Structure−Property Model for Membrane Partitioning of Oligopeptides

Lene Hjorth Alifrangis,*[,†] Inge T. Christensen,[#,‡] Anders Berglund,[§] Maria Sandberg,[||] Lars Hovgaard,[†] and Sven Frokjaer[†]

Departments of Medicinal Chemistry and Pharmaceutics, The Royal Danish School of Pharmacy, Universitetsparken 2, DK-2100 Copenhagen, Denmark

Received June 4, 1999

The aim of this study was to develop a structure−property model for membrane partitioning of oligopeptides using statistical design methods and multivariate data analysis. A set of 20 tetrapeptides with optional N-methylations at residues 2 and 4 was designed by a *D*-optimal design procedure. After synthesis and purification, the membrane partitioning abilities of the peptides were tested in two chromatographic systems with phospholipids as the stationary phase: immobilized artificial membrane chromatography (IAM) and immobilized liposome chromatography (ILC). The relationship between these measures and three different sets of calculated descriptors was analyzed by partial least-squares projection to latent structures (PLS). The descriptors used were the molecular surface area, Molsurf parameters, and Volsurf parameters. All three models were of good statistical quality and supported that a large hydrogen-bonding potential and the presence of a negative charge impair membrane partitioning, whereas hydrophobic parameters promote partitioning. The findings are in accordance with what has been found for absorption of known drugs and have implications for the design of peptide-like drugs with good oral bioavailability.

Introduction

Often the barrier for a new drug candidate to reach the market is deficient pharmacokinetic properties, particularly absorption from the intestinal epithelium,[1,2] rather than lack of potency. These poor statistics have augmented the effort to develop new screening strategies so as to incorporate pharmacokinetic and biopharmaceutical considerations in the optimization of lead compounds. On the basis of the structure of well-known drugs, a number of successful approaches for predicting absorption from molecular structure have been developed. Depending on the structural diversity of the molecules under study, the absorptive properties may be predicted from the size of the molecular polar surface area alone[3,4] or in combination with other factors.[5] Even absorption in humans has been modeled by these methods.[6,7] However, when it comes to peptides and peptidic drugs, the understanding is still somewhat lacking. The potential for this class of drugs is inherently growing as more natural, potent peptidic ligands are being discovered and the biotechnological capability is growing.[8] To rationally develop well-absorbed peptidic drugs, an improved comprehension of the relationship between the peptide structure and the absorption process is necessary.

Generally, the most important route of drug absorption is passive diffusion through the epithelial cells (transcellular transport). One of the crucial steps in this process is the partitioning of the drug from the extracellular aqueous environment, into − and through − the lipophilic cellular membrane. The present understanding of peptide absorption implies that the hydrogen-bonding potential is a major determinant for membrane permeation. The hydrogen-bonding potential is inversely correlated to membrane permeation, as there is a large energy penalty for breaking the hydrogen bonds between the solute and the aqueous environment before transport through the membrane.[9] It has been found that sequentially reducing the number of hydrogen bond donors by N-methylating the backbone amides within a homologous series increases epithelial permeation in Caco-2 cells[10] and rats.[11] Similarly, N-methylation has recently been used to improve the oral bioavailability in a novel series of growth hormone-stimulating peptides.[12] Additionally, conformational preferences,[13] size, and overall charge[14] also play a role. Eventually, all these factors, which are interrelated to some degree, depend on the amino acid sequence. In summary, the present evidence suggests that the factors known to govern absorption of known drugs, e.g. the polar surface area, which is related to the hydrogen-bonding potential, also influence peptide absorption and thus membrane partitioning. In support of this view, Barlow and Satoh[15] have estimated limiting polar and total surface areas for membrane permeation. However, these estimates may be further refined as they were based on the traditional "ideal" log $P_{octanol}$ of 2−3 for allowing membrane transport. Generally, a poor correlation between log $P_{octanol}$ and membrane transport has been found for peptides.[9,10] Recently, Stenberg et al.[16] have proposed a predictive model for Caco-2 cell permeability of dipeptides using a combination of polar and nonpolar surface areas. However, it is uncertain whether this model can be extrapolated to new peptidic drugs, which

* To whom correspondence should be addressed. Tel: +45 35 30 64 90. E-mail: lhk@mail.dfh.dk.
[#] Department of Medicinal Chemistry.
[†] Department of Pharmaceutics.
[‡] Present address: Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Maaloev, Denmark.
[§] Research Group for Chemometrics, Department of Organic Chemistry, Umeå University, S-901 87 Umeå, Sweden.
[||] Umetri AB, Box 7960, S-907 19 Umeå, Sweden.

10.1021/jm9910932 CCC: $19.00
Published on Web 12/17/1999

often have 3−6 amino acids in the basic structure. Thus, the aim of the present study is to develop a structure−property model for membrane partitioning for this group of peptides using statistical design methods and multivariate data analysis.

Methodology

De Novo Design of Model Peptides. Apart from the unspecific knowledge that more lipophilic molecules are better absorbed, no real "lead compound" exists with respect to membrane partitioning. As opposed to the specific molecular recognition involved in receptor binding, one end of the molecule is presumably not more important than the other in this nonspecific interaction. When no parts of the molecule need to be held fixed in order to fit into a pharmacophore model, it allows for simultaneous variation of all parts of the molecule. Thereby, the pitfalls of changing one factor at a time and finding false optima are avoided.[17]

A fixed number of 4 amino acids for all peptides was chosen in order to represent the majority of new peptidic or pseudopeptidic drugs typically having 3−6 amino acids in the basic structure. Furthermore, it keeps the number of necessary design factors at a minimum while allowing to see distinct secondary structure, such as β-turns, in solution. The number of peptides to be synthesized and tested was limited to 20. To study the interaction between sequence, β-turn formation, and N-methylations, an option for N-methylation of residues 2 and 4 was included in the design. At residue 4, the formation of an internal hydrogen bond in a β-turn is impaired by an N-methylation, which in return reduces the hydrogen-bonding potential directly. The amide hydrogen at residue 2 is not directly involved in stabilizing the β-turn. Rather, the consequence of an N-methylation in this position is a reduction of the hydrogen-bonding potential, but it probably also leads to a secondary alteration of the conformational preferences and, thus, has potential for stabilizing/destabilizing other conformations with internal hydrogen bonds.

Each amino acid in the peptides was described by three principal properties, z_1-z_3, thus adding up to 12 descriptors per peptide. These descriptors are part of an updated version of the *z*-scales previously published by Hellberg et al.[18] that now includes five principal properties (z_1-z_5) for 87 natural and unnatural amino acids.[19] The interpretation of the three major properties is the same as in the original *z*-scales: i.e. z_1 describes hydrophilicity, z_2 describes size/polarizability, and z_3 is interpreted as electronic effects.

Design Procedure. When varying 4 positions in a tetrapeptide using the 20 naturally occurring amino acids and multiplying with the 4 permutations of N-methylations, it adds up to 640 000 sequences. This large number constitutes the candidate set from which the 20 representative peptides of the training set should be selected. Due to the very reduced and constrained nature of the problem (see below), it was not feasible to use a fractional factorial design. Instead, a computer-generated *D*-optimal design was used.[20] A *D*-optimal algorithm is an exchange algorithm, which picks out experimental runs from the candidate set for which the determinant *D* of the $\mathbf{X'X}$ matrix is maximized.[21] To increase the probability of finding the global optima it has to be repeated several times. As opposed to the factorial designs, it utilizes the quantitative information contained in the design variables, not just the binary information of + and −.

Table 1. Principal Properties of the 20 Genetically Coded Amino Acids[19]

no.[a]	amino acid	z_1	z_2	z_3	no.[a]	amino acid	z_1	z_2	z_3
1	Asp	3.98	0.93	1.93	**11**	Ala	0.24	−2.32	0.60
2	Gln	1.75	0.50	−1.44	**12**	Arg	3.52	2.50	−3.50
3	Gly	2.05	−4.06	0.36	**13**	Asn	3.05	1.62	1.04
4	Leu	−4.28	−1.30	−1.49	**14**	Cys	0.84	−1.67	3.71
5	Lys	2.29	0.89	−2.49	**15**	Glu	3.11	0.26	−0.11
6	Phe	−4.22	1.94	1.06	**16**	His	2.47	1.95	0.26
7	Pro	−1.66	0.27	1.84	**17**	Ile	−3.89	−1.73	−1.71
8	Ser	2.39	−1.07	1.15	**18**	Met	−2.85	−0.22	0.47
9	Tyr	−2.54	2.44	0.43	**19**	Thr	0.75	−2.18	−1.12
10	Val	−2.59	−2.64	−1.54	**20**	Trp	−4.36	3.94	0.59

[a] Numbers **1**−**10**: amino acids used for the design. Numbers **11**−**20**: amino acids not used for the design.

Due to the excessive data handling of finding the *D*-optimal of a 640 000 × 12 matrix, the candidate set of 640 000 was reduced to 4410 by applying the constraints listed below:

a. Selection of 10 out of the 20 genetically coded amino acids based on the distribution of their principal properties, z_1-z_3 (cf. Table 1).

b. Minimum one aromatic residue per peptide for sensitive detection by fluorescence.

c. Maximum one ionizable residue of each sign per peptide to reduce overall charge.

d. Minimum three different amino acids per peptide.

e. No proline in positions 1 and 4.

f. Assignment of N-methylations by randomization after completion of design.

The experimental space under study is thus restricted by the vertexes made up of the highly irregular hypercube defined by the 4410 × 12 matrix.

In preliminary *D*-optimal designs, 18 out of 20 peptides contained at least one ionizable residue. It is very likely that the presence of a charge at physiological pH impairs diffusion across a lipophilic membrane, so a training set with 18 charged peptides would cover a narrow range of absorption properties. To overcome this, a stepwise procedure was used. In step 1, 13 neutral peptides were selected from the neutral sequences of the candidate set. These 13 sequences were used as inclusions in step 2 in which 4 monocharged peptides were added (finding the *D*-optimal of 17 peptides of which 13 are default). Finally, the 13 + 4 peptides were used as inclusions in step 3 selecting the 3 zwitterionic peptides. The *D*-optimal algorithm was repeated at least 40 times, and the training set with the combination of a good *G*-efficiency (>50%[21]) and a good distribution of polar surface area and van der Waals volume was passed on to the next step. The latter criterion was applied to ensure an adequate variation in two of the major factors determining absorption. It partially compensates for the fact that no direct information about the overall properties of the peptides is contained in the design when designing from the fragment descriptors. The final *G*-efficiency of 46.3% is slightly below 50%, which by rule of thumb is the threshold for an acceptable *D*-optimal design. Nevertheless, this is the best design obtainable for this strongly reduced and con-

Table 2. Characteristics of the 20 Model Peptides

peptide	sequence	MW	Mlogp[a]	log K_s[b]	log k'_{IAM}[c]	A1[d]	B1[e]	log $D_{7.4}$[f]	pK_{a1}[g]	pK_{a2}[h]
I	Ac-Gly-Leu-Asp-NMePhe-NH_2	505.58	−0.18	−0.15	−0.55	29.3	28.7		3.93	
II	Ac-Phe-Pro-Phe-Tyr-NH_2	613.73	1.19	1.54	1.20	35.9	35.4	1.64		
III	Ac-Gly-NMeGly-Leu-Phe-NH_2	447.55	0.09	0.59	−0.12	29.7	28.7	0.05		
IV	Ac-Lys-NMeGly-Tyr-NMeAsp-NH_2	550.63	−1.13	−0.54	−1.09	11.0	nd[i]		3.60	9.69
V	Ac-Tyr-Gly-Gly-Gln-NH_2	465.10	−1.90	0.23	−0.54	10.7	10.7	−2.59		
VI	Ac-Asp-Gln-Leu-Phe-NH_2	562.64	−0.86	−0.16	−0.69	26.2	26.2	>−2.50	3.97	
VII	Ac-Gly-NMePhe-Lys-Asp-NH_2	519.59	−0.90	−0.73	−1.09	16.0	17.7		3.78	10.31
VIII	Ac-Ser-NMePhe-Gly-Gly-NH_2	421.47	−1.30	0.12	−0.66	16.9	16.7	−1.51		
IX	Ac-Lys-NMePhe-Gly-NMeLeu-NH_2	532.15	0.41	0.57	0.05	26.4	28.0	−1.33		9.94
X	Ac-Gln-Gln-Phe-NMeGly-NH_2	534.03	−1.67	−0.19	−0.93	15.8	16.0			
XI	Ac-Phe-Gln-Lys-NMeLeu-NH_2	589.20	−0.29	0.19	−0.16	23.1	24.6			10.06
XII	Ac-Phe-Leu-Val-NMeLeu-NH_2	546.65	1.53	1.17	0.58	40.3	40.3	2.42		
XIII	Ac-Phe-NMeGln-Gln-NMeSer-NH_2	577.65	−1.96	−0.12	−0.82	18.0	17.7			
XIV	Ac-Gly-NMeAsp-Phe-NMeLys-NH_2	534.63	−0.69	−0.58	−1.05	18.1	19.8		nd[i]	nd[i]
XV	Ac-Val-NMeGln-Ser-NMePhe-NH_2	534.63	−0.89	0.14	−0.44	22.6	25.0			
XVI	Ac-Val-NMeTyr-Leu-Gln-NH_2	576.70	−0.22	0.39	0.07	26.3	26.2			
XVII	Ac-Ser-NMeLeu-Tyr-Gln-NH_2	564.70	−1.22	0.24	−0.23	19.6	19.5			
XVIII	Ac-Gln-Pro-Gln-Phe-NH_2	559.64	−1.65	−0.03	−0.75	17.8	17.9			
XIX	Ac-Leu-Val-Tyr-Gly-NH_2	491.60	0.06	0.39	0.03	23.8	23.7			
XX[j]	Ac-Gly-Ser-Tyr-NMeVal-NH_2	479.55	−1.08							

[a] Log $P_{octanol/water}$ calculated according to Moriguchi.[43] [b] Capacity factor on the ILC column. Average standard deviation of K_s was 0.02 log unit. [c] Capacity factor on the IAM column. Standard deviations were similar as for K_s. [d] Retention time in chromatography system A1 (cf. Experimental Section). [e] Retention time in chromatography system B1 (cf. Experimental Section). [f] Distribution coefficient between *n*-octanol and Hanks buffered saline solution, pH 7.4, determined by the shake flask method. [g,h] pK_a values determined by potentiometric titration.[33] [i] nd: not determined. [j] The N-methylation of valine was not accomplished, and thus, this peptide was excluded from the statistical analysis.

strained design problem. The sequences of the peptides are listed in Table 2 and are hereafter referred to by their roman numerals. The peptides were N-terminally acetylated and C-terminally amidated in order to reduce overall charge and reduce the risk of enzymatic degradation in biological systems.

Synthesis. The peptides were synthesized by solid-phase synthesis using Fmoc chemistry and a Rink amide resin as the solid support. HoBt (or HoAt) and EDC (or DIC) were used as coupling reagents. Where available, Fmoc N-methylated amino acids were used. Otherwise, an on-site N-methylation procedure was used.[22] Methylation was accomplished for serine, glutamine, tyrosine, and lysine, but not valine (in compound **XX** (cf. Table 2)). A resynthesis of peptide **XX** was not attempted. After purification by semipreparative HPLC, the purity was confirmed in two different chromatographic systems (as judged from the UV trace at 214 nm). The identity of the peptides was confirmed by LC−MS, PD−MS, and amino acid analysis.

Membrane Partitioning. The ability of the 19 peptides to interact with membrane phospholipids was assessed in two chromatographic systems: immobilized liposome chromatography (ILC)[23] and immobilized artificial membrane (IAM)[24] chromatography. The chromatographic capacity factors, K_s and k'_{IAM}, for the IL and IAM columns, respectively, were calculated from the retention time of the peptides recorded relative to a void volume marker. Theoretically, partitioning in the ILC system consisting of multi- and bilayered liposomes entrapped in a chromatographic gel is a better model of the cellular membrane than IAM chromatography where the phospholipid derivatives are forming a monolayer, being covalently linked to silica particles. However, from a practical point of view the IAM column is more attractive, as each ILC column needs to be individually prepared whereas the IAM column is commercially available. In case there is a difference between them when measuring partitioning of peptides − as has been suggested for small molecule drugs[23] − the two systems may contain complementary information relevant for the transcellular transport process.

In general, K_s is a good predictor for transport in Caco-2 cells as has been shown for a series of structurally diverse drugs.[23] As for IAM, the correlation to in vivo absorption is reasonable (Caco-2 permeability $R^2 = 0.56$, percent absorbed in rat perfusion $R^2 = 0.63$) for structurally diverse compounds.[24] This is superior over k'_w (the capacity factor from an octadecylsilyl column often used to estimate log $P_{octanol}$), but clearly, some properties relevant for membrane diffusion are not measured by k'_{IAM}. This may be due to the different composition of phospholipids between the immobilized membrane and living cells, interaction with other membrane components such as proteins, and perhaps also the fact that positive membrane interactions are not necessarily promoting diffusion across the membrane. Similar considerations hold for K_s.

Theoretical Characterization. The peptides were multivariately characterized by (1) calculated molecular surface properties: polar, nonpolar, and total water-accessible surface area; (2) Molsurf[25] parameters: physicochemical characteristics derived from a quantum mechanical wave function; and (3) Volsurf[26] parameters: characteristics derived from calculations of interaction energies with a water and a lipophilic probe using GRID[27] (cf. Experimental Section). The Volsurf descriptors are listed in Table 4.

Statistical Analysis. The membrane partitioning ability of the 19 peptides as measured by log K_s and log k'_{IAM} was related to the various molecular descriptors by means of partial least-squares projections to latent structures (PLS)[17] using the statistical software Simca-P.[28] As described above, peptide **XX** was not synthesized, and thus, it was excluded from the statistical analysis. Before analysis, all variables were scaled to unit variance and centered (autoscaling). To achieve a normal distribution of the log K_s values they were transformed as log(measured K_s + 0.3). The quality of the models was expressed as the explanatory power, R^2,

Table 3. List of Descriptors Used for the Multivariate Analysis

model no.	descriptors
1a	z_1, z_2, and z_3 for all four amino acids
1b	z_1 (hydrophilicity) for all four amino acids
2a	PWASA,[a] NPWASA, TWASA, P/NWASA, P/TWASA, charge (0/1), presence of N-methylations (0/1)
2b	PWASA, NPWASA, NPWASA^2, negative charge, N-methylation at position 4 (0/1)
3a	all Volsurf parameters[b]
3b	Volsurf parameters: V, V^2, W_5, Cw_5, D_1, D_2, ID_1, CP, CP^2
4a	all Molsurf parameters[c]
4b	Molsurf parameters: log P, log D, pK_a (acid), polarizability, Lewis base
ref	all parameters from models 1b, 2b, 3b, and 4b

[a] Cf. list of abbreviations and Supporting Information. [b] Cf. Table 4. [c] Cf. Supporting Information.

and the predictive power, Q^2, based on cross-validation[29] by leaving one-seventh of the data out.

Generally, variable selection was accomplished by excluding all variables with a VIP value (variable importance in the projection) below 1 and, subsequently, keeping only the variables, which induced an increase in the predictive value of the model. The VIP value is a weighted sum of squares of the PLS weights, w, taking into account the amount of explained y variance of each PLS dimension according to:[30]

$$\mathrm{VIP}(k) = [[\sum_a [(w_{ak}{}^2) \times \mathrm{SSY\%}(a)]]/\mathrm{SSY\%(cum)}] \times K$$

where VIP(k) is the VIP value for variable k, $w_{ak}{}^2$ is the squared PLS weight of variable k for component a, SSY%(a) is the percentage sum of squares of y explained by component a, SSY%(cum) is the total percentage explained sum of squares for the entire model, and K is the number of variables.

Furthermore, from variables with overlapping PLS loadings (signifying essential identical information content), a representative with the largest VIP value was chosen. Next, to account for undetected nonlinear relationships, it was attempted to include squared terms based on all the remaining variables. Again, only the squared terms with VIP > 1 were kept in the model. After this final optimization, the y-values were randomized 10 times and models using the same x-variables as the original model were built. The validity of the models was additionally tested by a permutation test.[30] When plotting R^2 and Q^2 as a function of the correlation coefficient between the original values and the predicted values, the intercept with the y-axis expresses to which degree these values rely on chance. Generally, if the model is valid, the intercept for Q^2 should be negative and for R^2 it should be below 0.3.[30]

Results and Discussion

Design. When designing from molecular fragment descriptors it is assumed that the sum of the parts adds up to give the properties of the entire peptide. Furthermore, it is anticipated that a statistically balanced composition of amino acids also leads to a balanced distribution of membrane partitioning abilities. However, the distribution of K_s and k'_{IAM} values (cf. Table 2) indicates that this is not the case. With the exception of the two peptides with the fewest hydrophilic side chains (**II** and **XII**), the peptides have capacity factors within a relatively narrow range. Even though the designed peptides have a balanced variation of residues, they seem to have a bias toward overall hydrophilic properties. The common hydrophilic scaffold of the backbone polar atoms may explain this bias. These atoms are major contributors to the total hydrogen-bonding potential, and so the observable effect of a particular side chain on partitioning is biased by this intrinsic hydrophilicity of a tetrapeptide. Beforehand, it was largely unknown how the nature and position of particular side chains can counteract this basic hydrophilicity, and thus it was not possible to include this information in the design. However, this training set provides a realistic basis for predicting the membrane partitioning properties of tetrapeptides in general. A generalizable model will be useful for optimizing the membrane partitioning abilities of new peptidic drugs, which use the natural peptide ligand as a starting structure.

ILC versus IAM. Despite the different structures of the stationary phase of the IL and IAM columns, the capacity factors from the two columns are highly correlated as shown in Figure 1. However, they are not numerically identical (cf. Table 2), which is probably due to different amounts of lipids available for partitioning. K_s is corrected for the amount of immobilized phospholipids, whereas k'_{IAM} is not. The apparent similarity of the columns may be explained by a combination of the similarity between the headgroups of the phospholipids (phosphatidylcholine) and the hydrophilicity of the peptides. If the peptides are predominantly partitioning to the vicinity of the headgroup region (which is similar), the thickness or number of phospholipid layers is of minor importance. If a larger range of capacity factors is included the difference is probably more pronounced.[23]

Model Building. As shown in Table 3, four models have been built from the various sets of parameters described in Methodology. The suffices a and b denote increasingly improved models in which the number of variables has been optimized and reduced to give the best model possible with a particular set of descriptors. In Table 5, the statistical quality of the models is shown. All models have good explanatory powers (R^2) and are low dimensional; i.e. they have only one or two components. The components (or latent variables) are linear combinations of the original variables. The influence from each variable is expressed as its loading. In the optimized versions of models 2−4, excellent predictive powers (Q^2) in the range 0.8−0.9 have been obtained for log K_s as well as for log k'_{IAM}. Model 1 employing the amino acid descriptors z_1−z_3 has the lowest quality of all the models before and after optimization. However, it is remarkable that a relatively good predictive power is obtained if using only the hydrophilicity (z_1-values) of the individual amino acids as in model 1b. Nevertheless, the poorer quality corroborates the above discussion that the properties of the individual amino acids are not sufficient for describing the whole peptide.

In Figure 2a−d the PLS loadings for all four, nonoptimized models are depicted. The parameters constituting the final, optimized models are highlighted with italics. In all the two-component models (Figure 2b−d), log K_s and log k'_{IAM} are located in the upper, right

Table 4. Definition and Significance of Volsurf Parameters

parameter	definition	VIP > 1 in model 3a	significance in model 3b
V	total water-accessible volume at 0.20 kcal/mol	V	V, V^2
S	total water-accessible surface area at 0.20 kcal/mol	none	
R	total volume/total surface area, V/S	none	
G	globularity; total surface area/surface area of sphere with volume V	none	
W_1-W_8	hydrophilic regions at 8 levels of interaction energy with water probe (−0.2, −0.5, −1.0, −2.0, −3.0, −4.0, −5.0, −6.0 kcal/mol)	W_2-W_7	W_5 (−3.0 kcal/mol)
Iw_1-Iw_8	hydrophilic integy moment at 8 energy levels as above; measure of the unbalance between the center of mass and the position of hydrophilic regions	none	
Cw_1-Cw_8	capacity factors at 8 energy levels as above; ratio between the hydrophilic regions and total surface area, W/S	Cw_1-Cw_8	Cw_5 (−3.0 kcal/mol)
D_1-D_8	hydrophobic regions at 8 levels of interaction energy with DRY probe (−0.2, −0.4, −0.6, −0.8, −1.0, −1.2, −1.4, −1.6 kcal/mol)	D_1, D_2	D_1, D_2
ID_1-ID_8	hydrophobic integy moments at 8 energy levels as for D_1-D_8; measure of the unbalance between the center of mass and the position of hydrophobic regions	ID_1, ID_2	ID_1
$E_{min1}-E_{min3}$	3 lowest energy minima for interaction energy with water probe	none	
D_{12}, D_{13}, D_{23}	distances between $E_{min1}-E_{min3}$	none	
HL_1, HL_2	hydrophilic−lipophilic balance; ratios W_4/D_3 and W_5/D_4	HL_1	
A	amphiphilic moment; length of a vector pointing from the center of the hydrophobic domain to the center of the hydrophilic domain	none	
CP	critical packing parameter defined as: (volume of hydrophobic part at −0.6 kcal/mol)/ [(surface of hydrophilic part at −3.0 kcal/mol)* (length of hydrophobic part at −0.6 kcal/mol)]	CP	CP, CP^2
POL	average molecular polarizability calculated from an additive method	POL	

Table 5. Statistical Quality of the Models

model no.	N^a	A^b	RMSE^c (log K_s)	RMSE^c (log k'_{IAM})	$R^{2\,d}$ (overall)	$Q^{2\,e}$ (overall)	$R^{2\,f}$ (intercept)	$Q^{2\,f}$ (intercept)
1a	12	1	0.32	0.36	0.69	0.23	0.55	0.03
1b	4	1	0.33	0.27	0.75	0.70	0.09	−0.10
2a	10	2	0.22	0.28	0.84	0.66	0.26	−0.13
2b	5	1	0.18	0.25	0.87	0.83	0.08	−0.19
3a	39	2	0.20	0.16	0.87	0.70	0.45	−0.12
3b	9	2	0.17	0.13	0.94	0.90	0.23	−0.14
4a	13	2	0.25	0.27	0.85	0.67	0.32	−0.12
4b	5	1	0.17	0.23	0.86	0.79	0.18	−0.16
ref	23	1	0.12	0.15	0.94	0.89	0.29	−0.19

[a] N: number of descriptors in the model (for details cf. Table 3). [b] A: number of significant PLS components determined by cross-validation. [c] RMSE: residual mean square error. [d] R^2 (overall): degree of variation explained by the model (combined for log K_s and log k'_{IAM}). [e] Q^2 (overall): predictive power (combined for log K_s and log k'_{IAM}). [f] R^2 and Q^2 (intercept): intercept with the y-axis after permuation of log K_s (cf. Methodology); permutation of k'_{IAM} gave similar results.

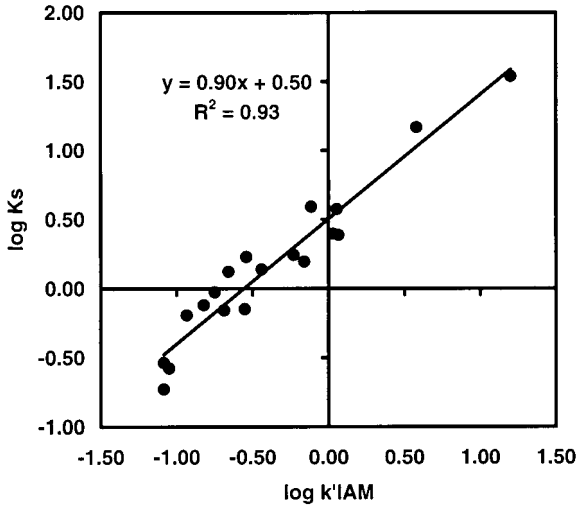

Figure 1. Correlation between the capacity factors log k'_{IAM} and log K_s, determined on an IAM column and an ILC column, respectively, for the 19 peptides.

quadrant. The parameters with influence on the first PLS component are found in the horizontal direction and on the second component in the vertical direction. Parameters far away from the origin in either direction have a large influence on the particular component and vice versa.

In all models, the first component may be interpreted as a hydrophilicity/hydrophobicity axis. Although the hydrophilic/hydrophobic factors are not loading in exactly the same direction in all models due to slightly different information content in these factors combined with the influence from the second component, it is possible to derive some general trends from the models. In the coefficient plots for the optimized models in Figure 2e−h, the parameters have been divided into two groups: those advantageous and those disadvantageous for membrane partitioning, having positive and negative regression coefficients, respectively. Regression coefficients from PLS are interpreted in the same way as regression coefficients in multiple linear regression with the important notion that they are not independent. The coefficients from e.g. model 3b define the linear relationship between log K_s and the Volsurf parameters:

$$\log K_s = 0.290 - 0.223W_5 - 0.310Cw_5 - 0.288ID_1 + 0.222V + 0.077V^2 + 0.194D_1 + 0.012D_2 + 0.040CP + 0.073CP^2$$

(scaled and centered variables).

The factors impairing membrane partitioning are related to hydrophilicity and hydrogen-bonding factors, such as the polar surface area (model 2) and large

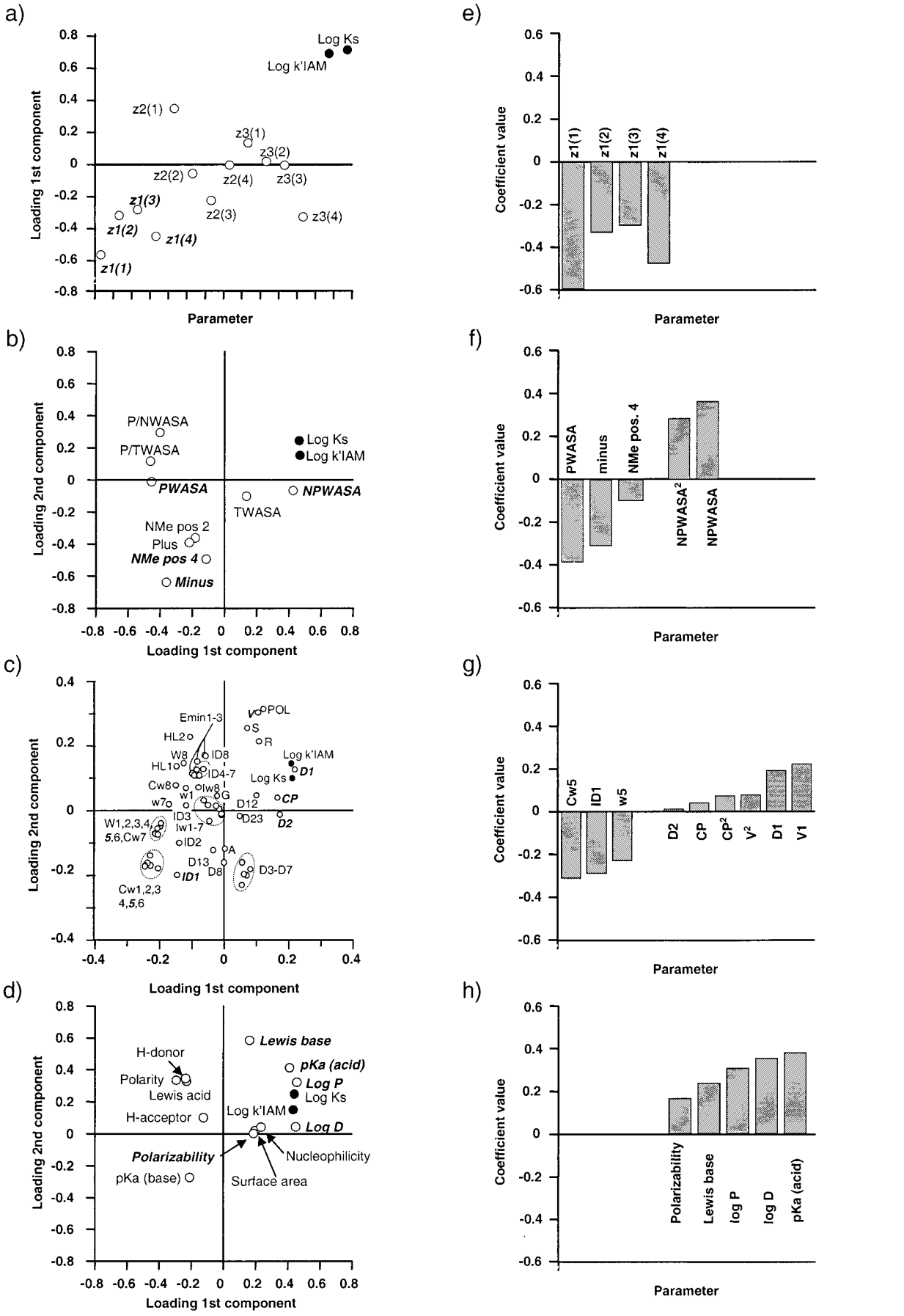

Figure 2. (a−d) PLS loadings for the models 1a, 2a, 3a, and 4a. (e−h) Plots of autoscaled PLS regression coefficients for models 1b, 2b, 3b, and 4b (optimized models).

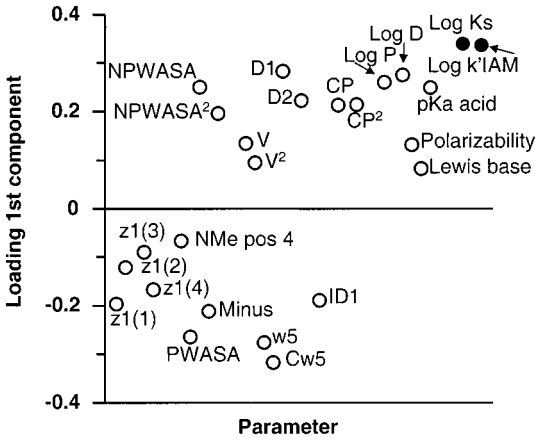

Figure 3. PLS loadings for the reference model with all the descriptors from the optimized models (1b, 2b, 3b, and 4b).

hydrophilic regions defined by the interaction with the water probe at −3.0 kcal/mol (model 3). After optimization, model 4 does not contain any factors impairing membrane partitioning although Figure 2d indicates a negative, nonsignificant influence of Lewis acid and hydrogen bond donor abilities. Partitioning is promoted by factors related to lipophilicity and size. In model 2, this is expressed as a large nonpolar surface area and in model 4 by high log *P*/log *D* values and polarizability. In model 3 these factors are modeled by large values of the critical packing factor and large hydrophobic areas measured by interactions with the DRY probe at levels −0.2 and −0.4 kcal/mol. In model 3b, the energy levels for the factors chosen for the model should not be interpreted too strictly. To reduce the number of factors, a few representatives with the largest VIP were chosen from the clusters of overlapping factors although they were almost equally good (cf. Figure 2c). Initially, it was attempted to build a reference model using all descriptors from the nonoptimized models and excluding nonsignificant variables as described in Methodology. However, the model was not superior to model 3b, and many of the minor factors in each model were not significant according to the VIP criteria (model not shown). Thus, to compare all the factors in the optimized models, a reference model was built using all the factors without further optimization. As can be seen from Figure 3, depicting the loadings of the reference model, the same factors as discussed above display positive and negative loadings and promote and impair partitioning, respectively.

The validity of the models is strongly substantiated by the fact that three different methods for molecular characterization point to the same overall properties as being important. As would be expected, a combination of hydrophilic and lipophilic factors is necessary to explain the partitioning behavior of the peptides. More importantly, the models relate to the relative importance of these factors. Within homologous series of compounds it has often been found that one single parameter such as the traditionally used log $P_{octanol}$ or the polar surface area suffices to explain the absorption/partitioning. However, it is well-known that for peptides (and several other classes) a poor correlation exists between log $P_{octanol}$ and absorption.[9,12,13] Also, if using more diverse compounds, a combination of hydrophilic and hydrophobic parameters is necessary as demonstrated in recent reports.[6,16] Using a linear combination of the polar and nonpolar surface areas, Stenberg et al. achieved a good predictive model for transport of dipeptides in Caco-2 cells[16] and Winiwarter et al. modeled the fraction of drugs absorbed in humans using a combination of the number of H-bond donors, polar surface area, and calculated log $P_{octanol}$.[6]

Effect of Charge. As judged from model 2, a negative charge is strongly unfavorable for partitioning, whereas a positive charge has a minor effect as it is not a significant term in the model. This finding is supported by model 4 in which partitioning is promoted by a high pK_a value for acids, i.e. acids which are not ionized at physiological pH. Analogously, the pK_a value for bases corresponding to a positive charge at pH 7.4 was not significant in the model.

A similar difference between the opposite charges was found by Pauletti et al.[14] studying IAM retention times of peptides carrying positive, negative, or no charges. Similarly, k'_{IAM} of bases has been found to be largely independent of the degree of protonization.[31] Furthermore, Austin et al.[32] found that primary amines have a much stronger affinity for liposomes than carboxylates with similar structure. Recently, Avdeef et al.[33] have contributed to a more profound understanding of this phenomenon. They found that the charged species of amphiphilic acids have less affinity for PC-based liposomes than bases have, compared to their corresponding neutral form. Their concept of ionized bases, better complementing the charge/H-bond structure of the PC membrane than ionizable acids, may also be used to explain the present findings: when an amine with its hydrophobic parts buried in the membrane becomes ionized, it moves within the membrane as the primary site of interaction changes from a hydrogen-bonding interaction between the neutral amine and the C═O group of the headgroup to an electrostatic interaction between the charged amine and the negative phosphate group.[33] However, moving a negative charge from the C═O group to the positively charged, more superficially positioned choline group is energetically less favorable. This was explained by the longer distance of movement and the requirement for a positive counterion located deeper within the membrane at the negative phosphate group.[33]

Effect of Overall Size. In model 2a, the total size as expressed by TWASA has a minor, positive effect on partitioning but is not significant in the optimized model. However, in models 3b and 4b the total size expressed as the molecular volume and the molecular surface area (which may be used interchangeably with polarizability in this case), respectively, has a major positive influence. These size measures may be interpreted as lipophilicity terms, which are not necessary in model 2 − suggesting that in this model this information is already more strongly expressed in the other factors (PWASA, NPWASA). In conclusion, it is probably not the size as such which promotes partitioning as the diffusion coefficient within the membrane is inversely related to the size.[34] Rather, the size is correlated with lipophilicity, and so an apparent positive effect is seen.

Effect of N-Methylation. Model 2a suggests that N-methylation in positions 2 and 4 impairs partitioning.

This is in contrast to what would be expected from studies in which increasing the number of N-methylations in a homologous series of peptides increased transepithelial transport in Caco-2 cells[10] and rats.[11] The N-methylations load in the same direction as the polar surface area (Figure 2b), and the effect may thus be interpreted as an increased hydrophilicity of the peptide. However, in this low-dimensional model, the effect is mixed up with effects of the molecular size parameters, and therefore, a detailed interpretation should be cautious. It may be speculated that a positive effect of N-methylation is only seen when the neighboring amino acids are lipophilic and/or the overall hydrophilicity of the peptide is low. To further elucidate the matter, a new, extended design with room to determine interaction effects should be made.

Critical Packing Factor. The larger RMSE (residual mean square error) of models 2 and 4 as compared to model 3 is mainly caused by a large residual of peptide **II**. This peptide has a K_s value at least 2 times larger than the rest of the set. Only when including at least one of two shape/hydrophobicity factors (*CP* and D_2) from the Volsurf set in models 2 and 4 is the residual of **II** decreased to the same level as in model 3. Due to the proline in position 2, **II** has a kink and is thus more folded and globular than the remaining peptides. For peptides being simultaneously globular and lipophilic, *CP* seems to be a good descriptor of their membrane partitioning abilities.

Effect of Conformation. The size of the polar and nonpolar surface area varies considerably with the conformation.[3] Using just one conformation for these calculations is nevertheless yielding a model of high quality, which suggests that it may not be necessary to include the conformational variability to explain the partitioning behavior. On the other hand, the fact that the molecular surface properties are missing the information equivalent to *CP* may be an effect from using a single conformation. Hypothetically, the conformationally dependent, Boltzmann averaged molecular surface properties[3,4,8] might give a better representation of the molecule, particularly **II**. To elucidate this, a conformational analysis has been performed for six of the peptides covering the entire range of K_s values (**II, V, VIII, IX, XII, XIIII**); 1000 different conformations were generated via high-temperature molecular dynamics (MD) as described in the Experimental Section. No significant difference was found between the PWASA of the single conformation used and the Boltzmann averaged PWASA. Based on this data set, it seems unlikely that the information contained in *CP* can be derived from the conformationally dependent molecular surface properties. However, the interpretation should be cautious as the conclusion is based mainly on **II** for which *CP* has a big contribution.

Comparison of Methods. Model 1b with the z_1 descriptors of the individual amino acids is a very fast method, but its predictive power is not sufficient for accurate predictions. Model 2 is also very rapid, as only one calculation is needed to derive the molecular surface properties of a compound. However, the information content of these properties may not be sufficient for all peptides, as was demonstrated in the case of **II**. Whether this is a general phenomenon or **II** is belonging to a special class of lipophilic proline-containing peptides still has to be shown. In previous studies, the Molsurf parameters used in model 4 have performed well for predicting blood−brain barrier permeability[35] and transport across Caco-2 cells.[36] In this study, Molsurf had a performance comparable to that of the molecular surface properties. Molsurf has the advantage of giving an output of traditional physicochemical properties, which are intuitively easy to comprehend. The major drawbacks are the CPU-time-demanding semiempirical and ab initio calculations. Furthermore, the calculated log *P* and log *D* values deviate up to 2 orders of magnitude from the experimental values (cf. Table 2 and Supporting Information). Unquestionably, Molsurf log *P* and log *D* are important lipophilic parameters as they are the most important factors in model 4. They may be used as such in the model but should not be interpreted as octanol/water partition and distribution coefficients in absolute terms.

With respect to computational effort, Volsurf is between the molecular surface approach and Molsurf. As Volsurf is based on GRID,[27] these parameters are mainly limited by the parametrization of GRID. The advantage of Volsurf as compared to the molecular surface approach is that the atoms are ranked according to the strength of potential hydrogen bonds. The polar and nonpolar surface areas are solely discriminated by the definition of hydrogen-bonding atoms, which are all given the same weight. Thus, for novel compounds with unknown H-bonding properties, the definition may be too crude, as exemplified previously.[3] If parametrization is a problem, then the Molsurf ab initio approach may be more appropriate despite its limitations.

Conclusion

Although a strong inverse relation to hydrogen-bonding factors and membrane partitioning was found, there is still a possibility that K_s underestimates the negative influence of hydrogen bonding on diffusion through the membrane. The lipophilic part of the peptides may be positioned within the membrane while the backbone amides are hydrogen bonding with the polar headgroups − aligning the backbone parallel to the plane of the bilayer.[37] This would promote membrane partitioning but not necessarily diffusion across the membrane. A finding like that would conform with the hydrogen-bonding potential hypothesis stating that the rate-determining step for membrane passage of peptides is the breaking of the hydrogen bonds between the solute and the phospholipid headgroups.[9−11]

It was possible to set up a statistical design using 19 (20) peptides as representatives for those of the 640 000 possible N-methylated tetrapeptides falling within the criteria for the constraints. On the basis of this rational design, three high-quality structure−property models for membrane partitioning of small peptides could be proposed. All three models indicated that the molecular properties known to be important for absorption of known drugs are also important for membrane partitioning of peptides. The models established here may be used for improving the membrane partitioning properties of new peptidic drugs. Although it is obvious that exchanging a hydrophilic amino acid with a hydrophobic one improves membrane partitioning, the

models can be used as guidelines to judge whether the improvement will be enough to reach a certain "target membrane partitioning" or if further changes are necessary. The modeling work presented here will be even more useful when future transport studies allow to combine theoretical descriptors, membrane partitioning, and absorption rate in a biological system into a complete model of peptide absorption.

Experimental Section

Abbreviations: IAM, immobilized artificial membrane (chromatography); ILC, immobilized liposome chromatography; k'_{IAM}, capacity factor from IAM column; K_s, capacity factor from ILC column; PWASA, polar water-accessible surface area; NPWASA, nonpolar water-accessible surface area; TWASA, total water-accessible surface area; NMe-2, N-methylation at residue 2; NMe-4, N-methylation at residue 4; PLS, partial least-squares projection to latent structures; R^2, explanatory power; Q^2, predictive power from cross-validation; VIP, variable importance in the projection; RMSE, residual mean square error; PBS, phosphate-buffered saline.

Synthesis of Peptides. As starting support for the synthesis, 0.1 mmol of 4-((2′,4′-dimethoxyphenyl)(Fmoc-amino)-methyl)phenoxy resin (Rink amide resin) (Novabiochem, Bad Soden, Germany) with a substitution capacity of 0.46−0.68 mmol/g (150−230 mg) was used. All reagents were added in 4× excess (0.4 mmol). The resin was placed in a 20-mL BioRad reactor tube, allowing for vacuum suction for all washing steps. Fmoc protection groups were cleaved off by 20% piperidine in DMF followed by washing with 4 × 5 mL of DMF. Subsequent amide couplings were achieved at room temperature with *N*-ethyl-*N*′-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) or *N*,*N*-diisopropylcarbodiimide (DIC) as coupling reagent and *N*-hydroxybenzotriazole (HoBt) or 1-hydroxy-7-azabenzotriazole (HoAt) as activating agent. The latter was used when acylating secondary amines (proline or N-methylated amino acids) in which case overnight coupling was used. For HoBt, the coupling time was ≥1 h. Prior to adding the Fmoc amino acid to the reactor it was stirred for 15 min in 3 mL of DMF with the carbodiimide and HoAt or HoBt as appropriate.

Commercially available Fmoc amino acids and Fmoc N-methylated amino acids were used. However, in the case of N-Me-Ser, N-Me-Tyr, N-Me-Lys, and N-Me-Gln, which are not commercially available, an on-site methylation procedure was used.[22] Briefly, after deprotecting the end amino acid of the growing peptide chain it was converted to an activated sulfonamide by means of *o*-nitrobenzenesulfonyl chlorides (*o*-NBS) under collidine catalysis. Methylation of the sulfonamide was achieved by adding methyl *p*-nitrobenzenesulfonate (MNBS) along with 1,3,4,6,7,8-hexahydro-1-methyl-2*H*-pyrimido[1,2-*a*]pyrimidine (MTBD) as deprotonating agent to promote selective methylation. Finally, the *o*-NBS group was cleaved off by 1,8-diazabicyclo[5.4.0]undecen-7-ene (DBU) and β-mercaptoethanol. After washing with 4 × 5 mL of DMF, the coupling continued with HoAt and carbodiimide as coupling reagents. When using the same procedure for Val in compound **XX**, methylation was not accomplished. This was probably due to steric hindrance between the β-branched valine and the bulky reagents used for the methylation.

N-Terminal acetylation was done by treatment with 100 μL (8× excess) of acetic acid anhydride in 3 mL of DMF for 1 h. Peptides **I**−**XII** were cleaved from the resin by 3.6 mL of TFA/phenol/ethanedithiol/thioanisole/water (40:3:1:2:2) for 1 h at room temperature and peptides **XIII**−**XVIIII** with 3 mL of TFA under the same conditions. Purification was done by semipreparative RP-HPLC. After purification, the peptides were lyophilized.

Analytical HPLC. Method A1: The column was equilibrated with 5% acetonitrile in a buffer consisting of 0.1 M ammonium sulfate, which was adjusted to pH 2.5 with 4 M sulfuric acid. After injection the sample was eluted by a gradient of 5−60% acetonitrile in the same buffer during 50 min.

Method B1: The column was equilibrated with 5% acetonitrile/0.1% TFA/water and eluted by a gradient of 5% acetonitrile/0.1% TFA/water to 60% acetonitrile/0.1% TFA/water during 50 min. The RP-HPLC analysis was performed using UV detection at 214, 254, 276, and 301 nm on a Vydac 218TP54 4.6-mm × 250-mm 5-μm C-18 silica column, which was eluted at 1 mL/min at 42 °C.

PDMS (plasma desorption mass spectrometry) analysis was performed on a Bio-ion (Applied Biosystems) system using a Californium 252 (Cf252) source on a nitrocellulose matrix.

Immobilized Liposome Column. The immobilized liposome column was made according to Beigi et al.[23] Liposomes were prepared by evaporation of diethyl ether from a solution of egg yolk phospholipids and hydration of the formed lipid film by 10 mM TRIS buffer, pH 7.4. 6 mL of liposome suspension (0.6 mg/mL) was mixed with 440 mg of dry Superdex 200 (Pharmacia Biotech, Sweden). After this mixture was degassed, the immobilization was completed by five freeze−thaw cycles (freezing at −70 °C for 5 min followed by thawing at 25 °C for 5 min). Nonimmobilized liposomes were removed by centrifugal washes. The material was packed to a total bed volume of 3.4 mL in a 4-mL glass column (HR 5/20, Pharmacia Biotech, Sweden). After the column was equilibrated with PBS buffer, pH 7.4, on a standard HPLC system, flow rate 0.5 mL/min, the retention time of the peptides was measured with UV detection at 210 nm and compared to that of the void volume marker, $K_2Cr_2O_7$. The amount of phospholipids immobilized on the column was determined by a modified phosphorus analysis[23] and used to calculate the capacity factor as follows:

$$K_s = (t_{R(peptide)} - t_{R(K_2Cr_2O_7)})/(\text{molar amount of phospholipids})$$

where t_R = retention time.

IAM Column. The IAM column was a commercially available 10-cm IAM.PC.DD column (Regis Technologies) eluted at 1 mL/min with 0.01 M PBS buffer, pH 7.4. The void volume marker was citric acid. The capacity factor k'_{IAM} was calculated as follows:

$$k'_{IAM} = (t_{R(peptide)} - t_{R(citric\ acid)})/t_{R(citric\ acid)}$$

where t_R = retention time.

Theoretical Characterization. The peptides were built in Sybyl 6.4[39] in an extended conformation and energy minimized with the Amber 4.0 force field as implemented in Sybyl. The water-accessible surface area of the peptides was calculated by the analytical algorithm Savol3.[40] The polar surface area was defined as the surface area of oxygen and nitrogen atoms and hydrogen atoms attached to them. The nonpolar surface area was calculated as the total minus the polar surface area. Molsurf descriptors were derived essentially as described by Norinder et al.[35] Briefly, in Spartan[41] geometry optimization was performed by the semiempirical method AM1,[42] followed by a single energy calculation at the quantum chemical 3-21G level. Molsurf uses the output from the 3-21G calculation to calculate the local ionization potential in a grid around the molecule and then translates that into physicochemical descriptors such as log *P*, pK_a, and hydrogen-bonding properties. Log *D* at pH 7.4 was calculated from log *P* and pK_a.

The basis for the Volsurf[26] descriptors is interaction energies calculated by the GRID program.[27] Using GRID version 16, the polar and hydrophobic properties were assessed as the isopotential contours at 8 energy levels for interaction energies with a water and a DRY probe. From the volumes of the isoenergetic contours, Volsurf derives the parameters listed in Table 4.

The log *P* value was calculated using the program mlogp.[43]

Conformational Analysis. 1000-ps high-temperature MD at 1000 K with sampling every picosecond was performed in

Sybyl 6.4[39] (Amber 4.0 force field), thus producing 1000 conformers. The conformers generated by the MD served as input for an energy minimization using the force field MM3* with water solvation (GB/SA continuum model) as implemented in Macromodel.[44] The performance of the conformational analysis was evaluated by examination of a Ramachandran plot for each residue, supplemented with plots of the χ_1 torsional angles. The Boltzmann averaged molecular surface parameters were calculated as described previously.[3]

Acknowledgment. The assistance of Per Lundahl and Farideh Beigi (Uppsala University, Sweden) with the ILC column is highly appreciated. Niels Langeland Johansen, Kjeld Madsen, and co-workers (Novo Nordisk A/S, Denmark) are thanked for their hospitality and advice during the synthesis work. The discussions of the statistical part with Lennart Eriksson, Erik Johansson, and Svante Wold (Umetric AB, Umeå, Sweden) are gratefully appreciated. Flemming Steen Jørgensen (The Royal Danish School of Pharmacy) is thanked for the discussion concerning the conformational analysis. The work was financially supported by a grant from Novo Nordisk A/S, Nycomed Danmark A/S, Fertin A/S, and the Danish Research Academy.

Supporting Information Available: Molecular surface descriptors, Volsurf descriptors, and Molsurf descriptors for all 20 peptides. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kansy, M.; Senner, F.; Gubernator, K. Physicochemical high throughput screening: parallel artifical membrane permeation assay in the description of passive absorption processes. *J. Med. Chem.* **1998**, *41*, 1007−1010.
(2) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3−25.
(3) Krarup, L. H.; Christensen, I. T.; Hovgaard, L.; Frokjaer, S. Predicting drug absorption from molecular surface properties based on molecular dynamics simulations. *Pharm. Res.* **1998**, *15*, 972−978.
(4) Palm, K.; Luthman, K.; Ungell, A.-L.; Strandlund, G.; Artursson, P. Correlation of drug absorption with molecular surface properties. *J. Pharm. Sci.* **1996**, *85*, 32−39.
(5) van de Waterbeemd, H.; Kansy, M.; Camenish, G.; Folkers, G.; Raevsky, O. A. Estimation of Caco-2 cell permeability using calculated molecular descriptors. *Quant. Struct.-Act. Relat.* **1996**, *15*, 480−490.
(6) Winiwarter, S.; Bonham, N. M.; Ax, F.; Hallberg, A.; Lennernäs, H.; Karlén, A. Correlation of human jejunal permeability (in vivo) of drugs with experimentally and theoretically derived parameters. A multivariate data analysis approach. *J. Med. Chem.* **1998**, *41*, 4939−4949.
(7) Palm, K.; Stenberg, P.; Luthmann, K.; Artursson, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* **1997**, *14*, 568−571.
(8) Langguth, P. Intestinal permeability and absorption of peptide drugs. In *Aspekte der Intestinalen Absorption und der Modellentwicklung in Pharmakokinetik und Pharmakodynamik*; Gramatte, T., Weiss, M., Eds.; W. Zuckschwerdt Verlag: München, Bern, Wien, New York, 1998; pp 45−55.
(9) Conradi, R. A.; Hilgers, A. R.; Ho, N. F. H.; Burton, P. S. The influence of peptide structure on transport across Caco-2 cells. *Pharm. Res.* **1991**, *8*, 1453−1460.
(10) Conradi, R. A.; Hilgers, A. R.; Ho, N. F. H.; Burton, P. S. The influence of peptide structure on transport across Caco-2 cells. II. Peptide modification which results in improved permeability. *Pharm. Res.* **1992**, *9*, 435−439.
(11) Karls, M. S.; Rus, B. D.; Wilkinson, K. F.; Vidmar, T. J.; Burton, P. S.; Ruwart, M. J. Desolvation energy: A major determinant of absorption, but not clearance, of peptides in rats. *Pharm. Res.* **1991**, *8*, 1477−1481.
(12) Hansen, T. K.; Ankersen, M.; Hansen, B. S.; Raun, K.; Nielsen, K. K.; Lau, J.; Peschke, B.; Lundt, B. F.; Thogersen, H.; Johansen, N. L.; Madsen, K.; Andersen, P. H. Novel orally active growth hormone secretagogues. *J. Med. Chem.* **1998**, *41*, 3705−3714.
(13) Knipp, G. T.; Vander Velde, D. G.; Siahaan, T. J.; Borchardt, R. T. The effect of beta-turn structure on the passive diffusion of peptides across Caco-2 cell monolayers. *Pharm. Res.* **1997**, *14*, 1332−1340.
(14) Pauletti, G. M.; Okumu, F. W.; Borchardt, R. T. Effect of size and charge on the passive diffusion of peptides across Caco-2 cell monolayers via the paracellular pathway. *Pharm. Res.* **1997**, *14*, 164−168.
(15) Barlow, D.; Sartoh, T. The design of peptide analogues with improved absorption. *J. Controlled Release* **1994**, *29*, 283−291.
(16) Stenberg, P.; Luthman, K.; Artursson, P. Prediction of membrane permeability to peptides from calculated dynamic molecular surface properties. *Pharm. Res.* **1999**, *16*, 205−212.
(17) Wold, S.; Sjöström M.; Carlson, R.; Lundstedt, T.; Hellberg, S.; Skagerberg, B.; Wikström, C.; Öhman, J. Multivariate design. *Anal. Chim. Acta* **1986**, *191*, 17−32.
(18) Hellberg, S.; Sjöström, M.; Skagerberg, B.; Wold, S. Peptide quantitative structure−activity relationships, a multivariate approach. *J. Med. Chem.* **1987**, *30*, 1126−1135.
(19) Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjöström, M.; Wold, S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J. Med. Chem.* **1998**, *41*, 2481−2491.
(20) Modde, version 3.0, Umetri AB, Umeå.
(21) Baroni, M.; Clementi, S.; Cruciani, G.; Kettaneh-Wold, N.; Wold, S. D-optimal designs in QSAR. *Quant. Struct.-Act. Relat.* **1993**, *12*, 225−231.
(22) Miller, S. C.; Scanlan, T. S. Site-selective N-methylation of peptides on solid support. *J. Am. Chem. Soc.* **1997**, *119*, 2301−2302.
(23) Beigi, F.; Gotschalk, I.; Hägglund, C. L.; Haneskog, L.; Brekkan, E.; Zhang, Y.; Österberg, T.; Lundahl, P. Immobilized liposome and biomembrane partitioning chromatography of drugs for prediction of drug transport. *Int. J. Pharm.* **1998**, *164*, 129−137.
(24) Pidgeon, C.; Ong, S.; Liu, H.; Qiu, X.; Pidgeon, M.; Dantzig, A. H.; Munroe, J.; Hornback, W. J.; Kasher, J. S.; Glunz, L.; Szczerba, T. IAM chromatography: An in vitro screen for predicting drug membrane permeability. *J. Med. Chem*, **1995**, *38*, 590−594.
(25) Molsurf − a generator of Chemical Descriptors for QSAR. *Computer-Assisted Lead Finding and Optimization*; van de Waterbeemd, H., Testa, B., Folkers, G., Eds.; Verlag Helvetica Chimica Acta: Basel, **1997**; pp 81−92. Molsurf version 3.11, Qemist AB, Karlskoga, Sweden, e-mail par.sjoberg@mbox309.swipnet.se.
(26) Cruciani, C.; Pastor, M.; Guba, W. Volsurf, a tool for handling 3D maps for QSAR studies. In preparation. Volsurf version 1.1, Multivariate Infometric Analysis, Perugia, Italy.
(27) GRID version 16, Molecular Discovery Ltd., Oxford, U.K.
(28) Simca, version 7.01, Umetri AB, Umeå, Sweden, 1998.
(29) Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Series B Methodol.* **1974**, *36*, 111−133.
(30) Eriksson, L.; Johansson, E.; Wold, S. Quantitative Structure−Activity relationship model validation. *Quantitative Structure−Activity Relationships in Environmental Sciences*; Setac Press, 1998; pp 381−397.
(31) Barbato, F.; La Rotonda, M. I.; Quaglia, F. Chromatographic indices determined on an immobilized artifical membrane (IAM) column as decriptors of lipophilic and polar interactions of 4-phenyldihydropyridine calcium-channel blockers with biomembranes. *Eur. J. Med. Chem.* **1996**, *31*, 311−318.
(32) Austin, R. P.; Davis, A. M.; Manners, C. N. Partitioning of ionizing molecules between aqueous buffers and phospholipid vesicles. *J. Pharm. Sci.* **1995**, *84*, 1180−1183.
(33) Avdeef, A.; Box, K. J.; Comer, J. E. A.; Hibbert, C.; Tam, K. Y. pH-metric log P10. Determination of liposomal membrane-water partition coefficients of ionizable drugs. *Pharm. Res.* **1998**, *15*, 209−215.
(34) Stein, W. D. The molecular basis of diffusion across cell membranes. *The Movement of Molecules across Cell Membranes*; Academic Press: New York, 1967; pp 65−125.
(35) Norinder, U.; Sjöberg, P.; Österberg, T. Theoretical calculation and prediction of brain-blood partitioning of organic solutes using MolSurf parametrization and PLS statistics. *J. Pharm. Sci.* **1998**, *87*, 952−959.
(36) Norinder, U.; Österberg, T.; Svensson, P. Theoretical calculation and prediction of Caco-2 cell permeability using MolSurf parametrization and PLS statstics. *Pharm. Res.* **1997**, *14*, 1786−1791.
(37) Jacobs, R. E.; White, S. H. The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the

insertion of transbilayer helices. *Biochemistry* **1989**, *28*, 3421−3437.
(38) Palm, K.; Stenberg, P.; Luthmann, K.; Artursson, P. Evaluation of dynamic polar molecular surface area as predictor of drug absorption: comparison with other computational and experimental predictors. *J. Med. Chem.* **1998,** *41*, 5382−5392.
(39) Sybyl version 6.4, Tripos Associates, St. Louis, MO.
(40) Pearlman, R. S.; Skell, J. M. *SAVOL3: Numerical and analytical algorithms for molecular surface area and volume*; College of Pharmacy, University of Texas: Austin, TX.
(41) Spartan version 4.0, Wavefunction Inc., Irvine, CA.
(42) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(43) (a) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple method of calculating octanol/water partition coefficient. *Chem. Pharm. Bull.* **1992**, *40*, 127−139. (b) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Hirano, H. Comparison of reliability of logP values for drugs calculated by several methods. *Chem. Pharm. Bull.* **1994**, *42*, 976−978. Mlogp spl script implemented by James F. Blake, Pfizer Inc., Central Research Division, e-mail blakefj@pfizer.com.
(44) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrikson, T.; Still, W. C. Macromodel − an integrated software system for modelling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.

JM9910932