10-31-1993

Computer assisted molecular modeling of thymidine nucleoside analog inhibitors of HIV-1 reverse transcriptase

Mark A. Fisher
New Jersey Institute of Technology

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

Computer Assisted Molecular Modeling of Thymidine Nucleoside Analog Inhibitors of HIV-1 Reverse Transcriptase

by
Mark A. Fisher

An active analog approach to receptor mapping was used to identify the three dimensional structural characteristics associated with a series of thymidine nucleoside analog inhibitors of HIV-1 reverse transcriptase (RT) that may be essential for their activity. Atomic substitutions at the 5 and 3' positions of nucleoside analogs confer global structural and electrostatic changes that result in either increased or diminished inhibitory activity. From a structural perspective, the activity differences can be attributed to the presentation of select atoms in three dimensional motifs that are common to all active compounds and absent or distorted in inactive/poorly active compounds.

The identification of these characteristics will complement more direct studies of the RT structure by providing a specific three dimensional orientation for substrate and inhibitor molecules at their receptor site. They can also serve as a three dimensional template for the screening of potentially active compounds; thus, aiding in the development and identification of new, more potent and selective inhibitor molecules. The characteristics identified are common to 15 thymidine nucleoside analog inhibitors of RT and have allowed the inference of a three dimensional map of the HIV-1 RT receptor site.

COMPUTER ASSISTED MOLECULAR MODELING OF THYMIDINE NUCLEOSIDE ANALOG INHIBITORS OF HIV-1 REVERSE TRANSCRIPTASE

by
Mark A. Fisher

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Biomedical Engineering

Biomedical Engineering Committee

October 1993

APPROVAL PAGE

Computer Assisted Molecular Modeling of Thymidine Nucleoside Analog Inhibitors of HIV-1 Reverse Transcriptase

Mark A. Fisher

Dr. David S. Kristol, Thesis Advisor — Date
Professor of Chemistry and Chemical Engineering,
NJIT

Dr. Janardan S. Yadav, Co-Advisor — Date
Academic Computing Services, Department of
Information Services and Technology, University of
Medicine and Dentistry of New Jersey (UMDNJ)

Dr. Mukund J. Modak, Co-Advisor — Date
Professor of Biochemistry and Molecular Biology,
UMDNJ-New Jersey Medical School

BIOGRAPHICAL SKETCH

Author: Mark A. Fisher

Degree: Master of Science in Biomedical Engineering

Date: October 1993

Undergraduate and Graduate Education:

- Bachelor of Science in Electrical Engineering
  New Jersey Institute of Technology, Newark, NJ, 1991
- Master of Science in Biomedical Engineering
  New Jersey Institute of Technology, Newark, NJ, 1993

Major: Biomedical Engineering

Presentations and Publications:

Fisher, M.A., Yadav, P., Arnold, E., Kristol, D.S., Modak, M.J. 1993. Identification of a "Pharmacophore" for Nucleoside Analog Inhibitors Directed at HIV-1 Reverse Transcriptase. *J. Molecular Recognition* (submitted August 1993).

Fisher, M.A., Yadav, P.N.S., Yadav, J., Laxminarayan, S., Kristol, D., Modak, M. A Computer Assisted Receptor Mapping Approach to the Design of Anti-AIDS Agents Directed at HIV Reverse Transcriptase. *19th Annual Northeast Bioengineering Conference Proceedings*. New Jersey Institute of Technology, July 1993.

Fisher, M.A., Yadav, P., Yadav, J. Arnold, E., Modak, M., Identification of "Pharmacophore" for Deoxythymidine Inhibitor Analogs of HIV RT: A Receptor Mapping Study. (Abstract) *Structure-Function of HIV and Novel Approaches to AIDS Therapy Conference*. Center for Advanced Biotechnology and Medicine-Rutgers Univ./UMDNJ, October 1992.

This thesis is dedicated to Mom, Pop, Rob, Chris, and Renée for their steadfast support throughout my academic career.

ACKNOWLEDGMENT

I wish to express my sincere gratitude and appreciation to my advisors Dr. Janardan Yadav, Professor Mukund Modak, and Professor David Kristol for providing an opportunity to work on a project as challenging and rewarding as that which is detailed in this work. Their thoughtful and productive guidance both inside and out of the laboratory has helped to foster my interest in the pursuit of scientific excellence.

A special thanks to my friend and colleague Dr. Prem N.S. Yadav, whose knowledge, wisdom, and constructive criticisms have contributed greatly to my appreciation of the intricacies of computer assisted molecular modeling. The thoughtful and constructive advice and expertise of the members of the Biochemistry and Molecular Biology Department at New Jersey Medical School, including: Stefan Serafianos, Shyamal Desai, Virendra Pandey, Neerja Kaushik, Richard Whipple, and Kajal Chowdhury was greatly appreciated.

This work would not be possible were it not for the facilities provided by Dr. Leslie Michelson and Dr. Swamy Laxminarayan of the Academic Computing Services Department at UMDNJ. The use of their well maintained academic computing center was also very much appreciated.

The author is grateful to the National Institutes of Health for funding this research through a Minority Research Supplement award under the umbrella of Grant # GM-36307.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
BACKGROUND INFORMATION

1.1 Introduction

HIV-1 RT is an RNA dependent DNA polymerase class of enzyme that is required for a productive viral infection. It is responsible for the conversion of a virus' single stranded RNA genome to a double stranded proviral DNA form that can be incorporated into the genetic information of a host cell. The catalysis of such a critical step in the viral life cycle has mandated the use of inhibitors specifically targeted at RT as anti-AIDS agents. Nucleoside analogs that lack the 3'-hydroxyl group (such as AZT, ddC, ddI, D4T, etc.) have been studied extensively as potential anti-viral agents. The incorporation of these compounds into a growing DNA chain results in chain termination due to their inability to form a 3',5'-phosphodiester link with the incoming nucleotide. All nucleoside analogs are assumed to act as competitive inhibitors of HIV-1 RT following intracellular conversion to their triphosphate (TP) form [1]. In many cases the inhibitor binding constant (AZT-TP, K_I=4.0nM) is stronger than that of the natural substrate (dTTP, K_I=500nM) [1]. For this reason, it is not fully established whether chain termination, reversible competitive binding, or both are primarily responsible for inhibitory activity. Although nucleoside analogs preferentially interact with RT at the TP level, they are recognized by other retroviral RTs and even cellular DNA polymerases. It is this nonspecific action that contributes to their toxic side effects. In therapeutically effective compounds the cytotoxic concentration is well above that required for antiviral activity.

The size, three dimensional shape, energetic, and electronic characteristics of a molecule play an important role in its recognition and interactions with

biological systems. It has been observed that nucleoside analogs that are structurally closely related have very different levels of anti-HIV activity [1,2,3,4]. This has prompted a search for the identification of the exact three dimensional structural characteristics that are responsible for the activity differences amongst these compounds. The correlation of bioactivity with three dimensional structural information will provide a molecular basis for the recognition of these compounds that can be extrapolated to the de novo design of more potent and specifically targeted RT inhibitors. I have limited this study to nucleosides of the thymidine and uridine base motif because the skeletal homology and competitive nature of these analogs suggest a common mode and location of binding at RT, i.e. the same three dimensional structural characteristics in each analog are recognized by the enzyme. With this assumption, the identification of these common features becomes a matter of comparing the structures of a number of inhibitors for common three dimensional spatial arrangements. There is a high degree of flexibility in nucleoside molecules which increases the potential for induced conformational changes during the formation of protein-ligand complexes. The potential for induced conformational changes in these compounds precludes conclusions which are based upon the comparison of static X-ray crystallographic structures. An approach which allows the comparison of a large number of compounds and their sterically allowed molecular conformations for common spatial orientations is most appropriate to this class of compound.

1.2 Historical Perspective

In the past, the computational demands of representing molecular structures in the "real world" medium of three dimensions limited structure-based studies to the use of hand held models. Recent advances in the development of high speed

computer hardware and software algorithms, which support the real time visualization and manipulation of molecular systems, have mandated the development of computer based molecular modeling systems for the analysis of three dimensional molecular characteristics. The use of molecular modeling protocols, which are based upon the hypothesis of a common structural basis for the activity of different compounds at a common receptor site, have been successfully utilized to identify three-dimensional features that are important for biological activity; and extrapolated to the design of potent new classes of biologically active compounds. There are essentially two approaches to the modeling of biomolecular systems each of which is dependent upon the amount and type of experimental data that is available. The direct approach is based upon experimental data associated directly with a macromolecular structure of interest and generally requires an fairly well resolved X-ray crystallographic structure to start. The indirect approach focuses upon experimental data related to ligands and is used to infer the three dimensional structural features of a macromolecules catalytic region based upon ligand complementarity.

An impressive example of the potential of the direct approach is the design of two new influenza virus salidase inhibitors using computer assisted molecular modeling protocols that improved upon the amount of structural and electronic complementarity between the enzyme receptor site and a bound inhibitor [5]. The study was begun with an X-ray crystallographic structure of a known inhibitor (2-deoxy-2,3-didehydro-D-N-acetyl-neuraminic acid), which resembles the product of the sialidase-catalyzed reaction, complexed with the influenza virus salidase enzyme. The salidase enzyme is a viral surface protein that assists in the cellular release of newly synthesized virus and the movement of virus through the mucus membrane of the respiratory tract. It is a glycohydrolase, which cleaves the terminal α-ketosidically linked sialic acids of glycoproteins,

glycolipids, and oligo-saccharides into sialic acid monomers. The X-ray crystallographic sialidase-inhibitor complex provides the bound molecular conformation of the inhibitor and its three dimensional position with respect to the amino acid residues in the sialidase receptor site. Functional group substitutions were made to the inhibitor which were energetically favored, i.e. that reduced the energy content of the enzyme-ligand complex. The most energetically favorable substitutions were used to design inhibitor analogs that would theoretically form preferred complexes. Two new sialic acid analogs (4-amino-2-deoxy-2,3-didehydro-D-N-acetylneuraminic acid and 4-guanidino-2-deoxy-2,3-didehydro-D-N-acetylneuraminic acid) were designed using this approach. These compounds were, subsequently, synthesized and assayed for sialidase inhibitory activity, and found to be the most potent sialidase inhibitors reported to date. Similar approaches have been used in the design of renin inhibitors [6,7] and peptide inhibitors of the HIV-1 protease [8].

In the absence of experimental data related to macromolecular structure (such as X-ray and biochemical analysis of structure and catalytic amino acid residues), which would provide both the volume and location of the receptor site along with the receptor bound orientation of substrate, an indirect modeling approach is required. Marshall et al. has proposed an "active analog approach" which is based upon the hypothesis that the enzymatic recognition of a series of competitively active analogs is due to their presentation of common three dimensional structural motifs [9]. This approach focuses on data associated with ligands and the structural complementarity that should exist between them and the receptors to which they bind. It has previously been used to identify the homologous three dimensional atomic patterns present in a series of active analogs, extrapolate them to the identification of new active analogs, and infer

the three dimensional receptor features of the macromolecules to which they bind [10,11]. The details of the approach are in Chapter 2.

1.3 Viral Life Cycle

The HIV-1 virus is member of the lentivirus retroviral subclass. This type of virus maintains its genetic information as a single stranded (ss) (+) sense RNA genome rather than double stranded (ds) DNA, as in eukaryotic cells. Unlike other RNA viruses, the retroviral lifecycle includes an obligatory intermediate period during which its genome is maintained as a provirus, integrated dsDNA. This intermediate state is possible due to the process of reverse transcription, which involves the production of a dsDNA copy of the viral genome based upon a ssRNA template. This flow of information in the reverse direction of normal transcription (rna → dna as opposed to dna → rna → protein) is catalyzed by a unique class of DNA polymerases called reverse transcriptases, which can conduct both RNA and DNA dependent DNA polymerization.

The HIV-1 virion consist of an inner core region which contains two distinct ssRNA genomes as well as several enzymes required for early viral replication events (such as reverse transcriptase, integrase, and protease). The inner core is surrounded by an icosahedral shaped nucleocapsid shell and glycoprotein envelope (see Figure 1). A productive HIV-1 infection requires a high affinity binding reaction between the viral envelope glycoprotein, gp120, and a cellular CD4 surface protein. The CD4 protein is present in low concentrations on the surface of monocytes, macrophages, and antigen presenting dendritic cells [12]; and in high concentrations on the surface of immature T lymphocytes as well as mature T-helper lymphocytes. A structural and electronic complementarity between the CD4 receptor of a cell and a virus' gp120 protein is responsible for bringing a virus particle into close enough

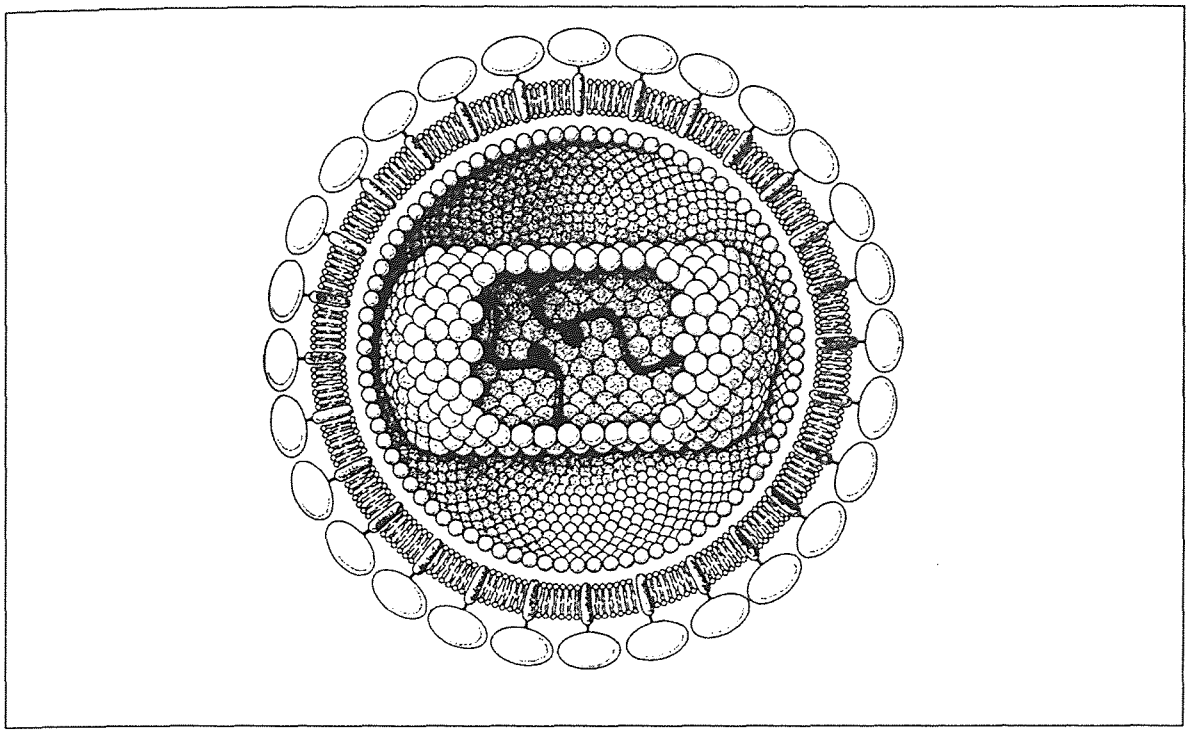


Figure 1: Schematic diagram of HIV virion [13].

proximity to a cell for a fusion reaction, mediated by the viral gp41 envelope glycoprotein, to facilitate viral entry (see Figure 2). In cells with an abundance of CD4 surface protein, highly efficient cell occurs due to the combined affects of CD4/gp120 binding and gp41 mediated fusion activity. In cells with low levels of CD4 protein, the primary dependence upon fusion activity results in less efficient viral entry. Following cell entry, a virus' glycoprotein envelope is degraded by cellular lysozymes and the nucleocapsid core moves toward the cell nucleus by a series of nuclear localization steps. During the period between cell entry and nuclear localization the process of reverse transcription occurs inside the viral core. This results in production of a dsDNA copy of the ssRNA genome. Upon reaching the nucleus, the core is degraded and a preintegration complex, which consist of the integrase protein complexed with the viral dsDNA, enters the nucleus. The integrase protein catalyzes the integration of viral dsDNA into the host cell genome, i.e. formation of provirus.

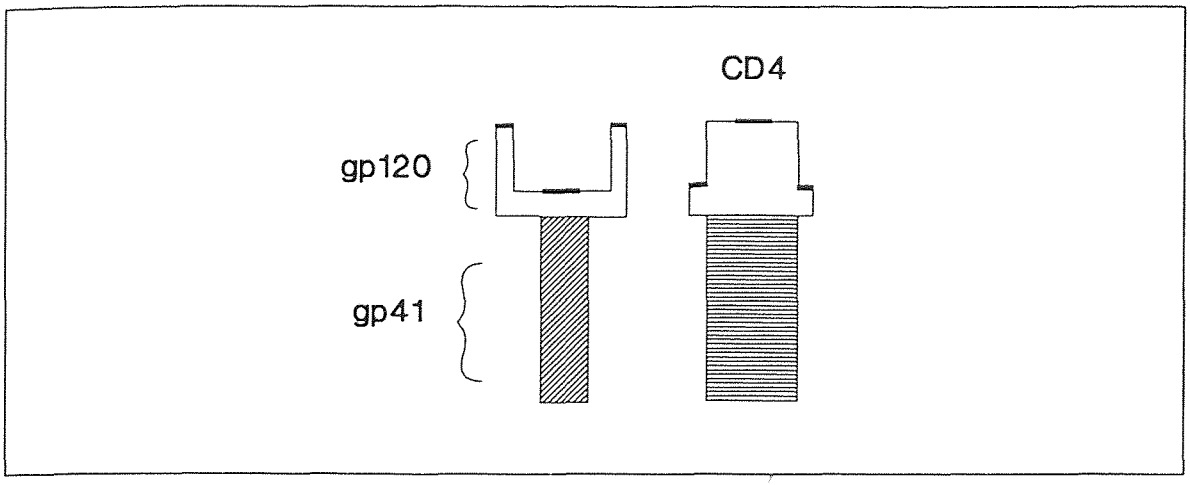


Figure 2: Schematic diagram ofCD4:gp120 complementarity and the relationship between gp120 and gp41.

Depending upon a number of host factors, the provirus will either enter a latent phase, and become dormant, or a lytic phase in which progeny virus is actively produced. Transcription of proviral DNA produces mRNA that can serve as either progeny viral genomes or templates for translation into a single polypeptide which is subsequently cleaved by the viral protease into individual viral proteins. Genomic mRNA associates with viral proteins at positions along the plasma membrane that have been altered to express viral envelope glycoproteins, such as gp120 and gp41, at the cell's surface. The budding of new virus occurs with the modified plasma membrane of the host cell serving as a glycoprotein envelope for progeny virus.

1.3.1 Cell Death

The budding process itself is not sufficient to cause cell death or diminished life span. It is the envelope glycoprotein mediated CD4/gp120 binding and fusion reactions that are toxic events. Two types of cell death are facilitated by these mechanisms. In cells presenting a high concentration of CD4 surface protein, fusion reactions occur between the gp120 of budding virions and the CD4 protein on a cell's surface. This results is a series of auto-fusion reactions along

the cell membrane which destroy its integrity and result in eventual death (see Figure 3).

The other type of cell death is due to the fusion a single infected cell, which presents viral envelope glycoproteins on its surface, with a number of normal cells presenting CD4 surface proteins. A single infected T cell can fuse with up to 500 uninfected CD4 presenting cells. This results in the formation of large multinucleated cell syncytia which have short life spans (see Figure 4) [12]. Evidence to substantiate these modes of cell death are observed as decreased viral cytopathic effects in virus mutants that have defective fusion , gp41, domains or in cells with low CD4 surface concentrations; and a prolific viral release from cells that have a low CD4 concentration without a loss of cell viability.

There are a variety of proteins unique to and required by the HIV-1 virus in order for it to establish a productive infection, such as the integrase, protease, RT, and a number of structural proteins.

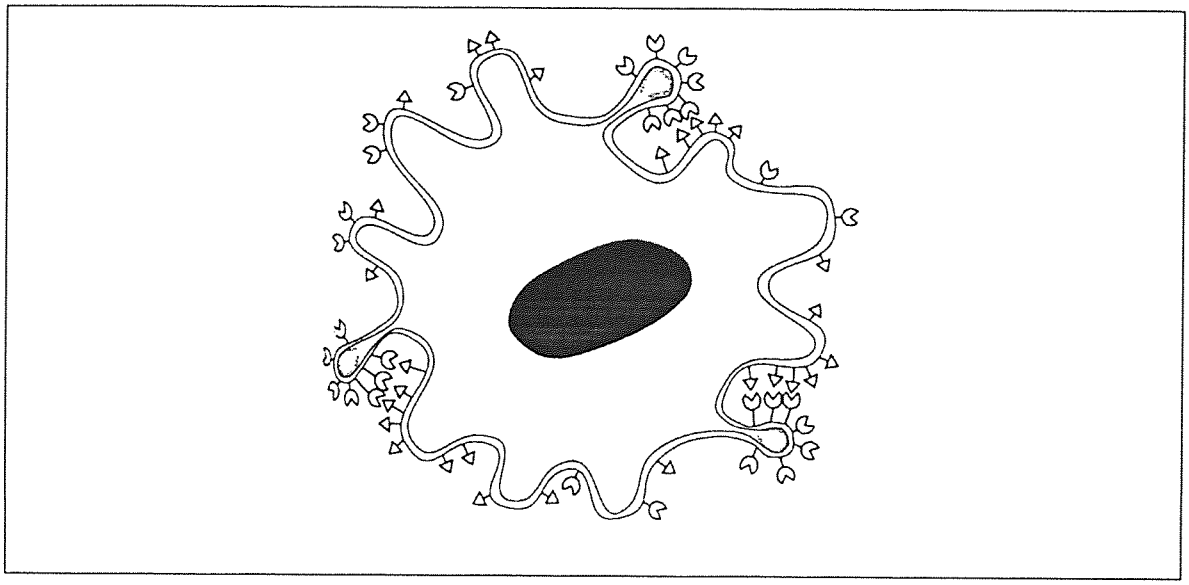


Figure 3: Auto-fusion reactions along cell membrane which result in destruction of membrane integrity and cell death [12].

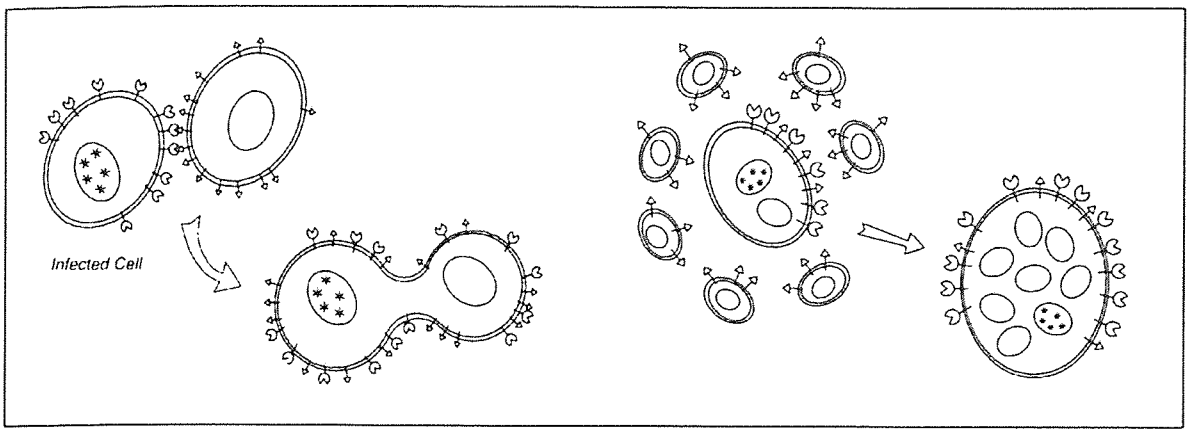


Figure 4: Aggregation of single infected cell with several uninfected cells to form multinucleated cell syncytia [12].

These proteins, for obvious reasons, are the chosen targets of antiviral therapy, since they are not required for normal cellular functioning. Currently, all of the proteins mentioned are under investigation as potential anti-AIDS agents. The RT enzyme, which is one of the most unique and critical enzymes used by HIV-1, has been the object of the most intense analysis.

1.4 Reverse Transcription

Reverse transcriptase is an RNA-dependent DNA polymerase class of enzyme that is crucial for the replication of retroviruses. It is responsible for the production of a dsDNA copy of a viruses (+)ssRNA genome (see Figure 5).

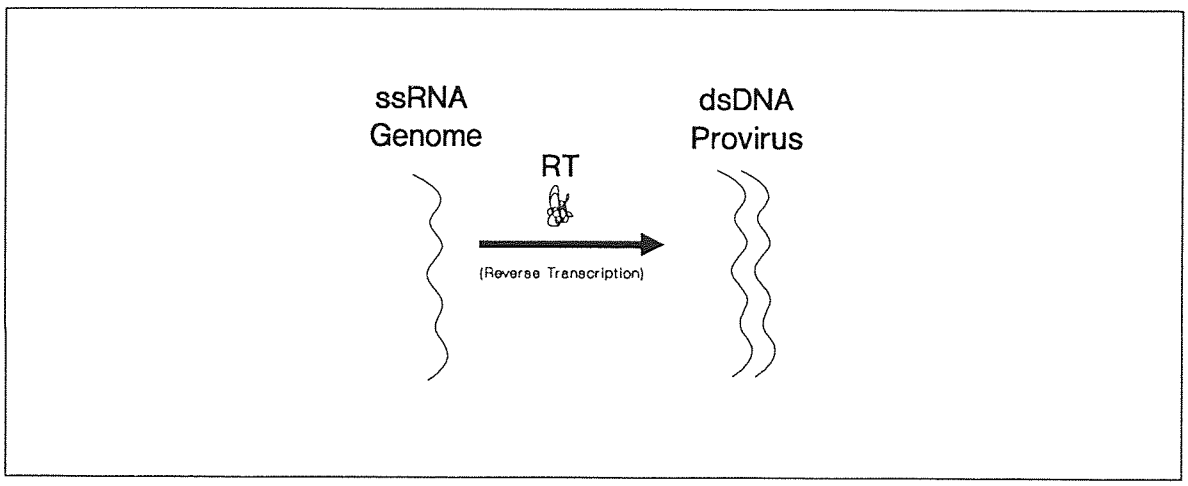


Figure 5: Input and output of reverse transcription.

It also catalyzes the production of long terminal repeat (LTR) sequences at both ends of the viral dsDNA, which is required for viral integration to the host genome and acts as a transcriptional promoter/enhancer region. RT like other DNA polymerases, synthesizes DNA in a 5' to 3' direction with the requirement of a primer strand for initiation. Under normal conditions a primer is provided by a primase enzyme which can add a short de novo RNA sequence to a single DNA strand. However, this enzyme is not available in the inner core region of a virus where reverse transcription takes place. A primer is provided by the association of host $tRNA^{lys}$, packaged in the virion during the preceding round of viral assembly, with a primer binding site (PBS) on the viral genome [13]. The PBS is a sequence of 18 nucleotides in the viral genome that is complementary to the 5' terminus of $tRNA^{lys}$. There are a number of intragenomic regions that serve as cofactors for reverse transcription and integration. In Figure 6 there is a schematic diagram of an RNA viral genome and some of the regions that are involved in reverse transcription.

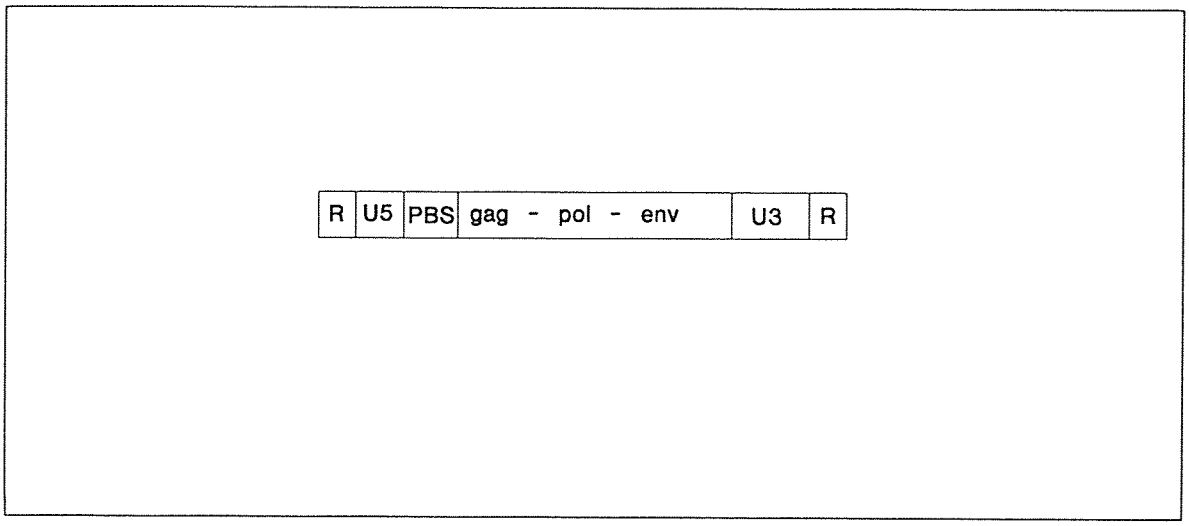


Figure 6: Schematic diagram of an RNA viral genome and the regions most relevant for reverse transcription.

Outline of reverse transcription (see Figure 7):

1. Host $tRNA^{lys}$ binds to the PBS region of the viral RNA genome and a (-) DNA strong stop copy of the 5'end of the viral genome is polymerized by RT. The region of the viral genome just copied is degraded by the ribonuclease H (RNase H).
2. The exposed 3' end of the strong stop (-)DNA base pairs with the other (+) ssRNA genome packaged in the virion and dissociates from the previous strand. The (-)DNA strand is extended to the PBS location on the second genome, which contains a stop codon.
3. The majority of the second RNA genome is degraded by RNase H, leaving small base paired RNA sequence in the vicinity of the 3' UTR. This piece of RNA serves as a primer for polymerization of a (+) DNA strong stop sequence and is, subsequently, degraded along with the $tRNA^{lys}$ by RNase H.
4. The PBS lip of the (+) DNA strong stop base pairs with the PBS region of the (-) DNA strand and makes an intergenomic jump to the 3' end of the (-) DNA strand. From this location, both strands are extended (5' to 3') to there full length and the resulting dsDNA contains LTRs at both ends.

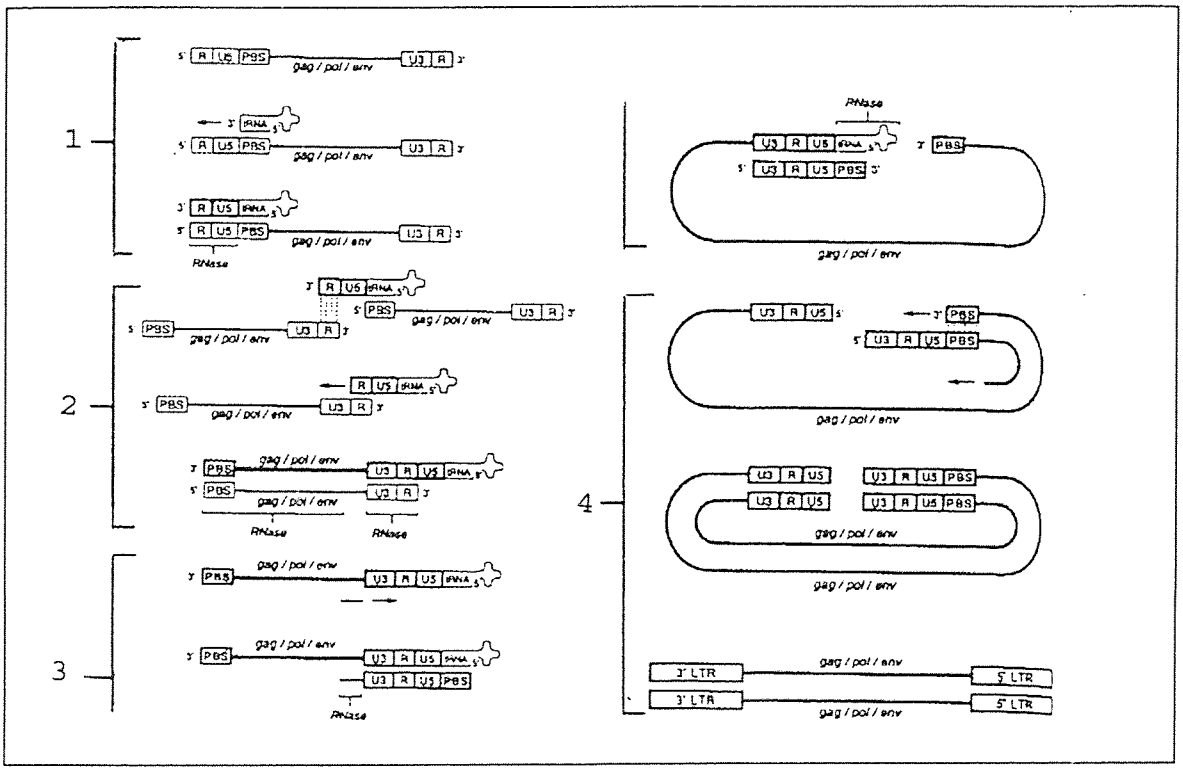


Figure 7: Graphical representation of reverse transcription events [12].

1.5 Reverse Transcriptase Structure

HIV-1 RT is a heterodimer protein that consist of a 66 kilodalton (kD) and 51 kiloDalton subunit. The 66kD subunit (p66) contains both the polymerase (pol) and RNase H domains. The heterodimer is a product of the proteolytic cleavage of a p66-p66 homodimer, which removes the RNase H domain from one p66 subunit, producing a 51 kD subunit (p51). The pol domain is present in both subunits but only catalytic in p66. A unique feature of the two subunits is that the tertiary structure of the pol domain in the p66 and p51 subunits are completely different, despite their identical primary and secondary features. The most visible difference between the two subunits is the presence of a large cleft in p66, which is large enough to accommodate an A- or B- form RNA-DNA hybrid between the RNase and pol domains, analogous to that found in the klenow fragment of E.coli DNA pol I [14].

The p66 subunit consist of five subdomains, four of which are in the pol domain and one in the RNase H domain. An anatomical resemblance between the four pol subdomains of p66 and a cupped hand led to their naming as fingers, palm, thumb, and connection regions by Steitz et al. with the connection region located between the pol and RNase H domains (see Figure 8). In Figure 9 there is a picture of a recently solved X-ray crystallographic structure of RT complexed with duplex DNA which was generously provided by Edward Arnold [15].

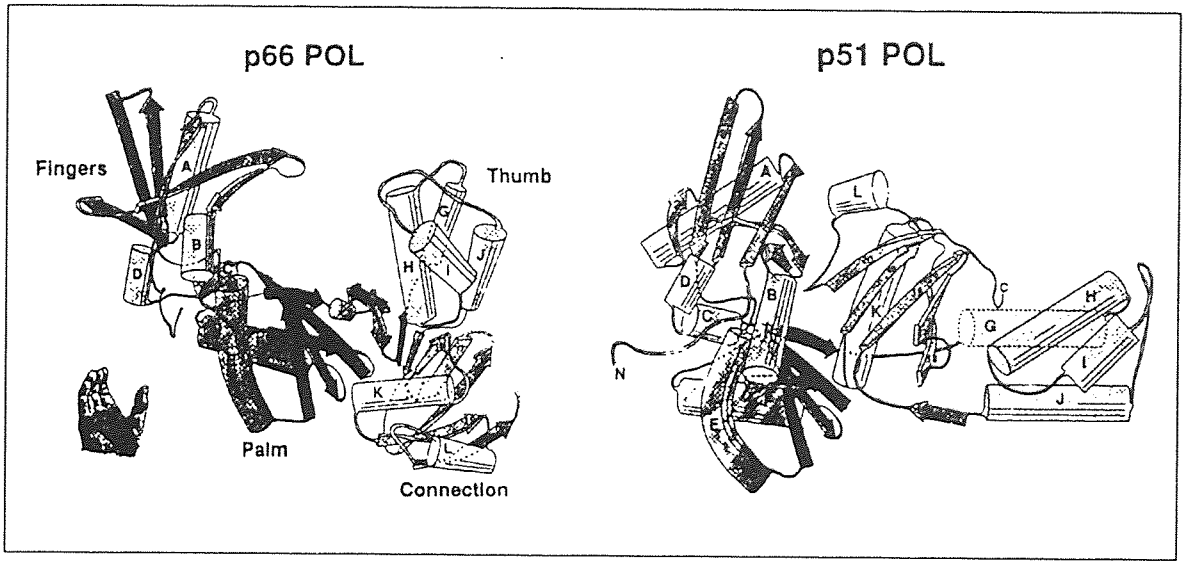
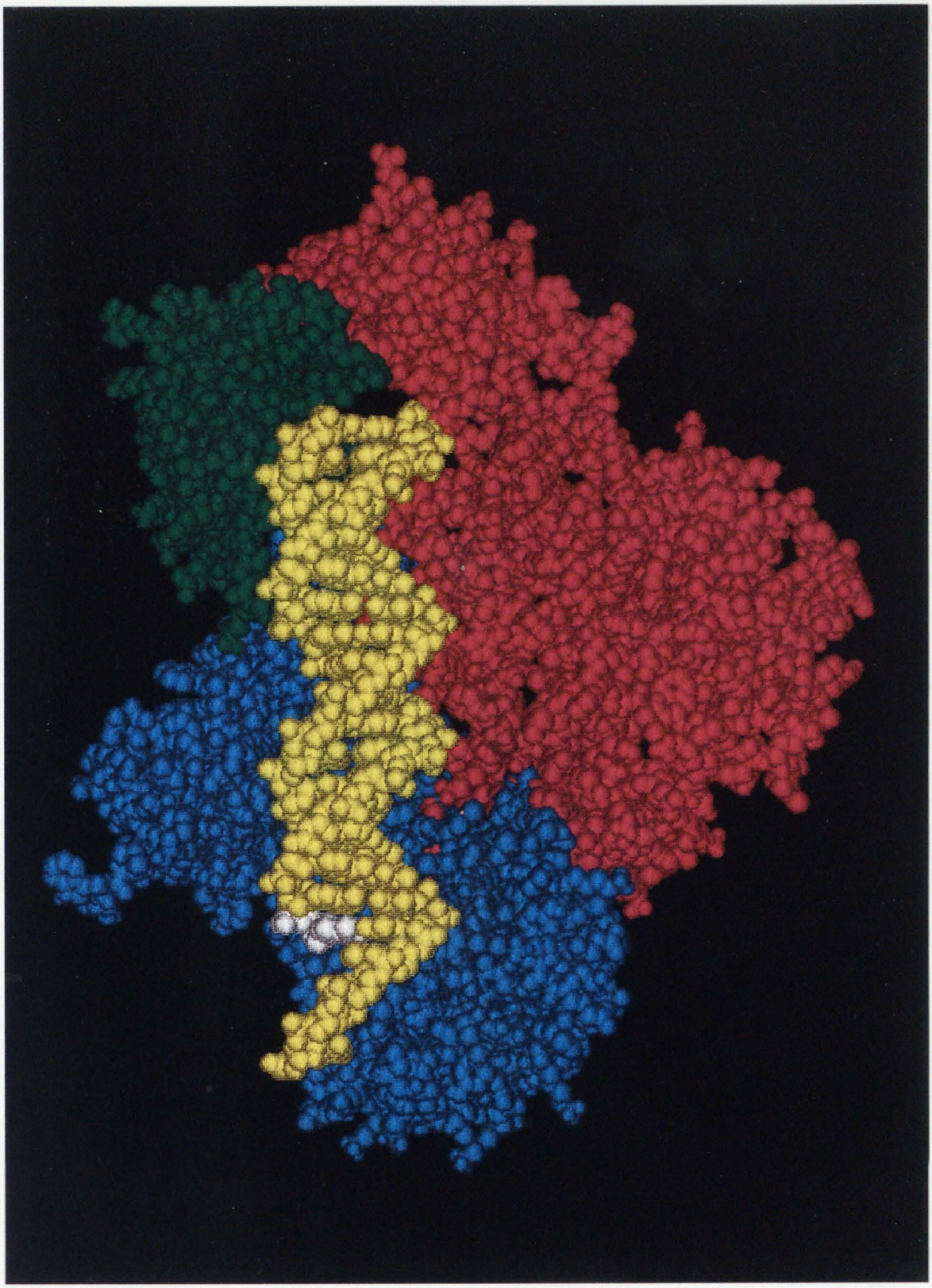


Figure 8: A cartoon structure of the p66 and p51 subunits of HIV-1 RT with anatomical subdomains. Notice the predominant DNA binding cleft in the p66 subunit which is absent in p51 [14].

Figure 9: Space filling rendition of recently solved X-ray crystal structure of HIV-1 RT [15]. The colored portions are p66 (cyan), p51 (magenta), dsDNA (yellow), and thymidine sustrate in vicinity of receptor site (white).



CHAPTER 2
METHODOLOGY

2.1 Active Analog Approach to Receptor Mapping

The active analog approach to receptor mapping is dependent upon the existence of a common pharmacophore for a series of compounds demonstrating competitive biological activity at the same protein. A pharmacophore is defined as the atoms or groups whose three dimensional spatial orientation is required for recognition by a given protein. It includes both the atoms and their three dimensional positions with respect to one and other.

The active analog approach provides an indirect way of describing a proteins receptor site based upon structural information associated with ligand molecules. In the absence of well resolved crystallographic data related to a protein of interest (as was the case with RT until recently[15]) this method provides a means of inferring a protein's three dimensional receptor features. The two premises upon which this approach is based are: 1) a common pharmacophore exists for a series of compounds which demonstrate competitive activity at the same receptor site and 2) the generation of all the sterically allowed conformations a molecule can assume will produce at least one that is bioactive, i.e. which presents a pharmacophore for which a receptor of appropriate size and shape exists in the protein of interest. From this perspective, the task of pharmacophore identification is that of generating the sterically allowed conformations for a series of active compounds and then analyzing them for common atomic spatial arrangements.

2.1.1 Receptor Mapping

The active analog approach presumes that any compound demonstrating the slightest activity at a given protein is capable of presenting the pharmacophore, which accounts for its recognition by the enzyme. From a structural perspective, highly active compounds are seen as presenting volumes complementary to the receptor cavity (see Figure 10A). While, poor activity or inactivity is seen as due to the presentation of accessory volumes, which occupy areas that conflict with protein volume at the receptor site, or an inability to present the required pharmacophoric pattern (see figure 10B).

The union of the active analog volumes superimposed along the pharamcophore is referred to as an enzyme excluded volume. It represents the minimum volume availability at the receptor site (see Figure 11A). The union volume of all inactive analogs superimposed along the pharmacophore contains regions that hypothetically conflict with protein volume at the receptor, a receptor essential volume (see Figure 11B). The intersection of the enzyme excluded and receptor essential volumes represents the volume associated with the inactive compounds that is in common with the enzyme excluded volume and, therefore, within the boundaries of the receptor site (see Figure 11C).

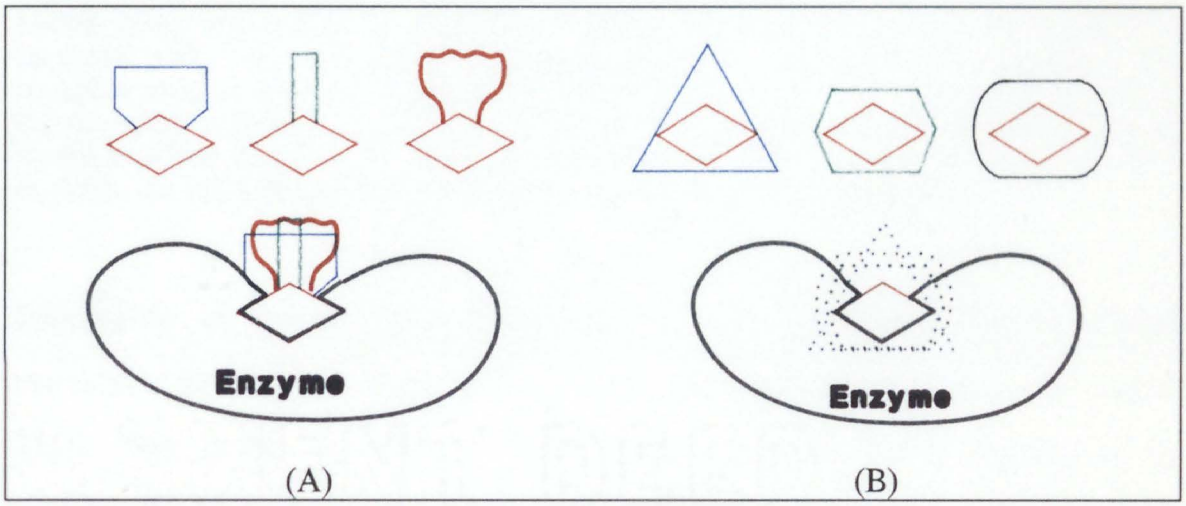


Figure 10: Example of active (A) and inactive (B) analogs at the same receptor site.

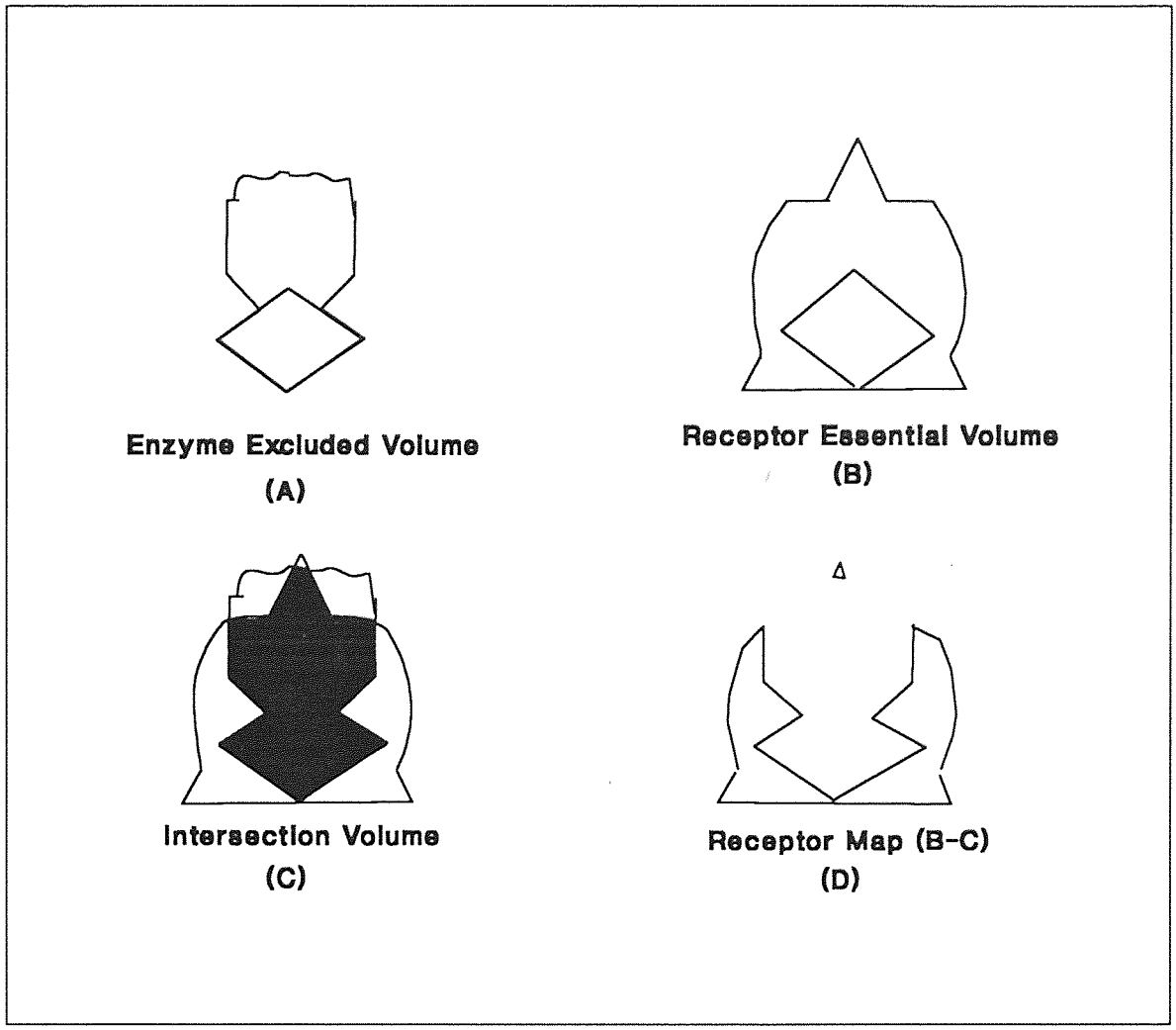
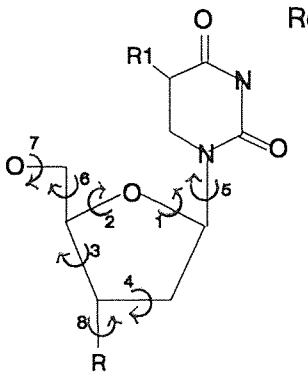


Figure 11: Receptor map example. Enzyme excluded volume is the net volume occupied by active analogs and, consequently, must be available at the receptor site. Receptor essential volume is the net volume occupied by inactive analogs which contains regions in conflict with enzyme volume at the receptor site. Intersection volume is the portion of the receptor essential volume within the receptor site, i.e.. not conflicting with enzyme volume at the receptor. The receptor map is the cavity produced by subtracting the intersection volume from the receptor essential volume. It represents the specific regions of volume excess in the inactive compounds which conflicts with protein volume at the receptor. It is, thus, an indirect map of the critical regions in the receptor site.

Subtraction of the intersection volume from the receptor essential volume provides a three dimensional cavity which constitutes a receptor map (see Figure 11D). This is the net volume of the inactive analogs that is outside of the receptor cavity and, consequently, which conflicts with protein volume at the receptor site.

2.1.2 Systematic Conformational Search

The systematic conformational search [22] is a tool for the investigation of the conformational space available to a compound via the rotation of flexible bonds (see Figure 12). It allows the definition of flexible bonds and their associated torsion angles, which when varied, result in the generation of the sterically allowed conformations that a compound can assume. The search algorithm checks the van der Waals contacts amongst non-bonded atoms, while scanning all possible combinations of torsional angle values. Only those conformations without van der Waals overlap between non-bonded atoms are acceptable. The search algorithm also provides a means of conformational discrimination based upon the logical intersection, union, or difference between search data (i.e. interatomic distances, torsional angles, conformational energy, etc.) from a number of molecules. The combinatorial attributes of the systematic conformation search allows the identification of the potentially pharmacophoric patterns in a series of compounds. The pharmacophore provides a common frame of reference for the orientation and superposition of active and inactive analogs at the receptor site.



Rotatable Bond Increments

Bond	Angle(°)
1	2
2	2
3	2
4	2
5	5
6	15
7	45
8	45

Figure 12: The search variables for the compounds under consideration.

2.1.3 Method of Comparison

The most desirable method of comparison of potential pharmacophoric patterns will allow the extrapolation of the pharmacophore to compounds from different molecular classes, which may demonstrate common biological activities. Interatomic distances were used as the metric for analysis and comparison of the conformational data associated with each analog in this study [9]. This method of comparison focuses upon the **orientation** of individual atoms with respect to each other, independent of the overall molecular framework of a compound. It allows the identification of pharmacophoric patterns that can be applied to structurally diverse molecules, i.e. outside the nucleoside family, which may demonstrate similar activity. Within this *orientation space*, a pharmacophore consisting of a narrow range of interatomic distances between specific atoms can be described in terms of an orientation map (OMAP) point, which has a dimensionality that is equivalent to the quantity of interatomic distances it contains. This reduces the comparison of three dimensional molecular features to the comparison of distances between atom pairs, which can be readily applied to structurally diverse compounds. The intersection of the OMAP points from a set of active analog's provides the common and potentially pharmacophoric atomic pattern(s) that are available (see Figure 13).

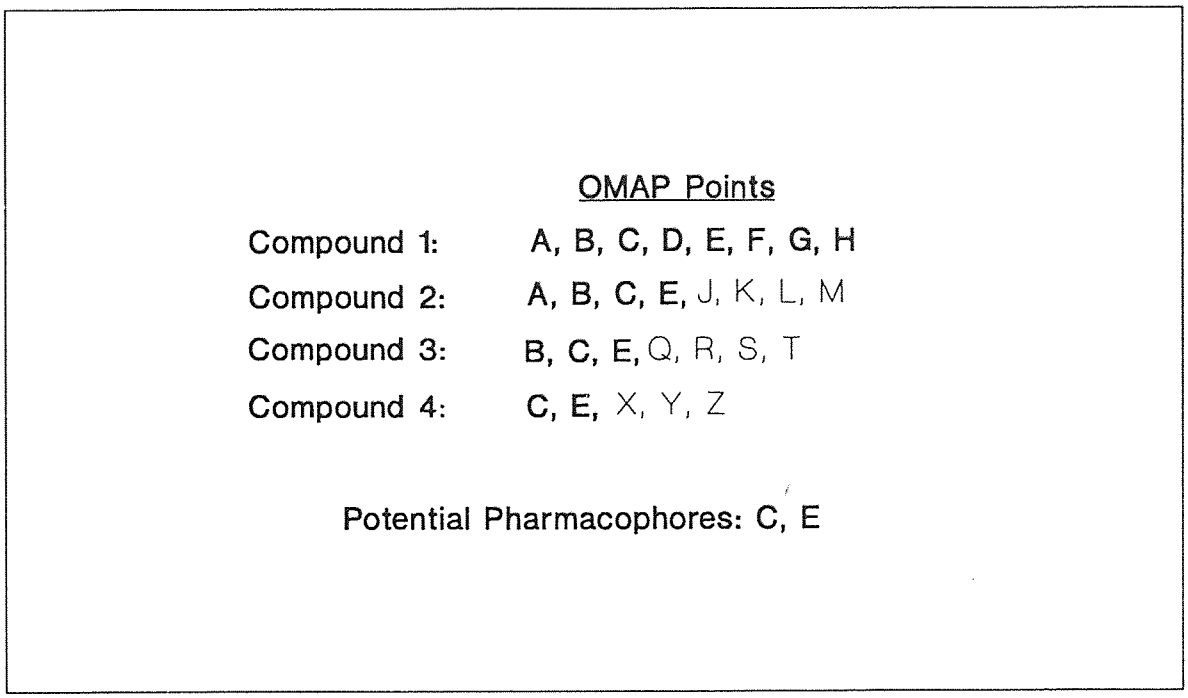


Figure 13: OMAP example. Compound 1 was searched without constraints and generated an OMAP point for every conformation (A-H). Compound 2 searched without constraints will generate all of the given OMAP points (A,B,C,E,J,K,L, and M), but with compound 1's OMAP as a constraint only those common to both (in bold). The searching of subsequent compounds with constraints results in the identification of the OMAP points common to all in compound 4 (C and E, in bold).

For example, an OMAP point is generated for each sterically allowed conformation of a compound and contains interatomic distance data related to predefined atom pairs. At the end of a systematic conformational search the resulting OMAP of each compound contains a point for each conformation and, thus, defines the entire set of sterically allowed spatial arrangements for the designated atoms with respect to each other. Since by definition a pharamcophore is a **common** spatial orientation of atoms in a series of active analogs, at least one point in each compound's OMAP should be common to all the analogs under consideration. The identification of the OMAP point(s) in each compound that are common to the series provides a set of potentially pharmacophoric patterns.

The usefulness of this approach is in the fact that although a pair of atoms may move throughout a common overall distance range in two compounds (minimum: 6.0 Å; maximum: 10.0 Å) the movement is discontinuous in each; and, the identification of common points inside this global range provides a means of conformational discrimination, by eliminating conformations without overlap in the segments that comprise an overall range (see Figure 14). An intersection is identified using a search and constrain approach, in which OMAP points are generated for a compound and used to constrain the conformations generated for a subsequent compound to those common to both. Using this approach, the last compound searched in a series will contain OMAP points common to itself and all previously searched compounds, i.e. an intersection.

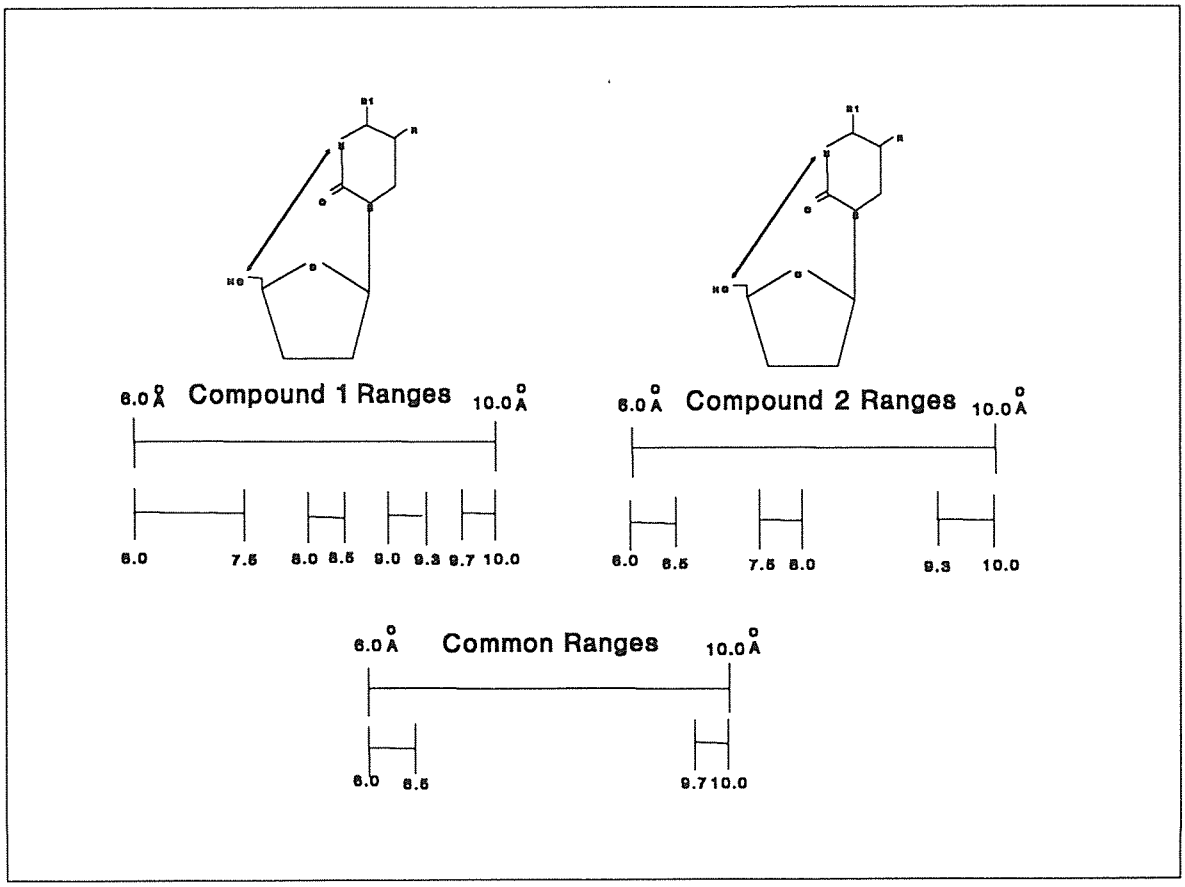


Figure 14: Example of distance range intersections as a means of identifying commonalities amongst compounds. The overall ranges for both compounds are identical, but when examined at a higher resolution one can see that there is room for discrimination between the two.

2.1.4 Search Resolution

The principal determinant of the quality of a systematic search is the resolution of interatomic distance measurements. This is analogous to the maximum amount of potential deviation that can occur between the pharmacophoric groups of two different analogs. The Grid Size parameter of the systematic conformational search defines the degree to which interatomic distances from different compounds must agree in order to be considered common. In simplest terms, it defines the increments with which interatomic distance measurement will be taken. Consequently, the Grid Size parameter quantitatively regulates the amount of pharmacophoric overlap between the selected atoms in different molecules.

For example, setting a grid size of 0.1 Å results in the measurement of interatomic distances at increments of 0.1 (i.e. 0.1, 0.2, 0.3, 0.4 Å, etc). Distances falling between grid points, such as 0.1 to 0.2 Å or 0.2 to 0.3 Å, are in a relative dead space and will map to a the same midpoint distance, i.e. 0.15 Å and 0.25 Å, respectively. The grid size thus provides a means of defining the level of accuracy with which interatomic distances will be measured. A grid size of 0.1 Å will map interatomic distances of 0.120, 0.144, and 0.151 Å from a molecule (A) and 0.149, and 0.168 Å from a molecule (B) to a common midpoint of 0.15 Å, i.e. all would be considered identical. A more accurate measurement would be made with a Grid Size of 0.01 Å, which would only map 0.144 Å from molecule (A) and 0.149 Å from molecule (B) to the same identical, at 0.145 Å (see Figure 15). The grid size parameter provides a means of defining the resolution with which the comparison of potential pharmacophoric patterns will be made.

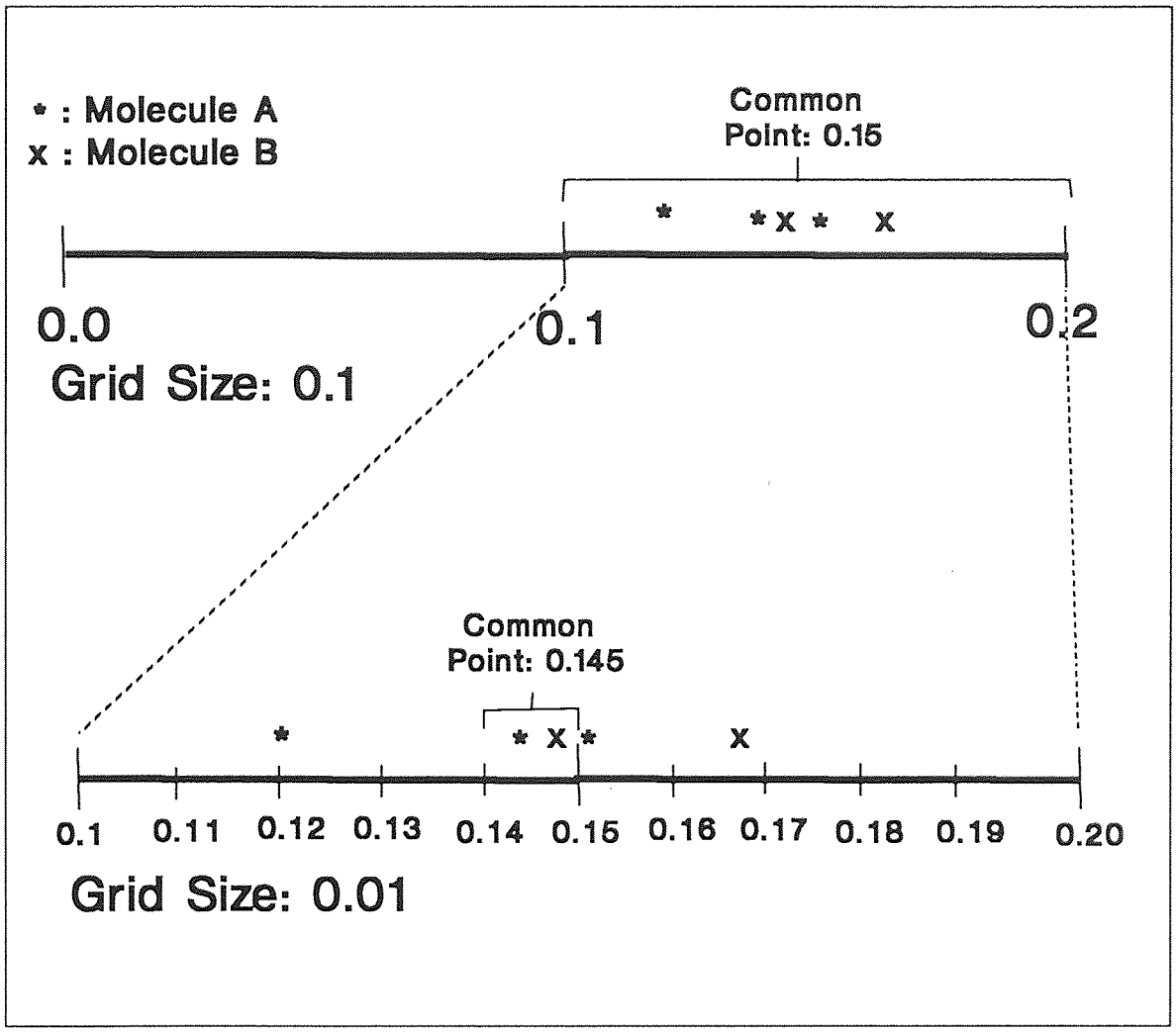


Figure 15: Search resolution example. Comparison of 0.1 and 0.01 grid sizes. The larger the grid size the greater the potential difference between interatomic distances which map to the same OMAP point. A grid size of 0.1 will consider interatomic distances within up to 0.1 Å of each other as identical (see top grid, all distances are will be mapped to 0.15 Å); whereas a grid size of 0.01 will provide a more stringent range of 0.01 Å for consideration as identical (see lower grid, only two distances within close proximity of each other are considered identical).

CHAPTER 3

EXPERIMENTAL RESULTS AND DISCUSSION

3.1 Selection of Thymidine Nucleoside Analog Inhibitors

The compounds given in Table 1A were chosen based upon their selectivity index (S.I.), which is the ratio of their cytotoxic concentration (CD_{50}) to their inhibitory (effective) concentration (EC_{50}). Well designed anti-AIDS drugs will maximize the numerator (CD_{50}) and minimize the denominator (EC_{50}). The S.I. provides a numerical representation of the preference a compound has for inhibition of viral enzymes over cellular enzymes. Although a S.I. is not a direct measure of the affinity of a compound for RT, it provides an measure of a its pharmacological effectiveness. A constant that is more indicative of the relative affinity of an inhibitor for a particular protein is K_I.

$$K_I = [\text{Enzyme}][\text{Inhibitor}]/ [\text{Enzyme-Inhibitor Complex}] \qquad (3.1)$$

From the equation it can be seen that the magnitude of K_I is representative of the amount of enzyme and inhibitor that are in a complex form. K_I is a constant that has a propensity to vary greatly from one study to the next depending upon the reaction conditions. For this reason, the results of studies involving the calculation of K_I are generally relative to the local conditions.

The relative values of K_I for five of the compounds under consideration were available from a single study [4]. There is apparently some correlation between these and the respective EC_{50} and S.I. values of the given compounds. In Table 1B the K_I values for these compounds are arranged in increasing order, i.e. decreasing affinity for RT. A relationship between K_I, EC_{50} and S.I. is most apparent when K_I is compared to a modified S.I. which involves the ratio

of S.I. to EC_{50}. This is analogous to a selectivity index that places a greater emphasis upon a compounds anti-viral activity (Modified S.I. = CD_{50} / $(ED_{50})^2$).

3.2 Software and Hardware

Molecular modeling studies were conducted on a Silicon Graphics Personal IRIS workstation interfaced with a main frame HP-UX 9000 computer. The SYBYL version 6.0 molecular modeling program from Tripos Associates [22], Inc. was

Table 1: Thymidine nucleoside active(+) and inactive(-) analogs with relevant concentrations in PMB, MT-4 (˜), and ATH8 (˜˜) cells [2,16].

	Compound	ED_{50}	CD_{50}	S.I.
* 1.	3'-Azido-2',3'-dideoxythymidine (AZT)	0.004	200	50,000 (+)
* 2.	3'-Azido-2',3'-dideoxyuridine (AzddU)	0.2	1000	5,000 (+)
3.	5-Chloro-3'-Fluoro-2',3'-dideoxythymidine (5Cl3FddU)	0.38	535	1,408 (+)
* 4.	5-ethyl-3'-Azido-2',3'-dideoxythymidine (5EtAzddU)	1	1000	1,000 (+)
* 5.	2',3'-dideoxythymidine (ddT)	0.2	>125	>625 (+)
6.	3'-Fluoro-2',3'-dideoxyuridine (3FddU)	0.04	16	400 (+)
7.	5-Chloro-3'-Azido-2',3'-dideoxythymidine (5ClAzddU)	0.72	213	296 (+)
8.	3'-Fluoro-2',3'-dideoxythymidine (3FddT)	0.001	0.197	197 (+)
9.	5-Bromo-3'-Azido-2',3'-dideoxythymidine (5BrAzddU)	1	>100	>100 (+)
10.	5-Iodo-3'-Azido-2',3'-dideoxyurirdine (5IAzddU)	1.1	>100	>91 (+)
* 11.	2'-Chloro-2',3'-dideoxyuridine (2CldduU)	81	500	6 (-)
* 12.	2',3'-dideoxyuridine (ddU)	210	625	3(-)
13.	5-ethyl-3'-Fluoro-2',3'-dideoxyuridine (5Et-3FddU)	330	625	2 (-)
14.	5-ethyl-3'-Iodo-2',3'-dideoxyuridine (5Et3IddU)	86	100	1.16 (-)

*: Crystal Structures; S.I.: Selectivity Index; ED_{50}: 50 % Inhibitory Concentration (μM); CD_{50}: 50 % Toxic Concentration (μM)

Table 1A: Available K_I values and modified S.I.

Compound	K_I	ED_{50}	S.I.	Modified S.I.
AZT	0.006	0.004	50,000	12,500,000
3FddT	0.007	0.001	197	197,000
AzddU	0.016	0.2	5,000	25,000
AzddEtU	0.12	1	1,000	1,000
3FddEtU	2.3	330	2	0.006

used for molecular visualization and manipulation. The MAXIMIN2 molecular mechanics program and MOPAC Semi-empirical Quantum Chemical program [23] from the Quantum Chemical Program Exchange (QCPE # 455) at Indiana University were the software packages used for geometry optimization of the structures under consideration. Crystallographic data was obtained from the Cambridge Crystallographic Database and generously provided by Patrick Van Roey (Medical Foundation of Buffalo Inc.).

3.3 Energy Minimization

The success of any computer aided molecular modeling study is dependent upon the accuracy of the starting geometries and method of calculating the relative energies of the compounds under consideration. The most reliable starting structure is one obtained from X-ray crystallographic data, which is analogous to a snapshot of the given compound at some instant of time. In the event that such data is not available, then software based construction of the compound is required. Regardless of the source of the starting geometry, all the compounds under consideration had their geometries optimized via a software based energy minimizer at the beginning of the study. The optimization of the manually generated structures eliminated bad electrostatic and/or van der Waals contacts that were present in the structure.

X-ray crystallographic geometries are products of atomic interactions within a matrix of similar molecules, which is usually not the medium of interest and can produce structural characteristics unique to that environment. There is also a of great deal of latitude with regard to the level of resolution that can be achieved for a given molecular system. The X-ray crystal structure can be used to evaluate the effectiveness of a molecular geometry optimizer. Significant alteration of an X-ray crystallographic structure following optimization is

generally indicative of a software geometry optimizer that is inadequate for the given system. There are essentially two approaches to geometry optimization, molecular mechanics and quantum mechanics, each of which has its own merits and weaknesses.

3.3.1 Molecular Mechanics Method

The molecular mechanics method of geometry optimization is based upon the experimentally based parameterization. It views a molecule as a collection of points (atoms), connected by springs (bonds), with variable elasticities (force constants). The process of parameterization involves the correlation of experimentally observed data to potential energy functions which are stored in molecular mechanics force fields. The value of a particular force field lies in the variety and quality of its parameterization. A force field consist of a variety of atom types for a given element. Each type of atom is designed to emulate specific structural motifs (i.e. Carbon atom types: C(aromatic), $C(sp^3)$, $C(sp^2)$, etc) based upon its experimentally derived potential energy functions.

With regard to geometry optimization, the molecular mechanics method assumes a set of optimal bond lengths and angles based upon its parameters sets. The energy of a molecule is defined according to the amount of deviation from these "ideal" structural motifs. The summation of the effects of these deviations over all bonds, valance angles, torsional angles, and non-bonded atomic interactions represents the total molecular energy. A Descent Series Iterative Method, which modifies the atomic coordinates of a structure from one iteration to the next in the direction of decreasing energy, is the means by which geometry optimization is achieved. This is generally a local energy minimum, which is dependent upon the starting geometry of the compound. The major problem with the molecular mechanics methods is that many of the force fields

have been developed for use in specific modeling areas and, therefore, may not be suited for a molecular system of interest.

The Tripos and AMBER force fields were available for use with the MAXIMIN2 molecular mechanics method. The effects of geometry optimization on each of the compounds under consideration using these two force fields were analyzed. The most disturbing problem presented by this method was the lack of appropriate parameters for nitrogen atoms in the azido configuration. This is a linear structure that can be represented as the average of two resonance forms (R-N=N=N and R-N-N≡N)[17]. In both force fields the only nitrogen atom parameters available that allowed interactions with other nitrogens were sp^3 hybridized and, consequently, resulted in tetrahedral azido geometries. The nitrogen atom parameter which was sp hybridized (linear) lacked the parameters necessary for interaction with other nitrogens. The only way to circumvent this lack of azido parameters was to define the azido region of a molecule as an aggregate. This abstains it from geometry optimization calculations and, thus, retains the geometry of the crystal structure ($\approx 171°$). Since the goal of this study was to realize the effects that substitutions on the sugar and base moieties have upon the overall nucleoside geometry, this method was not used.

3.3.2 Quantum Mechanics

The quantum mechanics approach differs from molecular mechanics in its method of energy calculation. The energy in quantum mechanics calculations, which is used to guide the geometry optimization of a compound, is derived from approximate solutions to the Schröedinger equation:

$$H\Psi = E\Psi \qquad (3.2)$$

Although, it can only be solved directly for a two particle system (hydrogen atom: 1 proton + 1 electron), fairly accurate approximations to its solution for

higher order systems have been achieved[18]. There are two approaches to this goal, ab initio and semi-empirical.

The Hartree-Frock equations are based upon the ab initio approach and approximates solutions for higher order systems by solving the Schroedinger equation directly for 1-electron functions and extrapolating these solutions to multi-electron systems (atoms molecules). The solution for molecules is approximated as the linear combination of the its' atomic orbitals, which are functions of the 1-electron systems they contain.

The semi-empirical approach does not attempt to calculate solutions for all the 1-electron functions of a molecule. Instead it considers the inner-core electrons as a static region and focuses upon explicit solutions for valence electrons only. It also ignores or replaces many of the integrals describing electron interactions with experimentally determined parameters. Thus, it can be viewed as almost a hybrid between quantum and molecular mechanics.

The advantage of the ab initio quantum mechanics method lies in the fact that it makes no assignment of parameters based upon experimental data and, therefore, can provide qualitative results related to molecular systems of contemporary interest, which may not correlate to previously determined experimental results. Its only drawback is in the requirement of a large amount of computation to arrive at a solution. The computational demands, both volume and time, have limited its application to systems of less than 20 atoms, unless a supercomputing platform is available. The advantages and disadvantages of the semi-empirical method are complementary to those of ab initio. In the parameterization of some of its computational task it has reduced its accuracy compared to ab initio, but can accomplish its solutions with increased speed and decreased computational requirements. The speed and accuracy of the semi-empirical method when compared to ab initio and molecular mechanics led to the

selection of the Mopac program as the method of geometry optimization for this study.

Mopac is a software package that contains four different semi-empirical methods (PM3, MNDO, and AM1). The most contemporary method is PM3 and the oldest is AM1. Each method was essentially designed to overcome deficiencies in prior methods. I analyzed each method and found that in all three the most striking inadequacies were an underestimation of sugar puckering and production of a non-planar geometry in the base ring. In compounds with definite 2' or 3' exo sugar conformations in crystal structures there was a tendency to reduce the degree of exo puckering. And, the torsion angles in the base ring of all compounds tended to deviate from values associated with planarity (0° and 180°). To remedy the base planarity problem I utilized the Mopac SYMMETRY option, which allows restriction of the range of variation in a dihedral angle, valence angle, or bond length to that of another respective angle or length in a molecule. Of the three methods available, AM1 produced the least amount of reduced sugar puckering and distorted base planarity; providing the optimized geometries that most closely resembled the crystal structures. Thus, the geometry of each compound under consideration was optimized using the AM1 method with SYMMETRY constraints.

3.4 Analysis of Sugar Puckering

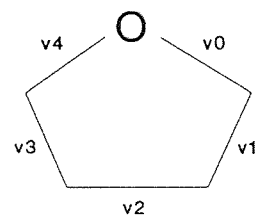
In a preliminary study, it was found that limiting the sugar ring to the crystal geometry resulted in very little structural homology amongst a set of only 5 thymidine nucleoside analogs [19]. This was a discouraging observation since the most important attribute of a pharmacophore is its applicability to a large number of active analogs. If at 5 compounds the amount of structural homology was

slight, the prospects for extension of the given homology to a larger number of compounds is not very good.

The fact that the geometry of a sugar ring is flexible and, therefore, could be possibly be induced to undergo conformational changes upon complexing with RT, prompted its inclusion as a conformational variable in this study (see Figure 12, bonds 1 to 4). To identify what values of angle rotation would be most appropriate to this region of a nucleoside, the types of sugar puckering that were produced by a number of different rotatable bond increments were analyzed. Sugar puckering is defined as the deviation of an atom(s) from a plane defined by three of atoms in the ring. The degree and type of puckering was analyzed by the calculation of a psuedorotation phase angle (P). This is a single angle, between 0 and 360°, which represents the cumulative effects that all of the torsion angles in a five membered ring have upon puckering [20]. It gives a qualitative idea of the net effect of changes in individual ring torsion angles upon the overall ring conformation. This variable is important because the type of puckering in the sugar ring has been suggested as a potential activity determinant in these compounds [1]. A mathematical model was developed using the Mathcad software package which accepts the individual torsion bond angles of a large number of conformations and computes P based upon the Saenger equation (see Appendix A)[20,21]:

Saenger Equation **(3.3)**:

$$P = \tan^{-1} \frac{[(v_4+v_1) - (v_3+v_0)]}{[2*v_2*(\sin 36° + \sin 72°)]}$$



3.5 Method of Comparison: Interatomic Distances

The interatomic distances (atom pairs) in Figure 16 were selected because they will most likely contribute to a pharmacophore for thymidine nucleoside analogs at the RT receptor site. These interatomic distances were selected because they involve heteroatom pairs, which may contribute to hydrogen bonding and electrostatic interactions at the receptor site, and atoms that are common to all the analogs, which allows correlation to the entire series of compounds. The selected interatomic distances were not all inclusive because those without variability between conformations (such as the distance between O4 and O2 of the base) do not provide conformational discrimination. The distance between the base center and O5' (interatomic distance H) was included because the position of the base ring will effect potential hydrophobic interactions at the receptor.

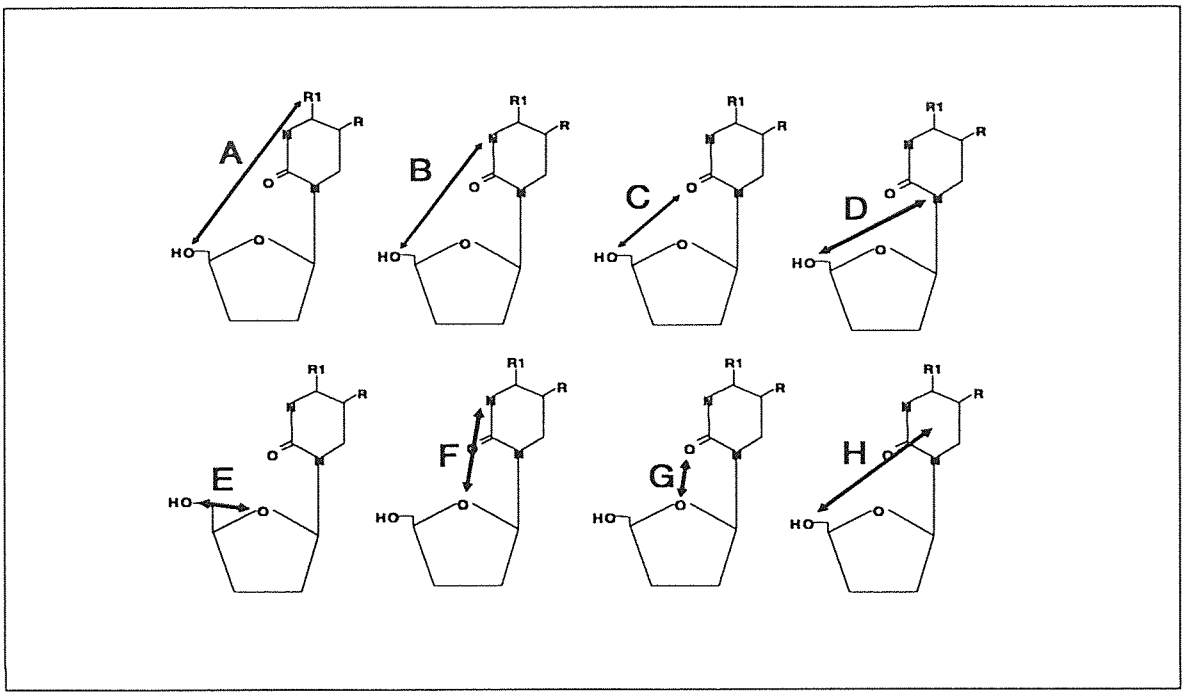


Figure 16: Interatomic distances (atom pairs) used as constraints in the systematic conformational search.

3.6 Systematic Conformational Search Results

The systematic conformational search was used to vary each compound along the flexible bonds defined in Figure 12 and compare the interatomic distances between the heteroatom pairs in Figure 15. All possible combinations of 1, 2, 3, 4, 5, 6, 7, and 8 interatomic distances were analyzed for simultaneous common orientations in all the compounds under consideration (see Appendix B). A set of seven interatomic distances (A,B,C,D,E,G, and H) with a single OMAP point as their intersection was identified and as the highest amount of structural homology and proposed as a potential pharmacophore (see Table 2, column A). The OMAP point identified here was used to constrain the mobility of all the compounds under consideration and identify the conformations in each that were capable of presenting the proposed pharmacophore (see Table 2, column B). The pharmacophore was applied to natural substrate (dT) and a completely inactive substrate (uridine) for further validation. The fact that dT (natural substrate) is capable of presenting the pharmacophore and uridine is not, supports the proposed pattern.

Table 2: Systematic Conformational Search OMAP Points

Compound	(A) OMAP Points	(B) Conformations
3'-Fluoro-2',3'-dideoxythymidine (3FddT)	12,730	152
2',3'-dideoxythymidine (ddT)	2,299	42
5-ethyl-3'-Azido-2',3'-dideoxythymidine (5EtAzddU)	99	352
3'-Azido-2',3'-dideoxyuridine (AzddU)	82	448
3'-Azido-2',3'-dideoxythymidine (AZT)	37	84
5-Chloro-3'-Fluoro-2',3'-dideoxythymidine (5Cl3FddU)	31	56
3'-Fluoro-2',3'-dideoxyuridine (3FddU)	19	56
5-Bromo-3'-Azido-2',3'-dideoxythymidine (5BrAzddU)	19	224
5-Chloro-3'-Azido-2',3'-dideoxythymidine (5ClAzddU)	3	28
5-Iodo-3'-Azido-2',3'-dideoxyurirdine (5IAzddU)	3	112
5-ethyl-3'-Fluoro-2',3'-dideoxyuridine (5Et-3FddU)	2	16
5-ethyl-3'-Iodo-2',3'-dideoxyuridine (5Et3IddU)	2	59
2',3'-dideoxyuridine (ddU)	1	30
2'-Chloro-2',3'-dideoxyuridine (2CldduU)	1	30
2'-deoxythymidine (dT: Natural Substrate)	1	--
Uridine (RNA)	0	--

3.7 Bioactive Conformation Selection

With the identification and generation of a pharmacophore and the pharmacophoric conformations, respectively, the next objective was to select a bioactive conformation from amongst those presenting the pharmacophore. The conformational data (torsion and P angles) associated with each compound were analyzed for similarities. The conformations of individual compounds were displayed superimposed along the pharmacophoric atoms (see Figure 17). This provided a context for the type of variation in each compound which is principally responsible for the generation of the pharmacophoric conformations. It was found that in all compounds the base and sugar geometries were relatively static and conformational variation was principally due to the rotation of bonds 7 (7 values, see Figure 12) and 8 (4 values), which together can generate up to 28 conformations. Analyzing the conformational data with the four sugar bonds replaced by their associated P angle values, it was found that the number of

conformations in many compounds were multiples of 28 in which the difference between multiples was due to small changes in sugar puckering (P angle). In the case of AZT, the 84 conformations generated were due to three different pseudorotation phase angles which each repeated the same 28 conformations due to rotation of bonds 7 and 8 (28 x 3 = 84 conformations). There was a difference of $\pm 2°$ between the three pseudorotation phase angles of AZT (176°, 175°, and 174°), resulting in three sets of structurally redundant conformations. The production of redundant conformations due to small negligible changes in P angle was a factor observed in all compounds with greater than 28 conformations. To reduce the amount of conformational redundancy, the conformations in which the P angle was within $\pm 5°$ of another conformation were neglected. This reduced the number of conformations in each compound to $\approx 1/3$ their original values.

With the number of conformations for each compound truncated and the principal conformational variations due to rotatable bonds 7 and 8, energetics were used to select the most appropriate positions of these bonds. The energy of each of the remaining conformations was calculated using the AM1 method from Mopac. In Figure 18 the energy with respect to bonds 7 and 8 reveals energy troughs at $\approx 180°$ for rotatable bond 7 and $\approx 171°$ for rotatable bond 8 (where applicable). The troughs observed for AZT were at similar positions in the other compounds (see Appendix C for other compounds). The lowest energy conformation of each compound was selected based upon these observations.

Figure 17: All 84 conformations of AZT superimposed along the pharmacophoric atoms.

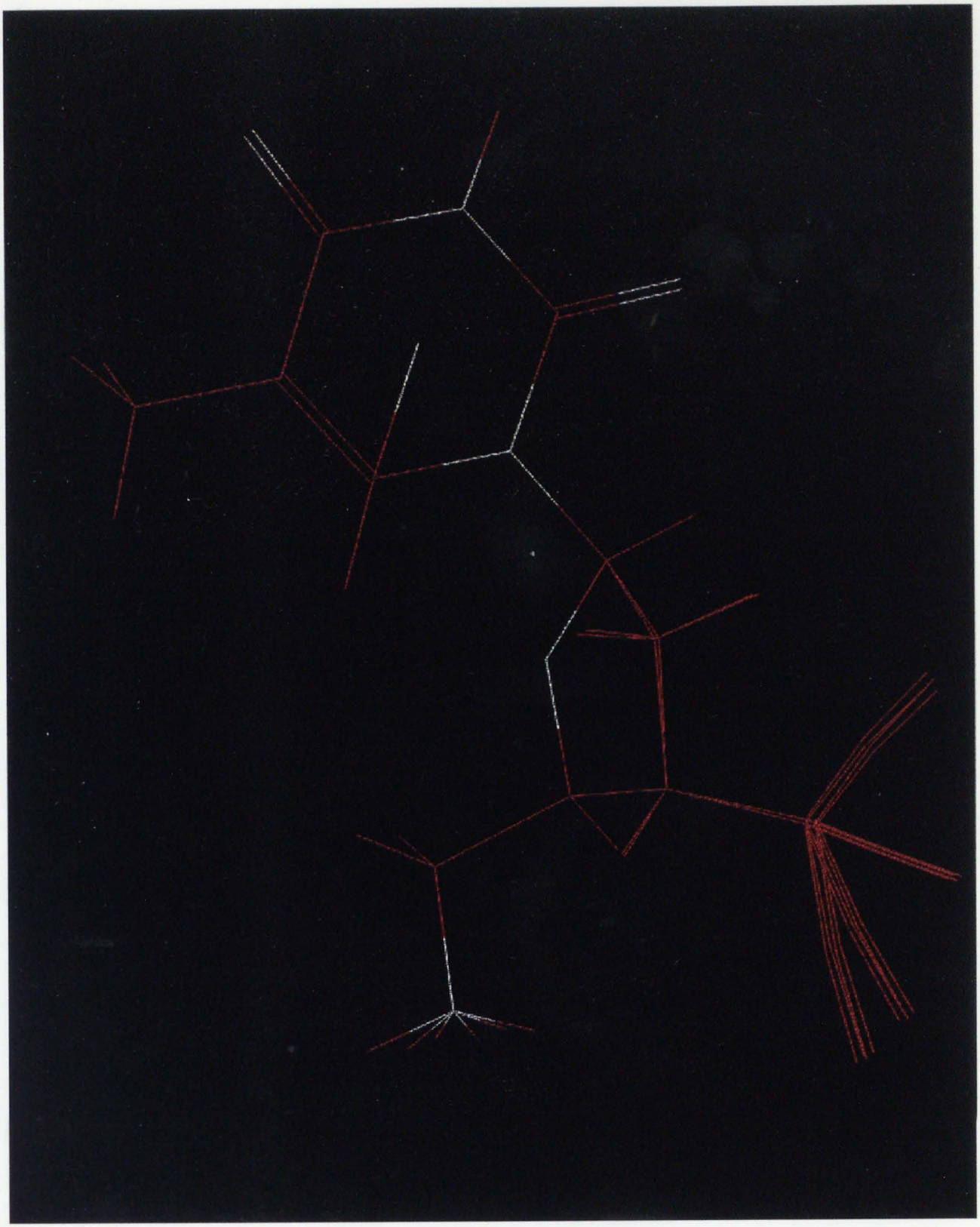
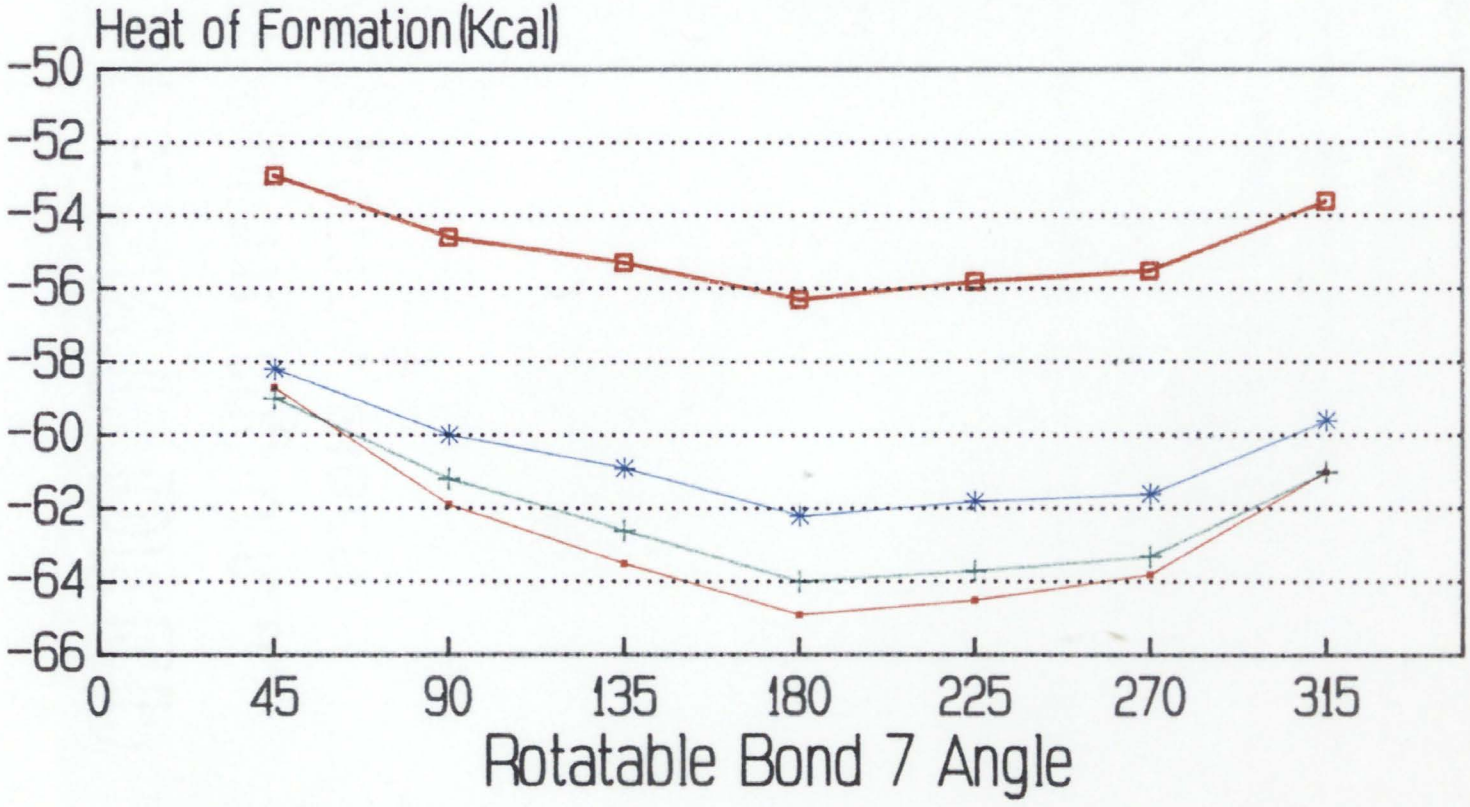


Figure 18: Graph of AM1 energy vs. rotatable bond 7 and 8 for AZT. Horizontal axis is rotatable bond 7, vertical axis is energy, and color represents position of rotatable bond 8. Notice the energy well at an angle of ° for rotatable bond 7 and the lowest energy graph for rotatable bond 8 (azido: 171.3°-red).

Energy vs. Rotatable Bond 7



Azido: (171.3) (126.3) (81.3) (306.3)

3'-Azido-2',3'-dideoxythymidine (AZT)

3.8 Receptor Mapping

With the bioactive conformation of each compound selected, the active and inactive compounds were superimposed along the pharmacophore and a union of the respective volumes was generated (see Figure 19). The intersection of these two volumes, which represents the portion of the inactive compounds located within the receptor site, was subtracted from the receptor essential (inactive) volume to produce a receptor map (see Figure 20). This is the net volume contribution of inactive analogs that is accessory to the enzyme excluded volume. It represents regions occupied by protein volume at the receptor site.

Figure 19: Active (left) and inactive (right) analogs superimposed along pharmacophoric atoms (white).

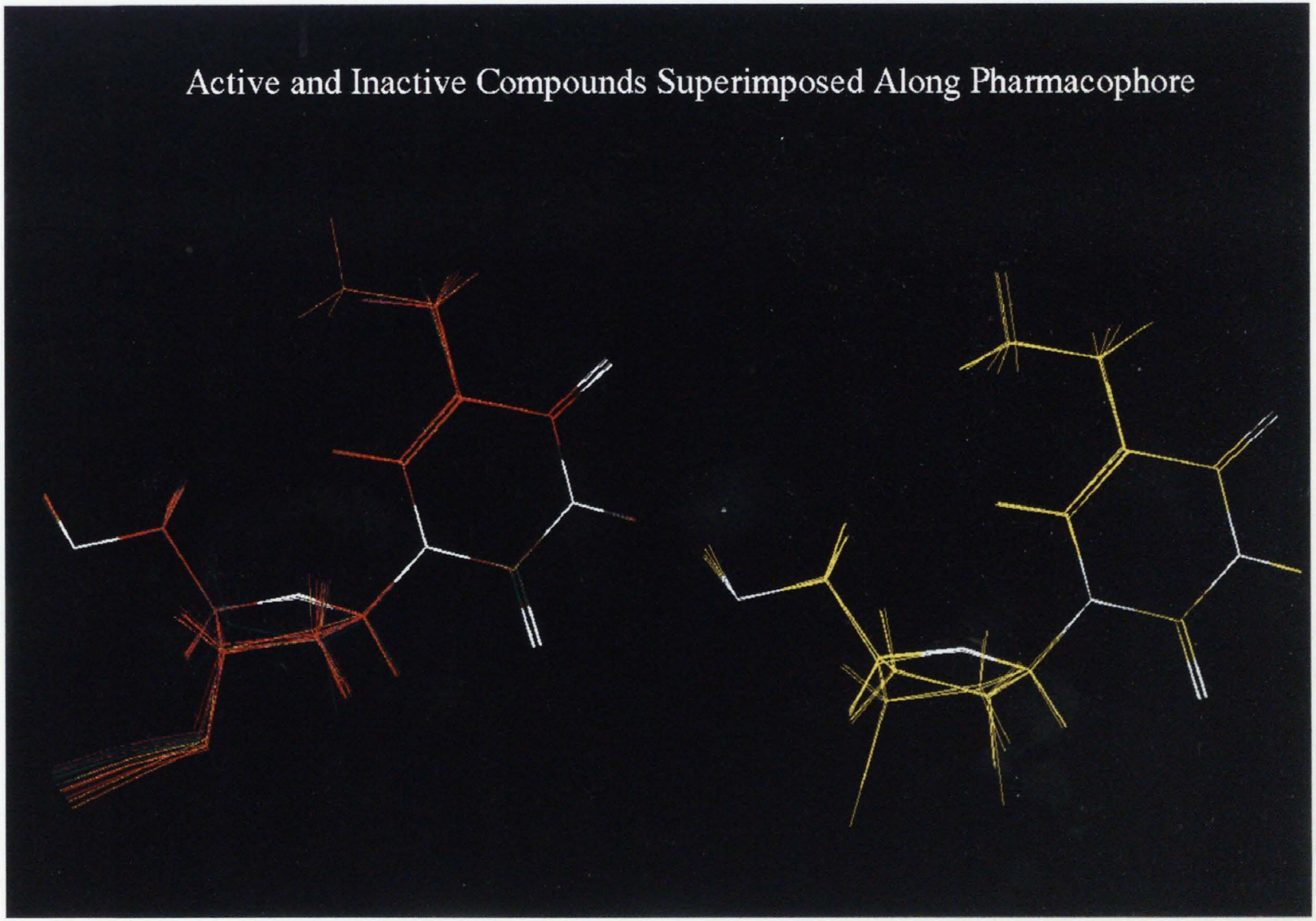
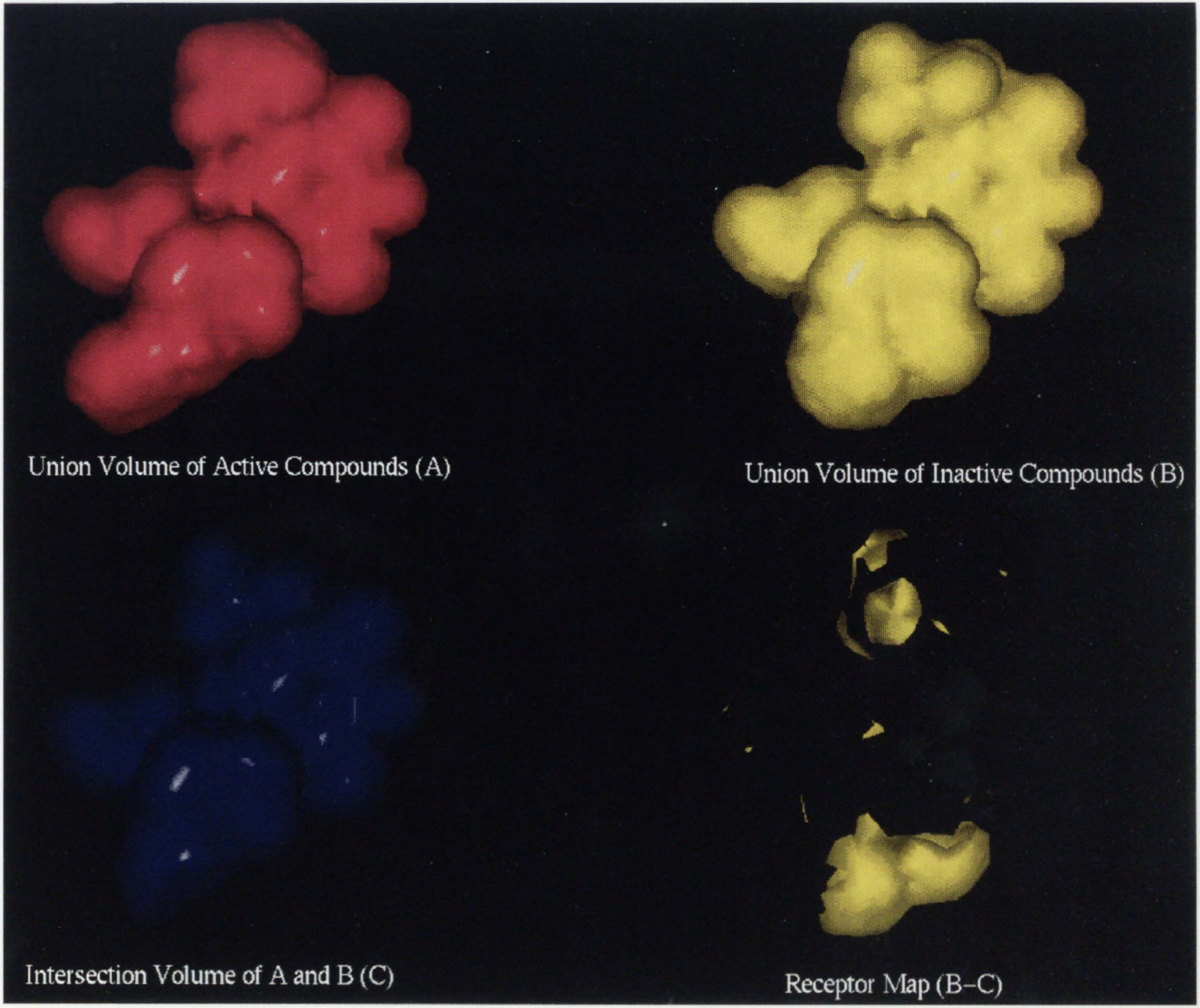

Active and Inactive Compounds Superimposed Along Pharmacophore

Figure 20: Union volume of active (magenta) and inactive (yellow) analogs superimposed along pharmacophore, Intersection volume of active and inactive analogs superimposed along pharmacophore (blue), and Receptor map (yellow).


Union Volume of Active Compounds (A)
Union Volume of Inactive Compounds (B)
Intersection Volume of A and B (C)
Receptor Map (B–C)

CHAPTER 4
CONCLUSION

In this study, A method of molecular analysis that is independent of any preconceived mechanistic notions was used to identify a unique pharmacophore for thymidine nucleoside analogs and infer a three dimensional receptor map of the receptor site of HIV-1 RT. The pharmacophore obtained in this is currently being extended to a larger number of active and inactive RT inhibitors to further validate it as the minimal structural requirement for thymidine nucleoside analog-like activity at RT. Along with providing a frame of reference for the orientation of these analogs at the RT receptor, It is hoped that this pharmacophore will provide a template for the screening and identification of new inhibitor molecules. A recently released adjunct to the SYBYL molecular modeling program [22] allows the application of pharmacophoric data as a three dimensional query constraint to the Cambridge Crystallographic Database, which has structural information on $>60{,}000$ compounds. This will allow the identification compounds that can present the pharmacophoric pattern and, thus, which may have inhibitory potential. The use of interatomic distances as the method of pharmacophore definition allows the identification of compounds that are capable of presenting the proposed pharmacophore outside of the thymidine nucleoside molecular framework.

The receptor map provides an idea of the areas on nucleoside analogs that are responsible for poor/in-activity at the RT receptor site. The map identified in this study, which shows a great deal of volumetric conflict in the region of the sugar ring, is in agreement with a number of contemporary opinions about the important regions on these compounds. This receptor map will be used in conjunction with biochemical and structural data associated with a recently

solved crystal structure of HIV-1 RT to identify the specific region on its surface with the greatest structural complementarity[15]. Each compound under consideration will be positioned in their bioactive conformations at the proposed receptor site and minimized as a protein-ligand complex. It is hoped that the information obtained from the continuation of this study will provide useful guidelines for the identification and synthesis of new inhibitors compounds and facilitate a better understanding of the stucture-function relationships between nucleoside analogs.

APPENDIX A

MATHCAD MODEL FOR THE CALCULATION OF THE PSEUDOROTATION PHASE ANGLE (P)

The Student Edition of MathCAD 2.0 For Educational Use Only

MCAD model for calculation of Pseudorotation Phase Angle(P) in sugar ring using individual torsional angles: V0(C4'-O4'-C1'-C2'), V1(O4'-C1'-C2'-C3'), V2(C1'-C2'-C3'-C4'), V3(C2'-C3'-C4'-O4'), & V4(C3'-C4'-O4'-C1')

Define Constants: $rad \equiv 1$

$$deg \equiv \frac{\pi}{180} \cdot rad$$

Import Angle Data: M := READPRN(filename) i := rows(M)

x := 1,2 ..i

The following WRITE/READ equations convert the angle file from sybyl (an i x 4 matrix) to a usable mcad format:

WRITE(M0) := $M^{\langle 0 \rangle}$	M0 := READ(M0)
WRITE(M4) := $M^{\langle 1 \rangle}$	$M4_x$:= READ(M4)
WRITE(M3) := $M^{\langle 2 \rangle}$	$M3_x$:= READ(M3)
WRITE(M2) := $M^{\langle 3 \rangle}$	$M2_x$:= READ(M2)

Generation of torsion angle in -40 to 40 degree range:

$$V0(x) := if\left[M0_x > 270, \left[M0_x - 359.9\right], M0_x\right]$$

$$V2(x) := if\left[M2_x > 270, \left[M2_x - 359.9\right], M2_x\right]$$

$$V3(x) := if\left[M3_x > 270, \left[M3_x - 359.9\right], M3_x\right]$$

$$V4(x) := if\left[M4_x > 270, \left[M4_x - 359.9\right], M4_x\right]$$

$$V1(x) := -(V0(x) + V2(x) + V3(x) + V4(x))$$

$$Y(x) := (V4(x) + V1(x)) - (V3(x) + V0(x))$$

$$X(x) := 2 \cdot V2(x) \cdot 1.5388$$

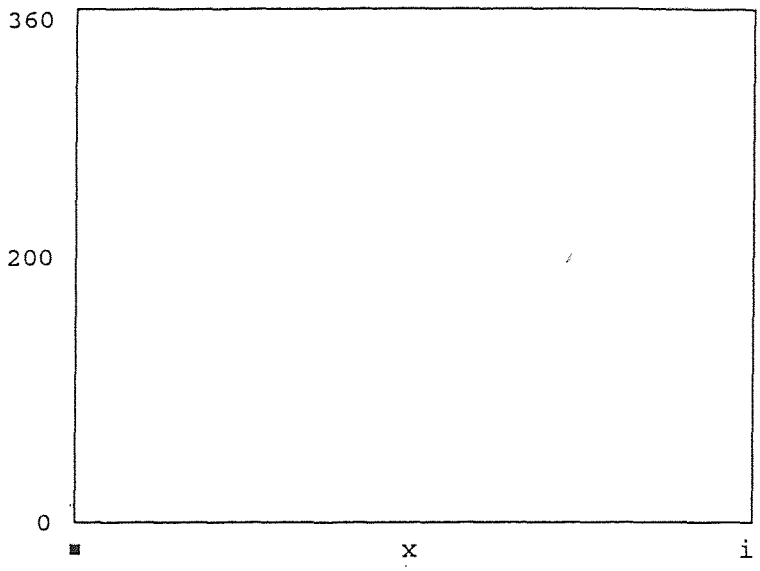
$$P(x) := atan\left[\frac{Y(x)}{X(x)}\right] \qquad P(x) := \frac{P(x)}{deg}$$

$$P(x) := if(V2(x) < 0, (P(x) + 180), P(x))$$

$$P(x) := if(P(x) < 0, (P(x) + 360), P(x))$$

APPENDPRN(AZT8744P) := P(x)

i =



APPENDIX B

SEARCH DATA FOR MUTIPLE PAIR CONSTRAINTS

Data for Two Heteroatom Pair Constraints

Lead Pair	Pair2	OMAP Points
A	B	147
	C	221
	D	117
	E	118
	F	0
	G	184
	H	111
B	C	152
	D	79
	E	98
	F	0
	G	177
	H	113
C	D	126
	E	49
	F	0
	G	130
	H	181
D	E	75
	F	0
	G	130
	H	90
E	F	0
	G	33
	H	99
F	G	0
	H	0
G	H	160

Data for 3 Heteroatom Pair Constraints

Lead Pair	Pair2	Pair3	OMAP Points
A	B	C	248
		D	135
		E	145
		G	300
		H	162
A	C	D	170
		E	105
		G	223
		H	226
A	D	E	52
		G	333
		H	130
A	E	G	325
		H	104
A	G	H	324
B	C	D	141
		E	90
		G	227
		H	198
B	D	E	88
		G	316
		H	121
B	E	G	309
		H	132
B	G	H	290
C	D	E	80
		G	209
		H	184
C	E	G	206
		H	99
C	G	H	220
D	E	G	198
		H	80
D	G	H	256
E	G	H	268

Data for 4 Heteroatom Pair Constraints

Pair1	Pair2	Pair3	Pair4	OMAP Points
A	B	C	D	159
			E	109
			G	189
			H	217
		D	E	50
			G	146
			H	131
		E	G	104
			H	109
		G	H	236
	C	D	E	27
			G	88
			H	171
		E	G	29
			H	78
		G	H	126
	D	E	G	151
			H	48
	E	G	H	253
B	C	D	E	39
			G	86
			H	161
		E	G	36
			H	82
		G	H	140
	D	E	G	103
			H	78
		G	H	197
	E	G	H	124
C	D	E	G	39
			H	58
		G	H	129

Pair1	Pair2	Pair3	Pair4	OMAP Points
	E	G	H	43
D	E	G	H	189

Data for 5 Heteroatom Pair Constraints

Pair1	Pair2	Pair3	Pair4	Pair5	OMAP Points
A	B	C	D	E	25
				G	70
				H	150
			E	G	26
			H	66	
			G	H	109
		D	E	G	33
			H	43	
			G	H	132
		E	G	H	64
	C	D	E	G	38
				H	47
			G	H	32
		E	G	H	22
	D	E	G	H	21
B	C	D	E	G	13
				H	35
			G	H	77
		E	G	H	32
	D	E	G	H	67
C	D	E	G	H	19

Data for 6 Heteroatom Pair Constraints

Pair1	Pair2	Pair3	Pair4	Pair5	Pair6	OMAP Points
A	B	C	D	E	G	7
					H	23
				G	H	64
			E	G	H	18
B	C	D	E	G	H	13

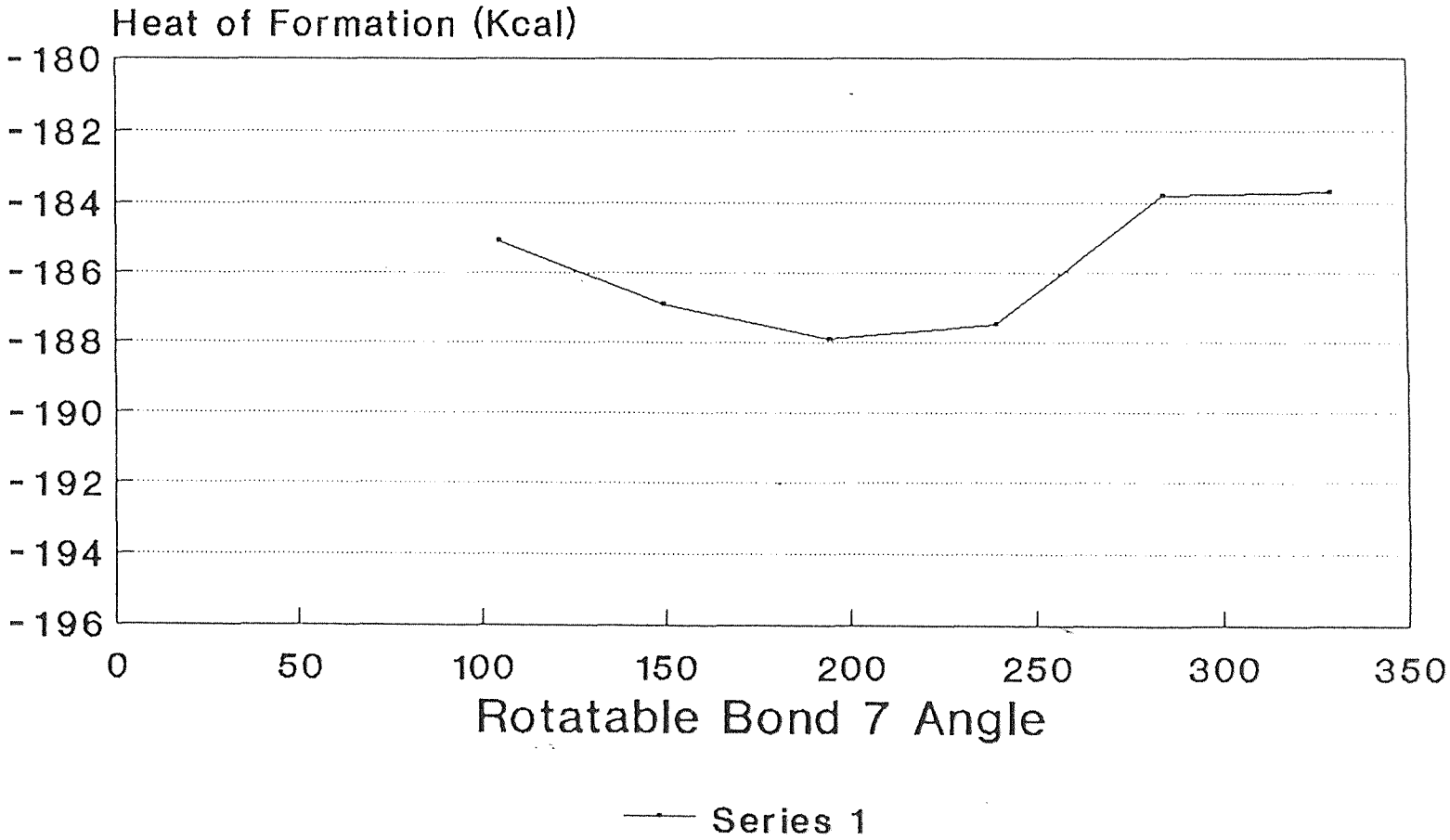
Data for 7 Heteroatom Pair Constraints

Pair1	Pair2	Pair3	Pair4	Pair5	Pair6	Pair7	OMAP Points
A	B	C	D	E	G	H	1

APPENDIX C

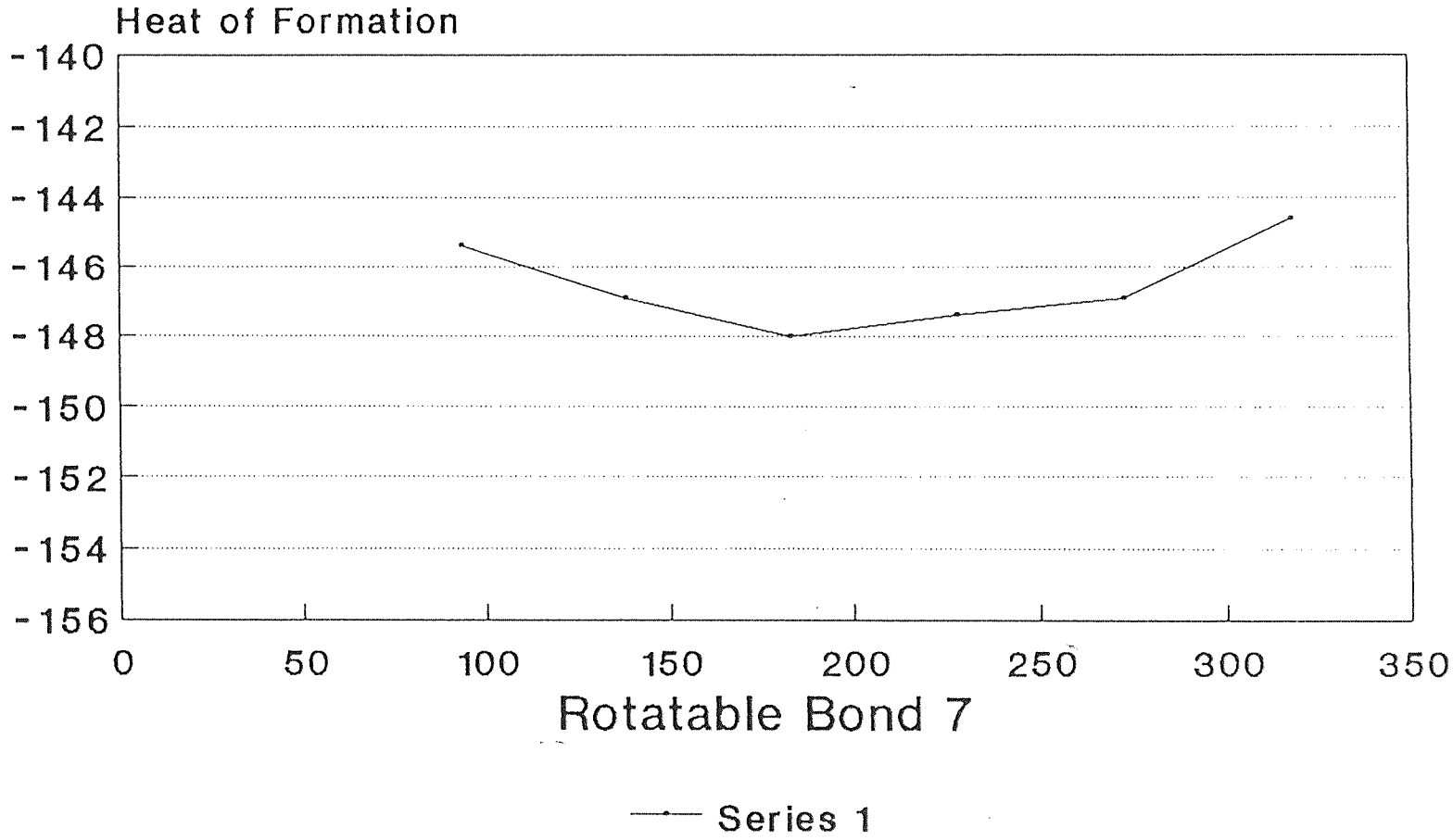
ENERGY VS ROTATABLE BOND 7 AND 8

Energy vs. Rotatable Bond 7



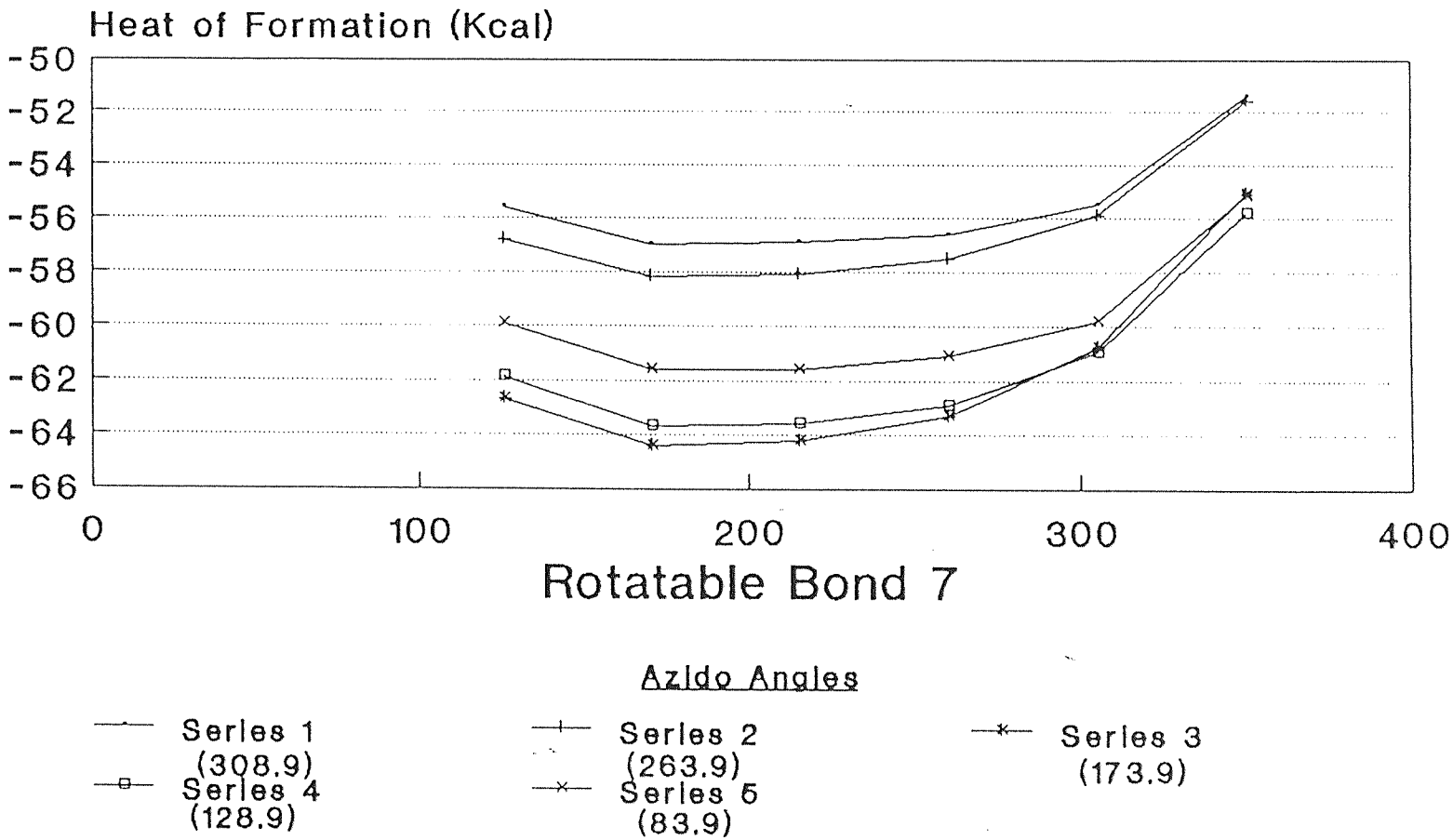
3'-Fluoro-2',3'-dideoxythymidine

Energy vs. Rotatable Bond 7



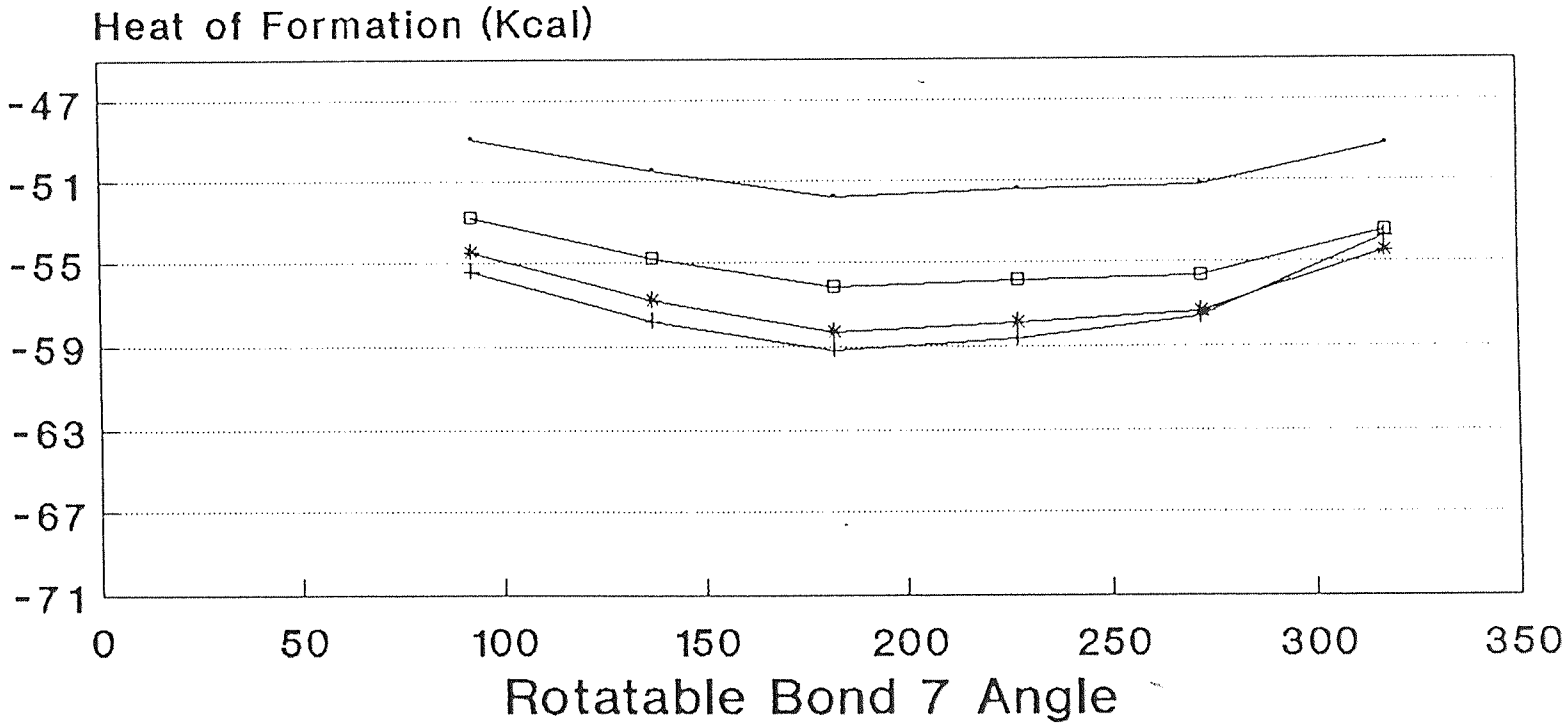
2',3'-dideoxythymidine

Energy vs. Rotatable bond 7



5-Ethyl-3'-Azido-2',3'-dideoxyuridine

Energy vs Rotatable Bond 7



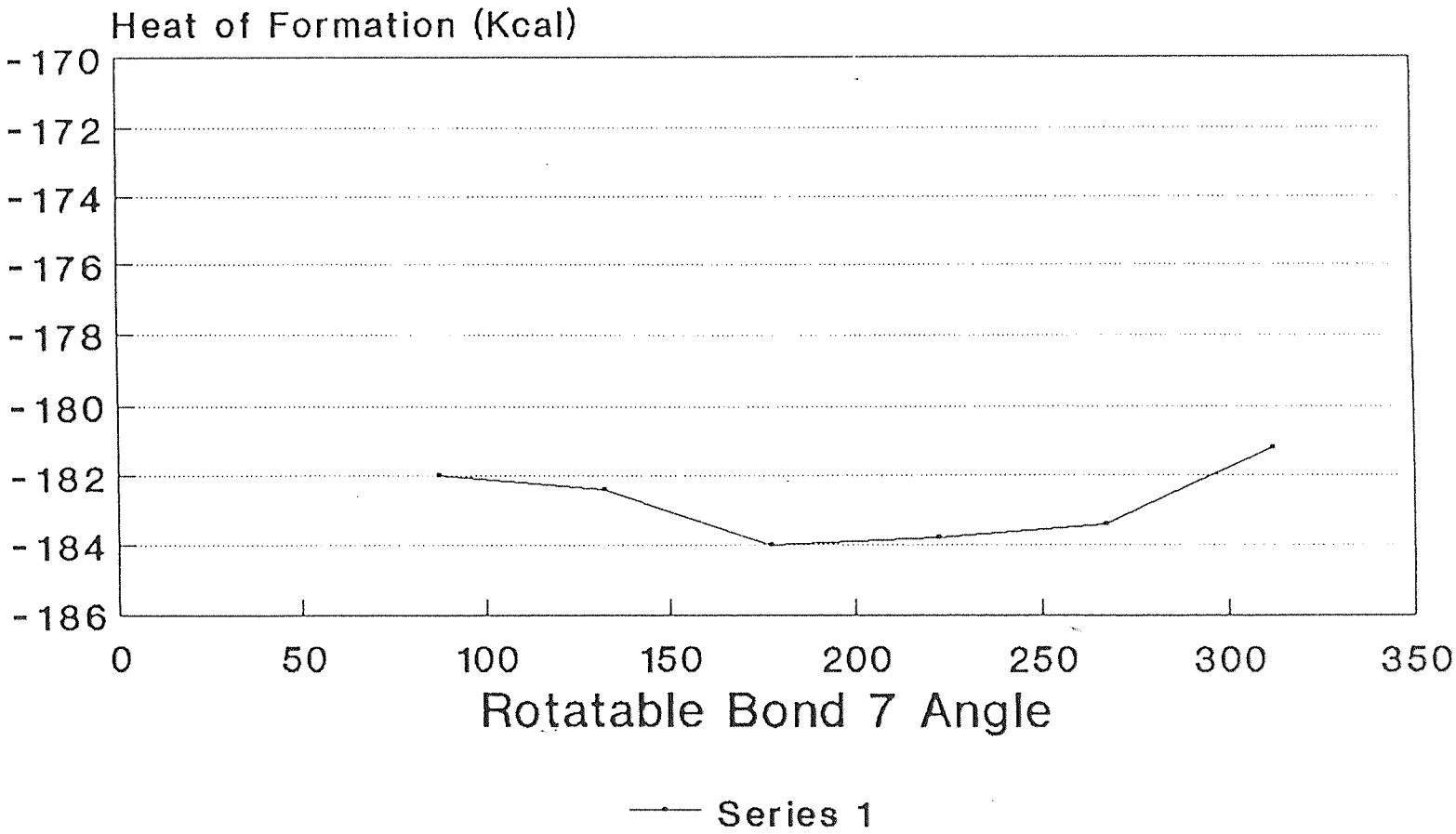
Azido Angles

Azido:

Series 1	Series 2	Series 3	Series 4
(302.8)	(167.8)	(122.8)	(77.8)

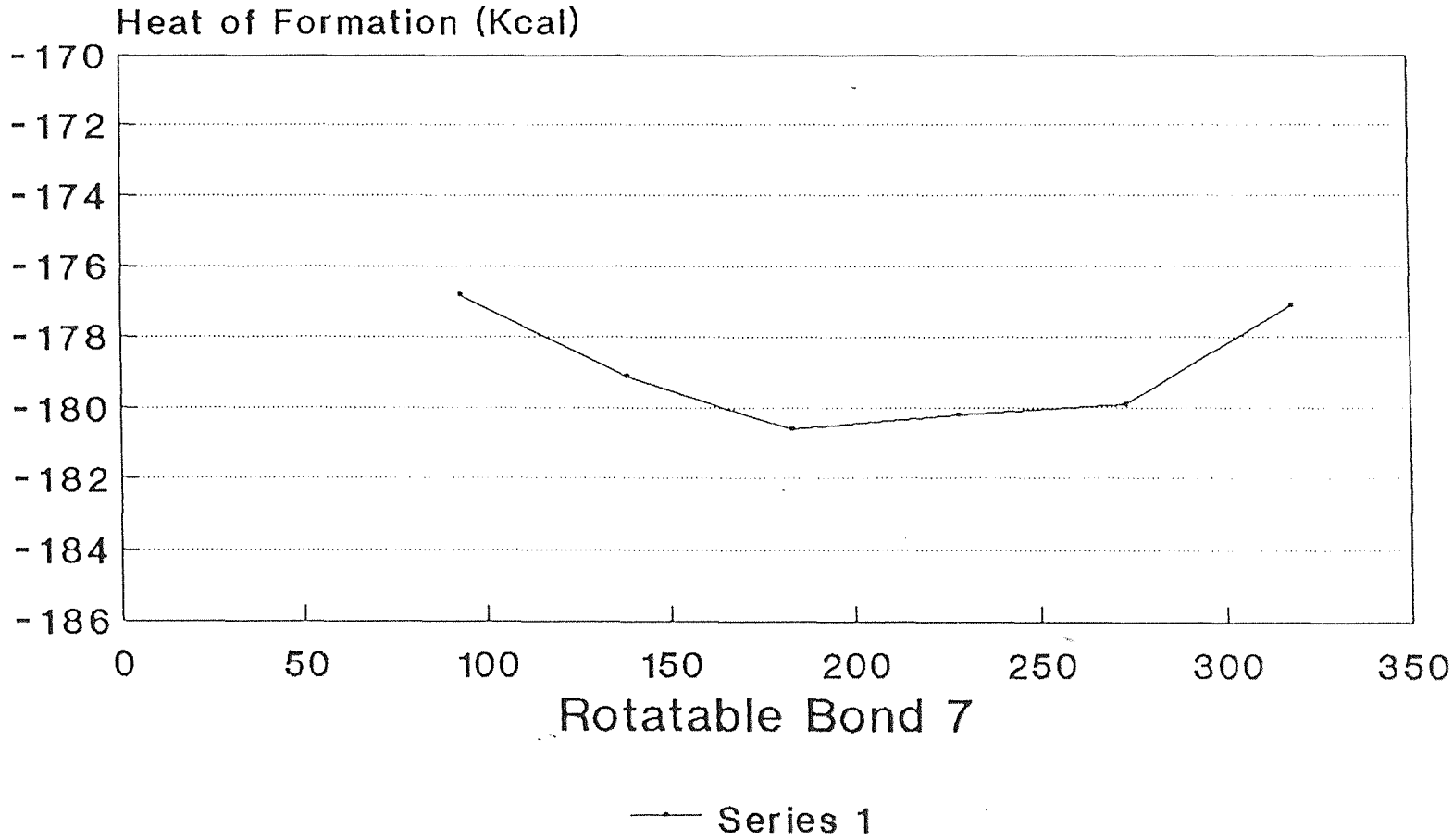
3'-Azido-2',3'-dideoxyuridine

Energy vs. Rotatable Bond 7



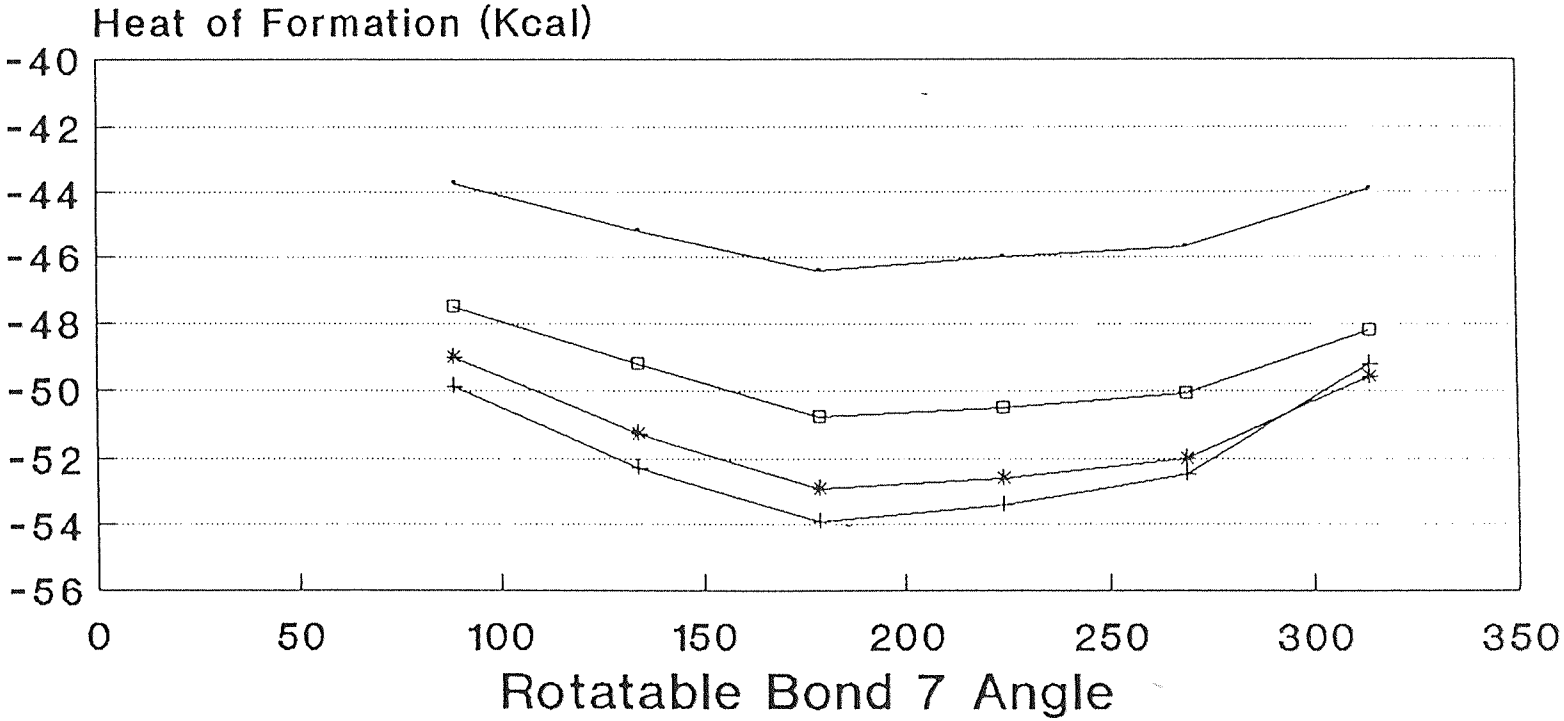
5-Chloro-3'-Fluoro-2',3'-dideoxythymidine

Energy vs. Rotatable Bond 7



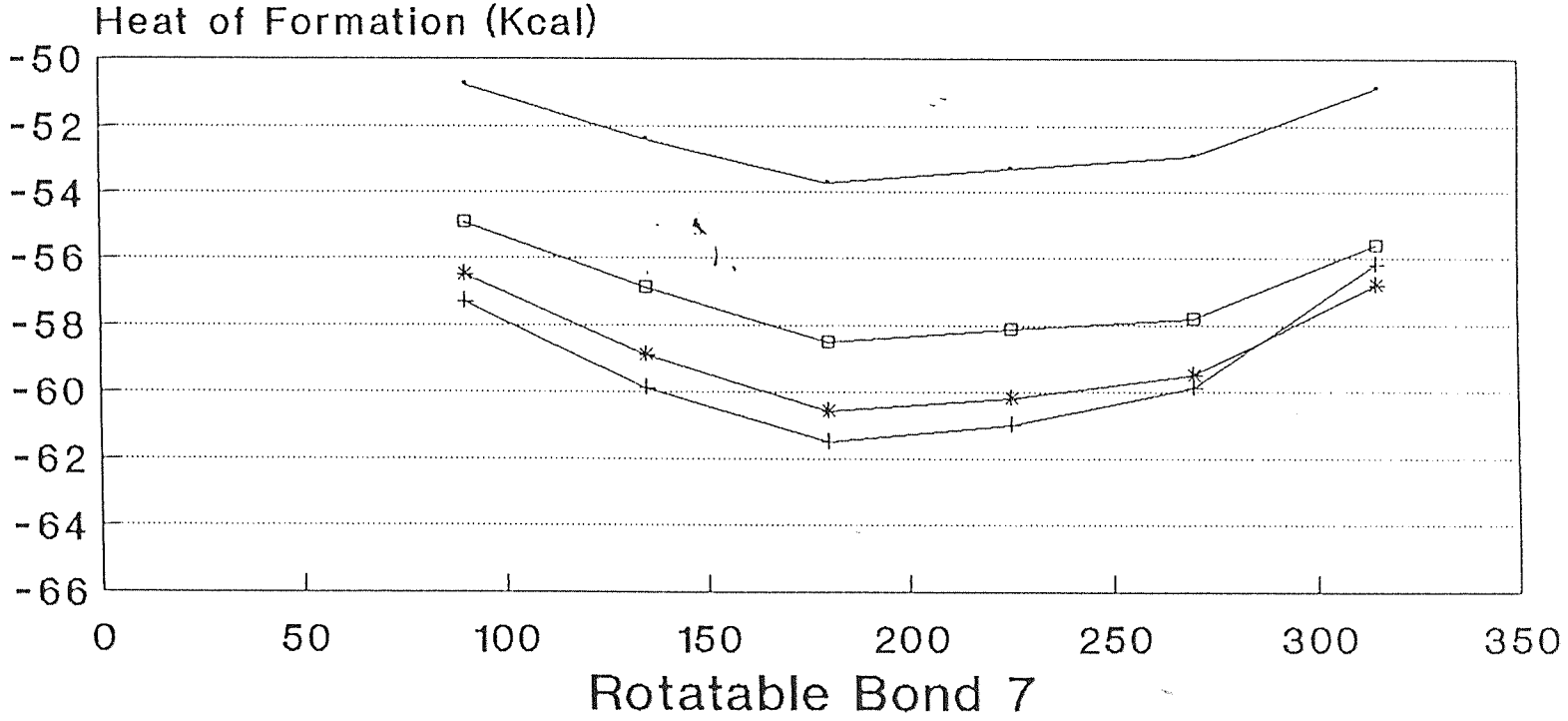
3'-Fluoro-2',3'-dideoxyuridine

Energy vs. Rotatable Bond 7



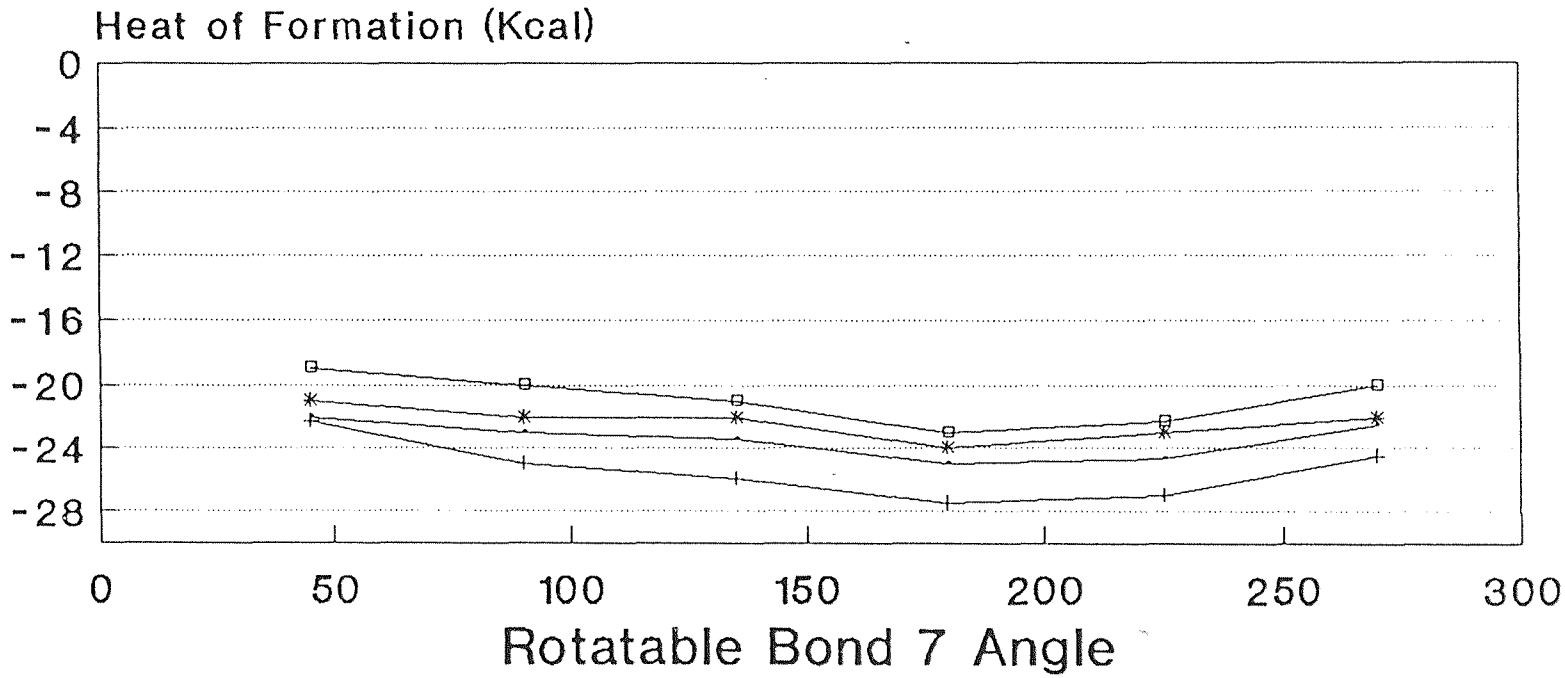
5-Bromo-3'-Azido-2',3'-dideoxythymidine

Energy vs Rotatable Bond 7



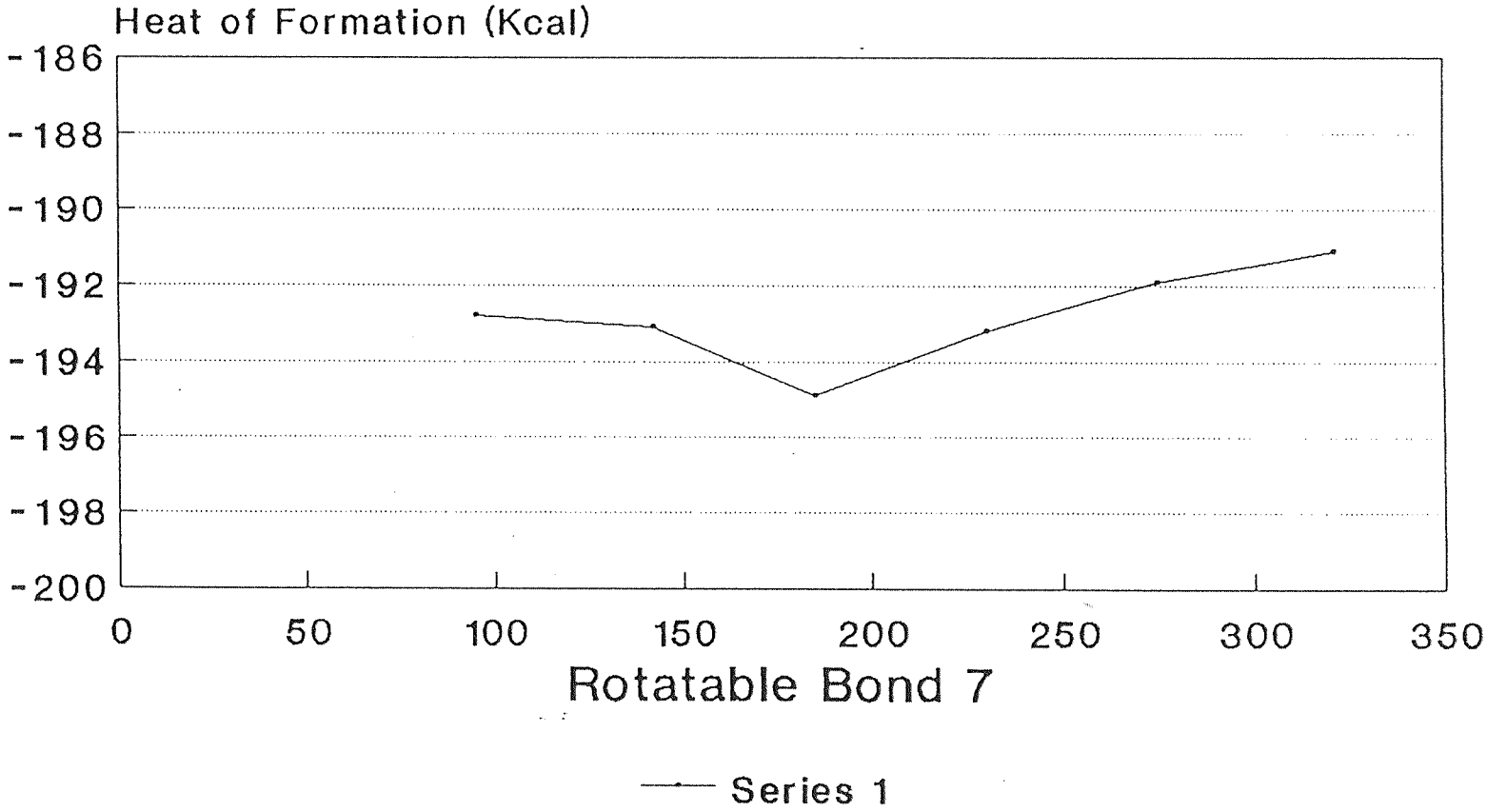
5-Chloro-3'-Azido-2',3'-dideoxyuridine

Energy vs Rotatable Bond 7



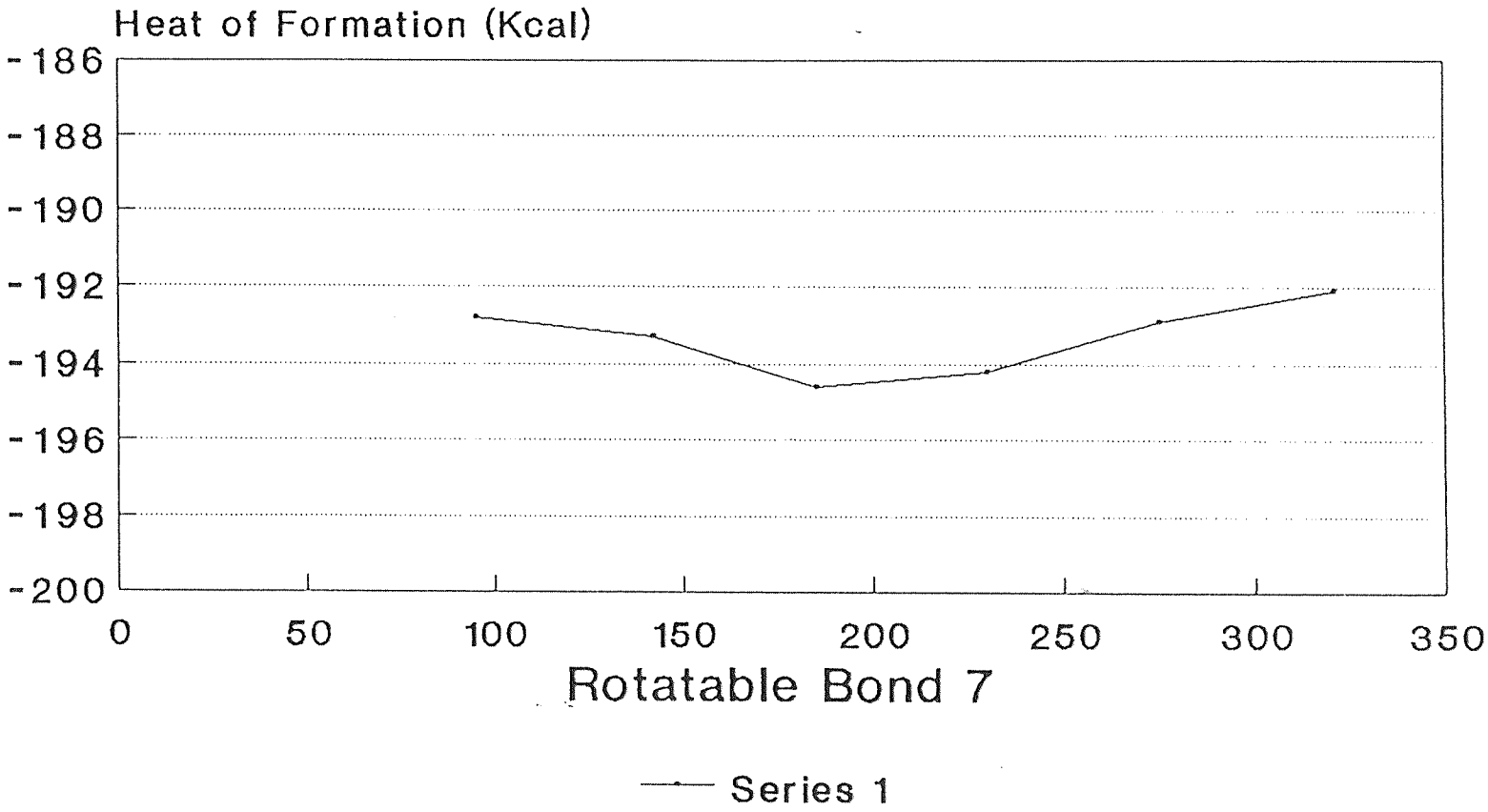
5-Iodo-3'-Azido-2',3'-dideoxyuridine

Energy vs Rotatable Bond 7



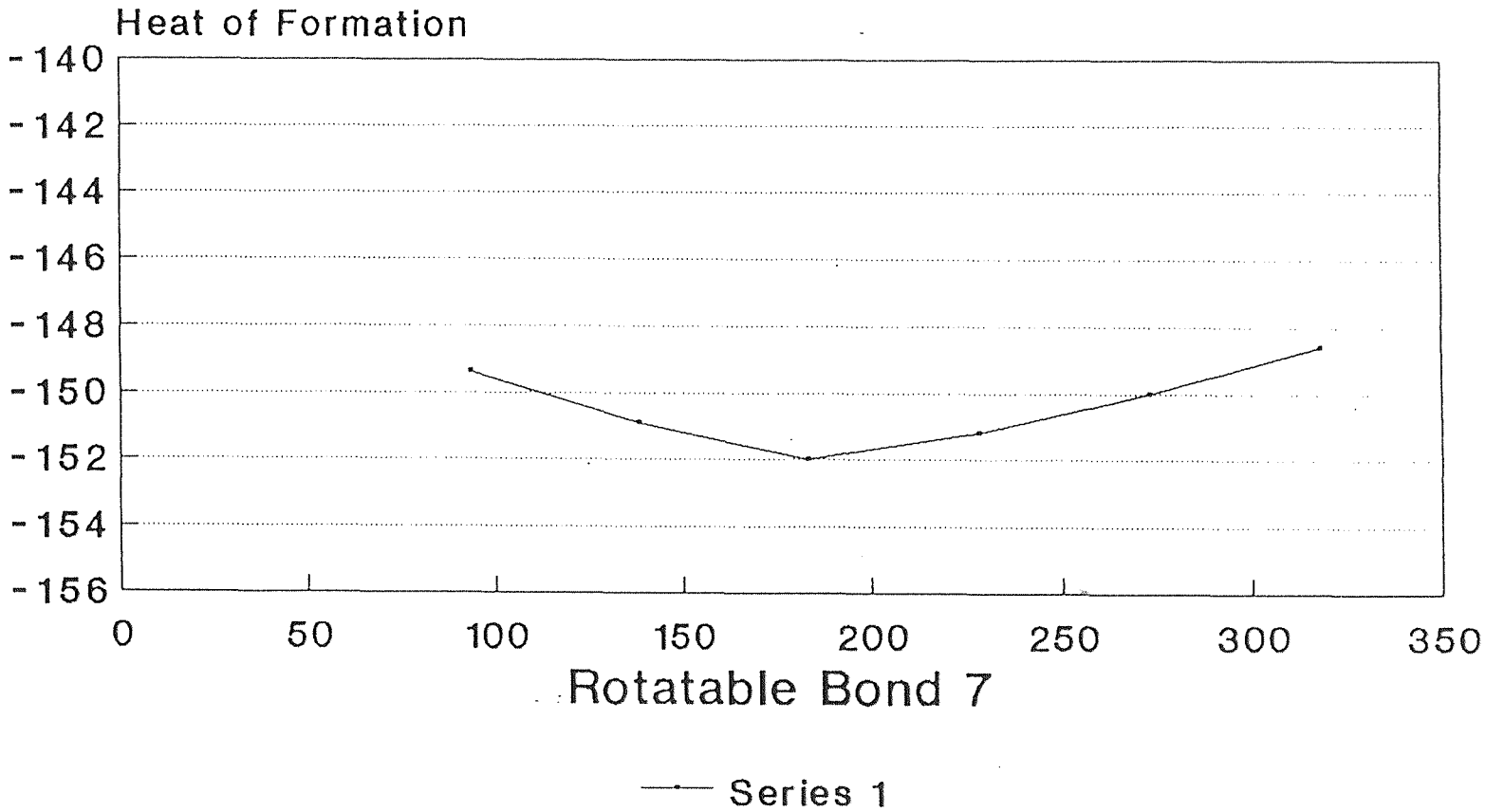
5-Ethyl-3'-Fluoro-2',3'-dideoxyuridine

Energy vs Rotatable Bond 7



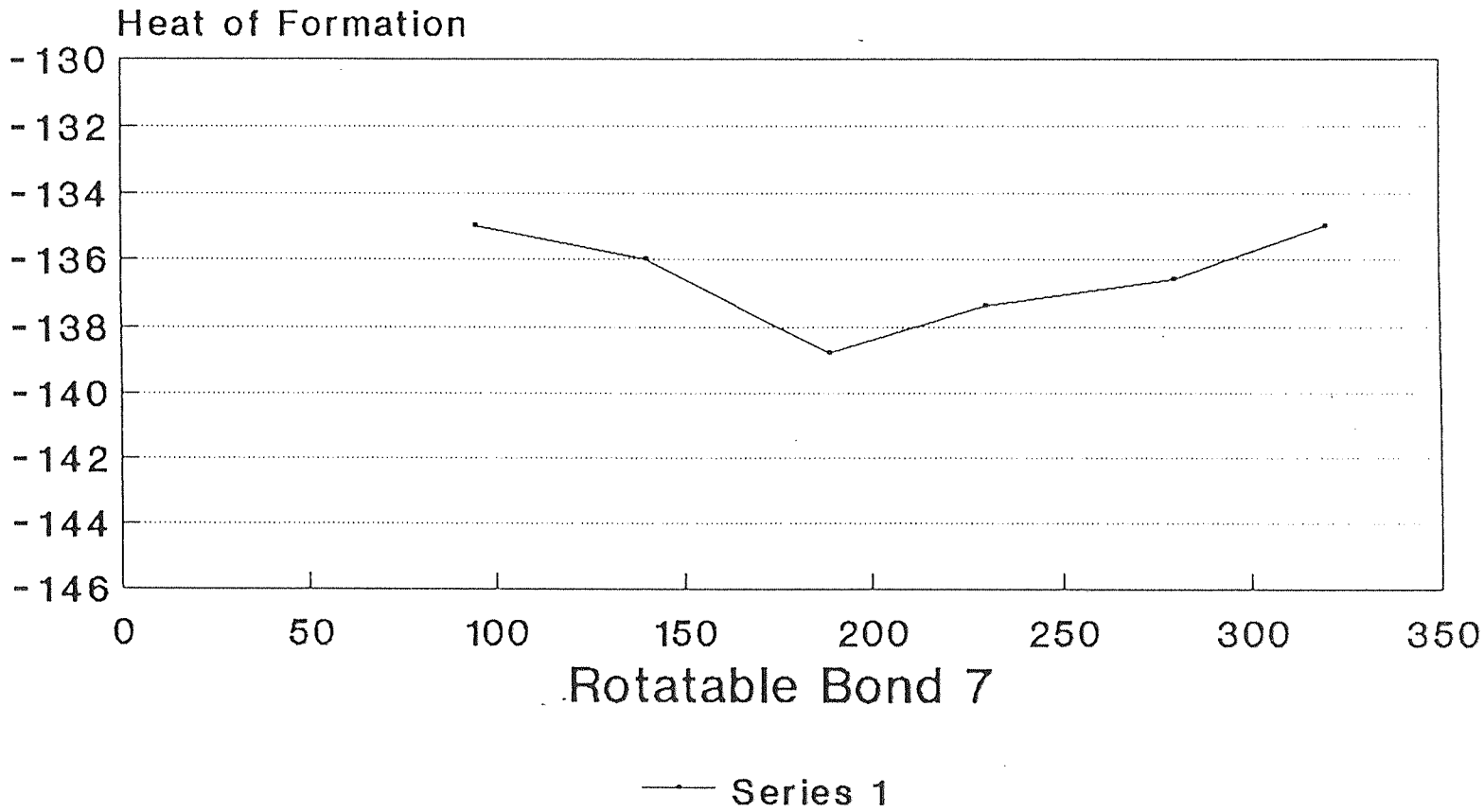
5-Ethyl-3'-Iodo-2',3'-dideoxyuridine

Energy vs Rotatable Bond 7



2',3'-dideoxyuridine

Energy vs Rotatable Bond 7



2'-Chloro-2',3'-dideoxyuridine

REFERENCES

1. Van Roey, P., Sallerno, J.M., Chu, C.K., Schinazi, R.F. 1989. Correlation Between Preferred Sugar Ring Conformation and Activity of Nucleoside Analogues Against Human Immunodeficeincy Virus. *Proceedings of National Academy of Sciences-USA*. 86: 3929-3933.

2. DeClercq, E. 1992. HIV Inhibitors Targeted at the Reverse Transcriptase. *AIDS Research and Human Retroviruses*. 8(2): 119-134.

3. Nasr, M., Cradock, J., Johnson, M.I. 1992. Computer Assisted Structure Activity Correlations of Halodideoxynucleoside Analogs as Potential Anti HIV Drugs. *AIDS Research and Human Retroviruses*. 8(2): 135-144.

4. Crabbe, M.J., Falkinbridge, S.S. 1991. Molecular Modeling of HIV-1 Reverse Transcriptase Inhibitors. *AIDS Research and Human Retroviruses*. 7(3): 261-264.

5. von Itzstein, M., Wu, W.Y., Kok, G.B, Pegg, M.S., Dyason, J.C., Jin, B., Phan, T.V., Smythe, M.L., White, H.F., Oliver, S.W., Colman, P.M, Varghese, J.N., Ryan, D.M., Woods, J.M., Bethell, R.C., Hotham, V.J., Cameron, J.M., Penn, C.R. 1993. Rational Design of Potent Sialidase-Based Inhibitors of Influenza Virus Replication. *Nature*. 363: 418-423.

6. Kuntz, I.D. 1992. Structure-Based Strategies for Drug Design and Discovery. *Science*. 257: 1078-1082.

7. Weber, A.E., Halgren, T.A., Doyle, J.J., Lynch, R.J., Sieyl, P.K., Parsons, W.H., Greenlee, W.T., Patchett, A.A. 1991. Design and Synthesis of P2-P1' Linked Macrocyclic Human Renin Inhibitors. *J. Medicinal Chemistry*. 34(9): 2692-2701.

8. Debouck, C. 1992. The HIV-1 Protease as a Therapeutic Target for AIDS. *AIDS Research and Human Retroviruses*. 8(2): 153-164.

9. Marshall, G.R., Barry, C.D., Bosshard, H.E., Dannkoehler, R.A., Dunn, D.A. 1979. The Conformational Parameter in Drug Design: The Active Analog Approach. *Computer Aided Drug Design, ACS Symposium Series 112*: 205-226.

10. Mayer, D., NAylor, C.B., Motoc, I., Marshall, G.R. 1987. A Unique Geometry of the Active Site of Agiotensin-Converting Enzyme Consistent with Structure Activity Studies. *J. Computer Aided Molecular Design*. 1: 3-16.

11. Lloyd, E.J., Andrews, P.R. 1986. A common Structural Model for Central Nervous System Drugs and Their Receptors. *J. American Chemical Society*. 29: 453-462.

12. Haseltine, W.A. 1991. Molecular Biology of the Human Immunodeficiency Virus type-1. *FASEB*. 5: 2349-2358.

13. Stryer, L. 1988. Biochemistry. W.H. Freeman and Company. New York: 851-879.

14. Kolstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., Steitz, T.A. 1992. Crystal Structure at 3.5 Å Resolution of HIV-1 Reverse Transcriptase Complexed with an Inhibitor. *Science*. 256: 1783-1790.

15. Jacobo-Molina, A., Ding, J., Nanni, R., Clark, A., Lu, X., Tantillo, C., Williams, R., Kamer, G., Ferris, A., Clark, P., Hizi, A., Hughes, S., Arnold, E. 1993. Crystal Structure of Human Immunodeficiency Virus Type 1 Reverse Transcriptase Complexed with Double Stranded DNA at 3.0 Å Resolution Shows Bent DNA. *Proceedings of the National Academy of Sciences-USA*. 90: 6320-6324.

16. Declercq, E. 1987. Targets for Antiviral Activity of Purimidine and Purine Nucleoside Analogues. *Nucleosides and Nucleotides*. 6(1): 197-207.

17. Camerman, A., Mastropaolo, Camerman, N. 1987. Azidothymidine: Crystal Structure and Possible Functional Role of the Azido Group. *Proceedings of the National Academy of Sciences-USA*. 84: 8239-8242.

18. Richards, W.G. 1989. *Computer Aided Molecular Design*. IBC Technical Services LTD.. Oxford: 43-50.

19. Fisher, M.A., Yadav, P.N.S., Yadav, J., Arnold, E., Modak, M. 1993. Identification of a "Pharmacophore" for Nucleoside Analog Inhibitors Directed at HIV-1 Reverse Transcriptase. *J. Molecular Recognition* (submitted August, 1993).

20. Saenger, W. 1984. Principles of Nucleic Acid Structure. Springer-Verley. New York: 9-101.

21. Mathcad. Mathsoft Incorporated, Cambridge, Massachusetts.

22. SYBYL: Molecular Modeling System. Tripos Associates, St. Louis, Missouri.

23. Quantum Chemistry Program Exchange, Department of Chemistry , Indiana University, Bloomington, Indiana.

24. Cohen, N.C., Blaney, J.M., Humblet, C., Gund, P., Barry, D.C. 1990. Molecular Modeling Software and Methods for Medicinal Chemistry. *J. Medicinal Chemistry*. 33(3): 883-894.

25. Beese, L.S., Derbyshire, V., Steitz, T.A. 1993. Structure of DNA Polymerase I Klenow Fragment Bound Duplex DNA. *Science*. 260: 352-355.

26. Baba, M., DeClercq, E., Tanaka, H. Ubasawa, M., Takashima, H., Sekiya, K., Nitta, I., Umizu, K., Nakashima, H., Mori, S., Shigeta, S., Walker, R.T., Miyasaka, T. 1991. Potent and Selective Inhibition of Human Immunodeficiency Virus Type 1 (HIV-1) by 5-Ethyl-6-Phenythiouracil Derivatives through their Interaction with the HIV-1 Reverse

Transcriptase. *Proceedings National Academy of Science-USA*. 88: 2356-2360.

27. Rabson, A.B. , Martin, M.A. 1985. Molecular Organization of the AIDS Retrovirus. *Cell*. 40: 477-480.

28. Cullen, B.R. 1991. Regulation of HIV-1 Gene Expression. *FASEB*. 5: 2361-2368.

29. Lowe, D.M, Parmar, V., Kemp, S.D., Larder, B.A. 1991. Mutational Analysis of Two Conserved Sequence Motifs in HIV-1 Reverse Transcriptase. *FEBS*. 282(2): 231-234.

30. Clark, M., Cramer III, R.D., Opdenbosch, N.V. 1989. Validation of General Purpose Tripos 5.2 Force Field. *J. Computational Chemistry*. 10(8): 982-1012.

31. Naruto, S., Motoc, I., Marshall, G.R. 1985. Computer-Assisted Analysis of Bioactivity: I.-Active Conformation of Histamine H_1 Receptor Antagonist. *Eur. J. Medicinal Chemistry*. 20(6): 529-532.

32. Hu, W.S., Temin, H.M. 1990. Retroviral Recombination and Reverse Transcription. *Science*. 250: 1227-1233.

33. Chu, C.K., Schinazi, R.F., Ahn, M.K., Ullas, G.V., Gu, Z.P. 1989. Structure Activity Relationships of Pyrimidine Nucleosides as Antiviral Agents for Human Immunodeficiency Virus Type 1 in Peripheral Blood Mononuclear Cells. *J. Medicinal Chemistry*. 32: 612-617.