A Novel Parameter (S_I) for Three-dimensional Shape Similarity between Groups: Correlation with Molecular Recognition and Biological Activity

Jun-ichi Kato,† Masato M. Ito,‡ Miho Tsuyuki, Shoji Skimizu, Yoshikatsu Kainami, Takashi Inakuma, Harumitsu Matsuoka, Takashi Isago, Koki Tajima and Tadashi Endo *
Department of Chemistry, College of Science and Engineering, Aoyama Gakuin University, Chitosedai, Setagaya-ku, Tokyo 157, Japan

A new parameter (S_I) representing the three-dimensional shape similarity between groups is defined as the degree of a maximal overlap between a pair of groups located within a three-dimensional grid. S_I, which ranges from 0 to 1, has been calculated for 105 pairs of groups including straight-chain (C_1–C_6), branched (C_4–C_6), and cyclic alkyl groups, and a phenyl group. A close relationship has been found to exist between S_I and (*i*) the degree of the discrimination between groups by a given group in the oxidation of a pair of associating thiols and (*ii*) relative biological activity (*i.e.*, effect on cat blood pressure) of a homologous series of benzilic esters.

Increasing attention has been focused on the role of molecular shape in chemical events such as chemical selectivity[1] and molecular recognition.[2] In an attempt to quantify molecular shape, STERIMOL[3] has been developed, consisting of five parameters. Since molecular shape is very difficult to represent as a single parameter, it would be most effective to quantify a relationship between molecular shapes (*e.g.*, shape similarity) rather than molecular shape itself.[4]

Simon *et al.* have defined a parameter, 'MSD' (minimal steric difference), as the non-overlapping molecular volumes at the maximal superposition degree of the two molecules.[5] A method was proposed for determining the degree of similarity among structures using graph theory.[6] An index of similarity, R_{AB},[7] between molecules (or molecular fragments) A and B was introduced in terms of their electron densities (ρ_A and ρ_B).§

Amoore has quantitatively evaluated similarity in molecular shape by subjecting silhouette photographs of molecular models of alarm pheromones to machine pattern recognition.[9] Further, Allinger[10] has proposed a parameter, *A*, to represent molecular-shape similarity, which is expressed as the ratio of overlapping volume (V_O) to the sum of V_O and non-overlapping volume (V_{NO}) when CPK molecular models of a pair of molecules are superimposed by the criterion that one superimposes the moments of inertia [eqn. (1)].

$$A = V_O/(V_O + V_{NO}) \qquad (1)$$

In order to elucidate the factors controlling molecular recognition, we examined the selectivity of a pair of associating thiols (**1** and **2**) during oxidation, with their structures and environment (*e.g.*, temperature and solvents) being systematically changed. As a result, it has been demonstrated that a certain geometrical shape as well as intermolecular association is required for molecular recognition.[11] On the basis of further investigation including (*i*) structure dependence of the formation of crystalline 1:1 complexes between a pair of acylurea derivatives and (*ii*) their X-ray analyses, we have proposed a *similarity recognition hypothesis*[2] in which the three-dimensional shape similarity between interacting groups in reacting molecules is responsible for more specific and precise molecular recognition than would otherwise be achieved.

The term 'shape similarity' is obscure and *qualitative*. Hence there is a need for quantification of the three-dimensional shape similarity between two groups. We report here (*i*) a novel method for quantifying the similarity in three-dimensional shape between groups including straight-chain, branched and cyclic alkyl groups, and a phenyl group and (*ii*) significance of the novel shape similarity parameter (S_I) mentioned above in the specific discrimination between groups by a given group and in biological activity of homologues.

Methods

Definition of S_I.—Assuming that an atom is represented by a sphere of appropriate van der Waals radius,[12,¶] a group can be considered to be an aggregate of these spheres. The two groups (G^1 and G^2) are so placed as to give maximal overlap in a three-dimensional grid in which the points (the grid intersections) are located with a constant spacing (I)‖. One investigates which of the points are involved in a given atom (Fig. 1). The distance (R_{PQ}) between point P (x_P, y_P, z_P) and atom Q (x_Q, y_Q, z_Q) is calculated from the usual Euclidean distance metric [eqn. (2)].

$$R_{PQ} = [(x_P - x_Q)^2 + (y_P - y_Q)^2 + (z_P - z_Q)^2]^{0.5} \qquad (2)$$

Point P can be regarded as involved in atom Q when $R_{PQ} \leqslant R_Q$, where R_Q is the van der Waals radius of atom Q, atom Q being part of group G^1. The number of the points involved in groups

---

† *Present address:* Department of Applied Chemistry, Tokyo University of Agriculture and Technology, Koganei, Tokyo 184, Japan.

‡ *Present address:* Institute of Life Sciences, Soka University, 1–236 Tangi-cho, Hachioji, Tokyo 192, Japan.

§ An *ab initio* method has been developed for the computation of this similarity index of valence electron density.[8]

¶ For a carbon atom involved in a benzene ring, the value of 1.77 Å is used.[13]

‖ The spacing between the grid intersections (I) is determined, with both its influence on the S_I and cpu time taken into consideration. Alterations in S_I values with I are small for $I \leqslant 0.5$ Å: S_I values for the pair Ph–n-C_5H_{11} are 0.53 for I = 0.5, 0.2 and 0.1 Å; those for the pair n-C_4H_9–n-C_5H_{11} are 0.83, 0.83 and 0.82 for I = 0.5, 0.2 and 0.1 Å, respectively. These results are in agreement with those of Stouch and Jurs.[13]

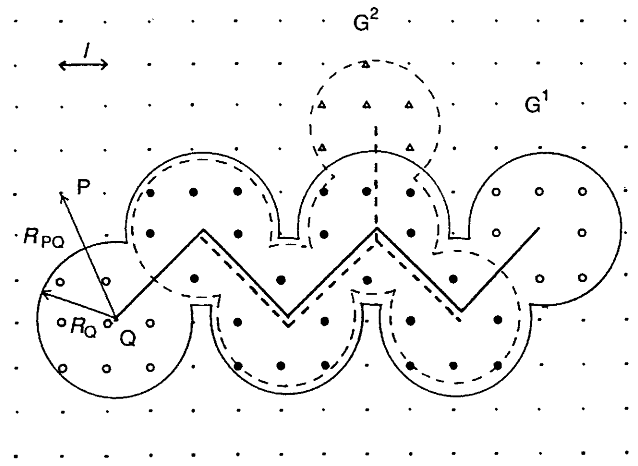


Fig. 1 A schematic drawing of a pair of groups (G^1 and G^2) superimposed in a three-dimensional grid in which the points are located with constant spacing (l) (a sectional diagram). For the symbols P, Q, R_{PQ}, and R_Q, see the text.

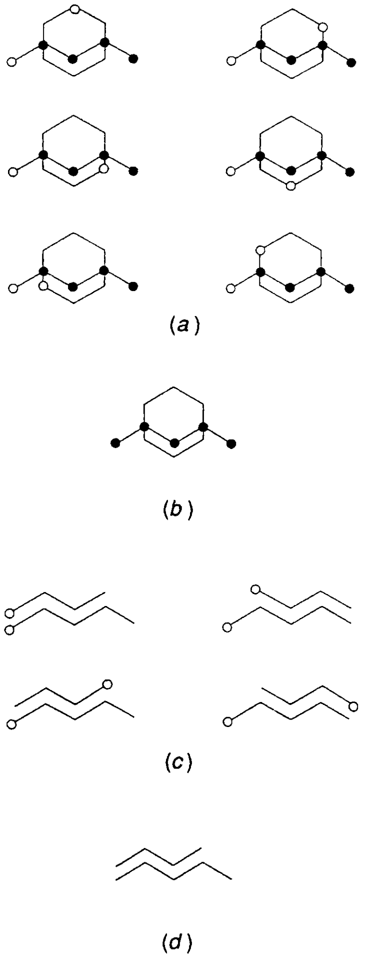


Fig. 2 Schematic drawings of the mode of overlapping between the paired groups [(a) and (c)] and between the corresponding paired molecules [(b) and (d)]. (a) n-Pentyl and phenyl groups; (b) n-pentane and benzene; (c) n-butyl and n-pentyl groups; (d) n-butane and n-pentane. The symbol ○ represents a radical centre.

G^1 and G^2 are expressed as N_1 and N_2, respectively. Therefore, the number of the points involved in either G^1 or G^2 (N_E) is given by eqn. (3) where N_B is the number of the points involved

$$N_E = N_1 + N_2 - N_B \qquad (3)$$

in both G^1 and G^2. In the light of parameter A,[10] we have defined the S_I as the ratio $N_B:N_E$ [eqn. (4)]. Parameter S_I thus takes values in the range 0–1.

$$S_I = N_B/N_E \qquad (4)$$

Calculation on 'Molecular' Basis.—There are two methods for calculating the S_I for a pair of groups (G^1 and G^2): method 1 is concerned with the paired *groups* (or *radicals*) themselves (G^1 and G^2), and method 2 with the corresponding paired *molecules* (G^1-H and G^2-H) in which the groups in question are each attached to a hydrogen atom. The former is more suitable for the definition of the S_I for the pair G^1–G^2, however its use causes a marked increase in the number of modes of overlap between G^1 and G^2 (Fig. 2). When S_I values for two paired groups [(1) Ph–n-C_5H_{11} and (2) n-C_4H_9–n-C_5H_{11}] were calculated by these two methods, the difference in S_I values between the two methods was less than 0.02.* Therefore, the S_I for a pair of *groups* (G^1 and G^2) is represented by the S_I for the corresponding pair of *molecules* (G^1-H and G^2-H) for computational convenience.

Results and Discussion

S_I Values.—Table 1 shows the S_I values for 105 pairs of groups (G^1 and G^2) including straight-chain alkyl, branched alkyl, and cyclic alkyl groups, and a phenyl group. Since S_I for the paired groups G^1–G^2 is represented by that for the corresponding paired molecules G^1-H–G^2-H for simplicity and since several different groups (G) afford a single molecule (G-H) (Table 2), S_I values for these groups (G) are all equal to one another. In Table 2 are listed seven sets of groups, for which each set has a single S_I value.

Of particular interest is the finding that when a cyclic group is used as one of the pair, S_I values for branched alkyl groups [$(CH_3)_2CH(CH_2)_k$] (k = 1–3) exceed those for the corresponding straight-chain alkyl groups. Another fascinating aspect of the data in Table 1 is that, among the alkyl groups examined, the cyclohexyl group most closely resembles the phenyl group. These results clearly indicate that S_I is specific for the three-dimensional shape of groups (G^1 and G^2).

Correlation of S_I with Selectivity in Oxidation of a Pair of Associating Thiols.—We have been studying factors controlling specific molecular recognition such as is achieved in biological systems, using a pair of associating model compounds (**1** and **2**), namely, acylurea derivatives and open-chain analogues of pyrimidine bases (*e.g.*, uracil and thymine).[2b] Thiols **1** and **2** each have three sites: (*i*) the reaction site (SH group) where a model reaction takes place, (*ii*) the binding site [–C(=O)NHC(=O)NH–, acylurea bond] the inner –NHC(=O)– unit of which participates in two NH ··· O intermolecular hydrogen bonds[14] which extend in the opposite direction to each other, and (*iii*) the recognition site (R^1 or R^2) that participates in the discrimination.

As a model reaction, oxidation with oxygen was chosen in connection with the correct pairing of half-cystine residues in proteins,† namely, specific S–S bond formation. Oxidation of a 1:1 mixture of **1** and **2** with oxygen in the presence of a catalytic amount of triethylamine gives one unsymmetrical (**4**) and two symmetrical disulphides (**3** and **5**) (Scheme 1). The selectivity (r), a measure of the degree of the recognition of **1** by **2** (or of **2** by **1**), is defined as the logarithmic quotient of the yield of

---

* S_I values for pair (1) are 0.55 (method 1) and 0.53 (method 2), those for pair (2) are 0.82 (method 1) and 0.83 (method 2).

† The correct pairing of half-cystine residues has been suggested to depend upon specific non-covalent bonds.[15]

Table 1 S_I values for 105 pairs of groups including straight-chain (C_1 to C_6), branched (C_4 to C_6), and cyclic alkyl groups, and a phenyl group[a]

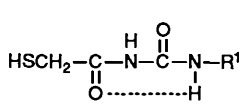
	C_1	C_2	n-C_3	n-C_4	n-C_5	n-C_6	i-C_4	i-C_5	i-C_6	neo-C_5	c-C_3	c-C_5	c-C_6	Ph
C_1	1.00	0.62	0.45	0.36	0.29	0.25	0.36	0.29	0.25	0.29	0.52	0.33	0.27	0.30
C_2	0.62	1.00	0.72	0.57	0.47	0.40	0.57	0.46	0.40	0.47	0.66	0.48	0.44	0.46
n-C_3	0.45	0.72	1.00	0.77	0.64	0.54	0.78	0.64	0.55	0.64	0.75	0.61	0.61	0.60
n-C_4	0.36	0.57	0.77	1.00	0.83	0.70	0.66	0.83	0.70	0.55	0.60	0.59	0.59	0.60
n-C_5	0.29	0.47	0.64	0.83	1.00	0.85	0.56	0.71	0.85	0.49	0.50	0.56	0.54	0.53
n-C_6	0.25	0.40	0.54	0.70	0.85	1.00	0.48	0.62	0.75	0.43	0.43	0.48	0.49	0.48
i-C_4	0.36	0.57	0.78	0.66	0.56	0.48	1.00	0.81	0.69	0.81	0.63	0.61	0.66	0.67
i-C_5	0.29	0.46	0.64	0.83	0.71	0.62	0.81	1.00	0.85	0.68	0.52	0.60	0.65	0.63
i-C_6	0.25	0.40	0.55	0.70	0.85	0.75	0.69	0.85	1.00	0.59	0.45	0.51	0.58	0.56
neo-C_5	0.29	0.47	0.64	0.55	0.49	0.43	0.81	0.68	0.59	1.00	0.52	0.60	0.66	0.60
c-C_3	0.52	0.66	0.75	0.60	0.50	0.43	0.63	0.52	0.45	0.52	1.00	0.55	0.51	0.53
c-C_5	0.33	0.48	0.61	0.59	0.56	0.48	0.61	0.60	0.51	0.60	0.55	1.00	0.67	0.69
c-C_6	0.27	0.44	0.61	0.59	0.54	0.49	0.66	0.65	0.58	0.66	0.51	0.67	1.00	0.79
Ph	0.30	0.46	0.60	0.60	0.53	0.48	0.67	0.63	0.56	0.60	0.53	0.69	0.79	1.00

[a] Abbreviations: C_1, methyl; C_2, ethyl; n-C_3, n-propyl; n-C_4, n-butyl; n-C_5, n-pentyl; n-C_6, n-hexyl; i-C_4, isobutyl; i-C_5, isopentyl; i-C_6, isohexyl; neo-C_5, neopentyl; c-C_3, cyclopropyl; c-C_5, cyclopentyl; c-C_6, cyclohexyl. See also Table 2.

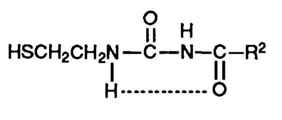
Table 2 Seven sets of groups each giving the same S_I value for a given group[a]

Set No.	Representative group	Other groups
1	n-propyl	isopropyl
2	n-butyl	*sec*-butyl
3	isobutyl	*tert*-butyl
4	n-pentyl	1-ethylpropyl, 1-methylbutyl
5	isopentyl	1,1-dimethylpropyl, 1,2-dimethylpropyl, 2-methylbutyl
6	n-hexyl	1-ethylbutyl, 1-methylpentyl
7	isohexyl	1,1-dimethylbutyl, 1,3-dimethylbutyl, 1-ethyl-2-methylpropyl, 2-methylpentyl

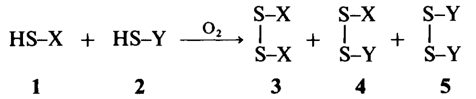
[a] Seven sets of groups in this table each give the same molecule. In Table 1, these sets of groups are each represented by the group in column 1 for simplicity.



1a $R^1 = Bu^n$
1b $R^1 = Ph$



2a $R^2 = (CH_2)_j CH_3$
2b $R^2 = (CH_2)_k CH(CH_3)_2$
2c $R^2 = cyclo\text{-}C_6H_{11}$



Scheme 1

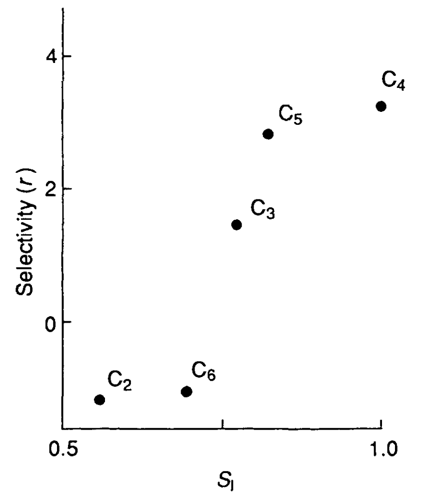


Fig. 3 Plot of the selectivity (r) in oxidation of **1a** and **2a** [at 35 °C in water–acetonitrile ($x_w = 0.42$)] against the S_I for the pair n-C_4H_9–$CH_3(CH_2)_j$ (j = 1–5). The carbon number in a straight-chain alkyl group (R^2) is shown in this figure for simplicity. Errors (three times the standard deviations) for r values range from ±0.04 to ±0.09 except for R^2 = n-C_3H_7 (±0.14) and n-C_6H_{13} (± 0.13).

unsymmetrical disulphide **4** to twice that of symmetrical disulphide **3** [eqn. (5)].

$$r = \ln\{[4]/(2[3])\} \qquad (5)$$

In Fig. 3 is plotted the selectivity (r)[16] in the oxidation of **1a** (R^1 = n-C_4H_9) and **2a** [$R^2 = (CH_2)_jCH_3$] (j = 1–5) for R^2 in water–acetonitrile [mole fraction of water (x_w) = 0.42] *versus* the S_I for the pair n-C_4H_9–G^2 ($G^2 = C_2H_5$ to n-C_6H_{13}). Both r and S_I exhibit a maximum at R^2 (= R^1) = n-C_4H_9 and at G^2 (= G^1) = n-C_4H_9, respectively, the correlation coefficient (ρ) for the r–S_I plot being 0.907. When R^2 and G^2 range from an ethyl to an n-pentyl group, the ρ value increases to 0.949 in the above solvent, and to 0.982 in water–ethanol ($x_w = 0.75$) [r = 2.16 ± 0.04 ($R^2 = C_2H_5$), 3.42 ± 0.10 (R^2 = n-C_3H_7), >6 (R^2 = n-C_4H_9),* and 3.97 ± 0.04 (R^2 = n-C_5H_{11})].

Fig. 4 shows the temperature dependence of the selectivity (r) (R^1 = Ph) as a function of the structures of R^2 in water–acetonitrile ($x_w = 0.42$). The order of the r for R^2 has proved to be the same as that of the S_I for the pair Ph–G^2, where G^2 is n-, *iso*-, or cyclohexyl.[17]† The ρ value for the r–S_I plot is 0.887 at 35 °C and 0.981 at 50 °C. The data in Figs. 3 and 4 support the similarity recognition hypothesis[2] described above.

---

* The error for the r value for R^2 = n-C_4H_9 is not shown, since the yield of symmetrical disulphide **3** cannot be determined owing to the limitations (0.1%) of sensitivity of the integrator employed.

† It has been found that the same is also true for R^2 = cyclo-C_5H_9, iso-C_5H_{11} and n-C_5H_{11} (in order of decreasing r and S_I values).[18]

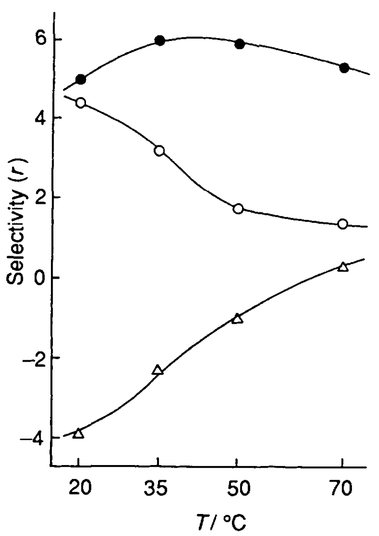
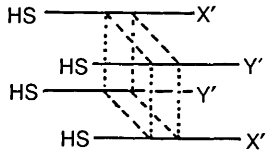


Fig. 4 Temperature dependence of the selectivity (r) in oxidation of **1b** and **2** [in water–acetonitrile (x_w = 0.42)] as a function of the structures of R^2. ●, R^2 = cyclo-C_6H_{11}; ○, R^2 = iso-C_6H_{13} [$(CH_2)_3CH(CH_3)_2$]; △, R^2 = n-C_6H_{13}. Errors (three times the standard deviations) for r values range from ±0.05 to ±0.23 except at 70 °C for R^2 = iso-C_6H_{13} (±0.37).



6

Fig. 5 Association scheme of a typical tetramer **6** formed by dimerisation of dimers: ----, hydrogen bonding responsible for the stabilisation of dimers; · · · ·, non-covalent weak interactions responsible for the stabilisation of tetramers. The symbols HS–X′ and HS–Y′ represent thiols **1** and **2**, respectively.

The Selectivity and Other Possible Factors Affecting It.—It has been shown that the oxidation proceeds through tetramers (reaction intermediates) such as **6** (Fig. 5),[2b,14a] which are formed by dimerisation of dimers produced from **1** and **2** through two NH···O intermolecular hydrogen bonds between the inner -NHCO- units in the -C(=O)NHC(=O)NH- groups,* and that the product ratio in this type of oxidation is kinetically controlled.[11]

The findings shown in Figs. 3 and 4 cannot be explained in terms of the following factors which influence chemical selectivity. Firstly, electronic effects of R^2 fail to account for the selectivity data, because the order of σ_P values[19] for R^2 [−0.15 (C_2H_5), −0.13 (n-C_3H_7), −0.16 (n-C_4H_9), and −0.15 (n-C_5H_{11})] differs sharply from that of the corresponding r values (Fig. 3).† Secondly, hydrophobic effects appear not to be responsible for the observed selectivity, since hydrophobicity increases progressively with increasing temperature[20] and π values‡ increase progressively with increasing number of carbon atoms[22] in straight-chain alkyl groups (π being a measure of hydrophobicity).§ Thirdly, steric effects cannot explain the selectivity data, because the order of ν values¶ for R^2 [0.56 (C_2H_5), 0.68 (n-C_3H_7), 0.68 (n-C_4H_9), 0.68 (n-C_5H_{11}), 0.73 (n-C_6H_{13}), 0.68 (iso-C_6H_{13}), and 0.87 (cyclo-C_6H_{11})][24] differs essentially from that of the corresponding r values (Figs. 3 and 4). Fourthly, the reactivity difference between thiols **1** and **2** is not responsible for the observed selectivity. This is because r values should become negative regardless of the structures of R^1 and R^2; if the selectivity depends upon the reactivity difference however, r shows large positive values (Figs. 3 and 4).

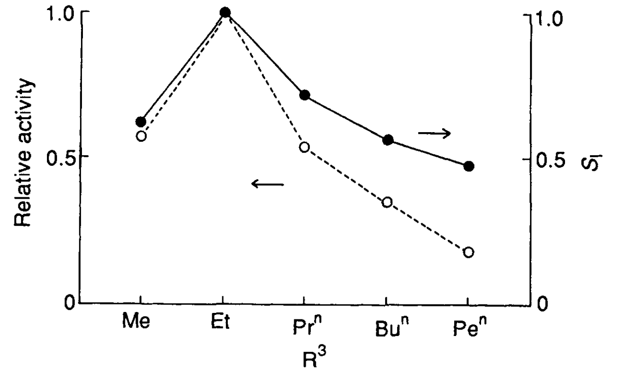
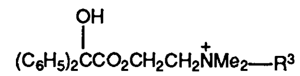


Fig. 6 The dependence (i) of relative activity (○) of benzilic esters **7** each having a straight-chain alkyl group as one of N-substituents (R^3) and (ii) of the S_I (●) for the pair C_2H_5–$CH_3(CH_2)_j$ (j = 0 to 4) upon the structures of R^3 (CH_3 to n-C_5H_{11}).



$$(C_6H_5)_2\overset{OH}{\underset{|}{C}}CO_2CH_2CH_2\overset{+}{N}Me_2\text{—}R^3$$

7

Correlation of S_I with Biological Activity.—The structure–activity relationship has been studied extensively and systematically.[25] Amoore has demonstrated that the more closely the molecular shape of a compound resembles that of benzaldehyde the more similar the compound is to benzaldehyde in odour (ρ = 0.90).[26] In an attempt to find a correlation of relative activities of bioactive substances with the shape similarity between groups involved in the substances, our attention was devoted to the structure dependence of an atropine-like activity[27] (*i.e.*, effect on cat blood pressure) of benzilic esters **7** each having a straight-chain alkyl group (C_1–C_5) as one of N-substituents (R^3). S_I was calculated for the pair C_2H_5–$CH_3(CH_2)_j$ (j = 0–4), an ethyl group giving the highest activity to esters **7**. As can be seen from the data in Fig. 6, S_I correlates closely with the relative activity of **7** (ρ = 0.973).

The relative activity, it seems, cannot be interpreted in terms of the hydrophobic,[22] electronic,[19] and steric[24] effects.‖ The new parameter, S_I, will be useful for a better understanding of specific and precise molecular recognition in chemical and biological systems and of relative biological activity of a homologous series of compounds, especially those showing a maximum for a given member.

Experimental

Molecules and Conformations.—The atomic coordinates for molecules were either determined using the data for bond distances and bond angles (Table 3) or taken from the Cambridge data base. In Table 4 are listed molecules, the

---

[*] Relative concentrations of tetramers are considered to control the selectivity; for more detailed discussions, see ref. 31.

[†] Since σ_P values are supposed to increase in the order iso-C_6H_{13} < n-C_6H_{13} < cyclo-C_6H_{11}, r values in Fig. 4 cannot be explained by electronic effects of R^2.

[‡] π is defined as: $\pi = \log P_X - \log P_H$, where P_H is the partition coefficient of a parent molecule (between octan-1-ol and water) and P_X is that for a derivative.[21]

[§] One reviewer expressed doubts about using hydrophobicity indices in solvent mixtures.

[¶] ν is defined from $\nu_X = r_{VX} - r_{VH} = r_{VX} - 1.20$, where r_{VX} and r_{VH} are the van der Waals radii of the X group and the hydrogen atom, respectively.[23]

[‖] The σ_P and ν values for a methyl group are reported to be −0.17 and 0.52, respectively.[19,24]

Table 3 Bond distances/Å or bond angles/°[a]

Compound	Bond distance	Bond angle[b]	Method[c]
Methane	C–H	1.0870	MW
Ethane	C–C	1.5351	MW
	C–H	1.0940	
	∠CCH	111.17	
Propane	C–C	1.532	ED
	C–H	1.107	
	∠CCC	112	
	∠HCH	107	
n-Butane	C–C	1.531	ED
	C–H	1.117	
	∠CCC	113.8	
	∠CCH	111.0	
Neopentane	C–C	1.537	ED
	C–H	1.114	
	∠CCH	112	
Cyclopropane	C–C	1.512	R
	C–H	1.083	
	∠HCH	114.0	
Cyclopentane	C–H	1.114	ED
	C–C	1.546	
	∠CCH	111.7	
Cyclohexane	C–C	1.536	ED
	C–H	1.119	
	∠CCC	111.3	
Benzene	C–C	1.399	ED
	C–H	1.101	IR

[a] *Kagaku Binran*, 3rd edn., Chemical Society of Japan, Maruzen, Tokyo, 1984. The data in this book are mainly taken from *Tables of Interatomic Distances and Configuration in Molecules and Ions* (Special Publication), Nos. 11 and 18, ed. L. E. Sutton, The Chemical Society, London, 1958 and 1965. [b] Bond angles not shown in this table were calculated in the usual way. [c] MW, microwave spectroscopic analysis; ED, electron diffraction; R, Raman spectroscopic analysis; IR, infrared spectroscopic analysis.

structural data of which were used for calculating S_I for the related groups. The longest straight chain in acyclic hydrocarbons is assumed to be antiperiplanar in conformation. Cyclopentane and cyclohexane are assumed to exist in planar and chair conformations, respectively.

Table 4 Molecules, structural data for which are used for calculating S_I for related groups

Group	Molecule
Methyl	Methane
Ethyl	Ethane
n-Propyl	Propane
n-Butyl	n-Butane
n-Pentyl	n-Butane
n-Hexyl	n-Butane
Isobutyl	n-Butane
Isopentyl	n-Butane
Isohexyl	n-Butane
Neopentyl	Neopentane
Cyclopropyl	Cyclopropane
Cyclopentyl	Cyclopentane
Cyclohexyl	Cyclohexane and n-butane[a]
Phenyl	Benzene

[a] As the ∠CCH for a cyclohexyl group, that for an n-butane is used.

Procedure for Determination of S_I.—Two groups (G^1 and G^2) are superimposed in such a way that there is maximum overlap between the two. This arrangement is termed *initial arrangement* $(1)_1$ [$IA(1)_1$]. A tentative S_I (^tS_I) value is calculated for the $IA(1)_1$ at $I = 0.5$ Å.* Starting from this arrangement, one of the paired groups is moved a little way along the *X*-, *Y*- or *Z*-axis, or rotated a little around the three respective axes, with the other group being fixed. These deviations give twelve new arrangements, for which ^tS_I values are calculated. Of these twelve, an arrangement affording a maximal ^tS_I value is termed $IA(1)_2$. Similar treatment yields the $IA(1)_3$, $IA(1)_4$, $IA(1)_5$, ··· and $IA(1)_n$. The calculation is performed until a ^tS_I value for $(1)_{n+1}$ is smaller than that for the $IA(1)_n$. This implies that a ^tS_I value for the $IA(1)_n$ is maximal among ^tS_I values for a set of arrangements $IA(1)_1$ to $IA(1)_{n+1}$. Similarly, for all other initial arrangements [$IA(2)_1$ to $IA(N)_1$], maximal ^tS_I values are determined which correspond to the arrangements $IA(2)_n$ to $IA(N)_n$, each having a possibility of giving the largest overlapping between G^1 and G^2. Of all the arrangements [$IA(1)_n$ to $IA(N)_n$], the arrangement which affords the largest ^tS_I value ($I = 0.5$ Å) is regarded as the *final arrangement*. Then, the *final* S_I value is calculated for this arrangement at $I = 0.2$ Å.*

Computation.—A computer program (SIMIL) implementing this method for calculating the S_I was written in Fortran 77. S_I was calculated on NEC ACOS 950 and SX-1EA systems at the Information Science Research Center, Aoyama Gakuin University, Japan.

General Procedures.—^1H NMR spectra were recorded with a JEOL GX-270 spectrometer. Chemical shifts (δ) are reported downfield from internal $SiMe_4$. The mass spectra were taken on a Hitachi RMU-6M spectrometer and exact mass data on a JEOL JMS-DX303 mass spectrometer. M.p.s were determined on a Yamato oil-immersion apparatus and are uncorrected. HPLC separations were conducted on a Waters Model 204 system including a UV detector attached to a Waters 740 Data Module (integrator).

Materials.—Water was purified through a Millipore Milli-Q water purification system followed by distillation. Acetonitrile was purified by distillation from CaH_2 and then from P_2O_5. Ethanol was purified by distillation from magnesium alkoxides.

Thiol **1** was prepared by reaction of the corresponding thioester [$MeC(=O)SCH_2C(=O)NHC(=O)NH-R^1$] with cysteamine[28] as described previously.[16,29]

Thiol **2** was prepared by addition of the corresponding acyl isocyanate to freshly sublimed cysteamine in tetrahydrofuran (THF) under argon at 0 °C as described previously,[30] their properties having already been reported.[16,29]

Preparation of Disulphides.—Symmetrical disulphides **3** were easily obtained by treatment of **1** with O_2 in the presence of Et_3N in MeCN at room temperature;[30] their properties have already been reported.[16,31]

Unsymmetrical disulphides **4** were prepared either by repeated recrystallisation of the oxidation mixture in cases where $r \geqslant 1$ or by reaction of a disulphide [$2,4-(NO_2)_2C_6H_3S-SCH_2CH_2NHC(=O)NHC(=O)-R^2$] with thiol **1** in the presence of silver acetate in dimethylformamide under argon as described previously[32] followed by recrystallisation of the reaction mixture in cases where $r < 1$. Disulphides **4** had the following properties [^1H NMR (270 MHz) spectra were measured in $(CD_3)_2SO$ at 0.01 mol dm^{-3}. Coupling constants are given in Hz].[16,33]

(**4bb**) (X^1 = Ph and Y^2 = iso-C_6H_{13}): m.p. 152.5–153.5 °C (acetonitrile); δ_H 0.83 (6 H, d, *J* 6), 1.07–1.15 (2 H, m), 1.44–1.53 (3 H, m), 2.25 (2 H, t, *J* 7, $COCH_2CH_2$), 2.80–3.00 (4 H, m,

---

* See page 134, footnote ||.

SCH_2CH_2N), 3.31–3.48 (2 H, m, SCH_2CO), 7.09–7.11 (1 H, m), 7.30–7.35 (2 H, m), 7.50–7.54 (2 H, m), 8.59 (1 H, s), 10.31 (1 H, s), 10.46 (1 H, s), and 10.77 (1 H, s) (Found: $M^+ + 1$, 441.1671. Calc. for $^{12}C_{19}{}^{1}H_{29}{}^{14}N_4{}^{16}O_4{}^{32}S_2$: $M^+ + 1$, 441.1631).

(4bc) (X^1 = Ph and Y^2 = cyclo-C_6H_{11}): m.p. 198.0–199.5 °C (acetonitrile); δ_H 1.09–1.78 (10 H, m), 2.25–2.35 (1 H, m, COCH), 2.82–2.94 (2 H, m, SCH_2CH_2N), 3.44–3.52 (2 H, m, SCH_2CH_2N), 3.71–3.80 (2 H, m, SCH_2CO), 7.05–7.13 (1 H, m), 7.30–7.36 (2 H, m), 7.51–7.54 (2 H, m), 8.59 (1 H, s), 10.24 (1 H, s), 10.32 (1 H, s), and 10.81 (1 H, s) (Found: $M^+ + 1$, 439.1495. Calc. for $^{12}C_{19}{}^{1}H_{27}{}^{14}N_4{}^{16}O_4{}^{32}S_2$: $M^+ + 1$, 439.1474).

Oxidation of a Pair of Thiols.—A mixture of **1** (0.50 mmol) and **2** (0.50 mmol) in solvent (12.5 cm^3) was stirred vigorously under oxygen for 15 min, in a well-stirred water bath thermostatted to ±0.1 °C for 20–50 °C and ±0.5 °C for 70 °C. To this mixture was added Et_3N (0.05 mmol), and vigorous stirring was continued for the time required to complete the oxidation (the oxidation was performed at least twice under the same conditions). When the oxidation was complete, the reaction mixture was evaporated to dryness. The yields of **3** and **4** were determined from their absorptions at 254 nm after separation of the three disulphides **3**–**5** in the mixture by HPLC [LiChrosorb CN with n-hexane–isopropyl alcohol: 95:5 (R^1 = Ph and R^2 = n-C_6H_{13}, iso-C_6H_{13}, and cyclo-C_6H_{11}) and 94:6 (R^1 = n-C_4H_9 and R^2 = C_2H_5 to n-C_6H_{13}) as the eluent]. The *r* values given in the text and in Figs. 3 and 4 represent the mean values of two or more experiments, and were reproducible to within the errors shown. Errors in *r* values shown in the text and figure captions are three times the standard deviations.

Acknowledgements

We thank Professor Yuji Ohashi for helpful discussions, Mr. Akihiko Kusai (JEOL) for measuring high-resolution mass spectra, and Mr. Toshio Nagata for valuable assistance.

References

1 K. Suslick, B. Cook and M. Fox, *J. Chem. Soc., Chem. Commun.*, 1985, 580; I. Yamaguchi, T. Joh and S. Takahashi, *J. Chem. Soc., Chem. Commun.*, 1986, 1412; B. R. Cook, T. J. Reinert and K. S. Suslick, *J. Am. Chem. Soc.*, 1986, **108**, 7281; N. Herron, G. D. Stucky and C. A. Tolman, *J. Chem. Soc., Chem. Commun.*, 1986, 1521; P. R. Carlier, W. S. Mungall, G. Schröder and K. B. Sharpless, *J. Am. Chem. Soc.*, 1988, **110**, 2978.

2 (*a*) T. Endo, H. Tasai, K. Miyazawa, M. Endo, K. Kato, A. Uchida, Y. Ohashi and Y. Sasada, *J. Chem. Soc., Chem. Commun.*, 1983, 636; (*b*) T. Endo, *Top. Curr. Chem.*, 1985, **128**, 91.

3 A. Verloop, W. Hoogenstraaten and J. Tipker, in *Drug Design*, ed. E. J. Ariëns, Academic Press, New York, 1976, vol. VII, p. 165.

4 I. Motoc, *Top. Curr. Chem.*, 1983, **114**, 93.

5 Z. Simon and Z. Szabadai, *Stud. Biophys.*, 1973, **39**, 123.

6 M. Randić and C. L. Wilkins, *J. Chem. Inf. Comput. Sci.*, 1979, **19**, 31.

7 R. Carbó, L. Leyda and M. Arnau, *Int. J. Quantum Chem.*, 1980, **17**, 1185.

8 P. E. Bowen-Jenkins, D. L. Cooper and W. G. Richards, *J. Phys. Chem.*, 1985, **89**, 2195; P. E. Bowen-Jenkins and W. G. Richards, *J. Chem. Soc., Chem. Commun.*, 1986, 133; E. E. Hodgkin and W. G. Richards, *J. Chem. Soc., Chem. Commun.*, 1986, 1342.

9 J. E. Amoore, G. Palmieri, E. Wanke and M. S. Blum, Science, 1969, **165**, 1266.

10 N. L. Allinger, *Pharmacol. Future Man, Proc. Int. Congr. Pharmacol., 5th*, 1972, **5**, 57.

11 T. Endo, A. Kuwahara, H. Tasai, T. Murata, H. Hashimoto and T. Ishigami, *Nature (London)*, 1977, **268**, 74.

12 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441.

13 T. R. Stouch and P. C. Jurs, *J. Chem. Inf. Comput. Sci.*, 1986, **26**, 4.

14 (*a*) T. Endo, Y. Takeda, T. Orii, Y. Kaneko and M. Kondo, *Chem. Lett.*, 1979, 1455; (*b*) A. Uchida, Y. Ohashi, Y. Sasada, Y. Kaneko and T. Endo, *Acta Crystallogr., Sect. C*, 1984, **40**, 115; (*c*) ref. 2(*b*).

15 C. B. Anfinsen, M. Sela and J. P. Cooke, *J. Biol. Chem.*, 1962, **237**, 1825.

16 T. Endo, T. Ueda, T. Isago, J. Kato, T. Suzuki, M. Misaka and M. M. Ito, *Bull. Chem. Soc. Jpn.*, 1987, **60**, 2298.

17 See also T. Endo, K. Tajima, M. Yamashita, M. M. Ito, J. Nishida and T. Ogikubo, *J. Chem. Soc., Chem. Commun.*, 1986, 1561.

18 T. Endo, unpublished results.

19 C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani and E. J. Lien, *J. Med. Chem.*, 1973, **16**, 1207.

20 G. Némethy and H. A. Scheraga, *J. Chem. Phys.*, 1962, **36**, 3401.

21 T. Fujita, J. Iwasa and C. Hansch, *J. Am. Chem. Soc.*, 1964, **86**, 5175.

22 C. Hansch, A. R. Steward, J. Iwasa and E. W. Deutsch, *Mol. Pharmacol.*, 1965, **1**, 205.

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

24 M. Charton, *Top. Curr. Chem.*, 1983, **114**, 57.

25 *Drug Design*, ed. E. J. Ariëns, Academic Press, New York, 1971–1980, vol. I–X.

26 J. E. Amoore, *Nature (London)*, 1971, **233**, 270.

27 H. R. Ing, *Fortschr. Arzneimittelforsch.*, 1964, **7**, 305.

28 T. Endo, K. Oda and T. Mukaiyama, *Chem. Lett.*, 1974, 443.

29 J. Kato, M. M. Ito, K. Hagino, N. Naitoh, M. Misaka, E. Fujishiro and T. Endo, *Bull. Chem. Soc. Jpn.*, 1988, **61**, 1419.

30 T. Endo, A. Okubo, Y. Kaneko, M. Uehara, H. Tasai, A. Sato, K. Nikki, N. Nakagawa and S. Kamei, *Bull. Chem. Soc. Jpn.*, 1982, **55**, 2224.

31 T. Endo, H. Takei, T. Isago, J. Kato, M. M. Ito, K. Guro, K. Akiyama, Y. Nakajima and H. Kataoka, *J. Am. Chem. Soc.*, 1989, **111**, 8166.

32 T. Endo, H. Tasai and T. Ishigami, *Chem. Lett.*, 1975, 813.

33 T. Endo, H. Takei, T. Isago, K. Guro, Y. Nakajima, J. Kato and M. M. Ito, *J. Chem. Soc., Perkin Trans. 2*, 1990, 1029.

Paper 0/01325K
Received 28*th March* 1990
Accepted 22*nd August* 1990