tetraeneN$_4$)NH$_3$ or between the cyano analogues. Further to this point, the labilities of the ammine complexes may be compared to: (a) $k_{exch} > 8 \times 10^6$ s^{-1} estimated for Co([14]aneN$_4$)(CH$_3$CN)$_2^{2+}$ by S. F. Lincoln and R. J. West, *Aust. J. Chem.*, **27**, 97 (1974); and (b) $k_{exch} > 10^4$ s^{-1} for Co(Me$_6$[14]4,11-dieneN$_4$)S$_2^{2+}$ (S = H$_2$O, CH$_3$OH, CH$_3$CN) and $k_{exch} > 10^5$ s^{-1} for Co([14]py(diene)N$_4$)(DMF)$_2^{2+}$ by L. L. Rusnak and R. B. Jordan, *Inorg. Chem.*, **15**, 709 (1976).
(56) J. F. Endicott, D. J. Halko W. M. Butler, and M. D. Glick, manuscript in preparation.
(57) From a variational treatment (see Appendix I): $\beta = \langle \chi_X | H | \chi_M \rangle$, $S = \langle \chi_X | \chi_M \rangle$, $\alpha_1 = \langle \chi_X | H | \chi_X \rangle$, $\alpha_2 = \langle \chi_L | H | \chi_L \rangle$, $\delta = \langle \chi_M | H | \chi_L \rangle$.
(58) E. M. Shustrovich, M. A. Porai-Koshits, and Y. A. Buslaev, *Coord. Chem. Rev.*, **17**, 1 (1975).
(59) A. R. Rossi and R. Hoffman, *Inorg. Chem.*, **14**, 365 (1975).
(60) D. H. Aue, H. M. Webb, and M. T. Bowers, *J. Am. Chem. Soc.*, **98**, 311 (1976); (b) *ibid.*, **98**, 318 (1976).
(61) J. E. Huheey, "Inorganic Chemistry", Harper and Row, New York, N.Y., 1972, p 220.
(62) (a) D. P. Rillema and J. F. Endicott, *Inorg. Chem.*, **15**, 1459 (1976); (b) B. Durham and J. F. Endicott, work in progress; (c) B. Durham, J. F. Endicott, and D. P. Rillema, 172nd National Meeting of the American Chemical Society, San Francisco, Calif., Aug. 1976, INOR 168.
(63) "Oxidants" used were Cl_2^-,[64] Br_2^-,[64] I_2^-,[64] CH_3,[27e] $CoCl^{2+}$,[62b] $MnCl^{2+}$,[62b] and $FeCl^{2+}$.[62b]
(64) S. D. Malone and J. F. Endicott, *J. Phys. Chem.*, **76**, 2223 (1972).
(65) For a discussion of the reactions limited by the rate of metal–water bond breaking see ref 8.
(66) Owing to the smaller M–Cl than M–O force constants the contributions of long Co–Cl–M bonds are relatively small.
(67) See also: (a) H. Taube, "Electron Transfer Reactions of Complex Ions in Solution", Academic Press, New York, N.Y., 1970; (b) A. Haim, *Acc. Chem. Res.*, **8**, 264 (1975).

Molecular Transforms: a Potential Tool for Structure–Activity Studies[1]

Leonard J. Soltzberg[2] and Charles L. Wilkins*

Contribution from the Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska 68588. *Received July 26, 1976*

Abstract: A chemical pattern recognition approach to the study of structure-activity relationships has been investigated. In this work, a generalized molecular transform function has been examined as a possible nonprejudicial structure encoding scheme. For the set of 114 tranquilizers and 72 sedatives used in this study, such a procedure is shown to be efficacious. These results are compared with those of a previous study, using substantially the same set of compounds, in which molecular fragment coding was used, and the present method is shown to be essentially as effective for the dichotomy of the tranquilizers and sedatives examined. Examination of the patterns generated suggests the possibility of extracting geometric "prototypes" for each activity class.

The objective of correlating molecular structure with chemical activity is one of the fundamental driving forces of chemical inquiry. In organic chemistry, structure-activity relationships have been successfully pursued by Hammett analysis, in the empirical domain, and by theoretical constructs, such as orbital symmetry rules.

In pharmacology, more elusive structure-activity relationships have also been pursued at both the empirical and the theoretical levels.[3] The complexity of metabolic processes virtually guarantees complexity in the relationship between the structure of a pharmacologic agent and the observable reaction of the recipient organism. Since the agent affects a system involving not one chemical reaction but several coupled reactions occurring in a physically heterogeneous medium, the relation between modifications in structure and the corresponding modifications in activity may well be obscure. It is in such situations, where the existence of a complex but genuine relationship is suspected, that the methods of pattern recognition can be most useful.

Chemical applications of pattern recognition have been extensively discussed in recent literature.[4] In pattern recognition analysis, a complex relationship within a group of patterns (representing chemical compounds, in the present case) is reduced to a readily understood measure such as nearness in a multidimensional space or to some other similarity measure based on multidimensional representation of the patterns. For reference, we contrast this approach with that of Hansch analysis and of Free-Wilson analysis, which typify two major avenues for current pharmacological structure-activity study.

One can view the search for structure-activity relationships very generally as a quest for functions of the form

$$\text{activity} = f(\text{molecular features})$$

Here, the definitions of the activity, of the functional relationship f (not necessarily an explicit analytic formula), and of the nature of the molecular features employed characterize any particular approach to the problem. In Hansch analysis, the biological activity is related to a dose level required to produce a standard effect; the function f is generally a polynomial of the second degree; and the molecular features are physicochemical data including octanol-water partition coefficients, Hammett-type ρ-σ electronic parameters, and, in some cases, a steric term.[5] The Free-Wilson method also defines activity on a quantitative scale of response, such as LD_{50}; here, the function f is a linear additive combination of "substituent contribution" terms; and the features themselves are simply the identities of the various substituents on some parent compound.[6] Both of these methods begin with a specific active structural nucleus, for which the magnitude of the activity can be modified by modifying various substituents. In contrast, the pattern recognition approach to structure-activity correlation takes the broader (and, perhaps, more naive) view that a diverse group of materials with similar *qualitative* biological activity may possess some common set of molecular characteristics which are responsible for the activity. In this case, the biological activity is defined as a qualitative type of action (for example, sedation, analgesia); the function f is a pattern discriminator capable of recognizing materials of a particular activity class; and the molecular features can be either physicochemical data or items from a predetermined list of descriptors, including fragment identities, molecular weight, topological features, and so forth. The first stage in such an analysis is to create a pattern classifier which can

recognize the activity class of a molecule on the basis of the given features. That such activity recognition can be achieved has already been demonstrated.[7,8] The second stage is to identify, by examining the classifier itself, those molecular characteristics which are responsible for the activity. This task has been attempted,[7] but has yet to be convincingly accomplished.

The purpose of the present work is to explore a novel source of physicochemical data which may be useful in seeking structure-activity relationships involving primarily molecular geometry. The relevance of the data to activity class is established by means of pattern recognition.

The Molecular Transform

A starting point for any structure-activity study must be the development of an adequate algorithm for encoding molecular structure.[1,9] Thus, it was our initial objective to find, for use in pattern recognition analysis, a way to describe molecules which would avoid the possible subjectivity of fragment-based descriptor lists; this point will be discussed further. Although infrared, nuclear magnetic resonance, and mass spectra have been employed as data sources for pattern recognition studies aimed at organic analysis,[4] mass spectra have been the only physicochemical source used for pharmacologic pattern recognition.[10]

Our original intent was to use x-ray structure factor data to characterize molecules, since accurate x-ray data contains detailed information about molecular shape and electron distribution. However, the structure factor data also contains information, irrelevant from the viewpoint of this work, regarding the crystal environment and derives from a specific orientation of the molecule with respect to the unit cell; thus, polymorphs of the same compound would have different structure factor tables. This difficulty led us to consider gas-phase electron diffraction, in which the data arise from individual molecules in all possible orientations; however, electron diffraction data for compounds of pharmacologic interest is not generally available. Nonetheless, one can compute a generalized scattering function from a *known* molecular structure. Such a function has been termed a "molecular transform" [11] and can be used as the functional basis for deriving the analytic scattering relationships of both x-ray and electron diffraction.

For the purposes of the present investigation, we computed molecular transform-like functions from x-ray derived three-dimensional atomic coordinates by means of a modified version of the equation used in electron diffraction studies for preparing theoretical scattering curves.

The general molecular transform is

$$G(\mathbf{S}) = \sum_{i=1}^{N} f_i \exp(2\pi i \mathbf{r}_i \cdot \mathbf{S}) \tag{1}$$

This represents the scattering in various directions $\mathbf{S}$ by a collection of N spherical scatterers located at points $\mathbf{r}_i$; the form factors f_i take into account the direction dependence of scattering from a spherical body of finite size. As presently employed in electron diffraction analysis, this relationship appears in a form based on that originally proposed by Wierl:[12]

$$I(s) = K \sum_{i=2}^{N} \sum_{j=1}^{i-1} f_i f_j \int_0^\infty P_{ij}(r) \frac{\sin sr}{sr} \, dr \tag{2}$$

Here, the independent variable s measures the scattering angle, as

$$s = 4\pi \sin(\theta/2)/\lambda \tag{3}$$

where θ is the scattering angle and λ is the wavelength of the electron beam; $I(s)$ is the intensity of the scattered radiation. The variable r represents interatomic distance, and $P_{ij}(r)$ is the probability distribution describing the vibrational variation in the distance between atoms i and j; f_i and f_j are the form factors for atoms i and j, and K is a collection of constants related to the experimental apparatus. For the present work, we have made the following simplifications:

$$K = 1 \tag{4}$$

$$f_i = Z_i \tag{5}$$

$$P_{ij}(r) = \delta(r - r_{ij}) \tag{6}$$

leading to:

$$I(s) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} Z_i Z_j \frac{\sin sr_{ij}}{sr_{ij}} \tag{7}$$

Equation 4 simply sets the experimental constants to unity, their being of no interest in the present context. Equation 5 is tantamount to assuming the atoms to be point scatterers, so that the form factors can be replaced by the atomic numbers of the atoms. With eq 6, we assume that the molecule is rigid and thus replace the vibrational distribution of distances $P_{ij}(r)$ with a δ function peaking at the average interatomic distance r_{ij}; this replacement eliminates the need for the integral over r. These simplifications do not impair the usefulness of the function in the present application, since we are interested in a unique coding of the molecular geometry, but not in the details of the molecular scattering or vibrational dynamics.

Experimental Section

In order to have a basis for comparison of the performance of pattern classifiers trained with molecular transforms, we chose a topic which had already been investigated by pattern recognition analysis. The distinction between sedative action and tranquilizing action in 66 psychotropic drugs was studied by Ting and coworkers,[10] who employed pattern classification schemes based on distance measure and on nonlinear mapping; the data utilized were low resolution mass spectra. The usefulness of that study was limited by the small number of patterns relative to the number of features used.[13,14] More recently, Stuper and Jurs have studied the same dichotomy, using a data base of 140 tranquilizers and 79 sedatives.[8] In this latter study, the patterns were characterized in terms of a predetermined list of 68 molecular descriptors including molecular weight, number of atoms of various types, and descriptors specifying the presence of specific structural moieties and topological features. Here, sufficient patterns were employed to assure nontrivial results, and the linear discriminant classifiers resulting from that work identified unknowns as tranquilizers or sedatives at the level of about 90% correct prediction.[15] In the present work, we show that the tranquilizer/sedative dichotomy can be effected in this same set of compounds using molecular transforms rather than fragment coding to characterize the substance.

From eq 7, it is evident that the computation of the molecular transform requires knowledge of all interatomic distances in a molecule. To obtain these distances, we used three dimensional atomic coordinate data, although it may indeed be possible to compute a usable molecular transform directly from x-ray structure factor data.[16] When available, coordinates from solved crystal structures were used to compute the molecular transforms. The structures for 18 of the compounds in the Stuper and Jurs study were found by means of the National Institutes of Health Chemical Information System; the data base was the Cambridge Crystallographic Data Centre Summer 1975 tape. For molecules whose solved structures were not available, coordinates were generated from those of similar "prototype" molecules with the use of a modification of the program ATCOOR[17] implemented on a Varian 620i computer with 16K core memory. For example, the coordinates of the phenothiazine ring in chlorpromazine were used as a basis for several other phenothiazine derivatives. The bond lengths, angles, and torsion angles required for this generation of coordinates were carefully selected from various sources[18,19] to produce, as nearly as possible, conformationally accurate coordinates. Suitable prototypes were not found for some of the compounds of interest, so that our final data set contained 186 compounds: 114 tranquilizers and 72 sedatives. These compounds are listed in Table I.

The use of coordinates computed in the above manner is expected

Table I. Compounds Included in Data Set[a,b]

			Tranquilizers				
1	A124	29	Mepazine	59	SKF 6333	101	Benzindopyrine
2	Acepromazine	30	Mesoridazine	60	T 412	104	IN 399
3	Aceprometazine	31	Methiomeprazine	61	Spiclomazine	105	Milipertine
4	Acetophenazine	32	Methophenazine	62	Thiethylperazine*	106	Oxypertine
5	Butaperazine	33	Methotrimeprazine	63	Thiopropazate	107	PI 11
6	Butyrylpromazine	34	Methoxypromazine*	64	Thioproperazine	108	Solpyertine
7	Carphenazine	35	Oxaflumazine	65	Thioridazide	109	Bromazepam
8	CB 1519	36	P 824	66	TPN 12	113	Cloazepam
9	CB 1658	37	P1030	67	Trifluoroperazine	117	Diazepam*
10	Chlorimipiphenine	38	Perazine	68	Trifluoroperazine sulfoxide	119	Lorazepam
11	Chlorproethazine	39	Perimetazine			121	Nitrazepam
12	Chlorpromazine*	40	Perphenazine	69	Trifluoropromazine	122	Nitrazepate
13	Chlorpromazine sulfoxide	41	Perphenazine sulfoxide	70	Triflutrimeprazine	123	Oxazepam
		42	Phenazin	71	Trimeprazine	125	Prazepam
14	CIBA 17040	43	Phenazine	72	Valeroylperazine	126	RO 5-2180
15	CPO 12	44	Pipamazine	73	WIN 13, 645-5	127	RO 53027
16	Cyamepromazine	45	Piperacetazine	88	Xanthiol	128	Sulazepam
17	Cyclophenazine	46	Piperidochlorpromazine	89	Bishomoreserpine	129	Temazepam
18	Dichlorpromazine	47	Prochlorperazine	90	Reserpedine	130	Tetrazepam
19	Dixyrazine	49	Promazine	91	Methyl-18-ketoreserpate	131	Aceperone
20	Ethylisobutrazine	50	Propiomazine	92	Raujemidine	132	AHR 1900
21	Fluorophenothiazine	51	Propiopromazine	93	Raunescine	133	FR 33
22	Fluphenazine decanoate	52	Ridazine	94	Renoxidine	134	Diphenchloxazine
23	Fluphenazine	53	RP 3300	95	Rescinnamine	135	Midaflur*
24	Fluphenazine enanthate	54	RP 4627	96	SU 5171	136	Prothipendyl
25	Heptylpromazine	55	RP 6696	97	8842	137	Trioxazine
26	Homophenazine	56	RP 9153	98	SU 10704	138	Captodiame
27	KS 33	57	SA 124	99	Raubasine	139	Phenyltoloxamine
28	MD 5051	58	SKF 5657	100	Benanserin	140	Cintriamide
			Sedatives				
1	Profenamine	20	Hexthal	40	CHI 38	62	RD 6030
2	Promethazine*	21	Hexobarbital	41	CHI 42	63	Chloral hydrate*
3	Cloxypendyl	22	Mephobarbital*	42	Clomethiazole	64	Dispranol
4	Fenoharman	23	Methabarbital*	44	ES 708	65	Ethinamate
5	Cannabigerol	24	Methitural	45	Ethinazone	66	Mebutamate
6	D 58SI	25	Methohexital	46	Glutethimide	67	Meprobamate
8	Allobarbital	26	Nealbarbitone	47	Homochlorcyclizine	68	Nisobamate
9	Alphenal	27	Pentobarbital	48	K 2004	69	Ethchlorvynol
10	Amobarbital*	28	Phenobarbital*	49	LB 50160	70	Methylpentynol
11	Aprobarbital	29	Probarbital	50	Mecloqualone	71	Petrichloral
12	Barbital*	30	Secobarbital	51	Methaqualone	72	Acetylcarbromal
13	Butalbitol	31	Talbutal	53	Oxypendyl	73	AEC
14	Butethal*	32	Thiamylal	55	Thalidomide*	74	Carbromal
15	Butallylonal	33	Thiopental	56	Ethomoxane	75	Bromisovalum
16	Cyclobarbital*	34	Vasalgin	57	Paraldehyde*	76	Ectylurea
17	Cyclopal	35	NSD 2023	59	Tricetamide	77	IPC
18	Febarbamate	36	Anileridine	60	Gaietamine	78	Valnoctamide
19	Heptabarbital*	38	CHI 21	61	RD 6020	79	Chlorethate

[a] Numbers correspond to the list of Stuper and Jurs.[8] [b] Compounds marked with * are those for which crystal structures were available.

to be less desirable than the use of coordinates from actual structure determinations. One might question whether errors introduced in the coordinate generation process might affect classification. Even with x-ray derived coordinates there is the possibility that molecular distortion induced by the demands of lattice packing might affect classification. However, tests indicated that neither of these problems was serious in the present case. To test the faithfulness of the generated coordinates in preserving molecular geometry, two molecular transforms were computed for chlorpromazine, one from crystallographic coordinates and the other from generated coordinates based on a prototype phenothiazine ring from promethazine. The two transforms were substantially the same, although there was some shifting of peaks at high s values. In another experiment, two transforms for methoxypromazine were included in a pattern recognition training set. One transform was computed from crystallographic coordinates and the other from coordinates generated from a prototype phenothiazine ring from chlorpromazine. Both patterns were correctly recognized by the linear discriminant. The effect of molecular distortion was examined by comparing the transforms of promethazine hydrochloride and promethazine hydrobromide, both computed from crystallographic coordinates. These two transforms were very similar, although noticeable differences were again present, particularly at high s values. The anions themselves were not included in the computation of the transforms nor were any hydrogen atoms. These omissions were found to have only a slight effect on the transforms, as expected from the large r_{ij}'s associated with the anions and the small Z_i for hydrogen. Finally, molecular transforms were computed for p-dichlorobenzene from the x-ray structures of both the triclinic and the monoclinic modifications. Here, the two transforms were qualitatively identical, although some differences in peak amplitude could be noticed. These observations indicate that our generated coordinates are adequate for the gross classifications we have tested in the present work. Nonetheless, the exclusive use of x-ray derived or energy-minimized coordinates might allow even cleaner separation of the categories or permit investigation of more subtle dichotomies.

After the atomic coordinates had been assembled, the molecular transform for each compound was computed according to eq 7 for 200 equally spaced s values from 1 to 31 $Å^{-1}$; this and subsequent computations were performed on an IBM 360/65 computer. The molecular transforms for two compounds are shown in Figure 1.

Reduction of the transforms to binary patterns was accomplished by identifying the zero crossings of the $I(s)$ curve (i.e., the s values

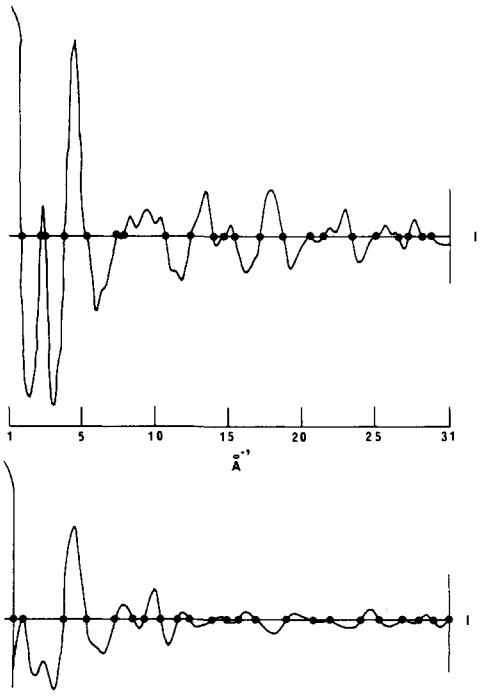

Figure 1. Molecular transforms for two of the compounds in this study, chlorpromazine (above) and barbital (below). Zero crossings are emphasized by enlarged dots. Abscissa represents the range s = 1-31 Å^{-1}.

Table II. Performance of Pattern Classifiers: Tranquilizer/Sedative Dichotomy

	No. of training patterns	No. of features	% recognition	No. of prediction errors	% prediction
Molecular Transform-Based Classifiers[a]					
Full data set	186	41	100		
"Leave-out-ten" runs					
1	176	50	100	0	
2	176	51	100	1	
3	176	48	100	0	
4	176	39	100	1	
5	176	56	100	0	
6	176	49	100	2	
7	176	43	100	3	
8	176	44	100	2	
9	176	51	100	1	
10	176	51	100	0	
Av of 10 "leave-out-ten" runs	176	48	100		90 (Z = 0)
Fragment, Environmental, and Geometric-Based Classifiers[b]					
Av of 20 "leave-out-ten" runs	209	39	100		89.5 (Z = 1.75)

[a] This work. [b] Reference 8.

at which $I(s)$ = 0) in the s range 1-31 Å^{-1}. For this purpose, the s range was divided into 100 equal intervals; a "1" was recorded in each interval containing a zero crossing, and "0" 's were recorded in all other intervals. Although amplitude data can be used in pattern recognition analysis, indications are that classifiers employing binary coding of spectra perform at least as well as those incorporating peak amplitudes.[20]

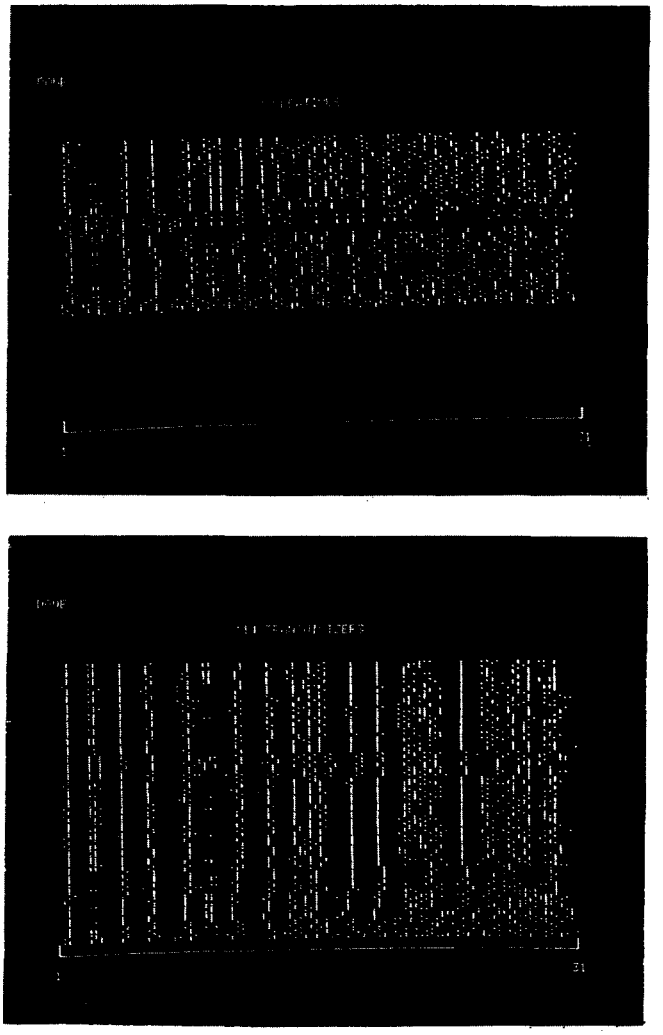

Figure 2. Graphic representation of the zero crossing binary coded molecular transforms. Each horizontal row corresponds to one compound; a dot is plotted in each interval containing a zero crossing of the molecular transform of that compound. Vertical bands suggest similarity within the class.

The 100-dimensional binary coded transforms were subjected to a pattern recognition analysis similar to that employed by Stuper and Jurs.[8] Weight vectors were developed by the error correction feedback method, with feature elimination accomplished by the weight-sign-change algorithm with initial weights equal to either 1 or −1. Each run began with 100 features, and the number of features was substantially reduced by the feature selection process. No threshold or "dead zone" was used in developing or applying the weight vectors; this corresponds to Z = 0 in the notation of Stuper and Jurs.

To discover the relevance of the molecular transforms to the tranquilizer/sedative dichotomy, two pattern recognition experiments were carried out. First, a weight vector was trained to classify the entire set of 186 compounds. Then, a series of ten "leave out ten" trials[8] was performed, in which various groups of ten patterns were left out of the training set and used to test the predictive ability of the resulting classifiers.

Results and Discussion

Table II summarizes the results of the pattern recognition tests on the tranquilizer/sedative data set when characterized by binary coded molecular transforms. The full data set was linearly separable with 41 features, and each of the 176-compound training sets was linearly separable with from 39 to 56 features. Average prediction for the 100 test compounds in the "leave out ten" sequence was 90%. This value corresponds to

ten prediction errors (seven tranquilizers and three sedatives misclassified). Because of the difference in the number of test compounds and the fact that our weight vectors were developed with $Z = 0$, these results are not strictly comparable with those of Stuper and Jurs. Nonetheless, it seems justified to conclude that the classifiers trained with molecular transforms perform at a level comparable with the performance level of those trained with the fragment, geometric, and environmental codes employed by Stuper and Jurs. Thus, the binary coded molecular transform contains sufficient information to make this particular two-class distinction about as well as was effected with a more tailored characterization of the molecules.

The step from a successful pattern classification to a structure-activity correlation is not trivial. The present case exemplifies some of the uncertainties involved. The tranquilizer/sedative dichotomy itself is not pharmacologically clear-cut. The classification used in this work is that of Stuper and Jurs,[8] which was based on a set of rules articulated in that paper; reference to the original source[21] indicates that for about 40 of their 219 compounds there is some ambiguity about the activity class. [The composition of the data set is also problematic in that two particular groups of compounds, phenothiazines (73/140) and barbiturates (27/79), predominate within the respective memberships of the tranquilizer and sedative classes.] The usefulness of a dichotomy based on such a test population has been questioned,[22] although not for the Stuper and Jurs study. The problem of incorporating information on inactive compounds has also been discussed.[8]

Beyond these problems concerning choice of topic and choice of data lie certain other fundamental questions. One such question is whether a meaningful structure-activity correlation can be derived at all from fragment code based classifiers. The difficulty here is that, by describing a molecule in terms of a predetermined list of molecular attributes, such as molecular weight, ring size, presence of carbonyl, and so forth, one may prejudice the possible results and thus draw misleading conclusions regarding which structural moieties are significant in producing activity. This type of error has already arisen in practice,[7] as Mathews has noted.[23] Some of the descriptors used in the Stuper and Jurs study are rather specific in this sense (for example, "presence of piperazinyl", "presence of dimethylamino")[8] and would be potential trouble spots if structure-activity correlations were pursued.

The use of molecular transforms is intended to obviate the problem of prejudice associated with fragment coding. Since the molecular transforms are derived purely from physical data, it is gratifying that classifiers based on them perform as well as ones based on a coding which, at least to a degree, is structured around molecular features prejudged by chemists as being likely to be of importance.

The remaining problem is how to extract from the transform-based classification scheme useful structure-activity relationships. Assuming for the moment that the tranquilizer/sedative dichotomy is pharmacologically meaningful and that the data set employed is sufficiently representative, then our pattern recognition experiments have shown that the molecular geometry as embodied in the binary coded molecular transforms is a reliable basis for distinguishing sedatives from tranquilizers. We would wish, then, to identify the geometric characteristics most responsible for the distinction. Figure 2 presents the binary coded transforms for both categories. The appearance of visually recognizable "bands" which differ between the two categories suggests the possibility of extracting a "prototype" sedative molecular transform and a "prototype" tranquilizer transform. Fourier inversion of such transforms would presumably give the interatomic distances characteristic of each prototype. Judicious examination of the feature selection made in the pattern classifier training process might further focus attention on certain geometric features important in the dichotomy; such features would, of course, have to be recognized from their characteristic patterns of interatomic distances. Ultimately, such indications would be used to guide pharmacologically motivated synthesis attempts.

Acknowledgment. This research was supported by the National Science Foundation under Grant MPS-74-01249. We thank Steven Kaberline for valuable technical assistance.

References and Notes

(1) A preliminary account of this work has appeared; L. J. Soltzberg and C. L. Wilkins, *J. Am. Chem. Soc.*, **98**, 4006 (1976).
(2) L. J. Soltzberg is a Visiting Associate Professor during the 1975–1976 academic year from Simmons College, Boston,
(3) See, for example: A. Albert, "The Selectivity of Drugs", Chapman and Hall, London, 1975.
(4) See, for example, P. C. Jurs and T. L. Isenhour, "Chemical Applications of Pattern Recognition", Wiley, New York, N.Y., 1975.
(5) W. P. Purcell, G. E. Bass, and J. M. Clayton, "Strategy of Drug Design: A Guide to Biological Activity", Wiley, New York, N.Y., 1973, pp 21–37.
(6) Reference 5, pp 89–94.
(7) B. R. Kowalski and C. F. Bender, *J. Am. Chem. Soc.*, **96**, 916 (1974).
(8) A. J. Stuper and P. C. Jurs, *J. Am. Chem. Soc.*, **97**, 182 (1975).
(9) W. E. Brugger, A. J. Stuper, and P. C. Jurs, *J. Chem. Inf. Comput. Sci.*, **16**, 105 (1976).
(10) K. H. Ting, R. C. T. Lee, G. W. A. Milne, M. Shapiro, and A. M. Guarino, *Science*, **180**, 417 (1973).
(11) H. Lipson and C. A. Taylor, "Fourier Transforms and X-Ray Diffraction", G. Bell and Sons, London, 1958, p 9.
(12) R. Wierl, *Ann. Phys. (Leipzig)*, **8**, 521 (1931).
(13) J. T. Clerc, P. Naegli, and J. Seibl, *Chimia*, **27**, 639 (1973).
(14) For a general discussion of this problem, see R. O. Duda and P. E. Hart, "Pattern Classification and Scene Analysis", Wiley, New York, N.Y., 1973, pp 69–70.
(15) Prediction capability was tested using the leave-out-ten procedure described in ref 8. A range of predictive capability from 85 to 92% correct was reported, depending on the preprocessing method used and upon the "dead zone" or prediction threshold used.
(16) Since the molecular transform as we have computed it depends only on interatomic distances and not on absolute atomic positions, it may be possible to use x-ray structure factor data directly to compute the molecular transform. The F_{obsd}^2 would be used to compute a Patterson function
$$\rho(xyz) = \frac{1}{V}\sum_h\sum_k\sum_l \left(F_{hkl}^{obsd}\right)^2 \exp(-2\pi i(hx + ky + lz),$$
a density function which has maxima at points corresponding to interatomic vectors r_{ij}. The amplitudes of the Patterson maxima are proportional to the product of the atomic numbers Z_iZ_j of the corresponding atoms. Thus the height and position of the Patterson peaks would supply the Z_iZ_j and r_{ij} information needed for the molecular transform computation. The potential difficulty would be the possible inclusion of some intermolecular distances. It might be possible to minimize this problem by using a restricted subset of the structure factor data, but such an approach would unfavorably affect the resolution of the Patterson map.
(17) J. E. Nordlander, Quantum Chemistry Program Exchange, Program Number 250. This program computes Cartesian coordinates for the atoms in a molecule from values of the bond lengths and a minimum set of bond angles and dihedral angles.
(18) "Structure Reports", Vol. 33–35, 37–39, International Union of Crystallography, Oosthoek, Scheltma, and Holkema, Utrecht, The Netherlands, 1968–1973, and references cited therein.
(19) P. J. Wheatley, *Phys. Methods Heterocycl. Chem.*, **5**, 1 (1972), and references cited therein.
(20) T. R. Brunner, C. L. Wilkins, T. F.-Lam, L. J. Soltzberg, and S. L. Kaberline, *Anal. Chem.*, **48**, 1146 (1976).
(21) E. Usdin and D. H. Efron, "Psychotropic Drugs and Related Compounds", 2nd ed, DHEW Publication No. (HSM) 72-9074, 1972.
(22) C. L. Perrin, *Science*, **183**, 551 (1974).
(23) R. J. Mathews, *J. Am. Chem. Soc.*, **97**, 935 (1975).