0026-895X/86/040436-11$02.00/0

MOLECULAR PHARMACOLOGY, 29:436–446

Quantitative Structure-Activity Relationships and Molecular Graphics in Ligand Receptor Interactions: Amidine Inhibition of Trypsin

MAURIZIO RECANATINI,[1] TERI KLEIN, CHUN-ZHENG YANG,[2] JUDITH McCLARIN, ROBERT LANGRIDGE, and CORWIN HANSCH

Department of Chemistry, Pomona College, Claremont, California 91711 (M.R., C.-Z.Y., C.H.), Computer Graphics Laboratory, University of California, San Francisco, California 94143 (T.K.), and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260 (J.M.)

Received October 7, 1985; Accepted December 18, 1985

SUMMARY

Quantitative structure-activity relationships have been formulated for four sets of amidine inhibitors of trypsin. The quantitative results from these equations are compared with qualitative models constructed from the X-ray crystallographic coordinates of a benzamidine bound to trypsin. The good agreement between the mathematical and graphics models provides further support for the use of substituent constants and regression analysis in the study of enzyme-ligand interactions.

Recent studies on ligand-enzyme interactions using QSARs (1) and color stereo computer graphics have shown the combination to be a powerful means for enhancing our understanding of how organic compounds react with bioreceptors (2–6). The method shows promise in the field of drug design as well as in basic science. Although there are few drug receptors whose structures are known in detail, the X-ray crystallographic structures of many enzymes have been determined. These enzymes make excellent models for studying how bioreceptors recognize and bind small molecules. In this report we consider benzamidines inhibiting trypsin.

An important step in blood coagulation is the activation of complement and fibrinolysis. These processes are mediated by serine proteases; hence, medicinal chemists have developed an interest in finding effective inhibitors of these enzymes. Soon after the discovery that benzamidine was a good inhibitor of trypsin (1), Baker and Erickson (7) undertook systematic studies of trypsin. Baker (8) has reviewed the early structure activity studies with this protease. Although there have been a number of studies of trypsin inhibition by benzamidines (8, 9), three studies have been published in which felicitous selection of substituents from the point of view of QSARs have been made (10–12). It is these studies with which we are concerned.

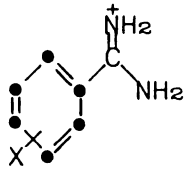
This research was supported by Grants GM-30362 (C. H.) and RR-1081 (R. L.) from the National Institutes of Health.

[1] Visiting Research Professor from the Institute of Pharmaceutical and Toxicological Chemistry, University of Bologna, Bologna, Italy.

[2] Visiting Scientist from the Institute of Hematology, Chinese Academy of Medical Science, Tianjin, China.

For construction of mathematical models, the data of Mares-Guia *et al.* (10), on 4-substituted benzamidines I are first considered. They derived Eq. 1 for the action of 10 benzamidines on bovine trypsin.

$$\log 1/K_i = -0.88\sigma + 4.40 \qquad (1)$$
$$n = 10, \quad r = 0.818$$



I

Eq. 1 accounts for 67% of the variance in log $1/K_i$. Mares-Guia *et al.* (10) also noted that there was a positive correlation between log $1/K_i$ and q, where q is the calculated net π-electron density at the central atom of each substituent. The π-electron densities were obtained from molecular orbital calculations.

An interesting feature of Eq. 1 is the meaning of the negative slope of the σ relationship which shows that electron release by substituents increases inhibitor potency. It is well known that the ionization constant of amidines is highly sensitive to the electronic effect of substituents. The range of pK_a' values for the substituted benzamidines of Mares-Guia *et al.* (10) was from 10.14 (4—NO_2) to 12.69 (4—OH). Since the K_i values were determined at pH 8, all of the amidines are essentially completely protonated (13, 14) and carry a positive charge on

ABBREVIATIONS: QSARs, quantitative structure-activity relationships; *MR*, molar refractivity; MIDAS, molecular interactive display and simulation; MS, molecular surface.

the amidine group which can interact with the carboxylate of Asp-189 (15). Hence, Mares-Guia *et al.* (10) concluded that the negative slope of Eq. 1 could not be related to the ionization of the benzamidines. They also considered the possibility that a charge transfer interaction occurred with the enzyme but rejected this possibility when they found no correlation with calculated ionization potentials. They considered the most likely interaction with respect to σ to be the dipole-dipole type of reaction between 4-substituents and the OH of what is now termed Ser-195 (15).

A second study of the inhibition of human trypsin by mostly 3-substituted benzamidines (a few 4-substituted congeners were used) was made by Andrews *et al.* (11). From their results they formulated correlation Eq. 2:

$$\log 1/K_i = 0.88MR_m - 0.24M_r + 4.77 \qquad (2)$$
$$n = 15, \quad r = 0.843, \quad s = 0.32$$

In this expression, MR_m is the molar refractivity of the *meta* substituents scaled by 0.1 and M_r is the molecular weight of 3- or 4-substituents, also scaled by 0.1. It is surprising that Andrews *et al.* (11) did not find a role for the electronic effect of substituents. The MR_m term applies only to meta substituents, but the M_r term applies to both *meta*- and *para*-substituents, so that it is surprising that the coefficient of −0.24 would account for the interactions of both *meta*- and *para*-substituents. From the analysis of Andrews *et al.* (11), the only effect of *para*-substituents would seem to be a negative effect on potency related to the size of the 4-substituent. The 3-substituents appear to have an ambivalent effect rationalized by positive MR_m and negative M_r affects.

In a third study, Labes and Hagen (12) derived Eq. 3:

$$\log 1/K_i = 0.21\pi - 0.43\sigma + 0.58HP + 1.39 \qquad (3)$$
$$n = 125, \quad r = 0.82, \quad s = 0.32$$

Although the quality of Eq. 3 is not high in terms of r or s, it does encompass an enormous variation in structure for 125 benzamidines. HP is an indicator variable assigned the value of 1 for certain carbonyl-containing substituents and a value of 0 for all others. Eq. 3 suggests a much smaller electronic effect than Eqs. 1 and 2 and also brings out the possibility of a hydrophobic effect via the π term.

Since the correlations of Eqs. 1–3 are not high and since they are radically different, it seemed important to reinvestigate the results of these studies in combination with computer graphics models of the benzamidines. A high resolution (1.8Å) X-ray crystallographic structure of bovine trypsin with a bound molecule of benzamidine has been published by Bode and Schwager (15).

Methods

Most of the substituent constants used in these studies were taken from Ref. 16, but for a number of substituents, values were lacking and had to be estimated. In the two series 3— and 4—$O(CH_2)_nCH_3$ of Table 3, the value of σ for the higher members was taken as the highest value of the longest known member. For the congeners 3— and 4—COOR (where R is a simple aliphatic group), the values of 3— and 4—$COOCH_3$ were used. For the series 3—COC_6H_4X, 4—$CH{=}CHCOC_6H_4X$, and 4—$NHCOC_6H_4X$, the effect of X was disregarded, and σ for the parent substituent was used. In the ketone series 4—CH_2COR the value for 4—CH_2COCH_3 was employed, and in the two series 4—$(CH_2)_2COOR$ and 4—$(CH_2)_2COR$, the value for 4—$(CH_2)_2COOH$ was used. Finally, in the two sets of ethers, 4—$O(CH_2)_nC_6H_5$ and 4—OCH_2COR, the value of σ for the first substituents was used for all of the others.

The MR values were scaled by 0.1 to keep them somewhat equiscalar with respect to π. Those whose values have not been reported were calculated (16). For example:

$$MR\ O(CH_2)_nCH_3 = MR\ OCH_3 + MR\ nCH_2 = 0.79 + n0.465$$

$$MR\ COCH_2OCH_3 = MR\ COCH_3 - MR\ H + MR\ OCH_3$$
$$= 1.12 - 0.10 + 0.79 = 1.81$$

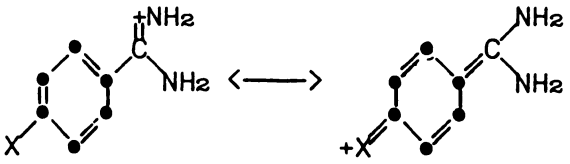
The π' values were calculated using nitrobenzene instead of benzene as the reference compounds:

$$\pi' = \log P\ X{-}C_6H_4{-}NO_2 - \log P\ C_6H_5{-}NO_2$$

For those cases where log P for the substituted nitrobenzene has not been reported, the values were calculated according to additivity principles (16–18).

QSAR of Trypsin Inhibition

In our previous analysis of enzyme-ligand interactions we have relied heavily on physical constants (16) to account for the electronic, hydrophobic, and steric interaction of substituents. Accordingly, we first made preliminary studies comparing σ and $\sigma+$ as suitable electronic parameters. It was thought that resonance interactions of the type shown in scheme II might play some part in the inhibitory action of the benzamidines.



II

In these studies σ proved to be superior to $\sigma+$ and, therefore, appears to rule out a role for through resonance between substituent and amidine.

In selecting suitable π values to model the hydrophobic effect of substituents, we derived π values from the nitrobenzene system since few partition coefficients of benzamidines have been reported.

$$\pi' = \log P_{x\text{—}C_6H_4\text{—}NO_2} - \log P_{C_6H_5\text{—}NO_2}$$

Nitrobenzene was selected because the strong electron-attracting effect of $NO_2(\sigma = 0.78)$ is similar to that of $-C(=NH_2^+)-NH_2$. Although σ for this group has not been reported, its positive charge assures it a high value. There is extensive evidence that strong electron-attracting groups modify π values for substituents attached to the benzene rings (17–19). The π constants of substituents containing lone pair electrons are normally increased in the presence of strong electron-attracting groups.

From the data in Table 1 from Mares-Guia and Shaw (1), we have derived Eqs. 4–6.

$$\log 1/K_i = -0.90(\pm0.51)\sigma + 4.46(\pm0.20) \qquad (4)$$
$$n = 10, \quad r = 0.820, \quad s = 0.279, \quad F_{1,8} = 16.4$$

$$\log 1/K_i = -0.75(\pm0.45)\sigma - 0.20(\pm0.21)B_4 + 4.89(\pm0.47) \qquad (5)$$
$$n = 10, \quad r = 0.902, \quad s = 0.225, \quad F_{1,7} = 5.33$$

$$\log 1/K_i = -0.78(\pm0.57)\sigma - 0.24(\pm0.50)MR_4 + 4.61(\pm0.37) \qquad (6)$$
$$n = 10, \quad r = 0.850, \quad s = 0.274, \quad F_{1,7} = 1.28$$

The 4—Br congener seems to be an outlier, and dropping this point yields the higher correlations of Eqs. 7 and 8.

$$\log 1/K_i = -0.83(\pm0.35)\sigma - 0.18(\pm0.16)B_4 + 4.79(\pm0.37) \quad (7)$$
$$n = 9, \quad r = 0.953, \quad s = 0.167, \quad F_{1,6} = 7.35$$

$$\log 1/K_i = -0.83(\pm0.42)\sigma - 0.29(\pm0.37)MR_4 + 4.59(\pm0.27) \quad (8)$$
$$n = 9, \quad r = 0.935, \quad s = 0.195, \quad F_{1,6} = 3.77$$

Eqs. 4–8 show that, in addition to the electronic effect of substituents on benzamidine potency, there is also a negative effect due to bulky *para* substituents. In deriving Eqs. 4–6, one data point, COO^-, was omitted by us as well as by Mares-Guia *et al.* (10). It is now well known that σ constants for charged groups are not constant and, hence, these substituents do not normally fit in well with neutral substituents in linear free energy relationships (20). Using σ for COO^-, it can be seen in Table 1 that this point is poorly fit. However, since COO^- is positioned in the enzyme so that it could interact with the H of Ser-195, it might be more appropriate to use σ for COOH instead of COO^-. In fact, a better fit is obtained by making this substitution.

Two types of steric parameters (B_4 and MR_4) have been used in Eqs. 5–8. B_4 is one of the steric parameters of Verloop *et al.* (21). It is the largest of the width parameters. Although one finds qualitatively the same effect using either MR_4 or B_4 (note that the subscript 4 on B refers to the type of parameter and not to the 4-position on the benzamidine), Eq. 8 is not very significant in terms of the F statistic. The correlation between σ and MR or B_4 is low ($r^2 \sim 0.17$). The correlation between MR and B_4 is high ($r^2 = 0.85$).

Analysis of the work of Andrews *et al.* (11) (Table 2) on the inhibition of human trypsin by 3- and 4-substituted benzamidines yields Eqs. 9–11.

$$\log 1/K_i = -0.44(\pm0.27)B_4 + 5.37(\pm0.49) \quad (9)$$
$$n = 14, \quad r = 0.713, \quad s = 0.395, \quad F_{1,12} = 12.4$$

$$\log 1/K_i = -0.49(\pm0.19)B_4 - 0.80(\pm0.58)\sigma + 5.48(\pm0.39) \quad (10)$$
$$n = 14, \quad r = 0.855, \quad s = 0.305, \quad F_{1,11} = 9.13$$

$$\log 1/K_i = -0.42(\pm0.20)B_4 - 0.71(\pm0.52)\sigma + 0.18(\pm0.18)\pi_3' + 5.30(\pm0.39) \quad (11)$$
$$n = 14, \quad r = 0.904, \quad s = 0.263, \quad F_{1,10} = 4.75$$

The stepwise development of Eqs. 9–11 shows the relative importance of the parameters. The steric effect modeled by B_4 (for *para*-substituents) is more important than σ which is the reverse of that found in Eqs. 4–6. Also, the coefficient with B_4 in Eq. 11 is larger than that in Eq. 5, which suggests that the steric interactions between 4-substituents and Ser-195 are greater with human trypsin than with bovine trypsin. A steric parameter for 3-substituents was not found to be of value (B_4 for H is used for all 3-substituents). The weighting factor for σ in Eqs. 5 and 11 are essentially identical as one might expect. The hydrophobic parameter π, which has been derived from the nitrobenzene system, *applies only to meta-substituents.* In the case of the 3,5—$(CH_3)_2$ congener, π for only one CH_3 group was employed. The second methyl group has been parameterized only for σ indicating that it does not make contact with Ser, or, if it does, some kind of compensation occurs since it is well fit by Eq. 11. Of the congeners tested by Andrews *et al.* (11), those containing a carboxyl group on the *ortho*-substituent were omitted in the derivation of Eqs. 9–11.

Of the five sterimol parameters of Verloop *et al.* (21), it was found that B_4 and L gave essentially the same result. Because of the collinearity between B_4 and L for the substituents in hand, it is not possible by statistical methods to say which is the best. However, it seems clear from the graphics that B_4 is the more significant parameter. Substituting MR_4 in Eq. 11 yields a slightly poorer correlation ($r = 0.887$).

If π constants from the benzene system are used in Eq. 11 in place of π' from the nitrobenzene system, a poorer correlation is obtained ($r = 0.899$).

In reevaluating the study by Labes and Hagen (12) (Eq. 8), we withheld in preliminary studies those congeners for which they had found it necessary to use an indicator variable *HP*. These substituents contain carbonyl groups. We also withheld basic substituents which would be protonated under the experimental conditions and two disubstituted congeners. The complexity of the substituents precluded us from studying the B parameters of Verloop *et al.* (21), since many of these were not available. The remaining congeners (Table 3) were divided into *meta*- and *para*-groups and the following QSARs were derived.

meta-Substituted Benzamidines

$$\log 1/K_i = 0.25(\pm0.09)\pi_3' + 1.40(\pm0.14) \quad (12)$$
$$n = 25, \quad r = 0.762, \quad s = 0.259, \quad F_{1,23} = 31.8$$

$$\log 1/K_i = 0.21(\pm0.09)\pi_3' - 0.79(\pm0.62)\sigma + 1.63(\pm0.22) \quad (13)$$
$$n = 25, \quad r = 0.823, \quad s = 0.231, \quad F_{1,22} = 6.98$$

Two classes of substituents in Table 3 are not well fit by Eq. 13: (3—CH_2CO_2R, 3—$CH_2COC_6H_4$-*X* and 3—CH═$CHCO_2R$, 3—CH═$CHCOC_6H_4$-*X*). Assigning an indicator variable of *I*-1 to the vinyl-containing substituents and *I*-2 to the others yields Eq. 14:

$$\log 1/K_i = 0.23(\pm0.07)\pi_3' - 0.75(\pm0.53)\sigma + 0.65(\pm0.21)I\text{-}1 + 0.43(\pm0.19)I\text{-}2 + 1.61(\pm0.18) \quad (14)$$
$$n = 37, \quad r = 0.909, \quad s = 0.204$$

The parameters common to Eqs. 13 and 14 are essentially identical, showing that the additional 12 derivatives behave in terms of π and σ like the 25 derivatives on which Eq. 13 is based after one makes a

TABLE 1

Parameters used to derive Eqs. 4–8 for the inhibition of bovine trypsin by benzamidines *X*—$C_6H_4C(=N^+H_2)NH_2$

X	log 1/K_i Observed[a]	log 1/K_i Calculated[b]	|Δlog 1/K_i|	σ	MR_4	B_4
1. 4—NH_2	5.08	5.02	0.06	−0.66	0.54	1.84
2. 4—OH	4.85	4.78	0.07	−0.37	0.28	1.93
3. H	4.78	4.69	0.09	0.00	0.10	1.00
4. 4—Br	4.72	4.32	0.40	0.23	0.89	1.95
5. 4—CH_3	4.58	4.61	0.03	−0.17	0.56	2.04
6. 4—F	4.37	4.57	0.20	0.06	0.09	1.35
7. 4—Cl	4.27	4.35	0.08	0.23	0.60	1.80
8. 4—OCH_3	4.23	4.52	0.29	−0.27	0.79	2.87
9. 4—$COOC_2H_5$	3.82	3.69	0.13	0.45	1.75	4.29
10. [c]4—COO^-	3.70	4.35	0.65	0.45	0.69	2.66
11. 4—NO_2	3.64	3.81	0.17	0.78	0.74	2.44

[a] From Ref. 6.
[b] Calculated by Eq. 5.
[c] This data point was omitted in deriving Eqs. 4–8.

TABLE 2

Parameters used to derive Eqs. 9–11 for the inhibition of human trypsin by benzamidines X—$C_6H_4C(=N^+H_2)NH_2$

X	log $1/K_i$		$\lvert\Delta \log 1/K_i\rvert$	σ	MR_3	π_3'	B_4
	Observed[a]	Calculated[b]					
1. 3—C_6H_5	5.40	5.17	0.23	0.06	2.53	1.89	1.00
2. 3—$CH_2C_6H_5$	5.20	5.32	0.12	−0.08	3.00	2.22	1.00
3. 3—NH_2	5.20	4.91	0.29	−0.16	0.54	−0.46	1.00
4. 3—$O(CH_2)_3OC_6H_5$	5.00	5.16	0.16	0.12	4.37	2.12	1.00
5. H	5.10	4.87	0.23	0.00	0.10	0.00	1.00
6. 3—OCH_3	4.90	4.84	0.06	0.12	0.79	0.31	1.00
7. 3—CH_2OH	4.50	4.76	0.26	0.00	0.72	−0.64	1.00
8. 3—NO_2	4.10	4.30	0.20	0.71	0.74	−0.36	1.00
9. 4—OCH_3	4.00	4.27	0.27	−0.27	0.10	0.00	2.87
10. 4—NO_2	3.90	3.70	0.20	0.78	0.10	0.00	2.44
11. 4—OC_2H_5	4.00	4.04	0.04	−0.24	0.10	0.00	3.36
12. 4—CH_2OH	4.10	4.15	0.05	0.00	0.10	0.00	2.70
13. 3,4—$(CH_3)_2$	5.10	4.71	0.39	−0.24	0.57	0.60	2.04
14. 3,5—$(CH_3)_2$	4.80	5.08	0.28	−0.14	0.57	0.60	1.00

[a] From Ref. 11.
[b] Calculated by Eq. 11.

correction for what appears to be the effect of a carbonyl group noted by Labes and Hagen (12). In both Eqs. 13 and 14 it is found that the coefficient with σ is in excellent agreement with the corresponding parameters in Eqs. 5–8, 10, and 11. Indeed, it is satisfying to find such good agreement on three different data sets from three different laboratories for complex reactions in which hydrophobic and steric interactions must be accounted for before the agreement can be seen. The Hammett equation (16) was not designed for the reaction of complex molecules interacting with heterogeneous surfaces.

Nevertheless, we have found striking agreement for the coefficients with σ, also where corrections were needed for steric and hydrophobic effects, with four cysteine hydrolases acting on a set of esters of the same structure, but with differences in the substituents (22).

The coefficient with the π' term in Eq. 14 agrees well with that in Eqs. 3 and 11. Because of the large variation in hydrophobicity of the *meta*-substituents used to derive Eq. 13, π' is the most important parameter and, because of less variation in π' for the substituents used for Eqs. 5–8, π' is not a very important term. Two data points [3—$COCH(CH_3)_2$ and 3—$NHCOC_6H_5$] are quite poorly fit and for this reason are not included in the analysis.

para-Substituted Benzamidines

$$\log 1/K_i = 0.27(\pm0.06)MR_4 + 0.85(\pm0.18) \qquad (15)$$
$$n = 55, \quad r = 0.778, \quad s = 0.329, \quad F_{1,53} = 81.2$$

$$\log 1/K_i = 0.24(\pm0.05)MR_4 - 0.76(\pm0.26)\sigma + 0.95(\pm0.15) \qquad (16)$$
$$n = 55, \quad r = 0.875, \quad s = 0.256, \quad F_{1,52} = 35.7$$

$$\log 1/K_i = -0.52(\pm0.45)MR_4 + 0.84(\pm0.50)\log(\beta \cdot 10^{MR_4} + 1) - 0.73(\pm0.24)\sigma + 1.41(\pm0.31) \qquad (17)$$
$$n = 55, \quad r = 0.900, \quad s = 0.235,$$
$$F_{2,50} = 5.77 \quad \text{optimum } MR_4 = 1.11$$

Using an indicator variable I-3 for 12 carbonyl-containing side-chains (see Table 3), Eq. 18 was formulated:

$$\log 1/K_i = -0.59(\pm0.52)MR_4 + 0.88(\pm0.55)\log(\beta \cdot 10^{MR_4} + 1) - 0.74(\pm0.23)\sigma + 0.51(\pm0.16)I\text{-}3 + 1.38(\pm0.30) \qquad (18)$$
$$n = 67, \quad r = 0.928, \quad s = 0.233, \quad \text{optimum } MR_4 = 1.03$$

In the stepwise development of Eq. 17, it is seen that (in contrast to Eqs. 6, 8, and 11) the most important term is MR_4 (Eq. 15) with a *positive* coefficient; however, the sign of this term becomes negative in the final equation (Eq. 17). That is, there is a small steric effect which is seen with small substituents (up to $MR_4 = 1$) but which is overshadowed by the positive MR effect of the large lengthy substituents. In Eq. 17 one sees a negative slope of −0.52 with MR_4 which brings out the same unfavorable effect on potency caused by the bulk of substituents until $MR_4 = 1.11$ (as usual, MR_4 has been scaled by 0.1 to make it somewhat more equiscalar with π). At this point the slope becomes positive: 0.32(−0.52 + 0.84). Thus, it appears that, after an initial bad contact with the enzyme, the more lengthy substituents make positive contact. Therefore, small substituents and the first part of large substituents exhibit a steric effect comparable to that found in Eqs. 6, 8, and 11. Substituents the size of C_2H_5, Br, etc., show only the negative effect. What is most gratifying to see is that, after the large amount of variance in log $1/K_i$ caused by MR_4 is separated, the same coefficient occurs with σ as in the other equations.

Eq. 18, which contains an additional 12 data points, accounted for by I-3, has essentially identical parameters. Note that the indicator variables I-1, I-2 in Eq. 14 and I-3 in Eq. 18 have coefficients close to that in Eq. 3. Labes and Hagen (12) have postulated that this increase in activity by about a factor of 3 is due to interaction between the OH of Ser-195 and the positively charged carbon of the carbonyl group in the side-chain of these substituents.

There is a significant difference between the intercepts of Eq. 14 and those of Eq. 18. The lower value with Eq. 18 for the *para*-isomers appears to account for part of the steric effect of these substituents—the rest is taken care of by the negative MR_4 term. If this difference is accounted for by the indicator variable, I-M (I-M = 1 for *meta*-substituents and 0 for *para*), then the two data sets can be merged (Eq. 19):

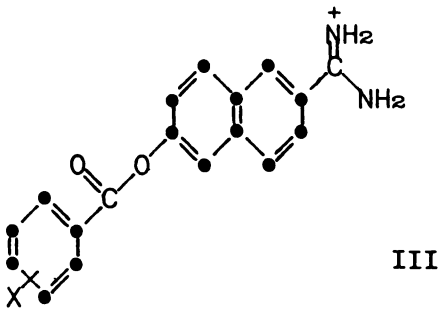
$$\log 1/K_i = -0.59(\pm0.49)MR_4 + 0.88(\pm0.52)\log(\beta \cdot 10^{MR_4} + 1) + 0.23(\pm0.07)\pi_3' - 0.74(\pm0.20)\sigma + 0.20(\pm0.30)I\text{-}M + 0.65(\pm0.22)I\text{-}1 + 0.43(\pm0.19)I\text{-}2 + 0.51(\pm0.15)I\text{-}3 + 1.38(\pm0.28) \qquad (19)$$
$$n = 104, \quad r = 0.924, \quad s = 0.222, \quad \text{optimum } MR_4 = 1.03$$

The parameters of Eq. 19 are consistent with those in all of the simpler equations showing the commonality of interaction with respect to σ, π_3', or MR_4 once other factors are accounted for. The squared correlation matrix for the parameters of Eq. 19 is as follows:

	σ	π_3'	MR_4	I-M	I-1	I-2	I-3
σ	1.000	0.024	0.175	0.164	0.010	0.001	0.063
π_3'		1.000	0.198	0.376	0.036	0.065	0.027
MR_4			1.000	0.527	0.048	0.069	0.147
I-M				1.000	0.091	0.131	0.072

I-1	1.000	0.004	0.007
I-2		1.000	0.009
I-3			1.000

Turning now to a quite different type of amidine (**III**) from data obtained by Aoyama *et al.* (23) in Table 4, we have derived Eqs. 20–22.



III

$$\log 1/K_i = -1.59(\pm 0.63)\sigma + 3.13(\pm 0.25) \quad (20)$$
$$n = 18, \quad r = 0.801, \quad s = 0.487, \quad F_{1,15} = 28.6$$

$$\log 1/K_i = -1.48(\pm 0.43)\sigma + 0.41(\pm 0.21)MR_4 + 2.64(\pm 0.28) \quad (21)$$
$$n = 18, \quad r = 0.922, \quad s = 0.326, \quad F_{1,14} = 20.7$$

Eqs. 20 and 21 are based only on *para*-derivatives III. In addition, substituents charged under experimental conditions [4—$OCH_2CH_2N(C_2H_5)_2$, 4—CH_2NH_2, and 4—$NHC(=NH)—NH_2$] were omitted.

Also, four poorly fit substituents were not included in deriving Eqs. 20 and 21: 4—$COOCH_3$, 4—$NHCOCH_3$, 4—$N(CH_3)_2$, and 4—$CH_2NHCOOCH_2C_6H_5$. There are no obvious explanations for these outliers.

Replacing MR_4 with π_4 in Eq. 21 yields a much poorer correlation ($r = 0.813$, $s = 0.489$), which indicates that the region where 4—*X* falls is not typically hydrophobic.

Analysis of the 3-substituents uncovered the fact that if *MR* (or π) for 3-substituents is set = 0, then 3- and 4-substituted congeners (with the exception of OCH_2O) can be included in a single equation (Eq. 22).

$$\log 1/K_i = -1.40(\pm 0.40)\sigma + 0.47(\pm 0.19)MR_4 + 2.59(\pm 0.24) \quad (22)$$
$$n = 21, \quad r = 0.915, \quad s = 0.322$$

A single 2-substituted congener was tested and it, too, was omitted. The parameters of Eq. 22 are essentially the same as those for Eq. 21, suggesting that 3-substituents do not contact the enzyme. The correlation between σ and *MR* is very low ($r^2 = 0.014$).

Molecular Graphics

The coordinates for the benzamidine and inhibitor-bovine trypsin complex have been determined by X-ray crystallography (15) and were obtained from the Brookhaven Protein Data Bank (24). Substituents on the benzamidine inhibitors were built using standard bond lengths and angles.

Coordinates were displayed and manipulated on an Evans and Sutherland Picture System 2 with the program MIDAS (25). Solvent-accessible surfaces for the protein were calculated with the program MS (26), and electrostatic surfaces were calculated with the program esp (27). As described previously, the benzamidine inhibitor is bound in a hydrophobic pocket in the active site and forms an electrostatic bond between the positively charged amidine and the negative carboxylate of Asp-189 at the back of the hydrophobic pocket. At the outer edge of the pocket is the catalytic Ser-195. The red surface represents hydrophobic area (carbon), the blue surface codes for polar space (nitrogen and oxygen), and yellow represents sulfur, α-carbon atoms, and C in C=0. In attempting to characterize the sulfur surface, we can make the following comparisons:

Compound	log *P*	Compound	log *P*
$CH_3CH_2CH_2CH_3$	2.89	CH_3CH_3	1.77
$C_2H_5(CH_2)C_2H_5$	3.39	CH_3SSCH_3	1.81
$C_2H_5(S)C_2H_5$	1.95	$CH_3CH_2CH_2CH_3$	2.89
$C_2H_5(O)C_2H_5$	0.89		
$C_2H_5(NH)C_2H_5$	0.58		

From any point of view a single sulfur atom in an aliphatic system is relatively hydrophilic compared to carbon. For example, the insertion of S into butane in diethylsulfide reduces log *P* from 2.89 to 1.95. Compared to pentane, the increase in hydrophilicity is even greater. In the case of the disulfide, if we compare ethane with dimethyl disulfide, the —S—S— linkage has little effect on log *P*. However, if we make the comparison of dimethyl disulfide with butane (the same number of non-hydrogen atoms), the —S—S— linkage is definitely hydrophilic. Although it is not nearly as polar as —S—, which is much more hydrophobic than —O— or —NH—, it does not behave like CH_2.

Turning now to the SH group, we can make the following comparison:

Compound	log *P*
$CH_3CH_2CH_2CH_2CH_3$	3.39
$CH_3CH_2CH_2CH_3$	2.89
$CH_3CH_2CH_2CH_2SH$	2.28
$CH_3CH_2CH_2CH_2OH$	0.88
$CH_3CH_2CH_2CH_2NH_2$	0.97

Comparing butanethiol to pentane, the former is definitely hydrophilic; however, it is in a different category from oxygen and nitrogen. Thus, it would seem that sulfur in its sulfide form cannot be called hydrophobic like carbon, nor can it be called hydrophilic like oxygen or nitrogen. For the present we suggest a single class (yellow) for the three forms of divalent sulfur found in proteins (—S—H, —S—, and —S—S—).

There is considerable evidence that, when methyl or methylene groups are placed next to electron-withdrawing groups, their hydrophobicity is greatly reduced (28). When electron-withdrawing groups are placed on both sides of a CH_2 group, it would seem likely that the CH_2 would be rendered nonhydrophobic. In order to test this idea, we compared log *P* for ICH_2I (2.30) to log *P* for I_2 (2.49) (16). The result suggests that CH_2 has in fact lost all of its hydrophobicity. Using *I* as the electronegative group has the advantage that hyperconjugative and enolization effects which might occur with COOR or CN are absent and only the inductive effect is present.

In accordance with the above findings we now distinguish three classes of surface: red (hydrophobic) formed by normal CH_3 and CH_2 groups, yellow (semihydrophilic) produced by divalent sulfur, α-carbons of amino acids, the carbon between the two nitrogens in histidine and C in C=0, and blue (hydrophilic) produced by oxygen and nitrogen. This is not to imply that there is not a continuum of surface from most hydrophobic to most hydrophilic which is the result of modulation (mostly of carbon) by nearby electronegative atoms. However, for the purpose of developing QSARs, it would be impractical at the present to make a finer categorization. To do so one would need atomic hydrophobic constants—something we have not been able to define. Even if such constants could be defined, to use them to characterize protein surfaces would require many times the number of data points which it is now found practical to work with.

As can be seen in Fig. 1, substituents in the 4-position on benzamidine are in close contact with the polar oxygen of Ser-195 (note the *blue surface*). The distance from a 4—Cl would be 2.85Å. The short distance between atom 1 of 4-substituents and the OH of Ser-195 is a source for steric effects and makes hydrogen bonding for 4—NH_2 or 4—OH a possibility. Planar groups such as NO_2, COO^-, or COOR would place an oxygen atom 1.6Å from the Ser OH. This steric effect could be mitigated by movement of the OH or rotation of the planar group out of the benzamidine plane or a combination of both.

The Ser OH is hydrogen bonded (note *dotted line*) to an adjacent

TABLE 3

Parameters used to derive equations 12–19 by X—$C_6H_4C(=N^+H_2)NH_2$

X	log 1/K_i		| Δlog 1/K_i |	MR_4	σ	π_3'	*I-M*	*I-1*	*I-2*	*I-3*
	Observed[a]	Calculated[b]								
1. H	1.46	1.41	0.05	0.10	0.00	0.00	0.00	0.00	0.00	0.00
2. 4—CH_3	1.52	1.38	0.15	0.57	−0.17	0.00	0.00	0.00	0.00	0.00
3. 4—C_2H_5	1.16	1.31	0.16	1.03	−0.15	0.00	0.00	0.00	0.00	0.00
4. 4—C_3H_7	1.42	1.34	0.08	1.50	−0.13	0.00	0.00	0.00	0.00	0.00
5. 4—C_4H_9	1.55	1.46	0.10	1.96	−0.16	0.00	0.00	0.00	0.00	0.00
6. 4—C_5H_{11}[c]	2.10	1.57	0.53	2.43	−0.15	0.00	0.00	0.00	0.00	0.00
7. 4—Cl	1.40	1.07	0.33	0.60	0.23	0.00	0.00	0.00	0.00	0.00
8. 4—Br	1.40	1.03	0.37	0.89	0.23	0.00	0.00	0.00	0.00	0.00
9. 4—NO_2	0.48	0.64	0.16	0.74	0.78	0.00	0.00	0.00	0.00	0.00
10. 4—OH	1.30	1.61	0.31	0.29	−0.37	0.00	0.00	0.00	0.00	0.00
11. 4—CHO	1.01	0.91	0.10	0.69	0.42	0.00	0.00	0.00	0.00	0.00
12. 4—CH_2Br	1.38	1.11	0.26	1.34	0.14	0.00	0.00	0.00	0.00	0.00
13. 4—$COOCH_3$	0.52	0.88	0.36	1.29	0.45	0.00	0.00	0.00	0.00	0.00
14. 4—$COOC_2H_5$	0.70	0.95	0.26	1.75	0.45	0.00	0.00	0.00	0.00	0.00
15. 4—$COOC_3H_7$	1.05	1.07	0.02	2.21	0.45	0.00	0.00	0.00	0.00	0.00
16. 4—COO—*i*—C_3H_7	0.92	1.07	0.15	2.22	0.45	0.00	0.00	0.00	0.00	0.00
17. 4—$COOC_8H_{17}$	1.80	1.73	0.07	4.55	0.45	0.00	0.00	0.00	0.00	0.00
18. 4—$COCH_3$	0.50	0.83	0.33	1.12	0.50	0.00	0.00	0.00	0.00	0.00
19. 4—$CONHCH_3$	0.85	0.97	0.11	1.46	0.36	0.00	0.00	0.00	0.00	0.00
20. 4—COC_6H_5	0.92	1.31	0.39	3.03	0.43	0.00	0.00	0.00	0.00	0.00
21. 4—$COCH_2OCH_3$	1.19	1.01	0.17	1.81	0.39	0.00	0.00	0.00	0.00	0.00
22. 4—OCH_3	1.50	1.41	0.08	0.79	−0.27	0.00	0.00	0.00	0.00	0.00
23. 4—OC_2H_5	1.00	1.39	0.39	1.25	−0.24	0.00	0.00	0.00	0.00	0.00
24. 4—OC_3H_7	1.05	1.47	0.42	1.71	−0.25	0.00	0.00	0.00	0.00	0.00
25. 4—OC_4H_9	1.42	1.63	0.21	2.17	−0.32	0.00	0.00	0.00	0.00	0.00
26. 4—OC_5H_{11}	1.70	1.77	0.07	2.63	−0.34	0.00	0.00	0.00	0.00	0.00
27. 4—OC_6H_{13}	1.77	1.90	0.13	3.09	−0.34	0.00	0.00	0.00	0.00	0.00
28. 4—OC_7H_{15}	2.00	2.03	0.03	3.56	−0.34	0.00	0.00	0.00	0.00	0.00
29. 4—OC_8H_{17}	2.22	2.16	0.06	4.02	−0.34	0.00	0.00	0.00	0.00	0.00
30. 4—OC_9H_{19}	2.40	2.30	0.10	4.49	−0.34	0.00	0.00	0.00	0.00	0.00
31. 4—$OC_{10}H_{21}$	2.30	2.43	0.13	4.95	−0.34	0.00	0.00	0.00	0.00	0.00
32. 4—$OC_{11}H_{23}$[c]	2.40	2.57	0.17	5.41	−0.34	0.00	0.00	0.00	0.00	0.00
33. 4—$OC_{12}H_{25}$[c]	2.16	2.70	0.55	5.88	−0.34	0.00	0.00	0.00	0.00	0.00
34. 4—$OCH_2C_6H_5$	2.04	1.85	0.19	3.23	−0.23	0.00	0.00	0.00	0.00	0.00
35. 4—$O(CH_2)_3C_6H_5$	2.17	2.12	0.05	4.16	−0.23	0.00	0.00	0.00	0.00	0.00
36. 4—$O(CH_2)_2C_6H_5$	1.85	1.98	0.13	3.68	−0.23	0.00	0.00	0.00	0.00	0.00
37. 4—$O(CH_2)_2OH$	1.32	1.40	0.08	1.43	−0.23	0.00	0.00	0.00	0.00	0.00
38. 4—$NHCOC_6H_5$	1.72	1.89	0.17	3.46	−0.19	0.00	0.00	0.00	0.00	0.00
39. 4—$NHCOC_6H_4$—4′—NO_2	2.22	2.07	0.15	4.10	−0.19	0.00	0.00	0.00	0.00	0.00
40. 4—$NHCOC_6H_4$—4′—NH_2	1.96	2.02	0.06	3.90	−0.19	0.00	0.00	0.00	0.00	0.00
41. 4—$CH_2OC_6H_5$	2.05	1.63	0.41	3.23	0.07	0.00	0.00	0.00	0.00	0.00
42. 4—$CH_2SC_6H_5$	1.82	1.90	0.08	3.89	−0.04	0.00	0.00	0.00	0.00	0.00
43. 4—CH_2COCH_3	1.40	1.19	0.21	1.50	0.07	0.00	0.00	0.00	0.00	0.00
44. 4—CH_2COCH_2Cl	1.40	1.31	0.09	2.06	0.07	0.00	0.00	0.00	0.00	0.00
45. 4—CH_2COCH_2OH	1.40	1.22	0.18	1.66	0.07	0.00	0.00	0.00	0.00	0.00
46. 4—$CH_2COC_6H_5$	1.92	1.70	0.22	3.47	0.07	0.00	0.00	0.00	0.00	0.00
47. 4—$CH_2COC_6H_4$—4′—CH_3	1.75	1.84	0.10	3.96	0.07	0.00	0.00	0.00	0.00	0.00
48. 4—$CH_2COC_6H_4$—4′—Cl	2.41	1.84	0.56	3.97	0.07	0.00	0.00	0.00	0.00	0.00
49. 4—$CH_2COC_6H_4$—4′—OH	1.48	1.76	0.27	3.66	0.07	0.00	0.00	0.00	0.00	0.00
50. 4—$CH_2COC_6H_4$—4′—OCH_3	2.05	1.91	0.15	4.18	0.07	0.00	0.00	0.00	0.00	0.00
51. 4—$CH_2COC_6H_4$—4′—OC_6H_5	2.92	2.47	0.45	6.14	0.07	0.00	0.00	0.00	0.00	0.00
52. 4—$CH_2COC_6H_4C_6H_5$	2.23	2.40	0.18	5.91	0.07	0.00	0.00	0.00	0.00	0.00
53. 4—$CH_2COCH_2OC_2H_5$	1.60	1.47	0.13	2.65	0.07	0.00	0.00	0.00	0.00	0.00
54. 4—CH=CH$COOCH_3$	1.70	1.40	0.30	2.29	0.03	0.00	0.00	0.00	0.00	0.00
55. 4—CH=CH$COOC_2H_5$	1.72	1.55	0.17	2.72	−0.01	0.00	0.00	0.00	0.00	0.00
56. 4—CH=CHCOC_6H_5	1.82	1.88	0.05	4.03	0.05	0.00	0.00	0.00	0.00	0.00
57. 4—CH=CHCOC_6H_4—4′—OCH_3	1.96	2.07	0.11	4.71	0.05	0.00	0.00	0.00	0.00	0.00
58. 4—CH=CHCOC_6H_4—3′—NO_2	1.72	2.06	0.34	4.66	0.05	0.00	0.00	0.00	0.00	0.00
59. 4—OCH_2COOCH_3	2.10	1.98	0.12	1.97	−0.18	0.00	0.00	0.00	0.00	1.00
60. 4—$OCH_2COOC_2H_5$	2.19	2.10	0.09	2.43	−0.18	0.00	0.00	0.00	0.00	1.00
61. 4—$OCH_2COC_6H_5$	2.30	2.46	0.16	3.72	−0.18	0.00	0.00	0.00	0.00	1.00
62. 4—$(CH_2)_2COOCH_3$	1.92	1.93	0.01	2.11	−0.07	0.00	0.00	0.00	0.00	1.00
63. 4—$(CH_2)_2COOC_2H_5$	1.85	2.09	0.23	2.69	−0.07	0.00	0.00	0.00	0.00	1.00
64. 4—$(CH_2)_2COCH_3$	2.38	1.92	0.46	2.05	−0.07	0.00	0.00	0.00	0.00	1.00
65. 4—$(CH_2)_2COC_6H_5$	2.57	2.47	0.10	4.02	−0.07	0.00	0.00	0.00	0.00	1.00
66. 4—$(CH_2)_2COC_6H_4$—4′—CH_3	2.52	2.60	0.08	4.49	−0.07	0.00	0.00	0.00	0.00	1.00
67. 4—$(CH_2)_2COC_6H_4$—4′—OCH_3	2.72	2.67	0.05	4.71	−0.07	0.00	0.00	0.00	0.00	1.00
68. 4—$(CH_2)_2COC_6H_4$-4′—Cl	2.46	2.61	0.16	4.52	−0.07	0.00	0.00	0.00	0.00	1.00

TABLE 3—*Continued*

X	log $1/K_i$ Observed[a]	log $1/K_i$ Calculated[b]	$\lvert\Delta \log 1/K_i\rvert$	MR_4	σ	π_3'	*I-M*	*I-1*	*I-2*	*I-3*
69. 4—$(CH_2)_2COC_6H_4$—4′—C_6H_5	2.82	3.17	0.35	6.46	−0.07	0.00	0.00	0.00	0.00	1.00
70. 4—$(CH_2)_2COC_6H_4$—4′—OC_6H_5	3.42	3.24	0.18	6.69	−0.07	0.00	0.00	0.00	0.00	1.00
71. 4-$SO_2NHC_6H_5$[c]	0.70	1.36	0.66	3.78	0.65	0.00	0.00	0.00	0.00	0.00
72. 3—CH_3	1.50	1.79	0.30	0.10	−0.07	0.60	1.00	0.00	0.00	0.00
73. 3—Cl	1.38	1.46	0.08	0.10	0.37	0.56	1.00	0.00	0.00	0.00
74. 3—NO_2	0.89	1.00	0.11	0.10	0.71	−0.36	1.00	0.00	0.00	0.00
75. 3—OH	1.48	1.55	0.07	0.10	0.12	0.15	1.00	0.00	0.00	0.00
76. 3—$COOCH_3$	1.06	1.36	0.31	0.10	0.37	0.14	1.00	0.00	0.00	0.00
77. 3—$COOC_2H_5$	1.19	1.44	0.26	0.10	0.37	0.50	1.00	0.00	0.00	0.00
78. 3—COO—*i*—C_3H_7[c]	1.00	1.55	0.55	0.10	0.37	0.97	1.00	0.00	0.00	0.00
79. 3—$COOCH_2C_6H_5$	1.52	1.76	0.24	0.10	0.37	1.90	1.00	0.00	0.00	0.00
80. 3—$COCH_3$	1.62	1.23	0.40	0.10	0.38	−0.43	1.00	0.00	0.00	0.00
81. 3—COC_6H_5	1.50	1.63	0.14	0.10	0.34	1.23	1.00	0.00	0.00	0.00
82. 3—$CONH_2$	1.38	1.15	0.23	0.10	0.28	−1.08	1.00	0.00	0.00	0.00
83. 3—$CONHCH_3$	0.82	1.15	0.33	0.10	0.35	−0.85	1.00	0.00	0.00	0.00
84. 3—$CONHCH_2C_6H_5$	1.82	1.55	0.27	0.10	0.35	0.92	1.00	0.00	0.00	0.00
85. 3—$CONHC_6H_5$	1.63	1.65	0.01	0.10	0.23	0.94	1.00	0.00	0.00	0.00
86. 3—COC_6H_4—4′—CH_3	2.00	1.78	0.22	0.10	0.34	1.88	1.00	0.00	0.00	0.00
87. 3—COC_6H_4—4′—OCH_3	2.00	1.68	0.32	0.10	0.34	1.43	1.00	0.00	0.00	0.00
88. 3—COC_6H_4—4′—Cl	1.89	1.80	0.08	0.10	0.34	1.99	1.00	0.00	0.00	0.00
89. 3—OCH_3	1.60	1.59	0.02	0.10	0.12	0.31	1.00	0.00	0.00	0.00
90. 3—OC_2H_5	1.57	1.70	0.13	0.10	0.10	0.75	1.00	0.00	0.00	0.00
91. 3—OC_4H_9	1.92	1.94	0.02	0.10	0.10	1.81	1.00	0.00	0.00	0.00
92. 3—OC_5H_{11}	1.96	2.06	0.10	0.10	0.10	2.34	1.00	0.00	0.00	0.00
93. 3—OC_6H_{13}	2.10	2.18	0.08	0.10	0.10	2.87	1.00	0.00	0.00	0.00
94. 3—OC_7H_{15}	2.19	2.30	0.12	0.10	0.10	3.40	1.00	0.00	0.00	0.00
95. 3—$O(CH_2)_2OH$	1.80	1.38	0.41	0.10	0.10	−0.65	1.00	0.00	0.00	0.00
96. 3—$CH_2SC_6H_5$	2.31	2.07	0.24	0.10	0.08	2.33	1.00	0.00	0.00	0.00
97. 3—$CH_2OC_6H_5$	2.05	1.94	0.10	0.10	0.06	1.69	1.00	0.00	0.00	0.00
98. 3—$NHCOC_6H_5$[c]	1.32	1.85	0.53	0.10	0.02	1.16	1.00	0.00	0.00	0.00
99. 3—CH_2COOCH_3	1.89	1.86	0.03	0.10	0.13	−0.32	1.00	0.00	1.00	0.00
100. 3—$CH_2COOC_2H_5$	1.85	1.98	0.13	0.10	0.13	0.21	1.00	0.00	1.00	0.00
101. 3—$CH_2COC_6H_5$	2.33	2.26	0.07	0.10	0.07	1.21	1.00	0.00	1.00	0.00
102. 3—$CH_2COC_6H_4$—4′—CH_3	2.35	2.40	0.06	0.10	0.07	1.86	1.00	0.00	1.00	0.00
103. 3—$CH_2COC_6H_4$—4′—OCH_3	2.37	2.31	0.06	0.10	0.07	1.43	1.00	0.00	1.00	0.00
104. 3—$CH_2COC_6H_4$—4′—Cl	2.36	2.43	0.08	0.10	0.07	2.00	1.00	0.00	1.00	0.00
105. 3—$CH_2COOCH_2C_6H_5$	2.35	2.24	0.11	0.10	0.13	1.34	1.00	1.00	1.00	0.00
106. 3—CH=CH$COOCH_3$	2.22	2.21	0.01	0.10	0.19	0.43	1.00	1.00	0.00	0.00
107. 3—CH=CH$COOC_2H_5$	2.26	2.33	0.07	0.10	0.19	0.96	1.00	1.00	0.00	0.00
108. 3—CH=CHCOC_6H_5	2.22	2.34	0.12	0.10	0.18	0.95	1.00	1.00	0.00	0.00
109. 3—CH=CHCOC_6H_4—3′—NO_2	2.19	2.28	0.09	0.10	0.15	0.59	1.00	1.00	0.00	0.00
110. 3—CH=CH$COOCH_2C_6H_5$	2.89	2.62	0.27	0.10	0.19	2.20	1.00	1.00	0.00	0.00

[a] From Ref. 12.
[b] Calculated by Eq. 19.
[c] These data points were omitted in deriving Eq. 12–19.

histidine which tends to hold it in place. Thus, we see that the steric effect of 4-substituents is the result of their unfavorable contact with the polar surface of the OH of Ser-195. However, breaking the hydrogen bond between the OH of Ser and histidine would allow rotation of the CH_2OH to relieve much of the steric effect.

The possible effect of electronic repulsion of substituents should be noted. For example, the repulsive force between oxygen of 4—OCH_3 and the Ser-195 oxygen could account for the fact that 4—OCH_3 is less active than 4—CH_3. However, since 4—OCH_3 is reasonably well fit in Tables 1–3, most of this repulsion must be accounted for by the steric terms B_4 or *MR*.

Fig. 2 looks directly into the hydrophobic hole into which the yellow wire model of the benzamidine is found. In the *meta*-position are placed two $O(CH_2)_5CH_3$ groups, one in green and the other in violet. The *outlines* of Ser-195 and Trp-215 are also shown. There is some ambiguity about how the *meta*-substituents may bind, but we favor the position shown by the *green* moiety on the *right* in Fig. 2. There appears to be more typical (*red*) hydrophobic space on the *right side*. The large tryptophane surface provides hydrophobic space. The coefficient of about 0.20 with π_3' in Eqs. 11 and 19 suggests that only partial desolvation of 3—*X* is occurring on the average. The complexity of binding is best appreciated by comparing the log $1/K_i$ values of 3—OR from C_1 to C_7 where it is seen that the increase per CH_2 is not that associated with 0.5π normally found for alkyl groups binding to a relatively smooth and uniform hydrophobic surface (1, 22). The surface onto which 3-substituents fall, whether it be the right or left side, is uneven, so that subtle steric effects beyond our level of resolution at present may account for the low coefficient with the π_3' term.

Small 3-substituents could interact with the side-chain of Gln-192 and partially with the polar amide moiety of Gln. A 3—NH_2, for example, would be 2.9Å from the amide oxygen of Gln-192.

Fig. 3 shows congener II with a 4—$OCH_2C_6H_5$ in *green* as a wire model. A *dotted yellow line* from the carbonyl group of the ester linkage indicates hydrogen bonding with the OH of Ser-195. The distance between the two groups is 2.7Å. The phenyl portion of the ester is slightly twisted out of coplanarity with the naphthyl ring. This allows the phenyl ring to make better hydrophobic contact with the enzyme surface. However, this also increases the distance for hydrogen bonding between the carbonyl oxygen and the H of Ser-195. Given the uncertainty in the X-ray coordinates and the possibility of surface changes when such large inhibitors bind, it does not seem profitable to speculate on what the optimum positioning of these would be.

The best way to rationalize the negative coefficient with the σ term in Eqs. 20–22 is to assume that it reflects hydrogen bonding between

TABLE 4

Parameters used to derive Eqs. 20–22

X	log $1/IC_{50}$		$\lvert \Delta \log 1/IC_{50} \rvert$	σ	MR_4
	Observed[a]	Calculated[b]			
1. H	2.40	2.64	0.24	0.00	0.10
2. 4—CH_3	3.00	3.10	0.10	−0.17	0.56
3. 4—*t*—C_4H_9	3.52	3.80	0.28	−0.20	1.96
4. 4—OCH_3	3.30	3.34	0.04	−0.27	0.79
5. 4—OC_4H_9	4.40	4.07	0.33	−0.32	2.17
6. 4—$OCH_2C_6H_5$	4.40	4.44	0.04	−0.23	3.22
7. 4—OH	3.52	3.24	0.28	−0.37	0.28
8. 4—NH_2	4.22	3.77	0.45	−0.66	0.54
9. 4—$N(CH_3)_2$[c]	3.52	4.49	0.97	−0.83	1.56
10. 4—$OCOCH_3$	3.40	2.75	0.65	0.31	1.25
11. 4—SCH_3	2.70	3.25	0.55	0.00	1.38
12. 4—F	2.52	2.55	0.03	0.06	0.09
13. 4—Cl	2.52	2.55	0.03	0.23	0.60
14. 4—Br	2.40	2.69	0.29	0.23	0.89
15. 4—$NHCOCH_3$[c]	2.52	3.30	0.78	0.00	1.49
16. 4—$COOCH_3$[c]	3.52	2.57	0.95	0.45	1.29
17. 4—$COCH_3$	2.70	2.42	0.28	0.50	1.12
18. 4—$CH{=}NOCH_3$	3.05	2.88	0.17	0.30	1.49
19. 4—SO_2NH_2	2.00	2.37	0.37	0.57	1.23
20. 4—CN	2.00	1.96	0.04	0.66	0.63
21. 4—NO_2	2.00	1.85	0.15	0.78	0.74
22. 4—$CH_2NHCOOCH_2C_6H_5$[c]	3.22	4.81	1.59	−0.05	4.53
23. 3—CH_3	2.70	2.69	0.01	−0.07	0.00
24. 3—CF_3	2.10	1.99	0.11	0.43	0.00
25. 3,4—di—CH_3	2.70	3.19	0.49	−0.24	0.56
26. 3,4—OCH_2O[c]	3.70	3.02	0.68	−0.16	0.44

[a] From Ref. 23.
[b] Calculated by Eq. 22.
[c] These data points were omitted in deriving Eq. 20–22.

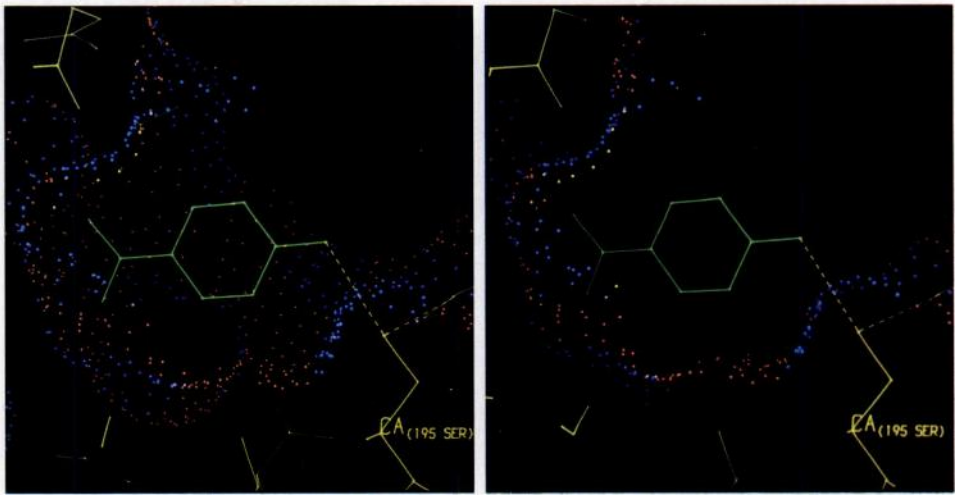


Fig. 1. Side view of the benzamidine binding in the hydrophobic pocket with the amidine group facing the carboxylate of Asp-189. The substituent in the *para*-position is the size of Cl, and the *dotted line* is to the oxygen of Ser-195. The other *dotted line* (to the *right*) shows hydrogen bonding between Ser and His. The *red* surface is hydrophobic, *blue* is polar, and *yellow* is intermediate.

the ester carbonyl oxygen and the H of the Ser-195 OH group. The more electron releasing the substituent, the higher the electron density on the carbonyl oxygen and, hence, the tighter the hydrogen bond.

A most interesting feature in Fig. 3 is that substituents in the 4-position of the ester moiety fall on a disulfide bridge between cysteine 42 and 58. The majority of the substituents in Table 4 are small and, hence, would largely be in contact with sulfur. It is therefore of particular interest that it is *MR* rather than π which best correlates the data. Phenomenologically, sulfur may turn out to behave more as hydrophilic than as hydrophobic space.

From a study of the graphics it is clear that 3-substituents cannot contact the trypsin surface. Even as it now stands in Fig. 3, there is not enough room for a 3-substituent to fit between the left side of the ring and the surface. It is likely that the ring is more perpendicular than we have shown it (to accommodate hydrogen bonding and to maintain coplanarity between the ring systems and their ester linkage), which would more effectively preclude contact of 3-substituents with the surface. The only effect of 3-substituents is the electronic effect on the carbonyl oxygen engaged in the hydrogen bonding with the Ser OH.

The most active substituents in Table 4 are 4—$OCH_2C_6H_5$ and 4—OC_4H_9. We have not considered the highly charged 4—$NHC(=NH)NH_2$ which is slightly more potent. As seen in Fig. 3, the phenyl moiety of 4—$OCH_2C_6H_5$ falls on the junction of *red*, *yellow*, and *blue* space. Hence one would expect it to be better correlated by *MR*

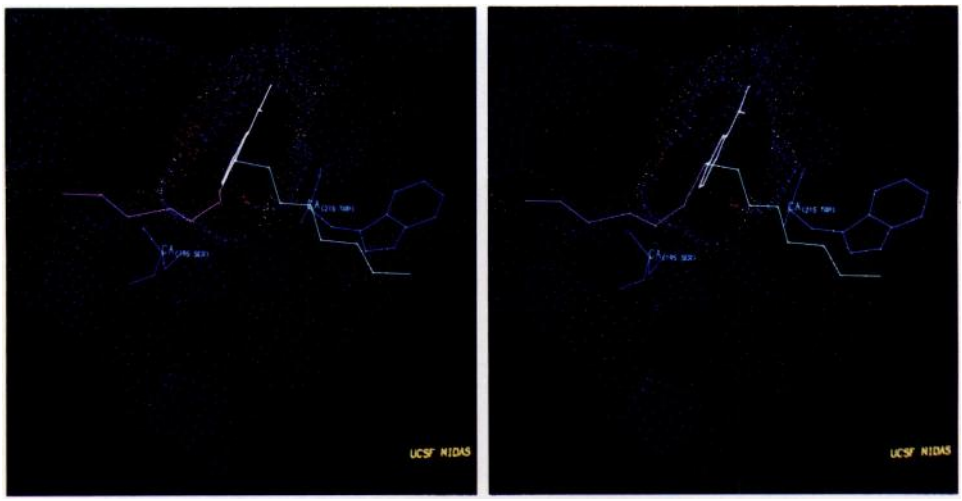


Fig. 2. End-on view of benzamidine in the hydrophobic pocket with $O(CH_2)_5CH_3$ in the 3- and 5-positions. On the *right* in *blue* one sees the skeleton of a tryptophane residue which creates a large hydrophobic surface.

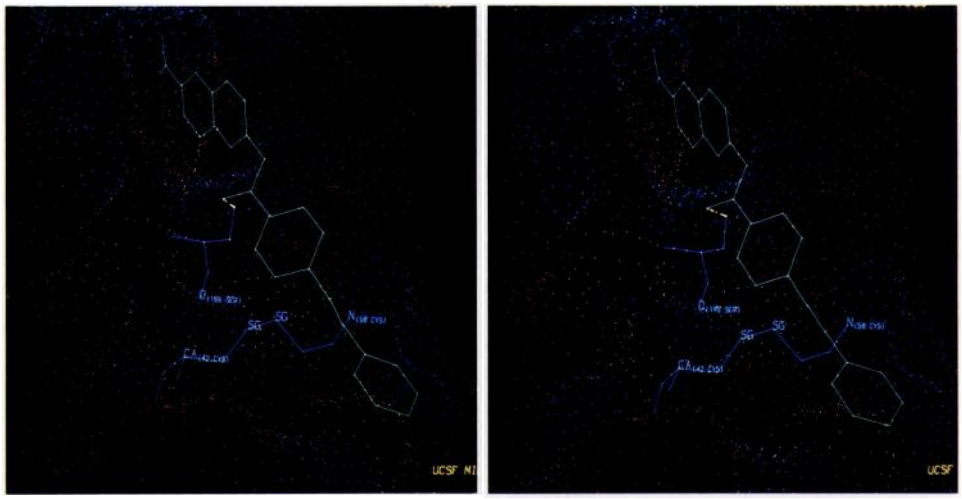


Fig. 3. Example of chemical structure II with an $OCH_2C_6H_5$ moiety in the *para*-position. Note the large amount of *yellow* surface created by the disulfide bridge (*SG-SG*).

than by π. The 4—$OCH_2C_6H_5$ congener is well fit using *MR*, whereas the 4—OC_4H_9 is more active than expected by slightly more than 1 SD. The third most active molecule is the 4—NH_2 derivative, the activity of which derives almost entirely from its electronic effect. One would expect 4—$N(CH_3)_2$ to be more active than 4—NH_2 but, in fact, we have had to omit it. If the rather wide $N(CH_3)_2$ group were twisted somewhat out of plane with the phenyl ring so as to inhibit its resonance, this could account for its lower than expected potency.

Discussion

A satisfying theme of unity throughout the QSARs for the three different sets of data for benzamidines from three different laboratories is the consistency of the coefficient with the σ term. This effect was overlooked (Eq. 2) by Andrews *et al.* (11) and was underestimated by Labes and Hagen (12) (Eq. 3). Clearly, electron release by substituents increases the inhibitory potency of the benzamidines. However, the role of electron release by substituents is not completely obvious. We agree with Mares-Guia *et al.* (10) that electron release by *X* would not significantly facilitate protonation of the amidine moiety and in this way promote binding. Since 3- and 4-substituents show the same electronic effect, it seems unlikely that *q*, as calculated by Mares-Guia *et al.* (10), has any specific relationship (such as hydrogen bonding) to the interactions of 4-substituents with Ser-195. In fact, *q* and σ are almost perfectly collinear ($r^2 = 0.993$).

Since calculation of the electrostatic surface potentials of the pocket into which the benzene ring fits shows that most of the surface bears a low negative charge, a charge transfer interaction seems to be out of question. Mares-Guia *et al.* (10) also reject the idea of a charge-transfer interaction.

One idea which has not been explored is the effect of substituents on the hydrophobicity of the benzene ring. It has been known for some time that strong electron withdrawal by electron-attracting groups can, in some instances, decrease hydrophobicity of neighboring parts of a molecule (Ref. 16, pp. 37–43). It has been noted that, in compounds of the type $RN^+(CH_3)_3$, the first several CH_2 groups of R do not show normal hydrophobicity (16). The hydrophobic effect of a CH_2 is a geometric function of its distance from the positive charge on the nitrogen atom. Hence, it is conceivable that reduction in hydrophobicity by the strong electron-attracting amidine group could be restored in part by strong electron-releasing groups in the *para*-position of the benzene ring. This point is worthy of further study.

One might expect that substituents which could hydrogen bond by proton donation to the OH of Ser-195 or oxygen of Gln-192 amide might show unexpectedly high activity. The best suited 4—OH is the second most active congener, but 4—NH_2, which is a less effective hydrogen bond donor, is the most potent congener in Table 1. At the 3-position (Table 2), an alkyl group is as active as the 3—NH_2. The activity of all these compounds is exactly what one would expect from their small size and their strong electron-releasing ability. The well predicted strongly electron-attracting NO_2 group produces the least active congener in both Tables 1 and 2 despite its small size.

One of the surprises in the development of QSARs during the past 20 years has been the paucity of examples in which an explicit role could be found for hydrogen bonding. The example covered by Eq. 22 is a notable exception which is understandable only because we have the X-ray crystallography and the molecular graphics. Most impressive examples which fail to show a role for hydrogen bonding are studies of the binding of miscellaneous compounds (hydrogen bonders and nonhydrogen bonders) to bovine serum albumin (29). The often *apparent* unimportance of hydrogen bonding in enzyme-ligand interactions can be accounted for in three ways. If we did not have crystallographic data it is unlikely that we would have associated the negative ρ value of Eq. 22 with hydrogen bonding. Many more examples such as Eq. 22 may exist. Very weak, nonspecific hydrogen bonding may simply be accounted for in the octanol-water π values. Finally, in the partitioning of ligands from an aqueous phase onto or into a macromolecule, substituents move from one kind of hydrogen bonding in the aqueous phase to a second with the macromolecule. The free energy change may sometimes be too small to detect. Fersht *et al.* (30) have recently shown via protein engineering techniques that specific hydrogen bonding between uncharged parts of ligands and proteins are weak—0.5–1.5 kcal/mol.

In the case of the OH of Ser-195, the OH or NH_2 groups of the benzamidine would exchange binding to water molecules for binding with the Ser OH. The overall free energy change may be too small for us to detect with our present methods. In the case of the naphthyl benzoates of Eq. 22, the geometry for hydrogen bonding may be so ideal that we can detect the effect.

Clearly, small *para*-substituents on the benzamidines do not contact hydrophobic space but, instead, collide with the polar OH of Ser-195. When π or a combination of π and *MR* do not correlate substituent effects while *MR* or steric parameters do, we have assumed this to be diagnostic of interaction occurring with polar enzymic space. Thus, small *para*-substituents are in line with this expectation as graphics allows us to visualize.

In the case of cysteine hydrolases acting on a set of esters (1, 2, 22), a situation similar to that of trypsin occurs in which an *MR* term accounts for *para*-substituents colliding head-on with either an amide moiety of Gln (papain) or an amino group of Lys (actinidin). In these instances, it was assumed that the positive coefficient with *MR* represents a buttressing effect of the flexible surface side-chains. This positive steric effect is not seen in the case of trypsin, and it seems likely that this may be due to a stiffening of the Ser—CH_2OH moiety by hydrogen bonding to the nearby His—57. This resistance gives way with groups having $MR > 1$ (e.g., above C_2H_5). In Table 3 the difference between log $1/K_i$ for 4—H and 4—C_2H_5 is 0.30. The steric effect of ethyl is offset by its electronic effect ($-0.75 \times -0.15 = 0.11$), so that the steric difference would be 0.41. This would correspond to a free energy difference of 0.56 kcal/mol needed to break the hydrogen bond, which is in the lower range found by Fersht *et al.* (30) for the energy of hydrogen bonding between uncharged groups in proteins.

Groups larger than ethyl in the *para*-position have a positive interaction characterized by a slope of 0.29 (Eq. 19), which in fact, is close to the slope of 0.23 with the π_3' term. Since π' and MR_4 for the 4-substituents longer than ethyl are very highly collinear, one cannot discern from the correlation equations which is the significant parameter.

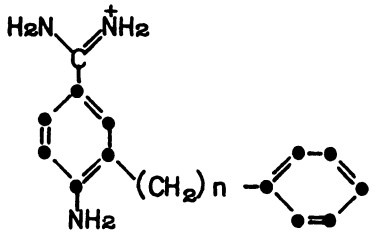
In conclusion, Eq. 19 rationalizes the discordant views of Eqs. 1–3 and shows that the data from three different sets of benzamidines inhibiting trypsin behave in a very consistent manner.

Furthermore, with Eq. 22, other features of trypsin-ligand interactions are characterized which provide a broader and more refined view of the region around the enzyme's active site. It is clear that both QSARs and X-ray crystallographic information, brought into focus via color stereo graphics, have a synergistic effect in the study of ligand-receptor interactions. The qualitative picture of the receptor obtained from the crystallography via graphics is invaluable in defining the geometry of the receptor and the nature of its surfaces. However, this tells us nothing about the electronic effect of substituents, nor does it tell us anything about the quantitative character of steric and hydrophobic interactions.

A most important point which several such QSAR-graphics analyses have now brought out is that, although there is undoubtedly flexibility in the structure of enzymes, the features of the active site obtained via QSARs from data obtained from studies *in solution* agree very well with the picture of the enzyme obtained from the crystallized form of the enzyme by X-ray crystallography.

The QSARs do bring out flexibility of ligand-enzyme interaction. Steric effects are modeled by the continuous variables *MR*, E_s, etc., which shows that contact between enzyme and ligand does not destroy activity completely but simply reduces it in proportion to substituent size. Probably both ligand and enzyme are gradually moved from their preferred positions.

Finally, the results of our analysis do suggest ways to make more potent benzamidine inhibitors of trypsin by taking advantage of the importance of electron release from the 4-position and hydrophobic binding where *meta*-substituents fall. Variations in **III** should produce log $1/K_i > 6$ (on the basis of results in Table 1 or 2).



III

Making the *n* in III, 3 or 4 should allow the phenyl of III to make contact with the large hydrophobic tryptophane surface.

References

1. Mares-Guia, M., and E. Shaw. Studies on the active center of trypsin. *J. Biol. Chem.* **240**:1579–1585 (1965).
2. Carotti, A., R. N. Smith, S. Wong, C. Hansch, J. M. Blaney, and R. Langridge. Papain hydrolysis of *X*-phenyl-*N*-methanesulfonyl glycinates: a quantitative

structure activity relationship and molecular graphics analysis. *Arch. Biochem. Biophys.* **229**:112–115 (1984).
3. Smith, R. N., C. Hansch, K. H. Kim, B. Omiya, G. Fukumura, C. D. Selassie, P. Y. C. Jow, J. M. Blaney, and R. Langridge. The use of crystallography, graphics and quantitative structure-activity relationships in the analysis of the papain hydrolysis of *X*-phenyl hippurates. *Arch. Biochem. Biophys.* **215**:319–328 (1982).
4. Blaney, J. M., C. Hansch, C. Silipo, and A. Vittoria. Structure-activity relationships of dihydrofolate reductase inhibitors. *Chem. Rev.* **84**:333–407 (1984).
5. Hansch, C., and J. M. Blaney. The new look to QSAR, in *Drug Design: Fact or Fantasy*? (K. H. R. Wooldridge, ed.). Academic Press, New York, 185–208 (1984).
6. Hansch, C., B. A. Hathaway, Z. R. Guo, C. D. Selassie, S. W. Dietrich, J. M. Blaney, R. Langridge, K. W. Volz, and B. T. Kaufman. Crystallography, quantitative structure-activity relationships, and molecular graphics in a comparative analysis of the inhibition of dihydrofolate reductase from chicken liver and *Lactobacillus casei* by 4,6-diamino-1,2-dihydro-2,2-dimethyl-1-(substituted-phenyl)-*s*-triazines. *J. Med. Chem.* **27**:129–143, (1984).
7. Baker, B. R., and E. H. Erickson. Irreversible enzyme inhibitors. CVI. Proteolytic enzymes. I. Bulk tolerances in trypsin-inhibitor complexes. *J. Med. Chem.* **10**:1123–1128 (1967).
8. Baker, B. R. *Design of Active-Site-Directed Irreversible Enzyme Inhibitors.* John Wiley & Sons, New York, 59 (1967).
9. Storzebecher, J., F. Markwardt, P. Richter, B. Voight, P. Walsmann, and G. Wagner. Synthetische Inhibitoren der Serinproteinasen. *Pharmazie* **31**:458–461 (1976).
10. Mares-Guia, M., D. L. Nelson, and E. J. Rogana. Electronic effects in the interaction of *para*-substituted benzamidines with trypsin: the involvement of the π-electronic density at the central atom of the substituent in binding. *J. Am. Chem. Soc.* **99**:2331–2336 (1977).
11. Andrews, J. M., D. P. Roman, Jr., D. H. Bing, and M. Cory. Inhibition of four human serine proteases by substituted benzamidines. *J. Med. Chem.* **21**:1202–1207 (1978).
12. Labes, D., and V. Hagen. Hansch-Analyse der Hemmwirkung von 3- und 4-substituierten Benzamidinen gegenüber Thrombin, Plasmin und Trypsin. *Pharmazie* **34**:649–653 (1979).
13. Charton, M. The application of the Hammett equation to amidines. *J. Org. Chem.* **30**:969–973.
14. Piskov, V. B., and V. P. Kasperovich. Synthesis and some physicochemical characteristics of benzamidines and their analogs. *J. Org. Chem. USSR* **14**:820–834 (1978).
15. Bode, W., and P. Schwager. The refined crystal structure of bovine β-trypsin at 1.8 Å resolution. *J. Mol. Biol.* **98**:693–717 (1975).
16. Hansch, C., and A. Leo. *Substituent Constants for Correlation Analysis in Chemistry and Biology.* Wiley-Interscience, New York (1979).
17. Leo, A. The octanol-water partition coefficient of aromatic solutes: the effect of electronic interactions, alkyl chains, hydrogen bonds, and *ortho*-substitution. *J. Chem. Soc. Perkin Trans. II* 825–838 (1983).
18. Fujita, T., J. Iwasa and C. Hansch. A new substituent constant, π, derived from partition coefficients. *J. Am. Chem. Soc.* **86**:5175–5178 (1964).
19. Fujita, T. Analysis and prediction of partition coefficients of *meta*- and *para*-disubstituted benzenes in terms of substituent effects. *J. Pharm. Sci.* **72**:285–289 (1983).
20. Hoefnagel, A. J., M. A. Hoefnagel, and B. M. Wepster. Substituent effects. 6. Charged groups: a simple extension of the Hammett equation. *J. Org. Chem.* **43**:4720–4745 (1978).
21. Verloop, A., W. Hoogenstraaten, and J. Tipker. Development and application of new steric substituent parameters in drug design, in *Drug Design* (E. J. Ariens, ed.), Vol. VIII. Academic Press, New York, 165–207 (1976).
22. Carotti, A., C. Ragueseo, and C. Hansch. Quantitative structure-activity relationships of cysteine hydrolases, ficin, hydrolysis of *X*-phenyl-*N*-methanesulfonyl glycinates. *Chem.-Biol. Interact.* **52**:279–288 (1985).
23. Aoyama, T., T. Okutome, T. Nakayame, T. Yaegashi, R. Matsui, S. Nunomora, M. Kurumi, Y. Sakurai, and S. Fujii. Synthesis and structure-activity of protease inhibitors IV. Amidinonaphthols and related acyl derivatives. *Chem. Pharm. Bull. (Tokyo)* **33**:1458–1471 (1985).
24. Brookhaven Protein Data Bank File: ICAC (May 22, 1976).
25. Langridge, R., T. E. Ferrin, I. D. Kuntz, and M. L. Connolly. Real-time color graphics in studies of molecular interactions. *Science (Wash. D. C.)* **211**:661–666 (1981).
26. Bash, P. A., N. Pattabiraman, C. Huang, T. E. Ferrin, and R. Langridge. Van der Waals surfaces in molecular modeling: implementation with real-time computer graphics. *Science (Wash. D. C.)* **222**:1325–1327 (1983).
27. Weiner, P. K., R. Langridge, J. M. Blaney, R. Shaeffer, and P. Kollman. Eectrostatic potential molecular surfaces. *Proc. Natl. Acad. Sci. USA* **78**:3754–3758 (1982).
28. Hansch, C., A. Leo, and D. Nikaitani. On the additive-constitutive character of partition coefficients. *J. Org. Chem.* **37**:3090–3092 (1972).
29. Helmer, F., K. Kiehs, and C. Hansch. The linear free-energy relationship between partition coefficients and the binding and conformational perturbation of macromolecules by small organic compounds. *Biochemistry* **7**:2858–2863 (1968).
30. Fersht, A. R., J. P. Shi, J. Knill-Jones, D. M. Lowe, A. J. Wilkinson, D. M. Blow, P. Brick, P. Carter, M. M. Y. Waye, and G. Winter. Hydrogen bonding and biological specificity analyzed by protein engineering. *Nature (Lond.)* **314**:235–238 (1985).

Send reprint requests to: Dr. Corwin Hansch, Department of Chemistry, Pomona College, Claremont, CA 91711.