Absolute Net Charge and the Biological Activity of Oligopeptides

Laavanya Parthasarathi,[†] Marc Devocelle,[‡] Chresten Søndergaard,[†] Ivan Baran,[†] Colm T. O'Dushlaine,[§] Norman E. Davey,[†] Richard J. Edwards,[†] Niamh Moran,[§] Dermot Kenny,[§] and Denis C. Shields*[,†]

UCD Conway Institute of Biomolecular and Biomedical Sciences, University College Dublin, Dublin 4, Republic of Ireland, Centre for Synthesis & Chemical Biology and Department of Medicinal and Pharmaceutical Chemistry and Department of Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland (RCSI), Dublin 2, Republic of Ireland

Received March 8, 2006

Sequences of human proteins are frequently prepared as synthetic oligopeptides to assess their functional ability to act as compounds modulating pathways involving the parent protein. Our objective was to analyze a set of oligopeptides, to determine if their solubility or activity correlated with features of their primary sequence, or with features of properties inferred from three-dimensional structural models derived by conformational searches. We generated a conformational database for a set of 78 oligopeptides, derived from human proteins, and correlated their 3D structures with solubility and biological assay activity (as measured by platelet activation and inhibition). Parameters of these conformers (frequency of coil, frequency of turns, the degree of packing, and the energy) did not correlate with solubility, which was instead partly predicted by two measures obtained from primary sequence analysis, that is, the hydrophobic moment and the number of charges. The platelet activity of peptides was correlated with a parameter derived from the structural modeling; this was the second virial coefficient (a measure of the tendency for a structure to autoaggregate). This could be explained by an excess among the active peptides of those which had either a large number of positive charges or in some cases a large number of negative charges, with a corresponding deficit of peptides with a mixture of negative and positive charges. We subsequently determined that a panel of 523 commercially available (and biologically active) peptides shared this elevation of absolute net charge: there were significantly lower frequencies of peptides of mixed charges compared to expectations. We conclude that the design of biologically active peptides should consider favoring those with a higher absolute net charge.

INTRODUCTION

The biological functions of a protein or peptide are often intimately dependent upon the conformations that the molecule can adopt. With the development of quantum chemical and molecular mechanical methods, theoretical modeling is possible. While the quantum chemical methods are too computationally intensive to explore a system with more than 100 atoms,[1] the use of stochastic conformational searches involving molecular mechanical principles is a viable alternative method to scan the conformational space of oligopeptides and explore the mechanism of their biological action.[2] Oligopeptides are frequently synthesized, and their biological activity is assessed. There are a number of interesting but difficult questions that arise in relation to the structures that such peptides adopt. First, the structure may influence solubility, affecting whether the peptide can be easily used in practice.[3,4] Second, the conformations that the peptide is free to adopt may well influence the peptide's biological activity. However, short peptides are likely to be extremely dynamic in solution. Linear oligopeptides represent a transient, globule-like structure which is considered to be an ensemble of interconverting populations of random coil and other conformational possibilities.[5] Therefore, it is unclear whether the set of conformers found by theoretical modeling will bear any relation to the main conformations adopted in reality. Clearly, even if the conformation obtained reflected the true conformation of the oligopeptide in solution, there remains the additional issue that this conformation may not reflect the conformation adopted in interaction with its ligand. Nevertheless, given the limited understanding of rules that may be usefully applied in designing oligopeptides that will be both soluble and active, we set out to determine if structural modeling provided any additional information regarding the solubility and activity, over and above that provided by predictions from primary amino acid sequences. We investigated a panel of oligopeptides for which we had information on the ease of solubilization and information on biological activity. They were modeled on the cytoplasmic regions of transmembrane proteins and were identified from an evolutionary sequence analysis of specificity-determining and conserved residues.[6] Because bioactive peptide motifs are often in disordered regions of protein surfaces,[7−10] it is of interest to investigate whether disorder-related aspects of oligopeptide structure might be associated with activity. The peptides we investigated here were

* Corresponding author phone: +353-1-716 6831; fax: +353-1-716 6701; e-mail: denis.shields@ucd.ie.
[†] University College Dublin.
[‡] Centre for Synthesis & Chemical Biology and Department of Medicinal and Pharmaceutical Chemistry, RCSI.
[§] Department of Molecular and Cellular Therapeutics, RCSI.

10.1021/ci0600760 CCC: $33.50
Published on Web 07/29/2006

palmitylated to facilitate membrane association. Circular dichroism analyses of oligopeptides reveal that they can adopt helical structures,[11−15] and NMR analyses of palmitylated peptides reveal that palmitylation induced helicity in the juxtamembrane portion of the peptide.[16]

Our objective was to determine whether computational structural modeling of oligopeptide conformers provided information regarding tertiary structure (over and above that suggested by computational methods using the primary sequence alone) that correlated with the solubility and activity of the oligopeptides. We also wished to determine if the property of such conformers to solubilize (as determined by second virial coefficient, B_{22}[17]) had any impact on the experimentally determined features of the peptides. The goal was to determine if any parameters, inferred from the best conformers or from the population of conformers for each peptide, would aid in the future design of novel oligopeptides with good properties of solubility and activity. Typically, such automatically generated conformer models are unlikely to be individually highly reliable but could have the capacity to identify overall trends that would give us clues about the structural constraints on solubility and activity. In addition to investigating parameters derived from the inferred 3D conformers, we also investigated a number of simpler parameters estimated from the primary sequences or from alignments of the primary sequences with biologically homologous sequences.

METHODOLOGY

Structural Features Derived from Conformational Modeling. Two approaches to modeling were considered, because the experimental evaluation of the peptide solubility and activity had been carried out on N-terminally palmitylated forms of 78 decameric peptides. The choice of the peptides followed a rationale described elsewhere.[6] In the first ("native peptide") model, the peptides were modeled as a simple decamer without any modification of N and C termini. The second model was closer to the states of the actual peptides, with C-amidation (corresponding to the experimental modification of the C terminus) and N-acetylation. N-acetylation is an approximation of the true experimental modification of the N terminus, which comprised an N-palmitylated group. While this model may not accurately reflect the true structure, the palmityl group is likely to be membrane-associated in functional assays and, therefore, cannot be accurately modeled in an aqueous force field. For the purposes of the analyses here, we are most interested in differences among peptides, so that even though the conditions of the modeling may not be entirely realistic, broad trends that emerge may well be relevant to the biological situation. Considering the differences within and between the "native" and "terminally modified" models permitted some consideration of the extent to which modification of the modeling parameters alters the general conclusions made. The conformational library of the two sets of peptides (the first set consisting of the native 10-mer sequences and the second set consisting of the *N*-acetylated and *C*-amidated 10-mer sequences) was created using the stochastic conformational search method available in the Molecular Operating Environment software (MOE).[18] In stochastic searches, new molecular conformations are generated by random rotation of the bonds. In the calculations, a bias of ∼30° was selected, so that dihedral angles were rotated by a random angle with a sum-of-Gaussians distribution with peaks at multiples of 30°. To make the search even more efficient, a simultaneous Cartesian perturbation of 0.0001 Å was also applied. The structures were subsequently optimized using the AMBER94 force field[19,20] employing Born continuum solvation model[21−23] as implemented in MOE. Each resulting conformation had been checked to determine if it had already been generated by comparing all atom positions using a predefined root-mean-square (RMS) tolerance (0.01 in this work). Conformations with energies > 100 kcal mol^{-1} above the respective minima were rejected. The minimum energy for each conformer was compared to an expectation derived from the energies for the 10 amino acids in the same force field minus the energy of nine water molecules eliminated during decamer formation, and the difference in energy (ΔE) was calculated as the difference between the expected value and the actual value obtained for the best conformer of the peptide.

The secondary structures of the peptide conformers obtained were calculated by the DSSP algorithm.[24] In a polypeptide, the main chain N−C_α and C_α−C bonds are relatively free to rotate. These rotations are represented by the torsion angles ϕ and ψ, respectively. The ϕ and ψ values for the peptide conformers were obtained using the RASTOP package.[25] On the basis of the ϕ and ψ values of each residue in the peptide conformers, the possible regions (favored, allowed, and disallowed) they span in the Ramachandran conformational map were identified. As a measure of how extended or compacted a given peptide structure might be, we defined the packing average as the mean pairwise distance in angstroms between each amino acid of the individual 10-mer peptides:

$$\text{packing average} = {}^1/_{45}\sum_{i=1}^{10}\sum_{j=i+1}^{10}[(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2]^{1/2}$$

where x, y, and z are the coordinates of the C_α atoms. From the individual packing average for each of the conformers of a peptide, the mean and standard deviation values were calculated. Among the peptides taken for this investigation, there are 26 pairs of peptides derived from alignable membrane-proximal regions of evolutionary-related transmembrane proteins (BLAST E value $< 1 \times 10^{-12}$). The 10-mer aligned peptide pairs had a mean percentage identity of 49% (standard deviation, SD, 20%). For these 26 pairs, the RMS deviation (RMSD) for the backbone was calculated[26] for the best conformer of the first peptide, compared to the conformer in the top 10 conformers of the second peptide with the most similar RMSD.

Method for Calculating the Second Virial Coefficient (B_{22}). The second virial coefficients were calculated from the interaction energy between two molecules of the same peptide using the McMillan−Mayer result[27]

$$B_{22} = \frac{-1}{16\pi^2 {M_w}^2}\int(e^{-w/kT} - 1)\, d\Omega$$

where M_w is the molecular weight, w is the potential of mean

Table 1. List of Predefined Grids and Number of Grid Points Used in the Calculations

domain	r_{min} [Å]	r_{max} [Å]	r_{step} [Å]	Δangle [deg]	calculations
1	0	5	1	45	64 000
2	5	30	0.5	18	48 400 000
3	30	50	0.5	30	3 386 880
4	50	100	2	45	320 000
				in total	52 170 880

force[28] acting between two peptide molecules in solution, k is the Boltzmann constant, T is the absolute temperature, and Ω indicates that the integration is carried out over all possible relative positions and orientations of the two molecules. The relative position of the two molecules is described using the spherical coordinates (r,θ,ϕ), and the relative orientation is described using the Euler angles (α,β,γ) giving a total of six dimensions. In practice, however, it is not possible to calculate the interaction energy for all possible relative orientations and positions of the two molecules. Instead, the interaction energy is calculated on four predefined grids as listed in Table 1.

Each domain spans from r_{min} to r_{max} in the r dimension that represents the molecular center−center distance. In each domain, the spacing between the grid points in the r dimension is given by r_{step} and by Δangle in the remaining five angular dimensions (θ, ϕ, α, β, and γ). The calculations are divided in the r dimension into four domains, allowing a different grid resolution in each domain. In the first domain (defined as all interactions for which the molecular center−center distance is less than 5 Å and where overlap between the peptide molecules therefore always occurs), the distance between grid points is 1 Å in the r dimension and 45° in each of the five remaining angular dimensions. This gives a total of $64\,000\{5[(360/45)]^3[(180/45)+1]^2\}$ grid points. In the second domain where the molecular center−center distance is in an interval [5−30 Å], all of the close interactions between the molecules are expected to happen, and hence, a high-resolution grid of a total of 48 400 000 grid points was applied. At larger molecular center−center distances (domains 3 and 4), a lower grid resolution is used as the interactions here are predicted to be long-range in nature and less dependent on the precise molecular positions. The interaction energy, w, was calculated for each of the grid points on the six-dimensional grid using the terms $w = w_{steric} + w_{nonelectrostatic} + w_{electrostatic}$. The steric interaction energy was only applied when the peptides were overlapping as $w_{steric} = \infty$, and otherwise, $w_{steric} = 0$. The electrostatic term of the interaction energy was modeled using the linearized Poisson−Boltzmann equation

$$\nabla\epsilon(\vec{r})\,\nabla\phi(\vec{r}) - \vec{\kappa}^2\phi(\vec{r}) + 4\pi\rho(\vec{r}) = 0$$

where ϵ is the relative permittivity, ϕ is the electrostatic potential, κ is the Debye−Hückel parameter, and ρ is the charge distribution of the peptide. The relative permittivity was set to 3 in the interior of the peptide and 78.4 outside the peptide. The partial charge for each atom was assigned using the AMBER force field parameter set for proteins and nucleic acids.[29] Protonation states were set according to standard pK_a values of free amino acids. The electrostatic potential around each peptide was found by solving the Poisson−Boltzmann equation using a finite difference method.[30] The electrostatic potential around the peptide was found on a Cartesian grid consisting of $256 \times 256 \times 256$ grid points separated by 1 Å. The electrostatic potential at a given point in space outside the peptide is then found by linear interpolation between the calculated points. The electrostatic interaction energy can then simply be found by multiplying the partial charge of each atom with the electrostatic potential at its position and summed up over all atoms. The nonelectrostatic interactions were modeled using a hybrid potential.[31] For interactions with a center−center distance of less than 6 Å, the Lennard-Jones potential was used, while interactions with greater separations were calculated using the Hamaker potential. The Lennard-Jones potential was calculated using parameters from the AMBER force field parameter.[29] The Hamaker potential was calculated using a theoretical Hamaker constant of $3.1kT$ corresponding to protein−water−protein interactions[32] and using molecular group volumes.[33] All of the calculated interaction energies were integrated as described by the McMillan−Mayer result to evaluate the second virial coefficient. First, the interaction energies were integrated in the r dimension using Simpson's rule, leaving five remaining dimensions to integrate. These remaining dimensions were integrated using Monte Carlo integration[34] by generating random points in the five-dimensional subspace and estimating the interaction energy at these points by linearly interpolating between the calculated interaction energies for the predefined grid points.

Structural Features Predicted from Primary Amino Acid Sequence. A variety of predicted features that might relate to peptide solubility and activity were estimated to compare alongside the structural predictions with solubility and activity of the peptides. The molecular weight, instability index, GRAVY, hydrophobicity score, theoretical pI, and aliphatic index of the peptides were calculated using EMBOSS.[35] The helical hydrophobic moment values for the peptides were calculated using the HMOMENT server.[36] Because antimicrobial peptides may disrupt membranes, we calculated measures related to the potential similarity to such peptides. The total hydrophobic ratio, total net charge, and Boman index[37] for the peptides were calculated using the antimicrobial peptide prediction server. Short signaling motifs often occur in disordered regions of peptides,[10] and therefore, we calculated a number of predictions of disorder, from the IUPRED,[38] DisEMBL,[39] and GLOBPLOT 2.1[40] servers. Functional motifs present in the peptides were analyzed using the ELM,[41] Scansite,[42] and Automotif[43] servers. The various secondary structural elements present in the peptides were predicted from the primary sequences, using the PSIPRED[44] (α-helix, β-sheet and coil), BTSVM[45,46](β-turn), and BHAIRPRED[47](β-hairpin) servers. The antigenic sites[48] overlapping the peptide sequences were predicted using the antigenic site prediction server. The helical content of the peptides was predicted using the AGADIR server.[11−15]

Comparison with Solubility and Activity. The activity data for the peptides is described in detail elsewhere.[6] Briefly, four different assays of platelet function (aggregation, ADP release, inhibition of thrombin-induced aggregation, and inhibition of thrombin-induced ADP release) were measured. p values from all four functional assays were obtained. "Activity" was defined as 1 if one of the four assays showed a significant effect for the peptide ($p < 0.05$); otherwise, it was defined as 0. The variable "solubility" was defined as a simple three-point scale, depending on whether they were

Table 2. Fraction of Residues Classified into Secondary Structural States

		native peptides		modified peptides	
secondary structure states[a]	definition	mean[b]	best conformer[c]	mean[b]	best conformer[c]
dssp(s)	random coils	0.35593	0.37948	0.35940	0.36923
dssp(t)	hydrogen-bonded turns	0.05841	0.06538	0.05114	0.06538
dssp(g)	3_{10}-helix	0.00317	0.00384	0.00317	0.00384
dssp(b)	residue in isolated β bridge	0.00039	0.00000	0.00086	0.00000
dssp(h)	α-helix	0.00003	0.00000	0.00027	0.02051
dssp(e)	extended strand	0.00000	0.00000	0.00002	0.00000
dssp(i)	π-helix	0.00129	0.00000	0.00004	0.00000

[a] As estimated by DSSP program. [b] Fraction for the peptide over all the conformers averaged over all the peptides. [c] Fraction for the best conformer of the peptide averaged over all the peptides.

Table 3. Spearman Correlation Coefficients among Pairs of Structural- and Sequence-Derived Variables

variables	number of conformers	ΔE	dssp(s) random coil	dssp(t) H-bonded turn	pack (mean)	SD of pack (mean)	B_{22}
			3D-Structural Variables[a,b]				
ΔE	−0.408						+0.321
dssp(s)	−0.503						
dssp(t)	+0.492		−0.578				
pack (mean)	+0.578	−0.367	−0.601	+0.367			
SD for pack (mean)	+0.902	−0.375	−0.472	+0.502	+0.568		
dssp(h)	+0.786	−0.383	−0.547	+0.566	+0.572	+0.707	
dssp(b)	+0.792	−0.394	−0.520	+0.605	+0.545	+0.702	
dssp(e)	+0.491		−0.432	+0.417	+0.368	+0.438	
dssp(i)	+0.619		−0.459	+0.461	+0.400	+0.500	
			Variables Based on Peptide Sequences[b]				
absolute net charge		+0.471					+0.854
Boman index		+0.348					
theoretical pI		+0.575					+0.605
mol. weight		+0.460					

[a] N-acetylated and C-amidated model. Similar correlation values are found for the native peptides (data not shown). [b] Only correlation coefficients with $P \leq 0.005$ and $\rho \geq 0.300$ are presented.

soluble in water (solubility of 1), whether MeOH was required to be added in order to achieve solubility (solubility of 2), and whether some DMSO was required to be added in order to achieve solubility (solubility of 3), with up to 100% DMSO being required for a number of peptides. The statistical analyses of the data were carried out with the STATA 8.0 package.[49]

RESULTS AND DISCUSSION

About 60% of modeled nonglycine residues fall outside the ϕ and ψ bond angles normally permitted in proteins.[50] Thus, the opportunity to adopt conventional α-helical structures is limited, and therefore, the proportion of peptide residues predicted by the 3D conformers to be in an α-helical conformation is very low (Table 2). The DSSP algorithm is able to make strong assignments that 36% of the residues are likely to be random coils [dssp(s)] but can only assign around 6% to hydrogen-bonded turns [dssp(t)]. The occurrence of so many illegal bond angles suggests that many of the conformations adopted may not represent well the true peptide conformations and reduces the power of DSSP to assign reasonable estimates of secondary structure. Bearing this in mind, we considered the overall trends across the data set.

The conformational features of the proline (Pro) rings in the Pro-containing peptides were analyzed on the basis of the endocyclic torsion angles. Two native peptides had one of the proline rings in them in a planar conformation, whereas others have the Pro rings in either up or down puckered conformations.[51] However, in most of the peptides, the conformation of the Pro rings differs between the native and modified peptides. This indicates that the proline conformations adopted are sensitive to the modifications at the N and C termini.

The 78 peptides fall largely into two groups: those where ΔE is negative, and where typically there are many alternative conformations possible, and those where it is positive and there are far fewer conformations. Table 3 indicates that there is a strong correlation of both the number of conformers and ΔE with the packing average, indicating that the peptides with more conformers have more extended structures; their conformations obtained may then be close to stochastic and may reflect greater fluidity of such peptides in solution. ΔE was strongly correlated with four parameters determined from the primary sequence, namely, the total net charge, Boman index, theoretical pI, and molecular weight (Table 3).

Correlations of Structural and Sequence Features with Solubility. It might have been predicted that more extended peptides should be more soluble, but we found that ΔE or packing were not correlated with solubility (Table 4). The B_{22} parameter did not correlate with the solubility of these short peptides (Table 4), despite being a good predictor of solubility for larger proteins.[52−54] The modeling of such higher-order structures is currently infeasible. What then accounts for the difference? It may be that substantial aggregates of a fairly large number of oligopeptides remain highly soluble, so long as the surface charge of the aggregate remains high. It is possible that the calculation of B_{22} is prone

Table 4. Spearman Pairwise Correlation of Activity and Solubility with 7 3D-Structural Variables

	native				modified[a]			
	activity[b]		solubility[c]		activity[b]		solubility[2]	
3D-structural variables	ρ	p	ρ	p	ρ	p	ρ	p
number of conformers	+0.252	0.026	−0.057	0.6190	+0.089	0.439	+0.040	0.730
ΔE	+0.070	0.544	−0.067	0.5615	+0.070	0.544	−0.067	0.561
dssp(s) random coil	−0.082	0.473	−0.002	0.9885	−0.113	0.323	−0.037	0.748
dssp(t) H-bonded turn	+0.042	0.717	−0.199	0.0814	+0.000	1.000	+0.182	0.110
packing (mean)	+0.041	0.722	+0.129	0.2598	+0.113	0.322	−0.047	0.680
SD for packing (mean)	+0.181	0.112	+0.049	0.6690	+0.186	0.102	+0.038	0.739
packing (best conformer)	−0.004	0.972	+0.091	0.4296	+0.153	0.182	−0.035	0.763
B_{22}	+0.331	0.003	+0.022	0.8490	+0.300	0.008	+0.153	0.182

[a] N-acetylated and C-amidated. [b] Activity is defined as 1 if one of the four assays showed a significant effect for the peptide ($p < 0.05$), otherwise it is defined as 0. [c] Solubility is defined as a simple three-point scale, depending on whether they were soluble in water (solubility of 1), whether MeOH was required to be added in order to achieve solubility (solubility of 2), and whether some DMSO was required to be added in order to achieve solubility (solubility of 3).

to error, assuming the conformational model selected. However, we noted that B_{22} is strongly correlated with the absolute net charge (Spearman $\rho = 0.85$ and $P = 0.00$) across the peptide data set (i.e., B_{22} is high for both those with a large number of positive and a large number of negative charges). Solubility showed a lack of correlation with either B_{22} ($p = 0.18$) or the absolute net charge ($p = 0.81$). Because it would be anticipated that peptides of mixed charge should associate more easily than those with a large absolute net charge, we conclude that the lack of association of solubility with either B_{22} or absolute net charge is unlikely to merely reflect problems with the computation of B_{22}.

The impact of features predicted from the primary sequences on solubility was assessed by a stepwise regression of the non-3D structural variables (ranked where appropriate) starting with a full model with all variables and then successively eliminating the least significant terms, with a threshold of $p = 0.05$. Only the hydrophobic moment and the number of charge centers in the peptide were significant independent predictors of increased solubility, with peptides insoluble in the two protic solvents (water and MeOH) having a low hydrophobic moment (Figure 1a). A strong hydrophobic moment may also have the property that the charge is distributed more evenly along the peptide, perhaps improving the solubility. It is striking that peptides of intermediate solubility (requiring MeOH to bring the peptide into solution) generally are more hydrophobic but have a higher hydrophobic moment than those that require DMSO (Figure 1b−d). No significant correlation is observed between the various disorder prediction scores and the solubility of the peptides.

The design of the peptides in this study included 26 pairs of peptides derived from homologous regions of proteins. This allows us to consider whether simple residue conservation or RMSD among models from pairs of evolutionarily related peptides is a better predictor of similarities in solubility. However, there was no correlation of RMSD with vehicle similarity ($\rho = 0.00$ and $p = 0.99$). Those peptide pairs with a strong residue identity tend more often to be soluble in the same vehicle. This is confirmed by a significant association ($\rho = 0.445$ and $p = 0.03$) between residue identity and solubility. This is likely to simply reflect the impact of residue composition on solubility, regardless of the particular modeled conformation.

Correlations of Structural and Sequence Features with Activity. ΔE and packing did not influence the activity (Table 4). We noted that, when the analysis was restricted to those peptides with a positive ΔE, there were weak correlations of activity with the energy of the best conformer. While this is suggestive, the statistical significance was weak, given the extent of multiple testing of alternative hypotheses. Somewhat surprisingly, the structural parameter B_{22} was correlated with activity (Table 4). A possible explanation is that peptides with higher B_{22} values are less likely to form larger aggregates, increasing their effective availability for interaction with ligands.

To investigate whether this correlation of B_{22} with activity could be more simply represented in terms of sequence characteristics, we considered the absolute net charge as a predictor variable, because B_{22} is closely related to the absolute net charge (i.e., molecules with a high B_{22} tend to either have a large net positive charge or else a large net negative charge). The parameters total net charge, theoretical pI, absolute net charge, helical content (Agadir prediction), and the ranked B_{22} (second virial coefficient) values for the two sets of peptides had significant pairwise correlations with activity ($p < 0.008$). However, stepwise regression indicated that absolute net charge alone was the most significant predictor, with the other variables not adding significantly to this model. This trend is seen in peptides solubilized in all three vehicles (Figure 2). We noted that the absolute net charge came out as a significant predictor of activity measured independently for each of the four individual assays that comprised the composite activity variable. Clearly, the potential absolute net charge is dependent to some extent on the number of charged residues in a peptide. Figure 3 illustrates the wider spread among the active peptides of absolute net charge for any given number of charge centers in the peptide. Put simply, it appears that there is an excess among active peptides of those with a larger number of positive charges and those with a larger number of negative charges. For example, those peptides with three charge centers show a strong preference for three positive charges among the active peptides, while peptides with two positive and one negative charge predominate among the inactive peptides. We would have anticipated that modeled conformers of peptides with a high net absolute charge would tend to have a more extended structure, because of the repulsive forces separating these same charges, while those with a

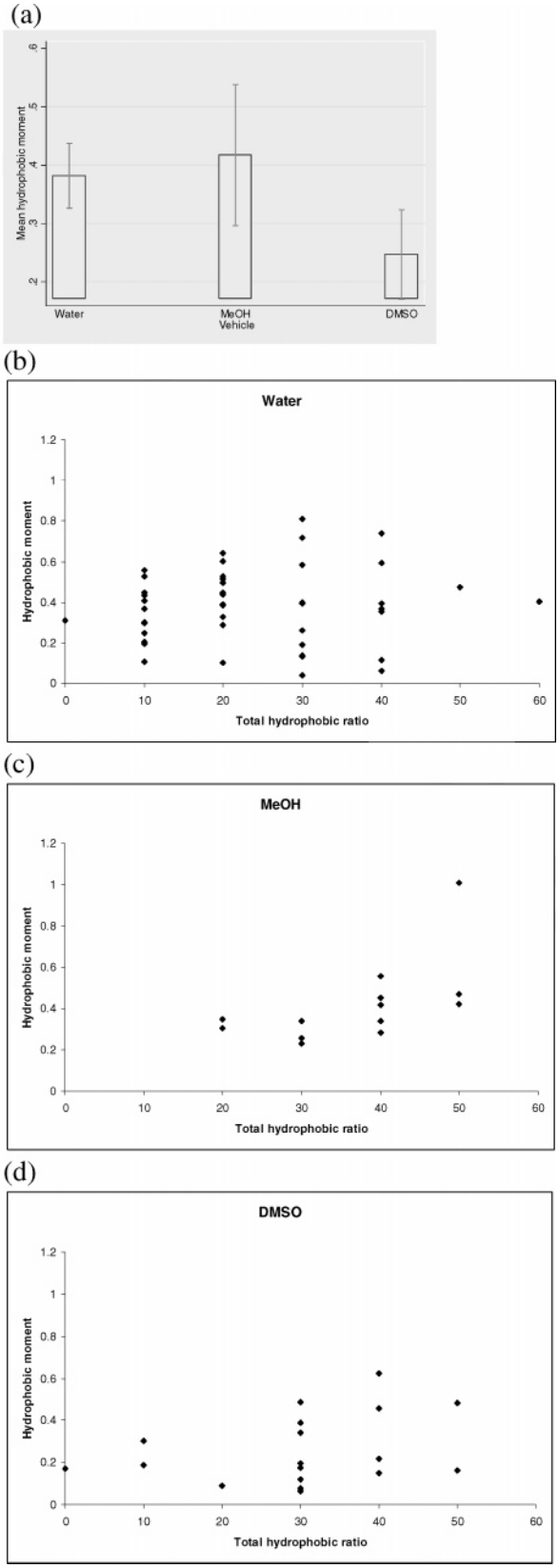

Figure 1. Impact on solubility of hydrophobic moment and total hydrophobic ratio. Hydrophobic moment is compared among vehicles (a) and its distribution shown with that of the total hydrophobic ratio for (b) water, (c) MeOH, and (d) DMSO.

mixture of charges might have a less extended structure, owing to the electrostatic stabilization between a negative and a positive charge. However, looking at peptides with three charge centers, in each of the 11 where all three were

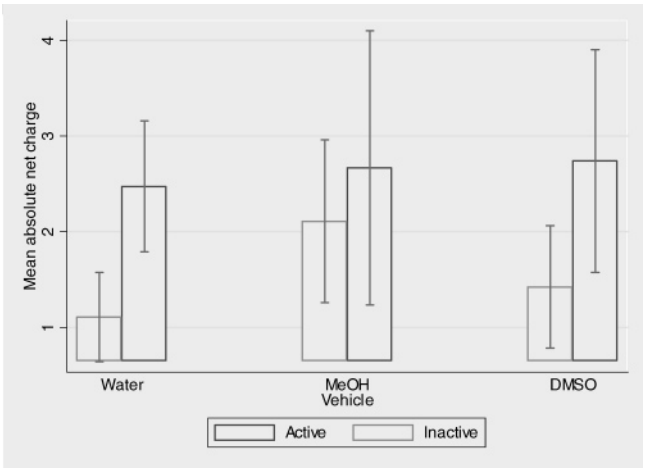

Figure 2. Impact of absolute net charge on the platelet activity of 78 peptides.

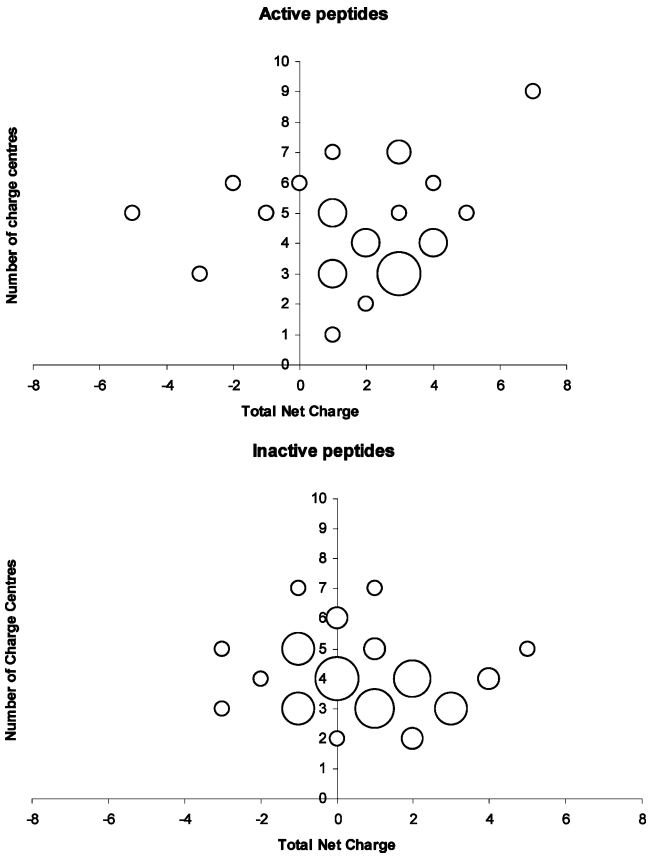

Figure 3. Overdispersion of total net charge in active and inactive platelet peptides (size of bubble is proportional to the number of peptides).

all positive (K or R), the ΔE values were positive, whereas the 13 peptides with a mixture of negative and positive charges included five with negative ΔE values. It is not clear if this is a chance observation, if the implicit solvation in the model is sufficient to overcome the stabilization of peptides of mixed charge, or if the calculation of ΔE tends to lead to an excessive estimate in the case of the bulkier positively charged amino acids.

Having established absolute net charge as a predictor variable of activity in this data set of 78 peptides, we then wished to determine if this is a general trend seen across many kinds of oligopeptides or if this is a specific observation to palmitylated peptides, or to peptides acting on platelet

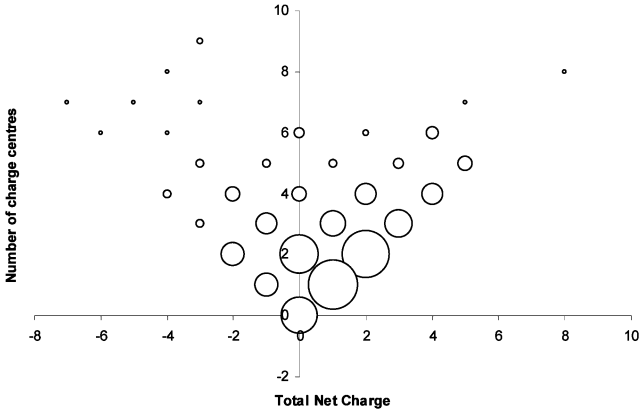

Figure 4. Overdispersion of total net charge in 523 commercially available oligopeptides (length 7−16) (size of bubble is proportional to the number of peptides).

signaling. Because commercially available peptides are usually experimentally active, we took a data set of 523 commercially available oligopeptides of length 7−16 (sequences obtained from the Web site of a commercial supplier, the American Peptide Company: www.americanpeptide-.com). We found a greater than expected frequency of peptides with a high absolute net positive charge (Figure 4). This effect was statistically significant for various numbers of charges. Of peptides with two charges, 84 were both positive, 21 both negative, and 56 had no net charge (Figure 4): this represented a significant departure from the binomial expectations of 77.9, 14.8, and 68.2, respectively [$p = 0.02$, χ-square with 1 degree of freedom (d.f.)]. The same pattern was seen in those with three charges, where observations of 4, 32, 21, and 25 were seen for −3, +3, −1, and +1, respectively, compared to expectations of 2.7, 25.7,17.2, and 36.4 ($p = 0.05$, 2 d.f.). A similar departure is seen for peptides with four charge centers ($p = 0.03$, 3 d.f.). For peptides with greater numbers of charges, visual inspection reveals a very similar excess of absolute net charge compared to expectations (Figure 4). While a subset of the bioactive peptides was known to have an excess of positive charge that directly related to their membrane disrupting activity,[55] these alone cannot account for the pattern seen in Figure 4. No significant correlation is observed between the various disorder prediction scores and the activity of the peptides.

Our analyses suggest that routine conformational modeling during the selection of peptides is not proven to provide any additional information that would help identify more soluble or more active peptides. However, the structural modeling led us to identify a simple correlation of absolute net charge with activity that was not only true for our own data set but also of relevance to a wider range of peptides. While the second virial coefficient (B_{22}) did not correlate as well in our data set with activity as did the simple absolute net charge, the models presented here of oligopeptides may only represent an approximate sample of the population of the conformers, given their great flexibility. It would be of interest to repeat this analysis in a structurally more constrained data set, of shorter or cyclized compounds. In such a data set, where there is greater confidence in the models, it will be easier to determine whether the second virial coefficient contains additional information regarding activity rather than simple absolute net charge. Our observation that the hydrophobic moment, as well as hydrophobicity, appears to predict the solubility of peptides suggests that the selection of soluble peptides for further studies should consider not only hydrophobicity or number of charges but also the distribution of charges and hydrophobicity along the peptides. While absolute net charge is correlated with bioactivity, we are somewhat surprised that it does not relate to experimental ease of solvation in water. Perhaps this indicates that, for oligopeptides, aggregates may often enter solution as easily as monomers. The main conclusion of our study is that the design of data sets of active peptides should consider favoring those peptides with a high absolute net charge. In combinatorial chemistry, this could be achieved by combining peptide syntheses from two separate pools of amino acids, one with an excess of negatively charged amino acids and the other with an excess of positively charged amino acids. In the design of biomimetic oligopeptides, charged groups could be added to the ends or substituted in the design. It will be very interesting to systematically determine if such modifications to peptide design protocols increase peptide bioactivity.

ACKNOWLEDGMENT

This work was supported by Science Foundation Ireland, by the Health Research Board, and by the Program for Research in Third Level Institutions administered by the Higher Education Authority. The authors acknowledge the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. We thank Xu Henan for programming assistance and Jens Nielsen, Edelmiro Moman, and Anthony Chubb for discussion.

REFERENCES AND NOTES

(1) Yang, Z. Z.; Zhang, Q. Study of Peptide Conformation in Terms of the ABEEM/MM Method. *J. Comput. Chem.* **2006**, *27*, 1−10.
(2) Groth, M.; Malicka, J.; Rodziewicz-Motowidlo, S.; Czaplewski, C.; Klaudel, L.; Wiczk, W.; Liwo, A. Determination of Conformational Equilibrium of Peptides in Solution by NMR Spectroscopy and Theoretical Conformational Analysis: Application to the Calibration of Mean-Field Solvation Models. *Biopolymers* **2001**, *60*, 79−95.
(3) Hilbich, C.; Kisters-Woike, B.; Reed, J.; Masters, C. L.; Beyreuther, K. Aggregation and Secondary Structure of Synthetic Amyloid Beta A4 Peptides of Alzheimer's Disease. *J. Mol. Biol.* **1991**, *218*, 149−163.
(4) Powers, M. E.; Matsuura, J.; Brassell, J.; Manning, M. C.; Shefter, E. Enhanced Solubility of Proteins and Peptides in Nonpolar Solvents through Hydrophobic Ion Pairing. *Biopolymers* **1993**, *33*, 927−932.
(5) Mayo, K. H. Recent Advances in the Design and Construction of Synthetic Peptides: For the Love of Basics or Just for the Technology of It. *Trends Biotechnol*. **2000**, *18*, 212−217.
(6) Edwards, R. J.; Moran, N.; Kenny, D.; Devocelle, M.; Kiernan, A.; Dunne, E.; Meade, G.; Park, S. D. E.; Foy, M.; Shields, D. C. Identification of Diverse Sets of Bioactive Oligopeptides from Human Proteins. *Nature Chem. Biol.* (Submitted for publication).
(7) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. Intrinsic Disorder in Cell-Signaling and Cancer-Associated Proteins. *J. Mol. Biol.* **2002**, *323*, 573−584.
(8) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically Disordered Protein. *J. Mol. Graphics Modell.* **2001**, *19*, 26−59.
(9) Wright, P. E.; Dyson, H. J. Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure−Function Paradigm. *J. Mol. Biol.* **1999**, *293*, 321−331.
(10) Neduva, V.; Russell, R. B. Linear Motifs: Evolutionary Interaction Switches. *FEBS Lett.* **2005**, *579*, 3342−3345.
(11) Muñoz, V.; Serrano, L. Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. *Nat. Struct. Biol*. **1994**, *1*, 399−409.

(12) Muñoz, V.; Serrano, L. Elucidating the Folding Problem of α-Helical Peptides Using Empirical Parameters, II. Helix Macrodipole Effects and Rational Modification of the Helical Content of Natural Peptides. *J. Mol. Biol.* **1994**, *245*, 275−296.
(13) Muñoz, V.; Serrano, L. Elucidating the Folding Problem of α-Helical Peptides Using Empirical Parameters III: Temperature and pH Dependence. *J. Mol. Biol.* **1994**, *245*, 297−308.
(14) Muñoz, V.; Serrano, L. Development of the Multiple Sequence Approximation within the Agadir Model of α-Helix Formation. Comparison with Zimm-Bragg and Lifson-Roig Formalisms. *Biopolymers* **1997**, *41*, 495−509.
(15) Lacroix, E.; Viguera, A. R.; Serrano, L. Elucidating the Folding Problem of α-Helices: Local Motifs, Long-Range Electrostatics, Ionic Strength Dependence and Prediction of NMR Parameters. *J. Mol. Biol.* **1998**, *284*, 173−191.
(16) Lockwood, N. A.; Haseman, J. R.; Tirrell, M. V.; Mayo, K. H. Acylation of SC4 Dodecapeptide Increases Bactericidal Potency Against Gram-Positive Bacteria, Including Drug Resistant Strains. *Biochem. J.* **2004**, *378*, 93−103.
(17) Guo, B.; Kao, S.; McDonald, H.; Asanov, A.; Combs, L. L.; Wilson, W. W. Correlation of Second Virial Coefficients and Solubilities Useful in Protein Crystal Growth. *J. Cryst. Growth* **1999**, *196*, 424−433.
(18) *Molecular Operating Environment*, version 2005.06; Chemical Computing Group Inc.: Montreal, Canada.
(19) Ripoll, D. R.; Scheraga, H. A. The Multiple-Minima Problem in the Conformational Analysis of Polypeptides. III. An Electrostatically Driven Monte Carlo Method: Tests on Enkephalin. *J. Protein Chem.* **1989**, *8*, 263−287.
(20) Ripoll, D. R.; Scheraga, H. A. On the Multiple-Minima Problem in the Conformational Analysis of Polypeptides. IV. Application of the Electrostatically Driven Monte Carlo Method to the 20-Residue Membrane-Bound Portion of Melittin. *Biopolymers* **1990**, *30*, 165−176.
(21) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(22) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(23) Schaefer, M.; Karplus, M. Comprehensive Analytical Treatment of Continuum Electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(24) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.
(25) *RASTOP*, v. 2.1; Valadon: La Jolla, CA.
(26) Maiti, R.; Van Domselaar, G. H.; Zhang, H.; Wishart, D. S. SuperPose: A Simple Server for Sophisticated Structural Superposition. *Nucleic Acids Res.* **2004**, *32*, W590−W594.
(27) McMillan, W. G.; Mayer, J. E. The Statistical Thermodynamics of Multicomponent Systems. *J. Chem. Phys.* **1945**, *13*, 276−305.
(28) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(29) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(30) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. Focusing of Electric Fields in the Active Site of Cu−Zn Superoxide Dismutase: Effects of Ionic Strength and Amino-Acid Modification. *Proteins: Struct., Funct., Genet.* **1986**, *1*, 47−59.
(31) Neal, B. L.; Asthagiri, D.; Lenhoff, A. M. Molecular Origins of Osmotic Second Virial Coefficients of Proteins. *Biophys. J.* **1998**, *75*, 2469−2477.
(32) Roth, C. M.; Neal, B. L.; Lenhoff, A. M. van der Waals Interactions Involving Proteins. *Biophys. J.* **1996**, *70*, 977−987.
(33) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441−451.
(34) Press, W. *Numerical Recipes in C − The Art of Scientific Computing*; Syndicate of the University of Cambridge: Cambridge, U. K., 2002.
(35) Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276−277.
(36) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C. The Hydrophobic Moment Detects Periodicity in Protein Hydrophobicity. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 140−144.
(37) Boman, H. G. Antibacterial Peptides: Basic Facts and Emerging Concepts. *J. Intern. Med.* **2003**, *254*, 197−215.
(38) Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. IUPred: Web Server for the Prediction of Intrinsically Unstructured Regions of Proteins Based on Estimated Energy Content. *Bioinformatics* **2005**, *21*, 3433−3434.
(39) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. Protein Disorder Prediction: Implications for Structural Proteomics. *Structure (Cambridge, MA, U.S.)* **2003**, *11*, 1453−1459.
(40) Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J. GlobPlot: Exploring Protein Sequences for Globularity and Disorder. *Nucleic Acids Res.* **2003**, *31*, 3701−3708.
(41) Puntervoll, P.; Linding, R.; Gemund, C.; Chabanis-Davidson, S.; Mattingsdal, M.; Cameron, S.; Martin, D. M.; Ausiello, G.; Brannetti, B.; Costantini, A.; Ferre, F.; Maselli, V.; Via, A.; Cesareni, G.; Diella, F.; Superti-Furga, G.; Wyrwicz, L.; Ramu, C.; McGuigan, C.; Gudavalli, R.; Letunic, I.; Bork, P.; Rychlewski, L.; Kuster, B.; Helmer-Citterich, M.; Hunter, W. N.; Aasland, R.; Gibson, T. J. ELM Server: A New Resource for Investigating Short Functional Sites in Modular Eukaryotic Proteins. *Nucleic Acids Res.* **2003**, *31*, 3625−3630.
(42) Obenauer, J. C.; Cantley, L. C.; Yaffe, M. B. Scansite 2.0: Proteome-Wide Prediction of Cell Signaling Interactions Using Short Sequence Motifs. *Nucleic Acids Res.* **2003**, *31*, 3635−3641.
(43) Plewczynski, D.; Tkacz, A.; Wyrwicz, L. S.; Rychlewski, L. AutoMotif Server: Prediction of Single Residue Post-Translational Modifications in Proteins. *Bioinformatics* **2005**, *21*, 2525−2527.
(44) McGuffin, L. J.; Bryson, K.; Jones, D. T. The PSIPRED Protein Structure Prediction Server. *Bioinformatics* **2000**, *16*, 404−405.
(45) Pham, T. H.; Satou, K.; Ho, T. B. Prediction and Analysis of Beta-Turns in Proteins by Support Vector Machine. *Genome Inf. Ser. n14 2003* **2003**, *14*, 196−205.
(46) Pham, T. H.; Satou, K.; Ho, T. B. Support Vector Machines for Prediction and Analysis of Beta and Gamma-Turns in Proteins. *J. Bioinf. Comput. Biol.* **2005**, *3*, 343−358.
(47) Kumar, M.; Bhasin, M.; Natt, N. K.; Raghava, G. P. BhairPred: Prediction of Beta-Hairpins in a Protein from Multiple Alignment Information Using ANN and SVM Techniques. *Nucleic Acids Res.* **2005**, *33*, W154−159.
(48) Kolaskar, A. S.; Tongaonkar, P. C. A Semi-Empirical Method for Prediction of Antigenic Determinants on Protein Antigens. *FEBS Lett.* **1990**, *276*, 172−174.
(49) *STATA8.0*; Stata Corporation: College Station, TX.
(50) Ramachandran, G. N.; Sasisekharan, V. Conformation of Polypeptides and Proteins. *Adv Protein Chem.* **1968**, *23*, 284−438.
(51) Ho, B. K.; Coutsias, E. A.; Seok, C.; Dill, K. A. The Flexibility in the Proline Ring Couples to the Protein Backbone. *Protein Sci.* **2005**, *14*, 1011−1018.
(52) George, A.; Wilson, W. W. Predicting Protein Crystallization from a Dilute Solution Property. *Acta Crystallogr.* **1994**, *50*, 361−365.
(53) Guo, B.; Kao, S.; McDonald, H.; Asanov, A.; Combs, L. L.; Wilson, W. W. Correlation of Second Virial Coefficients and Solubilities Useful in Protein Crystal Growth. *J. Cryst. Growth* **1999**, *196*, 424−433.
(54) Haas, C.; Drenth, J.; Wilson, W. W. Relation between the Solubility of Proteins in Aqueous Solutions and the Second Virial Coefficient of the Solution. *J. Phys. Chem. B.* **1999**, *103*, 2808−2811.
(55) Toke, O. Antimicrobial Peptides: New candidates in the Fight Against Bacterial Infections. *Biopolym. (Pept. Sci.)* **2005**, *80*, 717−735.

CI0600760