

Bioengineering of single-stranded DNA binding protein to unravel the role of liquid-liquid phase separation in bacterial genome maintenance

OTDK essay

biology MA

by:

Dávid Érfalvy

Supervisors:

Dr. Péter Ecsédi

Dr. Mihály Kovács

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

TERMÉSZETTUDOMÁNYI KAR

BIOLÓGIAI INTÉZET

Budapest, 2024



Table of contents

Abbreviations	3
Abstract	4
Introduction	4
Scientific Aims:	7
Materials and Methods	7
Design and construction of different EcSSB variants	7
Cloning of EcSSB variants	8
Expression and purification of EcSSB variants	9
Fluorescence polarization measurements	11
Turbidity measurements	13
Microscopy	14
Results	14
Expression and purification of EcSSB variants	14
Studying the ssDNA binding of SSB variants	14
RecQ and OB domain binding abilities of variants	16
Turbidity measurements	17
Microscopic analyses of formed protein droplets	19
Discussion	19
Author's contributions	21
Literature	22
Supplementary	23

Abbreviations

SSB: single-stranded DNA binding protein

EcSSB: *E. coli* SSB

hSSB: human SSB

PfSSB: *Plasmodium falciparum* SSB

LLPS: liquid-liquid phase separation

OB domain: oligonucleotide binding domain

IDL: intrinsically disordered linker

CTP: C-terminal peptide

ssDNA: single-stranded DNA

PEG: polyethylene-glycol

WT: wild type

flCTP: fluorescein labeled CTP

Abstract

It is now known that various membraneless cellular structures in eukaryotes (nucleolus, postsynaptic density, heterochromatin etc.) are liquid-liquid phase-separated (LLPS) condensates of proteins and nucleic acids. Such structures enable molecules to accumulate and function in special microenvironments, which could be especially advantageous for bacterial cells lacking inner membrane systems. Despite this, only a small number of proteins with LLPS propensity are known in prokaryotes. Our lab showed in 2020 that *E. coli* single-stranded DNA (ssDNA) binding protein (SSB) confers such propensity, highlighting a new context for its genome metabolic function. My goal is to unravel the importance of SSB-driven LLPS in bacterial physiology and stress response. As a first objective, I bioengineered EcSSB to retain its ssDNA- and protein-binding functions but abolish its LLPS propensity. I created and purified an array of protein variants affecting the SSB's intrinsically disordered linker (IDL) region, which is thought to be the main driver of LLPS. Many of these SSB variants showed selective reduction or loss of LLPS propensity, while retaining nucleic acid and protein binding capacities. We use these engineered variants to inspect if LLPS formation gives adaptive advantage to live bacterial cells under normal and stress conditions. Elucidation of the cellular mechanisms of SSB-driven LLPS is also advantageous in the development of new LLPS-inhibiting antimicrobial substances.

Introduction

Compartmentalization of the intracellular space is essential for the maintenance of every complex organism. It allows reactions demanding different environments, to take place simultaneously, molecules to be stored separately and released to the cytoplasm only when needed. In eukaryotes, independent organelles form from endomembrane systems, like the endoplasmic reticulum or the Golgi apparatus, entirely separating their insides away from the cytoplasm. Note here, that to a lesser extent prokaryotes can achieve a similar effect with the intrusion of their cell membrane (e.g. prokaryotic photosynthesis [1]), but compartmentalization via membranes in their case is less defined. In the last decades however, one phenomenon called liquid-liquid phase separation (LLPS), although known for more than a century (Kossel et al, 1897), has gained more attention as a strong tool to form more dynamic types of compartments without lipid bilayers [2, 3]. These condensates are generated by proteins and/or nucleic acids and can be observed in eukaryotes as membraneless organelles with various functions (**Fig. 1.**). For example, P-bodies are perinuclear formations of guanyl-specific ribonucleases (PGL1 and PGL3) associated with the nuclear pore complexes and are responsible for germline gene regulation and intergenerational small RNA inheritance [4]. Other examples are stress granules, which can be detected in response to a wide range of stresses and accumulate large amounts of mRNA stalled in translation initiation, many translation-related proteins and ribonucleoproteins [4] [5]. Despite our extensive knowledge in eukaryotes, we know of only few phase separating proteins in prokaryotes, although it may be of particular importance in their case.

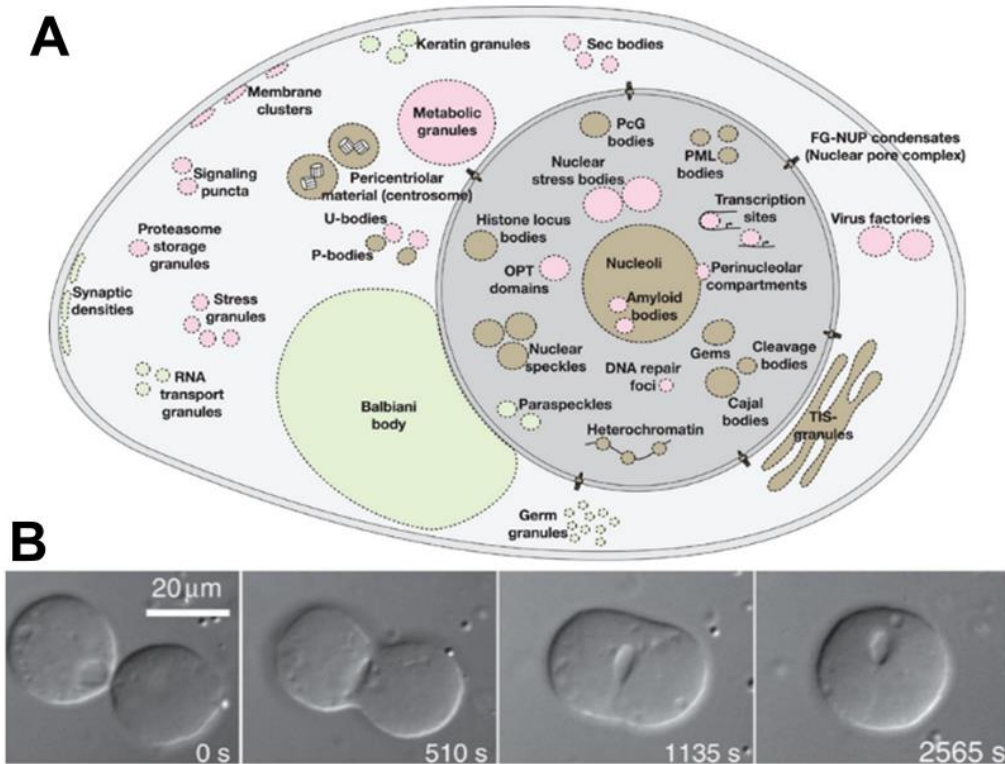


Fig. 1. The wide variety of membraneless organelles in eukaryotes and the appearance of LLPS droplets (A) LLPS enables the formation of very dynamic condensates and various membraneless organelles in eukaryotes [6]. (B) Spherical nucleoli from *Xenopus laevis* germinal vesicles coalesce into a larger sphere, an example of the quick dynamics of membraneless organelles[2].

In 2020 however, our research group suggested and proved that the *E. coli* single-stranded DNA (ssDNA) binding protein (SSB or EcSSB in case of *E. coli*) can form protein condensates under physiological conditions [7], thus expanding the number of known LLPS driver molecules in prokaryotes. SSBs are highly conserved, their homologues are present in the majority of bacteria and critical in genome maintenance. Although these proteins are extensively studied, our findings highlighted a new context for their genome metabolic function and raised the questions if SSB driven LLPS is presented in a wide range of bacteria and if LLPS confers any adaptive advantage to them. Answering these questions, starting with *E. coli*, is the main goal of my long-term research. In this work, I present my first steps towards my aims.

SSBs contain two main structural elements; the oligonucleotide-binding (OB) domain, responsible for single-stranded DNA/RNA binding, and the intrinsically disordered linker (IDL) driving protein-protein interactions. It is important to note that the very end of the IDL contains 8 highly conserved residues usually called the C-terminal peptide (CTP) (**Fig. 2.**). It currently seems that all three structural units are important for LLPS, at least in case of *E. coli*. The OB domains assemble into functional tetramers allowing multivalency, a general factor in LLPS propensitive proteins. CTPs within the tetramers can

form intra- and intermolecular interactions with the OB domains promoting protein association. Because of this, it was believed that the CTP – OB domain interaction drives droplet formation, however SSB variants with missing CTP can also form protein condensates when a crowder (e.g. polyethylene glycol) is present [8] [7]. Moreover, mutants lacking the IDL except the CTP completely lose their phase separation forming ability [7]. By examining the IDL of EcSSB, it is clear that the N-terminal end contains several glycine rich regions. Note here that these sequences are highly important in case of other proteins prone to form LLPS droplets [9] [10]. Moreover, arginine or aromatic residues in disordered regions can also promote phase separation and/or self-assembly of many proteins [11] and the IDL of EcSSB contains some of these residues at key positions [12]. These regions and residues were not studied separately before; thus, their impact is still not well understood. Interestingly, some SSB homologues can't perform LLPS, like in case of *Plasmodium falciparum* (PfSSB) (Uniprot code: Q8I415). This is presumably caused by the low amount of glycine and the high amount of negatively charged amino acids in the IDL [13]. Human SSB1 (hSSB1) (Uniprot code: Q9BQ15) (*unpublished data, Biorxiv: 2023.07.25.550517*) and SSB2 (hSSB2) (Uniprot code: Q96AH0) (*unpublished data, Biorxiv: 2023.09.15.557979*) can also form LLPS droplets, but with different mechanisms.

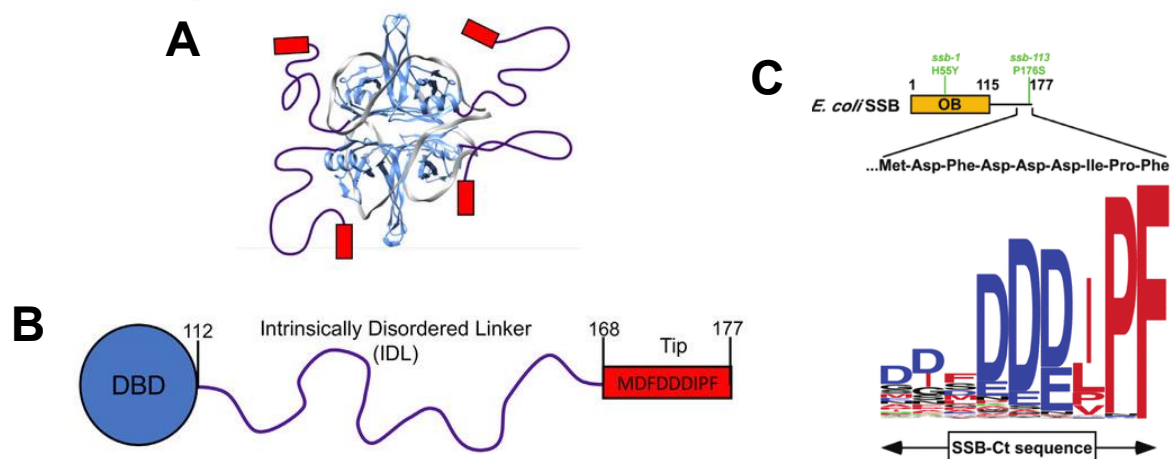


Fig. 2. Structures of the EcSSB in (A) tetrameric and (B) monomeric forms and (C) the LOGO sequence of the CTP. (A) The IDLs are schematically represented because they cannot be modeled in X-ray structural studies. (B) Numbers indicates the amino acid residues in the sequences. DBD is the DNA binding domain, what we rather name oligonucleotide binding domain, as it also binds RNA [14]. The red rectangle represents the C-terminal “acidic tip” motif of the protein, which is also called C-terminal peptide (CTP). (C) The sequence of the EcSSB CTP is displayed with its conservation across 280 eubacterial species represented as a LOGO, in which the height of the residue relates to its frequency at the given position. LOGO residues are colored to indicate the hydrophobic (red), the negatively charged (blue), the polar (black) and the positively charged (green) side chains. [15]

Scientific Aims

- Cloning, expressing and purifying different EcSSB constructs in which the IDL is modified.
- Observing the ssDNA and partner binding abilities of these variants.
- Studying their LLPS propensity and evaluate how IDLs from different organisms, glycine rich regions, aromatic and cationic residues impact the LLPS propensity of EcSSB.
- Applying these results to select variants carrying the nucleic acid and partner binding abilities of the WT protein, without the propensity of forming protein condensates. These constructs then can be introduced into the genome of *E. coli* to study if LLPS alone confers adaptive advantages to bacteria.

Materials and Methods

Design and construction of different EcSSB variants

To reach my scientific aims, I designed and produced several SSB variants (**Fig. 5.**). The most drastic changes were applied in cases of GG, (GGS)₄, EPE and EHE constructs, where the whole IDL of the EcSSB (except CTP) was eliminated (GG) or changed either to a GS linker ((GGS)₄) or to the sequences of *PfSSB* (EPE) or hSSB1 (EHE) IDLs. d1G and d2G are well-defined IDL deletion mutants to study the impact of glycine-rich regions. In d1G the sequence of 114-GGRQGGG-120 while in d2G 114-GGRQGGGAPAGGNIGGG-130 is deleted. I designed a specific mutant to study if cation- π interactions, formed by aromatic and cationic residues, are key for EcSSB-driven LLPS (R116D-W136D). Here, a tryptophan (W136) located in the central part of the IDL, and an arginine (R116) presented in the glycine rich region were substituted to aspartates. This not only diminishes the proposed cation- π interactions but potentially introduces ionic repulsions between IDLs. To determine the effects of CTP deficiency I used a previously produced construct missing the CTP (dC). Following the production of these EcSSB variants I could measure their DNA, RecQ and OB domain binding abilities together with their LLPS propensity and I could compare them with the WT protein.

Cloning of EcSSB variants

First, a His₆-tag containing modified pET29b vector was generated with a Tobacco Etch virus (TEV) cleaving site between the tag and the multicloning site, which was used later in case of all variants (**Fig. 3.**).



Fig. 3. Modification of pET29b vector A small DNA sequence shown here was inserted into pET29b using *NdeI* (CATATG site) and *XhoI* (CTCGAG site) restriction endonucleases. The yellow sections show the location of the His₆-tag and the Tobacco Etch Virus (TEV) cleavage site, while grey shows sequences coding GS linkers. An additional *BamHI* (GGATCC) site was also inserted, allowing the modular cloning of EcSSB variants.

To produce the complex sequences of EcSSB variants (precisely described in the Results section), different PCR based methods were applied. The plasmid containing the DNA sequence of PfSSB was bought from Addgene, while the sequences of EcSSB and hSSB1 were already in the laboratory. The sequences of oligos are in Table 1. (Supplementary) and were ordered from IDT.

QuickChange technique with Pfu ultra polymerase was used to produce the dC and R116D-W136D mutants. An additional 4% dimethyl-sulfoxide (DMSO) was applied to the reactions to increase its success rate. Mutations in R116D-W136D were introduced sequentially and not simultaneously.

Megaprimer PCR was applied to produce EPE, EHE, GG, d1G, d2G and (GGG)₄ variants. Here the constructs were produced in two steps. First, megaprimers were generated in one PCR reaction where the targeted IDL (e. g. human IDL with the sequence of CTP from EcSSB) sequences were amplified. This megaprimer was then used in a second PCR where the sequence of the OB domain (EcSSB) was connected (**Fig. 4.**). In case of megaprimer PCRs, KOD and polymerase was applied.

The prepared solutions were placed in a BLUE-RAY Turbocycler PCR device and were subjected to 1 min annealing at 55°C and 12 min elongation at 72 °C, for a total of 18 cycles in case of QuickChange and to 1 min elongation at 70 °C with 30 cycles when megaprimer PCRs were carried out.

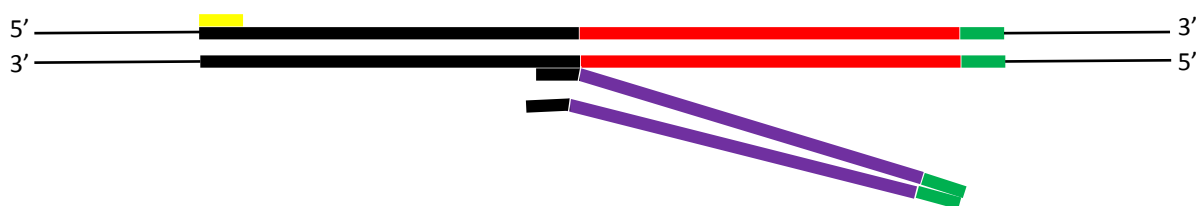


Fig. 4. Representation of megaprimer PCR The forward primer (yellow) is a conventional single stranded DNA oligo which can anneal with the DNA sequence of the OB domain of EcSSB (black). The megaprimers are double stranded oligos produced by a PCR run with to produce the needed IDL sequence. They contain a complementary sequence (black), which recognizes the end of the OB domain DNA sequence, and a specific IDL sequence (purple) with the CTP sequence (also from EcSSB similarly to the OB domain).

The produced DNA constructs were purified using the QUIAGEN QIAquick PCR purification kit and protocol. Constructs and the vector were then digested with BamHI-HF and XhoI-HF restriction endonucleases. Digestion was carried out for 1 h at 37 °C. Following digestion, a further purification of the constructs was performed. The resulting products were ligated into the N-terminal His₆-tag containing pET29b vector using T4 DNA ligase. For transformation, DH5α cells were heat-shocked for 1 min at 42°C. Transformed cells were inoculated onto Luria-Bertani (LB) medium plates containing ampicillin (100 µg/ml) and were grown overnight at 37°C. The next day, colonies were screened using PCR performed with Taq polymerase to determine whether the colonies contained the vectors with the inserted sequences or only the empty vector. Selected colonies were transferred to a test tube containing 3 ml LB and 100 µg/ml ampicillin and incubated overnight at 37°C. Minipreps were isolated from the grown cultures with the QIAprep Spin Miniprep Kit. Plasmids were sequenced by LGC Genomics.

Expression and purification of EcSSB variants

BER cells were transformed with plasmids containing the sequences of EcSSB variants and grew in LB containing 100 µg/ml ampicillin at 37°C. The expression was initiated using 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG), when the opacity of the cultures reached the absorbance of 0.8 at 600 nm, then they were incubated at 18°C ON. On the following day cells were pelleted (5000 g for 8 mins) then resuspended in a buffer containing 50 mM Tris pH 8, 500 mM NaCl, 30 mM imidazole and 3 mM NaN₃. Homogenization was performed by an ultrasound homogenizer (settings: duty cycle: 80%, output control: 8). After ultrasonication, the samples were centrifuged at 48000 g for 20 mins and the supernatants were applied to Ni affinity column. Here two buffers were used for the purification: a washing solution (buffer A: 50 mM Tris pH8, 500 mM NaCl and 30 mM imidazole) and an elution buffer (buffer B: completing buffer A with 250 mM imidazole). After the samples were applied and the column washed, the proteins were eluted slowly by a gradient, replacing buffer A with buffer B (1%

increase/min) and the elution peaks were collected. The gradient elution was necessary to prevent tetramers containing not only the target protein but the bulk WT EcSSBs. Tetramers with four His₆-tags eluted last, thus allowing us to separate heterotetramers from homotetramers. Collected samples were analyzed using SDS PAGE to check their homogeneity. Samples containing only the modified constructs were digested using TEV protease. The cleaved proteins were diluted until the imidazole concentration fell below 50 mM and were applied to reverse Ni column. The flowthrough was then added to Q HP (GE) anion exchange column for further purification. For this 20 mM TRIS, pH8 and 150 mM NaCl containing solution was used as buffer A. Buffer B contained 1 M NaCl and it was added in a gradient (1%/min). The eluted constructs were concentrated using Amicon Ultra centrifuge filters and stored at -80°C.

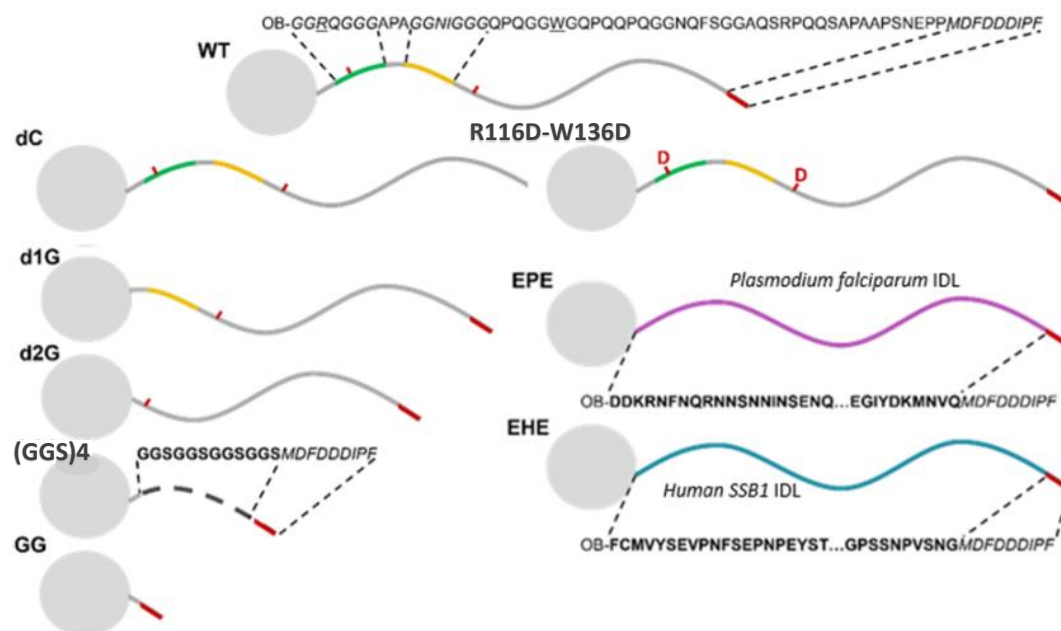


Fig. 5. Schematic representation of our EcSSB constructs: Grey circles and lines show the OB domains and the IDLs from EcSSB respectively, while red lines at the ends show the CTPs also from EcSSB. In EPE the IDL of PfSSB is in purple while in EHE the IDL from human is in teal. The green and orange sections show the glycine rich regions of EcSSB IDL. Residue substitutions are also marked. The sequences are also shown at constructs it helps understanding.

Fluorescence polarization measurements

Fluorescence polarization (FP) can be used to determine the binding constants and kinetics of chemical reactions by altering the rotational time of molecules, thereby causing the light emitted by a fluorophore to have unequal intensities along different axes of polarization. It requires a slower moving larger (proteins like EcSSB) and a fluorescently labeled faster moving smaller molecule (a peptide or short nucleic acid sequence). In principle, by using different polar filters, we can define the rotation speed of the labeled molecule. Faster rotation causes less light to be detected because of the more depolarized emission getting through polar filters less successfully, whereas slower rotation allows more light to be observed by the detector due to the emitted light staying more polarized. Since binding to a partner forms a slower moving larger complex, molecule associations allow more light to be detected causing signal increase, while dissociation has the opposite effect. (Fig. 6.)

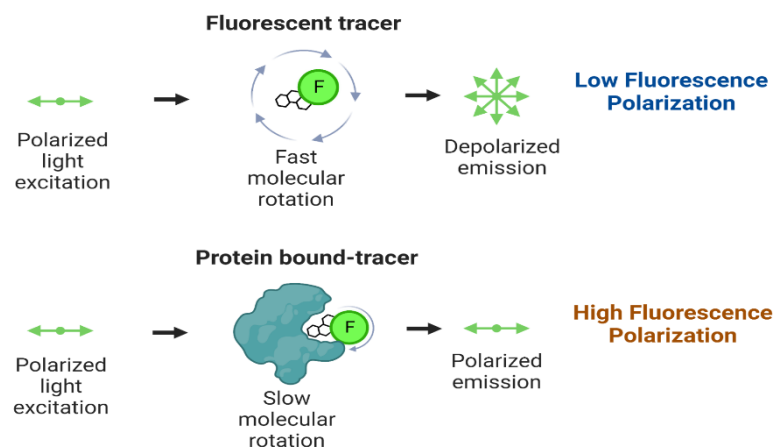


Fig. 6. The basics of fluorescence polarization (<https://bpsbioscience.com/product-types/biochemical-assay-kits-by-format-type/fluorescence-polarization>)

For DNA binding, fluorescein labeled single-stranded DNAs (36 and 79 nucleotides long) (Sigma-Aldrich) were used at 10 nM. The EcSSB constructs were applied in halving dilutions. (Fig. 7)

For the analysis of the binding capability of SSB variants to the RecQ helicase, which is one of the main interaction partner of EcSSB; competitive FP measurements were carried out. It is known that the last few residues of CTP binds the RecQ. Because of this, the RecQ – fluorescein labelled CTP (fICTP) (Sigma-Aldrich) interaction was determined first by using direct titration method (10 nM fICTP was titrated with RecQ halving dilutions starting from 10 μ M). During competitive measurements the determined RecQ – fICTP complex (RecQ at 0.65 μ M and fICTP at 10 nM) was titrated with the EcSSB variants, and the competition of the fICTP was followed. (Fig. 7.) We also performed a competitive titration similar to the RecQ measurements, only using the dC construct instead of RecQ as a central CTP binding molecule at 2,5 μ M constant concentration (fICTP stays the same). This took place after the direct titration probes between 10 nM fICTP and halving dilutions of dC starting from 20 μ M (tetrameric concentration). (Fig. 7.) (Fig. 9.)

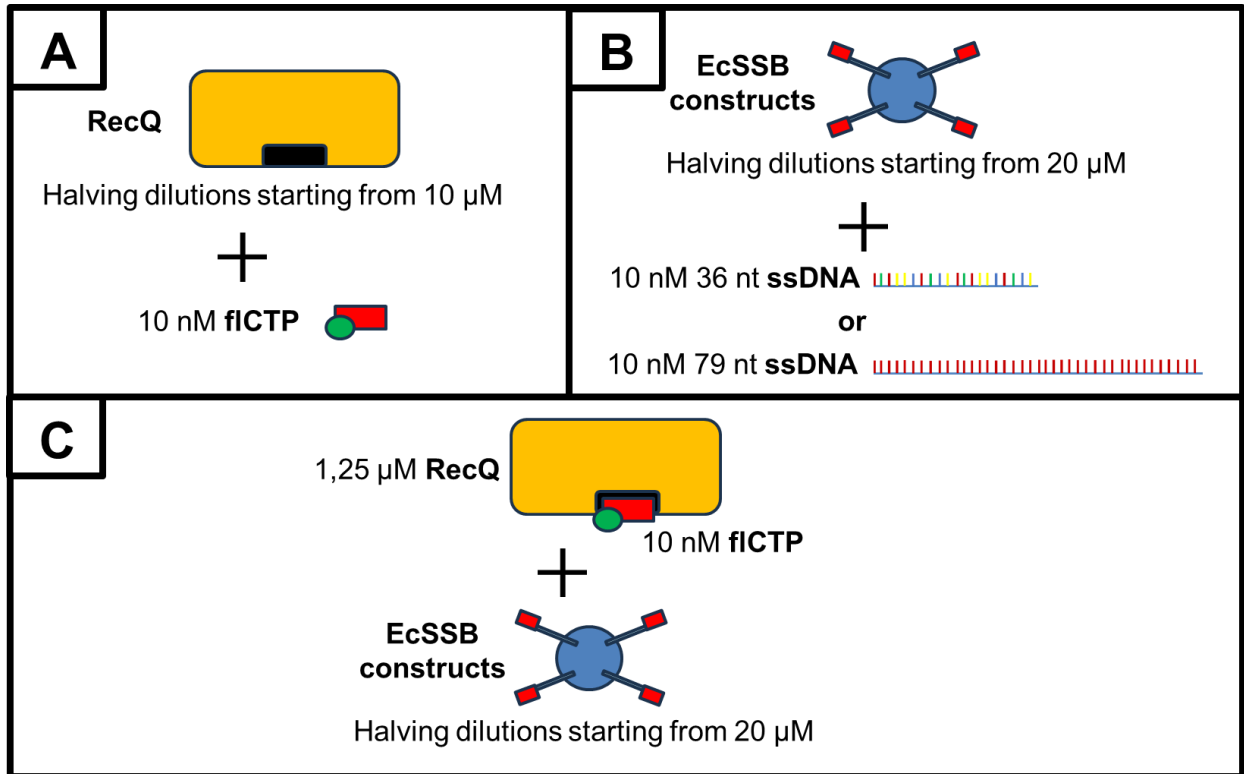


Fig. 7. Schematic representation of the conducted FP experiments: *A:* Preliminary direct FP for measuring the affinity between RecQ protein and fluorescently labeled CTP (fICTP). *B:* Direct FP for elucidating the affinity between our EcSSB constructs and 2 types of fluorescently labelled ssDNAs having different length and composition (the 36 nucleotides long is heterogenous nucleotide-wise, the 79 nucleotides long only consists of deoxy-thymidines). *C:* Competitive FP where fICTP competed against our EcSSB constructs for the same binding site of the RecQ protein

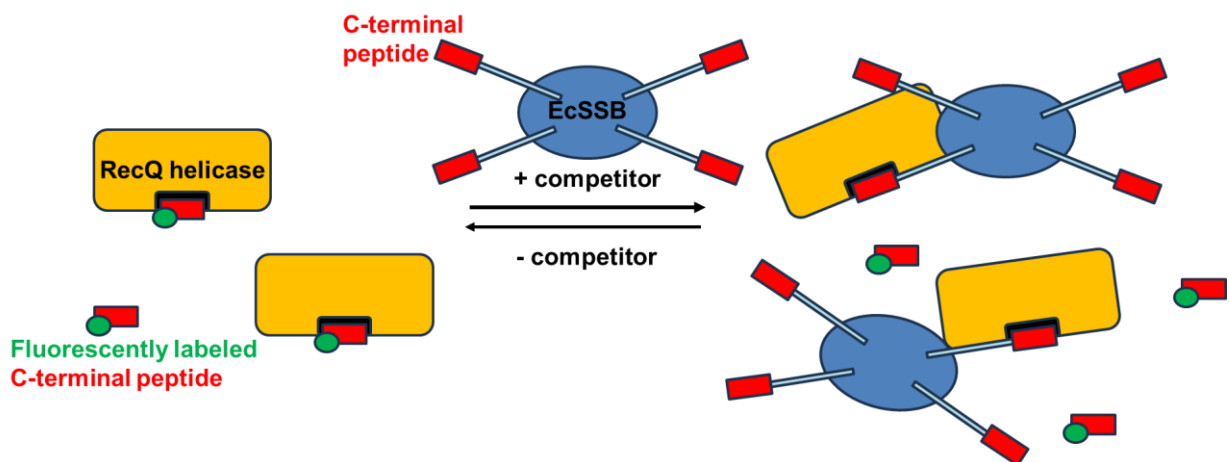


Fig. 8. Schematic representation of the competitive fluorescence polarization of the EcSSB tetramers' CTPs competing against fluorescently labeled CTPs bound to RecQ helicases. The blue EcSSB images are representative of all EcSSB variants previously mentioned except for the dC construct. All EcSSB are naturally found in tetramers; the four OB domains are represented by the blue oval, the light blue rods are the IDL regions and the CTPs are depicted as red rectangles.

These measurements were performed on a BioTek Synergy H4 Hybrid Reader measuring fluorescence polarization with the following settings: Normal, Excitation: 485/20, Emission: 528/20, Optics position: Top 510 nm, Gain: 105, Read height: 7 mm. The measurements were evaluated using Origin program.

In the case of direct titrations the K_d values were computed with the quadratic binding equation, but we used a different equation for competitive titration generated by Dr. László Radnai (Supplementary).

All measurements were carried out in triplicates and in a buffer containing 20 mM Tris-acetate pH 8,50 mM K-glutamate and 10 mM Mg-acetate. Errors are present with bars and show the standard deviation (SD) values.

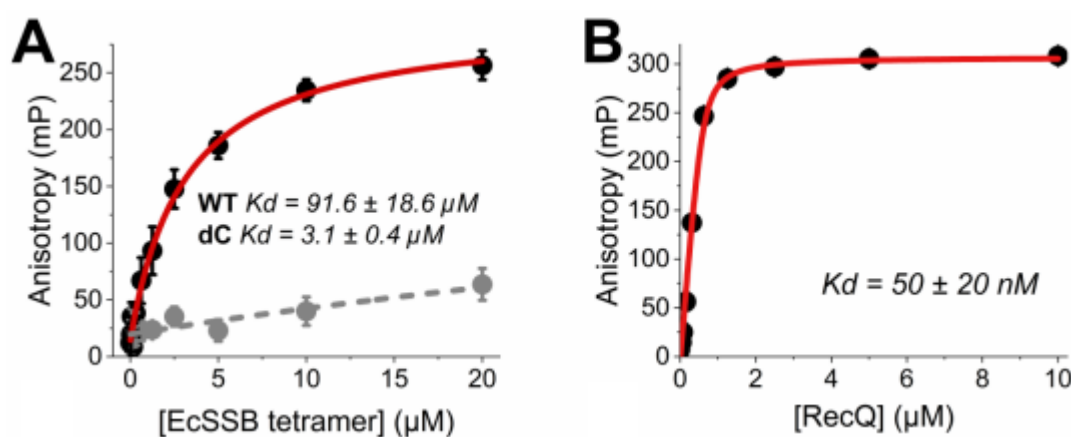


Fig. 9. Direct titrations during fluorescent polarization assays (A) Direct titration of dC (black) and WT (grey) (B) and recQ against flCTP. Fittings are in red and grey. Because of intra- and intermolecular OB domain – CTP interactions the interaction between the WT protein and the flCTP is weak.

Turbidity measurements

LLPS droplets scatter light allowing their quantification by measuring the OD₆₀₀ values of the samples. Dilution series from the EcSSB variants were generated starting from a tetramer concentration of 20 μM using the same buffer as in case of FP measurements. Turbidity measurements were performed on a TECAN Infinite F Nano+ machine, measuring absorbance at 600 nm (band width 10 nm) with a 25 flash/read setting. The whole reaction was repeated with the same buffer containing 3% polyethylene glycol (PEG 20K) as molecular crowder. All measurements were carried out in triplicates.

Microscopy

For the precise analyses of LLPS droplets, epifluorescence microscopy images were taken using a Nikon Eclipse Ti microscope. Here the EcSSB variants were mixed with 0.3 μM fluorescently labelled WT protein. Since the labelled WT molecule (it is true also for the labelled DNA and CTP) can be enriched in the forming protein droplets, they allow the fluorescent detection of condensates. The variants were applied in different concentrations in the same buffers (with and without 3% PEG 20K) as in the previous measurements. Images were taken using 488 nm light for excitation with a 100X objective.

Results

Expression and purification of EcSSB variants

Following the expression and purification process of all EcSSB constructs, the purity of the samples was analyzed using SDS PAGE electrophoresis (**Fig. 10.**). From this, we concluded that apart from an unexpected second band running separately from the EPE band -probably caused by ???? – the bands ran in their expected positions and their signal was very strong, proving that the purification process was successful.

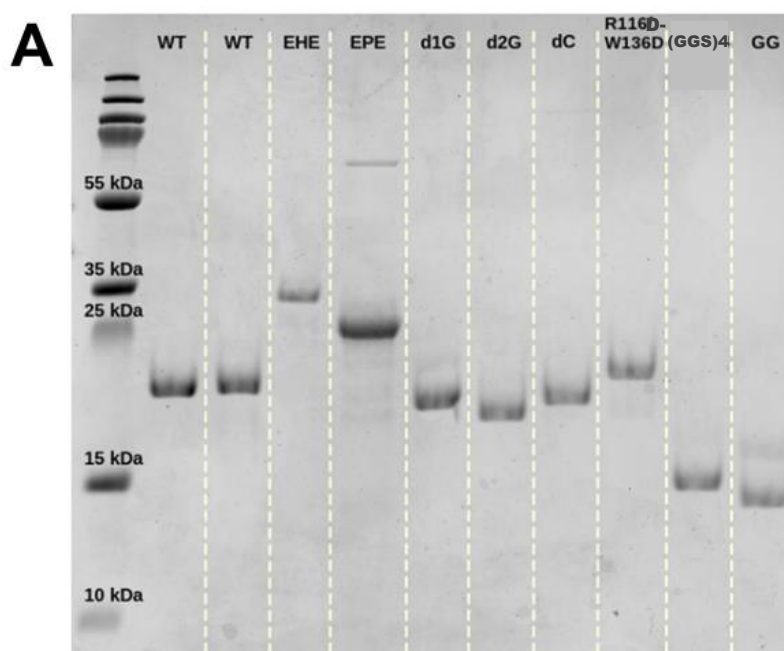


Fig. 10. SDS PAGE analysis of the purified EcSSB constructs. The first wild type protein was produced as in previous works [7] and presented as a control, while in case of the second WT sample the expression and purification shown in the method section was carried out.

Studying the ssDNA binding of SSB variants

To accurately describe the interactions between the different EcSSB constructs fluorescence polarization measurements were carried out with two types of fluorescently labeled ssDNAs (ssDNA36 and dT79). Even though the ssDNA36 and the dT79 differ in length allowing various binding modes, no significant differences were found between the variants and the WT protein. (**Fig. 11.**)

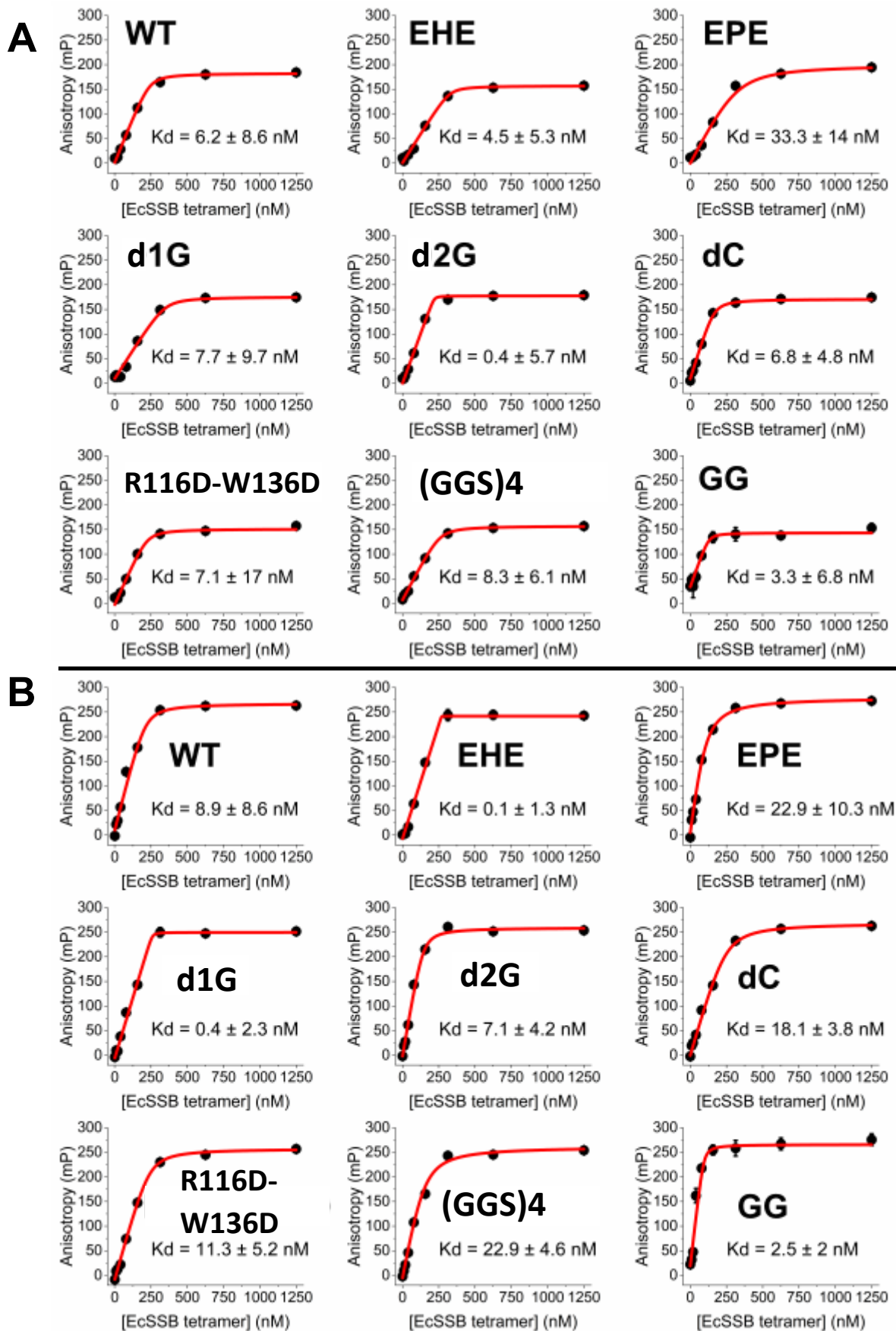


Fig. 11. Fluorescence polarization measurements studying the (A) ssDNA36 and (B) dT79 binding abilities of the different EcSSB variants. Error bars are inside the circle symbols most of the time

RecQ and OB domain binding abilities of variants

SSBs not only bind DNA, but they are important hub molecules promoting the assembly of other genome maintaining proteins by interacting with them. Thus, beside DNA binding, the bioengineered EcSSB variants should be able to bind their binding partners similarly to the WT protein since our goal is to lose only the LLPS propensity. We were interested in how changes in the IDL affect the protein binding behavior. For that, I used one of the main partner of EcSSB, namely the RecQ helicase. I found that the majority of the EcSSB variants bind the RecQ similarly to the WT ($K_d = 4,7 \pm 2,2$ nM), except the GG and the dC constructs. (**Fig. 12.**)

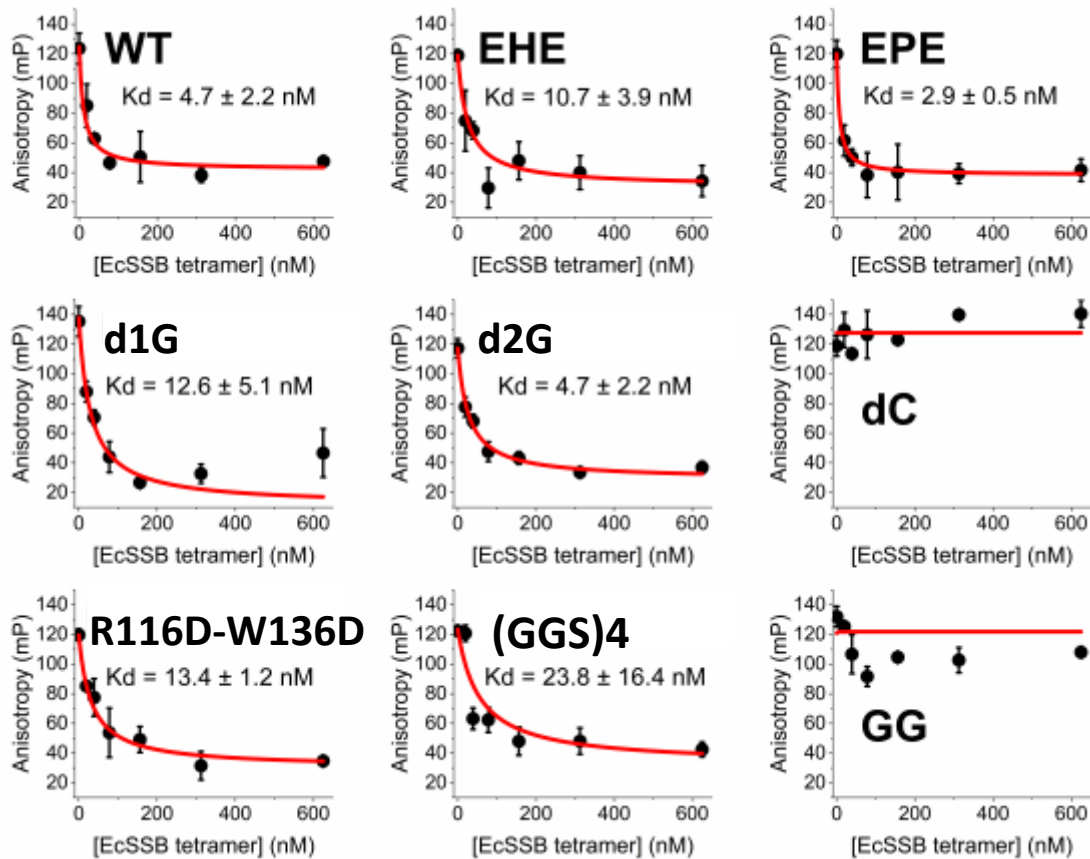


Fig. 12. Competitive titration of the RecQ – fICTP complex with the different EcSSB variants

I also studied the interactions between the dC construct and the other EcSSB variants. Since dC do not contain the CTP, this binding setup gave information about the intermolecular OB domain – CTP bindings. According to these measurements, there are 2 distinct populations of variants. The R116D-W136D and the EHE behave similarly to the WT protein. Interestingly the d1G, the d2G, the (GGs)4, the GG and the EPE constructs bind to the OB domain of dC with an even higher affinity than the WT SSB, presumably due to their favorable sterical and sequential properties beneficial for binding intermolecular OB domains. (**Fig. 13.**)

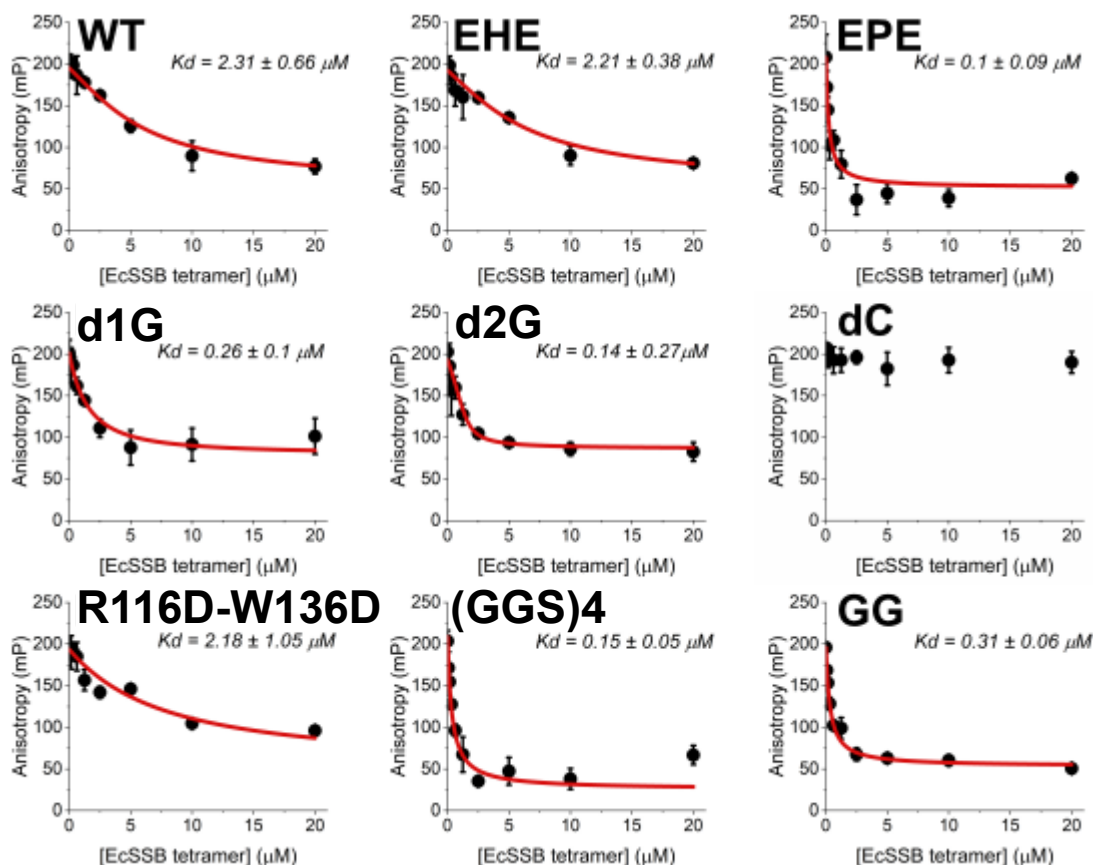


Fig. 13. Competitive titration of the dC - flCTP complex with the different EcSSB constructs

Turbidity measurements

A significant increase in absorbance was only measured in case of the WT SSB, while all other constructs showed little or no sign of that. However, with the addition of PEG as molecular crowder to the buffer (in order to better mimic the cellular environment), we observed an increase in turbidity not only for the WT protein, but mainly for the dC and the GG constructs, and to a lesser extent, also in case of the d1G and the d2G variants. The other mutants were not affected by the addition of PEG. (**Fig. 14.**)

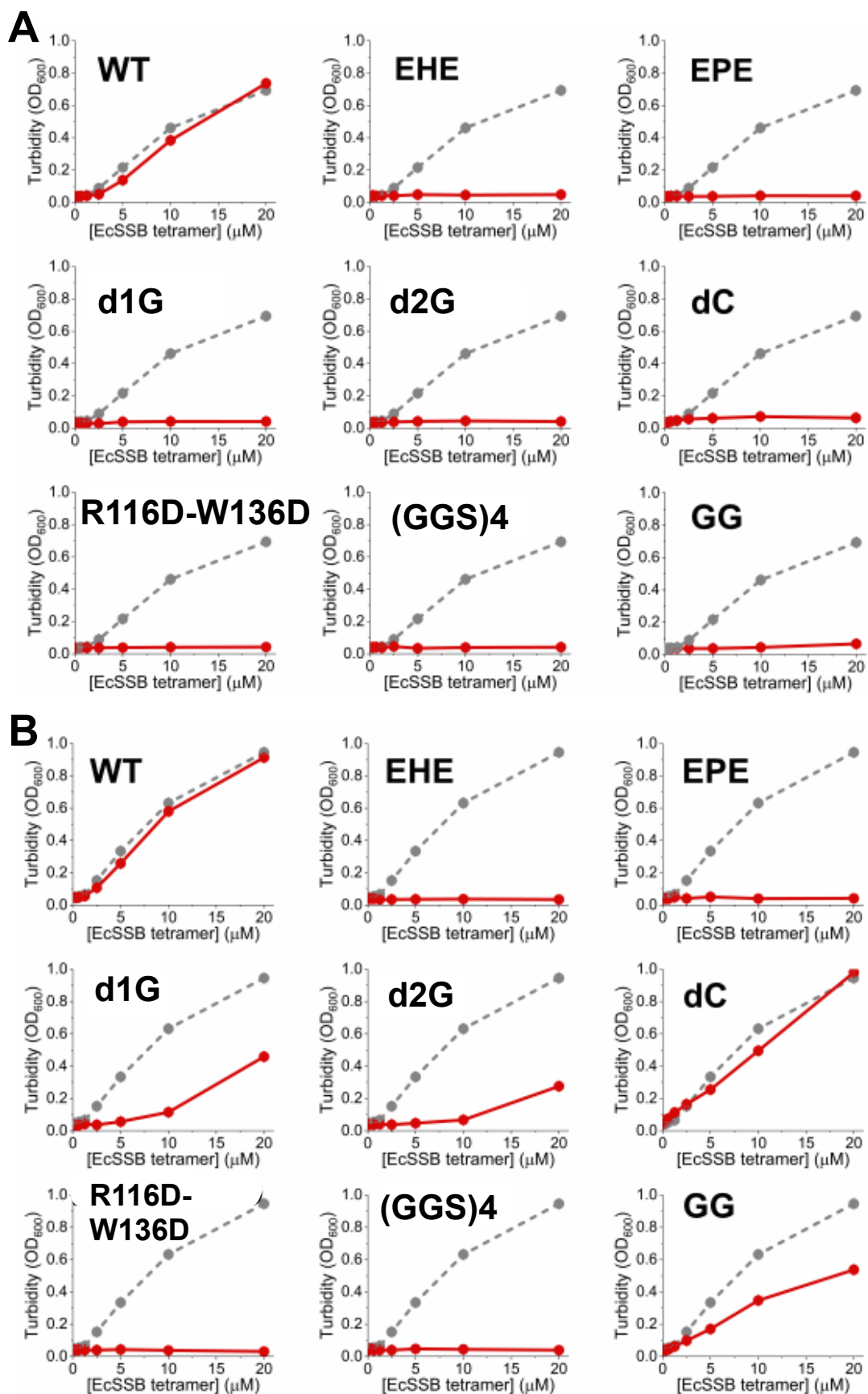


Fig. 14. Turbidity measurements of EcSSB WT and variants (A) without and (B) with the aid of molecular crowder PEG 20K. For easier comparison the absorbance curve of the WT protein is present in each case in grey. At the WT panel two measurements with the WT EcSSB is shown (red and grey)

Microscopic analyses of formed protein droplets

Since there was a possibility that the increase in turbidity values was due to aggregation rather than just LLPS, we also studied the variants using epifluorescence microscopy (20 μ M tetrames mixed with 0.3 μ M fWT protein). Here I found that the GG mutant formed only aggregates and was not able to phase separate under PEG, but all other SSBs that showed a change in absorbance formed phase separation droplets. (Fig. 15.)

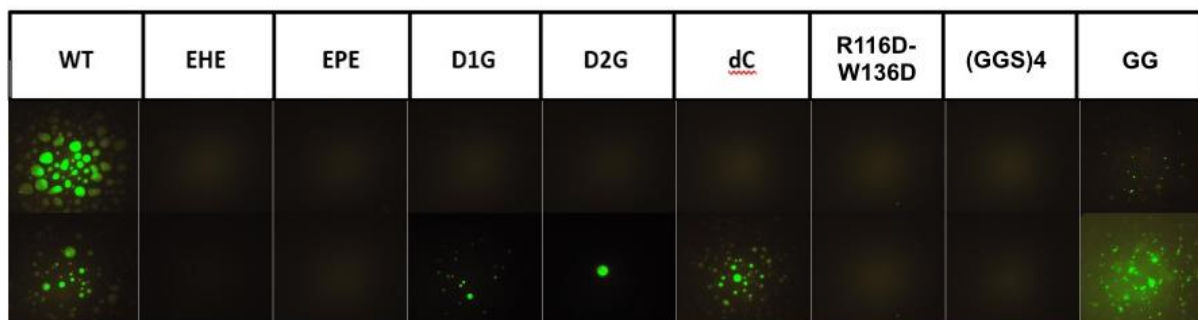


Fig. 15. Fluorescence microscopy images of *in vitro* forming LLPS condensates. In the first row, there is no polyethylene-glycol (PEG 20K) added to the buffer. The pictures in the second row were taken with the addition of 3% PEG 20K.

Discussion

LLPS as a phenomenon is known for a century but only recently has gained acceptance as an *in vivo* tool for compartmentalization. Moreover, several neurodegenerative diseases like Alzheimer' disease, Parkinson' disease, or amyotrophic lateral sclerosis (ALS) is now connected to dysfunctional LLPS [16]. By understanding the mechanism and importance of protein droplet formation *in vivo*, we can design biomolecules targeting condensations. These drugs could be utilized against these diseases or if LLPS turns out to be important for microbial survival, they can be used as antimicrobial agents.

My aim is to understand the functions of LLPS caused by EcSSB to ascertain if it is an excellent target for anti-LLPS drugs. For this, I bioengineered the protein to find constructs which retain the functions of the WT protein like DNA and partner binding but lacks its LLPS propensity. With the help of these variants the functions and adaptive advantages of LLPS in *E.coli* can be mapped and studied. The IDL variants which I produced here also helped to better understand the mechanisms of the LLPS of EcSSB. My recent findings show that the OB domain is specifically and exclusively responsible for oligonucleotide binding, partially as expected, as these interactions were unaffected in all constructs in which only the IDL and/or the CTP was modified. By looking at the RecQ helicase binding ability of the designed constructs, it turned out that apart from the dC and GG variants, all our constructs associated with the RecQ with similar affinities compared to the WT protein. This shows that only the availability of the CTP is essential for RecQ binding and as GG lacks the whole IDL, steric inhibition might appear. Any other change in the IDL did not affect this interaction. Note here that in search for

the best non LLPS propensive EcSSB variant, it would be highly advisable to expand this binding study with other important EcSSB partners, as various proteins could use the IDL differently.

Since LLPS forming ability was suggested to be connected to the intermolecular OB domain – CTP interactions, we assumed some coherence between the abilities of droplet formation and intermolecular OB domain binding. However, I found that EcSSB constructs most notoriously lacking LLPS propensity (GG, EPE, (GGS)4) had the highest intermolecular OB domain – CTP affinity, which also contradicts the assumption that only the OB domain – CTP interaction drives LLPS formation. In case of EPE, the highly charged sequence of the IDL could be the driver of stronger intermolecular OB domain binding ability, while in case of GG and (GGS)4 the shortened IDL might sterically promote that. This can also be the case with d1G and d2G constructs which also showed higher affinity to intermolecular OB domain binding compared to the WT protein. These results show that this interaction cannot be narrowed down only to the presence of the CTP like in case of RecQ.

	ssDNA ₃₆		dT ₇₉		RecQ		Intermolecular OB	
	K _d ± SD (nM)	p-value (N=3)	K _d ± SD (nM)	p-value (N=3)	K _d ± SD (nM)	p-value (N=3)	K _d ± SD (nM)	p-value (N=3)
WT	13 ± 7		25 ± 10		4.9 ± 1.4		2500 ± 600	
d1G	2.2 ± 2.5	0.57	5.5 ± 5.1	0.41	9.1 ± 4.4	0.16	220 ± 90	0.04
d2G	3.7 ± 4.5	0.21	10 ± 5	0.43	11 ± 3	0.33	110 ± 80	0.04
(GGS)4	11 ± 6	0.77	14 ± 7	0.70	17 ± 4	0.17	19 ± 16	0.002
GG	8.2 ± 14	0.44	5.4 ± 3.6	0.23	-	-	97 ± 26	0.004
R116D-W136D	10 ± 8	0.91	8.8 ± 5.7	0.50	17 ± 2	0.60	2500 ± 400	0.72
EPE	4.6 ± 3.8	0.43	21 ± 5	0.44	3.0 ± 0.6	0.25	57 ± 35	0.008
EHE	1.8 ± 2.6	0.22	1.6 ± 2.3	0.11	8.1 ± 3.9	0.2	2300 ± 400	0.77

Table 1. Interaction of EcSSB variants with ssDNA, RecQ helicase and intermolecular OB

domains. Dissociation constants (K_d) were determined in FP assays. K_d values refer to the interactions of EcSSB variants with ssDNA (ssDNA₃₆ or dT₇₉), RecQ helicase or intermolecular OB domains. Values reported are mean ± SD of the best-fit parameters of three independent measurements. T-probes were carried out for values of WT versus each variant. Note that K_d values below 10 nM convey significant fitting error due to applied fluorescent reporter concentrations of 100 nM (ssDNA oligonucleotides) and 25 nM (fICTP).

Strikingly, no variant was able to form LLPS condensates without the presence of a molecular crowder, which shows how fine-tuned this system is. Even with the presence of PEG, only dC showed very strong, while d1G and d2G moderate LLPS forming abilities. Also, I found that EcSSB without IDL (GG) is less stable compared to the WT protein and forms only aggregates when the concentration is high or a

crowder is present. These results are the most crucial in creating a new hypothesis about EcSSB LLPS propensity, as the so far assumed to be very important OB domain – CTP interactions between different tetramers are not specific requirements for LLPS in case of EcSSB. The CTP can only act as an initiator or enhancer for IDL – IDL interactions. Molecular crowders might have similar effects. Moreover, it is proven that the amount of glycines, and more importantly the presence of arginines and tryptophans in the IDL are significant for LLPS in case of EcSSB. It is important to note here that in the R116D-W136D mutant, the residues were exchanged to two negatively charged amino acids and the presence of these charges might also cause electrostatic repulsion between the IDLs. To better understand if the absence of arginine and tryptophan or the electrostatic repulsion causes the LLPS ability to disappear in case of the double mutant, not charged alanines can be introduced instead of aspartic acid. Since minor changes in the IDL caused loss of function, it is not surprising that the presence of IDLs from different species (EPE and EHE) do not allow the formation of EcSSB droplets.

It is essential to continue our research, as many parameters and factors which are uncertain are provided by the intracellular environment. Using the results showed in this work I found that R116D-W136D, D2G and EHE are very adequate candidates to use during *in vivo* studies. My next step is to introduce the sequences of these variants into the genome of *E.coli* with the deletion of the WT gene. Strains expressing only these non - LLPS propensive proteins can be used in survival assays or their GFP fused variants can be followed *in vivo* using fluorescence microscopy.

Author's contributions

- I cloned, expressed and purified the EcSSB constructs.
- I carried out fluorescence polarization and turbidity measurements.
- I used epifluorescence microscopy to study protein droplets.

Literature

<https://www.semanticscholar.org/paper/Ueber-die-basischen-Stoffe-des-Zellkerns.-Kossel/710f23965f2718f1e43817ac005412a6659c30de>

1. Nowicka, B. and J. Kruk, *Powered by light: Phototrophy and photosynthesis in prokaryotes and its evolution*. Microbiol Res, 2016. **186-187**: p. 99-118.
2. Shin, Y. and C.P. Brangwynne, *Liquid phase condensation in cell physiology and disease*. Science, 2017. **357**(6357).
3. Banani, S.F., et al., *Biomolecular condensates: organizers of cellular biochemistry*. Nat Rev Mol Cell Biol, 2017. **18**(5): p. 285-298.
4. Zhang, H., et al., *Liquid-liquid phase separation in biology: mechanisms, physiological functions and human diseases*. Sci China Life Sci, 2020. **63**(7): p. 953-985.
5. Peng, P.H., K.W. Hsu, and K.J. Wu, *Liquid-liquid phase separation (LLPS) in cellular physiology and tumor biology*. Am J Cancer Res, 2021. **11**(8): p. 3766-3776.
6. Spann, S., et al., *Biomolecular condensates in neurodegeneration and cancer*. Traffic, 2019. **20**(12): p. 890-911.
7. Harami, G.M., et al., *Phase separation by ssDNA binding protein controlled via protein-protein and protein-DNA interactions*. Proc Natl Acad Sci U S A, 2020. **117**(42): p. 26206-26217.
8. Tan, H.Y., et al., *The intrinsically disordered linker of E. coli SSB is critical for the release from single-stranded DNA*. Protein Sci, 2017. **26**(4): p. 700-717.
9. Kar, M., et al., *Glycine-Rich Peptides from FUS Have an Intrinsic Ability to Self-Assemble into Fibers and Networked Fibrils*. Biochemistry, 2021. **60**(43): p. 3213-3222.
10. Chong, P.A., R.M. Vernon, and J.D. Forman-Kay, *RGG/RG Motif Regions in RNA Binding and Phase Separation*. J Mol Biol, 2018. **430**(23): p. 4650-4665.
11. Murthy, A.C., et al., *Molecular interactions contributing to FUS SYGQ LC-RGG phase separation and co-partitioning with RNA polymerase II heptads*. Nat Struct Mol Biol, 2021. **28**(11): p. 923-935.
12. Kozlov, A.G., et al., *How Glutamate Promotes Liquid-liquid Phase Separation and DNA Binding Cooperativity of E. coli SSB Protein*. J Mol Biol, 2022. **434**(9): p. 167562.

13. Das, S., et al., *Comparative roles of charge, pi, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins*. Proc Natl Acad Sci U S A, 2020. **117**(46): p. 28795-28805.
14. Shinn, M.K., et al., *Are the intrinsically disordered linkers involved in SSB binding to accessory proteins?* Nucleic Acids Res, 2019. **47**(16): p. 8581-8594.
15. Shereda, R.D., et al., *SSB as an organizer/mobilizer of genome maintenance complexes*. Crit Rev Biochem Mol Biol, 2008. **43**(5): p. 289-318.
16. Wang, B., et al., *Liquid-liquid phase separation in human health and diseases*. Signal Transduct Target Ther, 2021. **6**(1): p. 290.

Supplementary

	WT	GG	(GGS)4	EHE	EPE	DeltaG	Delta2G	R116D-W136D
forward oligo	BamHI I- EcSSB B f: ctaggg atccatg gccagc agaggc gtaaac aagg	BamHI EcSSB f	BamHI EcSSB f	BamHI EcSSB f	BamHI EcSSB f	BamHI EcSSB f	BamHI EcSSB f	R116D quick f: ccatgcag atgctggg tggt W136D quick f: ggtcagc cgcaggg cgggtgac ggtcagc ctcagcag ccgc gatcagg gtggtggc gctccgg cagg,

revers e oligo	EcSS B- STOP -XhoI r: ctagctc gagtta gaacgg aatgtca tcate	GG- CTP- STOP- XhoI r: cagctcga gttagaac ggaatgtc atcatcaa agtccata ccacca gcatctgc atg	GGs4X -CTP mega r: cagctcga gttagaac ggaatgtc atcatcaa agtccata ctacctcc actacctc cact	EHE mega r: ctcgagtt agaacgg aatgtcat catcaaag tccatgcc gttactaa cagggttg ctggaag ggccagg cgggcct gcaggtgt gtggttg gctggctt cgagtgat tcgggtgc tgggtgg gtgggag ggagtatg aggggga tgtgggcc accaccg ggacctg gtggggc actcagtc cattcccat tctggttct cagaggc tgagag gcagcag agggtcc agtggtag gctggga agctgaa gggttgct	EPE mega r: ggcaccatg cagatgctg ctggatgac aaacgtaact tcaatcagc gcaacaaca gtaacaacat caactccga aaaccagca acatatcaac aacgaacac atcaacaac aacaacatc aacaacggt aacgattca tgccgctga acagcaacg ataaaatcat cgaagacaa agaattcac ggatcgctt ggatgacaa caatgaaga aaacaatttc cagagcaat tctgaaacct ttgataaaca ggaaggcat ctacgataaa atgaatgtcc aaatggactt tgatgatgac attccgttcta actcgag	DeltaG mega r: ggcaccatg cagatgctg gctccggca ggtggcaat atcgggtgt ggcagccg cagggcggt tggggtcag cctcagcag ccgcagggt ggcaatcag ttcagcggc ggcgcgca gtctcgccc gcagcagtc cgctccggc agcgccgtc taacgagcc gccgatgga ctttgatgat gacattccgt tctaactcga g	DeltaG2 G mega r: ggcaccatg cagatgctg cagccgcag ggcggttgg ggcagcct cagcagccg cagggtggc aatcagttca gcggcggc gcgcagtct cgcccgag cagtcgct ccggcagc gccgtctaa cgagccgc gatggacttt gatgatgac attccgttcta actcgag	R116D quick r: cctgccg gagcgcc accaccct gatcacca cccagcat ctgcatgg W136D quick r: gcggctg ctgaggct gaccgtca ccgccctg cggctga cc
-------------------	--	--	---	--	---	--	--	--

				gtcgttctg				
				caccgcct				
				tggtgggt				
				gcctgctg				
				gggtgctg				
				actctggg				
				tttggtca				
				ctgaagtt				
				aggaacct				
				cagaataa				
				accataca				
				gaacagc				
				atctgcat				
				gggtgcc				

Table 2: the primers used to create our modified EcSSB sequences

$$I = I_{\min} + (I_{\max} - I_{\min}) * (-\sqrt{\frac{((\frac{To * KdB + Bo * KdT}{KdB - KdT}) - (Bo + Do + KdB))^2 - 3 * (Bo * (Do - (KdT * (Bo + Do + KdB) + To * KdB) / (KdB - KdT)))^2}{9}} * \cos(\arccos(\frac{9 * ((\frac{To * KdB + Bo * KdT}{KdB - KdT}) - (Bo + Do + KdB)) * (Bo * (Do - (KdT * (Bo + Do + KdB) + To * KdB) / (KdB - KdT))) - 2 * ((\frac{To * KdB + Bo * KdT}{KdB - KdT}) - (Bo + Do + KdB))^3 - 27 * (Bo^2 * Do * KdT / (KdB - KdT)) / (2 * \sqrt{-(3 * (Bo * (Do - (KdT * (Bo + Do + KdB) + To * KdB) / (KdB - KdT))) - ((\frac{To * KdB + Bo * KdT}{KdB - KdT}) - (Bo + Do + KdB))^2)^3})}{3}) + \sqrt{3}) * \sin(\arccos(\frac{9 * ((\frac{To * KdB + Bo * KdT}{KdB - KdT}) - (Bo + Do + KdB)) * (Bo * (Do - (KdT * (Bo + Do + KdB) + To * KdB) / (KdB - KdT))) - 2 * ((\frac{To * KdB + Bo * KdT}{KdB - KdT}) - (Bo + Do + KdB))^3 - 27 * (Bo^2 * Do * KdT / (KdB - KdT)) / (2 * \sqrt{-(3 * (Bo * (Do - (KdT * (Bo + Do + KdB) + To * KdB) / (KdB - KdT))) - ((\frac{To * KdB + Bo * KdT}{KdB - KdT}) - (Bo + Do + KdB))^2)^3})}{3}) - ((\frac{To * KdB + Bo * KdT}{KdB - KdT}) - (Bo + Do + KdB)) / 3) / Bo$$

Competitive binding equation generated by Dr László Radnai