Multivariate Design, Synthesis, and Biological Evaluation of Peptide Inhibitors of FimC/FimH Protein−Protein Interactions in Uropathogenic *Escherichia coli*

Andreas Larsson,[†] Susanne M. C. Johansson,[†] Jerome S. Pinkner,[§] Scott J. Hultgren,[§] Fredrik Almqvist,[†] Jan Kihlberg,*[,†,‡] and Anna Linusson*[,†,‡]

Department of Chemistry, Organic Chemistry, Umeå University, SE−901 87 Umeå, Sweden, Medicinal Chemistry, AstraZeneca R&D Mölndal, SE−431 83 Mölndal, Sweden, and Department of Molecular Microbiology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, Missouri 63110

Received March 29, 2004

A peptide library targeting protein−protein interactions crucial for pilus assembly in Gram negative bacteria has been designed using statistical molecular design. A nonamer peptide scaffold was used, with seven positions being varied. The selection was performed in the building block space, and previously known structure−activity data were included in the design procedure. This resulted in a heavily reduced library consisting of 32 peptides which was prepared by solid-phase synthesis. The ability of the peptides to inhibit the protein−protein interaction between the periplasmic chaperone FimC and the pilus adhesin FimH was then determined in an ELISA. Novel peptides with the capability to inhibit the FimC/FimH protein−protein interaction to the same extent as the native FimC peptides were discovered. Multivariate QSAR studies of the response in the ELISA gave valuable information on the properties of amino acids which were preferred at the seven positions in the nonamer scaffold. This information can be used in attempts to develop optimized peptides and peptidomimetics that inhibit pilus assembly in pathogenic bacteria.

Introduction

Urinary tract infections (UTIs), including both pyelonephritis and cystitis, affect a large proportion of the world population and account for significant morbidity and high medical costs. It is estimated that one-third of American women will have at least one UTI before the age 65 and many will experience more than one infection per year.[1] Strains of uropathogenic *Escherichia coli* (UPEC) are the causative agents in the vast majority of all urinary tract infections.[1,2] In contrast to resident intestinal strains and other *E. coli* isolates, UPEC strains encode a number of virulence factors[3] that enable them to colonise the urinary tract and persist within the bladder for days to weeks, despite the presence of a highly effective host defense. In addition, it has been shown that UPEC strains can persist within mouse bladder tissue virtually unharmed during antibiotic treatments that effectively reduce bacterial titers within the urine.[4]

The ability of *E. coli* to adhere to host epithelial cells within the urinary tract appears to be the most essential virulence factor in UTI. UPEC display a variety of adhesins and adhesive organelles on the bacterial surface; two of the most important ones being P pili and type 1 pili which are rodlike, supramolecular protein appendages that extend from the bacterial surfaces. Adhesins at the tip of P and type 1 pili are responsible for the recognition and attachment to glycolipids of the globoseries[5,6] in the kidneys and mannose residues on glycoproteins[7,8] in the bladder, respectively. In the absence of pili, *E. coli* are unable to invade host uroepithelial cells and cause disease.[9] Type 1 pili are assembled by a chaperone/usher pathway where a periplasmic chaperone (FimC) helps to fold, stabilize, and transport pilus subunits from the inner cell membrane, through the periplasmic compartment to outer membrane assembly sites called ushers (FimD). The FimD usher aids in disassembling of the chaperone/subunit complex and incorporates the subunit into the growing pili.[10−12] Type 1 pili consist of several repeating immunoglobulin (Ig)-like subunits (FimA, FimF, FimG, and FimH) that all lack the C-terminal β-strand required to complete the Ig-fold. The subunits form pili through a donor strand exchange reaction, whereby every subunit donates its N-terminal extension to complete the Ig fold of its neighbor, thus forming a noncovalent Ig-like polymer.[13] The absence of the C-terminal β-strand makes folding of the subunits dependent upon the periplasmic chaperone FimC, which comprises two Ig-like domains.[14,15] As revealed by the structure of the FimC/FimH complex the chaperone donates its edge, G_1 β-strand to complete the Ig fold of the subunit in a process termed 'donor strand complementation' (Figure 1).[16,17] The adhesin, FimH, consists of two different Ig-like domains, one receptor binding domain that is involved in binding to mannose-derivatives in the bladder and one pilin domain which attaches FimH to the tip of the pilus. In the complex with FimC[16] a crevice between strand A″ and F of the pilin domain is filled by insertion of the G_1 β-strand of FimC parallel to the F strand producing an atypical immunoglobulin fold. The other subunits of type 1 pili (FimA, FimF, and

* To whom correspondence should be addressed. J.K.: phone: +46 90 786 6890, fax: +46 90 13 88 85, e-mail: jan.kihlberg@chem.umu.se. A.L.: phone: +46 90 786 6890, fax: +46 90 13 88 85, e-mail: anna.linusson@chem.umu.se.
[†] Umeå University.
[‡] AstraZeneca R&D Mölndal.
[§] Washington University School of Medicine.

10.1021/jm040818l CCC: $30.25
Published on Web 02/01/2005

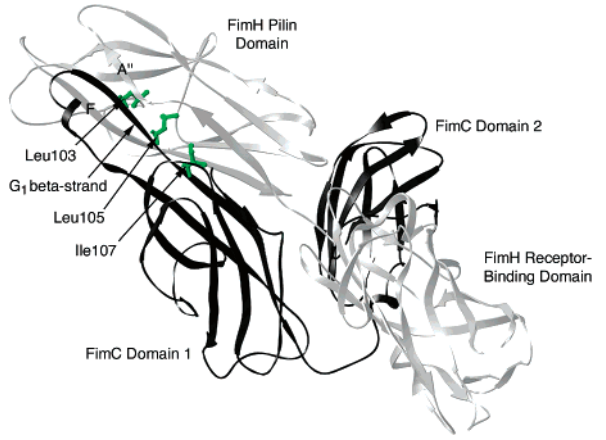


Figure 1. Donor strand complementation interactions in the FimC/FimH chaperone/adhesin complex (FimC in black and FimH in gray). The G_1 β-strand of FimC runs parallel to the F strand of FimH thereby completing the Ig fold of the FimH pilin domain in an atypical manner. The alternating hydrophobic residues of the G1 β-strand of FimC (Leu103, Leu105, Ile107) are buried in the hydrophobic core of the FimH pilin domain.

FimG) are assumed to form complexes with FimC in a similar manner.

Inhibition of one of the protein−protein interactions either in the pili or in any of the chaperone/subunit complexes by peptidomimetics could be an effective way of controlling UTIs caused by UPEC. Such an approach offers several advantages. To begin with, these protein−protein interactions constitute a novel drug target that is required for pathogenesis.[18] Furthermore, the pathway is conserved in a variety of pathogenic bacteria responsible for a wide range of diseases, such as UTIs, diarrhea, pneumonia, plague, and meningitis.[12] Finally, since a complex virulence mechanism is targeted, mutants will most likely be avirulent and the development of resistance is therefore less likely to occur.

Protein−protein interactions have been considered as attractive but difficult targets for drug development.[19−21] The chaperone/subunit complex however shows the unusual behavior where the binding epitope is concentrated in a single, short peptide found in a well-defined secondary structure element. This opens up possibilities for design of optimized peptide inhibitors of these protein−protein interactions and further development of such peptides into peptidomimetics. In fact, we recently reported initial results revealing that native peptides from FimC chaperone or type 1 pilus proteins were able to block FimC/FimH complexation.[22] Peptides have earlier been successfully used for inhibition of protein−protein interactions, e.g. in interactions between integrins and cell adhesion molecules,[23,24] in blocking of HIV protease[25]and HSV ribonucleotide reductase,[26] and activation of tumor suppressor p53 by interfering with oncoprotein Hdm2.[27]

A widely used strategy to assess both conformational and side chain requirements in an active peptide is to systematically scan the peptide by incorporation of the same amino acid at one position at a time and then compare the effect to the wild type.[28] Such alanine, serine, and proline scans may seem to give the experimentalists a structured way of working and a clear output. It should, however, be emphasized that this strategy is not the most efficient way of working and in some cases may lead to incorrect conclusions. Instead of changing properties for one position at a time, it is beneficial to vary different molecular properties and several positions at the same time and adopt a so-called statistical molecular design (SMD).[29−31] The clear advantage with this approach is that it gives the possibility of investigating more than one molecular property at several positions with the minimum number of experiments. It also gives information about potential interaction effects between properties at different positions and provides a solid base for structure−activity relationships.[32,33] SMD is used within the pharmaceutical industry to select building blocks (reactants) for parallel and combinatorial chemistry,[31,34−36] but has not been extensively used within the peptide community.[37−39]

In the present work a statistically designed peptide library of FimC/FimH complexation inhibitors has been synthesized and evaluated through biological testing and subsequent structure−activity relationship studies. The design was performed in the building block space, using previously acquired results[22] and the structure of FimC/FimH complex[16] in the selection procedure. The selected peptides were synthesized in parallel on solid phase, and their capability to inhibit FimC/FimH complex formation was assessed in an ELISA. The resulting designed library and response values were evaluated using multivariate methods.

Results and Discussion

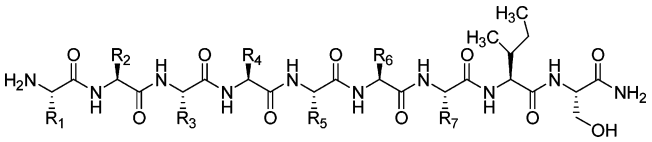
The Peptide Scaffold. When examining the crystal structure of the FimC/FimH complex for essential interactions,[16] it was found that the side chains of Leu103C and Leu105C in the G_1 β-strand were deeply buried in the lipophilic crevice created in the FimH pilin domain by the missing seventh strand (Figure 1). Ile107C was somewhat closer to the domain surface but still made extensive van der Waals contacts in the crevice of FimH. Ile108C and Ser109C were in contact with FimH but make only limited interactions.[16] The G_1 β-strand of the FimC chaperone therefore completes the immunoglobulin-like fold of the subunits in an atypical, parallel fashion. However, when aligning the sequence of the G_1 β-strand with the N-terminus of the pilus subunits, it was assumed that a short peptide corresponding to the G_1 β-strand would instead bind in the subunit's crevice in an antiparallel manner. Thus a common canonical immunoglobulin form would be produced in analogy with when the N-terminal peptides of the pilus subunits are bound in the crevice.[40] When bound in an antiparallel manner residues Leu105C and Ile107C would be deeply buried in the hydrophobic crevice of FimH. The N-terminal extensions of FimA, FimF, and FimG were consequently aligned to these two large and lipophilic FimC residues as a conserved character from the Clustal W alignment made by Choudbury et al.[16] (see Table 1).

Recent studies have shown that nonamer peptides corresponding to the G_1 β strand of FimC (Asn101 to Ser109) and the N-terminal extensions of subunits FimA, FimF, and FimG have the ability to inhibit subunit/chaperone complexation.[22] Therefore, a nonamer peptide scaffold in which the seven first positions were varied was selected for the statistical molecular design (Figure 2).

Table 1. Alignment of Native Peptide from the FimC G1 β-Strand with the N-Terminal Extensions of the Subunits FimA, FimF, FimG Using a Clustal W Alignment[16]

peptide	Pos1	Pos2	Pos3	Pos4	Pos5	Pos6	Pos7	Pos8	Pos9	
FimC (**34**)	Asn^{101}	Thr^{102}	Leu^{103}	Gln^{104}	Leu^{105}	Ala^{106}	Ile^{107}	Ile^{108}	Ser^{109}	-
FimA[a] (**35**)	-	Gly^{9}	Gly^{10}	Thr^{11}	Val^{12}	His^{13}	Phe^{14}	Lys^{15}	Gly^{16}	Gly^{17}
FimF[a] (**36**)	-	Asp^{1}	Ser^{2}	Thr^{3}	Ile^{4}	Thr^{5}	Ile^{6}	Arg^{7}	Gly^{8}	Tyr^{9}
FimG[a] (**37**)	-	Asp^{1}	Val^{2}	Thr^{3}	Ile^{4}	Thr^{5}	Val^{6}	Asn^{7}	Gly^{8}	Lys^{9}

[a] The subunit peptides were positioned with conserved large hydrophobic residues aligned with Leu105C and Ile107C.



Pos1	Pos2	Pos3	Pos4	Pos5	Pos6	Pos7
Asn*	Thr*	Leu*	Gln*	Leu*	Ala*	Phe*
Ala	Asp*	Gly*	Thr*	Gly	His*	Leu
Leu	Gly*	Ser*	Ala	Ser	Thr*	Ser
Lys	Lys	Phe	Asp	Phe	Asp	Trp
Thr	Phe		Lys		Leu	
	Val		Phe			

Figure 2. Building block selection for a nonamer peptide scaffold with seven positions varied and two held constant (Ile, Ser). The positions were screened for the physical properties size, lipophilicity and charge (positions 1, 2, 4, and 6), or size and lipophilicity (positions 3, 5, and 7). The selection was biased to include amino acids which were present in peptides believed to be binders. Underscored amino acids are found in the G_1 β-strand of FimC, whereas the amino acids marked with an asterisk correspond to those that can be found in one or more of the N-terminal extensions of the pilus subunits (FimA, FimF, and FimG).

Building Block Selection. The molecular properties selected as design variables were identified based on knowledge both from the FimC/FimH crystal structure and from alignment of the N-terminal peptides from FimA, FimF and FimG, which are involved in donor strand complementation between neighboring subunits in the pilus.[13,22] Proline was excluded from the candidate set due to it's propensity to initiate turn structures. According to the crystal structure residues 3, 5, and 7 of the nonamer peptide reside in a highly lipophilic environment within the FimH crevice, in contrast to residues 1, 2, 4, and 6 which interact with the surrounding solvent (cf. Figure 1). Therefore two different sets of design variables were employed. In position 1, 2, 4, and 6 three properties were varied; size, lipophilicity, and charge. A principal component analysis (PCA) of 19 of the naturally occurring amino acids (proline excluded) described by 80 molecular descriptors (Table 2) resulted in three principal properties describing the three properties, that were varied (Figure 3a−d). For position 3, 5, and 7, that point down into the lipophilic FimH crevice, the properties size and polarity were chosen as design parameters in the first screening design. Therefore, in addition to proline, all amino acids which are charged at pH 7.4 (Glu, Asp, Lys, Arg, and His) were excluded from the candidate set. The PCA resulted in two components describing size and polarity, which were used to select amino acids that cover this property space (Figure 3e,f). In total 18 factors were investigated in this screening design, three properties for positions 1, 2, 4, and 6, and two properties for positions 3, 5, and 7.

The selection of building blocks for the two sets of design variables was biased to include amino acids present in peptides that were believed to be binders (i.e. found in the N-termini of FimA, FimF and FimG) and complemented with diverse amino acids to spread the property space. Thus, Ala, Leu, Lys, and Thr were selected for position 1; Lys, Phe, and Val for position 2; Phe for position 3; Ala, Asp, Lys, and Phe for position 4; Gly, Ser, and Phe for position 5; Asp and Leu for position 6 and Leu, Ser and Trp for position 7 (Figure 2 and Figure 3a,c,e). In addition, the underscored amino acids in Figure 2, which are found in the G_1 β-strand of FimC were included in the selection. This was also the case for the amino acids marked with an asterisk (Figure 2) that are found in one or more of the three pilus subunits (FimA, FimF, or FimG). When amino acids with similar properties (i.e. Ile vs Leu) were found in the same position only one of them was selected.

The Final Library Selection. All combinations of the selected amino acids for the seven varied positions would result in 57 600 peptides. To reduce this number a second design was applied (Figure 4). The peptides were divided into two parts, N-terminal tripeptides and C-terminal tetrapeptides. All combinations of the selected amino acids for the different positions were generated. These sets were filtered on the criteria that both the tripeptides and the tetrapeptides should contain at least one amino acid that is present in the natural peptides from FimC, FimA, FimF, or FimG. This was done with the aim of tightening the design around the naturally occurring binders to increase the probability of designing and synthesizing active peptides useful for the QSAR modeling. A D-optimal design, using the selected amino acids as qualitative descriptors with linear and interaction terms, was applied to the two filtered blocks, resulting in 40 and 150 peptides, respectively. Qualitative descriptors were used to get a good spread of all the selected amino acids in the reduced sets. Finally, full length peptides were generated using the 40 selected tripeptides and 150 tetrapeptides resulting in a candidate set of 6000 heptapeptides. Peptides that were believed to cause problems during synthesis and purification were excluded iteratively from this candidate set and 32 peptides were then selected using D-optimal design with qualitative variables and linear terms (Table 3). One peptide containing amino acids corresponding to residues 101−109 in the G_1 β-strand of FimC, except for position 7 where Ile was changed for Leu, was also included as the reference point (peptide no. 33, Table 3).

QSAR of FimC/FimH Inhibition. All but three of the 33 selected peptides were successfully synthesized on solid phase and purified by reversed phase HPLC. The ability of the peptides to inhibit complexation between the FimC chaperone and the FimH adhesin was determined as averages of four replicates in a competitive ELISA at three concentrations 10, 50, and 200 μM (Table 3). The ELISA was carried out by

Table 2. List of the Structural Descriptors Used for Characterization of the Individual Amino Acid Residues[42]

No.	Abbreviation	Descriptors	No.	Abbreviation	Descriptors
1	VDistEq	Vertex Distance Equation	41	KierA1	Alpha Modified Shape Index
2	VDistMa	Vertex Distance Magnitude	42	KierA2	Alpha Modified Shape Index
3	weinerPath	Weiner Path Number	43	KierA3	Alpha Modified Shape Index
4	weinerPol	Weiner Polarity Number	44	KierFlex	Flexibility Index
5	a_aro	Number of aromatic atoms	45	apol	Atomic Polarizabilities
6	b_ar	Number of aromatic bonds	46	bpol	Atomic Polarizabilities
7	b_rotN	Number of rotatable bonds	47	dipole	Dipole Moment
8	b_rotR	Fraction of rotatable bonds	48	a_acc	Number of Hydrogen Bond Acceptors
9	chi0v	Atomic Valence Connectivity Index	49	a_acid	Number of Acidic Atoms
10	chi0v_C	Carbon Valence Connectivity Index	50	a_base	Number Basic Atoms
11	chi1v	Atomic Valence Connectivity Index	51	a_don	Number of Hydrogen Bond Donors
12	chi1v_C	Carbon Valence Connectivity Index	52	a_hyd	Number of Hydrophobic Atoms
13	Weight	Molecular Weight	53	vsa_acc	van der Waals Surface Areas of Hydrogen Bond Acceptors
14	chi0	Atomic Connectivity Index	54	vsa_acid	van der Waals Surface Areas of Acidic Atoms
15	chi0_C	Carbon Connectivity Index	55	vsa_base	van der Waals Surface Areas of Basic Atoms
16	chi1	Atomic Connectivity Index	56	vsa_don	van der Waals Surface Areas of Hydrogen Bond Donors
17	chi1_C	Carbon Connectivity Index	57	vsa_hyd	van der Waals Surface Areas of Hydrophobic Atoms
18	FCharge	Sum of Formal Charges	58	vsa_other	van der Waals Surface Areas of Other Atoms
19	VAdjEq	Vertex Adjacency Equation	59	vsa_pol	van der Waals Surface Areas of Polar Atoms
20	VAdjMa	Vertex Adjacency Magnitude	60	SlogP	Log of the Octanol/Water Partition Coefficient
21	zagreb	Zagreb Index	61	SMR	Molecular Refractivity
22	Q_PC+	Total Positive Partial Charge	62	ASA	Water Accessible Surface Area
23	Q_PC-	Total Negative Partial Charge	63	ASA+	Positive Partial Charge ASA
24	Q_RPC+	Relative Postitive Partial Charge	64	ASA−	Negative Partial Charge ASA
25	Q_RPC−	Relative Negative Partial Charge	65	ASA_H	Hydrophobic ASA
26	Q_VSA_FHYD	Fractional Hydrophobic van der Waals Surface Area	66	ASA_P	Polar ASA
27	Q_VSA_FNEG	Fractional Negative van der Waals Surface Area	67	FASA+	Fractional ASA+
28	Q_VSA_FPNEG	Fractional Polar Neagtive van der Waals Surface Area	68	FASA−	Fractional ASA−
29	Q_VSA_FPOL	Fractional Polar van der Waals Surface Area	69	FASA_H	Fractional Hydrophobic ASA
30	Q_VSA_FPOS	Fractional Positive van der Waals Surface Area	70	FASA_P	Fractional Polar ASA
31	Q_VSA_FPPOS	Fractional Polar Positive van der Waals Surface Area	71	TPSA	Total Polar Surface Area
32	Q_VSA_HYD	Total Hydrophobic van der Waals Surface Area	72	density	Molecular Mass Density
33	Q_VSA_NEG	Total Negative van der Waals Surface Area	73	vdw_area	van der Waals Surface Area
34	Q_VSA_PNEG	Total Polar Negative van der Waals Surface Area	74	vdw_vol	van der Waals Volume
35	Q_VSA_POL	Total Polar van der Waals Surface Area	75	glob	Globularity
36	Q_VSA_POS	Total Positive van der Waals Surface Area	76	std_dim1	Standard Dimension
37	Q_VSA_PPOS	Total Polar Positive van der Waals Surface Area	77	std_dim2	Standard Dimension
38	Kier1	Kappa Shape Index	78	std_dim3	Standard Dimension
39	Kier2	Kappa Shape Index	79	logP(o/w)	Log of the Octanol/Water Partition Coefficient
40	Kier3	Kappa Shape Index	80	PMI	Moment of Inertia (calculated from pmiZ/pmiX)

incubating mixtures of peptide and FimH on to 96-well plates previously coated with FimC, subsequent antibody detection of FimH gave a quantitative measure of the amount of formed FimC/FimH complex. The standard deviation for the biological measurements was 7.5% at 10 μM, 6.8% at 50 μM, and 6.2% at 200 μM. The structure−activity relationship of the peptides was evaluated using multi-Y PLS with the % inhibition at the three inhibitory concentrations (10, 50, and 200 μM) as the response. Each of the seven positions in the peptide scaffold was investigated separately using the molecular properties of the amino acids as described by the principal properties corresponding to the 18 factors in the building block design (cf. Figure 3). Eight of the 18 linear terms were excluded by an iterative procedure based on their coefficient values giving a valid QSAR model (2 PLS components, Table 4: Model 1). The plots over the calculated inhibition values versus the experimental ones (10, 50, and 200 μM) clearly showed that the residuals held nonlinear patterns. Therefore, square and interaction terms were added to the linear terms to address the nonlinearity. The interpretation of the main terms did not differ between the model based on solely linear terms (Model 1) and the one with nonlinear terms included (Model 2). It is important to remember that the synthesized set of peptides constitutes a screening design that supports linear factors, hence nonlinear terms need to be interpreted with caution. The final QSAR model (2 PLS components, Table 4: Model 2) showed a good correlation between the experimental and the calculated inhibition values (Figure 5). The model was successfully validated using an external test set consisting of peptides from FimA, FimF, FimG, and the wild-type FimC (Table 5). Evaluation of the test set in the competitive ELISA revealed that the FimA peptide had almost no inhibitory power, while the FimF and FimG peptides showed intermediate inhibition and the FimC peptide was the best inhibitor. The model predicts the same order of inhibitory power as found in the inhibition assay, with the FimA peptide being almost inactive, the FimF and FimG peptides having intermediate inhibition and the FimC peptide being the best inhibitor (Table 5). These findings further validate our previous alignment where the free FimC peptide

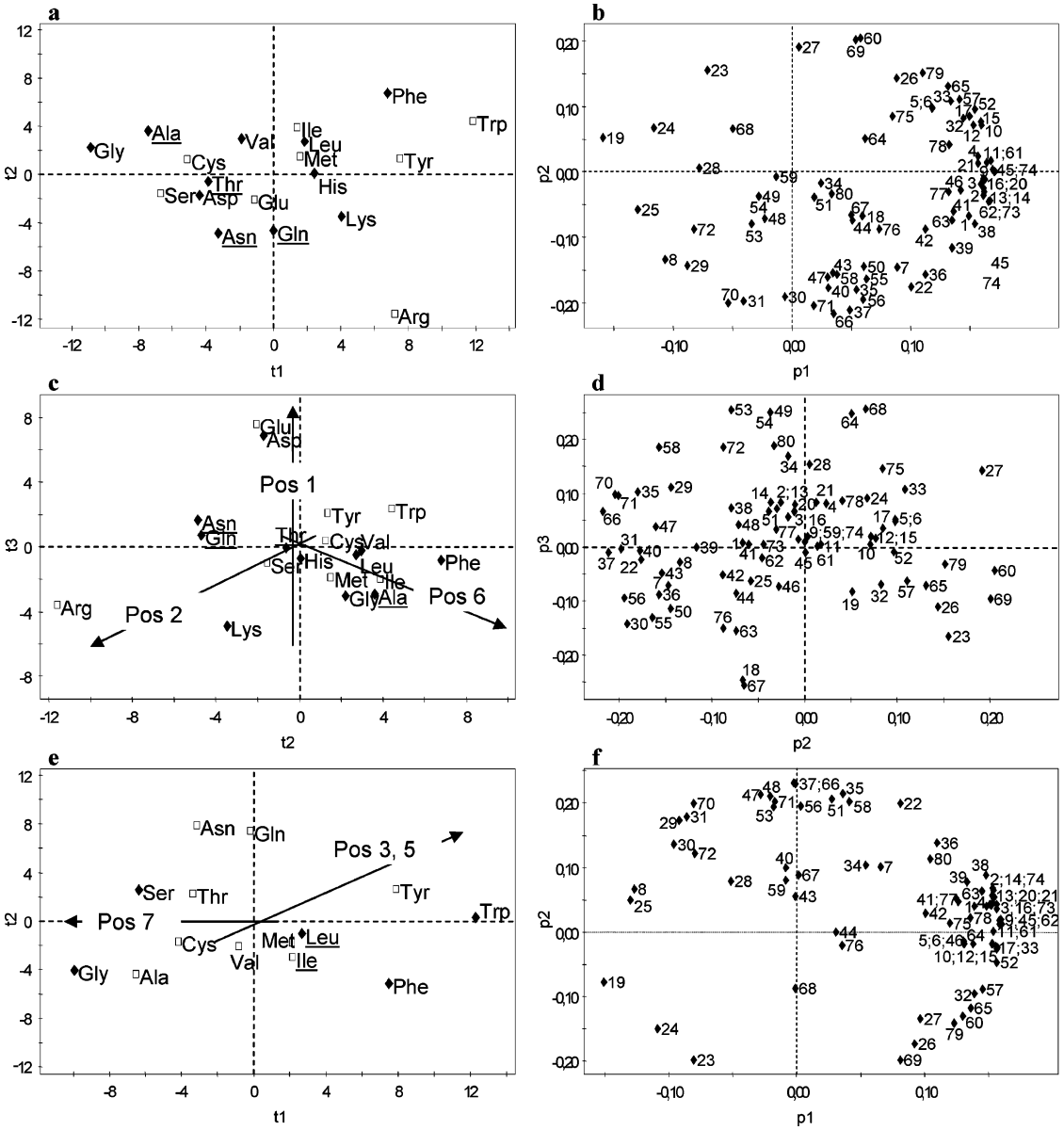


Figure 3. Score (a, c, e) and loading (b, d, f) plots of building block sets. Plots a−d correspond to the set used for positions 1, 2, 4 and 6 (proline excluded), whereas e and f correspond to the set used for positions 3, 5, and 7 (proline and charged amino acids excluded). Underscored amino acids are present in the G_1 β-strand of FimC. Amino acids marked with ♦ were included in the design; those marked with □ were not included. The arrows indicate the preferable direction shown by the linear terms in the QSAR model. The numbers in the loading plots (b, d, f) relate to the structural descriptors listed in Table 2.

was assumed to bind in the groove of FimH in an antiparallel fashion. The model contained 10 linear terms, three square terms and two interaction terms and the regression coefficients for the different peptide positions for inhibition at 50 μM are presented in Figure 6. The regression coefficients for the inhibition at 10 and 200 μM follow the same pattern as for 50 μM. Out of the seven peptide positions investigated, all but position 4 showed importance for the model. The amino acids Gln, Thr, Ala, Asp, Lys, and Phe that had been selected for position 4 have a good spread in principal properties (Figure 3a,c). However, from the QSAR model it was clear that properties such as size, lipophilicity, and charge at this position did not influence the ability of the peptides to inhibit FimC/FimH complexation.

Position 1 was investigated using Asn, Ala, Leu, Lys, and Thr, and the QSAR revealed that the size and the lipophilicity of these amino acids did not influence the potency of the peptides, i.e., the variables size (P1_size) and lipophilicity (P1_lipo) were not significant for position 1. However, the variable describing charge showed a significant effect in the model, by being positively correlated with inhibition of the complex formation (Figure 6). The score plot (Figure 3c) reveals that the positively charged Lys had a negative P1_charge value[41] and hence contributed to poor binders. Among the amino acids evaluated at this position Asn had the largest P1_charge value and therefore, according to the model, made the best contribution to inhibiting the complexation.

In the second position Thr, Asp, Gly, Lys, Phe, and Val were selected as building blocks in the design. Just as for the first position size was not an important variable. However, a hydrophilic, positively charged amino acid was preferred at this position, as indicated by significant values for the variables lipophilicity and

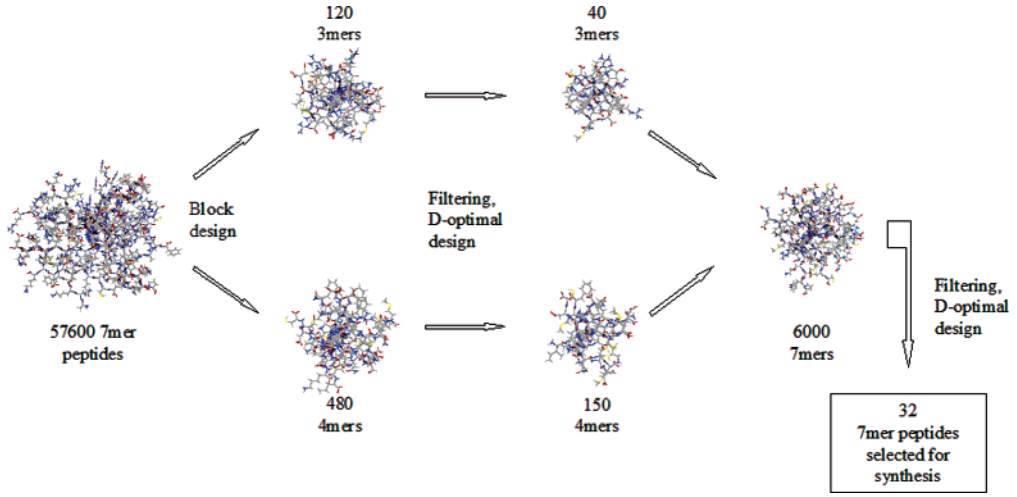


Figure 4. The design procedure. Due to the large number of possible heptamer peptides a block design was used. Three separate linear D-optimal designs were used, one in each filtered block and one on the filtered library of heptapeptides selecting 32 peptides for synthesis. Two different filtering procedures were utilized. First, the two blocks were biased to the native peptides (FimA, FimC, FimF, FimG) by requiring that at least one of the amino acids in each block should be present at the same position in one of the native peptides. Second, a manual filtering was made of the 6000 heptapeptides to improve synthetic feasibility.

Table 3. Amino Acid Sequences and Inhibition Data at Three Different Concentrations for the 32 Selected Peptides

										inhibition (%)[a]		
peptide	Pos1	Pos2	Pos3	Pos4	Pos5	Pos6	Pos7	Pos8	Pos9	10 μM	50 μM	200 μM
1	Thr	Gly	Gly	Ala	Gly	Ala	Leu	Ile	Ser	−12	17	3
2	Ala	Lys	Gly	Asp	Leu	Leu	Leu	Ile	Ser	3	13	14
3	Lys	Gly	Phe	Gln	Leu	Asp	Leu	Ile	Ser	−7	9	8
4	Asn	Lys	Phe	Phe	Leu	Ala	Leu	Ile	Ser	15	28	44
5	Asn	Val	Ser	Thr	Leu	His	Leu	Ile	Ser	18	30	44
6[b]	Leu	Phe	Leu	Thr	Phe	Leu	Leu	Ile	Ser	-	-	-
7	Lys	Asp	Leu	Ala	Ser	Thr	Leu	Ile	Ser	16	11	3
8	Leu	Val	Gly	Lys	Ser	His	Leu	Ile	Ser	1	0	5
9	Asn	Phe	Phe	Ala	Gly	His	Phe	Ile	Ser	−1	0	1
10	Asn	Asp	Leu	Lys	Gly	Leu	Phe	Ile	Ser	10	12	14
11	Lys	Val	Ser	Phe	Gly	Leu	Phe	Ile	Ser	20	17	29
12	Ala	Gly	Gly	Thr	Gly	Thr	Phe	Ile	Ser	−3	−1	−2
13	Lys	Lys	Leu	Lys	Leu	His	Phe	Ile	Ser	8	22	39
14[b]	Leu	Phe	Leu	Phe	Leu	Thr	Phe	Ile	Ser	-	-	-
15	Leu	Asp	Ser	Thr	Leu	Ala	Phe	Ile	Ser	10	25	37
16	Leu	Gly	Ser	Ala	Phe	Leu	Phe	Ile	Ser	19	29	40
17	Thr	Val	Leu	Asp	Phe	Asp	Phe	Ile	Ser	3	6	5
18	Asn	Thr	Gly	Gln	Phe	Ala	Phe	Ile	Ser	17	7	4
19	Asn	Gly	Ser	Asp	Ser	His	Phe	Ile	Ser	3	19	12
20	Ala	Phe	Ser	Lys	Ser	Asp	Phe	Ile	Ser	−5	−9	−6
21	Asn	Gly	Leu	Gln	Gly	Ala	Ser	Ile	Ser	10	19	24
22	Asn	Thr	Gly	Ala	Leu	Asp	Ser	Ile	Ser	2	11	15
23	Lys	Phe	Gly	Asp	Leu	Ala	Ser	Ile	Ser	4	1	7
24[b]	Thr	Lys	Ser	Gln	Phe	Thr	Ser	Ile	Ser	-	-	-
25	Ala	Asp	Phe	Phe	Phe	His	Ser	Ile	Ser	2	16	26
26	Thr	Thr	Phe	Thr	Ser	Leu	Ser	Ile	Ser	19	35	48
27	Leu	Thr	Phe	Asp	Gly	Thr	Trp	Ile	Ser	0	7	10
28	Ala	Thr	Ser	Phe	Gly	His	Trp	Ile	Ser	12	9	0
29	Thr	Asp	Gly	Gln	Leu	Leu	Trp	Ile	Ser	−8	1	4
30	Asn	Gly	Phe	Lys	Phe	Thr	Trp	Ile	Ser	2	−3	2
31	Ala	Val	Leu	Ala	Ser	Ala	Trp	Ile	Ser	0	8	18
32	Asn	Lys	Gly	Thr	Ser	Asp	Trp	Ile	Ser	3	8	10
33[c]	Asn	Thr	Leu	Gln	Leu	Ala	Leu	Ile	Ser	18	32	50
										24	42	59
										13	27	48

[a] The inhibitory power was obtained by measuring the ability to prevent FimC/FimH complexation in a competitive ELISA. [b] Peptides **6**, **14**, and **24** could not be successfully synthesized. [c] Peptide **33** corresponds to residues 101−109 of the wild-type FimC, except for having Leu instead of Ile at position 7 (107 in FimC). It was evaluated in triplicate.

charge, both of which were negatively correlated with the response.[41] The model indicates an interaction term between lipophilicity and charge, and such a term would strengthen the above preference even further. However, a square term of lipophilicity seems to be negatively correlated with the response, indicating that the optimum was within the tested region. These nonlinear terms were on the borderline of being significant, and taken together with the fact that the screening design supports linear terms, they should be treated with caution. In the design Lys was the amino acid best fulfilling the demand for hydrophilicity and charge.

Table 4. Summary of the Two Component PLS Models (multi-Y) Using Principal Properties of the Individual Amino Acids for Characterization of the Peptides and the % Inhibition at Three Different Peptide Concentrations as Responses

	model 1[a]			model 2[a]		
peptide concentration	R^2	Q^2(a)	Q^2(b)	R^2	Q^2(a)	Q^2(b)
10 μM	0.50	0.29	0.10	0.49	0.39	0.26
50 μM	0.78	0.50	0.38	0.79	0.59	0.60
200 μM	0.74	0.52	0.40	0.78	0.62	0.62

[a] Model 1: Ten linear terms included; Model 2: Ten linear terms and five interaction terms included; (a) Cross-validation using eight rounds; (b) Cross-validation using four rounds.

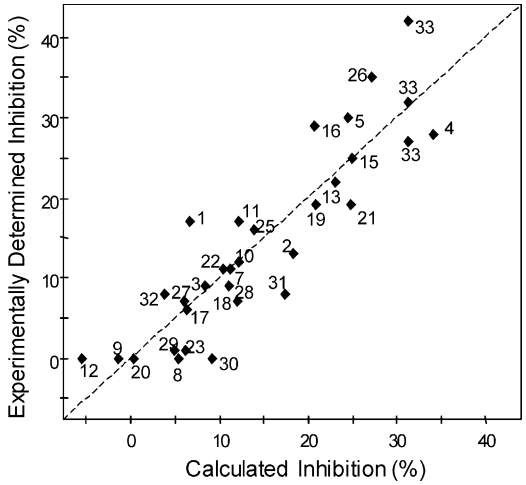


Figure 5. Calculated response values obtained using the QSAR (Model 2, Table 4) versus the experimentally determined values for inhibition of FimC/FimH complexation (expressed as % inhibition at 50 μM peptide concentration) for the synthesized peptides **1**−**33** listed in Table 3.

Table 5. Predicted and Experimental Inhibition Values for the Peptides in the External Test Set Using Model 2 (Table 4)

	predicted inhibition (%)			experimental inhibition (%)		
peptide[a]	10 μM	50 μM	200 μM	10 μM	50 μM	200 μM
FimC (**34**)	17	31	45	7	23	40
FimA[b] (**35**)	4	6	6	−3	1	3
FimF[b] (**36**)	12	20	27	−1	14	28
FimG[b] (**37**)	11	18	25	8	24	30

[a] Peptide sequences are given in Table 1. [b] In the model used for prediction, the term for position 1 was excluded due to missing values in the test set.

Positions 3 and 5 were believed to hold amino acids crucial for the β-strand formation, and a more cautious design was applied by varying only the two variables size and polarity. The amino acids incorporated at these positions were Leu, Gly, Phe, and Ser. The model based on the linear terms indicates that it is preferable to have large amino acids in these positions. However, it is important to point out that the selected set of amino acids was limited, and that the corner toward which the coefficients point was not included in the design. For position 3, polar residues also correlate positively with inhibitory power, although this was only scrutinized with Ser and should be further investigated. In addition, there were clear indications of nonlinearity as shown in an interaction term between size and polarity and a large negative square term for size. Position 5 also displayed a large negative square term for size. The negative correlation for the size square term at both positions indicates that the optimal choice of amino acids was actually within the current design.

In position 6 Ala, His, Thr, Asp, and Leu were used to investigate the properties size, lipophilicity, and charge. Here, lipophilicity was the most important property for influencing the ability of the peptides to inhibit FimC/FimH complexation. As long as the amino acid was lipophilic, the size appeared not to matter, indicating that both Ala and Leu were good choices at this position. Charged amino acids also showed a significant effect in the model, by being negatively correlated with the response. This leads to the conclusion that positively charged amino acids ought to increase the inhibitory power of the peptide (Figure 3b). Indeed all peptides having the negatively charged Asp at position 6 had no or only slight inhibition, whereas some of peptides with a His residue were among the best binders in the library.

In position 7, the varied amino acids were Phe, Leu, Ser, and Trp. The model show that in this position the size was crucial, and that small or intermediate sized amino acids were preferred. It is clear that Ser and Leu were good choices for this position. The polarity did not seem to have a large impact.

Examination of the crystal structure together with the interpretation of the QSAR model strengthen the assumption that short peptides corresponding to G_1 β-strands bind in the subunit's crevice in an antiparallel manner. The best match between QSAR and the crystal structure could be found if position 7 was placed in the pocket well defined by Ile181, Leu183, Val223, and Thr171 and position 6 between the lipophilic residues Val168 and Ile271. Accordingly, the nonpreferential position 4 was then placed in a solvent exposed region of the subunit and position 3 can be involved in π-stacking interactions with Phe276.

The above interpretation of the QSAR based on the screening design (Figures 3 and 6) showed clear directions of how to proceed toward improved inhibitory peptides in the next design step. In position 1 the best amino acid in this design, Asn, together with Tyr and Asp should be selected to further challenge the direction of the design toward charged amino acids. In position 2 the hydrophilic and positively charged Lys should be complemented by Gln and Arg. For position 3 and 5 there were indications of nonlinearity, and it would be interesting to expand the current investigation with large amino acids such as Tyr and Trp, in addition to the previously investigated Leu and Phe. For these positions it would also be interesting to map some nonnatural amino acids for a more thorough investigation of the property space. Position 4 should be held constant, e.g. as Gln. In position 6 Ala, His, or Leu appeared to be the best choices from the current investigation. The model indicates that Phe also could be a good choice, but considering histidines modest charge this position may also have to be evaluated with a more positively charged residue, e.g. Lys. For the final, 7th position, the model indicated that small or medium sized amino acids were beneficial for the inhibitory power. Therefore, it would be interesting to carry out a small investigation including Ala, Gly, Ser, and Thr at this position.

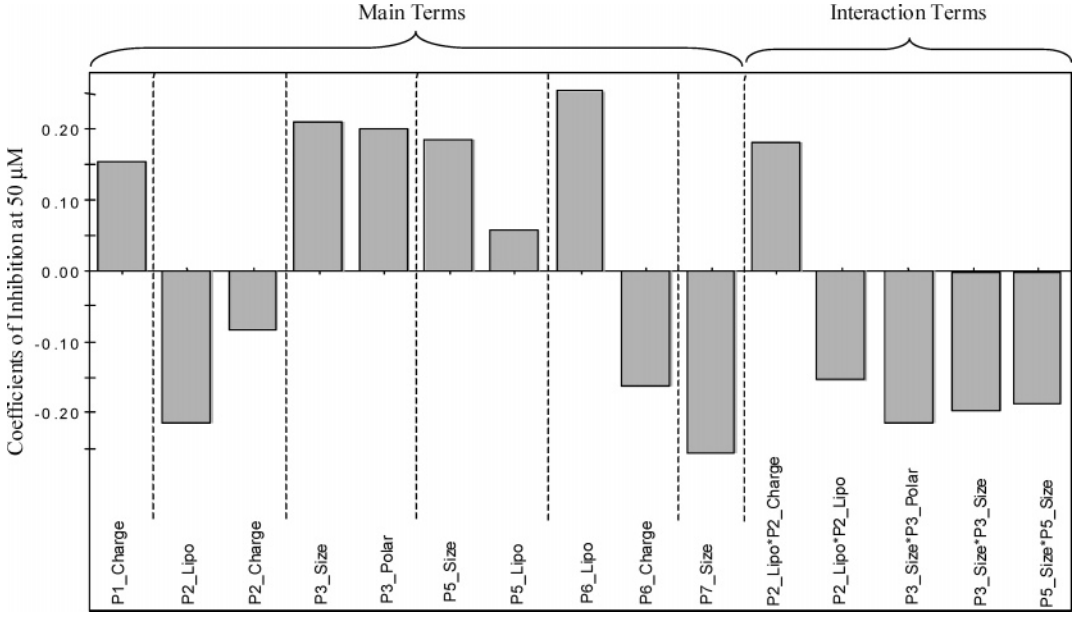


Figure 6. The model coefficients from the final QSAR (Model 2, Table 4) derived from % inhibition of FimC/FimH complexation at 50 μM peptide concentration. P1−P7 assigns from which position in the peptide the coefficient originates. In the label this is followed by a physical property, where Lipo stands for lipophilicity and Polar for polarity.

Conclusions

A library of nonamer peptides has been designed using statistical molecular design with the purpose of generating SAR for inhibition of the FimC/FimH complex, a protein−protein interaction crucial for pilus assembly in uropathogenic *Escherichia coli*. Seven of the positions in the nonamer scaffold were varied simultaneously investigating properties such as size, lipophilicity, and charge. A building block design procedure using prior information of native peptide inhibitors, the crystal structure of the FimC/FimH complex, as well as synthetic feasibility resulted in a library consisting of 32 peptides. After solid-phase synthesis the members of the peptide library was evaluated as inhibitors of FimC/FimH protein−protein complexation in an ELISA.

Novel peptides with the capability to inhibit the FimC/FimH protein−protein interaction to the same extent as the native FimC peptides were discovered among the 32 members of this initial, screening library. To our knowledge this is the first reported inhibition of the donor strand complementation mechanism by nonnative peptides. The discovery of biologically active nonnative peptides improves the possibilities to design peptides and peptidomimetics with optimized inhibitory powers.

The SMD approach of varying all positions at the same time together with the use of prior information resulted in a high-quality set of ligands for QSAR modeling, both in terms of structural variation and span in biological activity. The resulting model was verified by good predictions of an external test set of truncated native peptides and gave valuable information of the investigated properties for the amino acids at the seven varied positions. It was found that amino acids with negatively and positively charged side-chains were preferred at positions one and two, respectively. Large side-chains at positions three and five were predicted to enhance the inhibitory power, in agreement with the FimC/FimH crystal structure. These residues are located in the hydrophobic crevice of FimH. At positions six and seven, lipophilic and small/intermediate side-chains were preferred. Finally, position four in the peptide sequence did not have any significance for inhibition of the protein−protein interaction.

In summary, we have shown that an SMD approach can give highly informative QSAR data even when working with a greatly reduced peptide library, consisting of only 32 out of the 20^7 possible combinations. The multivariate QSAR model based on the designed library was successfully used for prediction and interpretation, giving valuable information about how to proceed toward improved inhibitory peptides in the next design step.

Experimental Section

Computational Methods. Characterization of Amino Acids. Both 2D and 3D descriptors were calculated from 3D structures of the individual amino acids with C-terminal amides and N-terminal acetylation. The structures were generated using the MOE protein builder module and subsequent energy minimization with the implemented Amber94 united-atom force field and an implicit solvent electrostatic correction model.[42] The amino acids were characterized with the MOE software, and the descriptors include properties of size, lipophilizity, polarizability, charge, flexibility, rigidity, and hydrogen-bonding capacities (Table 2). The structural descriptors were compressed by PCA, and the principal properties were further used in the modeling procedure.

Statistical Molecular Design and Data Analytical Methods. The individual amino acids were used as qualitative design variables and the building blocks were selected using D-optimal design.[43−45] The qualitative descriptors were set up so that for a model term with k levels there would be $k - 1$ expanded terms associated. For example, for a position A with five amino acids selected, there would be four ($5 - 1 = 4$) expanded terms representing that position. The D-optimal designs were performed using the MODDE software.[46] In the design of the final library the criterion that the selected peptides should have at least two out of seven varied positions identical with any of the wild-type subunits FimA, FimF, FimG, and FimC was added to the default criteria set by the program.[46]

Two data analytical methods were used: principal component analysis (PCA) for compression of the structural descriptors[47−49] and partial least squares projection to latent structures (PLS) to relate the structure descriptor matrix (**X**) to the activity matrix (**Y**).[50,51] The number of significant principal components from the PCA of the structural descriptor matrix was decided using their eigenvalues and the chemical interpretation of the loadings for corresponding components. The PCA calculations were made using the Simca software.[52]

In the PLS modeling, the peptides were described using principal properties of the amino acids in the sequence and % inhibition was used as the response. The PLS regression technique relates a latent variable in **X** to a latent variable in **Y**. The use of many responses simultaneously to determine a latent variable in **Y** based on their covariance, so-called multi-**Y**, gives the advantage of using more than one response to stabilize the modeling when based on experimental data containing noise.[53]

The number of linear terms was reduced by an interactive procedure excluding terms with low coefficient values until Q^2 decreased. To investigate indications of nonlinearity, square and interaction terms were added to the significant main factors. In a manner similar to the linear terms, the interaction terms were reduced based on their coefficient values until Q^2 decreased. The modeling was performed using the MODDE software.[46]

The number of significant components was decided by cross-validation[54] using two independent cross-validation rounds (four and eight classes) where the data had been left out throughout the modeling procedure. The cross-validated Q^2 was calculated according to:

$$Q^2 = 1 - \frac{\text{PRESS}}{\text{SS}}$$

PRESS is the predicted residual sum of squares when all objects have been left out of the modeling once and SS is the total sum of squares of **Y** corrected for the mean.[46] The final model was validated using an external test set. The models yielding the predicted residuals needed for the Q^2 calculations and the predictions of the external test set were made using the Simca software.[52]

Synthesis, Purification, and Characterization of the Library. The peptides **13**−**33** were synthesized as C-terminal amides by solid-phase peptide synthesis, in syringe-reactors on a cross-linked polystyrene resin grafted with poly(ethylene glycol) spacers (Argogel-Rink-NH−Fmoc, 150 μm, capacity 0.34 mmol/g). Peptides **1**−**12** and **34**−**37** were prepared on a cross-linked polystyrene resin with the Rink linker (PS−RINK−NH−Fmoc, 96 μm, capacity 0.9 mmol/g). The amount of resin was adjusted to theoretically generate 100 μmol of peptide (294 mg of Argogel-resin or 111 mg of PS-resin). Reagent solutions and DMF (dimethylformamide) for washing were added manually to the reactor. N^{α}-Fmoc (fluorenylmethoxycarbonyl) amino acids (Bachem AG, Bubendorf, Switzerland) with the following protective groups were used: *tert*-butyl for aspartic acid, *tert*-butyl for threonine and serine, triphenylmethyl (Trt) for asparagine, glutamine, and histidine, 2,2,5,7,8-pentamethylchroman-6-sulfonyl (Pmc) for arginine, and *tert*-butoxycarbonyl (Boc) for lysine and tryptophan. The N^{α}-Fmoc amino acids were activated in situ as 1-benzotriazolyl esters. Activation was performed by reaction of the appropriate N^{α}-Fmoc amino acid (400 μmol), 1-hydroxybenzotriazole (HOBT, 81 mg 600 μmol) for peptides **1**−**12** and **34**−**37**, or 7-aza-1-hydroxybenzotriazole (HOAt, 34 mg, 600 μmol) for peptides **13**−**33**, and *N*,*N*′-diisopropylcarbodiimide (60.4 μL, 390 μmol). The coupling reaction was monitored by bromophenol blue (37.5 μL of a 2 mM solution in DMF) and by the Kaiser test in necessary cases. N^{α}-Fmoc deprotection of the peptide resin was achieved by treatment with 20% piperidine in DMF during 10 min. A synthesizer block was used which made it possible to synthesize 11 peptides at the same time. After completion of the synthesis, the resin carrying the protected peptide was N^{α}-Fmoc deprotected, washed with CH_2Cl_2 several times, and dried under vacuum. The peptides were cleaved from the resin, and the side chains were deprotected by treatment with trifluoroacetic acid/water (95/5, 10 mL) or with trifluoroacetic acid/water/thioanisole/ethanedithiole (35/2/2/1, 40 mL), for 3 h at 40 °C followed by filtration (the latter cleavage method was used for peptides containing the protective groups Pmc or Trt, Peptides **3**−**5**, **8**−**10**, **13**, **18**, **19**, **21**, **22**, **24**, **25**, **28**−**30**, **32**−**37**). The resin was washed with acetic acid, before the filtrate was concentrated. Acetic acid was added again, and the solution was reconcentrated several times until the concentrated residue was dry. The residue was triturated with cold diethyl ether, which gave a solid crude peptide that was dissolved in a mixture of acetic acid and water and lyophilized. Purification was performed by preparative reversed phase HPLC (Beckman System Gold HPLC, detection at 214 nm) using a linear gradient of 0%-100% CH_3CN (0.1% TFA) in H_2O (0.1% TFA) during 60 min. For analytical HPLC a Kromasil C_8 column (100 Å, 5 μm, 25 × 4.6 mm, Hichrom Ltd., Berkshire, UK) was used with a flow rate of 1.5 mL/min. For preparative HPLC a Kromasil C_8 column (100 Å, 5 μm, 250 × 20 mm, Hichrom Ltd., Berkshire, UK) was used with a flow rate of 12 mL/min. High-resolution positive fast atom bombardment mass spectra (FAB-MS), recorded on a JEOL SX-102-A mass spectrometer (ions were produced by a beam of Xe atoms (6 keV) from a matrix of glycerol and thioglycerol), was used for characterization of the peptides. Synthesis and purification of peptides **6**, **14**, and **24** (see Table 3) did not succeed under the above conditions, probably because they were too lipophilic.

Thr-Gly-Gly-Ala-Gly-Ala-Leu-Ile-Ser-NH_2 (1). 67% yield. FAB-MS: 745 $(M + H)^+$, Calculated 745.

Ala-Lys-Gly-Asp-Leu-Leu-Leu-Ile-Ser-NH_2 (2). 54% yield. FAB-MS: 927 $(M + H)^+$, Calculated 927.

Lys-Gly-Phe-Gln-Leu-Asp-Leu-Ile-Ser-NH_2 (3). 62% yield. FAB-MS: 1018 $(M + H)^+$, Calculated 1018.

Asn-Lys-Phe-Phe-Leu-Ala-Leu-Ile-Ser-NH_2 (4). 64% yield. FAB-MS: 1050 $(M + H)^+$, Calculated 1050.

Asn-Val-Ser-Thr-Leu-His-Leu-Ile-Ser-NH_2 (5). 57% yield. FAB-MS: 981 $(M + H)^+$, Calculated 981.

Lys-Asp-Leu-Ala-Ser-Thr-Leu-Ile-Ser-NH_2 (7). 47% yield. FAB-MS: 945 $(M + H)^+$, Calculated 945.

Leu-Val-Gly-Lys-Ser-His-Leu-Ile-Ser-NH_2 (8). 47% yield. FAB-MS: 951 $(M + H)^+$, Calculated 951.

Asn-Phe-Phe-Ala-Gly-His-Phe-Ile-Ser-NH_2 (9). 68% yield. FAB-MS: 1037 $(M + H)^+$, Calculated 1037.

Asn-Asp-Leu-Lys-Gly-Leu-Phe-Ile-Ser-NH_2 (10). 28% yield. FAB-MS: 1004 $(M + H)^+$, Calculated 1004.

Lys-Val-Ser-Phe-Gly-Leu-Phe-Ile-Ser-NH_2 (11). 32% yield. FAB-MS: 995 $(M + H)^+$, Calculated 995.

Ala-Gly-Gly-Thr-Gly-Thr-Phe-Ile-Ser-NH_2 (12). 31% yield. FAB-MS: 808 $(M + H)^+$, Calculated 808.

Lys-Lys-Leu-Lys-Leu-His-Phe-Ile-Ser-NH_2 (13). 28% yield. FAB-MS: 1111 $(M + H)^+$, Calculated 1111.

Leu-Asp-Ser-Thr-Leu-Ala-Phe-Ile-Ser-NH_2 (15). 26% yield. FAB-MS: 964 $(M + H)^+$, Calculated 964.

Leu-Gly-Ser-Ala-Phe-Leu-Phe-Ile-Ser-NH_2 (16). 43% yield. FAB-MS: 952 $(M + H)^+$, Calculated 952.

Thr-Val-Leu-Asp-Phe-Asp-Phe-Ile-Ser-NH_2 (17). 24% yield. FAB-MS: 1054 $(M + H)^+$, Calculated 1054.

Asn-Thr-Gly-Gln-Phe-Ala-Phe-Ile-Ser-NH_2 (18). 74% yield. FAB-MS: 982 $(M + H)^+$, Calculated 982.

Asn-Gly-Ser-Asp-Ser-His-Phe-Ile-Ser-NH_2 (19). 72% yield. FAB-MS: 961 $(M + H)^+$, Calculated 961.

Ala-Phe-Ser-Lys-Ser-Asp-Phe-Ile-Ser-NH_2 (20). 39% yield. FAB-MS: 999 $(M + H)^+$, Calculated 999.

Asn-Gly-Leu-Gln-Gly-Ala-Ser-Ile-Ser-NH_2 (21). 11% yield. FAB-MS: 844 $(M + H)^+$, Calculated 844.

Asn-Thr-Gly-Ala-Leu-Asp-Ser-Ile-Ser-NH_2 (22). 37% yield. FAB-MS: 875 $(M + H)^+$, Calculated 875.

Lys-Phe-Gly-Asp-Leu-Ala-Ser-Ile-Ser-NH_2 (23). 81% yield. FAB-MS: 935 $(M + H)^+$, Calculated 935.

Ala-Asp-Phe-Phe-Phe-His-Ser-Ile-Ser-NH_2 (25). 73% yield. FAB-MS: 1068 $(M + H)^+$, Calculated 1068.

Thr-Thr-Phe-Thr-Ser-Leu-Ser-Ile-Ser-NH_2 (26). 45% yield. FAB-MS: 954 $(M + H)^+$, Calculated 954.

Leu-Thr-Phe-Asp-Gly-Thr-Trp-Ile-Ser-NH_2 (27). 14% yield. FAB-MS: 1037 $(M + H)^+$, Calculated 1037.

Ala-Thr-Ser-Phe-Gly-His-Trp-Ile-Ser-NH_2 (28). 51% yield. FAB-MS: 1003 $(M + H)^+$, Calculated 1003.

Thr-Asp-Gly-Gln-Leu-Leu-Trp-Ile-Ser-NH_2 (29). 63% yield. FAB-MS: 1030 $(M + H)^+$, Calculated 1030.

Asn-Gly-Phe-Lys-Phe-Thr-Trp-Ile-Ser-NH_2 (30). 17% yield. FAB-MS: 1097 $(M + H)^+$, Calculated 1097.

Ala-Val-Leu-Ala-Ser-Ala-Trp-Ile-Ser-NH_2 (31). 34% yield. FAB-MS: 915 $(M + H)^+$, Calculated 915.

Asn-Lys-Gly-Thr-Ser-Asp-Trp-Ile-Ser-NH_2 (32). 60% yield. FAB-MS: 1005 $(M + H)^+$, Calculated 1005.

Asn-Thr-Leu-Gln-Leu-Ala-Leu-Ile-Ser-NH_2 (33). 58% yield. FAB-MS: 970 $(M + H)^+$, Calculated 970.

FimC101−109 (34). 40% yield. FAB-MS: 970 $(M + H)^+$, Calculated 970.

FimA9−17 (35). 57% yield. FAB-MS: 930 $(M + H)^+$, Calculated 930.

FimF1−9 (36). 64% yield. FAB-MS: 1024 $(M + H)^+$, Calculated 1024.

FimG1−9 (37). 48% yield. FAB-MS: 945 $(M + H)^+$, Calculated 945.

Biological Evaluation. A stock solution of FimC (0.051 mg/mL) in phosphate-buffered saline (PBS; 120 mM NaCl, 2.7 mM KCl, 10 mM PBS, pH 7.4) was coated overnight onto microtiter wells (immulon 4HBX microtiter plates, Thermo Labsystems, Stockholm, Sweden) with 50 μL/well at 4 °C. The wells were then washed three times with PBS (with 10 mg/L NaN_3, 0.025% Tween) followed by blocking with 3% bovine serum albumin (BSA) in PBS (BSA-PBS) for 1 h at room temperature (RT). The wells were washed three times with PBS (with 10 mg/L NaN_3, 0.025% Tween) and incubated with a 50 μL solution of 3% BSA-PBS containing 2 μL peptide in DMSO (0.5, 2.5 and 10 nmol) and 2 μL FimH in 3 M urea, 20 mM MES, pH 6.8 (1 pmol), for 45 min at RT. 2 μL DMSO was used as a positive control, and 2 μL of 3 M urea, 20 mM MES, pH 6.8, was used as blank. The wells were washed three times with PBS (with 10 mg/L NaN_3, 0.025% Tween) and incubated with a 1:1000 dilution of mouse anti-FimH antiserum in 3% BSA-PBS for 45 min at RT. The wells were washed three times with PBS (with 10 mg/L NaN_3, 0.025% Tween) and then incubated with a 1:1000 dilution of goat antiserum to mouse IgG (immunoglobulin G) coupled to alkaline phosphatase (Sigma) in 3% BSA-PBS for 45 min at RT. The wells were washed three times with PBS (with 10 mg/L NaN_3, 0.025% Tween) and three times with developing buffer (10 mM diethanolamine, 0.5 mM $MgCl_2$). For developing, 50 μL of substrate (50 μL of filtered 1 mg/mL *p*-nitrophenyl phosphate: Sigma) in developing buffer was added. The reaction was incubated in the dark at RT for 1 h, and the absorbance at 405 nm was read (SpectraMax 340, Molecular Devices, Sunnyvale, CA).

The inhibition values were obtained as the amount of FimH binding to the coated FimC chaperone in the presence of peptide divided by the amount of FimH binding in the absence of peptide (all values obtained from averages of four replicates). Due to experimental variation and the calculation method, negative values were obtained for some of the peptides. In the modeling procedure these have been set to zero.

Acknowledgment. This work was funded by grants from the Swedish Research Council, the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine, Knut & Alice Wallenberg foundation, and J. C. Kempes Minnes Stipendiefond.

References

(1) Hooton, T. M.; Stamm, W. E. Diagnosis and treatment of uncomplicated urinary tract infection. *Infect. Dis. Clin. North Am.* **1997**, *11*, 551−581.
(2) Svanborg, C.; Godaly, G. Bacterial virulence in urinary tract infection. *Infect. Dis. Clin. North Am.* **1997**, *11*, 513−529.
(3) Johnson, J. R. Virulence factors in *Escherichia coli* urinary-tract infection. *Clin. Microbiol. Rev.* **1991**, *4*, 80−128.
(4) Mulvey, M. A.; Schilling, J. D.; Martinez, J. J.; Hultgren, S. J. Bad bugs and beleaguered bladders: Interplay between uropathogenic *Escherichia coli* and innate host defenses. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8829−8835.
(5) Kallenius, G.; Mollby, R.; Svenson, S. B.; Winberg, J.; Lundblad, A.; Svensson, S.; Cedergren, B. The Pk antigen as receptor for the haemagglutinin of pyelonephritic *Escherichia coli*. *FEMS Microbiol. Lett.* **1980**, *7*, 297−302.
(6) Leffler, H.; Eden, C. S. Chemical identification of a glycosphingolipid receptor for *Escherichia coli* attaching to human urinary tract epithelial cells and agglutinating human erythrocytes. *FEMS Microbiol. Lett.* **1980**, *8*, 127−134.
(7) Krogfelt, K. A.; Bergmans, H.; Klemm, P. Direct Evidence That the Fimh Protein Is the Mannose-Specific Adhesin of Escherichia-Coli Type-1 Fimbriae. *Infect. Immun.* **1990**, *58*, 1995−1998.
(8) Abraham, S. N.; Sun, D. X.; Dale, J. B.; Beachey, E. H. Conservation of the D-Mannose-Adhesion Protein among Type-1 Fimbriated Members of the Family Enterobacteriaceae. *Nature* **1988**, *336*, 682−684.
(9) Mulvey, M. A. Adhesion and entry of uropathogenic *Escherichia coli*. *Cell. Microbiol.* **2002**, *4*, 257−271.
(10) Hung, D. L.; Knight, S. D.; Woods, R. M.; Pinkner, J. S.; Hultgren, S. J. Molecular basis of two subfamilies of immunoglobulin-like chaperones. *EMBO J.* **1996**, *15*, 3792−3805.
(11) Thanassi, D. G.; Saulino, E. T.; Hultgren, S. J. The chaperone/usher pathway: a major terminal branch of the general secretory pathway. *Curr. Opin. Microbiol.* **1998**, *1*, 223−231.
(12) Hultgren, S. J.; Abraham, S.; Caparon, M.; Falk, P.; Stgeme, J. W.; Normark, S. Pilus and nonpilus bacterial adhesins − Assembly and function in cell recognition. *Cell* **1993**, *73*, 887−901.
(13) Sauer, F. G.; Pinkner, J. S.; Waksman, G.; Hultgren, S. J. Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* **2002**, *111*, 543−551.
(14) Pellecchia, M.; Guntert, P.; Glockshuber, R.; Wuthrich, K. NMR solution structure of the periplasmic chaperone FimC. *Nat. Struct. Biol.* **1998**, *5*, 885−890.
(15) Holmgren, A.; Branden, C. I. Crystal-structure of chaperone protein PapD reveals an immunoglobulin fold. *Nature* **1989**, *342*, 248−251.
(16) Choudhury, D.; Thompson, A.; Stojanoff, V.; Langermann, S.; Pinkner, J.; Hultgren, S. J.; Knight, S. D. X-ray structure of the FimC−FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* **1999**, *285*, 1061−1066.
(17) Sauer, F. G.; Futterer, K.; Pinkner, J. S.; Dodson, K. W.; Hultgren, S. J.; Waksman, G. Structural basis of chaperone function and pilus biogenesis. *Science* **1999**, *285*, 1058−1061.
(18) Stephens, C.; Shapiro, L. Bacterial protein secretion − a target for new antibiotics? *Chem. Biol.* **1997**, *4*, 637−641.
(19) Cochran, A. G. Antagonists of protein−protein interactions. *Chem. Biol.* **2000**, *7*, R85−R94.
(20) Berg, T. Modulation of protein−protein interactions with small organic molecules. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 2462−2481.
(21) Toogood, P. L. Inhibition of protein−protein association by small molecules: Approaches and progress. *J. Med. Chem.* **2002**, *45*, 1543−1558.
(22) Larsson, A.; Barnhart, M.; Stenström, T.; Pinkner, J. S.; Hultgren, S. J.; Almqvist, F.; Kihlberg, J. Manuscript in preparation.
(23) McDowell, R. S.; Blackburn, B. K.; Gadek, T. R.; McGee, L. R.; Rawson, T.; Reynolds, M. E.; Robarge, K. D.; Somers, T. C.; Thorsett, E. D.; Tischler, M.; Webb, R. R.; Venuti, M. C. From peptide to nonpeptide. 2. The De-novo design of potent, nonpeptidal inhibitors of platelet-aggregation based on a benzodiazepinedione scaffold. *J. Am. Chem. Soc.* **1994**, *116*, 5077−5083.
(24) McDowell, R. S.; Gadek, T. R.; Barker, P. L.; Burdick, D. J.; Chan, K. S.; Quan, C. L.; Skelton, N.; Struble, M.; Thorsett, E. D.; Tischler, M.; Tom, J. Y. K.; Webb, T. R.; Burnier, J. P. From peptide to nonpeptide. 1. The Elucidation of a bioactive conformation of the arginine-glycine-aspartic acid recognition sequence. *J. Am. Chem. Soc.* **1994**, *116*, 5069−5076.
(25) Fan, X. D.; Flentke, G. R.; Rich, D. H. Inhibition of HIV-1 protease by a subunit of didemnaketal A. *J. Am. Chem. Soc.* **1998**, *120*, 8893−8894.
(26) Moss, N.; Deziel, R.; Adams, J.; Aubry, N.; Bailey, M.; Baillet, M.; Beaulieu, P.; Dimaio, J.; Duceppe, J. S.; Ferland, J. M.; Gauthier, J.; Ghiro, E.; Goulet, S.; Grenier, L.; Lavallee, P.; Lepinefrenette, C.; Plante, R.; Rakhit, S.; Soucy, F.; Wernic, D.; Guindon, Y. Inhibition of Herpes-Simplex virus Type-1 ribonucleotide reductase by substituted tetrapeptide derivatives. *J. Med. Chem.* **1993**, *36*, 3005−3009.
(27) Bottger, A.; Bottger, V.; GarciaEcheverria, C.; Chene, P.; Hochkeppel, H. K.; Sampson, W.; Ang, K.; Howard, S. F.; Picksley, S. M.; Lane, D. P. Molecular characterization of the hdm2-p53 interaction. *J. Mol. Biol.* **1997**, *269*, 744−756.
(28) Marshall, G. R. A hierarchical approach to peptidomimetic design. *Tetrahedron* **1993**, *49*, 3547−3558.

(29) Hellberg, S.; Sjostrom, M.; Skagerberg, B.; Wold, S. Peptide Quantitative Structure−Activity-Relationships, a multivariate approach. *J. Med. Chem.* **1987**, *30*, 1126−1135.
(30) Linusson, A.; Wold, S.; Norden, B. Statistical molecular design of peptoid libraries. *Mol. Divers.* **1998**, *4*, 103−114.
(31) Linusson, A.; Gottfries, J.; Lindgren, F.; Wold, S. Statistical molecular design of building blocks for combinatorial chemistry. *J. Med. Chem.* **2000**, *43*, 1320−1328.
(32) Baroni, M.; Clementi, S.; Cruciani, G.; Kettanehwold, N.; Wold, S. D-Optimal designs in QSAR. *Quant. Struct.-Act. Relat.* **1993**, *12*, 225−231.
(33) Sjöström, M.; Eriksson, L. Applications of statistical experimental design and PLS modeling in QSAR. *Chemometric methods in molecular design*; VCH: Weinheim, 1995; pp 63−90.
(34) Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K.; Moos, W. H. Measuring diversity − Experimental design of combinatorial libraries for drug discovery. *J. Med. Chem.* **1995**, *38*, 1431−1436.
(35) Martin, E. J.; Critchlow, R. E. Beyond mere diversity: Tailoring combinatorial libraries for drug discovery. *J. Comb. Chem.* **1999**, *1*, 32−45.
(36) Lundstedt, T.; Clementi, S.; Cruciani, G.; Pastor, M.; Kettaneh, N.; Andersson, P. M.; Linusson, A.; Sjöström, M.; Wold, S.; Nordén, B. Intelligent combinatorial libraries. *Computer-Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry*; Verlag Helvetica Chimica Acta: Basel, 1997; pp 190−208.
(37) Mee, R. P.; Auton, T. R.; Morgan, P. J. Design of active analogues of a 15-residue peptide using D-optimal design, QSAR and a combinatorial search algorithm. *J. Pept. Res.* **1997**, *49*, 89−102.
(38) Alifrangis, L. H.; Christensen, I. T.; Berglund, A.; Sandberg, M.; Hovgaard, L.; Frokjaer, S. Structure−property model for membrane partitioning of oligopeptides. *J. Med. Chem.* **2000**, *43*, 103−113.
(39) Andersson, K.; Choulier, L.; Hamalainen, M. D.; van Regenmortel, M. H. V.; Altschuh, D.; Malmqvist, M. Predicting the kinetics of peptide-antibody interactions using a multivariate experimental design of sequence and chemical space. *J. Mol. Recognit.* **2001**, *14*, 62−71.
(40) Bork, P.; Holm, L.; Sander, C. The Immunoglobulin fold − Structural classification, sequence patterns and common core. *J. Mol. Biol.* **1994**, *242*, 309−320.
(41) Note that the PCA has rotated in such a way that positively charged residues gains a negative score-value for charge and vice versa.
(42) *MOE*; 2003.02 ed.; Chemical Computing Group Inc.: 1010 Sherbrooke St. West Suite 910 Montreal, Canada H3A 2R7.
(43) Mitchell, T. J. An algorithm for the construction of "D-optimal" experimental designs. *Technometrics* **1974**, *16*, 203−210.
(44) St. John, R. C.; Draper, N. R. D-optimality for regression designs: a review. *Technometrics* **1975**, *17*, 15−23.
(45) Dumouchel, W.; Jones, B. A simple bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics* **1994**, *36*, 37−47.
(46) *Modde 6.0*; Umetrics: Box 7960, S-907 19 Umeå, Sweden.
(47) Wold, H. Estimation of principal components and related models by iterative least squares. *Multivariate analysis*; Academic press: New York, 1966; pp 391−420.
(48) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37−52.
(49) Jackson, J. E. *A user's guide to principal components*; Wiley: New York, 1991.
(50) Wold, H. Soft modeling. The basic design and some extensions. *Systems under indirect observation*; North-Holland Publishing Company: Amsterdam, 1982; pp 1−53.
(51) Wold, S. PLS for multivariate linear modeling. *Chemometric methods in molecular design*; VCH: Weinheim, 1995; pp 195−218.
(52) *Simca-P 10.02*; Umetrics: Box 7960, S-907 19 Umeå, Sweden.
(53) Burnham, A. J.; MacGregor, J. F.; Viveros, R. Latent variable multivariate regression modeling. *Chemom. Intell. Lab. Syst.* **1999**, *48*, 167−180.
(54) Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. Royal Stat. Soc.* **1974**, *36 B*, 111−133.

JM040818L