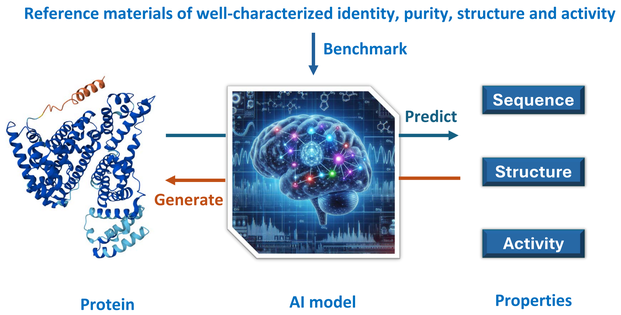
Characterization of *de novo* designed proteins: towards certified reference materials

Gustavo Martos[1], Andrea Briones[2], Ralf Josephs[1], Magali Bedu[1], Tiphaine Choteau[1], Max Ryadnov[2], Robert Wielgosz[1]

[1]Bureau International des Poids et Mesures (BIPM) and [2]National Physics Laboratory (NPL, UK)

Engineering biology is a critical technology of global significance. Confidence in the predictability and reproducibility of bioengineered systems and processes requires underpinning reference materials and methods.

The understanding of the principles governing the folding of primary amino acid sequences has allowed the engineering of *de novo* peptides and proteins with desirable functions, including self-assembling virus-like particles (VLPs). More recently, machine learning algorithms have arisen as powerful tools to accurately predict the 3D protein structures. The training of these models with larger, high-fidelity datasets opens the possibility to the AI-assisted design of peptides and proteins with promising applications in fields such as cell and gene therapy or vaccine and drug development. The reproducibility of the emerging manufacturing processes and the SI traceability of the materials' desired properties is essential to ensure their efficacy and safety. *De novo* peptide and protein standards of well-characterized identity, purity, structure and activity are therefore needed to benchmark AI-driven engineered proteins.



The CCQM Protein Analysis Working Group is running a series of interlaboratory comparisons to assess National Metrology Institutes' (NMIs) capabilities to characterize peptide and protein pure standards materials. The chemistry department of the BIPM and the National Physics Laboratory (UK) have collaborated to investigate a potential candidate material for such comparisons, i.e. the C_3(+) triskelion *de novo* peptide, capable of self-assembling into artificial virus-like capsids exhibiting strong antimicrobial activity (Fig. 1). The material could also be a candidate reference material for VLPs, such as gene-delivery products.

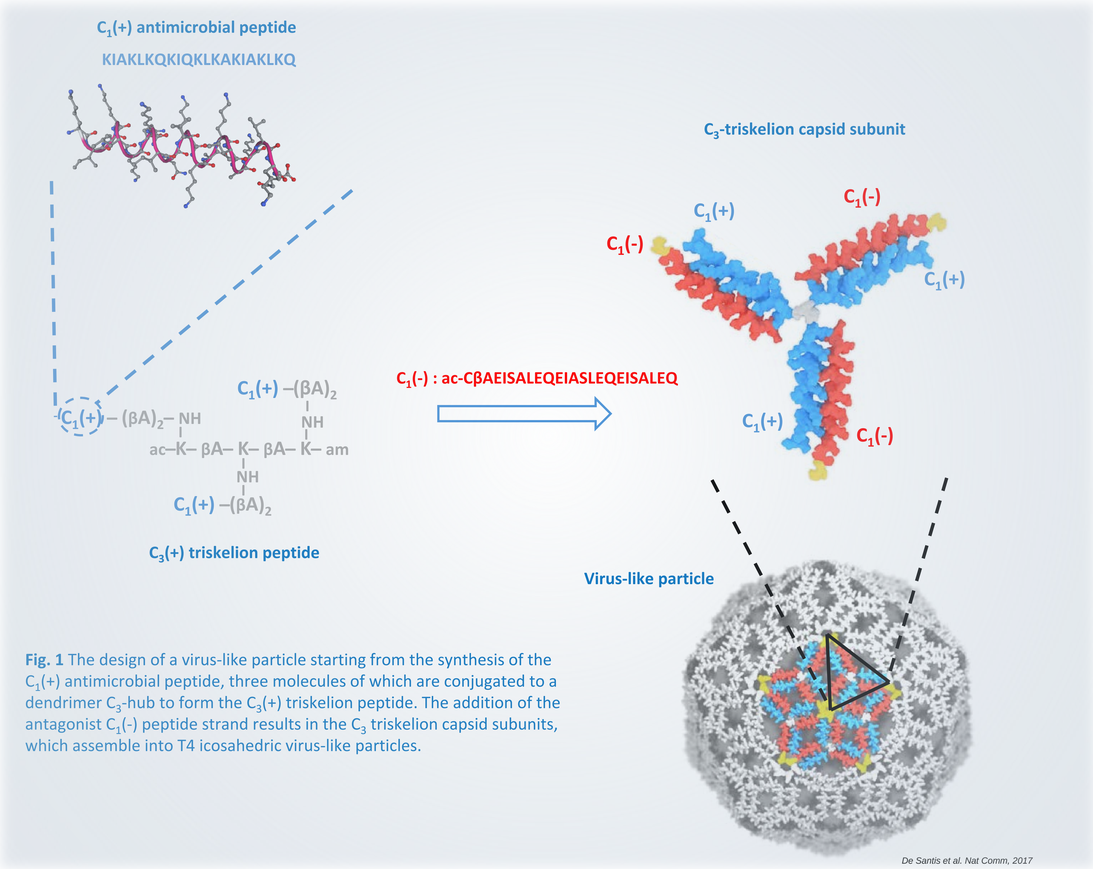


Fig. 1 The design of a virus-like particle starting from the synthesis of the C_1(+) antimicrobial peptide, three molecules of which are conjugated to a dendrimer C_3-hub to form the C_3(+) triskelion peptide. The addition of the antagonist C_1(-) peptide strand results in the C_3 triskelion capsid subunits, which assemble into T4 icosahedric virus-like particles.

De Santis et al. Nat Comm, 2017

The **methodology** developed at the BIPM to assign the purity of the C_3(+) triskelion included the mass balance method, qNMR and amino acid analysis. The identity and structurally related impurities were determined by LC-hrMS.

Mass balance

The determination of non-peptidic impurities and subtraction from the maximum purity value of 1000 mg/g allowed for the determination of the total peptide content.

Impurity components	*w (mg/g)*	*u (mg/g)*	*U (mg/g; k=2)*	Method	Traceability
H_2O	62	7.4	15	KFT	NIST SRM 2890
TFA	267	1.1	2.2	^{19}F-qNMR	BTMSB-F_4 internal standard
Cations	2	0.09	0.18	IC	KNO_3, $NaNO_3$, NH_4Cl, $CaCO_3$ CRMs
Peptide	669	7.5	15	1000-Σimp	

Table 1 Mass balance approach leading to the indirect quantification of total peptide content based on the determination of water by coulometric Karl Fischer titration, trifluoroacetic acid (TFA) by ^{19}F quantitative nuclear magnetic resonance (qNMR) and cations content by ion chromatography (IC).

High resolution mass spectrometry

High resolution and accuracy (<1 ppm) mass spectrometry allowed for the identity confirmation of the C_3(+) triskelion peptide and the detection of a major peptide impurity lacking 3 lysine residues. The ability of this impurity to form capsid subunits was not investigated so its contribution to the material impurity content was not considered.

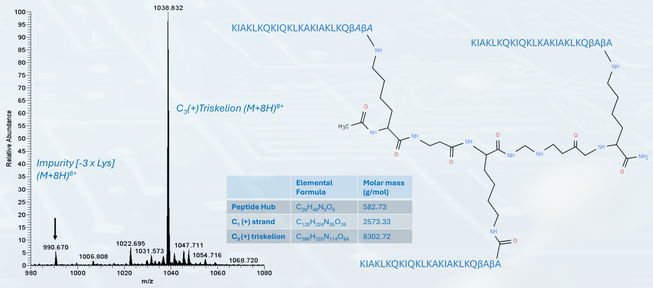


Fig. 3 High resolution mass spectrum of the C_3(+) triskelion material obtained on an LTQ-Orbitrap XL mass spectrometer (ThermoFisher) coupled to a liquid chromatography system (Agilent).

qNMR

^{1}H qNMR was based on the resonance signal of the ε-methylene protons of the C_3(+) triskelion lysine residues to establish the purity of the peptide by comparison to the resonance of the internal standard (Fig. 2). Other assigned signals were used for confirmation purposes and the estimation of measurement uncertainty.

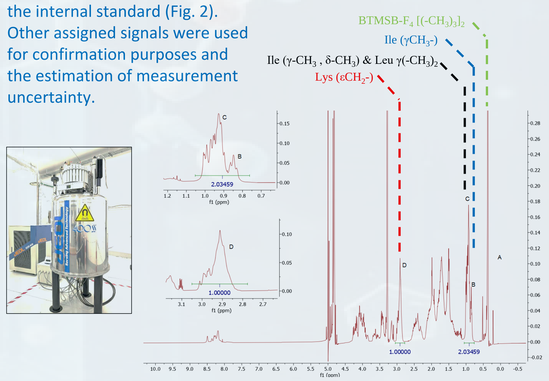


Fig. 2 ^{1}H NMR spectrum of C_3(+) triskelion peptide sample containing BTFSB-F_4 as internal standard. Acquisition was done on an ECS-400 MHz NMR instrument equipped with a direct type (Royal) automatic tuning probe (JEOL).

Purity Value assignment

The mass fraction values derived from qNMR and isotope-dilution amino acid analysis (AAA, Leu and Ile quantification) were used as confirmatory values of the mass balance purity assignment. The methodologies were consistent within uncertainties.

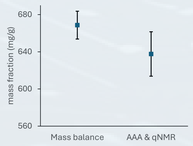


Fig. 4 Confirmation of mass balance purity assignment by AAA and qNMR.

Briones et al. iScience, 2022

The **future** holds great promise for custom-engineered proteins with targeted functions. The presented results demonstrated the possibility to value assign the mass fraction content with well-defined measurement uncertainty for a bioengineered material capable of self-assembling into VLPs. Reference materials of this type may be employed to confirm the purity of a commercial VLP, or virus-derived product as required by the manufacturer or a regulatory body; assist in the quantification of the encapsulation efficacy of a designed gene-delivery system; support validation of VLP performance in different sample matrices, *in vitro*, and in cell extracts and ultimately in live cells and tissues; and provide a route for measuring the amount of a desired material in target media with well-defined uncertainty. Candidate materials for future interlaboratory comparisons (Fig. 5) could allow National Metrology Institutes to test their measurement capabilities in this emerging field.

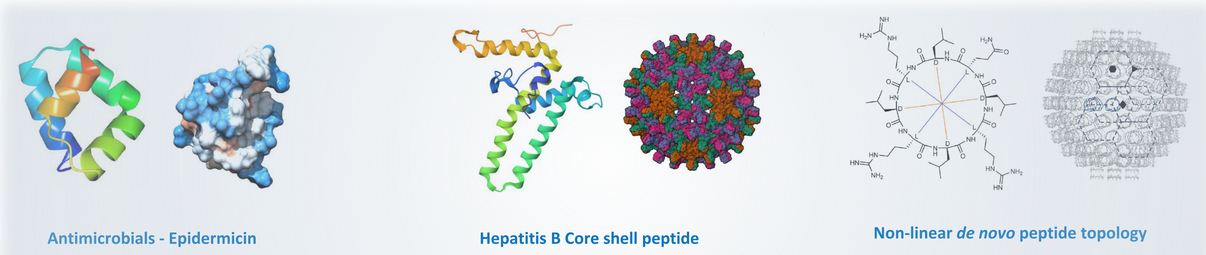


Fig. 5 Examples of protein topologies explored for the design of *de novo* virus-like particles with biological functions. Adapted from PDB entries 6SIG (left), 6HTX (middle) and from Noble et al. ACS Nano, 2024 (right).