Bottom-up instructive quality control in the biofabrication of smart protein materials

Fabián Rueda [1, 2, 3f]*, María Virtudes Céspedes* [3, 4f]*, Oscar Conchillo* [1]*, Alejandro Sanchez-Chardi* [5]*, Joaquin Seras-Franzoso* [1, 2, 3 £]*, Rafael Cubarsi* [6]*, Alberto Gallardo* [3, 4]*, Mireia Pesarrodona* [1, 2, 3]*, Neus Ferrer-Miralles* [1, 2, 3]*, Xavier Daura* [1,7]*, Esther Vázquez* [1, 2, 3]*, Elena García-Fruitós* [1, 2, 3 ¥]*, Ramón Mangues* [3, 4 *]*, Ugutz Unzueta* [3, 4]*, Antonio Villaverde* [1, 2, 3 *]

[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, 08193 Cerdanyola del Vallès, Spain.

[2] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, 08193 Cerdanyola del Vallès, Spain.

[3] CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Cerdanyola del Vallès, Spain.

[4] Biomedical Research Institute Sant Pau (IIB-SantPau), Hospital de la Santa Creu i Sant Pau, 08025 Barcelona, Spain.

[5] Servei de Microscòpia, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain.

[6] Departament de Matemàtica Aplicada IV, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain

[7] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

[f] Equally contributed

[£] Present address: Cibbim-Nanomedicine, Hospital Vall d'Hebron, Passeig de la Vall d'Hebron, 119-129, 08035 Barcelona, Spain

[¥] Present address: Department of Ruminant Production, IRTA, Institute of Agrifood Research and Technology. Torre Marimon, km 12,1 C-59, Caldes de Montbui, 08140, Barcelona, Spain

[*] corresponding authors.

A. Villaverde
e-mail, antoni.villaverde@uab.es
phone, +34 935813086
fax, +34 935812011

R. Mangues
e-mail, rmangues@santpau.cat
phone, +34 935537918
fax, +34 935537631

Nanostructured protein materials are gaining interest in biomedicine because of their biocompatibility, easy production and functional versatility. Merging structure and function in proteins allows designing protein composites with refined functions such as cell or tissue targeting. The basis of protein structure and biological activity is the attained spatial conformation, in a process tightly surveyed by the cell factory. However, at which extent the cell's quality control determines the architecture and biological performance of functional protein materials is a neglected issue. We demonstrate here that the activity at the systems level of a tumour-targeted protein-only nanoparticle is dramatically affected by key knock-out mutations in the quality control network of the producing bacteria, resulting in altered biodistribution patterns upon systemic administration. Therefore, since the conformational modulation at the molecular level determines the macroscopic biological performance, a tailored tuning of protein materials' activities might be approachable, in a bottom-up fashion, by the appropriate genetic adjustment of the cell factory's folding machinery.

Since the approval of insulin in 1981, [1] about 400 protein drugs, mainly produced in microbial cells, [2] have been authorized for use in humans. Apart from plain therapeutic cytokines, hormones, enzymes and antibodies, a plethora of more elaborated protein structures with different extents of complexity have been developed as nanoconjugates for drug delivery [3] including nab-paclitaxel, [4] denileukin difitox, [5] PEG-ADA [6] and pegaspargase. [7] Recent developments in the engineering of protein self-assembling [8] and the expanding catalogues of homing peptides [9] offer clues for the design and construction of smart protein nanostructures intended as functional substrates

in regenerative medicine [10] or as vehicles for the cell-targeted delivery of payload imaging agents and drugs. [11] Most of these applications are based on specific interactions between peptidic ligands displayed on the material's surface and surface-exposed receptors on the membrane of target cells, as aiming to internalization or signalling. Engineered protein materials are produced, as recombinant versions, in cell factories, mainly bacteria, [12] thus benefiting from the versatility and adaptability of biological fabrication. [13] In recombinant bacteria, disaggregation, folding and refolding, are executed by the quality control system (the chaperone-protease network), to minimize aggregation and to promote proper folding of engineered polypeptides. [14] How the quality control system does handle conventional soluble proteins is rather well stablished. [15] However, the cell's surveillance of bioactive, complex protein nanostructures performing specialized functions is a neglected issue, while it has a pivotal relevance in the context of emerging protein materials. [12] We have here analyzed the influence of the bacterial quality control on hyerarchical structural features and biological performance of smart protein materials of biomedical interest, illustated by a tumor-targeted, self-assembling nanoparticle produced by recombinant methods.

For that, we selected T22-GFP-H6, an engineered polypeptide (**Figure** 1A) that binds the cytokine receptor CXCR4 via the tumor-homing peptide T22. [16] This protein spontaneously self-assembles as nanoparticles of ~15 nm, that as observed by FESEM [17] (**Figure** 1B) organize as regular toroid (ring-shaped) materials,. When systemically administered in colorectal cancer mice models, these particles escape renal filtration and target primary tumour and metastatic

foci, through specific internalization in CXCR4$^+$ cells. [17, 18] Both the amino terminal T22 and the carboxy terminal H6 are involved in the interactions between building blocks that support nanoparticle formation. [19] Since T22 folds through two disulphide bonds, the fusion protein has been usually produced in *Escherichia coli* BL21 Origami B (TrxB$^-$, Gor$^-$) to facilitate disulphide bridge formation in a less reducing environment. [17]

To evaluate to which extent the protein production/folding machinery might have an impact on protein self-assembling and thus influence architectonic features and function of T22-GFP-H6 nanoparticles, the building block was produced in *E. coli* K-12 strains with knock-outed critical agents critical in different arms of the protein quality control. For that, we selected the main negative regulator of the whole quality control system and main disaggregase/foldase (the chaperone DnaK, JGT20 strain), the versatile ATPase ClpA (JGT4 strain) involved in ATP-dependent processes related with protein management, and the key cytosolic protease ClpP (JGT19 strain) that degrades misfolding-prone proteins. [20] As expected, recombinant proteins produced in these mutants exhibit altered proteolytic stability, solubility, aggregation profile and biological activity [20]. T22-GFP-H6 was also synthesized in the parental MC4100 *ara-D139 Δ(argF-lac)U169 rpsL150 relA1 flbB5301 deoC1 ptsF25 rbsR*). Origami B was kept as a reference for its ability to favour disulphide bridge formation. In all these strains, proteins were produced intracellularly and further purified from bacterial extracts by His-tag affinity chromatography. This resulted in the protein eluted into two separated peaks (P1 and P2, Supplementary **Figure** 1 A). Signs of differential proteolysis (Supplementary Figure 1 B) and variable protein yields (**Figure** 1 C)

5

were indicative of distinct protein management among the mutant set. Divergent patterns were observed when comparing Origami B and the K-12 derivatives, probably linked to enhanced disulphide bridge formation in the first case. The occurrence of two main peaks during imidazole elution indicated alternative conformations of the C-terminal hexahistidine, that affect the performance of T22-GFP-H6 in the Ni^{2+} chromatography (P1 generally showing less affinity by the matrix). Conformational variability was also supported by the spectrum of specific emission values of GFP fluorescence observed when T22-GFP-H6 particles were produced in alternative strains with aberrant quality control circuits (**Figure** 1 D).

Interestingly, there were no significant differences in the size (measured by DLS) and superficial charge (measured as Zeta potential) of the produced materials (**Figure** 1E, F), although P1 particles were slightly smaller than those found in P2 (an average of around 14 nm versus 16 nm). TEM and FESEM images confirmed the regular size and the ring-shaped nanoparticle organization, but also the slight differences between the material size in P1 and P2 fractions (**Figure** 2A). This indicated a robust self-assembling of building blocks, resulting in similar oligomers and oligomer-oligomer interactions irrespective of the strain used as factory. Despite the homogeneous geometry, the penetrability into $CXCR4^+$ HeLa cells was dramatically influenced by the genetic background, showing a wide variability when comparing data between bacterial strains. In addition, the uptake of P2 fractions was generally higher than that of P1's in the MC4100 background (**Figure** 2B). This fact supported the occurrence of conformational protein variants mostly represented by P1 and

P2 subpopulations but also modulated by the performance of the quality control in particular strains. While exhibiting similar geometries once assembled, the T22-homing peptide and the H6 tail of T22-GFP-H6 might be more available for cross-molecular interactions (with CXCR4 and Ni^{2+} respectively) in P2 than in P1 particles. H6, being a purification tag, is also a powerful endosomal escape agent [21], and its enhanced display would favour not only protein purification but also stability of internalized nanoparticles, as it seems to occur in P2 materials.

To identify any relevant physicochemical properties of the material that might influence cell penetrability we confronted internalization data (**Figure** 2B) with nanoparticle size and Zeta potential, with no observable dependences ($p > 0.1$ in both cases). However, when matching internalization and specific fluorescence data (representative of conformational status), a neat exponential trend of internalization (y) as a function of the specific fluorescence (x) was observed **Figure** 2C (left), corresponding to a constant decreasing uptake rate with increasing specific fluorescence following

Equation 1
$$\frac{1}{y}\frac{dy}{dx} = -A \; ; \; \; ln(|y|) + Ax = B$$

for positive parameters A and B, that are constants characteristic of the internalization process. In particular, the parameter B might be interpreted as a common molecular basis of cell uptake and capacity to emit fluorescence, related to protein conformation. In **Figure** 2C (right), the linear regression of the right-hand side of Eq. (1) is displayed, providing parameter values $A=$

$-18.7 \cdot 10^{-5} \pm 2.7 \cdot 10^{-5}$ mg/U and $B=5.66 \pm 0.24$, for a correlation coefficient

r=0.82 and a significance of $p=4.9 \cdot 10^{-7}$.

In summary, the most fluorescent nanoparticles among the whole set, mainly represented by P1, showed highly limited cell penetrability probably linked to a moderate exposure of H6 and T22 tails (T22-GFP-H6 produced in Origami B strain, being an exception). Contrarily, P2 materials were in general less fluorescent but more capable to penetrate target cells. This was linked to higher affinity of P2 materials by the chromatographic matrix, when comparing with P1's. Interestingly, different bacterial strains produced materials along a functional spectrum (covering a wide range of cell internalization potential and fluorescence emission; **Figure** 2C), whose ends might be defined by archetypal P1 and P2 fractions. This suggested a continuum of variants produced by bacterial mutants covering a wide range of cell internalization potential and fluorescence emission (**Figure** 2C). In this regard, some particle variants, mostly represented (but not exclusively) by P1, exhibited a specific fluorescence emission higher than the parental, unassembled protein GFP-H6, that was estimated to be 7,584 ± 834 Units/mg under the same recording settings. This indicated that the assembling of conformational isoforms of the building block might activate the GFP chromophore through particular cross-molecular interactions. Such particular architecture would restrict the solvent exposure of both end terminal tails. Nanoparticle versions with more exposed tails might be instead equally or slightly less fluorescent than the wild type GFP-H6.

This possibility was explored by modelling T22-GFP-H6 nanoparticles, resulting in two models that best fitted microscopic images and size measures (P1 and P2 in **Figure** 2D). P1 was characterized by less exposed terminal tails (**Figure** 2D, bottom) and by inter-molecular interactions between the overhanging ligand T22 and residues H148, M154, V163, V164, I167, S202, T203, E222 (amino acid numbering from [22]) of the adjacent GFP barrel (**Figure** 2E). These residues have been previously reported as involved in the modulation of the intensity of the fluorescence emission, [23] supporting the idea that T22 could modify their molecular environment, enhancing the fluorescence activity of GFP. [24, 25] In P2 materials, terminal tails are more available to external interactors and the GFP chromophore remains unaffected by internal cross-molecular contacts. Interestingly, the diameter of the modelled particles, which was not used as a modelling restraint, was 14.1 nm for P1 and 16.7 nm for P2, close to the experimental diameters and with a very similar difference (about 3 nm) between them. When exploring the set of (differently) cell-internalizing P2 particles, the suspected variability in the conformation of building blocks even within a single peak (**Figure** 2C) was confirmed by comparing the signal intensity at 218 nm on circular dichroism spectrum, where greater beta-sheet secondary structure signal is displayed by more compacted conformations (Supplementary **Figure** 1 C).

Geometry (size and shape) and charge of nanostructured materials administered systemically determine cell penetrability, local diffusion and biodistribution [26]. Being these parameters essentially homogeneous among the set of nanoparticles generated here, the potential of T22-GFP-H6 to bind

and internalize CXCR4+ cells *in vitro* is modulated by subtle conformational differences in the building blocks. To asses if the variable uptake *in vitro* could also influence the *in vivo* performance of the material, the biodistribution of P2 nanoparticles (those among the whole set, exhibiting efficient cell uptake *in vitro*) was evaluated in CXCR4$^+$ colorectal mouse models upon systemic administration. Under the homogeneous geometry of the P2 particle set, any altered biodistribution map should be uniquely attributed to fine structural variability determined by the bacterial genetic background. No fluorescence was observed in kidney or lung in any case, indicative of high stability of all particle isoforms, which remained assembled and dispersed *in vivo* (Table 1). The analysis of particle accumulation in CXCR4-overexpressing tumour tissue indeed revealed, at a first glance, heterogeneous tumour targeting (**Figure** 3 A). Such analysis was performed by determining CXCR4 expression in all individual tumours (**Figure** 3B and Supplementary **Figure** 2 B) and re-evaluating specific fluorescence emission by the IVIS equipment (Supplementary **Figure** 2D). We confirmed the differences in tumour uptake by determining immunohistochemically the amount of protein in the cell cytosol (**Figure** 3C). Accumulation in normal tissues (lacking or expressing low levels of CXCR4) was undetectable for all nanoparticle versions, except for an unexpected uptake in brain (**Figure** 4 A) restricted to endothelial cells (Figure 4 B), and exclusively observed in the material fabricated by JGT20 (DnaK$^-$) cells (Table 1). In this mutant, disaggregation and proteolysis of recombinant proteins is severely impaired, what results in a dramatic expansion of the spectrum of conformational protein variants and a less constrained targeting of the particles *in vivo*, when compared to reference strains. [20] In this regard, the material

produced in this genetic background shows an enhanced ability to reach the brain upon systemic administration, as compared to its uptake in tumor cells *in vivo*. This is indicative of an acquired capacity to internalize the CXCR4[+] microendothelia of the brain parenchyma. Conversely, nanoparticles produced in MC4100 show a high capacity to internalize epithelial tumor cells while are unable to undergo trancytosis through brain endothelial cells. This suggests different mechanisms of nanoparticle uptake in the brain and other organs, most likely determined by the blood-brain barrier (BBB). [27] In agreement, nanoparticles able to cross the BBB display improved transcytosis capacity in brain endothelial cells, while they are sorted differently in epithelial cells. [27, 28] The acquisition or not of BBB-crossing capacities in protein nanoparticles might have then subtle conformational bases, what would be of great relevance in the design of neurotropic vehicles and in identifying the rules for differential receptor-mediated internalization in epithelial tumor cells and other cell types.

Numerically, tumour targeting *in vivo* did not perfectly correlate with CXCR4+ cell penetration *in vitro*, but a trend was indeed observed. The materials obtained in the wild type K-12 bacteria were the most efficient in reaching the tumoral tissue (**Figure** 3 A), and those fabricated in absence of DnaK (JGT20) performed similarly to the particles produced in Origami B. Finally, the particles obtained in ClpP[-] (JGT19) targeted the tumour only poorly (**Figure** 3 A).

Data presented in the present study reveals a bottom-up instructive quality control in the biofabrication of smart protein complexes that determines not only

product properties at molecular level but also the macroscopic performance of the material (**Figure** 3 D, **Figure** 4). Subtle conformational variations in the oligomer, while keeping the whole particulate architecture alter the exposure of the overhanging tails, profoundly influencing the behaviour of the material in biological interfaces both *in vitro* and at systems level. For instance, using a DnaK-deficient strain for particle production might increase the conformational variability of the ligand and therefore, expand the spectrum of specific CXCR4+ target organs reached by the material upon delivery. Among other parameters that are convenient to be tuned in drug delivery, constricting of relaxing the biodistribution of a vehicle among target organs might offer interesting therapeutic opportunities for personalized medicine. Therefore, the engineering of the factory's quality control should allow the tuning and improvement of complex and specialized materials' functions in the targeting driven by structured homing peptides. Solving the causative link between particular mutations in the cell and the biological activities of the resulting material would permit a rational bottom-up design of material properties. Even assuming that this might be an unaffordable task, the empirical and systematic screening of the factory's genetic background regarding product properties would pave the way to the dramatic improvement of protein based materials for therapeutic applications.

Table 1. Biodistribution of CXCR4-targeted protein nanoparticles upon systemic administration.

	Whole Tumor [1]	Tumor sections	Liver	Kidney	Lung	Brain sections
Buffer	0.9 ± 0.15 [a, b, c]	1.5 ± 0.40 [d, e, f, g]	1.3 ± 0.11	0.52 ± 0.04	0.63 ± 0.04	1.027 ±0.07 [i]
Origami B (OmpT⁻, Lon⁻ TrxB⁻, Gor⁻)	1.5 ± 0.15 [a]	4.2 ± 0.75 [d, h]	1.3 ± 0.08	0.51 ±0.03	0.59 ± 0.05	0.96 ± 0.05 [j]
MC4100 (wt) [k]	2.7 ±1.10 [b]	6.9 ±1.07 [e, h]	1.80 ± 0.10	0.61 ± 0.07	0.92 ± 0.08	1.27 ±0.04 [k]
JGT4 (ClpA⁻)	1.29 ± 0.29	3.86 ±1.27	2.00 ± 0.26	0.69 ± 0.03	0.95 ± 0.12	1.25 ± 0.03 [l]
JGT19 (ClpP⁻)	0.91 ± 0.07	3.0 ± 0.10 [f]	1.96 ± 0.34	0.77 ± 0.03	0.78 ± 0.05	1.32 ±0.02 [m]
JGT20 (DnaK⁻)	1.47 ±0.14 [c]	4.27 ± 1.07 [g]	1.82 ± 0.06	0.89 ± 0.05	0.89 ±0.08	7.73 ± 1.53 [i, j, k, l, m]

[1] Fluorescence emitted by tumors and normal organs 5 h after the administration of 500 µg of each nanoparticle, expressed as mean ± se of Radiant efficiency (x e^9; [p/sec/cm^2/sr)]/µW/cm^2).

[a, b, c, d, e, f, g, h, i, j] Statistical significance between groups: a, p= 0.003 b, p=0.024; c, p=0.028; d, p= 0.028; e, p=0.025; f, p=0.025; g, p=0.009; h, p= 0.041; i, p= 019, j, p=0.003; k, p=0.025;l, p= 0.014; m, p=0.025.

[k] Wild type regarding protein quality control.

Reference List

[1]   *FDA. Drug Bull.* **1982**, *12*, 18.

[2]   N. Ferrer-Miralles, J. Domingo-Espin, J. L. Corchero, E. Vazquez, A. Villaverde, *Microb. Cell Fact.* **2009**, *8*, 17.

[3]   A. Z. Wang, R. Langer, O. C. Farokhzad, *Annu. Rev. Med.* **2012**, *63*, 185.

[4]   T. E. Stinchcombe, *Nanomedicine. (Lond)* **2007**, *2*, 415.

[5]   F. M. Foss, *Ann. N. Y. Acad. Sci.* **2001**, *941*, 166.

[6]   M. S. Hershfield, *Hum. Mutat.* **1995**, *5*, 107.

[7]   H. Wada, I. Imamura, M. Sako, S. Katagiri, S. Tarui, H. Nishimura, Y. Inada, *Ann. N. Y. Acad. Sci.* **1990**, *613*, 95.

[8]   N. Ferrer-Miralles, E. Rodriguez-Carmona, J. L. Corchero, E. Garcia-Fruitos, E. Vazquez, A. Villaverde, *Crit Rev. Biotechnol.* **2013**, *10.3109/07388551.2013.833163 [doi]*.

[9]   A. Gautam, P. Kapoor, K. Chaudhary, R. Kumar, G. P. Raghava, *Curr. Med. Chem.* **2014**, *21*, 2367.

[10] L. Cai, S. C. Heilshorn, *Acta Biomater.* **2014**, *10*, 1751.

[11] U. Unzueta, M. V. Cespedes, E. Vazquez, N. Ferrer-Miralles, R. Mangues, A. Villaverde, *Trends Biotechnol.* **2015**.

[12] J. L. Corchero, E. Vazquez, E. Garcia-Fruitos, N. Ferrer-Miralles, A. Villaverde, *Nanomedicine. (Lond)* **2014**, *9*, 2817.

[13] E. Vazquez, A. Villaverde, *Nanomedicine (Lond)* **2013**, *8*, 1895.

[14] C. J. Roberts, *Curr. Opin. Biotechnol.* **2014**, *30*, 211.

[15] J. Tyedmers, A. Mogk, B. Bukau, *Nat. Rev. Mol. Cell Biol.* **2010**.

[16] U. Unzueta, M. V. Cespedes, E. Vazquez, N. Ferrer-Miralles, R. Mangues, A. Villaverde, *Trends Biotechnol.* **2015**, *33*, 253.

[17] M. V. Cespedes, U. Unzueta, W. Tatkiewicz, A. Sanchez-Chardi, O. Conchillo-Sole, P. Alamo, Z. Xu, I. Casanova, J. L. Corchero, M. Pesarrodona, J. Cedano, X. Daura, I. Ratera, J. Veciana, N. Ferrer-Miralles, E. Vazquez, A. Villaverde, R. Mangues, *ACS Nano.* **2014**, *8*, 4166.

[18] U. Unzueta, M. V. Cespedes, N. Ferrer-Miralles, I. Casanova, Cedano JA, Corchero JL, J. Domingo-Espin, Villaverde A, R. Mangues, Vazquez E, *Int. J. Nanomedicine* **2012**, *7*, 4533.

[19]   U. Unzueta, N. Ferrer-Miralles, J. Cedano, X. Zikung, M. Pesarrodona, P. Saccardo, E. Garcia-Fruitos, J. Domingo-Espin, P. Kumar, K. C. Gupta, R. Mangues, A. Villaverde, E. Vazquez, *Biomaterials* **2012**, *33*, 8714.

[20]   E. Garcia-Fruitos, M. Martinez-Alonso, N. Gonzalez-Montalban, M. Valli, D. Mattanovich, A. Villaverde, *J. Mol. Biol.* **2007**, *374*, 195.

[21]   N. Ferrer-Miralles, J. L. Corchero, P. Kumar, J. A. Cedano, K. C. Gupta, A. Villaverde, E. Vazquez, *Microb Cell Fact.* **2011**, *10*, 101.

[22]   F. Yang, L. G. Moss, G. N. Phillips, Jr., *Nat. Biotechnol* **1996**, *14*, 1246.

[23]   The Fluorophore of Green Fluorescent Protein (GFP), in **2015**.

[24]   Y. Yan, G. Marriott, *Curr. Opin. Chem. Biol.* **2003**, *7*, 635.

[25]   D. M. Chudakov, M. V. Matz, S. Lukyanov, K. A. Lukyanov, *Physiol Rev.* **2010**, *90*, 1103.

[26]   R. E. Serda, B. Godin, E. Blanco, C. Chiappini, M. Ferrari, *Biochim. Biophys. Acta* **2010**.

[27] B.J. Spencer, I.M. Verma, *Proc. Natl. Acad. Sci. U. S. A* . **2007** *104*, 7594

[28] T. Su, T.T. Stanley. *J. Cell Sci*. **1998**, *111* (Pt 9), 1197
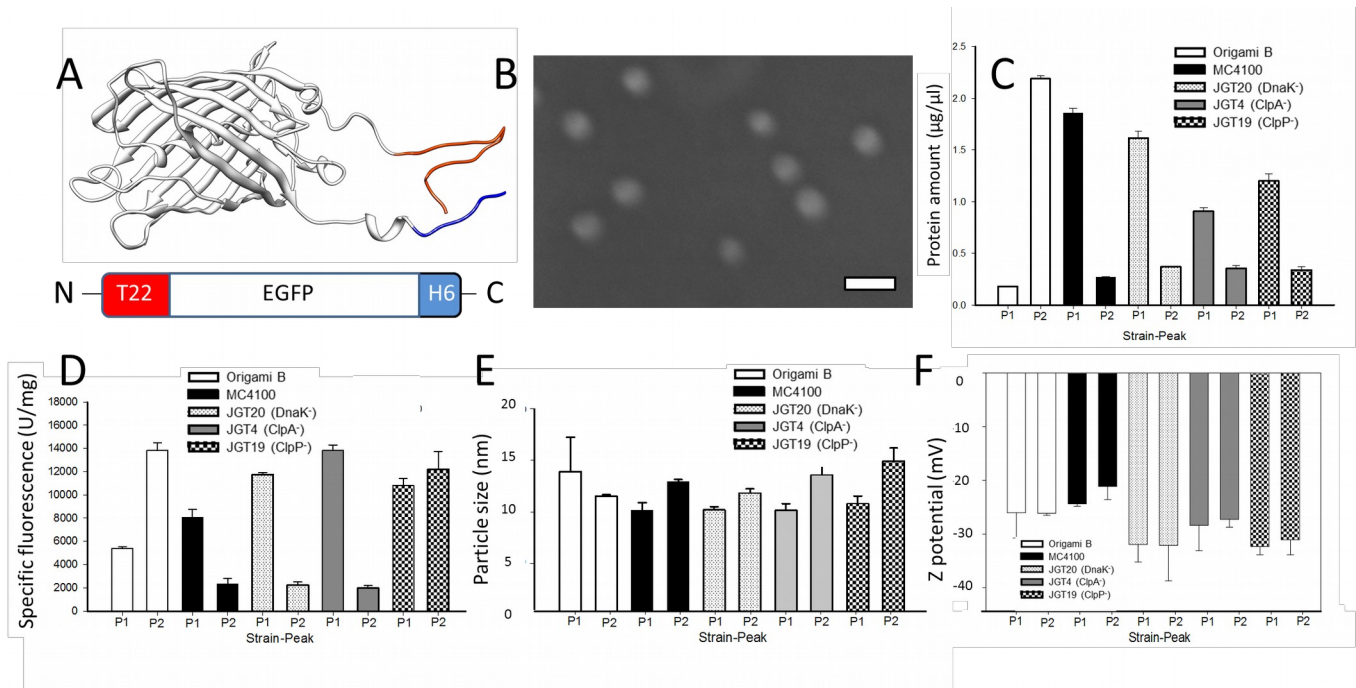
Figure 1. Biophysical properties of T22-empowered protein nanoparticles. A) Molecular modelling of a T22-GFP-H6 monomer, indicating the overhanging end terminal tails in orange (peptide T22) and in blue (hexahistidine), respectively. At the bottom, a cartoon summarizes the modular nature of the protein. Lengths of the modules are here shown as approximate. B) FESEM image of isolated T22-GFP-H6 nanoparticles produced in Origami B cells, upon purification from cell extracts. Bar indicates 50 nm. C) Amounts of T22-GFP-H6 determined upon chromatographic purification from bacterial cell extracts, in each of the observed peaks of elution (peak 1, P1; peak 2, P2). See Supplementary Figure 1A for more details. Protein production was induced in different strains under homogenous conditions. Specific fluorescence emission (D), size (E) and Zeta potential (F) of T22-GFP-H6 nanoparticles purified from different *E. coli* strains. MC4100 was used here as a wild type strain, while their isogenic derivatives JGT20 (DnaK⁻), JGT4 (ClpA⁻) and JGT19 (ClpP⁻) are knock-out mutants in main quality control functions. *E. coli* Origami B, that favours disulphide bridge formation, was also used as reference. P1 and P2 materials were analysed

separately. A summary of the experimental methods is offered in the Supplementary

Information file.
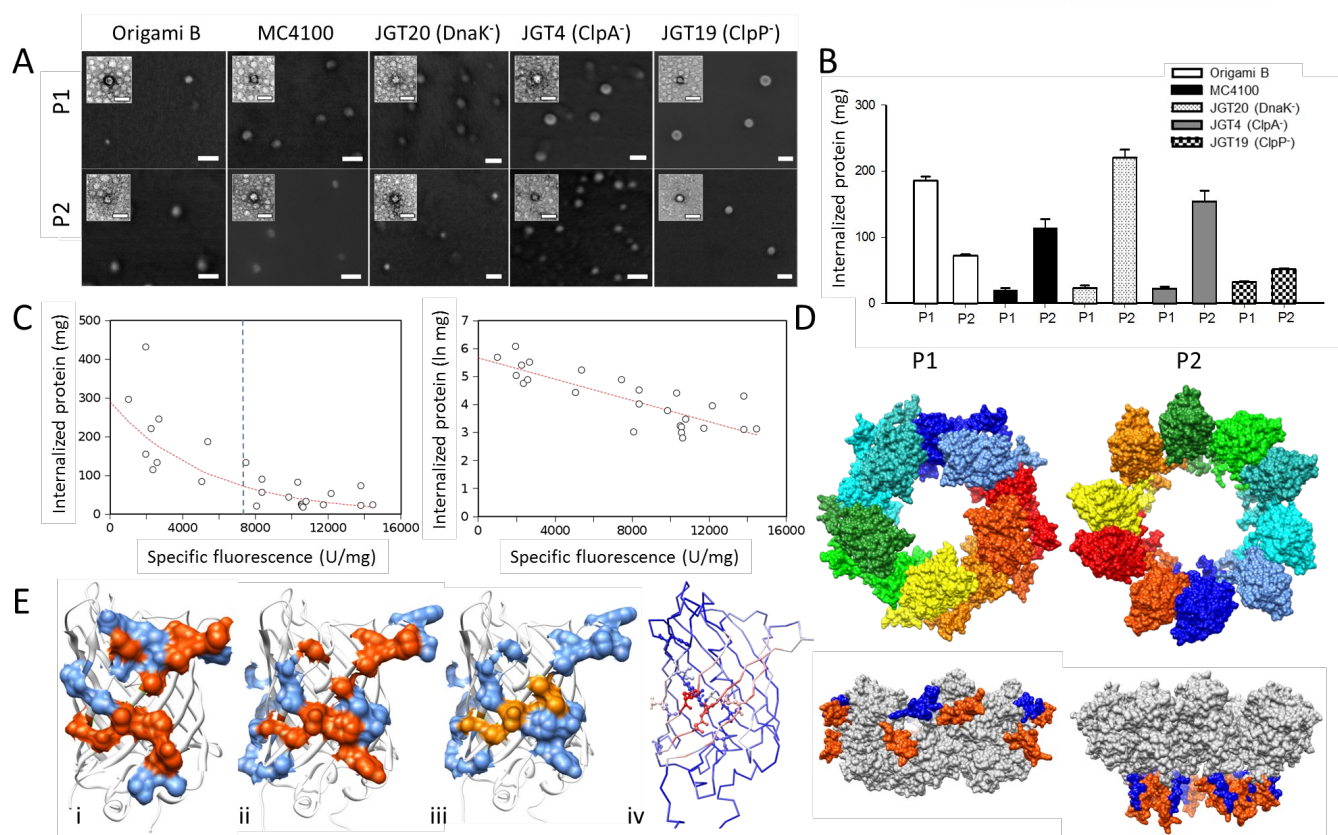
Figure 2. Biological performance of T22-empowered protein nanoparticles. A) Toroid-like organization of T22-GFP-H6 nanoparticles produced in different *E. coli* strains, and eluted as P1 and P2 fractions. Particles are visualized in wide FESEM fields. TEM images of individual particles are also depicted in the insets. Bars indicate 50 nm. Microscopy determination confirmed the slightly smaller size of P1 compared to P2 particles (in nm: TEM: 16.76±0.37 *vs* 16.79±0.38, *U*=4253.5, *p*=0.953; FESEM: 16.69±0.49 versus 17.56±0.41, *U*=4989.0, *p*=0.186). B) Internalization of T22-GFP-H6 nanoparticles in HeLa cells. Intracellular fluorescence was recorded upon harsh trypsin treatment to remove externally attached protein, and corrected by specific fluorescence of each nanoparticle to estimate the protein amount in mass. C) Materials' internalization (as raw data, left and as logarithm, right) is represented versus specific fluorescence of each particle. Both P1 and P2 materials were dialyzed against low and high salt buffers, and data were recorded for the full set of

samples. The vertical dashed line indicates the specific fluorescence of GFP-H6. D)

3D surface representations of archetypical P1 and P2 nanoparticles in a top view

(modelled with HADDOCK), obtained enforcing tail-barrel and tail-tail interactions

respectively. Each monomer is differently coloured. At the bottom, side view of the

same models with the T22 peptide coloured in orange and the hexahistidine tail in

blue. Nanoparticle diameters are 14.1 nm for P1 and 16.7 nm for P2. E) 3D

representation of the GFP segment of the T22-GFP-H6 monomer. From left to right: i)

Surface-accessible residues located at the GFP-dimer interface are shown in surface

representation; residues common to (ii) are shown in orange. ii) Surface-accessible

residues with low-penalty desolvation as predicted by EDP analysis are shown in

surface representation; residues common to (i) are shown in orange. iii) Residues

from the GFP barrel used to drive the docking with Haddock are shown in surface

representation; passive residues in blue, active in dark yellow. iv) Backbone coloured

in a gradient from blue to red according to the number of times (from 0 to 56,

respectively) in which a residue is found in an EDP-predicted binding site; residues

known to affect the fluorescence of the protein are shown in ball & stick

representation. For clarity, an expanded version of the model in iv panel is depicted in

the Supplementary **Figure** 3.

.
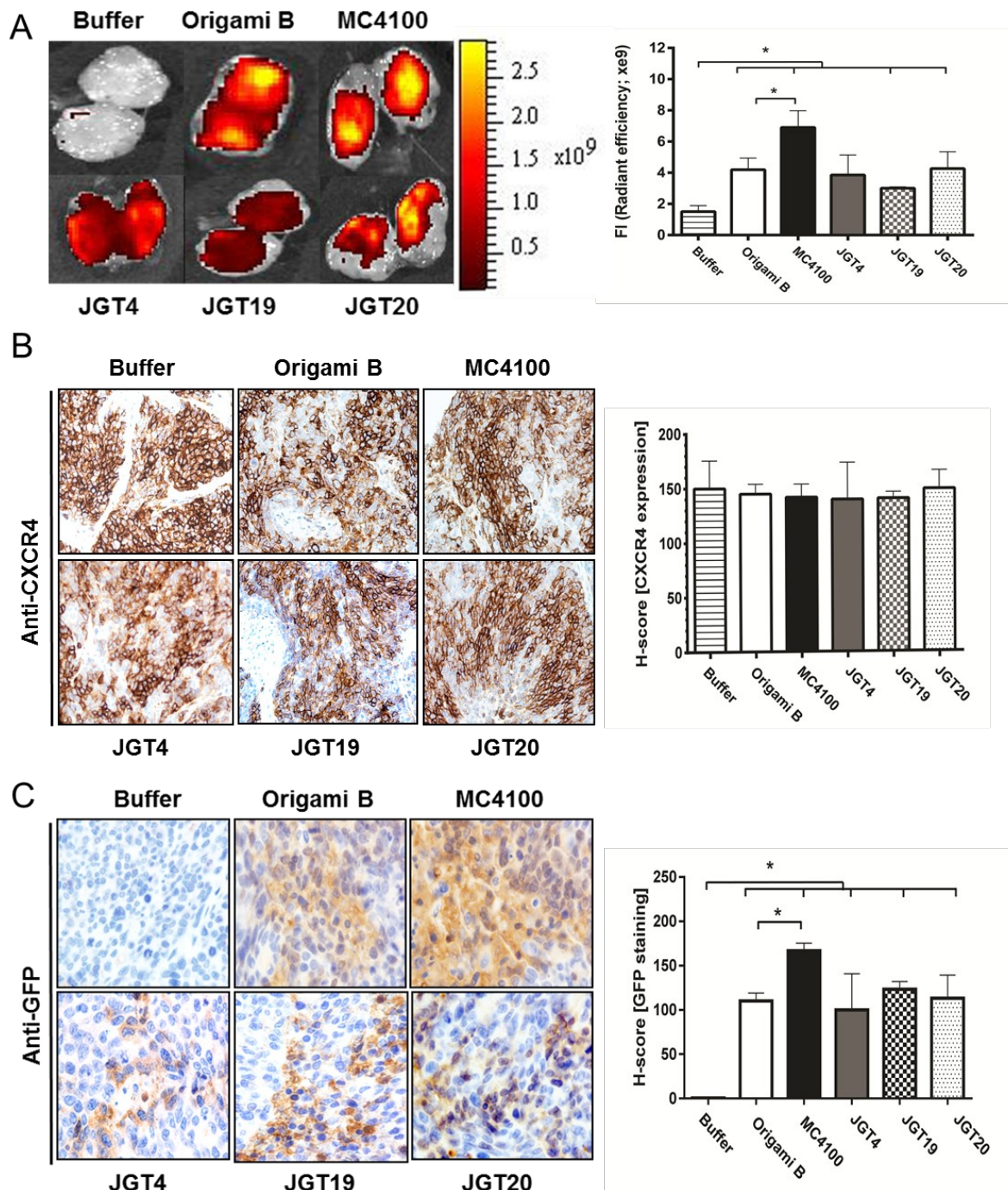
Figure 3. Tumor biodistribution of T22-GFP-H6 nanoparticle variants A) Representative pictures and quantitation of *ex vivo* fluorescence emission in sectioned tumors (detected by the IVIS Xenogen fluorimeter). The color scale bar identifies the intensity of fluorescence emitted by tumor tissue ranges from absent (grey color), low (brown), intermediate (red) or high (yellow color). B) High level of CXCR4 expression in tumors (deep brown color), showing no statistically significant

differences among groups. Displayed are representative colorectal tumor tissues of animals to be administered with the buffer or with materials produced in Origami-B, MC4100, JGT4, JGT19 or JGT20 bacterial strains. CXCR4 expression is detected immunohistochemically (IHC) using an anti-CXCR4 antibody (400 x magnification). C) Representative micrographs of nanoparticle internalization in the tumor cell cytosol, as detected by IHC using an antibody against the nanoparticle GFP domain. The antibody is coupled to peroxidase, and yields a brown precipitate when the target protein is detected), which allows the quantitation of nanoparticle internalization in the different groups (1000 x magnification). Preparations for immunohistochemistry are stained in brown after reacting with an antibody against the target protein and counterstained with hematoxilin, which stains in blue color the cell nuclei. The amount of target protein detected ranges from low (pale brown color) to high (intense brown color). (e.g. CXCR4 receptor in cell surface or GFP domain of the nanoparticle inside the cell cytosol). Notice the correlation between tumor emitted fluorescence (seen in panel A as range of colors) and the IHC detected nanoparticle internalization into the tumor cell cytosol (seen in panel C as range of brown color intensity).
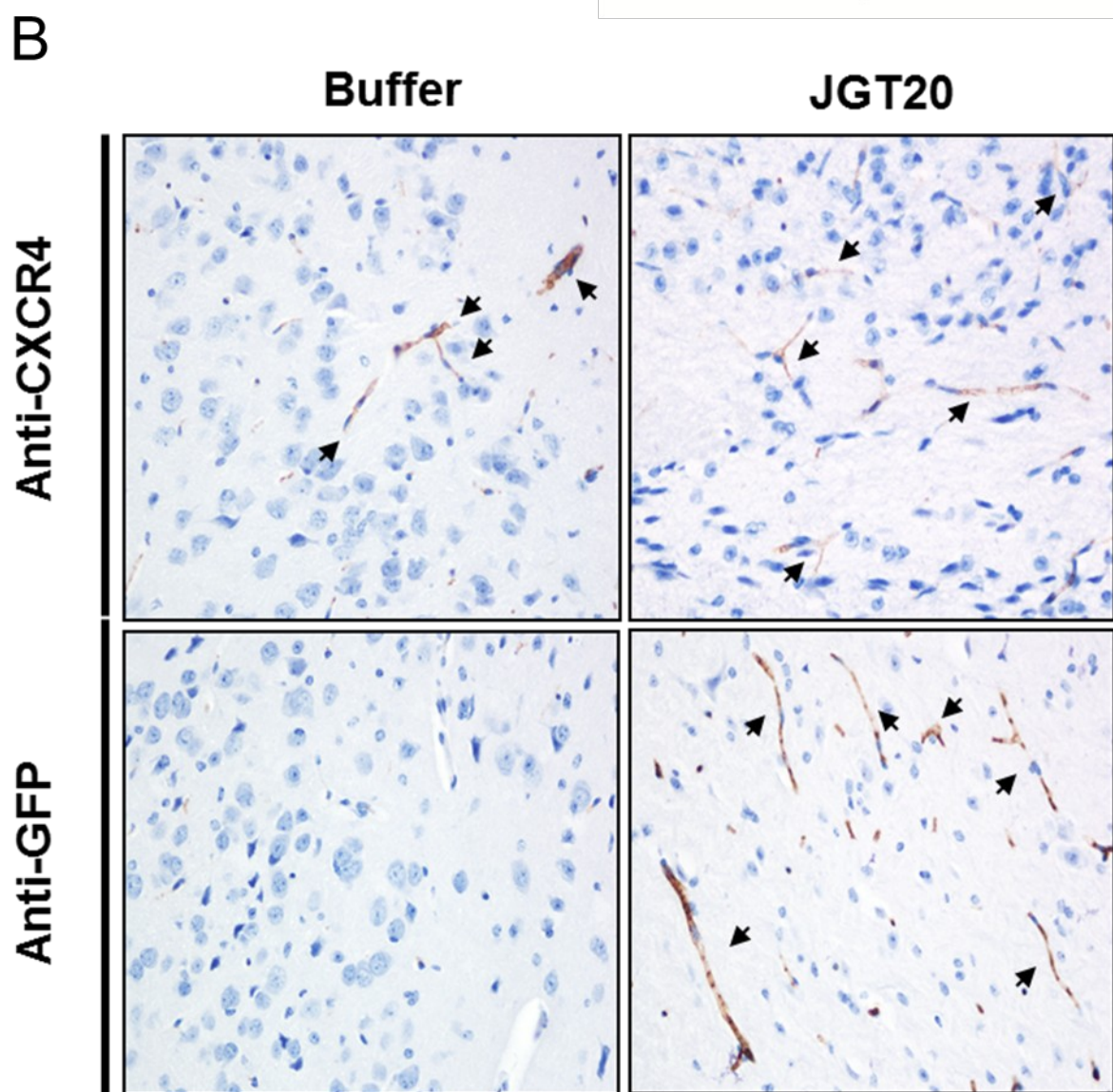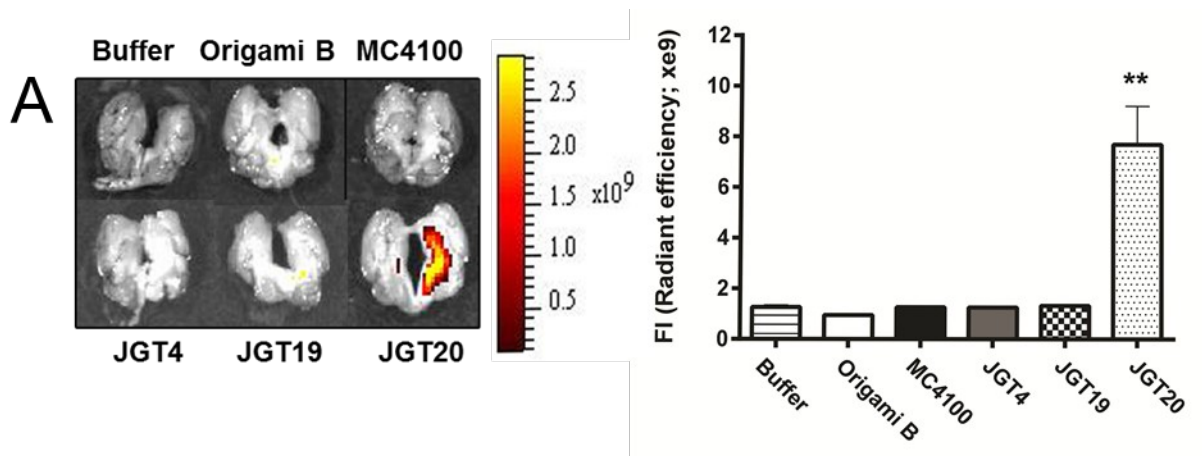
Figure 4. Biodistribution of T22-empowered protein nanoparticles in normal brain. A) Representative fluorescence recording images and quantitation of *ex vivo* fluorescence (measured using the IVIS fluorimeter) in sectioned brain after the administration of buffer or materials produced in Origami B, MC4100, JGT4, JGT19 or JGT20 bacterial strains. B) Representative micrographs of CXCR4 expression in brain microvascular endothelia (upper panel, black arrows), and nanoparticle uptake in CXCR4+ endothelia (lower panel, black arrows), 5 h after the administration of Buffer or JGT20-produced material, as measured by IHC with an anti-GFP antibody (400x magnification). Notice the similar CXCR4 expression in brain endothelia of Buffer and JG20 groups and the exclusive nanoparticle accumulation in the microvasculature in animals treated with nanoparticles produced in JGT20.

Supporting information

Bottom-up instructive quality control in the biofabrication of smart protein materials

Fabián Rueda [1, 2, 3 f]*, María Virtudes Céspedes* [3, 4 f]*, Oscar Conchillo* [1]*, Alejandro Sanchez-Chardi* [5]*, Joaquin Seras-Franzoso* [1, 2, 3 £]*, Rafael Cubarsi* [6]*, Alberto Gallardo* [3, 4]*, Mireia Pesarrodona* [1, 2, 3]*, Neus Ferrer-Miralles* [1, 2, 3]*, Xavier Daura* [1,7]*, Esther Vázquez* [1, 2, 3]*, Elena García-Fruitós* [1, 2, 3 ¥]*, Ramón Mangues* [3, 4 *]*, Ugutz Unzueta* [3, 4]*, Antonio Villaverde* [1, 2, 3 *]

[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, 08193 Cerdanyola del Vallès, Spain.

[2] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, 08193 Cerdanyola del Vallès, Spain.

[3] CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Cerdanyola del Vallès, Spain.

[4] Biomedical Research Institute Sant Pau (IIB-SantPau), Hospital de la Santa Creu i Sant Pau, 08025 Barcelona, Spain.

[5] Servei de Microscòpia, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain.

[6] Departament de Matemàtica Aplicada IV, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain

[7] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

[f] Equally contributed

[£] Present address: Cibbim-Nanomedicine, Hospital Vall d'Hebron, Passeig de la Vall d'Hebron, 119-129, 08035 Barcelona, Spain

[¥] Present address: Department of Ruminant Production, IRTA, Institute of Agrifood Research and Technology. Torre Marimon, km 12,1 C-59, Caldes de Montbui, 08140, Barcelona, Spain

[*]corresponding authors.

Strains, culture conditions and protein purification. We used the *E. coli* K12 strain MC4100 ([*araD139*], (*argF-lac*)169, λ⁻ *relA1*, *rpsL150*, *rbsR22*, *flb5301*, *deoC1*, *pstF25* Strep^R), and its derivatives JGT4 (ClpA⁻; *clpA::kan* Strep^R); JGT19 (*clpP::cat* Strep^R) and JGT20 (*dnak756 thr*::Tn10, Strep^R, Tc^R). Also, *Escherichia coli* Origami B (BL21, OmpT⁻, Lon⁻, TrxB⁻, Gor⁻, Strep^R, Tet^R, Novagen) was used as production control. K12 strains were transformed with pTrc99a (Ap^R, IPTG-inducible while Origami B was transformed with pET22b, both encoding the protein T22-GFP-H6. All strains were cultured in Luria-Bertani (LB) media. [1] Overnight cultures were inoculated in shake flasks containing 500 ml of LB with appropriate antibiotics and incubated at 37ºC and 250 rpm, growing up to 0.5 and 0.6 OD_{550} units. The expression of *T22GFPH6* gene was induced by the addition of 50 µl IPTG at 1 mM and cultures were incubated overnight at 20ºC and 250 rpm. Afterwards, cells were harvested by centrifugation (3,280 g, 4ºC, 40 min) and pellets were resuspended in 25 ml of Wash Buffer pH 8.0 (20 mM Tris-HCl, 500 mM NaCl and 20 mM imidazole), containing an EDTA-free protease inhibitor cocktail (Complete EDTA-free Roche Diagnostics, Indianapolis, USA). Cells were disrupted by pressuring at 1100 psi in a French press (Thermo FA-078A) and proteins were purified by His tag-affinity chromatography using 1 mL HiTrap Chelating HP column (GE Healthcare, Piscataway, NJ) through an AKTA purifier FPLC (GE Healthcare). Separations were made by linear gradient of Tris 20 mM, pH 8.0, 500 mM NaCl, and 500 mM imidazole. Fractions collected were dialyzed against $NaHCO_3$ 160mM pH 7.4 Buffer. Protein amounts were determined by Bradford's assay [2] and analyzed by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) and anti-GFP western blot.

Fluorescence, size particle and zeta potential. Fluorescence of purified protein was measured in a spectrometer Cary Eclipse (Varian, Mulgrave Australia) using 1 ml cuvettes, at 450 nm of excitation wavelength and 510 nm of emission wavelength. The volume and size distribution of nanoparticles in buffer $NaHCO_3$, as well as zeta potential, were measured by dynamic light scattering at 633 nm through a Zetasizer Nano ZS (Malvern Instruments Limited, Malvern, Worcestershire, UK) using disposable plastic cuvettes. Nanoparticles samples were analyzed by triplicate averaging fifteen single measurements.

Electron microscopy. Nanoparticles were analysed by transmission electron microscopy (TEM) and field emission scanning electron microscopy (FESEM). For TEM, protein samples were negatively stained with uranyl acetate by conventional methods [3] and observed in a Jeol 1400 microscope operating at 80 kV and equipped with a CCD Gatan Erlangshen ES1000W camera. For the quantification of nanoparticles size, 185 particles were measured using Gatan Digital Micrograph software. For FESEM, protein samples were directly deposited over silicon wafers, air dried and observed with an in-lens secondary electron detector through a Zeiss Merlin microscope operating at 2kV. For quantification of nanoparticle size distribution, a total of 214 nanoparticles were analysed with ImageJ software. Statistical differences of quantitative analyses between P1 and P2 were calculated by Mann-Whitney tests (U) using SPSS 15.0 software.

Protein internalization in cell culture. Protein internalization was analyzed in sub-confluent HeLa cell cultures in 12 well-plates (Nunclon™ Delta, Roskilde, Denmark). Briefly, medium was removed and cells were washed in PBS. Then 250 µl of 25 nM T22-GFP-H6 in OptiPro (Gibco, Paisley, UK), supplemented with L-Glutamine, were

added and incubated for 1 h at 37 ºC to allow cell binding and internalization. After incubation, trypsin digestion (1 mg/ ml for 15 min), was carried out. Trypsin was neutralized by the addition of 2 volumes of regular cell culture medium and samples were centrifuged at 1200 rpm for 5 min. Finally, pellets were resuspended in 300 µl of PBS and intracellular green fluorescence was analyzed by flow cytometry in a FACSCanto system (Becton Dickinson, Franklin Lakes, NJ), using a 15 W air-cooled argon-ion laser at 488 nm excitation for GFP. Typically, data were recorded in duplicate from 60,000 cell counts.

Modelling. The T22GFPH6 monomer was modelled by homology using Modeler 9v13 [4] and the following templates: the peptide polyphemusin I structure (pdb code 1RKK model 1) [5] for T22 (T22-GFP-H6 residues 2 to 19; 74 % identity); residues 40 to 49 of the globular domain of *Gallus gallus* histone H5 (pdb code 1HST) [6] (residues 17 to 26; 80 % identity) and the structure with pdb code 1QYO for GFP [7] (residues 27 to 262; 98 % identity). The histidine tail was modelled by Modeller's automodel function (residues 263 to 269). 500 models were generated and sorted by their DOPE score. [8] Models with "knots" [9] were removed and the one with the best per-residue score was selected. 20 models with best DOPE score were analysed with the Electrostatic-Desolvation-Profile method, [10] after removal of the N and C terminal tails, to predict the binding patch in the T22-GFP-GFP barrel. Residues from the patch with a surface accessible area greater than 40 % were used as Ambiguous Interaction Restraints (AIR) in HADDOCK. Those in the center of the patch, which were also more frequently predicted by EDP (**Figure** 2E), were selected as active residues while the rest were used as passive.

To construct the nanoparticle models several runs of HADDOCK were performed, enforcing C5 symmetry and using different active and passive residues as AIR. Three different combinations where used: T22-tail residues (1 to 25 in the T22-GFP-H6 monomer) as actives and His-tail residues (262 to 269) as passives; only T22 residues as actives; T22-tail residues as actives and EDP-predicted residues as actives and passives as previously explained. Histidines were protonated according to their pKas and pH 7.4 (same used for microscopy sample preparation) using the protonate3D [11] function from the MOE package. [12] All generated models where clustered as explained in the HADDOCK tutorial [13] and visually inspected with Rasmol [14], which was also used for measurements. 3D representations shown in figures have been generated with UCSF Chimera. [15] Diameters were calculated using the two barrel alpha-carbons farthest apart in the oligomer structure.

Nanoparticle biodistribution in the CXCR4+ tumor model. CXCR4-overexpressing SP5 human colorectal cancer line was implanted to generate subcutaneous tumors in Swiss nude mice (Charles River, France), as previously described. [3] When tumors reached ca. 500 mm^3 mice were randomly allocated to Origami B, MC4100, JGT4, JGT19, JGT20 or buffer-treated groups (N=3-5/group). The study was approved by the Institutional Animal Ethics Committee (protocol Nº DAAM: 8339). The experimental mice received a single 500 µg intravenous bolus of the corresponding nanoparticles in carbonate pH 7.5 buffer, whereas control mice received only buffer. At 5 h post-administration, the fluorescence emitted by the nanoparticles accumulated in the whole and slice sectioned tumor and normal tissues (kidney, lung, and heart, liver and brain) was measured *ex vivo* using IVIS® Spectrum equipment (Xenogen Biosciences, USA). The fluorescence signal was then digitalized, subtracting the autofluorescence, displayed as a pseudocolor overlay and

expressed as Radiant efficiency. Data was corrected by the specific fluorescence emitted by the different nanoparticles.

Histological and immunohistochemical (IHC) analysis. Tumors were fixed and paraffin-embedded, cut into 4 µm sections, processed as previously described [3, 16] and H&E stained for histological analysis by two independent observers. CXCR4 membrane expression and nanoparticle cell internalization in tumor and normal tissues was assessed by IHC using primary anti-CXCR4 (1:300; Abcam, UK) or anti-GFP (1:100; Santa Cruz Biotechnology, USA), and secondary HRP conjugated antibody, followed by chromogenic detection. [3] The percent of CXCR4-expressing cells in relation to the total cell number and their staining intensity was cuantified, scoring each from 0 to 3 (where 3 is the maximal intensity) and multiplying both values to obtain the H-score.

A

B

C
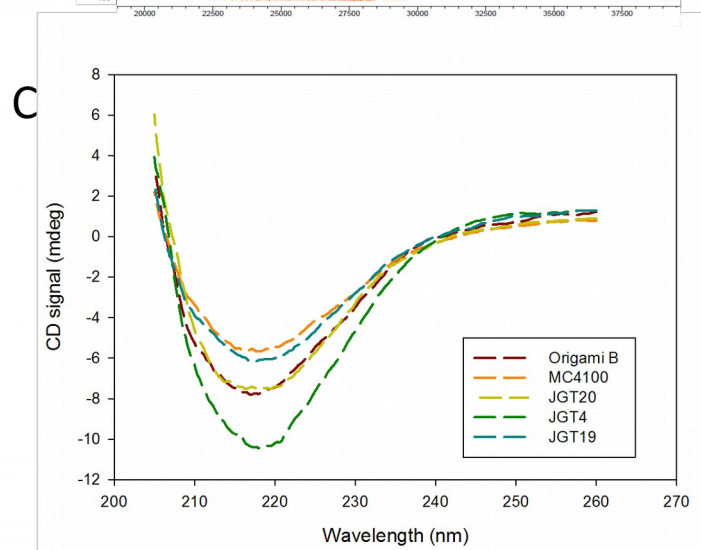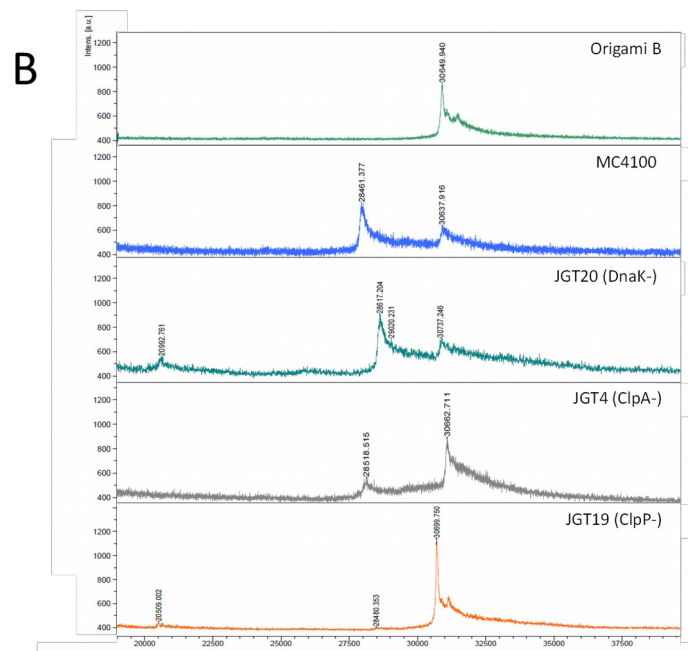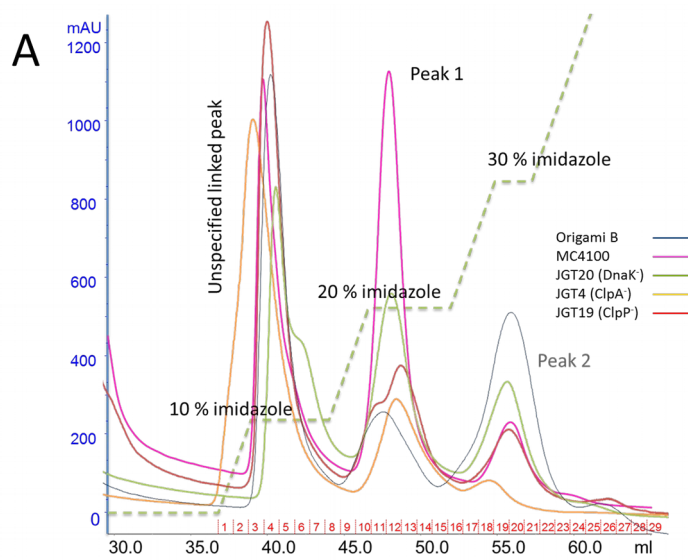
Supplementary Figure 1. Protein purification and preliminary characterization. A) Two protein peaks (P1 & P2) were observed in the separation of T22-GFP-H6 produced in different *E. coli* strains, by affinity chromatography against an imidazole concentration gradient. B) MS spectra of T22-GF-PH6 eluted in P2. Except for Origami B, the materials from all strains separated into two major peaks corresponding to the molecular weight of the full-length T22-GFP-H6 (30.6 KDa) and to a shorter species (~28.5 KDa). C) Circular Dichroism spectra from 260 nm to 205 nm of T22-GFP-H6 of P2 materials produced in different strains. A peak at 218 nm is observed corresponding to beta-sheet secondary structure signal.
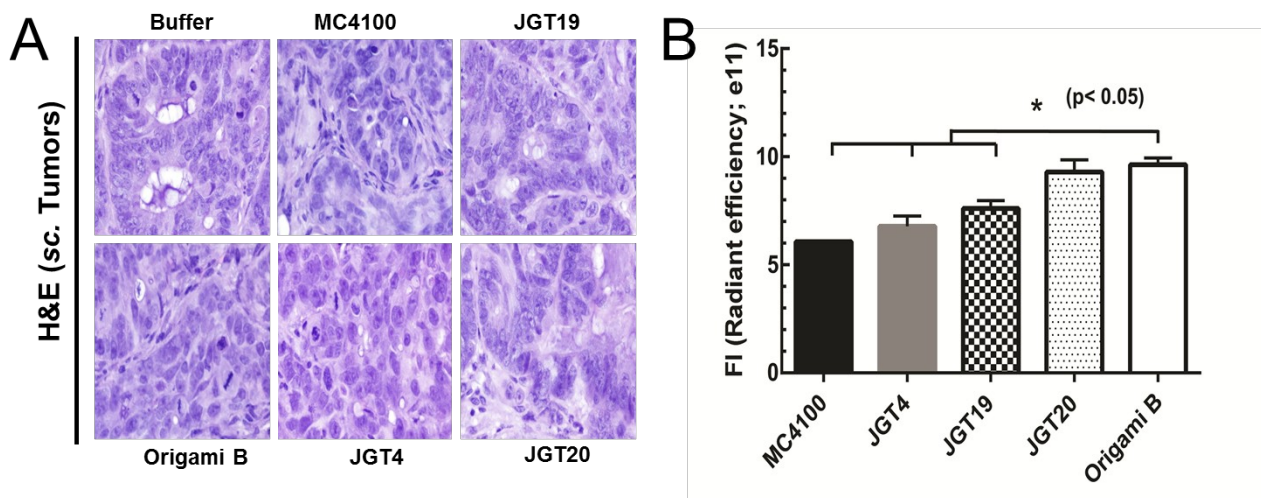
Supplementary Figure 2. A) Haematoxylin-eosin stained of SP5 subcutaneous colorectal tumors showed a similar architecture and histology among tumors. B) Significant differences in specific fluorescence among nanoparticle variants, as measured using the IVIS spectrum equipment.
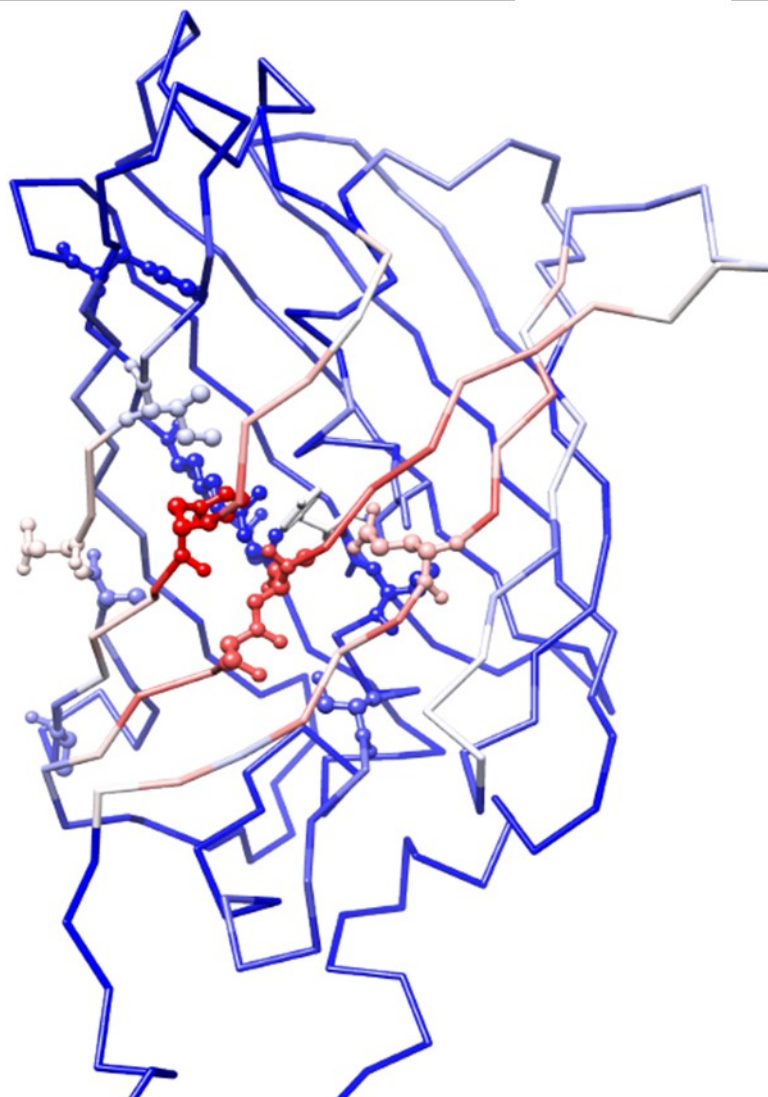
Supplementary Figure 3. 3D representation of the GFP segment of the T22-GFP-H6 monomer. Backbone coloured in a gradient from blue to red according to the number of times (from 0 to 56, respectively) in which a residue is found in an EDP-predicted binding site. Residues known to affect the fluorescence of the protein are shown in ball & stick representation.

Reference List

[1] J. Sambrook, E. Fritsch, T. Maniatis, *Molecular Cloning, A Laboratory Manual, ,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY **1989**.

[2] M. M. Bradford, *Anal. Biochem.* **1976**, *72*, 248.

[3] M. V. Cespedes, U. Unzueta, W. Tatkiewicz, A. Sanchez-Chardi, O. Conchillo-Sole, P. Alamo, Z. Xu, I. Casanova, J. L. Corchero, M. Pesarrodona, J. Cedano, X. Daura, I. Ratera, J. Veciana, N. Ferrer-Miralles, E. Vazquez, A. Villaverde, R. Mangues, *ACS Nano.* **2014**, *8*, 4166.

[4] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper, A. Sali, *Curr. Protoc. Bioinformatics* **2006**, *Chapter 5*, Unit.

[5] J. P. Powers, A. Rozek, R. E. Hancock, *Biochim. Biophys. Acta* **2004**, *1698*, 239.

[6] V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, R. M. Sweet, *Nature* **1993**, *362*, 219.

[7] D. P. Barondeau, C. D. Putnam, C. J. Kassmann, J. A. Tainer, E. D. Getzoff, *Proc. Natl. Acad. Sci. U. S. A* **2003**, *100*, 12111.

[8] M. Y. Shen, A. Sali, *Protein Sci.* **2006**, *15*, 2507.

[9] Modeller, https://salilab.org/modeller/tutorial/, in **2015**.

[10] S. Fiorucci, M. Zacharias, *Biophys. J.* **2010**, *98*, 1921.

[11] P. Labute, *Proteins* **2009**, *75*, 187.

[12] Molecular Operating Environment (MOE) software, MOE, Chemical Computing Group Inc., in **2015**.

[13] Haddock, http://haddock.chem.uu.nl/enmr/haddock-tutorial.php, in **2015**.

[14] R. A. Sayle, E. J. Milner-White, *Trends Biochem. Sci.* **1995**, *20*, 374.

[15] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605.

[16] M. Pesarrodona, N. Ferrer-Miralles, U. Unzueta, P. Gener, W. Tatkiewicz, I. Abasolo, I. Ratera, J. Veciana, S. S. Jr, A. Villaverde, E. Vazquez, *Int. J. Pharm.* **2014**, *473*, 286.

Graphical abstract
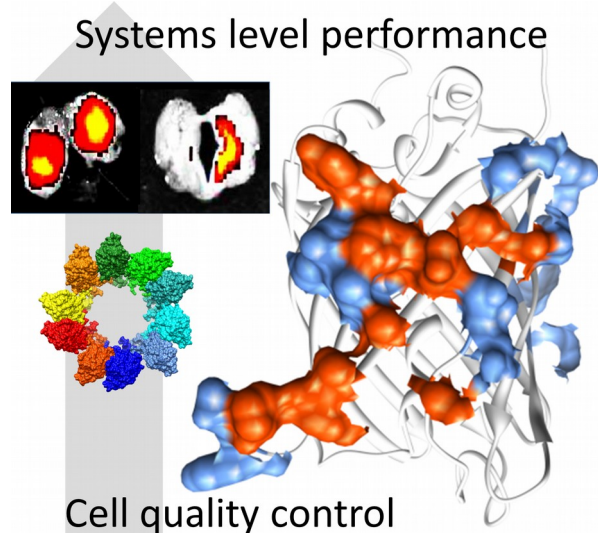
How the quality control of producing cells determines material's properties is a neglected but critical issue in the fabrication of protein biomaterials, which are unique in merging structure and function. The molecular chaperoning of protein's conformational status is revealed here as a potent molecular instructor of the macroscopic properties of self-assembling, cell-targeted protein nanoparticles, including biodistribution upon in *vivo* administration