CROATICA CHEMICA ACTA CCACAA **66** (1) 129–140 (1993)

ISSN 0011–1643
CCA–2124

Original Scientific Paper

Application of the Average Molecular Electrostatic Field in Quantitative Structure-Activity Relationships

Tibor Balogh and Gábor Náray-Szabó

Theoretical Chemistry Group, Department of Chemistry, Eötvös University Budapest, P. O. Box 32, H–1518 Budapest 112, Hungary

Received September 17, 1992

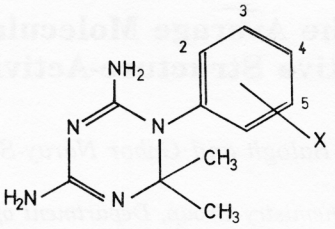
It is proposed to use the average molecular electrostatic field, F, as a descriptor in quantitative structure-activity relationships. F is thought to be proportional to the hydration ability of the molecule with large and small values corresponding to strong and weak hydration, respectively. QSAR equations, containing F, the molecular surface and the Coulombic interaction energy with the enzyme are derived to estimate catalytic efficiencies of various substrates of point mutants of subtilisin and to predict inhibitory potencies of substituted *s*-triazine derivatives on chicken liver dihydrofolate reductase.

INTRODUCTION

The three primary conditions for the successful binding of a ligand at the active site of a protein are steric, electrostatic (Coulombic) and hydrophobic complementarity.[1] The steric fit is the most important and it can be best illustrated by the lock-and-key analogy, even if considering the role of induced fit and protein dynamics. Coulombic complementarity corresponds to the electrostatic matching between the biopolymer and the ligand ensuring maximum attractive interaction. Hydrophobic complementarity represents minimization of dehydration free energies and can be formulated as the matching between regions of host and guest of similar polarity. A possibility to characterize the polarity of a certain region is the use of the electrostatic field, F, on the corresponding van der Waals surface. Regions with a large field strongly attract water molecules represented by point dipoles, while small field regions do not attract them and are therefore considered hydrophobic. The average value of F is considered to be proportional to the overall hydration ability of the molecule and, thus, a candidate for substitution of the π hydrophobic parameter of Hansch.[2]

Two examples will be presented in this paper, where F is used as a descriptor for hydration-dehydration effects governing enzyme-ligand interactions. The first deals with catalytic efficiencies ($\log k_{cat}/K_M$) for a number of subtilisin double mutants *vs.*

tetrapeptide substrates. The Glu-156 and/or Gly-166 side chains in the mutants while the P_1 subsite in the substrate were replaced by various amino acids.[3] Our second example is based on an extensive study by Hansch and coworkers on 4,6-diamino-1,2-dihydro-2,2′-dimethyl-1-(substituted phenyl)- *s*-triazine inhibitors (see **I** for the general formula) of chicken liver dihydrofolate reductase (DHFR).[4]



I

MODELS AND METHODS

In this study we do not deal with steric complementarity and consider the geometry fit between all ligands and the corresponding enzyme crevice as perfect and not influencing binding. DHFR inhibitors of which this is clearly not true were handled separately also by Hansch *et al.*[4] and dropped from the QSAR data set. Similarly, we also dropped combinations of mutant subtilisin and substrate with a steric conflict between side chains and, thus, used truncated sets in both studies. Coulombic effects were treated for subtilisin by calculating the variation of the electrostatic interaction energy between the ligands and the enzyme but no such calculations were done for the triazine inhibitors of DHFR.

In both cases, we focused on hydrophobic complementarity, which was quantified using F as the descriptor mainly responsible for the variation of the ligand-binding power. We got F as the average of the field vector lengths calculated on the dots of the van der Waals surface (for amino acid side chains) or in the reference points (for triazine derivatives). Further studies are needed to clarify whether hydration ability is sensitive or not to the orientation of the field vector on the dots or in reference points with respect to the molecular environment. Since we use F as a QSAR descriptor accounting for global hydration of the molecule, the problem does not seem to be of primary importance for the time being.

While F is assumed to account for the part of the dehydration energy change that comes from the breaking of the hydrogen bonds between solvent water molecules and the substituent, it does not describe the effect of cavity formation, which may be expressed as a function of the molecular surface (*cf. e.g.* Refs. 5 and 6). This quantity was calculated by the PCMODEL program.[7] Besides the overall molecular surface (S), PCMODEL calculates saturated apolar (S_{sa}), unsaturated apolar (S_{ua}) and polar (S_p) contributions, as well; these quantities were also used as descriptors. Note that $S_{sa} + S_{ua} + S_p = S$. Linear regression studies were done using the DrugIdea program.[8]

Subtilisin Mutants

Average molecular electrostatic fields for amino-acid side chains were calculated by using a set of atomic monopoles with charges fitted by Kollman and coworkers to *ab initio* quantum chemical electrostatic potentials.[9] A microcomputer program was written[10] for the generation of molecular van der Waals surface points as obtained by the Connolly algorithm.[11] We calculated electrostatic fields in 200 points on the dot surface from Coulomb's law, considering all atomic monopoles, and averaged them to get F values in Table I. For the calculation of side-chain surfaces we used a space increment value of 0.05 A and obtained the following values for S_{sa}, S_{sp} and S (note that S_{ua} was zero for all side chains considered in this study). Gly: 0, 0, 0 (by definition), Ala: 64.0, 0, 64; Asn: 46.4, 56.8, 103; Asp: 46.2, 50.5, 97; Gln: 71.5, 52.3, 124; Glu: 71.8, 48.2, 120; Ser: 49.4, 22.7, 72; Met: 134.9, 0, 135; Lys: 119.6, 24.3, 144.

Variation of the Coulombic interaction energy (E_{Coul}) was calculated using a three-dimensional model of the tetrahedral intermediate of the subtilisin-substrate complex obtained by molecular graphics (*cf.* Figure 1). We docked the substrate with Lys at site P_1 into the enzyme crevice (coordinates from the Protein Data Bank[12]); while no energy optimization was performed, only steric conflicts were avoided. Once the P_1 = Lys substrate was in place, other substrates were modeled simply by replacing Lys by Glu, Gln or Met, the side chains considered in the study of Wells and coworkers.[3] Pairs of mutant enzymes and substrates both possessing bulky side chains (Lys or Met) at the active site and also studied by Wells *et al.*,[3] were dropped from our data set because of steric conflicts not manageable in the present phase of investigation.

The Coulombic energy (E_{Coul}) was calculated as the interaction of the protein electrostatic potential with the monopoles representing the substrate. The former was obtained using our Bond Increment (BI) method[13], the latter from CNDO semiempirical molecular orbital calculations.[14] Since the BI method calculates the electrostatic potential as a simple sum of bond contributions with no polarization effects taken into account, the variation of E_{Coul} could be estimated by calculating the potential generated solely by the variable side chains and neglecting the protein environment which contributes to the overall value by a constant. We considered side chains Asp, Glu and Lys as ionized bearing −1, −1 and +1 electron unit charges, respectively.

TABLE I

Electrostatic fields averaged on the van der Waals surface of amino-acid side chains (V/nm)

side chain	F	side chain	F	side chain	F
charged		*polar*		*apolar*	
Asp^-	26.7	Asn	12.7	Met	5.6
Glu^-	22.8	Ser	11.8	Tyr	5.0
Lys^+	18.1	Arg(neutral)	11.8	Trp	4.4
Arg^+	15.8	Gln	10.9	Ala	1.3
		Cys	10.2	Val	1.3
		Thr	9.7	Leu	1.3
		His	9.0	Ile	1.2
		Lys(neutral)	7.3	Phe	0.6
				Gly	0.0

Triazine Inhibitors of DHFR

Hansch *et al.*[4] treated 114 different *s*-triazine derivatives in their QSAR equation where they used the hydrophobic parameter, obtained from group contributions, and the Hammett constant as descriptors. They found 12 outliers not obeying their equation, for which they gave reasonable steric or other explanation. We dropped these and six further derivatives [3–Br, 4–Br, 3–I, 4–I, 3–$CH_2SeC_6H_5$, and 4–$CSi(CH_3)_3$] for which we could not calculate F within our approximation. Since the hydrogen substituent was considered as a reference, it was also dropped from the data set. We derived multiple linear regression equations for the remaining 95 congeners.

Since electrostatic potential derived atomic charges were not available for this class of molecules, we adapted another methodology for the calculation of F values. Based on the BI method[13], we determined fields in certain reference points around substituent X. These are located along the AH bond (A=C,N,O) and hypothetical lone-pair directions at a distance from the van der Waals radius.[15] We postulate that AH bonds and lone pairs are fully transferable *i.e.* the value of F in a certain reference point depends only on the adjacent bond type and is independent of the other bonds in the molecule. The transferable bond increments of the molecular electrostatic field are displayed in Table II.

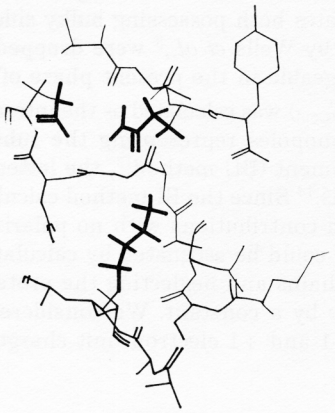


Figure 1. Geometric model of the tetrahedral intermediate of the subtilisin-substrate complex. Variable side chains are indicated by heavy lines.

TABLE II

Electrostatic field increments (Fi *in* V/*nm; for calculation see text), lp denotes lone pair*

bond type	*Fi*	bond type	*Fi*
C(sp^2 or sp^3)H	11.6	O(sp^3) lone pair	25.6
N(sp^2 or sp^3)H	17.5	S(sp^2) lone pair	15.5
O(sp^3)H	22.4	S(sp^3) lone pair	12.9
N(sp) lone pair	43.9	F(sp^3) lone pair	16.9
N(sp^3) lone pair	33.9	Cl(sp^3) lone pair	11.3
O(sp^2) lone pair	31.6	phenyl correction	–4.6

We calculated F for the substituent by averaging increments over the whole moiety, considering one reference point per each AH bond or N atom, two per each O or S atom and three per each F or Cl atom, respectively. A correction was introduced for substituents containing the phenyl substituent. Molecular surfaces were calculated keeping the 4,6-diamino-1,2-dihydro-2,2′-dimethyl-1-phenyl-*s*-triazine fragment in a fixed orientation, optimizing substituent geometries *in vacuo* and using a spacing increment value of 0.25 A.

RESULTS AND DISCUSSION

Subtilisin Mutants

In a previous paper[16] we proposed the following equation for the determination of specificities and Michaelis constants of subtilisin mutants *vs.* tetrapeptide substrates

$$\log X = a.F(E) + b.F(S) + c.q.F^{1/2}(E).F^{1/2}(S) + d \qquad (1)$$

where $X = k_{cat}/K_M$ or $1/K_M$, E and S refer to the enzyme side chains mutated and the substrate P_1 subsite, respectively. $F(E)$ was obtained as an arithmetic mean between average fields of Table I for the side chains in position 156 and 166. Purely heuristically we put $q = -1$ if both E and S are oppositely charged and $q = +1$ otherwise. Eq. (1) gives good correlation both for subtilisin double mutants treated in this study and trypsin Asp–189 mutants not used in the derivation of its analytical form. With regression parameters fitted separately to various sets of experimental $\log k_{cat}/K_M$ and $\log 1/K_M$ values the following correlation coefficients and standard errors were obtained (number of data points in parentheses): subtilisin k_{cat}/K_M 0.964, 0.35 (47); subtilisin $1/K_M$ 0.908, 0.31 (47); trypsin k_{cat}/K_M 0.993, 0.29 (7); trypsin $1/K_M$ 0.910, 0.23 (7).

In Eq. (1) the first two terms were taken to account for dehydration effects while the product term with the square root of F may stand for some approximate interaction energy between enzyme and substrate.[16] In order to refine our model, we considered the Coulombic interaction energy (E_{Coul}) in place of the above product term (see above). Using this term alone is insufficient for the description of mutation effects; only the changes due to charged side chains could be accounted for (*cf.* Table III and Figure 2).

In order to refine our regression equation, we considered F values and molecular surfaces for side chains, as well. Performing a stepwise regression analysis, we found the following equation as best reproducing specifities

$$\begin{aligned} \log k_{cat}/K_M = &- 0.0318(\pm 0.0103)F(156) - 0.0337(\pm 0.0059)F(166) - \\ &- 0.1649(\pm 0.0120)F(P1) - 0.0322(\pm 0.0049)S_{sa}(P1) - \\ &- 0.0031(\pm 0.0004)E_{Coul} + 11.33 \end{aligned} \qquad (2)$$

$$n = 47 \quad r = 0.957 \quad s = 0.41 \quad F = 90.3$$

$F(156)$, $F(166)$ and $F(P1)$ stand for average side-chain fields in the corresponding protein and substrate position. This equation is only slightly worse than Eq. (1) for which r and s are 0.964 and 0.35, respectively, but it is based on a much firmer theoretical basis. The calculated values are given in Table III and plotted in Figure 3.

TABLE III

Coulombic interaction energies (kcal/mol, upper row), by Eq. (2) estimated (middle row) and experimental (lower row) logK_{cat} / K_M values for Glu/X—156, Gly/Y—166 subtilisin double mutants acting on Ala—Z—Ala—Ala tetrapeptide substrates. X, Y and Z mean variable amino acid side chains

mutant	substrate			
X – Y	Z = Glu	Z = Gln	Z = Met	Z = Lys
	398	–19	1	–441
Glu-Asp	–	3.36	4.43	4.25
	–	3.02	3.81	4.21
	482	–32	1	–553
Glu-Glu	–	3.53	4.56	4.72
	–	3.06	3.86	4.48
	188	–9		–180
Glu-Asn	1.22	3.80	4.90	3.92
	1.62	3.85	5.02	4.25
	244	–19	11	–268
Glu-Gln	1.11	3.89	4.97	4.25
	1.20	4.36	5.54	4.10
	171	–5	0	–200
Gln-Asp	1.18	3.69	4.81	3.89
	1.30	3.40	5.03	4.41
	202	–9	0	–245
Ser-Asp	1.06	3.67	4.78	4.00
	1.23	3.41	4.67	4.24
	228	–16	1	–239
Glu-Met	1.34	4.06	–	–
	1.20	3.89	5.64	4.70
	204	–11	1	–204
Glu-Ala	–	4.19	5.29	4.38
	–	4.34	5.65	4.90
	204	–11	1	–204
Glu-Gly	1.60	4.23	5.33	4.42
	1.54	3.95	5.15	4.60
	26	–3	0	–28
Gln-Gly	2.53	4.58	5.71	4.26
	2.79	4.71	5.48	3.03
	7	–1	0	–8
Ser-Gly	2.55	4.55	5.69	4.17
	2.59	4.38	5.77	3.37
	17	–1	0	–17
Gln-Asn	2.12	4.15	5.28	3.80
	2.04	4.51	5.95	3.75
	18	0	0	–25
Ser-Asn	2.09	4.12	5.26	3.80
	1.91	4.57	5.72	3.68

It is interesting to notice that the coefficients of enzyme side chain fields, $F(156)$ and $F(166)$ are only 20% that of $F(P1)$, which indicates their reduced importance in

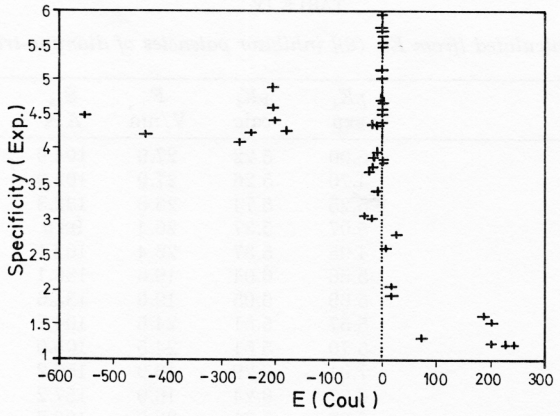


Figure 2. Plot of log k_{cat}/K_M *vs.* E_{Coul} for subtilisin double mutants.

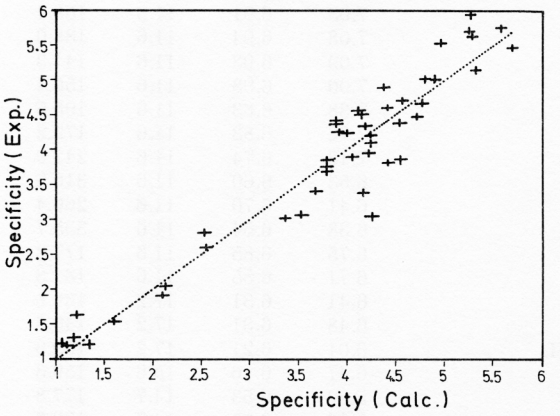


Figure 3. Plot of experimental *vs.* from Eq. (2) calculated log k_{cat}/K_M values for subtilisin double mutants.

determining log k_{cat}/K_M. Indeed, if we drop these descriptors from Eq. (2), we get the following one with only slightly worse statistical parameters.

$$\log k_{cat}/K_M = -0.1519(\pm 0.0160)F(P1) - 0.0173(\pm 0.0055)S_{sa}(P1) - \\ - 0.0027(\pm 0.0006)E_{Coul} + 8.33 \quad (3)$$

$$n = 47 \; r = 0.915 \; s = 0.55 \; F = 73.8$$

A straightforward explanation of the adequacy of Eq. (3) is that enzyme side chains in the specifity pocket are not hydrated before associating with the substrate. Therefore, their dehydration, supposedly accounted for by $F(156)$ and $F(166)$, does not

TABLE IV

Experimental and calculated [from Eq. (6)] inhibitor potencies of diamino-triazine inhibitors

No	Substituent	pK_i exp	pK_i calc	F V/nm	S_{sa} A.A	S A.A.	V
1	3–SO_2NH_2	5.00	5.22	27.9	104.9	275	6
2	4–SO_2NH_2	4.70	5.26	27.9	102.0	280	6
3	4–SO_2CH_3	5.25	5.74	23.0	132.3	296	7
4	3–$CONH_2$	5.07	5.37	26.4	98.0	254	5
5	4–$CONH_2$	4.95	5.37	26.4	102.1	262	5
6	3–$COCH_3$	5.56	6.04	19.6	136.1	268	6
7	4–$COCH_3$	5.69	6.05	19.6	133.6	267	6
8	3–OH	5.57	5.51	24.5	104.7	238	2
9	4–OH	5.70	5.51	24.5	105.0	238	2
10	3–CF_3	7.01	6.29	16.9	150.2	268	4
11	4–CF_3	6.77	6.24	16.9	157.2	263	4
12	4–NH_2	5.67	5.71	23.0	103.7	241	3
13	3–F	6.79	6.33	16.9	124.3	236	1
14	4–F	6.89	6.33	16.9	123.7	234	1
15	3–Cl	7.36	6.99	11.3	136.6	247	1
16	4–Cl	6.95	6.97	11.3	136.5	243	1
17	3,5–Cl_2	7.03	6.91	11.3	158.4	261	2
18	3–CH_3	7.08	6.94	11.6	138.0	247	4
19	4–CH_3	7.09	6.93	11.6	141.3	249	4
20	3–CH_2CH_3	7.00	6.93	11.6	156.4	274	7
21	4–$(CH_2)_3CH_3$	7.38	6.83	11.6	195.6	309	13
22	3,4–$(CH_2)_4$	7.72	6.83	11.6	179.2	283	12
23	3–$(CH_2)_5CH_3$	7.12	6.74	11.6	242.9	361	19
24	3–$(CH_2)_8CH_3$	6.53	6.60	11.6	310.6	429	28
25	4–$(CH_2)_8CH_3$	6.41	6.70	11.6	266.4	388	28
26	3–$(CH_2)_{11}CH_3$	6.38	6.54	11.6	335.7	453	37
27	3–$C(CH_3)_3$	6.75	6.85	11.6	177.3	286	13
28	4–$C(CH_3)_3$	6.71	6.85	11.6	181.3	292	13
29	3–OCH_3	6.41	6.31	17.2	139.5	265	5
30	4–OCH_3	6.48	6.31	17.2	140.6	267	5
31	3–OCH_3,4–OCH_3	6.01	6.21	17.2	169.9	286	10
32	3–OCH_2CH_3	6.47	6.45	15.6	159.3	283	8
33	3–$O(CH_2)_2CH_3$	5.92	6.53	14.7	177.8	302	11
34	4–$O(CH_2)_3CH_3$	5.90	6.55	14.7	179.8	297	11
35	3–$O(CH_2)_3CH_3$	6.20	6.52	14.1	215.2	340	14
36	3–$O(CH_2)_4CH_3$	6.28	6.50	13.8	229.2	345	17
37	3–$O(CH_2)_5CH_3$	6.30	6.54	13.5	240.4	366	20
38	4–$O(CH_2)_5CH_3$	6.46	6.57	13.5	225.8	350	20
39	3–$O(CH_2)_8CH_3$	6.55	6.48	12.9	300.2	425	29
40	3–$O(CH_2)_{10}CH_3$	6.56	6.42	12.7	337.5	462	35
41	4–$O(CH_2)_{10}CH_3$	6.03	6.49	12.7	307.3	432	35
42	3–$O(CH_2)_{11}CH_3$	6.38	6.38	12.6	362.1	487	38
43	4–$OCH_2)_{11}CH_3$	6.50	6.45	12.6	328.1	450	38
44	3–$O(CH_2)_{12}CH_3$	5.48	6.34	12.6	380.0	504	41
45	3–$O(CH_2)_{13}CH_3$	6.50	6.30	12.5	407.0	532	44
46	4–$O(CH_2)_2OC_6H_4$–4′–NH_2	6.76	7.03	13.0	166.3	367	21
47	3–$OCH_2C_6H_5$	6.93	7.33	10.1	159.0	340	15
48	4–$OCH2C_6H_5$	7.53	7.41	10.1	145.1	340	15
49	3–$OCH_2)_2OC_6H_5$	7.15	7.18	11.3	177.8	369	19
50	3–$O(CH_2)_2OC_6H_4$–3′–CH_3	7.02	7.15	10.7	210.8	394	22
51	3–$O(CH_2)_4OC_6H_5$	7.29	7.23	10.3	219.7	417	25
52	3–$O(CH_2)_4OC_6H_4$–3′–CF_3	7.54	7.03	10.8	260.6	445	28

TABLE IV *(continued)*

No	Substituent	pK_i exp	pK_i calc	F V/nm	S_{sa} A.A	S A.A.	V
53	3-$OCH_2C_6H_3$-3′-4′-Cl_2	6.78	7.35	9.0	198.5	372	16
54	4-$OCH_2C_6H_3$-3′-4′-Cl_2	7.14	7.40	9.0	190.9	372	16
55	3-$OCH_2C_6H_4$-4′-$CONH_2$	7.05	6.92	14.9	150.3	377	19
56	4-$OCH_2C_6H_4$-4′-$CONH_2$	7.30	6.98	14.9	136.9	373	19
57	4-$OCH_2C_6H_4$-4′-SO_2NH_2	7.49	6.85	16.5	140.1	398	20
58	4-$OCH_2C_6H_4$-4′-CH_2OH	7.35	7.12	12.1	169.4	367	19
59	3-CH_2O-c-C_6H_{11}	7.19	6.52	13.5	238.5	356	21
60	3-$CH_2NHC_6H_3$-3′,5′-$(CONH_2)_2$	6.98	6.77	17.4	137.8	401	24
61	3-$CH_2NHC_6H_4$-4′-SO_2NH_2	7.18	6.83	16.5	147.3	402	21
62	3-$CH_2OC_6H_5$	7.28	7.35	10.1	157.3	344	15
63	3-$CH_2OC_6H_4$-3′-Cl	7.18	7.35	9.5	180.2	359	15
64	3-$CH_2OC_6H_4$-3′-CN	7.59	7.04	13.1	161.3	366	16
65	3-$CH_2OC_6H_4$-3′-OCH_3	7.29	7.12	11.3	191.3	375	19
66	3-$CH_2OC_6H_4$-3′-CH_2OH	7.10	7.10	12.1	183.0	385	19
67	3-$CH_2OC_6H_4$-3′-CH_3	7.14	7.31	9.5	189.9	365	18
68	3-$CH_2OC_6H_4$-3′-CH_2CH_3	7.27	7.31	9.2	205.9	379	21
69	3-$CH_2OC_6H_4$-3′-$CH(CH_3)_2$	7.47	7.31	8.9	221.7	393	24
70	3-$CH_2OC_6H_4$-3′-$C(CH_3)_3$	7.24	7.29	8.6	236.9	404	27
71	3-$CH_2OC_6H_4$-3′-C_6H_5	6.79	7.50	9.2	177.0	386	25
72	3-$CH_2OC_6H_4$-3′-$NHCOCH_3$	7.64	6.91	13.4	183.6	376	22
73	3-$CH_2OC_6H_4$-3′-$NHCONH_2$	7.46	6.81	15.7	143.3	364	21
74	3-$CH_2OC_6H_4$-3′-$NHCSNH_2$	7.22	7.12	13.5	143.7	375	21
75	3-$CH_2OC_6H_4$-4′-$(CH_2)_4CH_3$	6.71	7.27	8.5	255.0	422	30
76	3-CH_2O-2-naphtyl	7.50	7.57	9.5	161.5	393	21
77	3-CH_2O-1-naphtyl	7.15	7.46	9.5	163.8	363	21
78	3-$CH_2SC_6H_5$	7.47	7.63	6.5	186.2	343	15
79	4-$CH_2SC_6H_5$	8.17	7.70	6.5	174.0	343	15
80	3-$CH_2SC_6H_4$-3′-CH_3	7.70	7.62	6.6	193.8	358	18
81	4-$CH_2SC_6H_4$-3′-CH_3	7.40	7.68	6.6	193.0	372	18
82	4-$CH_2SC_6H_4$-2′-CH_3	7.37	7.53	6.6	205.7	351	18
83	3-$SCH_2C_6H_5$	7.52	7.79	6.5	164.7	354	15
84	4-$SCH_2C_6H_5$	7.71	7.73	6.5	174.4	354	15
85	3-$SCH_2C_6H_4$-4′-Cl	7.55	7.72	6.5	185.1	368	15
86	4-$SCH_2C_6H_4$-4′-Cl	7.13	7.72	6.5	189.2	373	15
87	3-Cl,4-$OCH_2C_6H_4CON(CH_3)_2$	7.01	6.84	13.3	230.6	430	26
88	3-SO_2NH_2,4-Cl	5.66	5.88	22.9	113.8	302	7
89	3-NH_2,4-CH_2CH_3	6.50	6.49	15.9	142.2	276	10
90	3-$CH_2SC_6H_4$-4′-Cl	7.58	7.72	6.5	185.4	269	16
91	3-Cl,4-$SCH_2C_6H_5$	7.40	7.64	16.0	188.5	351	16
92	3-Cl,4-$CH_2SC_6H_5$	7.33	7.66	16.0	181.9	346	16
93	3-Cl,4-$O(CH_2)_8CH_3$	6.46	6.45	30.0	313.9	432	30
94	3-Cl,4-C_4H_8-C_6H_3-2′-Cl,4′-SO_2F	7.55	7.18	27.0	240.4	459	27

play an important (or any) role in substrate binding. This is understandable since no water molecules are seen in the pocket of substrate-free subtilisin by X-ray crystallography.[12]

Triazine Inhibitors of DHFR

Hansch and coworkers derived the following regression equation for the estimation of pK_i values of substituted triazine derivatives (*cf.* Formula I)

$$pK_i = 0.85(\pm 0.08)\pi - 1.04(\pm 0.14)\log(b \cdot 10^{\pi} + 1) + 0.57(\pm 0.49)\sigma + 6.36 \quad (4)$$

$$n = 101 \quad r = 0.910 \quad s = 0.29$$

where π is the sum of group hydrophobicity constants, b is a disposable parameter and σ is the Hammett constant. In a previous publication[17], we proposed using the average substituent electrostatic field as calculated from bond increments of Table II in a regression equation together with some other variables

$$pK_i = -0.090(\pm 0.010)F_s + 0.129(\pm 0.034)N + 0.884(\pm 0.184)I - 0.010(\pm 0.006)V + 7.88 \quad (5)$$

$$n = 107 \quad r = 0.764 \quad s = 0.56 \quad F = 35.8$$

where $N = N(\mathrm{Y})-N(\mathrm{XH})$ is the difference of the number of potential acceptor and donor atoms in a hydrogen bond with Y=uprotonated N, O, S, F, or Cl and X=N or O. I is an indicator variable with a value of 1 for substituents with a hydrogen-bonding ability and appropriately oriented in the binding pocket of DHFR. V is the number of substituent atoms supposed to be proportional to the molecular surface.

Descriptors N, I and V in Eq. (5) do not have very much physical meaning; they have to be considered as purely heuristic, empirical quantities. Instead, we derived regression equations similar to Eq. (2) with the only difference that E_{Coul} was not considered. Having 95 pK_i values in the data set, all possible forms of regression equations with F, S_{sa}, S_{ua}, S_{p} and S as descriptors failed to reproduce the activity of the 3–NO_2 derivative; it was underestimated by about 2 pK_i units. A possible explanation for this is similar to that given by Hansch *et al.*[4] for the 3–CN congener. It is straightforward to suppose that, like the cyano group, the nitro substituent also interacts with a water molecule bound to the enzyme and does not show up at the present resolution of crystallography. This interaction might enhance the binding power by ensuring a better electrostatic fit between DHFR and the inhibitor. In the following regression equations, we dropped the 3–NO_2 substituent from the data set and reduced it to 94 derivatives. Through stepwise regression analysis, we selected the following QSAR equation

$$pK_i = -0.1222(\pm 0.0082)F_s - 0.0057(\pm 0.0010)S_{\mathrm{sa}} + 0.0035(\pm 0.0010)S + 8.27 \quad (6)$$

$$n = 94 \quad r = 0.872 \quad s = 0.35 \quad F = 95.2$$

which again provides only slightly worse statistical parameters than Eq. (4) by Hansch *et al.*[4] Comparison of the calculated and experimental pK_i values is seen in Figure 4. The advantage of Eq. (6) over Eq. (4) seems to be that F_s can be estimated easily for a great number of derivatives and its approximate calculation based on Table II can be considerably refined. Work in this direction is in progress.

It has to be noticed that the standard error of the regression parameter of σ in Eq. (4) is quite large. Therefore, this term, considered to account for the interaction between the inhibitor molecules and the protein *i.e.* replacing E_{Coul}, can be easily dropped from the descriptors without significantly reducing the predictive power of the regres-

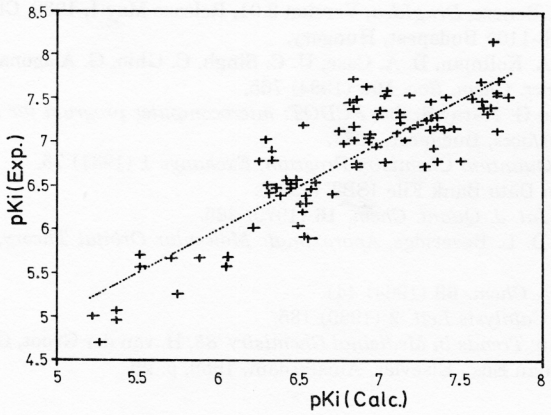


Figure 4. Plot of experimental *vs.* from Eq. (6) calculated pK_i values of substituted diaminotriazine derivatives on chicken liver DHFR.

sion equation. This justifies a posteriori why we did not consider this quantity in Eq. (6) and yet got a fair correlation. If descriptor S is dropped from the regression equation, its significance does not get strongly reduced and we get an expression analogous to Eq. (2).

$$\mathrm{p}K_i = -0.1294(\pm 0.0085)F_s - 0.0030(\pm 0.0007)S_{sa} + 9.07 \qquad (7)$$
$$n = 94 \quad r = 0.851 \quad s = 0.38 \quad F = 119.4$$

As mentioned above, V in Eq. (5) might be proportional to the molecular surface, S, or one of its components, S_{sa}. Indeed, this is the case and we obtain the following equations:

$$S = 6.623(\pm 0.181)V + 236.9 \qquad (8)$$
$$n = 94 \quad r = 0.967 \quad s = 17.3 \quad F = 1344$$

$$S_{sa} = 5.692(\pm 0.341)V + 94.3 \qquad (9)$$
$$n = 94 \quad r = 0.867 \quad s = 32.6 \quad F = 278.9$$

Accordingly, the use of V in Eq. (5) is justified by the present study.

REFERENCES

1. G. Náray-Szabó, *J. Mol. Graphics* **7** (1989) 76.
2. C. Hansch in: *Drug Design*, E. J. Ariens Ed., Academic Press, New York, 1971, p. 271.
3. J. A. Wells, D. B. Powers, R. R. Bott, T. P. Graycar, and D. A. Estell, *Proc. Natl. Acad. Sci. U.S.A.* **84** (1987) 1219.
4. C. Hansch, B. A. Hathaway, Z. R. Guo, C. Dias Selassie, S. W. Dietrich, J. M. Blaney, R. Langridge, K. W. Volz, and B. T. Kaufman, *J. Med. Chem.* **27** (1984) 129.
5. D. Eisenberg and A. D. McLahlan, *Nature* **319** (1986) 199.
6. C. J. Cramer and D. G. Truhlar, *J. Amer. Chem. Soc.* **113** (1991) 8305.
7. *Program POMODEL*, Serena Software, Bloomington, Indiana, USA, 1990.

8. A. Lopata and A. Bencze, *DrugIdea*, Version 2.01, Release May 1, 1989, CheMicro Limited, Salamon u. 13/a, H–1105 Budapest, Hungary.
9. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner, *J. Amer. Chem. Soc.* **106** (1984) 765,
10. I. Kolossváry and G. Náray-Szabó, *PCDOT: microcomputer program for the representation of molecular dot surfaces*, Budapest, 1987.
11. M. L. Connolly, *Quantum Chemistry Program, Exchange* **1** (1981) 75.
12. J. Kraut, Protein Data Bank File 1SBT (1987).
13. G. Náray-Szabó, *Int. J. Quant. Chem.* **16** (1979) 265.
14. J. A. Pople and D. L. Beveridge, *Approximate Molecular Orbital Theory*, McGraw Hill, New York, 1970.
15. A. Bondi, *J. Phys. Chem.* **68** (1964) 441.
16. G. Náray-Szabó, *Catalysis Lett.* **2** (1989) 185.
17. G. Náray-Szabó in: *Trends in Medicinal Chemistry '88*, H. van der Groot, G. Domány, L. Pallos, and H. Timmerman Eds., Elsevier, Amsterdam, 1989, p. 29.

SAŽETAK

Primjena prosječnoga elektrostatskog polja molekule u QSAR

Tibor Balogh i Gábor Náray-Szabó

Predložena je uporaba prosječnoga elektrostatskog polja molekule, F (za koje se uzima da je proporcionalno mogućnosti hidratacije molekule) kao deskriptora pri proučavanju kvantitativnih odnosa strukture i aktivnosti (QSAR). Izvedene su QSAR-jednadžbe kojima se, s pomoću deskriptora F, molekulske površine i energije Coulombskog međudjelovanja molekule i enzima, može procijeniti katalitička efikasnost raznih supstrata točkastih mutanata subtilizina i predvidjeti inhibitorska svojstva s-triazina prema dihidrofolat-reduktazi iz pileće jetre.