Journal of Computer-Aided Molecular Design, **12:** 557–561, 1998.
KLUWER/ESCOM

QSAR modeling with the electrotopological state indices: Corticosteroids

Carolina de Gregorio[a], Lemont B. Kier[b] & Lowell H. Hall[c]
[a]*University of Valencia, E-46100 Burjasot, Valencia, Spain;* [b]*Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23298, U.S.A.;* [c]*Department of Chemistry, Eastern Nazarene College, 23 East Elm Avenue, Quincy, MA 02170-2999, U.S.A.*

Received 10 March 1998; Accepted 5 June 1998

Key words: corticosteroids, E-state, electrotopological, QSAR, topological descriptor, topological superposition

Summary

A structure-activity analysis of a series of steroids binding to corticosteroid-binding globulin was made using the electrotopological state index for each atom in the molecule. Two indices were found to correlate well with the binding affinity. The indices encode structural characteristics in the A and the D rings of the steroids in the study. One of the indices was formulated as the difference between two indices in the A ring. The two were not intercorrelated, suggesting that the composite index signals the influence of structure changes in or near the A ring that can be monitored by the composite index. This is a new observation using this structure-activity method. It is suggested that this model makes some contributions towards detection of the pharmacophore.

Introduction

In QSAR models the representation of molecular structure has most often been accomplished with either physical properties such as partition coefficient or structure indices such as molecular connectivity chi indices. These structure representations have been largely based on the whole molecule. In this fashion the QSAR models offered little insight into drug pharmacophores. However the use of atom level indices may afford opportunity to probe the pharmacophore inherent in a set of closely related structures.

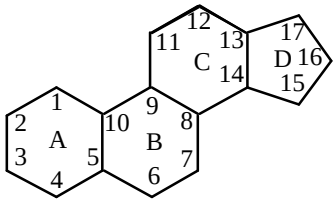
In drug design a critical step following the generation of a lead compound is the definition of a pharmacophore. A working definition of a pharmacophore is a pattern of atoms on an active molecule that is essential for that activity. The pattern includes the electronic structure of key atoms or bonds plus the spatial arrangement or topology of these features within the molecule. Classical approaches to the definition of a pharmacophore began with the search for more than one lead compound with significant activity. Comparisons among these molecules to find common features along with the intuition that heteroatoms are more important than carbon atoms was the line of reasoning taken to identify atoms and bonds constituting a pharmacophore. A second approach was the selective modification of a lead compound in which parts of a molecule were eliminated or changed in order to establish their importance. By altering different features in a lead compound it is possible to build a pattern that constitutes a putative pharmacophore. A third approach has been to synthesize molecules bearing some relationship to a lead compound, holding a possible pharmacophore constant while varying other, assumed inert parts of a molecule.

The role of the pharmacophore has never been more important in drug design. With the advent of combinatorial chemistry, the generation of potential lead compounds has entered the discovery scene on a very large scale. It is possible now, in a very short time to produce large numbers of molecules and to test them for selected activities using robotic systems. The definition of a pharmacophore takes on a correspondingly greater significance and priority in drug design. New ways to define the electronic and the topological aspects of atoms and features within molecules are now at a premium.

With the advent of all-valence electron molecular orbital theory [1] and high-speed computers, it became

possible to study simultaneously the electronic and conformation or topological structure of a lead compound. This led to the development of a general approach called receptor mapping [2]. This has evolved into the approaches used under the general term computer graphics to define and design molecules based upon a hypothetical pattern, the pharmacophore. From these efforts there has evolved the clear recognition that electronic and topological structure are essential ingredients in a definition of molecular structure and a pharmacophore. This proposition was recognized in the design of a unified index of atom-level fragments carrying information about electronic and topological structure called the electrotopological state (E-State) [3–6]. The appendix is devoted to a description of this index. The reader is also referred to an earlier article in this journal [7]. Increasing use of this index has demonstrated its utility in defining essential fragments of molecules in quantitative structure-activity studies [8–10]. Because the information generated by these indices is focused on sub-molecular fragments, especially at the atom level, it is logical that it may be exploited for the probing of a pharmacophore in a particular series of molecules. In the present study, we illustrate this ability of the E-state indices to shed light on a pharmacophore governing a particular biological activity.

To illustrate the use of the E-state to map a pharmacophore we have chosen a set of compounds that has been used in a number of studies in which structure-activity information has been the goal. This study is made on a series of 31 steroids with varying binding affinity for the corticosteroid-binding globulin (CBG) [11]. This same set of data was used by Cramer in his development of the comparative molecular field analysis (CoMFA) [12]. Later, Good compared Cramer's results with those obtained using data matrices derived from similarity studies using different indices [13–14]. Polanski proposed the use of molecular similarity indices derived from silhouettes and self-organizing maps using this group of compounds [15]. Wagener proposed a 3-D autocorrelation descriptor for this data set using neural networks [16]. Finally Kellogg employed these molecules in a pioneering work involving the use of the E-state and a new descriptor, the hydrogen E-State, as fields in a modified CoMFA study [7].



Structure I.

Biological data

A solid phase competitive binding assay has been used to measure the affinity constants, K_{aff}, for the binding of steroids to corticosteroid-binding globulin (CBG) in human plasma at 37 °C and pH 7.4 [17–19]. This information is expressed as the pK_{aff}.

Structure-activity analysis

Of the 31 compounds in the original data set only 30 were used in this study. The omitted molecule, aldosterone, presents an uncertain structure since some fraction of the molecule exists as a ketone whereas it is usually represented as a hemiacetal. The E-state indices were calculated for each atom in the set (numbered as shown in Structure I) and were regressed individually and in multiples against the pK_{aff} values for the 30 steroids. The structures are shown in Table 1 and the biological data is shown in Table 2.

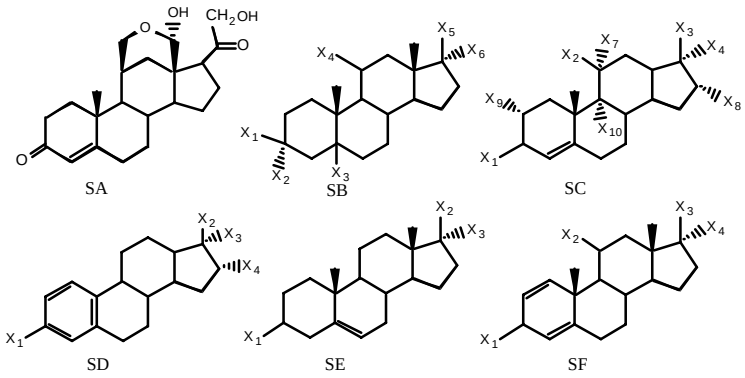
Results

Three E-State values were found to be of significance when regressed against the pK_{aff} values. A relating equation was found to be:

$$\begin{aligned} pK_{aff} = & 0.318(\pm 0.068)S(16) \\ & + 1.615(\pm 0.210)S(4) \\ & - 1.600(\pm 0.197)S(10) + 4.037(\pm 0.357) \\ & r^2 = 0.821,\ s = 0.492,\ n = 30,\ F = 40 \\ & r^2_{press} = 0.783,\ s_{press} = 0.541 \end{aligned}$$

The three E-State values are not significantly intercorrelated. It is of great interest that the coefficients for S(4) and S(10) are nearly identical but of opposite sign, yet they are not intercorrelated. It is possible that these indices reflect two parts of a structural singularity as is sometimes encountered in a classical Free–Wilson *de novo* analysis. As such, they relate to

Table 1. Structures of the corticosteroids used in this E-state analysis of corticosteroid binding globulin data for a two-variable model with E-state indices



id	Structure[a]	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	SB	OH	H	H[b]	H	OH	H				
2	SE	OH	OH	H							
3	SC	=O	H	=O				H	H	H	H
4	SB	H	OH	H[b]	H	=O					
5	SC	=O	OH	$COCH_2OH$	H			H	H	H	H
6	SC	=O	OH	$COCH_2OH$	OH			H	H	H	H
7	SC	=O	=O	$COCH_2OH$	OH				H	H	H
8	SE	OH	=O								
9	SC	=O	H	$COCH_2OH$	H			H	H	H	H
10	SC	=O	H	$COCH_2OH$	OH			H	H	H	H
11	SB	=O		H[b]	H	OH	H				
12	SD	OH	OH	H	H						
13	SD	OH	OH	H	OH						
14	SD	OH	=O		H						
15	SB	H	OH	H[c]	H	=O					
16	SE	OH	COMe	H							
17	SE	OH	COMe	OH							
18	SC	=O	H	COMe	H			H	H	H	H
19	SC	=O	H	COMe	OH			H	H	H	H
20	SC	=O[d]	H	OH	H			H	H	H	H
21	SF	=O	OH	$COCH_2OH$	OH						
22	SC	=O	OH	$COCH_2OCOMe$	OH			H	H	H	H
23	SC	=O	=O	COMe	H				H	H	H
24	SC	=O	H	$COCH_2OH$	H			OH	H	H	H
25	SC[e]	=O	H	OH	H			H	H	H	H
26	SC	=O	H	COMe	OH			H	OH	H	H
27	SC	=O	H	COMe	H			H	Me	H	H
28	SC[e]	=O	H	COMe	H			H	H	H	H
29	SC	=O	OH	$COCH_2OH$	OH			H	H	Me	H
30	SC	=O	OH	$COCH_2OH$	OH			H	H	Me	F

[a] Structures according to references [17–19].
[b] Of the 5-α steroid series.
[c] Of the 5-β steroid series.
[d] Assumed to be =O (testosterone) as indicated by reference [18] and not as -OH in table and further publications [compare reference [18] for mistakes in previous publications].
[e] H (hydrogen) instead of Me at C_{10} steroid skeleton.

a more general part of the molecule that is influencing its behavior. To test this possibility, we re-examined the relationship with pK_{aff} using the numerical difference between S(4) and S(10). Using this composite parameter, S(4, 10), we regressed the set of E-states against the binding affinity, pK_{aff}, to find the best two-variable equation:

$$\begin{aligned} pK_{aff} = \; & 0.319(\pm 0.065)S(16) \\ & +1.607(\pm 0.145)S(4,10) \\ & +4.051(\pm 0.228) \\ & r^2 = 0.821, s = 0.483, n = 30, F = 62 \\ & r^2_{press} = 0.792, s_{press} = 0.521 \end{aligned}$$

The predicted values from this equation are shown in Table 2.

Table 2. Observed and calculated binding data for the corticosteroid binding globulin data based on the two-variable E-state model

Obs	Structure[a]	PK_{obs}	PK_{calc}[b]	Res[c]	Pres[d]
1	SB	5.000	5.31	−0.31	−0.33
2	SE	5.000	5.30	−0.30	−0.33
3	SC	5.763	7.08	−1.32	−1.39
4	SB	5.613	5.26	0.35	0.38
5	SC	7.881	7.42	0.46	0.50
6	SC	7.881	7.34	0.54	0.58
7	SC	6.892	7.42	−0.53	−0.57
8	SE	5.000	5.26	−0.26	−0.28
9	SC	7.653	7.17	0.48	0.51
10	SC	7.881	7.10	0.78	0.83
11	SB	5.919	5.09	0.83	0.92
12	SD	5.000	5.15	−0.15	−0.17
13	SD	5.000	4.72	0.28	0.33
14	SD	5.000	5.11	−0.11	−0.12
15	SB	5.225	5.26	−0.04	−0.04
16	SE	5.225	5.34	−0.12	−0.13
17	SE	5.000	5.26	−0.26	−0.28
18	SC	7.380	7.19	0.19	0.21
19	SC	7.740	7.11	0.63	0.67
20	SC	6.724	7.12	−0.40	−0.42
21	SF	7.512	7.48	0.03	0.03
22	SC	7.553	7.36	0.19	0.20
23	SC	6.779	7.51	−0.73	−0.79
24	SC	7.200	7.42	−0.22	−0.23
25	SC	6.144	6.44	−0.29	−0.31
26	SC	6.247	6.67	−0.43	−0.46
27	SC	7.120	7.02	0.10	0.10
28	SC	6.817	6.50	0.32	0.34
29	SC	7.688	7.41	0.28	−0.29
30	SC	5.797	5.82	−0.03	−0.29

[a] See Table 1.
[b] See text for two-variable equation.
[c] $Res = PK_{obs} - PK_{calc}$.
[d] Predicted residual based on the leave-one-out method.

Discussion

The structure-activity relationship provides several pieces of information. At the molecular level, the equation model reveals two regions in the molecules that are strong contributors to the binding of these steroids at the corticosteroid-binding globulin (CBG) site.

Specifically, the information implicates the A and the D rings of the steroids in this study. In the case of the D ring, structure variations among the steroid series take the form of substituents on the 16 or 17 positions. The 17-substituents are substituted with hydroxyl, keto or acyl groups. The 16 position is occasionally substituted with a methyl or a hydroxyl group. Further, it is the variation in the position 16 E-state index which is related to binding even though other atoms in the D ring encode electronic and topological information related to the substitution pattern. The E-state for the 16 position carries the information about the substitution pattern in the D ring, the information which is significant for binding. Lower values for S(16) correspond to lower binding affinity. This E-state trend at 16 is a consequence of more electronegativity at or near the 16 and 17 positions. As an example, the presence of an acyl group on the 17 position leads to greater binding and a higher E-state than molecules with a hydroxyl group at position 17. These relationships between substitution patterns on the D ring and binding signaled by the 16 position E-state suggest the possibility of a pharmacophore feature associated with the D ring.

Another important feature in the equation model is the relationship of the binding affinity to an index formed from the difference between S(4) and S(10). This is another example of a composite index that we have used in our studies employing the E-state [5]. This index, labeled S(4, 10), carries electronic and topological information arising from the substitution pattern at the 4 and 10 positions but also, and perhaps more importantly at all positions in the A ring. Even though other A-ring atom E-state indices encode related information, it is the specific variation at these two sites which is related to binding. Like the example just described for S(16) signaling the effects on the D ring, the S(4,10) composite E-state index signals struc-

tural changes on the A ring governing binding affinity. Specifically, it is because other atom E-state indices do not contribute significantly to the QSAR model that it can be said that the S(4,10) composite index suggests the possibility of pharmacaphore contributions at these sites.

The general interpretation drawn from this model is the possibility that the pharmacophore may include contributions from the A ring as well as the D ring.

Other QSAR analyses of this data set have been indicated earlier. The statistical results shown here indicate that this model is better than those reviewed by Cramer [12]. Further, it is clear from this analysis that the use of the E-state is far simpler than that of the CoMFA method. No problems with developing three-dimensional structures arises. Further the troublesome alignment problem is completely avoided. The approach developed here may be called a topological superposition method [20]. The E-state indices are computed very rapidly in a straightforward manner from the molecule connection table. Structure information is easily available in modern data bases. The statistical analysis is quite straightforward, using standard multiple linear regression methods, now widely available. It is not necessary to display the structures on a graphics terminal although a field representation of the molecules may easily be accomplished [7] if desired. Finally, one may readily develop the structure interpretation. The specific atom E-state indices point to the areas of interest. The role of substituent and skeletal atoms is directly understood in terms of classical ideas of electronegativity and bonding schemes.

Conclusions

This study uncovers a new aspect of E-state identification of important structural features. A functional region of a molecule with potential as a pharmacophore may be illuminated by E-state indices, including a composite E-state index made up of the difference (or perhaps the sum) of two E-state indices. This is illustrated by a good correlation between the affinity of a series of steroids and two E-state indices associated with different parts of molecules in a series.

Acknowledgements

C. de G. thanks the University of Valencia for support to study at the Virginia Commonwealth University.

References

1. Hoffmann, R., J. Chem. Phys., 39 (1963) 1397.
2. Kier, L.B., Mol. Pharmacol., 3 (1967) 487.
3. Kier, L.B. and Hall, L.H., Pharm. Res., 7 (1990) 801.
4. Hall, L.H. and Kier, L.B., J. Chem. Inf. Comput. Sci., 31 (1991) 76.
5. Hall, L.H. and Kier, L.B., Quant. Struct.-Act. Relat., 10 (1991) 43.
6. Kier, L.B., Hall, L.H. and Frazer, J.W., J. Math. Chem., 7 (1991) 229.
7. Kellogg, G.E., Kier, L.B., Guillard, P. and Hall, L.H., J. Comput.-Aided Mol. Design, 10 (1996) 513.
8. Kier, L.B. and Hall, L.H., Adv. Drug Res., 22 (1992) 1.
9. Hall, L.H. and Kier, L.B., Med. Chem. Res., 2 (1992) 497.
10. Hall, L.H. and Kier, L.B., J. Chem. Inf. Comput. Sci., 35 (1995) 1039.
11. Mickelson, K.E., Forsthoefel, J. and Westphal, U., Biochemistry, 20 (1981) 6211.
12. Cramer, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
13. Good, A.C., Peterson, S.J. and Richards, W.G., J. Med. Chem., 36 (1993) 2929.
14. Good, A.C., So, S.S. and Richards, W.G., J. Med. Chem., 36 (1993) 553.
15. Polanski, J., J. Chem. Inf. Comput. Sci., 37 (1997) 553.
16. Wagener, M., Sadowski, J. and Gasteiger, J., J. Am. Chem. Soc., 117 (1995) 7769.
17. Dunn, J.F., Nisula, B.C. and Rodbart, D., J. Clin. Endocrin. Metab., 53 (1981) 58.
18. Mickelson, K.E., Forsthoefel, J. and Westphal, U., Biochemistry, 20 (1981) 6211.
19. Westphal, U., Steroid-Protein Interactions II, Springer Verlag, Berlin (1986).
20. Kier, L.B. and Hall, L.H., Molecular Structure Description: The Electrotopological State, Academic Press, San Diego, CA (in press).

Appendix

The E-state index value for an atom, i, in a molecule is given as S_i:

$$S_i = I_i + \Sigma \Delta I_{ij}. \quad (1)$$

The summation is over all other atoms j within the molecular skeleton. The term for the intrinsic state, I, of atom i in Equation 1 is:

$$I_i = (\delta_i^v + 1)/\delta_i, \quad (2)$$

where δ^v is the count of all valence electrons on a bonding atom other than to hydrogen. The δ value is the count of sigma electrons on a bonding atom other than to hydrogen. The perturbation term in Equation 1 is defined as:

$$\Delta I_{ij} = (I_i - I_j)/r_{ij}^2, \quad (3)$$

where r_{ij} is the number of atoms in the shortest path between atoms i and j.