Molecular Recognition Through Rational and Combinatorial Synthesis

by Edward Aurel Wintner
B.S., Chemistry
Yale University

Submitted to the Department of Chemistry
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

at the
Massachusetts Institute of Technology
June, 1996

Signature of Author ....
Department of Chemistry
May 20, 1996

\

Certified by ....
.... james ...., j.. \\
U Thesis Supervisor

Accepted by ..../.............................\..........
Professor Dietmar Seyferth
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:

Professor Gregory C. Fu ....

.... Chairman

Professor Julius Rebek, Jr. .....

.... upervisor

Professor John M. Essigmann ...

Molecular Recognition Through Rational and Combinatorial Synthesis

by
Edward Aurel Wintner

Submitted to the Department of Chemistry on May 20, 1996 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Organic Chemistry

Abstract

Molecular recognition is achieved through the rational and combinatorial synthesis of molecules.

Part I describes the synthesis of two systems of self replicating molecules: molecules designed to catalyze their own formation. The syntheses stem from a knowledge of the molecular recognition of adenine through hydrogen bonding with imide functionalities and through Π-stacking with aromatic surfaces. The first system, based on a mono-imide molecule, was shown to replicate poorly. However, insight from this result allowed the design of a second di-imide system that showed excellent and highly specific replication of product from its substrates. Furthermore, a previously developed self-replicating system was re-investigated. The possibility of non-specific chemical autocatalysis in the system was ruled out, confirming that autocatalysis is the result of the molecular recognition designed into the structures.

Part II describes a new method of combinatorial synthesis based on the combination of a polyfunctionalized core molecule with a set of diverse building blocks. The polysubstituted cores thus produced were analyzed by mass spectrometric techniques, and the novel combinatorial method was shown to produce the range of substituted compounds expected from mathematical models. Using the new method, mixtures ("libraries") of tens of thousands of molecules were produced and subsequently screened against the enzyme trypsin in an attempt to isolate compounds which specifically recognize the active site of the enzyme. Through a process of "iterative screening" in which active libraries were selected and smaller "sublibraries" were produced from the substituents of those active libraries, the libraries were "deconvoluted" to yield two final trypsin inhibitors with inhibition constants (K_is) of 9.4 and 72 μM.

To show that for *any* given target the combinatorial process is able to generate compounds with high molecular recognition for that target, the process was repeated. A new and more varied set of libraries was created spanning several hundred thousand different molecules, and these were screened against the Klenow Fragment of the enzyme DNA polymerase I. Through the same process of deconvolution, inhibitors were found that inhibited DNA polymerization at K_is of 25 and 1.5 μM. The outlook for even more potent polymerase inhibitors is also described.

Thesis Supervisor: Dr. Julius Rebek, Jr.
          Title: Camille Dreyfus Professor of Chemistry

to

My Mother and Father

in whose academic footsteps I happily walk,

and by whose loving example I am ever led

Acknowledgments

To begin, I must thank Debbie Kreutzer, because this Thesis simply would not have been completed without her help. Running the final biological assays for this work just days before my Defense and still finding time to edit and correct a full draft of the manuscript, Debbie has been not only a dedicated and gifted collaborator, but also a true friend. Her companionship through the latter three years of my MIT career has, very simply, made graduate life fun.

I turn next to the three "baymates" with whom I have had the privilege of sharing laboratory space during my stay in the Rebek Group. I arrived fresh out of college to be paired with Morgan "Mooseboy" Conn, and though I could not hope ever to rival his fastidious experimental technique or his elaborate system of filing, Morgan helped to turn me from a green "first year" into a full fledged graduate chemist. Morgan's general good character -- omitting several incidents of less than temperate demeanor -- made daily benchwork quite enjoyable. Our three and a half years together in 18-148 culminated in the collaborative effort of the diimide replicator detailed herein. Morgan was followed by fellow "fourth year" Yoko Kato, always amicable and ready to discuss a mechanism or NMR. Graduating a semester ahead of me, Yoko yielded her bench to postdoc Dmitry Rudkevich, beside whom it has been a pleasure to work. In addition to the wealth of chemical knowledge and the manifold stories of the former Soviet Union which he dispensed on a daily basis, in the darkest hours of this Thesis, Dmitry selflessly lent a hand at the keyboard when it seemed that the Experimental would never be compiled in time.

As much of the work in this Thesis was based on collaboration, a number of other chemists inside and outside the Rebek Group must be thanked here. First and foremost, I thank Dr. Thomas Carell for allowing me an ever expanding role in the combinatorial project upon which Professor Rebek and he first embarked. My work with Thomas opened for me a door into the growing field of Chemical Biology. Also on this project, we were fortunate enough to have the expert advice of Dr. Paul Vouros of Northeastern University and the excellent experimental contributions of his student Yuriy Dunayevskiy. Dr. Vouros is thanked for believing in the possibility of analyzing libraries of compounds through mass spectrometry when no one at MIT would, and Yuriy is thanked for realizing that goal. Furthermore, Yuriy went above and beyond the call of duty in helping to unravel the mystery of polymerase inhibition in Part II of this work.

The polymerase project itself was founded as a collaboration between our group and that of MIT Professor John Essigmann; he is thanked for his enthusiastic interest in combinatorial chemistry and for allowing Debbie Kreutzer to take a side track off from her own research. I also thank coworkers Robert Grotzfeld and Blake Hamann for donating several synthetic procedures in this endeavor. Fellow graduate student Jerry Shipps is a late addition to the combinatorial team; he was instrumental in synthesizing the last of the "xanthene mono acid" isomers for the polymerase project, and I now turn over to him my side of the three-way collaboration between Rebek, Essigmann, and Vouros.

On the self-replication project, I thank Dr. Belinda Tsao for her excellent collaboration, especially in the face of those who called into question the very

existence of replication. (On that note, I must thank Professor F. M. Menger of Emory University for giving us cause to publish two new papers on self-replication.) I thank Dr. Jong-In Hong for starting me out on self replication in the first place; it was through collaborating with him on the terphenyl replicator that I first learned the fine points of HPLC, a skill which has served me well in nearly every subsequent project.

To name only a few of the many other fine colleagues with whom I have had the opportunity to work in the Rebek Group, my thanks to Dr. Andy Sutherland on the combinatorial project, Ivan Huc on the replication project, Brendan O'Leary for NMR assistance, and Kent Pryor for his help in various moments of Macintosh hardware and software failure. It has been a great pleasure to work in an environment in which general camaraderie is maintained without sacrificing scientific rigor; I thank the members of the Rebek Group past and present and wish the current Group luck in their new labs at Scripps.

Scientific research does not exist in a vacuum, and I would like to thank my friends who gave me a life outside of MIT during my graduate career. Thanks especially to fellow J.E. roomies Larry, Dan, and Steve; Yale lives on. Thanks also to Jim Watt who took my tennis slot so that I could spend more time on my Thesis, and to my Tang roommates Eugene Bae, Andrew Barger, and Berlin Wu who helped provide a year of excellent food.

I cannot thank enough the members of my extended family for the support they have given me throughout my career and in fact my entire life; a warm thank-you to the Bunces, Calhouns, and Jaegers for constant encouragement, and my indebted gratitude to my Grandmothers Wintner and Calhoun *in memoriam*. I thank also B.J. Emmert, who was nearly family for my first 18 years of existence, for his friendship and the beginnings of a dream to make something special of my life.

To my brother Tom I give thanks for always having a close sibling with whom to talk and play sports; I wish him the best of luck in his own scientific endeavors at Harvard. To my parents, Claude and Martha, to whom I dedicate this work, there is no way to give the proper thanks, except perhaps to continue to make the most of their constant support and encouragement.

Finally, I wish to thank those people who have deeply influenced my career in chemistry: Mary Curtis of Belmont High School, Professor Jay Siegel of U.C.S.D., Dr. Jeff Carr of IBM's T.J. Watson Research Center, and Yale Professors Mike McBride, Harry Wasserman, and especially Alana Schepartz. At MIT, heartfelt thanks to Professor Dan Kemp for writing recommendations and Greg Fu for serving as my Thesis Committee Chairman with little advance notice.

I gratefully acknowledge the monetary support of the National Science Foundation in my first three graduate years, and the American Chemical Society, Organic Division for a Glaxo sponsored fellowship in my fourth year.

It remains only to thank Dr. Rebek, from whose lab all good things come. In the five years which I spent here at MIT, he provided an environment rich with the opportunity to learn and discover, an environment free from the pressures of an overbearing advisor, and an environment unique in its mission to apply the organic chemistry of molecular recognition to ever more complex problems of chemistry and biology.

Table of Contents

Part II.

Combinatorial Design of Molecular Recognition: Poly-Functionalized Core Molecules

Introduction

The field of molecular recognition in organic chemistry is the science of how molecules interact non-covalently with each other. Noncovalent forces lie at the heart of the "lock and key" mechanisms long recognized in nature by biologists, and as these mechanisms play a role in the regulation of nearly every living function, the understanding of molecular recognition has great consequence for science as a whole. In the motif of the lock and key, the single aspect of molecular recognition which defines and separates it from other facets of organic chemistry is its reversibility; the key can generally be removed from the lock. The forces which provide or deny a "fit" between two molecules include ionic interactions, hydrogen bonding, dipole-dipole interactions, Π-stacking, and van der Waals contacts.[1,2,3]

Ionic interactions, often termed "salt bridges" in the interaction of two organic molecules, represent the attractive force of two oppositely charged species. They can provide significant energetic stabilization of two molecules in contact, with the force of electrostatic interaction given by Coulomb's law and the dielectric constant D of the medium: $F = (q_1)(q_2) / (D{*}r^2)$.

Hydrogen bonding provides a force which is usually slightly weaker but can be highly specific and even directional. It denotes the energy gained when two atoms, most commonly nitrogen and/or oxygen, "share" the same proton. One atom, more closely bound to the hydrogen, is the proton-donor; the proton-acceptor has a partially negative charge which attracts the proton. A hydrogen bond, while far weaker than a covalent bond, nevertheless brings together the proton-donor and -acceptor closer than their van der Waals radii would normally allow. Hydrogen bonds, valued between -3 and -7 kcal/mol in the gas phase,[2] are thought to be able to

add -0.2 to -2 kcal/mol to the binding energy of two molecules in solution, depending on its polarity and dielectric constant.[4,5]

Dipole-dipole interactions, Π-stacking, and van der Waals contacts all result from the atomic proximity of two molecules close to the van der Waals radii of their atoms. At this range, an electrostatic force can arise between dipoles, or, as in the case of Π-stacking and van der Waals contacts, transiently induced oscillating dipoles. The net attractive effect of such contacts is generally between -0.3 and -1 kcal/mol in a given instance,[1,2] but a large interaction surface such as a long aromatic molecule may be worth several kcal/mol.[5]

The above forces are all individually irrelevant compared to the covalent bonds in molecules. However, if the shape of two molecules is such to position their charges, proton-donating and -accepting sites, and van der Waals surfaces in a complementary manner, the molecules will have a significant energetic affinity for each other.

With this knowledge in the hands of the organic chemist, the question then arises, how can complementarity between molecules -- molecular recognition -- be achieved through synthesis. In this Thesis, it is detailed how molecular recognition may be achieved through rational and combinatorial synthesis.

Part I.

Rational Design of Molecular Recognition: Abiotic Replicators

I.i An Introduction to Abiotic Replicating Systems

While the chemistry of molecules which reproduce themselves is by definition as old as life itself, the *synthetic* design of self-replicating systems is a relatively new field in chemistry. Work on molecules that produce copies of themselves began in the nineteen eighties and has of late become a significant facet of the broader field of molecular recognition. The Rebek group entered the field in 1989,[6,7] and ever since has sought to improve the ability of synthetic systems to replicate by rationally "engineering" the elements of molecular recognition involved. The excellent work of our many colleagues in the field, mentioned only briefly here, can best be read in their own words.[8-17]

I.i.1 Synthetic Replicating Systems in the Literature

Self-replication is one of the hallmarks of living organisms, and perhaps the ultimate goal in the study of self-replicating systems is to gain insight into the precepts and definitions of life. While the replication of life as we know it is now understood in a general molecular sense, it is clear that the DNA based cell is well evolved from what must have been the first living systems. One may speculate on the existence of very simple molecules or groups of molecules able to reproduce themselves, simple chemical cycles which could scarcely be termed "life." Modern

research into self-replication seeks to emulate such phenomena -- if only in a synthetic, abiotic environment -- with the goal of understanding the principles that must have governed the molecular transition from chaotic solution to the pockets of decreasing entropy known as ordered life.

Self-replication can be generally defined as the *directed* reproduction of an original by that original. *Directed* implies that some exchange of information takes place to ensure production of a copy of the original as opposed to any other product, and this generally translates to replication via a complementary "template effect." The enhancement of reactions by complementary surfaces lies at the heart of many processes in the laboratory and in nature. The most relevant biological template effect was revealed in Watson and Crick's structure of double-stranded DNA. It was clear to them that during replication, one strand of DNA acted as a template for the other. This feature led to the inspired experiments of Leslie Orgel and his coworkers at the Salk Institute,[11] and there, in 1986, Gunther von Kiedrowski showed that a short, self-complementary segment of DNA could act as a template for its own formation, even without the aid of enzymes.[8]

Complementary trideoxynucleotides CCG and CGG were coupled in the presence of a water-soluble condensation agent to form the self-complementary hexadeoxynucleotide by an autocatalytic process, thus creating the first abiotic self-replicating system. Subsequent improvements were made in the efficiency of autocatalysis; greater efficiency led to parabolic growth in the hexadeoxynucleotide concentration, reflecting the exponential nature of an autocatalytic process.[9] Several other abiotic self-replicating systems have since been described in the literature, most notably the reversible nucleic acid replicators of David Lynn[16] and T. R. Kelly's "bisubstrate reaction systems."[15] More generic template effects, leading to reduced activation entropies or stabilized intermediates, are responsible for the autocatalysis observed in a host of other processes.[17]

As shown above, molecular replication is simply an autocatalytic reaction where the product of a chemical transformation acts to catalyze that transformation through the *directed* production of copies of the molecule. We distinguish here between self-replication and other forms of autocatalysis; while there are many examples of the latter--the bromination of acetone and the formose reaction[18] being among the oldest examples--self-replication is a special subset of autocatalytic reactions in which molecular recognition plays a role. It must be noted, however, that self-complementarity is not *necessary* for self replication; other types of autocatalytic systems exist in which more general physical entities are created.

The replication of physical structures has been extensively explored by Luisi,[13,14] who has observed autocatalytic generation of micelles or reverse-micelles in both aqueous and organic media. The recognition event in these types of systems is the preferential binding of a substrate to the micelle whereby the exposure of the starting material to reagents, which are typically biphasic, is enhanced. The autocatalytic product in these systems, the micelle, is an aggregate of the reaction products with a loosely defined size and structure. This system differs from template based replication in that the latter is more strictly defined in its requirements of a complementary fit and stoichiometry in the recognition event. The difference is sufficiently great that "self-reproduction" has been proposed as the term for the behavior of the micelle-like systems.

I.i.2   The Emergence of Self-Replicating Systems in the Rebek Laboratory

While the most obvious (and glamorous) reason for exploring the world of self-replicating molecules is to search for insight into the origins of life, work in the Rebek group did not begin as an investigation into the so-called primordial soup of

prebiotic earth. Instead, our replicators grew out of a more general interest in molecular recognition of nucleic acids. Early on, it was discovered that derivatives of Kemp's triacid 1[19] (Figure 1) could be employed to create molecules with the ability to recognize certain substrates; the U-turn inherent in the Kemp's triacid moiety allowed the convergence of multiple functionalities at a single location.[20] In 1987, this U-turn was utilized to bring both hydrogen bonding and aryl stacking interactions to bear on the nucleic acid adenosine. Series of molecules such as **2a-2d** in Figure 1 allowed the evaluation of the energetics of aryl stacking versus hydrogen bonding in the complexation of 9-ethyl-adenine **3**.[5] From here, replicating molecules were just an idea away.
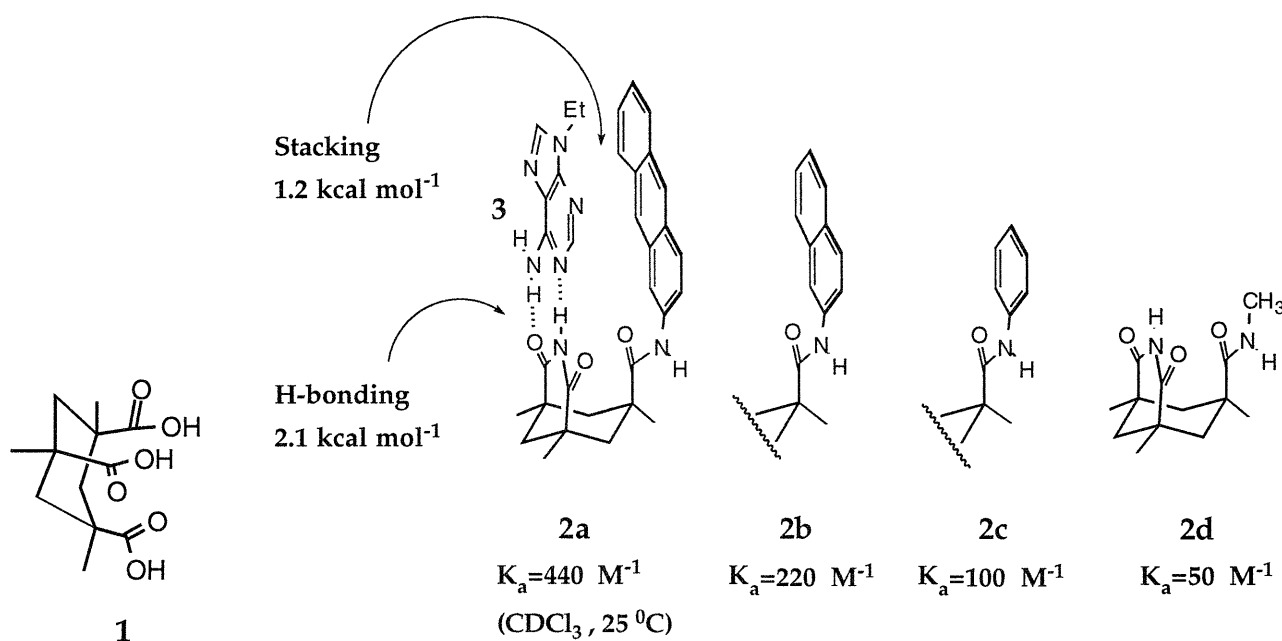

Figure 1. Energetics of aryl stacking and hydrogen bonding in the complexation of 9-ethyl-adenine 3.[13]

The key idea was that of self-complementarity: given a self-complementary molecule, one has in theory only to break the structure into two parts to produce a self-replicating system; alternatively stated, linking two complementary molecules in an appropriate way can give rise to a self-complementary, replicating structure.

In the schematic replicating system depicted in Figure 2, two complementary components (A) and (B) react in an intermolecular fashion to form template (T). Due to the self-complementary nature of the template, two additional units of (A) and (B) form a complex with the template (T:A:B). If the intermolecular forces of hydrogen bonding, aryl stacking, and van der Waals contacts are sufficient to anchor (A) and (B) on the template surface, an intracomplex reaction may take place to produce the dimer (T:T). Thus, the template produces a copy of itself and the molecules are called replicators. Once formed, the dimer may dissociate, and as long as components A and B remain, further replication of the template may proceed from the two new molecules of T.
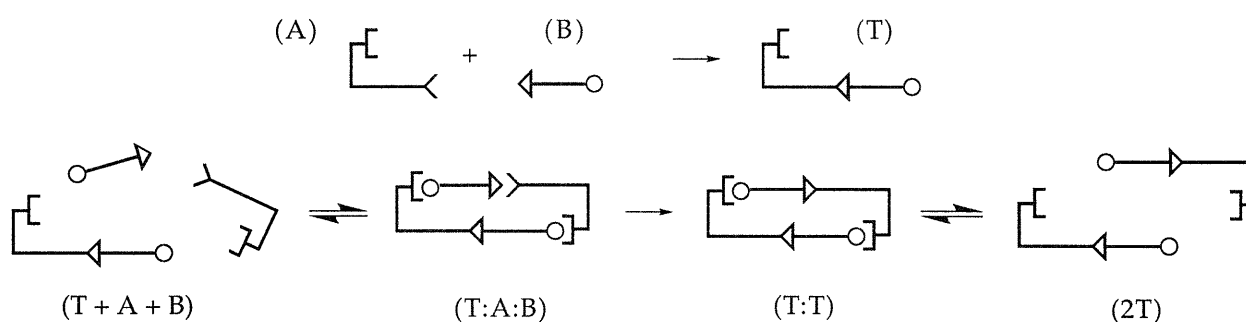

Figure 2. Schematic representation of self-complementary template based autocatalysis.

The viability of a replicator as pictured in Figure 2 hinges on the efficiency of the intracomplex reaction (T:A:B). If the reaction is more efficient than the background bimolecular reaction (A) plus (B), the system will show autocatalysis; the template (T) will act as a catalyst for its own formation. Efficiency within the complex (T:A:B) may arise from either of two sources. If the reaction has no intermediate, or formation of an intermediate is rate limiting, rate enhancement is derived from the reduction in entropy caused by bringing together the reagents on a

template. If breakdown of an intermediate is rate limiting (as is the case with our amide-forming replicators), rate enhancement is derived from template stabilization of an intermediate A-B such that product formation is favored over reversion to substrates.

I.i.3 The Naphthoyl System

As already stated, the first self-replicating system from the Rebek group evolved from work on the molecular recognition of adenine. By attaching an adenosine moiety to the aryl end of molecule **2b** in Figure 1, Tjama Tjivikua and James Nowick created self-complementary molecules such as **6**.[6,7] In the system, the Kemp's-imide-naphthoyl-pentafluorophenyl ester **5** reacts with 5'-amino-5'-deoxy-2',3'-isopropylidene-adenosine **4** to form the self-complementary autocatalytic template **6** (Figure 3).[6] The system was tested in chloroform with 1% triethylamine base at 21±1 °C, following the generation of template **6** by HPLC. The autocatalytic nature of the reaction was evident from the rate acceleration caused by seeding the reaction with its product (Figure 4).[7] At 8.2 mM initial concentrations of reactants, adding 20% of compound **6** produced an average 43% increase in initial rate of product formation, and adding 50% of compound **6** produced a 73% increase in initial rate. Addition of diacylaminopyridine **8**, a known hydrogen-bonding complement to imides,[21] *inhibited* the replication reaction (Figure 4),[7] showing the necessity of hydrogen bonding in the replication reaction.

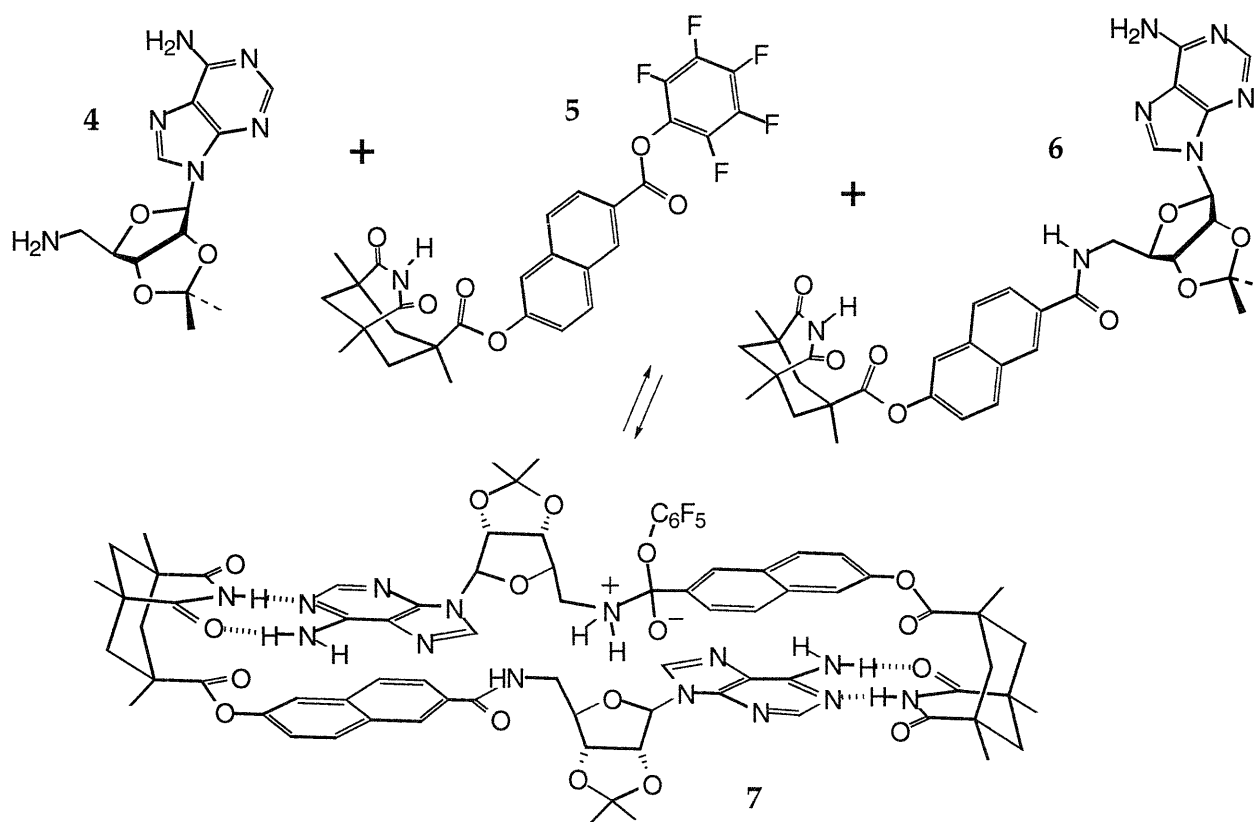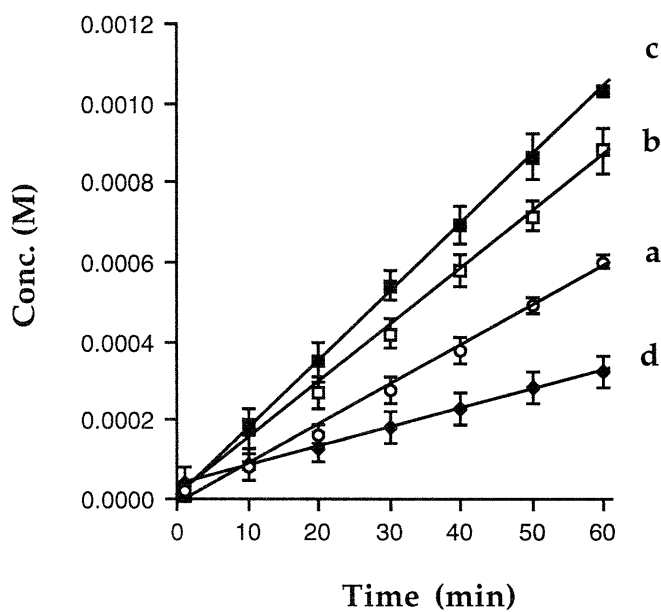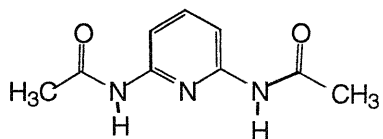Figure 3. An abiotic self-replicating system.

Figure 4.[7]   Concentration of product **6** formed in CHCl₃ as followed by HPLC, 21±1 °C.   8.2 mM initial concentrations of **4** and **5**, 1% TEA base added. a) Baseline reaction **4 + 5**;   b) Baseline reaction plus 0.2 equiv. product **6**; c) Baseline reaction plus 0.5 equiv. product **6**.   d) Baseline reaction plus 1.0 equiv. adenine binder **8**.

It should be noted that with varied amounts of added product, rate enhancements for this system are not directly proportional to product concentration, but rather to its square root. The theory behind this "square root law" was described by von Kiedrowski[8,9] to characterize nucleic acid replicators in which the autocatalytic entity exists largely in dimeric form. As shown in Figure 5 below, the rate-limiting step for template based autocatalysis is the irreversible formation of the dimer (T:T) from the complex (T:A:B), denoted by rate k. Added to the equation are the formation of (T:A:B) through equilibrium K_1, and the dissociation of the dimer through equilibrium K_2.

$$(T + A + B) \qquad (T:A:B) \qquad (T:T) \qquad (2T)$$
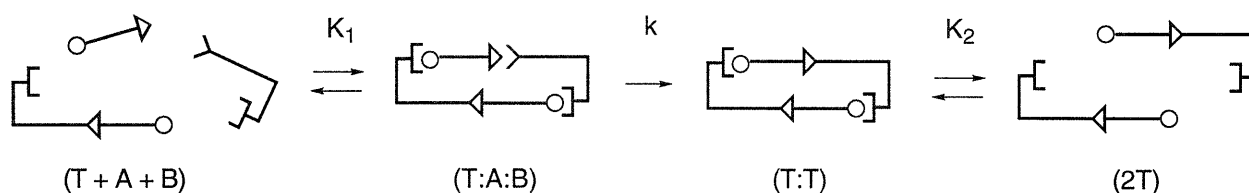
Figure 5. Equilibria K_1 and K_2 and rate constant k for template based autocatalysis.

K_2 may be expressed as:

Eqn. 1 $\qquad K_2 = [T:T] / [T]^2 \qquad$ or $\qquad [T] = [T:T]^{1/2} / [K_2]^{1/2}$

K_1 may be expressed as:

Eqn. 2 $\qquad K_1 = [T:A:B] / [A] [B] [T] \qquad$ or $\qquad [T:A:B] = K_1 [A] [B] [T]$

Substituting Eq. 1 for [T]:

Eqn. 3 $\qquad [T:A:B] = K_1*K_2^{-1/2} [A] [B] [T:T]^{1/2}$

The rate of formation of product (as the dimer) is by definition:

Eqn. 4 $\qquad d[T:T] / dt = k [T:A:B]$

Substituting Eq. 3 for [T:A:B]:

Eqn. 5 $\qquad d[T:T] / dt = k*K_1*K_2^{-1/2} [A] [B] [T:T]^{1/2}$

This is the rate of formation of dimeric template [T:T] in terms of concentrations of [A], [B] and template [T:T]. Thus, for the initial rate of template formation when [A] and [B] are nearly constant, and incorporating equilibrium constants K_1 and K_2, the apparent rate of formation of template reduces to:

Eqn. 6 $\qquad d[T:T] / dt = k_a [T:T]^{1/2}$

Thus, the rate of template formation is *dependent on the square root of template dimer present at the start of the reaction*. For a system in which the dimerization constant dictates that the template exists primarily in a dimeric state, [T:T] is simply half the concentration of "added template."

The naphthoyl replicating system follows the square root law nicely,[7] indicating that it is the monomeric form of the template which provides autocatalysis and that most of the template is dimerized in solution. As noted from Figure 4, addition of 0.2 and 0.5 equiv. of product **6** produced respectively 43% and 73% increases in the reaction rate of **4** plus **5**. As per the square root law:

$$43/73 = 0.59 \approx 0.63 = (0.2^{1/2})/(0.5^{1/2})$$

In general, the rate limiting step for ester aminolysis in aprotic solvents is the breakdown of the zwitterionic tetrahedral intermediate.[22-25] It is proposed that the autocatalysis observed in the naphthoyl system is the result of the product's ability to gather on its framework the two components of which it is formed and stabilize the tetrahedral intermediate thus created through noncovalent binding of its two ends (Figure 3, complex **7**). Lowering the activation energy of the intermediate subsequently lowers the energy of its transition state for release of pentafluorophenol, increasing the rate of amide formation. Furthermore, the steric "pincer" effect of the imide and adenine recognition sites may favor the elimination

of pentafluorophenol from the tetrahedral intermediate over reversion to substrates **4** and **5**.

I and my coworkers would later carry out a wide range of experiments in order to affirm that template effects were the true cause of catalysis in these reactions, and these results are detailed in Section **I.iii**. My first experiments in the Rebek group, however, were aimed at improving the original system.

I.ii   The Second Generation

As already stated, the key to a replicating system lies in the catalytic efficiency of the template, and ever since the creation of the first generation of replicator (Figure 3), it has been our goal to enhance the template process relative to the background reaction of amine and ester. As described below, this has been achieved by enhancing and adjusting the molecular recognition surfaces of the molecules involved.

I.ii.1   A Problem of Pathways

In the case of the naphthoyl derived system, there exist three background reactions of amine and ester. In addition to the pathway of random bimolecular addition, in both Hoogsteen and Watson-Crick binding modes, coupling between the amine and ester may occur within a *complex* of the *two precursors* (Figure 6). The initial product of this bimolecular pre-associative mechanism was postulated to be a *cis* amide, which isomerized to the *trans* amide, the active form of the template.[7]
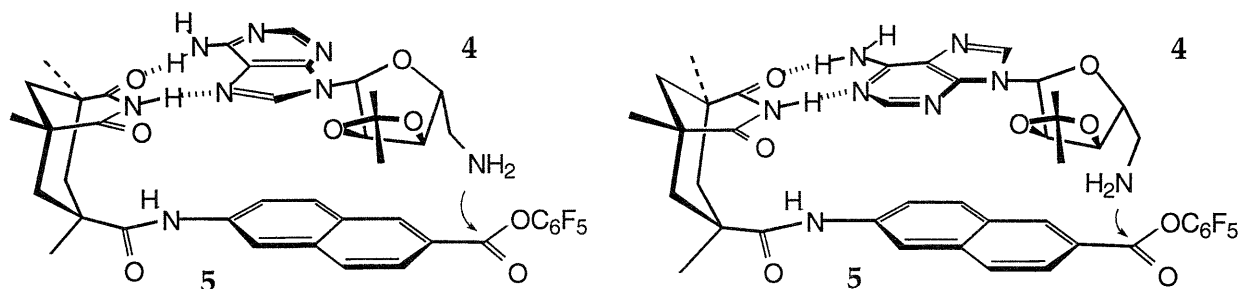
Figure 6. Hoogsteen- and Watson Crick-type intracomplex reactions in the naphthoyl

self-replicating system.

In fact, this appears to be the major background pathway for product formation. The magnitude of the bimolecular pre-associative effect is evident from the fact that at 8.2 mM, the initial rate of coupling of **4** plus **5** is 6.7 times faster than the initial rate of coupling of **4** plus **9**, in which the imide of **5** is N-methylated.[7] If no pre-associative pathway were active, **5** and **9** would react with **4** at the same initial rate, since a simple bimolecular reaction of ester and amine would not be dependent on the presence of a naphthylated imide. The pre-associative pathways operating in the naphthoyl system thus increase the rate of non-template catalyzed product formation, leading to a decrease in observed template efficiency relative to the background level of product formation.
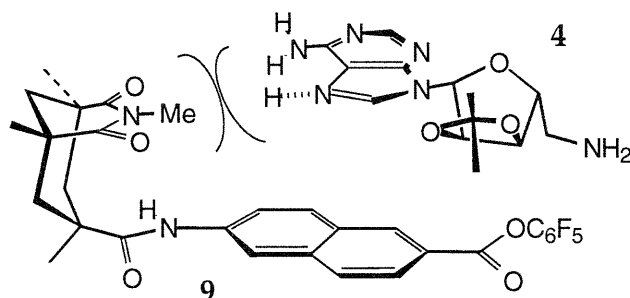

Figure 7. N-methylated PFP ester **9** has no pre-associative pathway available.

The work of Vince Rotello lengthened the spacer element in the system from a 2,6-naphthoyl to a 4,4'-biphenyl, partially solving the problem.[26] As shown in Figure 8, intracomplex reaction could now occur only if the adenosine were bound in the more-extended Watson-Crick mode. By reducing the amount of pre-associative catalysis, the effect of the template catalyzed pathway was enhanced: 20% of added biphenyl template increased the initial rate of coupling of **4** and **10** by 60%. Furthermore, the extended time course of the reaction revealed the gentle sigmoidal product growth curve expected of an efficient self-replicating system,[26] a feature which had already been observed for nucleic acid replicators.[9]
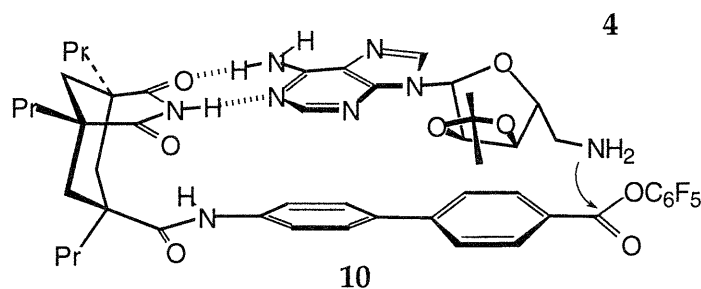
Figure 8. Intracomplex reaction in the biphenyl self-replicating system can only occur in Watson-Crick mode.

I.ii.2 The Terphenyl System

The next logical step was to remove the remaining pre-associative bimolecular pathway with a still longer spacer, and it was at this point that I joined the project, working with post-doc Jong-in Hong to synthesize and test the terphenyl derivative **11** (Scheme 1). Computer modeling showed that this spacer left only two reaction paths to formation of template: random bimolecular collision and template directed catalysis of the tetrahedral intermediate. As shown in Figure 9, no preassociative bimolecular path is available to the terphenyl molecule.

Figure 9. Intracomplex reaction is prohibited in the terphenyl system.

Scheme 1. Synthesis of the Terphenyl System

Synthesis of the terphenyl system was carried out starting from 4-amino-4″-carboxy-p-terphenyl **12**.[27] The benzyl ester **13** was formed through reaction with benzyl alcohol, triphenylphosphine, and diethylazodicarboxylate (DEAD) in THF. Subsequent to protection of the acid, the Kemp's triacid im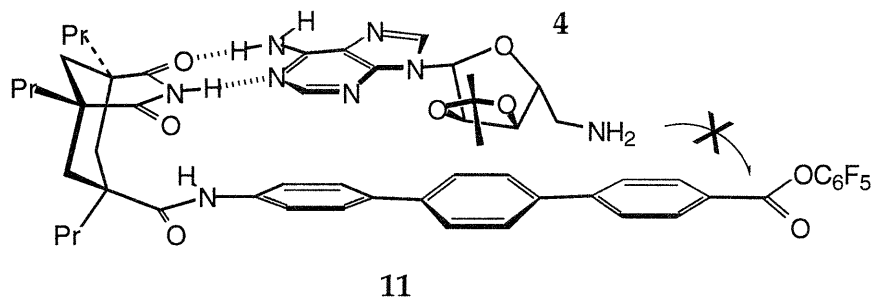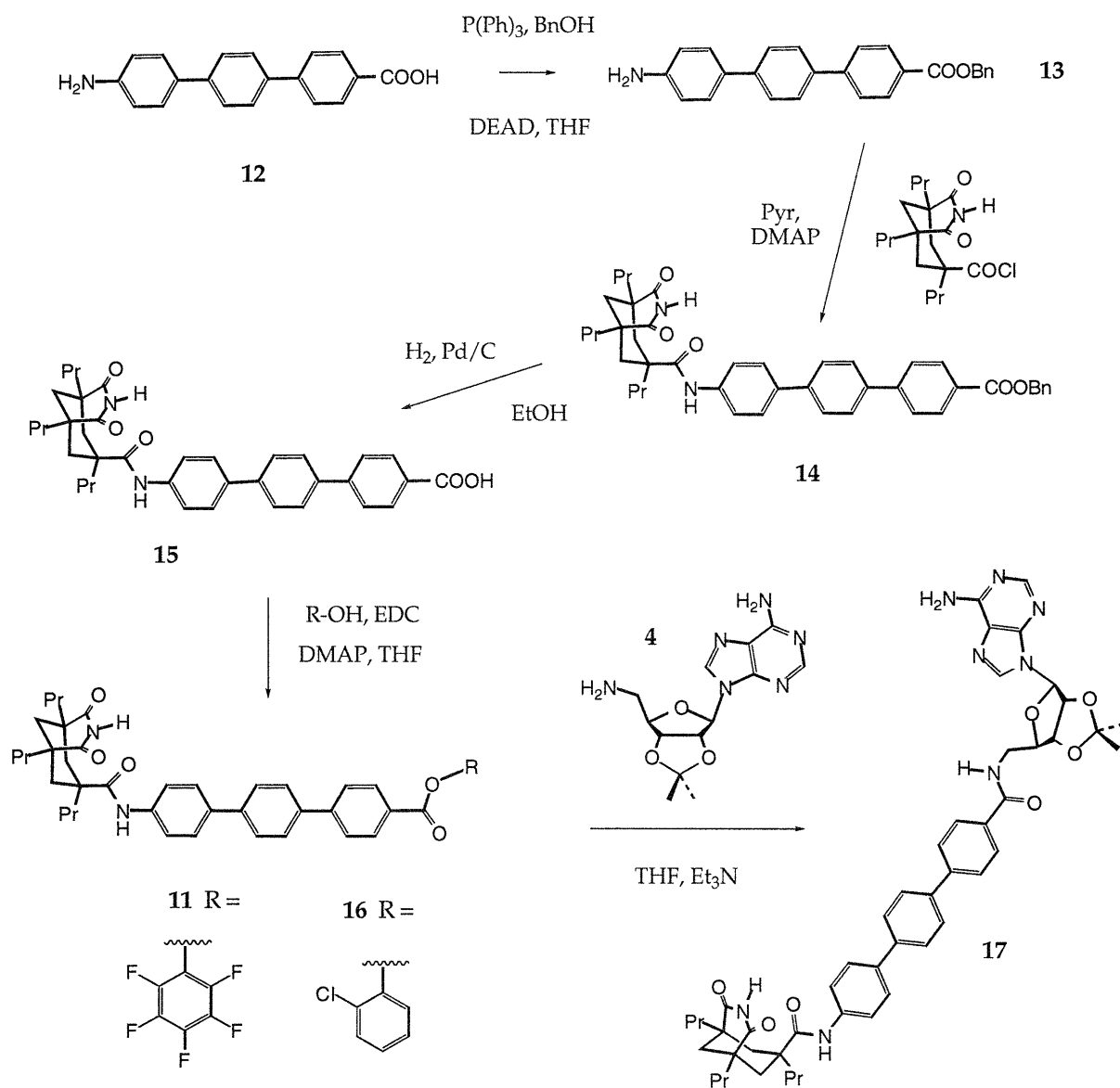ide moiety[5,19,26] was added (as the acid chloride) in pyridine with DMAP catalyst to give **14**. Deprotection (H$_2$, 10% Pd/C catalyst, EtOH) gave the acid **15**, which was the basis for formation of active esters **11** and **16** (1-ethyl-3-(3,3-dimethylaminopropyl)-carbodiimide, THF, cat. DMAP, pentafluorophenol/o-chlorophenol respectively). The template **17** was easily formed from **11** in THF with triethylamine.

Initial results were promising; as followed by HPLC, the rate of reaction of **4** plus **11** to form **17** was identical to the rate of reaction of **4** plus the N-methylated **18** (Figure 10). This indicated that indeed, the Kemp's imide played no role in the background reaction of **4** plus **11**; the pre-associative, bimolecular pathway had been shut down.
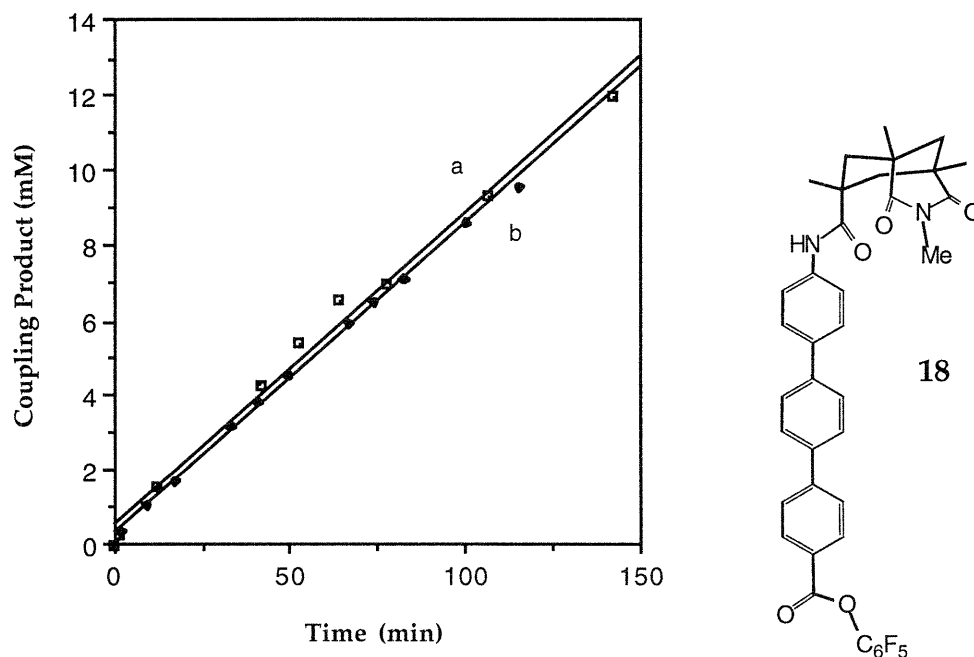

Figure 10. Plots of appearance of **17** and N-methylated **17** vs. time as determined by HPLC. Initial concentrations of **4**, **11** and **18** were 50 mM in CHCl$_3$ with 10 equiv. Et$_3$N added.

(a) Reaction of **4** and **18**.   (b) Reaction of **4** and **11**.

Unfortunately, addition of the terphenyl spacer took away not only the pre-associative bimolecular pathway, but also much of the efficiency of the replicative pathway. Template **17** was tested as a catalyst for adenosine **4** plus pentafluorophenyl ester **11**, but even at substrate concentrations of 50 mM, 0.35 equivalents of template led to only a 20% initial rate enhancement (Figure 11). The system was also tested with the less reactive *o*-chloro ester **16** over a period of 24 hr. to detect any sigmoidality in the formation of product **17**, but no upward curvature of the graph was seen.
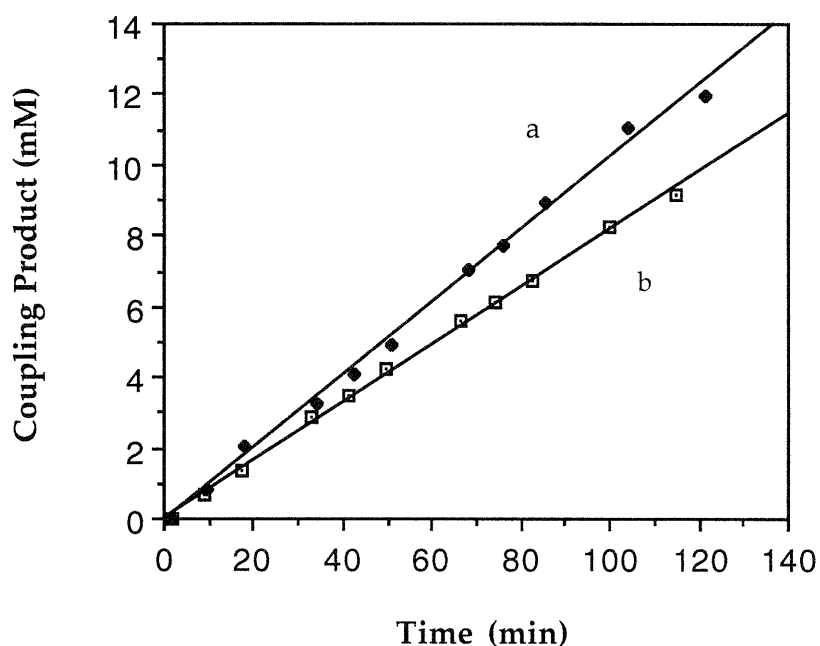
Figure 11. Plots of initial product formation of **17** vs. time as determined by HPLC. Initial concentrations of **4** and **11** were 50 mM in $CHCl_3$ with 10 eq. Et_3N added. (a) Reaction of **4** and **11** with 0.35 equiv. of template **17** added. (b) Reaction of **4** and **11** without added template.

Thus, the terphenyl system was a poor replicator, and binding studies revealed a possible reason: ^{1}H NMR titrations of the terphenyl molecule **14** with 5'-acetyl-2',3'-isopropylideneadenosine show a K_a for the complex of ~50 M^{-1}, (Figure 12) compared to a K_a of ~280 M^{-1} for the same titration of biphenyl imide

10.[21] Apparently, addition of a third phenyl ring creates an unfavorable geometry for the binding of the adenosine derivatives in question. While the theoretical design of the system was proven to be valid -- no pre-associative bimolecular pathways were active in coupling assays -- more practical concerns had clearly been overlooked.
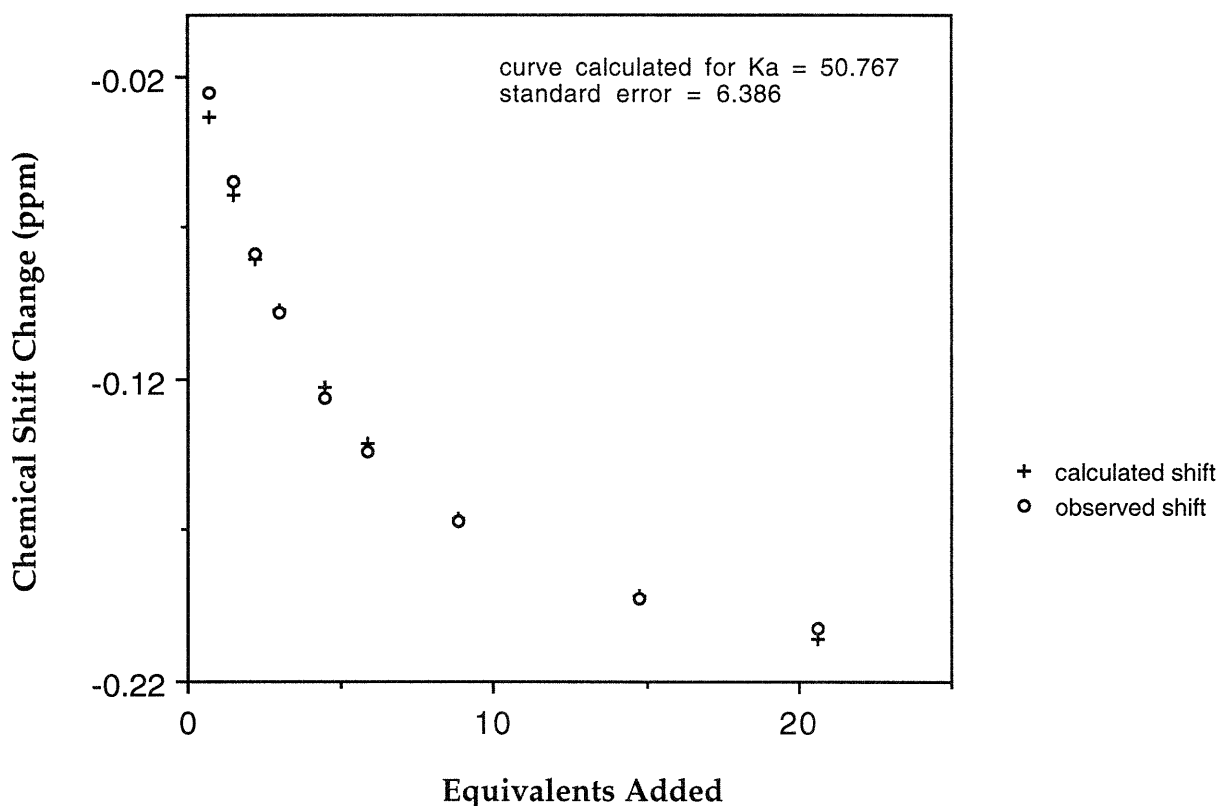
Figure 12. ^{1}H NMR titration of ester 14 with 5'-acetyl-2',3'-isopropylideneadenosine.

I.ii.3   The Second Generation Emerges

What was needed was a new adenosine binding structure that could hold its substrate fixed even in an elongated molecule. The spacer element in the receptor had to be of sufficient length to keep the amine and ester groups from reacting in a bimolecular complex, yet unlike the terphenyl system, the affinity of the

components for each other had to be high. These criteria were fulfilled by using a diaminocarbazole based diimide module developed by fellow graduate student Morgan Conn in his efforts at molecular recognition of polynucleotides;[28] structure **19** (Figure 13) had proven to be a nearly ideal complement to the purine nucleus of adenine.[29,30] The imides in molecule **19** chelate the purine through simultaneous Watson-Crick and Hoogsteen base-pairing, and the extended heterocyclic surface of the carbazole stacks against the purine. Not only is the binding affinity for adenosine derivatives extremely high ($K_a \sim 10^5$ M^{-1} in CDCl$_3$), but with triplex-like chelation of the adenine moiety,[30] the Watson-Crick/Hoogsteen conformational ambiguity present in the previous mono-imide naphthoyl and biphenyl systems is eliminated. Only one conformation -- in which both hydrogen bonding modes are satisfied simultaneously -- is available to complex **19**. Variable-temperature ^{1}H NMR spectra for a 5 mM solution (CDCl$_3$) showed two sharp imide peaks centered at 13 ppm at -55 °C; both Watson-Crick and Hoogsteen bound imides can be simultaneously observed when exchange is slowed.[30]
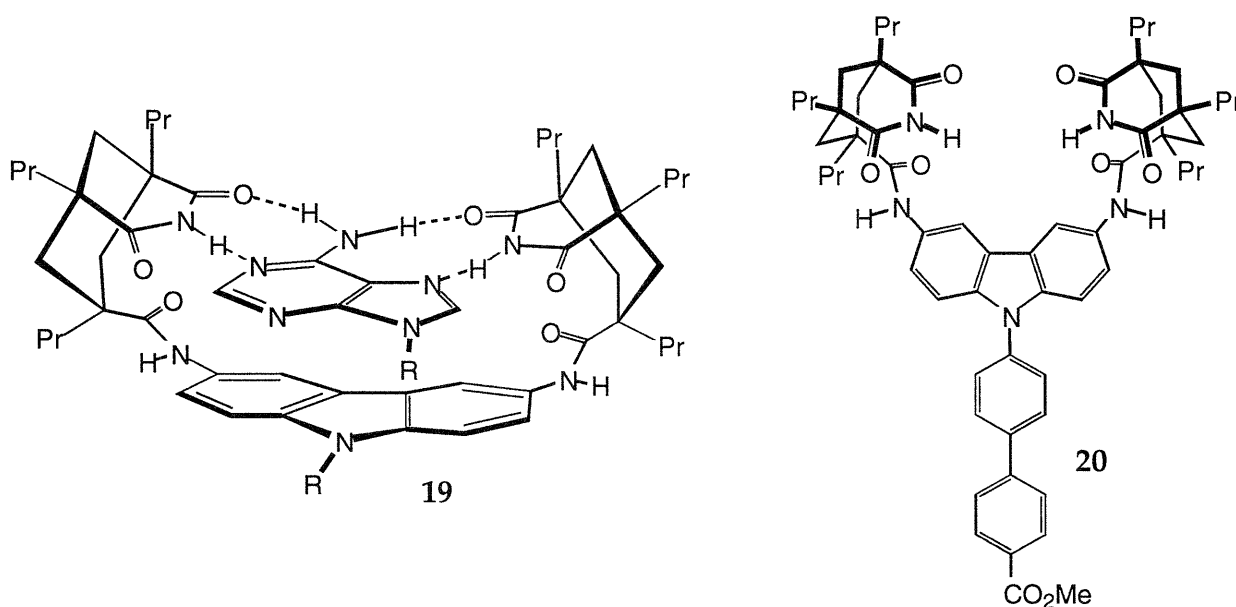
Figure 13. The carbazole-diimide backbone and its elongation with a biphenyl spacer.

A collaboration with Morgan was begun, and a second generation of self replicating molecules was born by the addition of a biphenyl substituent to N9 of the carbazole-diimide cleft,[31,32] shown as the methyl ester in structure **20**. Figure 14 shows the computer-predicted geometry of a bimolecular complex between **20** and amino adenosine **4**,[33] and from the figure it is clear that the amine and ester are separated by a significant distance, >5.5 Å in the model. This distance is well defined, as the diimide-bound adenine has only limited motion within the complex. Since the two reactive centers cannot approach each other within the complex, a bimolecular pre-associative pathway is ruled out. By addressing distance considerations, separation of ester and bound amine was achieved just as in the case of the terphenyl molecule (Figure 9), yet the diimide function of the new carbazole structure retained the ability to bind adenosine tightly.
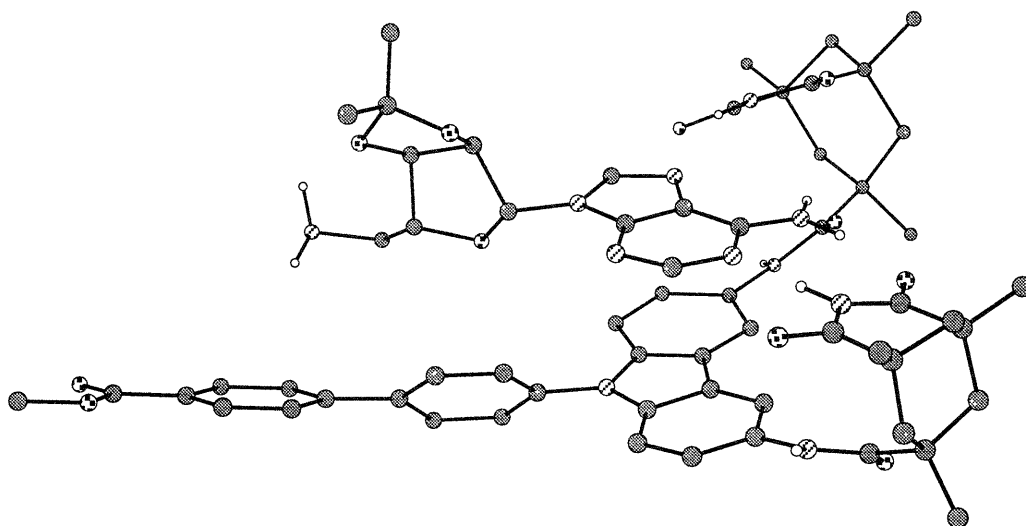
Figure 14. Computer-generated complex[33] between the biphenylcarbazole **20** and amino adenosine **4**.

The new self-replicating system, pictured in Figure 15, thus achieved precise positioning of amino adenosine and a pentafluorophenyl ester, and reaction of **4 + 21** to form the self-complementary template molecule **23** had to occur either in an unassisted intermolecular fashion or through a template catalyzed complex **23·21·4** (discussed below and modeled in Figure 18, *q.v.*).

21 R = pentafluorophenyl
22 R = 2,4-dinitrophenyl
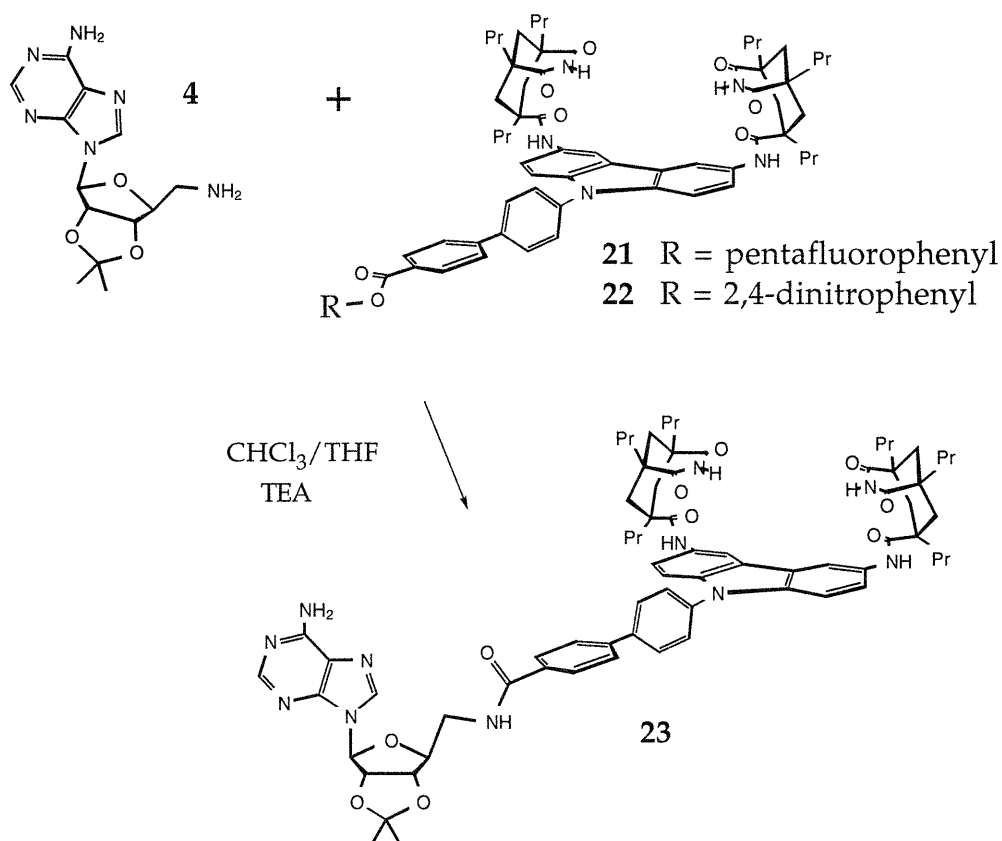
Figure 15. A diimide based replicator

The synthesis of the biphenyl diimide backbone was carried out primarily by Morgan.[31b] Synthesis began with 4'-iodobiphenyl-4-carboxylic acid and proceeded through Ullmann coupling[34] of methyl 4'-iodobiphenyl-4-carboxylate and carbazole. Nitration to the dinitro-carbazole followed by reduction (H_2/Pd-C, THF) gave the diamine, and subsequent condensation with the tripropyl derivative of Kemp's

triacid as the imide acid chloride[35] yielded the methyl ester **20.** Demethylation of **20** was achieved by S_N2 dealkylation with thiolate anion.[36] I added the final activated pentafluorophenyl and 2,4-dinitrophenyl esters by EDC coupling of acid and excess alcohol in THF (catalytic DMAP) to give **21** and **22,** and prepared the template **23** through coupling of amino-adenosine **4** in THF with Et_3N. I then ran all kinetic studies for the system.

Unlike previous systems in the group, kinetic studies of the coupling reaction of amine **4** with pentafluorophenyl ester **21** had to be performed in $CHCl_3/THF$ mixtures. THF was added to the reaction protocol in order to increase the rate of amide formation; for practical determination of initial rates by HPLC, it was desired to observe at least the first 5% of the total reaction within 100 min. Compared to the naphthoyl system, the background rate of which was greatly enhanced by pre-associative bimolecular effects as discussed, the background rate of formation of amide **23** was quite slow in $CHCl_3$, requiring 400 min. to attain 5% completion at similar reactant concentrations. (Compare 5% completion in 50 min. for the naphthoyl case, as shown in Figure 4.)

I.ii.4   Replication of the Second Generation

Once a suitable solvent system had been chosen, the reaction of **4** plus **21** to produce **23** was monitored by HPLC, and the system was found to be autocatalytic. At 6.2 mM in 13% $THF/CHCl_3$, 50% added **23** increased coupling of **4** plus **21** by an average of 53% (Figure 16, Table 1).[31,32] This was evidence of significant termolecular autocatalysis in a system where all other pathways to product formation -- except for the inevitable bimolecular background reaction -- could be excluded.

To prove that no other pathways were active in the carbazole-diimide replicator, controls were undertaken to assure that autocatalysis was due to molecular recognition and not to trivial chemical catalysis by some functionality of the template. The results are summarized in Figure 16 and Table 1.
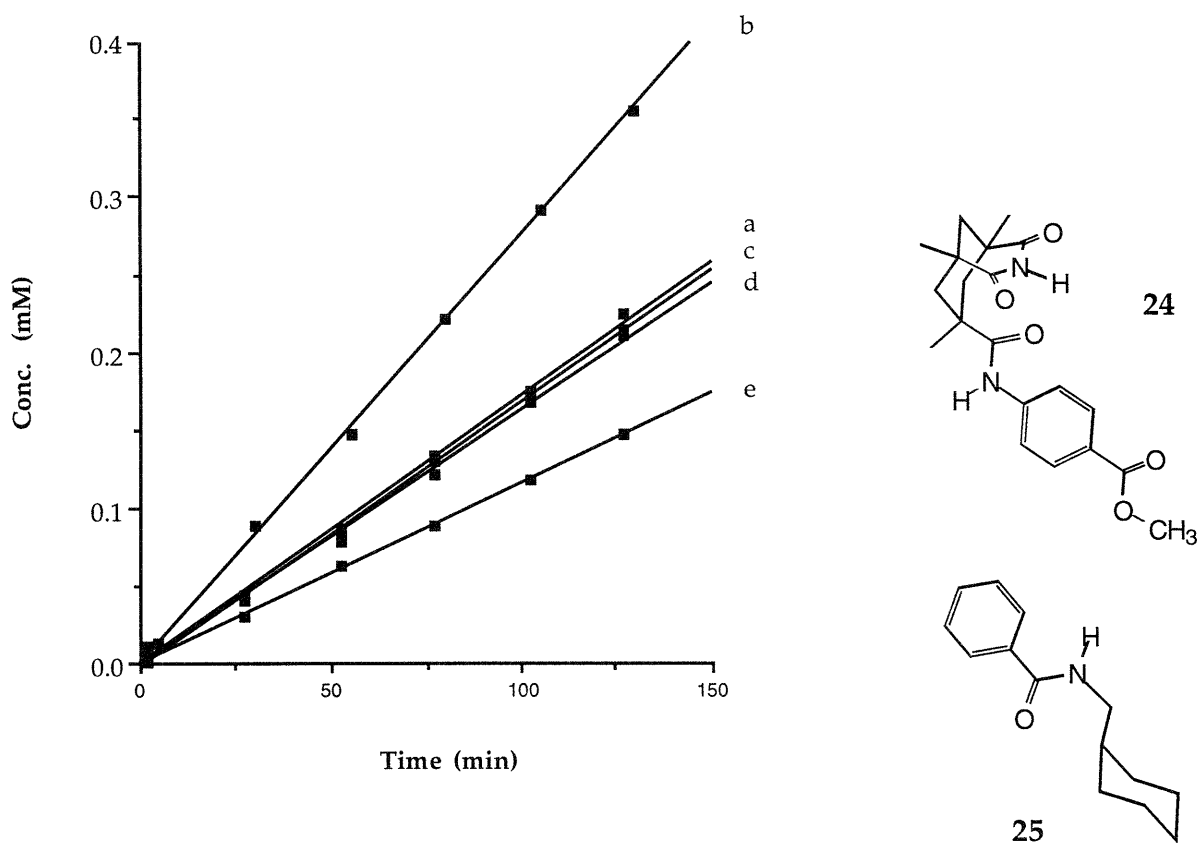
Figure 16.   Representative kinetic plots of the generation of product **23** as a function of time (initial 5% of reaction).   All reactions were performed at 6.2 mM initial concentrations of reactants **4** and **21** in 13% THF/CHCl3 with 1.0% TEA base added, 22±1 °C.   All individual slopes (reaction rates) are given in Table 4.   a) Baseline reaction (**4** + **21**); b) Baseline reaction plus 0.5 equiv. product **23**;   c) Baseline reaction plus 0.5 equiv. imide methyl ester **24**;   d) Baseline reaction plus 1.0 equiv. amide **25**;   e) Baseline reaction plus 0.5 equiv. diimide methyl ester **20**.

Table 1. Generation of product **23** as a function of time. All reactions were performed at 6.2 mM initial concentrations of reactants **4** and **21** in 13% $CHCl_3$/THF with 1.0% TEA base added, 22±1 °C.

Reaction	Additive	Individual Initial Rates of Product Formation (µM/min.)	Avg. Initial Rate of Product Formation (µM/min.)	Relative Rate
a	nothing	1.72 1.77 1.63 1.69 1.74 1.63 1.80 1.72	1.71 ± 0.06	1.00
b	product **23** (0.5 equiv.)	2.75 2.60 2.49 2.68	2.63 ± 0.11	1.54 ± 0.08
c	imide **24** (0.5 equiv.)	1.73 1.71 1.70 1.75	1.72 ± 0.02	1.01 ± 0.04
d	amide **25** (1.0 equiv.)	1.44 1.58 1.60 1.63	1.56 ± 0.08	0.91 ± 0.06
e	diimide **20** (0.5 equiv.)	1.26 1.05 1.15 1.27	1.18 ± 0.10	0.69 ± 0.06

The experiments revealed that the diimide function alone is not the source of the autocatalysis. This was established by control experiments with the diimide-methyl-ester **20** (Figure 13). This molecule did not catalyze the reaction of **21** with **4**; rather, inhibition resulted, probably as a consequence of its sequestering the amino adenosine in an unproductive complex. Additional experiments with **24** (Figure 16), which competes only poorly for adenosines, also supported the conclusion that an imide is an ineffective catalyst. This control molecule neither hindered nor enhanced the coupling of **4** with **21**.

Controls further showed that the phenyl amide portion of **23** does not catalyze the reaction. The coupling rate of **21** with **4** was not increased by the addition of the benzoyl derivative **25** (Figure 16), which presents a secondary amide function in a steric environment similar to that of **23**, but lacks recognition elements.

Finally, the control experiment performed in Figure 17 showed that neither the purine nor the ribose were effective as catalysts for an acylation reaction, and confirmed the result above that under these conditions, autocatalysis is not a general feature of ester aminolysis. The reaction of amine **4** with naphthoyl ester **26** (Figure 17) was followed by HPLC, and showed within experimental error that the amide product **27** -- which contains purine, ribose, and amide functionalities nearly identical to **23** -- did not significantly catalyze its own formation (Table 2). Interestingly, complexation of **4** with the diimide methyl ester **20** inhibited the reaction of **4** with **26** just as it did the reaction of **4** with **21**. This inhibition confirms the ability of the diimide **20** to sequester **4** in an unproductive complex.
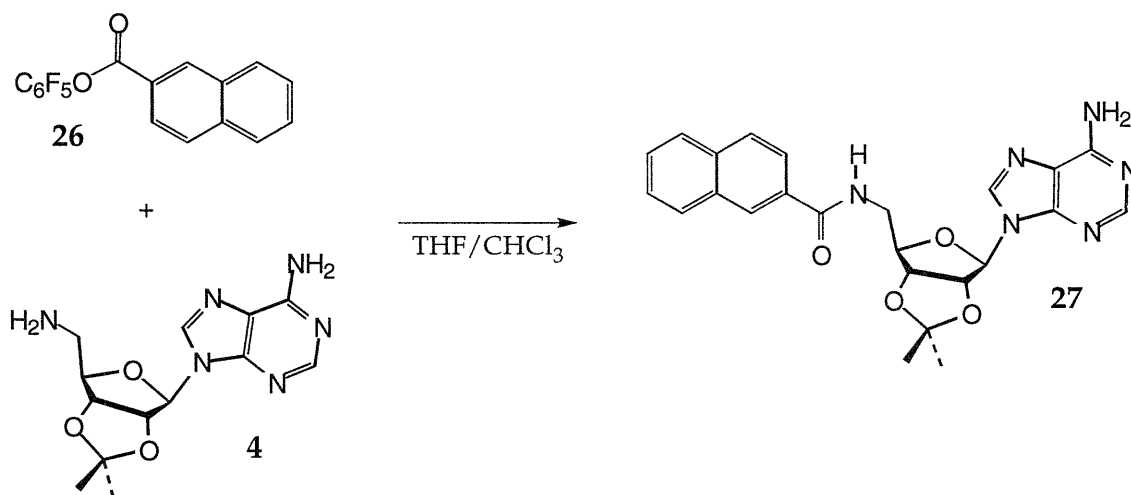
Figure 17. A non-replicating system.

Table 2. Generation of product **27**as a function of time. All reactions were performed at 6.2 mM initial concentrations of reactants **4** and **26** in 13% CHCl$_3$/THF with 1.0% TEA base added, 22±1 °C.

Additive	Individual Initial Rates of Product Formation (μM/min.)	Avg. Initial Rate of Product Formation (μM/min.)	Relative Rate
nothing	3.76 3.84 3.73	3.78 ± 0.06	1.00
amide **27**(0.5 equiv.)	4.11 4.05 4.07	4.08 ± 0.03	1.08 ± 0.02
diimide **20** (1 equiv.)	2.32 2.64 2.85	2.60 ± 0.27	0.69 ± 0.07

The above experiments all pointed toward self-replication. Separated, the individual features and functionalities of **23** are unable to account for the autocatalysis observed; the effect of the whole molecule is greater than the sum of its parts. The autocatalytic nature of **23** is most easily explained by postulating that the molecule serves as a template for its own replication: the initial reaction to form **23** is relatively slow, but once present, it can form a productive complex **23·21·4** stabilized by complementary recognition surfaces. Within the complex, hydrogen bonding and aryl stacking hold the tetrahedral intermediate in place, lowering the energy of the intermediate and thereby also lowering the energy of the transition state to the amide. The proposed tetrahedral intermediate for the complex **23·21·4** is modeled[33] in Figure 18.
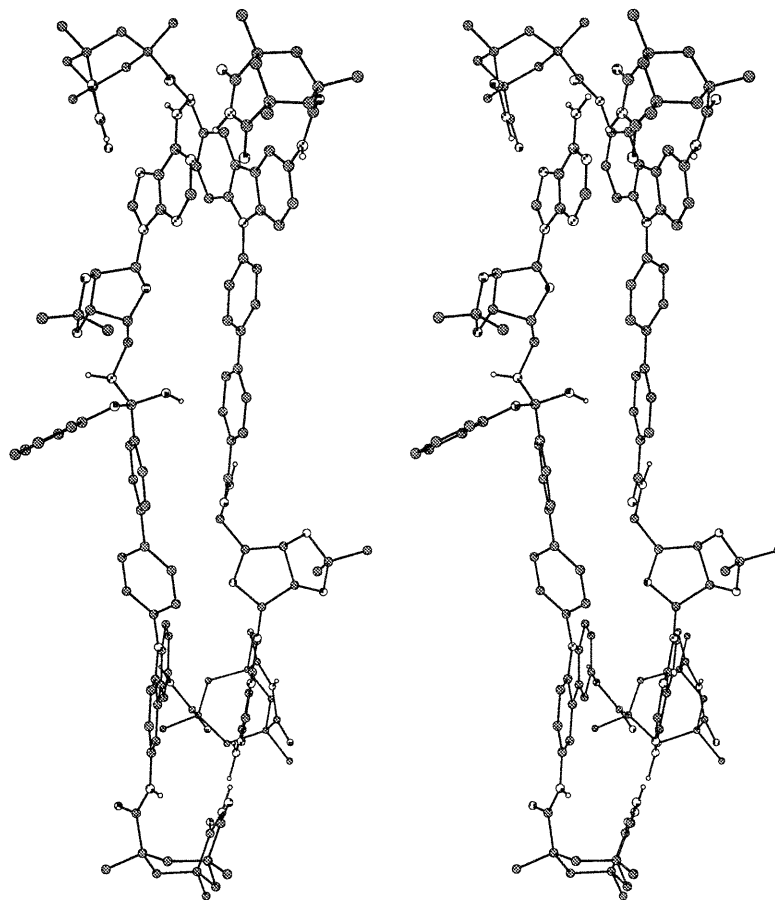
Figure 18. Computer generated stereo view[33] of the tetrahedral intermediate for the complex **23·21·4**, modeled as the neutral tautomer. Hydrogens attached to carbon have been omitted for clarity.

Thus, the molecular recognition incorporated into the biphenylcarbazole-diimide replicator achieved what the terphenyl system could not: the second generation showed efficient autocatalytic activity with an elongated structure which suppressed any pre-associative bimolecular pathways to acylation. Rational design of the adenine recognition site was shown by kinetic replication results to have been successful, and ^{1}H NMR titrations carried out by Morgan demonstrated clearly the improvements which had been made over the terphenyl system.

While slow exchange of the adenosine complex led to highly broadened ^{1}H NMR spectra during titration, the binding constant in $CDCl_3$ between the diimide-methyl-ester **20** (Figure 13) and 2',3'-isopropylidene-adenosine could be estimated at 10^5 M^{-1}. In THF-d_8, the association constant between the molecules was measured to be 576 M^{-1} with a downfield shift of the imide proton from 9.65 ppm to 12.87 ppm. With an association constant on the order of 10^5 M^{-1} in chloroform and 576 M^{-1} in THF, the binding of adenine by a biphenyl diimide moiety in 13% THF/CHCl3 was estimated at 86,000 M^{-1}. This indicates a far more favorable energy of binding than does the K_a of 50 M^{-1} seen for the terphenyl mono-imide.

In fact, it is postulated that the diimide replicator is actually *overly* capable of binding its complementary structures. In the absence of either negative or positive cooperativity, the dimerization interaction energy will be at least the sum of two individual association energies,[36] so the dimerization constant of template **23** in 13% THF/CHCl3 can be crudely estimated to be on the order of 10^9 M^{-1} (K_a^2). (The dimerization of **23** is so strong that dimerization occurs even in DMSO despite the highly competitive nature of this solvent for hydrogen bonding sites. A dimerization constant of 169 M^{-1} was calculated based on ^{1}H NMR dilution studies in DMSO-d_6)

With a dimerization constant of this magnitude, it can be estimated that in 13% THF/CHCl3 very little template (on the order of 1 μM)[38] is present as a monomer in solution; the system clearly suffers from severe product inhibition. That so little free template gave rise to a 54% increase in rate, however, indicates the success of **23** in positioning its substrates for reaction.

As titrations showed that the association constant between the two starting materials in the biphenyl diimide system is reduced from ~10^5 M^{-1} in CHCl3 to 576 M^{-1} in THF, an attempt was made to decrease dimerization of **23** by increasing solvent polarity, thereby enhancing the self-replicatory pathway. Thus, reactions

were attempted in 0 - 100% THF/CHCl3. Because the reaction was quite fast in THF and the initial rates of reaction had to be observed, experiments were conducted at 0.07 mM and followed by UV absorption. Using a UV spectrometer, data could be collected at intervals of seconds as compared to 15 minutes by HPLC.
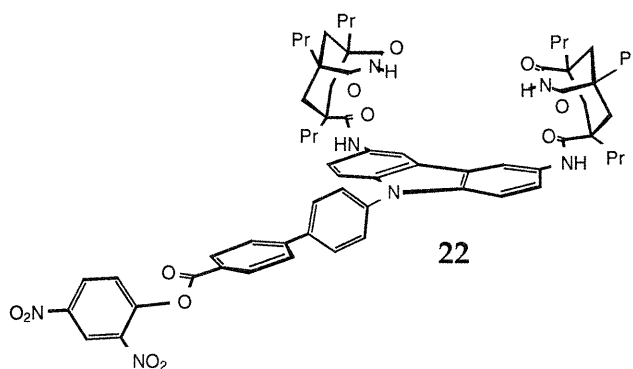
Figure 19. The 2,4-dinitro-biphenyl carbazole diimide used in UV/Vis kinetics studies.

Formation of **23** was followed in 0-100% THF/CHCl3, coupling **4** with the 2,4-dinitrophenyl ester **22** (Figure 19) and monitoring the release of 2,4-dinitrophenol at λ = 450nm. Reactions were then run again in the presence of 50% (0.035 mM) and 200% (0.14 mM) diimide template **23**, and these rates were compared to the uncatalyzed rate (Figure 20). Varying the percent makeup of the THF/CHCl3 solvent showed a catalytic peak for added **23** at ~15% THF, with a quick drop in catalysis thereafter. Thus, while NMR studies show that a more polar solvent does decrease dimerization of the template, higher polarity also enhances the rate of the background bimolecular reaction such that any gain in catalysis is "swamped out." One positive result of the experiment was to show that like the naphthyl system, the biphenyl diimide system correlates well with von Kiedrowski's square root law when differing amounts of template are added to catalyze the reaction (**I.i.2**). As seen from the data in Figure 20, the rate of product formation with 200% template

added is slightly more than double the rate of formation with 50% template added. The fact that theory is followed in the system lends support to the contention that most of the template is dimerized in solution and that the active catalytic species is the monomer **23**.

Figure 20. Catalytic enhancement of reaction rate in THF/CHCl$_3$ mixtures. Each bar shows rate enhancement of the reaction between **4** and the 2,4-dinitrophenyl ester **22** in the presence of 50% and 200% product **23**. The system was examined under dilute conditions (0.07 mM ester, 0.4 mM amine, 8 mM TEA) by monitoring the release of 2,4-dinitrophenol using UV/Vis spectroscopy.

I.ii.5 Speculation

Throughout our experiments with the diimide replicator, the slow background reaction of 4 plus 21 was noted, especially compared to the control reaction of 4 pl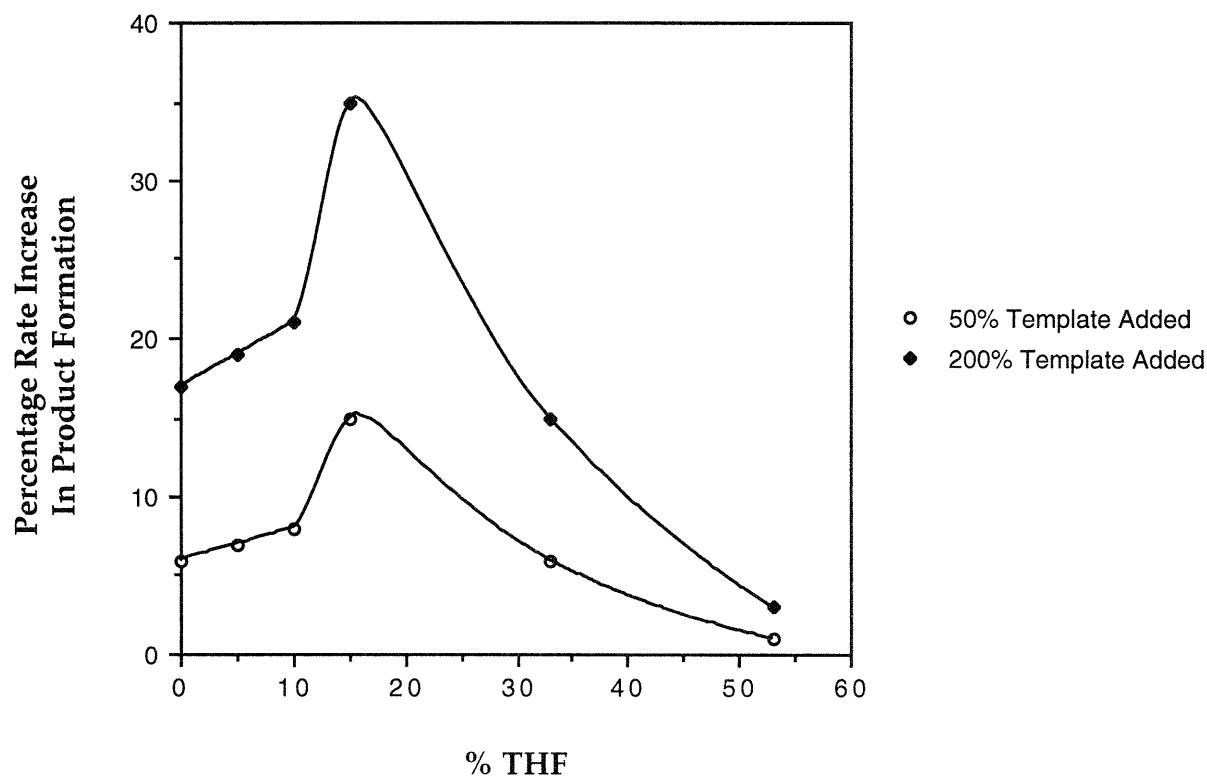us 26 in Figure 17 (which presumably gives a rate of acylation free of any complexation effects). The rate of formation of 23 (Table 4) compared with that of 27 (Table 5) under identical conditions revealed that recognition *slows* the rate of coupling by a factor of two; in the presence of 4, pentafluorophenyl ester species which are unable to complex 4 (*e.g.*, 26) are twice as reactive toward amines than is ester 21.

What are the consequences of complexation -- highly specific molecular recognition -- for replicators? Our results with the biphenyl diimide system suggest that a replicating species which binds tightly to its components will form copies of itself at a slower rate than those components would react free in solution. However, given that there may be any number of reactive components free in solution, the production of a specific combination of components may be greatly enhanced by a slow acting yet faithful replicative template.[39] For the same reasons that complexed components are slow to replicate, noncomplexed components are more exposed to side reactions. Structures that recognize each other and form complexes become stabilized, the surfaces in contact are protected from external, often harmful reagents, and the protected structures react primarily with molecules which are specifically complexed with them. Accordingly, the advantage of highly specific molecular recognition between components in evolution may have been less that it offered catalytic replication on an absolute kinetic scale and more that it offered catalytic replication compared to the sea of other reactions which those components might have undergone.

I.ii.6 Conclusion

The success of the biphenyl diimide system **4 + 21 → 23** in replication demonstrated the viability of molecular recognition through rational design. Through computer modeling of molecular distances and a knowledge of adenine's hydrogen bonding and aryl stacking properties, a problem of "engineering" in molecular recognition was solved: the specificity of binding necessary for replicatory autocatalysis was built into an elongated structure which eliminated the obscuring background pathway of a pre-associative bimolecular reaction. The highly specific complexation of the system not only achieved its goals, but also opened avenues of speculation on the importance of recognition in evolution.

I.iii   In Depth Evidence for Replication

Just as work was being completed on the "second generation" of replicators described in the previous section, a paper was published[40] that challenged the very basis of replication in the original replicating system characterized by Tjivikua and Nowick.[7] With the HPLC still warm from the kinetics of the biphenyl diimide system, I continued my collaboration with Morgan to re-investigate the original mono-imide naphthoyl system.

I.iii.1 Replication is Challenged in the Naphthoyl System

The naphthoyl based self-replicating system shown in Figure 21 was the first success of the Rebek group in the field of replication. In the two original papers detailing this work,[6,7] evidence for replication was essentially threefold:

1)   When seeded with its product, the system shows autocatalysis at 16.0 mM, 8.2 mM, and 2.2 mM concentrations of starting materials **4** and **5** in $CHCl_3$ at ambient temperature. As noted in Section **I**, under conditions of 8.2 mM, addition of 0.2 equiv. of product **6** produced a 43% average increase in the reaction rate, and addition of 0.5 equiv. of compound **6** produced a 73% increase in initial rate.

2)   No catalysis is observed when the reaction is seeded with control molecule **28**; in this case the rate of reaction is identical to that of **4** + **5** with nothing added. This result indicates the requirement of hydrogen bonding by the imide of **6** for autocatalysis to occur, which in turn hints at the involvement of the imide in a catalytic replicative complex such as **7**.

3)   As discussed in **I.i.3**, the reaction follows the theoretical "square root law" with respect to product **6**, indicating that the catalytic species is the monomer and that in solution it exists mostly in dimeric form.
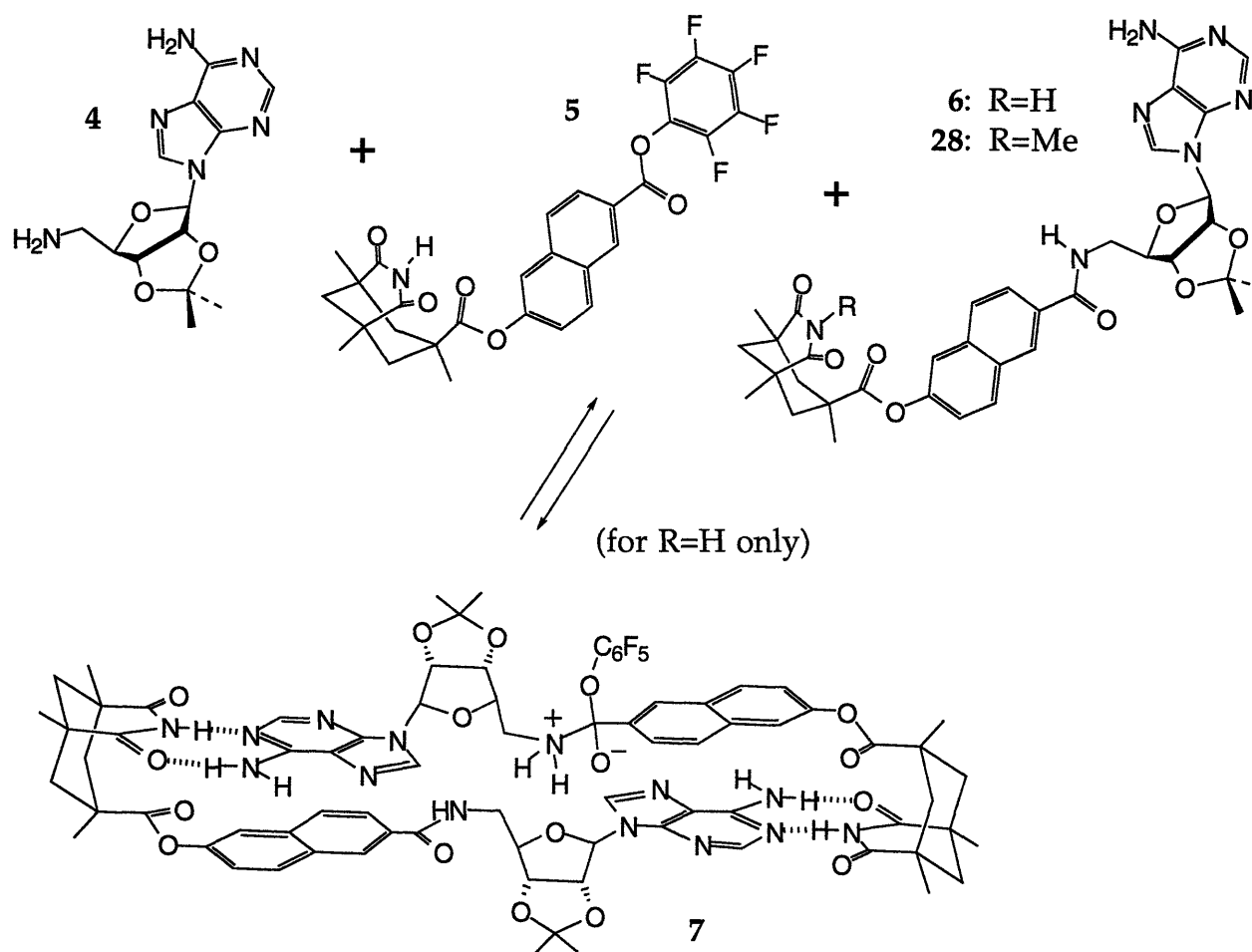
Figure 21. An abiotic self-replicating system.

While the above results clearly pointed toward autocatalytic replication as the cause of rate enhancement upon the addition of 6 to the reaction of 4 with 5, several other mechanisms of catalysis were not specifically excluded. Our detractors combined this lapse with their own investigation into the system to suggest that the rate enhancement seen in the system was not replication but rather mere chemical catalysis.[40,41]

They began by raising a question of "simple amide catalysis:"

[At substrate concentrations of 30 mM], "simple amides (*e.g.*, 2-naphthamide, acetamide) also catalyze the aminolysis of ester [5] by amine [4]. Since Rebek's template [6] is itself an amide, concern arose as to whether his catalysis might arise not from a template effect but instead from a more mundane amide acceleration." [40]

The mechanism postulated above is one in which the product 6 provides catalysis solely through its amide functionality as shown in complex 29 (Figure 22). The mechanism of 29 suggests that the hydrogen bonding properties of the imide and adenosine portions of 6 are superfluous to autocatalysis.
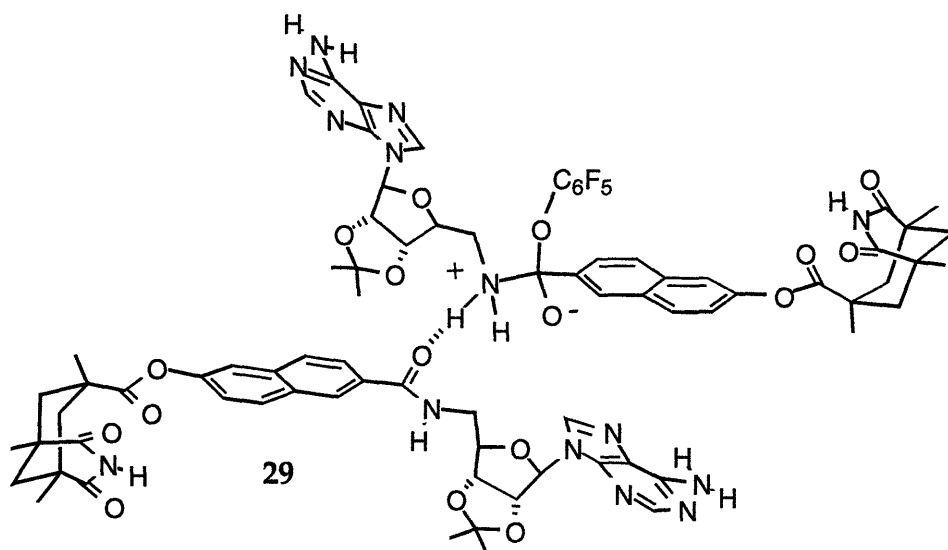
Figure 22. A proposed mechanism of amide catalysis.

Our detractors further postulated a second possible mechanism of catalysis in which the imide portion of **6** *does* bind to the adenosine of **4**, but in which actual catalysis is caused by the amide (complex **30**, Figure 23).[41] This hypothesis, which suggests that the hydrogen bonding properties of the adenosine portion of **6** are superfluous to autocatalysis, was based on their finding that the reaction of **4** with molecules **31** or **32** (Figure 23) is also catalyzed by **6**. The inference is based on the assumption that **32** and **33** serve as "non-hydrogen-bonding" analogs of ester **5**.
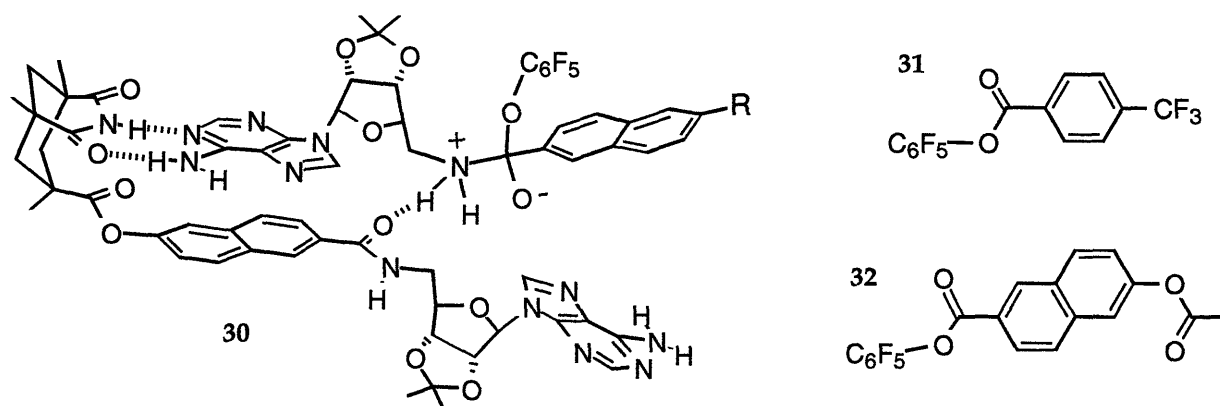
Figure 23. A proposed mechanism of amide catalysis with hydrogen bonding.

In contrast to the above mechanisms, we had asserted a mechanism based on complex **7** (Figure 21), in which both ends of the template hold their complementary substrates to stabilize the tetrahedral intermediate and subsequent transition state to dimeric **6**. In this model, stabilization of the transition state relies on molecular recognition *at both ends* in the form of hydrogen bonding and Π-stacking interactions, without participation of the amide.

All three of the above mechanisms make sense in that they direct catalysis at the breakdown of the zwitterionic tetrahedral intermediate -- the rate limiting step of ester aminolysis in aprotic solvents.[22-25] However, the means of catalysis in the first two depends upon the action of the amide of 6, while in the third case catalysis relies solely on the presence of hydrogen bonding and Π-stacking interactions -- molecular recognition. If it is the amide formed in the reaction of 4 plus 5 that is the catalyst for the further reaction of 4 and 5, the system is no more than another example of an autocatalytic chemical reaction; it is no more (or less) worthy of note than the autocatalytic bromination of acetone. If, on the other hand, it is the complementary shape and surface of the molecule formed by the reaction of 4 plus 5 which is the catalyst for the further reaction of 4 plus 5, then structure 6 is a template for *directed* reproduction of itself, and the reaction is an example of self-replication.

The challenge was thus put to us to determine which if any of the above three mechanisms are the cause of the autocatalysis observed in the naphthoyl system.

I.iii.2 Confirmation of Replication in the Naphthoyl System

In our control experiments for the diimide replicator (**I.ii.4**), we had previously found that in 13% THF/CHCl₃, a system of simple amide formation did not demonstrate significant autocatalytic behavior (Figure 17, Table 2). Furthermore, the fact that 28 shows no catalysis of the reaction of 4 + 5 argues that the amide moiety of 6 cannot alone be responsible for the autocatalysis observed. However, in light of the above claims of simple amide catalysis in our system, additional evidence against simple amide catalysis by product 6 (complex 29) was desired, and therefore amide formation was investigated in CHCl₃.

In collaboration with postdoctoral fellow Belinda Tsao, experiments were conducted in which ester **33** was coupled with amine **34** in $CHCl_3$.[42] Formation of product **35** could be easily followed by the appearance of amide methylene protons in ^{1}H NMR. The question of autocatalysis in the system was examined by adding amide **36** to the system; **36** is identical to the product of the reaction of **33** + **34**, except that the amide protons are deuterated, allowing formation of **35** to be monitored exclusively even in the presence of added **36**. As shown in Figure 24 and Table 3, even at 20 mM concentrations of pentafluorophenyl ester **33** and benzylamine **34**, amide **36** was not a significant catalyst of amide formation.
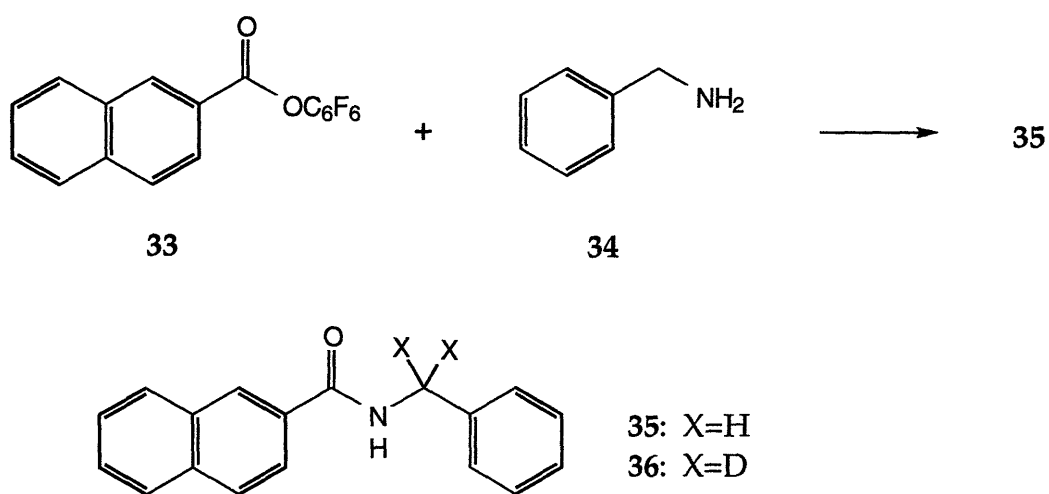
Figure 24. Molecules for amide formation control experiments.

Table 3.[42] Amide formation control experiments at 25 °C as followed by NMR. Coupling of **33** and **34** in CDCl₃ with or without addition of amide **36**. Initial velocities of reaction were determined through integration of the methylene peak of the product amide **35** at 4.72 ppm relative to the methylene peak of **34** at 3.88 ppm.

Conc. of Ester **33** and Amine **34** (mM)	Equiv. Amide **36**	Avg. Initial Rate of Formation of **35** (μM/min.)	Relative Rate
4	0	42	1
	0.5	42	1.00
8	0	84	1
	0.5	84	1.00
16	0	168	1
	0.5	174	1.04
20	0	258	1
	0.5	282	1.09

The data in Table 3 is strong evidence against a simple mechanism of amide catalysis as proposed in Figure 22. Since the amide of molecules **35** and **36** is quite similar to that of **6** in its immediate chemical environment, and yet shows no catalysis of ester aminolysis, it seems unlikely that **6** could act as suggested in complex **29**. In a further control experiment, addition of secondary amide **37** to the reaction of **4** plus **5** failed to provide any enhancement of the rate of formation of product (Table 4, Entry 3), more evidence that it is not the amide moiety of **6** which leads to autocatalysis in the system.
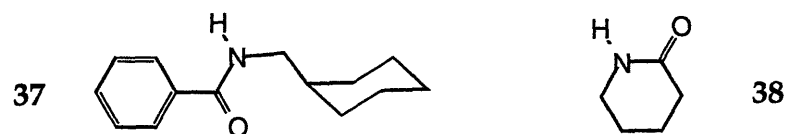
Figure 25. Control additives for the reaction of 4 + 5.

Table 4.[43] Effect of various additives on the formation of 6 in $CHCl_3$ as followed by HPLC. 2.2 mM initial concentrations of 4 and 5, 22±1 °C, 1% TEA base added.

Entry	Additive (0.5 Equiv.)	Avg. Initial Rate of Product Formation (µM/min.), ±5%	Percent of Baseline Rate
1	--	0.54	--
2	6	0.81	150%
3	37	0.52	96%
4	38	0.63	117%

Nevertheless, the fact remained that our detractors had found catalysis of 4 plus 5 with amides such as 2-naphthamide and acetamide. This apparent discrepancy was answered by adding valerolactam 38 as a control in the reaction of 4 with 5.[43] The action of *bifunctional* catalysts in acylation reactions and glucose mutarotation is well known,[17] and indeed, it was found that the *cis* amide valerolactam increased the initial rate of formation of 6 by 17% when added to the reaction of 4 and 5 (Table 4, Entry 4). Figure 26 depicts a possible catalytic role for valerolactam in the breakdown of the tetrahedral intermediate.[22] The acidic and basic sites on valerolactam can stabilize the zwitterionic tetrahedral intermediate and facilitate the required proton transfers for product release, ultimately regenerating the catalyst.
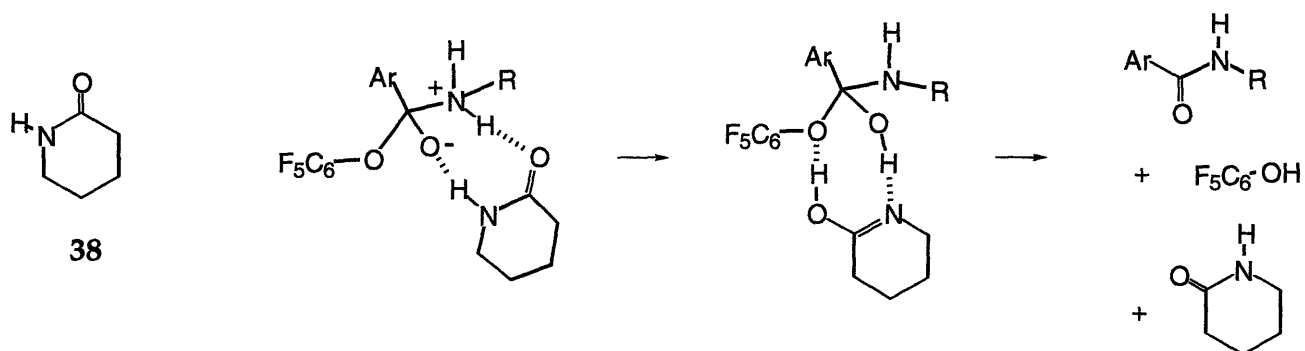
Figure 26. A possible role for valerolactam in the catalysis of ester aminolysis.[17]

Primary amides share the above bifunctional capability with valerolactam: their acidic and basic sites, being on the same edge of the molecule, can act in concert. Probably by the same mechanism as shown for valerolactam in Figure 26, primary amides such as 2-naphthamide and acetamide are able to catalyze the reaction of 4 plus 5.[40] *Trans* amides, however, such as those in molecules 6 or 36, have no such feature. The fact that primary amides enhance the reaction of 4 + 5 by operating through a bifunctional catalytic mechanism has no bearing on whether or not molecule 6 can be called a template for self replication.

While Su and Watson[24] showed that under certain conditions, small *trans* amides (even tertiary amides such as dimethyl acetamide) can hydrogen bond to the tetrahedral intermediates in related reactions and catalyze their breakdown to products, in light of the failure of secondary amides 37 and 36 to show catalysis in the reactions of 4 + 5 and 33 + 34, respectively, the *trans* amide of 6 can be ruled out as the sole contributor to the autocatalysis observed under our conditions. With complex 29 thus eliminated, evidence pointed to one of the two remaining mechanisms for the autocatalytic behavior of the naphthoyl system: either complex 30 or complex 7.

With the amide moiety of **6** accounted for, we next examined the remaining functional groups in the molecule for possible chemical catalysis. The imide, ribose, and purine functionalities all had to be considered as possible explanations for the autocatalysis observed; for example, imidazole is a well-known catalyst for acylation reactions, and the purine contains such a subunit. Using control molecules **3, 39,** and **28,** the potentially catalytic functions of the product molecule were tested *in the structural context of* **6,** and under the conditions where **6** acts as an autocatalyst.[43]

Table 5.[43] Effect of various additives on the formation of **6** in $CHCl_3$ as followed by HPLC. 2.2 mM initial concentrations of **4** and **5**, 22±1 °C, 1% TEA base added.

Entry	Additive (0.5 Equiv.)	Avg. Initial Rate of Product Formation (μM/min.), ±5%	Percent of Baseline Rate
1	--	0.54	--
2	**6**	0.81	150%
3	**3**	0.55	102%
4	**39**	0.56	104%
5	**28**	0.55	102%
6	**40**	0.50	93%
7	**41**	0.56	104%
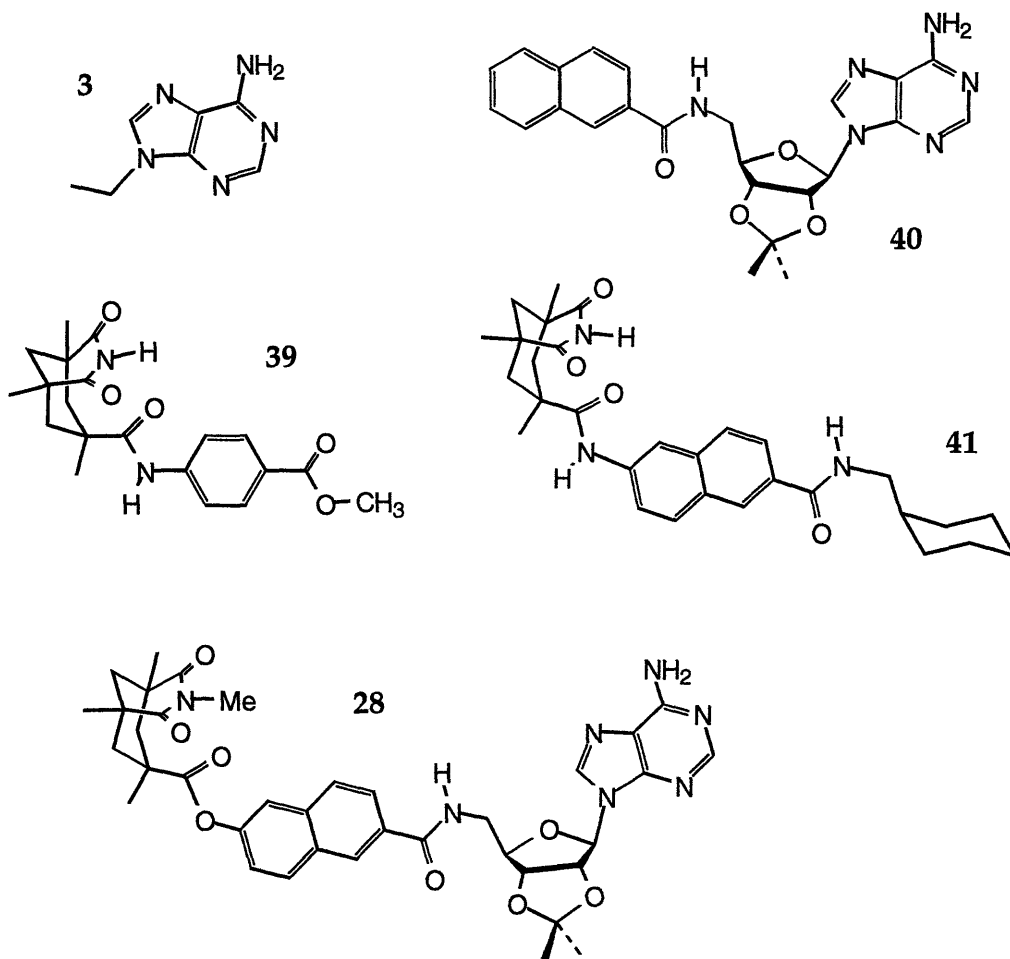
Figure 27. Control additives for the reaction of 4 + 5.

The question of purine catalysis was answered by the addition of 9-ethyl-adenine **3** (Table 5, Entry 3). The absence of catalysis in this experiment excluded the purine nucleus as the source of catalysis in Entry 2. The inability of control molecule **39** to catalyze the reaction allowed us to exclude the Kemp's-imide moiety of **6** as a source of direct chemical catalysis (Table 5, Entry 4).

A most telling control experiment involved the N-methylated imide **28**, confirming previous data under slightly different conditions.[6,7] In the presence of **28**, no rate enhancement in product formation was seen within the 5% experimental error (Table 1, Entry 5). The remarkable effect of this singular change -- substitution of a methyl group for the imide hydrogen of **6** -- excludes at once the

rest of the molecule's functionalities as independent sites of simple chemical catalysis. Since the imide itself had already been excluded as a chemical catalyst (control molecule **39**), this study pointed to base-pairing between the imide and adenine as necessary for catalysis by molecule **6**.

All of the data above, however, do not exclude a mechanism such as that of complex **30** in Figure 23. While mechanisms are ruled out in which either the imide, amide, or purine functions of **6** serve as *independent* sites of chemical catalysis, still at large are complexes in which the purine or imide serve in *combination* with the amide to position a substrate for catalysis of ester aminolysis. To prove that autocatalysis in the system was a result of complex **7**, a final pair of control molecules, **40** and **41**, was required.

Both molecules were tested as additives to the reaction of **4** plus **5**, but no enhancement of product formation was seen in either case (Table 5, Entries 6 and 7).[43b] The failure of either control molecule **40** or **41** to catalyze the reaction of **4** plus **5** excludes catalytic mechanisms **42** and **43** involving the amide and either one of the two base-pairing sites of **6** (Figure 28). Both complexes **42** and **43** propose that the amide chemically assists the breakdown of the tetrahedral intermediate, but as neither **40** nor **41** was a catalyst under these conditions, it was concluded that the full template **6** is necessary for autocatalysis. Merely positioning one of the two substrates (or one end of the tetrahedral intermediate) on the template backbone is not sufficient.
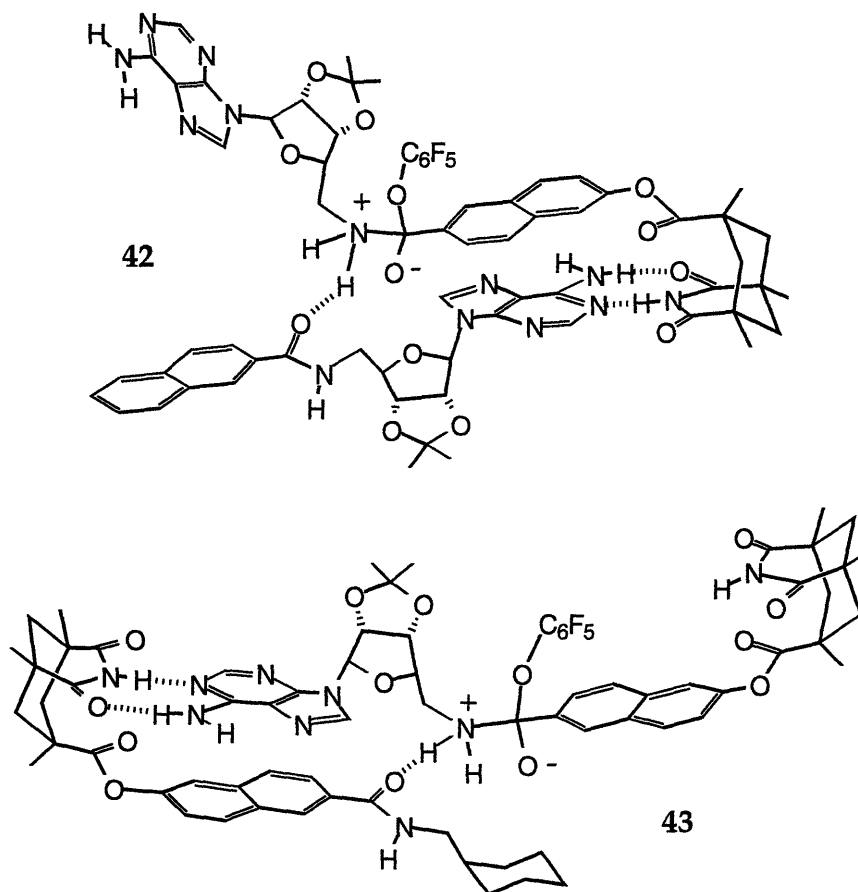
Figure 28. Two possible mechanisms of catalysis excluded by experiments with molecules 40 and 41.

The individual features and functionalities of 6 are, therefore, unable to account for the autocatalysis observed when 6 is added to the reaction of 4 + 5. Rather, just as in the case of the diimide replicator 23, the whole product molecule is more effective than the sum of its parts. The most economical explanation for these results is the mechanism of Figure 21, complex 7. Template catalyzed replication -- in which 6 binds 4 and 5, stabilizes the tetrahedral intermediate that forms, and favors its breakdown to an amide -- is the source of autocatalysis.

As in the case of excluding complex 29 above, in eliminating complex 30 we were nevertheless left without a good explanation for the results of our detractors: in this case, that molecule 6 could catalyze the reaction of 4 plus "non binding" esters 31 and 32 (Figure 23). Thus, the experiment shown in Figure 29 was

undertaken, investigating the effect of **6** on the coupling of **4** plus non-binding ester **44**. Cyclohexyl ester **44** has no ability to bind to the adenosine of molecule **6**, as it lacks any hydrogen bonding or Π-stacking surfaces (apart from the pentafluorophenyl function shared by all of the esters under consideration). Therefore, any catalysis by template **6** would support a mechanism of amide catalysis such as **30**.
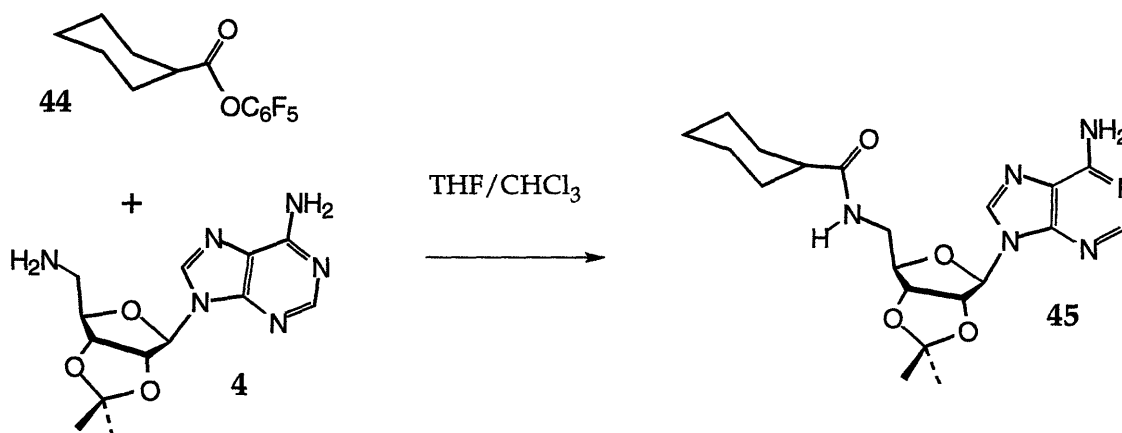
Figure 29. Control molecules used to test the proposed mechanism of complex **43**.

Table 6. Generation of product **45** as a function of time, as followed by HPLC. All reactions were performed at 2.0 mM initial concentrations of reactants **4** and **44** in CHCl$_3$ with 1.0% TEA base added, 22±1 °C.

Conc. of Ester **44** and Amine **4** (mM)	Equiv. Template **6**	Avg. Initial Rate of Formation of **45** (μM/min.)	Relative Rate
2.0	0	15.0	1
2.0	0.5	15.0	1.00
2.0	0.7	15.2	1.01

Following formation of **45** by HPLC at 2.0 mM, it was found that added **6** was unable to catalyze the ester aminolysis, of **4** + **44** (Table 6).[42] The experiment was repeated by NMR at 8 mM with 1 equiv. **6** added. Again, no catalysis by **6** was observed (Figure 30). The fact that **6** does not catalyze the coupling of **4** + **44** is further evidence against a mechanism such as complex **30** (Figure 23), and hints that the "non-binding" esters of Menger *et al.* (**31** and **32**) may not be devoid of recognition capabilities. Both **31** and **32** feature electron deficient aromatic surfaces, and Π-stacking of adenines with such surfaces can afford several kcal/mol in binding affinity under these conditions.[20] Structure **32** might further hydrogen bond through its acetyl carbonyl. In short, experiments with ester **44** indicate that **31** and **32** are ill-conceived control molecules for understanding the reaction in question: the coupling of **4** + **5** in the presence of **6**.
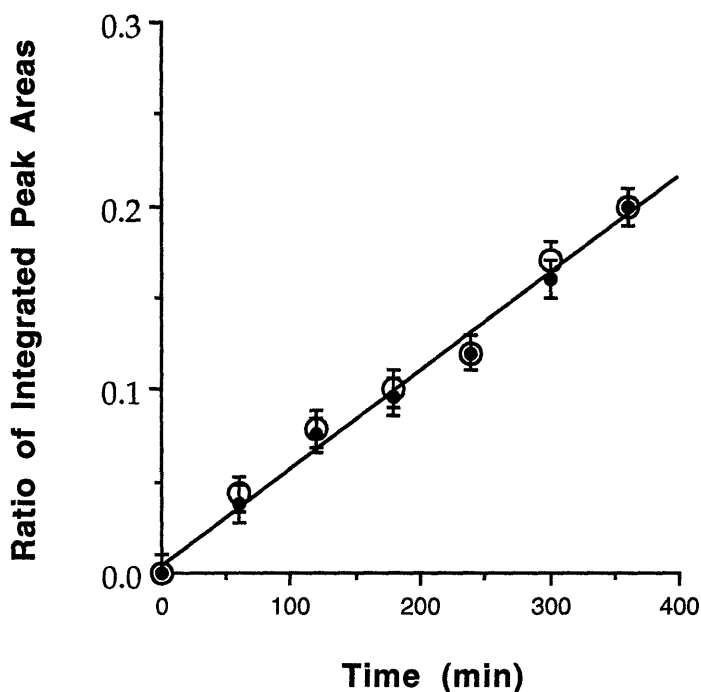
Figure 30. Graph of the ratio of product **45** to reactant amine **4** as a function of time, with (dark circles) or without (open circles) 1.0 equiv. added **6**. Reactions were performed at 8.0 mM initial concentrations of reactants **4** and **44** in $CHCl_3$ with 1.0% TEA base added, and followed by NMR at 25±0.3 °C. The rate of both reactions was 3.8 mM/min. Error bars reflect reproducibility.

The most plausible explanation, then, for the autocatalytic nature of **6** remains its ability to act as a template for its parts. The forces of hydrogen bonding and aryl stacking position the amine and active ester on the template's surface, stabilizing the tetrahedral intermediate formed (complex 7, Figure 21). Subsequent collapse to the amide bond is favored over reversion to a termolecular complex, and this results in an exact replica of the catalyst in the form of a template dimer. The reversible forces of molecular recognition which stabilize the dimer also permit its dissociation, and monomeric template is generated. Thus, the template, through specific noncovalent contacts, has produced a copy of itself, through the rational design of molecular recognition built into the structure, the naphthoyl system replicates.

This conclusion was recently corroborated by a third party; the contribution of David Reinhoudt *et al.* to the above question of self replication was in press at the time of this writing.[44] Personal communications, however allow us to state their finding that while "other pathways obscure the simple picture of a ternary complex [7] as the only complex that leads to rate enhancement," [their] "general conclusion is that self replication as defined by Rebek *et al.* operates in this system."

I.iii.3 Conclusion

The work above demonstrates that it is indeed possible to "engineer" molecular recognition in a rational manner. Beginning from a theoretical concept of complementarity between molecules and knowledge of the forces which enhance and detract from that complementarity, molecules were designed which at first simply recognized each other. Building upon this foundation, ever more complex molecules were synthesized by adding more and more function to the molecules, and the result was a system which -- depending on one's worldly outlook -- lies somewhere between a simple chemical reaction and the first stirrings of life.

Thus, the concept of molecular recognition through rational synthesis is borne out. The basic theoretical concepts of hydrogen bonding, aryl stacking, van der Waals forces, and their non-covalent brethren can be physically applied to achieve a desired goal. Of course, the results achieved in this way can only be as good as one's understanding of these basic concepts. There are many areas of molecular recognition, such as drug design, in which the problem to be solved (usually the molecular recognition of a large, three-dimensional protein surface) is *at present* somewhat beyond our theoretical calculation. In a case where theory is not sufficient to predict the effect of a given cause, a desired end may be achieved in one of two ways. The first is simple trial and error, a method which has a long but checkered history. The second is trial and error with a twist of rational input -- evolution of effect through the selection of ever more successful random permutations. The latter method forms the basis of progress in Nature, and also lies at the heart of combinatorial chemistry. It is this more chaotic side of chemistry which is the subject of Part II of this Thesis.

Part II.

Combinatorial Design of Molecular Recognition: Poly-Functionalized Core Molecules

II.i   An Introduction to Combinatorial Chemistry

In Part I, it was shown that molecular recognition could be synthetically manipulated in molecules, producing systems with desired capabilities (in this case, replication). Emphasis lay on the rational design of the molecules, Ångstrom by Ångstrom and hydrogen bond by hydrogen bond. As detailed above for the biphenyl diimide system, this process can be quite fruitful, yet it also has drawbacks: in computer modelling studies, the terphenyl system looked to have excellent self-replicatory characteristics, yet after much synthetic effort, the final product turned out to be ineffective as a catalyst. In short, the rational design of molecular recognition is far from a perfect science; it is only as good as one's foreknowledge of non-covalent properties in the compounds involved.

The recent emergence of the field of combinatorial chemistry has opened up a new world of synthesis which offers an alternative to rational design. Simply stated, combinatorial chemistry offers the chemist a means to select molecules with desired capabilities from a large pool of compounds, *without requiring any foreknowledge* of the properties of the molecules within that pool. In Part II of this thesis, the invention and application of a new process of combinatorial chemistry is detailed which allows molecular recognition to be focused on any desired target. In the case of biological targets, two examples of which we successfully investigated, this process becomes one of combinatorial drug design.

II.i.1   Combinatorial Chemistry as a New Tool

The screening of organic substances to discover lead compounds of pharmaceutical interest has traditionally involved the testing of individual compounds of high purity. Compounds are synthesized and assayed one at a time, a process which,

though highly successful, is labor intensive. The advent of techniques of combinatorial chemistry offers the means to generate a large number of different chemicals simultaneously, allowing rapid access to vast numbers of new chemical entities. Modern analytical methods allow the screening of such combinatorial "libraries" to select those species in the generated pool which possess desirable properties.[45]

The first attempts at combinatorial synthesis involved libraries of peptides on solid supports, and include the "pin-method" devised by Geysen et al.[46] and the "tea-bag" procedure invented in the Houghten group.[47] Both methods make use of the chemistry of Merrifield[48] and related peptide syntheses, which have been well optimized for solid supports. Geysen's peptides are synthesized on solid support pins which are dipped into reagents on 96 well microtiter plates; this allows peptides to be created in batches of 96 with a different peptide on each pin. Houghten's procedure encloses the solid support in "tea bags" which are then dipped into baths of reagents; several hundred tea bags may be operated on simultaneously, with each bag producing an individual pure compound.

The introduction of the "split-bead technology" of Furka et al.,[49,50] increased the parallel synthesis of peptides dramatically to several million compounds in a single library. In the synthesis, solid support polymer beads are coupled in groups to one of a set of amino acids as shown in Figure 31. All beads are then pooled and randomly split, and the new groups are again coupled to a single amino acid. After several rounds, all possible combinations of the given set of amino acids are created. Since a given bead only sees one of the amino acids at any given coupling stage, each bead is covered with a single pure peptide sequence, and the beads themselves can be individually assayed. As a result of the number of peptides which can be made in this way, the limiting factor in the successful screening of peptide libraries has become the characterization of active species; the amount of any individual compound present in a library of millions is too minute for its structure to be determined by conventional spectrometric methods.[51,52]
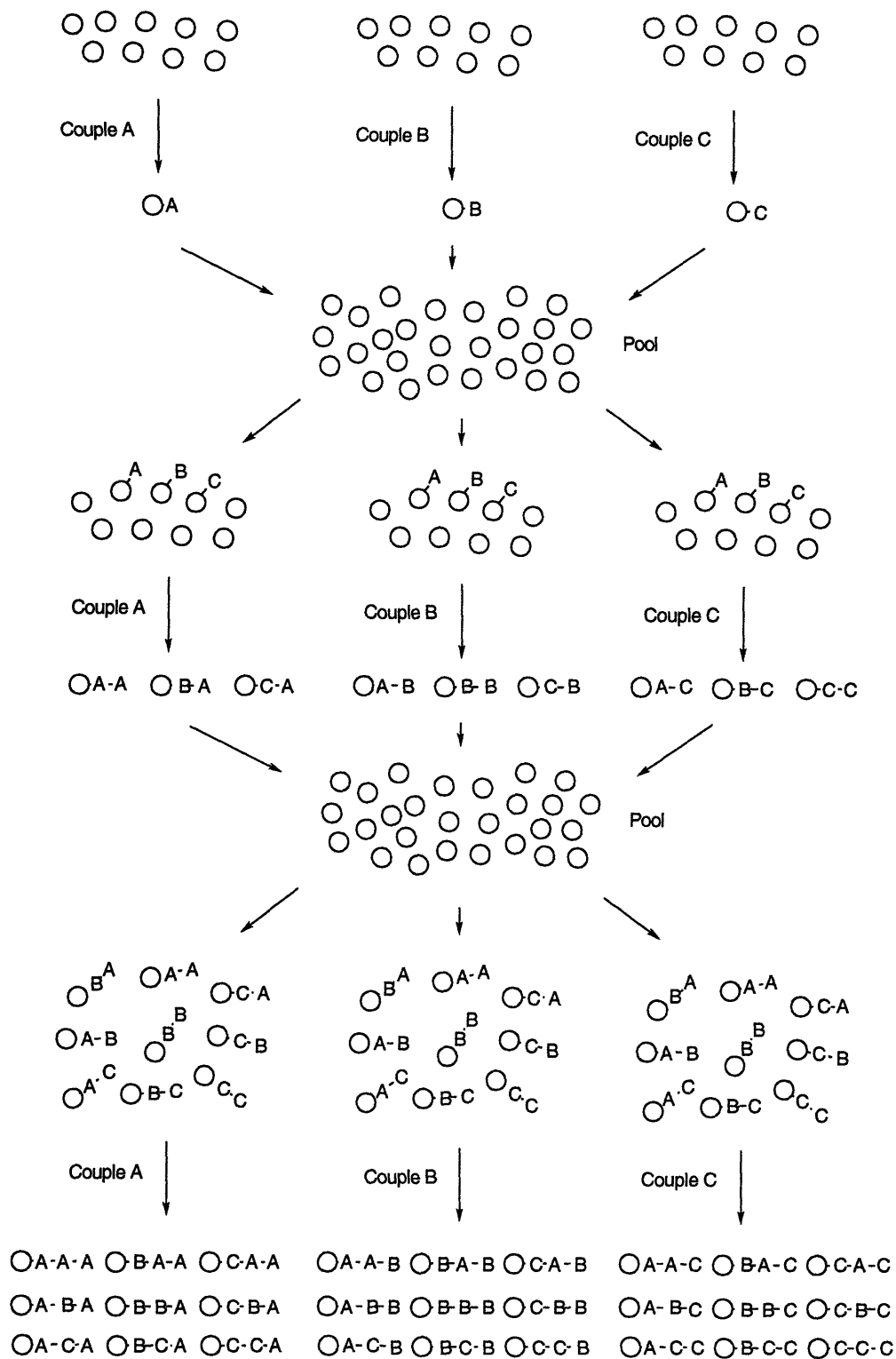
Figure 31. Schematic synthesis of peptides using the "Split Bead Method."[49,50]

Peptides, however, have several drawbacks as drug candidates, most notably their flexible, linear shape and their susceptibility to enzymatic cleavage *in vivo*. While the latter may be overcome by incorporating non-natural building blocks such as D-amino acids, the former characteristic is intrinsic to peptides. By nature, a peptide is a chain which must fold into a specific conformation to achieve a bioactive structure. Most of today's drugs, however, are much more rigid, with the positioning of their functionalities set by a well-defined carbon or heterocyclic skeleton. Ibuprofen (Figure 32) is a very simple example, with two flexible moieties attached to a benzene backbone. Such a skeleton overcomes much of the entropy barrier that looms before any unfolded peptide, instantly conferring an active shape to the drug's functional groups. A library of small, relatively rigid molecules -- molecules which do not have to be able to fold into a secondary structure to present their functionalities in a certain conformation -- will have a much higher chance of containing bioactive structures than a peptide library in which many of the structures -- lacking the ability to fold into a single energetically stable conformation -- exist as flexible "random coils."
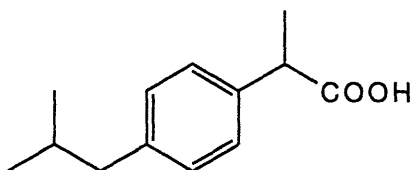
Figure 32. The drug Ibuprofen.

A major challenge to researchers in the field of combinatorial chemistry has thus become the application of combinatorial methods to produce small, drug-related molecules. The current list of successes includes benzodiazepines,[53] (*e.g.*, **46**) peptoids,[54] (*e.g.*, **47**) carbamates,[55,56](*e.g.*, **48**) and other synthetic compounds (*e.g.*, **49**) recently termed "diversomers;"[57,58] examples of each are noted in Figure 33. These combinatorial syntheses all begin to condense the functionality of the peptide chain into a more compact unit, increasing the preorganization of side chains to achieve a well-

defined conformation. With the introduction of poly-functionalized core molecules, the Rebek group recently added to the repertoire of methods by which small, conformationally rigid molecules may be combinatorially synthesized. The concept underlying our method is detailed in the following section.
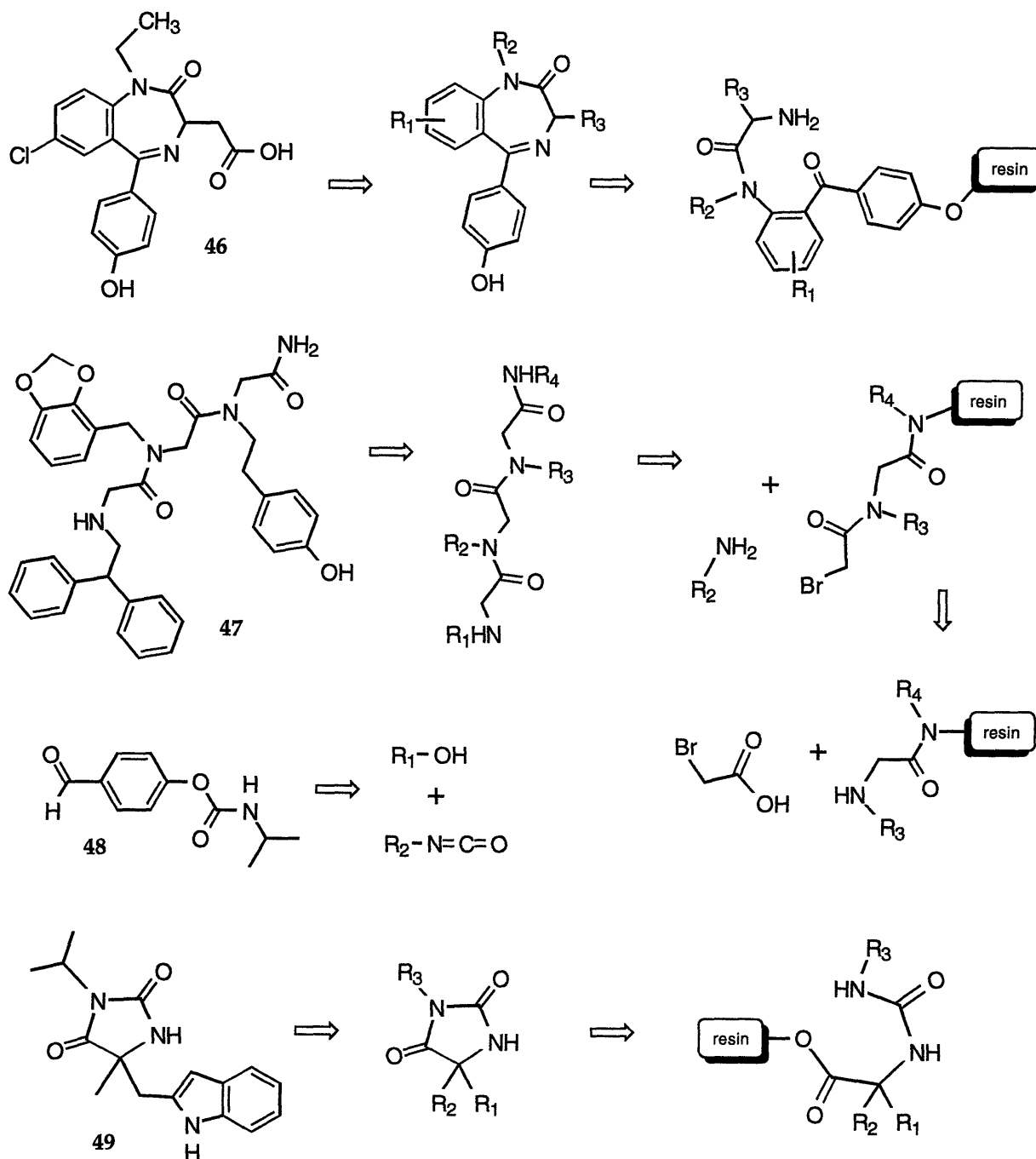
Figure 33. Some combinatorial methods of small molecule drug design described in the literature.[53-58]

II.i.2   Building Blocks on a Scaffold: Poly-functionalized Core Molecules

The ideal combinatorial method for the discovery of pharmaceutically valuable chemicals would produce large libraries of small organic molecules with a drug-related structure, yet would still lend itself to rapid synthesis, screening, and compound structure determination. In 1993 in our lab, postdoctoral fellow Thomas Carell began work on a combinatorial method which approached this goal.[59,60]

It should be noted here that the original idea for the addition of ligands to a poly-functionalized core molecule was conceived by Dr. Rebek and Thomas. I joined the project only after the first set of libraries had been created, when Thomas needed verification of their diversity by HPLC (**II.i.4**, Figure 6, *q.v.*). That said, under Thomas' tutelage I contributed happily to all aspects of the project thereafter, and thus I present the work of Section **II.ii** as a "second author." I later set out on my own foray into combinatorial chemistry, the results of which are detailed in section **II.iii**.

Poly-Functionalized
Core Molecule

Building Blocks

etc...

Random Generation of a Library of all Possible Variants
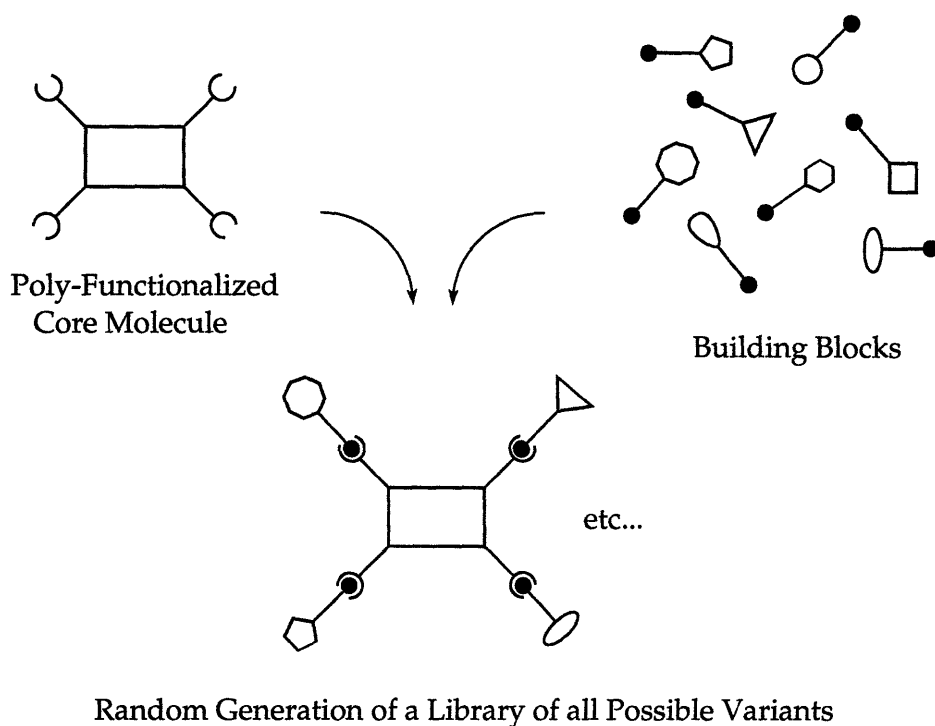
Figure 34.   Schematic representation of the procedure used to generate libraries of small organic molecules.

The idea behind our method of generating libraries of small organic molecules is summarized in Figure 34: combine a rigid core molecule supporting multiple reactive sites with a mixture of functionally diverse building blocks to produce a stochastic library of poly-functionalized structures. As illustrated in Figure 35, a molecule such as 9,9-dimethyl-xanthene-2,4,5,7,-tetra acid chloride 50 could be combined with four molar equivalents of an equimolar mixture of amines A-Z to produce tetra-substituted xanthene compounds A,A,A,A through Z,Z,Z,Z.
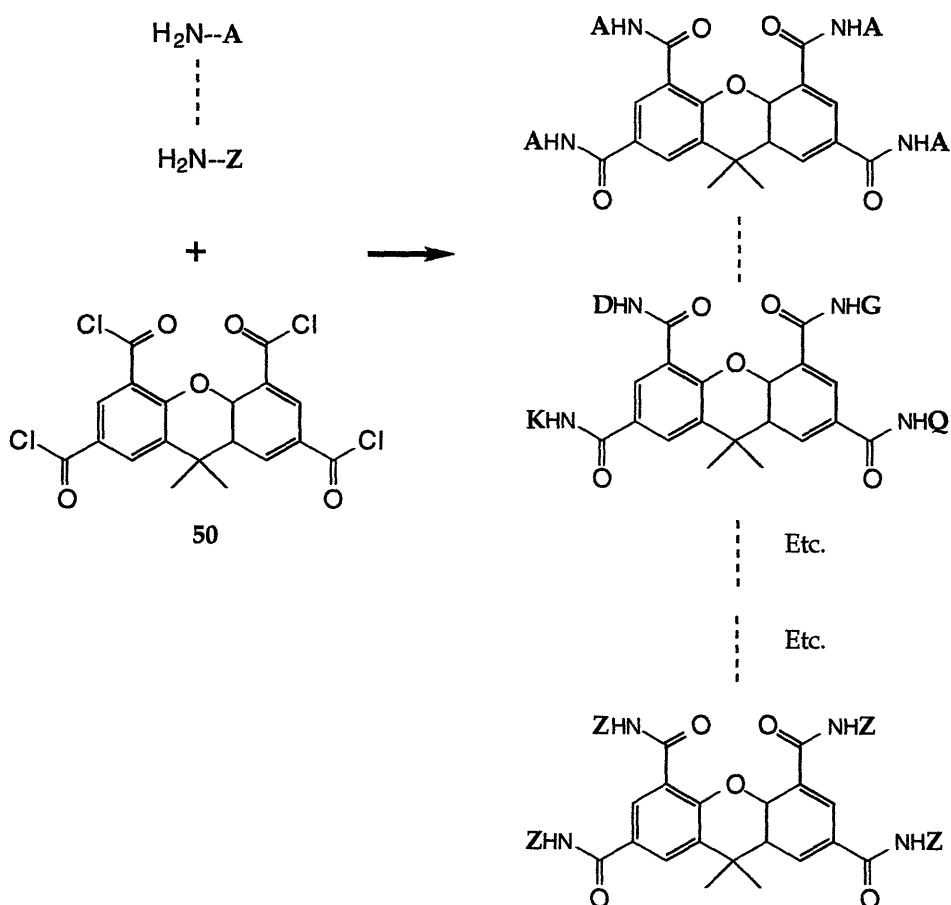
Figure 35. Schematic synthesis of a combinatorial library from xanthene core 50 and amines A - Z.

This method of library generation has several advantages. Firstly, it is a powerful method of generating molecular diversity; in a single combinatorial step, the xanthene core 50 above and twenty-six amines A-Z would produce theoretically 228,826

different xanthene tetra amides. Secondly, unlike some methods of generating peptide libraries (the split bead method, for example[49,50]), these compounds are not on a solid support but free in solution. This allows assaying of the compounds without worry of the solid support giving artifactual results. Thirdly, the molecules all have a rigid carbon backbone as defined by the core molecule. This scaffold determines the basic shape of the compounds, and could range from a very symmetrical cubane[61] core to a completely asymmetric carbon or heterocyclic skeleton. If a particular bioassay is known to be predisposed to a certain molecular shape or pharmacophore, a suitably configured core molecule may be chosen. Finally, this combinatorial strategy allows for the display of a nearly limitless variety of functionalities, as determined by the building blocks used. In the example in Figure 35, one could make use of hundreds of commercially available amines as building blocks. Thus, a broad scope of molecular recognition surfaces may be accessed in a single combinatorial step.

The above method of generating small-molecule libraries has two drawbacks, both of which are surmountable. Firstly, since compounds are generated in a single solution (not in wells or on beads as in some schemes), a method of screening must be employed which can assay a large *mixture* of molecules and eventually select one or more individual compounds of highest activity. As described later in detail, an iterative screening procedure modeled after Houghten[62] was used in this regard. A second problem is that not all building blocks will be compatible with the reactive sites of a given core molecule; for instance, lysine **51** (Figure 36) might react with xanthene tetra acid chloride at the lysine amino terminus, carboxy terminus, or at its amine side chain. This problem was easily circumvented by protecting potentially active sites during the combinatorial step and then deprotecting afterward in a second step.[63] Lysine, for example, can be introduced as **52**, and after reaction with the core, the *t*-Butyl and Boc groups can be removed with trifluoroacetic acid.

51   **52**
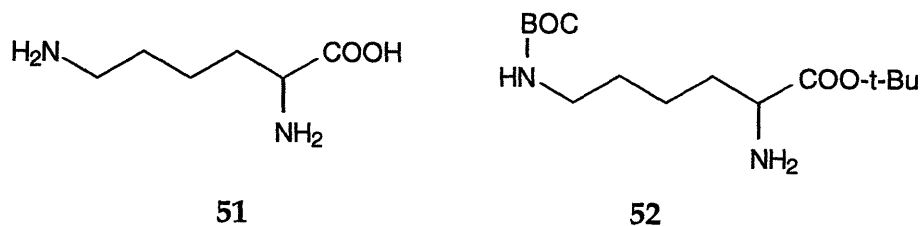
Figure 36. Lysine and its protected equivalent.

II.i.3   Xanthene Tetra Acid Chloride as a Core Molecule

In our initial studies, we used as a core 9,9-dimethyl-xanthene-2,4,5,7,-tetra acid chloride **50**. As shown in Scheme 2, this was prepared from commercial xanthenone **53** via methylation of the 9 position to give **54**, followed by tetrabromination to give **55**. Br/CN exchange to **56** and subsequent hydrolysis of the tetracyano compound provided 9,9-dimethyl xanthene-2,4,5,7-tetracarboxylic acid **57**, which was converted to the tetra acid chloride **50** with oxalyl chloride.
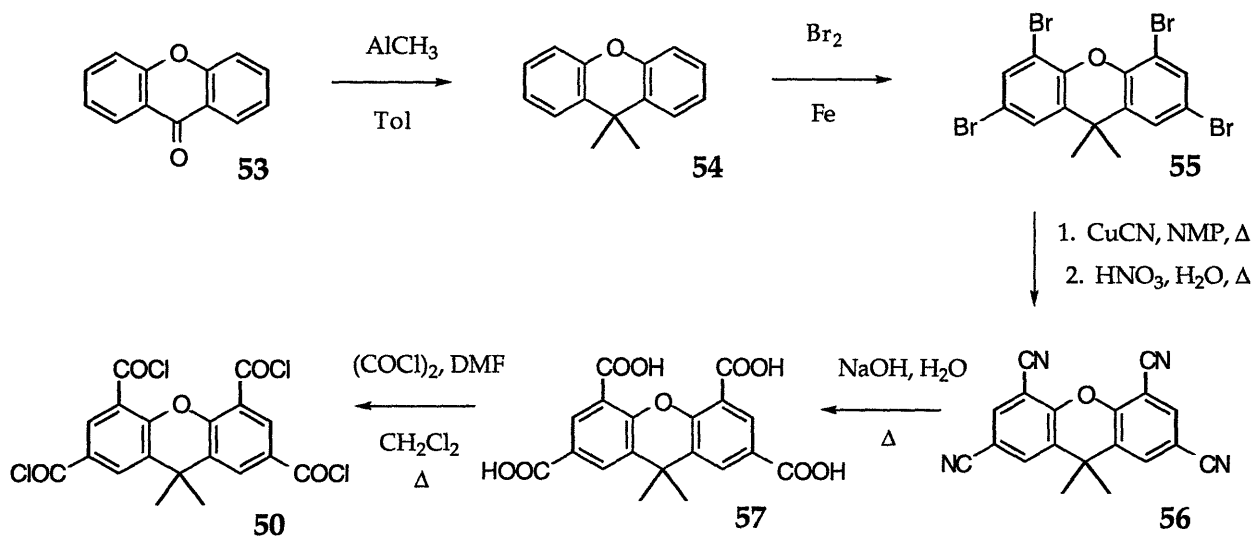
Scheme 2. Synthesis of the of xanthene core 50.

Combining a total 4 moles of amine nucleophiles for 1 mole of a tetra acid chloride core (*i.e.* 4/n moles of each amine for each of n amines), the theoretical number of different molecules created under ideal reaction conditions (identical reactivity of all building blocks and core sites) can be calculated from the combinatorial rule[64] and a set of symmetry factors. These factors depend on the symmetry of the core and are determined individually for each core; the higher the symmetry of the core structure, the fewer compounds are generated with a given set of building blocks. For example, twenty-one different building blocks combined with a highly symmetric cubane tetra acid chloride[61] core molecule would generate theoretically 16,611 possible compounds, while the same building blocks, when combined with the xanthene core, would generate theoretically 97,461 compounds.

Table 7 Calculation of molecular diversity for the xanthene tetra acid chloride core 50.

Combination Type	C_{2v} Symmetry Multiplier	Combinatorial Rule[64]	Combinations for m =			
			4	7	12	21
AAAA	1	$m!/1!(m-1)!$	4	7	12	21
AAAB/AABB	8	$m!/2!(m-2)!$	48	168	528	1680
AABC	18	$m!/3!(m-3)!$	72	630	3960	23940
ABCD	12	$m!/4!(m-4)!$	12	420	5940	71820
		Total Combinations	136	1225	10440	97461

Calculation of the theoretical number of tetra-functionalized compounds produced with xanthene core 50 and *m* building blocks is shown in Table 7 for m = 4, 7, 12, and 21. The total number of compounds equals the sum of the number of compounds with one (AAAA), two (AABB, AAAB), three (AABC), and four (ABCD) different building blocks.

Even under conditions of ideal reactivity, however, it should be noted that not all of the combinations in Table 7 would be produced in equal ratios. For example, given the symmetry of the xanthene core molecule, there will be two molecules of "xanthene-2,4,5,-A-7-B" for every one of "xanthene-2,4,5,7-tetra-A," since "xanthene-2-A-4,5,7-B" is equivalent to the former.

In practice, these ratios would also be affected by the reactivity and steric properties of the given amines and the four acid chloride core sites. Even though there exists in the library reaction mixture one amine for every one acid chloride, the most reactive, least hindered amine will condense preferentially with the most reactive, least hindered core site. While these secondary effects will clearly bias the ratio of compounds formed in a library, it was hoped that the facile reaction of acid chloride plus amine would limit the bias such that all compounds would be created at concentrations within at least an order of magnitude. This supposition was later confirmed by mass spectrometric analysis (see Section II.ii.2).

II.i.4 The First Libraries: Evidence for Molecular Diversity

To prepare libraries with the xanthene tetra acid chloride **50**, one equivalent of **50** was allowed to react with four equivalents of an amine mixture. In initial experiments, the amine mixtures contained equimolar mixtures of four to twenty-one of the building blocks listed in Table 8. With a few exceptions, these building blocks were L-amino acid derivatives, although clearly a wide variety of nucleophiles could be used in this context. The L-amino acids represent a natural set of biologically relevant, functionally diverse building blocks which are commercially available in their protected forms, and were therefore selected to establish a methodology for our scheme.

Table 8. List of the amine building blocks used to create libraries for HPLC analysis.

Amino Acid Building Blocks	Non-Amino Acid Building Blocks
1. L-alanine-methyl ester (Ala)	15. N-methylpyrrol-2-ethylamine
2. O^4-tert-butyl-L-aspartic acid methyl ester (Asp)	16. Furfurylamine
3. O^5-tert-butyl-L-glutamic acid methyl ester (Glu)	17. p-methoxybenzylamine
4. L-isoleucine-tert-butyl ester (Ile),	
5. L-leucine-methyl ester (Leu)	18. N^g-4-methoxy-2,3,6-trimethylbenzene-sulfonyl
6. N^ε-Boc-L-lysine-methyl ester (Lys)	-L-arginine-p-methoxy-benzylamide
7. L-methionine-methyl ester (Met)	19. S-trityl-L-cysteine-benzylamide
8. L-phenylalanine-methyl ester (Phe)	20. N^{im}-trityl-L-histidine-n-propylamide
9. L-proline-methyl ester (Pro)	21. L-valine-cyclohexylamide
10. O-tert-butyl-L-serine-methyl ester (Ser)	
11. O-tert-butyl-L-threonine-methyl ester (Thr)	
12. L-tryptophane-methyl ester (Trp)	
13. O-tert-butyl-L-tyrosine-methyl ester (Tyr)	
14. L-valine-methyl ester (Val).	

In order to extend the diversity of functional groups which were introduced into our libraries, three non-amino acid derived amines were used: p-methoxybenzylamine, furfurylamine and N-methylpyrrol-2-ethylamine. In addition, four amino acids were further functionalized at their carboxyl termini by reaction of the FMOC protected amino acids Fmoc-Arg(Mtr)-OH, Fmoc-Cys(Trt)-OH, Fmoc-His(Trt)-OH and Fmoc-Val-OH with p-methoxybenzylamine, benzylamine, n-propylamine and cyclohexylamine respectively (DMF, BOP, triethylamine). Subsequent cleavage of the FMOC protection groups yielded four novel building blocks: ArgA, CysA, HisA, ValA.

To ensure that all building blocks reacted with the core molecules in high yields, the xanthene core molecule 50 was reacted separately with the 21 amines in Table 8. Each amine listed gave the expected tetra-functionalized product in excellent yield within 30 min. reaction time. Additional functional groups in the building blocks other than the desired amines which might react with acid chlorides (*e.g.*, Ser side chain) were blocked with acid-labile protecting groups.[63] The hydrophobic nature of the libraries produced with the protected building blocks allowed their separation from unreacted amines by extraction with solutions of 1 M citric acid and water.

The obvious question arose: what had been created? What was in fact contained in those white powders? At this point in the project I was able to offer a ready solution based on my recent work with self replicating molecules: HPLC. Thus, as a qualitative check on the viability of the synthesis, protected libraries with theoretically increasing complexity were synthesized and examined on an analytical normal phase column. For this purpose, the xanthene tetra acid chloride 50 was condensed with four, seven, twelve and twenty-one different amines from Table 8, resulting in the preparation of mixtures L1 - L4 with theoretically 136, 1225, 10,440 and 97,461 different compounds respectively.[19] The HPLC chromatograms obtained from these mixtures are shown in Figure 37. They reveal changes in complexity and resolution which one would expect with increasing diversity, indicating that increasing the number of building blocks does in fact produce mixtures which are more and more diverse. The concept of combinatorial synthesis through poly-functionalized core molecules was upheld.
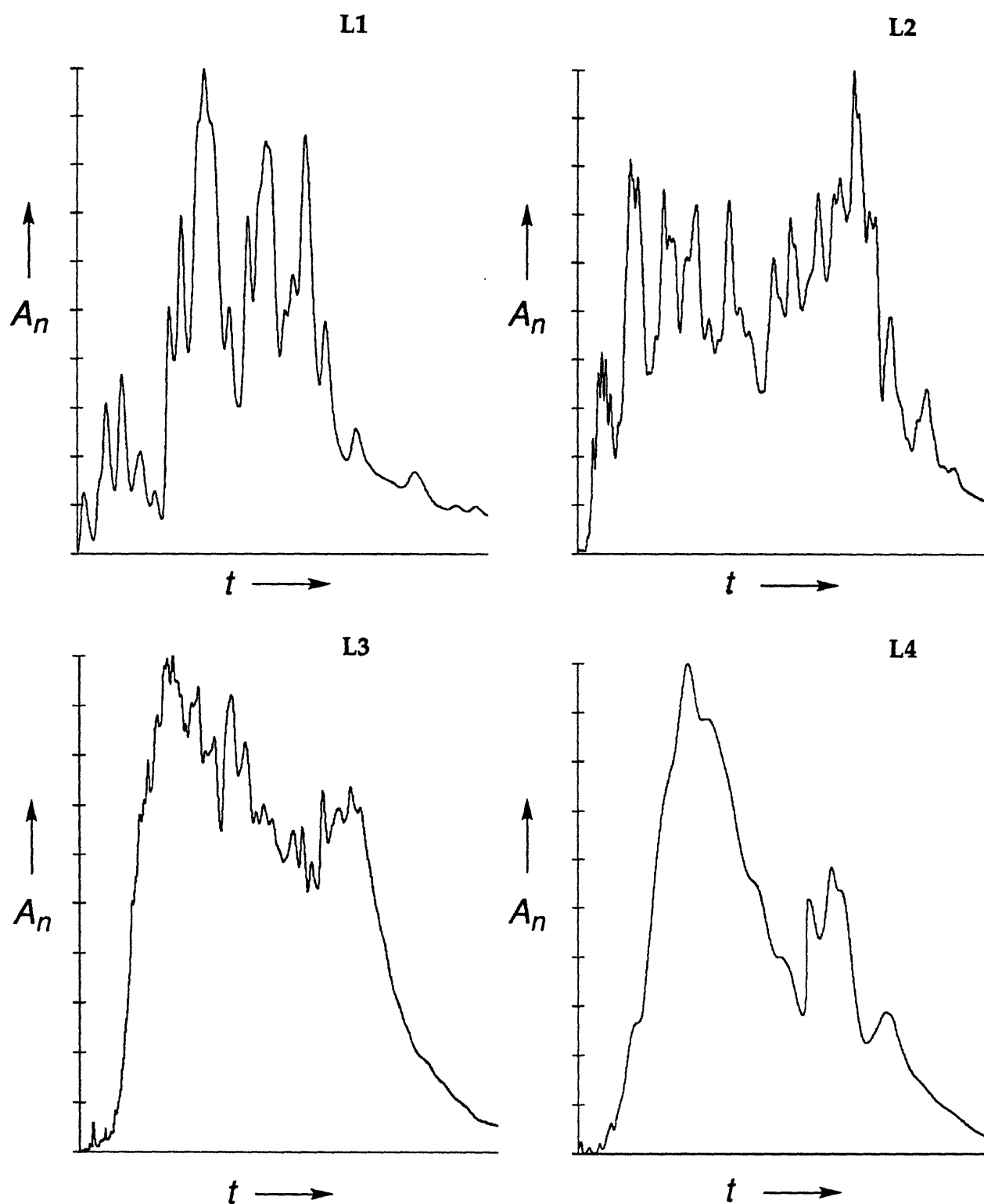
Figure 37. Analytical HPLC-chromatograms (normalized) obtained from libraries L1 - L4. Assuming ideal reaction conditions, library L1 contains theoretically 136 compounds, L2 contains 1225, L3 contains 10,440 and L4 contains 97,461. The libraries were constructed with the xanthene tetra acid chloride 50 and four, seven, twelve and twenty-one of the amines in Table 8.

II.ii   Synthesis, Analysis, and Screening of Water Soluble Libraries

Since many of the building blocks in Table 8 were protected with acid-labile protecting groups, it was hoped that treating the resulting libraries with acid would produce water soluble libraries. They were therefore treated with trifluoroacetic acid and precipitated with ether/n-hexane (1:1). After deprotection, however, the libraries which had been created with the building blocks in Table 8 were not found to be sufficiently water-soluble for screening purposes. It was assumed that a preponderance of methyl ester derivatives, which had given the protected libraries excellent organic phase solubility for HPLC analysis, lay at the root of the problem. The emphasis was thus shifted toward creating libraries which might eventually be screened in aqueous assays.

II.ii.1 Production of Water Soluble Libraries

A new set of 18 building blocks, listed in Table 9, was chosen specifically for properties of water solubility while retaining the diverse set of functionalities necessary to produce varied surfaces of molecular recognition. The new set of building blocks incorporated more t-butyl esters; unlike methyl esters, these would be easily cleaved under acidic deprotection conditions. The standard production of libraries thus became a three step procedure as detailed in Scheme 3:

1) Amine building blocks are coupled to the xanthene tetra acid chloride core **50** in CH_2Cl_2 with triethylamine base to take up HCl produced. A total of 4 moles of amine building blocks are used for every 1 mole of the tetra acid chloride.

2) Libraries in CH_2Cl_2 are washed with 1 M citric acid soln. and H_2O to remove triethylamine and any unreacted amines.

3) Libraries are deprotected with reagent K, a trifluoroacetic acid based reagent containing radical scavengers (trifluoroacetic acid, water, phenol, thioanisol, ethanedithiol (82.5 : 5 : 5 : 5 : 2.5).[65] This cleaves all t-butyl esters, Mtr, Trt, and Boc groups from the building blocks listed in Table 9 and converts the protected libraries into their water-soluble forms. The mixtures are then precipitated with ether/hexanes 1:1 to yield the libraries as powders.

Poly-functionalized core

+

Set of protected amines with sidechains R_n

$$CH_2Cl_2$$

triethylamine

wash CH_2Cl_2 with citric acid / H_2O to remove TEA·HCl and unreacted amines

Protected library

1. trifluoroacetic acid, scavengers[21]
2. precipitate with 1:1 ether/hexanes

Water soluble library
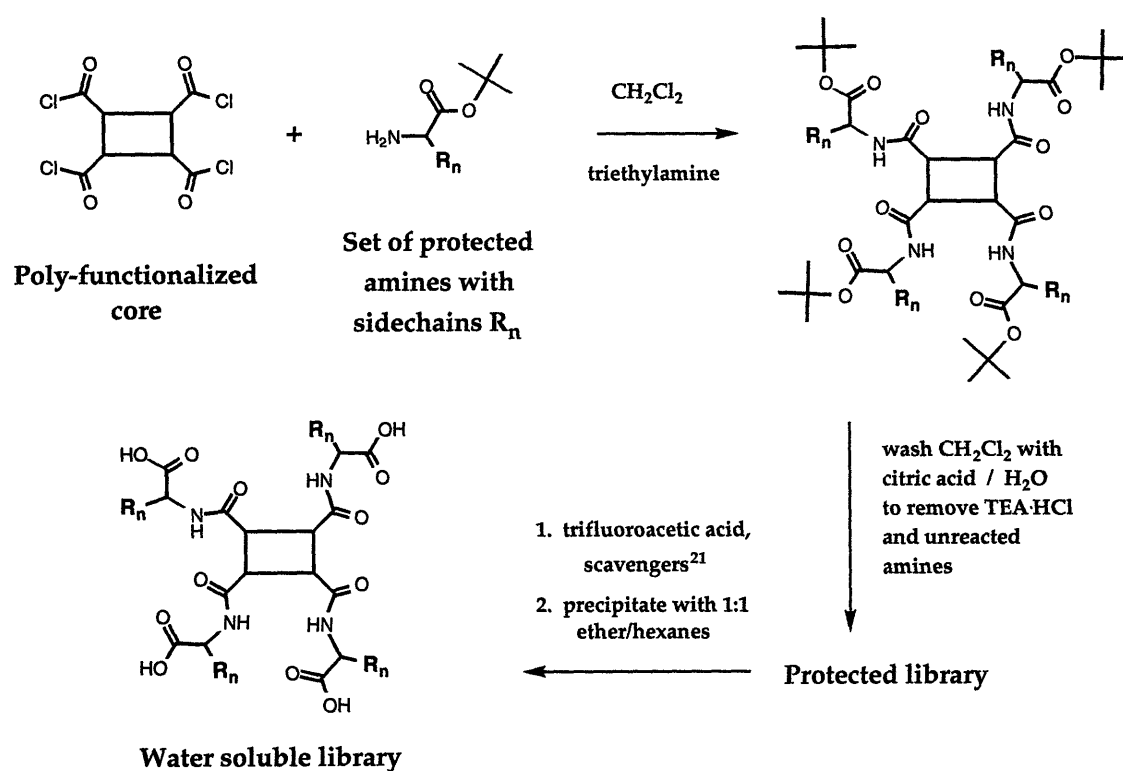
Scheme 3. Schematic synthesis of water soluble libraries.

Table 9: Building blocks used to prepare water soluble libraries.

	amino acid	protected reagent used
1	Ala	L-alanine-*tert*-butyl ester
2	Arg	N^g-4-methoxy-2,3,6-trimethylbenzene-sulfonyl-L-arginine
3	Asn	L-asparagine-*tert*-butyl ester
4	Asp	L-aspartic acid-β-*tert*-butyl-α-*tert*-butyl ester
5	Glu	L-glutamic acid-γ-*tert*-butyl-α-*tert*-butyl ester
6	Gly-OMe	Glycine-methyl ester
7	His	N^{im}-trityl-L-histidine
8	Ile	L-isoleucine-*tert*-butyl ester
9	Leu	L-leucine-*tert*-butyl ester
10	Lys-OMe	N^ε-Boc-L-lysine-methyl ester
11	Met-OMe	L-methionine-methyl ester
12	Phe	L-phenylalanine-*tert*-butyl ester
13	Pro	L-proline-*tert*-butyl ester
14	Ser	O-*tert*-butyl-L-serine-*tert*-butyl ester
15	Thr-OMe	O-*tert*-butyl-L-threonine-methyl ester
16	Trp-OMe	L-tryptophane-methyl ester
17	Tyr-OMe	O-*tert*-butyl-L-tyrosine-methyl ester,
18	Val	L-valine-*tert*-butyl ester

Using the procedure of Scheme 3, a library formed from **50** and the building blocks listed in Table 9 was found to be soluble in 10:1 water/dimethylsulfoxide, making the mixture viable for screening purposes. With 18 building blocks, the library contained theoretically 52,650 compounds. Before screening was begun against a biological assay, however, it was desired to know approximately what percentage of this theoretical number actually had been created in the combinatorial synthesis. While HPLC had provided a qualitative check of the organic soluble protected libraries, evidence of a more quantitative nature was needed to show that the deprotected libraries were as diverse as was predicted. Knowledge of the quality of the libraries

generated was essential to determine the likelihood of false negatives or false positives when screening a combinatorial library for useful compounds. In the case of false negatives, a potential ligand could be missed in the screening procedure either because it was not present or because it was present in a very low amount. False positives, on the other hand, could be caused by the activity of side products (about which more later -- see section **II.iii.4**).

II.ii.2 Analysis of Water Soluble Libraries

To establish that a sizable fraction of the expected compounds are produced in the above synthesis of water-soluble molecular libraries, it was decided to analyze several libraries by electrospray ionization (ESI) mass spectrometry.[60,66,67] In order to facilitate this goal, a collaboration was entered into with Paul Vouros of Northeastern University and his graduate student Yuriy Dunayevskiy. In discussion with our collaborators, it was concluded that direct observation of a library of many thousands of compounds was out of the question. Therefore, smaller yet representative libraries were constructed. Yuriy's mass spectrometric analysis of these model libraries allowed us to probe the effectiveness of the general synthesis in forming complex mixtures of molecules.[68]
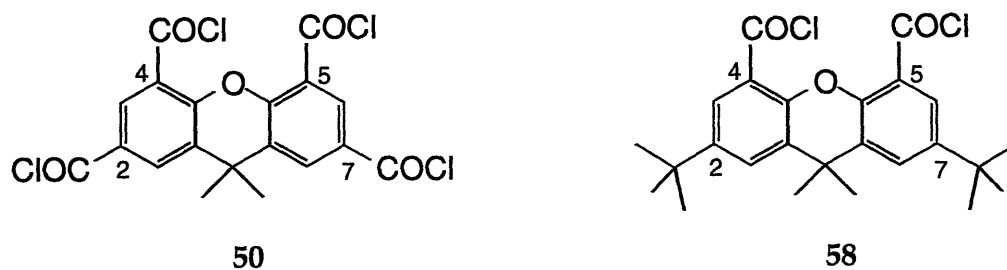
Figure 38. Truncated core-molecule **58** for mass spectrometric determination of model-library diversity, compared with full core molecule **50**.

Based on three criteria, the diacid chloride **58** [69] (Figure 38) was selected to serve as the core molecule for the model libraries. Firstly, condensation of **58** with nine amines from Table 9 leads to theoretically only 45 disubstituted compounds, a seemingly manageable number. Secondly, the two amides formed are at the four and five positions of the xanthene scaffold. As these positions are most susceptible to problems of steric crowding about the xanthene core, **58** provided a realistic test for determining any building block combinations which, because of their bulk, would be disfavored in the synthesis. Thirdly, the two *tert*-butyl groups occupying the two and seven position allowed examination of the precipitation behavior of highly hydrophobic library compounds in the final treatment with ether/*n*-hexane.

For the synthesis of the model libraries, five sets of building blocks were selected from the 18 building blocks listed in Table 9: three sets of eight amines and two of nine amines. Condensation of these sets of amines with the core molecule **58** was expected to give libraries of 36 and 45 compounds, respectively. With the help of a simple computer program which was written for the task, the building blocks were grouped in a set such that all of the compounds in the libraries produced would possess a unique molecular weight. This simplified individual detection of each compound present in the libraries. After their formation, the protected model libraries were washed and deprotected as shown in Scheme 3, and finally precipitated with ether and *n*-hexane to give white powders.

All model libraries were analyzed in both positive and negative ion modes. The ESI mass spectra showed negligible fragmentation under the ion optics settings used, and thus, taken together, the molecular ion peaks obtained from these measurements are a set of data directly correlated to the diversity of a given molecular library. The molecular ion peaks in the mass spectra were compared with the molecular weights expected for each model library, highlighting which compounds had been formed and which had not. Because the composition of the model libraries was chosen such that

most of the compounds were expected products in more than one model library, the presence or absence of most combinations was checked at least twice. Molecules were considered present if the signal-to-noise ratios of their molecular ion peaks exceeded 10:1 ; rough estimates of the concentration of each compound could be calculated by taking into account the ionization efficiencies of pure samples.

Results of the MS analysis of the model libraries are compiled in Figure 39. On the x- and y-axes are the abbreviations of the building blocks used. Each filled square in the charts represents the presence of one of the expected compounds in that model library. Because the truncated core-molecule possesses C_{2v} symmetry, only half of the possible building block combinations give rise to new compounds. The pattern of the squares indicates whether the corresponding compound was detected as its positive ion (grey), its negative ion (black) or in both modes (striped). An exemplary mass spectrum of model library M2 is shown in Figure 40.

Out of 198 expected compounds in the model libraries, 173 (87%) were detected.[67] Missing were predominantly those compounds which contained tryptophan residues; less than 50% of the expected Trp compounds were detected. The absence of Trp compounds may be due to degradation; Trp is an amino acid known to be vulnerable to strongly acidic deprotection conditions. However, as each model library contained some of the expected Trp-substituted xanthenes, this building block was included in the set of building blocks eventually used for large libraries. Outside of the compounds containing Trp, only two other combinations, the Gly-Gly and Tyr-Tyr compounds, were not detected above the threshold of 10:1 signal-to-noise ratio which had been set. It is thought that these compounds, being very hydrophobic, may not have precipitated well in the final ether/hexane precipitation step.
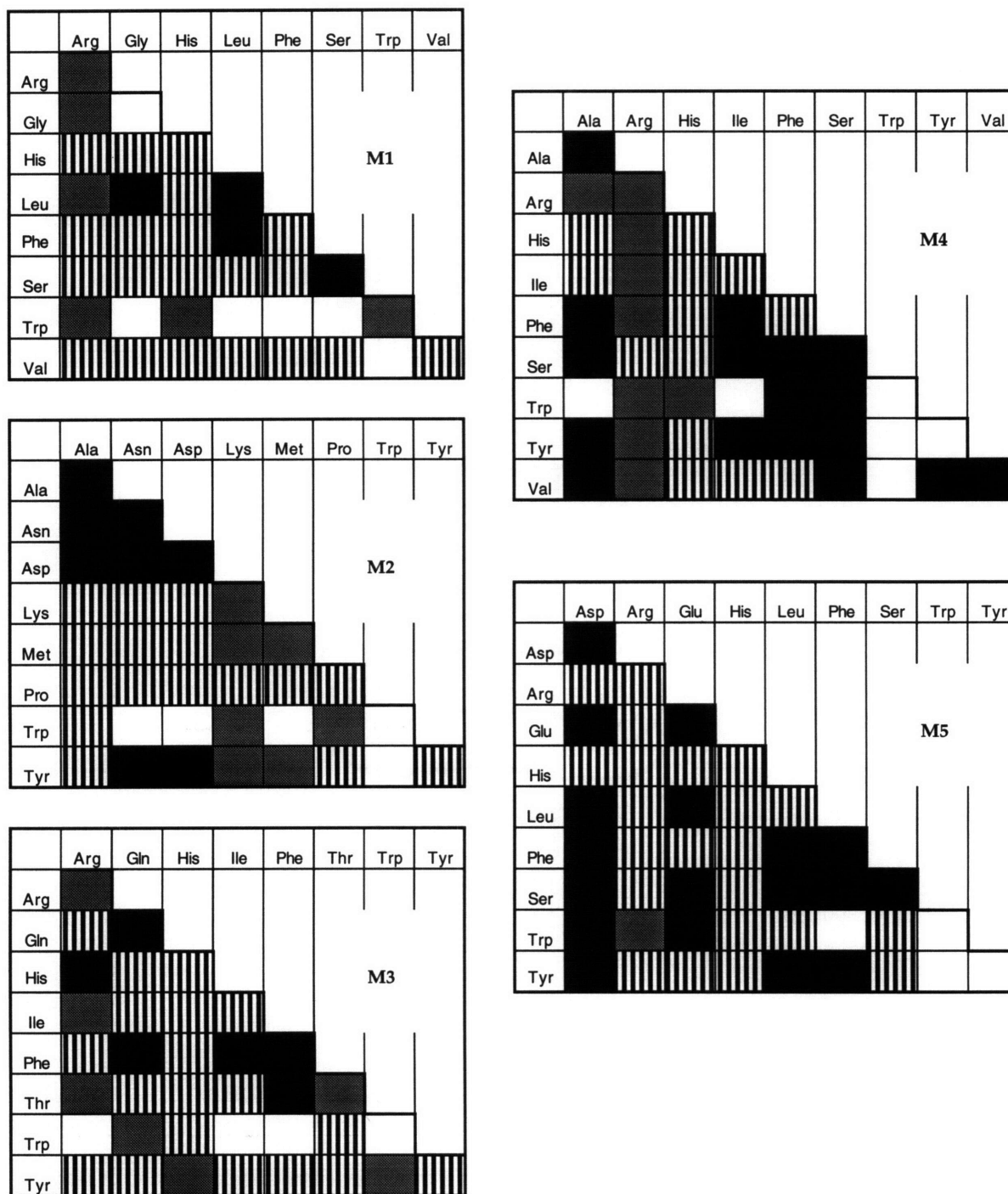
Figure 39. Mass spectrometric analysis of libraries: xanthene **58** disubstituted with various amino acids. Gray: detected as positive ion. Black: detected as negative ion. Striped: detected as both positive and negative ion. White: absent from the mixture.
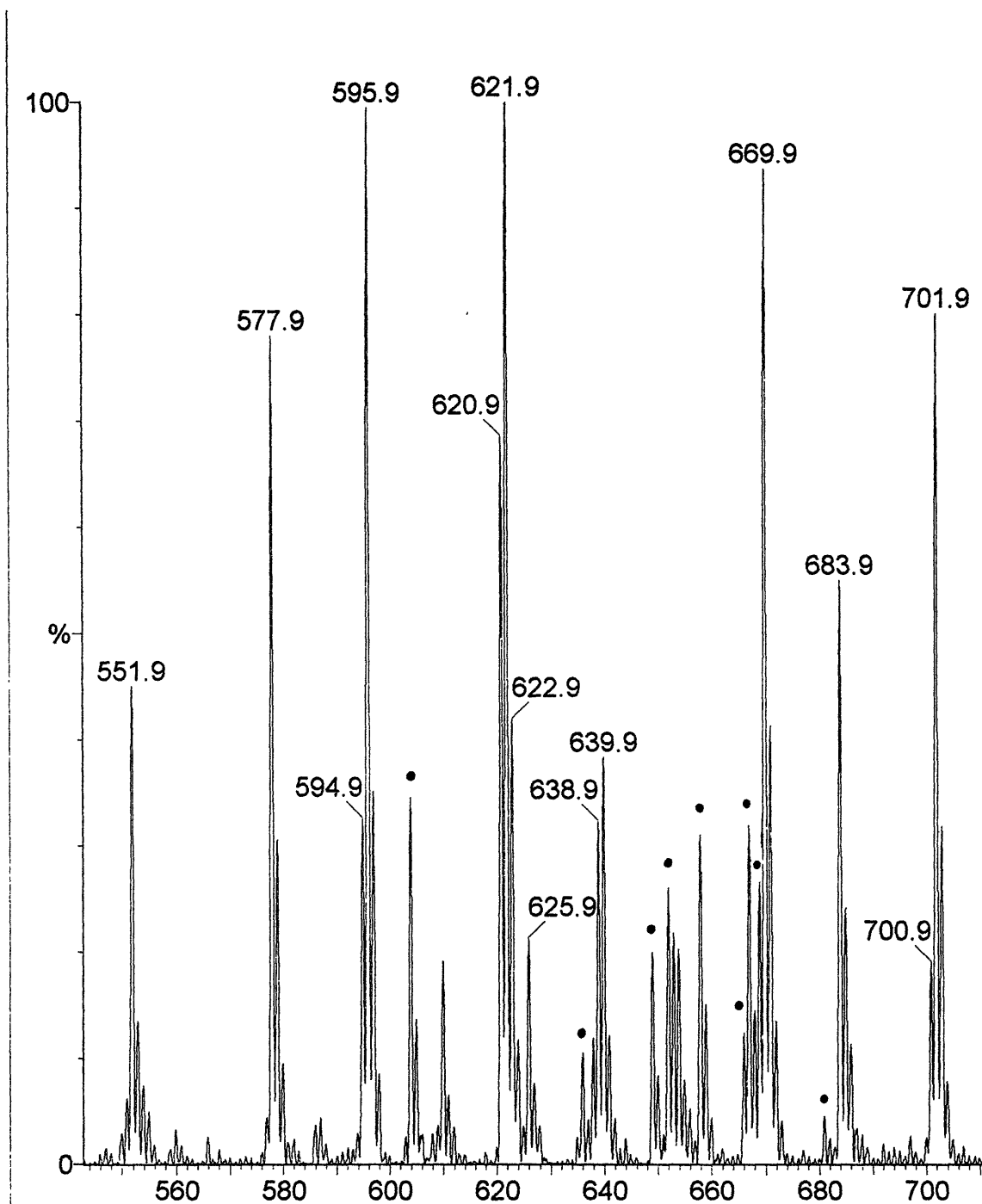
Figure 40.   Negative ion scan (m/z) of Library M2 -- xanthene 58 disubstituted with various amino acids as listed in Table 9.   Expected peaks are marked by their molecular weights (M -1) or by a dot (•) for clarity: Ala-Ala (552), Ala-Pro (578), Ala-Asn (595), Ala-Asp (596), Pro-Pro (604), Pro-Asn (621), Pro-Asp (622), Ala-Lys (623), Ala-Met (626), Asn-Asn (638), Asn-Asp (639), Asp-Asp (640), Pro-Lys (649), Pro-Met (652), Ala-Tyr (658), Asn-Lys (666), Asp-Lys (667), Asn-Met (669), Asp-Met (670), Ala-Trp (681), Pro-Tyr (684), Asn-Tyr (701), Asp-Tyr (702).   Not shown: Tyr-Tyr (764).   The peak at 610 does not match any of the expected products for library M2.

In the synthesis of a large mixture of molecules, the generation of some side products was inevitable. However, MS and MS/MS analyses of the model libraries indicate that relatively few side products were formed in the deprotected, ether/hexane precipitated material. In the six model libraries, only seven significant peaks were observed which did not match the expected molecular ion peaks for disubstituted xanthenes. The structures of the side products were tentatively assigned from MS/MS data, and include a xanthene-mono-histidine-mono-acid derivative and two mono-amino-acid-mono-alkylamine variants each of His, Lys, and Arg. (One such impurity, 58 substituted with Lys and methyl-ethyl-amine, is shown in Figure 42a. This substitution probably arose from an impurity in the triethylamine reagent used in the first step of library synthesis.) Overall, however, we estimate a level of side products below 10% of the total number of compounds in a given library.

Following the success of these experiments, chromatographic techniques coupled on-line to MS were used to add another dimension of resolving power to the system. Using capillary electrophoresis - electrospray ionization - mass spectrometry (CE-ESI-MS),[70,71] Yuriy was able to analyze a library synthesized from the diacid 58 plus all 18 amino acids from Table 9.

CE generates rapid, high resolution separations based on differences in the electrophoretic mobilities of charged species in an electric field in small-diameter fused-silica capillaries.[70,71] The inner walls of the bare silica capillary are negatively charged under aqueous conditions due to acidic silanol groups, and this causes the formation of an excess of positive charges of the electrophoretic buffer layer in contact with the inner wall. When voltage is applied across the capillary, electroosmotic flow (EOF) moves the sample from the positive end (anode) to the negative end (cathode). EOF becomes significant at pH's above 5 (buffer pH for these experiments was 7.9), and drives the analytes toward the cathode regardless of their charge. Thus, cations, neutrals and anions can be electrophoresed in a single run since they all move in the same direction.

The resultant mobility of the library compounds, and therefore their migration times, are determined by the difference between the mobility of the EOF and the electrophoretic mobility of the species: molecules bearing the most positive charge (*e.g.*, Lys derivatives) move with the front of the EOF, while molecules bearing the most negative charge (*e.g.*, Glu derivatives) run slowest.

Figure 41. Results of the mass spectrometric analysis of a library created with 58 plus the 18 building blocks in Table 9: Gray: detected by CE-MS. Checkered: detected by CE-MS/MS. White: absent from the mixture.
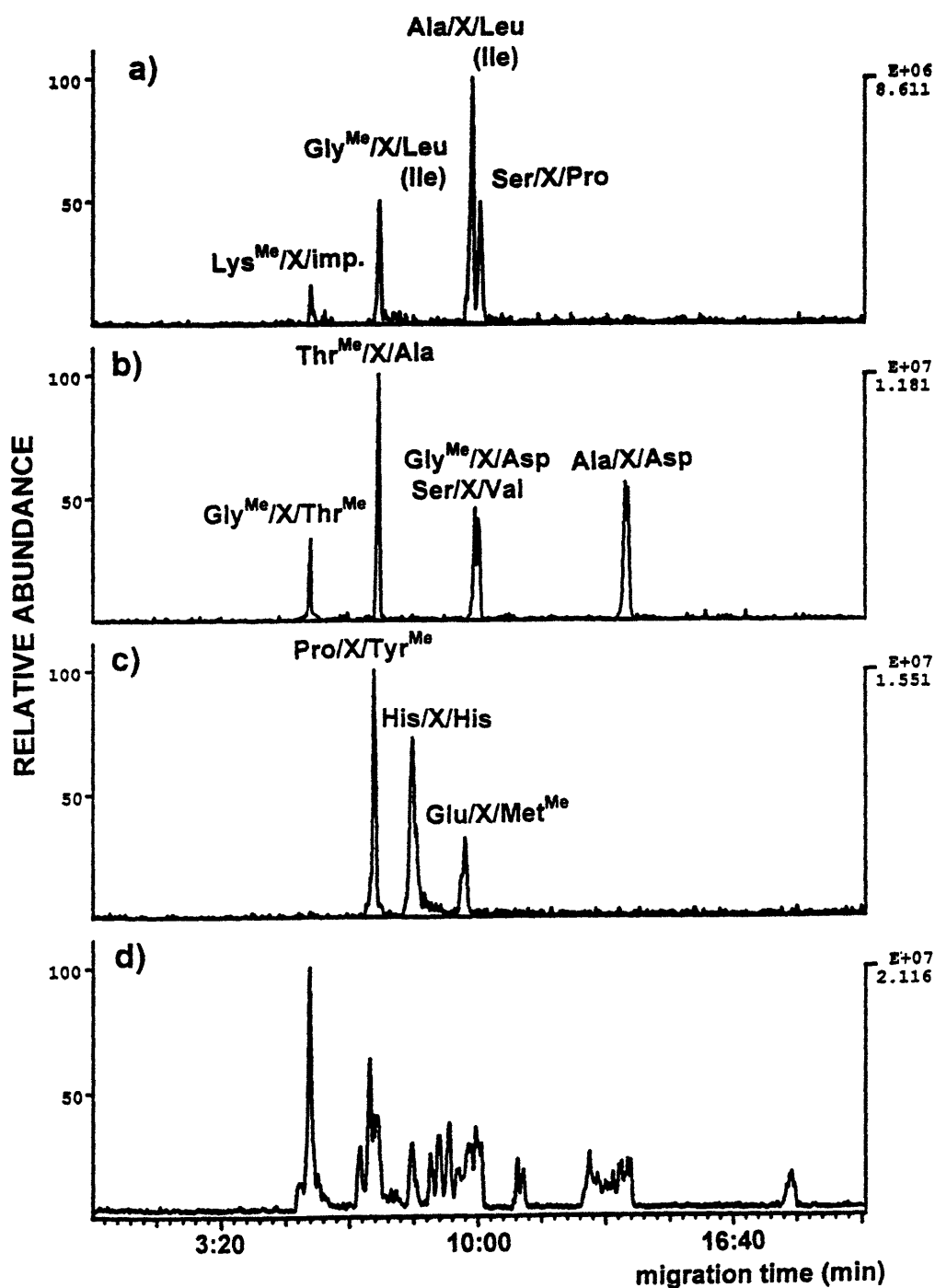
Figure 42. d: CE-MS electropherogram of a library created with 58 plus the 18 building blocks in Table 9. a-c: Selected ion traces: a) m/z = 595.5. b) m/z = 597.5. c) m/z = 685.5. The first peak of a corresponds to 58 substituted with Lys and methyl-ethyl-amine, probably present as a trace impurity in the triethylamine reagent used in synthesis.

Figure 41 shows the results of CE-ESI-MS analysis for the library created with **58** plus the 18 building blocks in Table 9, theoretically 171 compounds. Out of this number, 160 (94%) were found to be present; as in the smaller libraries, the only compounds missing were several of the Trp-containing molecules. Figure 42d shows the full CE-MS electropherogram for the library, with three selected molecular ion traces a, b, and c displayed above. During a CE-ESI-MS run, ESI-MS data is continually acquired over the entire mass range for the duration of the CE run; the MS signal is acquired at each mass range over the separation of the capillary. Figures 42a, b, and c demonstrate this MS signal at ion peaks m/z = 595.5, 597.5, and 685.5 respectively.

The first peak, comigrating with the EOF, consisted of non-charged derivatives which were substituted with methyl ester amino acids on both sites of the core molecule **58**. Most of the compounds with lysine methyl ester on one side and a neutral amino acid on the other also came out with this peak, since the positive charge of the lysine side chain balanced the negative charge on the carboxyl of the second building block. The next group of peaks (6 to 7.5 min.) consisted of various compounds substituted with arginine and also compounds containing both an amino acid methyl ester and a neutral amino acid (with a free carboxyl terminus). The third group in the electropherogram (from 8 to 10 min.) corresponded to those derivatives with two neutral amino acids, plus derivatives substituted with one acidic amino acid (e.g., Glu) and one neutral amino acid methyl ester. His-Asp and His-Glu migrated at 11 min., followed by another group of peaks containing compounds of one acidic and one neutral amino acid. The last peak consisted of three derivatives with negatively charged amino acids: Asp-Asp, Asp-Glu and Glu-Glu. Thus, library components were distributed within a rather wide migration time frame, allowing identification of the analytes in each group by subsequent MS.

While most of the compounds in the mixture with the same molecular weight were separated by CE (e.g., the analytes in Figures 42a and c), within each group of

peaks there were nevertheless several isobaric compounds. Happily, the molecular ions of the xanthene-diamino acids underwent collisionally induced fragmentation to form characteristic daughter ions, allowing unambiguous assignment of library compounds via MS/MS even in cases of isobaric molecular ion peaks. For example, identification of the pair of unresolved compounds Gly-Asp and Ser-Val at m/z = 597.5 in Figure 42b required the use of MS/MS analysis. The collision induced spectrum of the coeluting isobaric compounds contained the characteristic fragments of both molecules, confirming the presence of each in the model library.

The mass spectrometric analyses of the model libraries in Figures 39 and 41 reveal that most of the amines chosen as building blocks generate the expected condensation products with the truncated core-molecule **58**. Furthermore, taking into account the ionization efficiencies of the various pure compounds and the stoichiometry of A/A vs. A/B substitution patterns, those compounds detected were found to be present at concentrations within a single order of magnitute.[66,67]

The compounds absent from the libraries were primarily tryptophan-methyl ester derivatives. If one makes the reasonable assumption that reaction at the 2 and 7 positions of **50** is no more difficult than reaction at the 4 and 5 positions of **58** (see Figure 38 for numbering), it follows that most of the compounds expected in a large tetra-acid chloride based library will be formed. The compounds which are most probably absent in such libraries contain one or more Trp or multiple Gly building blocks, either because of acidic degradation or failure to precipitate in the final isolation step of ether/hexane. However, even if 75% of the expected Trp containing compounds are absent in a condensation of **50** with eighteen building blocks, the water-soluble library created still contains over 50,000 different molecules. With the above mass spectrometric data in hand, we were confident that our synthetic methodology produced highly diverse libraries of well defined composition, and we were able to put faith in the results of subsequent screening assays.

II.ii.3 Screening of Water Soluble Libraries Against Trypsin

Having developed a new combinatorial method and verified its effectiveness in the generation of large numbers of molecules, it remained only to prove that the compounds so produced were in some way useful. In the context of drug design, this translates to whether or not the libraries can be screened to select molecules which show biological activity in a given assay. Combinatorial drug design is dependent upon the generation of a diverse set of molecular recognition surfaces, such that at least one compound in a library will act as a complement to a desired target surface (*e.g.*, the active site of a protein). Would a library of our molecules provide a sufficient sampling of molecular recognition surfaces to generate compounds which could bind selectively to an enzyme?

Inhibition of the enzyme trypsin[72] was chosen as a test of the above system of combinatorial library generation. Trypsin is a digestive enzyme which is a member of the important class of serine proteases,[73,74] and is readily available commercially. Following a standard trypsin assay, we sought to isolate a compound able to inhibit the trypsin catalyzed cleavage of the amide bond in N^{α}-benzoyl-DL-arginine-*p*-nitroanilide[74,75] by screening libraries of compounds in an iterative selection process. Any combinatorial strategy must have some allied process of "deconvolution" by which desired molecules may be selected. The iterative selection procedure used herein was designed through modification and generalization of a screening method first employed by Houghten *et al.* for the identification of active peptides in a hexapeptide library solution.[62] As is described below in detail, the process involves the assay of libraries in which certain building blocks have been left out. If a library is active notwithstanding the omission of a given building block, it may be inferred that that building block is unnecessary for the generation of the active compound sought, and that building block may be discarded from the deconvolution procedure.

For screening purposes, we began with the water soluble library created by condensing the xanthene tetra acid chloride core **50** with the 18 amino acid derivatives listed in Table 9. **(II.ii.1)** This procedure gave a library of theoretically 52,650 different compounds. To evaluate the potency of the library, 2.5 mg in 50 μl DMSO and a blank of pure DMSO were incubated with a solution of bovine pancreatic trypsin in Tris buffer at pH 8.2, 1.0 ml total volume (see experimental). When N^{α}-benzoyl-DL-arginine-p-nitroanilide was added, the inhibitory activity of the library and blank could be correlated inversely to the rate of p-nitroaniline released from trypsin catalyzed hydrolysis. The rate of p-nitroaniline release was followed by UV absorbance. The results of this experiment are depicted graphically in Round 1 of Figure 43. Results are tabulated numerically in the experimental.

The xanthene based library in Round 1 caused a significant reduction of the enzyme's activity (34%). To determine which of the 18 building blocks were most responsible for this interaction, six sublibraries were synthesized. The 18 amines were grouped into the six sets Group 1 - Group 6 listed in Table 10: smaller hydrophobic side chains, larger hydrophobic side chains, basic side chains, hydroxyl side chains (plus Met), aromatic side chains, and acidic side chains (plus Asn).

Table 10. List of the amino acid sets into which building blocks were grouped for the generation of sublibraries in Round 2. (For the list of abbreviations see Table 9).

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Gly	Leu	Arg	Ser	Phe	Glu
Ala	Ile	Lys	Thr	Tyr	Asp
Val	Pro	His	Met	Trp	Asn
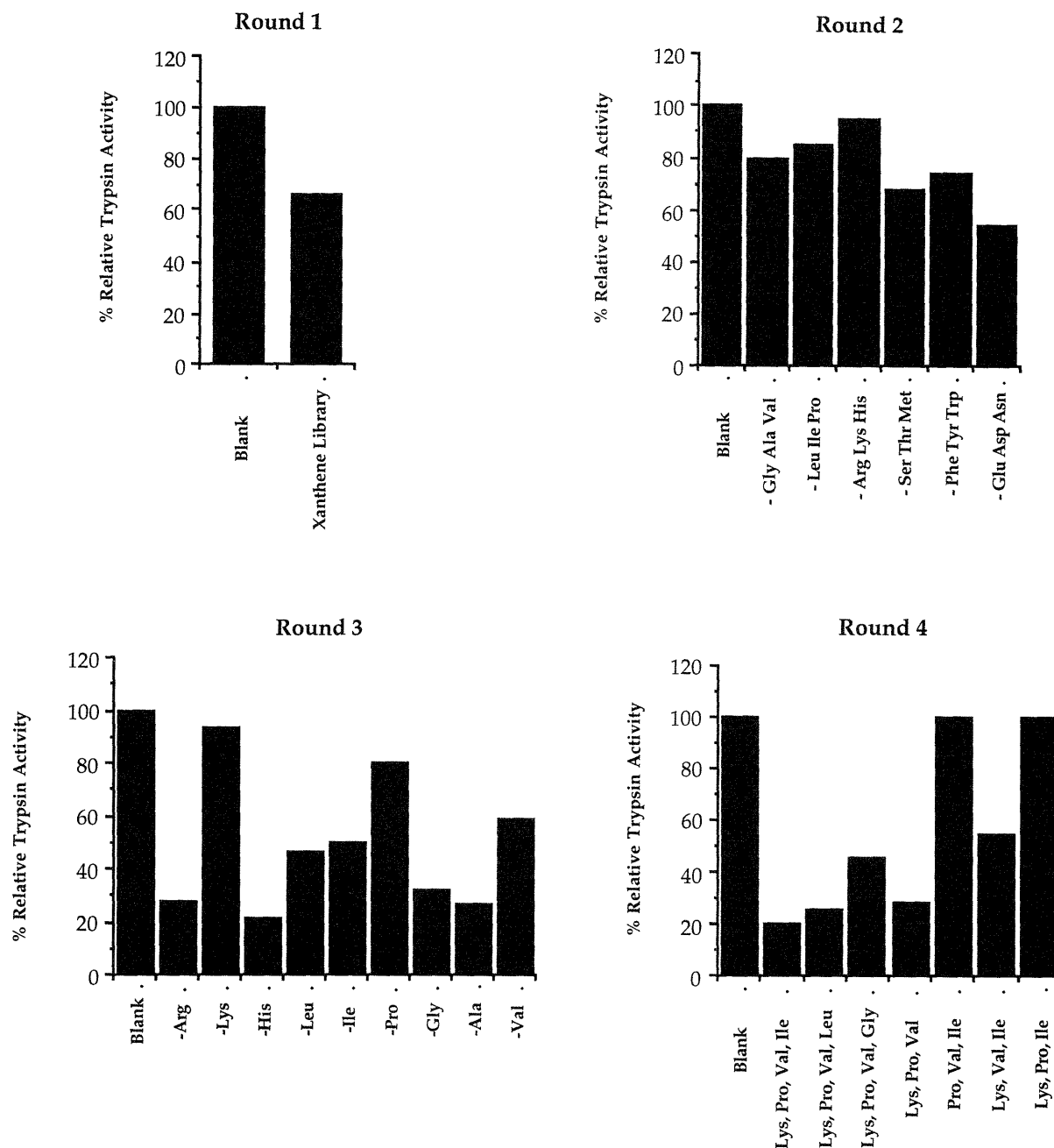
Figure 43. Percent of trypsin activity in the presence of added libraries relative to a blank containing no added library material. Each bar represents average trypsin activity for four measurements. The activity of the blank was set to 100%. Round 1: % of trypsin activity with the initial library constructed from core molecule **50** plus the 18 amino acid building blocks listed in Table 9. Round 2: % of trypsin activity with six sublibraries constructed from core molecule **50** and 15 of the 18 amino acid building blocks, each library missing three building blocks as noted. Round 3: % of trypsin activity with the nine sublibraries constructed from the core molecule **50** and eight of the nine building blocks Arg, Lys, His, Leu, Ile, Pro, Gly, Ala, and Val, each library missing one of the nine building blocks as noted. Round 4: % of trypsin activity with seven sublibraries constructed from the core molecule **50** and three or four of the building blocks Lys, Leu, Ile, Pro, and Val as noted.

Sublibraries were prepared with the xanthene core molecule **50** and fifteen building blocks *from five of the six groups* in Table 10. Thus, the first sublibrary was created with all building blocks *except* those in Group 1, the next was created with all building blocks *except* those in Group 2, and so on. Each sublibrary was calculated to consist of theoretically 25,425 different tetrasubstituted xanthenes. Sublibraries were screened just as the initial library, adding 2.5 mg of material to each run.

The screening results of the six mixtures are depicted in Round 2 of Figure 43, and show that the sublibrary generated *without* the basic side chain Group 3 did *not* inhibit enzyme activity. Thus, the presence of one or more of the three amino acids Arg, Lys and His was critical to the inhibitory potency of the library. This result was not entirely unexpected, given the known preference of trypsin for lysine and arginine at the carbonyl side of the cissile amide bond (often referred to as the P_1 position).[28] The next two most important groups were the aliphatic side chain Groups 1 and 2, containing Gly, Ala, Val and Leu, Ile, Pro respectively. The nine building blocks present in the three groups 1, 2, and 3 were therefore deemed to be most responsible for the presence of inhibitors in the initial library of Round 1.

To further narrow the field of possible inhibitors, nine new xanthene sublibraries were prepared using xanthene core **50** and eight of the nine building blocks already selected, with each sublibrary missing one of these nine amino acids. Each sublibrary in this round contained theoretically 2,080 compounds. The screening results obtained from this experiment are presented in Round 3 of Figure 43; a sublibrary which showed low inhibitory activity signaled that the building block omitted from that group was crucial for trypsin inhibition. For generation of an active trypsin inhibitor, lysine methyl ester appeared to be the most important building block, followed by proline, valine, isoleucine, and leucine.

Because a constant amount of library material -- 2.5 mg -- was again used for the assays of Round 3, and the sublibraries in Round 3 contain in theory only 2,080 compounds compared to 52,650 present in the initial library, building block selection has increased the content of inhibitors: 2.5 mg of material from the best of the sublibraries in Round 3 produce 79% inhibition, up from 34% in Round 1. Thus, the process of building block selection and sublibrary synthesis can be viewed as an "amplification" step.

In order to establish the most potent combination of the above five building blocks, xanthene sublibraries were constructed with three or four of these five -- it was by no means clear at this stage that the most potent molecule would contain four *different* building blocks. However, the trypsin assay results of Round 4 (Figure 43) confirmed that the most potent inhibitor was constructed from the four building blocks lysine, valine, proline and isoleucine. Omission or substitution of any one of these four building blocks gave libraries with a lower potency, even though each compound in a sublibrary constructed with three building blocks was present in *higher* concentration than each compound in a sublibrary constructed with four building blocks. Thus, the search for a trypsin inhibitor was narrowed to the twelve possible structural isomers of the xanthene core coupled to Lys, Ile, Pro, and Val. One or more of these twelve isomers had to be the inhibitor in question. (It should be emphasized that Lys here represents the amino acid methyl ester, as per Table 9, Section II.ii.1. Ile, Pro, and Val represent the corresponding free acids.)

To narrow the remaining possibilities for the structure of the most potent isomer from twelve to two, six more sublibraries were prepared using the dibenzylester xanthene diacid chloride derivative **61**. The new core molecule **61** was easily synthesized from the xanthene tetra acid chloride **50**.
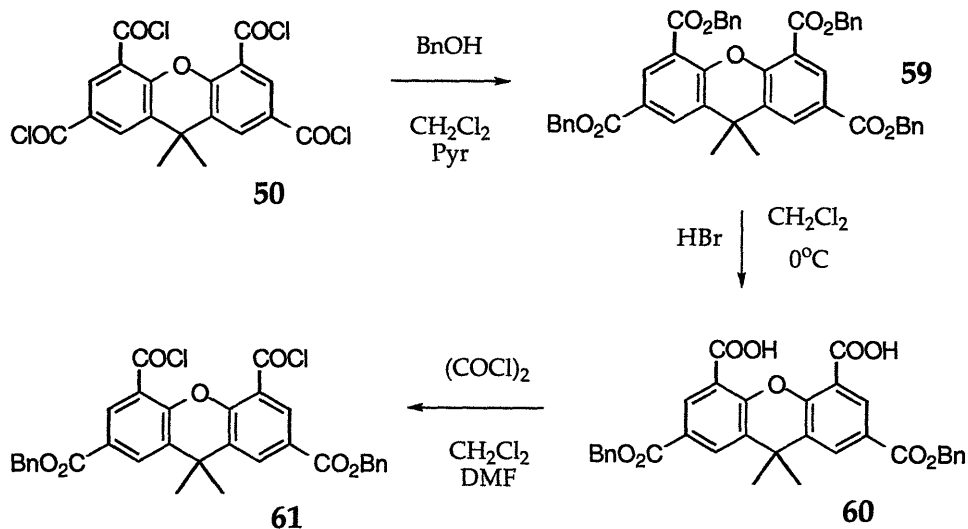
Scheme 4. Synthesis of the partially protected xanthene core 61.

As shown in Scheme 4, by reaction of 50 with benzylalcohol, the tetrabenzylester compound 59 was obtained, from which the two benzylester groups in the four and five positions were selectively removed by brief treatment with HBr in dichloromethane. The resulting xanthene dibenzylester diacid 60 was converted into the diacid chloride with oxalyl chloride in dichloromethane, leaving a core 61 which could be differentially functionalized on the "top" and "bottom" halves of the molecule.

Using xanthene 61, six new sublibraries were synthesized in the two step procedure outlined in Scheme 5. Compound 61 was treated in a first "randomization" step with two of the four protected amines Lys, Ile, Pro, and Val, followed by deprotection of xanthene positions two and seven by hydrogenolysis. Coupling of the resulting material with the two other building blocks (tris-pyrrolidino-benzotriazole-1-yl-oxy-phosphonium hexafluorophosphate, PyBOP) and deprotection of the building block protection groups with trifluoroacetic acid yielded six sublibraries, each with a unique distribution of the four selected building blocks around the xanthene core. Table 11 lists the building blocks used in the first and second randomization steps.
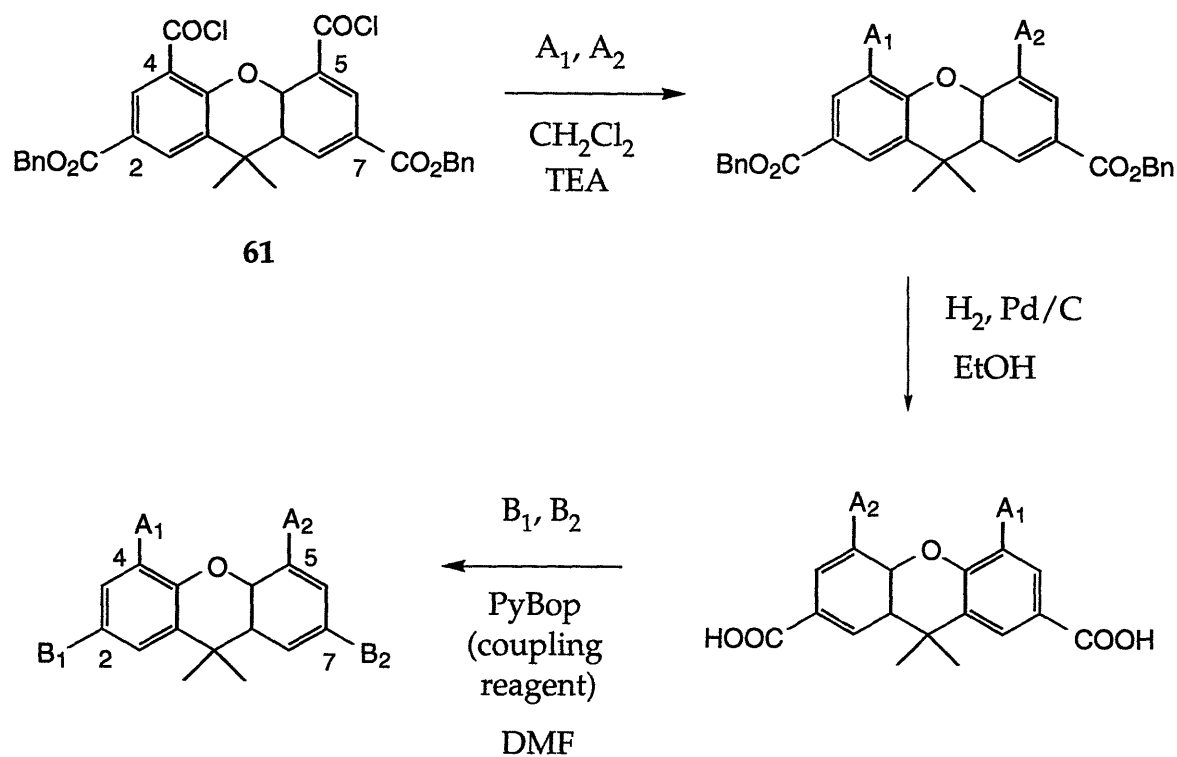
Scheme 5. Schematic synthesis of sublibraries for Round 5 (see also Table 11).

Table 11. List of the amino acid derivatives attached to xanthene positions four and five and positions two and seven in the generation of sublibraries for Round 5.

Sublibrary	Positions 4 and 5	Positions 2 and 7
1	Lys, Val	Ile, Pro
2	Lys, Ile	Val, Pro
3	Lys, Pro	Ile, Val
4	Val, Ile	Lys, Pro
5	Val, Pro	Lys, Ile
6	Ile, Pro	Lys, Val

The screening results of these six sublibraries are presented in Round 5, Figure 44. They revealed that only those compounds possessing Lys at the four or five positions and Pro at positions two or seven were active as trypsin inhibitors. Other arrangements of the four selected building blocks on the xanthene core were inactive. The most active library contained the Lys/Ile combination at xanthene positions four and five and Val/Pro at positions two and seven. This result narrowed the structure of a final most potent inhibitor to the two isomers 64 and 65 shown in Scheme 6.
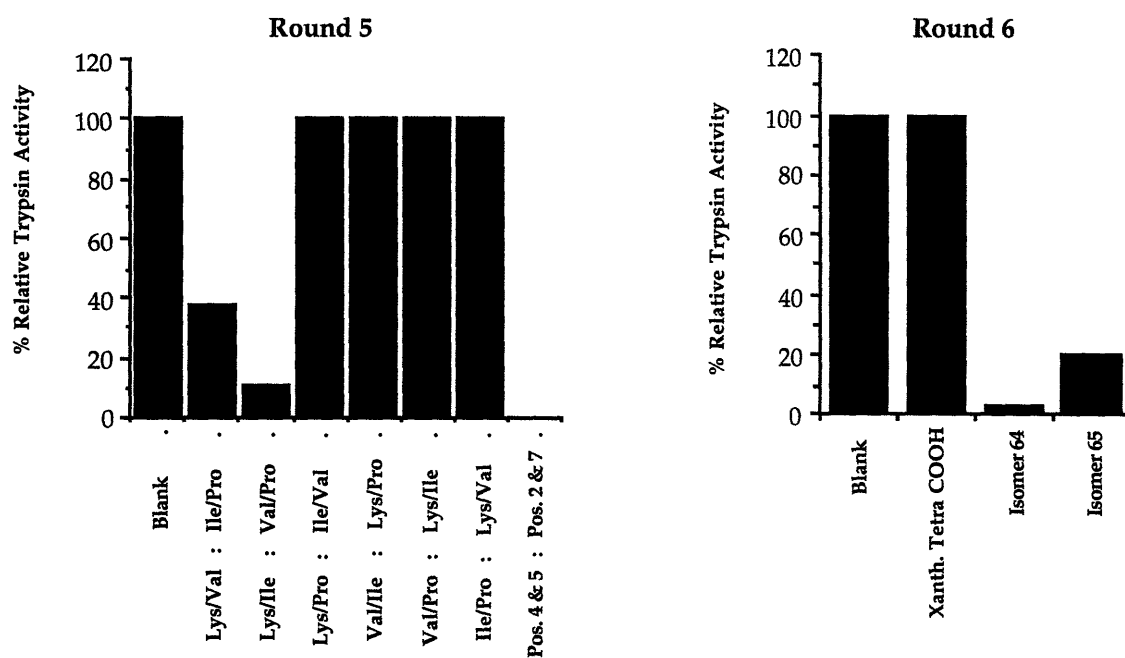
Figure 44. Round 5: % of trypsin activity with the six sublibraries constructed from core molecule 61. Sublibraries were variously substituted with Lys, Ile, Pro, and Val at xanthene positions 2, 4, 5, and 7 as noted. Round 6: % of trypsin activity with the two final inhibitors and 9,9-dimethyl-2,4,5,7-xanthene tetracarboxylic acid as a control.
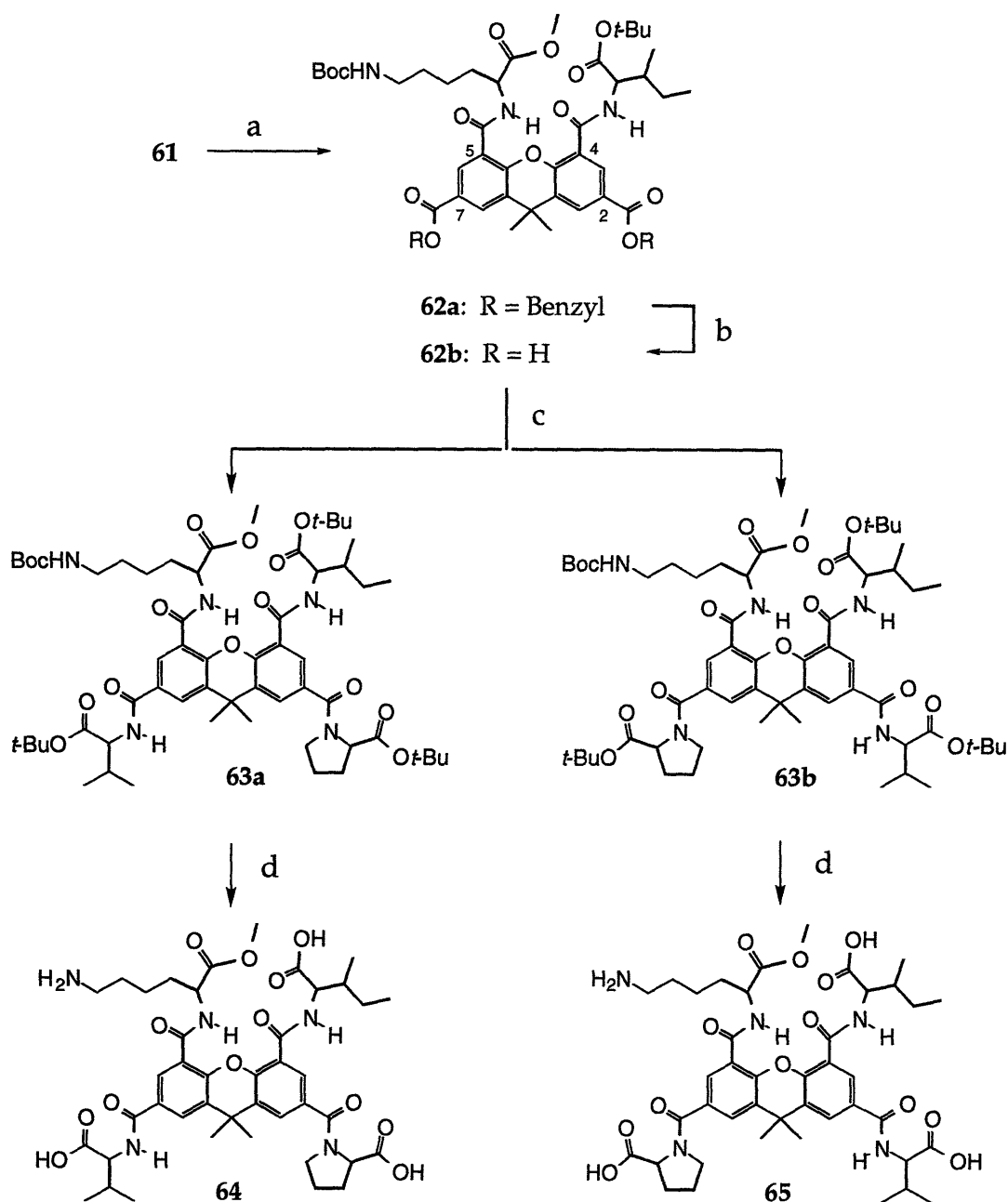
Scheme 6. Synthesis of the final trypsin inhibitor 64 and its isomer 65. a: H-Lys(Boc)-OMe, H-Ile-O-*t*-Bu, CH2Cl2, triethylamine. b: EtOAc/EtOH, 10% Pd/C, H2-atmosphere. c: H-Pro-O-*t*-Bu, H-Val-O-*t*-Bu, DMF, PyBOP, triethylamine, separate 63a and 63b by chromatography. d: CH2Cl2, Trifluoroacetic acid, thioanisol.

Compounds **64** and **65** were individually prepared via a four step synthesis from **61**, as outlined in Scheme 6. The dibenzyl protected diacid chloride **61** was allowed to react with a mixture of Lys and Ile and the mixed amide xanthene-monolysine-monoisoleucine compound **62a** isolated by flash chromatography (31% yield). Hydrogenolysis of the benzylester protecting groups yielded the monolysine-monoisoleucine xanthene diacid **62b**. Coupling of **62b** with a Pro/Val building block mixture yielded a set of four protected compounds from which two isomeric compounds **63a** and **63b** -- both containing all four building blocks Lys, Ile, Pro, and Val -- were isolated by flash chromatography and purified by normal phase preparative HPLC.

The assignment of the isomers to the structures of protected **63a** and **63b** was possible by evaluating two NOE measurements and a COSY spectrum of one of the isomers. Individual irradiation at the absorption frequencies (δ = 8.00 and δ = 7.93) of two aromatic xanthene protons connected to the same six-membered ring gave strong nuclear Overhauser effects with two NH-protons. These protons were assigned to the valine and isoleucine sub-structures through a COSY spectrum. This isomer therefore corresponded to protected compound **63b**, in which isoleucine and valine are connected to the same benzene ring.

In the final step of the synthesis, **63a** and **63b** were deprotected with trifluoroacetic acid in dichloromethane, and the products **64** and **65** were purified by reverse phase preparative HPLC. Screening of **64** and **65** in the trypsin assay (Round 6, Figure 44) revealed that both isomers are trypsin inhibitors, with **64** being the most potent. The K_i values of both compounds **64** and **65** were obtained by non-linear regression of kinetic data according to the equation for competitive inhibition and additionally by evaluation of Lineweaver-Burk plots.

For an enzyme such as trypsin which obeys Michaelis-Menten kinetics, rate of formation of product from substrate [S] may be expressed in the Michaelis-Menten equation as:[1,76]

Eqn. 7 $\qquad V = V_{max} [S] / ([S] + K_m)$

A plot of the above V vs. [S] is shown for inhibitor 64 in Figure 45. Kinetic data were obtained at six constant concentrations of inhibitor 64 by measurement of the rate of p-nitroaniline released by tryptic cleavage of the substrate N^{α}-benzoyl-DL-arginine-p-nitroanilide for increasing substrate concentrations.[76] The top curve of Figure 45 (open circles) was measured without inhibitor 64, while the bottom curve (closed triangles). was measured with the highest concentration of inhibitor.
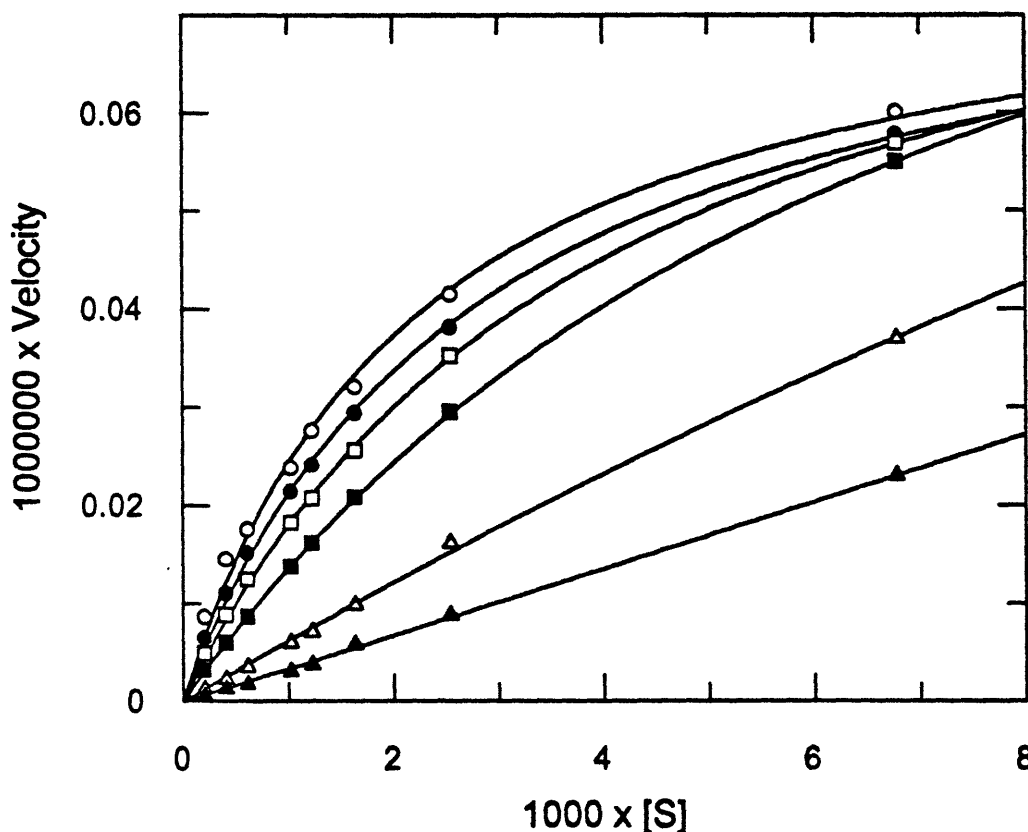
Figure 45. Plot of V (M/sec) vs. [S] (M) for inhibitor 64 obtained at 0, $1.08*10^{-6}$, $2.16*10^{-6}$, $5.40*10^{-6}$, $21.7*10^{-6}$, and $54.3*10^{-6}$ M concentrations of 64 Measurement is of the rate of p-nitroaniline released by tryptic cleavage of the substrate benzoyl-L-arginine-p-nitroanilide for increasing substrate concentrations.

To show that **64** competes with trypsin for substrate, the above data was graphed as a Lineweaver-Burk plot. Taking the reciprocal of both sides of Equation 7 gives:

Eqn. 8 $\qquad 1 \, / \, V = 1 \, / \, V_{max} + (\, K_m \, / \, V_{max} \,) * (\, 1 \, / \, S \,)$

A graph of $1/V$ vs. $1/S$ yields a straight line with an intercept of $1/V_{max}$ and a slope of K_m/V_{max}. If the *intercept* is unchanged in the presence or absence of inhibitor, then the inhibitor is competitive with the substrate in question. Thus, the plot in Figure 46 shows that **64** is a competitive inhibitor of trypsin. Note that the lines are now reversed from Figure 45, with the top curve (closed triangles) corresponding to the highest concentration of inhibitor and the bottom curve corresponding to the uninhibited reaction.
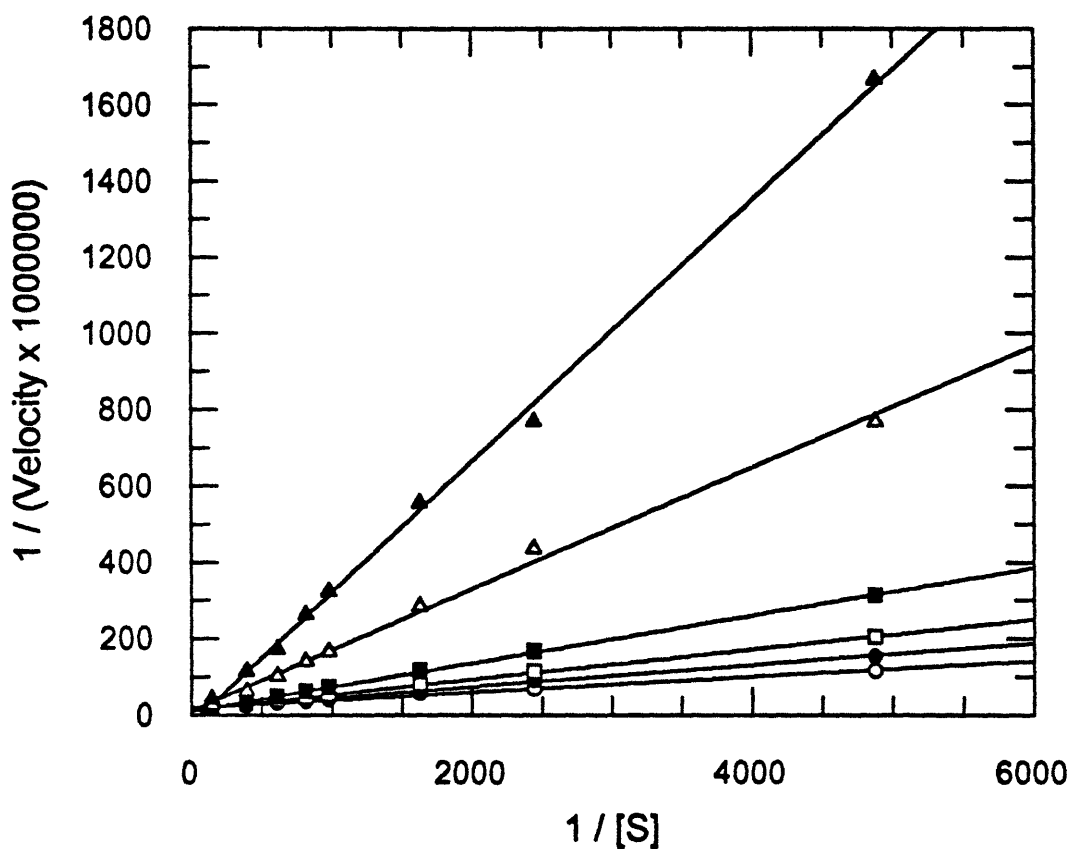

Figure 46. Plot of $1/V$ (sec/M) vs. $1/[S]$ (M^{-1}) for inhibitor **64** obtained at 0, $1.08*10^{-6}$, $2.16*10^{-6}$, $5.40*10^{-6}$, $21.7*10^{-6}$, and $54.3*10^{-6}$ M concentrations of inhibitor **64**.

In a Lineweaver-Burk plot, the ratio between the slope D_I of a line at inhibitor concentration [I] and the slope D_0 of the uninhibited reaction can be related to the K_i of enzyme inhibition as follows:

Eqn. 9 $\qquad D_I / D_0 = 1 + [I] / K_i$ $\qquad$ where $\qquad K_i = [E][I] / [EI]$

Based on the plot in Figure 46, a K_i of 9.4 ± 0.8 μM was found for inhibitor **64**. Through the same process, the K_i for **65** was found to be 72 ± 7 μM. Thus, structure **64** represents the most potent trypsin inhibitor that was selected from our initial xanthene library of 52,650 compounds, showing competitive inhibition and binding in the low micromolar range.

II.ii.4 Deconvolution of the Library: An Inspection of the Process

Through the above iterative screening procedure, we were able to extract from our libraries an inhibitor of trypsin. While this result is clearly propitious for a method of lead structure discovery, two main concerns must be addressed. Most importantly, one must ask if inhibitor **64** is truly the most active compound which was present in the initial xanthene library of 52,650 compounds; were there other compounds of interest in the library which were overlooked? Secondly, one must ask if the "hit" which was found was mere luck, or the inevitable result of a sound process of creating and deconvoluting diverse molecular libraries; in short, whether or not the method is repeatable for any given assay.

Inhibitors **64** and **65**, while reasonable inhibitors of trypsin ($K_i \sim 10$ μM), cannot mathematically account for the observed activity in the original xanthene based library of Round 1. All compounds were present in very low concentration (on the order of

0.1 µM), and thus, to produce the observed 38% reduction in activity, other inhibitory compounds beside **64** and **65** must have been present in the starting library of some 50,000 species. Either there existed in the library a broad range of compounds with lesser activity than **64**, or several highly active molecules were present which were passed over in the selection strategy.

In the selection strategy, only the most active building blocks or groups of building blocks were selected for the generation of further sublibraries. These choices guided us directly to the inhibitors **64** and **65**. The question then arises whether the described strategy automatically results in the isolation of the most active compound in a given library. Would a different initial grouping of the building blocks have resulted in the isolation of structurally different inhibitors? Could the strategy fail to select other compounds present in the library which possessed a higher inhibitory activity than **64**?

In order to address these questions, a computer program was written to simulate the activities observed in our process of deconvolution by sublibraries (code is given at the end of the experimental). The simplistic program was based on the assumptions that 1) inhibition was due to molecules binding at the active site of trypsin, and 2) each building block at a certain xanthene position added an incremental value to the "energy of binding" for that molecule. Based on the empirical evidence of molecules **64** and **65**, the values of "binding energy" for each building block at each xanthene position were adjusted until the program was able to roughly reproduce sublibrary activities. The output of the program is displayed in Table 12; prime notation, *e.g.*, Round 2', denotes computer generated results. The program reproduced the results of the actual screening rounds such that Groups 1-3 were selected over Groups 4-6 in Round 2', and Lys, Pro, Val, Ile, and Leu were selected from the sublibraries in Round 3'. The program reproduced Rounds 2 and 3 qualitatively, but not quantitatively.

Table 12. A(rel): Computer generated relative activity (%) for simulated xanthene libraries. Building blocks are omitted from the libraries as noted. An X denotes selection of a given library to influence the next round of screening.

Round 2'			Round 3'		
Library	A (rel)		Library	A (rel)	
- Gly, Ala, Val	81	X	- Arg	39	
- Leu, Ile, Pro	96	X	- Lys	100	X
- Arg, Lys, His	100	X	- His	32	
- Ser, Thr, Met	44		- Leu	57	X
- Phe, Tyr, Trp	49		- Ile	71	X
- Glu, Asp, Asn	39		- Pro	76	X
			- Gly	32	
			- Ala	42	
			- Val	86	X

The fact that the computer program was qualitatively successful using the simple parameters of incremental energy of binding gave several insights into our selection strategy. To begin, it seems likely that molecule 64 was merely the tip of a broad family of inhibitors; beneath the tip lay other less-potent inhibitors which nevertheless contributed greatly to the activity of the computer generated libraries. This explains the relatively high activity of the initial xanthene library in Round 1. Assuming the computer program mirrors reality, if one postulates a "molecular landscape" [31] in which valleys mark compounds that bind weakly and mountains mark families of inhibitors, our selection procedure is less a screening for individual inhibitors than it is a group selection procedure which provides a means of ascending a feature in a molecular landscape to its peak of activity.

With a working simulation in hand, the program was run again to model the success of the screening procedure when Round 2 was deconvoluted with different groupings of building blocks. In the simulated screening procedure, the combination Lys, Ile, Pro, Val was selected independently of how the building blocks were grouped. For example, when grouped as in Round 2″ (Table 13), four sublibraries p1, p2, p4, and p6 were selected, leading to sublibraries q1-q12 in Round 3″. The sublibraries q3, q4, q5, q8, and q10 were be selected from this round, once again converging to building blocks Ile, Leu, Val, Pro, and Lys (compare to actual screening Rounds 2 and 3). Thus, the program gave evidence that when given a single, well defined mountain in a molecular landscape, our screening procedure is not dependent on the grouping of building blocks to find the highest point of activity.

Table 13. A(rel): Computer generated relative activity (%) for simulated xanthene libraries. Building blocks are omitted from the libraries as noted. An X denotes selection of a given library to influence the next round of screening.

Round 2″			Round 3″		
Library	A (rel)		Library	A (rel)	
p1 -Phe, Met, Ile	76	X	q1 -Phe	47	
p2 -Leu, Val, Trp	92	X	q2 -Met	47	
p3 -Ala, Thr, Gly	46		q3 -Ile	74	X
p4 -Ser, Pro, Tyr	65	X	q4 -Leu	60	X
p5 -His, Asn, Glu	39		q5 -Val	85	X
p6 -Lys, Asp, Arg	100	X	q6 -Trp	47	
			q7 -Ser	40	
			q8 -Pro	77	X
			q9 -Tyr	39	
			q10 -Lys	100	X
			q11 -Asp	40	
			q12 -Arg	46	

When the computer program was changed to introduce a second peak in the molecular landscape (i.e., a second set of inhibitors with different building blocks was defined with an energy of binding equal to or greater than that of **64**), the landscape feature (and therefore the ultimate inhibitory molecule) which the computer selected *did* depend on initial groupings of building blocks. Simulations demonstrated that by employing our selection strategy, it was impossible to miss an active compound orders of magnitude more potent than compound **64**, but it was possible to overlook a compound with comparable or even slightly higher activity. The computer program thus affirmed that the described selection strategy of ever narrowing sublibraries will result in the isolation, if they exist, of *one or more* of the most active compounds present in a combinatorial library created from a core molecule and a set of building blocks. The screening procedure initially operates by selecting groups of molecules rather than individual compounds, but the groups selected become more and more focused, until a single highly potent structure is elucidated.

While the isolation of **64** shows that the selection strategy works, it is clear from analysis that some active structures could be missed in the process. The question then becomes, is this drawback a fatal flaw? We would argue that it is not. As long as a procedure is effective in identifying *at least one* of the active lead compounds in a library, and as long as the *potential for generating new libraries is limitless*, it is of little consequence that some other active compounds escape unnoticed. As an analogy, nature could never hope to sort through all 20^{200+} possible peptide structures to find *the best* enzyme for a given task, but it is able to select *useful* enzymes from this pool in a relatively short period of time.

The fact remains that the generation and iterative screening of a library synthesized from core molecule **50** and a set of 18 building blocks produced the trypsin inhibiting compound **64**. Thus, the 50,000 compounds in the original library provided a sufficient sampling of molecular recognition surfaces such that one among their number

could form a non-covalent bond to some portion of the active site of the enzyme. The only question left to be answered was that of reproducibility; could a similar combinatorial synthesis and deconvolution procedure be used to find an inhibitor of a second, unrelated enzyme, or was the discovery of **64** the purest chance? In the next section, this question is laid to rest.

II.iii   Synthesis and Screening of Expanded Libraries

Having tasted success in screening libraries composed of the xanthene core **50** and amino acid building blocks, it was desired to expand the range of "shape space" which the libraries sampled, offering a wider range of molecular recognition among the library compounds. This was done by creating a second core molecule, and by synthesizing libraries with a host of new building blocks.

II.iii.1 Generation of New Water Soluble Libraries

To add a new dimension to the creation of libraries described in the previous section, a new core molecule, 2,2',4,4'-biphenyl-tetracarboxylic acid chloride **66** was synthesized. Due to rotation at the phenyl-phenyl bond, modeling showed this core to present its four substituents in a more spherical orientation compared to the xanthene core **50**, in which all four carboxyl groups lie in a plane (Figure 47). Preparing libraries with **66** is no more (and no less) intrinsically likely to create useful surfaces of molecular recognition; it simply creates compounds with different surfaces than those prepared from xanthene core **50**.
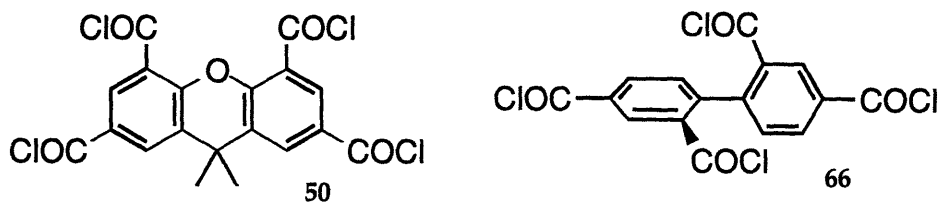
Figure 47. Polyfunctionalized cores **50** and **66** compared.

As shown in Scheme 7, the synthesis of **66**, based on a procedure from fellow graduate Robert Grotzfeld in our group, begins with permanganate oxidation of 4-bromo-*m*-xylene **67** to give 4-bromo-isophthalic acid **68**. The dimethyl ester **69** is formed in MeOH with catalytic H$_2$SO$_4$, and Ullmann coupling of this product with copper powder at 220 °C in a sealed tube gave the 2,2′,4,4′-biphenyl-tetra-methylester **70**. This was saponified to the tetra acid **71**, which was converted to the tetra acid chloride **66** by melting with PCl$_5$.
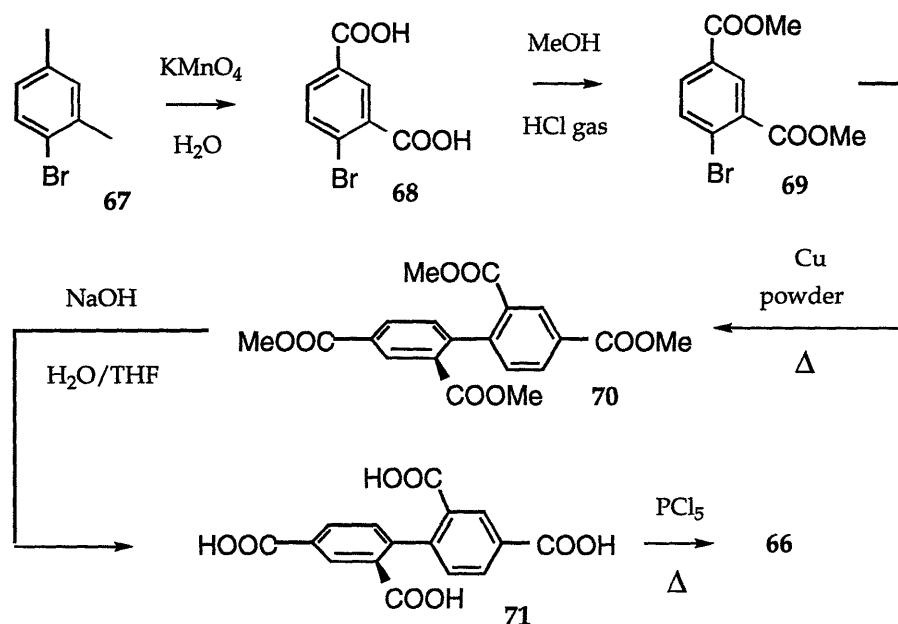
Scheme 7. Synthesis of the tetra acid chloride biphenyl core **66**.

With two core molecules in hand, each available for tetra-substitution, the pool of building blocks was the next target for expansion. In addition to the 18 L-amino acids already used, six D-amino acids were added (Leu, Lys, Phe, Pro, Ser and Tyr) as well as eleven non-amino-acid amines (Figure 48). The latter eleven primary amines were selected because they all added the functionality of at least one additional heteroatom while not requiring protecting groups; it was deemed that all other sites in these building blocks were unreactive compared to their primary amine functionalities.
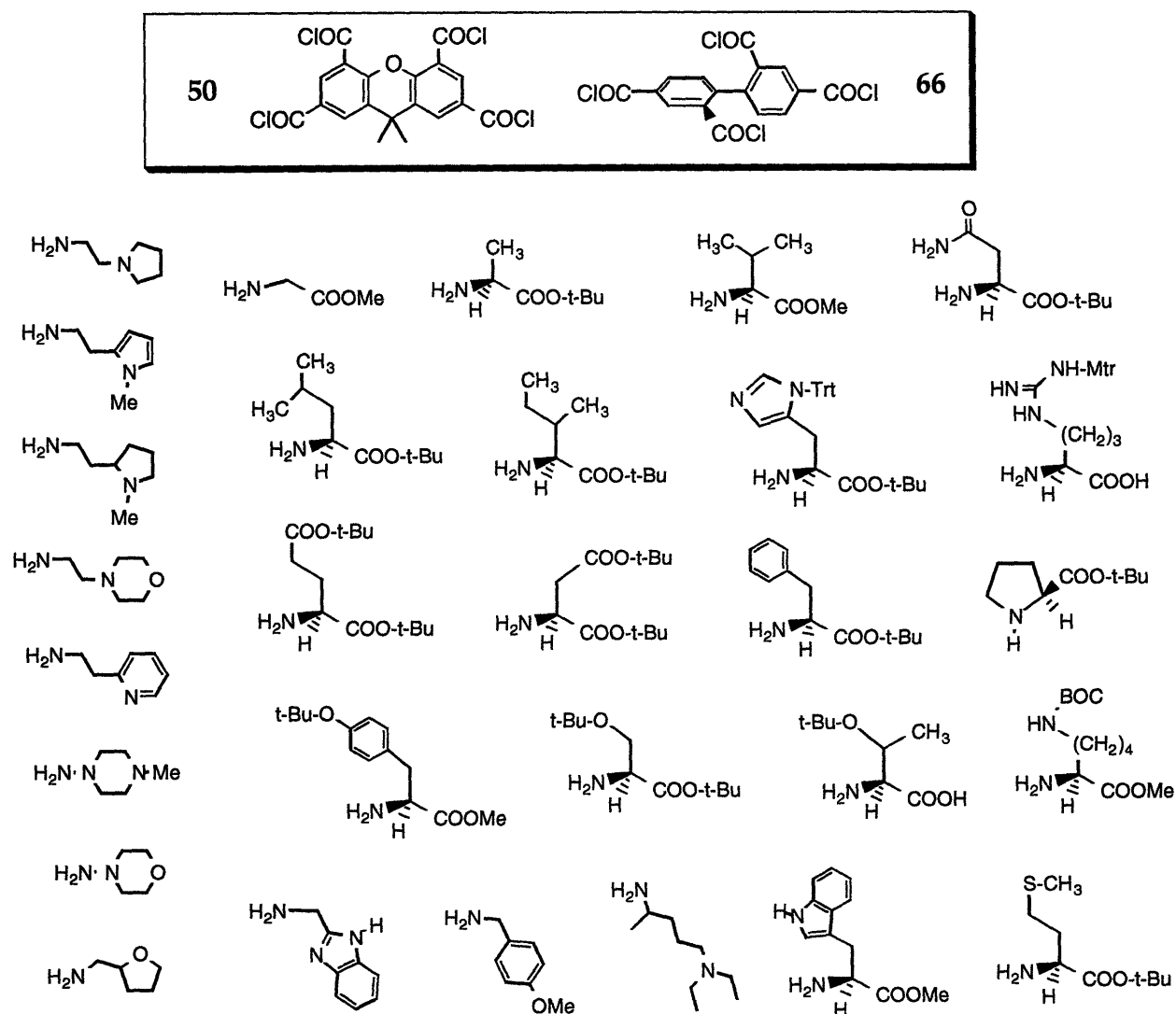

Figure 48. Building blocks and core molecules used for the generation of libraries.

From the pool in Figure 48, five groups of 14 amines and/or amino acids (D-or L-) were selected, and each of the five groups was combined with first the xanthene and then the biphenyl core to yield a total of 10 libraries. (See experimental for a list of building blocks in each library). As with the libraries in the previous section, these libraries were initially dichloromethane soluble due to their many hydrophobic protecting groups, and as before the libraries were washed with aqueous acid to remove any unreacted amines. The libraries were then deprotected with trifluoroacetic acid and precipitated with ether/hexane 1:1, yielding the water soluble libraries as powders.

With 14 building blocks and one tetra acid chloride core, each of the 10 libraries contained theoretically 19,306 different compounds. Having thus created five xanthene and five biphenyl libraries spanning well over 100,000 compounds in total, the next step was to design an interesting assay which could serve as a second test of our method of library generation from a poly-functionalized core and a set of building blocks.

II.iii.2 A New Screening Assay: Inhibition of DNA Polymerase I Klenow Fragment

Of the many interactions of enzymes in cells, those which regulate the workings of DNA are among the most critical. *Escherichia coli* DNA polymerase I[77] is one such enzyme, efficiently polymerizing deoxynucleoside triphosphates (dNTPs) to form the complementary strand of a single stranded primed DNA template.[1] The Klenow Fragment,[78,79] a proteolytic product of DNA polymerase I which retains polymerase activity and 3' to 5' exonuclease activity but lacks 5' to 3' exonuclease activity,[78] is well studied. Its mechanism of action is well

characterized,[80,81] its crystal structure has been known since 1985,[82-85] and it is widely used as a tool in molecular biology. However, very few Klenow Fragment inhibitors, the best of which is pyridoxal-5'-phosphate, are known;[86,87] pyridoxal-5'-phosphate exhibits a K_i of 32 µM and inhibits competitively with the binding of dNTPs to the enzyme. Given the inhibition found against trypsin, it was hoped that our libraries might be able to better this compound in activity, and at the same time prove the concept of our combinatorial chemistry. Thus, it was determined to create an assay for the inhibition of DNA polymerase I, and with it screen the new xanthene and biphenyl libraries created above.

While I had in mind a general idea of the assay that I wished to design, knowledge of the concepts of molecular biology does not confer on one the ability to successfully manipulate enzymes and DNA, let alone certify one for the use of radioactive labeling isotopes. Thus, a collaboration was begun with the group of John Essigmann. With the dedicated help of Toxicology graduate student Deborah Kreutzer, an assay was designed based on the inhibition of elongation of a DNA template (Template sequence is given in the experimental). A schematic representation of the assay is depicted in Figure 49.

In the assay, the substance (or library) to be tested is dissolved in DMSO and diluted to 10% DMSO with Tris buffer adjusted to pH 7.8. (The buffer also contains 5 mM Mg^{2+} salt requisite for polymerase activity.) DNA polymerase I Klenow Fragment (10 nM) in buffer is then incubated with several microliters of the above library solution to allow any active compounds to bind to the protein. After incubation of the polymerase, a DNA template (10 nM) is added which consists of 5'-^{32}P-radiolabeled DNA 17-mer annealed to a complementary unlabeled DNA 45-mer. Upon addition of deoxynucleotide triphosphates (dNTPs, 40 µM each), the polymerase begins to elongate the DNA 17-mer into a complete complementary 45-mer, *unless* the action of the enzyme is blocked by the substance being assayed.

DNA Polymerase I
Klenow Fragment

^{32}P labeled DNA 17-mer annealed
to unlabeled DNA 45-mer.

Pol I

Incubate 10 min.
Buffer, pH 7.8

5mM MgCl$_2$

+

Library of
compounds

Pol I

Mix with dATP, dCTP,
dGTP, dTTP

Run 5 min. at 25 °C

Quench with
0.1M EDTA

Run reaction products
on a denaturing gel

45-mer

17-mer

Pol I

OR

Pol I

Polymerization:
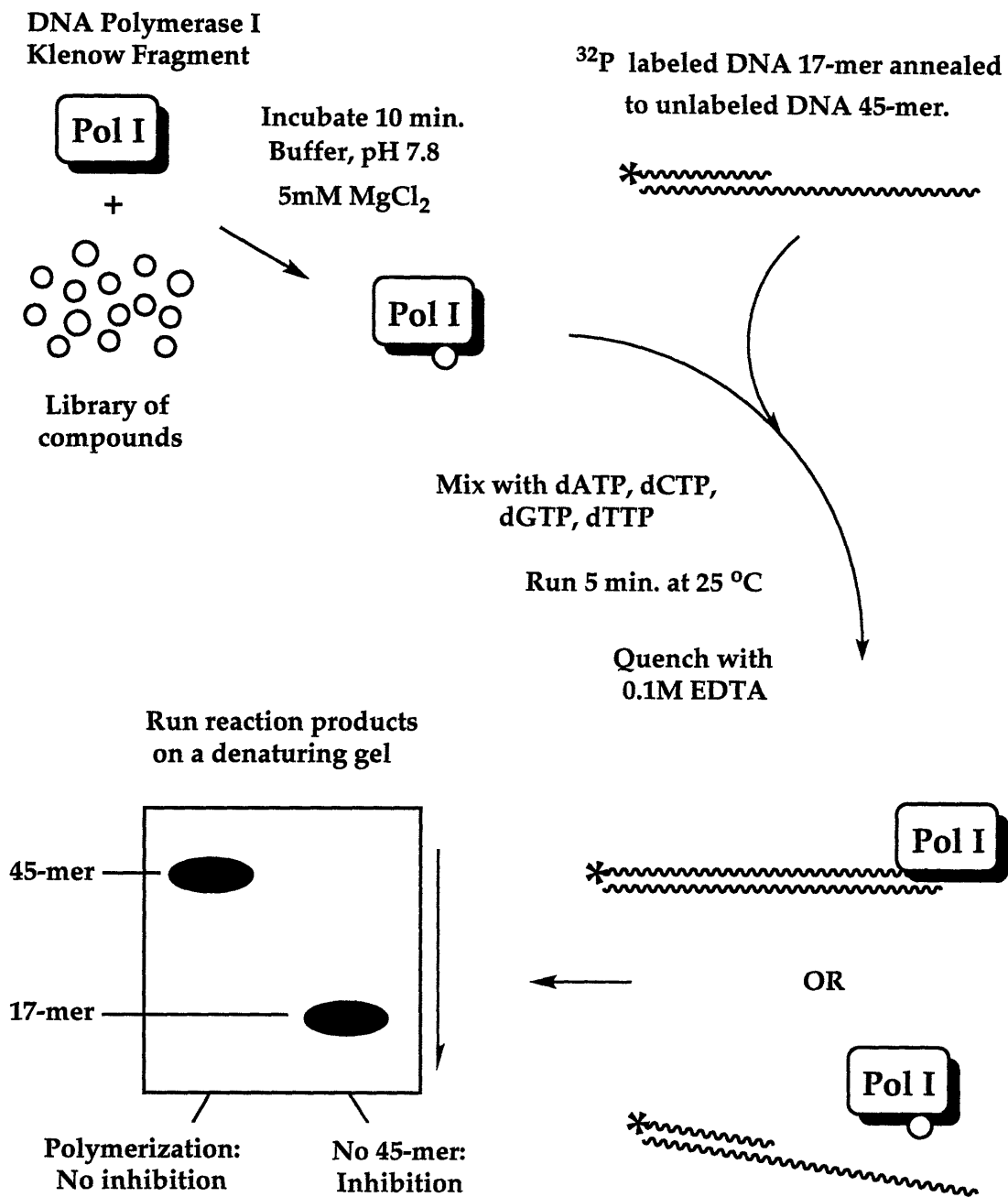No inhibition

No 45-mer:
Inhibition

Figure 49. A Schematic Assay for Polymerase Inhibition.

After five minutes, the reaction is quenched by removing Mg^{2+} from the reaction mixture with 0.1 M EDTA. The reaction solution is then electrophoresed through a denaturing polyacrylamide gel, which denatures all DNA duplexes into single strands and separates these strands based on their mobility through the gel. A DNA 45-mer runs much more slowly than a DNA 17-mer, and thus the 45mer will "stick" to the top of the gel while the 17mer runs down through it. ^{32}P imaging of the radiolabeled DNA reveals whether or not the labeled primer was extended to the full 45mer; if the substance being assayed inhibits polymerase activity completely, only the starting 17mer will be present.

The ten new libraries, five xanthene based and five biphenyl based, were tested in this assay. A 3.7 mg portion of each library was dissolved in 50 µl DMSO, the solutions were diluted to 500 µl with buffer, and 17 µl of each solution was incubated with DNA polymerase I. To these solutions were added labeled DNA template and dNTPs (to make a final volume of 25 µl), and the reactions were run 5 min. at 25 °C. Reactions were quenched with EDTA, and the products analyzed on a gel as described. The ^{32}P image of the gel is shown in Figure 50 (Round I). This and all subsequent biological work was contributed by Deborah Kreutzer in close collaboration, and it is only through her expertise in molecular biology that the project was able to progress.
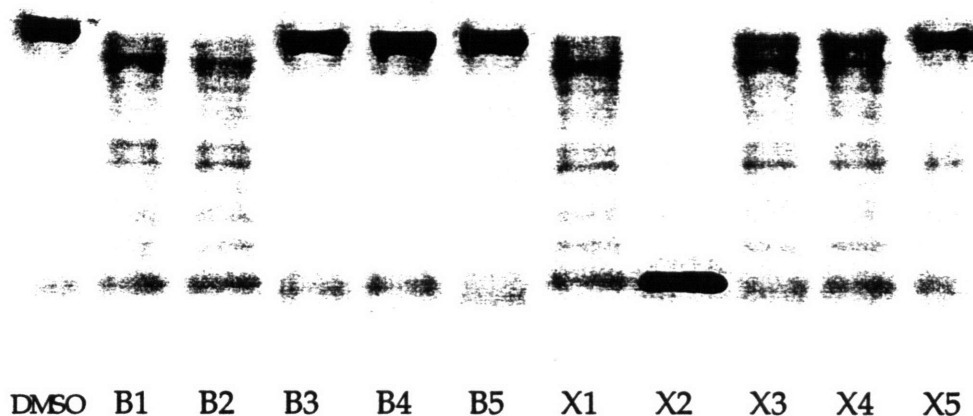
Figure 50. Polymerase screening Round I. Added: 17 µl aliquots of library solutions 3.7 mg in 500 µl.

In the results of Round I of the screening (Figure 50), several lanes on the gel show a lessening of the presence of the completed DNA 45-mer (top band). It is thus clear that several of the libraries inhibited the extension of the 17-mer (bottom band) to the 45-mer. One lane in particular, that of xanthene library 2 (**X2**), shut down polymerase activity nearly completely, showing only the starting DNA 17-mer. As noted, several other lanes -- corresponding to biphenyl libraries **B1** and **B2** and xanthene libraries **X1**, **X3**, and **X4** -- also show inhibition, though to a lesser extent. No less importantly, four of the ten libraries show complete elongation to the DNA 45-mer; even in the presence of the 19,000 compounds introduced by each of these libraries, the polymerase remains active. This indicates that the inhibition seen
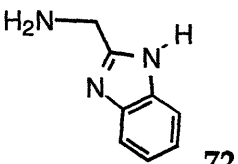
with the other six libraries is specific: observed inhibition is not due simply to the presence of 19,000 different "impurities" in a given sample, but rather to one or more individual molecules. This is not, perhaps, surprising, as the polymerase must remain active in the sea of molecules which are present in a cell.

II.iii.3 Initial Deconvolution of a Polymerase Inhibiting Library

The most inhibitory library from Round I above was xanthene library 2; this library was therefore selected as the most promising for deconvolution. An early control experiment showed that the library was equally as effective at inhibiting the polymerization of an entirely unrelated DNA template, in this case a 14-mer annealed to a 42-mer (see experimental for DNA sequence). This control supported an inhibitory interaction of library compounds with DNA polymerase I, and ruled out any specific inhibition through binding to the DNA template. With this initial result in hand, it was decided to proceed with deconvolution by the same iterative selection process used successfully in the case of the previous trypsin assay.

The 14 building blocks which were used to created **X2** are listed in Table 14. They include eight L-amino acids, five D-amino acids, and 2-(aminomethyl)-benzimidazole (**72**, henceforward referred to as AMB). To begin the iterative selection process, seven sublibraries of **X2** were synthesized by combining the xanthene tetra acid chloride core **50** with 12 of the 14 building blocks in Table 14. In each library, two building blocks were left out. The seven sublibraries were washed and deprotected as previously described (see Section **II.ii**), and after precipitation 3.7 mg of each were tested in the polymerase assay as described above for Round I. The results are shown in the gel in Figure 51 (Round II).

Table 14. Building blocks used to create xanthene library **X2**.

|	building block	protected reagent used
1	Arg	N^g-4-methoxy-2,3,6-trimethylbenzene-sulfonyl-L-arginine
2	Asn	L-asparagine-*tert*-butyl ester
3	Glu	L-glutamic acid-γ-*tert*-butyl-α-*tert*-butyl ester
4	Gly-OMe	Glycine-methyl ester
5	His	N^{im}-trityl-L-histidine
6	Ile	L-isoleucine-*tert*-butyl ester
7	Tyr-OMe	O-*tert*-butyl-L-tyrosine-methyl ester,
8	Val	L-valine-*tert*-butyl ester
9	AMB	2-(aminomethyl)-benzimidazole
10	D-Leu	D-leucine-*tert*-butyl ester
11	D-Lys-OMe	N$^\varepsilon$-Boc-D-lysine-methyl ester
12	D-Phe	D-phenylalanine-*tert*-butyl ester
13	D-Pro	D-proline-*tert*-butyl ester
14	D-Ser	O-*tert*-butyl-D-serine-*tert*-butyl ester
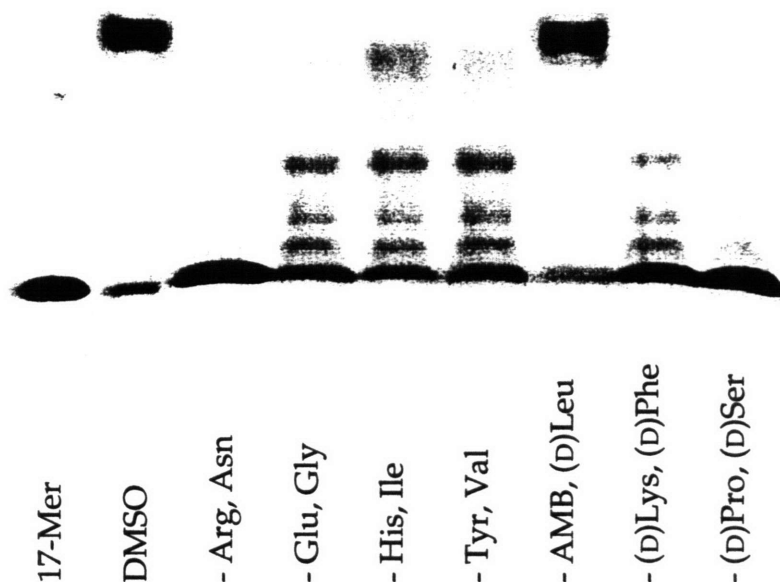
Figure 51. Polymerase screening Round II. Added: 17 µl aliquots of library solutions 3.7 mg in 500 µl.

The results of Round II show that without building blocks AMB or D-Leu, the library was inactive as an inhibitor; the full DNA 45-mer is present in this lane. Therefore, one or both of these building blocks is critical to the formation of polymerase-inhibiting xanthene derivatives. Omission of the pairs Glu/Gly, His/Ile, or Tyr/Val also resulted in a lessening of inhibitory activity, so these building blocks were carried on to the next round as well. Omission of Arg/Asn, D-Lys/D-Phe, or D-Pro/D-Ser had little effect on the ability of the resulting libraries to inhibit polymerization; very little of the DNA 17-mer is elongated in the presence of these libraries. The aforementioned six building blocks were therefore excluded from the next step of deconvolution.

For Round III, sublibraries were synthesized from the xanthene core **50** plus seven of the eight remaining building blocks. Each of the new sublibraries was thus missing one of the amines Glu, Gly, His, Ile, Tyr, Val, AMB, or D-Leu. The eight sublibraries were screened in the polymerase assay Round III, the resulting gel of which is shown in Figure 52.
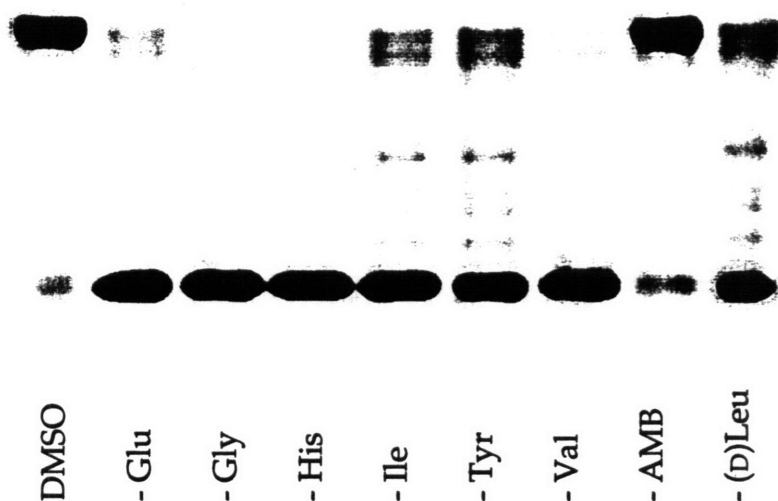
Figure 52. Polymerase screening Round III. Added: 14 µl aliquots of library solutions 3.7 mg in 500 µl.

The assay of eight sublibraries in Round III showed that the AMB building block was critical for the formation of inhibitory compounds; no inhibition of polymerization to the DNA 45-mer was seen in the absence of AMB. The omission of any of the building blocks Ile, Tyr, or D-Leu weakened the inhibitory activity of the resulting libraries, so these amines were also carried on for further study. Nearly complete inhibition was seen despite the absence of Glu, Gly, His, or Val in Round III; these building blocks were thus concluded to be unnecessary for the creation of inhibitory molecules.

The deconvolution was thus narrowed to four amines: AMB, Ile, Tyr, and D-Leu. To determine which of these building blocks were most critical to producing polymerase-inhibiting xanthene derivatives, sublibraries were synthesized with the xanthene core **50** and combinations of four, three, or two of the four amines in question. Val was kept in the building block pool as a control. Two gels from this round of screening are shown in Figure 53 (Round IV).
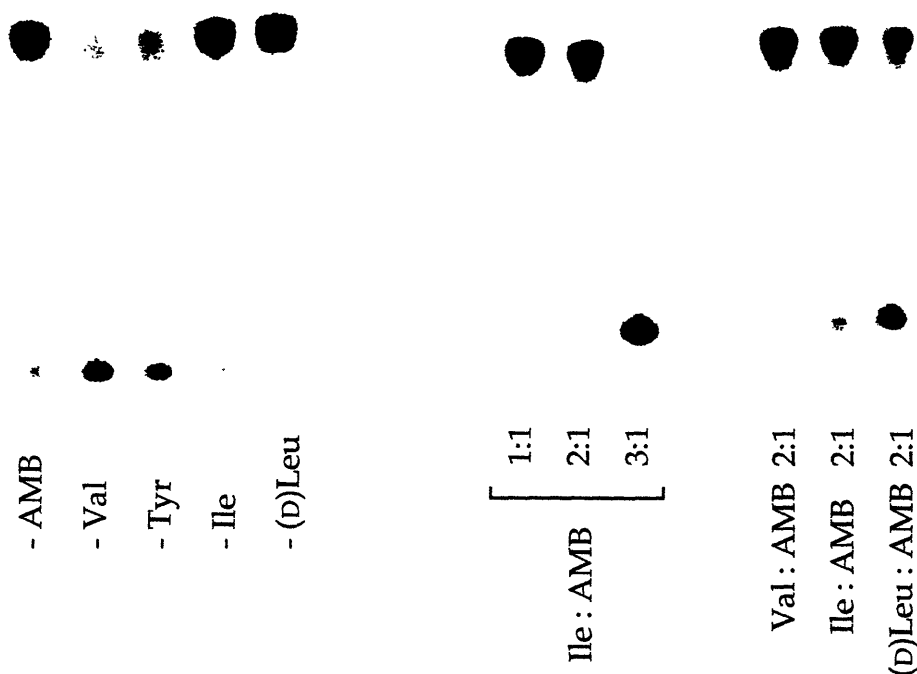
Figure 53. Polymerase screening Round IV. Added: 3 μl aliquots of library solutions 3.7 mg in 500 μl.

The first gel of Round IV shows assay results of libraries created with core **50** plus four of the five building blocks AMB, Val, Tyr, Ile, and D-Leu. With one amine left out in each library, omission of Val or Tyr did not affect inhibitory activity relative to omission of AMB, Ile, or D-Leu. It thus appeared that only AMB, Ile, and D-Leu were necessary for the creation of active compounds. In fact, excellent inhibition of polymerase activity was found by simply adding AMB and either of the

hydrophobic amino acids Ile or D-Leu. Round IV also determined what ratio of the four xanthene sites were occupied by each building block. As shown in the second gel of Round IV, the preferred ratio of amino acid to AMB was 3:1, and the preferred amino acid was D-Leu.

Having established a polymerase inhibiting library with the xanthene core and a 3:1 ratio of amines D-Leu and AMB, it was next assayed whether or not the non-natural D-configuration of the leucine building block was important. Two xanthene libraries **XD** and **XL** were created, one with AMB and three equiv. of D-Leu, and the second with AMB and three equiv. of L-Leu. Figure 54 shows a concentration course of added μl of these library solutions to the polymerase assay (as above, a library solution consists of 3.7 mg library material in 500 μl 10% DMSO/buffer pH 7.8). From the gel, it is clear that library **XL** is in fact more potent.
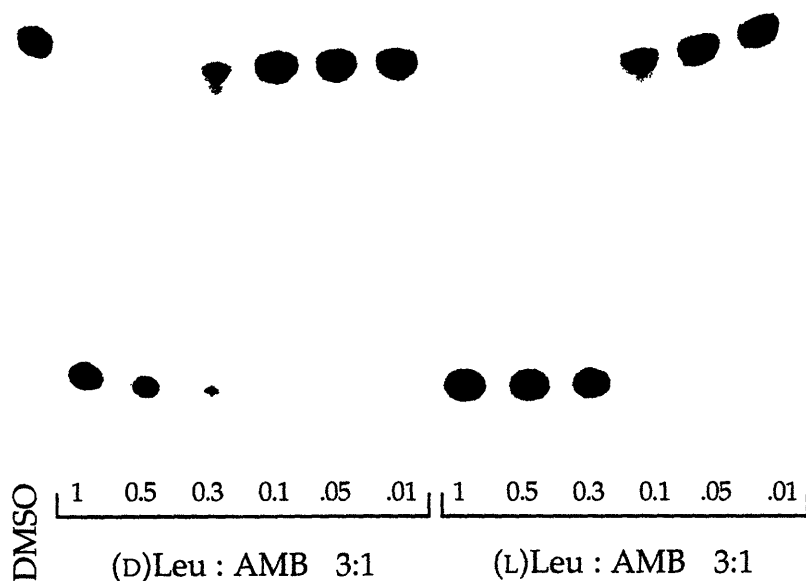
Figure 54. Concentration course of libraries XD and XL created with 50, 1 equiv. AMB, and 3 equiv. of either D-Leu or L-Leu screened against the Klenow Fragment of DNA polymerase I. μl amounts added are aliquots of library solutions 3.7 mg in 500 μl.
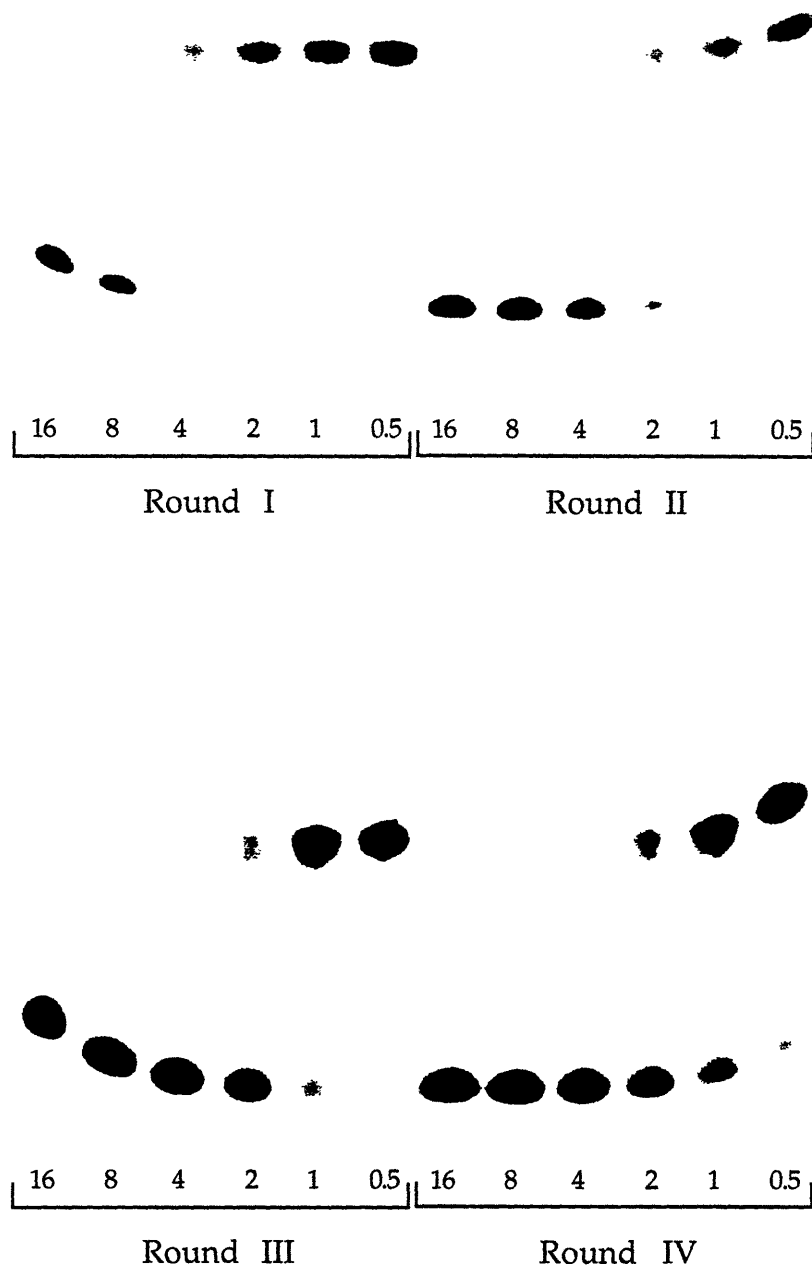
Figure 55. Concentration course of the most active libraries from Rounds I-IV, screened against the Klenow Fragment of DNA polymerase I. µl amounts added are aliquots of library solutions 3.7 mg in 500 µl.

To this point, library deconvolution had gone smoothly. Beginning with Xanthene library 2 of theoretically 19,000 compounds in Round I, iterative screening had progressed to libraries of approximately 10,000, 1000, and 150 compounds in Rounds II, III and IV, ending finally with a combination of just two different building blocks about a xanthene core. The "amplification" of activity through the rounds of screening can be seen from the gels in Figure 55. The figure shows a concentration course of the most active libraries from Rounds I - IV, adding aliquots of solutions made by dissolving a constant 3.7 mg of each library in 500 μl. As the rounds progress, the cutoff of library activity (appearance of the unextended DNA 17-mer, lower band) decreases from 8 to 4 to 2 to 1 μl of added library solution. As seen from Figure 54, this cutoff drops to 0.3 μl of added library solution for the final library created with core **50**, 1 equiv. AMB, and 3 equiv. of L-Leu (**XL**).

Thus, the method of library generation through a poly-functionalized core and a set of building blocks was again successful in creating active compounds. Coupled to the above process of iterative screening, a potent mixture of compounds had again been identified. As in the case of screening against trypsin, no further information could be learned at this point from the tetra acid chloride core **50**. Therefore, further deconvolution was attempted with the partially protected xanthene-2,7-dibenzyl-4,5-diacid chloride compound **61** (Figure 56).

61
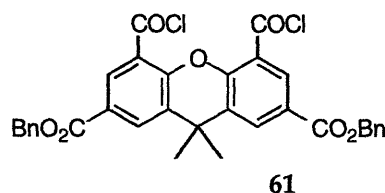
Figure 56. Partially protected xanthene core for isomer synthesis.

II.iii.4 Further Deconvolution of the Library: Unexpected Results.

With only two building blocks added to the xanthene core in library **XL**, there exist ten possible tetra-substituted compounds. Taking into account that the most favorable ratio of building blocks was 3:1 Leu/AMB, this leaves only the two isomers **73** and **74**, with AMB either at the "top" or the "bottom" position of the molecule (Figure 57).
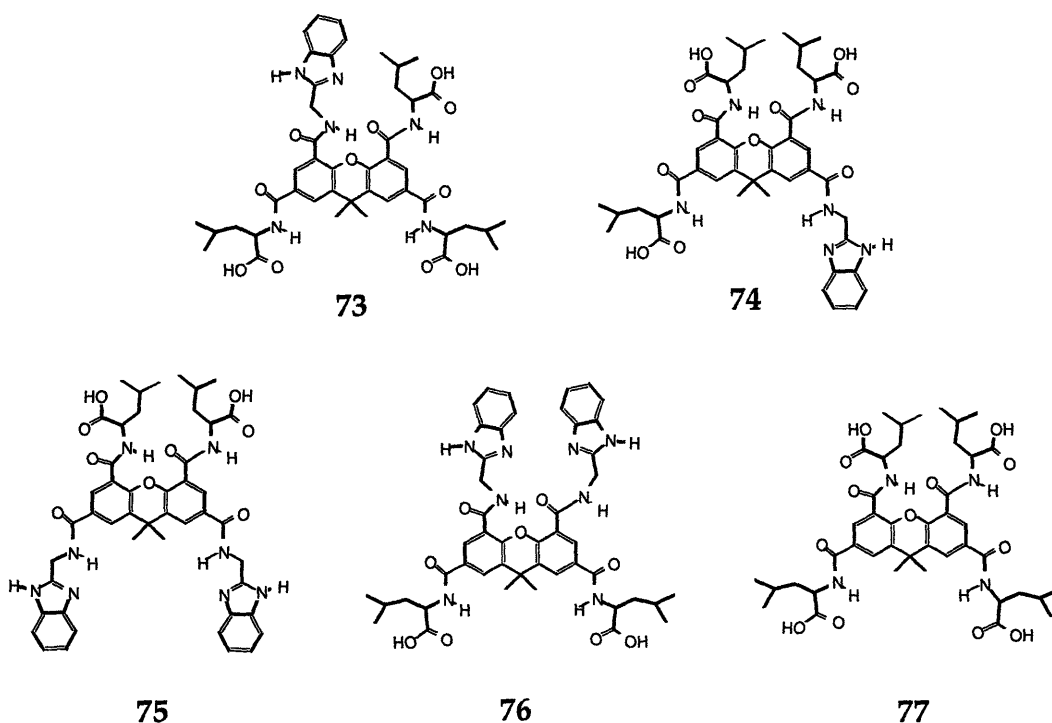
Figure 57. Xanthene based Leu/AMB derivatives. Compounds shown are with L-leucine.
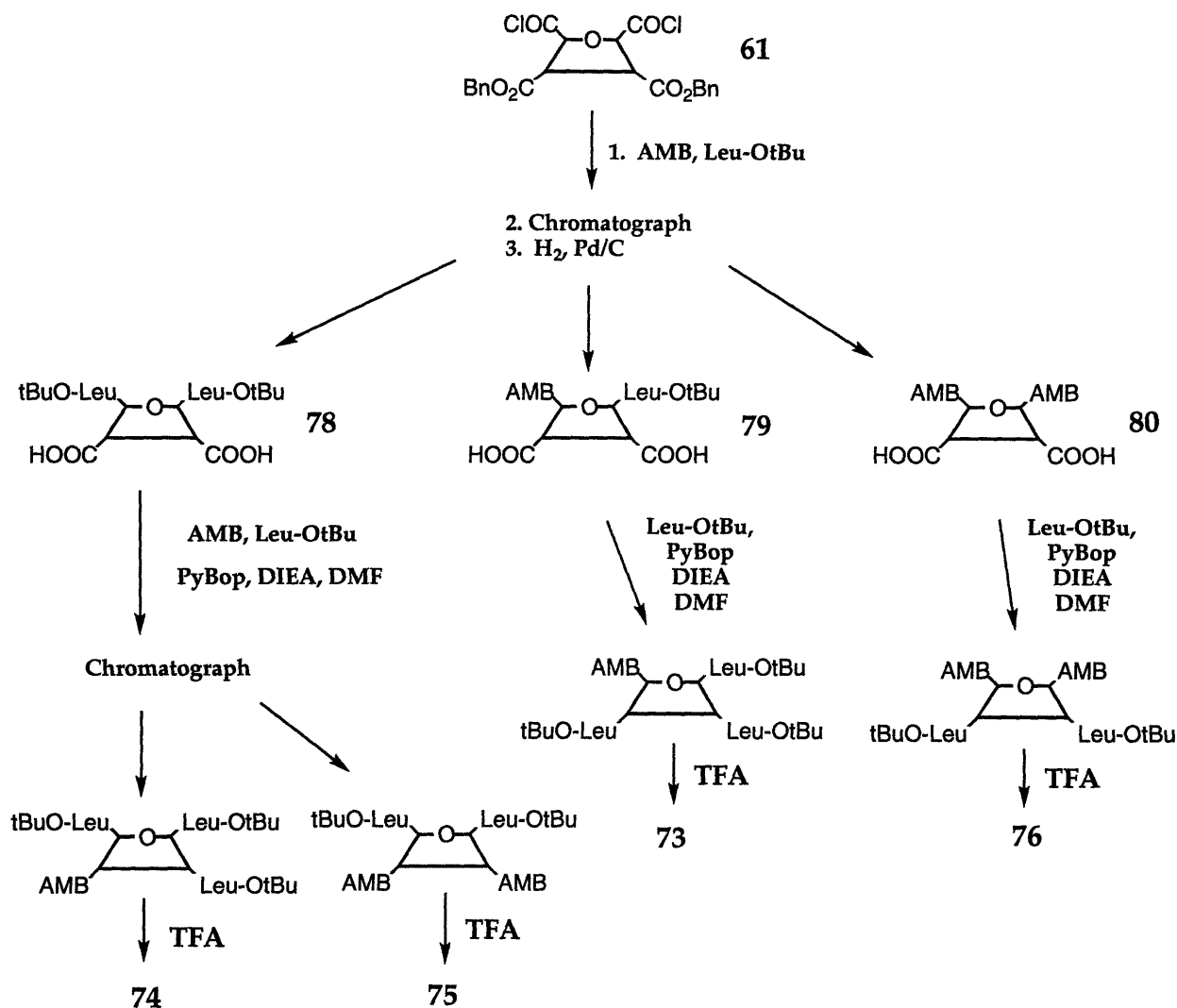
Scheme 8. Synthesis of Xanthene Leu/AMB derivatives.

Adapting the general synthetic route used to create trypsin inhibitors **64** and **65** (II.ii.2, Scheme 5), the synthesis depicted in Scheme 8 above was used to create **73** and **74** as well as controls **75** and **76**. Briefly, the xanthene diacid chloride core **61** was coupled with 1:1 mixture of Leu/AMB at its 4 and 5 positions. After chromatography and hydrogenation of the benzyl groups to give **78, 79,** and **80**, the second coupling was achieved with the reagent PyBop in DMF with diisopropyl-

ethylamine. Compound **78** was coupled separately to Leu/AMB or AMB/AMB to give **74** and **75** respectively after deprotection of the leucine *t*-Butyl groups. **79** and **80** were coupled to Leu/Leu to give, after deprotection, **73** and **76** respectively. Tetra-leucine control **77** was synthesized from core **50** in a single step.

Much to our surprise, none of the pure compounds pictured in Figure 57 were active! All signs in the deconvolution process pointed toward either **73** or **74** as one of the primary compounds responsible for the polymerase inhibition seen in the libraries, but neither molecule inhibited polymerization whatsoever. On the off chance of an allosteric inhibitory effect, all compounds in Figure 57 were assayed together, but the result was the same: total polymerization of the DNA template.

Some step in the process had clearly gone awry; it remained to determine what had gone wrong and what in fact the nature of the active compound was. Regardless of the failure of the compounds in Figure 57 to inhibit, the fact remained that combining xanthene tetra acid chloride **50** with one equiv. AMB and three equiv. L-leucine produced a highly inhibitory substance. The only possible conclusion was that something was being produced in the latter reaction as a side product to the chemistry expected.

The most obvious possibility was that the active compound was incompletely deprotected; perhaps one of the leucine *t*-butyl-ester groups had failed to come off during the trifluoroacetic acid deprotection step of the leucine carboxyls. Alternatively, some rearrangement might have taken place in the presence of TFA. A control library **XBn** was thus synthesized combining xanthene tetra acid chloride **50** with one equiv. AMB and three equiv. L-leucine-benzylester. After the usual washing step with 1 M citric acid and water, the library was deprotected in EtOH/EtOAC/TEA (60:39:1) with 10% palladium on carbon under H_2 atmosphere. This utterly orthogonal method of leucine deprotection resulted in a library identical to library **XL** in polymerase inhibition assays. Thus, excluding the citric

acid/H_2O washing step summarily, the formation of the inhibitory side product was pinpointed at the initial library synthesis: addition of amines to an acid chloride core in CH_2Cl_2 with triethylamine base.

To exclude any impurities (*e.g.*, diethylamine) which might have been introduced by the triethylamine, a library **XP** was synthesized through the identical procedure used to create active library **XL** except for the replacement of triethylamine with pyridine to soak up HCl in the initial amide forming reaction. The results showed that choice of base had no effect on the creation of an inhibitory library; **XP** behaved exactly as **XL** in the polymerase assay.

The last impurity considered was water; while the libraries were prepared with distilled dichloromethane under Ar, no extraordinary caution had been taken to keep the system dry. Furthermore, the 2-(aminomethyl)-benzimidazole reagent was used as it came from Aldrich Chemical: AMB·2HCl·H_2O. While it seemed unlikely that (less nucleophilic) water could compete against primary amines for reaction with an acid chloride, the possibility remained that the active xanthene compound was substituted with two leucines, one AMB and one unreacted acid.

To test this theory, a library **XW** was synthesized by combining xanthene tetra acid chloride core **50** with one equiv. AMB and only 2 equiv. L-Leucine. After 30 min. reaction time in CH_2Cl_2, the solution was quenched by the addition of water with vigorous stirring. After washing and deprotection, library **XW** was found to be *more* potent as a polymerase inhibitor than **XL**: a concentration course (Figure 58) showed that whereas **XL** inhibition cut off between 0.3 and 0.1 μl of library solution (3.7 mg in 500 μl), **XW** inhibition cut off between 0.1 and 0.03 μl. This result seemed to hint that the inhibitor might indeed contain an unsubstituted acid on the xanthene core.
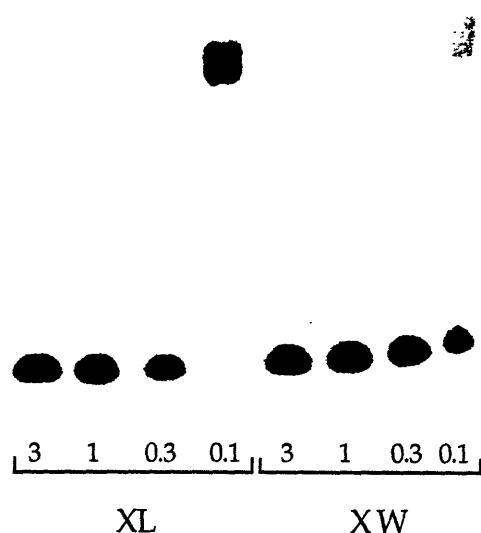
Figure 58. Concentration curve of xanthene libraries **XL** and **XW** screened against the Klenow Fragment of DNA polymerase I. μl amounts added are aliquots of library solutions 3.7 mg in 500 μl. Library **XL** was created with **50**, 1 equiv. AMB, and 3 equiv. of L-Leu. Library **XW** was created with **50**, 1 equiv. AMB, and 2 equiv. of L-Leu, followed by quenching with water.

There exist six different isomers of the xanthene core **50** substituted with one AMB, two leucines, and one free acid. A new set of these six compounds was thus synthesized using the dibenzyl core **61** and a new xanthene core **81** (Figure 59), developed by graduate student Jerry Shipps, which is further protected at its 5-position with a trichloroethanol group, removable with Zn/acetic acid.[88] The six compounds, two of which were synthesized pure and four of which were created as two pairs of isomers, are shown in Figure 60.
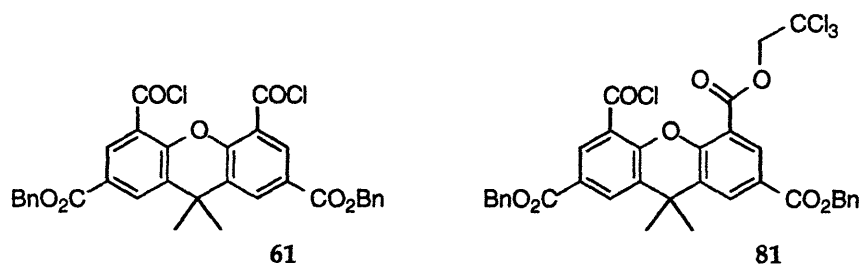
Figure 59. Partially protected xanthene cores for isomer synthesis.
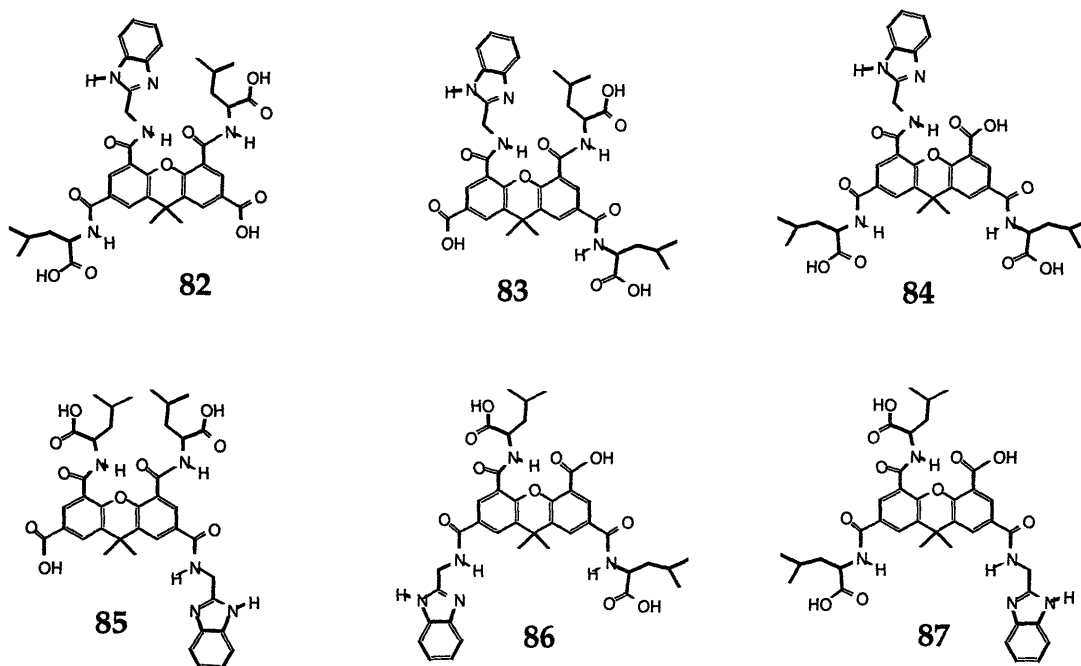
Figure 60. Xanthene based Leu/AMB derivatives. Compounds shown are with L-leucine.

The synthesis of the above six compounds is depicted in Scheme 9. As mentioned, those syntheses which involved the trichloroethanol (TCE) group were developed and performed by Jerry. To begin, the new core 81 was condensed (PyBOP coupling reagent) with either Leu or AMB; subsequent hydrogenation of the benzyl esters at xanthene positions 2 and 7 yielded 88 and 89. 88 was coupled with a 1:1 mixture of AMB and Leu, and after chromatography a mixture of two isomers was isolated. Removal of the TCE and t-Butyl protecting groups (Zn, HOAc followed by TFA) left the pair of isomers 86 and 87. Meanwhile, 89 was coupled with 2 equiv. Leu; deprotection of this product gave isomer 84.

The remaining three of six possible isomers were synthesized starting from the disubstituted xanthenes 78 and 79 already described above. (Scheme 8) 78 was condensed with AMB and 1 equiv. PyBOP, yielding isomer 85 upon t-Butyl deprotection (TFA). Similarly, 79 was condensed with Leu and 1 equiv. PyBOP; chromatography and deprotection produced a mixture of the final two isomers 82

and **83**. It should be noted that to avoid side products in a reaction in which one of two acids is being acylated, it is favorable to have PyBop as the limiting reagent and a slight excess of the desired amine.

The new set of six isomers (Figure 60) was screened for inhibition of DNA polymerization, but alas, none of the compounds showed any inhibitory activity in the assay. How, then, could we explain the enhanced activity of library **XW**, in which two and one equiv. Leu and AMB were added to one equiv. of the xanthene core? The only remaining possibility lay with the 2-(aminomethyl)-benzimidazole building block itself.
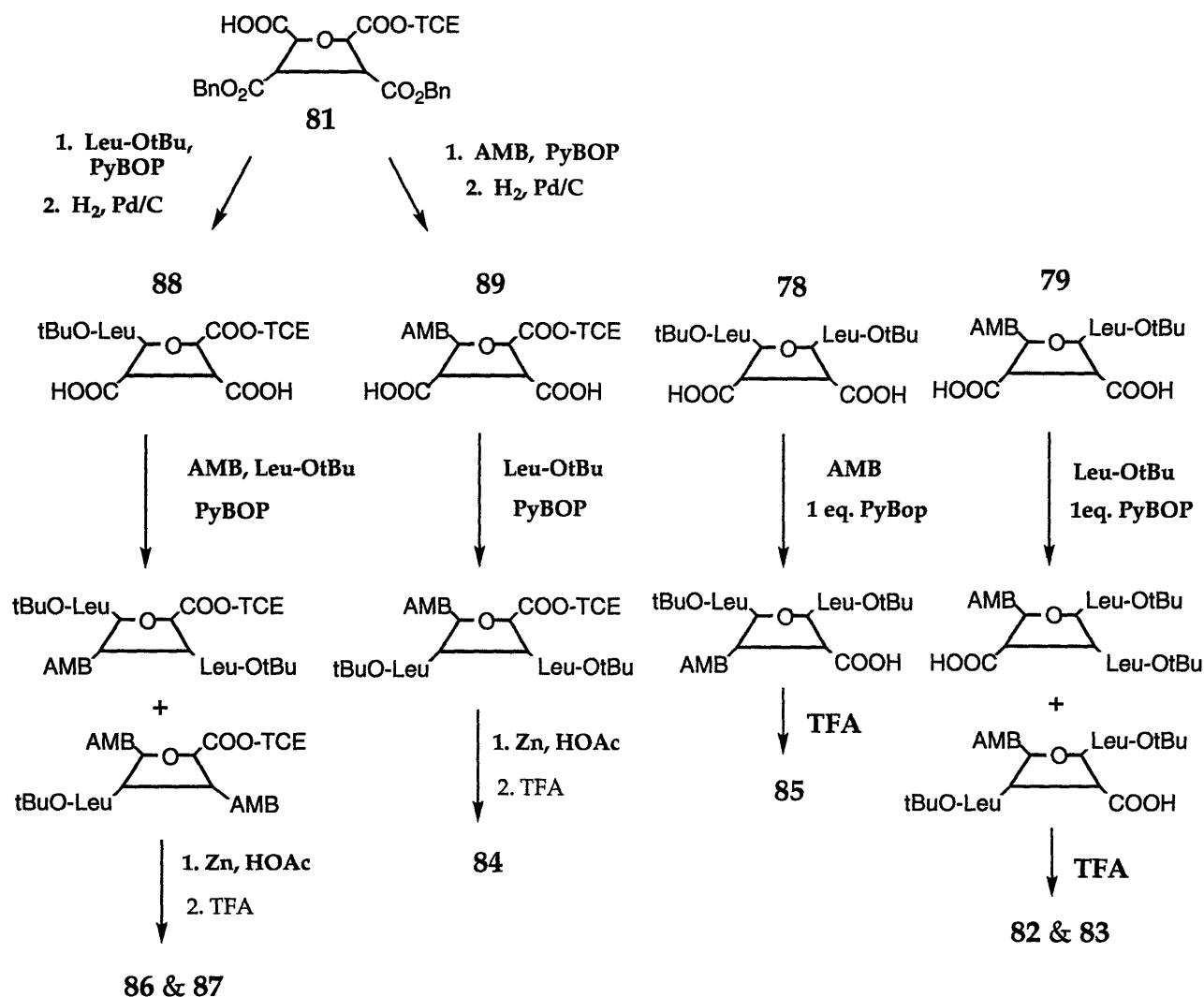
Scheme 9. Synthesis of Xanthene AMB/Leu/COOH derivatives.

It was postulated that after the primary amine of AMB reacted with a first xanthene acid chloride site, the imidazole nitrogen of the AMB could react with a second molecule of xanthene acid chloride, thus forming a macromolecular dimer of two xanthene cores. Naturally, with the xanthene tetra acid chloride core **50**, polymeric species could also arise from such a phenomenon. If such a dimer or polymer were to be the compound responsible for inhibitory activity in the assay, the case of library **XW** would be explained by the fact that during library formation, the ratio of primary amines to acid chloride sites was only 3:4, leaving a full equivalent of acid chloride sites for a second reaction with the AMB imidazole nitrogen. In contrast, all previous libraries (*e.g.*, library **XL**) had a 1:1 ratio of primary amines to acid chloride sites during their formation. Thus, it was possible that in the attempt to make a library **XW** containing xanthene molecules with unsubstituted xanthene acid functionalities (such as the molecules in Figure 60), we had in fact created xanthene dimers and polymers. The subsequent addition of water to library **XW** may have played no role whatsoever in conferring inhibitory activity on the library compounds.

To examine what compounds were in fact being produced in library **XW**, the library was re-synthesized and separated into twelve fractions by flash chromatography prior to leucine *t*-butyl deprotection (normal phase silica gel, 5-50% MeOH/CHCl₃). Each of the twelve fractions collected was then deprotected as usual with TFA and precipitated with ether/hexanes. The polymerase inhibition of the twelve fractions is shown in Figure 61, along with a picture of the TLC plate from chromatographic separation. Inhibitory material was found to be present in fractions **XW-2** through **XW-5**, with **XW-4** and **XW-5** showing extensive inhibition of DNA polymerization. Since the TLC shows good separation of compounds in each fraction, the inhibition seen is clearly the result of more that one active molecule.
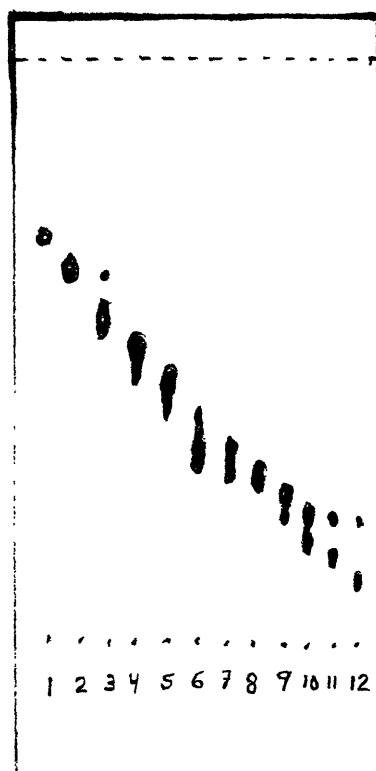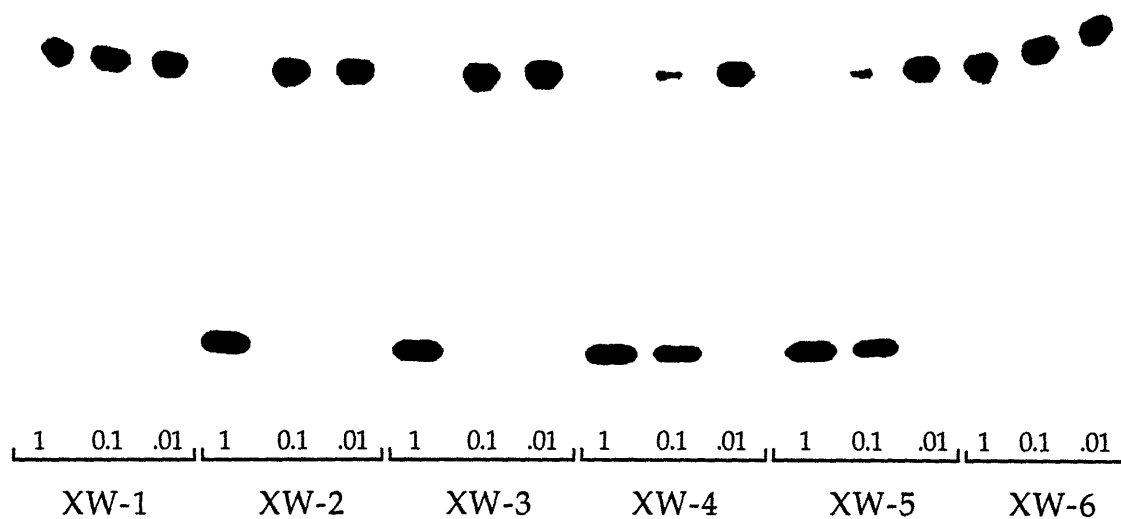
Figure 61. Mini-concentration curves of the first 6 of 12 fractions of xanthene library XW screened against the Klenow Fragment of DNA Polymerase I. 1, 0.1, and 0.01 μl amounts added are aliquots of library solutions 3.7 mg in 500 μl. A drawing of the TLC products of the separation is shown for comparison (10 % MeOH/CHCl₃). Fractions 7 through 12 showed no inhibition.
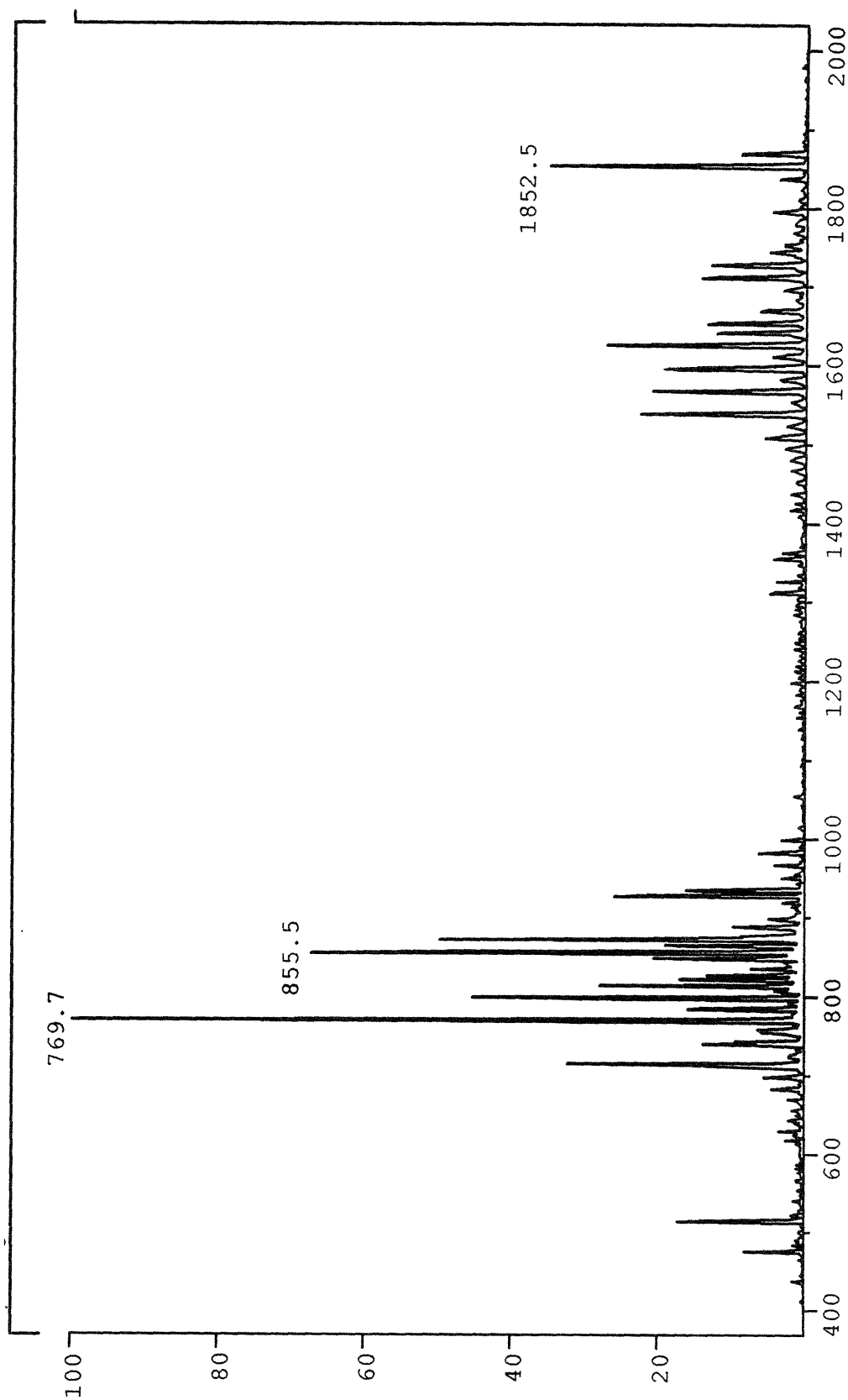
Figure 62. Electrospray ion chromatogram of fraction 4 of xanthene library **XW**.
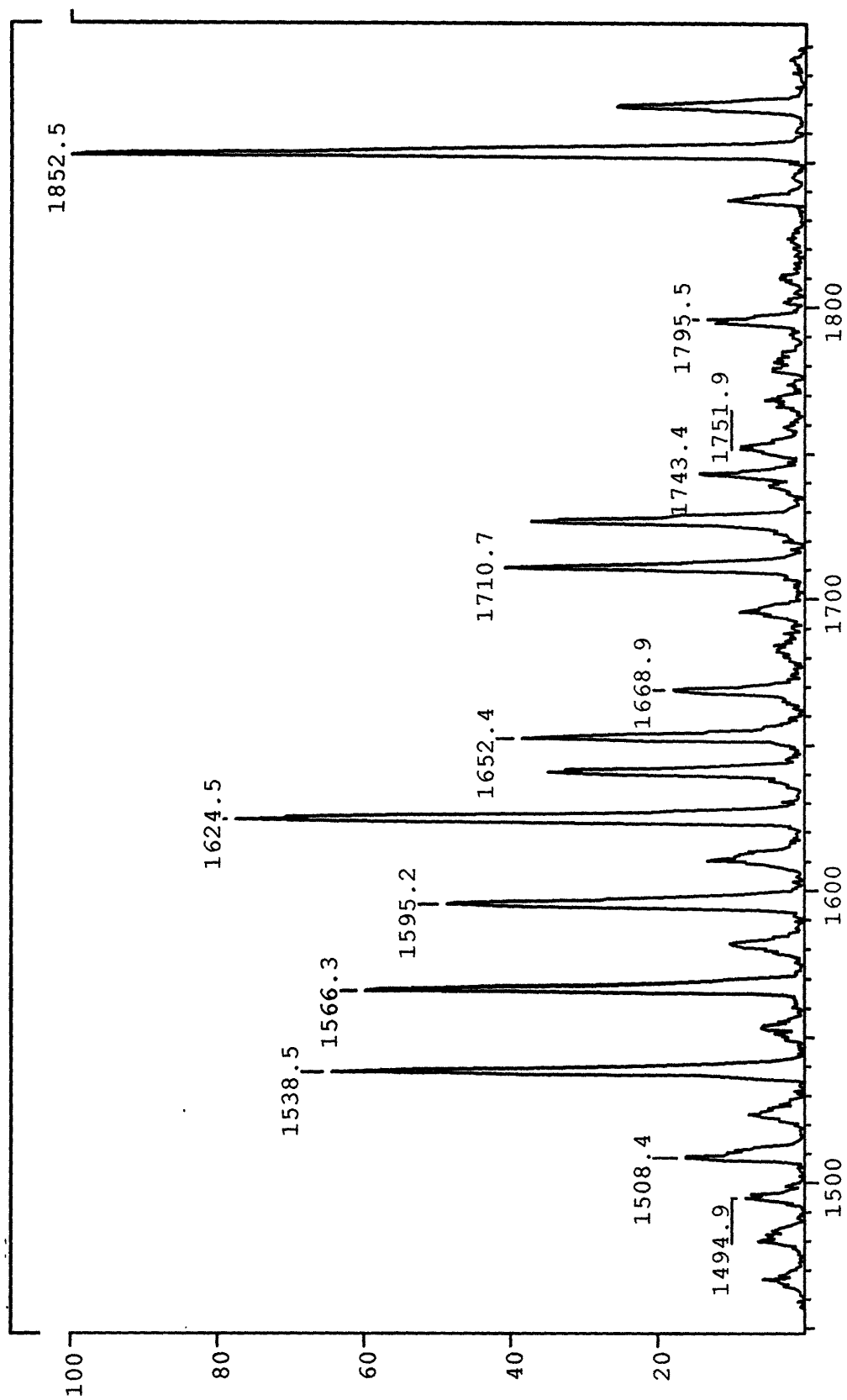
Figure 63. Electrospray ion chromatogram of fraction 4 of xanthene library XW, expanded.

The two highly inhibitory fractions **XW-4** and **XW-5** were examined by Electrospray Mass Spectrometry, again in collaboration with Yuriy Dunayevskiy. The ion chromatogram for **XW-4** is shown in Figures 62 and 63. In addition to the expected compounds in these fractions (molecular weight 855, for example, is the M+1 peak for isomer **74** , Figure 57), a host of compounds were seen in the region of xanthene dimers and trimers. Ion peaks 1595 and 1852, for example, could correspond to the M+1 peaks of dimer **90** and trimer **91** shown in Figure 64 (or any isomers thereof). Thus, the hypothesis that AMB might form dimers from xanthene acid chlorides was shown to have merit.
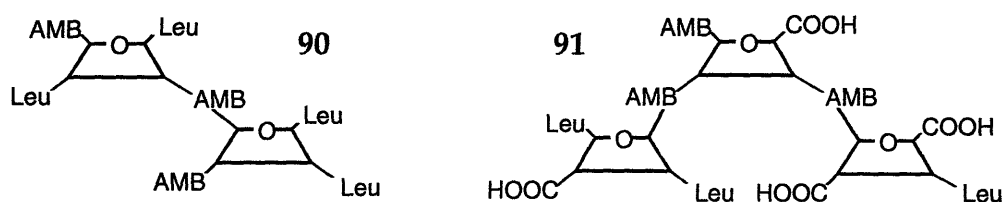
Figure 64. Schematic representation of xanthene dimer **90** and trimer **91**.

To this point, the above possibility of AMB polymerization had been improperly dismissed on the basis of the steric barrier to forming a bond between the AMB imidazole and a bulky xanthene acid chloride. Furthermore, initial tests run with AMB and one equiv. benzoyl chloride had shown excellent yield of product **92** (Figure 65). Re-examination of this reaction, however, showed a minor (<5%) side product of the di-acylated product **93**, and reaction of AMB with two equivalents of benzoyl chloride gave almost exclusively **93**. Clearly, then, AMB is capable of forming dimeric species under conditions of acylation. Computer modelling[52] of **93** (Figure 66) showed that while the aromatic ring of the second benzoyl group was forced out of planarity with the new amide bond, the amide itself

was able to retain a planar orientation, thus allowing a stable amide bond. Product **93**, for example, was chromatographed without degradation on silica, 5% MeOH/CHCl₃.
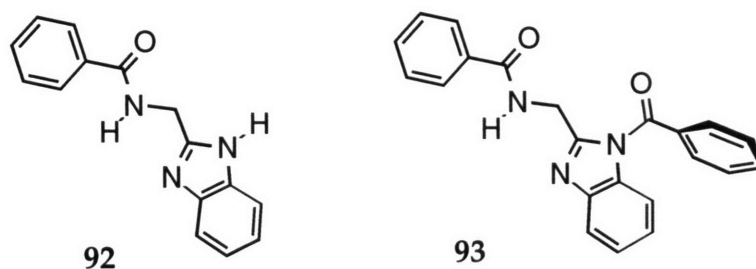
Figure 65. Mono- and di-acylation products of 2-(aminomethyl)-benzimidazole.
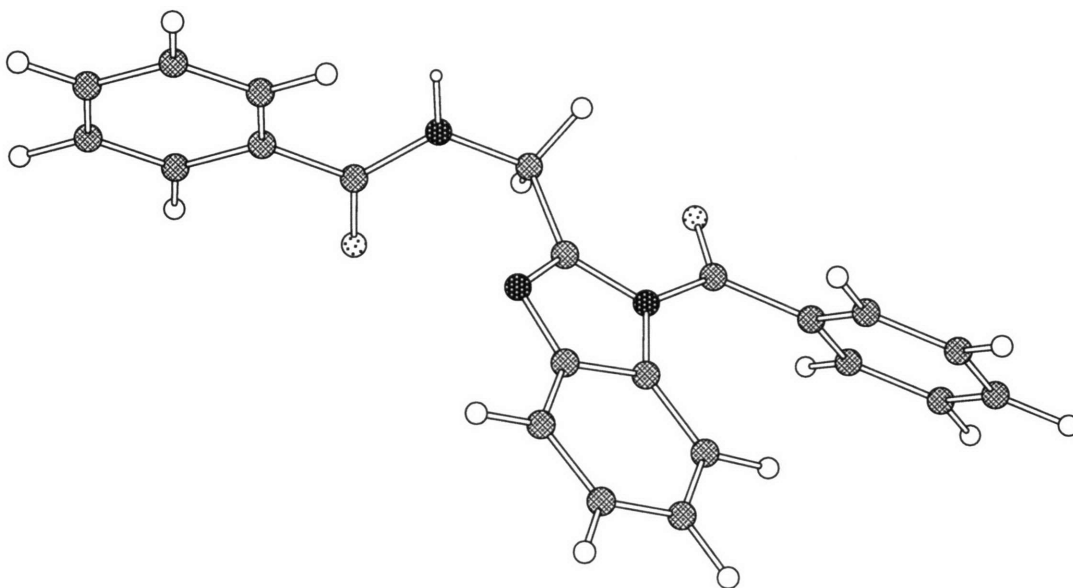
Figure 66. Computer model[52] of the di-acylation product of 2-(aminomethyl)-benzimidazole.

No hard synthetic evidence had yet been found, however, for the hypothesis that AMB dimers or trimers were the source of inhibition observed in the DNA polymerase assay. Therefore, **79** (see Scheme 8 above) was condensed with itself to form Library **XD** (Scheme 10). Compound **79** was simply stirred 30 min. in DMF

with 2 equiv. PyBop and diisopropylethylamine as a base. The reaction mixture was taken up in chloroform, washed with water to remove DMF, and the resulting material deprotected with TFA to remove leucine *t*-butyls. Library **XD** was precipitated with 1:1 ether/hexane as usual, and the resulting white powder assayed for polymerase inhibition.
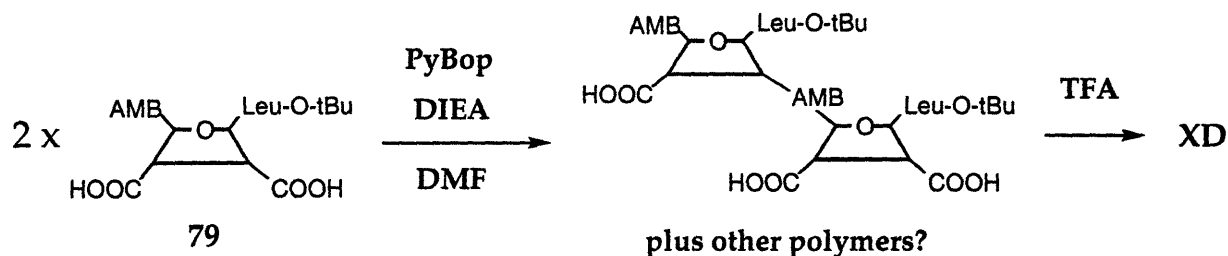
Scheme 10. Synthesis of library **XD**.

XD
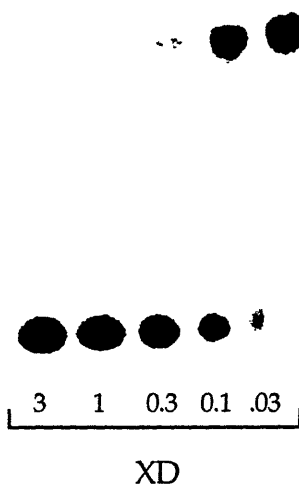
Figure 67. Concentration curve of xanthene library **XD** screened against the Klenow Fragment of DNA polymerase I. µl amounts added are aliquots of library solutions 3.7 mg in 500 µl. Library **XD** was created by allowing **79** to condense with itself in the presence of 2 equiv. PyBop.

As shown in Figure 67, tests of library **XD** confirmed our suspicions of dimeric or higher order active species. A concentration curve of **XD** in the assay revealed inhibition of elongation to the DNA 45-mer down to 0.1 µl of the standard library solution (3.7 mg in 500 µl). Thus, with only a single starting material and a single reagent, active compounds were created.

An electrospray mass spectrum of **XD** (Figure 68) taken by collaborator Yuriy Dunayevskiy showed that PyBop was a poor coupling reagent for acids and the AMB imide; most of the product was unreacted starting material (*t*-butyl deprotected **79**, M+1 = 629). The mono and bis diisopropylethyl-amine salts of deprotected **79** are also visible at 629 + 129 = 758 and 629 + 2*129 = 887, and the noncovalent dimer of deprotected **79** appears at M+1 = 1257. (This is normal for electrospray ionization; MS of pure *t*-butyl deprotected **79**, given in the experimental, also shows the noncovalent dimer at 1257).

However, a clear signal of the *covalent dimer* of deprotected **79** (dimer minus H₂O) can be seen at M+1 = 1239 (2*628 - 18 + 1). Perhaps because PyBop was a poor coupling reagent for AMB, no higher order polymers could be detected above the threshold of noise to m/z=3000. To complete the analysis of the spectrum, also visible are mono and bis diisopropylethyl-amine salts of the noncovalent dimer of deprotected **79** (M+1 = 1386, 1515). Finally, *t*-butyl deprotected **79**, its covalent dimer, and its non-covalent dimer all show an M+1+70 peak (699, 1309, 1327); the nature of the +70 adduct is not known, but its addition to all other peaks suggests a carboxylate salt much as with diisopropylethylamine.

Figure 68. Electrospray ion chromatogram of the xanthene library **XD**. Library **XD** was created by allowing **79** to condense with itself in the presence of 2 equiv. PyBop.
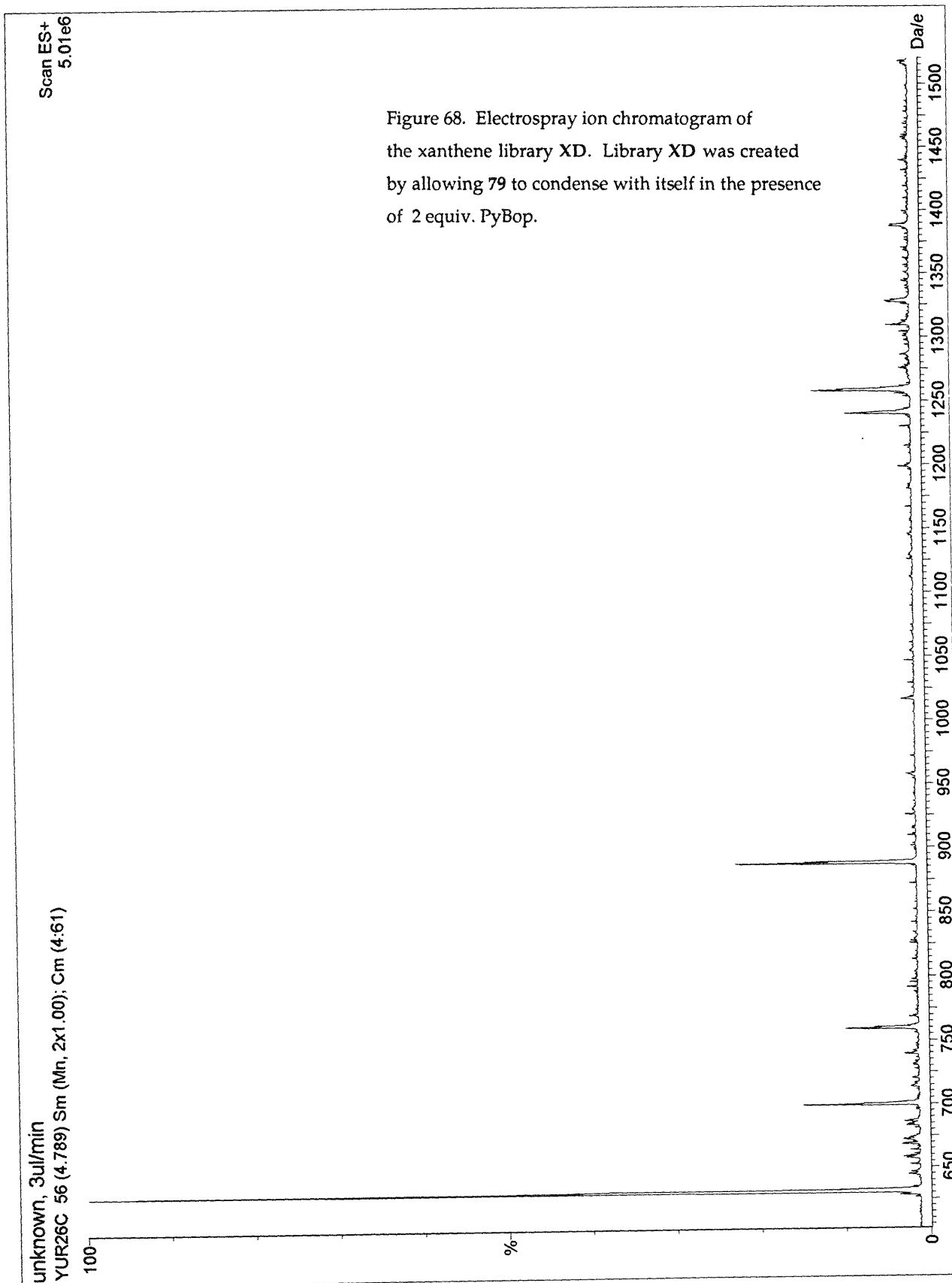
Given that control assays of the starting material **79** and its *t*-butyl deprotected product showed no polymerase inhibition, and given that the MS of library **X D** indicates only two covalently linked molecules -- deprotected compound **79** and the covalent dimer thereof -- it seems likely that one or both of the isomers **94** or **95** pictured in Figure 69 is the cause of polymerase inhibition in library **XD**. These particular molecules, however, cannot by themselves explain the inhibition of xanthene library **XW**, since its molecular weight is not seen in the mass spectrum of **XW** fraction 4 (Figures 62 and 63). Instead, the inhibitory activity of **94/95** in conjunction with the MS results of **XW-4** indicates a whole family of polymerase inhibitors based on the polymerization of xanthene by AMB.
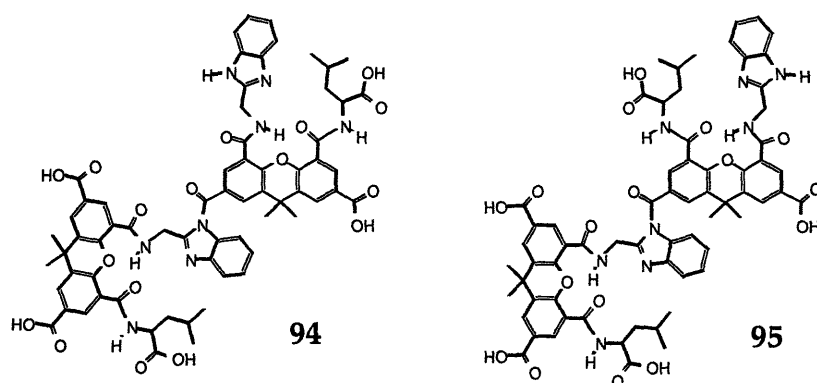
Figure 69. Molecules **94** and **95**, postulated as the polymerase inhibitors of library **XD**. Compounds shown are with L-leucine.

The aforementioned family of inhibitors is thus the end result of the deconvolution of the initial xanthene library 2 (**X2**). What can be said of these new molecules? To begin, there is clearly very specific molecular recognition between these compounds and the Klenow Fragment of DNA polymerase I (See also analysis of binding modes, Section **II.iii.5**.) As observed from Round IV of screening, the leucine side chain must make specific van der Waals contacts, because Leu>Ile>Val for the purposes of conferring inhibitory activity on the xanthene libraries. Thus, inhibition is not merely the result of a long chain of carboxylate groups; this was

confirmed by synthesizing the dimer **97** as shown in Scheme 11. Compound **97** was inactive as a polymerase inhibitor, leading to the conclusion that only a particular arrangement of Leu/COOH groups about a xanthene dimer suffice for inhibition of polymerase activity.
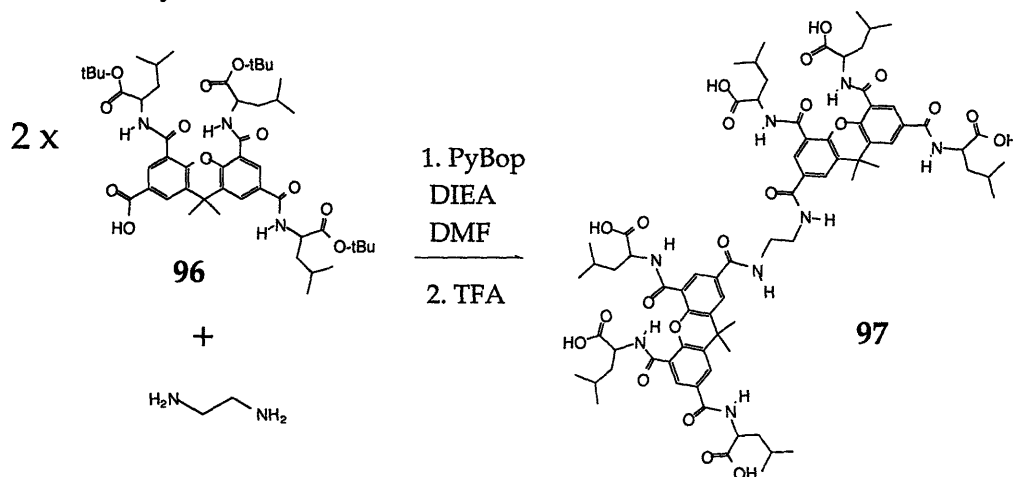
Scheme 11. Synthesis of the dimer **97**.
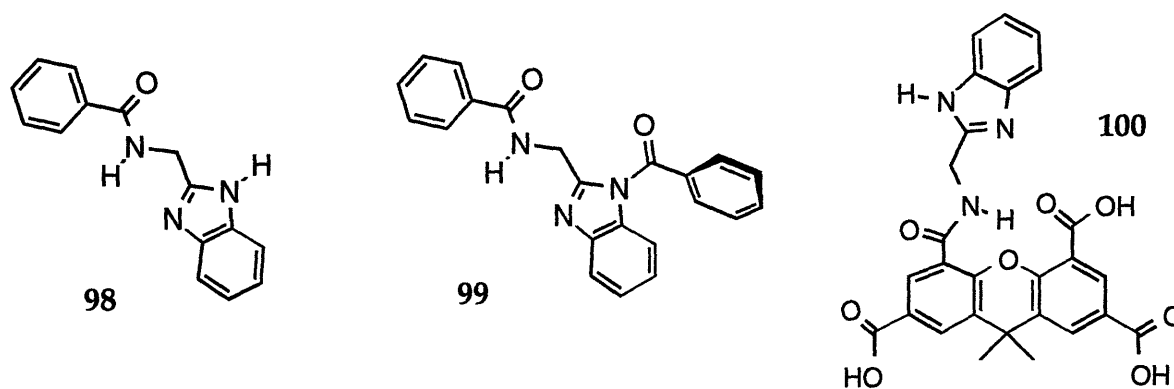
Figure 70. 2-(aminomethyl)-benzimidazole control molecules.

Control runs showed that neither of the AMB molecules **98, 99,** or **100** (Figure 70) showed any inhibition of DNA polymerase I, so it seems likely that the critical role of AMB is to connect Leu/COOH substituted xanthenes in a particular orientation such as **94** or **95**. Further evidence for this came from the fortuitous testing of the synthetic intermediates shown in Figure 71 which had been used in the synthesis of various xanthene isomers above (see Schemes 8 and 9).
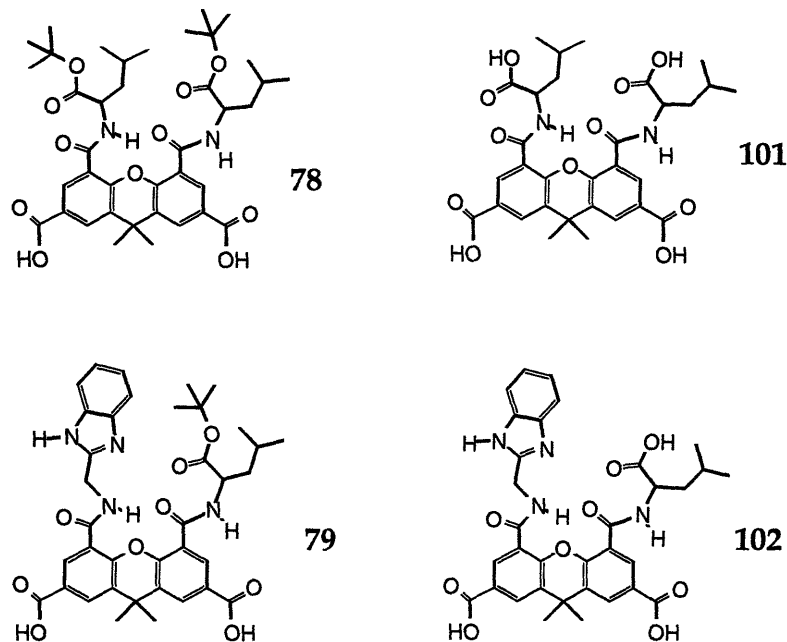
Figure 71. Disubstituted xanthene derivatives.

While assays showed molecules **79, 101**, and **102** to be inactive as inhibitors of DNA polymerase I, the di-(leucine-O-*t*-butyl) xanthene derivative **78** showed inhibition at 1 μl of the standard test solution (3.7 mg in 500 μl). Not to mention that the *t*-butyl esters of **78** would have been absent in any of the TFA treated xanthene libraries, pure compound **78** is less active than the mixture of compounds in library **XW** (inhibitory at 0.1 μl of test solution), so it clearly cannot account for the activity seen in the deconvolution process. However, the fortuitous discovery of inhibition by **78** allows speculation on the molecular recognition of inhibitory xanthene dimers such as **94** and **95** (Figure 69).

Comparing the molecules **94**, and **95** with **78**, they share a region of recognition as depicted in Figure 72. The leucine and one or both xanthene acids perhaps fill a binding site on the polymerase molecule. However, this interaction is clearly not sufficient for tight binding, as measured by the ineffectiveness of controls **79, 101**, and **102** (Figure 71) in inhibition assays. The evidence suggests a second site of binding leading off from positions R₁ and R₂ as marked in Figure 72. In the case

of **78**, this site is filled by R_1 = Leu-O-*t*-butyl and R_2 = *t*-butyl. In the case of **94/95**, R_2 = H, but R_1 is an entire additional xanthene core. Based on inhibition data of **78** and **XD** (**94** and **95**), the added xanthene core seems to fill this second site of binding more effectively than a pair of t-butyl esters, and beyond this the second xanthene of **94/95** supports an AMB Leucine pair which may further enhance binding.
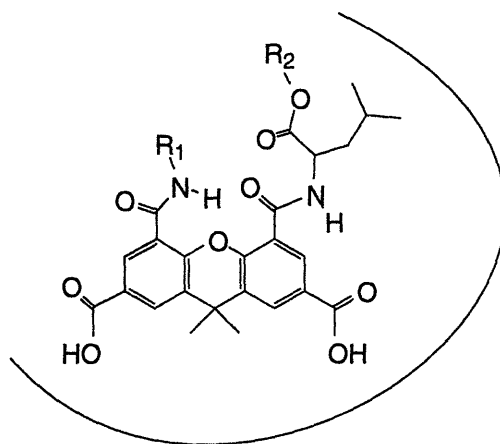
Figure 72. Proposed site of common recognition between molecules **78** and **94/95**.

In a process of combinatorial drug discovery (as opposed to rational design based on protein data), the researcher has no particular evidence of the mechanism of action of any active molecules which result from deconvolution. While speculation such as that of Figure 72 may be proposed, it was desired to examine -- in a more quantitative fashion -- the interaction of the above inhibitory compounds with the DNA polymerase I protein.

III.iii.5 Initial analysis of the Mode of Binding of New DNA Polymerase I Inhibitors

Based on X-ray crystalographic data, the Klenow Fragment of DNA polymerase I resembles a "right hand cupped as if to hold a rod."[82] As depicted scbematically in Figure 73, its polymerase activity resides within the deep cleft (the "palm" in this analogy) between two long alpha helices (the "thumb") and a larger protrusion of five alpha helices and one beta sheet (the "fingers"). A loosely ordered strand of 50 residues is attached to the end of the thumb, which may function to reversibly close the "grip" of the protein on DNA (see below).[82] At the "wrist" there exists a second functional domain thought to be responsible for the 3'-5' exonuclease activity of the Klenow Fragment, which likely allows the protein to "edit out" mismatched DNA basepairs.[82]
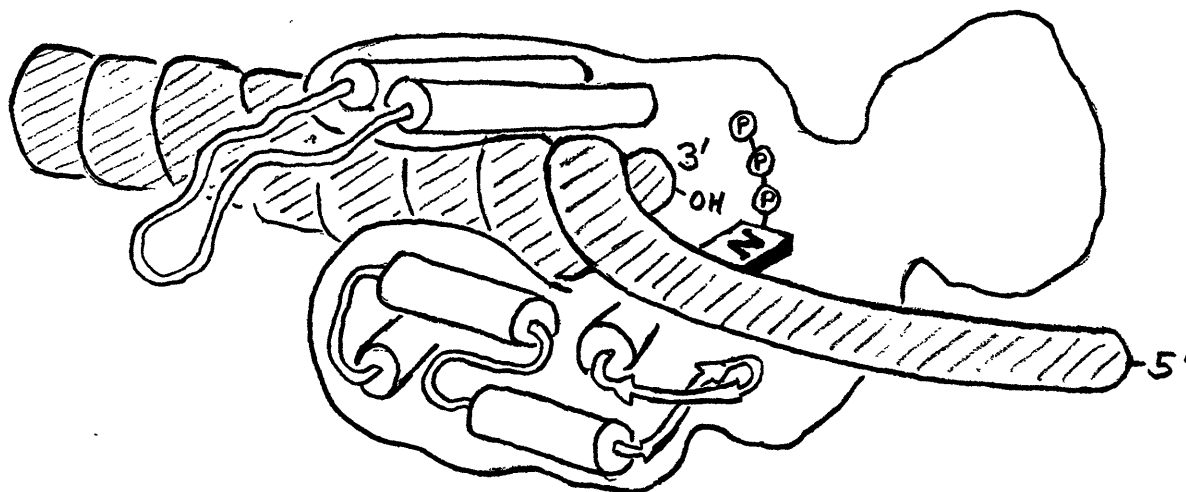
Figure 73. Hypothetical depiction of the Klenow Fragment of DNA polymerase I bound to a DNA template and a dNTP.[80,82,84,87]

Current models of the mechanism of polymerization place the DNA template within the "cupped hand" of the structure proposed above, with the site of dNTP binding to the growing DNA strand at "the base of the pinky finger."[80,82,84,87] The proposed minimal mechanism of polymerization is described in Equation 10

below:[80] the enzyme first binds the DNA template followed by the proper dNTP; subsequent pyrolysis of the triphosphate bond extends the 3' site of the growing DNA chain and leaves pyrophosphate (PP$_i$). After release of PP$_i$, the polymerase may "process" to the next base of the DNA template.

Eqn. 10:

$$E + DNA \rightleftharpoons E{:}DNA \xrightarrow[]{+dNTP} E{:}DNA{:}dNTP \rightleftharpoons E{:}DNA\text{-}N{:}PP_i \xrightarrow[]{-PP_i} E{:}DNA\text{-}N \longrightarrow elongation$$
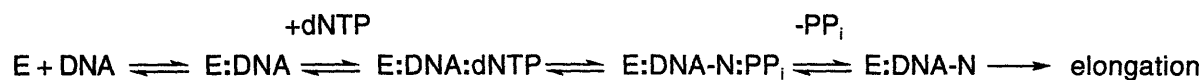
The molecule pyridoxal-5'-phosphate (P5P, 103) pictured in Figure 74 is able to inhibit the action of the Klenow Fragment of DNA polymerase I.[86,87] Studies in which the Klenow fragment was incubated with P5P in the presence of NaBH$_4$ covalently linked P5P to three lysines of the enzyme.[87] One of these lysines in particular is thought to be critical for dNTP binding during polymerization, and the covalently linked enzyme was found to be inactive. When the incubation experiment was repeated in the additional presence of dNTP, only two molecules of P5P were covalently linked to the enzyme, and the covalently linked enzyme was found to retain polymerization activity, although at a reduced rate. This indicates that the site of P5P inhibition is "in or around the dNTP binding site that is essential for polymerase activity."[87]
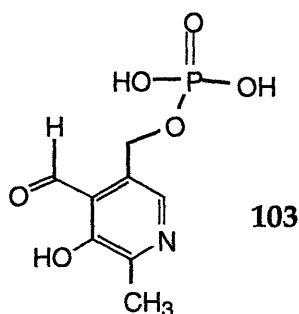
Figure 74. Pyridoxal-5-phosphate.

Studies performed by Deborah Kreutzer showed that addition of pyridoxal-5-phosphate to our polymerase assay showed inhibition of the elongation of DNA template to the DNA 45-mer, just as with our inhibitory libraries. Qualitative kinetic runs showed that increasing the concentration of dNTPs partially alleviated P5P inhibition, confirming that P5P interferes with the binding of dNTPs to the enzyme and not the binding of the DNA template. The mode of action of our library molecules seems to fit this pattern: initial studies with Fraction 5 of library XW indicated that its inhibitory activity was also decreased by increasing the concentration of dNTPs. Thus, it appears that our molecules also interfere with the dNTP binding step of Klenow Fragment polymerization.

Detailed kinetic studies of polymerization are beyond the scope of this Thesis (the simple equations of Michaelis Menton kinetics, used to determine the K_i of our trypsin inhibitors, do not apply for the more complex mechanism of Eqn. 10 above). However, a fair estimate of the K_is of our molecules may be obtained through comparison with the activity of P5P, known to have $K_i = 32\mu M$.[87] A plot of polymerase activity vs. inhibitor concentration is shown in Figure 75 for pyridoxal-5-phosphate, the pure compound 78, and fraction XW-5 of library XW. (Compound 78 and XW-5 are respectively the most potent substance and the most potent library isolated above). It should be noted that all inhibition could be reproduced after allowing the compounds to stand 24 h at RT in buffered solution, pH 7.8. Thus, our active compounds are relatively stable in an aqueous environment.
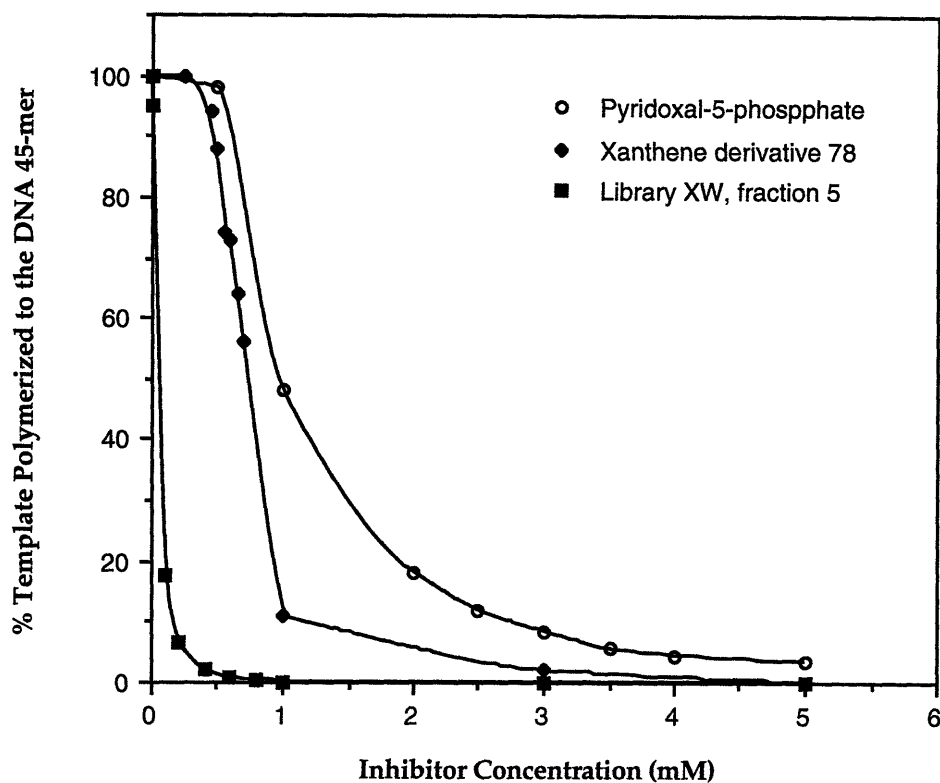
Figure 75. Plot of polymerase activity as a function of inhibitor concentration for P5P, 78, and XW-5.

The graph of inhibitor activity above clearly shows that both compound **78** and library **XW-5** are more potent inhibitors than pyridoxal-5-phosphate. Assume that the inhibitors bind at or near the site of dNTP incorporation and inhibit the reaction in the second step of Eqn. 10 (after the DNA template is bound). The rate V of enzyme catalysis can then be defined as:[1]

Eqn. 11: $V = k[ES]$

It follows that since in our assay (under conditions in which enzyme is saturated with substrate dNTPs), observed polymerization activity A (% polymerized template per a constant 5 minutes) is directly dependent on that limiting rate:

Eqn. 12: $A \propto k[ES]$

Again, under the assay conditions used, free enzyme is negligible, so:

Eqn. 13: $[E_{tot}] = [ES] + [EI]$ or $[ES] = [E_{tot}] - [EI]$

And substituting the definition of $K_i = [E]*[I] / [EI]$:

Eqn. 14: $[ES] = [E_{tot}] - ([E]*[I] / K_i)$

Since $A \propto k[ES]$, at conditions of equal observed polymerization activity A:

Eqn. 15: $k_1([E_{tot1}] - ([E_1]*[I_1] / K_{i1})) = k_2([E_{tot2}] - ([E_2]*[I_2] / K_{i2}))$

Excluding $[I_1]$ and $[I_2]$, all concentrations 1 and 2 are constant for the plots in Figure 74, and for the same enzyme, $k_1 = k_2$. Thus, *at conditions of equal observed polymerization activity*, Eqn. 15 reduces to:

Eqn. 16: $[I_1] / K_{i1} = [I_2] / K_{i2}$ or $[I_1] / [I_2] = K_{i1} / K_{i2}$

Extracting from Figure 74 the fact that equal polymerization activity is observed (50% template polymerization) at concentrations of 1.0 mM P5P, 0.75 mM **78**, and 0.05 mM **XW-5**, and given the known K_i of pyridoxal-5-phosphate is 32 μM, the K_i values for **78** and **XW-5** may be estimated at 25 μM and 1.5 μM respectively. Since **XW-5** is still a mixture of compounds, the actual K_i of the most active molecule in the library is estimated in the nanomolar region. The outlook for isolation of such a compound is promising, as described in the following section.

II.iii.6 Outlook for the Synthesis of Further DNA Polymerase I Inhibitors

At the time of this writing, everything in the Rebek laboratory is being marked with one of two labels: "MIT" or "Scripps." A short month from now, a convoy will take the entire lab to La Jolla, there to continue work under the auspices of the new Skaggs Institute for Chemical Biology. While a source of excitement to those escaping to the warmer clime of San Diego, the impending disappearance of the MIT laboratory has presented a formidable deadline for the completion of research. Thus, as always, there remains work to be done.

Most importantly, the only inhibitor above of DNA polymerase I which was isolable as a single isomer and unequivocably characterized was the di-(leucine-O-t-butyl) xanthene derivative **78**. Absolute synthetic confirmation of a xanthene dimer or trimer linked by AMB molecules is greatly to be desired as a final chapter to close out this project, and my colleague and collaborator Jerry Shipps has expressed interest in pursuing the work. In the Essigmann Group, collaborator Deborah Kreutzer has agreed to continue her excellent work on the biological side of the project, running polymerase assays as needed.

The condensation of dimers **94** and **95** from monomer **79** is impractical, as a) the coupling produces two isomers which will be difficult to separate, b) with PyBop as a coupling reagent, most of the product was unreacted starting material, and c) a more potent coupling reagent would run the risk of extreme polymerization, given that two acids exist for every AMB. What is needed is a synthesis of **94, 95**, and related polymers which will be efficient and regiospecific. One such synthesis is diagrammed in Scheme 12. By combination of one of the pentafluorophenyl (PFP) esters in Column I with one or two of the AMB derivatives in Column II, all of the possible dimers and trimers shown at the top of Scheme 12 may be created as their pure isomers.

R$_1$, R$_2$, R$_3$ = Leu or COOH   Or   R$_2$, R$_3$, R$_4$, R$_5$ = Leu or COOH

104

105

106

107

108

+

109

110

111
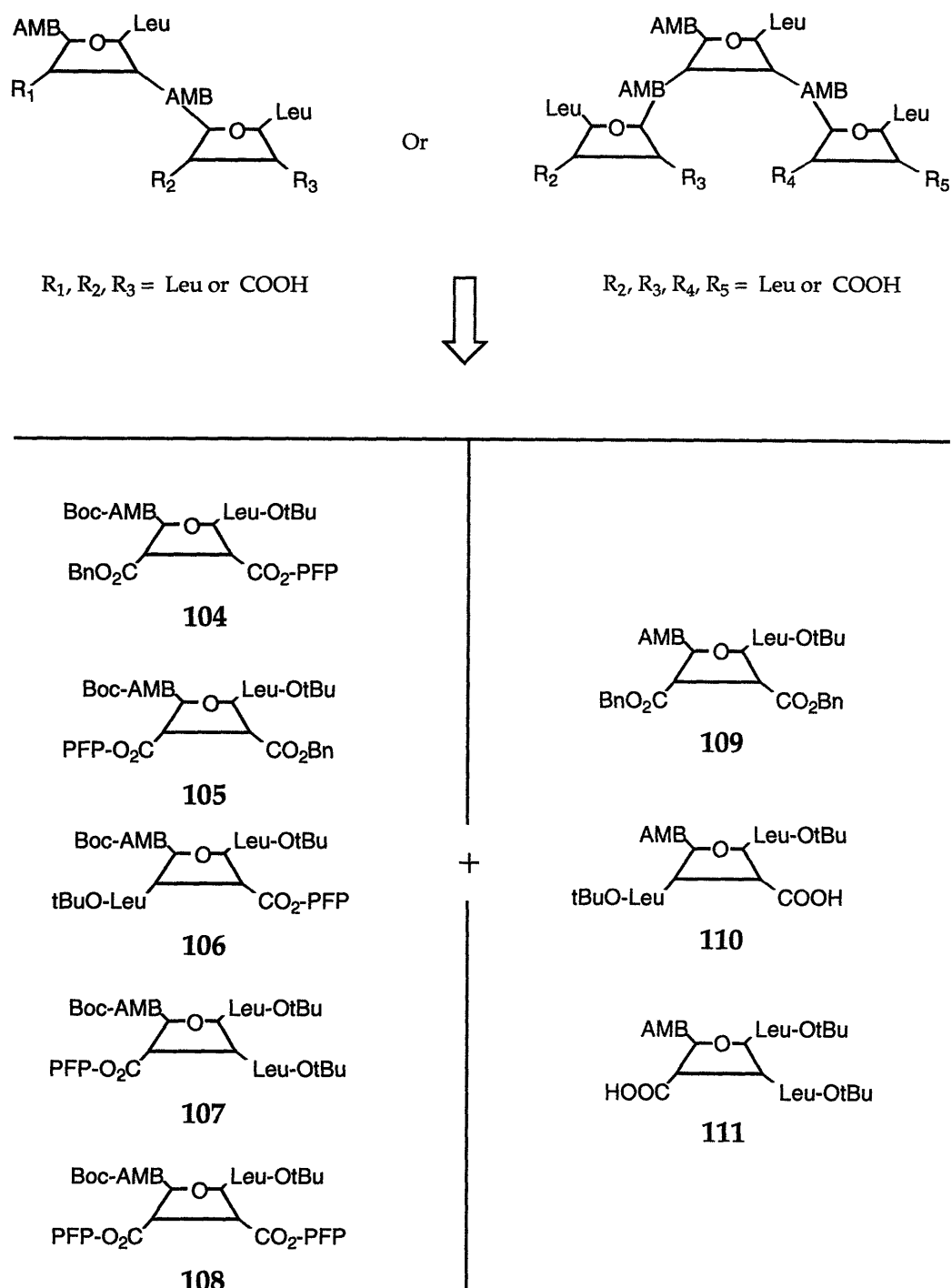
Scheme 12. Proposed regioselective formation of Xanthene polymers **94, 95,** and related molecules suggested as targets for future synthesis and assay against the Klenow Fragment of DNA polymerase I activity.

Syntheses of the molecules in Column II have been described in the text above. (Scheme 9) Syntheses of the molecules in Column I demand only three new steps: preparation of Boc-AMB/Leu-O-tBu xanthene derivatives, and introduction of the benzyl and/or PFP ester(s). As shown in Scheme 13, synthesis of Boc-protected AMB molecule **113** should be facilitated through initial formation of **109** as described above (Scheme 8), subsequent Boc protection of the AMB imide with di-*t*-butyl-dicarbonate to give **112**, and benzylester removal with H_2, Pd/C. The formation of PFP and/or benzyl esters requires condensation of xanthene acid and pentafluorophenol or benzyl alcohol in THF with the ester forming reagent 1-ethyl-3-(3,3-dimethylaminopropyl)-carbodiimide (EDC) plus triethylamine and catalytic 4-dimethylamino-pyridine (DMAP). As necessary for the given molecule, one or two equivalents of EDC may be used. If addition of a second Leu-O-t-butyl moiety is desired, this may be added prior to PFP ester formation with PyBop as described (Scheme 9). As noted above, to avoid side products in a reaction in which one of two acids is being acylated or esterified, it is favorable to have PyBop or EDC as the limiting reagent and a slight excess of the amine or alcohol in question.

Given the proven activity of compounds **78** and **94/95**, the pure compounds of Schemes 12 and 13 have an excellent chance of showing extensive inhibition against the Klenow Fragment of DNA polymerase I. With the continued effort of collaborator Jerry Shipps, it is hoped that the above methodology will expedite their synthesis.
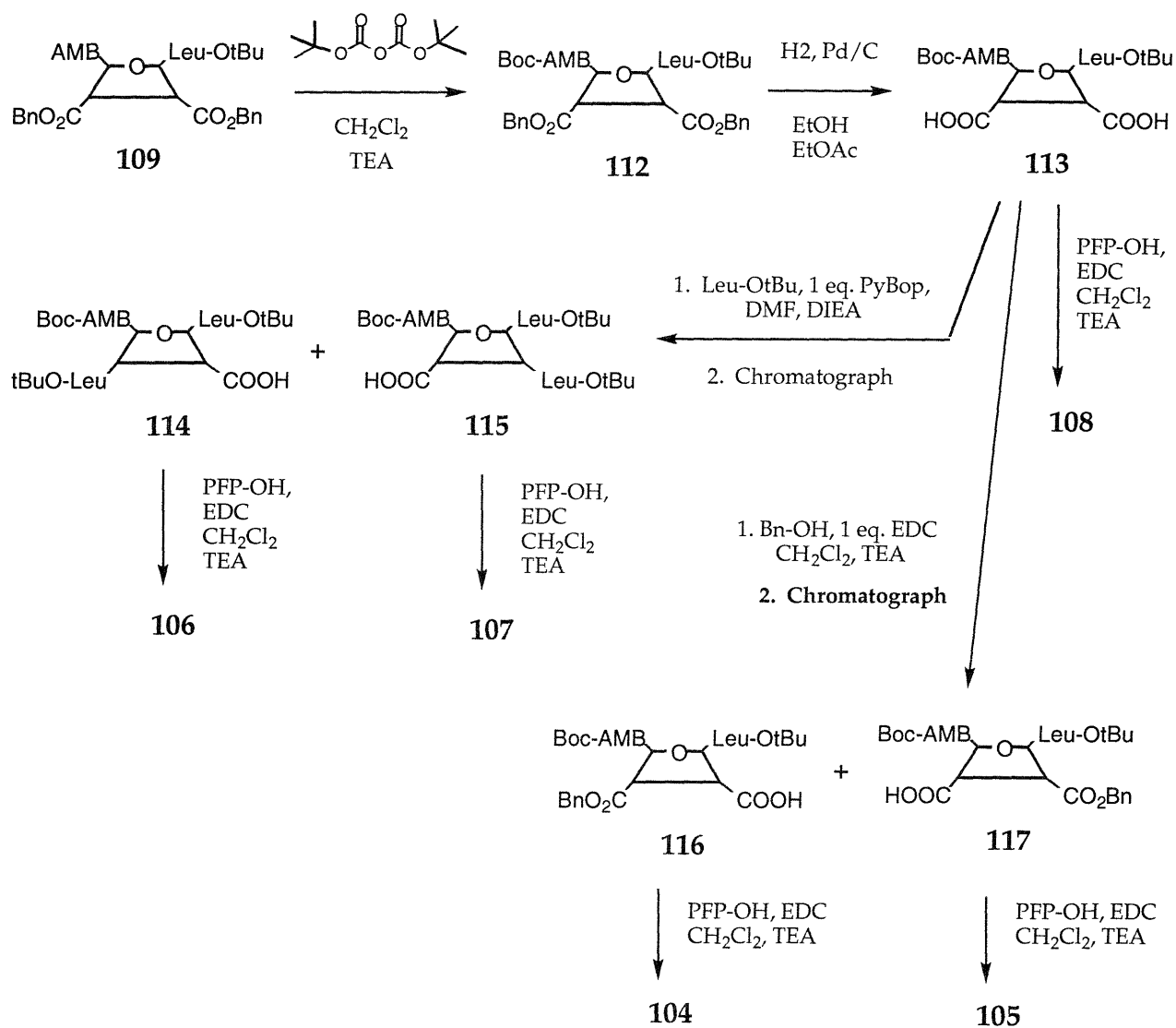
Scheme 13. Proposed formation of xanthene derivatives used in Scheme 12.

II.iii.7 Conclusion

What can be said of the combinatorial process which led to the family of polymerase inhibitors described above? Clearly, our second attempt at deconvoluting a library which showed activity in a biological assay was not as smooth as the first. The final inhibitors discovered were not, in fact, molecules which "should" have been in the library; they were not tetra-substituted xanthenes. Although it was presumed that a library of tetra substituted xanthene molecules was being screened, in truth we were deconvoluting a much larger library, containing at least dimeric and trimeric xanthene species. From this fact, one could dismiss our poly-functionalized core and building block approach to combinatorial chemistry as unreliable. Certainly, such a high level of generation of "side products" is unacceptable in the usual sense of organic synthesis.

However, taking a slightly larger view of the proceedings, one might conclude that in fact the process of combinatorial synthesis and deconvolution was highly successful: it determined a potent inhibitor of the DNA polymerase I Klenow Fragment *despite* the poor judgement of the human operator involved. In hindsight, it was clearly a mistake to add AMB to the pool of building blocks without a protecting group on its imidazole nitrogen, yet despite this human error, the selection process led us toward the active compound(s) 94/95. The method of library deconvolution through omission of building blocks and iterative screening was in fact capable of handling a much higher number of compounds than anticipated; counting the exponential increase of isomers for dimeric and trimeric xanthene species, many more than 19,000 compounds were generated.

Another cogent point in support of our combinatorial method is the following: regardless of what building blocks or core molecules are used in synthesis, the process will select the most active compounds, if any, which are produced. If the AMB building block had been Boc protected from the start, it seems unlikely from the data we collected that the initial library **X2** would have shown the exceptional inhibition which was observed already in Round I. The grouping of building blocks for the initial libraries (see experimental) was such that only libraries **B2** and **X2** contained AMB, and of the entire set of building blocks only AMB and Trp contain unprotected nitrogens (See Figure 48). (Trp was present only in **B5** and **X5**, both inactive.) However, the "mistake" of addition of unprotected AMB having been made, the deconvolution process of iterative selection was able to lock onto and follow the active species: the combinatorial method was unaffected by the fact that these species were not the intended tetra-substituted xanthenes. As introduced, combinatorial chemistry's great advantage over traditional methods of drug design is that it requires no foreknowledge of the molecular recognition -- or even the chemical makeup -- of the libraries synthesized.

The results of our experiment screening combinatorial libraries against DNA polymerase I Klenow Fragment can add one more insight into the general field of combinatorial chemistry. Each of the initial biphenyl and xanthene libraries, as designed, encompassed theoretically 19,306 molecules. If one imagines the gel of Round I without lane **X2**, inhibition of polymerase activity is slight; no outstanding polymerase inhibitors are present. One may postulate from this that perhaps, *on average*, a set of combinatorial compounds on the order of 10^4 molecules is not sufficient to guarantee the effective molecular recognition of a given target. When one extends the library to nearer 10^6 compounds however, as is easily the case if dimers and trimers are being formed in **X2**, potent molecular recognition is perhaps assured. Alternatively, one might explain this result in a more trivial fashion: that

the surface area for molecular recognition offered by dimeric or trimeric xanthene species is exponentially greater than that offered by monomeric tetra-substituted xanthenes.

In both the trypsin and polymerase experiments, molecular recognition was achieved through combinatorial synthesis. In the former case, starting from roughly 50,000 compounds, recognition was achieved to a K_i of 10 μM. In the latter case, starting from a library of compounds which probably numbered in the hundreds of thousands, recognition was achieved (so far) to a K_i of 1 μM.

Having proven the concept of our poly-functionalized molecules first through mass spectrometry and then through two solution-phase screening experiments against trypsin and the Klenow Fragment of DNA polymerase I, we are hopeful that our combinatorial methods may be a valuable tool in the search for potent therapeutic lead compounds, complementing the existing repertoire of combinatorial procedures to explore the landscape of molecular recognition in small molecules.

REFERENCES

(1) Stryer, L., ed. <u>Biochemistry</u>, 3rd edition, W. H. Freeman and Company: New York, 1988.

(2) Creighton, T. E. <u>Proteins: Structures and Molecular Properties</u>, W. H. Freeman and Company: New York, 1984.

(3) Saenger, W. <u>Principles of Nucleic Acid Structure</u>, Springer-Verlag: New York, 1984.

(4) Williams, D. H.; Searle, M. S.; Mackay, J. P.; Gerhard, U.; Maplestone, R. A. *Proc. Natl. Acad. Sci.* **1993**, *90*, 1172-1178.

(5) a) Askew, B.; Ballester, P.; Buhr, C.; Jeong, K. S.; Jones, S.; Parris, K.; Williams, K.; Rebek, J. Jr. *J. Am. Chem. Soc.* **1989**, *111*, 1082-1090.
b) Williams, K.; Askew, B.; Ballester, P.; Buhr, C.; Jeong, K. S.; Jones, S.; Rebek, J. Jr. *ibid* 1090-1094.

(6) Tjivikua, T.; Ballester, P.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1990**, *112*, 1249-1250.

(7) Nowick, J. S.; Feng, Q.; Tjivikua, T.; Ballester, P.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 8831-8839.

(8) von Kiedrowski, G. *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 932-935.

(9) von Kiedrowski, G.; Wlotzka, B.; Helbing, J.; Matzen, M.; Jordan, S. *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 423-426.

(10) Terfort, A.; von Kiedrowski, G. *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 654-656.

(11) a) Orgel, L. E. Cold Spring Harbor Symp. Quant. Biol. **1987**, *52*, 9-16.
b) Joyce, G. F. *ibid*, **1987**, *52*, 41-51.

(12) Orgel, L. E. *Nature* **1992**, *358*, 203-209.

(13) a) Bachmann, P. A.; Walde, P.; Luisi, P. L.; Lang, J. *J. Am. Chem. Soc.* **1991**, I, 8204-8209.

(14) Bachmann, P. A.; Luisi, P. L.; Lang, J. *Nature* **1992**, *357*, 57-59.

(15) a) Kelly, T. R.; Bridger, G. J.; Zhao, C. *J. Am. Chem. Soc.* **1990**, *112*, 8024-8034.
b) Kelly, T. R.; Zhao, C.; Bridger, G. J. *J. Am. Chem. Soc.* **1989**, *111*, 3744-3745.

(16) Goodwin, J. T.; Lynn, D. G. *J Am. Chem. Soc.* **1992**, *114*, 9197-9198.

(17) For reviews of template effects, see:
a) Anderson, S.; Anderson, H. L.; Sanders, J. K. M. *Acc. Chem. Res.* **1993**, *26*, 469-475.
b) Hoss, R.; Vögtle, F. *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 375-384.

(18) a) Mason, S. F. "Chemical Evolution: origins of the elements, molecules and living systems" Oxford: Clarendon Press, **1991**.
b) Weiss, A. H.; Socha, R. F.; Likholobov, V. A.; Sakharov, M. M. *Chemtech.* **1980**, *10*, 643-647. (The formose reaction involves the polymerization of formaldehyde.)

(19) a) Kemp, D. S.; Petrakis, K. S. *J. Org. Chem.* **1981**, *46*, 5140-5143.
b) An improved synthesis is given in: Rebek, J., Jr.; Askew, B.; Killoran, M.; Nemeth, D.; Lin, F.-T. *J. Am. Chem. Soc.* **1987**, *109*, 2426-2431.

(20) For a review see: Rebek, J. Jr. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 245-255.

(21) Hamilton, A. D.; Van Engen, D. *J. Am. Chem. Soc.* **1987**, *109*, 5035-5036.

(22) a) Rony, P. R. *J. Am. Chem. Soc.* **1969**, *91*, 6090-6096.
b) Swain, C. G.; Brown, J. F., Jr. *J. Am. Chem. Soc.* **1952**, *74*, 2538-2543.

(23) a) Menger, F. M.; Vitale, A. C. *J. Am. Chem. Soc.* **1973**, *95*, 4931-4935.
b) Menger, F. M.; Smith, J. H. *ibid* **1972**, *94*, 3824-3829.

(24) Su, C.; Watson, J. W. *J. Am. Chem. Soc.* **1974**, *96*, 1854-1857.

(25) Hogan, J. C.; Gandour, R. D. *J. Org. Chem.* **1992**, *57*, 55-61.

(26) Rotello, V.; Hong, J. I.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 9422-9423.

(27) Bizzari, P. C., Della Casa, C., Monaco, A. *Polymer* **1980**, *21*, 1065-1068.

(28) Galán, A.; de Mendoza, J.; Toiron, C.; Bruix, M.; Deslongchamps, G.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 9424-9425.

(29) Deslongchamps, G.; Galán, A.; de Mendoza, J.; Rebek, J., Jr. *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 61-63.

(30) Conn, M. M.; Deslongchamps, G.; de Mendoza, J.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1993**, *115*, 3548-3557.

(31) a) Conn, M. M.; Wintner, E. A.; Rebek, J. Jr. *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 1577-1579.
(b)   Conn, M. M. Ph. D. Thesis. Massachusetts Institute of Technology. Feb., **1994**, pp. 29-30.

(32) Wintner, E. A.; Conn, M. M.; Rebek, J. Jr. *J. Am. Chem. Soc.* **1994**, *116*, 8877-8884.

(33) MacroModel 3.5X. Mohamadi, F.; Richards, N. G.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440-467.

(34) For a review, see Lindley, J. *Tetrahedron* **1984**, *40*, 1433-1456.

(35) Jeong, K. S.; Muehldorf, A. V.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1990**, *112*, 6144-6145.

(36) For a review, see McMurry, J. *Org. React.* **1976**, *24*, 187-224.

(37) Breslow, R.; Halfon, S. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 6916-6918.

(38) Concentration of template monomer at 3.1 mM total template concentration (reaction b, Figure 16; 0.5 equiv. template added) was calculated by estimating a single binding event at 86,000 M^{-1} in 13% $THF/CHCl_3$ (between 10^5 and 576 M^{-1}) and taking the dimerization of the template to be the square of that value. Values were inserted into a model of autocatalysis developed by James Nowick (see details in reference 7). Concentration of template monomer may also be approximated using:
$$K_{dim} = [\text{template dimer}] / [\text{template monomer}]^2$$

(39) There are earlier examples of a template slowing down a reaction but improving the yield of the desired molecule; see Dietrich-Buchecker, C. O.; Sauvage, J. P. *New J. Chem.* **1992**, *16*, 277-285.

(40) Menger, F. M.; Eliseev, A. V.; Khanjin, N. A. *J. Am. Chem. Soc.* **1994**, *116*, 3613-3614.

(41) Menger, F. M.; Eliseev, A. V.; Khanjin, N. A.; Sherrod, M. J. *J. Org. Chem.* **1995**, *60*, 2870-2898.

(42) Wintner, E. A.; Tsao, B.; Rebek, J., Jr. *J. Org. Chem.* **1995**, *60*, 7997-8001.

(43) a) Wintner, E. A.; Conn, M. M.; Rebek, J., Jr. *Acc. Chem. Res.* **1994**, *27*, 198-203.

b) Conn, M. M.; Wintner, E. A.; Rebek, J. Jr. *J. Am. Chem. Soc.* **1994**, *116*, 8823-8824

(44) Reinhoudt, D. N.,; Rudkevich, D. M., de Jong, F.; *J. Am. Chem. Soc.* **1996**, *118*, *in press.*

(45) Jung, G.; Beck-Sickinger, A. G. *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 367 - 383.

(46) a) Geysen, H. M.; Meloen, R. M.; Barteling, S. J.; *Proc. Natl. Acad. Sci. USA* **1984**, *88*, 3998-4002.

b) Geysen, H. M.; Rodda, S. J.; Mason, T. J.; Tribbick, G.; Schoofs, P. G. *J. Imm. Meth.*, **1987**, *102*, 159 - 273.

(47) a) Houghten, R. A. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 5131-5135.
b) Houghten, R. A.; DeBray, S. T.; Bray, M. K.; Hoffmann, S. R.; Frizzell, N. D. *BioTechniques* **1986**, *4*, 522-528.

(48) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.

(49) a) Furka, Á.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Int. J. Pept. Prot. Res.* **1991**, *37*, 487-493.
b) Furka, Á.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Abstr. 14th Int. Congr. Biochem.*, Prague, Czechoslovakia 1988, Vol. 5, p. 47.

154

(50) a) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature (London)* **1991**, *354*, 82-84.

b) Salmon, S. E.; Lam, K. S.; Lebl, M.; Kondola, A.; Khattri, P. S.; Wade, S.; Pátek, M.; Kocis, P.; Krchnak, V.; Thorpe, D.; Felder, S. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11708-11712.

(51) Brummel, C. L.; Lee, I. N. W.; Zhous, Y.; Benkovic, S. J.; Winograd, N. *Science* **1994**, *264*, 399-402.

(52) See however the following methods of structure determination by coding schemes (molecular tags):

a) Tuerk, C.; Gold, L. *Science* **1990**, *24*, 505-510.

b) Beaudry, A. A.; Joyce, G. F. *Science* **1992**, *257*, 635-641.

c) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5381-5383.

d) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411-1418.

e) Nielsen, J.; Brenner, S.; Janda, K. D. *J. Am. Chem. Soc.* **1993**, *115*, 9812-9813.

f) Ohlmeyer, M. H. F.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G., Kobayashi, R.; Wigler, M.; Still, W. C. *Proc. Natl. Acad. Sci USA* **1993**, *90*, 10922.

g) Borchardt, A.; Still, W. C. *J. Am. Chem. Soc.* **1994**, *116*, 373-374.

h) Nestler, H. P.; Bartlett, P. A.; Still, W. C. *J. Org. Chem.* **1994**, *59*, 4723-4724.

(53) Bunin, B. A.; Ellman, J. A. *J. Am. Chem. Soc.* **1992**, *114*, 10997-10998.

(54) a) Zuckermann, R. N.; Martin, E. J.; Spellmeyer, D. C.; Stauber, G. B.; Shoemaker, K. R.; Kerr, J. M.; Figliozzi, G. M.; Goff, D. A.; Siani, M. A.; Simon, R. J.; Banville, S. C.; Brown, E. G.; Wang, L.; Richter, L. S.; Moos, W. H. *J. Med. Chem.* **1994**, *37*, 2678-2685.

b) Simon, R. S.; Martin, E. S.; Miller, S. M.; Zuckermann, R. N.; Blaney, J. M.; Moos, W. H. Techniques in Protein Chem. Part V, J. Wiley & Sons, San Diego: 1994.

c) Peptoids were recently highlighted: Kessler, H. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 543-544.

(55) Pirrung, C.; Chen, J. *J. Am. Chem. Soc.* **1995**, *117*, 1240-1245.

(56) a) Cho, C. Y.; Moran, E. J.; Cherry, S. R.; Stephaus, J. C.; Fodor, S. P. A.; Adams, C. L.; Sundaram, A.; Jacobs, J. W.; Schultz, P. G. *Science* **1993**, *261*, 1303-1305.

b) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. *Science* **1991**, *251*, 767-773.

c) Rozsnyai, L. F.; Benson, D. R.; Fodor, S. P. A.; Schultz, P. G. *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 759-761.

(57) DeWitt, S. H.; Kieley, J. S.; Stankovic, C. J.; Schroeder, M. C.; Reynolds Cody, D. M.; Pavia, M. R. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 6909 - 6913.

(58) a) Schroeder, M. C.; Reynolds Cody D. M.; Pavia, M. R. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 6909-6913.

b) The "Diversomer" topic was recently highlighted: Liskamp, R. M. J. *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 633-636.

(59) a) Carell, T.; Wintner, E. A.; Bashir-Hashemi, A.; Rebek, J., Jr. *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 2005-2007.

b) Carell, T.; Wintner, E. A.; Rebek, J., Jr. *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 2007-2110.

(60) a) Carell, T.; Wintner, E. A.; Sutherland, A. J.; Rebek, J., Jr.; Dunayevskiy, Y. M.; Vouros, P. *Chem. & Biol.* **1995**, *2*, 171-183.

(61) Bashir-Hashemi, A. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 612-613.

(62) a) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature (London)* **1991**, *354*, 84-86.

b) Pinilla, C.; Appel, J. R.; Blanc, P.; Houghten, R. A. *BioTechnique* **1992**, *13*, 901-905.

(63) For protection group strategies in peptide synthesis see:

a) Bodanszky, M. *Principles of peptide synthesis*, Springer Verlag, Heidelberg, 1984.

b) Bodanszky, M.; Bodanszky, A. *The practice of peptide synthesis*, Springer Verlag, Heidelberg, 1984.

(64) Weiss, N. A.; Hassett, M. J. *Introductory Statistics 3ed.* Addison-Wesley Publishing Comp. N. Y. 1991, p. 218-219.

(65) King, D.; Fields, C.; Fields, G. *Int. J. Pept. Prot. Res.* **1990**, *36*, 255-266.

(66) Dunayevskiy, Y. M.; Vouros, P.; Carell, T.; Wintner, E. A.; Rebek, J., Jr. *Anal. Chem.* **1995**, *67*, 2906-2915.

(67) Dunayevskiy, Y. M.; Vouros, P.; Wintner, E. A.; Shipps, G. W.; Carell, T.; Rebek, J., Jr. *Proc. Natl. Acad. Sci. USA* **1996**, *in press*.

(68) For an investigation of oligopeptide libraries with ion-spray mass spectrometry see: Metzger, J. W.; Wiesmüller, K-H.; Gnau, V.; Grünges, J.; Jung, G. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 894-896.

(69) Nowick, J. S.; Ballester, P.; Ebmeyer, F.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1990**, *112*, 8902-8906. The diacid 4 is commercially available from Aldrich Chemical Company.

(70) Smith, R. D.; Wahl, J.H.; Goodblett, D. R.; Hofstadler, S. A. *Anal. Chem.* **1995**, *65*, 574A-584A.

(71) Cai, J.; Henion J. *J. Chromatogr.* **1995**, *703*, 667-692.

(72) For the screening of a peptide library against trypsin see: Eichler, J.; Houghten, R. A. *Biochemistry* **1993**, *32*, 11035-11041.

(73) a) Holtz, J.; Goetz, R. M. *Arzneim. Forsch.* **1994**, *44(3a)*, 397-402.
b) Cody, J. R. *Drugs*, **1994**, *47*, 586-598.

(74) a) Erlanger, B. F.; Kokowsky, N.; Cohen, W. *Arch. Biochem. Biophys.* **1961**, *95*, 271-278.
b) Gaertner, H. F.; Puigserver, A. J. *Enzyme Microb. Technol.* **1992**, *14*, 150-155.

(75) Laskowski, M., Jr.; Kato, I. *Annu. Rev. Biochem.* **1990**, *49*, 593-626.

(76) Segel, I. H. *Enzyme Kinetics* John Wiley & Sons, Inc., N.Y. 1993.

(77) Kornberg, A *Science* **1969**, *163*, 1410.

(78) Klenow, H.; Henningsen, I. *Proc. Natl. Acad. Sci. USA* **1970**, 65, 168-175.

(79) Brutlag, D.; Atkinson, M. R.; Setlow, P.; Kornberg, A. *Biochem. Biophys. Res. Comm.* **1969**, *37*, 982-989.

(80) Kuchta, R. D.; Mizrahi, V.; Benkovic, P. A.; Johnson, A.; Benkovic, S. J. *Biochem*istry, **1987**, *26*, 8410-8417.

(81) Dahlberg, M. E.; Benkovic, S. J. *Biochemistry* **1991**, *30*, 4835-4843.

(82) Ollis, D. L.; Brick, P.; Hamlin, R.; Xuong, N. G.; Steitz, T. A. *Nature* **1985**, *313*, 762-766.

(83) Brick, P.; Ollis, D.; Steitz, T. A. *J. Mol. Bio.* **1983**, *166*, 453-455.

(84) Beese, L. S.; Derbyshire, V.; Steitz, T. A. *Science* **1993**, *260*, 352-355.

(85) Beese, L. S.; Friedman, J. M.; Steitz, T. A. *Biochemistry* **1993**, *32*, 14095-14101.

(86) Modak, M. J. *Biochemistry* **1976**, *15*, 3620-3626.

(87) Hazra, A. K.; Detera-Wadleigh, S.; Wilson, S. H. *Biochemistry* **1984**, *23*, 2073-2078.

(88) Woodward, R. B.; Heusler, K.; Gosteli, J.; Naegeli, P.; Oppolzer, W.; Ramage, R.; Ranganathan, S.; Vorbruggen H. *J. Am. Chem. Soc.* **1966**, *88*, 852-853.

(89) For standard molecular biology techniques see: Sambrook, J.; Fritsch, E. F.; Maniatas, T. <u>Molecular Cloning: A Laboratory Manual</u>, 2nd Edition, Cold Spring Harbor Laboratory Press: New York, 1989.

EXPERIMENTAL --- PART I

General.

All commercially available reagents were used without further purification. All solvents were purchased from Malinckrodt, and possessed analytical reagent quality. CH_2Cl_2 and THF were distilled for synthesis, DMSO and DMF were kept over oven dried molecular sieves. ^1H NMR spectra were obtained on Bruker AC-250, Varian XL-300, Varian UN-300 and Varian VXR-500 spectrometers. Mass spectra for product characterization were obtained on a Finnigan MAT 8200 system.

Terphenyl Replicator Related Compounds

4-amino-4"-carboxy-p-terphenyl (12)

For experimental procedure, see Bizzari et al. Polymer, 1980, 21, 1065-1068. [Ref 27]

4-amino-4"-benzyl-p-terphenyl carboxylate (13)

To a solution of crude 4-amino-4"-carboxy-p-terphenyl 12 (2.25 g, 7.8 mmol) and triphenylphosphine (3.19 g, 12 mmol) in 70 ml of THF was added benzyl alcohol (0.9 ml, 8.3 mmol), followed by diethyl azodicarboxylate (1.91 ml, 9.9 mmol). The reaction mixture was stirred for 2 h at RT. The resulting dark yellow solution was concentrated under reduced pressure and the residue was chromatographed on silica gel with CH_2Cl_2.

mp 186 - 189 °C.

IR (neat) 3031, 1708, 1400, 1269, 1216, 1103, 818, 772, 755, 699 cm^{-1}.

^1H NMR (300 MHz, CDCl$_3$) δ 8.13 (d, J = 7.5 Hz, 2 H), 7.68 (d, J = 7.5 Hz, 2 H), 7.65 (d, J = 7.8 Hz, 2 H), 7.62 (d, J = 7.8 Hz, 2 H), 7.38 (d, J = 7.5 Hz, 2 H), 6.76 (d, J = 7.5 Hz, 2 H), 7.47 - 7.34 (m, 5 H), 5.38 (s, 2 H), 3.74 (s, 2 H).

FABMS (3-NBA) calcd for [M+H]$^+$=380.1650, HRMS [M+H]$^+$=380.1646.

Tripropyl Kemp's imide terphenyl benzyl ester (14)

A solution of 4-amino-4"-benzyl-p-terphenyl carboxylate 13 (1.0 g, 2.64 mmol) and propyl Kemp's imide acid chloride (923 mg, 2.7 mmol) in 60 ml of pyridine was treated with catalytic amount of DMAP (10 mg, 0.08 mmol). The mixture was heated at reflux overnight and concentrated under reduced pressure. The residue was chromatographed on silica gel (eluted with CH_2Cl_2, 1% methanol in CH_2Cl_2, and 1.5% methanol in CH_2Cl_2).

mp 115 - 120 °C.

IR (neat) 3367, 2959, 1698, 1516, 1272, 1197, 1111, 755 cm^{-1}.

^{1}H NMR (300 MHz, CDCl$_3$) δ 8.14 (d, J = 8.5 Hz, 2 H), 7.68 (d, J = 8.5 Hz, 2 H), 7.66 (d, J = 5.7 Hz, 2 H), 7.63 (d, J = 5.7 Hz, 2 H), 7.55 (d, J = 9.3 Hz, 2 H), 7.52 (d, J = 9.3 Hz, 2 H), 7.47 - 7.30 (m, 5 H), 7.27 (s, 1 H), 5.38 (s, 2 H), 2.59 (d, J = 13.8 Hz, 2 H), 2.21 (d, J = 13.2 Hz, 1 H), 2.01 - 1.91 (m, 3 H), 1.53 - 1.20 (m, 12 H), 0.93 (t, J = 7.0 Hz, 6 H), 0.85 (t, J = 7.2 Hz, 3 H).

FABMS (glycerol and DMSO) exact mass for [M+H]$^+$=685.3641, HRMS [M+H]$^+$=685.3638.

Tripropyl Kemp's imide terphenyl acid (15)

A mixture of terphenyl benzyl ester **14** (285 mg, 0.416 mmol) and palladium on activated carbon, 10% (50 mg, catalytic) in ethanol was stirred under H$_2$ atmosphere (balloon) overnight. The mixture was filtered through a short pad of Celite washing with hot THF to give the acid as a white solid, 235 mg.

^{1}H NMR (250 MHz, DMSO$_{d6}$) δ 10.34 (s, 1H), 9.24 (s, 1H), 8.04 (d, J=8.2 Hz, 2 H), 7.75-7.85 (m, Ar, 6 H), 7.56-7.67 (m, Ar, 4 H), 2.63 (d, J=13.6 Hz, 2 H), 2.02 (d, J=13.5 Hz, 1 H), 1.72-1.82 (m, 2 H), 1.13 - 1.50 (m, 13 H), 0.89 (t, J=7.1 Hz, 6 H), 0.79 (t, J=6.9 Hz, 3 H).

FABMS (glycerol and DMSO) calcd for [M+H]$^+$=595.3172, HRMS [M+H]$^+$=595.3169

Tripropyl Kemp's imide terphenyl pentafluorophenyl ester (11)

The tripropyl Kemp's terphenyl acid **15** (190 mg, 0.32 mmol) in 7 ml of THF was treated successively with pentafluorophenol (70 mg, 0.38 mmol) and EDC (70 mg, 0.36 mmol) with cat. DMAP (5 mg, 0.04 mg). The mixture was concentrated under reduced pressure. The residue was chromatographed (eluted with CH$_2$Cl$_2$, 1% methanol in CH$_2$Cl$_2$, and 2% methanol in CH$_2$Cl$_2$) to give a white solid, 135 mg, 56%.

mp 135 - 140 °C.

IR (neat) 3372, 2961, 1759, 1698, 1520, 1254, 1050, 1004, 764 cm^{-1}.

^{1}H NMR (250 MHz, CDCl$_3$) δ 8.26 (d, J = 8.5 Hz, 2 H), 7.79 (d, J = 8.5 Hz, 2 H), 7.71 (d, J = 8.2 Hz, 2 H), 7.66 (d, J = 8.2 Hz, 2 H), 7.57 (d, J = 9.0 Hz, 2 H), 7.52 (d, J = 9.0 Hz, 2 H), 2.59 (d, J = 14.4 Hz, 2 H), 2.23 (d, J = 13.2 Hz, 1 H), 2.20 - 1.92 (m, 3 H), 1.54 - 1.17 (m, 12 H), 0.94 (t, J = 7.2 Hz, 6 H), 0.86 (t, J = 7.2 Hz, 3 H).

FABMS (3-NBA) calcd for [M+H]$^+$=761.3013, HRMS [M+H]$^+$=761.3008.

Tripropyl Kemp's imide terphenyl *o*-chloro ester (16): 130 mg (0.22 mmol) Kemps imide terphenyl carboxylic acid **15**, 42 mg (0.22 mmol) EDC·HCl, and 5 mg DMAP (0.04 mmol, cat.) were placed in a flask under Ar. 70 µl *o*-Chloro phenol in 5 ml dry THF were syringed in and the reaction allowed to stir overnight. The reaction solution was rotovapped to a yellow oil and chromatographed on silica gel (0-3% acetone in CH_2Cl_2) to yield a clear oil, and upon rotovapping with hexane this oil yielded a white solid. The solid was put under vacuum overnight. Yield 130 mg (0.18 mmol), 84%.

^{1}H NMR (250 MHz, $CDCl_3$) δ 8.29 (d, J=8.2 Hz, 2H), 7.77 (d, J=8.1 Hz, 2H), 7.67-7.73 (m, Ar, 4H), 7.47-7.59 (m, Ar, 4H), 7.20-7.30 (m, Ar, 4H), 2.58 (d, J=14.3 Hz, 2H), 2.23 (d, J=14.1 Hz, 1H), 1.96 (m, 2H), 1.23-1.52 (m, 13H), .94 (t, J=6.7, 6H), .86 (t, J=6.8, 3H). FABMS (3-NBA) calcd for $[M+H]^+$=705.3095, HRMS $[M+H]^+$=705.3094

Terphenyl template (17)

A mixture of terphenyl active ester **11** (135 mg, 0.18 mmol) and 5'-aminoadenosine (55 mg, 0.18 mmol) in 2 ml of THF was treated with triethylamine (0.5 ml). The reaction mixture was stirred overnight at RT and concentrated under reduced pressure. The residue was chromatographed (eluted with CH_2Cl_2, 1% methanol in CH_2Cl_2, 2%, 3%, and 4% methanol in CH_2Cl_2) to give a white solid (145 mg, 91% yield).

mp 180 - 184 °C.

IR (neat) 3338, 2960, 1694, 1644, 1516, 1490, 1213, 1096, 1004, 755 cm^{-1}.

NMR was taken by Jong-In Hong:

^{1}H NMR (300 MHz, $CDCl_3$) δ 12.66 (s, 1H), 8.03 (s, 1 H), 7.80 (br, d, 1 H), 7.74 (s, 1 H), 7.73 (d, J = 8.4 Hz, 2 H), 7.52 (d, J = 8.1 Hz, 2 H), 7.46 (d, J = 8.1 Hz, 2 H), 7.33 (d, J = 8.7 Hz, 2 H), 7.21 (d, J = 8.7 Hz, 2 H), 5.69 (d, J = 5.7 Hz, 1 H), 5.32 (t, J = 6.0 Hz, 1 H), 4.83 (d, J = 6.6 Hz, 1 H), 4.57 (s, 1 H), 4.54 (d, J = 9.6 Hz, 1 H), 3.37 (d, J = 13.5 Hz, 1 H), 2.64 (d, J = 14.4 Hz, 1 H), 2.59 (d, J = 14.4 Hz, 1 H), 2.31 (d, J = 12.3 Hz, 1 H), 2.02 (m, 2 H), 1.51 - 1.21 (m, 13 H), 0.98 (t, J = 5.1 Hz, 6 H), 0.82 (t, J = 7.2 Hz, 3 H). FABMS (3-NBA) calcd for $[M+H]^+$=883.4506, HRMS $[M+H]^+$=883.4509.

N-methyl tripropyl Kemp's imide terphenyl pentafluorophenyl ester (18):

This compound was provided by Postdoctoral Fellow Jong-in Hong. Synthesis is identical to that of the non-methylated compound **11**, except that N-methylated tripropyl Kemp's imide acid chloride [Ref. 26] is used in the acylation step of the 4-amino-4"-benzyl-p-terphenyl carboxylate **13**.

HPLC Kinetics

All reactions were performed at 50 mM initial concentrations of reactants **4** and **11** or **16** in $CHCl_3$ with with 1.0% TEA base. All solutions were freshly prepared for each kinetic run. Formation of product **17** was followed by HPLC at 254 nm on a Waters 600E instrument equipped with a Waters 490E UV detector. Product formation was followed for the first 5-10% of the reaction; initial rates were determined by linear fitting of the absolute size of the product peak areas. Kinetic runs showed excellent linearity and good convergence to the origin (See text Figures 10, 11).

HPLC separation was achieved using a Beckman Ultrasphere ODS (C-18) column (4.6 mm ID x 25 cm length). The elution system was:

14/86/0.2% H_2O/MeOH/TEA: 2 ml/min for 15 min.

Under these conditions, the retention times of amine **4**, product **17** and ester **11/16** were 1.5, 3.5, and 10.9 min, respectively.

Sample: ewa20
Acquired: 23-APR-92 15:54

Channel: detector 1
Method: C:\BASE\DATA\TJAMA

Filename: ewa20
Operator: ED Wintner.

volts

0.50   1.00   1.50   2.00   2.50

Amine

Template

Ester

x 10¹ minutes

0.20   0.40   0.60   0.80   1.00   1.20   1.40
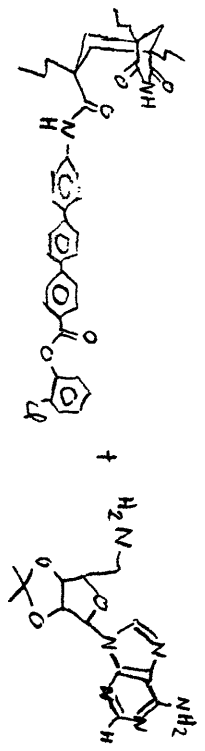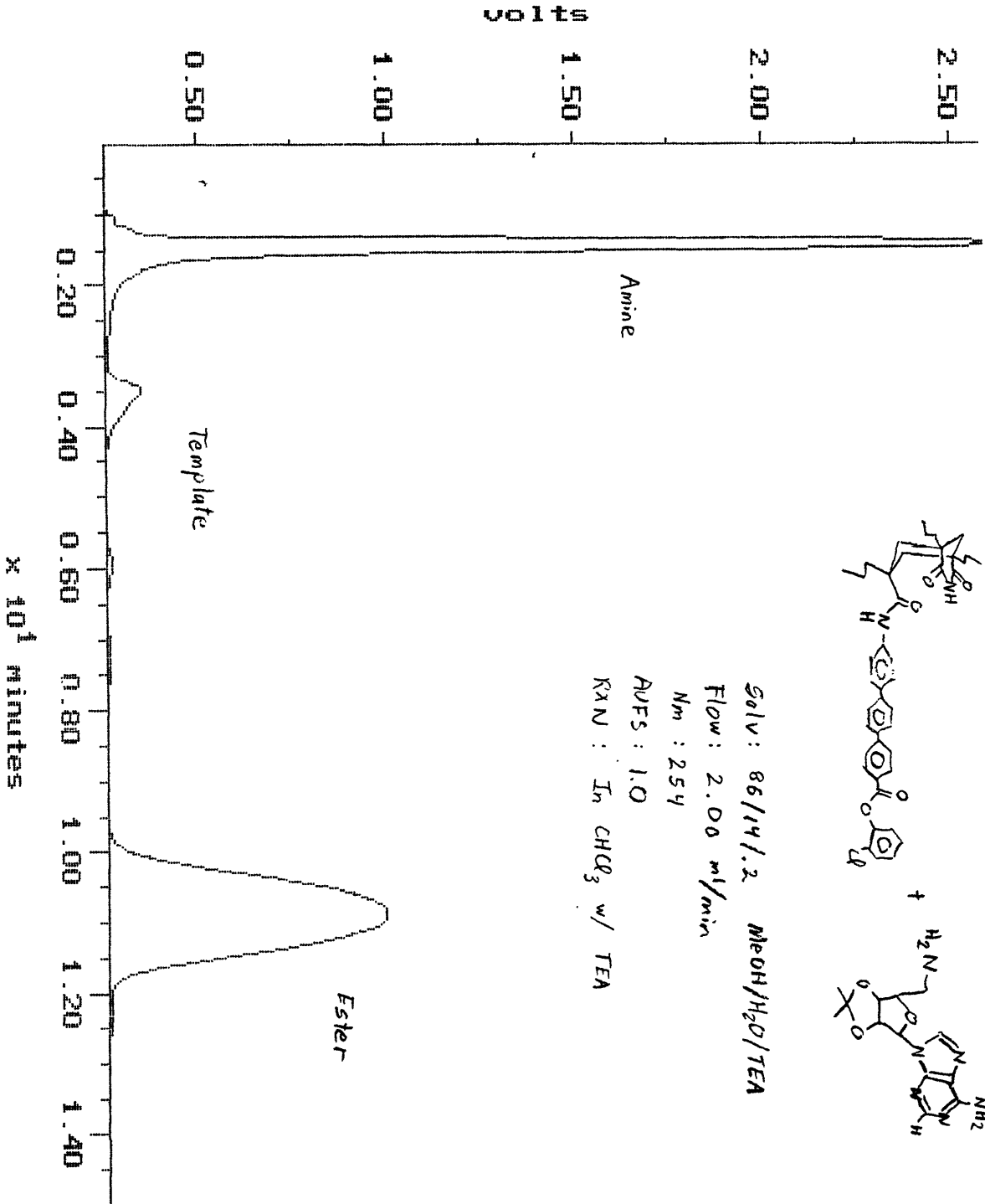
Solv: 86/14/.2 MeOH/H₂O/TEA

Flow: 2.00 ml/min

Nm: 254

AUFS: 1.0

RXN: In CHCl₃ w/ TEA

Biphenyl Diimide Related Compounds

Except for the compounds noted here, all synthesis of the biphenyl diimide replicator was undertaken by collaborator Morgan Conn. An experimental can be found in Ref. 32.

Tripropyl Kemp's imide carbzole biphenyl pentafluorophenyl ester (21)

65 mg (0.065 mmol) of the corresponding tripropyl Kemp's imide carbzole biphenyl acid (Morgan Conn, [Ref. 32]), 44 mg (0.24 mmol) of pentafluorophenol, 32 mg (0.17 mmol) of EDC [1-ethyl-3-(3-Dimethylaminopropyl)-carbodiimide], and 10 mg (0.081 mmol) of DMAP [4-dimethylaminopyridine] were stirred in dry THF under Ar for 18 hrs. The crude product mixture was concentrated to an oil and was chromatographed on silica with 1.5%-20% EtOAC/CH_2Cl_2 gradient elution. The fractions containing the ester were concentrated to a white solid, 60 mg (79%).

^{1}H NMR (250 MHz, $DMSO_{d6}$) δ 10.39 (imide,s,2H), 9.26 (amide, s,2H), 8.34-8.35 (Ar, 4H), 8.12 (d, J=8.5 Hz, 4H), 7.79 (d, J= 8.5 Hz, 2H), 7.52 (d, J=8.7 Hz, 2H), 7.41 (d, J= 8.7 Hz, 2H), 2.70(d, J= 12.7 Hz, 4H), 2.04 (d, J= 12.5 Hz, 2H), 1.7-1.8 (m,4H), 1.60 (m,4H), 1.0-1.5 (CH_2,m,22H), .88 (CH_3,t,12H), .73 (CH_3,t,6H). >95% purity by HPLC.

Tripropyl Kemp's imide carbzole biphenyl 2,4-dinitrophenyl ester (22)

70 mg (0.070 mmol) of the corresponding tripropyl Kemp's imide carbzole biphenyl acid (Morgan Conn, [Ref. 32]) were stirred in dry THF with 25 mg (0.136 mmol) of 2,4-dinitrophenol, 29 mg (0.151 mmol) of EDC [1-ethyl-3-(3-Dimethylaminopropyl)-carbodiimide], and 4.2 mg (0.034 mmol, cat.) of DMAP [4-dimethylaminopyridine]. Stirring was continued at RT under Ar for 24 hrs after which the crude product mixture was concentrated to an oil. The oil was chromatographed on silica with 2%-10% EtOAC/CH_2Cl_2 gradient elution. The fractions containing the desired ester were rotovapped with chloroform and hexanes to yield a white solid (60%).

^{1}H NMR (300 MHz, $DMSO_{d6}$) δ 10.37 (imide,s,2H), 9.24 (amide, s,2H), 8.95 (d, J=3.3 Hz, 1H), 8.74 (dd, J=3.3, 9.0, 1H), 8.25-8.35 (Ar,m,4H), 8.05-8.15 (Ar,m,4H), 8.03 (d, J=9 Hz, 1H), 7.78 (d, J=8.7 Hz, 2H), 7.51 (dd, J=1.8, 8.7 Hz, 2H), 7.40 (d, J=8.7 Hz, 2H), 2.69 (d, J=13.2 Hz, 4H), 2.04 (d, J=13.5 Hz, 2H), 1.8 (m,4H), 1.6 (m,4H), 1.0-1.5 (CH_2,m,22H), .9 (CH_3,t,12H), .8 (CH_3,t,6H). >95% purity by HPLC.

Biphenyl diimide template (23)

Tripropyl Kemp's imide carbzole biphenyl pentafluorophenyl ester **21** (50mg, 0.045 mmol), amino adenosine **4** (17 mg, 0.054 mmol) and DMAP (4 mg, 0.032 mmol) were refluxed 2h. in THF under Ar. The solution was concentrated to a brown solid and chromatographed on silica gel, 5% MeOH in CH_2Cl_2. The product was concentrated with hexanes to yield a yellow tinged solid (76%).

IR (KBr) 3374, 3214, 2958, 1696, 1647, 1491, 1465, 1204, 1079 cm^{-1};

^{1}H NMR (250 MHz, DMSO$_{d6}$) δ 10.40 (s, 2 H, imide), 9.24 (s, 2 H, amide), 8.75 (t, J=5.5 Hz, 1H), 8.35 (s, 1 H), 8.32 (s, 1 H), 8.12 (d, 2 H, J = 1.5 Hz), 8.03 (d, 2 H, J = 8.5 Hz), 7.98 (d, 2 H, J = 8.5 Hz), 7.90 (d, 2 H, J = 8.5 Hz), 7.72 (d, 2 H, J = 8.0 Hz), 7.49 (dd, 2 H, J = 9.0, 1.5 Hz), 7.44 (br s, 2 H, amine), 7.37 (d, 2 H, J = 8.5 Hz), 6.17 (d, 1 H, J = 2.5 Hz), 5.50 (dd, 1 H, J = 6.0, 2.5 Hz), 5.08 (dd, 1 H, J = 6.2, 3.2 Hz), 4.34 (m, 1 H), 3.56 (m, 2 H), 2.68 (d, 4 H, J = 13.5 Hz), 2.03 (d, 2 H, J = 12.5 Hz), 1.78 (m, 4 H), 1.54 (s, 3 H), 1.51 (m, 4 H), 1.32 (s, 3 H), 1.35-1.1 (m, 22 H), 0.87 (t, 12 H, J = 7.0 Hz), 0.80 (t, 6 H J = 7.0 Hz);

HRMS (FAB in 3-nitrobenzyl alcohol) calcd for $C_{74}H_{90}N_{11}O_{10}$ (M+H), 1292.6872; found, 1292.6860.

Trimethyl Kemp's imide phenyl methyl ester (24).

Reaction of methyl 4-aminobenzoate and Kemp's imide acid chloride was performed using previously published methodology [Askew *et al.*, Ref. 5a] to give **24** as a white solid.

mp 290-295 °C;

IR (KBr) 3357, 3189, 3097, 2978, 2930, 1720, 1689, 1596, 1529 cm^{-1};

^{1}H NMR (250 MHz, CDCl$_3$) δ 7.98 (d, 2 H, J = 8.7 Hz), 7.52 (d, 2 H, J = 8.7 Hz), 7.50 (s, 1 H), 3.90 (s, 3 H), 2.66 (d, 2 H, J = 13.7 Hz), 2.04 (d, 1 H, J = 13.7 Hz), 1.44 (d, 2 H, J = 13.3 Hz), 1.36 (d, 1 H, J = 3.8 Hz), 1.33 (s, 3 H), 1.31 (s, 6 H);

HRMS calcd for $C_{20}H_{24}N_2O_5$, 372.1685; found, 372.1683.

N-(Cyclohexylmethyl)benzamide (25)

Benzoyl chloride (100 µL, 0.86 mmol) was dissolved in anhydrous THF (5 ml) under argon. Cyclohexane methylamine (1.027 equiv) and TEA (2.5 equiv) were added, accompanied by formation of a white precipitate. After stirring at RT for 15 h, the reaction mixture was filtered and the filtrate concentrated. Pure amide 25 (155 mg, 83%) was isolated following chromatography (3:6:1 CHCl3:hexanes :EtOAc).

mp 109-110 °C

IR (KBr) 3344, 2914, 2847, 1636, 1548, 1492, 1442 cm^{-1}

^{1}H NMR (500 MHz, CDCl3) δ 7.760 (d, 2 H, J = 7.0 Hz), 7.490 (m, 1 H), 7.427 (m, 2 H), 6.168 (br s, 1 H, amine), 3.307 (t, 2 H, J = 6.5 Hz), 1.82-1.72 (m, 4 H), 1.70-1.65 (m, 1 H), 1.63-1.55 (m, 1 H), 1.30-1.13 (m, 3 H), 1.05-0.95 (m, 2 H)

HRMS calcd for $C_{14}H_{19}N_1O_1$, 217.1467; found, 217.1466.

HPLC Kinetics

All reactions of 4 + 21 were performed in teflon capped autoinjector vials at 6.2 mM initial concentrations of reactants in 13% THF/CHCl3 with 1.0% TEA base. A THF stock solution of 4 was prepared daily. All other solutions were freshly prepared for each kinetic run. Formation of products was followed by HPLC at 270 nm on a Waters 600E instrument equipped with a Waters 717 autosampler and a Waters 490E UV detector. Temperature inside the autosampler was constant at 22±1 °C in an individually thermostated room. Product formation was followed for the first 5-10% of the reaction; initial rates were determined by linear fitting of the absolute size of the product peak areas. Kinetic runs showed excellent linearity and good convergence to the origin (See text Figure 16 for individual runs). HPLC separation was achieved using a Beckman Ultrasphere SI column (4.6 mm ID x 25 cm length).

The elution system was 1% MeOH/CHCl3: 1.5 mL/min to 3.0 mL/min over 4 min Followed by 5% MeOH/CHCl3: 4 mL/min for 11 min.

Under these conditions, the retention times of ester 21 and product 23 were 2.1 and 8.2 min, respectively. Aminoadenosine 4 was retained on the column and removed by flushing the column with 1/10/89% TEA/MeOH/CHCl3 after each experiment (usually 6-8 injections).

Sample: edd18   Channel: detector 1   Filename: edd18
Acquired: 19-MAY-93 13:12   Method: C:\MAX\DATA1\EAWMET ·   Operator: mbp
Comments: Ultrasphere Si. 1% MeOH/CHCl3, 5% MeOH/CHC13

volts

0.40   0.60   0.80   1.00   1.20

0.00   0.20   0.40   0.60   0.80   1.00   1.20

x 10^1 minutes

PFP Ester

Template

Biphenyl-Carbazole-Diimide System
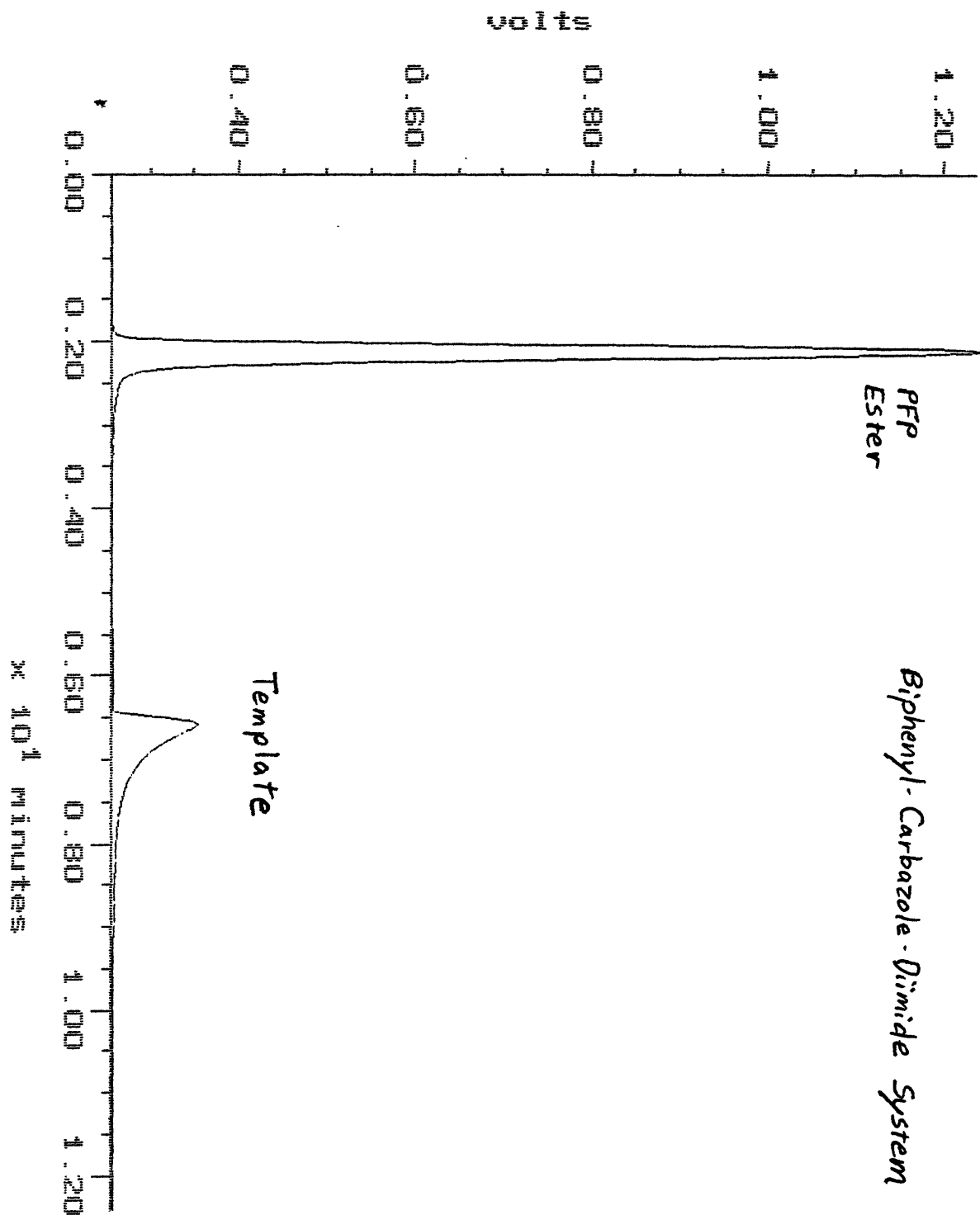
UV Kinetics

The graph in Figure 20 was created by experiments on a Perkin Elmer Lambda 12 UV/Vis spectrophotometer. A typical experiment is shown below (for 15% THF):

Stock Solutions A through D were made to the concentrations and % solvents shown in the table below. A large excess of triethylamine was added to ensure that all dinitrophenol released would be present as the dinitrophenolate (UV_{max} 450 nm). HPLC data had previously indicated that triethylamine excess did not alter the rate of formation of template. The microliter amounts shown below were added to their respective cuvettes, with adenosine **4** last at T=0.

Microliter Amounts Added:	A 0.15 mM Dinitro Ester **22** 100% CHCl$_3$	B 2.0 mM Adenosine **4** 50% THF/ CHCl$_3$	C 120 mM TEA 75% THF/ CHCl$_3$	D 0.53 mM Diimide Template **23** 100% CHCl$_3$	CHCl$_3$
Cuvette 1	700	300	100	0	400
Cuvette 2	700	300	100	100	300
Cuvette 3	700	300	100	400	0

This generated three cuvettes of 1.5 ml of 15% THF/CHCl$_3$ with the following molarities: 0.07 mM ester **22**, 0.40 mM amine **4**, 8.0 mM TEA, and 0, 0.035, and 0.14 mM Template **23**. The machine was autozeroed with a solution of 100 µl C, 150 µl THF, and 1250 µl CHCl$_3$.

UV data was accumulated starting from 5.0 min for 40 min, following the absorption at 450 nm every 1.0 min.

The data generated from this run is graphed below.

Dinitro Scorpion Kinetics in 15% THF/CHCl3
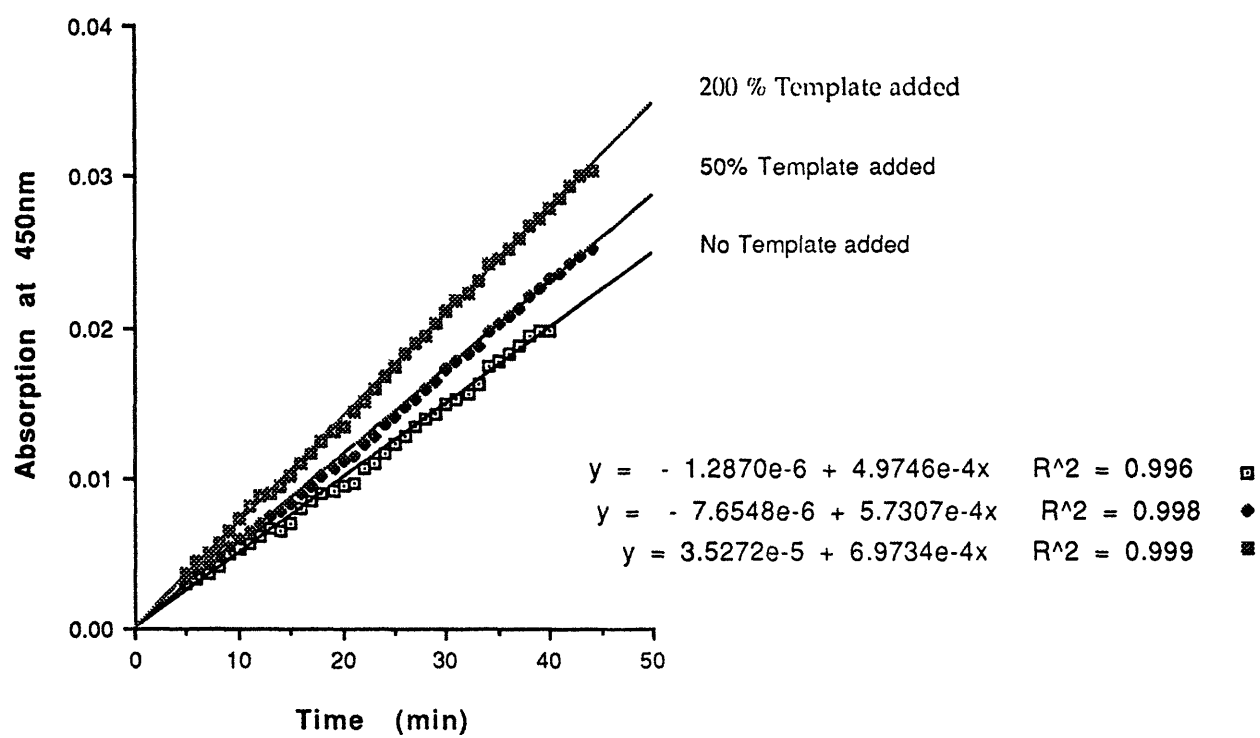
Naphthoyl Replicator Related Compounds

Unless listed here, all compounds were made according to the procedures published for the original naphthoyl system [see Nowick *et al.*, Ref. 7].

Deuterated benzyl amine

To an icecooled solution of $LiAlD_4$ (1 g, 0.024 mol) in 50 ml anhydrous THF was added a solution of benzylnitrile (2.2 ml, 0.022 mol) in 50 ml anhydrous THF dropwise under argon atmosphere. After addition was completed, the mixture was allowed to warm to room temperature and was stirred overnight. The mixture was then cooled with an icewater bath and excess $LiAlD_4$ was quenched with 10% NaOH. After fizzing had stopped, the white slurry was extracted three times with ethyl acetate. The organics were combined and dried over anhydrous $MgSO_4$. Solvent was removed by rotory evaporation and dried under vacuum. Yield = 52%.

^1H NMR (300 MHz, $CDCl_3$) δ 7.3 (m, 5H), 1.5 (s, 2H);

^{13}C NMR (300 MHz, $CDCl_3$) δ 143.81, 128.93, 127.49, 127.17, 46.30 (p, J = 20.4)

Deuterated 2-(((benzyl)amino)carbonyl)-naphthalene (36)

To a mixture of deuterated benzyl amine (0.567 g, 5.3 mmol), Et_3N (0.72 ml, 5.2 mmol) in anhydrous THF (50 ml) was slowly added a solution of 2-naphthoyl-chloride (1 g, 5.2 mmol) in 50 ml anhydrous THF at RT and under Ar atmosphere. This was stirred overnight and $Et_3N.HCl$ salt was removed by filtration through celite. The solvent was removed by rotory evaporation. Chromatography on silica gel with 1:1 mixture of ethyl acetate / hexanes as eluant gave the deuterated amide. Yield = 31%.

mp 133-135 °C (dec)

IR (KBr) 3289, 3054, 1636, 1624, 1535, 1504, 1400, 1316 cm^{-1}

^1H NMR (300 MHz, $CDCl_3$) δ 8.3 (s, 1H), 7.9 (m, 4H), 7.5 (m, 2H), 7.4 (m, 5H), 6.5 (s, 1H).

HRMS (EI) calcd for $C_{18}H_{13}D_2NO$, 463.1277; found, 463.1272.

N-(Cyclohexylmethyl)benzamide (37)

Benzoyl chloride (100 μL, 0.86 mmol) was dissolved in anhydrous THF (5 mL) under Ar. Cyclohexane methylamine (1.027 equiv) and TEA (2.5 equiv) were added, accompanied by formation of a white precipitate. After stirring at room temperature for 15 h, the reaction mixture was filtered and the filtrate concentrated. Pure amide 37 (155 mg, 83%) was isolated following chromatography (3:6:1 CHCl3:hexanes: EtOAc).

mp 109-110 °C

IR (KBr) 3344, 2914, 2847, 1636, 1548, 1492, 1442 cm^{-1}

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.760 (d, 2 H, J = 7.0 Hz), 7.490 (m, 1 H), 7.427 (m, 2 H), 6.168 (br s, 1 H, amine), 3.307 (t, 2 H, J = 6.5 Hz), 1.82-1.72 (m, 4 H), 1.70-1.65 (m, 1 H), 1.63-1.55 (m, 1 H), 1.30-1.13 (m, 3 H), 1.05-0.95 (m, 2 H);

HRMS calcd for C$_{14}$H$_{19}$N$_1$O$_1$, 217.1467; found, 217.1466.

Trimethyl Kemp's imide phenyl methyl ester (39).

Reaction of methyl 4-aminobenzoate and Kemp's imide acid chloride was performed using previously published methodology [Askew et al., ref. 5a] to give 39 as a white solid

mp 290-295 °C

IR (KBr) 3357, 3189, 3097, 2978, 2930, 1720, 1689, 1596, 1529 cm^{-1}

^{1}H NMR (250 MHz, CDCl$_3$) δ 7.98 (d, 2 H, J = 8.7 Hz), 7.52 (d, 2 H, J = 8.7 Hz), 7.50 (s, 1 H), 3.90 (s, 3 H), 2.66 (d, 2 H, J = 13.7 Hz), 2.04 (d, 1 H, J = 13.7 Hz), 1.44 (d, 2 H, J = 13.3 Hz), 1.36 (d, 1 H, J = 3.8 Hz), 1.33 (s, 3 H), 1.31 (s, 6 H)

HRMS calcd for C$_{20}$H$_{24}$N$_2$O$_5$, 372.1685; found, 372.1683.

5'-(((2-naphthyl)carbonyl)amino)-5'-deoxy-2',3'-isopropylideneadenosine (40)

Aminoadenosine 4 (51 mg, 0.17 mmol) [Ref. 7] and 2-naphthoyl chloride (37 mg, 0.19 mmol) were dissolved in anhydrous THF (10 mL) with an excess of TEA (9 equiv) under Ar, accompanied by the immediate formation of a white precipitate. The reaction was stirred at RT for 15 h, and filtered to remove TEA•HCl. After concentration, the residue was purified by flash chromatography (5% MeOH/CHCl3) to yield a the product as a white powder (76 mg, 0.165 mmol, 97%).

mp 145-150 °C (dec)

IR (KBr) 3322, 3172, 2928, 1644, 1598, 1533, 1474 cm^{-1}

^{1}H NMR (250 MHz, DMSO$_{d6}$) δ 8.836 (t, 1 H, J = 5.5 Hz), 8.433 (s, 1 H), 8.346 (s, 1 H), 8.073 (s, 1 H), 8.03-7.88 (m, 4 H), 7.65-7.55 (m, 2 H), 7.355 (br s, 2 H, amine), 6.166 (d, 1 H, J = 2.8 Hz), 5.500 (dd, 1 H, J = 6.3, 2.8 Hz), 5.091 (dd, 1 H, J = 6.3, 3.3 Hz), 4.351 (m, 1 H), 3.588 (m, 2 H), 1.530 (s, 3 H), 1.314 (s, 3 H)

HRMS (EI) calcd for C$_{24}$H$_{24}$N$_6$O$_4$, 460.1859; found, 460.1862.

Trimethyl Kemp's imide naphthyl cyclohexylamide (41)

The corresponding imide-napthyl-carboxylic acid (58 mg, 0.14 mmol) [Ref. 7], 1-ethyl-3-(3-dimethyl-1-aminopropyl)-carbodiimide (EDC, 40 mg, 0.21 mmol), and dimethylaminopyridine (DMAP, 5 mg, 0.04 mmol) were stirred in 8 ml anhydrous THF under Ar. Cyclohexylmethylamine (55 μl, 0.42 mmol) was added by syringe, and the solution stirred 24 hrs. The solution was evaporated and the crude solid purified by flash chromatography (40% EtOAC/Hex) to yield a clear oil. Product was precipitated from CHCl3 with hexanes to yield 41 (45 mg, 0.09 mmol, 64%) as a white powder.

mp 118-123 °C (dec)

IR (KBr) 3180, 2925, 2851, 1750, 1702, 1645, 1541, 1457, 1314, 1202, 1151 cm^{-1}

^{1}H NMR (300 MHz, DMSO-d_6) δ 10.861 (s, 1H,), 8.591 (t, 1H, J = 4.8 Hz), 8.455 (s, 1H), 8.055 (d, 1H, J = 9.0 Hz), 7.946 (s, 2H), 7.670 (d, 1H, J = 2.1 Hz), 7.316 (dd, 1H, J = 9.0, 2.1 Hz), 3.150 (t, 2H, J = 6.3 Hz), 2.520 (d, 2H, DMSO obsc.), 2.018 (d, 1H, J = 13.0 Hz), 1.55-1.80 (m, 6H), 1.502 (d, 1H, J= 12.9 Hz), 1.434 (d, 2H, J= 14.1 Hz), 1.384 (s, 3H), 1.175-1.225 (m, 3H), 1.151 (s, 6H), 0.5-1.0 (m, 2H)

HRMS (FAB+) calcd for C$_{30}$H$_{36}$N$_2$O$_5$, (M+H) 505.2702; found, 505.2706.

NMR Kinetics.

The NMR experiments shown in Table 3 were performed by Belinda Tsao. All ^{1}H NMR spectra were taken in CDCl$_3$ on a Varian Unity 300 MHz spectrometer with temperature control. Chemical shifts in parts per million are reported relative to residual solvent peak.

Coupling reactions of **33+34** were carried out at 25±0.3 °C adding benzyl amine **34** in CDCl$_3$ to a solution of naphthoyl pentafluorophenyl ester **33** in CDCl$_3$ and 0.01 equivalent of Et$_3$N with or without the deuterated amide **36**. Spectra were taken every 2 hours until at least 10% of the product was formed. Initial velocities of the reactions were determined through integration of the methylene peak of product amide **35** at 4.72 ppm relative to the methylene of benzyl amine **34** at 3.88 ppm.

Coupling reactions of **4 + 44** were carried out at 25±0.3 °C by adding adenosine amine **4** in CDCl$_3$ to a solution of cyclohexyl pentafluorophenyl ester **45** in CDCl$_3$ and 0.01 equivalent of Et$_3$N with or without 1.0 equiv. molecule **6**. Spectra were taken every hour until at least 10% of the product was formed. Initial velocities of the reactions were determined through integration of the C2 aromatic adenosine proton of the product **45** at 8.29 ppm relative to the C2 aromatic adenosine proton of the amine **4** at 8.35 ppm.

HPLC Kinetics

All reactions were performed in teflon capped autoinjector vials at 2.2 mM initial concentrations of reactants **4** and **5** in CHCl$_3$ with with 1.0% TEA base. A CHCl$_3$ stock solution of 5'-amino-5'-deoxy-2',3'-isopropylideneadenosine **4** was prepared daily. All other solutions were freshly prepared for each kinetic run. Formation of product **6** was followed by HPLC at 256 nm on a Waters 600E instrument equipped with a Waters 717 autosampler and a Waters 490E UV detector. Temperature inside the autosampler was constant at 22±1°C in an individually thermostated room. Product formation was followed for the first 5-10% of the reaction; initial rates were determined by linear fitting of the absolute size of the product peak areas. Kinetic runs showed excellent linearity and good convergence to the origin (representative individual data runs follow).

HPLC separation was achieved using a Beckman Ultrasphere ODS (C-18) column (4.6 mm ID x 25 cm length).

The elution system was: 30/70/0.1% H$_2$O/MeOH/TEA: 1 ml/min for 6 min, then at 2 ml/min for 4 min.

Followed by: MeOH with 0.1% TEA: 2.5 ml/min for 5 min

Followed by return to starting conditions: 30/70/0.1% H$_2$O/MeOH/TEA: 2 ml/min for 5 min.

Under these conditions, the retention times of amine **4**, product **6** and ester **5** were 3.3, 7.2, and 14.2 min, respectively.

Sample: ctest3   Channel: detector 1   Filename: ctest3
Acquired: 07-SEP-94 12:41   Method: C:\MAX\DATA1\EDMET   Operator: Ew
Comments: Ultrasphere ODS, H2O/MeOH/TEA 30/70/0.1, flush MeOH/TEA 100/0.1

x 10^{-1} volts

Sept 7, 1994   Naphthoyl System

1 mℓ/min   A

Amino Adenosine

6.0 min   A   2 mℓ/min

Template

10.0 min   B   2.5 mℓ/min

15.0 min   A   2 mℓ/min

PFP ester

x 10^{1} minutes

Solvent:
A - 30/70/0.1 H$_2$O/MeOH/TEA
B - 100/0.1 MeOH/TEA
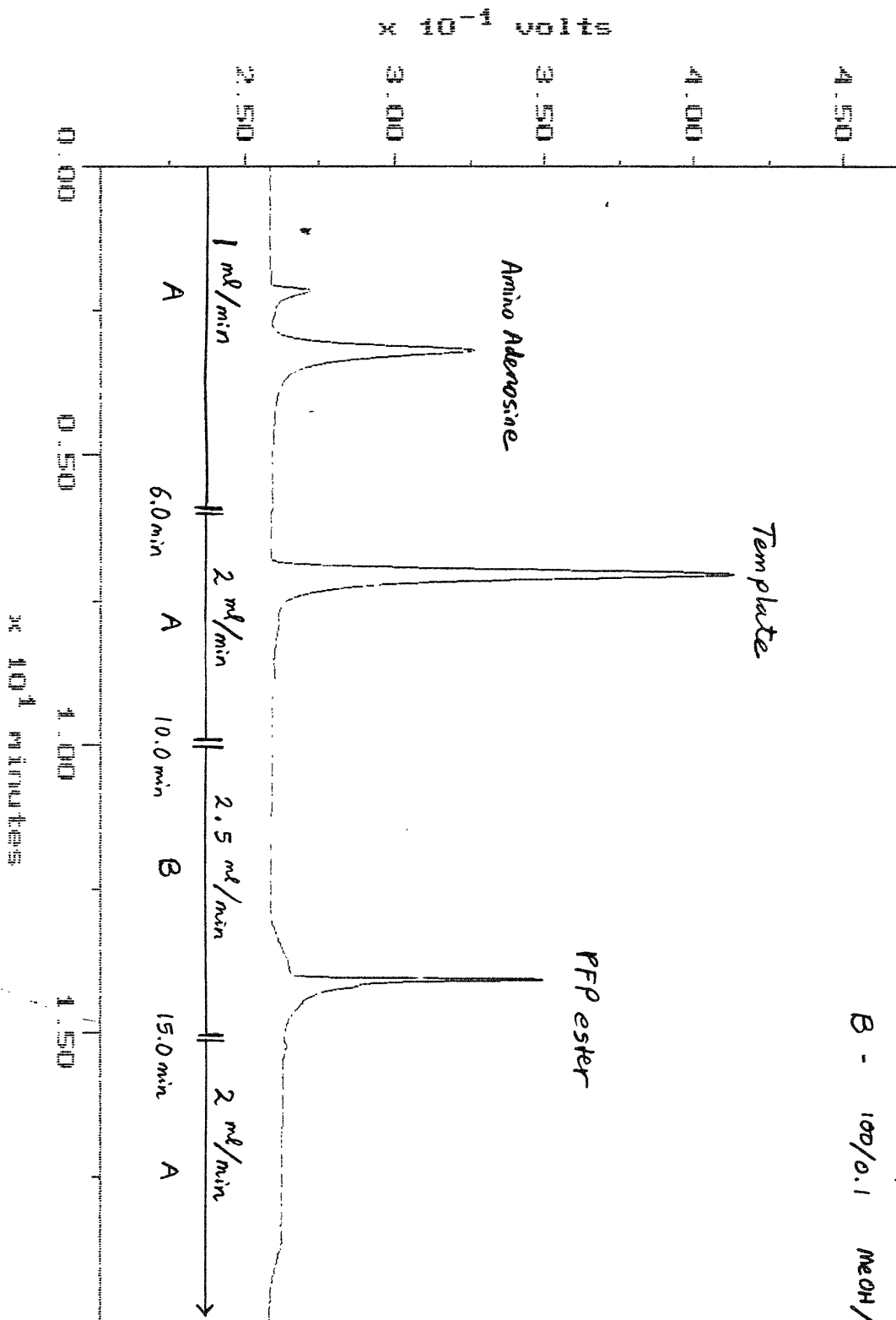
Formation of Naphthyl Template in CHCl3
2.2 mM Starting Conc. of Ester and Amine

Conc. (mM)

Time (min)
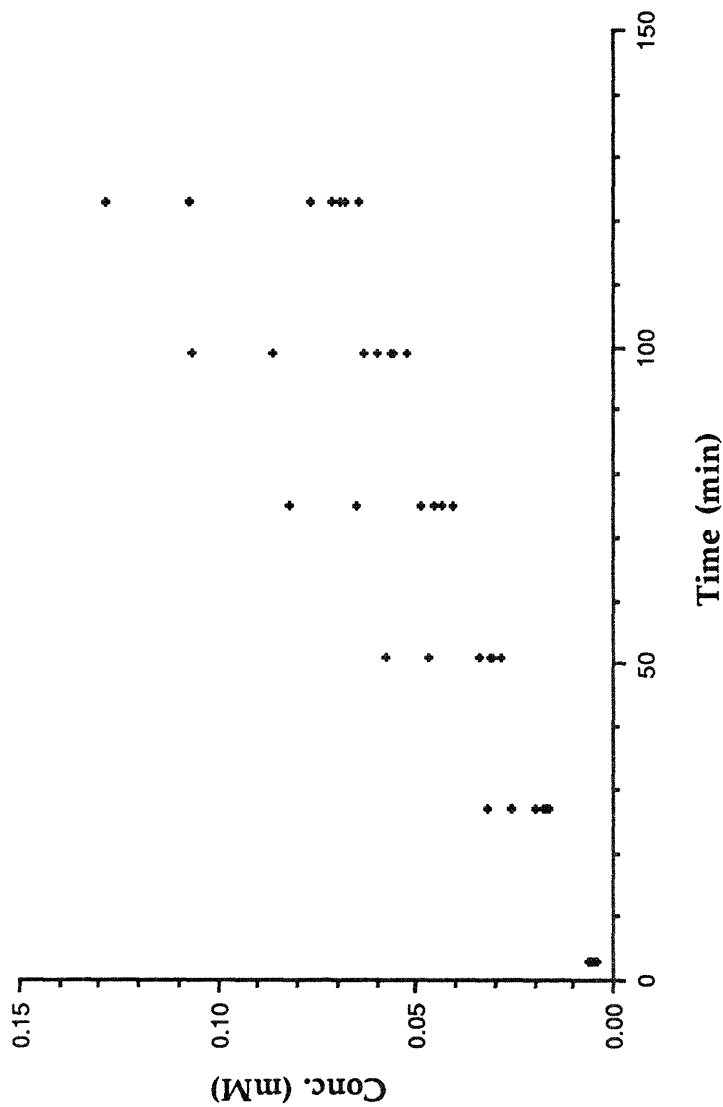
$y = 3.8550e\text{-}3 + 1.0293e\text{-}3x$	$R^2 = 0.999$	4 Eq. Valerolactam Added
$y = 3.4875e\text{-}3 + 8.3750e\text{-}4x$	$R^2 = 1.000$	50% Template Added
$y = 3.9050e\text{-}3 + 5.9357e\text{-}4x$	$R^2 = 1.000$	50% Phenyl Templ. Added
$y = 3.1200e\text{-}3 + 5.6000e\text{-}4x$	$R^2 = 0.999$	50% N-Me Phenyl Added
$y = 2.4100e\text{-}3 + 5.4429e\text{-}4x$	$R^2 = 1.000$	No Template Added
$y = 2.9575e\text{-}3 + 5.3321e\text{-}4x$	$R^2 = 1.000$	50% N-Me Templ. Added
$y = 2.9000e\text{-}3 + 5.0000e\text{-}4x$	$R^2 = 1.000$	50% Naphthoyl Aden. Added

Formation of Naphthyl Template in CHCl3
2.2 mM Starting Conc. of Ester and Amine

Conc. (mM)

Time (min)

▲ 4 Eq. Valerolactam Added	$y = 3.8550e\text{-}3 + 1.0293e\text{-}3x$	$R^2 = 0.999$
○ 50% Template Added	$y = 3.4875e\text{-}3 + 8.3750e\text{-}4x$	$R^2 = 1.000$
■ 50% Phenyl Templ. Added	$y = 3.9050e\text{-}3 + 5.9357e\text{-}4x$	$R^2 = 1.000$
✦ 50% N-Me Phenyl Added	$y = 3.1200e\text{-}3 + 5.6000e\text{-}4x$	$R^2 = 0.999$
◆ No Template Added	$y = 2.4100e\text{-}3 + 5.4429e\text{-}4x$	$R^2 = 1.000$
✦ 50% N-Me Templ. Added	$y = 2.9575e\text{-}3 + 5.3321e\text{-}4x$	$R^2 = 1.000$
▫ 50% Naphthoyl Aden. Added	$y = 2.9000e\text{-}3 + 5.0000e\text{-}4x$	$R^2 = 1.000$

EXPERIMENTAL --- PART II

General

Bovine pancreatic trypsin was purchased from Sigma. All protected amino acids were obtained from Advanced ChemTech and Novabiochem. Benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate (BOP) and PyBop (benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate) were purchased from Novabiochem. For preparative HPLC a Waters 600E system was used with a Waters 490I multiwavelength detector. Mass spectra for product characterization were obtained on a Finnigan MAT 8200 system. For the analysis of the diversity of the generated libraries by electrospray mass spectrometry, a VG Quatro instrument was used.

9,9-dimethyl-xanthene (54)

Compound **54** was obtained from fellow graduate Blake Hamann. His procedure from commercially available **53** is reproduced below by permission:

In a dry 1000 ml flask under Ar was suspended 45 g xanthone (1 eq.) in 300 ml toluene (dist., or fresh bottle). The suspension was cooled in an ice bath and 300 ml of a 2.0 M soln. of Me$_3$Al in toluene (2.5 eq.) were added dropwise by addition funnel. Do not use neat Me$_3$Al. After addition was complete, the reaction was stirred at 0°C for 2h, then allow to warm to RT stirring overnight. The solution was cannulated into 4 l of ice (white precip.), and the toluene layer was separated and extracted with 2x500 ml 3M HCl. The organic layer was dried over MgSO$_4$ and concentrated *in vacuo* to yield 9,9-dimethyl-xanthene as a yellow oil.

9,9-dimethyl-2,4,5,7-tetrabromoxanthene (55)

15.2 ml (47.36 g, 0.296 mol) Br$_2$ and 0.1 g Fe (as a catalyst) were placed in a 250 ml roundbottom flask. 100 ml CH$_2$Cl$_2$ were added and the mixture cooled with ice. 15.55 g (0.0739 mol) 9,9-dimethylxanthene **54** were added slowly, and after the addition of 50 further ml CH$_2$Cl$_2$ the reaction vessel was fitted with a water cooled condenser. The flask was stirred for 1 h. at 0 °C, for 2 h. at RT, and for 4 h. under reflux until the reaction was clear. The solution was washed with 200 ml of H$_2$O and the organic layer dried over MgSO$_4$. The solution was rotovapped to an orange-brown solid, which was refluxed in MeOH (200ml). The product was filtered off as a white solid and dried *in vacuo*. Yield 32.56 g, 83%.

^{1}H NMR (250 MHz,CDCl$_3$) δ 7.63 (d, J=2.0 Hz, 2 H), 7.44 (d, J=2.3 Hz, 2 H), 1.60 (s, 6 H). HRMS (EI) calcd for C$_{15}$H$_{10}$Br$_4$O, 521.74651; found, 521.7458.

9,9-dimethyl-2,4,5,7-tetracyanoxanthene (56)

25 g (0.048 mol) 9,9-dimethyl-2,4,5,7-tetrabromoxanthene **55** and 19.6 g (0.219 mol) CuCN were added to a 1 l three neck flask with condenser, mechanical stirrer, and thermometer. 150 ml N-methyl-pyrrolidinone was added, and the reaction refluxed at 210 °C for 4 h. The deep brown-green solution was poured over ice, then collected by vac. filtration and washed with H_2O to yield a grey, wet solid. This was placed in a 2L three neck flask with mechanical stirrer and stirred overnight in 500 ml H_2O. The flask was fitted with a condenser and stirred vigorously as 500 ml 20% nitric acid was added slowly by addition funnel (**Caution!**-NO_2 gas evolved-orange color). After 4 h. stirring, 200 ml conc. nitric acid was added slowly down the condenser. The solution turned green, and more gas was evolved. The solution was heated to reflux for 4 hrs, during which time a final 200 ml conc. nitric acid was added. The green solution was cooled in an icebath and filtered to give a beige solid. The product was taken on without purification.

Fellow Graduate Jerry Shipps later reported the following data for the same molecule:

^{1}H NMR (250 MHz, $DMSO_{d6}$) δ 8.56 (d, J= 1.9 Hz, 2 H), 8.48 (d, J= 1.8 Hz, 2 H), 1.67 (s, 6 H). ^{13}C NMR (62.9 MHz, DMSO) δ 151.60, 136.85, 136.54, 131.69, 116.02, 113.14, 108.28, 101.65, 34.27, 32.23.

HRMS (EI) calcd for $C_{19}H_{10}N_4O$, 310.0855; found, 310.0859.

9,9-dimethyl-2,4,5,7-xanthene tetracarboxylic acid (57)

The crude 9,9-dimethyl-2,4,5,7-tetracyanoxanthene **56** from above was suspended in 50 ml H_2O in a 1 l three neck flask with condenser and mechanical stirrer. 13 g (0.325 mol) NaOH were dissolved in 50 ml H_2O, and the solution was refluxed 36 h. The solution was allowed to cool, and was brought to pH 0 with conc. HCl. The aqueous solution was extracted with 3x1000 ml 1:1 THF/EtOAC, and the organic layers rotovapped to a tan powder. The solid was boiled in 4:1 EtOH/H_2O for 1 h., then allowed to cool and filtered. The off-white solid was dried *in vacuo* to give the pure product **57**, 11.8 g. Total yield of two steps from the tetra-bromo compound: 64%.

^{1}H NMR (250 MHz, $DMSO_{d6}$) δ 13.22 (br, 4 H), 8.26 (d, J = 1.8 Hz, 2 H), 8.16 (d, J = 1.8 Hz, 2 H), 1.67 (s, 6 H).

HRMS (FAB in 3-nitrobenzyl alcohol) calcd for $C_{19}H_{15}O_9$ (M + H), 387.0716; found, 387.0718.

9,9-dimethyl-2,4,5,7-xanthene-tetracarboxylic acid chloride (50)

5.0 g 9,9-dimethyl-2,4,5,7-xanthene-tetracarboxylic acid **57** (1 eq.) were suspended in 80 ml dry CH_2Cl_2. To the solution was added 4 ml oxalyl chloride (13 eq.) and 50 µl dimethyl formamide (cat.). Reflux was maintained until a clear brown solution was obtained, 4 h. The solution was concentrated and dried *in vacuo* overnight.

^{1}H NMR (250 MHz, CDCl$_3$): 8.70 (d, J= 2.1 Hz, 2 H), 8.40 (d, J= 2.2 Hz, 2 H), 1.7 (s, 6 H).

9,9-dimethyl-2,4,5,7-xanthene-tetracarboxylic acid tetrabenzyl ester (59)

To the tan solid **50** obtained above (1 eq.) was added 15 mg 4-dimethyl-amino-pyridine (cat.) and the reactants cooled in an icebath in a 500 ml flask with condenser. While stirring, 50 ml chilled CH_2Cl_2 was added, followed by 15 ml chilled benzyl alcohol (11 eq.), and finally 20 ml chilled pyridine. (Reaction may bubble violently for several minutes.) After addition of 20 ml further CH_2Cl_2, soln. was allowed to warm to RT and stir 24 hrs. The brown soln. was concentrated to a paste, sonicated with 200 ml MeOH, and filtered. Precipitate was washed with 2x100 ml MeOH, then re-filtered and dried *in vacuo* to yield the tetrabenzyl ester **59** as a white solid.

mp 142 - 143 °C (dec).

IR (KBr) 1729, 1709, 1440, 1376, 1326, 1304, 1245, 1119, 760, 733, 687 cm^{-1}

^{1}H NMR (300MHz, DMSO$_{d6}$): 8.33 (d, 2H, J=1), 8.19 (d, 2H, J=1), 7.32-7.49 (m, Ar, 20H), 5.38 (s, 4H), 5.30 (s, 4H), 1.69 (s, 6H)

9,9-dimethyl-2,4,5,7-xanthene-tetracarboxylic acid 2,7-dibenzyl ester (60)

In a three-neck flask with stirbar was dissolved 8.0 g 9,9-dimethyl-2,4,5,7-xanthene-tetraccarboxylic acid tetrabenzyl ester **59** in 250 ml CH_2Cl_2 and soln. chilled to 0 °C in an icebath. HBr was bubbled through the solution with stirring (keeping one neck open to vent HBr) until CH_2Cl_2 was saturated (~5 to 10 min). Flask was capped and stirred at 0 °C for 4 h. Depending on the amount of solvent used, product may precipitate during this time. Ar was bubbled through the solution 0.5 hr. to remove HBr, and solution was concentrated to a solid. The solid was dissolved in 30 ml THF, and the product precipitated by pouring into 100 ml hexanes and filtering. The product was washed with hexanes/ether and dried *in vacuo* to yield the dibenzyl diacid **60** as a white powder.

mp 239 - 241 °C (dec)

IR (KBr) 3111, 1718, 1612, 1437, 1257, 1123 cm^{-1}

^{1}H NMR (300 MHz, DMSO$_{d6}$): 13.2, (2H, broad), 8.30 (d, 2H, J=1), 8.18 (d, 2H, J=1), 7.3-7.5 (m, Ar, 10H), 5.38 (s, 4H), 1.68 (s, 6H)

Standard procedure for the preparation of acid chlorides.

0.5 mmol of the corresponding acid was dissolved in 20 ml of dichloromethane. 5 mmol oxalyl chloride were added in the case of 9,9-dimethyl-2,7-di-*tert*-butyl xanthene-4,5-dicarboxylic acid[69] and 9,9-dimethyl xanthene-2,4,5,7-tetracarboxylic acid 4,5-dibenzylester **60**, and 2.5 mmol oxalyl chloride were added for 9,9-dimethyl xanthene-2,4,5,7-tetracarboxylic acid **56**. After addition of two drops of dimethylformamide, the reaction mixture was heated to reflux for 4 h and concentrated *in vacuo* to yield **50** and **61** as tan solids and **58** as a yellow oil. The products were dried *in vacuo* overnight. The acid chlorides were used without further purification. For a given set of libraries, the total mmol of acid necessary for all libraries was reacted, and the resulting tetra acid chloride partitioned during synthesis (below) as a CH_2Cl_2 soln.

9,9-Dimethyl xanthene-2,4,5,7-tetracarboxylic acid chloride (50).

IR (KBr) 1776, 1740, 1585, 1230, 1270, 1163, 1128, 1013 cm^{-1}

^{1}H NMR (250 MHz, DMSO$_{d6}$) δ 8.25 (d, J = 1.9 Hz, 2 H, ar-H), 8.16 (d, J = 2.0 Hz, 2 H, ar-H), 1.67 (s, 6 H, CH$_3$);

MS (EI) *m/z* (%) = 458 (100, [M$^+$]), 423 (60), 407 (20), 381 (40), 318 (18).

9,9-Dimethyl-2,7-di-*tert*-butyl xanthene-4,5-dicarboxylic acid chloride (58).

IR (KBr) 3418, 2962, 1785, 1607, 1443, 1395, 1364, 1304, 1284, 1168 cm^{-1}

^{1}H NMR (250 MHz, DMSO$_{d6}$) δ 7.78 (d, J = 2.4 Hz, 2 H, ar-H), 7.72 (d, J = 2.4 Hz, 2 H, ar-H), 1.67 (s, 6 H, CH$_3$), 1.32 (s, 18 H, *tert*-butyl-H).

9,9-Dimethyl xanthene-2,7-dicarboxylic acid benzylester-4,5-dicarboxylic acid chloride (61).

IR (KBr) 2977, 2780, 2439, 1718, 1588, 1471, 1259, 1024 cm^{-1}

^{1}H NMR (250 MHz, DMSO$_{d6}$) δ 8.28 (d, J = 2.0 Hz, 2 H, ar-H), 8.17 (d, J = 2.0 Hz, 2 H, ar-H), 7.50 - 7.36 (m, 10 H, benzyl-H), 5.38 (s, 4 H, CH$_2$), 1.67 (s, 6 H, CH$_3$).

Standard procedure for the preparation of libraries.

0.22 mmol of the acid chloride core molecule was dissolved in 10 ml of dichloromethane. The amine mixture composed of an equimolar mixture of the desired amines (total molarity of amines = total molarity of acid chloride groups) was added and the reaction mixture was vigorously stirred. Triethylamine (0.5 ml) was added and the reaction mixture stirred for 2 h at RT The reaction mixture was diluted with 250 ml of dichloromethane and washed twice with 100 ml 1 M citric acid solution, and twice with water. The organic phase was separated, dried over $MgSO_4$ and concentrated *in vacuo* to afford a tan foam or an oil depending on library makeup.

Standard procedure for the deprotection of libraries.

100 mg of the library material was stirred with 6 ml reagent K (trifluoroacetic acid, water, phenol, thioanisol, ethanedithiol (82.5:5:5:5:2.5)[65] at RT for 4 h. The solution was concentrated *in vacuo* to yield a brown oil. The libraries were precipitated by addition of 15 ml of a mixture of diethylether/hexanes (1:1). The tan precipitate was filtered off and washed three times with ether/hexanes (1:1) and dried *in vacuo* overnight. If precipitation did not occur readily upon addition of ether/hexanes (1:1), the solution was concentrated again to an oil, dried *in vacuo*, taken up in 250 μl TFA, and sonicated as ether/hexanes (1:1) were added to the flask. Precipitates were then recovered as above.

Preparation of libraries L1 - L4 (Figure 37)

Description of the amine mixtures used:

No. of different tools	Tools A_n -NH2	No. of com- pounds
4	H-L-Trp-OMe, H-L-Val-OMe, H-L-Ala-OMe, H-L-Phe-OMe	136
7	H-L-Trp-OMe, H-L-Val-OMe, H-L-Ala-OMe, H-L-Phe-OMe, H-L-Met-OMe, H-L-Pro-OMe, H-L-Leu-OMe	1225
12	H-L-Ala-OMe, H-L-Phe-OMe, H-L-Met-OMe, H-L-Pro-OMe, H-L-Leu-OMe, H-L-Lys(Boc)-OMe, H-L-Ser(tBut)-OMe, H-L-His(Trt)-NHR, H-L-Asp(OtBut)-OMe, H-L-Glu(OtBut)-OMe, H-L-Thr(tBut)-OMe, Furfurylamine	10,968
21	H-L-Trp-OMe, H-L-Val-OMe, H-L-Ala-OMe, H-L-Phe-OMe, H-L-Met-OMe, H-L-Pro-OMe, H-L-Leu-OMe, H-L-Lys(Boc)-OMe, H-L-Ser(tBut)-OMe, H-L-His(Trt)-NHR4, H-L-Asp(OtBut)-OMe, H-L-Glu(OtBut)-OMe, H-L-Thr(tBut)-OMe, H-L-Ile-OtBut, H-L-Cys(Trt)-NHR2, H-L-Arg(Mtr)-NHR3, H-L-Tyr(tBut)-OMe, H-L-Val-NHR 4-Methoxybenzylamine, N-Methylpyrrol-2-ethylamine, Furfurylamine	99,141

HPLC chromatography for the libraries **L1-L4** was performed on a Waters 600I system with a Waters 490I UV detector and an Ultrasphere® SI column (4.6 mm i.d x 25 cm lengh, 80 Å, 5 μm) at a flow rate of 1.5 ml/min^{-1}. The gradient used was twofold: 0% to 100% B in 30 min to 100% C in 15 min (A: hexane; B: ethyl acetate; C: ethyl acetate/methanol 94:6).

Numerical Data for Figures 43 and 44

Round 1

Library	A (rel)	
Blank	100	
1 Xanthene	66	
		X

Round 2

Library	A (rel)	
Blank	100	
B1 -Gly,Ala,Val	79	X
B2 -Leu,Ile,Pro	85	X
B3 -Arg,Lys,His	95	X
B4 -Ser,Thr,Met	68	
B5 -Phe,Tyr,Trp	74	
B6 -Glu,Asp,Asn	54	

Round 2.5

Library	A (rel)	
Blank	100	
C1	31	X
C2	90	
C3	85	
C4	100	
C5	82	
C6	90	
C7	82	

Round 3

Library	A (rel)	
Blank	100	
D1 -Arg	28	
D2 -Lys	93	X
D3 -His	21	
D4 -Leu	46	X
D5 -Ile	50	X
D6 -Pro	80	X
D7 -Gly	32	
D8 -Ala	27	
D9 -Val	59	X

Round 4

Library	A (rel)	
Blank	100	
1 Lys, Pro, Val, Ile	20	X
2 Lys, Pro, Val, Leu	26	
3 Lys, Pro, Val, Gly	45	
4 Lys, Pro, Val	28	
5 Pro, Val, Ile	100	
6 Lys, Val, Ile	55	
7 Lys, Pro, Ile	100	

Round 5

Library	Positions 4 and 5	Positions 2 and 7	A (rel)	
Blank	--	--	100	
F1	Lys, Val	Ile, Pro	37	
F2	Lys, Ile	Val, Pro	11	X
F3	Lys, Pro	Ile, Val	100	
F4	Val, Ile	Lys, Pro	100	
F5	Val, Pro	Lys, Ile	100	
F6	Ile, Pro	Lys, Val	100	

Round 6

Library	A (rel)	
Blank	100	
G1 Tetra COOH	100	
G2 Isomer 64	3	Hit
G3 Isomer 65	20	

Table 1.[15] A(rel): Percent of trypsin activity in the presence of added libraries relative to a blank containing no added library material. Each value represents average trypsin activity for four measurements. Libraries were considered inhibitory if they reduced trypsin activity by > 10 %. The activity of the blank was set to 100%. An X denotes selection of a given library to influence the next round of screening.

Library solutions added in Rounds 1 - 4: 2.5 mg in 50 µl DMSO; Round 5: 0.5 mg in 50 µl DMSO; Round 6: 0.25 mg in 50 µl DMSO.

Round 1: % of trypsin activity with the initial library constructed from the core molecule **50** plus the 18 building blocks in Table 9.

Round 2: % of trypsin activity with the six sublibraries **B1 - B6** constructed from core molecule **50** and 15 of the 18 amino acid building blocks. **B1:** Gly, Ala, Val omitted. **B2:** Leu, Ile, Pro omitted. **B3:** Arg, Lys, His omitted. **B4:** Ser, Thr, Met omitted. **B5:** Phe, Tyr, Trp omitted. **B6:** Glu, Asp, Asn omitted.

Round 2.5: (data not graphed on bar charts in Figure 43) % of trypsin activity with the seven sublibraries **C1 - C7** constructed from the core molecule **50** and nine to twelve of the building blocks. **C1:** Arg, Lys, His, Leu, Ile, Pro, Gly, Ala, Val. **C2:** His, Leu, Ile, Pro, Ala, Val, Phe, Tyr, Trp, Ser, Thr, Met **C3:** Arg, Lys, His, Gly, Phe, Tyr, Trp, Glu, Asp, Asn **C4:** His, Leu, Ile, Pro, Ala, Val, Phe, Trp, Glu, Asp, Asn **C5:** Arg, Lys, His, Gly, Phe, Tyr, Trp, Ser, Thr, Met **C6:** His, Leu, Ile, Pro, Gly, Ala, Val, Phe, Trp, Glu, Asp, Asn **C7:** Arg, Lys, His, Pro, Gly, Trp, Glu, Asp, Asn.

Round 3: % of trypsin activity with the nine sublibraries **D1 - D9** constructed from the core molecule **50** and eight of the nine building blocks Arg, Lys, His, Leu, Ile, Pro, Gly, Ala, Val. **D1:** Arg omitted. **D2:** Lys omitted. **D3:** His omitted. **D4:** Leu omitted. **D5:** Ile omitted. **D6:** Pro omitted. **D7:** Gly omitted. **D8:** Ala omitted. **D9:** Val omitted.

Round 4: % of trypsin activity with the seven sublibraries **E1 - E7** constructed from the core molecule **50** and three or four of the building blocks Lys, Leu, Ile, Pro, Gly, Val. **E1:** Lys, Pro, Val, Ile **E2:** Lys, Pro, Val, Leu **E3:** Lys, Pro, Val, Gly **E4:** Lys, Pro, Val **E5:** Pro, Val, Ile **E6:** Lys, Val, Ile **E7:** Lys, Pro, Ile.

Round 5: % of trypsin activity with the six sublibraries **F1 - F6**. Sublibraries were variously substituted with Lys, Ile, Pro, and Val at xanthene positions 2, 4, 5, and 7.

Round 6: % trypsin activity with the two final inhibitors. **G1:** control run with 9,9-dimethyl-2,4,5,7-xanthene tetracarboxylic acid. **G2:** isomer **64**. **G3:** isomer **65**.

Screening procedure.

The procedure below was developed by Postdoctoral Fellow Thomas Carell.

All UV-measurements were performed on a Perkin-Elmer Lambda-2 spectrometer. 2.5 mg of trypsin (bovine pancreatic) was dissolved in 10 ml of an 0.005 M solution of HCL in water (trypsin stock solution). 50 mg of N-Benzoyl-D,L-arginine-p-nitroanilide were dissolved in 5 ml of dimethylsulfoxide (substrate stock solution). Library material was dissolved in 200 µl of dimethylsulfoxide (Amount of library material: see legend under Figures 6 and 7). 50 µl from this library stock solution was diluted with 500 µl of buffer (pH 8.2, 0.5 M Tris HCl, 0.04 M $CaCl_2$) and 10 µl of the trypsin stock solution. High buffer concentration was necessary to stabilize the pH in those experiments where 2.5 mg of library material was added; buffer concentration was constant for all screening and K_i determination experiments. The slightly cloudy mixtures were centrifuged for 5 min., and 400 µl of the clear supernatant was transformed into a 1.0 ml disposable UV-cuvette filled with 500 µl of buffer. (Only the supernatant was used to avoid UV absorption by cloudy mixtures.) The solutions were mixed and the UV-absorption at 405 nm was measured and noted (0-value). 50 µl of the substrate solution was added and the solutions mixed thoroughly. UV-absorption at 405 nm was measured at 10 min, 20 min, and 40 min, and the obtained absorption minus the 0-value was noted (activity value) and used for activity comparisons. Final activity percentage was obtained by comparison of activity value to the blank activity value measured without added library material (blank activity = 100%).

Computer Simulation:

The computer program, designed to simulate the relative inhibition of various xanthene libraries, followed the scheme below:

1. Based on empirical evidence, generate a "binding score" for all building blocks at a given xanthene position. For example, at xanthene position 5, give Ile +3, Leu or Val +2.5, any other hydrophobic building block +1, etc.

2. Based on the number of building blocks chosen, create a library of one each of all possible xanthene molecules containing 4 amino acids.

3. For each molecule in the library, sum the binding scores of the building blocks at positions 2, 4, 5, and 7.

4. Based on the total binding score, generate a score for the amount of inhibition which each molecule contributes. For example, inhibition score=$e^{(\text{binding score})}$.

5. Sum the inhibition scores of all molecules in the library to get the total reduction in trypsin activity.

6. Compare the reduction in trypsin activity to other libraries generated with the *same number of building blocks.*

7. If the comparison of library inhibition does not qualitatively match real life comparison of the same libraries, adjust the binding scores in step 1 and repeat.

8. Once the simulation can reproduce the results of actual libraries generated in rounds 1-4, use the program to simulate the results of other initial groupings of building blocks, or modify individual building block parameters to examine the possibility of missing equally or more potent building blocks with a given grouping scheme.

Standard procedure for the preparation of sublibraries F1 - F6.

First randomization step: 50 mg (0.0883 mmol) of **61** were added in six different 10 ml round bottom flasks together with 5 ml of dichloromethane and a mixture of two of the following four building blocks (0.0885 mmol each): H-Lys(Boc)-OMe (26.3 mg), H-Val-O*tert*-butyl (18.6 mg), H-Ile-O*tert*-butyl (19.8 mg) and H-Pro-O*tert*-butyl (18.4 mg). After the addition of 0.25 ml of triethylamine, the reaction vials were stirred for 1 h at RT, and each reaction mixture was diluted with 100 ml of dichloromethane and extracted twice with 100 ml of 1 M citric acid solution, once with 100 ml of saturated sodium hydrogen carbonate solution, and once with water. Each reaction mixture was dried with $MgSO_4$ and concentrated *in vacuo* to yield a tan foam. TLC analysis of the reaction mixtures showed the presence of three reaction products (silica gel, ethyl acetate/*n*-hexane 1:1). **Debenzylation:** Each of the above six reaction mixtures were dissolved in 4 ml of ethyl acetate and 1 ml of ethanol. After the addition of 10 mg 10% Pd/C the reaction mixtures were stirred in an H_2-atmosphere for 2 h at RT. Each reaction mixture was filtered through celite and concentrated *in vacuo* to yield a tan foam. **Second randomization step:** Each of the above six reaction mixtures were dissolved in 5 ml of dichloromethane. After the addition of 75 mg (0.17 mmol) benzotriazol-1-yloxytris-(dimethylamine)-phosphonium hexafluoro-phosphate (BOP), 0.25 ml of triethylamine, and 0.00885 mmol of the two missing building blocks [H-Lys(Boc)-OMe (26.3 mg), H-Val-O*tert*-butyl (18.6 mg), H-Ile-O*tert*-butyl (19.8 mg), or H-Pro-O*tert*-butyl (18.4 mg)], each reaction mixture was stirred at RT for 2 h. Each mixture was diluted with 100 ml of dichloromethane and extracted with 1 M citric acid solution and sodium hydrogen carbonate solution. After drying with $MgSO_4$ each reaction mixture was concentrated *in vacuo* to yield a tan oil. **Deprotection:** Each of the above six reaction products were stirred together with 3 ml of trifluoroacetic acid, 2 ml of dichloromethane, and five drops of thioanisol for 4 h at RT. The solvent was removed *in vacuo* and each library precipitated by addition of 25 ml of ice-cold diethylether. Each library was washed several times with cold ether and the solid dried *in vacuo*.

Molecules from Scheme 6:

The Molecules **62a-65** were synthesized by Postdoctoral Fellow Thomas Carell.
NMRs were taken with the help of Postdoctoral Fellow Belinda Tsao. I ran preparative HPLC on the compounds as noted.

9,9-Dimethyl xanthene-2,7-dicarboxylic acid benzylester-4-carboxylic acid (O-*tert*-butylisoleucine)amide-5-carboxylic acid-(O-methyl-N-benzyllysine)amide (62a).

266 mg (0.44 mmol) **5** were dissolved in 20 ml of dichloromethane. After addition of 131 mg (4.4 mmol) of O-methyl-N^ε-benzyl-lysine hydrochloride, 98 mg (4.4 mmol) O-*tert*-butyl-isoleucine hydrochloride and 1 ml of triethylamine, the reaction mixture was stirred for 1.5 h at RT. The solution was diluted with 200 ml of dichloromethane and washed twice with 1 M citric acid solution and once with 100 ml of water. After separation of the organic phase and drying with $MgSO_4$ the reaction mixture was concentrated *in vacuo* to give a white foam. The product was separated from the product mixture by flash chromatography on silica gel with ethyl acetate/n-hexane (1:1) as the second fraction (R_F = 0.66, TLC: silica gel; ethyl acetate/n-hexane 1:1). Yield: 135 mg (31 %) of **6**.

mp 90 - 100 °C (dec)

IR (KBr) 3340, 2974, 1736, 1654, 1526, 1395, 1368, 1253, 1155 cm^{-1}

^{1}H NMR (300 MHz, $DMSO_{d6}$) δ 8.83 (m, 2 H, NH), 8.30 (m, 3 H, ar-H), 8.08 (d, J = 1.8 Hz, 1H, ar-H), 7.55 - 7.30 (m, 10 H, benzyl-H), 6.70 (m, 1H, NH), 5.40 (s, 2H, CH_2), 5.38 (s, 2H, CH_2), 4.42 (m, 1 H, α-CH_{Lys}), 4.23 (m, 1 H, α-CH_{ILe}), 3.61 (s, 3 H, OCH_3), 2.88 (m, 2 H, CH_2), 1.88 - 1.80 (m, 3 H, Ile-H), 1.71 (s, 3 H, CH_3), 1.69 (s, 3 H, CH_3), 1.53 - 1.15 (m, 24 H, CH_2, *tert*-butyl), 1.00 - 0.98 (m, 6 H, CH_3)

HRMS (FAB in glycerol, DMSO) calcd for $C_{55}H_{67}N_3O_{13}$ (M + H), 978.47522; found, 978.47312.

9,9-Dimethyl xanthene-2,7-dicarboxylic acid-4-carboxylic acid (O-*tert*-butylisoleucine)amide-5-carboxylic acid (O-methyl-N-benzyllysine)amide (62b)

130 mg (1.3 mmol) of **6** was dissolved in 5 ml of ethyl acetate and 1 ml of ethanol. After the addition of 20 mg Pd/C (10%) the reaction suspension was stirred in an H_2-atmosphere for 2 h at RT. The catalyst was removed by filtration and the clear product solution was concentrated *in vacuo* to yield a white foam which was dried *in vacuo* overnight. Yield: 96 mg (98 %) of **7**.

mp 140 - 150 °C (dec)

IR (KBr) 3335, 2974, 1718, 1654, 1522, 1395, 1368, 1253, 1164 cm^{-1}

^{1}H NMR (300 MHz, CDCl$_3$) δ 8.96 (d, J = 7.5 Hz, 1 H, NH), 8.88 (d, J = 6.6 Hz, 1 H, NH), 8.30 (d, J = 2.6 Hz, 1 H, ar-H), 8.23 - 8.20 (m, 2 H, ar-H), 8.06 (d, J = 2.4 Hz, 1 H, ar-H), 6.76 - 6.72 (m, 1 H, NH), 4.48 - 4.41 (m, 1 H, α-CH), 4.24 (m, 1 H, α-CH), 3.62 (s, 3 H, OCH$_3$), 2.88 (m, 2 H, CH$_2$), 1.94 (m, 2 H, CH$_2$), 1.72 (s, 3 H, CH$_3$), 1.67 (s, 3 H, CH$_3$), 1.55 - 1.23 (m, 6 H, CH$_2$), 1.42 (s, 9 H, tert-butyl-H), 1.30 (s, 9 H, tert-butyl-H), 0.98 - 0.94 (m, 6 H, CH$_3$)

HRMS (FAB in glycerol, DMSO) calcd for C$_{41}$H$_{55}$N$_3$O$_{13}$ (M + H), 798.38132; found, 798.37952.

9,9-Dimethyl xanthene-4-carboxylic acid (O-*tert*-butylisoleucine)amide-2-carboxylic acid(O-*tert*-butylproline)amide-7-carboxylic acid (O-*tert*-butylvaline)amide-5-carboxylic acid (O-methyl-N-benzyllysine)amide (63a).

9,9-Dimethyl xanthene-4-carboxylic acid (O-*tert*-butylisoleucine)amide-7-carboxylic acid(O-*tert*-butylproline)amide-2-carboxylic acid (O-*tert*-butylvaline)amide-5-carboxylic acid (O-methyl-N-benzyllysine)amide (63b).

355 mg (0.48 mmol) of the diacid 7 and 430 mg (0.97 mmol) of Benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate (BOP) were dissolved in 25 ml of dimethylformaide. After the addition of 100 mg (0.48 mmol) of O-*tert*-butyl valine hydrochloride, 100 mg (0.48 mmol) of O-*tert*-butyl proline hydrochloride and 0.5 ml of triethylamine the reaction mixture was stirred at RT for 1 h. The solution was diluted with 200 ml of dichloromethane and washed twice with 100 ml of 1 M citric acid solution and three times with 100 ml of water. The organic phase was separated, dried with MgSO$_4$ and concentrated *in vacuo* to yield a clear oil. The two products **63a** and **63b** were separated from the product mixture by flash chromatography on silica gel with a n-hexane/ethylacetate gradient. After the elution of the first fraction with ethylacetate/n-hexane (1:1) the solvent was changed to ethylacetate/n-hexane (1.5:1) to elute **63a** as the second fraction (R$_F$ = 0.39). With ethylacetate/n-hexane (3:1) **63b** was obtained as the third fraction (R$_F$ = 0.26). Compounds were further purified by preparative HPLC: prep-Nova-Pak HR Silica (60 Å, 6 μm) column (19 x 300 mm), Solvent ethyl acetate/n-hexane (9:1), R$_{t,63a}$ = 11.5 min., R$_{t,63b}$ = 15 min.. TLC: silica gel, ethyl acetate/n-hexane (1:1).

Yield 158 mg (30 %) of **63a**

mp 90 - 100 °C (dec)

IR (KBr) 3346, 2974, 1738, 1652, 1520, 1394, 1368, 1252, 1153 cm^{-1}

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.98 (d, J = 7.0 Hz, 1 H, NH), 8.50 (d, J = 7.0 Hz, 1 H, NH), 8.50 (d, J = 2.0 Hz, 1 H, ar-H), 8.20 (d, J = 2.0 Hz, 1 H, ar-H), 7.79 (s, 1 H, ar-H), 7.67 (s, 1 H, ar-H), 6.84 (d, J = 7.5 Hz, 1 H, NH), 6.73 (br. d, J = 7.5 Hz, 1 H, NH), 4.82 - 4.79 (m, 1 H, α-CH$_{Lys}$), 4.62 (dd, J = 8.5 Hz, 5.0 Hz, 1 H, α-CH$_{Val}$), 4.59 - 4.50 (m, 2 H, α-CH$_{Pro}$ + α-CH$_{Ile}$), 3.72 (s, 3 H, OCH$_3$), 3.65 (m, 1 H), 3.48 (m, 1 H), 3.10 (m, 2 H, CH$_2$), 2.35 - 2.18 (m, 3 H), 2.16 - 1.85 (m, 5 H), 1.80 - 1.20 (m, 48 H), 0.98 (m, 12 H, CH$_3$)

HRMS (FAB in glycerol, DMSO) calcd for C$_{59}$H$_{87}$N$_5$O$_{15}$ (M + H), 1106.62770; found, 1106.62505.

Yield: 148 mg (28 %) of **63b**

mp 100 - 110 °C (dec)

IR (KBr) 3346, 2975, 1739, 1651, 1520, 1393, 1366, 1252, 1153 cm^{-1}

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.63 (d, J = 7.5 Hz, 1 H, NH), 8.29 (d, J = 1.5 Hz, 1 H, ar-H), 8.00 (d, J = 2.0 Hz, 1 H, ar-H), 7.93 (d, J = 2.0 Hz, 1 H, ar-H), 7.86 (d, J = 1.5 Hz, 1 H, ar-H), 7.08 (d, J = 7.0 Hz, 1 H, NH), 6.68 (d, J = 8.5 Hz, 1 H, NH), 4.86 - 4.78 (m, 1 H, α-CH$_{Lys}$), 4.64 (dd, J = 8.5 Hz, 5.0 Hz, 1 H, α-CH$_{Val}$), 4.56 (dd, J = 7.8 Hz, 5.0 Hz, 1 H, α-CH$_{Ile}$), 4.50 (dd, J = 8.5 Hz, 5.0 Hz, 1 H, α-CH$_{Pro}$), 3.80 - 3.55 (m, 2 H), 3.65 (s, 3 H, OCH$_3$), 3.17 - 3.00 (m, 2 H), 2.35 - 2.17 (m, 3 H), 2.16 - 1.80 (m, 5 H), 1.68 (s, 6 H, CH$_3$), 1.60 - 1.40 (m, 43 H), 1.00 (m, 12 H, CH$_3$)

HRMS (FAB in glycerol, DMSO) calcd for C$_{59}$H$_{87}$N$_5$O$_{13}$ (M + H), 1106.62770; found, 1106.62855.

9,9-Dimethyl xanthene-4-carboxylic acid-N-isoleucineamide-2-carboxylic acid-N-prolineamide-7-carboxylic acid-N-valineamide-5-carboxylic acid-N-lysineamide (64)

64 mg (0.059 mmol) of the protected precursor **63a** were dissolved in 2 ml of dichloromethane and were stirred with 2.5 ml trifluoroacetic acid and two drops of thioanisol for 4 h at RT. The reaction mixture was concentrated *in vacuo* to yield a brown oil. The product was precipitated upon addition of 50 ml of ice cold diethylether and subsequent sonification of the partly solid product for 5 min. The product was filtered off, washed two times with ice cold diethylether and dried *in vacuo* over night. **64** was further purified by preparative HPLC [prep-NovaPak HR C18 (60 Å, 6 μm) column (19 x 200 mm), solvent H$_2$O/acetonitrile/acetic acid (53:47:0.1), R$_t$ = 6.5 min.]. Yield: 45 mg (91 %) of **64**

mp 180 - 200 °C (dec)

IR (KBr) 3323, 2966, 1734, 1640, 1528, 1423, 1257, 1200. 1139 cm^{-1}

^{1}H NMR (90 °C, 500 MHz, DMSO-d$_6$) δ 8.88 (br.s, 1 H, NH), 8.41 - 8.25 (m, 3 H, ar-H, 2 NH), 8.21 (d, J = 2.5 Hz, 1 H, ar-H), 7.81 (m, 1 H, ar-H), 7.63 (m, 1 H, ar-H), 4.43 (m, 4 H, α-CH), 3.65 (s, 3 H, OCH$_3$), 3.65 - 3.55 (m, 2 H, CH$_2$), 2.80 (m, 2 H, CH$_2$), 2.50 (m, 2 H), 1.97 (m, 6 H), 1.73 (s, 3 H, CH$_3$), 1.70 (s, 3 H, CH$_3$), 1.60 (m, 3H), 1.48 (m, 2 H), 1.60 (m, 1 H), 1.05 - 0.85 (m, 12 H, CH$_3$); HRMS (FAB in glycerol, DMSO) calcd for C$_{42}$H$_{55}$N$_5$O$_{13}$ (M + H), 838.38746; found, 838.39772.

9,9-Dimethyl xanthene-4-carboxylic acid-N-isoleucineamide-7-carboxylic acid-N-prolineamide-2-carboxylic acid-N-valineamide-5-carboxylic acid lysineamide (65)

65 was obtained following the procedure for the preparation of **64**. 90 mg (0.83 mmol) of **63b** were converted into 67 mg (97 %) of **65**. **65** was purified by preparative HPLC [prep-NovaPak HR C18 (60 Å, 6 μm) column (19 x 200 mm), solvent H$_2$O/acetonitrile/acetic acid (53:47:0.1), R$_t$ = 6.5 min.]

mp 130 - 160 °C (dec)

IR (KBr) 3296, 2966, 1732, 1644, 1538, 1422, 1256, 1201, 1137 cm^{-1}

^{1}H NMR (90 °C, 500 MHz, DMSO-d$_6$) δ 9.00 (br.s, 1 H, NH), 8.38 (br.s, 1 H, NH), 8.28 (br.s, 1 H, NH), 8.20 (d, J = 2.0 Hz, 1 H, ar-H), 8.04 (s, 1 H, ar-H), 7.99 (s, 1 H, ar-H), 7.83 (s, 1 H, ar-H), 4.42 (m, 4 H, α-CH), 3.72 (s, 3 H, OCH$_3$), 3.58 (m, 2 H, CH$_2$), 2.82 (m, 2 H), 2.25 (m, 2 H), 2.10 - 1.85 (m, 6 H), 1.72 (s, 6 H, CH$_3$), 1.60 (m, 3 H), 1.48 (m, 2 H), 1.32 (m, 1 H), 1.05 (m, 9 H), 0.95 (t, J = 7.0 Hz, 3 H, CH$_3$)
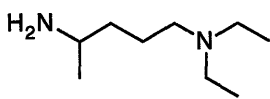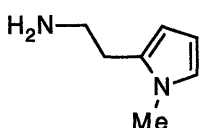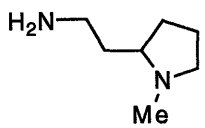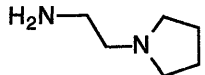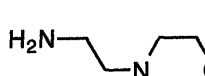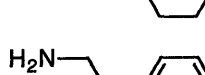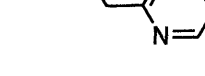
HRMS (FAB in glycerol, DMSO) calcd for C$_{42}$H$_{55}$N$_5$O$_{13}$ (M + H), 838.38746; found, 838.38696.

Experimental for DNA Polymerase I Screening

Building Blocks used to create libraries X1-X5 and B1-B5:

Non-Amino Acid Building Blocks

1 H_2N — ... — N(diethyl)

2 H_2N — (N-methylpyrrole)

3 H_2N — (N-methylpyrrolidine, Me)

4 H_2N — (pyrrolidine, N)

5 H_2N — (morpholine, N–O)

6 H_2N — (pyridine, N)

7 H_2N — (benzimidazole, N–H, N)

8 H_2N — N( )N–Me

9 H_2N — N( )O

10 H_2N — (tetrahydrofuran, O)

11 H_2N — (benzene)–OMe

Amino Acid Building Blocks

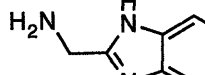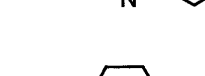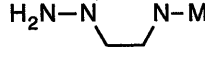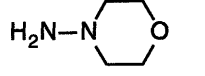|	AA	protected reagent used
12	(Ala)	L-alanine-t-butyl ester,
13	(Arg)	N^g-4-methoxy-2,3,6-trimethylbenzene -sulfonyl-L-arginine,
14	(Asn)	L-asparagine-t-butyl ester,
15	(Asp)	L-aspartic acid-β-t-butyl-α-t-butyl ester,
16	(Glu)	L-glutamic acid-γ-tert-butyl-α-t-butyl ester,
17	(His)	N^{im}-trityl-L-histidine,
18	(Ile)	L-isoleucine-t-butyl ester,
19	(Leu)	L-leucine-t-butyl ester,
20	(Lys)-OMe	N^{ε}-Boc-L-lysine-methyl ester,
21	(Phe)	L-phenylalanine-t-butyl ester,
22	(Pro)	L-proline-t-butyl ester,
23	(Ser)	O-t-butyl-L-serine-t-butyl ester,
24	(Tyr)-OMe	O-t-butyl-L-tyrosine-methyl ester,
25	(Val)	L-valine-t-butyl ester
26	(Gly)-OMe	Glycine-methyl ester,
27	(Met)-OMe	L-methionine-methyl ester,
28	(Thr)	O-t-butyl-L-threonine carboxylic acid
29	D-(Lys)	N^{ε}-Boc-D-lysine
30	D-(Trp)-OMe	D-tryptophane-methyl ester,
31	D-(Phe)	D-phenylalanine-t-butyl ester,
32	D-(Leu)	D-leucine-t-butyl ester,
33	D-(Pro)	D-proline-t-butyl ester,
34	D-(Ser)	O-t-butyl-D-serine-t-butyl ester,

The above building blocks were grouped in sets of 14 as follows:

X1, B1	12-25
X2, B2	7, 13, 14, 16, 17, 18, 24, 25, 26, 29, 31-34
X3, B3	1, 3, 5, 7, 10, 11, 12, 13, 15, 16, 17, 19, 20, 23
X4, B4	2, 4, 6, 8, 9, 12, 13, 15, 16, 17, 19, 20, 23, 25
X5, B5	1-11, 26, 27, 30

Polymerase Assay

Procedure for the polymerase assay is reproduced below courtesy of Deborah Kreutzer:

Escherichia coli DNA polymerase I Klenow Fragment (5 unit/µl, specific activity 20,000 units/mg) was from New England Biolabs and γ-^{32}P dATP (6000 Ci/mmol) was from New England Nuclear. Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer (Model 381) using Applied Biosystems reagents. The sequence of oligonucleotide primers and templates are shown below.

```
17-mer:   5'-GCTATGACCATGATTCA-3'
45-mer:   3'-CGATACTGGTACTAAGTCACCGTCGTTAAGTGACCGGCAGCAAAA-5'

14-mer:   5'-GCGATTCCCAGCTGAG-3'
42-mer:   3'-CGCTAAGTCGACTCATGATAGCCAGCTACGAACTGGACTCAG-5'
```

The primed DNA template was constructed by annealing 20 pmol of 5'-^{32}P labeled primer [ref 89] with 20 pmol of template in a buffer containing 20 mM Tris (pH 8.0), 5 mM MgCl$_2$ and 0.1 mM EDTA in a final volume of 400 µl. The primer-template solution was incubated at 100 °C for 5 min., then cooled slowly to RT for at least 2 hr.

The libraries were dissolved in DMSO (21.6 µl/mg) and then in the reaction buffer (108 µl/mg) which consisted of 10 mM Tris (pH 8.0), 5 mM $MgCl_2$ and 7.5 mM DTT. The pH of the libraries was adjusted to 7.8 by using 1 M KOH and 1 M HCl and the libraries were filtered through a 0.45 µm membrane.

Primer extension by the Klenow fragment was carried out in the reaction buffer with a maximum of 10% (v/v) DMSO. (At this concentration of DMSO, the activity of the polymerase was unaffected.) 0.25 units of polymerase were mixed with the library in a total volume of 18 µl, and the reaction was incubated at 0 °C for 10 min. To the polymerase-library mixture was added 0.25 pmol labeled primer-template and 40 µM of each of the four deoxynucleotides in a final volume of 25 µl. The primer extension reaction was incubated at RT for 5 min., and the reaction was stopped by the addition of EDTA to a final concentration of 85 mM. The reaction products were analyzed on a 10% denaturing polyacrylamide gel (run at 25 mAmp for 45 min.). The gel was exposed to a PhosphorImager plate and the plate analyzed on a Molecular Dynamics PhosphorImager (Model 400S). The extent of polymerase inhibition was determined by measuring the amount of radiolabeled primer (17-/14-mer) that was fully extended (45-/42-mer).

Biphenyl tetra acid chloride (66):

The synthesis of the biphenyl tetra acid **70** is a modification of a scheme from coworker Robert Grotzfeld.

4-bromo-isophthalic acid (68)

25g 4-bromo-m-xylene "tech" **67** (0.135 mol) and 105 g $KMnO_4$ (.67 mol) were refluxed in 1000 ml H2O 6 h. The purple solution was filtered (hot) and acidified (hot) to pH 0 with conc. HCl. The suspension was filtered (warm) to yield a white solid. Upon cooling, further precipitate formed in the filtrate, and this was filtered to yield a second crop of precipitate. Total isolated 11.8g, 36%.

^{1}H NMR (300 MHz, DMSO-d$_6$) δ 13.55 (br, s, 2H), 8.23 (s, 1H), 7.93 (d, J=8.7, 1H), 7.86 (d, J=8.7, 1H).

Calcd for $C_8H_5BrO_4$: 243.9372

4-bromo-isophthalic acid dimethyl ester (69)

11.5 g 4-bromo-isophthalic acid **68** (0.047 mol) were warmed in 400 ml MeOH (50 °C). HCl gas was bubbled through the solution 20 min, and the solution was stirred at 50°C 4h. Ar was bubbled through the soln. to remove HCl and the soln was concentrated to a white solid. The solid was taken up in 400ml ether/hexane 1:1 and washed with 0.5N NaOH and H$_2$O, 100 ml each. The organic layer was concentrated to a white solid and dried *in vacuo*. Yield 12.66 g, 98%.

^{1}H NMR (250 MHz, CDCl$_3$) δ 8.42 (d, J=2.2 Hz, 1H), 7.94 (dd, J=2.2, 8.5 Hz, 1H), 7.73 (d, J=8.5, 1H), 3.94 (s, 3H), 3.92 (s, 3H).

Calcd. for $C_{10}H_9BrO_4$: 271.9685

2, 2′, 4, 4′-biphenyl tetra acid tetramethyl ester (70)

7.0 g 4-bromo-isophthalic acid dimethyl ester **69** (.026 mol) and 4.8 g fine mesh Cu powder were mixed in a 20ml teflon-capped sealed tube with stirbar. The tube was heated to 195 °C for 1.5 hr, then to 235 °C for 2 hr. The cooled Cu paste was sonicated with acetone (1 L) and filtered. The brown filtrate was concentrated to a solid and chromatographed on silica gel (1-5% MeOH/CHCl$_3$, TLC R$_f$ 0.3 in 40% EtOAC/Hex). Product was recrystalized from hot CHCl$_3$ by pouring into 20/80 EtOAc/Hex, 1.7g, 34%.

^{1}H NMR (250 MHz, CDCl$_3$) δ 8.69 (d, J=1.8 Hz, 2H), 8.19 (dd, J=1.8, 7.8 Hz, 2H), 7.25 (d, J=7.8, 2H), 3.95 (s, 6H), 3.65 (s, 6H).

Calcd for $C_{20}H_{18}O_8$: 386.1002

2,2′,4,4′-biphenyl tetra acid (71)

2,2′,4,4′-biphenyl tetra acid tetramethyl ester 70 (950mg, 2.46 mmol) were disolved in 150 ml THF. To this was added 50 ml EtOH and 20 ml 4N NaOH. Clear soln. was stirred at RT 18h, then concentrated to remove organics. Aqueous layer was stirred a further 2 h at 40 °C, and a small amount of yellow precip was filtered off. Solution was adjusted to pH 8 (still clear) with conc. HCl and then poured into 100 ml conc HCL/H_2O 1:1 in a frit filter. White precipitate was collected, air dried with suction 2 h, and washed with ether, 735mg dry, 90%.

^{1}H NMR (300 MHz, DMSO-d_6) δ 13.1 (br, s, 4H), 8.47 (d, J=1.8 Hz, 2H), 8.10 (dd, J=1.8, 7.8 Hz, 2H), 7.25 (d, J=7.8, 2H).

Calcd for $C_{16}H_{11}O_8$ [M+H]: 331.0454 HRMS (FAB, 3-NBA) [M+H] 331.0453

2, 2′, 4, 4′-biphenyl tetra acid chloride (66)

400 mg 2,2′,4,4′-biphenyl tetra acid 71 (1.21 mmol) were pulverized with 2.5 g PCl_5 (12 mmol) in a mortar and pestle (use hood). The mixture was placed in a 10 ml flask and heated to a melt under H_2O condenser (150 °C) for 15 min. To the clear liquid was added 5 ml benzene and the soln. refluxed 1 hr. The solution was concentrated under vacuum at 50 °C to a white solid and dried *in vacuo* 48 h. The product was used without further purification.

^{1}H NMR (250 MHz, CDCl$_3$) δ 9.03 (d, J=1.8 Hz, 2H), 8.41 (dd, J=1.8, 8.3 Hz, 2H), 7.25 (d, J=8.3, 2H)

Calcd for $C_{16}H_{11}Cl_4O_4$: 406.9411.

Compounds from Scheme 8:

In the following procedures, 2-(aminomethyl)-benzimidazole is introduced as the di-hydrochloride hydrate, FW 220.10 (Aldrich), and the L-isomer of leucine-t-butyl ester is introduced as the hydrochloride, FW 223.8 (Novabiochem).

78, 79

2-(aminomethyl)-benzimidazole (123 mg, 0.56 mmol) and leucine-t-butyl ester (125 mg, 0.56 mmol) were stirred in 15 ml CH_2Cl_2 and 4 ml triethylamine. The solution was chilled and then added to a cold solution of 9,9-dimethyl xanthene-2,7-dicarboxylic acid benzylester-4,5-dicarboxylic acid chloride 61 (0.56 mmol) in 15 ml CH_2Cl_2. The reaction was stirred 2h, warming to RT. The solution was concentrated to a solid, which was taken up in 100ml CH_2Cl_2 and washed with 30 ml H_2O. The crude material was chromatographed with 1:39:60 MeOH/Hex/EtOAc followed by 5:20:75 MeOH/Hex/EtOAc to give the first two products by TLC (R_f = 0.95, 0.82 in 5:20:75 MeOH/Hex/EtOAc). Both fractions were concentrated to oils and stirred individually with 30 mg Pd/C 10% in 8ml EtOAc and 3 ml EtOH under H_2 atmosphere 4 h. The solutions were filtered with EtOAc wash and the filtrates rotovapped to white powders: diacid 78, 105 mg. diacid 79, 90 mg.

78

^{1}H NMR (250 MHz, DMSO-d_6) δ 13.24 (br s, 2 H), 8.96 (d, J = 7.0 Hz, 2 H), 8.26 (s, 4 H), 4.38 (m, 2 H), 1.77 (m, 12 H), 1.40 (s, 18 H), 0.93 (m, 12 H).
Calcd for [M+H]$^+$ $C_{39}H_{53}N_2O_{11}$: 725.3649, HRMS FAB(3NBA) [M+H]$^+$ found 725.3651

79

^{1}H NMR (250 MHz, DMSO-d_6) δ 13.66 (br s, 2 H), 12.22 (br s, 1 H), 9.51 (br t, 1 H), 9.15 (d, J = 7.2 Hz, 1 H), 8.50 (s, 1 H), 8.30 (s, 2 H), 8.19 (s, 1 H), 7.45 (m, 2 H), 7.13 (m, 2 H), 4.75 (m, 2 H), 4.37 (m, 1 H), 1.63 (m, 9 H), 1.38 (s, 9 H), 0.75 (m, 6 H).
Calcd for $C_{37}H_{40}N_4O_9$: 684.2795

80

2(aminomethyl)-benzimidazole (100 mg, 0.45 mmol) was stirred in 10 ml CH_2Cl_2 and 1 ml triethylamine. The solution was chilled and then added to a cold solution of 9,9-dimethyl xanthene-2,7-dicarboxylic acid benzylester-4,5-dicarboxylic acid chloride **61** (0.18 mmol) in 15 ml CH_2Cl_2. The reaction was stirred 2h, warming to RT. The solution was concentrated to a solid, which was taken up in 100ml CH_2Cl_2 and washed with 2x40 ml H_2O. The crude material was chromatographed with 2-5% $MeOH/CHCl_3$ to give 70 mg of a white solid. The solid was stirred with 40 mg Pd/C 10% in 20ml EtOAc, 8 ml EtOH, and 1 ml triethylamine (to keep the acid in solution) under H_2 atmosphere 4 h. The solution was filtered with EtOAc wash and the filtrate concentrated with hexanes to a white powder: diacid **80**, 79 mg as the triethylamine salt.

^{1}H NMR (300 MHz, DMSO-d_6) δ 9.61 (br t, 2 H), 8.40 (s, 2H) 8.29 (s, 2H), 7.47 (m, 4 H), 7.13 (m, 4 H), 4.64 (m, 4 H), 1.71 (s, 6 H).

Calcd for $C_{35}H_{28}N_6O_7$: 644.2019

76

The TEA salt of diacid **80** (79 mg, 0.1mmol) was stirred with 210 mg PyBop (benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate, 0.4 mmol) for 5 min in 4 ml DMF. To this was added 90 mg leucine-t-butyl ester (0.4 mmol) and 250 ml diisopropylethylamine and the solution stirred 3 h. The solution was diluted with 200 ml CH_2Cl_2 and washed with 4x100ml H_2O. The organics were dried over $MgSO_4$ and concentrated to a yellow oil, pure by TLC (Rf 0.5, 10% $MeOH/CHCl_3$). The oil was stirred 7 h in 1 ml CH_2Cl_2, 5 ml trifluoroacetic acid, and 1 drop thioanisole scavenger. The solution was concentrated to an oil and precipitated as a white solid by sonication with 1:1 hexane/ether: **76**, 113 mg.

^{1}H NMR (300 MHz, DMSO-d_6) δ 9.84 (br s, 2 H), 8.88 (d, J = 7.8 Hz, 2 H), 8.52 (s, 2H), 8.38 (s, 2 H), 7.63, 7.36 (m, 8 H), 4.88 (m, 4 H), 4.51 (m, 2 H), 1.75, 0.93 (m, 24 H).

Calcd. for $C_{47}H_{50}N_8O_9$: 870.3700, ESIMS [M+H]$^+$ found 871.5

73

203 mg PyBop (benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluoro-phosphate, 0.39 mmol) and diacid **79** (90 mg, 0.13 mmol) were stirred 5 min in 4 ml DMF. To this was added 75 mg leucine-t-butyl ester (0.33 mmol) and 250 ml diisopropylethylamine and the solution stirred 1 h. The solution was diluted with 200 ml CH_2Cl_2 and washed with 100 ml 1M citric acid and 2x200ml H_2O. The organic layer was concentrated to a yellow oil. The oil was stirred 7 h in 6 ml trifluoroacetic acid and 1 drop thioanisole scavenger. The solution was concentrated to an oil and precipitated as a white solid by sonication with 1:1 hexane/ether: **73**, 55 mg.

^1H NMR (300 MHz, DMSO-d_6) δ 9.81 (br t, 1 H), 9.34 (d, J = 7.5 Hz, 1 H), 8.85 (d, J = 7.5 Hz, 2 H), 8.53 (d, J = 2.7 Hz, 1 H), 8.37 (s, 2 H), 8.11 (d, J = 2.7 Hz, 1 H), 7.75, 7.53 (m, 4 H), 5.02 (m, 2 H), 4.48 (m, 3 H), 1.79 (m, 15 H), 0.91 (m, 18 H).

Calcd for [M+H]+ $C_{45}H_{55}N_6O_{11}$: 855.3928, HRMS FAB(3NBA) [M+H]+ found 855.3921

74, 75

Diacid **78** (210 mg, 0.29 mmol) was stirred with 453 mg PyBop (benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate, 0.87 mmol) for 5 min in 5 ml DMF. To this was added a solution in 3 ml DMF of 90 mg leucine-t-butyl ester (0.40 mmol), 89 mg 2(aminomethyl)-benzimidazole (0.40 mmol), and 250 ml diisopropylethylamine. The solution was stirred 2h, diluted with 200 ml CH_2Cl_2 and washed with 1 x 1M citric acid and 4 x 100 ml H_2O. The organics were dried over $MgSO_4$ and concentrated to a yellow oil, three products by TLC (Rf 0.05, 0.35, 0.95, in 5:20:75 MeOH/Hex/EtOAc). The crude material was chromatographed with 2:30:68 MeOH/Hex/EtOAc followed by 5:20:75 MeOH/Hex/EtOAc and 10:10:80 MeOH/ Hex/EtOAc to give all three products as oils, pure by TLC. The products having Rf = 0.05 and 0.35 were individually stirred 7 h in 2 ml CH_2Cl_2 / 5 ml trifluoroacetic acid. The solutions were concentrated to oils and precipitated as white solids by sonication with 1:1 hexane/ether: **74**, 84 mg; **75**, 79 mg.

74:

^1H NMR (300 MHz, DMSO-d_6) δ 9.58 (br s, 1 H), 9.03 (d, J = 6.4 Hz, 2 H), 8.83 (d, J = 7.8 Hz, 1 H), 8.37 (s, 2 H), 8.30, 8.26 (2 x s, 2 H), 7.67, 7.37 (m, 4 H), 4.86 (d, J = 4.2 Hz, 2 H), 4.43 (m, 3 H), 1.75 (m, 15 H), 0.92 (m, 18 H).

Calcd for [M+H]+ $C_{45}H_{55}N_6O_{11}$: 855.3928, HRMS FAB(3NBA) [M+H]+ found 855.3925

75:

^{1}H NMR (300 MHz, DMSO-d$_6$) δ 9.64 (br t, 2 H), 9.03 (d, J = 7.0 Hz, 2 H), 8.37 (d, J = 6.0 Hz, 4 H), 7.73, 7.44 (m, 8 H), 4.92 (m, 4 H), 4.42 (m, 2 H), 1.72 (m, 10 H), 0.91 (d, J = 5.5 Hz, 12 H).

Calcd. for C$_{47}$H$_{50}$N$_8$O$_9$: 870.3700, ESIMS [M+H]$^+$ found 871.5

77

Leucine-t-butyl ester (290 mg, 1.3 mmol) was stirred in 15 ml CH$_2$Cl$_2$ and 1 ml triethylamine. The solution was chilled and then added to a cold solution of 9,9-Dimethyl xanthene-2,4,5,7-tetracarboxylic acid chloride **50** (0.26 mmol) in 15 ml CH$_2$Cl$_2$. The reaction was stirred 3h, warming to RT. The solution was concentrated to a solid, which was taken up in 100ml CH$_2$Cl$_2$ and washed with 2 x 1M citric acid 100ml and 2 x 100 ml H$_2$O. The organic layer was dried over MgSO$_4$ and concentrated to a yellow oil. The oil was stirred in 1 ml CH$_2$Cl$_2$ / 5 ml trifluoroacetic acid and one drop thioanisole for 8 h. The solution was concentrated to a glass, disolved in 0.5 ml CHCl3, and precipitated with 1:1 ether/hexanes to yield a white solid: **77**, 185 mg.

^{1}H NMR (300 MHz, DMSO-d$_6$) δ 8.93 (d, J = 6.9 Hz, 2 H), 8.80 (d, J = 6.9 Hz, 2 H), 8.26 (dd, J = 6.6 Hz, J = 2.1 Hz, 4 H), 4.41 (m, 4 H), 1.74 (m, 18 H), 1.43, 1.40 (2 s, 36 H), 0.94 (m, 24 H).

Calcd. for C$_{43}$H$_{58}$N$_4$O$_{13}$: 838.4000, ESIMS [M+H]$^+$ found 839.6

Compounds from Scheme 9:

85

Diacid **78** (50 mg, 0.07 mmol) was stirred with 36 mg PyBop (benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate, 0.07 mmol) for 5 min in 3ml DMF. To this was added a solution in 2 ml DMF of 17 mg 2-(aminomethyl)-benzimidazole (0.08 mmol), and 250 ml diisopropylethylamine. The solution was stirred 1h, diluted with 200 ml CH$_2$Cl$_2$ and washed with 1 x 1M citric acid and 3 x 100 ml H$_2$O. The organic layer was concentrated to a solid with hexanes, and the crude material was chromatographed with 5-10% MeOH/CHCl3 to separate the first fraction (TLC Rf=0.9, 20% MeOH/CHCl3). This product was stirred 8 h in trifluoroacetic acid, the solution was concentrated to an oil, and then precipitated as a white solid by sonication with 1:1 hexane/ether: **85**, 25 mg.

^{1}H NMR (300 MHz, DMSO-d_6) δ 9.71 (br t, 1 H), 9.07 (m, 2 H), 8.30 (m, 4 H), 7.80 (m, 2 H), 7.55 (m, 2 H), 4.96 (d, J = 4.0 Hz, 2 H), 4.42 (m, 2 H), 1.75 (m, 12 H), 0.91 (d, J = 5.0 Hz, 12 H).

Calcd. for $C_{39}H_{43}N_5O_{10}$: 741.30096.

82/83

Diacid **79** (35 mg, 0.05 mmol) was stirred with 26 mg PyBop (benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate, 0.05 mmol) for 5 min in 2 ml DMF. To this was added a solution in 1 ml DMF of 11.5 mg leucine-t-butylester (0.05 mmol), and 100 ml diisopropylethylamine. The solution was stirred 1 h, diluted with 100ml CH_2Cl_2 and washed with 2 x 1M citric acid and 3 x 100 ml H_2O. The organic layer was concentrated to a solid with hexanes, and the crude material was stirred 7 h in trifluoroacetic acid. The solution was concentrated to an oil, and then concentrated to a solid with 1:2:2 CH_2Cl_2/hex/ether: **82/83** (as a mixture), 30 mg.

Compounds **84** and **86/87** were prepared by coworker Jerry Shipps. A sample of the trichloroethanol deprotection strategy is reproduced with his permission for compound **84**.

2,7-dileucine amide-4-AMB-5-xanthene carboxylic acid (84).

2,7-Dileucine(Ot-butyl)amide-4-AMB-5-(1,1,1)-trichloroethyl xanthene carboxylate (10 mg, 0.01 mmol) was dissolved in acetic acid (2 ml), zinc dust (excess, ~ 100 mg) added and the reaction stirred for six h with occasional sonication. TLC analysis showed that no starting material remained. To the oil was added TFA (1 ml). After six hours the TFA was evaporated, ether:hexanes added, causing a precipitate to form. A tan powder (7 mg, 0.09 mmol) was recovered.

^{1}H NMR (250 MHz, DMSO$_{d6}$): [note--spectrum was broad] δ 13.8 (br,s, 3 H, CO$_2$H), 12.6 (brs, 1 H, NH imide), 10.4 (br, 1 H, NH AMB amide), 8.87 (m, 2 H), 8.64 (br, 1 H), 8.48 (br, 1 H), 8.29 (br, 1 H), 8.26 (br, 1 H), 7.61 (m, 2 H), 7.35 (m, 2 H), 5.05 (m, 2 H), 4.50 (m, 2 H), 1.81-1.63 (m, 6 H), 1.65 (s, 6 H), 0.95-0.90 (m, 12 H).

Calcd. for $C_{39}H_{43}O_{10}N_5$: 741.30096.

Xanthene dimer synthesis

96

Diacid **78** (200 mg, 0.28 mmol) was stirred with 143 mg PyBop (benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate, 0.28 mmol) for 5 min in 3 ml DMF. To this was added a solution in 2 ml DMF of 73 mg leucine-t-butylester (0.33 mmol), and 100 ml diisopropylethylamine. The solution was stirred 1 h, diluted with 100 ml CH_2Cl_2 and washed with 2 x 100 ml 1M citric acid and 3 x 100 ml H_2O. The organic layer was concentrated to an oil, and the crude material was chromatographed on silica gel, 7% $MeOH/CHCl_3$ (TLC: product Rf=0.2, 10% MeOH/CHCl3). The purified material was obtained as a white foam, **96**, 175 mg.

^{1}H NMR (DMSO$_{d6}$, 300 MHz): δ 13.2 (br, s, 1 H), 8.98 (d, J=6.9 Hz, 1H), 8.92 (d, J=6.3 Hz, 1H), 8.80 (d, J=8.7 Hz, 1H), 8.271 (s, 3H), 8.248 (s, 1H), 4.35-4.44 (m, 3H), 1.6-1.9 (m, 15 H), 1.42 (s, 9H), 1.40 (s, 18H), 0.91-0.96 (m, 18H)
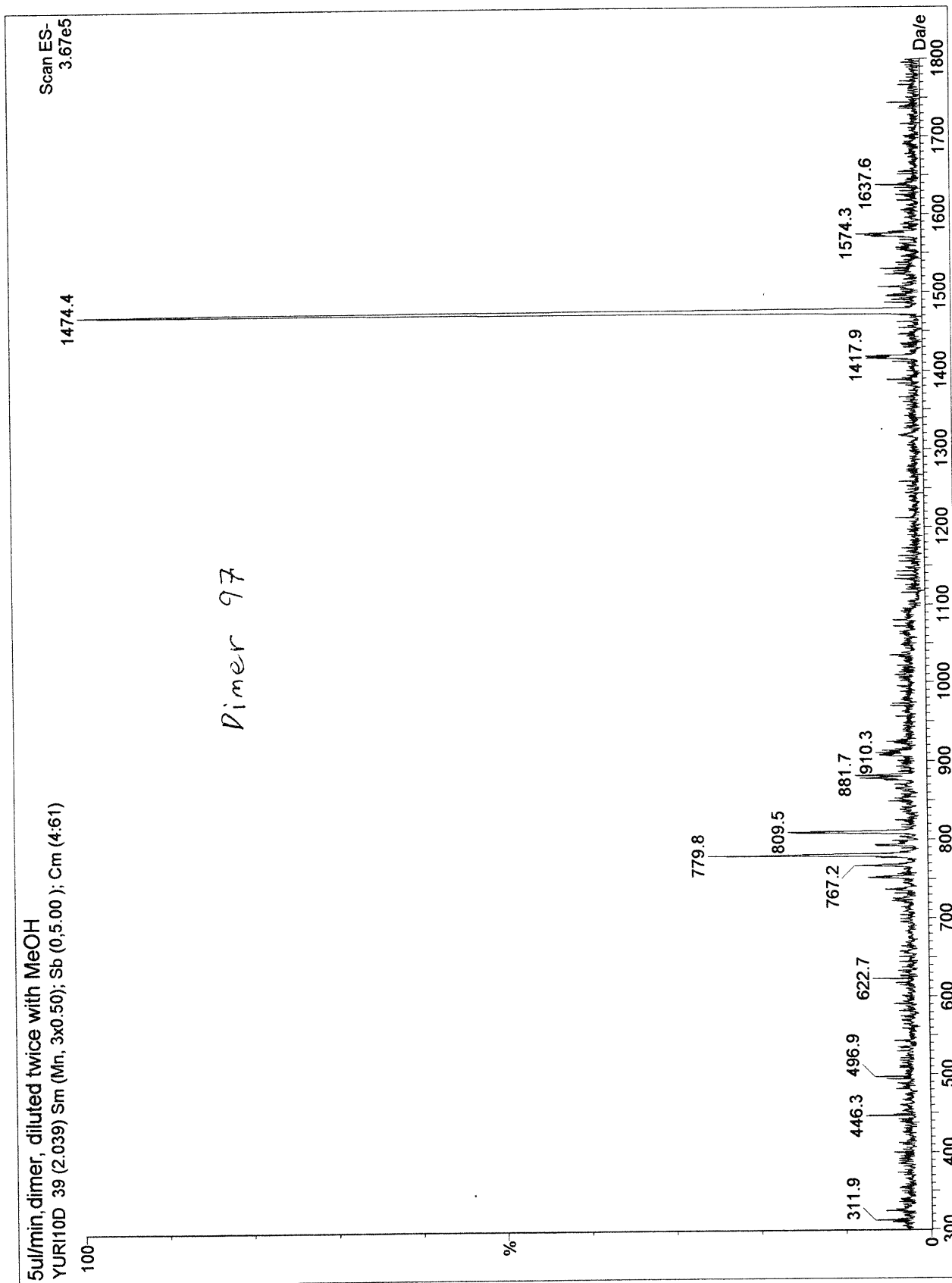
Calcd for $C_{49}H_{71}N_3O_{12}$: 893.5037

97

9,9-dimethyl-2-xanthene-4,5,7-tri-leucine-amide acid **96** (40 mg, 0.045 mmol) and 27 mg PyBop (benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluoro-phosphate, 0.53 mmol) were stirred 5 min in 2 ml DMF with 100 μl TEA. To this was added dropwise over 10 min 100 μl of a solution of 150 μl ethylenediamine in 10ml DMF (total 1.5 μl [1.34 mg] ethylenediamine added, 0.022 mmol). The solution was stirred 1 h, diluted with 100 ml CH_2Cl_2 and washed with 2 x 100 ml 1M citric acid and 3 x 100 ml H_2O. The organic layer was concentrated to an oil, then stirred 8 h in trifluoroacteic acid. The solution was concentrated, and the crude material was chromatographed on a short column of reverse phase C-18 silica gel (60-100% MeOH/H_2O) The purified material was obtained as a white solid: **97**, 19 mg.

^{1}H NMR (250 MHz, DMSO-d$_6$) δ 12.55 (br s, 6 H), 9.03 (d, J = 7.8 Hz, 2 H), 8.99 (d, J = 7.8 Hz, 2 H), 8.92 (br t, 2 H), 8.82 (d, J = 7.8 Hz, 2 H), 8.28 (d, J = 1.7 Hz, 8 H), 4.46 (m, 6 H), 3.51 (m, 4 H), 1.77 (m, 30 H), 0.92 (m, 36 H).

Calcd for $C_{76}H_{98}N_8O_{22}$: 1474.6795, ESIMS [M+H]$^+$ found 1474.4

5ul/min, dimer, diluted twice with MeOH
YURI10D 39 (2.039) Sm (Mn, 3x0.50); Sb (0,5.00); Cm (4:61)

Scan ES-
3.67e5

Dimer 97

AMB Leu-COOH

HOOC COOH

(t-Butyl deprotected 79)

SPEC: y0422a   30-APR-96   DERIVED SPECTRUM
Samp: ALOO
Mode: ESI +Q1MS LMR UP PROF   Start : 17:56:14
Oper: 1ul/min   Client: cubist   Inlet :
Base: 629.3   Inten : 36182016   Masses: 300 > 1500
Norm: 629.3   RIC : 246483491   #peaks: 6025
Peak: 300.00 mmu
Data: +/4>12

9
30

E+07
3.62

629.3

405.3   713.5   1257.8

100   80   60   40   20

400   600   800   1000   1200   1400

SPEC: y0422a
Samp: ALOO
Mode: ESI +Q1MS LMR UP PROF
Oper: 1ul/min   Client: cubist
Base: 629.3   Inten : 36182016
Norm: 1257.8   RIC   : 246483491
Peak: 300.00 mmu
Data: +/4>12

30-APR-96   DERIVED SPECTRUM
             Start : 17:56:14

             Inlet :
             Masses: 300 > 1500
             #peaks: 6025

9
30

t-Butyl deprotected 79 expanded

E+05
8.19

713.5
742.6
751.8
847.8
912.5
953.7
1032.4
1128.4
1257.8
1279.9
1377.1

800   1000   1200   1400

100   80   60   40   20
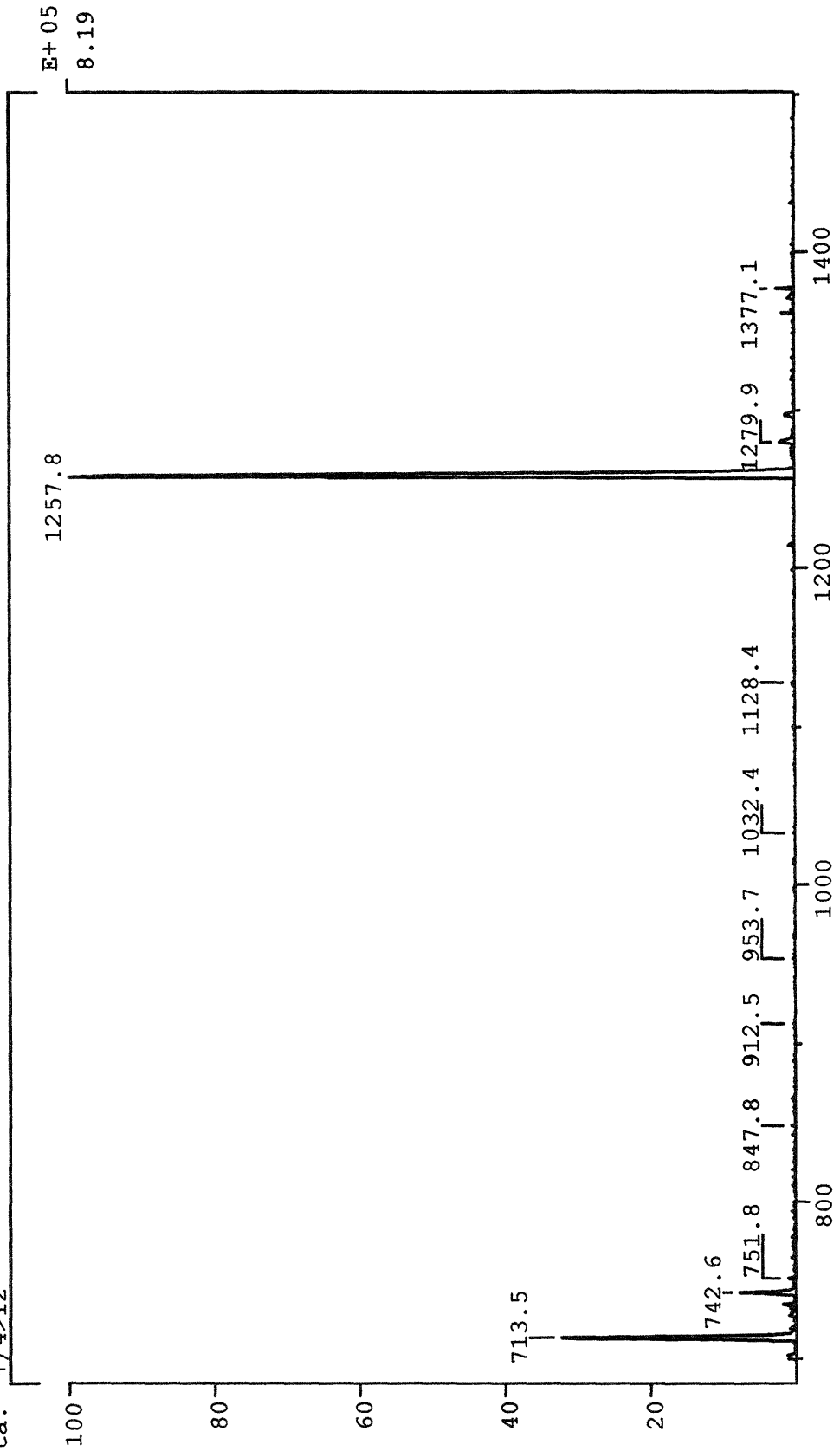
```
SCREEN 0
DIM libt(40)
DIM liba(40)
DIM libb(40)
DIM tools(40)
DIM rel(40)
DIM bb(30)
m = 0
REM setup of hit parameters
CLS
GOSUB 5000
PRINT "  Parameters for A site binders:";
LOCATE 14, 2: PRINT "  Position 1:        "
LOCATE 16, 2: INPUT "  Perfect fit building block"; onea
GOSUB 7000
LOCATE 16, 2: INPUT "     G value for perfect fit"; oneaa
GOSUB 7000
LOCATE 16, 2: INPUT "   Semi-fit building block 1"; oneax
GOSUB 7000
LOCATE 16, 2: INPUT "   Semi-fit building block 2"; oneay
GOSUB 7000
LOCATE 16, 2: INPUT "      G value for semi-fits"; oneaxy
GOSUB 7000
LOCATE 16, 2: INPUT "        Hydrophobic fit value"; oneap
GOSUB 7000
LOCATE 16, 2: INPUT "        Hydrophyllic fit value"; oneaq
GOSUB 7000
LOCATE 16, 2: INPUT "  Small side chain fit value"; onear

LOCATE 14, 2: PRINT "  Position 2:        "
GOSUB 7000
LOCATE 16, 2: INPUT "  Perfect fit building block"; twoa
GOSUB 7000
LOCATE 16, 2: INPUT "     G value for perfect fit"; twoaa
GOSUB 7000
LOCATE 16, 2: INPUT "   Semi-fit building block 1"; twoax
GOSUB 7000
LOCATE 16, 2: INPUT "   Semi-fit building block 2"; twoay
GOSUB 7000
LOCATE 16, 2: INPUT "      G value for semi-fits"; twoaxy
GOSUB 7000
LOCATE 16, 2: INPUT "        Hydrophobic fit value"; twoap
GOSUB 7000
LOCATE 16, 2: INPUT "        Hydrophyllic fit value"; twoaq
GOSUB 7000
LOCATE 16, 2: INPUT "  Small side chain fit value"; twoar

LOCATE 14, 2: PRINT "  Position 3:        "
GOSUB 7000
LOCATE 16, 2: INPUT "  Perfect fit building block"; thra
GOSUB 7000
LOCATE 16, 2: INPUT "     G value for perfect fit"; thraa
GOSUB 7000
LOCATE 16, 2: INPUT "   Semi-fit building block 1"; thrax
GOSUB 7000
LOCATE 16, 2: INPUT "   Semi-fit building block 2"; thray
GOSUB 7000
LOCATE 16, 2: INPUT "      G value for semi-fits"; thraxy
GOSUB 7000
LOCATE 16, 2: INPUT "        Hydrophobic fit value"; thrap
GOSUB 7000
LOCATE 16, 2: INPUT "        Hydrophyllic fit value"; thraq
GOSUB 7000
```

```
LOCATE 16, 2: INPUT "  Small side chain fit value"; thrar

LOCATE 14, 2: PRINT "  Position 4:      "
GOSUB 7000
LOCATE 16, 2: INPUT "  Perfect fit building block"; fora
GOSUB 7000
LOCATE 16, 2: INPUT "     G value for perfect fit"; foraa
GOSUB 7000
LOCATE 16, 2: INPUT "  Semi-fit building block 1"; forax
GOSUB 7000
LOCATE 16, 2: INPUT "  Semi-fit building block 2"; foray
GOSUB 7000
LOCATE 16, 2: INPUT "      G value for semi-fits"; foraxy
GOSUB 7000
LOCATE 16, 2: INPUT "       Hydrophobic fit value"; forap
GOSUB 7000
LOCATE 16, 2: INPUT "       Hydrophyllic fit value"; foraq
GOSUB 7000
LOCATE 16, 2: INPUT "  Small side chain fit value"; forar

LOCATE 16, 2: PRINT "                                        "
LOCATE 14, 2: INPUT "  A site binder multiplier"; facta
LOCATE 14, 2: PRINT "                             "

LOCATE 12, 2: PRINT "  Parameters for B site binders:";
LOCATE 14, 2: PRINT "  Position 1:      "
LOCATE 16, 2: INPUT "  Perfect fit building block"; oneb
GOSUB 7000
LOCATE 16, 2: INPUT "     G value for perfect fit"; onebb
GOSUB 7000
LOCATE 16, 2: INPUT "  Semi-fit building block 1"; onebx
GOSUB 7000
LOCATE 16, 2: INPUT "  Semi-fit building block 2"; oneby
GOSUB 7000
LOCATE 16, 2: INPUT "      G value for semi-fits"; onebxy
GOSUB 7000
LOCATE 16, 2: INPUT "       Hydrophobic fit value"; onebp
GOSUB 7000
LOCATE 16, 2: INPUT "       Hydrophyllic fit value"; onebq
GOSUB 7000
LOCATE 16, 2: INPUT "  Small side chain fit value"; onebr

LOCATE 14, 2: PRINT "  Position 2:      "
GOSUB 7000
LOCATE 16, 2: INPUT "  Perfect fit building block"; twob
GOSUB 7000
LOCATE 16, 2: INPUT "     G value for perfect fit"; twobb
GOSUB 7000
LOCATE 16, 2: INPUT "  Semi-fit building block 1"; twobx
GOSUB 7000
LOCATE 16, 2: INPUT "  Semi-fit building block 2"; twoby
GOSUB 7000
LOCATE 16, 2: INPUT "      G value for semi-fits"; twobxy
GOSUB 7000
LOCATE 16, 2: INPUT "       Hydrophobic fit value"; twobp
GOSUB 7000
LOCATE 16, 2: INPUT "       Hydrophyllic fit value"; twobq
GOSUB 7000
LOCATE 16, 2: INPUT "  Small side chain fit value"; twobr

LOCATE 14, 2: PRINT "  Position 3:      "
GOSUB 7000
LOCATE 16, 2: INPUT "  Perfect fit building block"; thrb
GOSUB 7000
LOCATE 16, 2: INPUT "     G value for perfect fit"; thrbb
GOSUB 7000
```

```
LOCATE 16, 2: INPUT "    Semi-fit building block 1"; thrbx
GOSUB 7000
LOCATE 16, 2: INPUT "    Semi-fit building block 2"; thrby
GOSUB 7000
LOCATE 16, 2: INPUT "         G value for semi-fits"; thrbxy
GOSUB 7000
LOCATE 16, 2: INPUT "         Hydrophobic fit value"; thrbp
GOSUB 7000
LOCATE 16, 2: INPUT "         Hydrophyllic fit value"; thrbq
GOSUB 7000
LOCATE 16, 2: INPUT " Small side chain fit value"; thrbr

LOCATE 14, 2: PRINT " Position 4:      "
GOSUB 7000
LOCATE 16, 2: INPUT " Perfect fit building block"; forb
GOSUB 7000
LOCATE 16, 2: INPUT "  ' G value for perfect fit"; forbb
GOSUB 7000
LOCATE 16, 2: INPUT "    Semi-fit building block 1"; forbx
GOSUB 7000
LOCATE 16, 2: INPUT "    Semi-fit building block 2"; forby
GOSUB 7000
LOCATE 16, 2: INPUT "         G value for semi-fits"; forbxy
GOSUB 7000
LOCATE 16, 2: INPUT "         Hydrophobic fit value"; forbp
GOSUB 7000
LOCATE 16, 2: INPUT "         Hydrophyllic fit value"; forbq
GOSUB 7000
LOCATE 16, 2: INPUT " Small side chain fit value"; forbr

LOCATE 16, 2: PRINT "                                      "
LOCATE 14, 2: INPUT "  B site binder multiplier"; factb
LOCATE 14, 2: PRINT "                              "
GOTO 10

5000 CLS : PRINT : PRINT : PRINT "   1: Phe    2: Met    3: Ile    4: Leu    5:
Val"
PRINT : PRINT "   6: Trp    7: Ala    8: Thr    9: Gly   10: Ser"
PRINT : PRINT "  11: Pro   12: Tyr   13: His   14: Asn   15: Glu"
PRINT : PRINT "  16: Lys   17: Asp   18: Arg": PRINT : PRINT
RETURN

6000 PRINT " Already chosen: "; bb(1); bb(2); bb(3); bb(4); bb(5); bb(6); bb(7)
; bb(8); bb(9); bb(10)
PRINT : PRINT "                    "; bb(11); bb(12); bb(13); bb(14); bb(15); bb(1
6); bb(17); bb(18); bb(19); bb(20)
PRINT : PRINT "                    "; bb(21); bb(22); bb(23); bb(24); bb(25); bb(2
6); bb(27); bb(28); bb(29); bb(30)
PRINT : PRINT
RETURN

7000 LOCATE 16, 20: PRINT "                    "
RETURN

10 CLS : acta = 0: actb = 0: k = 0
PRINT : PRINT : INPUT "  How many building blocks in this sub-library"; n
FOR x = 1 TO n
GOSUB 5000
GOSUB 6000
PRINT " Building block number"; x; : INPUT bb(x)
IF bb(x) > 18 OR bb(x) < 1 THEN LET k = 1
NEXT x

GOSUB 5000
GOSUB 6000
IF k = 1 THEN GOTO 60
```

```
FOR w = 1 TO n
FOR x = 1 TO n
FOR y = 1 TO n
FOR z = 1 TO n

LET ga = 0: LET gb = 0


IF bb(w) = onea THEN ga = ga + oneaa
IF bb(w) = oneax OR bb(w) = oneay THEN ga = ga + oneaxy
IF bb(w) < 8 THEN ga = ga + oneap
IF bb(w) > 12 THEN ga = ga + oneaq
IF bb(w) > 6 AND bb(w) < 12 THEN ga = ga + onear

IF bb(x) = twoa THEN ga = ga + twoaa
IF bb(x) = twoax OR bb(x) = twoay THEN ga = ga + twoaxy
IF bb(x) < 8 THEN ga = ga + twoap
IF bb(x) > 12 THEN ga = ga + twoaq
IF bb(x) > 6 AND bb(x) < 12 THEN ga = ga + twoar

IF bb(y) = thra THEN ga = ga + thraa
IF bb(y) = thrax OR bb(y) = thray THEN ga = ga + thraxy
IF bb(y) < 8 THEN ga = ga + thrap
IF bb(y) > 12 THEN ga = ga + thraq
IF bb(y) > 6 AND bb(y) < 12 THEN ga = ga + thrar

IF bb(z) = fora THEN ga = ga + foraa
IF bb(z) = forax OR bb(z) = foray THEN ga = ga + foraxy
IF bb(z) < 8 THEN ga = ga + forap
IF bb(z) > 12 THEN ga = ga + foraq
IF bb(z) > 6 AND bb(z) < 12 THEN ga = ga + forar

IF bb(w) = oneb THEN gb = gb + onebb
IF bb(w) = onebx OR bb(w) = oneby THEN gb = gb + onebxy
IF bb(w) < 8 THEN gb = gb + onebp
IF bb(w) > 12 THEN gb = gb + onebq
IF bb(w) > 6 AND bb(w) < 12 THEN gb = gb + onebr

IF bb(x) = twob THEN gb = gb + twobb
IF bb(x) = twobx OR bb(x) = twoby THEN gb = gb + twobxy
IF bb(x) < 8 THEN gb = gb + twobp
IF bb(x) > 12 THEN gb = gb + twobq
IF bb(x) > 6 AND bb(x) < 12 THEN gb = gb + twobr

IF bb(y) = thrb THEN gb = gb + thrbb
IF bb(y) = thrbx OR bb(y) = thrby THEN gb = gb + thrbxy
IF bb(y) < 8 THEN gb = gb + thrbp
IF bb(y) > 12 THEN gb = gb + thrbq
IF bb(y) > 6 AND bb(y) < 12 THEN gb = gb + thrbr

IF bb(z) = forb THEN gb = gb + forbb
IF bb(z) = forbx OR bb(z) = forby THEN gb = gb + forbxy
IF bb(z) < 8 THEN gb = gb + forbp
IF bb(z) > 12 THEN gb = gb + forbq
IF bb(z) > 6 AND bb(z) < 12 THEN gb = gb + forbr


acta = acta + facta ^ (ga)
actb = actb + factb ^ (gb)

v$ = INKEY$

LOCATE 20, 4: PRINT w, x, y, z;
```

```
NEXT z
NEXT y
NEXT x
NEXT w

FOR x = 1 TO 3
BEEP
FOR y = 1 TO 3000
NEXT y
NEXT x

20 v$ = INKEY$
25 IF v$ = "" THEN 20


30 CLS
acta = INT(acta): actb = INT(actb)
act = acta + actb: LOCATE 5, 1
PRINT : PRINT : PRINT "Total Inhibitory Activity:"; act
PRINT : PRINT "Inhibition from A site binders:"; acta
PRINT : PRINT "Inhibition from B site binders:"; actb
LET m = m + 1
LET libt(m) = act: LET liba(m) = acta: LET libb(m) = actb: LET tools(m) = n

40 v$ = INKEY$
45 IF v$ = "" THEN GOTO 40
IF v$ = "e" THEN END

47 CLS : PRINT : PRINT "Inhibitory Activity Summary:"
PRINT : PRINT : PRINT "              Tot. Act.:    Rel. Act.:      Act. A:        A
ct. B:    # Tools"
PRINT
LET big = libt(m)
FOR x = 1 TO m - 1
IF libt(x) < big THEN 50
LET big = libt(x)
50 NEXT x

FOR x = 1 TO m
LET rel(x) = INT((libt(x) / big) * 100)
PRINT "Library"; x; ":", libt(x), rel(x), liba(x), libb(x)
LOCATE 6 + x, 70: PRINT tools(x)
NEXT x

LET small = rel(m)
FOR x = 1 TO m - 1
IF rel(x) > small THEN 55
LET small = rel(x)
55 NEXT x

57 v$ = INKEY$
IF v$ = "" THEN GOTO 57
SCREEN 2

LOCATE 2, 5: PRINT "Activity of Trypsin Asssay"
FOR x = 1 TO m
LINE (11 + (x - 1) * 40, 50 - small + rel(x))-(4 + x * 40, 155), , BF
LOCATE 21, 5 * x - 2: PRINT x
NEXT x

60 v$ = INKEY$
65 IF v$ = "" THEN GOTO 60
SCREEN 0

IF v$ = "e" THEN END
IF v$ = "r" THEN 47
FOR x = 1 TO 30
LET bb(x) = 0
NEXT x
GOTO 10
```