Functional Group Contributions to Drug-Receptor Interactions

P. R. Andrews,* D. J. Craik, and J. L. Martin

Victorian College of Pharmacy Ltd, Parkville, Victoria, Australia 3052. *Received December 5, 1983*

The binding constants and structural components of 200 drugs and enzyme inhibitors have been used to calculate the average binding energies of 10 common functional groups. As expected, charged groups bind more strongly than polar groups, which in turn bind more tightly than nonpolar groups. The derived intrinsic binding energies (in kcal/mol) are (i) charged groups, CO_2^-, 8.2; PO_4^{2-}, 10.0; N^+, 11.5; (ii) polar groups, N, 1.2; OH, 2.5; CO, 3.4; O or S ethers, 1.1; halogens, 1.3; (iii) nonpolar groups, C (sp^2), 0.7; C (sp^3), 0.8. These values may be used to determine the goodness of fit of a drug to its receptor. This is done by comparing the observed binding constant to the average binding energy calculated by summing the intrinsic binding energies of the component groups and then subtracting two entropy related terms (14 kcal/mol for the loss of overall rotational and translational entropy and 0.7 kcal/mol for each degree of conformational freedom). Drugs that match their receptors exceptionally well have a measured binding energy that substantially exceeds this calculated average value--examples include diazepam and biotin. Conversely, if the observed binding energy is very much less than the calculated average value, then the drug apparently matches its receptor less well than average. Examples of this type include methotrexate and buprenorphine.

The experimentally observed binding of a drug to its receptor provides a measure of the total interaction between the two molecules but generally tells us little or nothing of the three-dimensional quality of their interaction. Is the match as good as that of hand and glove, or more like that of square peg and round hole? Morphine, for example, binds at 5 nM and is often regarded as a prototype drug for analgesic activity, but how well does it match its receptor? Butaclamol, being comparatively rigid, is widely used as a model for dopamine receptor antagonist activity, but how many of its functional groups actually interact with the receptor? Peptides such as the enkephalins are thought to bind to their receptors via several amino acids, but how many are actually required to account for the observed binding? Oxalate anion has an inhibition constant, K_I, for trans carboxylase[1] of 1.8×10^{-6} M, while methotrexate has a K_I of 10^{-11} for dihydrofolate reductase,[2] but which represents the best match to the corresponding active site? In conformationally flexible molecules, we know that the lowest energy conformation is not necessarily the biologically active form, since part of the drug–receptor binding may be used to perturb the drug conformation, but how much above the global energy minimum must we go to be sure of including the biologically active conformation?

To answer these questions, we need to have some way of estimating the potential bond strengths involved in the interaction between a drug and a reasonably matched receptor, but mechanisms for providing such estimates are presently far from satisfactory. It is possible, for example, to calculate the strengths of intermolecular interactions at various levels of approximation, using perturbation theory, but such calculations are not only time-consuming and inaccessible to the majority of workers but also quite unreliable in aqueous solution. For these reasons, most medicinal chemists prefer the simpler alternative of using standard values of the enthalpies of formation for different types of bond (ionic, hydrogen, van der Waals, etc.) to estimate approximate strengths for drug–receptor interactions.[3,4] Again, however, there are problems with this approach, including particularly the lack of any allowance for the entropic component of the interaction, the relatively large range of energies associated with each type of bond, and the uncertainty as to which of the possible bonds associated with any drug should be included in the interaction.

To overcome these difficulties, efforts have been made to estimate the strengths of interactions involving individual functional groups of the drug. One approach, which relies on finding pairs of compounds for which the difference in binding to a receptor or enzyme may be traced solely to the contribution of a single functional group, has been developed by Page and Jencks,[5,6] who refer to it as the anchor principle. It has the major advantage that the difference in binding of a drug molecule with, and without, the particular functional group, incorporates only the factors associated with that group while excluding the loss of overall rotational and translational entropy associated with the drug molecule (the anchor). It is limited, however, by the requirement for many pairs of compounds that differ only in the presence of a single functional group and in which the role of that functional group can be traced more or less exclusively to the provision of additional binding energy, rather than, for example, to conformational enhancement of biologically active forms.

In the present paper, we have attempted to overcome this limitation by using a series of 200 drugs and enzyme inhibitors, chosen somewhat subjectively on the basis of their apparent tight binding to their corresponding receptor sites, to provide a statistical estimate of the strengths of noncovalent bonds associated with each functional group in an average drug–receptor environment. For this purpose, since drugs acting on a wide range of different receptors are being used, the anchor principle cannot be applied directly, and an allowance for the loss of overall rotational and translational entropy must therefore be included in the analysis. This number has been taken, on the basis of the work of Page,[6,7] to be 14 kcal/mol under standard conditions (see below). Also included is an allowance for the loss of internal rotational entropy on binding, since it may be assumed that the bound conformation of a drug, at least at its optimal receptor, will be relatively fixed. This term is derived empirically from the data in the same way as the intrinsic

(1) D. B. Northrop and H. G. Wood, *J. Biol. Chem.*, 244, 5820 (1969).
(2) W. T. Colwell, V. H. Brown, J. T. Degraw, and N. E. Morrison, "Chemistry and Biology of Pteridines", R. L. Kisliuk, G. M. Brown, Eds., Elsevier, Amsterdam, 1979; pp 215–218.
(3) A. Albert, "Selective Toxicity", 6th ed., Wiley, New York, 1979, p 277.
(4) A. Korolkovas, "Essentials of Molecular Pharmacology", Wiley, New York, 1970, pp 137–186.
(5) M. I. Page and W. P. Jencks, *Proc. Natl. Acad. Sci. U.S.A.* 68, 1678 (1971).
(6) M. I. Page, *Angew. Chem. Int. Ed. Engl.*, 16, 449 (1977).
(7) M. I. Page, "Quantitative Approaches to Drug Design"; J. C. Dearden, Ed.; Elsevier, Amsterdam, 1983, p 109.

0022-2623/84/1827-1648$01.50/0

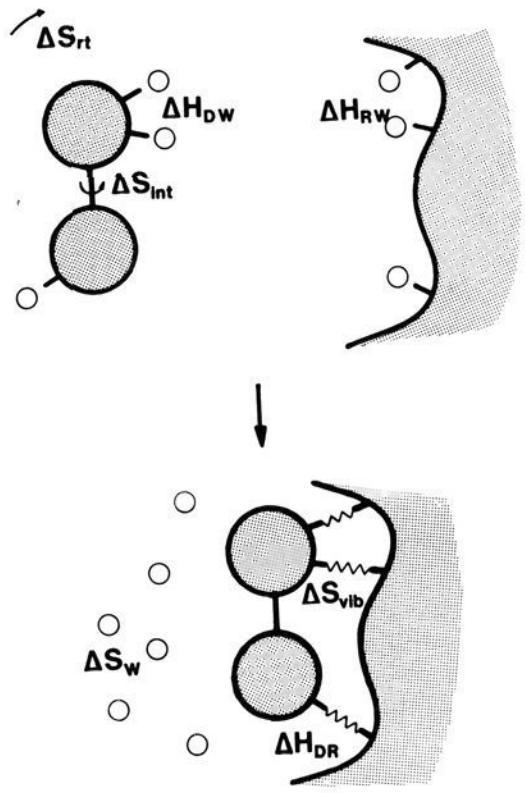


Figure 1. Binding interactions between a trifunctional drug and an optimal receptor. The small circles represent water molecules, the enthalpies of hydration of the drug and receptor being ΔH_{dw} and ΔH_{rw}, respectively. The free drug has an overall rotational and translational entropy of ΔS_{rt} and an internal entropy of ΔS_{int}. On binding, both terms are lost, but this unfavorable contribution may be compensated for by an increase in entropy due to loss of structured water on binding (ΔS_w), as well as an increase in entropy (ΔS_{vib}) due to new low-frequency vibrational modes associated with the drug–receptor noncovalent bonds. While ΔS_{rt} is essentially independent of the size of the drug (within limits), the other entropic and enthalpic terms depend on the number and nature of the functional groups present.

binding energies for individual functional groups, according to eq 1. Here, ΔG is the free energy of drug–receptor

$$\Delta G = T\Delta S_{rt} + n_{DOF}E_{DOF} + \sum n_X E_X \qquad (1)$$

binding determined from experimental binding or inhibition constants, $T\Delta S_{rt}$ is the loss of overall rotational and translational entropy of the bound drug molecule, n_{DOF} is the number of internal degrees of conformational freedom in the drug molecule, E_{DOF} is the corresponding energy associated with the change in entropy resulting from the loss of each such degree of conformational freedom, and E_X is the intrinsic binding energy associated with the functional group X, of which there are n_X present in the drug. It should be noted that each intrinsic binding energy E_X incorporates a number of terms, including the enthalpy of interaction between the functional group and its corresponding binding site on the receptor, the enthalpy changes associated with the removal of water of hydration from the functional group and its target site, and the subsequent formation of bonds between the displaced water molecules, the corresponding entropy terms associated with the displacement and subsequent bonding of water molecules, and the low-frequency vibrational entropy associated with the bonds formed between the functional group and its partner (Figure 1). It is apparent that these factors may be regarded, at least approximately, as properties of the functional group that are relatively independent of the groups to which the particular functional group is attached. Such intrinsic binding potentials may thus reasonably be used in an additive manner to provide an overall estimate of the drug–receptor interaction.

The application of this equation to the 200 molecules chosen for study leads to estimates of E_X for a range of common functional groups, as well as a value for E_{DOF}. These numbers are then shown to be useful in providing answers to questions of the type outlined above.

Method

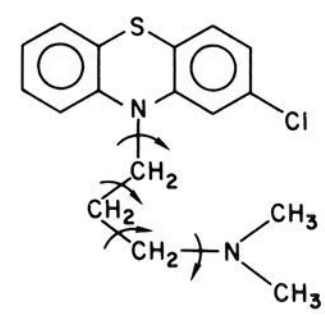
An extensive literature survey of binding studies provided the basic data for this project. Where separate studies of the binding of one drug to different receptors had been carried out, only the tightest binding interaction was considered. It was found that the relative strengths of binding were expressed in a number of different ways, the most direct being in terms of K_D values, which are simply dissociation constants for drug–receptor complexes. In many cases, literature binding data were reported in the form of IC_{50} values, i.e., the concentration of unlabeled drug required to cause 50% inhibition of the binding of a labeled drug. For our purposes these were converted to conventional equilibrium constants (K_I values), using the Cheng–Prusoff equation[8]

$$K_I = IC_{50}/(1 + D/K_D) \qquad (2)$$

in which D is the concentration of the labeled ligand and K_D is its dissociation constant. The third type of binding data encountered was for enzyme inhibitors. These were usually expressed in the form of inhibitory constants, K_I values, for the binding of competitive enzyme inhibitors. Binding data for irreversible inhibitors, or for compounds that bind covalently to enzymes, were not considered, since this study applies to noncovalent molecular interactions.

Having obtained binding data for a wide range of drugs, the next stage was to quantify those drugs in terms of a limited number of common structural units. The basic units chosen were as follows: degrees of internal rotational freedom, sp^2 carbons, sp^3 carbons, protonated nitrogens, neutral nitrogens, carboxyl groups, phosphates, hydroxy groups, carbonyls, ethers or thio ethers, and halogens. In the few cases where sp hybridized carbons were present these were coded as sp^2 carbons. The coding of a drug or compound into these units was carried out simply by counting the numbers of each type of structural unit in the compound, taking account of the following additional considerations. (1) The number of internal degrees of freedom was obtained by summing the number of possible internal rotations, excluding methyl group rotation and internal rotation in amide and other restricted systems. The aim here was to take account of those motions that would be lost on receptor binding. (2) The assignment of nitrogen types was based on literature pK_a values and the likelihood of protonation at physiological pH. Where pK_a values were not readily available, the protonation state of the nitrogen was either determined from the pK_a of a similar compound or predicted by using the method of Clark and Perrin.[9] Compounds containing nitrogens with pK_a values more than 2 pK_a units below physiological pH (7.4) were coded as neutral.

A simple example serves to illustrate the structural coding process. Chlorpromazine is the first entry in Table I and has the structure shown below.



(8) Y-C. Cheng and W. H. Prusoff, *Biochem. Pharmacol.*, **22**, 3099 (1973).

(9) J. Clark and D. D. Perrin, *Q. Rev., Chem. Soc.*, 295 (1964).

Table I. Structural Classification and Binding Energies

		number of functional groups											binding energy, kcal/mol		
no.	compd	DOF	C (sp^2)	C (sp^3)	N^+	N	CO_2^{2-}	PO_4^{2-}	OH	CO	O,S	Hal	obsd[a]	av[b]	diff[c]
1	chlorpromazine	4	12	5	1	1	0	0	0	0	1	1	12.3[u]	10.7	1.6
2	amitriptyline	3	14	6	1	0	0	0	0	0	0	0	12.3[u]	10.0	2.3
3	doxepin	3	14	5	1	0	0	0	0	0	1	0	13.1[u]	10.3	2.8
4	protriptyline	4	14	5	1	0	0	0	0	0	0	0	10.8[u]	8.5	2.2
5	mepyramine	7	11	6	1	2	0	0	0	0	1	0	12.7[u]	8.6	4.1
6	chlorpheniramine	5	11	5	1	1	0	0	0	0	0	1	12.8[u]	8.2	4.6
7	triprolidine	4	13	6	1	1	0	0	0	0	0	0	13.6[u]	9.8	3.8
8	carbinoxamine	6	11	5	1	1	0	0	0	0	1	1	12.3[u]	8.6	3.7
9	clozapine	1	13	5	1	3	0	0	0	0	0	1	12.3[u]	14.7	−2.4
10	mianserin	0	12	6	1	1	0	0	0	0	0	0	12.5[v]	11.8	0.7
11	desipramine	4	12	6	1	1	0	0	0	0	0	0	12.2[w]	9.1	3.1
12	nisoxetine	7	12	5	1	0	0	0	0	0	2	0	11.5[w]	7.2	4.2
13	WB-4101	8	12	7	1	0	0	0	0	0	5	0	12.9[x]	11.3	1.6
14	naphazoline	2	11	3	1	1	0	0	0	0	0	0	11.9[y]	7.4	4.5
15	oxymetazoline	4	7	9	1	1	0	0	1	0	0	0	12.2[z]	10.5	1.6
16	phentolamine	5	13	4	1	2	0	0	1	0	0	0	11.9[z]	11.2	0.6
17	noradrenaline	6	6	2	1	0	0	0	3	0	0	0	12.1[aa]	6.7	5.4
18	piperoxan	2	6	8	1	0	0	0	0	0	2	0	10.7[aa]	8.8	1.9
19	clonidine	1	7	2	1	2	0	0	0	0	0	2	12.3[aa]	8.2	4.0
20	*p*-aminoclonidine	2	7	2	1	3	0	0	0	0	0	2	12.7[aa]	8.7	3.9
21	14,304-18	1	9	2	1	4	0	0	0	0	0	1	12.4[ab]	10.7	1.7
22	lofexidine	3	7	4	1	1	0	0	0	0	1	2	11.8[ab]	8.4	3.4
23	tiamenidine	1	5	3	1	2	0	0	0	0	1	1	11.8[ab]	7.4	4.4
24	CP 18,534-1	1	9	2	1	2	0	0	0	0	1	2	11.6[ab]	10.7	0.9
25	ST600	1	7	3	1	2	0	0	0	0	0	1	11.4[ab]	7.7	3.6
26	guanabenz	2	8	0	1	3	0	0	0	0	0	2	12.3[ab]	7.8	4.5
27	Bay_α 6781	1	1	10	1	1	0	0	0	0	1	0	12.2[ab]	7.7	4.5
28	guanfacine	3	7	1	1	2	0	0	0	1	0	2	11.9[ab]	9.5	2.4
29	xylazine	1	7	5	1	1	0	0	0	0	1	0	11.3[ab]	7.9	3.4
30	yohimbine	3	8	12	1	1	0	0	1	1	1	0	11.9[ac]	18.7	−6.9
31	dibozane	4	12	10	1	1	0	0	0	0	4	0	11.5[ac]	16.6	−5.1
32	prazosin	5	12	6	1	4	0	0	0	1	3	0	13.1[ad]	18.6	−5.4
33	indoramin	5	14	7	1	2	0	0	0	1	0	0	11.5[ad]	15.2	−3.7
34	propranolol	7	10	6	1	0	0	0	1	0	1	0	12.8[ae]	8.1	4.7
35	timolol	8	2	11	1	3	0	0	1	0	3	0	13.0[ae]	11.5	1.5
36	alprenolol	9	8	7	1	0	0	0	1	0	1	0	12.9[ae]	6.1	6.7
37	dihydroalprenolol	9	6	9	1	0	0	0	1	0	1	0	13.1[ae]	6.3	6.8
38	spiperone	6	12	9	1	2	0	0	0	2	0	1	14.0[af]	19.4	−5.3
39	pimozide	7	18	9	1	2	0	0	0	1	0	2	12.8[af]	20.8	−8.0
40	haloperidol	7	12	8	1	0	0	0	1	1	0	2	12.3[af]	16.0	−3.7
41	domperidone	5	12	8	1	4	0	0	0	2	0	1	13.8[ag]	21.6	−7.8
42	ketanserin	5	12	7	1	2	0	0	0	3	0	1	12.8[ah]	21.9	−9.0
43	pizotifen	0	12	7	1	0	0	0	0	0	1	0	13.0[ah]	12.5	0.5
44	clopipazan	0	14	5	1	0	0	0	0	0	1	1	12.8[ah]	13.6	−0.8
45	cyproheptadine	0	16	5	1	0	0	0	0	0	0	0	12.8[ah]	12.6	0.1
46	metergoline	5	14	10	1	2	0	0	0	1	1	0	13.0[ah]	18.7	−5.6
47	methysergide	5	10	10	1	2	0	0	1	1	0	0	12.3[ah]	17.3	−5.0
48	pipamperone	7	6	13	1	2	0	0	0	2	0	1	12.4[ah]	17.7	−5.3
49	benperidol	6	12	8	1	2	0	0	0	2	0	1	12.2[ah]	18.6	−6.4
50	fluspirilene	7	18	10	1	2	0	0	0	1	0	2	11.6[ah]	21.6	−10.0
51	penfluridol	9	18	10	1	0	0	0	1	0	0	6	9.9[ah]	22.2	−12.4
52	astemizole	8	19	9	1	3	0	0	0	0	1	1	11.3[ah]	18.4	−7.1
53	nomifensine	2	12	4	1	1	0	0	0	0	0	0	8.8[ah]	8.9	−0.1
54	bufotenine	4	8	4	1	1	0	0	1	0	0	0	9.4[ah]	7.2	2.2
55	serotonin	4	8	2	1	1	0	0	1	0	0	0	8.9[ah]	5.6	3.3
56	tryptamine	3	8	2	1	1	0	0	0	0	0	0	7.9[ah]	3.8	4.2
57	mescaline	6	6	5	1	0	0	0	0	0	3	0	7.1[ah]	4.8	2.4
58	LSD	3	10	9	1	2	0	0	0	1	0	0	12.7[ai]	15.3	−2.7
59	apomorphine	2	12	5	1	0	0	0	2	0	0	0	11.9[ai]	13.5	−1.6
60	flupenthixol	7	14	9	1	1	0	0	1	0	1	3	11.5[ai]	18.3	−6.9
61	dopamine	5	6	2	1	0	0	0	2	0	0	0	11.8[ai]	4.9	6.9
62	GABA	4	0	3	1	0	1	0	0	0	0	0	10.9[aj]	5.3	5.6
63	muscimol	3	3	1	1	1	0	0	1	0	1	0	11.9[aj]	3.1	8.8
64	diazepam	1	13	2	0	2	0	0	0	1	0	1	11.5[ak]	3.1	8.5
65	lorazepam	2	13	1	0	2	0	0	1	1	0	2	11.6[ak]	5.4	6.2
66	bromazepam	1	12	1	0	3	0	0	0	1	0	1	10.3[ak]	2.7	7.5
67	mephobarbitone	2	6	4	0	2	0	0	0	3	0	0	7.3[al]	4.5	2.7
68	metharbitone	2	0	6	0	2	0	0	0	3	0	0	6.8[al]	1.9	4.9
69	DMBB[d]	4	0	9	0	2	0	0	0	3	0	0	10.0[al]	3.0	7.0
70	CHEB[e]	3	2	9	0	2	0	0	0	3	0	0	8.4[al]	5.1	3.3
71	picrotoxinin	2	2	11	0	0	0	0	1	2	3	0	8.7[am]	7.3	1.4
72	tutin	3	2	12	0	0	0	0	2	1	3	0	8.8[am]	6.6	2.3
73	phenytoin	2	12	1	0	2	0	0	0	2	0	0	11.2[an]	3.0	8.2
74	carbamazepine	0	14	0	0	2	0	0	0	1	0	0	5.9[ao]	1.5	4.4

Table I (Continued)

no.	compd	number of functional groups											binding energy, kcal/mol		
		DOF	C (sp^2)	C (sp^3)	N^+	N	CO_2^{2-}	PO_4^{2-}	OH	CO	O,S	Hal	obsd[a]	av[b]	diff[c]
75	oxalate	1	0	0	0	0	2	0	0	0	0	0	7.8[ap]	1.7	6.1
76	benzyl succinic acid	5	6	3	0	0	2	0	0	0	0	0	8.5[aq]	5.6	2.9
77	3-phenylpropionic acid	3	6	2	0	0	1	0	0	0	0	0	5.5[aq]	−2.0	7.5
78	NRDC 157	7	14	6	0	0	0	0	0	1	2	2	10.1[ar]	3.9	6.2
79	acetaldehyde ammonia	2	0	2	1	0	0	0	1	0	0	0	6.5[as]	0.2	6.3
80	acetate	0	0	1	0	0	1	0	0	0	0	0	3.1[as]	−5.0	8.1
81	ethylamine	1	0	2	1	0	0	0	0	0	0	0	2.2[as]	−1.6	3.8
82	pyridoxylalanine	7	5	5	1	1	1	0	2	0	0	0	9.2[at]	14.6	−5.4
83	3,4,5,6-tetrahydrouridine	6	0	8	0	2	0	0	4	1	1	0	9.0[au]	5.2	3.8
84	H_4-dUMP[f]	6	0	8	0	2	0	1	2	1	1	0	10.9[av]	10.2	0.7
85	acetopyruvate	3	0	2	0	0	1	0	0	2	0	0	9.1[aw]	0.5	8.6
86	phosphoglycolohydroxamic acid	5	1	1	0	1	0	1	2	0	0	0	10.9[ax]	0.3	10.6
87	2-phosphoglycolate	3	0	1	0	0	1	1	0	0	0	0	8.7[ay]	3.0	5.7
88	3-aminoenolpyruvate phosphate	4	2	0	1	0	1	1	0	0	0	0	10.5[az]	14.4	−3.9
89	TSP[g]	4	0	1	0	0	1	1	0	1	0	0	7.6[az]	5.7	2.0
90	CRDP[h]	12	0	5	0	0	1	2	3	0	0	0	8.2[ba]	17.6	−9.5
91	PALA[i]	6	0	3	0	1	2	1	0	1	0	0	12.3[bb]	15.3	−3.0
92	DIKEP[j]	7	0	4	0	0	2	1	0	1	0	0	10.9[bb]	14.2	−3.3
93	coformycin	6	4	7	1	3	0	0	4	0	1	0	15.0[bc]	16.5	−1.5
94	HRD[k]	6	0	9	0	2	0	0	4	1	1	0	10.5[bc]	6.0	4.5
95	PRBA[l]	6	0	6	0	2	0	1	2	3	1	0	15.1[bd]	15.4	−0.3
96	PPPA[m]	5	6	2	0	0	1	1	0	0	0	0	9.3[be]	6.6	2.7
97	SQ 14102	5	0	8	0	1	2	0	0	1	0	0	8.3[bf]	10.0	−1.7
98	allopurinol	0	4	0	1	3	0	0	0	1	0	0	12.5[bg]	7.2	5.4
99	alloxanthine	1	4	0	1	3	0	0	1	1	0	0	12.6[bg]	9.0	3.6
100	carfentanil	8	12	10	1	1	0	0	0	2	1	0	14.2[bh]	17.4	−3.2
101	R26800	6	12	10	1	1	0	0	0	1	0	0	13.8[bh]	14.3	−0.5
102	sufentanil	8	10	11	1	1	0	0	0	1	2	0	13.5[bh]	14.5	−1.0
103	R4837	8	18	10	1	2	0	0	0	2	0	0	13.6[bh]	21.7	−8.1
104	R6582	3	12	7	1	2	0	0	0	1	0	1	12.0[bh]	16.4	−4.4
105	dextromoramide	6	12	12	1	1	0	0	0	1	1	0	12.3[bh]	17.0	−4.7
106	methadone	7	12	8	1	0	0	0	0	1	0	0	11.8[bh]	10.9	0.9
107	R951	8	12	9	1	0	0	0	0	2	1	0	11.6[bh]	15.4	−3.9
108	ketobemidone	4	6	8	1	0	0	0	1	1	0	0	10.8[bh]	11.2	−0.5
109	etorphine	6	8	17	1	0	0	0	2	0	2	0	13.5[bh]	19.7	−6.2
110	buprenorphine	7	6	23	1	0	0	0	2	0	2	0	13.4[bh]	22.4	−9.1
111	ketazocine	3	6	11	1	0	0	0	1	1	0	0	11.4[bh]	14.3	−2.9
112	moxazocine	4	6	12	1	0	0	0	1	0	1	0	12.2[bh]	12.1	0.1
113	2,4-diamino-6-isobutylpteridine	4	6	4	1	5	0	0	0	0	0	0	12.3[bi]	8.1	4.3
114	2,4-diamino-6-benzylpteridine	4	12	1	1	5	0	0	0	0	0	0	12.3[bi]	9.9	2.5
115	pyrimethamine	4	10	2	1	3	0	0	0	0	0	1	9.7[bi]	8.2	1.5
116	methotrexate	11	12	5	1	7	2	0	0	1	0	0	12.9[bi]	30.5	−17.6
117	DMCPT[n]	4	13	2	1	4	0	0	0	0	0	1	13.2[bi]	11.5	1.7
118	DMPQ[o]	6	8	6	1	3	0	0	0	0	0	0	18.7[bi]	7.3	11.4
119	DAEP[p]	4	4	12	1	3	0	0	0	0	0	0	13.4[bj]	10.7	2.8
120	DDMP[q]	3	10	1	1	3	0	0	0	0	0	2	12.3[bj]	9.4	2.9
121	trimethoprim	7	10	4	1	3	0	0	0	0	3	0	11.5[bk]	9.6	1.8
122	2,3-diphosphoglycerate	6	0	2	0	0	1	2	0	0	0	0	6.8[bl]	11.8	−5.0
123	DBOA[r]	8	12	3	0	0	1	0	0	2	1	0	6.4[bl]	7.4	−1.0
124	biotin	5	0	8	0	2	1	0	0	1	1	0	20.5[bm]	4.0	16.4
125	isoleucine	4	0	5	1	0	1	0	0	0	0	0	7.1[bm]	6.9	0.2
126	2-methylbutamine	3	0	5	1	0	0	0	0	0	0	0	2.8[bm]	−0.6	3.4
127	2-phenylethylamine	3	6	2	1	0	0	0	0	0	0	0	5.5[bn]	1.2	4.3
128	amphetamine	3	6	3	1	0	0	0	0	0	0	0	6.8[bn]	2.0	4.8
129	phenylalaninol	5	6	3	1	0	0	0	1	0	0	0	7.1[bn]	3.2	3.9
130	phenylalaninamide	4	6	2	1	1	0	0	0	1	0	0	6.1[bn]	5.1	1.0
131	5′-deoxyadenosine	4	5	5	1	4	0	0	2	0	1	0	8.2[bo]	13.1	−4.9
132	*cis*-AMB[s]	4	2	2	1	0	1	0	0	0	0	1	5.9[bp]	7.2	−1.4
133	cycloleucine	2	0	5	1	0	1	0	0	0	0	0	4.9[bp]	8.3	−3.4
134	methoxinine	5	0	4	1	0	1	0	0	0	0	1	5.1[bp]	6.8	−1.6
135	phenol	1	6	0	0	0	0	0	1	0	0	0	2.7[bq]	−8.0	10.7
136	atropine	6	6	10	1	0	0	0	1	1	1	0	12.8[br]	12.5	0.3
137	arecoline	2	2	5	1	0	0	0	0	1	1	0	7.2[br]	5.9	1.2
138	benztropine	4	12	9	1	0	0	0	0	0	1	0	12.6[br]	11.4	1.2
139	scopolamine	6	6	10	1	0	0	0	1	1	2	0	12.9[br]	13.6	−0.7
140	thioridazine	4	12	9	1	1	0	0	0	0	2	0	11.0[br]	13.6	−2.6
141	pilocarpine	3	3	7	1	1	0	0	0	1	1	0	10.1[bs]	8.7	1.4
142	dexetimide	4	12	9	1	1	0	0	0	2	0	0	12.0[bs]	18.3	−6.2
143	oxotremorine	4	2	9	1	1	0	0	0	1	0	0	11.7[bs]	7.9	3.8
144	cytisine	0	4	6	1	1	0	0	0	1	0	0	12.1[bt]	9.6	2.5
145	nicotine	1	5	5	1	1	0	0	0	0	0	0	11.2[bt]	5.4	5.7
146	acetylcholine	4	0	6	1	0	0	0	0	1	1	0	11.1[bt]	4.0	7.1
147	carbachol	4	0	5	1	1	0	0	0	1	1	0	10.7[bt]	4.4	6.4
147	prostaglandin E_2	14	4	14	0	0	1	0	2	1	0	0	11.5[bu]	7.1	4.4

Table I (Continued)

no.	compd	number of functional groups											binding energy, kcal/mol		
		DOF	C (sp^2)	C (sp^3)	N^+	N	CO_2^{2-}	PO_4^{2-}	OH	CO	O,S	Hal	obsd[a]	av[b]	diff[c]
149	prostaglandin E_1	15	2	16	0	0	1	0	2	1	0	0	11.2[bu]	6.6	4.6
150	prostaglandin $F_{2\alpha}$	15	4	15	0	0	1	0	3	0	0	0	9.2[bu]	6.3	2.8
151	citrate	6	0	3	0	0	3	0	1	0	0	0	5.0[bv]	11.4	-6.4
152	malate	4	0	2	0	0	2	0	1	0	0	0	3.4[bv]	3.8	-0.4
153	1,2,4-benzenetricarboxylate	3	6	0	0	0	3	0	0	0	0	0	5.2[bv]	12.8	-7.5
154	progesterone	1	2	17	0	0	0	0	0	2	0	0	10.5[bw]	7.1	3.4
155	androstanolone	1	0	18	0	0	0	0	1	1	0	0	13.1[bx]	5.6	7.5
156	estradiol	2	6	12	0	0	0	0	2	0	0	0	13.2[bx]	3.4	9.8
157	dexamethasone	5	4	15	0	0	0	0	3	2	0	1	10.9[by]	13.0	-2.1
158	medroxyprogesterone acetate	3	2	19	0	0	0	0	0	3	1	0	10.2[by]	11.7	-1.5
159	dihydrotestosterone	1	0	18	0	0	0	0	1	1	0	0	6.8[bz]	5.6	1.2
160	corticosterone	4	2	17	0	0	0	0	2	2	0	0	11.6[ca]	10.0	1.6
161	aldosterone	4	2	17	0	0	0	0	2	2	1	0	12.7[ca]	11.1	1.6
162	tamoxifen	8	20	6	1	0	0	0	0	0	1	0	12.3[cb]	11.9	0.4
163	butenol 4-phosphate	5	2	2	0	0	0	1	1	0	0	0	4.1[cc]	-1.9	5.9
164	butynol 4-phosphate	5	2	2	0	0	0	1	1	0	0	0	3.4[cc]	-1.9	5.3
165	butanol 4-phosphate	6	0	4	0	0	0	1	1	0	0	0	3.3[cc]	-2.3	5.6
166	xylose 1-phosphate	5	0	5	0	0	0	1	3	0	1	0	6.2[cc]	5.2	1.0
167	fructose 1,6-diphosphate	9	0	6	0	0	0	2	3	0	1	0	10.2[cd]	13.3	-3.2
168	fructose 6-phosphate	8	0	6	0	0	0	1	4	0	1	0	5.7[ce]	6.5	-0.8
169	cytidine monophosphate	7	3	5	0	3	0	1	2	1	1	0	5.9[cf]	10.4	-4.5
170	phosphate	0	0	0	0	0	0	1	0	0	0	0	7.7[cg]	-4.0	11.7
171	camphor	0	0	9	0	0	0	0	0	1	0	0	11.4[ch]	-3.4	14.8
172	glucose	6	0	6	0	0	0	0	5	0	1	0	8.7[ci]	0.4	8.3
173	prenylamine	8	18	6	1	0	0	0	0	0	0	0	8.2[cj]	9.4	-1.2
174	cinnarizine	6	20	6	1	1	0	0	0	0	0	0	8.4[cj]	13.3	-5.0
175	ouabain	12	2	26	0	0	0	0	8	1	3	0	11.5[ck]	26.7	-15.2
176	*n*-acetylcolchinol	5	12	7	0	1	0	0	1	1	3	0	7.0[cl]	6.9	0.1
177	maysenine	3	14	11	0	2	0	0	1	2	3	1	7.2[cm]	18.7	-11.4
178	triac	5	12	1	0	0	1	0	1	0	1	3	9.3[cn]	7.5	1.9
179	tetraform	4	12	0	0	0	1	0	1	0	1	4	10.7[cn]	8.6	2.0
180	tetrac	5	12	1	0	0	1	0	1	0	1	4	11.4[cn]	8.8	2.7
181	thyroxine	7	12	2	1	0	1	0	1	0	1	4	10.3[cn]	19.7	-9.4
182	fenclofenac	4	12	1	0	0	1	0	0	0	1	2	9.2[co]	4.3	4.9
183	monohydroxyfenclofenac	5	12	1	0	0	1	0	1	0	1	2	8.7[co]	6.2	2.5
184	diclofenac	4	12	1	1	0	1	0	0	0	0	2	9.6[co]	14.7	-5.1
185	ibuprofen	4	6	6	0	0	1	0	0	0	0	0	8.3[cp]	0.5	7.9
186	naproxen	3	10	3	0	0	1	0	0	0	1	0	8.5[cp]	2.6	5.9
187	flurbiprofen	3	12	2	0	0	1	0	0	0	0	1	9.1[cp]	3.4	5.7
188	24,25-dihydroxy vitamin D_3	9	6	21	0	0	0	0	3	0	0	0	13.2[cq]	8.4	4.8
189	CPIB[t]	3	6	3	0	0	1	0	0	0	1	1	6.7[cr]	1.1	5.5
190	valproate	5	0	7	0	0	1	0	0	0	0	0	6.1[cs]	-3.6	9.7
191	bishydroxycoumarin	4	16	1	0	0	0	0	2	2	2	0	8.8[ct]	9.2	-0.4
192	warfarin	5	14	3	0	0	0	0	1	2	1	0	6.9[ct]	5.1	1.8
193	3-methyl-4-hydroxycoumarin	1	8	1	0	0	0	0	1	1	1	0	7.0[cu]	-1.3	8.3
194	DDT	3	12	2	0	0	0	0	0	0	0	5	7.6[cv]	0.4	7.2
195	dieldrin	0	2	10	0	0	0	0	0	0	1	6	6.1[cv]	4.2	1.9
196	lindane	0	0	6	0	0	0	0	0	0	0	6	6.1[cv]	-1.4	7.6
197	carbaryl	2	10	1	0	1	0	0	0	1	1	0	6.4[cv]	-1.9	8.3
198	carbofuran	2	6	5	0	1	0	0	0	1	2	0	4.6[cv]	-0.5	5.0
199	phenylbutazone	5	12	5	0	2	0	0	0	2	0	0	7.3[cw]	4.1	3.2
200	tetramethylammonium	0	0	4	1	0	0	0	0	0	0	0	3.6[cx]	0.6	3.0

[a] Calculated from observed binding constants (K_I or K_D values) with use of the equation $\Delta G_{obsd} = -RT \ln K_D$. [b] Value derived by summing AVERAGE functional group contributions and correcting for losses of entropy. [c] Difference between the observed and AVERAGE binding energy. Indicates the nature of the fit of the molecule to its binding site. Positive values indicate better than average fit; negative values worse than average fit. [d] 5-(1,3-Dimethylbutyl)-5-ethylbarbituric acid. [e] 5-(2-Cyclohexylideneethyl)-5-ethylbarbituric acid. [f] Tetrahydrodeoxyuridine 5'-monophosphate. [g] Tartronate semialdehyde phosphate. [h] 2-Carboxyribitol 1,5-diphosphate. [i] *N*-(Phosphonoacetyl)-L-aspartate. [j] 4,5-Dicarboxy-2-oxopentyl phosphonate. [k] 5-Hydroxy-1-β-D-ribofuranosyl-1,3-diazepin-2-one. [l] 1-(5-Phospho-β-D-ribofuranosyl)-barbituric acid. [m] 2-(Phosphoroyloxy)-3-phenylpropionic acid. [n] 2,4-Diamino-5-methyl-6-(*p*-chlorophenmethyl)-1,3,8-triazanaphthalene. [o] 2,4-Diamino-5-methyl-6-pentylquinazoline. [p] 2,4-Diamino-5-(1-adamantyl)-6-ethylpyrimidine. [q] 2,4-Diamino-5-(3,4-dichlorophenyl)-6-methylpyrimidine. [r] 4,4'-Diformyl-2-bibenzyloxyacetic acid. [s] *cis*-2-Amino-4-methoxybut-3-enoic acid. [t] α-(*o*-Chlorophenoxy)isobutyric acid. [u] V. T. Tran, R. Lebovitz, L. Toll, and S. H. Snyder, *Eur. J. Pharmacol.*, **70**, 501 (1981). [v] N. Brunello, D. Chuang, and E. Costa, *Eur. J. Pharmacol.*, **78**, 383 (1982). [w] R. Raisman, M. Sette, C. Pimoule, M. Briley, and S. Z. Langer, *Eur. J. Pharmacol.*, **78**, 345 (1982). [x] R. Raisman, M. S. Briley, and S. Z. Langer, *Eur. J. Pharmacol.*, 61, 373 (1980). [y] G. A. McPherson and R. J. Summers, *Clin. Exp. Pharmacol. Physiol.*, **9**, 77 (1982). [z] B. R. Rouot and S. H. Snyder, *Life Sci.*, **25**, 769 (1979). [aa] A. I. Salama, L. L. Lin, L. D. Repp, and D. C. U'Prichard, *Life Sci.*, **30**, 1305 (1982). [ab] R. J. Summers, B. Jarrott, and W. J. Louis, *Eur. J. Pharmacol.*, **66**, 233 (1980). [ac] K. D. Newman, L. T. Williams, N. H. Bishopric, and R. J. Lefkowitz, *J. Clin. Investig.*, **61**, 395 (1978). [ad] P. Barnes, J. Karliner, C. Hamilton, and C. Dollery, *Life Sci.*, **25**, 1207 (1979). [ae] K. E. J. Dickinson and S. R. Nahorski, *Eur. J. Pharmacol.*, **94**, 43 (1981). [af] P. M. Beart, M. Krstich, D. McDonald, and A. L. Gundlach, *Neuroscience Lett.*, **29**, 147 (1982). [ag] H. W. Rzezniczak, A. L. Gundlach, and P. M. Beart, *Neuroscience Lett.*, **30**, 63 (1982). [ah] J. E. Leysen, C. J. E. Niemegeers, J. M. Van Neuten, and P. M. Laduron, *Mol. Pharmacol.*, **21**, 301 (1982). [ai] P. Seeman, *Pharmacol. Rev.*, **32**, 229 (1981). [aj] S. J. Enna, J. W. Ferkaney, and P. Krogsgaard-Larsen in "GABA Neurotransmitters—Pharmacochemical, Biochemical and Pharmacological Aspects" (Alfred Benzoin Symposium 12), P. Krogsgaard-Larsen, J. Scheel-Kruger and H. Kofod, Eds., Munksgaard, 1978, pp 191–200. [ak] C. Braestrup and R. F. Squires, *Eur. J. Pharmacol.*, **48**, 263 (1978). [al] M. K. Ticku and R. W. Olsen, *Life Sci.*, **22**, 1643 (1978). [am] M. K. Ticku, M. Ban, and R. W. Olsen, *Mol. Pharmacol.*, 14, 391 (1978). [an] W. M. Burnham, L. Spero, M. M.

Table I footnotes (continued)

Okazaki, and B. K. Madras, *Can. J. Physiol. Pharmacol.*, **59**, 402 (1981). [ao] J. H. Skerritt, L. P. Davies, and G. A. R. Johnston, *Eur. J. Pharmacol.*, **82**, 195 (1982). [ap] Reference 1. [aq] L. D. Byers and R. Wolfenden, *J. Biol. Chem.*, **247**, 606 (1972). [ar] D. M. Soderlund, S. M. Ghiasuddin, and D. W. Helmuth, *Life Sci.*, **33**, 261 (1983). [as] J. D. Findlater and B. A. Orsi, *FEBS Lett.*, **35**, 109 (1973). [at] W. B. Dempsey and E. E. Snell, *Biochemistry*, **2**, 1414 (1963). [au] R. M. Cohen and R. Wolfenden, *J. Biol. Chem.*, **246**, 7561 (1971). [av] F. Maley and G. F. Maley, *Arch. Biochem. Biophys.*, **144**, 723 (1971). [aw] W. Tagaki, J. P. Guthrie, and F. H. Westheimer, *Biochemistry*, **7**, 905 (1968). [ax] K. D. Collins, *J. Biol. Chem.*, **249**, 136 (1974). [ay] F. C. Hartman, G. M. La Muraglia, Y. Tomozowa, and R. Wolfenden, *Biochemistry*, 14, 5274 (1975). [az] T. G. Spring and F. Wold, *Biochemistry*, **10**, 4655 (1971). [ba] M. Wishnick and M. D. Lane, *J. Biol. Chem.*, **245**, 4939 (1970). [bb] E. A. Swyryd, S. S. Seaver, and G. R. Stark, *J. Biol. Chem.*, **249**, 6945 (1974). [bc] V. E. Marquez, P. S. Liu, J. A. Kelley, J. S. Driscoll, and J. J. McCormack, *J. Med. Chem.*, **23**, 713 (1980). [bd] H. L. Levine, R. S. Brody, and F. H. Westheimer, *Biochemistry*, **19**, 4993 (1980). [be] W. Hofmann and M. Rottenberg in "Enzyme Inhibitors", U. Brodbeck, Ed., Verlag Chemie, Weinheim, 1980, pp 19–26. [bf] D. W. Cushman, H. S. Cheung, E. F. Sabo, and M. A. Ondetti, *Biochemistry*, **16**, 5484 (1977). [bg] S. Cha, R. P. Agarwal, and R. E. Parks, Jr., *Biochem. Pharmacol.*, **24**, 2187 (1975). [bh] J. E. Leysen, W. Gommeren, and C. J. E. Niemegeers, *Eur. J. Pharmacol.*, **87**, 209 (1983). [bi] Reference 2. [bj] W. R. Greco and M. T. Hakala, *Mol. Pharmacol.*, **18**, 521 (1980). [bk] H. S. Aschhoff and H. Vergin, *J. Antimicrob. Chemother.*, **5** (suppl. B), 19 (1979). [bl] C. R. Beddell, P. J. Goodford, D. K. Stammers, and R. Wootton, *Br. J. Pharmacol.*, **65**, 535 (1979). [bm] Reference 6. [bn] D. V. Santi and P. V. Danenberg, *Biochemistry*, **10**, 4813 (1971). [bo] D. V. Santi, P. V. Danenberg, and K. A. Montgomery, *Biochemistry*, **10**, 4821 (1971). [bp] J. R. Sufrin, J. B. Lombardini, and D. D. Keith, *Biochem. Biophys. Res. Commun.*, **106**, 251 (1982). [bq] I. Simonsson, B-H. Jonsson, and S. Lindskog, *Biochem. Biophys. Res. Commun.*, **108**, 1406 (1982). [br] C. P. Smith and F. P. Huger, *Biochem. Pharmacol.*, **32**, 377 (1983). [bs] M. Waelbroeck, P. Robberecht, P. Chatelain, and J. Christophe, *Mol. Pharmacol.*, **21**, 581 (1982). [bt] R. D. Schwartz, R. McGee, Jr., and K. J. Kellar, *Mol. Pharmacol.*, **22**, 56 (1982). [bu] R. Grant, K. Aktories, and K. H. Jacobs, *Mol. Pharmacol.*, **22**, 320 (1982). [bv] Z. B. Rose, *Adv. Enzym. Rel. Areas Mol. Biol.*, **51**, 211 (1980). [bw] A. Gal and A. Venetianer, *Biochem. Pharmacol.*, **32**, 919 (1983). [bx] I. Jung-Testas, F. Bayard, and E. E. Baulieu, *Nature (London)*, **259**, 136 (1976). [by] K. Kontula, T. Paavonen, T. Luukkainen, and L. C. Andersson, *Biochem. Pharmacol.*, **32**, 1511 (1983). [bz] M. Basset, G. DeFaye, and E. M. Chambaz, *Biochim. Biophys. Acta*, **491**, 434 (1977). [ca] S-M. Kurt Lee, M. A. Chekal, and A. I. Katz, *Am. J. Physiol.*, **244**, F504 (1983). [cb] J-C. Faye, S. Jozan, G. Redeuilh, E. E. Baulieu, and F. Bayard, *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 3158 (1983). [cc] W. P. Jencks, *Adv. Enzymol. Rel. Areas Mol. Biol.*, **43**, 219 (1975). [cd] C. B. Libby, W. A. Frey, J. J. Villafranca, and S. J. Benkovic, *J. Biol. Chem.*, **250**, 7564 (1975). [ce] N. P. B. Dudman, M. M. de Maine, and S. J. Benkovic, *J. Biol. Chem.*, **253**, 5712 (1978). [cf] T. A. Beyer, J. E. Sadler, J. I. Rearick, J. C. Paulson, and R. L. Hill, *Adv. Enzymol. Rel. Areas Mol. Biol.*, **52**, 23 (1981). [cg] D. Levine, T. W. Reid, and I. B. Wilson, *Biochemistry*, 8, 2374 (1969). [ch] B. W. Griffin and J. A. Peterson, *Biochemistry*, 11, 4740 (1972). [ci] B. I. Gold and J. Kyle-Lillegard, *Life Sci.*, **32**, 1865 (1983). [cj] P. Belleman, A. Schade, and R. Towart, *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 2356 (1983). [ck] A. R. Ford, J. K. Aronson, D. G. Grahame-Smith, and J. A. Rose, *Br. J. Clin. Pharmacol.*, 8, 115 (1979). [cl] M. H. Zweig and C. F. Chignell, *Biochem. Pharmacol.*, **22**, 2141 (1973). [cm] J. York, M. K. Wolpert-De Filippes, D. G. Johns, and V. S. Sethi, *Biochem. Pharmacol.*, **30**, 3239 (1981). [cn] T. A. Andrea, R. R. Cavalieri, I. D. Goldfine, and E. C. Jorgensen, *Biochemistry*, **19**, 55 (1980). [co] S. J. Capper, M. J. Humphrey, and A. B. Kurtz, *Clin. Chim. Acta*, **112**, 77 (1981). [cp] A. Kober and I. Sjoholm, *Mol. Pharmacol.*, **18**, 421 (1980). [cq] D. Somjen, G. J. Somjen, A. Harell, G. L. Mechanic, and I. Binderman, *Biochem. Biophys. Res. Commun.*, **106**, 644 (1982). [cr] N. P. Sollenne and G. E. Means, *Mol. Pharmacol.*, **15**, 754 (1979). [cs] A. Kober, Y. Olssen, and I. Sjoholm, *Mol. Pharmacol.*, 18, 237 (1980). [ct] J. H. Perrin, J. J. Vallner, and D. A. Nelson, *Biochem. Pharmacol.*, **24**, 769 (1975). [cu] Y. Adam, S. Kirkiacharian, and R. Boudet-Dalbin, *Eur. J. Med. Chem.*, 17, 327 (1982). [cv] B. P. Maliwal and F. E. Guthrie, *Mol. Pharmacol.*, **20**, 138 (1981). [cw] A. Rosen, *Biochem. Pharmacol.*, **19**, 2075 (1970). [cx] G. M. Steinberg, M. L. Mednick, J. Maddox, and R. Rice, *J. Med. Chem.*, 18, 1056 (1975).

As indicated, there are four internal rotational degrees of freedom, excluding terminal methyl group rotation. The two benzene rings together account for the 12 sp^2 carbons noted in Table I, while the side-chain methylene and methyl groups account for the five sp^3 carbons. The pK_a of chlorpromazine is 9.3 and hence at physiological pH, the terminal nitrogen (the more basic of the two nitrogens) is protonated. The other nitrogen is neutral. Finally, the sulfur and chlorine atoms each contribute one unit to the appropriate columns of Table I.

After structural coding of the compounds shown in Table I, a computer program[10] was used to perform a regression analysis of these data according to eq 1. The free energies of binding for use in the regression analysis were calculated from eq 3, where K represents K_D or K_I, using

$$\Delta G_{bind} = -RT \ln K \quad (3)$$

a standard temperature of 25 °C (298 K). Corrections to some of the binding data that were obtained at slightly different temperatures are entirely negligible in the context of later correlations.

In the initial stages of the project, regression analyses were carried out with use of a larger number of structural units than the 11 shown in Table I, e.g., separate categories were included for phenyl and amide functionalities. However, a trial-and-error approach showed that no significant improvement could be gained by varying the independent variables from this final set. A stepwise regression (using forward selection, backward elimination, and a combination procedure for the selection of variables[11]) generally confirmed the significance of each of the final independent variables.

Results

Free energies of binding and structural assignments for the 200 compounds used in this study are given in Table I. Compounds were chosen on the basis of availability of binding data, apparent tightness of binding, and diversity of structure. Attention to the last point has meant that the total numbers of each type of structural unit are fairly evenly distributed (except, of course, that there are significantly more sp^2 and sp^3 carbons than any other functionality) and that the final data set is not unduly biased toward one particular drug class or structural type.

The thesis of this work is that the intrinsic binding energies (E_X values) in eq 1 represent the binding of an average drug to an average receptor and that differences between the observed and the calculated average binding energies allow the degree of match of a particular drug to its receptor to be determined. As we have pointed out, one way of determining intrinsic binding energies for individual groups is to make use of the anchor principle.[5-7] However, in this study, where we examine a wide range of drugs and receptor types, insufficient comparative binding data are available for compounds differing by only one functional group. Instead, we have used a kind of mass anchor principle, based on linear regression, to obtain estimates of the intrinsic binding energies,. E_X. We stress that the use of a regression analysis to eq 1 is not meant to imply that there is a precise relationship between the observed

(10) Regression analyses were carried out on a PDP 11/34 minicomputer using a modified version of program REGRES, described by P. R. Bevington in "Data Reduction and Error Analysis for the Physical Sciences", McGraw-Hill, New York, 1969, Chapter 9. Stepwise regressions were carried out by using a standard statistics package, MINITAB, on a Cyber computer at the Royal Melbourne Institute of Technology.

(11) S. Chatterjee and B. Price, "Regression Analysis by Example", Wiley, New York, 1977, pp 201–202.

Table II. Intrinsic Binding Energies (kcal mol^{-1})

no.	group	energy	range[a]
1	DOF[b]	−0.7	−0.7−−1.0
2	C(sp^2)	0.7	0.6–0.8
3	C(sp^3)	0.8	0.1–1.0
4	N$^+$	11.5	10.4–15.0
5	N	1.2	0.8–1.8
6	CO_2^-	8.2	7.3–10.3
7	OPO_3^{2-}	10.0	7.7–10.6
8	OH	2.5	2.5–4.0
9	C=O	3.4	3.2–4.0
10	O,S	1.1	0.7–2.0
11	halogen	1.3	0.2–2.0

[a] Range of energies for six random 100-compound data sets. [b] Degrees of internal conformational freedom.

free energies of binding and the structural units we have selected. This would be the case only if each drug in the series bound each and every one of its structural features to an optimally matched receptor for that drug. The regression analysis is used merely as a vehicle for providing average estimates of the E_X values for each structural unit.

To determine the energy coefficients in eq 1, it is necessary to first fix the value of the entropy term, $T\Delta S_{rt}$. With use of data of Page, this value was calculated[12] to be −14 kcal mol^{-1}. A multiple regression analysis was then carried out, yielding the E_X values (kcal mol^{-1}) shown in eq 4. The free energies of binding calculated from eq 4

$$\Delta G = -14 - 0.7n_{DOF} + 0.7n_{C(sp^2)} + 0.8n_{C(sp^3)} + 11.5n_{N^+} + 1.2n_N + 8.2n_{CO_2^-} + 10.0n_{PO_4^{2-}} + 2.5n_{OH} + 3.4n_{C=O} + 1.1n_{O,S} + 1.3n_{Hal} \quad (4)$$

are shown in Table I, along with deviations from observed values for the 200 compounds studied. Clearly the differences are too large to allow eq 4 to be used in a predictive sense. We stress again that this is not our intention. Rather, the magnitudes of the deviations are to be used to quantify the nature of receptor binding for each drug. To emphasize the fact that eq 4 is not to be used as a predictive tool, we refer to ΔG values calculated from this equation as AVERAGE (Average Energy Resulting from All Group Energies) ΔG values. These reflect the expected binding energy of an average drug, based on its component parts. A positive deviation between the observed binding energy of a drug and its AVERAGE binding energy indicates that it binds better than "average", whereas a negative deviation indicates that the binding is weaker than "average".

In regression analyses for which the intercept has been arbitrarily fixed (in our case, to a known value), standard statistical parameters for evaluating the fit (e.g., the multiple correlation coefficient) and for determining the accuracy of derived regression coefficients (e.g., t tests) are not meaningful. An alternative measure of the accuracy of our intrinsic binding energies (coefficients in eq 4) was therefore obtained by running regression analyses on random subsets of the initial 200-compound data base to monitor variations in the derived coefficients. The ranges of E_X values for six separate sets of 100 compounds are shown in Table II. In terms of percentage errors, the smaller coefficients are less well defined than the larger values, but, in general, the ranges in E_X are fairly small. The values in Table II are certainly more tightly defined than have been previous estimates of intrinsic binding energies.

The derived coefficients do, of course, depend on the value determined for $T\Delta S_{rt}$. While we are confident that this term has a value close to 14 kcal mol^{-1}, we did investigate the effect of using other values. As the magnitude of this fixed term is decreased, the magnitudes of the derived E_X values decrease; however, they maintain the same relative ordering. This is true even if $T\Delta S_{rt}$ is allowed to be a free-fitting variable. This is entirely consistent with the fact that the E_X values represent a combination of enthalpic and entropic terms associated with functional group binding (see Figure 1) but not the entropic term associated with overall rotational and translational motion of the drug. $T\Delta S_{rt}$ is essentially independent of the number and nature of the functional groups present. The fact that all E_X values maintain the same order for different values of $T\Delta S_{rt}$ suggests that forcing this term to have an inappropriate value, or allowing it to be free-fitting, results in the regression analysis mixing some of this overall molecular entropy contribution into the E_X coefficients, giving them misleadingly small values.

Discussion

The intrinsic binding energies in Table II follow the expected trend that charged groups lead to stronger interactions than polar groups, which in turn are stronger than neutral groups such as sp^2 or sp^3 carbons. The magnitudes of the values for particular functional groups are generally in accord with previous qualitative estimates.[1,4,6]

It is pleasing to note that the regression analysis leads to a negative coefficient for the term related to the number of degrees of internal rotational freedom. Since the loss of these degrees of freedom on receptor binding results in an entropy loss that contributes in an unfavorable manner to the free energy of binding, then a negative sign in eq 4 would be expected. It is even more pleasing to note that the magnitude of this term agrees very closely with the theoretical value of 13–21 J K^{-1} mol^{-1} (equivalent to 0.9–1.5 kcal/mol^{-1} at 298 K) calculated by Page for the entropy associated with internal rotation in moderate-sized organic molecules.

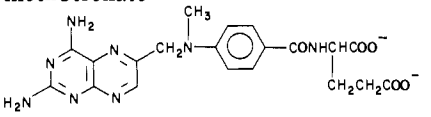
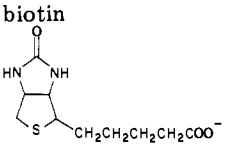
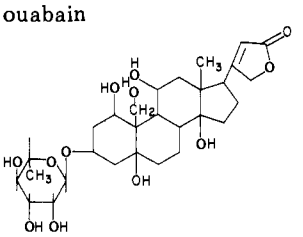
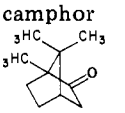
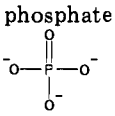
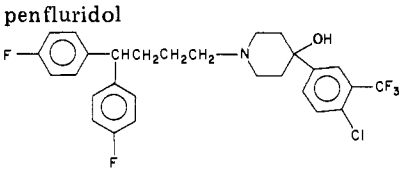
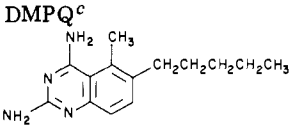
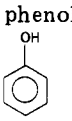
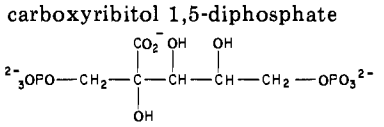
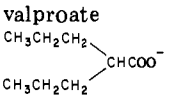
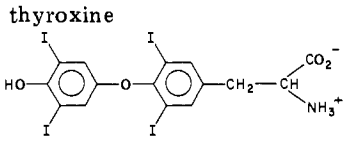
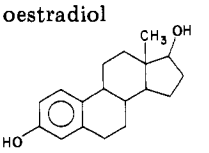
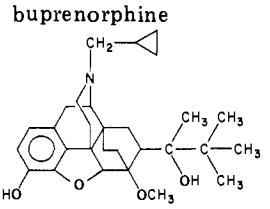
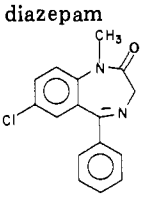
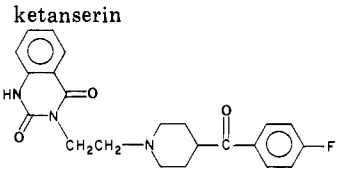
The primary use of AVERAGE binding energies derived from eq 4 is to deduce whether a particular drug represents a good or bad match to its receptor. Table III summarizes the structures that have particularly large deviations from average binding and leads to the following observations.

(1) The drugs with large positive deviations, i.e., those that are exceptionally well matched to their receptors, have a wide range of structural and electronic properties, ranging from relatively nonpolar molecules (e.g., camphor) to polar (e.g., diazepam) and charged (e.g., biotin) species. This implies that significantly higher than AVERAGE binding may be achieved for both van der Waals interactions and other noncovalent bond types when there is a close fit between the drug and the receptor.

(2) While charged groups are not essential for tight binding, they certainly enhance binding if geometrical requirements can be accommodated. This is particularly

(12) In the gas phase, an average-sized drug molecule has a translational entropy of 32.3 cal deg^{-1} mol^{-1} (135 J K^{-1} mol^{-1}) and a rotational entropy of 23.9 cal deg^{-1} mol^{-1} (100 J K^{-1} mol^{-1}). (Values in parentheses are those quoted by Page in Table 2 of ref 6.) A correction of −9.6 cal deg^{-1} mol^{-1} (−40 J K^{-1} mol^{-1}) must be applied to account for the transition from the gas phase to solution at a standard concentration of 1 M. Summing these three terms yields a value of 46.6 cal deg^{-1} mol^{-1} for the overall translational and rotational entropy of a drug in solution under standard conditions. It is difficult to estimate similar entropy terms for a large receptor protein; however, the derived value is unlikely to change significantly on drug binding and hence the net loss of entropy for the whole system of drug and receptor on binding will be approximately 46.6 cal deg^{-1} mol^{-1}. At 298 K, this corresponds to 14 kcal mol^{-1}.

Table III. A Selection of Compounds Calculated To Be either Poor or Exceptional Fits to Their Respective Binding Sites

poor fit compounds			exceptional fit compounds		
no.[a]	name and structure	diff,[b] kcal/mol	no.[a]	name and structure	diff,[b] kcal/mol
116	methotrexate 	-17.6	124	biotin 	+16.4
175	ouabain 	-15.2	171	camphor 	+14.8
			170	phosphate 	+11.7
51	penfluridol 	-12.4	118	DMPQ[c] 	+11.4
			135	phenol 	+10.7
90	carboxyribitol 1,5-diphosphate 	-9.5			
			190	valproate 	+9.7
181	thyroxine 	-9.4			
			156	oestradiol 	+9.8
110	buprenorphine 	-9.1			
			64	diazepam 	+8.5
42	ketanserin 	-9.0			
			75	oxalate 	+6.1

[a] Refers to compound number in Table I. [b] Difference between the observed and AVERAGE binding energies. [c] 2,4-Diamino-5-methyl-6-pentylquinazoline.

evident for the case of unhindered charged groups such as the phosphate anion, which has an observed binding almost double the AVERAGE binding energy for a phosphate group in a normal molecular environment. This is due in part to the increased charge on the phosphate anion relative to organic phosphate esters from which phosphate binding energies were derived, and in part to the unhindered nature of the free anion relative to phosphate esters. Similarly, the data in Table I for the oxalate anion suggest that both carboxyl groups are forming significantly stronger bonds (11 kcal/mol) than the AVERAGE value (8 kcal/mol). The binding of valproate also indicates the potential for stronger bonds involving carboxyl groups.

(3) The group of molecules with large negative deviations, i.e., those that are poor fits to their binding sites, are generally larger molecules with several degrees of conformational freedom and numerous polar groups. Noteworthy among these is the potent dihydrofolate reductase (DHFR) inhibitor methotrexate, which, despite an impressive dissociation constant of approximately 10^{-11} M, gives the worst receptor match of all of the compounds studied. Indeed, if all of the substituents in the methotrexate molecule were binding to DHFR, then an AVERAGE binding energy of 31 kcal/mol would be expected. This binding energy corresponds to an inhibition constant of 10^{-23} M. Obviously, all the structural features of methotrexate are not utilized in the binding of the molecule. Furthermore, it seems likely that the bound form of methotrexate is a high-energy conformation with a portion of the binding energy being utilized to achieve this form. These conclusions are consistent with the results of po-

Table IV. Calculation of AVERAGE Binding Energies for Representative CNS Drugs

	structural coding and energy contributions (kcal/mol)												binding energy, kcal/mol		
compd	$T\Delta S_{rt}$ −14.0	n_{DOF} −0.7	$n_{C(sp^2)}$ 0.7	$n_{C(sp^3)}$ 0.8	n_{N^+} 11.5	n_N 1.2	$n_{CO_2^-}$ 8.2	$n_{PO_4^{2-}}$ 10.0	n_{OH} 2.5	$n_{C=O}$ 3.4	$n_{O,S}$ 1.1	n_X 1.3	av[a]	actual[b]	diff[c]
morphine	−14.0	−1.4	5.6	7.2	11.5				5.0		1.1		15.0	11.2	−3.8
butaclamol	−14.0	−1.4	8.4	10.4	11.5				2.5				17.4	12.2	−5.2
desmethyldiazepam	−14.0	−0.7	9.1	0.8		2.4				3.4		1.3	2.3	11.0	+8.7

[a-c] See corresponding footnotes in Table I.

tential energy calculations,[13] X-ray structures,[14] and recent synthetic studies.[15]

(4) In contrast to methotrexate, another DHFR inhibitor, 2,4-diamino-5-methyl-6-pentylquinazoline, appears to be a very good fit to the binding site of the enzyme. The AVERAGE binding energy for this molecule is 7.3 kcal/mol (equivalent to a K_I of 10^{-6} M), yet its measured inhibition constant is 1.8×10^{-14} M (equivalent to a binding energy of 18.7 kcal/mol). It would seem that all the functional groups in this molecule are likely to be interacting with the active site, and it should therefore be a valuable lead compound for the design of new DHFR inhibitors.

The foregoing examples suggest some general principles for the application of intrinsic binding energies in drug design and structure–activity relationships. The first of these is that if the observed binding of a molecule is significantly stronger than the calculated AVERAGE binding energy, we may deduce that most functional groups in the drug molecule are interacting favorably with the receptor and that the drug probably acts in a low-energy conformation. Such a molecule should therefore provide a useful starting point in drug design. If, on the other hand, the observed binding of a drug is weaker than the calculated AVERAGE binding energy, we may conclude either that the functional groups in the drug are not all interacting with the receptor or that the drug is acting in a relatively high-energy conformation. This situation would therefore demand the synthesis of rigid analogues and/or the progressive deletion of functional groups to determine which are actually involved in binding.[14]

To illustrate these principles, we will consider three CNS active drugs not included among the 200 compounds in Table I, viz., morphine, butaclamol, and desmethyldiazepam. The calculation of the AVERAGE binding energies in these molecules is given, together with their observed dissociation constants, in Table IV. In the case of morphine, the observed binding energy is close to, but less than, the AVERAGE binding energy. Since morphine is a rigid molecule, none of its potential binding energy is being used in conversion to a higher energy conformation, and the data therefore suggest that morphine is making somewhat less than average use of the intrinsic binding potential of its functional groups. This finding is fully consistent with the fact that substantial portions of the morphine structure can be removed without significantly reducing binding. A similar situation applies to butaclamol, where the AVERAGE binding energy is 5 kcal/mol greater than that observed. Since butaclamol is also a comparatively rigid molecule, this implies that a significant portion of the molecule is not interacting with the receptor.

Table V. AVERAGE Binding Energies of Representative Amino Acid Residues

classification	name	structure	AVERAGE binding energy, kcal/mol
nonpolar	glycine	$HCHCO$ (NH)	4.0
	leucine	$(CH_3)_2CHCH_2CHCO$ (NH)	5.8
polar	serine	$HOCH_2CHCO$ (NH)	5.9
	tyrosine	$HOC_6H_4CH_2CHCO$ (NH)	9.4
charged	glutamic acid	$HOOCCH_2CH_2CHCO$ (NH)	11.7
	lysine	$H_3\overset{+}{N}CH_2CH_2CH_2CH_2CHCO$ (NH)	15.2

Thus butaclamol, which has been widely used in dopamine receptor mapping studies,[17] may represent a poor guide in the development of dopamine receptor antagonists.

In contrast to morphine and butaclamol, the calculated AVERAGE binding energy for desmethyldiazepam is some 9 kcal/mol less than that observed. It is thus very likely that desmethyldiazepam acts in a low-energy conformation and that the basic benzodiazepine structure contains the optimal arrangement of receptor binding substituents. In this context it is noteworthy that no major alteration in the benzodiazepine structure over the last 10 years has resulted in a significant increase in potency.

A further application of the concepts developed here is to the emerging class of biologically active peptides. Table V gives the AVERAGE binding energy for a series of representative amino acids *within* a polypeptide molecule (i.e., calculated from the intrinsic binding energies in Table II without allowing for charged carboxyl or amino terminals). The AVERAGE binding energy of a cyclic peptide can be calculated simply by summing the individual binding energies of the component amino acids and subtracting the 14 kcal/mol allowance for loss of overall rotational and translational entropy. The calculation for noncyclic peptides is similar, except that a further 15.1 kcal/mol must be added to account for the charged terminal groups. In the case of large peptides such as insulin, this calculation clearly results in massive AVERAGE binding energies. Indeed, the data in Table V suggest that the observed binding of virtually any biologically active

(13) M. J. Spark, D. A. Winkler, and P. R. Andrews, *Int. J. Quant. Chem., Quant. Biol. Symp.*, **9**, 321 (1982).
(14) D. A. Matthews, R. A. Alden, J. T. Bolin, D. J. Filman, S. T. Freer, R. Hamlin, W. G. J. Hol, R. Kisliuk, E. J. Pastore, L. T. Plante, N. Xuong, and J. Kraut, *J. Biol. Chem.*, **253**, 6946 (1978).
(15) J. R. Piper, J. A. Montgomery, F. M. Sirotnak, and P. L. Chello, *J. Med. Chem.* **25**, 182 (1982).
(16) We stress that in this paper we examine only one aspect of drug action, that of optimal receptor binding. In many cases, those parts of a drug that do not contribute to binding may, in fact, be necessary for other purposes, such as transport or chemical stability. Nevertheless, binding and receptor matching are fundamental to drug action, and any procedure that enhances our understanding of these features should have wide application.
(17) L. G. Humber, "Chronicles of Drug Discovery", J. S. Bindra, D. Lednicer, Eds., Wiley, New York, 1982, Vol. 1, Chapter 3.

peptide can be accounted for by interaction with no more than four or five amino acid residues. In Leu-enkephalin, for example, the observed binding can be explained solely in terms of the tyrosine and phenylalanine residues, in complete accord with the conclusions of structure–activity data.[18,19]

Registry No. 1, 50-53-3; 2, 50-48-6; 3, 1668-19-5; 4, 438-60-8; 5, 91-84-9; 6, 113-92-8; 7, 486-12-4; 8, 486-16-8; 9, 5786-21-0; 10, 24219-97-4; 11, 50-47-5; 12, 53179-07-0; 13, 613-67-2; 14, 835-31-4; 15, 1491-59-4; 16, 50-60-2; 17, 51-41-2; 18, 59-39-2; 19, 4205-90-7; 20, 66711-21-5; 21, 59803-99-5; 22, 31036-80-3; 23, 31428-61-2; 24, 76280-95-0; 25, 28125-87-3; 26, 5051-62-7; 27, 24248-22-4; 28, 29110-47-2; 29, 7361-61-7; 30, 146-48-5; 31, 7762-32-5; 32, 19216-56-9; 33, 26844-12-2; 34, 525-66-6; 35, 26839-75-8; 36, 13655-52-2; 37, 60106-89-0; 38, 749-02-0; 39, 2062-78-4; 40, 52-86-8; 41, 57808-66-9; 42, 74050-98-9; 43, 15574-96-6; 44, 60085-78-1; 45, 129-03-3; 46, 17692-51-2; 47, 361-37-5; 48, 1893-33-0; 49, 2062-84-2; 50, 1841-19-6; 51, 26864-56-2; 52, 68844-77-9; 53, 24526-64-5; 54, 487-93-4; 55, 50-67-9; 56, 61-54-1; 57, 54-04-6; 58, 50-37-3; 59, 58-00-4; 60, 2709-56-0; 61, 51-61-6; 62, 56-12-2; 63, 2763-96-4; 64, 439-14-5; 65, 846-49-1; 66, 1812-30-2; 67, 115-38-8; 68, 50-11-3; 69, 2964-06-9; 70, 22173-64-4; 71, 17617-45-7; 72, 2571-22-4; 73, 57-41-0; 74, 298-46-4; 75, 144-62-7; 76, 884-33-3; 77, 501-52-0; 78, 55700-98-6; 79, 75-39-8; 80, 64-19-7; 81, 75-04-7; 83, 18771-50-1; 84, 32017-56-4; 85, 5699-58-1; 86, 51528-59-7; 87, 13147-57-4; 88, 35752-42-2; 89, 75521-69-6; 90, 27442-42-8; 91, 51321-79-0; 92, 60949-21-5; 93, 11033-22-0; 94, 92315-28-1; 95, 13484-63-4; 96, 92315-29-2; 97, 63250-34-0; 98, 315-30-0; 99, 2465-59-0; 100, 59708-52-0; 101, 53758-22-8; 102, 56030-54-7; 103, 59708-47-3; 104, 6440-26-2; 105, 357-56-2; 106, 76-99-3; 107, 4310-87-6; 108, 469-79-4; 109, 14521-96-1; 110, 52485-79-7; 111, 36292-69-0; 112, 58239-89-7; 113, 51395-54-1; 114, 51583-02-9; 115, 58-14-0; 116, 59-05-2; 117, 6015-76-5; 118, 70997-40-9; 119, 47035-30-3; 120, 7761-45-7; 121, 738-70-5; 122, 138-81-8; 123, 60698-89-7; 124, 58-85-5; 125, 73-32-5; 126, 96-15-1; 127, 64-04-0; 128, 300-62-9; 129, 16088-07-6; 130, 5241-58-7; 131, 4754-39-6; 132, 92418-74-1; 133, 52-52-8; 134, 4385-91-5; 135, 108-95-2; 136, 51-55-8; 137, 63-75-2; 138, 86-13-5; 139, 51-34-3; 140, 50-52-2; 141, 92-13-7; 142, 21888-98-2; 143, 70-22-4; 144, 485-35-8; 145, 54-11-5; 146, 51-84-3; 147, 51-83-2; 148, 363-24-6; 149, 745-65-3; 150, 551-11-1; 151, 77-92-9; 152, 97-67-6; 153, 528-44-9; 154, 57-83-0; 155, 521-18-6; 156, 50-28-2; 157, 50-02-2; 158, 71-58-9; 159, 521-18-6; 160, 50-22-6; 161, 52-39-1; 162, 10540-29-1; 166, 19993-20-5; 167, 488-69-7; 168, 643-13-0; 169, 63-37-6; 170, 14265-44-2; 171, 76-22-2; 172, 50-99-7; 173, 390-64-7; 174, 298-57-7; 175, 630-60-4; 176, 38838-26-5; 177, 52978-30-0; 178, 51-24-1; 179, 2055-97-2; 180, 67-30-1; 181, 51-48-9; 182, 34645-84-6; 183, 77182-38-8; 184, 15307-86-5; 185, 15687-27-1; 186, 22204-53-1; 187, 5104-49-4; 188, 40013-87-4; 189, 17413-79-5; 190, 99-66-1; 191, 66-76-2; 192, 81-81-2; 193, 15074-17-6; 194, 50-29-3; 195, 60-57-1; 196, 58-89-9; 197, 63-25-2; 198, 1563-66-2; 199, 50-33-9; 200, 51-92-3; morphine, 57-27-2; butaclamol, 51152-91-1; desmethyldiazepam, 1088-11-5; glycine, 56-40-6; leucine, 61-90-5; serine, 56-45-1; tyrosine, 60-18-4; glutamic acid, 56-85-9; lysine, 56-40-6.

(18) F. A. Corin and G. R. Marshall, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 5179 (1977).
(19) A. P. Feinberg, I. Creese, and S. H. Snyder, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 4215 (1976).

3-Substituent Effect and 3-Methylene Substituent Effect[1] on the Structure–Reactivity Relationship of 7β-(Acylamino)-3-cephem-4-carboxylic Acid Derivatives Studied by Carbon-13 and IR Spectroscopies[2]

Junko Nishikawa* and K. Tori†

Shionogi Research Laboratories, Shionogi & Co., Ltd., Fukushima-ku, Osaka 553, Japan. *Received January 5, 1984*

Relationships between the chemical reactivity of 3-substituted cephalosporins or 3-methylene-substituted cephalosporins and several parameters observed by ^{13}C NMR and IR spectroscopies are described. Among 3-substituted cephalosporins, the values of δ(C-3) and δ(COO) of ^{13}C NMR spectra are correlated with the logarithms of the rate constants for alkaline hydrolysis (log k_{obsd}) when substituents at the 3-position are classified into two groups, i.e., OR substituents and others. Among the 3-methylene-substituted cephalosporins, the difference values of the ^{13}C chemical shifts for C-3 and C-4, Δδ(4–3), are correlated with log k_{obsd}. The β-lactam $\nu_{C=O}$ value of the solution IR spectra is a good index for the prediction of a significant change of the β-lactam reactivity resulting from modification of a 3-substituent or a 3-methylene substituent. From analysis of these observed parameters, both resonance and inductive effects of the substituent at the 3-position were found to affect the chemical reactivity of the β-lactam ring in cephalosporin, while only the inductive effect of the substituent at the 3′-position was found to affect the β-lactam reactivity.

β-Lactam antibiotics, such as penicillins, cephalosporins, and oxacephalosporins, inhibit biosynthesis of bacterial cell walls by acylating and thereby inactivating transpeptidases and carboxypeptidases.[3] Because the antibacterial activity of an antibiotic depends on the acylation of those enzymes by the β-lactam ring of the antibiotic, the chemical reactivity that represents the acylating ability of the β-lactam ring is an important factor affecting the antibacterial activity. Thus, much interest has been attached to investigation of the structure–reactivity relationship of cephalosporins and penicillins as the first stage in the prediction of antibacterial activity.

A number of parameters have been proposed as indicators of the β-lactam reactivity, for example, the IR

† Deceased.

(1) Cephalosporins were examined from two viewpoints according to the structure of the substituent at the 3-position, i.e., one with the substituent at the 3-position (called the 3-substituted cephalosporin in this paper) and the other with a methylene group at the 3-position (called 3-methylene-substituted cephalosporin in this paper). Here, we define 3-methylene substituent as a group at the 3′-position of a cephalosporin (i.e., the substituent is on the methylene at the 3-position) and a 3-substituent as a whole group, e.g., CH_2R or R′, i.e., a direct 3-substituent, attached to the 3-position of either a cephalosporin or a direct 3-substituted cephalosporin, respectively.
(2) Some part of this study has been reported in a communication: Nishikawa, J.; Tori, K. *J. Antibiot.* 1981, *34*, 1641.

0022-2623/84/1827-1657$01.50/0