Molecular Mechanics and Molecular Shape. Part 4.[1-3] Size, Shape, and Steric Parameters

Amatzya Y. Meyer
Department of Organic Chemistry, Hebrew University, Jerusalem 91904, Israel

To deal with steric effects of substituents on reaction rates, two types of descriptor are proposed. Shape is characterised by the ratio (G) of the substituent's surface area to its volume. Bulk is characterised by the volume (V^a) of the portion of the substituent that is within 0.3 nm of the reaction centre. For alkyl substituents, rate constants for several series of reactions (including the classical series, acid hydrolysis of the esters RCO_2Et) correlate fairly well with V^a and G. Taft's steric parameters (E_s) correlate less well. Values of E_s for hetero-substituents are found to be contaminated by non-steric factors. It is possible to construe whether the contaminant acts to hinder or to enhance reaction.

It seems timely to confront the geometrical attributes of simple molecules[1-3] with the long established parameters of steric effect.[4] In this paper I assess briefly some variants of the traditional approach, and then describe and examine an exploratory version of an alternative.

Steric effects of substituent groups on reaction rates or equilibria are usually examined in terms of 'Taft's steric parameters'[4] (E_s) or variants thereof.[5,6] In essence, these 'parameters' are processed observable data, originally derived by averaging relative rate constants for several series of reactions. In terms of the basic series,[7] the measured constants (k) for acid hydrolysis of the esters RCO_2Et (where R varies) are divided by the constant for the standard substrate (k_0 for R_0 = Me), and the logarithm is taken [equation (1)]. To examine steric effects in other reactions, say acid alcoholyses of the esters RCO_2Et,[8,9a] two sets of empirical numbers have to be compared. Constants (k') are measured for the other reaction, and a linear correlation is sought between the series of log k' and the series of E_s [equation (2)]. A good correlation reveals an analogy:[10] under given experimental conditions, ester hydrolysis and (say) ester alcoholysis respond in an analogous way to substituent change. Such a correlation does not provide a molecular interpretation of steric effects, nor does it guarantee that the rate of the other reaction depends solely on steric factors: k and k' may have similar contributions from non-steric factors.[9b] In fact, the vast literature on this topic is pervaded[6,11] by uncertainty as to the 'extent of stericity' of the so-called steric parameters.

$$E_s = \log(k/k_0)_A \quad (1)$$

$$\log k' = a \log(k/k_0)_A + b \quad (2)$$

$$E_s = \psi r + h \quad (3)$$

$$v = r - 1.20 \, (r \text{ in Å}) \quad (4)$$

$$\log k' = \psi' v + h' \quad (5)$$

$$\text{S.D.} = (\text{M.W.}/V) - 0.29 \, (V \text{ in Å}^3) \quad (6)$$

Other scales of steric parameters have been reviewed.[5] In the main, these also are processed empirical data, except that the reactions chosen,[12] or the reaction conditions,[13] are not those of Taft (S^0, E_s'), or that the protocol for processing is different[6,14] (E_s^c, E_s^c).

Charton's steric parameters[15] are in principle different, since they seek to relate effects to a cause. Starting from an observation[4] that van der Waals radii of several chemical residues quantitatively parallel the corresponding E_s values [equation (3)], he adopted the radii themselves (after some processing) as parameters[16] [v in equation (4)]. Regressions on kinetic constants could then follow [equation (5)]. When this approach was developed, almost two decades ago, three types of shortcoming were unavoidable, as follows. (*a*) Since most chemical residues are not spherical, there is more than one way to define their radius. Charton enumerated three choices[15] and other workers endorsed more.[17] The door is thus opened to a proliferation of parameter sets, all reasonable but leading to different 'reaction constants' [ψ' in equation (5)]. (b) Charton's trigonometric procedure[18] applies strictly only to residues of types –A and $-AB_3$. To obtain v values for residues of more complex types, equation (3) has to be, and has been, worked backwards. In fact, perusal of the literature reveals a vicious circle, run back and forth to produce v values from E_s and E_s values from v.[13] In practice, then, estimation of v is based precisely on the stratagem one wished to supplant. (*c*) It may happen at the time of need that there are not enough data to estimate radii for a given group. This was the case with nitro and phenyl in the sixties.[15]

Updating Charton's Approach.—With present experience in dealing with the size and shape of molecules,[1-3,19-21] and present computational capabilities, Charton's approach can be updated. To this end, the molecule is depicted as a geometrical solid of sharply defined boundaries, and descriptors of size and shape are evaluated. One can then examine to what extent these descriptors sustain correlations with rate or other constants.

The descriptor of size that we examine here (V^a) is a delimited portion of the volume of the substituent. Delimitation is suggested by Taft's observations[4] that consequent γ-methylations of a residue (*i.e.* changes in a remote location) affect its E_s value only slightly. The selected descriptor of shape (G) is the substituent's surface-to-volume ratio. Its intervention is suggested by the non-additivity of steric influences that compounded substituents exert.[4] It seems that an index of type G is what Dash and Behera were looking for when they defined the steric density parameter (S.D.) as the ratio of molecular weight (M.W.) to volume [equation (6)].[22] In a series of congeners, G indeed increases with M.W.

Hansch calls for steric parameters that are 'uniformly and generally defined from geometrical considerations alone'.[6] The parameters V^a and G fulfill this requirement and have other helpful features: they come from computation, obviating the dependence on empirical data [(*b*) above]; they are readily calculable[1-3] (*c*); and the plurality of choice is much reduced (*a*).

Index of Size.—In a set of molecules RX, where R varies, let R be the 'residue' and RH its 'parent.' The 'anchor atom' is that

Table 1. Data for sample alkyl residues[a]

	R	RH	V^a	E_s	V	G
(**1a**) Bu^i	$CH_2CH(CH_3)_2$	$CH(CH_3)_3$	0.0526	−2.17	0.0791	13.45
(**1b**) Bu^t	$C(CH_3)_3$	$CH(CH_3)_3$	0.0716	−2.78	0.0791	13.45
(**2a**) Pentyl	$CH_2[CH_2]_3CH_3$	$CH_3[CH_2]_3CH_3$	0.0479	−1.64	0.0960	13.26
(**2b**) Neopentyl	$CH_2C(CH_3)_3$	$C(CH_3)_4$	0.0575	−2.98	0.0955	13.06

[a] V^a (volume in the anchor sphere) and V (total volume of RH) in nm^3; G (ratio of total surface area to total volume) in nm^{-1}; $CH_2[CH_2]_3CH_3$ in extended conformation.

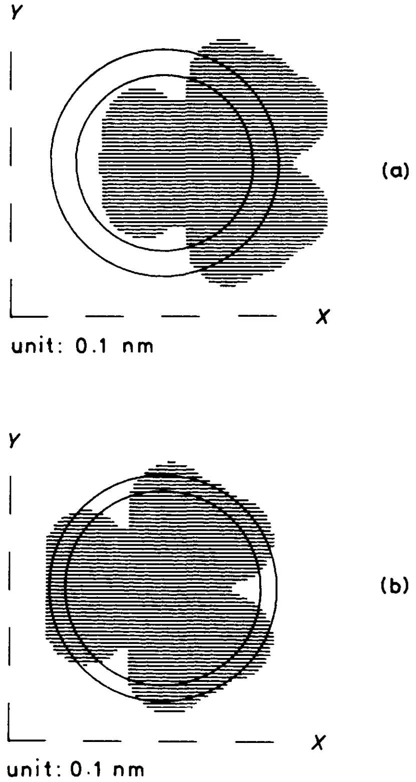

Figure 1. Sections in the van der Waals body of 2-methylpropane, parallel to the first principal plane, showing anchor sphere (outer circle, r = 0.3 nm) and anchor volume (inner circle); (*a*) through $C^{(1)}$, $r(V^a)$ = 0.232 nm; (*b*) through $C^{(2)}$, $r(V^a)$ = 0.258 nm

atom in RH by which R is linked to X. To illustrate, the parent of both $R_{(1a)} = CH_2CH(CH_3)_2$ and $R_{(1b)} = C(CH_3)_3$ is 2-methylpropane, $CH(CH_3)_3$; the anchor-atoms, respectively, are $C^{(1)}$ and $C^{(2)}$. Figure 1 shows sections through the van der Waals body of this parent, in planes parallel to the first principal plane and passing through the anchor atoms. Carbon atoms $C^{(1)}$ [in Figure 1(a), representing $R_{(1a)}$] and $C^{(2)}$ [Figure 1(b), $R_{(1b)}$] are circled.

Hypothesis. The steric effect of a residue is related to the volume of that portion of its body that is close to the anchor atom. In fact, portions distant from the point of attachment have only a slight effect, as can be gleaned from the E_s values of a few unbranched alkyl residues: ethyl, −1.31; propyl, −1.60; butyl, −1.63; pentyl, −1.64 (see Appendix for a more extensive tabulation). The significance of volume close to the anchor atom is illustrated by the E_s values of $CH_2CH(CH_3)_2$ and $C(CH_3)_3$, −2.17 and −2.78. Figure 1 shows indeed how much more encumbered is the anchor region of the latter (**1b**) than that of the former (**1a**).

We have not sought as yet the best mode of cutting. Also, for the time being, we perform the calculations on the parent molecule RH, not on the residues R. In this way, one sole data file (as it comes out of the molecular-mechanical optimisation) can be used in sequence for all conceivable anchors in a molecule. The error introduced is almost independent of residue (0.10—0.15 nm^3 per molecule) and does not interfere with linear correlations.

We chose to consider the volume of RH that is contained in a sphere about the anchor atom. This is the volume in the 'anchor sphere' (V^a). Now, if the radius of the sphere (r^a) is taken too small, V^a is not sensitive enough to variations in RH. On the other hand, if r^a is taken too large, a large portion of a molecule or all of it is circumscribed by the sphere, and again there is no discrimination. The compromise we chose is r^a = 0.3 nm, giving behaviour within series resembling that of Taft's steric parameters. For example, there is a significant difference between CH_2CH_3 and $CH_2CH_2CH_3$, but only a very small difference between $CH_2CH_2CH_3$ and $CH_2CH_2CH_2CH_3$ (see Appendix). In Figure 1(a) and (b), the outer circles are drawn about anchor atoms at r^a = 0.3 nm. It is readily noticed that the anchor-sphere of the tertiary isomer [Figure 1(b)] contains much more filled volume than that of the isobutyl system [Figure 1(a)]. In each case, the inner circle is a great circle of a sphere of volume equal to V^a; *i.e.* its radius $r(V^a)$ is $(3V^a/4\pi)^{1/3}$. Relevant numbers are cited in Table 1.

Index of Shape.— In Figure 1 and entries (**1a**) and (**1b**) of Table 1, we compared two isomeric residues derived from the same parent. Now we compare two isomeric residues derived from isomeric parents. These are $R_{(2a)} = CH_2[CH_2]_3CH_3$, derived from pentane, and $R_{(2b)} = CH_2C(CH_3)_3$, derived from 2,2-dimethylpropane. As the data in Table 1 show, their V^a values differ by 0.0096 nm^3, and their E_s values by 1.34 units. This contrasts with values for $R_{(1a)}$ and $R_{(1b)}$, where the difference in V^a was twice as large, while the difference in E_s was only half as much.

Being isomers, pentane and 2,2-dimethylpropane have an almost identical total volume (V), 0.096 nm^3.[1] Their shapes, by contrast, differ considerably. The extended conformation of pentane is elongated, with minimal overlapping between contiguous groups and consequently a comparatively large surface area (S). 2,2-Dimethylpropane is a spherical top, with maximal overlapping and a comparatively small surface area. The large difference in shape, as well as the small difference in V^a, is stressed in Figure 2.

The ratio $G = S/V$ measures the molecular surface area per unit volume. For stereoisomers, which have an almost equal volume, G is a descriptor of globularity:[1-3] the more globular species has the lower G. For congeners of non-equal volume, G reflects the relative compactness. The many cases cited in the Appendix illustrate that G diminishes with (*a*) number of atoms in the residue (*cf.* propyl, butyl, pentyl); (*b*) globularity of residue (*cf.* pentyl, cyclopentyl); (*c*) extent of branching (*cf.* pentyl, isopentyl, t-pentyl).

Hypothesis. It is reasonable to assume that, in a chemical

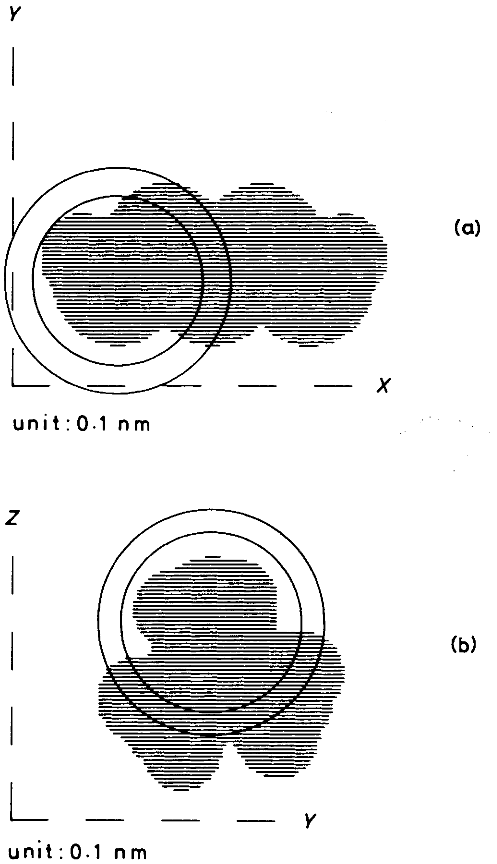

Figure 2. Sections in van der Waals body of two isomers, illustrating globularity; (*a*) pentane, anchor at $C^{(1)}$, cut through carbon plane, $r(V^a) = 0.225$ nm; (*b*) 2,2-dimethylpropane, anchor at $C^{(1)}$, cut through $C^{(1)}$ parallel to third principal plane, $r(V^a) = 0.239$ nm

reaction, the three structural features that push G down would act to hinder the approach of a reagent or to encumber a transition state. Hence, G could be the descriptor of shape we need. In fact, one calculates lower G values for residues of more negative E_s.

As with V^a, and for the same technical reasons, we calculate G for RH rather than for R. This attributes the same G to all residues R derived from a given parent.

Kinetic Coefficients.—In this section we consider alkyl residues only. To examine the two hypotheses, we performed linear regression analyses based on V^a and G for several series of rate constants k and of $\log k$ from the literature. The series, by necessity, are short (6—10 members). It was strange to find that, in some series, k correlates somewhat better with V^a and G than does $\log k$. Also, correlations of k are always of the expected form, $k = -aV^a + bG + c$ (a and b positive), *i.e.* bulk and complexity push k down. Correlations of $\log k$ are erratic in this sense. To examine the contribution of non-steric factors,[23] analyses were repeated with inclusion of Taft's polar parameters[4] σ^* or, when for k, with antilog(σ^*).

It is not our intention to give a detailed account of the analyses (k and $\log k$, with σ^* or without, variegated series), and only some of the derived equations are chosen as illustration. Correlations are fair to good. The quantities given with each equation are, in this order: number of points, multiple linear regression coefficient, standard deviation of residuals, value calculated and value reported for the residue $CH_2CH_2CH_2CH_3$ (Bu). Bu was chosen because it is intermediate in size, shape and E_s, and occurs in most series. Predictions for other residues are readily reconstructed by using V^a and G values from the Appendix in the correlation equations. The estimated regression parameters are given with their standard errors. These are not small, but usually there is only one point that deviates seriously from the reported value.

Equations (7)—(9) were obtained for acid hydrolysis at 25 °C of esters RCO_2Et in 70% dioxane:[7] the backbone of Taft's work. The offending point in the correlations for $10^5 k$ is the ethyl residue. As for saponification, it has been suspected[12,24] that the rates for RCO_2Me in aqueous 40% dioxane at 35 °C are sensitive to hyperconjugation, whereas those for $MeCO_2R$ are not. Analysis with exclusion of hyperconjugation parameters[14] [equations (10) and (11)] does not reveal any striking difference between the two cases. If anything, reactions of RCO_2Me are more sensitive to the shape of R than those of $MeCO_2R$. The troublesome points are Et for RCO_2Me and Pr^i for $MeCO_2R$. Effects of size and shape are stressed by comparing $MeCO_2R$ [aqueous 40% dioxane; equation (10)] with[25] $PhCO_2R$ [aqueous 60% dioxane; equation (12)]. In the latter case, V^a and G have less effect than in the former, presumably because Ph in the substrate poses more severe steric requirements than does Me. Et is the deviating point for $PhCO_2R$. The outcome of including polar parameters is shown in equations (13) and (14). Steric requirements of the approaching reagent are illustrated by alcoholyses at 25 °C of β-naphthyl esters:[8] $C_{10}H_7OCOR + R'OH$, where $R'OH = CH_3OH$, $CH_3CH_2CH_2OH$, or $CH_3CHOHCH_3$. Equations (15)—(17) show that as R′OH becomes more congested (see Me, Pr, and Pr^i in the Appendix), the size and shape of the substituent R gradually lose their role in affecting the rate.

$$\log(10^5k) = -44.84(9.21)V^a - 0.086(0.115)G + 3.566(1.980)$$
$$(RCO_2Et\text{, acid hydrolysis: }10,\ 0.920,\ 0.245,\ 0.254,\ 0.253) \quad (7)$$

$$10^5k = -76.99(13.61)V^a + 0.501(0.170)G - 1.220(2.927)$$
$$(RCO_2Et\text{: }10,\ 0.974,\ 0.362,\ 1.89,\ 1.79) \quad (8)$$

$$10^5k = -7.407(4.733)\text{antilog}(\sigma^*) - 134.65(38.87)V^a + 0.829(0.261)G + 2.414(3.534)$$
$$(RCO_2Et\text{: }10,\ 0.982,\ 0.329,\ 1.72,\ 1.79) \quad (9)$$

$$k = -253.7(71.8)V^a + 2.993(0.915)G - 23.50(16.18)$$
$$(MeCO_2R\text{, saponification: }8,\ 0.982,\ 1.372,\ 4.97,\ 5.38) \quad (10)$$

$$k = -196.6(99.4)V^a + 4.083(1.266)G - 40.66(22.38)$$
$$(RCO_2Me\text{, saponification: }8,\ 0.970,\ 1.899,\ 5.33,\ 5.96) \quad (11)$$

$$k = -21.22(9.50)V^a + 0.285(0.103)G - 2.52(1.86)$$
$$(PhCO_2R\text{, saponification: }9,\ 0.961,\ 0.170,\ 0.328,\ 0.289) \quad (12)$$

$$k = 3.43(2.19)\text{antilog}(\sigma^*) + 8.97(2.10)V^a + 0.206(0.105)G - 5.47(2.51)$$
$$(PhCO_2R\text{: }9,\ 0.974,\ 0.152,\ 0.304,\ 0.289) \quad (13)$$

$$\log k = 1.09(3.64)\sigma^* - 58.51(21.75)V^a - 0.170(0.086)G + 4.79(1.54)$$
$$(PhCO_2R\text{: }9,\ 0.983,\ 0.139,\ -0.452,\ -0.539) \quad (14)$$

$$k = -5.76(1.51)V^a + 0.049(0.019)G - 0.27(0.33)$$
$$(CH_3OH\text{: }6,\ 0.986,\ 0.029) \quad (15)$$

$$k = -1.17(0.25)V^a + 0.014(0.006)G - 0.107(0.098)$$
$$(CH_3CH_2CH_2OH\text{: }6,\ 0.977,\ 0.009) \quad (16)$$

$$k = -0.07(0.03)V^{a} + 0.000\,94(0.000\,38)G - 0.0085(0.0066)$$
$$(CH_3CHOHCH_3\text{: } 6, 0.974, 0.0006) \quad (17)$$

In brief, the explained variations are 85% or higher; in some contexts as high as 95%. We take this to imply that steric effects of substituents can be quantitatively related to their size and shape, but that our particular choice of attributes needs to be improved. Pathways to improvement come readily to mind: usage of atomic radii other than 'van der Waals'[2] or of an anchor radius other than 0.3 nm, severing H from RH, paying more attention to the conformation of R, and more.

Steric Parameters.—At least by definition, the traditional steric parameters are logarithms of kinetic constants [equation (1)], and not calculable descriptors of geometry. To check to what extent the parameters correlate with the descriptors, we carried out regression analyses based on V^{a} and G for several types of alkyl parameters: Taft's E_s, on Hansch and Leo's scale;[26] Hancock's E_s^{c};[14] and Dubois' E_s'.[13] Correlations for E_s and antilog(E_s), based on the nineteen residues listed in the Appendix, are cited in equations (18) and (19). It transpires that V^{a} and G account for a considerable portion of the variation in E_s, but not for a preponderant portion. The correlations in equations (7)—(17) look better than those in equations (18) and (19).

Various explications come to mind. One is that the quality of equations (7)—(17) is only apparent, since the series examined contained few members. Such a suspicion would undermine much previous work, including that of Charton. Another is that details in the evaluation of V^{a} and G have to be retouched. Yet, the parameters E_s cannot be clear of fault (see before). (*a*) They may contain non-steric contributions that do not evolve in parallel with steric factors. As an illustration of one type, cycloheptyl and butyl have the same polar parameter[26] (−0.13), even though the former is incontestibly more encumbered than the latter. As an illustration of another type, ethyl and cyclobutyl have the same E_s (−1.3), and this cannot be interpreted on steric grounds alone. Note that cycloheptyl and butyl do not occur together in any of the series to which equations (7)—(17) refer; neither do ethyl and cyclobutyl. (*b*) The parameters E_s were assembled from different sources, derived from variegated reactions or by roundabout ways, and cannot all have the same significance. Taft advocated the use of average values on the grounds that small specific effects and consequences of experimental error would be reduced.[27] The outcome is that the E_s list does not constitute a uniform series, unlike the regressands in each of equations (7)—(17). Compare the quality of equation (8) with that of equation (18), bearing in mind that the latter purportedly refers to the reaction characterised by the former!

As it happens, reservation (*a*) is stronger than (*b*). In fact, when E_s values are replaced by Hancock's E_s^{c} parameters,[14] the correlations improve immediately [equations (20) and (21)], whereas there is no improvement with Dubois' E_s' parameters. One might recall that E_s' values all refer to the same series of reactions. Value of E_s^{c}, by contrast, are obtained by correcting E_s for non-steric factors: $E_s^{c} = E_s + 0.306(N - 3)$, where N is the number of hydrogen atoms on the anchor atom.

$$E_s = -71.38(10.81)V^{a} - 0.316(0.152)G + 6.202(2.554)$$
$$(19, 0.890, 0.435, -1.50, -1.63) \quad (18)$$

$$\text{antilog}(E_s) = -0.76(0.29)V^{a} + 0.0048(0.0042)G - 0.0023(0.0696)$$
$$(19, 0.794, 0.0118, 0.0265, 0.0234) \quad (19)$$

$$E_s^{c} = -86.93(8.77)V^{a} - 0.323(0.124)G + 6.709(2.074)$$
$$(19, 0.950, 0.353, -1.84, -1.94) \quad (20)$$

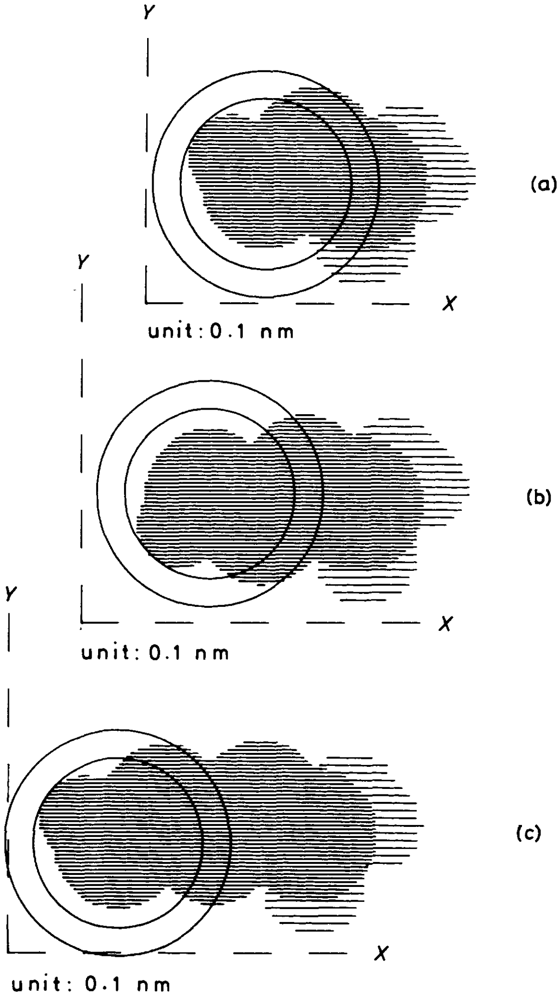

Figure 3. Sections in the main-atom plane through the van der Waals body of three carboxylic acids, stressing the spatial relation between anchor sphere (centres at ω-carbon atom, r = 0.3 nm) and hetero-grouping O_2H (spaced shading); (*a*) propanoic, $r(V^{a})$ = 0.227 nm, regions overlap; (*b*) butanoic, $r(V^{a})$ = 0.226 nm, regions tangential; (*c*) pentanoic, $r(V^{a})$ = 0.225 nm, regions distant

$$\text{antilog}(E_s^{c}) = -0.32(0.17)V^{a} + 0.0083(0.0024)G - 0.0833(0.0401)$$
$$(19, 0.881, 0.0068, 0.0136, 0.0115) \quad (21)$$

Now to hetero-residues. It is known that the 'steric parameters' E_s of some hetero-residues encode resonance on top of purely steric factors.[9b,28] Perusal of the numbers (Appendix) shows that this cannot apply to all. Can $CH_2CH_2CH_2CO_2H$ (E_s = −1.65) be so much less crowded than $CH_2CH_2CO_2H$ (− 2.21)? One might say, as shown in Figure 3, that reaction rates are affected by the proximity of the hetero-assembly O_2H to the reaction centre. In $[CH_2]_2CO_2H$ the anchor sphere englobes somes of O_2H, retarding reaction, whereas it is just tangential to O_2H in $[CH_2]_3CO_2H$ and far from it in $[CH_2]_4CO_2H$. But this is an argument in terms of distance-dependent effects, field or inductive perhaps,[29] not in terms of size or shape.

It has been stated that the steric mechanism should be identical in two structures of the same size and shape, regardless of the particular elements of which they are formed.[9c] To see that E_s values of hetero-residues do not verify this tenet,

Table 2. Geometrical attributes of hetero-residues[a]

	OCH_3	CH_2OH	CH_2F	CF_3
V	0.0328	0.0337	0.0294	0.0339
S	0.523	0.542	0.482	0.545
G	15.97	16.09	16.36	16.07
$\sigma(ab)$	0.132	0.134	0.128	0.150
$\sigma(ac)$	0.133	0.138	0.118	0.117
$\sigma(bc)$	0.081	0.078	0.078	0.117
E_s	−0.55	−1.21	−1.48	−2.40

[a] Calculations for the van der Waals bodies[1,3] of the residues themselves, not of parent molecules; cross-sectional areas σ in three planes of the principal co-ordinate system;[3] V in nm^3 per molecule, S and σ in nm^2 per molecule, G in nm^{-1}.

consider OCH_3, CH_2OH, CH_2F, and CF_3. The first three are 'classical isosteres':[30] CH_2 is isosteric with O, and F with OH. Descriptors for these residues (not the parent structures!) are listed in Table 2. The residues OCH_3, CH_2OH, and CF_3 are similar by any steric criterion; CH_2F is somewhat smaller and less compact than the others. And yet, their E_s values are widely different, spanning a range of 1.85 units, with CH_2F out of order.

Using results for alkyl residues, we can at least distinguish 'well behaved' from 'ill behaved' cases, and identify the sense of ill behaviour. The Appendix cites all tabulated[26] E_s values for residues of three constitutive classes: alcohols, organic halogenides, and carboxylic acids (nos. 20—47). For each of these, V^a and G were calculated, and E_s was predicted from equation 19; it was then compared with the tabulated value.[26]

When hetero-residues are arranged sequentially by literature values, the 'well behaved' cases fall in the middle of the list (CH_2OH to CMe_2OH; say, from −1 to −2 E_s units). Near the origin (F to Br), literature values are too low in absolute value; far from it (CH_2CH_2Cl to CBr_3), they are too high. It is indeed as if the experimental E_s parameters are contaminated by non-steric contributions. In some cases, these enhance reactivity, pulling E_s up (F to Br); in other cases, they retard reaction, pushing E_s down (CH_2CH_2Cl to CBr_3).

Conclusion.—The notion of 'steric effects' is problematic in the sense that parameters meant to measure them may be contaminated by non-steric factors, including distance-dependent effects. For alkyl residues, the traditional parameters (kinetic coefficients or tabulated E_s values) can still be correlated fairly well with descriptors of size and shape. For hetero-residues, the descriptors indicate the extent and sense of intervention of the non-steric factors.

The foregoing study is exploratory. I concentrated on identifying appropriate types of descriptor rather than on optimising the descriptor of each type. Technical details can be altered and correlations improved. But what we already have is good enough to vindicate the intuitive expectation: the steric effect of a substituent has to do with its shape and with the bulk of its portion that is close to the reaction-centre, and the interdependence is of a simple form. And conversely: since V^a and G come from straightforward computation, one has a simple way to forecast effects that are of purely steric origin.

Table 3. Steric descriptors and parameters for selected residues

No.	Residue	V^a	G	E_s
1	Me	0.0284	16.42	−1.24
2	Et	0.0431	14.83	−1.31
3	Pr	0.0478	14.05	−1.60
4	Bu	0.0479	13.57	−1.63
5	Pentyl	0.0479	13.26	−1.64
6	Isopentyl	0.0508	13.09	−1.59
7	Bu^i	0.0526	13.45	−2.17
8	Pr^i	0.0574	14.05	−1.71
9	Neopentyl	0.0575	13.06	−2.98
10	C_4H_7	0.0579	13.10	−1.30
11	C_5H_9	0.0616	12.63	−1.75
12	Bu^s	0.0621	13.57	−2.37
13	C_6H_{11}	0.0625	12.34	−2.03
14	C_7H_{13}	0.0661	12.04	−2.34
15	1-Ethylpropyl	0.0667	13.26	−3.22
16	Bu^t	0.0716	13.45	−2.78
17	CMe_2Et	0.0758	13.09	−3.41
18	$1\text{-}MeC_6H_{10}$	0.0787	12.24	−3.27
19	$CHMeCMe_3$	0.0892	12.71	−4.57
20	F	0.0122	21.35	−0.46
21	OCH_3	0.0339	15.97	−0.55
22	Cl	0.0254	16.53	−0.97
23	$CHOHCH_3$	0.0497	14.67	−1.15
24	Br	0.0329	15.32	−1.16
25	CH_2OH	0.0352	15.97	−1.21
26	I	0.0408	14.14	−1.40
27	CH_2F	0.0309	16.25	−1.48
28	CH_2Cl	0.0404	14.77	−1.48
29	CH_2Br	0.0434	14.23	−1.51
30	$[CH_2]_4CO_2H$	0.0480	13.08	−1.56
31	CH_2I	0.0455	13.58	−1.61
32	$[CH_2]_3F$	0.0472	13.99	−1.64
33	$[CH_2]_3CO_2H$	0.0481	13.33	−1.65
34	$[CH_2]_3Cl$	0.0477	13.45	−1.72
35	$CHClCH_3$	0.0550	14.02	−1.74
36	CMe_2OH	0.0639	13.89	−1.95
37	CH_2CH_2Cl	0.0468	14.02	−2.14
38	$CH_2CH_2CO_2H$	0.0487	13.72	−2.21
39	CH_2CH_2Br	0.0478	13.62	−2.24
40	CH_2CH_2I	0.0482	13.00	−2.26
41	CF_3	0.0354	15.93	−2.40
42	C_6H_5	0.0610	11.83	−2.41
43	$CHICH_3$	0.0594	13.00	−2.60
44	CH_2CMe_2OH	0.0550	13.32	−2.98
45	CCl_3	0.0643	13.33	−3.30
46	$CHOHCMe_3$	0.0637	13.08	−3.45
47	CBr_3	0.0729	12.38	−3.67

Appendix.—Molecules were constructed computationally by tracing a van der Waals sphere[20] about each atom. Volume[1] and surface area[3] were evaluated as described elsewhere. To obtain V^a, scan of the van der Waals body was limited to points within the anchor sphere. Geometries of alkanes came from MM2,[31] those of halides[32] and alcohols[33] from our force field. Propanoic acid was attributed its ED geometry.[34] Butanoic and pentanoic acid were constructed by appending a methyl or ethyl group to propanoic acid. Figures were traced by running the pen in parallel lines across the section and making it 'draw' when within the van der Waals body, and 'move' otherwise. Values of E_s were taken from Hansch and Leo's tabulation,[26] and are on their scale (lower by 1.24 units than on Taft's scale[4]). Values of σ^* were also taken from Hansch and Leo. Statistical analyses were performed by the IMSL routines BECOVM and RLMUL.[35]

Table 3 lists V^a (nm^3 per molecule), G (nm^{-1}), and E_s for the 47 residues alluded to in the main text.

References

1 Part 1, A. Y. Meyer, *J. Chem. Soc., Perkin Trans. 2*, 1985, 1161.

2 Part 2, A. Y. Meyer, *J. Mol. Struct.*, 1985, **124**, 93.

3 Part 3, A. Y. Meyer, *J. Comput. Chem.*, 1986, **7**, 144.

4 R. W. Taft, in M. S. Newman, 'Steric Effects in Organic Chemistry,' Wiley, New York, 1956, p. 600.

5 R. Gallo, *Prog. Phys. Org. Chem.*, 1983, **14**, 115.

6 S. H. Unger and C. Hansch, *Prog. Phys. Org. Chem.*, 1976, **12**, 91.

7 G. Davies and D. P. Evans, *J. Chem. Soc.*, 1940, 339.

8 M. Harfenist and R. Baltzly, *J. Am. Chem. Soc.*, 1947, **69**, 362.

9 J. E. Leffler and E. Grunwald, 'Rates and Equilibria of Organic Reactions,' Wiley, New York, 1963, (*a*) Table 7—18; (*b*) p. 228; (*c*) p. 231.

10 S. Wold and M. Sjöström, in 'Correlation Analysis in Chemistry,' ed. N. B. Chapman and J. Shorter, Plenum Press, New York, 1978.

11 H. Mager, P. P. Mager, and A. Barth, *Tetrahedron*, 1979, **35**, 1953.

12 U. Berg, R. Gallo, G. Klatte, and J. Metzger, *J. Chem. Soc., Perkin Trans. 2*, 1980, 1350.

13 J. A. MacPhee, A. Panaye, and J. E. Dubois, *Tetrahedron*, 1978, **34**, 3553.

14 C. K. Hancock, E. A. Meyers, and B. J. Yager, *J. Am. Chem. Soc.*, 1961, **83**, 4211.

15 M. Charton, *J. Am. Chem. Soc.*, 1969, **91**, 615.

16 M. Charton, *J. Am. Chem. Soc.*, 1975, **97**, 1552.

17 E. Kutter and C. Hansch, *J. Med. Chem.*, 1969, **12**, 647.

18 M. Charton, *Prog. Phys. Org. Chem.*, 1971, **8**, 235.

19 M. L. Connolly, *J. Am. Chem. Soc.*, 1985, **107**, 1118.

20 A. Gavezzotti, *J. Am. Chem. Soc.*, 1985, **107**, 962.

21 M. Santavý and J. Kypr, *J. Mol. Graphics*, 1984, **2**, 47.

22 S. C. Dash and G. B. Behera, *J. Indian Chem. Soc.*, 1980, **57**, 542.

23 W. A. Pavelich and R. W. Taft, *J. Am. Chem. Soc.*, 1957, **79**, 4935.

24 C. K. Hancock, B. J. Yager, C. P. Falls, and J. O. Schreck, *J. Am. Chem. Soc.*, 1963, **85**, 1297.

25 C. K. Hancock and C. P. Falls, *J. Am. Chem. Soc.*, 1961, **83**, 4214.

26 C. Hansch and A. J. Leo, 'Substituent Constants for Correlation Analysis in Chemistry and Biology,' Wiley, New York, 1979.

27 R. W. Taft, *J. Am. Chem. Soc.*, 1952, **74**, 2729.

28 J. Shorter, in 'Advances in Linear Free Energy Relationships,' ed. N. B. Chapman and J. Shorter, Plenum Press, London, 1972.

29 R. D. Topsom, *Acc. Chem. Res.*, 1983, **16**, 292.

30 C. W. Thornber, *Chem. Soc. Rev.*, 1979, **8**, 563.

31 N. L. Allinger, *J. Am. Chem. Soc.*, 1977, **99**, 8127.

32 A. Y. Meyer, *J. Mol. Struct.*, 1977, **40**, 127; 1978, **49**, 383; 1983, **94**, 95; A. Y. Meyer and N. Ohmichi, *ibid.*, 1981, **73**, 145.

33 A. Y. Meyer, *J. Mol. Struct.*, 1983, **105**, 143.

34 J. L. Derissen, *J. Mol. Struct.*, 1971, **7**, 81.

35 International Mathematical and Statistical Libraries, Inc., 'Reference Manual,' 7500 Bellaire Bd., Houston (Texas), 1982.

Received 26*th November* 1985; *Paper* 5/2076