Journal of Pharmaceutical Sciences

JUNE 1970
VOLUME 59 NUMBER 6

REVIEW ARTICLE

α-Chymotrypsin: A Case Study of Substituent Constants and Regression Analysis in Enzymic Structure–Activity Relationships

CORWIN HANSCH and EUGENE COATS

Keyphrases □ α-Chymotrypsin—structure–activity relationship □ Structure–activity relationship—α-chymotrypsin substituent constants □ Regression analysis—α-chymotrypsin structure–activity relationship □ Inhibitors—α-chymotrypsin □ Substrates—α-chymotrypsin

One of the central problems in drug research is the development of a systematic approach for the elucidation of structure–activity relationships. The problem is so pressing and so complex that it must be approached on many levels simultaneously. In this report the authors are concerned with purified enzyme, one of the simplest systems. The concept of drug-receptor sites, which has developed in recent years (1), is much like that of the active sites on enzymes. In some instances, enzymes are the active sites. Thus it seemed reasonable to make a survey of the structure–activity work on one of the most thoroughly studied enzymes, chymotrypsin.

The hydrolytic enzymes have been the subject of the most extensive studies of structure–activity interrelation. Although an enormous amount of work has been devoted to all aspects of such enzymes, the understanding of their mechanism of action is still quite incomplete (2–4). Of this class of enzymes the most extensively studied is chymotrypsin. Its modes of action in the hydrolysis of esters and amides have been reviewed from various points of view (2–9). The present survey is from a particular point of view, namely that of extrathermodynamic structure–activity relationships.

In surveying the literature, the authors have selected for analysis only those sets of data for which suitable substituent constants are available. Even with this limitation, the data are so voluminous and scattered that they cannot claim to have included all possible examples. Their object has been to try to characterize in a gross fashion some common characteristics of substrates and inhibitors, which have been defined by the constants K_m and K_i. This effort is not intended to be the final definitive study but rather an initial study which hopefully will encourage others to undertake better designed experiments to explore more thoroughly substituent effects for which, at present, ideal data are lacking.

Considerable evidence now supports the view that chymotrypsin operates by what is called a double-displacement mechanism (2, 4, 10, 11); that is, the enzyme is first acylated and then the acyl intermediate is hydrolyzed to products:

$$\mathrm{RC(=O)—OR' + enzyme \rightarrow RC(=O)—enzyme \rightarrow RCOOH + enzyme + R'OH} \quad \text{(Eq. 1)}$$

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} ES' \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P_2 \qquad \text{(Eq. 2)}$$

(ES' is the acyl intermediate; P_1 is the first product released.)

$$ES' + H^+ \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} ES'H \qquad K_i = \frac{k_{-4}}{k_4} \qquad \text{(Eq. 3)}$$

The kinetic implications of this mechanism have been worked out by Gutfreund and Sturtevant (10, 11). In the early stages of the reaction, it is customary to assume that $k_2[ES] \gg k_{-2}[ES'][P_1]$ and that $k_3[ES'] \gg k_{-3}[E][P_2]$. If a large amount of substrate is present, then it may be assumed that $[S] = [S]_0$ where $[S]_0$ is the initial substrate concentration. Application of the steady-state approximation to ES yields:

$$\{(k_2 + k_3')[S]_0 + k_3K_m\}[E] = k_3'K_m[E]_0 + \frac{k_3'K_m[S]_0[E]_0}{K_m + [S]_0} \exp\left[-\left(\frac{(k_2 + k_3')[S]_0 + k_3'K_m}{K_m + [S]_0}\right)t\right] \quad \text{(Eq. 4)}$$

where $[E]_0 = [E] + [ES] + [ES']$ is the total enzyme concentration, t represents time, $K_m = (k_{-1} + k_2)/k_1$, and $k_3' = k_3K_i/(K_i + [H^+])$. The rates of product appearance are:

$$V = \frac{d[P_1]}{dt} = \frac{k_2}{K_m}[E][S]_0 \quad \text{(Eq. 5)}$$

$$\frac{d[P_2]}{dt} = k_3'[E]_0 - k_3'\frac{K_m + [S]_0}{K_m} \quad \text{(Eq. 6)}$$

Measurements which are made at times large enough so that the exponential term in Eq. 3 can be neglected can be treated using Eq. 6:

$$V = \frac{d[P_1]}{dt} = \frac{d[P_2]}{dt} = \frac{k_2k_3'[S]_0[E]_0}{(k_2 + k_3')[S]_0 + k_3'K_m} \quad \text{(Eq. 7)}$$

Lineweaver-Burk plots at different values of $[S]_0$ yield the apparent constant:

$$K_{m(\text{app.})} = \frac{k'_3}{k_2 + k'_3}K_m \quad \text{(Eq. 8)}$$

From the development of Eq. 8, it is clear that the apparent K_m with which the authors are concerned is a complex constant. The problem is even more complicated if one assumes two or more binding sites to be involved which might or might not lead to productive reactions (2). In fact, the picture is so complex that for structure–activity studies, one must proceed with an open mind and examine what empirical evidence is available simply to improve one's ability to design better experiments with which to make still closer approximations. The dangers of relying on $K_{m(\text{app.})}$ as a meaningful constant are apparent from Eq. 8, and they have also been stressed by Bender and Kézdy (2).

While considerable effort (2, 9, 10, 12) has been made to define the kinetic parameters associated with Eqs. 1–3, the present consideration will be limited to the hopefully less complex parameters $K_{m(\text{app.})}$ and K_i. Despite many theoretical points for concern about the complex nature of K_m, Neurath and Hartley (13) have summarized considerable evidence to show that K_m is a close approximation to the simple binding constant K_s. This review further supports this idea and attempts to characterize more sharply the enzymic areas surrounding the catalytic site.

The general approach to the formulation of extrathermodynamic models has been well analyzed for nonenzymic reactions by Leffler and Grunwald (14). For enzymic processes, hydrophobic forces become extremely important, and these must play an important part in any model (15–18). The approach in such studies has been to factor effects of substituents on rate or equilibrium constants into free energy-related terms as follows:

$$-\Delta G = RT \ln K \quad \text{(Eq. 9)}$$

$$\delta_X \Delta G = \delta_X \Delta G_{\text{hydrophobic}} + \delta_X \Delta G_{\text{electronic}} + \delta_X \Delta G_{\text{steric}} \quad \text{(Eq. 10)}$$

That is, the effect of substituent X on the free energy change in a rate or equilibrium process characterized by k is factored in operational terms as shown in Eq. 10. Extrathermodynamic numerical solutions to Eq. 10 can be obtained by the use of suitable substituent constants:

$$\delta_X \log K = a\pi_X + b\sigma_X + cE_{s_X} + d \quad \text{(Eq. 11)}$$

Other free energy-based parameters such as polarizability may also be employed in Eq. 11, so the pertinent substituent parameters determining biological response can be sorted out *via* regression analysis. In Eq. 11, the parameter π is defined (19) as: $\pi_X = \log P_X - \log P_H$, where P_X is the octanol–water partition coefficient of a derivative and P_H is that of a parent molecule. Hydrophobic binding is operationally defined by the octanol–water reference system. Electronic effects of X, represented by Hammett's σ-parameter and its various modifications (14), are related to highly specific electronic effects not contained in π. To represent the steric effects of X, Taft's E_s parameter (20) can be used (21, 22). Here again, as with π and σ, there is overlap between π and E_s. Steric effects are highly specific effects such as those involved in the formulation of E_s. These parameters have been successfully employed in the delineation of the roles of substituents in a variety of systems (18) including a number of interest here, namely, simple proteins (23) and enzymes (16).

While the parameters σ and E_s have been extensively studied and their use justified by many good correlations in relatively simple systems, partition coefficients have been much less studied. That they can be used to correlate quantitatively binding constants of small molecules to proteins is illustrated (17) in Eq. 12:

$$\log 1/C = 0.75(\pm 0.07)\log P + 2.30(\pm 0.15) \quad \begin{matrix} n & r & s \\ 42 & 0.960 & 0.159 \end{matrix} \quad \text{(Eq. 12)}$$

In Eq. 12, C is the molar concentration of organic compound necessary to produce a 1:1 complex with purified bovine serum albumin, and P is the octanol–water partition coefficient. This linear free energy relationship correlates the affinities of 42 miscellaneous (phenols, anilines, naphthalenes, alcohols, *etc.*) organic compounds for a hydrophobic site on serum albumin. This equation and many similar ones (17, 18, 23) appear to justify the octanol–water partition coefficient as a meaningful parameter for estimating hydrophobic character of a molecule. The relative constant π is the corresponding parameter for a substituent (19, 24).

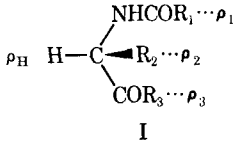
While the present authors have used octanol–water as a reference system, Scholtan has shown in a rather extensive study that the isobutanol–water system appears to yield comparable results (25). Wildnauer and

Canady (26) have recently shown that for certain chymotrypsin inhibitors, aliphatic hydrocarbons such as pentane, hexane, heptane, and water serve as suitable reference systems. They also correlated inhibitor potency with surface area of the inhibitor (26).

Since α-chymotrypsin has been readily available in crystalline form for some time, it has been the subject of extensive investigation. The detailed studies of Niemann, Bender, Baker, and others (see 27–32), dealing with the selectivity and inhibition of the enzyme, along with recent sequence analyses and X-ray crystallographic work (33) provide data for structure–activity analysis. A variety of data is most important; a correlation of one or two sets of data for such a complex process as enzyme substrate interaction leaves one with the feeling that other workers using different molecules might find quite different results. It is only after many sets of quite different data can be treated with a consistent result that one can place much confidence in such extrathermodynamic correlations.

One unfortunate observation of the present survey is that most of the studies of substrates and inhibitors were designed without giving much thought to the present state of the ability to treat substituent effects in quantitative terms; that is, instead of varying one of two parts of substrate or inhibitor molecules with substituents for which reliable physicochemical parameters are known, workers have often studied a small set of congeners in which gross changes have been made which preclude treatment by present methods of analysis. It is hoped that a result of the analyses in this report will be the encouragement of better designed studies for *quantitative* structure–activity analysis.

In attempting to understand substituent effects on the interaction of substrates and inhibitors with α-chymotrypsin, the model of Hein and Niemann (30, 31) is of help in orienting the discussion. Actually, their nomenclature will be employed to designate the space around an asymmetric center held on the active site of the enzyme. As will become apparent, the results of this analysis do not support the necessity of postulating the large number of microscopic binding constants in which each of the less than perfect modes of interaction plays a definite part in holding substrate to enzyme. The view taken here is similar to that of Bender and Kézdy (2) that the 12 possible modes of interaction between, say, an L-substrate and the enzyme violate the rule of "scientific simplicity," and this complexity should not be invoked until absolutely necessary. The general picture of the Hein-Niemann model is shown in Structure I. In this generalized model, $NHCOR_1$ represents the *N*-acyl portion; R_2 is the side chain in the α-position; COR_3 is the ester or amide bond that is hydrolyzed; and ρ_1, ρ_2, ρ_3, and ρ_H are the areas of the active site with which the four substituents on the α-carbon interact. Structure I depicts the most specific



I

Table I—L-$RCHCO_2CH_3$ ($NHCOCH_3$) as Chymotrypsin Substrates

R	E_s[a]	π[b]	σ^*[a]	log $1/K_m$ Obs.[c]	log $1/K_m$ Calcd.[d]
Methyl	0.00	0.50	0.00	−2.87	−2.69
Ethyl	−0.07	1.00	−0.10	−1.72	−1.99
Propyl	−0.36	1.50	−0.12	−1.01	−1.28
Butyl	−0.39	2.00	−0.13	−0.83	−0.57
Pentyl	−0.40	2.50	−0.16	−0.21	0.14
Hexyl[e]	−0.40	3.00	−0.17	−0.47	
Isopropyl	−0.47	1.30	−0.19	−2.05	−1.56
Isobutyl	−0.93	1.80	−0.13	−0.58	−0.85
Benzyl	−0.38	2.03[f]	0.22	−0.10	−0.53
Cyclohexylmethyl	−0.98	2.89[g]	−0.06	0.72	0.69

[a] From *Reference 20.* [b] From *Reference 19.* [c] From *Reference 50; K_m* is in m*M*. [d] Calculated using Eq. 13. [e] Not included in derivation of Eq. 13. [f] Benzyl value of 2.63 minus 0.60 for folding effects (24) observed on measured log *P* for *N*-acetylphenylalanine methyl ester. [g] Calculated using measured log *P* for cyclohexanol minus hydroxyl plus methyl. See text for discussion.

of the 12 possible interactions, with all of the others resulting in a less productive or nonproductive complex termed "wrong-way binding."

Recent experimental evidence now makes it possible to indicate some of the amino acids of the enzyme which are involved in the portions of the active site, as proposed by Hein and Niemann.

ρ_H Area: On the basis of results (34, 35) with α-methylamino acids, it has been concluded that although the hydrolysis rate is lowered, the methyl group has little effect on binding. Hence, the ρ_H area is presumed to be open to solvent.

ρ_1 Area: On the basis of unique dialkylation of the active site by *p*-nitrophenyl-*N*-bromoacetyl-α-aminoisobutyrate, methionine 192 has been identified as being reactive in this area of the active site (36). The ester portion of this reagent was first hydrolyzed by the enzyme, affording *p*-nitrophenol and acylated serine 195. Then the bromoacetyl moiety, now fixed to the enzyme, reacted with methionine 192.

ρ_2 Area: Evidence for a specific role for an amino acid residue interaction with a substrate in this area has not been found. There is some evidence to suggest that a tryptophan residue in this region might participate in a charge transfer reaction (37). X-ray crystallographic studies indicate (33) that tryptophan 215 lies near methionine 192.

ρ_3 Area: By irreversible alkylation studies with diisopropyl fluorophosphate and other reagents, a serine residue has been implicated in this portion of the active site (38). From amino acid sequence studies and X-ray crystallography, it is now fairly certain (33, 39) that this is serine 195. Histidine 57 has also been identi-

Table II—$C_6H_5CH_2CH_2CO_2R$ as Chymotrypsin Substrates

R	π[a]	log $1/K_m$ Obs.[b]	log $1/K_m$ Calcd.[c]
Methyl	0.50	3.27	3.27
Ethyl	1.00	3.36	3.37
Propyl	1.50	3.48	3.48

[a] From *Reference 19.* [b] From *Reference 26; K_m* is in *M*. [c] Calculated using Eq. 14.

Table III—$C_6H_5CONHCH_2CO_2R$ as Chymotrypsin Substrates

R	E_s[a]	π[b]	log $1/K_m$ Obs.[c]	log $1/K_m$ Calcd.[d]
Methyl	0.00	0.50	−0.51	−0.51
Ethyl	−0.07	1.00	−0.36	−0.34
Propyl	−0.36	1.50	−0.28	−0.25
Isopropyl	−0.47	1.30	−0.31	−0.37
Butyl	−0.39	2.00	−0.04	−0.06
Isobutyl	−0.93	1.80	−0.38	−0.36

[a] From *Reference 20.* [b] From *Reference 19.* [c] From *Reference 46; K_m* is in m*M*. [d] Calculated using Eq. 15.

fied (40–42) by alkylation and X-ray crystallography in this portion of the active site, and it lies only 4 Å from the serine 195 residue (33).

METHOD

The general approach is to account for substituent effects on the binding of substrate by enzyme, as represented by log $1/K_m$ or log $1/K_i$, by the linear combination of free energy-based substituent constants (Eq. 11). In Eq. 11, the constants *a*, *b*, *c*, and *d* are found by the method of least squares. (See *Reference 43* for a discussion of this technique and its use in regression analysis.) While E_s has been defined (19) to represent intramolecular steric repulsions, it has been found to be of use in intermolecular interactions as well (21, 22). In testing the value of E_s in chymotrypsin correlations, the authors have also tested Hancock's corrected parameter (44), defined as: $E_s^c = E_s + 0.306(n - 3)$, where n represents the number of α-hydrogens on the substituent alkyl group. In assaying electronic effects of substituents, σ, as well as its variations (14) σ_I, σ^+, and σ^-, has been employed.

For hydrophobic interactions, π or log P has been used. In most of the examples, a large portion of the substrate or inhibitor was held constant, and π can be used to represent the hydrophobic effect of the substituents. For example, in Table I the $-CHCO_2CH_3$ portion
$NHCOCH_3$
of the substrates is constant, and π has been used for each of the R functions attached to the α-carbon. For each CH_3 or CH_2 group, the value of 0.50 for π was used to estimate π for higher alkyl functions. The additive constitutive character of π and log P greatly simplifies such analyses (19, 24).

From the data in Tables I–X, the equations in Table XX have been derived for enzyme substrate binding; from the data in Tables XI–XIX, the equations in Table XXI have been derived for enzyme inhibitor interaction. Only the most significant equations, as determined (43) by an F test ($\alpha \leq 0.10$), are given. In several of the examples, so few data points are available that one cannot be sure that higher order (45) equations are not needed.

Table IV—$C_6H_5CONHCH_2CO_2R$ as Chymotrypsin Substrates

R	π[a]	log $1/K_m$ Obs.[b]	log $1/K_m$ Calcd.[c]
Methyl	0.50	2.57	2.56
Ethyl	1.00	2.63	2.63
Propyl	1.50	2.77	2.75
Isopropyl	1.30	2.66	2.69

[a] From *Reference 19.* [b] From *Reference 26; K_m* is in *M*. [c] Calculated using Eq. 16.

Table V—$XCH(CH_2COOEt)_2$ as Chymotrypsin Substrates

X	π[a]	σ_I	log $1/K_m$ Obs.[b]	log $1/K_m$ Calcd.[c]
—OH	−1.16	0.29	−2.00	−1.91
—$NHCOCH_3$	−1.21[d]	0.24	−1.82	−1.94
—$OCOCH_3$	−0.27	0.33	−1.57	−1.45
—H	0.00	0.00	−1.21	−1.31

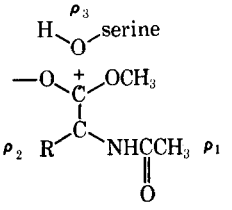
[a] Aliphatic values from *Reference 24.* [b] From *Reference 47; K_m* is in *M*. [c] Calculated using Eq. 17. [d] Aliphatic $CONH_2$ plus 0.50 for methyl group.

The following log P values in the octanol–water reference system have been measured for the first time: *N*-acetyl-DL-phenylalanine methyl ester, 0.92 ± 0.01; ethyl propyl *p*-nitrophenylphosphonate, 2.20 ± 0.01; ethyl furoate, 1.52 ± 0.01; azulene, 3.20 ± 0.02; ethyl isonicotinate, 1.43 ± 0.02; ethyl nicotinate, 1.32 ± 0.01; ethyl picolinate, 0.87 ± 0.01; ethyl anthranilate, 2.57 ± 0.01; and ethyl 2-thiophenecarboxylate, 2.33 ± 0.03. The standard deviations are from four separate determinations using different volume ratios of solvents, except for ethyl picolinate where only three determinations were made. To calculate π-values for the acyl functions of Table VIII, the ethoxy value of −0.23 derived from diethyl ether was subtracted from the measured log P values for the ethyl esters.

The use of group polarizability (P_E) as reflected in atomic molecular refractivities in quantitative correlations was suggested by the recent study of Agin *et al.* (59). They showed that, as a first approximation, binding of small molecules to macromolecules should be logarithmically related to αI, where α is the polarizability and I is the ionization potential. It is found (49) that little is lost in correlation if I is neglected. Niemann and Hein also considered the importance of polarizability (52). Although there is a large amount of cocorrelation between P_E and π since both are dependent on molar volume, the proper selection of derivatives shows enough independence so that they may be used to characterize enzymic binding areas. For example, $r^2 = 0.3$ for the correlation between π and P_E for the substituents in Table IX. For the 21 groups of Table VII, the cocorrelation is $r^2 = 0.4$.

RESULTS

Equation 13 in Table XX correlates the data of Table I. Using the Hein-Niemann model, the substrate fit to the site of action can be pictured as follows:



II

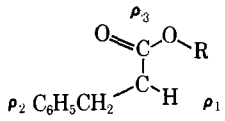
Table VI—$RCOOC_6H_4NO_2$ as Chymotrypsin Substrates

R	π^a	E_s^b	pH 5.92 log k_2/K_m Obs.[c]	pH 5.92 log k_2/K_m Calcd.[d]	pH 7.99 log k_2/K_m Obs.[c]	pH 7.99 log k_2/K_m Calcd.[e]	pH 8.90 log k_2/K_m Obs.[c]	pH 8.90 log k_2/K_m Calcd.[f]
Methyl	0.50	0.00	2.56	2.62	3.13	3.30	3.12	3.14
Ethyl	1.00	−0.07	2.80	2.89	3.53	3.51	—	—
Propyl	1.50	−0.36	2.81	2.78	3.42	3.39	3.25	3.18
Pentyl	2.50	−0.40	3.55	3.49	3.97	3.96	3.80	3.74
Isopropyl	1.30	−0.47	2.49	2.42	3.28	3.09	2.79	2.88
Isobutyl	1.80	−0.93	1.99	2.01	2.60	2.71	2.33	2.45
tert-Butyl	1.98[g]	−1.54	1.38	1.07	2.16	1.90	1.86	1.57
Neopentyl	2.48[h]	−1.74	0.82	1.12	1.69	1.92	1.36	1.56

[a] From *Reference 19*. [b] From *Reference 20*. [c] From *Reference 48*. [d] Calculated using Eq. 19. [e] Calculated using Eq. 20. [f] Calculated using Eq. 21. [g] From measured log *P* for *tert*-butylbenzene. [h] Calculated by adding 0.50 for methyl to the measured *tert*-butyl value.

In Structure II, the α-hydrogen is not shown; it is below the plane of the page. Since adding terms in E_s or σ to Eq. 13 does not result in a significant reduction in the variance of the data, enzyme substrate binding as defined by K_m depends linearly on the hydrophobic character of R as defined by the octanol–water reference system. In deriving Eq. 13, two cases where R = H or hexyl have not been included. Hein and Niemann pointed out that when R = hexyl, a sharp break occurs in both K_m and K_0, indicating that groups as long as hexyl do not fit into the ρ_2 area without some steric hindrance. The length of the chain may be involved since the cyclohexylmethyl group, which is just as lipophilic as hexyl (compare π-values, Table I), is well fit by Eq. 13. However, other explanations are possible (see *Discussion*). This same kind of break comes in data from inhibitor studies (see Eqs. 34 and 35), although at a somewhat longer chain length. The case where R = H is more active than Eq. 13 would predict. Why this is so is not clear. However, in this example, no binding possibility for the ρ_2 area exists.

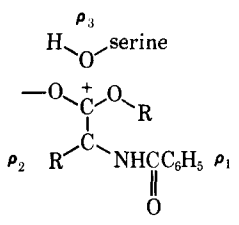
The esters of hydrocinnamic acid of Table II are correlated by Eq. 14. These derivatives can be visualized as fitting the model site as in Structure III:



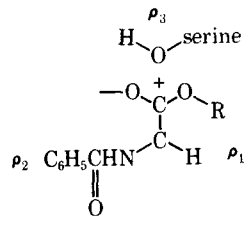
III

Although only three derivatives are in this set, they have been included simply because the hydrophobic character of the area as characterized by the coefficient with π in Eq. 14 can be compared to that in Eqs. 15 and 16. The mean coefficient with π for these three equations is 0.29 ± 0.1. The difference between this dependence of binding on π and that of Eq. 13 indicates the pronounced difference between the ρ_2 and ρ_3 areas.

Two sets of data, in which the alkyl group of the ester function pictured in Structures IV*a* and IV*b* as binding to the ρ_3 area is varied, are correlated by Eqs. 15 and 16.



IV*a*



IV*b*

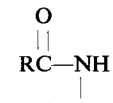


Table VII—R′—CH(L)—$COOCH_3$ as Chymotrypsin Substrates

Acyl-R	Alkyl-R′	π-R′[a]	P_{E-R}[b]	−log $1/K_m$ Obs.[c]	−log $1/K_m$ Calcd.[d]
Benzoyl	Methyl	0.50	25.10	−0.99	−1.12
2-Quinolinyl	Methyl	0.50	41.00	0.66	0.19
2-Furoyl	Methyl	0.50	17.00	−1.69	−1.78
2-Theophenoyl	Methyl	0.50	22.71	−1.18	−1.31
Nicotinyl	Methyl	0.50	23.00	−1.57	−1.29
Isonicotinyl	Methyl	0.50	23.00	−1.46	−1.29
Picolinyl	Methyl	0.50	23.00	−1.25	−1.29
Acetyl	Methyl	0.50	5.72	−2.87	−2.71
o-Aminobenzoyl	Methyl	0.50	29.40	−0.67	−0.76
Chloroacetyl	Propyl	1.50	10.58	−0.70	−0.93
Benzoyl	Propyl	1.50	25.10	0.07	0.27
Acetyl	Isopropyl	1.30	5.72	−2.05	−1.61
Chloroacetyl	Isopropyl	1.30	10.58	−1.64	−1.21
Benzoyl	Isopropyl	1.30	25.10	−0.66	−0.01
Acetyl	Propyl	1.50	5.72	−1.01	−1.33
Acetyl	Benzyl	2.03[e]	5.72	−0.10	−0.60
Acetyl	Isobutyl	1.80	5.72	−0.58	−0.92
Acetyl	Ethyl	1.00	5.72	−1.72	−2.02
Acetyl	Butyl	2.00	5.72	−0.83	−0.64
Acetyl	Pentyl	2.50	5.72	−0.21	0.05
Acetyl	Cyclohexyl-methyl	2.89[f]	5.72	0.72	0.59

[a] Taken from *Reference 19* unless otherwise noted. [b] Calculated from refractive indexes and density or from atomic refractivity; see *Reference 49*. [c] From *References 30* and *50–52*. K_m is in m*M*. [d] Calculated *via* Eq. 22. [e] Benzyl π-value of 2.63 minus 0.60 for folding interactions (24). [f] Calculated by taking the measured log *P* for cyclohexanol, subtracting the π-value for aliphatic hydroxyl, and adding the π-value for methylene.

Table VIII—$CH_3CH(L)COOCH_3$ as Chymotrypsin Substrates

NHAcyl

Acyl	π^a	P_E[b]	−log $1/K_m$ Obs.[c]	−log $1/K_m$ Calcd.[d]
Furoyl	1.75	17.0	−1.69	−1.90
Theophenoyl	2.56	22.7	−1.18	−1.31
Nicotinyl	1.55	23.0	−1.57	−1.28
Isonicotinyl	1.66	23.0	−1.46	−1.28
Picolinyl	1.10	23.0	−1.25	−1.28
Benzoyl	2.75	25.1	−0.99	−1.06
2-Quinolinyl	2.45	41.0	0.66	0.58
2-Aminobenzoyl	2.80	29.4	−0.67	−0.62

[a] See section on *Method*. [b] Calculated using values from *Reference 49*. [c] From *Reference 52*. [d] Calculated using Eq. 23.

Table IX—Acyl-$NHCH_2CO_2CH_3$ as Chymotrypsin Substrates

Acyl	P_E[a]	log $1/K_m$ Obs.[b]	log $1/K_m$ Calcd.[c]
Acetyl	5.72	−1.48	−1.82
Propionyl	10.34	−1.58	−1.63
Isobutyryl	14.96	−1.66	−1.44
Isopentanoyl	19.58	−1.38	−1.25
Chloroacetyl	10.58	−1.82	−1.63
Dichloroacetyl	15.44	−1.34	−1.42
Trifluoroacetyl	6.08	−1.72	−1.81
Phenylacetyl	29.72	−0.90	−0.83
Benzoyl	25.10	−0.88	−1.02
p-Aminobenzoyl	29.40	−0.92	−0.84
Nicotinyl	23.00	−1.49	−1.11
Isonicotinyl	23.00	−1.24	−1.11
2-Furoyl	17.00	−1.29	−1.36
Indole carbonyl	36.47	−0.14	−0.55

[a] Calculated from atomic refractivities or refractive indexes and densities; see *Reference 49.* [b] From *Reference 53.* [c] Calculated using Eq. 24.

In this example the amide moiety might fit into either the ρ_1 or ρ_2 area. In Eq. 15, where a better selection and larger number of derivatives were tested, a dependence is found on both π and E_s. The positive sign of the coefficient with E_s indicates that large functions hinder binding in the ρ_3 area as measured by $1/K_m$. This means that as R becomes larger, interaction of the electron-deficient sp^2 carbon of the ester group with a nucleophile, say serine, becomes unfavorable. Binding is promoted by increasing values of π but to a lesser extent than at ρ_2 (compare slopes in Eqs. 13 and 15). In Eq. 16, the dependence of $1/K_m$ on π is in rough agreement with that of Eqs. 14 and 15. In Eq. 17, good agreement is found between log $1/K_m$ and π. So few points are available that no assessment of the roles of E_s or σ can be made.

In a recent study of *p*-nitrophenyl esters of fatty acids, Milstien and Fife (48) found that a plot of log k_2/K_m *versus* E_s gave a fair correlation. Such a correlation is expressed in numerical form in Eq. 18. It is seen in Eq. 19 that when hydrophobic bonding is also taken into account, a much better correlation results. The coefficient with E_s is larger than that of Eq. 15, as one would expect, since R is closer to the electron-deficient carbonyl carbon in molecules of Table VI than in those of Table III. One would expect more hindrance to binding of the kind depicted in Structure IV. Studies by Milstien and Fife were made at various hydrogen-ion concentrations. Equations 20 and 21, derived from data obtained from runs made under more basic conditions, yield the same quality correlations as at the lower pH. Under each of the three different conditions of pH, the dependence of log k_2/K_m on π and E_s is the same. While there are small differences in the coefficients with these terms from equation to equation, these variations are well within the 95% confidence intervals. There is a significant difference between the intercept of the low pH (Eq. 19) and equations (Eqs. 20 and 21) for work at the higher pH. This is in line with the well-known higher activity of chymotrypsin at higher pH. The work of Milstien and Fife, expressed in units of k_2/K_m, is not directly comparable with the other work using K_m and K_i. It has been included to show that although activity of the enzyme does vary with pH, the relative substituent effects do not; at least this is true for the range of changes and pH range investigated by Milstien and Fife. This would imply that large changes in the geometry of the reaction site, as it is defined by these esters, do not occur with changes in pH. The variations in activity with pH might be attributable to the relative

Table X—Acyl-$NHCH(CH_3)COOCH_3$ as Chymotrypsin Substrates

Acyl	P_E[a]	log $1/K_m$ Obs.[b]	log $1/K_m$ Calcd.[c]
Furoyl	17.00	−1.69	−1.75
Theophenoyl	22.70	−0.87	−1.04
Nicotinyl	23.00	−0.78	−1.00
Isonicotinyl	23.00	−1.43	−1.00
Picolinyl	23.00	−1.23	−1.00
Benzoyl	25.10	−0.52	−0.74
o-Aminobenzoyl	29.40	−0.20	−0.20

[a] See *Reference 49.* [b] From *Reference 52.* [c] Calculated according to Eq. 25.

Table XI—R—$C_6H_4OCH_2COX$ as Chymotrypsin Inhibitors

R(X = CH_3)	π[a]	σ	log $1/I/S$ Obs.[b]	log $1/I/S$ Calcd.[c]
H	0.00	0.00	−1.77	−1.96
4-NO_2	0.24	0.78	−1.30	−1.41
3-NO_2	0.11	0.71	−1.40	−1.55
2-NO_2	−0.23	1.24[d]	−1.48	—[e]
4-CN	−0.32	0.66	−2.22	−1.92
3-CN	−0.30	0.56	−2.05	−1.95
4-OCH_3	−0.04	−0.27	−1.92	−2.12
4-CH_3	0.52	−0.17	−1.96	−1.63
3-CH_3	0.51	−0.07	−1.77	−1.59
4-Cl	0.70	0.23	−1.30	−1.30
3-Cl	0.76	0.37	−0.74	−1.19
2-C_6H_5	2.13	0.00[d]	−0.90	—[e]
3,4-Di-Cl	1.46	0.60	−0.57	−0.52
2,3-Di-Cl	1.35	1.05[d]	−0.41	—[e]
3,4-Benzo	1.34	0.17	−0.95	−0.82
R(X = C_6H_5)				
H	0.00	0.00	−1.40	−1.10
4-OCH_3	−0.04	−0.27	−0.98	−1.25
4-Cl	0.70	0.23	−0.40	−0.43

[a] From *Reference 19.* [b] From *Reference 54.* [c] Calculated using Eq. 26. [d] *Ortho*-values for σ taken from *Reference 55.* [e] *Ortho*-substituents omitted because of unreliable substituent constants.

Table XII—Miscellaneous Chymotrypsin Inhibitors

Compound	log *P*	log $1/K_i$ Obs.[a]	log $1/K_i$ Calcd.[b]
Phenol	1.46	−0.54	−0.52
3-Methoxyphenol	1.58	−0.30	−0.40
4-Methylphenol	1.94	−0.03	−0.05
3-Methylphenol	2.02	0.05	0.03
2-Chlorophenol	2.15	0.07	0.16
2-Bromophenol	2.35	0.21	0.35
4-Chlorophenol	2.39	0.42	0.39
4-Bromophenol	2.59	0.68	0.58
2,4-Dichlorophenol	3.08	1.14	1.06
2,4-Dibromophenol	3.48	1.35	1.45
Acetonitrile	−0.34	−2.86	−2.87
Cyclohexanol	1.23	−1.46	—
Benzene	2.13	−0.49	−0.46
N,N-Dimethylaniline	2.31	−0.22	−0.28
Chlorobenzene	2.84	0.02	0.24
Bromobenzene	2.99	0.29	0.38
Acridine	3.40	0.85	0.78
Naphthalene	3.37	0.96	0.75

[a] From *Reference 56.* [b] Calculated using Eq. 29.

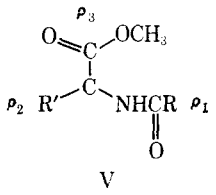
Table XIII—Aromatic Acids as Chymotrypsin Inhibitors

Compound	log P	pK_a	—log $1/K_i$— Obs.[a]	Calcd.[b]
Benzoic acid	1.81	4.17	1.99	2.31
3-Methylbenzoic acid	2.37	4.27	2.42	2.80
4-Methylbenzoic acid	2.27	4.37	2.43	2.71
3-Phenylpropionic acid	1.84	4.66	2.79	2.33
4-Phenylbutyric acid	2.42	4.76	3.22	2.85
2-Naphthoic acid	3.19	4.16	3.86	3.53
4-*tert*-Butylbenzoic acid	3.79[c]	4.41	3.89	4.07

[a] From *Reference 26*; K_m is in *M*. [b] Calculated using Eq. 30. [c] Calculated by adding *tert*-butyl π-value of 1.98 (from *tert*-butylbenzene log *P*) to the measured log *P* for benzoic acid.

protonation of the critical imidazole moiety in the ρ_3 area.

Equation 22 correlates a set of substrates (Table VII) in which changes are being made simultaneously in two positions which would affect binding in the ρ_2 and ρ_1 areas of Structure V:



V

For this set of congeners, poor correlations were obtained using $\Sigma\pi$ for R′ and R. Factoring R into two terms, $\pi_R + \pi_R'$, improved the correlation; however, from many preliminary calculations, it was observed that π does not correlate substituent effects well for the ρ_1 area. This area does not appear to be hydrophobic in character. The best correlations for this area were obtained using the group polarizability (P_E). Thus it would seem that dispersion forces and steric factors are most important for binding in this area.

In Eq. 23, using π instead of P_E yielded a correlation with r of only 0.556! For this set of substrates, where only the acyl portion is being varied, rather large changes in R result in relatively small changes in K_m. In fact, there is only a 10-fold change in binding for the molecules in Table VIII; thus the ρ_1 area does not appear to be apolar in character, nor does it appear to be sterically demanding, since a wide variety of modifications binding in this area are quite active substrates. Equations 24 and 25 are two further examples where better correlations are obtained with P_E than with π. Using π in Eq. 24 instead of P_E gave a correlation with

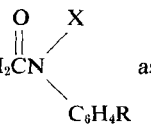
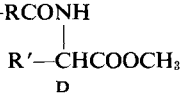


Table XIV—$C_6H_5OCH_2CN$ as Chymotrypsin Inhibitors

X	R	π_R[a]	σ_R	—log $1/I/S$— Obs.[b]	Calcd.[c]
Me	H	0.00	0.00	−1.70	−1.29
H	H	0.00	0.00	−1.11	−1.29
Me	4-NO_2	0.24	0.78	−0.78	−1.00
H	4-Cl	0.70	0.23	−0.18	−0.44
H	4-Br	1.02	0.23	−0.23	−0.05
H	3-Cl	0.76	0.37	−0.43	−0.37

[a] From *Reference 19*. [b] See *b*, Table XI. [c] Calculated using Eq. 31.

Table XV—RCONH as Chymotrypsin Inhibitors

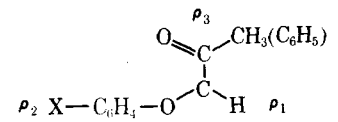


R	R′	P_{E-R}[a]	π_R'[b]	—log $1/K_i$— Obs.[c]	Calcd.[d]
CH_3	$(CH_3)_2CH$	5.72	1.30	−2.30	−2.55
$ClCH_2$	$(CH_3)_2CH$	10.58	1.30	−2.22	−2.11
C_6H_5	$(CH_3)_2CH$	25.10	1.30	−0.72	−0.82
CH_3	C_3H_7	5.72	1.50	−2.05	−1.97
$ClCH_2$	C_3H_7	10.58	1.50	−1.77	−1.54
C_6H_5	C_3H_7	25.10	1.50	−0.25	−0.24
CH_3	$C_6H_5CH_2$	5.72	2.03[e]	−0.36	−0.45

[a] See *Reference 49*. [b] From *Reference 19*. [c] From *Reference 51*. [d] Calculated using Eq. 33. [e] Calculated from benzyl π-value of 2.63 minus folding contribution of 0.60.

$r = 0.627$, while the same procedure with Eq. 25 gave a correlation with $r = 0.709$. The mean coefficient with P_E in the four substrate examples (Eqs. 22–25) is 0.088 ± 0.035.

The equations in Table XXI summarize the structure–activity relationships of the data from Tables XI–XIX on chymotrypsin inhibitors. One of the best designed sets of inhibitors comes from the work of Baker *et al.* (54). The results are summarized in Eq. 26. Most of his derivatives were methyl ketones; however, a few were phenyl ketones. The constant X was given a value of 0 for the methyl ketones and a value of 1 for the phenyl ketones. Both the π- and σ-parameters are necessary to obtain a good correlation. The fit of these inhibitors to the model of the active site can be depicted as Structure VI:



VI

The dependence of the inhibiting power of these derivatives on hydrophobic binding is strong, although somewhat different from Eq. 13. This is probably due to the fact that $1/K_m$ and I_{50} are not strictly comparable ways of comparing binding affinities. The positive coefficient with σ in Eq. 26 means that electron-withdrawing substituents also promote binding. Since π differs slightly from system to system and this difference is related to σ, part of the effect of σ may simply be that

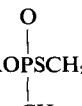


Table XVI—ROPSCH$_2$CH$_2$SEt as Chymotrypsin Inhibitors

R	π[a]	—log K_i— Obs.[b]	Calcd.[c]
Methyl	0.50	−0.47	−0.59
Ethyl	1.00	−0.34	−0.02
Propyl	1.50	0.53	0.55
Butyl	2.00	1.26	1.11
Pentyl	2.50	1.86	1.68
Hexyl	3.00	2.44	2.25
Heptyl	3.50	2.52	2.81
Octyl[d]	4.00	—	—

[a] From *Reference 19*. [b] From *Reference 57*. [c] Calculated using Eq. 34. [d] This point not employed in the regression; see text for discussion.

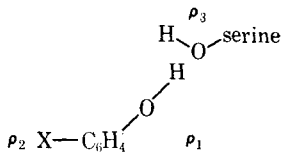
Table XVII—$ROP(=O)(CH_3)-SCH_2CH_2S^+(CH_3)Et\ CH_3SO_4^-$ as Chymotrypsin Inhibitors

R	π[a]	log K_i Obs.[b]	log K_i Calcd.[c]
Methyl	0.50	0.27	0.00
Ethyl	1.00	0.14	0.64
Propyl	1.50	1.24	1.28
Butyl	2.00	2.09	1.92
Pentyl	2.50	2.71	2.57
Hexyl	3.00	3.47	3.21
Heptyl	3.50	3.55	3.85
Octyl[d]	4.00	3.42	—

[a] From *Reference 19.* [b] From *Reference 57.* [c] Calculated using Eq. 35. [d] This point was not used in the regression; see text for discussion.

of correcting π; this could, in fact, explain part of the difference between Eqs. 26 and 13. However, in the present case this difference must be small, since π-constants are from the phenoxyacetic acid system which, of course, is quite closely related to the phenoxyacetones. Whether the phenyl ring is really fitting into the same area (ρ_2) as the R group of Eq. 13 is, of course, open to question. It is possible that the inhibitors bind in another part of the enzyme and bring about their effects allosterically. The dummy parameter X of Eq. 26 is simply a technique (43) for taking into account the stereoelectronic difference between the methyl and phenyl functions. Using P_E in Eq. 26 instead of π yields a poorer correlation (r = 0.779), indicating that this set more closely resembles compounds of Eq. 13 than molecules binding to the ρ_1 area (Eqs. 22–25).

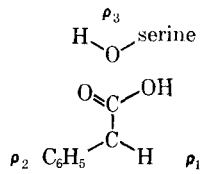
An interesting set of data is that of Berezin *et al.* (56) which gives rise to Eqs. 27–29. Equation 27 correlates the 10 phenols of Table XII, and Eq. 28 correlates the rest of the molecules in Table XII. Cyclohexanol was omitted in the regression analysis because of uncertainty in its hydrogen-bonding ability. The slopes of Eqs. 27 and 28 are quite similar, indicating a common hydrophobic mechanism of action. However, the difference in the intercepts indicates the phenols to be about eight times as effective on an isolipophilic basis. Assuming this to be due to the strong hydrogen-bonding ability of the phenols, the term, HB, has been added for hydrogen bonding, and Eqs. 27 and 28 have been combined into Eq. 29. For phenols, HB is assigned a value of 1; the other molecules were given a value of 0. Cyclohexanol does not fit into the hydrogen-bonding group; that is, giving it a value of 1 for HB results in a poorly calculated log $1/K_i$. However, using a value of 0 for hydrogen bonding results in a good calculated log $1/K_i$. The similar dependence of inhibitory activity of the compounds represented by Eqs. 13 and 29, as indicated by the coefficients associated with the π and log P terms, would suggest a common mode of binding by the inhibitors and substrates. The importance of hydrogen bonding might be rationalized as in Structure VII:



VII

It is possible that the acidic proton of the phenols aids in holding the inhibitors between areas ρ_3 and ρ_2, although there are many other ways in which the phenolic OH could participate in binding inhibitor to enzyme. Structure VII is meant to be suggestive and not to imply that only serine must be involved. Since the slopes of Eqs. 27–29 are the same, one would assume the same mechanism of inhibition as far as the hydrophobic contribution of the inhibitor is concerned. Using P_E in Eq. 29 in place of log P results in a much poorer correlation (r = 0.870). This information also supports binding in the ρ_2 rather than the ρ_1 area.

A smaller set of congeners, acting in a parallel manner to those of Eqs. 26–29, is the group of acids in Table XIII correlated by Eq. 30. This group of acids can be fit to the model as in Structure VIII:



VIII

Confidence limits on the coefficient with the log P term of Eq. 30 are large. While it cannot be said that the dependence on hydrophobic bonding is that of about 1 as was found for substrates and inhibitors reacting with the ρ_2 area, it is reasonably close. The positive coefficient with the pKa term indicates that the more unionized acids are more effective. The poorer correlation with this set (note the confidence intervals and standard deviation) may be due to steric effects involved with the different side-chain lengths between the carboxyl group and the aromatic ring.

Table XVIII—C_6H_5COR as Chymotrypsin Inhibitors

R	π[a]	log $1/K_i$ Obs.[b]	log $1/K_i$ Calcd.[c]
Acetophenone	0.50	3.27	3.27
Propiophenone	1.00	3.36	3.37
Butyrophenone	1.50	3.48	3.48

[a] Values for alkyl portion only. [b] From *Reference 26;* K_m is in *M.* [c] Calculated using Eq. 36.

Table XIX—Hydrocarbons as Chymotrypsin Inhibitors

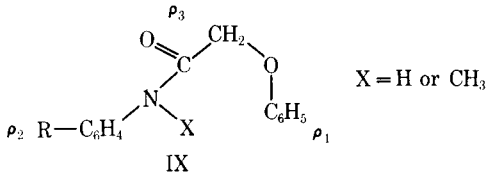
Compound	log P	log $1/K_i$ Obs.[a]	log $1/K_i$ Calcd.[b]
Azulene	3.20	4.22	3.50
Benzene	2.13	2.09	1.93
Toluene	2.69	2.55	2.75
Naphthalene	3.37	3.94	3.75
Chlorobenzene	2.84	3.10	2.97
Indene	3.33	3.67	3.70
Pentane	2.50	2.25	2.47
Cyclohexene	2.22	2.07	2.06
Ethylbenzene	3.15	2.75	3.43
Anthracene	4.45	5.27	5.34

[a] From *References 32* and *58;* K_m is in *M.* [b] Calculated using Eq. 37.

Table XX—Correlation of Structure and Activity of Chymotrypsin Substrates

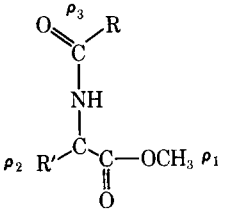
Type Substrate	"Best" Equation	n	r	s	Conditions	Eq. No.
$RCHCO_2CH_3$ $NHCOCH_3$ from Table I	$\log 1/K_m = 1.419(\pm0.40)\pi - 3.409(\pm0.74)$	9	0.955	0.350	In water at 25°, pH 7.90 and 0.10 M in sodium chloride	13
$C_6H_5CH_2CH_2CO_2R$ from Table II	$\log 1/K_m = 0.210(\pm0.22)\pi + 3.160(\pm0.24)$	3	0.997	0.012	0.1 M KCl; 3.3×10^{-3} M Tris buffer; pH 6.9; 25°	14
$C_6H_5CONHCH_2CO_2R$ from Table III	$\log 1/K_m = 0.406(\pm0.18)\pi + 0.400(\pm0.30)E_s - 0.714(\pm0.19)$	6	0.972	0.047	In water at pH 7.0 and 25°	15
$C_6H_5CONHCH_2CO_2R$ from Table IV	$\log 1/K_m = 0.251(\pm0.31)\pi + 3.343(\pm0.36)$	4	0.925	0.055	0.1 M KCl; 3.3×10^{-3} M Tris buffer; pH 6.9; 25°	16
$RCH(CH_2COOEt)_2$ from Table V	$\log 1/K_m = 0.518(\pm0.61)\pi - 1.308(\pm0.52)$	4	0.932	0.152	In water at pH 7.8, 25° and 0.1 M NaCl	17
$RCOC_6H_4NO_2$ from Table VI	$\log k_2/K_m = 1.164(\pm0.65)E_s + 3.101(\pm0.60)$	8	0.872	0.460	25° in 4.68% CH_3CN, pH 5.92	18
	$\log k_2/K_m = 1.762(\pm0.42)E_s + 0.789(\pm0.40)\pi + 2.225(\pm0.52)$	8	0.981	0.201	pH 5.92	19
	$\log k_2/K_m = 1.513(\pm0.42)E_s + 0.632(\pm0.39)\pi + 2.983(\pm0.51)$	8	0.976	0.198	pH 7.99	20
	$\log k_2/K_m = 1.620(\pm0.45)E_s + 0.627(\pm0.42)\pi + 2.823(\pm0.59)$	7	0.982	0.196	pH 8.90	21
RCONH R'—CHCOOCH$_3$ L from Table VII	$\log 1/K_m = 1.382(\pm0.27)\pi R' + 0.082(\pm0.02)P_{E-R} - 3.876(\pm0.58)$	21	0.934	0.331	In water at 25°; pH 7.90 and 0.10–0.05 M NaCl	22
L CH_3CH—$COOCH_3$ $NHCOR$ from Table VIII	$\log 1/K_m = 0.103(\pm0.023)P_E - 3.653(\pm0.62)$	8	0.975	0.179	In water at 25°; pH 7.90 and 0.10 M NaCl	23
RCONH $CH_2CO_2CH_3$ from Table IX	$\log 1/K_m = 0.042(\pm0.015)P_E - 2.068(\pm0.31)$	14	0.873	0.225	In water at 25°; pH 7.90 and 0.50 M NaCl	24
RCONH CH_3—$CHCO_2CH_3$ D from Table X	$\log 1/K_m = 0.125(\pm0.077)P_E - 3.887(\pm1.822)$	7	0.882	0.270	In water at 25°; pH 7.90 and 0.10 M NaCl	25

The amides in Table XIV, described by Eq. 31, can fit the Hein-Niemann model as in Structure IX. Unfortunately, only six derivatives are available to assess complex substituent changes. As Eq. 31 shows, π_R is an important factor, and the slope of this equation is in agreement with that of Eqs. 13 and 29. Using P_E instead of π yields, as expected, a poorer correlation (r = 0.829). So few data points are available that no assessment of the role of X could be made. The differences in activity when X is H or CH_3 are small, so that a reasonable correlation can be achieved with the single-variable Eq. 31.

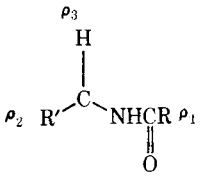


IX

The esters correlated by Eqs. 32 and 33 are comparable to those of Eq. 22, except that D-isomers of Eqs. 32 and 33 are used as inhibitors. Again, using P_E for the acyl function and π for the α-alkyl group gives the best correlation. The higher coefficient with π_R of Eq. 33 indicates binding of this function in the ρ_2 area. Two models of binding (Structures X*a* and X*b*) can be used to rationalize these results:



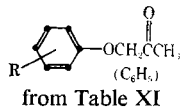
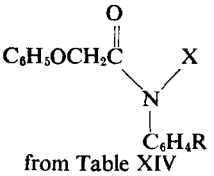
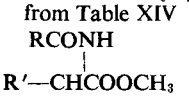
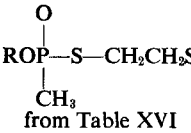
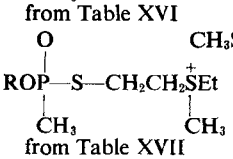
X*a*



X*b*

If binding occurs so that R′ is in the ρ_2 area and the α-H is in the ρ_H area, then the fit of Structure X*a* is obtained. The geometry here is quite different from that of Structure II and may be responsible for the difference in slope for the R term of Eq. 13 and the R′ term of Eq. 33. If this is indeed the way binding occurs, the fact that P_E gives a better correlation than π would imply that ρ_3 is similar to ρ_1 (see *Discussion*). The arrangement in Structure X*b* is based on the premise that R′ must fit into ρ_2 and R into the ρ_1 area. This places the α-H in the ρ_3 area and the ester function in the ρ_H below the plane of the page. It is unfortunate that more data points are not available so that more

Table XXI—Correlation of Structure and Activity of Chymotrypsin Inhibitors

Type Inhibitor	"Best" Equation	n	r	s	Conditions	Eq. No.
 from Table XI	$\log 1/I/S = 0.798(\pm0.28)\pi + 0.459(\pm0.45)\sigma + 0.868(\pm0.40)X - 1.964(\pm0.24)$	15	0.913	0.261	0.05 M Tris buffer; pH 7.4; 10% dimethyl sulfoxide	26
Misc. Compds. from Table XII	$\log 1/K_i = 0.950(\pm0.11)\log P - 1.883(\pm0.26)$	10	0.990	0.089	None given	27
	$\log 1/K_i = 0.996(\pm0.10)\log P - 2.596(\pm0.26)$	8	0.995	0.139		28
	$\log 1/K_i = 0.977(\pm0.06)\log P + 0.592(\pm0.12)HB - 2.537(\pm0.18)$	17	0.994	0.111		29
Aromatic acids	$\log 1/K_i = 0.942(\pm0.58)\log P + 0.960(\pm1.78)pKa - 3.660(\pm8.23)$	7	0.917	0.361	0.1 M KCl; 3.3×10^{-3} M Tris buffer; pH 6.9; 25°	30
 from Table XIV	$\log 1/I/S = 1.216(\pm0.85)\pi_R - 1.289(\pm0.51)$	6	0.893	0.297	0.05 M Tris buffer; pH 7.4; 10% dimethyl sulfoxide	31
 from Table XV	$\log 1/K_i = 0.900(\pm0.85)\pi_{R+R'} - 3.824(\pm2.40)$	7	0.771	0.630	In water at 25°; pH 7.90 and 0.10 M NaCl	32
	$\log 1/K_i = 2.874(\pm0.90)\pi_{R'} + 0.089(\pm0.02)P_{E-R} - 6.793(\pm1.49)$	7	0.984	0.193		33
 from Table XVI	$\log K_i = 1.133(\pm0.24)\pi - 1.151(\pm0.53)$	7	0.984	0.242	25°; pH 7.60	34
 from Table XVII	$\log K_i = 1.284(\pm0.32)\pi - 0.643(\pm0.71)$	7	0.978	0.326	25°; pH 7.60	35
C_6H_5COR from Table XVIII	$\log 1/K_i = 0.31(\pm0.37)\pi + 2.53(\pm0.40)$	3	0.996	0.020	0.1 M KCl; $3.3 \times 10^{-3}M$ Tris buffer; pH 6.9; 25°	36
Hydrocarbons from Table XIX	$\log 1/K_i = 1.473(\pm0.43)\log P - 1.209(\pm1.31)$	10	0.942	0.379	0.1 M KCl; $3.3 \times 10^{-3}M$ Tris buffer; pH 6.9; 25°	37

weight could be placed on this equation and the coefficients could be made sharply defined.

Equations 34 and 35, based on data from Tables XVI and XVII, correlate inhibition by two sets of phosphonates. In each set, the authors have omitted the derivative where R = octyl, since, as with Eq. 13, a break occurs at this point. The slopes of Eqs. 34 and 35 are close to those of the other equations where it is expected that binding in the ρ_2 area is occurring. From a comparison of the intercepts of Eqs. 34 and 35, it is clear that the onium compounds are more effective, despite the fact that the onium compounds are, as a whole, less hydrophobic. It may be that the positively charged sulfur aids in binding by interaction with an electron-rich species of the ρ_1 area and, in this fashion, compensates for the lower hydrophobic character.

The three ketones correlated by Eq. 36 are included to show that dependence of inhibitory power on π when the ρ_3 area is presumably involved is essentially the same as for substrates (Eqs. 14–16).

The correlation obtained with the hydrocarbons in Eq. 37 is not as good as that found by Wildnauer and Canady (26) with the molecular area of the inhibitor. The slope of Eq. 37 is somewhat higher than those for other sets binding in the ρ_2 area. Because of the rather large confidence interval on this slope, one cannot be sure this difference is real.

DISCUSSION

While the sets of data analyzed in this review were by no means ideally designed for assaying the relative importance of hydrophobic, electronic, and steric effects of substituents operating on substrates in the ρ_1, ρ_2, and ρ_3 areas of the enzyme, the overall view obtained with grossly different molecular species from many different laboratories gives a consistent picture.

Hydrophobic binding in the ρ_2 area is defined by the slopes of the hydrophobic terms in Eqs. 13, 22, 26, 29, 31, 34, 35, and 37. The mean and standard deviation for the eight values is 1.21 ± 0.23. There does not appear to be any significant difference for the slopes for the substrates and those for the inhibitors. This finding supports the idea of a common area and mechanism of binding for the two classes of reactants. Inhibition viewed in these terms would appear to be simple occupation of the binding site by the inhibitor.

The data on the ρ_3 area are also gratifyingly uniform. Comparable dependence on hydrophobic binding is seen in the coefficients with the hydrophobic terms in Eqs. 14–16 and 36. The mean and standard deviation for these four sets is 0.29 ± 0.1. Thus the importance of hydrophobic binding in the ρ_3 area is about one-fourth that of the ρ_2 area. Again, it is seen that substrates and inhibitors show the same dependence on π.

Steric inhibition of binding by the sp^2 carbon of the carbonyl group can be quite important, as shown by Eqs. 15 and 19–21. It is of interest to compare the coefficients of E_s in Eqs. 15 and 19–21 with those obtained for the hydrolysis of esters under homogeneous conditions. Taft (20) has defined E_s as:

$$\log k/k_0 = \delta E_s \qquad \text{(Eq. 38)}$$

where k_0 refers to the unsubstituted ester and k refers to the rate of acidic hydrolysis of a corresponding substituted acetate: X—CH_2COOR. Taft (20) found a value of δ of 0.30 for the alkaline hydrolysis of acetate esters

$$\begin{matrix} & & \mathrm{O} & \\ & & \| & \\ \mathrm{RCH_2O} & - & \mathrm{C} & -\mathrm{CH_3} \end{matrix}$$

of the type RCH_2OCCH_3. This compares with the value of 0.40 in Eq. 15. Present evidence indicates that the geometry of the transition state for acid and alkaline hydrolysis of esters is the same, so that one would expect the same dependence on δ. In Eqs. 19–21, an average value of δ of 1.5 is found. This compares with the values of 1.4, 1.7, and 1.9 found for the methanolysis, *n*-propanolysis, and isopropanolysis, respectively, of esters in which the R group of RCO_2—β—$C_{10}H_7$ was varied (20). Methanolysis of *l*-menthyl esters ($RCO_2C_{10}H_{16}$) showed a δ value of 1.7. Thus the serine moiety of chymotrypsin seems to make about the same steric demands on substrates as simple alcohols in homogeneous organic reactions. Of course, in Eqs. 19–21, the effect of E_s on the two constants, K_m and k_2, is being considered; while in Eq. 15, it is the effect on K_m that is correlated. The result with Eq. 15 implies that the formation of *ES* as characterized by K_m must involve the conversion of the sp^2 carbon of the ester into an sp^3 carbon in the *ES* complex; that is, since the $sp^2 \rightarrow sp^3$ change is involved in the homogeneous hydrolysis of esters, and since the δ values are so close for the enzymic and homogeneous processes, one would surmise that tetrahedral transition states are involved in each case. Although the steric effects in Eqs. 19–21 are involved with two processes, the result with Eq. 15 suggests that the effect of E_s may be primarily on K_m rather than k_2.

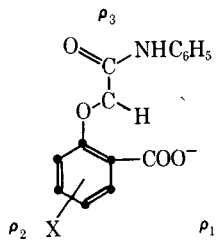
One of the most interesting aspects of this survey is the finding that binding in the ρ_2 area is characterized by π, while that in the ρ_1 area is not well correlated by π but correlates with polarizability as defined by P_E. While such an effect was not anticipated in undertaking the analysis, in retrospect it seems quite logical. This is the area in which a peptide bond of a protein molecule is bound. The binding of such a polarizable group would be facilitated by a polarizing atmosphere.

No doubt, better results would be obtained if, instead of the average polarizability that is embodied in molar refractivity, one could employ directional values. These values for polarization along the three mutually perpendicular directions are at present known for only a relatively few simple species (60).

It is the authors' interpretation of the correlation data that the area must be polar in character and charge interactions aid in holding polarizable groups in this area. One cannot completely rule out desolvation of the acyl moiety with binding in the area since there is some correlation with π; however, this would appear to be of secondary importance. Inhibitors and substrates appear to show the same dependence on P_E (compare Eqs. 22 and 33).

One might expect that the ρ_3 area would also be polar in character. The low coefficients with π in Eqs. 14–16 and 36 suggest that it is not lipophilic. Because only alkyl groups on the congeners are binding in this area, the same quality of correlation is obtained using π or P_E. The results with Eq. 33 in which P_E gives a better correlation than π can be interpreted as in Structure X*a* to mean that the character of the ρ_3 and ρ_1 areas is similar. It would be worthwhile to study a set of congeners binding in this area which has more variance in polarizability.

The partition coefficient of the whole molecule, or even a constant fraction of it, may not be the decisive feature in the binding process. This is illustrated in the recent study by Baker (61) of compounds in Structure XI. With the —COO^- function in the *ortho*-position, an active series of inhibitors is obtained. Only a twofold loss over that of the parent compound occurred. However, placing a —COO^- function in the *para*-position of the phenoxyacetones resulted in a huge loss in activity. That the binding of the phenoxy moiety



XI

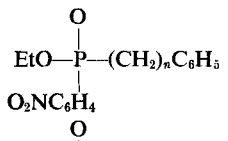
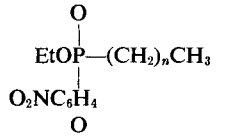
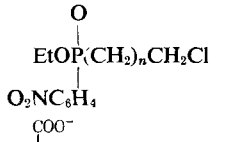
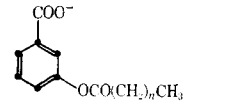
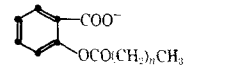
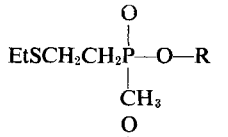
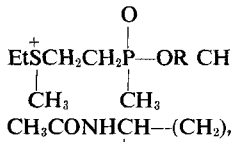
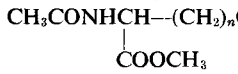
in Structure XI occurs in the ρ_2 area is supported by Eq. 39:

$$\log I_{50} = 0.70(\pm 0.25)\pi + 2.29(\pm 0.18) \quad \overset{n}{4} \quad \overset{r}{0.993} \quad \overset{s}{0.043} \qquad \text{(Eq. 39)}$$

The slope of Eq. 39 is quite close to that of Eq. 26. One finds a constant increase in inhibitory power for each unit of hydrophobicity of X. Apparently, in the *ortho*-position, the —COO^- function can remain free of the area and retain its solvation shell. In fact, it may aid binding by interaction with the polarizing ρ_1 area. This may also be true for the molecules correlated by Eq. 35. However, when in the 4-position, desolvation of the —COO^- appears to be necessary for the phenoxy ring to move into the area.

A point that is not completely clear is the break in activity which occurs when a certain degree of bulkiness or chain length is reached in a set of congeners. This has been mentioned in connection with Eqs. 13, 24, and 35. Other data are available on this point; however, it is not clear whether the break is a function of

Table XXII—Relation between Lipophilic Character of Side Chain and Change in Character of Binding

Compound	π-Value at which Break Occurs	Reaction Constant	Reference
	3.63	I_{50}	62
	3.50	I_{50}	62
	3.39	I_{50}	62
	3.50	Rel. rate hydrolysis	63
	3.50	$V_{max.}$	64
	3.50	K_i	57
	3.50	K_i	57
	2.50	K_m	50

chain length, hydrophobicity, or total molecular volume of the chain. There does seem to be a rather close correlation between the break in activity and the π-value; this is illustrated by the examples in Table XXII. Except for the last example, a close relationship is found between the degree of hydrophobic character and the break in the activity parameter. This break appears to occur at about the same point, regardless of the type functional group involved (*e.g.*, phosphate or acetate). The break also occurs at the same π-value, whether one is considering binding (K_m or K_i) or rate of hydrolysis ($V_{max.}$ or relative rate). This indicates that when a lipophilic moiety of sufficient size is present in the substrate or inhibitor, a decrease in expected activity occurs in a variety of different processes. This could be attributed to a kind of micelle formation[1] on the enzyme of the substrate. While most of the side chains are simple alkyl groups, the two examples where a phenyl and a chlorine moiety are included in the side chain are well in line with the alkyl function, despite the fact that the geometry of the phenyl group, its polarizability, *etc.*, is quite different from the alkane groups. A recheck of Set 8 may be worthwhile since the authors pointed out that a break in activity occurred when the α-alkyl group contained five carbons but, using a cyclohexylmethyl function, gives normal activity (see Table I). This was taken as evidence that chain length was crucial for the area. The other data of Table XXII make this seem unlikely.

Although only a few studies such as the present for chymotrypsin have been made on purified enzymes (16, 18), some data at hand for comparison indicate that other enzymes have hydrophobic pockets which show the same relationship between log P and various activities. Three such examples are:

Inhibition of NADH Oxidase Activity by Barbiturates (45)—

$$\log I_{50} = 1.107 \log P + 1.237 \quad \begin{matrix} n & r & s \\ 6 & 0.921 & 0.261 \end{matrix} \quad \text{(Eq. 40)}$$

Inhibition of Adenosine Deaminase by 9-(1-Hydroxy-2-alkyl)adenines (65)—

$$-\log (I/S)_{0.5} = 0.932\pi - 0.483 \quad \begin{matrix} n & r & s \\ 6 & 0.987 & 0.157 \end{matrix} \quad \text{(Eq. 41)}$$

Relative Rate of Hydrolysis of p-*Nitrophenyl Esters by Serum Esterase (16)*—

$$\log \text{rate} = 0.950 \log P + 3.503 E_s - 0.469 \quad \begin{matrix} n & r & s \\ 6 & 0.976 & 0.497 \end{matrix} \quad \text{(Eq. 42)}$$

In each of the three examples, enzymic activity shows the same dependence on hydrophobic character as defined by the octanol–water reference system. This suggests, but by no means is firm evidence, that the enzymic hydrophobic sites are quite similar to octanol in terms of polarity. This kind of hydrophobic site is quite different from that of serum albumin, hemoglobin, or whole serum when log P is taken as the reference. In many examples of the binding of quite different kinds of organic compounds to serum protein or homogenized tissue, the coefficient with log P or π falls in the range 0.5–0.7 (17). While different ways are used in expressing the binding constants which are not strictly comparable, in 24 such examples (18) a mean and standard deviation of slope of 0.58 ± 0.11 were found.

The importance of having suitable reference standards for determining the relationship between apolar character and mode of interaction of organic compounds with biochemical systems can be further extended. Narcosis under proper conditions is a completely reversible process. Simple binding of organic compounds to nerve tissue causes narcosis, so that the molar concentration of drug producing a standard narcosis can be *roughly* compared to K_m or K_i values from purified enzymes. The following equation illustrates the dependence of such reversible processes on log P.

Tadpole Narcosis by Miscellaneous Organic Compounds—

$$\log 1/C = 0.96 \log P + 0.75 \quad \begin{matrix} n & r & s \\ 44 & 0.967 & 0.303 \end{matrix} \quad \text{(Eq. 43)}$$

In 18 such examples, using various organisms, a mean slope of 1.10 ± 0.14 was found (18). Other processes (17, 18) show the same as well as different dependencies on log P. Too few examples are available at present to draw conclusions of deep significance.

The present study has successfully pulled together a wide variety of data on chymotrypsin substrates and

[1] The authors are indebted to Professor R. Nelson Smith for a number of discussions in which this idea developed.

inhibitors. This extrathermodynamic definition of the characteristics of the areas of interaction in the enzyme is quite consistent, especially when allowance is made for the fact that the information comes from a variety of different laboratories using different experimental techniques with quite a diverse group of organic compounds. It is hoped that this survey of the structure–activity relationship of chymotrypsin will encourage others working in this area to design more suitable substrates and inhibitors which can be used to define the intermolecular interactions more precisely. It seems likely that this approach, tested on chymotrypsin, can be applied to other enzyme substrate or enzyme inhibitor interactions. Also, it seems reasonable to expect that nonenzymic sites of drug action can be mapped using this approach.

REFERENCES

(1) E. J. Ariëns, "Molecular Pharmacology," vol. I, Academic, New York, N. Y., 1964.

(2) M. L. Bender and F. J. Kézdy, *Ann. Rev. Biochem.*, **34,** 49 (1965).

(3) L. Cunningham, in "Comprehensive Biochemistry," vol. 16, M. Florkin and E. H. Stotz, Eds., Elsevier, Amsterdam, The Netherlands, 1965, p. 85.

(4) T. C. Bruice and S. Benkovic, "Bioorganic Mechanisms," vol. I, W. A. Benjamin, New York, N. Y., 1966, p. 212.

(5) D. E. Koshland, in "The Enzymes," vol. I, P. D. Boyer, H. Lardy, and K. Myrbäck, Eds., Academic, New York, N. Y., 1959, chap. 7.

(6) E. L. Smith, in "The Enzymes," vol. IV, P. D. Boyer, H. Lardy, and K. Myrbäck, Eds., Academic, New York, N. Y., 1959, chap. 1.

(7) P. Desnuelle, in "The Enzymes," vol. IV, P. D. Boyer, H. Lardy, and K. Myrbäck, Eds., Academic, New York, N. Y., 1959, chap. 5.

(8) B. R. Baker, "Design of Active-Site-Directed Irreversible Enzyme Inhibitors," Wiley, New York, N. Y., 1967.

(9) W. P. Jencks, "Catalysis in Chemistry and Enzymology," McGraw-Hill, New York, N. Y., 1969, pp. 218–229.

(10) H. Gutfreund and J. M. Sturtevant, *Biochem. J.*, **63,** 656 (1956).

(11) H. Gutfreund and J. M. Sturtevant, *Proc. Nat. Acad. Sci.*, **42,** 719(1956).

(12) G. H. Dixon and H. Neurath, *J. Biol. Chem.*, **225,** 1049 (1957).

(13) H. Neurath and B. S. Hartley, *J. Cell. Comp. Physiol.*, **54,** Suppl. 1, 179(1959).

(14) J. E. Leffler and E. Grunwald, "Rates and Equilibria of Organic Reactions," Wiley, New York, N. Y., 1963.

(15) C. Hansch and T. Fujita, *J. Amer. Chem. Soc.*, **86,** 1616 (1964).

(16) C. Hansch, E. W. Deutsch, and R. N. Smith, *ibid.*, **87,** 2738 (1965).

(17) C. Hansch, *Accounts Chem. Res.*, **2,** 232(1969).

(18) C. Hansch, in "Drug Design," vol. I, E. J. Ariëns, Ed., in the series "Medicinal Chemistry," Academic, New York, N. Y., to be published.

(19) T. Fujita, J. Iwasa, and C. Hansch, *J. Amer. Chem. Soc.*, **86,** 5175(1964).

(20) R. W. Taft, in "Steric Effects in Organic Chemistry," M. S. Newman, Ed., Wiley, New York, N. Y., 1956, p. 644.

(21) E. Kutter and C. Hansch, *J. Med. Chem.*, **12,** 647(1969).

(22) E. Kutter and C. Hansch, *Arch. Biochem. Biophys.*, **135,** 126(1970).

(23) F. Helmer, K. Kiehs, and C. Hansch, *Biochemistry*, **7,** 2858(1968).

(24) C. Hansch and S. M. Anderson, *J. Org. Chem.*, **32,** 2583 (1967).

(25) W. Scholtan, *Arzneim.-Forsch.*, **18,** 505(1968).

(26) R. Wildnauer and W. J. Canady, *Biochemistry*, **5,** 2885 (1966).

(27) M. L. Bender and F. J. Kézdy, *J. Amer. Chem. Soc.*, **86,** 3704(1964).

(28) M. L. Bender, F. J. Kézdy, and C. R. Gunter, *ibid.*, **86,** 3714(1964).

(29) B. R. Baker, in "Design of Active-Site-Directed Irreversible Enzyme Inhibitors," Wiley, New York, N. Y., 1967, p. 129.

(30) G. E. Hein and C. Niemann, *J. Amer. Chem. Soc.*, **84,** 4487(1962); *ibid.*, **84,** 4495(1962).

(31) C. L. Hamilton, C. Niemann, and G. Hammond, *Proc. Nat. Acad. Sci.*, **55,** 664(1966).

(32) J. L. Miles, D. A. Robinson, and W. J. Canady, *J. Biol. Chem.*, **238,** 2932(1963).

(33) P. B. Sigler, D. M. Blow, B. W. Mathews, and R. Henderson, *J. Mol. Biol.*, **35,** 143(1968).

(34) W. B. Lawson and H. J. Schramm, *Biochemistry*, **4,** 377 (1965).

(35) B. R. Baker, "Design of Active-Site-Directed Irreversible Enzyme Inhibitors," Wiley, New York, N. Y., 1967, p. 55.

(36) W. B. Lawson and H. J. Schramm, *J. Amer. Chem. Soc.*, **84,** 2017(1962).

(37) D. S. Sigman and E. R. Blout, *ibid.*, **89,** 1747(1967).

(38) N. K. Schaeffer, S. C. May, and W. H. Summerson, *J. Biol. Chem.*, **202,** 67(1953).

(39) B. S. Hartley and D. L. Kauffman, *Biochem. J.*, **101,** 229 (1966).

(40) Y. Nakagawa and M. L. Bender, *J. Amer. Chem. Soc.*, **91,** 1566 (1969).

(41) G. Schoellman and E. Shaw, *Biochemistry*, **2,** 252(1963).

(42) E. G. Ong, E. Shaw, and G. Schoellman, *J. Biol. Chem.*, **240,** 694(1965).

(43) N. R. Draper and H. Smith, "Applied Regression Analysis," Wiley, New York, N. Y., 1966.

(44) C. K. Hancock, E. A. Meyers, and B. J. Yager, *J. Amer. Chem. Soc.*, **83,** 4211(1961).

(45) C. Hansch and S. M. Anderson, *J. Med. Chem.*, **10,** 745 (1967).

(46) B. Zerner and M. L. Bender, *J. Amer. Chem. Soc.*, **86,** 3669(1964).

(47) S. G. Cohen and J. Crossley, *ibid.*, **86,** 1217(1964).

(48) J. B. Milstien and T. H. Fife, *Biochemistry*, **8,** 623(1969).

(49) A. Leo, C. Hansch, and C. Church, *J. Med. Chem.*, **12,** 766(1969).

(50) J. B. Jones, T. Kunitake, C. Niemann, and G. E. Hein, *J. Amer. Chem. Soc.*, **87,** 1777(1965).

(51) J. B. Jones, C. Niemann, and G. E. Hein, *Biochemistry*, **4,** 1735(1965).

(52) J. R. Rapp, C. Niemann, and G. E. Hein, *ibid.*, **5,** 4100(1966).

(53) J. P. Wolf, III, and C. Niemann, *ibid.*, **2,** 82(1963).

(54) B. R. Baker and J. A. Hurlbut, *J. Med. Chem.*, **10,** 1129 (1967).

(55) G. B. Barlin and D. D. Perrin, *Quart. Rev.*, **20,** 75(1966).

(56) I. V. Berezin, A. V. Levashov, and K. Martinek, "Vth Federated European Biological Meeting, Prague," 1968.

(57) A. A. Aaviksaar, E. V. Rozengart, P. F. Sikk, and R. A. Herbst, *Org. Reactiv.*, **4,** 1059(1968).

(58) A. J. Hymes, D. A. Robinson, and W. J. Canady, *J. Biol. Chem.*, **240,** 134(1965).

(59) D. Agin, L. Hersh, and D. Holtzman, *Proc. Nat. Acad. Sci.*, **53,** 952(1965).

(60) C. K. Ingold, "Structure and Mechanism in Organic Chemistry," 2nd ed., Cornell University Press, Ithaca, N. Y., 1969, p. 152.

(61) B. R. Baker and J. A. Hurlbut, *J. Med. Chem.*, **11,** 1054 (1968).

(62) E. L. Becker, T. R. Fukuto, B. Boone, D. C. Canham, and E. Boger, *Biochemistry*, **2,** 72(1963).

(63) H. J. Hofstee, *Biochim. Biophys. Acta*, **24,** 211(1957).

(64) H. J. Hofstee, *ibid.*, **32,** 182(1959).

(65) H. F. Schaeffer, R. N. Johnson, E. Oldin, and C. Hansch, to be published.

ACKNOWLEDGMENTS AND ADDRESSES

Received from the *Department of Chemistry, Pomona College, Claremont, CA 91711*