 25^{o} C. Minutes of incubation time are indicated above each lane. (b) Plots of results in (a): The fractions of the expected 24-mer fully-extended control product (∇) and an aborted 19-mer product (1 nucleotide before MC lesion) (□) are plotted as a function of incubation time. The rates for the two products are different, the MC lesion is not bypassed and a 19-mer aborted product (1 nucleotide before lesion) is accumulated at a slower rate than in the case of the same reaction run with Klenow (exo-) DNA polymerase (see Fig. 43 b). (c) PAGE of primer extension of 2,7-DAM-adducted template/primer complex 19 (scheme 4). Reaction conditions: Same as in (a). (d) Plots of results in (c): The percent of original primer remained un-extended during the primer extension reaction is plotted as a function of incubation time for the complex 19 using the data presented in 41 (c). The 2,7-DAM lesion (closed triangles) is bypassed with a slower rate than the control undamaged template (closed circles).

The primer strand in the MC complex **17** was again elongated only to a 19-mer, indicating a complete block of elongation one nt before the MC adduct **7** in the template (Figure 41-a). This polymerase exhibited a low activity for elongation of the starting 15-mer primer in complex **17** relative to elongation of the control, nonadducted complex and also relative to elongation in complex **17** by Klenow (exo-) DNA polymerase. In sharp contrast, the primer strand in the 2,7-DAM-dG-N7 complex **19** was elongated to full its 24 nt length, indicating bypass of the adduct **8** in the template, albeit at a lower efficiency relative to the control, nonadducted template (Figure 41 b and Figure 42). A similar result was obtained using complex **18** as the substrate for the T-7 (exo-) polymerase as shown in figures 35-40. This substrate yielded, in addition, products of overextension of the primer by 1, 2, or 3 nts beyond the template terminus, most likely originating from primer slippage. Figure 42 presents a detailed mathematical fit of the kinetics data presented in Figure 41. Since the data were best fitted to a double exponential decay model, we hypothesized that the primer extension reaction proceeded with a change in the conformation of the DNA polymerase as it was binding to the template/primer substrate.

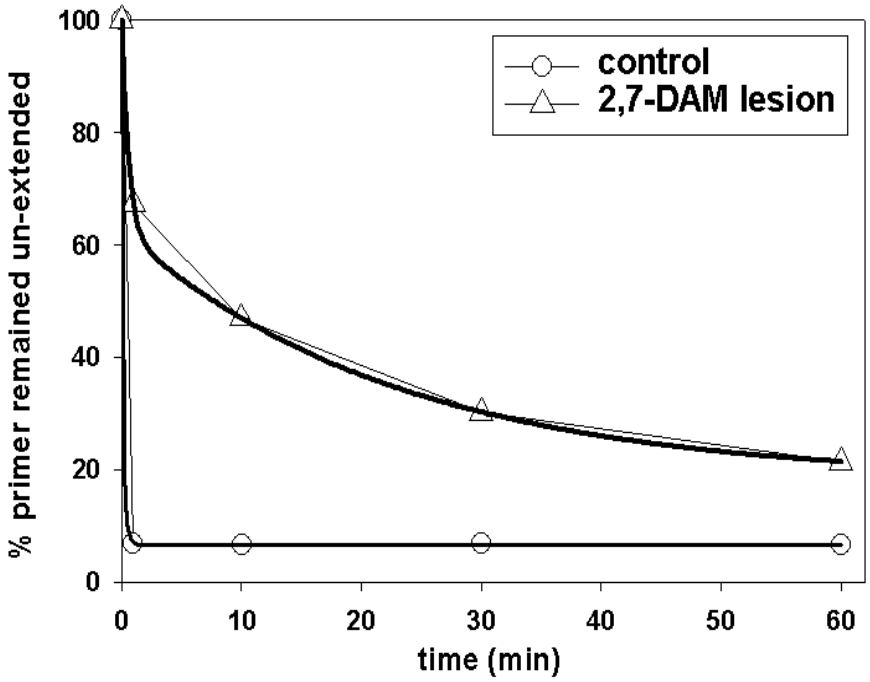


Figure 42: Quantitative analysis of primer extension on 2,7-DAM-adducted template of complex 18 (Scheme 4) by T7 (exo-) DNA polymerase. Double-exponential decay fit of experimental data presented in figure 41 (d) to the equation f=Y(0)+ a*exp(-b*x)+c*exp(-d*x). The % primer remained un-extended follows a double exponential decay with a faster rate (d = 1.2080) for the control than for the 2,7-DAM alkylated template (d= 0.0527) probably due to a lower affinity of the polymerase for the alkylated substrate and a loss of processivity near the bulky lesion.

2.3.6.2. Primer Extension kinetics (TLS) by Klenow exo- polymerase.

The primer strand in the 2,7-DAM complex **19** (scheme 4) containing adduct **8** was elongated past the 2,7-DAM adduct **8** (scheme 3) using DNA polymerse Klenow exo-. This is concluded from the presence of 18-24 nt long primer products at gradually increasing relative intensities of the fully extended 24-mer as the time of the incubation was increased from 1 to 60 min (Figure 43 c). This extension reaction was conducted at a 5:1 stoichiometry of the duplex **19** : enzyme ratio. Another extension reaction was run at 1:1 stoichiometry, which resulted in a slightly different pattern, in which some accumulation of intermediates between 17- and 23-mers was detected (Figure 43 d). In any case, both results indicate that the 2,7-DAM-dG-N7 adduct **8** is bypassed by Klenow (exo-) DNA polymerase, in contrast to the MC-dG-N^2 adduct **7**. It is notable, however, that the rate of disappearance of the starting 15-mer primer is much slower than that in the nonadducted control complex or in the MC-adducted complex 17 as can be seen from all results of primer extension reactions presented in Figures 44-53 performed with DNA polymerase Klenow exo- using templates constructs 18 or 21 (see scheme 4).

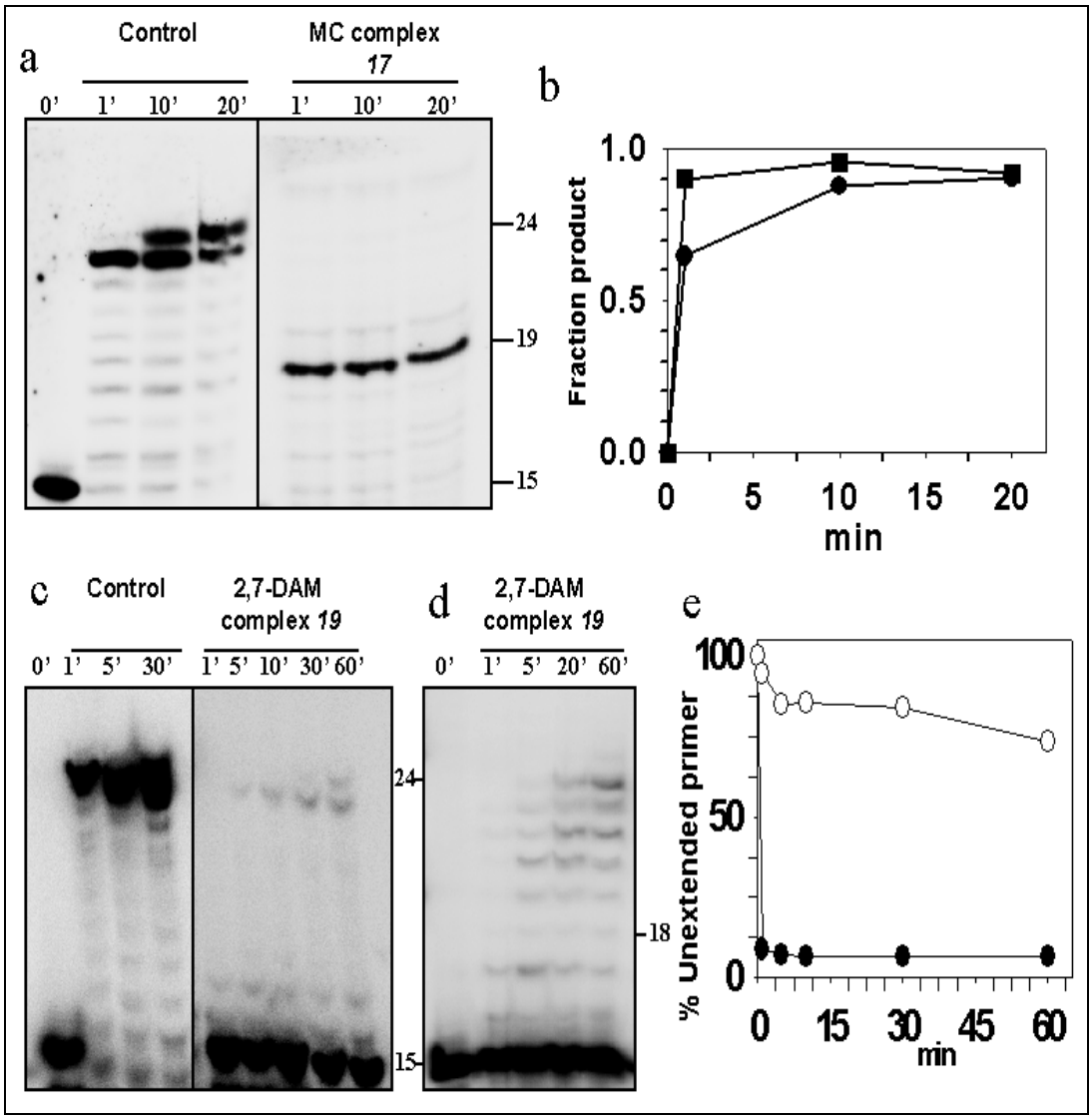
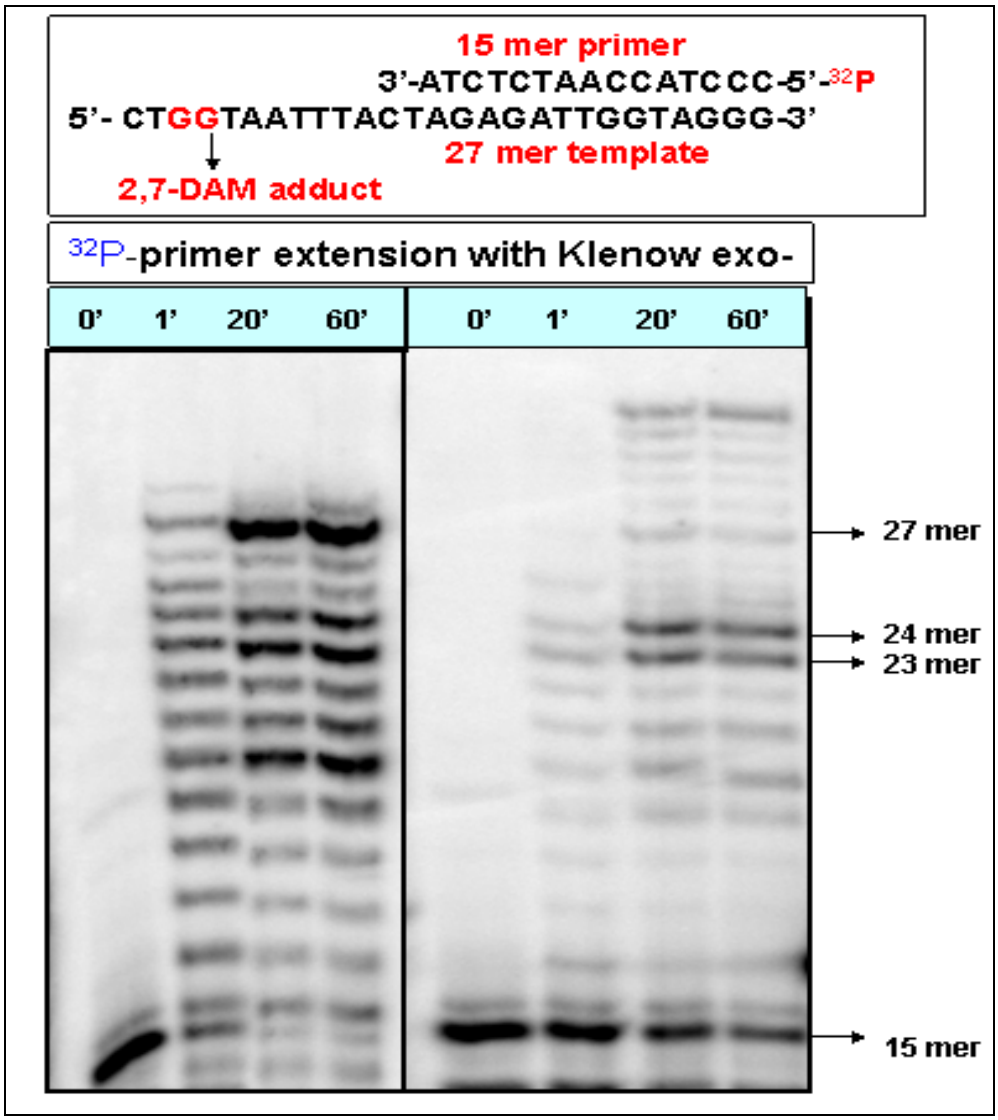


Figure 43: PAGE analysis of primer extension of MC and 2,7-DAM-adducted templates by Klenow (exo-) DNA polymerase. (a) PAGE of primer extension in MC-adducted template/primer complex 17. Reaction conditions: 250 μM dNTP (each), 200 nM complex 17, 40 nM Klenow (exo-) DNA polymerase, 25^{o}C. Minutes of incubation time are indicated above each lane. (b) Plots of results in (a): The fraction of the expected 24-mer fully extended product (control) and aborted 19-mer product (1 nucleotide before MC lesion) is plotted as a function of incubation time. (Closed circles: 24-mer control product; closed squares: 19-mer abortive product.) (c) PAGE of primer extension of 2,7-DAM-adducted template/primer complex 19. Reaction conditions: 250 μM dNTP (each), 200 nM complex 19, 40 nM Klenow (exo-) DNA polymerase, (S:E ratio 5:1); 25^{o} C. Minutes of incubation time are indicated above each lane (d) same as (c), except used 40 nM complex (S:E ratio 1:1). (e) Plots of results in (c) and (d): The percent of original primer remained un-extended during the primer extension reaction is plotted as a function of incubation time. The 2,7-DAM lesion (closed triangles) is bypassed with slower rate as compared with the control undamaged template (closed circles).

A



[Control] template **[G*-2,7-DAM]-alkylatd template**

B

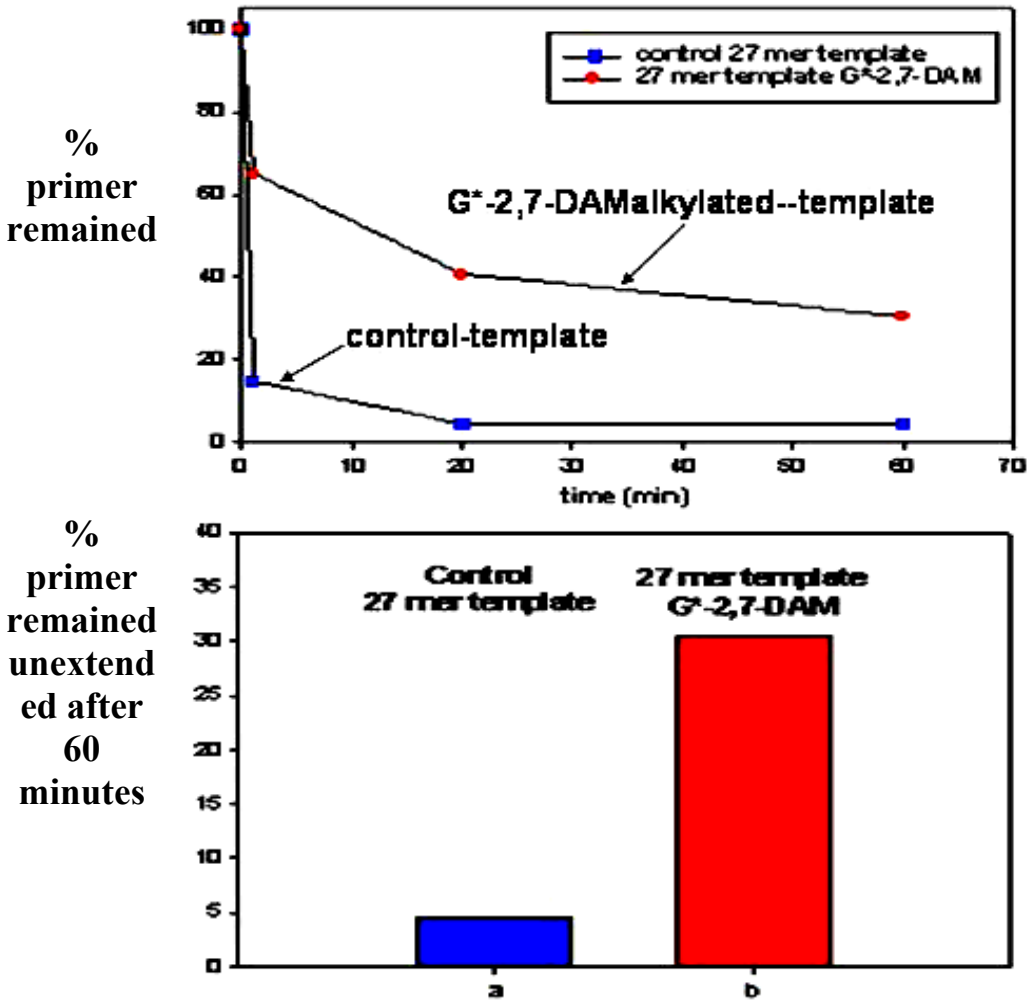
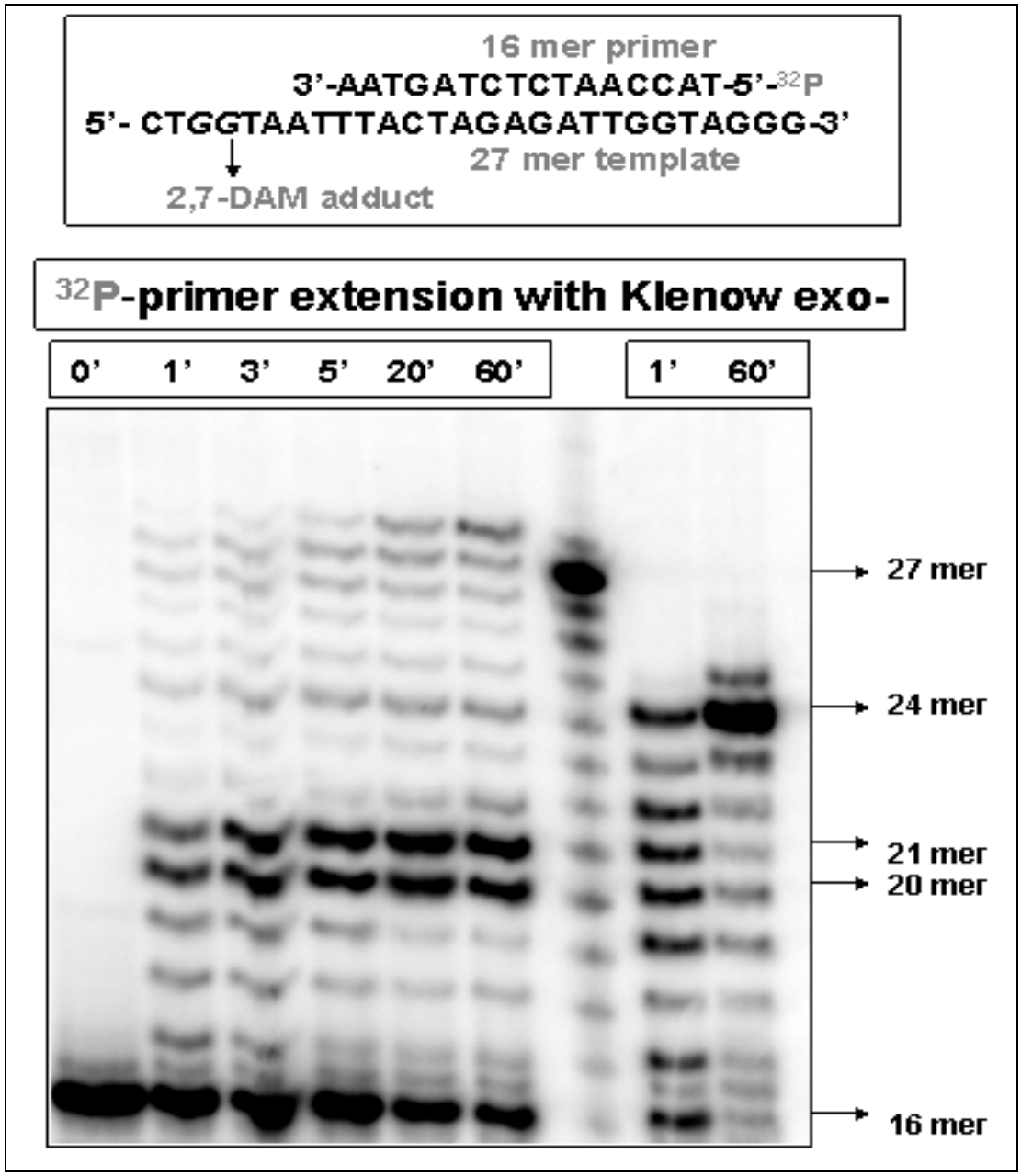


Figure 44. Translesion synthesis (TLS) using a 27-mer-G*-2,7-DAM (construct 21, scheme 4) and Klenow exo- polymerase at 500 μM each dNTPs. (A) The primer is less extended when is annealed with the alkylated template as compared with the control, non-alkylated template. (B) The quantitative analysis shows the same %primer remained un-extended kinetics as is shown in figures 40, 42 and 43 (e). The rate of primer extension is slower in the case of alkylated template and the yield of the primer remained-un-extended is higher than in the control non-alkylated template, suggesting that the polymerase has a lower affinity for the alkylated template and is loosing processivity for the DNA substrate during the primer extension reaction.

A



[G*-2,7-DAM] template **[Control]-template**

B

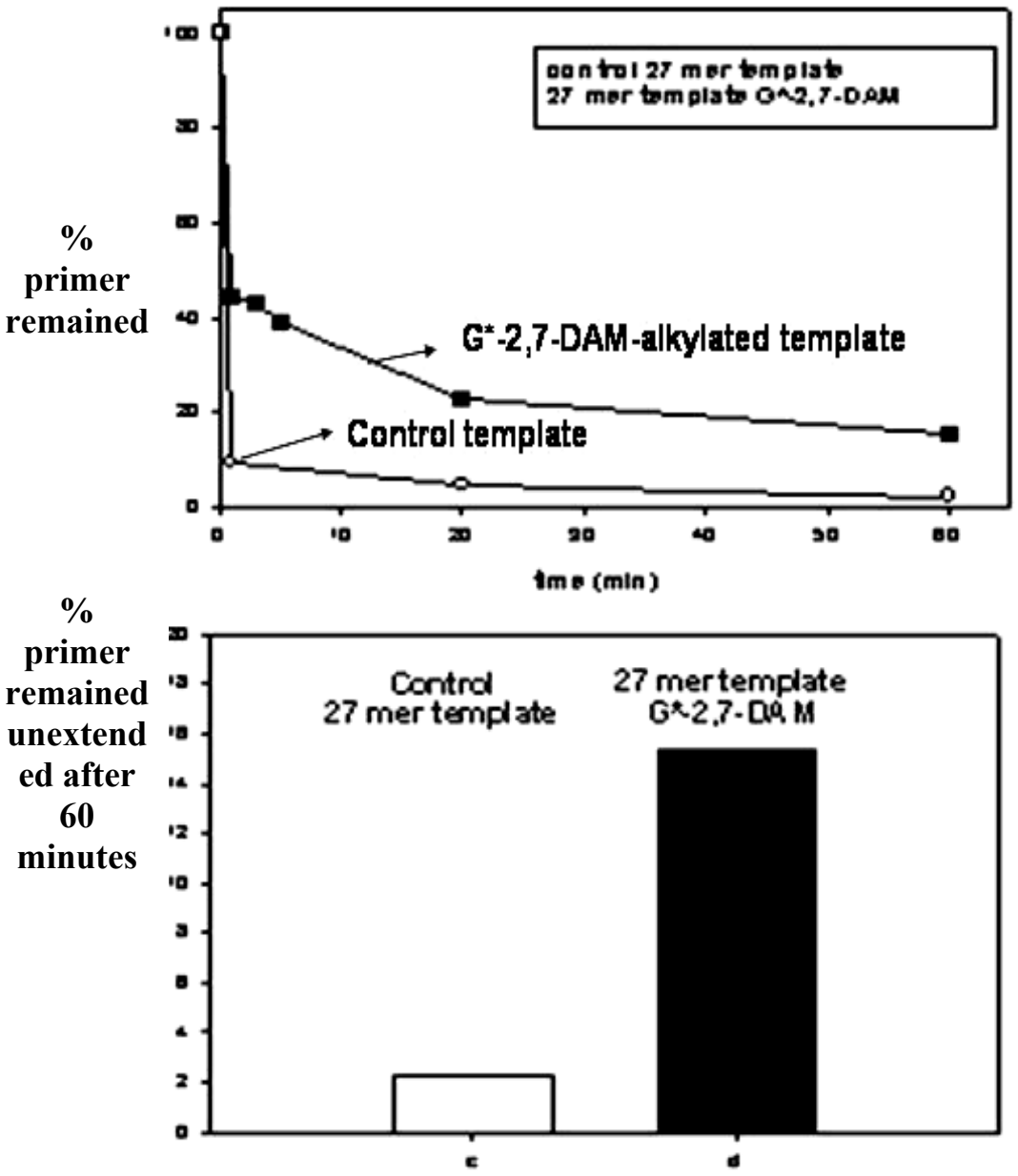


Figure 45. Translesion synthesis (TLS) using a 27-mer-G*-2,7-DAM (construct 18, scheme 4) and Klenow exo- polymerase at 500 μM each dNTPs. (A) The results are showing that the primer is less extended when is annealed with the alkylated template as compared with the control, non-alkylated template. (B) The rate of primer extension is slower in the case of alkylated template and the yield of the primer remained-un-extended is higher than in the control non-alkylated template.

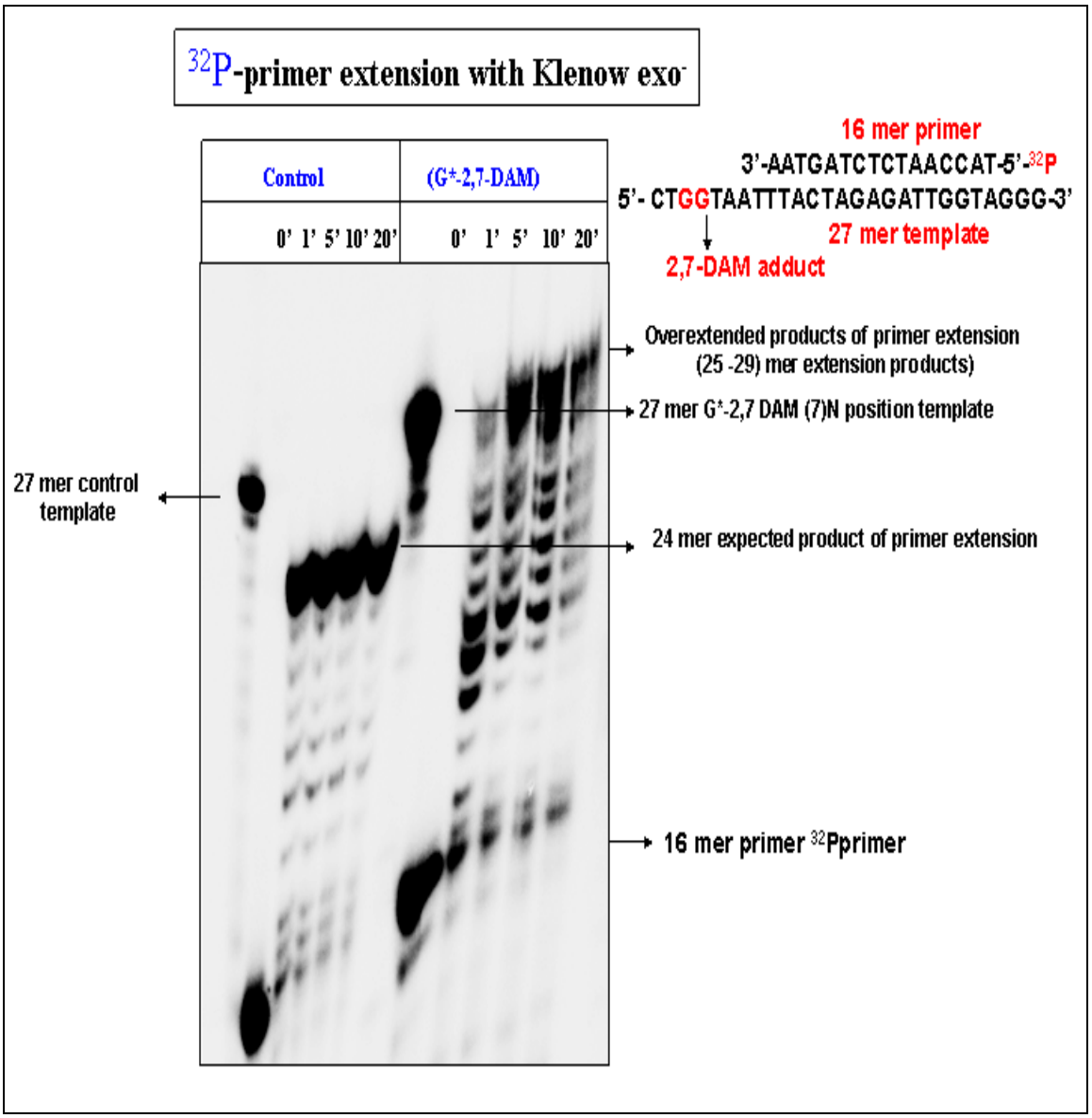


Figure 46. Translesion synthesis (TLS) using a 27-mer-G*-2,7-DAM (construct 18, scheme 4) and Klenow exo- polymerase at 500 μM each dNTPs (independent experiment). The rate of primer extension is slower in the case of alkylated template and the yield of the primer remained-un-extended is higher than in the control non-alkylated template.

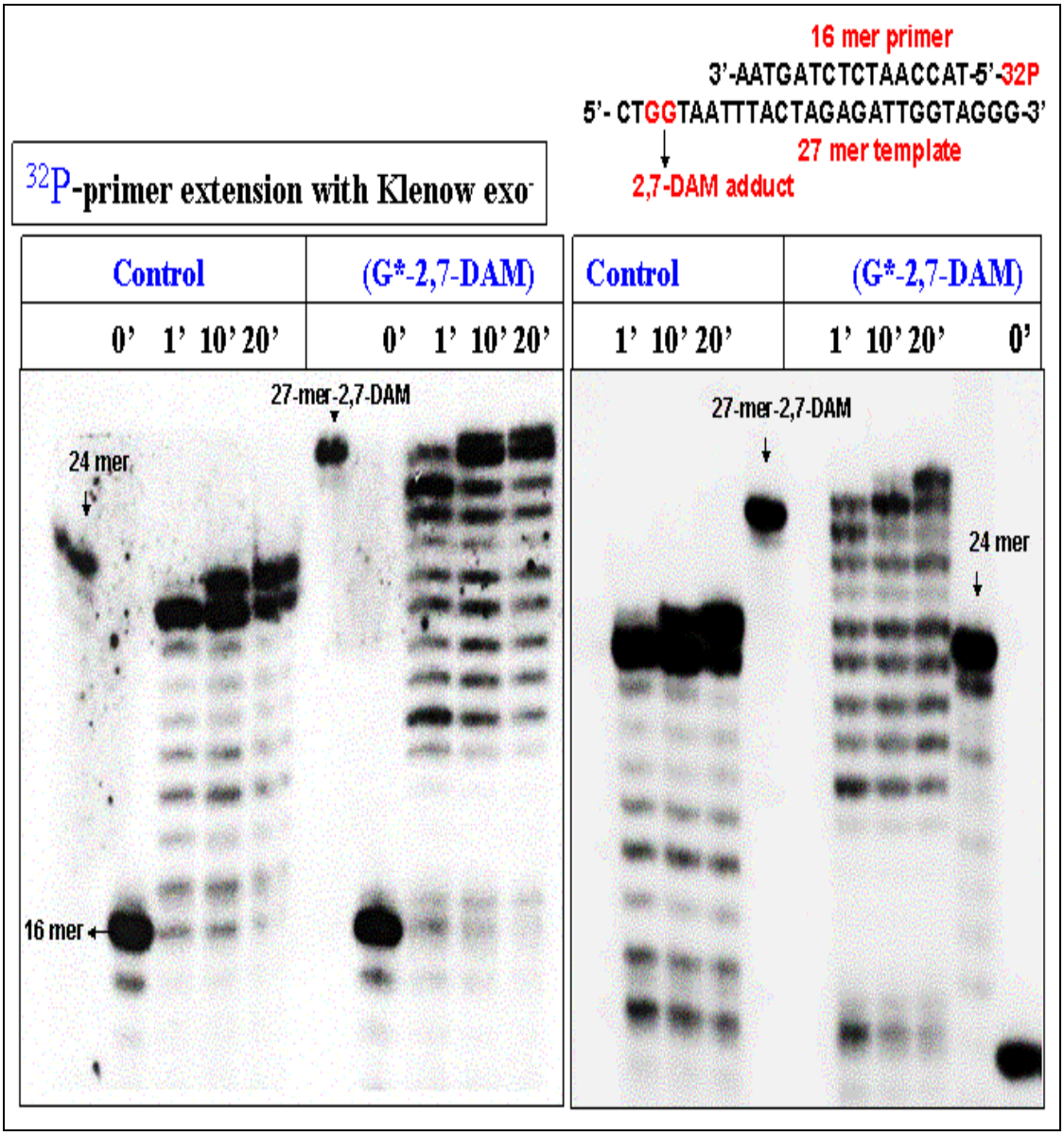
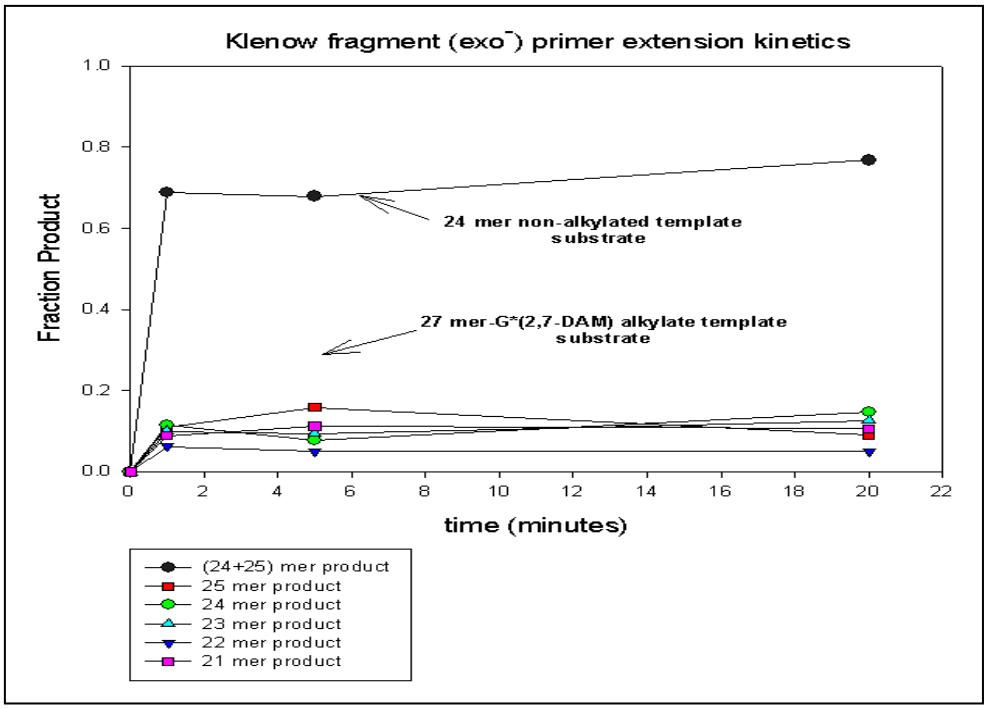


Figure 47. Translesion synthesis (TLS) using a 27-mer-G*-2,7-DAM (construct 18, scheme 4) and Klenow exo- polymerase at 300 μM each dNTPs (independent experiment). The rate of primer extension is slower in the case of alkylated template and the yield of the primer remained-un-extended is higher than in the control non-alkylated template.

(A)



(B)

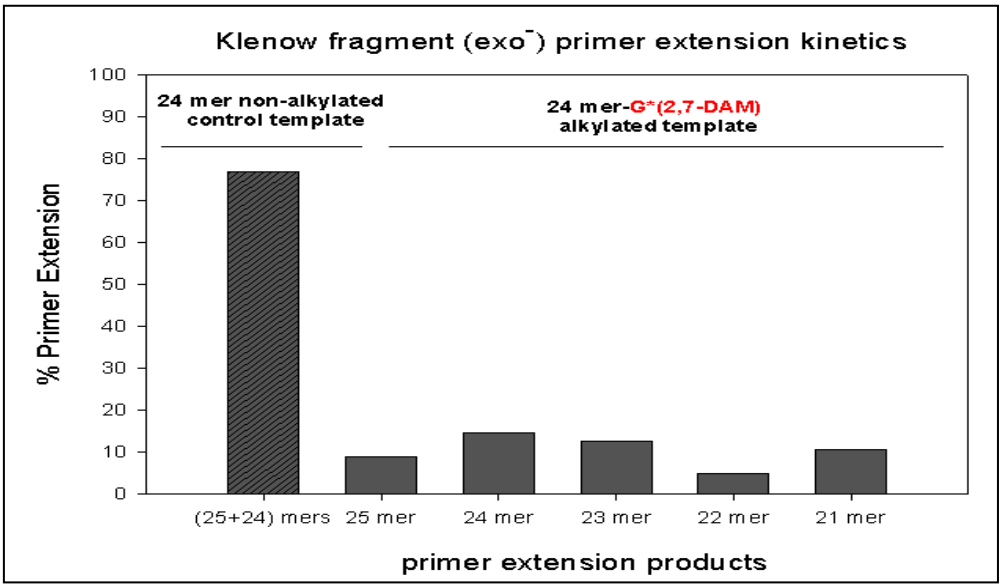


Figure 48: Quntitative measurements of TLS (translesion synthesis) performed with Klenow exo- polymerase and 27-mer-2,7-DAM template (substrate 18 in scheme 4) at 250 μM final concentration of each dNTP. The data from Figure 46 were quantified as fraction of product (A) or % primer remained (B) unextended using the Phosphorimager and the Image Quant as described in Specific Methods section.

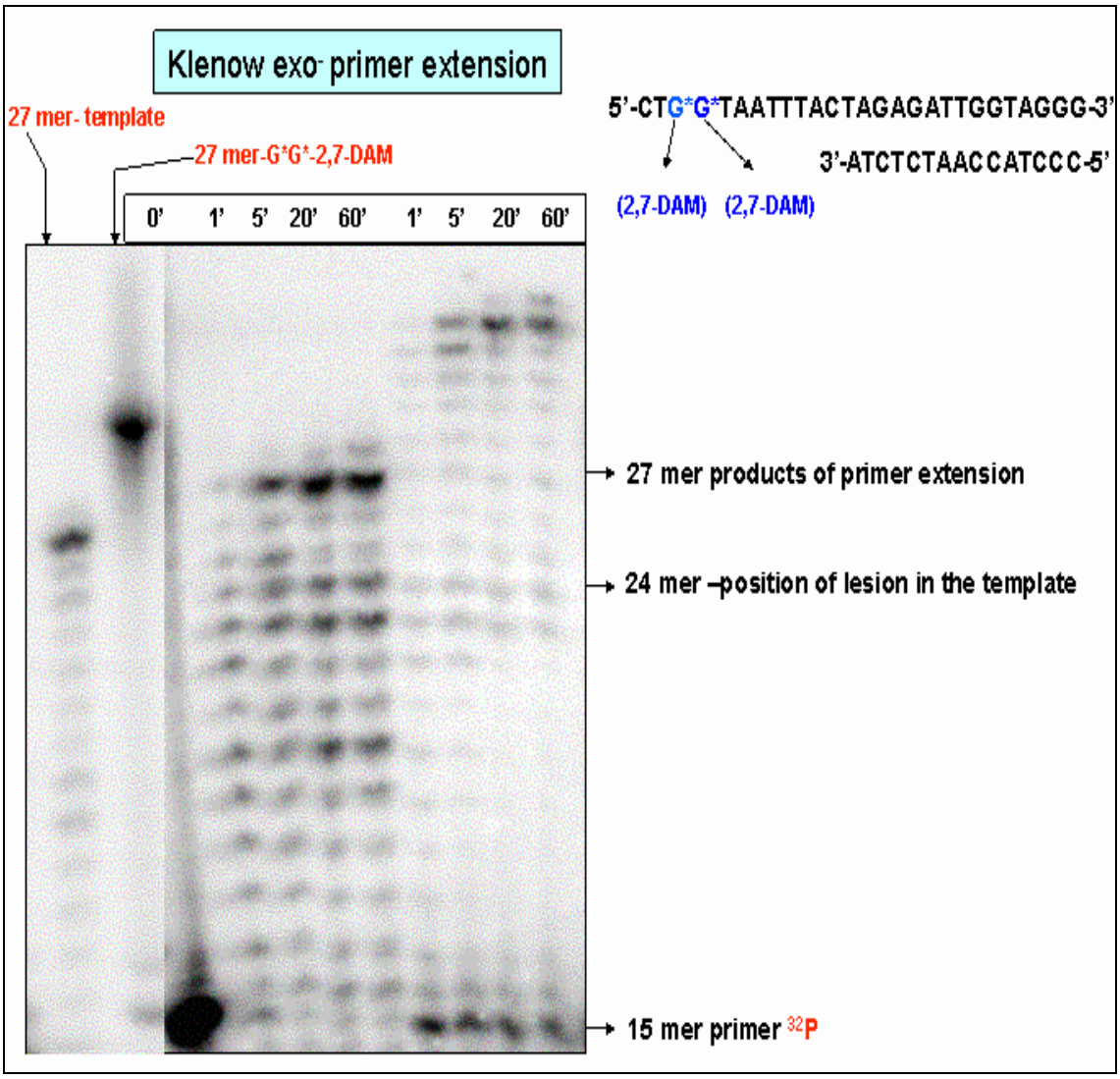


Figure 49: Translesion synthesis (TLS) using a (27-mer-G*-2,7-DAM)2 (construct 24, scheme 4) and Klenow exo- polymerase at 500 μM each dNTPs. The rate of primer extension is slower in the case of double alkylated template and the yield of the primer remained-un-extended is higher than in the control non-alkylated template. These results are consistent with the those obtained for the 27-mer template containing one single G alkylated with 2,7-DAM.

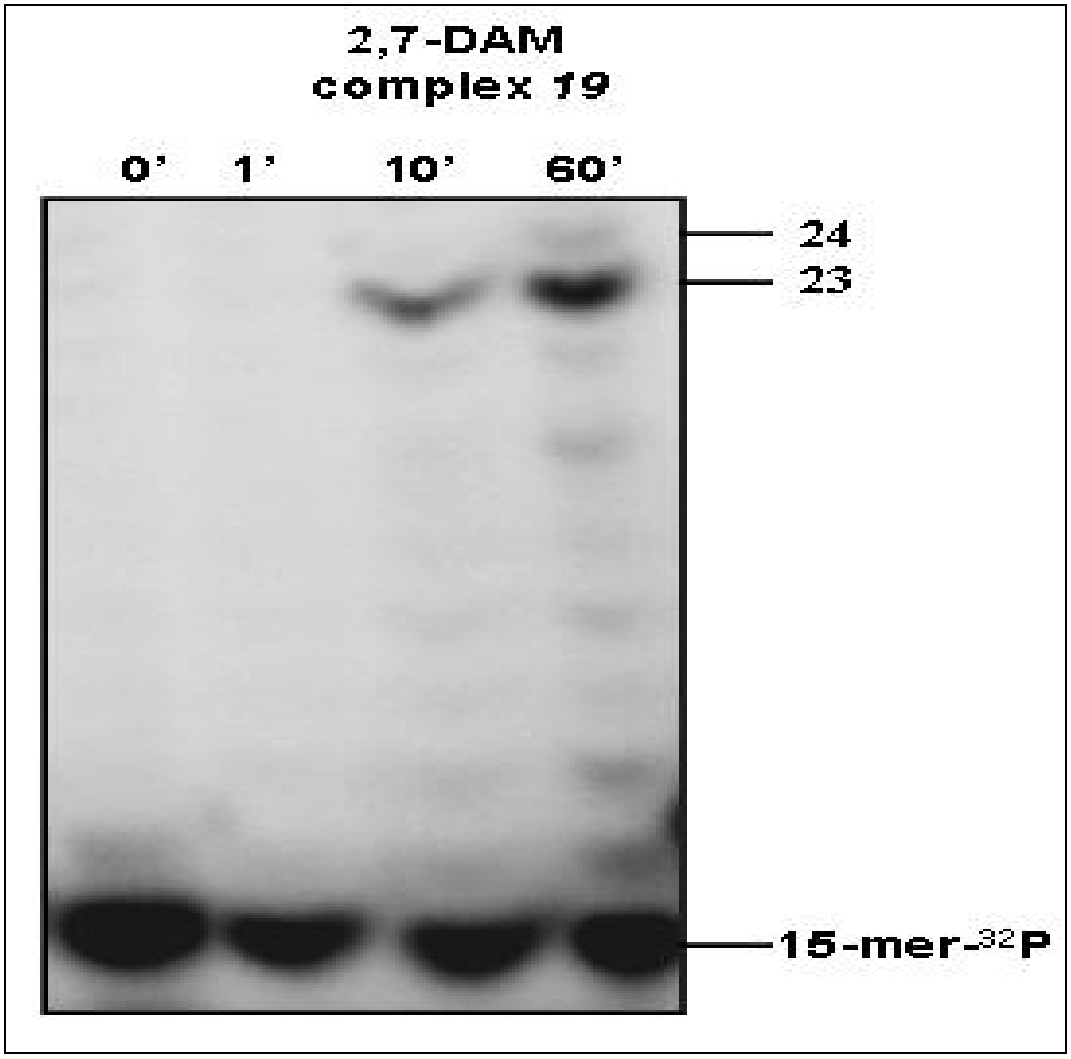


Figure 50. PAGE analysis of primer extension (TLS) on 2,7-DAM-adducted template of complex *19* by Klenow (exo-) DNA polymerase at 37°. The increased in the temperature had a better effect upon the TLS reaction *in vitro*.

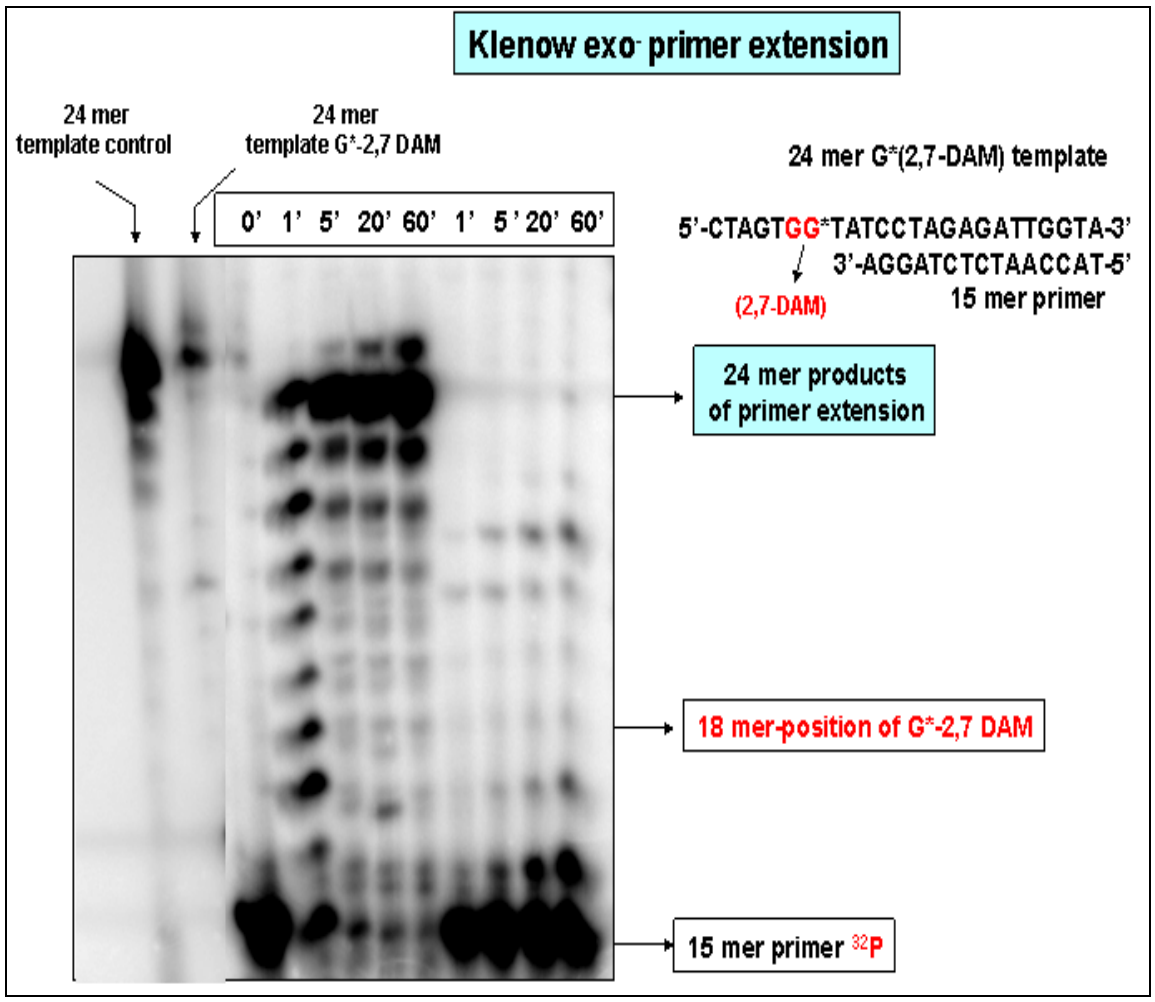


Figure 51: The primer extension reaction using a 27-mer-G*-2,7-DAM and Klenow exo- polymerase but a different template construct (the drug is moved toward the midle of template) (construct 18 scheme 4). The ratio substrate (primer/template): enzyme is 6:1. The lesion with 2,7-DAM is bypassed but with slower rate with respect to the control template. Also, at this enzyme: substrate ratio the enzyme is less efficient in extending fully the primer (21-22 mers instead of 24 mers are the final products of the reaction).

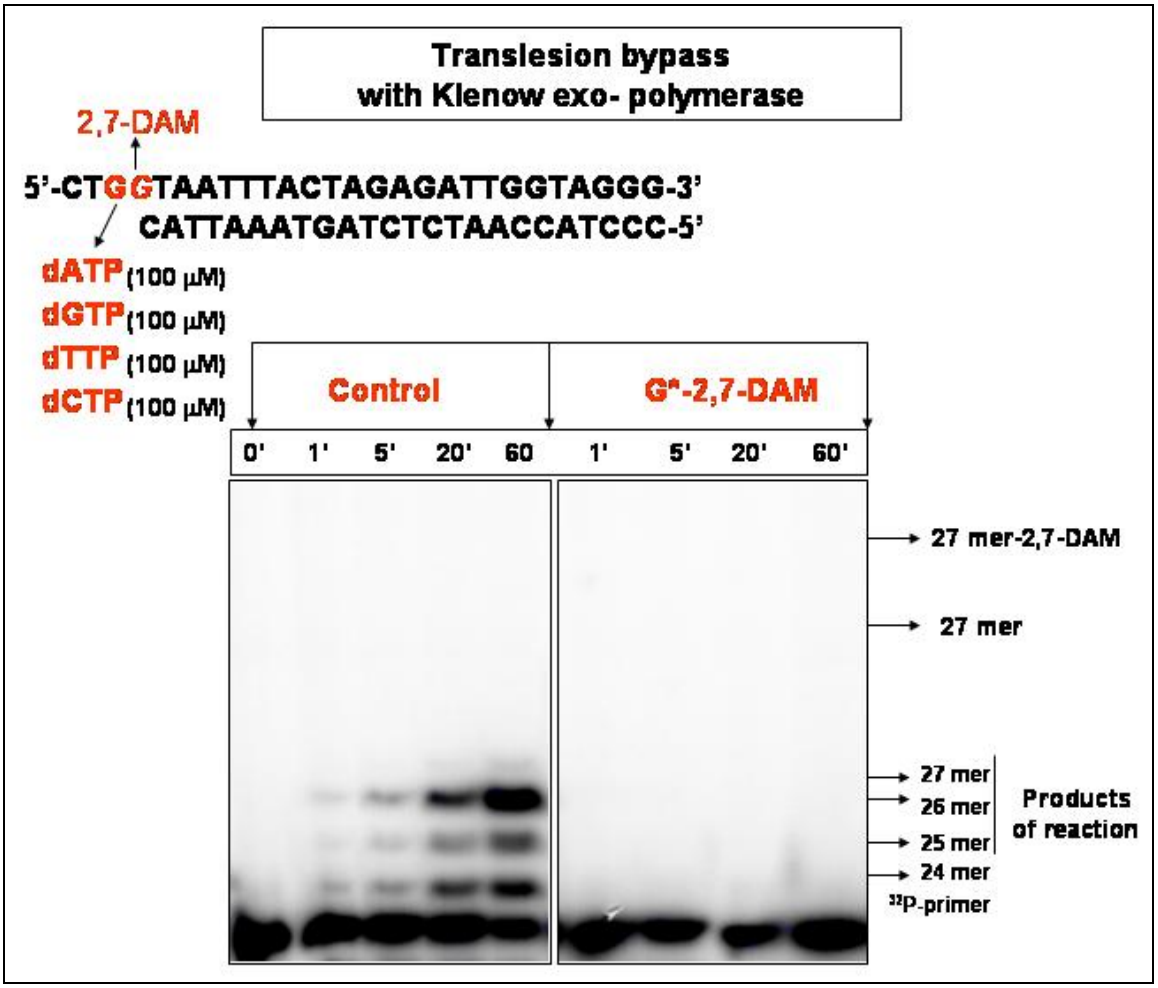


Figure 52: The primer extension reaction using a 27-mer-G*-2,7-DAM (construct (18)-scheme 4) and Klenow exo- polymerase at lower concentration of dNTPs (100 μM instead of 500 μM). At lower dNTPs concentration (i.e. below 200 μM) the lesion with G*-2,7-DAM was not bypassed anymore, suggesting that the 2,7-DAM drug within the template is decreasing the relative affinity for the dNTPs (i.e. the Km for dNTPs is increased).

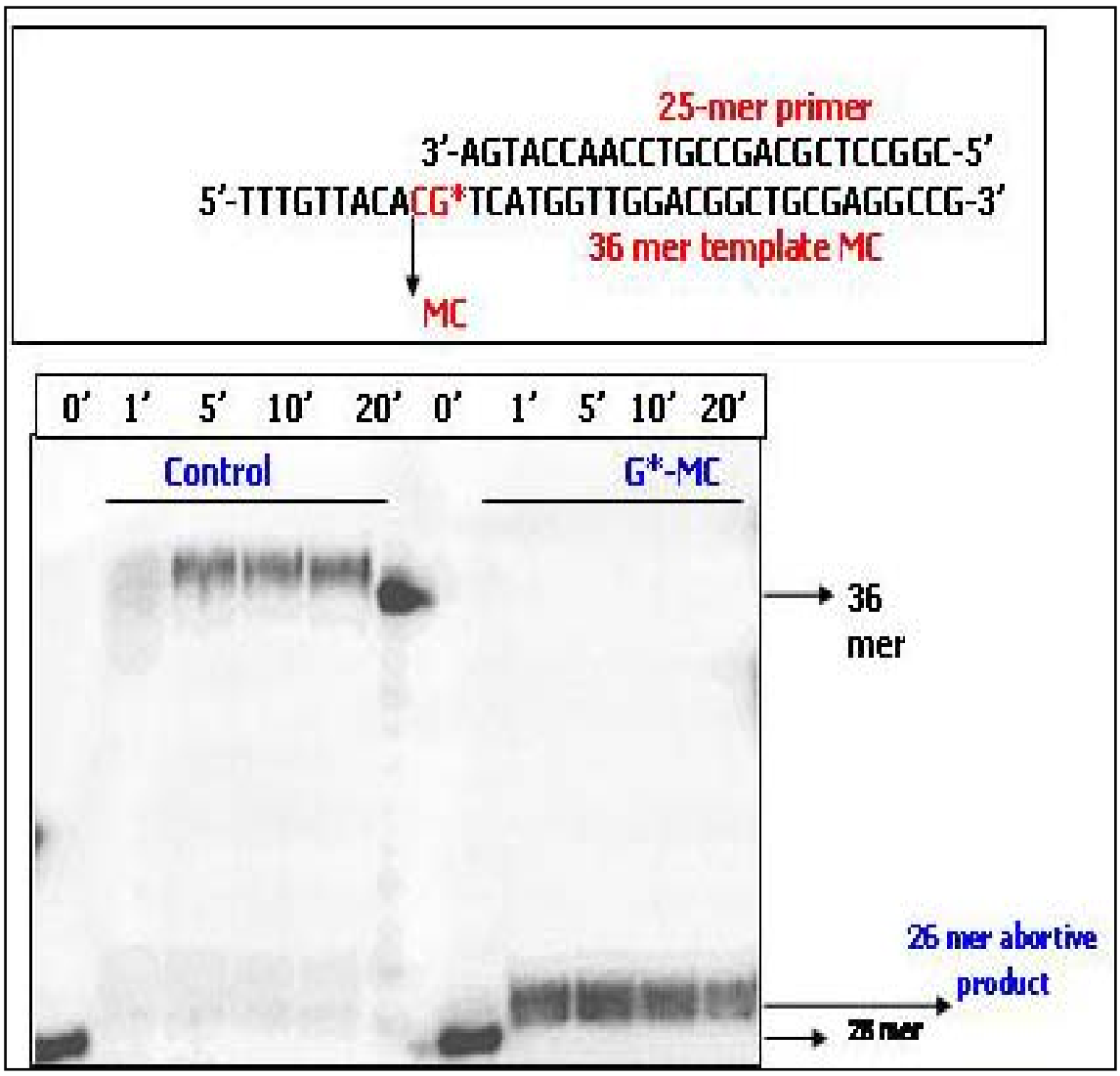


Figure 53: Klenow exo- polymerase primer extension of a dG-N^2-mitomycin C alkylated 36 mer template. The reaction of primer extension is inhibited by the presence of MC drug, after one nucleotide incorporation opposite the lesion.

2.3.6.3. Primer Extension kinetics (TLS) by DNA polymerase eta.

The comparison between the TLS of N7-G-2,7-DAM and TLS of N^2-dG-MC were performed with the error prone DNA polymerase eta and showed that a bulky minor groove adduct (N^2-dG-MC) is completely inhibited the polymerase one nucleotide before lesion, while the major groove adduct (N7-G-2,7-DAM) is allowing lesion bypass.

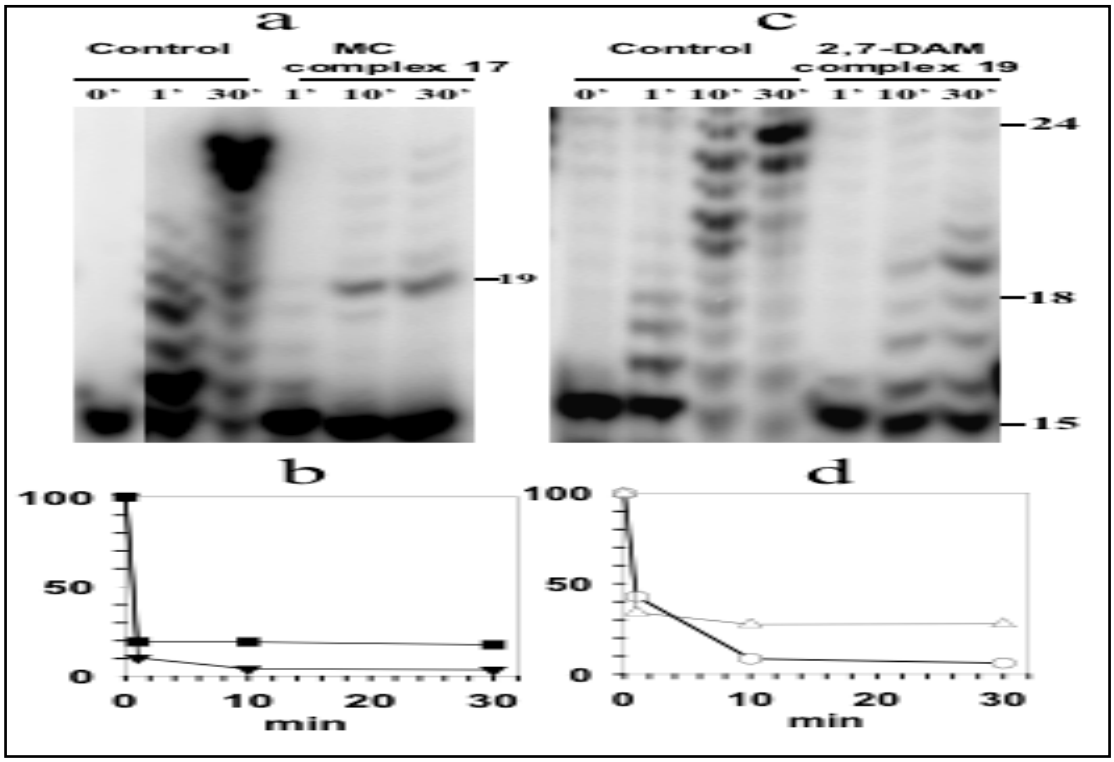


Figure 54: PAGE analysis of primer extension on MC- and 2,7-DAM-adducted templates by DNA polymerase eta. (a) PAGE of primer extension of MC-adducted template/primer complex 17. Reaction conditions: 250 μM dNTP (each), 200 nM complex 17, 40 nM DNA polymerase eta, 25 ^{o}C. (b) Plots of results in (a): The percent of original primer remained un-extended during primer remained un-extended during prime extension reaction is plotted as a function of incubation time for the complex 17.

The MC lesion was not bypassed and a 19-mer abortive product (1 nucleotide before lesion) is accumulated with the same rate as the 24-mer product of control template: Closed triangles: non-alkylated control 25-mer; closed squares: adducted template/primer aborted 19-mer. (c) PAGE of primer extension of 2,7-DAM-adducted template/primer complex 19. Reaction conditions: as in (a). (d) Plots of results from (c): The percent of original primer remained un-extended during the primer extension reaction is plotted as a function of incubation time for the complex 19 using the data presented in figure 54 (c). The 2,7-DAM lesion (open triangles) is bypassed with a slower rate as compared with the control undamaged template (open circles).

These results would be the basis for further structural investigations of complexes between eta DNA polymerase and substrates containing adducted templates with these two structurally different drugs. In advance, the molecular modeling studies using the Dpo4 DNA polymerase (39) in complex with the constructs used for primer extension and single nucleotide incorporation kinetics may bring new structural basis for the observed differences in the translesion synthesis of the two studied monoadducts.

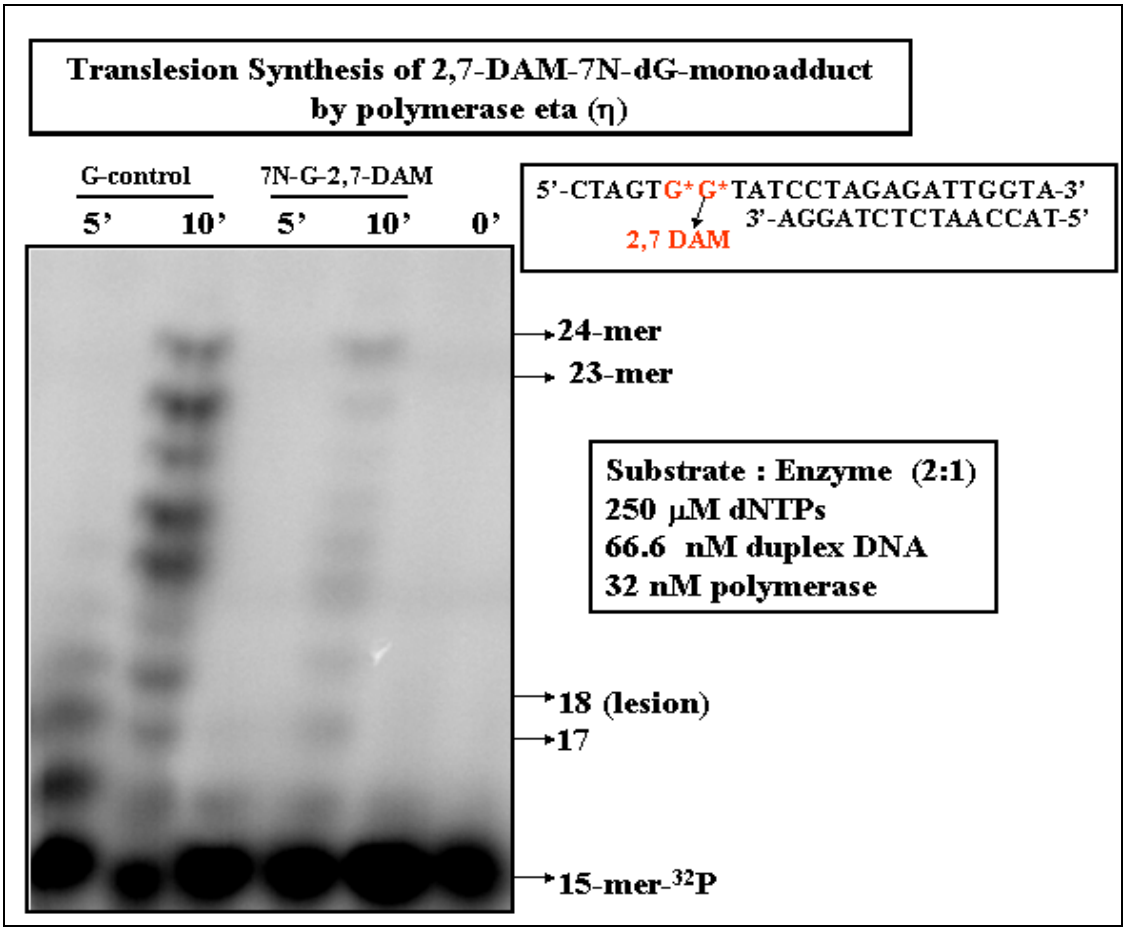


Figure 55: PAGE analysis of primer extension on 2,7-DAM-adducted template (construct 19-scheme 4) by DNA polymerase eta. The conditions for running the translesion synthesis (TLS) are inserted within the same figure. At smaller substrate: enzyme ratios (such as 2:1 or 1:1) the N7-G*-2,7-DAM lesion is more easily bypassed by polymerase eta which showed in general a lower efficiency of bypassisng this bulky adduct (see figure 54 c).

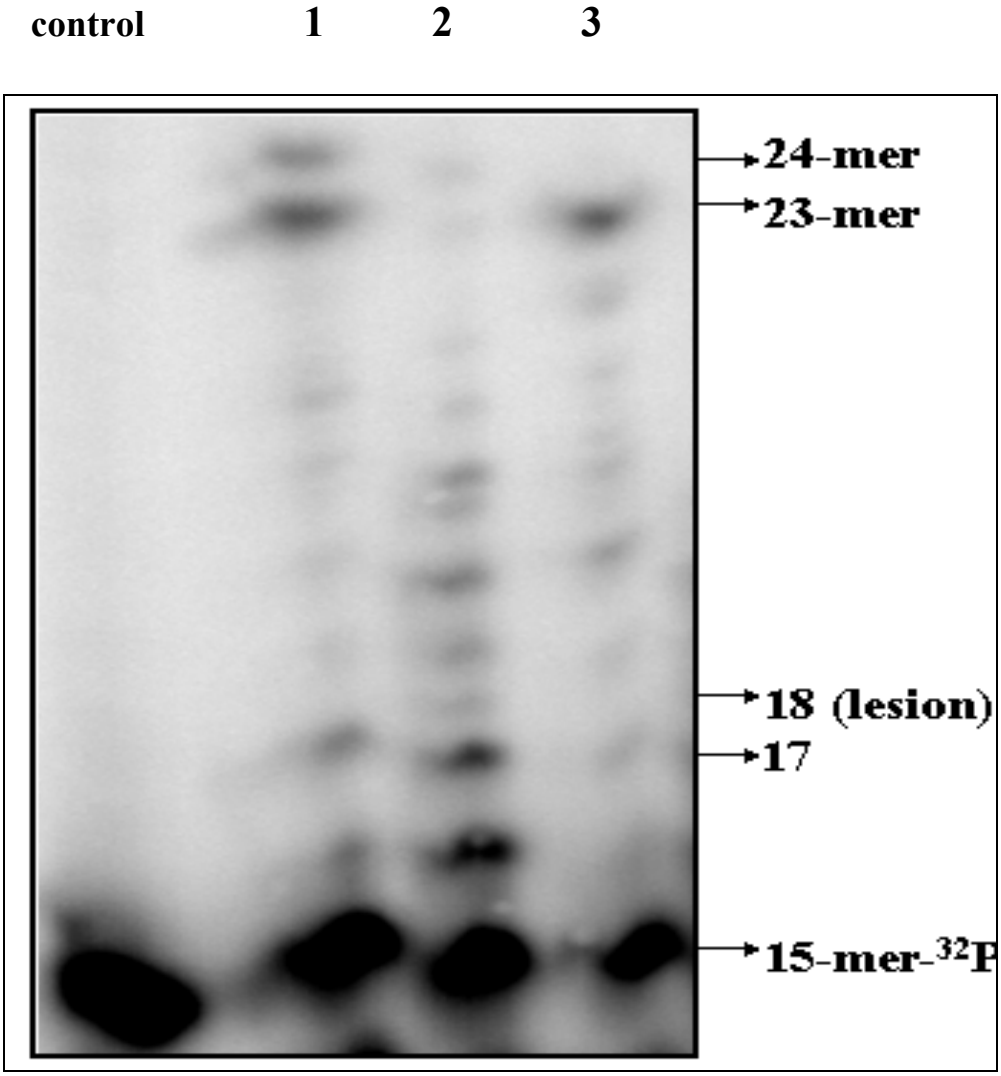


Figure 56: PAGE analysis of primer extension on 2,7-DAM-adducted template of complex *19* by T7 exo$^-$, Klenow exo$^-$ and eta DNA polymerases. The translesion synthesis (TLS) were run at 250 μM final concentration of each dNTP, 89 nM final duplex DNA concentration and a substrate to enzyme (S:E) ratio of 3:1 for 20 minutes. The lower substrate: enzyme ratio increased the efficiency of N7-G-2,7-DAM adduct bypass by all three DNA polymerases: 1- Klenow exo- polymerase; 2- T7-exo- polymerase; 3- eta polymerase.

In the MC-dG-N^2 complex **17**, polymerase η extended the primer strand only to 19-mer, one nt before the MC adduct **7** in the template strand, at relatively low efficiency

as compared to control primer extension (Figure 54 a). In contrast, in the 2,7-DAM-dG-N7 complex **19**, the primer strand was extended two nts past the 2,7-DAM adduct **8**, to 20-mer, with surprisingly high efficiency. A product corresponding to the fully extended primer is formed only to a low extent (Figure 54 b). In these reactions, a ratio of 5:1 (S:E) was employed. In analogous studies of other bulky adducts, polymerase η showed similarly low efficiencies of primer extension under such conditions (46-48). However, using a lower (2:1) or 3:1 S:E ratios, we observed substantially more efficient TLS and the extension to 23 nt length of the primer (figure 55 and figure 56 respectively).

2.3.6.4. Primer Extension kinetics (TLS) by Klenow + DNA polymerase.

The primer extension reaction of adducted N7-G-2,7-DAM templates was performed at high concentration of dNTPs (1mM each) when the lesion was bypassed by Klenow exo+ with the production of fully extended primers (24 mers for construct 23-scheme 4). However, we observed the degradation of the alkylated template as compared with the control non-alkylated template suggesting that the lesion may induce improper alignment of the primer with the template and misspaired base pairs may appear opposite the guanine alkylated with 2,7-DAM.

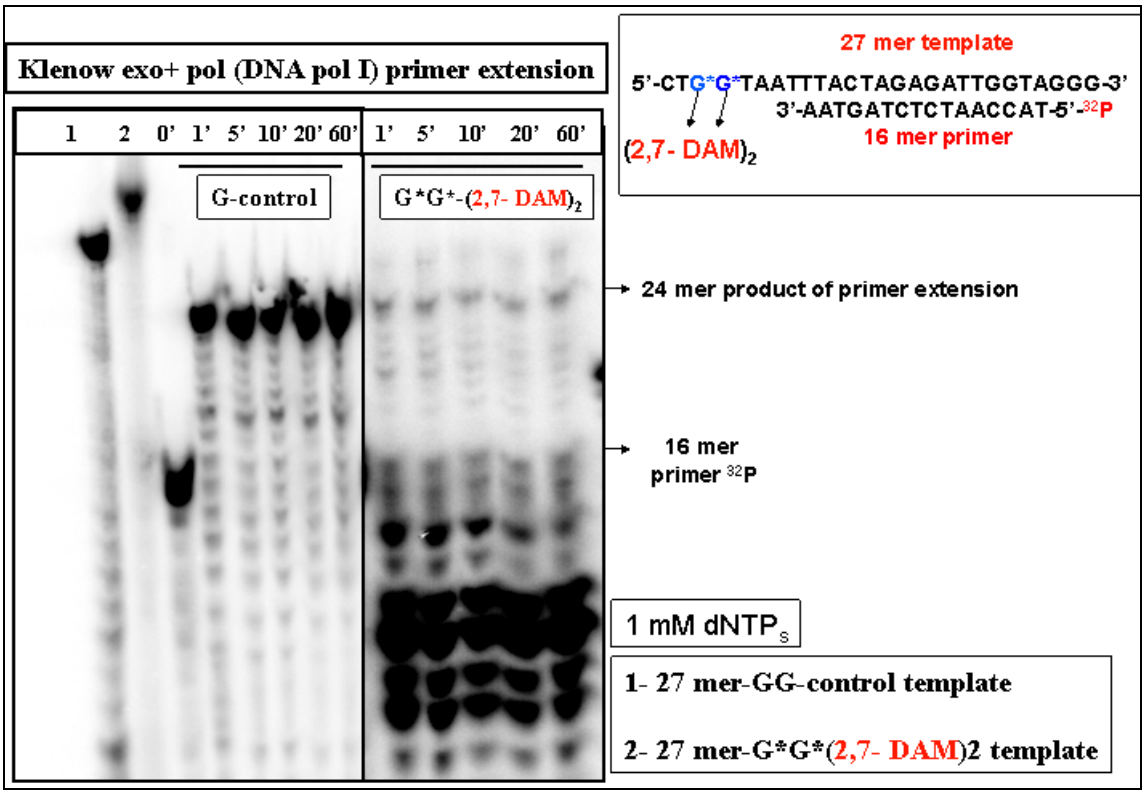


Figure 57: PAGE analysis of primer extension on 2,7-DAM-adducted template by Klenow exo$^+$ DNA polymerases at 1 mM final dNTPs concentration. The translesion synthesis pass the N7-G-2,7-DAM lesion was obtained only on higher dNTP concentration (> 1mM each dNTP). At low concentration of dNTPs (below 500 μM) we didn't observed any translesion synthesis (see figure 3, APPENDIX VII).

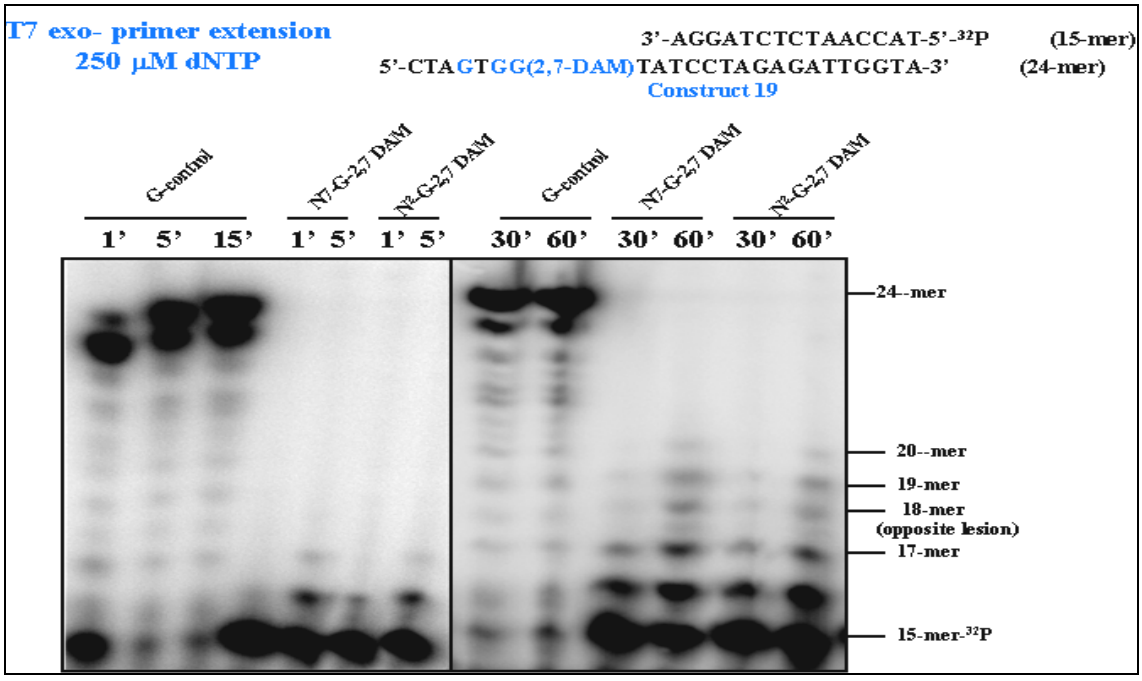
Additional primer extension reaction using all constructs presented in scheme 4 and all four DNA polymerases (T7 exo-, Klenow exo-, Klenow exo+ and eta) are presented in APPENDIX VII-Figures 87-89.

2.3.6.5. Comparison of translesion synthesis (TLS) by Klenow exo- and T7 exo- DNA polymerases on adducted templates containing the N7-dG- and N^2-dG-2,7-DAM.

The translesion synthesis (TLS) on adducted templates containing the adducts N7-dG-2,7-DAM and N^2-dG-2,7-DAM were performed on constructs containing the N7- dG-2,7-DAM (construct 19-scheme 4) and on independent constructs containing N^2-dG-2,7-

DAM monoadduct obtained through ligation of small 12-mer oligonucleotides into 24 mer templates as described in Materials and Methods. Two polymerases were shown the TLS: T7 exo- and Klenow exo- polymerase (figures 58-60). TLS performed by polymerase eta on templates adducted with N^2-dG-2,7-DAM is under current investigation. We don't have yet the strong evidences for the position of N^2-dG-2,7-DAM within the templates, thus we can propose just a global interpretation for the rate of primer extension reaction observed in the case of the templates adducted with N^2-dG-2,7-DAM. We have however some preliminary data supporting the fact that the adduct at N2-dG is almost 50% distributed in between the adjacent guanines. These preliminary results for the mapping of the N2-dG-2,7-DAM position and the results from primer extension reactions presented in figure 58 would suggest that the N2-dG-2,7-DAM lesion is also bypassed, but with slower rate and less efficiently than the N7-G-2,7-DAM lesion (based on the %primer remained unextended during primer extension reaction).

(A)



(B)

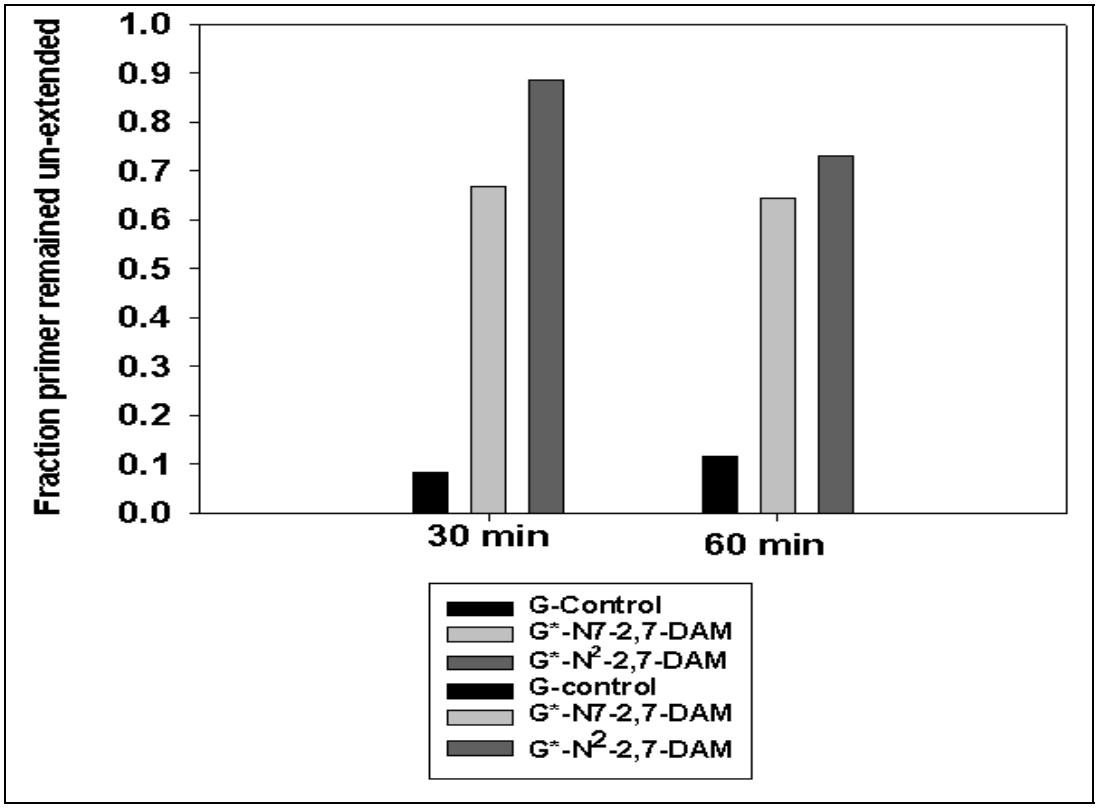


Figure 58: The primer extension reactions on templates adducted with N7-dG- and N^2-2,7-DAM by T7 exo- polymerase. The primer extension reaction was performed in vitro using the assay described in Materials and Methods for primer extension reaction for the construct template/primer-19 at 250 μM of each dNTP. (B) Quantification of gels in figure 56 (A) shows that at two different time points the fraction of primer remained un-extended in the case of N^2-dG-2,7-DAM adduct is higher than in the case of N7-dG-2,7-DAM adduct. These results suggest that N^2-dG-2,7-DAM adduct is decreasing the affinity of T7 exo- polymerase for the substrate more than the N7-dG-2,7-DAM adduct does, explaining the slower rate of primer extension observed for the N^2-dG-2,7-DAM lesion than for the N7-G-2,7-DAM lesion .

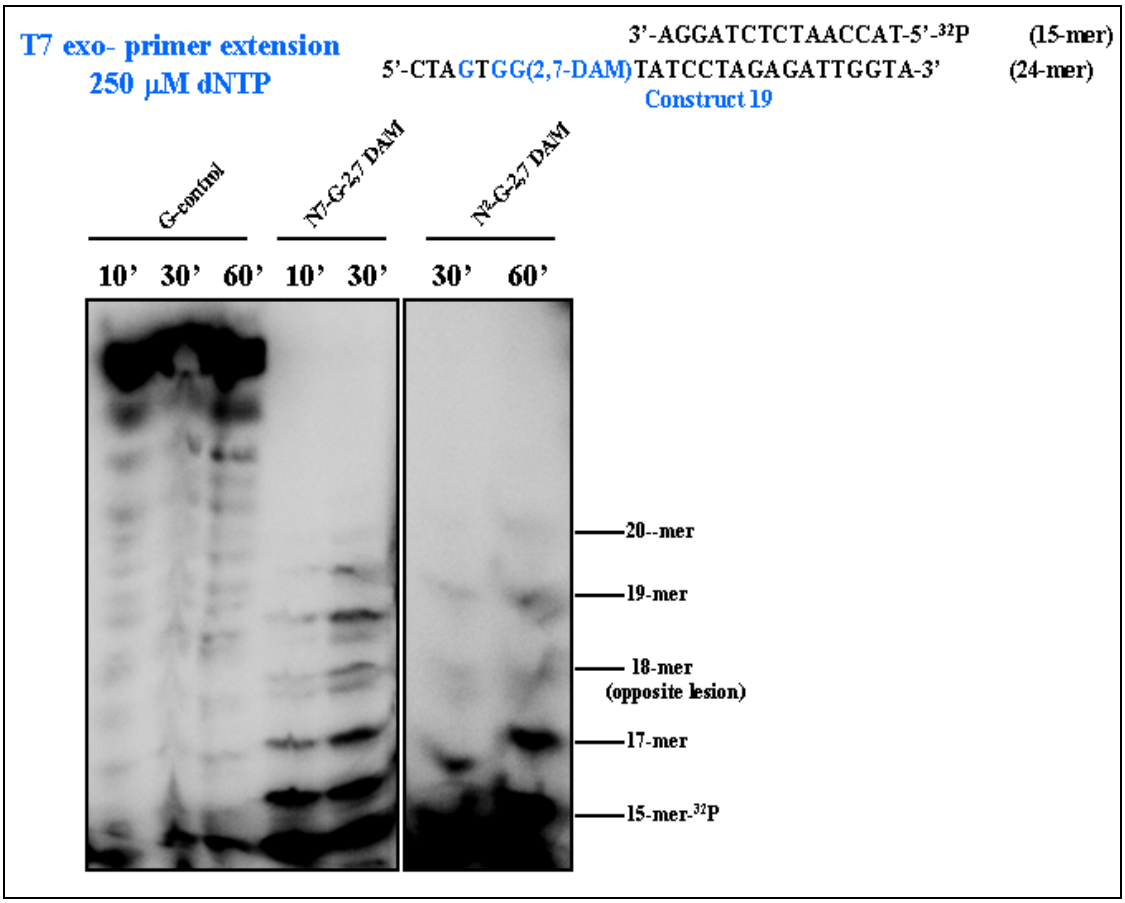


Figure 59: The primer extension reactions on templates adducted with N7-dG- and N^2-2,7-DAM by T7 exo- polymerase (independent experiment). The primer extension reaction was performed in vitro using the assay described in Materials and Methods for primer extension reaction for the construct template/primer-19 at 250 μM of each dNTP.

The fraction of primer remained un-extended in the case of N^2-dG-2,7-DAM adduct is higher than in the case of N7-dG-2,7-DAM adduct. These results suggest that N^2-dG-2,7-DAM adduct is decreasing the affinity of T7 exo- polymerase for the substrate more than the N7-dG-2,7-DAM adduct does which is reflected in a less effcicient primer extension reaction for the -dG-2,7-DAM lesion. Figure 60 shows the primer extension reaction performed with Klenow exo- DNA polymerase were it can be seen that the N^2-dG-2,7-DAM lesion is even more less efficient replicated than the N7-G-2,7-DAM lesion.

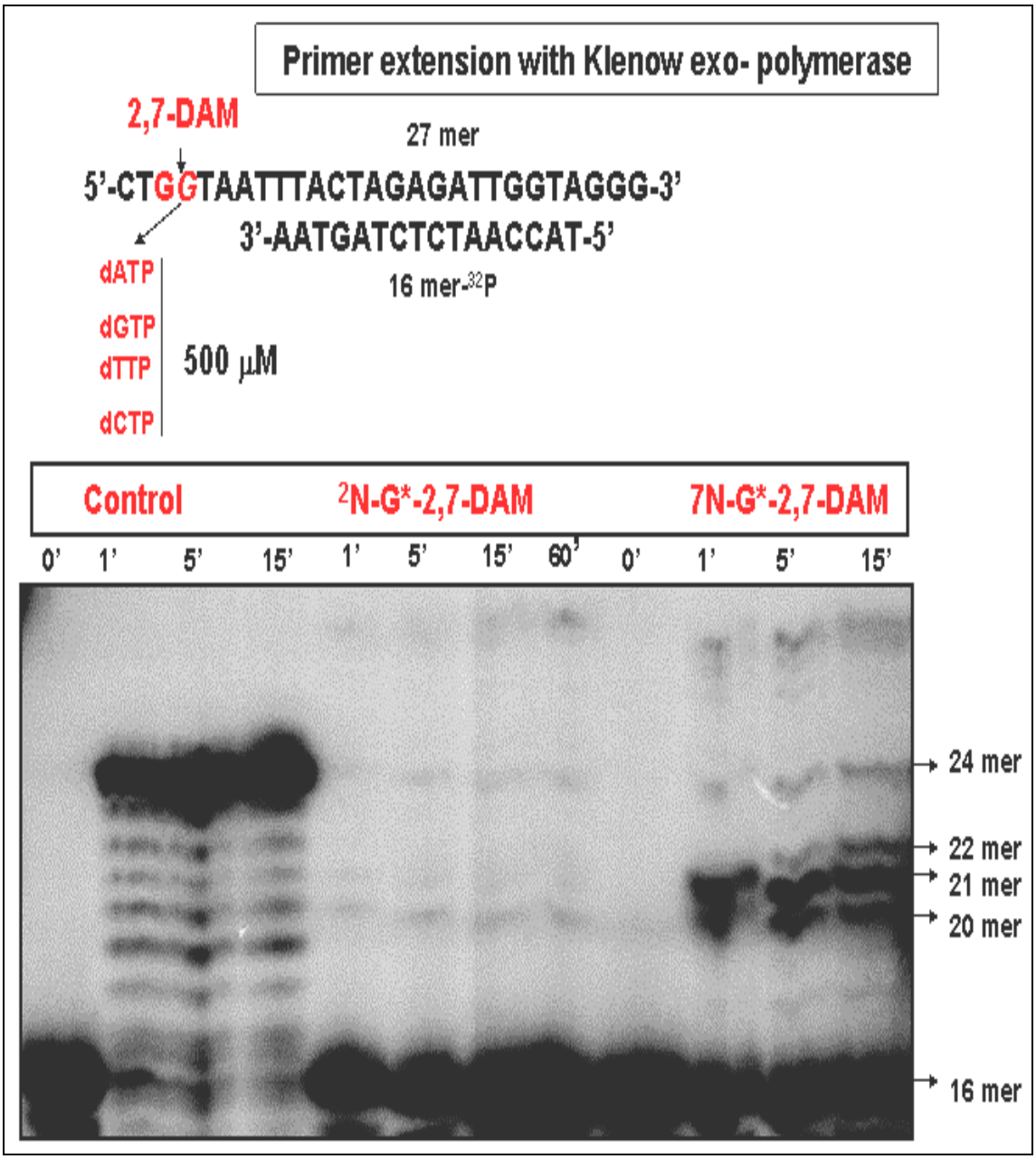
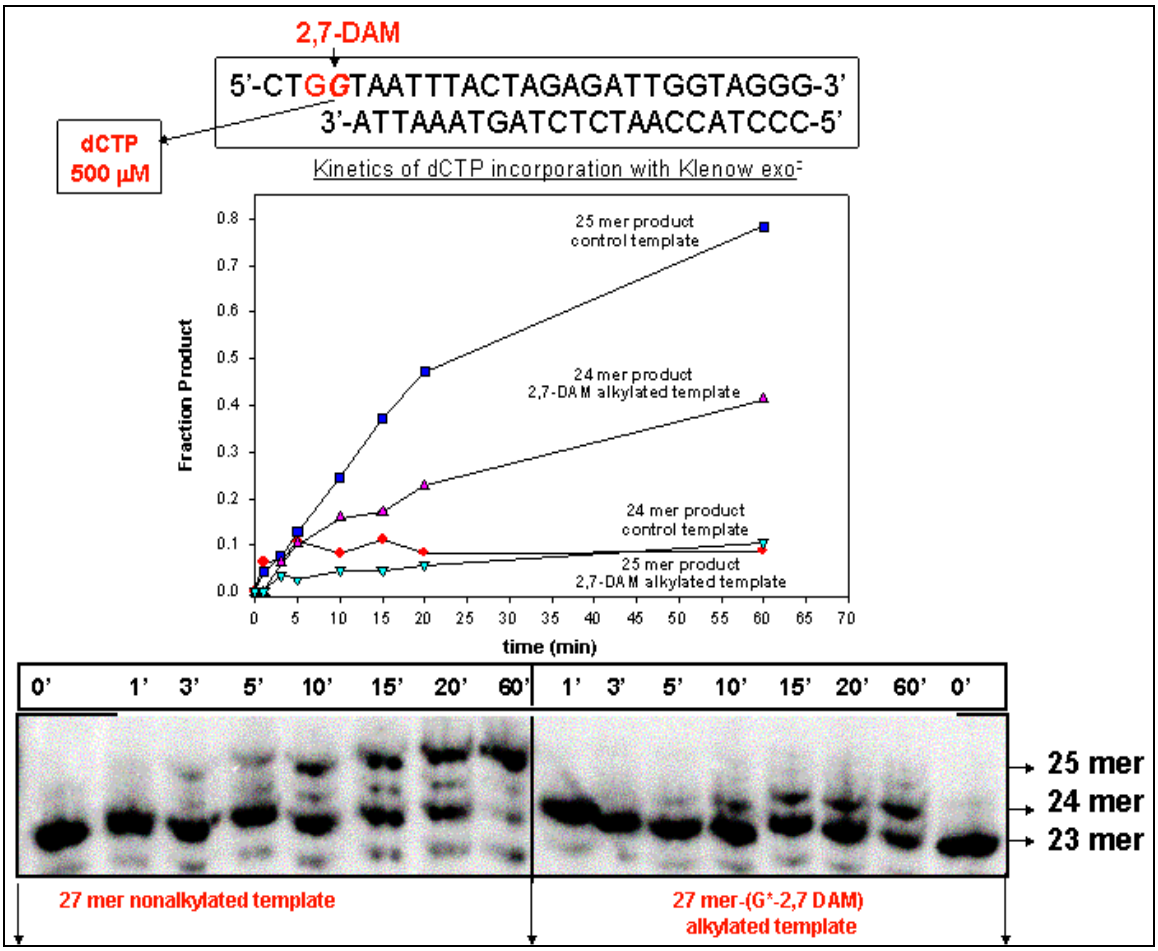


Figure 60: The primer extension reactions on templates adducted with N7-dG- and N^2-2,7-DAM by Klenow exo- polymerase. The primer extension reaction was performed in vitro using the assay described in Materials and Methods for primer extension reaction for the construct template/primer-19 at 250 μM of each dNTP. The fraction of primer remained un-extended in the case of N^2-dG-2,7-DAM adduct is higher than in the case of N7-dG-2,7-DAM adduct.

2.3.7. Single nucleotide incorporation opposite 2,7-DAM-dG-N7 Adduct.

These experiments were designed to test the fidelity of the translesion bypass observed in the case of the 2,7-DAM-adducted (**8**) (scheme 3) templates, described in scheme 4. The alkylated primer-template complex **21** and the control nonalkylated complex were incubated with only a single dNTP at a time at high concentration (500 μM) and Klenow (exo-) DNA polymerase.

(A)



(B)

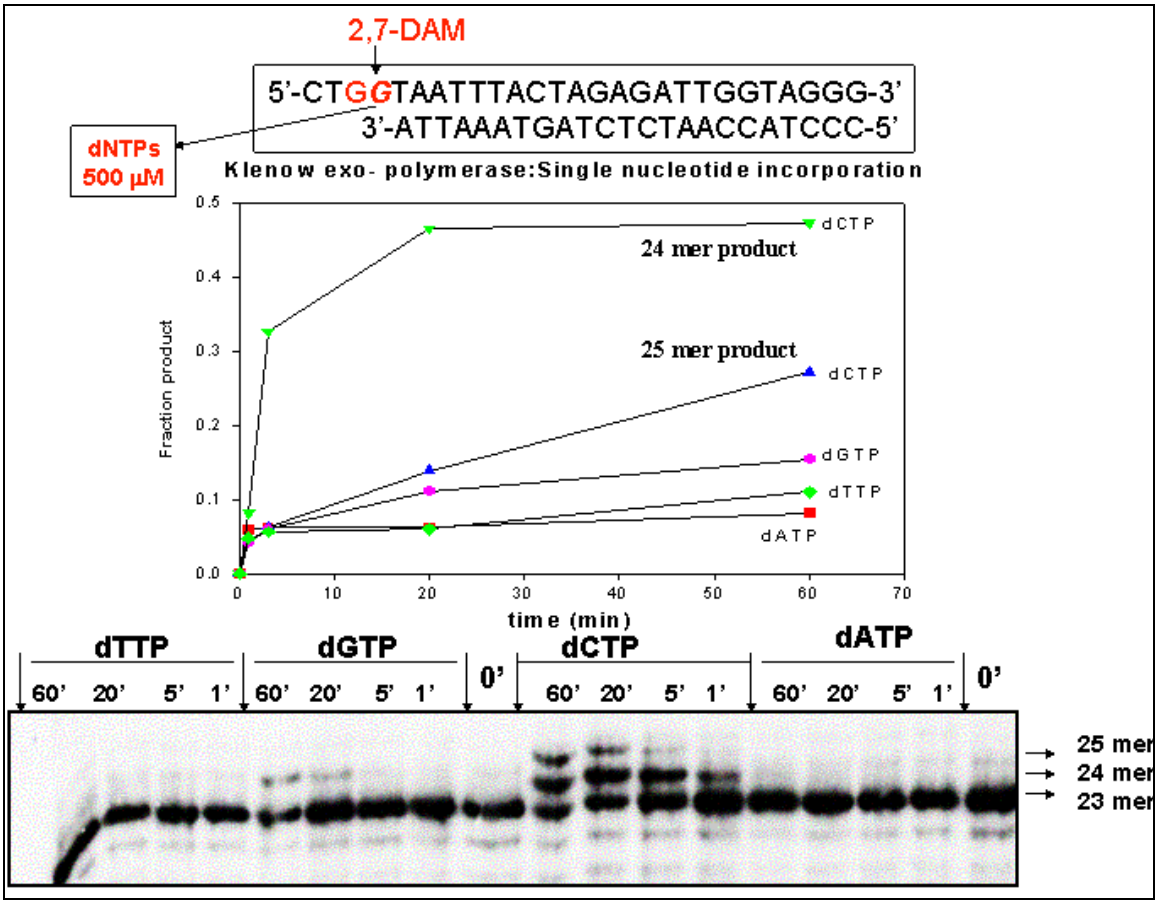


Figure 61: Single-nucleotide incorporation kinetics opposite the 2,7-DAM-dG-N7 lesion. dCTP was having the highest frequency of incorporation both opposite 2,7-DAM alkylated guanine in the template. The kinetics for dCTP incorporation opposite the lesion shows that the second dCTP incorporation opposite G beyond the lesion in the template has a slower rate than in the case of control, non-alkylated template (B). These results are consistent with primer extension kinetics which showed the loss of processivity for polymerase as is approaching the lesion.

The control non-alkylated complex incorporated two C units into the primer strand (25-mer). A small amount of primer intermediate (24-mer) incorporating one single C unit was observed at the end of the 60-min incubation period Figure 61 (A). No appreciable incorporation of the other 3 nucleotides was observed (data not shown). The alkylated complex 21 incorporated two C units at a lower overall efficiency than the

control (figure 61 (B)). The 24-mer one-C-incorporated intermediate was accumulated in a two-fold proportion relative to the 25-mer two-C-product, during the 60 min incubation (figure 61 B). Figure 61 (B) shows that the nucleotides A and T were not incorporated. The base G was incorporated to a relatively low extent opposite the G(4) position of the template.

The important conclusion of these results is that the 2,7-DAM-guanine-N7 adduct is not mutagenic in the Klenow (exo-) DNA polymerase replication system.

As in the case of primer extension reaction, we didn't observed any single nucleotide incorporation events at low concentration of dNTP (100 μM) (figure 62).

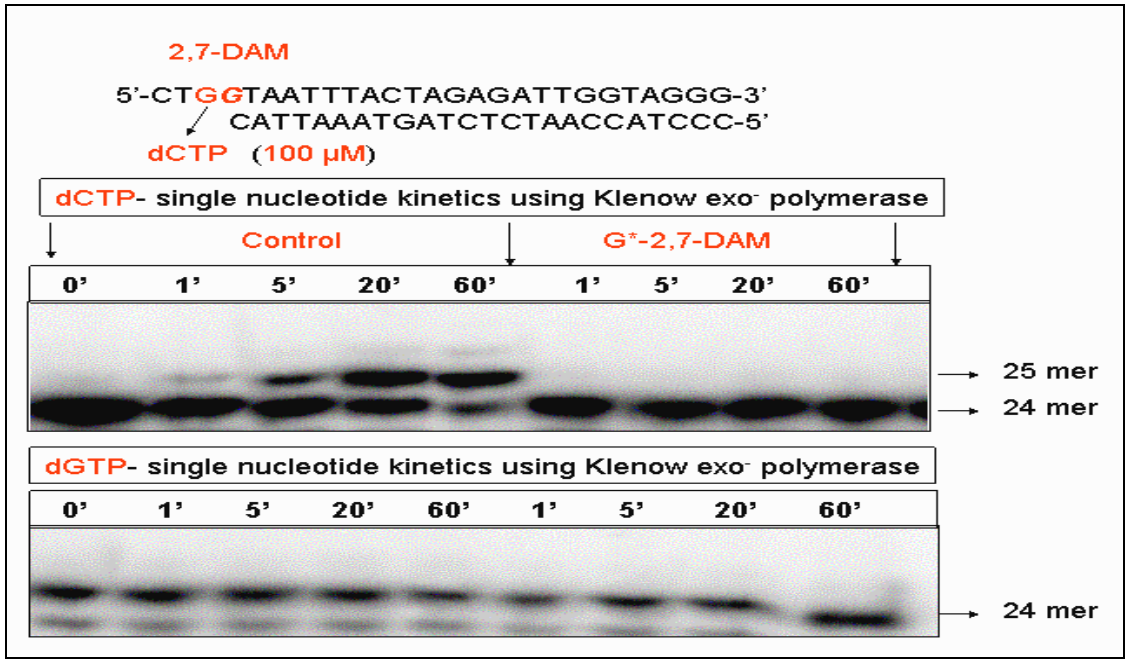


Figure 62: Single nucleotide kinetics for dCTP and dGTP incorporation opposite the N7-G-2,7-DAM lesion by Klenow exo- polymerase at lower concentration of dNTPs (100 μM). At lower dCTP concentration (i.e. below 300 μM) there is no single nucleotide incorporation opposite the lesion N7-G*-2,7DAM, suggesting that the 2,7-DAM drug is decreasing the relative affinity for the dCTP, i.e. the Km for dCTP is increased. In the case of dGTP, there was no incorporation of nucleotide opposite the lesion at 100 μM.

In Figure 63, the lesion N7-G-2,7-DAM was bypassed with lower efficiency and slower rate than the control nonalkylated template, supporting the data presented in Figure 61 (b). The dNTP with the highest efficiency of incorporation one nucleotide beyond the lesion was dCTP. dATP, dGTP and dTTP were not incorporated beyond the lesion with high efficiency (data not shown) confirming the lack of mutagenicity (in vitro) of the N7-G-2,7-DAM monoadduct. Altogether, the results from figures 61 (b) and 63 supports the view that N7-G-2,7-DAM is not an easily bypassed lesion but is definitely non-mutagenic when the lesion bypass reactions are performed with replicative DNA polymerases (such as T7exo- DNA polymerase) or with Klenow exo-.

By contrary, when the lesion bypassing experiments are performed in vitro with DNA polymerases which are known to bypass the lesions with high error during DNA synthesis, such as DNA polymerase eta, the N7-lesion is prone to be mutagenic, due to the lack of high fidelity of bypass of DNA-polymerase eta. As figure 64 shows, DNA polymerase eta incorporated with high efficiency dGTP opposite the N7-G-2,7-DAM lesion but with lower efficiency than in the control nonalkylated template. dGTP was incorporated opposite dGTP in the control nonalkylated template, but with lower efficiency than dCTP, the normal expected dNTP to be incorporated opposite dGTP.

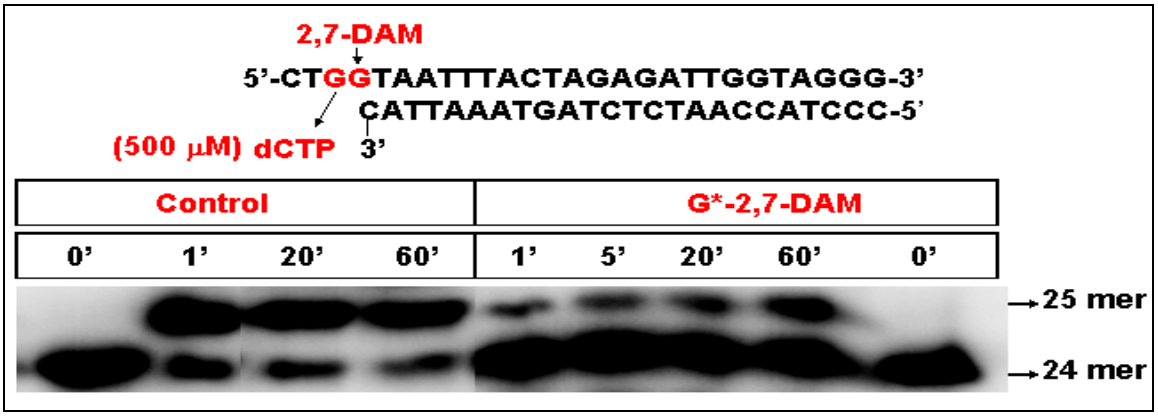
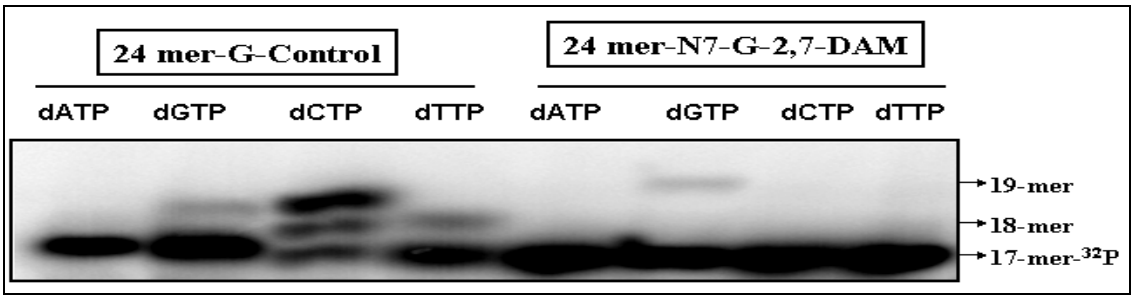


Figure 63: Single-nucleotide incorporation of dCTP one nucleotide beyond the N7-G*-2,7-DAM lesion by Klenow exo- polymerase. dCTP had the highest frequency of incorporation beyond 2,7-DAM alkylated guanine in the template construct . The rate of dCTP incorporation one nucleotide beyond lesion is slower than in the control, non-alkylated template, given rise to a lower yield of expected extended product (as compared with control). No appreciable amounts of dATP, dGTP and dTTP were observed to be incorporated in this primer/template construct.



Δt = 30 minutes
66.65 nM primer
100 nM template
32 nM polymerase eta
[Substrate:Enzyme] ratio = 2:1

3'-ATAGGATCTCTAACCAT-5'-^{32}P (17-mer)
5'-CTAGTGG(2,7-DAM)TATCCTAGAGATTGGTA-3' (24-mer)
Construct 22

Figure 64: Single nucleotide incorporation opposite N7-dG-2,7-DAM lesion by DNA eta polymerase. The conditions for running the single nucleotide incorporation are described in the insert below figure 64. The control experiment showed that polymerase eta is error-prone in incorporation one single dNTP at a time, with the following order of frequency of dNTPS incorporation: dCTP>> dTTP>dGTP>>>>dATP (in agreement with other published pol eta single nucleotide incorporation experiments (40-50). However, in the case of N7-dG-2,7-DAM adducted template the only dNTP incorporated opposite the lesion is dGTP.

2.3.8. Single nucleotide incorporation Opposite MC-dG-N^2 Adduct.

Despite the previous studies which showed that MC is non-mutagenic both in vitro and in vivo (17-Chapter 1) our last studies of single nucleotide incorporation opposite the N^2-dG-MC lesion showed that at higher concentrations of dNTPs, 500 μM, there is a nonspecific incorporation of all dNTPs opposite this lesion contained within a 36 mer template (figure 65). At 100 μM there is no dNTP incorporated opposite the N^2-dG-MC lesion (figure 66), results which are in agreement with the translesion synthesis were the extension of the primer is blocked o nucleotide before the lesion (figure 41). An independent experiment run at 100 μM of each dNTP but with a 24 mer-N^2-dG-MC lesion construct we observed a preference for dATP incorporation opposite this lesion. The order of dNTPs incorporation in this case was dATP>dCTP>dGTP>dTTP (figure 67).

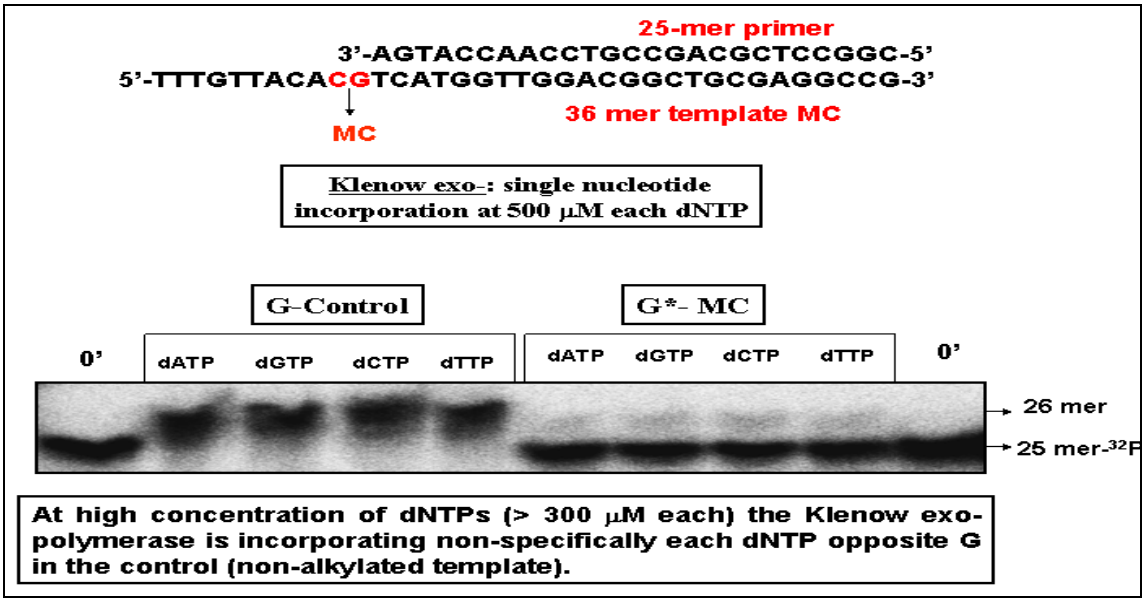


Figure 65: Single nucleotide incorporation opposite the G*-MC lesion using Klenow exo- polymerase and 500 μM each dNTP. The reaction of single nucleotide incorporation were performed for 30 minutes using the conditions described in materials and methods.

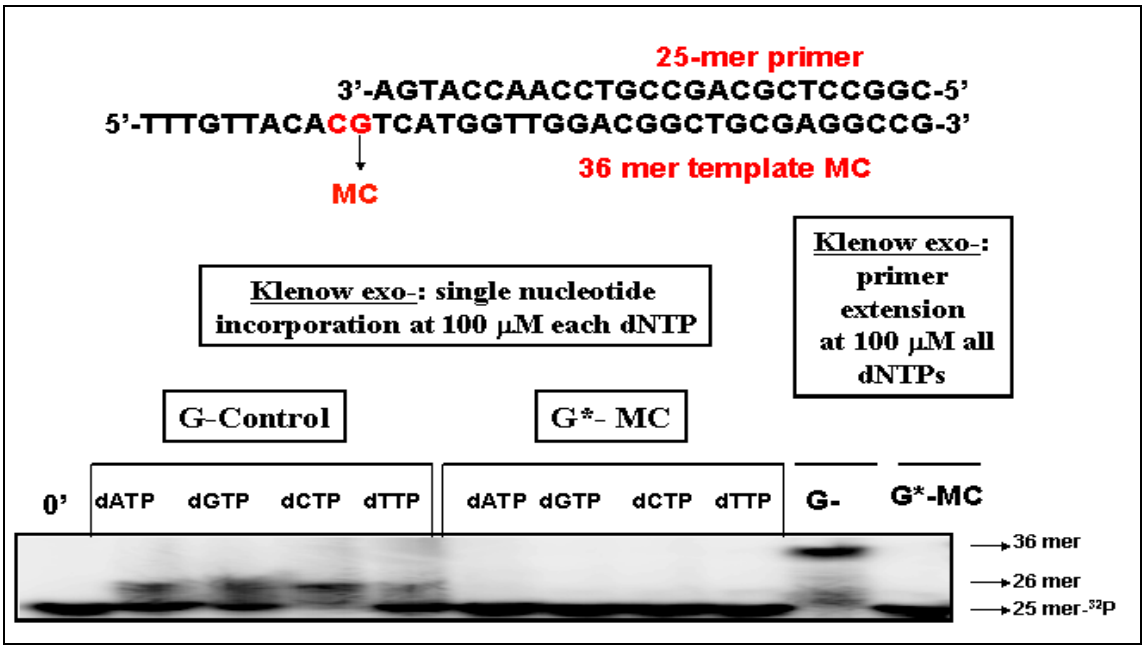


Figure 66: Single nucleotide incorporation opposite the G*-MC lesion using Klenow exo- polymerase and 100 μM each dNTP. The reaction of single nucleotide incorporation were performed for 30 minutes using the conditions described in materials and methods.

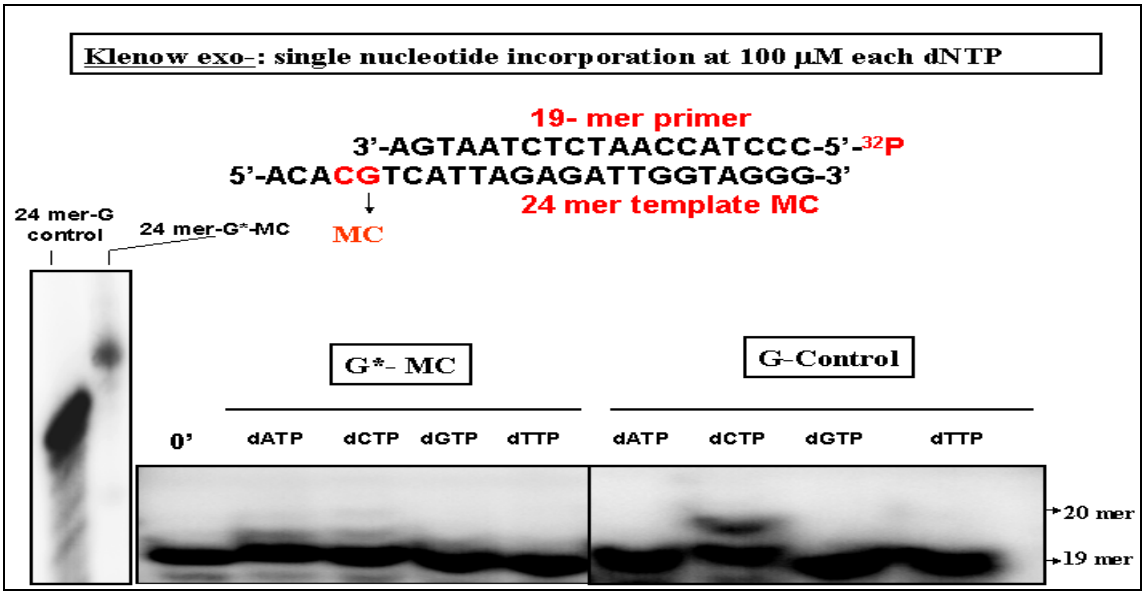
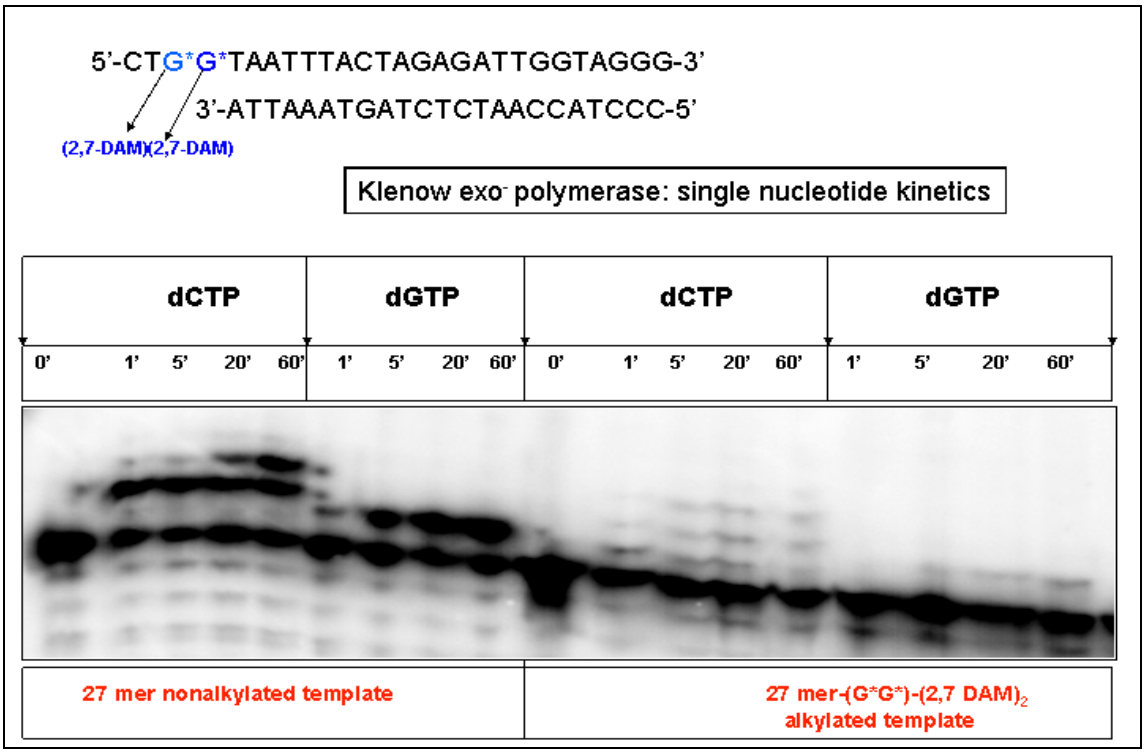
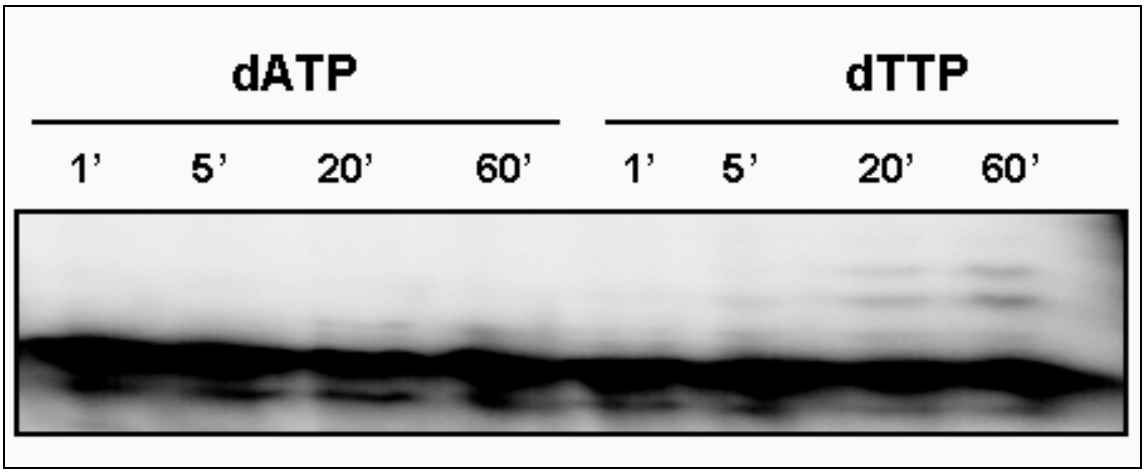


Figure 67: Single nucleotide incorporation opposite the G*-MC lesion using Klenow exo- polymerase and 100 μM each dNTP. The reaction of single nucleotide incorporation were performed for 30 minutes using the conditions described in Materials and Methods.

Since we always obtained double-2,7-DAM alkylated (at 2 Gs) 12 mer constructs (see scheme 3) we generated trhough standard ligation experiments 27-mer templates (see scheme 4) containing 2 Gs alkylated at N7 position with 2,7-DAM. The presence of the double alkylated 12 mer was determined by ESI-MS (-) mode and by Maxam Gilbert hot piperideine treatment for G-line as explained in Materials and Methods (Section 2.2.3.4). We conducted single nucleotide kinetics of dNTP incorporation opposite the double alkylated lesion (see constructs in scheme 3 and in figure 68) and we determined that the double adduct N7-G-N7G-(2,7-DAM)2 is mutagenic. The frequency of each dNTP incorporation opposite the double (N7-2,7 DAM-G-)$_2$ lesion was: dTTP > dCTP > dGTP>>> dATP (figure 68 (A) and (B) (independent experiment)).

(A)





(B)

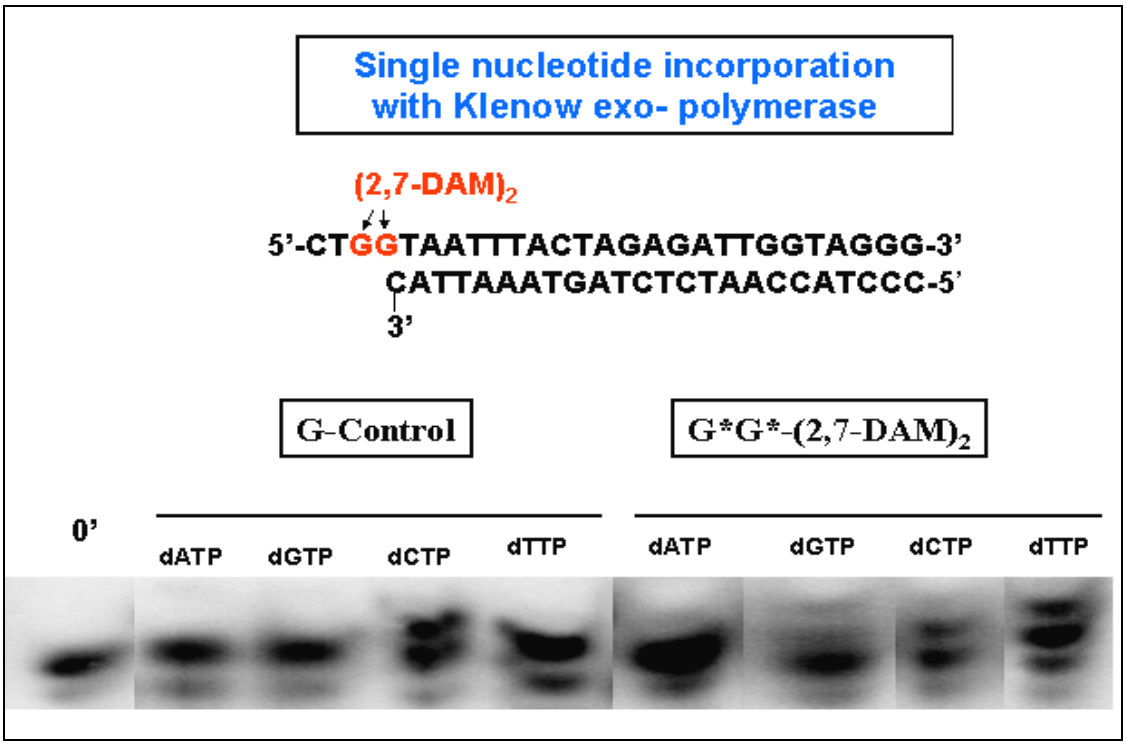


Figure 68: Single-nucleotide incorporation kinetics opposite the (G*-N7-2,7-DAM)$_2$ lesion (construct 24)using Klenow exo- polymerase and 500 μM each dNTP. The frequency of incorporation opposite the double (N7-2,7 DAM-G-)$_2$ lesion was:

dTTP > dCTP > dGTP>>> dATP (A) and (B) (independent experiment).

DISCUSSION

The major goal of the research addressed the relationship between the structure of Mitomycin C (MC) and 2,7-Diaminomitosene (2,7-DAM) monoadducts and their interaction with polymerases in an *in vitro* polymerase assay system. Specifically, we investigated the effect of MC and 2,7-DAM adducted DNA template upon the kinetics of DNA polymerases in an *in vitro* system. Initial cell biology experiments showed higher rate of survival for the cells treated with 2,7-DAM as compared with the cells treated with MC, suggesting that mitomycin C (MC) was a very cytotoxic drug while 2,7-DAM is non cytotoxic (1-20-Chapter 1) .

Based on these experimental evidences we initially hypothesized that a primer extension reaction performed *in vitro* just in the presence of the alkylated DNA templates/primer substrates and the DNA polymerase (no other proteins from the replication machinery involved) is expected to result in a translesion bypass of the 2,7-DAM lesion and a stop of the polymerase before the lesion containing guanine alkylated with mitomycin C.

Initially, during the first stages of this research, steady-state kinetics were performed using 10 fold molar excess of DNA substrates (annealed primer/template) vs. enzyme concentration and 500 μM of each dNTP in order to determine if the lesion in the template containing 2,7DAM is bypassed by the DNA polymerases. Primer extension experiments were conducted with 27 mer and 24 mer construct templates having one guanine (G) alkylated either with MC or with 2,7-DAM were used together with control non-alkylated templates. In addition, templates containing 2 guanines both alkylated with 2,7 DAM were also used to investigate the lesion bypass by polymerases.

We investigated first the translesion synthesis of the N7-G-2,7-DAM and N^2-dG-2,7-DAM adducted templates by DNA replicative polymerases T7 exo- and Klenow exo- after we confirmed that the templates were homogeneosly pure by PAGE analysis (see Figures 31-34). We showed for the first time that both T7 exo- and Klenow exo- DNA polymerase are bypassing the N7-G-2,7-DAM bulky lesion (figures 35-52). The quantitative analysis of primer extended showed that the rate of bypass was slower in the case of templates containing alkylated guanines as compared with the non-alkylated templates. The results showed that in the case of templates containing N7-G-2,7-DAM there is an accumulation of shorter length DNA products during the primer extension and translesion bypass. 17, 18, 20, 21, and 22 mers are accumulated in the reaction such that the final yield for the expected product (24 mer+/-25 mer) is just 5-10 % in the case of the 24 mer templates (figures 36, 38, 40, 41, 48). These shorter DNA products correspond to the positions one nucleotide before lesion, one nucleotide opposite the lesion and one and two nucleotides beyond the lesion. We are proposing that the DNA polymerases are losing the processivity as they are approaching the lesion, a phenomenon that may explain the accumulation of shorter length DNA products (also reported in other primer extension reactions of bulky DNA-adducts (1-50).

In addition, the primer extension reaction for the 27 template was accompanied by over-extension with 5 more nucleotides beyond the expected length of 27 mer, suggesting that there may be a slippage of the primer during the extension reaction, primarily due to the presence of the drug very near the 5'-end of the alkylated template (figures 37-42).

It is important to mention that both replicative DNA polymerases had very good kinetics of primer extension with control substrates, suggesting that there was no error in

the polymerase assay. In most cases the % primer extension was between (75-85)% in the case of control in all of the *in vitro* assays for primer extension.

All the primer extension kinetics were performed in high excess of each dNTP (300 μM and 500 μM) and there was no primer extension at lower dNTPs concentration (100 μM dNTPs). Since the Km for each dNTP is around 50 μM if normal (non-alkylated) templates are used in all polymerase reactions (2-5), the above results suggest that Km for the dNTPs is highly increased in the case of the N7-2,7-DAM (G)-alkylated template, thus the DNA polymerases have lower affinity for these chemically alkylated substrates. For most replicative DNA polymerases, the loss in affinity for the substrates is correlated many cases with the lost of processivity (1-20).

The mitomycin C (MC) monoadducts are in general not bypassed as was the case of both 24 mer-G*(MC) and 36 mer-G*(MC) adducted templates. Thus DNA polymerase (Klenow fragment (exo-) was blocked one nucleotide before the alkylated G*(MC) in the case of 24 mer-G*(MC) template, producing high yield (80%) of abortive product 19 mer (the mitomycin C lesion has the 20th position) (figure 41). In the case of the 36 mer-**G**(MC) the DNA polymerase (Klenow fragment, exo-) was blocked after it incorporated 1 nucleotide opposite the G-N^{2-}alkylated guanine and the rate of incorporation of one single nucleotide opposite the lesion was also higher than in the case of the control (non-alkylated) template (figure 53). Independently than Klenow exo-, T7 exo- polymerase was also inhibited in bypassing the N^2-dG-MC lesion 1 nucleotide before lesion, but at a slower rate than in the control experiments (figure 41). Thus the mitomycin C monoadducts are a strong block for the replication, a result consistent with already published data which are shown the highly cytotoxic effect of mitomycin C (in mammalian cancer cells) (10, 11-Chapter 1). The mitomycin G-^{2}N-monoadduct was

shown independently to be a block of replication in an in vitro system using primer/mitomycin C-^{2}N-G-alkylated template substrates and T7 and Klenow exo^{-} polymerases by other group of researchers (17-Chapter 1).

In the case of the templates containing one single G-N^{2}-alkylated with mitomycin C quantitative analysis of %primer extension data supports the idea that in the case of mitomycin C monoadducts, the rate of primer extension is the same for both the non-alkylated and the alkylated templates. In some cases, the rate for the extension of the G-mitomycin C alkylated template was higher than for the non-alkylated template suggesting a possible higher affinity of DNA polymerase for the alkylated template, making the mitomycin C monoadduct to behave like a "suicide" substrate (figure 41).

The primer extension reaction conducted with DNA polymerase eta showed for the first time that DNA polymerase eta is performs the translesion synthesis of N7-dG-2,7-DAM lesion is true with less efficiency than T7 exo- and Klenow exo-DNA polymerases. Although the structural details and biological function of the translesion bypass type DNA polymerase η differ from those of the T7 and Klenow fragment polymerases (40-48), the observed differential extent of the bypass of the 2,7-DAM-dG-N7 and MC-dG-N^{2} adducts in the cell free system (figures 54-56) follows the same adduct structure-activity pattern obtained using the other two DNA polymerases (9-Chapter 1).

It is known (40-65) that pol eta is an error-prone polymerase, thus it was important to perform single nucleotide incorporation at end-point kinetics opposite N7-dG-2,7-DAM monoadduct to test the mutagenicity of this lesion as it is bypassed by DNA polymerase eta . The control experiment showed that polymerase eta is error-prone in incorporation one single dNTP at a time, with the following order of frequency of

dNTPS incorporation: **dCTP>> dTTP>dGTP>>>>dATP** (in agreement with other published pol eta single nucleotide incorporation experiments (46)). However, in the case of N7-dG-2,7-DAM adducted template, the only dNTP observed as being incorporated opposite the lesion is dGTP, with a lower efficiency than in the control template, probably due to the presence of the drug in the G-template opposite the incoming dNTP which could allow favorable hydrogen-bonds or electrostatic interactions between the drug and the incoming dNTP (10-15).

An interesting case was the translesion bypass obtained with Klenow exo + polymerase during TLS of $(N7\text{-}dG\text{-}2,7\text{-}DAM)_2$ monoadducts. At lower concentration of dNTP (< 1 mM) the polymerase was not bypassing the lesion. In addition, extended primers were degraded in shorter DNA fragments, suggesting that the extended primer does not form the right hydrogen bonding between the incoming dNTP and the base alkylated in the template. This misspaired primer-template substrate would allow dNTP missincorporation events which are detected by the Klenow exo+ editing function resulting in degradation of the wrong extended primer (figure 89, APPENDIX VII). At high concentration of dNTP (1 mM) we observed that the lesion N7-dG-2,7-DAM was bypassed resulting in fully extended primer (figure 57). The Klenow exo+ activity toward substrates containing N7-2,7-DAM alkylated templates suggests structural deficiencies in annealed primer/template complex such as nucleotide missincorporation events at the end of the extended primer during the replication or the slippage of primer.

Each of the four polymerases had a different extension efficiency: Klenow exo- had the highest efficiency of bypassing the N7-G-2,7-DAM lesion followed by T7 exo- and eta polymerases, with the lowest efficiency being observed for Klenow exo+. In all

cases substrate: enzyme ratios (S:E) of 1: 1, 2:1 and 3:1 were allowing a higher efficiency for translesion synthesis of both bulky adducts than S:E ratios of 5:1 or higher.

Additional single-nucleotide kinetics using ^{32}(P)-primers with the free 3'-OH end stopping one nucleotide before the lesion or having at the 3'-OH a cytosine base-pairing with the opposite guanine-2,7-DAM lesion in the template were used to investigate the frequency of single nucleotide incorporation opposite the lesion and the frequency of single nucleotide incorporation one nucleotide beyond the lesion. In both cases dCTP had the highest frequency of incorporation opposite 2,7-DAM alkylated guanine in the template. Very low (<1%) frequency of dGTP incorporation opposite the lesion containing 2,7-DAM was also observed. It is important to mention that the single-nucleotide kinetics experiments were performed with Klenow exo$^-$ polymerase which has intrinsically the tendency to incorporate dGTP opposite G in the control non-alkylated template at high dGTP concentrations (>250 μM). No detectable levels for incorporation of dATP or dTTP opposite the 2,7-DAM alkylated guanine in the template were observed during all the single-nucleotide kinetics experiments (figure 61). These results suggest a lack of depurination of the 2,7-DAM (G) alkylated template during the single-nucleotide kinetics. Thus, we can argue that the lesion bypass during primer extension experiments was allowed strictly by the presence of 2,7-DAM in the template and was not due to the presence of an apurinic site. The presence of apurinic sites due to the depurination events (a phenomenon encountered primarily for the lesions containing alkylated guanine at N7 position in the purine ring (50-57)) is totally excluded both in the case of primer extension and single nucleotide kinetics experiments since the apurinic sites are characterized by a high frequency of dATP incorporation opposite the apurinic site, which did not our system (50-57).

In contrast with the single 2,7-DAM-monoadducted template the double 2,7-DAM-monoadduct template (construct **24**) showed translesion synthesis, i.e. the double adduct was bypassed but with the accumulation of overextended products of extension, suggesting that the double adduct impared the right alignment and hydrogen bonding between the template and the primer. The DNA polymerase fidelity of bypassing the double adduct is lower than for the single N7-G-2,7-DAM monoadduct. In single nucleotide incorporation events opposite the double lesion (figure 68) we observed dTTP with the highest frequency of incorporation opposite the lesion suggesting the mutagenic potential of this lesion. The dNTP frequency of incorporation opposite the double (N7-2,7 DAM-G-)$_2$ lesion was: dTTP > dCTP > dGTP>>> dATP .

In the case of kinetics of lesion bypass experiment, a ^{32}(P)-primer having at the 3'-OH a cytosine base-pairing with the opposite guanine-2,7-DAM in the template was annealed with the 27 mer template, and the kinetics for the next dCTP incorporation beyond this lesion was determined. The results suggest a very slow rate for the incorporation of dCTP opposite the next G after the alkylated-G in the template (see figure 63).

In contrast, the G-MC-adducted templates were shown in some cases (see figures 65 and 67) the potential for being mutagenic with the following order of frequency of incorporation of dNTPs opposite the lesion: dATP > dCTP >> dGTP>>> dTTP a result which is in contrast with earlier findings were the T7exo- and Klenow exo- DNA polymerases were used in single nucleotide incorporation assays (17-Chapter 1).

In the field of DNA damage bypass these are one of the few examples which show that the dG-N^2-monoadducts fully inhibit the translesion bypass while the dG-N7-monoadducts are bypassed by the DNA polymerase eta. These results are opening new

questions related to the structure of primer/damaged template and their molecular recognition by eta polymerase.

In order to propose a model that could explain the molecular mechanism of action of mitomycin C and its derivative 2,7-DAM, computation modeling can be performed to allow the docking of the primer / (G*) MC or 2,7-DAM alkylated template into the active site of Klenow fragment, T7 exo- and Dpo4 DNA polymerases (an error-prone translesion synthesis DNA polymerase proposed to be structurally similar with eta polymerase). Since both the N7-G-2,7-DAM and N^2-dG-2,7-DAM monoadducts are bulky, it is of interest to perform the first docking experiments using the Dpo4, which is part of Y family of polymerases, known to be error-prone during translesion synthesis. Y-family DNA polymerases are believed to replace replicative DNA polymerases that are stalled at sites of bulky lesions such as those caused by the binding of polycyclic aromatic hydrocarbon (PAH) metabolites to DNA (46-48). Such adducts, if not repaired, largely block primer extension by replicative DNA polymerases (47- 49) but can be bypassed by the Y-family polymerases (50). Most of members of this latter family lack an intrinsic proofreading exonuclease, exhibit low processivity, and replicate undamaged DNA with low efficiency and fidelity (48-49). Translesion synthesis by Y-family DNA polymerases can be error-prone and may result in mutations, which, if present in critical cell cycle control genes such as oncogenes and tumor suppressor genes, can lead to cancer initiation (46-49). A bulky DNA adduct at the active site is bypassed more easily by the Y-family DNA polymerases, and the spacious active site results in a relaxed geometric selection for the incoming 2'-deoxynucleotide 5'-triphosphate (dNTP) (48, 49), thus compromising the efficiency and fidelity of DNA replication. The lesion bypass

ability, accuracy, and efficiency of these polymerases vary significantly and depend on the type of DNA lesion (46-65).

Crystal structures of Dpo4 with unmodified DNA and the incoming nucleotide cocrystallized with the protein provide excellent models for investigating the structural features that determine lesion bypass efficiency and fidelity (41). These structures solved for Dpo4 showed two modes of entry for a dNTP (45, 46). They can be used to perform docking experiments of the primer/adducted MC and 2,7-DAM-templates using molecular dynamics simulation and AMBER 6.0 force-field as used by others to study details of the interactions between the alkylated templates and the Dpo4-DNA polymerase which could explain the different translesion synthesis observed in biochemical assays (41-45).

The translesion synthesis (TLS) on adducted templates containing the adducts N7-dG-2,7-DAM and N^2-dG-2,7-DAM were performed on constructs containing the N7- dG-2,7-DAM (construct 19-scheme 4) and on templates constructs containing N^2-dG-2,7-DAM monoadduct. Two polymerases were shown the TLS of the N^2-dG-2,7-DAM monoadduct: T7 exo- and Klenow exo- polymerase (figures 60-63). We assumed that N^2-dG-2,7-DAM monoadduct can be located on either of the three guanines in the template. The ESI-MS (-) data showed the incorporation of one single drug within the structure. The degree of primer extension for the template adducted with N^2-dG-2,7-DAM was less efficient than for the templates adducted with N7-dG-2,7-DAM adduct both with Klenow exo- and T7 exo- polymerase (figures 60-63). Klenow exo- polymerase was less efficient than T7 exo- in performing the primer extension reaction on templates adducted with N^2-dG-2,7-DAM lesion. It is important to mention that these are just global observations with respect to the primer extension reaction and we cannot

say anything about the bypass of N^2-dG-2,7-DAM lesion since we don't have yet evidence for the position of the monoadduct within the template.

Further experiments are going on to compare the TLS of N7-dG-2,7-DAM and N^2-dG-2,7-DAM adducted templates using in vitro primer extension and single nucleotide incorporation performed by DNA polymerase eta (η).

CONCLUSION

In conclusion, the lesion with N7-G-2,7-DAM guanine is bypassed by four DNA polymerases we have investigated, Klenow exo-, T7exo-, eta, Klenow exo+ and is not mutagenic. All four DNA polymerases lost thir processivity when it encounters the lesion during the primer extension reactions as reflected in the accumulation of abortive products in addition to the fully extended primer product. Thus, the lesion containing one or two G-alkylated with 2,7-DAM is clearly affecting the normal rate of primer extension. However, the N7-G-2,7-DAM lesion is bypassed resulting in lower yield for the final products of primer extension. The results from the in vitro assay of primer extension are partially in agreement with the "in vivo" experiments which showed the lack of cytotoxicity of 2,7-DAM monoadducts (9-Chapter 1).

The results from single-nucleotide kinetics experiments suggest a lack of mutagenicity for the (N7)-2,7-DAM monoadduct since the highest frequency of single nucleotide incorporation was obtained with dCTP, the normal nucleotide expected to be incorporated opposite G. This lack of mutagenicity of the N7-G--2,7-DAM monoadduct is consistent with the new findings which showed that the adduct was only weakly toxic and generated ~50% progeny as compared to control when it was inserted into an M13

plasmid and its replication was screened in repair-competent *Escherichia coli* (9-Chapter 1). No mutant was isolated after analysis of more than 4000 progeny phages from SOS-induced or uninduced host cells; therefore, it was estimated that the mutation frequency of 2,7-DAM-dG-N7 was less than 2×10^{-4} in *E. coli*. In addition, to determine if N7-G--2,7-DAM might be mutagenic in mammalian cells, it was incorporated into a single-stranded shuttle phagemid vector, pMS2, and replicated in simian kidney (COS-7) cells. Analysis of the progeny showed that mutational frequency of a site specific 2,7-DAM-dG-N7 was not higher than the spontaneous mutation frequency in simian kidney cells (9-Chapter 1). Thus, our in vitro experiments of primer extension and single nucleotide incorporation opposite the N7-G-2,7-DAM lesion are in agreement with these in vivo findings which are supporting the non-mutagenic potential of this lesion and its lack of cytotoxicity.

We do not have X-Ray data or molecular models of the MC and 2,7-DAM adducted templates annealed with primers and any DNA polymerases investigated. These in vitro biochemical studies raised structural questions related to the structural factors that may be responsible for the observed differences between the 2,7-DAM-dG-N7 adduct **8** and the MC-dG-N^2 adduct **7** as replication blocks. The three-dimensional solution structure of the G-N7 adduct **8** in duplex DNA (18-Chapter 1) indicated that the drug moiety is not intercalated and lies in the major groove of a relatively nonaltered B-DNA structure. In contrast, the mitomycin drug moiety of the G-N^2 adduct **7** lies snugly in the minor groove of duplex DNA (19-Chapter 1). Although solution structures of primer-template junctions at mitomycin adducts have not been determined experimentally, it was speculated based on the adducted duplex DNA structures above

that at this junction the Watson-Crick H-bonding interface of the G-N7 adduct in the template and the incoming dNTP is unobstructed; furthermore, the drug moiety of the adduct has no tendency to stack with the incoming dNTP and the guanine when this basepair is formed (9-Chapter 1). Rather, the drug moiety stays out of the way in the incipient major groove during the TLS step at the active site. In contrast, the H-bonding interface of the G-N^2 adduct in the template is directly obstructed by the bound drug moiety and formation of the closed conformation of the DNA polymerase is sterically prevented (9-Chapter 1).

A similar scenario was proposed for earlier precedent, which is analogous to the present findings (38): The ethylene dibromide-derived bulky glutathione adducts at the N7- and N^2-positions of guanine differ considerably in their behavior in TLS systems of all four DNA polymerases tested in their system (38). The N7 adduct allowed relatively efficient full primer extension and selective incorporation of dCTP opposite the adduct, while the N^2 adduct strongly blocked replication and only dATP was incorporated opposite to the adduct. The authors suggested a similar explanation to what we proposed above and speculated that the weak blocking effect of the G-N7 adduct is due to direct interaction of the bulky glutathione residue with the polymerases. On the basis of these considerations and our present results, it appears that bulky G-N7 adducts, which intercalate in duplex DNA, inhibit DNA replication and its fidelity while groove-binding G-N7 adducts are less inhibitory (19-Chapter 1).

APPENDIX I

1. Synthesis of an internal N7-2,7-DAM monoadduct standard by chemical reductive alkylation of dG with 2,7-DAM. In order to confirm the structural identity of the N7-2,7-DAM monoadduct of Guanine, the nucleoside **dG** (commercially available) was also alkylated with 2,7-DAM, under the same experimental conditions described for the oligonucleotide duplex (I) (**C6/C7**) (see Materials and Methods). In **figure 1** the HPLC reverse chromatography on a C18 analytical column presents the pattern of elution for the major adducts that are obtained by alkylation of dG with 2,7-DAM, under reductive conditions. The reverse HPLC was run with an analytically C18 column, 100 angstroms, Microsorb, and the buffer system was: Buffer A: 0.02 M NH_4Acetate, pH=5.5, Buffer B:30% acetonitrile in 70% 0.02 M NH_4Acetate. The gradient applied was 20-60% in 60 minutes, with a flow rate of 1ml/min.

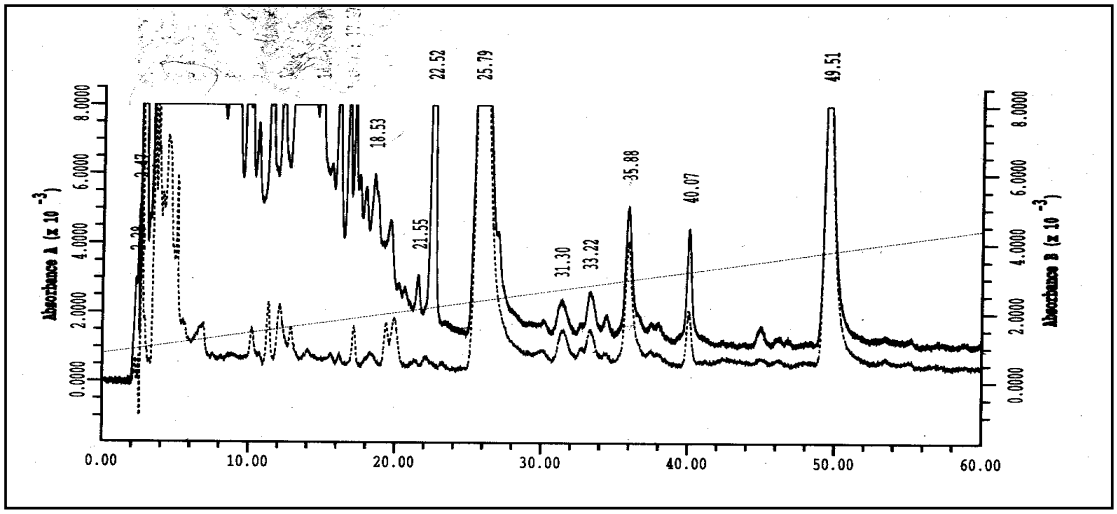


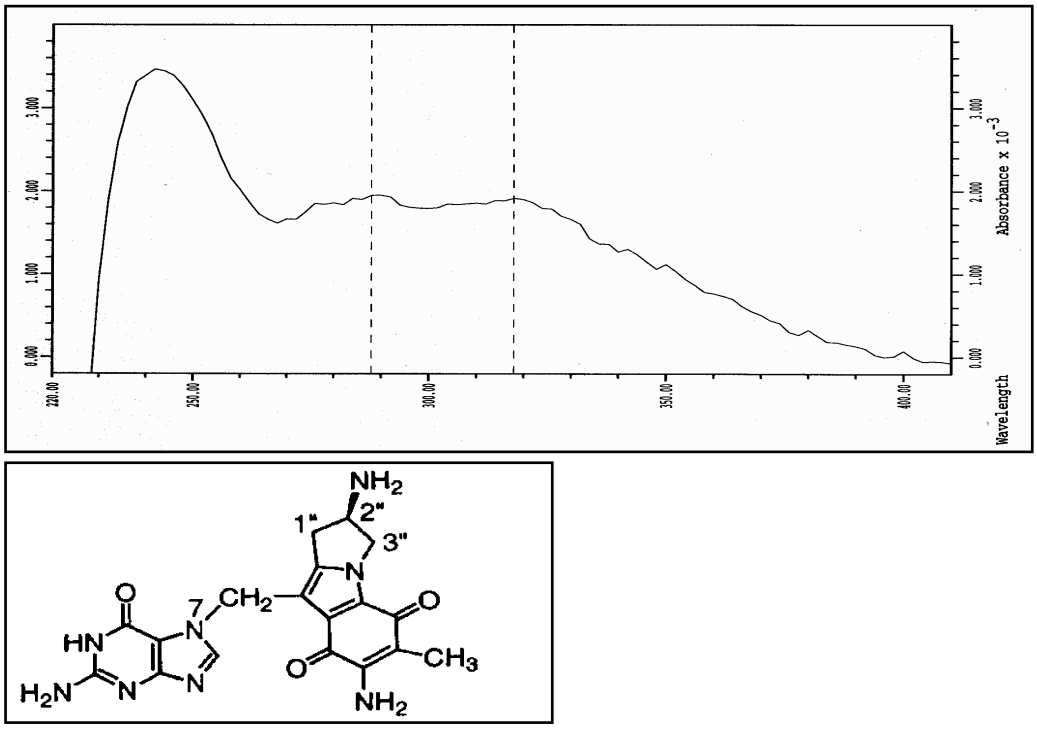
Figure 69 . **The reverse-phase HPLC of reaction mixture containing dG alkylated with 2,7-DAM and free excess of drug**. A parallel control was run in which the drug itself was injected in the column (data not shown), in order to identify the peaks that are corresponding to the drug.

As can be seen from **figure 1**, many adducts are formed, the major peaks being eluted at 31.30 minutes, 33.22 minutes, 35.88 minutes and 40.07 minutes. The peaks at 25 and

49.51 minutes were assigned to the drug itself. An inspection of the spectral characteristics showed that the peak at 35 minutes correspond to the N7-G monoadduct while the peak at 40.07 minutes correspond to another adduct of guanine with 2,7 DAM, originally identified by Y. Palom as being the N2-guanine monoadduct (23-Chapter 1).

The structure of **N7-2,7-DAM** monoadduct is presented below (figure 1 (B)) together with the spectral characteristics that were obtained from the **scan** (between 200-400 nm) of the peak at **35 minutes**.

Structure and UV-scan of **N7-G-2,7-DAM** monoadduct at 35.5 min retention time.



(B) The structure and the spectrum of the N7-G monoadduct with 2,7-DAM.

The spectral characteristic of the N7-G monoadduct is the ratio of 1 between the absorbance at 285 nm and the absorbance at 314 nm. It is important to mention that the N7-guanine monoadduct have specific characteristics, i.e. by heating it depurinates,

releasing the dG which has the 2,7-DAM drug covalently linked to the N-7 position in the guanine ring.

2. Alkylation of duplex (II)- 5'-CTAGTGGTATCC-3'/(CI)-3'-TCACCATAGG-5' (CIII) by 2,7-DAM in the conditions that favors the production of N7-dG-2,7-DAM and N^2-dG-2,7-DAM monoadducts.

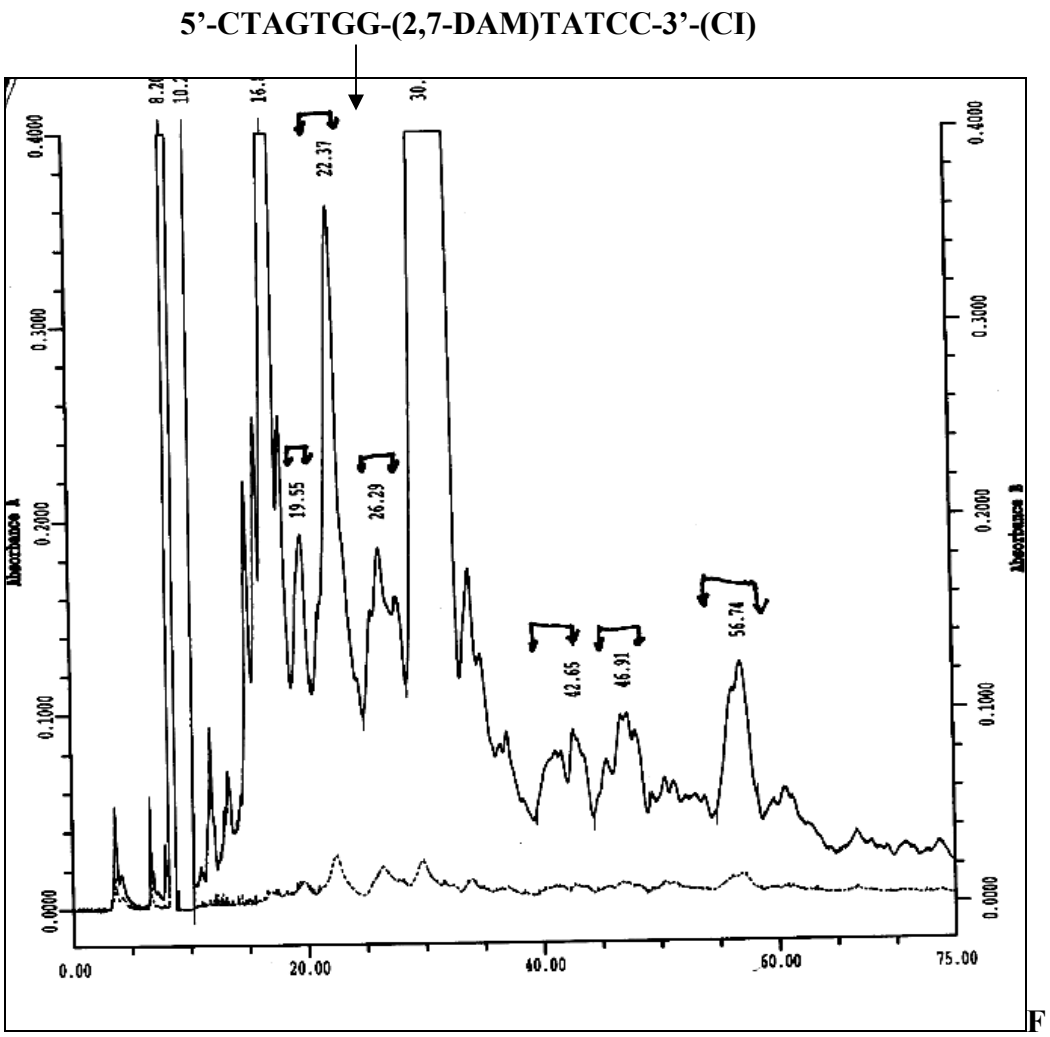
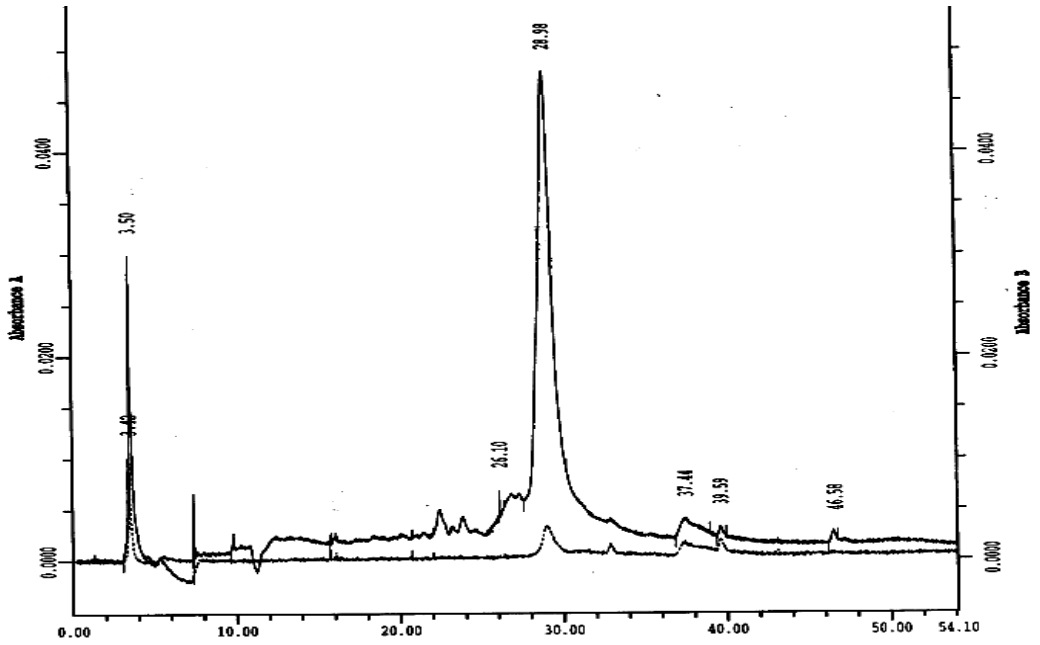
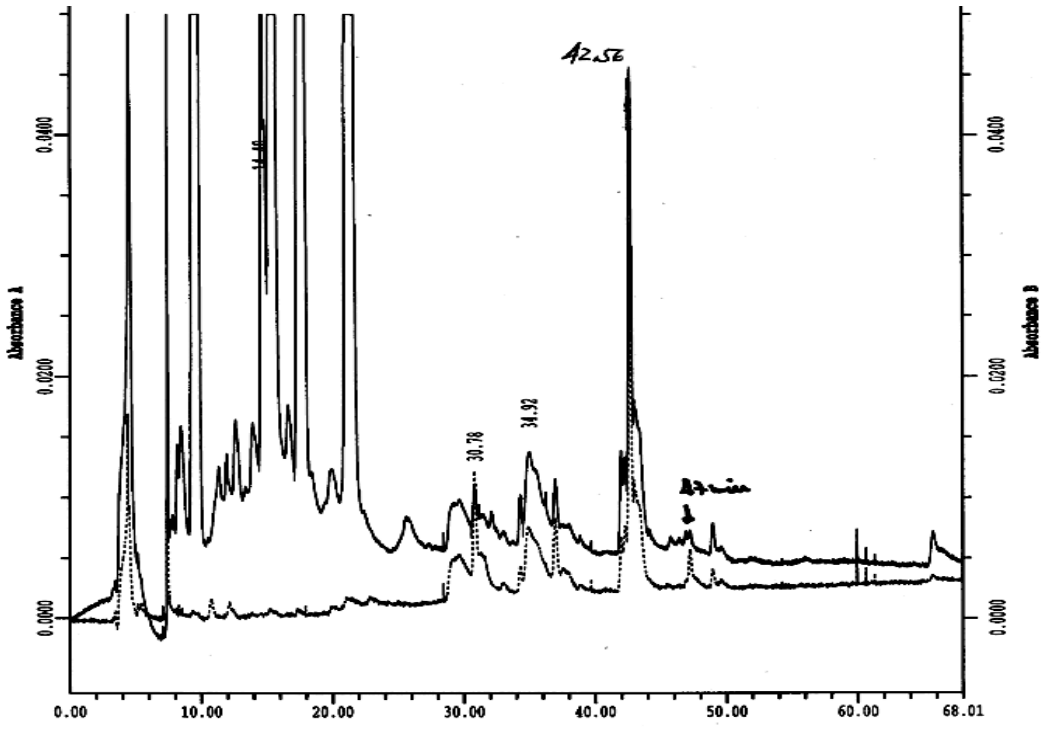


Figure 70: Semipreparative run of the duplex (II)- 5'-CTAGTGGTATCC-3'/(CI)-3'-TCACCATAGG-5' (CIII) alkylated with 2,7-DAM: hydrogenation: 25 minutes; argon: 90 minutes.

(A)


3.50
26.10
28.98
37.44
39.59
46.58
Absorbance A
Absorbance B
0.0400
0.0200
0.0000
0.00
10.00
20.00
30.00
40.00
50.00
54.10

(B)


42.58
30.78
34.92
Absorbance A
Absorbance B
0.0400
0.0200
0.0000
0.00
10.00
20.00
30.00
40.00
50.00
60.00
68.01

(C)

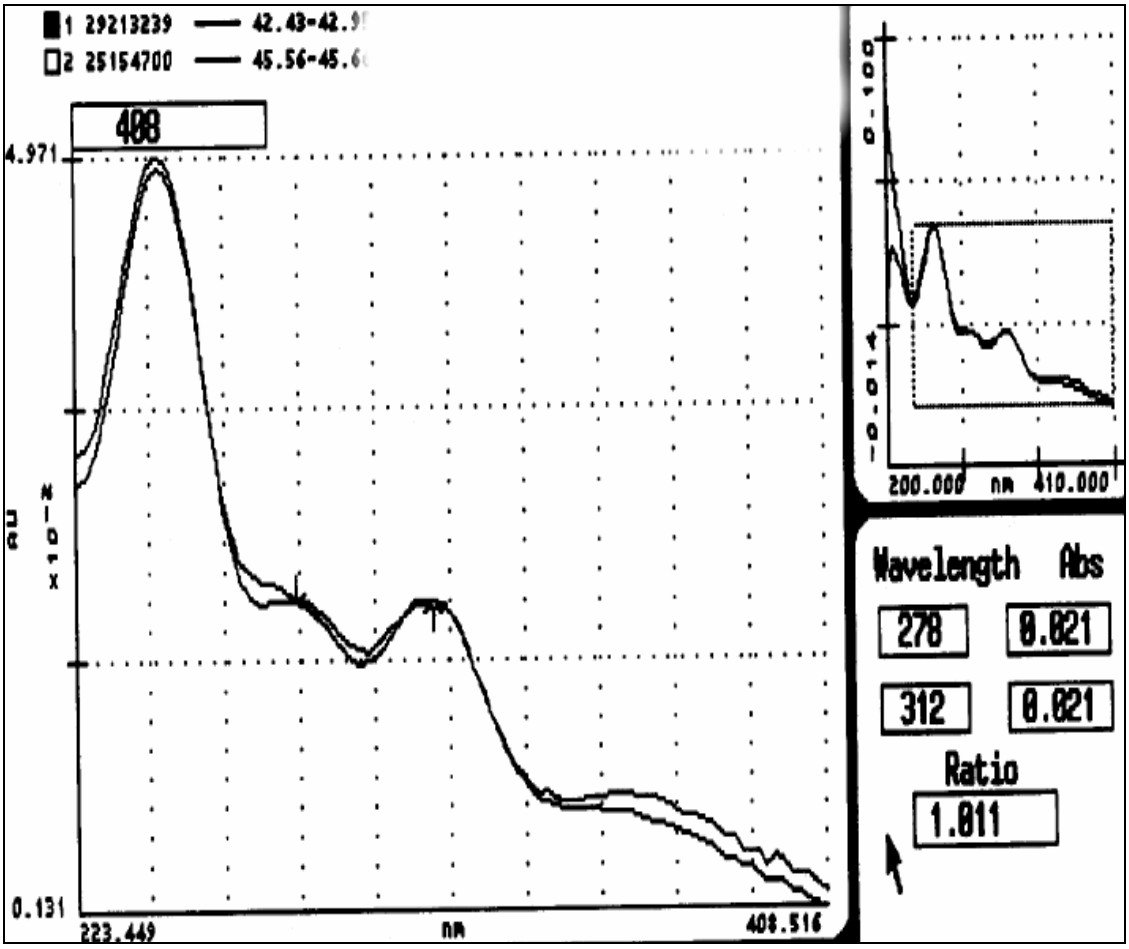


Figure 71: The purified N^2-dG-2,7-DAM adducted C-I-5'-CTAGTGGTATCC-3'oligonucleotide analysis by reversed-phase HPLC (A). SVD treatment of pure C-I-5'-CTAGTGGTATCC-3' oligonucleotide for nucleoside analysis. The adduct eluted at 42.56 minutes in **(B)** shows the UV-characteristics of N^2-dG-2,7-DAM monoadduct (C).

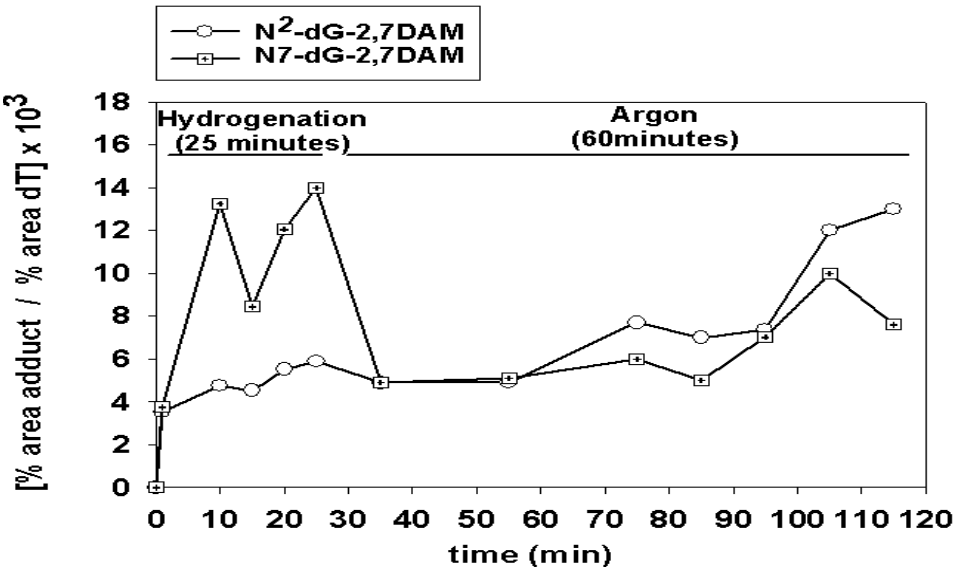
APPENDIX II

In **Figure 72 (A)** the % adduct area of N7-G-2,7-DAM monoadduct and % of N^2-dG-2,7-DAM relative to the % of dT is presented as a function of alkylation reaction time. In general the N7-G-2,7-DAM monoadduct is instable supporting the hypothesis that it can be decomposed under anaerobic conditions into a reactive species of 2,7-DAM which can in turn activate the NH2 (N2) position of guanine giving rise to N2-dG-2,7-DAM monoadduct. However, and independent pathway for producing the N2-dG-monoadduct cannot be excluded. In contrast, the N2-dG-2,7-DAM monoadduct is more stable and is formed at a slower rate than N7-2,7-DAM monoadduct. Also it can be seen that other adducts (eventually of guanine-see figure 23 (C) their UV scans) labeled z, w, W are also formed in parallel with N2-dG-2,7-DAM adduct, possibly competing with the intermediate reactive species which gives rise also to N2-dG-2,7-DAM adduct.

In Figure 72 **(B-C)** the SVD treatment of a reaction mix at two different data points under Argon is presented as an example for the assignment of the adducts N7-dG-2,7-DAM and N2-dG-2,7-DAM. In all cases the reaction mix treated with SVD and alkaline phosphatase is heated to fully release the N7-dG-2,7-DAM. The assignment of the adducts is performed based on authentic standards of N7 and N2-dG-2,7-DAM and on their UV-characteristics **(figures 20-23).**

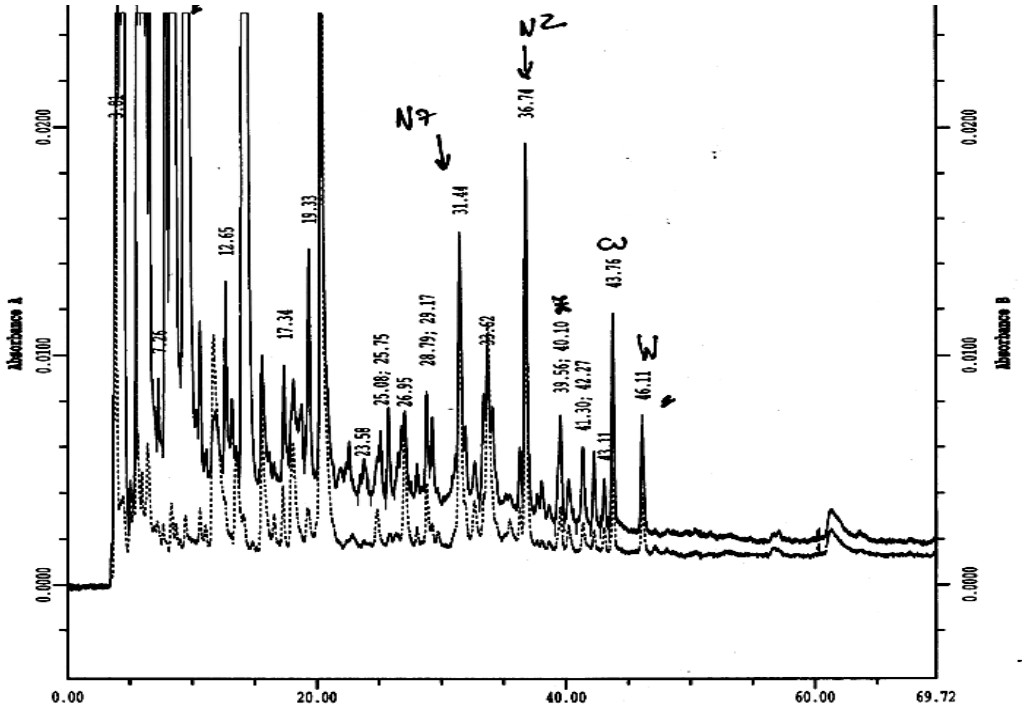
In **figure 72 (D)** the ratio N7/N2-dG-2,7-DAM is plotted as a function of alkylation reaction time. This graph suggests indirectly that N7-dG-2,7-DAM is converted into N2-dG-2,7-DAM but there are not direct proofs for this mechanism yet.

(A)



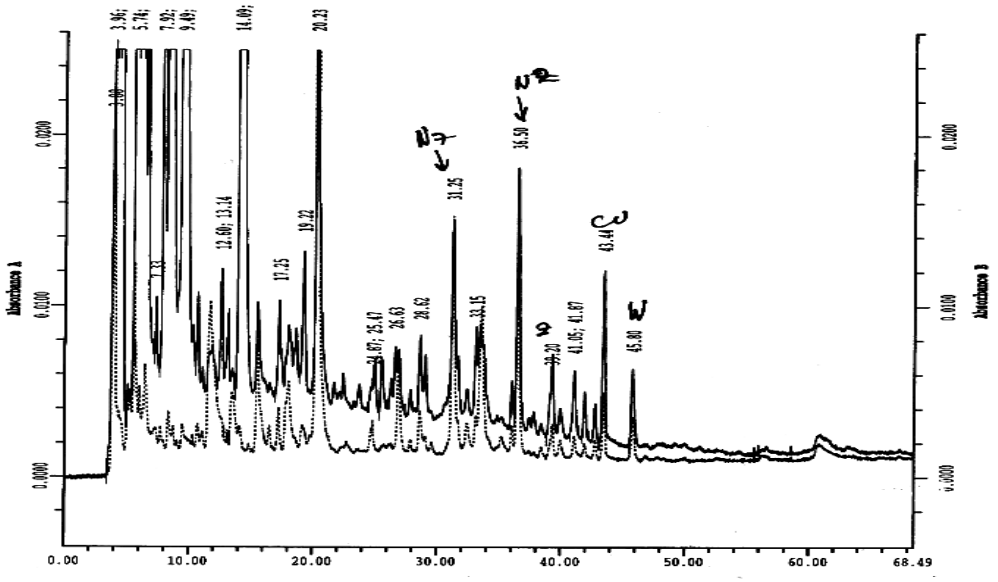
(B)

80 minutes Argon (after 25 minutes hydrogenation)



90 minutes Argon (after 25 minutes hydrogenation)

(C)



(D)

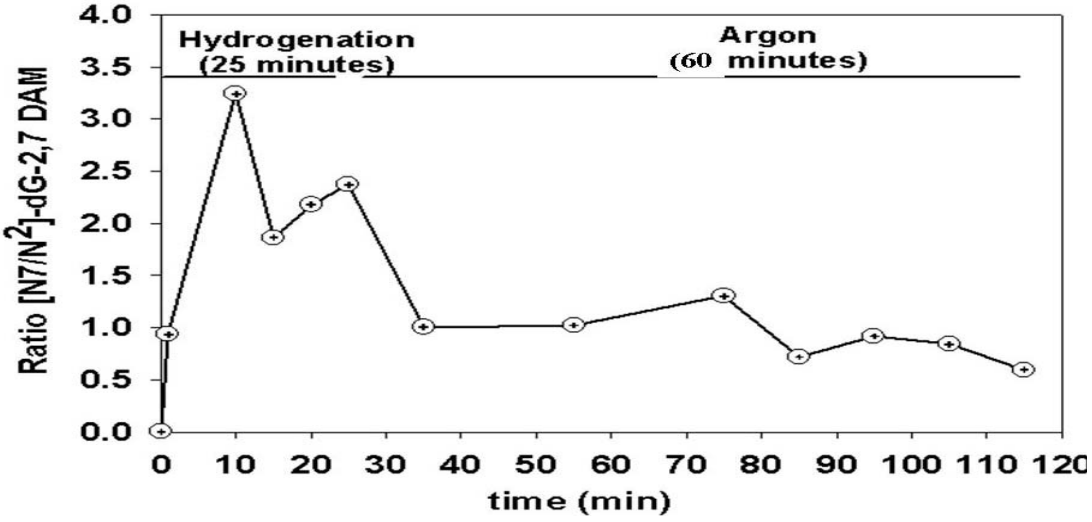


Figure 72: The progress of a reaction mix during alkylation of DNA duplex (I): 5'-CTGG*TAATTTAC-3' (C6 oligonucleotide) / 3'-GACCATTAA-5' (C7 oligonucleotide) under conditions which favor the production of N^2-dG-2,7-DAM monoadduct; hydrogenation time: 25 minutes and Argon: 90 minutes.

APPENDIX III

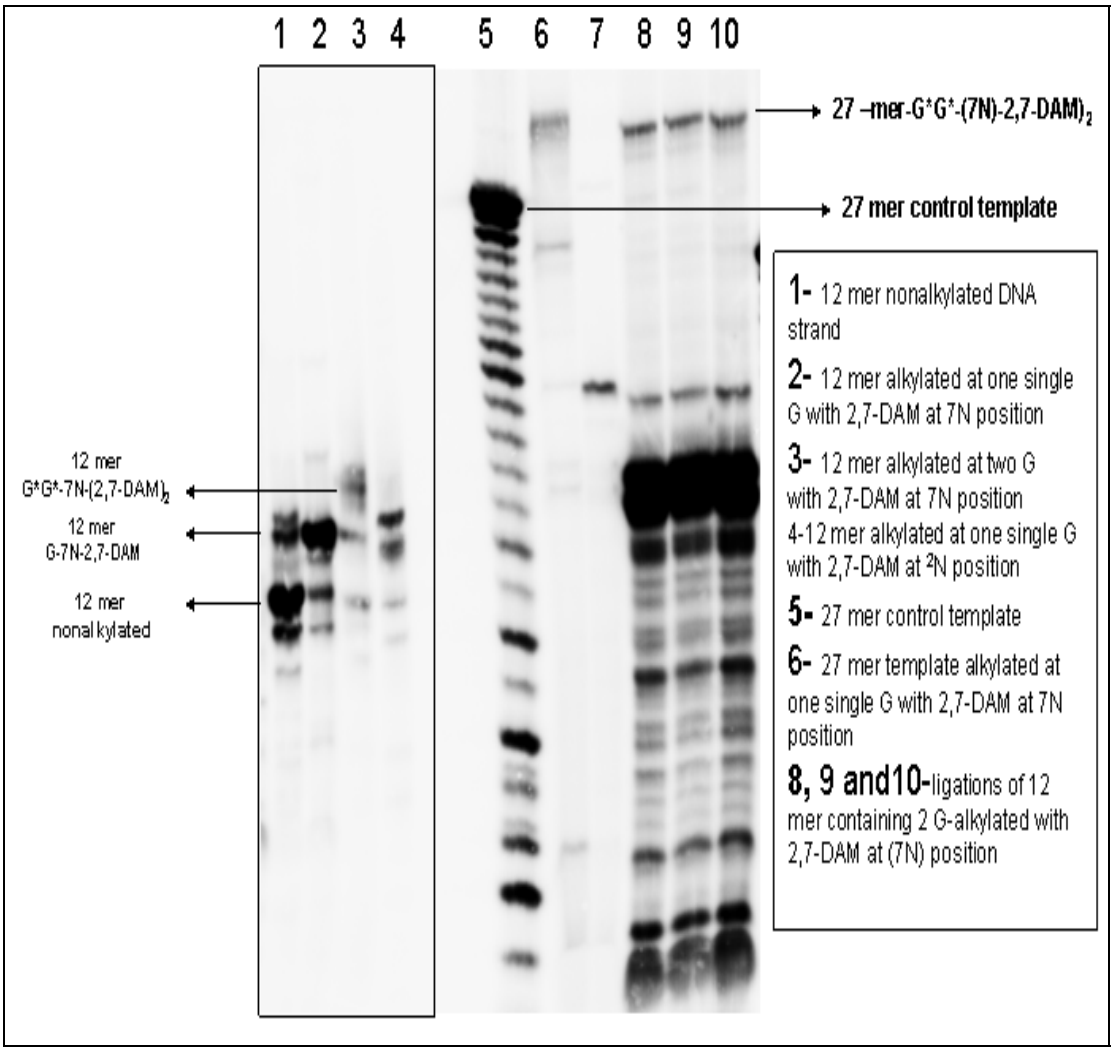


Figure 73. Sequencing gel analysis of ligation mix for ligation of a 12 mer containing 2 guanines alkylated with 2,7-DAM into a 27 mer template (27-mer-G*G*-(2,7DAM)$_2$) (scheme 4-II).

[5'-CTAGTGG*(2,7-DAM) TATCCTAGAGATTGGTA-3']

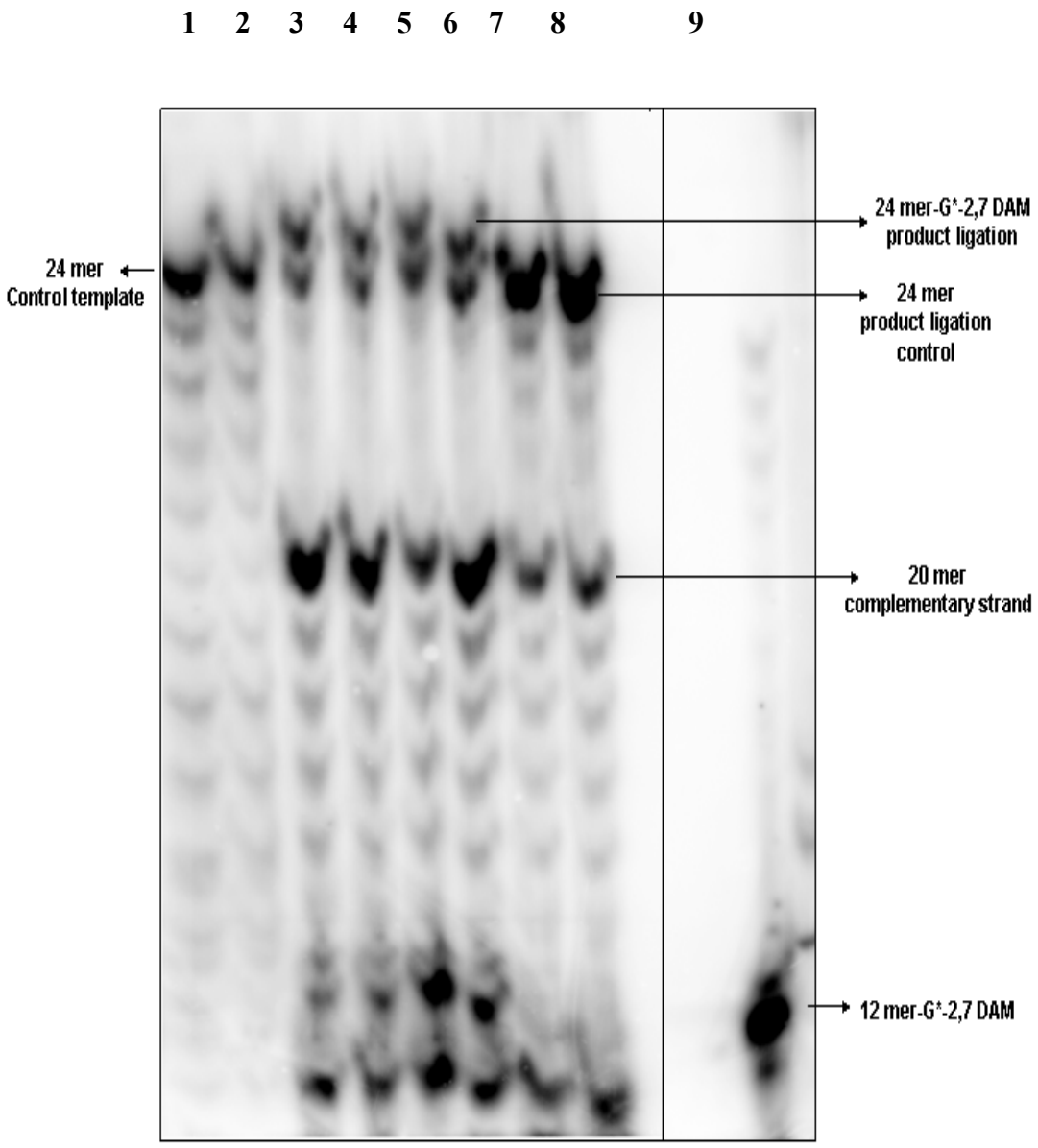


Figure 74. Sequencing gel analysis of ligation products from reactions involving the ligation of 12 mer adduct alkylated at one guanine with 2,7-DAM at (N7) position (scheme 4-II).

APPENDIX IV

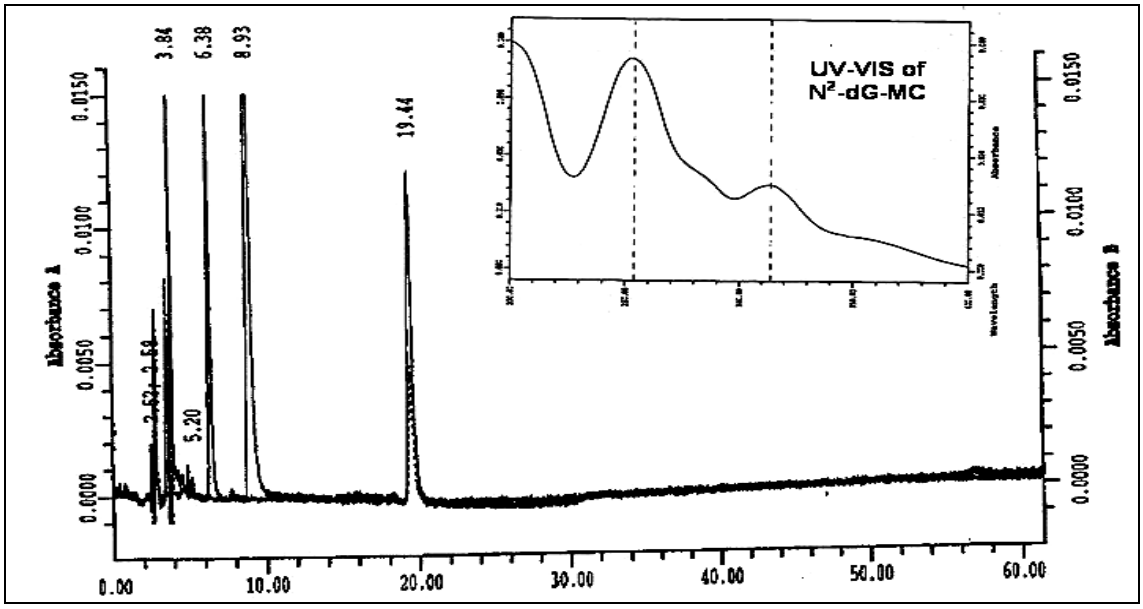


Figure 75: Reversed-phase HPLC analytical run for the alkylated 24 mer-G*MC oligonucleotide digested with snake venom phosphodiesterase and alkaline phosphatase. The peak at 19.4 minutes corresponds to the nucleoside N^2-dG-mitomycin C monoadduct which has the structure shown in the inserted UV scan.

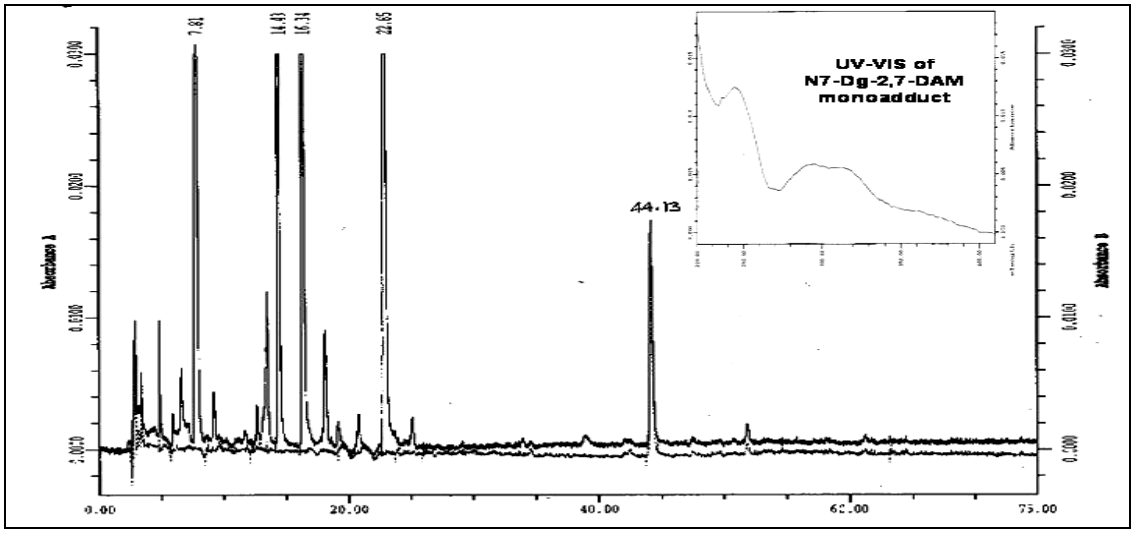


Figure 76: **Reversed-phase HPLC analytical run for the depurinated alkylated 24 mer-G*-2,7-DAM oligonucleotide digested with snake venom phosphodiesterase and alkaline phosphatase.** The peak at 44.13 minutes corresponds to the nucleoside N7-G-2,7-DAM monoadduct (8-figure 1) which has the structure shown in the inserted UV scan.

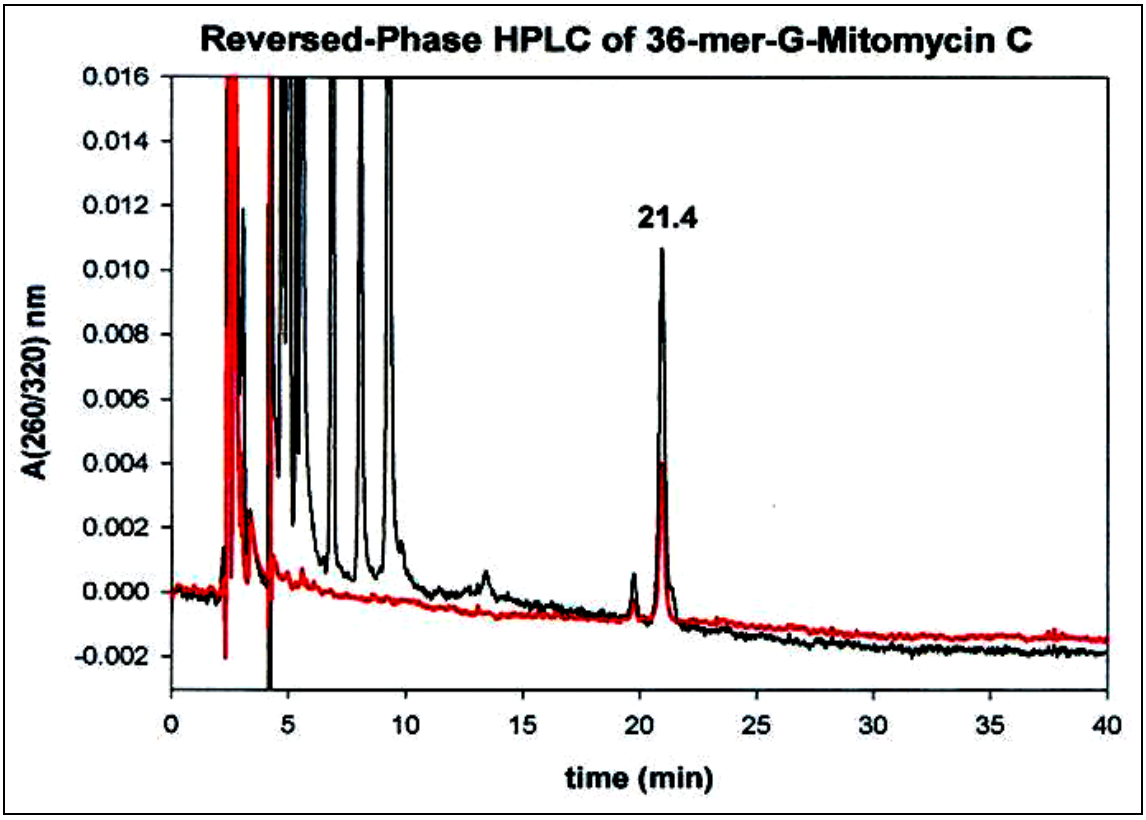


Figure 77: Reversed-phase HPLC analytical run for the alkylated 36 mer-G*MC oligonucleotide digested with snake venom phosphodiesterase and alkaline phosphatase. The peak at 21.4 minutes corresponds to the nucleoside N^2-dG-mitomycin C monoadduct which has the expected structure as is shown in figure 75-Appendix IV.

Adducted templates oligonucleotides were digested with snake venom phosphodiesterase (SVD) and alkaline phosphatase and subjected to HPLC analysis of nucleoside pattern as described in Materials and Methods (section 2.2.3.3.).

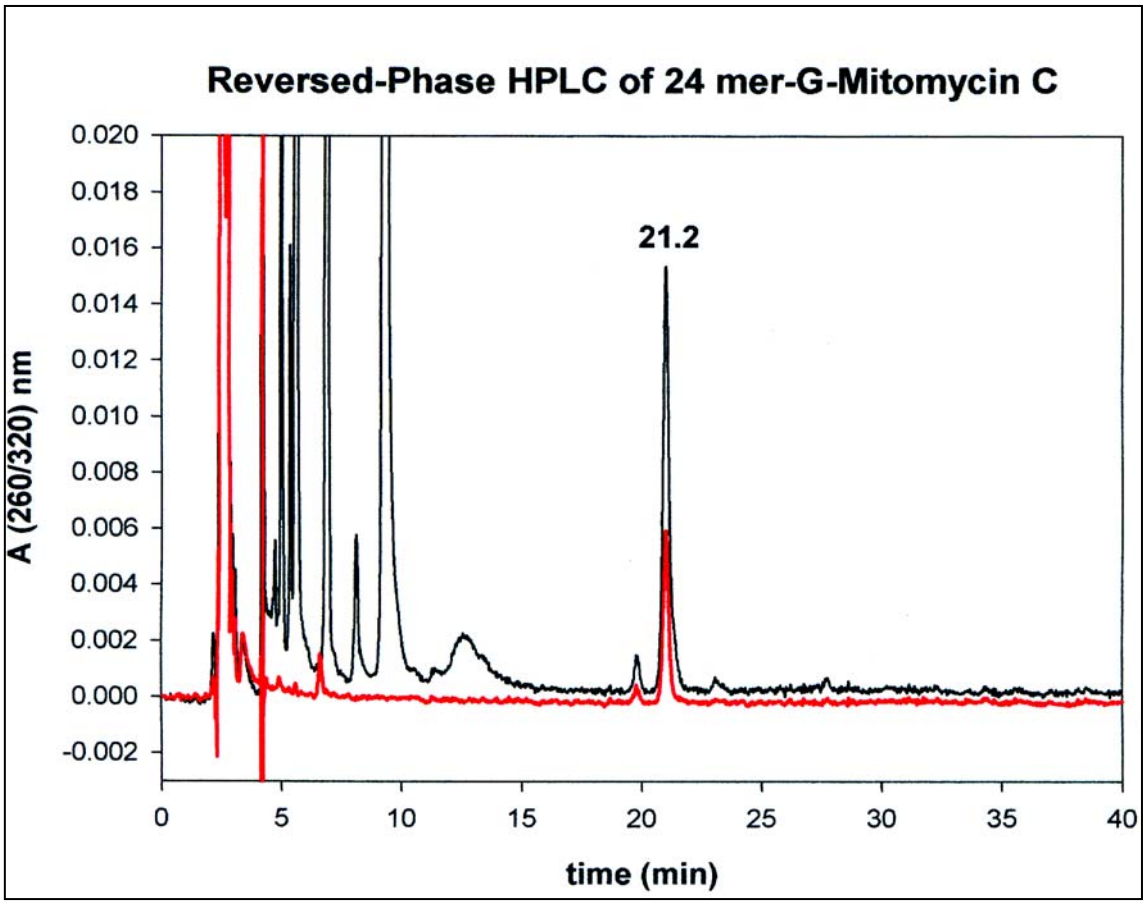


Figure 78: Analytical Reversed-phase HPLC for the alkylated 24 mer-G*MC oligonucleotide digested with snake venom phosphodiesterase and alkaline phosphatase. The peak at 21.2 minutes corresponds to the nucleoside N^2-dG-mitomycin C monoadduct which has the expected structure as is shown in figure 1-Appendix IV.

APPENDIX V

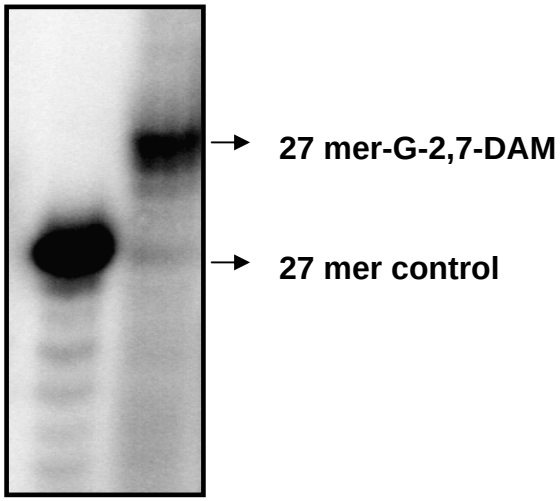


Figure 79: PAGE analysis of the construct template (15) (scheme 3).

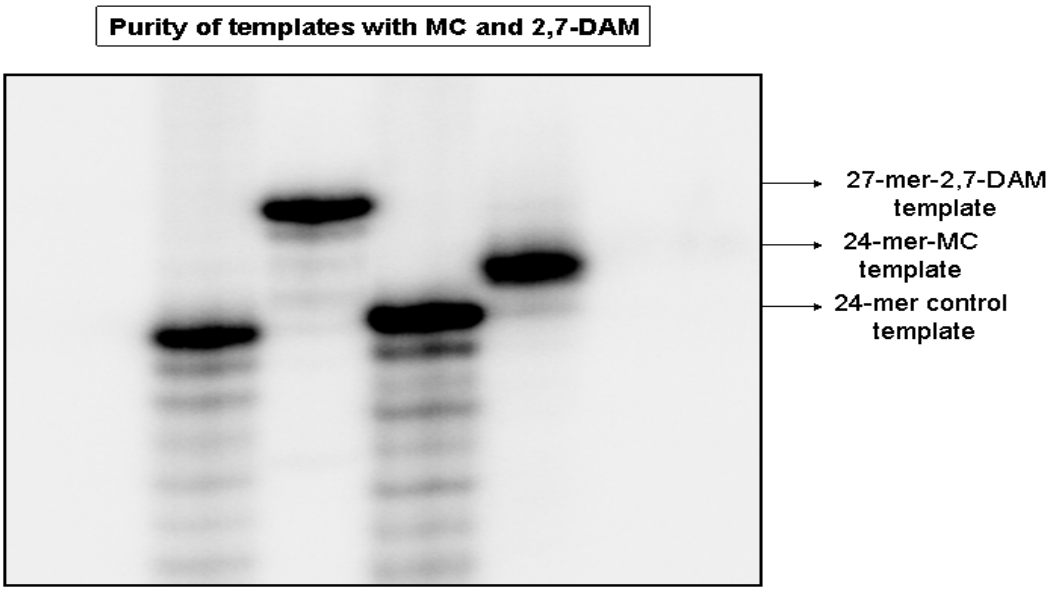


Figure 80: Purity control for the 27 mer-G*(2,7-DAM) and 24-mer–G*(MC) using PAGE analysis.

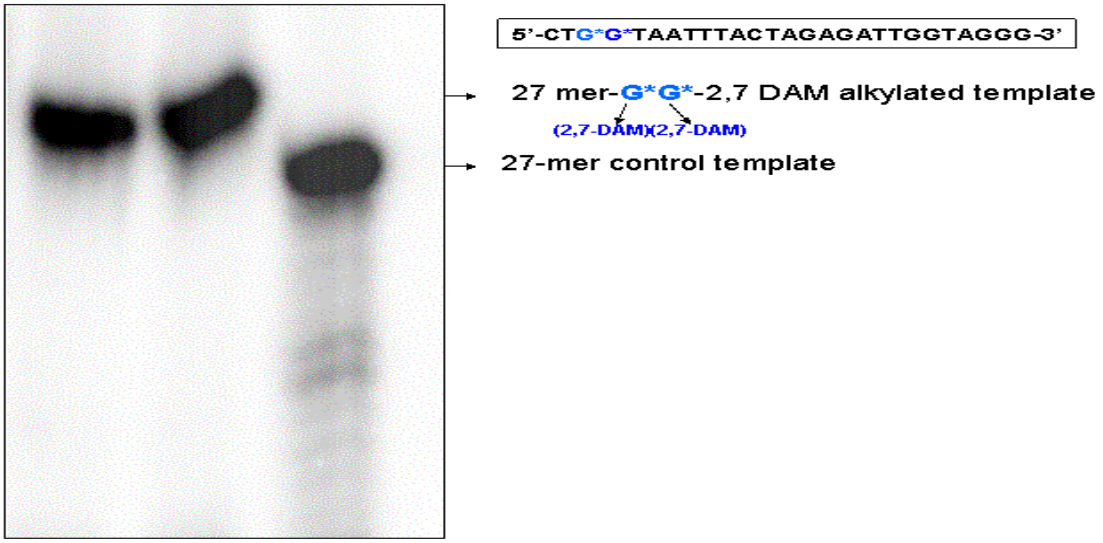


Figure 81. Purity control for the template 27-mer-G*G*-(2,7-DAM)$_2$ [the template containing 2 guanines each alkylated at 7N position with 2,7-DAM).

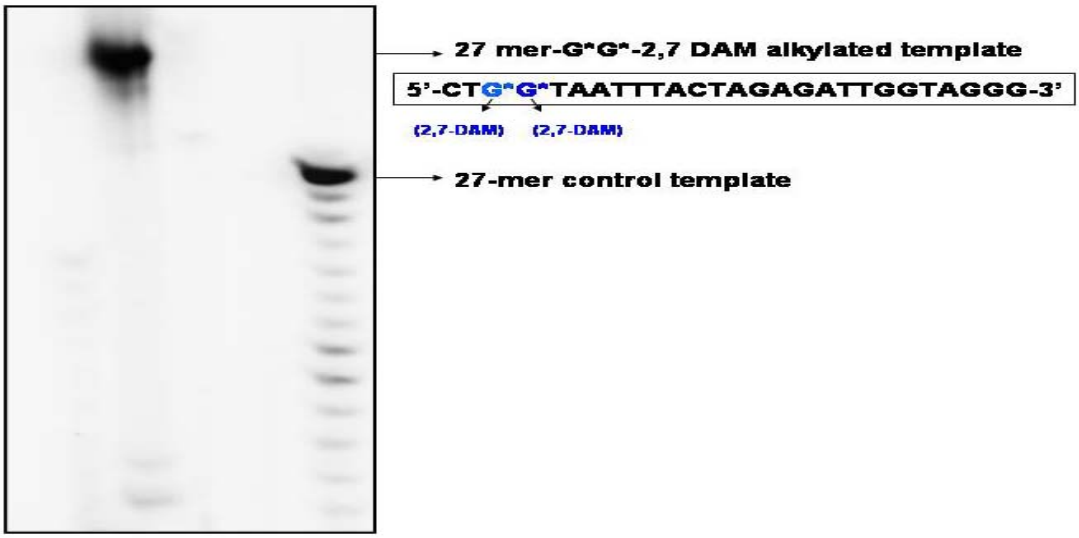


Figure 82: Purity control for the template 27-mer-G*G*-(2,7-DAM)$_2$ [the template containing 2 guanines each alkylated at 7N position with 2,7-DAM]-independent experiment).

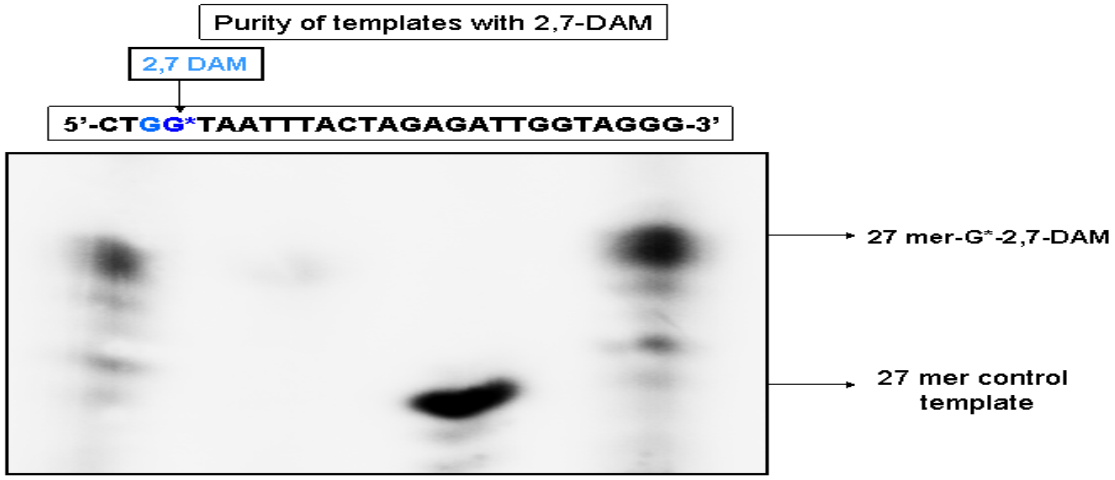


Figure 83: Purity control for the template 27-mer-G*-(2,7-DAM)$_2$-template containing 1 guanine each alkylated at 7N position with 2,7-DAM-independent experiment.

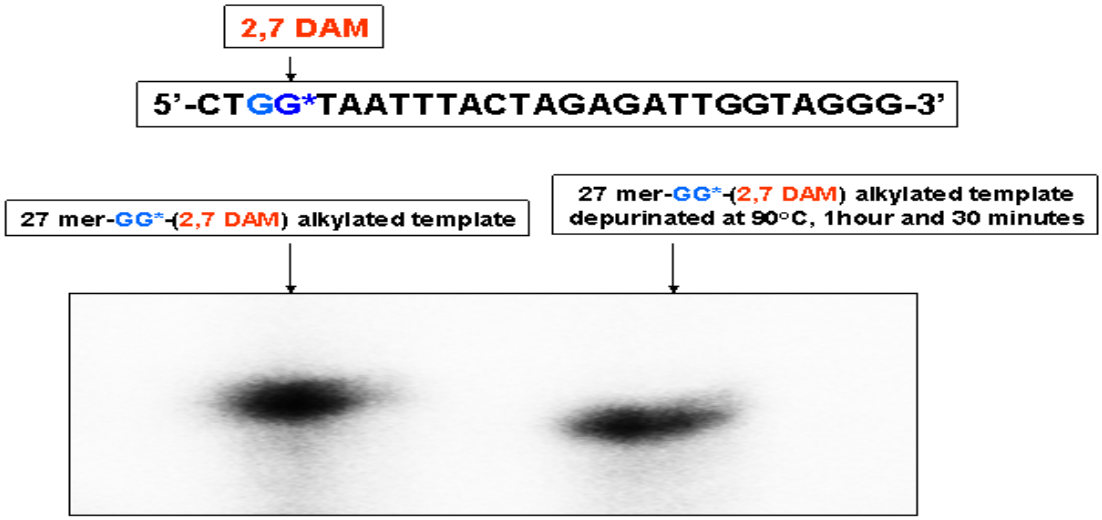


Figure 84: Depurination control for the template 27-mer-G*-2,7-DAM monitored by high resolution sequencing PAGE analysis. In order to test if the 2,7-DAM alkylated is depurinated an aliquot of the template was subjected to depurination by heating at 90^{o}C 1 hour and 30 minutes. The sequencing gel presents the original 27 mer-G*-2,7 DAM template and the depurinated template. It can be seen that the depurinated template (which lost a G-2,7-DAM) migrated faster than the control template (containing the 2,7-DAM drug). This result showed that the template which was subjected to primer extension and single-nucleotide kinetics assay was not depurinated.

APPENDIX VI

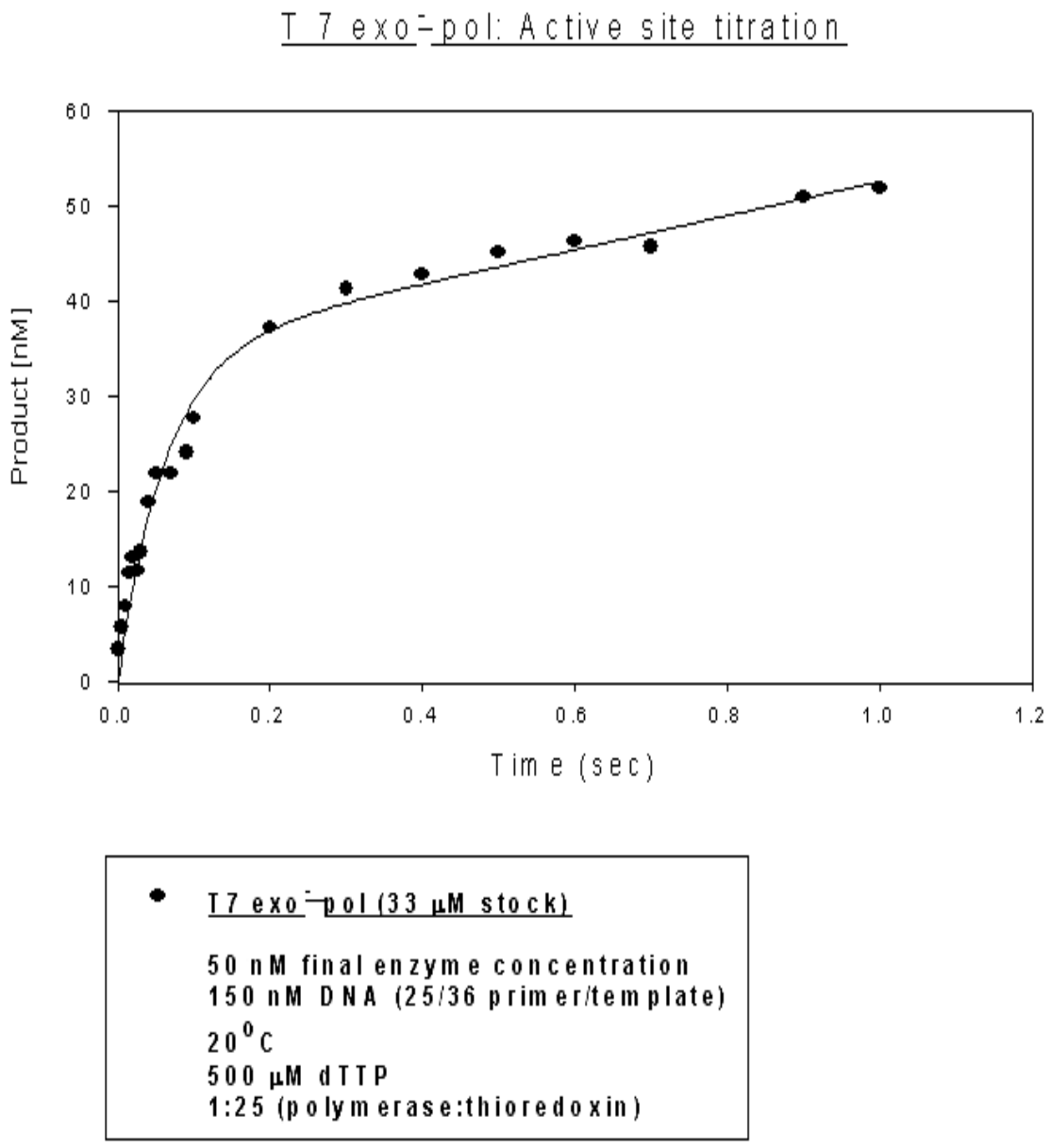


Figure 85: Pre-steady state kinetics for active site titration of T7 exo- DNA polymerase (quench flow).

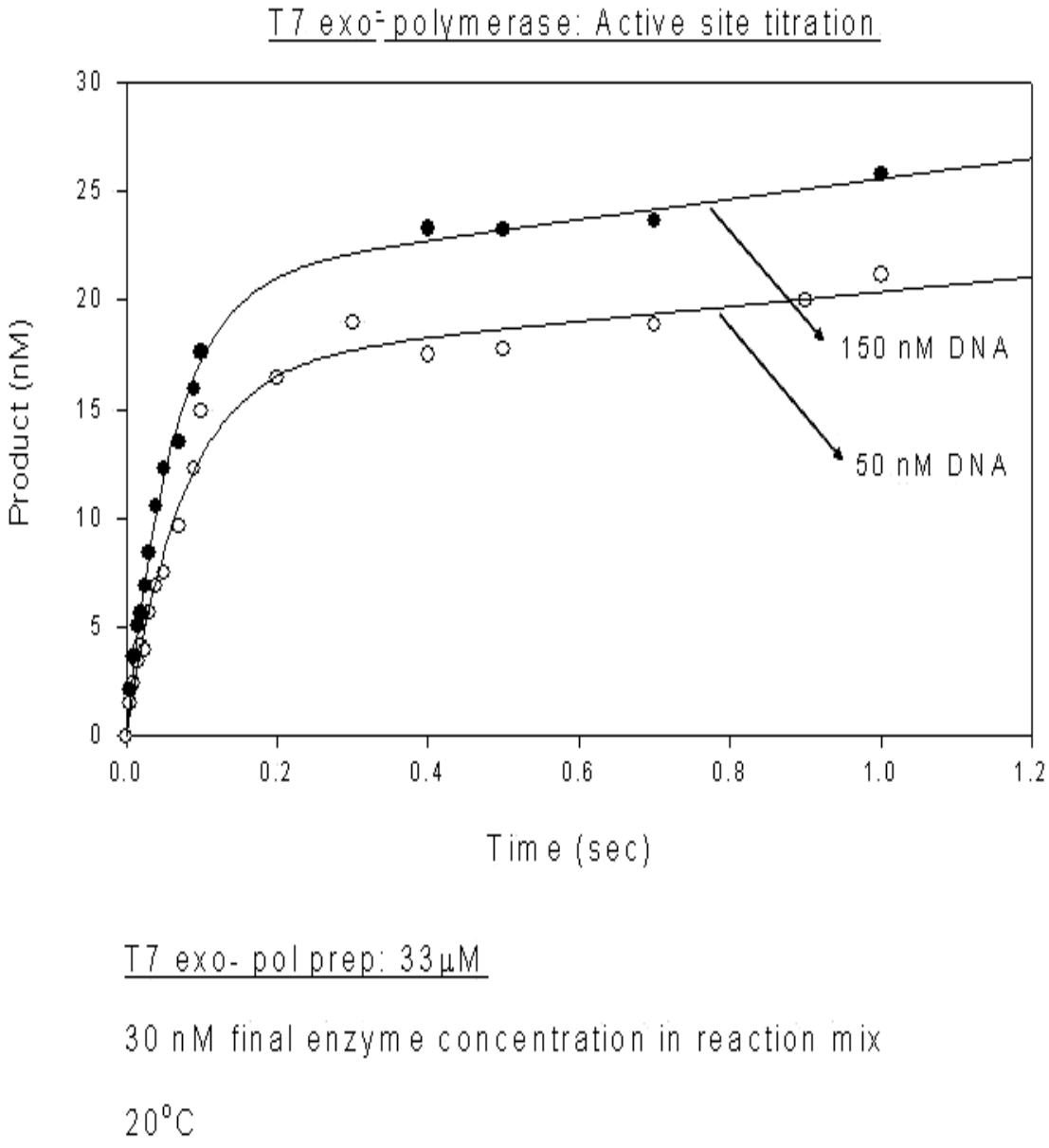


Figure 86: Pre-steady state kinetics for active site titration of T7 exo- DNA polymerase (quench flow) (independent experiment).

APPENDIX VII

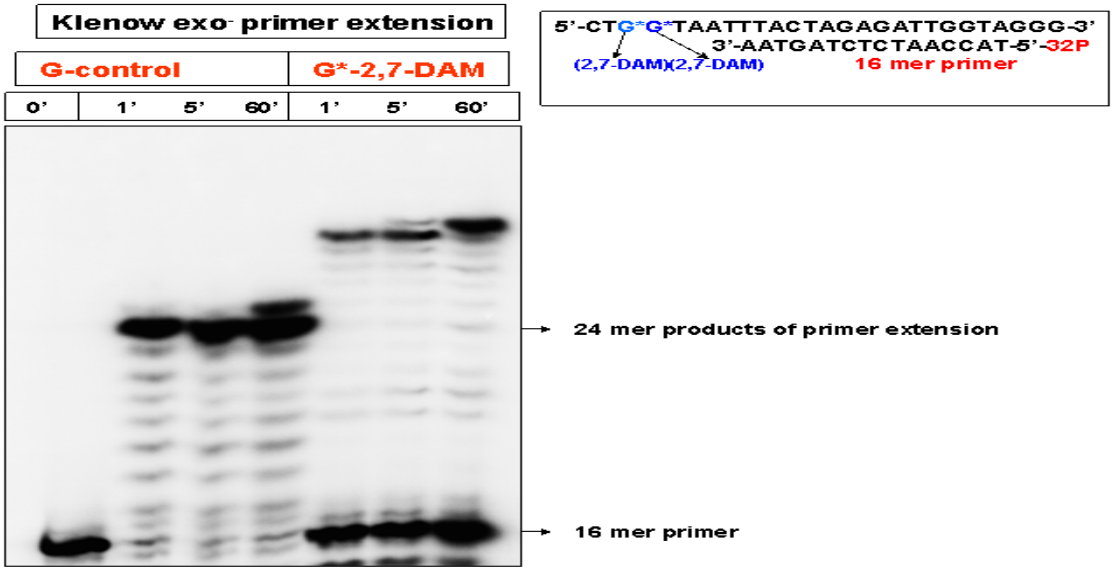


Figure 87: PAGE analysis of primer extension on 2,7-DAM-adducted template of complex *23* by Klenow (exo-) DNA polymerase.

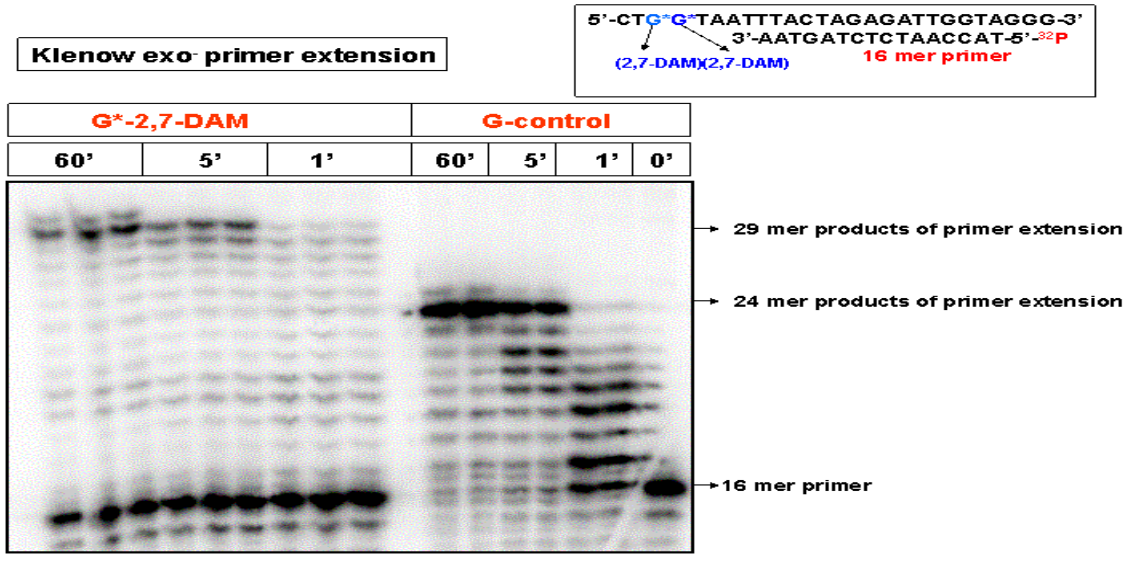


Figure 88: PAGE analysis of primer extension on 2,7-DAM-adducted template of complex *23* by Klenow (exo-) DNA polymerase.

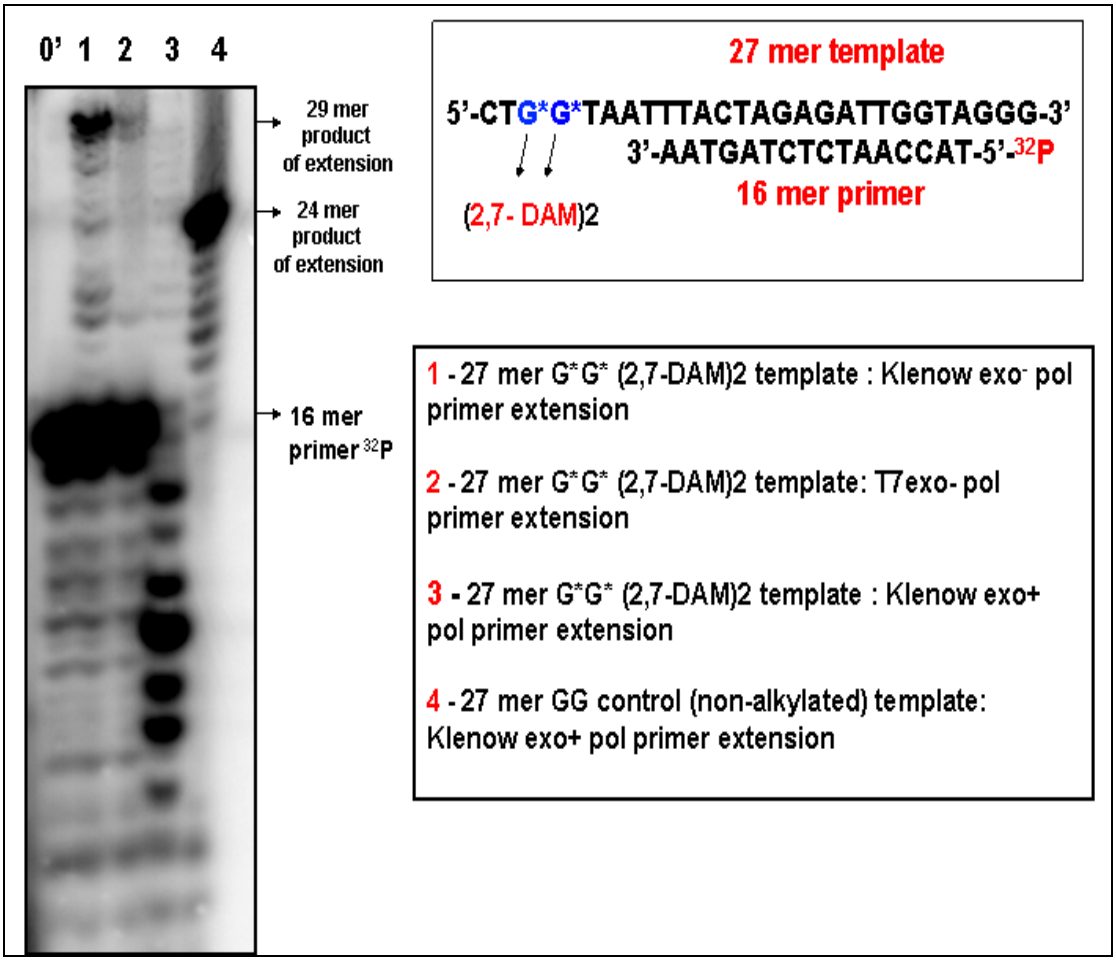


Figure 89: PAGE analysis of primer extension on (2,7-DAM-adducted)2 template of complex *23* by Klenow (exo+) DNA polymerase 250 μM dNTP. At low concentration of each dNTP the N7-G*-2,7-DAM lesion is not bypassed by Klenow exo+. Klenow exo- has a higher efficiency of lesion bypass than T7 exo- DNA polymerase.

REFERENCES

1. Christner, D. F., Lakshman, M. K., Sayer, J. M., and Dipple, A. (1994) Primer extension by various polymerases using oligonucleotide templates containing stereoisomeric benzo[a]pyrene-deoxyadenosine adducts. *Biochemistry* **33,** 14297-14305.

2. Abbotts, J., Gupta, D. N., Zon, G., and Wilson, S. H. (1988) Studies on the mechanism of *Escherichia coli* DNA polymerase I large fragment. Effect of template sequence and substrate variation on termination of synthesis. *J. Biol. Chem.* **263,** 15094-15103.

3. Chary, P., and Lloyd, R. S. (1995) In vitro replication by prokaryotic and eukaryotic polymerases on DNA templates containing site-specific and stereospecific benzo[a]pyrene-7,8-dihydrodiol-9,10-epoxide adducts. *Nucleic Acids Res.* **23,** 1398-1405.

4. Hruszkewycz, A. M., Canella, K. A., and Dipple, A. (1991) DNA polymerase-mediated nucleotide incorporation adjacent to hydrocarbon-deoxyadenosine and hydrocarbon-deoxyguanosine adducts. *Carcinogenesis* **12,** 1659-1663.

5. Hruszkewycz, A. M., and Dipple, A. (1991) Bypass of a hydrocarbon adduct in an oligonucleotide template mediated by mispairing adjacent to the adduct. *Carcinogenesis* **12,** 2185-2187.

6. Hruszkewycz, A. M., Canella, K. A., Peltonen, K., Kotrappa, L., and Dipple, A. (1992) DNA polymerase action on benzo[a]pyrene-DNA adducts. *Carcinogenesis* **13,** 2347-2352.

7. Shibutani, S., Margulis, A. L., Geacintov, N. E., and Grollman, A. P. (1993) Translesion synthesis on a DNA template containing a single stereoisomer of dG-(+)- or dG-(-)-*anti*-BPDE (7,8-dihydroxy-*anti*-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene). *Biochemistry* **32,** 7531-7541.

8. Belguise-Valladier, P., Maki, H., Sekiguchi, M., and Fuchs, R. P. P. (1994) Effect of single DNA on *in vitro* replication with DNA polymerase III holoenzyme. Comparison with other polymerases. *J. Mol. Biol.* **236,** 151-164.

9. Comess, K. M., Burstyn, J. N., Essigmann, J. M., and Lippard, S. J. (1992) Replication inhibition and translesion synthesis on templates containing site-specifically placed *cis*-diamminedichloroplatinum(II) DNA adducts. *Biochemistry* **31,** 3975-3990.

10. Basu, A. K., Hanrahan, C. J., Malia, S. A., Bizanek, R., and Tomasz, M. (1993) Effect of site-specifically located mitomycin C-DNA mono adducts on *in vitro* DNA synthesis by DNA polymerases. *Biochemistry* **32,** 4708-4718.

11. Thrall, B. D., Mann, D. B., Smerdon, J. J., and Springer, D. L. (1992) DNA polymerase, RNA polymerase and exonuclease activities on a DNA sequence modified by benzo[a]pyrene diolepoxide. *Carcinogenesis* **13,** 1529-1534.

12. Reardon, D. B., Bigger, C. A., and Dipple, A. (1990) DNA polymerase action on bulky deoxyguanosine and deoxyadenosine adducts. *Carcinogenesis* **11,** 165-168.

13. Latham, G. J., and Lloyd, R. S. (1994) Deoxynucleotide polymerization by HIV-1 reverse transcriptase is terminated by site-specific styrene-oxide adducts after translesion synthesis. *J. Biol. Chem.* **269,** 28527-28530.

14. Beard, W. A., and Wilson, S. H. (1993) Kinetic analysis of template-primer interactions with recombinant forms of HIV-1 reverse transcriptase. *Biochemistry* **32,** 9745-9753.

15. Alekseyev, Y. O., and Romano, L. J. (2000) In vitro replication of primer-templates containing benzo [a] pyrene adducts by exonuclease-deficient *Escherichia coli* DNA polymerase I (Klenow fragment): effect of sequence context on lesion bypass. *Biochemistry 39,* 10431-10438.

16. Zhuang, P., Kolbanovskiy, A., Amin, S., and Geacintov, N. E. (2001) Base sequence dependence of in vitro translesional DNA replication past a bulky lesion catalyzed by the exo(-) Klenow fragment of Pol I. *Biochemistry 40,* 6660-6669.

17. Shibutani, S., Margulis, L. A., Geacintov, N. E., and Grollman, A. P. (1993) Translesional synthesis on a DNA-template containing a single stereoisomer of DG-(+)anti-BPDE or DG-(-)-anti-BPDE (7,8-dihydroxy-anti-9,10-epoxy-7,8,9,10-tetrahydrobenzo [a] pyrene). *Biochemistry 32,* 7531-7541.

18. Alekseyev, Y. O., Dzantiev, L., and Romano, L. J. (2001) Effects of benzo [a] pyrene DNA adducts on *Escherichia coli* DNA polymerase I (Klenow fragment) primer-template interactions: evidence for inhibition of the catalytically active ternary complex formation. *Biochemistry 40,* 2282-2290.

19. Lipinski, L. J., Ross, H. L., Zajc, B., Sayer, J. M., Jerina, D. M., and Dipple, A. (1998) Effect of single benzo [alpha] pyrene diol epoxide-deoxyguanosine adducts on the action of DNA polymerases in vitro. *Int. J. Oncol. 13,* 269-273.

20. Kunkel, T. A., Pavlov, Y. I., and Bebenek, K. (2003) Functions of human DNA polymerases eta, kappa and iota suggested by their properties, including fidelity with undamaged DNA templates. *DNA Repair 2,* 135-149.

21. Prakash, S., Johnson, R. E., and Prakash, L. (2005) Eukaryotic translesion synthesis: specificity of structure and function. *Annu. Rev. Biochem. 74,* 317-353.

22. Kanuri, M., Finneman, J., Harris, C. M., Harris, T. M., and Lloyd, R. S. (2001) Efficient nonmutagenic replication bypass of DNAs containing beta-adducts of styrene oxide at adenine N-6. *Environ. Mol. Mutagen. 38,* 357-360.

23. Latham, G. J., McNees, A. G., DeCorte, B., Harris, C. M., Harris, T. M., Odonnell, M., and Lloyd, R. S. (1996) Comparison of the efficiency of synthesis past single bulky DNA adducts in vivo and in vitro by the polymerase III holoenzyme. *Chem. Res. Toxicol. 9,* 1167-1175.

24. Latham, G. J., Harris, C. M., Harris, T. M., and Lloyd, R. S. (1995) The efficiency of translesion synthesis past single styrene oxide DNA-adducts in vitro is polymerase-specific. *Chem. Res. Toxicol. 8,* 422-430.

25. Latham, G. J., Zhou, L., Harris, C. M., Harris, T. M., and Lloyd, R. S. (1993) The replication fate of R-styrene and S-styrene oxide adducts on adenine N6 is dependent on both the chirality of the lesion and the local sequence context. *J. Biol. Chem. 268,* 23427-23434.

26. Mitra, G., Pauly, G. T., Kumar, R., Pei, G. K., Hughes, S. H., Moschel, R. C., and Barbacid, M. (1989) Molecular analysis of O-6-substituted guanine-induced mutagenesis of Ras oncogenes. *Proc. Natl. Acad. Sci. U.S.A. 86,* 8650-8654.

27. Singer, B., Chavez, F., Goodman, M. F., Essigmann, J. M., and Dosanjh, M. K. (1989) Effect of 3' flanking neighbors on kinetics of pairing of dCTP or dTTP opposite O-6-methylguanine in a defined primed oligonucleotide when *Escherichia coli* DNA-polymerase-I is used. *Proc. Natl. Acad. Sci. U.S.A. 86,* 8271-8274.

28. Snow, E. T., Foote, R. S., and Mitra, S. (1984) Base-pairing properties of O-6-methylguanine in template DNA during in vitro DNA-replication. *J. Biol. Chem. 259,* 8095-8100.

29. Ellison, K. S., Dogliotti, E., Connors, T. D., Basu, A. K., and Essigmann, J. M. (1989) Site-specific mutagenesis by O-6-alkylguanines located in the chromosomes of mammalian-cells-influence of the mammalian O-6-alkylguanine-DNA alkyltransferase. *Proc. Natl. Acad. Sci. U.S.A. 86,* 8620-8624.

30. Abbotts, J., Sengupta, D. N., Zmudzka, B., Widen, S. G., Notario, V., and Wilson, S. H. (1988) Expression of human DNA polymerase-beta in *Escherichia coli* and characterization of the recombinant enzyme. *Biochemistry 27,* 901-909.

35. Eckert, K. A., and Hile, S. E. (1998) Alkylation-induced frameshift mutagenesis during in vitro DNA synthesis by DNA polymerases alpha and beta. *Mutat. Res. 422,* 255-269.

36. Yasui, M., Matsui, S., Ihara, M., Laxmi, Y. R. S., Shibutani, S., and Matsuda, T. (2001) Translesional synthesis on a DNA template containing N-2-methyl-2'-

deoxyguanosine catalyzed by the Klenow fragment of *Escherichia coli* DNA polymerase I. *Nucleic Acids Res. 29,* 1994-2001.

37. Terashima, I., Matsuda, T., Fang, T. W., Suzuki, N., Kobayashi, J., Kohda, K., and Shibutani, S. (2001) Miscoding potential of the N-2-ethyl-2'-deoxyguanosine DNA adduct by the exonuclease-free Klenow fragment of *Escherichia coli* DNA polymerase I. *Biochemistry 40,* 4106-4114.

38. Kim, M.-S., and Guengerich, F. P. (1998) Polymerase blockage and misincorporation of dNTPs opposite the ethylene dibromide-derived DNA adducts *S*-[2-(*N*7-guanyl)ethyl]glutathione, *S*-[2-(N^2-guanyl)ethyl]glutathione, and *S*-[2-(O^6-guanyl)ethyl]glutathione. *Chem. Res. Toxicol. 11*, 311-316.

39. Lewis, W., Meyer, R. R., Simpson, J. F., Colacino, J. M., and Perrino, F. W. (1994) Mammalian DNA polymerase-alpha, polymerase-beta, polymerase-gamma, polymerase-delta, and polymerase-epsilon incorporate Fialuridine (Fiau) monophosphate into DNA and are inhibited competitively by Fiau triphosphate. *Biochemistry 33,* 14620-14624.

40. Perrino, F. W., and Loeb, L. A. (1989) Differential extension of 3' mispairs is a major contribution to the high fidelity of calf thymus DNA polymerase-α. *J. Biol. Chem. 264,* 2898-2905.

41. Trincao, J., Johnson, R. E., Escalante, C. R., Prakash, S., Prakash, L., and Aggarwal, A. K. (2001) Structure of the catalytic core of *S. cerevisiae* DNA polymerase eta: implications for translesion DNA synthesis. *Mol. Cell 8,* 417-426.

42. Washington, M. T., Johnson, R. E., Prakash, L., and Prakash, S. (2003) The mechanism of nucleotide incorporation by human DNA polymerase eta differs from that of the yeast enzyme. *Mol. Cell. Biol. 23,* 8316-8322.

43. Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L., and Aggarwal, A. K. (2004) Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing. *Nature 430,* 377-380.

44. Kaguni, L. S., and Clayton, D. A. (1982) Template-directed pausing in in vitro DNA-synthesis by DNA polymerase-alpha from *Drosophila melanogaster* embryos. *Proc. Natl. Acad. Sci. U.S.A. 79,* 983-987.

45. Weismanshomer, P., Dube, D. K., Perrino, F. W., Stokes, K., Loeb, L. A., and Fry, M. (1989) Sequence specificity of pausing by DNA-polymerases. *Biochem. Biophys. Res. Commun. 164,* 1149-1156.

46. Johnson, R. E., Washington, M. T., Prakash, S., and Prakash, L. (2000) Fidelity of human DNA polymerase eta. *J. Biol. Chem. 275,* 7447-7450.

47. Zhang, Y. B., Yuan, F. H., Wu, X. H., Taylor, J. S., and Wang, Z. G. (2001) Response of human DNA polymerase iota to DNA lesions. *Nucleic Acids Res. 29,* 928-935.

48. Haracska, L., Prakash, L., and Prakash, S. (2003) A mechanism for the exclusion of low-fidelity human Y-family DNA polymerases from base excision repair. *Genes Dev. 17,* 2777-2785.

49. Creighton, S., Bloom, L. B., and Goodman, M. F. (1995) Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *DNA Replication 262*, 232-256.

50. Matsuda, T., Bebenek, K., Masutani, C., Hanaoka, F., and Kunkel, T. A. (2000) Low fidelity DNA synthesis by human DNA polymerase-eta. *Nature 404,* 1011-1013.

51. Tissier, A., McDonald, J. P., Frank, E. G., and Woodgate, R. (2000) Pol iota, a remarkably error-prone human DNA polymerase. *Genes Dev. 14,* 1642-1650.

52. Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S., and Prakash, L. (2000) Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. *Nature 406,* 1015-1019.

53. Zhang, Y. B., Yuan, F. H., Wu, X. H., and Wang, Z. G. (2000) Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase iota. *Mol. Cell. Biol. 20,* 7099-7108.

54. Wang, J., Sattar, A., Wang, C. C., Karam, J. D., Konigsberg, W. H., and Steitz, T. A. (1997) Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. *Cell 89,* 1087-1099.

55. Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., and Angerer, B. (1999) Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc. Natl. Acad. Sci. U.S.A. 96,* 3600-3605.

56. Zhao, Y. X., Jeruzalmi, D., Moarefi, I., Leighton, T., Lasken, R., and Kuriyan, J. (1999) Crystal structure of an archaebacterial DNA polymerase. *Structure 7,* 1189-1199.

57. Rodriguez, A. C., Park, H. W., Mao, C., and Beese, L. S. (2000) Crystal structure of a p7l alpha family DNA polymerase from the hyperthermophilic archaeon *Thermococcus sp. 9 degrees N-7. J. Mol. Biol. 299,* 447-462.

58. Lehmann, A. R. (2002) Replication of damaged DNA in mammalian cells: new solutions to an old problem. *Mutat. Res. 509,* 23-34.

59. Prakash, S., and Prakash, L. (2002) Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. *Genes Dev. 16,* 1872-1883.

60. Zhou, B. L., Pata, J. D., and Steitz, T. A. (2001) Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol. Cell 8,* 427-437.

61. Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001) Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell 107,* 91-102.

62. Silvian, L. F., Toth, E. A., Pham, P., Goodman, M. F., and Ellenberger, T. (2001) Crystal structure of a DinB family error-prone DNA polymerase from *Sulfolobus solfataricus*. *Nat. Struct. Biol. 8,* 984-989.

63. Uijon, S. N., Johnson, R. E., Edwards, T. A., Prakash, S., Prakash, L., and Aggarwal, A. K. (2004) Crystal structure of the catalytic core of human DNA polymerase kappa. *Structure 12,* 1395-1404.

64. Washington, M. T., Minko, I. G., Johnson, R. E., Wolfle, W. T., Harris, T. M., Lloyd, R. S., Prakash, S., and Prakash, L. (2004) Efficient and error-free replication past a minor-groove DNA adduct by the sequential action of human DNA polymerases iota and kappa. *Mol. Cell. Biol. 24,* 5687-5693.

65. Haracska, L., Prakash, S., and Prakash, L. (2000) Replication past O-6-methylguanine by yeast and human DNA polymerase eta. *Mol. Cell. Biol. 20,* 8001-8007.

CHAPTER 3

Thermal melting and thermodynamic analysis of stability of a DNA duplex containing the 2,7-DAM-G-monoadduct

3.1. INTRODUCTION

In order to study the stability of chemically modified deoxyoligonucleotides spectroscopical methods are primarily used either to gain details of high structural resolutions, as is the case of nuclear magnetic resonance (NMR) (1-4), either to analyze the thermodynamic parameters, such as enthalpy, entropy and free energy for the folded state of the chemically modified DNA duplex (5-21). We use UV spectroscopy to investigate the stability of a self-complementary DNA duplex **d(G-T-G-*G*(2,7-DAM)-T-A-T-A-C-C-A-C)** chemically alkylated with 2,7-DAM in order to determine its thermodynamic stability by performing reversible melting experiments. It is well known from other studies of DNA interaction with different drugs, (which are binding either reversible or chemically to specific bases) that the drug has primarily a destabilizing effect upon the DNA duplex, most of the time due to the decrease in the level of stacking interactions (1-10). Since we have already determined the NMR structure of the sel-complementary duplex **d(G-T-G-*G*(2,7-DAM)-T-A-T-A-C-C-A-C)** and we knew that the drug is oriented in the major groove, but without being intercalated within the DNA bases (19-Chapter 1) it was interesting to perform melting studies which could reveal the stability of the DNA duplex. This chapter presents the original contribution to the thermodynamic analysis of melting of a DNA duplex containing 2,7-DAM alkylated on N7-guanine. The duplex containing alkylated guanine had lower thermal stability and a

more favorable enthalpy but much less favorable entropy in the duplex state, relative to the nonadducted parent duplexes.

3.2. MATERIALS AND METHODS

3.2.1. UV Spectroscopy and Melting Studies.

Oligonucleotide concentrations were determined using formula: ε260(strand) = number of purines × 14 000 + number of pyrimidines × 7000. In the case of the alkylated oligonucleotide a correction for ε260(strand) was done by adding 5189 (ε260 of 2,7-DAM). ε260(duplex) was assumed as equal to 2 × ε260(strand). The ε_{260} for the duplex of the nonalkylated oligonucleotide was 252,000 M^{-1} cm^{-1} while the ε260 for the alkylated duplex was 262, 378 M^{-1} cm^{-1}. The concentration of the oligonucleotide (strand) was 2.2 μM in all melting experiments. The synthesis of the alkylated DNA self-complementary duplex used in melting studies was performed according to published data (18-Chapter 1). Thermal melting curves at 260 nm were determined using a Cary-3 spectrophotometer. Temperature control was obtained through a jacketed cell holder connected to a water bath and controlled by the software of the Cary-3. The samples were heated gradually in stoppered cuvettes of 1 cm path length. The temperature range was 20-85 °C. The data were acquired with 0.8 °C increments/min at each 15 s, such that a total of 260 data points were acquired for each sample. The melting curve data were fitted by sigma plot nonlinear regression analysis to a sigmoid curve. Three independent experiments were carried out on each sample, and the resulting T_ms were averaged. Immediately prior to analysis the samples were incubated at 60 °C for 3 min and then slowly cooled at room temperature to provide the annealing.

3.2.2. Thermodynamic Analysis of the Melting Curves. Melting temperature T_m and the thermodynamic parameters that characterize the helix-coil transition of the DNA duplexes were determined by employing the built-in Cary-3 software. The method of choice for the calculation of the thermodynamic parameters was direct application of the van't Hoff equation (*21*).

Our chosen method for the analysis was the direct application of the van't Hoff equation (22, 23). Using this analysis of the melting curves, an important assumption was made, that for a dodecamer duplex the melting process is bimolecular, corresponding to the duplex-coil melting transition. The Cary-3 software is creating first the curve $\alpha(T)$ (the fraction of the duplex that remains in the initial state, in this case in the duplex form, at temperature T. The plot $\alpha(T)$ versus T is generated using the slopes (m) and intercepts (b) from the fitted baselines as follows:

$\alpha(T) = [A_u(T)-A(T)]/[A_u(T)-A_L(T)]=[m_uT+b_u-A(T)]/[(m_u-m_L)T+b_u-b_L]$

where $A_u(T)= m_uT+b_u$ and $A_L(T) = m_LT + b_L$

The T_m is defined as the temperature at which $\alpha=0.5$. The association constant K(T) is calculated using the formula $K(T) = \alpha(T)/nC^{n-1}[1-\alpha(T)]^n$; were C represents the total concentration of the duplex (in this case for the self-complementary duplex was 1.1 μM for both free and the alkylated oligonucleotide. A plot of lnK(T) versus 1/T is used to determine $\Delta H^0 = -R\delta lnK(T)/ \delta(1/T)$. The free energy change associated with the melting transitions is also determined from the van't Hoff equation using the formula:

$\Delta G^0 = \Delta H^0 / R (1/T-1/T_m) + (n-1)ln(nC/2)$. The values for the entropy change is determined using the formula $\Delta G^0 = \Delta H^0 - T_m\Delta S$. In these equations "n" represents the

number of molecules (strands) in equilibrium with the duplex. In the case of the bimolecular process, n=2.

Independent of the built-in software, the Nonlinear Regression analysis of the Sigma-Plot software was used to fit the melting curves to a sigmoid curve with 4 parameters [$f=Y_{(0)} + a/[1+\exp(-(X-X_{(0)})/b]$ where the independent and the dependent variables are as follows:

$X_{(0)}$= the value of X for the Y value 50% of the distance from the minimum to the maximum of smoothed data for the sigmoid shaped curve.

$Y_{(0)}$ = min (Y) =minimum of Y of the smoothed data for the sigmoid shaped curve.

A = max (y) – min (y) = the distance from the minimum to the maximum of smoothed data for the sigmoid shaped curve.

$b = X_{(75)}-X_{(25)}$ = the region of the sigmoid curve where Y varies linearly with X.

$X_{(75)}$= the X value for the Y value 75% of the distance from the minimum to the maximum of smoothed data for a sigmoid curve.

$X_{(25)}$ = the X value for the Y value 25% of the distance from the minimum to the maximum of smoothed data for a sigmoid curve.

The statistical analysis of Sigma Plot was used to assess the goodness of the fit. Each independent fit was used to determine Tm. The Tm values determined from Nonlinear Rgression of the melting curves were within 1% errorwith those determined using the software provided by Cary-UV spectrophotometer assuming a two-state transition model for the melting curves.

3.4. RESULTS

3.4.1. Synthesis of 2,7-Diaminomitosene Adducts with d(G-T-G-G-T-A-T-A-C-C-A-C) for melting studies.

We synthesized the 2,7-diaminomitosene adduct with d(G-T-GG(4)-T-A-T-A-C-C-A-C) (duplex III) and purified it using reverse-phase C18-HPLC. The G4-position as the adduct site was established as described in the Materials and Methods section 3.2 (see Appendix VIII for HPLC analysis of oligonucleotide d(G-T-G-G(4)-T-A-T-A-C-C-A-C)) . This adduct complemented with itself as observed by imino proton resonances in the NMR spectrum run in H_2O (19-Chapter 1). The self-complementary duplex was used in the melting studies.

3.4.2. Melting Studies and Thermodynamic Analysis of the Melting Curves. One potential problem that arises in the case of guanyl-N7 adducts is sensitivity toward thermal depurination (1-18, 28). To minimize the experimental artifacts due to this effect, melting experiments were conducted such that the temperature increase was relatively fast (0.80 °C min^{-1}) to avoid excessive incubation at each temperature step, although in a range acceptable for short oligonucleotide duplexes. Figure 1 presents the melting curves of the control duplex and 2,7-DAM-adducted duplex III. Depurination was not significant, as seen by their complete reversibility and by HPLC analysis of the melted sample (data not shown). The melting temperatures are lower for the alkylated duplex III (two adducts per duplex) as compared with the control, with ΔT_m = 8.90 ± 0.70 °C. The reduction in the thermal stability of other nonintercalating N7-guanyl adducts is in the range of ~11 °C (*27*), which is not far from that found for the 2,7-DAM adduct. The hyperchromicity for the nonadducted duplex was

15.5% while that for the adducted duplex III was 7.6%. The lower hyperchromicity of the adducted duplex could be due to helical perturbations centered about the adducted bases. The shape of the melting curves is compatible with a two-state cooperative transition between the helix and the coil states both for the nonalkylated and for the adducted duplexes. The results of van't Hoff analysis of the melting curves based on this assumption are presented in Table 1. The transition enthalpy for the adducted duplex III facilitates duplex formation by -29.4 kcal/mol over that of the parent counterpart, while the transition entropy is less favorable, as a difference of -98 cal mol^{-1} K^{-1} was observed. In the case of other major groove-located guanyl-N7 adducts the thermodynamic analysis of melting curves also showed that enthalpy of the alkylated duplex is more favorable, while the entropy is less favorable, suggesting that the decrease in the thermal stability of the adducted duplex is due to the entropic component (27, 28). It has been also speculated that the more favorable coil-to-helix transition enthalpy of the adducted duplex may be due to the positive charge imposed on the imidazole ring by N7-guanine alkylation (28). The ΔG of the coil-to-helix transition of the adducted duplex is only slightly lower than that of the parent duplex ($\Delta\Delta G^{\circ}_{25\,^{\circ}C}$ = -0.3 kcal/mol); this is also analogous to the other G-N7 alkylated oligonucleotides (27, 28).

Table 5: Effect of the 2,7-DAM-Guanine-N7 Adduct 4 in the Self-Complementary d(GTGGTATACCAC) Adduct Duplex 5 on Thermal and Thermodynamical Duplex Stability, Spectroscopically Determined T_m, and Calculated Changes of the van't Hoff Transition Enthalpy (ΔH°), Entropy (ΔS°), and Free Energy ($\Delta G_{25}{}^{\circ}$) for the Coil-to-Helix Transition[a]

Oligonucleotide	T_m ($^{\circ}$C)	ΔH° (kcal/mol)	ΔS° (cal mol^{-1} K^{-1})	$\Delta G^{\circ}_{25\,^{\circ}C}$ (kcal/mol)
d(GTGGTATACCAC)	51.0 ± 0.4	-49.2 ± 2.0	-125.7 ± 5.9	-11.7 ± 0.2
d(GTG[M]GTATACCAC)[b] (**III**)	42.1 ± 0.4	-78.6 ± 6.1	-223.2 ± 1.9	-12.1 ± 0.3

[a] The thermodynamic parameters were extracted from the melting curves (figure 1) by direct application of the van't Hoff equation (*21*) using the Cary-3 software.[b] [M]G denotes the adducted guanine residue (see also Appendix IX for details of thermodynamic analysis using the built-in software provided by Cary-UV-Visible spectrophotemer).

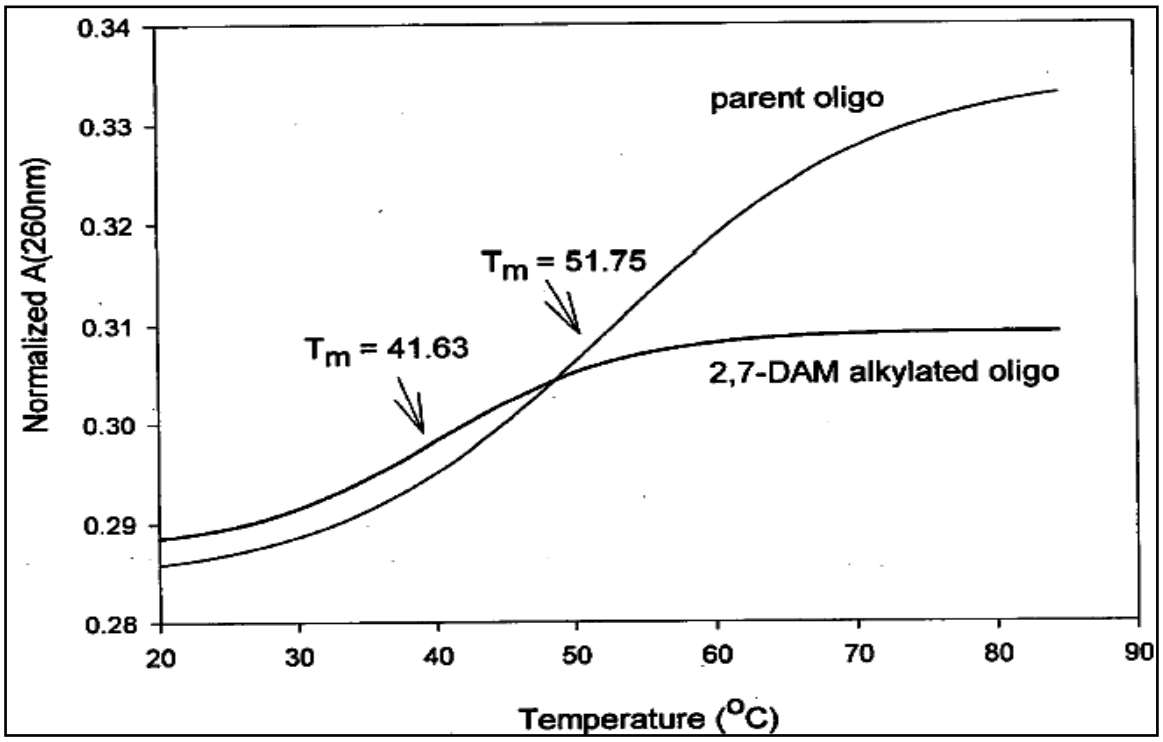


Figure 90: The melting curves for the parent and the alkylated (G-T-G-G(2,7-DAM)-T-A-T-A-C-C-A-C). The details for running of melting studies are described in specific methods section (see also Appendix IX for thermodynamic analysis of melting curves using the Cary-UV-Visible built-in software).

3.5. DISCUSSIONS AND CONCLUSION

This original study of thermal stability of the 2,7-DAM-DNA duplex **III** (Figure 67 and Table 5) reveals properties of adduct duplex **III** strikingly similar to those of duplexes containing simpler nonintercalating guanine-N7 alkyl groups: lower thermal stability and a more favorable enthalpy but much less favorable entropy in the duplex state, relative to the nonadducted parent duplexes (28). On the other hand, the guanine-N7-aflatoxin B_1 adduct,

in which the drug moiety is intercalated, shows increased thermal stability, characteristic of intercalated adducts (29). A structural feature of the adduct duplex **III**, namely, confinement of the G-N7 (+) electrostatic interaction in the groove of the duplex (see reference 18 and figure 2 in Chapter 1), is consistent with the observed decrease of transition entropy resulting in decreased thermal stability over that of the parent duplex.

APPENDIX VIII

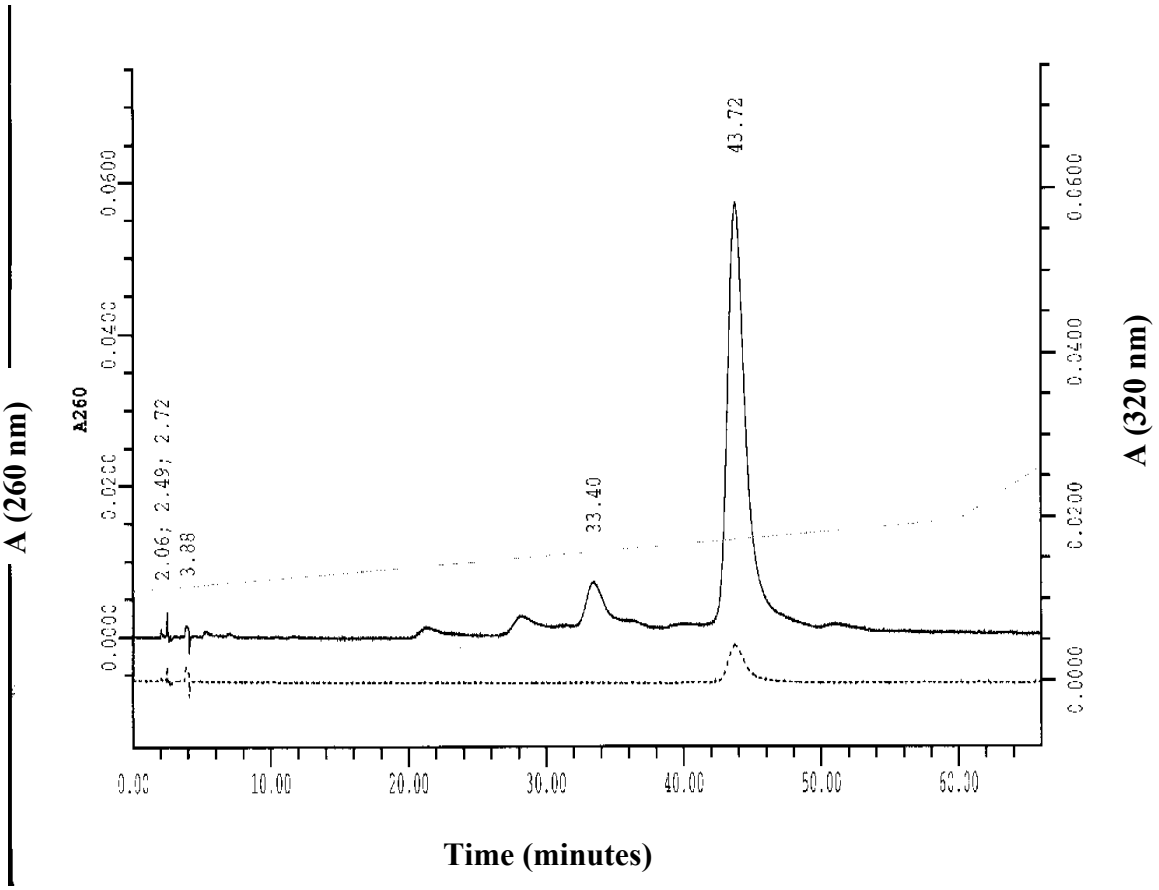


Figure 91. HPLC analysis of self-complementary oligonucleotide d(G-T-G-G(4)-T-A-T-A-C-C-A-C)) used in thermal melting studies

APPENDIX IX

Table 6: Thermodynamic analysis of melting curves using the Cary-UV-Visible built-in software. (Tm is the melting temperature for the nonalkylated (nonalk) or for the alkylated DNA (alk)).

Table 1

Data 1

	-1- Tmnonalk	-2- Tmalk	-3- dHnonalk	-4- dHalk	-5- dSnonalk	-6- dSalk	-7- dGnonalk	-8- dGalk
1	51.75	41.63	-50.03	-74.65	-127.90	-210.90	-49.91	-49.25
2	50.64	41.85	-45.48	-90.61	-114.60	-261.40	-47.33	-53.12
3	50.77	42.78	-52.13	-70.49	-134.60	-197.30	-50.20	-48.88

Table 2

Statistics - Data 1

	-1- Tmnonalk	-2- Tmalk	-3- dHnonalk	-4- dHalk	-5- dSnonalk	-6- dSalk	-7- dGnonalk	-8- dGalk
Mean	51.0533	42.0867	-49.2133	-78.5833	-125.7000	-223.2000	-49.1460	-50.4163
Std.Dev	0.6068	0.6104	3.3994	10.6211	10.1799	33.7738	1.5826	2.3443
Std.Err	0.3503	0.3524	1.9626	6.1321	5.8774	19.4993	0.9137	1.3535
95% Conf	1.5043	1.5132	8.4269	26.3290	25.2353	83.7232	3.9233	5.8115
99% Conf	3.4267	3.4471	19.1963	59.9771	57.4858	190.7206	8.9372	13.2385
Size	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000
Total	153.1600	126.2600	-147.6400	-235.7500	-377.1000	-669.6000	-147.4380	-151.2490
Min	50.6400	41.6300	-52.1300	-90.6100	-134.6000	-261.4000	-50.1990	-53.1150
Max	51.7500	42.7800	-45.4800	-70.4900	-114.6000	-197.3000	-47.3260	-48.8830
Min.Pos	50.6400	41.6300	--	--	--	--	--	--
Missing	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Other	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

REFERENCES

1. Seo, K.Y., Nagalingam, A., Tiffany, M., Loechler, E.L. (2005) Mutagenesis studies with four stereoisomeric N^2-dG benzo[a]pyrene adducts in the identical 5'-CGC sequence used in NMR studies: G→T mutations dominate in each case *Mutagenesis*, **20**, 441–448.

2. Cho, B.P. (2004) Dynamic conformational heterogeneities of carcinogen–DNA adducts and their mutagenic relevance *J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev*, . **22**, 57–90.

3. Patel, D.J., Mao, B., Gu, Z., Hingerty, B.E., Gorin, A., Basu, A.K., Broyde, S. (1998) Nuclear magnetic resonance solution structures of covalent aromatic amine–DNA adducts and their mutagenic relevance *Chem. Res. Toxicol*, . **11**, 391–407.

4. Isaacs, R.J. and Spielmann, H.P. (2004) A model for initial DNA lesion recognition by NER and MMR based on local conformational flexibility *DNA Repair*, **4**, 455–464 .

5. Geacintov, N.E., Broyde, S., Buterin, T., Naegeli, H., Wu, M., Yan, S., Patel, D.J. (2002) Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery *Biopolymers*, **65**, 202–210.

6. Arghavani, M.B., SantaLucia, J., Romano, L.J. (1998) Effect of mismatched complementary strands and 5'-change in sequence context on the thermodynamics and structure of benzo[*a*]pyrene-modified oligonucleotides *Biochemistry*, **37**, 8575–8583.

7. Marky, L.A. and Breslauer, K.J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves *Biopolymers*, **26**, 1601–1620.

8. Tanaka, F., Kameda, A., Yamamoto, M., Ohuchi, A. (2004) Thermodynamic parameters based on a nearest-Neighbor model for DNA sequences with a single-bulge loop *Biochemistry*, **43**, 7143–7150.

9. Allawi, H.T. and SantaLucia, J. (1998) Thermodynamics of internal C.T mismatches in DNA *Nucleic Acid Res*, . **26**, 2694–2701.

10. Brown, K., Harvey, C.A., Turteltaub, K.W., Shields, S.J. (2003) Structural characterization of carcinogen-modified oligodeoxynucleotide adducts using matrix-assisted laser desorption/ionization mass spectrometry *J. Mass Spectrom*, **38**, 68–79.

11. Sagi, J., Hang, B., Singer, B. (1999) Sequence-dependent repair of synthetic AP sites in 15-mer and 35-mer oligonucleotides: role of thermodynamic stability imposed by neighbor bases *Chem. Res. Toxicol*, . **12**, 917–923.

12. Sagi, J., Guliaev, A.B., Singer, B. (2001) 15-mer DNA duplexes containing an abasic site are thermodynamically more stable with adjacent purines than with pyrimidines *Biochemistry*, **40**, 3859–3868.

13. Van Houte, L.P., Westra, J.G., Retel, J., Van Grondelle, R. (1988) A spectroscopic study of the conformation of poly d(G-C).poly d(G-C) modified with the carcinogenic 2-aminofluorene *Carcinogenesis*, **9**, 1017–10127.

14. Van Houte, L.P.A., Bokma, J.T., Lutgerink, J.T., Westra, J.G., Retel, J., Van Grondelle, R., Blok, J. (1987) An optical study of the conformation of the aminofluorene–DNA complex *Carcinogenesis*, **8**, 759–766.

15. Van Houte, L.P.A., Westra, J.G., Retel, J., Van Grondelle, R. (1991) A circular dichroism study on the conformation of d(CGT) modified with N-acetyl-2-aminofluorene or 2-aminofluorene *J. Biomol. Struct. Dyn*, **9**, 45–59.

16. Marky, L.A., Rentzeperis, D., Luneva, N.P., Cosman, M., Geacintov, N.E., Kupke, D.W. (1996) Differential hydration thermodynamics of stereo isomeric DNA-benzo[*a*]pyrene adducts derived from diol epoxide enantiomers with different tumorigenic potentials *J. Am. Chem. Soc*, . **118**, 3804–3810.

17. Sagi, J., Chenna, A., Hang, B., Singer, B. (1998) A single cyclic *p*-benzoquinone adduct can destabilize a DNA oligonucleotide duplex *Chem. Res. Toxicol*, . **11**, 329–334.

18. Gelfand, C.A., Plum, G.E., Grollman, A.P., Johnson, F., Breslauer, K.J. (1998) The impact of an exocyclic cytosine adduct on DNA duplex properties: significant thermodynamic consequences despite modest lesion-induced structural alterations *Biochemistry*, **37**, 12507–12512.

19. Hang, B., Sagi, J., Singer, B. (1998) Correlation between sequence-dependent glycosylase repair and the thermal stability of oligonucleotide duplexes containing 1, N^6-ethenoadenine *J. Biol. Chem*, . **273**, 33406–33413.

20. Sagi, J., Perry, A., Hang, B., Singer, B. (2000) Differential destabilization of the DNA oligonucleotide double helix by a T·G mismatch, 3,N^4-ethenocytosine, 3,N^4-ethanocytosine, or an 8-(hydroxymethyl)-3,N^4-ethenocytosine adduct incorporated into the same sequence contexts *Chem. Res. Toxicol*, . **13**, 839–845.

21. Gelfand, C.A., Plum, G.E., Grollman, A.P., Johnson, F., Breslauer, K.J. (1998) Thermodynamic consequences of an abasic lesion in duplex DNA are strongly dependent on base sequence *Biochemistry*, **37**, 7321–7327.

22. Puglisi, J. D., and Tinoco, I., Jr. (1989) *Methods Enzymol. 180*, 304-325.

23. Plateau, P., and Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310-7311.

24. Gopalakrishnan, S., Stone, M. P., and Harris, T. M. (1989) *J. Am. Chem. Soc. 111*, 7232-7239.Singer, B., and Grunberger, D. (1983) *Molecular Biology of Mutagens and Carcinogens*, pp 15-44, Plenum Press, New York.

25. Boudsocq F; Iwai S; Hanaoka F; Woodgate R. *Nucleic acids research*, (2001), Nov 15), 29(22), 4607-16.

26. Hartley, J. A., and Wyatt, M. D. (1997). In *Drug-DNA Interaction Protocols, Methods in Molecular Biology* (Fox, K. R., Ed.) Vol. 90, pp 147-156, Humana Press, Totowa, NJ.

27. Gopalakrishnan, S., Harris, T. M., and Stone, M. P. (1990). *Biochemistry 29*, 10438-10448.

28. Bailey, E. A., Iyer, R. S., Stone, M. P., Harris, T. M., and Essigmann, J. M. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 1535-1539.

29. Gralla J. and Crothers D. M. (1973). *J. Mol. Biol.*, 78, 301-319.

PART II

Structure-based design, synthesis, structure-conformation and structure-activity relationships studies of D-Phe-Pro-D-Arg-P1'-$CONH_2$ tetrapeptides with inhibitory activity for thrombin

ABSTRACT

This original research part II of the thesis presents the structure-activity relationship (SAR) of tetrapetides from series D-Phe-Pro-DArg-P1'-CONH2 with reversible inhibitory activity toward thrombin. The P1' position was varied with D and L amino acids. The significant differences between the inhibitory constants (Kis) of tetrapeptides from the series D-Phe-Pro-D-Arg-P1'-CONH2 suggest that the interaction between the amino acid at P1' position and the S1' subpocket in thrombin is very specific. There is a 2 to 500 fold experimentally determined difference between the Kis of different peptide inhibitors and our in vitro inhibition assay for thrombin proved that the P1' position requires small hydrophobics and polar amino-acids. In addition, there is a significant change in the affinity for interaction with thrombin as the configuration for the same amino-acid in P1' position is changing from L to D. Specifically, a switch from L-Thr into D-Thr in P1' was correlated with a 13 fold increase in the inhibitory activity. Similarly, a switch from L-Ala to D-Ala in P1' was increasing the affinity 8 fold. These differences in the binding affinities upon switching from L into D of amino-acids in P1' were confirmed both kinetically and through isothermal titration calorimetry (ITC). In the case of ITC the heat released upon binding to thrombin for peptides with L-Ala in P1' is lower,1.2 kcal/mole of injectant, than the heat released during the titration of the same peptide into thrombin but with D-Ala in P1', 6.5 kcal/mole of injectant. The structural basis for this favorable switch in the affinity was further investigated through molecular modeling of docked peptides into the thrombin template 1ABJ.pdb. The results from molecular modeling suggest that the D-Thr in P1' forms different hydrogen bonds with specific amino-acids from the insertion loop and the S1' subpocket of thrombin which are

lacking in the analog peptide containing L-Thr in P1'. The extra hydrogen bonds between the peptide ligand and the amino acids side chains from the active site of thrombin might be responsible for the observed increased affinity for peptide containing D-amino acids in P1'. These results strongly support our original structure-based design of peptides as reversible inhibitors for thrombin and provides a new lead peptide inhibitors which can be used to design further more potent anticoagulants.

INTRODUCTION

The vascular system, in the healthy state, is under the strict homeostatic control of the coagulation system. The soluble component of the coagulation system, known as the coagulation cascade, is comprised of a series of soluble proteases and their regulatory cofactors (Figure 92). Alpha-thrombin is the last enzyme of this cascade that converts fibrinogen to fibrin I and fibrin II by limited proteolysis. Alpha-thrombin is generated by the proteolytic cleavage of prothrombin by the prothrombinase complex which includes factor V, a phospholipid surface, and factor Xa, the serine-protease immediately preceding thrombin in the coagulation cascade. In the presence of Ca^{2+}, alpha-thrombin is also involved in the activation of factor XIII to generate factor XIIIa. Factor XIIa stabilizes the blood clot by cross-linking fibrin through N^{ε}-(γ-glutamyl)lysine transamide bridges. The polymerized fibrin forms the general scaffold for the growth of the thrombus.

The enzymatic activity of thrombin is known to be crucial not only for the arrest of blood flow but also for the ensuing growth of the thrombus. Therefore, the inhibition of the enzymatic activity of alpha-thrombin is considered a viable mechanistic approach for the development of anticoagulant inhibitors (figure 92, (1-13)).

Alpha-thrombin is a prolate ellipsoidal glycoprotein molecule made up of a short A chain (36 residues) and a long B chain (259 residues) cross-linked by one disulfide bond. Amino acids Ser195, His57, and Asp102 (chymotrypsin numbering system) form a catalytic triad, which is integral to the catalytic activity of all serine proteases. These residues are located at the entrance to the substrate binding pocket, and their geometry is stabilized by hydrogen bonds. Serine proteases hydrolyze peptide bonds via the formation of tetrahedral state intermediates. Stabilization of the transition state

intermediate occurs through the formation of hydrogen bonds between the oxyanion intermediate and the amido groups of residues Gly193 and Ser195. The substrate binding sites in the enzyme involved in precise interactions are referred to as Sn, …S3, S2, S1, S1', S2', S3', …Sn' sites, and the amino acid residues of the substrate or inhibitor that occupy these sites are referred to as Pn, …P3, P2, P1, P1', P2', P3',…Pn', respectively. These complementary sites were shown through many X-Ray structures of thrombin with peptide-derived inhibitors to allow the specific alignment of the substrate/inhibitor with the catalytic triad and the oxyanion hole for enzymatic specificity (14, 15). In trypsin-like serine proteases, Asp189 is at the bottom of the primary S1 susbstrate binding site, and forms a salt bridge with the guanidino group of P1 Arg residues in the substrate/inhibitor. Thrombin is a serine-protease with Arg-Xxxx or Lys-Xxxx specificity at P1 position within the substrate sequence H2N…-P3-P2-P1-P1'-P2'-P3'-…COOH with preference for arginine (were P1-P1' represents the scissile bond). The molecular basis for the enhanced affinity for arginine has been elucidated by Bode et al. (14). A large majority of the active site-directed synthetic inhibitors contain guanidino or an amidino groups to mimic the interaction of the natural substrate with the enzyme (15-20). The charged guanidine or amidino groups form strong ionic interactions with the carboxylate group of Asp189 in the specificity pocket of the enzyme (S1) (14, 17-20).

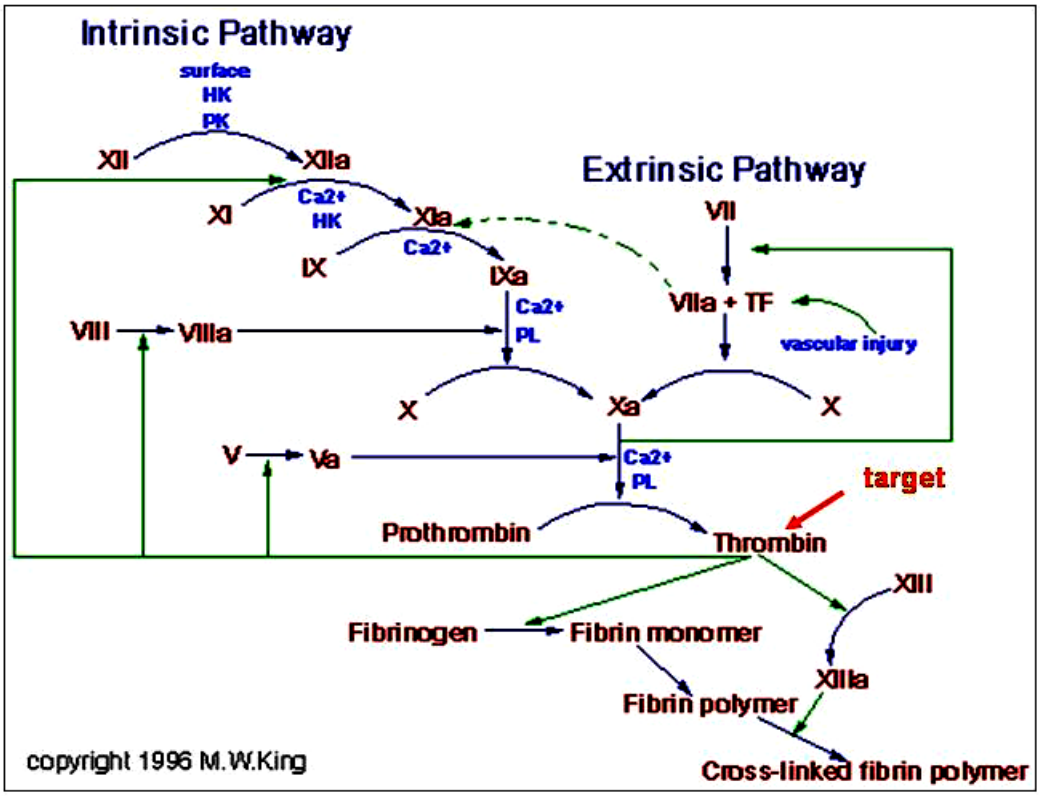


Figure 92: The coagulation cascade showing the intrinsic and the extrinsic pathways together with the most important serine-proteases involved in blood clotting (5).

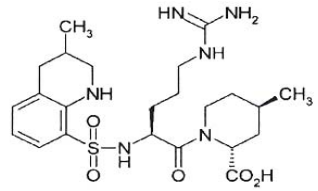
As can be seen from Figure 92 thrombin is the last in a cascade of trypsin-like plasma serine-proteases, which by catalyzing the conversion of fibrinogen to fibrin, activation of FXIII and inducing platelets aggregation is a key enzyme in haemostasis and thrombus formation.

It has been a long term goal in antithrombotic therapy to design and discover drugs that could inhibit thrombin directly and could provide better efficacy and safety as compared with other antithrombotic drugs such as heparin and coumarins (16). Potent and selective inhibitors of thrombin were originally designed based on the D-Phe-Pro-Arg sequence. Two major small molecular weigh inhibitors were emerging as lead in the structure-activity relationship for thrombin. The first is of the argatroban and NAPAP

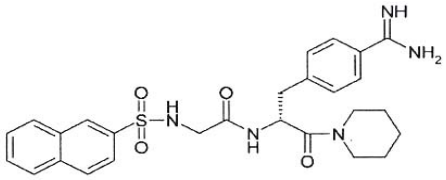
type (Scheme 7). The reactivity of thrombin toward peptides of arginine chloromethyl ketone and peptide substrates is also highly dependent on the nature of the amino acids residues in the flanking peptide sequence (P2 and P3 binding sites). The most effective inhibitor of thrombin in the chloromethyl ketone series is H-(D)Phe-Pro-ArgCH_2Cl, which, when compared with other arginine chloromethyl ketones, indicates differences in reactivity greater than 4 orders in magnitude obtained by variations in amino acid residues in the P2 and P3 residues. In the above mentioned sequence ((D)Phe-Pro-Arg) the P1=Arg, P2=Pro, and P3=D(Phe) (15-20).

Another well known class of inhibitors of alpha-thrombin are peptides containing α-aminoboronic acids with neutral side chains. These boronic acid based inhibitors are highly effective, slow binding, having association constants in the order of pM range. One of the most known boronic acid derivative with anti-coagulant activities is Ac-(D)Phe-Pro-boroArg-OH. The mechanism of inhibition by boronic acids has been already elucidated, and involves the formation of tetrahedral complex with the active site serine in a manner analogous to that expected for the tetrahedral intermediate formed during normal substrate hydrolysis. However, the synthesis of boronic acids peptides requires a complex procedure, and only few laboratories in the world are involved in this synthesis, making the cost of these anti-coagulant drugs being high (21-26). Also, clinical administration of these drugs have shown secondary effects, like severe hypotension in some cases and respiratory depression (21).

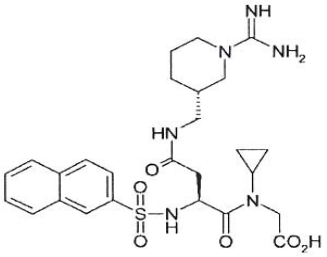
Scheme 7: Major types of thrombin inhibitors (14-19).



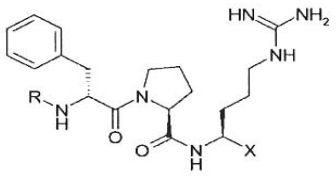
argatroban



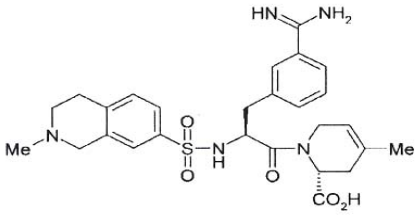
NAPAP



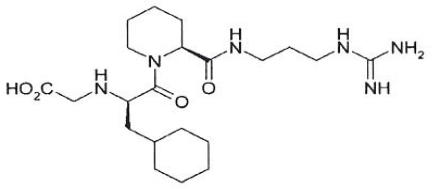
napsagratran



PPACK R = H; X = -$COCH_2Cl$
DuP-714 R = Ac; X = -$B(OH)_2$
efegatran R = Me; X = -CHO



UK-156,406



inogatran

One of the most recently developments in the field of rational drug design with anti-coagulant activities involves the molecular design of peptides or small molecules that can reversibly bind to the active site of thrombin or other coagulation factors (mostly targeted being factors VIIa, IXa, Xa) and thus can have a competitive type of inhibition (27-29). The struggle is for designing and development of reversible inhibitors with nanomolar and even picomolar association binding constants. These peptides which mimics the natural substrate have the advantage that are soluble in H_2O and thus can pass different physiological barriers, as compared with mostly small organic molecules or other small

derivatives of peptides with anti-coagulant properties (like chloromethyl ketones) which are less soluble in H_2O and thus their utilization is limited (28-31).

One of the most powerful tool in finding small peptides (less then 15-20 amino acids) as possible reversible inhibitors of thrombin or other targeted coagulation factors is the commercialized phage display libraries. The advantage of using the commercialized kit of the phage display is that multiple libraries can be screen fast such that the probability of finding a sequence of amino acids with effective inhibition toward these class of serine proteases is increased (32-38). One very important aspect of this approach is that if no specific selection is applied, the probability of finding active-site directed binding peptides is decreased (38-40).

The disadvantage of using such phage-display libraries is that no D-amino acids can be incorporated in these peptides naturally displayed by the phage, so the sequence "space" that can be actively involved in the inhibition process is limited. Thus, Phage Display Libraries can be used in parallel with the molecular design or structure-based design (SBD) strategies that are using docking experiments coupled with combinatorial chemistry approaches (41-45). Using synthetic methods the amino acids sequences containing D-amino acids can be further explored for their potential anti-coagulant activities (46-53).

The **Structure-Activity-Relationship (SAR)** for designing and development of peptides with anti-coagulant activities can be presented schematically in the Figure 93. The major goal of this research was focused on designing and development of new peptides reversible inhibitors for thrombin. Our major experimental approach involved a combination between computational modeling (using the docking software **"SCULPT"**),

combinatorial chemistry applied to solid phase synthesis of peptide libraries and in vitro screening of peptides for their inhibitory activity toward thrombin.

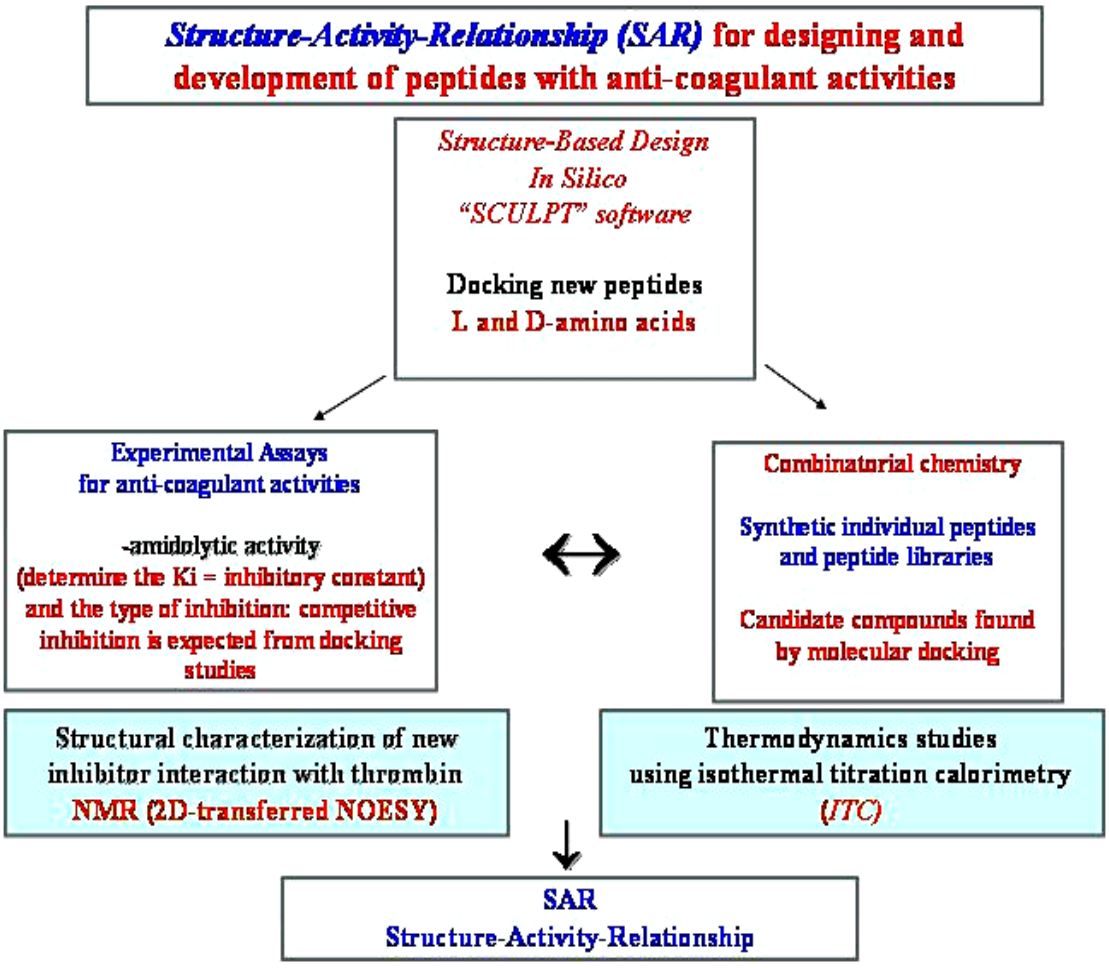


Figure 93: Strategy employed for finding peptides inhibitors of thrombin.

Based on the strategy of discovering of peptides inhibitors presented in figure 2 molecular modeling provides the information for chemical synthesis of peptides and experimental evaluation of the inhibitory constant. Initially, a defined criteria (such as the predicted free energy of interaction between thrombin and different peptides) was used to generate "leading compounds" groups from molecular docking experiments. This process can work in parallel with the selection of phage display libraries from where a

specific peptide sequence can be chosen to increase the reliability of the computational modeling. This sequence is supposed to prove a very good inhibitory potential (lower μM – nM range). The major criteria for selection of leading compounds from computational modeling is the based on the predicted free energy (ΔG) of interaction between the designed peptides and the enzyme of interest (like thrombin or other coagulation factors). The free energy criteria assumed that energies which are below "–50 kcal/mol" are predicting a very tight interaction between the docked ligand and the protein of interest (in the order of nM and pM range) (54-58). However, a critique evaluation of these values of the predicted free energies of interaction should be considered since the accuracy of this prediction varies from one software to another and depends on the type of force-field is implemented in the docking software (54-65). It is well known from the field of structure-based design (SBD) that the initial docking experiments are based on the sterical fitness between the ligand and the active site of the targeted enzymes. Thus, the van der Waals and Lenard Jones potentials dominate the free energy functions which are used to describe the binding of ligands to the protein targets during first stages of "crude" docking (55, 57, 60).

The X-Ray crystallographic studies of human alpha-thrombin complexed with hirudin and synthetic peptide inhibitors have revealed many details of thrombin-inhibitor interactions at the molecular level. For example hirudin, produced in the salivary glands of the blood sucking leech ***Hirudo medicinalis,*** prevents the clotting of ingested blood and is the most potent natural thrombin inhibitor with a Kd reported as low as 20fM. Hirugen and hirulogs are examples of peptide inhibitors designed to mimic thrombin-hirudin interactions. The crystal structure of an enzyme / ligand complex is a very good target for the design of new inhibitors. However, the possible lack of correspondence

between the crystal structure and the dynamic structure of the interacting species in solution, at the same time scale as the biochemical reactions is a major concern in inhibitor design (41, 42, 47, 57, 59).

It is important to use more than one single software in docking experiments till the validation of new leading group compounds is reliable (60). As an alternative, if one single docking software is used in the beginning of the *in silico* screening of a large number of compounds then the docking experiments have to be performed by using independent templates having high atomic resolution of their X-Ray solved structures (< 2.5 Å) (55, 59, 61).

This original research thesis presents the structure-based design rational of small peptides reversible inhibitors for thrombin together with the structure-activity relationship derived from in vitro kinetics of thrombin inhibition, thermodynamic analysis of binding of lead peptide to protein target and from more advanced structural investigations such as 2-D-transferred NOESY experiments. We used the docking software *SCULPT* provided by MDL (66) to perform all the docking experiments and we assessed the free energy of interaction between the peptides and thrombin using the built-in advanced molecular mechanics force-field (MMFF) (56, 58). The docking of all peptides of well defined sequences was done using the same template protein with known 3D structure, 1ABJ.pdb. However, it was assumed that the active site can have minor perturbation in the conformation from one 3D structure to another. Thus, in later experiments more than one template was used in docking experiments (such as 1AI8.pdb, 1HAI.pdb). All the templates used in docking experiments were characterized by high atomic resolution of their X-Ray structures (<2.5 Å) and by less than 0.5 Å rsmd (root

mean square deviation) between the backbone and the side chains of the most important amino acids involved in ligand binding.

As presented in figure 93 a feedback from combinatorial chemistry and in vitro screening of lead compounds for their inhibitory activity against thrombin increased the accuracy of SAR for the predicted models from docking experiments and speed the SAR at specific variable positions P1', P2' and P3' within the original designed peptide sequence (1)-D-Phe-Pro-Arg (P1)-D-Pro(P1')-P2'-P3'-$CONH_2$. Originally a well defined SAR was found for the two sequence spaces which were investigated (1)-D-Phe-Pro-Arg (P1)-D-Pro(P1')-P2'-P3'-$CONH_2$ and (2)-D-Phe-Pro-D-Arg(P1)-P1'-P2'-P3'-$CONH_2$ as derived from the in vitro kinetics of thrombin inhibition. Both peptide sequences showed inhibition of thrombin activity in an in vitro assay using the S2238 (H-D-Phe-Pip-Arg-pNA) as substrate. New lead compounds emerged from these two major sequences and new peptide libraries were designed by maintaining constant the tripeptide D-Phe-Pro-D-Arg- and varying P1' position with amino acids analogs of Phe.

Synthesis of individual peptides and libraries of peptides was performed using F-moc manual synthesis of amide-tetrapeptide libraries on a Rink-Amide resin incorporating phenylalanine (Phe) or Phe analogs in the P3 position such as trans/(cis) cinnamic, dihydrocinnamic acids, D-Naphthylalanine (D-Nal), Phe constrained analog 1,2,3,4- (D)-tetrahydroisoquinoline-3-carboxylic acid [(D) Tic], 3-amino-2-naphtoic acid, D-3,3-diphenylalanine, D-3,5-difluorophenylalanine, D-3-benzothienylalanine, (R, S)-3-amino-3(1-naphtyl)-propionic acid and 1,2,3,4-tetrahydronorharman-3-carboxylic acid. The libraries were obtained using the partition-mixing procedure coupled with parallel synthesis. Ten variation in the P1' position were performed using natural and un-natural amino-acids that were shown very good inhibitory activity for thrombin in the sequence

space DPhe (P3)–Pro (P2)–DArg-P1', i.e. P1' = D-Ser, D-Cys, D-Ala, L/D-Thi, D-3-benzothienylalanine, D-His, D-Pro, D-Thr, D-Asn and D-Gln. The trans-isomers of the cinnamoyl-tetrapeptides were at least five fold more active than the cis-cinnamoyl isomers while the release of double bond constrain through dihydrocinnamic acid had at least two fold better activity than the cis-isomers but less activity than the trans-isomers. In addition the L-Thi was used to scan positions P2 and P1'. The new inhibitor D-Phe-Pro-DArg-(L)Thi has a Ki of 8.6 μM and was discovered as being a new lead compound in the series D-Phe-Pro-D-Arg-P1', containing an unnatural amino acid analog in the P1' position. The order of activity for the peptides containing analogs of Phe in the P3 position is D-Nal > D-Phe > Dihydrocinnamics > Transcinnamic > D-Tic > L-Tic > ciscinnamics >>tetrahydroharman acid-compounds with conserved residues at P2=Pro and P1=D-Arg and variable L/D- amino-acids at P1'. Lead compounds having the experimentally determined inhibitory constant (Ki) between 16.5 –0.9 μM are potential competitive inhibitors for thrombin. In order to predict if the peptides with constrained conformation are more potent inhibitors the advanced MM3 force-field was used to minimize individual tetrapeptides and the backbone dihedral angles phi and psi were predicted to favor in most cases beta turns and beta hairpin conformation. Circular dichroism investigations suggests that the D-Arg- in *i+2* position followed by D-amino acids (polar and neutral like D-Thr, D-Gln, D-Ser and D-Ala) or L-Pro in *i+3* position favor beta turn or alpha helix structures in solution at low and neutral pH. Replacement of D- with L-amino acids in *i+3* position was accompanied by a significant lost in the beta turn or alpha helix structure with a shift toward a disorganized structure. Also, the aromatics (L-Phe, L-Tyr and L-Trp) in *i+3* position are disturbing the beta structure while the replacement of D-Phe with L/D-Tic is accompanied by a shift toward beta-

strand-like structure. SAR (structure-activity relationship) suggests that tetrapeptides which adopt beta turn - beta hairpin or alpha helix conformation in solution are more active toward inhibiting thrombin.

In addition, isothermal titration calorimetry (ITC) was used to determine the thermodynamic of binding of peptide inhibitors to thrombin. Both enthalpic and entropic factors are predicted to be important contributors for the free energy of binding between the peptides and thrombin as determined from ITC data suggesting that the hydrophobic interactions together with hydrogen bonding are governing the ligand affinity for thrombin. The ITC results were in agreement with the kinetics of thrombin inhibition with respect to the experimentally determined association constant, mostly in the order of 10-0.8 μM for the best inhibitors.

2D-transferred NOESY experiments were used to further assed if the peptide inhibitors are binding to the active site. Binary complexes between the peptides D-Phe-Pro-D-Arg-Gly-$CONH_2$, D-Phe-Pro-D-Arg-Ala-$CONH_2$ and thrombin and ternary complexes between these peptides-thrombin-thromstop (low nM inhibitor) were showing building-up of negative NOEs in complexes between thrombin-peptides while the presence of thromstop inhibitor decreased the negative NOEs in ternary complexes. These results are the first structural evidence for peptides binding more than 80% to the active site of thrombin in the absence of any X-Ray data.
The original SAR presented at P1' position in the sequence space D-Phe-(P3)–Pro-(P2)–DArg-P1'-$CONH_2$ is the first independent evidence for the a specific amino acids requirements in P1', such as polar (L-Ser, L-Thr, L-Cys), small (Gly, Ala, Val) or bulky hydrophobics (L-Ile, L-Thi). An independent group of investigations (67) showed through X-Ray experiments that the P1' position can be occupied by bulkier groups (such

as L-Ile and L-Thi) than expected from the natural substrates for thrombin (i.e. Gly, Ala, Val, Ser or Thr) since the side chain of Lys60F, which defines the S1's subpocket is displaced significantly and is forming a large nonpolar S1' subsite to accommodate the bulky P1' residue (67). Our SAR studies for the P1' position showed that there are significant differences in the binding affinity for thrombin of tetrapeptides from the series D-Phe-Pro-D-Arg-P1'-$CONH_2$ (ranging from 5-500 fold) suggesting that the interaction between the amino acid at P1' position and the S1' pocket in thrombin is very specific tolerating only a limited type of amino acids as described above (67). In addition, there was a significant change in the affinity for interaction with thrombin as the configuration for the same amino-acid in P1' position is changing from L to D. Specifically, a switch from L-Thr into D-Thr in P1' was correlated with a 13 fold increase in the inhibitory activity. The structural basis for this favorable switch in the affinity was further investigated through molecular docking and modeling. We found that D-configuration, at least in the case of D-Thr at P1' position allows more favorable hydrogen bonding networks to be established between the peptide-ligand and side-chains within the the active site of thrombin. This research study is presenting molecular docking experiments which support the hypothesis that tetrapeptides differing in one single amino acid at P1' position adopt different conformations into active site of thrombin which may be part of the factors determining the significant differences between their inhibitory potential.

Chapter 1

Molecular docking and structure-based design of peptides with potential inhibitory activity against thrombin

1.1. Rational selection of thrombin-templates for docking

Our major goal was to design new peptide inhibitors directed toward the active site of thrombin, such that a competitive inhibition could be obtained. The availability of many crystallographically determined complexes of thrombin with the modified tripeptide PPACK (D-Phe-Pro-Arg-chloromethylketone) provided the 3D structure as target for docking of new peptide inhibitors, a process known in the field of drug design and discovery as Structure-Based-Design (SBD) (41-55). The major structural features of the active site of thrombin are conserved in between many crystallographically determined complexes thrombin/inhibitors. These major structural features refer to the aryl binding pocket (S3 pocket) largely defined by Trp 215, Leu99, Ile174, Tyr60A and Trp60D. In all PPACK/thrombin complexes the amino acid from the P3 position (D-Phe) is docked into the S3 pocket. The P2 fragment (Pro) is inserted into the hydrophobic pocket (S2 pocket) defiend by the Leu59-Asn62 insertion loop of the enzyme. The side chain of the P1 group (Arg) is inserted into the specificity pocket (S1 pocket) and forms the salt linkages with Asp189 of the enzyme. The S1 pocket of alpha-thrombin is known to be very similar to that of trypsin-like enzymes. The pocket is enclosed on one side by residues Gly 216-Gly 219 in a twisted beta-strand arrangement in relation to the arginine

side chain of the inhibitor. The pocket is closed off at the bottom by the Asp 189 which is involved in salt bridges to the guanidine group of the P1 arginine residue. This latter interaction provides a significant fraction of the binding energy for the inhibitor. The pocket is closed on the other side by residues 190-193, forming an irregular strand. Gln 193 folds over the P1 side chain and closes the pocket in the front. The back part of the pocket is enclosed by the side-chains of Tyr 228, Leu 160, Val 138, and Ala 190. Thrombin has also the unique insertion loops B (Tyr60-A-Ile60-I) and C (Thr149-A-Ala149-E), contributing to the specificity of substrate binding (68-70). It was already shown earlier that Lys60F side chain excludes bulky amino acids at the P1' position (mutagenesis studies revealed that the replacement of Lys-60-F with Ala enhanced the affinity of the bulky amino acids such as Leu at P1') (83). The P2' position in the natural substrates and peptide inhibitors for thrombin is very diverse tolerating Phe, leu, Val, Ile, Pro, Gly, His, Glu and Asn; however, the kcat/Km of synthetic substrates showed a preference for bulky hydrophobic P2' residues such as Phe and Trp with a total exclusion of acidic amino acids such as Glu and Asp (80-86). P3' position in substrates and inhibitors is very diverse, tolerating Arg, Asn, His, Leu, Gly and Asp; however the higher catalytic efficiency of thrombin (highest kcat/Km) was determined for basic and bulky hydrophobic amino acids, while the acidic amino acids Asp and Glu were excluded from P3' position (70-75).

Originally the PDB (Protein Data Bank) (Brookhaven) was searched for X-ray complexes between thrombin and different small peptide or non-peptidic derived inhibitors and more than 300 solved structures were retrieved (including NMR structures).

The selection of thrombin templates used in the docking experiments was determined by the accuracy of the X-Ray structures based on screening of the structural data through PROCHECK data base (PROCHECK v.3.5.40). This data base is presenting the main structural features of the X-ray structure as specific plots such as Ramachandran plot, Ramachandran plots by residue type, Chi1-Chi2 plots, main-chain parameters, side-chain parameters, residue properties, main-chain bond length distributions, main-chain bond angle distributions, RMS distances from planarity, distorted geometry plots. The main parameters checked by PROCHECK for each input of X-ray data are: covalent geometry, planarity, dihedral angles, chirality, non-bonded interactions, main-chain hydrogen bonds, disulphide bonds, stereochemical parameters, parameter comparisons, residue-by-residue analysis. For docking experiments we used 3-5 protein templates based on their overall atomic resolution (below 2.5 Å) and their specific structural accuracy of thrombin active site subpockets S1', S2', S3' (73). The subpockets S3, S2, S1 of thrombin are well solved at atomic resolution in most the structures containing small peptides or non-peptides inhibitors which are occupaying just these subpockets. However, the S1', S2' and S3' subpockets are not solved at high atomic resolution in some thrombin complexes with small molecules inhibitors (68-86). The complexes of thrombin with bivalent inhibitors are the best templates for those docking experiments which require testing of additional binding sites for the inhibitor at S1', S2' and S3', i.e. were the specificity of residues at P1', P2' and P3' in the peptide inhibitor is the major subject in SAR studies. In these complexes of thrombin and bivalent inhibitors the ligands are derived from PPACK and are blocking the S3, S2, S1 pockets within the active site but contain in addition a –(Gly)n –linker covalently linked to a peptide segment blocking the S1', S2' and S3' pockets within the fibrinogen recognition exosite (73).

One of the most important questions raised during ***molecular docking*** is related to the conformation of the active site of the protein, which is used as template for all docking experiments. The issue of having the same conformation of the active site for all known 3D structures is an open question. In the case of thrombin and thrombin complexes with different inhibitors it has been shown that the structure of the hirulog 1-thrombin complex, with an N-terminal D-Phe-Pro-Arg targeted to bind in the active site, has practically identical interactions with that of the PPACK (*D-Phe-Pro-Arg-chloromethyl ketone*) -thrombin complex in the active site (73-86). Since the structure of an enzyme/ligand complex is a valuable basis for the design of novel inhibitors, a preliminary comparison of the crystallographically determined structures of different enzyme/ligand complexes is necessary before starting any docking experiments. In the case of alpha-thrombin/inhibitor complexes, the geometries of the enzyme active site among these complexes are similar as already described by others (73). The difference in thrombin conformation between thrombin/hirudin and thrombin/PPACK complexes is limited to the short autolysis loop spanning residues Lys 145-Gly 150 (68-86). In the case of the specificity pocket of alpha-thrombin (the S1 pocket), the geometry of this pocket was shown to be essentially invariant in the crystal structures of these complexes whether or not the pocket is occupied (73). Therefore, this specificity pocket in alpha-thrombin was one of the targets for designing of novel inhibitors (80-86).

Other structural feature important for high affinity of interaction between thrombin- ligand inhibitor requires a dual hydrogen bond (H-bond) (usually antiparallel) made between inhibitors and Gly216 at the S3 site in thrombin. Loss of the H-bond between Gly216-CO and inhibitor-NH can result in a loss of 3-100 fold in affinity while

loss of the Gly216-NH H-bond interaction with inhibitor-CO can reduce activity by up to 10,000-fold (69, 75). An additional important structural features of thrombin active site is Glu192. Thrombin shares this acidic residue with only the serine protease Protein C, whereas the majority of trypsin-like serine proteases have a glutamine at this position. This residue is a major determinant for thrombin specificity since a thrombin mutant Glu192→Gln had the decreased ability to activate Factor Xa (80-85). It was showing later through molecular modeling studies that some amino acids at P1' position in the peptide inhibitors derived from substrates [P3-P2-P1-P1'-$CONH_2$] are establishing important hydrogen bonds with their side chains and the O atoms from the carboxylate side chain of Glu192 making this amino acid in the active site of thrombin a significant determinant for the affinity of binding of peptides inhibitors (68-86). In figure 94 the details of the active site of thrombin template 1ABJ in complex with PPACK are presented. The pdb structure was downloaded from PDB data bank and the modeling was performed using the software ***SCULPT***-from MDL.

1ABJ.pdb

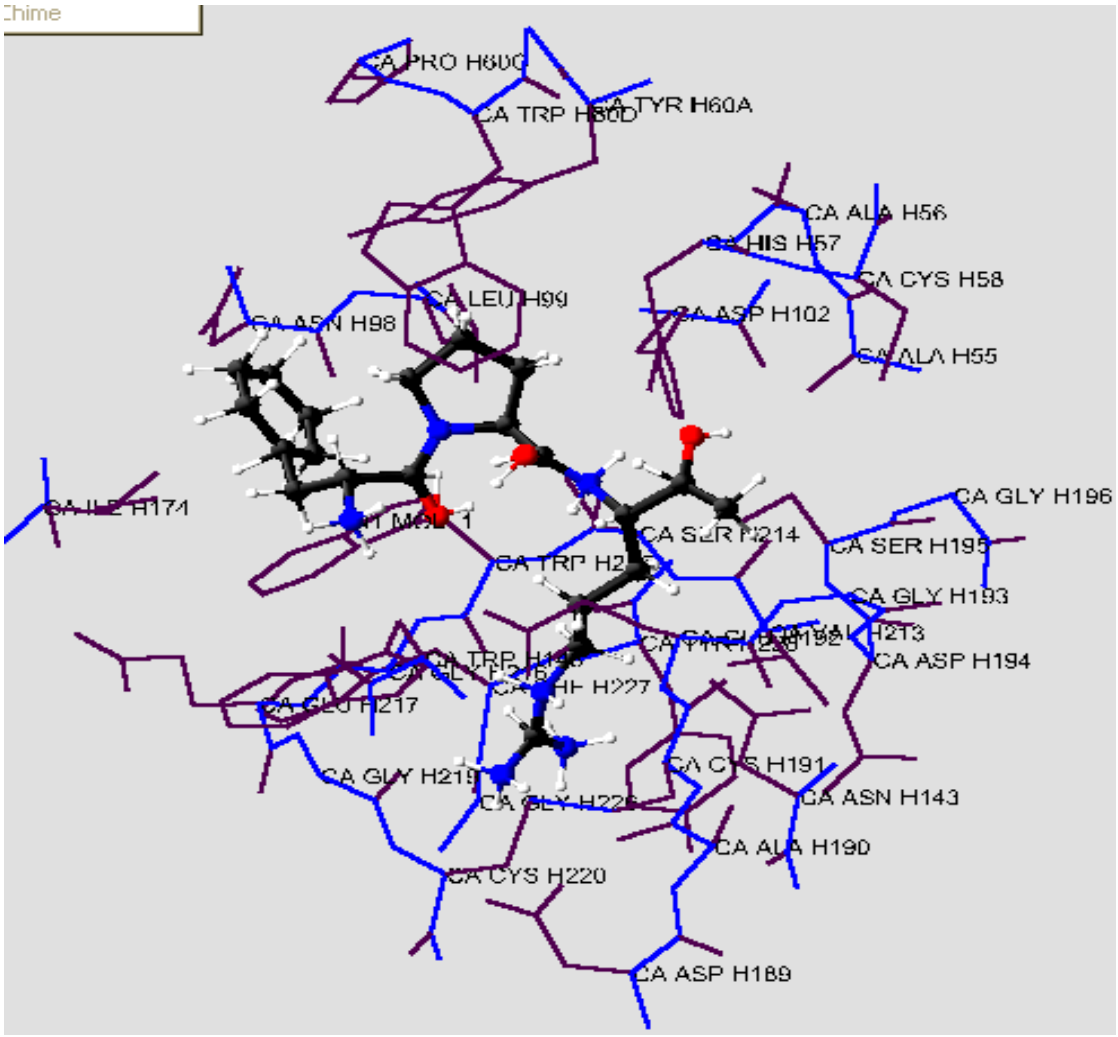


Figure 94. The active site details of the template 1ABJ.pdb (thrombin in complex with PPACK) (73). The pdb coordinates were downloaded from PDB and the complex was modeled using the software *SCULPT* from MDL (66). The amino acids located within the major subpockets of thrombin are labeled as presented in the original reference (73).

<u>Aryl pocket (S3)</u> Trp215, Leu99, Ile174,

Tyr60A, Trp60D; *dPhe (P3) docked into S3*[(1)]

<u>Hydrophobic pocket (S2)</u>
Leu59-Asn62 insertion loop; *Pro (P2) docked into S2*[(1)]

<u>Specificity pocket (S1)</u> *Gly216-Gly219*

Asp189: **salt bridge with Arg (P1);** ***Arg (P1) docked into S1; Residues192-193, Tyr228, Leu160, Val138 and Ala190*** [(1)]

The structure 1ABJ was checked for the accuracy of X-Ray data using the software PROCKECK (figure 95).

1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	203	83.2%*
Additional allowed regions	[a,b,l,p]	41	16.8%
Generously allowed regions	[~a,~b,~l,~p]	0	0.0%
Disallowed regions	[XX]	0	0.0%
Non-glycine and non-proline residues		244	100.0%
End-residues (excl. Gly and Pro)		6	
Glycine residues		23	
Proline residues		16	
Total number of residues		289	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and *R*-factor no greater than 20.0 a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

2. G-Factors

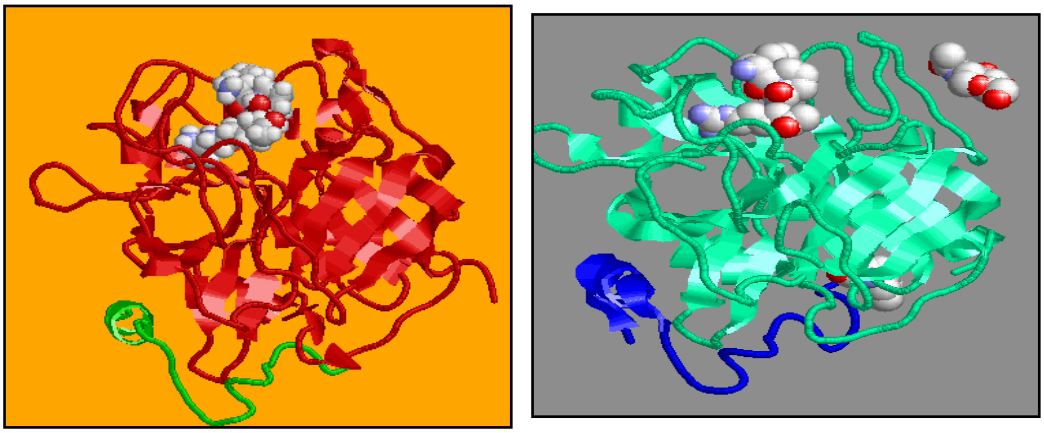
Parameter	Score	Average Score
Dihedral angles:-		
Phi-psi distribution	-0.56*	
Chi1-chi2 distribution	-0.54*	
Chi1 only	-0.49	
Chi3 & chi4	0.33	
Omega	0.14	
		-0.22
Main-chain covalent forces:-		
Main-chain bond lengths	-0.09	
Main-chain bond angles	-2.32**	
		-1.38**
OVERALL AVERAGE	-0.59*	

Figure 95: Summary of PROCKECK (87) analysis of 1ABJ.pdb. G-factors provide a measure of how unusual, or out-of-the-ordinary, a property is (Values below -0.5* - unusual; Values below -1.0 - highly unusual).**

It is important to mention that the main-chain bond-lengths and bond angles are compared with the Engh & Huber (1998) (87) ideal values derived from small-molecule data. Therefore, structures refined using different restraints may show apparently large deviations from normality, as is the case for the structure 1ABJ.pdb. However, the template 1 ABJ.pdb used in original docking experiments and development of lead peptide compounds has 2.4 Å resolution with very well solved S1', S2' pockets within the active site of thrombin. Thus, this structure was passing the structural criteria for being suitable in designing inhibitors were the peptide requires well defined P1' and P2' positions within the sequence space [P3-P2-P1-P1'-P2'-$CONH_2$].

Some models of X-Ray structures of thrombin complexes with PPACK or other small molecule inhibitors used as templates for docking experiments are presented in Figure 96. Additional X-Ray structures were used in docking experiments and tested for their atomic resolution and structural accuracy at the required pockets within active site of thrombin (S3, S2, S1, S1', S2' and S1'). The X-ray structures together with their references are presented in Table 1.

1ABJ.pdb **1HAI.pdb**



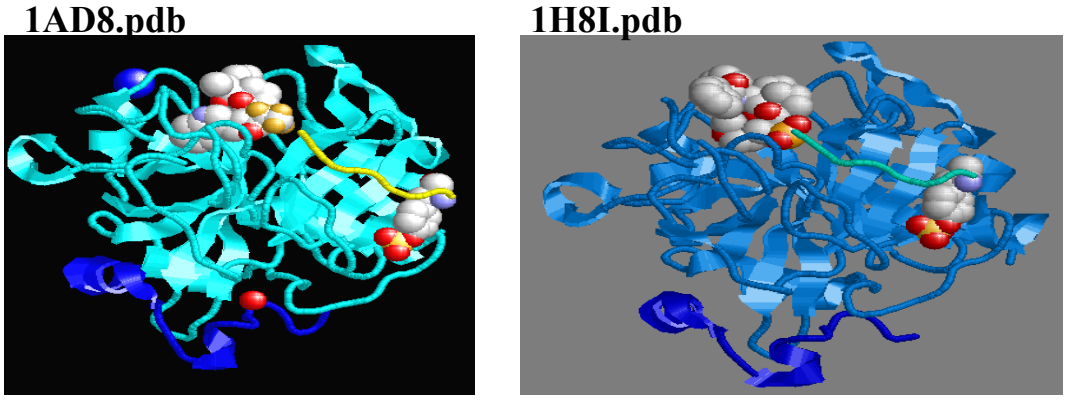


Figure 96: Models of X-Ray structures of thrombin complexes with PPACK or other small molecule inhibitors used as templates for docking experiments. All structures were downloaded from PDB and modeled using the software Rasmol (88).

The superposition of more than 2 templates used in docking experiments showed some degree of deviation of some side-chains in the active site of thrombin, however they were not significant (less than 0.5 Å-quantitative) (data not shown) supporting the fact that structure-based design in the case of thrombin was accurate enough if more than 2 templates for docking experiments are used. However, the predicted free energies for the interaction between thrombin and peptides were slightly different (5-8 kcal/mol difference) when docking of the same peptide was done into each of the active site of the two independent templates (see table 8). These differences in the free energy are relatively significant if we consider the accuracy of the software *SCULPT* which is *within* $\pm$ 5-8 kcal/mol, suggesting that is always a good control to perform docking within different X-ray known structures for the same target (thrombin in our case).

1.2. *In silico* screening of peptides with potential anti-thrombin activity using the docking software "*SCULPT*".

The strategy of designing new peptides with potential inhibitory activity against thrombin was based on discovering original leading compounds characterized by a high affinity of interaction with thrombin. The free energy of interaction between thrombin and each ligand was calculated with the built-in molecular mechanics force-field (MMFF) provided by the docking software *SCULPT*. These new leading compounds with a specific sequence were further developed by adding extra natural or unnatural amino acids in the next position (from N to C terminus), such that residues with different chemistry of the side chain were screened (aromatics, hydrophobic, acidic, basic, polar). After each round of minimization of the structure of the complex between thrombin and the new lead compound new sequences were developed. In the original experimental investigations we focused on those complexes that were characterized by a free energy of interaction between thrombin and the new peptides below "-50 kcal/mol". This is not a real free energy for what might be a complex between thrombin and an inhibitor peptide, since for an dissociation constant in the range of nM-pM we expect free energy of interaction in the order of (-10kcal/mol) – (-20kcal/mol)". Therefore, a free energy of "-50 kcal/mol" correspond to a dissociation constant far lower than the pM range and thus is not a reliable value as generated by the software SCULPT which is employing molecular force-field 94 (MMFF94) for predicting the free energy of interaction between protein target and the ligand. We still considered leading group compounds those with lower free energy, i.e. below "-50 kcal/mol" since we assumed that this free energy of

interaction might qualitatively represent a very stable complex between inhibitor and thrombin.

Since the structure of an enzyme/ligand complex is a valuable basis for the design of novel inhibitors, a preliminary comparison of the crystallographically determined structures of different enzyme/ligand complexes is necessary before starting any docking experiments (41-50). In the case of alpha-thrombin/inhibitor complexes, the geometries of the enzyme active site among these complexes are similar as already described by others (68-86). In the PDB (Protein Data Bank) there are more than 300 structures thrombin/inhibitors solved by X-ray diffraction, thus making it difficult to choose a template for docking experiments. Based on these structural data, we used 3-5 X-Ray solved structures of thrombin in complex with PPACK or other small molecule ligand inhibitors (such as benzamidine compounds (42)) to generate an *in silico* combinatorial peptide libraries with potential inhibitory activity for thrombin (see section 1.1).

In *SCULPT*, the force-field is provided with potential energies that are used to model: explicit hydrogen bonds; variable dihedral angles; van der Waals interactions; electrostatic interactions; user-applied tugs (springs). Van der Waals interactions are modeled with a modified Lennard-Jones function between atoms within 6Å of each other. Electrostatic interactions are modeled with a Coulomb model using a distance-dependent dielectric interaction between atoms within 10 Å of each atom. In addition to van der Waals interactions the free energy of interaction between the designed peptides and thrombin was including the electrostatic forces in the late steps of docking when a refined structure between thrombin and the inhibitor peptide was generated.

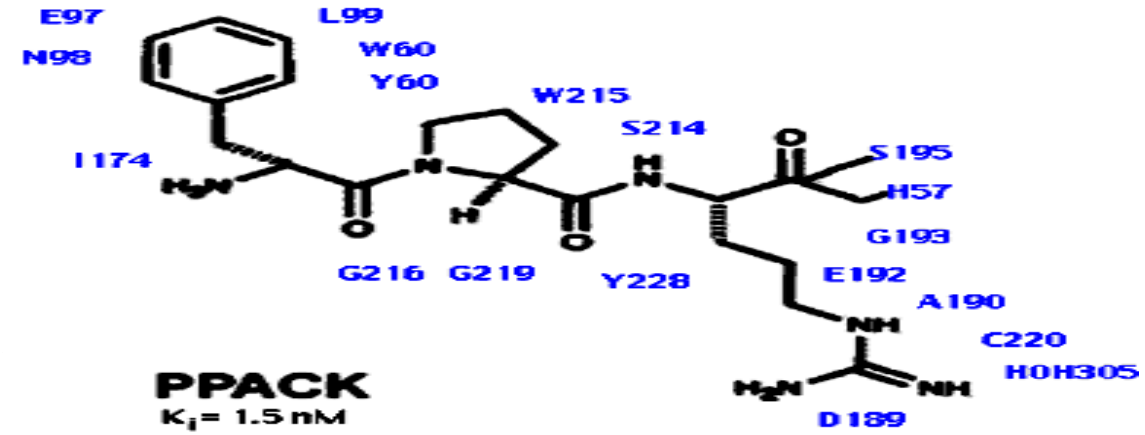
In addition to this initial criteria new scoring functions for the evaluation of lead compounds were developed based on the structural complementarity between the peptide

ligand and the the thrombin active site such as the normalized complementarity between the ligand and thrombin, as determined from ligand-protein contact analysis performed with the Ligand Protein Contacts (LPC) software developed originally by Sobolev (89-93) and provided by the server running at Weizmann Institute (Israel).

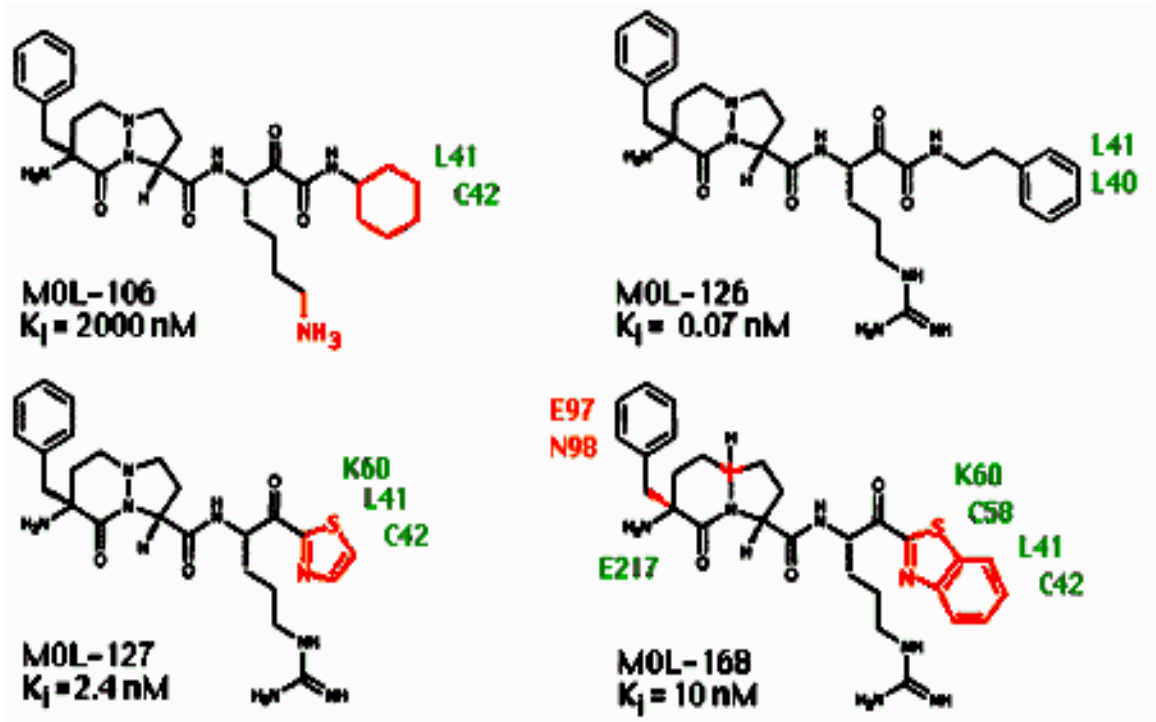
1.3. Peptide sequences used to generate new candidate inhibitors for thrombin.

During the first stages of docking experiments the structural details of some very well known peptides and peptido-mimetic inhibitors of thrombin were chosen as a starting point for developing further candidate compounds. As shown in Figure 97 (A-E), the X-Ray data show specific interactions between the peptido-mimetic compounds in the active site of thrombin. The salt bridge between D189 of thrombin binding pocket and the R (Arg) at the P1 position of peptidomimetic inhibitors was proposed to make a large contribution to the free energy of interaction of the thrombin-inhibitor complex, and thus is a structural requirement for the major classes of peptides derived inhibitors for thrombin (27-31). As mentioned earlier other structural feature important for high affinity of interaction between thrombin and ligand inhibitor require a dual hydrogen bond (H-bond) (usually antiparallel) between inhibitors and Gly216 at the S3 site in thrombin. Loss of the H-bond between Gly216-CO and inhibitor-NH can result in a loss of 3-100 fold in affinity while loss of the Gly216-NH H-bond interaction with inhibitor-CO can reduce activity by up to 10,000-fold (27-31).

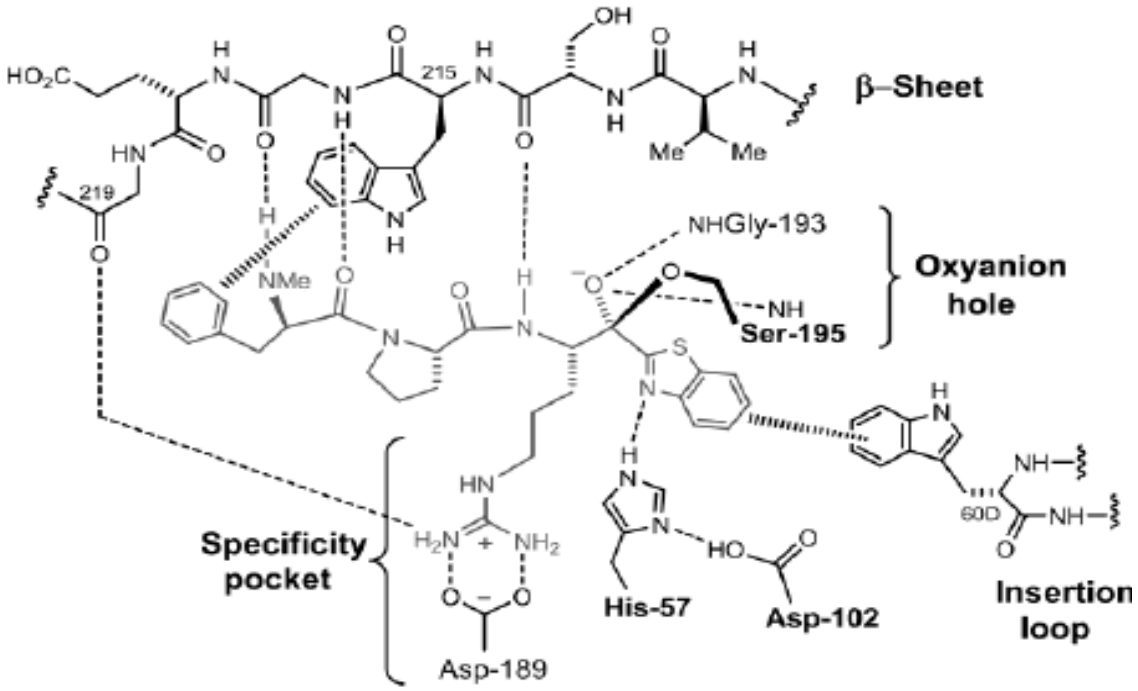
A


E97
N98
L99
W60
Y60
W215
S214
I174
S195
H57
G193
G216
G219
Y228
E192
A190
C220
HOH305
D189
PPACK
K_i = 1.5 nM

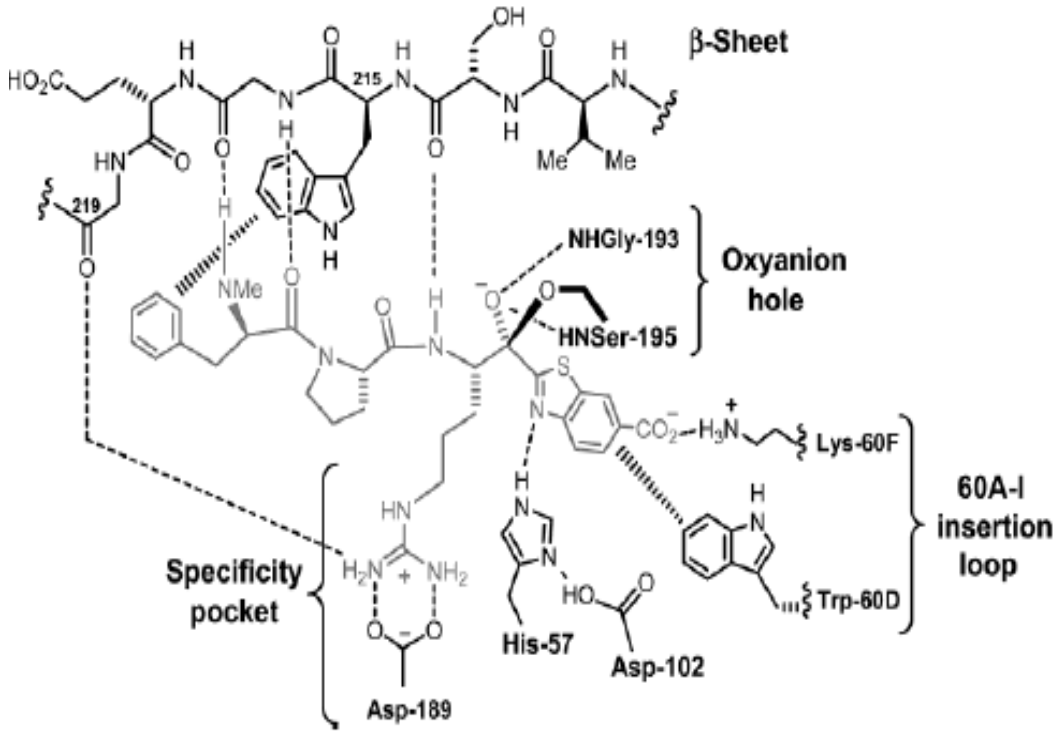
B


L41
C42
MOL-106
K_i = 2000 nM
NH_3
L41
L40
MOL-126
K_i = 0.07 nM
K60
L41
C42
MOL-127
K_i = 2.4 nM
E97
N98
K60
C58
L41
C42
E217
MOL-168
K_i = 10 nM

C


β–Sheet
NHGly-193
Oxyanion hole
NH
Ser-195
Specificity pocket
His-57
Asp-102
Insertion loop
Asp-189

D


β-Sheet
NHGly-193
Oxyanion hole
HNSer-195
Lys-60F
60A-I insertion loop
Trp-60D
Specificity pocket
His-57
Asp-102
Asp-189

E

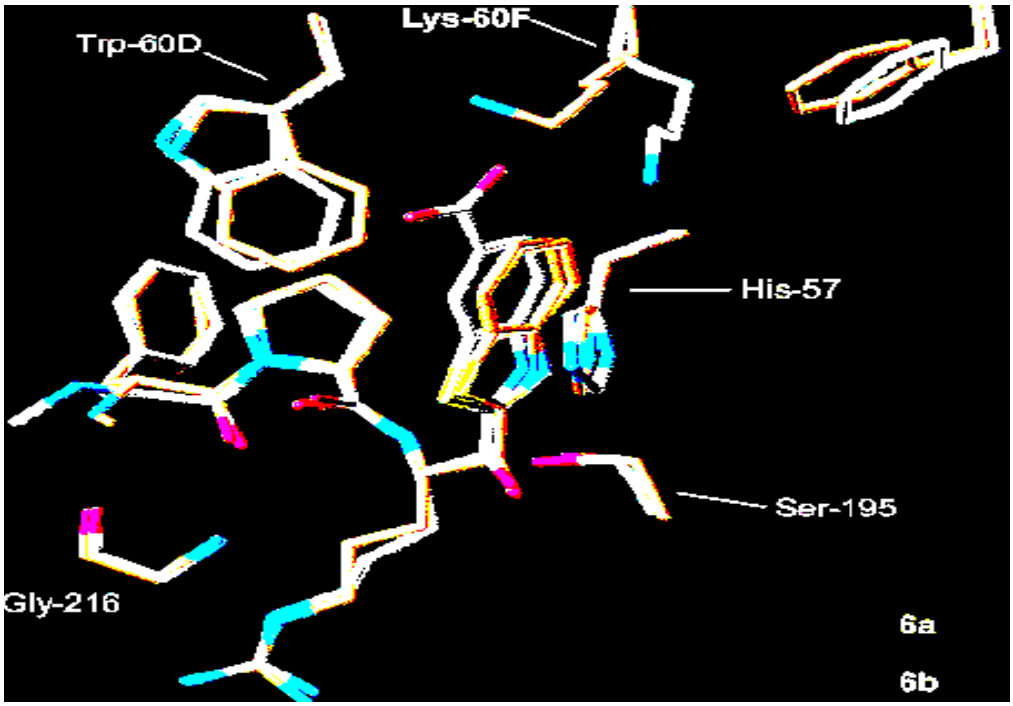


Figure 97 (A-E): The major structural features characterizing peptides inhibitors interaction with thrombin (27, 28, 80,81).

During the initial docking experiments specific criteria were used to generate "lead compounds" using the software "*SCULPT*". The major criteria for selecting of lead compounds from docking experiments was based on the predicted free energy (ΔG) of interaction between the designed peptides and thrombin as described in section 1.2.

substrate	specificity site[a]	
fibrinogen A 12–24	GGGVR	GPRVVERH
fibrinogen A 15–27	VRGPR	VVERHQSA
fibrinogen B 10–22	FFSAR	GHRPLDKK
prothrombin 151–163	MVTPR	SEGSSVNL
prothrombin 282–294	FFNPR	TFGSGEAN
factor V 706–718	ALGIR	SFRNSSLN
factor V 1014–1026	PLSPR	TFHPLRSE
factor V 1541–1553	AWYLR	SNNGNRRN
factor VIII 368–380	FIQIR	SVAKKHPK
factor VIII 736–748	AIEPR	SFSQDSRH
factor VIII 1309–1371	TQSKR	ALKQFRLP
factor VIII 1685–1697	NQSPR	SFQKKTRH
factor XIII 33–42	GVVPR	GVNLQEF
factor XIII 509–518	EGVMK	SRSNVDM
platelet receptor 37–50	TLDPR	SFLLRNPN
antithrombin III 389–399[c]	VIAGR	SLNPNR
factor VII 148–157[b]	KPQGR	IVGGKV
factor XI 365–377[c]	KIKPR	IVGGTASR
protein C8–17[c]	QVDPR	LIDGKV

-P1------P1'-

Figure 98: Peptide sequence around the cleavage site of the thrombin-susceptible bonds (-P1-P1'-) in proteins from human plasma (67).

The **strategy** of designing new peptides inhibitors for thrombin was based on the substrate specificity of thrombin, which requires that the carboxyl side of the scissile bond to have the amino acids Arg or Lys (see Figure 98). In the following amino acid sequence…N-**P4-P3-P2-*P1-P1'*-P2'-P3'-P4'-C…, -*P1-P1'*-** is the scissile bond (**P1 has to be Arg or Lys**). The two sequences used for designing peptide inhibitors are shown below:

Peptide sequence (1) $_2$HN-X-D-Phe-Pro-Arg-D-Pro-P2'-P3'-…CONH$_2$

Peptide sequence (2) $_2$HN -X-D-Phe-Pro-D-Arg-P1'-P2'-P3'…. CONH$_2$

Peptides with **D-Pro in the P1'** position and **D-Arg instead of Arg in the P1** position are expected to inhibit the hydrolysis of peptide, allowing these sequences to function as potential inhibitors. Variations in positions P1', P2', P3' and P4' with a selection of the

20 naturally encoded amino acids in the L- and D- forms were used to generate the *in silico* library of peptides with potential inhibitory activity against thrombin.

1.4. MATERIALS AND METHODS

1.4.1. MATERIALS

The software *SCULPT* used in docking experiments was purchased from MDL (66). The structures were draw with ISIS Draw 2.2.1 from MDL and the molecular models were generated with Rasmol (88) and SPDBViewer (94). All softwares were operated on the PC machine with Windows XP operating system.

1.4.2. METHODS

Docking and structure-based design. Molecular Modeling. Molecular docking was carried out using the SCULPT software package provided by MDL. The strategy of designing new peptides inhibitors consist in drawing the polypeptides within ISIS DRAW 2.2.1 and then importing them into three-dimensional presentation using SCULPT for docking and minimization. The procedure for the docking process is described below:

1. The initial initial ".pdb" file of interest containing the atomic coordinates was downloaded from PDB , such as the "**PPACK/thrombin**" complex template (**1ABJ.pdb**) for docking of new peptides into active of thrombin.
2. The file was opened within SCULPT and all the water molecules were removed from it. The file was resaved. The minimization process was done using just the command for ***Van der Waals*** forces and the lowest free energy for the ***inhibitor/protein complex*** was recorded. This is the control free energy of interaction between the inhibitor and the protein templates with known 3D structure.

3. The inhibitor was deleted from the new file in **SCULPT** and saved separately, as "protein.sc3" file (for example "**thrombin.sc3**"). This contained just the template protein which is used later for docking experiments.
4. The initial file containing the protein/inhibitor complex was opened and the protein was deleted , leaving just the inhibitor. This new file containing just the inhibitor alone was saved separately either as **".mol" or ".sc3** " file formats.
5. The new peptides sequences were drawn as single peptide chain in **ISIS DRAW**, using the amino acid templates. These new designed peptides were copied and then opened individually each on top of the initial peptide inhibitor used as templates in **SCULPT** program. The function ***"Paste Align"*** of the software **SCULPT** was used to superimpose the new designed inhibitor over the old inhibitor template. After this superposition was done the old inhibitor was deleted; thus, the new inhibitor was left in the same position as the old inhibitor within the active site of protein template. The protein file (protein.sc3) containing the template protein was then opened together with the new inhibitor using the merge file box provided by **SCULPT** software.
6. The protein was ***"freezed***" while the inhibitor was ***"thawed"*** (i.e. allowed conformational freedom) using the corresponding functions from the SCULPT software. This ensured the flexible conformational search for the ligand in the active site but limited the conformational states of the protein itself (rigid docking experiments). The minimization procedures were using ***Van der Waals*** interactions for finding the best fit of the ligand into the active. In additional the electrostatic interactions provided by the MMFF force field from SCULPT were used together with the Van der Waals force-field to assess the interaction between the peptides and

the template-thrombin. The rounds of minimization were performed till the free energy **(ΔG)** of interaction between the ligand and the target protein remained constant.

7. The new complex between the ***protein template – new inhibitor*** was saved as an "**.sc3"** file. This predicted free energy of interaction between protein and the new peptide inhibitor was further recorded and used as a criteria for selection of major lead compounds.

In most cases the electrostatic force was applied during minimization procedure after the crude docking model was obtained using just the van der Waals interactions in assessing the free energy of interaction between thrombin template and the peptide ligands.

The new theoretical model of the protein/new inhibitor complex was saved as ".**pdb**" file separately and opened in any program that accept **pdb** files, such as **Rasmol** and **SPDV Viewer**.

1.5. RESULTS

1.5.1. Molecular docking and structure-based design of peptide libraries as potential inhibitors for thrombin.

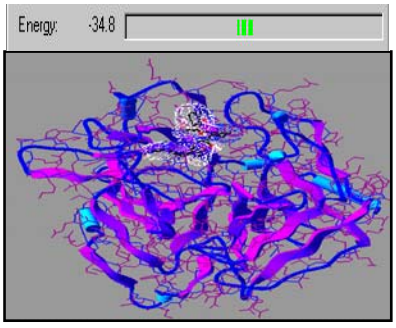
Initial molecular docking experiments were used to generate a candidate group of compounds (with both L- and D- amino acids) that were characterized by a predicted free energy (ΔG), based on Van der Waals interactions with thrombin in the range from -20 to -50 kcal/mol. The thrombin template used in all molecular docking experiments was the X-Ray determined structure of thrombin in complex with the irreversible inhibitor PPACK (D-Phenyl-Prolyl-D-Arginyl-chloromethylketone) **(PDB code: 1ABJ.pdb).**

Several compounds having a Gly at the P1' position (in the series ***dPhe-Pro-dArg-Gly)*** have a predicted ΔG below "*-20* kcal/mol. These were further used to generate pentapeptides and hexapeptides having variable P2'and the P3' positions, and a constant P1' position. The **X-position** was also varied but the peptides show lower affinity for thrombin. The peptides were docked into active site of thrombin by taking the X-ray coordinates of the protein template from the ".**pdb" file** and deleting the initial inhibitor PPACK. The results from strategy of docking experiments are shown on **Figure 99**.

Figure 99: Docking experiments: experimental design and *in silico* SAR (Structure Activity Relationship) of peptides reversible inhibitors for thrombin.

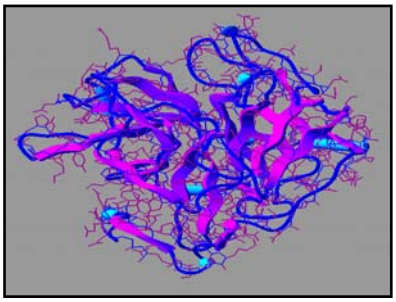
Strategy used for docking experiments using software "*Sculpt*"

1. PDB: Imported coordinates for Thrombin/PPACK 1ABJ.pdb in "*SCULPT*".



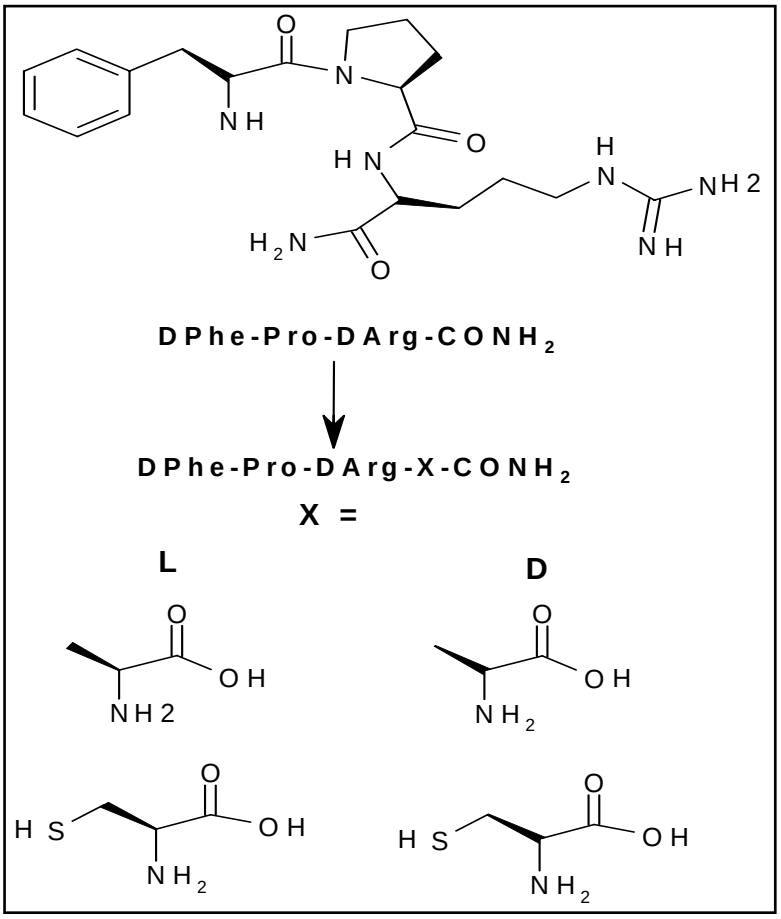
2. Minimized Thrombin/PPACK using "*SCULPT*" (molecular –mechanics algorithm).

3. Deleted inhibitor and preserve the X-Ray coordinates of protein *template* THROMBIN.



4. Deleted the protein in a separate file and maintain the coordinates of initial INHIBITOR (PPACK).sc3

5. Drew the structure of new peptide inhibitor using the software **"ISIS DRAW 2.2.1".** and save the files with the ".skc" extension.



In addition to D-Phe, other Phe analogs were used to perform trials for new peptides with inhibitory activity for thrombin (figure 100).

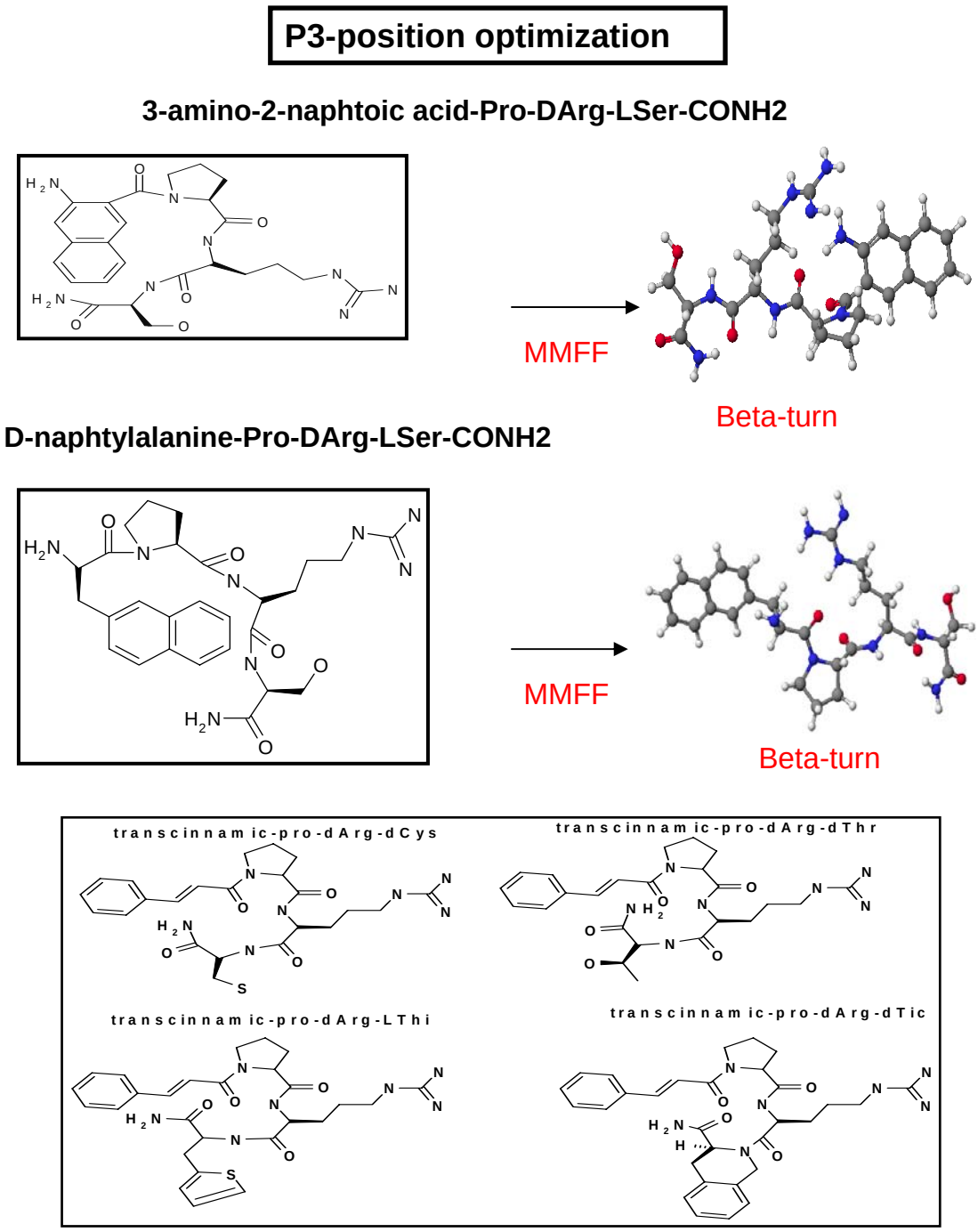
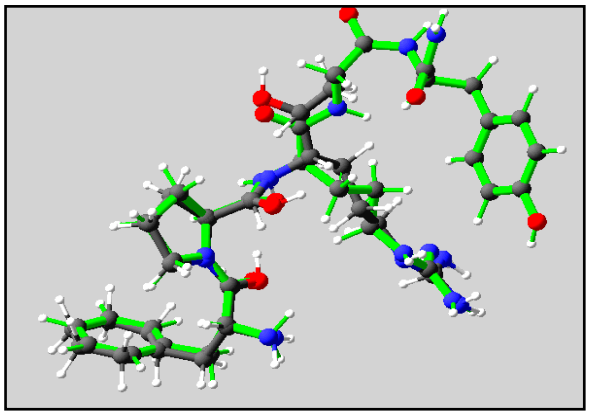


Figure 100: Phe-analogs used as trials for the P3 position within the D-Phe-Pro-D-Arg-P1'-CONH2 peptide. Other Phe analogs used in P3, P2 or P1' positions are described in Materials and Methods and presented in section . These peptides containing natural or unnatural amino acids were minimized using the advanced molecular mechanics force field (MMFF) provided by SCULPT and their final conformation assessed by inspection of their dihedral angles (phi and psi) (data not shown). Based on dihedral angles analysis (data not shown) we classified the conformation of the new compounds as being beta turns, beta sheet or alpha helix (68-86).

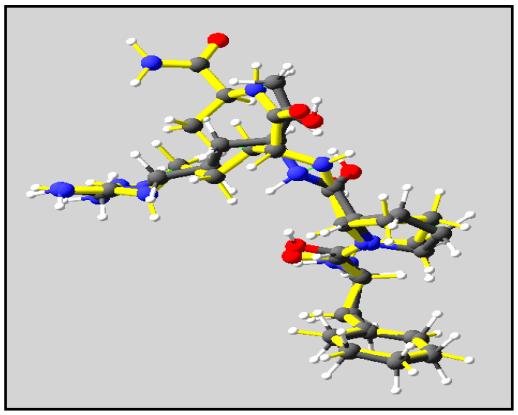
6. Superimposed the new inhibitor peptide over the old inhibitor (PPACK) using the function *"Paste and Align"* from "*SCULPT*". Deleted the old inhibitor and generated a new file for new inhibitor: [New Inhibitor.sc3]

Example of new designed peptides inhibitors superimposed over the old inhibitor, PPACK (the skeleton of the old inhibitor is presented in grey colors) are shown below:

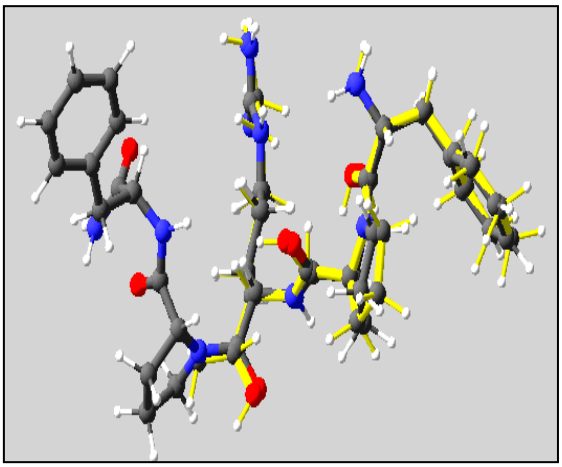
D-Phe-Pro-D-Arg-Tyr-CONH$_2$



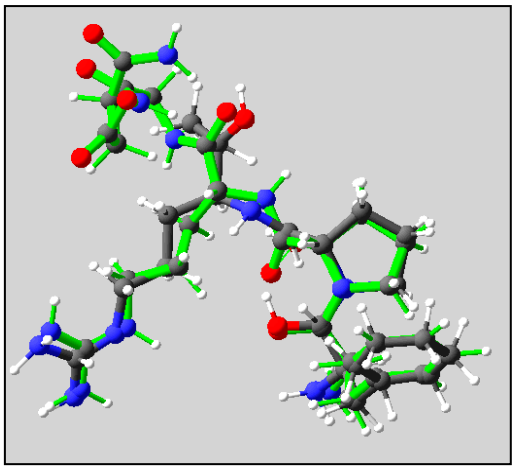
D-Phe-Pro-D-Arg-Ala -CONH$_2$



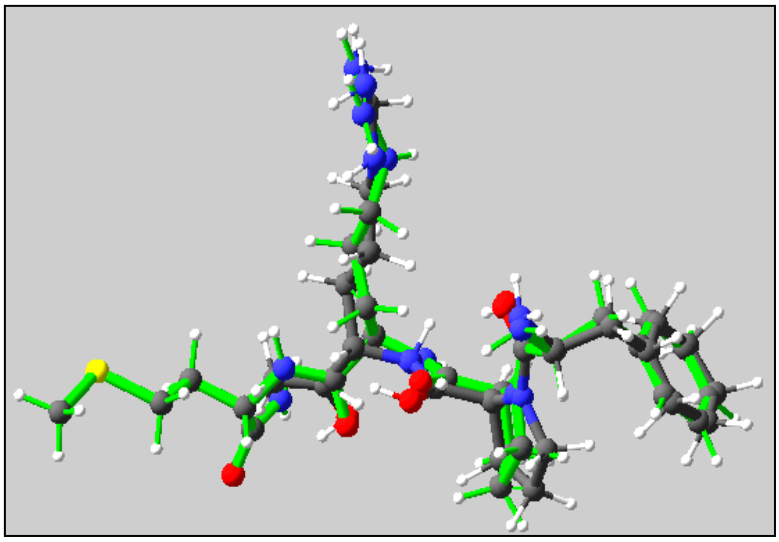
D-Phe-Pro-Arg-D-Pro-D-Arg-CONH$_2$



D-Phe-Pro-D-Arg-Gly-Asp-CONH$_2$

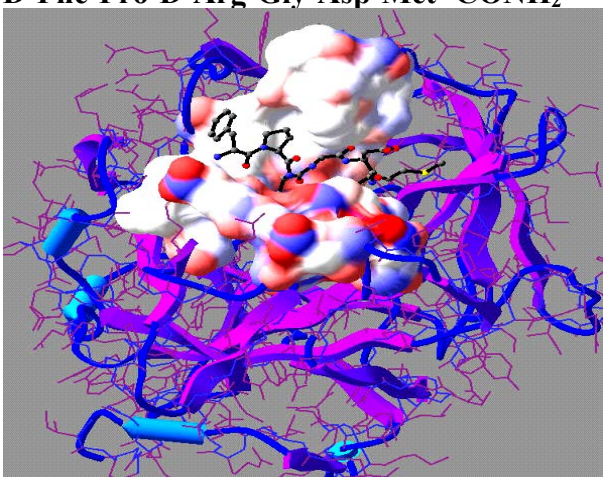


D-Phe-Pro-D-Arg-Met -$CONH_2$

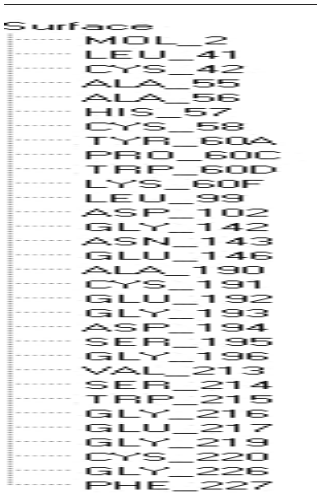


7. **Superimposed the template file [THROMBIN.sc3] over the [New Inhibitor.sc3]** using the ***MERGE FILE*** option from **"*SCULPT*".** Used the options ***FREEZE*** for protein and ***THAW*** for inhibitor to further perform **the *MINIMIZATION*** procedure which predicted the free energy of interaction ΔG (kcal/mol) **between thrombin and new peptide inhibitor** (molecular mechanics (MM) algorithm). Example of computer generated complexes between the new designed peptide inhibitors in complex with thrombin are shown together with the **active site surface of thrombin (within 12 Å of ligand).**

D-Phe-Pro-D-Arg-Gly-Asp-Met -$CONH_2$

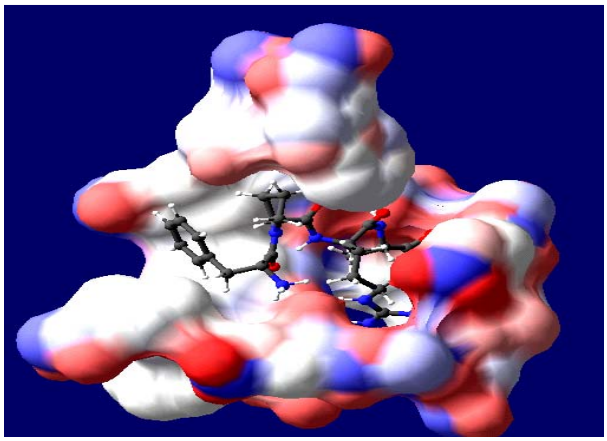


12 Å surface within ligand



D-Phe-Pro-D-Arg-*Ala*-$CONH_2$

ΔG =- 69.4 [kcal/mol]



All peptide compounds containing other amino-acids analogs at P3, P2 and P1' positions were docked into active site of thrombin using the same software "***SCULPT"*** and the free energy of interaction with thrombin was assessed using the same strategy described above. Tabel 7 shows the original peptide libraries and the ***scoring functions*** used to determine the best lead compounds based on free energy of interaction with thrombin as is predicted by "S*CULPT*".

Table 7: the original peptide libraries generated *in silico* through docking experiments using SCULPT and the free energy of interaction between ligand and target protein as predicted by the docking software. The free energy was predicted based on the van der Waals interactions between each peptide and thrombin and is reflecting the relative sterical fitness between each ligand and thrombin.

Inhibitor tried (from N to C terminus)					Free Energy of interaction
original: ppack (d-phe-pro-arg)					
d-phe	pro	Arg	Pro	Trp	124.4
d-phe	pro	Arg	Phe		243.4
d-phe	pro	Arg	Ala		-44.5
d-phe	pro	Arg	Gly		-54.1
d-phe	pro	Arg	d-pro	d-tyr	-68.3
d-phe	pro	Arg	d-pro	d-phe	-61.8
d-tyr	pro	Arg	d-pro	d-tyr	-61.5
d-phe	ala	Arg	d-pro	d-tyr	-59.2

Leu	d-phe	pro	Arg	d-pro	d-tyr		-59.1
acetyl	d-phe	pro	Arg	d-pro	d-tyr		-58.2
	d-phe	pro	Arg	d-pro	d-trp		-57
d-leu	d-phe	pro	Arg	d-pro	d-tyr		-54.7
d-lys	d-phe	pro	Arg	d-pro	d-tyr		-52.7
Lys	d-phe	pro	Arg	d-pro	d-tyr		-51.6
formyl	d-phe	pro	Arg	d-pro	d-tyr		-51.4
d-phe	d-phe	pro	Arg	d-pro	d-tyr		-51.4
Z	d-phe	pro	Arg	d-pro	d-tyr		-50.7
Arg	d-phe	pro	Arg	d-pro	d-tyr		-50
d-arg	d-phe	pro	Arg	d-pro	d-tyr		-49.4
Phe	d-phe	pro	Arg	d-pro	d-tyr		-47.2
carbamino	d-phe	pro	Arg	d-pro	d-tyr		-46.2
	d-phe	pro	Arg	d-pro	phe		-41.1
	d-phe	pro	Arg	d-pro	Trp		-40.1
d-phe	phe	pro	Arg	d-pro	d-tyr		-39
Gly	d-phe	pro	Arg	d-pro	d-tyr		-30
	d-phe	pro	Arg	d-pro	Tyr		-28.5
	d-phe	pro	Arg	d-pro	d-tyr	Gly	16.2
Bz	d-phe	pro	Arg	d-pro	d-tyr		666.4
	d-phe	pro	d-arg				-29.6
	d-phe	val	d-arg	gly	asp		-32.5
	d-phe	val	d-arg	gly			-23
	d-phe	pro	d-arg	gly	phe		-43.1
	d-phe	pro	d-arg	gly	asp		-52.1
	d-phe	pro	d-arg	gly	Arg		-43.7
	d-phe	pro	d-arg	gly	Val		-37.8
	d-phe	pro	d-arg	gly	asn		-51.6
	d-phe	pro	d-arg	gly	d-asp		-38.9
	d-phe	pro	d-arg	gly	Glu		-20
	d-phe	pro	d-arg	gly	ser		-47.2
	d-phe	pro	d-arg	gly	Cys		-45
	d-phe	pro	d-arg	ala	asp		-51.9
	d-phe	pro	d-arg	d-ala	asp		-45.2
	d-phe	pro	d-arg	val	asp		-39.2
	d-phe	pro	d-arg	gly	asp	Ala	-58.9
	d-phe	pro	d-arg	gly	asp	Met	-56.3
	d-phe	pro	d-arg	gly	asp	Lys	-55.4
	d-phe	pro	d-arg	gly	asp	Gly	-51.5
	d-phe	pro	d-arg	gly	asp	Ile	-46.9
	d-phe	pro	d-arg	gly	asp	d-ala	-46.5
	d-phe	pro	d-arg	gly	asp	Cys	-45.3
	d-phe	pro	d-arg	gly	asp	Glu	-44.3
	d-phe	pro	d-arg	gly	asp	Phe	-30.4
	d-phe	pro	d-arg	phe			-42.7
	d-phe	pro	d-arg	glu			-28.5
	d-phe	pro	d-arg	arg			-34.5

d-phe	pro	d-arg	Tyr		-47.9
d-phe	pro	d-arg	Trp		-44.6
d-phe	pro	d-arg	Tyr	phe	-53.2
d-phe	pro	d-arg	Tyr	Arg	-45.6
d-phe	pro	d-arg	d-tyr		-36.5
d-phe	pro	d-arg	Tyr	Ala	-52.2

The *in silico* SAR was further extended to peptides containing amino acids analogs at P3 position, or unnatural amino acids analogs at P1' position. The docking experiments were performed as described in Materials and Methods for each variation in P3 or P1' with different amino acids analogs and the free energy of interaction between peptides and thrombin was assessed based on van der Waals interactions (sterical fitness) or based on both van der Waals and electrostatic interactions (Tables 8 and 9).

Table 8: The predicted free energy of interaction between thrombin and different peptides with amino-acids analogs determined from docking experiments performed with the software *"SCULPT".* The free energy was predicted based on the van der Waals interactions and is reflecting the relative sterical fitness between ligand and the target protein 1ABJ.pdb.

	Sequence N → C	**ΔG (Kcal/mol)**
1	**$^{\alpha}$N-CBZ-Pro-D-Arg-Gly-$CONH_2$**	**-71**
2	**Transcinnamic-Pro-D-Arg-Ser-$CONH_2$**	**-80.0**
3	**L-Tic-Pro-D-Arg-Cys-$CONH_2$**	**-88.9**
4	**L-Thi-Pro-D-Arg-Gly-$CONH_2$**	**-79.3**
5	**L-Tic-L-Thi-D-Arg-$CONH_2$**	**-68.3**
6	**DPhe-Pro-D-Arg-DThr-$CONH_2$**	**-90.6**
7	**L-Tic-Pro-D-Arg-DAla-$CONH_2$**	**-85.4**
8	**D-Phe-Pro-D-Arg-DCys-$CONH_2$**	**-88.8**
9	**Transcinnamic-Pro-D-Arg-Cys-$CONH_2$**	**-85.8**
10	**D-Phe-Pro-D-Arg-Met-$CONH_2$**	**-110.0**

11	**L-Tic-Pro-D-Arg-Thi-$CONH_2$**	**-85.2**
12	**Transcinnamic-Pro-D-Arg-Ile-$CONH_2$**	**-84.6**
13	**D-Phe-Pro-D-Arg-D-Ser-$CONH_2$**	**-92.0**
14	**D-Phe-Pro-D-Arg-D-Gln-$CONH_2$**	**-84.7**
15	**L-Tic-Pro-D-Arg-Gly-$CONH_2$**	**-83**
16	**D-Phe-Pro-D-Arg-Thi-$CONH_2$**	**-98.2**
17	**D-Phe-Pro-D-Arg-Thr-$CONH_2$**	**-80.2**
18	**$^{\alpha}$N-CBZ-Tyr-Pro-Arg-D-Pro-$CONH_2$**	**-114.3**
19	**L-Tic-Pro-D-Arg-$CONH_2$**	**-56.7**
20	**$^{\alpha}$N-CBZ-Pro-D-Arg-Ala-$CONH_2$**	**-73.2**
21	**$^{\alpha}$N-CBZ-Pro-Arg-D-Pro -$CONH_2$**	**+16,163**
22	**$^{\alpha}$N-CBZ-Pro-Arg-D-Pro-Gly-$CONH_2$**	**+ 559.7**
23	**$^{\alpha}$N-CBZ-Pro-D-Arg -Gly-$CONH_2$**	**-71**
24	**D-Phe-Pro-D-Arg-$CONH_2$**	**-56.6**
25	**Transcinnamic-Pro-D-Arg-$CONH_2$**	**-63.3**
26	**$^{\alpha}$N-CBZ-Phe-Pro-Arg-D-Pro-$CONH_2$**	**-116.4**
27	**$^{\alpha}$N-CBZ-Pro-D-Arg-D-Ala-$CONH_2$**	**-81.7**
28	**$^{\alpha}$N-CBZ-Tyr-Pro-Arg-D-Pro-$CONH_2$**	**-114.3**
29	**Dnaphthylalanine-Pro-D-Arg-Gly-$CONH_2$**	**-81.8**
30	**Dnaphthylalanine-Pro-D-Arg-DAla-$CONH_2$**	**-82.6**
31	**D-tetrahydroharman(Tpi)-Pro-D-Arg-DAla-**	**-80.0**
32	**D-tetrahydroharman(Tpi)-Pro-D-Arg-Ile**	**-71.3**
33	**Transcinnamic-Pro-D-Arg-D-Ala-$CONH_2$**	**-78.3**
34	**Ciscinnamic-Pro-D-Arg-Gly-CONH2**	**-74.3**
35	**Ciscinnamic-Pro-D-Arg-Ile-CONH2**	**-86.4**
36	**Ciscinnamic-Pro-D-Arg-DAla-CONH2**	**-74.0**
37	**Ciscinnamic-Pro-D-Arg-Ser -CONH2**	**-82.7**
38	**Ciscinnamic-Pro-D-Arg-CONH2**	**-67.0**
39	**Dihydrocinnamic-Pro-D-Arg-Gly-CONH2**	**-73.3**
40	**Dihydrocinnamic-Pro-D-Arg-Cys-CONH2**	**-80.3**

41	**D-Tic-Pro-D-Arg-Gly-CONH2**	**-83.0**
42	**D-Tic-Pro-D-Arg-Cys-CONH2**	**-83.8**

Table 9:The predicted free energy of interaction between thrombin and different peptides *trans*cinnamic and *dihydro*cinamic analogs of Phe (P3) determined from docking experiments performed with the software *"SCULPT".* The free energy was predicted based on the van der Waals and electrostatic interactions between ligand and the target protein 1ABJ.pdb.

Inhibitor (Tetrapeptides)	**Binding Energy (van der Waals + Electrostatics (kcal/mol)**
***Trans*-cinnamoyl-Pro- D-Arg-D-Pro**	**-138.0**
***Trans*-cinnamoyl-Pro-D-Arg-D-His**	**-148.3**
***Trans*-cinnamoyl-Pro-D-Arg-D-Ala**	**-120.5**
***Trans*-cinnamoyl-Pro-D-Arg-D-Ser**	**-172.0**
***Trans*-cinnamoyl-Pro-D-Arg-D-Cys**	**-110.0**
***Trans*-cinnamoyl-Pro-D-Arg-D-Thr**	**-151.3**
***Trans*-cinnamoyl-Pro-D-Arg-L-Thi**	**-165.5**
***Trans*-cinnamoyl-Pro-D-Arg-D-Tic**	**-115.0**
***Trans*-cinnamoyl-Pro-D-Arg-D-Asn**	**-132.9**
***Trans*-cinnamoyl-Pro-D-Arg-D-Gln**	**-148.5**
***Trans*-cinnamoyl-Pro-D-Arg-D-3-Benzothienylalanine**	**-163.9**
***Trans*-cinnamoyl-Pro-D-Arg-D-3,5-DifluoroPhenylalanine**	**-166.5**
***Trans*-cinnamoyl-Pro-D-Arg-Gly**	**-132.0**
Dihyrocinnamoyl-Pro-D-Arg-D-Pro	**-183.5**
Dihyrocinnamoyl-Pro-D-Arg-D-His	**-164.7**
Dihyrocinnamoyl-Pro-D-Arg-D-Ala	**-152.2**
Dihyrocinnamoyl-Pro-D-Arg-D-Ser	**-165.1**
Dihyrocinnamoyl-Pro-D-Arg-D-Cys	**-114.7**
Dihyrocinnamoyl-Pro-D-Arg-D-Thr	**-167.3**
Dihyrocinnamoyl-Pro-D-Arg-L-Thi	**-183.2**
Dihyrocinnamoyl-Pro-D-Arg-D-Tic	**-146.7**
Dihyrocinnamoyl-Pro-D-Arg-D-Asn	**-157.4**
Dihyrocinnamoyl-Pro-D-Arg-D-Gln	**-195.7**
Dihyrocinnamoyl-Pro-D-Arg-D-3-Benzothienylalanine	**-167.8**
Dihyrocinnamoyl-Pro-D-Arg-Gly	**-122.9**

We were interested to validate the accuracy of the proposed models peptides-thrombin complexes by performing an independent docking experiment using another template 1AI8.pdb which was also checked through PROCHECK for its structural accuracy (see section 1.2). The results from table 10 are showing that overall there are relatively low differences between the predicted free energy of interaction between peptides docked in 1ABJ.pdb and 1AI8.pdb.

Table 10: The comparison between the predicted free energy of interaction between two different thrombin templates (1ABj.pdb and 1AI8.pdb) and lead peptide inhibitors. The docking experiments were performed with the software *"SCULPT".* The free energy was predicted based on the van der Waals interactions and is reflecting the relative sterical fitness between ligand and the target protein 1ABJ.pdb (± 5-10-kcal/mol is within the sensitivity of software SCULPT).

Sequence N → C	**ΔG (Kcal/mol) (1ABJ.pdb template)**	**ΔG (Kcal/mol) (1AI8.pdb template)**
D-Phe-Pro-D-Arg-Ala-$CONH_2$	**-77.3**	**-77.2**
D-Phe-Pro-D-Arg-D-Ala- $CONH_2$	**-80.3**	**-75.2**
D-Phe-Pro-D-Arg- Gly-$CONH_2$	**-76.6**	**-75.0**
D-Phe-Pro-D-Arg-L-Arg- $CONH_2$	**-87.2**	**-83.2**
D-Phe-Pro-D-Arg-l-Glu- $CONH_2$	**-87.8**	**-73.1**
D-Phe-Pro-D-Arg-L-His- $CONH_2$	**-88.2**	**83.4**
D-Phe-Pro-D-Arg-L-Phe- $CONH_2$	**-87.1**	**-78.6**

D-Phe-Pro-D-Arg-L-Pro $CONH_2$	**-76.4**	**-78.8**
D-Phe-Pro-D-Arg-D-Pro- $CONH_2$	**-70.0**	**-76.0**
D-Phe-Pro-D-Arg-L-Trp- $CONH_2$	**-86.4**	**-96.8**
D-Phe-Pro-D-Arg-L-Tyr- $CONH_2$	**-73.9**	**-80.1**
D-Phe-Pro-D-Arg-L-Val- $CONH_2$	**-83.3**	**-82.4**
D-Phe-Pro-D-Arg-L-Thr- $CONH_2$	**-78.6**	**-79.6**
D-Phe-Pro-D-Arg-D-Thr- $CONH_2$	**-90.4**	**-79.5**
D-Phe-Pro-D-Arg-L-Ser- $CONH_2$	**-90.8**	**-82.5**
D-Phe-Pro-D-Arg-D-Ser-$CONH_2$	**-99.4**	**-80.4**
D-Phe-Pro-D-Arg-L-Cys- $CONH_2$	**-88.0**	**-80.8**
D-Phe-Pro-D-Arg-D-Cys- $CONH_2$	**-104.5**	**-77.6**
D-Phe-Pro-D-Arg-L-Gln- $CONH_2$	**-94.2**	**84.5**
D-Phe-Pro-D-Arg-D-Gln- $CONH_2$	**-104.6**	**-87.1**
D-Phe-Pro-D-Arg-Ile- $CONH_2$	**-83.0**	**86.1**
D-Phe-Pro-D-Arg-L-Thi- $CONH_2$	**-98.0**	**-91.2**
D-Phe-Pro-D-Arg-L-Met- $CONH_2$	**-109.0**	**-107.0**

In order to visualize the environment of each ligand within the active site of thrombin, the pdb files from docking experiments were analyzed with the software SCULPT. In figure 10 the details of the active site of thrombin within 10 Å of some of the designed peptide

inhibitors are presented as molecular models performed with the software SPDB Viewer. The main hydrogen bonding between peptide and amino acids within the subpockets S3, S2, S1 and S1' in the active site of thrombin are presented as dashed lines. It can be seen that the most important hydrogen bonds such as between Asp 189 in the S1 pocket of thrombin and guanidinium group of D-Arg in the ligand is conserved through different classes of inhibitors. Of importance to note is the difference in the hydrogen patterning between these two keys groups as the amino acid in P1' position in the peptide is changing from L to D, such as is the case of L-Thr and D-Thr (figure 101 A, B, C and D). L-Thr in P1' is predicted to have less favorable hydrogen bonds than D-Thr, i.e. 2 hydrogen bonds instead of three (compare figure 101 A with 101 C and 101 D and 101E).

Figure 101 (A) D-Phe-Pro-D-Arg-D-Thr-$CONH_2$ (depicted in green color)

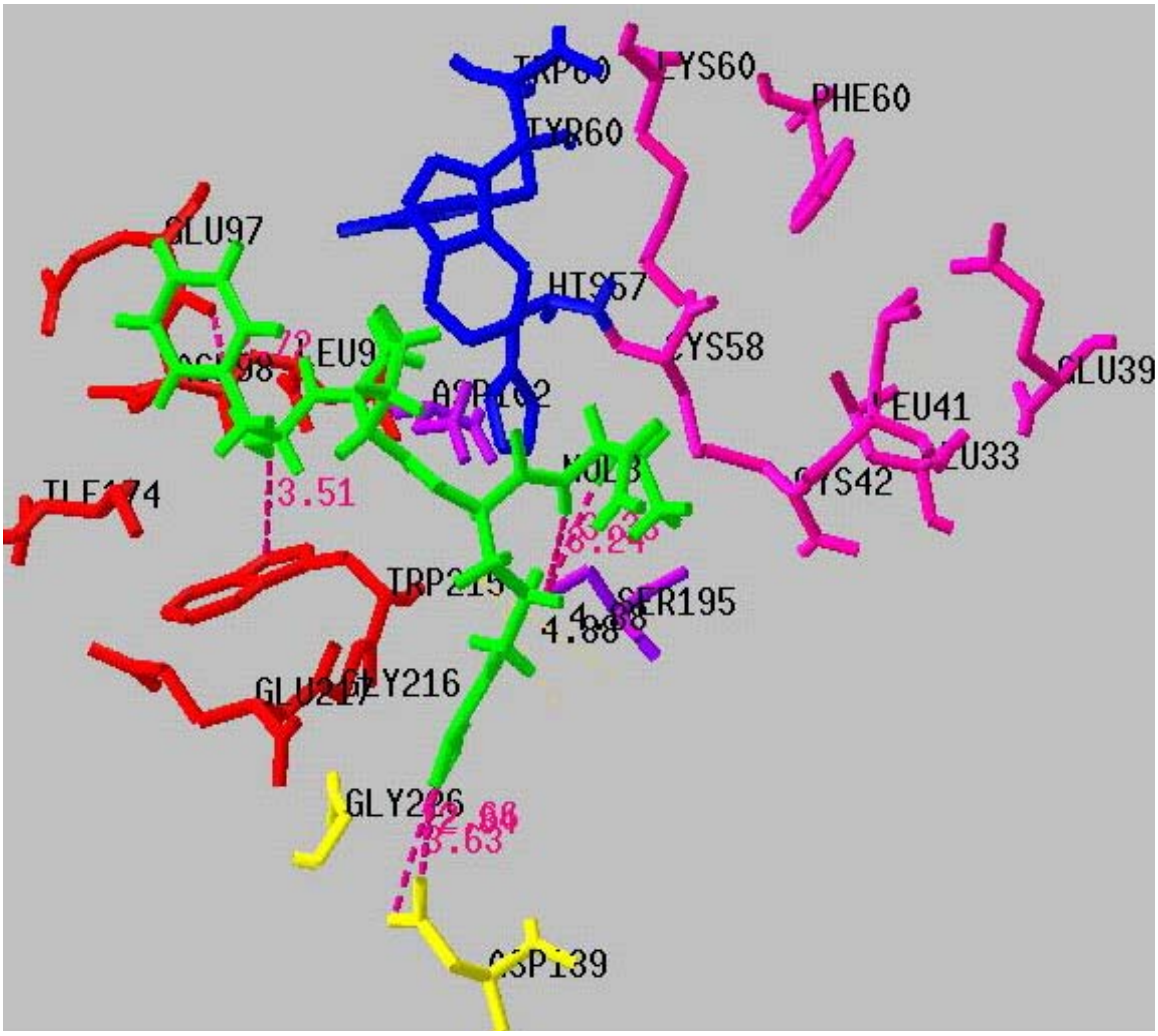


Figure 101 (B) D-Phe-Pro-D-Arg-D-Thr-$CONH_2$ (Van der Waals representation of peptide ligand as colored in green).

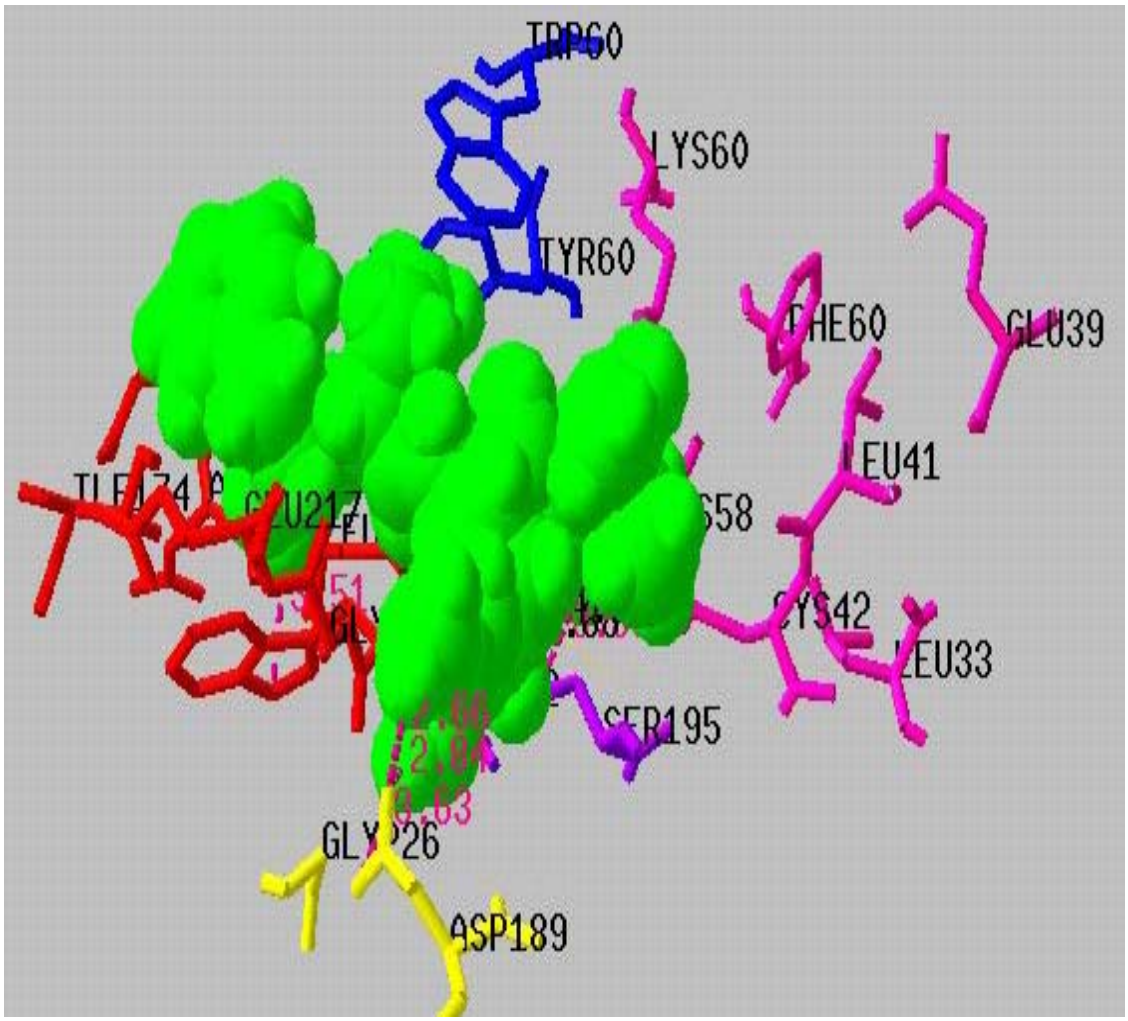


Figure 101 (C) D-Phe-Pro-D-Arg-L-Thr-CONH2 (blue color).

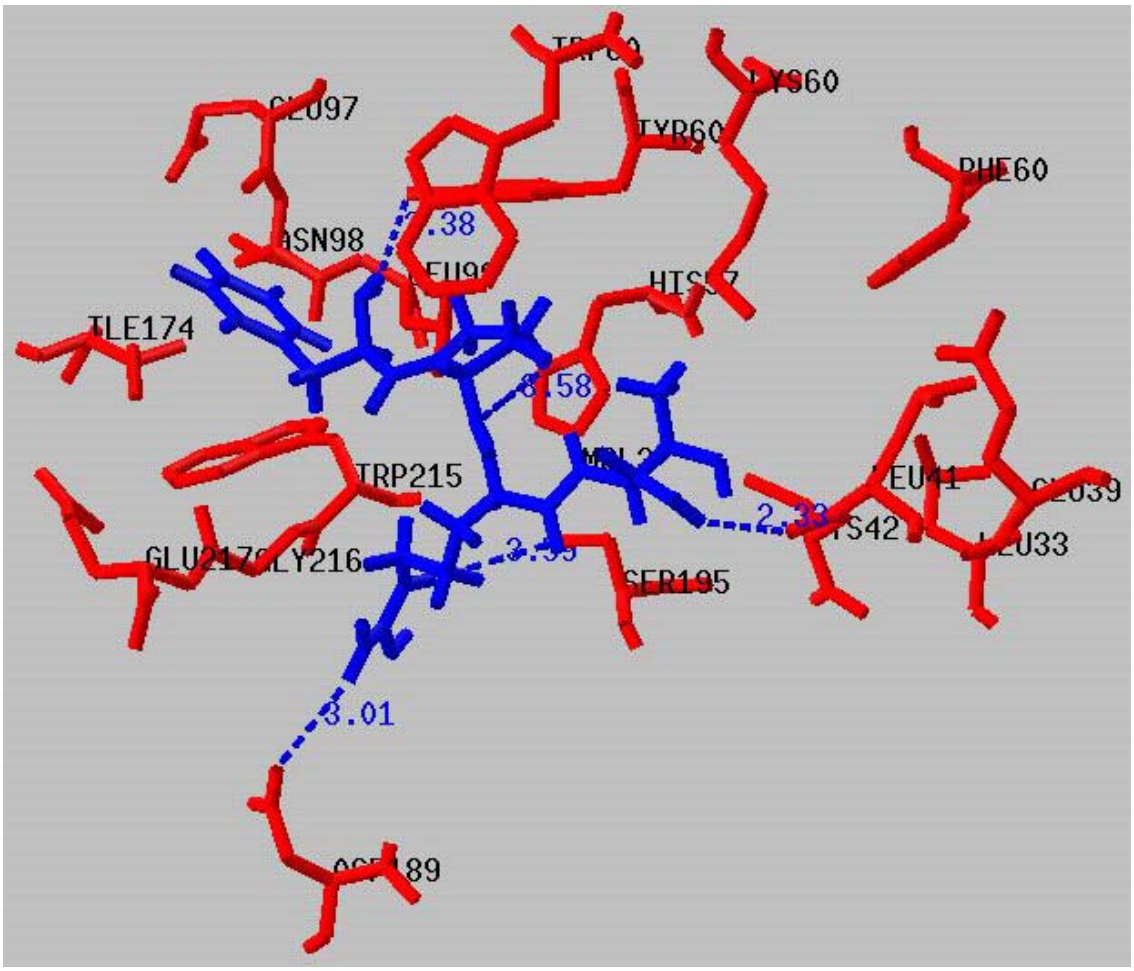
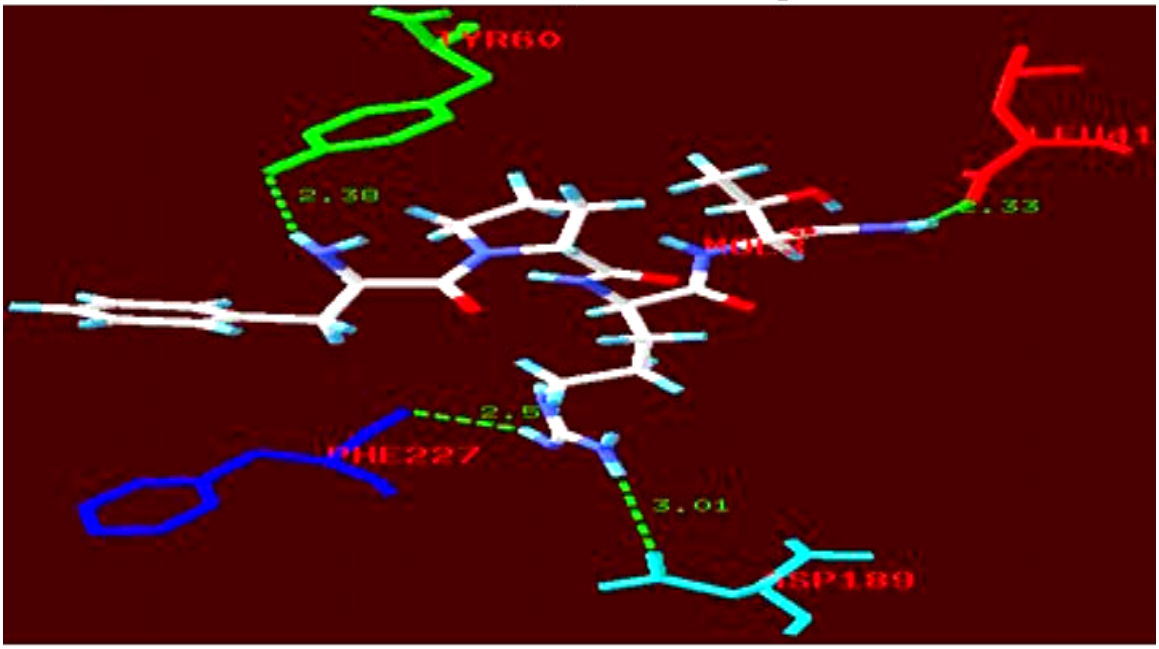


Figure 101 (D) (the experimentally determined Kis (inhibitory constants) are indicated for each peptide in uM (micromolar).

D-Phe-L-Pro-D-Arg-L-Thr-$CONH_2$ = 12.5 uM



D-Phe-L-Pro-D-Arg-D-Thr-$CONH_2$ = 0.920 uM

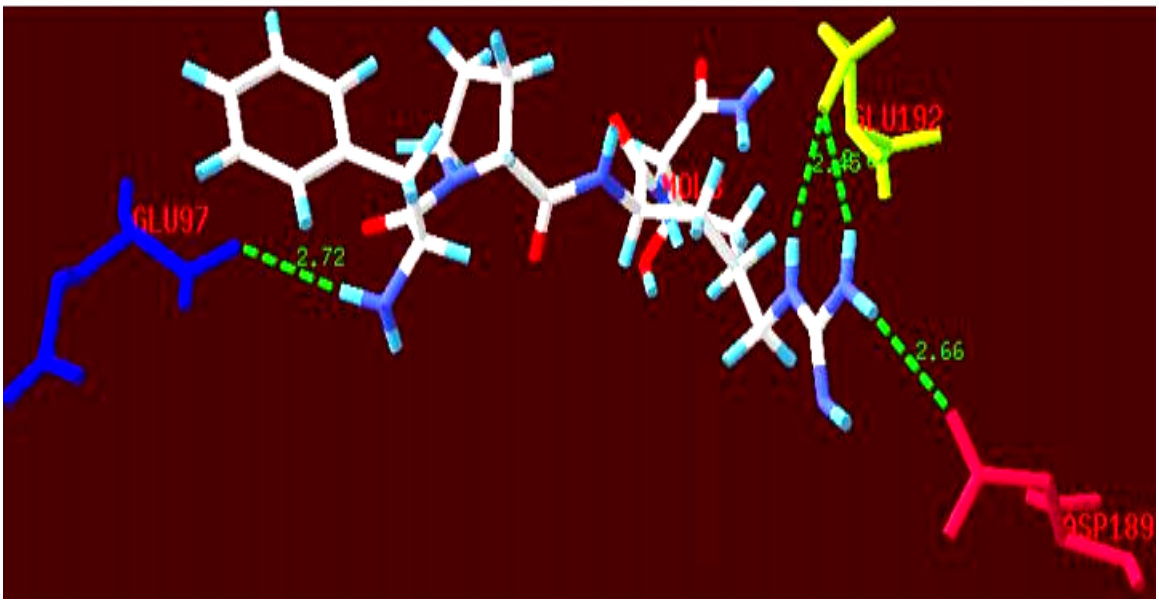


Figure 101 (E): superimposed L-Thr (P1') and D-Thr (P1') tetrapeptides from 10 (D) showing the different conformation adopted by each tetrapeptide as the configuration of the amino acid at P1' position is shifted from L to D.

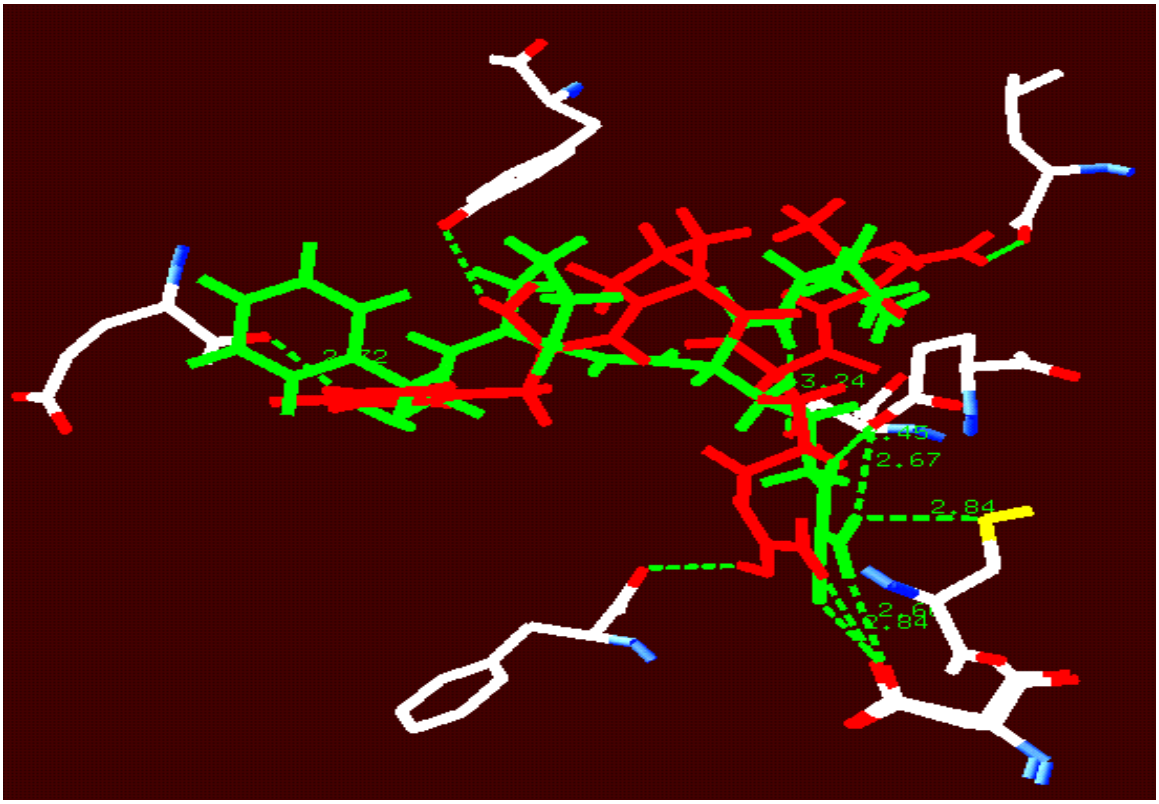


Figure 101 (A-E): Molecular modeling of peptides docked into active site of thrombin 1ABJ.pdb. The models were generated with SPDB-Viewer using the pdb files generated by the software SCULPT during docking experiments.

Additional docking experiments were performed with unnatural amino acids scanning P3, P2 and P1' positions as presented in Tables 8-9. A very promising amino acid analog at P1' was L-thienylalanine (L-Thi) which was predicted to occupy the S1' subpocket of thrombin's active site with a free energy of interaction of -95 kcal/mol (based on SCULPT docking experiments using the Van der Waals force-field) (Figure 11). This amino-acid analog of phenylalanine was scanned by others (67) but within different sequence space and was found to be very well fitted sterically in the small hydrophobic pocket S1' (based on X-ray diffraction of complexes between the peptide inhibitor containing L-Thi at P1' position and thrombin (67).

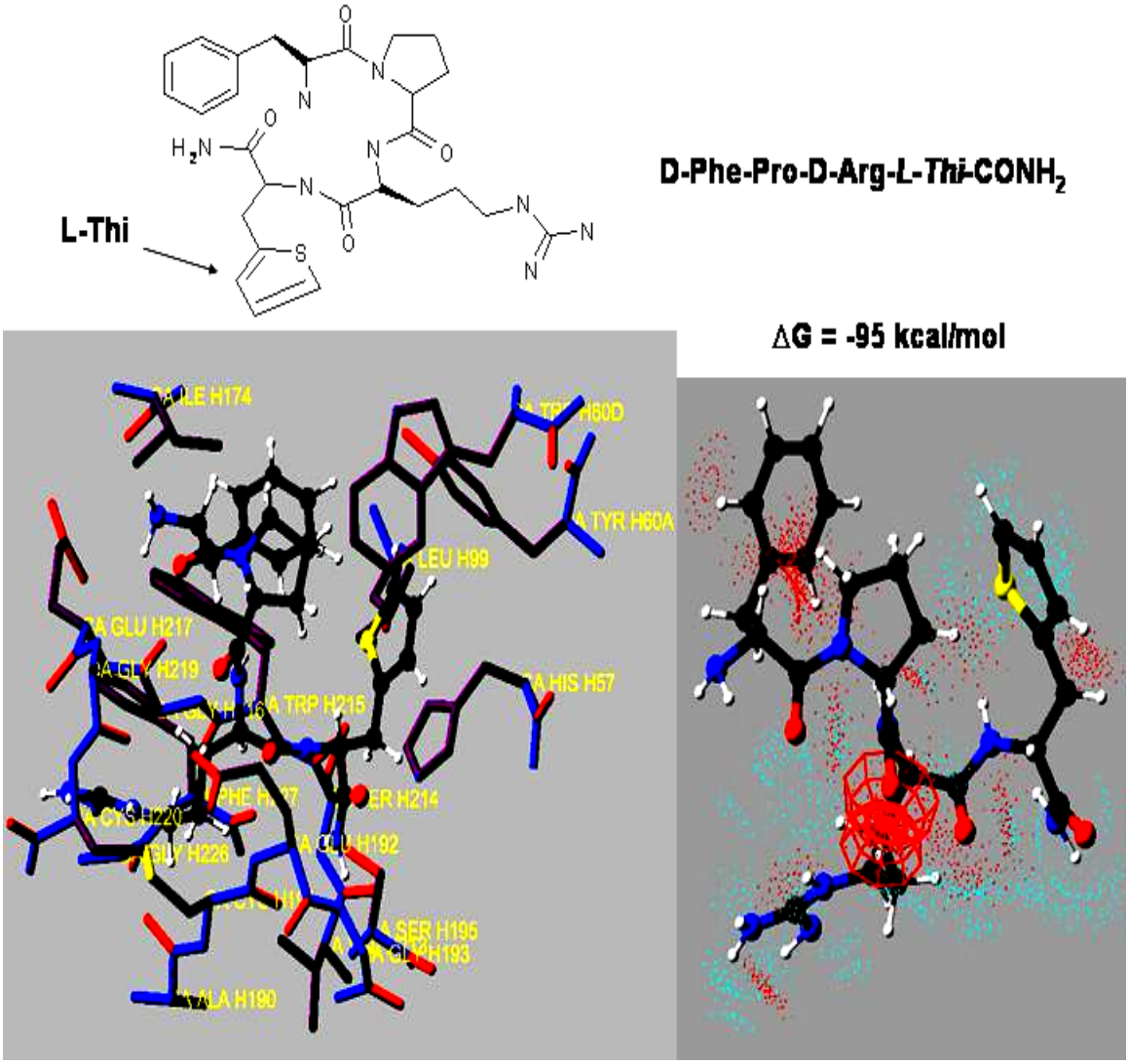


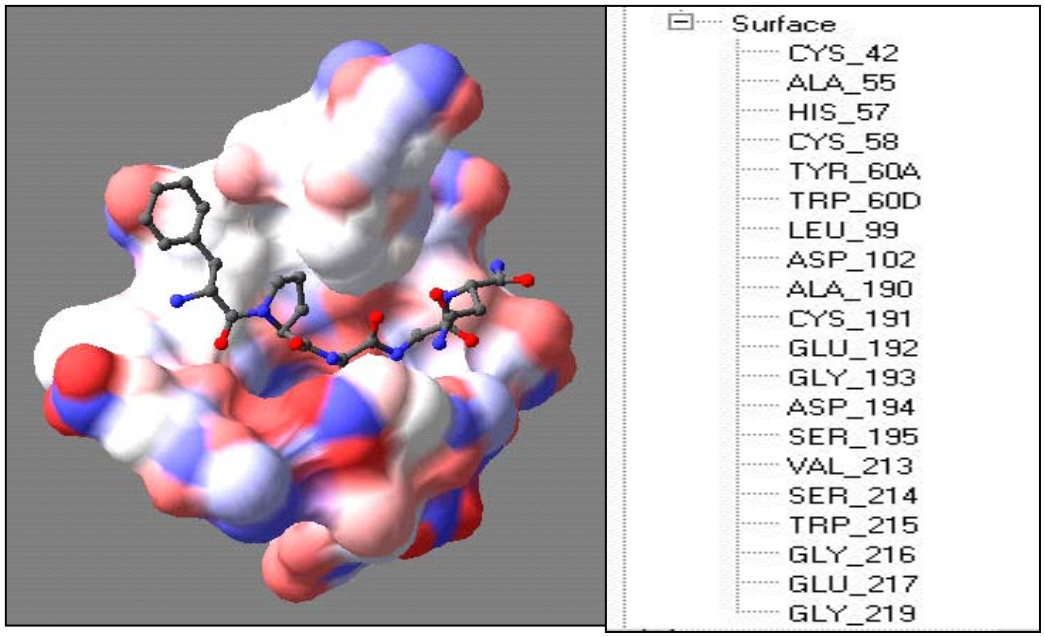
Figure 102. Molecular modeling of a new peptide inhibitor containing L-Thi (thienylalanine) docked into active site of thrombin 1ABJ.pdb. The models were generated with SPDB-Viewer using the pdb files generated by the software SCULPT.

Other examples of computer generated complexes between the new designed peptide inhibitors in complex with thrombin are presented in figure 103 (A), (B), (C) and figure 104 (A-F). The electrostatic surface within 10-12 Å of the ligand was generated with the software "SCULPT". The amino acids from active site of thrombin which were found within 10-12 Å of the ligand are those corresponding to the S3, S2, S1, S1' subpockets of the thrombin as they were described in many X-ray structures (67-81).

(A)

D-Phe-Pro-D-Arg-Gly-Asn-$CONH_2$

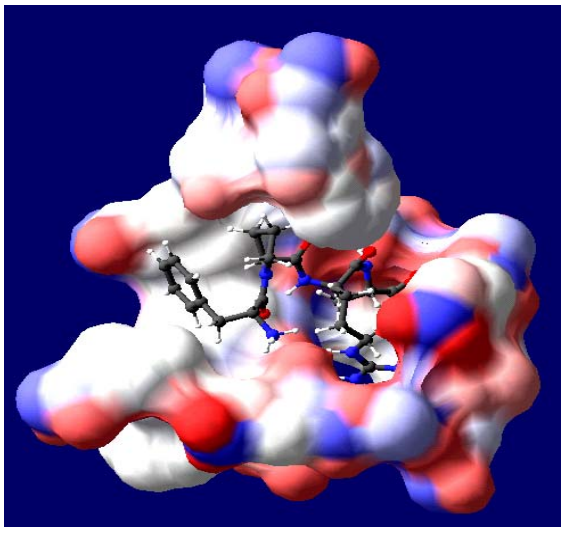
Active-site Electrostatic Surface within 10 Å from the ligand



B)

D-Phe-Pro-DArg-*Ala*-$CONH_2$

ΔG =- 69.4 [kcal/mol]



C)

D-Phe-Pro-D-Arg-*Gly*-$CONH_2$

ΔG = - 70.5 [kcal/mol]

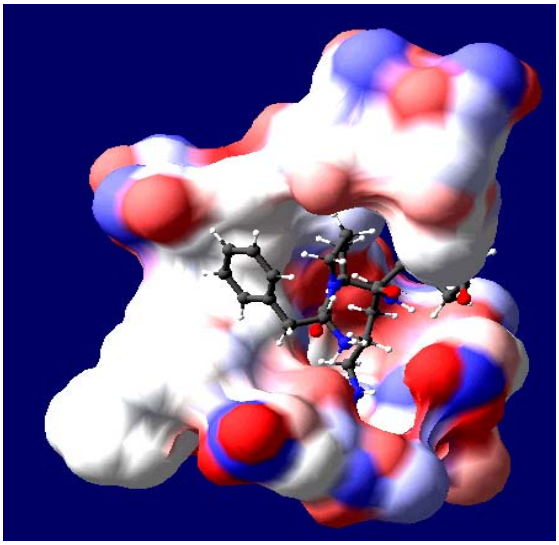
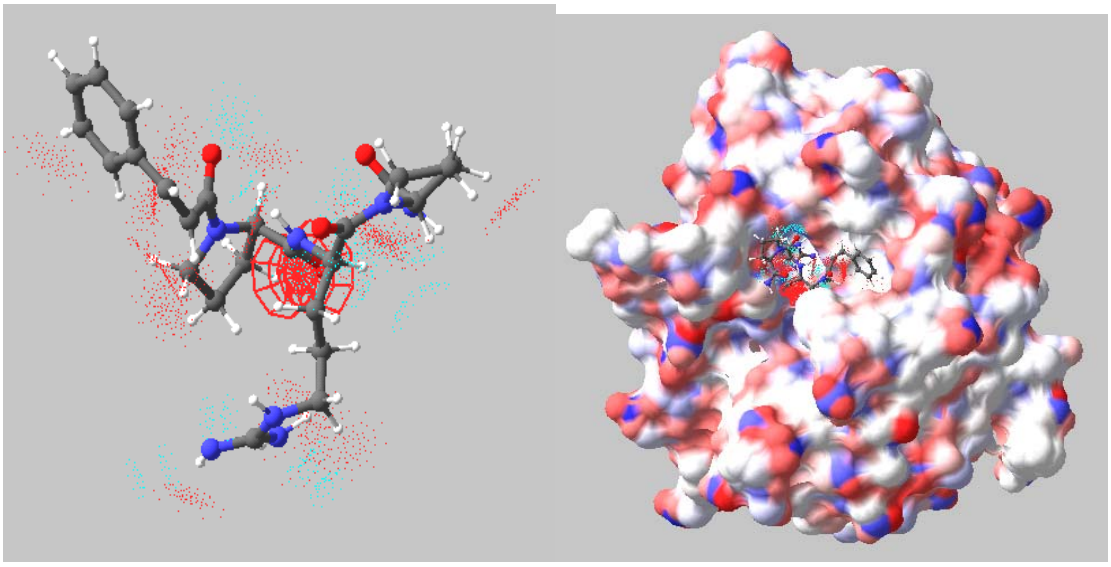
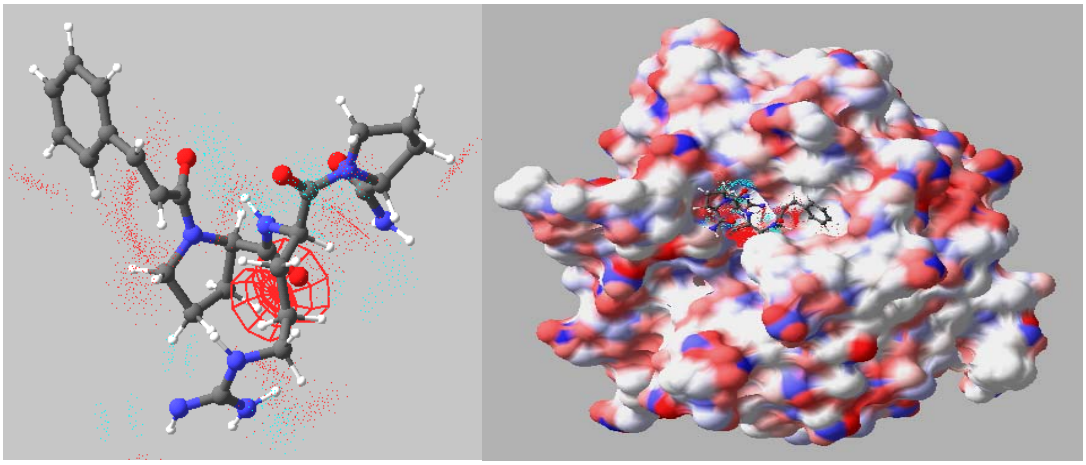


Figure 103. Electrostatic surface representation (within 10-12 Å of the docked peptide ligand within the active site of thrombin). The electrostatic potential was generated with the docking software SCULPT using the built-in MMFF94 force-field.

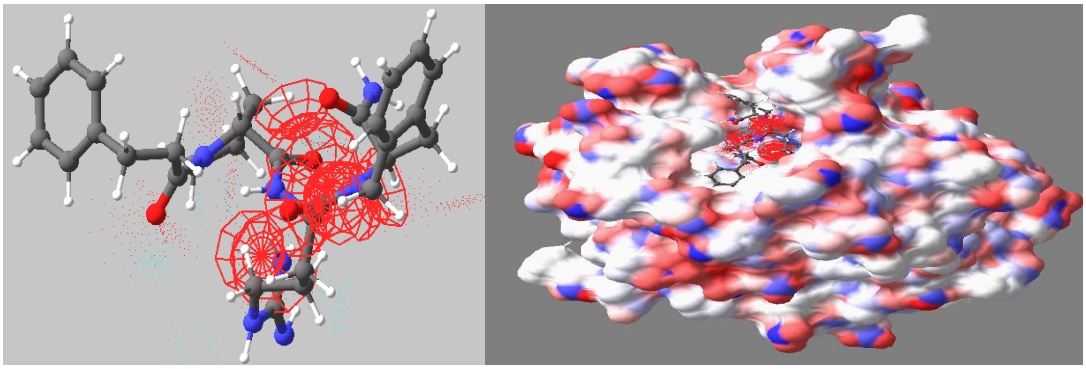
(A)



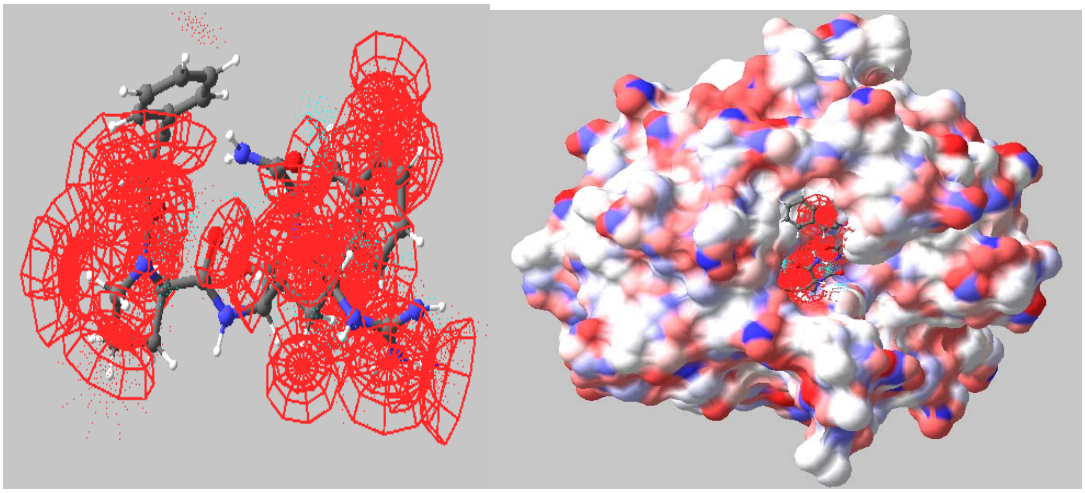
(B)



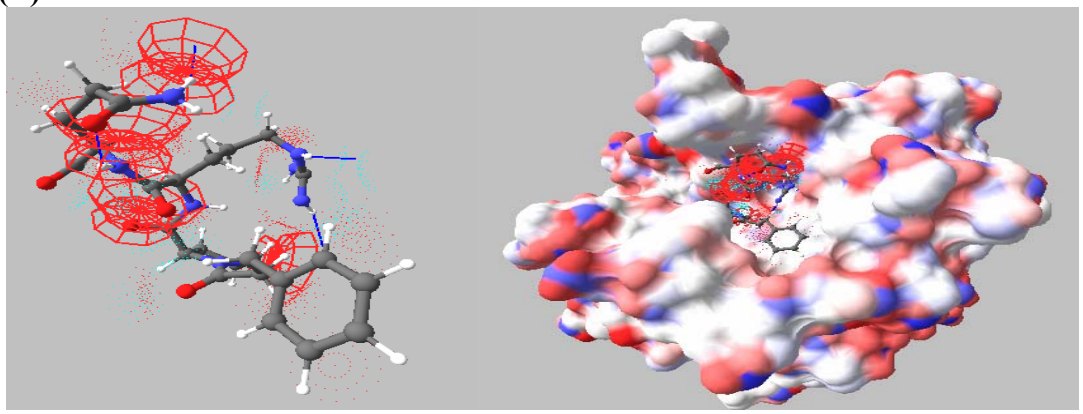
(C)



(D)



(E)



(F)

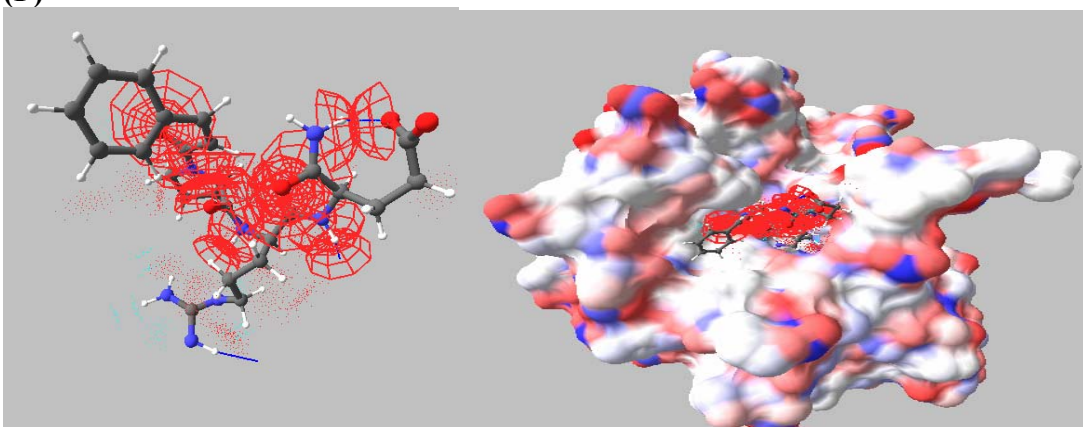


Figure 104: Molecular models of tetrapeptides with transcinnamic and dihydrocinnamic acids at P3 position in the ligand peptide docked into active site of thrombin. (A) *Trans*-cinnamoyl-Pro-Arg-D-Pro-$CONH_2$ (B) Dihydrocinnamoyl-Pro-Arg-D-Pro-$CONH_2$ (C) Dihydrocinnamoyl-Pro-Arg-D-Tic-$CONH_2$ (D) *Trans*-cinnamoyl-Pro-Arg-D-Tic-$CONH_2$ (E) Dihydrocinnamoyl-Pro-Arg-D-Gln-$CONH_2$ (F) *Trans*-cinnamoyl-Pro-Arg-D-Gln-$CONH_2$ (C and D) has one of the lowest binding energies; (E and F) has one of the highest binding energies (table 9).

In order to assess the validity of the predicted free energy, we performed "positive control" type of docking experiments in which a well known competitive inhibitor for thrombin ("*Thromstop"*) was docked into the active site of thrombin. We expected a negative free energy as predicted by SCULPT. Figure 105 showed that this was indeed a good control experiment, since the software SCULPT predicted a favourable free energy of interaction between thrombin and the inhibitor. Also, we performed control docking experiments to assess the selectivity of peptide inhibitors for thrombin as compared with other serine-proteases from coagulation cascade (like Factor X-a). In these type of experiments the peptide inhibitor specific for thrombin, with a very good predicted free energy of interaction with thrombin (< -50 kcal/mol) was docked into the active site of factor Xa. We expected a lower affinity between these peptides and Factor Xa. This would be predicted as higher free energy of interaction between the peptide and Factor X-

a (toward more positive values, i.e. less favorable sterical fitness between ligand and protein template) (table 11).

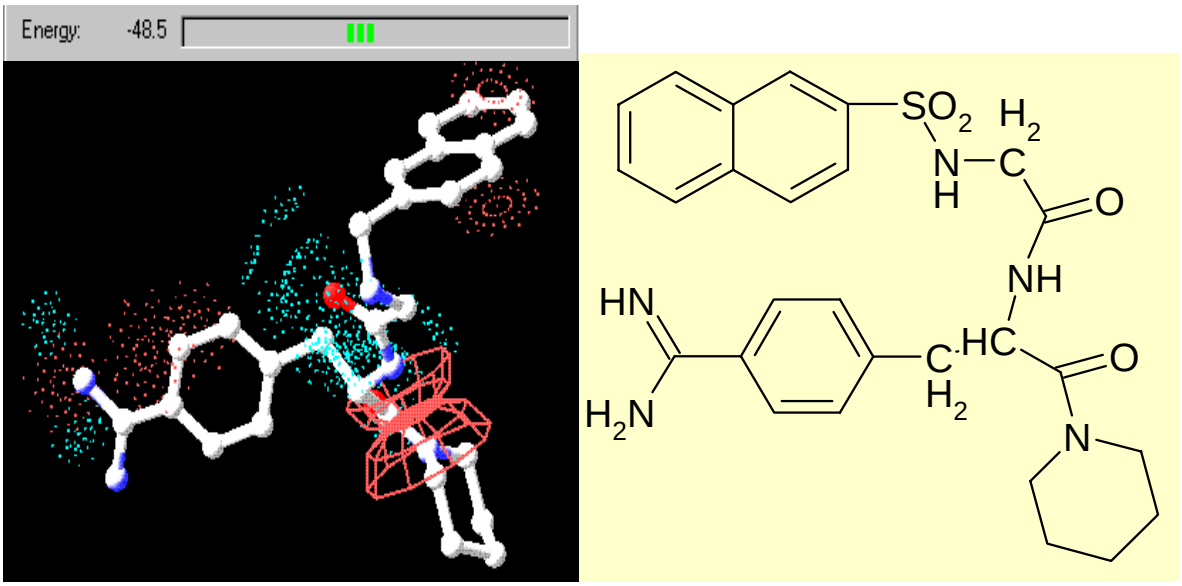


Figure 105. The "SCULPT" presentation of the model of 'Thromstop' (Ki = 25 nM for bovine thrombin) docked into active site of thrombin. The commercially available inhibitor "Thromstop" (N-α-NAPAP) was docked into active site of thrombin, using the same thrombin template as for the other peptide inhibitors, i.e. 1ABJ.pdb.

In Table 11 all peptides docked in 1ABJ.pdb and 1AI8.pdb presented in table 5 were docked within the active site of Factor X-a and the free energy of their interaction with Factor X-a was predicted as described in Materials and Methods. These results are confirming that the designed peptides are more selective for thrombin than for Factor X-a, since their free energy of interaction with thrombin has lower negative values than energy predicted for their interaction with Factor X-a. It can be seen that there no positive values but the free energy is raising toward more positive values with 15-20 kcal/mol (compare tables 10 and 11). Altogether, the results from docking experiments are supporting "SCULPT" as a reliable docking software in predicting at least a favorable

(negative) vs. a non-favorable (toward more positive) free energy of interaction between the ligand and the target protein.

Table 11: Docking-control experiment for selectivity of peptides designed as inhibitors for thrombin. The predicted free energy of interaction between Factor X-a (1FAX.pdb) and lead peptide inhibitors is presented in kcal/mol. The docking experiments were performed with the software "*SCULPT*". The free energy was predicted based on the van der Waals interactions and is reflecting the relative sterical fitness between ligand and the target protein 1FAX.pdb (± 5 kcal/mol is within the accuracy of software SCULPT).

Sequence N → C	**ΔG (Kcal/mol) (1FAX.pdb template)**
D-Phe-Pro-D-Arg-Ala-$CONH_2$	**-56.1**
D-Phe-Pro-D-Arg-D-Ala- $CONH_2$	**-58.5**
D-Phe-Pro-D-Arg- Gly-$CONH_2$	**-59.2**
D-Phe-Pro-D-Arg-L-Arg- $CONH_2$	**--75**
D-Phe-Pro-D-Arg-l-Glu- $CONH_2$	**-78.8**
D-Phe-Pro-D-Arg-L-His- $CONH_2$	**-83.4**
D-Phe-Pro-D-Arg-L-Phe- $CONH_2$	**-72.9**
D-Phe-Pro-D-Arg-L-Pro $CONH_2$	**-65.8**
D-Phe-Pro-D-Arg-D-Pro- $CONH_2$	**-63.3**
D-Phe-Pro-D-Arg-L-Trp- $CONH_2$	**-70.0**
D-Phe-Pro-D-Arg-L-Tyr- $CONH_2$	**-68.4**

D-Phe-Pro-D-Arg-L-Val- $CONH_2$	**-67.5**
D-Phe-Pro-D-Arg-L-Thr- $CONH_2$	**-64.8**
D-Phe-Pro-D-Arg-D-Thr- $CONH_2$	**-71.6**
D-Phe-Pro-D-Arg-L-Ser- $CONH_2$	**-76.5**
D-Phe-Pro-D-Arg-D-Ser-$CONH_2$	**-77.3**
D-Phe-Pro-D-Arg-L-Cys- $CONH_2$	**-66.5**
D-Phe-Pro-D-Arg-D-Cys- $CONH_2$	**-69.7**
D-Phe-Pro-D-Arg-L-Gln- $CONH_2$	**-80.7**
D-Phe-Pro-D-Arg-D-Gln- $CONH_2$	**-78.4**
D-Phe-Pro-D-Arg-Ile- $CONH_2$	**-68.4**
D-Phe-Pro-D-Arg-L-Thi- $CONH_2$	**-75.7**
D-Phe-Pro-D-Arg-L-Met- $CONH_2$	**-62.7**

1.5. Scoring Functions for discovery of lead peptides inhibitors for thrombin.

Scoring Function (I): In order to have a quantitative value for the docked structures the free energy of interaction between different peptides and thrombin was considered as the *first scoring function* to rank the peptides inhibitors. Thus, the initial *in silico* library of peptides contained compounds characterized by a free energy of interaction with thrombin template of less than "-10 kcal/mol". Compounds with positive

free energy of interaction were discarded, but for validation reasons some of them were chemically synthesized and their activity was tested. This free energy (ΔG) (kcal/mol) for the interaction between peptides and thrombin was calculated primarily based on Van der Waals interactions between ligands and thrombin.

Since specific contacts between ligand and protein are known from solved X-ray structures (68-86 and figures 94 and 101), new scoring functions were developed by assessing the structural details of the peptides docked into the active site.

Scoring Function (II): It is already known that the salt bridge between Asp 189 (from the specificity pocket in the active site of thrombin) and the Arg in the P1 position of the inhibitors contributes to a big fraction from the energy of interaction between thrombin and different inhibitors (41-53). Based on these structural data we proposed the distances between (D/L) Arg (at P1 position) in the peptides and the Asp 189 in the specificity pocket S1 of thrombin as an important parameter for predicting the structural fitness of the peptide ligand into the active site of thrombin (this was the second scoring function criteria for discovery of lead compounds).

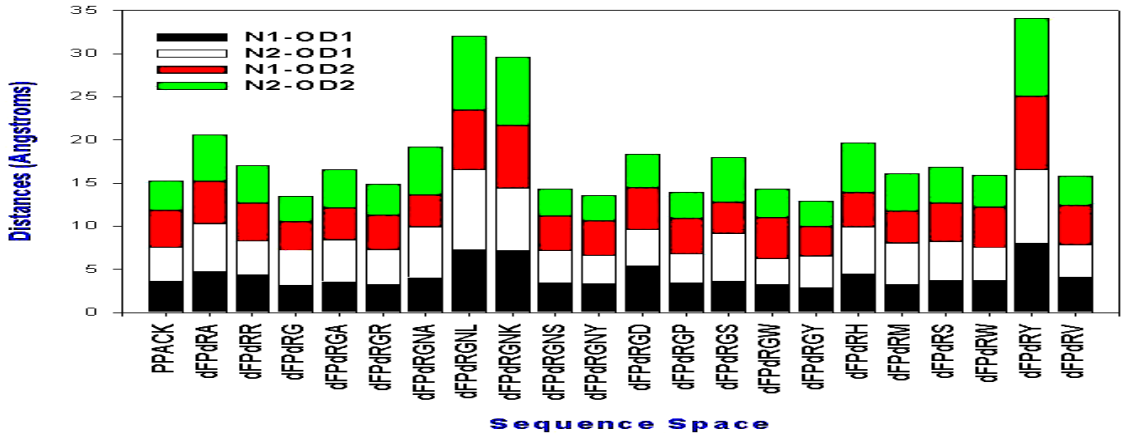


Figure 106: Scoring function II (the interatomic distance between Asp 189 in thrombin and D-Arg at P1 position in peptide).

As can be seen from Figure 106 the majority of tetrapeptides are predicted to have very similar distances between Asp (189) from thrombin (OD1 and OD2 of the carboxyl group) and D-Arg from peptide inhibitor (N1 and N2 of the guanidinium group) as compared with similar distances from the control inhibitor, PPACK. However, the tetrapeptide D-Phe-Pro-D-Arg-Tyr is predicted to have higher than 5 Å interatomic distance (Asp 189-D-arg (in P1-peptide) so this peptide is predicted to have less binding interaction with thrombin. This was a successful prediction performed with SCULPT since the experimental results demonstrated the aromatic amino acids in P1' position are having at least 10 fold less inhibitory activity (and thus less affinity) toward thrombin as compared with Gly or other polar amino acids at P1' (see section 3.2 and Table 18). Two heaxapeptides D-Phe-Pro-dArg-Gly-Asn-Leu and D-Phe-Pro-D-Arg-Gly-Asn-Lys have more than 7.0 Å interatomic distance between the above mentioned atoms (the distances between OD1 and OD2 of the carboxyl group from Asp 189 and N1 and N2 of the guanidinium group from the D-Arg) suggesting that he hexapeptides are less potent in binding thrombin, again a result which was supported by experimental data (table 13 and section 3.2). All hexappetides studied were at least 50 fold less effective than the tetrapeptides in inhibiting thrombin. Based on these good correlations between computation docking models and the experimental results, the salt bridges with 3-4 Å interatomic distance between Asp 189-thrombin and-D-Arg (P1-peptide) were proposed to predict a good affinity between ligand and thrombin and were used as a reliable scoring function in assesing the lead compounds.

Scoring Function (III): Another scoring function was developed using the analysis of the structural details of the docked peptides with respect to the % favorable contacts and normalized complementarity between the ligand (peptide inhibitor) and thrombin

active site. The % favorable contacts and normalized complementarity were analyzed using the algorithm developed by Sobolev V., Wade R.C., Vriend G. and Edelman M (88-91). The complementarity function is defined as $CF=S_l-S_i-E$ where S_l is the sum of all surface areas of legitimate atomic contacts between ligand and receptor, S_i is the sum of all surface areas of illegitimate atomic contacts, and E is a repulsion term. Legitimacy depends on the hydrophobic-hydrophilic properties of the contacting atoms. In order to define it, for each inter-atomic contact, eight atom classes have been introduced (88-91):

I Hydrophilic - N and O that can donate and accept hydrogen bonds

(e.g., oxygen of hydroxyl group of Ser. or Thr)

II Acceptor - N or O that can only accept a hydrogen bond

III Donor - N that can only donate a hydrogen bond

IV Hydrophobic - Cl, Br, I and all C atoms that are not in aromatic rings and do not have a covalent bond to
a N or O atom

V Aromatic - C in aromatic rings irrespective of any other bonds formed by the atom

VI Neutral - C atoms that have a covalent bond to at least one atom of class I or two or more atoms from class II or III; atoms; S, F, P, and metal atoms in all cases

VII Neutral-donor - C atoms that have a covalent bond with only one atom of class III

VIII Neutral-acceptor - C atoms that have a covalent bond with only one atom of class II

For each pair of contacts the state of legitimacy is shown below:

+,legitimate

-, illegitimate (88-91).

All the lead compounds having predicted free energy of interaction with thrombin less than "-50 kcal/mol) were analysed using the above mentioned scoring functions, by sending *ONLINE* the "pdb" files of the peptides docked into active site of thrombin to the

server located at Weizmann Institute (Israel) which is providing the structural analysis of the ligand-protein. **Figure 107** presents the results from LPC analysis of peptides-thrombin complexes in the case of control PPACK-thrombin **(A)** and three different peptides inhibitors**:** D-Phe-Pro-D-Arg-Ala-$CONH_2$ (B) D-Phe-Pro-D-Arg-Gly-$CONH_2$ (C) and D-Phe-Pro-Arg-D-Pro-D-Arg-$CONH_2$ (D).

				Specific contacts			
Residue		Dist	Surf	HB	Arom	Phob	DC
41H	LEU	5.2	1.3	-	-	-	+
42H	CYS*	5.2	5.9	-	-	-	-
57H	HIS*	3.3	49.2	+	+	+	-
60H	TRP*	3.4	101.4	+	-	-	+
97H	GLU	3.1	29.2	-	-	-	-
98H	ASN*	3.4	28.3	-	-	-	-
99H	LEU*	3.6	32.2	-	-	+	+
174H	ILE*	3.6	27.1	-	-	+	-
189H	ASP*	3.3	28.9	+	-	-	-
190H	ALA*	3.4	28.7	+	-	-	+
191H	CYS*	3.5	22.8	-	-	-	+
192H	GLU*	3.6	46.3	+	-	+	+
193H	GLY*	3.1	30.1	-	-	-	+
195H	SER*	3.3	26.5	-	-	-	+
213H	VAL*	4.3	12.4	-	-	-	+
214H	SER	3.8	4.6	+	-	-	+
215H	TRP*	3.2	85.5	+	+	+	+
216H	GLY*	3.3	42.3	+	-	-	+
217H	GLU*	4.4	6.0	+	-	+	+
219H	GLY	3.2	25.3	+	-	-	-
220H	CYS*	4.2	0.8	-	-	-	-
225H	TYR	5.3	0.6	+	-	-	-
226H	GLY*	3.9	10.1	-	-	-	-
227H	PHE	4.8	0.2	+	-	-	-
228H	TYR*	5.4	1.0	+	-	-	-

48% "FAVORABLE CONTACTS"
52% DISTURBING CONTACTS

Theoretical maximum (Å2)	691
Actual value (Å2)	400
Normalised complementarity	0.58

Figure 107 (A): control experiment (A): ligand (PPACK) in complex with thrombin (1ABJ.PDB) entry.

Residue		Dist	Surf	Specific contacts HB	Arom	Phob	DC
41H	LEU	5.7	1.0	+	-	-	-
42H	CYS*	5.7	4.4	-	-	-	-
57H	HIS*	2.8	47.3	-	-	+	+
60H	LYS*	3.0	124.4	+	-	-	+
96H	TRP	5.5	2.0	-	-	-	-
97H	GLU	3.2	28.9	-	-	-	-
98H	ASN*	3.5	24.7	-	-	-	-
99H	LEU*	3.4	45.3	-	-	+	-
174H	ILE*	3.6	27.1	-	-	+	-
189H	ASP*	4.1	9.3	+	-	-	-
190H	ALA*	3.2	42.7	+	-	-	+
191H	CYS*	3.6	17.8	+	-	-	+
192H	GLU*	3.0	66.6	+	-	-	+
193H	GLY*	3.9	9.2	+	-	-	-
195H	SER*	3.1	22.6	+	-	-	+
213H	VAL*	4.2	8.1	-	-	-	+
214H	SER	4.1	1.6	-	-	-	+
215H	TRP*	3.2	55.9	+	+	+	+
216H	GLY*	3.0	56.2	+	-	-	+
217H	GLU*	5.0	7.5	-	-	+	+
219H	GLY	2.9	27.3	+	-	-	-
220H	CYS*	3.5	8.5	-	-	-	+
226H	GLY*	4.7	4.6	-	-	-	

48% "FAVORABLE CONTACTS"
52% DISTURBING CONTACTS
dPhe-Pro-dArg-*Gly*-CONH2

Theoretical maximum (Å2)	711
Actual value (Å2)	420
Normalised complementarity	0.59

Figure 107 (B): Structural complementarity for peptide inhibitor D-Phe-Pro-D-Arg-*Gly*-$CONH_2$.

Residue		Dist	Surf	Specific contacts HB	Arom	Phob	DC
57H	HIS*	3.5	17.6	+	-	+	-
60H	TRP*	3.4	109.9	-	-	+	+
96H	TRP	5.2	1.6	-	-	-	-
97H	GLU	3.2	26.7	-	-	-	-
98H	ASN*	3.6	21.5	-	-	-	-
99H	LEU*	3.5	65.7	-	-	+	-
174H	ILE*	3.1	36.1	-	-	+	-
189H	ASP*	4.6	8.1	+	-	-	-
190H	ALA*	3.6	12.9	+	-	-	+
191H	CYS*	2.5	67.0	+	-	-	+
192H	GLU*	2.6	40.8	-	-	+	+
193H	GLY*	2.9	15.8	+	-	-	-
194H	ASP*	3.7	1.8	-	-	-	+
195H	SER*	2.5	33.5	+	-	-	+
213H	VAL*	3.0	26.8	-	-	+	+
214H	SER	2.7	41.0	+	-	-	+
215H	TRP*	3.1	55.0	+	+	+	+
216H	GLY*	2.8	64.8	+	-	-	+
217H	GLU*	4.2	10.9	-	-	+	+
219H	GLY	3.5	11.9	+	-	-	+
220H	CYS*	3.2	16.0	-	-	-	-
226H	GLY*	4.7	4.0	-	-	-	-

46% "FAVORABLE CONTACTS"
54% DISTURBING CONTACTS
dPhe-Pro-dArg-*Ala*-CONH2

Theoretical maximum (Å2)	668
Actual value (Å2)	379
Normalised complementarity	0.57

Figure 107 (C): Structural complementarity for peptide inhibitor **D-Phe-Pro-D-Arg-*Ala*-CONH$_2$.**

Specific contacts

Residue		Dist	Surf	HB	Arom	Phob	DC
17H	VAL*	2.6	24.2	-	-	-	+
57H	HIS*	3.0	31.6	+	-	+	+
60H	TRP*	2.9	128.3	+	+	+	+
96H	TRP	3.4	22.0	-	-	-	-
97H	GLU*	2.8	39.9	-	-	-	+
98H	ASN*	3.9	6.5	-	-	-	-
99H	LEU*	3.0	64.7	-	-	+	+
148H	TRP*	4.8	11.1	+	-	-	-
174H	ILE*	3.5	37.9	-	-	+	+
189H	ASP*	2.2	58.2	-	-	-	+
190H	ALA*	2.0	73.7	+	-	-	+
191H	CYS*	2.6	36.8	+	-	-	+
192H	GLU*	2.6	77.8	+	-	+	+
193H	GLY	4.8	2.7	-	-	-	+
194H	ASP*	4.1	0.8	-	-	-	+
195H	SER*	3.5	13.2	+	-	-	+
213H	VAL*	3.1	24.3	-	-	-	+
214H	SER	2.9	18.7	-	-	-	+
215H	TRP*	3.1	27.9	-	+	+	+
216H	GLY*	2.6	63.5	+	-	-	+
217H	GLU*	3.8	10.3	+	-	+	+
219H	GLY*	2.8	59.1	+	-	-	+
220H	CYS*	2.3	54.3	-	-	-	+
221H	ASP*	2.3	27.5	+	-	-	+
225H	TYR	5.9	0.2	-	-	-	-
226H	GLY*	4.7	7.4	-	-	-	-
228H	TYR*	5.2	0.9	+	-	-	-

22% "FAVORABLE CONTACTS"
78% DISTURBING CONTACTS
D-Phe-Pro-Arg-D-Pro-*D-Arg*-CONH2

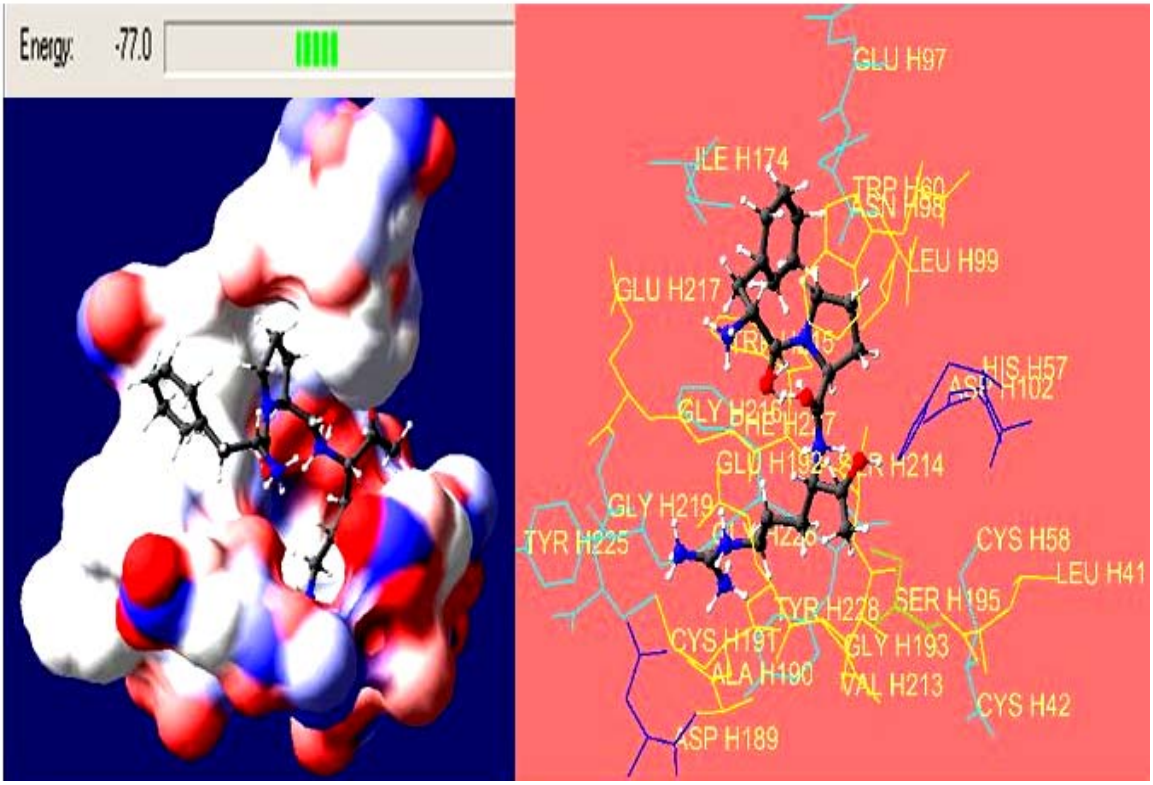
Theoretical maximum (Å2)	927
Actual value (Å2)	328
Normalised complementarity	0.35

107-(D): Structural complementarity for peptide inhibitor D-Phe-Pro-Arg-D-Pro-*D-Arg*-$CONH_2$

In addition to the quantitative analysis of ligand-protein interactions presented in Figure 107**, Figure 108** presents the molecular models containing the predicted favorable and disturbing contacts for some of the peptide-thrombin complexes analyzed through LPC.

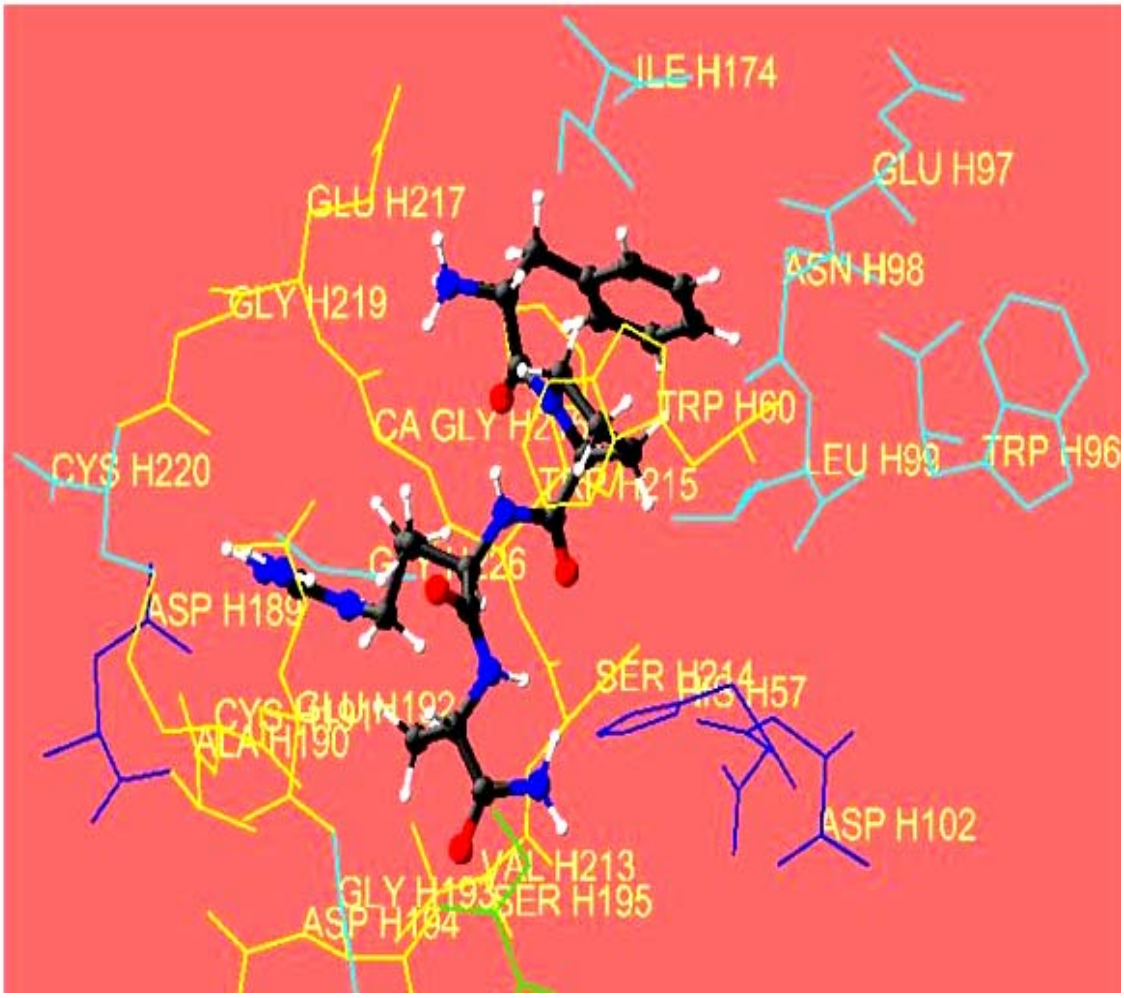
(A) **(B)**

D-Phe-Pro-Arg-chloromethylketone (PPACK) (original template 1ABJ.pdb)



Color code:
Non-favorable contacts: blue-cyan
Favorable contacts: yellow
Catalytic important residues: dark-blue

C) D-Phe-Pro-D-Arg-*Ala*-$CONH_2$
ΔG =- 69.4 [kcal/mol]



D-Phe-Pro-D-Arg-*Gly-CONH*$_2$
ΔG = - 70.5 [kcal/mol]

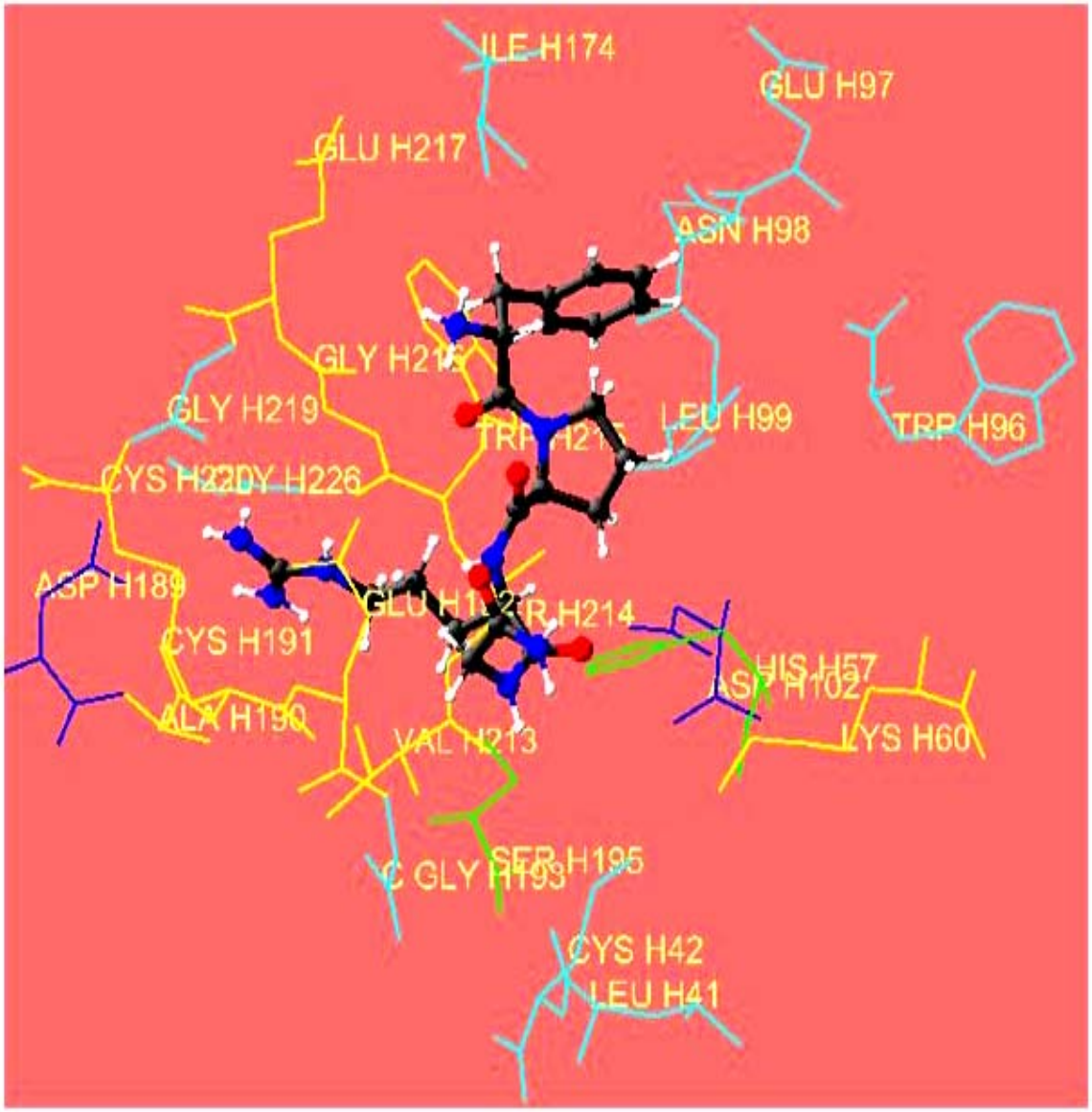


Figure 108: Molecular model for some peptide-thrombin complexes showing details of the active site and of favorable and unfavorable contacts as determined from LPC (ligand-protein-contact) analysis of each ligand-thrombin complex.

As can be seen from the figure 108 (A, B, C and D) the tetrapeptides D-Phe-Pro-D-Arg-Gly-$CONH_2$ and D-Phe-Pro-D-Arg-Ala-$CONH_2$ are predicted to have almost the same normalized complementarity with the control (around 0.58) suggesting that they can

be lead compounds. By contrary the pentapeptide **D-Phe-Pro-Arg-D-Pro-**D-Arg-$CONH_2$ was predicted to have a low normalized complementarity **(0.35<0.57 (control))**, suggesting that this is a less active compound.

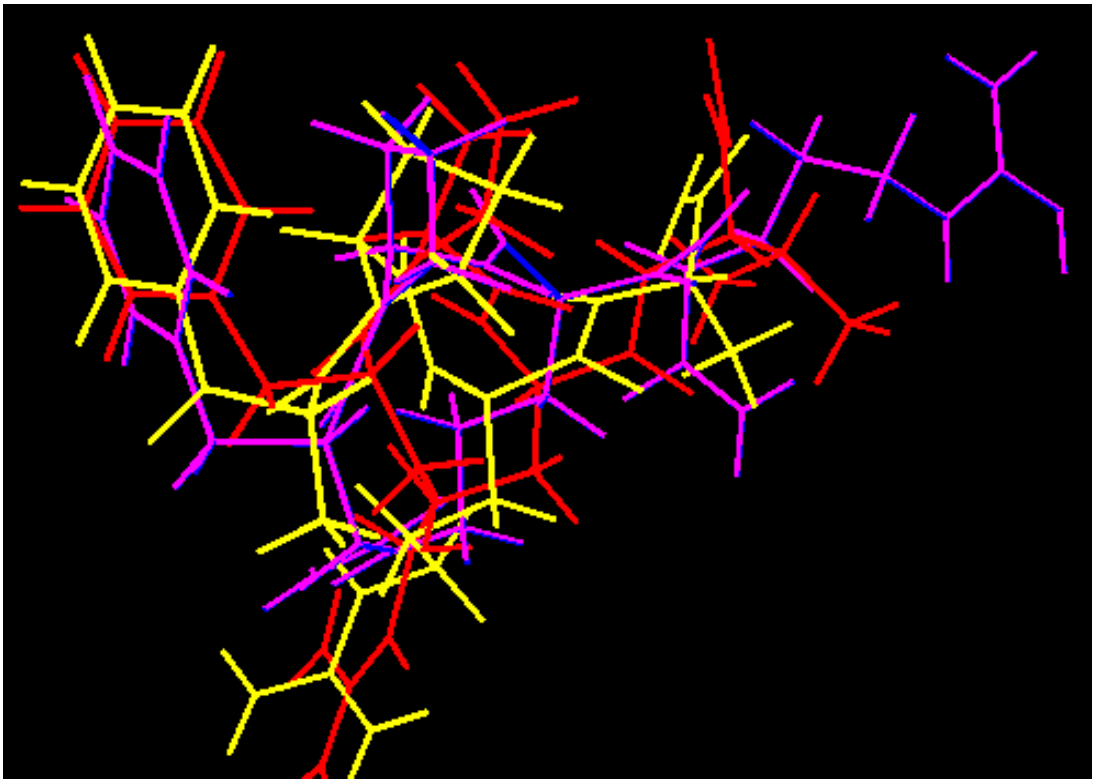
It is important to mention that all the above peptides were predicted to have negative free energy of interaction (favorable) with thrombin. These results suggests that the scoring functions based on structural fitness between ligand and target protein, especially based on normalized complementarity between ligand and the protein target as predicted by LPC analysis are important in assessing the overall potential of a peptide inhibitor.

1.5.3. Analysis of different conformations adopted by peptide inhibitors docked into active site of thrombin.

The thrombin template used in docking experiments was 1ABJ.pdb. The peptides were docked into active site of thrombin using the software "Sculpt" and further minimized using the molecular mechanics (MM) force–field provided by the same software. The template was further deleted and the individual peptides were save as individual .pdb files, preserving their conformation adopted in the active site of thrombin after docking and minimization. The software SPDB Viewer was used to superimpose different peptides differing in one single amino acid position in the sequence space, thus allowing the visualization of different conformations adopted in the active site of thrombin by peptide inhibitors with similar sequence space. A structure-activity relationship (SAR) for tetrapetide, pentapeptide and hexapeptide libraries was performed. In figure 109 (A, B, C and D) tetrapeptides from series D-Phe-Pro-D-Arg-P1'-$CONH_2$ are compared with respect to their conformations adopted in the active site of thrombin (P1'

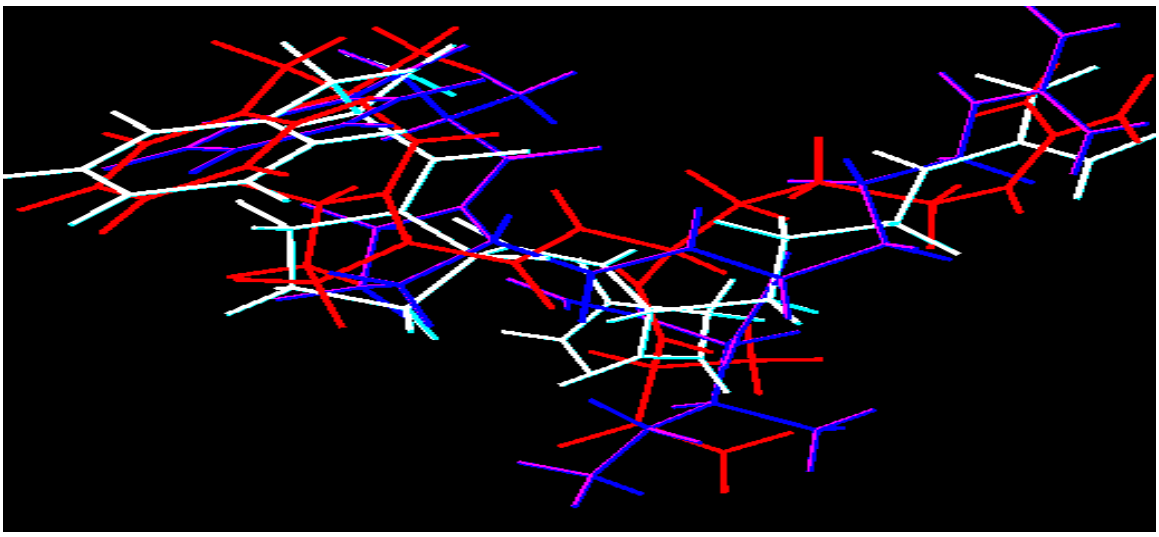
was varied with different amino acid combinations) together with the experimentally determined inhibitory constant (Ki) as described later in Section 3.2 and table 8.

Figure 109 (A)



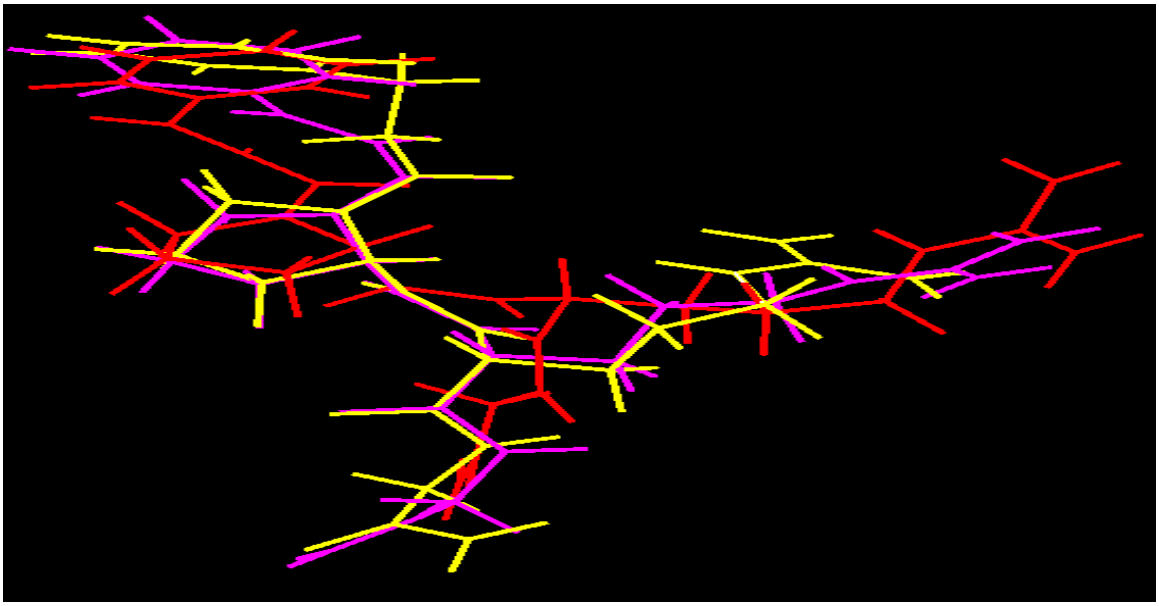
D-Phe-Pro-D-Arg-Val-CONH2------red Ki = 56.32 ± 9.59 μM
D-Phe-Pro-D-Arg-Ala-CONH2-----yellow Ki = 16.64 ± 0.01 μM
D-Phe-Pro-D-Arg-Arg-CONH2-----violet Ki = 91.09 ± 21.99 μM

Figure 109 (B)



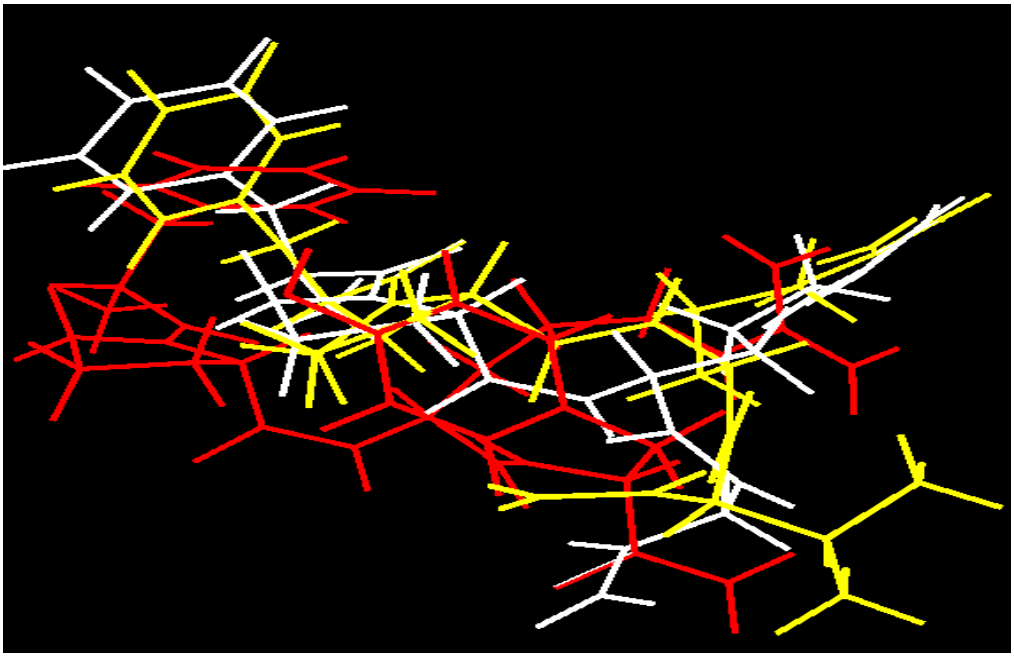
D-Phe-Pro-D-Arg-Ala-CONH2…red **Ki = 16.64 ± 0.01 μM**
D-Phe-Pro-D-Arg-Gly-CONH2…green **Ki = 5.89 ± 0.33 μM**
D-Phe-Pro-D-Arg-Ile-CONH2…blue **Ki = 6.33 ± 0.03 μM**

Figure 109 (C)



***Trans*cinnamic-Pro-D-Arg-Gly-CONH2…blue-violet** **Ki = 6.6 μM**
***Cis*cinnamic-Pro-D-Arg-Gly-CONH2….red** **Ki = 32.5 μM**
***Dihydro*cinnamic-Pro-D-Arg-Gly-CONH2….yellow** **Ki = 102.5 μM**

Figure 109 (D)



DPhe-Pro-DArg-Gly-CONH2… blue Ki = 5.89 ± 0.33 μM
DPhe-Pro-DArg-Val-CONH2… yellow Ki = 56.32 ± 9.59 μM
DPhe-Pro-DArg-Tyr-CONH2…. red Ki = 50.5 ± 5.0 μM

Figure 109 (A, B, C and D): Modeled conformations adopted by different peptides into active site of 1ABJ.pdb together with the experimentally determined inhibitory constant (Ki) as described later in Section 3 and table 13.

1.6. DISCUSSIONS

The analysis of the structural models together with the Ki values obtained experimentally suggest that the tetrapeptides differing in one single amino acid at the P1' position are adopting different conformations into the active site of thrombin, and these different conformations might be related to significant differences in their inhibitory potential, as revealed by their Ki values (Figure109). These findings suggest that P1' position has a significant contribution to the binding energies for the tetrapeptides having the sequence D-Phe-Pro-D-Arg-P1'-$CONH_2$. Also, significant differences between the

Ki_s of tetrapeptides of the series D-Phe-Pro-D-Arg-P1'-$CONH_2$ (varying from 500 fold) (section 3) suggest that the interaction between the amino acid at the P1' position and the S1' pocket in thrombin is very specific. These results are consistent with the X-Ray data, which showed that S1' pocket of thrombin is a small cavity lined by amino acids from the back side of the apolar S2 pocket (His57, Tyr60A and Trp60D) and the side chain of Lys60F (11). The S1' pocket can accommodate small polar amino acid side chains as revealed by P1' amino acids in a number of natural thrombin substrates (P1' in most physiological substrates is occupied by Gly, Ser, Thr, Ile, Leu and Val). The computational model of the active-site docked tetrapeptides from the series D-Phe-Pro-DArg-P1' and their experimentally determined Ki values (see further table 13) support the X-Ray data since the best inhibitors are having L/D amino acids with small side chain (like Gly) or with apolar side-chain (like Cys, Ser, Thr) or more bulky amino acids such as Ile or analogs of phenylalanine like thienylalanine (L-Thi). Independent studies performed by (67) showed that P1' can accommodate bulky hydrophobics even though in natural substrates these residues are found only when the substrate is complexed with another co-factor (68-86). This cofactors are proposed to induce a conformational change in thrombin such that the small pocket P1' will become larger than is in the uncomplexed state (based on X-Ray studies (70-75) and will accommodate bulkier residues such as Ile, Leu, Norleucine (Nle) and Val beside the most usual encountered residues at P1' which are Thr, Ser, Ala or Gly (see also section 1.1). The computation models of trans/cis and dihydrocinnamic derivatives (in P3) support in most cases the prediction that peptides with trans-cinnamic in P3 position better fit to the active site than peptides with cis-cinnamic acid while dihydrocinnamic acid in P3 position adopt a

favorable conformation depending on the amino acid in P1' (better than transcinnamics peptides with the same amino acids in P1') (see Tables 9 and 14).

We have thus proposed and validated with our own experimental data new scoring functions for discovery of new peptides inhibitors for thrombin one of them being based on quantitative analysis of ligand-protein interaction, namely the % normalized complementarity between the ligand peptide and the thrombin target (based on LPC analysis of docked peptides). The comparison between the predicted computation models and the experimental results showed that in some cases (like in the case of peptide D-Phe-Pro-D-Arg-Tyr-$CONH_2$) a good correlation between a weak normalized complementarity and the experimental results (Ki> 50 μM). Also, the peptides from the series D-Phe-Pro-Arg-D-Pro-P1'-CONH2 show a relative good correlation between the experimental data of Ki>100 μM and their low normalized complementarity (<0.4) as compared with control normalized complementarity (>0.5) (Table 12).

Table 12: Correlation between the predicted % favorable contacts, normalized complementarity and the experimental determined inhibitory activity for some peptides (see table 13 for details about Ki).

Peptides	**% favorable contacts**	**normalized complementarity**	**Ki (μM)**
dPhe-Pro-dArg-*Gly*-$CONH_2$	48% FAVORABLE	0.59	5.89 ± 0.33
dPhe-Pro-dArg-*Ala*-$CONH_2$	46% FAVORABLE	0.57	16.64 ± 0.01
dPhe-Pro-Arg-*dPro* -*dArg*-$CONH_2$	22% "FAVORABLE	0.35	980.8
dPhe-Pro-Arg *dPro* -*dPhe*$CONH_2$	*35%* FAVORABLE	0.44	2, 362

Another proposed scoring function was the distance between Asp 189 in thrombin and D-Arg in peptide inhibitors (figure 96). Taken just by itself, the distance between Asp189 (specificity pocket of thrombin) and Arg (P1 position from the peptide) is not sufficient to predict possible lead compounds. We are proposing that all possible interactions between the peptide and the thrombin would be investigated and analyzed in terms of structural fitness. The % normalized complementarity was one of the best scoring function for ranking the reversible peptides inhibitors as a function of their structural complementarity within S3, S2, S1, S1' subpockets of thrombin.

These scoring functions proved to correlate with the experimentally determined dissociation constants for some compounds (see Table 12); however we couldn't find >60% correlation with all the experimental values, suggesting that other factors are contributing to the free energy of interaction between peptides and thrombin such as possible local induced fit at the active site of enzyme upon ligand binding. All docking experiments were performed *in vacuo*, without water of hydration and with the frozen protein target, thus limiting the real environment of the ligand by some artificial geometrical constraints imposed by the active site. In reality, the active site is flexible, can have induced fit as suggested by X-Ray data (68-86) so that the S1' subpocket in thrombin can be bigger than usual and can accommodate bulky hydrophobic amino acids side chains such as Ile, Leu even though in natural substrates this subpocket is fitting small and polar amino acids such as Gly, Ala, Ser, Thr.

Structural evidence shows that the side chain of Lys60F defining the S1' cavity in the active site of thrombin is moved significantly thereby forming a larger groove which can accommodate bulky residues (67). Based on these X-ray data we can further improve the docking experiments by leaving free of constraints 12 Å of the active site of

thrombin making significant contacts with the ligand (i.e. leaving "melted" the active site of thrombin during docking experiments); this experiment would mimics the induced-fit due to ligand binding. Thus the normalized complementarity between ligand and the target protein would improve as a scoring function for a more reliable prediction of the structural fitness between the peptide inhibitors and the target protein (91-93).

Other novel approaches to molecular docking involve the screening of fragments of small molecules, "fragment-based drug discovery" (95). We will use our original designed sequence of peptides to further extend the binding site of the ligand into the active site of thrombin by performing in silico fragment docking of shorter peptide sequences derived from the original D-Phe-Pro-D-Arg-X-CONH2, such as the tripeptide D-Phe-Pro-D-Arg or the dipeptide Pro-D-Arg and extend the fragments linked at N or C-terminus with different amino-acids analogs and further screen toward more powerful drug-like molecules that can fill other pockets within the active site of thrombin like S4 and S2' or S3'. Using this approach of fragment screening and linking of fragments the chances of discovering more potent inhibitors for thrombin will increase and novel drug-like molecules containing both peptide and small organic molecule fragments would score in the low nanomolar inhibitory constant. This approach was used recently by a group of investigators at Astex Therapeutics in UK to discover thrombin inhibitors in the low nM range potency which were selective primarily for binding to S2-S4 pockets in the active site of thrombin, bypassing thus the requirement for having specific interactions at S1 pocket (95).

Another important contribution to future discovery and lead optimization of peptide-derivative reversible inhibitors for thrombin will have to reconsider the guanidinium group in P1 position which is key to inhibitor potency through salt-bridge

formation with Asp189 in the S1 specificity pocket. The compounds with guanidinium or amidine groups in P1 demonstrated good in vivo efficacy, but their use was restricted by poor oral bioavailability. A new class of compounds were discovered in the last years to be highly potent inhibitors of thrombin in the low nanomolar range of their Kis but they are lacking the highly basic group in P1, making them better drug-like molecules, i.e. higher oral bioavailability. These compounds contain hydrophobic groups in P1 position such as chlorophenyl and they are binding through aromatic interactions with Tyr 228 side chain, deeply in the S1 pocket (95). We are further investigating through molecular docking and chemical synthesis and structure-activity relationship a new series of peptide-based inhibitors with D-amino-acids at P1 position which have aromatic and halogenated derivatives of Phe, Tyr, Trp or even His-derivatives (imidazole ring would bring less basicity at P1 position than guanidinium group). These compounds are expected to enhance their interactions at S1 position through hydrophobic or aromatic interactions and are predicted to have a higher oral bioavailability.

REFERENCES

1. Riester, Daniel; Wirsching, Frank; Salinas, Gabriela; Keller, Martina; Gebinoga, Michael; Kamphausen, Stefan; Merkwirth, Christian; Goetz, Ruediger; Wiesenfeldt, Martin; Stuerzebecher, Joerg; Bode, Wolfram; Friedrich, Rainer; Thuerk, Marcel; Schwienhorst, Andreas. Thrombin inhibitors identified by computer-assisted multiparameter design. Proceedings of the National Academy of Sciences of the United States of America (2005), 102(24), 8597-8602.

2. Stahl, Martin; Mauser, Harald; Tsui, Mark; Taylor, Neil R. A Robust Clustering Method for Chemical Structures. Journal of Medicinal Chemistry (2005), 48(13), 4358-4366.

3. Nantermet, Philippe G.; Burgey, Christopher S.; Robinson, Kyle A.; Pellicore, Janetta M.; Newton, Christina L.; Deng, James Z.; Selnick, Harold G.; Lewis, S. Dale; Lucas, Bobby J.; Krueger, Julie A.; Miller-Stein, Cynthia; White, Rebecca B.; Wong, Bradley;

McMasters, Daniel R.; Wallace, Audrey A.; Lynch, Joseph J.; Yan, Youwei; Chen, Zhongguo; Kuo, Lawrence; Gardell, Stephen J.; Shafer, Jules A.; Vacca, Joseph P.; Lyle, Terry A. P2 pyridine N-oxide thrombin inhibitors: a novel peptidomimetic scaffold. Bioorganic & Medicinal Chemistry Letters (2005), 15(11), 2771-2775.

4. Costanzo, Michael J.; Almond, Harold R., Jr.; Hecker, Leonard R.; Schott, Mary R.; Yabut, Stephen C.; Zhang, Han-Cheng; Andrade-Gordon, Patricia; Corcoran, Thomas W.; Giardino, Edward C.; Kauffman, Jack A.; Lewis, Joan M.; de Garavilla, Lawrence; Haertlein, Barbara J.; Maryanoff, Bruce E. In-Depth Study of Tripeptide-Based □-Ketoheterocycles as Inhibitors of Thrombin. Effective Utilization of the S1' Subsite and Its Implications to Structure-Based Drug Design. Journal of Medicinal Chemistry (2005), 48(6), 1984-2008.

5. Steinmetzer, Torsten; Stuerzebecher, Joerg. Progress in the development of synthetic thrombin inhibitors as new orally active anticoagulants. Current Medicinal Chemistry (2004), 11(17), 2297-2321.

6. Stuerzebecher, Joerg; Hauptmann, Joerg; Steinmetzer, Torsten. Thrombin. Proteinase and Peptidase Inhibition (2002), 178-201.

7. Morrissette, Matthew M.; Stauffer, Kenneth J.; Williams, Peter D.; Lyle, Terry A.; Vacca, Joseph P.; Krueger, Julie A.; Lewis, S. Dale; Lucas, Bobby J.; Wong, Bradley K.; White, Rebecca B.; Miller-Stein, Cynthia; Lyle, Elizabeth A.; Wallace, Audrey A.; Leonard, Yvonne M.; Welsh, Denise C.; Lynch, Joseph J.; McMasters, Daniel R. Low molecular weight thrombin inhibitors with excellent potency, metabolic stability, and oral bioavailability. Bioorganic & Medicinal Chemistry Letters (2004), 14(16), 4161-4164.

8. Lu, Tianbao; Markotan, Thomas; Coppo, Frank; Tomczuk, Bruce; Crysler, Carl; Eisennagel, Stephen; Spurlino, John; Gremminger, Lisa; Soll, Richard M.; Giardino, Edward C.; Bone, Roger. Oxyguanidines. Part 2: discovery of a novel orally active thrombin inhibitor through structure-based drug design and parallel synthesis. Bioorganic & Medicinal Chemistry Letters (2004), 14(14), 3727-3731.

9. Peterlin-Masic, Lucija; Kranjc, Andreja; Marinko, Petra; Mlinsek, Gregor; Solmajer, Tomaz; Stegnar, Mojca; Kikelj, Danijel. Selective 3-Amino-2-pyridinone acetamide thrombin inhibitors incorporating weakly basic partially saturated heterobicyclic P1-Arginine mimetics. Bioorganic & Medicinal Chemistry Letters (2003), 13(19), 3171-3176.

10. Nantermet, Philippe G.; Selnick, Harold G. Potent thrombin inhibitors via a 20-membered ring olefin metathesis macrocyclization. Tetrahedron Letters (2003), 44(11), 2401-2404.

11. De Filippis, Vincenzo; Colombo, Giorgio; Russo, Ilaria; Spadari, Barbara; Fontana, Angelo. Probing the Hirudin-Thrombin Interaction by Incorporation of Noncoded

Amino Acids and Molecular Dynamics Simulation. Biochemistry (2002), 41(46), 13556-13569.

12. Reiner, John E.; Siev, Daniel V.; Araldi, Gian-Luca; Cui, Jingrong Jean; Ho, Jonathan Z.; Reddy, Komandla Malla; Mamedova, Lala; Vu, Phong H.; Lee, Kuen-Shan S.; Minami, Nathaniel K.; Gibson, Tony S.; Anderson, Susanne M.; Bradbury, Annette E.; Nolan, Thomas G.; Semple, J. Edward. Non-covalent thrombin inhibitors featuring P3-heterocycles with P1-monocyclic arginine surrogates. Bioorganic & Medicinal Chemistry Letters (2002), 12(8), 1203-1208.

13. St-Denis, Yves; Levesque, Sophie; Bachand, Benoit; Edmunds, Jeremy J.; Leblond, Lorraine; Preville, Patrice; Tarazi, Micheline; Winocour, Peter D.; Siddiqui, M. Arshad. Novel bicyclic lactam inhibitors of thrombin: highly potent and selective inhibitors. Bioorganic & Medicinal Chemistry Letters (2002), 12(8), 1181-1184.

14. Bode, Wolfram. X-ray crystal structures of thrombin in complex with D-Phe-Pro-Arg and with small benzamidine- and arginine-based "non-peptidic" inhibitors. Advances in Experimental Medicine and Biology (1993), 340.

15. Dahlgren, Anders; Johansson, Per-Ola; Kvarnstrom, Ingemar; Musil, Djordje; Nilsson, Ingemar; Samuelsson, Bertil. Novel Morpholinone-Based D-Phe-Pro-Arg Mimics as Potential Thrombin Inhibitors: Design, Synthesis, and X-ray Crystal Structure of an Enzyme Inhibitor Complex. Bioorganic & Medicinal Chemistry (2002),

16. Kokko, K. P.; Arrigoni, C. E.; Dix, T. A. Selectivity enhancement induced by substitution of non-natural analogues of arginine and lysine in arginine-based thrombin inhibitors. Bioorganic & Medicinal Chemistry Letters (2001), 11(14), 1947-1950.

17. Greenidge, P. A.; Weiser, J. A comparison of methods for pharmacophore generation with the catalyst software and their use for 3D-QSAR: application to a set of 4-aminopyridine thrombin inhibitors. Mini-Reviews in Medicinal Chemistry (2001), 1(1), 79-87.

18. Mlinsek, G.; Novic, M.; Hodoscek, M.; Solmajer, T. Prediction of Enzyme Binding: Human Thrombin Inhibition Study by Quantum Chemical and Artificial Intelligence Methods Based on X-ray Structures. Journal of Chemical Information and Computer Sciences (2001), 41(5), 1286-1294.

19. Fox, Thomas; Haaksma, Eric E. J. Computer based screening of compound databases: 1. Preselection of benzamidine-based thrombin inhibitors. Journal of Computer-Aided Molecular Design (2000), 14(5), 411-425.

20. Clare, Brian W.; Scozzafava, Andrea; Briganti, Fabrizio; Iorga, Bogdan; Supuran, Claudiu T. Protease inhibitors. Part 2. Weakly basic thrombin inhibitors incorporating sulfonyl-aminoguanidine moieties as S1 anchoring groups: synthesis and structure-activity correlations. Journal of Enzyme Inhibition (2000), 15(3), 235-264.

21. Kettner, Charles; Knabb, Robert M. Peptide boronic acid inhibitors of thrombin. Advances in Experimental Medicine and Biology (1993), 340 (Design of Synthetic Inhibitors of Thrombin), 109-18.

22. Elgendy, Said; Deadman, John; Claeson, Goran. New peptide boronic acid inhibitors of thrombin. Advances in Experimental Medicine and Biology (1993), 340(Design of Synthetic Inhibitors of Thrombin), 173-8.

23. Wienand, Anette; Ehrhardt, Claus; Metternich, Rainer; Tapparelli, Carlo. Design, synthesis and biological evaluation of selective boron-containing thrombin inhibitors. Bioorganic & Medicinal Chemistry (1999), 7(7), 1295-1307.

24. Mitchell T J; Knabb R M; Christ D D; Farmer A R; Reilly T M Analysis of the thrombin inhibitor DuP 714 by an enzyme-linked immunosorbent assay. Blood coagulation & fibrinolysis : an international journal in haemostasis and thrombosis (1994 Aug), 5(4), 517-21.

25. Kaiser B; Callas D; Hoppensteadt D; Mallinowska K; Fareed J Comparative studies on the inhibitory spectrum of recombinant hirudin, DuP 714 and heparin on thrombin and factor Xa generation in biochemically defined systems. Thrombosis research (1994 Mar 1), 73(5), 327-35.

26. Badimon J J; Weng D; Chesebro J H; Fuster V; Badimon L Platelet deposition induced by severely damaged vessel wall is inhibited by a boroarginine synthetic peptide with antithrombin activity. Thrombosis and haemostasis (1994 Apr), 71(4), 511-6.

27. Su, Zhengding; Vinogradova, Anna; Koutychenko, Anatol; Tolkatchev, Dmitri; Ni, Feng. Rational design and selection of bivalent peptide ligands of thrombin incorporating P4-P1 tetrapeptide sequences: from good substrates to potent inhibitors. Protein Engineering, Design & Selection (2004), 17(8), 647-657.

28. Szewczuk, Zbigniew; Gibbs, Bernard F.; Yue, Shi Yi; Purisima, Enrico O.; Konishi, Yasuo. Conformationally restricted thrombin inhibitors resistant to proteolytic digestion. Biochemistry (1992), 31(38), 9132-40.

29. Maraganore, John M.; Fenton, John W., II. Thrombin inhibition by synthetic hirudin peptides. Advances in Experimental Medicine and Biology (1990), 281(Fibrinogen, Thromb., Coagulation, Fibrinolysis), 177-83.

30. Schaerer, Kaspar; Morgenthaler, Martin; Seiler, Paul; Diederich, Francois; Banner, David W.; Tschopp, Thomas; Obst-Sander, Ulrike. Enantiomerically pure thrombin inhibitors for exploring the molecular-recognition features of the oxyanion hole. Helvetica Chimica Acta (2004), 87(10), 2517-2538.

31. Raffler, Nikolai A.; Schneider-Mergener, Jens; Famulok, Michael. A novel class of small functional peptides that bind and inhibit human α-thrombin isolated by mRNA display. Chemistry & Biology (2003), 10(4), 369.

32. Rajagopal, Srivats; Meza-Romero, Roberto; Ghosh, Indraneel. Dual surface selection methodology for the identification of thrombin binding epitopes from hotspot biased phage-display libraries. Bioorganic & Medicinal Chemistry Letters (2004), 14(6), 1389-1393.

32. Campos, Ivan T. N.; Silva, Melissa M.; Azzolini, Simone S.; Souza, Adriana F.; Sampaio, Claudio A. M.; Fritz, Hans; Tanaka, Aparecida S. Evaluation of phage display system and leech-derived tryptase inhibitor as a tool for understanding the serine proteinase specificities. Archives of Biochemistry and Biophysics (2004), 425(1), 87-94.

33. Stoop, A. Allart; Craik, Charles S. Engineering of a macromolecular scaffold to develop specific protease inhibitors. Nature Biotechnology (2003), 21(9), 1063-1068.

34. Meiring, Muriel S.; Litthauer, Derek; Harsfalvi, Jolan; van Wyk, Veronica; Badenhorst, Philip N.; Kotze, Harry F. In vitro effect of a thrombin inhibition peptide selected by phage display technology. Thrombosis Research (2002), 107(6), 365-371.

35. Tanaka, Aparecida S.; Silva, Melissa M.; Torquato, Ricardo J. S.; Noguti, Maria A. E.; Sampaio, Claudio A. M.; Fritz, Hans; Auerswald, Ennes A. Functional phage display of leech-derived tryptase inhibitor (LDTI): construction of a library and selection of thrombin inhibitors. FEBS Letters (1999), 458(1), 11-16.

36. Wirsching, Frank; Opitz, Thomas; Dietrich, Rudiger; Schwienhorst, Andreas. Display of functional thrombin inhibitor hirudin on the surface of phage M13. Gene (1997), 204(1/2), 177-184.

37. Markland, William; Roberts, Bruce L.; Ladner, Robert C. Selection for protease inhibitors using bacteriophage display. Methods in Enzymology (1996), 267(Combinatorial Chemistry), 28-51.

38. Markland, William; Ley, Arthur Charles; Ladner, Robert Charles. Iterative Optimization of High-Affinity Protease Inhibitors Using Phage Display. 2. Plasma Kallikrein and Thrombin. Biochemistry (1996), 35(24), 8058-8067.

39. van Meijer, Marja; Roelofs, Yvonne; Neels, Jaap; Horrevoets, Anton J. G.; van Zonneveld, Anton-Jan; Pannekoek, Hans. Selective screening of a large phage display library of plasminogen activator inhibitor 1 mutants to localize interaction sites with either thrombin or the variable region 1 of tissue-type plasminogen activator. Journal of Biological Chemistry (1996), 271(13), 7423-8.

40. Rajagopal Srivats; Meza-Romero Roberto; Ghosh Indraneel Dual surface selection methodology for the identification of thrombin binding epitopes from hotspot biased phage-display libraries. Bioorganic & medicinal chemistry letters (2004 Mar 22), 14(6), 1389-93.

41. Malikayil J A; Burkhart J P; Schreuder H A; Broersma R J Jr; Tardif C; Kutcher L W 3rd; Mehdi S; Schatzman G L; Neises B; Peet N P Molecular design and characterization of an alpha-thrombin inhibitor containing a novel P1 moiety. Biochemistry (1997 Feb 4), 36(5), 1034-40.

42. Obst U; Banner D W; Weber L; Diederich F Molecular recognition at the thrombin active site: structure-based design and synthesis of potent and selective thrombin inhibitors and the X-ray crystal structures of two thrombin-inhibitor complexes. Chemistry & biology (1997 Apr), 4(4), 287-95.

43. Morenweiser R; Auerswald E A; van de Locht A; Fritz H; Sturzebecher J; Stubbs M T Structure-based design of a potent chimeric thrombin inhibitor. Journal of biological chemistry (1997 Aug 8), 272(32), 19938-42.

44. Brady S F; Stauffer K J; Lumma W C; Smith G M; Ramjit H G; Lewis S D; Lucas B J; Gardell S J; Lyle E A; Appleby S D; Cook J J; Holahan M A; Stranieri M T; Lynch J J Jr; Lin J H; Chen I W; Vastag K; Naylor-Olsen A M; Vacca J P Discovery and development of the novel potent orally active thrombin inhibitor N-(9-hydroxy-9-fluorenecarboxy)prolyl trans-4-aminocyclohexylmethyl amide (L-372,460): coapplication of structure-based design and rapid multiple analogue synthesis on solid support. Journal of medicinal chemistry (1998 Jan 29), 41(3), 401-6.

45. Li M; Lin Z; Johnson M E Structure-based design and synthesis of novel thrombin inhibitors based on phosphinic peptide mimetics. Bioorganic & medicinal chemistry letters (1999 Jul 19), 9(14), 1957-62.

46. Dahlgren Anders; Johansson Per Ola; Kvarnstrom Ingemar; Musil Djordje; Nilsson Ingemar; Samuelsson Bertil Novel morpholinone-based D-Phe-Pro-Arg mimics as potential thrombin inhibitors: design, synthesis, and X-ray crystal structure of an enzyme inhibitor complex. Bioorganic & medicinal chemistry, (2002 Jun), 10(6), 1829-39.

47. Pfau Roland. Structure-based design of thrombin inhibitors. Current opinion in drug discovery & development (2003 Jul), 6(4), 437-50.

48. Lu Tianbao; Markotan Thomas; Coppo Frank; Tomczuk Bruce; Crysler Carl; Eisennagel Stephen; Spurlino John; Gremminger Lisa; Soll Richard M; Giardino Edward C; Bone Roger Oxyguanidines. Part 2: Discovery of a novel orally active thrombin inhibitor through structure-based drug design and parallel synthesis. Bioorganic & medicinal chemistry letters (2004 Jul 16), 14(14), 3727-31.

49. Hanessian, Stephen; Balaux, Elise; Musil, Djorde; Olsson, Lise-Lotte; Nilsson, Ingemar. Exploring the chiral space within the active site of α-thrombin with a constrained mimic of D-Phe-Pro-Arg - design, synthesis, inhibitory activity, and x-ray structure of an enzyme-inhibitor complex. Bioorganic & Medicinal Chemistry Letters (2000), 10(3), 243-247.

50. Danilewicz, John C.; Abel, Stuart M.; Brown, Alan D.; Fish, Paul V.; Hawkeswood, Edward; Holland, Stephen J.; James, Keith; McElroy, Andrew B.; Overington, John; Powling, Michael J.; Rance, David J. Design of Selective Thrombin Inhibitors Based on the (R)-Phe-Pro-Arg Sequence. Journal of Medicinal Chemistry (2002), 45(12), 2432-2453.

51. Isaacs, Richard C. A.; Newton, Christina L.; Solinsky, Mark G.; Naylor-Olsen, Adel M. L-376,062. A potent, selective, noncovalent thrombin inhibitor bearing a novel imidazole P1 ligand. Book of Abstracts, 217th ACS National Meeting, Anaheim, Calif., March 21-25 (1999), MEDI-005.

52. Semple, J. Edward; Rowley, David C.; Owens, Timothy D.; Minamni, Nathaniel K.; Uong, Theresa H.; Brunck, Terence K. Potent and selective thrombin inhibitors featuring hydrophobic, basic P3-P4-aminoalkyllactam moieties. Bioorganic & Medicinal Chemistry Letters (1998), 8(24), 3525-3530.

53. Maryanoff, Bruce E.; Hecker, L. R.; Schott, M. R.; Yabut, S. C.; Zhang, H. -C.; Andrade-Gordon, P.; Giardino, E. C.; Kauffman, J. A.; Lewis, J. M.; Costanzo, Michael J. In-depth study of tripeptide-based acylheterocycles as inhibitors of thrombin. Effective utilization of the S1' subsite and its implications to protein structure-based drug design. Book of Abstracts, 216th ACS National Meeting, Boston, August 23-27 (1998), MEDI-021.

54. Grueneberg, Sven. A QSAR model derived from a homology model: A strategy to include structural information in ligand-based design. QSAR & Combinatorial Science (2005), 24(4), 517-526.

55. Kekenes-Huskey, Peter M.; Muegge, Ingo; von Rauch, Moriz; Gust, Ronald; Knapp, Ernst-Walter. A molecular docking study of estrogenically active compounds with 1,2-diarylethane and 1,2-diarylethene pharmacophores. Bioorganic & Medicinal Chemistry (2004), 12(24), 6527-6537.

56. Kosinsky, Yuri A.; Volynsky, Pavel E.; Lagant, Philippe; Vergoten, Gerard; Suzuki, Ei-Ichiro; Arseniev, Alexander S.; Efremov, Roman G. Development of the force field parameters for phosphoimidazole and phosphohistidine. Journal of Computational Chemistry (2004), 25(11), 1313-1321.

57. Wu, Guosheng; Vieth, Michal. SDOCKER: A Method Utilizing Existing X-ray Structures To Improve Docking Accuracy. Journal of Medicinal Chemistry (2004), 47(12), 3142-3148.

58. Cho, Art E.; Guallar, Victor; Berne, Bruce J.; Friesner, Richard A. Importance of electric charges in molecular docking: QM/MM approach. Abstracts of Papers, 228th ACS National Meeting, Philadelphia, PA, United States, August 22-26, 2004 (2004), COMP-239.

59. Liu, Hao-Yang; Kuntz, Irwin D.; Zou, Xiaoqin. Pairwise GB/SA Scoring Function for Structure-based Drug Design. Journal of Physical Chemistry B (2004), 108(17), 5453-5462.

60. Friesner, Richard A.; Banks, Jay L.; Murphy, Robert B.; Halgren, Thomas A.; Klicic, Jasna J.; Mainz, Daniel T.; Repasky, Matthew P.; Knoll, Eric H.; Shelley, Mee; Perry, Jason K.; Shaw, David E.; Francis, Perry; Shenkin, Peter S. Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. Journal of Medicinal Chemistry (2004), 47(7), 1739-1749.

61. Wu, Guosheng; Robertson, Daniel H.; Brooks, Charles L., III; Vieth, Michal. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm. Journal of Computational Chemistry (2003), 24(13), 1549-1562.

62. Fernandez-Recio, Juan; Totrov, Maxim; Abagyan, Ruben. ICM-DISCO docking by global energy optimization with fully flexible side-chains. Proteins: Structure, Function, and Genetics (2003), 52(1), 113-117.

63. Wang Renxiao; Lai Luhua; Wang Shaomeng Further development and validation of empirical scoring functions for structure-based binding affinity prediction. Journal of computer-aided molecular design, (2002 Jan), 16(1), 11-26.

64. Nagata, Hiroshi; Mizushima, Hiroshi; Tanaka, Hiroshi. Concept and prototype of protein-ligand docking simulator with force feedback technology. Bioinformatics (2002), 18(1), 140-146.

65. Mueller, Luciano; Langley, David R.; Cheney, Daniel L. Lead docking protocol featuring improved sampling and scoring. Abstracts of Papers, 228th ACS National Meeting, Philadelphia, PA, United States, August 22-26, 2004 (2004), COMP-245.

66. MDL: http://www.mdli.com/products/sculpt.html

67. Slon-Usakiewicz, Jacek J.; Sivaraman, J.; Li, Yunge; Cygler, Miroslaw; Konishi, Yasuo. Design of P1' and P3' Residues of Trivalent Thrombin Inhibitors and Their Crystal Structures. Biochemistry (2000), 39(9), 2384-2391.

68. Noeteberg, Daniel; Brnalt, Jonas; Kvarnstroem, Ingemar; Linschoten, Marcel; Musil, Djordje; Nystroem, Jan-Erik; Zuccarello, Guido; Samuelsson, Bertil. New proline mimetics: Synthesis of thrombin inhibitors incorporating cyclopentane- and cyclopentenedicarboxylic acid templates in the P(2) position. Binding conformation investigated by x-ray crystallography. Journal of Medicinal Chemistry (2000), 43(9), 1705-1713.

69. Stubbs M T; Oschkinat H; Mayr I; Huber R; Angliker H; Stone S R; Bode W The interaction of thrombin with fibrinogen. A structural basis for its specificity. European journal of biochemistry / FEBS (1992 May 15), 206(1), 187-95.

70. Turk D; Sturzebecher J; Bode W Geometry of binding of the N alpha-tosylated piperidides of m-amidino-, p-amidino- and p-guanidino phenylalanine to thrombin and trypsin. X-ray crystal structures of their trypsin complexes and modeling of their thrombin complexes. FEBS Letters (1991 Aug 5), 287(1-2), 133-8.

71. Sturzebecher J; Vieweg H; Wikstrom P; Turk D; Bode W Interactions of thrombin with benzamidine-based inhibitors. Biological chemistry Hoppe-Seyler (1992 Jul), 373(7), 491-6.

72. Brandstetter H; Turk D; Hoeffken H W; Grosse D; Sturzebecher J; Martin P D; Edwards B F; Bode W Refined 2.3 A X-ray crystal structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA. A starting point for improving antithrombotics. Journal of molecular biology (1992 Aug 20), 226(4), 1085-99.

73. Qiu X; Padmanabhan K P; Carperos V E; Tulinsky A; Kline T; Maraganore J M; Fenton J W 2nd Structure of the hirulog 3-thrombin complex and nature of the S' subsites of substrates and inhibitors. Biochemistry (1992 Dec 1), 31(47), 11689-97.

74. Bode W; Turk D; Karshikov A The refined 1.9-A X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. Protein science : a publication of the Protein Society (1992 Apr), 1(4), 426-71.

75. Platt E. The comparison of an interim tertiary predicted model of bovine thrombin and the x-ray structure of human thrombin. Advances in experimental medicine and biology (1993), 340 79-81.

76. Arni R K; Padmanabhan K; Padmanabhan K P; Wu T P; Tulinsky A Structure of the non-covalent complex of prothrombin kringle 2 with PPACK-thrombin. Chemistry and physics of lipids (1994 Jan), 67-68 59-66.

77. Stubbs M T; Bode W The clot thickens: clues provided by thrombin structure. Trends in biochemical sciences (1995 Jan), 20(1), 23-8.

78. Vijayalakshmi J; Padmanabhan K P; Mann K G; Tulinsky A The isomorphous structures of prethrombin2, hirugen-, and PPACK-thrombin: changes accompanying activation and exosite binding to thrombin. Protein Science : a publication of the Protein Society (1994 Dec), 3(12), 2254-71.

79. Bergner A; Bauer M; Brandstetter H; Sturzebecher J; Bode W. The X-ray crystal structure of thrombin in complex with N alpha-2-naphthylsulfonyl-L-3-amidino-phenylalanyl-4-methylpiperidide: the beneficial effect of filling out an empty cavity. Journal of Enzyme Inhibition (1995), 9(1), 101-10.

80. Bertrand J A; Oleksyszyn J; Kam C M; Boduszek B; Presnell S; Plaskon R R; Suddath F L; Powers J C; Williams L D Inhibition of trypsin and thrombin by amino(4-amidinophenyl)methanephosphonate diphenyl ester derivatives: X-ray structures and molecular models. Biochemistry (1996 Mar 12), 35(10), 3147-55.

81. Nardini M; Pesce A; Rizzi M; Casale E; Ferraccioli R; Balliano G; Milla P; Ascenzi P; Bolognesi M Human alpha-thrombin inhibition by the active site titrant N alpha-(N,N-dimethylcarbamoyl)-alpha-azalysine p-nitrophenyl ester: a comparative kinetic and X-ray crystallographic study. Journal of molecular biology (1996 May 24), 258(5), 851-9.

82. Nienaber V L; Mersinger L J; Kettner C A Structure-based understanding of ligand affinity using human thrombin as a model system. Biochemistry (1996 Jul 30), 35(30), 9690-9.

83. Rezaie A R; Olson S. T. Contribution of lysine 60f to S1' specificity of thrombin. Biochemistry (1997 Feb 4), 36(5), 1026-33.

84. Burkhard P; Taylor P; Walkinshaw M.D. An example of a protein ligand found by database mining: description of the docking method and its verification by a 2.3 A X-ray structure of a thrombin-ligand complex. Journal of molecular biology (1998 Mar 27), 277(2), 449-66.

85. Mathews, I. I.; Tulinsky, A. Active-site mimetic inhibition of thrombin. Acta Crystallographica, Section D: Biological Crystallography (1995), D51(4), 550-9.

86. Strickland, Corey L.; Fevig, John M.; Galemmo, Robert A., Jr.; Wells, Brian L.; Kettner, Charles A.; Weber, Patricia C. Biochemical and crystallographic characterization of homologous non-peptidic thrombin inhibitors having alternate binding modes. Acta Crystallographica, Section D: Biological Crystallography (1998), D54(6, Pt. 2), 1207-1215.

87. Engh & Huber. Who checks the checkers? Four validation tools applied to eight atomic resolution structures. EU 3-D Validation Network. (1998), J Mol Biol. 276(2):417-36.

88. RasMol, created by Roger Sayle: http://www.umass.edu/microbio/rasmol/index2.htm

89. Sobolev, Vladimir; Wade, Rebecca C.; Vriend, Gert; Edelman, Marvin. Molecular docking using surface complementarity. Proteins: Structure, Function, and Genetics (1996), 25(1), 120-129.

90. Sobolev, Vladimir; Edelman, Meir. Computational tools for analysis of structural data in the PDB. Folding & Design (1996), 1(Suppl.), S59.

91. Eyal, Eran; Najmanovich, Rafael; Mcconkey, Brendan J.; Edelman, Marvin; Sobolev, Vladimir. Importance of solvent accessibility and contact surfaces in modeling

side-chain conformations in proteins. Journal of Computational Chemistry (2004), 25(5), 712-724.

92. McConkey, Brendan J.; Sobolev, Vladimir; Edelman, Marvin. The performance of current methods in ligand-protein docking. Current Science (2002), 83(7), 845-856.

93. Sobolev, Vladimir; Sorokine, Anatoli; Prilusky, Jaime; Abola, Enrique E.; Edelman, Marvin. Automated analysis of interatomic contacts in proteins. Bioinformatics. (1999), 15(4), 327-332.

94. SWISS PDB Viewer, GlaxoSmithKline: http://www.expasy.ch/spdbv/.

95. Howard N., Abell C., Blakemore W., Chessari G., Congreve M., Howard S., Jhoti H., Murray C.W., Seavers L.C.A., Montfort R.L.M. Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. J.Med.Chem. 2006, 49, 1346-1355.

Chapter 2

Solid Phase Peptide Synthesis of individual and of peptide libraries

2.1. INTRODUCTION

Large combinatorial libraries of random peptides have een used for a variety of applications that include analysis of protein-protein interactions, epitope mapping, and drug targeting. The major obstacle in screening these libraries is the loss of specific but low affinity binding peptides during washing steps. Loss of these specific binders often results in isolation of peptides that bind nonspecifically to components used in the selection process (1-10). To avoid these problems, we constructed focused libraries of peptides derived from the original sequences: D-Phe(P3)-Pro(P2)-D-Arg(P1)-X(P1')-CONH2 and D-Phe(P3)-Pro(P2)-Arg(P1)-D-Pro(P1')-Y(P2')-CONH2 using a modified 'one-bead-one-compound' approach. The chemical synthesis of 10 peptides at a time for each variation in X or Y within the sequence spaces presented above was achieved by

coupling in parallel at X and Y positions followed by “pool and split” synthetic method. The “split” step was performed at P3 position were further variations were performed with unnatural amino acids, analogs of Phe. Steric and aggregation effects may appear due to sequences that are hard to synthesize, thus optimizing the coupling steps (10-20) become a very important achievement during synthesis of peptide libraries. Owing to the robustness of amide bond chemistry, the ability to explore extensive chemical diversity by incorporation of unnatural and natural amino acids, and the ability to explore conformational diversity, through the incorporation of various constraints, arrays of conformationally constrained peptides were synthesized to gain mainly hairpin and beta-turn structures (10-18). We took advantage of the well defined sequence space containing D-amino acids at specific positions (such as D-Arg at i+3 position and D-Pro at i+4 position) and synthesized individual peptides or peptide libraries after which the individual components were tested for their secondary structure elements using circular dichroism (CD) (Chapter 5). The presence of beta turns was confirmed by circular dichroism investigations for most peptides tested in which D-Arg occupied the i+3 position within the sequence space D-Phe(P3)-Pro(P2)-D-Arg(P1)-X(P1’)-CONH2. This chapter presents the solid phase synthesis with modified F-moc chemistry approach of C-teminal amidated peptides which were further tested for their inhibitory potential against thrombin and were found to inhibit thrombin at 0.9- 1000 μM (Chapter 3).

2.2. MATERIALS AND METHODS

2.2.1. Solid phase polypeptide synthesis (SPPS), based on Fmoc chemistry synthesis of individual designed peptides.

The individual peptides both with natural and unnatural amino-acids analogs were synthesized using the 432A Synergy Personal Peptide Synthesizer from Applied Biosystems. The F-moc protected amino acids (both L and D) were purchased from NovaBiochem as powder. Each amino acid was packed in a separate cadrige provided by Applied Biosystems, such that the molar ratio between the amino acid and the resin with free NH_2 groups (10 μmole sites) was 4:1. Each synthesis yielded approximately 25 μmole yield for each tetra, penta or hexapeptide. The AM-NH2-resins from ***Applied Biosystems*** or Rink-Amide-resins from Novabiochem were used to generate amidated peptide. The substitution level varied from 0.46-0.8 mmoles F-moc for one g of amidated resin. Fmoc chemistry involved the following steps (1):

(1) Deprotection of the Fmoc (base labile) with piperidine-20 minutes (2) activation of carboxyl group using HBTU 8 mmol (2-(1 H-benzotriazol-1-yl)-1,1,3,3-tetrameethyluronium hexafluorophosphate)/HOBT; (3) coupling of the amino acids using DIEA (0.4 M N,N-dilsopropylethylamine) solved in N-methypyrrolidone (NMP)or Dimethylformamide (DMF); tetrahydrofuran was used as a drying agent between couplings; DMF was used as solvent throughouth the whole synthesis. The peptide was cleaved from the resin and the side-chains were deprotected after the last F-moc was removed with 20% piperidine using 50 μL of ethanedithiol (EDT), 50 μL of thioanisole (as scavengers), and 900 μL TFA (trifluoroacetic acid) (for each 0.2 mg of resin). Precipitation of the polypetide was done with 15 mL of methyl tert- butyl ether (MTBE).

All the scavengers and TFA were filtered out using a glass filter (medium size pores) and the peptides were washed four times with 10 mL of MTBE, for removal of organic reagents. Solubilization of peptides was performed with 20 mL of de-ionized water and 10 mL of acetonitrile. These steps in F-moc solid peptide synthesis were performed also for the libraries of peptides which were synthesized using one-bead-one peptide chain method. For libraries of peptides with variable P1' position within the sequence space D-Phe-Pro-Arg-D-Pro-P1'-CONH2, 5 or 10-parallel coupling of each amino acid in P1' were performed in separated vials, after which the mixing of all vials was performed in one single beaker and the synthesis further proceeding with the coupling of the common amino-acids. The coupling was performed always with 4 fold excess of the F-moc protected amino acid with respected to the free deprotected NH2 group on the resin support. The yield of individual synthesized peptides was further determined by high-pressure liquid chromatography (HPLC) and electrospray ionization mass spectroscopy (ESI-MS).

2.2. Reversed Phase (RP) High Pressure Liquid Chromatography (HPLC) of individual peptides and of peptide libraries.

RP-HPLC was used to determine the purity of peptides using a Waters 2695 HPLC system with built in Photo Diode Array (PDA) system. The buffer system contained**:** ***buffer A*****:** 0.1% TFA in de-ionized water (aqueous phase); ***buffer B:*** 100% CH_3-CN (acetonitrile) (organic phase). The following gradient was used for peptide separation in the reversed phase mode: 0% - 55% buffer B (acetonitrile) in 40 minutes, 55% - 0% acetonitrile in 15 min with a total run time of 55 min, at a flow rate of 1 ml/min. The reversed-phase columns C_{18}, 100 Å and $C_{8,}$ 300 Å, both from Varian, Microsorb Rainin

were commonly used during the analytical control of peptide purity. Some peptides were further purified using semipreparative columns, C18, 300 Å from Varian-Rainin. The elution of the peptides was monitored using the PDA of the HPLC in the wavelength range of 205 nm – 320 nm.

The concentration of each peptide in water was determined from UV absorption measurements. The molar extinction coefficient for each residue was considered to be around 300 at 230 nm (an average value including side-chain contributions and spillover from the peptide **Π→Π* transition**) (10). In addition the concentration of pure peptides was determined from weighing known the molecular weight for each peptide as determined from the amino acid composition.

2.3. Electrospray Ionization-(+) mode, Mass Spectroscopy analysis of peptides.

The theoretical molecular weight of peptides was calculated from the amino acid sequence and the LCQ Finnigan mass spectrometer was used to confirm it. The mass scans were acquired in electrospray positive mode (ESI-(+)) using a flow rate of 0.2 ml/min during direct injection of peptide samples into the ion source. The ESI spectrometer was tunned to acquire the 230 ^{o}C temperature of the capillaries before each injection. A combination of acetonitrile, methanol and water were used to dissolved all peptide samples before injection into the ion source. The exact mass for each peptide was calculated from amino acid sequence, using the **EXPASY** tool for proteomic analysis from Swiss Prot data bank. Positive ionization mode electrospray (ESI) mass spectroscopy was performed for each peptide.

2.4. RESULTS AND DISCUSSIONS

The individual peptides were synthesized mostly with the automatic Peptide Synthesizer from Applied Biosystems while the peptide libraries were synthesized manually using the "split and pool" method. Each peptide was tested by reversed-phase HPLC and ESI-MS (+) mode for their purity and mass (Table 13). All peptides synthesized were homogeneous pure and with the expected molecular weight. In general these peptides were directly used in the kinetic assay with a couple of exceptions when the re-purification by HPLC was necessary to be applied to increase the peptide purity.

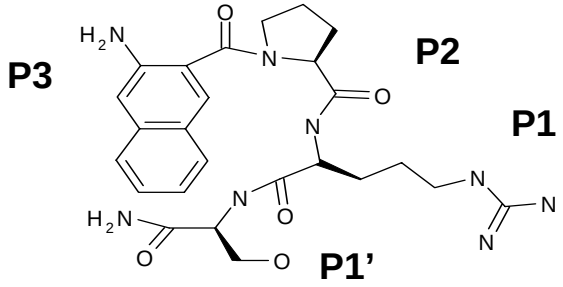
Table 13: Representative individual peptides synthesized with their theoretical and experimental mass together and their purity as determined by reversed-phase HPLC.

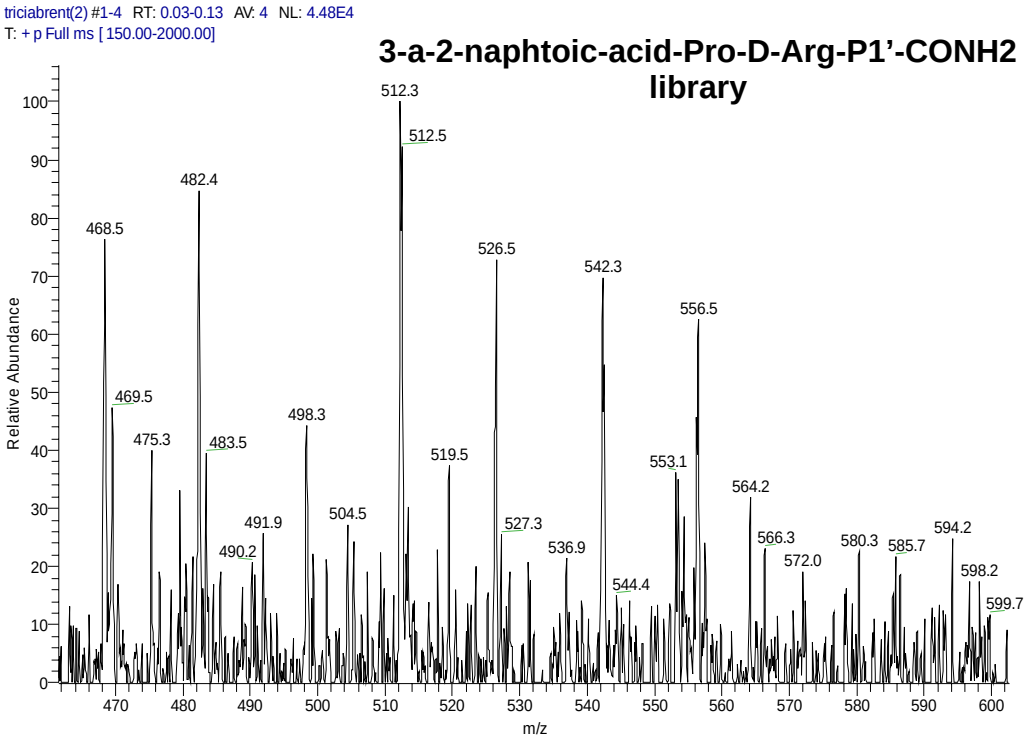
	Sequence N → C	**Theoretical M.W. ($g.mol^{-1}$)**	**Experimental M.W. ($g.mol^{-1}$)**	**HPLC: Purity in percentage**
1	**$^{\alpha}$N-CBZ-Pro-DArg-Gly-$CONH_2$**	**462.5**	**462.30**	**100**
2	**Transcinnamic-Pro-DArg-Ser-$CONH_2$**	**488.5**	**488.45**	**95**
3	**L-Tic-Pro-DArg-Cys-$CONH_2$**	**533.7**	**533.33**	**98**
4	**L-Thi-Pro-DArg-Gly-$CONH_2$**	**481.6**	**482.40**	**90**
5	**L-Tic-L-Thi-DArg-$CONH_2$**	**486.58**	**486.41**	**60**
6	**Dphe-Pro-DArg-DThr-$CONH_2$**	**519.61**	**519.47**	**100**
7	**L-Tic-Pro-DArg-DAla-$CONH_2$**	**501.7**	**501.40**	**98**
8	**Dphe-Pro-DArg-DCys-$CONH_2$**	**521.65**	**521.40**	**100**
9	**Transcinnamic-Pro-DArg-Cys-$CONH_2$**	**504.6**	**504.40**	**94**
10	**DPhe-Pro-DArg-Met-$CONH_2$**	**549.7**	**549.40**	**99**
11	**L-Tic-Pro-DArg-Thi-$CONH_2$**	**583.7**	**583.40**	**98**
12	**Transcinnamic-Pro-Darg-Ile-$CONH_2$**	**514.6**	**514.40**	**98**
13	**Dphe-Pro-DArg-DSer-$CONH_2$**	**565.58**	**565.47**	**100**
14	**Dphe-Pro-DArg-DGln-$CONH_2$**	**546.63**	**546.40**	**100**
15	**L-Tic-Pro-DArg-Gly-$CONH_2$**	**487.6**	**487.4**	**100**
16	**DPhe-Pro-DArg-Thi-$CONH_2$**	**571.7**	**571.40**	**85**
17	**DPhe-Pro-DArg-Thr-$CONH_2$**	**519.61**	**519.50**	**95**
18	**L-Tic-Pro-DArg-$CONH_2$**	**430.5**	**430.40**	**100**
19	**$^{\alpha}$N-CBZ-Pro-DArg Ala-$CONH_2$**	**476.6**	**476.6**	**60**
20	**DPhe-Pro-DLys-$CONH_2$**	**375.47**	**390.3**	**100**

21	**Transcinnamic-Pro-DArg-$CONH_2$**	**401.47**	**401.33**	**100**
22	**$^{\alpha}$N-CBZ-Pro-DArg-DAla-$CONH_2$**	**476.6**	**476.40**	**100**
23	**Dnaphthylalanine-Pro-DArg-Gly-$CONH_2$**	**525.3**	**525.3**	**98**
24	**Dnaphthylalanine-Pro-DArg-$CONH_2$**	**468.1**	**468.3**	**98**
25	**D-tetrahydroharman-Pro-Darg-DAla-**	**540.1**	**540.33**	**90**
26	**D-tetrahydroharman-Pro-Darg-Ile**	**582.2**	**582.35**	**90**
27	**Transcinnamic-Pro-DArg-DAla-$CONH_2$**	**472.6**	**472.20**	**95**

Figures 110 (A, B, C) are the ESI-MS (positive mode) of different peptide libraries synthesized by maintaining the same amino acid in the P3 position but varying amino acids at P1' position. Their corresponding theoretical and experimentally determined m/z ratio is also included. The synthesis was conducted to assure equimolar yield for each peptide variation. In most cases the yield of each peptide within the library mix was determined by the efficiency of coupling in P1' position. Amino-acids such as L-Pro or L-His were less efficient in coupling. Double coupling procedures were applied in the cases were the yield of coupling was small. The Kaizer test (1) and other important coupling or deprotection procedures were performed as already published and reviewed in many F-moc solid phase peptide synthesis reviews (1-3). As can be seen from Figure 110 the peptide libraries were successfully synthesized using one-bead-one peptide method with parallel couplings in P1' and further mixing in one pool of the whole peptide library since the ESI-MS scan of the whole library mixture shows all expected m/z ratios as predicted by calculated theoretical molecular mass for each individual peptide.

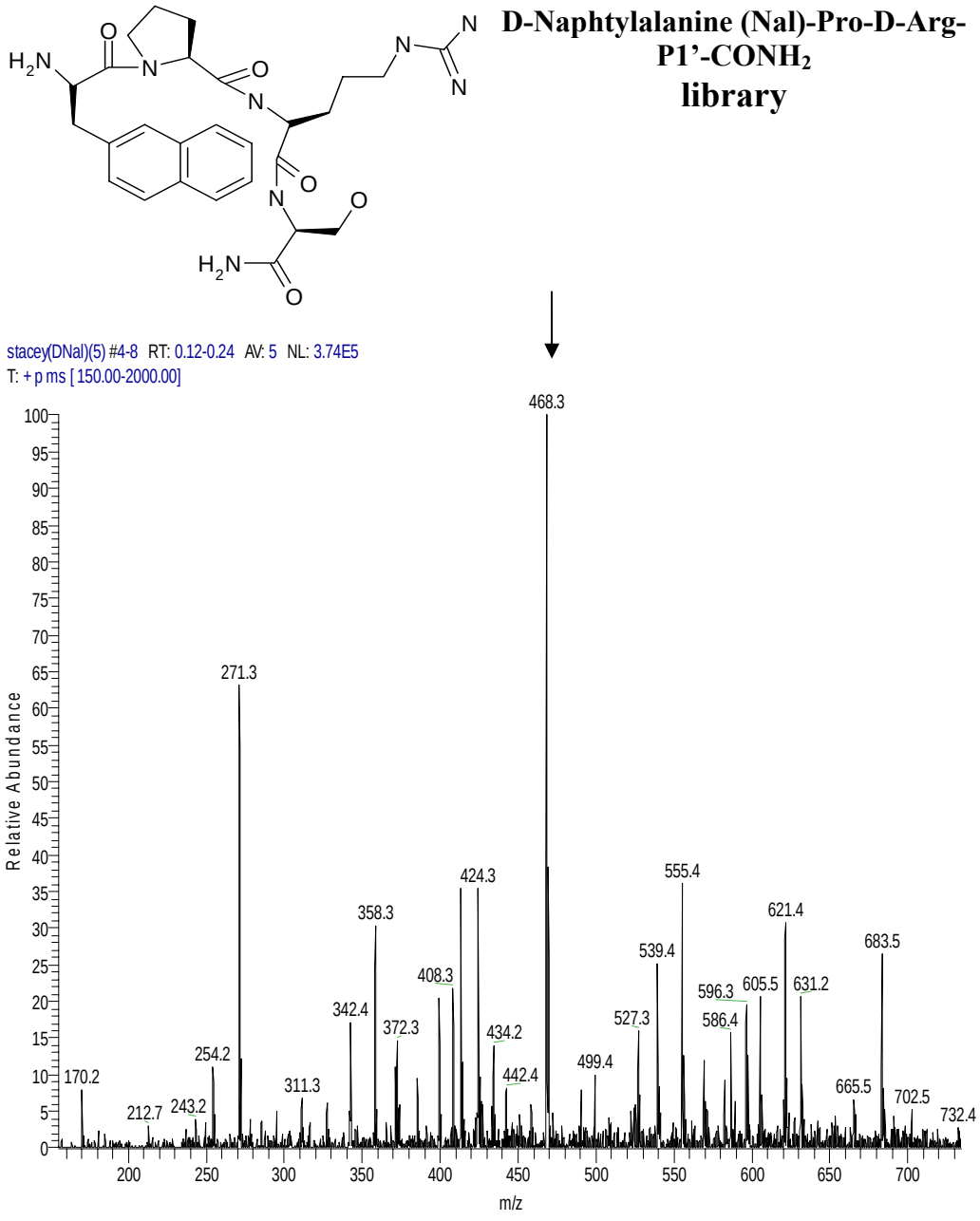
Figure 110 (A):

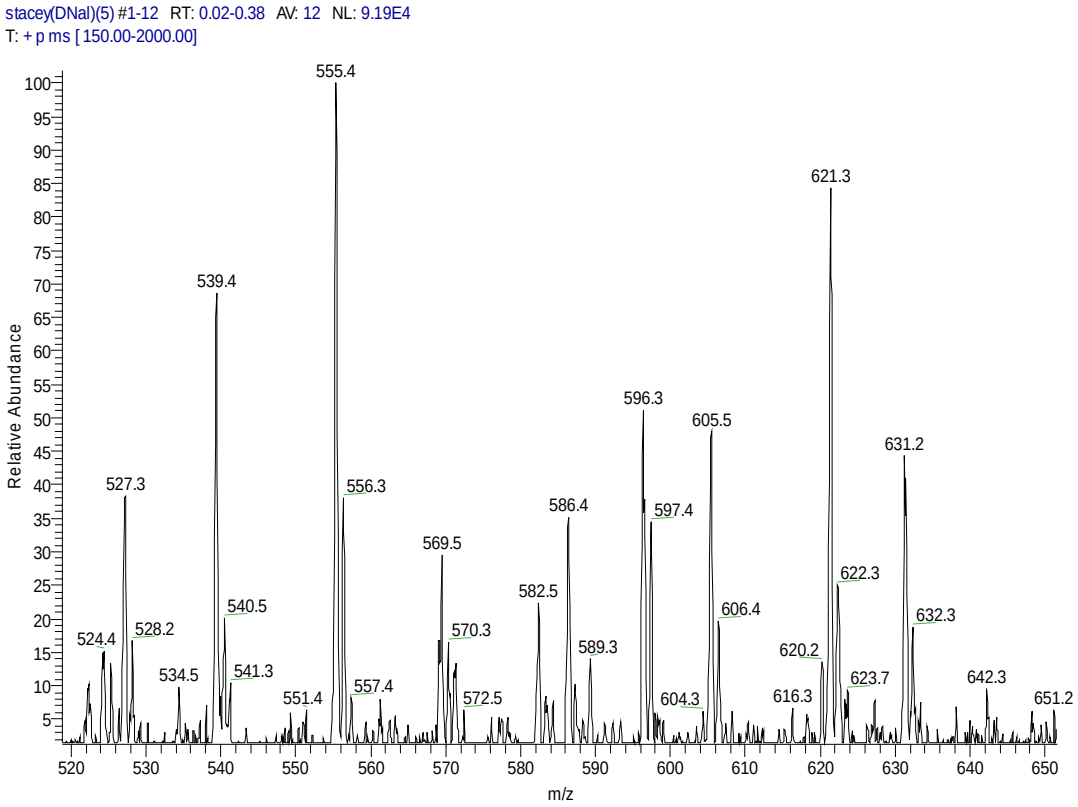




3-a-2-naphtoic-acid-Pro-DArg-P1'	Theoretical Mass	Experimental Mass
L-Ser	529.62	526.5
L-Cys	545.69	542.3
L-Thr	543.65	542.2
L-Ala	513.62	512.3-514.3
L-Gly	499.60	498.3

Figure 110 (B):


D-Naphtylalanine (Nal)-Pro-D-Arg-P1'-CONH2
library
stacey(DNal)(5) #4-8 RT: 0.12-0.24 AV: 5 NL: 3.74E5
T: + p ms [150.00-2000.00]
Relative Abundance
m/z
468.3
271.3
424.3
555.4
358.3
621.4
539.4
683.5
408.3
596.3
605.5
631.2
342.4
372.3
527.3
586.4
434.2
254.2
499.4
442.4
170.2
311.3
665.5
702.5
212.7
243.2
732.4



D-Nal-Pro-DArg-P1'	Theoretical Mass	Experimental Mass
L-Ser	553.68	555.4
L-Cys	569.75	569.5-570.3
L-Ala	537.68	539.4-540.5
D-Gln	594.74	596.3-597.4
D-His	603.73	605.5-606.4
D-Thr	567.71	569.5
D-Tyr	629.72	631.2
D-Asn	581.27	582.5
L-Thi	618.80	621.3

Figure 110 (B). Zoomed section of the ESI_MS scan.

Figure 110 (C) :

***Trans*-cinnamic/*Dihydro*-cinnamic-Pro-D-Arg- P1'-$CONH_2$**

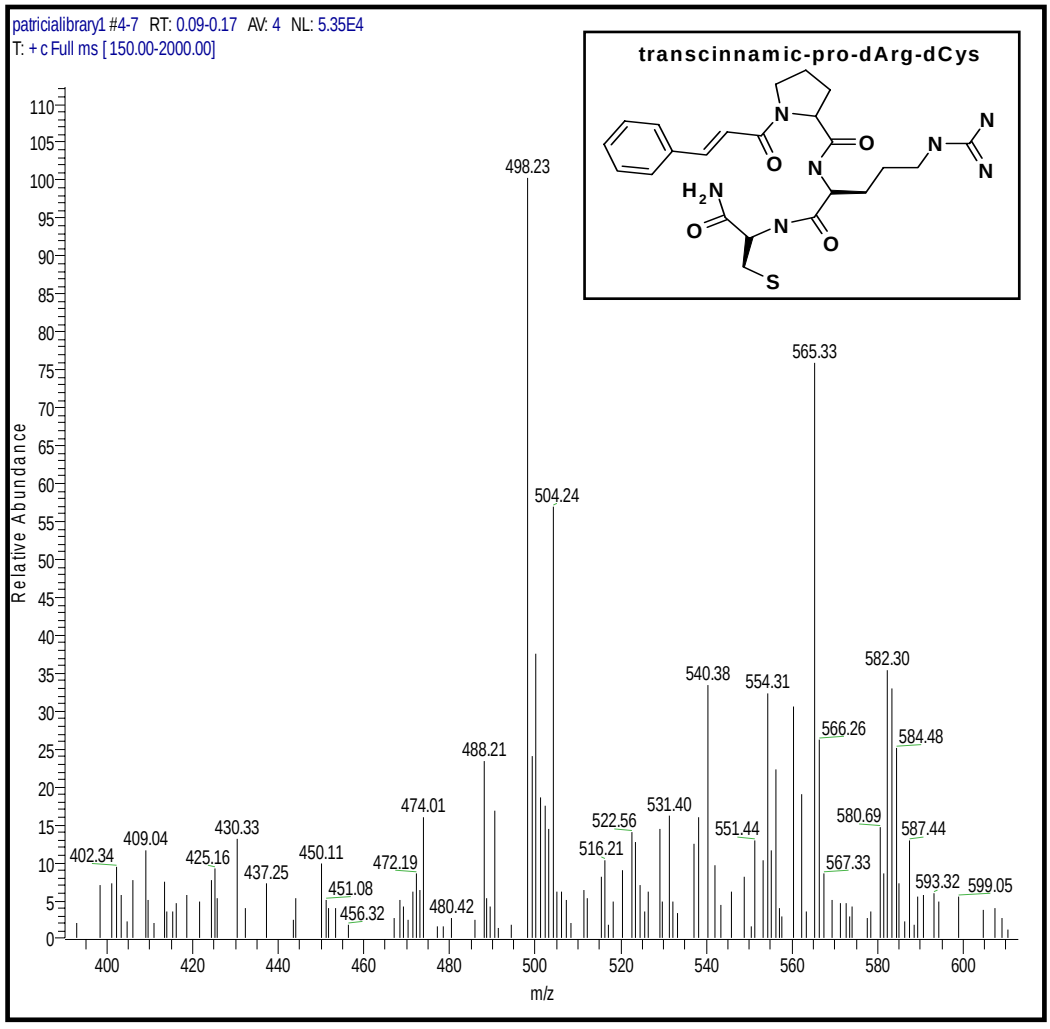


Figure 110 (A-C): ESI-MS (+) mass spectrometry of peptide libraries synthesized using "one-bead-one compound" synthetic procedure.

Some peptide libraries were partially purified by reversed phase HPLC and the masses of their mixtures were re-tested by ESI-(+). As can be seen from figures 111 and 112 the yield of some peptides in these libraries are enhanced as compared with starting material (figure 110-C).

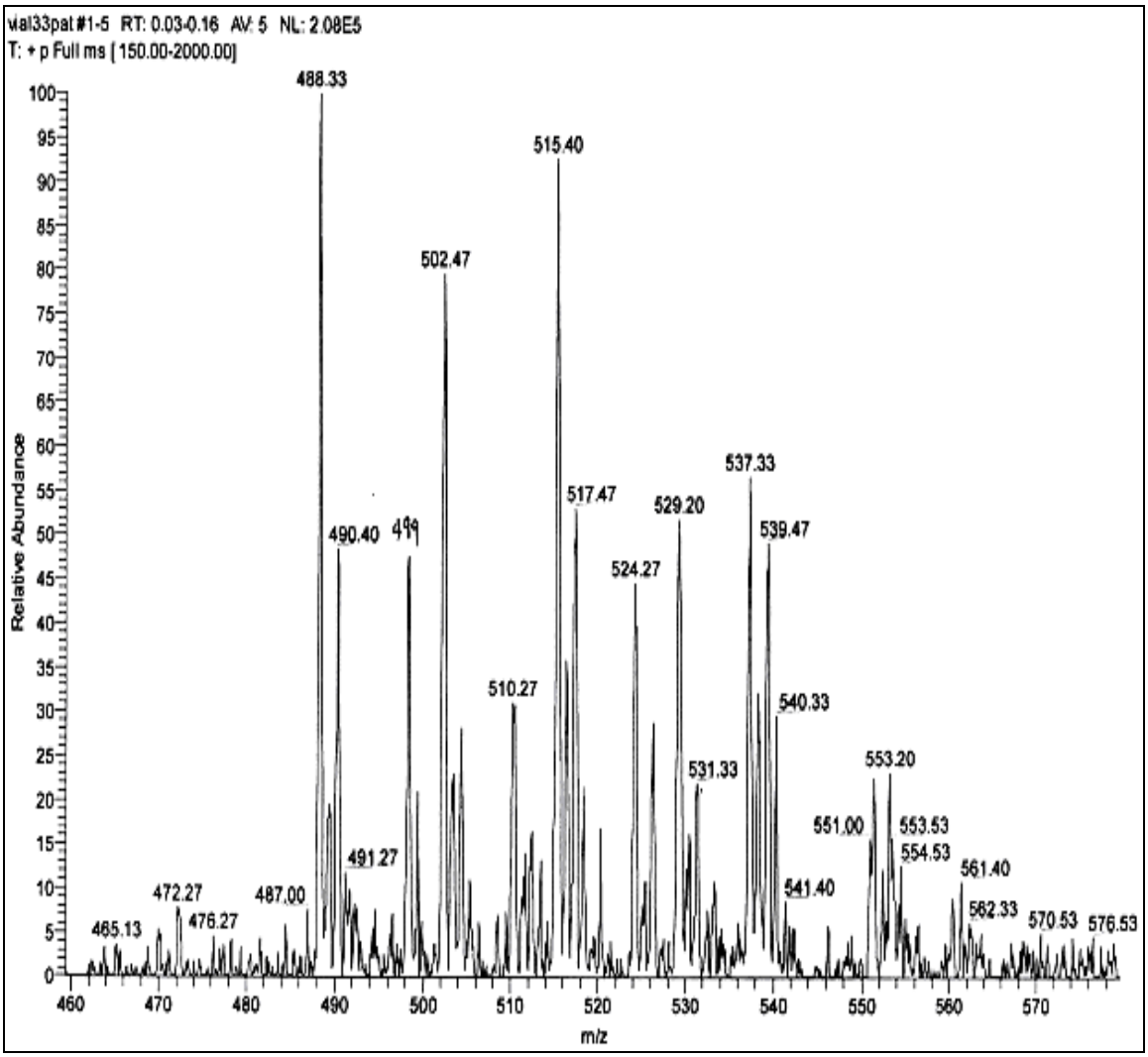


Figure 111. Mass spectroscopy result of a mixture of peptides purified by HPLC: *Trans*-cinnamoyl-Pro-D-Arg-D-Ser-$CONH_2$ (MW = 487.60, m/z = 488.33), Dihydrocinnamoyl-Pro-D-Arg-D-Ser-$CONH_2$ (MW = 489.62, m/z = 490.40)., *Trans*-cinnamoyl-Pro-D-Arg-D-Pro-$CONH_2$ (MW = 497.64, m/z = 499.00), *Trans*-cinnamoyl-Pro-D-Arg-D-Thr-$CONH_2$ (MW = 501.63, m/z = 502.47),*Trans*-cinnamoyl-Pro-D-Arg-D-Asn-$CONH_2$ (MW = 514.63, m/z = 515.40), Dihydrocinnamoyl-Pro-D-Arg-D-Asn-$CONH_2$ (MW = 516.65, m/z = 517.47) ,*Trans*-cinnamoyl-Pro-D-Arg-D-Gln-$CONH_2$(MW = 528.66, m/z = 529.05), Dihydrocinnamoyl-Pro-D-Arg-D-Gln-$CONH_2$ (MW = 530.68, m/z = 531.40), Dihydrocinnamoyl-Pro-D-Arg-D-Cys-$CONH_2$ (MW =506.69, m/z = 529.69 (+Na, MW = 23), *Trans*-cinnamoyl-Pro-D-Arg-L-Thi-$CONH_2$ (MW = 553.77, m/z = 554.31), *Trans*-cinnamoyl-Pro-D-Arg-D-His-$CONH_2$(MW = 537.67, m/z = 537.33-539.47), *Dihydrocinnamic*-cinnamoyl-Pro-D-Arg-D-His-$CONH_2$(MW = 539.69, m/z = 540.38), *Trans*-cinnamoyl-Pro-D-Arg-D-Thr-$CONH_2$(MW = 501.63, m/z = 524.63 (524.27, + Na (MW =23)).

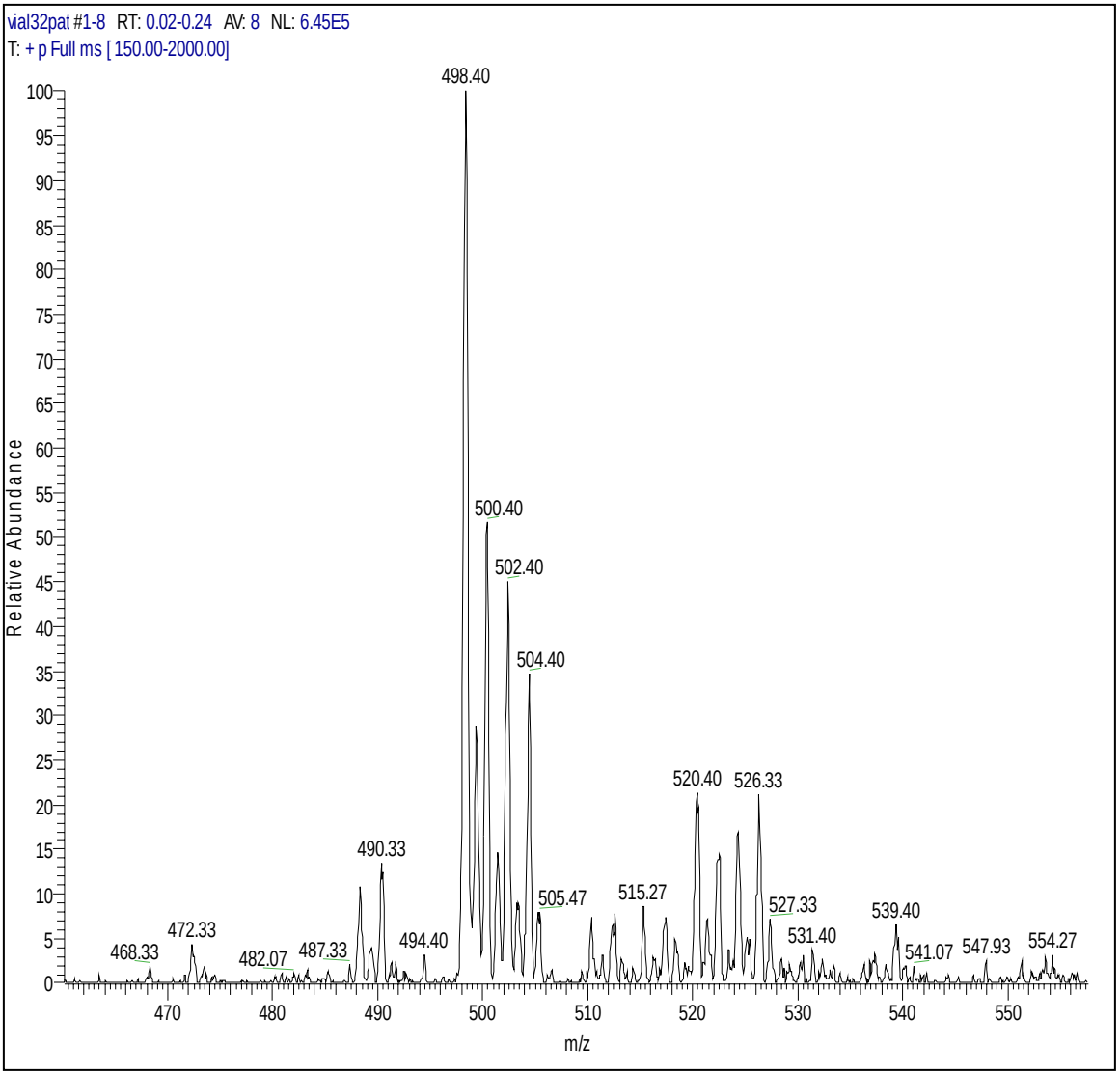


Figure 112: Mass spectroscopy result of a mixture of peptides: *Trans*-cinnamoyl-Pro-D-Arg- D-Pro-CONH2 (MW = 497.64, m/z = 498.23), *Dihydro*cinnamoyl-Pro-D-Arg-D-Pro-CONH2 (MW = 499.66, m/z = 500.40), *Trans*-cinnamoyl-Pro-D-Arg-D-Thr-CONH2 MW = 501.63, m/z = 502.22), *Dihydro*cinnamoyl-Pro-D-Arg-D-Thr-CONH2 (MW = 503.65, m/z = 504.24).

Other original ESI-(+)-MS of peptide libraries and of purified peptides are presented in APPENDIX X.

The peptide libraries were subjected to reversed-phase HPLC for further purification of individual peptides. In Appendix XI some of the HPLC preparative runs and analytical HPLC control of purity of some peptides are presented. The structure-

activity relationship was performed only for purified peptides using in vitro kinetics-assay of thrombin inhibition (Chapter-3).

REFERENCES

1. Chen, W.C., White P.D. In F-moc Solid Phase Peptide Synthesis; Chen, W.C., White P.D., Eds.; Oxford University Press, New York, 2000; Chap 3, 41-76.

2. Aggarwal S, Harden JL, Denmeade SR. 2006. Synthesis and screening of a random dimeric peptide library using the one-bead-one-dimer combinatorial approach. Bioconjug Chem. 17(2):335-40.

3. Choe Y, Leonetti F, Greenbaum DC, Lecaille F, Bogyo M, Bromme D, Ellman JA, Craik CS. 2006. Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. J. Biol Chem.

4. Okochi M, Nakanishi M, Kato R, Kobayashi T, Honda H. 2006. High-throughput screening of cell death inducible short peptides from TNF-related apoptosis-inducing ligand sequence. FEBS Lett. 580(3):885-9.

5. Kopecky EM, Greinstetter S, Pabinger I, Buchacher A, Romisch J, Jungbauer A. 2005. Combinatorial peptides directed to inhibitory antibodies against human blood clottingfactorVIII. Thromb Haemost. 94(5):933-41.

6. Nakamura C, Inuyama Y, Goto H, Obataya I, Kaneko N, Nakamura N, Santo N, MiyakeJ. 2005. Dioxin-binding pentapeptide for use in a high-sensitivity on-bead detection assay. Anal Chem. 77(23):7750-7.

7. Wang W, McMurray JS, Wu Q, Campbell ML, Li C. 2005. Convenient solid-phase synthesis of diethylenetriaminepenta-acetic acid (DTPA)- conjugated cyclic RGD peptide analogues. Cancer Biother Radiopharm. 20(5):547-56.

8. Perlman ZE, Bock JE, Peterson JR, Lokey RS. 2005. Geometric diversity through permutation of backbone configuration in cyclic peptide libraries. Bioorg Med Chem Lett. 15(23):5329-34.

9. Shin DS, Kim DH, Chung WJ, Lee YS. 2005. Combinatorial solid phase peptide synthesis and bioassays. J Biochem Mol Biol. 38(5):517-25.

10. Stephenson KA, Banerjee SR, Sogbein OO, Levadala MK, McFarlane N, Boreham DR, Maresca KP, Babich JW, Zubieta J, Valliant JF.

2005. A new strategy for the preparation of peptide-targeted technetium and rhenium radiopharmaceuticals. The automated solid-phase synthesis, characterization, labeling, and screening of a peptide-ligand library targeted at the formyl peptide receptor. Bioconjug Chem. 16(5):1189-95.

11. Bourne GT, Nielson JL, Coughlan JF, Darwen P, Campitelli MR, Horton DA, Rhumann A, Love SG, Tran TT, Smythe ML. 2005. A convenient method for synthesis of cyclic peptide libraries. Methods Mol Biol. 298:151-65.

12. Aina OH, Marik J, Liu R, Lau DH, Lam KS. 2005. Identification of novel targeting peptides for human ovarian cancer cells using "one-bead one-compound" combinatorial libraries. Mol Cancer Ther. 4(5):806-13.

13. Hilpert K, Volkmer-Engert R, Walter T, Hancock RE. 2005. High-throughput generation of small antibacterial peptides with improved activity. Nat Biotechnol. 23(8):1008-12.

14. Falciani C, Lozzi L, Pini A, Bracci L. 2005. Bioactive peptides from libraries. Chem Biol. 12(4):417-26.

15. Gosalia DN, Salisbury CM, Ellman JA, Diamond SL. 2005. High throughput substrate specificity profiling of serine and cysteine proteases using solution-phase fluorogenic peptide microarrays. Mol Cell Proteomics. 4(5):626-36.

16. Teixido M, Albericio F, Giralt E. 2005. Solid-phase synthesis and characterization of N-methyl-rich peptides. J Pept Res. 65(2):153-66.

17. Pastor MT, Mora P, Ferrer-Montiel A, Perez-Paya E. 2004. Design of bioactive and structurally well-defined peptides from conformationally restricted libraries. Biopolymers. 76(4):357-65.

18. Sasubilli R,Gutheil WG. 2004. General inverse solid-phase synthesis method for C-terminally modified peptide mimetics. J Comb Chem. 6(6):911-5.

Chapter 3

Kinetics analysis of thrombin inhibition by synthetic peptides

3.1. INTRODUCTION

The search for small, orally bioavailable, non-peptidic synthetic inhibitors of thrombin has been intense in many pharmaceutical companies (1-10). Two small molecular weight inhibitor types are emerging as structure-activity relationships were primarily investigated in the last years. The first is argatroban and NAPAP type (11-12), were the lipophilic groups on either side of the basic P1 chain pack together to interact with the hydrophobic S2 site (13-15). Napsagatran (Ro 46-6240) developed by Hilpert et al. (16) has a more complex P1 residue (see Scheme 7) but belongs to the same group as argatroban and NAPAP. The only interaction with the catalytic residue Ser 195 is via a hydrogen bond to the carboxylate function in both argatroban and napsagatran. Unfortunately, none of these compounds is orally active due to either poor absorption from the gastrointestinal tract and/or rapid clearance via the bile (17). A second inhibitor type is based on the substrate-derived irreversible chloromethyl ketone inhibitor PPACK and includes compounds such as DuP-714 and efegatran (1-17). These compounds interact covalently with the hydroxyl group of catalytic Ser-195 residue. Though oral activity has been claimed for these compounds, the selectivity toward thrombin is weaker due to the conserved catalytic Ser 195 within the active site of different serine proteases (18-25).

In addition to the above mentioned groups of thrombin inhibitors, potent and selective inhibitors of thrombin were derived based on the D-Phe (P3)-Pro (P2)-Arg (P1) sequence in which Arg residue was replaced by bezamidine or by 1-amidinopiperidine

which conserved the salt-bridge with Asp-189 from the active site of thrombin (18-22). An important difference between thrombin and trypsin is the replacement of Ser 190 in trypsin by alanine in thrombin. This renders the S1 pocket of thrombin bigger and more hydrophobic. As part of an effort to generate highly selective thrombin inhibitors some research groups were exploring other substituents at P1 position while maintaining conserved the hydrophobic moieties at P3 and P2 position. Thus, the P1 position was probed with substituted five-membered hetaryl amidines which gave a different special display of the substituents than the often used paraaminomethylbenzamidines. An additional small substituent on the hetrocyclic ring was expected to result in a disfavorable interaction with the Oγ atom of Ser190 in trypsin thereby increasing the selectivity toward thrombin. Indeed, the compounds were shown more than 1000 fold higher selectivity toward thrombin than trypsin (25-30).

We choose to develop peptide-inhibitors using the original sequence D-Phe (P3)-Pro (P2)-Arg (P1) but maintaining constant the Pro at P2 and switching the L-Arg into D-Arg to obtain a nonhydrolyzable scissile bond -D-Arg-X- . We also scanned the P3 positions with different amino-acids analogs or with trans, cis and dyhidrocinnamic acid in an effort to increase the lipophilicity for the S3 pocket and to investigate other conformational constraints at P3 position which would enhance the affinity of the peptide ligand for the target thrombin. The structure-activity-relationship (SAR) for 66 peptides derived from the sequence D-Phe (P3)-Pro (P2)-D-Arg (P1) is the subject of the actual chapter. The peptides were screened for their structural fitness within the active site of thrombin by molecular docking (Chapter 1). In all doking experiments the carbonyl carbon from D-Arg (P1) was shown to have more than 4.5 Å interatomic distance from the Oγ of Ser195 from the active site of thrombin. It is very well known from X-ray data

(31-40) that the carbonyl carbon of L-Arg from P1 is within less than 3 Å to Oγ of Ser195 in compounds were the scissile bond L-Arg-X is hydrolysable (40-50). Based on other X-ray studies an interatomic distance higher than 4.5 Å between carbonyl carbon of D-Arg (P1) and Oγ of Ser195 from thrombin would render the scissile bond D-Arg-X nonhydrolyzable (45-50). Based on these structural evidences and based on our own docking experiments we synthesized peptides with D-Arg in P1 and investigated the P1', P2' and P3' positions within the peptide for best sequence space that would confer both protease stability and higher inhibitory potential. The P3 position was scanned with Phe analogs (D-Naphthylalanine (Nal), L and D-tetraisoquinoline (Tic), D-tetrahydroharman acid (D-Tpi), L-beta-2-thienylalanine (L-Thi)) and with *trans*, *cis* and *dihydro*-cinnamic acid. All peptides were non-hydrolyzable as demonstrated by their mass spectroscopy analysis after preincubation for 1 hour at room temperature with thrombin, which showed the intact peptides, supporting the structural predictions from docking experiments.

3.2. MATERIALS AND METHODS

All kinetics experiments for thrombin inhibition were performed using the colorimetric assay with **S2238** [H-D-Phenylalanyl-L-pipecolyl-L-Arginine-p-nitro-anilide dihydrochloride (**H-D-Phe-Pip-Arg-pNA.2HCl)]** (Chromogenix) as the specific chromogenic substrate for thrombin. The experiments were conducted at 25^{o}C, using a thermostated Cary-UV-VIS Spectrophotometer in phosphate buffer, pH=7.46, of 0.1 I (ionic strength), containing 0.2 M NaCl and 50 μg/ml BSA (bovine serum albumin). The addition of BSA was required by the fact that thrombin is adhering on plastic cuvettes through electrostatic interactions, decreasing the final rate of reactions up to 100 fold as compared with the reaction run in the presence of BSA. The final reaction volume was

1ml. The hydrolysis of the chromogenic substrate was monitored by measuring changes in absorbance at 405 nm using the built-in Kinetics software of Cary-UV-VIS spectrophotometer. Each kinetic reaction was performed to completion during 12-20 minutes and the signal for changing absorbance at 405 nm was recorded each 1 second. Kinetics of thrombin inhibition were performed using 2 different substrate concentrations:

(1) In one set of experiments the reactions were performed in **pseudo-first order** kinetics conditions (i.e. at substrate concentration below the Km, in most cases 0.7 Km, or near Km). Of course the most rigurous pseudo-first order kinetics are run at substrate concentration below Km, but the signal at 405 nm should be reliable such that the fit of the experimental data to the pseudo-first order mathematical model can be done. The Km of bovine thrombin for S2238 is around 4 μM as determined experimentally using the steady-state assay. Therefore, the pseudo-first order kinetics were conducted at 2.8 μM (0.7 Km). The pseudo-first order kinetics were used to determine the observed rate of hydrolysis of substrate (**kobs**), from which the inhibitory constant was determined **(K_I = [I] / ($kobs_{uninhib}$/$kobs_{inhibited}$ -1) (I)**, where [I] = inhibitor concentration, $kobs_{uninhib}$ = observed pseudo-first order rate in the absence of inhibitor and $kobs_{inhibited}$ = observed pseudo-first order rate in the presence of [I] concentration of inhibitor). The Ki were derived from this equation by assuming competitive inhibition model. In this case using the relationship, v= kcat [E] [S]/(Km (1+ [I]/Ki) +[S]) or

[E]/v = 1/kcat + (Km (1+ [I]/Ki))/(kcat [S]) it follows that a plot of [E]/v vs. [I] at constant [S] will be linear if competitive inhibition holds. Furthermore, under the conditions [S] <<< Km, the pseudo-first-order rate constant k equals v/[E]. Therefore,

k^{-1} vs. [I] should be linear and Ki can be obtained by using the equation:

$1+[I]/Ki = [E]_i/v_i \ . \ v_0/[E]_0 = k_0/k_i$ from were the equation $K_i = [I]/(k_o/k_i -1)$ which is the same with the equation (I) were k_i and k_0 are the pseudo-first order constants determined in the presence and in the absence of the inhibitor, respectively.

Each kinetic trace was fitted to the first order reaction mathematical model $A_t = A_0e^{-kt}$ where $A_t =$ the absorbance at 405 nm at any time during 20 minutes of run, $A_0=$ the absorbance at 405 nm at time t=0, and k= the observed pseudo-first order rate constant. The fit was performed using the Nonlinear Regression built-in software of the Cary UV-spectrophotometer, and the **kobs** obtained from the fit were compared for the reactions performed in the presence of inhibitor with those obtained for reactions performed in the absence of inhibitor.

(II) In another set of experiments the kinetics were run at five times Km (5Km) or eight times Km (8Km) concentration of substrate S2238 **(saturation kinetics**) in order to determine the type of inhibition for each peptide using Lineweaver-Burk or Eadie-Hofstee plots. The reactions were allowed to go to completion, during 12-20 minutes and the slopes at different time points were used to determine the velocities at that point. The difference between the absorbance at 405 nm after 20 minutes of run (which is proportional with the total amount of substrate present in the reaction mixture and represent the infinite absorbance value) and the absorbance at 405 nm taken in the middle of the slope for each data point is proportional to the substrate concentration in reaction mixture at that point. In all calculations a molar extinction coefficient of 8,800 $[M^{-1}cm^{-1}]$ at 405 nm was used for the reaction product (pNA). We assumed that upon hydrolysis of S2238 substrate for each molecule of S2238 one molecule of para-nitroanilide was

liberated. The progress curve analysis was performed using the software provided by the Cary-UV-VIS spectrophotometer. The following formulas were used to calculate the substrate concentrations (in μM) and velocity (in μM/min) at any point on the progress curve:

(1) S (substrate) = $\{[A_{00}-A_{midle}]/8{,}800\} \times 10^6$ (μM) were A_{00} represents the infinite absorbance value (i.e. the absorption after the reaction proceeded to completion).

(2) V (velocity) = $\{(\Delta A/min)/8{,}800\} \times 10^6$ (μM/min) were ΔA/min is the variation of absorbtion/min (i.e. is the slope in the point on the kinetic curve were the velocity is calculated).

(3) A_{midle} (Absorbtion in the midle of 1 minute interval at which the slope is calculated) is calculated using the formula:

$A_{midle} = \{[A_{stop}-A_{start}]/2 + Astart\}$;

The exact concentration of thrombin was determined from the UV absorption spectrum using the molar extinction coefficient for thrombin (65,735 $M^{-1}cm^{-1}$) at 279 nm.

The final concentration of thrombin used in all kinetics experiments was determined from a preliminary pseudo-first order kinetics experiment in which the final concentration of substrate S2238 was kept constant (0.7 Km = 2.8 μM), while the final concentration of thrombin was varied between 10nM-40nM. A plot of kobs as a function of enzyme concentration determined that 30 nM thrombin final concentration was a reliable concentration since is in the linear range of kobs = f(enzyme concentration) and thus a linear relationship between the A (405 nm) and the enzyme concentration is expected (Figure 113).

The effect upon Km and Vmax were determined by assuming different models for thrombin inhibition. For the most peptide-inhibitors tested at low concentration of peptide a competitive inhibition model was assumed and in this case Km' (i.e. Km in the presence of the inhibitor) was increased with the value Km multiplied by (1+[I]/Ki), were the Km is the thrombin' Km for S2238 while the Vmax remained unchanged since the peptide is binding to the free enzyme E. At higher concentration of peptide (> 10 μM) both a non-competitive and an uncompetitive model for inhibition was assumed. In the case of non-competitive inhibition the peptide could bind both to free enzyme E and to the complex ES (enzyme-substrate) and in this case Km remained unaffected while the Vmax is decreased and is given by equation Vmax/(1+[I]/Ki) were Vmax is the maximal velocity in the absence of the inhibitor, [I] is the concentration of peptide inhibitor and Ki is the inhibitory constant. In the case of uncompetitive competition the peptide inhibitor could bind to the ES complex only, thus both Km and Vmax were decreased by (1+[I]/Ki) value.

In order to confirm the validity of the assay for determining the inhibition constant, the steady-state saturation kinetics was used to evaluate the K_i for inhibition of thrombin by the commercially available inhibitor ("*Thromstop*") (K_i = 25nM for bovine thrombin).

3.3. RESULTS

Kinetics of thrombin inhibition by synthetic peptides

The preliminary steady-state kinetics were conducted to determine the Km and the kcat of bovine thrombin for the substrate S2238. The Km value was confirmed to be

in the range already published (1-17), i.e. 4.16 $\pm$ 0.63 μM and the Vmax was also in the range published, i.e. 7.60 (μM/min). The kcat (Vmax/Et) was however for fold lower than the published value for thrombin (4.21 [sec^{-1}]), suggesting that the enzyme was not fully active. However, for steady-state kinetics the % activity of thrombin is less important, thus all the kinetics experiments were conducted with this bovine thrombin inhibitor.

In the first set of experiments the linearity of the observed pseudo-first rate of hydrolysis (kobs) as a function of the enzyme concentration was determined by maintaining the substrate concentration constant (0.7 Km) and varying the concentration of enzyme between 10 nM and 30 nM. In Figure 113, the graph describing the linearity of kobs with the enzyme concentration is presented. The experimental data were fitted to a linear regression analysis. It can be seen that when the enzyme concentration was between 10 nM and 30 nM there was a linear relationship between kobs and the enzyme concentration. This experiment determined us to choose 30 nM, the enzyme concentration of thrombin at which all kinetics experiments were performed.

The pseudo-first order kinetics experiments with different peptides inhibitors were conducted at concentrations of S2238 substrate below Km (0.7 Km), or near Km (2Km) as already described in *"Materials and Methods"* and the concentration of each peptide was varied between 1-500 μM.

The individual kinetic traces with some of the peptides that were tested for the inhibitory activity are presented in the Figure 114 (A-T). The kinetic raw data were imported from Cary-UV spectrophotometer by use of the program Sigma Plot and the nonlinear fit of the kinetics traces to first order reaction ($A_t = A_0e^{-kt}$) was confirmed with the software NonLin Regression package. For the most potent inhibitors with inhibitory

constant Ki (K_I = [I] / ($kobs_{uninhib}$/$kobs_{inhibited}$ -1)) in the range of 100-150 μM the linearized function of the first order fit was used to confirm the kobs obtained from nonlinear regression analysis of kinetic traces. The linearized form of the first order mathematical model is described by the equation: ln (A_{00}-A_t) = lnA_0-(kobs) t, where A_{00} and A_t have the significance as described in *"Materials and Methods"* (figure 114, H and I).

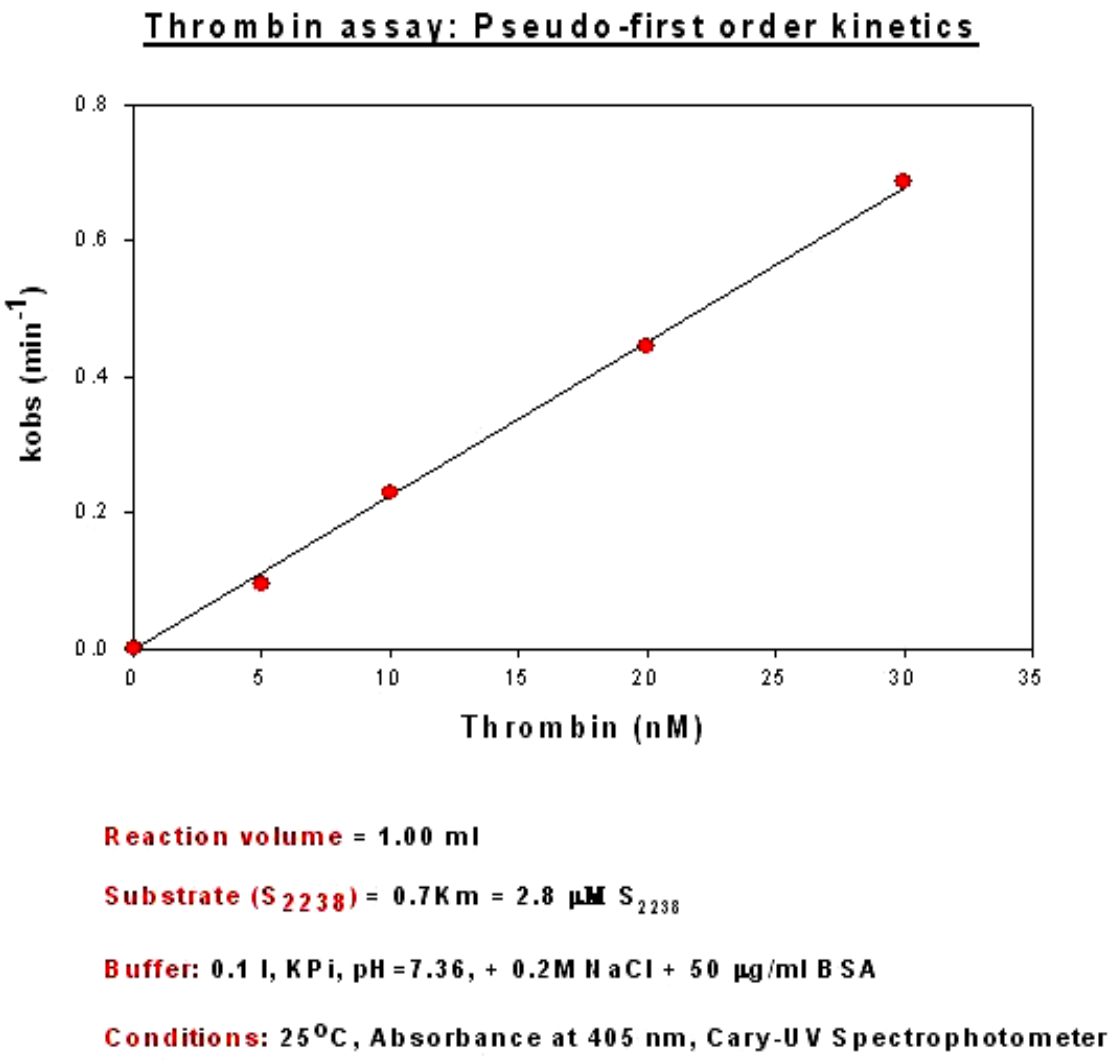
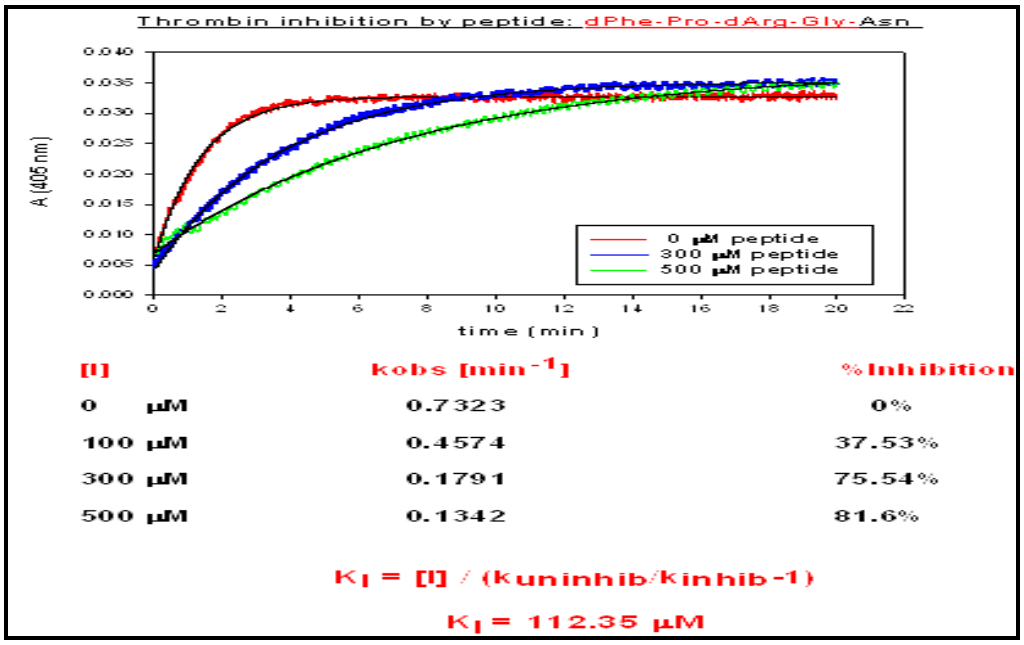


Figure 113: The linearity of kobs with the enzyme concentration over the range 10 nM-40 nM.

Figure 114 (A-S): The pseudo-first order kinetics for thrombin inhibition by different peptides.

A

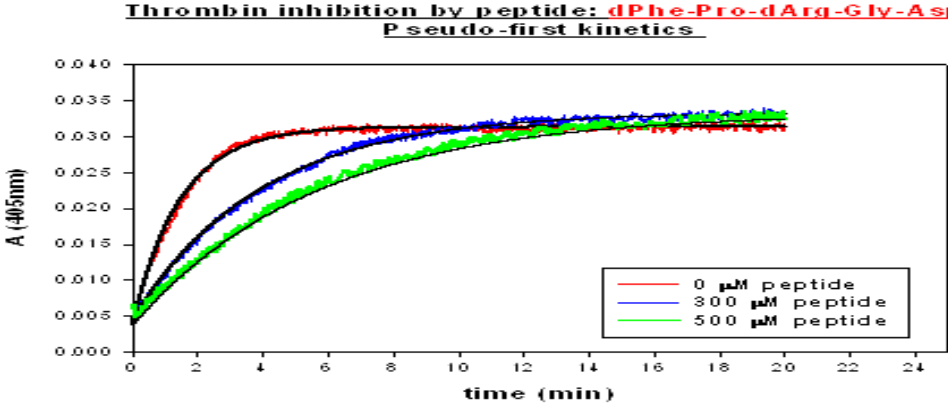


[I]	kobs [min^{-1}]	%Inhibition
0 µM	0.7323	0%
100 µM	0.4574	37.53%
300 µM	0.1791	75.54%
500 µM	0.1342	81.6%

K_I = [I] / (kuninhib/kinhib-1)

K_I = 112.35 µM

B

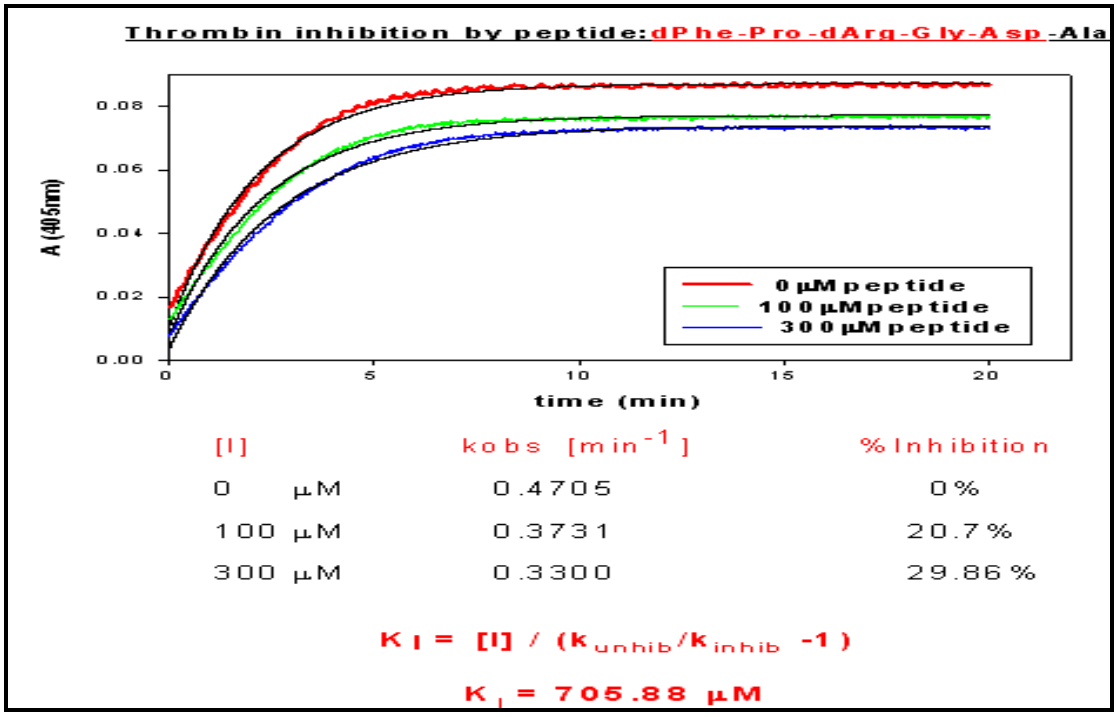


[I]	kobs [min^{-1}]	% Inhibition
0 µM	0.7034	0%
100 µM	0.4235	39.8%
300 µM	0.2534	64%
500 µM	0.1669	76.3%

K_I = [I] / (kuninhib/kinhib -1)

K_I = 156.25 µM

C

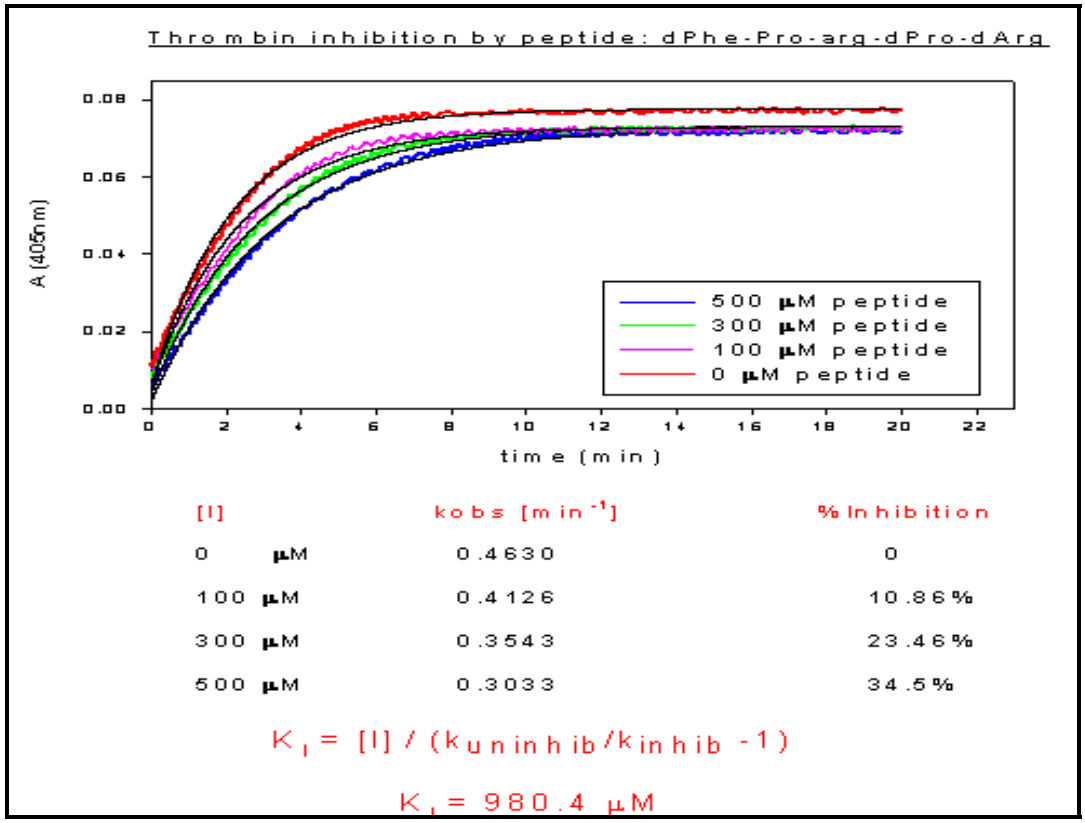


[I]	kobs [min^{-1}]	% Inhibition
0 µM	0.4705	0%
100 µM	0.3731	20.7%
300 µM	0.3300	29.86%

$$K_I = [I] / (k_{unhib}/k_{inhib} - 1)$$

$$K_I = 705.88\ \mu M$$

D

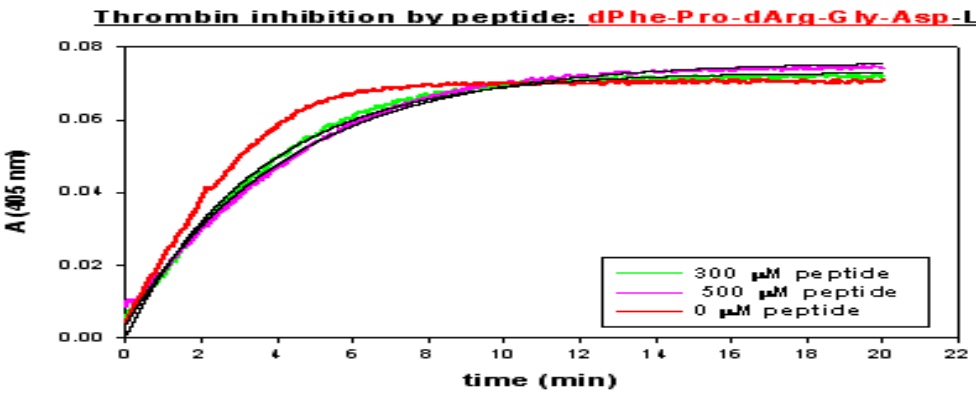


[I]	kobs [min^{-1}]	% Inhibition
0 µM	0.4630	0
100 µM	0.4126	10.86%
300 µM	0.3543	23.46%
500 µM	0.3033	34.5%

$$K_I = [I] / (k_{uninhib}/k_{inhib} - 1)$$

$$K_I = 980.4\ \mu M$$

E

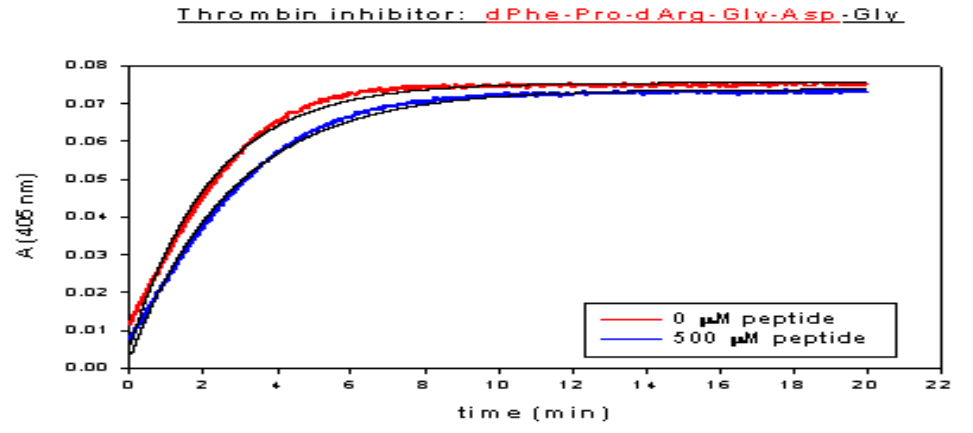


[I]	kobs [min^{-1}]	% Inhibition
0 µM	0.4175	0%
100 µM	0.3535	15.32%
300 µM	0.2847	31.80%
500 µM	0.2339	42.77%

$K_I = [I] / (k_{uninhib}/k_{inhib} - 1)$

$K_I = 643.77\ \mu M$

F

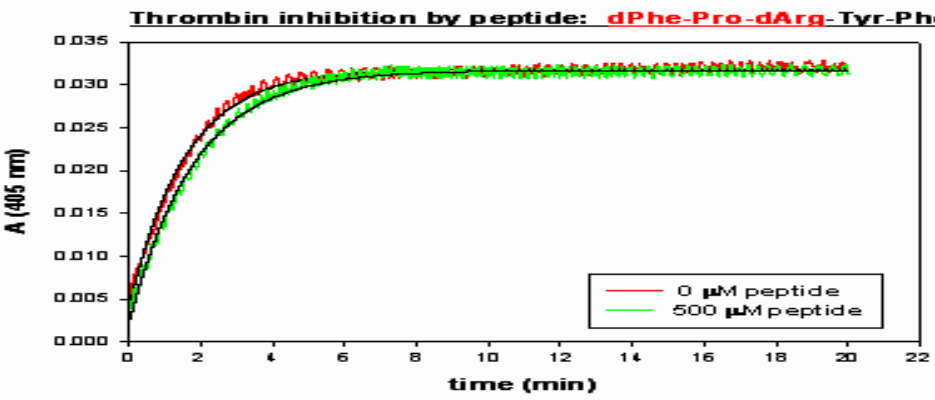


[I]	kobs [min^{-1}]	% Inhibition
0 µM	0.4502	0%
100 µM	0.4147	7.88%
300 µM	0.3837	14.77%
500 µM	0.3571	20.67%

$K_I = [I] / (k_{uninhib}/k_{inhib} - 1)$

$K_I = 1.73\ mM$

G

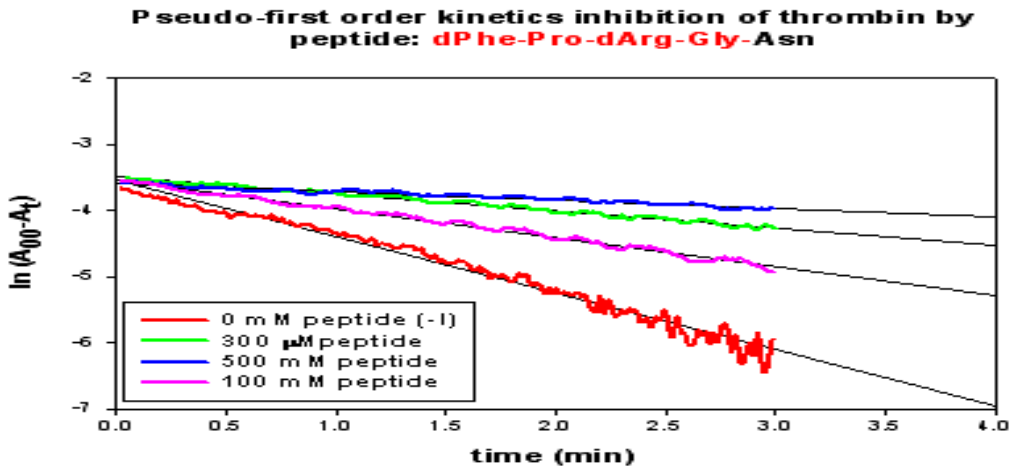


[I]	kobs [min^{-1}]	% Inhibition
0 µM	0.6444	0%
300 µM	0.5910	8.3%
500 µM	0.5546	14%

$K_I = [I] / (k_{uninhib}/k_{inhib} - 1)$

$K_I = 3.33$ mM

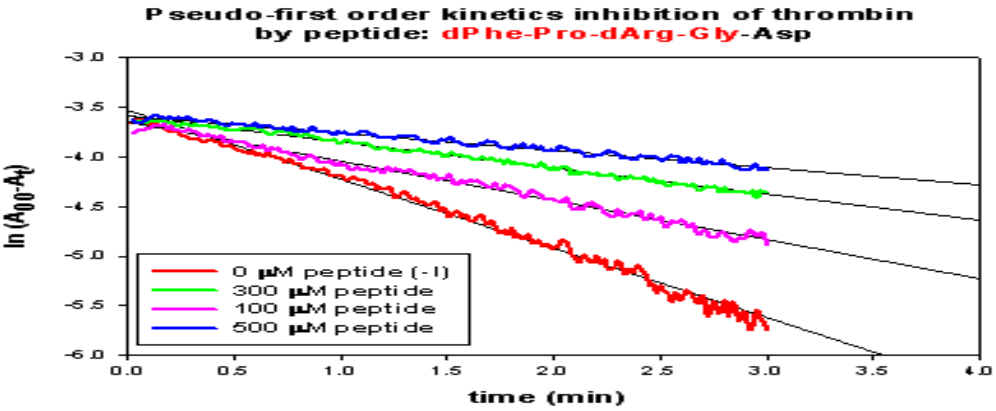
H



[I]	kobs [min^{-1}]	% Inhibition
0 µM	0.8516	0%
100 µM	0.4346	49.0%
300 µM	0.2548	70.07%
500 µM	0.1246	85.01%

$K_I = 104.3$ µM

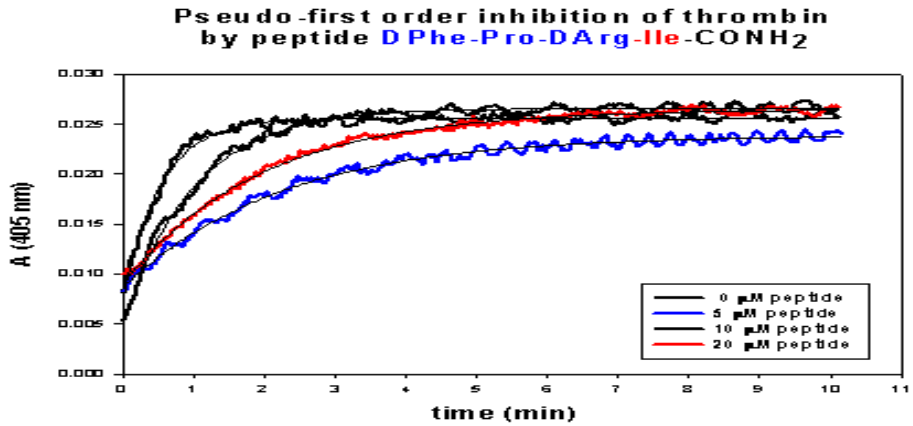
I



[I]	kobs [min^{-1}]	% Inhibition
0 µM	0.6962	0 %
100 µM	0.3934	43.5%
300 µM	0.2630	62.22%
500 µM	0.1737	75.05%

K_I = 130.04 µM

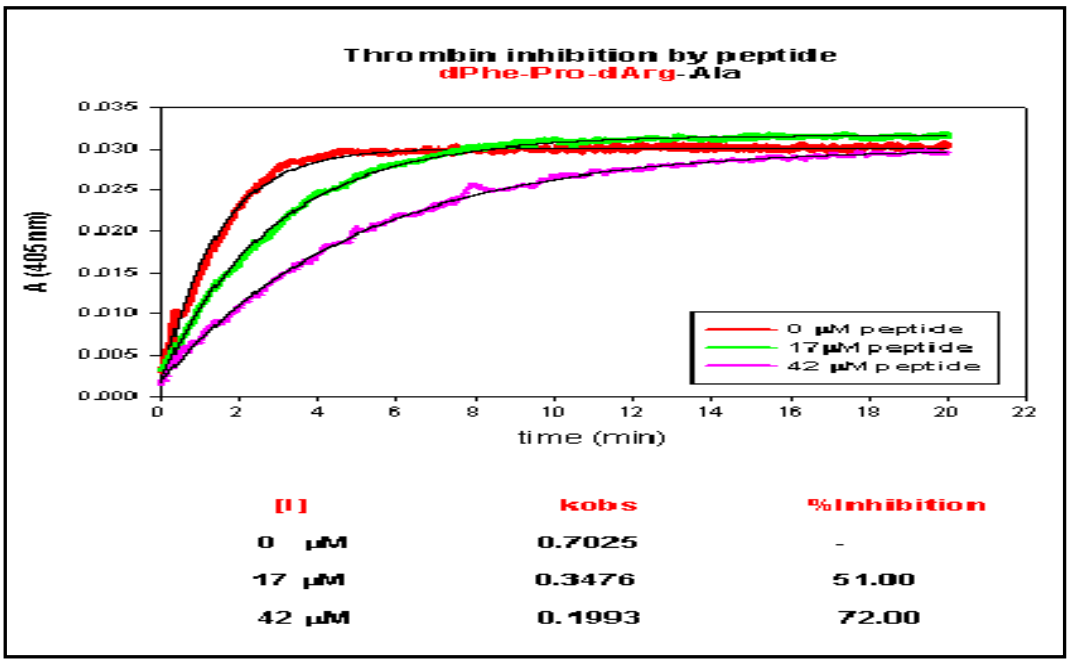
J



[I] [µM]	kobs (sec^{-1})	% Inhibition
0	1.885 ± 0.0250	-
5	1.007 ± 0.0094	46.57
10	0.5147 ± 0.0039	72.7
20	0.4495 ± 0.0058	76.2

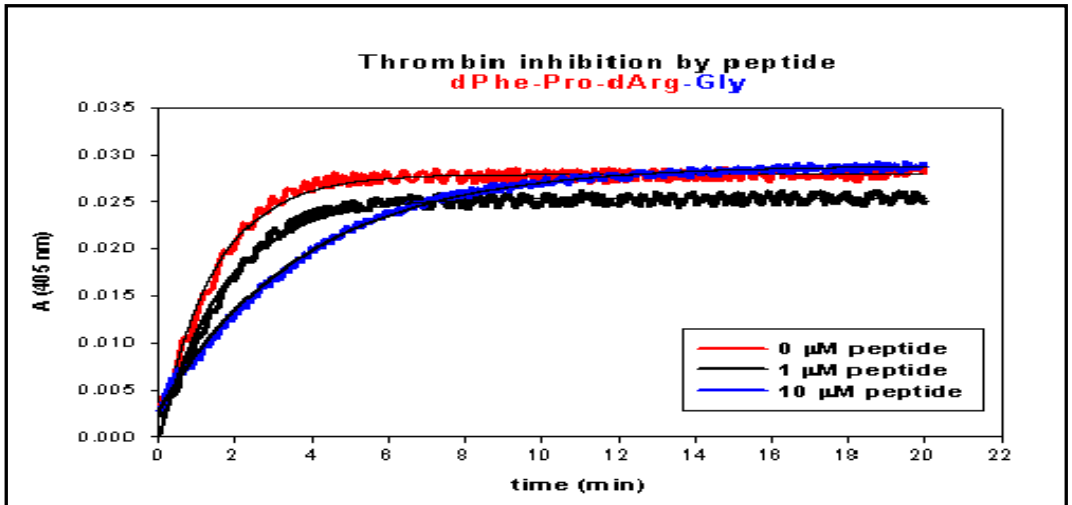
Ki = 6.20 µM

K



[I]	kobs	%Inhibition
0 µM	0.7025	-
17 µM	0.3476	51.00
42 µM	0.1993	72.00

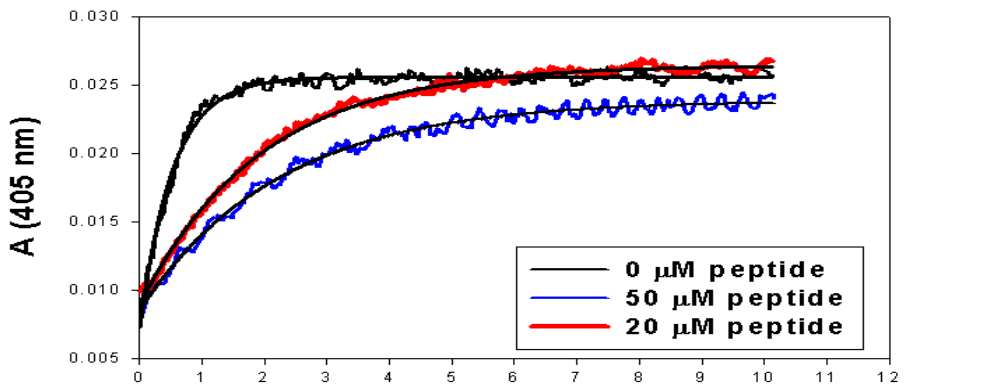
L **Ki = 16.64 µM**



[I]	k_{obs}	%Inhibition
0	0.6877	-
1 µM	0.5829	15.2
10 µM	0.2638	61.6
50 µM	0.0729	89.4

K_i = 6.25 µM

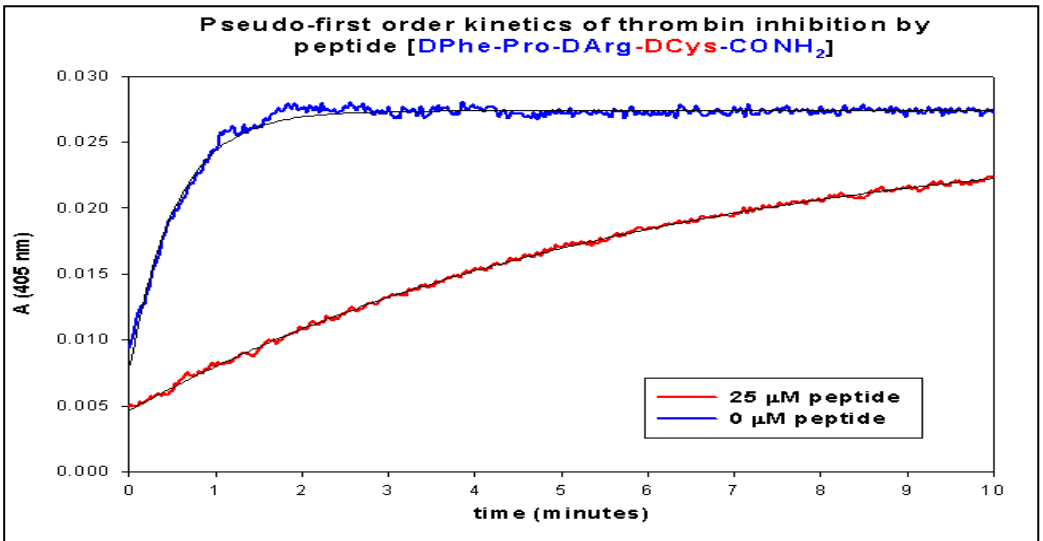
M



[I] [µM]	kobs (sec^{-1})	% Inhibition
0	1. 885 ± 0.0250	-
20	0.5147 ± 0.0039	73%
50	0.4495 ± 0.0058	77%

Ki = 11. 57 µM

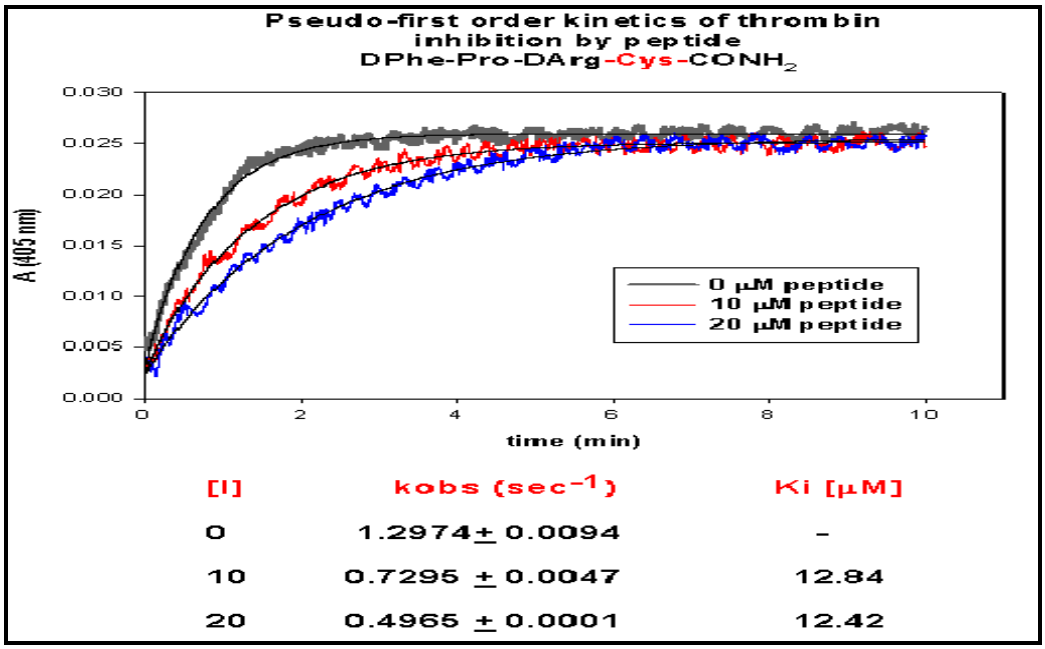
N



0 µM peptide	Kobs (min^{-1})
	1.8598
25 µM peptide	0.1679

Ki = 2.48 µM

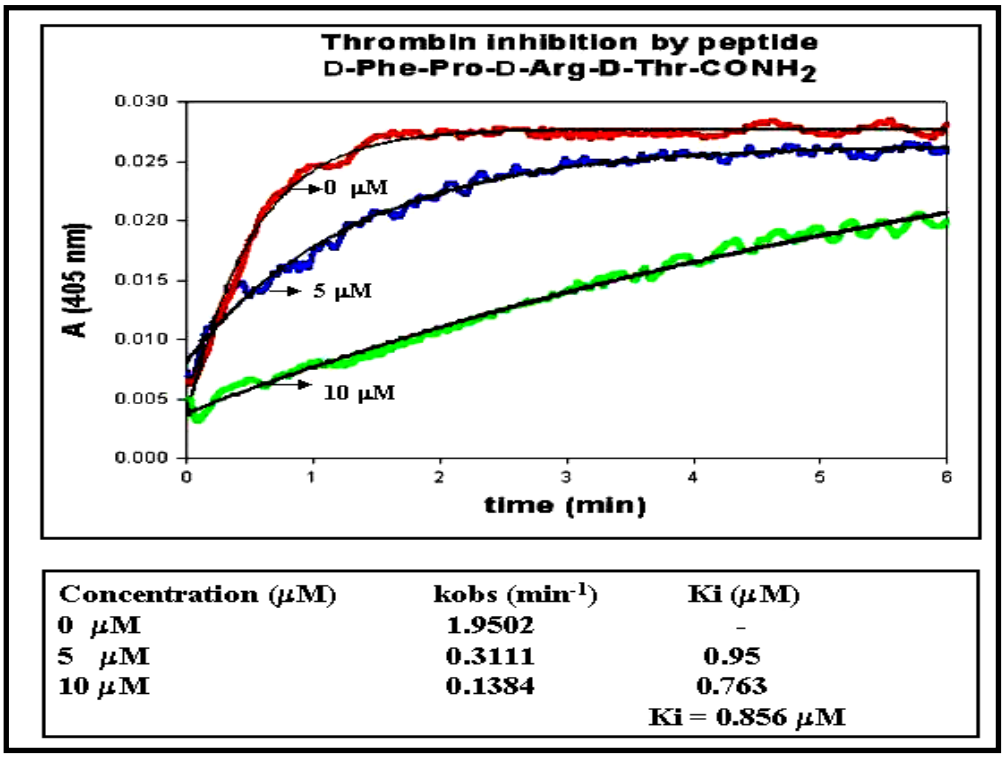
O



[I]	kobs (sec^{-1})	Ki [µM]
0	1.2974± 0.0094	-
10	0.7295 ± 0.0047	12.84
20	0.4965 ± 0.0001	12.42

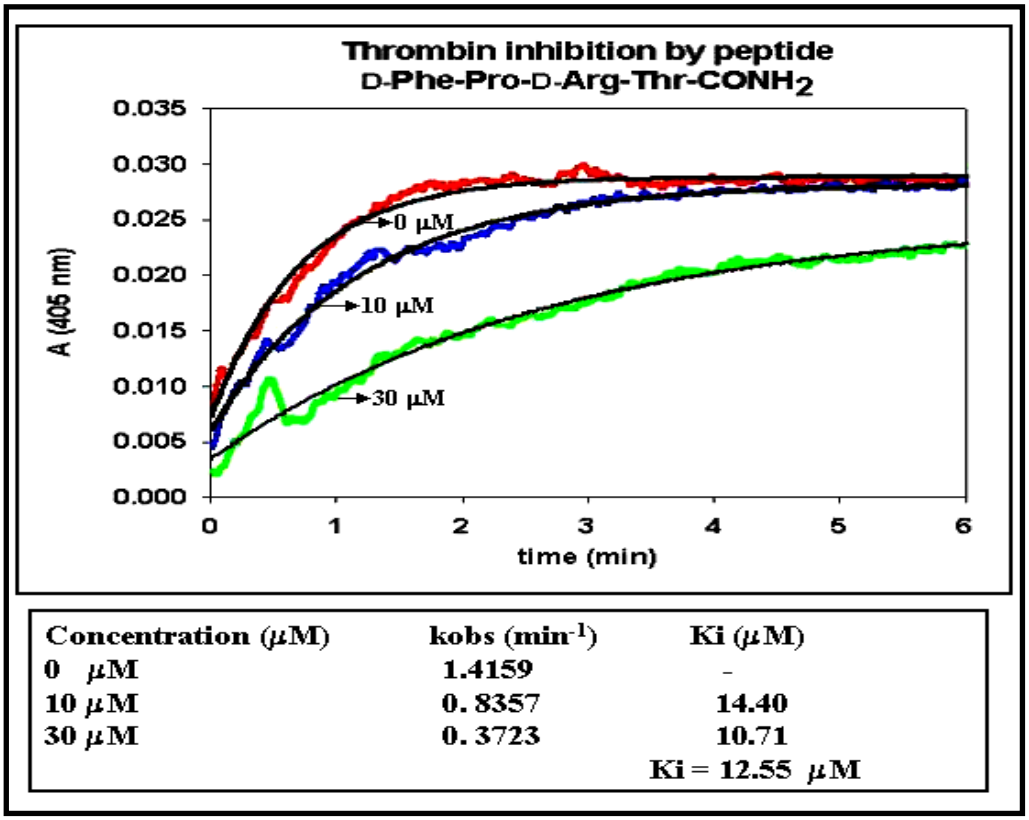
Ki = 12.4 µM

P



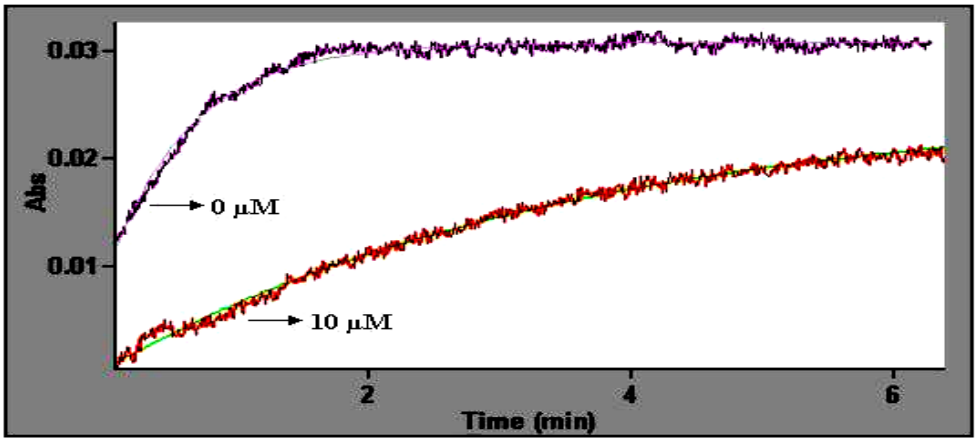
Concentration (µM)	kobs (min^{-1})	Ki (µM)
0 µM	1.9502	-
5 µM	0.3111	0.95
10 µM	0.1384	0.763
		Ki = 0.856 µM

Q



Concentration (µM)	kobs (min^{-1})	Ki (µM)
0 µM	1.4159	-
10 µM	0. 8357	14.40
30 µM	0. 3723	10.71
		Ki = 12.55 µM

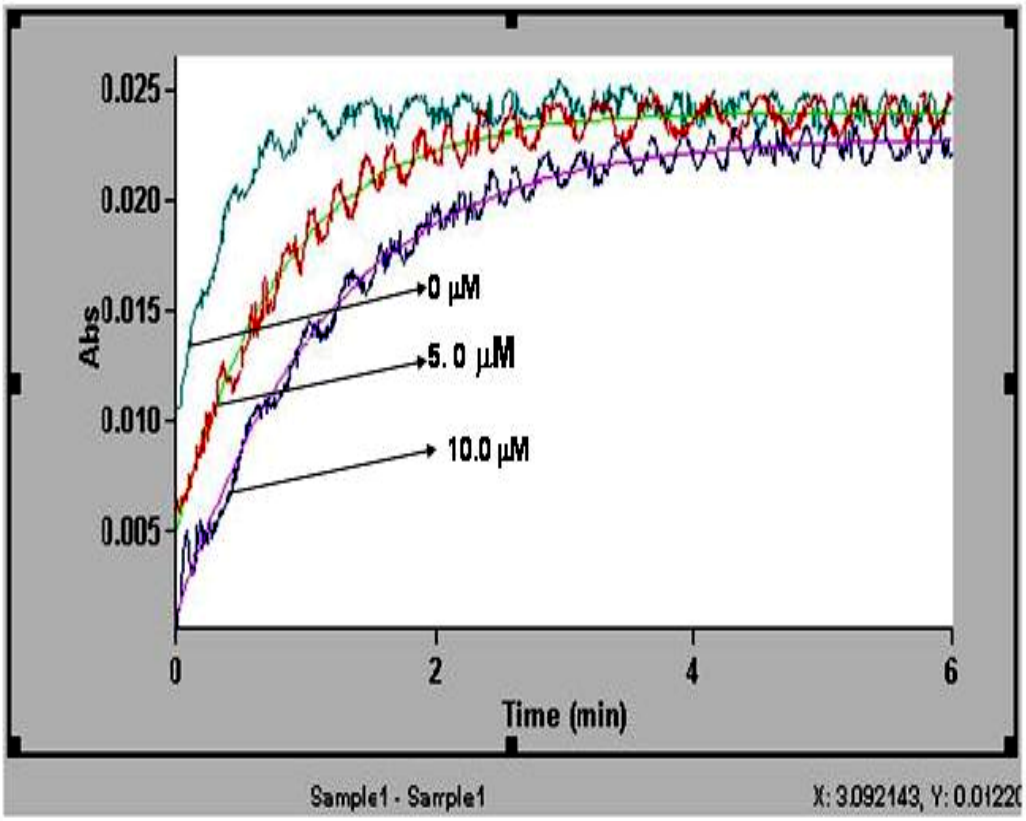
R



D-Phe-Pro-D-Arg-Gly-$CONH_2$

Concentration peptide	kobs (min^{-1})	Ki (µM)
0 µM	1.57 90	-
10 µM	0.2842	2.19 µM

S



Pseudo-first order kinetics fit of kinetic data for the thrombin inhibition by peptide [DPhe-Pro-Darg-Thi-$CONH_2$]

observed rate constant

Sample	Start (min)	Stop (min)	k (min-1)	Factor*k	A0	A inf	S.D.	
Sample1	0.000	6.000	1.2013	1.2013	-0.0025	0.0164	0.0005	0 µM peptide
Sample1	0.000	6.000	1.1973	1.1973	0.0049	0.0240	0.0007	5.0 µM peptide
	0.000	6.000	0.8473	0.8473	0.0012	0.0229	0.0006	10.0 µM peptide

Ki = 8.6 µM

Some peptides showed a variation in Ki (for most cases an increased Ki) with increasing concentration of peptide concentration in the inhibition assay mix (figure 114-115). It is well documented that many peptides are aggregating into macromolecular complexes, a process determined by ionic strength and pH (40). The reason we observed

these aggregation phenomena only in some peptides and not in all cases studies may be related with their sequence. Primarily, the high content of aromatic amino acids in some peptides sequences may enhance the aggregation through pi-pi stacking interactions between the aromatic rings. D-Phe-Pro-D-Arg-D-Pro-Phe-$CONH_2$ showed an increased K_I as the concentration of peptide was increased in the final reaction assay: at 1 μM concentration of peptide Ki was 1.88 μM; at 5 μM concentration of peptide-Ki was 93.34 μM and at 10.0 μM concentration of peptide-Ki was 169.00 μM. These results suggest that the peptide may aggregates at higher concentration which can cause a lost in affinity for thrombin and thus we obtained a higher Ki. Other peptides such as D-Phe-Pro-D-Arg-D-Pro-*Ala*-$CONH_2$ and D-Phe-Pro-D-Arg-D-Pro-*Trp*-$CONH_2$ showed the same effect, i.e. an increase in Ki with increased peptide concentration (D-Phe-Pro-D-Arg-D-Pro-*Trp*- $CONH_2$ had Ki = 64.41 μM at 20.0 μM in the inhibition assay and Ki = 107.98 μM at 50 μM peptide concentration in the inhibition assay; D-Phe-Pro-D-Arg-D-Pro-*Ala*-$CONH_2$ had a Ki = 58.66 μM at 20 μM final peptide concentration in the reaction mix and a Ki = 146.64 μM at 50 μM final peptide concentration in the reaction mix**).** For most peptides, the Ki remained constant at higher concentration of peptide and the thrombin inhibition was mostly following a competitive model (see progress curve analysis in Figure 117).

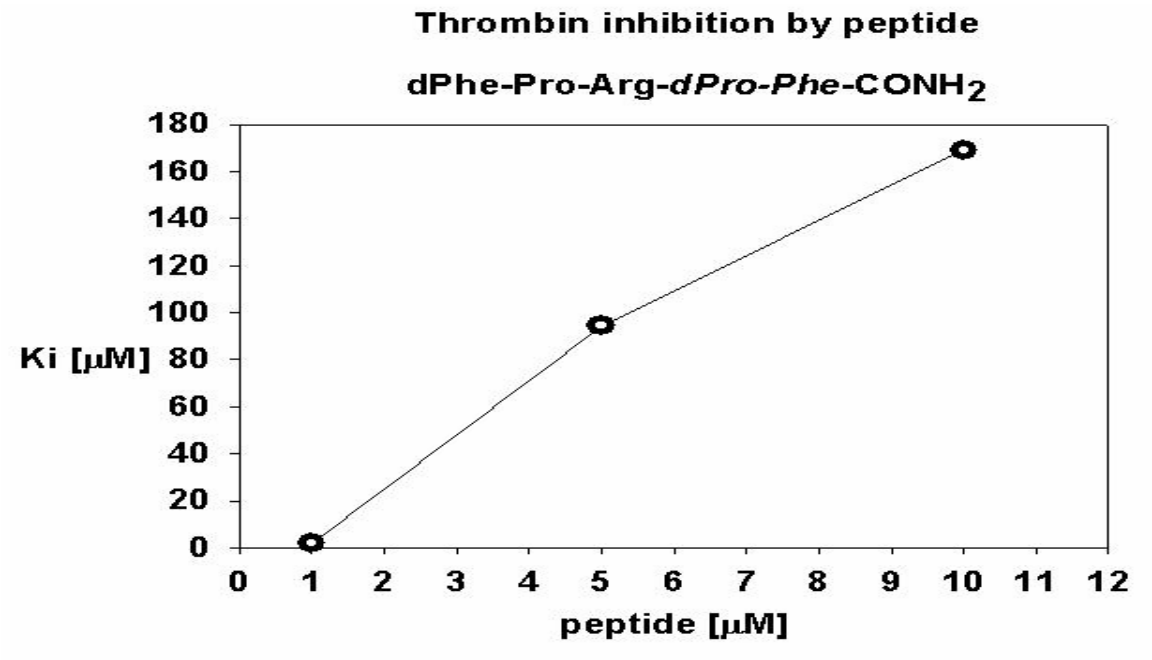


Figure 115: Variation of Ki with increasing concentration of peptide in the assay mix.

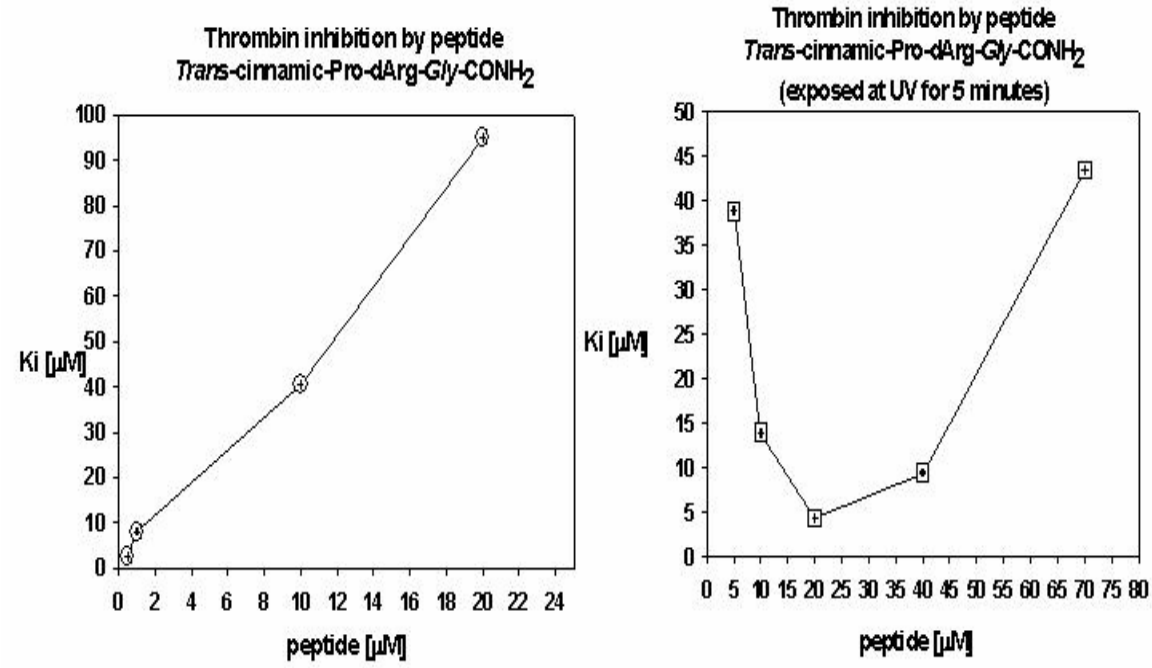


Figure 116: Variation of Ki with increasing concentration of peptide in the assay mix.

A detailed presentation of all Kis of different peptides screened for their inhibitory activity against thrombin is presented in tables 14-15.

Table 14. Hexapeptides, pentapeptides and tetrapeptides inhibitors for thrombin and their experimentally determined Kd = Ki (μM). Ki was determined from pseudo-first order reactions of thrombin inhibition as described in Materials and Method (Section 3.2.)

Peptide library N $CONH_2$	**Ki (μM)**
D-Phe-Pro-D-Arg-*Arg*	**91.09 ± 21.99**
D-Phe-Pro-D-Arg-*Ala*	**16.64 ± 0.01**
D-Phe-Pro-D-Arg-*Gly*	**5.89 ± 0.33**
D-Phe-Pro-D-Arg-*Glu*	**555.68 ± 89.77**
D-Phe-Pro-D-Arg-*His*	**122.08 ± 0.06**
D-Phe-Pro-D-Arg-*Ser*	**17.60 ± 0.62**
D-Phe-Pro-D-Arg-*Trp*	**65.55 ± 6.85**
D-Phe-Pro-D-Arg-*Tyr*	**50.5 ± 5.0**
D-Phe-Pro-D-Arg-*Pro*	**89.50 ± 3.94**
D-Phe-Pro-D-Arg-*Cys*	**12.61 ± 0.22**
D-Phe-Pro-D-Arg-*Gln*	**43.06 ± 6.78**
D-Phe-Pro-D-Arg-*Lys*	**66.56 ± 19.73**
D-Phe-Pro-D-Arg-*Ile*	**6.33 ± 0.03**
D-Phe-Pro-D-Arg-*D-Ala*	**2.06 ± 0.03**
D-Phe-Pro-D-Arg-*D-Pro*	**489.26 ±112.95**
Gly-Arg-D-Pro-Phe	**1138.95 ± 482.27**
D-Phe-Pro-D-Arg-*Val*	**56.32 ± 9.59**
D-Phe-Pro-D-Arg-*Gly-Asp*	**156.25**
D-Phe-Pro-D-Arg-*Gly-Asn*	**112.35**
D-Phe-Pro-D-Arg-*Gly-Asp-Ala*	**705.88**
D-Phe-Pro-D-Arg-*Gly-Asp-Met*	**993.4**
D-Phe-Pro-D-Arg-*Gly-Asp-Lys*	**643.77**
D-Phe-Pro-D-Arg-*Gly-Asp-Gly*	**1,730**
D-Phe-Pro-D-Arg-*Tyr-Phe*	**3,333**
D-Phe-Pro-Arg-*D-Pro-D-Phe*	**2,362**

D-Phe-Pro-Arg-*D-Pro-D-Arg*	**980.4**
D-Phe-Pro-Arg-*D-Pro D-Tyr*	**3,639**

Table 15. Tetrapeptide library containing natural and unnatural amino acids analogs and their experimentally determined Kd = Ki (μM).

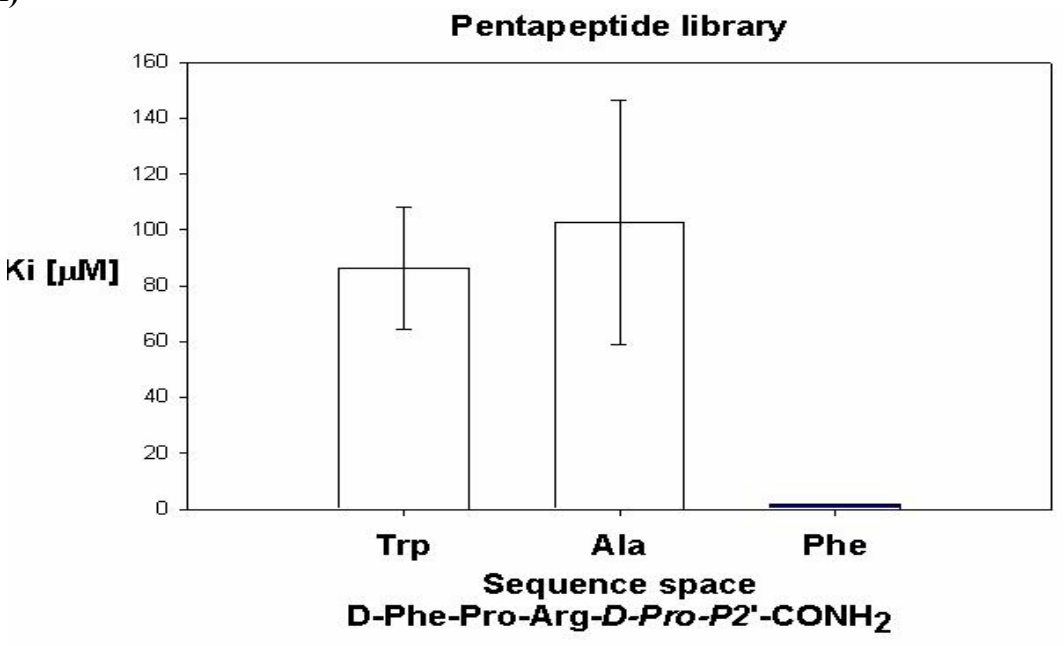
	Sequence N → C	Ki (dissociation constant) (μM)	Ka (apparent association constant) = 1/Ki (L/Mol)
1	**L-Tic-Pro-DArg-Cys-$CONH_2$**	**1,050**	**0.952×10^3**
2	**L-Tic-L-Thi-DArg-$CONH_2$**	**562.5**	**1.77×10^3**
3	**L-Tic-Pro-DArg-DAla-$CONH_2$**	**121.25**	**8.24×10^3**
4	**L-Thi-Pro-DArg-Gly-$CONH_2$**	**175**	**5.71×10^3**
5	**L-Tic-Pro-D-Arg-Thi-$CONH_2$**	**312.7**	**3.19×10^3**
6	**D-Phe-Pro-D-Arg-DCys-$CONH_2$**	**2.4**	**416.6×10^3**
7	**D-Phe-Pro-D-Arg-Cys-$CONH_2$**	**12.51**	**80.0×10^3**
8	**D-Phe-Pro-D-Arg-Met-$CONH_2$**	**37**	**27.03×10^3**
9	**D-Phe-Pro-D-Arg-DThr-$CONH_2$**	**0.940**	**1063.8×10^3**
10	**D-Phe-Pro-D-Arg-Thr-$CONH_2$**	**12.5**	**80.0×10^3**
11	**D-Phe-Pro-D-Arg-DSer-$CONH_2$**	**12.3**	**81.3×10^3**
12	**D-Phe-Pro-D-Arg-Ser-$CONH_2$**	**17.06**	**58.6×10^3**
13	**D-Phe-Pro-D-Arg-Gly- $CONH_2$**	**5.89**	**169.7×10^3**
14	**D-Phe-Pro-D-Arg-Thi-$CONH_2$**	**8.16**	**122.5×10^3**
15	**D-Phe-Pro-D-Arg-DGln-$CONH_2$**	**19.31**	**51.8×10^3**
16	**D-Phe-Pro-D-Arg-Gln-$CONH_2$**	**43.06**	**23.22×10^3**
17	**D-Phe-Pro-D-Arg-Ala- $CONH_2$**	**16.64**	**60.1×10^3**
18	**D-Phe-Pro-D-Arg-DAla- $CONH_2$**	**2.06**	**485.4×10^3**

19	**D-Naphthylalanine(Nal)-Pro-DArg-$CONH_2$**	**11.62**	**86.05 x 10^3**
20	**D-Naphthylala(Nal)-Pro-DArg-Gly-$CONH_2$**	**2.71**	**369.01 x 10^3**
21	**D-tetrahydroharman(D-Tpi)-Pro-DArg-D-Ala**	**520.5**	**1.92 x 10^3**
22	**D-tetrahydroharman(DTpi)-Pro-D-Arg-Ile**	**147.5**	**6.77 x 10^3**
23	**Transcinnamic-Pro-D-Arg-Gly-$CONH_2$**	**6.23**	**160.5 x10^3**
24	**Ciscinnamic-Pro-D-Arg-Gly- $CONH_2$**	**31.5**	**31.7 x 10^3**
25	**Transcinnamic-Pro-D-Arg-Ile-$CONH_2$**	**161.8**	**6.18 x 10^3**
26	**Ciscinnamic-Pro-D-Arg-Ile-$CONH_2$**	**53.6**	**18.6 x 10^3**
27	**Transcinnamic-Pro-D-Arg-Ser-$CONH_2$**	**83**	**12.05 x 10^3**
28	**Ciscinnamic-Pro-D-Arg-Ser-$CONH_2$**	**117.6**	**8.5 x 10^3**
29	**Transcinnamic-Pro-D-Arg-Cys-$CONH_2$**	**30.5**	**32.8 x 10^3**
30	**Ciscinnamic-Pro-D-Arg-Cys-$CONH_2$**	**394**	**2.5 x 10^3**
31	**Transcinnamic-Pro-D-Arg-D-Ala-$CONH_2$**	**44.6**	**22.42 x 10^3**
32	**Ciscinnamic-Pro-D-Arg-D-Ala-$CONH_2$**	**69.4**	**14.4 x 10^3**
33	**Transcinnamic-Pro-D-Arg-$CONH_2$**	**240**	**4.1 x10^3**
34	**Dihydrocinnamic-Pro-D-Arg-Cys-$CONH_2$**	**64**	**15.6 x 10^3**
35	**Dihydrocinnamic-Pro-D-Arg-Gly-$CONH_2$**	**105.8**	**9.45 x 10^3**
36	**D-Tic-Pro-D-Arg-Gly-$CONH_2$**	**171.2**	**5.84 x 10^3**

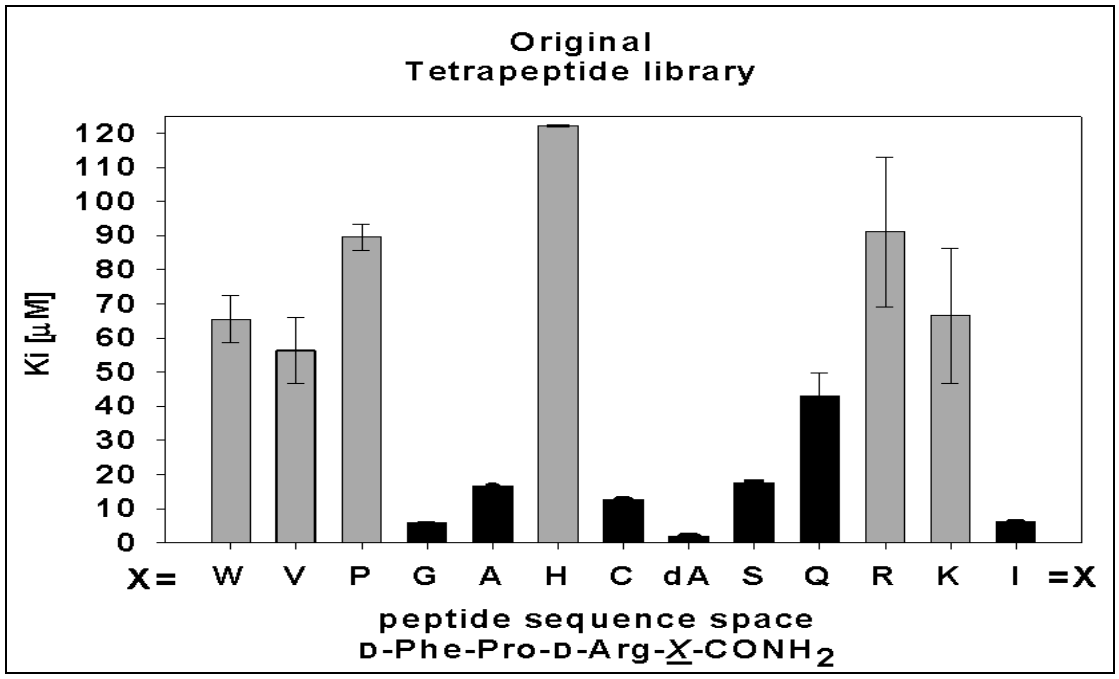
The structure-activity relationships presented in Tables 14 and 15 showed that the amino-acid at P1' position was a major factor in determining the inhibitory potential of each peptide for thrombin. Also, the P3 position within the peptide ligand was an important factor in determining the inhibitory potential and only D-naphthylalanine (D-Nal) was showing at least two fold higher inhibitory activity than D-Phe when Gly was occupying P1' position while P2 (L-Pro) and P1 (D-Arg) remained constant. Figure 117 (A-Q) present a comparison between different peptide sequences with respect to their inhibitory activity against thrombin contrasting the P1' and P3 positions with respect to their effect upon the Ki (inhibitory constant) (the 1/Ki = Kassociation is presented for some graphs). In addition the screening of pentapeptide libraries with the sequence D-Phe-Pro-Arg-D-Pro-P2'-CONH2 is presented for the SAR at P2' (Figure 117 (A)).

Figure 117 (A-S). Comparison of SAR for peptides with different sequence space where containing single amino acid variations (for details related to the Ki of each peptide refer to tables 14 and 15).

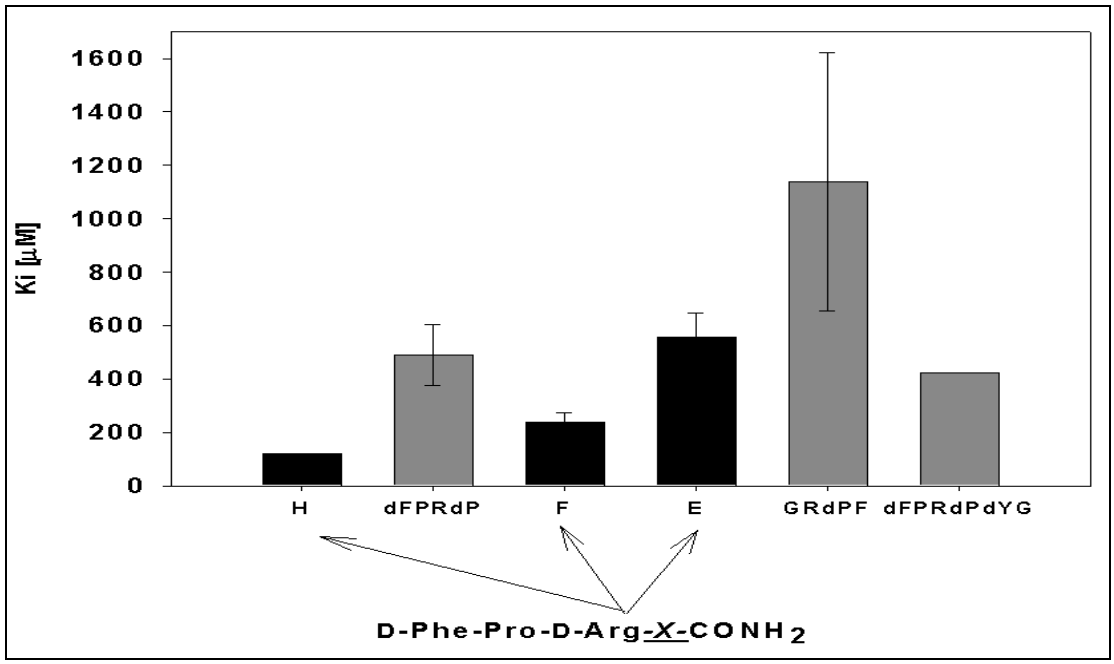
(A)



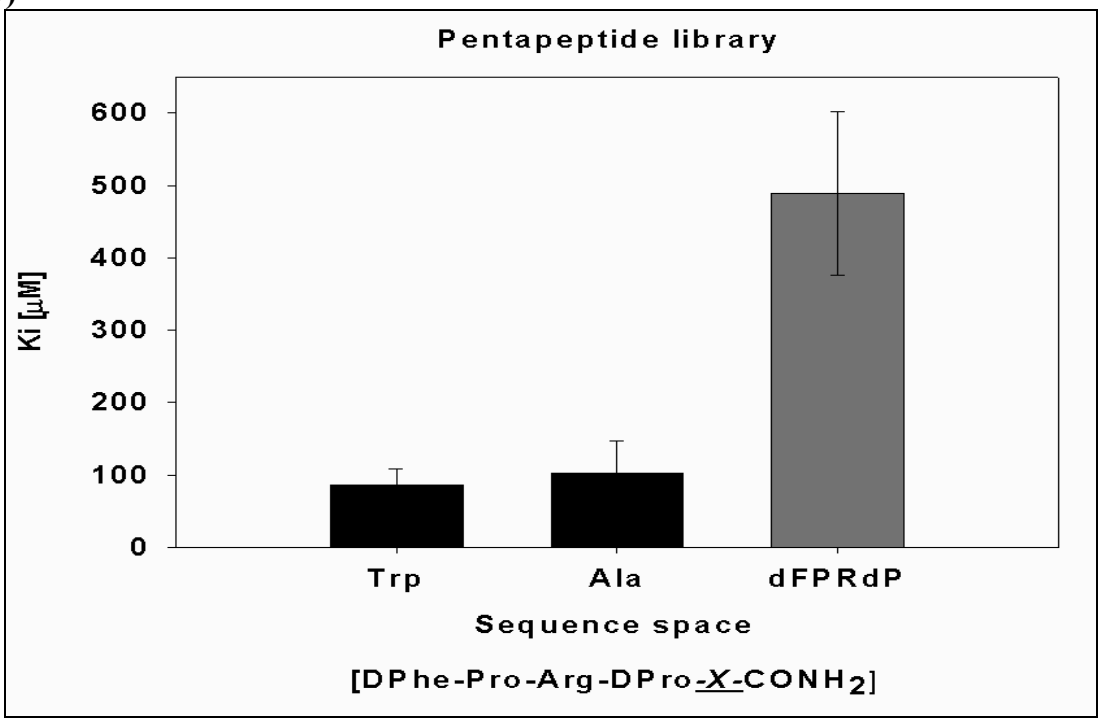
(B)


Original
Tetrapeptide library
Ki [µM]
120
110
100
90
80
70
60
50
40
30
20
10
0
X= W V P G A H C dA S Q R K I =X
peptide sequence space
D-Phe-Pro-D-Arg-X-CONH2

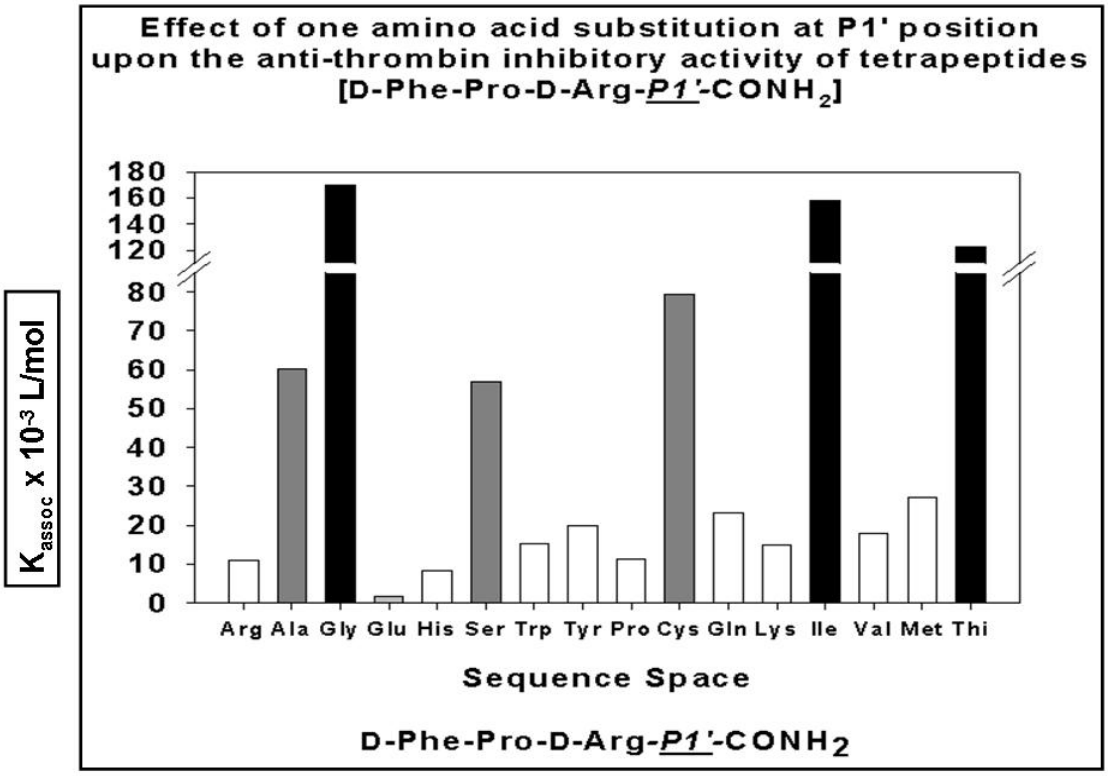
(C)


Ki [µM]
1600
1400
1200
1000
800
600
400
200
0
H
dFPRdP
F
E
GRdPF
dFPRdPdYG
D-Phe-Pro-D-Arg-X-CONH2

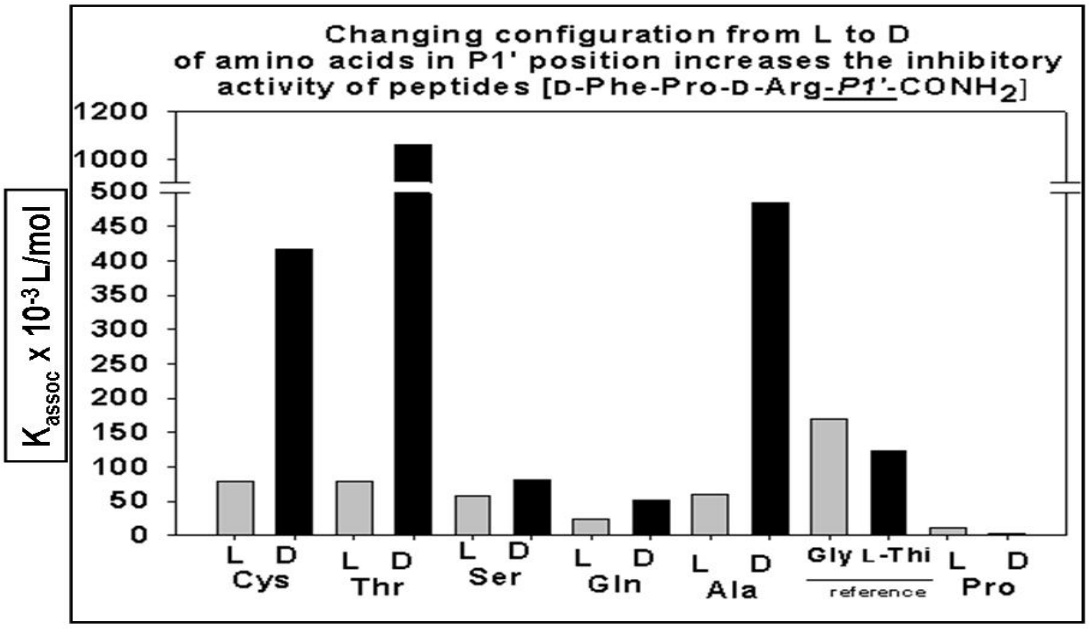
(D)


Pentapeptide library
600
500
400
300
200
100
0
Ki [μM]
Trp
Ala
dFPRdP
Sequence space
[DPhe-Pro-Arg-DPro-X-CONH2]

(E)

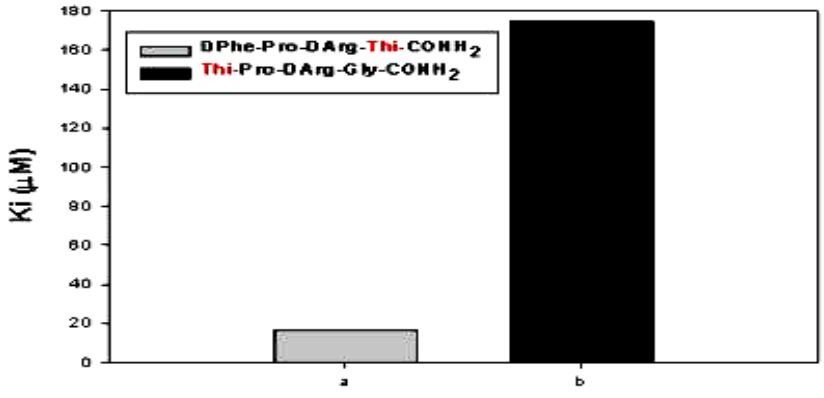

Effect of one amino acid substitution at P1' position
upon the anti-thrombin inhibitory activity of tetrapeptides
[D-Phe-Pro-D-Arg-P1'-CONH2]
180
160
140
120
80
70
60
50
40
30
20
10
0
Kassoc x 10-3 L/mol
Arg Ala Gly Glu His Ser Trp Tyr Pro Cys Gln Lys Ile Val Met Thi
Sequence Space
D-Phe-Pro-D-Arg-P1'-CONH2

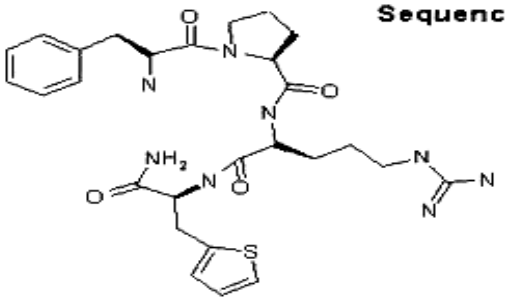
(F)



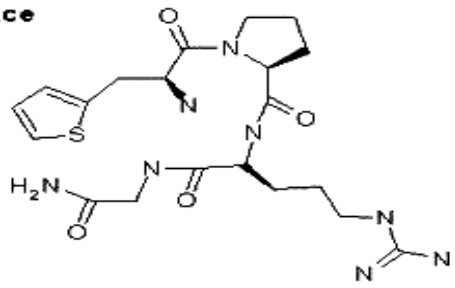
(G)

L-Thi (thienyl-ala) scanning on positions P3 and P1'



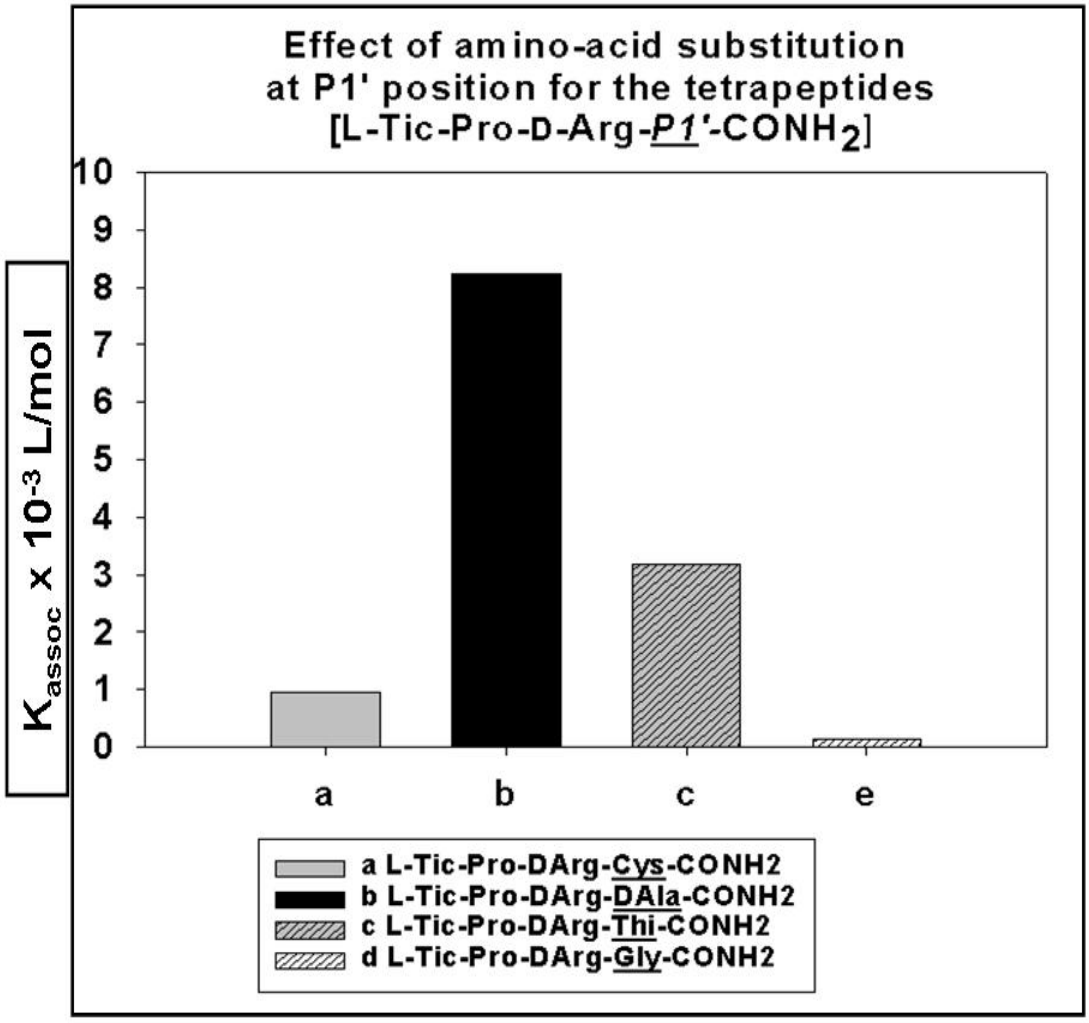


(1) Dphe-Pro-DArg-Thi-$CONH_2$

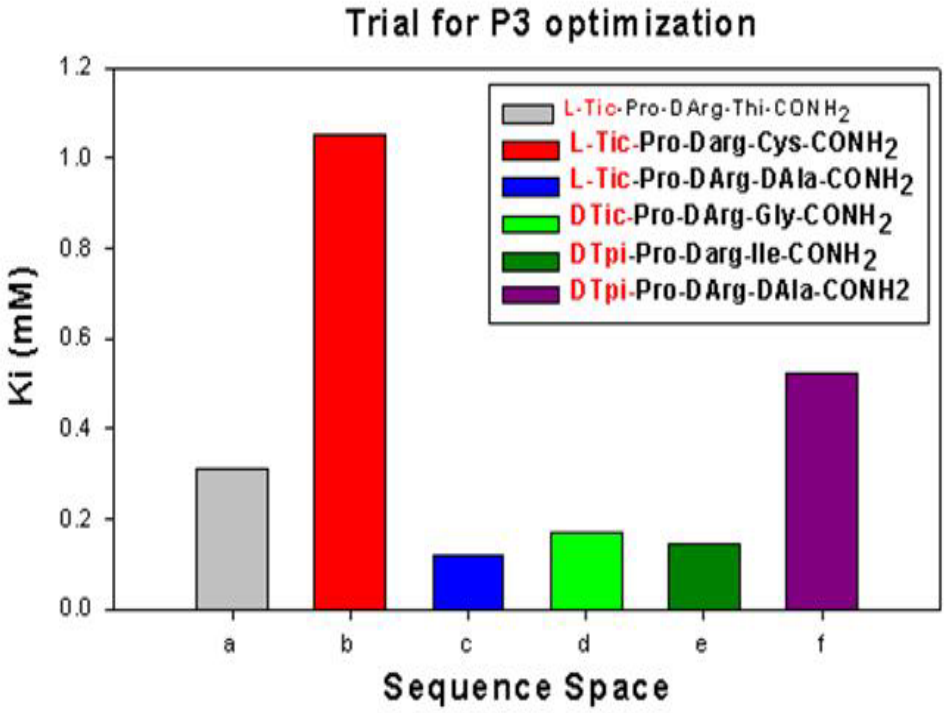


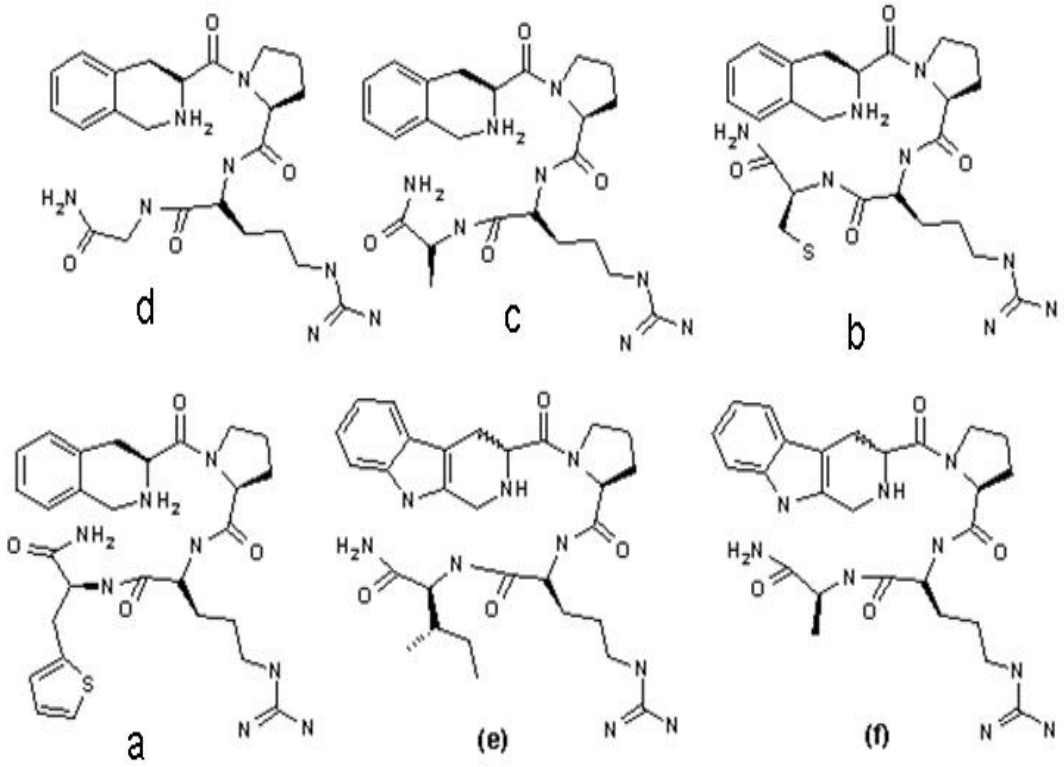
(2) Thi-Pro-DArg-Gly-$CONH_2$

(H)

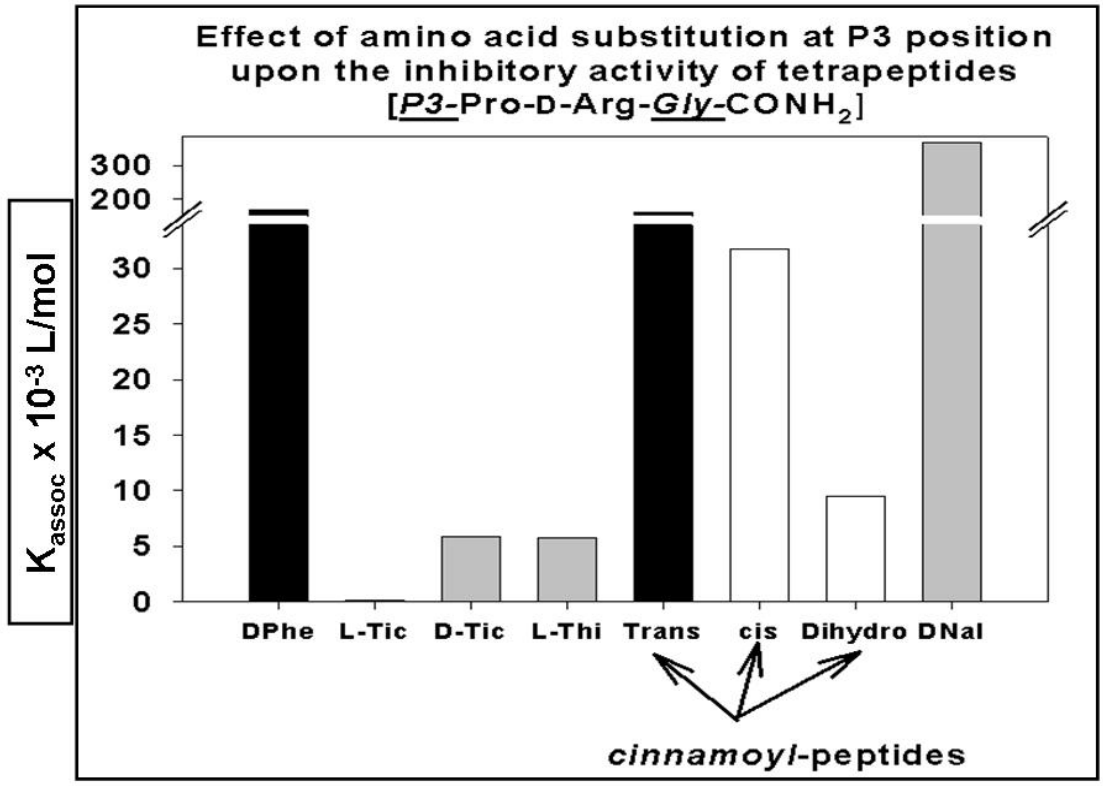

Effect of amino-acid substitution
at P1' position for the tetrapeptides
[L-Tic-Pro-D-Arg-P1'-CONH2]
Kassoc x 10-3 L/mol
10
9
8
7
6
5
4
3
2
1
0
a
b
c
e
a L-Tic-Pro-DArg-Cys-CONH2
b L-Tic-Pro-DArg-DAla-CONH2
c L-Tic-Pro-DArg-Thi-CONH2
d L-Tic-Pro-DArg-Gly-CONH2

(I)

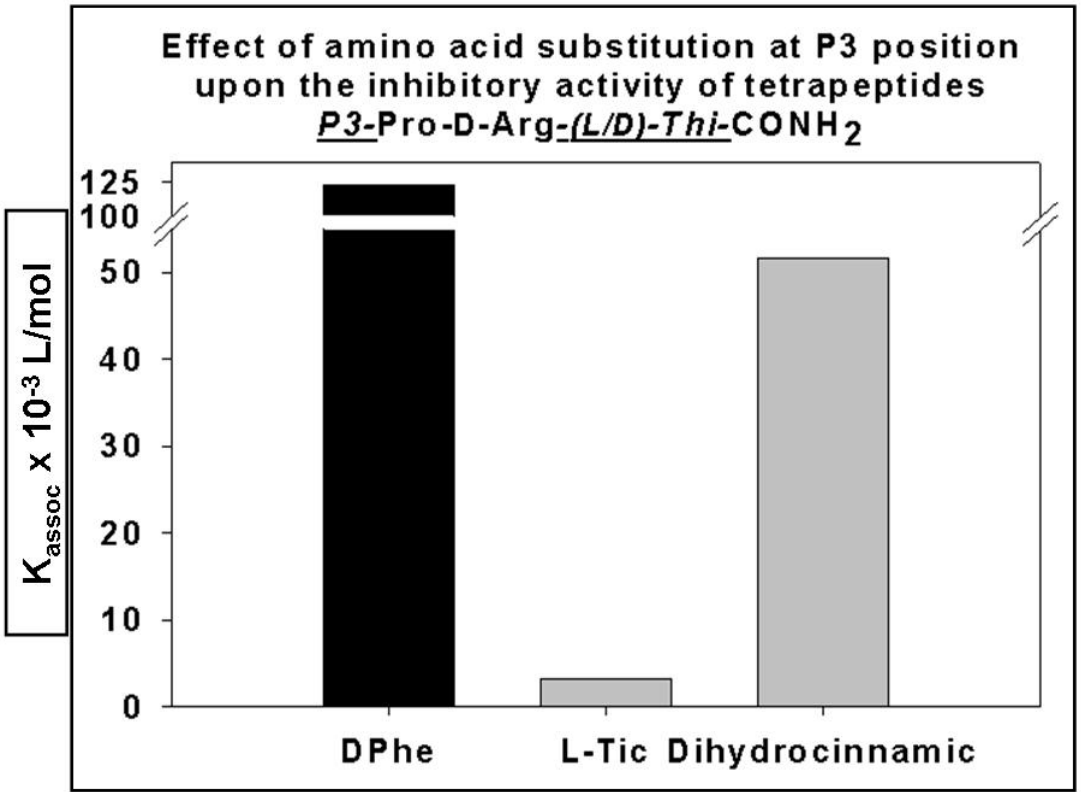

Trial for P3 optimization
Ki (mM)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
a
b
c
d
e
f
Sequence Space
L-Tic-Pro-DArg-Thi-CONH2
L-Tic-Pro-Darg-Cys-CONH2
L-Tic-Pro-DArg-DAla-CONH2
DTic-Pro-DArg-Gly-CONH2
DTpi-Pro-Darg-Ile-CONH2
DTpi-Pro-DArg-DAla-CONH2


d
c
b
a
(e)
(f)

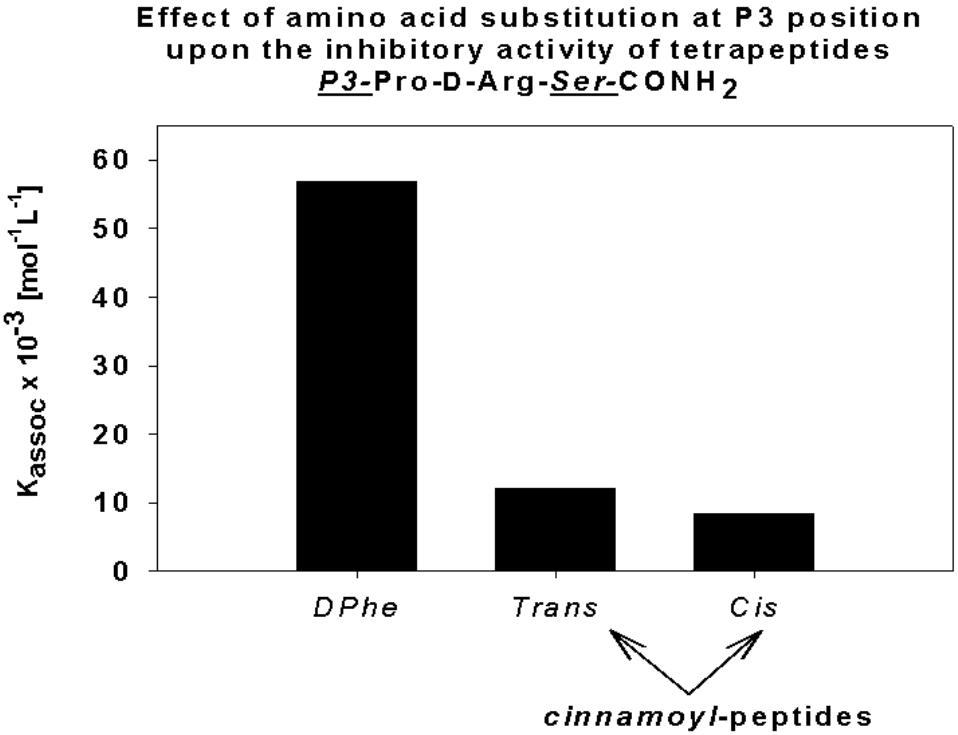
(J)


Effect of amino acid substitution at P3 position
upon the inhibitory activity of tetrapeptides
[P3-Pro-D-Arg-Gly-CONH2]
Kassoc x 10-3 L/mol
300
200
30
25
20
15
10
5
0
DPhe
L-Tic
D-Tic
L-Thi
Trans
cis
Dihydro
DNal
cinnamoyl-peptides

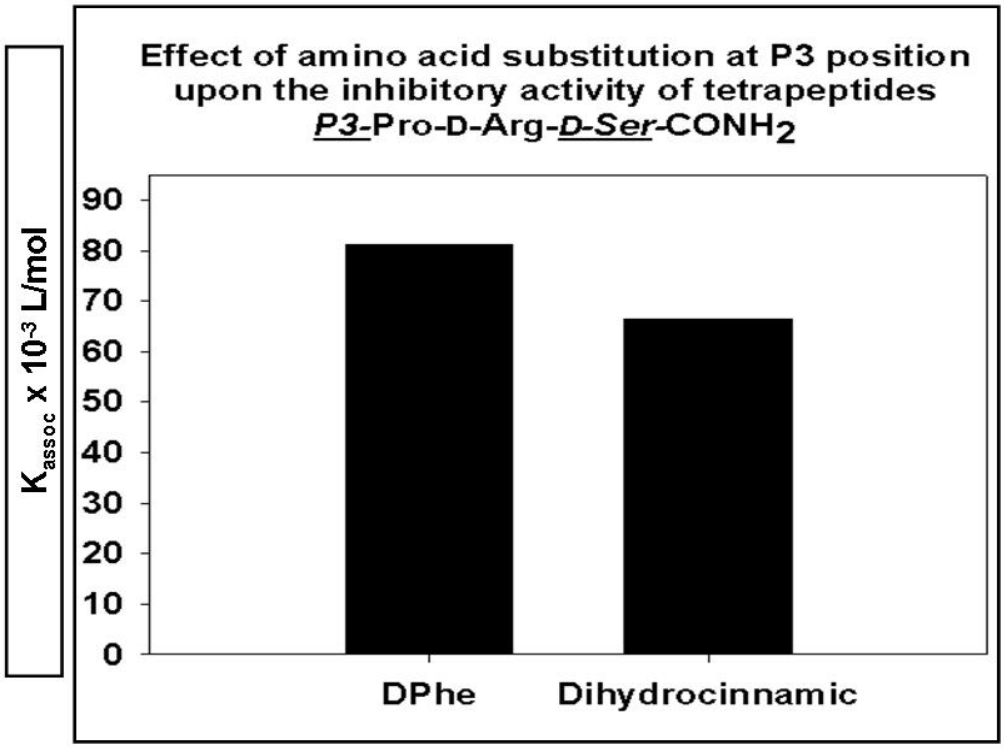
(K)


Effect of amino acid substitution at P3 position
upon the inhibitory activity of tetrapeptides
P3-Pro-D-Arg-(L/D)-Thi-CONH2
Kassoc x 10-3 L/mol
125
100
50
40
30
20
10
0
DPhe
L-Tic
Dihydrocinnamic

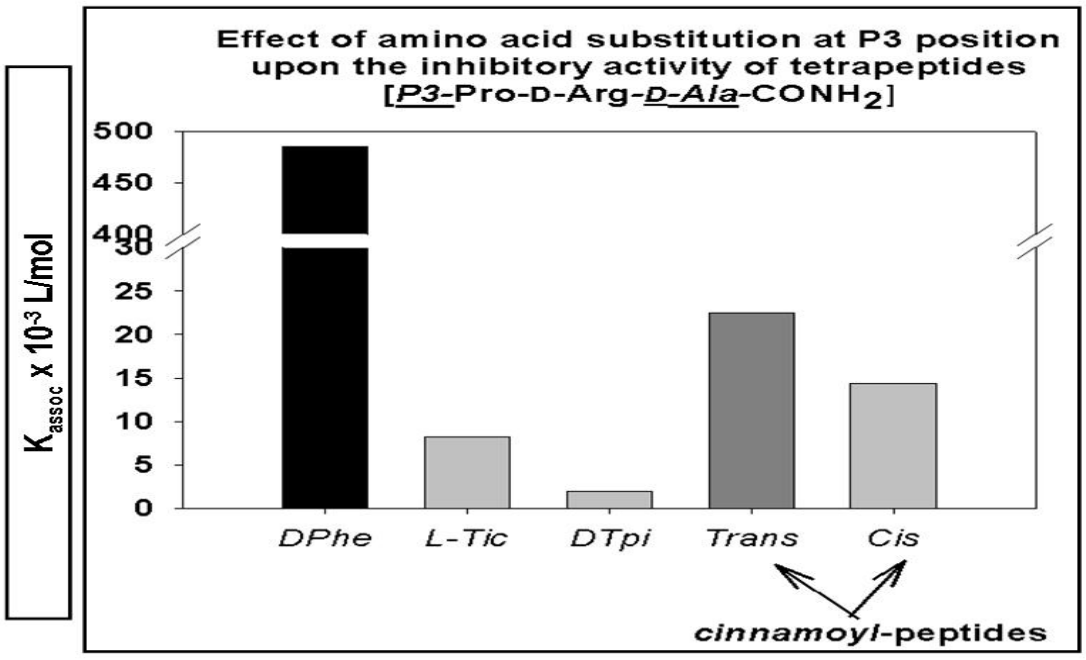
(L)


Effect of amino acid substitution at P3 position
upon the inhibitory activity of tetrapeptides
P3-Pro-D-Arg-Ser-CONH2
Kassoc x 10-3 [mol-1L-1]
60
50
40
30
20
10
0
DPhe
Trans
Cis
cinnamoyl-peptides

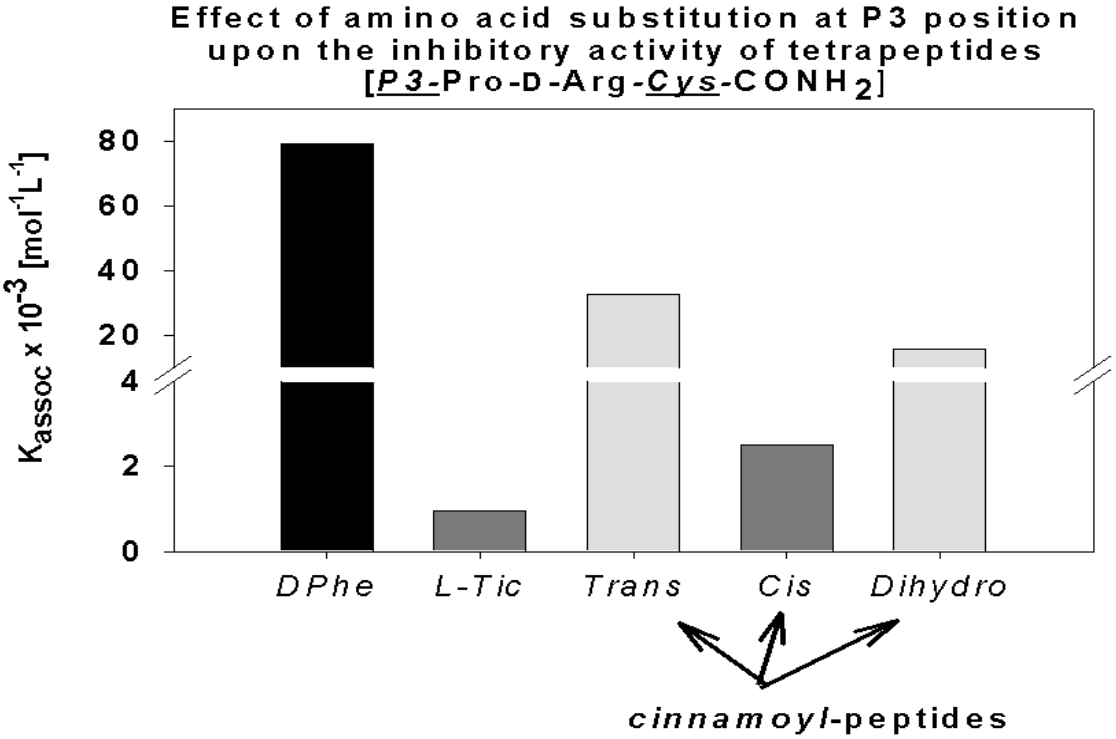
(M)


Effect of amino acid substitution at P3 position
upon the inhibitory activity of tetrapeptides
P3-Pro-D-Arg-D-Ser-CONH2
Kassoc x 10-3 L/mol
90
80
70
60
50
40
30
20
10
0
DPhe
Dihydrocinnamic

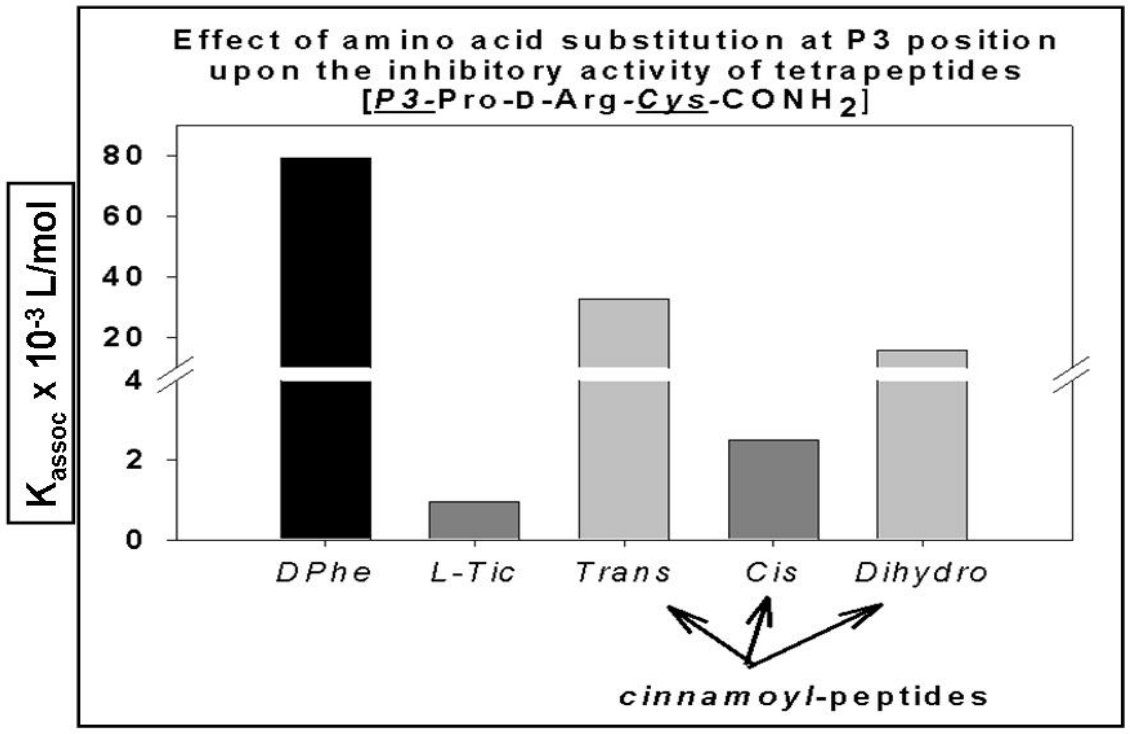
(N)


Effect of amino acid substitution at P3 position
upon the inhibitory activity of tetrapeptides
[P3-Pro-D-Arg-D-Ala-CONH2]
Kassoc x 10-3 L/mol
500
450
400
30
25
20
15
10
5
0
DPhe
L-Tic
DTpi
Trans
Cis
cinnamoyl-peptides

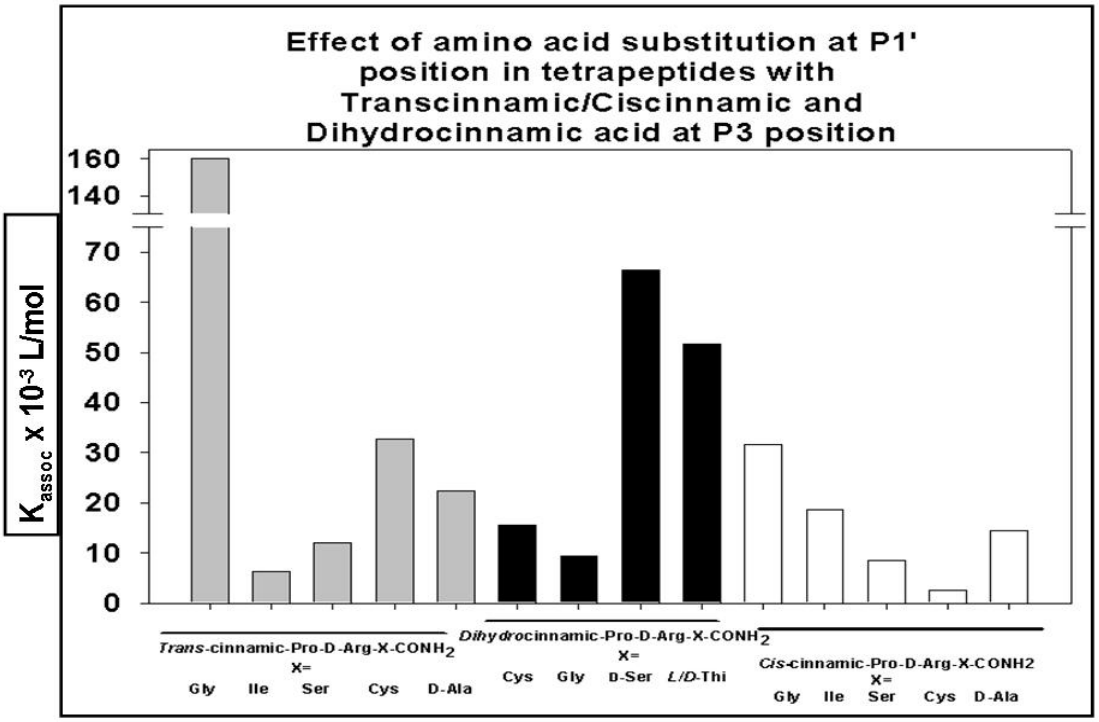
(O)


Effect of amino acid substitution at P3 position
upon the inhibitory activity of tetrapeptides
[P3-Pro-D-Arg-Cys-CONH2]
Kassoc x 10-3 [mol-1L-1]
80
60
40
20
4
2
0
DPhe
L-Tic
Trans
Cis
Dihydro
cinnamoyl-peptides

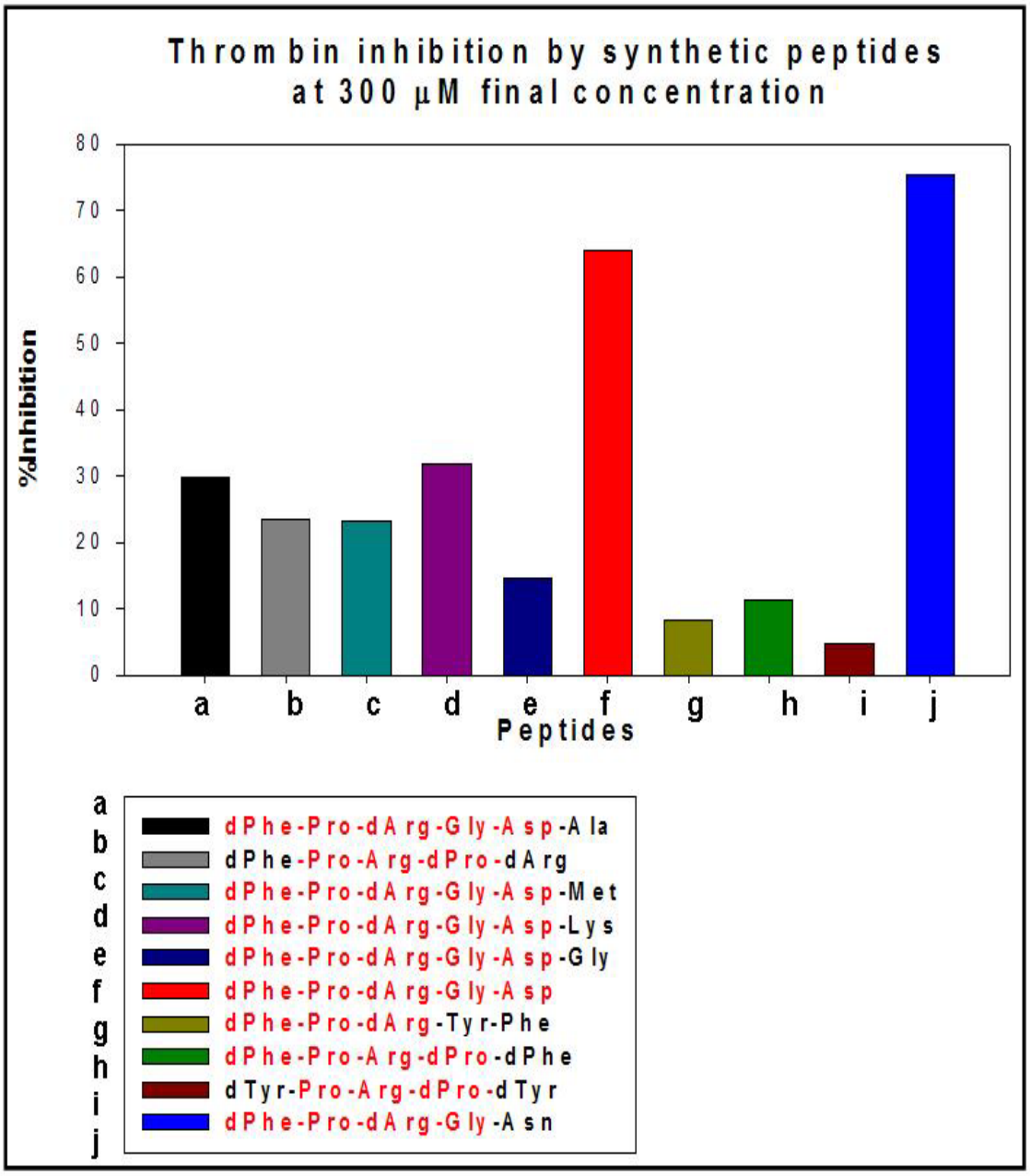
(P)


Effect of amino acid substitution at P3 position
upon the inhibitory activity of tetrapeptides
[P3-Pro-D-Arg-Cys-CONH2]
Kassoc x 10-3 L/mol
80
60
40
20
4
2
0
DPhe
L-Tic
Trans
Cis
Dihydro
cinnamoyl-peptides

(Q)


Effect of amino acid substitution at P1'
position in tetrapeptides with
Transcinnamic/Ciscinnamic and
Dihydrocinnamic acid at P3 position
Kassoc x 10-3 L/mol
160
140
70
60
50
40
30
20
10
0
Trans-cinnamic-Pro-D-Arg-X-CONH2
X=
Gly
Ile
Ser
Cys
D-Ala
Dihydrocinnamic-Pro-D-Arg-X-CONH2
X=
Cys
Gly
D-Ser
L/D-Thi
Cis-cinnamic-Pro-D-Arg-X-CONH2
X=
Gly
Ile
Ser
Cys
D-Ala

(S)

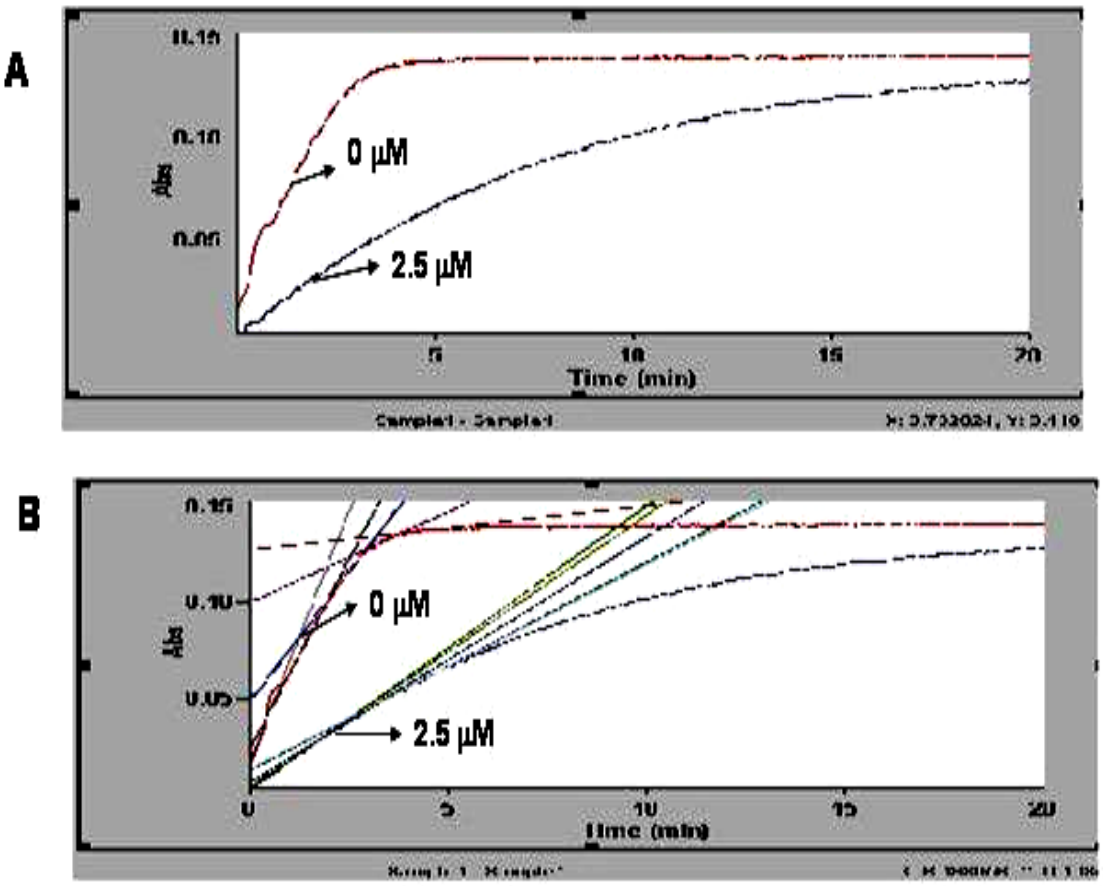


In order to determine the mechanism of thrombin inhibition by different lead peptides progress curve analysis was performed for all kinetics of inhibition run at 3Km or 5Km of substrate S2238 concentration (saturation conditions) (**Figure 118).** The kinetics of

thrombin inhibition were run at 1 ml final volume, in the buffer for thrombin assay (0.1 I (ionic strength) potassium phosphate buffer, pH 7.48, containing 0.2 M NaCl and 50 μg/ml BSA as additive) as described in Materials and Methods on Section 3.2. Eadie-Hofstee or Lineweaver Burk plots were performed for all major lead compounds in order to determine the effect upon Km and Vmax and the models for thrombin inhibition were assumed to be competitive at low concentration of peptide or non-competitive or uncompetitive at higher concentration of peptide (> 10 μM) as presented in Materials and Methods (Section 3.2).

A) Progress curve analysis of thrombin inhibition by peptide

[D-Phe-Pro-D-Arg-*D-Ala*-$CONH_2$]



Zero Order Result Table
Factor 1.00

0 µM inhibitor

Sample	Start (min)	Stop (min)	Slope (Abs/min)	Factor* Slope	A Start	A Stop	S.D.
Sample1	0.000	1.000	0.0520	0.0520	0.0151	0.0627	0.0039
Sample1	1.000	2.000	0.0394	0.0394	0.0627	0.1013	0.0004
Sample1	2.000	3.000	0.0269	0.0269	0.1013	0.1275	0.0009
Sample1	3.000	4.000	0.0094	0.0094	0.1275	0.1363	0.0006
Sample1	4.000	5.000	0.0023	0.0023	0.1363	0.1382	0.0004

2.5 µM inhibitor

	0.000	1.000	0.0148	0.0148	0.0025	0.0180	0.0009
Sample1	1.000	2.000	0.0141	0.0141	0.0180	0.0318	0.0005
Sample1	2.000	3.000	0.0127	0.0127	0.0318	0.0449	0.0005
Sample1	3.000	4.000	0.0108	0.0108	0.0449	0.0558	0.0003
Sample1	4.000	5.000	0.0106	0.0106	0.0558	0.0663	0.0003

B)

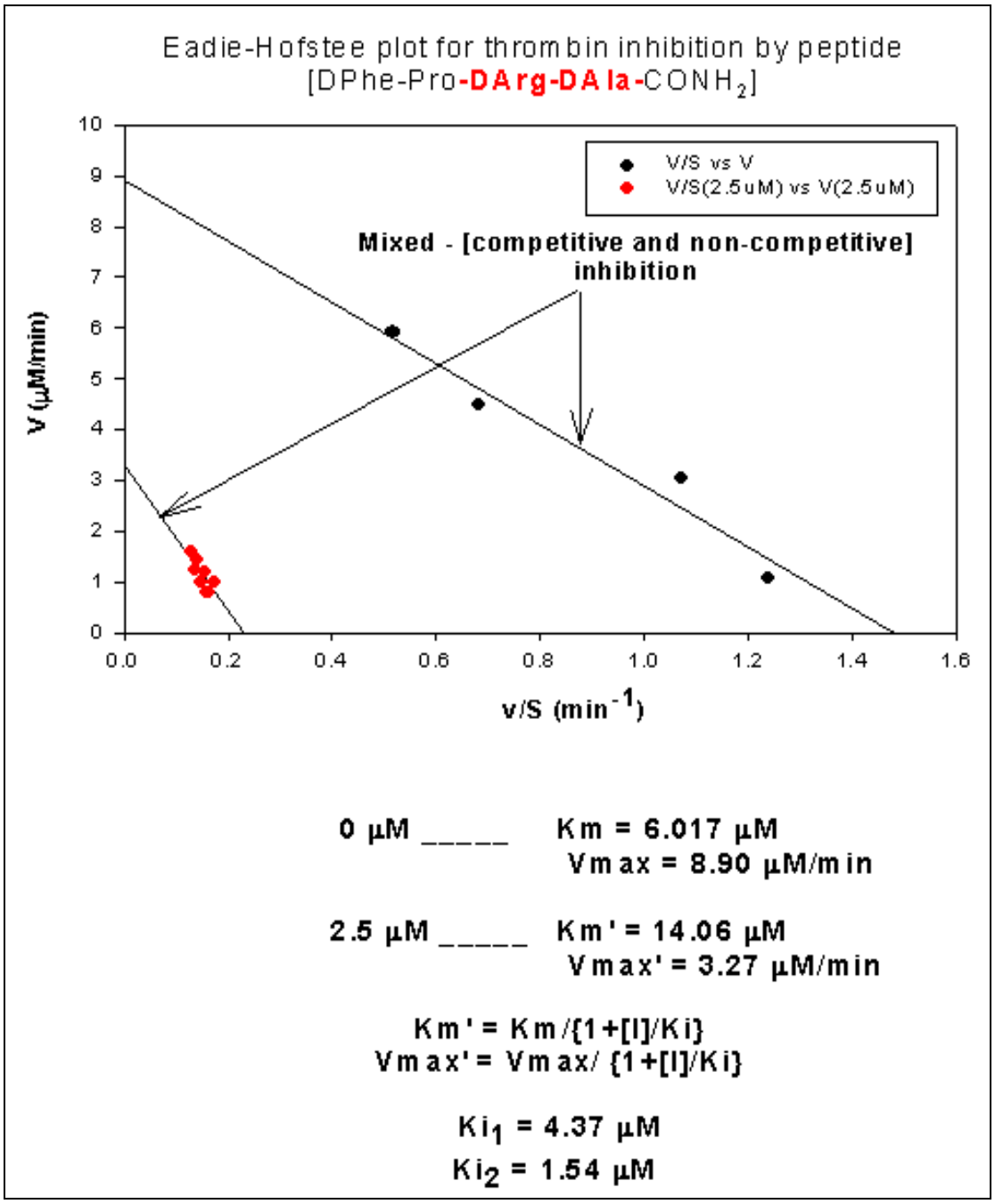
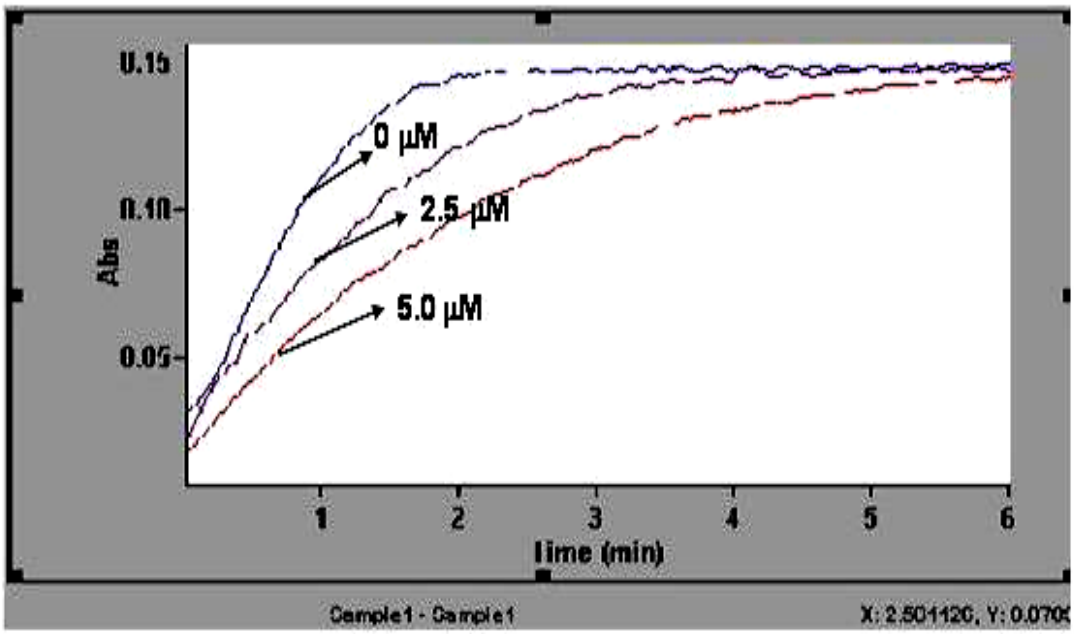


Figure 118 (A-B). Progress curve analysis of kinetics for thrombin inhibition run at saturation concentration of substrate S2238 in the presence of different peptide inhibitors. The concentration of substrate and the velocity on different points of each curve were determined as described in Materials and Methods and presented as Eadie-

Hofstee or Lineweaver-Burk plots. The vmax and Km were calculated for each inhibitor based on the inhibition plots.

C) Progress curve analysis of thrombin inhibition by peptide [D-Phe-Pro-D-Arg-*D-Cys*-$CONH_2$]



Zero Order Result Table
Factor 1.00

Sample	Start (min)	Stop (min)	Slope (Abs/min)	Factor* Slope	A Start	A Stop	S.D.
0 µM peptide inhibitors							
Sample1	0.000	1.000	0.0884	0.0884	0.0321	0.1114	0.0015
Sample1	1.000	2.000	0.0338	0.0338	0.1114	0.1458	0.0024
Sample1	2.000	3.000	0.0020	0.0020	0.1458	0.1475	0.0005
Sample1	3.000	4.000	0.0003	0.0003	0.1475	0.1477	0.0004
Sample1	4.000	5.000	0.0003	0.0003	0.1477	0.1489	0.0005
2.5 µM peptide inhibitors							
	0.000	1.000	0.0599	0.0599	0.0234	0.0833	0.0018
Sample1	1.000	2.000	0.0385	0.0385	0.0833	0.1216	0.0012
Sample1	2.000	3.000	0.0180	0.0180	0.1216	0.1390	0.0008
Sample1	3.000	4.000	0.0050	0.0050	0.1390	0.1450	0.0007
Sample1	4.000	5.000	0.0000	0.0000	0.1450	0.1466	0.0009
5.0 µM peptide inhibitors							
	0.000	1.000	0.0494	0.0494	0.0182	0.0650	0.0007
Sample1	1.000	2.000	0.0318	0.0318	0.0650	0.0979	0.0008
Sample1	2.000	3.000	0.0222	0.0222	0.0979	0.1210	0.0007
Sample1	3.000	4.000	0.0140	0.0140	0.1210	0.1341	0.0007
Sample1	4.000	5.000	0.0072	0.0072	0.1341	0.1408	0.0004

D)

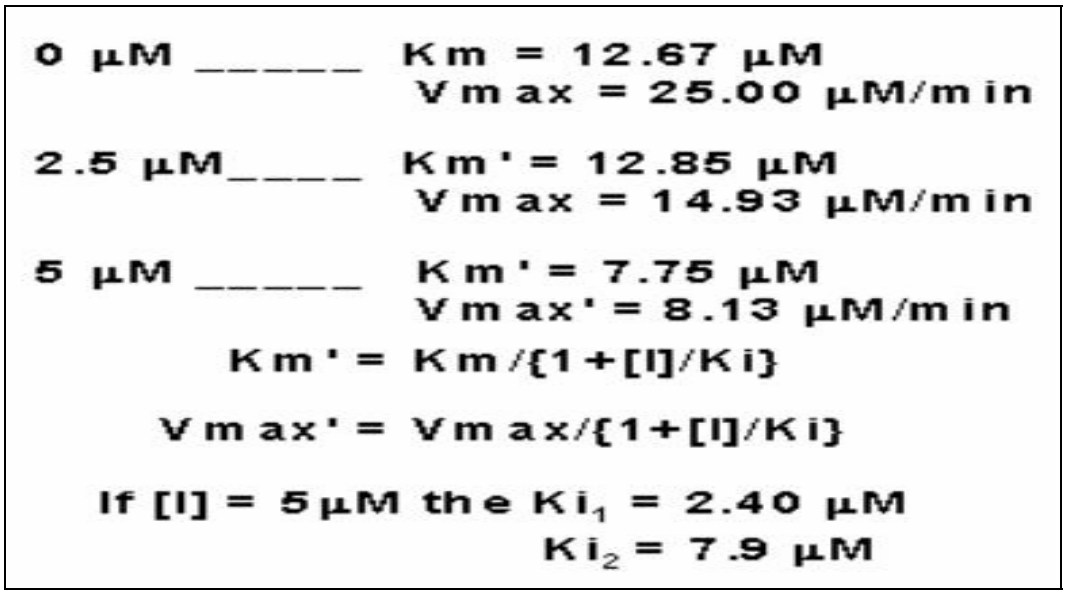


Figure 118 (C-D). Progress curve analysis of kinetics for thrombin inhibition run at saturation concentration of substrate S2238 in the presence of different peptide inhibitors. The concentration of substrate and the velocity on different points of each curve were determined as described in Materials and Methods and presented as Eadie-Hofstee or Lineweaver-Burk plots. The vmax and Km were calculated for each inhibitor based on the inhibition plots.

The progress curve analysis for thrombin inhibition by peptide **[D-Phe-Pro-D-Arg-*D*-*Cys*-CONH$_2$]** was used to obtain the Eadie-Hofstee plot. From Eadie-Hofstee the inhibitory effect of peptide was determined by assessing the effect upon Km and Vmax, from which the Ki was calculated assuming the competitive or mixed inhibition model (1-10).

As can be seen from the results of thrombin inhibition a mixed competition model can be proposed in which the peptide binds with high affinity to the active site (competitive inhibition) (Ki_1 = 2.4 uM) but it binds independently to another binding site (it can overlap partially with the active site) in the ES (enzyme-substrate) complex (Ki_2=7.9 uM). In this model the Km remains almost unchanged as can be seen from the above results (12.85 uM) only at low concentration of peptide (below 5 uM). As the

concentration of peptide is increasing above 5 uM the mixed inhibition behavior became more powerful and we obtained an effect upon both the Km and the Vmax.

The same type of mixed inhibition behavior was obtained for the peptide D-Phe-Pro-D-Arg-Ser-CONH2 (i.e. competitive and uncompetitive inhibition). The peptide affected both the Km and vmax during the kinetics of thrombin inhibition (Figure 118 E) again at lower concentration of peptide being mostly competitive inhibition while at higher concentration behaving mostly as uncompetitive inhibition (peptide is binding to ES complex and both Km and vmax are decreased).

E)

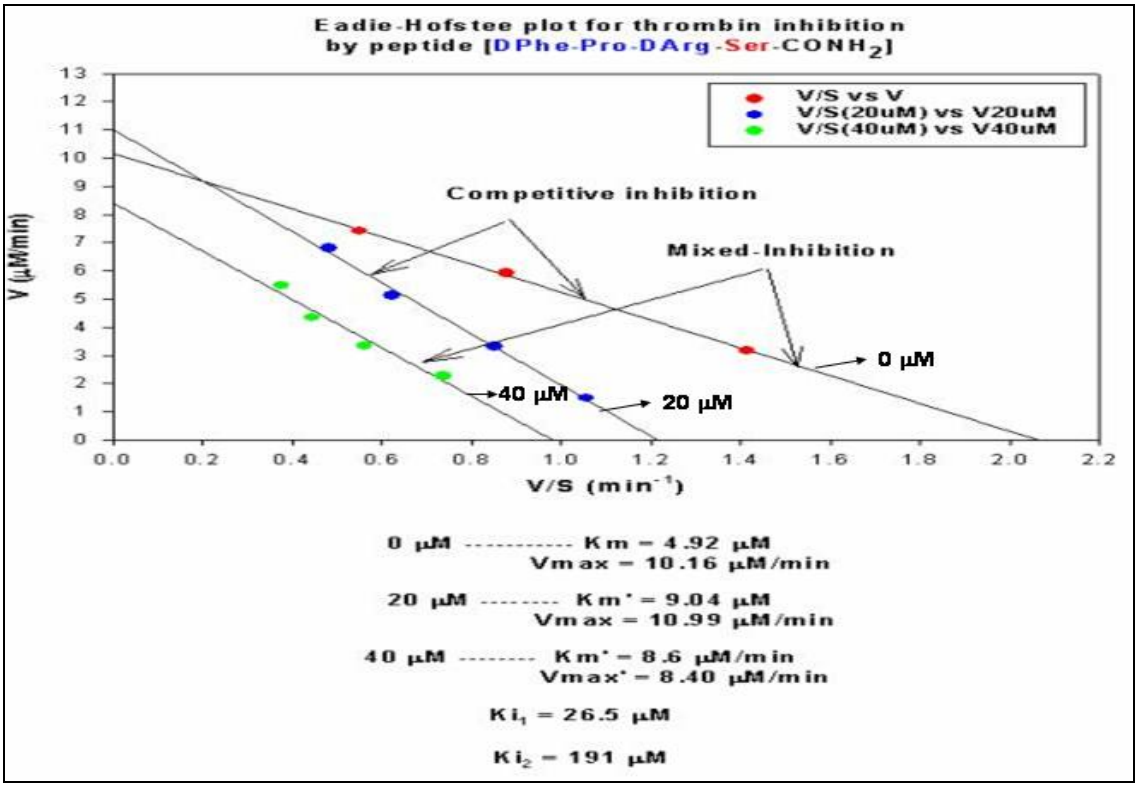
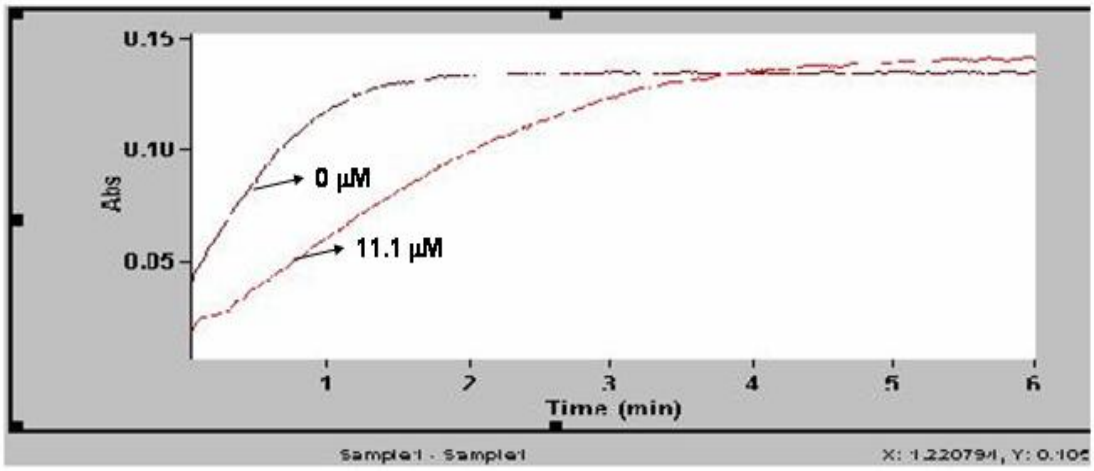


Figure 118 (E). Progress curve analysis of kinetics for thrombin inhibition run at saturation concentration of substrate S2238 in the presence of different peptide inhibitors. The concentration of substrate and the velocity on different points of each curve were determined as described in Materials and Methods and presented as Eadie-Hofstee or Lineweaver-Burk plots. The vmax and Km were calculated for each inhibitor based on the inhibition plots.

F)

[D-Phe-Pro-D-Arg-D-Ser-$CONH_2$]



0 µM peptide inhibitor

Sample1	0.000	1.000	0.0803	0.0803	0.0392	0.1181	0.0030
Sample1	1.000	2.000	0.0147	0.0147	0.1181	0.1337	0.0014
Sample1	2.000	3.000	0.0010	0.0010	0.1337	0.1352	0.0003
Sample1	3.000	4.000	-0.0003	-0.0003	0.1352	0.1349	0.0003
Sample1	4.000	5.000	0.0002	0.0002	0.1349	0.1354	0.0003

5.6 µM peptide inhibitor

	0.000	1.000	0.0533	0.0533	0.0169	0.0682	0.0005
Sample1	1.000	2.000	0.0455	0.0455	0.0682	0.1135	0.0006
Sample1	2.000	3.000	0.0224	0.0224	0.1135	0.1358	0.0007
Sample1	3.000	4.000	0.0063	0.0063	0.1358	0.1424	0.0005
Sample1	4.000	5.000	0.0012	0.0012	0.1424	0.1437	0.0003

11.1 µM peptide inhibitor

	0.000	1.000	0.0431	0.0431	0.0167	0.0606	0.0010
Sample1	1.000	2.000	0.0388	0.0388	0.0606	0.0993	0.0005
Sample1	2.000	3.000	0.0242	0.0242	0.0993	0.1239	0.0004
Sample1	3.000	4.000	0.0115	0.0115	0.1239	0.1359	0.0005
Sample1	4.000	5.000	0.0039	0.0039	0.1359	0.1393	0.0005

0 µM _____ Km = 4.507µM
Vmax = 9.197 [µM/min]

5.6 µM _____ Km' = 7.1 µM
Vmax = 9.81 [µM/min]

11.1 µM ____ Km' = 6.5 µM
Vmax = 8.041 [µM/min]

$Km' = Km/(1+[I]/Ki)$

For [I] = 5.6 µM ------ Ki = 15.33 µM

Figure 118 (F). Progress curve analysis of kinetics for thrombin inhibition run at saturation concentration of substrate S2238 in the presence of different peptide inhibitors. The concentration of substrate and the velocity on different points of each curve were determined as described in Materials and Methods and presented as Eadie-Hofstee or Lineweaver-Burk plots. The vmax and Km were calculated for each inhibitor based on the inhibition plots.

G)

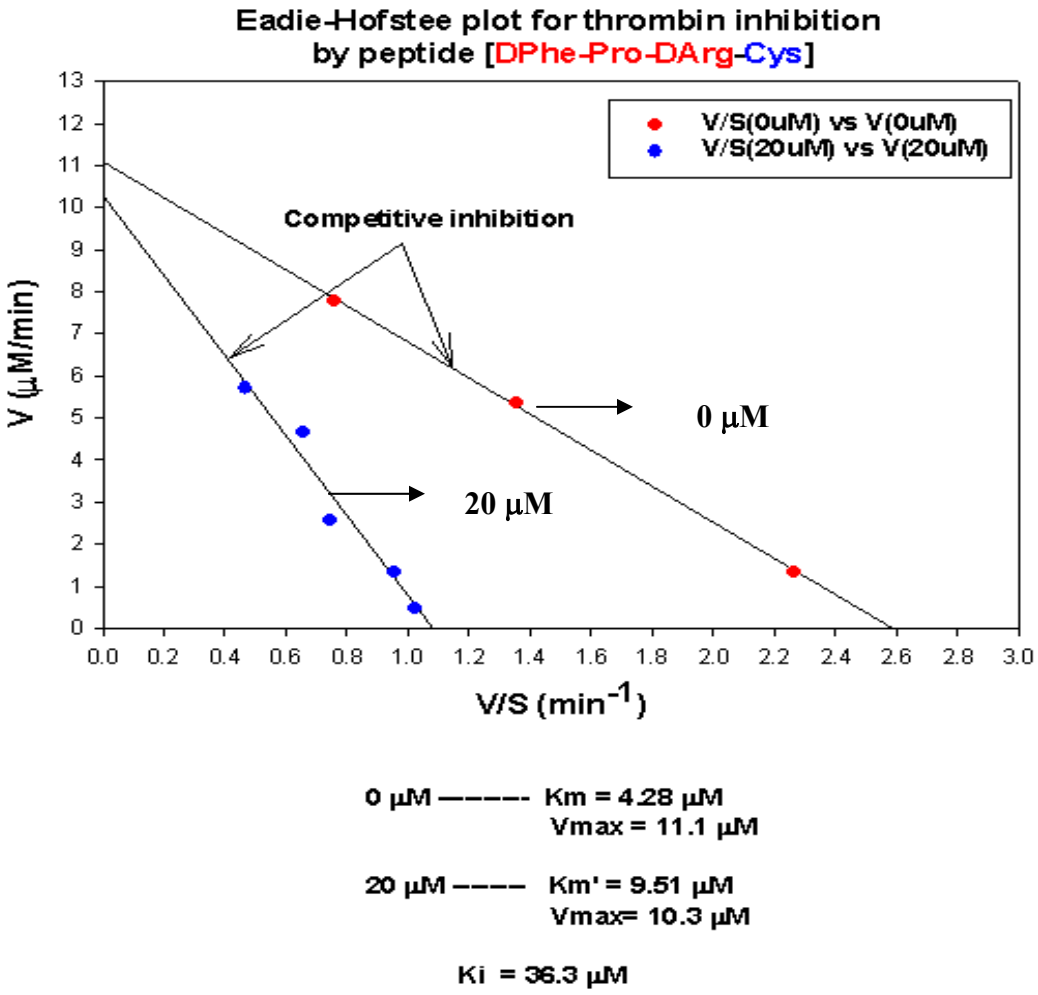
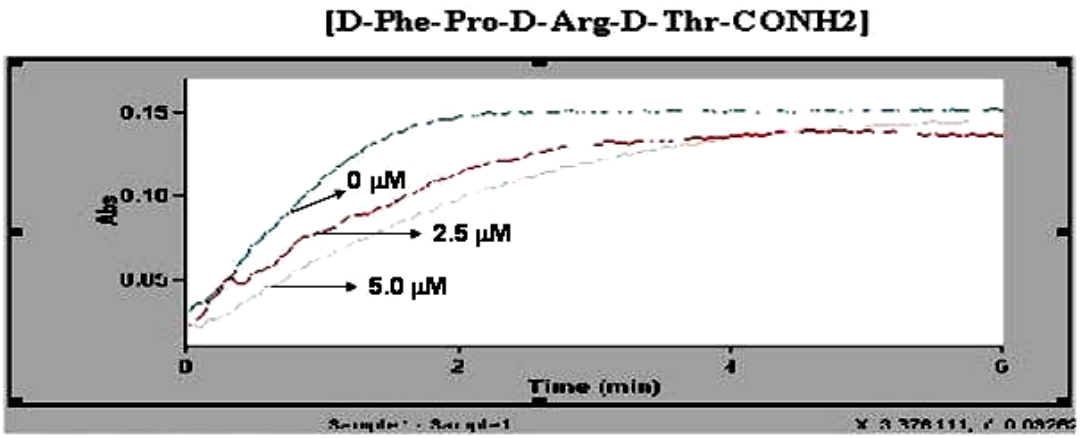


Figure 118 (G). Progress curve analysis of kinetics for thrombin inhibition run at saturation concentration of substrate S2238 in the presence of different peptide inhibitors. The concentration of substrate and the velocity on different points of each curve were determined as described in Materials and Methods and presented as Eadie-Hofstee or Lineweaver-Burk plots. The vmax and Km were calculated for each inhibitor based on the inhibition plots.

H)



Zero Order Result Table
Factor 1.00

Sample	Start (min)	Stop (min)	Slope (Abs/min)	Factor* Slope	A Start	A Stop	S.D.
0 µM peptide inhbitor							
Sample1	0.000	1.000	0.0842	0.0842	0.0311	0.1098	0.0016
Sample1	1.000	2.000	0.0367	0.0367	0.1098	0.1470	0.0025
Sample1	2.000	3.000	0.0026	0.0026	0.1470	0.1499	0.0005
Sample1	3.000	4.000	0.0003	0.0003	0.1499	0.1507	0.0004
Sample1	4.000	5.000	0.0000	0.0000	0.1507	0.1509	0.0004
2.5 µM peptide inhibitor							
	0.000	1.000	0.0549	0.0549	0.0227	0.0786	0.0030
Sample1	1.000	2.000	0.0347	0.0347	0.0786	0.1132	0.0009
Sample1	2.000	3.000	0.0171	0.0171	0.1132	0.1305	0.0007
Sample1	3.000	4.000	0.0039	0.0039	0.1305	0.1357	0.0004
Sample1	4.000	5.000	0.0029	0.0029	0.1357	0.1391	0.0005
5.0 µM peptide inhibitor							
	0.000	1.000	0.0453	0.0453	0.0230	0.0631	0.0013
Sample1	1.000	2.000	0.0351	0.0351	0.0631	0.0979	0.0006
Sample1	2.000	3.000	0.0224	0.0224	0.0979	0.1202	0.0006
Sample1	3.000	4.000	0.0136	0.0136	0.1202	0.1350	0.0005
Sample1	4.000	5.000	0.0065	0.0065	0.1350	0.1413	0.0007

0 µM _____ Km =9.8 µM
Vmax' = 19.78 µM/min

2.5 µM _____ Km = 9.64 µM
Vmax' = 12.45 µM/min

5.0 µM _____ Km = 8.66 µM
Vmax' = 8.42 µM/min

$$V_{max}' = V_{max}/\{1+[I]/K_i\}$$

$$K_i = 4.26\ \mu M$$

Figure 118 (H). Progress curve analysis of kinetics for thrombin inhibition run at saturation concentration of substrate S2238 in the presence of different peptide inhibitors. The concentration of substrate and the velocity on different points of each curve were determined as described in Materials and Methods and presented as Eadie-Hofstee or Lineweaver-Burk plots. The vmax and Km were calculated for each inhibitor based on the inhibition plots.

I)

Eadie-Hofstee plot for thrombin inhibition by peptide

[D-Phe-Pro-D-Arg-Ala-CONH2]

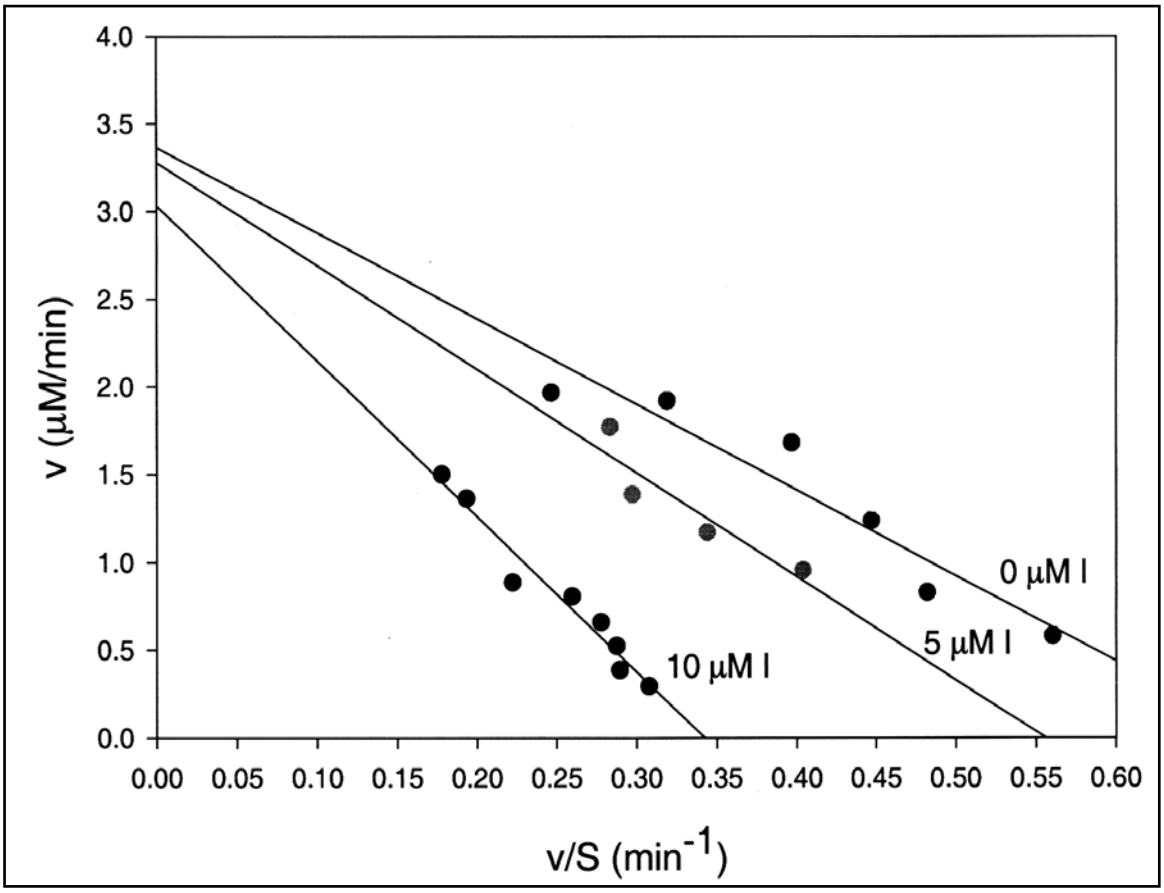


Figure 118 (I). Progress curve analysis of kinetics for thrombin inhibition run at saturation concentration of substrate S2238 in the presence of different peptide inhibitors. The concentration of substrate and the velocity on different points of each curve were determined as described in Materials and Methods and presented as Eadie-Hofstee or Lineweaver-Burk plots. The vmax and Km were calculated for each inhibitor based on the inhibition plots. The Ki calculated from the slope of the curve for 5 μM inhibitor was around 17 μM which was the same with the Ki calculated at pseudo-first order kinetics (see table 8). At 10 μM inhibitor a mixed type of inhibition is observed. In this case the two Kis were: Ki (I) = 26.31 μM and Ki (II) = 8.53 μM.

J)

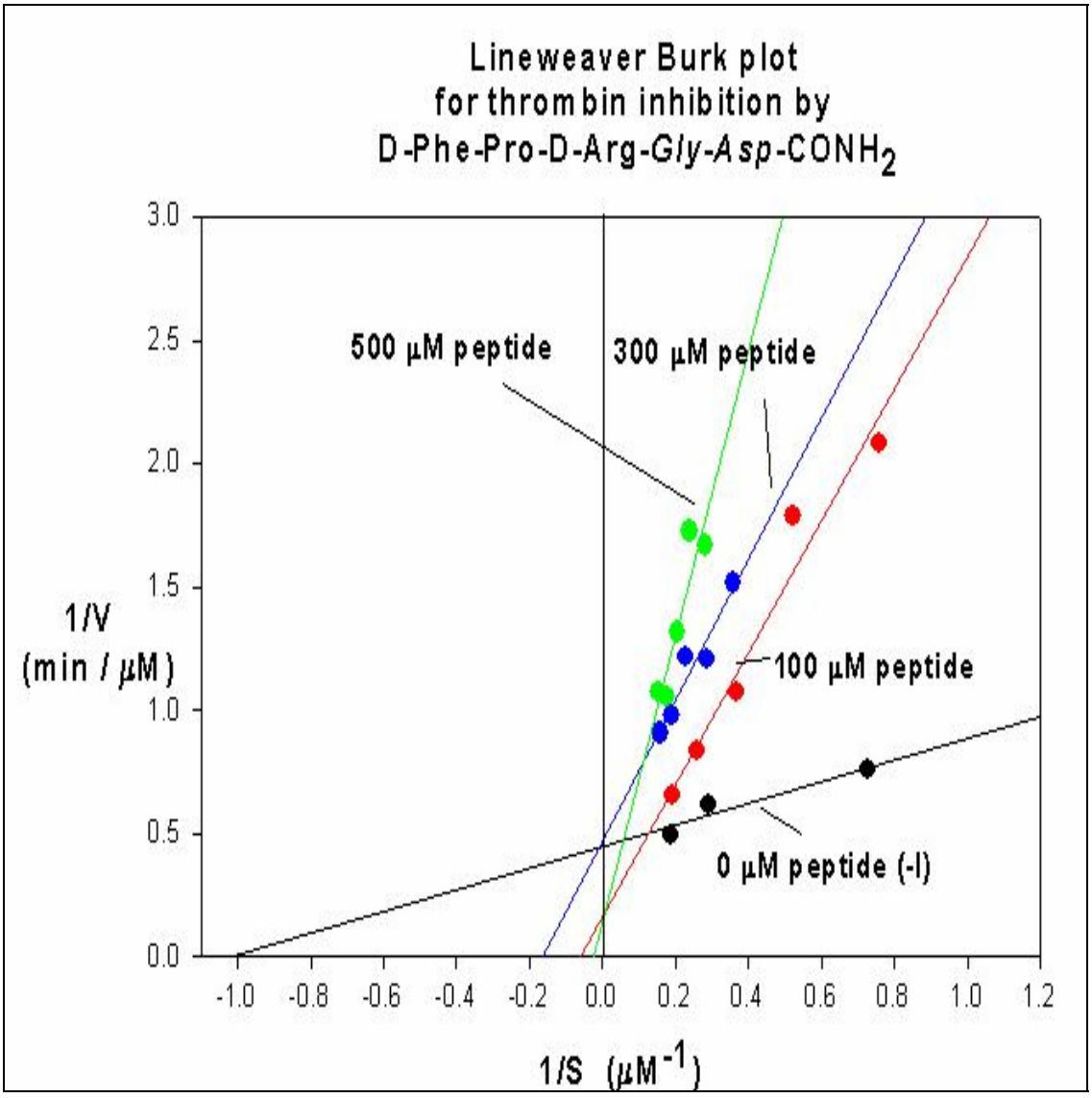


Figure 118 (J). Progress curve analysis of kinetics for thrombin inhibition run at saturation concentration of substrate S2238 in the presence of different peptide inhibitors. The concentration of substrate and the velocity on different points of each curve were determined as described in Materials and Methods and presented as Eadie-Hofstee or Lineweaver-Burk plots. The vmax and Km were calculated for each inhibitor based on the inhibition plots.

K)

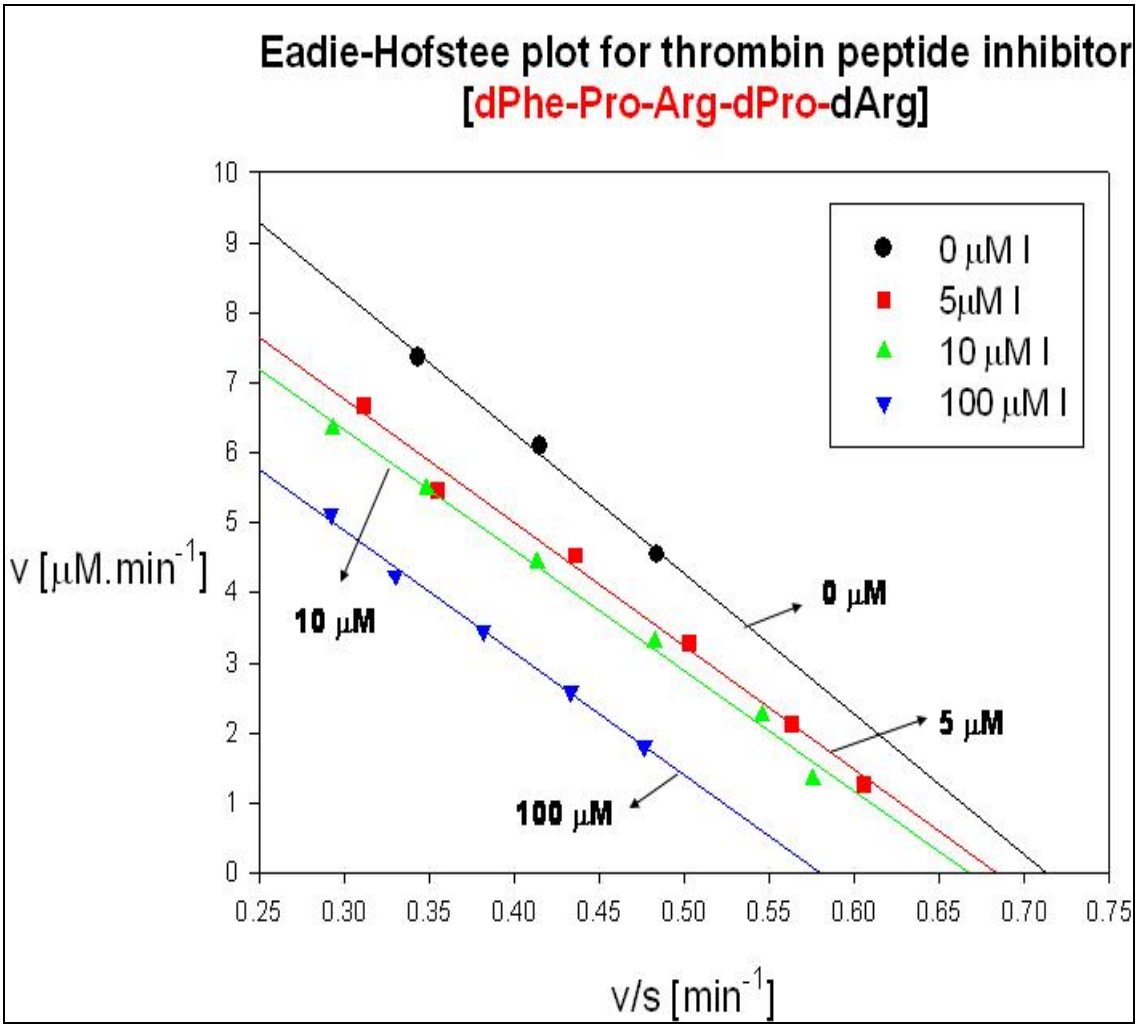


Figure 118 (K). Progress curve analysis of kinetics for thrombin inhibition run at saturation concentration of substrate S2238 in the presence of different peptide inhibitors. The concentration of substrate and the velocity on different points of each curve were determined as described in Materials and Methods and presented as Eadie-Hofstee or Lineweaver-Burk plots. The vmax and Km were calculated for each inhibitor based on the inhibition plots.

L)

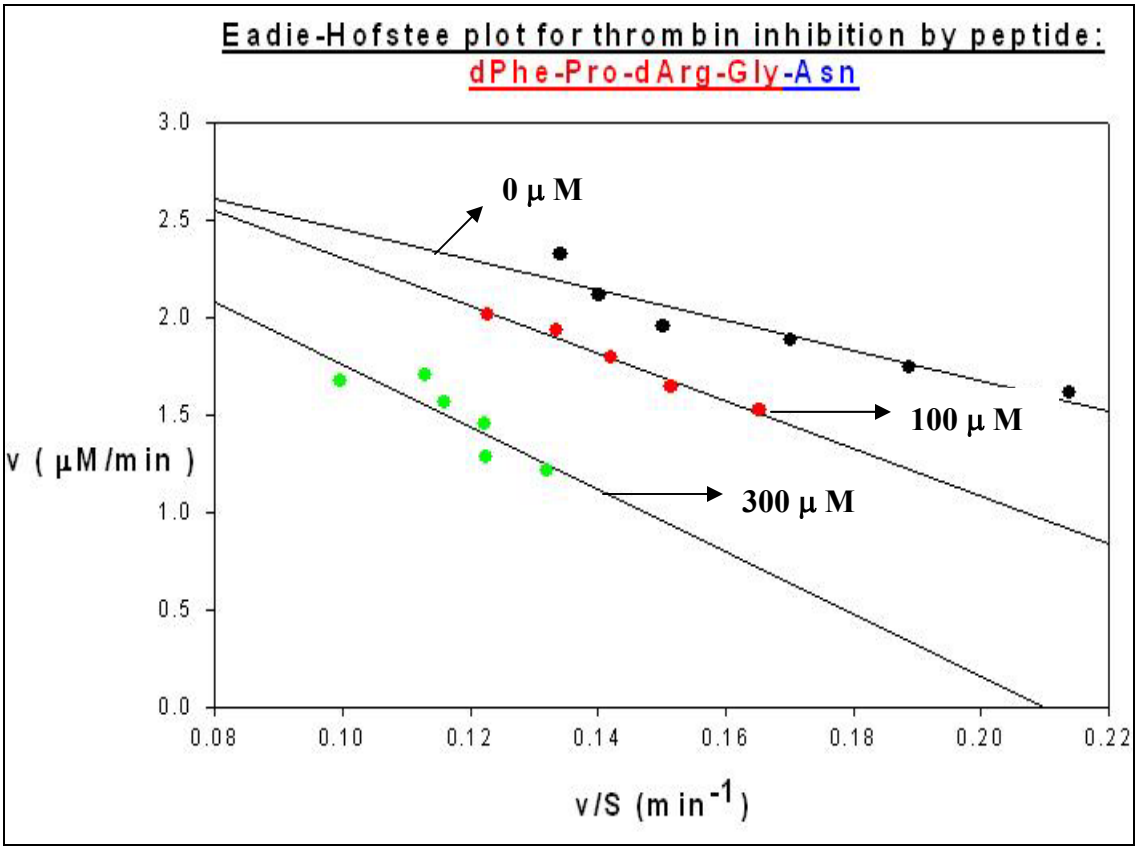


Figure 118 (A-L). Progress curve analysis of kinetics for thrombin inhibition run at saturation concentration of substrate S2238 in the presence of different peptide inhibitors. The concentration of substrate and the velocity on different points of each curve were determined as described in Materials and Methods and presented as Eadie-Hofstee or Lineweaver-Burk plots. The vmax and Km were calculated for each inhibitor based on the inhibition plots.

3.4. DISCUSSIONS

The main purpose of this research investigation was to discover peptides with potential inhibitory activity against thrombin. We were interested primarily in finding reversible peptides inhibitors directed against the active site of thrombin which would allow the inhibition of blood clotting.

Originally we discovered that synthetic pentapeptides such as D-Phe-Pro-D-Arg-Gly-Asp and D-Phe-Pro-D-Arg-Gly-Asn are having 8 fold more inhibitory effect than the peptides D-Tyr-Pro-Arg-D-Pro-D-Tyr and D-Phe-Pro-D-Arg-Tyr-Phe (figure 117 (S) and sequences D-Phe-Pro-D-Arg-Gly-Asp and D-Phe-Pro-D-Arg-Gly-Asn have the highest %inhibition of thrombin at 300 μM final concentration, more then 60% inhibition in the case of D-Phe-Pro-D-Arg-Gly-Asp and more than 70% inhibition of thrombin in the case of D-Phe-Pro-D-Arg-Gly-Asn sequence (Figure 117 (S)).

Another set of sequences like peptides D-Phe-Pro-D-Arg-Gly-Asp-Ala, D-Phe-Pro-D-Arg-Gly-Asp-Lys and D-Phe-Pro-Arg-D-Pro-D-Arg and D-Phe-Pro-D-Arg-Gly-Asp-Met have more then 30% inhibition or more then 20% inhibition of thrombin respectively at 300 μM final concentration of peptide in the reaction mixture. Other sequences like D-Phe-Pro-D-Arg-Tyr-Phe or D-Tyr-Pro-Arg-D-Pro-D-Tyr have very small inhibitory activity (less than 10%) at the same final concentration of inhibitor (300 μM).

All peptides showed to be resistant to hydrolysis by thrombin since the preincubation of thrombin (30 nM) with each peptide (20 μM) for 30 minutes at room temperature in the same buffer used for kinetics of thrombin inhibition didn't resultated in the production of tripeptide D-Phe-Pro-D-Arg-COOH (MW= 418.5) as would have been expected if hydrolysis would have taken place (i.e. D-Phe-Pro-D-Arg-X-CONH2 has the scissile bond -D-Arg-X- stable to hydrolysis). The results from ESI-MS analysis of the reaction mixture thrombin+peptide are presented in Appendix XII.

From these original data we concluded that there is a structure-activity relationship (SAR) between the specific amino acid sequences and the inhibitory activity. From this initial study we hypothesized that the peptides which have the sequence D-Phe-

Pro-D-Arg-Gly-Asp(Asn) are better inhibitors then those that are conserving the D-Phe-Pro-D-Arg sequence but have aromatic amino acids in the next two consecutive positions from N to C terminus (like Tyr and Phe).

These original SAR studies in the pentapeptides and hexapeptides series supported the hypothesis that the sequence space D-Phe-Pro-D-Arg-P1'-CONH2 is more effective than the sequence D-Phe-Pro-L-Arg-D-Pro-P2'. We proceeded further and we synthesized tetrapeptide libraries using the sequence space D-Phe-Pro-D-Arg-P1' were the P1' position was occupied with all structural classes of natural L-aminoacids, such that an SAR at P1' position was investigated. We were successful in finding a Structure-Activity Relationship (SAR) between specific amino acid sequences and the inhibitory activity. The peptides D-Phe-Pro-D-Arg-*Gly*-$CONH_2$, D-Phe-Pro-D-Arg-*Ala*-$CONH_2$, D-Phe-Pro-D-Arg-*D-Ala*-CONH2, D-Phe-Pro-D-Arg-*D-Cys*-$CONH_2$, D-Phe-Pro-D-Arg-*Cys*-$CONH_2$, D-Phe-Pro-D-Arg-*Ser*-$CONH_2$, D-Phe-Pro-D-Arg-*D-Ser*-$CONH_2$, D-Phe-Pro-D-Arg-*Ile*-$CONH_2$, D-Phe-Pro-D-Arg-*Thr*-$CONH_2$, D-Phe-Pro-D-Arg-D-*Thr*-$CONH_2$ are lead inhibitors having Ki values below 15 μM (see table 14). The new inhibitor D-Phe-Pro-DArg-*(L)Thi*-$CONH_2$ has a Ki of 8.6 μM and was discovered as being a new lead compound in the series D-Phe-Pro-DArg-P1' and contains an unnatural amino acid analog in the P1' position.

These results support the view that the structure of the amino-acid at P1' position in the sequence context D-Phe-Pro-DArg-*P1'*-$CONH_2$ may induce significant changes in the conformation of the peptide inhibitor bound reversibly to the active site or to an independent site of thrombin possible near the active site groove. We are hypothesizing that these changes in the conformation of the peptides are accompanied by local changes in the conformation of thrombin (induced fit) such that amino acids and amino acids

analogs with bulky hydrophobic side chains are fitting in P1' position (such as L-Ile and L-Thi) even though they are occupying the P2' positions in natural substrates (such L-Ile) (40-50). Other group of researchers (Usakiewicz J.J.S et al., 2000) (51) have shown using bivalent peptide inhibitors which are blocking active site and the fibrinogen binding site that the L-Thi in P1' position induces a conformation changes in thrombin such that the Lys^{60F} side chain of thrombin is moving significantly and is forming a large nonpolar S1' subsite to accommodate larger bulky residues in P1' (X-Ray structures of the bivalent peptide inhibitors with thrombin were the basis for the observed structural changes (51). The proposed changes in the conformation both in the inhibitors and in the active site of thrombin itself may be responsible for the 5-500 fold determined differences in their inhibitory activity (see Tables 14-15). The molecular models of the conformation adopted by the peptides into active site of thrombin support also the experimental data (see figure 109 (A-D)) by showing the change in the conformation of the peptide backbone and of specific side chains of the residues at P3-P1' positions as one single amino acid at P1' is changing.

Also, significant differences between the Kis of tetrapeptides from the series DPhe-Pro-DArg-P1' suggest that the interaction between the amino acid at the P1' position and the S1' pocket in thrombin is very specific. These results are consistent with the X-ray data which showed that the S1' pocket of thrombin is a small cavity lined with amino acids of the back side of the apolar S2 pocket (His57, Tyr60A and Trp60D) and the side chain of Lys60F (51). This S1' pocket can accommodate small polar amino acid side chains as observed from P1' amino acids in a number of natural thrombin substrates (P1' in most physiological substrates is occupied by Gly, Ser, Thr, Ile, Leu and Val). The computational model of the active-site docked tetrapeptides from the series D-Phe-Pro-D-

Arg-P1' and their experimentally determined Ki values (see tables 14-15 and figure 109) support the X-ray data since the best inhibitors are having L/D amino acids with small side chain (like Gly) or with polar (like Cys, Ser, Thr) or small nonpolar side chains (Ala).

The most important experimental findings are related to the change in the affinity as the configuration of the amino-acid in the P1' position changes from L to D. The switch from L to D in the P1' is related to a 13 fold increase in the affinity of tetrapeptides from the series D-Phe-Pro-D-Arg-P1'-$CONH_2$. These experiments are the first in the field of reversible peptides inhibitors for thrombin to show that a switch in the configuration from L to D for the amino-acids mostly encountered at P1' in the physiological thrombin's substrates (such as L-Ser, L-Thr and L-Ala) causes at least a 10 fold increase in the affinity for thrombin (figure 117 E-F). The structural basis for this favorable switch in the affinity was further investigated through molecular modeling experiments and as described earlier on section 1.1. we observed a more favorable hydrogen bonding network for the D-amino acid in P1' than for the L-isomer (see figure 15 D).

These differences in the binding affinities upon switching from L into D for amino-acids in P1' was confirmed both kinetically (by measuring the observed rates in pseudo-first order kinetics-see figures) and through ITC data (the heat released for peptides with L-Ala in P1' is lower (1.2 kcal/mole of injectant) than the heat released during the titration of the same peptide but with D-Ala in P1' (4.5 kcal/mole of injectant) (see figures 120-123 and Appendix XIII).

We were also developing peptides with new P3-analogs of D-Phe (see figure 25). The SAR for P3 position (table 15) showed that the tetrapeptides from series D-Phe-Pro-

DArg-P1' are the best together with the peptides from series D-Nal (naphthylalanine)-Pro-DArg-P1'. Thus, D-Nal substitution at P3 position increases the affinity of a tetrapeptide 2 fold (compare the compound with compound). The order of activity for the peptides containing analogs of D-Phe in the P3 position is: D-(Nal) (naphtyl-alanine) > (D)Phe> Dihydrocinnamic>Transcinnamic > >(D)Tic> L-Thi> (L)Tic >D-Tpi with conserved residues at P2=Pro and P1=D-Arg. New peptides libraries containing D-Nal at P3 position and variable favorable D/L natural or unnatural amino acids at P1' position are under current investigation.

One of the most important mechanistically findings suggest that the experimentally determined thrombin inhibition is not 100% competitive but rather has around 10% non-competitive and uncompetitive, mixed inhibition behavior. It is possible that the peptides are binding to the ES complex with higher affinity than to free enzyme (E), at higher concentration than their Ki (figure 118). However, the results have to be interpreted with caution since the peptides might bind partially overlapping the binding sites of the substrate, a process which would explain the mixed inhibition behavior of these peptides (see figures 115, 116). Also, it is possible that some of these peptides are assembling in higher molecular complexes at higher concentration (> 2 their Ki) and this phenomena could be responsible for the deviation from the competitive type of inhibition behavior. We found inhibitors which were showing an increased Ki with increasing concentration of peptides, such as the inhibitors D-Phe-Pro-L-Arg-*D-Pro-L-Phe*-CONH2 or the trans and cis-cinnamic-peptide derivatives (see figures 115, 116). However, other inhibitors which were showing mixed inhibition behavior were not necessary showing the increased in Ki as the concentration of peptide was increasing in the reaction mixture, such is the case of peptides D-Phe-Pro-D-Arg-*D-Ala*-CONH2 and

D-Phe-Pro-D-Arg-*D-Ser*-CONH2 (see figure 118). These contradictory results supports the view that the mixed inhibition behavior is not due entirely to the assembly of peptides in higher macromolecular complexes but to binding of individual peptides to other sites nearby the active site of thrombin. Further structural studies such as X-ray of binary complexes between these lead peptides and thrombin will respond to this question raised from kinetics of thrombin inhibition.

REFERENCES

1. Bode, W.; Mayr, I.; Baumann, U.; Huber, R.; Stone, S. R.; Hofsteenge, J. The refined 1.9 Å crystal structure of α-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J*. **1989**, *8*, 3467-3475.

1. Bode, W.; Turk, D.; Karshikov, A. The refined 1.9 Å X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human α-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Sci*. **1992**, *1*, 426-471.

2. Berliner, L. J., Ed. *Thrombin: Structure and Function*; Plenum Press: New York, 1992.

3. Davie, E. W.; Fugikawa, K.; Kisiel, W. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* **1991**, *30*, 10363-10370.

4. Maffrand, J. P. Direct thrombin inhibitors. *Nouv*. *Rev*. *Fr*. *Hematol*. **1992**, *34*, 405-419.

5. Talbot, M. D.; Butler, K. D. Potential clinical uses of thrombin inhibitors. *Drug News Perspect*. **1990**, *3*, 357-363.

6. Das, J.; Kimball, S. D. Thrombin active site inhibitors. *Bioorg*. *Med*. *Chem*. **1995**, *3*, 999-1007.

7. Scarborough, R. M. Anticoagulant strategies targeting thrombin and factor Xa. *Annu*. *Rep*. *Med*. *Chem*. **1995**, *30*, 71-80.

8. Kimball, S. D. Thrombin active site inhibitors. *Curr*. *Pharm*. *Des*. **1995**, *1*, 441-468.

9. Fareed, J.; Callas, D. D. Pharmacological aspects of thrombin inhibitors: a developmental perspective. *Vessels* **1995**, *1*, 15-24.

10. Lefkovits, J.; Topol, E. J. Direct thrombin inhibitors in cardiovascular medicine. *Circulation* **1994**, *90*, 1522-1536.

11. Betschmann, P.; Lerner, C.; Sahli, S.; Obst, U.; Diederich, F. Molecular recognition with biological receptors: structure-based design of thrombin inhibitors. *Chimia* **2000**, *54*, 633-639.

12. Menear, K. Direct thrombin inhibitors: Current status and future prospects. *Expert Opin. Invest. Drugs* **1999**, *8*, 1373-1384.

13. Ripka, W. C.; Vlasuk, G. P. Antithrombotics/serine proteases. *Annu. Rep. Med. Chem.* **1997**, *32*, 71-89.

14. Sanderson, P. E. J. Small, noncovalent serine protease inhibitors. *Med. Res. Rev.* **1999**, *19*, 179-197.

15. Kimball, S. D. Challenges in the development of orally bioavailable thrombin active site inhibitors. *Blood Coagulation Fibrinolysis* **1995**, *6*, 511-519.

16. Kettner, C.; Shaw, E. D-Phe-Pro-ArgCH_2Cl-a selective affinity label for thrombin. *Thromb. Res.* **1979**, *14*, 969-973.

17. Hauptmann, J.; Markwardt, F. Studies on the anticoagulant and antithrombotic action of an irreversible thrombin inhibitor. *Thromb. Res.* **1980**, *20*, 347-351.

18. Rydel, T. J.; Ravichandran, K. G.; Tulinsky, A.; Bode, W.; Huber, R.; Roitsch, C.; Fenton, J. W., II. The structure of a complex of recombinant hirudin and human α-thrombin. *Science* **1990**, *249*, 277-280.

19. Bajusz, S.; Széll, E.; Bagdy, D.; Barabás, E.; Horvath, G.; Dioszegi, M.; Fittler, Z.; Szabo, G.; Juhasz, A.; Tomori, E.; Szilagyi, G. Highly active and selective anticoagulants: D-Phe-Pro-Arg-H, a free tripeptide aldehyde prone to spontaneous inactivation, and its stable *N*-methyl derivative, D-MePhe-Pro-Arg-H. *J. Med. Chem.* **1990**, *33*, 1729-1735.

20. Tomori, É.; Széll, E.; Barabás, É. High-performance liquid chromatograhy of a new tripeptide aldehyde (GYKI-14166), correlation between the structure and activity. *Chromatographia* **1984**, 437-442.

21. Bajusz, S.; Barabás, E.; Fauszt, I.; Feher, A.; Horvath, G.; Juhasz, A.; Szabo, A. G.; Széll, E. Active site-directed thrombin inhibitors: α-hydroxyacyl-prolyl-arginals.

New orally active stable analogs of D-Phe-Pro-Arg-H. *Bioorg*. *Med*. *Chem*. **1995**, *3*, 1079-1089.

22. Shuman, R. T.; Rothenberger, R. B.; Campbell, C. S.; Smith, G. F.; Gifford-Moore, D. S.; Gesellchen, P. D. Highly selective tripeptide thrombin inhibitors. *J*. *Med*. *Chem*. **1993**, *36*, 314-319.

23. Maryanoff, B. E.; Qiu, X.; Padmanabhan, K. P.; Tulinsky, A.; Almond, H. R., Jr.; Andrade-Gordon, P.; Greco, M. N.; Kauffman, J. A.; Nicolaou, K. C.; Liu, A.; Brungs, P. H.; Fusetani, N. Molecular basis for the inhibition of human α-thrombin by the macrocyclic peptide cyclotheonamide A. *Proc*. *Natl*. *Acad*. *Sci*. *U.S.A.* **1993**, *90*, 8048-805.

24. Maryanoff, B. E.; Greco, M. N.; Zhang, H.-C.; Andrade-Gordon, P.; Kauffman, J. A.; Nicolaou, K. C.; Liu, A.; Brungs, P. H. Macrocyclic peptide inhibitors of serine proteases. Convergent synthesis of cyclotheonamides A and B via a late-stage primary amine intermediate. Study of thrombin inhibition under diverse conditions. *J*. *Am*. *Chem*. *Soc*. **1995**, *117*, 1225-1239.

25. Maryanoff, B. E.; Zhang, H.-C.; Greco, M. N.; Glover, K. A.; Kauffman, J. A.; Andrade-Gordon, P. Cyclotheonamide derivatives: synthesis and thrombin inhibition. Exploration of specific structure-function issues. *Bioorg*. *Med*. *Chem*. **1995**, *3*, 1025-1038.

26. Costanzo, M. J.; Maryanoff, B. E.; Hecker, L. R.; Schott, M. R.; Yabut, S. C.; Zhang, H.-C.; Andrade-Gordon, P.; Kauffman, J. A.; Lewis, J. M.; Krishnan, R.; Tulinsky, A. Potent thrombin inhibitors that probe the S_1' subsite: tripeptide transition state analogs based on a heterocycle-activated carbonyl group. *J*. *Med*. *Chem*. **1996**, *39*, 3039-3043.

27. Matthews, J. H.; Krishnan, R.; Costanzo, M. J.; Maryanoff, B. E.; Tulinsky, A. Crystal structures of thrombin with thiazole-containing inhibitors: probes of the S_1' binding site. *Biophys*. *J*. **1996**, *71*, 2830-2839.

28. Recacha, R.; Costanzo, M. J.; Maryanoff, B. E.; Carson, M.; DeLucas, L. J.; Chattopadhyay, D. Crystal structure of human α-thrombin complexed with RWJ-51438 at 1.7 Å: unusual perturbation of the 60A-60I insertion loop. *Acta Crystallogr*., *Sect*. *D*: *Biol*. *Crystallogr*. **2000**, *56*, 1395-1400.

29. Giardino, E. C.; Costanzo, M. J.; Kauffman, J. A.; Li, Q. S.; Maryanoff, B. E.; Andrade-Gordon, P. Antithrombotic properties of RWJ-50353, a potent and novel thrombin inhibitor. *Thromb. Res.* **2000**, *98*, 83-93.

30. Sorbera, L. A.; Bayes, M.; Castaner, J.; Silvestre, J. Melagatran and ximelagatran: anticoagulant thrombin inhibitor. *Drugs Future* **2001**, *26*, 1155-1170.

31. Gustafsson, D.; Nystrom, J.-E.; Carlsson, S.; Bredberg, U.; Eriksson, U.; Gyzander, E.; Elg, M.; Antonsson, T.; Hoffmann, K.-J.; Ungell, A.-L.; Sorensen, H.; Nagard, S.; Abrahamsson, A.; Bylund, R. The direct thrombin inhibitor melagatran and its oral prodrug H 376/95: Intestinal absorption properties, biochemical and pharmacodynamic effects. *Thromb. Res.* **2001**, *101*, 171-181.

32. Sarich, T. C.; Eriksson, U. G.; Mattsson, C.; Wolzt, M.; Frison, L.; Fager, G.; Gustafsson, D. Inhibition of thrombin generation by the oral direct thrombin inhibitor ximelagatran in shed blood from healthy male subjects. *Thromb. Haemostasis* **2002**, *87*, 300-305. Hopfner, R. Ximelagatran (AstraZeneca). *Curr. Opin. Invest. Drugs* **2002**, *3*, 246-251.

33. Shuman, R. T.; Rothenberger, R. B.; Campbell, C. S.; Smith, G. F.; Gifford-Moore, D. S.; Paschal, J. W.; Gesellchen, P. D. Structure-activity study of tripeptide thrombin inhibitors using α-alkyl amino acids and other conformationally constrained amino acid substitutions. *J. Med. Chem.* **1995**, *38*, 4446-4453.

34. Claeson, G.; Philipp, M.; Agner, E.; Scully, M. F.; Metternich, R.; Kakkar, V. V.; DeSoyza, T.; Niu, L. H. Benzyloxycarbonyl-D-Phe-Pro-methoxypropylboroglycine: A novel inhibitor of thrombin with high selectivity containing a neutral side chain at the P1 position. *Biochem. J.* **1993**, *290*, 309-312.

35. Deadman, J. J.; Elgendy, S.; Goodwin, C. A.; Green, D.; Baban, J. A.; Patel, G.; Skordalakes, E.; Chino, N.; Claeson, G.; Kakkar, V. V.; Scully, M. F. Characterization of a class of peptide boronates with neutral P1 side chains as highly selective inhibitors of thrombin. *J. Med. Chem.* **1995**, *38*, 1511-1522.

36. Claeson, G.; Cheng, L.; Chino, N.; Deadman, J.; Elgendy, S.; Kakkar, V. V.; Scully, M. F.; Philipp, M.; Lundin, R.; Mattson, C. Novel peptide mimetics as highly efficient inhibitors of thrombin based on modified D-Phe-Pro-Arg sequences. *Pept.: Chem., Struct. Biol., Proc. Am. Pept. Symp.* **1992**, 824-825.

37. Levy, O. E.; Semple, J. E.; Lim, M. L.; Reiner, J.; Rote, W. E.; Dempsey, E.; Richard, B. M.; Zhang, E.; Tulinsky, A.; Ripka, W. C.; Nutt, R. F. Potent and selective thrombin inhibitors incorporating the constrained arginine mimic L-3-piperidyl(N-guanidino)alanine at P1. *J. Med. Chem.* **1996**, *39*, 4527-4530.

38. Semple, J. E.; Rowley, D. C.; Brunck, T. K.; Ha-Uong, T.; Minami, N. K.; Owens, T. D.; Tamura, S. Y.; Goldman, E. A.; Siev, D. V.; Ardecky, R. J.; Carpenter, S. H.; Ge, Y.; Richard, B. M.; Nolan, T. G.; Hakanson, K.; Tulinsky, A.; Nutt, R. F.; Ripka, W. C. Design, synthesis, and evolution of a novel, selective, and orally bioavailable class of thrombin inhibitors: P1-Argininal derivatives incorporating P3-P4 lactam sulfonamide moieties. *J. Med. Chem.* **1996**, *3*, 4531-4536.

39. Abelman, M. M.; Ardecky, R. J.; Nutt, R. F. Preparation of methionine sulfone and S-substituted cysteine sulfone derivatives as inhibitors of thrombin or factor Xa. PCT Int. Appl. WO 9528420, 1995.

40. Malley, M. F.; Tabernero, L.; Chang, C. Y.; Ohringer, S. L.; Roberts, D. G. M.; Das, J.; Sack, J. S. Crystallographic determination of the structures of human α-thrombin complexed with BMS-186282 and BMS-189090. *Protein Sci*. **1996**, *5*, 221-228.

41. Kaiser, B.; Hauptmann, J. Pharmacology of synthetic thrombin inhibitors of the tripeptide type. *Cardiovasc. Drug Rev.* **1992**, *10*, 71-87.

42. Wu, Q.; Sheehan, J. P.; Tsiang, M.; Lentz, R. S.; Birktoft, J. J.; Sadler, J. E. Single amino acid substitutions dissociate fibrinogen-clotting and thrombomodulin-binding activities of human thrombin. *Proc*. *Natl*. *Acad*. *Sci*. *U.S.A.* **1991**, *88*, 6775-6779.

43. Qui, X.; Padmanabhan, K. P.; Carperos, V. E.; Tulinsky, A.; Kline, T.; Maraganore, J. M.; Fenton, J. W., II. Structure of the hirulog 3-thrombin complex and nature of the S' subsites of substrates and inhibitors. *Biochemistry* **1992**, *31*, 11689-11697.

44. Krishnan, R.; Tulinsky, A.; Vlasuk, G. P.; Pearson, D.; Vallar, P.; Bergum, P.; Brunck, T. K.; Ripka, W. C. Synthesis, structure and structure-activity relationships of divalent thrombin inhibitors containing an α-keto amide transition state mimetic. *Protein Sci*. **1996**, *5*, 422-433.

45. Mathews, I. I.; Tulinsky, A. Active site mimetic inhibition of thrombin. *Acta Crystallogr*. **1995**, *D51*, 550-559.

46. Tulinsky, A.; Qiu, X. Active Site and Exosite Binding of α-Thrombin. *Blood Coagulation Fibrinolysis* **1993**, *4*, 305-312.

47. Tulinsky, A. Molecular interactions of thrombin. *Semin*. *Thromb*. *Hemostasis* **1996**, *22*, 117-124.

48. Hakansson, K.; Tulinsky, A.; Abelman, M. M.; Miller, T. A.; Vlasuk, G. P.; Bergum, P. W.; Lim-Wilby, M. S. L.; Brunck, T. K. Crystallographic structure of a peptidyl keto acid inhibitor and human α-thrombin. *Bioorg*. *Med*. *Chem*. **1995**, *3*, 1009-1017.

49. Williams, J. W.; Morrison, J. F. The kinetics of reversible tight-binding inhibition. *Methods Enzymol*. **1979**, *63*, 437-467. (b) Morrison, J. F. The slow-binding and slow, tight-binding inhibition of enzyme-catalyzed reactions. *Trends Biochem*. *Sci*. **1982**, *7*, 102-105.

Chapter 4

Isothermal titration calorimetry (ITC) studies of thrombin binding with peptides inhibitors

4.1. INTRODUCTION

Isothermal Titration Calorimetry (ITC) is a rapid, multi-probe SAR technique for confirming the success of lead optimization cycles without needing to determine the structure of every complex in this iterative process (1-20). The mechanism of action, and alterations to it, can be readily assessed using ITC. Different binding mechanisms show characteristic ITC profiles that can be represented schematically as in the Figure 119.

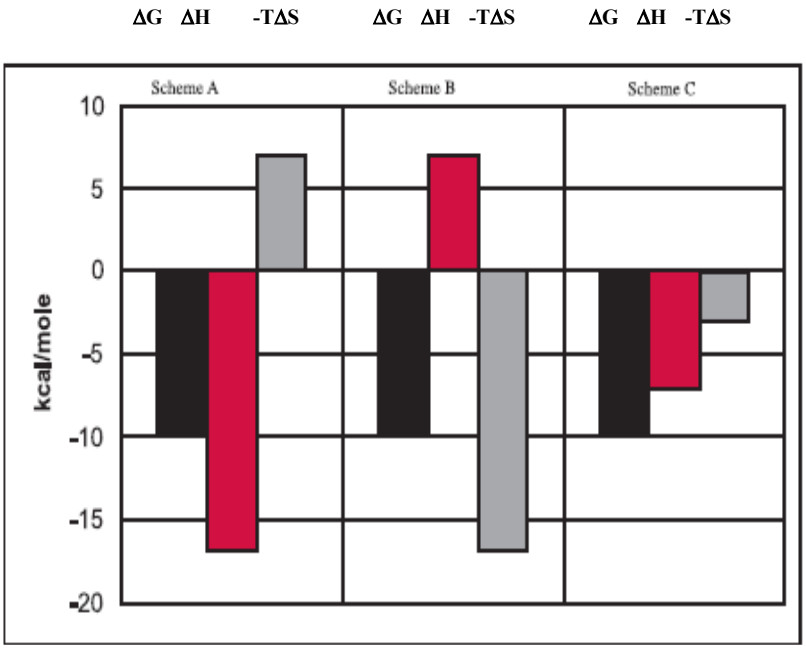


Figure 119: ITC derived thermodynamic signatures for different ligands characterized by different noncovalent interactions with the same target protein shown as changes in the free energy (ΔG), change in the enthalpy (ΔH) and change in the entropy (ΔS). Scheme A: good hydrogen bonding and a conformational change; Scheme B: binding dominated by hydrophobic interactions; Scheme C: favorable hydrogen bondings and hydrophobic interactions (1-10).

Each section in Figure 119 represents an interaction driven by different mechanisms but with identical affinities. These histograms are particularly powerful when comparing related leads. For instance, the successful addition of a hydrophobic group for improved affinity would manifest itself in the ITC data as an increase in the negative value of ΔG and $-T\Delta S$ (scheme B). An additional H-bond would characteristically increase the negative ΔG and ΔH but make the $-T\Delta S$ more positive (scheme A) (11-15).

One major goal in drug design is to make drugs which bind to their target with the highest binding affinity. Higher affinity is known to results in lower dosage requirements, greater specificity, better drug efficacy, reduced side effects, and less drug resistance. The thermodynamic signatures for each drug binding to its own target can be determined by ITC and a QSAR (quantitative structure-activity relationship) can be derived based on change in the enthalpy (ΔH) and change in the entropy (ΔS) characterizing the binding interaction. It has been already shown (Figure 119-scheme A) that hydrogen bonding formation and favorable van der Waals interactions between drug and target results in a favorable enthalpy of binding. Hydrogen bonding formation is a result of optimal placement of hydrogen bond donor and acceptor groups on the drug and target, and is highly directional and specific (21-25). In the same time hydrogen bond donors and receptors may be in close proximity in a crystal structure but may have little effect on binding affinity, due to conformational effects (20-21). A common strategy in drug design is the addition of non-polar groups to increase the hydrophobic interactions between target and compound. However, hydrophobic interactions are non-specific compared with hydrogen bonding. Freire's research demonstrated that drugs with favorable enthalpy and entropy have highest affinity, due to specific hydrogen bond

formation and van der Waals interactions as well as hydrophobic interactions with the target (20-24). A better strategy for identifying lead compounds is proposed to be related to finding compounds with favorable ΔH, and further optimize these drugs by addition of hydrophobic groups. This could result in a highly-specific drug with tight binding affinity (21,23, 25).

4.2. MATERIALS AND METHODS

The ITC experiments for some lead compounds were performed at 22^0- 25^0C, using an VP-ITC microcalorimeter from Microcal and the data were fitted using the built-in Microcal-Origin software. The buffer for titration was the same buffer used in kinetics studies, e.g. 0.1 I (ionic strength) potassium phosphate buffer, pH 7.46, containing 0.2 M NaCl and 50 μg/ml BSA as additive). All samples were degassed before the titration using the Thermo-Vac (sample degassing and thermostat) associated with the instrument. All experiments were performed by having in the syringe the ligand (peptide) and in the cell the thrombin (bovine, from Sigma), both equilibrated with the same buffer. The ratio between the peptide (ligand-L) and protein (P) (i.e. L/P) was varied in between different titration experiments from 0.01-0.05 (as molar ratios) in order to determine the best signal/noise ratio for the heat released/injection during titrations. Other parameters were changed such as the amount of protein used in the cell (between 8-80 μM) and the concentration of peptide stock in the syringe (0.5-2 mM peptide). The variations of these parameters resulted in better assessment of the factor "C" which is described by the following equation: C= Ka [M]n, were Ka is the association constant for any macromolecule interacting with a ligand, [M] is the molarity of the protein and n is the stoichiometry of binding. In ITC (25) C is reported to be 5<C<100 in order to achieve a

good signal/noise, i.e. the concentration of macromolecule [M] is varied so that enough heat is released upon binding a specific ligand such that the integration of the heat released is accurate within the 1-5% experimental error. In our experiments we varied [M] so that C was always between 20-50, assuming that n=1 and varying Ka in the limits determined by the results from kinetics for each peptide inhibitor. 15-25 injections were performed for each peptide individually, having the reference power stable at 15 μcal/sec as suggested by the company specifications for the VP-ITC instrument. The stirring of the syringe was also constant for all experiments at 300 rpm, and the spacing between injections was standardized between 500-800 seconds such that the baseline was achieved after each injection. The volume of ligand was between 5-15 μL as determined by the L/P ratio achieved during titration.

4.3. RESULTS

All titration experiments were characterized by both exothermic and endothermic heat of binding and saturability, confirming the kinetic data which showed that peptides are specific inhibitors for thrombin. A nonspecific binding of peptides to the target thrombin would have been characterized by nonsaturable binding event. In control experiments the peptides were titrated directly into the buffer or the buffer was titrated into buffer (figure 124). The heat released upon titration of peptides into buffer was very small and nonspecific (Figure 124). As can be seen from Figure 120 and 122 the titration experiment determined the exothermic and the endothermic effect for the heat of binding reaction in the case of peptide [D-Phe-Pro-D-Arg-D-Ala-$CONH_2$]. Depending on the L/P ratio during titration the binding was exclusively exothermic (figure 120) or was having both exothermic and endothermic components (figure 122). The thermodynamic

signatures characterizing each of the binding process are presented in Figures 121 and 132.

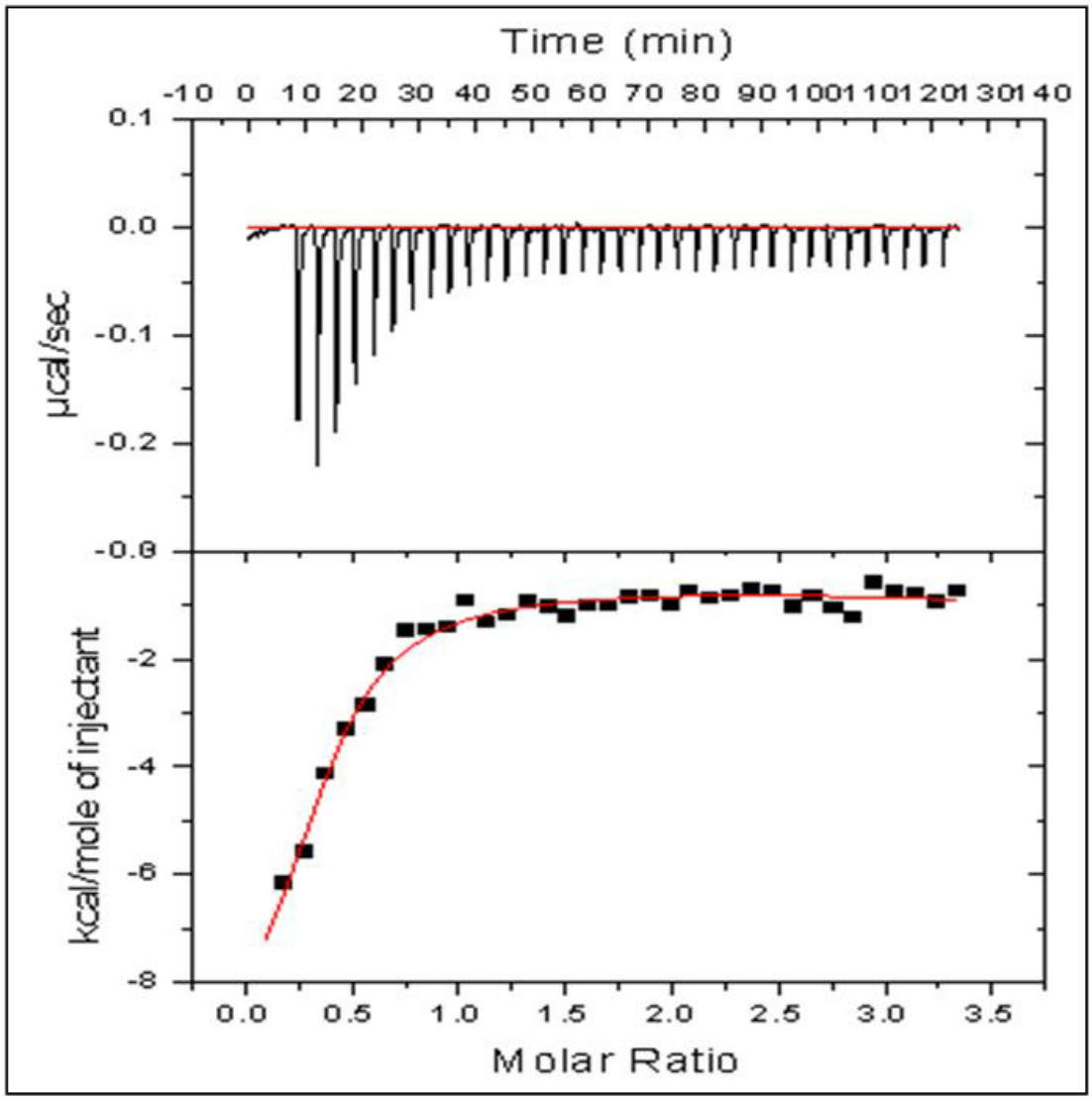


Figure 120. Titration of peptide [D-Phe-Pro-D-Arg-D-Ala-CONH$_2$] into thrombin using 1.03 mM peptide (ligand-L) in the syringe, 0.0084 mM thrombin (protein-P) in the cell and a ratio of L/P of 0.01 as increment during the titration. The raw data are shown as filled data points while the fitted result to "sequential binding site model' is shown as a continuous line. The area underneath each injection peak (top panel) is equal to the total heat released for that injection. When this integrated heat is plotted against the molar ratio of ligand added to macromolecule in the cell, a complete binding isotherm for the interaction is obtained (bottom panel). The molar ratio between ligand and protein for each injection during the titration is plotted on x-axis while the y-axis shows both the changed in the heat/second (µcal/sec) and the transformed data as kcal/mole of injectant.

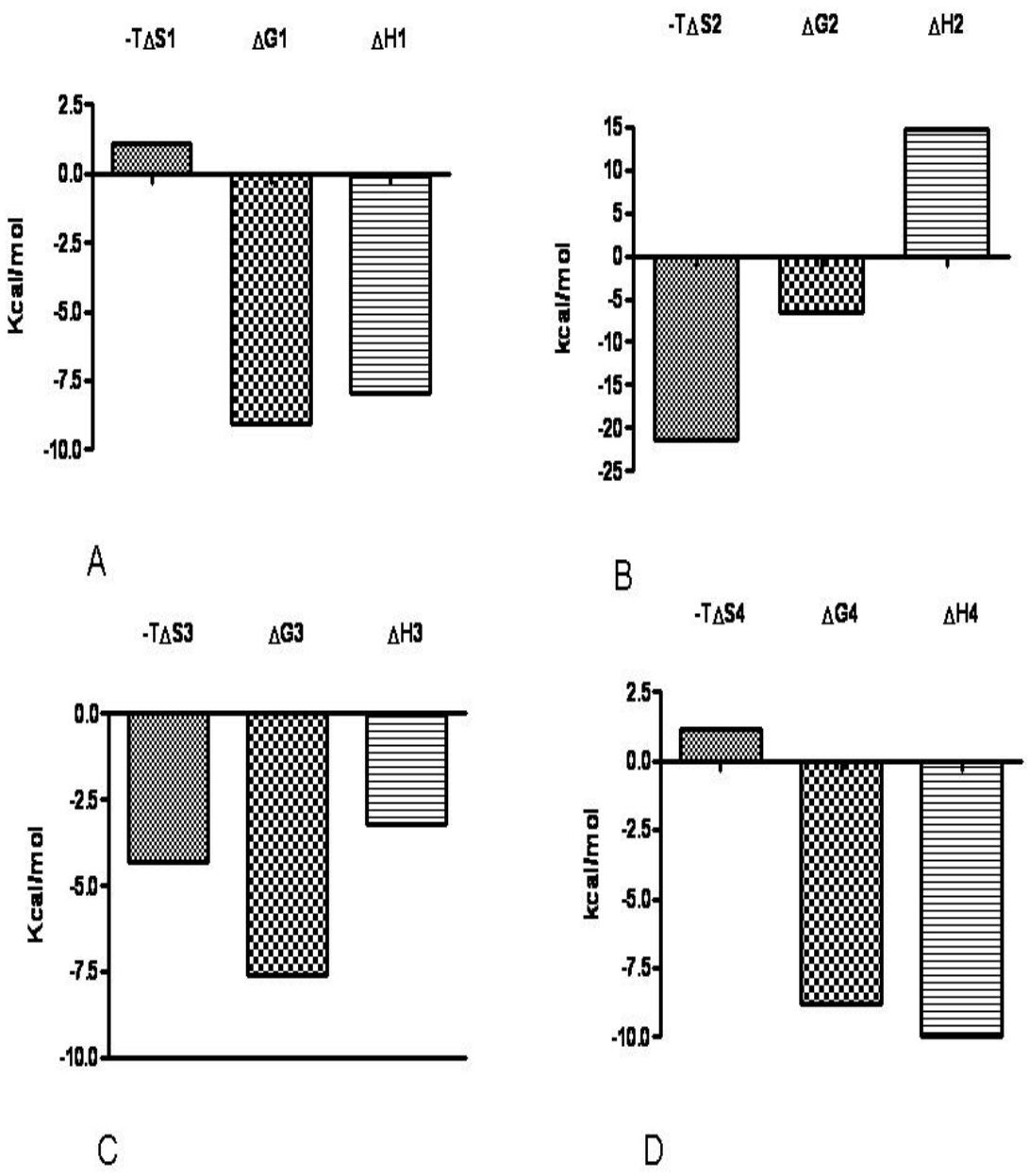


Figure 121. The ITC signatures for the peptide [D-Phe-Pro-D-Arg-D-Ala-$CONH_2$] titrated into thrombin (Figure 120) as are described for each of the 4 sequential binding sites.

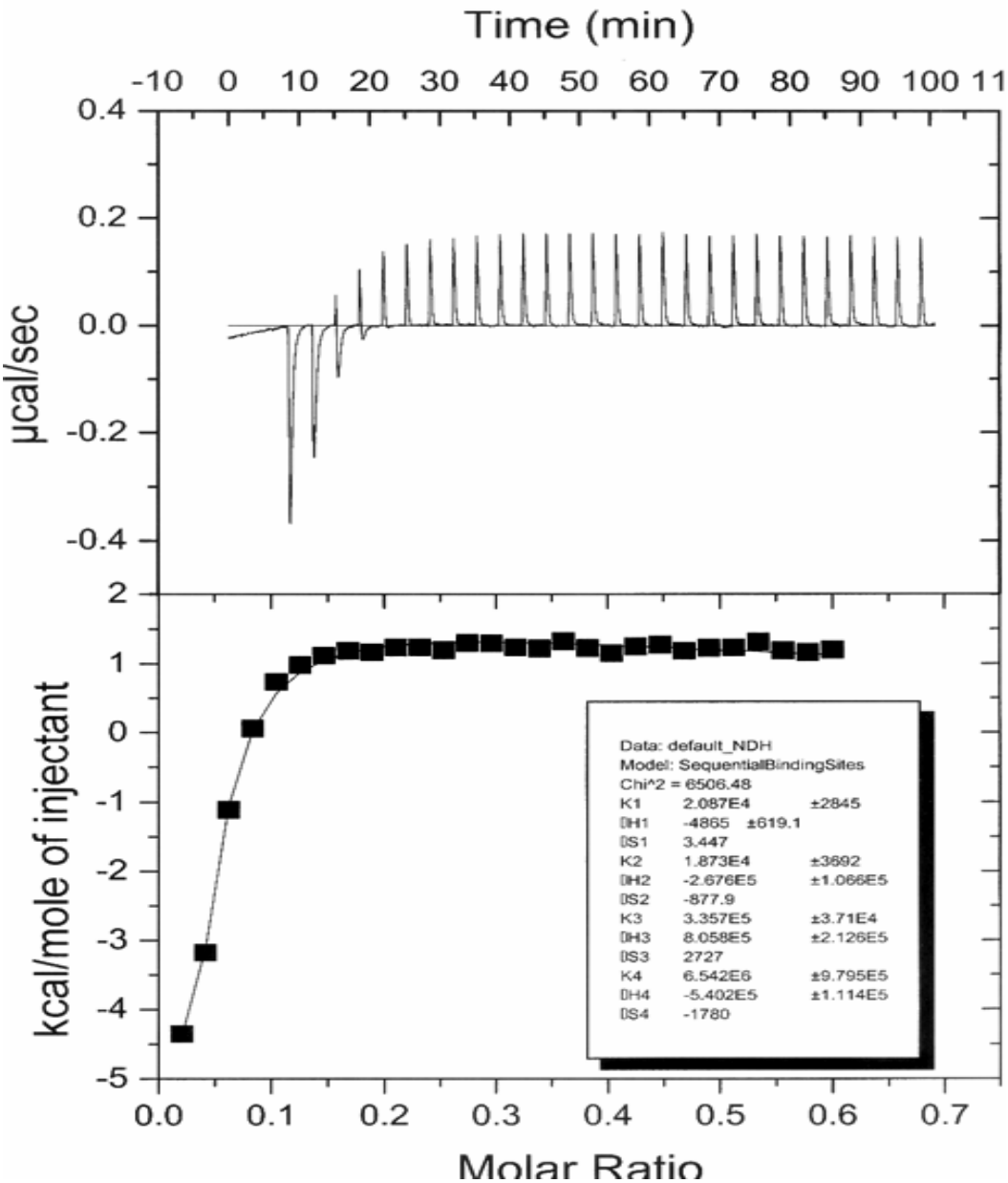


Figure 122. Titration of peptide [D-Phe-Pro-DArg-D-Ala-CONH2] into thrombin using 1.03 mM peptide (L) in the syringe, 0.012 mM thrombin (P) in the cell and a L/P: 0.02 increments. The raw data are shown as filled data points while the fitted result to "sequential binding site model' is shown as a continuous line. The area underneath each injection peak (top panel) is equal to the total heat released for that injection. When this integrated heat is plotted against the molar ratio of ligand added to macromolecule in the cell, a complete binding isotherm for the interaction is obtained (bottom panel). The molar ratio between ligand and protein for each injection during the titration is plotted on x-axis while the y-axis shows both the changed in the heat/second (µcal/sec) and the transformed data as kcal/mole of injectant. The inset picture on the lower panel shows the thermodynamic parameters resulted from the fit of experimental data to a "sequential binding site" model (4 sequential binding sites in this case).

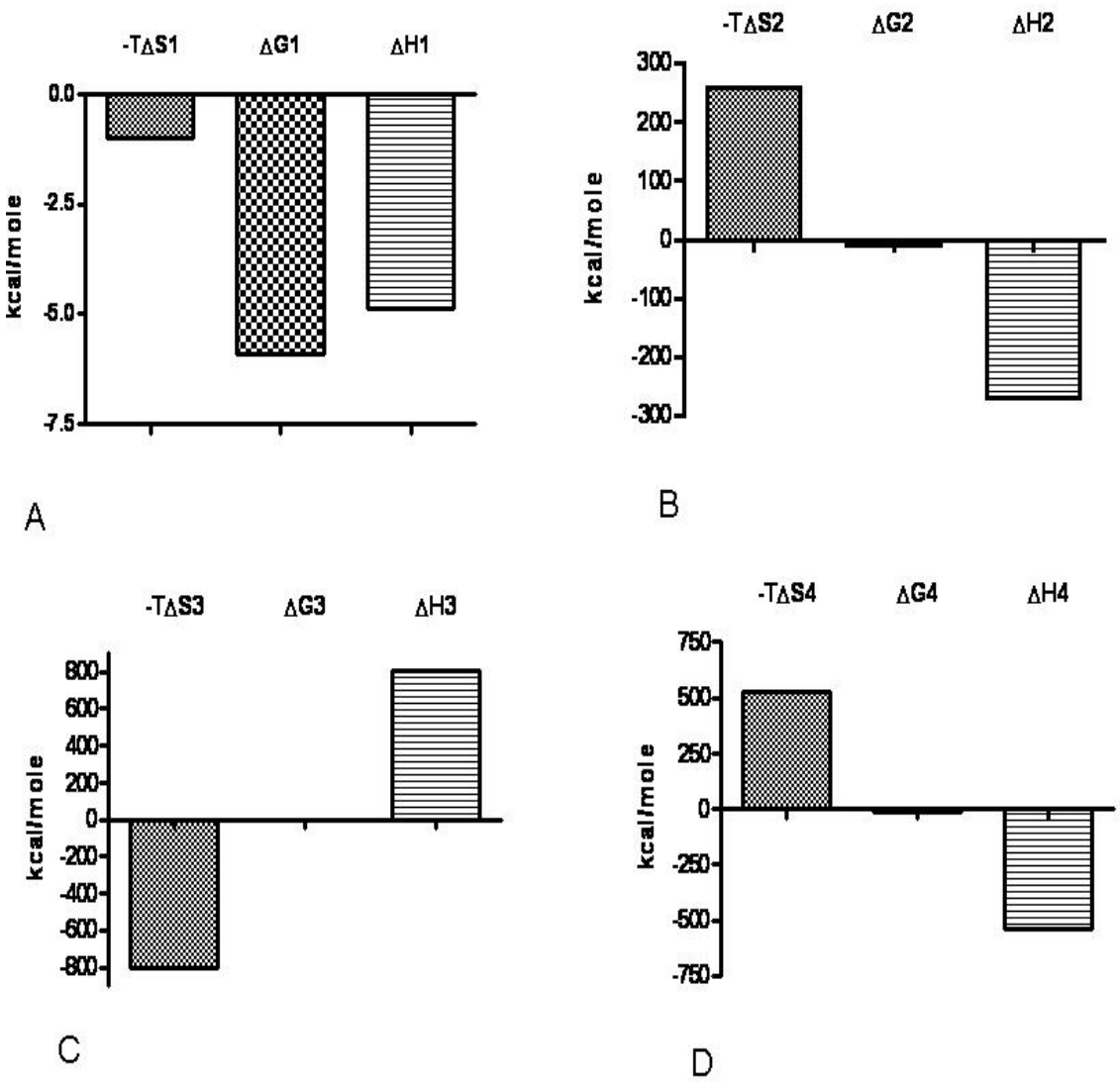


Figure 123. The ITC signatures for the peptide [D-Phe-Pro-D-Arg-D-Ala-$CONH_2$] titrated into thrombin (Figure 122) as are described for each of the 4 sequential binding sites.

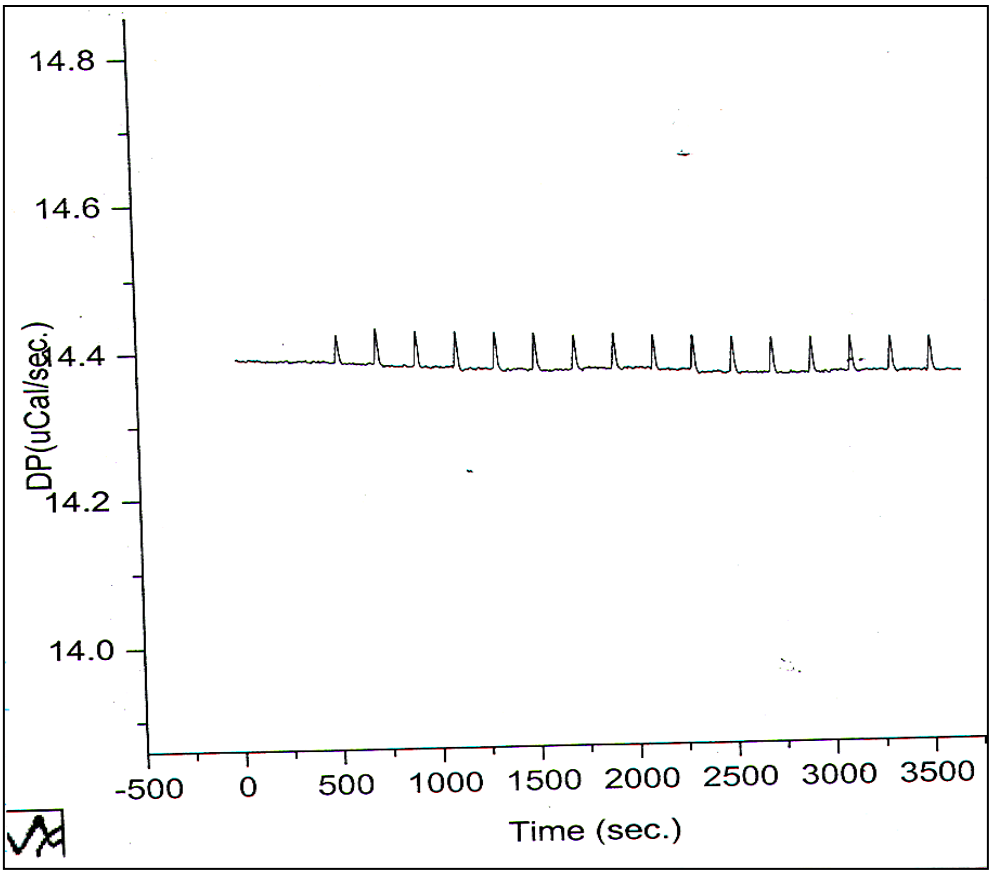


Figure 124. Titration of peptide [DPhe-Pro-D-Arg-D-Ala-CONH2] into buffer at 25°C.

4.3. DISCUSSIONS

For most cases the affinity constant determined from ITC experiments was at the same order of magnitude with the affinity constant determined from kinetics studies. From example in the case of the peptide [D-Phe-Pro-D-Arg-D-Ala] the equilibrium constant determined from kinetics was around 485 x 10^3 L/mol while one of the equilibrium constant determined from ITC experiments was 335.7 x 10^3 L/mol. In the case of peptide [D-Phe-Pro-D-Arg-Cys] the equilibrium constants (association constants) determined from ITC were 72.1x10^3, 111.1 x 10^3 and 140 x10^3 (for 3 sequential binding

sites) while the equilibrium constant determined from kinetics was 80 x 10^3 L/mol, suggesting a good agreement within 5% experimental error with the affinity constant determined from ITC data. In conclusion the ITC titration experiments were very reliable method to assess the binding affinities between different peptide inhibitors and thrombin. Furthermore the detailed thermodynamic analysis of the ITC binding isotherm (as shown in figures 121 and 123) showed that all titration experiments were characterized by both exothermic and endothermic heat of binding and saturability, confirming the kinetic data which showed that peptides are specific inhibitors for thrombin (additional ITC data are presented in Appendix XIII). A nonspecific binding of peptides to the target thrombin would have been characterized by nonsaturable binding event. In control experiments the peptides were titrated directly into the buffer or the buffer was titrated into buffer. The heat released upon titration of peptides into buffer was very small and nonspecific (Figure 124).

The data were best fitted to a "sequential binding sites" suggesting that the peptides may bind to thrombin with a positive allosteric mechanism, in which the peptide binding to one site may increase the affinity of peptide to another site on thrombin. Since the kinetic data showed both competitive and mixed inhibition we assumed that the "sequential binding sites" are related to overlapping sites of the active site binding pocket, spanning the subpockets S3, S2, S1 and S1' for which the peptide inhibitors were designed to interact. Thus, the overlapping site may results from conformational changes on the active site pocket induced by peptide binding. This hypothesis was partially confirmed by the ITC results (Figure 121-A, compared with Figure 119) were the positive ($-T\Delta S$) term was accompanied by negative free energy (ΔG) and enthalpy (ΔH).

Other fitting procedures were tried, like fitting to two independent binding sites or one single-binding sites, but the statistical analysis of the binding curves based on Chi^2) were shown that the models were not reliable (the fitting procedure is performed till Chi^2 achieve the lowest value during consecutive iterations procedure).

The fitted data to "sequential" binding sites are consistent with the kinetics results which showed that the peptides are not binding only to the active site (competitive inhibition) but there is at least one more independent or "linked" binding site since the kinetics were shown mixed inhibition at higher peptide concentration (10Ki) (Figure 118).

REFERENCES

1. I.R. McKinnon, L. Fall, A. Parody-Morreale and S.J. Gill, A twin titration microcalorimeter for the study of biochemical reactions, *Anal. Biochem.* **139** (1984), pp. 134–139.

2. T. Wiseman, S. Williston, J.F. Brandts and L.N. Lin, Rapid measurement of binding constants and heats of binding using a new titration calorimeter, *Anal. Biochem.* **179** (1989), pp. 131–137.

3. E. Freire, O.L. Mayorga and M. Straume, Isothermal titration calorimetry, *Anal. Chem.* **62** (1990), pp. 950A–959A.

4. M. El Harrous, S.J. Gill and A. Parody-Morreale, Description of a new Gill titration calorimeter for the study of biochemical reactions: I. Basic response of the instrument, *Meas. Sci. Technol.* **5** (1994), pp. 1065–1070.

5. M. El Harrous, O.L. Mayorga and A. Parody-Morreale, Description of a new Gill titration calorimeter for the study of biochemical reactions: II. Operational characterization of the instrument, *Meas. Sci. Technol.* **5** (1994), pp. 1071–1077.

6. A. Velazquez-Campoy, O. Lopez-Mayorga and M.A. Cabrerizo-Vilchez, Development of an isothermal titration microcalorimetric system with digital control and dynamic power Peltier compensation: I. Description and basic performance, *Rev. Sci. Instrum.* **71** (2000), pp. 1824–1831.

7. A. Velazquez-Campoy, O. Lopez-Mayorga and M.A. Cabrerizo-Vilchez, Development of an isothermal titration microcalorimetric system with digital control and dynamic power Peltier compensation: II. Characterization and operation mode. Myoglobin adsorption onto polymeric latex particles, *Rev. Sci. Instrum.* **71** (2000), pp. 1832–1840.

8. A. Velázquez-Campoy and E. Freire, Isothermal titration calorimetry: measuring intermolecular interactions. In: R. Simpson, Editor, *Proteins and proteomics: a laboratory manual*, Cold Spring Harbor Laboratory Press, New York (2003), pp. 882–892.

9. A. Velazquez-Campoy, S.A. Leavitt and E. Freire, Characterization of protein–protein interactions by isothermal titration calorimetry, *Methods Mol. Biol.* **261** (2004), pp. 35–54.

10. H. Naghibi, A. Tamura and J.M. Sturtevant, Significant discrepancies between van't Hoff and calorimetric enthalpies, *Proc. Natl. Acad. Sci. U. S. A.* **92** (1995), pp. 5597–5599.

11. Y. Liu and J.M. Sturtevant, Significant discrepancies between van't Hoff and calorimetric enthalpies. II, *Protein Sci.* **12** (1995), pp. 2559–2561.

12. Y. Liu and J.M. Sturtevant, Significant discrepancies between van't Hoff and calorimetric enthalpies. III, *Biophys. Chemist.* **64** (1997), pp. 121–126.

13. J.B. Chaires, Possible origin of differences between van't Hoff and calorimetric enthalpy estimates, *Biophys. Chemist.* **64** (1997), pp. 15–23.

14. J.R. Horn, J.F. Brandts and K.P. Murphy, Van't Hoff and calorimetric enthalpies: II. Effects of linked equilibria, *Biochemistry* **41** (2002), pp. 7501–7507.

15. I. Luque and E. Freire, Structure-based prediction of binding affinities and molecular design of peptide ligands, *Methods Enzymol.* **295** (1998), pp. 100–127.

16. I. Luque and E. Freire, Structural parameterization of the binding enthalpy of small ligands, *Proteins* **49** (2002), pp. 181–190.

17. A. Velazquez-Campoy, M.J. Todd and E. Freire, HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity, *Biochemistry* **39** (2000), pp. 2201–2207.

18. M.J. Todd, I. Luque, A. Velazquez-Campoy and E. Freire, Thermodynamic basis of resistance to HIV-1 protease inhibition: calorimetric analysis of the V82F/I84V active site resistant mutant, *Biochemistry* **39** (2000), pp. 11876–11883.

19. A. Velazquez-Campoy, I. Luque, M.J. Todd, M. Milutinovich, Y. Kiso and E. Freire, Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor, *Protein Sci.* **9** (2000), pp. 1801–1809.

20. A. Velazquez-Campoy, Y. Kiso and E. Freire, The binding energetics of first- and second-generation HIV-1 protease inhibitors: implications for drug design, *Arch. Biochem. Biophys.* **390** (2001), pp. 169–175.

21. A. Velazquez-Campoy and E. Freire, Incorporating target heterogeneity in drug design, *J. Cell. Biochem., Suppl.* **37** (2001), pp. 82–88.

22. E. Freire, Designing drugs against heterogeneous targets, *Nat. Biotechnol.* **20** (2002), pp. 15–16.

23. A. Velazquez-Campoy, S. Vega and E. Freire, Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes, *Biochemistry* **41** (2002), pp. 8613–8619.

24. H. Ohtaka, A. Velazquez-Campoy, D. Xie and E. Freire, Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease, *Protein Sci.* **11** (2002), pp. 1908–1916.

25. M.L. Doyle and P. Hensley, Tight ligand binding affinities determined from thermodynamic linkage to temperature by titration calorimetry, *Methods Enzymol.* **295** (1998), pp. 88–99.

Chapter 5

Structure-conformation analysis of peptides inhibitors for thrombin using circular dichroism spectroscopy

5.1. INTRODUCTION

With a view towards better understanding the structural requirements of the peptide inhibitors of thrombin the conformation of some peptide lead inhibitors (see tables 14 and 15 in Chapter 3) has been studied by circular dichroism (CD). All peptides exhibited a high dependence of structure on the sequence of amino-acids at P3 and P1' positions (see tables 14 and 15-Chapter 3). The tetrapeptides tend to adopt beta-turns in solvents like H2O and phosphate buffer at pH=7.4 with the same composition as the buffer used for kinetics of thrombin inhibition (see Chapter 3). We wanted to propose a structure-conformation relationship which could explain why some peptides were more potent in inhibiting thrombin when changes at P3 and P1' positions within the peptide chain were performed (as reflected in the experimentally determined Kis (see tables 14 and 15). We hypothesized that the amino-acids at the specified positions could affect the conformation of the peptide by imposing specific intramolecular hydrogen bonds which could give rise to small alpha-helices or beta turns. Other known peptides such as spontaneously folding beta-hairpin peptide (Lys-Lys-Tyr-Thr-Val-Ser-Ile-Asn-Gly-Lys-Lys-Ile-Thr-Val-Ser-Ile) and related cyclic (cyclo-Gly-Lys-Tyr-Ile-Asn-Gly-Lys-Ile-Ile-Asn) and linear (Ser-Ile-Asn-Gly-Lys) were reported to fold in known secondary structure elements such as beta-hairpin (1-15). The beta-hairpin peptides displayed a CD minimum at 216 nm (14). These are the first CD data reported on peptides with the

sequence space D-Phe-Pro-D-Arg-P1'-CONH2 and are showing a good correlation between the structure of folded peptide and its in vitro potential for inhibiting thrombin.

5.2. MATERIALS AND METHODS

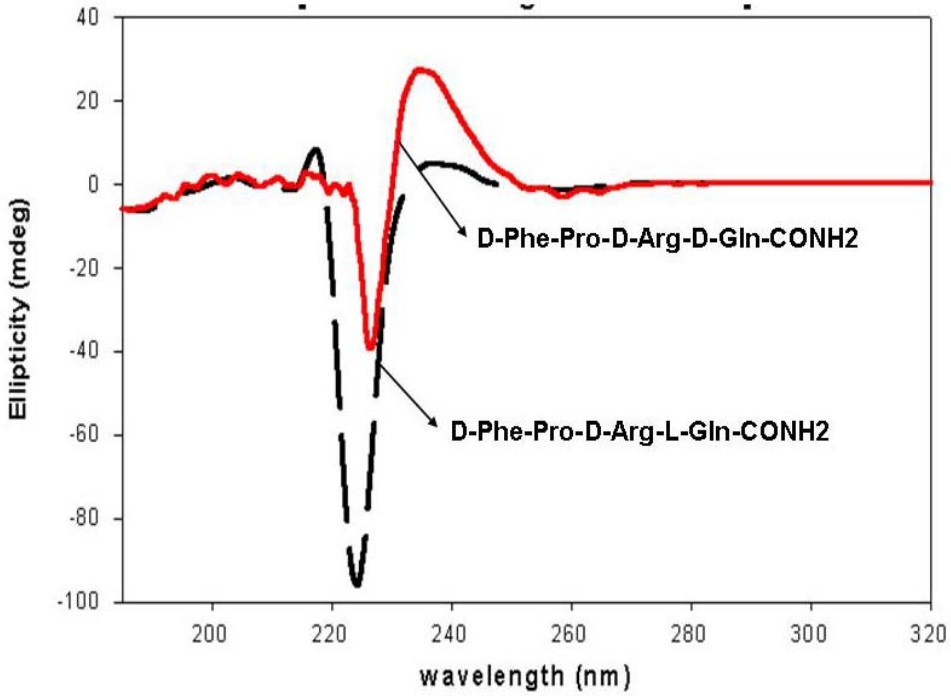
All CD data were collected using a Jasco Spectropolarimeter using the following parameters: scanning mode: 195 nm-320 nm, resolution: 0.5 nm, response time: 1 second. The data were collected as 15 scans and the final spectra are presented as an average over 15 independent accumulations. The peptides were solubilized in 30 mM KH_2PO_4 buffer, pH=5.5 or in the phosphate buffer used in all kinetics studies (pH=7.46, of 0.1 I (ionic strength), containing 0.2 M NaCl. The spectra didn't show any changes with the shift in pH from pH=5.5 to pH 7.46 for all compounds tested, as was expected based on their amino-acid sequence. The Arg side chain remained protonated for all the conditions tested while the amino acids at P1' position didn't show any change in their protonation status either through the range of pH tested, since they were hydrophobics and polar neutral with pKa of side chain above pH 7.46. The spectra presented in Figure 32 are for compounds solubilized in phosphate buffer without salt at pH 5.5. The presence of salt had no effect upon the spectral features.

5.3. RESULTS

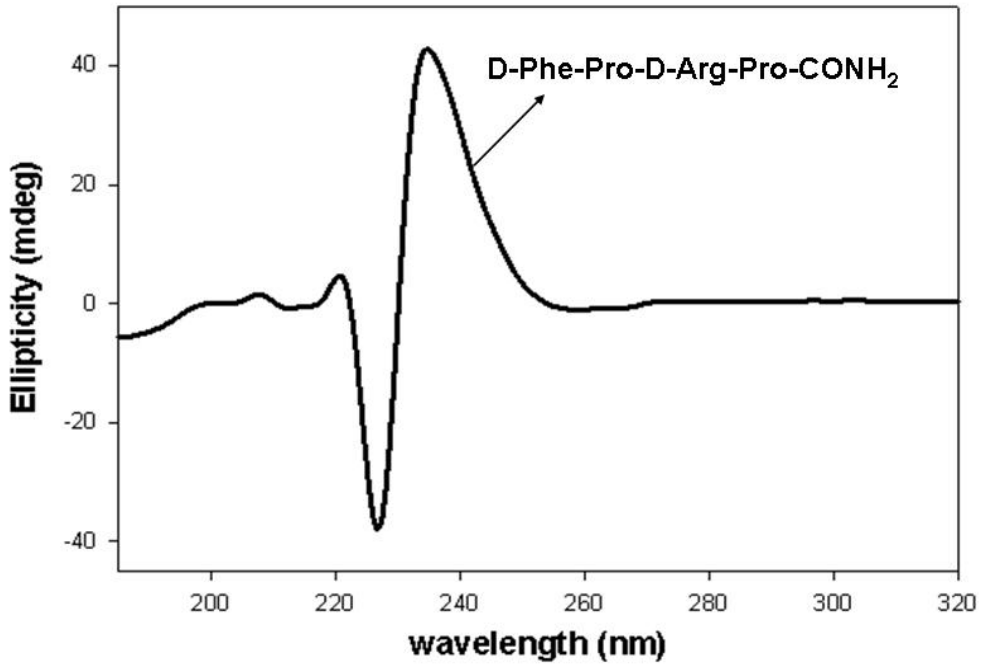
The peptides were dissolved in 30 mM phosphate buffer at pH 5.5 or in the same buffer used in kinetic studies for all CD experiments as described in the Materials and Methods and the spectra were collected as 15 independent accumulations which were further averaged using the built-in software provided by Jasco Spectropolarimeter.

Figure 125 presents the CD scans of some peptides tested in kinetic studies. A SAR between the global conformation adopted in solution and their inhibitory potential was under current investigation.

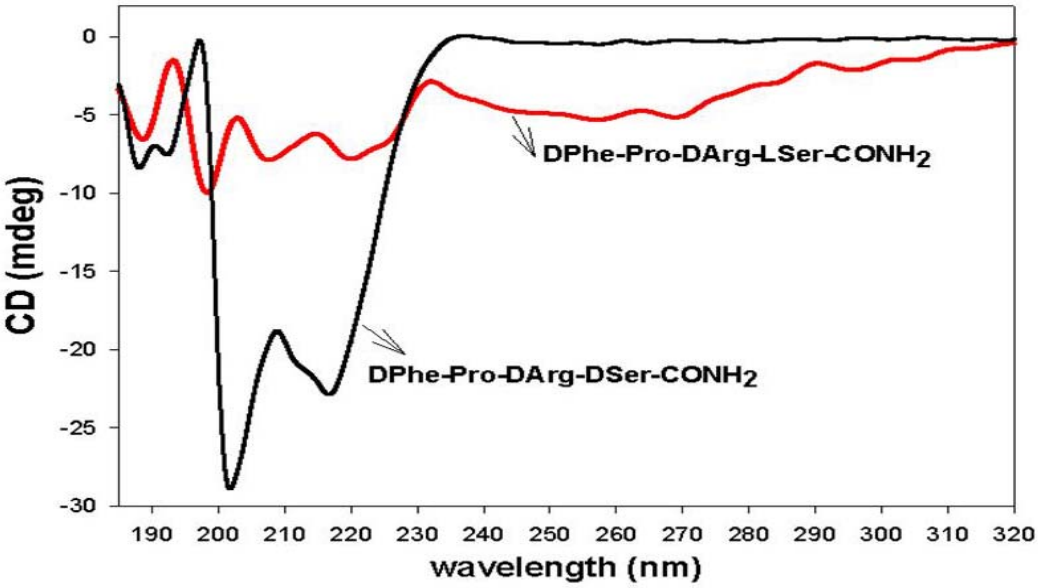
A



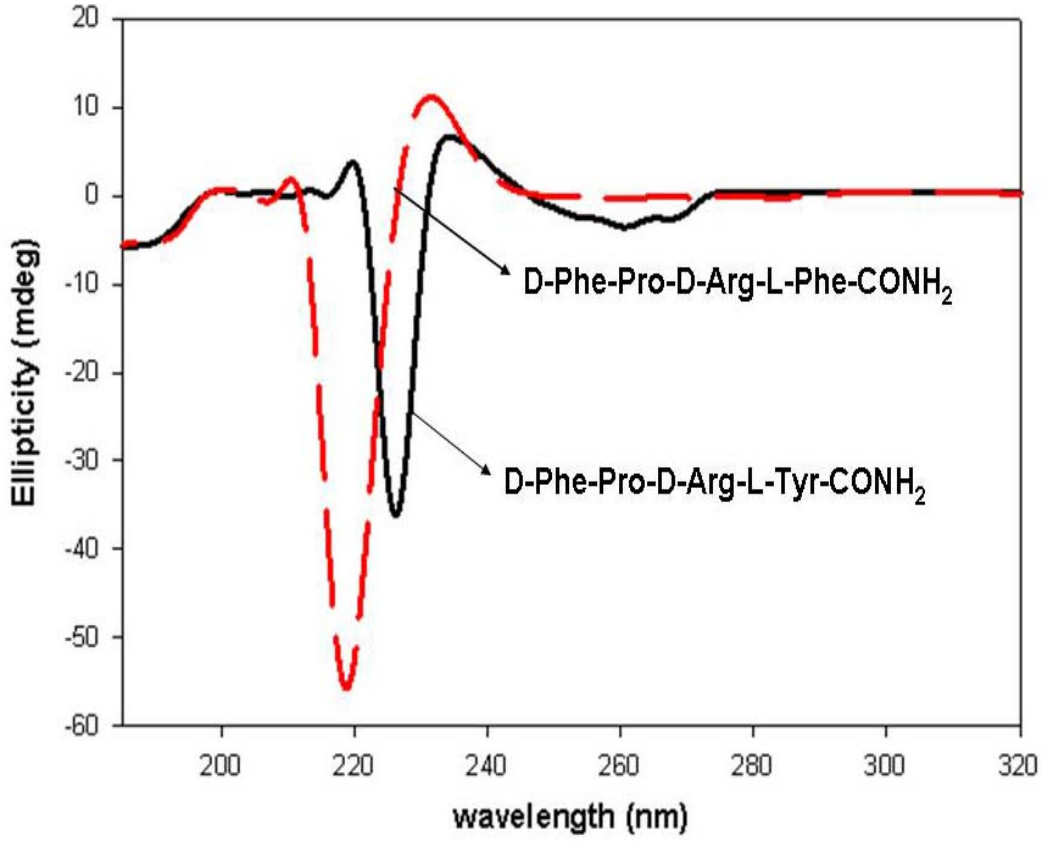
B


D-Phe-Pro-D-Arg-Pro-CONH2
Ellipticity (mdeg)
wavelength (nm)
40
20
0
-20
-40
200
220
240
260
280
300
320

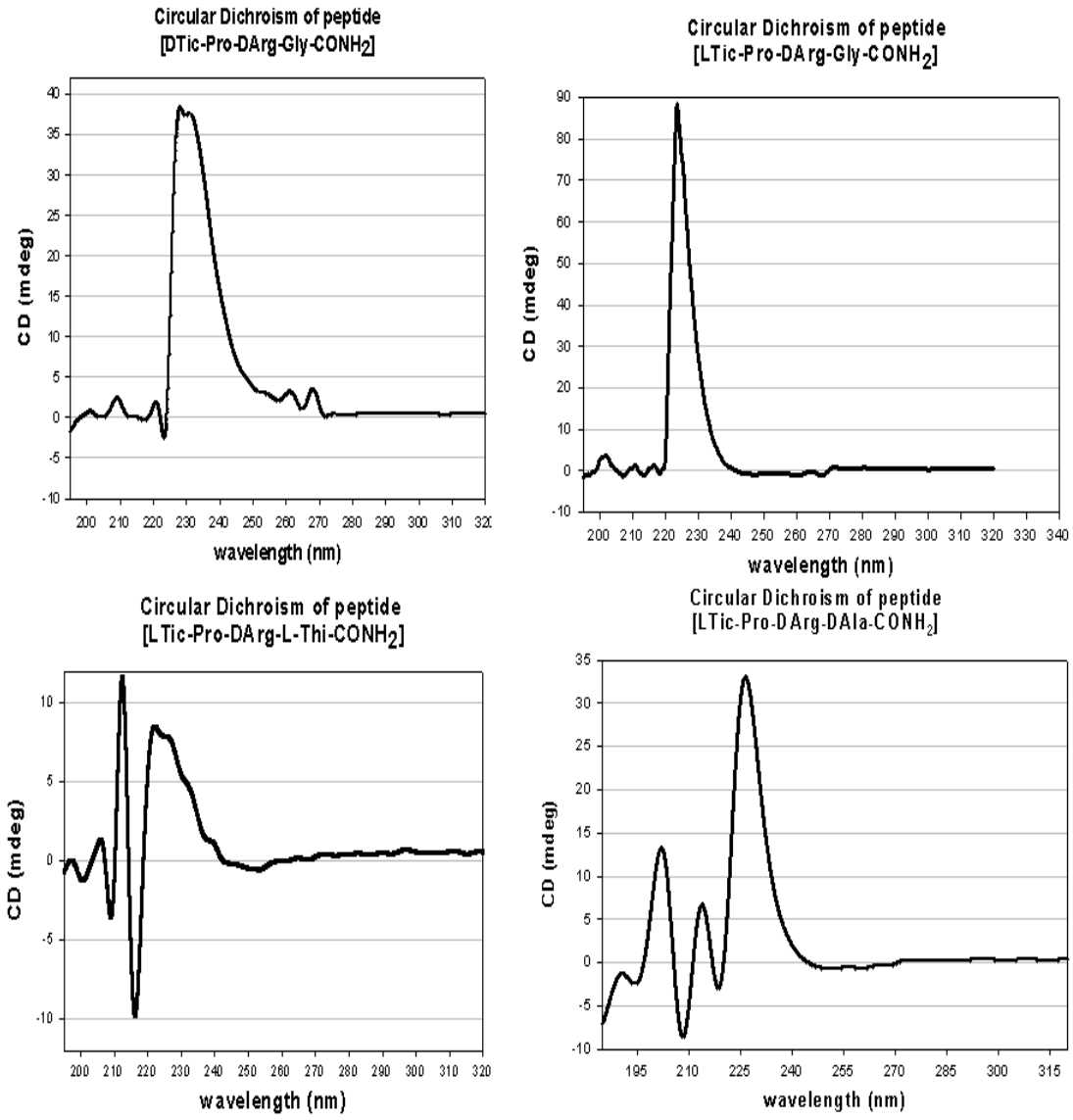
C


DPhe-Pro-DArg-LSer-CONH2
DPhe-Pro-DArg-DSer-CONH2
CD (mdeg)
wavelength (nm)
0
-5
-10
-15
-20
-25
-30
190
200
210
220
230
240
250
260
270
280
290
300
310
320

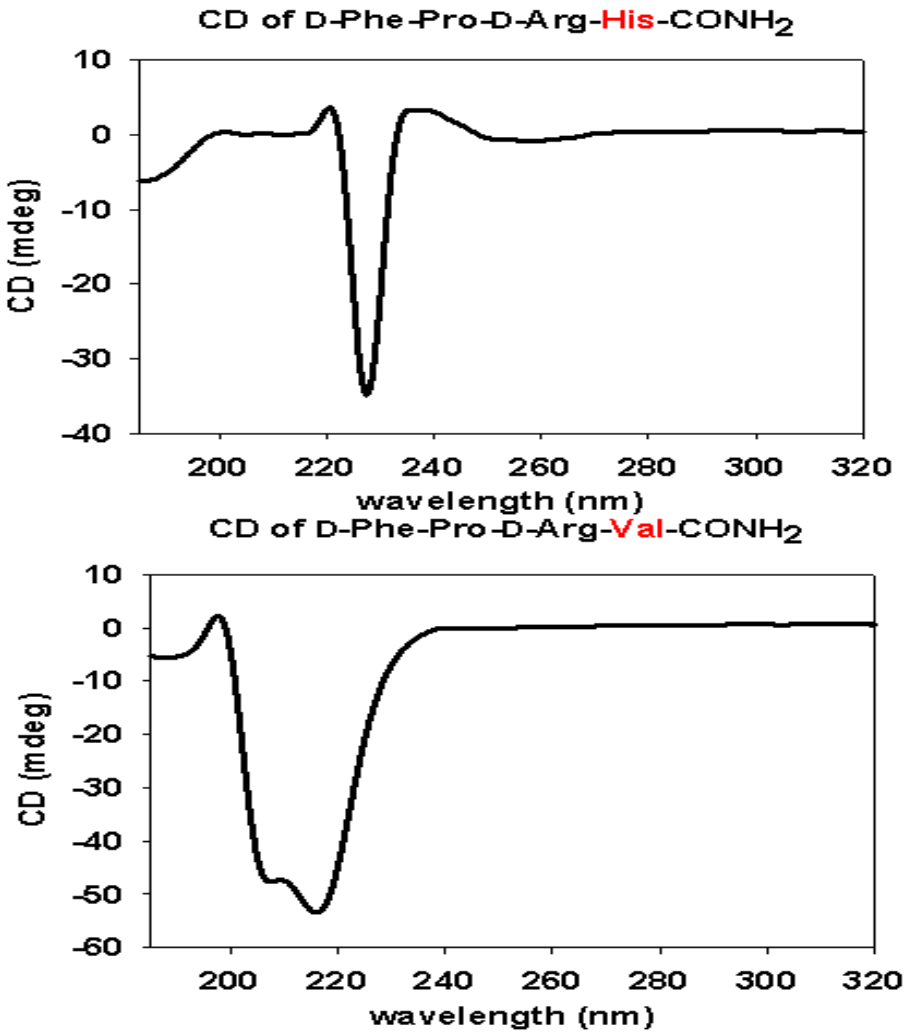
D


20
10
0
-10
-20
-30
-40
-50
-60
Ellipticity (mdeg)
200
220
240
260
280
300
320
wavelength (nm)
D-Phe-Pro-D-Arg-L-Phe-CONH2
D-Phe-Pro-D-Arg-L-Tyr-CONH2

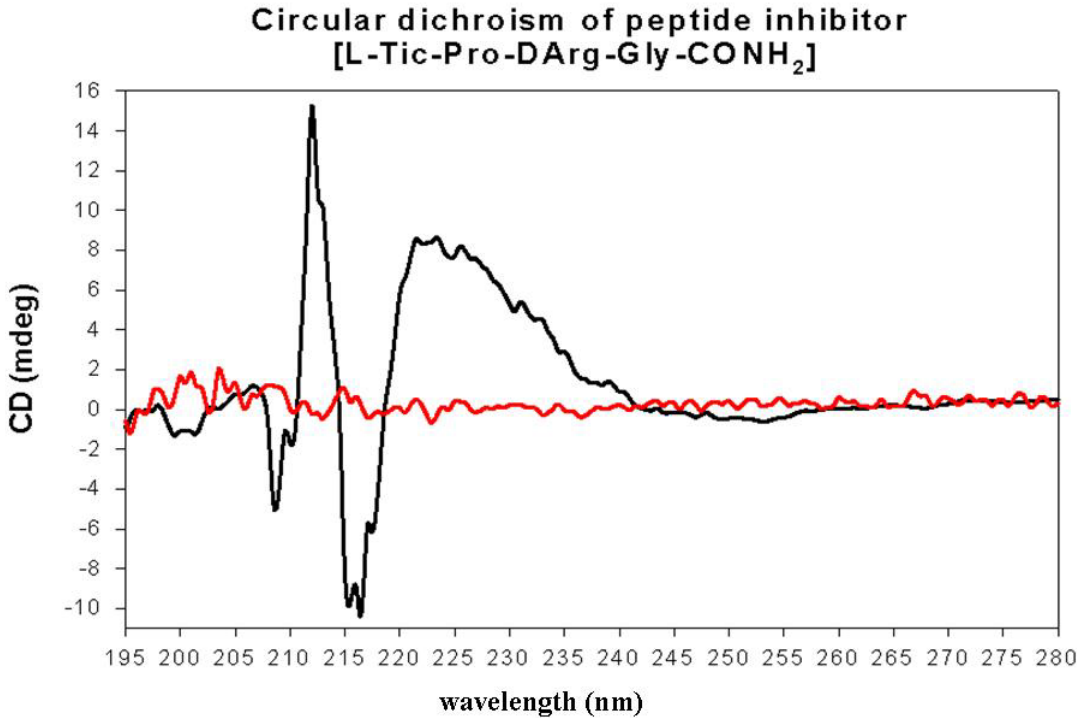
E


Circular Dichroism of peptide
[DTic-Pro-DArg-Gly-CONH2]
CD (mdeg)
wavelength (nm)
Circular Dichroism of peptide
[LTic-Pro-DArg-Gly-CONH2]
CD (mdeg)
wavelength (nm)
Circular Dichroism of peptide
[LTic-Pro-DArg-L-Thi-CONH2]
CD (mdeg)
wavelength (nm)
Circular Dichroism of peptide
[LTic-Pro-DArg-DAla-CONH2]
CD (mdeg)
wavelength (nm)

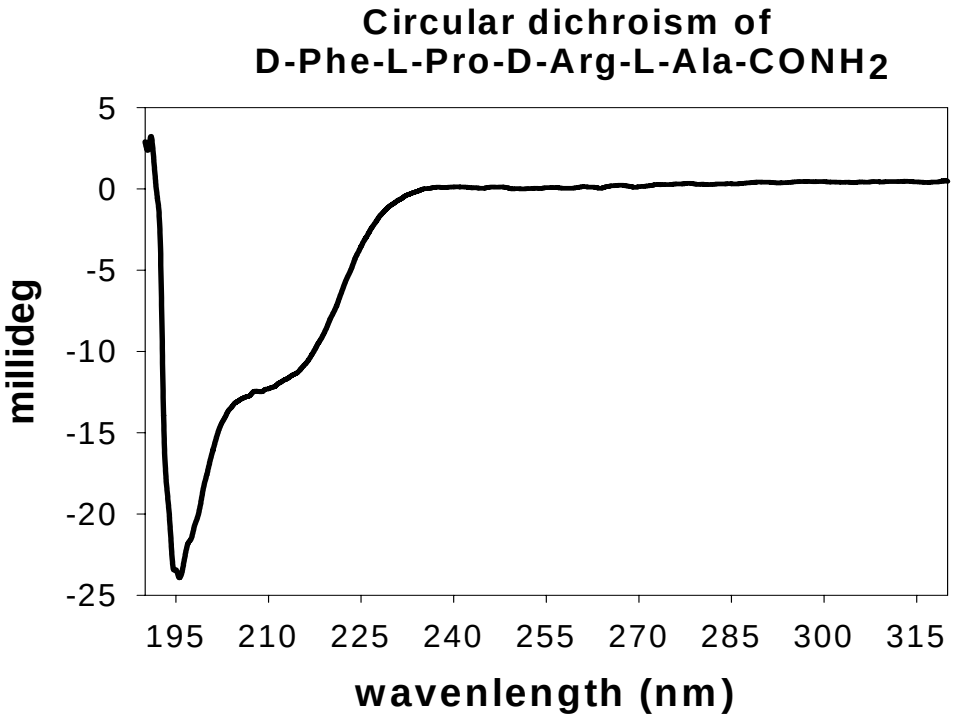
F


CD of D-Phe-Pro-D-Arg-His-CONH2
CD (mdeg)
10
0
-10
-20
-30
-40
200
220
240
260
280
300
320
wavelength (nm)
CD of D-Phe-Pro-D-Arg-Val-CONH2
CD (mdeg)
10
0
-10
-20
-30
-40
-50
-60
200
220
240
260
280
300
320
wavelength (nm)

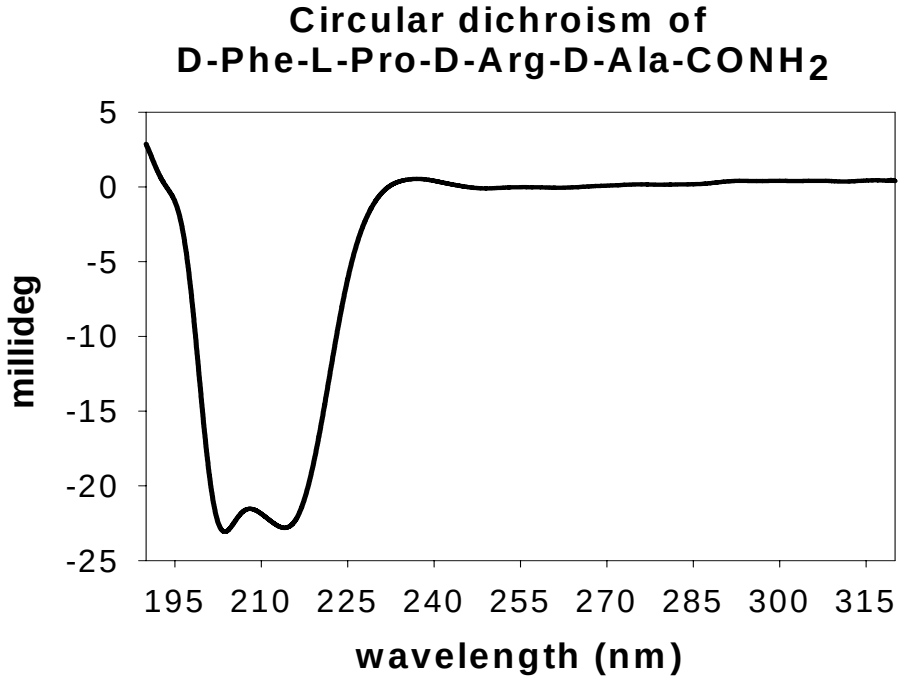
G


Circular dichroism of peptide inhibitor
[L-Tic-Pro-DArg-Gly-CONH2]
CD (mdeg)
16
14
12
10
8
6
4
2
0
-2
-4
-6
-8
-10
195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280
wavelength (nm)

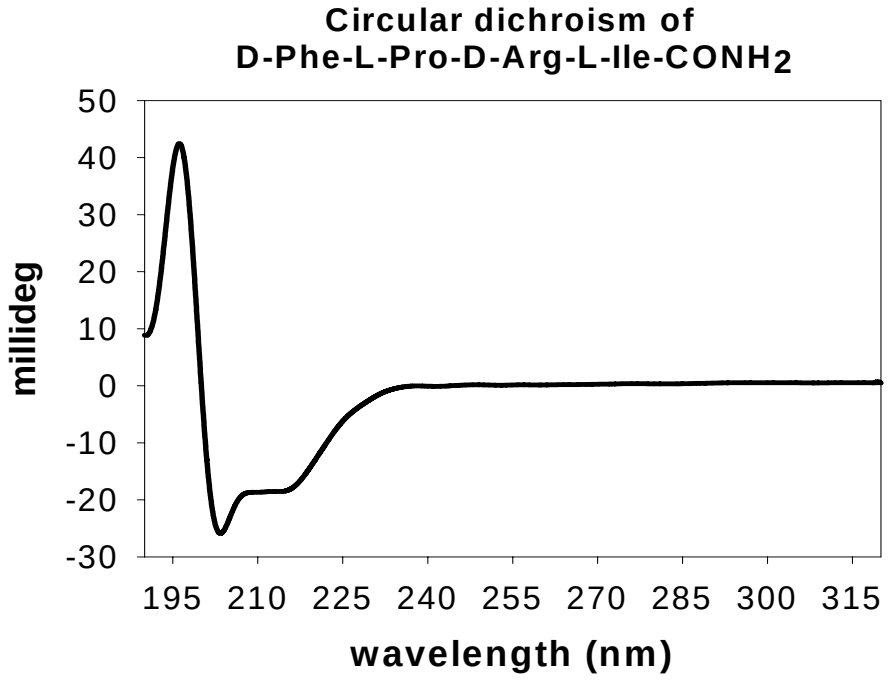
H


Circular dichroism of
D-Phe-L-Pro-D-Arg-L-Ala-CONH2
millideg
5
0
-5
-10
-15
-20
-25
195 210 225 240 255 270 285 300 315
wavenlength (nm)

I


Circular dichroism of
D-Phe-L-Pro-D-Arg-D-Ala-CONH2
millideg
5
0
-5
-10
-15
-20
-25
195 210 225 240 255 270 285 300 315
wavelength (nm)

J


Circular dichroism of
D-Phe-L-Pro-D-Arg-L-Ile-CONH2
millideg
50
40
30
20
10
0
-10
-20
-30
195 210 225 240 255 270 285 300 315
wavelength (nm)

K

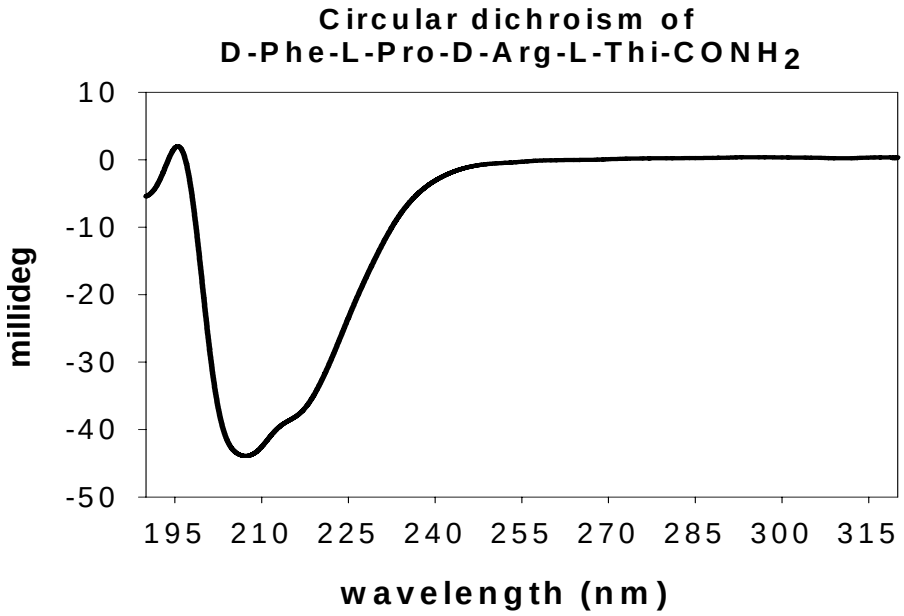


Figure 125(A-K): Circular dichroism (CD) of different peptides tested as being lead inhibitors for thrombin. All spectra were collected in the buffer used for thrombin inhibition kinetics without BSA and were corrected for any buffer effect by subtracting the buffer from the original sample spectrum (in G the spectrum of the sample is superimposed over the buffer spectrum (horizontal, red line); the buffer didn't have significant absorption in the region 195-320 nm).

5.4. DISCUSSIONS

Our circular dichroism investigations suggest that the D-Arg- in i+2 position followed by D-amino acids (polar and neutral hydrophobic like D-Thr, D-Gln, D-Ser and D-Ala) or L-Pro in i+3 position favors beta turn (I and III), beta hairpin or alpha-helical structures in solution at low and neutral pH. The literature reported characteristics for type II β-turn are: weak, red-shifted negative nπ*-near 225 nm; strong positive band with maximum between 200-205 nm; strong negative band between 180-190 nm (1-20)

while the characteristics of type I and III β-turn resemble α-helix, with a negative nπ* band and a negative ππ*-couplet (15-21) (see the CD of peptides D-Phe-Pro-D-Arg-Val-CONH2, Dhe-Pro-D-Arg-D-Ser-CONH2 and figure 125 (C, H, I). Replacement of D- with L-amino acids in i+3 position was accompanied by a significant lost in the beta turn structure with a shift toward an unorganized like structure with the exception of L-amino acids with high alpha helix propensity (such as L-Ile). The aromatics (L-Phe, L-Tyr and L-Trp) in i+3 position are disturbing the beta turn and alpha helix structure while the replacement of D-Phe with L/D-Tic at i-th position is accompanied by a shift toward beta-strand-like structure (see figure 125 (D, E and G).

Some peptides however were shown the alpha helical characteristic (such as D-Phe-Pro-D-Arg-Ile-CONH2, D-Phe-Pro-D-Arg-Thi-CONH2) supporting the view that a pre-organized secondary structure is more active toward inhibiting thrombin (the above mentioned peptides with induced secondary structure are one of the lead compounds inhibitors for thrombin, with Ki in the range of 8-1 μM (see tables 14 and 15).

REFERENCES

1. Blanco, F.J., Jimenez, M.A., Herranz, J., Rico, M., Santoro, J., and Nieto, J.L. 1993. NMR evidence of a short linear peptide that folds into a ß-hairpin in aqueous solution. *J. Am. Chem. Soc.* **115:** 5887–5888.

2. Blanco, F.J., Jimenez, M.A., Pineda, A., Rico, M., Santoro, J., and Nieto, J.L. 1994a. NMR solution structure of the isolated N-terminal fragment of protein-G B1 domain. Evidence of trifluoroethanol induced native-like ß-hairpin formation. *Biochemistry* **33:** 6004–6014.

3. Blanco, F.J., Rivas, G., and Serrano, L. 1994b. A short linear peptide that folds into a native stable ß-hairpin in aqueous solution. *Nat. Struct. Biol.* **1:** 584–590.

4. Blasie, C.A. and Berg, J.M. 1997. Electrostatic interactions across a ß-sheet. *Biochemistry* **36:** 6218–6222.

5. Chakrabartty, A. and Baldwin, R.L. 1995. Stability of -helices. *Adv. Protein Chem.* **46:** 141–176.

6. Chen, P.Y., Lin, C.K., Lee, C.T., Jan, H., and Chan, S.I. 2001. Effects of turn residues in directing the formation of the ß-sheet and in the stability of the ß-sheet. *Protein Sci.* **10:** 1794–1800.

7. Cochran, A.G., Skelton, N.J., and Starovasnik, M.A. 2001a. Tryptophan zippers: Stable, monomeric ß-hairpins. *Proc. Natl. Acad. Sci.* **98:** 5578–5583.

8. Cochran, A.G., Tong, R.T., Starovasnik, M.A., Park, E.J., McDowell, R.S., Theaker, J.E., and Skelton, N.J. 2001b. A minimal peptide scaffold for ß-turn display: Optimizing a strand position in disulfide-cyclized ß-hairpins. *J. Am. Chem. Soc.* **123:** 625–632.

9. de Alba, E., Jimenez, M.A., Rico, M., and Nieto, J.L. 1996. Conformational investigation of designed short linear peptides able to fold into ß-hairpin structures in aqueous solution. *Fold. Des.* **1:** 133–144.

10. de Alba, E., Rico, M., and Jimenez, M.A. 1997. Cross-strand side-chain interactions versus turn conformation in ß-hairpins. *Protein Sci.* **6:** 2548–2560.

11. Dyson, H.J., Bolinger, L., Feher, V.A., Osterhout, Jr., J.J., Yao, J., and Wright, P.E. 1998. Sequence requirements for stabilization of a peptide reverse turn in water solution—Proline is not essential for stability. *Eur. J. Biochem.* **255:** 462–471.

12. Falcomer, C.M., Meinwald, Y.C., Choudhary, I., Talluri, S., Milburn, P.J., Clardy, J., and Scheraga, H.A. 1992. Chain reversals in model peptides: Studies of cystine-containing cyclic peptides. 3. Conformational free energies of cyclization of tetrapeptides of sequence. *J. Am. Chem. Soc.* **114:** 4036–4042.

13. Favre, M., Moehle, K., Jiang, L., Pfeiffer, B., and Robinson, J.A. 1999. Structural mimicry of canonical conformations in antibody hypervariable loops using cyclic peptides containing a heterochiral diproline template. *J. Am. Chem. Soc.* **121:** 2679–2685.

14. Freund, S.M., Wong, K.B., and Fersht, A.R. 1996. Initiation sites of protein folding by NMR analysis. *Proc. Natl. Acad. Sci.* **93:** 10600–10603.

15. Gellman, S.H. 1998. Minimal model systems for ß sheet secondary structure in proteins. *Curr. Opin. Chem. Biol.* **2:** 717–725.

16. Gill, S.C. and von Hippel, P.H. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182:** 319–326.

17. Grishina, I.B. and Woody, R.W. 1994. Contributions of tryptophan side chains to the circular dichroism of globular proteins: Exciton couplets and coupled oscillators. *Faraday Discuss.* **99:** 245–262.

18. Gu, H., Kim, D., and Baker, D. 1997. Contrasting roles for symmetrically disposed ß-turns in the folding of a small protein. *J. Mol. Biol.* **274:** 588–596.

19. Gunasekaran, K., Ramakrishnan, C., and Balaram, P. 1997. ß-hairpins in proteins revisited: Lessons for de novo design. *Protein Eng.* **10:** 1131–1141.

20. Hutchinson, E.G. and Thornton, J.M. 1994. A revised set of potentials for ß-turn formation in proteins. *Protein Sci.* **3:** 2207–2216.

21. Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22:** 2577–2637.

Chapter 6

Fourier Transformed Infrared (FTIR) microscopy of *trans*-and *cis*-cinnamoyl peptides

6.1. INTRODUCTION

FTIR microscopy of solid peptides films containing *trans* and *cis*-cinnamic acid at P3 position in the sequence space D-Phe-Pro-D-Arg-P1'-CONH2 was used to characterize the conversion of *trans*–cinnamoyl peptides into their *cis*-cinnamoyl isomer. It has been shown that the aliphatic double bond stretch vibration of the *trans*-cinnamoyl peptide can be shifted 3-4 wavenumbers upon the conversion into the *cis*-isomer (1-10). Our FTIR results are showing that this wavenumber shift is reproducible in peptides were the *trans*-cinnamic acid was incorporated through an amide bond during the solid phase synthesis of tetrapeptides *trans*-cinnamoyl-Pro-D-Arg-P1'-CONH2. The results of FTIR spectra are showing that *trans*-cinnamoyl peptides are having the double bond stretch vibration band (located between 1586-1648 cm^{-1}) (10-15) shifted 3-10 wavenumbers upon exposure to UV light and conversion into *cis*-isomer. This structural evidence was further used to interpret the structure-activity relationship of *trans* and *cis*-cinnamoyl peptides. We found a lower inhibitory activity for the cis-cinnamoyl peptides than for the trans-cinnamoyl peptides (figure 117 (Q) suggesting that the *cis*-isomer may adopt a constrained conformation within the active-site of thrombin with less favourable contacts within S3' subpocket.

6.2. MATERIALS AND METHODS

Fourier transform infrared (FTIR) studies of solid peptide films

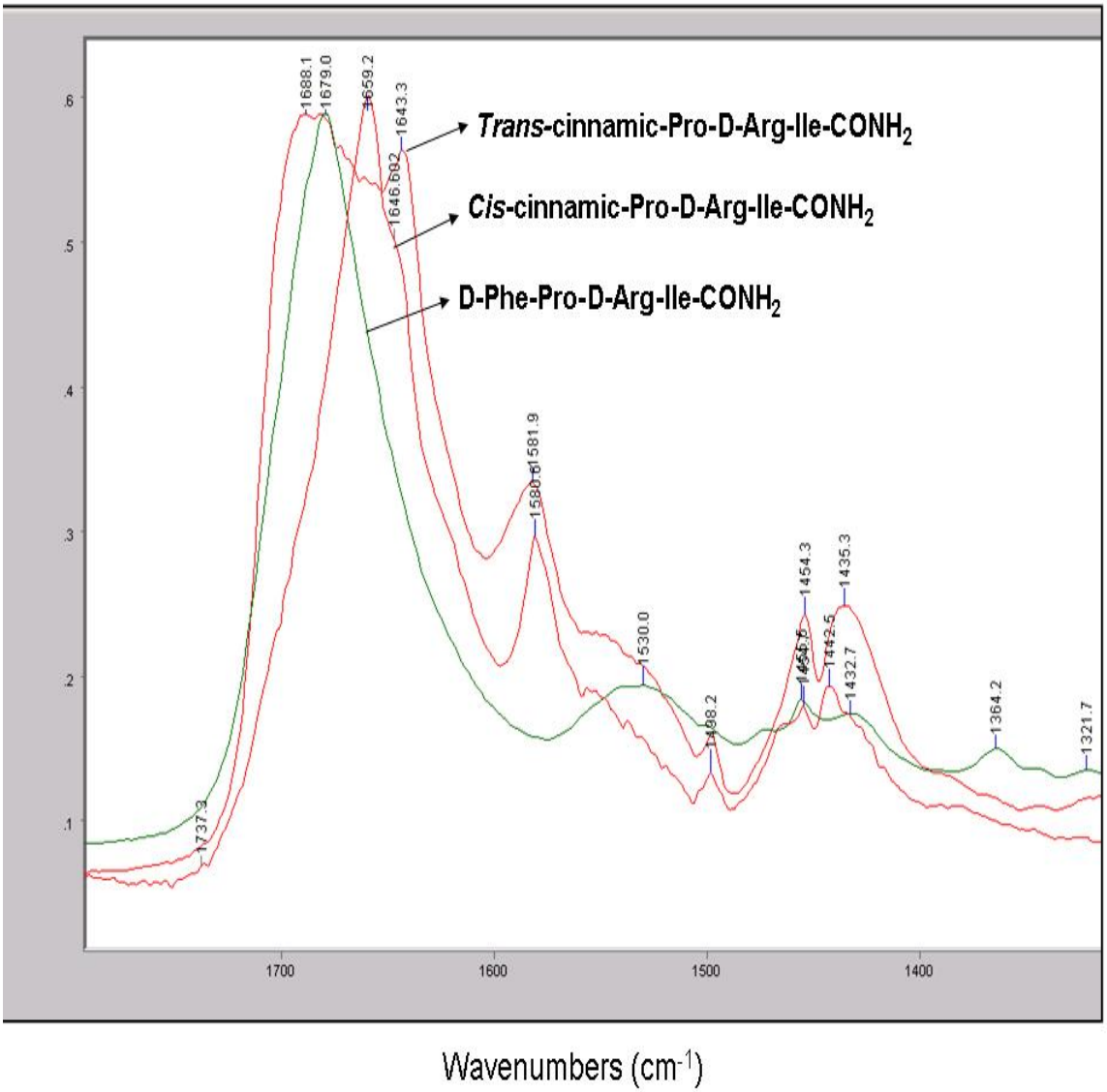
The *cis*-cinnamoyl peptides were obtained from *trans*-cinnamoyl peptides by exposing the solution of *trans*-cinnamoyl peptide to UV-light (354 nm) for 5 minutes. The solution of each peptide sample was spotted onto microscope slides (Low-6 microscope slides-Kenley technologies) which are reflecting back in the IR region. A Perkin Elmer Spectrum Spot light FTIR microscope (with imaging system) was used to collect the FTIR spectra. All scans were collected at 2 cm^{-1} spectral resolution and at least 8-10 scans were averaged and corrected for water interference by subtracting the background of water vapors. The FTIR spectra were further deconvoluted to assess the presence of the double bond stretch using Grams software (from Thermo-Galactic Applications) and a mixed Gaussian and Lorentzian fitting procedure. The goodness of the fit was determined based on the chi^2 analysis. As a control pure *trans*-cinnamic acid was used to assign the double bond stretch for *trans* and *cis* isomer.

6.3. RESULTS

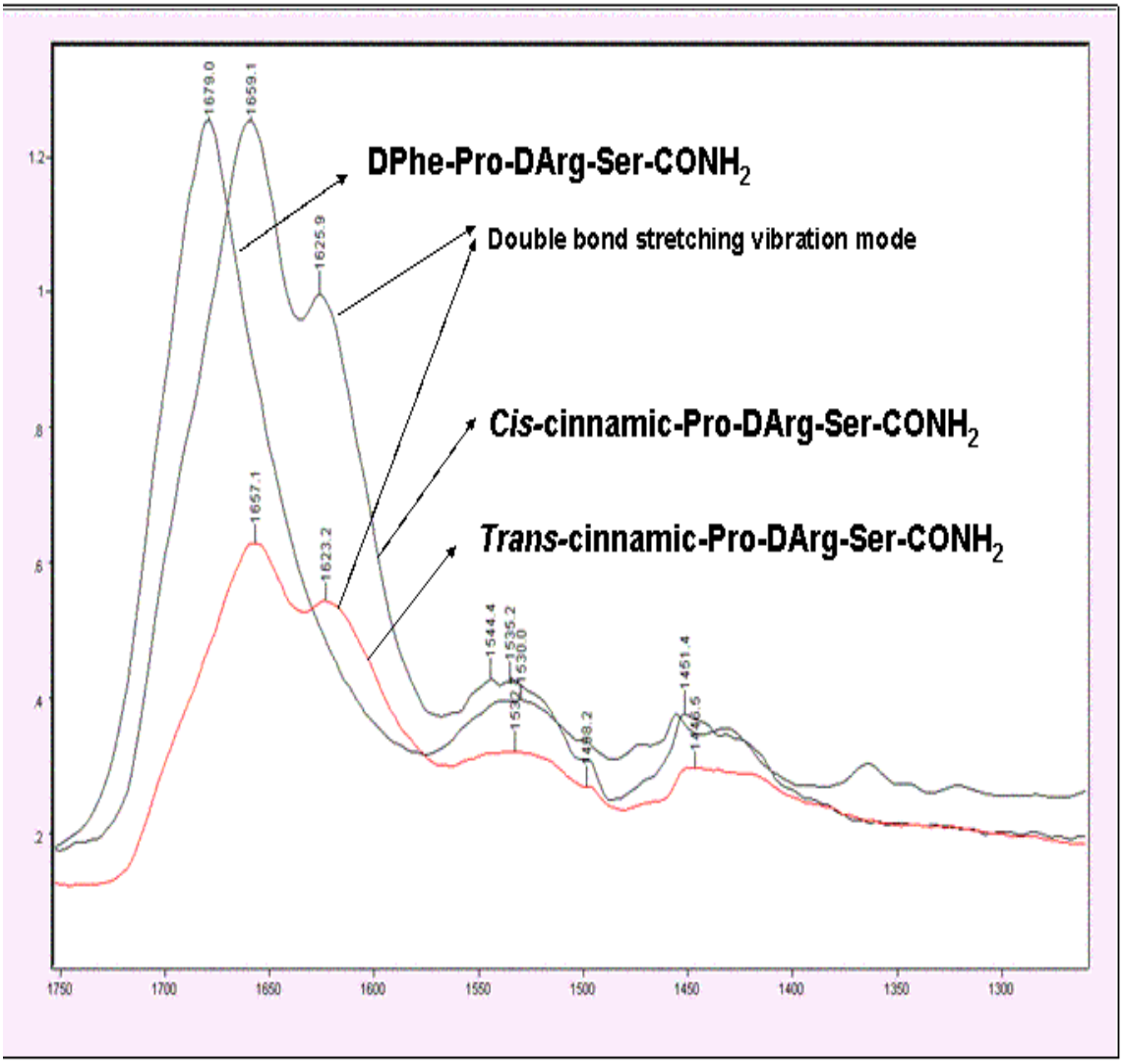
As already reported in the literature there is a 3-10 wavenumbers shift between *trans* and *cis* isomers of cinnamic acid in the region 1586-1648 cm-1) (5-10). Based on our knowledge these are the first FTIR data of *trans/cis*-cinnamoyl peptides characterized by a different inhibitory activity against thrombin. These results are showing that the FTIR microscopy applied to solid films of peptides is a very sensitive method to assess fine structural changes such as the conversion between *trans* and *cis*-isomer of cinnamic acid within the backbone of a peptide. Figure 126 shows the FTIR spectra for the major

pepides withtrans and cis-cinnamic acid at P3 position (see also Chapter 3 for kinetic of thrombin inhibition by these peptides). The calculated spectral curves resulted from fitting to a mixed Gaussian and Lorentzian functions. During fitting procedure the wavenumber intervals for the Amide I region of the peptide backbone was (1595-1688 cm^{-1}), the Amide II region (1500-1600 cm-1) and the aliphatic double bond stretch region (1583-1643 cm-1). The results are showing that in the case of cis-isomer the double bond stretch vibration mode is shifted to higher wavenumbers. In some cases we observed an opposite shift in the wavenumbers, i.e. the trans-isomer has the double bond stretch mode shifted toward higher wavenumbers. We don't have an explanation for these observed differences in the wavenumber shifts, however there was a consistency in having always 3-10 wavenumber shift of the aliphatic double bond stretch between the two cinnamoyl-peptide isomers, supporting the conversion of trans into cis isomer upon UV irradiation, eventhough the conversion amy not have been 100%. The SAR studies (see Section 3- related to the kinetic analysis of the data, figure 117 (Q)) supported the double bond constrain imposed by the *trans* or *cis*-cinnamic acid as a significant factor in determining a different affinity for thrombin of trans vs. cis-cinnamoyl peptides. The cis-cinnamoyl peptides have in some cases five fold less affinity than the corresponding trans-cinnamoyl peptides (table 15). The dihydrocinnamic-peptides (with more flexibility in adopting different conformations due to the release of the double bond constrain) were more active than the cis-cinnamoyl peptides in inhibiting thrombin and depending on the amino acid at P1' position more or less active than the trans-cinnamoyl peptides. Overall the inhibitory activity toward thrombin was as follows: *Trans*-cinnamoyl $\geq$ dihydrocynnamic > *cis*-cinnamoyl-peptides (see table 15 and figure 117).

A)

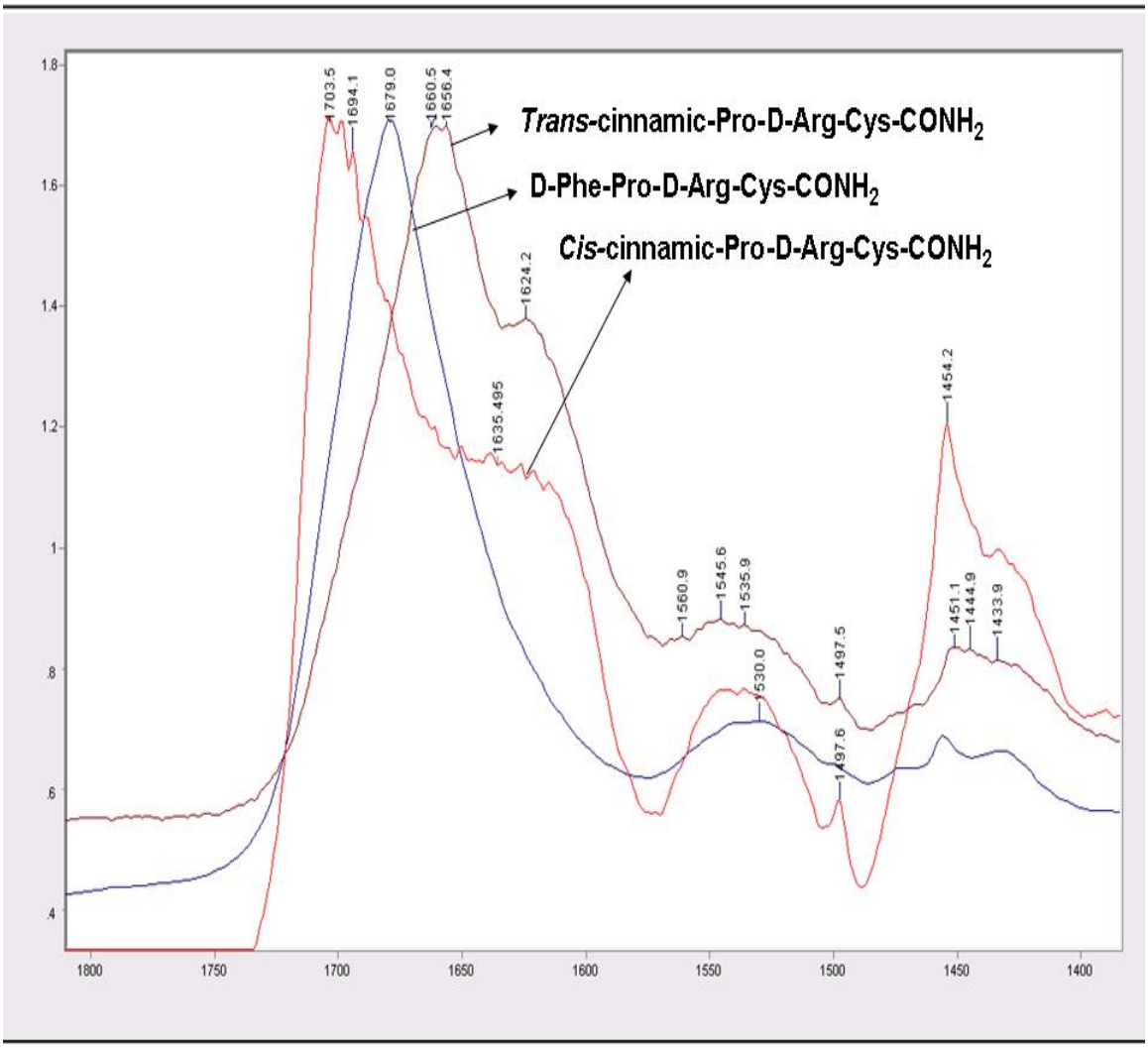

Trans-cinnamic-Pro-D-Arg-Ile-CONH2
Cis-cinnamic-Pro-D-Arg-Ile-CONH2
D-Phe-Pro-D-Arg-Ile-CONH2
1688.1
1679.0
1659.2
1643.3
1646.602
1581.9
1530.0
1454.3
1435.3
1442.5
1432.7
1364.2
1321.7
1737.3
1700
1600
1500
1400
Wavenumbers (cm-1)

B)

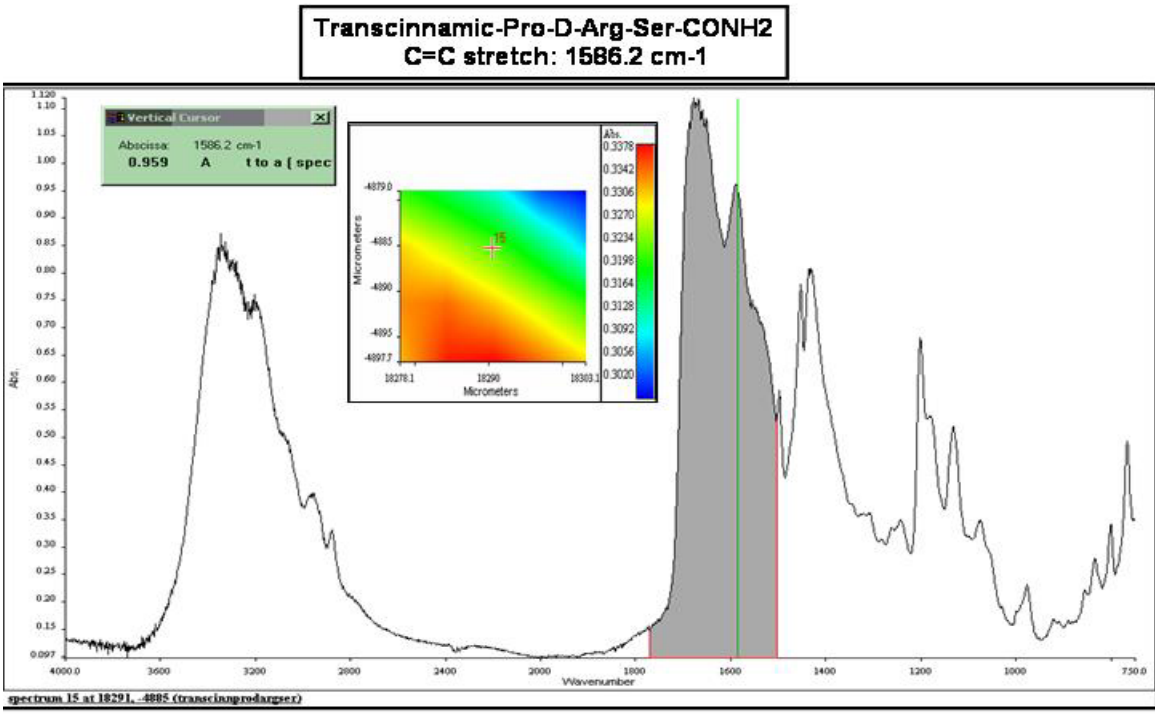


Wavenumber (cm^{-1})

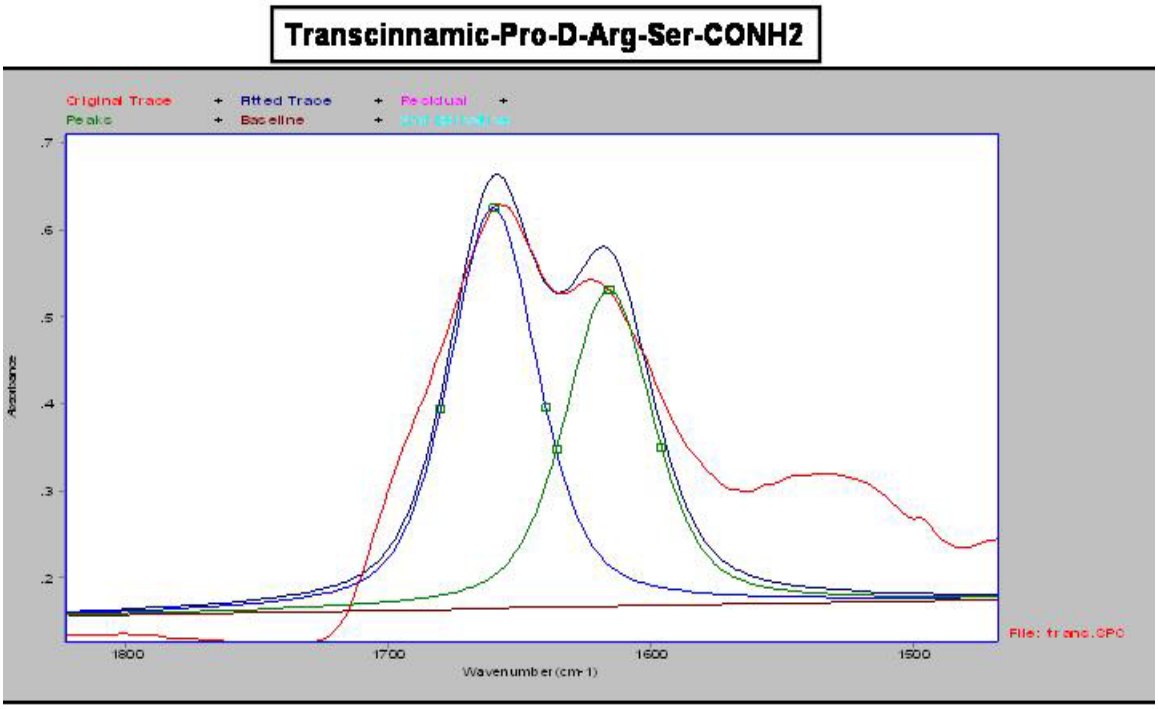
C)


1703.5
1694.1
1679.0
1660.5
1656.4
1624.2
1635.495
1560.9
1545.6
1535.9
1530.0
1497.5
1497.6
1454.2
1451.1
1444.9
1433.9
Trans-cinnamic-Pro-D-Arg-Cys-CONH2
D-Phe-Pro-D-Arg-Cys-CONH2
Cis-cinnamic-Pro-D-Arg-Cys-CONH2
1.8
1.6
1.4
1.2
1
.8
.6
.4
1800
1750
1700
1650
1600
1550
1500
1450
1400
Wavenumbers (cm-1)

D)



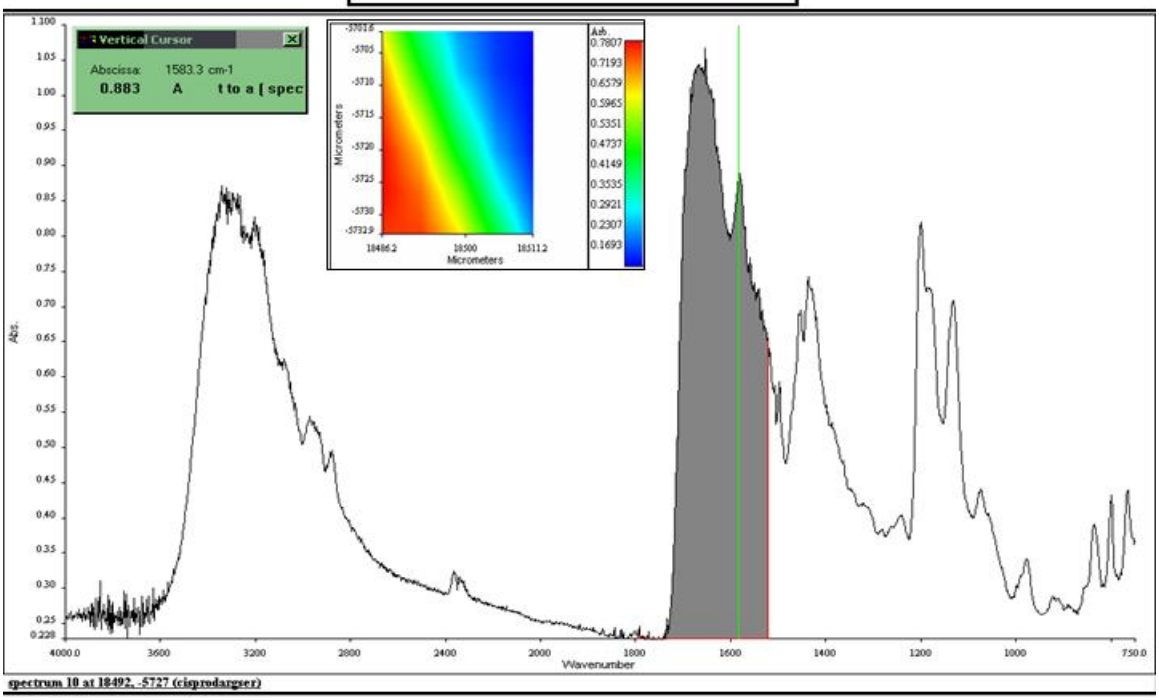
E)



Mixed G+L:	1664	cm-1	Amide I (Gaussian + Lorentzian fit)
Mixed G+L:	1618	cm-1	C = C stretch

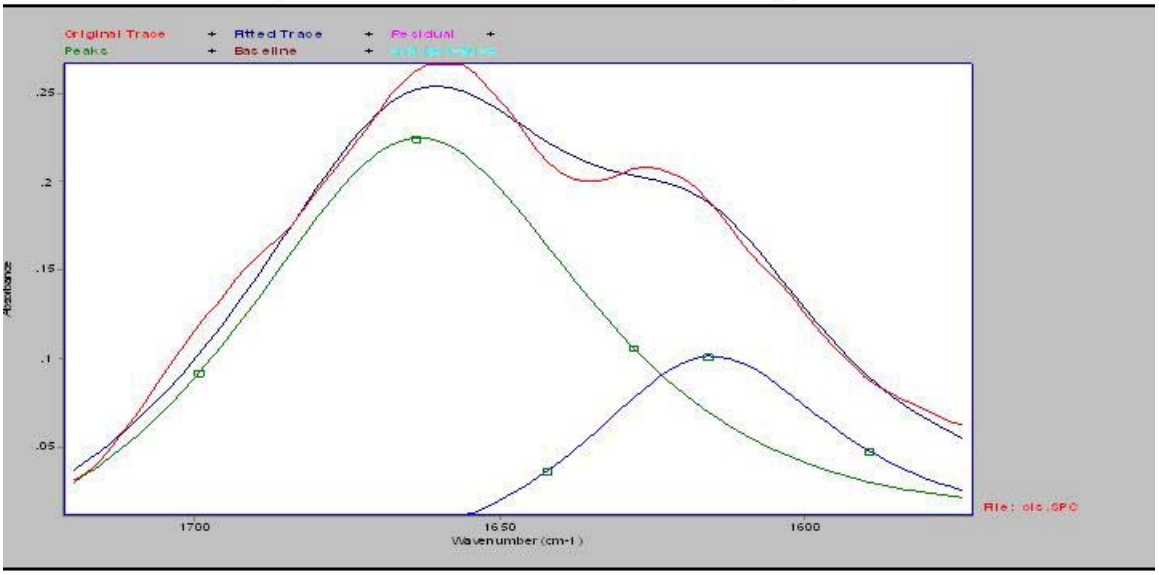
F)

Ciscinnamic-Pro-DArg-Ser-CONH2
C=C stretch: 1583.3 cm-1



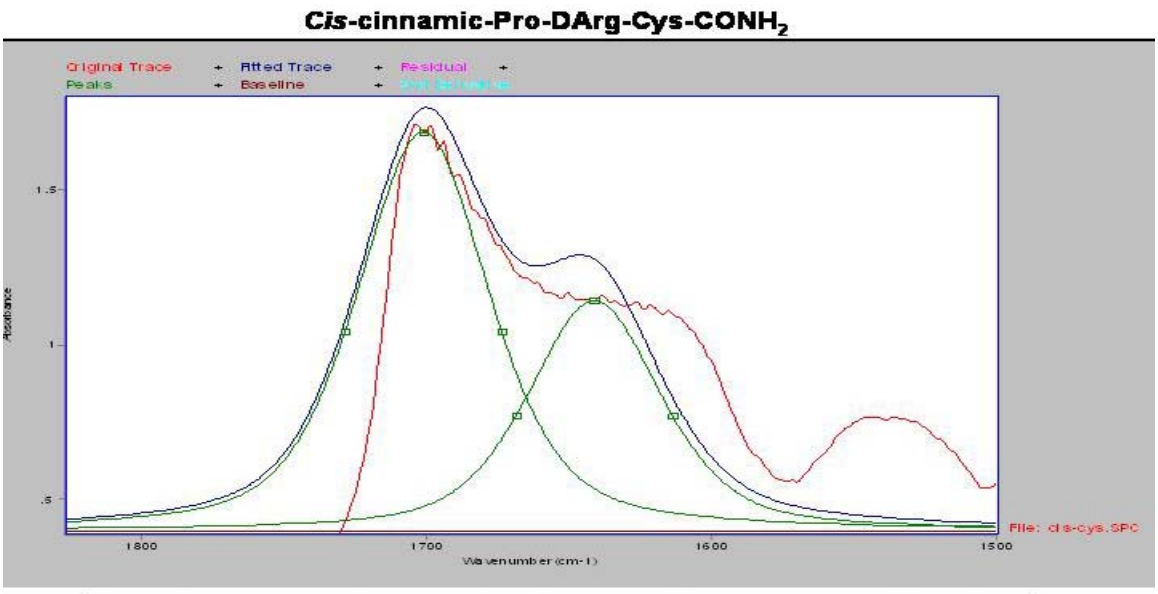
G)

Cis-cinnamic-Pro-D-Arg-Ser-CONH$_2$



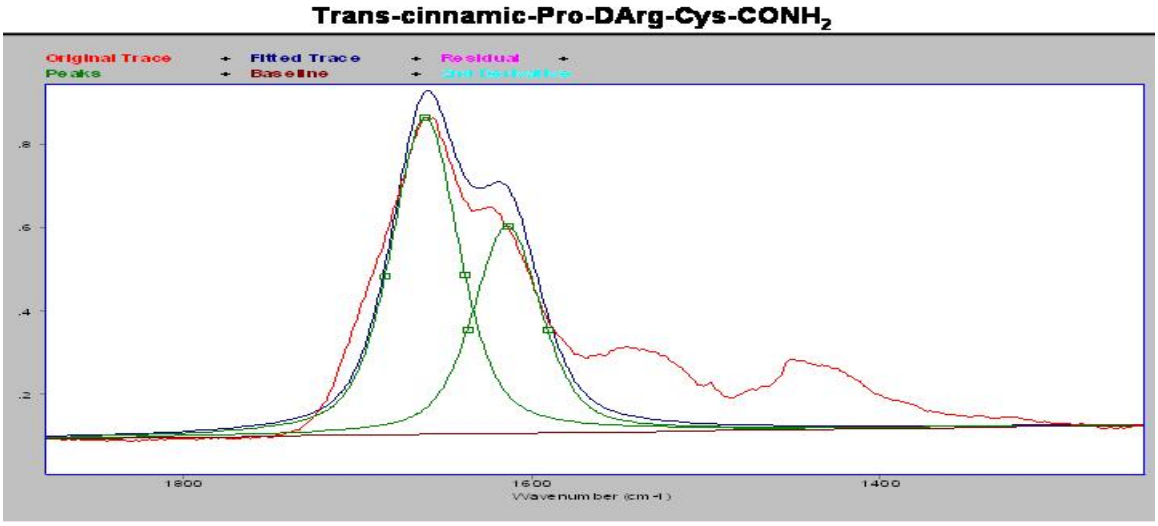
Mixed G+L:	**1663**	**cm-1**	**Amide I (Gaussian + Lorentzian fit)**
Mixed G+L:	**1615**	**cm-1**	**C = C stretch**

H)



Mixed G+L:	1701	cm-1	Amide I (Gaussian + Lorentzian fit)
Mixed G+L:	1641	cm-1	C = C stretch

I)



Mixed G+L:	1661	cm-1	Amide I (Gaussian + Lorentzian fit)
Mixed G+L:	1614	cm-1	C = C stretch

Figure 126 (A-I). FTIR spectra for transcinnamic-peptides and cis-cinnamic-peptides.

REFERENCES

1. Vass E, Hollosi M, Besson F, Buchet R. (2003). Vibrational spectroscopic detection of beta- and gamma-turns in synthetic and natural peptides and proteins. 2003 Chem Rev.;103(5):1917-54.

2. Fasman, G. D. In *Prediction of Protein Structure and the Principles of Protein Conformation*; Fasman, G. D., Ed.; Plenum Press: New York, 1986; pp 193-316.

3. Sternberg, M. F. E.; Bates, P. A.; Kelley, L. A.; MacCallum, R. M. *Curr. Opin. Struct. Biol.* **1999**, *9*, 368.

4. Koehl, R.; Levitt, M. *Nat. Struct. Biol.* **1999**, *6*, 108.

5. Jones, D. T.; Taylor, W. R.; Thorton, J. M. *Nature* **1992**, *358*, 86.

6. Richardson, J. S. *Adv. Protein Chem.* **1981**, *34*, 167.

7. Milner-White, E. J. *J. Mol. Biol.* **1990**, *216*, 385.

8. Milner-White, E. J.; Ross, B. M.; Ismail, R.; Belhadj-Mostefa, K.; Poet, R. *J. Mol. Biol.* **1988**, *204*, 777.

9. (a) Pavone, V.; Gaeta, A. L.; Nastri, F.; Maglio, O.; Isernia, C.; Saviano, M. *Biopolymers* **1996**, *38*, 705.

10. (a) Li, W.; Liu, Z.; Lai, L. *Biopolymers* **1999**, *49*, 481.

11. Fiser, A.; Kinh Gian Do, R.; Šali, A. *Protein Sci.* **2000**, *9*, 1753.

12. Galaktionov, S.; Nikiforovich, G. V.; Marshall, G. R. *Biopolym. (Peptide Sci.)* **2001**, *60*, 153.

13. Sreerama, N.; Woody, R. W. In *Circular Dichroism*; Berova, N., Nakanishi, K., Woody, R. W., Eds.; Wiley-VCH: New York, 2000; Chapter 21, pp 601-620.

14. Venyaminov, S. Yu.; Vassilenko, K. S. *Anal. Biochem.* **1994**, *222*, 176.

15. Perczel, A.; Hollósi, M. *Circular Dichroism and the Conformational Analysis of Biomolecules*; Fasman, G. D., Ed.; Plenum Publishing Co.: New York, 1996; Chapter 9, pp 285-380.

Chapter 7

Solution 2-D-transferred NOESY- NMR experiments of binary complexes [thrombin-inhibitor] and ternary complexes [thrombin-peptide inhibitor-thromstop]

7.1. INTRODUCTION

The observation of trNOEs relies on different tumbling time (τ_c) of free and bound molecules. Low or medium molecular weight molecules (MW <1000-2000) have a short correlation time ($\tau_c < 10^{-10}$s) and, as a consequence, such molecules exhibit positive NOEs (maximum value +0.5) depending on their molecular weight, shape and field strength. Large molecules, have much longer correlation time ($\tau_c > 10^{-5}$ s) however and thus exhibit strongly negative NOEs (maximum value of -1). When a small molecule is interacting with a large molecular weight protein and becomes bound to it, it behaves as a part of the large molecule and adopts the corresponding NOE behavior, i.e. it shows negative NOEs, so-called trNOEs (an equilibrium between the free the bound ligand is assumed). Specifically, the information of the bound ligand (having the macromolecule-like property) is transferred to the free ligand through cross-relaxation events. These trNOEs reflect the bound conformation of the ligand. Thus, binding the ligand to a target protein receptor can be easily distinguished by looking at the sign and the size of the observed NOEs, specifically "negative" trNOEs are detected if the ligand is binding to the macromolecule. The trNOEs experiment can be designed such that trNOEs originating from the bound state of the ligand can be discriminated from the trNOEs of the free ligand in solution. In practice, the change in sign (from positive to negative) of NOE of a small molecule is a quick, certain indication that the binding has occurred (1-10).

Figure 127 presents a schematic representation of the difference between the NOESY and the tr-NOESy experiments in the case of a small molecule ligand (containing 4 different protons (a, b, c and d). The part A correspond to the small positive NOEs from cross-diagonal peaks in the case of free ligand in solution (the "positive" NOEs have a different color (sign) than the diagonal). Part B correspond to the situation of ligand interacting with the target macromolecule when the large "negative" NOEs are building-up (the NOEs have the same sign with the diagonal). The arrows shows the assigned protons which are involved in the transferred-NOEs (6-11).

A **B**

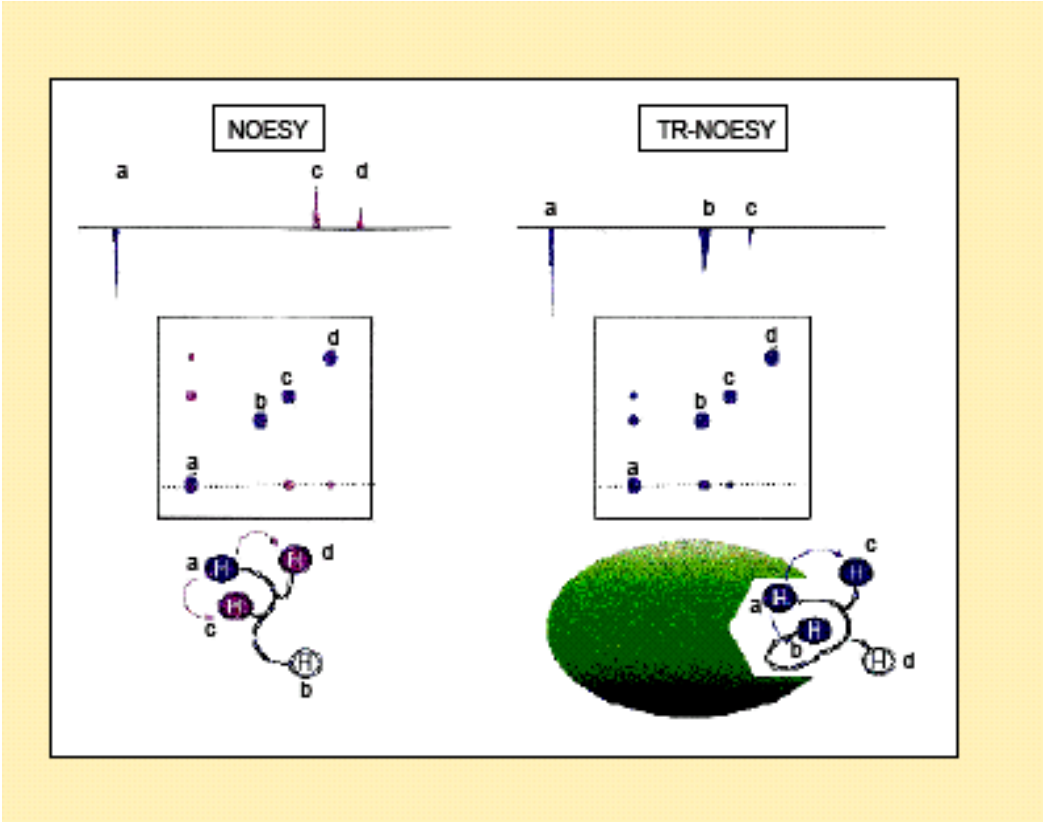


Figure 127: Schematic representation of a 2D-tr-1H-NOESY experiment .

Both the 1D and 2D transferred NOE experiments were used extensively to determine the conformation of weak to medium affinity ligands when bound to a large protein. Recently the method has been proposed as a tool for performing NMR screening (12-15). There are two major types of tr-NOEs experiments:

1. Intramolecular magnetization transfer NOESY.
2. Intermolecular magnetization transfer NOESY.

The major factors which limit the tr-NOESY type of experiments are binding affinities between the ligand and the macromolecule (which should be in the order of uM to mM range, i.e. the ligand should be in fast exchange with the receptor molecule) and the concentration of ligand with respect to the receptor molecule. A molar ratio of 10:1 to 20:1 ligand : receptor is expected to assure that the bound ligand magnetization is transferred to the excess free ligand through chemical exchange, and thus, the NMR information characterizing the bound structure is observed via the sharp resonances of the free ligand (12-15). Therefore the method of tr-NOESY is not suitable for the identification of ligands with poor solubility or high affinity binding constants (i.e. Kd (dissociation constants) in the nM range). Another disadvantage is the presence in the spectrum of strong diagonal peaks that hamper the observation of cross-peaks between ligand resonances having similar chemical shifts. In addition, the strong diagonal peaks may introduce t1 noise or baseline problems that interfere with the observation of weak cross peaks (1-6).

7.2. MATERIALS AND METHODS

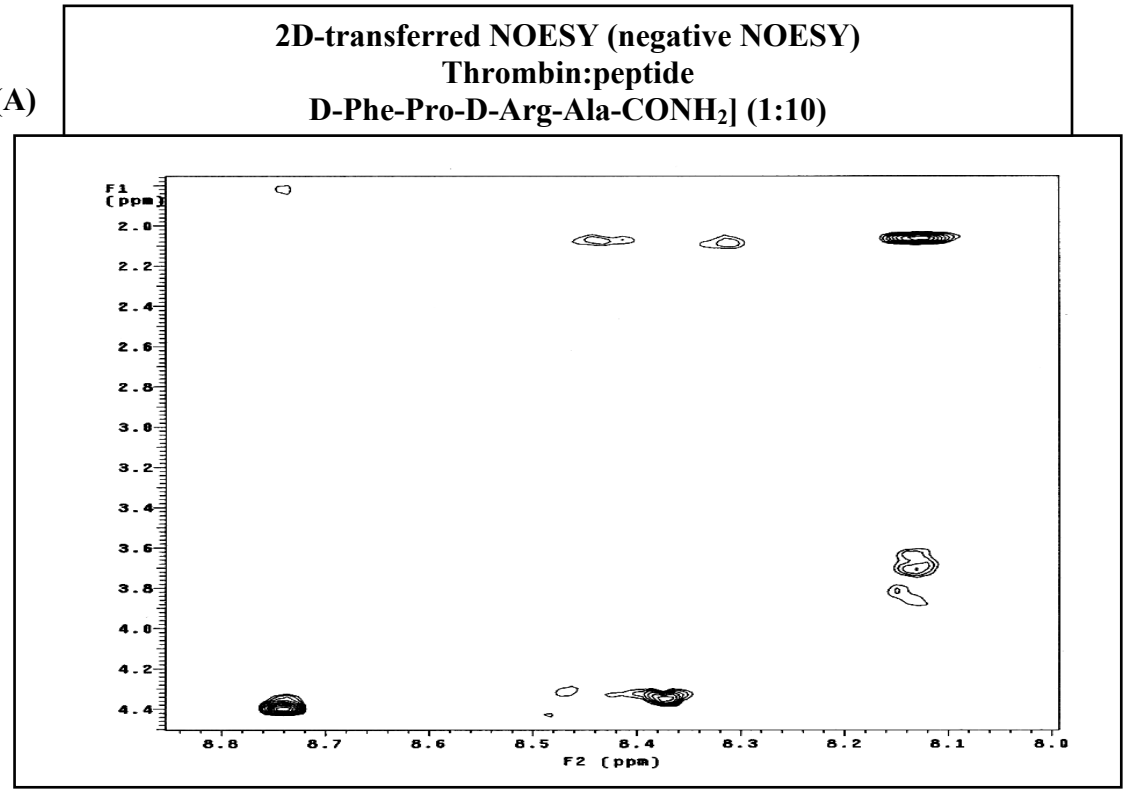
2-D-transferred NOESY experiments were performed for two tetrapeptides D-Phe-Pro-D-Arg-Ala-$CONH_2$ and D-Phe-Pro-D-Arg-Gly-$CONH_2$ which have the Kis at low μM and are suitable for these kinds of experiments (8, 9). The 2D-transferred

NOESY experiments were conducted at 25^{o}C, with 10% D2O in all samples, in a buffer 30 mM phosphate, containing 0.2 M NaCl, pH 5.6. The sequence pulse and the mixing time (120 ms) were very similar with those described for other thrombin-peptides system (8, 9). The NMR experiments were conducted both in the presence and in the absence of a strong inhibitor (thromstop) in order to test if the peptides which were shown good inhibitory activity are competing for the same sites with thromstop, i.e. at the active site.

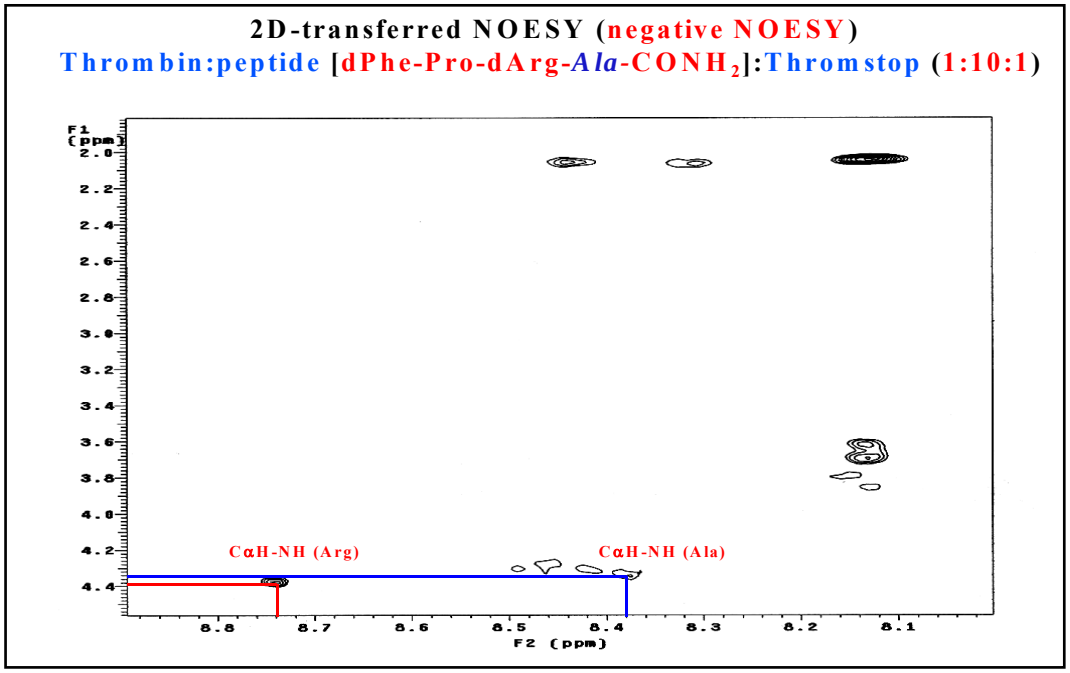
7.3. RESULTS AND DISCUSSIONS

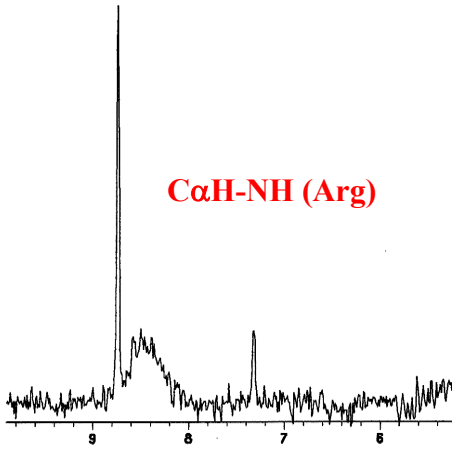
The NMR results showed in Figure 128 support the hypothesis of having competition for the active sites for both D-Phe-Pro-D-Arg-Ala-$CONH_2$ and D-Phe-Pro-D-Arg-Gly-$CONH_2$ peptides. However, this competition is not 100%. Specifically the intensity of the negative NOEs built up in the complex thrombin-peptides were diminished in the ternary complex thrombin-peptide-thromstop (figure 128 (A, B, C and E and F). Part of the peptides molecules are binding to other sites than the active-sites, a result which is consistent with the kinetic data (Figure 118-Chapter 3). Based on Wurtrich (5-6) assignments of 1H chemical shift for different amino-acids within a peptide sequence space, we were able to assign the transfer of NOE in between CαH-NH (D-Arg) and CαH-NH of the amino acid after Arg within the sequence space D-Phe-Pro-D-Arg-P1'-CONH2, i.e. CαH-NH of Ala or Gly. Based on transferred NOE experiments D-Arg and and the amino acids at P1' position (Gly or Ala for the peptides we investigated) are involved in binding to the active site of thrombin since their negative NOEs were decreased in intensity by competition with the more potent inhibitor Thromstop.

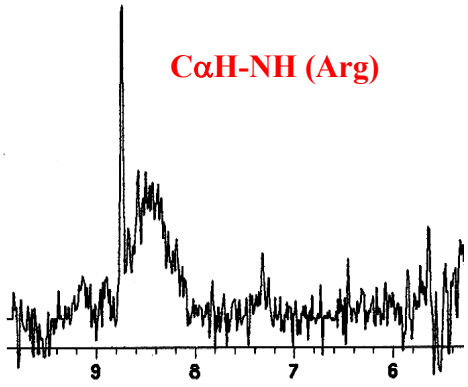
(A)


2D-transferred NOESY (negative NOESY)
Thrombin:peptide
D-Phe-Pro-D-Arg-Ala-CONH2] (1:10)
F1 (ppm)
F2 (ppm)

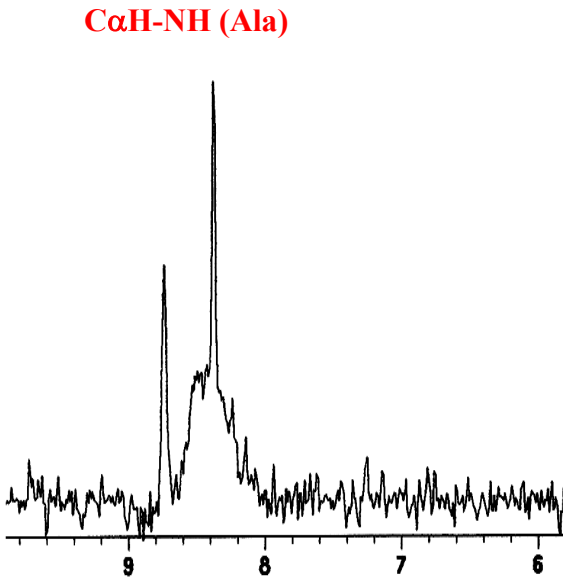
(B)

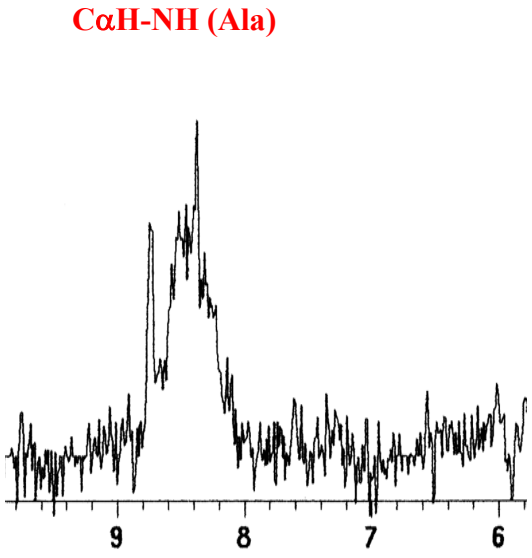

2D-transferred NOESY (negative NOESY)
Thrombin:peptide [dPhe-Pro-dArg-Ala-CONH2]:Thromstop (1:10:1)
CαH-NH (Arg)
CαH-NH (Ala)
F1 (ppm)
F2 (ppm)





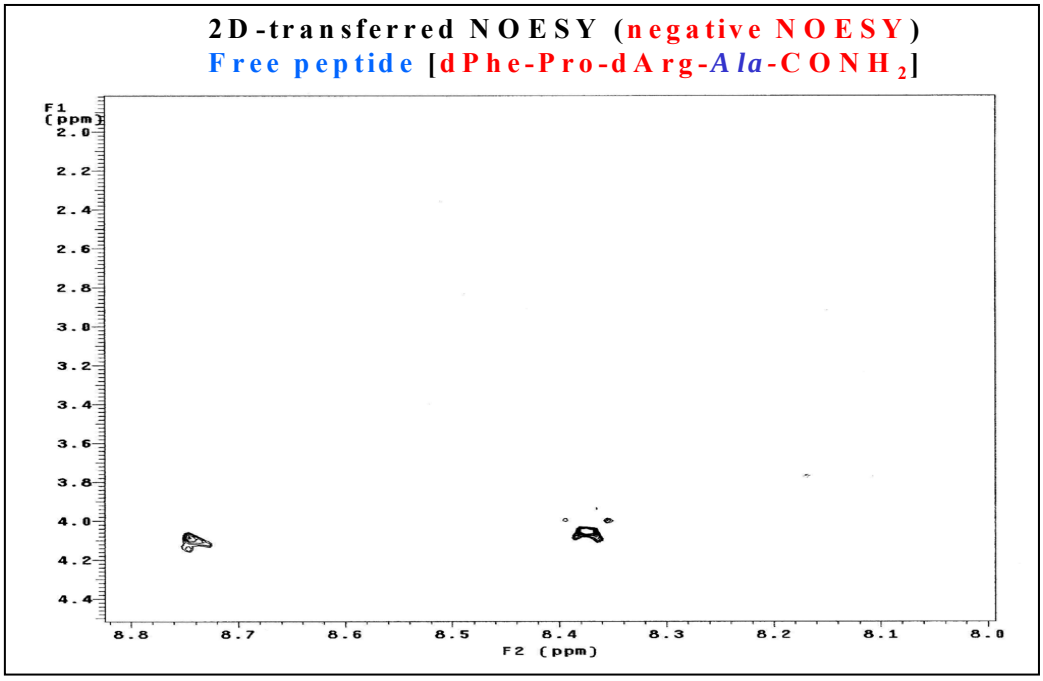
(C)

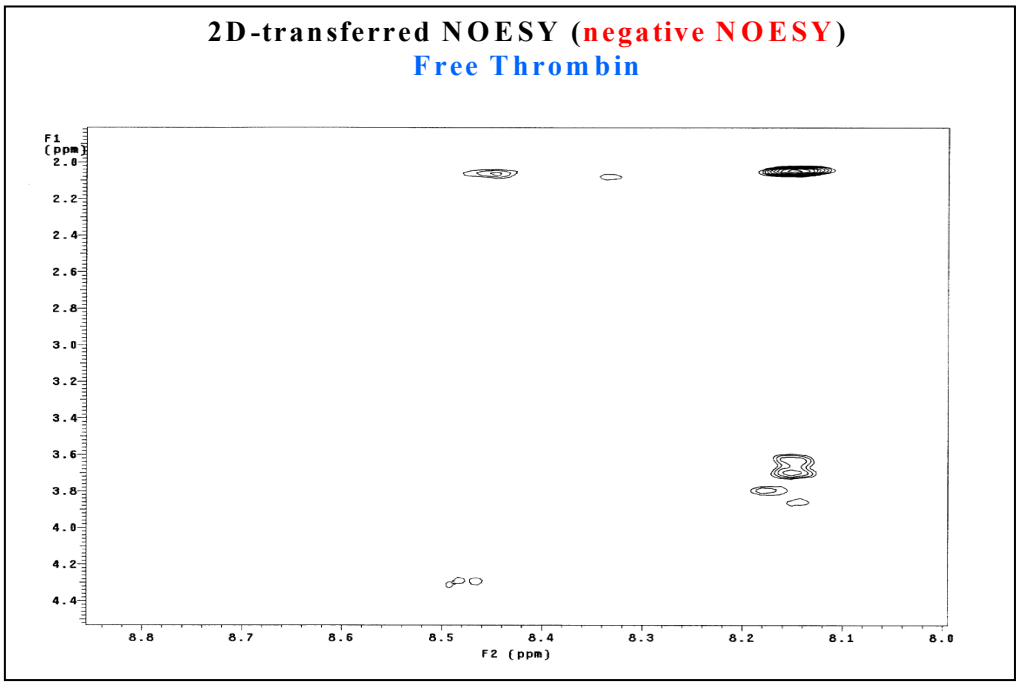




(D)

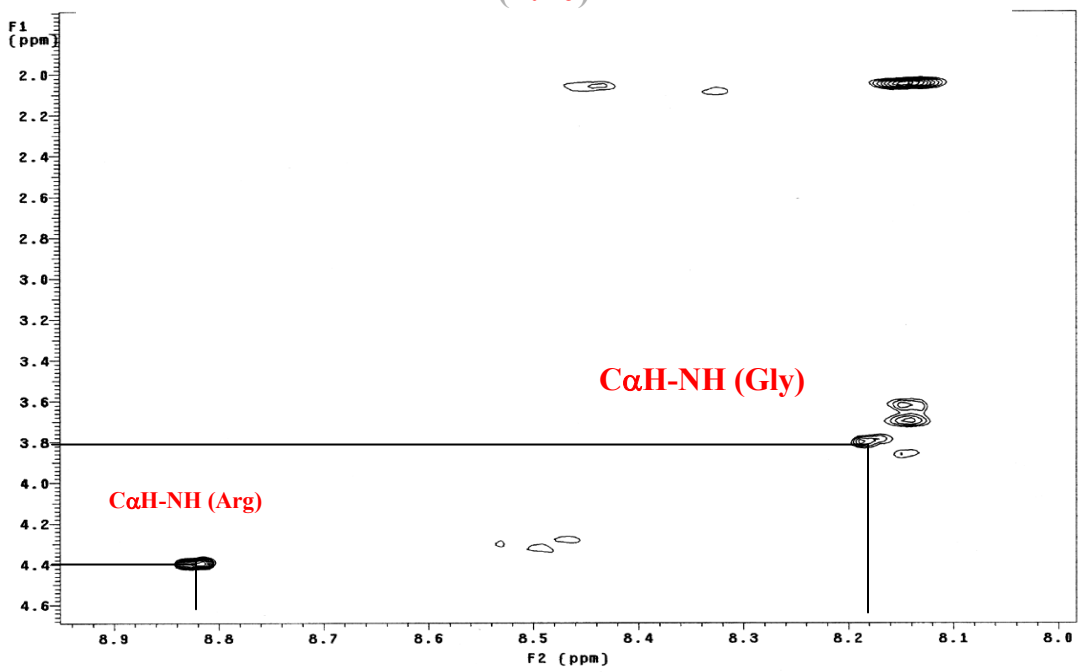
Control experiments



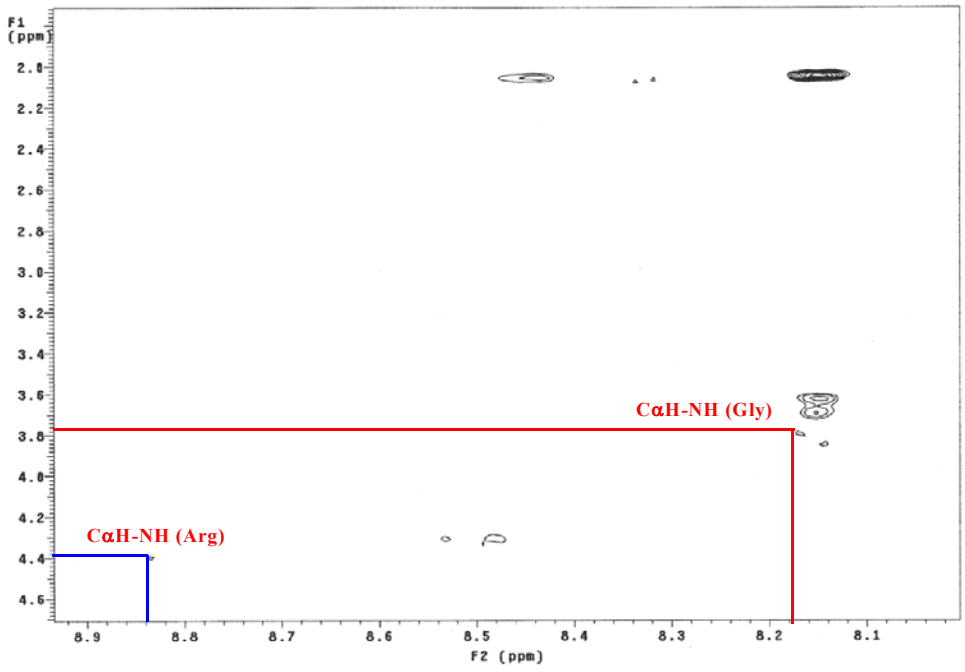


(E)

2D-transferred NOESY (negative NOESY)
Thrombin : peptide [dPhe-Pro-dArg-Gly-$CONH_2$)
(1:10)



2D-transferred NOESY (negative NOESY)
Thrombin:peptide [dPhe-Pro-dArg-*Gly*-$CONH_2$]:Thromstop (1:10:1)



(F)

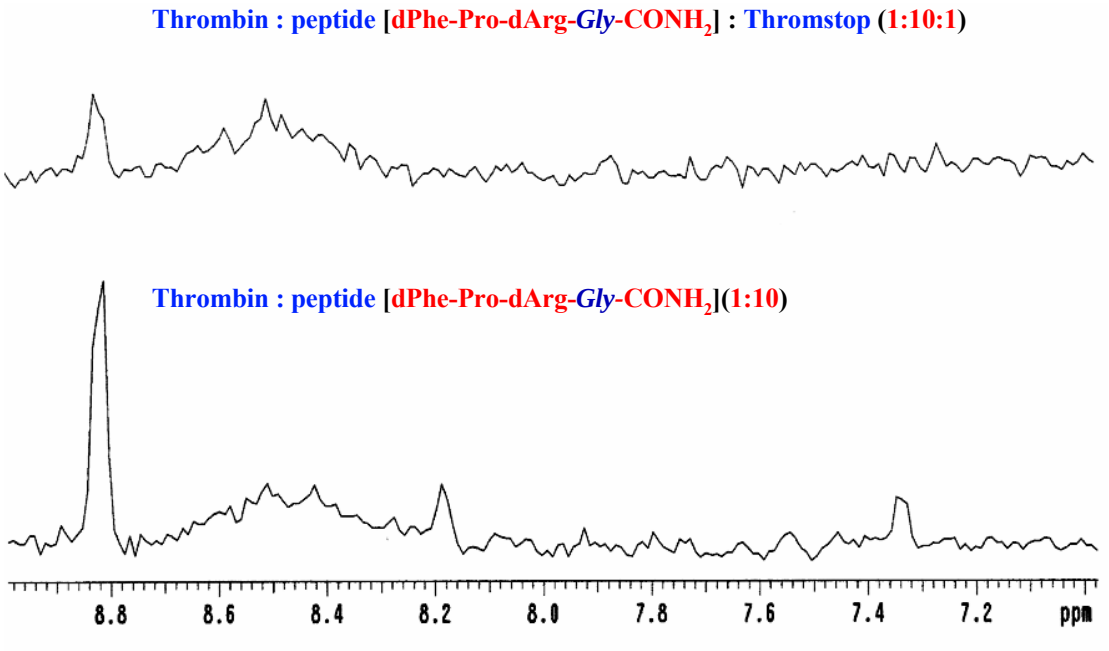
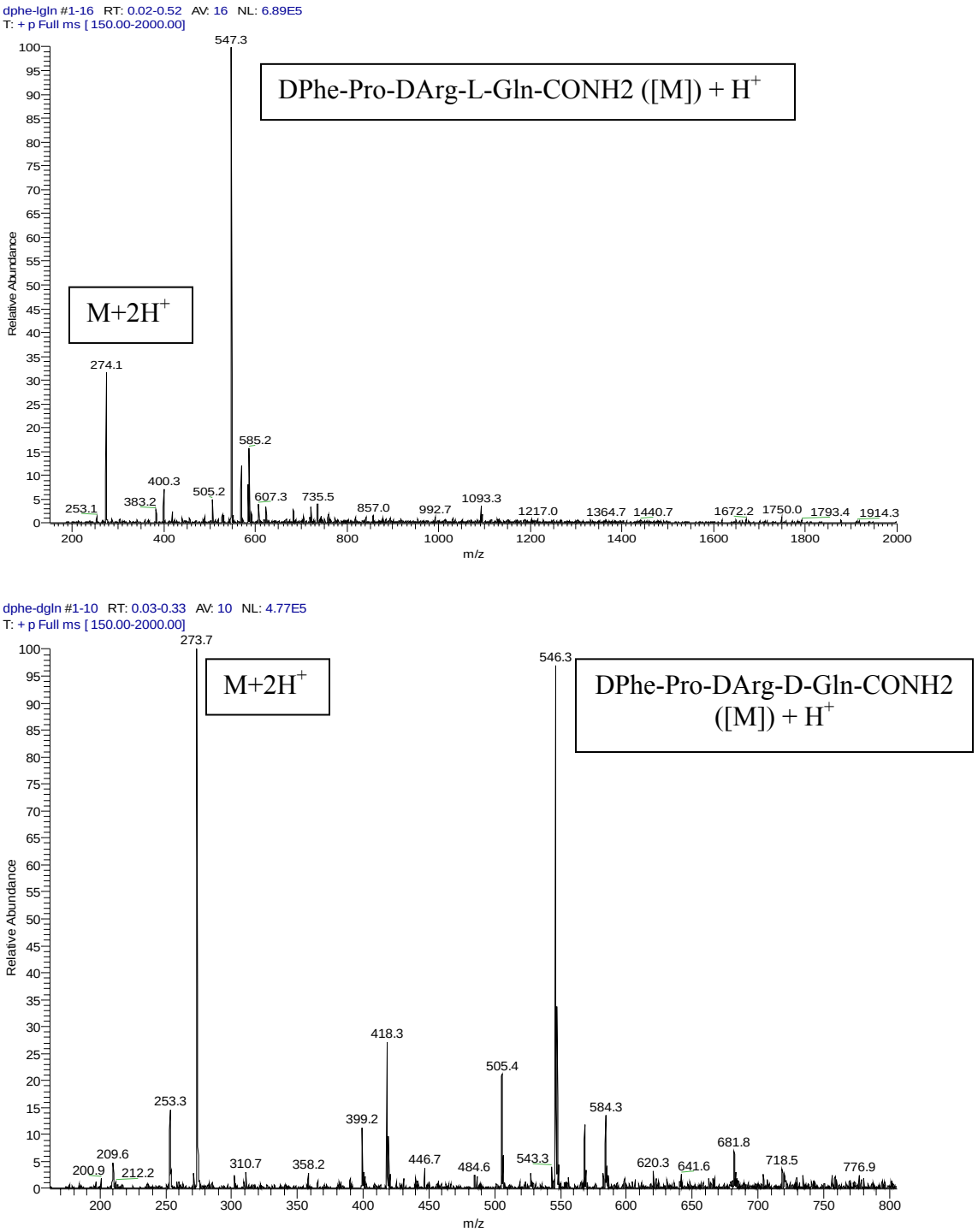
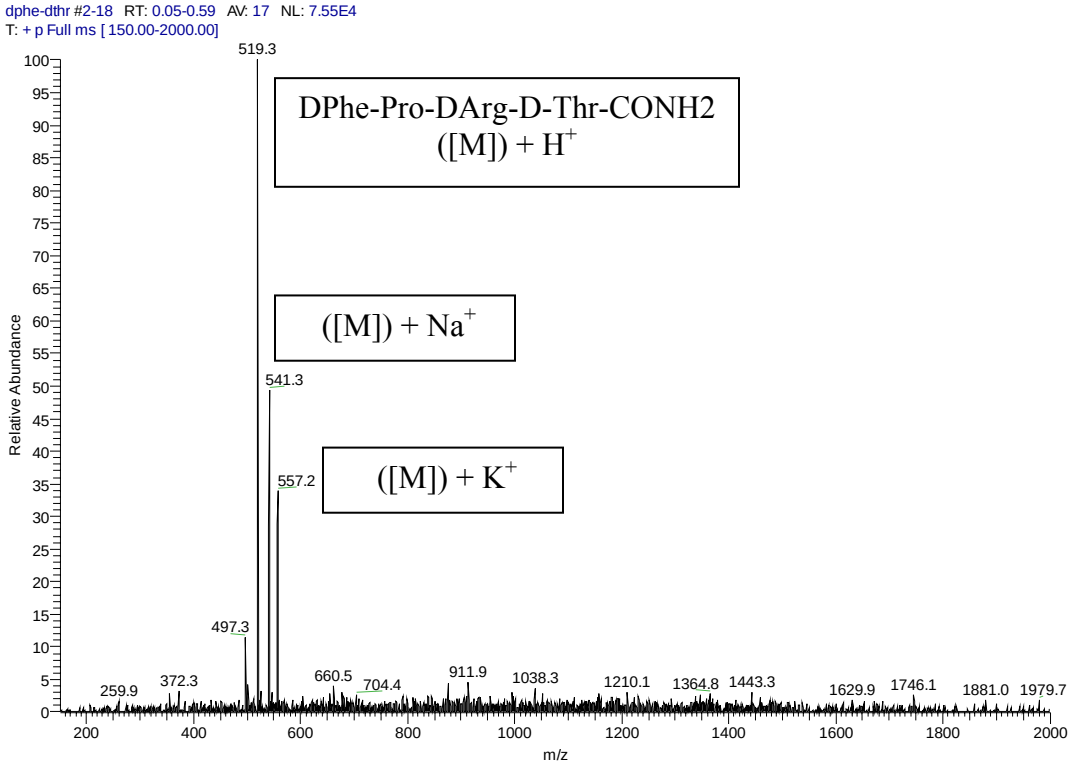


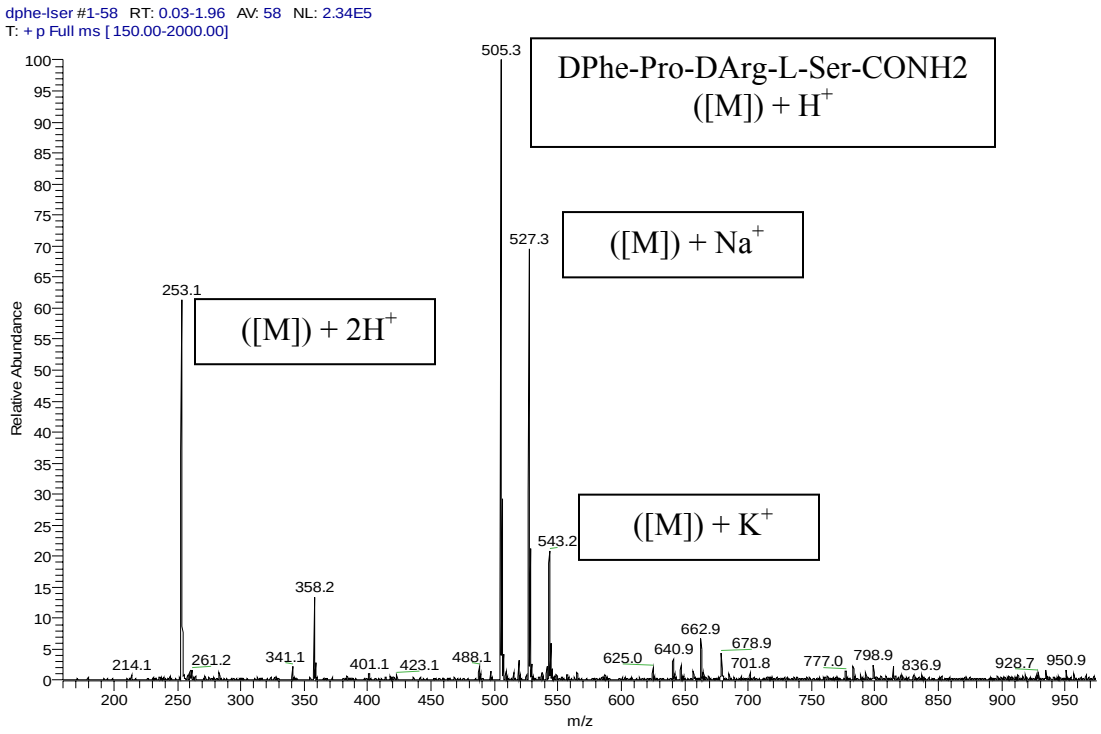
Figure 128. (A-F). Transferred NOESY experiments for binary complexes (thrombin-peptide) and ternary complexes (thrombin:peptide:thromstop)

APPENDIX X

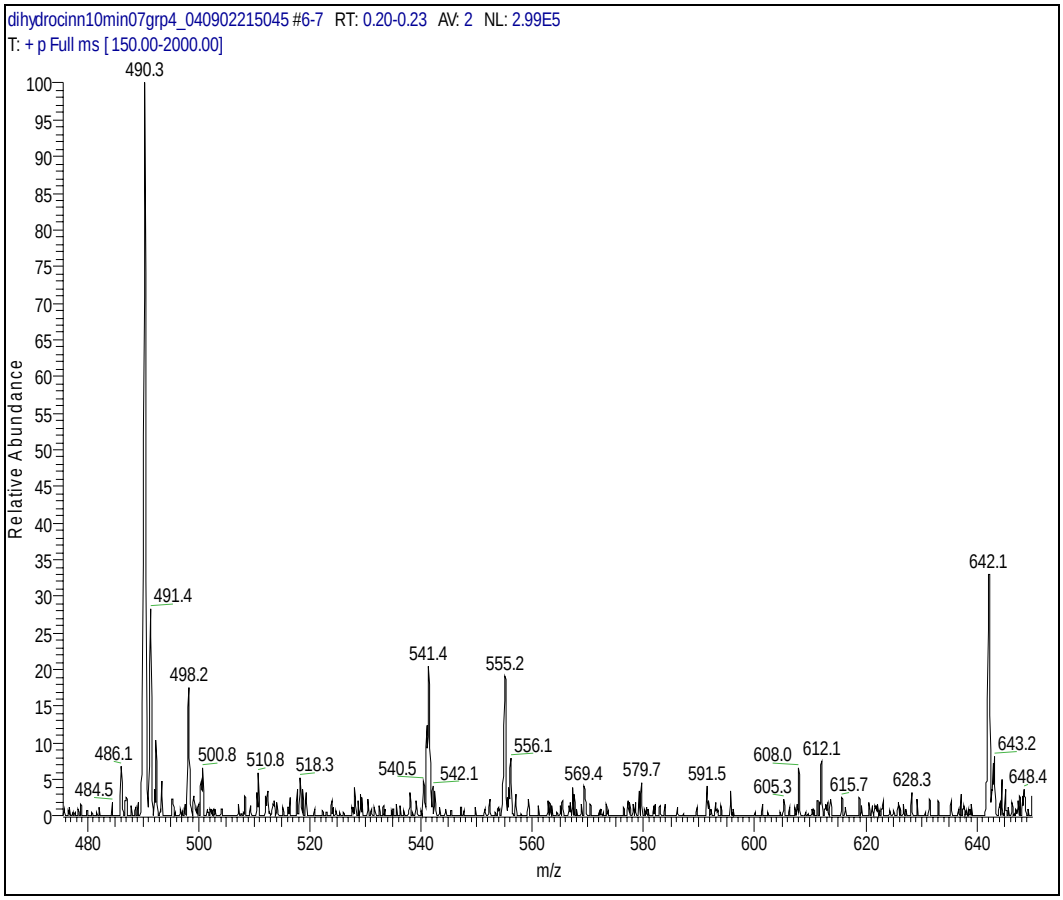
Figure 129. ESI-MS of some purified peptides from a mixture of peptide libraries, which were used in kinetic studies.




dphe-dthr #2-18 RT: 0.05-0.59 AV: 17 NL: 7.55E4
T: + p Full ms [150.00-2000.00]
DPhe-Pro-DArg-D-Thr-CONH2
([M]) + H+
([M]) + Na+
([M]) + K+
519.3
541.3
557.2
497.3
259.9
372.3
660.5
704.4
911.9
1038.3
1210.1
1364.8
1443.3
1629.9
1746.1
1881.0
1979.7
Relative Abundance
m/z


dphe-lser #1-58 RT: 0.03-1.96 AV: 58 NL: 2.34E5
T: + p Full ms [150.00-2000.00]
DPhe-Pro-DArg-L-Ser-CONH2
([M]) + H+
([M]) + Na+
([M]) + 2H+
([M]) + K+
505.3
527.3
253.1
543.2
358.2
214.1
261.2
341.1
401.1
423.1
488.1
625.0
640.9
662.9
678.9
701.8
777.0
798.9
836.9
928.7
950.9
Relative Abundance
m/z

Dihydrocinnamoyl-Pro-D-Arg-D-Ser-$CONH_2$



Dihydrocinnamoyl-Pro-D-Arg-L-Thi-$CONH_2$

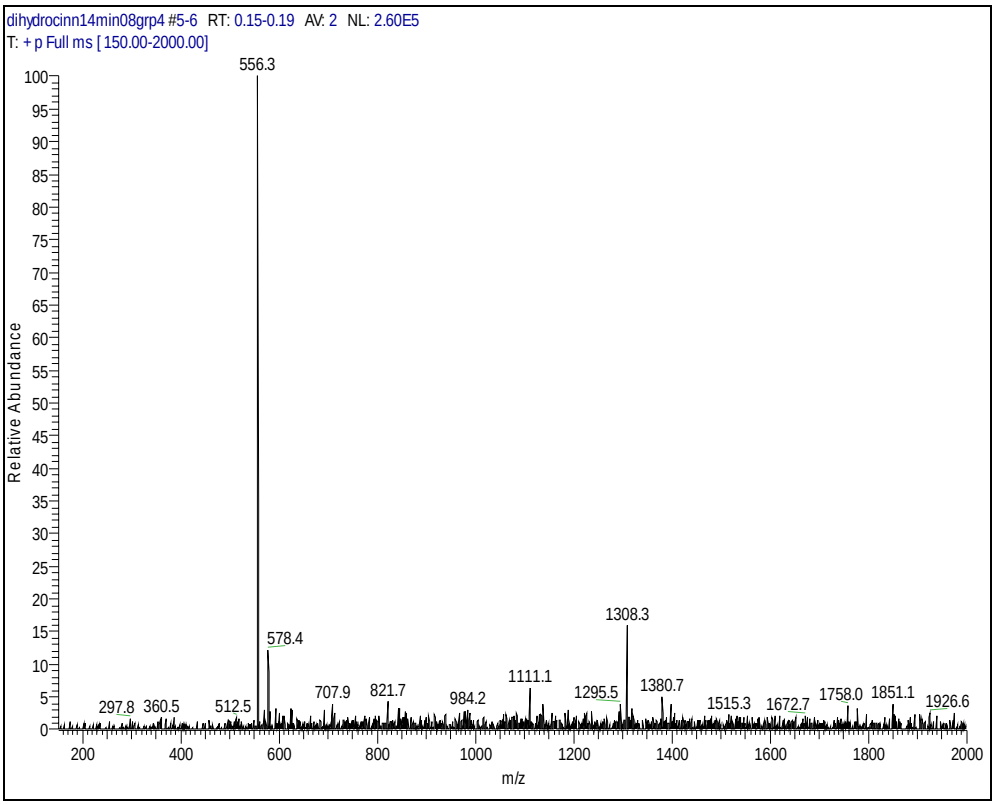


Table 16: Candidate tetrapeptides synthesized in this study with their calculated and experimental molecular weights.

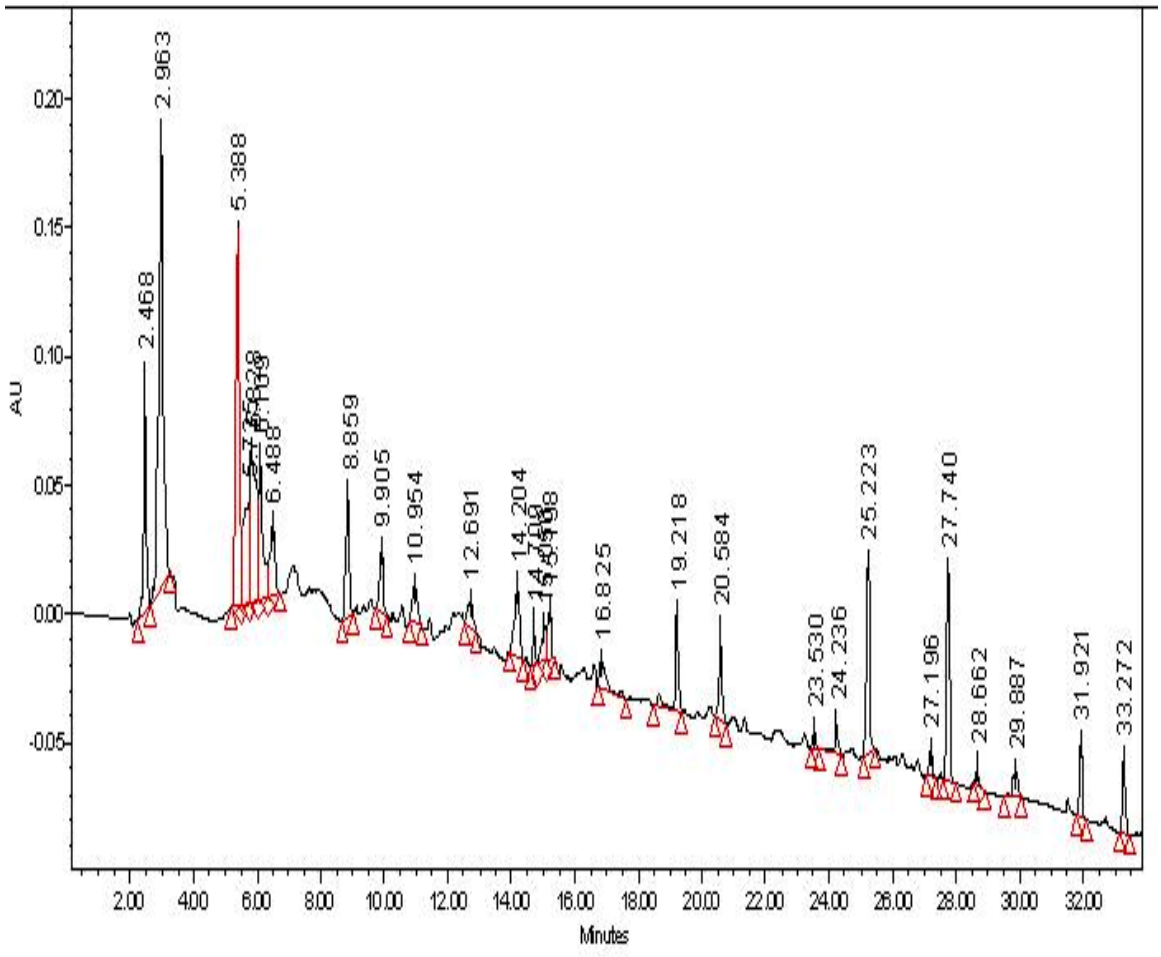
Inhibitor (Tetrapeptides)	**Theoretical M.W. ($g.mol^{-1}$)**	**Experimental Mass (m/z) ± [H^+]**
***Trans*-cinnamoyl-Pro- D-Arg-D-Pro-$CONH_2$**	**497.64**	**498.23**
***Trans*-cinnamoyl-Pro-D-Arg-D-His-$CONH_2$**	**537.67**	**538.28**
***Trans*-cinnamoyl-Pro-D-Arg-D-Ala-$CONH_2$**	**471.60**	**472.19**
***Trans*-cinnamoyl-Pro-D-Arg-D-Ser-$CONH_2$**	**487.60**	**488.21**
***Trans*-cinnamoyl-Pro-D-Arg-D-Cys-$CONH_2$**	**504.67**	**506.40 ± H^+**
***Trans*-cinnamoyl-Pro-D-Arg-D-Thr-$CONH_2$**	**501.63**	**502.22**
***Trans*-cinnamoyl-Pro-D-Arg-L-Thi-$CONH_2$**	**553.77**	**554.31**
***Trans*-cinnamoyl-Pro-D-Arg-D-Tic-$CONH_2$**	**559.76**	**560.23**
***Trans*-cinnamoyl-Pro-D-Arg-D-Asn-$CONH_2$**	**514.63**	**515.45**
***Trans*-cinnamoyl-Pro-D-Arg-D-Gln-$CONH_2$**	**528.66**	**529.05**
***Trans*-cinnamoyl-Pro-D-Arg-D-3-Benzothienylalanine**	**640.81**	**681.00** ($+H^+$, **K^+**)
***Trans*-cinnamoyl-Pro-D-Arg-D-3,5-**	**620.69**	**644.7 ($+H^+$,Na^+)**

DifluoroPhenylalanine		
***Trans*-cinnamoyl-Pro-D-Arg-Gly- $CONH_2$**	**512.52**	**513.46**
Dihyrocinnamoyl-Pro-D-Arg-D-Pro-$CONH_2$	**499.66**	**500.28**
Dihyrocinnamoyl-Pro-D-Arg-D-His-$CONH_2$	**539.69**	**540.38**
Dihyrocinnamoyl-Pro-D-Arg-D-Ala-$CONH_2$	**473.62**	**474.01**
Dihyrocinnamoyl-Pro-D-Arg-D-Ser-$CONH_2$	**489.62**	**490.48**
Dihyrocinnamoyl-Pro-D-Arg-D-Cys-$CONH_2$	**506.69**	**505.87**
Dihyrocinnamoyl-Pro-D-Arg-D-Thr-$CONH_2$	**503.65**	**504.24**
Dihyrocinnamoyl-Pro-D-Arg-L-Thi-$CONH_2$	**555.79**	**556.30**
Dihyrocinnamoyl-Pro-D-Arg-D-Tic-$CONH_2$	**561.78**	**562.20**
Dihyrocinnamoyl-Pro-D-Arg-D-Asn-$CONH_2$	**516.65**	**517.34**
Dihyrocinnamoyl-Pro-D-Arg-D-Gln-$CONH_2$	**530.68**	**531.40**
Dihyrocinnamoyl-Pro-D-Arg-D-3-Benzothienylalanine	**642.83**	**644.79**
Dihyrocinnamoyl-Pro-D-Arg-Gly-$CONH_2$	**514.59**	**515.27**

APPENDIX XI

Figure 130. Reversed-Phase HPLC analysis of libraries of synthetic peptides

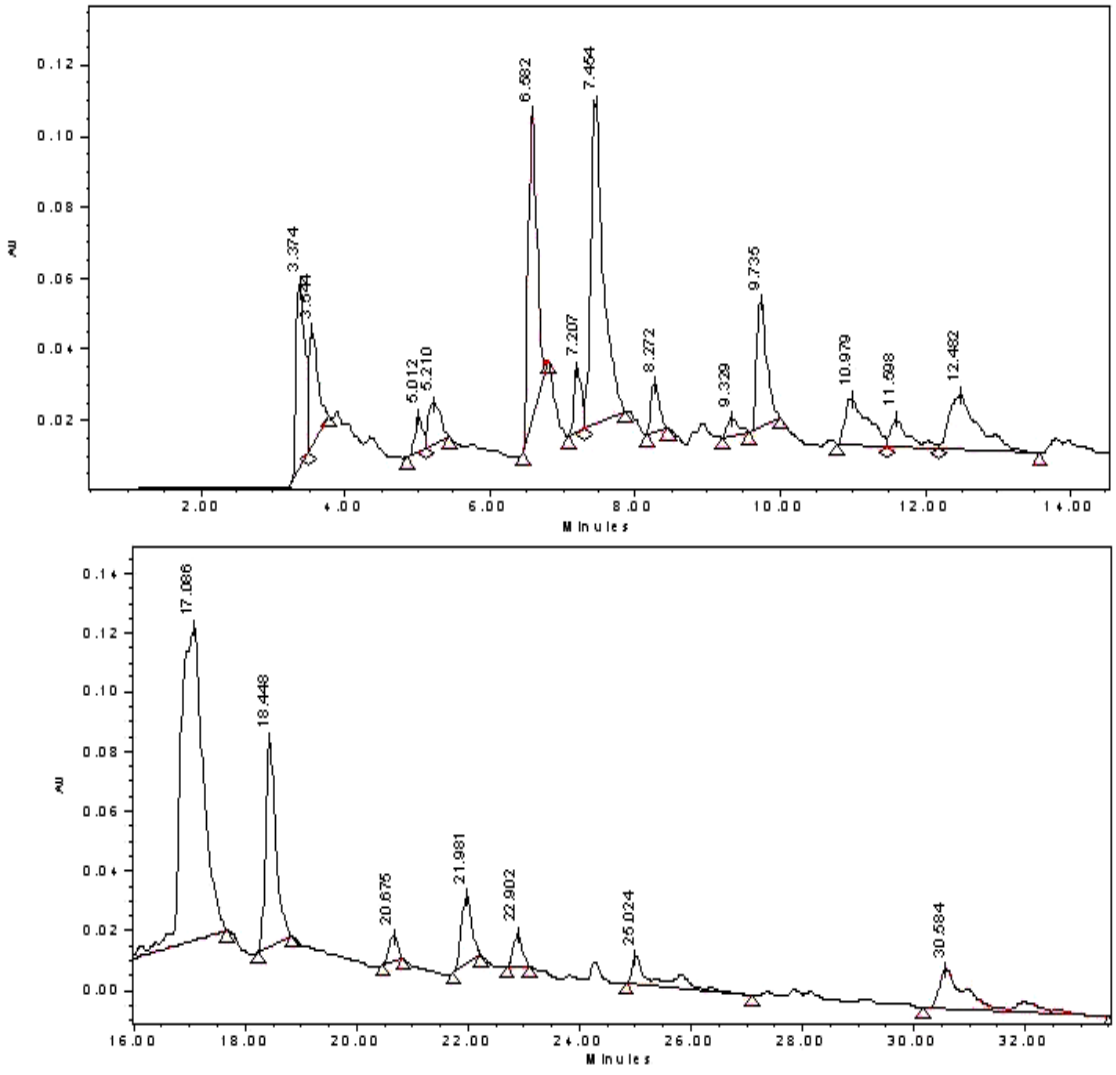
(A) [Transcinnamic-Pro-D-Arg-P1'-CONH2]



Each individual peak was collected and then subjected to ESI-MS(+) analysis to detrmien the mass of the peptide.

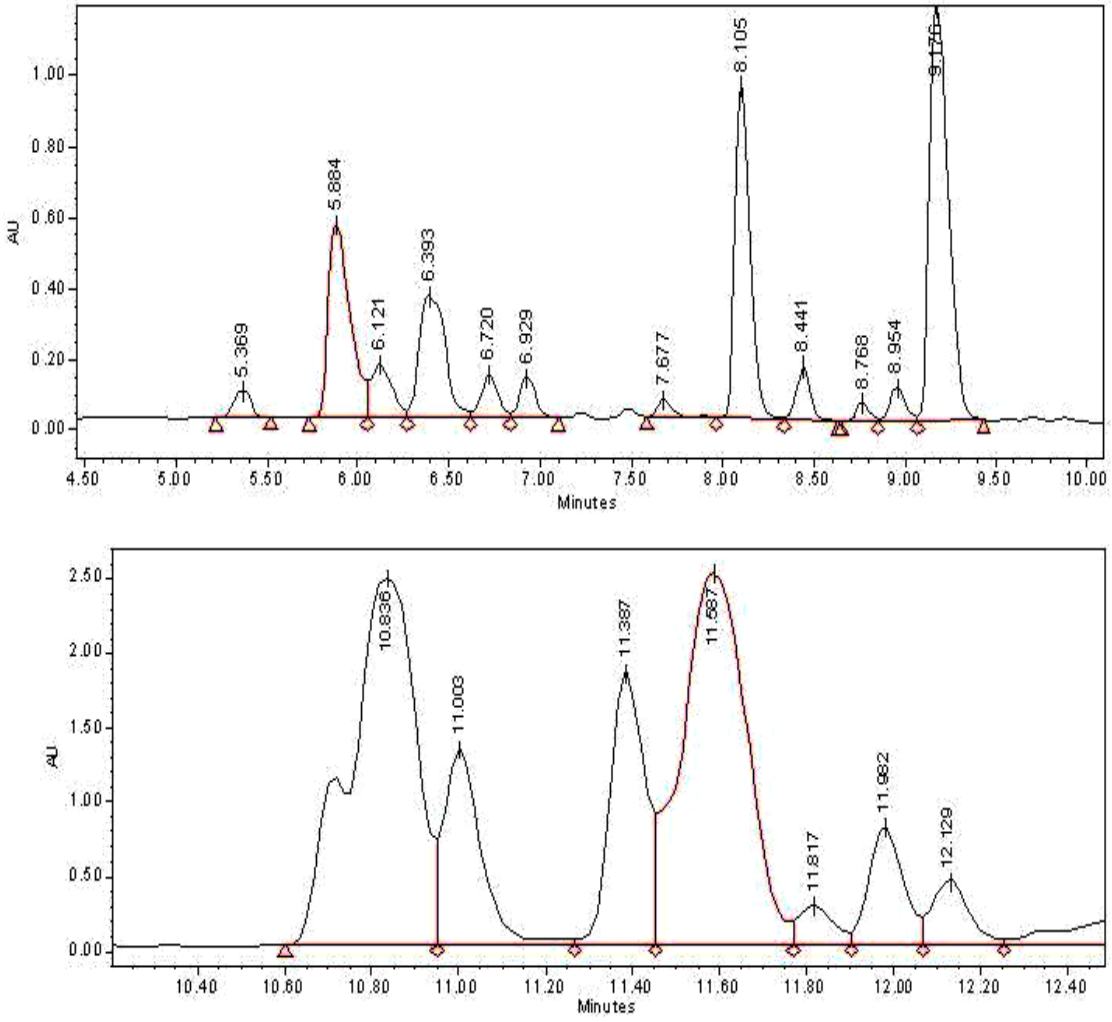
Figure 130-(B) RP-HPLC of tetrapeptide library

D-Tic-Pro-D-Arg-P1'-$CONH_2$



Each individual peak was collected and then subjected to ESI-MS(+) analysis to detrmien the mass of the peptide.

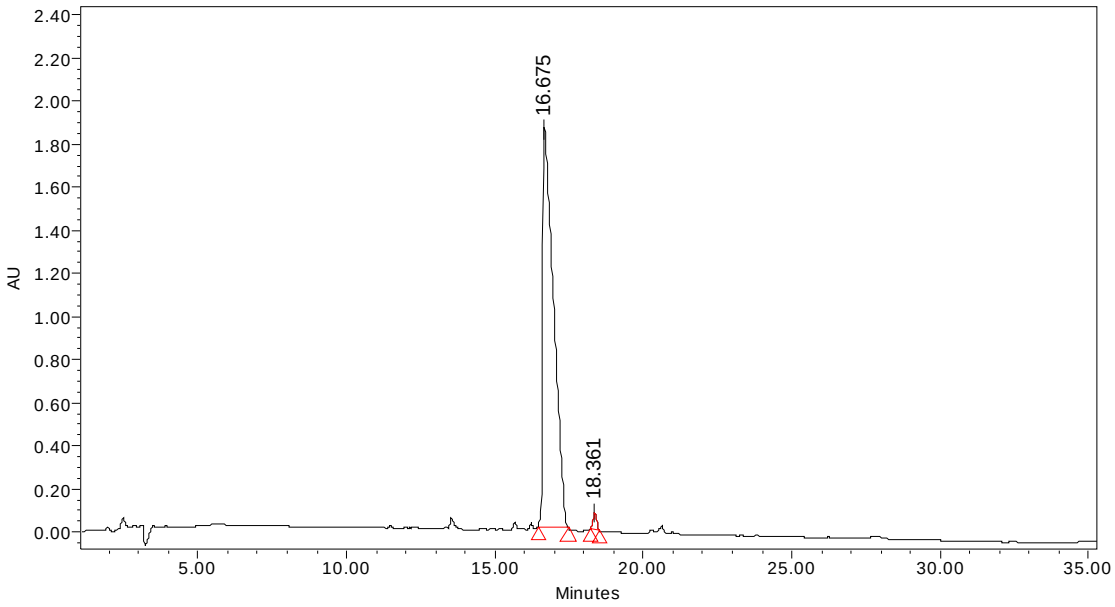
Figure 130-C: RP-HPLC of tetrapeptide library D-Nal-Pro-D-Arg-P1'-$CONH_2$



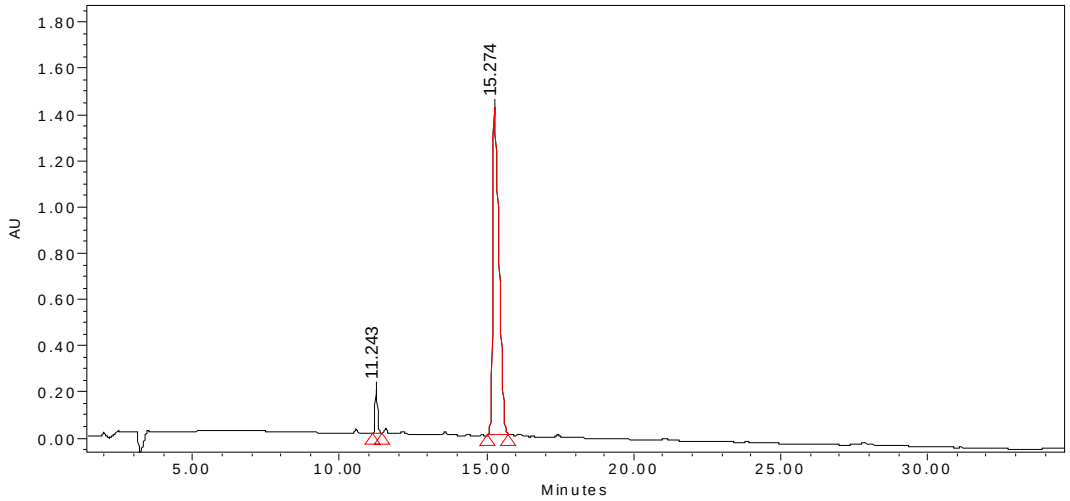
Each individual peak was collected and then subjected to ESI-MS(+) analysis to detrmien the mass of the peptide.

Figure 131 (A-J): HPLC-for some of the peptides inhibitors for thrombin

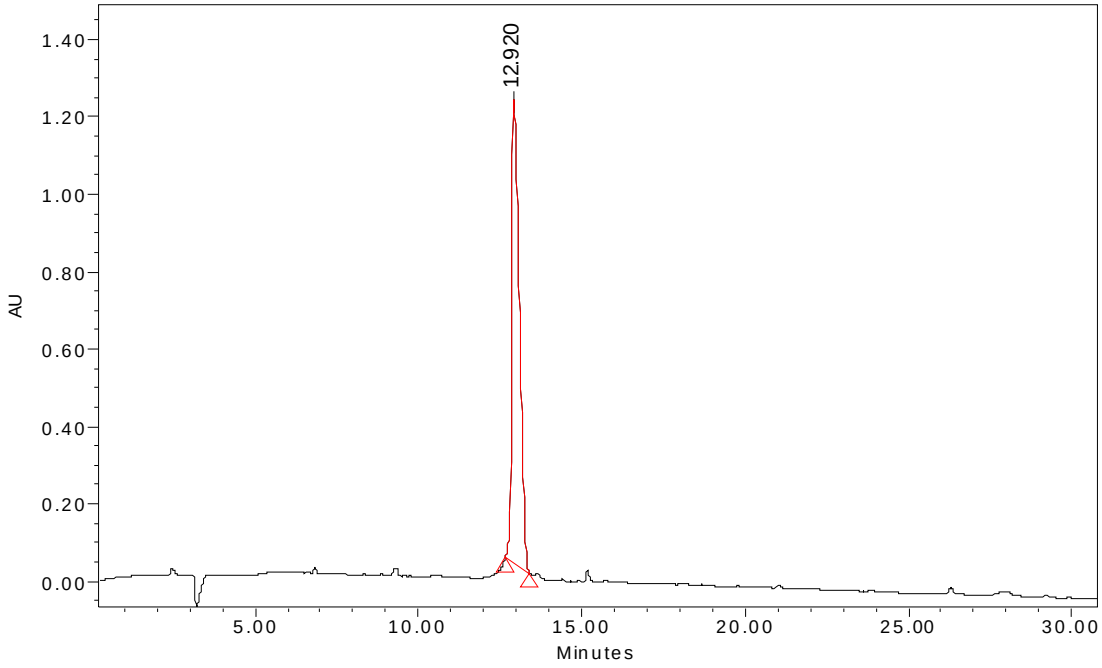
131-A: DPhe-Pro-DArg-DSer-$CONH_2$



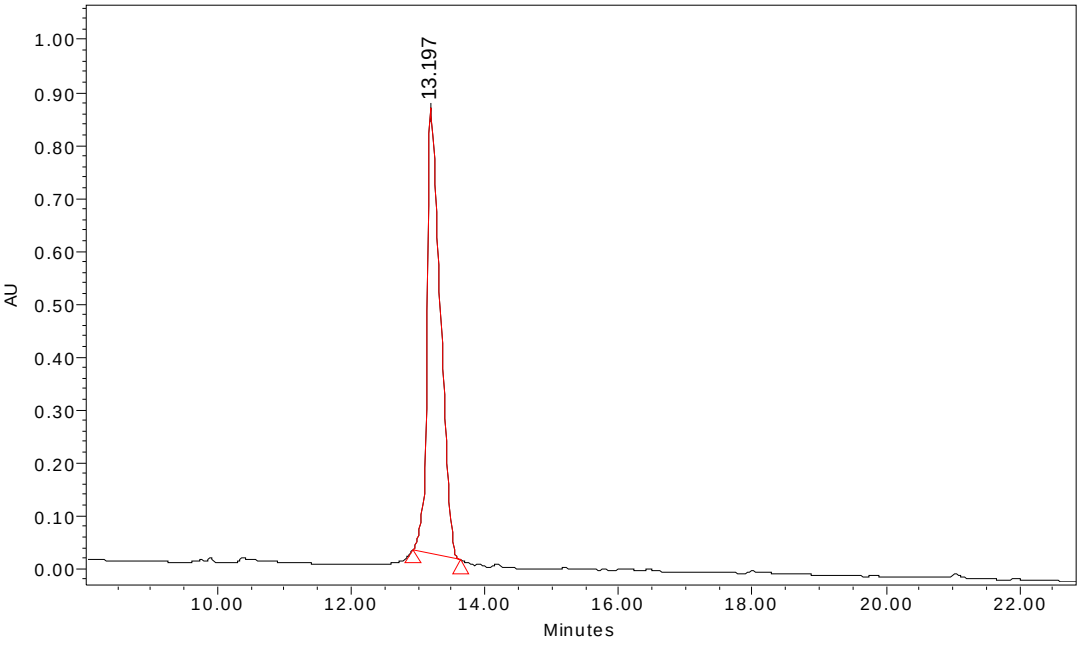
131-B: L-Tic-Pro-D-Arg-L-Thi-$CONH_2$



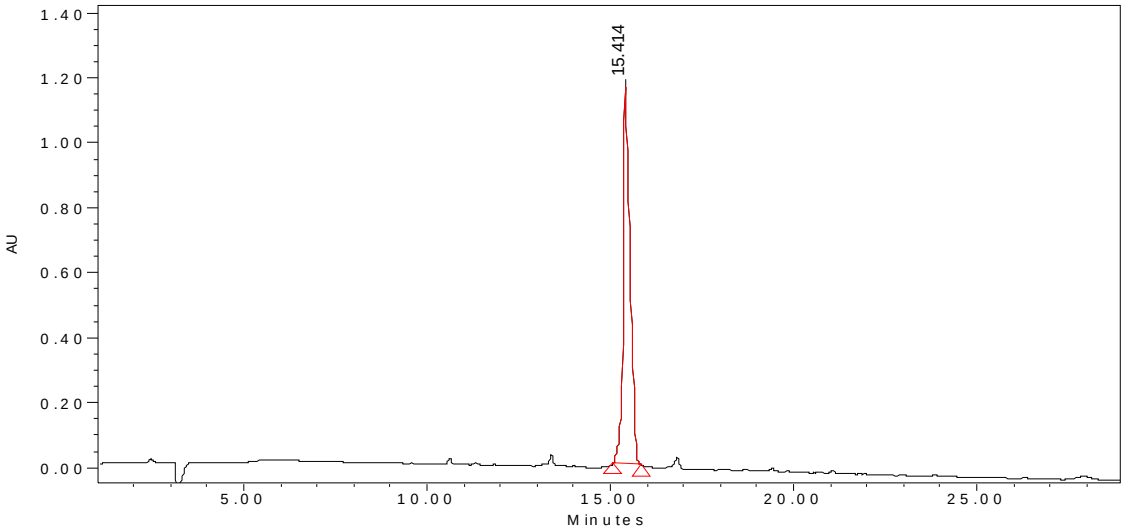
131-C: D-Phe-Pro-D-Arg-D-Gln-$CONH_2$



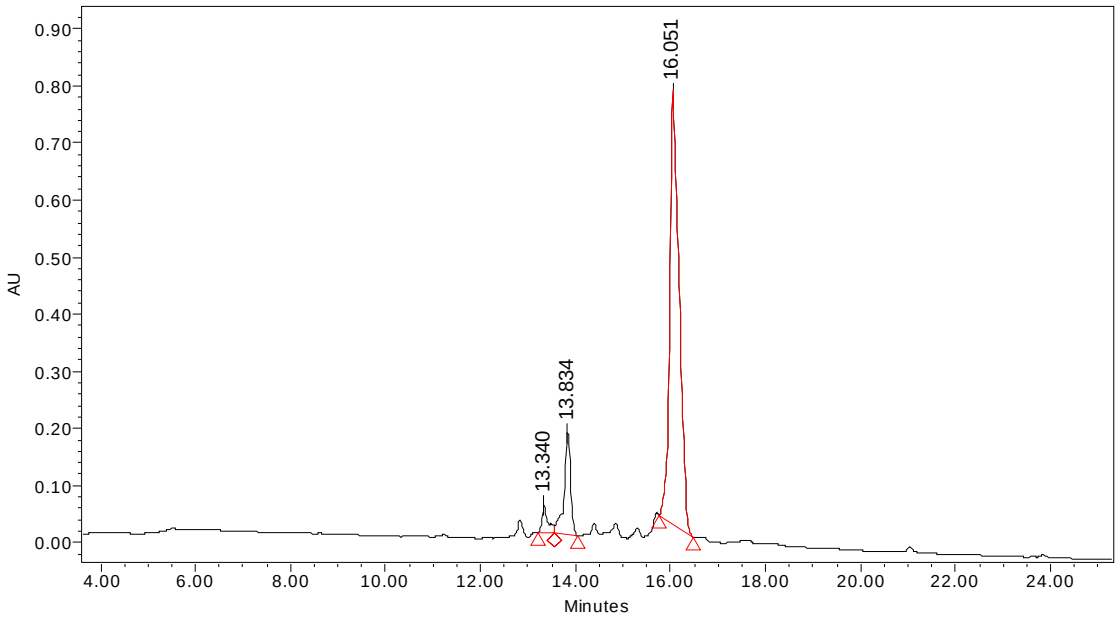
131-D: D-Phe-Pro-D-Arg-D-Thr-$CONH_2$



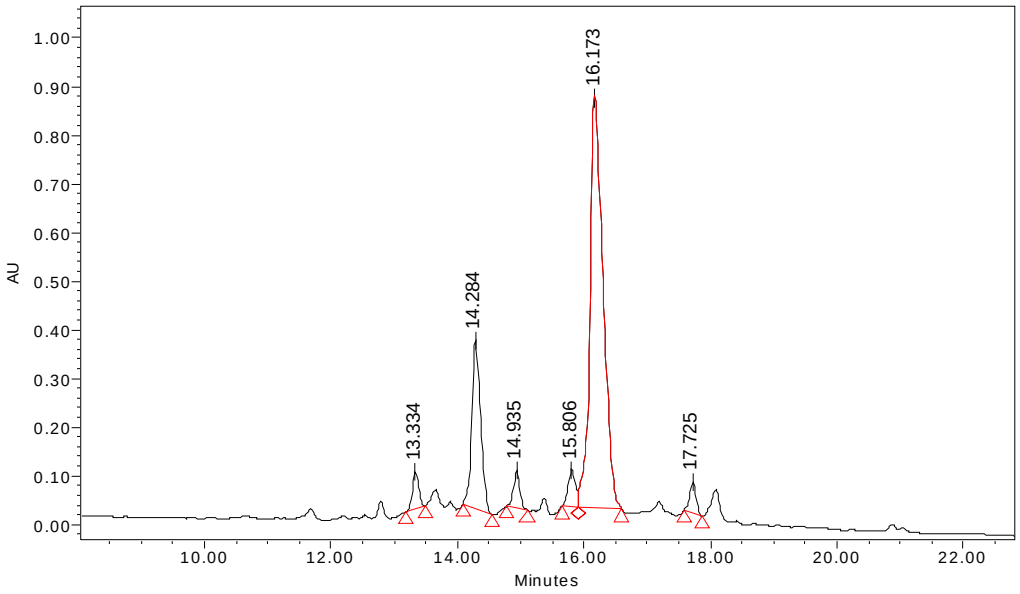
131-E: D-Phe-Pro-D-Arg-L-Thi-$CONH_2$



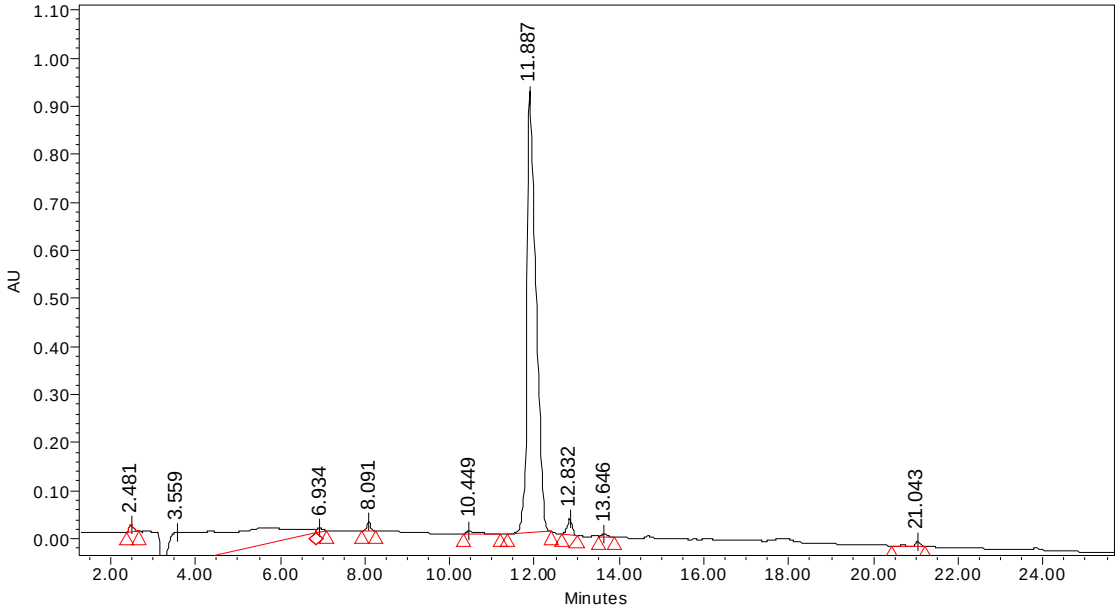
131-F: D-Phe-Pro-D-Arg-L-Cys-$CONH_2$



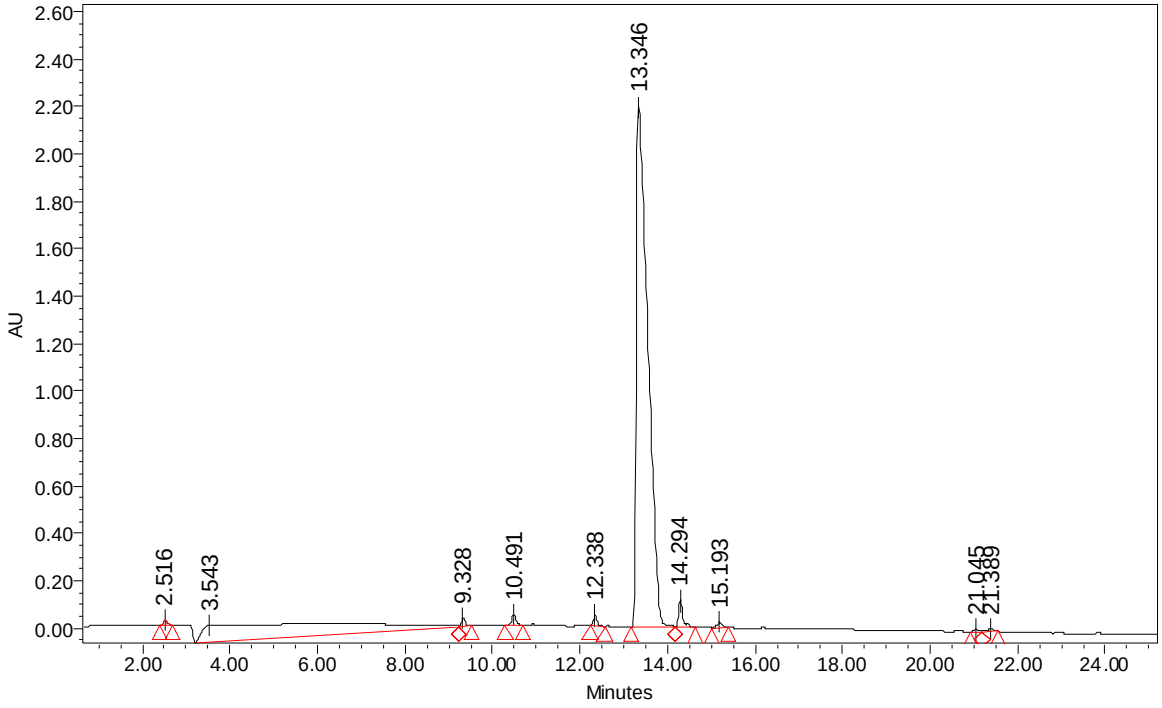
131-G: D-Phe-Pro-DArg-D-Cys-$CONH_2$



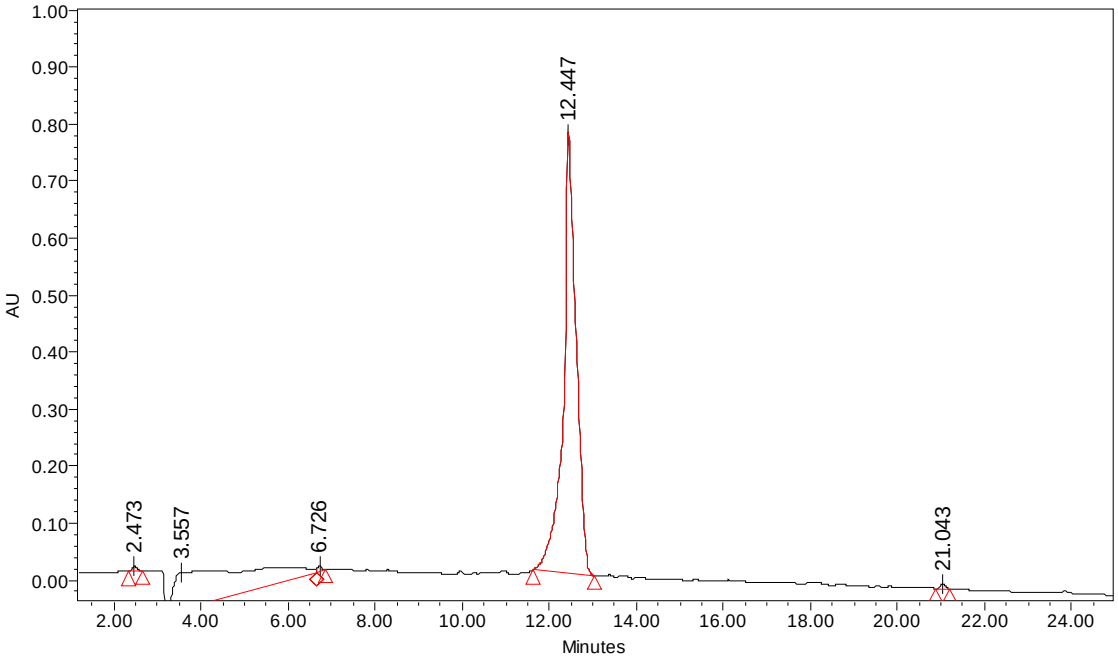
131-H: LTic-Pro-DArg-DAla-$CONH_2$



131-I: L-Tic-L-Thi-DArg-CONH$_2$



131-J: D-Tic-Pro-D-Arg-Gly-CONH$_2$

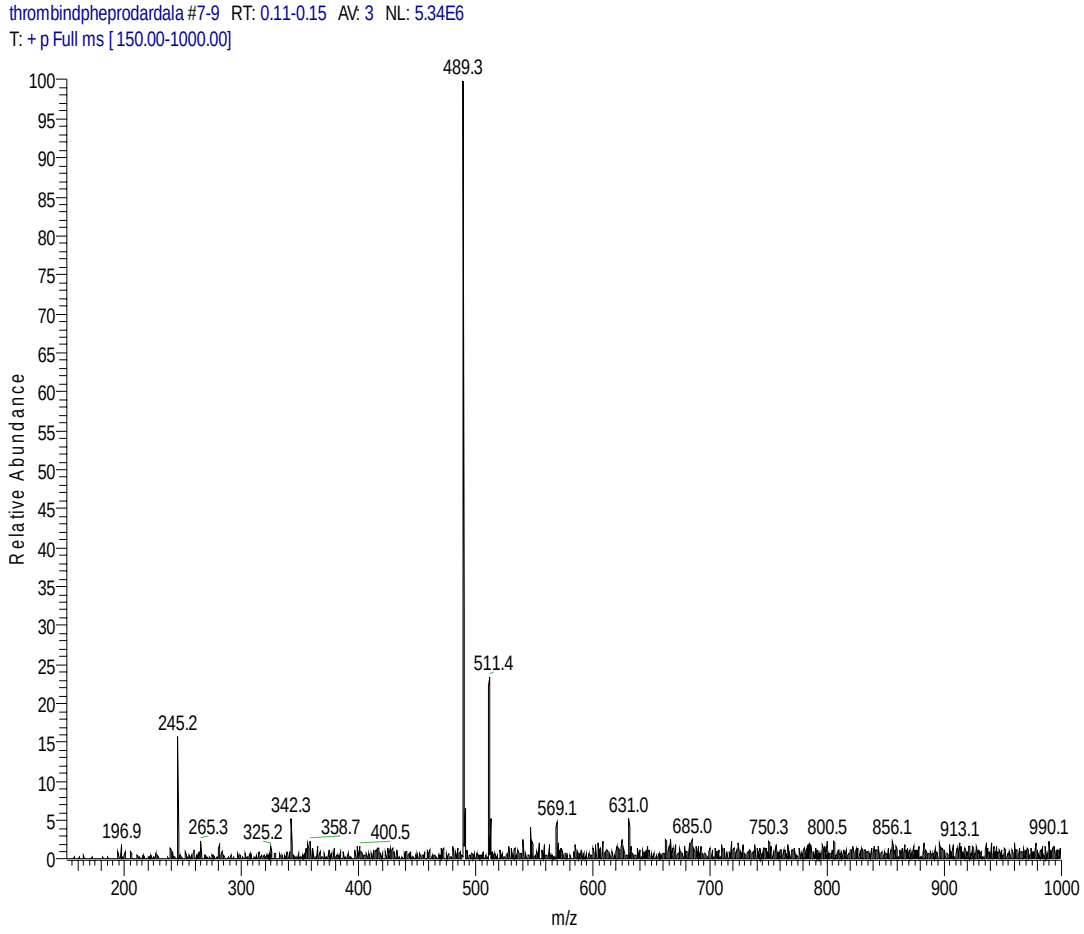


APPENDIX XII

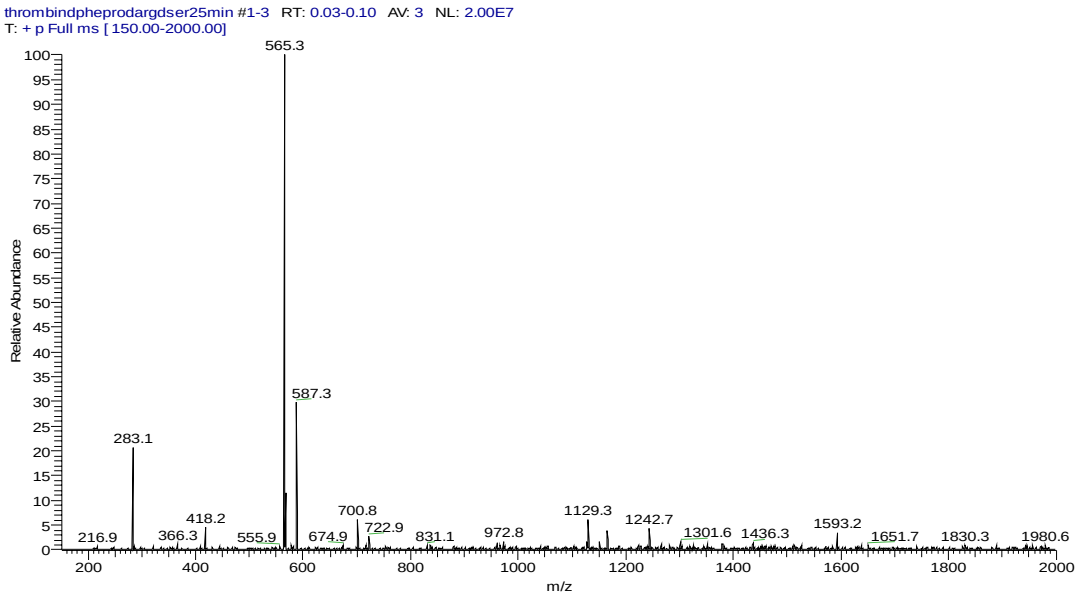
Figure 132: Control for stability to hydrolysis of peptides inhibitors using ESI-(+) mass spectroscopy.

The expected MW of tripeptide which would result by hydrolysis of peptide bond D-Arg-X [sequence space **D-Phe-Pro-D-Arg-X-CONH2]** is 418.5 The absence of the 419.5 m/z ratio from the ESI-MS scan demonstrates that the peptide was stable to hydrolysis by thrombin.

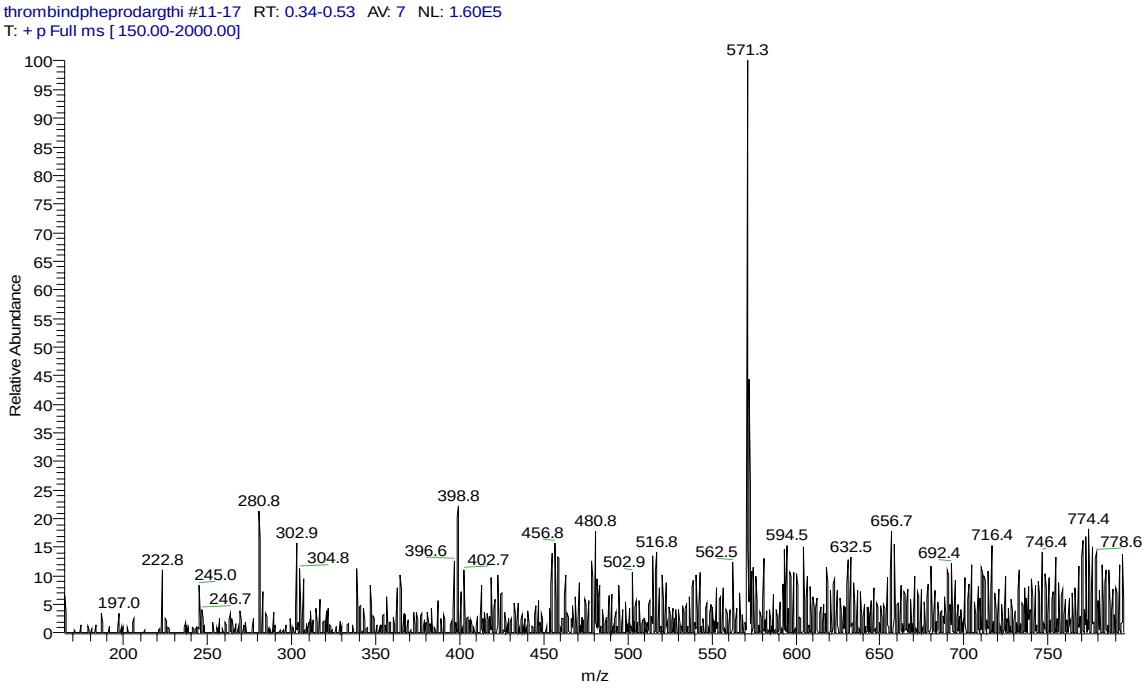
(A) Thrombin (30 nM) + peptide D-Phe-Pro-D-Arg-D-ALA-CONH2 (20 μM) (MW=488.3, m/z =489.3) incubated for 30 minutes at room temperature in the same buffer used for kinetic studies (see Materials and Methods Section 3.2).



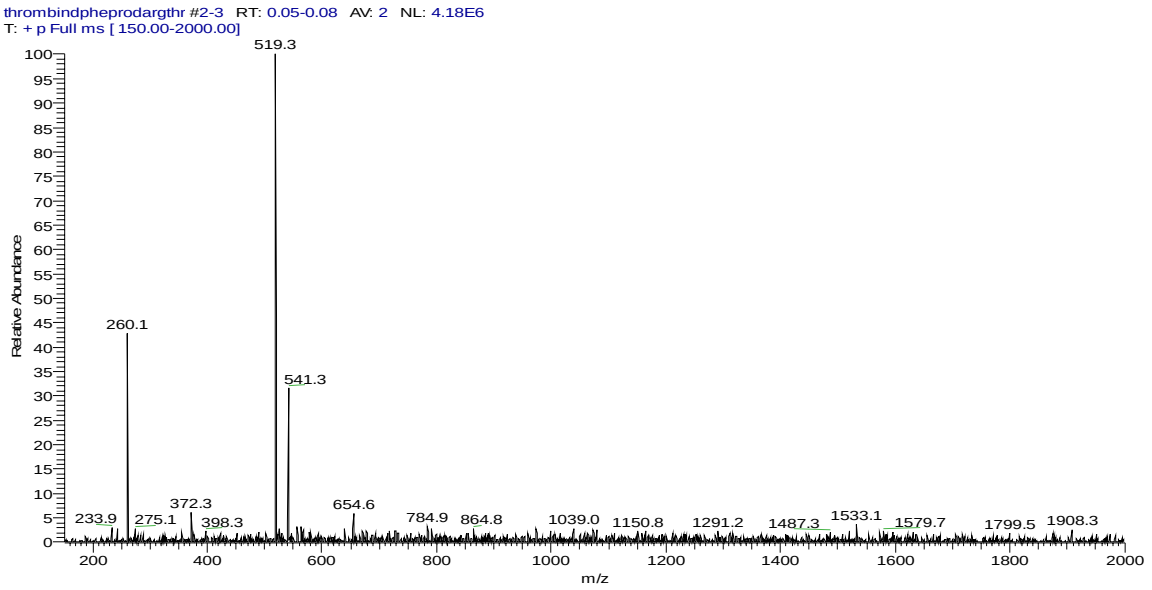
(B) Thrombin (30 nM) + peptide D-Phe-Pro-D-Arg-D-Ser-CONH2 (20 μM) (MW=564.6, m/z = 565.3) incubated for 30 minutes at room temperature in the same buffer used for kinetic studies (see Materials and Methods Section 3.2).



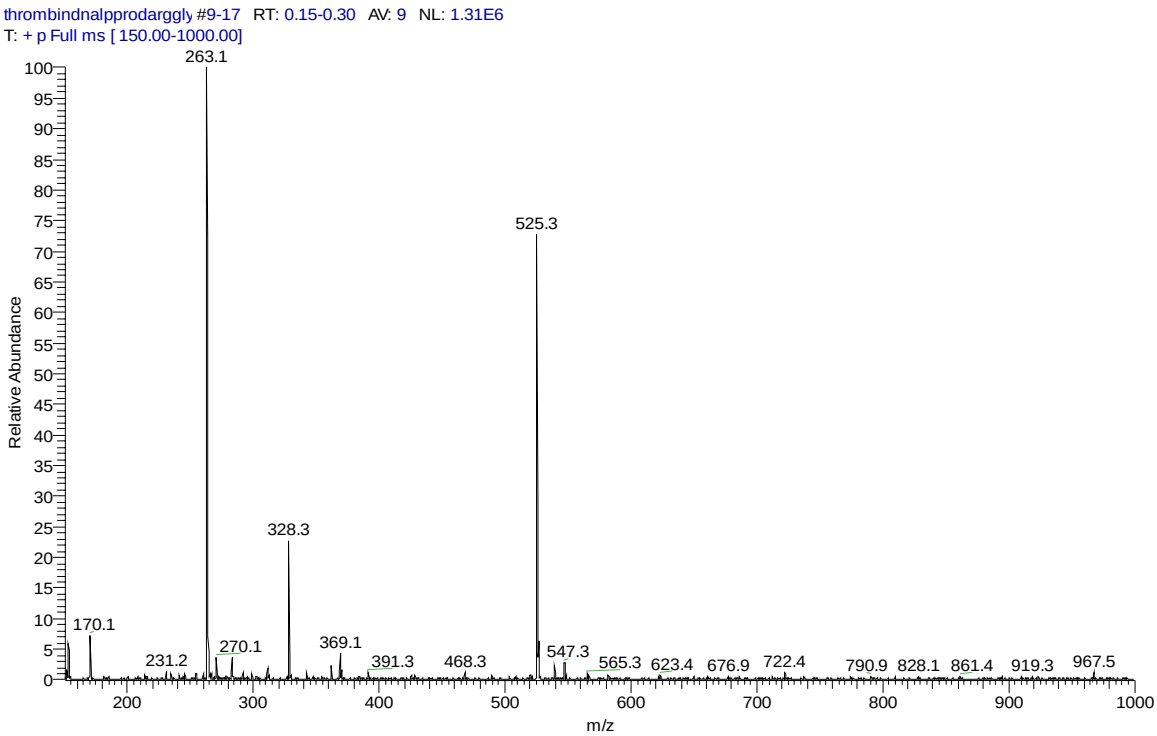
(C) Thrombin (30 nM) + peptide D-Phe-Pro-D-Arg-Thi-CONH2 (20 μM) (MW=571.6, m/z = 571.3) incubated for 30 minutes at room temperature in the same buffer used for kinetic studies (see Materials and Methods Section 3.2).



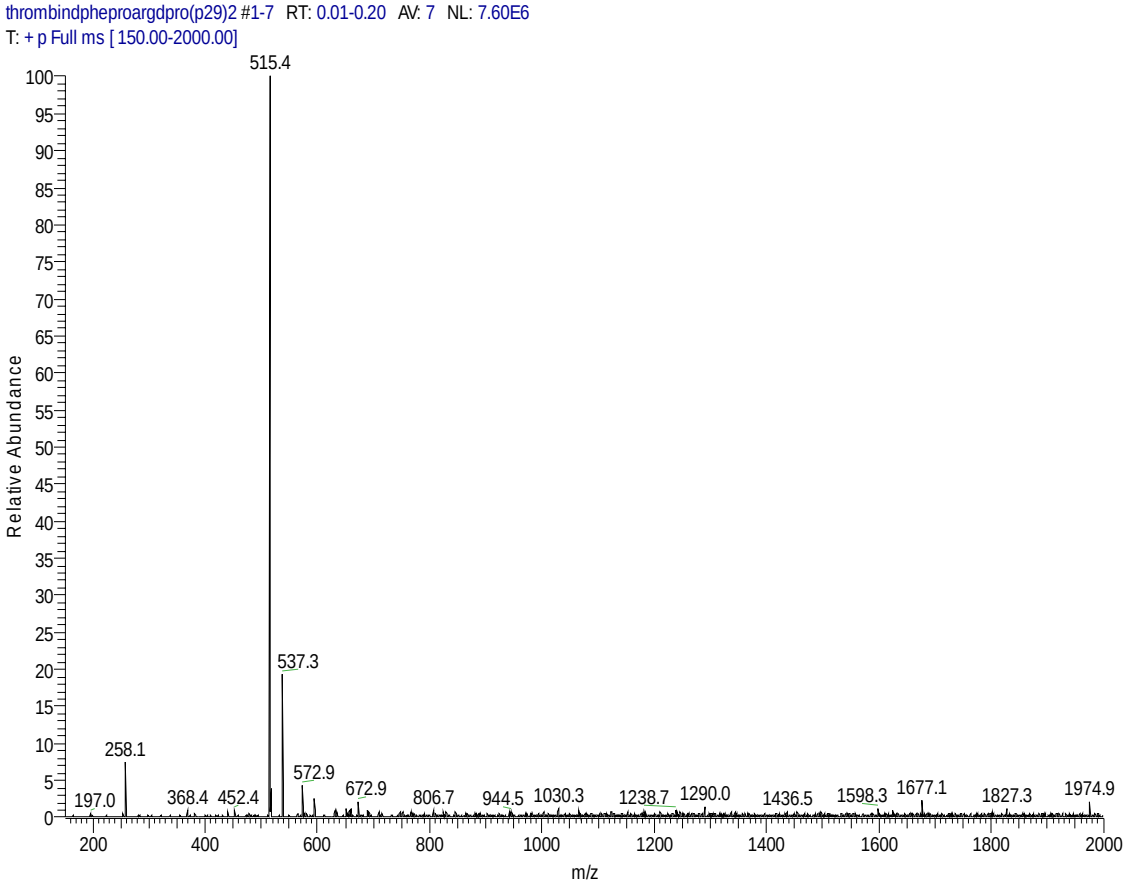
(D) Thrombin (30 nM) + peptide D-Phe-Pro-D-Arg-Thr-CONH2 (20 μM) (MW=518.4, m/z = 519.4) incubated for 30 minutes at room temperature in the same buffer used for kinetic studies (see Materials and Methods Section 3.2).



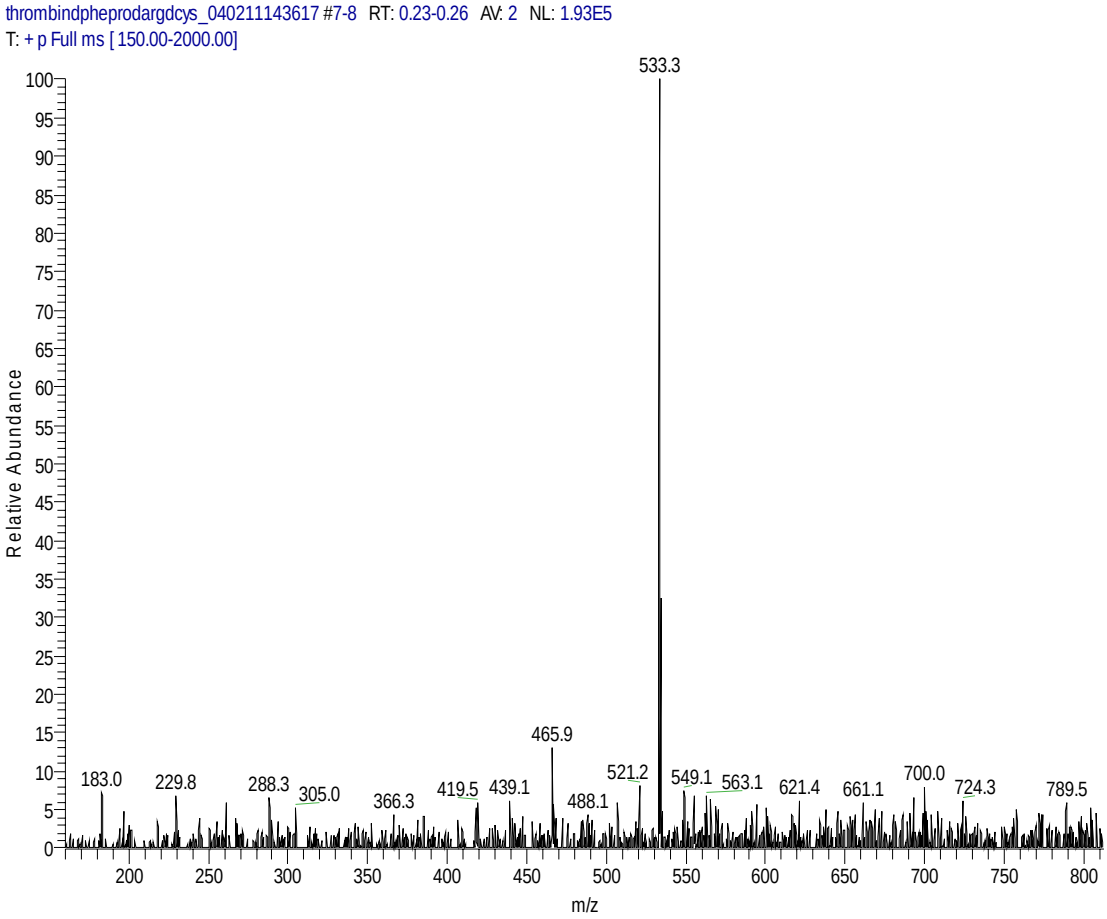
(E) Thrombin (30 nM) + peptide D-Nal-Pro-D-Arg-Gly-CONH2 (20 μM) (MW=524.4, m/z = 525.3) incubated for 30 minutes at room temperature in the same buffer used for kinetic studies (see Materials and Methods Section 3.2).



(F) Thrombin (30 nM) + peptide D-Phe-Pro-Arg-D-Pro-CONH2 (20 μM) (MW=514.4, m/z = 515.4) incubated for 30 minutes at room temperature in the same buffer used for kinetic studies (see Materials and Methods Section 3.2).



(G) Thrombin (30 nM) + peptide D-Phe-Pro-D-Arg-D-Cys-CONH2 (20 μM) (MW=532.4, m/z = 533.4) incubated for 30 minutes at room temperature in the same buffer used for kinetic studies (see Materials and Methods Section 3.2).



APPENDIX XIII

Figure 133. Titration of peptide [D-Phe-Pro-D-Arg-Ala-CONH2] into thrombin.

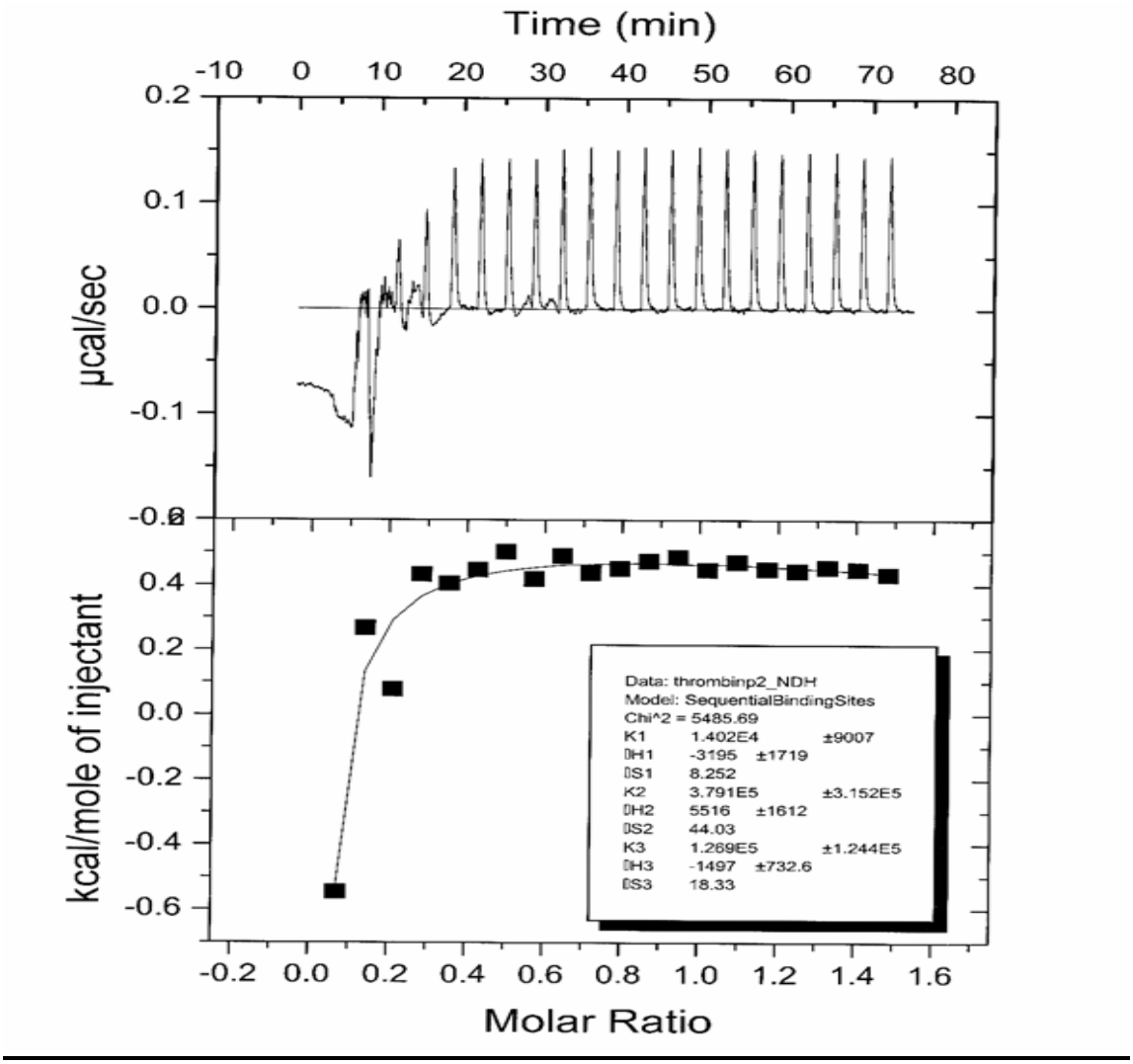


Figure 134. Titration of peptide [DPhe-Pro-DArg-Gly-CONH2] into thrombin.

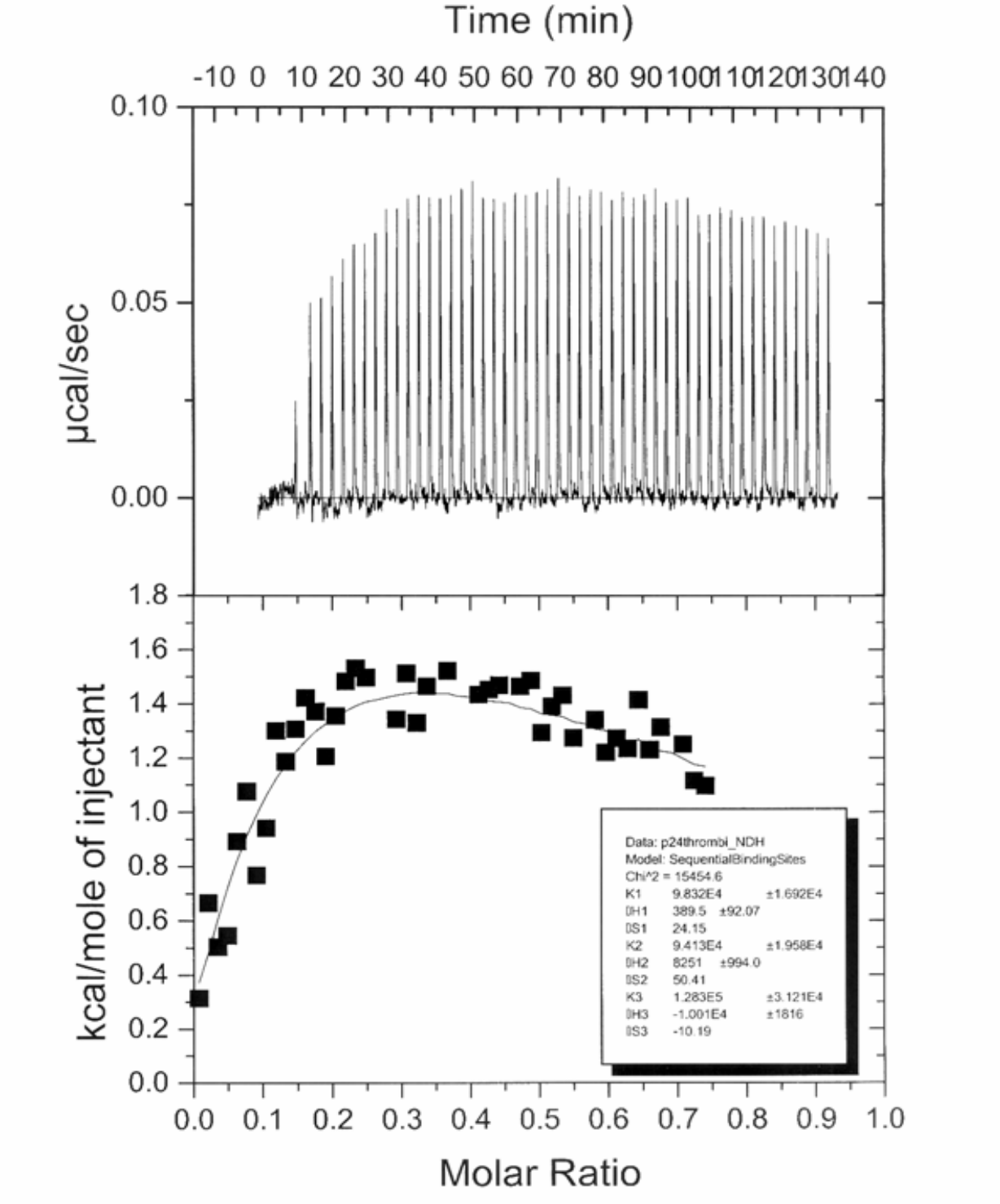
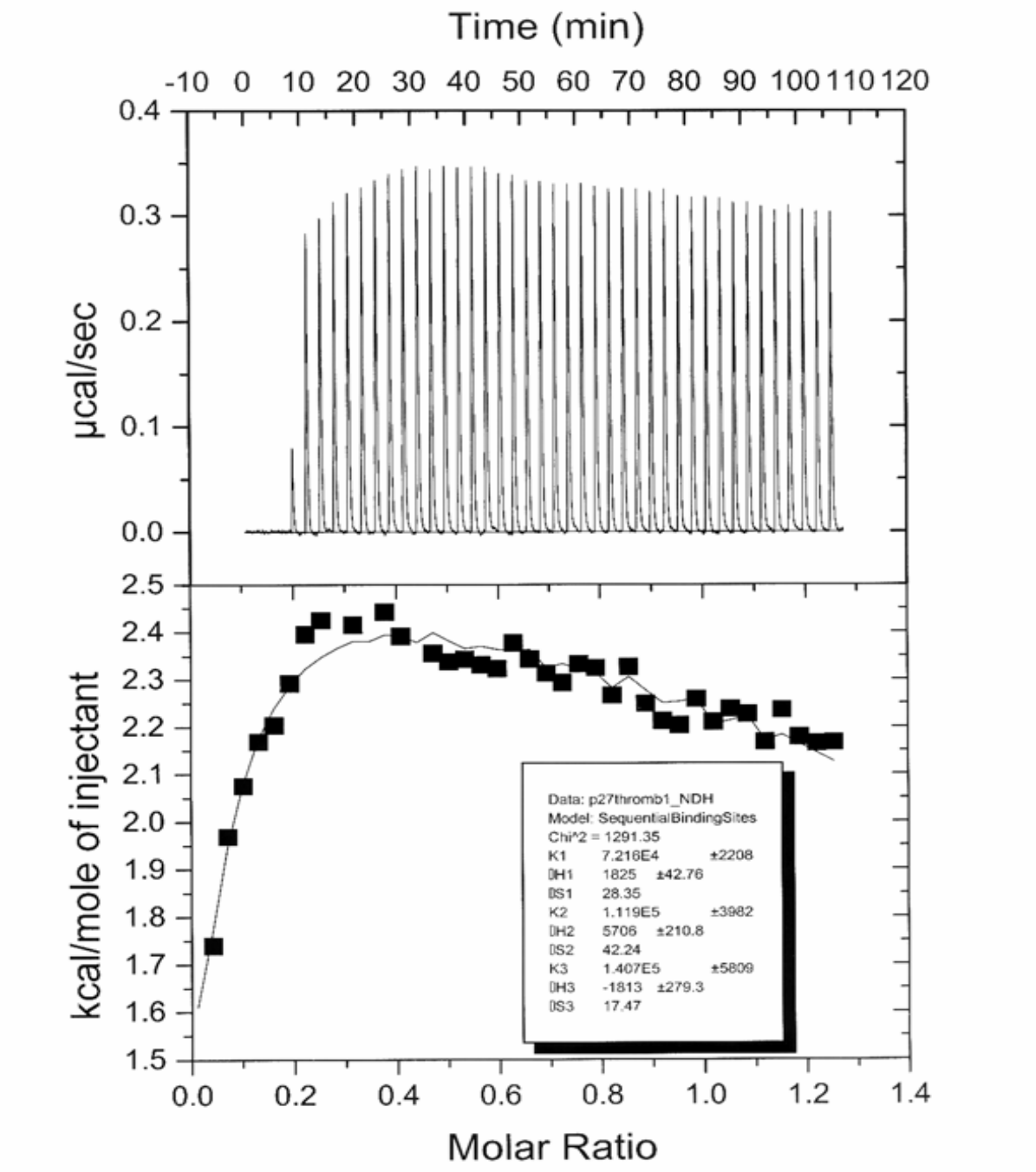


Figure 135. Titration of peptide [DPhe-Pro-DArg-Cys-CONH2] into thrombin.



REFERENCES

1. V.J. Basus, Proton nuclear magnetic resonance assignments. *Methods Enzymol.* **177** (1989), pp. 132–149.

2. C. Blackmar, V.L. Healy, R. Hrabal, F. Ni and E.A. Komives, Structure/activity of the region of thrombomodulin that binds to thrombin. *Bioorg. Chem.* **23** (1995), pp. 519–527.

3. W. Braun and N. Go, Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J. Mol. Biol* **186** (1985), pp. 611–626.

4. D.G. Davis, Elimination of baseline distortions and minimization of artifacts from phased 2D NMR spectra. *J. Magn. Reson.* **81** (1989), pp. 603–607.

5. Wüthrich, K. 1986. *NMR of proteins and nucleic acids./ITL J. Wiley, New York.*

6. Wüthrich, K., Billeter, M., and Braun, W. 1984. Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton–proton distances. *J. Mol. Biol.* **180:** 715–740.

7. C. Griesinger, G. Otting, K. Wuthrich and R.R. Ernst, Clean TOCSY for 1H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110** (1988), pp. 7870–7872.

8. T. Hayashi, M. Zushi, S. Yamamoto and K. Suzuki, Further localization of binding sites for thrombin and protein C in human thrombomodulin. *J. Biol. Chem.* **265** (1990), pp. 20156–2015.

9. R. Hrabal, E.A. Komives and F. Ni, Structural resiliency of an EGF-like subdomain bound to its target protein, thrombin. *Protein Sci.* **5** (1996), pp. 195–203.

10. D.S. Wishart, B.D. Sykes and F.M. Richards, The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31** (1992), pp. 1647–1651.

11. G. Wagner, W. Braun, T.F. Havel, T. Shaumann, N. Go and K. Wutrich, Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196** (1987), pp. 611–639.

12. J. Srinivasan, S. Hu, R. Hrabal, Y. Zhu, E.A. Komives and F. Ni, Thrombin-bound structure of an EGF subdomain from human thrombomodulin determined by transferred nuclear overhauser effects. *Biochemistry* **33** (1994), pp. 13553–13561.

13. G.T. Montelione, K. Wuthrich, A.W. Burgess, E.C. Nice, G. Wagner, K.D. Gibson and H.A. Scheraga, Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization with restraints. *Biochemistry* **31** (1992), pp. 236–249.

14. G.T. Montelione, M.E. Winkler, L.E. Burton, E. Rinderknecht, M.B. Sporn and G. Wagner, Sequence-specific 1H-NMR assignments and identification of two small antiparallel beta-sheets in the solution structure of recombinant human transforming growth factor alpha. *Proc. Natl Acad. Sci. USA* **86** (1989), pp. 1519–1523.

15. A. Kumar, R.R. Ernst and K. Wuthrich, A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95** (1980), pp. 1–6.

Autobiographical Statement

Cristina C. Clement was born in Bucharest, Romania, on 24th April 1967. She attended the public schools in that country and she graduated with Bachelor of Science (B.S.) in Biochemistry in 1991 from University of Bucharest, School of Biological Sciences. In 1995 she came to USA as Ph.D. student to the University of Missouri-Kansas City from were she graduated with Master in Cell and Molecular Biology in December 1998. In August 1998 she was admitted to the Ph.D. Program in Biochemistry at City University of New York (CUNY) were she was pursuing her Ph.D. thesis in the labs of Professor Maria Tomasz at Hunter College and Professor Manfred Philipp at Lehman College.