Peptide Design

International Edition: DOI: 10.1002/anie.201504018
German Edition: DOI: 10.1002/ange.201504018

Multidimensional Design of Anticancer Peptides**

*Yen-Chu Lin, Yi Fan Lim, Erica Russo, Petra Schneider, Lea Bolliger, Adriana Edenharter, Karl-Heinz Altmann, Cornelia Halin, Jan A. Hiss, and Gisbert Schneider**

Abstract: *The computer-assisted design and optimization of peptides with selective cancer cell killing activity was achieved through merging the features of anticancer peptides, cell-penetrating peptides, and tumor-homing peptides. Machine-learning classifiers identified candidate peptides that possess the predicted properties. Starting from a template amino acid sequence, peptide cytotoxicity against a range of cancer cell lines was systematically optimized while minimizing the effects on primary human endothelial cells. The computer-generated sequences featured improved cancer-cell penetration, induced cancer-cell apoptosis, and were enabled a decrease in the cytotoxic concentration of co-administered chemotherapeutic agents in vitro. This study demonstrates the potential of multidimensional machine-learning methods for rapidly obtaining peptides with the desired cellular activities.*

The computational design of peptides with desired biological activities is a (re)emerging research topic that will profoundly affect the future of chemical biology and drug discovery.[1] Designer peptides with selective cell-targeting and membrane-lytic properties have great potential for both drug delivery and lead structure development. Herein, we present the development and practical application of a technique that merges the concepts of machine-learning for property and activity prediction with multidimensional peptide optimization. The approach readily identified new sequences that exert selective cellular anticancer activity through both direct membrane disruption and the induction of apoptosis.

Anticancer peptides (ACPs) represent an exploratory class of anticancer agents. Although ACPs share certain structural features and molecular modes of action with antimicrobial peptides (AMPs), for example, an often overall positive charge and amphiphilic structure, not all known AMPs are also ACPs.[2] The oncolytic effects of ACPs may partly be explained by direct membrane-disruptive mechanisms without relying on surface-receptor interaction.[2b] Cell-penetrating peptides (CPPs) are structurally related to ACPs and are able to enter the cytosol without causing membranolysis at low concentrations.[3] CPPs have been widely investigated as cargo delivery vehicles.[4] However, not all peptides with antitumor or cell-penetrating activity are selective for cancer cells.[5] Systematic analysis revealed that their net positive charge, hydrophobicity, and hydrophobic moments are apparently insufficient predictors of selectivity, and that AMPs and ACPs cannot be differentiated from CPPs based on simple structural features alone.[6] Certain CPPs (e.g. TAT peptides) also show potent antimicrobial activity,[7] and certain AMPs, such as LL-37, have potential uses for drug delivery.[8] Tumor-homing peptides (THPs) have recently attracted much attention because of their strong affinity towards specific receptors or markers that are overexpressed in cancer cells.[9] They are rich in arginine and cysteine residues,[10] including the well-investigated RGD (Arg-Gly-Asp) and NGR (Asn-Gly-Arg) motifs.[11] THPs may be conjugated with ACPs, and the resulting chimeras have been shown to induce apoptosis in cultured cancer cells and to inhibit tumor growth in mice.[12]

In this study, we pursued a multidimensional peptide design approach based on support vector machine (SVM) classifiers.[13] We aimed to modify the sequence of an AMP template (Decoralin),[14] which we found to possess anticancer activity, by merging features of both CPPs and THPs into the template. We hypothesized that once modified, the designed 11-mers might penetrate cancer cells in concentrations that are lower than the critical concentration for lysing the membrane. Such cationic designer peptides could then exert their potential cytotoxicity by targeting the mitochondrion, an organelle with a net negatively charged membrane, and eventually induce apoptosis.[15] Instead of coupling an AMP to a CPP sequence, as carried out by Liu et al.,[16] which yields a high-mass peptide, we here propose a computational approach for generating minimalist peptides.

We pursued supervised SVM learning to obtain predictive models for membrane-active peptides. As a training set for model building, we collected annotated ACPs ($N = 51$), CPPs ($N = 481$), and THPs ($N = 669$) up to a length of 16 amino acids from publicly accessible databases and represented all peptides by correlated residue properties (PPCALI descriptor)[17] for computational analysis. This procedure led to a real-valued vector representation of each amino acid sequence, which takes potential residue cooperativity into account. Then we trained separate SVM classifiers for each group of peptides (Table 1). The SVM models turned out to be predictive, robust, and competitive.[18] The SVM_{ACP} model showed the highest variance of classification accuracy among the three classifiers. Despite its only moderate Matthews correlation coefficient ($mcc = 0.59 \pm 0.20$; *mcc* values are in $[-1,1]$ with zero indicating random performance), which is

[*] Dr. Y.-C. Lin, Y. F. Lim, E. Russo, Dr. P. Schneider, L. Bolliger, A. Edenharter, Prof. Dr. K.-H. Altmann, Prof. Dr. C. Halin, Dr. J. A. Hiss, Prof. Dr. G. Schneider
Department of Chemistry and Applied Biosciences
Swiss Federal Institute of Technology (ETH)
Vladimir-Prelog-Weg 4, 8093 Zurich (Switzerland)
E-mail: gisbert.schneider@pharma.ethz.ch

[**] We thank Sarah Haller for technical assistance. ETH Zürich and the Swiss National Science Foundation (grants 205321-134783, 206021-133768) supported this research financially.

The supporting information for this article includes experimental details and a further peptide design study with TAT_{48-60} as the template. It is available on the WWW under http://dx.doi.org/10.1002/anie.201504018.

Table 1: Performance of SVM models in 10-fold cross-validation, presented as the mean ± SD. *mcc* = Matthews correlation coefficient.

Classifier model	*mcc*	Specificity	Sensitivity
SVM_{ACP}	0.59 ± 0.20	0.94 ± 0.04	0.77 ± 0.16
SVM_{CPP}	0.76 ± 0.06	0.92 ± 0.02	0.81 ± 0.04
SVM_{THP}	0.72 ± 0.06	0.90 ± 0.01	0.82 ± 0.03

likely a result of the small number of ACP training samples, the high specificity (94 ± 4%) of the SVM_{ACP} model motivated us to use it for peptide design.

Prior to computational sequence design, we tested the cytotoxicity of the design template, Decoralin, on breast adenocarcinoma cells (MCF7) and primary human dermal microvascular endothelial cells (HDMEC). Decoralin was active against MCF7 cells (EC_{50} = 19 ± 1 μM) and displayed only mild HDMEC cytotoxicity (EC_{50} = 132 ± 1 μM; Figure 1A). We used glycine scanning to investigate which Decoralin residues were the most critical for the observed cytotoxicity against MCF7 cells. We opted for glycine instead of the more common alanine scanning because we aimed at probing the individual residue positions with a conformationally more flexible (impact on turn formation) and less hydrophobic (impact on lipid membrane interaction and amphiphilicity) probe. Glycine mutation of positions 1, 4, and 11 of the 11-residue template largely maintained peptide activity (Figure 1B), which suggests that the terminal residues are not critical for the observed cytotoxic effect. These results are consistent with the reported diverse roles of the C- and N-termini of membrane-active peptides.[19] Based on the outcome of the glycine scan, we computationally generated 20^3 = 8000 variations of Decoralin residues 1, 4, and 11 by allowing each of the 20 standard amino acids in these three positions. We chose this permutation strategy because we wanted to modulate peptide activity through the smallest possible sequence modification with a high chance of retaining membrane activity.

We computed SVM prediction scores for all of the peptides generated. Similar to our earlier work,[20] we calculated the scores as the distance of a peptide (represented by the PPCALI descriptor) to the separating hyperplane in SVM Hilbert space, and interpreted the score values as an indicator of the trustworthiness of the predictions but not as an actual quantitative index of peptide activity.[21] 146 of the 8000 modified sequences received favorable scores with both the SVM_{ACP} and SVM_{CPP} models (Figure 1C), and 196 sequences were classified as both ACP and THP (Figure 1D). The wild-type Decoralin sequence itself received a perfect SVM_{ACP} score. Even if we deleted Decoralin from the training data, Decoralin was consistently predicted as ACP but not as CPP or THP. Aiming at dual high scores for ACP + CPP and ACP + THP features, we selected 11 (AC1–11) and seven (AT1–7) Decoralin derivatives, respectively, for further analysis.

We synthesized and tested peptides AC1–11 for cytotoxicity against four different cancer cell lines, as well as HDMEC cells to represent normal non-transformed human cells (Table 2). Decoralin and all 11 derivatives displayed nonspecific activity against all of cancer cell lines and negligible toxicity against HDMECs ($EC_{50} \geqslant 100$ μM). Notably, AC1–5 displayed greater activity against the cancer cell lines than Decoralin, and AC2 showed a two-fold improvement.

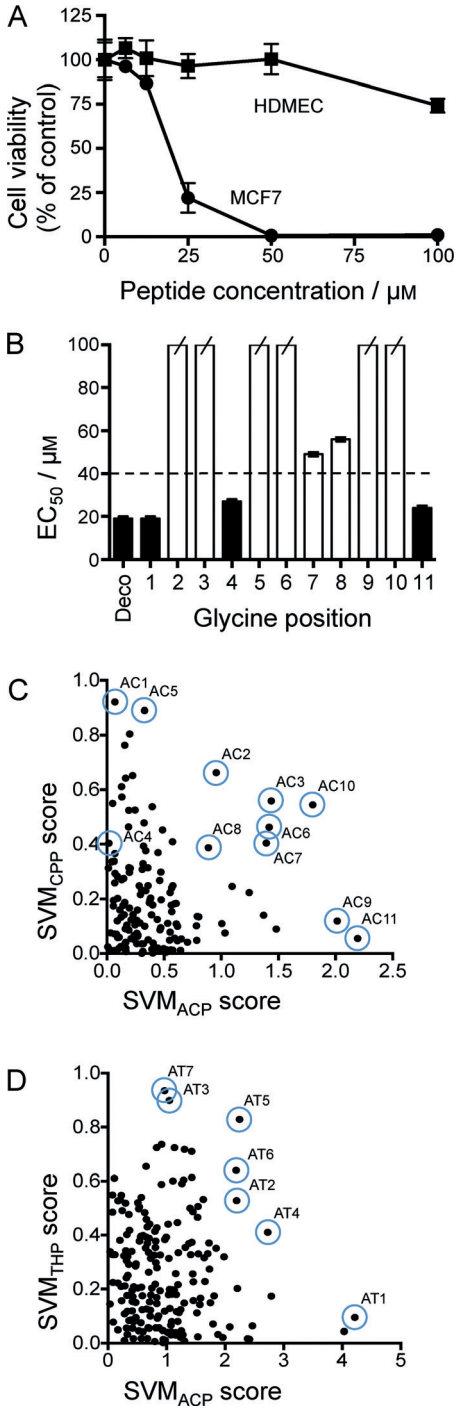

Figure 1. A) Cytotoxicity of Decoralin against MCF7 and HDMEC cells. Cells were treated for 24 h, followed by 1 h MTT treatment. B) MCF7 cytotoxicity of the glycine-walk variations of Decoralin (Deco). The data presented in (A) and (B) are the mean ± SD (N = 3 independent experiments, three technical replicates each). C, D) Prediction scores for the computer-generated sequences. Scores were computed as distance to the respective SVM hyperplane. The highlighted peptides were selected for synthesis and testing.

Table 2: Cytotoxicity of Decoralin and derivatives computationally predicted as ACP + CPP (AC1–11) or ACP + THP (AT1–7).[a]

Identifier	Sequence	MCF7	A549	LU1205	Jurkat	HDMEC
Decoralin	$SLLSLIRKLIT_{NH2}$	19 ± 1	28 ± 1	18 ± 1	40 ± 1	132 ± 1
AC1	$RLLRLIRKLII_{NH2}$	11 ± 1	21 ± 1	9 ± 1	39 ± 1	141 ± 1
AC2	$RLLRLIRKLIL_{NH2}$	9 ± 1	14 ± 1	7 ± 1	25 ± 1	125–250
AC3	$LLLLLIRKLIK_{NH2}$	11 ± 1	22 ± 1	16 ± 1	25 ± 1	125–250
AC4	$YLLYLIRKLIL_{NH2}$	13 ± 1	29 ± 1	14 ± 1	22 ± 1	256 ± 2
AC5	$RLLRLIRKLIR_{NH2}$	14 ± 1	27 ± 1	16 ± 1	27 ± 1	138 ± 7
AC6	$QLLQLIRKLII_{NH2}$	22 ± 1	29 ± 1	21 ± 1	35 ± 1	171 ± 1
AC7	$QLLQLIRKLIM_{NH2}$	33 ± 1	33 ± 1	24 ± 1	46 ± 1	142 ± 2
AC8	$MLLMLIRKLIL_{NH2}$	16 ± 1	58 ± 1	31 ± 1	31 ± 1	262 ± 1
AC9	$QLLQLIRKLIR_{NH2}$	53 ± 1	60 ± 1	38 ± 1	75 ± 1	260 ± 1
AC10	$QLLQLIRKLIK_{NH2}$	50-100	54 ± 1	39 ± 1	69 ± 1	≈250
AC11	$QLLQLIRKLIY_{NH2}$	50-100	61 ± 1	29 ± 1	58 ± 1	349 ± 1
AT1	$LLLQLIRKLIL_{NH2}$	10 ± 1	30 ± 1	12 ± 1	53 ± 1	154 ± 1
AT2	$TLLLLIRKLIL_{NH2}$	38 ± 1	79 ± 1	16 ± 1	30 ± 1	348 ± 1
AT3	$RLLLLIRKLIL_{NH2}$	24 ± 1	35 ± 1	18 ± 1	40 ± 1	160 ± 1
AT4	$LLLLLIRKLIL_{NH2}$	25 ± 1	> 100	48 ± 1	50 ± 1	278 ± 1
AT5	$QLLLLIRKLIL_{NH2}$	63 ± 1	> 100	44 ± 1	39 ± 1	271 ± 1
AT6	$QLLLLIRKLIV_{NH2}$	77 ± 1	> 100	53 ± 1	> 100	339 ± 1
AT7	$CLLLLIRKLIL_{NH2}$	> 100	> 100	76 ± 1	> 100	≈500

[a] The half maximal effective concentration (EC_{50}) values (μM) are given as the mean ± SD (N = 3 independent experiments with three technical replicates each).

In an attempt to investigate whether the improved anticancer activity of some of the designed peptides might at least in part be caused by intracellular toxicity, potentially benefitting from enhanced cell-penetration, we coupled fluorescein isothiocyanate (FITC) to the N-termini of Decoralin and AC1–5 and performed fluorescence microscopy and flow cytometry analysis with the peptide-treated cells. The microscopy images suggest that these peptides might indeed enter MCF7 cells when applied at low concentration (1 μM; Figure 2A). Flow cytometry analysis corroborated this preliminary observation and demonstrated that AC1, AC2, and AC5 show significantly ($p < 0.05$) greater MCF7 cell penetration than Decoralin, but no improved uptake in lung carcinoma (A549) cells (Figure 2B,C). AC2 resulted in significantly ($p < 0.0001$) more early apoptotic cells (5.2% vs. 1.2%, Annexin V-$FITC^+$/7-AAD^-) and late-apoptotic and/or necrotic cells (15% compared with 3.5%, Annexin V-$FITC^+$/7-AAD^+) than Decoralin, whereas both peptides produced an overall comparable level of damaged cells (19%, Annexin V-$FITC^-$/7-AAD^+; Figure 2D). The higher percentage of cells in the Annexin V-FITC-negative and 7-AAD-positive populations of the Decoralin-treated cells suggests that the cell membranes in most of the dead cells were fully destroyed. We thus did not detect staining of these samples, which is consistent with observations for other membrane-lytic peptides.[12b] After treatment with AC2, the cells were found in all phases, thus suggesting that AC2 kills MCF7 cells by both membrane-lytic and intracellular mechanisms.

Decoralin itself displayed pronounced cell-penetrating activity in MCF7 and A549 cells (Figure 2B,C), thus corroborating previous observations that ACPs could show mixed membrane-lytic and cell-penetrating activities.[22] Our machine-learning models were apparently capable of identifying key features of CPPs and allowed us to transfer these to the Decoralin derivatives. Moreover, the designed peptides retained anticancer activity. We successfully tuned the Decoralin sequence according to these objectives and confirmed the additional cell-penetrating ability in MCF7 cells. In AC1, AC2, and AC5, two serine residues of the Decoralin sequence are replaced by arginine, which is in line with previous work by Nakase et al., who showed that introducing an arginine residue into an ACP could improve membrane permeation.[23] However, their arginine-substituted peptides lost cytotoxicity against HeLa cells, thus suggesting a delicately balanced structure–activity relationship between ACPs and CPPs. Our peptide design strategy adequately addressed this issue through the SVM models.

Not all ACPs are sufficiently selective for therapeutic application.[5,24] We therefore wanted to optimize Decoralin derivatives for selectivity toward cancer cells. We synthesized seven peptides with predicted anticancer and tumor-homing properties (AT1–7, Figure 1C), all of which showed increased selectivity for cancer cells over HDMEC cells compared to Decoralin, and AT1–4 retained strong cytotoxicity against several cancer cell lines (Table 2). AT2 and AT4 showed selective toxicity against late-stage human melanoma (LU1205) and MCF7 cells but a negligible effect on HDMEC cells, which indicates the successful combination of ACP and THP features by our computational peptide design approach.

Poor tumor penetration of anticancer drugs can be a limiting factor for their efficacy, and co-administration with ACPs or other tumor-specific penetrating peptides has been suggested as a promising solution for clinical use.[25] A549 and MCF7 cells showed different sensitivities to doxorubicin ($EC_{50} = 450 \pm 1$ nM and 22 ± 1 nM, respectively) and epothilone C ($EC_{50} = 49 \pm 1$ and 14 ± 2 nM, respectively). We co-incubated A549 and MCF7 cells with mixtures of ACPs (Decoralin or AC2) and either doxorubicin or epothilone C. The incubation time was 72 h. While we did not observe significant synergism between the peptides and the chemotherapeutics in MCF7 cells, we observed strong and significant ($p < 0.0001$, extra sum-of-squares F test) synergistic effects in A549 cells with Decoralin or AC2 in combination with doxorubicin ($EC_{50} = 260 \pm 1$ and 127 ± 1 nM, respectively) or epothilone C ($EC_{50} = 31 \pm 1$ and 21 ± 1 nM, respectively; Figure S1 in the Supporting Information). The designed peptide AC2 outperformed Decoralin in terms of EC_{50} values, thus supporting the hypothesis that optimized ACPs can enhance the cytotoxic effect of chemotherapeutic agents.

Peptide design guided by machine-learning models enabled the identification of tailored peptide sequences with the desired potent and selective cellular anticancer activity. Our

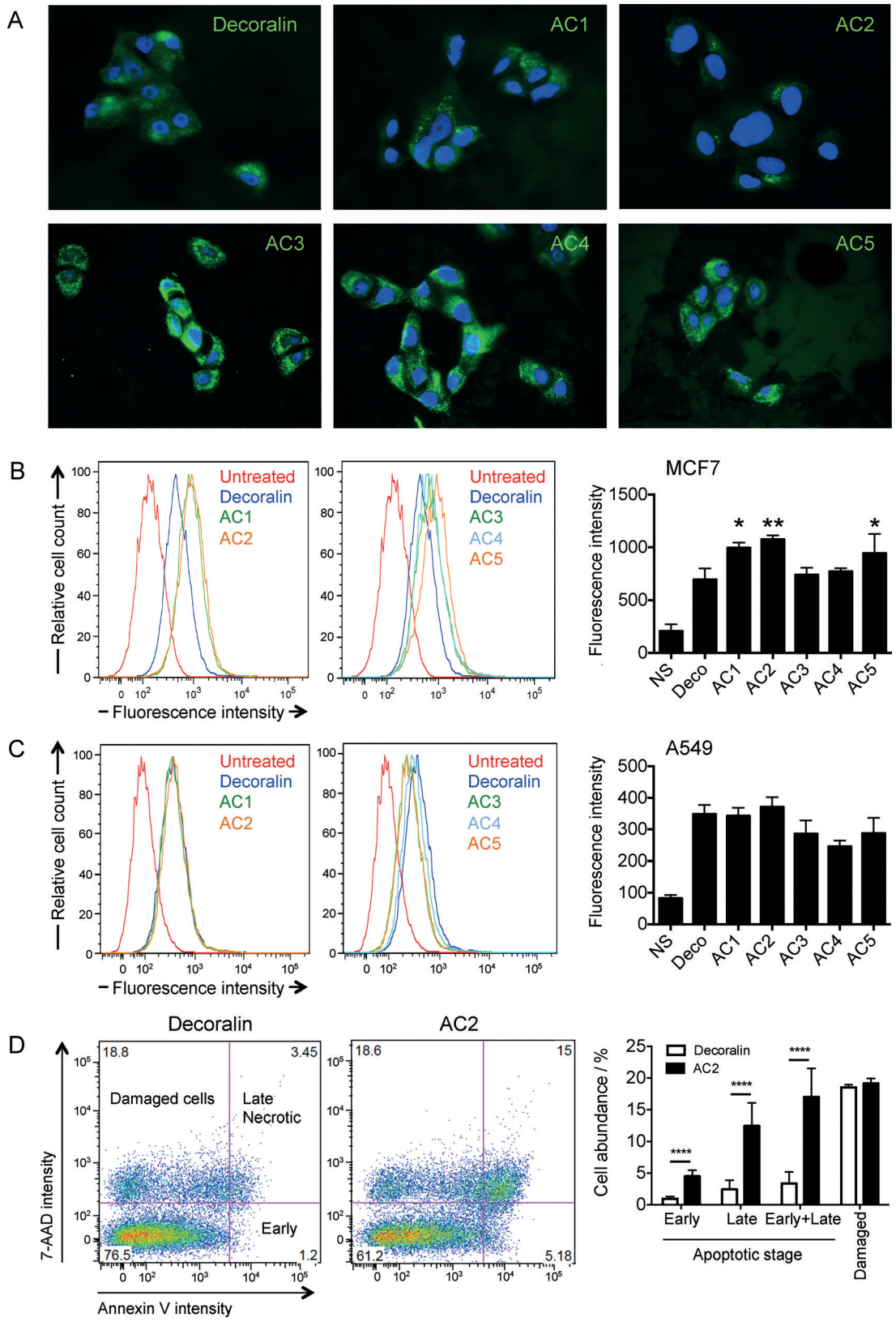

Figure 2. A) Fluorescence microscopy images of MCF7 cells treated with 1 μM of FITC-labeled Decoralin and peptides AC1–7 (green: peptides; blue: nuclei). Representative flow cytometry results are shown for MCF7 (B) and A549 (C) cells (fluorescence intensity given as mean ± SD, $N = 3$). The statistics compare Decoralin (Deco) with peptide treatments (NS: not stimulated, no peptide); * $p < 0.05$, ** $p < 0.01$. D) Flow cytometry experiments for cellular apoptosis. MCF7 cells were treated with 25 μM of Decoralin or AC2 for 20 min and stained with 7-AAD and Annexin V-FITC. The numbers in each quadrant give the percentage of cells. The histogram shows the percentages of cells in different apoptotic states as the mean ± SD of $N = 3$ independent experiments with three replicates each (**** $p < 0.0001$).

prediction models were accurate in extracting activity-related features from known ACPs and CPPs, which allowed us to transfer these features to new peptides. Starting from a template sequence, the approach facilitated the design of derivatives with significantly improved properties with minimal synthetic effort needed. This multidimensional computational approach essentially extends conventional peptide de novo design aiming at a single objective and offers an alternative to engineering high-molecular-weight peptides with potential solubility problems.

Keywords: cancer · molecular design · drug discovery · machine learning · lipid membranes

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 10370–10374
Angew. Chem. **2015**, *127*, 10512–10516

[1] a) N. Tsomaia, *Eur. J. Med. Chem.* **2015**, *94*, 459–470; b) C. D. Fjell, J. A. Hiss, R. E. Hancock, G. Schneider, *Nat. Rev. Drug Discov.* **2012**, *11*, 37–51; c) *De Novo Molecular Design* (Ed.: G. Schneider), Wiley-VCH, Weinheim, **2013**.
[2] a) J. S. Mader, D. W. Hoskin, *Expert Opin. Invest. Drugs* **2006**, *15*, 933–946; b) F. Harris, S. R. Dennison, J. Singh, D. A. Phoenix, *Med. Res. Rev.* **2013**, *33*, 190–234.
[3] F. Madani, S. Lindberg, U. Langel, S. Futaki, A. Graslund, *J. Biophys.* **2011**, *2011*, 414729.
[4] C. Bechara, S. Sagan, *FEBS Lett.* **2013**, *587*, 1693–16702.
[5] S. Riedl, D. Zweytick, K. Lohner, *Chem. Phys. Lipids* **2011**, *164*, 766–781.
[6] a) S. T. Henriques, M. N. Melo, M. A. Castanho, *Biochem. J.* **2006**, *399*, 1–7; b) E. Eiriksdottir, K. Konate, U. Langel, G. Divita, S. Deshayes, *Biochim. Biophys. Acta Biomembr.* **2010**, *1798*, 1119–1128.
[7] W. L. Zhu, S. Y. Shin, *J. Pept. Sci.* **2009**, *15*, 345–352.
[8] a) X. Zhang, K. Oglecka, S. Sandgren, M. Belting, E. K. Esbjorner, B. Norden, A. Graslund, *Biochim. Biophys. Acta Biomembr.* **2010**, *1798*, 2201–2208; b) K. Splith, I. Neundorf, *Eur. Biophys. J.* **2011**, *40*, 387–397.
[9] a) W. Arap, R. Pasqualini, E. Ruoslahti, *Science* **1998**, *279*, 377–380; b) N. Svensen, J. G. Walton, M. Bradley, *Trends Pharmacol. Sci.* **2012**, *33*, 186–192.
[10] P. Kapoor, H. Singh, A. Gautam, K. Chaudhary, R. Kumar, G. P. Raghava, *PLoS One* **2012**, *7*, e35187.
[11] a) R. Pasqualini, E. Koivunen, E. Ruoslahti, *Nat. Biotechnol.* **1997**, *15*, 542–546; b) R. Pasqualini, E. Koivunen, R. Kain, J. Lahdenranta, M. Sakamoto, A. Stryhn, R. A. Ashmun, L. H. Shapiro, W. Arap, E. Ruoslahti, *Cancer Res.* **2000**, *60*, 722–727.
[12] a) Y. Chen, X. Xu, S. Hong, J. Chen, N. Liu, C. B. Underhill, K. Creswell, L. Zhang, *Cancer Res.* **2001**, *61*, 2434–2438; b) M. Sioud, A. Mobergslien, *Biochem. Pharmacol.* **2012**, *84*, 1123–1132.
[13] a) C. Cortes, V. Vapnik, *Machine Learning* **1995**, *20*, 273–297; b) E. Byvatov, U. Fechner, J. Sadowski, G. Schneider, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1882–1889.
[14] K. Konno, M. Rangel, J. S. Oliveira, M. P. Dos Santos Cabrera, R. Fontana, I. Y. Hirata, I. Hide, Y. Nakata, K. Mori, M. Kawano, H. Fuchino, S. Sekita, J. R. Neto, *Peptides* **2007**, *28*, 2320–2327.
[15] See Ref. [2a].
[16] S. Liu, H. Yang, L. Wan, J. Cheng, X. Lu, *Cancer Biother. Radiopharm.* **2013**, *28*, 289–297.
[17] C. P. Koch, A. M. Perna, S. Weissmuller, S. Bauer, M. Pillong, R. B. Baleeiro, M. Reutlinger, G. Folkers, P. Walden, P. Wrede, J. A. Hiss, Z. Waibler, G. Schneider, *ACS Chem. Biol.* **2013**, *8*, 1876–1881.
[18] A. Gautam, K. Chaudhary, R. Kumar, A. Sharma, P. Kapoor, A. Tyagi, G. P. Raghava, Open source drug discovery consortium, *J. Transl. Med.* **2013**, *11*, 74.
[19] a) D. W. Hoskin, A. Ramamoorthy, *Biochim. Biophys. Acta Biomembr.* **2008**, *1778*, 357–375; b) Y.-C. Lin, J. A. Hiss, P. Schneider, P. Thelesklaf, Y. F. Lim, M. Pillong, F. M. Koehler, P. S. Dittrich, C. Halin, S. Wessler, G. Schneider, *ChemBioChem* **2014**, *15*, 2225–2231; c) C. S. Lee, W. C. Tung, Y. H. Lin, *J. Pept. Sci.* **2014**, *20*, 438–445.
[20] E. Byvatov, B. C. Sasse, H. Stark, G. Schneider, *ChemBioChem* **2005**, *6*, 997–999.
[21] J. Platt, *Advances in Large Margin Classifiers*, MIT Press, Cambridge, **1999**, pp. 61–74.
[22] a) H. Xu, C. X. Chen, J. Hu, P. Zhou, P. Zeng, C. H. Cao, J. R. Lu, *Biomaterials* **2013**, *34*, 2731–2737; b) K. Ohara, M. Kohno, T. Hamada, K. Kawakami, *Peptides* **2013**, *50*, 28–35.
[23] I. Nakase, S. Okumura, S. Katayama, H. Hirose, S. Pujals, H. Yamaguchi, S. Arakawa, S. Shimizu, S. Futaki, *Chem. Commun.* **2012**, *48*, 11097–11099.
[24] E. Kondo, K. Saito, Y. Tashiro, K. Kamide, S. Uno, T. Furuya, M. Mashita, K. Nakajima, T. Tsumuraya, N. Kobayashi, M. Nishibori, M. Tanimoto, M. Matsushita, *Nat. Commun.* **2012**, *3*, 951.
[25] a) N. Papo, Y. Shai, *Cell. Mol. Life Sci.* **2005**, *62*, 784–790; b) K. N. Sugahara, T. Teesalu, P. P. Karmali, V. R. Kotamraju, L. Agemy, D. R. Greenwald, E. Ruoslahti, *Science* **2010**, *328*, 1031–1035.

Received: May 4, 2015
Published online: June 26, 2015

