Identification, Characterization and Synthesis of Natural Parasitic Cysteine Protease Inhibitors – More Potent Falcitidin Analogs

Stephan Brinkmann[1λ], Sandra Semmler[1λ], Christian Kersten[2], Maria A. Patras[1], Michael Kurz[3], Natalie Fuchs[2], Stefan J. Hammerschmidt[2], Jennifer Legac[4], Peter E. Hammann[5], Andreas Vilcinskas[1,6], Philip. J. Rosenthal[4], Tanja Schirmeister[2], Armin Bauer[3,], Till F. Schäberle[1,6*]*

1 Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany

2 Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg University Mainz, 55128 Mainz, Germany

3 Sanofi-Aventis Deutschland GmbH, R&D, 65926 Frankfurt am Main, Germany

4 Department of Medicine, University of California, San Francisco, 94143 California, United States

5 Evotec International GmbH, 37079 Göttingen, Germany

6 Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, 35392 Giessen, Germany

λ Authors contributed equally

* Correspondence: Armin.Bauer@sanofi.com; Till.Schaeberle@ime.fraunhofer.de

KEYWORDS, Bacteroidetes, *Chitinophaga*, falcipain, rhodesain, α-chymotrypsin, malaria, pentapeptide, natural products, protease inhibitors, parasitic diseases, antimalarial agents, sleeping sickness, *Plasmodium falciparum*, *Trypanosoma*

ABSTRACT

Protease inhibitors represent a promising therapeutic option for the treatment of parasitic diseases such as malaria and human African trypanosomiasis. Falcitidin was the first member of a new class of inhibitors of falcipain-2, a cysteine protease of the malaria parasite *Plasmodium falciparum*. Using a metabolomics dataset of 25 *Chitinophaga* strains for molecular networking enabled identification of over 30 natural analogs of falcitidin. Based on MS/MS spectra, they vary in their amino acid chain length, sequence, acyl residue, and *C*-terminal functionalization; therefore, they were grouped into the four falcitidin peptide families A-D. The isolation, characterization and absolute structure elucidation of two falcitidin-related pentapeptide aldehyde analogs by extensive MS/MS spectrometry and NMR spectroscopy in combination with advanced Marfey's analysis was in agreement with the *in silico* analysis of the corresponding biosynthetic gene cluster. Total synthesis of chosen pentapeptide analogs followed by *in vitro* testing against a panel of proteases revealed selective parasitic cysteine protease inhibition and additionally low-micromolar inhibition of α-chymotrypsin. The pentapeptides investigated here showed superior inhibitory activity compared to falcitidin.

INTRODUCTION

Parasitic diseases such as malaria and sleeping sickness (human African trypanosomiasis, HAT) are poverty-associated diseases that have a massive impact on human life, especially in tropical areas.[1,2] Protozoan parasites of the *Trypanosoma* genus, transmitted by tsetse flies, are responsible for HAT.[2] With fewer than 3000 cases in 2015, the number of reported cases of HAT has reached a historically low level in recent years.[2] On the contrary, 229 million malaria cases and 409,000 deaths were estimated in 2019.[3] *Plasmodium falciparum* is the most virulent malaria parasite,

causing nearly all deaths[3] as well as most drug-resistant infections.[4,5] Resistance to most available antimalarials, including artemisinin-based combination regimens that are the standard treatment for falciparum malaria, is of great concern. Hence, there is a great need to discover new active compounds to serve as lead structures for the development of new antimalarial drugs.[4,5]

Papain-family cysteine proteases represent interesting therapeutic targets against several infectious diseases, including malaria (falcipains) and HAT (rhodesain).[6] Synthetic or natural product-based small molecule inhibitors were investigated and proved falcipains and rhodesain to be promising targets.[7,8] However, developing compounds that target these cysteine proteases into drugs has proved challenging, in part due to the difficulty of achieving selectivity.[9] The acyltetrapeptide falcitidin and its synthetic analogs were described as first members of a new class of cysteine protease inhibitors. In a falcipain-2 assay the natural compound showed an IC_{50} value in the low micromolar range and the optimized synthetic analogs a sub-micromolar IC_{50} activity against *Plasmodium falciparum* in a standard blood-cell assay. Instead of reactive groups that covalent-reversibly or irreversibly bind to active-site cysteines, these compounds share a *C*-terminal amidated proline.[10,11]

High structural diversity in combination with diverse biological activities make natural products a valuable source in the search for new lead candidates for drug development. Technological improvements in resolution and accuracy of spectrometry methods followed by complex and large dataset analysis allowed the application of metabolomics techniques based on e.g. ultra-high-performance liquid chromatography in line with high-resolution tandem mass spectrometry (UHPLC-HRMS/MS) for the discovery and characterization of these metabolites present in a given sample.[12] Data visualization and interpretation using tandem mass spectrometry (MS/MS)

networks (molecular networking) allows automatic annotation of MS/MS spectra against library compounds as well as the identification of signals of interest and their analogs.[13]

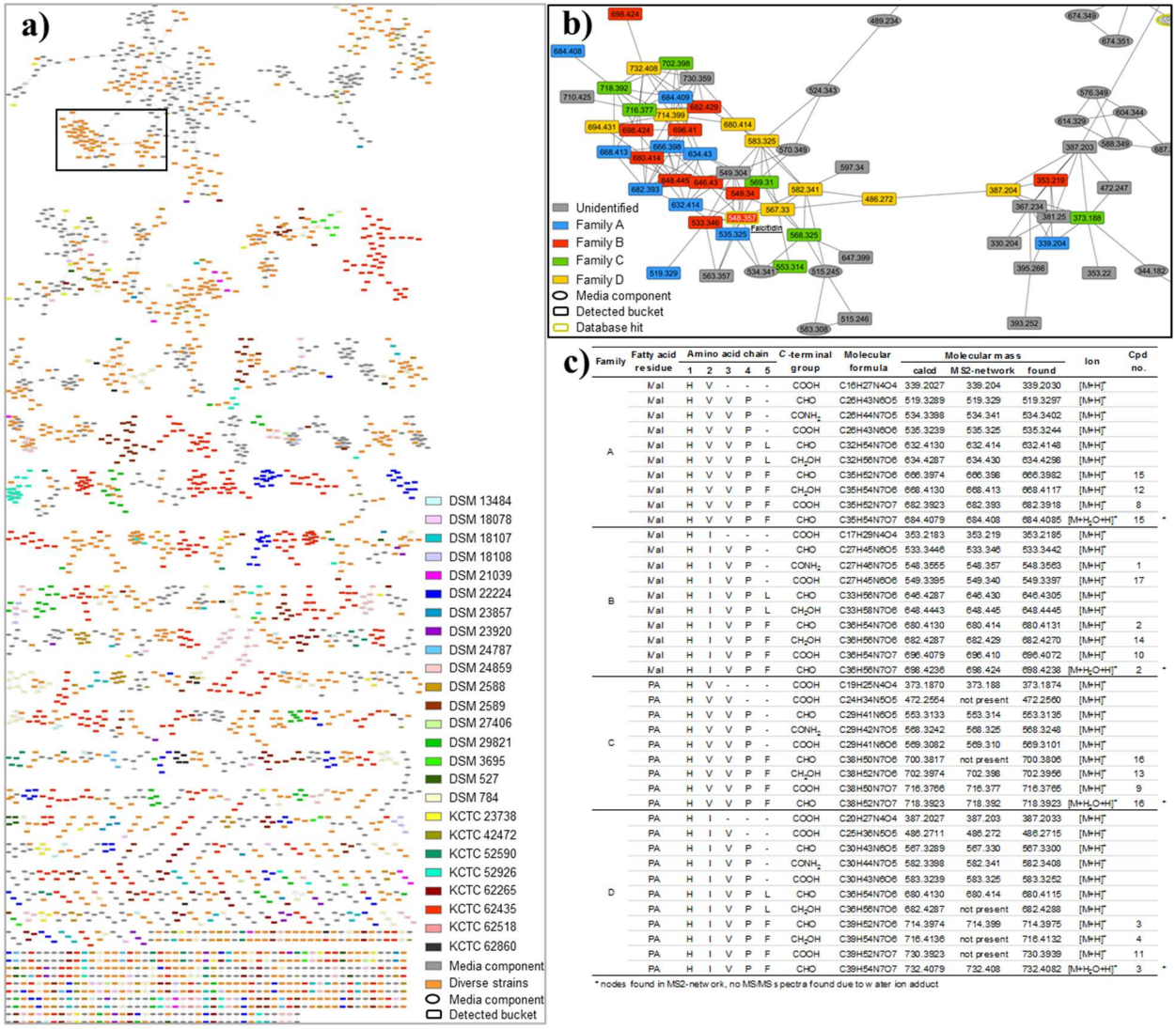
In this study, molecular networking was applied to analyze a previously generated dataset based on bacterial organic extracts originated from 25 *Chitinophaga* strains (phylum Bacteroidetes) in which falcitidin was annotated.[14] We aimed to investigate if additional natural analogs were produced by these strains. In total, more than 30 natural analogs were discovered, biosynthesized by *C. eiseniae* DSM 22224, *C. dinghuensis* DSM 29821, and *C. varians* KCTC 52926. They were classified into four peptide families (A-D) of falcitidin-like natural products based on variations in their amino acid chain length, sequence, acyl residue, and *C*-terminal functionalization. The analysis of MS/MS spectra revealed each family to contain truncated di- or tripeptides and larger pentapeptides, including molecules with classical *C*-terminal aldehyde moieties which are supposed to covalent-reversibly react with the active-site cysteine and serine residues of proteases.[9] A gene encoding a multimodular non-ribosomal peptide synthetase, responsible for the production of the pentapeptide aldehydes, was identified *in silico* in selected *Chitinophaga* genomes. After isolation and structure elucidation of two natural aldehyde analogs, total synthesis of all four pentapeptide aldehydes and their carboxylic acid and alcohol derivatives was successfully achieved using a solid-phase peptide synthesis (SPPS) split approach, followed by functional *C*-terminal group interconversion. This enabled activity testing against a panel of proteases: the cysteine proteases cathepsin B and L, falcipain-2 and 3, and rhodesain, as well as the serine proteases α-chymotrypsin, matriptase-2, and the transpeptidase of *Staphylococcus aureus* sortase A. Besides a low-micromolar IC_{50} activity against α-chymotrypsin, all four tested pentapeptide aldehydes had improved inhibitory activity, compared to the reference falcitidin, against falcipain-2, and three of them also showed activity against rhodesain. In summary, the

pentapeptide aldehydes display specific inhibition of the parasitic cysteine proteases falcipain-2 and -3 compared to no inhibition of human cathepsin B and L.

RESULTS AND DISCUSSION

Discovery of falcitidin analogs. Exploration of the genomic and metabolomic potential of the Bacteroidetes genus *Chitinophaga* revealed a high potential to find chemical novelty.[14] In the framework of a previous metabolomics' study of *Chitinophaga*, an inhibitor of the antimalarial target falcipain-2, falcitidin (**1**),[10] was detected.[14] Originally wrongly described as myxobacterium-derived tetrapeptides produced by *Chitinophaga* sp. Y23, falcitidin and its synthetic analogs are first members of a new class of cysteine protease inhibitors without a reactive group that covalent-reversibly or irreversibly binds to the active site cysteine residue.[10,11] Here, molecular networking was carried out to analyze the chemical profiles of all 25 previously cultured *Chitinophaga* strains (Figure 1a). This facilitated the discovery of more than 30 natural analogs of falcitidin (**1**, *m/z* 548.3563 $[M+H]^+$, molecular formula $C_{27}H_{46}N_7O_5$) (Figure 1b) biosynthesized by *C. eiseniae* DSM 22224, *C. dinghuensis* DSM 29821, and *C. varians* KCTC 52926. Analysis of the corresponding MS/MS spectra led to the assignment of four peptide families that differed in the amino acid chain at position 2 and the acyl residue of the peptide. Members of families A and B shared a fragment ion of *m/z* 222.1237 $[M+H]^+$ corresponding to the molecular formula $C_{11}H_{16}N_3O_2^+$. The fragment can be assigned to the described isovaleroyl (iVal) residue attached to the first amino acid histidine.[10] In contrast, a fragment ion of *m/z* 256.1084 $[M+H]^+$, corresponding to the molecular formula $C_{14}H_{14}N_3O_2^+$, was detected in peptides of families C and D (Figure S1-4). A difference within the acyl moiety was assumed due to the identical number of three nitrogen atoms, also indicating a histidine at position one. Moreover, the MS/MS fragmentation pattern of

family members A and B as well as C and D varied in position 2 of the amino acid chain, either carrying valine or isoleucine. Within each peptide family, members differed in chain length from two to five amino acids and their *C*-terminal functional group (Figure 1c). Interestingly, a corresponding falcitidin analog with the unusual *C*-terminal amidated ($CONH_2$-group) proline, indicated by a neutral loss of 114.0794 Da, was found in each family. Falcitidin (iVal-H-I-V-P-$CONH_2$) itself belongs to family B. Its analogs with the same neutral loss were assigned to the UHR-ESI-MS ion peaks at *m/z* 534.3402 $[M+H]^+$ ($C_{26}H_{44}N_7O_5$, family A), *m/z* 568.3248 $[M+H]^+$ ($C_{29}H_{42}N_7O_5$, family C), and *m/z* 582.3408 $[M+H]^+$ ($C_{30}H_{44}N_7O_5$, family D). Additionally, within all families tetrapeptide-analogs carrying a *C*-terminal unmodified (COOH-group) proline and analogs with an aldehyde moiety (CHO-group) indicated by a neutral loss of 99.0686 Da (C_5H_9NO) were found. Besides these tetrapeptide analogs of falcitidin, truncated di- or tripeptides with *C*-terminal COOH-groups and larger pentapeptide analogs were predicted based on MS/MS spectra. The spectra of the pentapeptides, which were found in all families, revealed either phenylalanine or leucine/isoleucine in the *C*-terminal position. Besides an unmodified *C*-terminal phenylalanine (COOH-group), additional analogs carrying a *C*-terminal phenylalaninal (CHO-group) and phenylalaninol (CH_2OH-group) were predicted. The last two analogs were proposed based on the neutral losses of 149.0841 Da ($C_9H_{11}NO$) and 151.1006 Da ($C_9H_{13}NO$), respectively. For analogs with a leucine/isoleucine at position five only analogs with a CHO- and CH_2OH-group were detected, with neutral losses of 115.1031 Da ($C_6H_{13}NO$) and 117.1164 Da ($C_6H_{15}NO$), respectively (Figure 1c, Figure S1-4).



Family	Fatty acid residue	Amino acid chain 1	2	3	4	5	C-terminal group	Molecular formula	Molecular mass calcd	MS2-network	found	Ion	Cpd no.
A	Val	H	V	-	-	-	COOH	C16H27N4O4	339.2027	339.204	339.2030	[M+H]+	
	Val	H	V	V	P	-	CHO	C26H43N6O5	519.3289	519.329	519.3297	[M+H]+	
	Val	H	V	V	P	-	$CONH_2$	C26H44N7O5	534.3398	534.341	534.3402	[M+H]+	
	Val	H	V	V	P	-	COOH	C26H43N6O6	535.3239	535.325	535.3244	[M+H]+	
	Val	H	V	V	P	L	CHO	C32H54N7O6	632.4130	632.414	632.4148	[M+H]+	
	Val	H	V	V	P	L	CH_2OH	C32H56N7O6	634.4287	634.430	634.4298	[M+H]+	
	Val	H	V	V	P	F	CHO	C35H52N7O6	666.3974	666.398	666.3982	[M+H]+	15
	Val	H	V	V	P	F	CH_2OH	C35H54N7O6	668.4130	668.413	668.4117	[M+H]+	12
	Val	H	V	V	P	F	COOH	C35H52N7O7	682.3923	682.393	682.3918	[M+H]+	8
	Val	H	V	V	P	F	CHO	C35H54N7O7	684.4079	684.408	684.4085	$[M+H_2O+H]^+$	15 *
B	Val	H	I	-	-	-	COOH	C17H29N4O4	353.2183	353.219	353.2185	[M+H]+	
	Val	H	I	V	P	-	CHO	C27H45N6O5	533.3446	533.346	533.3442	[M+H]+	
	Val	H	I	V	P	-	$CONH_2$	C27H46N7O5	548.3555	548.357	548.3563	[M+H]+	1
	Val	H	I	V	P	-	COOH	C27H45N6O6	549.3395	549.340	549.3397	[M+H]+	17
	Val	H	I	V	P	L	CHO	C33H56N7O6	646.4287	646.430	646.4305	[M+H]+	
	Val	H	I	V	P	L	CH_2OH	C33H58N7O6	648.4443	648.445	648.4445	[M+H]+	
	Val	H	I	V	P	F	CHO	C36H54N7O6	680.4130	680.414	680.4131	[M+H]+	2
	Val	H	I	V	P	F	CH_2OH	C36H56N7O6	682.4287	682.429	682.4270	[M+H]+	14
	Val	H	I	V	P	F	COOH	C36H54N7O7	696.4079	696.410	696.4072	[M+H]+	10
	Val	H	I	V	P	F	CHO	C36H56N7O7	698.4236	698.424	698.4238	$[M+H_2O+H]^+$	2 *
C	PA	H	V	-	-	-	COOH	C19H25N4O4	373.1870	373.188	373.1874	[M+H]+	
	PA	H	V	V	-	-	COOH	C24H34N5O5	472.2554	not present	472.2560	[M+H]+	
	PA	H	V	V	P	-	CHO	C29H41N6O5	553.3133	553.314	553.3135	[M+H]+	
	PA	H	V	V	P	-	$CONH_2$	C29H42N7O5	568.3242	568.325	568.3248	[M+H]+	
	PA	H	V	V	P	-	COOH	C29H41N6O6	569.3082	569.310	569.3101	[M+H]+	
	PA	H	V	V	P	F	CHO	C38H50N7O6	700.3817	not present	700.3806	[M+H]+	16
	PA	H	V	V	P	F	CH_2OH	C38H52N7O6	702.3974	702.398	702.3956	[M+H]+	13
	PA	H	V	V	P	F	COOH	C38H50N7O7	716.3766	716.377	716.3765	[M+H]+	9
	PA	H	V	V	P	F	CHO	C38H52N7O7	718.3923	718.392	718.3923	$[M+H_2O+H]^+$	16 *
D	PA	H	I	-	-	-	COOH	C20H27N4O4	387.2027	387.203	387.2033	[M+H]+	
	PA	H	I	V	-	-	COOH	C25H36N5O5	486.2711	486.272	486.2715	[M+H]+	
	PA	H	I	V	P	-	CHO	C30H43N6O5	567.3289	567.330	567.3300	[M+H]+	
	PA	H	I	V	P	-	$CONH_2$	C30H44N7O5	582.3398	582.341	582.3408	[M+H]+	
	PA	H	I	V	P	-	COOH	C30H43N6O6	583.3239	583.325	583.3252	[M+H]+	
	PA	H	I	V	P	L	CHO	C36H54N7O6	680.4130	680.414	680.4115	[M+H]+	
	PA	H	I	V	P	L	CH_2OH	C36H56N7O6	682.4287	not present	682.4288	[M+H]+	
	PA	H	I	V	P	F	CHO	C39H52N7O6	714.3974	714.399	714.3975	[M+H]+	3
	PA	H	I	V	P	F	CH_2OH	C39H54N7O6	716.4136	not present	716.4132	[M+H]+	4
	PA	H	I	V	P	F	COOH	C39H52N7O7	730.3923	not present	730.3939	[M+H]+	11
	PA	H	I	V	P	F	CHO	C39H54N7O7	732.4079	732.408	732.4082	$[M+H_2O+H]^+$	3 *

* nodes found in MS2-network, no MS/MS spectra found due to water ion adduct

Figure 1. MS^2-networking analysis revealed more than 30 natural falcitidin analogs. **A**: Complete MS^2-network based on extracts of 25 *Chitinophaga* strains. **B**: Falcitidin network color-coded according to the four molecule families A-D. **C**: Overview of molecular features of all identified natural falcitidin analogs.

Natural compound isolation and spectroscopic analysis. In order to confirm the proposed structures of the pentapeptide falcitidin analogs, their functional *C*-terminal groups and the unknown acyl chain of peptide families C and D, isolation and spectroscopic analysis of natural pentapeptides with *m/z* 680.4131 $[M+H]^+$ (**2**, $C_{36}H_{54}N_7O_6$), *m/z* 714.3975 $[M+H]^+$ (**3**,

$C_{39}H_{52}N_7O_6$), and *m/z* 716.4132 $[M+H]^+$ (**4**, $C_{39}H_{54}N_7O_6$) was achieved. Isolation was performed by adsorption of the metabolites on XAD16N resin and sequential C18-RP-HPLC followed by C18-RP-UHPLC fractionations starting from a 20 L fermentation of *C. eiseniae* DSM 22224. Based on the analysis of several two-dimensional NMR spectra, including DQF-COSY, TOCSY, ROESY, multiplicity edited-HSQC, and HMBC experiments, the structure of the isolated compound **2** was determined to be the phenylalaninal extended version of falcitidin (iVal-H-I-V-P-F-CHO), determined from the additional aromatic signals δ_H = 7.27 and 7.23 ppm, as well as the aldehyde protons at δ_H = 9.47 and 9.41 ppm and δ_C = 200.3 ppm (Table 1, Figure S19). The presence of two signals for several protons and carbon atoms indicated an epimerization of the stereogenic center of the *C*-terminal phenylalaninal. The analysis of the NMR data of the isolated compound **3** revealed the same peptide sequence. However, instead of the isovaleroyl moiety, compound **3** contains a phenyl acetyl (PA) residue at the *N*-terminus, leading to the final sequence PA-H-I-V-P-F-CHO (Table 2, Figure S21). Again, the protons of the terminal aldehyde moiety gave rise to a set of two signals (e.g. δ_H of aldehyde proton = 9.46 and 9.41 ppm), indicating the epimerization of its stereogenic center. The NMR sample of compound **4** contained a mixture of **4** and aldehyde **3**, not allowing for a complete assignment of the NMR signals. The phenylalaninol moiety, as indicated by MS/MS results, was confirmed by the presence of a second methylene group adjacent to C_α (δ_H = 3.27 ppm, δ_{13C} = 61.8 ppm), which in the COSY spectrum is correlated with an exchangeable proton at 4.7 ppm (OH). The reduction of the aldehyde function also leads to a significant high-field shift of the H_α and the amide proton (Figure S15), verifying the structure of compound **4** to be PA-H-I-V-P-F-CH_2OH (Figure 2).

Interestingly, UHPLC control measurements of the pure pentapeptide samples after NMR spectroscopy revealed the presence of mixtures of the pentapeptide and its corresponding

tetrapeptide analog with *C*-terminal amidated proline. The NMR samples were dissolved in DMSO, afterwards dried *in vacuo*, resolved in MeOH and measured. For example within the control measurement of compound **3**, the falcitidin analog of family D (*m/z* 582.3408 $[M+H]^+$, $C_{30}H_{44}N_7O_5$) was detected (Figure S5). Chemically, the decomposition of the pentapeptide to the unexpected amide version of the proline could be based on two hypothetical mechanisms: i) base-catalysis by the imidazole moiety of the histidine[15] or ii) an oxidative decomposition *via* an *N*,*O*-acetal-like intermediate.[16]

Table 1. NMR data comparison for **2** in DMSO-d_6 for natural isolated one (700 MHz, 176 MHz) and synthetic one (600 MHz, 100 MHz). Due to the presence of two diastereomers (see above) two sets of signals are obtained. Therefore, for some signals two chemical shift values are given.[a]

Position	**δ ^{13}C [ppm]**		**δ ^{1}H [ppm]; (m, ∫, *J*)**	
	natural isolated	synthetic	natural isolated	synthetic
Phenylalaninal				
CHO	200.3	200.3	9.47, 9.41	9.47 (s, 0.2H) 9.41 (s, 0.2H)
NH	-	-	8.38	8.36 (d, 0.2H, *J* = 7.7 Hz) 8.34 (d, 0.2H, *J* = 7.0 Hz)
α-CH	59.5	59.7 / 59.5	4.30	4.35-4.24 (m, 2H, overlay) 4.24-4.16 (m, 2H, overlay)
β-CH$_2$	33.5	33.6 / 33.3	3.13, 2.74	3.15-3.07 (m, 1H) 2.91-2.83 (m, 2H, overlay) 2.78-2.69 (m, 2H, overlay)
γ-C$_{quart}$	137.6	137.6	-	-
CH$_{arom}$	129.3, 128.1, 126.3 126.2	129.3, 129.2, 129.1, 128.9, 128.2, 128.1, 126.24, 126.19	7.23, 7.27, n.a.	7.29-7.12 (m, 5H)
Proline				
CO	~172.0	172.1 / 172.0	-	-
α-CH	~59.1	59.1 / 59.0	4.29	4.35-4.24 (m, 2H, overlay)
β-CH$_2$	29.3	29.31 / 29.28	1.92, 1.54	2.02-1.83 (m, 5H, overlay) 1.83-1.70 (m, 2H, overlay)

γ-CH_2	~24.4	24.44 / 24.38	1.79, 1.74	1.60-1.51 (m, 1H) 2.02-1.83 (m, 5H, overlay) 1.83-1.70 (m, 2H, overlay)
δ-CH_2N	~47.0	47.00 / 46.96	3.75, 3.53	3.78-3.68 (m, 1H) 3.57-3.47 (m, 1H)
Valine				
CO	169.6	169.58 / 169.55	-	-
NH	-	-	8.05	8.00-7.94 (m, 2H, overlay)
α-CH	55.8	55.7	4.25	4.35-4.24 (m, 2H, overlay)
β-CH	29.7	29.7	1.98	2.02-1.83 (m, 5H, overlay)
γ-CH_3	19.0 18.5	19.03 / 19.00 18.45 / 18.41	0.88 0.86	0.90-0.84 (m, 6H, overlay)
Isoleucine				
CO	170.7	170.7	-	-
NH	-	-	7.82	7.69-7.63 (m, 1H)
α-CH	~56.3	56.32 / 56.29	4.25	4.24-4.16 (m, 2H)
β-CH	~36.9	36.9 / 36.8	1.65	1.70-1.60 (m, 2H)
γ-CH_2	24.0	24.0	1.25 0.93	1.30-1.22 (m, 1H) 0.98-0.90 (m, 1H)
γ-CH_3	15.2	15.2	0.71	0.76-0.67 (m, 6H, overlay)
δ-CH_3	11.0	11.0	0.74	0.76-0.67 (m, 6H, overlay)
Histidine				
CO	170.0	170.9	-	-
NH	-	-	8.16	8.00-7.94 (m, 2H, overlay)
α-CH	~51.7	52.7	4.71	4.61-4.55 (m, 1H)
β-CH_2	~27.9	29.9*	3.03 2.85	2.91-2.83 (m, 2H, overlay) 2.78-2.69 (m, 2H, overlay)
γ-C_{quart}	n.a.	134.2	-	-
δ-CH_{arom}	~116.7	n.o.	~7.2	6.76 (s, 1H)
ε-CH_{arom}	133.8	134.5	~8.7	7.50 (s, 1H)
ε-NH	-	-	-	n.o.
Isovaleroyl				
CO	171.6	171.4	-	-
CH_2	44.3	44.5	1.96	2.02-1.83 (m, 5H, overlay)
CH	25.4	25.5	1.89	2.02-1.83 (m, 5H, overlay)
CH_3	22.2	22.2	0.80	0.81 (d, 3H, J = 6.4 Hz)
CH_3	22.1	22.1	0.77	0.78 (d, 3H, J = 6.4 Hz)

[a] n.o. - not observed; n.a. - not assigned due to line broadening; * - data obtained from HMBC or HSQC spectrum

Table 2. NMR data comparison for **3** in DMSO-d_6 for natural isolated one (500 MHz, 126 MHz) and synthetic one (600 MHz, 100 MHz).[a]

Position	**δ ^{13}C [ppm]**		**δ ^{1}H [ppm]; (m, ∫, *J*)**	
	natural isolated (major)	synthetic	natural isolated (major)	synthetic
Phenylalaninal				
CHO	200.3, 200.5	200.3	9.46, 9.41	9.47 (s, 0.2H) 9.41 (s, 0.2H)
NH	-	-	8.38, 8.36	8.36 (d, 0.25H, *J* = 7.7 Hz) 8.34 (d, 0.25H, *J* = 7.0 Hz)
α-CH	59.5, 59,8	59.1 / 59.0	4.29, 4.17	4.24-4.16 (m, 2H, overlay)
β-CH$_2$	33.5, 33,3	33.5 / 33.3	3.12/2.74, 3.09/2.85	3.15-3.06 (m, 1H, overlay) 2.93-2.83 (m, 2H, overlay) 2.78-2.71 (m, 1H, overlay)
γ-C$_{quart}$	137.7	137.6	-	-
CH$_{arom}$	129.3, 129.2 128.1, 128.2 126.2, 126.3	129.3, 129.2, 129.14, 129.09, 129.0, 128.2, 128.1, 128.0, 126.25, 126.20, 126.16	7.21, 7.24, 7.18	7.29-7.13 (m, 10H, overlay)
Proline				
CO	172.0	172.1 / 172.0	-	-
α-CH	59.0, 59.1	59.5	4.29, 4.30	4.33-4.24 (m, 2H, overlay)
β-CH$_2$	29.3, 29.4	29.32 / 29.29	1.91/1.54, 1.97/1.74	2.02-1.84 (m, 2H, overlay) 1.84-1.69 (m, 2H, overlay) 1.59-1.51 (m, 1H)
γ-CH$_2$	24.4, 24.5	24.5 / 24.4	1.78/1.74, 1.91/1.80	2.02-1.84 (m, 2H, overlay) 1.84-1.69 (m, 2H, overlay)
δ-CH$_2$N	46.98, 47.02	47.01 / 46.97	3.75/3.50, 3.75/3.54	3.78-3.68 (m, 1H) 3.57-3.47 (m, 1H)
Valine				
CO	169.58	169.61 / 169.58	-	-
NH	-	-	7.98, 7.96	7.95 (t, 1H, *J* = 9.0 Hz)
α-CH	55.8	55.8	4.26	4.33-4.24 (m, 2H, overlay)
β-CH	29.7	29.7	1.97	2.02-1.84 (m, 2H, overlay)
γ-CH$_3$	19.0 18.5	19.03 / 19.00 18.5 / 18.4	0.87, 0.85	0.94-0.82 (m, 7H, overlay)
Isoleucine				
CO	170.75	170.7	-	-
NH	-	-	7.76	7.77-7.71 (m, 1H)
α-CH	56.39	56.40 / 56.37	4.21	4.24-4.16 (m, 2H, overlay)
β-CH	36.79	36.8 / 36.7	1.64	1.69-1.59 (m, 1H)
γ-CH$_2$	23.9	23.9	1.22, 0.88	1.27-1.17 (m, 1H)

γ-CH_3	15.3	15.3	0.69	0.94-0.82 (m, 7H, overlay) 0.75-0.66 (m, 6H, overlay)
δ-CH_3	11.0	11.0	0.72	0.75-0.66 (m, 6H, overlay)
Histidine				
CO	170.8	170.8	-	-
NH	-	-	8.29	8.26 (d, 1H, *J* = 8.1 Hz)
α-CH	52.9	52.9	4.58	4.62-4.55 (m, 1H)
β-CH_2	n.a.	30.1*	2.89, 2.75	2.93-2.83 (m, 2H, overlay) 2.78-2.71 (m, 1H, overlay)
γ-C_{quart}	n.a.	n.o.	-	-
δ-CH_{arom}	n.a.	n.o.	6.75	6.76 (s, 1H)
ε-CH_{arom}	134.5	134.5	7.49	7.51 (s, 1H)
ε-NH	-	-	-	n.o.
Phenylacetyl				
CO	169.9	169.9	-	-
CH_2 a	42.0	42.0	3.43	3.45 (d, 1H, *J* = 14.1 Hz) 3.42 (d, 1H, *J* = 14.1 Hz)
C_{quart}	136.2	136.2	-	-
CH_{arom}	128.97, 128.05, 126.18	129.3, 129.2, 129.14, 129.09, 129.0, 128.2, 128.1, 128.0, 126.25, 126.20, 126.16	7.17, 7.22, 7.19	7.29-7.13 (m, 10H, overlay)

[a] n.o. - not observed; n.a. - not assigned due to line broadening; * - data obtained from HMBC or HSQC spectrum

Finally, advanced Marfey's analysis was conducted to determine the absolute configuration of the amino acids.[17] Total hydrolysis of a sample containing **3** and **4**, followed by chemical derivatization with N_α-(2,4-dinitro-5-fluorophenyl)-L-valinamide (Marfey's reagent) and UHPLC-MS comparison to reference substrates confirmed the literature known D-His-L-Val/Ile-L-Val-L-Pro configuration[10] and the incorporation of L-phenylalanine at position five by verifying the presence of L-phenylalaninol (Figure S5). All stereogenic centers were further confirmed by total synthesis and comparison of NMR data to natural isolated compounds (Table 1 and 2).

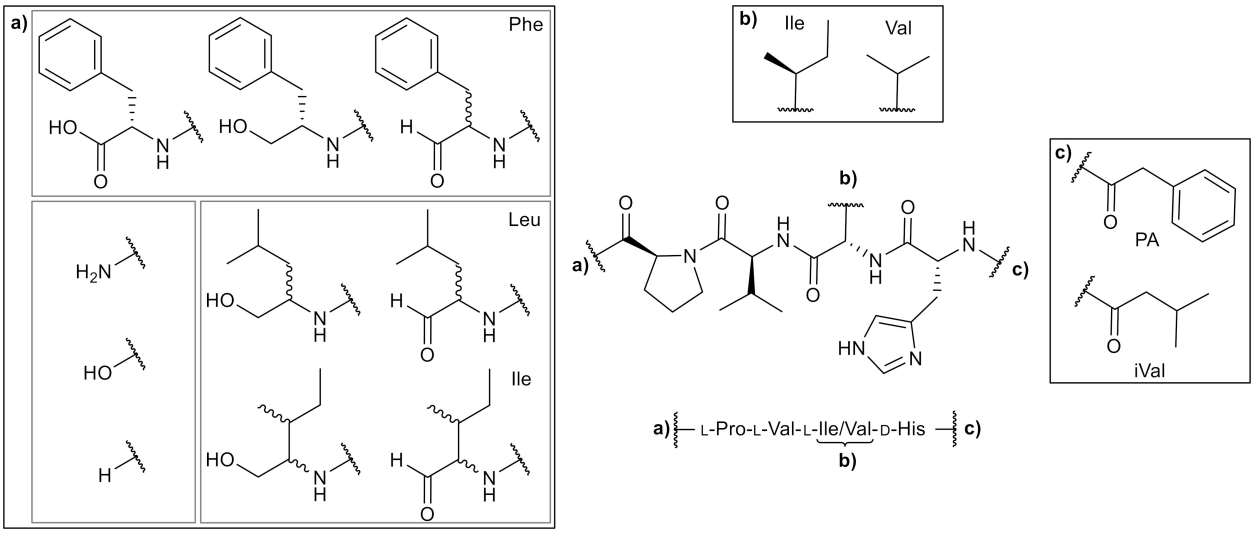


Figure 2. Overview of chemical structures of natural falcitidin analogs.

Identification of the biosynthetic gene cluster. The publicly available genomes of all three producers, *C. eiseniae* DSM 22224 (FUWZ01000000), *C. dinghuensis* DSM 29821 (QLMA01000000), and *C. varians* KCTC 52926 (JACVFB010000000), were scanned with antiSMASH[18] for NRPS-type biosynthetic gene clusters (BGCs) matching the structural features of the molecules. The number and predicted substrate specificity of the A-domains, as well as precursor supply and post-assembly modifications were taken into account. BGCs congruent to the pentapeptide structure were identified in each case. Furthermore, an epimerization domain to catalyze the conversion of L- to D-amino acids is positioned in agreement with the determined stereochemistry of the molecules. Interestingly, manual search of other publicly available *Chitinophaga* genomes revealed the presence of similar BGCs in genomes of *C. varians* Ae27 (JABAIA010000000) and *C. niastensis* DSM 24859 (PYAW00000000), the latter also being included in the molecular networking analysis. However, no production was observed for *C. niastensis*. MAFFT alignment[19] of all five BGCs allowed clear cluster boarder prediction and revealed variations in the upstream region. In conclusion, falcitidin analog biosynthesis is encoded

by a single NRPS core gene 18.5 kbp in length. No additional genes encoding proteins involved in further modifications or transportation are conserved between all five BGCs (Figure 3). The NRPS starts with a starter condensation (C-starter) domain responsible for the addition of iVal or PA to the peptide core. A terminal reductase domain (TD) at the *C*-terminal end in four BGCs should be responsible for the reductive release process.[20] The presence of alcohol and aldehyde pentapeptides suggests a similar reduction as e.g. shown in the biosynthesis of the siderophore myxochelin.[21–23] The (peptidyl)acyl thioester attached to the carrier protein could be reduced first to an aldehyde and then to an alcohol via a four-electron reduction during the product release.[24,25] The BGC in the genome of *C. varians* Ae27 carries a NAD_binding_4 domain instead of the TD. This domain family is sequence-related to the *C*-terminal region of the male sterility protein in arabidopsis species[26] and a jojoba acyl-CoA reductase.[27] The latter is known to catalyze a similar reduction reaction, with the formation of a fatty alcohol from a fatty acyl substrate.[27]

A BGC congruent to the pentapeptide structure and the presence of the unusual tetrapeptide analog after NMR study of its corresponding pentapeptide aldehyde points toward falcitidin and its identified analogs from families A, C and D with their unusual *C*-terminal amidated proline to be degradations products. Biochemically a cleavage by a carboxypeptidase in the presence of ammonia could result in the *C*-terminal amide group of the proline, which can explain the tetrapeptides in the extract. Chemically, the unusual *C*-terminal amide group instead of the expected carboxylic acid could be based either on a base-catalyzed[16] or an oxidative decomposition.[15] These hypotheses and their mechanisms will need to be evaluated in further studies.

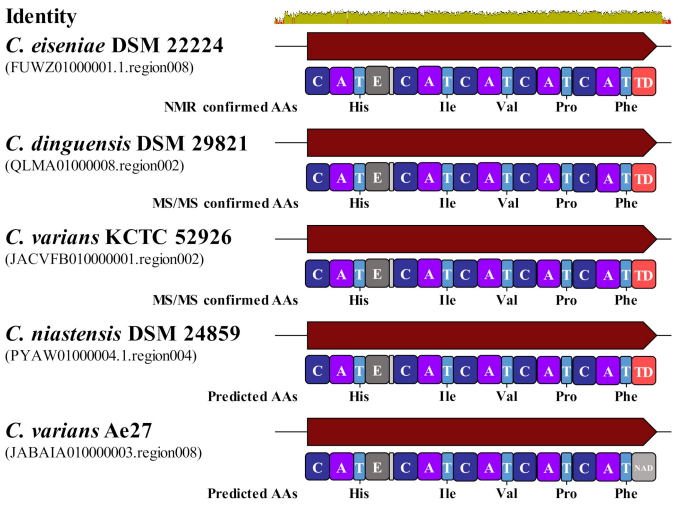
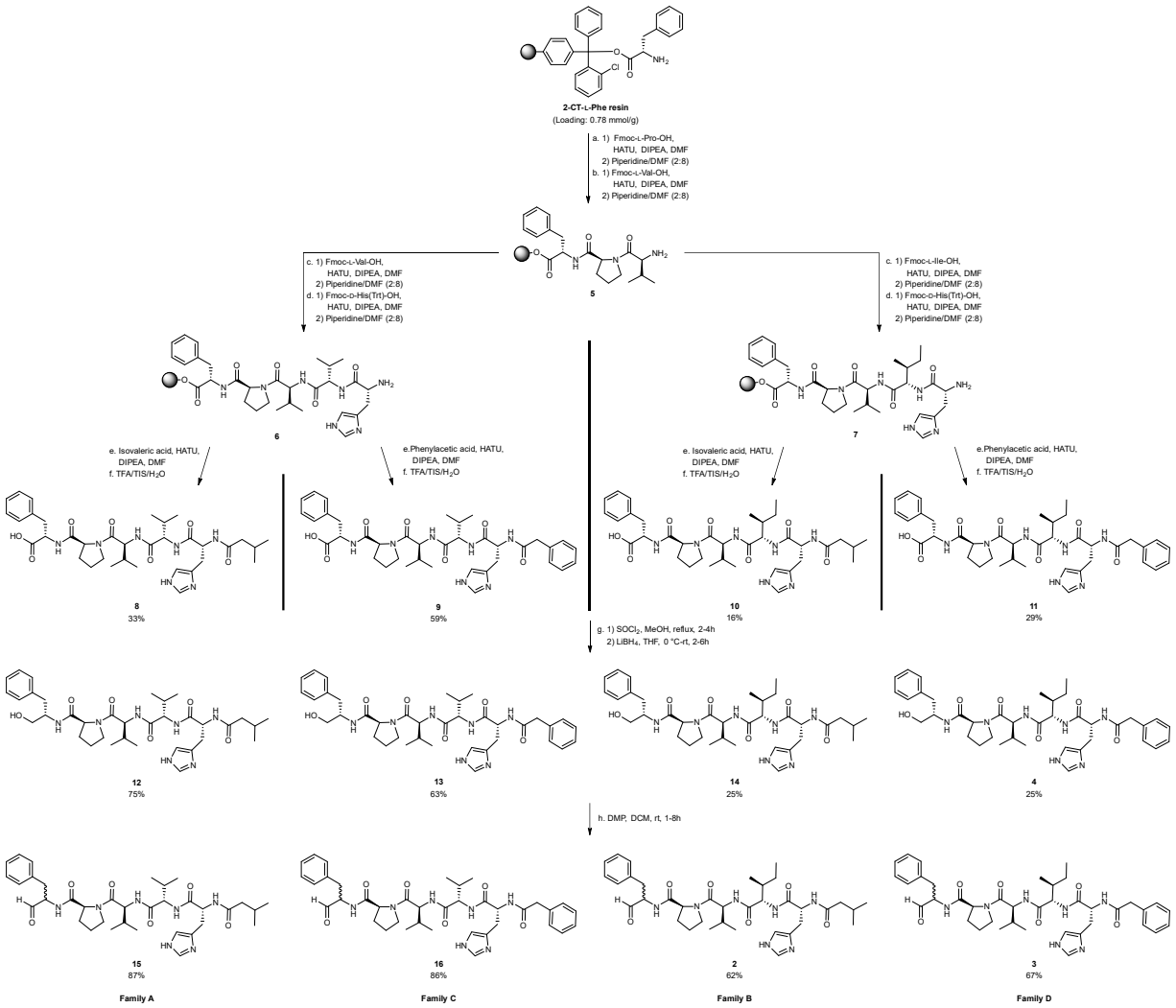


Figure 3. Biosynthetic gene clusters for falcitidin-like pentapeptides found in various *Chitinophaga* strains. Nucleotide alignment of all five BGCs using MAFFT alignment[19] allowed clear cluster boarder prediction and revealed variations in the upstream region. AA, amino acids; C, Condensation domain; A, Adenylation domain; T, Peptidyl-carrier protein domain; E, Epimerization domain; TD, Terminal reductase domain; NAD, NAD_binding_4 domain.

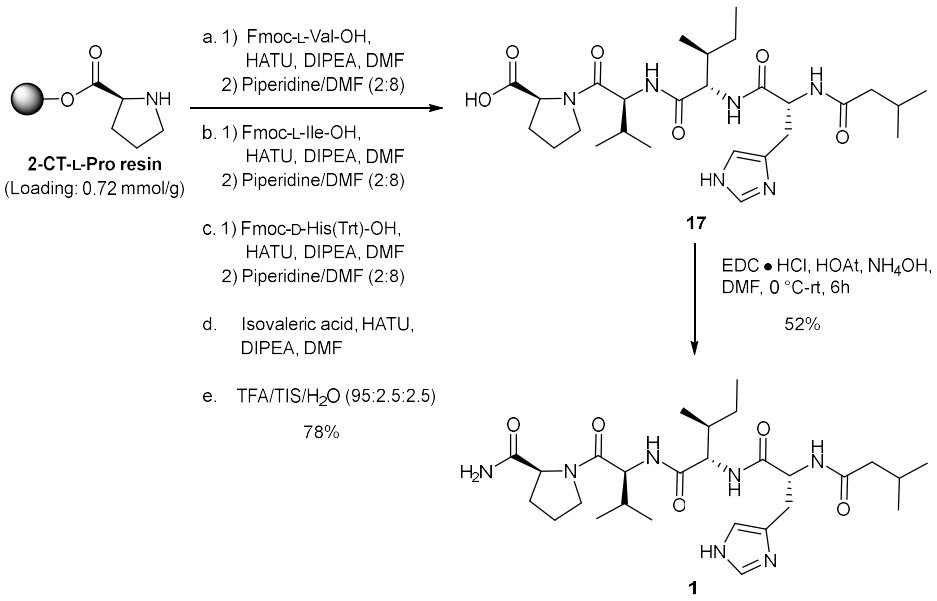
Total synthesis of pentapeptide falcitidin analogs. The scarcity of natural material and difficulties in isolation due to co-elution of aldehydes and alcohol analogs made it necessary to synthesize the most promising pentapeptide phenylalanine-aldehydes for biological testing. The application of a split approach solid-phase peptide synthesis (SPPS)[28] followed by functional group interconversion gave access to a variety of falcitidin pentapeptide analogs of family A, B, C and D (Scheme 1).



Scheme 1. SPPS split approach (a-f) and functional group interconversions (g-h) to chosen falcitidin pentapeptide analogs of family A, B, C and D. Standard SPPS conditions apply and reactions were carried out at room temperature if not noted otherwise. Reduction of the acid to the alcohol was achieved *via* the corresponding methyl ester. The aldehyde was obtained by Dess-Martin oxidation; the stereogenic center of phenylalanine could not be retained. Detailed conditions can be found in the Supporting Information.

Splitting the synthesis after the third amino acid and for the attachment of the two different fatty acids allows the synthesis of all four chosen acid analogs (**2**, **3**, **15**, and **16**) based on 2-chlorotrytil-L-Phe resin (2-CT-L-Phe). The functionalization of the peptide acid yields the corresponding methyl ester, the alcohol and the aldehyde. Reduction to the alcohol was done *via* the methyl ester, after direct reduction of the acid could not be achieved. The methyl ester was gained by conversion with thionyl chloride in methanol, followed by the reduction with $LiBH_4$ in THF. For the oxidation to the aldehyde Dess-Martin periodinane (DMP) was chosen as a very mild reagent. We found, that small traces of methanol as a stabilizer in DCM as the solvent made the reaction take longer, up to eight hours instead of one, and was responsible for incomplete conversion. Epimerization of the stereogenic center of the aldehydes could not be avoided, nor was it necessary to be avoided based on the data for the natural isolated compounds. They were obtained as diasteromers for the phenylalanine moiety of roughly equal proportions based on NMR data (Figure S17, S18, S20, S22). Overall yields of the SPPS for the four different chains varied between 16-59%. The reduction yielded the corresponding alcohols over two steps with 25-75% yield and oxidation to the aldehydes achieved 62-87% yield (Scheme 1).

For reference purposes, falcitidin (**1**) was synthesized using the same SPPS approach instead of the literature known liquid phase method.[11] This afforded acid analog **17** with a yield of 78% and, after amidation, falcitidin (**1**) with a 41% over all yield (Scheme 2). Additionally, the acid analog **17** and falcitidin (**1**) themselves present great functional groups for further derivatization and the introduction of different war heads, like a nitrile or azide group, to increase the potency as a potential inhibitor.[29,30]



Scheme 2. Synthesis of falcitidin (**1**) *via* acid analog **17**.

Bioactivity. Falcitidin (**1**) was previously reported to display an IC_{50} of 6 µM against falcipain-2.[10] Therefore, the falcitidin pentapeptide aldehyde analogs **2**, **3**, **15** and **16** were tested in a similar *in vitro* assay against falcipain-2 together with **1** as control. All four aldehydes were active with an IC_{50} of 41.5 µM (**15**) to 23.7 µM (**2**). However, the originally described activity of **1** could not be reproduced, with an IC_{50} >50 µM. This indicated differences in assay sensitivity, which made it difficult to put the aldehyde activities in line with the literature. Aldehydes **3** and **15** displayed activities also against the related *P. falciparum* cysteine protease falcipain-3 (66% sequence identity with falcipain-2), with an IC_{50} of 45.4 µM and 42.5 µM, respectively. Inhibition of both falcipains is of importance, since falcipain-3 is able to compensate for knockout of falcipain-2.[31] A counter screen against human cysteine proteases cathepsin B and L and sortase A of *Staphylococcus aureus* as a surrogate for unrelated cysteine proteases revealed no activity for pentapeptide aldehydes **2**, **3**, **15** and **16** as well as the *C*-terminal acid (**11**) and alcohol (**4**) of **3**, thus demonstrating selectivity over these off-targets. Activities with an IC_{50} of 57.7 µM (**2**) to

17.1 µM (**15**) were also observed for the falcipain-homologue cysteine protease of *Tryoanosoma brucei rhodesiense* rhodesain. Most interestingly, aldehydes **3**, **15** and **16** displayed higher activities, with IC_{50}s of 3.7 µM (**15**) and 1.5 µM (**16**), against α-chymotrypsin, which was tested as a prototype of human serine proteases, while the serine membrane protease matriptase-2 (TMPRSS6) was not inhibited (Table 3). To further elucidate these observations, molecular docking studies were performed.

Table 3. Overview of protease activity data. n.d. = not determined; - = not active

No.	Peptide sequence							IC_{50} [µM]							
								Rhod	Falcipain-2	Flacipain-3	CatB	CatL	α-CT	SrtA	TMPRSS6
1	iVal	H	I	V	P	-	$CONH_2$	-	>50	>50	-	-	-	-	-
17	iVal	H	I	V	P	-	COOH	-	n.d.	n.d.	-	-	-	-	-
11	PA	H	I	V	P	F	COOH	-	n.d.	n.d.	-	-	-	-	-
4	PA	H	I	V	P	F	CH_2OH	-	n.d.	n.d.	-	-	-	-	-
2	iVal	H	I	V	P	F	CHO	57.7	33.8	45.4	-	-	-	-	-
15	iVal	H	V	V	P	F	CHO	17.1	42.5	-	-	-	2.2	-	-
3	PA	H	I	V	P	F	CHO	25.4	23.7	42.5	-	-	3.7	-	-
16	PA	H	V	V	P	F	CHO	-	38.2	-	-	-	1.5	-	-

Docking studies. Due to the high number of rotatable bonds and the associated degrees of conformational freedom in the molecules under elucidation, docking of the full-length peptides is challenging.[32–34] Hence, truncated tripeptides with an *N*-terminal acetyl (ace)-cap to avoid a non-present charge of full-length inhibitors (**2**, **3**, **4**, **11**, **15** and **16**) were used for docking studies. First, a conventional non-covalent docking was performed. However, aldehydes are known electrophilic warheads and able to form covalent-reversible hemithioacetal adducts with catalytic cysteine residues.[35] Therefore, the predicted poses with special attention to the distance between the nucleophilic sulfur of the catalytic cysteine residue and the electrophilic carbon of the aldehyde

moiety were elucidated, and an additional covalent docking was performed. The tripeptides mimic the orientation of the protease substrates by addressing S3-S1 (Figure 4a and b). Based on these results (Table S15), the low inhibitory potency of the *C*-terminal alcohol (**4**) and acid (**11**) moieties as well as that for falcitidin cannot simply be explained by non-covalent interactions, as the docking scores are within a similar range or even higher when compared to the aldehydes. However, based on predicted binding poses the covalent reaction between aldehydes and catalytic cysteine residues, which might contribute to higher affinity, seems likely for molecules with proteinogenic L-Phe as P1 residue, with a distance of 3.0 Å between the electrophilic carbon atom and the nucleophilic sulfur of Cys for both falcipain-2 and falcipain-3, and the aldehyde oxygen coordinated by hydrogen bonds in the oxyanion hole, but rather unlikely for D-Phe (distance of 7.3 and 5.0 Å for falcipain-2 and falcipain-3, respectively). Additionally, only small differences between covalent and non-covalent binding modes were observed, indicating that no larger conformational changes need to take place during or after reaction. While testing against α-chymotrypsin was conducted to demonstrate selectivity over human off-target serine proteases, the high affinity is reasonable, as inhibitors with aromatic moieties deeply buried in the S1 pocket of the protease were reported previously, as well as cleavage preference after Phe.[36–38] Aldehydes are able to form covalent reversible hemiacetal adducts with serine[39]; moreover, an increased potency is indicated by proximity of the aldehyde to the catalytic Ser-195 residue in the predicted binding modes for both L- and D-Phe at P1 (Figure 4c, Table S15).

Falcitidin (**1**) was found to display poor whole-cell activity against chloroquine-sensitive *P. falciparum* strain 3D7 (IC_{50} >10 μM), while a structure-activity-relationship (SAR) study identified a synthetic trifluoromethyl analog displaying sub-micromolar IC_{50} activity.[11] Together with our results, this allows further development of the SAR, with the aim to increase both potency

and selectivity with an improved peptidic recognition sequence. However, no general selectivity-issue against serine proteases can be expected, as matriptase-2 was not inhibited by the compounds. Additionally, the docking studies revealed a rather unlikely distance for covalent binding of D-Phe-analogs. With all compounds having been tested as mixtures of D- and L-Phe due to the aldehyde's natural epimerization of its stereogenic center,[40,41] further analogs with e.g. other warheads preventing epimerization such as a nitrile group[29,30] should ideally be synthesized and tested in L-configurations only.

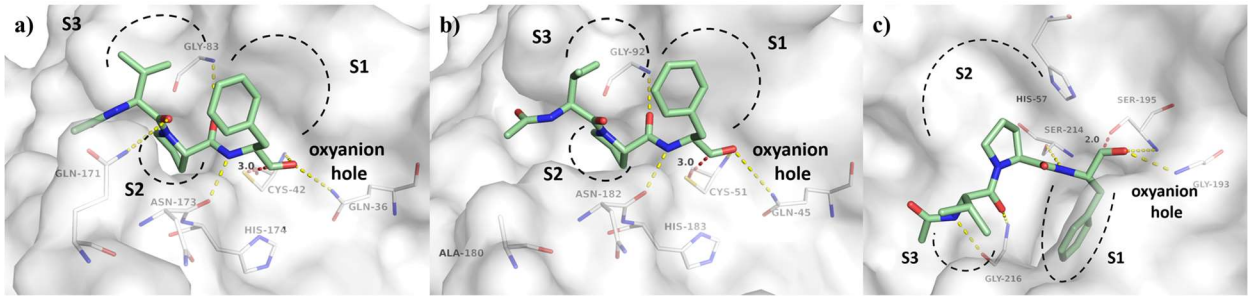


Figure 4. Non-covalent docking predicted binding modes of truncated peptidic protease inhibitor ace-Val-Pro-Phe-aldehyde (green carbon atoms) in complex with falcipain-2 (**A**, PDB-ID 3BPF), falcipain-3 (**B**, PDB-ID 3BPM) and chymotrypsin (**C**, PDB-ID 1AFQ). Proteases are depicted as white transparent surface, for clear view only residues forming polar interactions (yellow dashed lines) and catalytic Cys/Ser and His residues are labeled and depicted as lines. Substrate binding sites S1-S3 are schematically indicated. The distance between nucleophilic sulfur (Cys-42 in falcipain-2 and Cys-51 in falcipain-3) or oxygen (Ser-195 in chymotrypsin) of the catalytic center to the electrophilic carbon atom of the aldehyde is depicted as a dashed red line and labeled with its distance in Å.

CONCLUSION

In this study, metabolic networking analysis of *Chitinophaga* strains led to the discovery of over 30 *N*-acyl oligopeptides, structurally related to falcitidin, an inhibitor of the antimalarial cysteine protease falcipain-2.[10] Isolation and structure elucidation of two novel natural pentapeptide aldehydes validated the MS/MS fragmentation pattern analysis. A BGC congruent to the pentapeptide structure was identified and indicated falcitidin and its tetrapeptide analogs carrying the *C*-terminally amidated proline to be degradation products from the described pentapeptide aldehydes. Total synthesis gave access to the most promising aldehyde analogs and allowed their biological profiling. A selective *in vitro* activity against parasitic cysteine proteases rhodesain, falcipain-2 and falcipain-3, together with a low-micromolar IC_{50} inhibition of the serine protease α-chymotrypsin was observed. This forms the basis for future studies to develop optimized derivatives with increased potency and selectivity against targeted proteases.

EXPERIMENTAL SECTION

General Experimental Procedures. For all UHPLC-QTOF-UHR-MS and MS/MS measurements a quadrupole time-of-flight spectrometer (LC-QTOF maXis II, Bruker Daltonics, Bremen, Germany) equipped with an electrospray ionization source in line with an Agilent 1290 infinity II LC system (Agilent Technologies, CA, Unites States) was used. C18 RP-UHPLC [ACQUITY UPLC BEH C18 column (130 Å, 1.7 µm, 2.1 x 100 mm)] was performed at 45 °C with the following linear gradient (A: H_2O, 0.1% HCOOH; B: CH_3CN, 0.1% HCOOH; flow rate: 0.6 mL/min): 0 min: 95% A; 0.30 min: 95% A; 18.00 min: 4.75% A; 18.10 min: 0% A; 22.50 min: 0% A; 22.60 min: 95% A; 25.00 min: 95% A. A 50 to 2000 *m/z* scan range at 1 Hz scan rate was used to acquire mass spectral data. The injection volume was set to 5 µL. MS/MS experiments

were performed at 6 Hz and the top five most intense ions in each full MS spectrum were targeted for fragmentation by higher-energy collisional dissociation at 25 eV using N_2 at 10^{-2} mbar. Precursors were excluded after 2 spectra, released after 0.5 min and reconsidered if the intensity of an excluded precursor increased by factor 1.5 or more. Data were analyzed using the Bruker DataAnalysis 4.0 software package. Specific rotation was determined on a digital polarimeter (P3000, A. Krüss Optronic GmbH, Germany). Standard wavelength was the sodium D-line with 589 nm. Temperature, concentration (g/100 mL) and solvent are reported with the determined value.

NMR Spectroscopy. NMR spectra of natural isolated falcitidin analogs were acquired on a Bruker AVANCE 700 spectrometer (700 MHz for ^{1}H, 176 MHz for ^{13}C) and a Bruker AVANCE 500 spectrometer (500 MHz for ^{1}H, 126 MHz for ^{13}C). Both instruments were equipped with a 5 mm TCI cryo probe. For structure elucidation and assignment of proton and carbon resonances 1D-^{1}H, 1D-^{13}C, DQF-COSY, TOCSY (mixing time 80 ms), ROESY (mixing time 150 ms), multiplicity edited-HSQC, and HMBC spectra were acquired. NMR spectra of synthesized molecules were recorded on an AVANCE III HD 600 spectrometer (600 MHz for ^{1}H, 151 MHz for ^{13}C) from Bruker Biospin (Bruker Biospin GmbH, Rheinstetten, Germany). ^{1}H- and ^{13}C-chemical shifts are reported in ppm and were referenced to the corresponding residual solvent signal (DMSO-d_6: δ_C = 39.52 ppm, δ_H = 2.50 ppm). δ_C shifts marked with an * (asterisk) were not observed in the ^{13}C NMR spectrum, but were obtained either from HMBC or HSCQ data.

MS/MS Networking. Molecular networking was performed following established protocols.[13,42] In brief, parent ions are represented by a list of fragment mass/intensity value pairs within the raw data (*.d files) converted with MSConvert (ProteoWizard package32) into plain text files (*.mgf). These ions are included in the final network once they share at least six fragments

(tolerance Δppm 0.05) with at least one partner ion.[43] Deposited compounds from an *in silico* fragmented[44] commercial database (Antibase 2017[45]) as well as our *in-house* reference compound MS/MS database were included in the final network to highlight known NPs. A visualization of the network was constructed in Cytoscape v3.6.0.[46] Edges were drawn between scan nodes with a cosine similarity >0.7[47].

Strain Fermentation and Purification of Falcitidin Analogs. A pre-culture (R2A, 100 mL in 300 mL Erlenmeyer flask) of *C. eiseniae* DSM 22224 was inoculated from plate (R2A) and incubated at 28 °C with agitation at 180 rpm for 3 days. A 20 L fermentation in medium 3018 (1 g/L yeast extract, 5 g/L casitone, pH 7.0) inoculated with 2% (v/v) pre-culture was carried out separate 2 L flasks filled with 500 mL culture volume at 28 °C with agitation at 180 rpm for 4 days. The culture broth was subsequently freeze-dried using a delta 2-24 LSCplus (Martin Christ Gefriertrockungsanlagen GmbH, Osterode am Harz, Germany). The sample was extracted with one-time culture volume CH_3OH, evaporated to dryness using rotary evaporation under reduced pressure, and resuspened in 3 L of 10% CH_3OH/H_2O. The extract was loaded onto a XAD16N column (1 L bed volume) and eluted step-wise with 10%, 40%, 60%, 80%, and 100% CH_3OH (two-times bed volume each). The 80 and 100% fractions containing falcitidin analogs were pooled and the sample was adjusted to 200 mg/mL in methanol to achieve further separation using preparative C18-RP-HPLC (Synergi 4 µm Fusion-RP 80 Å (250 x 21.2 mm)) by eluting in a linear gradient increasing from 25% to 75% CH_3CN (+0.1% HCOOH) in 22 min. Fractions of interest were concentrated to 100 mg/mL for semi-preparative C18-RP-HPLC (Synergi 4 µm Fusion-RP 80 Å (250 x 10 mm)) using a linear gradient from 15% to 50% CH_3CN (+0.1% HCOOH) in 29 min. Final purification of samples of interest (100 mg/mL) were achieved using UHPLC on a ACQUITY UPLC BEH C18 column (130 Å, 1.7 µm, 100 x 2.1 mm) eluting in a isocratic gradient

of 27.50% CH_3CN (+0.1% HCOOH) in 18 min. In total, isolation yielded 1.5 mg of **2**, 1 mg of **3**, and 2 mg of a mixture of **3** and **4**.

(2*R*/*S*)-1-((3-methylbutanoyl)-D-histidyl-L-isoleucyl-L-valyl)-*N*-((*S*)-1-oxo-3-phenylpropan-2-yl)pyrrolidine-2-carboxamide (**2**). Amorphous, white powder; see Table 1 for ^{1}H and ^{13}C NMR data; positive HR-ESIMS *m/z* 680.4131 $[M+H]^+$, calculated mass for $C_{36}H_{54}N_7O_6^+$ 680.4130; Δ= 0.15 ppm.

(2*R*/*S*)-*N*-((*S*)-1-oxo-3-phenylpropan-2-yl)-1-((2-phenylacetyl)-D-histidyl-L-isoleucyl-L-valyl)pyrrolidine-2-carboxamide (**3**). Amorphous, white powder; see Table 2 for ^{1}H and ^{13}C NMR data; positive HR-ESIMS *m/z* 714.3975 $[M+H]^+$, calculated mass for $C_{39}H_{52}N_7O_6^+$ 714.3974; Δ= 0.14 ppm.

Advanced Marfey's Analysis. The absolute configuration of all amino acids was determined by derivatization using Marfey's reagent.[17] Stock solutions of amino acid standards (50 mM in H_2O), $NaHCO_3$ (1 M in H_2O), and N_α-(2,4-dinitro-5-fluorophenyl)-L-valinamide (L-FDVA, 70 mM in acetone; Sigma Aldrich, St. Louis, MO, Unites States) were prepared. Commercially available and synthesized standards were derivatized using molar ratios of amino acid to L-FDVA and $NaHCO_3$ (1/1.4/8). After stirring at 40 °C for 3 h, 1 M HCl was added to obtain concentration of 170 mM to end the reaction. Samples were subsequently evaporated to dryness and dissolved in DMSO (final concentration 50 mM). L- and D-amino acids were analyzed separately using C18 RP-UHPLC-MS (A: H_2O, 0.1% HCOOH; B: CH_3CN, 0.1% HCOOH; flow rate: 0.6 mL/min). A linear gradient of 15-75% B in 35 min was applied to separate all amino acid standards. Total hydrolysis of the peptide sample containing **3** and **4** was carried out by dissolving 250 µg in 6 M DCl in D_2O and stirring for 7 h at 160 °C. The sample was subsequently evaporated to dryness. Samples were dissolved in 100 µL H_2O, derivatized with L-FDVA and analyzed using the same parameters as described before.

Fluorometric Assays. Rhodesain (Rhod),[48–50] *Staphylococcus aureus* sortase A (SrtA)[51] and human matriptase-2 (TMPRSS6)[52] were expressed and purified as published previously, cathepsins B and L (CatB, CatL; human liver, Calbiochem) and α-chymotrypsin (α-CT; Sigma Aldrich, St. Louis, MO, Unites States) were purchased. For these proteases except SrtA, fluorescence increase upon cleavage of the fluorogenic substrates was monitored without incubation with a TECAN Infinite F200 Pro fluorimeter (excitation $\lambda = 365$ nm; emission $\lambda = 460$ nm) in white, flat-bottom 96-well microtiter plates (Greiner bio-one, Kremsmünster, Austria) with a total volume of 200 µL. Inhibitors and substrates were prepared as stock solutions in DMSO to a final DMSO-content of 0.5%. Inhibitors were screened at final concentrations of 20 µM and eventually at 1 µM. IC_{50} values were determined for compounds for which an inhibition of >50% at a concentration of 20 µM was observed. All assays were performed in technical triplicates and normalized to the activity of DMSO instead of the inhibitors by measuring the increase of fluorescence signal over 10 min. The data were analyzed using GraFit V 5.0.13[53] (Erithracus Software, Horley, UK, http://www.erithacus.com/grafit/).

Cbz-Phe-Arg-AMC (Bachem, Bubendorf BL, Switzerland) was used as a substrate for Rhod, CatB and CatL. The enzymes were incubated at room temperature in enzyme incubation buffer (Rhod: 50 mM sodium acetate pH 5.5, 5 mM EDTA, 200 mM NaCl and 2 mM DTT; CatB/L: 50 mM TRIS-HCl pH 6.5, 5 mM EDTA, 200 mM NaCl, 2 mM DTT) for 30 min. 180 µL assay buffer (Rhod: 50 mM sodium acetate pH 5.5, 5 mM EDTA, 200 mM NaCl and 0.005% Brij35; CatB/L: 50 mM TRIS-HCl pH 6.5, 5 mM EDTA, 200 mM NaCl and 0.005% Brij35) were added to the 96-well plates, afterwards the respective enzyme in enzyme incubation buffer (5 µL; to yield final concentrations for Rhod 0.01 µM, CatB 0.1 µM, CatL 0.2 µM) followed by 10 µL DMSO

(control) or inhibitor solution in DMSO and finally substrate (5 µL; final concentrations for Rhod 10 µM, CatB 100 µM and CatL 6.5 µM) were added.[54]

Transpeptidation efficacy of SrtA was performed *in vitro* as described previously.[51] Briefly, SrtA was diluted in assay buffer (50 mM TRIS-HCl pH 7.50, 150 mM NaCl) to a final concentration of 1 µM. The FRET-pair substrate Abz-LPETG-Dap(Dnp)-OH (Genscript, Piscataway, NJ, Unites States) and the tetraglycine (Sigma Aldrich, St. Louis, MO, Unites States) were added at 25 µM and 0.5 mM, respectively.

α-Chymotrypsin (final concentration 0.4 µM) was dissolved in assay buffer containing 50 mM TRIS-HCl pH 8.0, 100 mM NaCl, 5 mM EDTA. Suc-Leu-Leu-Val-Tyr-AMC (Bachem, Bubendorf BL, Switzerland) was used as a substrate at a final concentration of 52.5 µM.[55]

Proteolytic activity of matriptase-2 was measured in a final concentration of 2.5 nM Enzyme in 180 µL reaction buffer (50 mM TRIS-HCl (pH 8.0), 150 mM NaCl, 5 mM $CaCl_2$, 0.01% (v/v) $T_{x\text{-}100}$). After addition of the inhibitors, reaction was started without further incubation by adding the substrate Boc-Leu-Arg-Arg-AMC (Bachem. Bubendorf BL, Switzerland, $K_M = 36.1 \pm 5.8$ µM) to a final concentration of 100 µM.

The recombinant enzymes falcipain-2 and falcipain-3 were expressed and purified as previously described.[56] Stock solutions of the compounds, substrate and the positive control E-64 (Sigma Aldrich, St. Louis, MO, Unites States) were prepared at 10 mM in DMSO. The compounds were incubated in 96-well white flat bottom plates with 30 nM of recombinant falcipain-2 or 3 at room temperature in assay buffer (100 mM sodium acetate, pH 5.5 with 5 mM DTT) for 10 min. After incubation, the fluorogenic substrate Z-Leu-Arg-AMC (R&D Systems, Minneapolis, MN, Unites States) was added at a concentration of 25 µM in a final assay volume of 200 µL. Fluorescence was monitored with a Varioskan Flash (Thermo Fisher Scientific Inc., Waltham, MA, Unites

States) with excitation 355 nm and emission 460 nm. The IC_{50} was calculated using GraphPad Prism (GraphPad Software, San Diego, CA, Unites States) based on a sigmoidal dose response curve.

Molecular docking. Molecular docking was performed against falcipain-2 (complex with E-64, PDB-ID 3BPF),[57] falcipain-3 (complex with Leupeptin, PDB-ID 3BPM)[57] and chymotrypsin (complex with D-Leucyl-L-phenylalanyl-p-fluorobenzylamide PDB-ID 1AFQ).[36] Conventional non-covalent template-based docking was performed with HYBRID v3.3.0.3 (OpenEye Scientific Software, Santa Fe, NM, US, http://www.eyesopen.com).[58,59] The receptor was prepared using the make_receptor tool version 3.3.0.3 under default settings for potential field generation around the reference ligand for 3BPM and 1AFQ. As the complexed ligand E-64 of 3BPF does not reach toward S1 of falcipain-2, the potential field was generated using leupeptin from the aligned structure of 3BPM (falcipain-3) for which a similar binding behaviour for falcipain-2 and falcipain-3 can be expected.[60] 800 ligand conformers per molecule for docking were generated using omega pose (OMEGA v3.1.0.3, OpenEye Scientific Software, Santa Fe, NM, US, http://www.eyesopen.com).[61] Covalent docking was performed using MOE 2020.09.[62] Using the covalent reaction of acetalization from aldehyde and Cys/Ser (for E-64 re-docking from epoxide to beta-hydroxythioether) a rigid docking with GB/VI scoring was applied for initial placement of 50 poses from which the ten best scoring ones were refined using the ASE scoring function. Docking setups were validated by re-docking of crystallographic reference ligands by pose-inspection and RMSD calculation (Table S15).

ASSOCIATED CONTENT

Supporting Information.

The Supporting Information is available free of charge at LINK.

MS/MS-based assignment of peptide sequences, Marfey's analysis, detailed description of all syntheses, ^{1}H and ^{13}C spectra and data of all synthesized compounds and the natural isolated ones, docking results. (PDF)

AUTHOR INFORMATION

Corresponding Authors

Armin Bauer − *Sanofi-Aventis Deutschland GmbH, R&D, 65926 Frankfurt am Main, Germany*; orcid.org/0000-0002-2171-1333; Email: Armin.Bauer@sanofi.com

Till F. Schäberle − *Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany; Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, 35392 Giessen, Germany*; orcid.org/0000-0001-9947-8079; Phone: +49-641-97-219140; Email: Till.Schaeberle@ime.fraunhofer.de

Authors

Stephan Brinkmann − *Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany*; orcid.org/0000-0003-4299-1079

Sandra Semmler − *Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany*

Christian Kersten − *Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg University Mainz, 55128 Mainz, Germany*; orcid.org/0000-0001-9976-7639

Maria A. Patras − *Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany*; orcid.org/0000-0002-6874-0932

Michael Kurz − *Sanofi-Aventis Deutschland GmbH, R&D, 65926 Frankfurt am Main, Germany*

Natalie Fuchs − *Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg University Mainz, 55128 Mainz, Germany*

Stefan J. Hammerschmidt − *Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg University Mainz, 55128 Mainz, Germany*; orcid.org/0000-0002-0769-8435

Jennifer Legac − *Department of Medicine, University of California, San Francisco, 94143 California, United States*

Peter E. Hammann − *Evotec International GmbH, 37079 Göttingen, Germany*

Andreas Vilcinskas − *Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany; Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, 35392 Giessen, Germany*; orcid.org/0000-0001-8276-4968

Philip. J. Rosenthal − *Department of Medicine, University of California, San Francisco, 94143 California, United States*; orcid.org/0000-0002-7953-7622

Tanja Schirmeister − *Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg University Mainz, 55128 Mainz, Germany*; orcid.org/0000-0002-4587-5076

Author Contributions

[λ] S. Brinkmann and S. Semmler contributed equally.

Funding Sources

This work was financially supported by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK) via the state initiative for the development of scientific and economic excellence for the LOEWE Center for Insect Biotechnology and Bioresources. Sanofi-Aventis

Deutschland GmbH and Evotec International GmbH contributed in the framework of the Sanofi-Fraunhofer Natural Products Center of Excellence/Fraunhofer-Evotec Natural Products Center of Excellence.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors would like to thank Jens Glaeser, Sören M. M. Schuler, Yolanda Kleiner and Christoph Pöverlein for valuable discussions. We thank Christoph Hartwig, Victoria Öhler and the NMR department of the Justus-Liebig-University Giessen for technical assistance. We further thank OpenEye Scientific for free academic software licenses and Torsten Steinmetzer and co-workers from the Philipps University of Marburg for the kind gift of matriptase-2 coding plasmid for recombinant protein expression.

References

(1) Ashley, E. A.; Pyae Phyo, A.; Woodrow, C. J. Malaria. *The Lancet* **2018**, *391* (10130), 1608–1621. DOI: 10.1016/S0140-6736(18)30324-6.

(2) Büscher, P.; Cecchi, G.; Jamonneau, V.; Priotto, G. Human African trypanosomiasis. *The Lancet* **2017**, *390* (10110), 2397–2409. DOI: 10.1016/S0140-6736(17)31510-6.

(3) World Health Organization. *World malaria report 2020: 20 years of global progress and challenges.*; World Health Organization, 2020.

(4) Haldar, K.; Bhattacharjee, S.; Safeukui, I. Drug resistance in Plasmodium. *Nature reviews. Microbiology* **2018**, *16* (3), 156–170. DOI: 10.1038/nrmicro.2017.161. Published Online: Jan. 22, 2018.

(5) Conrad, M. D.; Rosenthal, P. J. Antimalarial drug resistance in Africa: the calm before the storm? *The Lancet Infectious Diseases* **2019**, *19* (10), e338-e351. DOI: 10.1016/S1473-3099(19)30261-0.

(6) Otto, H.-H.; Schirmeister, T. Cysteine Proteases and Their Inhibitors. *Chemical reviews* **1997**, *97* (1), 133–172. DOI: 10.1021/cr950025u.

(7) Ettari, R.; Previti, S.; Tamborini, L.; Cullia, G.; Grasso, S.; Zappalà, M. The Inhibition of Cysteine Proteases Rhodesain and TbCatB: A Valuable Approach to Treat Human African Trypanosomiasis. *Mini reviews in medicinal chemistry* **2016**, *16* (17), 1374–1391. DOI: 10.2174/1389557515666160509125243.

(8) Rosenthal, P. J. Falcipain cysteine proteases of malaria parasites: An update. *Biochimica et biophysica acta. Proteins and proteomics* **2020**, *1868* (3), 140362. DOI: 10.1016/j.bbapap.2020.140362. Published Online: Jan. 9, 2020.
(9) Drag, M.; Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nature reviews. Drug discovery* **2010**, *9* (9), 690–701. DOI: 10.1038/nrd3053.
(10) Somanadhan, B.; Kotturi, S. R.; Yan Leong, C.; Glover, R. P.; Huang, Y.; Flotow, H.; Buss, A. D.; Lear, M. J.; Butler, M. S. Isolation and synthesis of falcitidin, a novel myxobacterial-derived acyltetrapeptide with activity against the malaria target falcipain-2. *The Journal of antibiotics* **2013**, *66* (5), 259–264. DOI: 10.1038/ja.2012.123. Published Online: Jan. 23, 2013.
(11) Kotturi, S. R.; Somanadhan, B.; Ch'ng, J.-H.; Tan, K. S.-W.; Butler, M. S.; Lear, M. J. Diverted total synthesis of falcitidin acyl tetrapeptides as new antimalarial leads. *Tetrahedron Letters* **2014**, *55* (11), 1949–1951. DOI: 10.1016/j.tetlet.2014.02.008.
(12) Krug, D.; Müller, R. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **2014**, *31* (6), 768–783. DOI: 10.1039/c3np70127a. Published Online: Apr. 25, 2014.
(13) Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; Felicio, R. de; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C. Molecular networking as a dereplication strategy. *Journal of natural products* **2013**, *76* (9), 1686–1699. DOI: 10.1021/np400413s. Published Online: Sep. 11, 2013.
(14) Brinkmann, S.; Kurz, M.; Patras, M. A.; Hartwig, C.; Marner, M.; Leis, B.; Billion, A.; Kleiner, Y.; Bauer, A.; Toti, L.; Pöverlein, C.; Hammann, P. E.; Vilcinskas, A.; Glaeser, J.; Spohn, M. S.; Schäberle, T. F. *Genomic and chemical decryption of the Bacteroidetes phylum for its potential to biosynthesize natural products*, 2021. DOI: 10.1101/2021.07.30.454449.
(15) Tan, S.; Li, F.; Park, S.; Kim, S. Transition metal-free aerobic oxidative cleavage of the C–N bonds of α-amino esters. *Org. Chem. Front.* **2019**, *6* (23), 3854–3858. DOI: 10.1039/C9QO01033E.
(16) DOERING, W. E.; CHANLEY, J. D. The autoxidation of quininone. *J. Am. Chem. Soc.* **1946**, *68*, 586–588. DOI: 10.1021/ja01208a017.
(17) Bhushan, R.; Brückner, H. Marfey's reagent for chiral amino acid analysis: a review. *Amino acids* **2004**, *27* (3-4), 231–247. DOI: 10.1007/s00726-004-0118-0. Published Online: Oct. 22, 2004.
(18) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research* **2019**, *47* (W1), W81-W87. DOI: 10.1093/nar/gkz310.
(19) Katoh, K.; Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **2013**, *30* (4), 772–780. DOI: 10.1093/molbev/mst010. Published Online: Jan. 16, 2013.
(20) Du, L.; Lou, L. PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **2010**, *27* (2), 255–278. DOI: 10.1039/B912037H.
(21) Gaitatzis, N.; Kunze, B.; Muller, R. In vitro reconstitution of the myxochelin biosynthetic machinery of Stigmatella aurantiaca Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. *Proceedings of the National Academy of Sciences* **2001**, *98* (20), 11136–11141. DOI: 10.1073/pnas.201167098.

(22) Silakowski, B.; Nordsiek, G.; Kunze, B.; Blöcker, H.; Müller, R. Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium Stigmatella aurantiaca Sga1511This article is dedicated to Prof. Dr. E. Leistner on the occasion of his 60th birthday. *Chemistry & Biology* **2001**, *8* (1), 59–69. DOI: 10.1016/S1074-5521(00)00056-9.

(23) Li, Y.; Weissman, K. J.; Müller, R. Myxochelin Biosynthesis: Direct Evidence for Two- and Four-Electron Reduction of a Carrier Protein-Bound Thioester. *J. Am. Chem. Soc.* **2008**, *130* (24), 7554–7555. DOI: 10.1021/ja8025278.

(24) Chhabra, A.; Haque, A. S.; Pal, R. K.; Goyal, A.; Rai, R.; Joshi, S.; Panjikar, S.; Pasha, S.; Sankaranarayanan, R.; Gokhale, R. S. Nonprocessive 2 + 2e- off-loading reductase domains from mycobacterial nonribosomal peptide synthetases. *Proceedings of the National Academy of Sciences* **2012**, *109* (15), 5681–5686. DOI: 10.1073/pnas.1118680109. Published Online: Mar. 26, 2012.

(25) Barajas, J. F.; Phelan, R. M.; Schaub, A. J.; Kliewer, J. T.; Kelly, P. J.; Jackson, D. R.; Luo, R.; Keasling, J. D.; Tsai, S.-C. Comprehensive Structural and Biochemical Analysis of the Terminal Myxalamid Reductase Domain for the Engineered Production of Primary Alcohols. *Chemistry & Biology* **2015**, *22* (8), 1018–1029. DOI: 10.1016/j.chembiol.2015.06.022. Published Online: Jul. 30, 2015.

(26) Aarts, M. G.; Hodge, R.; Kalantidis, K.; Florack, D.; Wilson, Z. A.; Mulligan, B. J.; Stiekema, W. J.; Scott, R.; Pereira, A. The Arabidopsis MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. *The Plant journal : for cell and molecular biology* **1997**, *12* (3), 615–623. DOI: 10.1046/j.1365-313x.1997.00615.x.

(27) METZ, J. I.M.; LASSNER, M. Reprogramming of Oil Synthesis in Rapeseed: Industrial Applications. *Ann NY Acad Sci* **1996**, *792* (1 Engineering P), 82–90. DOI: 10.1111/j.1749-6632.1996.tb32494.x.

(28) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154. DOI: 10.1021/ja00897a025.

(29) Serim, S.; Haedke, U.; Verhelst, S. H. L. Activity-based probes for the study of proteases: recent advances and developments. *ChemMedChem* **2012**, *7* (7), 1146–1159. DOI: 10.1002/cmdc.201200057. Published Online: Mar. 19, 2012.

(30) Gehringer, M.; Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *Journal of medicinal chemistry* **2019**, *62* (12), 5673–5724. DOI: 10.1021/acs.jmedchem.8b01153. Published Online: Jan. 25, 2019.

(31) Sijwali, P. S.; Rosenthal, P. J. Gene disruption confirms a critical role for the cysteine protease falcipain-2 in hemoglobin hydrolysis by Plasmodium falciparum. *Proceedings of the National Academy of Sciences* **2004**, *101* (13), 4384–4389. DOI: 10.1073/pnas.0307720101. Published Online: Mar. 15, 2004.

(32) Andrusier, N.; Mashiach, E.; Nussinov, R.; Wolfson, H. J. Principles of flexible protein-protein docking. *Proteins* **2008**, *73* (2), 271–289. DOI: 10.1002/prot.22170.

(33) Pons, C.; Grosdidier, S.; Solernou, A.; Pérez-Cano, L.; Fernández-Recio, J. Present and future challenges and limitations in protein-protein docking. *Proteins* **2010**, *78* (1), 95–108. DOI: 10.1002/prot.22564.

(34) Knaff, P. M.; Kersten, C.; Willbold, R.; Champanhac, C.; Crespy, D.; Wittig, R.; Landfester, K.; Mailänder, V. From In Silico to Experimental Validation: Tailoring Peptide

Substrates for a Serine Protease. *Biomacromolecules* **2020**, *21* (4), 1636–1643. DOI: 10.1021/acs.biomac.0c00240. Published Online: Apr. 2, 2020.
(35) Silva, D. G.; Ribeiro, J. F. R.; Vita, D. de; Cianni, L.; Franco, C. H.; Freitas-Junior, L. H.; Moraes, C. B.; Rocha, J. R.; Burtoloso, A. C. B.; Kenny, P. W.; Leitão, A.; Montanari, C. A. A comparative study of warheads for design of cysteine protease inhibitors. *Bioorganic & medicinal chemistry letters* **2017**, *27* (22), 5031–5035. DOI: 10.1016/j.bmcl.2017.10.002. Published Online: Oct. 3, 2017.
(36) Kashima, A.; Inoue, Y.; Sugio, S.; Maeda, I.; Nose, T.; Shimohigashi, Y. X-ray crystal structure of a dipeptide-chymotrypsin complex in an inhibitory interaction. *European journal of biochemistry* **1998**, *255* (1), 12–23. DOI: 10.1046/j.1432-1327.1998.2550012.x.
(37) Stoll, V. S.; Eger, B. T.; Hynes, R. C.; Martichonok, V.; Jones, J. B.; Pai, E. F. Differences in binding modes of enantiomers of 1-acetamido boronic acid based protease inhibitors: crystal structures of gamma-chymotrypsin and subtilisin Carlsberg complexes. *Biochemistry* **1998**, *37* (2), 451–462. DOI: 10.1021/bi971166o.
(38) Rawlings, N. D.; Waller, M.; Barrett, A. J.; Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* **2014**, *42* (Database issue), D503-9. DOI: 10.1093/nar/gkt953. Published Online: Oct. 23, 2013.
(39) Cho, S.; Dickey, S. W.; Urban, S. Crystal Structures and Inhibition Kinetics Reveal a Two-Stage Catalytic Mechanism with Drug Design Implications for Rhomboid Proteolysis. *Molecular cell* **2016**, *61* (3), 329–340. DOI: 10.1016/j.molcel.2015.12.022. Published Online: Jan. 21, 2016.
(40) Lyons, B.; Kwan, A. H.; Jamie, J.; Truscott, R. J. W. Age-dependent modification of proteins: N-terminal racemization. *The FEBS journal* **2013**, *280* (9), 1980–1990. DOI: 10.1111/febs.12217. Published Online: Mar. 20, 2013.
(41) Kajita, R.; Goto, T.; Lee, S. H.; Oe, T. Aldehyde stress-mediated novel modification of proteins: epimerization of the N-terminal amino acid. *Chemical research in toxicology* **2013**, *26* (12), 1926–1936. DOI: 10.1021/tx400354d. Published Online: Dec. 3, 2013.
(42) Allard, P.-M.; Péresse, T.; Bisson, J.; Gindro, K.; Marcourt, L.; van Pham, C.; Roussi, F.; Litaudon, M.; Wolfender, J.-L. Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. *Analytical chemistry* **2016**, *88* (6), 3317–3323. DOI: 10.1021/acs.analchem.5b04804. Published Online: Mar. 1, 2016.
(43) Riyanti; Marner, M.; Hartwig, C.; Patras, M. A.; Wodi, S. I. M.; Rieuwpassa, F. J.; Ijong, F. G.; Balansa, W.; Schäberle, T. F. Sustainable Low-Volume Analysis of Environmental Samples by Semi-Automated Prioritization of Extracts for Natural Product Research (SeaPEPR). *Marine drugs* **2020**, *18* (12). DOI: 10.3390/md18120649. Published Online: Dec. 17, 2020.
(44) Allen, F.; Greiner, R.; Wishart, D. Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics* **2015**, *11* (1), 98–110. DOI: 10.1007/s11306-014-0676-4.
(45) Laatsch, H. *AntiBase: The Natural Compound Identifier*, 5. Auflage; Wiley-VCH, 2017.
(46) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **2003**, *13* (11), 2498–2504. DOI: 10.1101/gr.1239303.
(47) Marner, M.; Patras, M. A.; Kurz, M.; Zubeil, F.; Förster, F.; Schuler, S.; Bauer, A.; Hammann, P.; Vilcinskas, A.; Schäberle, T. F.; Glaeser, J. Molecular Networking-Guided Discovery and Characterization of Stechlisins, a Group of Cyclic Lipopeptides from a

Pseudomonas sp. *Journal of natural products* **2020**, *83* (9), 2607–2617. DOI: 10.1021/acs.jnatprod.0c00263. Published Online: Aug. 21, 2020.
(48) Previti, S.; Ettari, R.; Cosconati, S.; Amendola, G.; Chouchene, K.; Wagner, A.; Hellmich, U. A.; Ulrich, K.; Krauth-Siegel, R. L.; Wich, P. R.; Schmid, I.; Schirmeister, T.; Gut, J.; Rosenthal, P. J.; Grasso, S.; Zappalà, M. Development of Novel Peptide-Based Michael Acceptors Targeting Rhodesain and Falcipain-2 for the Treatment of Neglected Tropical Diseases (NTDs). *J. Med. Chem.* **2017**, *60* (16), 6911–6923. DOI: 10.1021/acs.jmedchem.7b00405. Published Online: Aug. 11, 2017.
(49) Johé, P.; Jaenicke, E.; Neuweiler, H.; Schirmeister, T.; Kersten, C.; Hellmich, U. A. Structure, interdomain dynamics, and pH-dependent autoactivation of pro-rhodesain, the main lysosomal cysteine protease from African trypanosomes. *The Journal of biological chemistry* **2021**, *296*, 100565. DOI: 10.1016/j.jbc.2021.100565. Published Online: Mar. 18, 2021.
(50) Jung, S.; Fuchs, N.; Johe, P.; Wagner, A.; Diehl, E.; Yuliani, T.; Zimmer, C.; Barthels, F.; Zimmermann, R. A.; Klein, P.; Waigel, W.; Meyr, J.; Opatz, T.; Tenzer, S.; Distler, U.; Räder, H.-J.; Kersten, C.; Engels, B.; Hellmich, U. A.; Klein, J.; Schirmeister, T. Fluorovinylsulfones and -Sulfonates as Potent Covalent Reversible Inhibitors of the Trypanosomal Cysteine Protease Rhodesain: Structure-Activity Relationship, Inhibition Mechanism, Metabolism, and In Vivo Studies. *J. Med. Chem.* **2021**, *64* (16), 12322–12358. DOI: 10.1021/acs.jmedchem.1c01002. Published Online: Aug. 11, 2021.
(51) Barthels, F.; Marincola, G.; Marciniak, T.; Konhäuser, M.; Hammerschmidt, S.; Bierlmeier, J.; Distler, U.; Wich, P. R.; Tenzer, S.; Schwarzer, D.; Ziebuhr, W.; Schirmeister, T. Asymmetric Disulfanylbenzamides as Irreversible and Selective Inhibitors of Staphylococcus aureus Sortase A. *ChemMedChem* **2020**, *15* (10), 839–850. DOI: 10.1002/cmdc.201900687. Published Online: Mar. 25, 2020.
(52) Steinmetzer, T.; Schweinitz, A.; Stürzebecher, A.; Dönnecke, D.; Uhland, K.; Schuster, O.; Steinmetzer, P.; Müller, F.; Friedrich, R.; Than, M. E.; Bode, W.; Stürzebecher, J. Secondary amides of sulfonylated 3-amidinophenylalanine. New potent and selective inhibitors of matriptase. *Journal of medicinal chemistry* **2006**, *49* (14), 4116–4126. DOI: 10.1021/jm0512721.
(53) Leatherbarrow, R. J. *GraFit Version 5*; Erithacus Software Ltd., Horley, U.K., 2006.
(54) Schirmeister, T.; Kesselring, J.; Jung, S.; Schneider, T. H.; Weickert, A.; Becker, J.; Lee, W.; Bamberger, D.; Wich, P. R.; Distler, U.; Tenzer, S.; Johé, P.; Hellmich, U. A.; Engels, B. Quantum Chemical-Based Protocol for the Rational Design of Covalent Inhibitors. *J. Am. Chem. Soc.* **2016**, *138* (27), 8332–8335. DOI: 10.1021/jacs.6b03052. Published Online: Jul. 1, 2016.
(55) Ludewig, S.; Kossner, M.; Schiller, M.; Baumann, K.; Schirmeister, T. Enzyme kinetics and hit validation in fluorimetric protease assays. *Current topics in medicinal chemistry* **2010**, *10* (3), 368–382. DOI: 10.2174/156802610790725498.
(56) Sijwali, P. S.; Brinen, L. S.; Rosenthal, P. J. Systematic optimization of expression and refolding of the Plasmodium falciparum cysteine protease falcipain-2. *Protein expression and purification* **2001**, *22* (1), 128–134. DOI: 10.1006/prep.2001.1416.
(57) Kerr, I. D.; Lee, J. H.; Pandey, K. C.; Harrison, A.; Sajid, M.; Rosenthal, P. J.; Brinen, L. S. Structures of falcipain-2 and falcipain-3 bound to small molecule inhibitors: implications for substrate specificity. *Journal of medicinal chemistry* **2009**, *52* (3), 852–857. DOI: 10.1021/jm8013663.
(58) McGann, M. FRED pose prediction and virtual screening accuracy. *Journal of chemical information and modeling* **2011**, *51* (3), 578–596. DOI: 10.1021/ci100436p. Published Online: Feb. 16, 2011.

(59) McGann, M. FRED and HYBRID docking performance on standardized datasets. *Journal of computer-aided molecular design* **2012**, *26* (8), 897–906. DOI: 10.1007/s10822-012-9584-8. Published Online: Jun. 5, 2012.

(60) Lisk, G.; Pain, M.; Gluzman, I. Y.; Kambhampati, S.; Furuya, T.; Su, X.-Z.; Fay, M. P.; Goldberg, D. E.; Desai, S. A. Changes in the plasmodial surface anion channel reduce leupeptin uptake and can confer drug resistance in Plasmodium falciparum-infected erythrocytes. *Antimicrobial agents and chemotherapy* **2008**, *52* (7), 2346–2354. DOI: 10.1128/AAC.00057-08. Published Online: Apr. 28, 2008.

(61) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *Journal of chemical information and modeling* **2010**, *50* (4), 572–584. DOI: 10.1021/ci100031x.

(62) *Molecular Operating Environment (MOE)*; 2020.09; Chemical Computing Group ULC; 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2020. https://www.chemcomp.com/index.htm.