Eur Biophys J (2003) 32: 589–598
DOI 10.1007/s00249-003-0302-2

ARTICLE

James Wallace · Frederick Harris · David A. Phoenix

A statistical investigation of amphiphilic properties of C-terminally anchored peptidases

Received: 28 January 2003 / Accepted: 19 March 2003 / Published online: 30 April 2003

Abstract A number of DD-peptidases have been reported to interact with the membrane via C-terminal amphiphilic α-helices, but experimental support for this rests with a few well-characterized cases. These show the C-terminal interactions of DD-carboxypeptidases to involve high levels of membrane penetration, DD-endopeptidases to involve membrane surface binding and class C penicillin-binding proteins to involve membrane binding with intermediate properties. Here, we have characterized C-terminal α-helices from each of these peptidase groups according to their amphiphilicity, as measured by mean $<\mu H>$, and the corresponding mean hydrophobicity, $<H>$. Regression and statistical analyses showed these properties to exhibit parallel negative linear relationships, which resulted from the spatial ordering of α-helix amino acid residues. Taken with the results of compositional and graphical analyses, our results suggest that the use of C-terminal α-helices may be a universal feature of the membrane anchoring for each of these groups of DD-peptidases. Moreover, to accommodate differences between these mechanisms, each group of C-terminal α-helices optimizes its structural amphiphilicity and hydrophobicity to fulfil its individual membrane-anchoring function. Our results also show that each anchor type analysed requires a similar overall balance between amphiphilicity for membrane interaction, which we propose is necessary to stabilize their initial membrane associations. In addition, we present a methodology for the prediction of C-terminal α-helical anchors from the classes of DD-peptidases analysed, based on a parallel linear model.

Keywords C-terminal amphiphilic α-helix · DD-peptidase · Membrane anchoring · Penicillin-binding protein · Regression with categorical predictors

D. A. Phoenix (✉)
Dean's Office, Faculty of Science,
University of Central Lancashire, Preston, PR1 2HE, UK
E-mail: daphoenix@uclan.ac.uk
Tel.: +44-1772-893481
Fax: +44-1772-894981

J. Wallace
Department of Physics, Astronomy and Mathematics,
University of Central Lancashire, Preston, PR1 2HE, UK

F. Harris
Department of Forensic and Investigative Science,
University of Central Lancashire, Preston, PR1 2HE, UK

Introduction

An increasing number of proteins have been reported which interact with the membrane via amphiphilic α-helices. These are secondary structural elements characterized by an ordered segregation of hydrophilic and hydrophobic amino acid residues about the α-helical long axis. This segregation of residues allows such α-helices to partition into membranes so that the predominantly apolar face of the anchor penetrates the membrane lipid core and its polar face interacts with the membrane lipid headgroup region (Phoenix and Harris 1998, 2002; Phoenix et al. 2002). Structure–function relationships for these α-helices were established by the work of Segrest et al. (1990), who showed that four major classes of lipid-interactive amphiphilic α-helices could be discerned, based on biological function. These classes were distinguished by their spatial arrangements of charged amino acid residues (Segrest et al. 1992), mean levels of hydrophobicity, $<H>$, and the magnitude of their hydrophobic moment, $<\mu H>$ (Eisenberg et al. 1982a), which provides a measure of α-helix amphiphilicity (Eisenberg et al. 1984a, 1986). Based on similar criteria, structure–function relationships have since been established for several other classes of amphiphilic α-helix (Picot and Garavito 1994; Wendt et al. 1997; Johnson and Cornell 1999; Harris et al. 2000; Phoenix and Harris 2002; Phoenix et al. 2002; Tossi and Sandri 2002).

Phoenix (1990) first proposed that C-terminal amphiphilic α-helices might function as membrane

anchors when studying a number of *Escherichia coli* peptidases. These enzymes are known to bind to the periplasmic face of the cytoplasmic membrane. However, hydropathy plot analysis showed no conventional hydrophobic anchoring sequences, no obvious sites for covalent modification were apparent, and the membrane binding mechanisms of these enzymes were unclear (Gittins et al. 1993). Phoenix (1990) further showed that the C-terminal regions of these enzymes possessed the periodicity in hydrophobic residues associated with lipid-interactive amphiphilic α-helical structure. The participation of these C-terminal amphiphilic α-helices (CAHs) in membrane anchoring was later confirmed experimentally (Phoenix and Harris 1998; Phoenix et al. 2002).

The enzymes analysed in the study of Phoenix (1990) were low molecular mass penicillin-binding proteins (PBPs). These are a family of DD-peptidases involved in bacterial cell wall biosynthesis and the use of CAHs in their membrane anchoring mechanisms appears to be generally regarded as universal (Ghuysen 1994, 1997). However, with the advent of sequenced prokaryotic genomes, membership of this family is increasing. Concomitantly, the numbers of DD-peptidases reported to utilize CAHs in their membrane interactions is growing. These enzymes fall into three distinct groups: DD-carboxypeptidases, DD-endopeptidases and class C PBPs (Granier et al. 1992). Each of these peptidase groups has differing biological requirements (Song et al. 1998) and it has been suggested that there may be functional differences between the membrane anchoring mechanisms employed by their CAHs (Phoenix and Harris 1998, 2002; Phoenix et al. 2002). Experimental support for this conjecture, and indeed the general use of CAHs by DD-peptidases, is nevertheless limited to studies on a few members of each of the DD-peptidase groups reported. The prototypes for these studies appear to be the DD-carboxypeptidases and the class C PBPs of *E. coli* (Nelson and Young 2001; Harris et al. 2002; Nelson et al. 2002; Phoenix et al. 2002) and *Bacillus* spp (Waxman and Strominger 1981a, 1981b, 1981c; Buchanan and Ling 1992; Pedersen et al. 1998). However, although these experimental results are supported by some isolated theoretical analyses (Phoenix 1990; Pewsey et al. 1996; Roberts et al. 1997; Harris 1998; Phoenix and Wallace 2000; Harris et al. 2002), no such analysis of DD-peptidase CAHs appears to have been conducted on a systematic basis.

In this study we have investigated the universality of the use of CAHs as peptidase-membrane anchors by statistical and conventional graphical analyses of their structural amphiphilicity. We show that CAHs of the same DD-peptidase group possess similar amphiphilic properties. We further show that these properties differ, on average, between the various DD-peptidase groups and that these differences can be related to experimentally demonstrated functional differences between the membrane interactions of these CAHs. Our results also suggest that an appropriate balance between amphiphilicity and hydrophobicity may be a general requirement to stabilize the initial membrane associations of biologically active amphiphilic α-helices. Finally, we demonstrate that the unique parallel linear relationships established between the expected values of $<\mu H>$, denoted E($<\mu H>$), and $<H>$, for the peptidase classes analysed, can be used for predicting the anchoring mechanism of newly identified homologues.

Methods

The data set and hydrophobic moment analysis

A data set (Appendix 1) comprising the C-terminal sequences of putative α-helical anchors from 18 DD-carboxypeptidases, 5 DD-endopeptidases and 9 class C PBPs was compiled from the Swissprot data bank (http://www.expasy.ch/sprot/; accessed 11.03.03). These sequences were then characterized according to their hydrophobic moment, a hydrophobicity-based property introduced by Eisenberg et al. (1982a) to represent amphiphilicity. The hydrophobic moment assumes amino acid residues to be periodically spaced, at equal intervals, along the axial backbone of a protein. It is further assumed that, for a putative α-helical protein sequence, consecutive residues subtend an incremental angle of 100°, in a plane perpendicular to this axis. The hydrophobicity values of successive amino acids in these sequences are then treated as vectors and summed in two dimensions. The hydrophobic moment is defined as the absolute value of the resultant of these summed vectors. For a given sequence, it is conventional to compute the mean hydrophobic moment, $<\mu H>$, defined as:

$$\langle \mu H \rangle \frac{1}{L} \left\{ \left(\sum_{i=1}^{L} H_i \cos(i100^\circ) \right)^2 + \left(\sum_{i=1}^{L} H_i \sin(i100^\circ) \right)^2 \right\}^{\frac{1}{2}} \quad (1)$$

where H_i is the hydrophobicity for the i-th residue in the C-terminal sequence, according to the Eisenberg consensus hydrophobicity scale (Eisenberg et al. 1982b), and L represents window size. To characterize α-helical structure, it is usual practice to scan a given amino acid sequence with a moving window of 11 residues and determine the window with maximum hydrophobic moment. Without loss of generality, the first residue of this window was taken as $i=1$. For this window, the corresponding mean hydrophobicity was also computed, which is defined as:

$$\langle H \rangle = \frac{1}{L} \sum_{i=1}^{L} H_i \quad (2)$$

Initial data analysis

The CAHs of our data set were characterized according to hydrophobic moment analysis and the windows of maximum amphiphilicity with corresponding values of $<\mu H>$ and $<H>$ computed. Hydrophobic moment plots of $<\mu H>$ versus $<H>$ (Eisenberg et al. 1984b) were constructed for the CAHs of the complete data set, for the CAHs of DD-carboxypeptidases compared to the CAHs of the other peptidase groups in the data set, and for each individual group of CAHs in the data set.

As represented by windows of maximal amphiphilicity, dotplots of amino acid residue frequency against hydrophobicity value (equivalently amino acid, as the Eisenberg consensus scale has no redundant values present) were obtained for the CAHs of the full data set. To facilitate direct data comparisons, similar dotplots were constructed for the individual groups of CAHs in the data set and a Loess plot (Cleveland 1979), which provides a locally weighted regression scatter plot of the frequencies, scaled for number of residues present, was obtained.

Regression analysis with categorical predictors

First-order regression models for $<\mu H>$ versus $<H>$, with additional indicator variables present (Neter et al. 1996) to represent the three categorical groups of CAHs, were fitted to the data set. The general model employed is given by:

$$\langle \mu H \rangle_i = \beta_0 + \beta_1 \langle H \rangle_i + \sum_{j=1}^{m} \beta_{j+1} I_j + \sum_{j=1}^{m} \beta_{1j} I_j \times \langle H \rangle_i + \varepsilon_i \quad (3)$$

where, $i = 1, ..., n$ is an index for the i-th observation from n, β_0 is the common intercept regression coefficient, β_1 is the regression coefficient for the gradient associated with $<H>$, $\beta_2, ..., \beta_{m+1}$ are the regression coefficients for the m indicator variables, I_j, $j = 1, ..., m$ and the $\beta_{11}, ..., \beta_{1m}$ are the regression coefficients for the interaction, $I_j \times <H>$, of the j-th indicator variable with the mean hydrophobicity for the optimum window of the i-th peptidase. The error, ϵ_i, for the i-th observation is considered to be random and, in the general case, to follow an unspecified distribution, $\epsilon_i \approx (0, \sigma^2)$, with a mean value of zero and with constant variance, σ^2, for all $i = 1, ..., n$.

The regression coefficients associated with the indicator variables and interaction terms have a particular interpretation. The β_j, $j = 2, ..., m+1$, coefficients for the m indicator variables, representing the $m+1$ categories being modelled, represent the difference in the intercept value of the corresponding category to that for a selected baseline category. Similarly, the β_{1k} coefficients represent the difference in the gradient for the $k+1$-th category, again relative to that of the chosen baseline category, which in all cases was the group of DD-carboxypeptidase CAHs.

Two-category regression

The CAHs for our data set were classified into two distinct groups. The CAHs of DD-carboxypeptidases were designated as the baseline group, with the CAHs from the remaining categories forming the other group, and a single indicator variable, I, was created to reflect this. Regression analyses were then conducted on a model with an interaction present and with this absent, based on this constraint. Using the notation of Eq. (3), the full model, where interaction is present, is given by:

$$\langle \mu H \rangle_i = \beta_0 + \beta_1 \langle H \rangle_i + \beta_{11} I \times \langle H \rangle_i + \varepsilon_i \quad (4)$$

where:

$$I = \begin{cases} 0 & i \text{ from baseline category} \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

When $I = 0$, Eq. (4) reduces to:

$$\langle \mu H \rangle_i = \beta_0 + \beta_1 \langle H \rangle_i + \varepsilon_i \quad (6)$$

and we have a simple regression equation. However, when $I = 1$, after rearrangement, Eq. (4) now becomes:

$$\langle \mu H \rangle_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_{11}) \langle H \rangle_i + \varepsilon_i \quad (7)$$

Equation (7) is also that for a simple linear regression relationship. We are now able to see from Eqs. (6) and (7) that β_2 and β_{11} represent the difference in the intercept and gradient regression coefficients, respectively, for the second category to that of the baseline category, for a simple linear model.

Further regression analysis was also carried out on the two-category model without interactions present. The regression model for this is given by:

$$\langle \mu H \rangle_i = \beta_0 + \beta_1 \langle H \rangle_i + \beta_2 I + \varepsilon_i \quad (8)$$

By analogy with Eqs. (6) and (7), when $I = 0$, Eq. (8) reduces to Eq. (6) and we have a simple regression equation. However, when $I = 1$, Eq. (8) now becomes:

$$\langle \mu H \rangle_i = (\beta_0 + \beta_2) + \beta_1 \langle H \rangle_i + \varepsilon_i \quad (9)$$

Clearly, from Eqs. (8) and (9) the expected values of $<\mu H>$, $E(<\mu H>)$, represent two parallel lines with a common gradient given by β_1. The regression coefficient β_2 is the difference in intercept for the second category, represented by the indicator variable value $I = 1$, relative to the baseline category, represented by $I = 0$. In the degenerate case where $\beta_2 = 0$, these are coincident lines.

The two models were fitted to the data and residual analyses were conducted on the standardized residuals to check for normality and homoscedasticity of the errors. Confidence intervals and the corresponding tests for non-zero values for all the regression coefficients were also obtained, under the normality assumption.

Three-category regression

A more detailed linear model with categorical predictors present was also fitted to the data. Again, the CAHs of DD-carboxypeptidases were designated as the baseline category, with the CAHs from the remaining categories forming separate groupings. Two indicator variables, I_1 and I_2, were created to represent this. Regression analyses were then conducted on a model with all first-order interactions present and on an additional model with these absent. Using the notation of Eq. (3), the full model, where interactions are present, is given by:

$$\langle \mu H \rangle_i = \beta_0 + \beta_1 \langle H \rangle_i + \beta_2 I_1 + \beta_3 I_2 + \beta_{11} I_1 \times \langle H \rangle_i + \beta_{12} I_2 \times \langle H \rangle_i + \varepsilon_i \quad (10)$$

The reduced model with no interactions present is given by:

$$\langle \mu H \rangle_i = \beta_0 + \beta_1 \langle H \rangle_i + \beta_2 I_1 + \beta_3 I_2 + \varepsilon_i \quad (11)$$

The two models were fitted to the data and the associated standardized residuals were subsequently checked for satisfaction of the distributional assumptions. Confidence intervals for all the regression coefficients for both the models were also obtained and corresponding tests for non-zero values were also undertaken.

Bootstrap regression of three-category model

The statistical requirements of a generalized linear model are that the observations are independent, with the errors following a member of the exponential family of distributions. If normality is assumed, the variance of the errors is further required to be homoscedastic. Whilst violation of the normal theory model does not invalidate the least squares regression equation, the associated statistical tests for the regression coefficients could be unreliable.

Notwithstanding the known robustness of regression analysis to minor departures from these (Draper and Smith 1998), with small sample sizes the statistical tests of the distributional assumptions can be considerably underpowered. More importantly, tests for the regression parameters are also underpowered and corresponding confidence intervals over-conservative.

To verify the three-category regression model with no interaction terms present, as expressed in Eq. (11), a bootstrap regression in pairs simulation (Freedman 1981; Efron and Tibshirani 1993) was undertaken on the $<\mu H>_i$, $<H>_i$ paired observations, $i = 1,..., n$. Here, however, resampling was further stratified for our categories, with sampling with replacement of 18, 5 and 9 from the three categories, respectively. One thousand bootstrap replicates of each of the regression coefficients β_0, β_1, β_{11} and β_{12} were obtained and 95% bootstrap percentile confidence intervals were calculated for each of these to provide corroborations with those obtained under the normality assumption.

Regression on randomized sequences

To establish whether any identified linear model is either an artefact of the mathematical relationship between $<\mu H>$ and $<H>$ or even specific to the classes of CAHs analysed, a Monte Carlo randomization simulation study (Manly 1997) was conducted for values of $<H>$ within the surface-active range.

The order of the primary sequences of amino acid residues present in the original optimum $<\mu H>$ 11-residue windows of our data set were randomized. The new $<\mu H>$ for each window was then calculated and only those $<\mu H>$ and $<H>$ pairs ($<H>$ remains unchanged) whose corresponding $<\mu H>$ was within the range (0.25, 0.75), consistent with that of the original data set, were retained. The simple, two-predictor model with a common gradient detailed in Eq. (4) was then fitted to the 32 randomized pairs and the regression coefficients so obtained were recorded. This process was repeated 1000 times iteratively and 95% percentile confidence intervals were constructed for the two regression coefficients β_0 and β_1.

Results

Initial data analysis

Dotplots of the overall amino acid counts associated with the windows of maximal amphiphilicity for the groups of peptidase CAHs were constructed. Inspection of the dotplot for the total number of CAHs in our data set revealed a high abundance of strongly hydrophobic residues ($H > 1.0$). This was balanced by a substantial number of charged and polar residues ($H < 0$), particularly the strongly positively charged residues arginine ($H = -2.53$) and lysine ($H = -1.5$). These properties are consistent with the amphiphilic nature of the CAHs of our data set. As the counts represented by these dotplots are influenced by the relative number of protein sequences available, a Loess density plot (Cleveland 1979), scaled for group size, is presented in Fig. 1.

From this density plot, the relative abundances of charged and polar residues are seen to be similar for each group of CAHs. However, the relative abundances of strongly hydrophobic residues in these groups show clear differences with the rank order: DD-carboxypeptidases > class C PBPs > DD-endopeptidases being present.

A scatterplot of $<\mu H>$ against $<H>$ with each group of CAHs individually identified is given in Fig. 2. From this plot it is most evident that an overall negative trend is present in the data. Whilst this is also observed for each of the three groups of CAHs, the linear relationship for the smaller, DD-endopeptidase group is less well defined, with a high level of scatter also being present. However, it is shown later that the corresponding gradient shifter, β_{12}, for the regression with categorical variables is not significant, confirming the consistency of the negative trend. It would also appear that the gradient in each case is similar, but that there may well be relative displacements in $E(<\mu H>)$. This is most noticeable for the CAHs of DD-carboxypeptidases versus those of class C PBPs.

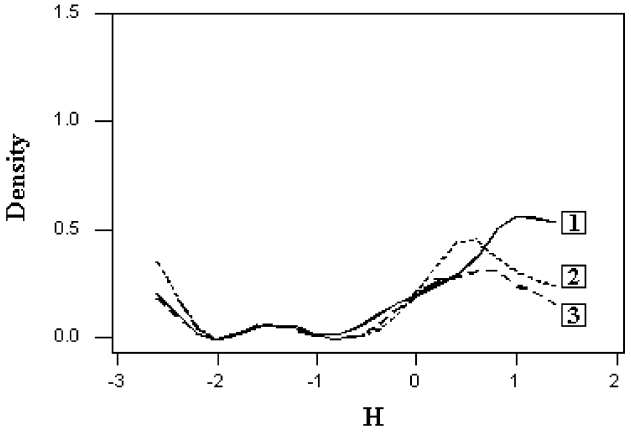

Fig. 1 Loess plot showing the relative amino acid residue abundances in the CAHs of (*1*) DD-carboxypeptidases, (*2*) class C PBPs and (*3*) DD-endopeptidases

Two-category regression

The coefficients for the regression of $<\mu H>$ on $<H>$ with two categorical predictors and interaction term present are given in Table 1. The *P*-value for each is also included. The residuals appeared to follow a normal distribution ($P = 0.421$, Anderson-Darling test; $P > 0.15$, Kolmogorov-Smirnov test). Bartlett's test for homoscedasticity of the residuals was also non-significant at the 10% level ($P = 0.433$) and there was no evidence of any influential observations being present (Cook's distance < 0.12 for each fitted value). A plot of the residuals against the fitted values for the model, identified by category, also showed a random distribution with no

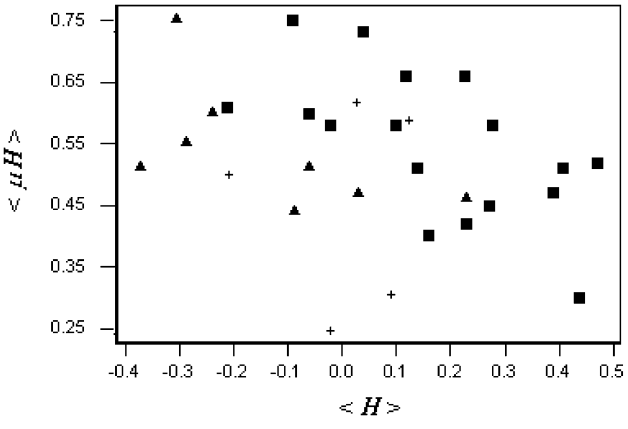

Fig. 2 Scatterplot of $<\mu H>$ against $<H>$ for the CAHs of DD-carboxypeptidases (*squares*), class C PBPs (*triangles*) and DD-endopeptidases (*plusses*). For each of the first two of these groups of CAHs, two sequences identified by hydrophobic moment analysis showed 100% homology (see Appendix 1), and therefore each of these groups contains a data point, degenerate for two sequences

Table 1 Two-category regression with interaction

Regression coefficient	Regression coefficient value	*P*-value
β_0	0.618	< 0.0005
β_1	−0.360	0.015
β_2	−0.126	0.013
β_{11}	−0.013	0.953

systematic departures. The normal theory model assumptions therefore appear to hold. The regression coefficients β_0, β_1 and β_2 are all significant at the 5% level (Table 1), but there is strong evidence from the data that β_{11} is not statistically different from zero. There is, however, considerable variability about the fitted line, with an adjusted coefficient of determination value of 0.23. A linear model with different intercept values for the two categories but with a common gradient therefore provides a reasonable fit to the data, albeit with some degree of spread being present.

The reduced model with the interaction term, representing different gradients removed, was fitted to these data. The regression coefficient values and associated *P*-values for these are given in Table 2. As would be expected from the analysis of the residuals from the fuller model, the residuals appeared to follow a normal distribution ($P = 0.402$, Anderson-Darling test; $P > 0.15$, Kolmogorov-Smirnov test). Bartlett's test for homoscedasticity of the residuals was also non-significant at the 10% level ($P = 0.675$) and there was no evidence of any influential observations being present (Cook's distances < 0.14). A plot of the residuals against the fitted values for the model, identified by category, again showed a random distribution with no systematic departures. The adjusted coefficient of determination was 0.253. As all the regression coefficients present in the model are highly significant, the model provides a reasonable fit to the data but with high variability being present.

Three-category regression

The coefficients for the regression of $<\mu H>$ on $<H>$ with three categorical predictors and first-order interaction terms present are given in Table 3 together with the associated *P*-values for each term. The regression coefficients β_0, β_1, β_2 and β_3 are all significant at the 7% level, but there is strong evidence from the data that the first-order interaction coefficients β_{11} and β_{12} are not statistically different from zero.

Table 2 Two-category regression with no interaction

Regression coefficient	Regression coefficient value	*P*-value
β_0	0.618	< 0.0005
β_1	−0.365	0.001
β_2	−0.126	0.011

Table 3 Three-category regression with first-order interactions

Regression coefficient	Regression coefficient value	*P*-value
β_0	0.618	< 0.0005
β_1	−0.360	0.019
β_2	−0.164	0.015
β_3	−0.116	0.070
β_{11}	−0.400	0.392
β_{12}	−0.014	0.954

The reduced three-category model, with no interaction terms present, was next fitted to the data. The results of this are given in Table 4. The regression coefficients β_0, β_1, β_2 and β_3 are all significant at the 5.9% significance level. There is, however, still considerable variability about the fitted line, with an adjusted coefficient of determination value of 0.23. The normality assumption appears to be reasonable ($P = 0.253$ Anderson-Darling test; $P > 0.15$, Kolmogorov-Smirnov test) and the homoscedasticity assumption also seems to be satisfied ($P = 0.126$, Bartlett's test and from inspection of the spread of residuals plotted against the fitted values of the model). Furthermore, no unduly influential observations are present (Cook's distances < 0.14). A parallel linear model with different intercept values for the three categories, but with a common gradient, therefore provides a reasonable fit to the data. Again, a degree of spread about these parallel fitted lines is present. A graph of the proposed parallel model is given in Fig. 3.

Bootstrap regression of the three-category model

Percentile confidence intervals of 95% for the 1000 bootstrap replicates of the regression coefficients β_0, β_1, β_2 and β_3 for the stratified bootstrap regression in pairs are given in Table 5. From these confidence intervals, and by inspection of corresponding histograms (data not shown), it can be concluded that β_0, β_1, β_2 and β_3 are significantly different from zero and should be included in the linear model. The stratified bootstrap regression in pairs resampling thus confirms and reinforces the results obtained directly from classical regression analysis.

Table 4 Three-category regression with no interactions

Regression coefficient	Regression coefficient value	*P*-value
β_0	0.614	< 0.0001
β_1	−0.337	0.005
β_2	−0.159	0.012
β_3	−0.106	0.059

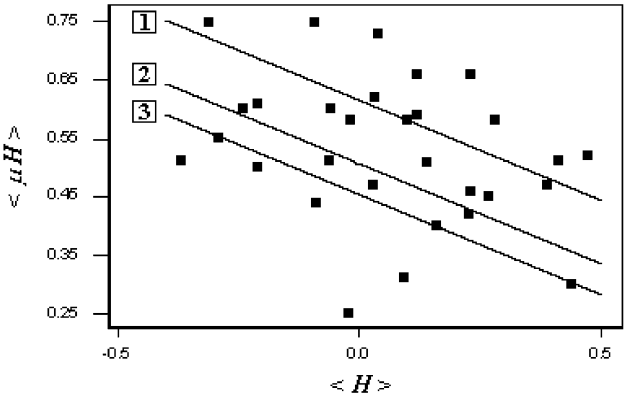

Fig. 3 Three-category parallel model for the CAHs of (*1*) DD-carboxypeptidases, (*2*) class C PBPs and (*3*) DD-endopeptidases

Table 5 95% percentile bootstrap confidence intervals for β_0, β_1, β_2 and β_3

Regression coefficient	95% confidence interval
β_0	(0.57, 0.67)
β_1	(−0.54, −0.18)
β_2	(−0.31, −0.014)
β_3	(−0.19, −0.035)

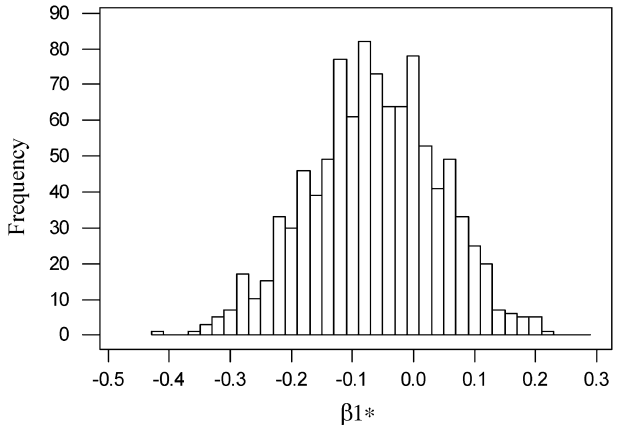

Fig. 4 Histogram of 1000 bootstrap replicates, β_1*, for β_1 from bootstrap regression in pairs for 34 randomized optimum sequences

Regression on randomized sequences

The histogram for the 1000 bootstrap replicates of the regression gradient coefficient, β_1, for the bootstrap regression in the pairs study of the 32 randomized sequence pairs is given in Fig. 4. To demonstrate the statistical power, the corresponding 95% percentile confidence intervals for both β_0 and β_1 are given in Table 6. From the confidence intervals, and by inspection of the histogram, we have no evidence from the 32 randomize pairs data that β_1 is significantly different from zero. The data can confidently be considered to be random. This is an important result and demonstrates that the linear relationships that have been identified are not an artefact of the mathematical relationship between $<\mu H>$ and $<H>$, given in Eq. (3). Furthermore, we are also able to compare the 95% confidence interval for β_0 (Table 6), derived from the randomization simulation study, with the point estimate for β_0, the corresponding increment, β_2, and the gradient, β_1, from the two-category regression analysis with no interaction (Table 2). From these it is readily seen that E($<\mu H>$) for both categorical groups is greater than the mean value, as represented by β_0 from the randomized sequence model, over the majority of the range of values for $<H>$. Whilst it should be expected that randomization would lead to an overall loss of amphiphilicity, these result provide strong evidence that the relationship between $<\mu H>$ and $<H>$ also reflects biological function and organization.

Table 6 95% percentile bootstrap confidence intervals for β_0 and β_1 for 32 pairs from randomized optimum windows

Regression coefficient	95% confidence interval
β_0	(0.26, 0.37)
β_1	(−0.28, 0.130)

Characterization of CAHs using the linear parallel model

Based on the statistical parallel linear model for the three categorical predictors, with no interactions present, it is possible to provide a methodology for putative classification of the C-terminal anchoring mechanisms employed by newly identified DD-peptidases. Initially, a simple regression of $<\mu H>$ onto $<H>$ for the new set of data must be conducted. Provided the regression coefficient for the gradient has a value within the range of the 95% confidence interval of that for the three-category parallel model, the methodology is applicable. Next, the $<\mu H>$, $<H>$ and corresponding additional indicator variable for these new data are introduced to increment the original design. This incremented model is now a four categorical predictor variable problem. This is fitted to the revised full data set and the value of the additional regression coefficient associated with the fourth categorical variable is noted. As a parallel model is being imposed, the value of this will quantify the difference in the average level of amphiphilicity present for the new DD-peptidases relative to the baseline, the DD-carboxypeptidases group. The anchoring mechanism for the new DD-peptidases is then predicted to be that associated with the group whose E($<\mu H>$) equation is the nearest. To test the efficacy of this methodology, a parametric bootstrap study (Efron and Tibshirani 1993) was conducted. In this, five ($<\mu H>$, $<H>$) pairs were randomly generated with $<H>$ values within the range for the three groups, and with $<\mu H>$ based on the empirical relationships which have been established. The errors were assumed to follow a normal distribution, with the mean square errors from the original regression being used as the common standard deviation. The four categorical variable model was then fitted to this augmented data set and the putative prediction noted. This process was repeated 1000 times for each of the three peptidase groups, representing the three fundamental anchoring mechanism categories. The process was also repeated for ten new, randomly generated pairs. The empirical probability of correct classification for the five and ten simulated peptidases is given in Table 7. The predictive power is particularly impressive, especially for the larger data sets, and we would also expect similar improvements when the original data set is enlarged with the future identification of additional members of the classes. Given the increasing numbers of CAHs being identified, and the levels of homology being seen, this would provide a useful tool for directing future experimental investigations.

Table 7 Probability of successful classification of anchoring type for sample sizes $n = 5$ and $n = 10$: P_0 for DD-carboxypeptidase; P_1 for DD-endopeptidase; P_{11} for class C

	$n = 5$	$n = 10$
P_0	0.875	0.950
P_1	0.705	0.773
P_{11}	0.574	0.691

Discussion

It appears to be generally accepted that DD-peptidases interact with the membrane via CAHs (Ghuysen 1994, 1997). DD-peptidases fall into three subclasses and it has been suggested that the CAHs of each of these subclasses may use functionally different mechanisms of membrane interaction (Phoenix and Harris 1998, 2002; Phoenix et al. 2002). However, experimental support for these observations is limited (Waxman and Strominger 1981a, 1981b, 1981c; Buchanan and Ling 1992; Pedersen et al. 1998; Nelson and Young 2001; Harris et al. 2002; Nelson et al. 2002; Phoenix et al. 2002).

Here, the C-terminal sequences of a number of DD-carboxypeptidases, class C PBPs and DD-endopeptidases were analysed. Inspection of the corresponding $<\mu H>$ and $<H>$ for the CAHs of each of these peptidase groups suggested that a negative parallel linear relationship for the expected values of $<\mu H>$, $E(<\mu H>)$, as a function of $<H>$ would provide a good model for these data (Fig. 3). Regression analysis and bootstrap regression resampling confirmed this suggestion, with further randomization simulation studies clearly demonstrating that these relationships were not artefactual. Rather, these relationships were a direct consequence of the sequential ordering of the residues forming these CAHs and, thus, their amphiphilicity. These observations imply biological function and clearly show that the CAHs of the individual peptidase groups analysed possess residue arrangements with similar amphiphilic properties. This would appear to support the view that CAHs are generally used by each of these peptidase groups for membrane interaction.

The linear relationships shown in Fig. 3 are parallel with a common gradient, $\beta_1 = -0.337$. In contrast, β_1 for the α-helices of transmembrane proteins is low, owing to the predominantly hydrophobic nature of their membrane interactions (Daman et al. 2001), which imposes a severe constraint, and no corresponding relationship has been established for the α-helices of soluble proteins (Phoenix et al. 1998). This shows that the membrane interaction of these amphiphilic α-helices has a common characteristic in that β_1 lies within a limited negative range. However, since these interactions lead to differing biological outcomes, with the only common factor appearing to be initial association with the membrane, we speculate that restrictions on the value of β_1 reflect the fact that a consistent balance between amphiphilicity and hydrophobicity is required to accommodate these initial associations. Furthermore, the different intercept values observed indicate that appropriate "barrier" levels of amphiphilicity are required to initiate and stabilize membrane interactions for the various anchoring mechanisms used by these CAHs.

Inspection of Fig. 3 reveals that, for a given $<H>$, the corresponding $E(<\mu H>)$ for CAHs follows the rank order: DD-carboxypeptidases > class C PBPs > DD-endopeptidases. To ascend this hierarchy, the CAHs of successive peptidase groups must optimize their residue arrangements to increase average levels of amphiphilicity, for each level of average hydrophobicity within the observed range. The relative numbers of strongly charged and polar residues possessed by these groups of CAHs are comparable, but the number of strongly hydrophobic residues follows the same rank order as that for $E(<\mu H>)$ identified above (Figs. 1, 3). This correlation suggests that hydrophobicity is the major driving force behind the membrane binding of DD-carboxypeptidase CAHs, and is in accord with the results of biochemical studies on *E. coli* PBP5 and PBP6. The CAHs of these DD-carboxypeptidases possess well-defined strongly hydrophobic apolar faces (Phoenix et al. 2002) and each has been shown by spectroscopic studies (Harris et al. 2002) and monolayer investigations (Harris et al. 1998) to exhibit high levels of membrane core penetration. Moreover, consistent with the high amphiphilicity of these CAHs (Phoenix et al. 2002), these same studies showed membrane penetration to be stabilized by electrostatic interactions between the charged residues of the CAHs and the membrane lipid headgroup region. Most recently, it has been suggested that the structural amphiphilicity and high levels of membrane penetration exhibited by these CAHs are necessary for membrane anchoring function. The recently solved structure of PBP5 (Davies et al. 2001) showed the protein's C-terminal anchoring region to have a surface location, diametrically opposed to the active site, thus maximizing access of this site to cell wall substrate. However, in the absence of high levels of C-terminal membrane anchoring, PBP5 rapidly diffused through the *E. coli* periplasm and caused cell death through unregulated cell wall lysis (Phoenix and Harris 1998; Nelson and Young 2001; Nelson et al. 2002).

The CAHs of class C PBPs and DD-endopeptidases are lower in the hierarchy of $E(<\mu H>)$ shown in Fig. 3 and their residue arrangements include relatively higher numbers of residues with lower hydrophobicity (Fig. 1). These arrangements give rise to the same average levels of hydrophobicity as the CAHs of DD-carboxypeptidases but lower average levels of amphiphilicity (Fig. 3). In contrast to the CAHs of DD-carboxypeptidases, those of class C PBPs and DD-endopeptidases show boundaries between their polar and non-polar faces, which are diffuse and give no clear segregation of hydrophobic and hydrophilic residues (Phoenix et al.

2002). This indicates that these latter CAHs achieve lower average levels of amphiphilicity through the use of residue arrangements, which are optimized to promote oppositions of hydrophobic and hydrophilic residues but with lower resolution. Implications from this are that electrostatic stabilization may play a greater role in the membrane interactions of CAHs from class C PBPs and DD-endopeptidases, and that these CAHs may show lower levels of membrane penetration than DD-carboxypeptidases. This proposition is consistent with the results of studies on the membrane interactions of the *E. coli* proteins PBP7 and PBP4, respectively the best-characterized members of the class C PBPs and DD-endopeptidases (Romeis and Holtje 1994; Henderson et al. 1995; Phoenix and Harris 1998). Biophysical studies on both peptidases (Henderson et al. 1995; Harris et al. 1997) and the PBP4 C-terminal α-helix (Harris et al. 1997, 2002) have shown their strength of membrane interaction to be lower than that of *E. coli* DD-carboxypeptidases and to be governed by electrostatic interactions, with only minor contributions from hydrophobic forces. Consistent with this observation, there are clear differences between the membrane associations of PBP4 and PBP7. PBP4 binds to a specific membrane bound receptor with the protein's C-terminal membrane anchoring stabilizing this binding (Harris et al. 1997). In contrast, the binding of PBP7 to the membrane is readily disrupted by ionic perturbants with a soluble form of the protein known (Romeis and Holtje, 1994).

In conclusion, each group of CAHs in our data set showed a negative linear relationship between $E(<\mu H>)$ and the corresponding $<H>$. Analysis of these relationships clearly suggested that the CAHs of each peptidase group in our data set optimized their structural amphiphilicity to fulfil the membrane

Table 8 DD-carboxypeptidases analysed

Organism	Protein	C-terminal sequence	$<\mu H>$	$<H>$
Bascillus stearothermophilus	PBP5	KAWFVLSMRAVGGLFVDLWTSV	0.47	0.39
Bascillus subtilis	PBP5	VLTMRSIGGFFAGIWGSIVDTVTGWF	0.52	0.47
Bascillus subtilis	PBP5*	KPKKQFFETFKSIFLNAAGGAKWSI	0.66*	0.12*
Bascillus subtilis	P38422	AKEDMKKAGFISFLKRTMGDWTKFK	0.61	−0.21
Streptococcus pneumoniae	PBP3	DKKIEKAFFLKVWWNQFVRFVNEKL	0.73	0.04
Streptococcus lactis	Q9CDF8	AKKTVPRSSAPKVFWNHFVNFVNEKL	0.58	0.28
Staphylococcus aureus	P72355	GGTCLVAGLALIVHMIINRLFRKRK	0.75	−0.09
Mycobacterium tuberculosis	Q10828	TYWDQAATLFDWGFALNPQASVGSL	0.40	0.16
Yersinia pestis	PBP5	NEVKEGGFFSRMVDYIKLMFHRWFG	0.58	0.10
Escherichia coli	PBP5	QEIPEGNFFGKIIDYIKLMFHHWFG	0.66*	0.23*
Escherichia coli	PBP6	EGGFFGRVWDFVMMKFHQWFGSWFS	0.51	0.41
Escherichia coli	PBP6b	PLVTLESVGEGSMFSRLSDYFHHKA	0.58	−0.02
Haemophilusinfluenza	PBP5	NDVGEAGIFGKLWDWLVLTVKGLFS	0.45	0.27
Salmonella typhimurium	PBP6b	PLVTLESGVKGGMFSRLSDYFQHKA	0.60	−0.06
Salmonella typhimurium	PBP6a	EGGFFSRMWDFVLMKLHQWFGSWFS	0.51	0.14
Salmonella typhimurium	PBP5	QEIPEGNFFGKIIDYIKLMFHHWFG	0.66	0.23
Pasteurella multocida	DacA	QEVQEGGIFGKAWDWLVLTVKSLFD	0.42	0.23
Streptococcus facealis	Q9EXN3	ASMRTIFLSAAVLSFLCAVGLFLLF	0.30	0.44

Table 9 DD-endopeptidases analysed

Organism	Protein	C-terminal sequence	$<\mu H>$	$<H>$
Escherichia coli	PBP7	PAAALSYKKQKAAQMAAAGQTAQND	0.25	−0.02
Vibrio cholera	Q9KL77	LSSEAKQYKKQRSQEQIAKISDYKS	0.50	−0.21
Haemophilus influenza	PBP7	VAGEAQYEDGYDEVGFNTLIQKLSK	0.62	0.03
Salmonella typhimurium	PBP7	SYKKQKAAQMAAASASAGAQTAQND	0.31	0.09

Table 10 Class C PBPs analysed

Organism	Protein	C-terminal sequence	$<\mu H>$	$<H>$
Salmonella typhimurium	PBP4	ADQRNRRIPLVRFESRLYKDIYQNN	0.75*	−0.31*
Escherichia coli	PBP4	ADQRNRRIPLVRFESRLYKDIYQNN	0.75*	−0.31*
Actinomadura sp (strain R39)	P39045	PEGARMMRGPVQGSGELECSWVQAC	0.60	−0.24
Bascillus subtilis	PBP4a	LLNGLIDEEDGKDIEDQIAVILANQ	0.51	−0.37
Synechocystis sp (strainPCC6803)	PBP4	LREAMKQMVLWTAQVEKCQPSDQGR	0.47	0.03
Mycobacterium tuberculosis	PBP4	NEAGPNGRNAMDALATKLWFCGCTT	0.44	−0.09
Neisseria gonorrhoeae	Q85665	FVAKNIISGGDGWLDAKLMCKERRA	0.46	0.23
Synechocystis sp (strain PCC6803)	P74173	LNAGPMDKVLLGDPARNLTPPPSES	0.51	−0.06

anchoring function of that group. The linear relationships describing the groups of CAHs analysed showed identical gradients, suggesting a common balance between amphiphilicity and hydrophobicity to be necessary for the membrane interactions of these amphiphilic structures. Our analyses clearly demonstrate that this balance is not required for membrane penetration by these CAHs and we have suggested that it may be required to stabilize their initial membrane associations. Moreover, when taken with the results of other authors (Picot and Garavito 1994; Wendt et al. 1997; Johnson and Cornell 1999; Harris et al. 2000; Phoenix and Harris 2002; Phoenix et al. 2002; Tossi and Sandri 2002), we postulate that such a requirement will be common to a range of membrane interactive amphiphilic α-helices, and may therefore be a key determinant of their ability to interact with membranes. Finally, we have introduced a methodology based on the parallel linear model, which enables accurate classification of the likely C-terminal anchoring mechanism used by newly identified DD-peptidase classes.

Appendix

The data set used in this study included the DD-carboxypeptideases (Table 8), DD-endopeptidases (Table 9) and class C PBPs (Table 10). In each case the source organism and recognized protein name are given, but where no such name exists, the Swissprot accession code is given. Also shown for each protein are the C-terminal 25 amino acid residues, the most amphiphilic α-helical segment identified within this sequence by hydrophobic moment analysis (emboldened residues), and the corresponding computed values of $<\mu H>$ and $<H>$ for this segment. A pair of segments with 100% homology and identical $<\mu H>$ and $<H>$ were identified in the DD-carboxypeptidase group and class C PBPs, respectively. These led to degenerate data points in subsequent analyses and are indicated by the superscript *.

References

Buchanan CE, Ling ML (1992) Isolation and sequence analysis of *dacB*, which encodes a sporulation-specific penicillin-binding protein in *Bacillus subtilis*. J Bacteriol 174:1717–1725

Cleveland WS (1979) Robust locally-weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829–836

Daman A, Wallace J, Phoenix D, Harris F (2001) Hydrophobic moment and hydrophobicity distributional properties and classification for membrane interacting transmembrane proteins. Proc 4th Eur Symp Protein Soc 10:141

Davies C, White SW, Nicholas R A (2001) Crystal structure of a deacylation-defective mutant of penicillin-binding protein 5 at 2.3-Å resolution. J Biol Chem 276:616–623

Draper NR, Smith H (1998) Applied regression analysis, 3rd edn. Wiley, New York

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, New York

Eisenberg D, Weiss RM, Terwilliger TC, Wilcox W (1982a) Hydrophobic moment and protein structure. Faraday Symp Chem Soc 17:109–120

Eisenberg D, Weiss RM, Terwilliger TC (1982b) The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature 299:371–374

Eisenberg D, Wesson M, Wilcox W (1984a) In: Fasman G (ed) Prediction of protein structure and principles of protein conformation. Plenum, New York, pp 635–646

Eisenberg D, Schwarz E, Komaromy M, Wall R (1984b) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 179:125–142

Eisenberg D, Wilcox W, McLachian A (1986) Hydrophobicity and amphiphilicity in protein structure. J Cell Biochem 3:11–17

Freedman DA (1981) Bootstrapping regression models. Ann Stat 9:1218–1228

Ghuysen JM (1994) Molecular structures of penicillin-binding proteins and β-lactamases. Trends Microbiol 3:372–380

Ghuysen JM (1997) Penicillin-binding proteins. Wall peptidoglycan assembly and resistance to penicillin: facts, doubts and hopes. Int J Antimicrob Agents 8:45–60

Gittins RG, Phoenix DA, Pratt JM (1993) Multiple mechanisms of membrane anchoring of *Escherichia coli* penicillin-binding proteins. FEMS Microbiol Rev 13:1–12

Granier B, Duez C, Lepage S, Englebert S, Dusart J, Dideberg O, Van Beeumen J, Frere J-M, Ghuysen J-M (1992) Primary and predicted secondary structures of the Actinomadura R39 extracellular DD-peptidase, a penicillin-binding protein (PBP) related to the *Escherichia coli* PBP4. Biochem J 282:781–788

Harris F (1998) PhD thesis. University of Central Lancashire, Preston, UK

Harris F, Phoenix DA (1997) Membrane binding of *Escherichia coli* penicillin-binding protein 4 is predominantly electrostatic in nature and occurs at a specific binding site. Protein Pept Lett 5:63–66

Harris F, Phoenix DA (1998) An investigation into the lipid interactions of peptides corresponding to the C-terminal anchoring domains of *Escherichia coli* penicillin-binding proteins 4, 5 and 6. Biochim Biophys Acta 1415:10–22

Harris F, Wallace J, Phoenix DA (2000) Use of hydrophobic moment plot methodology to aid the identification of oblique orientated α-helices. Mol Membr Biol 17:201–207

Harris F, Brandenberg K, Phoenix DA, Seydel U (2002) Investigations into the mechanisms used by the C-terminal anchors of *Escherichia coli* penicillin-binding proteins 4, 5, 6 and 6b for membrane interaction. Eur J Biochem 269:5821–5829

Henderson TA, Templin M, Young KD (1995) Identification and cloning of the gene encoding penicillin-binding protein 7 of *Escherichia coli*. J Bacteriol 177:2074–2079

Johnson JE, Cornell R (1999) Amphitropic proteins: regulation by reversible membrane interactions (review). Mol Membr Biol 16:217–235

Manly FJ (1997) Randomisation, bootstrap and Monte Carlo methods in biology, 2nd edn. Chapman & Hall, London

Nelson DE, Young KD (2001) Contributions of PBP 5 and DD-carboxypeptidase penicillin binding proteins to maintenance of cell shape in *Escherichia coli*. J Bacteriol 183:3055–3064

Nelson DE, Ghosh AS, Paulson A, Young KD (2002) Contribution of membrane-binding and enzymatic domains of penicillin binding protein 5 to maintenance of uniform cellular morphology of *Escherichia coli*. J Bacteriol 184:3630–3639

Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) Applied linear statistical models, 4th edn. Irwin, Chicago

Pedersen LB, Murray T, Popham DL, Setlow P (1998) Characterization of *dacC*, which encodes a new low-molecular-weight penicillin-binding protein in *Bacillus subtilis*. J Bacteriol 174:4967–4973

Pewsey AR, Phoenix DA, Roberts MG (1996). Monte Carlo analysis of potential C-terminal membrane interactive α-helices. Protein Pept Lett 3:185–192

Phoenix DA (1990) Investigation into structural features of the *Escherichia coli* penicillin-binding protein 5 C-terminal anchor. Biochem Soc Trans 18:948–949

Phoenix DA, Harris F (1998) Amphiphilic α-helices and lipid interactions. In: Phoenix DA (ed) Protein targeting and translocation. Portland Press, London, pp 19–36

Phoenix DA, Harris F (2002) The hydrophobic moment and its use in the classification of amphiphilic structures (review). Mol Membr Biol 19:1–10

Phoenix DA, Wallace J (2000) Analysis of the membrane interactive potential of the *Escherichia coli* PBP6b C-terminus. Protein Pept Lett 7:99–104

Phoenix DA, Stanworth A, Harris F (1998) The hydrophobic moment plot and its efficacy in the prediction and classification of membrane interactive proteins and peptides. Membr Cell Biol 12:101–110

Phoenix DA, Harris F, Daman OA, Wallace J (2002) The prediction of amphiphilic α-helices. Curr Protein Pept Sci 3:201–221

Picot D, Garavito RM (1994) Prostaglandin H synthase: implications for membrane structure. FEBS Lett 346:21–25

Roberts MG, Phoenix DA, Pewsey AR (1997) An algorithm for the detection of surface-active α-helices with the potential to anchor proteins at the membrane interface. CABIOS 13:99–106

Romeis T, Holtje J-V (1994) Penicillin-binding protein 7/8 of *Escherichia coli* is a DD-endopeptidase. Eur J Biochem 224:597–604

Segrest JP, De Loof H, Dohlman JG, Brouillette CG, Anantharamaiah GM (1990) Amphipathic helix motif: classes and properties. Proteins Struct Funct Genet 8:103–117

Segrest JP, Venkatachalapathi YV, Srinivas SK, Gupta KB, De Loof H, Anatharamaiah GM (1992) Role of basic amino acid residues in the amphipathic helix: the snorkel hypothesis. In: Balaram P, Ramaseshan S (eds) Molecular conformation and biological interactions. Indian Academy of Sciences, Bangalore, India, pp 597–635

Song J, Xie G, Elf PK, Young KD, Jensen RA (1998) Comparative analysis of *Pseudomonas aeruginosa* penicillin-binding protein 7 in the context of its membership in the family of low-molecular-mass PBP's. Microbiology 144:975–983

Tossi A, Sandri L (2002) Molecular diversity in gene-encoded, cationic antimicrobial polypeptides. Curr Pharmaceut Des 8:743–761

Waxman DJ, Strominger JL (1981a) Limited proteolysis of the penicillin-sensitive D-alanine carboxypeptidase purified from *Bacillus subtilis* membranes. Active water-soluble fragments generated by cleavage of a COOH-terminal membrane anchor. J Biol Chem 256:2059–2066

Waxman DJ, Strominger JL (1981b) Primary structure of the COOH-terminal membranous segment of a penicillin-sensitive enzyme purified from two bacilli. J Biol Chem 256:2067–2077

Waxman DJ, Strominger JL (1981c) High-molecular-weight penicillin-binding proteins from membranes of bacilli. J Bacteriol 148:950–955

Wendt KU, Poralla K, Schulz GE (1997) Structure and function of a squalene cyclase. Science 277:1811–1815