Journal of Peptide Science
J. Peptide Sci. **10**: 329–335 (2004)
Published online 21 January 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.553

JOURNAL OF PEPTIDE SCIENCE

Prediction of Antibiotic Activity and Synthesis of New Pentadecapeptides based on Lactoferricins

TORE LEJON,[a]* TRINE STIBERG,[a] MORTEN B. STRØM[b] and JOHN S. SVENDSEN[a]

[a] Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway
[b] Department of Medicinal Chemistry, Institute of Pharmacy, Faculty of Medicine, University of Tromsø, N-9037 Tromso, Norway

Received 18 September 2003
Accepted 25 October 2003

Abstract: The antibacterial activity against *Escherichia coli* and *Staphylococcus aureus* has been studied for a number of modified pentadecapeptides based on lactoferricins of different origin. The peptides were classified by multivariate methods and quantitative structure–activity relationships (QSAR) were developed using theoretically derived variables for the amino acids. For the modified peptides based on bovine lactoferricin (LFB) a model was calculated and used for prediction of new peptides that were then tested for antibacterial activity in order to improve peptide activity and to check the validity of the model. Models were also calculated including lactoferricins of different origin. Theories of the mechanism of action of the peptides are briefly discussed. Copyright © 2004 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: lactoferrin; lactoferricin; pentadecapeptides; antibacterial activity; QSAR; predictions

INTRODUCTION

Lactoferricins, found in a variety of mammals, exhibit antibiotic activity [1,2] and have served as a starting point in the search for antibiotics with a wider range of action, specifically against multi-resistant bacteria [3] and as drugs against cancer [4]. Since the native lactoferrin protein is a large molecule with 689 amino acid residues, early studies were directed towards finding the smallest peptide necessary for this activity. Our laboratories have previously shown that pentadecamers, based on murine lactoferricin, containing amino acid residues 17–31 exhibit good antibacterial activity [3], and that in order to achieve high activity it is necessary to incorporate tryptophan [5], or other large aromatic amino acid residues [6–8], into positions 6 and 8. These results were based on screening of antibiotic activity by varying the substitution pattern, e.g. by alanine scans. Later, when more data were available, multivariate methods were employed in order to model quantitative structure–activity relationships (QSARs), based on the assumption that individual molecules can be described by physico-chemical variables, and that the model obtained can be related to the (biological) activity of the molecule.

The first models that were calculated made use of macroscopic peptide descriptors, e.g. α-helicity, lipophilicity and HPLC retention times, as structural descriptors, and the calculated models revealed a good correlation between observed and predicted activities for the studied peptides. Even though it is feasible to create a model from measured or calculated peptide properties, as shown by Strøm *et al.* [9] that gives a good correlation between peptide properties and antibacterial activities, it does not imply that it is straightforward to use this model for designing peptides with enhanced antibacterial activities. This is because peptide descriptors describe macroscopic properties of 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.

peptide, e.g. α-helicity or HPLC retention time, but it is difficult to predict the effect of amino acid substitution on these properties. The design of new peptides would thus rely heavily on educated guesses of what the result of a substitution would be, resulting in new peptides having to be synthesized before obtaining an answer. This approach is therefore both expensive (input of chemicals and solvents) and time-consuming, since in addition to synthesis and purification, testing is necessary for each compound.

Instead the problem to be solved is how to describe molecules in such a way that the information can be used directly to design new compounds. As a consequence of this the study was extended and the theoretically derived amino acid descriptors, developed by Hellberg *et al.* [10,11], were used to investigate whether these descriptors would model peptide properties. By using this approach it was possible to model peptide properties, e.g. α-helicity, lipophilicity and HPLC retention times, in addition to theoretically derived properties such as different measures for charge localization and Eisenberg α-helix propensities and, most importantly, antibacterial activity [12].

An advantage of being able to use amino acid descriptors is that the properties modelled for new peptides can be directly translated into new peptides. The second advantage is that new peptides can be tested in the model before they are synthesized, minimizing the effort in synthesis, purification and analysis. Thirdly, different multivariate methods have proven to be of much help, especially since they are iterative in the sense that all new information can be used immediately in order to derive new and better models. The models that were developed proved to be very good at describing macroscopic properties but the ultimate test for a model must be its predictive power. The present study includes a number of lactoferricin peptides found in a variety of species and the antibiotic response is predicted for pentadecapeptides of bovine origin.

MATERIALS AND METHODS

Multivariate Analysis

The Simca-P 10.0 program package from Umetrics, Umeå, Sweden was used for all calculations. The theoretically derived *z*-scales were centred and either scaled to unit variance or used without scaling. The logarithm of the MIC-values was used as the biological response and was centred and used in scaled or un-scaled mode. The peptides chosen for the analysis were 25 bovine lactoferricins (LFB) (including retro-LFB), 18 murine lactoferricins (LFM), two caprine lactoferricins (LFC), three human lactoferricins (LFH) and two porcine lactoferricins (LFP) (Table 1).

Peptide Synthesis

Peptides were synthesized, purified and analysed as previously reported and the purity of all peptides was found to be >98% in all cases [13].

All the peptides were synthesized on a 9050 *Plus* PepSynthesizer (Milligen) by Fmoc solid phase peptide synthesis, as described [13]. In brief, the carboxylic acid group was preactivated with Pfp-esters or *in situ* activated with the coupling reagent HBTU in DMF. In the case of coupling with Pfp-esters, 1.3 eq. of 1-HOBt was added to catalyse the reaction, whereas 2.4 eq. of DIPEA was added when coupling with HBTU (1 eq.). A four-fold excess of amino acids was employed during every coupling step. The acid labile protecting groups were deprotected during cleavage of the peptide from the solid support upon treatment with Reagent K [14] (82.5% TFA, 5% anisole, 2.5% EDT, 5% water and 5% phenol) for less than 3 h. Cysteine was irreversibly protected in all peptides with an Acm-group. The peptides were purified on a RP-HPLC C18-column (Delta-Pak™ C18, 100Å, 15 µm, 25 × 100 mm, Waters Corporation) with a mixture of water and acetonitrile (containing 0.1% TFA) as the mobile phase and UV-detection at 254 nm. All the peptides were analysed for impurities on an analytical RP-HPLC C18-column (Delta-Pak™ C18, 100Å, 5 µm, 3.9 × 150 mm, Waters Corporation) with an isocratic mixture of water and acetonitrile (containing 0.1% TFA) as the mobile phase. The integrity of the peptides was checked by positive ion electro spray ionization mass spectrometry on a VG Quattro quadrupole mass spectrometer, and the purity of the peptides was ensured to be above 98% before biological screening.

Peptide Testing

The minimal inhibitory concentration (MIC) against *E. coli* and *S. aureus* were tested for all peptides using earlier described antibacterial assays [15].

Table 1 Amino Acid Sequences and Antimicrobial Activity (μM) of the 15 Amino Acid Residue of Bovine (LFB residues 17–31; FKCRRWQWRMKKLGA), Mouse (LFM residues 16–30; EKCLRWQNEMRKVGG), Goat (LFC residues 17–31; SKCYQWQRRMRKLGA), Human (LFH residues 18–32; TKCFQWQRNMRKVRG) and Pig (LFP residues 17–31; SKCRQWQSKIRRTNP) Lactoferricins

Number	Name	*E. coli*	*S. aureus*	Number	Name	*E. coli*	*S. aureus*
1	LFB	24	48	25	LFM	500	500
2	LFB A1	35	100	26	LFM W8	500	500
3	LFB K1	29	49	27	LFM W8 Y13	500	500
4	LFB A3	13	51	28	LFM A1 W8	391	500
5	LFB W3	10	10	29	LFM A9 W8	500	500
6	LFB F4	10	100	30	LFM A1, 9 W8	134	500
7	LFB F4 K1	10	100	31	LFM R1 W8	37	374
8	LFB W4, 10	2	5	32	LFM R9 W8	100	500
9	LFB W3,14	4	4	33	LFM A1 R9 W8	31	257
10	LFB W14	8	10	34	LFM A9 R1 W8	31	257
11	LFB A14	12	77	35	LFM R1, 9 W8	10	37
12	LFB A2	40	100	36	LFM A1 W8 Y13	151	500
13	LFB A4	35	100	37	LFM A9 W8 Y13	500	500
14	LFB A5	61	100	38	LFM A1, 9 W8 Y13	88	500
15	LFB A6	100	100	39	LFM R1 W8 Y13	19	73
16	LFB A7	15	75	40	LFM R9 W8 Y13	63	500
17	LFB A8	100	100	41	LFM A1 R9 W8 Y13	25	40
18	LFB A9	28	100	42	LFM A9 R1 W8 Y13	12	25
19	LFB A10	70	100	43	LFM R1, 9 W8 Y13	10	12
20	LFB A11	35	100	44	LFC	252	500
21	LFB A12	25	100	45	LFC W8	174	500
22	LFB A13	25	100	46	LFH	500	500
23	LFB W3,7,14	10	10	47	LFH W8	74	500
24	LFB Retro	39	97	48	LFH W8 Y13	55	500
				49	LFP	500	500
				50	LFP W8	219	500
				51	LFB A4, 14, R3, W1, 10	>10	>10
				52	LFB W1, 3, 4, 10, 14	>10	4

RESULTS AND DISCUSSION

Prediction of Peptide Activity

An earlier paper showed that the biological response, in addition to other peptide properties, can be modelled from the *z*-values [12], and the same approach was used for predicting the activity of new peptides and to test if the model is valid outside the range studied. At this time only a limited number of LFB peptides was available for analysis and the first calculations were made on the peptides 1–11 (Table 1). As dependent variables only the antibiotic response of the peptides against *E. coli* and *S. aureus* was chosen. The calculated model used four components (third component not significant) to explain 93% of the variation in the *y*-variables using 84% of the variation in the *x*-variables. In order to evaluate the model, the predicted vs observed biological response can be plotted and it is clear that both log MIC *E. coli* ($r^2 = 0.957$) and log MIC *S. aureus* ($r^2 = 0.924$) are well described by the calculated model (Figure 1a,b). A closer inspection of the loadings for the model reveals that the variables having the largest influence on the model are z_1 for positions 1, 3, 4 and 14, z_2 for positions 10 and 14 and z_3 for position 4. From the papers by Hellberg *et al.* it is known that z_1 contains information about hydrophilicity, z_2 about bulk and z_3 about electronic properties [10,11]. A complication in the evaluation of the results is that substitutions in a certain position will not affect the biological responses equally, i.e. a substitution that improves the response for one activity might

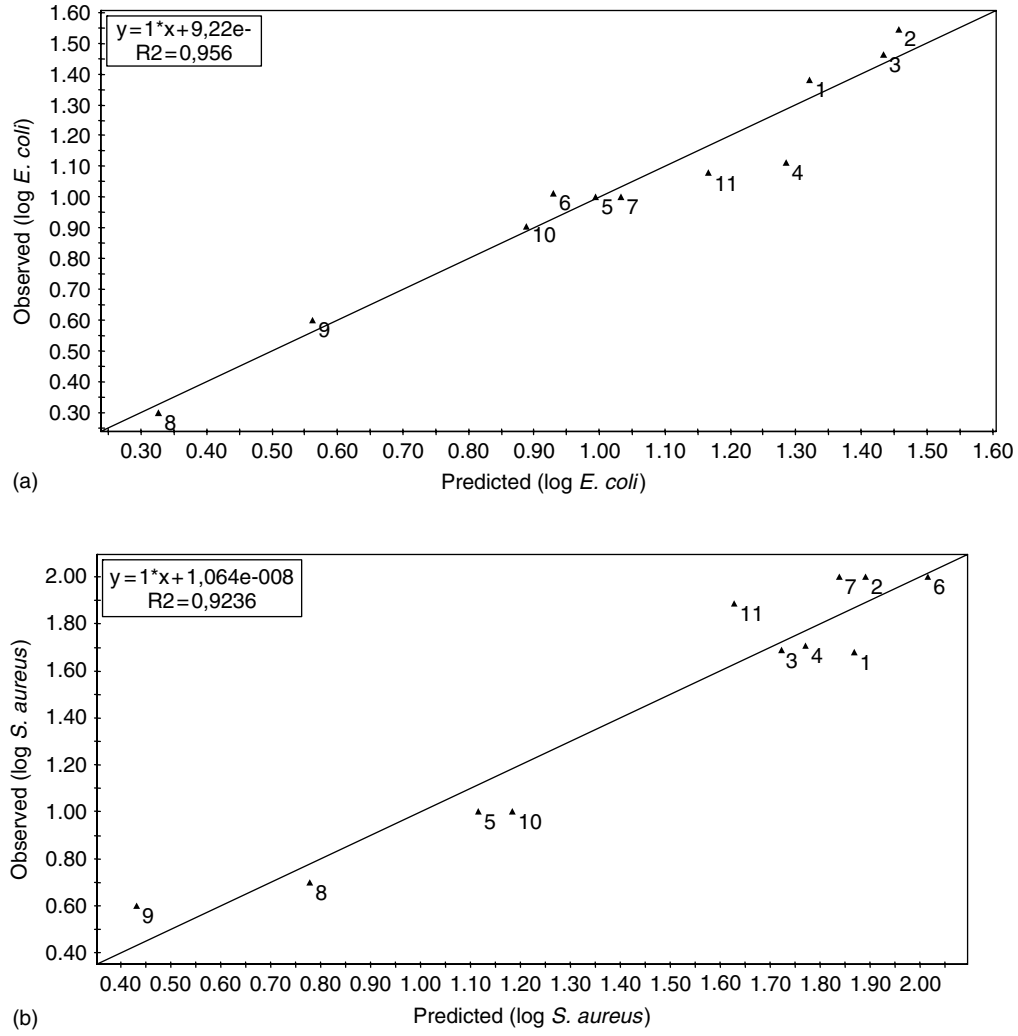

Figure 1 (a) Observed *vs* predicted log (MIC *E. coli*), (b) Observed *vs* predicted log (MIC *S. aureus*).

give an unwanted response for another. At this stage the strength of PLS is manifested in that this model can be used for predicting the activity of new peptides before synthesis and subsequent testing. The first new peptide that was tested in the model was LFB A4, 14, R3, W1, 10, i.e. a small lipophilic amino acid in positions 4 and 14, a large lipophilic amino acid in positions 1 and 10 and a large charged amino acid in position 3. This peptide was predicted to have a MIC value for *E. coli* of 3.1 μg/ml (log MIC = 0.49) and MIC value for *S. aureus* of 5.7 μg/ml (log MIC = 0.75). For both bacteria the biological testing showed MIC values of >10 μg/ml. A second peptide was also synthesized containing tryptophan in all five substituted positions (LFB W1, 3, 4, 10, 14) with predicted MIC values of 0.3 μg/ml (log MIC = −0.47) and 0.2 μg/ml (log MIC = −0.71) and measured values of >10 μg/ml for *E. coli* and 4 μg/ml for *S. aureus*, respectively. None of the responses were thus as good as predicted, but the activities for the second peptide were as good as the best results from the original data set. This is somewhat surprising since several of the responses in the original data set were outside the measured range (MIC > 100 μg/ml), being treated as though the responses were the same even if substitutions had been undertaken. The results from this analysis made it necessary to synthesize more peptides in order to investigate whether a model with more peptides included would lead to a model that predicted new peptide activities better. In this respect an alanine scan was performed, substituting one amino acid at a time with an alanine. Calculating a model including all of these

peptides gave a five-component model explaining 93% of the variation in biological response using 70% of the variation in X. As for the previous model, the correlation between predicted and observed responses was very good. For *E. coli* $r^2 = 0.966$ while $r^2 = 0.905$ for *S. aureus*. Using the same peptides as above for testing the model showed that the model still predicted lower MIC values than what was actually observed (Table 2). The result indicated that the model was not good enough for predictions, even if biological responses used in the calculations fitted the model well. For *S. aureus* the results predicted were closer to the observed ones. This can be explained as a result of still too few peptides being included in the calculations or that the variation in substitution pattern for the new peptides was much larger than what is reflected in the work-set. At this point a number of new peptides had been synthesized, and it was of interest to analyse these to see if they could be included in the model. Therefore a PCA was performed on the 50 peptides of the entire data set, including *z*-variables as well as biological responses.

Principal Component Analysis

If data were centred without any scaling of the variables, the calculations resulted in a model with four components (second component not significant according to cross validation) explaining more than 70% of the variation in the data. From the plot of the scores for *t* [1] *vs t* [2] (Figure 2) it was clear that the peptides of different origin appeared as clusters with the individual peptides having similar scores. The exception to this was the peptides derived from LFB that were not as close together as the others and three of the peptides were even slightly outside the confidence limit of 95%. A closer inspection revealed that two of these peptides were substituted with tryptophan in position 14, the third being retro-LFB, which has a completely different sequence to the native peptide. It should also be noted that LFB W14 is closer to these two peptides than to the rest of the LFB peptides. In Figure 3 the loadings for the first two components are plotted and it is evident that positions 1 and 4 were important for the first component while position 14 was responsible for the variation in the second dimension, which explains

Table 2 Activity Against *E. coli* and *S. aureus* in μg/ml

	Model 1[a]		Model 2[b]		Model 3[c]		Activity[d]	
	E. coli	*S. aureus*	*E. coli*	*S. aureus*	*E. coli*	*S. aureus*	*E. coli*	*S. aureus*
LFB A4, 14, R3, W1, 10	3.1	5.7	2.5	7.3	3.2	10.2	>10	>10
LFB W1, 3, 4, 10, 14	0.3	0.2	0.2	0.4	0.4	1.0	>10	4

[a] Original LFB work-set including peptides 1–11.
[b] Extended LFB work-set including peptides 1–23.
[c] Work-set consisting of extended LFB work-set and LFM peptides 25–43.
[d] Measured MIC value.

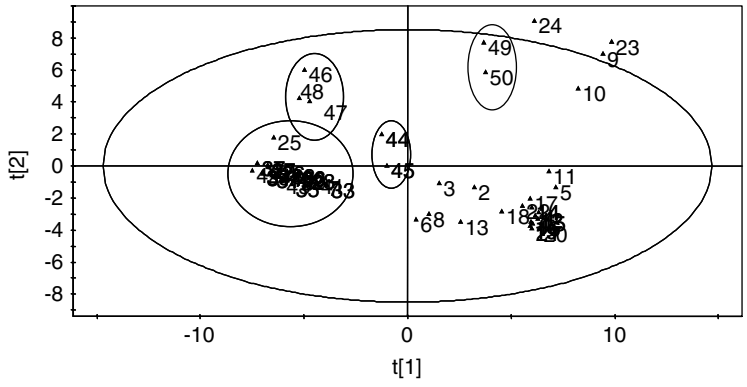

Figure 2 First two components from PCA of entire data set.

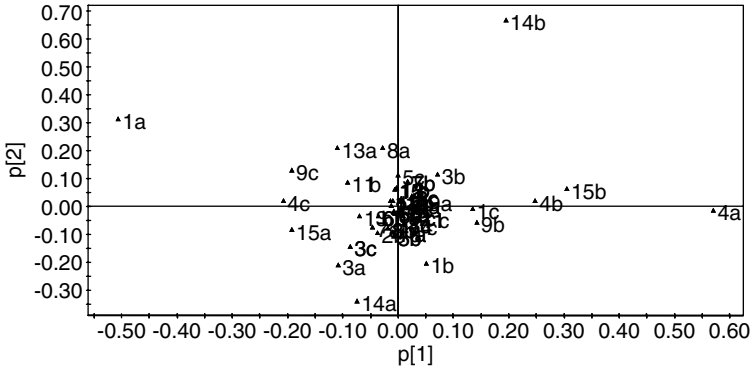

Figure 3 First two loadings from PCA of entire data set.

why the peptides that are substituted in position 14 were outliers.

In order to let all the variables have the same influence on the calculations, a new PCA was performed in which the variables were scaled to unit variance. This resulted in a model explaining only 54% of the variation using four components. Retro-LFB is clearly outside the confidence limits while the difference between the peptides of different origin was not as clear as in the previous calculation. When a new model was calculated, excluding retro-LFB, a model explaining 41% of the variation using three significant variables was obtained. In this model the two peptides derived from porcine lactoferricin did not fit the model and were found outside the 95% confidence limit. The result was almost identical if the *z*-variables were scaled to unit variance and the biological responses were used as the logarithm of the MIC-value. All the above calculations reveal that the inter-species difference in peptide composition is large and that peptides of different origin should, *a priori*, be treated as separate groups.

PLS for Extended Data Set

A new PLS model was, however, calculated including all synthesized LFM peptides to see if it was possible to model activity using peptides of different origin. As expected, the model resulting from these calculations was not as good as the previous ones as judged from the correlation between observed and predicted responses (*E. coli* $r^2 = 0.79$ and *S. aureus* $r^2 = 0.75$) but the predicted values were closer to the observed ones for the test peptides, with the response for *S. aureus* being quite close. Thus, this indicates that including peptides in which variation in substitution pattern is large will lead to models that predict biological responses better than smaller, more homogenous, data sets.

CONCLUSIONS

The computational methods employed in this study, combined with the theoretically derived *z*-scales, prove that peptides of different origin can be classified into different groups and that the entire set of peptides can fairly accurately predict the activity of new peptides. That activities are not more accurately predicted can be explained in part by the selection of the new peptides to be tested, since more substitutions were undertaken simultaneously in these peptides compared with the model peptides, i.e. by making fewer changes in the new compounds should result in more accurate models. However, by making more substitutions a larger variation in amino acid content can be investigated and, if needed, be included in new models. If shorter peptides were to be used this problem would probably be of minor importance, but even for long peptides it is possible to use this method for the design of peptides with improved properties.

Another possible explanation for the incomplete predictive ability could also be due to differences in the modes of action of peptides when the amino acid composition is changed. For example, the proposed 'carpet-mechanism' can be valid for the originally tested peptides, but other mechanisms, possibly intracellular, could be more important in the new peptides. Work is in progress in order to address these issues.

Acknowledgement

Alpharma AS is greatly acknowledged for financial support to Trine Stiberg, Morten B. Strøm and John S. Svendsen.

REFERENCES

1. He J, Furmanski P. Sequence specificity and transcriptional activation in the binding of lactoferrin to DNA. *Nature* 1995; **373**: 721–724.
2. Vorland LH. Lactoferrin: a multifunctional glycoprotein. *APMIS* 1999; **107**: 971–981.
3. Rekdal Ø, Andersen J, Vorland LH, Svendsen JS. Construction and synthesis of lactoferricin derivatives with enhanced antibacterial activity. *J. Pept. Sci.* 1999; **5**: 32–45.
4. Yang N, Stensen W, Svendsen JS, Rekdal Ø. Enhanced antitumor activity and selectivity of lactoferrin-derived peptides. *J. Pept. Res.* 2002; **60**: 187–197.
5. Strøm MB, Rekdal Ø, Svendsen JS. Antibacterial activity of 15-residue lactoferricin derivatives. *J. Pept. Res.* 2000; **56**: 265–274.
6. Haug BE, Svendsen JS. The role of tryptophan in the antibacterial activity of a 15-residue bovine lactoferricin peptide. *J. Pept. Sci.* 2001; **7**: 190–196.
7. Haug BE, Skar ML, Svendsen JS. Bulky aromatic amino acids increase the antibacterial activity of 15-residue bovine lactoferricin derivatives. *J. Pept. Sci.* 2001; **7**: 425–432.
8. Haug BE, Andersen J, Rekdal Ø, Svendsen JS. Synthesis of a sulphonylated trytophan: the antibacterial activity of bovine lactoferricin peptides containing trp(2-Pmc). *J. Pept. Sci.* 2002; **8**: 307–313.
9. Strøm MB, Rekdal Ø, Stensen W, Svendsen JS. Increased antibacterial activity of 15-residue murine lactoferricin derivatives. *J. Pept. Res.* 2001; **57**: 127–139.
10. 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.
11. Hellberg S, Sjöström M, Skagerberg B, Wold S. Peptide quantitative structure-activity relationships, a multivariate approach. *J. Med. Chem.* 1987; **30**: 1126–1135.
12. Lejon T, Strøm MB, Svendsen JS. Antibiotic activity of pentadecapeptides modelled from amino acid descriptors. *J. Pept. Sci.* 2001; **7**: 74–81.
13. Strøm MB, Rekdal Ø, Svendsen JS. The effects of charge and lipophilicity on the antibacterial activity of undecapeptides derived from bovine lactoferricin. *J. Pept. Sci.* 2002; **8**: 36–43.
14. King DS, Fields CG, Fields GB. A cleavage method which minimizes side reactions following Fmoc solid phase peptide synthesis. *Int. J. Peptide Protein Res.* 1990; **36**: 255–266.
15. Amsterdam D. Susceptibility testing of antimicrobials in liquid media. In *Antibiotics in Laboratory Medicine*, 4th edn, Lorian V (ed.) Williams and Wilkins: Baltimore, MD, 1996; 52–111.