Supporting information for

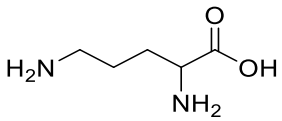
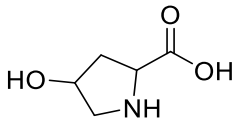
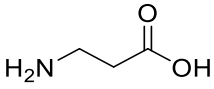
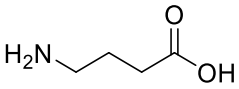
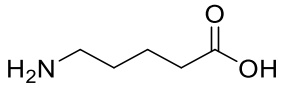
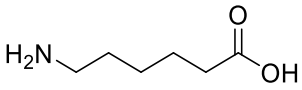
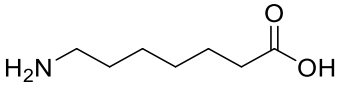
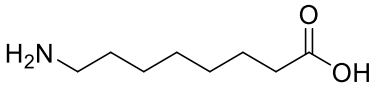
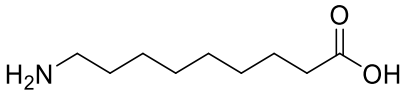
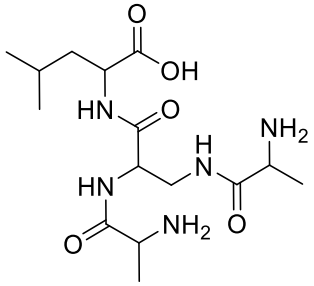
Populating Chemical Space with Peptides using a Genetic Algorithm

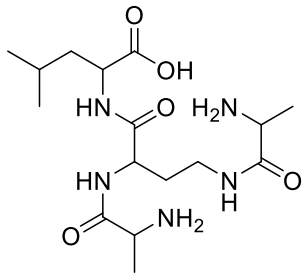
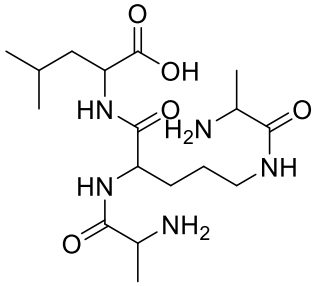
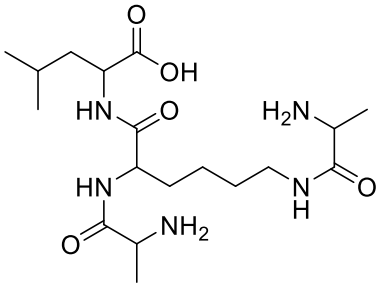
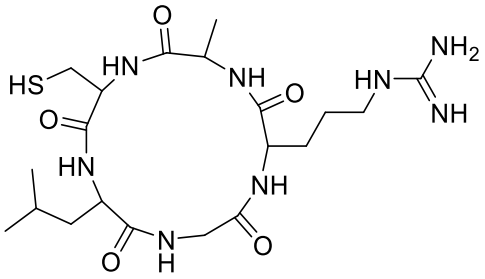
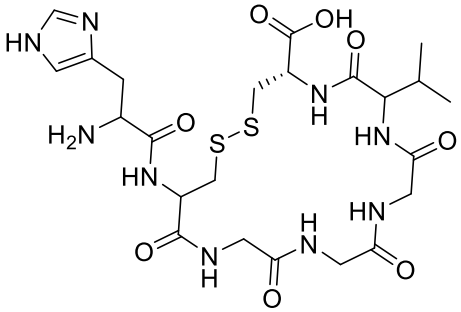
Alice Capecchi, Alain Zhang and Jean-Louis Reymond*

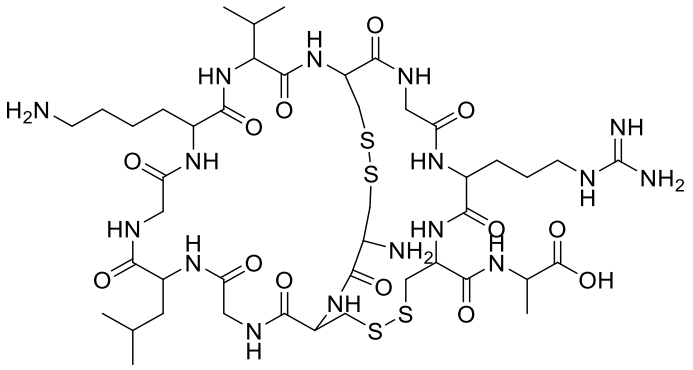
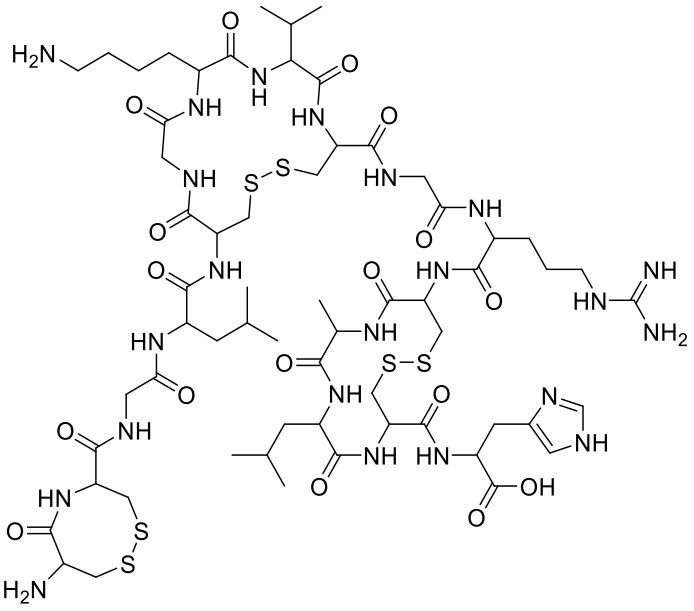
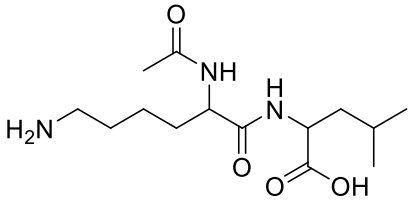
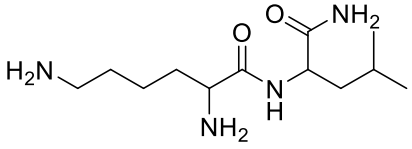
Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland

Email: jean-louis.reymond@dcb.unibe.ch

Table S1. PDGA recognized symbols and their correspondent internal character.

Symbol[a)]	**id**[b)]	**Description**	**Example**
Orn	O	Ornithine	
Hyp	Z	Hydroxyproline	
bAla	!	Beta-alanine	
Gaba	?	Gamma-aminobutyric acid	
a5a	=	Delta-aminopentanoic acid	
a6a	%	Epsilon-aminohexanoic acid	
a7a	$	Zeta-aminoheptanoic acid	
a8a	@	Eta-aminooctanoic acid	
a9a	#	Theta-aminononaanoic acid	
Dap	1	2,3-diaminopropionic acid as branching unit	Ala-Dap-Leu 
Dab	2	2,4-diaminobutyric acid as branching unit	Ala-Dab-Leu

			
BOrn	3	Ornithine as branching unit	Ala-BOrn-Leu 
BLys	4	Lysine as branching unit	Ala-BLys-Leu 
cy	X	Amide bond head-to-tail cyclization. It is always placed at the beginning (left, N terminus) of the sequence.	Cy-Arg-Ala-Cys-Leu-Gly 
Cys1	Ä	First pair of cyclizes cysteines, Always in pair, never next to each other.	His-Cys1-Gly-Gly-Gly-Val-Cys1 

Cys2	Ö	Second pair of cyclizes cysteines. They are always present in pair, never next to each other, present only if Cys1 is already part of the sequence.	Cys1-Cys2-Gly-Leu-Gly-Lys-Val-Cys1-Gly-Arg-Cys2-Ala 
Cys3	Ü	Third pair of cyclizes cysteines. They are always present in pair, never next to each other, present only if Cys1 and Cys2 are already part of the sequence.	Cys1-Cys1-Gly-Leu-Cys2-Gly-Lys-Val-Cys2-Gly-Arg-Cys3-Ala-Leu-Cys3-His 
Ac	&	N-terminus acetylation. It is always placed at the beginning (N-terminus, left) of the sequence	Ac-Lys-Leu 
NH2	+	C-terminus amide. It is always placed at the end (C-terminus, right) of the sequence	Lys-Leu-NH2 

[a] PDGA input. [b] Internally used characters. [c] Lower case.

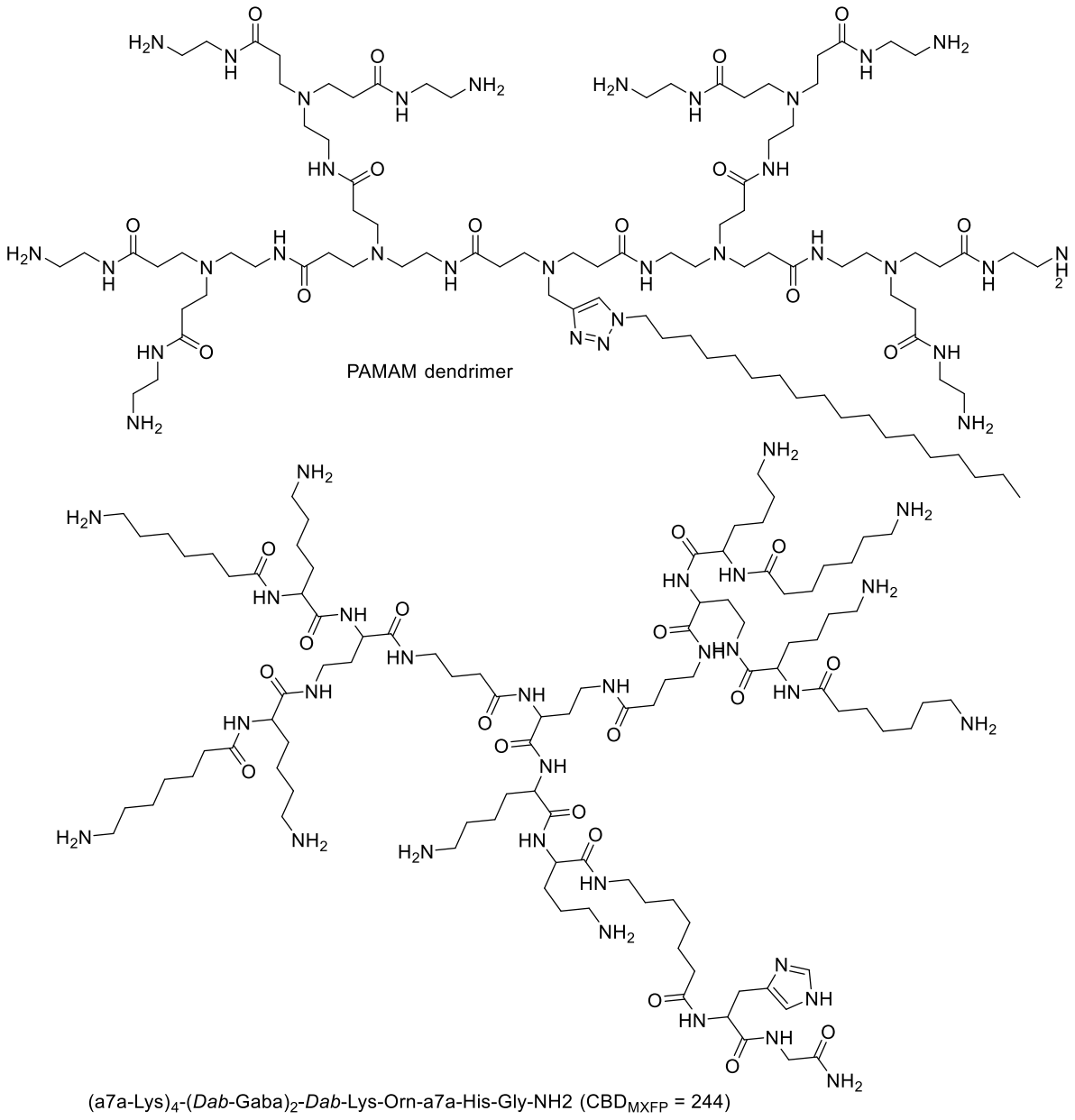


Figure S1. Structures of the non-peptidic PAMAM target and its best analog.

Table S2. SMARTS HBA/HBD and charge assignment.

Description	**SMARTS**	**HBA/HBD** [a) b)]	**Charge** [b)]
Aliphatic carbon	[C]	0	0
Nitrogen	[#7]	1	0
Tertiary nitrogen	[NX3;H0]	0	0
Aliphatic oxygen	[O]	2	0
Thiol	[SH1X2]	1	0
Hydroxyl at aliphatic carbon	[$([OHX2]-[CX4])]	3,0	0,0
Tyrosine	N[CX4H1]([CH2X4][cX3]1[cX3H][cX3H][cX3]([OHX2,OH0X1-])[cX3H][cX3H]1)[CX3]=[OX1]	1,0,0,0,0,0,0,3,0,0,0,2	0,0,0,0,0,0,0,0,0,0,0,0
Proline	N1[CX4H]([CH2][CH2][CH2]1)[CX3](=[OX1])	0,0,0,0,0,0,2	0,0,0,0,0,0,0
Histidine	N[CX4H]([CH2X4][#6X3]1:[$([#7X3H+,#7X2H0+0]:[#6X3H]:[#7X3H]),$([#7X3H])]:[#6X3H]:[$([#7X3H+,#7X2H0+0]:[#6X3H]:[#7X3H]),$([#7X3H])]:[#6X3H]1)[CX3](=[OX1])	1,0,0,0,2,0,1,0,0,2	0,0,0,0,0,0,0,0,0,0
Charged nitrogen	[$([NH2X3]-[CX4]),$([N]=[CX3])]	1,0	1,0
Carbonyl	[$([O]=[C])]	2,0	0,0
Carboxyl	[OH,O-]-[C](=O)	2,0,2	-1,0,0
Ether	[OX2]([CX4])[CX4]	2,0,0	0,0,0
Phenol	[OH1X2]-[c]	3,0	0,0

[a)] 0 = no hydrogen donor or acceptor site; 1 = donor site; 2 = acceptor site; 3= donor and acceptor site. [b)] When more values are present, they refer to the SMARTS atom in the correspondent position.

Table S3. Input parameters and excluded building blocks (bb).

Target	**N**	**M**	**G**	**Treshold CBD**	**Topology**	**Excluded bb** [a)]	**Time limit**
Indolicidin	50	1	0.8	300	linear	Hyp, Orn, bAla, Gaba, a5a, a6a, a7a, a8a, a9a, Ac	24 h
Indolicidin SEQSIM	50	1	0.8	5	linear	Hyp, Orn, bAla, Gaba, a5a, a6a, a7a, a8a, a9a, Ac	6 h
Tyrocidine A	50	1	0.8	300	cyclic	bAla, Gaba, a5a, a6a, a7a, a8a, a9a, Ac	24 h
ω-conotoxin-MVIIA	50	1	0.8	300	polycyclic	Hyp, Orn, bAla, Gaba, a5a, a6a, a7a, a8a, a9a, Ac	72 h
G3KL	50	1	0.8	300	dendrimer	Hyp, Orn, bAla, Gaba, a5a, a6a, a7a, a8a, a9a, Ac	48 h
Acetyl-CoA	50	1	0.8	300	linear	Ac	24 h
Epothilone A	50	1	0.8	300	cyclic	Ac	24 h
Cholic Acid	50	1	0.8	300	cyclic	Ac	24 h
α-cyclodextrin	50	1	0.8	300	cyclic	Ac	24 h
PAMAM dendrimer	50	1	0.8	300	dendrimer	Ac	24 h

[a)] For a definition of the mentioned building blocks refer to Table S1.