Notes

A Fujita-Ban Structure-Activity Analysis of 44 Steroids

J. B. Justice, Jr.

Department of Chemistry, Emory University, Atlanta, Georgia 30322. *Received August 15, 1977*

A Fujita–Ban structure–activity analysis has been performed on 44 steroids with a common pregn-4-ene-3,20-dione structure. The measured biological activity was glycogen deposition in the liver, expressed relative to the activity of cortisol. Substituents are at positions 2, 6, 9α, 11β, 16α, 17, and 21. Substituents of positions 2, 6, and 17 accounted for 60% of the total variance. The substituents with significance at the $p < 0.01$ level were 6α-hydrogen, 6α-methyl, 2α-hydrogen, 9α-fluorine, and 9α-hydrogen.

In the present paper the Fujita–Ban method of structure–activity relationship (SAR) analysis is applied to a set of 44 glucocorticoids in an attempt to quantitate some of the structure–activity relationships qualitatively known in steroids. A data base is also created for the comparison of some of the more well-known methods of structure–activity analysis. The study of steroid activity by mathematical techniques is particularly useful as a vehicle for the analysis and development of SAR methods because of the extensive knowledge available on action and metabolism of steroids. By comparing the results obtained through statistical analysis of structure–activity relationships in a series of similar steroids to the metabolic patterns of these molecules, some meaning may be given to the significance of certain structural features which enhance or inhibit activity. Such knowledge should be of benefit in the design of molecules which affect various cellular and physiological functions.

Data Set. Table I lists the specific structures and associated biological activities.[1,2] The experimental procedure used in ref 1 and 2 to determine biological activity was designed to measure glucocorticoid activity. It was based upon a comparison of the effect of the test compound with that of a standard glucocorticoid, cortisol, on glycogen deposition in the liver. The test compounds were administered to rats by a subcutaneous route 5 days after the animals had been bilaterally adrenalectomized. The amount of glucose in the liver was then measured and from the glucose concentration, the potency of the test compound relative to cortisol was determined. The standard error in the assay, expressed as percent relative potency, was 11.0%. A more complete description of the experimental procedure can be found in ref 1 and 2.

Figure 1 illustrates the pregn-4-ene-3,20-dione structure that is common to the molecules in the structure–activity study. The framework contains variations at seven positions: 2α, 6α, 9α, 11β, 16α, 17, and 21. The number and type of substituent at each of these positions are listed in Table II.

Method. The 44 structures listed in Table I were analyzed for substituent effect on glycogenic activity by the Fujita–Ban method.[3] In this method, the biological activity is assumed to arise from linearly independent contributions at each position in the molecule.

The Fujita–Ban coefficients can be derived from Free–Wilson[4] coefficients by a linear transformation in which the origin of the coordinate space is changed from the absence of any substituent to that of the substituent hydrogen. Instead of the constant term being the overall average of the biological activities, it becomes the activity of an all-hydrogen-substituted molecule. Thus the hydrogen coefficients of Table III are all zero. Kubinyi and Kehrhahn have shown how the Fujita–Ban coefficients may be obtained more directly.[5,6]

The 44 structures were encoded into 17 dimensional vectors, each variable of the vector containing a one or zero indicating the presence or absence of the nonhydrogen substituents.

Regression analysis was performed on the log of relative glycogenic potency (cortisol potency = 1) to yield the results in Table III.

Results and Discussion

The regression equation accounted for 75.4% of the total variance and had a multiple correlation coefficient of 0.868. The F for the overall regression was 4.69, which was significant at the 0.001 level. Table III lists the regression coefficients, standard error, and F for each substituent. Position 2α accounted for 27.0% of the total variance and 35.8% of the variance accounted for by the regression equation. Positions 17 and 6 were also significant contributors to the regression equation, accounting for 18.2 and 15.4% of the total variance, respectively. Together these three positions accounted for 60% of the total variance and 80% of the "explained" variance.

If one examines the substituents at these positions, it is found that for the 2α position, methyl is the most positive substituent but is not significant statistically. A double bond is almost as positive ($p < 0.025$) while hydrogen has a negative effect ($p < 0.005$). At the second most active position, carbon 17, the only positive substituent is hydroxyl ($p < 0.10$). Acetate is negative but not significant. Hydrogen was the most negative substituent ($p < 0.05$). The third most active position is 6α. A combination of double bond and methyl group is the most positive structural variation at carbon 6; however, this is not statistically significant. The statistically significant substituents are methyl ($p < 0.005$), a positive substituent, and hydrogen ($p < 0.001$), a negative one. It is interesting that a double bond occurs at both extremes of position 6α, positively with a methyl group and negatively with a hydrogen. It is as if the double bond has enhanced the effect of hydrogen and methyl functionalities.

The remaining positions, 21, 11β, 9α, and 16α, accounted for 6.2, 5.4, 2.0, and 1.1% of the total variance, respectively. At position 21, hydroxyl, the most positive substituent (0.705, $p < 0.025$), and hydrogen, the most negative (−1.208, $p < 0.025$), were the only statistically significant substituents. At 11β, the only significant structural variation of those tested was hydrogen (−4.482, $p < 0.025$). The 11β-hydrogen had the largest coefficient of the data set, indicating that the activity was greatly reduced by the

0022-2623/78/1821-0465$01.00/0

Table I. Structures of 44 Pregn-4-ene-3,20-diones

Compd no.	Log BA obsd	Log BA calcd	X_1, 2α			X_2, 6α					X_3, 9α			X_4, 11β			X_5, 16α			X_6, 17			X_7, 21			
			H	Me	DB[a]	H	DB,Me	DB	Me	F	H	F	Cl	H	O	OH	H	Me	OH	H	Ac	OH	H	Ac	F	OH
1	1.93	1.91			1				1			1				1	1					1				1
2	1.36	1.01			1	1						1				1	1					1		1		
3	1.15	1.26			1				1			1				1	1					1				1
4	1.12	1.05		1		1						1				1	1					1		1		
5	1.07	0.58	1							1	1					1	1					1		1		
6	1.03	1.09			1	1						1				1			1			1				1
7	0.93	0.33			1	1						1				1			1			1	1			
8	0.92	0.92	1				1				1					1	1					1		1		
9	0.90	0.57	1			1						1				1	1					1				1
10	0.89	0.77			1				1			1				1	1			1			1			
11	0.85	0.57	1						1		1					1	1					1		1		
12	0.78	0.35			1	1					1					1	1					1		1		
13	0.71	0.40		1		1					1					1	1					1		1		
14	0.64	0.52		1		1					1					1	1					1				1
15	0.53	0.52	1			1						1				1			1			1				1
16	0.43	0.48			1	1					1					1	1					1				1
17	0.37	1.14			1				1		1					1	1					1		1		
18	0.36	−0.06	1						1		1					1	1					1	1			
19	0.34	0.06	1						1			1				1	1			1			1			
20	0.32	0.39			1	1					1				1		1					1				1
21	0.29	−0.03	1						1		1					1	1					1			1	
22	0.26	0.26			1	1					1				1		1					1		1		
23	0.20	0.53			1				1		1					1	1					1			1	
24	0.20	0.70	1						1		1					1	1					1				1
25	0.12	−0.69	1						1		1				1		1			1			1			
26	0.10	0.10	1			1							1			1	1					1		1		
27	0.08	0.25			1			1			1					1	1					1				1
28	0.06	0.07	1							1		1				1	1			1			1			
29	0.00	−0.09	1			1					1					1	1					1				1
30	−0.01	−0.21		1		1						1			1		1			1			1			
31	−0.10	0.48	1						1		1				1		1					1		1		
32	−0.11	0.42		1		1						1				1	1					1	1			
33	−0.11	−0.12	1			1						1			1			1		1			1			
34	−0.15	−0.32	1					1			1					1	1					1				1
35	−0.28	−0.12		1		1						1				1	1			1			1			
36	−0.31	−0.72	1			1					1				1		1			1						1
37	−0.46	−0.18	1			1					1				1		1					1				1
38	−0.55	−0.63	1			1					1					1	1			1						1
39	−0.60	−0.60	1							1	1			1				1				1				1
40	−0.70	−0.58	1			1						1				1	1				1		1			
41	−0.80	−0.21	1			1					1					1	1					1		1		
42	−1.15	−0.68	1							1	1				1		1			1			1			
43	−1.30	−0.73	1			1						1				1	1			1			1			
44	−1.30	−0.77	1			1						1				1			1	1			1			

[a] DB = double bond, Me = methyl, Ac = acetate, O = carbonyl.

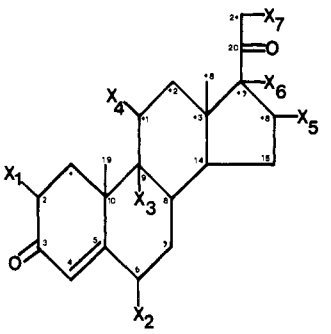


Figure 1. Pregn-4-ene-3,20-dione structure indicating positions of substitution.

Table II. Number and Type of Substituent at Each Carbon

Substituent	Position						
	2α	6α	9α	11β	16α	17	21
Hydrogen	25	25	26	1	38	11	14
Methyl	6	12			2		
Double bond	13[a]	2[b]					
Double bond, methyl		1[b]					
Hydroxyl				34	4	31	16
Carbonyl				9			
Fluoro		4	17				2
Chloro			1				
Acetate						2	12

[a] The double bond is a 1,2 double bond. [b] A 6,7 double bond.

presence of hydrogen. Position 9α had two significant substituents, fluorine, the most positive (0.905, $p < 0.01$), and hydrogen (−0.596, $p < 0.01$), the most negative.

None of the substituents at 16α had any statistical significance.

Table III indicates that hydrogen, by itself or with a double bond, is the most negative substituent in every position except 16α, where it and hydroxyl are approximately equal and close to zero. The hydrogen coefficient was statistically significant at all positions except 16α, where no substituents were significant. The observed vs.

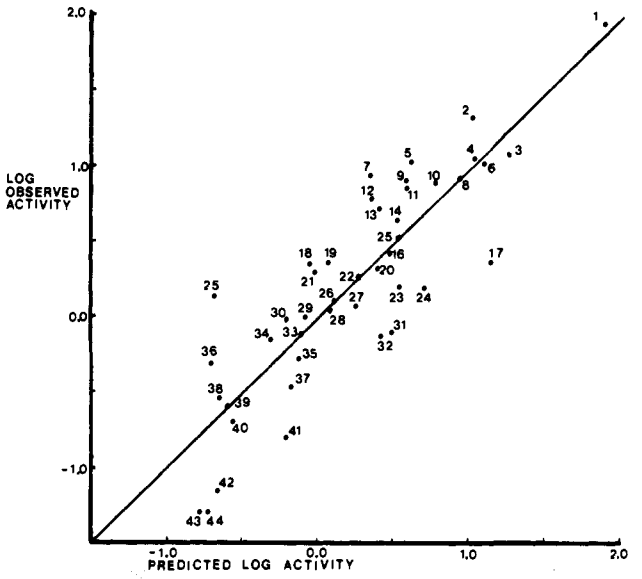


Figure 2. Graph of calculated vs. observed activity.

predicted activity is plotted in Figure 2. The standard error was 0.65. Figure 2 indicates that several structures have extreme deviations from the regression line. Positive deviation results from activity not accounted for by the regression equation. The structures having the greatest positive deviations are **25**, **7**, **4**, and **36**. The structures having the greatest negative deviation are **17**, **23**, **41**, **43**, and **31**. The most negative deviations tend to occur among the less active compounds, while the extreme positive deviations cover the range of activities. Neither set of structures exhibits any obvious pattern to explain the deviations. If one examines the residuals for less extreme but more systematic variation, a consistent positive deviation is observed at the upper end of the activity scale. Evidently some interaction effects are taking place among the substituents most contributing to activity or a logarithmic transformation was not appropriate. A second weaker trend occurs at the bottom of the activity scale

Table III. Regression Coefficients for Substituents

Carbon	Substituent	Fujita-Ban regression coeff	Std error	F^a
2	Double bond	0.564	0.122	6.61[e]
	α-Methyl	0.611	0.217	2.74
	α-Hydrogen	0.0	0.071	12.19[c]
6	Double bond, methyl	1.129	0.479	2.99
	α-Methyl	0.782	0.137	12.39[c]
	α-Fluorine	0.793	0.276	3.17
	Double bond	−0.108	0.344	2.36
	α-Hydrogen	0.0	0.080	14.13[b]
9	α-Fluorine	0.652	0.137	8.27[d]
	α-Chlorine	0.312	0.489	0.01
	α-Hydrogen	0.0	0.088	8.59[d]
11	β-Hydrogen	0.0	0.787	6.12[e]
	Carbonyl	1.920	0.165	0.03
	β-Hydroxyl	2.010	0.051	1.61
16	α-Methyl	0.702	0.501	1.81
	α-Hydroxyl	0.043	0.277	0.06
	α-Hydrogen	0.0	0.040	0.49
17	Hydrogen	0.0	0.189	4.14[f]
	Acetate	0.162	0.382	0.39
	Hydroxyl	0.536	0.076	4.01
21	Hydrogen	0.0	0.178	6.32[e]
	Acetate	0.630	0.151	1.47
	Fluorine	0.022	0.349	1.47
	Hydroxyl	0.754	0.117	6.82[e]
Constant		−3.388		

[a] F (for overall regression) = 4.69, significant at 0.001 level. [b] $p < 0.001$. [c] $p < 0.005$. [d] $p < 0.01$. [e] $p < 0.025$. [f] $p < 0.05$.

where the five least active compounds have activities less than that predicted by the regression equation. Both of these trends suggest that a linear model is inadequate to completely account for the systematic variation in activity with structure.

Relationship to Metabolism. The data indicate that positions 2, 6, and 17 are particularly significant to biological activity. Since one of the ways to increase the activity of a molecule is to prevent its degradation, it is interesting to compare these results to the metabolism of cortisol.

In the main metabolic pathway of cortisol, the molecule is first transformed at the Δ^4 double bond to the 11β,-17α,21-trihydroxy-5α- and -5β-pregnane-3,20-diones. These two products are in turn metabolized at the C_3 carbonyl to 3α- and 3β-hydroxyl groups. The third step in this path is at C_{20} where again a carbonyl is reduced to a hydroxyl. Other initial transformations of cortisol are to 6β-hydroxycortisol, 2α-hydroxycortisol, and the 20α- and 20β-tetroles.[7] Carbons 2 and 6 are sites of alternate pathways. Carbon 6 is also adjacent to the first degradation step in the main pathway, while 2 and 17 are adjacent to the second and third steps, respectively.

Carbon 6 is adjacent to the double bond that is opened to the 5α- or 5β-hydrogen. If the increase in activity is related to prevention of bond opening, then it appears that the most effective blockage is created by moving the methyl substituent from the α position to a position between α and β, created by the 6,7 double bond. The bond itself (more properly hydrogen and the bond) has the most negative coefficient, indicating that it is not the presence of the double bond that increases activity but the spatially directing effect which it has that is important. At position 2, the position controlling the greatest variance, a methyl group promotes activity while hydrogen inhibits activity. This could be a size effect but more substituents are needed before any conclusions can be drawn.

Conclusion

A set of 44 glucocorticoids has been analyzed for structure–activity relationships by the Fujita–Ban approach. Seven different positions on the molecule were analyzed for their contribution to the variation in biological activity. Positions 2α, 17, and 6α accounted for 60% of the total variance. The three positions accounting for the greatest variance were adjacent to metabolic points in the molecular framework.

Acknowledgment. The author thanks D. Jane Taylor, NCI, for providing the data upon which this work was performed and the Emory University Computing Center for computer time.

References and Notes

(1) A. G. Hilgar and D. J. Hummel, Ed., "Endocrine Bioassay Data", Issue 1, Cancer Chemotherapy National Service Center, National Cancer Institute, 1964.
(2) A. G. Hilgar and L. C. Trench, Ed., "Endocrine Bioassay Data", Issue 2, General Laboratories and Clinics, National Cancer Institute, 1968.
(3) T. Fujita and T. Ban, *J. Med. Chem.*, **14**, 148 (1971).
(4) S. M. Free and J. W. Wilson, *J. Med. Chem.*, **7**, 395 (1964).
(5) M. Kubinyi and O. Kehrhahn, *J. Med. Chem.*, **19**, 578 (1976).
(6) M. Kubinyi, *J. Med. Chem.*, **19**, 587 (1976).
(7) M. H. Briggs and J. Brotherton, "Steroid Biochemistry and Pharmacology", Academic Press, New York, N.Y., 1970.

Molecular Structure of 1-(2-Chloroethyl)-3-(*trans*-4-methylcyclohexyl)-1-nitrosourea

H. Warren Smith, Arthur Camerman,*

Departments of Medicine (Neurology) and Pharmacology, University of Washington, Seattle, Washington 98195

and Norman Camerman

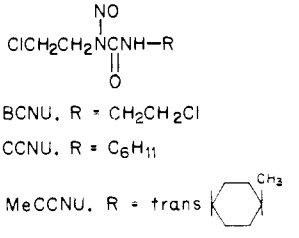
Department of Biochemistry, University of Toronto, Toronto, M5S 1A8, Ontario, Canada. *Received October 10, 1977*

The three-dimensional structure of 1-(2-chloroethyl)-3-(*trans*-4-methylcyclohexyl)-1-nitrosourea (MeCCNU, NSC-95441), an effective antitumor agent, has been determined by single-crystal x-ray diffraction. MeCCNU crystallizes in monoclinic space group $P2_1/c$, with cell dimensions a = 12.387, b = 10.810, and c = 10.198 Å, β = 102.62°, and Z = four molecules per unit cell. The structure was solved by direct phasing procedures and refinement by anisotropic least squares converged at a discrepancy index R = 0.065. The cyclohexyl ring is in the chair conformation with the plane of the nitrosourea moiety twisted approximately 90° from the cyclohexyl ring. The carbon–nitrogen bonds of the urea group are significantly asymmetric.

One of the most promising groups of compounds to be developed by the Chemotherapy Program of the National Cancer Institute is the 1-(2-chloroethyl)-3-alkyl-1-nitrosoureas.[1,2] The nitrosoureas are effective antitumor agents in the treatment of lymphomas[3] and solid tumors[4] in man. They are highly lipid soluble, cross the blood–brain barrier rapidly,[5,6] and exhibit unusual delayed bone marrow toxicity.[3]

The nitrosoureas are chemically reactive compounds that decompose nonenzymatically at relatively rapid rates under physiological conditions. Studies of their decomposition products reveal the formation of a highly reactive 2-chloroethyldiazene hydroxide and isocyanates,[7] with the former intermediate becoming an alkylating species, possibly in part as a 2-chloroethyl carbonium ion.[8] The mechanism of antitumor action of these compounds is not fully understood, but it is believed to involve passive diffusion of the intact nitrosourea and possibly the isocyanate across the cell membrane.[9] The isocyanates are capable of interacting with amino acids and proteins,[10–12]

Chart I. Chemical Structures of Clinically Useful Nitrosoureas



0022-2623/78/1821-0468$01.00/0