The Prediction of Bradykinin Potentiating Potency of Pentapeptides. An Example of a Peptide Quantitative Structure-activity Relationship

Sven Hellberg, Michael Sjöström and Svante Wold*

Research Group for Chemometrics, Umeå University, S-901 87 Umeå, Sweden

Hellberg, Sven, Sjöström, Michael and Wold, Svante, 1986. The Prediction of Bradykinin Potentiating Potency of Pentapeptides. An Example of a Peptide Quantitative Structure-activity Relationship. – Acta Chem. Scand. B 40: 135–140.

The variation in amino acid sequence, in a set of bradykinin potentiating pentapeptides, is described by three variables per amino acid position. The variables were derived from a principal components analysis of a property matrix for the 20 coded amino acids. The resulting structure descriptor matrix describes the observed activity of the peptides to 97 % by means of a multivariate partial least squares (PLS) model. It is demonstrated that this quantitative structure-activity relationship (QSAR) can be used to predict the activity of new peptide analogs.

Introduction

The increasing interest in natural and synthetic peptides and their activity in various biological processes is an incentive to find relationships between the variation in peptide amino acid sequence, and the measured peptide activity.[1] We here demonstrate such a relationship for one set of peptides. The relationship is based on the same principle of quantitative analogy models that previously have been shown to apply in structure-reactivity relationships in organic chemistry, *e.g.*, the Hammett relationship.[2,3] Here the modelling rests on two cornerstones:

1. The characterisation of each individual amino acid by three scales (z_1, z_2 and z_3),[4] see Table 1. These scales are derived by a principal components analysis[5] of a matrix consisting of 20 properties for the 20 coded amino acids (Table 2). We are presently working with an extension of these scales to non-coded amino acids. This characterisation is similar to the analysis of Sneath[6] who extracted four scales from a matrix of qualitative data (20 amino acids × 134 descriptors). We prefer our scales z_1, z_2 and z_3 because they are based on quantitative (continuous) properties, see Table 3. Cramer[21] has derived scales (BCDEF) in a similar way for common organic compounds.

2. A statistical projection method, called PLS,[22] which allows the modelling of a $(n\times1)$ vector y (here peptide activity) as a combination of K $(n\times1)$ vectors of structure descriptors x_k $(k=1,K)$. These models apply also in the case when the number of descriptors (K) is larger than the number of compounds (n).

Mathematical modelling

For a set of n peptides, the structural variation is described by z_1, z_2 and z_3 in each varied position. Then the matrix X is modelled as $X = 1^*\bar{x} + T^*P + E$ where T is the low-dimensional $(n\times A)$ score matrix, P the corresponding $(A\times K)$ loading (or weight) matrix and E the residuals. Simultaneously y is modelled as $y = 1^*\bar{y} + T^*B' + f$ where T is the same matrix as above, b is a (1^*A) coefficient row vector ($'$ designate the transpose) and f is a residual vector. This PLS modelling is similar to principal components regression (PCR),[5] where X first is subjected to a principal components analysis and then y is modelled as a

*To whom correspondence should be addressed.

Table 1. Descriptor scales z_1, z_2 and z_3 for amino acids. The first three score vectors of a principal components analysis of a table with 20 properties for the 20 coded amino acids. z_1 is mainly related to hydrophilicity, z_2 contains additional information from the size and hydrophobicity/hydrophilicity scales while z_3 contains information from pK_{COOH}, p*I* and ^1H NMR variables.

Amino acid[a]		z_1	z_2	z_3
Ala	A	−0.24	−1.74	−0.39
Val	V	−2.03	−1.03	−2.33
Leu	L	−2.90	−0.22	−1.01
Ile	I	−3.22	−0.59	−1.79
Pro	P	−1.07	−0.82	0.83
Phe	F	−3.66	0.34	0.83
Trp	W	−4.41	2.18	2.32
Met	M	−2.20	−0.41	−0.38
Lys	K	2.76	3.25	−1.57
Arg	R	2.85	4.59	−2.00
His	H	2.00	0.61	1.63
Gly	G	2.41	−4.08	−2.47
Ser	S	1.78	−1.80	−0.45
Thr	T	0.99	−0.80	−1.38
Cys	C	0.90	−2.17	2.30
Tyr	Y	−2.25	1.87	0.12
Asn	N	2.57	0.19	2.00
Gln	Q	1.59	1.02	0.08
Asp	D	2.10	−1.19	3.27
Glu	E	2.10	0.60	0.39

[a]Symbols and abbreviations are in accordance with the recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature.

combination of the first major score vectors. In PLS, however, the projection and the "regression" steps are made simultaneously in a more efficient way.[22] The method is easily extended to model multivariate data *Y*, *i.e.*, if the activity is measured as L variables giving a $n \times L$ activity matrix Y instead of the $n \times 1$ activity vector y.

For new peptides the activity y can be predicted according to the scheme given in ref. 22*c*.

The y-model is similar to an ordinary regression model and has an analogous interpretation. The variance of the residuals f in relation to the variance of y informs about how much the model "explains" of the data. The latent variables t_a (columns in T; $a=1,2,\ldots,A$) are combined with the coefficients b_a (vector elements in b; $a=1,2,\ldots,A$). Since the t-values are approximately combinations of the original x-variables with the coefficients p (elements in P), the influence of the k:th x-variable is simply $\sum_a (p_{ak} * b_a)$.

Table 3. Variables used to characterize the amino acids.

Var. No.	Ref. No.	Property
1	–	Molecular weight
2	7	pK_{COOH} (COOH on C_α)
3	7	pK_{NH2} (NH_2 on C_α)
4	8	p*I*, pH at the isoelectric point
5	9	Substituent van der Waals volume
6	10	^1H NMR for C_α-H (cation)
7	10	^1H NMR for C_α-H (dipolar)
8	10	^1H NMR for C_α-H (anion)
9	11,12	^{13}C NMR for C=O
10	11, 12	^{13}C NMR for C_α-H
11	11, 12	^{13}C NMR for C=O in tetrapeptide
12	11, 12	^{13}C NMR for C_α-H in tetrapeptide
13	13	R_f for 1-*N*-(4-nitrobenzofurazono)-amino acids in ethyl acetate/pyridine/water
14	14	Slope of plot $1/(R_f-1)$ versus mole % H_2O in paper chromatography
15	15	d*G* of transfer from organic solvent to water
16	16	Hydration potential or free energy of transfer from vapor phase to water
17	17	R_f, salt chromatography
18	18	Log *P*, partition coefficient for amino acids in octanol/water
19	19	Log *D*, partition coefficient at pH 7.1 for acetylamide derivatives of amino acids in octanol water
20	20	$dG=RT\ln f$; f=fraction buried/accessible amino acids in 22 proteins

Table 2. Measurements from which the z_1-z_3 scales are calculated.

	Variables (see Table 3)																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
ALA	89	2.35	9.87	6.11	13.7	4.20	4.11	3.32	176.5	51.3	171.8	48.0	5.1	7.0	0.87	1.94	0.89	−2-89	−1.52	0.3
VAL	117	2.29	9.72	5.96	34.1	3.98	3.60	3.05	175.0	61.3	170.4	57.0	8.5	5.6	1.87	1.99	0.85	−2.08	−0.61	0.6
LEU	131	2.36	9.60	5.98	44.4	4.07	3.72	3.27	176.3	54.4	171.7	50.5	10.0	4.9	2.17	2.28	0.73	−1.61	−0.13	0.5
ILE	131	2.32	9.76	5.94	44.4	4.05	3.66	3.10	174.9	60.6	170.6	56.2	9.3	4.9	3.15	2.15	0.76	−1.72	−0.03	0.7
PRO	115	1.99	10.60	6.30	30.7	4.45	4.11	3.46	175.0	61.3	170.4	59.1	4.9	6.6	2.77	–	0.82	−2.50	−1.34	−0.3
PHE	165	2.58	9.24	5.48	56.1	4.38	3.98	3.51	174.7	57.0	170.9	53.6	9.6	5.0	2.87	−0.76	0.52	−1.63	−0.04	0.5
TRP	204	2.38	9.39	5.89	74.8	4.43	4.05	3.59	175.2	56.1	171.2	53.1	9.2	5.3	3.77	−5.88	0.20	−1.75	0.42	0.3
MET	149	2.28	9.21	5.74	45.0	4.24	3.85	3.23	175.0	54.9	170.9	51.5	8.7	5.3	1.67	−1.48	0.76	−1.84	0.60	0.4
LYS	146	2.20	8.90	9.59	61.5	4.08	3.46	3.24	175.1	55.0	171.2	51.8	1.3	10.1	1.64	−9.52	0.97	−4.44	−2.82	−1.8
ARG	174	2.18	9.09	11.15	77.3	4.13	3.21	3.19	174.9	54.8	171.0	51.7	2.0	9.1	0.85	−19.92	0.88	−4.20	−2.84	−1.4
HIS	155	1.78	8.97	7.47	45.1	4.45	3.98	3.51	174.6	55.0	170.0	51.8	1.6	8.4	0.87	−10.27	0.83	−4.15	−1.70	−0.1
GLY	75	2.34	9.60	5.97	3.5	3.94	3.55	3.22	173.2	42.2	168.7	42.1	4.1	7.9	0.10	2.39	0.92	−3.25	−1.83	0.3
SER	105	2.11	9.15	5.68	18.3	4.22	3.84	3.35	172.8	57.1	169.7	54.9	3.1	7.5	0.07	−5.06	0.96	−3.30	−1.87	−0.1
THR	119	2.15	9.12	5.64	28.5	4.42	3.58	3.10	173.7	61.2	169.9	58.0	3.5	6.6	0.07	−4.88	0.92	−2.91	−1.57	−0.2
CYS	121	1.71	8.33	5.07	25.0	4.31	3.98	3.56	173.5	57.0	169.7	51.6	–	11.5	1.52	−1.24	0.85	−2.49	−0.29	0.9
TYR	181	2.20	9.11	5.66	59.1	3.32	3.93	3.44	174.7	57.0	171.0	53.9	8.0	5.7	2.67	−6.11	0.49	−2.42	−0.87	−0.4
ASN	132	2.02	8.80	5.41	32.7	4.38	4.00	3.59	175.7	52.6	170.9	49.3	0.6	10.0	0.09	−9.68	0.89	−3.41	−2.41	−0.5
GLN	146	2.17	9.13	5.65	42.7	4.18	3.77	3.27	175.9	55.3	170.9	51.7	1.4	8.6	0.00	−9.38	0.82	−3.15	−2.05	−0.7
ASP	133	1.88	9.60	2.77	30.0	4.39	4.09	3.56	175.2	52.6	170.3	49.2	0.7	13.0	0.66	−10.95	0.87	−0.43	−2.60	−0.6
GLU	147	2.19	9.67	3.22	40.2	4.16	3.81	3.23	175.3	55.4	170.9	51.4	1.8	12.5	0.67	−10.20	0.84	−4.19	−2.47	−0.7

The loadings p_k display directly this influence of the x-variables.

We have earlier used the PCR and PLS approaches to model the biological activity of beta-adrenergic agents,[23] anesthetics[24] and halogenated hydrocarbons.[25]

Results

The example concerns a set of fifteen bradykinin potentiating pentapeptides investigated by Ufkes *et al.*[26] who modified the peptides in all five positions, most extensively in position 3. The biological activity is modelled as the logarithm of the relative activity index compared to peptide no. 1 (Val-Glu-Ser-Ser-Lys). This data set has previously been analysed by Schaper[27] who used an additive scheme, the so-called Fujita-Ban approach. That approach does not allow predictions outside the investigated amino acids, which severely limits the utility of the results.

We describe the change of amino acids in each of the five positions by z_1, z_2 and z_3. Thus we get an X-matrix with 15 columns (five positions $\times 3$ z-values) and fifteen rows (fifteen investigated peptides). It would also be possible to describe each amino acid position in the pentapeptides by the 20 properties and relate the resulting 15×100 matrix to the biological activity. However, the approach using the z-values as descriptors gives equivalent results and is more convenient. Prior to the data analysis the descriptor and biological activity data were scaled to unit variance.

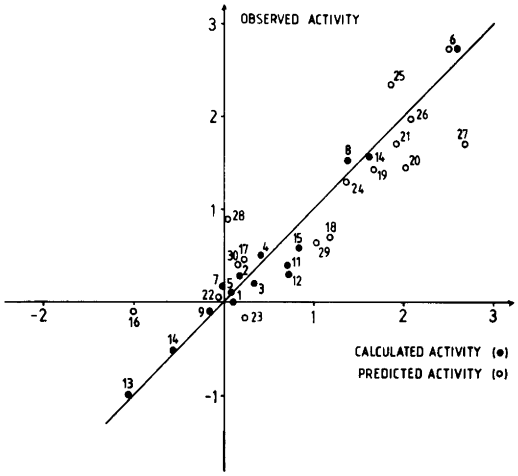

Fig. 1. Observed, calculated (no. 1–15) and predicted (No. 6, 16–30) activities for the pentapeptides in the example. The peptides are, in order 1–30: VESSK, VESAK, VEASK, VEAAK, VKAAK, VEWAK, VEANP, VEHAK, VAAAK, GEAAK, LEAAK, FEAAK, VEGGK, VEFAK, VELAK, AAAAA, AAYAA, AAWAA, VEWAA, VAWAK, VKWAA, VWAAK, VAAWK, EKWAP, VKWAP, RKWAP, VEWVK, PGFSP, FSPFR, RYLPT.

Table 4. Loadings (p_{ak}) for the descriptor variables in the PLS model.

Var. No.	Pos. No.	Descriptor	p_{1k}	p_{2k}
1	1	z_1	−0.07	−0.32
2	1	z_2	0.07	0.30
3	1	z_3	0.04	−0.05
4	2	z_1	0.02	0.21
5	2	z_2	0.01	0.04
6	2	z_3	0.02	0.29
7	3	z_1	−0.36	−0.06
8	3	z_2	0.48	0.18
9	3	z_3	0.45	0.12
10	4	z_1	−0.34	0.40
11	4	z_2	0.39	−0.45
12	4	z_3	0.39	−0.45
13	5	z_1	0.04	0.16
14	5	z_2	0.04	0.16
15	5	z_3	−0.04	−0.16

The PLS analysis gives two highly significant components, b_1=0.46 and b_2=0.31, explaining 97% (81% and 16% respectively) of the variance in the activity, see Fig. 1. The loadings p_k for the 15 variables used in the model are given in Table 4. From Table 4 it is seen that position number 3 has the main influence on the activity. The influence of the other positions 1, 2, 4 and 5 is not negligible, however.

Subsequently we found in the literature another set of one inactive and fifteen active bradykinin potentiating pentapeptides published by Ufkes *et al.*[28] These data refer to the same test system. On the basis of our previously developed model, predictions of the activity for these peptides were calculated, see Fig. 1. As see from Fig. 1 the differences between the predicted and measured activities are small. In addition the inactive peptide (Gly-Gly-Gly-Gly-Gly) was predicted to have a relative activity index of 0.001 compared to the reference peptide (No. 1), *i.e.* it is correctly predicted to be inactive. To our

knowledge this is the first peptide-QSAR which has demonstrated a significant *predictive* capability.

To further test the predictive power of the model, the most active peptide (No. 6, Val-Glu-Trp-Ala-Lys) was removed and a new model calculated from the remaining fourteen peptides. Then the activity of No. 6 was predicted by inserting its z-values into the model. As seen in Fig. 1 the prediction is precise. This is even more remarkable since Trp is not represented in position 3 in the "training" peptides; No. 6 is the only peptide having Trp in this position.

This model can also be used to construct peptides predicted to be highly active. On the basis of the loadings p_k of the model and the z-values in Table 1 we propose that the peptide Ile-Asn-Trp-Ala-Lys will be more potent than any peptide included in the set. With its z-values inserted in the model the predicted activity is 3.20. Hence, it is predicted to be more potent than the "optimal" pentapeptide <Glu-Lys-Trp-Ala-Pro (<Glu = pyroglutamic acid) proposed by Ufkes *et al.*[28] Other, more potent peptides can also be predicted from the model.

Discussion

In conclusion, we have shown that the same principles as used in the extrathermodynamic relationships in physical organic chemistry, can be used to model and predict the biological activity of peptides. By deriving three "scales" for amino acids, and using these scales in multivariate models, remarkably precise relationships have been found in a number of examples of which we here, for brevity, report only one. A full report will be given later.

The fact that bilinear models based on the description of the individual amino acids well describe the activity shows that, at least here, the conformational variation between the peptides either has no effect on the activity, or, in some way is accounted for by the model. If this is valid for many peptide sets, which we are currently investigating, this considerably simplifies the picture of peptide activity and also warrants some modifications of the interpretation of the influence of conformation on activity.

Acknowledgements. Grants from the Swedish Natural Science Council (NFR), the Swedish Council for Planning and Coordination of Research (FRN) and the National Swedish Board for Technical Development (STU) are gratefully acknowledged.

References

1. Zeelen, F. J. *Chemtech* (1983) 419.
2. Hammett, L. P. *Physical Organic Chemistry* 2nd. Ed., McGraw-Hill, New York 1970.
3. *a.* Wold, S. and Sjöström, M. In Chapman, N. B. and Shorter, J., Eds., *Correlation Analysis in Chemistry,* Plenum Publ. Co., 1978, p. 1.
   b. Sjöström, M. and Wold, S. *Acta Chem. Scand. B 35* (1981) 537.
4. Sjöström, M. and Wold, S. *J. Mol. Evol.* Accepted for publication.
5. Mardia, K. V., Kent, J. T. and Bibby, J. M. *Multivariate Analysis,* Academic. Press, London 1982.
6. Sneath, P. H. A. *J. Theoret. Biol. 12* (1966) 157.
7. *The Merck Index,* the ninth edition, 1977.
8. *Handbook of Biochemistry,* CRC, 1968.
9. Seydel, J. K. and Schaper, K. J. *Chemische Struktur und biologische Aktivität von Wirkstoffen,* Verlag Chemie, Weinheim 1979.
10. Roberts, G. C. K. and Jardetzky, O. *Adv. Protein Chem. 24* (1970) 447.
11. Horsley, W., Sternlicht, H. and Cohen, J. S. *J. Am. Chem. Soc. 92* (1970) 680.
12. Rosenthal, S. N. and Fendler, J. H. *Adv. Phys. Org. Chem. 13* (1976) 279.
13. Aboderin, A. A. *Int. J. Biochem. 2* (1971) 537.
14. Woese, C. R., Drugre, D. H. and Saxinger, S. A. *Proc. Natl. Acad. Sci. USA 55* (1966) 966.
15. Jones, D. D. *J. Theor. Biol. 50 (1975) 167.*
16. *Wolfenden, R., Andersson, L., Cullis, P. M. and Southgate, C. C. B. Biochemistry 20* (1981) 849.
17. Weber, A. L. and Lacey, J. C., Jr. *J. Mol. Evol. 11* (1978) 199.
18. Pliška, V., Schmidt, M. and Fauchère, J.-L. *J. Chromatogr. 216* (1981) 79.
19. Fauchère, J.-L. and Pliška, V. *Eur. J. Med. Chem. 18* (1983) 369.
20. Janin, J. *Nature 277* (1979) 491.
21. *a.* Cramer III, R. D. *J. Am. Chem. Soc. 102* (1980) 1837.
    b. Cramer III, R. D. *J. Am. Chem. Soc. 102* (1980) 1849.
22. *a.* Wold, H. In Jöreskog, K. G. and Wold, H., Eds., *Systems Under Indirect Observation,* North-Holland Publ. Co., Amsterdam 1982, part II, p. 1.;
    b. Wold, S., Ruhe, A., Wold, H. and Dunn III, W. J. *SIAM J. Sci. Stat. Comput. 5* (1984) 735;
    c. Wold, S., Albano, C., Dunn III, W. J., Esbensen, K., Hellberg, S., Johansson, E. and Sjöström, M. In Martens, H. and Russwurm, H. Jr. Eds. *Food Research and Data Analysis,* Proc. IUFOST

Conf., Applied Science Publ., London 1983, pp. 147–188.
23. Dunn III, W. J., Wold, S. and Martin, Y. C. *J. Med. Chem. 21* (1978) 922.
24. Hellberg, S., Wold, S., Dunn III, W. J., Gasteiger, J. and Hutchings, M. G. *Quant. Struct.-Act. Relat. 4* (1985) 1.
25. Dunn III, W. J., Wold, S., Edlund, U., Hellberg, S. and Gasteiger, J. *Quant. Struct.-Act. Relat. 3* (1984) 131.
26. Ufkes, J. G. R., Visser, B. J., Heuver, G. and van der Meer, C. *Eur. J. Pharmacol. 50* (1978) 119.
27. Schaper, K. J. *Eur. J. Med. Chem. 15* (1980) 449.
28. Ufkes, J. G. R., Visser, B. J., Heuver, G., Wynne, H. J. and van der Meer, C. *Eur. J. Pharmacol. 79* (1982) 155.

Received April 29, 1985.