

Are Hugo Pripp · Tomas Isaksson · Leszek Stepaniak ·
Terje Sørhaug

Quantitative structure-activity relationship modelling of ACE-inhibitory peptides derived from milk proteins

Received: 5 April 2004 / Revised: 8 July 2004 / Published online: 15 September 2004
© Springer-Verlag 2004

Abstract Quantitative structure-activity relationship (QSAR) modelling was performed on peptides derived from milk proteins that inhibit angiotensin-I-converting enzyme (ACE). Physico-chemical descriptors expressed hydrophobicity, size and charge of side chains of the two most external amino acids in N- or C-terminal position. Models were estimated with partial least squares regression and validated with full cross-validation. A relationship ($R=0.73$, $p<0.001$) was found between hydrophobicity and positively charged amino acid in C-terminal position, size of amino acid next to C-terminal position and ACE-inhibition of peptides up to six amino acids in length. When longer peptides were included the relationship between C-terminal structure and activity decreased, reflecting the likely influence by steric effects. No relationship between N-terminal structure and inhibition activity was found. These biochemical interpretations were supported by findings from QSAR-modelling using so-called z-scales developed by Jonsson et al. (1989, *Quant. Struct.-Act Relat.* 8, 204–209) for amino acids.

Keywords QSAR · ACE-inhibition · Peptides · Milk proteins

Introduction

Angiotensin-I-converting enzyme (ACE) is by definition associated with the renin-angiotensin system, which regulates peripheral blood pressure. The enzyme can increase blood pressure by converting angiotensin I to the potent vasoconstrictor, angiotensin II, and catalyse the degradation of bradykinin and enkephalins. Inhibition of ACE

may therefore exert an antihypertensive effect and potent synthetic inhibitors of ACE are used extensively in the treatment of hypertension in humans [1]. Several peptides derived from milk proteins by enzymatic cleavage have also been found to be potent inhibitors of ACE [2, 3]. Thus, foods enriched with such inhibitory peptides could be targeted towards consumers as functional food with milk-derived nutraceuticals that reduce blood pressure. A composition of peptides in the foods giving a high inhibitory potency would be crucial to obtain claimed health effects. Taking into account the large amount of theoretical possible peptides, i.e. 400 dipeptides, 8,000 tripeptides, 160,000 quatripeptides, 3.2×10^6 pentapeptides etc., the examination of all possible peptides to find highly efficient inhibitors would be a daunting task. At present, the main strategy has been to identify and characterise inhibitory peptides isolated from enzymatic digests of proteins, but methods describing the relationship between peptide structure and ACE-inhibition are needed to predict theoretical inhibitory potential of food protein hydrolysates.

One approach is to develop statistical models that predict the relationship between structure (e.g. amino acid sequence) and activity (e.g. ACE-inhibition). The approach is referred to as quantitative structure-activity relationship (QSAR) modelling. Besides contributing to a biochemical understanding of which peptides show activity, it provides a tool to predict the amino acid sequence of peptides that would give a potent inhibitory potential. Using physico-chemical variables to describe the chemical structure of active components, data analysis can reveal the relationship between activity and structure. Besides a suitable dataset of active compounds, an optimal set of descriptor variables is critical in QSAR modelling. Descriptor variables such as hydrophobicity/-philicity, molecular mass and shape, charge and electronic properties of the individual amino acids are good candidates. The development within the field of chemometrics has introduced powerful regression-type techniques such as partial least square regression (PLSR) to find relationships between variables [4]. QSAR-modelling makes possible the

A. H. Pripp (✉) · T. Isaksson · L. Stepaniak · T. Sørhaug
Department of Chemistry, Biotechnology and Food Science,
Agricultural University of Norway,
Chr. Magnus Falsens vei 1, P.O. Box 5003, 1432 Ås, Norway
e-mail: are-hugo.pripp@ikbm.nlh.no
Tel.: +47-64-948592
Fax: +47-64-947720

identification of relationships between variables describing the chemical structure of peptides and their activity.

A recent development of QSAR on peptides is the use of amino acid "z-scores" obtained by principal component analysis (PCA) of property data [5, 6]. The z-scores have been interpreted as related to hydrophilicity (z1-score), side-chain bulk (z2-score) and electronic properties (z3-score) of amino acids. The z-scores have been proven to be useful for modelling a number of biological effects of small peptides, among them inhibition of ACE activity by dipeptides [7]. Since a QSAR-modeling study of ACE-inhibition activity of dipeptides has been proven successful and provided valuable information, it seems warranted to extend it to larger peptides and with variables more directly linked to physico-chemical properties than the z-scores.

The objective of this study was to use hydrophobicity/-philicity, molecular size and charge properties as descriptors of the amino acids in QSAR-modelling of ACE-inhibiting peptides derived from milk proteins and compare these results with modelling using z-scores. Based on obtained models, biochemical interpretation of the relationship between structure and activity may be expected.

Materials and methods

The primary structure of ACE inhibitory peptides derived from milk proteins and their respective inhibition of activity (Table 1) expressed as peptide concentration (μM) required to inhibit ACE by 50% ($\text{IC}_{50\%}$) was obtained from a review by Fitzgerald and Meisel [2]. Peptides up to a length of eight amino acids and with a specific measurement of $\text{IC}_{50\%}$ in μM were used in the modelling. Five continuous and three categorical variables describing the physico-chemical properties of amino acids were used (Table 2). Each amino acid at a given position in the peptides was thus described by eight variables. Special focus was given to the two external amino acids in N- or C-terminal position further called N- or C-terminal

Table 1 Peptide fragments from milk proteins and their ACE inhibition expressed as $\log \text{IC}_{50\%}$ ($\mu\text{mol/l}$) (from literature review by Fitzgerald and Meisel [2])

No.	Peptide	$\log \text{IC}_{50\%}$	No.	Peptide	$\log \text{IC}_{50\%}$
1	LW	1.70	19	HIRL	3.06
2	YL	2.09	20	FFVAP	0.78
3	LF	2.54	21	AVPYP	1.90
4	IR	2.84	22	KVLPVP	0.70
5	YP	2.86	23	TTMPLW	1.20
6	YG	3.18	24	YKVPQL	1.34
7	RL	3.39	25	LAYFYP	1.81
8	VAP	0.30	26	YQQPVL	2.45
9	IPP	0.70	27	GLDIQK	2.76
10	VPP	0.95	28	LAHKAL	2.79
11	PLW	1.56	29	VAGTWY	2.83
12	FGK	2.20	30	AVPYPQR	1.18
13	PYP	2.34	31	ALPMHIR	1.63
14	AVP	2.53	32	FPEVFGK	2.15
15	HIR	2.98	33	YFPFGPIP	2.70
16	FVAP	1.00	34	KVLPVPQ	3.00
17	YLLF	2.24	35	SLVLPVPE	1.59
18	YGLF	2.87	36	DAYPSGAW	1.99

Table 2 Physico-chemical properties of amino acids used as descriptor variables

Variable no.	Property	Ref.
1	Hydrophilicity score	[14]
2	Hydropathy score	[15]
3	Side-chain hydrophobicity (J/mol)	[16]
4	Van der Waals volume (cm^3/mole)	[4]
5	pI	[16]
6	Positively charged side chains	[16]
7	Negatively charged side-chains	[16]
8	Aromatic groups in side-chains	[16]

dipeptide, respectively. The N- or C-terminal dipeptide could then be described by a 17-term model (Eq. 1).

$$y = b_0 + \sum_{j=1}^2 \sum_{i=1}^8 b_{ij} x_{ij}, \quad (1)$$

where y is the modelled property (ACE-inhibition), $j=1,2$ is the position of the amino acid in the N- or C-terminal dipeptide, $i=1, \dots, 8$ is the descriptor variables (Table 2) and consequently x_{ij} is descriptor variable number i in position j .

The amino acid z-scores used were those reported by Jonsson et al. [5]. Using the three z-scores for each amino acid, properties to a dipeptide region could be described by a six-term model, where the terms z_{11} , z_{21} and z_{31} were scores for the amino acid in position 1 and z_{12} , z_{22} and z_{32} were scores for the amino acid in position 2 giving the following model

$$y = b_0 + b_{11}z_{11} + b_{21}z_{21} + b_{31}z_{31} + b_{12}z_{12} + b_{22}z_{22} + b_{32}z_{32} \quad (2)$$

Modelling of the ACE inhibition based on the physico-chemical descriptor variables or the z-scores for amino acids was performed calculating partial least square regression (PLSR) with full cross validation. The predictive accuracy of a model is described by the multivariate correlation coefficient (R) after full cross-validation and the root mean square error of prediction (RMSEP), which is defined for a response variable y_i as

$$\text{RMSEP} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (3)$$

where y_i is the measured data from analysis of sample number i , \hat{y}_i is the predicted value obtained by full cross validation, and N is the number of objects used in the prediction. In order to compare the influence of measurement uncertainty on the predictive ability of the models, the relative ability of prediction (RAP) is used [8]. RAP was originally developed in connection with sensory analysis and takes into account the level of experimental error in the reference data. It is defined as

$$\text{RAP} = \frac{\text{SD}_{\text{tot}}^2 - \text{RMSEP}^2}{\text{SD}_{\text{tot}}^2 - \text{SD}_{\text{ref}}^2} \quad (4)$$

where SD_{tot} is the standard deviation for the data-set, SD_{ref} is a standard error that indicates the uncertainty of the analysis and RMSEP is the root mean square error of prediction (after full cross-validation). By removal of SD_{ref} , the RAP equation will express the squared multivariate correlation coefficient [9].

Partial least square regression was performed with the Unscrambler software, version 8.0 (Camo Process AS, Oslo, Norway) and statistical test of regression coefficient with the Minitab Statistical Software, release 13.1 (Minitab Inc., State College, PA, USA).

Results and discussion

Fitzgerald and Meisel [2] reviewed properties of milk protein hydrolysates and bioactive peptides and compiled a group of ACE-inhibitory peptides derived from milk proteins (Table 1). A set of variables describing physico-chemical properties of amino acids (Table 2) was chosen. Previous research on structure-activity relationship between ACE-inhibitory peptides have pointed out that hydrophobic amino acid residues and charged side groups influence inhibitory potential. To cover different effects of hydrophobicity three descriptive variables were included in the QSAR-modelling. The Van der Waals volume describes molecular volume and thereby steric effects of amino acid side-chains. Initially, the molecular weight of amino acids was also included in the models, but since van der Waals volume and molecular mass of amino acids are highly correlated ($r=0.96$) this descriptor was later neglected in the modeling work. Three categorical variables describing aromatic (i.e. Trp, Tyr, Phe) and positively (i.e. Lys, Arg, His) or negatively (i.e. Asp, Glu) charged side chains were also included with a value of 1 for presence of given type of amino acid and value of 0 for absence.

Modelling work with the data set revealed that a logarithmic transformation of $IC_{50\%}$ -values improved the models. This is in agreement with previous QSAR-modelling using so-called z-scores to amino acids on a set of ACE-inhibiting dipeptides [7]. The ACE-inhibition was also in that study expressed as a logarithmic transformation of $IC_{50\%}$ -values. The data set of ACE-inhibitory peptides derived from milk proteins is made up of different research work over a 15-year period, and the ACE inhibition was measured using different assays and expressions of measurement uncertainty are lacking in several of the studies. It was therefore difficult to derive the precise uncertainty measurement based on those data. However, recently in our laboratory a study on ACE-inhibition by ethanol-soluble peptides from fish was conducted using an extract of rabbit lung acetone powder as source of ACE and furanacryloyl-Phe-Gly-Gly (FAPGG) as substrate [10]. The standard deviation between replicates for $\log IC_{50\%}$ based on the data from that study was 0.06. At such a low measurement uncertainty, the difference between RAP and squared multivariate correlation coefficient was neglectable. RAP is one if measurement uncertainty is equivalent to the value of RMSEP.

QSAR-models on N- or C-terminal dipeptides for subsets of ACE-inhibiting peptides with increasing length were calculated by PLSR and the predictive ability assessed by the multivariate correlation coefficient (R) (Table 3). A highly significant correlation ($p<0.001$) was found for the QSAR-model based on the C-terminal dipeptide for peptides up to six amino acids in length. The low R for the data set comprising only di- and tripeptides are likely a result of a low number of samples used in the model. When the data set was increased with longer peptides, the prediction ability improved. The lower R-value when peptides with seven and eight amino acids

Table 3 The multivariate correlation coefficient (R) for QSAR-models on the two most external amino acids (aa) in the N- or C-terminal position using physio-chemical descriptor variables (Table 2) for subsets (n =number of peptides in subset) including peptides of increasing length

Peptide length	n	R			
		N-terminal position		C-terminal position	
2-3 aa	15	0.22	NS	0.11	NS
2-4 aa	19	0.32	NS	0.69	^c
2-5 aa	21	0.41	NS	0.73	^c
2-6 aa	29	-0.40	NS	0.69	^c
2-7 aa	33	-0.12	NS	0.52	^b
2-8 aa	36	-0.16	NS	0.50	^b

NS Not significant

^a Significant on 5% level

^b Significant on 1% level

^c Significant on 0.1% level

were included, might reflect that for smaller peptides the ACE inhibition potential is primarily a result of the C-terminal structure, but as length increases steric effects that are not expressed by this QSAR-model begins to interfere with the results. In contrast to C-terminus, in which the composition of the two most external amino acids had a clear influence on the ACE inhibitory potency, the two amino acids at N-terminus had no apparent influence on inhibition.

Eight variables described each of two terminal amino acids and regression model with 17 terms were generated (Eq. 1). Work was undertaken to optimise the model using fewer terms. Martens and Martens [11] developed a modified “Jack-knife” method to identify collineary and noisy variables that can be left out to improve the model. An improved prediction (RMSEP=0.60, $R=0.73$, $p<0.001$) was obtained with a three-term model using side-chain hydrophobicity (x_1), positively charged side chain for amino acid (x_2) in C-terminal position and van der Waals volume for the amino acid (x_3) next to C-terminal position. Peptide LAHKAL was so poorly predicted by this model that it was considered as an outlier. The QSAR model is given in Eq. 5 and a plot of predicted versus measured values is given in Fig. 1.

$$\log IC_{50\%} = 1.46 - 9.29 \cdot 10^{-5}x_1 + 0.52x_2 + 3.21 \cdot 10^{-2}x_3 \quad (5)$$

Based on the regression coefficient of the weighted variables, the three terms are of relatively similar importance for predicting ACE inhibition by peptides. A biochemical interpretation of this QSAR-model is that increased side chain hydrophobicity of the amino acid and absence of positive charge in the C-terminal position enhance ACE-inhibitory potential (decreased $\log IC_{50\%}$), while increased side chain size of the amino acid next to the C-terminal position decreases ACE-inhibitory potential. Such an interpretation of the QSAR-model is supported by previous research on the structure-activity relationship of ACE-inhibitory peptides pointing out that the C-terminal tripeptide region of the substrate influ-

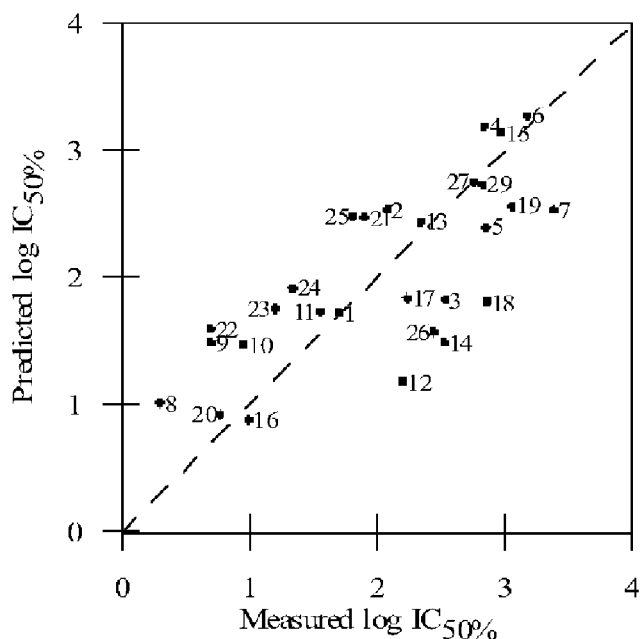


Fig. 1 Predicted versus measured values for inhibition of angiotensin-I-converting enzyme by peptides up to six amino acid residues in length derived from milk proteins using the QSAR-model (Eq. 5) based on hydrophobicity, van der Waals volume and positively charged side chains of the two last amino acids comprising C-terminus (RMSEP=0.60, R=0.73, $p < 0.001$). Numbers correspond to peptides in Table 1

ences binding to ACE significantly and that peptides containing hydrophobic amino acid residues in the C-terminal region display a high potency for inhibition [12, 13]. Some studies have suggested that the presence of amino acids with a positively charged side group contributes significantly to ACE inhibitory potency [2] The effect of positively charged side groups needs further investigation. Omitting that variable from the model presented in Eq. (5) somewhat reduced the predictive ability (RMSEP=0.61, R=0.71, $p < 0.001$) without altering the relative effect and biochemical interpretation of the two remaining variables.

As an alternative to QSAR-modelling on specific physico-chemical descriptor variables for the amino acids, principal component analysis (PCA) as performed on large amount of structural and property data for amino acids indicated that three principal components accounted for over 80% of the variance in the data [5, 6]. The three first PCs represented properties closely aligned with the chemical concepts of hydrophilicity (z_1), molecular size (z_2) and electronic properties (z_3). Using this approach, ACE inhibition by peptides derived from milk proteins (Table 1) were examined for structure-activity relationships. The same approach for QSAR-modelling as presented in Table 3, but using z-scores [5] instead of physico-chemical variables was explored (Table 4). Similar to the models with physico-chemical variables, highly significant predictive ability was found for peptides up to six amino acids in length based on the prop-

Table 4 The multivariate correlation coefficient (R) for QSAR-models on the two most external amino acids (aa) in the N- or C-terminal region using z-scores (from reference [5] for subsets (n = number of peptides in subset) including peptides of increasing length

Peptide length	n	R	
		N-terminal position	C-terminal position
2-3 aa	15	0.33	0.76
2-4 aa	19	0.32	0.84
2-5 aa	21	0.45	0.85
2-6 aa	29	-0.28	0.67
2-7 aa	33	-0.14	0.50
2-8 aa	36	-0.20	0.44

NS Not significant

^a Significant on 5% level

^b Significant on 1% level

^c Significant on 0.1% level

erties of the dipeptide region in the C-terminal position. Again, no correlation between the N-terminal structure and ACE-inhibition was found; except for peptides up to five amino acid residues at 5%-level, but taken into account the lack of significant correlation for the other subset of peptides this is likely a coincidence. These findings support that ACE-inhibiting potential for smaller peptides are linked to the composition of the C-terminal region, but not the N-terminal, and as peptide length increases other steric effects not related to the C-terminal structure begins to interfere.

Even though similar predictive ability of the QSAR-models was obtained using either z-scores or physico-chemical properties of amino acids, a disadvantage with the z-scores is that the biochemical interpretation of structure-activity relationships becomes more difficult. For the dataset containing peptides from two to six amino acids in length PLSR gave the following model:

$$\log \text{IC}_{50\%} = 2.21 + 0.03z_{1c1} - 0.1z_{2c1} - 0.21z_{3c1} - 0.07z_{1c2} + 0.12z_{2c2} - 0.17z_{3c2}, \quad (6)$$

where z_{1c1} , z_{2c1} and z_{3c1} are the z_1 , z_2 and z_3 scores, respectively, for the amino acid in C-terminal position and z_{1c2} , z_{2c2} and z_{3c2} are the z_1 , z_2 and z_3 scores, respectively, for the amino acid next to the C-terminal position. A plot of predicted versus measured values is given in Fig. 2. Based on the modified Jack-knife method developed by Martens and Martens [11] to identify important terms in PLSR-models, z_{3c1} , z_{2c2} and z_{3c2} were identified. As found using physico-chemical variables in the QSAR-modelling, the peptide LAHKAL was poorly predicted after full cross validation supporting the view that it can be regarded as an outlier. Modelling ACE-inhibition of dipeptides using z-scores has been done by Hellberg et al. [7], who found that hydrophobicity (z_1) and size (z_2) of the C-terminal position followed by electronic effects (z_3) and hydrophobicity (z_1) of the N-terminal position in the dipeptides were the most important factors of the QSAR-model. Studies on peptides longer than dipeptides were not presented.

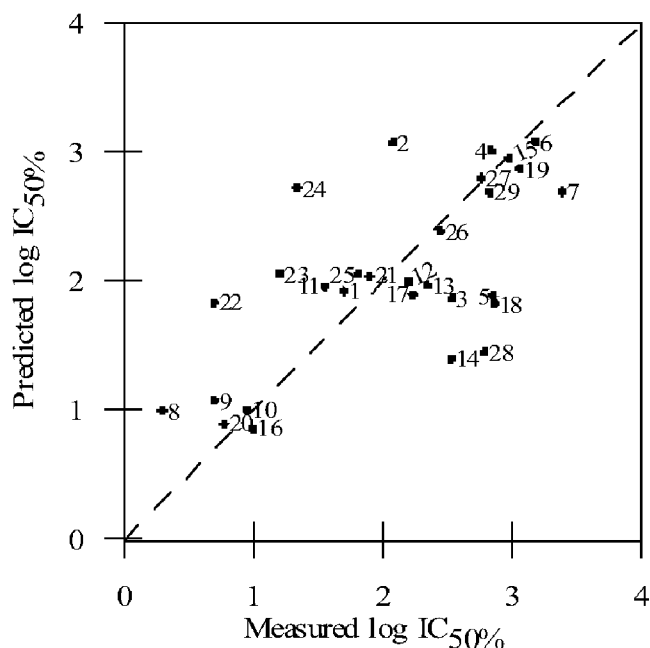


Fig. 2 Predicted versus measured values for inhibition of angiotensin-I-converting enzyme by peptides up to six amino acid residues in length derived from milk proteins using the QSAR-model (Eq. 6) based on z-scores of two last amino acids comprising C-terminus (RMSEP=0.65, R=0.67, $p < 0.001$). Numbers correspond to peptides in Table 1

QSAR-modelling of ACE-inhibitory peptides derived from milk proteins found a relationship between the compositions of the C-terminal region and inhibiting potency. Such a relationship was not apparent for the N-terminal region. A correlation was found between ACE-inhibition and structural properties related to hydrophobicity, positive charge and molecular volume of the amino acids at the C-terminal region covering the two last amino acids for peptides containing up to six amino acid residues. For longer peptides, steric effects not taken into account in the QSAR-model may be important. Measuring ACE-inhibition of peptides predicted to have different

inhibitory potential should further validate the findings of the QSAR-model. Predicting bioactivity of peptides can help to identify food proteins containing encrypted peptides of potential for functional foods.

Acknowledgements This work was funded by the Norwegian Research Council through project "Proteolysis in Foods: Quality and Biological Functions".

References

- Wyvratt MJ, Patchet AA (1985) *Med Res Rev* 5:485–531
- Fitzgerald RJ, Meisel A (2003) Milk protein hydrolysates and bioactive peptides. In: Fox PF, McSweeney PLH (eds) *Advanced dairy chemistry, vol 1, proteins*, 3rd edn. Kluwer Academic/Plenum Publishers, New York, pp 675–698
- Gobbetti M, Stepaniak L, De Angelis M, Corsetti A, Di Cagno R (2002) *Crit Rev Food Sci* 42:223–239
- Esbensen KH (2001) *Multivariate data analysis—in practice*. Camo Process AS, Oslo, Norway
- Jonsson JM, Eriksson L, Hellberg S, Sjöström M, Wold S (1989) *Quant Struct–Act Relat* 8:204–209
- Hellberg S, Sjöström M, Skagerberg B, Wold S (1987) *J Med Chem* 30:1126–1135
- Hellberg S, Eriksson L, Jonsson J, Lindgren F, Sjöström M, Skagerberg B, Wold S, Andrews P (1991) *Int J Pept Protein Res* 37:414–424
- Martens M, Martens H (1986) *Appl Spectrosc* 40:303–310
- Martens H, Naes T (1987) *Multivariate calibration by data compression*. In: Williams PC, Norris K (ed) *Near-infrared technology in the agricultural and food industries*. Assoc Cer Chem, St. Paul, pp 57–87
- Sollie LA (2003) *Bioactive peptides in fish*. MSc Thesis, Agricultural University of Norway
- Martens H, Martens M (2000) *Food Qual Pref* 11:5–16
- Cheung HS, Feng-Lai W, Ondetti MA, Sabo EF, Cushman DW (1980) *J Biol Chem* 255:401–407
- Ondetti MA, Cushman DW (1982) *Ann Rev Biochem* 51:283–308
- Hopp TP, Woods KR (1981) *Proc Natl Acad Sci USA* 78:3824–3828
- Kyte J, Doolittle R (1982) *J Mol Biol* 157:105–132
- Cheftel JC, Cuq J-L, Lorient D (1985) *Amino acids, peptides and proteins*. In: Fennema OR (ed) *Food chemistry*. Marcel Dekker Inc, New York, pp 245–367