Journal of Peptide Science
J. Peptide Sci. **8**: 302–306 (2002)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.403

Simple Parameterization of Non-proteinogenic Amino Acids for QSAR of Antibacterial Peptides

TORE LEJON,* JOHN S. SVENDSEN and BENGT E. HAUG

Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway

Received 22 March 2002
Accepted 13 April 2002

Abstract: The antibacterial activity of bovine lactoferricin-(17–31)-pentadecapeptide against *Escherichia coli* and *Staphylococcus aureus* is sensitive to substitution of the Trp residues, and synthetic peptides with phenylalanine and any of eight non-proteinogenic aromatic amino acids greatly affected antibiotic activity. Using simple size-related descriptors for the new amino acids it is possible to develop quantitative structure–activity relationships (QSARs) that can be used as tools in the search for more active peptides.

Keywords: pentadecapeptides; amino acids; antibacterial activity QSAR

INTRODUCTION

In the development of structure–activity relationships of peptides, numerous methods have been used to describe the variation in activity as a function of structure. Based on predictive ability, the best models are probably those derived from (measured) parameters describing macroscopic properties of the peptides, e.g. hydrophobicity, charge and α-helicity, since these properties reflect the actual behaviour of the peptide [1]. The drawback of such models is that new improved peptides will be described by the same input descriptors used in the modelling. This will again lead to difficulties in translation of these properties into amino acid sequences, and thereby construction of novel peptides.

One way of overcoming this problem is to use amino acid descriptors [2,3] rather than peptide descriptors, a valuable approach in our studies of antibiotic peptides [4]. By not measuring the properties of the peptide but instead using variables based on measured or theoretical properties of the individual amino acid residues, information about peptide properties are bound to disappear to some extent, but the advantages will fully compensate for this. First of all, new peptides can be 'tested' in the model before they are synthesized, saving synthesis and work-up time. Secondly, descriptors do not have to be measured for each new compound; and — maybe the most important aspect of all — the predicted peptides will be described as having certain amino acid properties, as described by their z_1, z_2 and z_3 values, and not as having peptide properties. However, the use of this method is also limited, as useful descriptors are only available for a small number of amino acids, with the coded amino acids being by far the most studied.

There are examples of non-coded amino acids for which descriptors have been developed that are comparable to those for the coded amino acids [5], but for most researchers such studies are not useful, since amino acids to be incorporated into new peptides are often not parameterized. Consequently, descriptors for the novel amino acids to be used in the peptides have to be developed. In most cases this would cost too much in relation to what can be gained; further, difficulties in the use of descriptors for novel amino acids arise from the

* Correspondence to: Tore Lejon, Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway;
e-mail: tore.lejon@chem.uit.no
Contract/grant sponsor: Alpharma AS (Norway).
Contract/grant sponsor: Norwegian Research Council.

fact that descriptor data are measured in different laboratories. If the modified amino acids are merely present in the same positions in the molecules, only descriptors for these are necessary and the remaining (natural) amino acids can be assumed constant and not affected by the substitutions.

Bovine lactoferricin-(17–31)-pentadecapeptide, FKCRRWQWRMKKLGA (LFB) has previously been shown to exhibit high antibiotic activity against both *E. coli* and *S. aureus* [6]. It is clear from structure–activity studies that the Trp residues at positions 6 and 8 are crucial for activity [1], and incorporation of novel aromatic amino acids at these positions has been used successfully to further increase the activity [7–9]. In addition, it is known from studies in which descriptors for amino acids have been developed, that the largest influence on the principal properties comes from (i) lipophilicity and (ii) steric parameters [2]. In the present work we examine the possibility of modelling structure–activity relationships based on simple parameters, e.g. size of the side chains, for the non-coded amino acids employed.

METHODS

Calculations of molecular volumes were performed using MacroModel Version 6.5 [10] (Table 1) as described by Haug *et al.* [8]. The program package

Table 1 Data for LFB Analogues Used in the Study

Peptide	Amino acid at position 6 and 8[a]		Position 6[b]			Position 8[b]			MIC (μM)	
			V (Å^3)	*L* (Å)	*W* (Å)	*V* (Å^3)	*L* (Å)	*W* (Å)	*E. coli*	*S. aureus*
LFB	Trp	Trp	129.4	5.4	4.7	129.4	5.4	4.7	24	48
1	Phe	Trp	100.0	4.3	2.4	129.4	5.4	4.7	25	148
2	Trp	Phe	129.4	5.4	4.7	100.0	4.3	2.4	12	148
3	Phe	Phe	100.0	4.3	2.4	100.0	4.3	2.4	76	151
4	Bal	Trp	135.0	5.4	4.8	129.4	5.4	4.7	7.2	36
5	Trp	Bal	129.4	5.4	4.7	135.0	5.4	4.8	7.2	17
6	Bal	Bal	135.0	5.4	4.8	135.0	5.4	4.8	9.5	12
7	1-Nal	Trp	142.1	5.2	5.0	129.4	5.4	4.7	24	72
8	Trp	1-Nal	129.4	5.4	4.7	142.1	5.2	5.0	9.6	48
9	1-Nal	1-Nal	142.1	5.2	5.0	142.1	5.2	5.0	4.8	24
10	2-Nal	Trp	142.6	6.5	5.0	129.4	5.4	4.7	9.6	36
11	Trp	2-Nal	129.4	5.4	4.7	142.6	6.5	5.0	4.8	24
12	2-Nal	2-Nal	142.6	6.5	5.0	142.6	6.5	5.0	4.8	9.6
13	Bip	Trp	172.3	8.7	2.4	129.4	5.4	4.7	12	7.1
14	Trp	Bip	129.4	5.4	4.7	172.3	8.7	2.4	7.1	7.1
15	Bip	Bip	172.3	8.7	2.4	172.3	8.7	2.4	5.8	1.4
16	Dip	Trp	172.4	4.3	7.1	129.4	5.4	4.7	3.6	17
17	Trp	Dip	129.4	5.4	4.7	172.4	4.3	7.1	4.8	17
18	Dip	Dip	172.4	4.3	7.1	172.4	4.3	7.1	3.5	7
19	Ath	Trp	185.4	5.1	7.3	129.4	5.4	4.7	5.9	16
20	Trp	Ath	129.4	5.4	4.7	185.4	5.1	7.3	4.7	7.1
21	Ath	Ath	185.4	5.1	7.3	185.4	5.1	7.3	4.6	3.4
22	Tbt	Trp	325.1	7.6	8.9	129.4	5.4	4.7	5.6	4.5
23	Trp	Tbt	129.4	5.4	4.7	325.1	7.6	8.9	5.6	2.2
24	Tbt	Tbt	325.1	7.6	8.9	325.1	7.6	8.9	8.3	3.1
25	Tpc	Trp	365.0	7.7	12.4	129.4	5.4	4.7	3.2	4.3
26	Trp	Tpc	129.4	5.4	4.7	365.0	7.7	12.4	6.4	3.2
27	Tpc	Tpc	365.0	7.7	12.4	365.0	7.7	12.4	9.6	2.9

[a] Abbreviations; Ath, β-(anthracen-9-yl)alanine; Bal, β-(benzothien-3-yl)-alanine; Bip, β-(4,4′-biphenyl)alanine; Dip, β-diphenylalanine; LFB, lactoferricin B residues 17–31; MIC, minimal inhibitory concentration; 1-Nal, β-(naphth-1-yl)alanine; 2-Nal, β-(naphth-2-yl)alanine; Tbt, β-(2,5,7-tri-*tert*-butyl-indol-3-yl)alanine; Tpc, β-[2-(2,2,5,7,8-pentamethyl-chroman-6-sulfonyl)-indol-3-yl]alanine, i.e. Trp(2-Pmc).
[b] Size parameters given as *V*, volume; *L*, length; *W*, width Ref. [8].

Simca-P 9.0 from Umetrics (Umeå, Sweden) was used for all calculations. Size variables used as descriptors for the amino acids were centred and scaled to unit variance. The logarithms of the MIC values were used as dependent variables.

RESULTS AND DISCUSSION

The nine aromatic amino acids were incorporated into positions 6 and/or 8 of LFB as replacements of tryptophan and the resulting 27 peptides were tested for their antibiotic activity against *E. coli* and *S. aureus* (Table 1) [7–9]. Also, since selectivity is often of interest, the ratio between responses for *E. coli* and *S. aureus* were added as a measure for selectivity. As dependent variables, x-variables, the size of the amino acid side chains were used. Originally four size parameters were used, length, width, area and volume, but since all the amino acids, with the exception of Tpc, are rigid flat molecules area and volume will always be correlated. For this reason, the area of the amino acids was excluded, as variables that are correlated will enforce each other and will influence the model more than their importance would account for. The six parameters used as x-variables were thus the length, width and volume of side chains of the amino acids in positions 6 and 8. The model calculated using this data set was not significant and there was no good correlation between the actual and the predicted responses for the data set. Analysis of the data revealed that it was mainly the peptides that were substituted in both positions that did not fit the model, and therefore a new model was calculated using only peptides in which either the 6- or the 8-position had been substituted, compared with LFB. This new model explained 74% of the variation in the y-variables, and the predictive power of the model was good with $Q^2(\text{cum}) = 0.48$, resulting in good correlation between predicted and observed activities against *S. aureus* (Figure 1) and predicted and observed selectivity (Figure 2), while the result for prediction of *E. coli* activities was poorer. It should be noted that a Q^2 value >0.5 is considered to reflect a high predictivity of the model. The 48% predictive power of our model is surprisingly good considering both the simple descriptors used and the fact that the MIC values obtained represent discrete titre steps rather than continuous variables.

In the calculations performed no sequence information was included in the variables describing the peptides, and only peptides in which tryptophan in the 6- and 8-positions had been replaced by one aromatic amino acid were used for modelling (in addition to LFB). Problems dealing with sequence information in peptides in connection with multivariate analysis have been briefly discussed elsewhere [11], so we wanted to investigate if further

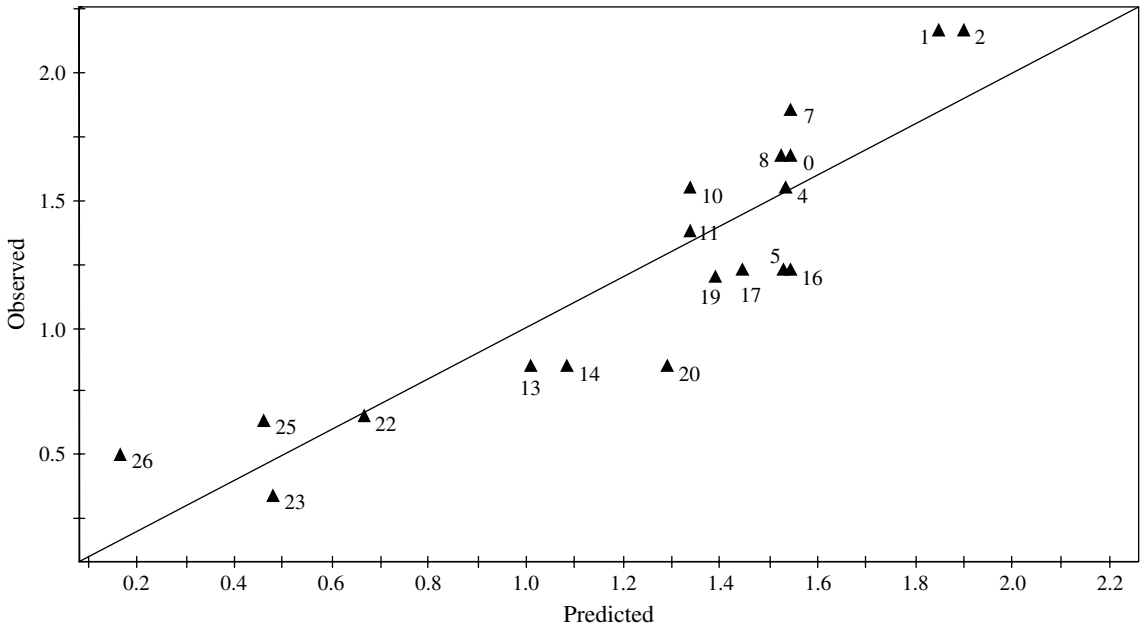

Figure 1 Predicted vs observed log MIC *S. aureus*.

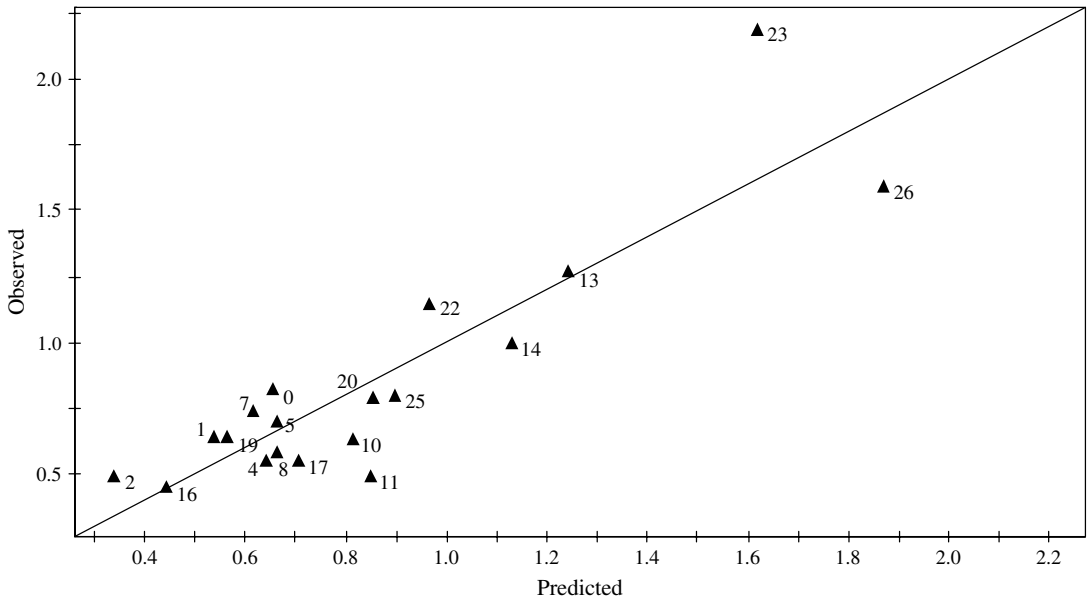

Figure 2 Predicted vs observed ratio log MIC *E. coli*/log MIC *S. aureus*.

Table 2 Simplified Descriptor–Response Matrix

Xxx[a]	Size parameters[b]			MIC [Xxx^6]-LFB[a]		MIC [Xxx^8]-LFB[a]		MIC [Xxx^{6,8}]-LFB[a]	
	V (Å^3)	*L* (Å)	*W* (Å)	*E. coli*	*S. aureus*	*E. coli*	*S. aureus*	*E. coli*	*S. aureus*
1 Phe	100.0	4.3	2.4	25	148	12	148	76	151
2 Bal	135.0	5.4	4.8	7.2	36	7.2	17	9.5	12
3 1-Nal	142.1	5.2	5.0	24	72	9.6	48	4.8	24
4 2-Nal	142.6	6.5	5.0	9.6	36	4.8	24	4.8	9.6
5 Bip	172.3	8.7	2.4	12	7.1	7.1	7.1	5.8	1.4
6 Dip	172.4	4.3	7.1	3.6	17	4.8	17	3.5	7
7 Ath	185.4	5.1	7.3	5.9	16	4.7	7.1	4.6	3.4
8 Tbt	325.1	7.6	8.9	5.6	4.5	5.6	2.2	8.3	3.1
9 Tpc	365.0	7.7	12.4	3.2	4.3	6.4	3.2	9.6	2.9

[a] Xxx denotes any of the aromatic amino acids used as replacements for tryptophan.
[b] Size parameters given as *V*, volume; *L*, length; *W*, width Ref. [8].

simplification of the data was possible, e.g. by only including information about the aromatic amino acids used to replace the tryptophan residue(s). In this model the descriptors for the remaining tryptophans were thus excluded (Table 2).

The model from these calculations explained 72% of the variation in the y-variables, but the predictive ability was low, as seen from the Q^2(cum) value of only 23%. The variables responsible for the low predictive ability of the model were the log MIC values against *E. coli* for [Xxx^8]-LFB and [Xxx^{6,8}]-LFB, respectively. As for the first data set, all the peptide responses with two substitutions made were excluded from further calculations. A new model was thus obtained that was able to explain 77% of the variation in the y-variables. This model had a high predictive ability (Q^2(cum) = 0.49), and the predicted variables were in good agreement with the observed, except for *E. coli* activities (examples are given in Table 3).

CONCLUSIONS

Size parameters can be used as variables in models for structure–activity relationships, using PLS, in

Table 3 Predicted vs Observed log MIC Values for [Xxx^6]-LFB

Xxx	*E. coli*		*S. aureus*		Ratio	
	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.
Phe	1.398	1.180	2.170	1.916	0.644	0.572
Bal	0.857	1.039	1.556	1.598	0.551	0.664
1-Nal	1.380	1.018	1.857	1.601	0.743	0.639
2-Nal	0.982	1.046	1.556	1.398	0.631	0.826
Bip	1.079	1.196	0.851	1.066	1.268	1.245
Dip	0.556	0.853	1.230	1.591	0.452	0.467
Ath	0.771	0.848	1.204	1.431	0.640	0.586
Tbt	0.748	0.661	0.653	0.666	1.145	0.973
Tpc	0.505	0.437	0.633	0.445	0.797	0.901

studies of antibiotic peptides based on bovine lactoferricin. The limitation is that only LFB peptides in which one Trp residue has been replaced by another aromatic amino acid can be modelled.

Acknowledgement

Alpharma AS (Norway) is gratefully acknowledged for financial support to John S. Svendsen and the Norwegian Research Council for support to Bengt E. Haug.

REFERENCES

1. Strøm MB, Rekdal Ø, Svendsen JS. Antibacterial activity of 15-residues lactoferricin derivatives. *J. Peptide Res.* 2000; **56**: 265–274.
2. Hellberg S, Sjöström M, Wold S. The prediction of bradykinin potentiating potency of pentapeptides. An example of a peptide quantitative structure–activity relationship. *Acta Chem. Scand. B* 1986; **40**: 135–140.
3. Hellberg S, Sjöström M, Skagerberg B, Wold S. Peptide quantitative structure–activity relationships, a multivariate approach. *J. Med. Chem.* 1987; **30**: 1126–1135.
4. Lejon T, Strøm MB, Svendsen JS. Antibiotic activity of pentadecapeptides modelled from amino acid descriptors. *J. Peptide Sci.* 2001; **7**: 74–81.
5. Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J. Med. Chem.* 1998; **41**: 2481–2491.
6. Rekdal Ø, Andersen J, Vorland LH, Svendsen JS. Construction and synthesis of lactoferricin derivatives with enhanced antibacterial activity. *J. Peptide Sci.* 1999; **5**: 32–45.
7. Haug BE, Svendsen JS. The role of tryptophan in the antibacterial activity of a 15-residue bovine lactoferricin peptide. *J. Peptide Sci.* 2001; **7**: 190–196.
8. Haug BE, Skar ML, Svendsen JS. Bulky aromatic amino acids increase the antibacterial activity of 15-residue bovine lactoferricin derivatives. *J. Peptide Sci.* 2001; **7**: 425–432.
9. Haug BE, Andersen J, Rekdal Ø, Svendsen JS. Synthesis of sulphonylated tryptophan — antibacterial activity of bovine lactoferricin peptides containing Trp(2-Pmc). *J. Peptide Sci.* 2002; **8**: 307–313.
10. Mohamdi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still WC. MacroModel — an integrated software system for modelling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* 1990; **11**: 440–467.
11. Lejon T, Strøm MB, Svendsen JS. Is information about peptide sequence necessary in multivariate analysis? *Chemom. Intell. Lab. Syst.* 2001; **57**: 93–95.

