Designing Linkers

CuRe

iGEM

A comprehensive guide by the EPFL 2021 iGEM team

What you need to know

Rigid linkers
Flexible linkers
Semi-flexible linkers

EPFL

What you need to know

Polymerization of a protein or complex is often a key element in the success or improvement of a project. Indeed, linker peptides not only serve to connect the protein moieties, but also provide several other functions, such as maintaining cooperative inter-domain interactions, leading to the proper folding of the whole complex, preserving biological activity of original proteins, and sometimes adding new active functions.

However, creating a complex of multiple bio-macromolecules joined by a linker is anything but easy and requires a rational and articulate design, aiming to meet certain specifications and limitations sometimes given by the complex itself. Consequently, choosing a certain linker rather than another can not only lead to the proper functioning of the complex and the idea behind it but, if chosen correctly, allows to bring to the project several advantages and added values. Similarly, choosing an unsuitable linker may result in several undesirable outcomes, including mis-folding of the fusion proteins, low yield in protein production, or impaired bioactivity. Therefore, the choice of the linker can bring several advantages to the idea or design of the project, but it requires appropriate attention in the development and the decision more suitable to their own needs.

During the design of our project, we committed a significant part of our time to the choice and selection of linkers that characterized our dimers. This work is the result of several brainstorming sessions, consultations with experts and bibliographic research.
As a result, we are pleased to report and share our ideas and results here, hoping they will be a prosperous resource and inspiration for future iGEM teams.

All of our linkers worked, were compatible with cloning, and resulted in the expression of the interested complex. Our final choice was seven linkers. Four of these are variants of linkers that are pretty common and well known in the field. However, the other three represent our real breakthrough and innovation.

To express different features in all our dimers, we developed a precise design for each linker. To do so, we mainly played with four factors: **length, hydrophobicity, amino acid residues, and secondary structure**. The last three could also be categorized under the macro concept of flexibility. Therefore, length and flexibility were the main actors of our design.

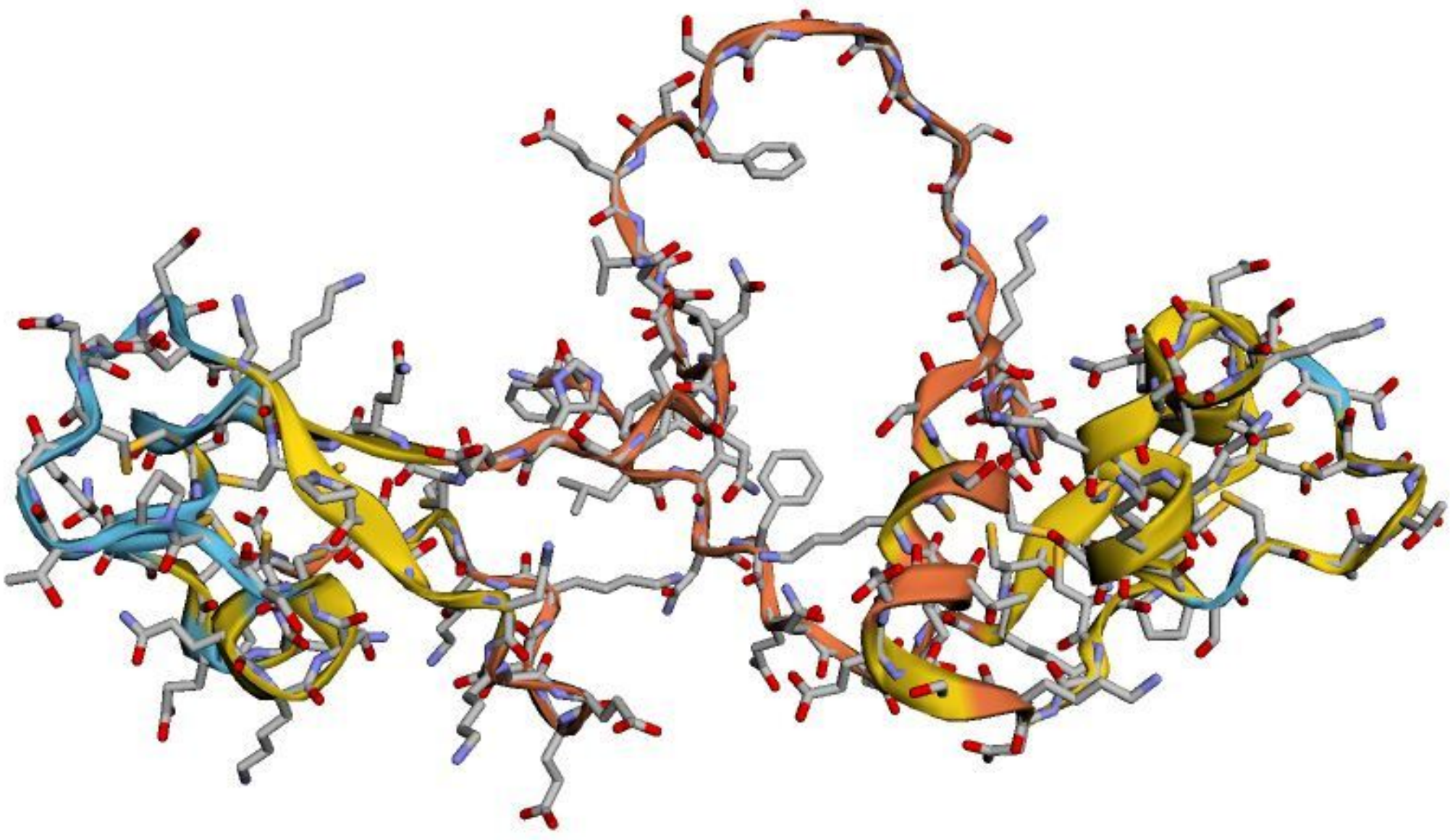
Regarding the length, it determines the distance between the two proteins, thus allowing greater or lesser interaction between the two. In order to obtain a tradeoff between the lack of interaction and the size of the complex, we chose sequences that varied between 14 and 20 amino acids, length at the threshold between an intermediate and a long sequence.

Flexibility (or rigidity) is instead a parameter used to empirically describe the ability of a protein sequence to maintain a specific conformation. This parameter, generally, is directly related to the amino acids nature, hence to the structure of the linker, which would also affect the orientations of the fused proteins.
Moreover, previous studies have shown that the flexibility of a linker plays a crucial role in the function of fusion proteins, indicating the importance of linker flexibility in the construction of the complex.
In order to observe the behavior of our complex in different conformations and observe the possible added value brought by each of them, we decided to experimentally test **flexible, semi-rigid and rigid** structures.

Flexible linkers

The most commonly used flexible linkers have sequences consisting primarily of stretches of Glycine and Serine residues (“GS” linker). An example of the most widely used flexible linker has the sequence of (Gly-Gly-Gly-Gly-Ser)n. By adjusting the copy number “n”, the length of this GS linker can be optimized to achieve appropriate separation of the functional domains, maintain necessary inter-domain interactions, and allow for proper folding of the fusion proteins.
In addition, although flexible linkers do not have rigid structures, they can serve as passive linkers to keep a general distance between functional domains.
The incorporation of Serine or Threonine can maintain the stability of the linker in aqueous solutions by forming hydrogen bonds with the water molecules, and therefore reduces the unfavourable interaction between the linker and the protein moieties.

Therefore, regarding the flexible linker, the final choice has been oriented towards a combination of Glycine and Serine, in two different combinations of GGGGS, (GGGGS)n, with n=3,4.



Protein structure prediction of a dimer with a flexible linker made with the AlphaFold prediction algorithm [8]

Rigid linkers

While flexible linkers have the advantage of connecting the functional domains passively and permitting a certain degree of movement, the lack of rigidity of these linkers can be limiting. There are several examples in the literature where flexible linkers resulted in poor expression yields or loss of biological activity. (rx) The ineffectiveness of flexible linkers in these instances was attributed to the poor separation of the protein domains or insufficient reduction of their interference with each other. In these situations, rigid linkers have been successfully applied to keep a fixed distance between the domains and maintain their independent functions, separating the functional domains more efficiently than the flexible linkers.

Rigid linkers exhibit relatively stiff structures by adopting α±-helical structures or by containing multiple Pro residues. Indeed, Pro is a unique amino acid with a cyclic side chain that causes a very restricted conformation. Furthermore, the lack of amide hydrogen on Pro typically prevents the formation of hydrogen bonds with other amino acids, and therefore reduces the interaction between the linkers and the protein domains. As a result, the inclusion of Pro residues might increase the stiffness and structural independence of the linkers.

Furthermore, in this case, the length of the linkers can be easily adjusted by changing the copy number to achieve an optimal distance between domains. As a result, rigid linkers are chosen when the spatial separation of the domains is critical to preserve the stability or bioactivity of the fusion proteins.

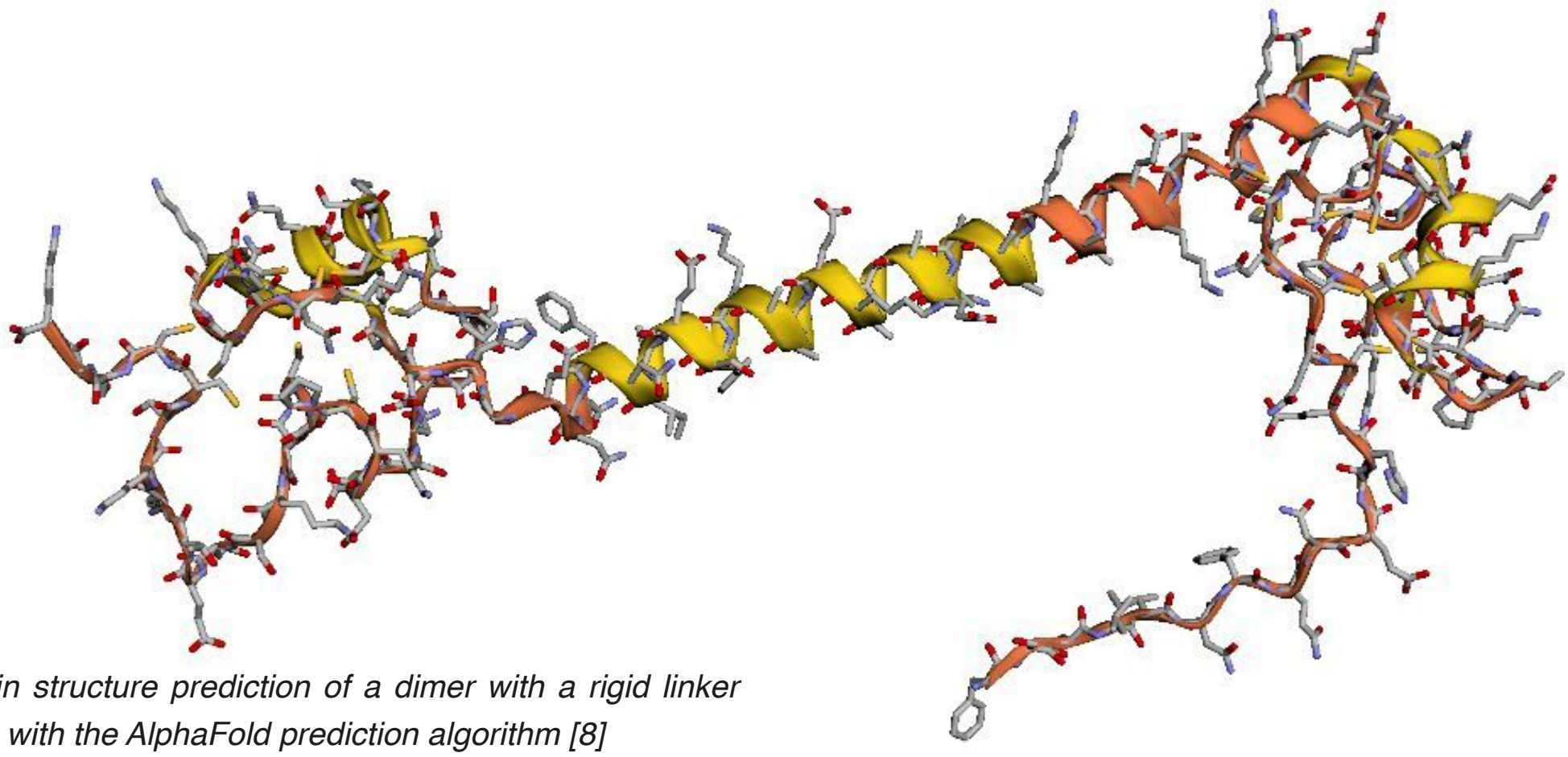
In this case, the most common rigid linker is the alpha helix-forming linkers with the sequence of (EAAAK)n, which shows an approximately 80 % helicity with n = 3 . As suggested by George and Heringa, many natural linkers exhibited α±-helical structures. Indeed, this latter is characterized by high rigidity and stability, with intra-segment hydrogen bonds and a closely packed backbone.

The α±-helix is a rigid and stable structure, with intra-segment hydrogen bonds and a closely packed backbone. Furthermore, some α±-helical conformations form rapidly during folding, which prevents non-native interactions of the domains with the linker and maintains the proper distance between the two moieties. Therefore, linkers in an α±-helix structure might also serve as rigid spacers to effectively separate protein domains and reduce their unfavourable interactions.

Therefore, the stiff α±-helical linkers may act as simple rigid spacers between protein domains.

Another type of rigid linkers, as explained above, has a Pro-rich sequence, (XP)n, with X designating any amino acid, preferably Ala, Lys, or Glu. Therefore, the presence of Pro in non-helical linkers can increase the stiffness and allows for effective separation of the protein domains as well.

Therefore, regarding the rigid linkers, we decided to use either a combination of EAAAK, interested in the resulting α±-helix structure, or a combination of AP interested in the properties of the poly-proline sequence.



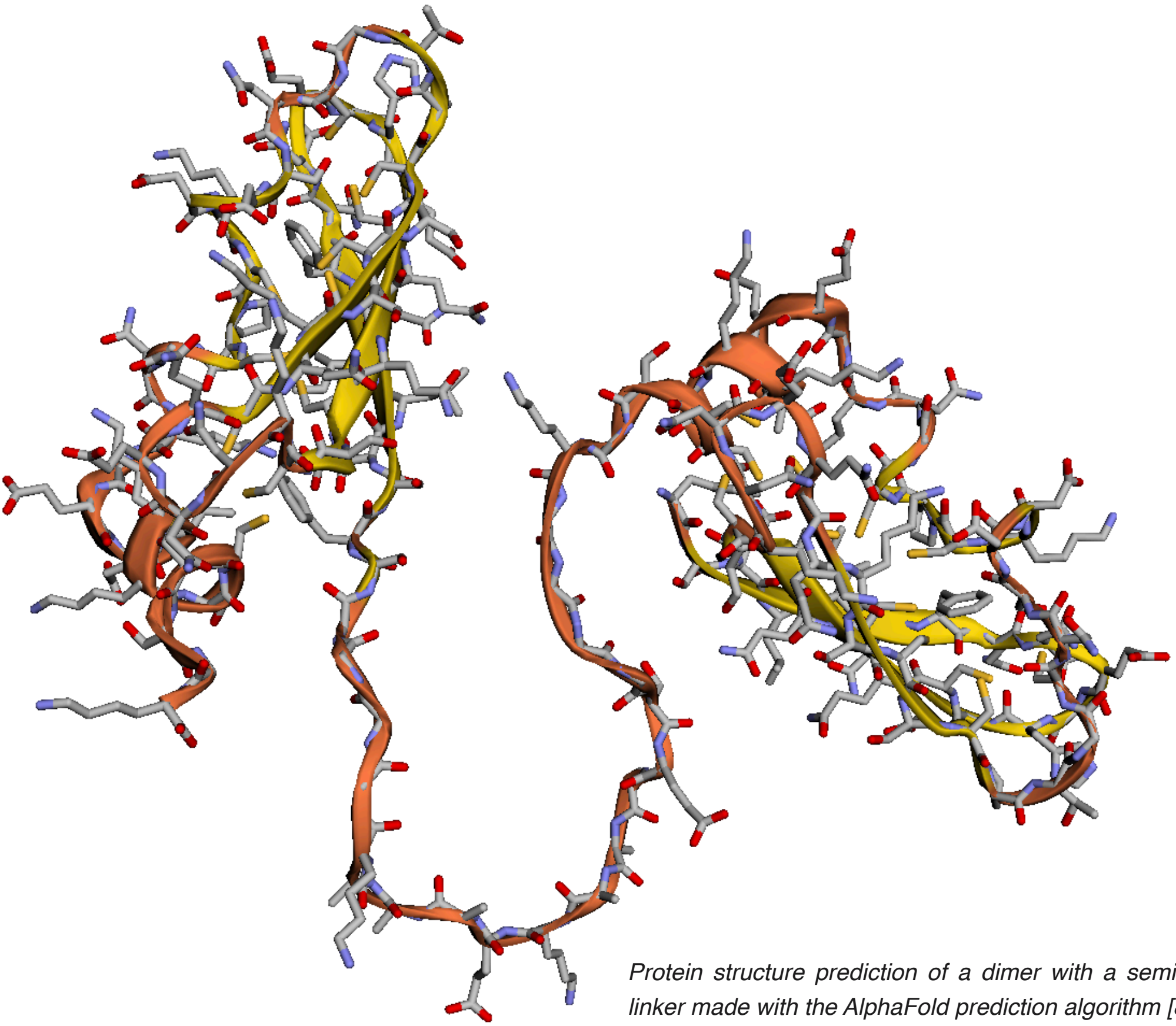
Protein structure prediction of a dimer with a rigid linker made with the AlphaFold prediction algorithm [8]

Semi-rigid linkers

Finally, recombining the entire building units with these significantly different flexibilities can potentially extend the linker mobility range for fusion protein design. Such an approach to generating linkers would benefit from using stable conformational sequences as the entire building units, thereby eliminating unwanted complexity in predicting conformations for various sequences and perturbations between the linker region and connected domains. In addition, as linker flexibility is dependent mainly on its conformation, which is in turn determined by the primary sequence, the linker library by recombining the rigid and flexible building units with a determined total length, can be expected to have widely controllable flexibility, capable of changing from rigidity to flexibility.

Therefore for the semi-flexible (or semi-rigid) conformation, we decided to combine the design choices used for flexible and rigid linkers. In particular, the main idea is to separate the two functional domains at a sufficient distance to prevent any interaction between the two and maintain their independent functions while allowing extreme flexibility in the area close to the C and N terminus of the fusion proteins. In this way, the complex will be highly rigid in the centre and particularly flexible in the vicinity of the two proteins allowing further degrees of freedom, proper folding, thus further facilitating the process of copper uptake. Accordingly, the structures used for the semi-rigid linker are:

GGGGS(EAAAK)nGGGS, with n=1,2, alpha helicity of around 15-20% and 25-30% respectively,



Protein structure prediction of a dimer with a semi-rigid linker made with the AlphaFold prediction algorithm [8]

Conclusion

In summary, linkers can adopt various structures and exert diverse functions to fulfil the application of fusion proteins and complexes, mainly based on their flexibility and length. The flexible linkers are often rich in small or hydrophilic amino acids such as Gly or Ser to provide structural flexibility and have been applied to connect functional domains that favour interdomain interactions or movements. In cases where sufficient separation of protein domains is required, rigid linkers may be preferable. By adopting α±-helical structures or incorporating Pro, the rigid linkers can efficiently keep protein moieties at a distance. Both flexible and rigid linkers are stable *in vivo* and do not allow joined proteins to separate. Finally, to obtain a tradeoff between the property of the flexible linkers and the one of the rigid linkers, the use of a semirigid linker, like the one here proposed, designed and implemented, could be the most suitable choice for most of the applications. In particular, being able to maintain a net distance between the two proteins, to avoid their interaction and incorrect folding, and at the same time allowing several degrees of freedom and preserving the correct biological activity is the ideal goal in the design of fused proteins. This is precisely the application our semi-rigid linkers have been designed for.

References

1. Klein JS, Jiang S, Galimidi RP, Keeffe JR, Bjorkman PJ. Design and characterization of structured protein linkers with differing flexibilities. *Protein Eng Des Sel*. 2014;27(10):325-330. doi:10.1093/protein/gzu043

2. Ryoichi A, Hiroshi U, Atsushi K, Noriho K, Teruyuki N, Design of the linkers which effectively separate domains of a bifunctional fusion protein, *Protein Engineering, Design and Selection*, Volume 14, Issue 8, August 2001, Pages 529–532, https://doi.org/10.1093/protein/14.8.529

3. George RA, Heringa J. An analysis of protein domain linkers: their classification and role in protein folding. Protein Eng. 2002 Nov;15(11):871-9. doi: 10.1093/protein/15.11.871. PMID: 12538906.

4. Argos P. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. J Mol Biol. 1990

5. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev*. 2013;65(10):1357-1369. doi:10.1016/j.addr.2012.09.039

6. Williamson MP. "The structure and function of proline-rich regions in proteins." *Biochemical journal* 297.Pt 2 (1994): 249. doi: 10.1042/bj2970249

7. Li, G., Huang, Z., Zhang, C. *et al.* Construction of a linker library with widely controllable flexibility for fusion protein design. *Appl Microbiol Biotechnol* 100, 215–225 (2016). https://doi.org/10.1007/s00253-015-6985-3

8. Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2