Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 1275–1283
Bioorganic & Medicinal Chemistry

QSAR studies on some thiophene analogs as anti-inflammatory agents: enhancement of activity by electronic parameters and its utilization for chemical lead optimization

Ajay D. Pillai,[a] Shubha Rani,[b] Parendu D. Rathod,[c] Franklin P. Xavier,[c] Kamala K. Vasu,[c,*] Harish Padh[d] and Vasudevan Sudarsanam[c]

[a]*Department of Pharmaceutical Chemistry, M.S Ramaiah College of Pharmacy, M.S Ramaiah Nagar, Mathikere, Bangalore 560 054, Karnataka, India*
[b]*Department of Biostatistics, B.V. Patel Pharmaceutical Education and Research Development Centre, Thaltej-Gandhinagar Highway, Thaltej, Ahmedabad 380 054, Gujarat, India*
[c]*Department of Medicinal Chemistry, B.V. Patel Pharmaceutical Education and Research Development Centre, Thaltej-Gandhinagar Highway, Thaltej, Ahmedabad 380 054, Gujarat, India*
[d]*Department of Biotechnology, B.V. Patel Pharmaceutical Education and Research Development Centre, Thaltej-Gandhinagar Highway, Thaltej, Ahmedabad 380 054, Gujarat, India*

Received 8 August 2004; revised 6 November 2004; accepted 8 November 2004
Available online 8 December 2004

Abstract—Small molecule heterocycle is an integral part of new drug discovery in anti-inflammatory research. In our previous papers we reported the synthesis of thiophene analogs substituted at the fifth position with α-oximino propionic ester moiety and the fact that such new chemical entities exhibit anti-inflammatory activity in male/female Sprague–Dawley rats. In this paper we report the quantitative structure activity relationship (QSAR) studies of a series of 43 thiophene analogs. The analogs when subjected to cluster analysis technique led to the formation of four homogeneous groups. The cluster analysis technique grouped the 2-anilino-5-substituted-4-methyl-thiophene-3-carboxylic acid methyl esters as one homogeneous group. The clusters were individually taken up for a Hansch type of QSAR study with 10 molecular descriptors. The QSAR equations generated were cross validated by the leave out one method. The studies gave an insight into the dominant role played by electronic properties like energy of the lowest unoccupied molecular orbital (E_{LUMO}) and dipole moment (dipole) in modulating the anti-inflammatory activity. From the QSAR studies a three point pharmacophore has been established for designing novel anti-inflammatory molecules.

1. Introduction

The drugs used to treat the symptoms of inflammation are called anti-inflammatory agents. The most widely accepted mode of action of nonsteroidal anti-inflammatory drugs (NSAID's) is the inhibition of the cyclooxygenase (COX-1) enzyme involved in arachidonic acid metabolism.[1] However with the identification of another isoform of constitutive COX-1, namely COX-2 (inducible) and inhibition of COX-2 proving beneficial in clinical situations had led to the introduction of many potent anti-inflammatory agents called COX-2 inhibitors.[2–4] However in view of the polygenic and multifactorial nature of inflammatory diseases with a number of mediators exacerbating the disease process, many alternative pathways and other rate limiting steps are recently being targeted in the design of novel anti-inflammatory drugs. Noteworthy of these is the emergence of many candidates incorporating the oximino feature in their structure, particularly candidates having anti-inflammatory activity centered on Phospholipase A_2.[5–10] A more recent approach in anti-inflammatory research has already been the development of nitric oxide releasing NSAID's.[11–14]

Keywords: Pharmacophore; QSAR; Electronic parameter.
* Corresponding author. Tel.: +91 079 2743 9375; fax: +91 079 2745 0449; e-mail: perd@perdcentre.com

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.11.016

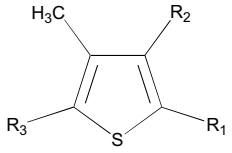
In our previous paper[15] we reported the design, synthesis, and pharmacological evaluation of novel tetra substituted thiophenes as anti-inflammatory agents. In continuation with the work, and to establish a pharmacophore for designing better anti-inflammatory agents, we thought it is essential to carry out a QSAR study, for these thiophene analogs[15,16] to perceive the importance of the molecular properties, which are critical in accentuating the biological activity. The present work deals with the QSAR studies of some thiophene analogs as anti-inflammatory agents. The results culminated in the generation of a novel pharmacophore for designing better anti-inflammatory agents.

2. Results and discussion

2.1. Development of QSAR equations

A QSAR study was carried out for the thiophene analogs ($n = 43$) Table 1, making use of a variety of interme-

Table 1. Biological activity elicited by the tetra substituted thiophenes and their corresponding cluster membership generated by K-mean cluster analysis



C	R_3	R_2	R_1	$P(B)$	Cluster membership
1	–CO–CH_2–$COOCH_2CH_3$	$COOCH_3$	–NH–CO–Ph	0.79 ± 0.12 (53.4)	4
2	*Z*–CO–C(=N–OH)–COO–CH_2–CH_3	$COOCH_3$	–NH–CO–Ph	0.74 ± 0.11 (50.4)	3
3	*Z*–CO–C(=N–OCH_3)–COO–CH_2–CH_3	$COOCH_3$	–NH–CO–Ph	0.72 ± 0.09 (49)	3
4	–CO–Ph–*p*-Cl	$COOCH_3$	–NH–CO–Ph	0.89 ± 0.13 (60)	4
5	–CO–Ph–*p*-SO_2CH_3	$COOCH_3$	–NH–CO–Ph	0.40 ± 0.05 (27)	3
6	3,5-Dimethyl-4-methoxy-2-pyridyl–	$COOCH_3$	–NH–CO–Ph	0.31 ± 0.07 (21)	4
7	3,5-Dimethyl-4-methoxy-2-pyridyl–	$COOCH_3$	–NH–CH_3	0.47 ± 0.14 (32)	4
8	–CO–Ph–*p*-SO_2CH_3	$COOCH_3$	–NH–CH_3	0.89 ± 0.16 (60)	3
9	–CO–Ph–*p*-SCH_3	$COOCH_3$	–NH–CH_3	0.89 ± 0.11 (60)	4
10	*Z*–CO–C(=N–OH)–COO–CH_2–CH_3	$COOCH_3$	–NH–CH_3	0.63 ± 0.09 (43)	3
11	*Z*–CO–C(=N–OH)–COO–CH_2–CH_3	$COOCH_3$	–NH–C_2H_5	0.45 ± 0.12 (31)	3
12	CO–CH_2–$COOCH_2CH_3$	$COOCH_3$	–NH–C_2H_5	0.94 ± 0.14 (64)	3
13	–CO–Ph–*p*-SO_2CH_3	$COOCH_3$	–NH–C_2H_5	0.56 ± 0.08 (38)	3
14	–CO–Ph–*p*-SCH_3	$COOCH_3$	–NH–C_2H_5	0.28 ± 0.04 (19)	4
15	3,5-Dimethyl-4-methoxy-2-pyridyl–	$COOCH_3$	–NH–C_2H_5	0.41 ± 0.07 (28)	4
16	–CO–Ph–*p*-SCH_3	$COOCH_3$	–NH–COOEt	0.32 ± 0.09 (22)	3
17	–CO–CH_2–CO–NH–Ph	$COOCH_3$	–NH–COOEt	0.07 ± 0.01 (5)	3
18	CO–C(=N–OCH_3)–COO–CH_2–CH_3	$COOCH_3$	–NH–COOEt	0.32 ± 0.11 (22)	3
19	–CO–Ph–*p*-SO_2CH_3	$COCH_3$	–NH–COOEt	0.80 ± 0.12 (54)	3
20	–CO–Ph–*p*-SO_2CH_3	CO–NH–Ph	–NH–COOEt	0.72 ± 0.11 (48.6)	3
21	–CO–Ph–*p*-Cl	CO–NH–Ph	–NH–C_2H_5	0.37 ± 0.02 (25)	4
22	–CO–Ph–*p*-SCH_3	CO–NH–Ph	–NH–Ph–Cl-*p*	0.62 ± 0.10 (42)	1
23	–CO–Ph–*p*-Cl	CO–NH–Ph	–NH–Ph–Cl-*p*	0.44 ± 0.06 (30)	1
24	–CO–Ph–*p*-NH–CO–CH_3	$COOCH_3$	–NH–Ph–Cl-*p*	0.56 ± 0.09 (38)	4
25	–CO–Ph–*p*-SO_2CH_3	$COOCH_3$	–NH–Ph–Cl-*p*	0.89 ± 0.12 (60)	4
26	*Z*–CO–C(=N–OH)–COO–CH_2–CH_3	$COOCH_3$	–NH–Ph–Cl-*p*	1.11 ± 0.17 (75)	3
27	*Z*–CO–C(=N–OCH_3)–COO–CH_2–CH_3	$COOCH_3$	–NH–Ph–Cl-*p*	0.29 ± 0.08 (20)	3
28	C(=N–OH)–$COOCH_2CH_3$	$COOCH_3$	–NH–Ph–Cl-*p*	0.41 ± 0.05 (28)	4
29	–C=(N–OH)–CO–CH_3	$COOCH_3$	–NH–Ph–Cl-*p*	0.34 ± 0.07 (23)	4
30	–CO–CH_2–CO–NH–Ph	$COOCH_3$	–NH–Ph–Cl-*p*	0.78 ± 0.10 (53)	4
31	CO–CH_2–$COOCH_2CH_3$	$COOCH_3$	–NH–Ph–Cl-*p*	0.51 ± 0.09 (34.7)	3
32	–CO–$COOCH_2CH_3$	$COOCH_3$	–NH–Ph	0.03 ± 0.001 (2)	2
33	C(=N–OH)–$COOCH_2CH_3$	$COOCH_3$	–NH–Ph	0.44 ± 0.09 (30)	2
34	*Z*–CO–C(=N–OH)–COO–CH_2–CH_3	$COOCH_3$	–NH–Ph	0.69 ± 0.10 (46.6)	2
35	*Z*–CO–C(=N–OCH_3)–COO–CH_2–CH_3	$COOCH_3$	–NH–Ph	0.89 ± 0.09 (60)	2
36	CO–CH_2–$COOCH_2CH_3$	$COOCH_3$	–NH–Ph	0.06 ± 0.01 (4.16)	2
37	–CO–CH_2Cl	$COOCH_3$	–NH–Ph	0.41 ± 0.1 (27.7)	2
38	–CO–Ph–SO_2CH_3-*p*	$COOCH_3$	–NH–Ph	0.44 ± 0.09 (30)	2
39	–CO–Ph–SCH_3-*p*	$COOCH_3$	–NH–Ph	0.33 ± 0.05 (22.7)	2
40	–CO–Ph–Cl-*p*	$COOCH_3$	–NH–Ph	0.06 ± 0.03 (4.16)	2
41	–C(NOH)–Ph–SO_2CH_3-*p*	$COOCH_3$	–NH–Ph	0.04 ± 0.001 (3)	2
42	–C=(N–OH)–CO–CH_3	$COOCH_3$	–NH–Ph	0.22 ± 0.04 (15)	2
43	–COO–$CH(CH_3)_2$	$COOCH_3$	–NH–Ph	0.04 ± 0.001 (3)	2

C denotes the experimental candidate. Forty three novel designed tetra substituted thiophenes were taken up for QSAR study. R_1, R_2, and R_3 are the different substitutions in the thiophene nucleus. P = Paw volume and B = % protection given to the rat paw, in carrageenin induced rat paw edema model. Control reading is 2.25 ± 0.07. The last column denotes the cluster membership of each candidate, generated by K-mean cluster analysis.

Table 2. Haloketones (H) used for the synthesis of novel designed thiophenes

H	IUPAC nomenclature	Structure
HI	(*Z*)-γ-Chloro-β-oxo-α-hydroxyimino butyric acid, ethyl ester	Cl–CH_2–CO–C(=N–OH)–COO–CH_2–CH_3
HII	(*Z*)-γ-Bromo-β-oxo-α-methoxyimino butyric acid, ethyl ester	Br–CH_2–CO–C(=N–OCH_3)–COO–CH_2–CH_3
HIII	γ-Bromo-β-oxo-butyric acid, ethyl ester	Br–CH_2–CO–CH_2–$COOCH_2CH_3$
HIV	β-Bromo-α-hydroxyimino propionicacid, ethyl ester	Br–CH_2–C(=N–OH)–$COOCH_2CH_3$
HV	β-Bromo-pyruvic acid, ethyl ester	Br–CH_2–CO–$COOCH_2CH_3$
HVI	4-(α-Bromo acetyl)1-phenyl methyl sulfone	Br–CH_2–CO–Ph–*p*-SO_2CH_3
HVII	4-(α-Bromo acetyl)-thioanisole	Br–CH_2–CO–Ph–*p*-SCH_3
HVIII	4-(α-Bromo acetyl)-1-chloro benzene	Br–CH_2–CO–Ph–*p*-Cl
HIX	4-(α-Bromo acetyl)-1-acetyl amino benzene	Br–CH_2–CO–Ph–*p*-NH–CO–CH_3
HX	(3,5-Dimethyl-4-methoxy)-2-chloromethyl pyridine	See Scheme 1
HXI	Isopropyl chloro acetate	Cl–CH_2–COO–$CH(CH_3)_2$
HXII	Dichloro acetone	Cl–CH_2–CO–CH_2Cl
HXIII	β-Bromo diacetyl monoxime	Br–CH_2–C=(N–OH)–CO–CH_3
HXIV	γ-Bromo-acetoacetanilide	Br–CH_2–CO–CH_2–CO–NH–Ph

HX is shown in Scheme 1.

Table 3. Type of molecular descriptors selected for the study

Molecular descriptors	Type
Heat of formation (HOF)	Thermodynamic
Electronic energy (EE)	Electronic
Dipole moment (dipole)	Electronic
Ionization potential (IP)	Electronic
Energy of the highest occupied molecular orbital (E_{HOMO})	Electronic
Energy of the lowest unoccupied molecular orbital (E_{LUMO})	Electronic
Logarithm of the partition coefficient (log *P*)	Lipophilic
Molar refractivity (MR)	Thermodynamic
Molar volume (MV)	Steric
Polarizability (POLAR)	Electronic

The descriptors fell into four categories namely electronic, steric, lipophilic, and thermodynamic.

diates (Table 2), to produce structural diversity. The biological activity was first regressed with the molecular descriptors selected for the study (Tables 3 and 4). Since many descriptors selected for the study were nonorthogonal, multiple linear regression (MLR) of these descriptors with the biological activity gave problem of multicollinearity. A principal component analysis (PCA) was carried out for finding the prominent components, which explained the variance of the data set. Thus ionization potential, molar refractivity, and polarizability were found as the highest weighted descriptors from the principle scores of the normalized Eigen vectors of the first three major principal components (Table 5) and as those pertinent for describing biological activity. A multiple linear regression analysis (MLR) of biological activity with ionization potential, molar refractivity, and polarizability as descriptors or with principle components Z1, Z2, and Z3 itself serving as descriptors gave the problem of multicollinearity again.

The heterogeneity of the candidates selected for the study was a hindrance for a successful generation of a regression equation for the training set $n = 43$. This was an impetus for searching and exploiting data analysis techniques like cluster analysis, which helps to create homogeneous groups of compounds from a heterogeneous data set. A cluster analysis was performed using K-mean cluster technique to have homogeneous groups among the training set. Finally after many permutations with nine descriptors, excluding the electronic energy, we could get four clusters comprising 2, 12, 16, and 13 candidates, respectively (Table 1). Two compounds that fell into the first cluster were compound number **22** and **23**. The 2-anilino-5-substituted-4-methyl-thiophene-3-carboxylic acid methyl esters were clustered separately into the second cluster. All the clusters generated were further taken up individually for QSAR study and the equations generated were validated using leave one out method.

2.2. Development of QSAR equation for Cluster II

To sort out the orthogonal descriptors prior to performing a MLR, descriptors of compounds belonging to Cluster II were subjected to principal component analysis. Heat of formation (HOF), energy of the highest occupied molecular orbital (E_{HOMO}) and molar refractivity (MR) were selected as main descriptors based on the principle scores of the normalized Eigen vectors of the first three principal components. A Hansch type QSAR was performed for this cluster using the above three descriptors utilizing the multiple linear regression analysis as the statistical technique, which resulted in Eq. 1 as follows,

Table 4. Values of the descriptors for individual compounds in the training set

C	HOF (kcal)	EE (eV)	Dipole (Debye)	IP (eV)	E_{HOMO} (eV)	E_{LUMO} (eV)	Log *P*	MR (cm^3)	POLAR (cm^3)	MV (cm^3)	*B*	Log *B*
1	−91.5	−35,516.2	5.12	9.03	−9.03	−0.878	2.95	100.23	40.24	296.4	53.4	1.72
2	−174.09	−40,074.9	6.2	9.03	−9.03	−0.924	3.03	104.68	41.78	301.9	50.3	1.70
3	−167.5	−42,695.8	7.05	9.05	−9.05	−0.983	3.29	109.45	43.75	326.3	49	1.69
4	−73.8	−36,534.02	2.35	9.09	−9.09	−0.981	5.08	109.35	43.8	302	60	1.77
5	−127.17	−43,479.9	5.7	9.2	−9.2	−1.279	3.23	118.07	46.23	331.8	27	1.43
6	−76.6	−40,074.7	3.74	8.7	−8.7	−0.714	4.93	113.12	45.51	327	21	1.32
7	−75.3	−27,460.5	1.61	8.4	−8.4	−0.337	3.62	90.06	35.62	267.5	32.0	1.50
8	−126.2	−30,723.5	7.66	8.86	−8.86	−1.122	1.92	95.01	36.39	272.4	60	1.77
9	−67.04	−26,107.4	4.48	8.31	−8.31	−0.581	3.66	94.39	36.07	258.5	60	1.77
10	−170.86	−28,556.7	4.6	8.7	−8.7	−0.683	1.72	81.62	31.42	235.9	43	1.63
11	−177.4	−30,567.4	4.51	8.76	−8.76	−0.682	2.06	86.37	33.25	252	30.9	1.49
12	−196.2	−26,111.2	7.22	8.78	−8.78	−0.642	1.98	81.92	32.18	253.5	64	1.80
13	−133.1	−32,792.5	7.57	8.91	−8.91	−1.138	2.26	99.76	38.22	288.9	38	1.57
14	−73.5	−28,059.3	4.55	8.3	−8.3	−0.56	4	99.14	37.9	274.8	19	1.27
15	−82.2	−29,521.9	1.55	8.43	−8.43	−0.343	3.96	94.81	37.45	284	28	1.44
16	−148.16	−34,205.8	3.02	8.4	−8.4	−0.836	3.99	103.6	40.48	293.9	22	1.34
17	−187.9	−38,256.13	1.88	8.81	−8.81	−1.191	2.63	103.36	41.92	298	4.99	0.69
18	−243.75	−39,246.9	3.59	9.17	−9.17	−0.951	1.98	86.38	37.93	296.2	21.9	1.34
19	−157.9	−36,639	3.35	9.35	−9.35	−1.335	1.75	103.1	40.11	302.5	54	1.73
20	−126.04	−47,861.2	6.17	8.98	−8.98	−1.418	3.24	125.95	49.52	349.9	48.6	1.68
21	−72.8	−40,520.8	4.7	8.88	−8.88	−1.108	5.1	117.23	47.32	320	25	1.39
22	40.05	−43,832.6	1.04	8.3	−8.3	−0.846	7.17	139.21	54.66	356.1	42	1.62
23	−13.67	−36,464.6	2.79	8.5	−8.5	−0.787	6.69	112.92	44.76	310.7	30.0	1.47
24	−77.2	−39,737.3	4.27	8.7	−8.7	−0.951	4.65	117.66	47.4	322.2	38	1.57
25	−100.5	−40,716.5	5.9	8.8	−8.8	−1.23	4.45	118.11	46.26	330.9	60	1.77
26	−145.16	−38,390.6	4.2	8.6	−8.6	−1.015	4.24	104.72	41.44	297.9	75.2	1.87
27	−137.14	−40,824.6	4.6	8.64	−8.64	−0.975	4.51	109.49	43.41	322.3	20.0	1.30
28	−119.8	−34,764.3	3.86	8.7	−8.7	−1.014	4.62	99.25	39.28	284.6	28	1.44
29	−68.62	−30,143.9	2.21	8.48	−8.48	−0.809	3.86	93.4	36.9	262.1	23	1.36
30	−81.6	−39,775.4	3.5	8.75	−8.75	−1.163	4.82	117.25	47.32	320	53	1.72
31	−163.8	−33,823.2	5	8.79	−8.79	−1.132	4.17	100.27	40.12	295	34.7	1.54
32	−144.62	−30,158.5	4.1	8.7	−8.7	−1.007	3.67	90.92	267.1	36.34	2.00	0.30
33	−112.83	−32,583.3	3.05	8.45	−8.45	−0.631	4.06	94.45	275.3	37.45	30.0	1.47
34	−138.54	−36,307.4	5.27	8.59	−8.59	−0.873	3.72	99.91	288.6	39.62	46.6	1.66
35	−130.5	−38,693.6	5.53	8.56	−8.56	−0.834	3.95	104.68	313	41.59	60	1.77
36	−156.11	−31,843.3	3.17	8.57	−8.57	−0.791	3.61	95.46	283.6	38.18	4.16	0.61
37	−74.28	−24,539.3	3.85	8.61	−8.61	−0.774	3.81	84.36	242	33.88	27.7	1.44
38	−94	−38,637.1	7.38	8.69	−8.69	−1.174	3.89	113.31	319	44.35	30.0	1.47
39	−34.7	−33,751.6	5.3	8.34	−8.34	−0.745	5.63	112.68	301.3	44.03	22.7	1.35
40	−40.45	−31,946.5	4.57	8.56	−8.56	−0.839	5.74	104.58	289.1	41.74	4.60	0.66
41	−62.04	−41,570.4	5.36	8.72	−8.72	−1.005	4.28	116.84	327.5	46.73	3.00	0.47
42	−61.79	−28,181.3	4.13	8.38	−8.38	−0.63	3.3	88.6	252.8	35.08	15	1.17
43	−124.2	−28,569.9	3.17	8.6	−8.6	−0.808	4.45	88.89	270.4	36.31	3.00	0.47

Log *B* denotes the logarithm of the biological activity (% protection from inflammation of the rat paw in carrageenin induced rat paw edema model), C represents the compound selected for QSAR study. *B* represents the biological activity or % protection to inflammation given to the rat paw, in carrageenin induced rat paw edema model.

Table 5. Eigenvalues and eigenvectors of correlation matrix, $n = 43$ generated from principal component analysis, where *n* represents the number of compounds in the training set

	Eigen value	Proportion	Cum. Prop.	Df
Vector1	3.980	0.39	0.39	54
Vector2	2.74	0.27	0.67	44
Vector3	1.77	0.17	0.85	35
Vector4	0.65	0.06	0.91	27
Vector5	0.40	0.040	0.95	20
Vector6	0.25	0.02	0.98	14
Vector7	0.15	0.01	0.99	9
Vector8	0.02	0.002	0.99	5
Vector9	0.001	0.0001	1	2
Vector10	1.45E−017	1.45E−016	1	0

Df denotes degree of freedom.

$$\log B = 18.47 - 0.002(\mathrm{HOF}) + 2.20(E_{\mathrm{HOMO}}) + 0.012(\mathrm{MR})$$
$$n = 12 \quad r = 0.44 \quad \mathrm{SE} = 0.55$$
$$F_{\mathrm{Calculated}} = 0.72 \quad P > 0.05 \qquad (1)$$

where n represents the number of data points, r is the multiple correlation coefficient, SE is the standard error of the estimate, and 'F' is the F-statistic ratio. The regression was not significant and the residuals were very high.

As the correlation matrix from PCA showed HOF, dipole, E_{LUMO}, MR, and $\log P$ as orthogonal descriptors, a MLR was performed using these five descriptors, which gave Eq. 2.

$$\log B = 2.46 - 0.005(\mathrm{HOF}) + 0.66(\mathrm{dipole}) + 3.92(E_{\mathrm{LUMO}}) + 0.03(\log P) - 0.01(\mathrm{MR})$$
$$n = 12 \quad r = 0.88 \quad \mathrm{SE} = 0.32$$
$$F_{\mathrm{Calculated}} = 4.59 \quad P \leqslant 0.05 \qquad (2)$$

Regression was significant, however it gave high residuals for compounds **33**, **37**, and **42**. As a result, a MLR was done excluding these three compounds from the data set using all the five descriptors that gave Eq. 3.

$$\log B = 2.82 - 0.004(\mathrm{HOF}) + 0.72(\mathrm{dipole}) + 4.19(E_{\mathrm{LUMO}}) - 0.06(\log P) - 0.01(\mathrm{MR})$$
$$n = 9 \quad r = 0.99 \quad \mathrm{SE} = 0.08$$
$$F_{\mathrm{Calculated}} = 86.7 \quad P \leqslant 0.05 \qquad (3)$$

The equation was significant with low residuals. The correlation matrix showed that $\log P$ is highly inter correlated with HOF. Therefore a MLR was performed by removing $\log P$, which was correlated with HOF and Eq. 4 was obtained as follows;

$$\log B = 2.0 - 0.005(\mathrm{HOF}) + 0.71(\mathrm{dipole}) + 4.06(E_{\mathrm{LUMO}}) - 0.013(\mathrm{MR})$$
$$n = 9 \quad r = 0.99 \quad \mathrm{SE} = 0.07$$
$$F_{\mathrm{Calculated}} = 142.9 \quad P \leqslant 0.05 \qquad (4)$$

Eq. 4 suggests a high value of E_{LUMO} for the compounds to improve its activity. The significance of E_{LUMO} indicates, high electrophilicity of the compounds, and there by accepting electrons to its lowest unoccupied molecular orbital, would help them to improve the biological activity. The importance of dipole in modulating the biological activity, may be due to more number of carbonyl groups ($^{+}$C–O^{-}), where a permanent polarization is seen due to the electro negativity difference between the atoms. This permanent polarization may result in a dipole–dipole type of interaction with the target. The carbonyl carbon at the third and fifth position, serve as the positive centers in this series. The compounds might be involved in making fruitful binding interactions with the amino acids present at the catalytic site of the enzyme, through a hydrogen type of bonding. The molecular properties like E_{LUMO} and dipole play a critical role in modulating the activity profile for these classes of compounds.

Table 6. Validation of QSAR equations generated by leave one out method

Compound	A	B
Validation of Eq. 4 by leave one out method		
33	1.47	1.37
34	1.66	1.56
35	1.77	1.88
36	0.61	0.56
39	1.35	1.45
Validation of Eq. 6 by leave one out method		
2	1.7	1.6
3	1.69	1.7
5	1.43	1.49
8	1.77	1.97
11	1.49	1.45
Validation of Eq. 10 by leave one out method		
6	1.32	1.38
7	1.5	1.48
14	1.27	1.18
15	1.44	1.45
21	1.39	1.37

A and B represents the observed $\log B$ and predicted $\log B$.

The validation of Eq. 4 was carried out using leave one out method. Thus each candidate was left out one by one and the equations there in generated could predict the biological activity close to that of the experimentally observed value thus validating Eq. 4 (Table 6).

2.3. Development of QSAR equations for Cluster III

The third cluster was subjected to a PCA, wherein dipole, ionization potential, and polarizability were selected based on the same principles described above. MLR of these descriptors with biological activity gave Eq. 5. Regression was not significant. The residuals were high for compounds **13**, **19**, and **26**.

$$\log B = 0.69 + 0.10(\mathrm{dipole}) + 0.077(\mathrm{IP}) - 0.009(\mathrm{polar})$$
$$n = 16 \quad \mathrm{SE} = -0.24 \quad r = 0.65$$
$$F_{\mathrm{Calculated}} = 2.97 \quad P > 0.05 \qquad (5)$$

A MLR excluding these three candidates gave Eq. 6. The equation was found to be highly significant with low standard error of mean.

$$\log B = 2.16 + 0.15(\mathrm{DIPOLE}) - 0.12(\mathrm{IP}) - 0.009(\mathrm{polar})$$
$$n = 13 \quad \mathrm{SE} = 0.14 \quad r = 0.90$$
$$F_{\mathrm{Calculated}} = 13.2 \quad P \leqslant 0.05 \qquad (6)$$

An increment in the dipole value is again recommended for better activity of the candidates by Eq. 6. The carbonyl carbon at the second position of the 2-carbethoxy

amino thiophenes and 2-benzoyl amino thiophenes, imino carbon at the fifth position as in the oximino analogs and carbonyl carbon at the third position serve as the positive centers in the designed series. It is obvious that compounds **2**, **3**, **8**, **13**, **20**, and **38**, which had either an oximino function or a methyl sulfonyl function at the fifth position of the molecule had useful activity profile. The incorporation of both functions simultaneously at the fifth position as in compound **41** showed decrease in the activity profile. This supports our earlier finding[15] that aromatic oximes are less preferred over aliphatic oximes at the fifth position of the thiophenes as anti-inflammatory agents. A plausible explanation for the lesser activity exhibited by the 2-carbethoxy amino series might be that, the carbethoxy group gets metabolized to the corresponding acid followed by a decarboxylation to amine, which also suggests that a primary amine is not preferred at the second position of the thiophenes. This view is supported by the fact that 2-benzoyl amino thiophenes were superior in eliciting the anti-inflammatory activity compared to the 2-carbethoxy thiophenes. The studies probably reflect the importance of a secondary amino group at the second position of the thiophene in modulating the anti-inflammatory activity profile of the candidates.

The validation of Eq. 6 was carried out using leave one out method. Thus each candidate was left out one by one and the equations there in generated could predict the biological activity close to that of the experimentally observed value thus validating Eq. 6 (Table 6).

2.4. Development of QSAR equations for Cluster IV

A MLR was performed with HOF, dipole, E_{LUMO}, log P, MV, and IP selected from PCA results for the fourth cluster, on the same lines as described earlier. Since MLR revealed multicollinearity and the nonorthogonality of log P with rest of the descriptors, a MLR was performed excluding log P, which resulted in Eq. 7.

$$\begin{aligned}\log B = &-1.005 + 0.0009(\mathrm{HOF}) + 0.0262(\mathrm{dipole})\\ &- 0.100(E_{\mathrm{LUMO}}) - 0.002(\mathrm{MV})\\ &+ 0.36(\mathrm{IP})\\ n = 13 \quad &\mathrm{SE} = -0.21 \quad r = 0.50\\ F_{\mathrm{Calculated}} = 0.47 \quad &P > 0.05 \end{aligned} \tag{7}$$

As the residual values for compounds **9** and **30** were high, they were removed from the training set and QSAR was performed ($n = 11$), which resulted in Eq. 8. The r value improved to 0.72.

$$\begin{aligned}\log B = &-3.3 - 0.002(\mathrm{HOF}) - 0.005(\mathrm{dipole})\\ &+ 0.083(E_{\mathrm{LUMO}}) - 0.0001(\mathrm{MV})\\ &+ 0.54(\mathrm{IP})\\ n = 11 \quad &\mathrm{SE} = -0.17 \quad r = 0.72\\ F_{\mathrm{Calculated}} = 1.099 \quad &P > 0.05 \end{aligned} \tag{8}$$

The residuals also improved, except for compound **25**, which showed a high residual. A MLR was again carried out after deleting candidate number **25**, which gave Eq. 9. The residual values improved. The r value significantly improved from 0.72 to 0.85.

$$\begin{aligned}\log B = &-4.4 - 0.002(\mathrm{HOF}) - 0.012(\mathrm{DIPOLE})\\ &+ 0.26(E_{\mathrm{LUMO}}) - 0.002(\mathrm{MV})\\ &+ 0.77(\mathrm{IP})\\ n = 10 \quad &\mathrm{SE} = -0.12 \quad r = 0.85\\ F_{\mathrm{Calculated}} = 2.269 \quad &P > 0.05 \end{aligned} \tag{9}$$

However the regression analysis gave a high residual for candidate 24. Hence MLR was carried out for $n = 9$ excluding compound **24**, which resulted in Eq. 10.

$$\begin{aligned}\log B = &-5.5 - 0.0004(\mathrm{HOF}) - 0.0055(\mathrm{DIPOLE})\\ &+ 0.41(E_{\mathrm{LUMO}}) - 0.0047(\mathrm{MV})\\ &+ 1.0(\mathrm{IP})\\ n = 9 \quad &\mathrm{SE} = -0.037 \quad r = 0.99\\ F_{\mathrm{Calculated}} = 32.7 \quad &P \leqslant 0.05 \end{aligned} \tag{10}$$

The validation of Eq. 10 was carried out using leave one out method. Thus each candidate was left out one by one and the equations generated could predict the biological activity close to that observed validating the generated Eq. 10 (Table 6).

Later on when we removed HOF and dipole, the regression Eq. 11 generated was similar to Eq. 10, which proved these molecular descriptors redundant in modulating the activity for compounds of this cluster.

$$\begin{aligned}\log B = &-5.5 + 0.42(E_{\mathrm{LUMO}}) - 0.0049(\mathrm{MV})\\ &+ 1.0(\mathrm{IP})\\ n = 9 \quad &\mathrm{SE} = -0.037 \quad r = 0.99\\ F_{\mathrm{Calculated}} = 80.7 \quad &P \leqslant 0.05 \end{aligned} \tag{11}$$

2.5. Outcome of the QSAR studies

Taking into consideration of all the QSAR results, the importance given to the molecular descriptors, E_{LUMO} and dipole in modulating the biological activity profile, is well reflected. The biological activity chart clearly demarcates the oximino and methyl sulfonyl analogs in exhibiting better anti-inflammatory activity. More over the lesser activity profile exhibited by the 2-carbethoxy amino analogs give a remote indication of the probable involvement of a secondary amino group at the second position of the thiophene in modulating the anti-inflammatory activity profile of the candidates. Based on these findings, we propose a three point pharmacophore as shown in Figure 1, in designing better anti-inflammatory agents.

3. Conclusion

A series of tetra substituted thiophene analogs previously designed, synthesized, and characterized were subjected to a QSAR study to delineate the importance of the molecular descriptors pertinent in the modulation of the anti-inflammatory activity elicited by the candi-

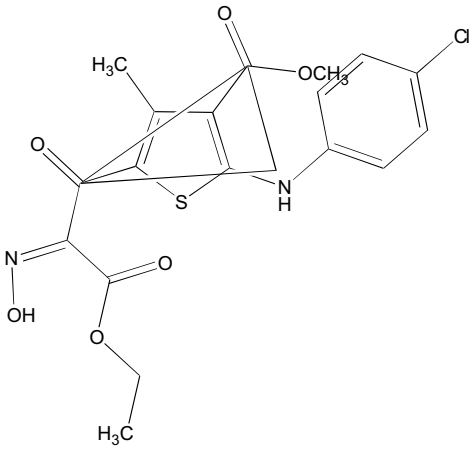


Figure 1. The proposed three point pharmacophore generated from the QSAR study.

dates in the in vivo mode. Owing to the orthogonality of the descriptors selected, K-mean cluster, a cluster analysis technique was utilized for the formation of homogeneous clusters before taking them further for a QSAR. The QSAR model used was Hansch type model. The statistical technique used was multiple linear regression analysis. From the QSAR studies, the molecular descriptors like dipole and E_{LUMO} proved important in defining the activity of the candidates. A logical reason for the dominance of these two parameters might be the more number of carbonyl groups in the designed series. The oximino analogs and the methyl sulfonyl analogs were dominating the biological activity chart and a close look at them reveal the high E_{LUMO} and dipole moments for these candidates. Based on these studies we disclose a novel three point pharmacophore for designing better anti-inflammatory agents. The studies on whether these oxime analogs at physiological pH release nitric oxide, which is gastro protective, is in progress. This is relevant in view of the fact that some 2-oximino amides have been reported as NO donors.[25] If this is true, then it is really worth looking at these oxime analogs as potent leads for the treatment of chronic inflammatory conditions like rheumatoid arthritis with cytoprotective ability.

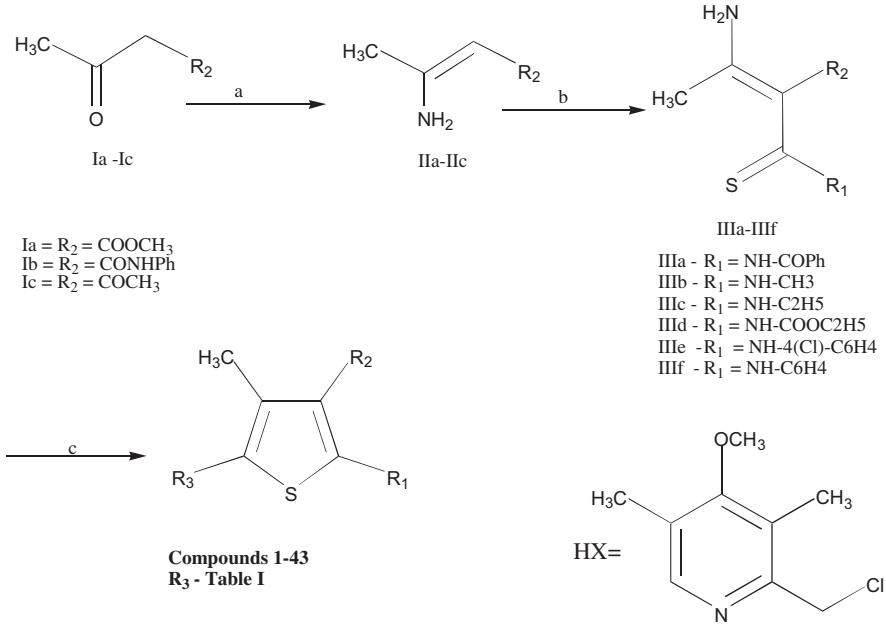
4. Experimental

4.1. Material and methods

Isothiocyanates were synthesized using the modified Kaluza method.[16] α-Haloketones (HI–HXIV) were synthesized using reported procedure.[16] Carrageenin was purchased from Qualigen Mumbai.

4.2. Synthesis of tetra substituted thiophenes

The sequence utilized to synthesize the thiophene analogs (Scheme 1) was similar to the method reported by Rajappa et al.[16] The reaction was brought about by reacting equimolar amount of an enamine of the formula **IIa**–**IIc** with alkyl, aryl, aroyl, or carbethoxy isothiocyanates, to give the corresponding adduct of the formula **IIIa**–**IIIf**, followed by reaction with α-halo carbonyl or halomethylene compounds of the formula (HI–HXIV, Table 2) to yield tetra substituted thiophenes **1**–**43** (Scheme 1). All the compounds were characterized by IR, Mass, and ^{1}H NMR data.



Scheme 1. Scheme for the synthesis of tetra substituted thiophenes taken for QSAR study. (3,5-Dimethyl-4-methoxy)-2-chloromethyl pyridine was the only halomethylene compound used. (a) Ammonia, (b) alkyl, aryl, aroyl, and carbethoxy isothiocyanates, diethyl ether, (c) haloketones and halomethylene compounds (HI–HXIV), acetonitrile.

4.3. Pharmacology

Anti-inflammatory activity[15] was done using carrageenin induced rat paw edema method in rats.

Sprague–Dawley (male/female) rats weighing 150–250 g were used for the edema test. Animals were divided into 44 groups comprising six rats per group. One group was kept as the control and remaining 43 groups (test group) were used to determine the anti-inflammatory activity elicited by the 43 candidates, respectively. Rats were put on fast for 18 h prior to the experiment. The standard drug, Ibuprofen (100 mg/kg body weight) and the test drugs (100 mg/kg body weight) were given orally as a suspension, in 0.1% sodium CMC as vehicle. One hour later, 0.1 mL of 1% carrageenin solution in saline was injected in the sub-plantar region of the right hind paw of each rat. After 3 h of the carrageenin injection, the reduction in the paw volume compared to vehicle control was measured using plethysmometer. The institutional ethics committee, constituted by the Ministry of Social Justice and Empowerment, Government of India, approved the experimental protocol.

4.4. QSAR methods

4.4.1. Data set and parameters. Training set is the set of molecules whose biological activity is regressed with its molecular descriptor values. Our training set consisted of 43 novel thiophene analogs, synthesized and biologically evaluated, for anti-inflammatory activity in carrageenin induced rat paw edema in vivo model.[15] The biological activity was expressed as $\log B$ where, $B = \%$ protection from inflammation given to the rat paw. Descriptor is any molecular property which is characteristic of a molecule and can be utilized to determine new QSAR relationship. The number of descriptors selected for the study was 10, which fell into four categories as electronic, steric, lipophilic, and thermodynamic (Table 3).

4.4.2. Cluster analysis. Before any data is subjected to analysis, it is of utmost importance to explore the data. Mathematically and graphically, data can be examined by many existing techniques. One among them is called the cluster analysis.[17] Cluster analysis is the generic term applied for a wide variety of procedures that can be used to create a classification of a heterogeneous data into relatively homogeneous groups. The aim may be considered to be that of examining homogeneity of the data or detecting some unusual data points or identifying patterns or indicating potentially interesting relationships in the data. Our data was analyzed using K-mean cluster analysis. Clustering was done using the software package SPSS[18] and KyPlot.[19]

The values of the molecular descriptors were calculated using Cambridge software package Chemoffice program version 4 and 5[20] (Table 4). The values for the descriptors like Heat of formation, dipole moment, electronic energy, ionization potential, energy of the highest occupied molecular orbital, and energy of the lowest unoccupied molecular orbital were calculated after minimizing the energy of the molecule. Molar refractivity and logarithm of the partition coefficient were calculated using software package Cambridge software, pro version 4. Polarizability and molar volume were calculated using software packages ACD labs, chemsketch version 3.5.

4.4.3. Multiple regression analysis. The QSAR model used was Hansch. Multiple regression analysis was carried out and statistical quality of the equations[21,22] were justified by parameters like correlation coefficient r, standard error of the estimate (SE), variance ratio (F) at specified degrees of freedom (df), and constant terms of regression equation: regression coefficients and intercepts. Significance of the regression coefficients was justified by t-test. Predictor variables with higher P values were removed in developing the equation to get more acceptable equation with statistical quality.

4.4.4. Validation of regression equations developed by quantitative structure activity relationship (QSAR) studies. Validation methods are needed to establish the predictiveness of the regression equation generated by the model. Leave one out cross validation was used for the validation of the equation.[23,24] That means, one candidate is left out from the training set at a time and a regression is carried out. This should not show much change in the residuals and the new generated equation should be able to predict the activity of the left over candidate with in close proximity.

Acknowledgements

The authors are thankful to the CDRI Lucknow in providing the ^{1}H NMR data. The authors also wish to thank AVIK pharmaceutical, Mumbai for providing various reagents for the synthesis. The IC grant provided by the industrial commission of Gujarat for carrying out this work is acknowledged.

References and notes

1. John, A. B.; Jeremy, H. T. Drugs for Rheumatic Diseases. *Essentials of Pharmacology*, 3rd ed.; Harper and Row: Philadelphia, 1817; pp 304–317.
2. Seibert, K. et al. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12013–12017.
3. Chan, C. C. et al. *J. Pharmacol. Exp. Ther.* **1995**, *274*, 1531–1537.
4. Futaki, N.; Yoshikawa, K.; Hamasaka, Y.; Aria, I.; Higuchi, S. *Gen. Pharmacol.* **1993**, *24*, 105–110.
5. Grau, M. *Drugs Future* **1982**, *7*, 745–746.
6. Krame, J. B.; Capiris, T.; Sircar, J. C.; Connor, D. T.; Bornemeier, D. A.; Dyer, R. D.; Kuipers, P. J.; Kennedy, J. A.; Wright, C. D.; Okonkwo, G. C. N.; Lesch, M. E.; Schrier, D. J.; Boschelli, D. H. *Bioorg. Med. Chem.* **1995**, *3*, 403–410.
7. Nicholson, C. D. et al. *Ann. Rep. Med. Chem.* **1994**, *29*, 75.
8. Riet, B. V. et al. Amidox. NSC-343341. *Drugs Future* **1991**, *16*(11), 990–991.
9. Castaner, J. et al. *Drugs Future* **1992**, *17*, 193–196.
10. Wikel, J. H. et al. *J. Med. Chem.* **1980**, *23*, 368–372.
11. Information update. *Drugs Future* **1998**, *23*, 80.

12. Information update. *Drugs Future* **1999**, *24*, 90.
13. Information update. *Drugs Future* **2000**, *25*, 84.
14. Information update. *Drugs Future* **2001**, *26*, 72.
15. Pillai, A. D.; Rathod, P. D.; Franklin, P. X.; Padh, H.; Vasu, K. K.; Sudarsanam, V. *Biochem. Biophys. Res. Commun.* **2004**, *317*, 1067–1074.
16. Pillai, A. D., et al. Synthesis of tri and tetra substituted thiophenes and their process thereof. Indian Patent Application Number 951, 2001.
17. Everitt, B. S.; Graham, D. *Applied Multivariate Data Analysis*; Oxford University Press: New York, 1992.
18. SPSS version 7.5 for windows.
19. KyPlot version 5.1.2600.0 for windows.
20. ChemOffice, version 5.0 for windows.
21. Rawlings, J. O. *Applied Regression Analysis: A Research Tool*; Wadsworth: CA, USA, 1988; p 327.
22. Darper, N. R.; Smith, H. *Applied Regression Analysis*; John Wiley and Sons: New York, 1981; p 294.
23. Schaper, K. J. *Quant. Struct. Act. Relat.* **1999**, *18*, 354.
24. Tetko, I. V.; Tanchuk, V. Y.; Villa, A. E. *J. Chem. Int. Comput. Sci.* **2001**, *41*, 1407.
25. Keefer, L. F. *Pharm. News* **2000**, *7*(3), 27–32.