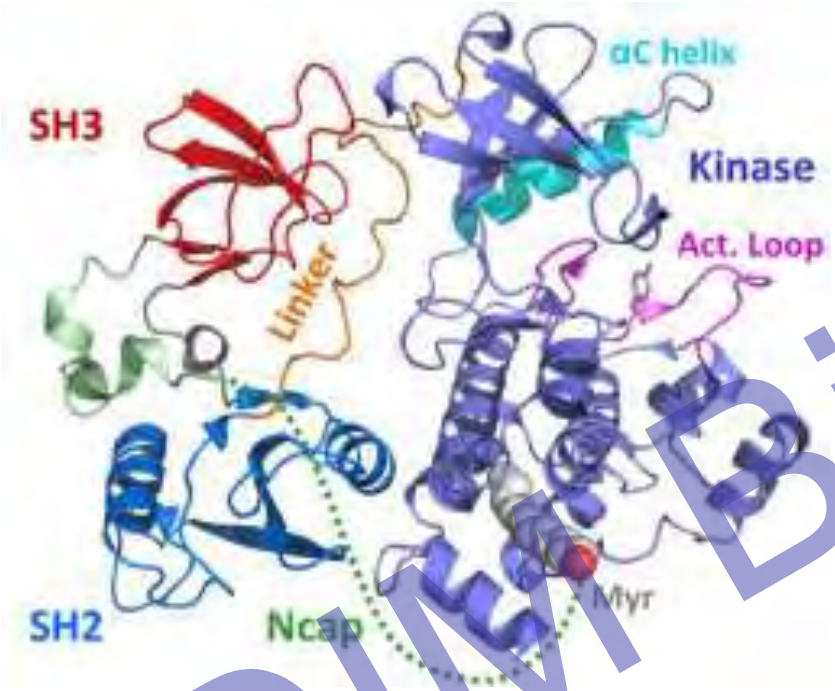
Inteins as molecular tools for synthetic biology

Barbara Di Ventura
Molecular and Cellular Engineering Group
Signaling Research Centers BIOSS and CIBSS
Institute of Biology II
University of Freiburg

Proteins are fascinating molecules

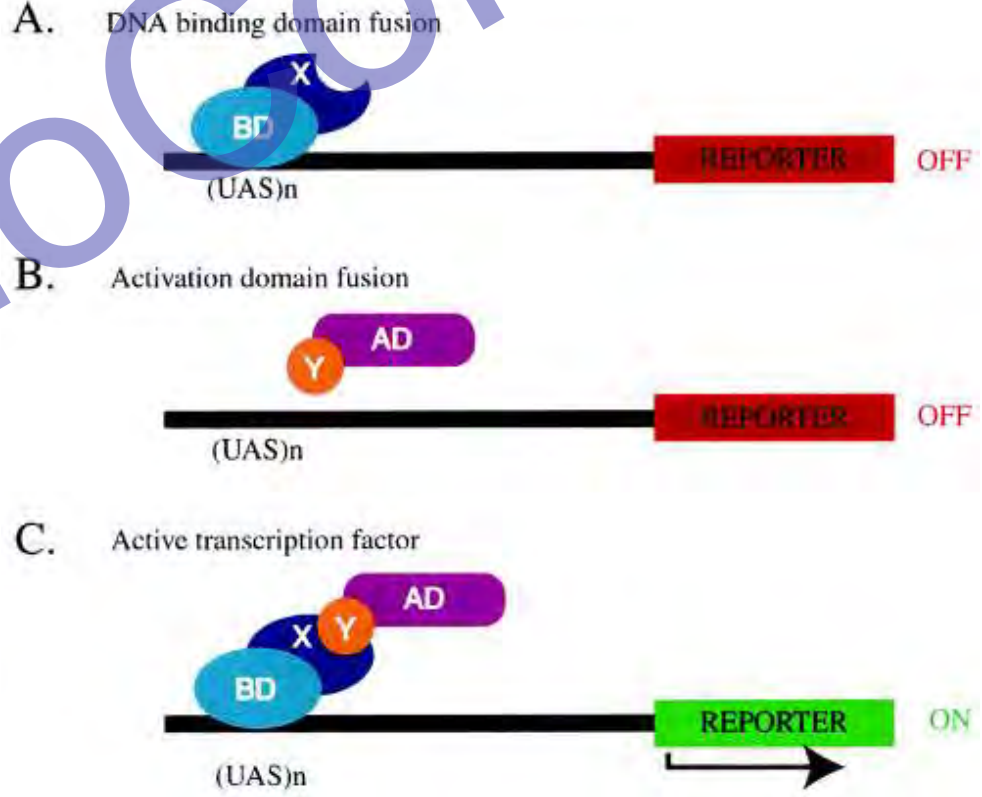
c-Abl-1b protein kinase



Panjarian *et al.*, JBC, 2013

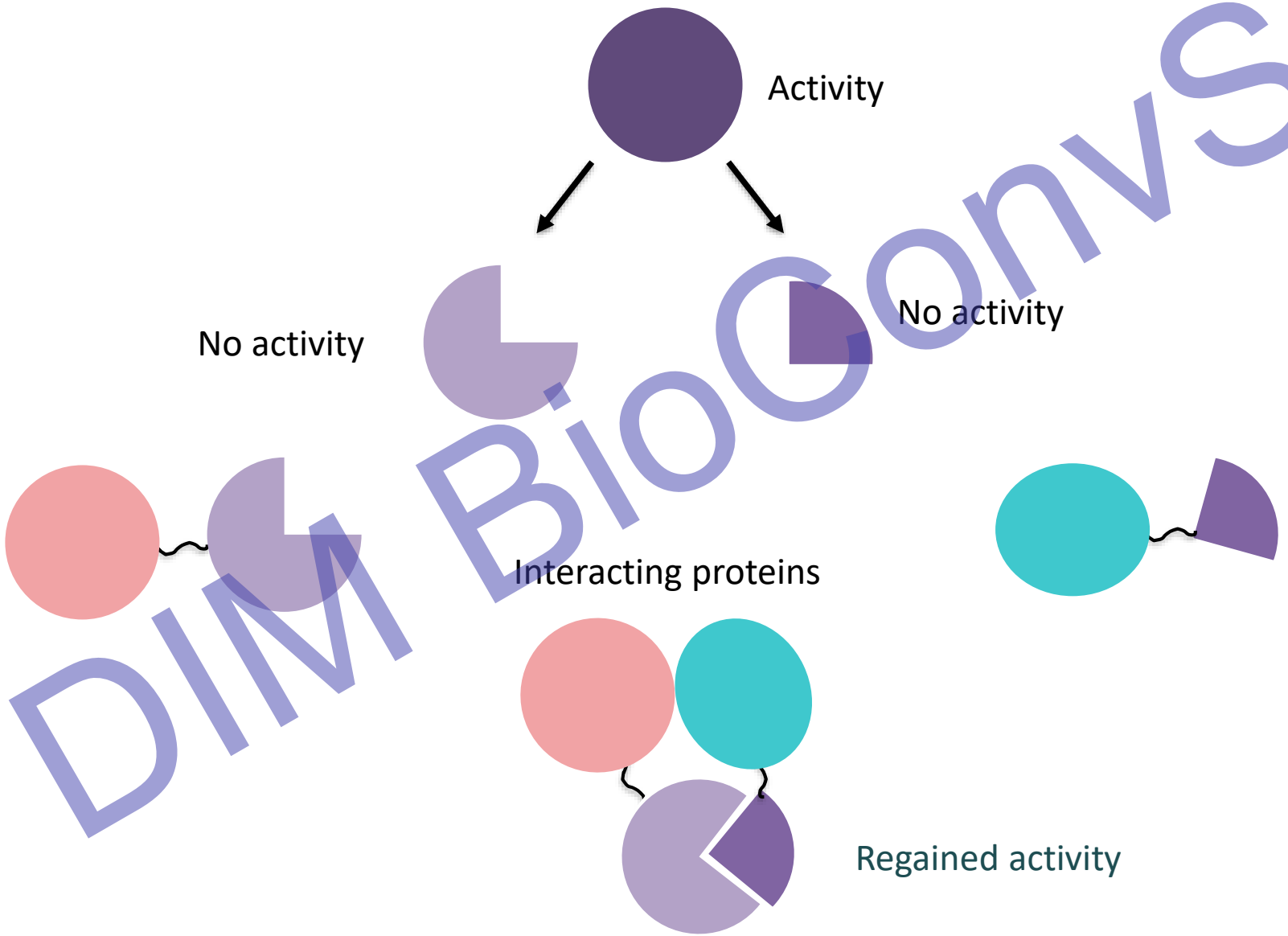
Protein domains can be *extracted* and brought back in close physical proximity using interacting protein partners

The yeast two-hybrid technology

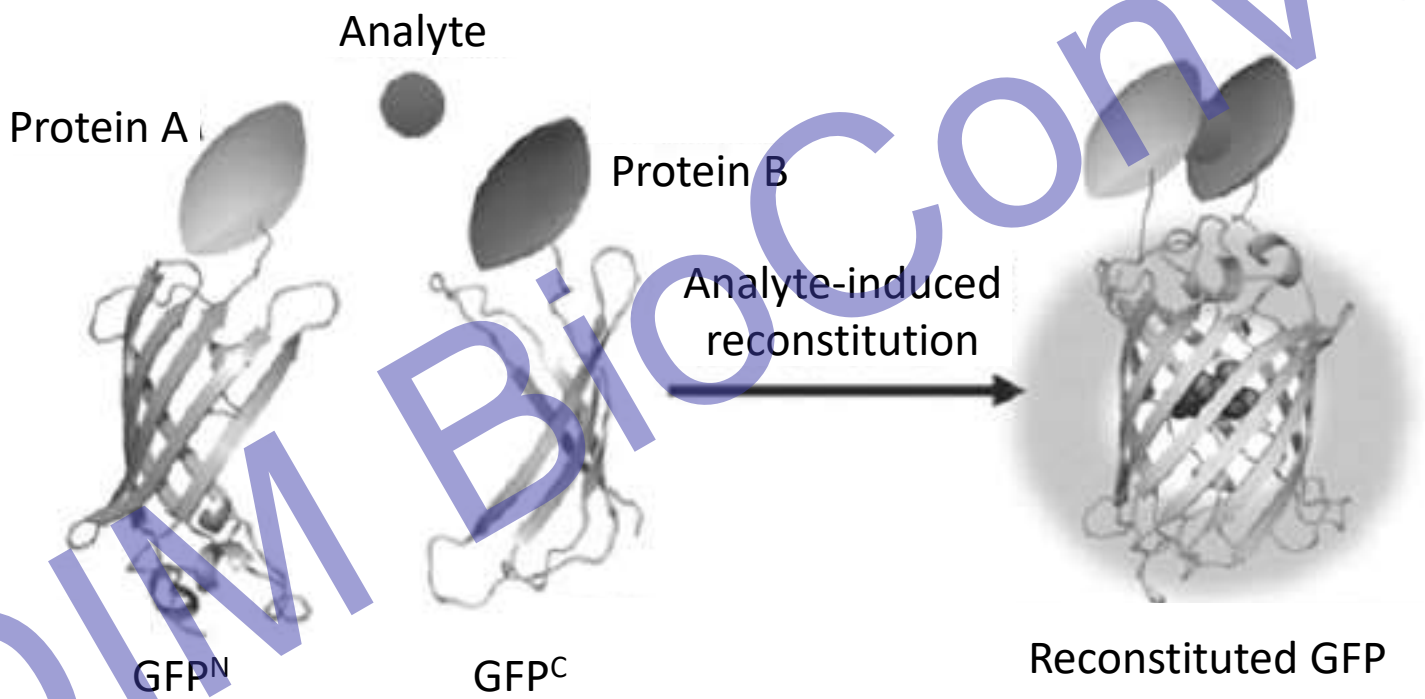


Stephens and Banting,Traffic, 2002

Proteins can be split into two or more parts

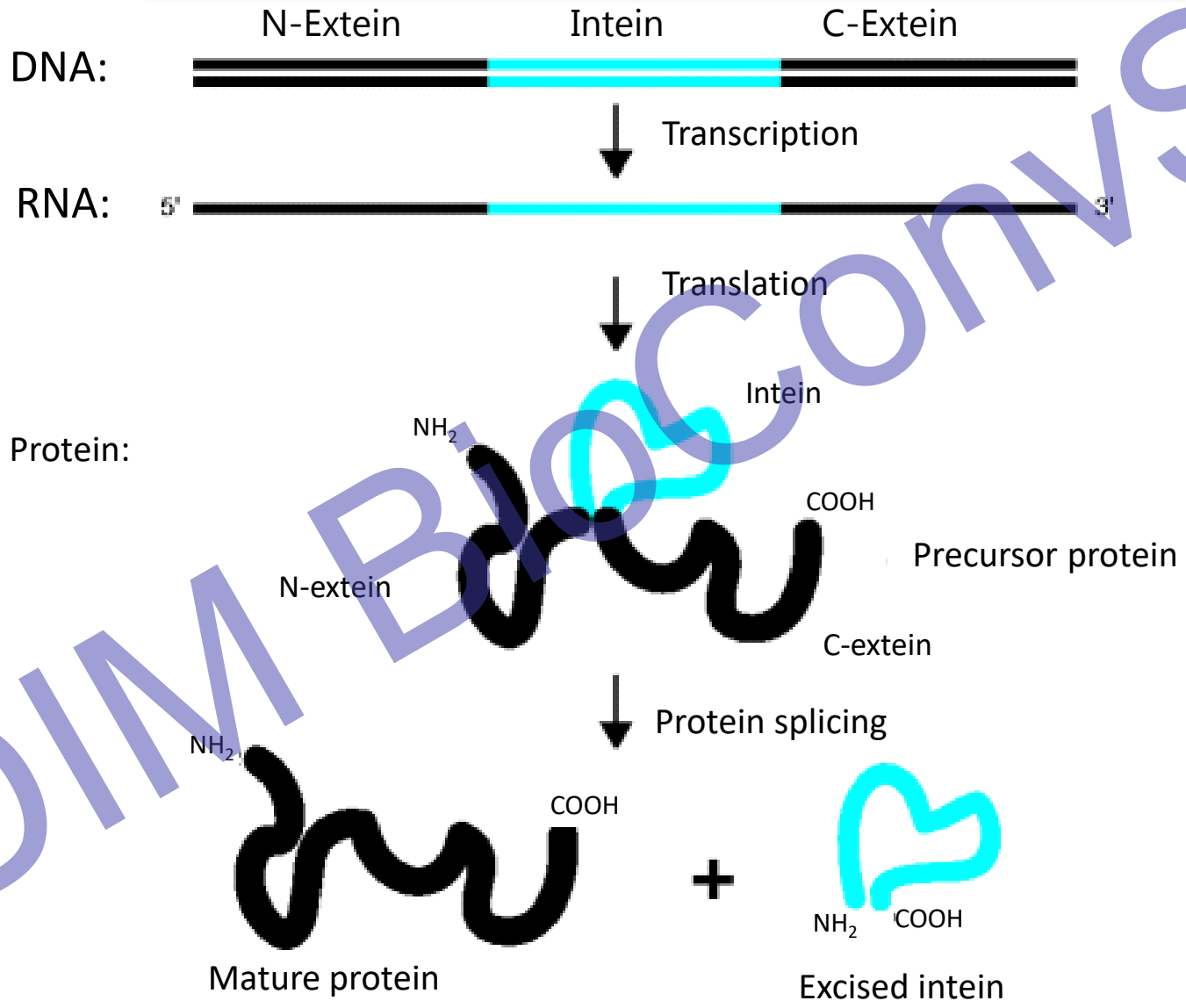


Reconstitution of split GFP is a widely used method to study protein-protein interactions



Wang *et al.*, ChemBioChem, 2009

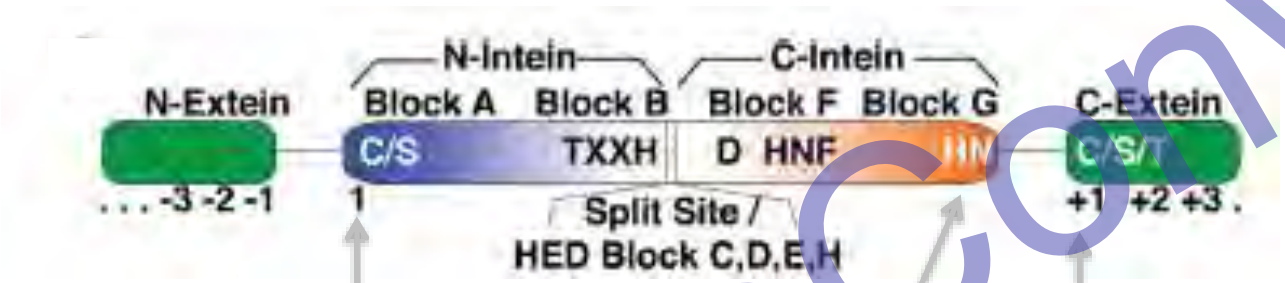
Inteins are special proteins



Elleuche and Pöggeler, Appl. Microbiol. Biotechnol., 2010

Protein splicing is a multistep process

Highly conserved motifs in the active site are needed for the splicing reaction



Erylmatz *et al.*, J Biol Chem (2014), 289: 14506-14511

Cys or Ser at the intein N-terminus

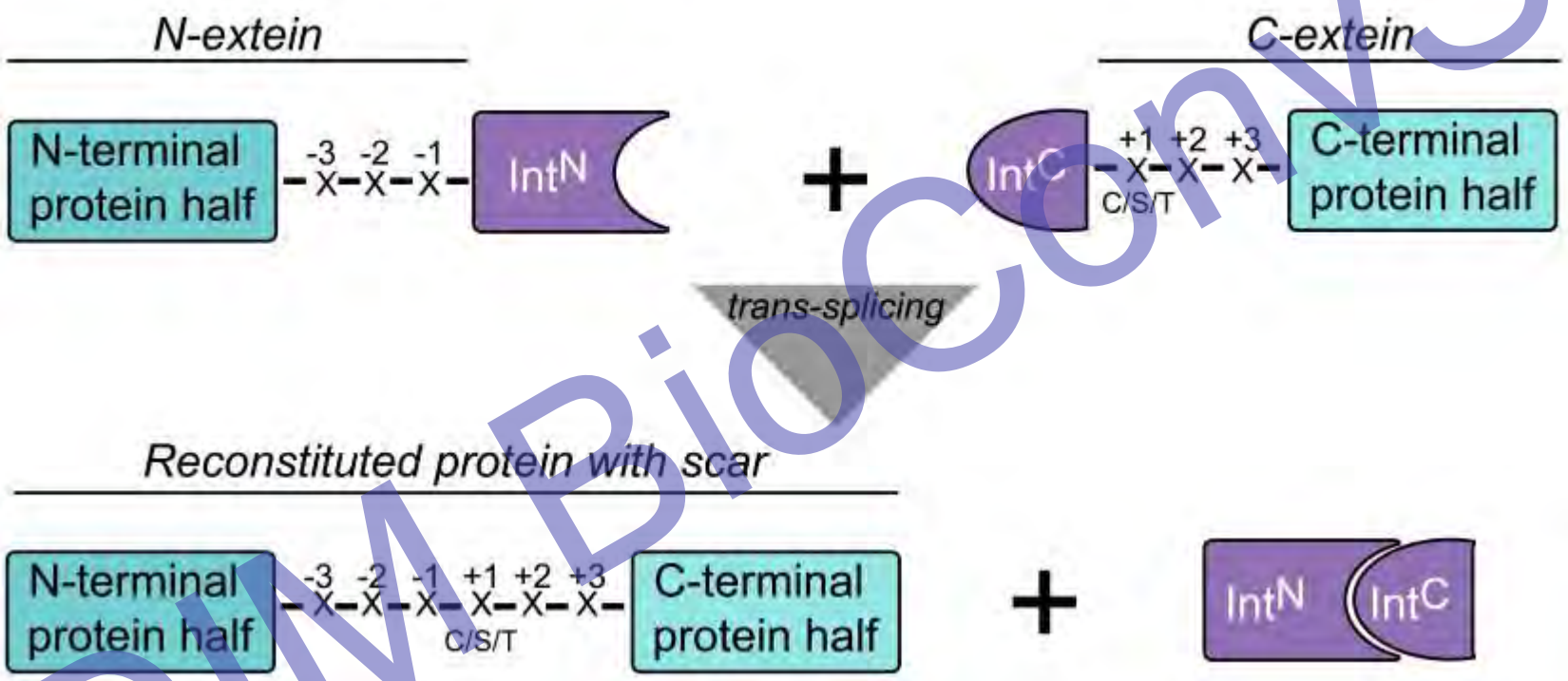
In vitro protein splicing showed it is a SELF-CATALIZED reaction

Invariant Asn at the intein C-terminus

Hydroxyl or thiol containing residues (Cys, Ser or Thr) at position +1

Split inteins can reconstitute proteins

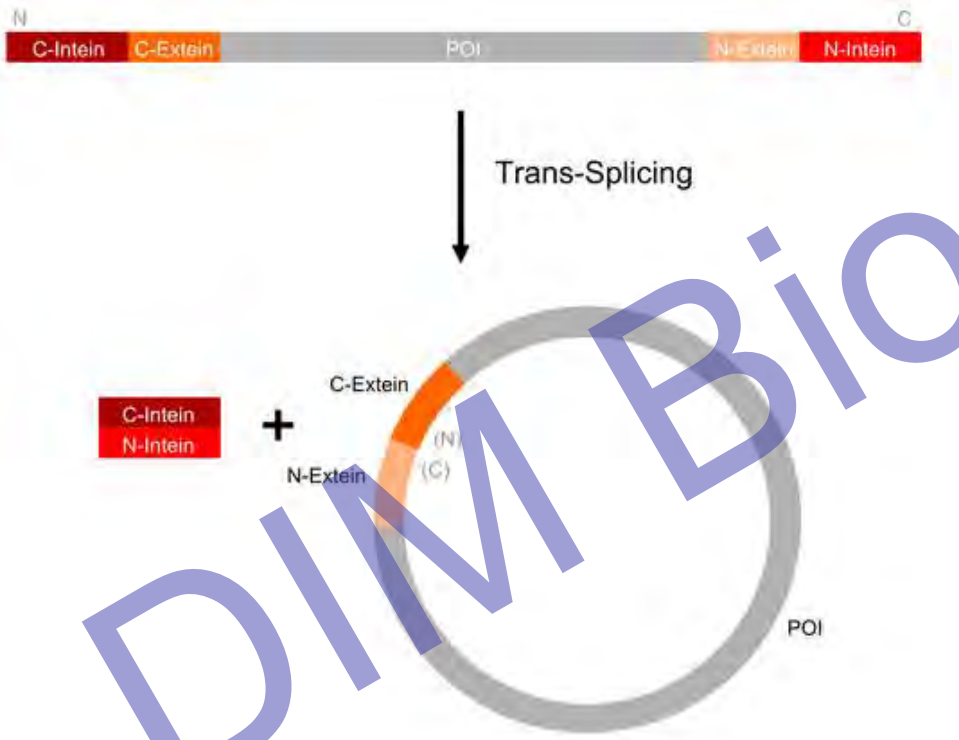
Inteins naturally exist as ***contiguous*** or ***split*** inteins



Split inteins can be used to circularize proteins

Circular proteins do not exist in nature

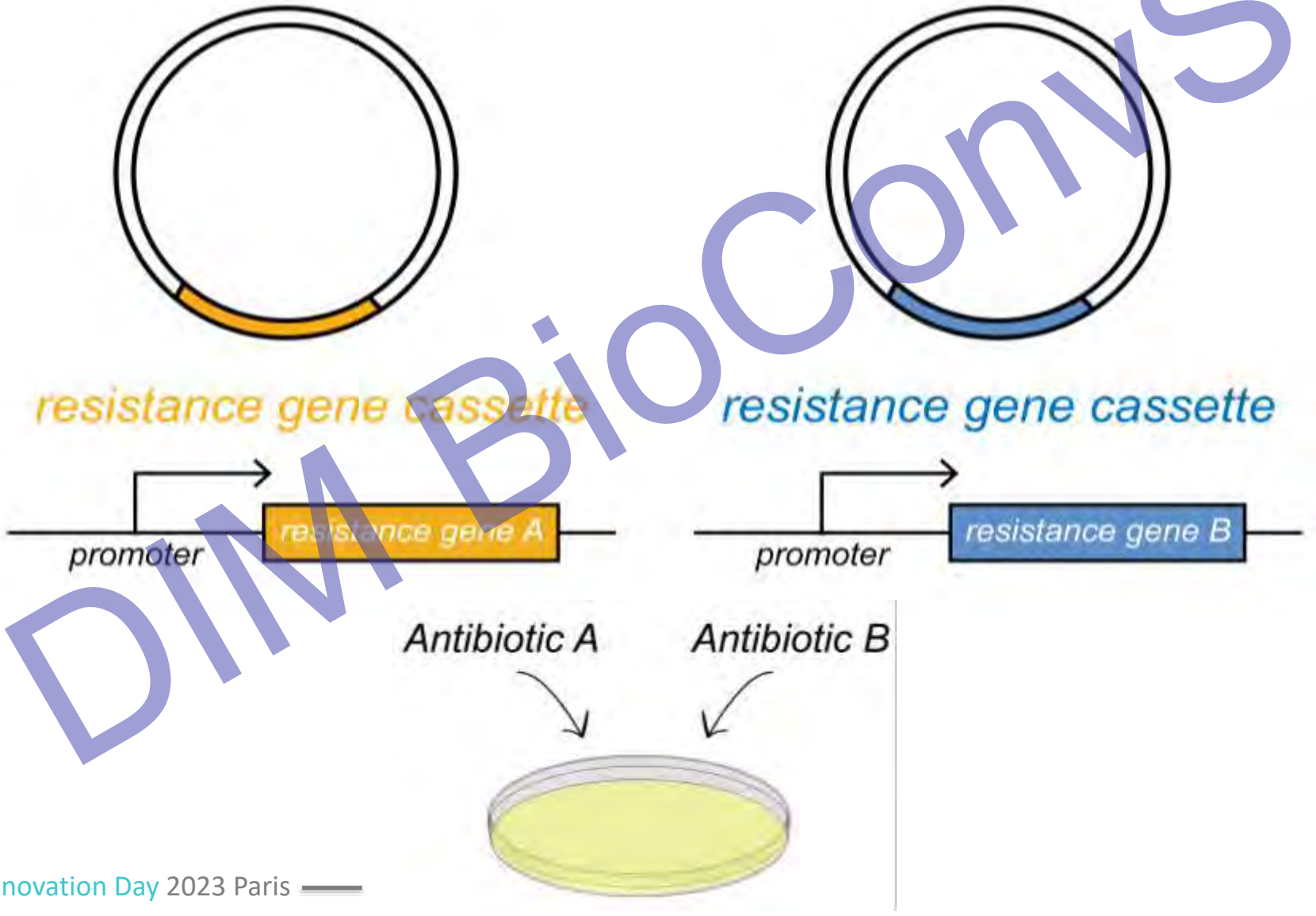
But we can make them using split inteins



Circular proteins are more resistant towards aggregation, proteolytic cleavage and have higher thermostability

Waldhauer et al., Mol BioSystems, 2015

In synthetic biology and microbiology two plasmids are often used

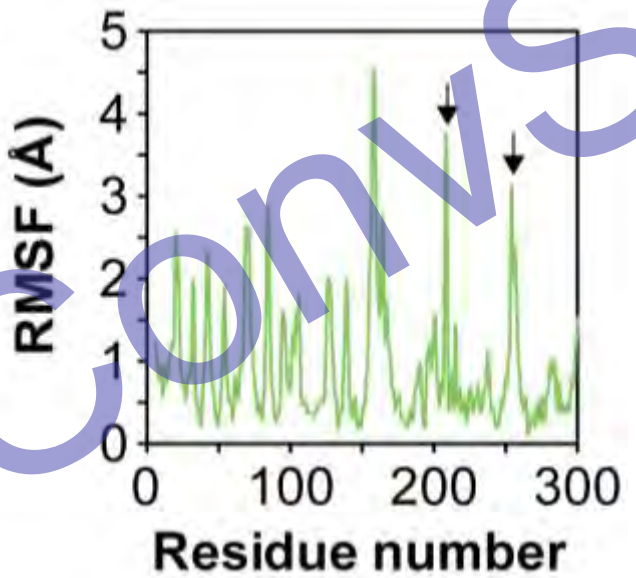


Using one antibiotic could be advantageous

Navaneethan Palanisamy

We used a rational approach to split enzymes conferring resistance towards antibiotics

Splitting in flexible regions is advisable

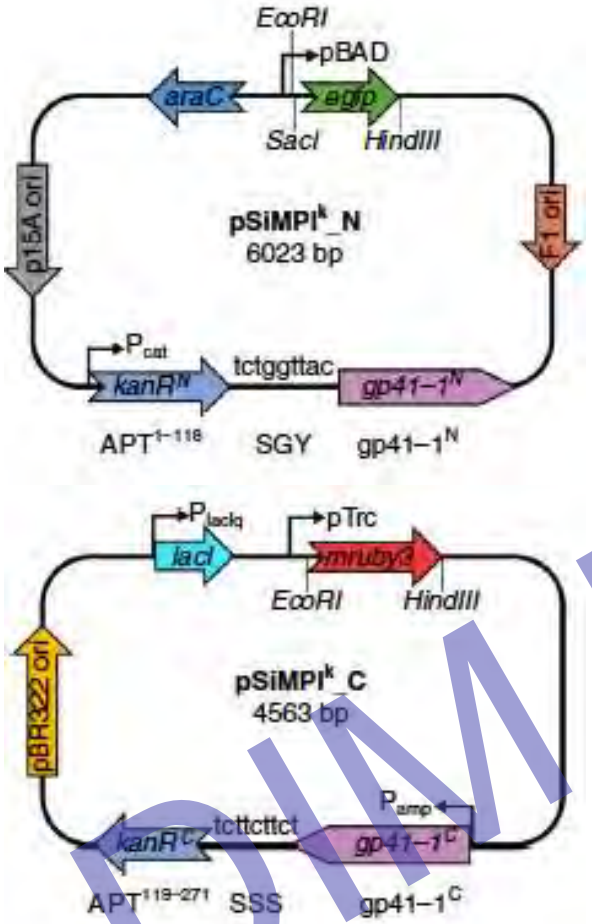


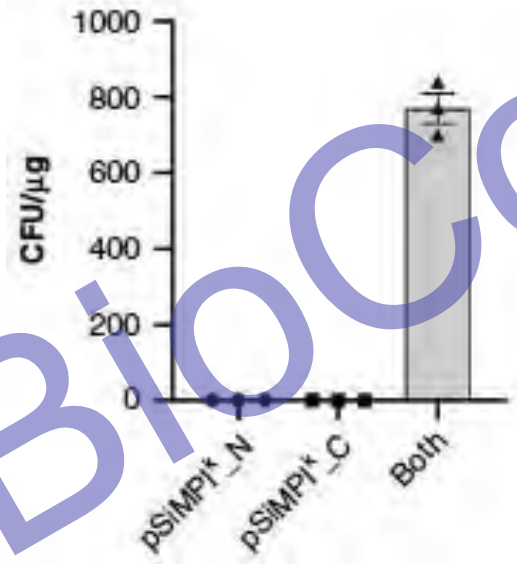
We used the webserver CABS-flex 2.0 to calculate the root mean squared fluctuation of Ca atoms

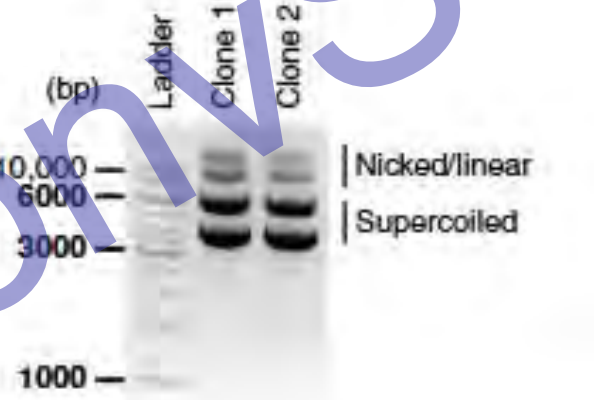
Additional criteria for the splice site

- Surface-exposed
- Not located in/near active site
- Located in a region of low conservation

We started with kanamycin resistance







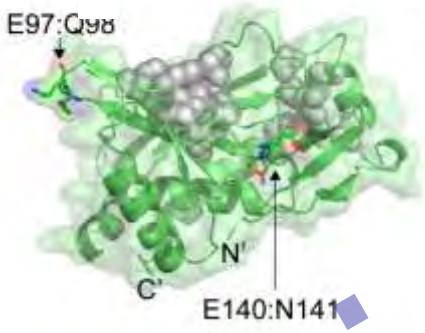
Palanisamy *et al.*, Nature Commun., 2019

SiMPl

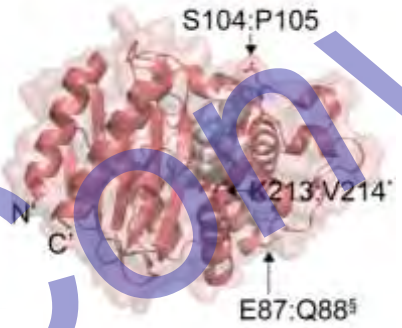
Split intein-mediated selection of cells containing two plasmids

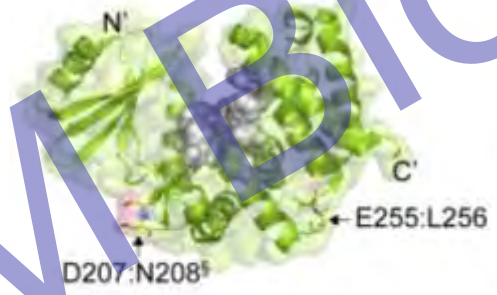
We created the SiMPl toolbox

Chloramphenicol acetyltransferase



Beta-lactamase





Hygromycin B phosphotransferase

Palanisamy *et al.*, Nature Commun., 2019

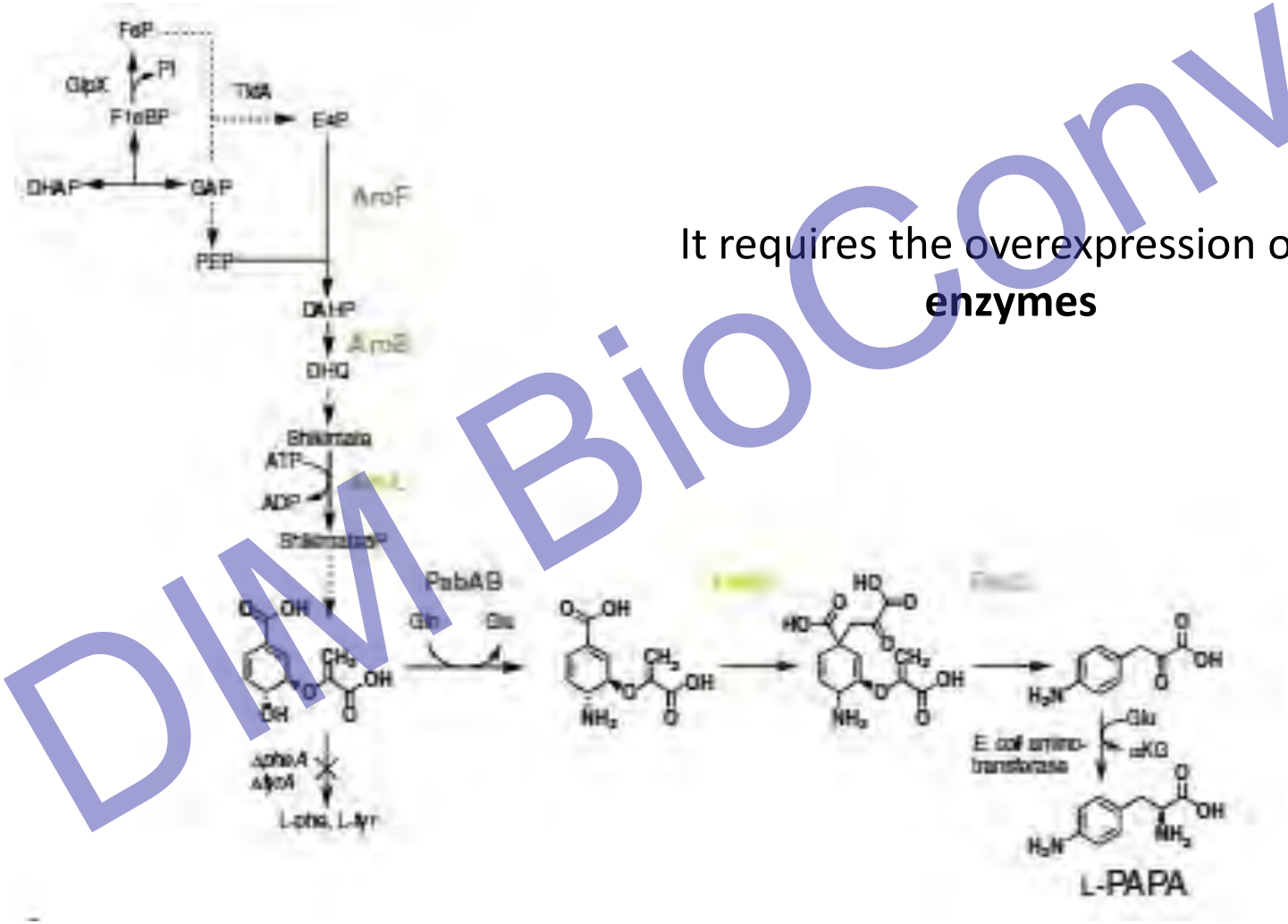
Puromycin acetyltransferase

Palanisamy *et al.*, ACS Omega, 2020

Spectinomycin/streptomycin

The use of SiMPl helps improve yield of valuable compounds

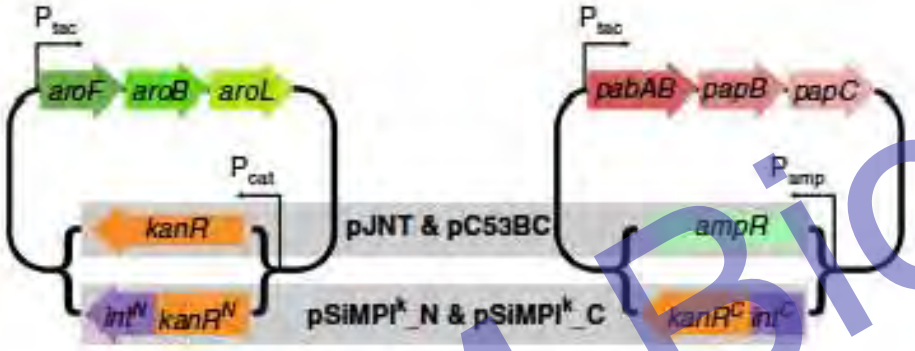
L-PAPA production from *E. coli*



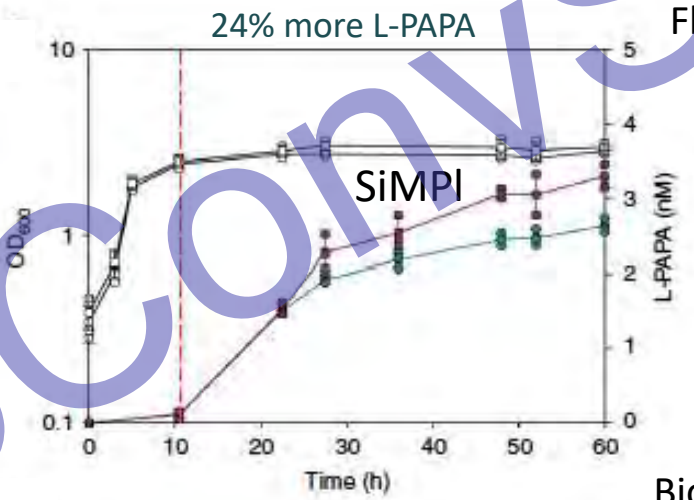
It requires the overexpression of **6 enzymes**

The use of SiMPl helps improve yield of valuable compounds

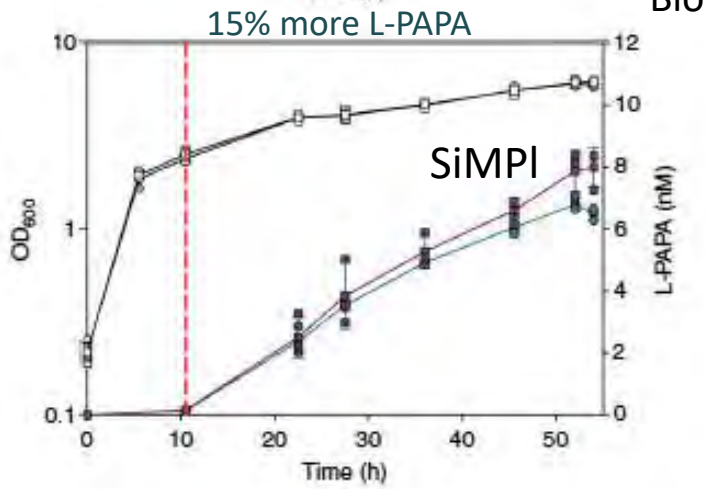
We constructed SiMPl versions of the plasmids to compare them with the original ones



Flasks

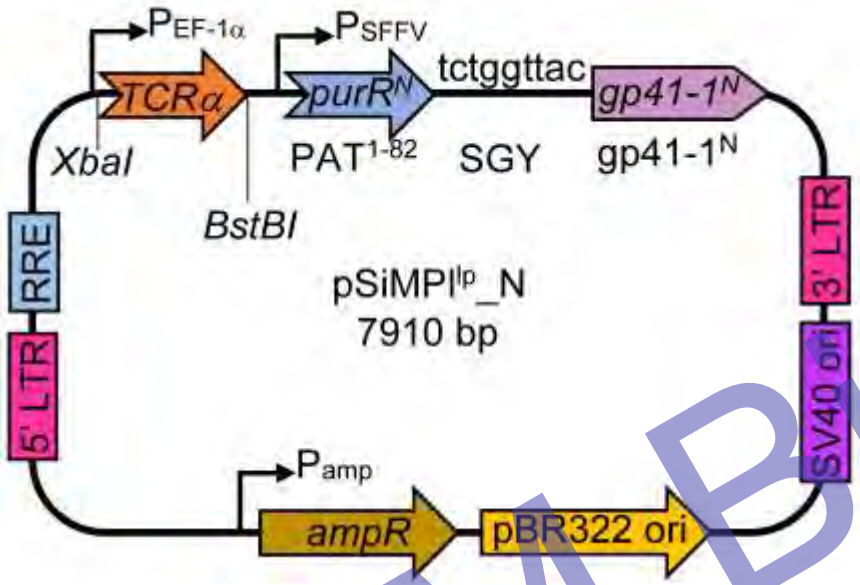


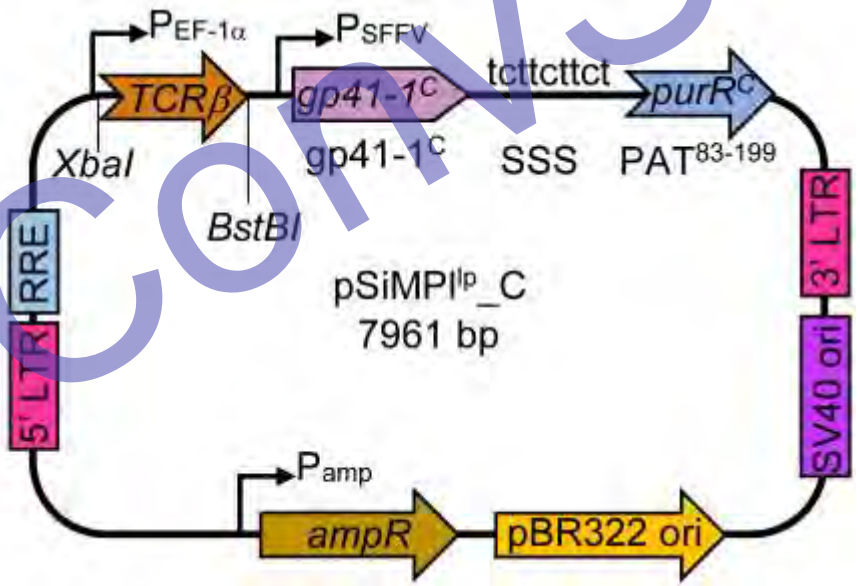
Bioreactor

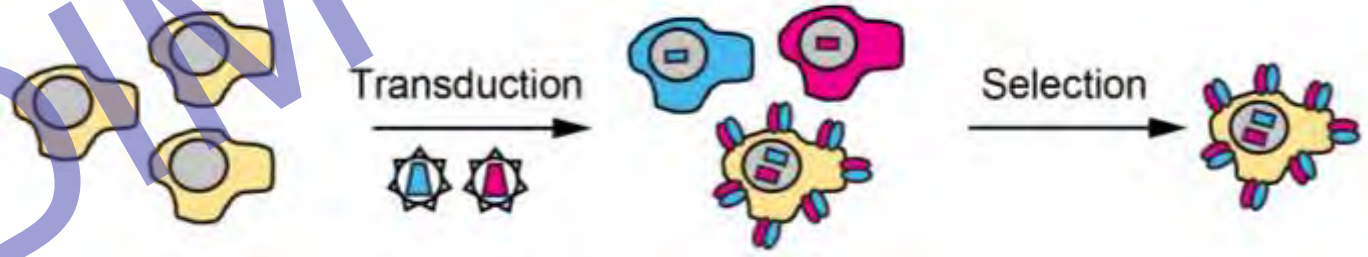


We constructed SiMPl lentiviral vectors

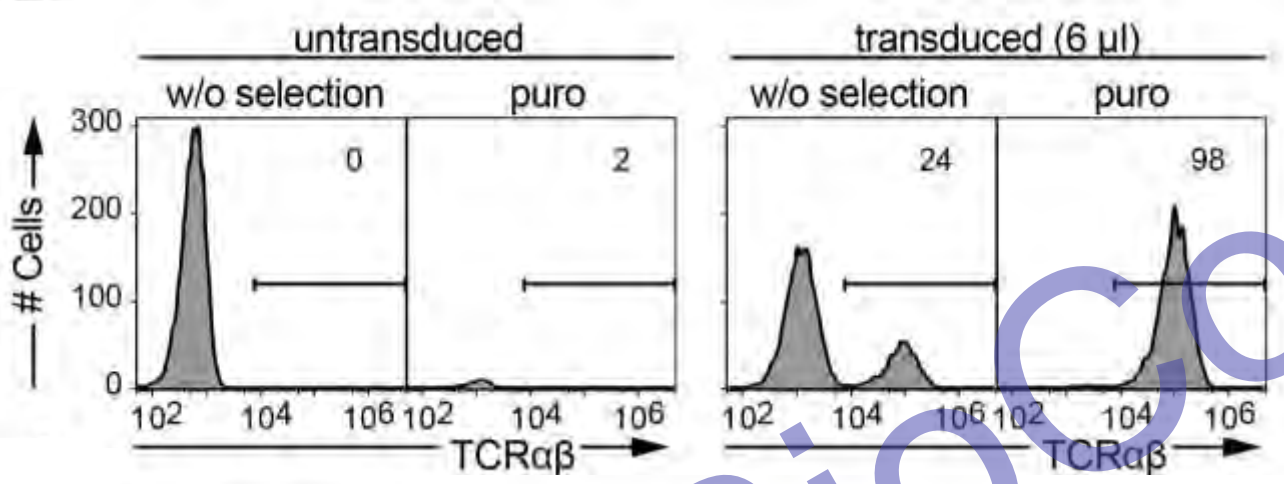
Can we select Jurkat T cells expressing murine TCR?





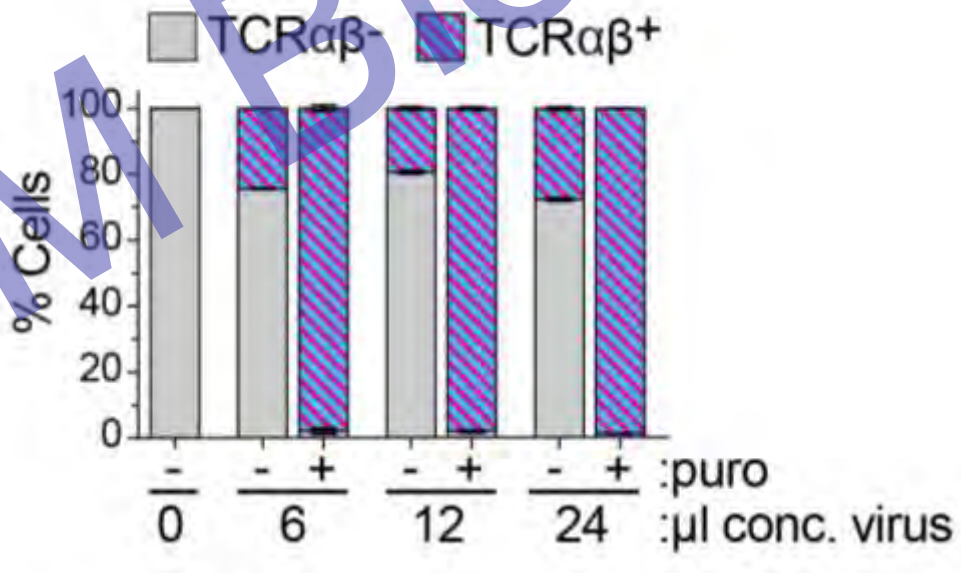


We get 100% pure cell population

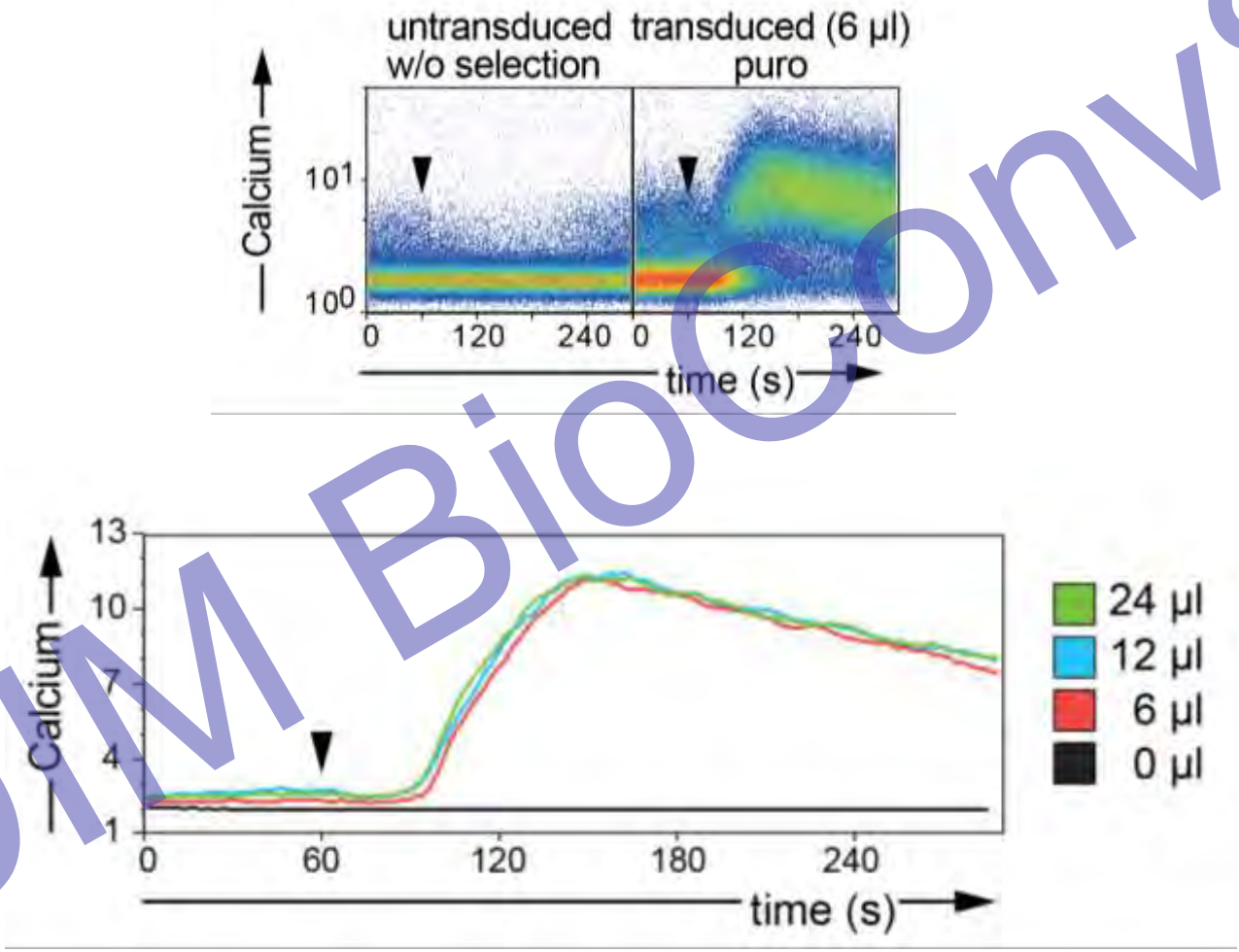


Claudia Juraske

Anna Morath



The selected cells are functional



Palanisamy *et al.*, Nature Commun., 2019

SiMPl is based on the naturally split gp41-1 intein

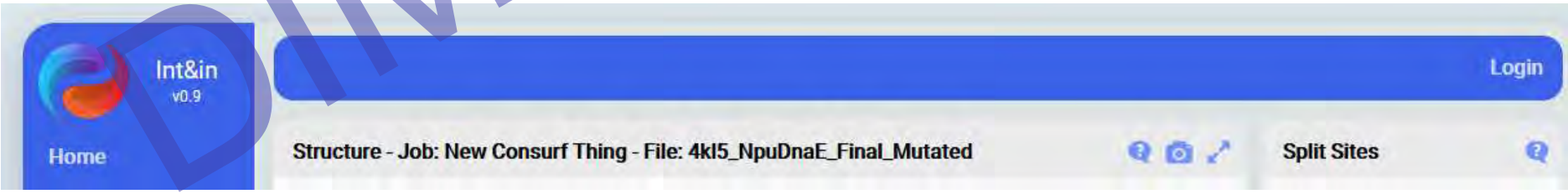
Splitting different inteins could be necessary

or

Splitting an intein at different positions

Mirko Schmitz
&
Mehmet Öztürk

We developed the web server **Int&in** to predict active split sites in inteins

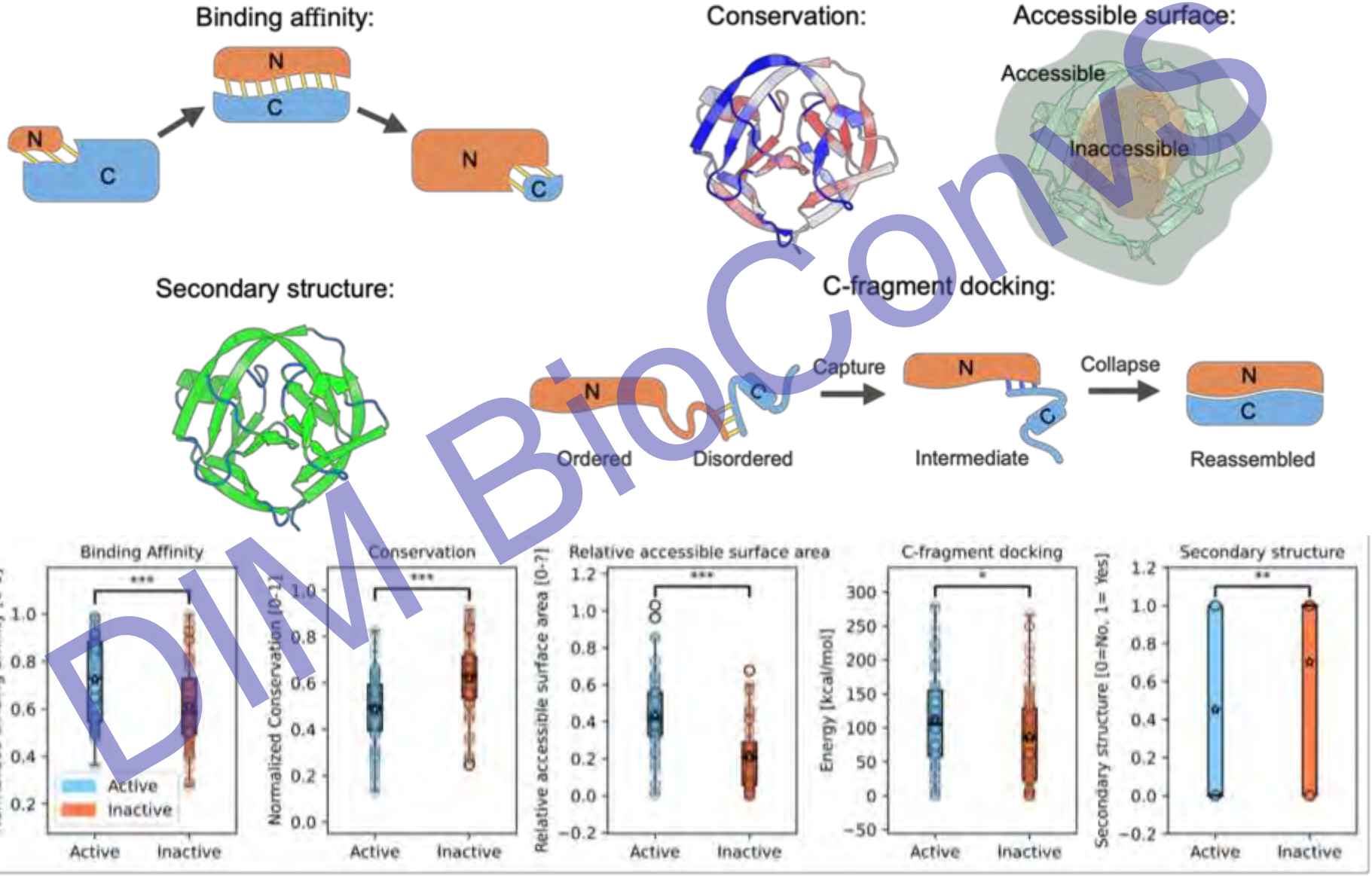


We gathered information on functionality for a large set of randomly distributed sites

Experimentally active

Experimentally inactive

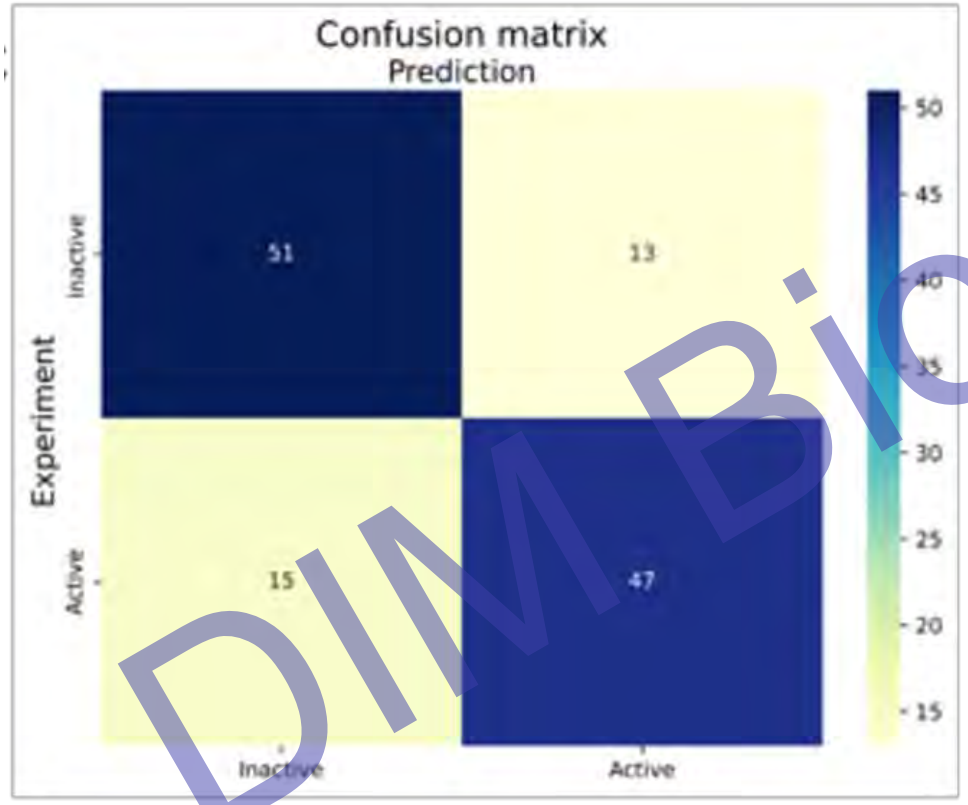
To create a predictive model we extracted structural and biochemical features



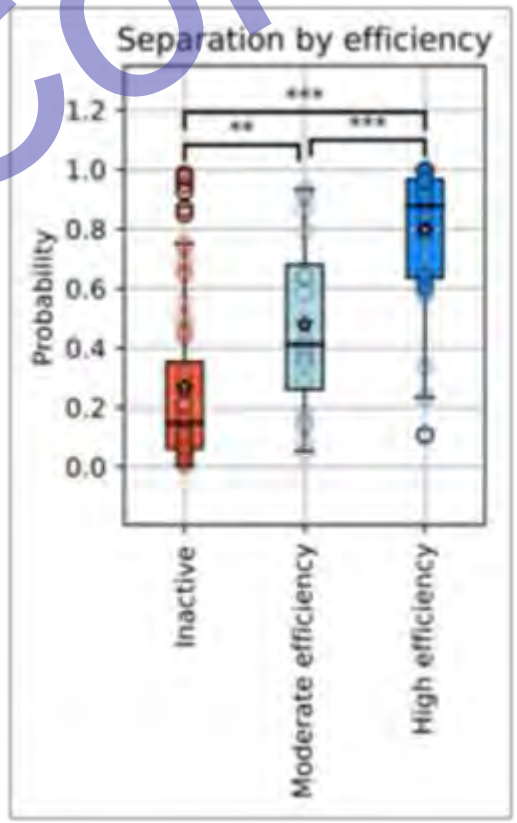
We can predict with high accuracy if a split site is active or inactive

Gaussian Naïve Bayes model

Training data set



The model distinguishes between different splicing efficiencies

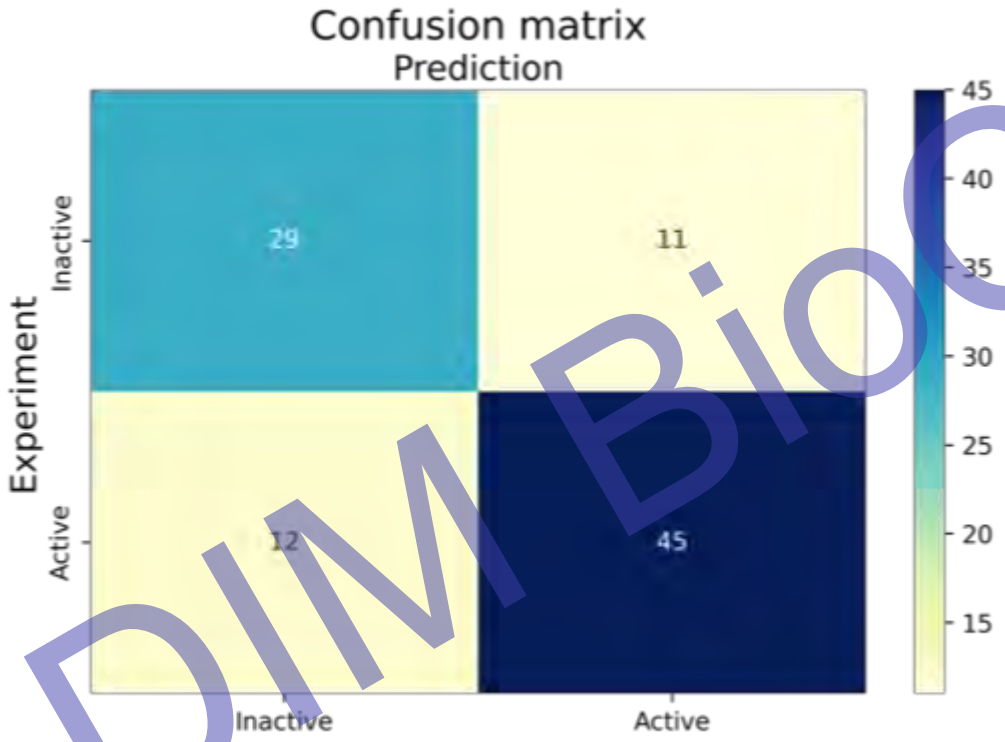


Accuracy: 0.78
Precision: 0.78
Recall: 0.76
MCC: 0.56

We tested the model on a data set taken from the literature

Testing data set: 97 split sites (57 active, 40 inactive), 41 different inteins

Aranko, Wlodawer and Iwai, Nature's recipe for splitting inteins. *Protein Eng Des Sel* **27**, 263-271 (2014)

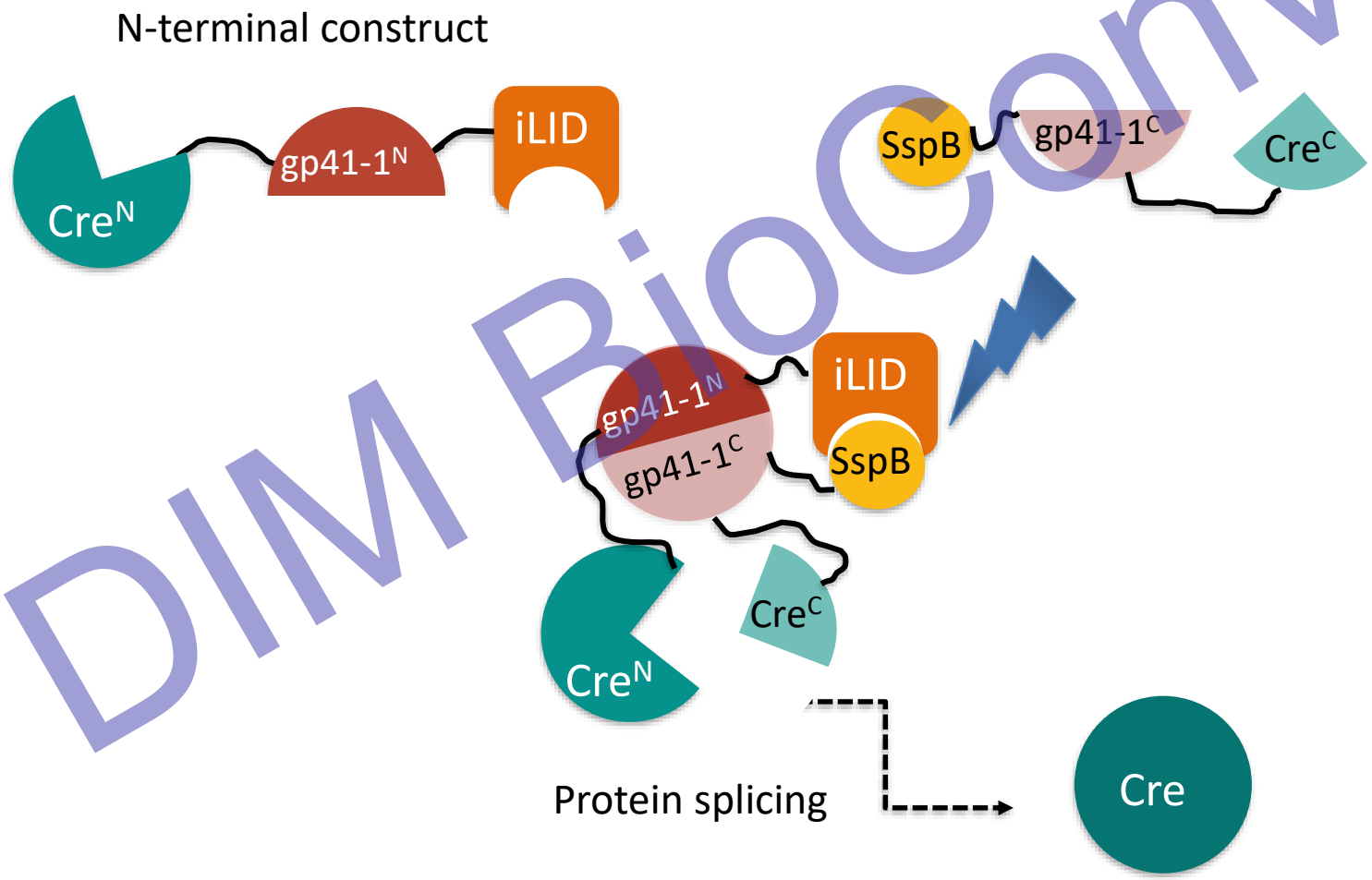


Accuracy: 0.76
Precision: 0.78
Recall: 0.79
MCC: 0.51

Submitted & on bioRxiv

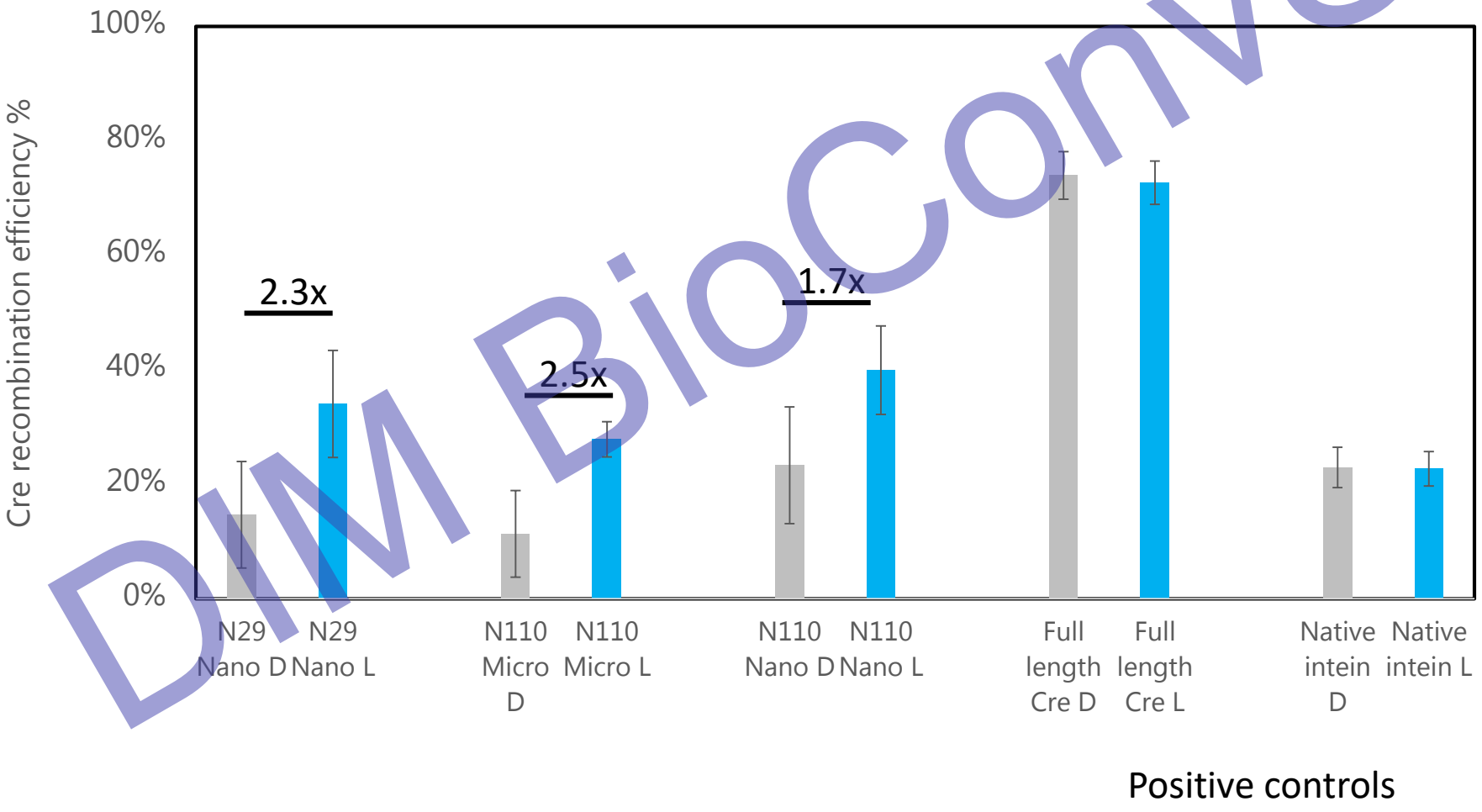
We employed Int&in to engineer novel conditional inteins

Test application: control the activity of the Cre recombinase in mammalian cells with blue light



We get higher recombination for cells exposed to blue light

Active Cre → EGFP gets expressed



Take-home messages

Inteins are super cool proteins!!

Circular proteins have interesting properties. They can be created using split inteins

Rational approach can be used to split enzymes conferring resistance towards antibiotics – 100% accuracy!

With SiMPl you can select cells containing two plasmids with a single antibiotic

SiMPl can be useful in metabolic engineering projects in bacteria

Int&in: a web server to predict active split sites in inteins

Int&in can be used to construct novel conditional inteins

Controlling Cre reconstitution with light is possible (but challenging)

Many thanks to…

All members of the Di Ventura lab

Anna Morath, Claudia Juraske and Wolfgang Schamel (Uni Freiburg)

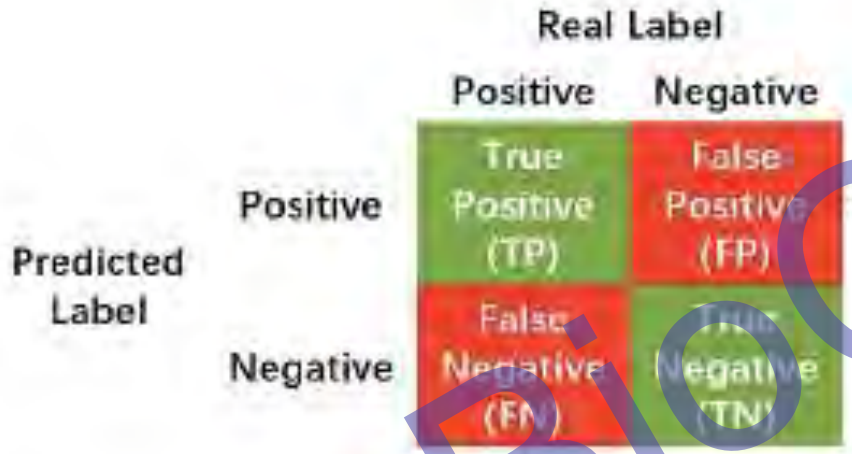
Jung-Won Youn and Georg Sprenger (Uni Stuttgart)

Karsten Voigt (Freiburg University)

Some definitions

MCC = Matthews Correlation Coefficient

Precision is how good the model is at predicting a specific category.



$$\text{Precision} = \frac{\sum TP}{\sum TP + FP}$$

Accuracy tells you how many times the ML model was correct overall.

$$\text{Recall} = \frac{\sum TP}{\sum TP + FN}$$

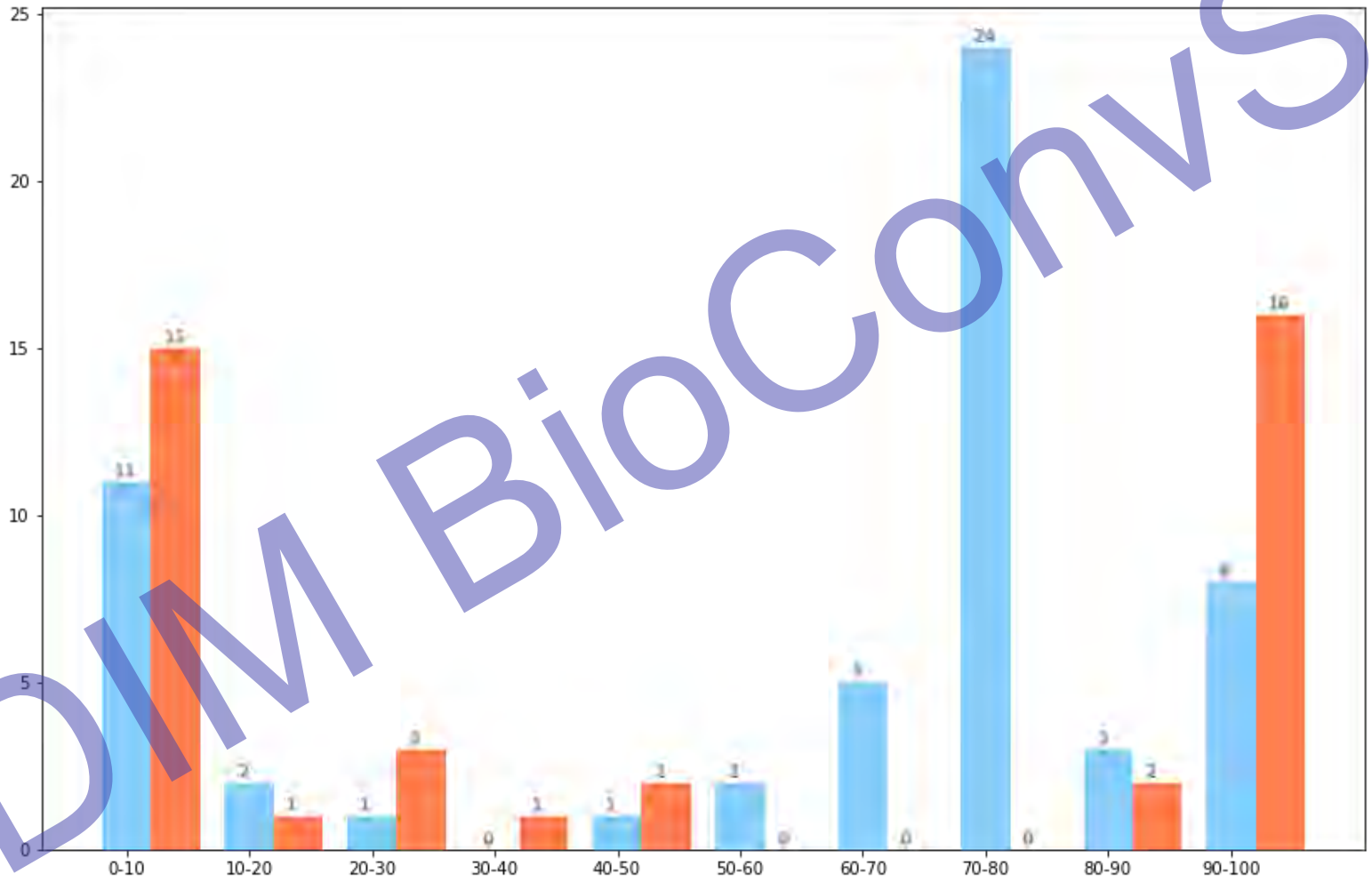
$$\text{Accuracy} = \frac{\sum TP + TN}{\sum TP + FP + FN + TN}$$

Recall tells you how many times the model was able to detect a specific category.

Ma *et al.*, IEEE ACCESS, 2023

The data found in the literature are biased towards active sites and certain positions

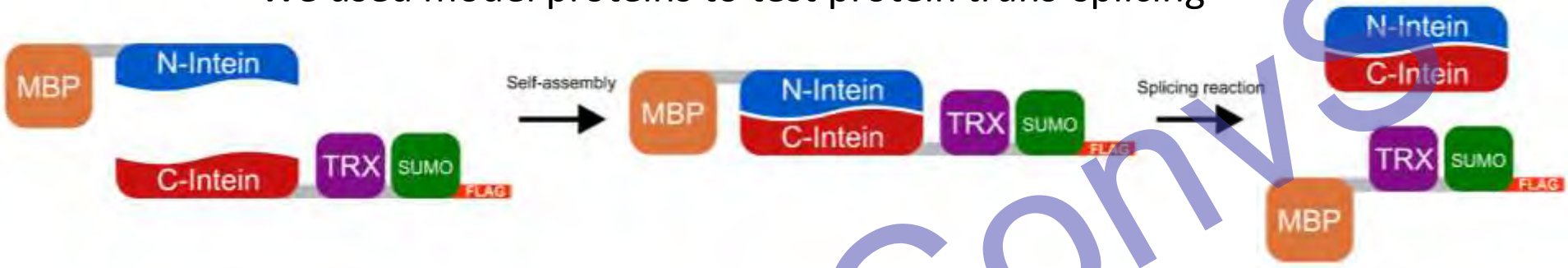
57 active vs 40 inactive sites



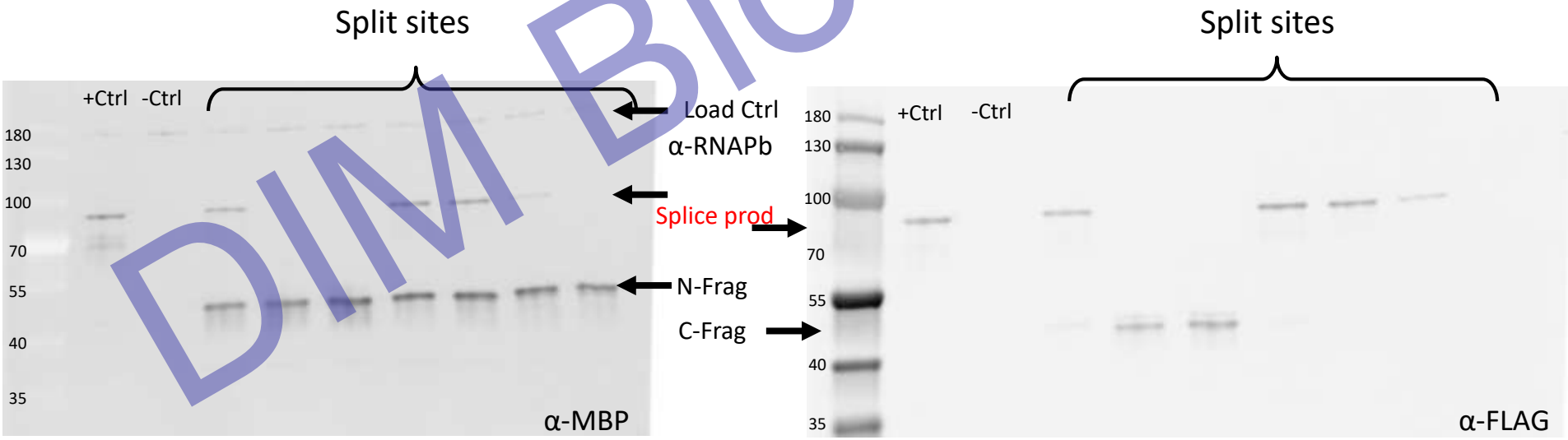
submitted

We decided to generate our own data set

We used model proteins to test protein *trans*-splicing

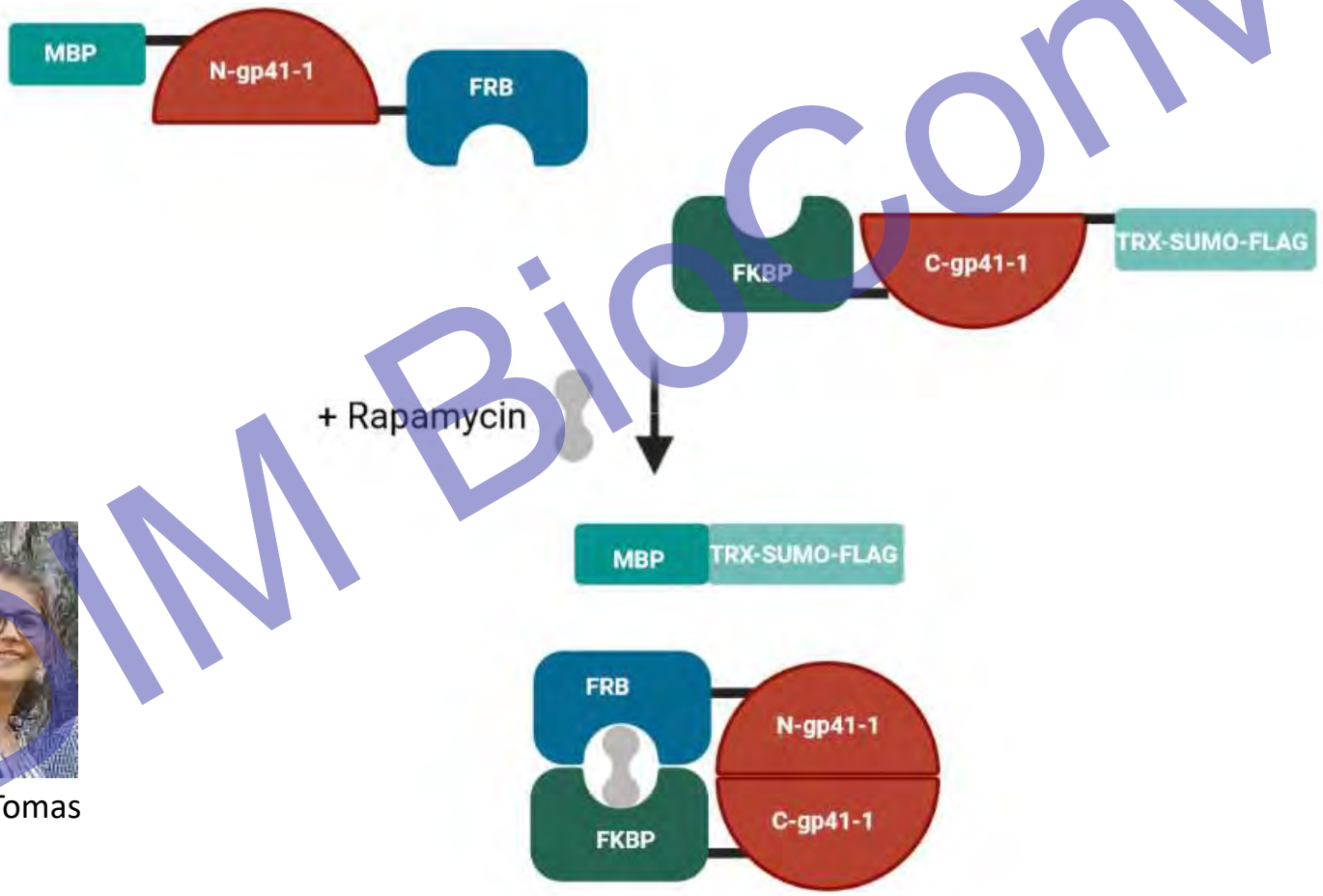


We performed Western blotting to quantify the efficiency of splicing



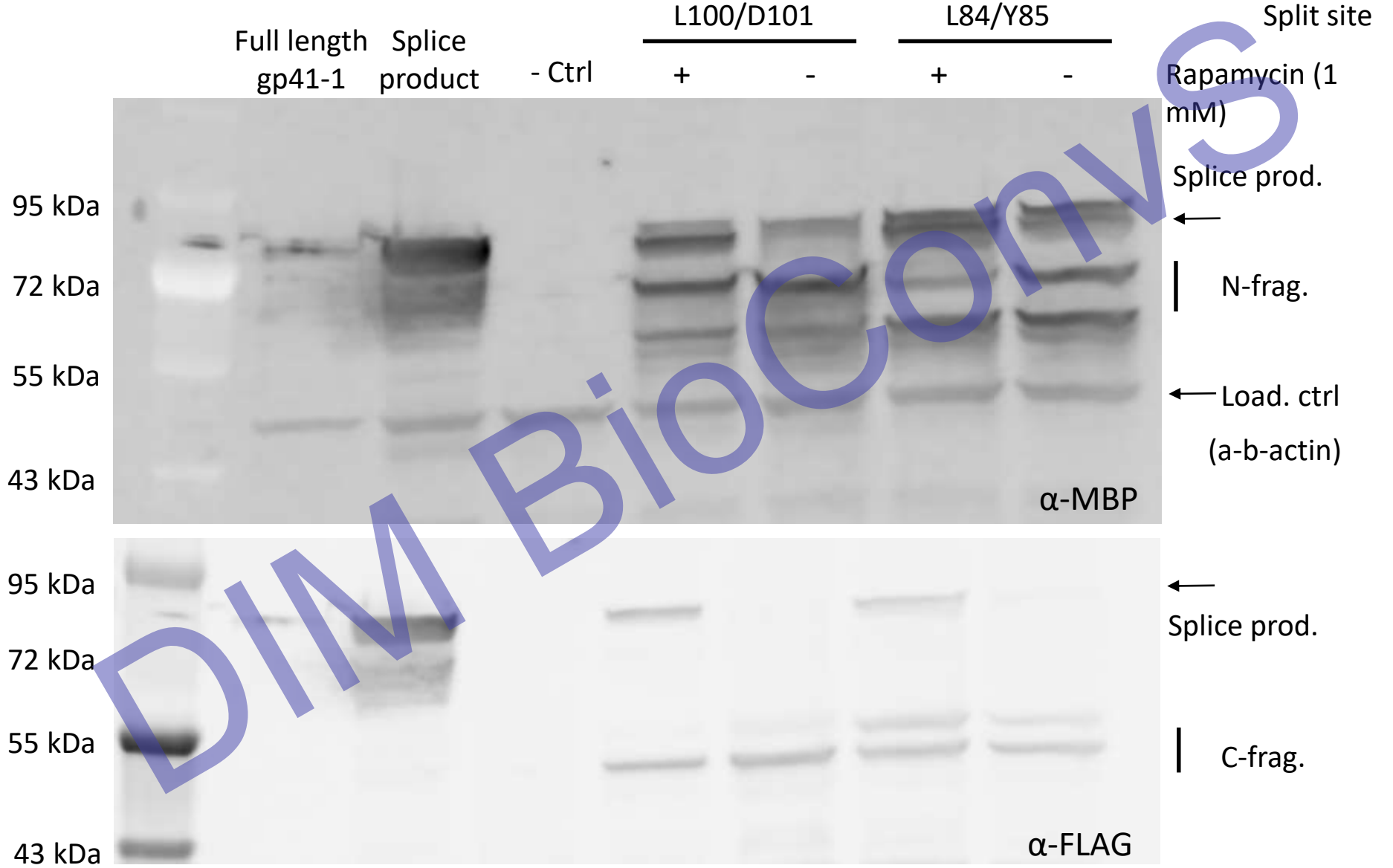
We can also explicitily predict sites at which the fragments have low affinity

We should be able to control the splicing reaction by regulating the interaction between the two intein fragments

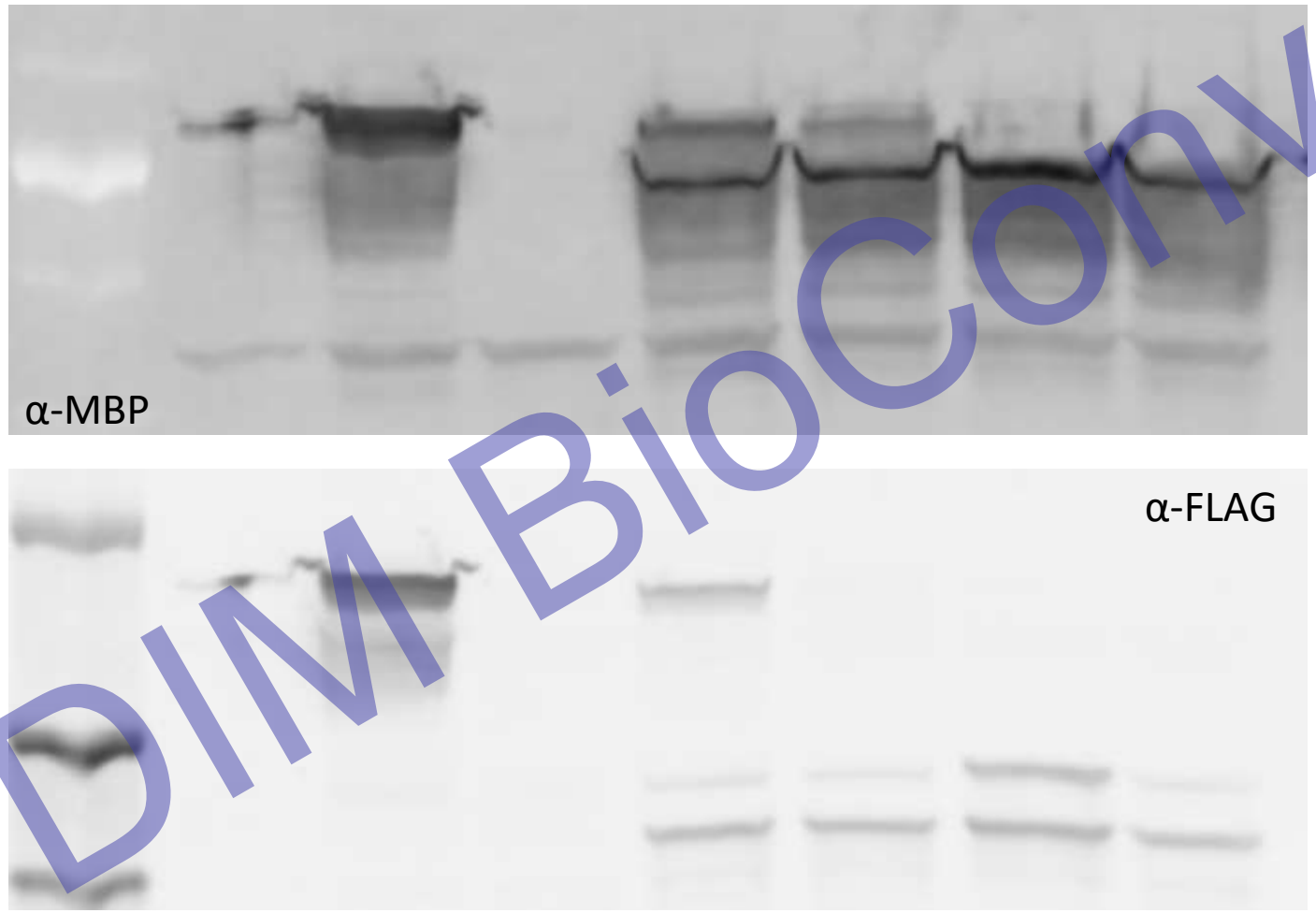


Franziska Tomas

We can increase splicing by the addition of rapamycin



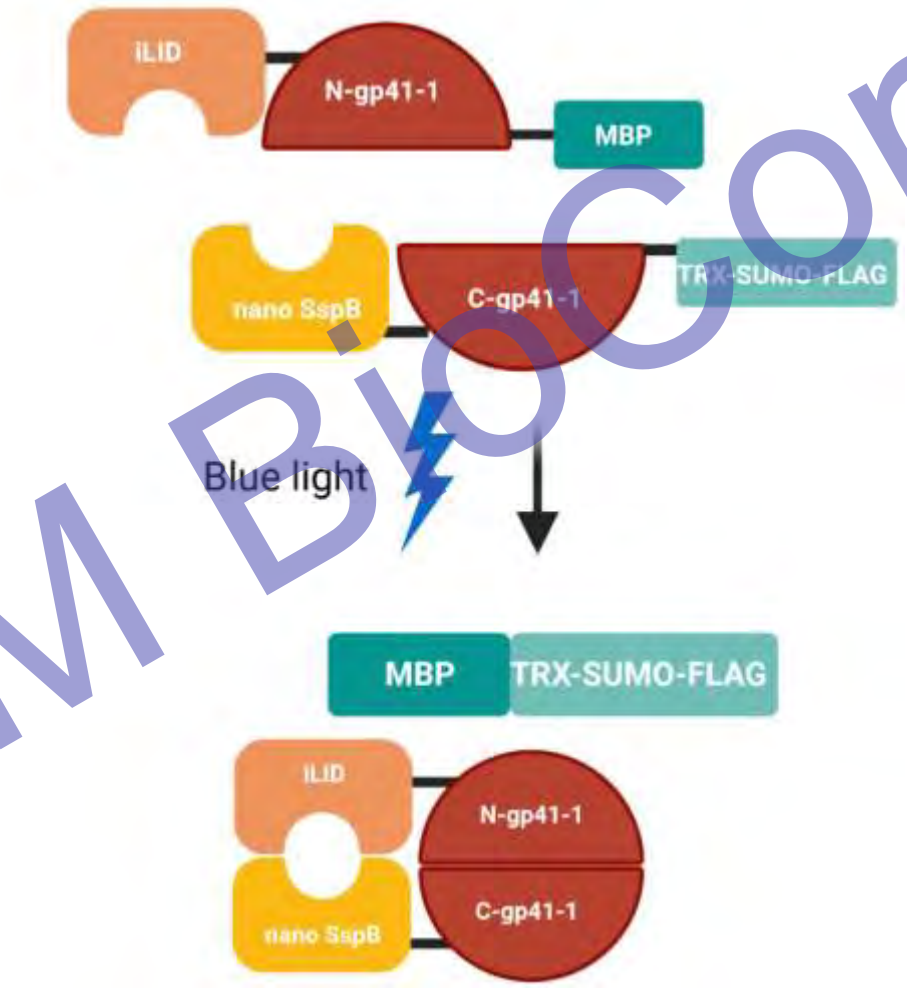
We can increase splicing by the addition of rapamycin



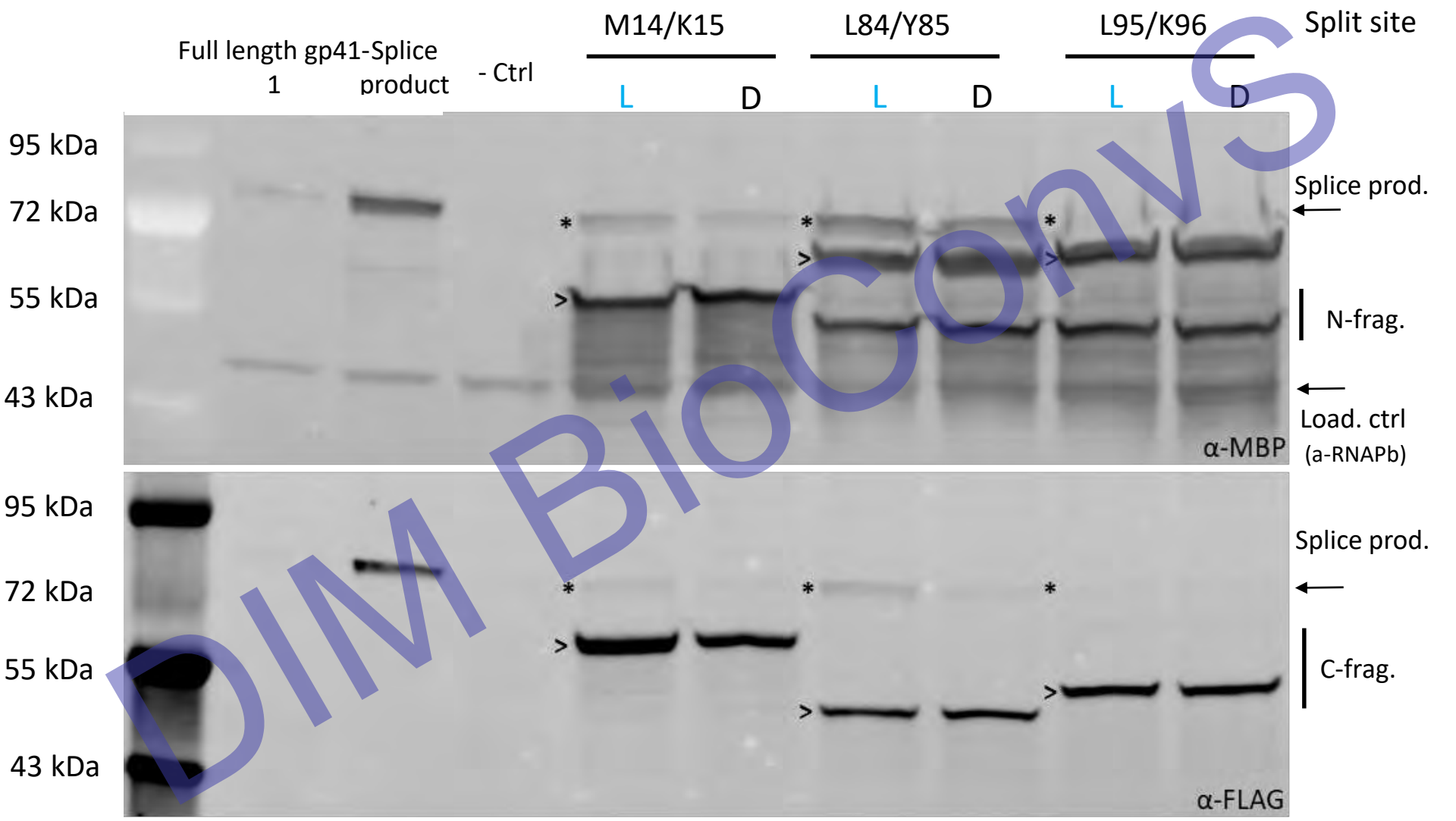
By using blue light-inducible dimerizers we should control the reaction with light

We selected the iLID system

Guntas *et al.*, PNAS (2015), 112: 112-117



We can increase splicing by illuminating the cells with blue light



We can increase splicing by illuminating the cells with blue light