STUDIES ON HUMAN SERINE PROTEASES: TRYPSIN, PLASMIN, THROMBIN AND $C\bar{1}s$

DANIEL PAUL ROMAN, JR.

THESES FOR THE DEGREE OF M. S.

MICHIGAN STATE UNIVERSITY

1977

This is to certify that the

thesis entitled

STUDIES ON HUMAN SERINE PROTEASES:

TRYPSIN, PLASMIN, THROMBIN, AND Cl_s

presented by

Daniel Paul Roman, Jr.

has been accepted towards fulfillment
of the requirements for

Master's degree in Microbiology

Richard A. Stein
Major professor

Date December 14, 1977

O-7639

ABSTRACT

STUDIES ON HUMAN SERINE PROTEASES: TRYPSIN
PLASMIN, THROMBIN AND Cls

by

Daniel Paul Roman, Jr.

A series of substituted benzamidines were investigated as inhibitors of the hydrolysis of N^{α}-carbobenzoxy-L-Tyrosine-para-nitrophenyl ester (N^{α}-Z-L-Tyr-p-Np) by the human proteases trypsin, plasmin, $C\bar{1}s$ and thrombin. The inhibition constants (K_i) were measured and correlated with various substituent parameters to try to determine some of the factors that controlled the magnitude of K_i. The most important single variable involved in the correlations for trypsin, thrombin and $C\bar{1}s$ was R, the Swain and Lupton resonance parameter, resulting in equations of the form $\log 1/K_i = k_1R + k_2$ where k_1 and k_2 are calculated constants from regression analysis. For $C\bar{1}s$, the inclusion of π_{NO_2} and $(\pi_{NO_2})^2$, measures of substituent hydrophobicity derived from the nitrobenzene series, results in a significantly better correlation. In the case of plasmin, molar refractivity (MR) and π (benzene series) were the most significant single variables. It does not appear that any electronic effect plays a significant role in benzamidine binding to plasmin, thus making it very different from the other three enzymes. A comparison

between human trypsin and thrombin and their bovine counterparts indicates to us that differences do exist, most possibly in the hydrophobic regions of these enzymes. The equations generated for human plasmin in this work and in the work by Coats (1973) are similar, indicating that a mass synthesis approach to inhibitor design is unnecessary.

STUDIES ON HUMAN SERINE PROTEASES: TRYPSIN PLASMIN, THROMBIN AND C1S

by

Daniel Paul Roman, Jr.

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTERS OF SCIENCE

Department of Microbiology and Public Health

1975

ACKNOWLEDGEMENTS

I am grateful to the Department of Microbiology and Public Health for their partial financial support, and to the Center for Blood Research for allowing me to use its facilities during the course of this work.

I would also like to express my appreciation to Dr. Michael Cory, SRI, who did the biological correlations and to the following people: Dr. Judith Andrews, Gail Conneely, Dr. Dianne Kenney, Dr. Fabian Lionetti and Stephen Hunt for their assistance in reading the thesis and discussing the data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The human body contains hundreds of enzymes and proteins which vary widely in their structure and function. One important class of enzymes is the serine proteases. These all have a serine residue essential for the catalytic hydrolysis of both natural and synthetic ester substrates. The following study was carried out on four enzymes; thrombin, plasmin, trypsin and $Cl\bar{s}$ which differ in their physiological roles but are similar in two important aspects. First, that they are all serine proteases, and second their hydrolytic action is almost exclusively directed toward l-lysyl and l-argininyl bonds in polypeptides. The anionic binding site which all the enzymes are known or presumed to possess is a possible starting point for a comparison of the nature of the substrate specificity of these enzymes.

Much is known about the biological roles and the physical-chemical properties of these enzymes. Trypsin has been the most studied of the group and much physical data has been accumulated, culminating recently in the three dimensional structure of bovine trypsin as determined by X-ray crystallographic techniques. Biologically, this pancreatic enzyme is responsible for the activation of all the other zymogens of pancreatic tissue. The fact that trypsin will hydrolyze virtually any peptide bond adjacent to a basic amino acid is probably critical to its role as a digestive enzyme.

Plasmin is functionally similar to trypsin. It is an endopeptidase and will hydrolyze a wide variety of peptide bonds adjacent to an

argininyl or lysyl residue. Its main role is the dissolution of insoluble fibrin polymers; however, it also hydrolyzes a number of other naturally occurring proteins. Thrombin is an enzyme of the coagulation sequence and its major substrate is fibrinogen. Thrombin and other coagulation enzymes react on fibrinogen to give fibrin polymers and a stable fibrin clot. As with plasmin and trypsin, the bond in fibrinogen which is cleaved is an arginine bond, indicating an anionic binding site in thrombin.

$\bar{C1s}$, a subunit of the first component of complement, Cl, is again a serine protease which is responsible for the cleavage of C4 and C2, early acting components of the complement sequence. Cl initiates the complement system which acts as a biological amplifier of the immune response. The action of $\bar{C1s}$ on C4 and C2 also results in the generation of phlogistic peptides important in immune hypersensitivity states. $\bar{C1s}$ is inhibited by substituted benzamidines and guanidines, suggesting an anionic binding site adjacent to the catalytic site.

The esterolytic activity of each enzyme was determined with a simple spectrophotometric assay procedure. The same substrate was used for each of the enzymes so that comparable inhibition constants could be accurately determined. Inhibition constants (K_i) for a series of substituted benzamidines was determined and the K_i correlated with various substituent constants. The equations gnerated were used to compare the four enzymes to determine the factors which control substrate specificity.

This thesis is divided into two parts. The first part is a literature review, in which pertinent information on the nature of trypsin, thrombin, plasmin and $\bar{C1s}$, the kinetics of the reaction involved, and on the use of the Hansch multiparameter regression analysis in correlating biological activity of inhibitors with substituent parameters is discussed. The second part deals with the correlation of K_i with substituent

parameters and a determination of factors important in determining the biological activity of these compounds. A comparison is made of the human and bovine systems for thrombin, trypsin and plasmin while $C1\bar{s}$ is compared to whole guinea pig complement to determine if any significant differences are seen between the species.

LITERATURE REVIEW

Part I

Trypsin, Thrombin, Plasmin and $C1\bar{s}$

Introduction

Human serine proteases comprise an important class of enzymes in the body. Their wide range of functions and specificities are important in contributing to the health or, unfortunately, to the detriment of the individual. Volumes of data have been collected on trypsin and the so called "trypsin-like" enzymes. This review will be concerned with detailing some of the significant similarities and differences of four enzymes of the trypsin class, trypsin, thrombin, plasmin and $C1\bar{s}$. The second part of this review will be a description of the Hansch multiparameter regression analysis technique and its application, not just to drug design but, to developing ideas concerning possible structural differences in enzyme binding sites.

Trypsin

Trypsin has been one of the most widely studied enzymes in the class of serine proteases. Serine proteases are so named because these enzymes are inactivated by the reaction of a serine residue in the active site of these molecules by diisopropylfluorophosphate (DIFP). Trypsin is inhibitable by this compound (1). Another important feature in the mechanism of action of trypsin is the presence of a charge relay system consisting of aspartate, histidine and serine residues. The structural evidence for this system has

been shown by X-ray crystallographic data (2). The inactivation of trypsin by tosyl lysine chloromethyl ketone (TLCK), which reacts with the active site histidine, and other experiments have helped confirm the importance of an intact charge relay system (3).

Human trypsin has been purified from pancreatic tissues by adsorption to Sepharose-Trasylol^{R} columns and then subsequent elution with a pH gradient (4). Trypsin is secreted by the pancreas as the proenzyme trypsinogen. Trypsinogen is activated to trypsin by the hydrolysis of a single peptide bond near the N-terminus. Trypsin can undergo further hydrolysis and at least five varieties of trypsin are known. Comparisons of human trypsin with trypsins from other species indicate many similarities including molecular weight, amino acid composition, pH optimum and inhibition by several synthetic and naturally occurring inhibitors (5,6).

Trypsin exhibits a marked preference for lysinyl and argininyl side chains in the hydrolysis of peptide bonds (7,8,9); however, it can hydrolyze some neutral substrates (10). Although trypsin is relatively specific in its requirement for basic amino acid residues, its specificity is broad and it will hydrolyze virtually any peptide, amide, or ester which contains a lysinyl or argininyl side chain. Since trypsin is a digestive enzyme and is responsible for the activation of the other zymogens in the pancreatic tissue, this wide hydrolytic activity takes on a special significance.

A wide variety of substrates have been used to determine trypsin activity and a number of different types of inhibitors are also known (11). The development of an active site directed titrant p-nitrophenyl-p'-guanidinobenzoate (PNPGB) by Chase and Shaw (12) has allowed the accurate determination of the amount of active enzyme in any solution. The importance of an active site titrant has been shown in a recent example by Moroi et al.

(15) which demonstrates that α-1-antitrypsin combines in a 1:1 ratio with trypsin instead of a 1:2 complex as has been described elsewhere (13,14).

Thrombin

Thrombin is a serine protease and both its esteratic and clotting activities can be blocked with DIFP (16,17). Thrombin's physiological role is the formation of fibrin monomers from fibrinogen following the activation of prothrombin to thrombin via interaction with thromboplastin, factor V, factor Xa and Ca^{++} (18). Thrombin is purified by a combination of ion exchange chromatography, ethanol fractionation and gel filtration. Combination of the above techniques with affinity chromatography on p-chlorobenzylamido-ε-amino caproyl agarose or mNH_2-benzamidino agarose results in the recovery of protein that is 90% active based on active site titration with PNPGB (19,20).

Thrombin consists of two chains, A and B. It is the B chain that contains the reactive serine residue analogues to trypsin (21). Although no X-ray crystallographic data is available, there is a high degree of homology between the B chain and chymotrypsin A and B, trypsin and elastase (21). Because of this homology the existence of the charge relay system can be postulated. The fact that thrombin is inhibited by DIFP and also be tosyl lysine chloromethyl ketone (TLCK), which reacts with the trypsin active site histidine residue (22,23), is further evidence for the presence of this system.

Thrombin is much more specific in its activity than trypsin or plasmin. The major physiological substrate for thrombin is fibrinogen. The formation of fibrin monomers is brought about through the cleavage of an arginine-glycine bond (24). Thrombin can also hydrolyze thrombosthenin M (25) and the two hormones secretin (26) and cholecystokinin-pancreozymin (27); all

three are cleaved at a susceptible argininyl peptide bond. Although no lysine containing polypeptide is known to be a substrate for thrombin, it can hydrolyze N-carbobenzoxy-L-lysine-p-nitrophenyl ester (28) and p-toluene sulfonyl-L-lysine methyl ester (29). Thrombin, like trypsin, can be titrated with p-nitrophenyl-p'-guanidinobenzoate (30), p-nitrophenyl-m'-guanidinobenzoate can also be used as an active site titrant (30).

Thrombin can be inhibited by benzamidine and benzylamines, although they are poor inhibitors of thrombin compared to their inhibition of plasmin and trypsin (31).

Plasmin

Plasmin like thrombin and trypsin is a serine protease inhibitable by DIFP. It is also specifically inactivated by TLCK which reacts with a single histidine in the active center (32,33,34). Plasmin's physiological function is the lysis of blood clots via the proteolysis of fibrin. Plasmin is purified as plasminogen by affinity chromatography on lysine-Sepharose columns (35). The plasminogen obtained is then activated to plasmin by the addition of urokinase (36). The activation of plasmin, molecular weight 75,400 (37), occurs via the cleavage of a single arginyl valine bond (33,34,38). Reduction and alkylation of plasmin results in the formation of two major polypeptide chains, a heavy chain of approximately 49,000 and a light chain of 26,000. The active site serine is contained in the light chain (32,33,34).

Plasmin is similar to trypsin in its specificity for lysine and arginine esters (39). Its action on S-sulfo fibrinogen is also similar to that of trypsin, since only lysine and arginine bonds are cleaved; however, whereas, trypsin will cleave 80% of the bonds of this molecule, plasmin will hydrolyze only 50% (40,41). Plasmin can be titrated with p-nitrophenyl-p'-

guanidinobenzoate, and will hydrolyze p-nitrophenyl-m'-guanidinobenzoate analog as well (30). Plasmin is inhibited by substituted benzamidines (42,43).

$C1\bar{s}$

Cls is a subunit of the first component of complement, Cl. Complement is a system of eleven proteins that act as biological amplifiers in immunologically mediated reactions and are also important in hypersensitivity disease states. Cls exists in serum as a proenzyme, which is activated upon interaction of Cl with antigen-antibody aggregates (44,45,46,47). Cls can also be activated by plasmin and trypsin (48). C4 and C2, two early acting components of the complement sequence are the only two protein substrates known for $C1\bar{s}$ (49,50,51). The action of $C1\bar{s}$ on C4 and C2 is highly restricted. In each molecule only one peptide bond is cleaved (49,50,52). It is important to point out that the nature of the bonds cleaved in C2 and C4 is not known, so that a comparison with trypsin, plasmin and thrombin as a protease with specificity towards basic amino acid side chains must be based on other evidence.

$C1\bar{s}$ has been purified by a variety of methods including differential salt fractionation, ion exchange chromatography and gel filtration (53,54). More recently, affinity chromatography coupled with preparative electrophoresis has provided relatively high yields of purified enzyme (55).

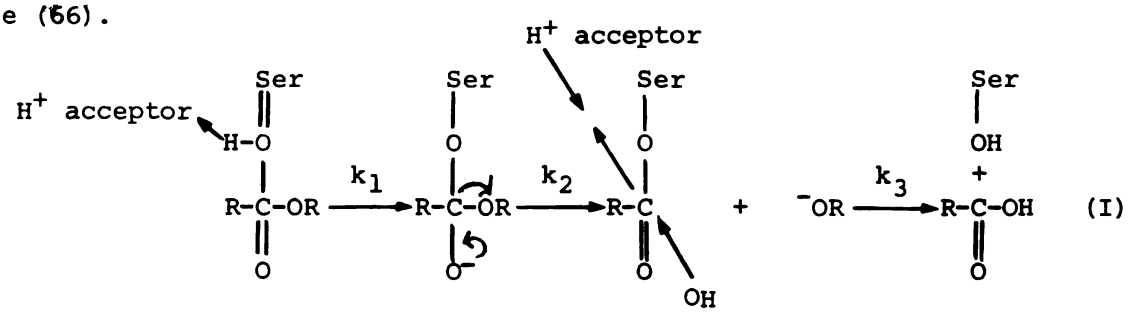
$C1\bar{s}$ is a serine protease inhibitable by DIFP (56). Sulfonyl fluoride compounds which react with the serine hydroxyl group of serine proteases are known to irreverisbly inactivate $C1\bar{s}$ (57). Inhibition of $C1\bar{s}$ by benzamidine, pyridinium and guanidinium compounds is presumptive evidence for an anionic binding site (57,58). Further evidence for an anionic site is provided by the fact that $C1\bar{s}$ will hydrolyze N-tosyl-l-arginine methyl

ester (60) and N-carbobenzoxy-L-lysine-p-nitrophenyl ester (Roman, D.P., this work). Although the preceeding data can be used to tentatively classify $C1\bar{s}$ as a "trypsin-like" enzyme, important differences must be noted. $C1\bar{s}$ does not react with p-nitrophenyl-p'-guanidinobenzoate (61). Phenyl methane sulfonyl fluoride which inhibits a number of proteases will not inactivate $C1\bar{s}$ (62). TLCK and TPCK, two reagents that react with active site histidines involved in the charge relay system do not inactivate $C1\bar{s}$ (61).

$C1\bar{s}$ is similar to plasmin in its activation mechanism. The conversion of Cls to $C1\bar{s}$ results in the formation of two polypeptide chains, a heavy chain (M.W. 70,000) and a light chain (M.W. 30,000). It has been shown that the DFP reactive serine residue is in the light chain of the $C1\bar{s}$ molecule (63,64,65). The properties of these enzymes are summarized in Table 1.

Enzyme Kinetics

Before one can investigate the nature of the active sites of enzymes, an appreciation of the mechanism of action of proteolytic enzymes and the evaluation of kinetic parameters which can be used in such studies should be developed. The four enzymes under study are all serine proteases and the hydrolysis of substrates proceeds via a double displacement mechanism. The following scheme shows the reaction of a substrate with a serine protease (66).



In the first step of the reaction enzyme (E) plus a substrate (S) combine to form an ES complex, presumably in the form of a tetrahedral intermediate (I).

Table 1. Properties of Trypsin, Plasmin, Thrombin and $C1\bar{s}$

Enzyme	M.W.	Active Site Serine Sequence	Peptide Bond Specificity	Protein Substrates	Inhibitors DFP	Inhibitors TLCK	Inhibitors Benzamidine	Molecular Titrant
Trypsin	23,000	Bovine CQGDSGGPV	Arg, Lys	All	Yes	Yes	Yes	PNPGB
Plasmin	H-chain 68,000 L-chain 25,000	Human CQGDSGGPL	Arg, Lys	Includes: Casein, $C1\bar{s}$, Insulin, Fibrin + Others	Yes	Yes	Yes	PNPGB
Thrombin	H-chain 36,000 L-chain 17,000	Bovine CQGDSGGPF	Arg	Fibrinogen, Secretin, Cholecystokinin, Pancreozymin	Yes	Yes	Yes	PNPGB
$C1\bar{s}$	H-chain 70,000 L-chain 30,000	Human CGKDSGEGR	Unknown	C4, C2	Yes	No	Yes	None Known

N-Z-L-Tyr-p-Np, N-Z-L-Lys-p-Np and Tosyl-Arg-OMe are representative substrates for all four enzymes.

The next step is the release of the alcoholic or amino portion of the substrate leaving an acylated enzyme intermediate, ES'. The final step, if $k_3>0$, is the return to free enzyme by the addition of H_2O and the release of the acidic moiety of the substrate. The main feature of the reaction of the enzyme with the substrate is the participation of a catalytic triad of the enzyme which provides the necessary charge relay and nucleophilic attack groups for hydrolysis. The triad consists of aspartate, histidine and serine residues. The charge relay system shown for α-lytic protease and assumed for other serine proteases is shown in scheme II (67).

Some ways to investigate the above reaction mechanisms are to use specific acylating or alkylating reagents that can react with the active site serine or histidine residues. Two popular reagents are diisopropylfluorophosphate and N^{α}-tosyl-L-lysine chloromethyl ketone, serine and histidine, acylating and alkylating reagents, respectively. Another method for examining the reaction of these enzymes, particularly the specificity controlling features, is through the use of low molecular weight organic inhibitors. The use of competitive inhibitors can provide useful information concerning the nature of the binding sites of these enzymes. In order to carry out such studies, a knowledge of the basic kinetics of the reactions of the enzymes under study is necessary.

The following reaction scheme is necessary for the development of the kinetic parameters which can be used for investigating the binding and active sites of the serine proteases (68,69,70).

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} ES' + P_1 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P$$

$$ES' + H^+ \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} ES'H^+ \qquad K_i = \frac{k_{-4}}{k_4}$$

If one assumes that in the early stages of the reaction $k_2[ES] >> k_{-2}[ES][P_1]$, $k_3[ES] >> k_3[E][P_2]$ and $[S] = [S]_o$ where $[S]_o$ is the initial substrate concentration. Then application of the steady state approximation to ES yields

$$\{(k_2 + k'_3)[S]_o + k'_3K_m\}[E] =$$

$$k'_3K_m[E]_o + \frac{k_2K_m[S]_o[E]_o}{K_m + [S]_o} \exp \frac{\{-[(k_2 + k_3)\, S_o + k'_3K_m]t\}}{K_m + [S]_o}$$

where $[E]_o = [E] + [ES] + [ES']_t$; $k'_3 = k_3K_i/(K_i + [H]^+)$; $K_m = \frac{k_{-1} + k_2}{k_1}$

At small values of t, if sufficiently high concentrations of $[S]_o$ are used such that $(k_2 + k'_3)[S]_o >> k'_3K_m$; $k_2 + k'_3$ and K_m can be evaluated by means of a Lineweaver-Burk plot of the following equation at various substrate concentrations, where K is the first order rate constant for the reaction

$$k = \frac{(k_2 + k'_3)[S]_o + k'_3K_m}{K_m + [S]_o}$$

Measurements of the rate of appearance of P_1 or P_2 at large values of t so that the exponential term is very small allows one to calculate the apparent constants $K_{m(app)} = \frac{k'_3 \quad K_m}{k_2 + k'_3}$ and $k_{cat}[E]_o$ from the following equation

$$V = \frac{dP_1}{dt} = \frac{dP_2}{dt} = \frac{k_2k'_3[S]_o[E]_o}{(k_2 + k'_3[S]_o + k'_3K_m}$$

$V_{max} = (k_{cat}[E]_o)$ is thus determined.

With the kinetic constants established, it is now possible to examine effects of inhibitors on the reaction of an enzyme and its substrate. This study is to examine the effect of structural changes of synthetic inhibiors on competitive inhibition constants. Once competitive inhibition has been shown, K_i values can be readily determined. The usual method for the determination of K_i (assumed to be the enzyme-inhibitor dissociation constant) is

to plot 1/v vs [I] at two or more substrate concentrations (71). The point of intersection of these lines will be at a value of $[I] = -K_i$, $1/v = 1/v_{max}$. Although it is theoretically possible to substantiate that competitive inhibition is taking place with the above method, because of experimental error some ambiguity does remain in differentiating competitive from mixed inhibition. In mixed inhibition plots of 1/v vs [I] at different substrate concentrations yields intersecting lines that intersect at a point $i = -K_i$, $1/v = [1-(K_i/K'_i)]/v_{max}$ (72). Competitive inhibition is an extreme case where $K'_i \to \infty$, thus giving the required intercepts. In order to further demonstrate competitive inhibition, one should make plots of s/v vs [I] at different substrate concentrations. The observed pattern of lines, parallel for competitive inhibition, delineates the mode of inhibition (72). Once competition has been shown it is only necessary to run inhibition experiments at one substrate concentration, provided v_{max} has already been determined. A line drawn parallel to the x-axis at a height of $1/v_{max}$ will intersect the other line at $i = -K_i$ (71). It must be stressed that this is valid for competitive inhibitors only.

K_i is taken to be a measure of the dissociation constant of the enzyme-inhibitor complex. Because K_i is an equilibrium constant, this value can be used to get a value for the free energy (ΔG) of binding of the inhibitor to the enzyme from the relation $\Delta G = -RT\ln K_i$. This fact also forms the basis for the extra thermodynamic approach to the study of structure activity relationships. In most cases, a biological activity (BA) such as K_i, that has a basis, can be represented as $BA = f(\Delta H, \Delta S, \Delta E)$. In particular, for K_i it is assumed that ΔG can be factored into polar electronic and steric parameters $\Delta G_x = \delta_x \Delta G_{hydrophobic} + \delta_x \Delta G_{electronic} + \delta_x \Delta G_{steric}$. As will be seen later, the substituent parameters used to develop equations

describing changes in K_i are derived from changes in ΔG for a particular process (73). ΔH, ΔG and ΔE are thermodynamic state variables and can be represented by other state variables. These variables do not define the nature of the interaction, however, they do give general ideas as to the forces involved and are a starting point for more extensive investigation.

Concluding Remarks

The preceding sections have attempted to review some of the properties of the four enzymes under study related to: inhibitors, molecular titrants, activation mechanism and substrates, as well as, some of the kinetic features of the enzymes. Table 1 is a tabulation of some of the more important properties of these enzymes which can be used as a reference for the similarities and differences of these proteins.

Part II

Multiparameter Regression Analysis

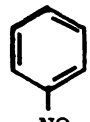
Introduction

The development of quantitative structure activity relationships (QSAR) for biochemical systems is of growing importance to the biochemist and medicinal chemist. The separation to electronic, steric and hydrophobic forces and their contributions to specific chemical interactions are important to the elucidation of biochemical mechanisms as well as to drug design. This section will deal first with a description of some of the parameters commonly used in QSAR. The second section will discuss the rationale of such analyses and some of the systems to which this approach has been applied.

Parameters in QSAR

Partition Coefficients, π

Simply stated, the partition coefficient P is the ratio of concentrations of a solute distributed between two separate phases, in this case $C_{octanol}/C_{water}$ (66). The value of P gives an indication of the relative hydrophobicity of a molecule and its derivatives. π is derived from the partition coefficient by the following relation: $\pi_X = \log P_X - \log P_H$ (67) where P_X is the partition coefficient of the substituted molecule and P_H of the unsubstituted molecule. π_X is therefore the value for a particular molecular fragment. Most π values remain relatively constant between one solute system and another provided there are no special intramolecular electronic or steric interactions between substituents and the parent molecule. The choice of the series from which the π values are chosen will depend on the

class of compounds being tested in a particular system. For example, are π values obtained from a nitrobenzene series or a phenylacetic acid series more representative of the π values for substituted benzamidines (e.g. should π_x for  be derived from  or )? In practice one usually

obtains π values for a few of the benzamidine derivatives to be tested, then following a comparison with a standard series, the rest of the π values can be taken from the appropriate standard reference (66).

Hammett σ Constants

σ constants are obtained for benzene substituents in a fashion analogous to that of obtaining π. σ can be defined by the relation: $\log K_x/K_H = \sigma\rho$ (68). σ values are calculated from a standard reaction series (i.e. ionization of benzoic acids) and ρ is assumed to equal 1.0. From this standard series the σ values calculated can then be applied to other reactions and correlations of their effects on reaction rates or equilibria can be determined. σ values are a measure of the electronic effect of the substituent on a particular reaction. The most common parameters are σ_m and σ_p, values for meta and para substituents, respectively. Other σ values based on different parent compounds, σ_I (69), σ^+ (70), σ_R (69), σ^- appear frequently in the literature. The slope ρ is a measure of the effect of σ on a particular reaction under consideration. The value of ρ_{obs} is obtained by plotting $\log K_x$ vs σ. A positive slope indicates that in the transition state the reaction center is electron rich, whereas a negative slope indicates an electron deficient center relative to the initial state.

Swain and Lupton F and R (72)

Since the development of the Hammett relationship and the application of σ constants to different sets of reactions, more than 20 σ sets have been developed to explain and correct deviations of correlations of reaction constants vs σ. Thus, the convenience of predicting reaction constants has been destroyed and casts serious doubts as to the validity of the Hammett equation. In order to reduce the need for so many σ sets and to restore the general nature of the equation, two constants F and R have been developed. F and R are a factorization of σ constants into pure field and pure resonance effects. The relationship for any σ constant in terms of F and R can be expressed by $\sigma = fF + rR$, where f and r are empirical weighting factors which are independent of the substituent but depend on the series being investigated (σ_m, σ_p, σ_I, σ_R, etc.), and F and R are the field and resonance constants which differ for each substituent.

Nys and Rekker, f (73)

The hydrophobic fragmental constant, f, is a parameter which is obtained in a manner similar to π. The relationship used for the determination of f is $\log P(sx) = f(s) + f(x)$. The main difference between this relationship and the relationship for obtaining π, $\log P(SH) = \log P(sx) + \pi(x)$, is that f represents the contribution of a particular fragment to the total lipophilicity; whereas π represents the lipophilicity of a group X upon substitution for hydrogen in the present compound. For example:

Using π: $\log P(C_6H_5\text{-}CH_2\text{-}CH_2\text{-}C_6H_5) = \log P(C_6H_6) + 2\pi(CH_3) + \pi(C_6H_5)$

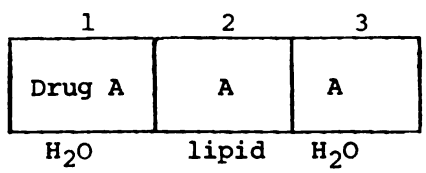
Using f: $\log P(C_6H_5\text{-}CH_2\text{-}CH_2\text{-}C_6H_5) = 2f(C_6H_5) + 2f(CH_2)$

Molar Refractivity M.R. (74)

Molar refractivity is a parameter based partially on the size of a molecule. It is dependent on the molecular weight of the compound of interest, which can also be used in regression analyses. Molar refractivity can be considered as a measure of the bulk steric effect of a molecule involved in intermolecular reactions in contrast to E_s which is more of an intramolecular steric effect.

Use of the Hansch Technique

The Hammett equation has been shown to be an important relationship in predicting the substituent effects on rates and equilibria in organic reactions (75). However, for biological systems, although limited success has been seen with the Hammett equation, the need for other substituent parameters to explain variance in the data has become apparent. From the work of Meyer (76), Overton (77) and others (78,79), evidence has been provided that the relative hydrophobic nature of substituents is important. A linear combination of two constants to give an equation: (Eq. 1) $\log 1/BR = K_1\sigma + K_2\pi + K_3$; where BR is some biological response (I_{50}, inhibition constant K_I, LD_{50}, etc.) and K_1, K_2 and K_3 are constants obtained by least squares has resulted in better correlations than the use of σ constants alone (80). However, even this has not seemed to be enough. The hypothesis that a parabolic relationship existed between BR and π led to an important and generally useful equation: (Eq. 2) $\log 1/BR = K_1 \log P^2 + K_2 \log P + K_3\sigma + K_4$ (80). In retrospect, the form of equation 2 translates into the following scheme which has been carried out on a 3 compartment system by Hansch:



Scheme I

This scheme represents the movement of a Drug A from an aqueous compartment 1 into a lipid or other hydrophobic compartment such as a cell membrane or protein-binding site followed by passage back into the aqueous phase. If a drug is too hydrophobic it will never get into compartment 2 and if too lipophilic the drug will enter compartment 2 and become trapped there. Therefore, a balance of lipophilicity and hydrophilicity is important in obtaining maximally active compounds.

There are many examples where biological activity is related parabolically to π which one should expect according to Scheme I (81,82,83,84,85). However, what one logically expects to find in biological systems is not always true and there are also hundreds of examples where a linear relation in π is sufficient to explain the variance in the data and the addition of π^2 does not improve the correlation (81).

Multiparameter regression analyses are not difficult to perform provided one has access to a large computer and the appropriate programs. A strict mathematical treatment of the solution of such equations is beyond the scope of this review and the reader is referred to a book by Daniel and Wood (86). The basic technique is based on the familiar linear regression analysis where equations of the form $y = mx + b$ are fitted. The theory is the same in that the coefficients of the independent variables are found by minimizing the sum of the squares of the deviations, the differences between each observed value y_j and the fitted value Y_j.

LIST OF REFERENCES

LIST OF REFERENCES

1. Jansen, E. F., M. D. F. Nutting, R. Jang, and A. K. Balls. 1949. J. Biol. Chem. 179, 189.

2. Krieger, M., L. M. Kay, and R. M. Stroud. 1974. J. Mol. Biol. 83, 209.

3. Beeley, J. G., and H. Neurath. 1968. Biochemistry 7, 1239.

4. Johnson, D., and J. Travis. 1976. Anal. Biochem. 72, 573.

5. Roberts, R. C., and J. Travis. 1969. Biochemistry 8, 2884.

6. Mallory, P. A., and J. Travis. 1964. Biochemistry 12, 2847.

7. Neurath, H., and G. W. Schwert. 1950. Chem. Rev. 46, 69.

8. Baird, J. B., E. F. Curragh, and D. T. Elmore. 1965. Biochem. J. 96, 733.

9. Baines, N. J., J. B. Baird, and D. T. Elmore. 1964. Biochem. J. 90, 470.

10. Sanborn, B., and G. E. Hein. 1968. Biochemistry 7, 3616.

11. Walsh, K. A. 1970. Methods Enzymol. 19, 60.

12. Chase, T., and E. Shaw. 1967. Biochem. Biophys. Res. Comm. 29, 508.

13. Johnson, D., R. Pannell, and J. Travis. 1974. Biochem. Biophys. Res. Comm. 57, 584.

14. Pannell, R., D. Johnson, and J. Travis. 1974. Biochemistry 13, 5439.

15. Moroi, M., and M. Yamasaki. 1974. Biochem. Biophys. Acta 359, 130.

16. Gladner, J. A., and K. Laki. 1956. Arch. Biochem. Biophys. 62, 501.

17. Miller, K. D., and H. van Vunakis. 1956. J. Biol. Chem. 223, 227.

18. Seegers, W. H. 1973. Ser. Haemat. 6, 36.

19. Hixson, H. F., and A. H. Nishikawa. 1973. Arch. Biochem. Biophys. 154, 501.

20. Thompson, A. R., and E. W. Davie. 1971. Biochem. Biophys. Acta 250, 210.

21. Magnusson, S. 1971. The Enzymes, Vol. III (Boyer, P. D., Ed.), Academic Press, NY, p. 278.

22. Baird, J. B., and D. T. Elmore. 1968. FEBS Letters 1, 343.

23. Shaw, E., W. Mares-Guia, and W. Cohen. 1965. Biochemistry 4, 2219.

24. Bailey, K., and F. R. Bettelhein. 1955. Biochem. Biophys. Acta 18, 495.

25. Cohen, I., Z. Bohak, A. de Vries, and W. Katchalski. 1969. Europ. J. Biochem. 10, 388.

26. Mutt, V., S. Magnusson, J. E. Jorpes, and E. Dahl. 1965. Biochemistry 4, 2358.

27. Mutt, V., and J. E. Jorpes. 1968. Europ. J. Biochem. 6, 156.

28. Curragh, E. F., and D. T. Elmore. 1964. Biochem. J. 93, 163.

29. Kezdy, F. J., L. Lorand, and K. D. Miller. 1965. Biochemistry 4, 2302.

30. Chase, T., and E. Shaw. 1969. Biochemistry 8, 2212.

31. Markwardt, F., H. Landmann, and P. Walsmann. 1968. Europ. J. Biochem. 6, 502.

32. Groskopf, W. R., L. Summaria, and K. C. Robbins. 1969. J. Biol. Chem. 244, 3590.

33. Summaria, L., B. Hsieh, W. R. Groskopf, K. C. Robbins, and G. H. Barlow. 1967. J. Biol. Chem. 242, 5046.

34. Groskopf, W. R., B. Hsieh, L. Summaria, and K. C. Robbins. 1969. J. Biol. Chem. 244, 359.

35. Deutsch, D. G., and E. T. Mertz. 1970. Fed. Proc. 29, 647.

36. White, W. F., G. H. Barlow, and M. M. Mozan. 1966. Biochemistry 5, 2160.

37. Barlow, G. H., L. Summaria, and K. C. Robbins. 1969. J. Biol. Chem. 249, 1138.

38. Robbins, K. C., L. Summaria, B. Hsieh, and R. J. Shah. 1967. J. Biol. Chem. 242, 2333.

39. Sherry, S., N. Alkjaersig, and A. P. Fletcher. 1966. Thromb. Diath. Haemorrh. 16, 18.

40. Blomback, B., M. Blomback, B. Hessel, and S. Iwanaga. 1967. Nature 215, 1445.

41. Iwanaga, S., P. Wallen, N. T. Grondahl, A. Henschen, and B. Blomback. 1967. Biochem. Biophys. Acta 147, 606.

42. Markwardt, F., P. Walsmann, and H. G. Kazmirowski. 1969. Pharmazie 24, 400.

43. Markwardt, F., P. Walsmann, and H. Landmann. 1970. Pharmazie 25, 551.

44. Lepow, I. H., L. Wurz, O. D. Ratnoff, and L. Pillemer. 1964. J. Immunol. 73, 146.

45. Lepow, I. H., and L. Pillemer. 1955. J. Immunol. 75, 63.

46. Lepow, I. H., O. D. Ratnoff, and L. Pillemer. 1956. Proc. Soc. Exp. Biol. Med. 92, 111.

47. Lepow, I. H., G. B. Naff, E. W. Todd, J. Penskey, and C. F. Hinz. 1963. J. Exp. Med. 117, 983.

48. Ratnoff, O. D., and G. B. Naff. 1967. J. Exp. Med. 125, 337.

49. Patrick, R. A., S. B. Taubman, and I. H. Lepow. 1970. Immunochemistry 7, 217.

50. Budzko, D. B., and H. J. Muller-Eberhard. 1971. Immunochemistry 7, 227.

51. Stroud, R. M., K. F. Austen, and M. M. Mayer. 1965. Immunochemistry 2, 219.

52. Stroud, R. M., M. M. Mayer, J. A. Miller, and A. T. McKenzie. 1966. Immunochemistry 3, 163.

53. Muller-Eberhard, H. J. 1969. Adv. Immunol. 8, 1.

54. Haines, A. L., and I. H. Lepow. 1964. J. Immunol. 92, 456.

55. Assimeh, S. N., D. H. Bing, and R. Painter. 1974. J. Immunol. 113, 225.

56. Baker, B. R., and M. Cory. 1971. J. Med. Chem. 14, 805.

57. Bing, D. H., M. Cory, and M. Doll. 1974. J. Immunol. 113, 584.

58. Bing, D. H. 1969. Biochemistry 8, 4503.

59. Westfall, S., G. H. Wirtz, and J. Canady. 1972. Immunochemistry 9, 1221.

60. Roman, D. P., J. A. Andrews, and D. H. Bing, this work.

61. Bing, D. H., personal communication.

62. Bing, D. H., J. L. Mernitz, and S. E. Spurlock. 1972. Biochemistry 11, 4264.

63. Sakai, K., and R. M. Stroud. 1973. J. Immunol. 111, 296.

64. Barlas, T., G. K. Scott, and J. E. Fothergill. 1973. Biochem. Soc. Trans. 1, 1219.

65. Fothergill, J. E. 1974. Workshop on Complement - Abstract - 2nd International Congress of Immunology.

66. Leo, A., C. Hansch, and D. Elkins. 1971. Chem. Rev. 71, 525.

67. Fujita, T., J. Isawa, and C. Hansch. 1964. J. Am. Chem. Soc. 86, 5175.

68. Hammett, L. P. 1970. Physical Organic Chemistry, 2nd Edition, McGraw Hill, NY.

69. Taft, R. W., and I. C. Lewis. 1958. J. Am. Chem. Soc. 80, 2436.

70. Brown, H. C., and Y. Okamoto. 1958. J. Am. Chem. Soc. 80, 4979.

71. Biggs, A. I., and R. A. Robinson. 1961, 388.

72. Swain, C. G., and E. C. Lupton. 1968. J. Am. Chem. Soc. 90, 4328.

73. Nys, G. G., and R. F. Rekker. 1974. Europ. J. Med. Chem. 9, 361.

74. Moore, W. J. 1962. Physical Chemistry, 3rd Edition, Prentice Hall, NJ, p. 561.

75. Jaffe, H. H. 1953. Chem. Rev. 53, 191.

76. Meyer, H. 1899. Arch. Exp. Pathol. Pharmakol. 42, 109.

77. Overton, E. 1899. Vierteljahresschr. Naturforsch. Ges. Zeurich. 44, 88.

78. Meyer, K. H., and H. Hemmi. 1935. Biochem. Z. 277, 39.

79. Hansch, C. 1971. Drug Design, Vol. I (Ariens, E. J., Ed.), Academic Press, NY, p. 271.

80. Hansch, C. 1969. Accounts Chem. Res. 2, 232.

81. Hansch, C. 1972. Adv. Chem. Ser. 114, 20.

82. Hansch, C., A. R. Steward, S. M. Anderson, and D. Bentley. 1968. J. Med. Chem. 11, 1.

83. Penniston, J. T., L. Beckett, D. L. Bentley, and C. Hansch. 1969. Mol. Pharmacol. 5, 333.

84. McFarland, J. W. 1970. J. Med. Chem. 13, 1192.

85. Higuchi, T., and S. S. Davis. 1970. J. Pharm. Sci. 59, 1376.

86. Daniel, C., F. S. Wood. 1971. Fitting Equations to Data, Wiley-Interscience, NY.

APPENDIX A

APPENDIX A

General Theory of Substituent Constants and Multiparameter Regression Analyses

In Part II of the Literature Review I dealt with a qualitative description of the substituent constants that were used in this study and the method employed for the calculation of the correlation equations. I am now going to present a somewhat more quantitative approach that should help to explain where these constants are derived and to explain why these variables can be used to explore the nature of the inhibition of serine proteases. The parameters to be discussed will be π, the hydrophobic parameter and the electronic parameters σ, F, and R.

The rationale behind the correlation of K_i with substituent constants is based on the fact that interactions in biological systems are subject to the same laws of thermodynamics which govern simple organic reactions. The substituent constants used in this work are derived from simple organic systems and reflect the effect of substituents on equilibrium constants. The inhibition constants, K_i, obtained for each of the benzamidine derivatives studied is defined as the dissociation constant for the enzyme-inhibitor complex. Therefore, $1/K_i$ is the association constant for the same complex. K_i is an equilibrium constant and is related to ΔG by the thermodynamic relationship $\Delta G = -RT\ln K_i$. While the interactions which effect ΔG for biological systems are more complex than those of simple organic systems, these interactions are chemical interactions. A substituent on benzamidine

affects the nature of its chemical interaction whether it is a simple organic or a complex biological equilibrium.

The Hydrophobic Parameter, π

The parameter π, as with the case of σ is best examined in extra thermodynamic terms. Consider the following equilibria for the reaction of movement of a solute between two solvent systems where H and L stand for hydrophilic and lipophilic solvents, respectively.

$$X - \overset{H}{\phi} - H \qquad H - \overset{L}{\phi} - H \qquad (1)$$

$$X - \overset{H}{\phi} - H \qquad X - \overset{L}{\phi} - H \qquad (2)$$

The free energy change associated with these reactions will be for (1)

$$\Delta G = 2.3RT \log \frac{(H\text{-}\phi\text{-}H)^{L}}{(H\text{-}\phi\text{-}H)^{H}} = 2.3RT \log P_H \qquad (3)$$

for (2)

$$\Delta G = 2.3RT \log \frac{(X\text{-}\phi\text{-}H)^{L}}{(X\text{-}\phi\text{-}H)^{H}} = 2.3RT \log P_X \qquad (4)$$

where $\frac{(H\text{-}\phi\text{-}H)^{L}}{(H\text{-}\phi\text{-}H)^{H}}$ and $\frac{(X\text{-}\phi\text{-}H)^{L}}{(X\text{-}\phi\text{-}H)^{H}}$ are the partition coefficients P_H and P_X, respectively. It should be noted that P_H and P_X are not the thermodynamic partition coefficients since those are the ratios of mole fractions $(X_a{}^{L}/X_a{}^{H}) = P'$, however, the ratio of concentration is more more commonly used and is therefore used to develop π values.

Now subtracting Eq. 3 from Eq. 4 gives

$$\Delta G_X - \Delta G_H = 2.3RT \log P_X/P_H \qquad (5)$$

$$\frac{\Delta G_X - \Delta G_H}{2.3RT} = \log P_X - \log P_H \qquad (6)$$

is then defined as $\log P_X - \log P_H$ and is a measure of the substituent-contribution to the difference of the change in free energy in the two molecules as they are partitioned. From thermodynamics it is known that

free energy changes are additive and this combined with observation leads to the fact that if polysubstituted molecules are used the value of π_{total} for a molecule will be $\pi_{total} = \Sigma\pi_i$. This is strictly only for an ideal case; however, this relation is often times true in practice.

Hammett, σ, and Swain and Lupton, F and R

The derivation of σ is very similar to that of π. It is based on differences in the free energies of reaction of unsubstituted and substituted compounds.

$$\Delta G_X = 2.303RT \log K_X$$

$$\Delta G_H = 2.303RT \log K_H$$

Therefore

$$\frac{\Delta G_X - \Delta G_H}{2.303} = \log K_X/K_H$$

where σ is the substituent constant and ρ is the reaction constant, a measure of the sensitivity of the equilibrium or reaction to changes in the substituent. The σ values defined by Hammett represent the effect of each substituent on the ionization of benzoic acids under standard conditions. The reaction constant, ρ, is set equal to 1.00 and $\sigma = \log K_X/K_H$. The determination of any two rate or equilibrium constants can then be used to determine ρ and the rates of reaction of other substituted compounds can be predicted. This relationship is also very valuable in determining the nature of the electronic characteristics of the transition state. A consideration of a plot of log K vs. σ shows us that if ρ is positive the rate will be increased by electron withdrawing substituents and when ρ is negative the rate will increase with electron donating substituents. A number of other substituent constants have been derived to explain deviations which occur by utilizing the original Hammett constants. In an effort to restore the

generality of using substituent constants to explain variations in rate data and to try to combine all of the various types of constants seen into one general set of constants, Swain and Lupton devised the parameters F and R.

The parameters, F and R, are defined to be measures of the field and resonance potentials of a substituent. The calculation of F and R rests on the assumption that any σ set can be represented as a linear combination of two other σ_i sets and that one can ignore any factors other than field and resonance for a substituent kept three or more atoms distant from the reaction center. This leads to the equation

$$\sigma_i = fF + rR$$

where f and r are weighting factors dependent on reaction series and F and R are the field and resonance terms dependent on the substituent.

For the calculation of F values the parameter σ' was used initially and the correlation of $\sigma' = a\sigma_m + b\sigma_p$ was done. The F values (which are calculated values of σ') could then be determined from $F = a\sigma_m + b\sigma_p$. The term σ' was used as a measure of pure field effect since it is based on the ionization of 4-substituted bicyclo [2.2.2] octane-1-carboxylic acids. In this series, there is no resonance because there is no intervening conjugation or unsaturation between the substituent and the reaction center.

To obtain R values the equation $\sigma_p = \alpha F + R$ was used. Before the application of the above equation, field effect may be removed from σ_p by assuming that for the trimethylammonium ion substituent R = 0.00. The value α, the weighting factor for field effects, is readily calculated [σ_p(trimethylammonium ion) = αF] and from this the rest of the R values may be obtained.

Multiple Parameter Regression Analysis (31)

The following section will deal with the theory and derivation of general equations for the least squares technique. In least squares analysis the coefficients of the equation are obtained by minimizing the sum of squares of the differences between the observed value y_j and the expected value, Y_j. The following least squares solution will be for three independent variables.

The estimated equation is

$$Y = b_o + b_1x_1 + b_2x_2 + b_3x_3 \qquad (1)$$

Since b_o is always of the form

$$b_o = \bar{y} - \sum_{i=1}^{K} b_i\bar{x}_i; \text{ for K constants} \qquad (2)$$

We need only find the values of b_1, b_2 and b_3 since upon substitution of (2) into (1) we get (3).

$$Y - \bar{y} = b_1(x_1-\bar{x}_1) + b_2(x_2-\bar{x}_2) + b_3(x_3-\bar{x}_3) \qquad (3)$$

The solution for the coefficients are obtained by minimizing the sum of squares of the residuals

$$[rr] \equiv \sum_{j=1}^{N} (y_j-Y_j)^2 \qquad (4)$$

y_j is observed value
Y_j is fitted value

Substituting for Y_j (from (3))

$$[rr] = \sum_{j=1}^{N} (y_j-\bar{y}) - b_1(x_{1j}-\bar{x}_1) - b_2(x_{2j}-\bar{x}_2) - b_3(x_{3j}-\bar{x}_3)^2 \qquad (5)$$

Now differentiating (5) with respect to each b_i and setting equal to zero leads to the following sets of equations:

$$\begin{aligned} b_1[11] + b_2[12] + b_3[13] &= [1y] \\ b_2[21] + b_2[22] + b_3[23] &= [2y] \\ b_3[31] + b_2[32] + b_3[33] &= [3y] \end{aligned} \qquad (6)$$

where

$$[n] \equiv \Sigma(x_{1j}-\bar{x}_1)^2;$$

$$[21] \equiv [12] \equiv \Sigma(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2);$$

$$[1y] \equiv \Sigma(x_{1j}-\bar{x}_1)(y_j-\bar{y}); \text{ etc.}$$

to illustrate this solution

$$\frac{drr}{db_1} = 2\Sigma[(y_j-\bar{y}) - b_1(x_{1j}-\bar{x}_1) - b_2(x_{2j}-\bar{x}_2) - b_3(x_{3j}-\bar{x}_3)](x_{1j}-\bar{x}_1)$$

where

$$(x_{1j}-\bar{x}_1) = \frac{db_1(x_{1j}-\bar{x}_1)}{db_1}$$

Continuation of this process leads to the equations (6). All that is left is to solve the simultaneous linear equations which may be done by the method of determinants.

Quantitative Structure Activity Relationships: Inhibition of Human $C1\bar{s}$, Thrombin, Trypsin and Plasmin by Substituted Benzamidines

by

Daniel P. Roman, Jr. and David H. Bing

(Manuscript submitted to Biochemistry)

ABSTRACT

The inhibition constants for a series of substituted benzamidines were measured and correlated with various substituent parameters to try to determine some of the factors that control the magnitude of K_i. The most important single variable involved in the correlations for trypsin, thrombin and $Cl\bar{s}$ is the Swain and Lupton resonance parameter, resulting in equations of the form $1/K_i = k_1R + k_2$ where k_1 and k_2 are calculated constants. For $Cl\bar{s}$ the inclusion of πNO_2 and $(\pi NO_2)_2$ measures of substituent hydrophobicity derived from the nitrobenzene series, results in a significantly better correlation. In the case of plasmin, molar refractivity (MR) and benzene series were the most significant single variables. It does not appear that any electronic effect plays a significant role in benzamidine binding to plasmin, thus making it very different from the other three enzymes. A comparison between human trypsin and thrombin and their bovine counterparts indicates that differences exist, most probably in the hydrophobic regions of these enzymes. The equations generated for human plasmin in this work and in the work by Coats (1973), are similar, indicating that a mass synthesis approach to inhibition design is unnecessary.

$C1\bar{s}$, trypsin, plasmin and thrombin are all serine proteases, and all appear to possess an anionic binding site (Perlmann and Lorand, 1970; Muller-Eberhard, 1968). $C1\bar{s}$, a subunit of the first component of complement, C1, is the enzyme responsible for the hydrolysis of C4 and C2, two early acting components in the complement sequence. Complement is a group of eleven serum proteins which act as biological amplifiers in the immune response. Thrombin (EC 3.4.4.13) is a proteolytic enzyme whose main physiological activity is the conversion of fibrinogen to fibrin which can then react to form a fibrin clot. Plasmin (EC 3.4.4.14) is another serum protease whose primary function is the dissolution of the fibrin clot. Plasmin has also been shown to activate $C1\bar{s}$ (Ratnoff and Naff, 1967), and to catalyze formation of prekallikrein activating fragments from activated Hageman factor (Austen, 1971). Trypsin (EC 3.4.4.4), one of the most studied of the serine proteases, is a pancreatic digestive enzyme, which is responsible for the activation of the other digestive enzymes in the pancreas. These enzymes are known to hydrolyze basic and certain aromatic amino acid esters, in particular, N^{α}-Z-L-Tyr-Np, N^{α}-Z-L-Lys-p-Np and N^{α}-Z-L-Lys-p-Np (Kezdy et al., 1965; Silverstein, 1973; Bender et al., 1965; Martin et al., 1958).

Extensive data exists on the inhibition of the bovine enzymes, thrombin, trypsin and human plasmin by substituted benzamidines (Markwardt et al., 1968, 1969, 1970). Baker and his coworkers compiled

---

*The terminology used for the complement proteins is that suggested in the Bull. Wld. Hlth. Org. "Nomenclature of Complement," Immunochemistry 7, 137, 1970. Other abbreviations used in this paper are: N^{α}-carbobenzoxy-L-Tyrosine-p-nitrophenyl ester, N^{α}-Z-L-Tyr-p-Np; N^{α}-carbobenzoxy-L-Lysine-p-nitrophenyl ester, N^{α}-Z-L-Lys-p-Np.

considerable data on the inhibition of whole guinea pig complement by substituted benzamidines (Baker and Erickson, 1969; Baker and Cory, 1969a, 1969b, 1971a, 1971b). Coats (1973) has reported multiple parameter regression analyses of these data, and has compared trypsin, thrombin and plasmin suggesting that differences in the anionic binding sites of these enzymes do exist. Hansch et al. (1974) have developed quantitative structure activity relationships (QSAR) from the data derived by Baker et al. (1969a, 1969b, 1971a, 1971b) and a fairly definitive direction is indicated in making more pharmacologically active compounds. The following study was undertaken to examine the inhibition of these enzymes by a series of substituted benzamidines and to relate the inhibition constants to the physical-chemical properties of the substituents.

MATERIALS AND METHODS

Organic Compounds. N^{α}-Z-L-Tyr-p-Np was obtained from Nutritional Biochemical Corp. N^{α}-Z-L-Lys-p-Np was prepared from N^{α}-carbobenzoxy-N^{ϵ}-carbo-t-butyloxy-L-lysine-p-nitrophenyl ester (Sigma Chemical Co.) according to previously described procedures (Bender et al., 1965a). All of the above esters were determined to be greater than 99% homogeneous as indicated by the presence of less than 1% p-nitrophenol in 10^{-3} M solutions of these compounds at pH 8.1, 410 nm. Benzamidine, 3-aminobenzamidine and 3-nitrobenzamidine were purchased from Aldrich Chemical Co., 4-Nitrobenzamidine was prepared according to previously described procedures (Pinner and Gradenwitz, 1897). The remainder of the benzamidine compounds were the generous gift of Dr. Michael Cory, SRI, Menlo Park, CA, and the preparation of the compounds has been described (Cory, 1971).

Enzymes. Highly purified human $C1\bar{s}$ was prepared by affinity chromatography by the method of Assimeh et al. (1974). Human plasmin and human thrombin were the gift of Dr. Robert Rosenberg, Beth Israel Hospital, Boston, MA. Plasmin was purified on lysine-Sepharose as previously described (Deustsch and Mertz, 1970). Thrombin was purified by the method of Rosenberg et al. (1970). Human trypsin was kindly donated by Dr. James Travis, Department of Biochemistry, University of Georgia, Athens, GA. This protein was purified from human pancreas by affinity chromatography of Trasylol^{R}-Sepharose (Johnson and Travis, 1975).

Protein concentrations were measured by the method of Lowry et al. (1951), and bovine serum albumin used to determine a standard curve.

Enzymatic Analysis. All readings were made on a Gilford Model 240N spectrophotometer. The esterolytic activity of the samples was determined using N^{α}-Z-L-Tyr-p-Np as described by Bing (1969). A stock solution of 10^{-3} M N^{α}-Z-L-Tyr-p-Np was made in acetone and 0 to 50 µl of the solution was added to a solution of .005 ionic strength Tris-HCl containing 0.09 M NaCl, pH 8.1 containing the protein to be studied. The final volume of the assay solution was 1.0 ml. The production of p-nitrophenol was measured at 410 nm and recorded continuously on a Heath SR255B recorder. For the assay using N^{α}-Z-L-Lys-p-Np, a stock solution of 10^{-3} M N^{α}-Z-L-Lys-p-Np was made in 50% acetone-H_2O and 10 to 50 µl of this solution was added to a 1.0 ml sample of protein in 0.02 ionic strength acetate solution pH 6.0 containing 0.085 M NaCl. For the human trypsin assay with both substrates, 50 mM $CaCl_2$ was included in the assay buffer. Production of p-nitrophenol was measured at 340 nm and recorded continuously on a Heath Model SR255B recorder. A blank with no protein was run with each assay to correct for the rate of spontaneous hydrolysis of N^{α}-Z-L-Tyr-p-Np. The rate of spontaneous hydrolysis was less than 1% of the enzyme catalyzed hydrolysis of N^{α}-Z-L-Lys-p-Np. All analyses were carried out between 25 and 26°C. The molar extinction coefficient of p-nitrophenol at 340 nm, pH 6.0 was determined to be 5.85×10^3.

Inhibition was measured at a substrate concentration of 3×10^{-4} with the inhibitor concentration adjusted to give approximately 10 to 80% inhibition. Plots of 1/v vs. inhibitor concentration were made. K_i values were calculated using the slopes of these plots and previously

determined V_{max} (Dixon, 1953). K_i is assumed to be an estimate of the dissociation constant for the enzyme-inhibitor complex. Competitive inhibition was shown by the methods of Cornish-Bowden (1974) and Dixon (1953). Plots were made of s/v vs. inhibitor concentration at three different substrate concentrations, and 1/v vs. inhibitor concentration at two substrate concentrations. Plots of s/v vs. I for the four enzymes are linear indicating competitive inhibition.

In order to investigate the nature of the factors responsible for the magnitude of K_i exerted by some substituent X, the following general equation was used:

$$\log 1/K_i = k + k'x_1 + k''x_2 + k'''x_3$$

where k, k', k'', and k''' are constants which are fixed for a given system and derived from multiparameter regression analysis and x_1, x_2 and x_3 are various substituent parameters. Table I is a list of the parameters that were used in the biological correlations. The QSAR were done on the SRI CDC 3600 computer with a multiparameter regression analysis program.

Table 1. Substituent Constants Used in Quantitative Structure Activity Relationships

---

σ, σ_m, σ_p - Electronic Parameters - Hammett Constants[a]

Swain and Lupton Parameters[b]
F - Field Effect
R - Resonance Effect

π, π_{NO_2} - Hydrophobicity (Benzene, Nitrobenzene Series, respectively)[c]

F - Hydrophobic Fragmental Constant developed by Nys and Rekker[d]

M.W. - Molecular Weight

M.R. = P_E - Molar Refractivity, Polarizability[e]

---

[a]Hammett (1970)

[b]Swain and Lupton (1968)

[c]Leo et al. (1971)

[d]Nys and Rekker (1974)

[e]Hansch et al. (1974)

RESULTS AND DISCUSSION

Reaction of Human Trypsin, Plasmin, Thrombin and $C1\bar{s}$ with N^{α}-Z-L-Tyr-p-Np. The rate of the reaction of the four enzymes with 3×10^{-4} M N^{α}-Z-L-Tyr-p-Np was proportional to enzyme concentration. The reaction was linear for at least three to five minutes depending on the amount of enzyme used. The deviation from linearity beyond five minutes was probably due to a buildup of product and competitive inhibition between N^{α}-Z-L-Tyr and N^{α}-Z-L-Tyr-p-Np. The kinetic parameters for the reaction of $C1\bar{s}$, trypsin, plasmin and thrombin are shown in Table II. Figure 1 shows the Lineweaver-Burk plots of the reaction of each of the enzymes with N^{α}-Z-L-Tyr-p-Np.

Reaction of Human $C1\bar{s}$, Plasmin, Trypsin and Thrombin with N^{α}-Z-L-Lys-p-Np. As in the case of N^{α}-Z-L-Tyr-p-Np, the reaction of the four enzymes with 3×10^{-4} M N^{α}-Z-L-Lys-p-Np was linear with enzyme concentration. The assay of these enzymes was attempted at pH 8.1 in order to get a better comparison of the relative affinities and k_{cat} values of $C1\bar{s}$, thrombin, trypsin and plasmin for N^{α}-Z-L-Tyr-p-Np and N^{α}-Z-L-Lys-p-Np. However, at pH 8.1 the lysine ester is very unstable and the spontaneous rate of hydrolysis was almost equal to hydrolysis in the presence of enzyme. At pH 6.0 the enzymatic rate of hydrolysis was linear for five to six minutes and the spontaneous rate of hydrolysis of N^{α}-Z-L-Lys-p-Np was negligible. The values for $K_{m(app)}$,

Table 2. Kinetic Parameters for Reaction of Human $C1\bar{s}$, Tyrpsin, Plasmin, and Thrombin with N^{α}-Z-L-Tyr-p-Np and N^{α}-Z-L-Lys-p-Np

	N^{α}-Z-L-Tyr-p-Np			N^{α}-Z-L-Lys-p-Np		
Enzyme	$K_{m(app)}$ ($\times 10^5$ M)	V_{max} ($\times 10^6$ M/min/mg	k_{cat} (sec^{-1})	$K_{m(app)}$ ($\times 10^5$ M)	V_{max} ($\times 10^6$ M/min/mg)	k_{cat} (sec^{-1})
Trypsin	.933	97.3	37.3	3.09	25.1	96.2
Plasmin	8.43	.388	.486	2.0[a]	—	40[a]
Thrombin	1.39	6.40	3.83	2.1[b]	—	.72[b]
$C1\bar{s}$	14.01	1.5	3.76	1.35	.35	.60

[a]Silverstein (1973)

[b]Kezdy *et al.* (1965)

Figure 1. Lineweaver-Burk plot of the reaction of N^{α}-Z-L-Tyr-p-Np with human $C1\bar{s}$, trypsin, plasmin and thrombin. Assays were performed as described in Methods. Ordinate, reciprocal of initial rate in min $mmole^{-1}$; abscissa, reciprocal of concentration of substrate 1. $mole^{-1}$.

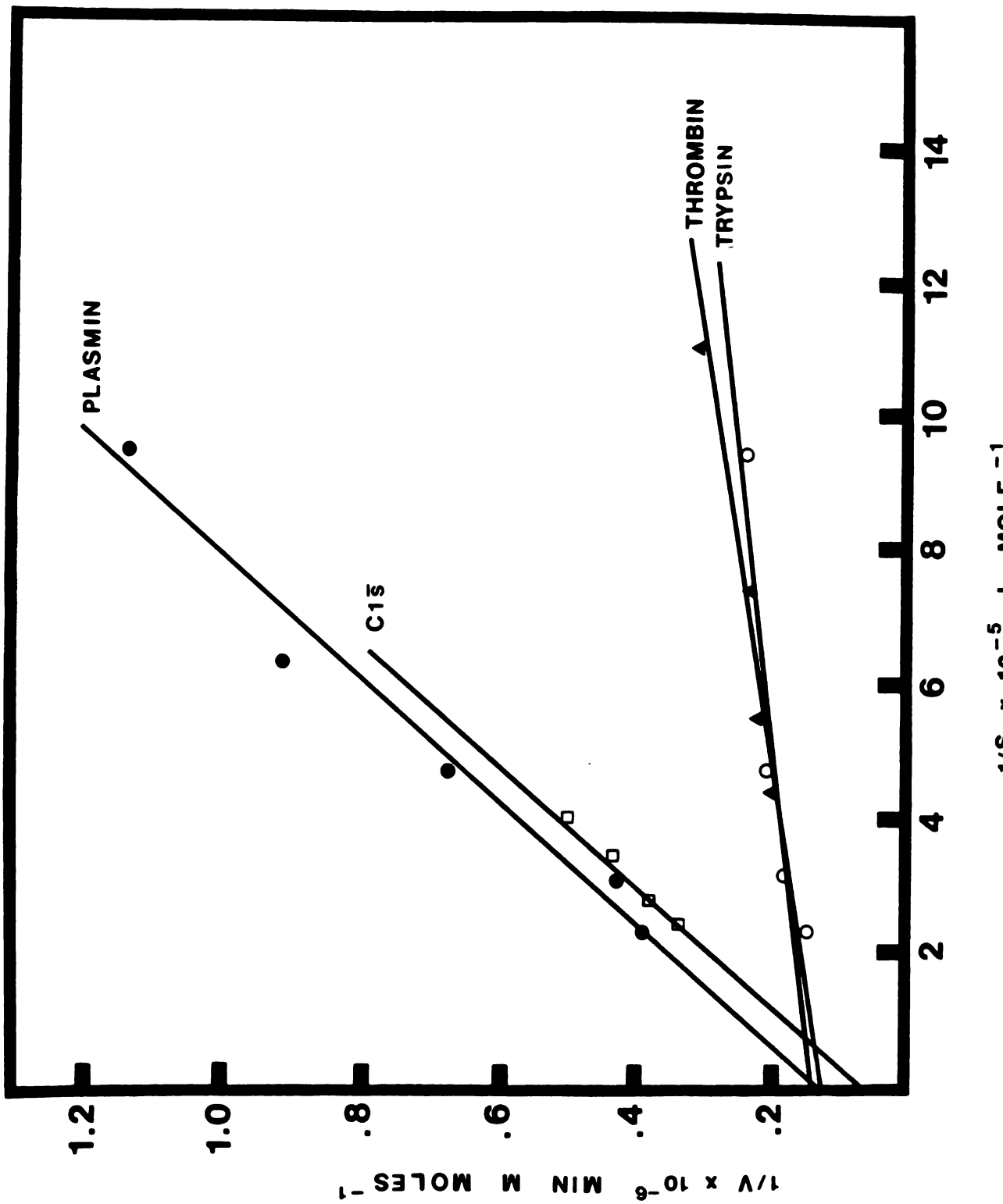

PLASMIN
C1s̄
THROMBIN
TRYPSIN
1.2
1.0
.8
.6
.4
.2
2
4
6
8
10
12
14
1/V x 10⁻⁶ MIN M MOLES⁻¹
1/S x 10⁻⁵ L MOLE⁻¹

Figure 1.

V_{max} and k_{cat} for $C1\bar{s}$ and trypsin and $K_{m(app)}$ and k_{cat} for thrombin (Kezdy et al., 1965) and plasmin (Silverstein, 1973) are shown in Table II. Figure 2 contains plots for trypsin and $C1\bar{s}$ for their reaction with N^{α}-Z-L-Lys-p-Np.

The hydrolysis of some synthetic neutral and cationic substrates by bovine trypsin (Bender et al., 1965; Martin et al., 1958), human plasmin (Silverstein, 1973), thrombin (Kezdy et al., 1965) and $C1\bar{s}$ (Bing, 1969) have been reported. This part of the study was designed to accumulate $K_{m(app)}$, V_{max} and k_{cat} values for the hydrolysis of N^{α}-Z-L-Tyr-p-Np and N^{α}-Z-L-Lys-p-Np by the human enzymes, trypsin, plasmin, thrombin and $C1\bar{s}$.

Table III contains the values of K_i obtained for each of the four enzymes. Table IV contains the values for π, π_{NO_2}, f, F, R and MR which were used in the correlations.

In the correlations obtained for all the enzymes, except plasmin, R, the Swain and Lupton resonance parameter (1968) was the most significant single variable. Only thrombin produced an equation that was significant at the .95 level of the F test with R alone. All the other enzymes were dependent on more than one term to give equations that are considered significant at the .90 level or better. The best single variable equation for plasmin was a correlation in MR. Correlations with π and D gave equations which were similar in their F test values. D, is a dummy variable used to determine if 4-position isomers had a special effect on the inhibition of the enzymes compared to 3-position isomers. D equals 1 for 4-position and zero for 3-position substituents. Table V is a summary of the most significant equations obtained for each enzyme.

Figure 2. Lineweaver-Burk plot of the reaction of N^{α}-Z-L-Lys-p-Np with human $C1\bar{s}$ and trypsin. Assays were performed as described in Methods. Ordinate, reciprocal of initial rate in min. $mmole^{-1}$; abscissa, reciprocal of concentration of substrate in l. $mole^{-1}$.

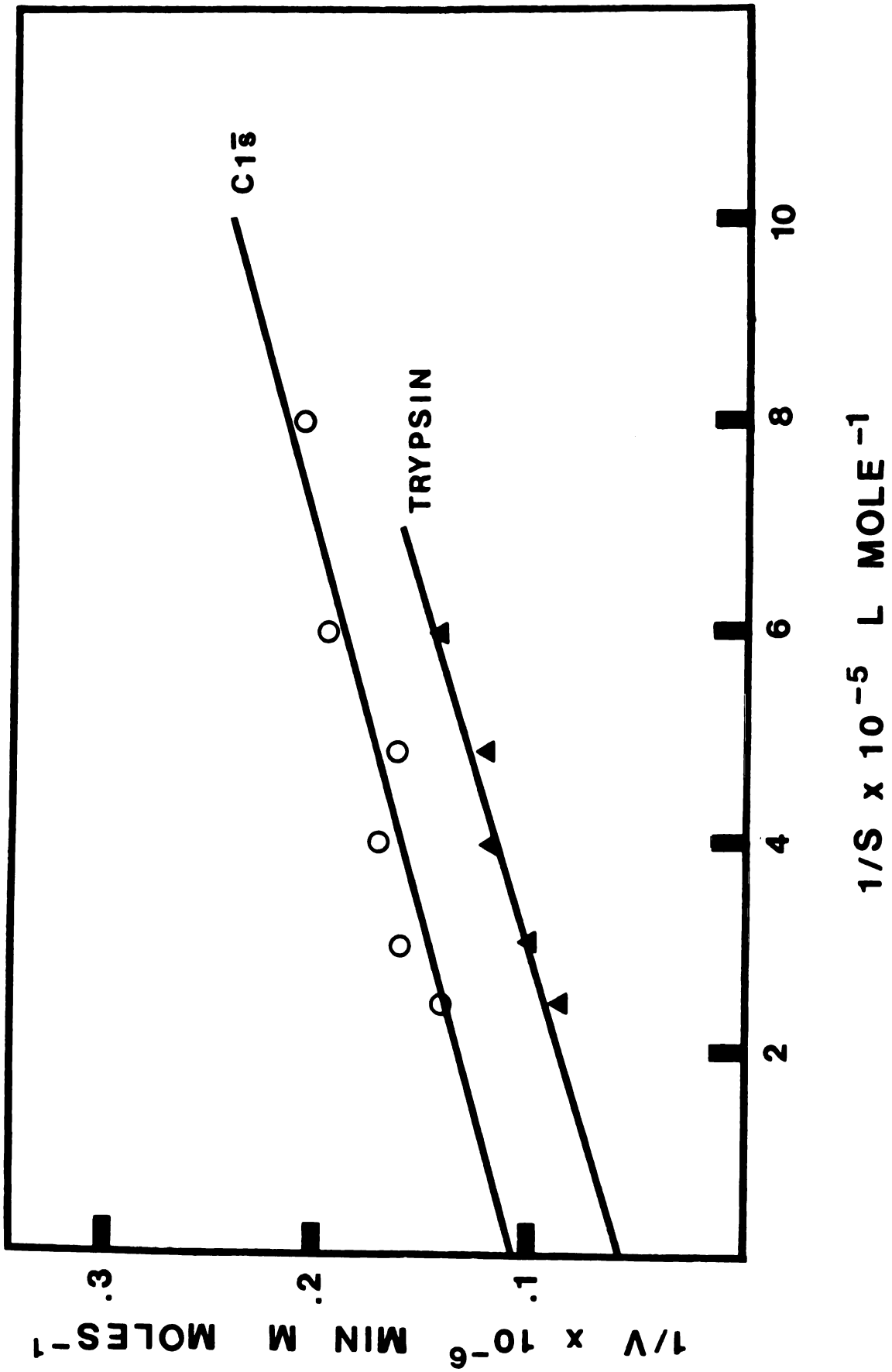


Figure 2.

Table 3. Substituents Constants for X

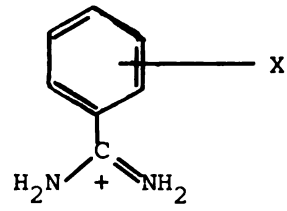
X	σ[a]	F[a]	R[a]	f[b]	M.R.[a]	π(Benzene)[c]	π(NO_2Benzene)[c]
H	0.00	0.00	0.00	0.00	.130	0.00	0.00
3-NH_2	-.16	.02	-.68	-.91	.542	-1.23	-.48
3-Br	.39	.44	-.17	1.17	.888	-.86	-.86
3-MeO	.12	.26	-.51	.244	.787	-.02	.30
3-Phenyl	.06	.08	-.08	1.9	2.54	1.96	1.96
3-Benzyl	-.08	-.08	-.01	2.42	3.00	2.01	2.30
3,4-diMe	-.24	-.08	-.26	1.40	1.13	1.12	1.01
3,5-diMe	-.07	-.04	-.13	1.40	1.13	1.12	1.15
4-NO_2	.78	.67	.16	-.09	.736	-.28	-.28
3-NO_2	.17	.67	.16	-.09	.736	-.28	-.28
2-Me	-.17	-.04	-.13	.70	.565	.56	.52
3-CF_3	.43	.38	.19	1.28	.502	.88	1.11
3-COOH	.37	.33	.15	-.40	.693	-.32	-.02
3-COOEt	.37	.33	.15	1.23	1.747	.54	.54
3-phenoxypropoxy	.10	.25	-.57	1.48	.437	2.23	2.23
4-MeO	-.27	.26	-.51	.24	.787	.18	.18
3-CH_2OH	0.00	0.00	0.00	.91	.719	.65	.65
4-CH_2OH	0.00	0.00	0.00	.91	.719	.60	.60

[a]Hansch *et al.* (1973)

[b]Nys and Rekker (1973, 1974)

[c]Leo *et al.* (1971)

Table 4. Benzamidine Inhibitors Tested for Inhibition



Log 1/Ki

X	Trypsin	Plasmin	Thrombin	$Cl\bar{s}$
H	5.11	3.28	3.14	3.21
3-NH_2	5.20	3.99	5.18	3.24
3-Br	N.T.	2.13	N.T.	4.01
3-MeO	4.87	3.53	3.42	3.77
3-Phenyl	5.36	3.52	2.00	3.64
3-Benzyl	5.20	3.86	3.15	3.19
3,4-di-Me	5.08	2.48	2.96	3.93
3,5-di-Me	4.83	3.57	4.17	4.09
4-NO_2	3.87	2.44	2.63	1.03
3-NO_2	4.10	2.95	2.69	2.79
2-Me	3.77	3.47	2.48	2.77
3-CF_3	3.87	4.38	2.34	2.30
3-COOH	2.32	3.37	N.T.	3.22
3-CO_2Et	3.97	N.T.	2.59	N.T.
3-phenoxypropoxy	4.97	4.67	4.28	4.95
4-MeO	4.06	N.T.	N.T.	No Inh.
3-CH_2OH	4.52	2.61	2.22	2.01
4-CH_2OH	4.08	3.17	1.97	No Inh.
4-OH	N.T.	No Inh.	No Inh.	No Inh.
4-EtO	N.T.	No Inh.	N.T.	No Inh.

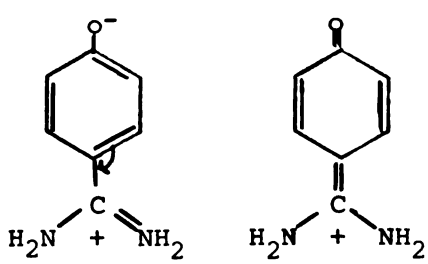
Table 5. Correlation Equations for Trypsin, Plasmin, Thrombin and $C1\bar{s}$ (All Compounds)

	n	s	r	F
Trypsin				
log 1/Ki = 4.24 - 1.33 R	17	.70	.50	1.25
= 4.54 - 1.39 σ	17	.70	.49	1.23
= 4.70 - 1.58 F	17	.70	.49	1.23
= 4.20 + .170$(\pi NO_2)^2$	17	.70	.49	1.23
= 4.32 - 1.23σ + .15$(\pi NO_2)^2$	17	.63	.65	1.44
= 4.46 - 1.26 F + .13$(\pi NO_2)^2$	17	.65	.62	1.34
= 4.05 - 1.23 R + .16$(\pi NO_2)^2$	17	.62	.67	1.51
$C1\bar{s}$				
log 1/Ki = 2.91 - 2.25 R	15	.78	.61	1.38
= 2.70 + .45π_{NO_2} - 2.1 R	15	.64	.78	2.07
= 2.75 + .97πNO_2 - 1.95 R - .33$(\pi NO_2)^2$	15	.60	.82	2.32
Plasmin				
log 1/Ki = 3.49 - .79 D	16	.65	.54	1.17
= 2.99 + 2.93 M.R.	16	.64	.47	1.20
= 3.16 + 3.00πNO_2	16	.64	.47	1.20
Thrombin				
log 1/Ki = 2.73 - 2.52 R	15	.62	.75	2.59

n, number of data prints
s, standard deviation from regression line
r, correlation coefficient
F, value F statistic

Thrombin was the only enzyme studied that was significantly correlated with only one variable. Addition of other terms did not significantly improve the equation in R alone. In fact, no other equation, single or multivariable, produced a more significant equation as judged by the F statistics.

The correlation indicated that inhibition of $Cl\bar{s}$ by benzamidines was dependent on both R and π_{NO_2}. R is the most important single parameter. A combination of π_{NO_2} and R was still insufficient to explain the variance of the data, however, the addition of $(\pi_{NO_2})^2$ resulted in an equation that was significant at the .95 level of the F test. The data on the inhibition of $Cl\bar{s}$ by 4-substituted benzamidines deserves additional comment. No inhibition of $Cl\bar{s}$ was observed for 4-OH, 4-OMe and 4-OEt. Only 4-NO_2 and 4-CH_2OH produced measurable levels of inhibition and the expected values of K_i for these compounds were much smaller than those observed. Since only these two K_i values for 4-substituted benzamidines were available, a correlation of D with log $1/K_i$ was not significant. Two explanations of the data are possible. First, interaction might affect the binding of the para compounds. An anomaly exists in the fact that the 3,4-dimethyl compound is a very good inhibitor. The other explanation is that the positive charge of the amidine is destroyed by resonance through the ring as shown:



Resonance destruction of the positive charge most probably accounts for the failure of 4-hydroxy benzamidine to inhibit $C1\bar{s}$ as well as plasmin and thrombin. The other 4-substituted benzamidines inhibit plasmin, trypsin and/or thrombin to varying degrees, but all 4-substituted compounds are poor inhibitors of $C1\bar{s}$. Therefore, we propose a steric interaction is interfering with the binding of these compounds to $C1\bar{s}$. Baker and Erickson (1969) have reported that 4-substituted benzamidines are not inhibitory against whole guinea pig complement. Further investigations with 4-halogen and 4-CF_3 substituted compounds could be done to determine the possible limits of substituent size in the para position of benzamidines for the inhibition of $C1\bar{s}$.

The trypsin data did not correlate well with any of the variables tested. The equations for trypsin given in Table IV were judged to be the most significant. They are presented with some reservation since none were judged to be significant at even the .90 level of the F test; however, some interesting points can be made. The most important single variable equation for trypsin is R. However, the equations in F and σ are very close in terms of the F test, and it is difficult to assess what kind of electronic interaction is influencing the inhibition. One more interesting equation is one that involves $(\pi_{NO_2})^2$. A plot of log $1/K_i$ vs. $(\pi_{NO_2})^2$ gives a parabolic curve in which log $1/K_i$ increases with the absolute value of π_{NO_2}. In the case of trypsin, the reason for this can possibly be explained based on the nature of enzyme specificity. Trypsin is a digestive enzyme which will hydrolyze virtually all x-lys-x' or x-arg-x' bonds; thus, the location for binding x must be able to accomodate a wide range of lipo- or hydrophobic groups. If the hydrophobic pocket that binds benzamidine is long

enough to accomodate the substituent, then the walls of the crevice may show different degrees of hydrophobicity. If one wall is hydrophobic it could interact better with a hydrophobic substituent thus giving an increase in binding of such molecules. If the other wall is relatively hydrophilic, polar groups could interact there just as readily. The differential hydrophobicity in the region directly to the rear of the benzamidine (relative to the active site serine) (Krieger et al., 1974) then explains the dependence of log $1/K_i$ on $(\pi_{NO_2})^2$, since then binding will be expected to increase with the absolute value of $\pi(NO_2)$.

Similarly, the equations generated for plasmin (see Table IV) were insufficient to explain the variance of the data. Examination of the equations tells us something about the binding procket of plasmin. First, the best equation was a function of MR. An equation as a function of π was similar to the MR equation in terms of the F statistic. This similarity is not unusual since MR and π display a good deal of collinearity. This means that their values are dependent on some of the same factors such as, molecular weight and electronic character, and that these values increase or decrease in a parallel manner. If these equations are indeed indicative of the nature of the binding site of plasmin and no electronic terms are involved, then we suggest that plasmin is different from the other three enzymes studied and K_i is dependent on only hydrophobic or molecular size parameters. The one final equation for plasmin is one that involves the dummy variable, D. This indicates that 4-position isomers affect the inhibition in a negative manner, as was the case with Cl^-.

It is clear that R is the single most important factor which can affect the magnitude of K_i in three of the four enzymes. Two questions

can be asked concerning the effect of R on log $1/K_i$. First, why should R be significant? The resonance interaction of a meta substituent with the amidine center will be minimal because classical resonance structures do not accomodate this interaction. The second question to be asked is why should log $1/K_i$ show a negative dependence on R, indicating that the better electron donors give the better benzamidine inhibitors. This increased density in the ring and the charged moiety will tend to delocalize the positive charge on the amidine group and the interaction of this charge by the anionic site in the enzyme will be less favorable. Since the proposed mechanism of inhibition involves formation of an ion pair between the cationic benzamidine and the anionic site, the delocalization of the positive charge should result in a decrease in inhibition. This is the opposite of what was observed.

The resonance effect and the negative dependence on R can be explained as follows. The mechanism for classical resonance from a meta substituent does not allow for a direction of the electrons to the amidine center but rather a resonance distribution of the electron pair around the benzene ring. The increased electron density in the benzene ring might allow for the formation of a π-π interaction between the enzyme and the benzamidine. Evidence for this type of interaction is found in bovine trypsin. East and Trowbridge (1968) report the observation of a difference spectrum on the binding of benzamidine to bovine trypsin indicating some perturbation of the $\pi \rightarrow \pi^*$ transition for the benzene ring. The formation of a π-π interaction between the benzamidine ring and the enzyme could elicit a secondary electronic effect on the amidine group by feeding electron density into the enzyme and thus withdrawing it from the amidine center. This would make the amidine more positive and increase the ionic bond strength.

The low correlations with all compounds for trypsin and plasmin and to a lesser extent $C1\bar{s}$ were somewhat distressing. However, it is our feeling that some other variable, perhaps one unknown at this time, is necessary to explain the data. For example, calculations of electron density around the ring or charge transfer through π-π interaction from the benzamidine ring and its substituents to the enzyme surface could be important. This is not unrealistic considering the nature of the dependence of the equations on R, and suggests that present variables are insufficient to explain dependence of structure on log $1/K_i$ except at a very rudimentary level.

In order to determine if significant equations could be obtained and if there might be some effect that we have not considered, compounds with large deviations between observed and calculated log $1/K_i$ were eliminated from the correlations. Table VI lists the various equations obtained for each of the enzymes and the compounds that were eliminated from the computation. The equations generated without these compounds gave better correlations and the significance of the equations also increased. Trypsin, thrombin and $C1\bar{s}$ gave equations with F test values of .975, .975 and .90, respectively in which the most significant variable was R. Plasmin remained distinct from the other enzymes in that the two best equations, significant at the .90 level, were functions of MR or π.

Coats (1973) and subsequently Hansch (1974) investigated quantitative structure activity relationships for bovine thrombin and trypsin, human plasmin and the whole guinea pig complement system. With the data presented in this paper, it is now possible to make a comparison between the human and bovine enzymes. The equations generated in this study indicate that the human enzymes are different from the corresponding bovine enzymes.

Table 6. Correlation Equations for $Cl\bar{s}$, Trypsin, Thrombin, Plasmin (Selected Compounds)

	n	s	r
Plasmin			
Eliminate 3-NH_2, 3,4-diMe, 4-NO_2, 3-CF_3, 3-Br			
log 1/Ki = 3.2 + .35π	11	.32	.81
= 3.00 + .32 M.R.	11	.32	.81
Trypsin			
Eliminate 3-Phenyl, 3-Benzyl, H, 4-MeO, 3-COOH			
log 1/Ki = 4.1 - 1.3 R	11	.25	.88
$Cl\bar{s}$			
Eliminate 3-NH_2, 2-Me, 3-CH_2OH			
log 1/Ki = 3.01 - 3.04 R	12	.61	.80
Thrombin			
Eliminate 2-Me, 4-CH_2OH			
log 1/Ki = 2.87 - 2.55 R	13	.45	.87

The major factors controlling the inhibition of human trypsin and thrombin seem to be electronic in nature; however, π_{NO_2} or P_E appears to be important in the inhibition of the bovine enzymes. Coats (1973) also found that the combination of σ_m with π_{NO_2} or P_E resulted in a significant improvement in the correlation of the data and therefore electronic effects are important. We do not question the contribution of σ_m to inhibition by benzamidines since we have found R to be the most important single variable for human trypsin and thrombin. However, we do question the validity of the equations generated by Coats (1973) for the following reasons. First, the range of σ_m constants that were used by Coats (1973) in his correlations for trypsin was narrow (.19 units), and five of the eight σ_m constants had the same value. The second objection is that the formation of a two variable equation for trypsin is at the limit of data points per variable. Topliss and Costello (1972) have suggested that at least five to six data points per parameter should be available in order to avoid chance correlations.

A comparison between our equations and the ones calculated by Coats (1973) for human plasmin are very similar, if one ignores the equation in σ_m. We have the same objections to this equation as for the trypsin equations. The similarity of the equations generated in this paper to those developed by Coats (1973) help to point out the applicability of multiparameter regression analysis. Markwardt *et al.* (1968, 1969, 1970) evaluated a number of alkoxy benzamidines different from the substituted benzamidines we have studied. Even so, the equations 1 to 4 below show that the same direction for inhibitor synthesis is indicated. This also shows that a mass synthesis approach to inhibitor design is not really necessary. For example, the more hydrophobic inhibitors will, to a

point, be the best. Coats (1973) saw that the inhibition for 3-OC_7H_{11} benzamidine did fall off. Although this is an isolated case, we agree that this may be near the upper level of hydrophobicity acceptable and that even more hydrophobic groups would eventually lead to the inclusion of π^2 term.

Plasmin (from Coats 1973):

$\log 1/K_i = .14\pi + 3.64$	n=8	s=.14	r=.80	1
*$\log 1/K_i = .2P_E + 3.60$	n=8	s=.13	r=.85	2

Plasmin (this work):

$\log 1/K_i = .35\pi + 3.2$	n=11	s=.32	r=.81	3
$\log 1/K_i = .32\ MR + 3.00$	n=11	s=.32	r=.81	4

A comparison of the $C1\bar{s}$ data with the data for the inhibition of guinea pig complement is made with some reservations as we are analyzing one enzyme, and guinea pig complement consists of at least eleven proteins. However, an inspection of two of the equations for $C1\bar{s}$ and whole complement showed them to be very similar with respect to the values of their coefficients.

$C1\bar{s}$ (this work):

$\log 1/K_i = 2.91 + .45\pi_{NO_2} - 2.1R$	n=15	s=.64	r=.78	5

Whole guinea pig complement (Coats 1973):

$\log 1/I_{50} = 2.56 + .26\pi_{NO_2} - 1.23\sigma_m$	n=25	s=.29	r=.83	6

Combination of this observation with other data that shows $C1\bar{s}$ to be inhibited by a wide variety of benzamidines (Bing, 1969) supports the conclusion that a main site of action of whole complement inhibition by benzamidines is in the $C1\bar{s}$ molecule.

---

*Note: We have multiplied the P_E coefficient by ten to make it more equiscalar with our MR values since we have multiplied our MR values by .1 to put then in a range with π.

We have shown that when a subset of all compounds tested is used in the biological correlations that three enzymes, trypsin, thrombin and $C1\bar{s}$ are very similar while plasmin is different. However, when all of the compounds are included, the correlations, even though somewhat less significant, do begin to point out differences among the enzymes studied. We believe that the decreased significance is a result of some unknown parameter which is required for good correlations. The usefulness of the equations lies in the fact that they do indicate a direction for the synthesis of new compounds that could show selective specificity for the enzymes studied.

ACKNOWLEDGEMENTS

The authors would like to thank Ms. Sandra Spurlock and Dr. Judith Andrews for their interest, help and criticism of our work, and Ms. Rachelle Rosenbaum for typing the manuscript.

LIST OF REFERENCES

LIST OF REFERENCES

Assimeh, S. N., Bing, D. H., and Painter, R. (1974), J. Immunol. 113, 225.

Austen, K. F. (1971), Hopkins J. Med. 128, 57.

Baker, B. R., and Cory, M. (1969a), J. Med. Chem. 12, 1049.

Baker, B. R., and Cory, M. (1969b), J. Med. Chem. 12, 1053.

Baker, B. R., and Cory, M. (1971a), J. Med. Chem. 14, 119.

Baker, B. R., and Cory, M. (1971b), J. Med. Chem. 14, 808.

Baker, B. R., and Erickson, E. H. (1969), J. Med. Chem. 12, 408.

Bender, M. L., Kezdy, F. J., and Feder, J. (1965), J. Am. Chem. Soc. 87, 4953.

Bender, M. L., Killheffer, J. V., and Roeske, R. W. (1965), Biochem. Biophys. Res. Comm. 19, 161.

Bing, D. H. (1969), Biochemistry 8, 4503.

Coats, E. A. (1973), J. Med. Chem. 16, 1102.

Cornish-Bowden, A. (1974), Biochem. J. 137, 143.

Deutsch, G. E., and Mertz, E. T. (1970), Science 170, 1095.

Dixon, M. (1953), Biochem. J. 55, 170.

East, E. J., and Trowbridge, C. G. (1968), Arch. Biochem. Biophys. 125, 334.

Hammett, L. P. (1970), Physical Organic Chemistry, 2nd Ed., McGraw-Hill, NY.

Hansch, C., Leo, A., Unger, S. H., Kim, K. H., Nikaitani, D., and Lien, E. J. (1973), J. Med. Chem. 16, 1207.

Hansch, C., and Yoshimoto, M. (1974), J. Med. Chem. 17, 1160.

Johnson, D., and Travis, J. (1976), Anal. Biochem. 72, 573.

Kezdy, F. J., Lorand, L., and Miller, K. D. (1965), Biochemistry 4, 2302.

Krieger, M., Kay, L. M., and Stroud, R. M. (1974), J. Mol. Biol. 83, 209.

Leo, A., Hansch, C., and Elkins, D. (1971), Chem. Res. 71, 525.

Lowry, O. H., Rosebrough,, N. J., Farr, A. L., and Randall, R. J. (1951), J. Biol. Chem. 193, 265.

Markwardt, F., Landmann, H., and Walsmann, P. (1968), Eur. J. Biochem. 6, 502.

Markwardt, F., Walsmann, P., and Kazmirowski, H. G. (1969), Pharmazie 24, 400.

Markwardt, F., Walsmann, P., and Landmann, H. (1970), Pharmazie 25, 551.

Martin, C. J., Bolukow, J., and Axelrod, A. E. (1958), Biochem. Biophys. Acta 27, 430.

Muller-Eberhard, H. J. (1968), Advanced Immunology, Vol. 8, Dixon, Jr., F. J., and Kunkel, H.C., Eds., Academic Press, NY.

Nys, G. G., and Rekker, R. F. (1973), Chim Ther. 8, 521.

Nys, G. G., and Rekker, R. F. (1974), Eur. J. Med. 9, 361.

Rosenberg, R. D., and Waugh, D. F. (1970), J. Biol. Chem. 245, 5049.

Silverstein, R. M. (1973), Thrombosis Res. 3, 729.

Swain, C. G., and Lupton, E. C. (1968), J. Am. Chem. Soc. 90, 4328.

Topliss, J. G., and Costello, R. J. (1972), J. Med. Chem. 15, 1068.

Walsh, K. A., and Wilcox, P. E. (1970), Methods in Enzymology 19, 31.

CONCLUDING REMARKS

CONCLUDING REMARKS

The data presented in this thesis demonstrate: 1) The human enzymes, trypsin, plasmin, thrombin and $C1\bar{s}$ hydrolyze both neutral (N^{α}-Z-L-Tyr-p-Np) and cationic (N^{α}-Z-L-Lys-p-Np) substrates and are inactivated by PMNAB; 2) The binding of neutral and cationic substrates for hydrolysis most likely occurs in the anionic binding site; 3) The binding sites of the human enzymes are different from one another as determined by QSAR techniques; 4) Bovine trypsin and thrombin differ from their human counterparts; 5) The main site of action of substituted benzamidines on whole guinea pig complement is probably on the $C1\bar{s}$ subunit of this enzyme; 6) The most common variables presently used in QSAR are not sufficient to determine an accurate description concerning the factors responsible for influencing the binding of substituted benzamidines by the four enzymes studied.

Multiparameter regression analysis is a useful alternative technique for the design of active site directed reversible and irreversible inhibitors. Most of the investigations on developing these types of inhibitors have relied on luck, intuition and educated guesses. The studies in this thesis have shown that after analyzing a small number of inhibitors, a direction for further inhibitor design can be suggested. QSAR developments have raised the design of inhibitors to a point somewhere over the educated guess phase. Further developments in this area, such as,

the use of new parameters (e.g. charge transfer complex calculation, or molecular orbital-charge density calculations) offer the promise of a more rigorous absolute approach to inhibitor design.

We have also used the data obtained in the QSAR analysis to support our contention that the binding site of neutral aromatic and cationic substrates are identical. We have investigated the use of N^{α}-Z-L-Lys-p-Np in the determination of K_i by substituted benzamidines. The values of K_i for both N^{α}-Z-L-Tyr-p-Np and N^{α}-Z-L-Lys-p-Np agree well within experimental error. This of course should be predicted since K_i is independent of the substrate used provided that the same mechanism of inhibition is seen. This is further evidence that the lysine and tyrosine esters bind at the same location.

We cannot stress the importance of the QSAR technique enough. It is true that it appears that the present state of the art is insufficient to analyze specific systems, such as enzyme active sites. However, refinement and development of this technique to include new constants and approaches will be very important in expanding our knowledge concerning the nature of enzyme specificity. From a practical standpoint, QSAR will be important in designing new effective drugs for use in disease states. It is possible to perform dual QSAR on drugs and examine both effectiveness and toxicity. Figure 1 is an example where one can plot toxicity and effectiveness vs. some parameter and allow one to maximize the effectiveness while arriving at an acceptable level of toxicity. This analysis is critical to drug design whether it be by intuition, luck, guesswork or QSAR.

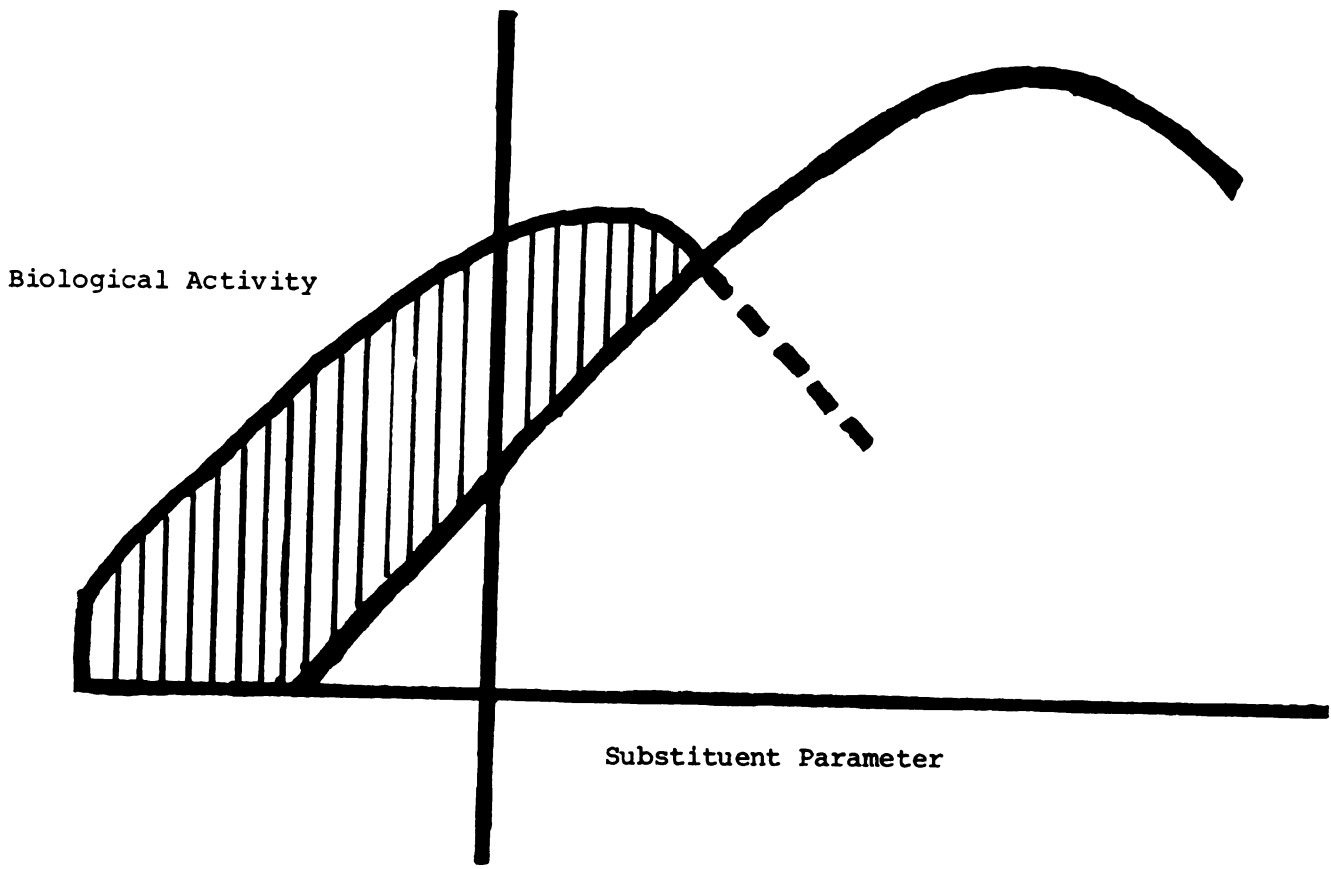


Figure 1.