SCUOLA NORMALE SUPERIORE

RELAZIONE ATTIVITA' ANNUALE DEI PERFEZIONANDI/DOTTORANDI – TERZO/QUARTO ANNO
REPORT ON THE PHD ACTIVITY – THIRD/FORTH YEAR

NOME E COGNOME NAME AND SURNAME	Elisa Martino
DISCIPLINA PHD COURSE	Nanosciences

CORSI FREQUENTATI CON SOSTENIMENTO DI ESAME FINALE ATTENDED COURSES (WITH FINAL EXAM)	**VOTAZIONE RIPORTATA MARK**	**NUMERO DI ORE HOURS**
----	---	---

CORSI FREQUENTATI SENZA SOSTENIMENTO DI ESAME FINALE ATTENDED COURSES (ATTENDANCE ONLY)	**NUMERO DI ORE HOURS**
----	---

ALTRE ATTIVITÀ FORMATIVE (SEMINARI, WORKSHOP, SCUOLE ESTIVE, ECC.) – DESCRIZIONE OTHER PHD ORIENTED ACTIVITIES (SEMINARS, WORKSHOPS, SUMMER SCHOOLS, ETC) – DESCRIPTION	**NUMERO DI ORE HOURS**
23/10/2019 - Dr. Sacha De Carlo - IIT Seminar – "The current state of Micro-Electron Diffraction"	2
28/04/2020 (online) - Prof. Fiorella Meneghetti – "An overview of X-ray diffraction methods in medicinal chemistry"	4
22-25/09/2020 – Online School GeCry School - From Gene to Protein Crystal Structure	34

SCUOLA
NORMALE
SUPERIORE

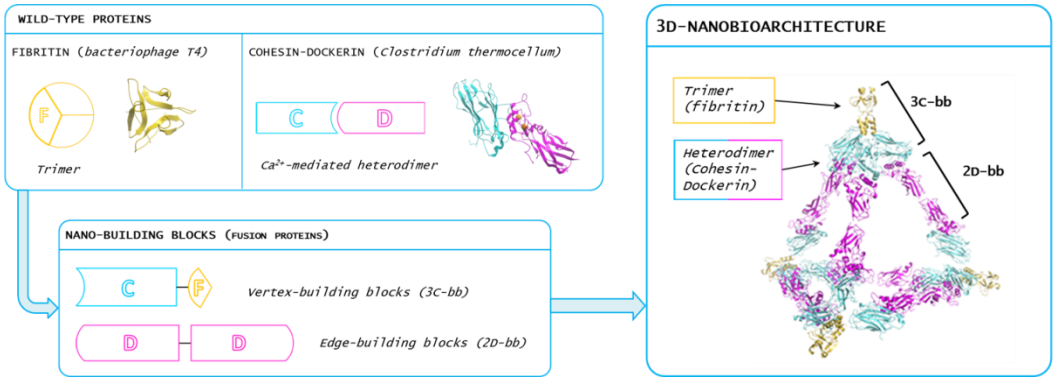
ATTIVITÀ DI RICERCA SVOLTA (MAX. 8.000 CARATTERI)*
RESEARCH ACTIVITY (MAX. 8000 CHARACTERS)

Full report attached.

Summary:

My PhD project aims to generate a novel bioinspired scaffold by simple engineered protein-scaffolds. Such structures are designed to self-assemble upon stimulation into desired well-ordered and stable multicomponent nanoarchitectures.

Specifically, we are developing a scalable tetrahedral protein cage, exploiting the fusion between two distinct oligomeric subunits. These consist in three protein domains (Fibritin, Cohesin and Dockerin), whose association in C2+C3 symmetry axes provides a final tetrahedral bioarchitecture having scalable ~16 nm-edges.



In the past year, I investigated the optimization and development of our designed protein cage by different assemblies of the nano-building blocks I had produced during my second year.

The first assembly strategy tested, comprising a chemical crosslinking reaction as crucial part of the assembly, proved to be much more challenging than expected (see full report, section 2.1). The chemical crosslinking process turned out to be not very efficient Even if some improvement to this approach was obtained, the final desired structure had not been isolated and identified yet. Therefore, we are focusing on the crystallization and structural characterization of the single subunits, with the aim of investigating their atomic features and to use a structure-based redesign for successful crosslinking.

In parallel to the further optimization of the first method, we started the development of non-chemical "biomimetic" approach (see full report, section 2.2). We developed many new nano-building blocks (2D-bb), and we have started to test them for the final assembly process. We have not yet identified a building block giving unambiguously the tetrahedral assembly, but we have planned, in the next months, to analyze the newly obtained assemblies by negative staining and to select the most promising ones, before proceeding to a structural characterization.

For the next year, our intention is to try to conclude the optimization of the final assembled protein cage. We still have some strategies left to test, other than the ones that we are currently investigating.

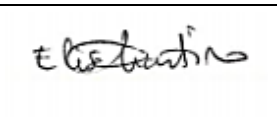
*se si intende sottoporre una relazione di ricerca più estesa, utilizzare il campo per una descrizione sintetica e allegare il documento in formato .pdf

If you are going to submit a longer report, please fill the box with a synthetic abstract and attach a document in pdf format

SCUOLA
NORMALE
SUPERIORE

EVENTUALI PUBBLICAZIONI PUBLICATIONS (IF AVAILABLE)

NOME DEL RELATORE THESIS ADVISOR
Dr. Gianpiero Garau (CNI@NEST) Dr. Stefano Luin (SNS)

DATA DATE	15/10/2020	FIRMA SIGNATURE	

ENGINEERING OF A SCALABLE TETRAHEDRAL NANO-BIOARCHITECTURE

1. Introduction

Oligomeric protein complexes are widely diffused in nature (e.g. viral capsids, bacterial microcompartments). This natural protein assemblies have inspired protein engineers to design and generate novel 3D structural arrangements having symmetrical bioarchitectures. The engineered biostructures can adopt "cage-like" properties, to encapsulate bio-organic compounds and large biomolecules for specific applications in therapy, sensing and biocatalysis.[1] During the design of novel protein biocages, it is hard to predict the optimal interfaces between oligomeric components, as engineered proteins often show ineffective folding process, limited expression and proper assembly. Moreover, protein assemblies are difficult to crystallize and characterize structurally, due to their porous architecture.[2,3]

My PhD project aims to generate a novel bioinspired scaffold by simple engineered protein-scaffolds. Such structures are designed to self-assemble upon stimulation into desired well-ordered and stable multicomponent nanoarchitectures.
Specifically, we are developing a scalable tetrahedral protein cage, exploiting the fusion between two distinct oligomeric subunits (Figure1). These consist in three protein domains (Fibritin, Cohesin and Dockerin), whose association in C2+C3 symmetry axes provides a final tetrahedral bioarchitecture having scalable ~16 nm-edges.

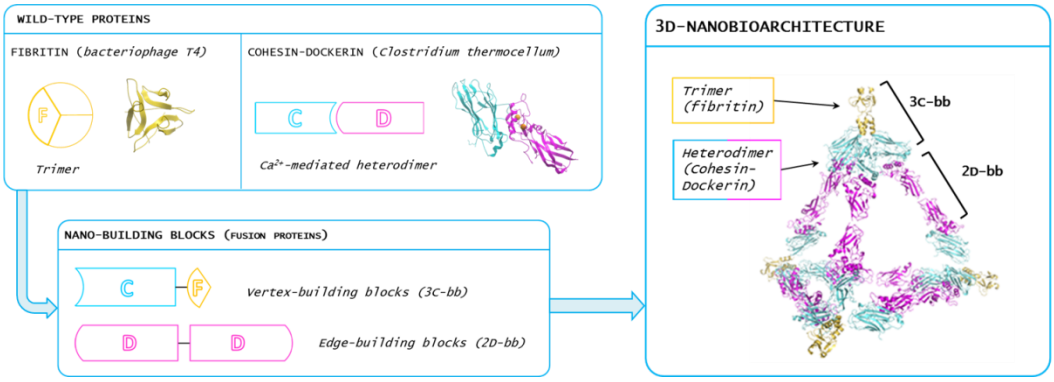


Figure 1 Project outline: 3D nano-bioarchitecture design

2. Results and Discussion

Last year I focused on the development of the supramolecular assembly of the nano-building blocks I had produced during my second year. Such building blocks were designed to assemble in an overall tetrahedral fold. The vertex building block (3C-bb) is composed of the Cohesin protein engineered with a trimeric domain (Fibritin),[4] giving overall a trimeric subunit, which should constitute the vertex of the final pyramid. The edge building-block (2D-bb) is a dimeric subunit, with two units of Dockerin protein joined together to face in opposite directions. The final 3D assembly is given by the specific interface-recognition of the two building-blocks between the Cohesin and Dockerin subunits, which form a calcium-mediated heterodimer with very high affinity ($K_D \sim 10^{-9}$).[5,6]

In our initial design of the 2D-bb, two dockerin units were joined together by chemical linkage (dimaleimide-thiol crosslinking). This initial approach had the advantage of an easier design, but also to impart a higher rigidity to the junction between the two dockerin units, through a cysteine-mediated covalent bond between the two moleculat units. This was supposed to favour the proper orientation of the dockerin towards the 3C-bb, to provide the final desired 3D assembly.

We have extensively tested chemical crosslinking as a tool for the construction of the nano-architecture (see section 2.1 Chemically crosslinked architecture). The chemical crosslinking process turned out to be not very efficient, and we were not able to isolate and characterized the final assembled structure. Therefore, we are focusing on the crystallization and structural characterization of the single subunits, with the aim of investigating their atomic features and to use a structure-based redesign for successful crosslinking.

As a backup strategy for the generation of the 2D-bb module, we started the development of non-chemical "biomimetic" approach. We engineered the dockerin protein with a dimerization domain at the N-terminus, in order to constitute an homodimer by heterologous expression. We are currently screening several dimerization domains, in order to find the best to trigger the desired final structure (see section 2.2 Design and Screening of homodimeric-dockerins). We have not yet identified a dimeric domain giving unambiguously the tetrahedral assembly, but we are still testing our possibilities.

2.1 Chemically crosslinked architecture

The use of the chemical crosslinking for the formation of the 2D-bb provides the possibility to direct the assembly in two different ways (Figure 2):

A) Heterodimer-driven assembly: 1) the chemical crosslinking on the isolated dockerin is first performed, therefore the 2D-bb is synthesized previously than the assembly with the 3C-bb takes place; 2) then, 3C-bb and 2D-bb are mixed as the final assembly step. The assembly is therefore driven by the heterodimer formation between the cohesin and dockerin domains.

B) Crosslinking-driven assembly: 1) the dockerin monomer and the 3C-bb are initially mixed, to form the heterodimers; 2) then, the chemical crosslinking is conducted. In this way, the 3D assembly is mediated by chemical crosslinking, which locks the final 3D structure.

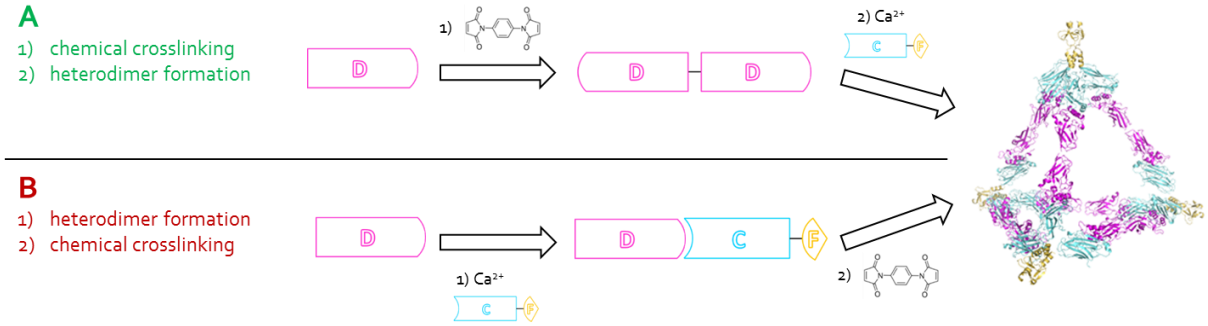


Figure 2 Assembly strategies with chemical crosslinking

The first approach (A) was extensively studied, under many conditions of pH, buffer, temperature and concentration. The most critical parameter for the outcome of the assembly was the concentration. In all cases, though, it was clear that the assembly was not giving exclusively the desired structure, but instead a mixture of high molecular weight structures (Figure 3).

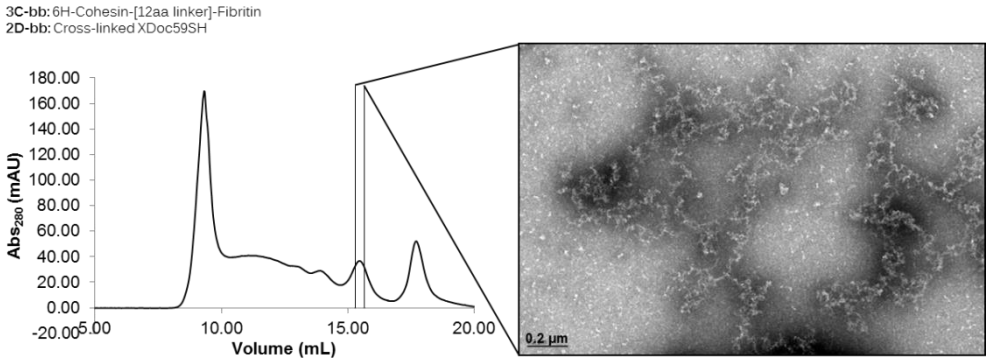


Figure 3 Heterodimer-driven assembly: size exclusion chromatography profile of the assembled sample (left) and negative staining (right)

The observation coming from the size-exclusion chromatography was confirmed by negative staining analysis on the most promising sample obtained through this approach: many heterogeneous, high molecular weight structures are observed (Figure 3).

Unexpectedly, the second approach (B) showed very different results (Figure 4). The formation of high molecular weight structures was essentially eliminated, even if the formation of a predominant species is not obtained neither in this case. Optimization of the crosslinking reaction is currently underway.
It is interesting to notice that in this case the assembly is conducted in a intramolecular-like way: after the first two units of dockerin react, they bring close together other four unreacted units, which are more likely to react with each other, rather than with other "free" dockerins in solution. This might explain why the formation of higher molecular weight species is suppressed with this assembly approach.

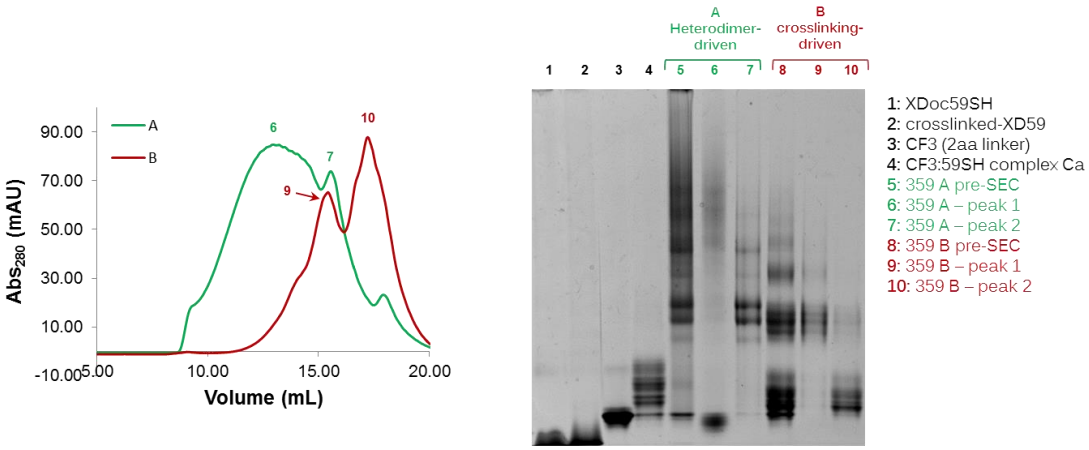


Figure 4 Comparison between assembly methods A (Heterodimer-driven) and B (Crosslinking-driven). Size-exclusion chromatography profile (left) and native-PAGE (right).

Since neither of the crosslinking approaches is giving exclusively the desired product, we decided to structurally investigate the single subunits in order to better understand their orientation and, therefore, to adjust the design for a better efficiency of assembly.
We focused on the crystallization of the intermediate assembled-structure between 3C-bb and three (monomeric) dockerin units (in the presence of Calcium ions). This is the substrate for the crosslinking reaction in the strategy of assembly B. The elucidation of the structure could help us to finely understand the orientation of the single subunits in the assembled-structure, and therefore, to optimize the overall assembly process.
Thus far, we have screened more than 200 different matrix conditions and obtained successfully protein crystals in three different crystallization conditions for forthcoming structural investigation by X-rays diffraction.

2.2 Design and Screening of homodimeric-dockerins for "biomimetic" assembly approach

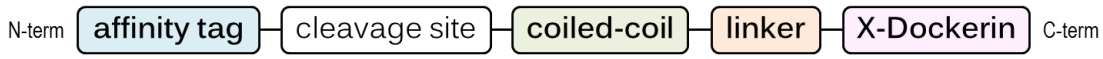
Since the previous approaches were not fully successful, we attempted in parallel to generate the nano-structure by a 2D-bb re-design, to expand the number of assembly strategies. We replaced the chemical crosslinking approach with a "biomimetic" approach. The assembly of the 2D-bb subunit in this case is driven by the formation of a dimeric antiparallel helix along the straight edge of the tetrahedral pyramid. By genetically fusing a homo-dimeric domain to dockerin, the formation of a dimeric dockerin would occur by simple heterologous expression of this new chimeric protein.

Among all the possible dimeric domains, we focused the coiled-coils protein family. These domains are small sequences (usually 30-40 amino acids) which form amphipathic alpha-helices, thus driving the oligomerization of the peptide. Coiled-coiles are very well-known protein structural elements, and have been extensively studied and engineered.

In order to properly orient the two Dockerin units in the 2D-bb for the final assembly, the dimerization needs to takes place in an antiparallel fashion, therefore, we selected three families of coiled-coil, which have been previously used for higher-order assemblies:

1) Bcr: oligomerization domain of Bcr-Ab1 oncoprotein, 39 amino acids-long helix, crystal structure is known;[7]
2) MyoX: dimerization antiparallel domain of the MyosinX, composed of one main dimerization domain of 30 amino acids, with an additional smaller helix to stabilize the assembly, NMR structure is known;[8]
3) APHs: "synthetic" coiled-coil sequences, with no structural characterization (APH, APH2, APH3, APH4).[9,10]

We engineered several dimeric structures, appending the coiled-coil sequence (22-24 kDa) to the N-terminal of the Dockerin. Different constructs were designed varying the length of the linker separating the dimerization domain and the dockerin, and varying the affinity tag for the optimization of the purification process:



Construct Name	Affinity tag	Coiled-Coil	Linker
6H-BD	6xHis	Bcr	none (direct fusion)
MBP-BD	MBP	Bcr	none (direct fusion)
6H-BhD	6xHis	Bcr	[EAAAK] - **h**elical
MBP-BhD	MBP	Bcr	[EAAAK] - **h**elical
6H-BgsD	6xHis	Bcr	[GS] – Bam restriction site
6H-M10D	6xHis	MyoX	none (truncation at 910 position of MyoX sequence: direct fusion with XDoc)
MBP-M10D	MBP	MyoX	none (truncation at 910 position of MyoX sequence: direct fusion with XDoc)
6H-M30D	6xHis	MyoX	20 AA (truncation at 930 position of MyoX sequence)
MBP-M30D	MBP	MyoX	20 AA (truncation at 930 position of MyoX sequence)
6H-M20D	6xHis	MyoX	10 AA (truncation at 920 position of MyoX sequence)
6H-AD	6xHis	APH	none (direct fusion)
MBP-AD	MBP	APH	none (direct fusion)

Table 1 2D-bb constructs with coiled-coil domains

The expression of the first engineered construct, 6H-BD, proved it to be very insoluble. We thus focused on the generation of a set of 2D-bb constructs with a purification tag, which can increase the protein fusion solubility, the Maltose Binding Protein (MBP). With this strategy, the engineered construct carries, starting from the N-terminal in order: 1) the MBP, 2) a protease recognition site for the cleavage of the purification tag, 3) the coiled-coil sequence, 4) a short linker of variable length, 5) the Dockerin unit.
In all cases, the MBP-fused constructs were soluble and were produced in very high yields. However, when the MBP is removed by TEV digestion, all of the 2D-bb so far tested, tended to precipitate and only a few micrograms of the expression product remained soluble, so no assembly test with the 3C-bb was possible.

Therefore, I tried to produce simple constructs with only a 6-histidine chain (6xHis) as purification tag instead of MBP. Most of the expressed and purified proteins tested so far, showed good solubility and final purity. Although the expression yields are not very high (in the order of a few mg per litre of bacterial culture), sufficient sufficient quantities of pure protein for the following steps of characterization and structure assembly in the presence of the 3C-bbs module could be obtained.

From size exclusion chromatography (SEC), almost all the obtained 2D-bbs showed to form the expected dimeric structure in solution (Figure 5, left).
Despite the molecular weight of all the constructs are very similar (22-24 kDa – monomer), we can observe some anomalies:
1) 6H-M10D have a chromatographic profile similar to the monomeric dockerin reference (18 kDa), therefore it looks like a monomeric species in solution, but its native-PAGE profile looks very similar to the dimeric 6H-M30D;
2) 6H-BhD shows a chromatographic profile which reflects a structure with higher size than the others (confirmed by the native-PAGE);
3) from SEC the monomeric reference (X-Dockerin) have a very different profile from all proteins except 6H-M10D. Surprisingly, native-PAGE showed for this protein a profile that is very similar to that obtained for the MyoX coiled-coil;
4) the construct 6H-APH had very low solubility (see line 2 native PAGE – Figure 5).

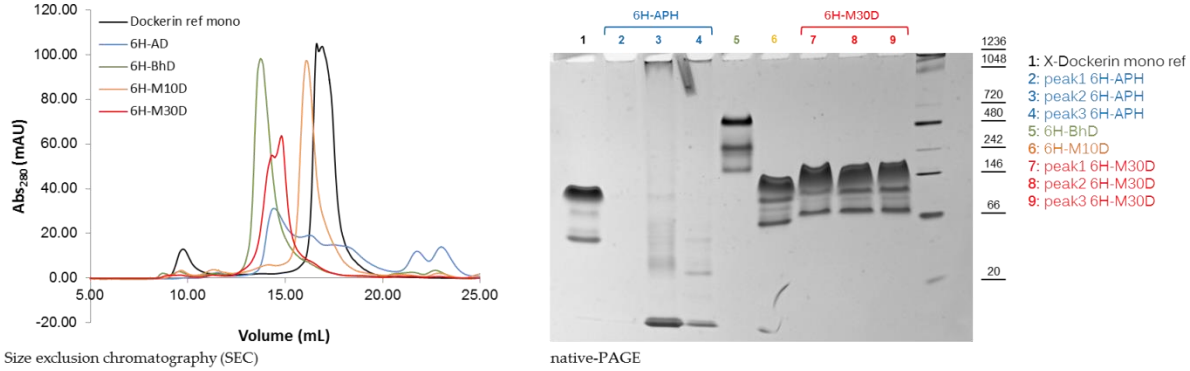


Figure 5 Size exclusion chromatography of 2D-bb with fused different coiled-coils (left) and native-PAGE of the corresponding fractions (right)

Finally, we also have performed preliminary assembly tests with the proteins 6H-M10D, 6H-M30D and 6H-BhD. From SEC (Figure 6), it is clear that the situation is very different from the one observed with the chemical crosslinking method. The "biomimetic assembly" approach resulted in very defined peaks of apparently homogeneous material. Also, the formation of undesired higher-molecular assemblies was suppressed. I reported only the 6H-M30D:CF3 profile for a clearer picture, but both M10D and BhD show similar profile.

However, it is difficult to unambiguously predict if the desired assembly is formed from the elution volume, since we are observing a multi-protein porous assembly. Therefore, we have started the production of novel samples for the characterization by negative staining, in order to gain a better understanding of the 3d shape and quality of our samples, before proceeding to a more accurate structural characterization (single particle cryo-EM and/or X-ray diffraction crystallography).

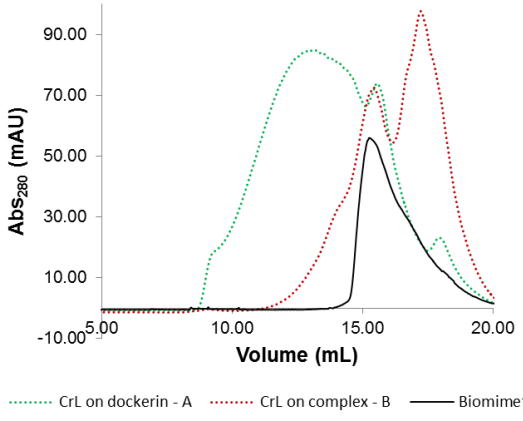


Figure 6 Confront between "crosslinking-based" and "biomimetic-besed" assembly methods

3. Conclusions and Outlook

In the past year, I investigated the optimization and development of our designed protein cage by different assembly strategies.

The first strategy, comprising a chemical crosslinking reaction as crucial part of the assembly, proved to be much more challenging than expected. Even if some improvement to this approach was obtained (development of strategy B), the final desired structure had not been isolated and identified yet.

Therefore, in parallel to the further optimization of the first method, we sought a different strategy for the assembly, through a more "biomimetic" approach. We developed many new 2D-bb, and we have started to test them for the final assembly process. We have planned, in the next months, to analyze these assemblies by negative staining and to select, among all the constructs produced, the most promising ones, before proceeding to a structural characterization.

For the next year, our intention is to try to conclude the optimization of the final assembled protein cage. We still have some strategies left to test, other than the ones that we are currently investigating:

- Among the 2D-bb developed for the biomimetic approach, some of them have the MBP tag to the N-terminal. We observed that these are not able to dimerize, probably due to steric hindrance given by the affinity tag. We could exploit this "hiding" of the dimerization domain to form first the complex with the 3C-b (in analogy with the crosslinking method B). Then, once formed the heterodimer between cohesin and dockerin, we could remove the MBP and the dimerization of the cohesin domain should be triggered, giving a "coiled-coils-driven" strategy of assembly.
- A further possibility could be to exploit the trimerization site as the final driving force of the assembly. Thus far, we have never considered to alter the 3C-bb construct. However, since it forms immediately very stable trimers, it might drive efficiently the final assembly step of the nano-biostructure. Analogously to the previous point, we could mask the trimerization domain with a hindered tag (such as MBP), form the heterodimer with the dockerin, and finally trigger the assembly of the cage by removal of the tag. In this way, we could have a "trimerization-driven" strategy of assembly.

Finally, as a complementary methodology to investigate the nanomaterial assembly process, we aim also to functionalize the 2D-bb and 3C-bb with fluorescent tags (fluorescent proteins or even small fluorophores to avoid altering too much the assembly), to gain a more immediate read-out of the complex formation. We could exploit FRET-based strategies to better appreciate differences among the assembly methods, since SEC and native-PAGE could not give us a fully understanding of the situation. This is of course a more "indirect" approach than any structural characterization, but it could give us immediate information, and therefore be used as a "screening" method to test the assemblies.

4. References

(1) Kobayashi, N.; Arai, R. Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. *Curr Opin Biotechnol* **2017**, *46*, 57-65.
(2) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **2016**, *353*, 389-394.
(3) Rudroff, F.; Mihovilovic, M. D.; Gröger, H.; Snajdrova, R.; Iding, H.; Bornscheuer, U. T. Opportunities and challenges for combining chemo- and biocatalysis. *Nature Catalysis* **2018**, *1*, 318–325.
(4) Tao, Y.; Strelkov, S. V.; Mesyanzhinov, V. V.; Rossmann, M. G. Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure* **1997**, *5*, 789–798.
(5) Adams, J. J.; Pal, G.; Jia, Z.; Smith; S. P. Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proc Natl Acad Sci* **2006**, *103* (2) , 305-310.

(6) Jindou, S.; Kajino, T.; Inagaki, M.; Karita, S.; Beguin, P.; Kimura, T.; Sakka, K.; Ohmiya, K. Interaction between a type-II dockerin domain and a type-II cohesin domain from Clostridium thermocellum cellulosome. *Biosci Biotechnol Biochem* **2004**, *68*(4), 924-6.
(7) Taylor, C.M.; Keating, A. E. Orientation and oligomerization specificity of the Bcr coiled-coil oligomerization domain. *Biochemistry* **2005**, *44*(49), 16246-16256.
(8) Lu, Q.; Ye, F.; Wei, Z.; Wen, Z.; Zhang, M. Antiparallel coiled-coil-mediated dimerization of myosin X. *Proc Natl Acad Sci* **2012**, *109*(43), 17388-9.
(9) Gurnon, D.G.; Whitaker, J.A.; Oakley, M.G. Design and characterization of a homodimeric antiparallel coiled coil. *J Am Chem Soc* **2003**, *125*(25), 7518-9.
(10) Negron, C.; Keating, A. E. A set of computationally designed orthogonal antiparallel homodimers that expands the synthetic coiled-coil toolkit. *J Am Chem Soc* **2014**, *136*(47), 16544–16556.