

# Data-mining unveils structure–property–activity correlation of viral infectivity enhancing self-assembling peptides

Received: 27 February 2023

Accepted: 1 August 2023

Published online: 23 August 2023

 Check for updatesKübra Kaygisiz<sup>1</sup>, Lena Rauch-Wirth<sup>2</sup>, Arghya Dutta<sup>3,5</sup>, Xiaoqing Yu<sup>4</sup>, Yuki Nagata<sup>4</sup>, Tristan Berau<sup>3,6</sup>, Jan Münch<sup>2</sup>, Christopher V. Synatschke<sup>1</sup>✉ & Tanja Weil<sup>1</sup>✉

Gene therapy via retroviral vectors holds great promise for treating a variety of serious diseases. It requires the use of additives to boost infectivity. Amyloid-like peptide nanofibers (PNFs) were shown to efficiently enhance retroviral gene transfer. However, the underlying mode of action of these peptides remains largely unknown. Data-mining is an efficient method to systematically study structure–function relationship and unveil patterns in a database. This data-mining study elucidates the multi-scale structure–property–activity relationship of transduction enhancing peptides for retroviral gene transfer. In contrast to previous reports, we find that not the amyloid fibrils themselves, but rather  $\mu\text{m}$ -sized  $\beta$ -sheet rich aggregates enhance infectivity. Specifically, microscopic aggregation of  $\beta$ -sheet rich amyloid structures with a hydrophobic surface pattern and positive surface charge are identified as key material properties. We validate the reliability of the amphiphilic sequence pattern and the general applicability of the key properties by rationally creating new active sequences and identifying short amyloidal peptides from various pathogenic and functional origin. Data-mining—even for small datasets—enables the development of new efficient retroviral transduction enhancers and provides important insights into the diverse bioactivity of the functional material class of amyloids.

The stable and safe integration of genetic information is an emerging challenge in cell engineering, e.g., for gene therapy. Retroviral gene transfer is the commonly used method to integrate genetic payloads into cells. Retroviral vectors can deliver therapeutic transgenes with low immunogenicity and hold great promise for the treatment of genetic disorders in recent *ex vivo* clinical trials<sup>1</sup> and have been approved for chimeric antigen receptor (CAR) T-cell therapy<sup>2</sup>.

However, low transduction efficiencies of the viral vectors at target cells resulting from low virus–cell attachment rates are still a major challenge that limits therapeutic efficacy<sup>3,4</sup>.

Various materials such as polymers, lipids, and peptides have been introduced as additives to increase viral transduction efficiency mainly by facilitating colocalization of viruses and cells<sup>5</sup>. Interestingly, the transduction-enhancing ability of amyloid peptide nanofibers

<sup>1</sup>Department Synthesis of Macromolecules, Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany. <sup>2</sup>Institute of Molecular Virology, Ulm University Medical Center, Meyerhofstraße 1, 89081 Ulm, Germany. <sup>3</sup>Department Polymer Theory, Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany. <sup>4</sup>Department Molecular Spectroscopy, Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany. <sup>5</sup>Present address: Institute of Biochemistry II, Faculty of Medicine, Goethe University, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany. <sup>6</sup>Present address: Institute for Theoretical Physics, Heidelberg University, Philosophenweg 19, 69120 Heidelberg, Germany.

✉ e-mail: [synatschke@mpip-mainz.mpg.de](mailto:synatschke@mpip-mainz.mpg.de); [weil@mpip-mainz.mpg.de](mailto:weil@mpip-mainz.mpg.de)

(PNF) was first found in seminal fluid in the context of human immunodeficiency virus 1 (HIV-1) infection<sup>6</sup>. This finding resulted in intense research on peptide fibrils to boost viral infectivity<sup>7–11</sup>. For example, the 12-mer peptide enhancing factor C (EF-C, QCKIKQIINMWQ) corresponding to residues 417–428 of the HIV-1 envelope glycoprotein 120 (gp120) was found in a random screening approach and spontaneously self-assembled into positively charged amyloid fibrils, which strongly enhance retroviral transduction<sup>12</sup>. Due to their versatile material properties, amyloid nanostructures have attracted broad interest for various therapeutic applications, such as gene therapy<sup>12</sup> or as additives in vaccines<sup>13</sup>.

Many peptides and proteins have a propensity to assemble into hierarchically ordered amyloid structures and Nature uses these structures in both, pathogenic and functional contexts<sup>14,15</sup>. A hypothesis that distinguishes pathogenic from functional amyloids is based on the kinetics of their assembly process. It has been proposed that oligomeric species rather than mature, thermodynamically stable amyloid fibrils are related to cytotoxicity<sup>16–18</sup>. Oligomers that prevail for a long time in solution can interact and also refold other monomeric proteins thus acting as multipliers for proteins misfolding and aggregation<sup>19,20</sup>. In contrast, reports on functional amyloids typically show very fast and quantitative assembly without long lag times, thus preventing the formation of cytotoxic oligomers<sup>21,22</sup>. The tensile strength and stability of amyloid fibrils are attributed to their high cross  $\beta$ -sheet structures, which are mainly stabilized by hydrogen bonding and hydrophobic interactions<sup>23</sup>. The defined molecular structure of the peptide building blocks enables modifications of the sequence at the level of single amino acids, which gives access to tuneable assembly features and materials properties from the nano- to the microscale (structure–property), resulting in programmable bioactivity (property–activity)<sup>24</sup>.

The observation that certain functional amyloids are biocompatible and exhibit low cytotoxicity combined with virus-binding properties makes them a promising class of materials for retro- and lentiviral gene therapy<sup>12,25,26</sup>. In previous studies, we and others have shown via electron and fluorescence microscopy that virus-particles associate to fibrillar structures<sup>12,26,27</sup>. However, a fundamental understanding of the corresponding molecular as well as nanoscopic property–activity relationship is limited by inherent multi-scale and multi-parameter complexity of the supramolecular self-assembling systems. Data-mining and regression models are efficient tools to study structure–function relationship and unveil underlying design principles and patterns in a database<sup>28,29</sup>. However, so far, the relationship between properties and bioactivity of peptides has been elucidated mainly for antimicrobial<sup>30</sup> or antiviral peptides<sup>31</sup>, while they are still elusive for self-assembling nanostructured peptides with transduction enhancing properties.

We have designed and synthesized a peptide library consisting of 163 sequences to study structure–property–activity relationships for infectivity enhancement of nanostructures composed of short peptides. Based on the known and commercially available self-assembling 12-mer peptide EF-C (Fig. 1A)<sup>12</sup>, a smaller library was reported by us containing optimized peptide sequences with improved bioactivity<sup>26,32</sup>. However, a fundamental understanding of the structure–property–activity relationship of amyloid peptides and their unique infectivity is still lacking. Therefore, we expand the peptide library by scrambling and substituting amino acids, point mutations, peptide sequence truncations, extensions with bioactive epitopes, and random peptide sequences (Table S2). The resulting peptide library includes peptides with sequence lengths ranging from 4 to 19 amino acids (Table S3, Fig. 1B), whose bioactivity, aggregation, and nanostructure formation were characterized in a quantitative HIV-1 infection assay as well as by transmission electron microscopy (TEM), light scattering count rate, zeta-potential measurements, and

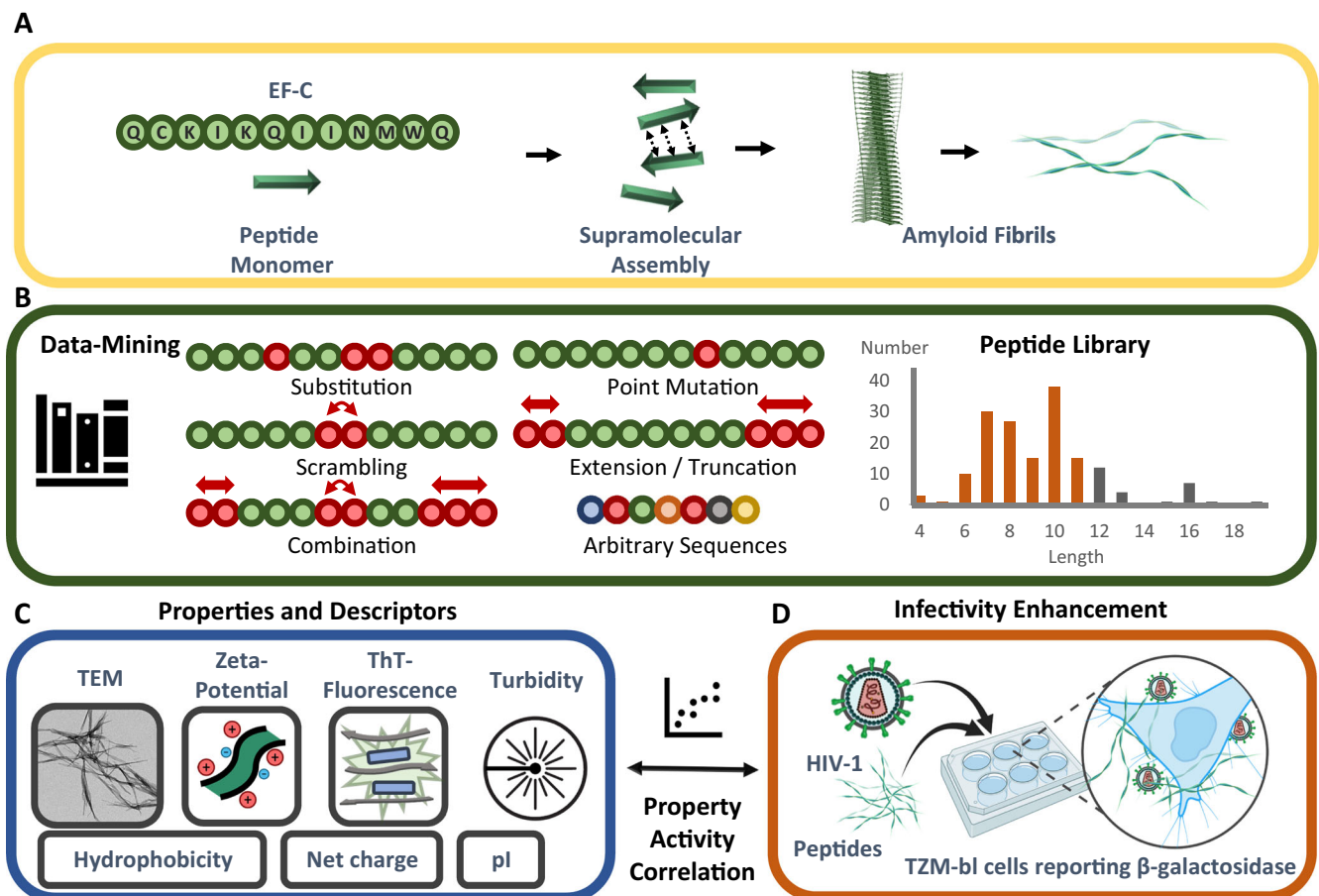
Thioflavin T (ThT) fluorescence assays. Collecting this multi-parameter dataset provided the physicochemical properties for morphology, microscopic aggregation, surface charges, and  $\beta$ -sheet amyloid structures, respectively. Together with bioinformatic parameters, these acquired data were applied in a property–activity correlation analysis using multiple linear regression. Microscopic aggregation of already formed amyloid structures with positive surface charges was identified as the key material characteristics for boosting viral transduction. We demonstrated the general understanding and universal applicability of the key properties for pathogenic and functional amyloids, unrelated to EF-C. To the best of our knowledge, this is the first structure–property–activity relationship of infectivity-enhancing amyloid peptide nanomaterials, that enables not only an understanding but also the rational design and discovery of therapeutic and pathogenic peptide sequences. We envision that the presented study provides a fundamental understanding of this important structural class.

## Results

### Characterization of a peptide library based on EF-C derivatives

To unveil the structure–property–activity relationship of infectivity enhancing peptides, we performed a data-mining approach. This procedure involves data-generation by creating a peptide library, processing and analyses of the peptide measurements, establishing a model for the structure–property–activity relationship, and finally interpretation and prediction based on the established relationship. The peptide library was synthesized via routine solid phase peptide synthesis according to Fmoc-protocol<sup>33</sup> and their purities were confirmed by liquid chromatography mass spectrometry (LC-MS, Fig. S35). Unless stated otherwise, the sequences were obtained in >95% purity by LC-MS from commercial suppliers (Table S3). To induce fibril formation, the peptides were first dissolved in dimethyl sulfoxide (DMSO) and then diluted in phosphate-buffered saline (PBS). The physicochemical properties of the peptides were analyzed on multiple length scales (Fig. 1C). On nanometer scale, the morphologies of molecular assemblies were investigated by TEM and were categorized in (1) fibrillar and (2) non-fibrillar structures (SI Section 1.1, Fig. S1, Fig. S33). Fibril formation was further assessed by the amyloid sensitive ThT dye. Enhancement of fluorescence intensity was accounted to ThT binding to  $\beta$ -sheet rich fibrils (SI Section 1.1)<sup>34</sup>. Both fibril formation and ThT-activity of peptides were evaluated as binary features (fibrils/no fibrils and ThT-active/not ThT-active, Table S5).  $\beta$ -Sheet content was determined via Fourier transform infrared spectroscopy (FT-IR) analysis (Fig. S34) and efficiency to assemble into nanostructures was determined via the conversion rate (CR, Table S5) assay as reported earlier<sup>26</sup>. On the micrometer scale, zeta-potential measurements were conducted to gain information on surface charge of peptide assemblies. Another parameter taken from the zeta-potential measurements was the derived count rate of scattered light, which gives information on turbidity of the sample due to aggregation of fibrillar assemblies (SI Section 1.1, Table S5)<sup>35</sup>. The microscopic aggregation of PNF was further characterized for selected examples via brightfield and fluorescence microscopy (Fig. S32). In this study, we use the term microscopic aggregation to describe the fiber-fiber interaction into less ordered micrometer-sized particles and exclude the highly ordered  $\mu$ m-sized crystalline alignments that have been previously studied for amyloid aggregation<sup>36–38</sup>. Bioinformatic parameters for the library are calculated with amino acid properties to yield global peptide properties like hydrophobicity or isoelectric point (SI Section 1.2, Table S4)<sup>39,40</sup>.

As an indicator of biological activity, the transduction enhancement of the self-assembled peptides was assessed by determining HIV-1 infection rates of TZM-bl cells (Fig. 1D). This cell type is a widely applied model cell line for high throughput analysis



**Fig. 1 | Schematic visualization of this work to identify property-activity relationships of peptides enhancing retroviral transduction enhancement.** **A** EF-C is a 12-mer peptide sequence, derived from the glycoprotein gp120 of the human immunodeficiency virus 1 (HIV-1), that can assemble via supramolecular interactions into amyloid fibrils with high cross- $\beta$  sheet content. **B** Amino acids of the sequence of EF-C were systematically substituted, point mutated, or scrambled. Sequence extensions and truncations have been conducted by repeating terminal amino acids or introducing bioactive peptides. Randomly generated peptides were added to the library with no connection to EF-C. The resulting library consists of 163

peptides, mostly comprising fewer than 12 residues (Table S3). **C** Experimental methods, such as TEM (Fig. S33), zeta-potential, ThT fluorescence (Table S5), turbidity analysis (Table S8) as well as bioinformatic characterization such as hydrophobicity, net charge and isoelectric point (pI) (Table S4) were applied to correlate with **(D)** the infectivity enhancement of HIV-1 on T2M-bl cells by the peptides, as quantified via a chemiluminescence based assay (Table S5). HIV-1 and peptides were incubated together before they were added to cells. Created with BioRender.com. Source data for **B** is provided within the Source Data file.

of HIV-1 infection, because it expresses the HIV-1 receptors CD4 and CCR5 and reports  $\beta$ -galactosidase upon successful infection<sup>41</sup>. HIV-1 was mixed with different concentrations of peptides and subsequently used to infect T2M-bl cells. The infection rates were assessed with a chemiluminescence-based assay, which allowed quantification of the infection rates relative to virus-only infectivity. For example, a 1.3  $\mu$ M concentration of EF-C fibrils enhanced HIV-1 infection by a factor of  $42 \pm 7$ , as compared to infection rates of the virus only control (Table S5). In this study, EF-C at 1.3  $\mu$ M concentration was selected as a reference, to compare quantitative infection data obtained for every peptide among various measurements (Table S5). Additionally, the cell viability was determined for a representative selection of 77 peptides at a concentration of 1.3  $\mu$ M and revealed no toxicity (Figs. S2, S3).

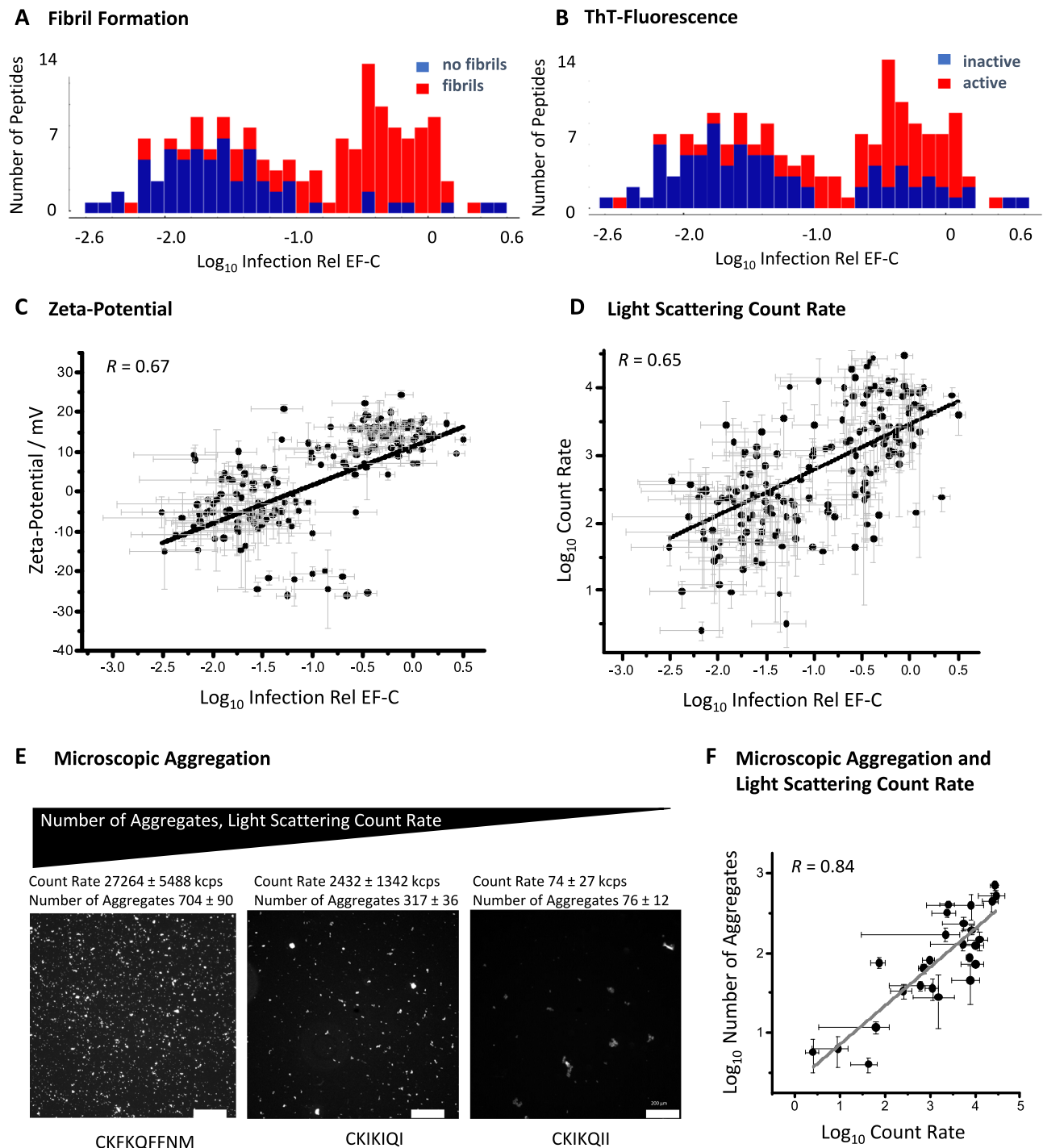
The herein investigated peptide library based on EF-C derivatives consists of peptides, which are similar in sequence (Table S2), but show a wide variety of biological activity (*i.e.* infectivity enhancement), which allow to study structure-property-activity relationships. Peptides showing an  $n$ -fold infectivity enhancement of <4-fold relative to virus-only infectivity or <10% relative to EF-C are considered as inactive, which is the case for 49% of the peptide library (Table S5).

### Microscopically aggregated amyloid fibrils and positive surface charge highly correlate with transduction enhancement

To elucidate the structure-property-activity relationship, we applied linear regression on the experimental and bioinformatic descriptors and evaluated the Pearson correlation coefficient (Pearson  $R$ ). Notably, for most of the descriptors, the logarithm of the infection enhancement relative to EF-C ( $\text{Log}_{10}$  Infection Rel. EF-C) correlates with a higher Pearson  $R$  compared to absolute infection enhancement values. To enable linear regression, we therefore applied logarithmic scale for the infectivity relative to EF-C for all the following correlations.

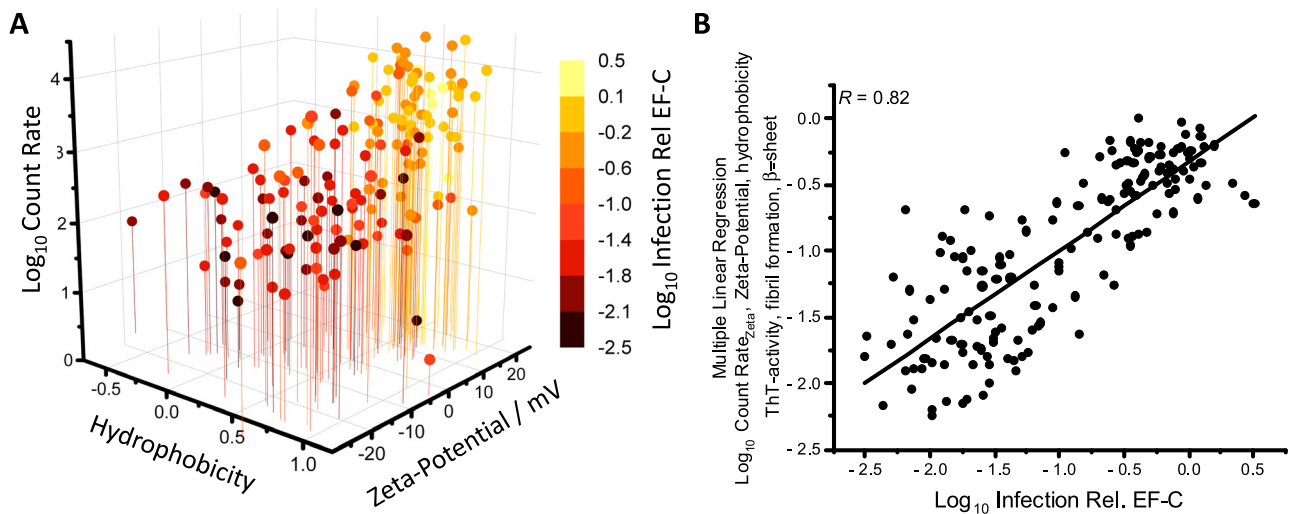
Fibril formation (Fig. 2A, Fig. S5  $R = 0.59$ ), ThT-fluorescence (Fig. 2B, Fig. S5  $R = 0.40$ ), and hydrophobicity according to Fauchère hydropathy scale (Fig. S5,  $R = 0.40$ ) correlate strongly with infectivity enhancement. Our data indicate that a high hydrophobicity is important to achieve quantitative assembly of peptides into fibrils (Fig. S6). The importance of the latter is further demonstrated with a representative selection of 80 peptides (Table S5), which we additionally characterized regarding the conversion of monomers to nanotubes (Fig. S5, conversion rate,  $R = 0.50$ , Fig. S7 B, Table S5).

Previous reports indicated that the zeta-potential representing the surface charge of the fibrils plays a crucial role for transduction enhancement<sup>6,12,25,27</sup>. Plotting the zeta-potential against  $\text{Log}_{10}$  Infection



**Fig. 2 | Correlation of experimental parameters with  $\text{Log}_{10}$  Infection Rel EF-C of a library containing 163 peptides.** **A** Histogram of  $\text{Log}_{10}$  Infection Rel EF-C showing fibril forming (red) and non-fibril forming (blue) peptide distribution analyzed via TEM (Fig. S33). **B** Histogram of  $\text{Log}_{10}$  Infection Rel EF-C showing ThT-active (red) and ThT-inactive (blue) peptide distribution (Table S5). **C** Zeta-potential plotted against  $\text{Log}_{10}$  Infection Rel EF-C. Mean values are displayed with error bars indicating standard deviation from triplicate measurements (Table S5). Linear fit with Pearson correlation factor  $R = 0.67$ . **D** Count rate of scattered light plotted against  $\text{Log}_{10}$  Infection Rel EF-C. Linear fit with Pearson correlation factor  $R = 0.65$ . Mean values are displayed with error bars indicating standard deviation from triplicate measurements (Table S5). **E** Selection of peptides with high and low count rate showing relationship between number of particles and count rate of

scattered light (Table S8). Peptide fibrils were preincubated at 1 mg/mL one day and diluted to 0.1 mg/mL with ThT solution (50  $\mu\text{M}$ ) just before measurement (Fig. S32). Number of aggregated fibrils ( $> 10 \mu\text{m}^2$ ) were quantified via fluorescence microscopy in an area of  $1330 \mu\text{m} \times 1330 \mu\text{m}$  in triplicate measurements via automated counting (3D object counter ImageJ), scale bar  $200 \mu\text{m}$ . **F**  $\text{Log}_{10}$  Count rate from zeta-potential measurements of representatively selected peptides plotted against averaged number of aggregated particles in an area of  $1330 \mu\text{m} \times 1330 \mu\text{m}$  (Table S8). Mean values are displayed with error bars indicating standard deviation from triplicate measurements, linear fit (solid line) demonstrates a high correlation with a Pearson  $R$  of 0.84. Source data for **A–D** and **F** is provided within the Source Data file.



**Fig. 3 | Multiple parameter analysis to evaluate property-activity relationship between physicochemical and bioinformatic descriptors and infectivity enhancement.** **A** Four-dimensional plot visualizes that a positive zeta-potential (Table S5), high hydrophobicity (Table S4), and high count rate of scattered light (Table S5) are corresponding to infectivity enhancement (color coded  $\text{Log}_{10}$  Infection Rel. EF-C, Table S5). **B** Multiple linear regression (constrained linear least-squares problems solution) with the equation:  $y = A_0 + A_1 \cdot x_1 + A_2 \cdot x_2 + [\dots] + A_5 \cdot x_5$  with best performing single parameters of 163 peptides based on EF-C

(Fig. S5, Table S4, S5):  $\text{Log}_{10}$  Count Rate, zeta-potential, hydrophobicity, ThT-activity, fibril formation,  $\beta$ -sheet and  $\text{Log}_{10}$  Infection Rel EF-C as dependent target variable show an improved linear correlation by 22% compared to the best correlating feature zeta-potential. Linear Fit with Pearson correlation factor  $R$  is 0.82. The found relationship is  $\text{Log}_{10}$  Infect Rel EF-C =  $-2.33462 + 0.02128$  (zeta-potential) +  $0.29879$  ( $\text{Log}_{10}$  Count Rate) +  $0.27355$  (fibril formation) +  $0.26241$  (hydrophobicity) +  $0.10744$  (ThT-activity) +  $0.00356$  ( $\beta$ -sheet). Source data for Fig. 3 A, B is provided within the Source Data file.

relative to EF-C shows indeed the strongest linear correlation ( $R = 0.67$ , Fig. 2C, Fig. S5, Table S5). Another comparably high correlating property is microscopic aggregation of the peptides, which was measured via the count rate of scattered light during zeta-potential measurement ( $\text{Log}_{10}$  Count Rate,  $R = 0.65$  Fig. 2D, Fig. S5, Table S5, Table S8). The timescale for the aggregation of peptides is within seconds as can be observed via TEM and turbidity from light scattering (Fig. S4). Although there is no suitable model for correlating the scattering intensity of light with the size of PNFs (SI Section 1.1), our data shows that the logarithmic number of  $\mu\text{m}$ -sized aggregated fibrils determined by fluorescence microscopy (Fig. 2E, Fig. S32) correlates linearly with the logarithmic scattering intensity (Fig. 2F, Table S8), which further validates the relationship between the microscopic aggregation and infectivity.

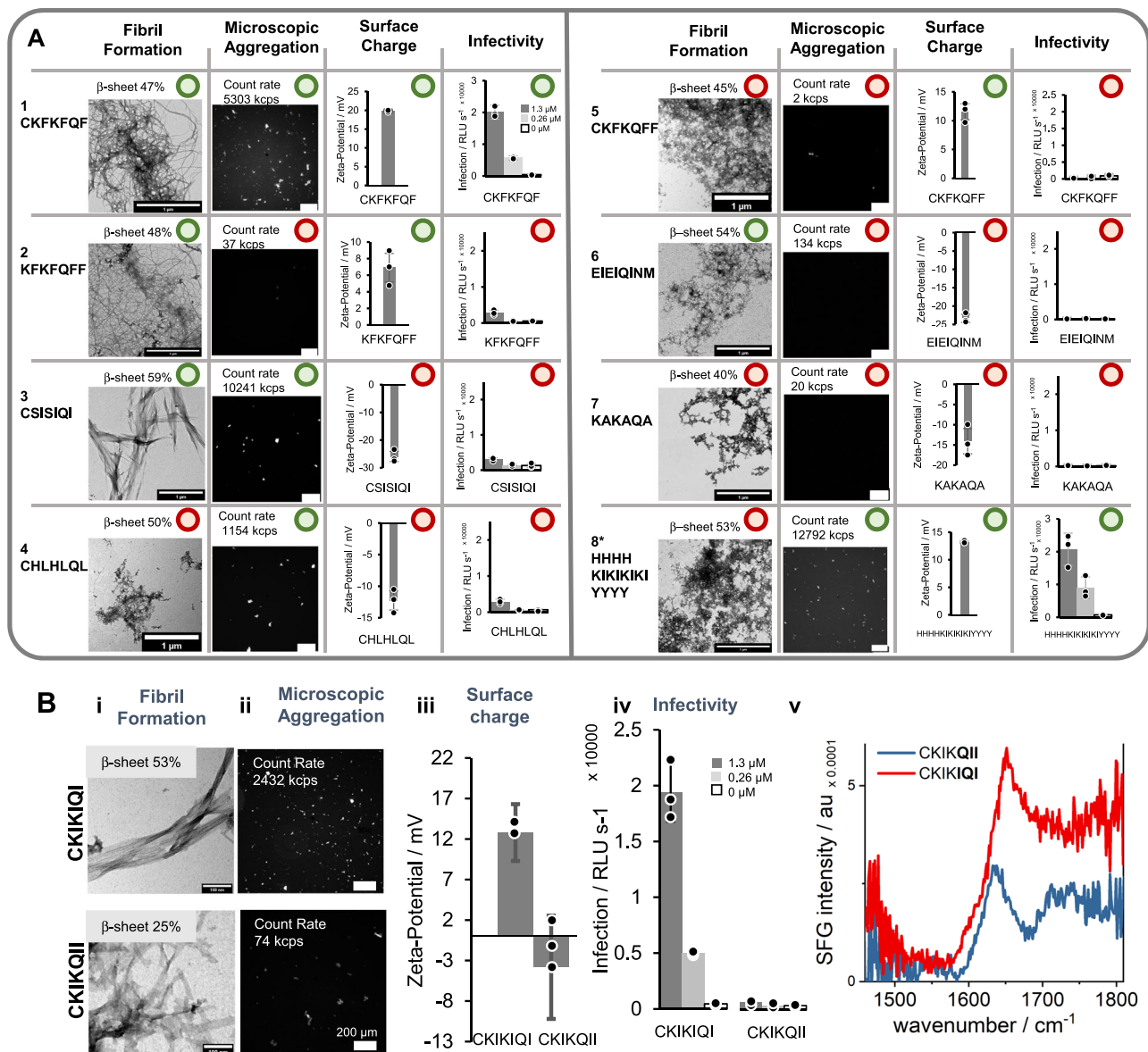
Despite the high correlation of zeta-potential and count rate with infectivity enhancement (Fig. 2C, D, Table S5), some data points deviate from the regression line. Moreover, not all bioactive peptides also form fibrils—for example, 7 infectivity-enhancing peptides from the library (Fig. 2A) form  $\beta$ -sheet rich amorphous aggregates instead. Interestingly, these peptides show a high calculated hydrophobicity due to large amounts of tryptophan (W) and tyrosine (Y) residues (Fig. S16), which indicates that multiple parameters are codependent of each other and cannot be covered from single parameter models to describe the property-activity relationship sufficiently.

### Synergistic effect of hydrophobicity, charge, and microscopic aggregation

We hypothesized that several conditions must be fulfilled to observe peptide-virus-cell interactions, which ultimately leads to enhanced viral transduction. So, in a next step, we considered multiple descriptors at once to identify the relevant physicochemical properties. Plotting the best correlating continuous features (zeta-potential, count rate, hydrophobicity) against  $\text{Log}_{10}$  infection relative to EF-C directly visualizes that a positive zeta-potential, high count rate and high hydrophobicity are coinciding with infectivity enhancement (Fig. 3A, Table S5).

In fact, applying multiple linear regression on the best correlating features revealed a superior Pearson correlation parameter ( $R = 0.82$ ) compared to the single features alone, an increase by 22% compared to best correlating single feature zeta-potential ( $R = 0.67$ ) (Fig. 3B, Table S5). A strong effect of multi-parameter considerations is observed for the  $\beta$ -sheet content, which was previously identified as a critical feature for infectivity in a small data set<sup>26</sup>. In our extended peptide library, we observe that the presence of  $\beta$ -sheet rich structures is crucial for activity only if we apply multi-parameter regression ( $R = 0.91$ , Fig. S10D, E) but not from single-parameter correlations ( $R = 0.29$ , Fig. S5, Table S5). This is because peptides with similar  $\beta$ -sheet content are not enhancing transduction efficiency to the same extent (Fig. S7) because they either do not show positive surface charge, do not form fibrils or do not aggregate (Fig. S10A–C).

Applying multi-parameter regression has its strongest impact on correlation of the bioinformatic parameters. For example, the highest single-parameter bioinformatic correlation can be found for hydrophobicity ( $H_{\text{fauchere}}$ ,  $R = 0.40$ , Fig. S5). However, in combination with all other calculated descriptors the correlation coefficient increases by roughly 70% (Fig. S11,  $R = 0.67$ , Table S4). The advantage in applying multiple linear regression is, that each contribution is weighted and reveals the importance of combined properties for the activity. In this way a statistically significant ( $p < 0.05$ ) correlation between zeta-potential, count rate, fibril formation, and infectivity enhancement (Fig. S10D) for the experimental parameters and hydrophobicity ( $H_{\text{fauchere}}$ ), aliphatic index (AI), hydrophobic moment index (HMI), instability index (II), graph shape index (GSI), upsilon steric parameter (USP), polarizability (P) and normalized van der Waals Volume (Norm vdW Vol.) (Fig. S11) for the bioinformatic parameters are identified. We propose that next to hydrophobicity, also other global peptide features (P, AI, HMI, USP, Norm. vdW Vol., GSI and II) correlate significantly with infectivity enhancement ( $p < 0.05$ ) but with a comparably small correlation coefficient (Figs. S5 S10, and S11). These parameters calculate different aspects of the amino acid side chain hydrophobicity, charge, sterical demand, and the sequence amphiphilicity (SI Section 1.2).



**Fig. 4 | Evaluation of global property-activity relationship for infectivity enhancement with selected peptides as case studies to gain insight into the mode of action.** Internal order is evaluated by TEM (Fig. S33, Table S5, fibril formation) and FT-IR (Fig. S34, Table S5 β-sheet content), surface charge by zeta-potential and infectivity is shown in absolute infection rates of peptides at 1.3, 0.26 and 0 μM (Table S5). All displayed error bars indicate standard deviation from triplicate measurements. **A** Overview of all possible combinations of physicochemical features, which correlate most strongly with infectivity enhancement with examples for each possible combination. 1 The majority of peptides (75%), which enhance infectivity have high internal order, aggregate into μm-sized aggregates, and have a positive zeta-potential, as shown exemplarily for the peptide CKFKFQF. 2 Without microscopic aggregation there is no infectivity enhancement even though other prerequisites are fulfilled such as fibril formation and positive zeta-potential shown for KFKFQFF. 3 Peptide fibrils which aggregate to clusters but have a negative surface charge are not enhancing viral infection, shown for CSISIQI.

However, if the peptides are highly hydrophobic slight infection enhancement can be observed (e.g. for EIEIQINMWQ, Fig. S15). Peptides which do not fulfill at least two of the features (assembly, aggregation, charge) cannot enhance infectivity as exemplarily shown for 4 (CHLHLQL or CKFKFQF with EGCG (Fig. S21), 5 CKFKQFF 6 EIEIQINM, 7 KAKAQA. 8\* Peptides with a high amount of hydrophobic amino acids (W, Y) can enhance transduction although they are not assembling into fibrils if they form β-sheet rich, positively charged aggregates e.g. shown for HHHHKIKIKI-KIYYYY. **B** Comparison of physicochemical properties and activity of CKIKIQI and CKIKQII. i TEM micrographs and β-sheet content determined by FT-IR (Fig. S34), scale bar 1 μm. ii Fluorescence microscopy of ThT-stained peptide fibrils with their respective light-scattering count rate (Table S5), scale bar 200 μm. iii Zeta-potential iv Absolute infection rates of peptides at 1.3, 0.26 and 0 μM concentration. v SFG spectroscopy at the amide I and II region. Source data for **A**, **B** is provided within the Source Data file.

### A differentiated perspective on the mode of action

To showcase the underlying codependences of the internal order (β-sheet and fibril formation), microscopic aggregation, and zeta-potential in the observed property-activity relationship, we exemplarily selected individual peptides (Fig. 4A, B, SI Section 6).

The most prevalent case (38% of the library, 75% of all active peptides, Table S5) for infectivity enhancement are peptides

assembling into fibrils with β-sheet secondary structure, which further aggregate into positively charged μm-sized aggregates (for example CKFKFQF, Fig. 4A-1).

The importance of these parameters is shown by the lack of activity when removing one of the properties (Fig. 4A). For example, if peptides assemble into positively charged fibrils but do not aggregate into μm-sized clusters (e.g. KFKFQFF, Fig. 4A-2), no strong infectivity

enhancement can be observed. In our library, cases where an assembly into fibrils but no aggregation into  $\mu\text{m}$ -sized particles occur, are mostly observed for peptides, which do not contain *N*-terminal cysteine at the first position (Fig. S13, SI Section 6.1). For example, the peptides CKIKIQI and KIKIQIC only differ in the position of cysteine within their sequence, but CKIKIQI shows stronger aggregation and infectivity enhancement than KIKIQIC (Fig. S14). In contrast to lysine, cysteine is an amino acid with a hydrophobic residue, and might increase the surface hydrophobicity of the fibrillar structure when positioned at the *N*-terminus. The thiol group of the cysteine side chain can further form disulfide bonds, which might increase inter-fibril attachment and aggregation.

Small changes in molecular sequence can induce a strong change in microscopic aggregation properties as observed for example for CKIKQII and CKIKIQI. Sum-frequency generation (SFG) measurements can give information on water interaction of fibrils, which is dependent on surface hydration and hydrophobicity properties of fibrils<sup>42,43</sup>. To investigate changes in water-fibril interaction the peptides CKIKIQI and CKIKQII were selected as examples since they are composed of same amino acids in different order, form fibrils and differ significantly in their aggregation behavior (Fig. 4B). The SFG measurements show that while both peptides accumulate at the air/water interface, the signal intensity of CKIKIQI is higher than CKIKQII. This indicates that CKIKIQI is more ordered or may be present at higher concentrations at the air/water interface (Fig. 4B). Both effects can be traced back to a higher surface hydrophobicity and lateral interaction of fibrils in  $\mu\text{m}$ -sized aggregates, which in turn are a prerequisite for infectivity enhancement.

Noteworthy, not every parameter in this relationship must be fulfilled to observe any infectivity enhancement. For example, while the majority (90%) of infectivity-enhancing peptides (Infection relative to EF-C > 10%) forms fibrils, some of the active peptides show non-fibrillar morphologies (Fig. 2A). These peptides contain a high amount (> 20%) of hydrophobic amino acids such as W or Y and form amorphous amyloid aggregates with high  $\beta$ -sheet content (Table S5). Additionally, they fulfill other prerequisites such as positive zeta-potential and microscopic aggregation (Fig. S16). Fibril formation is strongly connected to aggregation as it likely facilitates entanglement but appears not to be an indispensable prerequisite for infectivity enhancement (Figs. 4A–8, S16, and S19).

As already extracted from the global correlation, the surface charge plays a key role. Only weak infectivity enhancement is observed for peptides with negative zeta-potential even if the fibrils with high  $\beta$ -sheet content are aggregating (Fig. 4A–3, Table S5). However, as it also can be extracted from the single parameter correlation with zeta-potential (Fig. 2A, Table S5), five peptides with negative zeta-potential show a moderate infectivity enhancement larger than 10% Rel EF-C ( $\text{Log}_{10}$  Infect > -1) (Fig. S15). We hypothesize that these peptide fibrils can slightly enhance infectivity because of their extraordinarily high hydrophobicity (> 0.6) and microscopic aggregation of fibrils, which favor interactions with viruses and cells despite the electrostatic repulsion. The importance of considering hydrophobic interactions in addition to electrostatic attraction is further emphasized by a recent report from us, in which we applied machine learning approach to target different sequence spaces<sup>44</sup>. An overall lower hydrophobicity—be it by substituting phenylalanine (F) or isoleucine (I) amino acids with alanine (A) or leaving out hydrophobic amino acids—results in loss of activity (SI Section 6.3–6.4, Figs. S17–S20), further emphasizing its impact.

Activity is also lost, when more than one of the criteria (fibril formation, positive zeta-potential, aggregation) we identified is not fulfilled (Fig. 4A–4–7). Peptides that only show  $\mu\text{m}$ -sized aggregation (Fig. 4A–4), positive zeta-potential (Fig. 4A–5), or  $\beta$ -sheet rich-fibrils (Fig. 4A–6) do not enhance viral infectivity. To illustrate that the

activity of a peptide can be solely traced back to physicochemical properties while the sequence stays same, we added a chaotropic substance (Fig. S21)<sup>45</sup>. Adding epigallocatechin gallate (EGCG) to CKFKFQF drastically decreases infectivity enhancement,  $\beta$ -sheet, and surface charge while microscopic aggregation still takes place (Fig. S21). This once more demonstrates that positive surface charge of microscopically large aggregates with a  $\beta$ -sheet structure are prerequisites for activity.

In summary, our findings show that higher hydrophobicity of assembled peptides drives microscopic aggregation, which is also a prerequisite for virus-cell membrane interaction and colocalization additional to electrostatic attraction. Thereby, we conclude that multiple conditions including (I) a positive surface charge, (II) ordered structure, (III) microscopic aggregation are required for highly efficient retroviral transduction enhancement (Fig. 5).

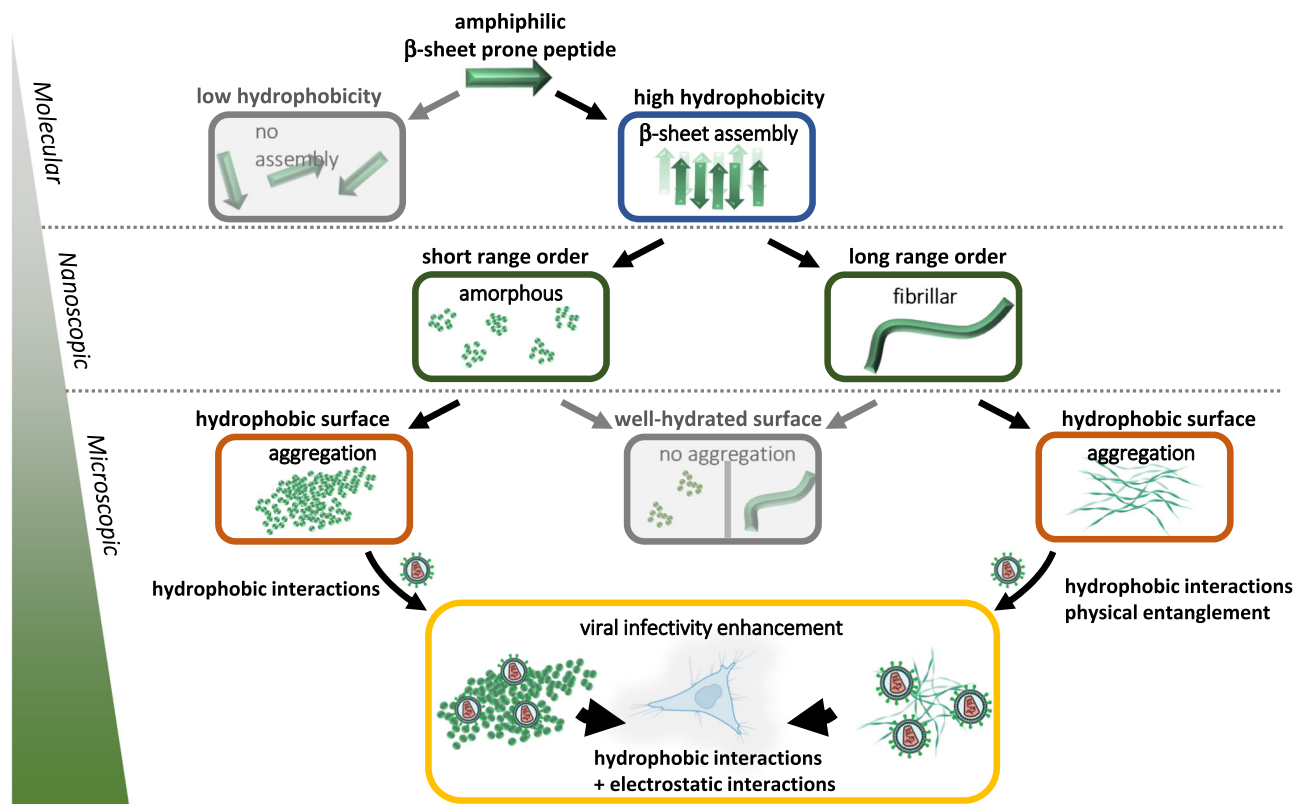
### Design rules: alternating, amphiphilic sequences patterns

We wondered if a sequence pattern exists, that underlies the identified key physicochemical properties and results in infectivity enhancement. To find common design rules which are preserved in active peptides, the sequences in the library were simplified by coarse-graining the amino acids according to their side chain hydrophathy (Hydrophilic P, Hydrophobic H) and charge (-, 0, +). The distribution of recurring patterns in highly active peptides (Infect. Rel. EF-C > 70%) was counted (SI Section 7, Fig. S22A)<sup>46</sup>. The majority (55%) of the active sequences and 44 peptides in the library (163 peptides, Table S3) were composed of alternating amphiphilic sequences  $\text{P}^+\text{H}^0\text{P}^0\text{H}^0\text{P}^0\text{H}^0\text{P}^0$  or  $\text{P}^+\text{H}^0\text{P}^+\text{H}^0\text{P}^0\text{H}^0\text{P}^0$  (Fig. 6A).

Interestingly, an alternating amphiphilic sequence does not guarantee high transduction efficiency: 8 out of 44 peptides have an alternating amphiphilic sequence but do not show transduction enhancement. This is likely due to their low sequence hydrophobicity (Fig. S23A). To systematically study the impact of hydrophobicity on activity, amino acids from highly active sequences (W, M, N, I, C, Q, K, F and Y, Fig. S22B, Fig. 6A, B) and from weakly active peptides (S, G, R, H, T, V, P, A and L, Figs. S22B and S23B)<sup>46</sup> were used to representatively create peptides which share the same amphiphilic pattern ( $\text{P}^+\text{H}^0\text{P}^0\text{H}^0\text{P}^0\text{H}^0\text{P}^0$  or  $\text{P}^+\text{H}^0\text{P}^+\text{H}^0\text{P}^0\text{H}^0\text{P}^0$ ) and test their infectivity enhancement. Noteworthy, we observed that most of the herein-found amino acids from highly active sequences and weakly active sequences have been known to promote and break  $\beta$ -sheet fibril formation, respectively (SI Section 7).

This approach was exceptionally effective in generating new and active sequences: Eight of the 10 newly created peptides composed of the identified pattern and amino acids prevalent in highly active sequences show infectivity enhancement (Fig. 6B, Table S6). Out of these only the W-rich peptide RFWFWFW does not form fibrils (Figs. 6B and S25) but fulfills all other important requirements for infectivity enhancement—especially a high hydrophobicity (Fig. 6B). Reducing the number of W by one decreases hydrophobicity (1.58  $\rightarrow$  1.18) and results in loss of the infectivity enhancing properties as shown for RINFWFW, Fig. 6B and E). A high amount of W is related to more disordered fibrils but increased activity of the peptides—as also observed for other examples such as KIKIWIW  $\rightarrow$  KIWIWIW (Fig. S25) or C-terminal extension with W (Fig. S16 and S20). These examples once more emphasize the codependences of multiple descriptors for infectivity-enhancing peptides and enable the creation of new active sequences in a reliable fashion. Strikingly, in this way a cysteine- and glutamine-rich peptide KCQCQCQ was found, which showed activity in the range of EF-C while being half the length of EF-C (Fig. 6B, Table S6).

The active peptides all share the identified key-properties for infectivity enhancement, which emphasizes the importance of an



**Fig. 5 | Schematic overview of proposed mode of action over multiple scales from molecular to microscopic level.** Molecular—Amphiphilic peptides with certain hydrophobicity can form  $\beta$ -sheet rich structures. Nanoscopic—Depending on the hydrophobicity of the  $\beta$ -sheet structures, short-range ordered amorphous

aggregates or long-range ordered fibrils can form and further aggregate into  $\mu$ m-sized particles. Microscopic— $\beta$ -sheet rich aggregates can bind viruses and interact with cellular membrane by electrostatic and hydrophobic interactions. Created with BioRender.com.

amphiphilic sequence pattern with sufficiently hydrophobic amino acids and cationic amino acids for aggregation of self-assembled peptide fibrils.

### Key properties for infectivity enhancement translate to functional and pathogenic amyloids

The fact that aggregation is necessary to observe transduction enhancement may shed light on amyloids in pathogenic contexts. In Nature, microscopic aggregation of amyloidal fibrils is well known as so-called amyloid-plaques, which results in deposition of non-soluble proteins in neurodegenerative diseases as well as facilitating adhesion of bacterial biofilms on substrates<sup>14</sup>. Amyloid plaques from amyloid A $\beta$  occur in several neurodegenerative diseases and may have a connection to neuroinflammation processes. It is speculated that amyloid A $\beta$  is produced by the immune system to combat infections<sup>47,48</sup>. In a downstream process, these amyloid fibrils might form micron-sized aggregates (amyloid plaques) that facilitate infections, leading to stronger infectious reactions—which eventually produce more amyloid fibrils in a vicious cycle. We hypothesized that the property-activity findings from the EF-C library can be translated to other amyloid peptides from functional and pathogenic contexts. To test key-properties for infectivity enhancement, we selected amyloidal peptides from protein databases and literature reports based on published data from transmission electron micrographs and ThT-fluorescence assays (Table S1).

Three of the 24 peptides that were selected, namely ISKLEYSNFSVRY, LANWMCLAKW and VHDCVNITIK, showed infectivity enhancement of retroviral transduction in the range of 30–40% relative to EF-C (Fig. 7A, Table S7). These newly found transduction-enhancing peptides have a positive zeta-potential (Fig. 7B, Table S7, +9.3 mV, +14.7 mV, +9.3 mV, respectively), microscopically aggregate

(Fig. 7B, Fig. S29-B, Table S7), assemble into fibrils (Figs. S36, S37, and 7C–E) with a high  $\beta$ -sheet amount (Figs. S30 and S39, Table S7) and are ThT-active (Fig. S29–A). We studied the interaction of fibrils with HeLa cells by applying confocal fluorescent laser scanning microscopy. The interaction of the amyloidal fibrils with HeLa cells is clearly visible in Fig. 7C–E and shows that these newly discovered infectivity-enhancing peptides act in a comparable fashion as EF-C by associating with cellular membranes and facilitating viral uptake through colocalization<sup>12</sup>.

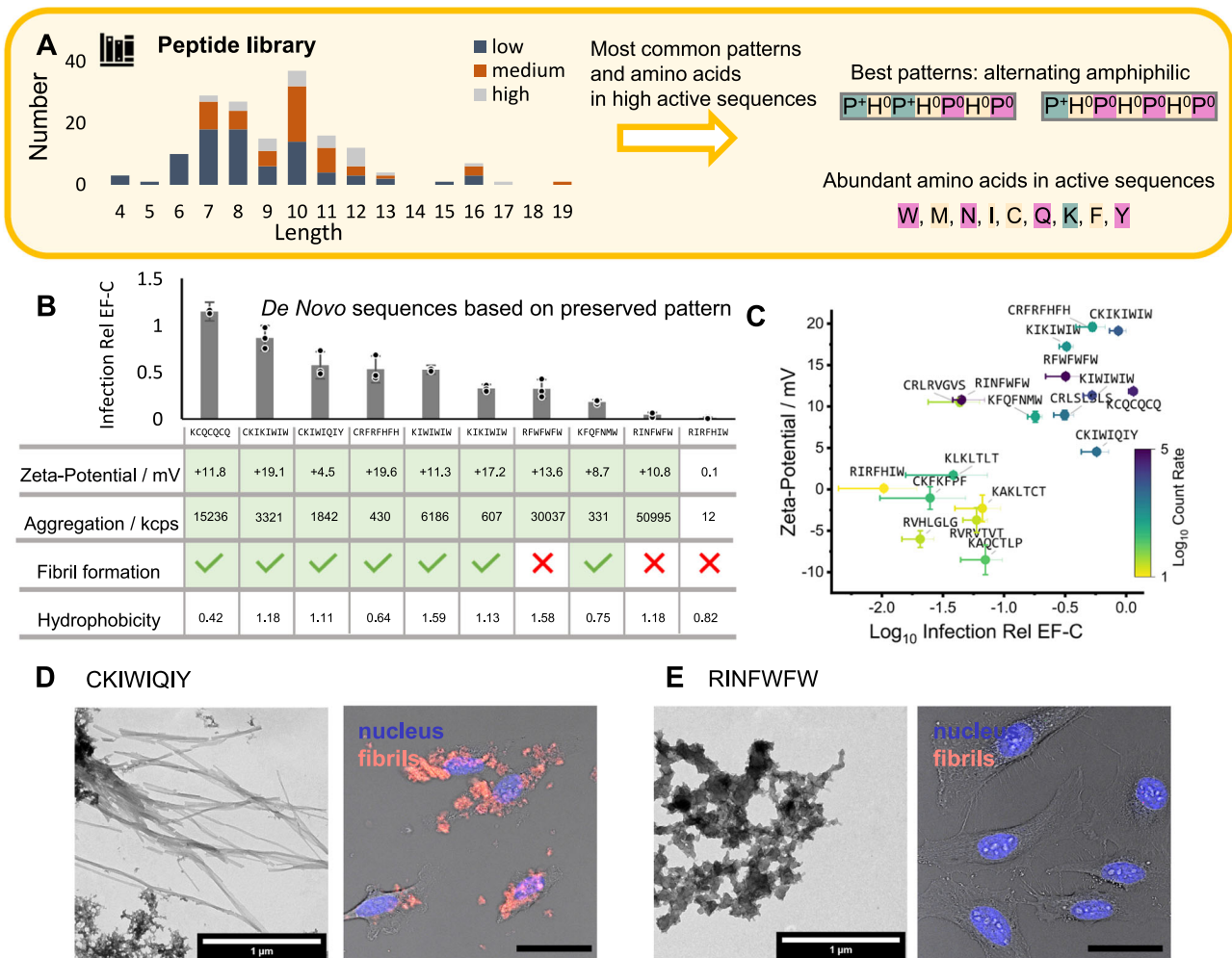
Interestingly, four fibril-forming peptides IKFLSVN, GYVIK, RQGNINIVA, and MKVIFLKDVKG, that are derived from peroxidase III<sup>49</sup>, serum amyloid P<sup>50</sup>, insects' chorion<sup>51</sup>, and ribosomal protein<sup>52</sup>, respectively, have positive zeta-potential in a similar range, but do not enhance infectivity. This is because these peptides do not aggregate to microscopic clusters as visible from low light scattering count rates (Figs. 7B and S29–B) and microscopy data (Fig. S31).

In summary, studying endogenous amyloidal peptides confirms the identified key-properties for transduction enhancement and further emphasizes the importance of microscopic fibrillar aggregates beside the electrostatic interactions to mediate binding of the positively charged fibrils to the negatively charged cell surface.

### Discussion

In this study, we aimed to understand the mode of action of short peptides acting as adjuvants to increase viral infectivity. We hypothesized that certain physicochemical properties are responsible for the infection-enhancing properties of peptides. To test this, we applied a data-mining approach that includes the creation of a suitable dataset, the evaluation of corresponding data, and interpretation of found relationships. Therefore, we created a peptide library based on known





**Fig. 6 | Preserved pattern in infectivity enhancing peptides can be applied to create new active sequences confirming property-activity relationship found for original library.** **A** Coarse graining infectivity enhancement in low (<10%) medium (10-70%) and high (>70%, Infection Rel EF-C) and coarse graining amino acids according to their hydrophilicity (H = hydrophobic; P = hydrophilic) and side chain charge at pH 7.4 (+ = positive, - = negative, 0 = neutral) enables identification of most abundant sequence pattern in highly active peptides (Table S5) that is an alternating amphiphilic motif  $P^+H^0P^+H^0P^0H^0P^0$  and  $P^+H^0P^0H^0P^0H^0P^0$ . Source code is available under the following link <https://doi.org/10.5281/zenodo.8079726>.

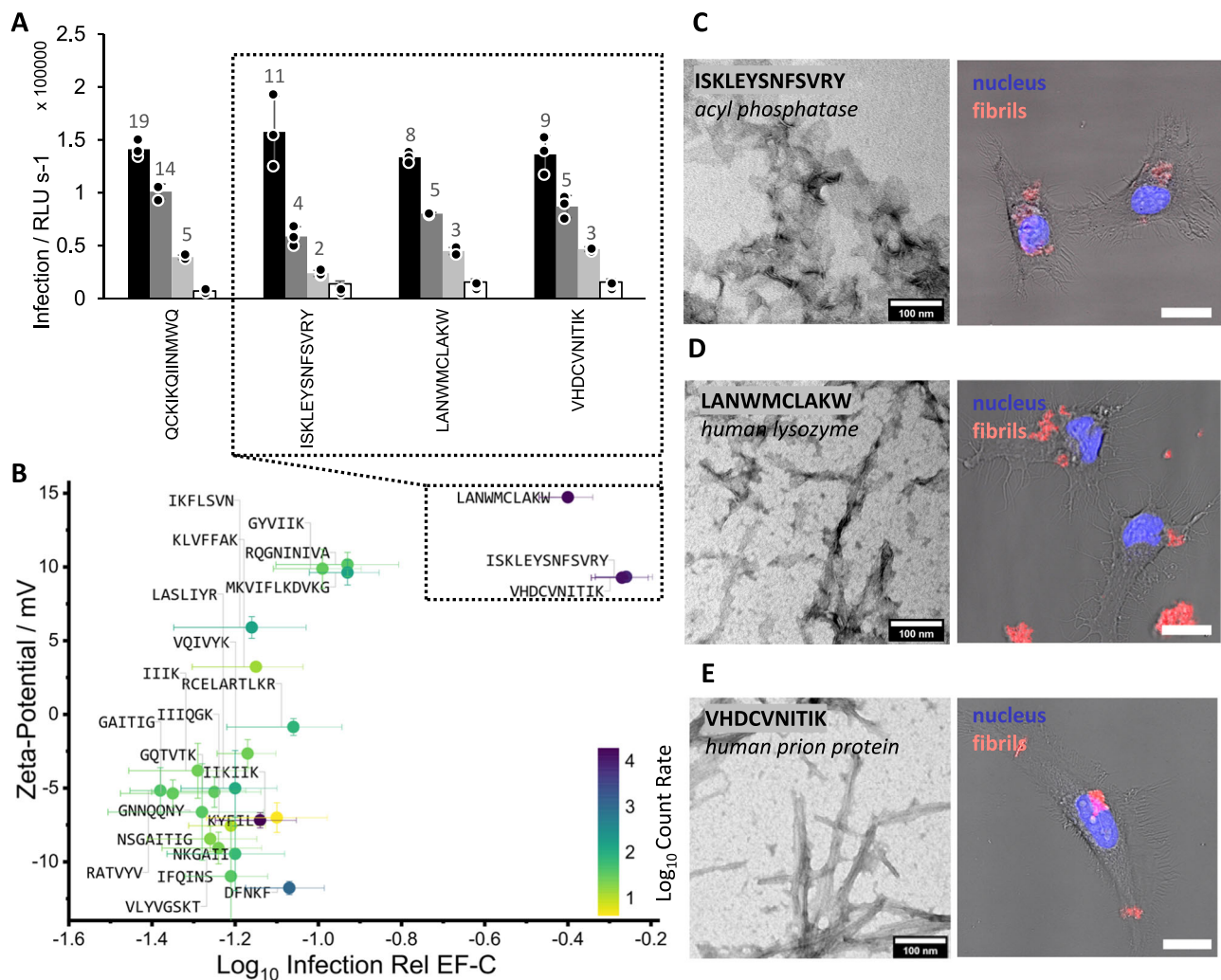
**B** New peptides are created by combining the preserved pattern  $P^+H^0P^+H^0P^0H^0P^0$  with the identified best-performing amino acids. 8 of 10 newly created peptides are enhancing infectivity (>10% Infection Rel EF-C at 1.3 μM, Table S6) and overall confirm the identified property-activity relationship. Error bars indicate standard deviation from triplicate measurements. **C** Graphical visualization of the most important property-activity relationship zeta-potential plotted versus Log<sub>10</sub> Infection relative to EF-C. The color code represents the Log<sub>10</sub> Count Rate from light scattering, which can be traced back to microscopic aggregation (Table S6). Mean

values are displayed with error bars indicating standard deviation from triplicate measurements. **D** Selected representative example for fibril morphology (TEM, Fig. S25, scale bar 1 μm) and cell interaction (confocal microscopy, scale bar 50 μm) of an active peptide (CKIWIQIY), which is based on the preserved pattern. Representative confocal image shows merged measurement from transmission, Proteostat (red, fibrils) and Hoechst (blue, cell nuclei) channels. **E** Selected representative example for amorphous morphology (TEM, scale bar 1 μm) and no cell interaction (confocal microscopy, scale bar 50 μm) of an inactive peptide (RINFWFW), which is based on the preserved pattern. Confocal image shows merged measurement from transmission, Proteostat (red, fibrils) and Hoechst (blue, cell nuclei) channels. CKIWIQIY and RINFWFW were both susceptible for amyloid sensitive dye staining (Fig. S24), but RINFWFW does not attach to cells in accordance with its low activity. TEM measurements were conducted once with at least three microscopy images recorded for each peptide sample, confocal images have been prepared in biological triplicates and recorded at least at three different sites. Source data for **A–C** is provided within the Source Data file.

infectivity-enhancing peptide EF-C. This enables tracing back differences in activity to specific physicochemical properties because sequences stay relatively similar.

Since the interactions between fibrils, viruses and cells occur on multiple length scales we focus on applying complementary methods to characterize multiple physicochemical properties for a large number of peptides, thus circumventing scale limitations of individual characterization methods. For example, TEM is a qualitative high-resolution method for structural characterization on a nanometer scale, but it cannot quantify reliably fibril polymorphism and it lacks information on intermolecular interactions, such as β-sheet content, which was therefore determined via FT-IR spectroscopy. It is

well-known that amyloid polymorphism can greatly influence its bioactivity<sup>53,54</sup>. However, the systematic study of different fibril morphologies is outside the scope of this study. To detect the formation of β-sheet rich fibrils in bulk, we utilized the amyloid-sensitive ThT dye. However, it should be noted that ThT-fluorescence has been associated with false-positive results (SI Section 1.1). All these methods cannot give information on microscopic aggregation properties, for which we applied microscopy and light scattering. Bioinformatic properties gave information for elucidating general effects of structural changes on infectivity enhancement. Multiparameter analyses enabled us to correlate multi-scale dependencies of interaction of materials with cells and viruses.



**Fig. 7 | Infectivity enhancing properties of selected amyloid peptides from pathogenic and functional origin.** **A** Infection rates of EF-C, and peptides ISKLEYSNFSVRY, LANWMCLAKW, VHDCVNITIK at 6.5, 1.3, 0.26, 0  $\mu$ M concentration. Numbers at the end of bars indicate n-fold infectivity enhancement compared to virus-only infectivity (0  $\mu$ M) (Table S7). Error bars indicate standard deviation from triplicate measurements. **B** Zeta-potential plotted against Log<sub>10</sub> Infection Rel EF-C. Color indicates Log<sub>10</sub> Count Rate by scattered light. Three peptides (box) that show identified important physicochemical properties that have positive zeta-potential, show fibril formation, and aggregation are enhancing infectivity (Table S7). Mean values are displayed with error bars indicating standard deviation from triplicate

measurements. TEM micrograph (1 mg/mL, PBS, Fig. S36, S37 scale bar 100 nm) and confocal microscopy (scale bar 20  $\mu$ m) of acyl phosphatase derived peptide ISKLEYSNFSVRY (**C**), human lysozyme derived peptide LANWMCLAKW (**D**) and human prion protein derived peptide VHDCVNITIK (**E**). 20  $\mu$ g/mL preassembled peptides (stained with Proteostat, red) were incubated with Henrietta Lacks (HeLa) cells (nucleus stained with Hoechst, blue). For single channels see Fig. S28. TEM measurements were conducted once with at least three microscopy images recorded for each peptide sample, confocal images have been prepared in biological triplicates and recorded at least at three different sites. Source data for **A, B** is provided within the Source Data file.

We observed that the interactions of peptides with cells and viruses are mostly driven by three codependent properties: (I) ordered structure (II) microscopic aggregation and (III) positive surface charge. While the importance of zeta-potential and fibril formation was already suggested for EF-C PNF in previous reports<sup>12,26,55</sup>, microscopic aggregation—which we found to correlate strongly with infectivity—was not systematically studied nor identified as a necessary feature before.

We hypothesize that  $\mu$ m-sized aggregates can accelerate infectivity by physical entanglement with cellular protrusions, which are actively transporting the fibril/virus complexes to the cell surface, where virions can enter the cells<sup>55</sup>. Interestingly, for aggregates larger than 10  $\mu$ m<sup>2</sup> it is not the size, but rather the number of aggregates correlating with infectivity enhancement. Fibrils aggregating to few very large aggregates might simply not offer enough particles for virus-cell interactions or may be too big to get efficiently engaged by protrusions. In contrast, fibrils aggregating to many small  $\mu$ m-sized aggregates provide a high number of dispersed particles for viruses to

attach to and subsequently interact with cell membranes. Noteworthy, there are no obvious changes in fibril formation and aggregation properties of the fibrils in the presence of virions, based on our own data and earlier reports<sup>12,25</sup>. The importance of well-distributed aggregates rather than aggregate size is emphasized by findings from cumulative surface area of total aggregates, where we do not observe any difference between the surface area of “active” aggregates and “inactive” aggregates (SI Section 4.3, Fig. S8). Interestingly, the formation of certain aggregate size and number appears to be an intrinsic feature of the non-covalent self-assembly properties of the molecular sequence, e.g. after mechanical downsizing aggregates by pressure or ultrasonication, their original size is recovered immediately in solution (Fig. S9), preventing size-dependent activity investigations of the same peptide.

The strong correlation observed for  $\mu$ m-sized aggregation further emphasizes a fundamental common principle for infectivity enhancement of peptides: The hydrophobic surface properties may

not only be a driving force for peptides to form fibrils<sup>56</sup>, increase fibril-fibril association<sup>57,58</sup> which leads to  $\mu\text{m}$ -sized aggregates, but also a prerequisite to bind and capture viral particles. An interplay between hydrophobicity, electric charge, and anions in buffer is well-known to control the formation of hierarchical structures, such as amorphous aggregates or amyloid fibrils<sup>59–61</sup>. Hydrophobicity as a prerequisite for efficient gene delivery was recently shown for polyplexes and polymers for non-viral applications<sup>62,63</sup>. The interaction of amphiphilic peptides with cellular lipid membranes was reported to be facilitated by electrostatic and hydrophobic interactions of the peptides<sup>64</sup> and multivalent binding to the amyloid surface reduced binding affinity to cells and infectivity of semen-derived enhancer of viral infection (SEVI) amyloid fibrils<sup>65</sup>. Thus, we conclude that, a high hydrophobicity can facilitate interaction of peptide fibrils with cell membranes and thereby increase infectivity enhancement as shown systematically in our peptide library.

Although, an alternating amphiphilic sequence pattern with hydrophobic and cationic amino acids is well-known for amyloids<sup>66</sup>, and it is found in our library prevalent for active sequences, activity is not limited to these sequence motifs as long as the identified physicochemical properties are fulfilled. For example, we found the same relationship between microscopic aggregation and infectivity enhancement for structurally different endogenous short self-assembling amyloid peptides as exemplified by the three newly discovered highly effective peptides ISKLEYSNFSVRY, LANWMCLAKW, and VHDCVNITIK. Interestingly, these peptides are derived from different biological contexts. While the sequences ISKLEYSNFSVRY and LANWMCLAKW occur naturally in acyl phosphatase and human lysozyme, the peptide VHDCVNITIK is a sequence from the pathogenic human prion protein<sup>52</sup>. It is important to note that while these protein fragments can enhance infectivity, their physicochemical properties cannot be simply extrapolated to the parent protein, as shown exemplary for lysozyme and LANWMCLAKW (Fig. S40). However, it is not recommended to use peptides derived from the original pathogenic origin for therapeutic purposes because they may act as potential seeds for disease-related aggregation events. Amyloid fibrils that aggregate expose fewer terminal ends that are known to disrupt cellular membranes<sup>67</sup>. The fibrillar aggregates reported in this study do not show toxicity but good cell viability for HeLa cells (Fig. S2) as well as other cell types for similar fibrils, as reported elsewhere<sup>32,68,69</sup>. Microscopic aggregation has been previously observed in the non-diseased state of endogenous semen protein prostatic acid phosphatase<sup>70</sup>, which suggests a broader biological significance of micrometer-sized aggregation for amyloids beyond simple disease-related associations.

Pioneering works have put much efforts in identifying amyloid fiber evolution mechanisms on a molecular and nm-size level, e.g. for oligo and protofilament formation<sup>71–73</sup>. On the inter-fiber mesoscale lateral aggregation was traced back H bonds<sup>74</sup> and VdW polar interaction modes<sup>75</sup> and to hydrophobic interactions<sup>76,77</sup>. Interestingly, experiments and simulations showed that dynamic, flexible fibrils are less prone to align into multi-fiber strands<sup>37,78</sup>. However, the exact structural determinants for the  $\mu\text{m}$ -sized aggregation of amyloid fibrils still remain elusive, and the majority of studies, reporting on mechanistic insights of amyloid formation, focus on aggregation in the nm-length scales (fibril formation)<sup>79</sup> and rarely on  $\mu\text{m}$ -sized aggregation. One of the reasons for the lack of studies that deal with understanding the multi-scale aggregation phenomena stems from the difficulty of relating molecular information to microscopic properties, which is especially challenging for self-assembly processes occurring within short time frames. Our data indicates that high  $\beta$ -sheet content of amyloid peptides, and hydrophobicity are not only necessary for interaction with viral particles and cells but also favor aggregation of fibrils into  $\mu\text{m}$ -sized fibrillar aggregates. The higher surface hydrophobicity of CKIKIQI compared to CKIKQII which is observed from SFG measurements (Fig. 4B–v) may be

arising from a better intermolecular  $\beta$ -sheet packing of peptide monomers in the fibrils due to localization of the more hydrophobic amino acid isoleucine in the center of the peptide. Furthermore, a high propensity to form  $\beta$ -sheet structures does not necessarily lead to lateral aggregation of fibrils into microscopic aggregates<sup>52</sup>. For example, the sequence GNNQQNY forms  $\beta$ -sheet rich amyloid fibrils but with well hydrated external faces, which presumably hinder them to form microscopic aggregates (Table S7)<sup>75</sup>. The interplay between polar and non-polar interactions, so-called hydrophobic patches, at the fibril surface can determine the fate of ordered or less ordered fiber-fiber aggregation across length scales<sup>36,80</sup> and might explain why some amyloid fibrils form unordered large ( $> 10 \mu\text{m}^2$ ) aggregates while others stay isolated or form well-aligned bundles. Our data shows that regular arrangement of polar and non-polar amino acids facilitates stable  $\beta$ -sheet rich structures and SFG measurements indicate that hydrophobic surface interface dominates for microscopically aggregating peptides. Therefore, surface hydrophobicity likely is a key driving force for formation and  $\mu\text{m}$ -sized aggregation<sup>81</sup> of amyloid fibrils and an important characteristic besides charge-driven interactions to associate with viruses and cells<sup>82</sup>. The fewer water-accessible areas may result in higher fibril aggregation which may be at the same time the driving force for cell and virus interaction. However, structural evidence of these interactions is needed to support these hypotheses, which is beyond the scope of this study. Due to the lack of experimental methods to show the assembly dynamics and peptide-water interaction on very short timescales ( $\sim \text{ns}$ ) which is necessary to explain molecular dynamics, our current efforts focus on molecular simulations which may connect structural information to physicochemical properties underlying intermolecular interactions which eventually result into microscopic aggregation.

The specific factors that determine the ordered or unordered aggregation of micrometer-sized particles are currently unknown, but they are important not only for understanding biological activity but also for a wide range of diseases. We believe that multi-scale and multiparameter considerations in mechanistic origins of microscopic aggregation are not only important for studying pathogenesis, such as neurodegenerative diseases but also for infectivity enhancement and understanding possible connections between them.

This study represents the first investigation of the multi-scale structure-property-activity correlation of self-assembling amyloid peptides for in vitro retroviral transduction enhancement. In contrast to previous reports, we find that not the amyloid fibrils itself but rather their  $\mu\text{m}$ -sized aggregated states facilitate retroviral gene delivery. By applying a data-mining approach, we could perform a multi-parameter regression and unveil crucial co-dependencies in the peptide properties that govern their activity and act on length scales from nano to micrometers: Amphiphilic, hydrophobic sequences form microscopic aggregates with a high  $\beta$ -sheet content and a positively charged surface, all of that required for efficient retroviral transduction enhancement. The fibril-fibril aggregation and the fibril-virus interaction are both driven by the hydrophobicity and amphiphilicity of the peptide sequence. Importantly, we exploited these structure-property-activity findings to rationally create novel active sequences. Moreover, by successfully applying our findings to screen for active amyloid peptides from endogenous proteins, we validated the universal applicability of our study. In that way, our study contributes to a fundamental understanding on the emergence of bioactivity of microscopically aggregating amyloid fibrils.

## Methods

### Materials

Fmoc-protected amino acids and Wang resin were purchased from Novabiochem®. OymaPure®, *N*-ethyl-diisopropylamine, piperidine and trifluoroacetic acid were obtained from Carl Roth. Dimethylformamide (DMF for peptide synthesis), diethyl ether and dimethylsulfoxid was

purchased from Acros Organics. Acetonitrile was purchased from Fisher Scientific. Uranyl acetate and fluorecamine were purchased from Merck. Lysozyme was purchased from Amresco. PBS, Thioflavin T and  $\alpha$ -cyano-4-hydroxycinnamic acid were purchased from Sigma Aldrich. Proteostat® was purchased from Enzo Life Sciences. All chemicals are listed in Supplementary Data 1 and were used as received unless explicitly stated otherwise.

### Data-mining

The herein applied data-mining procedure covers data generation, processing, data extraction and evaluation, model development, prediction, and interpretation to evaluate the identified property–activity relationship.

The first step involved data generation, where 163 distinct peptides based on an infectivity-enhancing peptide EF-C (QCKIK-QIINMWQ) were synthesized (Table S2), six physicochemical and one biological properties were measured (see SI Section 1.1 and method section below). The collected data was then processed, as summarized in Table 1. Twelve bioinformatic properties including peptide net charge, hydrophobicity, isoelectric point, aliphatic index, hydrophobic moment index, boman index, instability index, graph shape index, upsilon steric parameter, smoothed upsilon steric parameter, polarizability, and normalized van der Waals volume were calculated via the “peptides” package in R (see SI Section 1.2)<sup>40</sup>.

The next step was extraction and evaluation of the collected data: For an initial screening every descriptor was summarized in a table with the software Excel (Version 2304 Build 16.0.16327.20200, Microsoft 365) and imported to the data-mining software Orange3 (Orange3-v.3.34.0, Anaconda) to find the best correlating descriptor combinations via the functions “scatter plot → find informative projections” and “pearson correlations”. The Pearson correlation coefficients for every descriptor were calculated via the function “correlation plot” of the software OriginPro 2021 v.9.8.0.200 (OriginLab). The property activity relationship was identified using the function “constraint multiple regression” analysis (constrained linear least-squares problems solution with the equation:  $y = A_0 + A_1 \cdot x_1 + A_2 \cdot x_2 + \dots + A_5 \cdot x_5$ ) of the software OriginPro 2021 v.9.8.0.200 (OriginLab). The model was evaluated using various statistical metrics such as the Pearson correlation coefficient and the statistically significant descriptors were identified by  $p < |t|$ ,  $t = 0.05$ .

The established linear regression model was validated with de novo peptides predicted from a pattern analysis technique and with existing peptides from a database screening.

The pattern analysis aims to find amino acid patterns and amino acids that are prevalent in high-active peptides (see SI Section 7). Briefly, amino acids were first classified into four categories based on their hydrophilicity and charge at pH 7.4 using the Kyte–Doolittle hydrophathy scale<sup>83</sup>. The infectivity enhancement was categorized in low (<4-fold virus only, <10% relative to EF-C), medium (10–70% relative to EF-C), high (>70% relative to EF-C). All possible 7-mer patterns were then generated from these coarse-grained amino acids and matched against existing patterns in the peptide library. To identify patterns that are more prevalent in high-active sequences compared to medium- or low-active sequences, an activity index of a pattern  $p$  ( $AI(p)$ , Eq. (1)) was calculated by subtracting the sum of its relative numbers in medium- and low-active sequences from its relative number in high-active sequences as in:

$$AI(p) = \frac{N_{\text{high}}(p)}{\sum N_{\text{high}}} - \frac{N_{\text{medium}}(p)}{\sum N_{\text{medium}}} - \frac{N_{\text{low}}(p)}{\sum N_{\text{low}}} \quad (1)$$

For the amino acid analysis, the absolute abundance of amino acids in one category (low, medium, high) in the whole peptide library was first determined. The relative abundance of amino acids “*Rel AA*” was then calculated by determining the number “*N*” of peptides

containing a certain amino acid “*AA*” relative to the total amount of peptides of one category (Eq. (2)).

$$Rel AA = \frac{N(\text{peptides containing } AA)_{\text{category}}}{\sum \text{peptide}_{\text{category}}} \quad (2)$$

All pattern and amino acid analyses were conducted using Python scripts. Source code is provided<sup>46</sup>.

The prediction and evaluation of new peptides were conducted by retranslating the coarse-grained best-performing patterns with best-performing amino acids and worst-performing amino acids to yield 10 peptide sequences designed to be active and 8 peptides designed to be inactive. To avoid human bias, the selection of amino acids from the pool of best-performing and worst-performing ones was conducted randomly.

For the data-base screening 538 short (<20 amino acids) peptides across various sources (176 peptides from PDB<sup>84</sup>, 260 peptides from a report<sup>52</sup> and 102 peptides from Waltz database<sup>85</sup>) were collected by their bioinformatic properties (hydrophobicity, isoelectric point, net charge, calculated the “peptides” package in R<sup>40</sup> v 4.0.3) and by any reported information on structure formation (TEM, FT-IR and ThT-fluorescence). A list of all considered peptides from the above-mentioned databases is summarized in Supplementary Data 2. Ultimately, 24 different peptides (SI Section 8) were selected based on fibril formation by TEM and preferably by net charge (>0), hydrophobicity (>0.2),  $\beta$ -sheet formation by ThT fluorescence.

### Solid-phase peptide synthesis and characterization

Peptide sequences were purchased from Phtd Peptides industrial Co. limited with purity of  $\geq 95\%$ . Some peptides were synthesized from us (marked with \* in Table S3) by using an automated microwave peptide synthesizer (CEM, Liberty Blue™) at a 0.1 mmol scale using Fmoc protected Wang resin according to the standard coupling strategy. Briefly, the resin was swollen in DMF for 1 h and subsequently the Fmoc protecting group was cleaved with a piperidine solution (20 vol% in DMF) by microwaving at 155 W, 75 °C for 15 s and at 30 W, 90 °C for 50 s. The resin was washed three times with DMF and Fmoc protected amino-acid (5 equiv. relative to the resin loading capacity), DIC (5 equiv.) and Oxyma Pure® (10 equiv.) were dissolved in DMF and added to the reaction vessel. After microwaving at 170 W, 75 °C for 15 s and at 30 W, 90 °C for 110 s the resin was washed with DMF. Repeating this procedure for all required amino acids yields the desired peptide sequence. In the final step, the Fmoc protecting group was cleaved by microwaving at 155 W, 75 °C for 15 s and at 30 W, 90 °C for 50 s with piperidine solution (20% in DMF) and the resin was washed with DCM. The peptide was cleaved off the resin through treatment with 5 mL of trifluoroacetic acid/water/triisopropylsilane (95% / 2.5% / 2.5%) for 2 h and precipitated in cold diethyl ether (40 mL). The precipitate was centrifuged at 3434 x g for 15 min and dissolved in water and 0.1% TFA and purified via HPLC using a gradient of water and acetonitrile containing 0.1% TFA or 0.1% NH<sub>4</sub>OH for peptides containing mainly basic or acidic amino acids, respectively. HPLC runs for purification were conducted in preparative scale using a Shimadzu system with the following modules: DGU-20A5R, LC-20AP, CBM-20A, SPD-M20A, SIL-10AP, FRC-18A and a Phenomenex Gemini® 5  $\mu$ m NX-C18, 110 Å, 150  $\times$  30 mm, 5  $\mu$ m pore size with a flow rate of 25 mL/min. The peptides were lyophilized to obtain a white solid. Purity (Fig. S35 and Table S3, Supporting Information) was analyzed by MALDI-TOF-MS (rapifleX MALDI-TOF/TOF, Bruker MALDI Synapt G2-SI, Waters) and HPLC-ESI-MS on a Shimadzu LC-2020 Single Quadrupole MS instrument equipped with the modules LC-20AD, SIL-20A, SPD-20A, CTO-using a Kinetex EVOC18 100 Å LC 50  $\times$  2.1 mm column with 2.6  $\mu$ m pore size. An acetonitrile/water mixture with additional 0.1% formic acid was used as eluent.

**Table 1 | Summary of pre-processing methods to obtain output data for data-mining from the readout data of experimental descriptors**

Descriptor	Method	Readout	Pre-processing	Output
Infectivity enhancement	$\beta$ -galactosidase assay	$\beta$ -galactosidase luminescence (infection rate intensity) in relative light units per second (RLU / s)	Calculation of n-fold infection in presence of peptides relative to virus only infection in absence of peptides. Calculation of ratio between n-fold infection of peptides relative to n-fold infection of EF-C (infection Rel. EF-C).	Logarithmic infectivity relative to EF-C ( $\text{Log}_{10}$ Rel Infect EF-C)
Fibril formation	TEM	Micrographs	Qualitative evaluation of high aspect ratio nanostructures (fibrils)	Fibril formation observed (1) or not observed (0)
Monomer to nanostructure conversion	Conversion Rate	Fluorescamine fluorescence intensity	Fluorescence intensity ratio before and after filtration	Percentual fluorescence intensity ratio (conversion rate)
Amyloid structure	ThT-fluorescence	Thioflavin-T fluorescence intensity	N-fold fluorescence intensity of peptides compared to buffer control sample	ThT-active (N-fold >2), or ThT-inactive (N-Fold <2).
$\beta$ -sheet content	FT-IR	Infrared absorbance (1300–1800 $\text{cm}^{-1}$ ) of peptide samples	Data-smoothing, normalization, calculation of second derivative of absorbance intensity, calculation of peak integral ratios between peaks at 1625-1635 $\text{cm}^{-1}$ ( $\beta$ -sheet) and approx. 1650-1700 $\text{cm}^{-1}$ (native structure).	Percentual $\beta$ -sheet content
Charge	Zeta-Potential	Zeta-Potential in mV	-	Zeta-Potential in mV
Microscopic aggregation	Light Scattering Count Rate	Light Scattering Count Rate in kilo counts per second (kcps)	-	Light Scattering Count Rate in kilo counts per second (kcps)
Microscopic aggregation	Widefield Microscopy	Microscopy images	Count area size and number of aggregates (>10 $\mu\text{m}^2$ ) via 3D objects counter (ImageJ)	Aggregate area size and number in a microscopy image.

## Nanofiber-formation

To induce fibril-formation, peptides were dissolved in DMSO (10 mg/mL stock solution) and diluted in PBS (DBPS w/o calcium chloride and magnesium chloride, Sigma Aldrich) to a final incubation concentration of 650  $\mu\text{M}$  or 1 mg/mL as indicated in the text. Other concentrations described in the text were achieved by further dilution of preformed PNFs. The solution was incubated at room temperature for 24 h in order to obtain fibrils. Literature-derived PNFs were prepared according to literature procedures (see SI Section 7).

## Amyloid fibril characterization

**ATR-FT-IR.** For ATR FT-IR spectroscopy measurements, 200  $\mu\text{L}$  of the respective 1 mg/mL PNF solution was lyophilized and the resulting powder used for measurement. All spectra were recorded on a Bruker Tensor 27 spectrometer with a diamond crystal as ATR element (PIKE MiracleTM, spectral resolution 2  $\text{cm}^{-1}$ ). Every sample was measured with 64 scans. Data were analyzed with OriginLab software and calculation of  $\beta$ -sheet content was conducted according to a previous report<sup>26</sup>. **TEM.** For TEM measurements 5  $\mu\text{L}$  of PNF fibrils (1 mg/mL) were placed on copper grids coated with carbon and formvar layer (300 mesh, Plano GmbH). After 10 min incubation time, the grids were staining with 4% uranyl acetate solution for 2.5 min and washed with water. Measurements were performed on a Jeol 1400 electron microscope with 120 kV acceleration voltage. **CR Assay.** For the determination of the amount of peptide monomer converting into nanostructures the amine reactive dye fluorescamine was used analogous to a previous reported protocol<sup>26</sup>. Briefly, 100  $\mu\text{L}$  of the preformed PNF solution was centrifuged in a Vivaspin 500 tube (3 kDa MWCO) to separate fibers from free peptide monomer (16200 x g, 4  $^{\circ}\text{C}$ , 45 min). The filtrate and 100  $\mu\text{L}$  of the non-filtered PNF sample were lyophilized and dissolved in 30  $\mu\text{L}$  DMSO. 10  $\mu\text{L}$  of the filtrate and non-filtered DMSO samples were submitted in a black UV Star<sup>®</sup> 384 microliter well-plates (Greiner bio-one) and 3  $\mu\text{L}$  of the fluorescamine solution (10 mg/mL, DMSO) was added to measure the fluorescence enhancement. After 20 min of incubation at RT, fluorescence was recorded on an Infinite<sup>®</sup> M1000 PRO microplate reader (Tecan) at  $\lambda_{\text{em}} = 470$  nm upon excitation at  $\lambda_{\text{ex}} = 365$  nm with 10 nm bandwidths and multiple reads per well (3x3). Measurements were repeated in triplicates and averaged with standard deviation. All values were calculated as n-fold fluorescence enhancement (DMSO only as a reference was set to 1). The conversion rate CR was defined according to the equation Eq. (3).

$$CR = 100 - \left( \frac{100 \times \text{fluorescence intensity (filtrate)}}{\text{fluorescence intensity (original)}} \right) [\%] \quad (3)$$

**ThT-assay.** 4  $\mu\text{L}$  of 1 mg/mL peptide fibrils solution is added to a 20  $\mu\text{L}$ , 50  $\mu\text{M}$  ThT-solution in PBS. The mixture is incubated for 15 min at RT to allow binding of ThT dye with  $\beta$ -sheet rich structures. For reference PBS (4  $\mu\text{L}$ ) was added instead of peptide fibril solution. The samples were placed in black UV Star<sup>®</sup> 384 microliter well-plates (Greiner bio-one). Fluorescence spectra were recorded on an Infinite<sup>®</sup> M1000 PRO microplate reader (Tecan) at  $\lambda_{\text{em}} = 488$  nm upon excitation at  $\lambda_{\text{ex}} = 440$  nm with 10 nm bandwidths and multiple reads per well (3x3). Measurements were repeated in triplicates and averaged with standard deviation. A peptide fibril was considered as ThT-active if the fluorescence intensity was at least twice as strong compared to the control (PBS containing 10% DMSO). **Zeta-Potential.** For the determination of the surface charge of aggregated peptides zeta-potential measurements were conducted with a Zetasizer Nano ZS, Malvern Instruments. Unless stated otherwise the peptides were dissolved from DMSO (10 mg/mL) in PBS (pH 7.4) to concentration of 1 mg/mL and incubated for 1d at RT. Just before the measurement the 60  $\mu\text{L}$  of the peptides were further diluted in 600  $\mu\text{L}$  KCl (aq, 1 mM) in a 1 mL disposable

folded capillary cells (DTS-1060, Zetasizer Nano series, Malvern). The zeta potential of the peptides was derived from the electrophoretic mobility based on the Smoluchowski formula. All measurements were averaged over two individually performed experiments and conducted in triplicates for each sample. **Aggregate analysis.** The derived count rate of scattered light at 633 nm, 173° from the zeta-potential measurement was used as information on the light scattering intensity and turbidity of the sample as an indicator for microscopic aggregation<sup>35</sup>. Furthermore, fluorescence microscopy was conducted by staining 1  $\mu$ L preformed fibrils (1 mg/mL) with 9  $\mu$ L ThT (50  $\mu$ M). Brightfield microscopy was performed simultaneously to evaluate aggregation also for non ThT-active fibrils. The measurements were performed on a Leica DMI8 microscope with 10x air objective and equipped with a Leica MCI70 HD camera with a filter setting  $\lambda_{em} = 527/30$  nm upon excitation at  $\lambda_{ex} = 480/40$  nm) and processed with the Leica LASX software. To quantify size and number of microscopic aggregation of particles > 10  $\mu$ m<sup>2</sup> at least 3 independently measured images were analyzed via the 3D object counter analysis tool of ImageJ, v 1.53f51. **Sum-frequency generation.** The SFG experiments were carried out with a femtosecond Ti: Sapphire amplified laser system (Spitfire Ace, Spectra-Physics, 800 nm, 40 fs, 1 kHz) with 5 W output power. We used 2 W to pump an optical parametric amplifier (TOPAS, light conversion) with a non-collinear difference frequency generation stage to generate broadband IR pulses. Another 1 W of the laser output passed through an etalon to generate narrowband visible pulses (10 cm<sup>-1</sup>). The incident angles of the visible and IR beams were 36° and 41° with respect to the surface normal, respectively. The visible (13  $\mu$ J) and IR (5  $\mu$ J) pulses were overlapped spatially and temporally at the sample position. Subsequently, the generated SFG pulse was guided into a spectrograph (Acton SP 300i, Princeton Instruments), and detected with an EMCCD (Newton, Andor Technology). To avoid IR absorption of water vapor, we purged N<sub>2</sub> for the measurement. All spectra were collected in the ssp (denoting s-, s- and p-polarized SFG, visible and IR, respectively) polarization combination. The spectra were normalized to the non-resonant signal taken from z-cut quartz after subtracting a background spectrum.

**Virus–Peptide Interaction.** T2M-bl cell line and R5-tropic HIV-1 stock plasmid (pBRNL4.39-92TH014) were obtained from the National Institutes of Health AIDS Research and Reference Reagent Program. Human wild-type HeLa cell line were obtained from abcam, ab260075, LOT: GR3292155-1. The cell lines were not further authenticated, and no commonly misidentified cell lines are used within this study. Infectious R5-tropic HIV-1 stocks and infectivity assays were prepared analogous to a previous report<sup>26</sup>. Briefly, 10,000 T2M-bl cells were seeded into 96-well plates in 180  $\mu$ L DMEM one day prior to infection experiment. Peptides were incubated two days prior to experiment at 650  $\mu$ M concentration in PBS to generate fibrils. One day before the experiment the peptide samples were diluted to 130  $\mu$ M in PBS. On the day of the infection experiment, 40  $\mu$ L of 130  $\mu$ M, 26  $\mu$ M, 5.2  $\mu$ M and 0  $\mu$ M peptide samples were mixed 1:1 (v/v) with 40  $\mu$ L of HIV-1 (pre-diluted 1:125 (v/v) in cell culture medium), resulting in fibril concentration of 65  $\mu$ M, 13  $\mu$ M, 2.6 and 0  $\mu$ M during virion treatment. After 10 min incubation to allow binding of virus particles to peptide fibrils, 20  $\mu$ L of the virus-peptide fibril mixture were added to 180  $\mu$ L of cells resulting in final cell culture peptide fibril concentrations of 6.5  $\mu$ M, 1.3  $\mu$ M, 0.26 and 0  $\mu$ M, respectively. The effect of peptide fibrils (final concentration on cells 6.5, 1.3, 0.26, 0  $\mu$ M) on HIV-1 infection was studied 3 days post infection via a luminescence assay for  $\beta$ -galactosidase, which was reported by T2M-bl cells upon infection. For the property–activity correlation a peptide concentration of 1.3  $\mu$ M was used as a standard condition and EF-C was always used as a reference to compare infectivity enhancement between peptides. The HIV-1 infection assay was conducted in three technical replicates and reproduced at least once.

The cell viability after addition of peptides to T2M-bl cells was studied via the CellTiter–Glo assay. To this end, 10,000 cells were seeded and on the next day serial diluted peptides were added. After 3 days the supernatant was removed and 100  $\mu$ L CellTiter–Glo Reagent 1:1 diluted in PBS was added. After 10 min 50  $\mu$ L was transferred to white microplate and luminescence was recorded by Orion microplate luminometer.

Confocal laser scanning microscopy studies were performed for the visualization of the cell–peptide interaction. HeLa cells were seeded one day prior to conducting the assay (40,000/well) in an 8–well IBIDI slide. 4  $\mu$ L of the preformed peptide fibrils (1 mg/mL) were diluted with 4  $\mu$ L Proteostat (Enzo Life Science, 1  $\mu$ L stock in 999  $\mu$ L PBS) and further diluted with medium to receive a final peptide concentration of 20  $\mu$ g/mL. The nucleus of the HeLa cells was stained with Hoechst 33342 (NucBlue™, Thermo Fisher Scientific). The peptide solution mixture was transferred to the HeLa cells and incubated for 30 min at 37 °C before washing three times with PBS. The interaction of fibril clusters with cells was monitored after 30 min incubation time on a Stellaris 8 confocal laser scanning microscope (Leica) equipped with a 20x air objective and laser excitation wavelength of 405 nm (Hoechst) and 561 nm (Proteostat).

### Statistics and reproducibility

All virus assays were conducted in technical triplicates. All tested peptides for the virus infectivity assay were conducted at four different concentrations (6.5  $\mu$ M, 1.3  $\mu$ M, 0.26  $\mu$ M, 0  $\mu$ M). To verify reproducibility of the experimental findings the infection assays were at least repeated independently in biological duplicates. The physicochemical characterization methods for peptide self-assembly and morphology, that are determined via CR Assay, ThT–Assay and Zeta-Potential have been conducted in technical triplicates with two independent repetitions for Zeta-Potential and ThT-Assay measurements. Unless stated otherwise the aggregate analysis via widefield microscopy and TEM analysis have been conducted once with at least three microscopy images recorded for each peptide sample. ATR–FT–IR measurements have been conducted once for each peptide sample. The sum–frequency generation experiments have been conducted in two independent repetitions. To ensure comparability within individual measurements, the reference peptide “EF-C” was included in all assays. All replication attempts were successful.

No statistical method was used to predetermine sample size. No data were excluded from the analyses and the experiments were not randomized because it is not relevant for this study, as infectivity enhancement is independent from the order of samples tested. The investigators were not blinded to allocation during experiments and outcome assessment because it is not relevant for this study, as the results from infection assays are read out automated via a multi-plate reader.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The authors declare that the main data supporting the findings of this study are available within the paper and its associated Supplementary Information. Source data for manuscript and supplementary figures are provided with this paper. Other databases that are used for this study are “The Protein Data Bank” (PDB, <https://www.rcsb.org/>)<sup>84</sup>, a report by Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J. and Serrano, L (<https://doi.org/10.1038/nbt1012>)<sup>52</sup> and the WALTZ-DB 2.0 database (<http://waltzdb.switchlab.org/home>)<sup>85</sup>. A summary of peptides considered for evaluation from these databases is provided in a supplementary file. Source data are provided with this paper.

## Code availability

The computer codes and the corresponding input data for amino acid and peptide pattern analyses are openly available at the following data repository <https://doi.org/10.5281/zenodo.8079726><sup>46</sup>.

## References

- Yoder, K. E., Rabe, A. J., Fishel, R. & Larue, R. C. Strategies for targeting retroviral integration for safer gene therapy: advances and challenges. *Front. Mol. Biosci.* **8**, 418 (2021).
- U.S. Food & Drug Administrations, Approval Letter - KYMRIAH (tisagenlecleucel). <https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/kymriah-tisagenlecleucel> (2017).
- Labbé, R. P., Vessillier, S. & Rafiq, Q. A. Lentiviral vectors for T cell engineering: clinical applications, bioprocessing and future perspectives. *Viruses* **13**, 1528 (2021).
- Vormittag, P., Gunn, R., Ghorashian, S. & Veraitch, F. S. A guide to manufacturing CAR T cell therapies. *Curr. Opin. Biotechnol.* **53**, 164–181 (2018).
- Kaygisiz, K. & Synatschke, C. V. Materials promoting viral gene delivery. *Biomater. Sci.* **8**, 6113–6156 (2020).
- Münch, J. et al. Semen-derived amyloid fibrils drastically enhance HIV infection. *Cell* **131**, 1059–1071 (2007).
- Arnold, F. et al. Naturally occurring fragments from two distinct regions of the prostatic acid phosphatase form amyloidogenic enhancers of HIV infection. *J. Virol.* **86**, 1244–1249 (2012).
- Roan, N. R. et al. Peptides released by physiological cleavage of semen coagulum proteins form amyloids that enhance HIV infection. *Cell Host Microbe* **10**, 541–550 (2011).
- Zhang, L. et al. A novel modified peptide derived from membrane-proximal external region of human immunodeficiency virus type 1 envelope significantly enhances retrovirus infection. *J. Pept. Sci.* **20**, 46–54 (2014).
- Lee, Y.-H. & Ramamoorthy, A. Semen-derived amyloidogenic peptides-Key players of HIV infection. *Protein Sci.* **27**, 1151–1165 (2018).
- Castelletto, V. & Hamley, I. W. Amyloid and hydrogel formation of a peptide sequence from a coronavirus spike protein. *ACS Nano* **16**, 1857–1867 (2022).
- Yolamanova, M. et al. Peptide nanofibrils boost retroviral gene transfer and provide a rapid means for concentrating viruses. *Nat. Nanotechnol.* **8**, 130–136 (2013).
- Shrimali, P. C. et al. Amyloidogenic propensity of self-assembling peptides and their adjuvant potential for use as DNA vaccines. *bioRxiv* 2022.09.09.507367 <https://doi.org/10.1101/2022.09.09.507367> (2022).
- Knowles, T. P. J. & Buehler, M. J. Nanomechanics of functional and pathological amyloid materials. *Nat. Nanotechnol.* **6**, 469–479 (2011).
- Al-Garawi, Z. S., Morris, K. L., Marshall, K. E., Eichler, J. & Serpell, L. C. The diversity and utility of amyloid fibrils formed by short amyloidogenic peptides. *Interface Focus* **7**, 20170027 (2017).
- Cremades, N. et al. Direct observation of the interconversion of normal and toxic forms of  $\alpha$ -synuclein. *Cell* **149**, 1048–1059 (2012).
- Klein, W., Stine, W. & Teplow, D. Small assemblies of unmodified amyloid  $\beta$ -protein are the proximate neurotoxin in Alzheimer's disease. *Neurobiol. Aging* **25**, 569–580 (2004).
- Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81**, 678–699 (2003).
- Dear, A. J. et al. Identification of on- and off-pathway oligomers in amyloid fibril formation. *Chem. Sci.* **11**, 6236–6247 (2020).
- Michaels, T. C. T. et al. Dynamics of oligomer populations formed during the aggregation of Alzheimer's A $\beta$ 42 peptide. *Nat. Chem.* **12**, 445–451 (2020).
- Pellarin, R. & Caflich, A. Interpreting the aggregation kinetics of amyloid peptides. *J. Mol. Biol.* **360**, 882–892 (2006).
- Fowler, D. M. et al. Functional amyloid formation within mammalian tissue. *PLoS Biol.* **4**, 0100–0107 (2006).
- Zapadka, K. L., Becher, F. J., Gomes dos Santos, A. L. & Jackson, S. E. Factors affecting the physical stability (aggregation) of peptide therapeutics. *Interface Focus* **7**, 20170030 (2017).
- Sinha, N. J., Langenstein, M. G., Pochan, D. J., Kloxin, C. J. & Saven, J. G. Peptide design and self-assembly into targeted nanostructure and functional materials. *Chem. Rev.* **121**, 13915–13935 (2021).
- Kirti, S. et al. Amyloid fibrils with positive charge enhance retroviral transduction in mammalian cells. *ACS Biomater. Sci. Eng.* **5**, 126–138 (2019).
- Sieste, S. et al. Supramolecular peptide nanofibrils with optimized sequences and molecular structures for efficient retroviral transduction. *Adv. Funct. Mater.* **31**, 2009382 (2021).
- Schütz, D. et al. Negatively charged peptide nanofibrils from immunoglobulin light chain sequester viral particles but lack cell-binding and viral transduction-enhancing properties. *ACS Omega* **6**, 7731–7738 (2021).
- Fernandez-Lozano, C., Gestal, M., Munteanu, C. R., Dorado, J. & Pazos, A. A methodology for the design of experiments in computational intelligence with multiple regression models. *PeerJ* **4**, e2721 (2016).
- Yamashita, H. et al. Machine learning-based amino acid substitution of short peptides: acquisition of peptides with enhanced inhibitory activities against  $\alpha$ -amylase and  $\alpha$ -Glucosidase. *ACS Biomater. Sci. Eng.* **6**, 6117–6125 (2020).
- Giguère, S. et al. Machine learning assisted design of highly active peptides for drug discovery. *PLOS Comput. Biol.* **11**, e1004074 (2015).
- Chowdhury, A. S., Reehl, S. M., Kehn-Hall, K., Bishop, B. & Webb-Robertson, B. J. M. Better understanding and prediction of antiviral peptides through primary and secondary structure feature importance. *Sci. Rep.* **10**, 1–8 (2020).
- Schilling, C. et al. Sequence-optimized peptide nanofibers as growth stimulators for regeneration of peripheral neurons. *Adv. Funct. Mater.* **29**, 1809112 (2019).
- Merrifield, R. B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).
- Krebs, M. R. H., Bromley, E. H. C. & Donald, A. M. The binding of thioflavin-T to amyloid fibrils: Localisation and implications. *J. Struct. Biol.* **149**, 30–37 (2005).
- Yoshimura, Y. et al. Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. *Proc. Natl Acad. Sci. USA.* **109**, 14446–14451 (2012).
- Adamcik, J. et al. Understanding amyloid aggregation by statistical analysis of atomic force microscopy images. *Nat. Nanotechnol.* **5**, 423–428 (2010).
- Bolisetty, S., Adamcik, J. & Mezzenga, R. Snapshots of fibrillation and aggregation kinetics in multistranded amyloid  $\beta$ -lactoglobulin fibrils. *Soft Matter* **7**, 493–499 (2011).
- Ridgley, D. M., Ebanks, K. C. & Barone, J. R. Peptide mixtures can self-assemble into large amyloid fibers of varying size and morphology. *Biomacromolecules* **12**, 3770–3779 (2011).
- Fauchère, J.-L., Charton, M., Kier, L. B., Verloop, A. & Pliska, V. Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int. J. Pept. Protein Res.* **32**, 269–278 (2009).
- Osorio, D., Rondón-Villarreal, P. & Torres, R. Peptides: a package for data mining of antimicrobial. *Pept. R. J.* **7**, 4 (2015).
- Takeuchi, Y., McClure, M. O. & Pizzato, M. Identification of gammaretroviruses constitutively released from cell lines used for human immunodeficiency virus research. *J. Virol.* **82**, 12585–12588 (2008).
- Fu, L., Ma, G. & Yan, E. C. Y. In situ misfolding of human islet amyloid polypeptide at interfaces probed by vibrational sum frequency generation. *J. Am. Chem. Soc.* **132**, 5405–5412 (2010).

43. Hosseinpour, S. et al. Structure and dynamics of interfacial peptides and proteins from vibrational sum-frequency generation spectroscopy. *Chem. Rev.* **120**, 3420–3465 (2020).
44. Kaygisiz, K. et al. Inverse design of viral infectivity-enhancing peptide fibrils from continuous protein-vector embeddings. *Biomater. Sci.* **11**, 5251–5261 (2023).
45. Zhao, J. et al. Epigallocatechin-3-gallate (EGCG) inhibits fibrillation, disaggregates amyloid fibrils of  $\alpha$ -synuclein, and protects PC12 cells against  $\alpha$ -synuclein-induced toxicity. *RSC Adv.* **7**, 32508–32517 (2017).
46. Kaygisiz, K. et al. Data-mining unveils structure-property-activity correlation of viral infectivity enhancing self-assembling peptides. *Zenodo.org* <https://doi.org/10.5281/ZENODO.8079727> (2023).
47. Fulop, T. et al. Can an infection hypothesis explain the beta amyloid hypothesis of Alzheimer's Disease? *Front. Aging Neurosci.* **10**, 224 (2018).
48. Hur, J. Y. et al. The innate immunity protein IFITM3 modulates  $\gamma$ -secretase in Alzheimer's disease. *Nature* **586**, 735–740 (2020).
49. Medini, K. et al. Controlling gelation with sequence: Towards programmable peptide hydrogels. *Acta Biomater.* **43**, 30–37 (2016).
50. Kurnellas, M. P., Adams, C. M., Sobel, R. A., Steinman, L. & Rothbard, J. B. Amyloid fibrils composed of hexameric peptides attenuate neuroinflammation. *Sci. Transl. Med.* **5**, 179ra42 (2013).
51. Maurer-Stroh, S. et al. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods* **7**, 237–242 (2010).
52. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J. & Serano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–1306 (2004).
53. Aubrey, L. D. et al. Quantification of amyloid fibril polymorphism by nano-morphometry reveals the individuality of filament assembly. *Commun. Chem.* **3**, 125 (2020).
54. Qiang, W., Yau, W.-M., Lu, J.-X., Collinge, J. & Tycko, R. Structural variation in amyloid- $\beta$  fibrils from Alzheimer's disease clinical subtypes. *Nature* **541**, 217–221 (2017).
55. Schütz, D. et al. Viral transduction enhancing effect of EF-C peptide nanofibrils is mediated by cellular protrusions. *Adv. Funct. Mater.* **31**, 2104814 (2021).
56. Betush, R. J., Urban, J. M. & Nilsson, B. L. Balancing hydrophobicity and sequence pattern to influence self-assembly of amphipathic peptides. *Pept. Sci.* **110**, e23099 (2018).
57. De Mello, L. R. et al. Amyloid-like self-assembly of a hydrophobic cell-penetrating peptide and its use as a carrier for nucleic acids. *ACS Appl. Bio Mater.* **4**, 6404–6416 (2021).
58. Zhao, R. et al. Measurement of amyloid formation by turbidity assay—seeing through the cloud. *Biophys. Rev.* **8**, 445–471 (2016).
59. Frederix, P. W. J. M. et al. Exploring the sequence space for (tri-) peptide self-assembly to design and discover new hydrogels. *Nat. Chem.* **7**, 30–37 (2015).
60. Strandberg, E. et al. Phosphate-dependent aggregation of [KL]n peptides affects their membranolytic activity. *Sci. Rep.* **10**, 12300 (2020).
61. Marek, P. J., Patsalo, V., Green, D. F. & Raleigh, D. P. Ionic strength effects on amyloid formation by amylin are a complicated interplay among debye screening, ion selectivity, and hofmeister effects. *Biochemistry* **51**, 8478–8490 (2012).
62. Zhang, Z. et al. Dose-independent transfection of hydrophobized polyplexes. *Adv. Mater.* **33**, 2102219 (2021).
63. Hack, F. J. et al. Indole, phenyl, and phenol groups: the role of the comonomer on gene delivery in guanidinium containing methacrylamide terpolymers. *Macromol. Rapid Commun.* **42**, 2000580 (2021).
64. Yang, Y., Jalali, S., Nilsson, B. L. & Dias, C. L. Binding mechanisms of amyloid-like peptides to lipid bilayers and effects of divalent cations. *ACS Chem. Neurosci.* **12**, 2027–2035 (2021).
65. Capule, C. C., Brown, C., Olsen, J. S., Dewhurst, S. & Yang, J. Oligovalent amyloid-binding agents reduce SEVI-mediated enhancement of HIV-1 infection. *J. Am. Chem. Soc.* **134**, 905–908 (2012).
66. West, M. W. et al. De novo amyloid proteins from designed combinatorial libraries. *Proc. Natl Acad. Sci. USA.* **96**, 11211–11216 (1999).
67. Milanese, L. et al. Direct three-dimensional visualization of membrane disruption by amyloid fibrils. *Proc. Natl Acad. Sci.* **109**, 20455–20460 (2012).
68. Kaygisiz, K. et al. Photoinduced amyloid fibril degradation for controlled cell patterning. *Macromol. Biosci.* **23**, 2200294 (2023).
69. Ender, A. M. et al. Cell-instructive surface gradients of photo-responsive amyloid-like fibrils. *ACS Biomater. Sci. Eng.* **7**, 4798–4808 (2021).
70. Usmani, S. M. et al. Direct visualization of HIV-enhancing endogenous amyloid fibrils in human semen. *Nat. Commun.* **5**, 1–8 (2014).
71. Chiti, F. & Dobson, C. M. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
72. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
73. Knowles, T. P. et al. Role of intermolecular forces in defining material properties of protein nanofibrils. *Science* **318**, 1900–1903 (2007).
74. Mossou, E. et al. The self-assembling zwitterionic form of L-phenylalanine at neutral pH. *Acta Crystallogr. Sect. C. Struct. Chem.* **70**, 326–331 (2014).
75. Nelson, R. et al. Structure of the cross- $\beta$  spine of amyloid-like fibrils. *Nature* **435**, 773–778 (2005).
76. Do, T. D., Kincannon, W. M. & Bowers, M. T. Phenylalanine oligomers and fibrils: The mechanism of assembly and the importance of tetramers and counterions. *J. Am. Chem. Soc.* **137**, 10080–10083 (2015).
77. Eckes, K. M., Mu, X., Ruehle, M. A., Ren, P. & Suggs, L. J.  $\beta$  sheets not required: combined experimental and computational studies of self-assembly and gelation of the ester-containing analogue of an Fmoc-dipeptide hydrogelator. *Langmuir* **30**, 5287–5296 (2014).
78. Paparcone, R., Cranford, S. W. & Buehler, M. J. Self-folding and aggregation of amyloid nanofibrils. *Nanoscale* **3**, 1748 (2011).
79. Nguyen, P. H., Sterpone, F. & Derreumaux, P. Aggregation of disease-related peptides. In *Progress in Molecular Biology and Translational Science* vol. 170 435–460 (Elsevier, 2020).
80. Adamcik, J. & Mezzenga, R. Adjustable twisting periodic pitch of amyloid fibrils. *Soft Matter* **7**, 5437 (2011).
81. Elsayy, M. A. et al. Modification of  $\beta$ -sheet forming peptide hydrophobic face: effect on self-assembly and gelation. *Langmuir* **32**, 4917–4923 (2016).
82. Wu, Y. et al. Patchy amphiphilic dendrimers bind adenovirus and control its host interactions and in vivo distribution. *ACS Nano* **13**, 8749–8759 (2019).
83. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
84. Berman, H. M. et al. The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
85. Louros, N. et al. WALTZ-DB 2.0: an updated database containing structural information of experimentally determined amyloid-forming peptides. *Nucleic Acids Res.* **48**, D389–D393 (2020).

## Acknowledgements

The study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projekt Nummer 316249678—SFB 1279



(A02, A03, A05, C01). A.D. acknowledges support by BiGmax, the Max Planck Society's Research Network on Big-Data-Driven Materials-Science. T.B. acknowledges support from the Emmy Noether program of the Deutsche Forschungsgemeinschaft (DFG).

### Author contributions

KK conceived the idea, designed, and carried out the experiments, analyzed data, performed the formal analysis, and wrote the original draft. LRW designed, and carried the experiments related to virus transduction, analyzed data, performed the formal analysis, and revised the manuscript. AD, XY, YN analyzed data, performed the formal analysis, and revised the manuscript. TB, CVS, JM, and TW conceptualized and supervised, provided the funding resources, and revised the manuscript. The final manuscript was approved by all authors.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

J.M. and L.R.W. hold patents and K.K., L.R.W., C.V.S., J.M. and T.W. submitted a patent for application of peptide fibrils for retroviral transduction enhancement. The remaining authors declare no competing interests. The authors have no relevant non-financial interests to disclose.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-40663-6>.

**Correspondence** and requests for materials should be addressed to Christopher V. Synatschke or Tanja Weil.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023