Vol. 99, No. 2, 1981 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
March 31, 1981 Pages 451-457

A STUDY OF THE APPLICABILITY OF QSAR CALCULATION FOR PEPTIDE HORMONES

László Nádasdi and Kálmán Medzihradszky

Institute of Organic Chemistry, Eötvös Lorand University,
Budapest, Muzeum krt. 4/B, 1088 Hungary

Received February 4, 1981

SUMMARY - A quantitative structure-activity relationship was calculated for the hypothalamic hormone LH-RH. A very good correlation ($R > 0.95$) and statistically significant equations at 95% confidence limit were obtained using information on the lipophilic nature (π) and steric characteristics (Y_Y) of the amino acid side chains.

During the last two decades there has been an enormous development in the field of quantitative structure-activity (QSAR) calculations for biologically active compounds. The methods of calculation can be classified into two fundamental groups: (i) the "extrathermodynamic method" renewed and developed by Hansch (1), using physicochemical or quantum mechanical parameters, and (ii) the so-called "de novo" method introduced by Free and Wilson (2), using computed contributions of individual substituents.

Among the surprisingly few attempts for QSAR calculations of peptide hormones, one of the first can be attributed to Sneath (3), who found moderate correlations between the similarities or differences in the structures and the activities of oxytocin, vasopressin and angiotensin. Later Simon (4) simplified this method and found somewhat more meaningful but still rather poor correlation between the structures and activities of these hormones. Recently Kaurov (5) elaborated a characterization method for amino acids starting from similar considerations: the complex properties of the side chains were described with only one number. The calculated data correlated well with the biological activity only of some hypophyseal hormone analogs. Another kind of structure-activity approach was reported by Rudinger (6), who established a lipophilic character - biological activity relationship

0006-291X/81/060451-07$01.00/0

for oxytocin analogs varied in position 4. Likewise, Jorgensen and Weinkam (7) found a similar relationship for angiotensin analogs substituted in position 5, after separation of the substituents according to their branching on the β-carbon atom.

The low number of QSAR calculations for peptide hormones may be attributed to the fact that modification of the peptide chain at a certain point changes not only the well-defined physicochemical properties, but also the conformation of the molecule, which is important in the binding to the receptor and, as a consequence, in eliciting the biological response. On the other hand, the correct characterization by one or more parameters of the rigidity or flexibility of the peptide chain seems to be insolvable at present.

The physicochemical properties of the amino acids which are of major importance for QSAR calculations can be classified as follows: (i) the lipophilic character, playing a decisive role in the action of many drugs; (ii) steric properties characterized by the volume, geometry and branching of the side chains; (iii) the electronic characteristics, acidic/basic nature and H-bridge forming ability of the molecule.

In this paper we report our results on the QSAR calculation for luteinizing hormone-releasing hormone (Glp-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH_2, LH-RH) analogs varied in position 6 (table 1), which have been synthesized and tested in vivo by Coy et al. (8,9). Although on the basis of this work only a limited number of compounds were available, the fact that their biological activity was determined in the same laboratory makes their comparison more reliable. The main point for the selection of the appropriate physicochemical properties was that they should be easily expressed by simple parameters and treated theoretically with sufficient accuracy.

CALCULATION PROCEDURE AND RESULTS

For the sake of simplicity we did not consider analogs containing L-amino acids in position 6 (all of them exhibit very low activity); we also omitted the parent hormone, Gly^6-LH-RH. As a rule, we calculated the multiple regression coefficients (R) for all the one-, two- and three-membered combinations

TABLE I

Luteinizing hormone - releasing potencies of LH-RH analogs
(measured in vivo on immature male rats Coy et.al.(8,9))

Substituent in pos. 6	Biological response[x] (%)	lg BR_{obs}	Calculated lg BR values and deviations					
			eq.1	Δ	eq.6	Δ	eq.7	Δ
Gly (LH-RH)	100.0	2.000						
D-Ala	708.2	2.850	2.576	-0.274	2.867	0.017	2.873	0.023
D-Leu	942.1	2.974	2.881	-0.093	2.892	-0.082	2.967	-0.007
D-Glu	188.7	2.276	2.451	0.175	2.410	0.134	2.270	-0.006
D-Phe	1075	3.031	2.933	-0.098	2.948	-0.083	3.024	-0.007
D-Trp	1329	3.124	2.933	-0.191	2.950	-0.174	3.026	-0.098
D-Phe(Me_5)	2267	3.355	3.409	0.054	3.482	0.127	3.337	-0.018
D-Phg	255.2	2.407	2.822	0.415	2.420	0.013	2.425	0.018
D-Phe(F_5)	1208	3.082	3.093	0.011	3.128	0.046	3.178	0.096

[x]Calculated on molar basis

D-Phe(Me_5) = D-pentamethylphenylalanine, D-Phg = D-phenylglycine,
D-Phe(F_5) = D-pentafluorophenylalanine

of the selected properties (not more than 8 at the same time). This was followed by checking the equations having high "R" values statistically, by the "F" and "Student t" tests. As the criterium of acceptability of a regression equation, 95% confidence limit has been chosen. The computation has shown that in one-membered combinations it is only the lipophilicity which can describe the structure-activity relationships with good statistics (equation 1, see explanation of symbols and their numerical values in Table II; BR = biological response).

(1) $\lg BR = 2.491 + 0.2085.\Pi$

$n = 8 \quad R = 0.803 \quad s = 0.237 \quad F = 10.88 \quad t(0) = 17.003 \quad t(\Pi) = 3.299$

Using the molecularpolarizability, no good fit was attained. This parameter is one of the electronic properties of the molecule, giving information about the magnitude of the London dispersion forces existing between the molecules. Statistical analysis of equation 2 shows that it cannot be accepted; its moderate value is due to the close relation between the Π and MR values

(2) $\lg BR = 2.455 + 0.01580.MR$

$n = 8 \quad R = 0.625 \quad s = 0.310 \quad (?)F = 3.85 \quad t(0) = 9.970 \quad (?)t(MR) = 1.962$

(see the intercorrelation matrix in Table III). Analysis of the observed and calculated lg BR values according to equation 1 (see also Table I) makes it clear that the large deviations for the D-Ala and D-Phg analogs (positive and negative, respectively) are the consequence of the absence (D-Ala) or presence (D-Phg) of steric hindrance near the α-carbon atom. Starting from this observa-

TABLE II

Physicochemical properties of some amino acid side chains

	$\Pi^{(a)}$	MR(b)	$V_{\gamma,c}$ (c)	$Y_{\gamma,c}$ (d)	$V_{\gamma,mv}$ (e)	$Y_{\gamma,mv}$ (d)
Ala	0.407	5.65	39.61	0.000	33.36	0.000
Leu	1.874	19.62	51.05	1.000	50.04	1.000
Glu	-0.189	16.52	51.05	1.000	50.04	1.000
Phe	2.121	30.01	51.12	1.006	43.43	0.604
Trp	2.121	39.85	51.05	1.000	43.43	0.604
Phe(Me_5)	4.405	53.13	51.12	1.006	43.43	0.604
Phg	1.591	25.36	62.15	1.971(f)	46.89	0.811
Phe(F_5)	2.890	28.92	51.12	1.006	43.43	0.604

(a) $\Pi_X = \lg P_X - \lg P_{Gly}$, where P denotes the partition coefficient of the corresponding amino acid in octanol/water system; lg P values were calculated from Rekker's hydrophobic fragmental constants (10).
(b) molar refractivity indices (molecular polarizability) were calculated from Vogel's data (11)
(c) the sum of the volumes of the α-γ atoms were calculated from the atomic distances and van der Waals radii employed for amino acids by Scheraga (12); the overlapping volumes were subtracted
(d) the branching indices were calculated according to equation 3.
(e) these values were computed from the following data: V(C,aliphatic)=16.68
V(C,aromatic) =10.07
(f) when corrected: $Y'_{\gamma,c}$ (Phg)=2.276

tion, we derived a steric parameter (Y_γ) for α-amino acids (X) according to equation (3)

$$Y_\gamma(X) = \frac{V_\gamma(X) - V_\gamma(Ala)}{11.44} \qquad (3)$$

where V_γ denotes (Table 2) the volumes of the β and γ atom of the side chain in addition to that of the α -CH group (e.g. $CH-CH_2-C$ for leucine). The constant in eq(3) corresponds to the difference between the volumes of the C-H and C-C groups. Two sets of Y_γ values were computed; the first one ($Y_{\gamma,c}$) was based on V_γ values calculated from the atomic distances and van der Waals radii. To obtain the second set ($Y_{\gamma,mv}$), the group volumes calculated from molar volumes of 60 aliphatic and aromatic hydrocarbons by the least square method were used. Combination of these parameters with the lipophilicity gave equations 4 and 5:

(4) $\lg BR = 2.798 + 0.2404.\Pi - 0.3687.Y_{\gamma,c}$

n = 8 R = 0.954 s = 0.132 F = 24.84 t(0) = 24.377
t(Π) = 6.658
$t(Y_{\gamma,c})$ = -3.799

(5) $\lg BR = 2.756 + 0.2084.\Pi - 0.4054.Y_{\gamma,mv}$

n = 8 R = 0.876 s = 0.210 F = 8.21 t(0) = 13.164
t(Π) = 3.715
(?)$t(Y_{\gamma,mv})$ = -1.615

TABLE III

Intercorrelation matrix of the variables

	Π	Π^2	MR	$Y_{\gamma,c}$	$Y'_{\gamma,c}$	$Y_{\gamma,mv}$
Π	1	0.927	0.869	0.233	0.185	0.000
Π^2		1	0.853	0.110	0.063	0.021
MR			1	0.372	0.310	0.165
$Y_{\gamma,c}$				1	0.993	0.690
$Y'_{\gamma,c}$					1	0.630
$Y_{\gamma,mv}$						1

The advantage of the $Y_{\gamma,c}$ parameter is obvious (equation 4); application of the $Y_{\gamma,mv}$ parameter led, however, not only to a worse fit, but the statistics of equation 5 was also poor (low $t(Y_{\gamma,mv})$ value). It is noteworthy that the molecularpolarizability, even when combined with the Y_{γ} parameter, did not give a good fit (R=0.858), nor a statistically acceptable equation. Considering the difference in the directions of the γ-atoms in Phg and other amino acids, and employing the corresponding correction by dividing the $Y_{\gamma,c}$ (Phg) value by a factor of cos 30° (the bond directions around a tetrahedral Cβ atom being 2x60°, looking from the Cα-Cβ axis, while in the case of Phg they are 2x90°), a better correlation was obtained:

(6) $\lg BR = 2.772 + 0.2338.\Pi - 0.3181.Y'_{\gamma,c}$

$n = 8 \quad R = 0.957 \quad s = 0.127 \quad F = 26.90 \quad t(0) = 26.254 \quad t(\Pi) = 6.787 \quad t(Y'_{\gamma,c}) = -3.987$

Combination of the lipophilicity with the aromatic character, marked only with a single dummy variable (I_{ar}) also failed to give either a good fit (R=0.828) or acceptable statistical values.

Although the investigation of three-membered combinations of parameters for 8 points may be criticized from the statistical point of view, we present the following equation which is significant at 95% confidence limit:

(7) $\lg BR = 2.174 + 0.4078.\Pi - 0.04150.\Pi^2 - 0.3660.Y'_{\gamma,c}$

$n = 8 \quad R = 0.989 \quad s = 0.071 \quad F = 61.64 \quad t(0) = 44.310 \quad t(\Pi) = 7.602 \quad t(\Pi^2) = -3.473 \quad t(Y'_{\gamma,c}) = -7.853$

Other parameter combinations gave no acceptable equations. For example the Π, $Y_{\gamma,c}$ and I_{ar} combination had only R = 0.961 and $t(I_{ar}) < t(4,0.975)$.

DISCUSSION

The quantitative structure-activity analysis presented clearly shows that in the case of LH-RH analogs substituted in position 6 by D-amino acids,

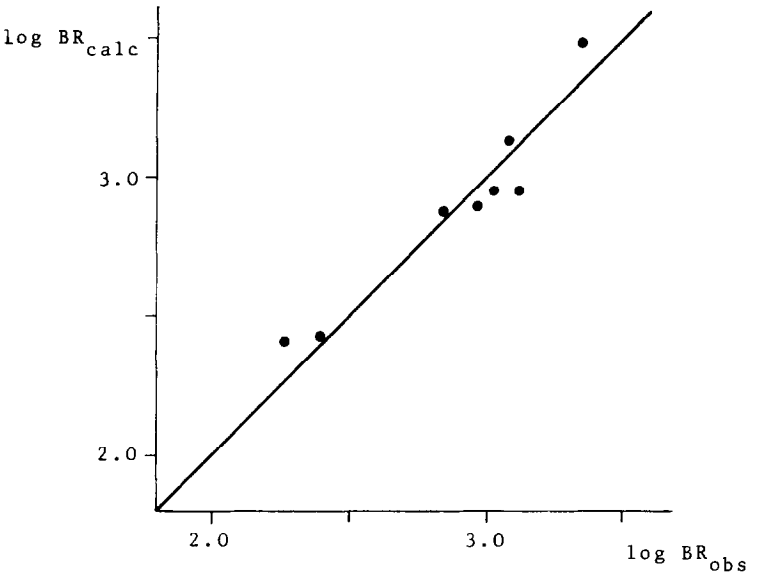

Figure 1. - Graphical analysis of equation 6. Solid line represents the ideal fit (R=1.000).

lipophilicity is the most important physicohcemical property regulating the biological activity of the molecule. It alone is responsible for 64.5% of the lg BR variance. The transport and/or binding to a receptor are regulated mainly by this factor. However, as the correlation using the lipophilicity alone is not sufficient, the queston arises, whether steric or other properties play a significant role in this case. Steric parameters may act by several means; our results clearly demonstrate the usefulness of considering one of them, $Y_{\gamma,c}$, by measuring the steric requirements of the individual amino acid side chains in the close vicinity of the C^{α}-atom. The lipophilicity and the $Y_{\gamma,c}$ parameters together can account for 90.8% (or 91.6%) of the lg BR variance (see diagram in Fig.1). The Y_{γ} value is probably characteristic of the felxibility of the peptide chain near to the C^{α} -atom, which varies according to the nature of the side groups. On this basis we conclude that highly lipophilic D-amino acids in position 6 of the LH-RH are favourable for the biological activity, whereas residues bearing a substitution or even more a branching on the C^{β} atom act in the opposite direction, by making the

peptide chain more rigid. The aromatic character of the side chains does not seem to have any influence on the biological action. Equation 7, in spite of its excellent fit and good statistics should be treated carefully because of uncertainties in the determination of the lipophilicity of the Phe(Me_5) and Phe(F_5) side chains. As Π_{opt}, calculated from equation 7, is higher than the lipophilicity of any of the amino acids listed in Table I, it is very likely that an LH-RH analog containing a more lipophilic residue in position 6 will possess higher biological activity.

ACKNOWLEDGEMENT - The authors are greatly indebted to Dr.J.Seprodi, for valuable discussions during this work and for drawing their attention to this remarkable series of LH-RH analogs.

REFERENCES

1. Hansch, C. (1969) Acc. Chem. Res. 2, 232-239.
2. Free, S.M. and Wilson, J.W. (1964) J. Med. Chem. 7, 395-9.
3. Sneath, P.H.A. (1966) J. Theor. Biol. 12, 157-195.
4. Simon, Z. (1968) Rev. Roum. Biochim. 5, 319-324.
5. Kaurov, O.A. (1978) Bioorg. Khim. 4, 604-618.
6. Rudinger, J. in "Endocrinology 1971" (S. Taylor, ed.) 12. William Heinemann Medical Books, London (1972)
7. Jorgensen, E.C. and Weinkam, R.J. in "Peptides 1973", Proc. 11th Eur. Pept. Symp., (H. Nesvadba, ed.), 311-323.
8. Coy, D.H., Vilchez-Martinez, J.A. and Schally, A.V., in "Peptides 1976", Proc. 14th Eur. Pept. Symp., (A. Loffet, ed.), 463-469; and personal communication from this laboratory.
9. Coy, D.H., Vilchez-Martinez, J.A. and Meyers, C.A. in Clinical Neuroendocrinology: A Pathophysiological Approach, 1979 (G. Tol, ed.), 83-88.
10. Rekker, R.R., Pharmacochemistry Library, Vol. 1. (1977) The Hydrophobic Fragmental Constants (Ed. W.Th. Nauta and R.F. Rekker)
11. Vogel, A.I. (1948) J. Chem. Soc. 1833-18 55.
12. Scheraga, H.A. (1968) Adv. in. Phys. Org. Chem. 6, 103-184.