Quantitative Structure-Activity Relationship of Chymotrypsin-Ligand Interactions

Corwin Hansch,*

Department of Chemistry, Pomona College, Claremont, California 91711

Ciro Grieco, Carlo Silipo,* and Antonio Vittoria

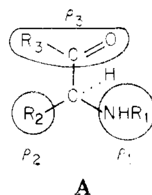
Institute of Pharmceutical and Toxicological Chemistry, University of Naples, Naples, Italy. Received January 31, 1977

Quantitative structure–activity relationships (QSAR) have been formulated for the interactions of a variety of ligands with chymotrypsin. The parameters K_m, k_2, k_3, $k_{cat.}$, and K_i are found to be strongly dependent on molar refractivity as well as steric and electronic character of the substituents in structures of the type $R_2CH(COOR_3)NHCOR_1$ where R may be H. A model for binding of D and L esters is presented which gives a consistent view of the binding step, acylation, and deacylation. The model suggests new avenues for exploration.

One of the great problems in medicinal chemistry is that of developing an understanding of the forces involved in the interaction of organic compounds with the so-called drug-receptor sites in living systems. Except in the case of enzymes, opportunities to work directly with isolated receptors are still severely limited. Our group has become more and more concerned with the formulation of enzymic structure–activity relationships[1] to increase our general understanding of the mechanism of interaction of small molecules with macromolecular systems and, in addition, to learn how to modulate enzyme activity first in vitro and then in vivo. Chymotrypsin is an attractive system, on which we have made some initial studies,[2] because its mechanism of action in the hydrolysis of esters and amides has been extensively investigated[3,4] and is now understood in a general sense. We believe that enzymic QSAR provide valuable insight into enzyme regulation and, therefore, are important in developing lead compounds in drug research.[5] Of greatest interest to us are compounds containing a chiral center which so often occurs in natural products.

The difference in biological activity between stereoisomers has long fascinated bioorganic chemists. Hydrolases such as chymotrypsin, although selective, do show large differences in activity, even with enantiomorphs, and thus lend themselves well to studies of the mode of action of the macromolecular receptors.

Hein and Niemann[6] developed a system of nomenclature which can be used to discuss the space around the site where a molecule containing a chiral center binds. It is illustrated in structure A with the basic L form of the

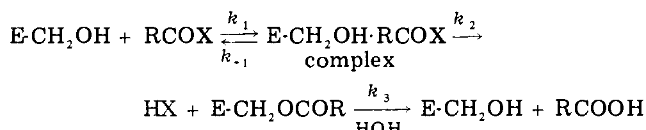


A

α-amino acid skeleton. The symbols ρ_1, ρ_2, and ρ_3 refer to enzymic space into which the three groups attached to the α-carbon atom fall. The α-H, which in the above drawing falls below the plane of the page, is said to be in ρ_H space.

Our interest in this report is to make a sharper definition of ρ_1, ρ_2, and ρ_3 space by means of correlation analysis[1b] of the perturbations of K_i, K_m, k_2, k_3, and $k_{cat.}$ by the substituents R_1, R_2, and R_3. In our first such attempt[2a] we reached the conclusion that binding in ρ_1 space correlated with MR (molar refractivity of the substituent[7]) and, hence, was not truly hydrophobic in character. It was also concluded at that time that binding in ρ_2 space correlated with the hydrophobic parameter π and suggested hydrophobic space. Unfortunately, for the data

Scheme I



$$\text{E-CH}_2\text{OH} + \text{RCOX} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \underset{\text{complex}}{\text{E-CH}_2\text{OH}\cdot\text{RCOX}} \xrightarrow{k_2}$$

$$\text{HX} + \text{E-CH}_2\text{OCOR} \xrightarrow[\text{HOH}]{k_3} \text{E-CH}_2\text{OH} + \text{RCOOH}$$

considered at that time, π and MR were highly collinear vectors for R_2. In later work[2e] with amide sets where π and MR for R_2 were more orthogonal, it became clear that binding in ρ_2 space is better correlated by MR than π. Niemann and his colleagues also noted correlation between binding and MR for a very limited set of congeners.[8]

The meaning behind the correlation of binding with MR is not entirely clear. One can simply say (when π and MR are reasonably orthogonal) that MR space is different from π space. Our present working hypothesis is that MR space consists of predominantly polar amino acid residues.

One faces a dilemma in the development of correlation equations via regression analysis. One can either study small sets of closely related congeners with the aim of obtaining tight correlations with low standard deviations, or one can attempt to develop QSAR encompassing large sets of congeners of more diverse structure. In the latter case, one is usually forced to accept data of varying quality from a variety of sources. In any case, the greater the diversity of structures, the poorer the correlation is apt to be. This problem has recently been discussed and analyzed by Exner.[10] In this report we are more interested in taking the broad view to obtain a general "map" of the "forest" without being too concerned about the "individual or even small groves of trees".

The so-called double displacement mechanism of hydrolysis of chymotrypsin can be written as shown in Scheme I. In Scheme I, $-CH_2OH$ is the nucleophilic moiety of Ser-195 of the enzyme. Chymotrypsin reacts with a wide variety of amides, peptides, and esters to first form the Michaelis complex which is followed by acylation of the enzyme. The acylated enzyme then reacts with water or another nucleophile to yield the regenerated enzyme and an acid.

The complex parameters which we wish to discuss are K_m and $k_{cat.}$.

$$K_m = K_s \frac{k_3}{k_2 + k_3} \qquad k_{cat.} = \frac{k_2 k_3}{k_2 + k_3}$$

Relatively little data are available for which K_s ($K_s = k_{-1}/k_1$) has been evaluated (see ref 11) so that we shall be concerned only with $K_{m(app)}$.

$$K_{m(app)} = \frac{k_{-1} + k_2}{k_1}$$

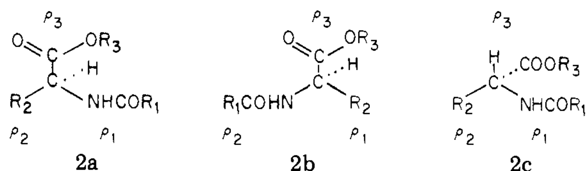
If one can make the assumption that the catalytic steps are very slow compared to the formation of the enzyme–substrate complex, then $1/K_m$ can be viewed as a simple binding constant. Neurath and Hartley[12] first presented evidence to show that this appears to hold for chymotrypsin. The difficulty of considering K_m as a meaningful constant has been discussed by Bender and Kēzdy.[3]

QSAR of the Michaelis Constant [$K_{m(app)}$]. In our phenomenological approach we shall assume for a first approximation that $1/K_m$ can be taken as a simple binding constant. We have formulated eq 1 from the data of Table

$$\log 1/K_m = 1.09\ (\pm 0.11)\ \text{MR-2} + 0.80\ (\pm 0.11)\ \text{MR-1} + 0.52\ (\pm 0.13)\ \text{MR-3} - 0.63\ (\pm 0.26)\ I\text{-1} + 1.26\ (\pm 0.28)\ \sigma^* - 0.057\ (\pm 0.013)\ \text{MR-1}\cdot\text{MR-2}\cdot\text{MR-3} - 1.61\ (\pm 0.47) \quad (1)$$

$n = 71,\ r = 0.979,\ s = 0.332$

I on the L form of N-acyl esters undergoing hydrolysis by chymotrypsin. In this equation, MR-1, MR-2, and MR-3 refer to the molar refractivities of $-NHCOR_1$ (or $-NHSO_2R_1$), R_2, and OR_3 falling in ρ space as shown in structures in **2a–c**. I-1 is an indicator variable assigned



L ester as substrate D ester as substrate D ester as inhibitor

the value of 1 when $R_2 = -CH(CH_3)_2$ and the value of 0 for all other cases. The Taft polar parameter (σ^*) applies only to R_3. The figures in parentheses are the 95% confidence intervals, n represents the number of data points used to derive the equation, r is the correlation coefficient, and s is the standard deviation from the regression. The positive coefficients with the MR terms show that binding in ρ_1, ρ_2, and ρ_3 space is proportional to the molar refractivity of the groups falling in these regions up to the point where the cross-product term becomes significant. Substituting the hydrophobic parameters π-1, π-2, and π-3 for the corresponding MR constants produces a much poorer equation ($r = 0.928$, $s = 0.611$). It is apparent from the correlation matrix (Table III) that π-3 and MR-3 are so collinear that no decision can be made about the character of ρ_3 space from this equation. Although MR-1 and MR-2 are rather collinear with π-1 and π-2, the results of eq 1, taken with our correlation on amides[2e] where π and MR are not so collinear, strongly suggest that ρ_1 and ρ_2 space are not typically hydrophobic. This assumes, of course, that correlation with π is diagnostic of hydrophobic space. Interaction in ρ_2 space is stronger than interaction in ρ_1 space (per unit of MR), also noted with amides, and binding in ρ_3 space is weakest of all. Note that the MR values used in this report have been scaled by 0.1 to make them more equiscalar with π.

The negative coefficient with I-1 brings out the fact that, other factors being equal, valine derivatives are about four times (antilog of 0.62) more poorly bound than the other R_2 groups of Table I. The positive coefficient with σ^* shows that electron-withdrawing groups on R_3 favor binding. There is considerable collinearity (Table III) between σ^* and MR-3 so that in fact σ^* may not be quite as important as its position in Table II would suggest. A better selection of R_3 with respect to σ^* should be investigated.

A most interesting term is the cross product MR-1·MR-2·MR-3. Its negative coefficient reveals the fact that placing too much bulk in ρ_1, ρ_2, or ρ_3 space results in poorer binding. This appears to occur with one large group in a given site or by the combination of two or three smaller groups in two or three sites; that is, the three spaces, ρ_1, ρ_2, and ρ_3, are interrelated. Crowding in one space affects binding in the others.

Equation 1 accounts for 96% of the variance in $\log 1/K_m$ which we feel is about as good a fit of the data as one can expect since the K_m values are from many different laboratories and have been obtained under a variety of experimental conditions. In some instances, varying amounts of organic solvents have been added to aid solution of solutes and parameters have been measured at different pH values. K_m is relatively insensitive for neutral substrates in the pH range 5–8; however, k_2 and k_3 are quite sensitive to pH. Equation 1 contains six variables so that there is a ratio of about 12 data points/variable. Chance correlation[14] is diminished beyond reasonable doubt. Table II displays the development of eq 1 and the statistical justification (F statistic) of each of the variables. While there are a few rather poorly correlated points which produce a somewhat higher standard deviation than one would like to see, eq 1 does bring order to a large amount of structural variation in 71 esters. We shall use it as an archetypal QSAR for comparison with less well developed chymotrypsin QSAR.

One data point (**72**) in Table I was not used in the formulation of eq 1; this N-formyl analogue, as well as a few other formyl analogues, is not well predicted by our equations. Not enough of these points are available for special parameterization to search out the nature of this effect. This is an area worthy of further study.

We have not employed charged ligands (e.g., lysine analogues) in the development of eq 1. While lysine analogues do no fit eq 1 well, it is noteworthy that compounds of the type[13] $C_6H_5CH_2CH(NH_3^+)COOR$ do fit.

Table IV contains data on the same type of esters as those correlated by eq 1 except that these have the unnatural D configuration. Although the geometry is unnatural, they are still capable of being hydrolyzed by the enzyme. The QSAR is given in eq 2. For compounds 1–8,

Table IV, compounds 1–8

$$\log 1/K_m = 1.38\ (\pm 0.57)\ \text{MR} - 2.49\ (\pm 1.8) \quad (2a)$$

$n = 8,\ r = 0.924,\ s = 0.234$

Table IV, compounds 9–15

$$\log 1/K_m = 0.47\ (\pm 0.41)\ \text{MR-2} + 4.87\ (\pm 0.71) \quad (2b)$$

$n = 7,\ r = 0.792,\ s = 0.302$

Table IV, compounds 1–15

$$\log 1/K_m = 0.47\ (\pm 0.31)\ \text{MR-2} + 1.38\ (\pm 0.59)\ \text{MR-1} + 1.83\ (\pm 0.72)\ I\text{-1} + 2.76\ (\pm 1.9) \quad (2c)$$

$n = 15,\ r = 0.993,\ s = 0.267$

R_2 and R_3 are constant; hence the single-variable eq 2a correlates the data. For compounds **9–15**, R_1 is constant so that the data are correlated by eq 2b. Although eq 2b is statistically significant ($F_{1,5} = 8.7$; $F_{1,5;\alpha=0.05} = 6.6$), the correlation is poor. Data point 11 is quite poorly fit. It if is dropped, one obtains the same slope but now $r = 0.880$ and $s = 0.235$. For convenience in comparative purposes, eq 2a and 2b can be merged via an indicator variable into eq 2c.

The correlation coefficient for eq 2c has little real meaning since, as we have noted in another instance,[9] it

Table I. Parameters Used in the Formulation of Eq 1

No.	Log $1/K_m$ Obsd	Log $1/K_m$ Calcd	$\lvert\Delta \log 1/K_m\rvert$	MR-1	MR-2	MR-3	I-1	σ^*	R_1	R_2	R_3	Ref
1	0.21	0.55	0.34	1.49	0.56	0.79	0	0.00	L-NHCOMe	Me	OMe	19d
2	0.60	0.64	0.04	1.49	0.56	1.25	0	−0.10	L-NHCOMe	Me	OEt	19h
3	0.75	1.00	0.25	1.49	1.50	1.71	1	−0.19	L-NHCOMe	i-C_3H_7	O-i-C_3H_7	19g
4	0.88	1.56	0.68	2.80	0.56	0.79	0	0.00	L-NHCO-furyl-H_4	Me	OMe	19c
5	0.95	0.88	0.07	1.49	1.50	0.79	1	0.00	L-NHCOMe	i-C_3H_7	OMe	19a
6	0.96	0.93	0.03	1.49	1.50	1.25	1	−0.10	L-NHCOMe	i-C_3H_7	OEt	19g
7	1.28	1.03	0.25	1.49	1.03	0.79	0	0.00	L-NHCOMe	Et	OMe	19a
8	1.31	1.46	0.15	2.67	0.56	0.79	0	0.00	L-NHCO-furyl	Me	OMe	19c
9	1.37	1.24	0.13	1.98	1.50	0.79	1	0.00	L-NHCOCH_2Cl	i-C_3H_7	OMe	19b
10	1.43	1.86	0.43	3.18	0.56	0.79	0	0.00	L-NHCO-3-pyridyl	Me	OMe	19c
11	1.51	1.81	0.30	1.49	1.75	1.25	0	−0.10	NHCOMe[b]	COOEt	OEt	19m
12	1.51	1.63	0.12	1.49	1.50	1.71	0	−0.19	L-NHCOMe	C_3H_7	O-i-C_3H_7	19g
13	1.54	1.86	0.32	3.18	0.56	0.79	0	0.00	L-NHCO-4-pyridyl	Me	OMe	19c
14	1.64	2.16	0.52	1.49	2.11	1.25	0	−0.10	L-NHCOMe	CH_2COOEt	OEt	19h
15	1.72	1.74	0.02	1.49	1.50	1.75	1	0.39	L-NHCOMe	i-C_3H_7	OCH_2CH_2Cl	19g
16	1.74	1.86	0.12	3.18	0.56	0.79	0	0.00	L-NHCO-2-pyridyl	Me	OMe	19c
17	1.82	1.94	0.12	3.28	0.56	0.79	0	0.00	L-NHCO-2-thienyl	Me	OMe	19c
18	1.99	1.51	0.48	1.49	1.50	0.79	0	0.00	L-NHCOMe	C_3H_7	OMe	19a
19	2.01	2.08	0.07	3.46	0.56	0.79	0	0.00	L-NHCOPh	Me	OMe	19c
20	2.17	1.98	0.19	1.49	1.96	0.79	0	0.00	L-NHCOMe	C_4H_9	OMe	19a
21	2.22	2.14	0.08	3.46	0.56	1.25	0	−0.10	L-NHCOPh	Me	OEt	19j
22	2.30	1.87	0.43	1.98	1.50	0.79	0	0.00	L-NHCOCH_2Cl	C_3H_7	OMe	19b
23	2.33	2.42	0.09	3.90	0.56	0.79	0	0.00	L-NHCOPh-2-NH_2	Me	OMe	19c
24	2.34	2.32	0.02	3.46	1.50	0.79	1	0.00	L-NHCOPh	i-C_3H_7	OMe	19b
25	2.42	1.98	0.44	1.49	1.96	0.79	0	0.00	L-NHCOMe	i-C_4H_9	OMe	19a
26	2.48	2.58	0.10	1.49	2.54	1.25	0	−0.10	L-NHCOMe	Ph	OEt	19m
27	2.53	2.92	0.39	1.49	2.89	0.79	0	0.00	L-NHCOMe	C_6H_{13}	OMe	19a
28	2.76	3.14	0.38	1.49	3.00	2.17	0	−0.21	L-NHCOMe	CH_2Ph	R-O-sec-C_4H_9	19n
29	2.79	2.45	0.34	1.49	2.42	0.79	0	0.00	L-NHCOMe	C_5H_{11}	OMe	19a
30	2.85	2.51	0.34	3.46	1.03	0.79	0	0.00	L-NHCOPh	Et	OMe	19j
31	2.90	3.04	0.14	1.49	3.00	0.79	0	0.00	L-NHCOMe	CH_2Ph	OMe	19a
32	2.96	3.03	0.07	1.49	3.00	1.25	0	−0.10	L-NHCOMe	CH_2Ph	OEt	19f
33	3.04	3.14	0.10	1.49	3.00	2.17	0	−0.21	L-NHCOMe	CH_2Ph	S-O-sec-C_4H_9	19n
34	3.07	2.95	0.12	3.46	1.50	0.79	0	0.00	L-NHCOPh	C_3H_7	OMe	19b
35	3.10	2.37	0.72	1.49	1.50	1.75	0	0.39	L-NHCOMe	C_3H_7	OCH_2CH_2Cl	19g
36	3.15	3.21	0.06	1.49	3.18	1.25	0	−0.10	L-NHCOMe	CH_2Ph-4-OH	OEt	19f
37	3.28	3.51	0.23	1.49	3.00	3.60	0	−0.21	L-NHCOMe	CH_2Ph	S-O-CH(Me)-c-C_6H_{11}	19n
38	3.31	3.51	0.20	1.49	3.00	3.60	0	−0.21	L-NHCOMe	CH_2Ph	R-O-CH(Me)-c-C_6H_{11}	19n
39	3.31	3.33	0.02	1.49	1.50	3.40	1	1.14	L-NHCOMe	i-C_3H_7	OPh-4-NO_2	19aa
40	3.32	3.22	0.10	1.49	0.56	3.40	0	1.14	L-NHCOMe	Me	OPh-4-NO_2	19aa
41	3.38	4.00	0.62	2.67	3.18	0.79	0	0.00	L-NHCO-furyl	CH_2Ph-4-OH	OMe	19j
42	3.49	3.22	0.27	1.49	3.18	0.79	0	0.00	L-NHCOMe	CH_2Ph-4-OH	OMe	19i
43	3.66	3.28	0.38	5.01	0.56	0.79	0	0.00	L-NHCO-2-quinolyl	Me	OMe	19c
44	3.72	3.17	0.55	1.49	3.13	0.79	0	0.00	L-NHCOMe	CH_2-c-C_6H_{11}	OMe	19a
45	4.02	4.29	0.27	1.49	4.23	0.79	0	0.00	L-NHCOMe	CH_2-indolyl	OMe	19k
46	4.05	4.24	0.19	1.49	4.23	1.25	0	−0.10	L-NHCOMe	CH_2-indolyl	OEt	19f
47	4.22	3.85	0.36	1.49	3.00	3.36	0	0.11	L-NHCOMe	CH_2Ph	S-O-CH(Me)Ph	19n
48	4.30	3.85	0.44	1.49	3.00	3.36	0	0.11	L-NHCOMe	CH_2Ph	R-O-CH(Me)Ph	19n
49	4.46	4.35	0.11	3.46	3.00	0.79	0	0.00	L-NHCOPh	CH_2Ph	OMe	19j
50	4.57	4.42	0.15	1.82	3.00	2.77	0	0.60	L-NHSO_2Me	CH_2Ph	OPh	19e
51	4.62	5.16	0.54	1.49	3.00	3.40	0	1.14	L-NHCOMe	CH_2Ph	OPh-4-NO_2	19k
52	4.66	4.34	0.32	3.46	3.18	1.25	0	−0.10	L-NHCOPh	CH_2Ph-4-OH	OEt	19j
53	4.70	5.08	0.38	4.19	0.56	3.40	0	1.14	L-NHCOOCH_2Ph	Me	OPh-4-NO_2	19l
54	4.70	4.36	0.34	1.82	3.00	3.33	0	0.46	L-NHSO_2Me	CH_2Ph	OPh-4-Me	19e
55	4.74	4.51	0.23	3.46	3.18	0.79	0	0.00	L-NHCOPh	CH_2Ph-4-OH	OMe	19j
56	4.75	4.71	0.04	4.19	1.50	3.40	1	1.14	L-NHCOOCH_2Ph	i-C_3H_7	OPh-4-NO_2	19l
57	4.82	4.26	0.56	1.82	3.00	3.45	0	0.36	L-NHSO_2Me	CH_2Ph	OPh-4-OMe	19e
58	4.87	5.34	0.47	4.19	1.49	3.40	0	1.14	L-NHCOOCH_2Ph	CH_2CONH_2	OPh-4-NO_2	19l
59	4.88	4.72	0.16	1.82	3.00	3.27	0	0.75	L-NHSO_2Me	CH_2Ph	OPh-4-Cl	19e
60	4.89	5.01	0.12	1.82	3.00	3.78	0	0.90	L-NHSO_2Me	CH_2Ph	OPh-4-COMe	19e
61	5.06	5.17	0.11	1.82	3.00	3.40	0	1.09	L-NHSO_2Me	CH_2Ph	OPh-3-NO_2	19e
62	5.24	5.47	0.23	4.19	1.96	3.40	0	1.14	L-NHCOOCH_2Ph	C_4H_9	OPh-4-NO_2	19l
63	5.32	5.47	0.15	4.19	1.96	3.40	0	1.14	L-NHCOOCH_2Ph	i-C_4H_9	OPh-4-NO_2	19l
64	5.37	5.23	0.13	1.82	3.00	3.40	0	1.14	L-NHSO_2Me	CH_2Ph	OPh-4-NO_2	19e
65	5.40	5.66	0.26	4.19	4.23	3.27	0	0.75	L-NHCOOCH_2Ph	CH_2-indolyl	OPh-4-Cl	19o
66	5.57	5.34	0.23	4.19	1.50	3.40	0	1.14	L-NHCOOCH_2Ph	C_3H_7	OPh-4-NO_2	19l
67	5.68	5.60	0.08	4.19	4.23	3.78	0	0.90	L-NHCOOCH_2Ph	CH_2-indolyl	OPh-4-COMe	19o
68	5.70	6.14	0.44	1.49	4.23	3.40	0	1.14	L-NHCOMe	CH_2-indolyl	OPh-4-NO_2	19k
69	5.74	5.21	0.53	4.19	1.03	3.40	0	1.14	L-NHCOOCH_2Ph	Et	OPh-4-NO_2	19l
70	5.92	6.09	0.17	4.19	4.23	3.40	0	1.14	L-NHCOOCH_2Ph	CH_2-indolyl	OPh-4-NO_2	19o
71	6.32	5.75	0.57	4.19	3.00	3.40	0	1.14	L-NHCOOCH_2Ph	CH_2Ph	OPh-4-NO_2	19l
72	4.88[a]	2.90	1.98	1.03	0.56	3.40	0	1.14	L-NHCOH	Me	OPh-4-NO_2	19l

[a] This point not used in deriving eq 1. [b] Symmetric compound; see ref 19m.

Table II. Development of Eq 1

Intercept	σ*	MR-2	MR-1	MR-1·2·3	MR-3	*I*-1	*r*	*s*	$F_{1,X}$[a]
2.62	2.28						0.728	1.08	77.6
1.18	2.05	0.70					0.888	0.734	82.9
−0.14	1.60	0.84	0.48				0.938	0.557	51.1
−0.81	1.95	1.08	0.66	−0.034			0.949	0.508	14.5
−1.93	1.20	1.14	0.85	−0.059	0.54		0.972	0.386	49.4
−1.62	1.26	1.09	0.80	−0.057	0.52	−0.63	0.979	0.332	23.7

[a] $F_{1,60;\alpha=0.001} = 12$.

Table III. Squared Correlation Matrix of Variables Associated with Eq 1

	MR-1	MR-2	MR-3	π-1	π-2	π-3	σ*	*I*-1	MR-1·2·3	MR-L	MR-S
MR-1	1.00	0.05	0.02	0.88	0.13	0.00	0.14	0.01	0.18	0.52	0.05
MR-2		1.00	0.12	0.04	0.69	0.14	0.02	0.04	0.31	0.10	0.54
MR-3			1.00	0.01	0.12	0.91	0.61	0.00	0.49	0.09	0.11
π-1				1.00	0.13	0.00	0.12	0.00	0.17	0.41	0.07
π-2					1.00	0.14	0.01	0.02	0.16	0.02	0.25
π-3						1.00	0.43	0.00	0.41	0.07	0.10
σ*							1.00	0.00	0.39	0.14	0.08
I-1								1.00	0.02	0.10	0.00
MR-1·2·3									1.00	0.26	0.60
MR-L										1.00	0.07
MR-S											1.00

Table IV. Parameters Used in the Formulation of Eq 2

No.	Log $1/K_m$ Obsd	Calcd	$\lvert\Delta \log 1/K_m\rvert$	MR-1	MR-2	MR-3	*I*-1	R_1	R_2	R_3	Ref
1	1.21	1.38	0.17	2.80	0.56	0.79	0	D-NHCO-furyl-H_4	Me	OMe	19c
2	1.31	1.20	0.11	2.67	0.56	0.79	0	D-NHCO-furyl	Me	OMe	19c
3	1.57	1.90	0.33	3.18	0.56	0.79	0	D-NHCO-4-pyridyl	Me	OMe	19c
4	1.77	1.90	0.13	3.18	0.56	0.79	0	D-NHCO-2-pyridyl	Me	OMe	19c
5	2.16	2.04	0.12	3.28	0.56	0.79	0	D-NHCO-2-thienyl	Me	OMe	19c
6	2.22	1.90	0.32	3.18	0.56	0.79	0	D-NHCO-3-pyridyl	Me	OMe	19c
7	2.48	2.29	0.19	3.46	0.56	0.79	0	D-NHCOPh	Me	OMe	19c
8	2.80	2.90	0.10	3.90	0.56	0.79	0	D-NHCOPh-2-NH_2	Me	OMe	19c
9	5.08	5.13	0.05	4.19	0.56	3.40	1	D-$NHCOOCH_2Ph$	Me	OPh-4-NO_2	19l
10	5.13	5.58	0.45	4.19	1.50	3.40	1	D-$NHCOOCH_2Ph$	*i*-C_3H_7	OPh-4-NO_2	19l
11	5.50	5.58	0.08	4.19	1.50	3.40	1	D-$NHCOOCH_2Ph$	C_3H_7	OPh-4-NO_2	19l
12	5.62	5.79	0.17	4.19	1.96	3.40	1	D-$NHCOOCH_2Ph$	*i*-C_4H_9	OPh-4-NO_2	19l
13	5.71	5.36	0.35	4.19	1.03	3.40	1	D-$NHCOOCH_2Ph$	Et	OPh-4-NO_2	19l
14	5.84	5.57	0.27	4.19	1.49	3.40	1	D-$NHCOOCH_2Ph$	CH_2CONH_2	OPh-4-NO_2	19l
15	6.42	6.29	0.13	4.19	3.00	3.40	1	D-$NHCOOCH_2Ph$	CH_2Ph	OPh-4-NO_2	19l
16	5.21[a]	0.76	4.45	1.03	0.56	3.40	1	D-NHCOH	Me	OPh-4-NO_2	19l
17	3.60[a]	0.08	3.52	1.03	3.00	0.79	0	D-NHCOH	CH_2Ph	OMe	19d

[a] These points not used in deriving eq 2.

Table V. Development of Eq 2

Intercept	*I*-1	MR-1	MR-2	*r*	*s*	$F_{1,X}$[a]
1.94	3.67			0.967	0.519	187
−2.49	2.31	1.38		0.985	0.362	14.7
−2.76	1.83	1.38	0.47	0.993	0.267	11.0

[a] $F_{1,11;\alpha=0.005} = 12.2$.

is a by-product of merging two sets of data rather far apart in data space. This great artificial increase in variance which is largely accounted for by *I*-1 (see Table V) tends to confuse the issue which is more clearly seen in eq 2a and 2b. Equation 2 is very important because it supports Niemann's concept of "wrong-way" binding. While the 2b side of the equation is not a sharp correlation, it strongly suggests that R_2 is not binding in ρ_2 space. Taken alone, eq 2 would be of little value; however, with strong support of eq 2a and the comparison with eq 1, it clearly points out an area of great importance for further study.

It is clear from the correlation matrix of Table VI that π-2 and MR-2 are reasonably orthogonal, while MR-1 and π-1 are compromised as independent variables by high

Table VI. Squared Correlation Matrix of Variables Considered in the Derivation of Eq 2

	MR-1	MR-2	MR-3	π-1	π-2	*I*-1	σ*
MR-1	1.00	0.41	0.78	0.59	0.10	0.78	0.78
MR-2		1.00	0.52	0.39	0.41	0.52	0.52
MR-3			1.00	0.75	0.12	1.00	1.00
π-1				1.00	0.09	0.75	0.75
π-2					1.00	0.12	0.12
I-1						1.00	1.00
σ*							1.00

collinearity. Substituting π-1 and π-2 for MR-1 and MR-2 in eq 2 gives a poorer correlation ($r = 0.968$, $s = 0.555$), again pointing to the nonhydrophobic character of ρ_1 and ρ_2 space.

The D esters of Table VII acting as inhibitors lead to the QSAR of eq 3 which, unfortunately, is based on only four

$$\log 1/K_i = 1.42\,(\pm 0.25)\,\text{MR-2} + 1.07\,(\pm 0.27)\,\text{MR-1} - 0.16\,(\pm 0.08)\,\text{MR-1}\cdot\text{MR-2}\cdot\text{MR-3} - 2.71\,(\pm 0.81) \quad (3)$$

$n = 12$, $r = 0.988$, $s = 0.207$

Table VII. Parameters Used in the Formulation of Eq 3

No.	Log $1/K_i$ Obsd	Calcd	$\lvert\Delta \log 1/K_i\rvert$	MR-1	MR-2	R_1	R_2	R_3	Ref
1	0.71	0.73	0.02	1.49	1.50	D-NHCOMe	i-C_3H_7	OMe	19b
2	0.77	1.16	0.39	1.98	1.50	D-NHCOCH$_2$Cl	i-C_3H_7	OMe	19b
3	0.95	0.73	0.22	1.49	1.50	D-NHCOMe	C_3H_7	OMe	19b
4	1.23	1.16	0.07	1.98	1.50	D-NHCOCH$_2$Cl	C_3H_7	OMe	19b
5	2.28	2.45	0.17	3.46	1.50	D-NHCOPh	i-C_3H_7	OMe	19b
6	2.40	2.44	0.04	1.49	3.18	D-NHCOMe	CH$_2$Ph-4-OH	OEt	19p
7	2.64	2.58	0.06	1.49	3.00	D-NHCOMe	CH$_2$Ph	OMe	19b
8	2.75	2.45	0.30	3.46	1.50	D-NHCOPh	C_3H_7	OMe	19b
9	3.10	3.15	0.05	3.18	3.18	D-NHCO-3-pyridyl	CH$_2$Ph-4-OH	OEt	19p
10	3.24	3.15	0.09	1.49	4.23	D-NHCOMe	CH$_2$-indolyl	O-i-C_3H_7	19q
11	3.60	3.62	0.02	1.49	4.23	D-NHCOMe	CH$_2$-indolyl	OEt	19q
12	4.05	4.09	0.04	1.49	4.23	D-NHCOMe	CH$_2$-indolyl	OMe	19q

Table VIII. Development of Eq 3

Intercept	MR-2	MR-1	MR-1·2·3	r	s	$F_{1,X}$[a]
0.25	0.80			0.839	0.653	23.9
−1.64	0.98	0.69		0.954	0.379	20.7
−2.71	1.42	1.07	−0.16	0.988	0.207	22.2

[a] $F_{1,8;\alpha=0.005} = 14.7$.

data points/variable. Qualitatively, the coefficients of eq 3 parallel those of eq 1 although they are all somewhat larger than those of eq 1. The addition of a term in MR-3 to eq 3 does not improve the correlation. MR-3 is little more than an indicator variable for these data since, outside of the cases where R_3 = OMe or OEt, there is only one other example of R_3 (i.e., isopropyl). The fact that MR-3 only occurs in the cross-product term with a negative coefficient means that it has only a negative influence on binding. Since the relative size of the coefficients with MR-2 and MR-1 parallels those of eq 1, this suggests that binding may be occurring as in structure **2c**. R_3 would be forced into ρ_H space in this mode of binding. It is logical to expect ρ_H space to be limited so that a negative coefficient with MR-3 is not unexpected. A better selection of R_3 functions must be studied before eq 3 can be taken seriously.

An alternative QSAR for the D esters of Table III is formulated in eq 4. The standard deviation of eq 4 is

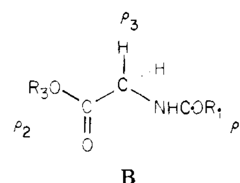
$$\log 1/K_i = 1.18 (\pm 0.25) \text{ MR-2} + 0.73 (\pm 0.27) \text{ MR-1} - 1.06 (\pm 0.92) \text{ MR-3} - 1.20 (\pm 0.96) \quad (4)$$
$$n = 12, r = 0.976, s = 0.292$$

much higher than eq 3 so that one tends to favor eq 3; however, with more data with a better range of R_3 groups, it is likely that an MR-3 term could be added to eq 3. The development of eq 3 is shown in Table VIII and the correlation matrix is given in Table IX. Substitution of π-1, π-2, and π-3 for the corresponding MR terms in eq 3 yields a much poorer correlation (r = 0.935, s = 0.476).

Another type of ester whose hydrolysis by chymotrypsin has been studied extensively is the glycinates of Table X whose QSAR is

$$\log 1/K_m = 0.48 (\pm 0.10) \text{ MR-1} + 0.69 (\pm 0.10) \text{ MR-3} + 0.44 (\pm 0.23) \sigma^* - 0.20 (\pm 0.30) \quad (5)$$
$$n = 42, r = 0.981, s = 0.235$$

Comparing eq 5 with our archetypal QSAR (eq 1) suggests that R_3 might bind in ρ_2 space and that NHCOR$_1$ binds in ρ_1 space as shown in structure B. This picture of



B

binding is speculative and, since the normal chiral center is absent in the glycinates, it is more hazardous to use eq 5 to assess the character of ρ_1 and ρ_2 space. In any case, it is shown in Table XI that the high collinearity between MR-3 and π-3 precludes any judgment about the space with which these vectors are concerned. The variables MR-1 and π-1 are less collinear. The poorer correlation (r = 0.929, s = 0.445) when π-1 and π-3 replace MR-1 and MR-3 in eq 5 again suggests nonhydrophobic space.

The possibility was explored that the addition of cross-product or exponential terms in MR-1 and MR-3 to eq 5 would improve the correlation (Table XII); no improvement was found, as might be anticipated from eq 1. The cross-product term in this equation shows that the ρ_1, ρ_2, and ρ_3 spaces are interrelated and that placing less bulk in one allows greater bulk in another. Since the glycines are missing groups larger than H for R_2, more bulk can be accommodated by the two binding areas.

Equation 6 has been derived from the parameters of

$$\log 1/K_i = 1.06 (\pm 0.20) \text{ MR} - 1.49 (\pm 0.85) \quad (6)$$
$$n = 7, r = 0.987, s = 0.132$$

Table XIII for aromatic acids acting as inhibitors.

Table IX. Squared Correlation Matrix of Variables Considered in the Derivation of Eq 3

	MR-1	MR-2	MR-3	π-1	π-2	π-3	MR-1·2·3
MR-1	1.00	0.16	0.03	0.90	0.20	0.03	0.03
MR-2		1.00	0.47	0.28	0.87	0.49	0.50
MR-3			1.00	0.10	0.29	0.98	0.71
π-1				1.00	0.30	0.10	0.00
π-2					1.00	0.29	0.31
π-3						1.00	0.75
MR-1·2·3							1.00

Table X. Parameters Used in the Formulation of Eq 5

No.	Log $1/K_m$ Obsd	Log $1/K_m$ Calcd	$\lvert\Delta \log 1/K_m\rvert$	MR-1	MR-3	σ^*	R_1	R_3	Ref
1	1.02	1.34	0.32	1.49	1.25	−0.10	NHCOMe	OEt	19t
2	1.07	1.21	0.14	1.79	0.79	0.00	$NHSO_2CH_3$	OMe	19r
3	1.16	1.30	0.14	1.98	0.79	0.00	$NHCOCH_2Cl$	OMe	19r
4	1.19	1.38	0.19	2.13	0.79	0.00	$NHCOOC_2H_5$	OMe	19r
5	1.27	1.04	0.23	1.43	0.79	0.00	$NHCOCF_3$	OMe	19r
6	1.34	1.52	0.18	2.42	0.79	0.00	NHCO-*i*-C_3H_7	OMe	19r
7	1.42	1.29	0.12	1.96	0.79	0.00	NHCOEt	OMe	19r
8	1.50	1.89	0.39	3.18	0.79	0.00	NHCO-3-pyridyl	OMe	19s
9	1.51	1.07	0.44	1.49	0.79	0.00	NHCOMe	OMe	19r
10	1.62	1.74	0.12	2.88	0.79	0.00	NHCO-*i*-C_4H_9	OMe	19r
11	1.66	1.55	0.11	2.48	0.79	0.00	$NHCOCHCl_2$	OMe	19r
12	1.71	1.64	0.07	2.67	0.79	0.00	NHCO-2-furyl	OMe	19r
13	1.75	1.89	0.14	3.18	0.79	0.00	NHCO-4-pyridyl	OMe	19s
14	1.98	1.91	0.07	3.23	0.79	0.00	$NHCOCH_2SC_2H_5$	OMe	19r
15	2.00	1.89	0.11	3.18	0.79	0.00	NHCO-2-pyridyl	OMe	19s
16	2.08	2.29	0.21	4.01	0.79	0.00	NHCOPh-4-NH_2	OMe	19r
17	2.10	2.23	0.13	3.88	0.79	0.00	$NHCOCH_2Ph$	OMe	19r
18	2.12	2.02	0.10	3.46	0.79	0.00	NHCOPh	OMe	19r
19	2.62	2.92	0.30	3.46	2.17	−0.13	NHCOPh	O-*i*-C_4H_9	19t
20	2.64	2.30	0.34	3.46	1.25	−0.10	NHCOPh	OEt	19t
21	2.69	2.57	0.11	3.46	1.71	−0.19	NHCOPh	$OCH(CH_3)_2$	19t
22	2.70	2.88	0.18	3.46	2.17	−0.21	NHCOPh	R-O-*sec*-C_4H_9	19n
23	2.72	2.61	0.11	3.46	1.71	−0.12	NHCOPh	OC_3H_7	19t
24	2.86	2.54	0.32	4.52	0.79	0.00	NHCO-3-indolyl	OMe	19r
25	2.95	2.92	0.03	3.46	2.17	−0.13	NHCOPh	OC_4H_9	19t
26	2.95	2.88	0.07	3.46	2.17	−0.21	NHCOPh	S-O-*sec*-C_4H_9	19n
27	3.40	3.37	0.02	1.49	3.40	1.14	NHCOMe	OPh-4-NO_2	19t
28	3.66	3.76	0.10	3.46	3.23	0.11	NHCOMe	R-$OCH(CH_3)Ph$	19n
29	3.77	3.76	0.01	3.46	3.23	0.11	NHCOPh	S-$OCH(CH_3)Ph$	19n
30	3.77	3.77	0.00	3.46	3.21	0.16	NHCOPh	OPh-4-NH_2	19e
31	3.85	3.70	0.15	3.46	2.76	0.72	NHCOPh	OPh-4-F	19e
32	3.87	3.66	0.21	3.46	2.77	0.60	NHCOPh	OPh	19e
33	3.87	3.98	0.11	3.46	3.33	0.46	NHCOPh	OPh-4-Me	19e
34	3.87	3.82	0.05	2.41	3.40	1.14	$NHCOC_3H_7$	OPh-4-NO_2	19l
35	3.94	4.07	0.13	3.46	3.27	0.75	NHCOPh	OPh-4-Cl	19e
36	3.95	3.60	0.35	1.95	3.40	1.14	NHCOEt	OPh-4-NO_2	19l
37	3.99	4.02	0.03	3.46	3.45	0.36	NHCOPh	OPh-4-OMe	19e
38	4.04	4.53	0.49	3.43	3.33	1.73	NHCOPh	OCH_2-4-pyridyl	19t
39	4.06	4.33	0.27	3.46	3.40	1.14	NHCOPh	OPh-4-NO_2	19t
40	4.17	4.31	0.14	3.46	3.40	1.09	NHCOPh	OPh-3-NO_2	19e
41	4.51	4.49	0.02	3.46	3.78	0.90	NHCOPh	OPh-4-COMe	19e
42	5.31	4.68	0.62	4.19	3.40	1.14	$NHCOOCH_2Ph$	OPh-4-NO_2	19l
43	4.80[a]	3.15	1.65	1.03	3.40	1.14	NHCOH	OPh-4-NO_2	19l

[a] This point not used in deriving eq 5.

Table XI. Development of Eq 5

Intercept	MR-3	MR-1	σ^*	r	s	$F_{1,X}$[a]
0.94	0.92			0.928	0.435	248.4
−0.24	0.84	0.44		0.972	0.275	60.8
−0.20	0.69	0.48	0.44	0.981	0.235	15.6

[a] $F_{1,30;\alpha=0.001} = 13.3$.

Table XII. Squared Correlation Matrix of Variables Considered in the Derivation of Eq 5

	MR-1	MR-3	σ^*	π-1	π-3
MR-1	1.00	0.07	0.00	0.63	0.08
MR-3		1.00	0.55	0.13	0.89
σ^*			1.00	0.02	0.51
π-1				1.00	0.16
π-3					1.00

Substituting log P for MR in eq 6 gives a poor fit (r = 0.867, s = 0.403), again indicating interaction in nonhydrophobic space. π and MR are moderately orthogonal (r^2 = 0.68). We interpret the slope of eq 6 to mean that binding is occurring in ρ_2 space.

A result similar to that of eq 6 comes from the data of Table XIV on the inhibition of chymotrypsin by a set of miscellaneous aromatic compounds. Substitution of log P for MR in eq 7 gives a much poorer correlation (r =

Table XIII. Parameters Used in the Formulation of Eq 6

Log $1/K_i$ Obsd	Log $1/K_i$ Calcd	$\lvert\Delta \log 1/K_i\rvert$	MR	Inhibitor	Ref
1.99	1.92	0.07	3.23	Benzoic acid	19u
2.42	2.42	0.00	3.70	*m*-Toluic acid	19u
2.43	2.42	0.01	3.70	*p*-Toluic acid	19u
2.79	2.90	0.11	4.15	Hydrocinnamic acid	19u
3.22	3.39	0.17	4.62	4-Phenyl-*n*-butyric acid	19u
3.86	3.66	0.20	4.87	2-Naphthoic acid	19u
3.89	3.89	0.00	5.09	4-*tert*-Butylbenzoic acid	19u

$$\log 1/K_i = 1.02\ (\pm 0.25)\ \mathrm{MR} - 0.38\ (\pm 0.89) \qquad (7)$$

n = 11, r = 0.952, s = 0.325

0.844, s = 0.569). The slopes of eq 6 and 7 are identical. The more negative intercept of eq 6 brings out the fact that the acids are poorer inhibitors. The fit of eq 7 is slightly poorer than that found using the molecular area (r = 0.966, s = 0.273).

QSAR of k_2 Acylation. The acylation step in chymotrypsin hydrolysis has not been studied extensively in

Table XIV. Parameters Used in the Formulation of Eq 7

Log $1/K_i$ Obsd	Calcd	$\lvert\Delta \log 1/K_i\rvert$	MR	Inhibitor	Ref
2.07	2.31	0.24	2.64	Cyclohexene	19cc
2.08	2.29	0.21	2.62	Benzene	19bb
2.25	2.19	0.06	2.52	Pentane	19cc
2.55	2.77	0.22	3.09	Toluene	19bb
2.74	3.23	0.49	3.54	Ethylbenzene	19bb
3.02	2.94	0.08	3.25	Nitrobenzene	19bb
3.09	2.72	0.37	3.04	Chlorobenzene	19bb
3.67	3.16	0.51	3.47	Indene	19cc
3.93	3.88	0.05	4.17	Naphthalene	19bb
4.21	3.89	0.32	4.18	Azulene	19bb
5.26	5.47	0.21	5.73	Anthracene	19bb

a structure–activity sense. The best such study is from Berezin and his colleagues. We have formulated eq 8 from

$$\log k_2 = -0.52\ (\pm 0.22)\ \text{MR-1} + 1.10\ (\pm 0.25)\ \text{MR-2} - 1.56\ (\pm 0.50)\ I\text{-1} + 0.42\ (\pm 0.85) \quad (8)$$

$n = 18, r = 0.971, s = 0.399$

their data in Table XV. Equation 8 rationalizes 94% of the variance in log k_2. There are a few poorly fit points and this is to be expected. Berezin et al. note that when k_2 and k_3 differ by more than an order of magnitude, one cannot expect reliable values for k_2 because of lack of precision in experimental technique. Equation 8 correlates L esters of the type in structure **2a**. As in our archetypal eq 1, we find a slope of about 1 for MR-2 and a negative coefficient with I-1 for cases where $R_2 = CH(CH_3)_2$. The effect of the isopropyl group is almost ten times more deleterious in the acylation step (k_2) than in the binding process ($1/K_m$). No term appears for R_3; one would expect an electronic and possibly a steric effect with R_3. However, there is so little variation in R_3 that these effects cannot be assayed properly.

The coefficient with MR-1 is opposite in sign from that in eq 1. This could be interpreted to mean that R_1 is desorbed from ρ_1 space in the acylation step or that large groups hinder the movement of the substrate necessary in the acylation process of hydrolysis. The development of eq 8 is given in Table XVI. Although π and MR are rather collinear (see Table XVII), the substitution of π for MR in eq 8 yields a poorer correlation ($r = 0.916$, $s = 0.664$).

Three sets of data (Table XVIII) from the laboratory of Kirsch provide further perspective on k_2.

acylation with $X\text{-}C_6H_4COOC_6H_4\text{-}4\text{-}NO_2$, pH 7

$$\log k_2 = 0.45\ (\pm 0.43)\ \Sigma\sigma^+ + 3.54\ (\pm 0.19) \quad (9)$$

$n = 14, r = 0.553, s = 0.319$

$$\log k_2 = 0.62\ (\pm 0.24)\ \Sigma\sigma^+ + 0.59\ (\pm 0.23)\ \Sigma\text{MR} + 3.09\ (\pm 0.20) \quad (10)$$

$n = 14, r = 0.907, s = 0.168$

acylation with $X\text{-}C_6H_4COOC_6H_4\text{-}4\text{-}NO_2$, pH 6

$$\log k_2 = 0.67\ (\pm 0.43)\ \Sigma\text{MR} + 2.75\ (\pm 0.34) \quad (11)$$

$n = 13, r = 0.713, s = 0.268$

$$\log k_2 = 0.83\ (\pm 0.25)\ \Sigma\text{MR} + 0.49\ (\pm 0.21)\ \Sigma\sigma^+ + 2.57\ (\pm 0.20) \quad (12)$$

$n = 13, r = 0.933, s = 0.144$

One data point [4-$CH(Me)_2$] was omitted in the formulation of eq 8 and 9 because it was poorly fit. Although the isopropyl group in valine is usually poorly fit in ρ_2 space, the isopropyl group is well fit by eq 12 and 14. At pH 7, the best single-variable equation was that in σ^+ although ΣMR produced almost as good an equation. The electronic parameter σ^+ gave slightly better correlations than σ. We assume that the $X\text{-}C_6H_4$ moiety is binding in ρ_2 space; however, the coefficient with this term is rather low in eq 10 but closer to normal in eq 12. It appears that a conformational change at the lower pH leads to better binding. In the study at pH 6, the best single-variable equation is that in ΣMR.

acylation with $X\text{-}C_6H_4COOC_6H_3\text{-}2,4\text{-}(NO_2)_2$, pH 7

$$\log k_2 = 0.67\ (\pm 0.59)\ \sigma^+ + 5.38\ (\pm 0.27) \quad (13)$$

$n = 14, r = 0.582, s = 0.441$

$$\log k_2 = 0.97\ (\pm 0.25)\ \sigma^+ + 0.96\ (\pm 0.27)\ \text{MR-}p + 4.87\ (\pm 0.18) \quad (14)$$

$n = 14, r = 0.949, s = 0.178$

Table XV. Parameters Used in the Formulation of Eq 8

No.	Log k_2 Obsd	Calcd	$\lvert\Delta \log k_2\rvert$	MR-1	MR-2	I-1	R_1	R_2	R_3	Ref
1	−1.16	−0.75	0.40	3.46	0.56	0	L-NHCOPh	Me	OEt	19j
2	−1.04	−1.28	0.24	3.46	1.50	1	L-NHCOPh	$i\text{-}C_3H_7$	OMe	19j
3	−0.75	−0.25	0.50	1.49	1.50	1	L-NHCOMe	$i\text{-}C_3H_7$	$i\text{-}OC_3H_7$	19j
4	−0.49	−0.51	0.02	1.98	1.50	1	L-$NHCOCH_2Cl$	$i\text{-}C_3H_7$	OMe	19j
5	−0.39	−0.24	0.15	3.46	1.03	0	L-NHCOPh	Et	OMe	19j
6	−0.26	−0.25	0.01	1.49	1.50	1	L-NHCOMe	$i\text{-}C_3H_7$	OEt	19j
7	−0.01	−0.25	0.24	1.49	1.50	1	L-NHCOMe	$i\text{-}C_3H_7$	OMe	19j
8	0.62	0.28	0.34	3.46	1.50	0	L-NHCOPh	C_3H_7	OMe	19j
9	0.94	0.79	0.15	1.49	1.03	0	L-NHCOMe	Et	OMe	19j
10	1.55	1.31	0.24	1.49	1.50	0	L-NHCOMe	C_3H_7	OMe	19j
11	1.66	1.94	0.28	3.46	3.00	0	L-NHCOPh	CH_2Ph	OMe	19j
12	1.82	2.55	0.73	2.67	3.18	0	L-NHCO-furyl	CH_2Ph-4-OH	OMe	19j
13	2.01	1.81	0.20	1.49	1.96	0	L-NHCOMe	C_4H_9	OMe	19j
14	2.40	2.14	0.26	3.46	3.18	0	L-NHCOPh	CH_2Ph-4-OH	OEt	19j
15	2.42	2.96	0.54	1.49	3.00	0	L-NHCOMe	CH_2Ph	OEt	19j
16	2.56	2.14	0.42	3.46	3.18	0	L-NHCOPh	CH_2Ph-4-OH	OMe	19j
17	2.90	2.96	0.06	1.49	3.00	0	L-NHCOMe	CH_2Ph	OMe	19j
18	3.70	3.16	0.54	1.49	3.18	0	L-NHCOMe	CH_2Ph-4-OH	OEt	19j
19	−1.03[a]	−0.24	0.79	1.49	0.10	0	NHCOMe	H	OEt	19j
20	−0.38[a]	−1.26	0.88	3.46	0.10	0	NHCOPh	H	OMe	19j
21	−0.31[a]	−0.24	0.07	1.49	0.10	0	NHCOMe	H	OMe	19j

[a] These points not used in deriving eq 8.

Table XVI. Development of Eq 8

Intercept	MR-2	I-1	MR-1	r	s	$F_{1,X}$[a]
−1.81	1.39			0.844	0.832	39.6
−0.97	1.15	−1.26		0.917	0.640	12.0
0.42	1.10	−1.56	−0.52	0.971	0.399	24.6

[a] $F_{1,14;\alpha=0.001} = 17.1$; $F_{1,14;\alpha=0.005} = 11.1$.

Table XVII. Squared Correlation Matrix of Variables Considered in the Derivation of Eq 8

	MR-1	MR-2	π-1	π-2	I-1
MR-1	1.00	0.00	1.00	0.01	0.06
MR-2		1.00	0.00	0.80	0.14
π-1			1.00	0.01	0.05
π-2				1.00	0.14
I-1					1.00

In the case of eq 9–12, MR gives a slightly better correlation than MR-p (i.e., setting MR-m = 0). However, in the case of eq 13 and 14, a definite improvement is seen by setting MR-m = 0. It appears that meta substituents make rather poor contact with the enzyme. A more critical look at meta substituents is needed since there are mostly only para derivatives in the above data sets. The coefficients with the MR term in eq 12 and 14 are what one would expect for binding in ρ_2 space. Not only is the slope of eq 10 out of line, the quality of fit is bad. It may be that the experimental results supporting eq 10 are not as good as the other results. In these equations, σ^+ and MR are quite orthogonal ($r^2 \sim 0.1$).

QSAR of k_3 Deacylation. In a previous study[2b] of the deacylation of chymotrypsin, eq 15 was formulated for

$$\log k_3 = 2.20 (\pm 0.60) \sigma^* + 1.01 (\pm 0.40) E_s + 0.37 (\pm 0.22) \pi - 2.08 (\pm 0.48) \quad (15)$$

$n = 13$, $r = 0.969$, $s = 0.327$

substrates of the type $RCOOC_6H_4$-4-NO_2. Since this analysis, additional data have appeared on similar molecules. Equation 16 has been derived from the parameters

$$\log k_3 = 2.09 (\pm 0.34) \sigma^* + 1.21 (\pm 0.27) E_s + 0.34 (\pm 0.10) \mathrm{MR} - 0.95 (\pm 0.71) I\text{-}1 - 1.91 (\pm 0.30) \quad (16)$$

$n = 36$, $r = 0.975$, $s = 0.320$

Table XVIII. Parameters Used in the Formulation of Eq 9–12

No.	Log k_2 Obsd	Log k_2 Calcd	$\lvert\Delta \log k_2\rvert$	ΣMR	σ^+	X	Ref
			A. X-C_6H_4COO-C_6H_4-4-NO_2, pH 7				
1	2.97	3.13	0.16	0.89	−0.78	4-OMe	19v
2	3.11	3.16	0.05	0.19	−0.07	4-F	19v
3	3.16	3.21	0.05	0.20	0.00	H	19v
4	3.29	3.43	0.14	0.66	−0.07	3-Me	19v
5	3.31	3.28	0.02	0.66	−0.31	4-Me	19v
6	3.56	3.57	0.01	0.70	0.11	4-Cl	19v
7	3.59	3.77	0.18	0.60	0.52	3-CF_3	19v
8	3.60	3.42	0.18	0.19	0.35	3-F	19v
9	3.77	3.75	0.02	0.70	0.40	3-Cl	19v
10	3.94	3.93	0.01	0.73	0.66	4-CN	19v
11	3.95	3.82	0.13	0.60	0.61	4-CF_3	19v
12	3.96	4.07	0.11	0.84	0.79	4-NO_2	19v
13	3.97	3.57	0.40	1.13	−0.30	4-C_2H_5	19v
14	4.07	4.14	0.07	2.06	−0.26	4-t-C_4H_9	19v
15	4.44[a]	3.86	0.58	1.60	−0.28	4-i-C_3H_7	19v
			B. X-C_6H_4COO-C_6H_4-4-NO_2, pH 6				
16	2.64	2.74	0.10	0.20	0.00	H	19v
17	2.74	2.69	0.04	0.19	−0.07	4-F	19v
18	2.78	2.93	0.15	0.89	−0.78	4-OMe	19v
19	2.85	2.97	0.12	0.66	−0.31	4-Me	19v
20	3.17	2.90	0.27	0.19	0.35	3-F	19v
21	3.22	3.32	0.10	0.60	0.52	3-CF_3	19v
22	3.27	3.21	0.06	0.70	0.11	4-Cl	19v
23	3.35	3.50	0.15	0.73	0.66	4-CN	19v
24	3.41	3.37	0.04	0.60	0.61	4-CF_3	19v
25	3.43	3.35	0.08	0.70	0.40	3-Cl	19v
26	3.50	3.37	0.13	1.13	−0.30	4-C_2H_5	19v
27	3.55	3.66	0.11	0.84	0.79	4-NO_2	19v
28	3.87	3.77	0.10	1.60	−0.28	4-i-C_3H_7	19v
			C. X-C_6H_4COO-C_6H_3-2,4-$(NO_2)_2$, pH 7				
29	4.75	4.90	0.15	0.10	−0.07	3-Me	19v
30	4.83	4.89	0.06	0.09	−0.07	4-F	19v
31	4.86	4.96	0.10	0.10	0.00	H	19v
32	4.96	4.87	0.09	0.79	−0.78	4-OMe	19v
33	5.22	5.11	0.11	0.56	−0.31	4-Me	19v
34	5.25	5.30	0.05	0.10	0.35	3-F	19v
35	5.47	5.35	0.12	0.10	0.40	3-Cl	19v
36	5.47	5.47	0.00	0.10	0.52	3-CF_3	19v
37	5.53	5.57	0.04	1.03	−0.30	4-Et	19v
38	5.55	5.55	0.00	0.60	0.11	4-Cl	19v
39	5.94	6.11	0.17	0.63	0.66	4-CN	19v
40	5.99	6.04	0.05	1.50	−0.28	4-i-C_3H_7	19v
41	6.17	6.34	0.17	0.74	0.79	4-NO_2	19v
42	6.40	5.94	0.46	0.50	0.61	4-CF_3	19v

[a] This point not used in deriving eq 9 and 10.

Table XIX. Parameters Used in the Formulation of Eq 16. Hydrolysis of $RCOO\text{-}C_6H_4\text{-}4\text{-}NO_2$

	Log k_3								
No.	Obsd	Calcd	$\lvert\Delta \log k_3\rvert$	MR	σ^*	E_s	I-1	R	Ref
1	0.47	0.35	0.12	1.05	1.05	-0.24	0	$ClCH_2$	19w
2	0.39	0.65	0.26	0.10	0.49	1.24	0	H	19w
3	-0.06	-0.12	0.06	2.88	0.56	-0.30	0	$CH_3(CH_2)_2CONHCH_2$	19w
4	-0.10	0.05	0.15	1.86	0.85	-0.37	0	ICH_2	19w
5	-0.12	-0.55	0.43	1.96	0.56	-0.39	0	$CH_3CONHCH_2$	19w
6	-0.25	-0.39	0.14	2.44	0.56	-0.40	0	$CH_3CH_2CONHCH_2$	19w
7	-0.39	0.21	0.60	4.56	0.57	-0.51	0	$C_6H_5CH_2OCONHCH_2$	19w
8	-0.46	-0.64	0.18	1.21	0.52	-0.19	0	CH_3OCH_2	19w
9	-0.47	-0.90	0.43	4.57	-0.04	-0.38	0	β-$(C_8H_6N)(CH_2)_2$	19w
10	-0.61	-0.63	0.02	1.49	0.58	-0.36	0	$HCONHCH_2$	19w
11	-0.73	-1.03	0.30	3.46	0.08	-0.38	0	$C_6H_5(CH_2)_2$	19w
12	-0.83	-0.92	0.09	5.03	-0.05	-0.51	0	β-$(C_8H_6N)(CH_2)_3$	19w
13	-0.87	-1.15	0.28	3.93	0.02	-0.51	0	$C_6H_5(CH_2)_3$	19w
14	-1.21	-1.61	0.40	2.89	-0.15	-0.30	0	$CH_3(CH_2)_5$	19z
15	-1.27	-1.53	0.26	1.98	0.14	-0.48	0	$Cl(CH_2)_3$	19w
16	-1.37	-1.46	0.09	2.44	0.05	-0.40	0	$Cl(CH_2)_4$	19w
17	-1.44	-1.91	0.47	2.42	-0.16	-0.40	0	$CH_3(CH_2)_4$	19x
18	-1.47	-1.49	0.02	3.36	-0.17	-0.30	0	$CH_3(CH_2)_6$	19z
19	-1.65	-0.89	0.76	3.00	0.22	-0.38	0	$C_6H_5CH_2$	19w
20	-1.67	-1.67	0.00	1.52	0.39	-0.90	0	$Cl(CH_2)_2$	19w
21	-1.68	-1.14	0.54	4.39	-0.05	-0.51	0	$C_6H_5(CH_2)_4$	19w
22	-1.77	-1.99	0.22	1.96	-0.13	-0.39	0	$CH_3(CH_2)_3$	19z
23	-2.00	-1.72	0.28	0.56	0.00	0.00	0	Me	19w
24	-2.10	-1.86	0.24	1.03	-0.10	-0.07	0	Et	19x
25	-2.21[a]	-1.02	1.19	2.88	0.13	-0.30	0	$CH_3CONH(CH_2)_3$	19w
26	-2.22	-2.09	0.13	1.50	-0.12	-0.36	0	$CH_3(CH_2)_2$	19x
27	-2.45	-2.37	0.08	1.50	-0.19	-0.47	0	$(CH_3)_2CH$	19w
28	-2.52[a]	-1.02	1.50	4.11	-0.02	-0.38	0	β-$(C_8H_6N)CH_2$	19w
29	-2.69	-2.98	0.29	2.93	0.96	-2.58	1	$4\text{-}CF_3\text{-}C_6H_4$	19y
30	-2.85	-2.63	0.22	3.17	1.09	-2.58	1	$3\text{-}NO_2\text{-}C_6H_4$	19y
31	-2.90	-2.64	0.25	1.96	-0.13	-0.93	0	$(CH_3)_2CHCH_2$	19x
32	-3.05	-3.41	0.36	2.53	0.82	-2.58	1	$3\text{-}F\text{-}C_6H_4$	19y
33	-3.44[a]	-2.52	0.92	3.17	1.14	-2.58	1	$4\text{-}NO_2\text{-}C_6H_4$	19y
34	-3.48	-3.38	0.10	3.04	0.75	-2.58	1	$4\text{-}Cl\text{-}C_6H_4$	19y
35	-3.69	-3.74	0.05	1.96	-0.30	-1.54	0	$(CH_3)_3C$	19w
36	-3.69	-3.87	0.18	2.54	0.60	-2.58	1	C_6H_5	19y
37	-3.82	-3.55	0.27	2.42	-0.17	-1.74	0	$(CH_3)_3CCH_2$	19x
38	-4.27	-4.00	0.27	3.00	0.46	-2.58	1	$4\text{-}CH_3\text{-}C_6H_4$	19y
39	-4.39	-4.15	0.24	3.17	0.36	-2.58	1	$4\text{-}CH_3O\text{-}C_6H_4$	19y

[a] These points not used in deriving eq 16.

in Table XIX. In eq 16, MR has been used instead of π as in eq 15 in line with our new findings in correlating K_m and k_2. Using π in eq 16 results in a poorer correlation ($r = 0.956$, $s = 0.424$). The indicator variable I-1 is given a value of 1 for those cases where benzoates are involved. No aromatic esters were studied in the formulation of eq 15. The negative coefficient with I-1 suggests possible steric hindrance in deacylation not well accounted for by E_s of the phenyl ring. An indication of the E_s role of I-1 can be seen in Table XXI where the high collinearity between these vectors is apparent.

It is satisfying to see that the coefficients with σ^* and E_s as well as the intercepts of eq 15 and 16 are in close agreement. The development of eq 16 is given in Table XX and the relative orthogonality of the vectors in Table XXI. Three data points (**25**, **28**, and **33**) have not been used in the derivation of eq 16; for some reason that is not obvious, they are poorly fit.

Data on deacylation of compounds of the type in structure **2** are assembled in Table XXII. Their QSAR is

$$\log k_3 = 0.75\ (\pm 0.14)\ \text{MR-2} - 1.79\ (\pm 0.28)\ I\text{-1} - 1.48\ (\pm 0.26)\ I\text{-2} - 0.31\ (\pm 0.30) \qquad (17)$$

$$n = 33,\ r = 0.977,\ s = 0.289$$

In this expression, I-1 takes the value of 1 for D isomers and 0 for L isomers. The negative coefficient with this term shows that D isomers are about 65 times more difficult to deacylate than L isomers. I-2 is assigned a value of 1 when R_2 = isopropyl. As usual, this group yields less active congeners.

R_3 is not present in the deacylation step so no term is needed for this group. Note, however, that no term occurs for R_1 in eq 17. This is not entirely unexpected since the 0.80 MR-1 term in eq 1 becomes -0.52 MR-1 in the acylation step (eq 8). The obvious conclusion from eq 8 and 17 is that R_1 has been removed from ρ_1 space in the acylation step and in the deacylation step is, for practical

Table XX. Development of Eq 16

Intercept	E_s	σ^*	MR	I-1	r	s	$F_{1,X}$[a]
-0.75	1.12				0.805	0.814	62.6
-0.99	1.38	1.65			0.928	0.519	50.7
-1.79	1.53	1.86	0.35		0.969	0.350	40.4
-1.91	1.21	2.09	0.34	-0.95	0.975	0.320	7.4

[a] $F_{1,30;\alpha=0.001} = 13.2$; $F_{1,30;\alpha=0.01} = 7.6$.

Table XXI. Squared Correlation Matrix of Variables Considered in the Derivation of Eq 16

	E_s	σ^*	MR	π	I
E_s	1.00	0.13	0.08	0.10	0.79
σ^*		1.00	0.01	0.11	0.30
MR			1.00	0.50	0.03
π				1.00	0.05
I-1					1.00

Table XXII. Parameters Used in the Formulation of Eq 17

No.	Log k_3 Obsd	Log k_3 Calcd	$\|\Delta \log k_3\|$	MR-2	*I*-1	*I*-2	R_1	R_2	R_3	Ref
1	−2.10	−2.45	0.35	1.50	1	1	D-$NHCOOCH_2Ph$	*i*-C_3H_7	OPh-4-NO_2	19l
2	−1.65	−0.98	0.67	1.49	1	0	D-$NHCOOCH_2Ph$	CH_2CONH_2	OPh-4-NO_2	19l
3	−1.45	−1.32	0.12	1.03	1	0	D-$NHCOOCH_2Ph$	Et	OPh-4-NO_2	19l
4	−1.05	−1.68	0.63	0.56	1	0	D-$NHCOOCH_2Ph$	Me	OPh-4-NO_2	19l
5	−0.96	−0.63	0.33	1.96	1	0	D-$NHCOOCH_2Ph$	*i*-C_4H_9	OPh-4-NO_2	19l
6	−0.93	−0.66	0.27	1.50	0	1	L-$NHCOOCH_2Ph$	*i*-C_3H_7	OPh-4-NO_2	19l
7	−0.82	−0.97	0.15	1.50	1	0	D-$NHCOOCH_2Ph$	C_3H_7	OPh-4-NO_2	19l
8	−0.74[a]	0.15	0.89	3.00	1	0	D-$NHCOOCH_2Ph$	CH_2Ph	OPh-4-NO_2	19l
9	−0.68	−0.66	0.02	1.50	0	1	L-$NHCOCH_2Cl$	*i*-C_3H_7	OMe	19j
10	−0.68	−0.66	0.02	1.50	0	1	L-NHCOMe	*i*-C_3H_7	*i*-C_3H_7	19j
11	−0.68	−0.66	0.02	1.50	0	1	L-NHCOMe	*i*-C_3H_7	OEt	19j
12	−0.68	−0.66	0.02	1.50	0	1	L-NHCOMe	*i*-C_3H_7	OMe	19j
13	−0.66	−0.66	0.00	1.50	0	1	L-NHCOPh	*i*-C_3H_7	OMe	19j
14	−0.22	0.11	0.33	0.56	0	0	L-NHCOPh	Me	OEt	19j
15	0.18	0.46	0.28	1.03	0	0	L-NHCOPh	Et	OMe	19j
16	0.22	0.46	0.24	1.03	0	0	L-NHCOMe	Et	OMe	19j
17	0.43	0.46	0.03	1.03	0	0	L-$NHCOOCH_2Ph$	Et	OPh-4-NO_2	19l
18	0.48	0.11	0.37	0.56	0	0	L-$NHCOOCH_2Ph$	Me	OPh-4-NO_2	19l
19	0.77	0.82	0.05	1.50	0	0	L-NHCOPh	C_3H_7	OMe	19j
20	0.77	0.82	0.05	1.50	0	0	L-NHCOMe	C_3H_7	OMe	19j
21	0.91	0.82	0.09	1.50	0	0	L-$NHCOOCH_2Ph$	C_3H_7	OPh-4-NO_2	19l
22	0.92	1.16	0.24	1.96	0	0	L-$NHCOOCH_2Ph$	*i*-C_4H_9	OPh-4-NO_2	19l
23	1.10	1.16	0.06	1.96	0	0	L-$NHCOOCH_2Ph$	C_4H_9	OPh-4-NO_2	19l
24	1.20	1.16	0.04	1.96	0	0	L-NHCOMe	*i*-C_4H_9	OMe	19u
25	1.25[a]	2.86	1.61	4.23	0	0	L-$NHCOOCH_2Ph$	CH_2-indolyl	OC_6H_4-4-NO_2	19l
26	1.28	1.16	0.12	1.96	0	0	L-NHCOMe	C_4H_9	OMe	19j
27	1.53	0.81	0.72	1.49	0	0	L-$NHCOOCH_2Ph$	CH_2CONH_2	OPh-4-NO_2	19l
28	1.59	1.94	0.35	3.00	0	0	L-$NHCOOCH_2Ph$	CH_2Ph	OPh-4-NO_2	19l
29	1.96	1.94	0.02	3.00	0	0	L-NHCOPh	CH_2Ph	OMe	19j
30	1.97	1.94	0.03	3.00	0	0	L-NHCOMe	CH_2Ph	OEt	19j
31	2.04	1.94	0.10	3.00	0	0	L-NHCOMe	CH_2Ph	OMe	19j
32	2.08	2.08	0.00	3.18	0	0	L-NHCOPh	CH_2Ph-4-OH	OMe	19j
33	2.12	2.08	0.04	3.18	0	0	L-NHCOPh	CH_2Ph-4-OH	OEt	19j
34	2.30	2.08	0.22	3.18	0	0	L-NHCOMe	CH_2Ph-4-OH	OEt	19j
35	2.30	2.08	0.22	3.18	0	0	L-NHCO-furyl	CH_2Ph-4-OH	OMe	19j
36	−1.14[a]	−1.68	0.54	0.56	1	0	D-NHCOH	Me	OPh-4-NO_2	19l
37	−0.61[a]	0.11	0.72	0.56	0	0	L-NHCOH	Me	OPh-4-NO_2	19l

[a] These points not used in deriving eq 17.

Table XXIII. Development of Eq 17

Intercept	MR-2	*I*-1	*I*-2	*r*	*s*	$F_{1,X}$[a]
−1.64	1.13			0.722	0.907	33.8
−0.97	0.92	−1.61		0.861	0.678	25.4
−0.31	0.75	−1.79	−1.48	0.977	0.289	136

[a] $F_{1,29;\alpha=0.001} = 13.4$.

Table XXIV. Squared Correlation Matrix of Variables Considered in the Derivation of Eq 17

	MR-2	*I*-1	*I*-2	π-2
MR-2	1.00	0.08	0.04	0.42
I-1		1.00	0.00	0.09
I-2			1.00	0.00
π-2				1.00

purposes, out of contact with the enzyme.

Two congeners (**8** and **25** in Table XXII) have not been used in the formulation of eq 17. It is not clear why these points are so poorly fit. As usual, formyl derivatives (**36** and **37**) are poorly fit.

The substitution of π-2 for MR-2 in eq 17 yields a poorer correlation as usual ($r = 0.910$, $s = 0.562$). In this case, π and MR are moderately orthogonal (Table XXIV). The development of eq 17 is given in Table XXIII.

Overall Hydrolysis ($k_{cat.}$). Most studies of chymotrypsin hydrolysis have not been concerned with the measurement of k_2 and k_3 but report K_m and $k_{cat.}$ values. Although $k_{cat.}$ is a complex parameter involving at least k_2 and k_3, we felt an analysis would be worthwhile of its relationship to chemical structure. Table XXV contains $k_{cat.}$ values from many different laboratories obtained under varying experimental conditions. Despite the inhomogenicity of the data and the complexity of $k_{cat.}$, eq 18 does a reasonable job of bringing order to 57 analogues

$$\log k_{cat.} = 1.79\ (\pm 0.33)\ \text{MR-2} - 0.24\ (\pm 0.08)\ (\text{MR-2})^2 - 1.45\ (\pm 0.26)\ I\text{-1} - 0.01\ (\pm 0.009)\ \text{MR-1}\cdot\text{MR-2}\cdot\text{MR-3} - 1.51\ (\pm 0.31) \quad (18)$$

$$n = 57,\ r = 0.959,\ s = 0.313$$

of esters of the type in structure **2**. In eq 18, *I*-1 = 1 for R = isopropyl. It takes the same value when C_6H_5 is attached as

or carboxylate is on the α-carbon. The steric similarity is obvious as Cohen et al. have noted.[19m] MR-1 occurs only in the negative cross-product term, indicating a constant negative effect for R_1. The dominant role is played by MR-2 (Table XXVI) as we have seen from the correlation equations for the simpler parameters k_2 and k_3. The similarity between cross-product terms in eq 1 and 18 suggests that $k_{cat.}$ may not be entirely independent of K_m. It is quite apparent from these results that when R_2 reaches a certain size [−0.23 $(\text{MR-2})^2$ becomes significant], catalysis rate drops off. This may also occur with R_2

Table XXV. Parameters Used in the Formulation of Eq 18

No.	Log $k_{cat.}$ Obsd	Log $k_{cat.}$ Calcd	$\|\Delta \log k_{cat.}\|$	MR-1	MR-2	I-1	R_1	R_2	R_3	Ref
1	−2.40[a]	−0.86	1.54	5.01	0.56	0	L-NHCO-2-quinolyl	Me	OMe	19c
2	−1.19	−1.04	0.15	3.46	1.50	1	L-NHCOPh	i-C_3H_7	OMe	19j
3	−1.16	−0.66	0.50	3.46	0.56	0	L-NHCOPh	Me	OEt	19j
4	−1.15	−0.63	0.52	3.18	0.56	0	L-NHCO-2-pyridyl	Me	OMe	19c
5	−1.08	−0.80	0.28	1.49	1.50	1	L-NHCOMe	i-C_3H_7	O-i-C_3H_7	19g
6	−0.90	−0.86	0.04	1.98	1.50	1	L-$NHCOCH_2Cl$	i-C_3H_7	OMe	19j
7	−0.88	−0.80	0.08	1.49	1.50	1	L-NHCOMe	i-C_3H_7	OEt	19g
8	−0.82	−0.80	0.02	1.49	1.50	1	L-NHCOMe	i-C_3H_7	OMe	19a
9	−0.64	−0.66	0.02	3.46	0.56	0	L-NHCOPh	Me	OMe	19c
10	−0.64	−0.80	0.16	1.49	1.50	1	L-NHCOMe	i-C_3H_7	OCH_2CH_2Cl	19g
11	−0.62	−0.72	0.10	3.90	0.56	0	L-NHCOPh-2-NH_2	Me	OMe	19c
12	−0.49	−0.05	0.43	3.46	1.03	0	L-NHCOPh	Et	OMe	19j
13	−0.40	−0.58	0.18	2.80	0.56	0	L-NHCO-furyl-H_4	Me	OMe	19c
14	−0.38	−0.63	0.25	3.18	0.56	0	L-NHCO-4-pyridyl	Me	OMe	19c
15	−0.34	−0.64	0.30	3.28	0.56	0	L-NHCO-2-thienyl	Me	OMe	19c
16	−0.29	−0.56	0.27	2.67	0.56	0	L-NHCO-furyl	Me	OMe	19c
17	−0.24	−0.63	0.39	3.18	0.56	0	L-NHCO-3-pyridyl	Me	OMe	19c
18	−0.15	−0.57	0.42	1.49	1.75	1	NHCOMe[b]	COOEt	OEt	19m
19	−0.05	−0.04	0.00	1.49	2.54	1	L-NHCOMe	Ph	OEt	19m
20	0.02	0.19	0.17	1.49	1.03	0	L-NHCOMe	Et	OMe	19a
21	0.11	−0.42	0.53	1.49	0.56	0	L-NHCOMe	Me	OMe	19d
22	0.39	0.45	0.06	3.46	1.50	0	L-NHCOPh	C_3H_7	OMe	19j
23	0.43	0.70	0.27	1.49	1.50	0	L-NHCOMe	C_3H_7	OMe	19a
24	0.46	0.70	0.24	1.49	1.50	0	L-NHCOMe	C_3H_7	O-i-C_3H_7	19g
25	0.57	0.70	0.13	1.49	1.50	0	L-NHCOMe	C_3H_7	OCH_2CH_2Cl	19g
26	0.70	1.09	0.39	1.49	1.96	0	L-NHCOMe	i-C_4H_9	OMe	19a
27	0.79	1.59	0.80	1.49	2.89	0	L-NHCOMe	C_6H_{13}	OMe	19a
28	0.92	1.09	0.17	1.49	1.96	0	L-NHCOMe	C_4H_9	OMe	19a
29	1.03	1.26	0.23	4.19	4.23	0	L-$NHCOOCH_2Ph$	CH_2-indolyl	OPh-4-NO_2	19o
30	1.05	1.26	0.21	4.19	4.23	0	L-$NHCOOCH_2Ph$	CH_2-indolyl	OPh-4-COMe	19o
31	1.06	1.26	0.20	4.19	4.23	0	L-$NHCOOCH_2Ph$	CH_2-indolyl	OPh-4-Cl	19o
32	1.13	1.39	0.26	1.49	2.42	0	L-NHCOMe	C_5H_{11}	OMe	19a
33	1.18	1.65	0.47	1.49	3.13	0	L-NHCOMe	CH_2-c-C_6H_{11}	OMe	19a
34	1.29	1.62	0.33	1.49	3.00	0	L-NHCOMe	CH_2Ph	R-CH(Me)-c-C_6H_{11}	19n
35	1.37	1.62	0.25	1.49	3.00	0	L-NHCOMe	CH_2Ph	S-CH(Me)-c-C_6H_{11}	19n
36	1.44	1.60	0.16	1.49	4.23	0	L-NHCOMe	CH_2-indolyl	OMe	19k
37	1.48	1.60	0.12	1.49	4.23	0	L-NHCOMe	CH_2-indolyl	OPh-4-NO_2	19k
38	1.49	1.37	0.11	3.46	3.00	0	L-NHCOPh	CH_2Ph	OMe	19j
39	1.49	1.58	0.09	1.82	3.00	0	L-$NHSO_2Me$	CH_2Ph	OPh-3-NO_2	19e
40	1.51	1.58	0.07	1.82	3.00	0	L-$NHSO_2Me$	CH_2Ph	OPh	19e
41	1.54	1.58	0.04	1.82	3.00	0	L-$NHSO_2Me$	CH_2Ph	OPh-4-OMe	19e
42	1.54	1.58	0.04	1.82	3.00	0	L-$NHSO_2Me$	CH_2Ph	OPh-4-COMe	19e
43	1.55	1.62	0.07	1.49	3.00	0	L-NHCOMe	CH_2Ph	S-sec-C_4H_9	19n
44	1.56	1.58	0.02	1.82	3.00	0	L-$NHSO_2Me$	CH_2Ph	OPh-4-Me	19e
45	1.56	1.58	0.02	1.82	3.00	0	L-$NHSO_2Me$	CH_2Ph	OPh-4-Cl	19e
46	1.56	1.58	0.02	1.82	3.00	0	L-$NHSO_2Me$	CH_2Ph	OPh-4-NO_2	19e
47	1.70	1.60	0.10	1.49	4.23	0	L-NHCOMe	CH_2-indolyl	OEt	19f
48	1.70	1.51	0.18	2.67	3.18	0	L-NHCO-furyl	CH_2Ph-4-OH	OMe	19j
49	1.72	1.62	0.10	1.49	3.00	0	L-NHCOMe	CH_2Ph	OMe	19a
50	1.84	1.62	0.22	1.49	3.00	0	L-NHCOMe	CH_2Ph	OEt	19j
51	1.89	1.62	0.27	1.49	3.00	0	L-NHCOMe	CH_2Ph	OPh-4-NO_2	19k
52	1.93	1.42	0.51	3.46	3.18	0	L-NHCOPh	CH_2Ph-4-OH	OEt	19j
53	1.96	1.42	0.54	3.46	3.18	0	L-NHCOPh	CH_2Ph-4-OH	OMe	19j
54	1.98	1.62	0.36	1.49	3.00	0	L-NHCOMe	CH_2Ph	R-sec-C_4H_9	19n
55	2.02	1.62	0.40	1.49	3.00	0	L-NHCOMe	CH_2Ph	R-CH(Me)Ph	19n
56	2.04	1.62	0.42	1.49	3.00	0	L-NHCOMe	CH_2Ph	S-CH(Me)Ph	19n
57	2.07	1.66	0.41	1.49	3.18	0	L-NHCOMe	CH_2Ph-4-OH	OMe	19i
58	2.28	1.66	0.62	1.49	3.18	0	L-NHCOMe	CH_2Ph-4-OH	OEt	19f

[a] This point not used in deriving eq 18. [b] Symmetric compound; see ref 19m.

Table XXVI. Development of Eq 18

Intercept	MR-2	I-1	$(MR\text{-}2)^2$	MR-1·2·3	r	s	$F_{1,X}$[a]
−0.99	0.74				0.816	0.617	110
−0.65	0.66	−1.17			0.899	0.473	39.8
−1.58	1.90	−1.47	−0.28		0.954	0.326	60.2
−1.51	1.79	−1.45	−0.24	−0.10	0.959	0.313	5.5

[a] $F_{1,40;\alpha=0.001} = 12.6$; $F_{1,40;\alpha=0.05} = 4.2$.

relatively small but R_1 and R_3 quite large, as is nicely brought out by the negative cross-product term. Undoubtedly, large R groups produce distortions in the active site and slow catalysis.

One data point (1) is very poorly fit and has been omitted in the derivation of eq 18. The substitution of π for MR in eq 18 yields a poorer correlation (r = 0.894, s = 0.469). MR and I-1 are highly orthogonal.

Table XXVII. Parameters Used in the Formulation of Eq 19

No.	Log $k_{cat.}$ Obsd	Log $k_{cat.}$ Calcd	$\lvert\Delta \log k_{cat.}\rvert$	MR-1	MR-3	*I*-1	R_1	R_3	Ref
1	0.39[a]	-1.19	1.58	3.18	0.79	1	NHCO-2-pyridyl	OMe	19s
2	0.27	-0.28	0.55	1.49	3.40	0	NHCOMe	OPh-4-NO_2	19t
3	-0.01[a]	-1.05	1.04	2.67	0.79	1	NHCO-2-furyl	OMe	19r
4	-0.13	-0.25	0.12	3.46	3.23	1	NHCOPh	S-OCH(Me)Ph	19n
5	-0.21	-0.21	0.00	3.46	3.33	1	NHCOPh	OPh-4-Me	19e
6	-0.25	-0.16	0.09	3.46	3.45	1	NHCOPh	OPh-4-OMe	19e
7	-0.25	-0.20	0.05	3.43	3.33	1	NHCOPh	OCH_2-4-pyridyl	19t
8	-0.28	-0.18	0.10	3.46	3.40	1	NHCOPh	OPh-3-NO_2	19e
9	-0.28	-0.02	0.25	3.46	3.78	1	NHCOPh	OPh-4-COMe	19e
10	-0.28	-0.45	0.17	3.46	2.76	1	NHCOPh	OPh-4-F	19e
11	-0.30	-0.26	0.04	3.46	3.21	1	NHCOPh	OPh-4-NH_2	19e
12	-0.31	-0.18	0.13	3.46	3.40	1	NHCOPh	OPh-4-NO_2	19e
13	-0.31	-0.24	0.07	3.46	3.27	1	NHCOPh	OPh-4-Cl	19e
14	-0.31	-0.44	0.13	3.46	2.77	1	NHCOPh	OPh	19e
15	-0.34[a]	-1.19	0.85	3.18	0.79	1	NHCO-3-pyridyl	OMe	19r
16	-0.60	-0.89	0.29	3.46	3.23	0	NHCOMe	R-OCH(Me)Ph	19n
17	-0.62	-0.69	0.07	3.46	2.17	1	NHCOPh	S-O-*sec*-C_4H_9	19n
18	-0.76	-0.69	0.06	3.46	2.17	1	NHCOPh	OC_4H_9	19t
19	-0.78	-0.69	0.08	3.46	2.17	1	NHCOPh	O-*i*-C_4H_9	19t
20	-0.80	-1.19	0.39	3.18	0.79	1	NHCO-4-pyridyl	OMe	19r
21	-0.88	-1.54	0.66	2.13	0.79	0	$NHCOOC_2H_5$	OMe	19r
22	-0.90	-1.27	0.37	3.46	0.79	1	NHCOPh	OMe	19t
23	-0.90	-0.89	0.01	3.46	1.71	1	NHCOPh	OC_3H_7	19t
24	-1.01	-1.08	0.07	3.46	1.25	1	NHCOPh	OEt	19t
25	-1.04	-0.69	0.34	3.46	2.17	1	NHCOPh	S-O-*sec*-C_4H_9	19n
26	-1.07	-1.35	0.28	1.43	0.79	0	$NHCOCF_3$	OMe	19r
27	-1.25	-1.56	0.31	4.52	0.79	1	NHCO-3-indolyl	OMe	19r
28	-1.26	-1.50	0.24	1.98	0.79	0	$NHCOCH_2Cl$	OMe	19r
29	-1.32	-0.89	0.43	3.46	1.71	1	NHCOPh	$OCH(CH_3)_2$	19t
30	-1.39	-1.42	0.03	4.01	0.79	1	NHCOPh-4-NH_2	OMe	19r
31	-1.42	-1.18	0.24	1.49	1.25	0	NHCOMe	OEt	19i
32	-1.60	-1.75	0.15	2.88	0.79	0	NHCO-*i*-C_4H_9	OMe	19r
33	-1.64	-1.64	0.00	2.48	0.79	0	$NHCOCHCl_2$	OMe	19r
34	-1.68	-1.62	0.06	2.42	0.79	0	NHCO-*i*-C_3H_7	OMe	19r
35	-1.74	-1.49	0.24	1.96	0.79	0	NHCOEt	OMe	19r
36	-1.89	-1.37	0.52	1.49	0.79	0	NHCOMe	OMe	19r
37	-1.92	-1.45	0.47	1.79	0.79	0	$NHSO_2Me$	OMe	19r
38	-2.22	-2.02	0.20	3.88	0.79	0	$NHCOCH_2Ph$	OMe	19r
39	-2.27	-1.84	0.43	3.23	0.79	0	$NHCOCH_2SC_2H_5$	OMe	19r

[a] These points not used in deriving eq 19.

Table XXVIII. Development of Eq 19

Intercept	MR-3	*I*-1	MR-1	*r*	*s*	$F_{1,X}$[a]
-1.88	0.50			0.859	0.340	95.7
-1.93	0.43	0.28		0.878	0.322	4.8
-1.29	0.42	0.64	-0.27	0.905	0.291	8.5

[a] $F_{1,30;\alpha=0.001} = 13.3$; $F_{1,30;\alpha=0.05} = 4.2$.

The parameter $k_{cat.}$ has also been determined for a number of acylglycine esters. Equation 19 has been derived from the data of Table XXVII.

$$\log k_{cat.} = 0.42\ (\pm 0.10)\ \text{MR-3} - 0.27\ (\pm 0.19)\ \text{MR-1} + 0.64\ (\pm 0.34)\ I\text{-1} - 1.29\ (\pm 0.49) \quad (19)$$

$n = 36$, $r = 0.905$, $s = 0.291$

The indicator variable *I*-1 is given the value of 1 when R_1 is an aromatic moiety and 0 when R_1 is aliphatic. The correlation of eq 19 is lower than many of our others in terms of *r*; however, *s* is in line with other results such as eq 18 which brings out the lower variance in $k_{cat.}$ for the glycines. Most of the correlation depends on MR-3. *I*-1 and especially MR-1 contribute little to the reduction of the variance. As with eq 18, we find a small negative role for MR-1. The substitution of π for MR in eq 19 as usual yields a poorer correlation ($r = 0.859$, $s = 0.351$). Three data points not used in the formulation of eq 19 all contain a heterocyclic ring. The development of eq 19 is given in Table XXVIII and the correlation among the variables in Table XXIX.

Table XXIX. Squared Correlation Matrix of Variables for Eq 19

	MR-1	MR-3	*I*-1
MR-1	1.00	0.11	0.57
MR-3		1.00	0.27
I-1			1.00

Discussion

Ligand Binding ($1/K_m$). It is assumed that for a first approximation, $1/K_m$ can be taken as a binding constant. The correlation of binding by eq 1 is surprisingly good for such a relatively simple equation. Except for the isopropyl group interacting in ρ_2 space (−0.63 *I*-1), no special steric effects were uncovered which highlight the large amount of flexibility around the active site. Valine analogues are about four times less well fit than other R_2 groups.

The results of eq 1 can be compared with our recent study[2e] on D and L amides of acylamino acids [$R_2CH(CONH_2)NHCOR_1$] correlated by eq 20. In this equation,

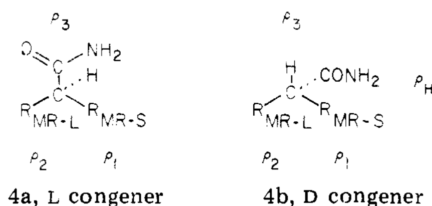
$$\log 1/K = 0.72\ \text{MR-L} + 0.23\ \text{MR-S} + 0.32\ I\text{-1} + 0.31\ I\text{-2} - 1.06 \quad (20)$$

$n = 45$, $r = 0.928$, $s = 0.235$

K stands for K_m (L substrates) or K_i (D inhibitors and glycine analogues). MR-L and MR-S refer to the larger

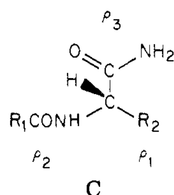
and smaller of R_1 and R_2. The indicator variable *I*-1 is assigned the value of 1 for cases where $CONHNH_2$ occurs instead of $CONH_2$. The hydrazides bind about twice as well as the corresponding amides. *I*-2 takes the value of 1 for D inhibitors and glycine analogues which appear to bind about twice as strongly as the L stereoisomers.

Equation 20 is of special interest since two different types of stereoisomers, one acting as a substrate and the other as an inhibitor, fit the same equation. This suggests the following type of binding for the range of congeners studied so far (see structures **4a** and **4b**).



4a, L congener 4b, D congener

In the case of the D analogues covered by eq 20, MR of R_2 is always greater than MR of $NHCOR_1$ (i.e., R-S = NHCOR) so that all of the D congeners fit the above picture with $NHCOR_1$ falling in ρ_1 space and R_2 falling in ρ_2 space. This would place the amide in ρ_H space and explain its inability to function as a substrate. All of the L isomers but three fall in the class where MR of R_2 > MR of $NHCOR_1$ so normal binding (structure **4a**) in ρ_1 and ρ_2 space occurs. In three cases, MR of $NHCOR_1$ > MR of R_2 and, since these molecules are well fit by eq 20, "wrong-way" binding is inferred, as shown in structure C.



C

Apparently this kind of wrong-way binding still allows hydrolysis to occur while wrong-way binding of the type shown in structure **4b** does not allow *amides* to undergo hydrolysis. Unfortunately, of the three examples where "wrong-way" binding of structure C seems to occur, two have MR-L and MR-S values differing by only 10%. In a single case (L-nicotinylalaninamide), MR-L = 3.23 and MR-S = 0.57. This is a good test of eq 20 since this point is very well fit; however, we are testing other derivatives of this type to more firmly establish the ideas behind eq 20.

In the formulation of eq 1, we studied the use of MR-L and MR-S instead of MR-1 and MR-2. Our results shed little light on the problem because of high collinearity (Table III) between these. The problem is also made more difficult in the case of the esters by the small difference in the coefficients of MR-1 and MR-2. In the case of the amides, a greater difference between the coefficients of the MR terms indicates a greater selectivity in binding than with the esters. We are making a set of esters in which the collinearity among these vectors will be minimized. This should resolve the dilemma.

The D form of the much more easily hydrolyzed esters behaves differently than the amides. Some D esters act as substrates, some as inhibitors. It is of interest that those acting as substrates (Table IV) have, in general, MR-1 ≫ MR-2 while those acting as inhibitors (Table VII) have MR-2 ≫ MR-1. With D substrates binding as shown in structure **2b**, the larger R_1CONH– moiety preempts the better binding space. This results in an arrangement suitable for hydrolysis; however, when the better binding R_2 preempts ρ_2 space, the D configuration binds as in structure **2c**. This parallels the amide binding of structure **4b** and results in inhibition. If the above hypotheses are correct, one might be able to make D-amide substrates by using small R_2 groups and large $NHCOR_1$ groups. This does not appear to have been tested and we are pursuing this possibility. The above hypotheses about binding give a reasonably self-consistent view which offers many ideas for checking their validity.

The acidic and neutral inhibitors correlated by eq 6 and 7 fit in with the results of eq 1. In each of these examples, the slope is essentially 1, suggesting the same kind of binding in ρ_2 space as eq 1.

The correlation of glycinates ($R_1CONHCH_2COOR_3$) is good in terms of fit. The meaning of the coefficients with MR-1 and MR-3 is not easily deduced. The premise underlying the binding of structure B is that, since ρ_2 appears to be the most effective binding site and MR-3 has the largest coefficient, MR-3 binds in ρ_2 space. Since the coefficient with MR-3 in eq 5 is about 30% smaller than those of eq 1, 6, and 7, this could be taken to mean that binding is occurring in ρ_3 space. The coefficient with MR-3 in eq 1 is rather close to that of eq 5 (see discussion of $k_{cat.}$ for glycinates).

Acylation (k_2). The results of eq 8 seem reasonable when compared with eq 1. The slope of MR-2 is identical, indicating the crucial role of R_2 in ρ_2 space. Since the acylation step is the most demanding sterically, one would expect and finds that the isopropyl group has an even greater negative coefficient than in eq 1. The negative coefficient with MR-1 in the acylation step shows that large groups hinder acylation. It seems that as the carbonyl group assumes the necessary tetrahedral character of the transition state, large R_1 groups are a drag on the attainment of the necessary geometry. Even though the correlation coefficient of eq 8 is high, the standard deviation is higher than that of eq 1 despite the fact that all of the data for eq 8 come from the same laboratory. This supports the view that steric interactions in the acylation process are most critical.

The nitrophenyl esters of eq 9–14 do not add much to our general perspective of the acylation step except in that they give clear support for the importance of MR rather than π. In these examples, r^2 for the correlation between parameters ranges from 0.27 to 0.42. The two vectors are more orthogonal in these equations than in any of the other examples. π gives a much poorer result in each of these three examples.

Deacylation (k_3). The correlation equation for deacylation of congeners of the type $RCOOC_6H_4$-4-NO_2 (eq 16) has an MR term with coefficient of 0.34. This is lower than that found for the chiral acyl esters of eq 17. It may be that this is due to a poor quality fit in ρ_2 space. A less likely possibility is that R is binding in ρ_3 space.

Since MR for the 4-phenyl moiety is in general larger than MR of R, the phenyl group might be sequestered by ρ_2 space, placing R in ρ_3 space. This hypothesis does not square with the results of eq 12 and 14. For the present we have no convincing explanation for the small coefficient of MR in eq 16.

In the deacylation of eq 17, it is unlikely that R_2 is not in ρ_2 space. Although the coefficient with this term is lower than eq 1, it is positive. It is hard to find an explanation for this other than that R_2 is holding the substrate in a favorable conformation for attack by the nucleophile and displacement from the enzyme.

The importance of favorable geometry in this step is brought out by the variable *I*-1. Its negative coefficient

shows that, on the average, D-acyl moieties are 65 times more difficult to displace than L analogues. As usual in ρ_2 space, the isopropyl group shows its deleterious effect.

Overall Hydrolysis ($k_{cat.}$). Equation 18 for the hydrolysis of chiral esters brings out what we have noted with the simpler parameters. The most significant feature is the great importance of MR in ρ_2 space. Although $k_{cat.}$ may not be entirely independent of K_m since both have similar cross-product terms, the fact that only negative MR-1 and MR-3 terms occur in eq 18 in contrast to the positive terms in these variables in eq 1 establishes a quite different substituent effect on $k_{cat.}$. The negative MR-1 effect in eq 18 is what one would expect from the results on the acylation step (eq 8). There is a gratifying self-consistency throughout the QSAR of K_m, k_2, k_3, $k_{cat.}$, and K_i.

Equation 19 ($k_{cat.}$ for glycinates) also shows a negative coefficient for MR-1. The positive coefficient with MR-3 must be associated with the necessary binding in ρ_2 space. This supports the kind of binding pictured in structure B.

General Conclusions

The work on ligand interactions with chymotrypsin is so voluminous that many workers tend to publish new data with little or no attempt to fit their results into the context of the numerous previous studies. In this paper we have made a serious effort to ferret out all of the constants which have been determined under reasonably similar conditions for the simple *N*-acylamino acid esters. We plan a second publication on certain types of inhibitors, such as phosphate esters.

Our correlation equations, based heavily on the extensive studies of Carl Niemann as well as a number of important recent studies, provide a rather good, self-consistent[15] picture of the nonspecific role of substituents acting in ρ_1, ρ_2, and ρ_3 space. Niemann suspected that MR of the substituents must play an important role in K_m but QSAR was not sufficiently developed for him to see clearly what was involved.

In a number of instances, more specific effects are established via terms in σ, E_s, and indicator variables.

The most important general conclusion is that ligands appear to bind in relatively few ways, as indicated in structures **2**, B, **4**, and C. Of course, since we have not obtained perfect correlations, we cannot rule out small contributions from "wrong-way" binding; however, our view contrasts sharply with that of Hamilton et al.[6b] who have postulated that one must consider all possible interactions between the four substituents with four types of space: ρ_1, ρ_2, ρ_3, and ρ_H. They have derived binding constants for such interactions which, of course, produce a good fit of the data because of the large number of variables. While "wrong-way" binding can play an important role sometimes as we have noted, the model of Hamilton et al. is overly complicated and impractical to work with. Our much simpler model leaves 4–8% of the variance in the different parameters unaccounted for, which seems to be a small price to pay for the simpler picture. One must also bear in mind that a fair fraction of the unexplained variance may simply be due to noise in the data. The parameters are from many different laboratories and have been measured under a variety of experimental conditions.

Another point of interest which comes out of the correlation analyses, especially with K_m, is the large amount of flexibility in the enzyme. The fact that binding is linearly dependent on MR-1, MR-2, and MR-3 in eq 1 over such a range of bulky groups argues for flexibility. There are certain exceptions, for example, the poor interaction of valine residues. In general, however, aliphatic and aromatic groups alike are well fit in the same QSAR. This linearity holds until the cross product of MR terms becomes too large. At that point, there is a gradual decrease dependent on total bulk of the three groups. A cut-off point has not been found where binding drops to 0 because of lack of bulk tolerance in ρ_1, ρ_2, and ρ_3 space. In eq 1, the cross-product term MR-1·MR-2·MR-3 indicates that chymotrpsin appears to be flexible enough so that bulk tolerance can be achieved by some partitioning of strain among the three regions.

A significant feature in our results is the positive coefficient which always occurs with MR-2 in the correlation of K_m, k_2, k_3, or $k_{cat.}$. It seems natural that in correlating log $1/K_m$ one should find a positive coefficient with MR-2. This suggests binding in ρ_2 space via dispersion forces, the production of a necessary conformational change, or both. The same effect appears to be necessary in the acylation step. It was surprising to find a large positive coefficient with MR-2 in the deacylation step (k_3). Here one might expect that desorption of R_2 from ρ_2 space would be important and that large groups might hinder this process, resulting in a *negative* coefficient with MR-2. Since this does not occur, it suggests that the primary role of R_2 may be that of causing and maintaining an essential conformational change in the enzyme which is not only essential for acylation but also is crucial in the deacylation step. Such an "induced fit" might facilitate the approach of water molecules in the hydrolysis of the acyl enzyme.

Brot and Bender[16] concluded from a study of $k_{cat.}/K_m$ values that bindings in ρ_1 and ρ_2 were independent processes. Our results show that this is true only up to a point. The cross-product term in eq 1 establishes the limits of such independence. They also concluded that $k_{cat.}/K_m$ is the best parameter for structure–activity studies. It is now clear that by taking into account the importance of MR, good self-consistent correlations can be found for K_m, k_2, or k_3. In fact, the study of these simpler parameters one at a time yields important information which cannot be easily seen in the more complex parameters.

During the past 20 years many studies have been made of chymotrypsin–ligand interactions. The majority of these have attempted to relate small structural changes with activity via σ and E_s. It became clear in our first analysis of such work[2a] that serious progress could not be made without taking account of the nonspecific forces (π and MR). The evidence from our present study shows that MR is a parameter of overwhelming importance. It is now abundantly clear that progress in enzymic structure–activity work cannot be made without the use of π and MR or other suitable parameters to account for hydrophobic and dispersion forces.

The molar refractivity is defined as (mol wt/d)[$(n^2 - 1)/(n^2 + 2)$]. Since the range in the refractive index (n) is small, MR in essence is an adjusted molar volume. Parachor (mol wt/$d{\cdot}\gamma^{1/4}$) is a similar parameter which has also been employed in QSAR studies. In parachor, the molar volume is adjusted by use of surface tension (γ). We attempted to see if parachor used instead of MR would yield different results. For the data at hand, parachor and MR are almost perfectly collinear so that no judgment can be made about their relative value.

Exactly what correlation with MR rather than π means is not clear. In introducing MR into biochemical correlations, Pauling and Pressman[17] suggested that this parameter could be employed to assess the role of dispersion forces. In addition to the binding of ligands by such forces,

the effective volume of the substituent is also modeled by MR. The volume of ligand or certain parts of it may play a crucial role in modulating enzymic processes.

Our disposition is to assume that π truly models partitioning with the concomitant desolvation of substituents into enzymic space. This means creating a cavity in the enzyme with possible attendant conformational changes. When good correlations are obtained with MR rather than π, we assume that desolvation is not the main driving force. Since MR is closely related to molar volume, one explanation is that its role is the production of a necessary conformational change. Highly polar groups such as OH, $CONH_2$, etc., might not be desolvated in this process.

MR might also be related to a second kind of hydrophobic bonding suggested by Franks.[18] In this process, two solvated groups are postulated to "freeze" together without loss of water between them. This process would depend heavily on dispersion forces. We have suggested this possibility in other enzymic QSAR.[2e] We have also noted that the active site in chymotrypsin is surrounded by primarily hydrophilic rather than hydrophobic residues. This is also true of one binding site in papain where we believe that MR is operative. It is possible that the basic difference in correlations based on π or MR relates to amino acid residues at the binding site.

The results in this paper, along with others,[1,2] help to establish the fact that one can normally expect with the tools presently at hand to formulate good quantitative expressions to describe how macromolecular receptors recognize their substrates. We find, as others have also noted,[20] that hydrogen bonding is, in general, not an important factor in binding substrate and ligand. Moreover, evidence is beginning to accumulate to indicate that it is quite feasible to deal with the problem of different kinds of activity shown by stereoisomers. This can be accomplished via separate equations (e.g., eq 1, 2, and 3) or in single equations (such as eq 17 and 20).

Now that it is clear that two nonspecific parameters (π and MR) are needed to rationalize the interactions of ligands with macromolecules in vitro, experiments must be designed to explore the implications of MR for in vivo interaction of ligands with enzymes whose in vitro QSAR has been established. We are currently exploring two enzyme systems from this point of view.

Method

All of the values for K_m, K_i, k_2, k_3, and $k_{cat.}$ are from the papers in ref 19. The values for π and MR are largely from our recent compilation or calculated by combination of these values.[7] The π values for $NHCOR_1$ are simply log P values of amides, R_1CONH_2. It was necessary to measure two new values.

$$\log P_{C_6H_5CH_2OCONH_2} = 1.20$$
$$\log P_{C_5H_4N\text{-}3\text{-}CONH_2} = -0.37$$

The π values for $-OR_3$ are from CH_3OR_3. Ingold has discussed the nature of MR.[21] Quaile has listed many values for parachor and discussed its additivity.[22] Experimental values were used[23] for the aromatic structures of pyridine and benzene. All MR values have been scaled by 0.1. Most of the σ^* values are from the table of Leffler and Grunwald,[24] others are from Nagai et al.[25] In the case of eq 16, we employed the value used by Dupaix et al.[19w]

In the beginning of our analysis, we thought that the electronic effect of R_1 on the $-NHCO-$ might influence ligand interaction. To explore this, we used pK_a values of the corresponding acids as a parameter but found no significant improvement in results.

The steric parameters are from Taft's listing.[26]

It must be emphasized that the use of triple cross-product terms in eq 1 and 18 is strictly a phenomenological approach.

In developing the various correlation equations where different sets of molecules from different laboratories were involved, we normally developed correlation equations for these sets individually (where enough data points were available) and only merged the sets after we were satisfied that parallel results were being obtained. We have not used stepwise regression analysis but have studied all possible regression equations.[27]

Acknowledgment. This investigation was supported by U.S. Public Health Service Research Grant No. CA-11110 from the National Cancer Institute. We wish to thank Professor R. Nelson Smith for valuable discussions.

References and Notes

(1) (a) C. Hansch, E. W. Deutsch, and R. N. Smith, *J. Am. Chem. Soc.*, **87**, 2738 (1965); (b) C. Silipo and C. Hansch, *ibid.*, **97**, 6849 (1975); (c) C. Hansch, *Adv. Pharmacol. Chemother.*, **13**, 45 (1975); (d) C. Hansch and M. Yoshimoto, *J. Med. Chem.*, **19**, 71 (1976).
(2) (a) C. Hansch and E. Coats, *J. Pharm. Sci.*, **59**, 731 (1970); (b) C. Hansch, *J. Org. Chem.*, **37**, 92 (1972); (c) R. N. Smith and C. Hansch, *Biochemistry*, **12**, 4924 (1973); (d) R. N. Smith, C. Hansch, and T. P. Poindexter, *Physiol. Chem. Phys.*, **6**, 323 (1974); (e) M. Yoshimoto and C. Hansch, *J. Org. Chem.*, **41**, 2269 (1976).
(3) M. L. Bender and F. J. Kézdy, *Annu. Rev. Biochem.*, **34**, 49 (1965).
(4) (a) D. M. Blow, *Acc. Chem. Res.*, **9**, 145 (1976); (b) W. P. Jencks, "Catalysis in Chemistry and Enzymology", McGraw-Hill, New York, N.Y., 1969, pp 218–229.
(5) C. Hansch, *J. Med. Chem.*, **19**, 1 (1976).
(6) G. E. Hein and C. Niemann, *J. Am. Chem. Soc.*, **84**, 4487, 4495 (1962); (b) C. L. Hamilton, C. Niemann, and G. Hammond, *Proc. Natl. Acad. Sci. U.S.A.*, **55**, 664 (1966).
(7) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, **16**. 1207 (1973).
(8) J. R. Rapp, C. Niemann, and G. Hein, *Biochemistry*, **5**, 4100 (1966).
(9) C. Hansch and D. F. Calef, *J. Org. Chem.*, **41**, 1240 (1976).
(10) O. Exner, *Collect. Czech. Chem. Commun.*, **40**, 2781 (1975).
(11) I. V. Berezin, N. F. Kazanskaya, and A. A. Klyosov, *FEBS Lett.*, **15**, 121 (1971).
(12) H. Neurath and B. S. Hartley, *J. Cell. Comp. Physiol.*, **54**, Suppl. 1, 179 (1959).
(13) J. E. Purdie and N. L. Benoiton, *Can. J. Biochem.*, **48**, 1058 (1970).
(14) J. G. Topliss and R. J. Costello, *J. Med. Chem.*, **15**, 1066 (1972).
(15) S. H. Unger and C. Hansch, *J. Med. Chem.*, **16**, 745 (1973).
(16) F. E. Brot and M. L. Bender, *J. Am. Chem. Soc.*, **91**, 7187 (1969).
(17) L. Pauling and D. Pressman, *J. Am. Chem. Soc.*, **67**, 1003 (1945).
(18) F. Franks in "Water", Vol. 4, F. Franks, Ed., Plenum Press, New York, N.Y., 1974, Chapter I.
(19) (a) J. B. Jones, T. Kunitake, C. Niemann, and G. E. Hein, *J. Am. Chem. Soc.*, **87**, 1777 (1965); (b) J. B. Jones, C. Niemann, and G. E. Hein, *Biochemistry*, **4**, 1735 (1965); (c) J. R. Rapp, C. Niemann, and G. E. Hein, *ibid.*, **5**, 4100 (1966); (d) G. E. Hein and C. Niemann, *J. Am. Chem. Soc.*, **84**, 4495 (1962); (e) A. Williams, *Biochemistry*, **9**, 3383 (1970); (f) M. L. Bender, *J. Am. Chem. Soc.*, **84**, 2582 (1962); (g) J. B. Jones and C. Niemann, *Biochemistry*, **1**, 1093 (1962); (h) S. G. Cohen, J. Crossley, and E. Khedouri, *ibid.*, **2**, 820 (1963); (i) D. W. Ingles and J. R. Knowles, *Biochem. J.*, **99**, 275 (1966); (j) I. V. Berezin, N. F. Kazanskaya, and A. A. Klyosov, *FEBS Lett.*, **15**, 121 (1971); (k) B. Zerner, R. P. M. Bond, and M. L. Bender, *J. Am. Chem. Soc.*, **86**, 3674 (1964); (l) J. J. Bĕchet, A. Dupaix, and C. Roucous, *Biochemistry*, **12**, 2566 (1973); (m) S. G. Cohen, R. M. Schultz,

and S. Y. Weinstein, *J. Am. Chem. Soc.*, **88**, 5315 (1966); (n) Y. Y. Lin, D. N. Palmer, and J. B. Jones, *Can. J. Chem.*, **52**, 469 (1974); (o) R. E. Williams and M. L. Bender, *Can. J. Biochem.*, **49**, 210 (1971); (p) R. J. Foster and C. Niemann, *J. Am. Chem. Soc.*, **77**, 3370 (1955); (q) H. T. Huang and C. Niemann, *ibid.*, **74**, 101 (1952); (r) J. P. Wolf III and C. Niemann, *Biochemistry*, **2**, 82 (1963); (s) J. P. Wolf III, R. A. Wallace, R. L. Peterson, and C. Niemann, *ibid*, **3**, 940 (1964); (t) B. Zerner and M. L. Bender, *J. Am. Chem. Soc.*, **86**, 3369 (1964); (u) R. Wildnauer and W. J. Canady, *Biochemistry*, **9**, 2885 (1966); (v) C. D. Hubbard and J. F. Kirsch, *ibid.*, **11**, 2483 (1972); (w) A. Dupaix, J. J. Béchet, and C. Roucous, *ibid.*, **12**, 2559 (1973); (x) T. H. Fife and J. B. Milstien, *ibid.*, **6**, 2901 (1967); (y) M. Caplow and W. P. Jencks, *ibid.*, **1**, 883 (1962); (z) K. Martinek, V. N. Dovoska, and S. D. Varfolomeev, *Biokhimiya*, **37**, 1245 (1972); (aa) M. S. Silver and M. S. Matta, *Arch. Biochem. Biophys.*, **151**, 62 (1972); (bb) J. L. Miles, D. A. Robinson, and W. J. Canady, *J. Biol. Chem.*, **238**, 2932 (1963); (cc) A. J. Hymes, D. A. Robinson, and W. J. Canady, *ibid.*, **240**, 134 (1965).

(20) C. Chothia and J. Janin, *Nature (London)*, **256**, 705 (1975).

(21) C. K. Ingold, "Structure and Mechanism in Organic Chemistry", 2nd ed, Cornell University Press, Ithaca, N.Y., 1969, pp 142–152.

(22) O. R. Quaile, *Chem. Rev.*, **53**, 439 (1953).

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

(24) J. E. Leffler and E. Grunwald, "Rates and Equilibria of Organic Reactions", Wiley, New York, N.Y., 1963, p 222.

(25) Y. Nagai, H. Matsumoto, and H. Watanabe, *Bull. Chem. Soc. Jpn.*, **45**, 2560 (1972).

(26) R. W. Taft in "Steric Effects in Organic Chemistry", M. S. Newman, Ed., Wiley, New York, N.Y., 1956, p 598.

(27) J. Y. Fukunaga, C. Hansch, and E. E. Steller, *J. Med. Chem.*, **19**, 605 (1976).

Improved Delivery through Biological Membranes. 4. Prodrugs of L-Dopa[1]

Nicholas Bodor,* Kenneth B. Sloan, Takeru Higuchi,

INTERx Research Corporation, Lawrence, Kansas 66044

and Kunihiro Sasahara

Sankyo Co., Ltd., Tokyo, Japan. *Received March 24, 1977*

Various classes of transient derivatives of L-Dopa have been synthesized, systematically protecting one or more of the main sites of metabolism in the molecule: the carboxy function, the amino, and/or the catechol system. The derivatives studied include carboxy esters, phenol esters, amides, peptides, and various combinations of these functions. A number of these derivatives effectively prevent the metabolism of L-Dopa prior to and/or during the absorption process, resulting in a significantly better bioavailability of the drug. In vivo studies using dogs showed up to 2.5-fold increase in L-Dopa blood levels. The metabolism as well as toxicity aspects of the prodrugs is also discussed.

L-Dopa (L-3,4-dihydroxyphenylalanine) is still generally accepted as the first drug of choice in the management of Parkinsonism. Long-term therapy with L-Dopa is, however, associated with a number of therapeutic problems.[2,3] The most serious limitations of L-Dopa can be summarized as follows: poor bioavailability, wide range of interpatient variations of plasma levels, unpredictable therapeutic response, and various side effects. The main factors responsible for these problems are the physical–chemical properties of the drug substance: low water solubility resulting in incomplete dissolution at and prior to the absorption site, low lipid solubility resulting in unfavorable partition, and the high susceptibility of the drug molecule to chemical and enzymatic degradation.

L-Dopa is usually administered orally and, in man as in dog, the material in solution appears to be well adsorbed, primarily in the small bowel by a special carrier transport mechanism. But, in fact, the drug is extensively metabolized in the gastrointestinal tract and/or during its first passage through the liver, so that relatively little arrives in the blood as intact L-Dopa. This metabolism of L-Dopa is unfavorable to its therapeutic intent.[4] Also, individual differences in the degree of breakdown during the passage of the drug through the gastrointestinal tract may be responsible for the highly variable blood levels observed in patients receiving similar doses.[5]

The peripheral side effects appear to be due to one or more biotransformation products rather than L-Dopa itself. Eighteen metabolites of L-Dopa were detected in the urine,[7] but the major metabolism during and prior to absorption involves primarily decarboxylation and conjugation. L-Aromatic amino acid decarboxylase has a high activity in the gastric mucosa.[8] Also, conjugation of L-Dopa and its metabolites takes place predominantly in the gastrointestinal–hepatic system.[9] L-Dopa is rapidly and continuously metabolized in blood, since only 5–8% of it is protein-bound, making it very susceptible to metabolic processes.[5] Finally, the remaining intact L-Dopa is rapidly taken up by the brain and localized in the caudate nucleus.[10,11]

Although peripheral DC inhibitors, such as L-α-methyl-Dopa hydrazine, reduced the L-Dopa requirements by up to 80%, a number of other problems related to L-Dopa therapy are still unsolved, primarily the complex problem of bioavailability of the drug, including the dissolution–absorption–metabolism processes, prior to delivery to the blood. The other still existing major problem is the side effects caused by L-Dopa. Although a close correlation between some dyskinesias and plasma L-Dopa concentration has been found,[12] other forms of involuntary movements are not associated with high plasma L-Dopa concentrations.[6]

In order to improve the therapeutic value of L-Dopa, we have carried out a systematic study of transient derivatives (prodrugs) of the molecule aimed at solving one or more of the above problems. Based on the previous evaluation of the physical–chemical properties of L-Dopa, an ideal prodrug of L-Dopa should be soluble in water and in lipids, completely adsorbed from the gastrointestinal tract without any chemical degradation or metabolism, and thus deliver L-Dopa intact in the blood stream, at a reproducible therapeutic level. A circulating prodrug which is transformed to L-Dopa might also alter the peripheral metabolism of L-Dopa. In addition, a prodrug of L-Dopa coadministered with a decarboxylase inhibitor could combine the best aspects of these two approaches. But