The Journal of Experimental Biology 201, 245–258 (1998)
Printed in Great Britain © The Company of Biologists Limited 1998
JEB1143

MODELLING QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIPS BETWEEN ANIMAL BEHAVIOUR AND ENVIRONMENTAL SIGNAL MOLECULES

KENNETH A. BROWNE[1], MARIO N. TAMBURRI[2] AND RICHARD K. ZIMMER-FAUST[1,*]

[1]*Department of Biology, University of California, Box 951606, Los Angeles, CA 90095-1606, USA* and [2]*Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA*

*Author for correspondence (e-mail: zf@protos.lifsci.ucla.edu)

Accepted 29 October 1997: published on WWW 22 December 1997

Summary

Quantitative structure–activity relationships (QSARs) between the physicochemical properties of environmental signal molecules and animal behaviour have been determined. Past work has shown that oyster and barnacle larval settlement and mud crab abdominal pumping (for larval dispersal) are stimulated by small peptide cues. In all the peptides examined that were active at ecologically relevant concentrations, arginine or lysine was found at the carboxy terminus, but the amino acids found at preceding positions were highly variable. We used the multivariate partial least squares algorithm to relate composite properties for the hydrophilicity, size and charge of each amino acid and the sequence position to oyster, barnacle and crab behaviour patterns. From the information in these QSAR models, the apparent variability in amino acid sequences eliciting behavioural responses was explained in each case, and more potent peptide analogues are hypothesized on the basis of untested amino acid sequences. Remarkably, these peptide signals are all structurally related to the carboxy-terminal sequence of mammalian C5a anaphylatoxin, a potent white blood cell chemoattractant. Even more striking is the fact that these different animal species should rely on apparently similar environmental signal molecules when residing within a common habitat (southeastern US estuaries). Through the physicochemical properties of amino acids, the current QSAR models clearly differentiate between the optimal sequences for eliciting oyster, barnacle and mud crab behaviour. Thus, QSARs provide a novel and powerful method not only for relating the physicochemical properties of molecules to animal behaviour but also for differentiating responses to chemicals by individuals of different species.

Key words: chemical cue, peptide, C terminus, quantitative structure–activity relationship (QSAR), partial least squares (PLS), oyster, *Crassostrea virginica*, settlement, anaphylatoxin.

Introduction

Pheromones and hormones are celebrated examples of chemical signal molecules. One strategy for producing chemical signals is to biosynthesize a novel substrate for each cue needed, as with anthopleurine, a sea anemone alarm pheromone (Howe and Sheikh, 1975), or adrenaline, the mammalian ‘fight-or-flight’ hormone (Baulieu and Kelly, 1990). A more cost-effective method is to use a polymeric system with different repeating units. Peptides are a well-known class of polymer signal molecules and they are a logical choice within organisms and aquatic environments for several reasons. First, because of the charged nature of the terminal primary amine and carboxylic acid groups at neutral pH, peptides are water-soluble and are not volatile. Second, the machinery (enzymes), templates (DNA, through mRNA) and structural units (amino acids) for producing peptides are already available in every living organism (Lehninger *et al.* 1993). Third, using the 20 coded amino acids available in eukaryotic systems, an incredible variety of information can be presented in a short amino acid sequence: with as few as five amino acids, there are 20^5=3.2 million possible unmodified peptides. Finally, intra- and extracellular proteases can rapidly degrade peptides to their constituent amino acids, thus terminating signal initiation (but not necessarily propagation; Hughes, 1978). Thus, peptides are excellently suited as intercellular and organismal signal molecules in aqueous environments.

Two well-known examples of peptide cues in mammalian systems are peptide hormones and anaphylatoxins. Small hormones such as vasopressin and oxytocin produce opposing responses (antidiuretic *versus* diuretic effects, respectively, among others) even though seven of their nine residues are identical (König, 1993). The anaphylatoxins (C3a, C4a and C5a) are amino-terminal proteolytic cleavage products (74–77 residues) of complement proteins. Despite having a sequence homology of 30–39 %, including an essential carboxy-terminal (C-terminal) arginine, C3a, C4a and C5a each evoke specific and distinct host defensive and inflammatory responses (Moon *et al.* 1981; Hugli, 1981). Subtle differences in amino acid

sequence can result in compounds that induce very different activities within a single organism.

Peptides are emerging as important chemical cues in marine environments. Recent work with oyster larval settlement (Tamburri *et al.* 1992; Turner *et al.* 1994; Zimmer-Faust and Tamburri, 1994) and with larval dispersal by ovigerous crabs (Forward *et al.* 1987; Rittschof *et al.* 1989; Pettis *et al.* 1993) indicates that small peptides from conspecifics or the environment are inducing agents. In addition, chemical cues evoking barnacle larvae to settle and metamorphose are produced by juvenile and adult conspecifics (Crisp and Meadows, 1962; Raimondi, 1988; Crisp, 1990). Experimental analyses have led investigators to conclude that one, or more, water-insoluble glycoproteins could be the inducing agent (Larman and Gabbott, 1975; Larman *et al.* 1982; Larman, 1984), although settlement is evoked by a variety of sticky substances. These glycoprotein inducers, called 'arthropodins', are believed to be homologous with compounds described from insect and crustacean cuticles. Remarkably, there has been a prelimary report of a seawater-soluble form of arthropodin for *Balanus amphitrite* (Rittschof, 1985, 1993; Tegtmeyer and Rittschof, 1989). As in the case of oysters, the soluble barnacle larval settlement cue is thought to be a peptide. While these cues clearly belong to different size classes (approximately 20–40 amino acids for the barnacle factor and fewer than 10 amino acids for the oyster and crab factors), each may have an obligate basic amino acid at its C terminus (e.g. arginine or lysine). Synthetic peptide analogues have previously been tested for each of these cases (three for oyster larvae, 13 for barnacle larvae and 21 for ovigerous crabs) in order to help define the active centre. In each case, the synthetic analogues have narrow effective concentration ranges (generally 2–4 log units), are potent at nanomolar and lower doses and have an obligate basic amino acid at their C terminus. As in mammalian systems, the differences in information contained in peptides are utilized by marine animals and manifested as behavioural responses.

In the present paper, we describe the behaviour of oyster larvae in response to solutions containing one of 53 small peptides. A quantitative structure–activity relationship (QSAR) is developed from the physicochemical properties of these molecules and their relative bioactivities to explain the chemical nature of the effect of amino acid side-chains on organismal behaviour. Likewise, behavioural data for barnacle larvae and ovigerous mud crabs in the presence of peptides are modelled using QSARs. Using these relationships, we extrapolate to potentially more active peptides without needing to synthesize and test all possible combinations (8420 possibilities for di-, tri- and tetrapeptides with arginine at the C terminus). Finally, the idealized sequences are compared with the mammalian complement system, and receptor site environments are proposed.

Materials and methods

Larval cultures

Seventeen-day-old oyster (*Crassostrea virginica* Gmelin) larvae, from Chesapeake Bay oysters spawned at the Virginia Institute of Marine Science, were shipped *via* overnight courier to our laboratory. They were maintained at 0.5–1.0 larvae ml^{-1} in a 1:1 mixture of oceanic and artificial seawater (ASW) medium at 25 ‰ salinity, pH 8.0, in an incubator at 25 °C with a 12 h:12 h L:D cycle (light on at 07:00 h). Before use, the culture medium was membrane-filtered (0.22 μm) and autoclaved for 15 min at 150 °C and 101.3 kPa. Cultures were actively aerated with air bubbled through sterile Pasteur pipettes. The culture medium was changed daily to preclude the build-up of pathogenic bacteria. Marine flagellates (*Isochrysis galbana* and *Pavlova lutheri*) were provided as food at 2.5×10^4 cells ml^{-1} once each day.

Oyster larvae are typically more than 250 μm long and have pigmented eyespots when they are competent to settle and metamorphose (Galtsoff, 1964; Coon *et al.* 1990). All experiments were begun within 6 h of 95 % of the larvae in the culture developing eyes, and experiments were run for 36 h thereafter. These conditions were met for larvae aged between 19 and 21 days post-fertilization. The mean length of larvae tested was 315±4 μm (mean ± S.E.M., N=100).

Settlement bioassays

Settlement is defined here as the potentially reversible attachment of the larval foot to a substratum. This is different from the irreversible developmental events of metamorphosis. The procedures outlined below are much as previously described (Tamburri *et al.* 1992; Zimmer-Faust and Tamburri, 1994). Between 1 and 2 h before testing, the larvae were gently filtered from the culture medium, rinsed with sterile ASW, then placed in a separate container of ASW at a density of 0.5 larvae ml^{-1}. This procedure removed the larvae from their microalgal food, the exudates of which influence locomotory behaviour (Zimmer-Faust *et al.* 1997). Sixty larvae in 5 ml of ASW were transferred into a cylindrical Plexiglas microcosm (30 mm in diameter and 42 mm high) containing 25 ml of a test or control solution. The solution containing the larvae was gently stirred and placed in a darkened room. Larval movements in the horizontal plane within 1 mm of the chamber bottom were monitored with an infrared-sensitive video camera (mounted beneath the microcosm) after the fluid had been allowed to come to rest for approximately 15 s. Illumination was provided by a light source (≥820 nm) oriented at 90 ° to the axis of the video camera. The field of view was 6.75 mm×6.70 mm, approximately 5 % of the total cross-sectional area of the microcosm.

Fifty-three peptides of low molecular mass (L-isomers) were individually bioassayed for their influence on oyster larval settlement. All peptides were purchased from Sigma Chemical Company, Research Plus, Inc. or BACHEM California, Inc. or were custom-synthesized at the UCLA Peptide Synthesis Core Facility. The effect of each compound was tested on larvae at 2 log unit concentration intervals from 10^{-6} to 10^{-18} mol l^{-1}. Between five and eight replicate trials were performed for each chemical and control solution in each experiment. In each trial, the number of larvae that settled in the video field was counted

during 3 min of exposure to a control or an experimental solution. The percentage of all larvae tested that settled per trial was calculated by assuming that the video field provided a representative sample of the entire microcosm. Our experimental design was constrained by the small size of the larvae. We therefore used the largest video field that allowed unmistakable observation of larval foot extension and substratum attachment.

The bioassays with all of the compounds were conducted over three consecutive summers with several batches of oyster larvae each summer. To account for potential batch-to-batch and summer-to-summer variability, all groups of trials included a positive control condition (glycyl-glycyl-L-arginine, GGR) as well as an ASW control. Since the absolute number of larvae settling may vary between groups of trials, the percentage settlement responses to the compounds studied were normalized between the ASW control at the low end (0) and GGR at the high end (1.0).

Quantitative structure–activity relationships (QSARs)

Relationships between the physicochemical properties of peptides and oyster (*Crassostrea virginica*) and barnacle (*Balanus amphitrite*) larval settlement and ovigerous crab (*Rhithropanopeus harrisii*) abdominal pumping were found using the partial least squares, or projection to latent structures, technique (PLS; Wold, 1982; Dunn *et al.* 1984; Geladi and Kowalski, 1986). With PLS, a model is constructed which connects the variation in chemical structure to the variation in biological activity through linear relationships of latent variables (LV). PLS models are calculated to simultaneously minimize residuals and yield optimally correlated latent variables (Geladi and Kowalski, 1986). Three advantages of PLS over other multivariate statistical algorithms lie in (1) its ability to use more descriptors and activities than data points (Dunn *et al.* 1984), (2) its decreased likelihood of generating chance correlations (Topliss and Edwards, 1979; Wold and Dunn, 1983) and (3) its strong predictive capabilities (Glen *et al.* 1989).

Univariate biological data (dependent variables) were based on the normalized activities of oyster larvae in solutions containing 10^{-8} mol l^{-1} peptide; this was the concentration at which most active peptides showed maximal response. Molar threshold concentrations (negative logarithmic transformations) were used as the dependent variables for barnacle larval settlement (Tegtmeyer and Rittschof, 1989) and crab abdominal pumping behaviour (Forward *et al.* 1987; Rittschof *et al.* 1989; Pettis *et al.* 1993). Only compounds whose sequences terminated in arginine (for oyster larvae), or for which response thresholds were less than micromolar concentrations (for barnacle larvae and ovigerous crabs), were used for the analyses.

For multivariate descriptor values (independent variables), we used the three sets of 'principal properties' for amino acids previously extracted by principal component analysis of a multiproperty matrix (Hellberg *et al.* 1987; 29 properties ranging from molecular mass, pK_a values and octanol/water partition coefficients to high-performance liquid chromatography retention times and nuclear magnetic resonance chemical shifts). These 'principal properties' tentatively correspond to side-chain hydrophilicity (z_1), size (z_2) and electronic properties (z_3). Two possible drawbacks of using amino acid descriptors to model peptides are that both neighbouring groups and peptide conformation can change the descriptor values. It has been found that, as long as the descriptors are derived in an internally consistent manner, these caveats introduce acceptably small errors for short peptides (Collantes and Dunn, 1995).

Compounds were scored for the physicochemical properties of the amino acid moieties and the positions of the amino acids in the peptides. The C terminus is considered the 'parent' position; thus, the peptide labels are decremented from the carboxy to the amino terminus (C-5, C-4, C-3, C-2, C-1, C) even though their sequence is shown from the amino to the carboxy terminus. Using PLS_Toolbox 1.5.2 (Eigenvector Technologies) interfaced with MATLAB 4.2c.1 (The MathWorks, Inc.), the scores from many different compounds with varying bioactivities were developed into a composite model which describes essential physicochemical characteristics at each amino acid position (loadings). When examining peptides that were less than the maximum length (six amino acids), physicochemical data for the remaining positions were filled with zeros (all components of the matrix must contain the same number of elements), indicating that there was no information in that portion of the sequence. To test the legitimacy of zero-filling, models for a subset of the oyster larval data (tripeptides) were developed with and without zero-filled amino acid positions; both models yielded exactly the same results. Since we had relatively few compounds available for modelling, leave-one-out cross-validation (LOO; Lorber and Kowalski, 1988) was used followed by a predictive sum of squares (PRESS) calculation to estimate prediction errors. Successful models had a maximum number of latent variables that corresponded to a minimum PRESS and a maximum percentage variance (above the noise) in biological activity captured by the model. Data points that had high Studentized residual ($\geqslant 2$ for data points significant at the 96 % confidence level) or leverage (a measure of the influence that a data point has on the model; should be $<3p/N$ where p is the number of latent variables and N is the number of data points) values were removed from the analysis (Lorber *et al.* 1988); there were never more than two or three of these calibration outliers.

Results

Qualitative examination of oyster larval responses to peptides

The settlement behaviour of competent oyster (*Crassostrea virginica*) larvae was evaluated in the presence of a wide range of concentrations of 53 short peptides. The percentage of larval settlers in response to each compound was normalized between 0 for the artificial seawater (ASW)

Table 1. *Oyster larval settlement responses to peptide solutions at 10^{-8} mol l^{-1}*

Experiment	Sequence[a]	Percentage settlement[b]	Normalized response
1	GGR	56.7±5.1***	1.000
	DGR	49.7±7.0**	0.833
	RGR	36.8±7.0*	0.533
	LGR	28.2±3.5	0.333
	ASW control	14.1±7.0	0
2	GGR	51.6±4.2***	1.000
	GGGR	44.9±6.7**	0.833
	GDR	43.3±4.2**	0.800
	GRR	31.1±3.6**	0.500
	GFR	23.7±5.5	0.320
	GGYR	16.7±3.6	0.150
	ASW control	10.6±5.1	0
3	GGR	50.6±4.4***	1.000
	GGK	19.0±5.1	0.175
	ASW control	12.3±4.5	0
4	GGR	59.9±6.1***	1.000
	TKPR	28.2±3.9	0.320
	TRKR	24.7±4.5	0.245
	LWMR	20.8±8.0	0.165
	HLGLAR	17.0±5.1	0.080
	KRTLRR	17.0±5.0	0.080
	ASW control	13.1±5.1	0
5	GGR	67.3±8.3***	1.000
	RSR	35.9±5.7	0.340
	RR	27.9±6.7	0.175
	KR	27.9±6.7	0.175
	ASW control	19.5±6.7	0

[a]Strings of one-letter codes for amide-linked amino acids (peptides); peptides are labelled from the amino to the carboxy terminus (left to right); A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; P, proline; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine. Peptides tested that gave no response above the ASW control level: GR, RG, GG, AR, RK, DR, ER, IR, LR, KK, MR, FR, TR, YR, GGH, GGL, GGD, GGG, RRR, GPR, RGG, RRRR, APGPR, RRRRR, DSDPR, LPPSR, YIGSR, VPDPR, YGGFLR, YGGFMRR, RPGFSPFR, GRGLSLSR, KRGHPGKR, GWMDFSPFR, GGGGGGGGGR. ASW, artificial sea water

[b]Values are means ± S.E.M., N=5–8. Statistical analysis was performed on the number of larvae that settled per trial in the video field. The significance of the settlement response was determined by applying a two-way analysis of variance (with compound and dosage as class variables) followed by pairwise comparisons between each compound at 10^{-8} mol l^{-1} and the ASW control, using Bonferroni's correction: *P<0.05; **P<0.01; ***P<0.001.

control and 1.0 for GGR, the most potent peptide tested. The effect of peptide length on oyster larval settlement is an important factor to consider. Of the 16 dipeptides tested, only RR and KR induced any settlement response at 10^{-8} mol l^{-1} (Table 1); these peptides were ineffective at higher and lower concentrations. The tripeptides, followed by the tetrapeptides, with an arginine at the C terminus, were the most potent oligomers studied; many of these compounds stimulated oyster larvae to settle when present at concentrations as low as 10^{-10} mol l^{-1}; no compounds tested induced significant settlement activity when present at lower concentrations (Table 1; Fig. 1). The pentapeptides tested had no effect on larval settlement, and only two of the hexapeptides were slightly active. Longer peptides were not active. The optimal peptide length within a homologous series is clearly illustrated by varying the number of glycine residues preceding the C-terminal arginine; only sequences containing three and four amino acids were active (Fig. 2E).

Most of the active peptides studied (18 of 53) displayed dose–response curves that increased to a maximum then decreased (bell-shaped curves); the maximum settlement-inducing response for the active peptides occurred at 10^{-8} mol l^{-1} (Fig. 1). TKPR is an exception, with maximal activity at 10^{-6} mol l^{-1}; the difference in activity from that at 10^{-8} mol l^{-1} is not significant. For a series of peptides varying only in their C terminus, GGR is by far the most active, with GGK generating approximately 18% as much response (Fig. 2A). Polar, neutral and negatively charged amino acids (histidine, glycine, leucine and aspartic acid) at the C terminus lead to inactive structures. A single positively charged peptidic arginine is critical for inducing settlement, two arginine residues incur some activity, but more eliminate the potency of peptides (Fig. 2B). At the C-2 position, small neutral, negatively and positively charged amino acids (glycine, aspartic acid, arginine and leucine) are all active (Fig. 2C). The C-1 position is similarly tolerant of glycine, phenylalanine, aspartic acid and arginine (Fig. 2D); only proline at this position eliminates activity. Considerable variability in lipophilicity, charge and size is allowed at all but the C terminus of bioactive peptides; this is not easily reconciled by visual inspection of the data.

Quantitative evaluation of oyster larval responses to peptides

To define the oyster larval settlement activity data better, a quantitative structure–activity relationship (QSAR) was developed. The partial least squares (PLS) technique was used to relate the biological activity to physicochemical descriptors for each of the amino acids at every position in the 17 different active peptides ending with a C-terminal arginine (Table 1; GGK was not included in the analysis because it was found to be a calibration outlier). The first two latent variables captured 95.0% of the variance in the biological activity (=85.9%+9.1%); subsequent latent variables do not contribute significantly to the model. The PLS loadings are shown for each sequence position and descriptor in Fig. 3A. The magnitudes of the loadings reflect the relative importance of a given descriptor at that position. For example, glycine is a very small amino acid so its size descriptor is large and negative (z_2); tryptophan is large and it has a large positive size descriptor (Fig. 4). Aspartic acid is negatively charged at

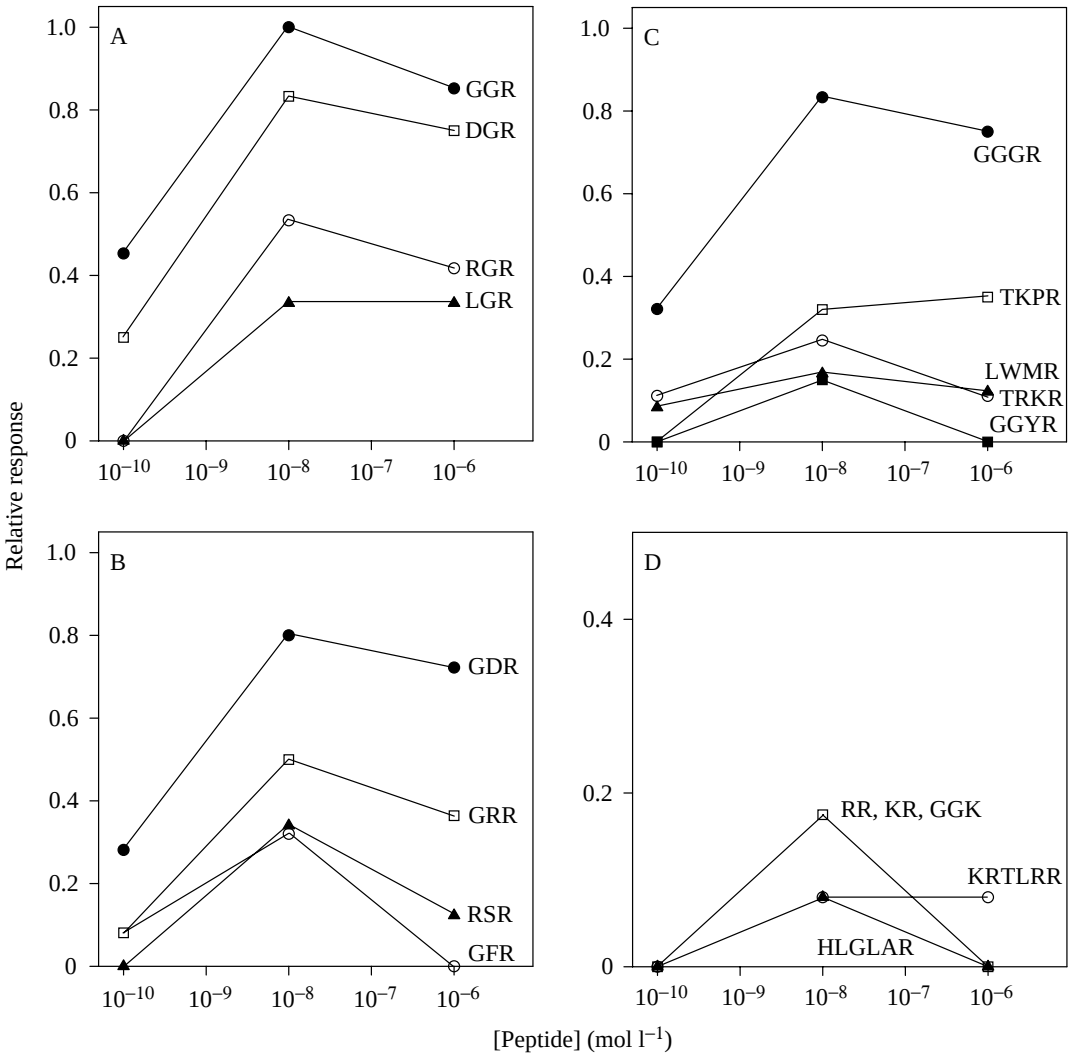

Fig. 1. The relationships between peptide concentrations (mol l^{-1}) and relative oyster larval settlement responses. Data are shown only for peptides at concentrations $\geqslant 10^{-10}$ mol l^{-1} since lower doses had no significant activity relative to sea water (control). Normalized responses are based on mean values from at least five trials with each compound and concentration. (See Table 1 for the percentage of larvae settled at a peptide concentration of 10^{-8} mol l^{-1}). (A) Tripeptides of the homologous series *X*-glycyl-L-arginine. GGR-normalized responses are representative of data from the five different sets of trials. (B) Tripeptides of the series glycyl-*X*-L-arginine (RSR is an exception). (C) Tetrapeptides of the series *X*-*X*-*X*-L-arginine. (D) Peptides giving very low responses. Note that the ordinate scale has been changed for clarity. The zero values correspond to the normalized control (ASW, 25‰, pH 8.0) responses.

neutral pH and it has a large positive charge descriptor (z_3) while arginine, which is positively charged, has a large negative charge descriptor. The neutral amino acids glycine and tryptophan have very small charge contributions. The loadings can now be interpreted. The C terminus is the most important amino acid position, and the loadings clearly define an arginine-type moiety; large hydrophilicity and size contributions (z_1 and z_2) and a large negative charge contribution (z_3, a very positively charged residue). The C-1 and C-2 positions have similar characteristics; small (negative z_2), with very little charge (near zero z_3) but still hydrophilic (positive z_1), much like glycine. The C-3 position is filled by relatively small (negative z_2) and neutral (small z_3) amino acids. The loadings at positions C-4 and C-5 have almost no size contributions (near zero z_2), which suggests that no amino acids are favoured at these positions. The loadings complement the qualitative analysis.

The quantitative information in the loadings can be used to check the predictive ability of the QSAR model. The predicted larval settlement activity data are back-calculated from the amino acid sequences of the tested peptides and the first two latent variables. The resulting predicted responses are plotted against the observed biological activity values; the data points cluster tightly about the optimal observed *versus* predicted line (Fig. 3B). GGR is both predicted and found to be the most active peptide. The right-hand portion of the plot is dominated by tripeptides followed by di- and tetrapeptides to the left, and finally hexapeptides. This is the same pattern suggested by the qualitative examination of these peptides (Figs 1, 2) and in the PLS loadings. Likewise, the peptides on the right of the plot

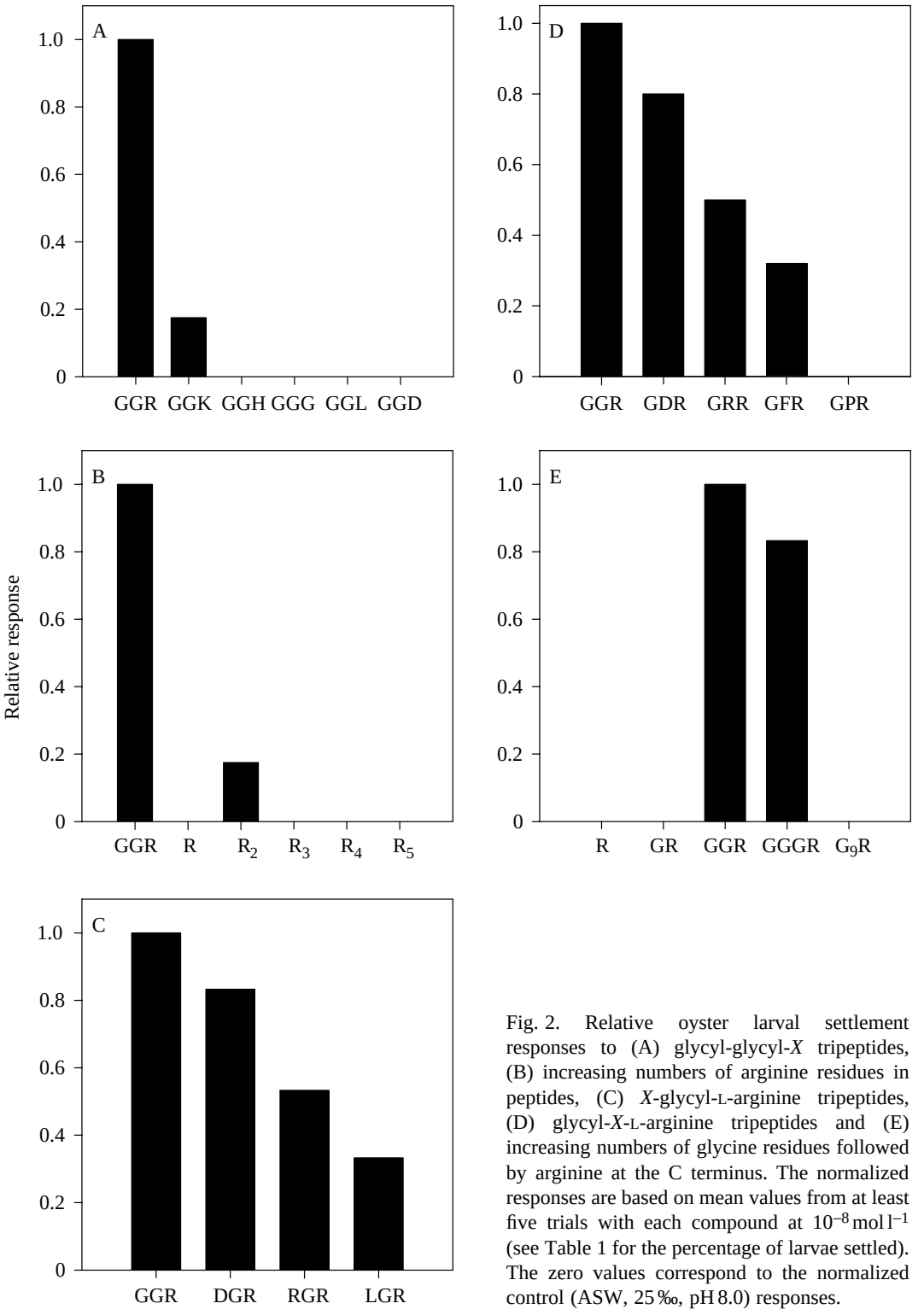

Fig. 2. Relative oyster larval settlement responses to (A) glycyl-glycyl-*X* tripeptides, (B) increasing numbers of arginine residues in peptides, (C) *X*-glycyl-L-arginine tripeptides, (D) glycyl-*X*-L-arginine tripeptides and (E) increasing numbers of glycine residues followed by arginine at the C terminus. The normalized responses are based on mean values from at least five trials with each compound at 10^{-8} mol l^{-1} (see Table 1 for the percentage of larvae settled). The zero values correspond to the normalized control (ASW, 25 ‰, pH 8.0) responses.

all contain at least one glycine residue in one of the positions preceding the C-terminal arginine. The current QSAR is clearly effective at modelling oyster larval settlement activity on the basis of the physicochemical properties of amino acids and their positions in peptides.

So far, the QSAR has only modelled settlement responses that were already known. QSARs are developed not only to explain known biological activity with respect to chemical structure but also to predict novel structures with increased activity that can be tested. This has been done through computer simulations with all possible dipeptides (20 combinations), tripeptides (400 combinations) and

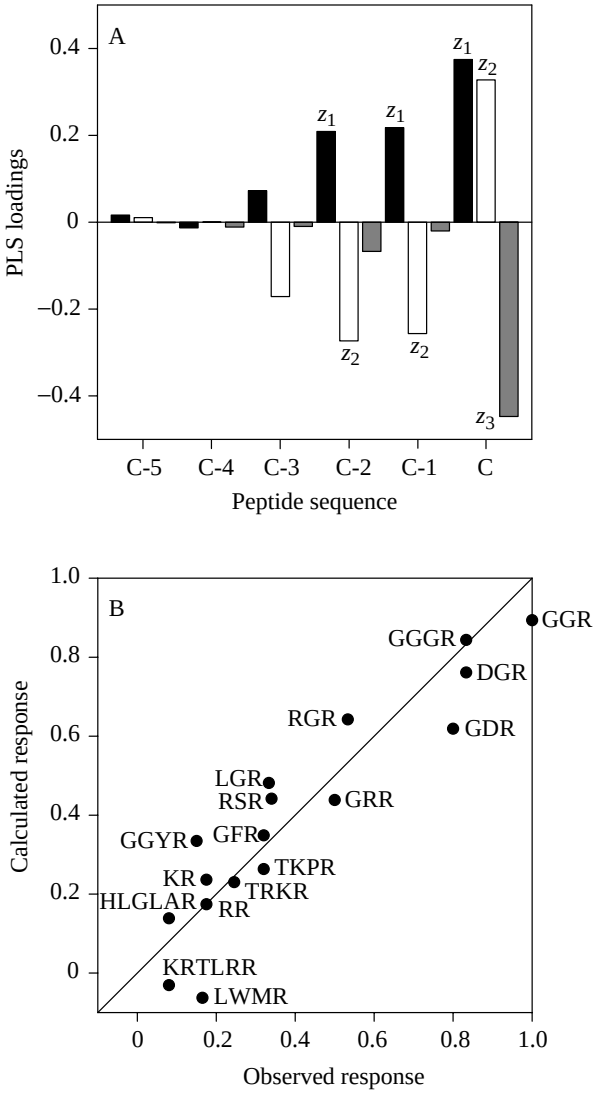

Fig. 3. Partial least squares (PLS) model data for oyster larval settlement activity in the presence of different peptides. This model captures 95.0 % of the variance in biological activity with two latent variables. (A) Plot of loadings *versus* physicochemical descriptors (z_1, z_2 and z_3) and sequence position for the first two model dimensions. The magnitudes of the loadings shown are composites of the percentages of the first and second loadings. (B) Plot of calculated *versus* observed normalized responses.

tetrapeptides (8000 combinations) containing arginine at their C terminus. All dipeptides were predicted to be less than 63 % as active as GGR. The most active tripeptide not already tested, GSR, was predicted to be approximately 77 % as active as GGR (Table 2). Sequences of the form *X*-GGR, where *X* is a lipophilic, neutral amino acid (W, Y, F, L, I, P, M and V), were the only peptides with calculated activities greater than that of GGR (0.893, as predicted by the QSAR model; Fig. 3B). Significantly, the addition of these amino acids to the amino terminus of GGR places the predicted peptides within the size range of the natural oyster larval settlement cue.

Table 2. *Predicted behavioural responses to untested peptides based on quantitative structure–activity relationships*

Animal	Sequence[a]	Predicted[b] response
Oyster larva	WGGR	0.948
	FGGR	0.934
	YGGR	0.920
	IGGR, LGGR, MGGR, PGGR	0.910
	VGGR	0.895
	RGGR	0.891
	GSR	0.766
Barnacle larva	IIR	11.4
	LIR	11.2
	VIR	11.2
	ILR	11.1
	LLR, VLR, IVR, FIR	10.9
	LVR, VVR	10.8
	IR	10.2
Mud crab	GIR	13.1
	GFR	12.9
	GWR	12.4
	GVR	12.2
	GIK	12.1
	GLK, GFK	11.8
	FR	10.3

[a]Amino acid codes are the same as in Table 1.

[b]Predicted settlement responses for oyster larvae are normalized between 0 and 1; predicted responses for barnacle larval settlement and mud crab abdominal pumping are –log(threshold concentration).

Only the top seven predicted values are shown.

Quantitative evaluation of barnacle larval responses to peptides

Barnacle (*Balanus amphitrite*) larval settlement has been preliminarily reported in response to waterborne cues released by adult conspecifics (Rittschof, 1985; Tegtmeyer and Rittschof, 1989), in addition to many other factors (for a review, see Pawlik, 1992). As in oysters, the waterborne barnacle larval settlement cue may be a peptide that contains either arginine or lysine at its C terminus but, unlike the oyster cue, this peptide seems to be large, in the size range 3000–5000 g mol^{-1}. To explore the specificity of peptide sequence to response, a series of di- and tripeptides with different combinations of amino acids were tested for their ability to induce barnacle larval settlement (Tegtmeyer and Rittschof, 1989). Because the behavioural threshold concentrations in the absence of statistical analysis were reported, these results must be considered with some caution. Only synthetic peptide analogues with arginine or lysine at the C terminus were shown to yield threshold responses, and even then not all lysine-terminal peptides enhanced settlement. The dose–response curves for active peptide analogues were bell-shaped over a 1000-fold concentration range, with discrete

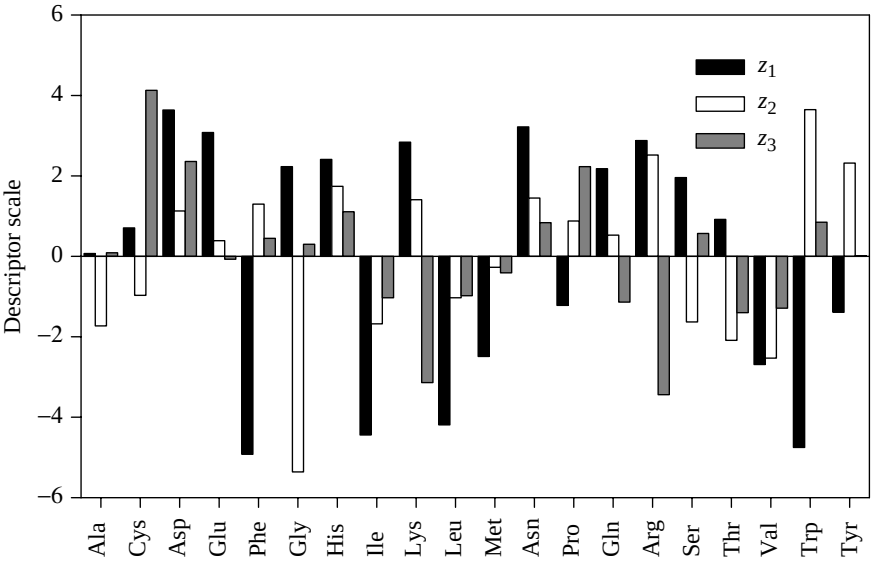

Fig. 4. Physicochemical descriptor scales used for the amino acids (Hellberg *et al.* 1987). The term z_1 refers to hydrophilicity, z_2 refers to size and z_3 refers to electronic properties. Note that for tyrosine the z_3 descriptor is too small to be seen.

doses yielding maximal responses. Qualitatively, barnacle larvae respond to synthetic peptides with a basic C terminus, a neutral or basic residue in the immediately preceding position, and a neutral amino acid or no residue at the C-2 position (Tegtmeyer and Rittschof, 1989).

A QSAR was attempted between the barnacle larval settlement responses (molar threshold concentrations) and the physicochemical properties of the six active peptides. This QSAR was not predictive of the observed responses in spite of the fact that 98.6 % of the variance in activity was captured by two latent variables (Fig. 5A), and it continued to fail with mean centring or autoscaling of the data. Individual peptides were systematically omitted from the model with no improvement. The QSAR only became predictive when the lysine-terminal pair of peptides (HK and GHK) was omitted from the model. Using the four peptides with arginine at the C terminus yielded a predictive QSAR model with two latent variables capturing 99.9 % of the biological activity (=94.0 %+5.9 %; Fig. 5B). Examination of the cumulative loadings from the first two model dimensions (Fig. 5C) indicates that a large, hydrophilic, positively charged amino acid (such as arginine) is optimal at the C terminus. The C-1 and C-2 positions are filled by small neutral residues, with C-2 being slightly hydrophobic. The quantitative descriptions (loadings) of the amino acid side-chains at each position reflect the predicted activities of the peptides as well as the qualitative examination.

The barnacle larval settlement QSAR can be used to predict additional peptide sequences with greater activity than those already tested. All dipeptides (20) and tripeptides (400) with C-terminal arginine sequences have been examined. One dipeptide (IR) was predicted to be approximately three times more potent than the most active tested compounds. Thirty-five tripeptides were predicted to be between three and 50 times more potent than LR and LGR, the most active peptides assayed (Table 2; Fig. 5B). Each of these tripeptides was of the form *X*-*Z*-R where *X* is neutral and lipophilic and *Z* is L, I, V or F. These are the types of residue that one would intuitively predict from the PLS loadings.

Quantitative evaluation of ovigerous crab responses to peptides

Ovigerous mud crabs (*Rhithropanopeus harrisii*) respond to water-soluble cues released from brooded embryos by flexing their abdomens repeatedly (Forward and Lohmann, 1983). This pumping action both liberates and disperses newly hatched larvae from their egg capsules. Proteolytic and size-fractionation analyses of sea water incubated with recently hatched crab embryos indicate that the cue is a peptide of less than 500 g mol^{-1} (Rittschof *et al.* 1985).

The ability of amino acids and short peptides to induce abdominal pumping by mud crabs was examined (Forward *et al.* 1987; Rittschof *et al.* 1989; Pettis *et al.* 1993). The only peptides that were active had arginine or lysine at the C terminus. Most of these compounds generated dose-dependent responses similar to those found with oyster and barnacle larvae; when viewed over several orders of magnitude in concentration, distinct maximum response levels were flanked by decreased responses at higher and lower concentrations. Qualitatively, peptides with a neutral aliphatic amino acid followed by a C-terminal arginine were the most active, followed by peptides with lysine at the C terminus. Thus, the compound(s) responsible for inducing mud crabs to disperse their larvae appears to be closely related to those substances inducing oyster and barnacle larvae to settle.

The physicochemical properties of 11 synthetic analogues, ranging from one to five amino acids in length, were related to the molar threshold response concentrations required for ovigerous mud crabs to begin abdominal pumping. Three peptides, AR, IGR and LGR, were not used because they had high Studentized residuals. Also, the three ‘superpotent’ tripeptides GIR, GFR and GMR were excluded because of

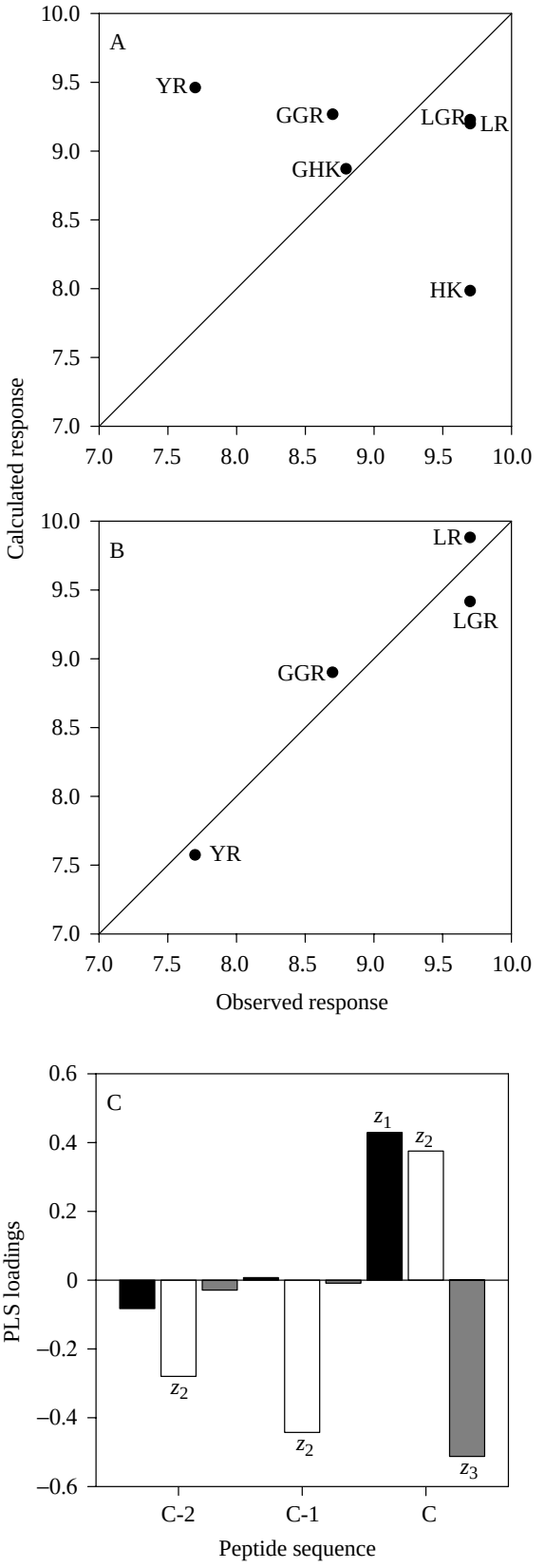

Fig. 5. Partial least squares (PLS) model data for barnacle larval settlement activity in the presence of different peptides. (A) Plot of calculated *versus* observed molar threshold concentrations (negative logarithmic transformation) for all six active peptides. This model accounts for 98.6 % of the variance in biological activity with two latent variables. (B) Plot of calculated *versus* observed molar threshold concentrations (negative logarithmic transformation) for arginine-containing peptides. This model captures 99.9 % of the variance in biological activity with two latent variables. (C) Plot of loadings *versus* physicochemical descriptors (z_1, z_2 and z_3) and sequence position for the first two model dimensions with arginine-containing peptides. The magnitudes of the loadings shown are composites of the percentages of the first and second loadings.

their unprecedentedly low threshold concentrations (10^{-17} to 10^{-21} mol l^{-1}) that require further confirmation. The remaining compounds were divided into two classes, those with arginine and those with lysine at their C terminus. Successful QSARs were calculated for each (data not shown). Examination of both QSARs suggested that the lysine- and arginine-terminal peptides were extremes of a continuum. A new predictive QSAR was developed, spanning seven orders of magnitude in threshold concentration, which included the peptides from both data sets. Two latent variables captured 99.7 % of the variance in biological activity (Fig. 6A). The lysine-terminal peptides occupied the low-activity positions in the QSAR (lower left-hand portion of the plot), followed by the arginine-terminal peptides at higher activities (right-hand side of the plot). The PLS loadings from the first two model dimensions reflect the qualitative and quantitative evaluations of peptide activities; a large hydrophilic, positively charged residue is at the C-terminal position, and relatively small uncharged amino acids are at the C-1 and C-2 positions, with C-1 also being hydrophobic (Fig. 6B).

The composite QSAR has been used to model the abdominal pumping responses of untested peptides. The effects of 40 dipeptides and 800 tripeptides containing arginine and lysine at their C terminus were computer-simulated. No dipeptides were predicted to have activities greater than IR, the most active dipeptide tested (Rittschof *et al.* 1989). Only one tripeptide not already tested, GIR, was predicted to have more activity than GLR, the most active compound in the QSAR (Table 2). There were several tripeptides with predicted threshold activities close to or below 10^{-12} mol l^{-1}, including three with lysine at their C terminus (GFR, GWR, GVR, GIK, GFK and GLK). They are all of the form G-*X*-*B*, where *X* is a hydrophobic residue and *B* is a basic residue. All of these sequences are within the purported size range of the crab cue (<500 g mol^{-1}). As with the oyster and barnacle larval settlement analyses, the predicted peptides for ovigerous mud crabs mirror the qualitative analysis and the information in the PLS loadings.

Discussion

The influence of peptides on oyster, barnacle and mud crab behaviour

Fifty-three peptides were tested for their effects on competent oyster larval settlement. Of these, only 18 small

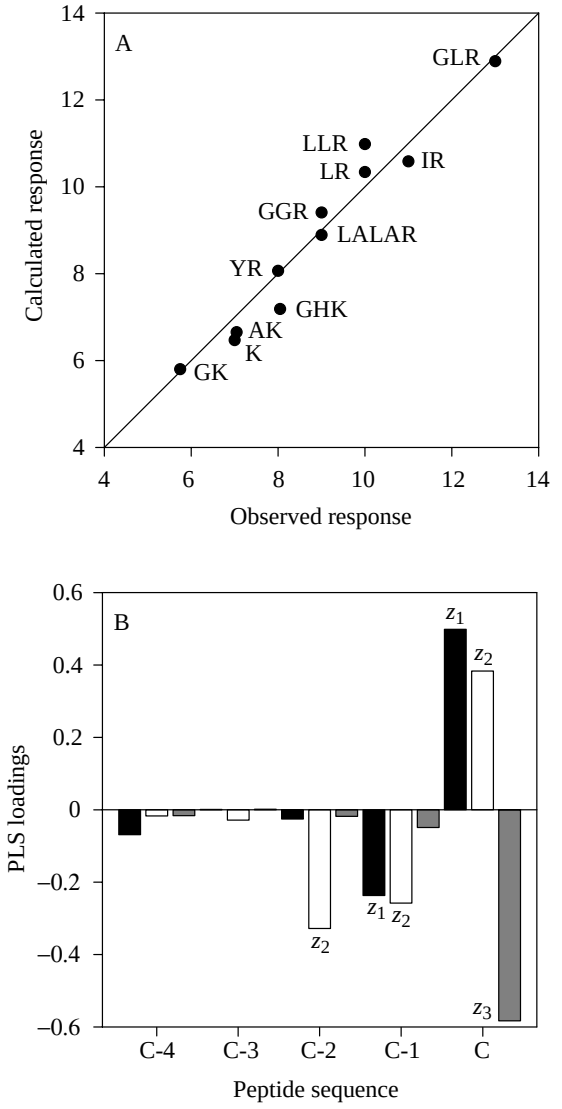

Fig. 6. Partial least squares (PLS) model data for ovigerous crab abdominal pumping activity in the presence of different peptides. This model captures 99.7 % of the variance in biological activity with two latent variables. (A) Plot of calculated *versus* observed molar threshold concentrations (negative logarithmic transformation). (B) Plot of loadings *versus* physicochemical descriptors (z_1, z_2 and z_3) and sequence position for the first two model dimensions. The magnitudes of the loadings shown are composites of the percentages of the first and second loadings.

peptides (2–6 residues) induced settlement responses at submicromolar concentrations such as may be found in natural estuarine environments (Zimmer-Faust *et al.* 1997). Using threshold responses to characterize oyster larval settlement results in most active peptides having indistinguishable response values (approximately 10^{-10} mol l^{-1}). Instead, activities are classified according to the percentage response at a constant peptide concentration (10^{-8} mol l^{-1}, the maximal activity for most peptides). Glycyl-glycyl-L-arginine (GGR) was the most potent compound examined. Replacing the C-terminal amino acid of GGR with lysine, glycine, histidine or leucine left little if any activity; this is consistent with the loss of activity in oyster-conditioned water after treatment with carboxypeptidase B or arginase, enzymes that selectively cleave lysine or arginine from the C terminus of a peptide or that hydrolyze L-arginine to L-ornithine and urea, respectively (Zimmer-Faust and Tamburri, 1994). Inverting the sequence (RGG) or emphasizing arginine (RRR) stifled all activity, indicating that the presence of arginine in the molecule is not sufficient to induce settlement. While adding a glycine to the amino terminus (GGGR) induced only a slight decrease in activity, removing a glycyl monomer (GR) led to an inactive compound. In tri- and tetrapeptides, a glycine residue was important at the C-1 or C-2 position. Peptides with five or more amino acids had little, if any, activity even when they contained an arginine at their C terminus. This is reflected in the small PLS loadings at positions C-4 and C-5; the size loading z_2 is near zero and, hence, residues are not important at these positions. Thus, only peptides in a very narrow range of molecular masses (250–600 g mol^{-1}) or amino acid lengths (3–4) are effective inducers of oyster larval settlement. As long as there is an arginine residue at the C terminus of a peptide, biological activity remains in spite of considerable sequence variability at the C-1, C-2 and C-3 positions of tri- and tetrapeptides. Both the size and molecular properties of the stimulatory synthetic peptides and the natural inducer(s), as previously characterized, are in remarkable agreement.

The responses of barnacle larvae and ovigerous mud crabs to small basic peptide cues mirrored the pattern seen with oyster larvae. Barnacle larval settlement was stimulated when peptides (2–3 amino acids) were present at 10^{-8} to 10^{-10} mol l^{-1} (Tegtmeyer and Rittschof, 1989). For synthetic analogues active in barnacle settlement, the residue positions preceding the C-terminal arginine or lysine emphasized glycine and aliphatic amino acids. The effect of length cannot be compared with the data from oyster larvae since only peptides with two and three amino acids were tested. Ovigerous mud crabs also responded to peptides (1 –5 amino acids) but the effective concentrations ranged from 10^{-13} to more than 10^{-6} mol l^{-1}, over seven orders of magnitude (Rittschof *et al.* 1989). Again, activities were sorted by amino acid sequence: arginine at the C terminus was more active than lysine, and peptides with a hydrophobic residue preceding the C-terminal position were more active than those with other substitutions. Threshold responses segregated small peptides into sequence classes for inducing barnacle larval settlement and ovigerous crab larval dispersal.

The selectivity of peptide sequence for activity in oysters, barnacles and mud crabs was clearly different. This was despite all active peptides having obligate C-terminal basic residues and some overlap in the remainder of their sequences. For example, GGR was the most active peptide assayed for oyster larvae but falls near the centre of the tested scale for barnacle larvae and ovigerous crabs. Likewise, LR and LGR were the most active peptides for barnacle larvae, while LGR had approximately half the activity of the most potent oyster cue, and LR was approximately three orders of magnitude less

active than the most potent cue for stimulating mud crab abdominal pumping. The current QSAR models defined optimal sequences for eliciting patterns of oyster, barnacle and mud crab behaviour which differentiated peptides into physicochemical sequence classes: (1) oyster larvae settled in response to a three- to four-residue small, neutral and slightly hydrophilic–small hydrophilic–small hydrophilic–large hydrophilic, positively charged motif; (2) barnacle larvae settled in the presence of small neutral–small neutral and slightly hydrophobic–large hydrophilic, positively charged di- or tripeptides; (3) ovigerous mud crabs responded to two- to three-residue small neutral–small hydrophobic–large hydrophilic, positively charged peptides. These results suggest that individuals of different species can overlap spatially and temporally, yet selectively respond to appropriate signal molecules produced by conspecifics rather than being maximally stimulated by similar cues from other species.

Typical dose–response curves are expected to be sigmoidal or hyperbolic in shape (Tallarida and Jacob, 1979). The dose–response curves of most peptides were not sigmoidal or hyperbolic in oyster larval settlement over the entire range of concentrations tested but, rather, were bell-shaped (inverted parabolas). This phenomenon has been observed previously in behavioural bioassays of marine animals such as shrimp with adenosine 5′-monophosphate (Carr and Thompson, 1983), and barnacle larvae (Tegtmeyer and Rittschof, 1989) and mud crabs (Forward *et al.* 1987; Rittschof *et al.* 1989; Pettis *et al.* 1993) with natural and synthetic peptides. Mammalian physiological responses, such as neutrophil chemoattraction stimulated by C5a anaphylatoxin and its peptide analogues (Ember *et al.* 1992; Kawai *et al.* 1992), guinea-pig ileum contraction induced by decyl trimethylammonium ions (Stephenson, 1956) and gastric acid secretion induced by histamine (Code, 1982), also have bell-shaped dose–response curves. The bell-shaped dependence could be due to desensitization of chemosensory receptor cells. Cell responses are down-regulated as receptor proteins become unavailable by exclusion from membrane surfaces through receptor–ligand complex internalization. Evidence for such internalization is based on a previous study (Jesaitis *et al.* 1983) in which it was demonstrated that the chemotactic peptide *N*-formyl-Met-Leu-Phe disappeared from the outer membrane of human granulocytes and subsequently reappeared in the Golgi apparatus while bound to the receptor protein.

Quantitative structure–activity relationships were derived between peptide sequences and the behavioural responses of oyster and barnacle larvae and ovigerous mud crabs. In each case, the QSARs quantify what we already know about the abilities of structural analogues to influence animal behaviour by relating the physicochemical characteristics of signal molecules to their biological activity. PLS was specifically used in the current studies because it simultaneously finds factors that capture the greatest amount of variance in predictor variables (biological activity) with a single factor that best correlates predictor and predicted variables (Geladi and Kowalski, 1986; Lorber *et al.* 1987). This makes the predictive ability of the QSAR models very robust.

The PLS models for oyster and barnacle larvae and ovigerous crabs were predictive of their biological activities induced by peptides. Each model captured at least 95 % of the variance in behavioural activity. Generally, if one peptide was observed to have a higher activity than another, it was also predicted to have a higher activity. The PLS loadings describe graphically the essential features of the peptide sequences required for optimal response. Activities were calculated for over 400 to more than 8000 untested peptides for each animal system. In every case, at least one peptide was predicted to be more active than any of those already tested (although caution should be taken with the barnacle example since there were so few data points). The predicted peptides fit the parameters of the PLS loadings and are prime candidates for future synthesis and testing. Thus, in addition to quantifying relationships between structure and activity, QSARs allow the rational design of analogues with greater potencies based on the activities of relatively few compounds.

The relationship between marine invertebrate peptide cues and mammalian anaphylatoxins

The cleavage of the complement proteins C3, C4 and C5 by sequence-specific serine proteases yields the anaphylatoxins C3a, C4a and C5a, respectively. The anaphylatoxins elicit a wide range of immunological responses such as lysosomal enzyme release, increased vascular permeability, smooth muscle contraction, platelet aggregation and white blood cell chemotaxis (Hugli, 1981). The C-terminal sequence of each 74–78 residue anaphylatoxin is neutral hydrophobic–neutral hydrophilic–leucine–neutral–arginine (Moon *et al.* 1981); the C-terminal arginine is essential for activity (Bokisch *et al.* 1969; Bokisch and Müller-Eberhard, 1970; Gerard and Hugli, 1980). Extensive studies with short structural analogues of C3a and C5a indicate that the minimum amino acid length required to induce anaphylactic responses is 5–8 residues, although responses are generally much reduced compared with those of the native sequences (Hugli and Erickson, 1977; Caporale *et al.* 1980; Kawai *et al.* 1992). Some short synthetic peptides have demonstrated activity equal to or surpassing that of the native anaphylatoxins when hydrophobic residues are situated near their amino terminus, presumably because of secondary interactions with the receptor binding site (Ember *et al.* 1991, 1992). Significantly, the peptides predicted to be more active than GGR in inducing oyster larval settlement all have a hydrophobic residue at the C-3 position. The anaphylatoxins appear to be part of a ‘common theme’, together with the signal molecules involved in oyster and barnacle settlement and mud crab abdominal pumping, minimally described as peptides terminating in neutral–basic or neutral–neutral–basic residues (Hugli, 1981; Rittschof, 1990, 1993).

To induce a biological response, a compound must interact with a receptor binding site. A receptor can have varying degrees of specificity, and there can be multiple receptors that bind a molecular species. The nanomolar and lower threshold

concentrations of anaphylatoxins required to stimulate receptors indicate that the binding sites have a high affinity for their signal molecules (Gerardy-Schahn *et al.* 1988; Sanderson *et al.* 1995). Evidence for C3a, C4a and C5a having specific receptors arises from the different suites of responses from each anaphylatoxin on a single tissue (Fukuoka and Hugli, 1988). On the basis of synthetic analogue studies, the anaphylatoxin binding sites are highly sequence-selective (Unson *et al.* 1984; Ember *et al.* 1991, 1992; Sanderson *et al.* 1994, 1995). Strong correlations between different pharmacological activities within a set of C5a analogues support the hypothesis of a common receptor recognition site in a given tissue (Sanderson *et al.* 1994, 1995).

Many of the same arguments for C5a receptors can be used for C-terminal arginine-containing peptides interacting with oyster and barnacle larvae and mud crabs. These peptides are generally stimulatory at concentrations below 10^{-8} mol l^{-1}, indicating that the compounds have a high affinity for their receptors. QSAR analysis ranks the different synthetic analogues both by their physicochemical properties and by their biological activity. Subtly different peptide sequences induce different responses in oysters, barnacles and mud crabs, suggesting these marine animals have distinct but related receptors for C-terminal arginine peptides.

The receptors responsible for transmitting behavioural signals in oyster and barnacle larvae and mud crabs have yet to be isolated and characterized. Nevertheless, our QSAR analyses allow us to speculate regarding the structures of receptor binding sites. Similarities between the anaphylatoxins and the oyster/barnacle/crab-stimulating peptides and between their biological responses (concentration ranges, bell-shaped concentration dependence, active structural motifs) suggest that there may be homology among their respective receptors. Sequence analysis of the receptor protein for C5a indicates that it belongs to the rhodopsin superfamily (Boulay *et al.* 1991; Gerard and Gerard, 1991). Members of the rhodopsin receptor family bind a variety of molecular species, ranging from catecholamines and lipids to modestly sized proteins. By analogy, the receptors mediating behavioural responses in oyster and barnacle larvae and mud crabs may also belong to the rhodopsin class. An alternative hypothesis is that receptors for these C-terminal arginine peptides arose from a serine protease mutation (Rittschof, 1990, 1993). Trypsin, a well-known serine protease (Desnuelle, 1960), specifically hydrolyzes proteins on the carboxyl side of arginine or lysine, and the substrate is bound more tightly than the product. All of the C-terminal arginine synthetic peptides and natural cues tested could have arisen from digestion of a precursor protein by a trypsin-like enzyme. Minor modifications can alter trypsin into a form that binds trypsin digest products more tightly than the substrates, as with the removal of a single hydroxyl group in the case of anhydrotrypsin (Yokosawa and Ishii, 1977). Isolation of the receptor molecules is necessary to distinguish between the two possible progenitor proteins that could have given rise to the C-terminal arginine peptide receptors in oysters, barnacles and mud crabs.

Irrespective of receptor origin, the PLS loadings of the optimal peptide sequences can be used to form conceptual models for oyster larva and mud crab receptors (there is not enough information to do the same with the barnacle larva model). The magnitudes of the loadings can be matched with the appropriate amino acid descriptors (Fig. 4) such that opposite charges attract and hydrophobic moieties minimize their contact with water. A conceptual receptor model already proposed for mud crabs (Pettis *et al.* 1993) differs little from that presented here. One or two negatively charged residues such as aspartate or glutamate salt pair with the dispersed positive charge of the C-terminal guanidinium cation. Set apart from these residues is a positively charged arginine or lysine to stabilize the terminal carboxylate of the peptide. Further back along the receptor binding site, one or more large hydrophobic residues (e.g. tryptophan, phenylalanine) exclude water from the small hydrophobic C-1 functional group. Any large neutral group will effectively fill the space between the receptor and the C-2 amino acid. The portions of oyster, mud crab and C3a (Unson *et al.* 1984) receptors that bind the C-terminal arginine peptides are expected to be nearly identical. The oyster larval receptor provides large polar moieties from amino acids such as asparagine, glutamine or tyrosine to hydrogen-bond with the small polar residues at C-1 and C-2. The C-3 amino acid is larger and less hydrophilic than that at position C-1 or C-2, thus favouring hydrophobic residues from the receptor. Like their PLS loadings, these two receptor models are distinct and they are different from the model proposed for the C3a binding site (Unson *et al.* 1984).

Quantitative structure–activity relationships provide a novel approach to establishing the stimulatory effects of diverse molecules on biological responses, including animal behaviour. The argument can be made that a QSAR analysis of synthetic analogues may be even more useful foro understanding biological activity than identification of the native signal molecules because it ‘maps’ the essential structural features of the active moieties. This is, in fact, what we have done for oyster and barnacle larval settlement and ovigerous mud crab abdominal pumping behaviour. The different animal models were remarkably similar to qualitative evaluations of the anaphylatoxins. Our findings thus add evidence to support Haldane’s (1955) theory of an evolutionary link between more primitive external receptors functioning in chemical communication and internal receptors for mammalian neuro- and immunoactive agents.

We gratefully acknowledge the National Science Foundation (OCE 94-16749) and the UCLA Academic Senate for supporting this work. K. Kurkowski and M. W. Luckenbach graciously provided oyster larvae, and D. W. Schar gave expert assistance with larval settlement assays.

References

BAULIEU, E.-E. AND KELLY, P. A. (1990). (eds). *Hormones: From Molecules to Disease*. New York: Chapman & Hall.

BOKISCH, V. A. AND MÜLLER-EBERHARD, H. J. (1970). Anaphylatoxin inactivator of human plasma: Its isolation and characterization as a carboxypeptidase. *J. clin. Invest.* **49**, 2427–2436.

BOKISCH, V. A., MÜLLER-EBERHARD, H. J. AND COCHRANE, C. G. (1969). Isolation of a fragment (C3a) of the third component of human complement containing anaphylatoxin and chemotactic activity and description of an anaphylatoxin inactivator of human serum. *J. exp. Med.* **129**, 1109–1130.

BOULAY, F., MERY, L., TARDIF, M., BROUCHON, L. AND VIGNAIS, P. (1991). Expression cloning of a receptor for C5a anaphylatoxin on differentiated HL-60 cells. *Biochemistry* **30**, 2993–2999.

CAPORALE, L. H., TIPPETT, P. S., ERICKSON, B. W. AND HUGLI, T. E. (1980). The active site of C3a anaphylatoxin. *J. biol. Chem.* **255**, 10758–10763.

CARR, W. E. S. AND THOMPSON, H. W. (1983). Adenosine 5′-monophosphate, an internal regulatory agent, is a potent chemoattractant for a marine shrimp. *J. Comp. Physiol. A Sens. Neural Behav. Physiol.* **153**, 47–53.

CODE, C. F. (1982). Histamine receptors and gastric secretion. In *Pharmacology of Histamine Receptors* (ed. C. R. Ganellin and M. E. Parsons), pp. 217–235. Boston: Wright.

COLLANTES, E. R. AND DUNN III, W. J. (1995). Amino acid side chain descriptors for quantitative structure–activity relationship studies of peptide analogues. *J. med. Chem.* **38**, 2705–2713.

COON, S. L., FITT, W. K. AND BONAR, D. B. (1990). Competency and delay of metamorphosis in the Pacific oyster, *Crassostrea gigas* (Thunberg). *Mar. Biol.* **106**, 379–387.

CRISP, D. J. (1990). Gregariousness and systematic affinity in some North Carolinian barnacles. *Bull. mar. Sci.* **47**, 516–525.

CRISP, D. J. AND MEADOWS, P. S. (1962). The chemical basis of gregariousness in cirripedes. *Proc. R. Soc. Lond. B* **156**, 500–520.

DESNUELLE, P. (1960). Trypsin. In *The Enzymes*, 2nd edn, vol. 4 (ed. P. D. Boyer, H. Lardy and K. Myrbäck), pp. 119–132. New York: Academic Press.

DUNN III, W. J., WOLD, S., EDLUND, U., HELLBERG, S. AND GASTEIGER, J. (1984). Multivariate structure–activity relationships between data from a battery of biological tests and an ensemble of structure descriptors: The PLS method. *Quant. Struct.–Act. Relat.* **3**, 131–137.

EMBER, J. A., JOHANSEN, N. L. AND HUGLI, T. E. (1991). Designing synthetic superagonists of C3a anaphylatoxin. *Biochemistry* **30**, 3603–3612.

EMBER, J. A., SANDERSON, S. D., TAYLOR, S. M., KAWAHARA, M. AND HUGLI, T. E. (1992). Biologic activity of synthetic analogs of C5a anaphylatoxin. *J. Immunol.* **148**, 3165–3173.

FORWARD, R. B. AND LOHMANN, K. J. (1983). Control of egg hatching in the crab *Rhithropanopeus harrisii* (Gould). *Biol. Bull. mar. biol. Lab., Woods Hole* **165**, 154–166.

FORWARD, R. B., RITTSCHOF, D. AND DE VRIES, M. C. (1987). Peptide pheromones synchronize crustacean egg hatching and larval release. *Chem. Senses* **12**, 491–498.

FUKUOKA, Y. AND HUGLI, T. E. (1988). Demonstration of a specific C3a receptor on guinea pig platelets. *J. Immunol.* **140**, 3496–3501.

GALTSOFF, P. S. (1964). The American oyster, *Crassostrea virginica* Gmelin. *Fishery Bull. Fish. Wildl. Serv. U.S.* **64**, 1–480.

GELADI, P. AND KOWALSKI, B. R. (1986). Partial least-squares regression: A tutorial. *Analyt. chim. Acta* **185**, 1–17.

GERARD, C. AND HUGLI, T. E. (1980). Amino acid sequence of the anaphylatoxin from the fifth component of porcine complement. *J. biol. Chem.* **255**, 4710–4715.

GERARD, N. P. AND GERARD, C. (1991). The chemotactic receptor for human C5a anaphylatoxin. *Nature* **349**, 614–617.

GERARDY-SCHAHN, R., AMBROSIUS, D., CASARETTO, M., GRÖTZINGER, J., SAUNDERS, D., WOLLMER, A., BRANDENBURG, D. AND BITTER-SUERMANN, D. (1988). Design and biological activity of a new generation of synthetic C3a analogues by combination of peptidic and non-peptidic elements. *Biochem. J.* **255**, 209–216.

GLEN, W. G., DUNN III, W. J. AND SCOTT, D. R. (1989). Principal components analysis and partial least squares regression. *Tetrahedron Comput. Methodol.* **2**, 349–376.

HALDANE, J. B. S. (1955). Animal communication and the origin of the human language. *Scient. Prog.* **43**, 385–401.

HELLBERG, S., SJÖSTRÖM, M., SKAGERBERG, B. AND WOLD, S. (1987). Peptide quantitative structure–activity relationships, a multivariate approach. *J. med. Chem.* **30**, 1126–1135.

HOWE, N. R. AND SHEIKH, Y. M. (1975). Anthopleurine: A sea anemone alarm pheromone. *Science* **189**, 386–388.

HUGHES, J. (1978). Opioid peptides. In *Neuronal Information Transfer* (ed. A. Karlin, V. M. Tennyson and H. J. Vogel), pp. 283–293. New York: Academic Press.

HUGLI, T. E. (1981). The structural basis for anaphylatoxin and chemotactic functions of C3a, C4a, and C5a. *CRC Crit. Rev. Immunol.* **1**, 321–366.

HUGLI, T. E. AND ERICKSON, B. W. (1977). Synthetic peptides with the biological activities and specificity of human C3a anaphylatoxin. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1826–1830.

JESAITIS, A. J., NAEMURA, J. R., PAINTER, R. G., SKLAR, L. A. AND COCHRANE, C. G. (1983). The fate of an *N*-formylated chemotactic peptide in stimulated human granulocytes. *J. biol. Chem.* **258**, 1968–1977.

KAWAI, M., QUINCY, D. A., LANE, B., MOLLISON, K. W., OR, Y.-S., LULY, J. R. AND CARTER, G. W. (1992). Structure–function studies in a series of carboxyl-terminal octapeptide analogues of anaphylatoxin C5a. *J. med. Chem.* **35**, 220–223.

KÖNIG, W. (1993). *Peptide and Protein Hormones: Structure, Regulation, Activity*; *A Reference Manual*. New York: VCH.

LARMAN, V. N. (1984). Protein extracts from some marine animals which promote barnacle settlement: Possible relationship between a protein component of arthropod cuticle and actin. *Comp. Biochem. Physiol.* **77**B, 73–81.

LARMAN, V. N. AND GABBOTT, P. A. (1975). Settlement of cyprid larvae of *Balanus balanoides* and *Elminius modestus* Darwin. *J. mar. Biol. Ass. U.K.* **55**, 183–190.

LARMAN, V. N., GABBOTT, P. A. AND EAST, J. (1982). Physico-chemical properties of the settlement factor proteins from the barnacle *Balanus balanoides*. *Comp. Biochem. Physiol.* **72**B, 329–338.

LEHNINGER, A. L., NELSON, D. L. AND COX, M. M. (1993). *Principles of Biochemistry*, 2nd edn. New York: Worth Publishers.

LORBER, A. AND KOWALSKI, B. R. (1988). *A Cross-Validatory Approach to Error Estimation in the Multiple Linear Regression Model*. Center for Process Analytical Chemistry: Publication Announcement no. 55.

LORBER, A., VELTKAMP, D. J. AND KOWALSKI, B. R. (1988). *Outliers Analysis in Multivariate Calibration*. Center for Process Analytical Chemistry: Publication Announcement no. 57.

LORBER, A., WANGEN, L. E. AND KOWALSKI, B. R. (1987). A theoretical foundation for the PLS algorithm. *J. Chemometrics* **1**, 19–31.

MOON, K. E., GORSKI, J. P. AND HUGLI, T. E. (1981). Complete

primary structure of human C4a anaphylatoxin. *J. biol. Chem*. **256**, 8685–8692.

PAWLIK, J. R. (1992). Chemical ecology of the settlement of benthic marine invertebrates. *Oceanogr. mar. Biol. A. Rev*. **30**, 273–335.

PETTIS, R. J., ERICKSON, B. W., FORWARD, R. B. AND RITTSCHOF, D. (1993). Superpotent synthetic tripeptide mimics of the mud-crab pumping pheromone. *Int. J. Peptide Protein Res*. **42**, 312–319.

RAIMONDI, P. T. (1988). Settlement cues and determination of the vertical limit of an intertidal barnacle. *Ecology* **69**, 400–407.

RITTSCHOF, D. (1985). Oyster drills and the frontiers of chemical ecology: Unsettling ideas. *Am. Malacol. Bull. Special Ed.* **1**, 111–116.

RITTSCHOF, D. (1990). Peptide-mediated behaviours in marine organisms: Evidence for a common theme. *J. chem. Ecol.* **16**, 261–272.

RITTSCHOF, D. (1993). Body odors and neutral-basic peptide mimics: A review of responses by marine organisms. *Am. Zool.* **33**, 487–493.

RITTSCHOF, D., FORWARD, R. B., JR AND MOTT, D. D. (1985). Larval release in the crab *Rhithropanopeus harrisii* (Gould): Chemical cues from hatching eggs. *Chem. Senses* **10**, 567–577.

RITTSCHOF, D., FORWARD, R. B., JR, SIMONS, D. A., REDDY, P. A. AND ERICKSON, B. W. (1989). Peptide analogs of the mud crab pumping pheromone: Structure–function studies. *Chem. Senses* **14**, 137–148.

SANDERSON, S. D., KIRNARSKY, L., SHERMAN, S. A., EMBER, J. A., FINCH, A. M. AND TAYLOR, S. M. (1994). Decapeptide agonists of human C5a: The relationship between conformation and spasmogenic and platelet aggregatory activities. *J. med. Chem*. **37**, 3171–3180.

SANDERSON, S. D., KIRNARSKY, L., SHERMAN, S. A., VOGEN, S. M., PRAKASH, O., EMBER, J. A., FINCH, J. A. AND TAYLOR, S. M. (1995). Decapeptide agonists of human C5a: The relationship between conformation and neutrophil response. *J. med. Chem*. **38**, 3669–3675.

STEPHENSON, R. P. (1956). A modification of receptor theory. *Br. J. Pharmac*. **11**, 379–393.

TALLARIDA, R. J. AND JACOB, L. S. (1979). *The Dose–Response Relation in Pharmacology*. New York: Springer-Verlag.

TAMBURRI, M. N., ZIMMER-FAUST, R. K. AND TAMPLIN, M. L. (1992). Natural sources and properties of chemical inducers mediating settlement of oyster larvae: A re-examination. *Biol. Bull. mar. biol. Lab., Woods Hole* **183**, 327–338.

TEGTMEYER, K. AND RITTSCHOF, D. (1989). Synthetic peptide analogs to barnacle settlement pheromone. *Peptides* **9**, 1403–1406.

TOPLISS, J. G. AND EDWARDS, R. P. (1979). Chance factors in studies of quantitative structure–activity relationships. *J. med. Chem.* **22**, 1238–1244.

TURNER, E. J., ZIMMER-FAUST, R. K., PALMER, M. A., LUCKENBACH, M. AND PENTCHEFF, N. D. (1994). Settlement of oyster (*Crassostrea virginica*) larvae: Effects of water flow and a water-soluble chemical cue. *Limnol. Oceanogr*. **39**, 1579–1593.

UNSON, C. G., ERICKSON, B. W. AND HUGLI, T. E. (1984). Active site of C3a anaphylatoxin: Contributions of the lipophilic and orienting residues. *Biochemistry* **23**, 585–589.

WOLD, H. (1982). Soft modeling: The basic design and some extentions. In *Systems Under Indirect Observation: Causality, Structure, Prediction*, vol. 2 (ed. K. G. Jöreskog and H. Wold), pp. 1–54. New York: North Holland.

WOLD, S. AND DUNN III, W. J. (1983). Multivariate quantitative structure–activity relationships (QSAR): Conditions for their applicability. *J. Chem. Inf. Comput. Sci.* **23**, 6–13.

YOKOSAWA, H. AND ISHII, S. (1977). Anhydrotrypsin: New features in ligand interactions revealed by affinity chromatography and thionine replacement. *J. Biochem*. **81**, 647–656.

ZIMMER-FAUST, R. K. AND TAMBURRI, M. N. (1994). Chemical identity and ecological implications of a waterborne, larval settlement cue. *Limnol. Oceanogr.* **39**, 1075–1087.

ZIMMER-FAUST, R. K., TAMBURRI, M. N. AND DECHO, A. W. (1997). Chemosensory ecology of oyster larvae: Benthic–pelagic coupling. In *Zooplankton: Sensory Ecology and Physiology* (ed. D. K. Hartline, P. Lenz, J. E. Purcell and D. L. MacMillan), pp. 37–50. New York: Gordon and Breach.