The above experiment was repeated using 4.8 g (0.03 mol) of bromine in 20 ml of carbon tetrachloride. Work-up gave 0.4 g of a brown gum which, after filtration through an alumina column, using chloroform as eluent, gave 0.3 g of a yellow oil which had the same retention time on glc analysis as 4,5-dibromo-*o*-xylene. Its ir spectrum was superimposable on that of authentic 4,5-dibromo-*o*-xylene.

Reaction of Thallium(I) Bromide with Bromine in Acetonitrile. Preparation of Thallium(III) Bromide.—A suspension of 0.28 g (0.001 mol) of thallium(I) bromide in 20 ml of acetonitrile was stirred vigorously while 0.16 g (0.001 mol) of bromine in 10 ml of acetonitrile was added. The bromine was decolorized, and the solid dissolved to give a solution of thallium(III) bromide. Dilution of an aliquot of this solution (0.1 ml made up to 10 ml) gave a solution whose uv spectrum showed an intense absorption band at 275 nm, with a shoulder at 300 nm.

Registry No.—Bromine, 7726-95-6; thallium(III) acetate, 2570-63-0; phenylthallium diacetate, 20425-82-5; 4-*o*-xylylthallium diacetate, 31947-39-4; thallium(III) bromide, 13701-90-1.

Structure–Activity Relationship in the Chymotrypsin Hydrolysis of *p*-Nitrophenyl Esters[1]

CORWIN HANSCH

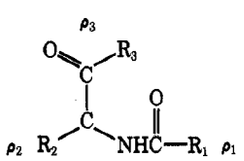
Department of Chemistry, Pomona College, Claremont, California 91711

Received February 26, 1971

The structure–activity relationship in chymotrypsin substrates is examined using the substituent constants σ^*, E_s, and π for the evaluation of electronic, steric, and hydrophobic effects on the relative rates of reaction. It is found that hydrophobic forces (defined by π) play a positive role in the deacylation step.

We have been interested in studying substituent effects (hydrophobic, electronic, and steric) on enzyme substrate interactions.[2] The present paper analyzes substituent effects on chymotrypsin hydrolysis from the work of Dupaix, Béchet, and Roucous[3] and compares this with earlier studies. The structure–activity problem with chymotrypsin has been approached from many points of view.[4,5] In this report our primary purpose is to consider the role of hydrophobic forces in the hydrolysis step *via* extrathermodynamic correlations. There appears to be a role for these forces independent of specific steric and electronic effects of substituents.

In a recent study[6] of chymotrypsin substrates and inhibitors, the Hein–Niemann[7,8] model of the active site was employed in an analysis of the structure–activity relationship. This model pictures four sections in space into which the four substituents attached to the α carbon of an amino acid moiety of a protein or peptide would fit. This is depicted as in I. In I the hydrogen



I

atom on the α carbon is not shown. It is in the space below the plane of the page. The ρ_H region into which the hydrogen fits is assumed to be occupied only by solvent. The ρ_1, ρ_2, and ρ_3 regions have quite different binding characteristics for substrates and inhibitors. We[6] have attempted a characterization of these areas using linear modeling techniques[9] employing four types of physicochemical parameters: hydrophobic ($\log P$, π), electronic (σ), steric (E_s), and polarizability (P_E). Our general model[10] is formulated in eq 1. In eq 1, k may

$$\log k = k_1\pi + k_2\sigma + k_3E_s + k_4 \quad (1)$$

be a rate or equilibrium constant and the disposable parameters, k_1–k_4, are evaluated *via* the method of least squares.[9] The parameter π is obtained from octanol–water partition coefficients[11,12] and is an operationally defined "hydrophobic bonding" constant analogous to the familiar Hammett constant.[13] E_s is Taft's steric parameter.[13] Atomic refractivities[14] have been used as a measure of polarizability.[15]

In a review of the chymotrypsin literature it was found[6] that, for eight sets of substrates and inhibitors with hydrophobic groups attached to an α carbon, the coefficient with the hydrophobic term (π or $\log P$) in eq 1 or its simpler forms had a mean value with standard deviation of 1.21 ± 0.23. The dependent variable in these correlations was either $\log 1/K_m$, $\log K_i$, or $\log 1/C$. K_i is an inhibition constant and C is the molar concentration causing 50% inhibition. There was no apparent difference in the coefficients for substrates or inhibitors. This high coefficient (1.21) was suggested to be a ρ_2 area characteristic. For four sets of congeners with groups fitting the ρ_3 area, the mean coefficient with the apolar term was 0.29 ± 0.1. Binding in the ρ_3 area was found to be quite different; the ρ_3 area does not behave by our operationally defined hydrophobic

(1) This work was supported by Grant CA 11110 from the National Institutes of Health.

(2) (a) C. Hansch, E. W. Deutsch, and R. N. Smith, *J. Amer. Chem. Soc.*, **87**, 2738 (1965); (b) C. Hansch, *Accounts Chem. Res.*, **2**, 232 (1969).

(3) A. Dupaix, J. J. Béchet, and C. Roucous, *Biochem. Biophys. Res. Commun.*, **41**, 464 (1970).

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

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

(6) C. Hansch and E. Coats, *J. Pharm. Sci.*, **59**, 731 (1970).

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

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

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

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

(11) T. Fujita, J. Iwasa, and C. Hansch, *ibid.*, **86**, 5175 (1964).

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

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

(14) R. L. Shriner, R. C. Fuson, and D. Y. Curtin, "The Systematic Identification of Organic Compounds," 5th ed, Wiley, New York, N. Y., 1964.

(15) D. Agin, L. Hersh, and P. Holtzman, *Proc. Nat. Acad. Sci., U. S.*, **53**, 952 (1965).

TABLE I

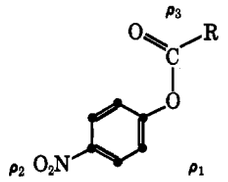
DATA EMPLOYED IN FORMULATION OF EQ 2-4

		$O_2NC_6H_4OCOR$				Log k_3		pK_1''	
No.	R	π^a	σ^{*b}	E_s^b	P_E^c	Obsd[b]	Calcd[d]	Obsd[b]	Calcd[e]
1	$ClCH_2$	0.89	1.05	−0.24	10.58	0.42	0.33	6.90	6.96
2	H	0.00	0.49	1.24	1.10	0.18	0.27	7.60	7.62
3	ICH_2	1.50	0.85	−0.37	18.50	−0.24	−0.01	6.95	7.00
4	CH_3OCH_2	0.03	0.64	−0.19	11.86	−0.47	−0.84	7.16	7.14
5	$ClCH_2CH_2$	1.39	0.38	−0.90	15.20	−1.68	−1.62	7.13	7.04
6	$C_6H_5CH_2$	2.63	0.21	−0.38	29.83	−1.73	−1.01	7.41	7.26
7	CH_3	0.50	0.00	0.00	5.72	−2.00	−1.88	7.40	7.46
8	$(CH_3)_2CH$	1.30	−0.19	−0.47	14.92	−2.47	−2.48	7.34	7.40
9	$(CH_3)_3C$	1.68	−0.30	−1.54	19.58	−3.74	−3.66	7.04	7.13
10	$Cl(CH_2)_3$	1.89	0.14	−0.40	19.82	−1.29	−1.46	7.33	7.29
11	$Cl(CH_2)_4$	2.39	0.05	−0.40	24.40	−1.35	−1.47	7.41	7.32
12	$C_6H_5CH_2CH_2$	3.13	0.08	−0.38	34.45	−0.75	−1.11	7.26	7.32
13	$C_6H_5(CH_2)_3$	3.63	0.02	−0.45	39.07	−0.92	−1.12	7.33	7.32
14	$C_6H_5(CH_2)_4$	4.13	0.02	−0.45	43.69	−1.73[f]	−0.93	7.73[f]	7.32
15	H_2NCH_2	0.00	0.49	1.24		−0.46	−0.31[g]	7.01	7.00[h]
16	L-CH_3CHNH_2	0.50	0.00	0.00		0.28	0.04[g]	7.21	7.23[h]
17	L-$(CH_3)_2CHCHNH_2$	1.80	−0.13	−0.93		0.83	0.95[g]	7.59	7.57[h]
18	L-$C_6H_5CH_2CHNH_2$	2.63	0.22	−0.38		1.57	1.53[g]	7.53	7.55[h]

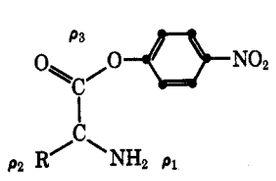
[a] From ref 11 and 12. [b] From ref 3. [c] From ref 14. [d] Calculated using eq 4. [e] Calculated using eq 14. [f] These values not used in deriving eq 4 and 14. [g] Calculated using eq 12. [h] Calculated using eq 16.

standard as other enzymic systems we have studied.[16] Substituents binding in the ρ_1 area are not well correlated by π but do correlate well with polarizability (6).

Dupaix, *et al.*,[3] employing a well-designed set of substrates (II and III), analyzed the structure–activity



II

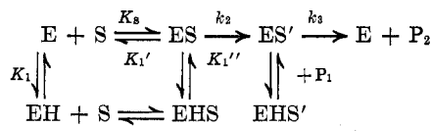


III

relationship in chymotrypsin hydrolysis in terms of electronic (σ^*) and steric (E_s) parameters but did not attempt to include hydrophobic terms in their linear free energy relationship. In II, one set of congeners has been cast in the Hein–Niemann model as we have previously done with structures of this type.[6] A second set of congeners derived from L-amino acids is formulated in III which previous work[6] indicates to be most likely.

Dupaix, *et al.*,[3] schematically represent their study in Scheme I, where S is the substrate, E and EH are the

SCHEME I



active and inactive forms of the enzyme, ES and EHS are the corresponding enzyme substrate complexes, ES′ and EHS′ the corresponding forms of the acyl–enzyme intermediate, and P_1 and P_2 are the hydrolysis products *p*-nitrophenol and carboxylic acid, respectively. Dupaix, *et al.*,[3] reported values of k_3 and pK_1'' values for the 18 compounds listed in Table I.

(16) H. F. Schaeffer, R. N. Johnson, E. Oldin, and C. Hansch, *J. Med. Chem.*, **13**, 452 (1970).

Results and Discussion

First, considering the larger set of congeners (compounds 1–14), we have derived, *via* the method of least squares, eq 2–4. Of the four independent variables of

$$\log k_3 = 2.349(\pm 1.06)\sigma^* - 1.852(\pm 0.49) \quad (2)$$
$$n = 13;\ r = 0.827;\ s = 0.668$$

$$\log k_3 = 1.890(\pm 0.87)\sigma^* + 0.792(\pm 0.58)E_s - 1.458(\pm 0.47) \quad (3)$$
$$n = 13;\ r = 0.915;\ s = 0.503$$

$$\log k_3 = 2.201(\pm 0.60)\sigma^* + 1.012(\pm 0.40)E_s + 0.374(\pm 0.22)\pi - 2.067(\pm 0.48) \quad (4)$$
$$n = 13;\ r = 0.969;\ s = 0.327$$

Table I, σ^* gave the highest correlation with log k_3 (eq 2). The best two-variable equation is eq 3, which is statistically a significant improvement over eq 2 ($F_{1,10}$ = 9.4). Adding a term in π or P_E to eq 3 gives the same result because of the large amount of cocorrelation between these two parameters (r^2 = 0.935). Equation 4 with a π term represents a statistically significant reduction in the variance when compared with eq 3 ($F_{1,9}$ = 14.6). The overall significance of eq 4 is high ($F_{3,9}$ = 136; $F_{3,9\ \alpha.005}$ = 8.7). The positive coefficient with σ^* means that k_3 is increased by electron-withdrawing substituents, while the positive coefficients with E_s mean that large groups slow down the reaction. The positive slope of the π term means that reaction is favored by hydrophobic bonding. This is surprising for the hydrolysis step. In a previous study[2a] of k_3 in a hydrolysis by emulsin, it was found that the coefficient with the hydrophobic term was negative, as one might expect. In the emulsin work, k_3 refers to the overall hydrolysis so that the meaning is somewhat different than in Scheme I. The value of the coefficient with π in eq 4 is about what we have found for binding in the ρ_3 area as characterized by $1/K_m$ or K_i correlations. It seems likely that the value of the π coefficient in eq 4 being near that found in equations correlating binding is coincidental. However, the positive coefficient is most interesting in that it indicates that some kind of apolar interaction promotes deacylation. Exactly how

TABLE II

DEACYLATION OF CHYMOTRYPSIN. DATA USED IN DERIVATION OF EQUATIONS

No.	R	E_s[a]	π[b]	σ^*[a]	Log k_3 Obsd[c]	Log k_3 Calcd[d]	Log k_3' Obsd[c]	Log k_3' Calcd[e]
1	CH_3	0.00	0.50	0.00			−2.46	−2.51
2	CH_3CH_2	−0.07	1.00	−0.10	−2.10	−2.16	−2.21	−2.25
3	$CH_3CH_2CH_2$	−0.36	1.50	−0.12	−2.22	−2.21	−2.36	−2.32
4	$CH_3(CH_2)_4$	−0.40	2.50	−0.13	−1.44	−1.48	−1.60	−1.66
5	$(CH_3)_2CH$	−0.47	1.30	−0.19	−2.61	−2.54	−2.74	−2.63
6	$(CH_3)_2CHCH_2$	−0.93	1.80	−0.13	−2.90	−2.85	−3.03	−2.95
7	$(CH_3)_3C$	−1.54	1.68	−0.30	−3.80	−3.88	−3.85	−3.93
8	$(CH_3)_3CCH_2$	−1.74	2.18	−0.17	−3.82	−3.79	−3.87	−3.87

[a] From ref 13. [b] From ref 11 and 12. [c] From ref 17. [d] Calculated using eq 6. [e] Calculated using eq 9.

apolar interactions aid deacylation is not clear. It could be that a conformational change in the enzyme is produced so that the histidine moiety is more favorably positioned to displace the acyl function, or binding the acyl function in a hydrophobic cleft could aid in breaking the acyl linkage by a kind of mechanical stretching action. Still another possibility is simply that of better orienting of substrate so that deacylation is sterically favored.

The above work of Dupaix, *et al.*,[3] can be compared with that of Fife and Milstein.[17] From their data on deacylation in Table II, eq 5 and 6 have been derived.

$$\log k_3 = 1.259(\pm 0.66)E_s - 1.707(\pm 0.65) \quad n = 7;\ r = 0.909;\ s = 0.404 \tag{5}$$

$$\log k_3 = 0.793(\pm 0.18)\pi + 1.539(\pm 0.14)E_s - 2.842(\pm 0.28) \quad n = 7;\ r = 0.998;\ s = 0.072 \tag{6}$$

Equation 6 is quite a significant improvement over eq 5 ($F_{1,4} = 155.5$). All possible combinations of π, E_s, and σ^* did not yield an equation with a lower standard deviation than eq 6. Omitting molecule 3 of Table II, Fife and Milstein[17] plotted log k_3 against E_s to obtain a rather good correlation ($r = 0.988$). Thus it occurred to us that possibly molecule 3 was behaving in a different manner from the others in Table II. Omitting this datum point, eq 7 and 8 are obtained. Not only

$$\log k_3 = 1.106(\pm 0.24)E_s - 1.967(\pm 0.26) \quad n = 6;\ r = 0.988;\ s = 0.133 \tag{7}$$

$$\log k_3 = 0.572(\pm 0.43)\pi + 1.413(\pm 0.26)E_s - 2.607(\pm 0.49) \quad n = 6;\ r = 0.998;\ s = 0.058 \tag{8}$$

does eq 8 have essentially the same shape as eq 6, the additional term in eq 8 over eq 7 is very significant statistically ($F_{1,3} = 18.5$; $F_{1,3\ \alpha,025} = 17.4$) even though only three degrees of freedom remain. The evidence that apolar forces are involved in k_3 is quite compelling.

Using Fife and Milstein's data for k_3' measured at pH 7.7 gives a set of points including one more value (Table II). From these data, eq 9 and 10 have been formulated.

$$\log k_3' = 1.065(\pm 0.58)E_s - 2.032(\pm 0.53) \quad n = 8;\ r = 0.878;\ s = 0.410 \tag{9}$$

$$\log k_3' = 0.718(\pm 0.16)\pi + 1.472(\pm 0.15)E_s - 2.869(\pm 0.22) \quad n = 8;\ r = 0.996;\ s = 0.085 \tag{10}$$

Equation 10 is an excellent correlation, quite comparable to eq 6. However, including the additional molecule makes even more apparent the importance of apolar forces in k_3. The additional term in eq 10 is quite significant compared to eq 9 ($F_{1,5} = 135$) ($F_{2,5} = 302$).

Milstein and Fife[18] have measured the parameter k_2/K_m in the acylation of chymotrypsin. Equation 11

$$\log k_2/K_m = 1.51(\pm 0.42)E_s + 0.63(\pm 0.39)\pi + 2.98(\pm 0.51) \quad n = 8;\ r = 0.976;\ s = 0.198 \tag{11}$$

results from our earlier analysis of these data.[6] Here, as in the other work by Fife and Milstein, no role can be seen for σ^*. No doubt this is because sufficient variation in this parameter was not present in the derivatives. The forces involved in the k_2 step are quite similar to those of the k_3 step (compare eq 10 and 11). The coefficients with the E_s and π terms are quite similar. That the weighting factors are the same with E_s is not unexpected. Again it is surprising that apolar forces play a role in this step. Unfortunately, the confidence intervals on the π term in eq 11 are not so tight as in eq 10. It is possible that apolar effects are not so important in k_2 as in k_3.

Compounds 15–18 in Table I are structurally so different from 1–14 that they have been treated separately, as shown in eq 12. Equation 12 is significant ($F_{1,2} = 44$; $F_{1,2\ \alpha,025} = 38.5$). The correlations with E_s ($r = 0.801$) and σ^* ($r = 0.443$) were very much poorer. Adding a second variable is rather meaningless for so few data points; however, doing so did not in any case give an equation with a lower standard deviation than eq 12. The coefficient with the π term in eq 12 is in agreement with that found in eq 4, 6, 8, 10, and 11. Even though little data are available, this group appears to behave mechanistically different than compounds 1–14. However, more sterically hindering functions might establish the need for an E_s term.

$$\log k_3 = 0.70(\pm 0.45)\pi - 0.31(\pm 0.73) \quad n = 4;\ r = 0.978;\ s = 0.219 \tag{12}$$

Two types of initial binding have been depicted in structures II and III corresponding to our earlier work.[6] The coefficients with the π term would suggest that different apolar forces are involved with k_2 and k_3 than with K_i and K_m.

The substituent effects on K_1'' are quite different from those on k_3, as seen from eq 13 and 14. The

$$pK_1'' = 0.190(\pm 0.18)E_s + 7.316(\pm 0.12) \quad n = 13;\ r = 0.568;\ s = 0.175 \tag{13}$$

$$pK_1'' = 0.292(\pm 0.09)E_s - 0.407(\pm 0.14)\sigma^* + 7.458(\pm 0.08) \quad n = 13;\ r = 0.932;\ s = 0.081 \tag{14}$$

(17) T. H. Fife and J. B. Milstein, *Biochemistry*, **6**, 2901 (1967).

(18) J. B. Milstein and T. H. Fife, *ibid.*, **8**, 623 (1969).

"best" single-variable equation is that in eq 13. The "best" two-variable equation is eq 14, which is quite significantly better than eq 13 ($F_{1,10} = 41$). The overall significance of eq 14 is high ($F_{2,10} = 33$; $F_{2,10\ \alpha.005} = 9.4$). It is interesting and logical that apolar forces have no influence on pK_1''. Adding terms in π or P_E to eq 14 does not result in improved correlation. The steric effect of the substituent on pK_1'' is qualitatively the same as for k_3, only much smaller in magnitude. The electronic effect is in the opposite direction.

As with k_3, a different story results from the analysis of molecules 15–18 of Table I. This is summarized in eq 15 and 16. Although a good correlation in terms of

$$pK_1'' = -0.28(\pm 0.29)E_s + 7.33(\pm 0.23) \quad n = 4;\ r = 0.945;\ s = 0.109 \tag{15}$$

$$pK_1'' = -0.45(\pm 0.65)E_s + 0.65(\pm 2.2)\sigma^* + 7.23(\pm 0.41) \quad n = 4;\ r = 0.996;\ s = 0.002 \tag{16}$$

r, eq 15 is not statistically significant ($F_{1,2} = 2.1$; $F_{1,2\ \alpha.1} = 8.5$). However, eq 16 is significant ($F_{2,1} = 69.6$; $F_{2,1\ \alpha.1} = 49.5$). Even though shown to be significant by the F test, very little confidence can be placed in eq 16 because of the few data points. It is simply of interest to note that it is quite different from eq 4.

Compound 14 was not included in deriving any of the above equations. The values predicted by eq 4 and 14 are wide of the mark. This is to be expected, since there are at least seven known examples[6] where a break in the structure–activity relationship of chymotrypsin substrates and inhibitors occurs when π for the side chain is 3.50 or greater. Compound 14 is yet another such example.

The type of derivatives selected by Dupaix, *et al.*,[3] in Table I constitute by far the best set yet studied from the point of view of the physical organic chemist trying to separate substituent effects. Their work, when taken with that of Fife and Milstein, clearly shows that apolar forces are important in the k_3 step.

Acknowledgment.—I wish to thank Professor Neal Cornell for suggesting the present study and for helpful discussions. I also thank Professor T. H. Fife for an enlightening discussion.

Clarification of the Acid-Catalyzed Reaction of Glyoxal with Carbamate Esters

GRAHAM F. WHITFIELD,* RODNEY JOHNSON, AND DANIEL SWERN

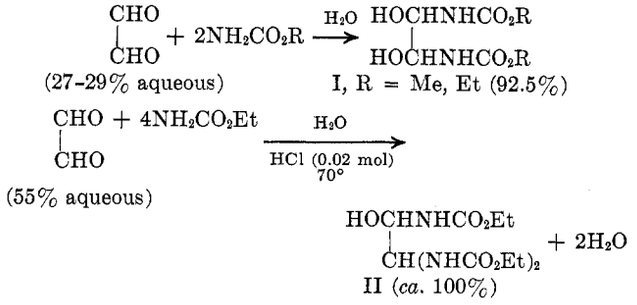
Department of Chemistry and Fels Research Institute, Temple University, Philadelphia, Pennsylvania 19122

Received May 21, 1971

The reaction of glyoxal with ethyl carbamate in the presence of concentrated hydrochloric acid gives 1,1,2,2-tetra(carbethoxyamino)ethane (VI), in contrast with the reports of Pauly and Sauter and Gaylord who claimed that the reaction product was glyoxal bis(carbethoxyimide) (III). The mechanism of formation of the tetracarbamate (VI) is discussed, and it is shown that certain parallels exist between the reactions of glyoxal with carbamate esters and with primary amines.

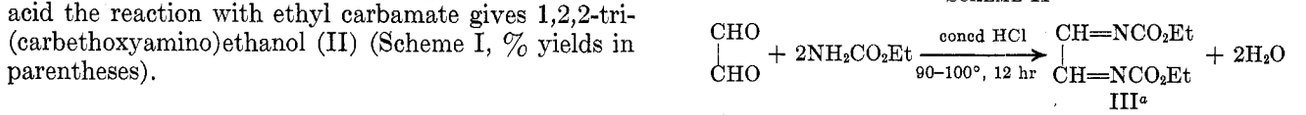
The reaction of glyoxal with carbamate esters yields a variety of products depending on the acidity or basicity of the reaction medium. For example,[1] treatment of neutralized glyoxal with carbamate esters (1:2 molar ratio) is reported to give 1,2-di(carbalkoxyamino)-ethane-1,2-diol (I), whereas in the presence of dilute acid the reaction with ethyl carbamate gives 1,2,2-tri-(carbethoxyamino)ethanol (II) (Scheme I, % yields in parentheses).

SCHEME I



There is some doubt, however, as to the structure of the product formed in stronger acid solution. The reaction of glyoxal with ethyl carbamate in the presence of concentrated hydrochloric acid was investigated first by Pauly and Sauter[2] who reported that the product was an insoluble, microcrystalline powder; the structure assigned was that of glyoxal bis(carbethoxyimide) (III) (Scheme II). No melting point was reported for the

SCHEME II



III[a]

[a] *Anal.* Calcd for $C_8H_{12}N_2O_4$: C, 47.99; H, 6.04; N, 13.99. Found: C, 47.74; H, 6.4; N, 14.07.

product, but the microanalysis was correct for the assigned structure.

A similar result was reported by Gleim[3] who examined the reaction of glyoxal with allyl carbamate in the presence of concentrated hydrochloric acid. The reaction product was claimed to be glyoxal bis(carb-allyloxyimide) (IV); however, the microanalysis was not in good agreement with this structure (Scheme III).

For some obscure reason, the abstract literature[4] subsequently referred to compound III as "carbamic acid, *N,N'* acetylene bis-, diethyl ester," $EtO_2CNHC{\equiv}CNHCO_2Et$ (V), and this prompted Gaylord[5] to re-

(1) British Patent 801,991 (B.A.S.F.); *Chem. Abstr.*, **53**, 7019*a* (1959).

(2) H. Pauly and H. Sauter, *Chem. Ber.*, **63**, 2063 (1930).

(3) C. E. Gleim, *J. Amer. Chem. Soc.*, **76**, 107 (1954).

(4) Subject Index, *Chem. Abstr.*, 3rd Decennial Index, 1927–1936.

(5) N. G. Gaylord, *J. Org. Chem.*, **20**, 547 (1955).