# Computational Design and Optimization of G-CSFR Modulators

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Timo Ullrich

aus Schorndorf

Tübingen

2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

| | |
|---|---|
| Tag der mündlichen Qualifikation: | 01.02.2024 |
| Dekan: | Prof. Dr. Thilo Stehle |
| 1. Berichterstatter/-in: | Prof. Dr. Andrei Lupas |
| 2. Berichterstatter/-in: | Prof. Dr. Boris Maček |

## Acknowledgments

Before I acknowledge specific individuals, I would like to express my gratitude to several groups and organizations that played important roles during my PhD journey. I want to thank all the members of Patrick Müller's lab, previously from the FML (now University of Konstanz), Department 1 at the MPI, and the Skokowa/Welte lab at the UKT for the great atmosphere and the help they offered me. Of course, I also want to express my gratitude to all the people behind the scenes who were responsible for running the organization and maintaining the institute. Finally, I want to thank the IMPRS for their financial support and the RST for their invaluable support throughout every phase of my PhD.

I would like to express my gratitude to several individuals who have played crucial roles during my PhD journey. In particular, I want to thank the following individuals for their supervision and support.

Mohammad ElGamacy, who dedicated considerable time to providing scientific guidance, teaching me computational and wet-lab techniques, and guiding me through my PhD. Julia Skokowa, who offered valuable scientific input during frequent focus meetings, was part of my Thesis Advisory Committee (TAC), supported my projects, introduced me to her lab and its members, and provided me with the opportunity to be part of her research group. Boris Maček, my official university supervisor, and Heike Brötz-Oesterhelt, both of whom were part of my TAC. Patrick Müller, who convinced me during the IMPRS interviews to pursue my PhD on topics related to his lab's work, particularly in the design and optimization of novel therapeutic agents. I also appreciate his scientific input during my time at the FML. Andrei Lupas, for providing me with the opportunity to work in his department and participate in excellent scientific gatherings, especially in Alpbach. I'm particularly grateful for Andrei's immediate assistance during the transition from the FML to department 1 and for being part of my TAC. I'd also like to express my appreciation to Kateryna Maksymenko, Matej Milijas-Jotic, and Valeriia Hatskovska for their work-related assistance and for being wonderful companions throughout this journey.

Finally, I want to express my gratitude to my partner, family, and friends for the great time and support outside of the workplace.

## Table of Contents

## Abbreviations

| | |
|---|---|
| 2D | Two-dimensional |
| ADA | Anti-drug antibody |
| AKT | Protein kinase B |
| AML | Acute myeloid leukemia |
| APC | Antigen-presenting cell |
| B4_st2 | Boskar4 short tandem of 2 domains |
| BAM | Beta-barrel assembly machinery |
| boa | Bivalent-orientation antiparallel |
| bop | Bivalent-orientation parallel |
| bv | Boskar4 variant |
| CD | Circular dichroism |
| CFU | Colony-forming unit |
| CNTF | Cytokine ciliary neurotrophic factor |
| CRH | Cytokine receptor homologous |
| DARPin | Ankyrin repeat protein |
| dpf | days post-fertilization |
| *E. coli* | *Escherichia coli* |
| $EC_{50}$ | Half-maximal effective concentration |
| EHEC | Enterohemorrhagic *Escherichia coli* O157:H7 |
| EPO | Erythropoietin |
| EPOR | Erythropoietin receptor |
| ERK | Extracellular signal-regulated kinase |
| FACS | Fluorescent activated cell sorting |
| FNIII | Fibronectin type III |
| G-CSF | Granulocyte colony-stimulating factor |
| G-CSFR | Granulocyte colony-stimulating factor receptor |
| GI | Gastrointestinal |
| GO | Gene Ontology |
| H1/2/3/4 | Helix 1/2/3/4 |
| HSC | Hematopoietic stem cell |
| HSPCs | Hematopoietic stem and progenitor cells |
| $IC_{50}$ | Half-maximal inhibitory concentration |
| Ig | Immunoglobulin |
| IL-10Rβ | Interleukin-10 receptor beta |
| IL-22 | Interleukin-22 |
| IL-22Rα | Interleukin-22 receptor alpha |
| IL-2Rβ | Interleukin-2 receptor β |
| IL-3 | Interleukin-3 |
| IL-3Rα | Interleukin-3 receptor α |
| IL-6 | Interleukin-6 |
| IM | Inner membrane |
| IMAC | Immobilized metal affinity chromatography |
| inter-TMD | Inter-transmembrane domain |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |

| | |
|---|---|
| JAK | Janus kinase |
| $k_a$ | Association rate |
| $k_d$ | Dissociation rate |
| $K_D$ | Equilibrium dissociation constant |
| LB | Lysogeny broth |
| LIFR | Leukemia inhibitory factor receptor |
| MAPK | Mitogen-activated protein kinase |
| MDSC | Myeloid-derived suppressor cells |
| MHCI | Major histocompatibility complexes class I |
| MHCII | Major histocompatibility complexes class II |
| nanoDSF | Nanoscale-differential scanning fluorimetry |
| novokine | Cytokine with novel function |
| $OD_{600}$ | Optical density at 600 nm |
| OM | Outer membrane |
| OMP | Outer membrane protein |
| ori | Orientation rigging design |
| PBS | Phosphate-buffered saline |
| PCA | Principal component analysis |
| PDB | Protein Data Bank |
| PEG | Polyethylene glycol |
| Pegfilgrastim | Pegylated granulocyte colony-stimulating factor |
| PI3K | Phosphoinositide 3-kinase |
| pSTAT3 | Phosphorylated signal transducer and activator of transcription 3 |
| pSTAT5 | Phosphorylated signal transducer and activator of transcription 5 |
| rhG-CSF | Recombinant human granulocyte colony-stimulating factor |
| RMSD | Root mean square deviation |
| RT | Room temperature |
| SCF | Stem cell factor |
| SCN | Severe congenital neutropenia |
| sd | standard deviation |
| SDS-PAGE | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SEC | Size exclusion chromatography |
| SOCS | Suppressor of cytokine signaling |
| SPR | Surface plasmon resonance |
| STAT | Signal transducer and activator of transcription |
| synthekine | Synthetic cytokine |
| VHH | Variable domain of heavy chain of heavy chain-only antibodies |
| γc | Common γ-chain |

## Zusammenfassung

In Anbetracht der essentiellen Rolle, die Proteine bei der Aufrechterhaltung lebenswichtiger Prozesse spielen, ist es nicht überraschend, dass Proteine ebenso mit einem breiten Spektrum von Krankheiten assoziiert sind. Umgekehrt können Proteine entsprechend auch als Therapeutika verwendet werden. So haben sich beispielsweise sekretierte Wachstumsfaktoren, die so genannten Zytokine, als vielversprechende Kandidaten für proteinbasierte Therapeutika erwiesen, vor allem aufgrund ihrer stark immunmodulatorischen Eigenschaften. Ein wesentlicher Bestandteil des angeborenen Immunsystems sind Neutrophile, deren Differenzierung durch den hämatopoetischen Teilprozess, die Granulopoese, reguliert wird. Dieser Prozess erfordert ein komplexes Zusammenspiel verschiedener Zytokine und ihrer entsprechenden Rezeptoren. Ein entscheidender Akteur in diesem Zusammenspiel ist der Granulozyten-Kolonie-stimulierende Faktor Rezeptor (engl. Granulocyte-Colony Stimulating Factor Receptor, G-CSFR), der durch seinen nativen Liganden, G-CSF, aktiviert wird. Im Gegensatz zu den meisten anderen Zytokinen wird G-CSF in seiner nativen Form aufgrund seines Sicherheitsprofils und seiner Fähigkeit, die Zahl der Neutrophilen im Blut zu erhöhen, klinisch eingesetzt.

Die Anwendung von G-CSF ist jedoch durch seine Stabilität, seine Produktionskosten und seine native Aktivität auf G-CSFR begrenzt. Darüber hinaus gibt es nur sehr beschränkte Kenntnisse über nicht-native G-CSFR-Modulatoren. Diese könnten nicht nur für das Verständnis von G-CSFR-bedingten Krankheiten, sondern auch für die Entwicklung innovativer therapeutischer Anwendungen entscheidend sein. Um das ungenutzte Potential des klinisch bedeutsamen G-CSFR über seine native Aktivität hinaus zu erschließen, haben wir uns zum Ziel gesetzt, mit Hilfe von Proteindesigntechniken neuartige Liganden zu entwickeln, die die Aktivität des G-CSFR modulieren können. Neben der Anpassung der Rezeptoraktivität sollen diese Liganden im Vergleich zu ihrem nativen Gegenstück G-CSF auch eine höhere Stabilität und eine effizientere Produktion aufweisen.

Zu diesem Zweck habe ich ein kürzlich entwickeltes hyperthermostabiles *de novo* designtes G-CSFR-Bindungsmodul genutzt und es mit *in silico* und *in vitro* Hochdurchsatzmethoden optimiert, um ein breites Spektrum von Varianten mit erhöhter Bindungsaffinität zu erhalten. Ich zeige, dass diese verbesserten Bindungsmodule zur Herstellung von G-CSFR-Agonisten verwendet werden können,

die in zellbasierten Assays G-CSF-Aktivität erzielen, und dass sie auch zur Herstellung von Liganden verwendet werden können, die die G-CSFR-Aktivität durch Abstimmung der Rezeptorgeometrie modulieren können. Diese Liganden zeichnen sich durch eine fein abgestimmte intrazelluläre Signalübertragung, transkriptomische Aktivität und primäre Stammzelldifferenzierung aus und zeigen *in vivo* Aktivität in Zebrafisch- und Mausmodellen. Neben der Entwicklung von Agonisten zeige ich, dass diese verbesserten Bindungsmodule auch zur Herstellung kompetitiver G-CSFR-Antagonisten mit nanomolarer Hemmwirkung verwendet werden können.

Soweit ich weiß, ist diese Arbeit der erste Nachweis, dass die Aktivität des G-CSFR durch designte Liganden, die nicht-native Rezeptor-Geometrien induzieren, modelliert werden kann. Darüber hinaus werden in dieser Studie eine Reihe von Bindungsmodulen mit unterschiedlichen Affinitäten für G-CSFR zusammen mit den neu entwickelten Liganden vorgestellt, die die Grundlage für eine systematische Untersuchung der Modulation der G-CSFR-Aktivität bilden. Diese Ergebnisse sind vielversprechend, um die Entwicklung innovativer Proteintherapeutika voranzutreiben.

## Abstract

Considering the indispensable role played by proteins in maintaining vital life processes, it is not surprising that proteins are linked to a broad spectrum of diseases. Conversely, proteins can be leveraged for effective therapeutic interventions. For example, secreted growth factors known as cytokines have emerged as promising candidates for protein-based therapeutics, primarily due to their potent immunomodulatory properties. An integral element of the innate immune system consists of neutrophils, the maturation of which is coordinated by a hematopoietic subprocess known as granulopoiesis. This process requires a complex interplay among various cytokines and their corresponding receptor molecules. A critical player in these interactions is the granulocyte-colony stimulating factor receptor (G-CSFR), which is activated by its native ligand, G-CSF. Unlike most other cytokines, G-CSF has found clinical use in its native form due to its favorable safety profile and its ability to increase the number of neutrophils in the blood.

However, G-CSF application is restricted by its stability, production cost and native activity on G-CSFR. Additionally, there is very limited knowledge regarding non-native G-CSFR modulators, which could be key not only for understanding G-CSFR related diseases, but also for the development of innovative therapeutic applications. In order to unlock the untapped potential of the clinically significant G-CSFR beyond its native activity, my objective was to employ protein design techniques to craft novel ligands capable of modulating G-CSFR activity. In addition to customizing receptor activity, these designs offer enhanced stability and more efficient production compared to their native counterpart G-CSF.

To this end, I utilized a recently developed hyper-thermostable *de novo* designed G-CSFR binding module and optimized it with *in silico* and *in vitro* high-throughput methods to obtain a broad spectrum of variants with enhanced binding affinity. I demonstrate that these enhanced binding modules can be utilized to generate G-CSFR agonists that achieve G-CSF activity in cell-based assays and can also be used to create ligands capable of modulating G-CSFR activity by tuning receptor geometry. These ligands featured fine-tuned intracellular signaling, transcriptomic activity, and primary stem cell differentiation and exhibit *in vivo* activity in zebrafish and mouse models. In addition to designing agonists, I show that these enhanced binding modules

can be used to generate competitive G-CSFR antagonists with nanomolar inhibitory activity.

To the best of my knowledge this work is the first demonstration that G-CSFR activity can be tuned by the design of ligands inducing non-native receptor geometries. Additionally, this study presents an array of binding modules with diverse affinities to G-CSFR, along with the newly designed ligands, providing the foundational components for systematic investigation of G-CSFR activity modulation. These findings hold significant promise for advancing the development of innovative protein therapeutics.

## List of Publications

### *Publications not included as chapter into this dissertation*

Maksymenko K., Maurer A., Aghaallaei N., Barry C., Borbara´ n-Bravo N., **Ullrich T**., Dijkstra T. M.H., Hernandez Alvarez B., Muller P., Lupas A. N., Skokowa J., ElGamacy M. (2023). The design of functional proteins using tensorized energy calculations. Cell Reports Methods, 3(8), 100560. https://doi.org/10.1016/j.crmeth.2023.100560

### *Preprints and manuscripts*

**Ullrich T.**, Pollmann C., El-Riz M., Ritter M., Haaf J., Aghaallaei N., Tesakov I., Maksymenko K., Hatsovska V., Kandabarau S., Klimiankou M., Lengerke C., Welte K., Hernandez-Alvarez B., Müller P., Lupas A., Piehler J., Skokowa J., ElGamacy M. (2023). Tuning of granulopoietic signaling by *de novo* designed agonists. BioRxiv, https://doi.org/10.1101/2023.11.25.568662.

**Ullrich T.**, Klimenkova O., Hartmann M. D., Skokowa J., ElGamacy M. (2023). *De novo* design of G-CSFR-inhibiting proteins.

# 1 Introduction

Cytokines are protein-based secreted growth factors that activate their corresponding receptors, an essential element of various aspects of development, homeostasis, and immunity. Following their discovery starting in the nineteen fifties [1], expectations of their potential use as therapeutics rose continually over the next decades, culminating in the attempt to recombinantly produce a comprehensive array of cytokines. Disillusion came during clinical trials, where most of the tested cytokines were shown to have strong adverse effects [2]. Only a few candidates proved to be comparably safe and found high clinical usage. Among these candidates, the granulocyte colony-stimulating factor (G-CSF) stands out as a prominent immunotherapeutic agent that has been extensively utilized since its approval in the 1990s for the clinical use in cancer patients undergoing chemotherapy [3]. Its success in the field has resulted in a substantial global market value which is thought to exceed five billion US-dollars by 2030 [4]. Till today, G-CSF is used in its native form with only minor modifications.

In spite of the substantial clinical benefits associated with G-CSF, it faces inherent limitations, including low stability under neutral pH conditions, rapid systemic clearance, susceptibility to proteolytic degradation, and high production costs. Additionally, G-CSF function is limited to form a predetermined complex with its corresponding receptor (G-CSFR), which, in turn, restricts it to a particular pattern of native receptor activity. In my work, I address these challenges by pioneering the untapped potential of a new generation of G-CSFR modulators to enhance clinical utility through protein design, leveraging the significant progress in this field over the past three decades.

One of the possibilities that protein design offers is the creation of binding modules that specifically address their target while being smaller and more stable than their native counterpart. These conformationally stable and versatile building blocks can then be further optimized with high-throughput methods to serve as a basis for the design of novel ligands with tailored receptor activity. Furthermore, ligands designed for high stability possess the capacity for subsequent retroactive redesign, a capability limited in native ligands due to their inherent lower conformational stability.

An important area for further investigation lies in the development of ligands that are capable to induce non-native receptor activity. This approach is crucial for exploring a

broad spectrum of receptor activation patterns, thereby facilitating the development of innovative G-CSFR agonists.

Additionally, given the involvement of G-CSFR activity in a broad spectrum of diseases, ranging from leukemia to autoimmune disorders, clinical interest extends beyond activation to encompass inhibition. In strong contrast to the potentially extensive range of applications, the repertoire of available G-CSFR antagonists remains severely limited. Hence, I aim to broaden the spectrum of accessible G-CSFR inhibitors, facilitating the exploration of innovative therapeutic applications.

Therefore, my overarching goal was to leverage protein design techniques in order to design novel ligands with enhanced biophysical properties that can purposefully modulate the activity of the clinically relevant G-CSFR, aiming to fully unlock its untapped potential. In the subsequent paragraphs, I will present a concise outline of this dissertation, which addresses the aforementioned challenges.

**Chapter 2, Cytokines and protein design,** delves into the background of cytokines in general and in the context of protein design. Initially, an overview of cytokines and their receptors will be presented, with a particular emphasis on G-CSF. Subsequently, the significance of protein design in the realm of cytokine design will be underscored. Furthermore, I will present strategies related to the design of idealized binding modules, that were developed in our group, in conjunction with the considerations for establishing a computationally-driven optimization pipeline tailored to these designed binders.

**Chapter 3, Tuning of granulopoietic signaling by *de novo* designed agonists,** includes how I optimized hyper-thermostable G-CSFR binding modules that I subsequently used to design potent G-CSFR agonists. In addition, I present novel ligands that were intentionally designed to enable non-native ligand-receptor geometries and demonstrate G-CSFR tunability. These ligands were able to modulate intracellular signaling and primary stem cell differentiation and exert graded control over key target genes.

**Chapter 4, *De novo* design of G-CSFR-inhibiting proteins,** covers how I developed highly potent G-CSFR antagonists, utilizing protein design and high-throughput optimization methods. This constitutes the first case of antagonizing the clinically relevant G-CSFR utilizing protein-based, small, epitope-targeted, and hyper-

thermostable proteins. Figure 1 summarizes the core concept of my research and delineates the content addressed in Chapters 3 and Chapter 4.

**Chapter 5, Discussion,** elaborates on various aspects of modulation and inhibition of G-CSFR activity and the exceptional stability of the novel ligands, underscoring their potential as protein therapeutics.

**Chapter 6, Conclusion,** summarizes the key findings and insights acquired throughout my studies.

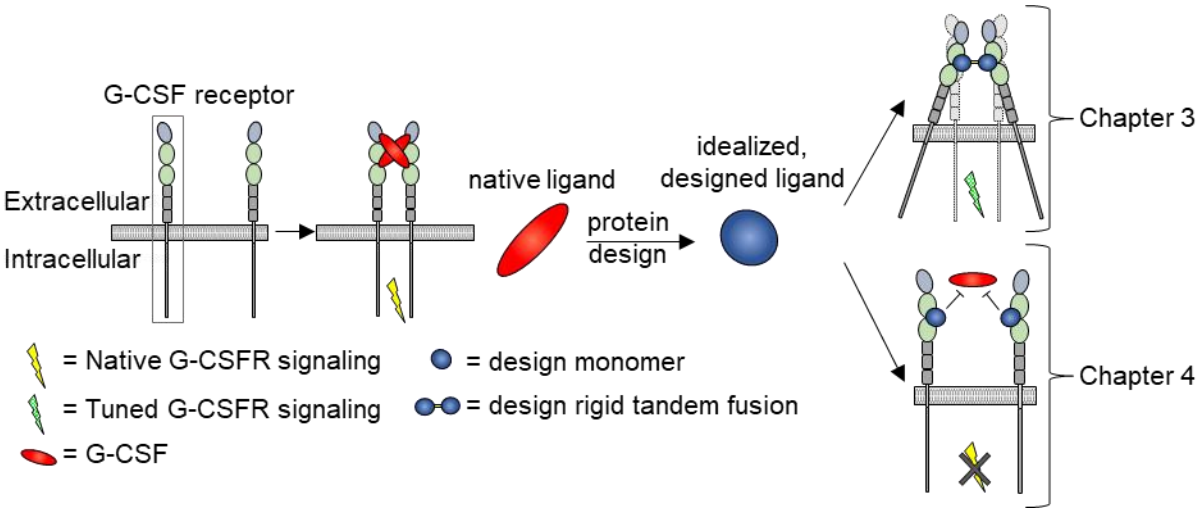



***Figure 1:*** *Schematic overview of the principles of G-CSFR activation and the respective mode of action of the designed ligands (details on the background can be found in Chapter 2). Chapter 3 exemplifies the development of activity tuning agonists while Chapter 4 illustrates the design of novel G-CSFR antagonists.*

# 2 Cytokines and protein design

## *2. 1 Cytokines and their receptors*

### *2.1.1 Essential players of development, homeostasis, and immunity*

Cytokines are immune cell secreted growth factors that play a crucial role in various aspects of development, homeostasis, and immunity by activating their corresponding receptors [5]. The term “cytokine” has its roots in the Greek words “cyto” and “kine”, which mean “cell” and “movement” [6], respectively, and was introduced by Stanley Cohen in 1974, referring to cellular substances that trigger chemotaxis and activate immune cells [7]. Cytokines exhibit diverse modes of action, serving as autocrine effectors by acting on the producing cells, paracrine effectors by influencing nearby cells, or endocrine effectors by exerting their effects on distant cells throughout the body [8]. Currently, approximately 200 cytokines have been identified, with many of them having their functions elucidated [9]. They exhibit cellular pleiotropy, the ability of one molecule to stimulate multiple distinct cell types and tissues [10].

In general, cytokines promote an antiviral, anti-tumor, or anti-parasitic state, facilitating cell migration and adhesion, regulating inflammation, stimulating cell division and differentiation, and influencing immune cell activity [8, 11]. For example, they coordinate the development and homeostasis of regulatory T cells [12] or regulate inflammatory processes [13]. Numerous cytokine drugs utilized clinically or undergoing investigation in clinical trials are based on molecules that modulate both innate and adaptive immunity [6].

However, cytokines are just one side of the story; they have corresponding receptors that act like decoders, converting extracellular signals into intracellular signals. Specific cytokines bind to distinct epitopes on cytokine receptors' ectodomains, forming signaling complexes [14]. Hence, the functional diversity of cytokines is also mirrored in the diversity of cytokine receptor complexes, which can demonstrate varying degrees of complexity. For instance, some cytokines, such as G-CSF, are thought to initiate relatively straightforward homo-dimeric receptor complexes [15]. On the other hand, cytokines like interleukin-3 (IL-3) give rise to more intricate hetero-octamer receptor complexes [16].

### *2.1.2 Regulation of granulopoiesis*

Granulopoiesis is part of the hematopoiesis and describes the production of granulocytes in the bone marrow [17]. Granulocytes play a vital role in the immune system and harbor specific granules within their cytoplasm [18]. These granules are secretory vesicles containing a mixture of cytotoxic molecules, such as the protease neutrophil elastase. This protease, for instance, enhances the clearance of gram-negative bacteria [19]. Different types of granulocytes are described, namely basophils, eosinophils, and neutrophils [20].

Neutrophils are the most abundant white blood cells in mammals. Thus, they play an essential role for the innate immune system [21]. Neutrophils are also recognized for their impact on the adaptive immune system, acting as antigen-presenting cell (APC) [22, 23]. Typically, they derive from hematopoietic stem cells (HSCs) through the interplay of intracellular and extracellular factors. Intracellularly, transcription factors like C/EBP-α regulate the precise timing of gene activation specific to their lineage [24]. Extracellular stimuli that influence granulopoiesis are predominantly mediated by growth factors, with G-CSF playing a central role as the primary driver of granulopoiesis across various mammalian organisms [25, 26]. Figure 2 illustrates the development of HSCs to mature neutrophils in the homeostatic granulopoiesis.

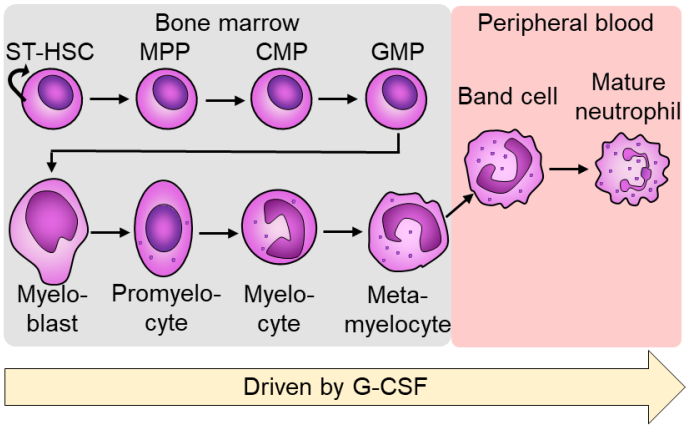



***Figure 2**: Development of neutrophils from short-term hematopoietic stem cells (ST-HSCs). MPP, multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte-macrophage progenitor. Figure adopted from [27].*

Another type of granulopoiesis is emergency granulopoiesis, predominantly occurring during pathogenic infections. Pathogens are detected by hematopoietic and resident cells through Toll-like receptor signaling [28], initiating emergency granulopoiesis in HSCs and neutrophil progenitors [29]. The substantial release of G-CSF with around 800 ng/L during acute bacterial infections [30] emphasizes its important role as a major driving force behind emergency granulopoiesis.

### *2.1.3 G-CSF and its receptor*

#### *2.1.3.1 Protein complex structure*

G-CSF, a class 1 cytokine is a four helical bundle with an “up-up-down-down” architecture (Figure 3 A). Its corresponding receptor (G-CSFR), a transmembrane protein comprising 813 amino acids, belongs to the class I cytokine receptor family and is encoded by the CSF3R gene. It contains an immunoglobulin (Ig)-like domain, a cytokine receptor homologous (CRH) domain, and three fibronectin type III (FNIII) domains. G-CSF binds to the ectodomains of G-CSFR with two separate binding sites, a high-affinity site, which is located on helix A and helix C, enabling interactions with both CRH domains. The second binding site is a low-affinity site found on a small additional helix (helix E) within the long connection loop between helix A and helix B, which specifically targets the Ig-like domain (Figure 3 B). The resulting 2:2 G-CSF:G-CSFR complex is illustrated in Figure 3 C [15, 31]. An alternative 2:2 crystal structure complex [32], lacking the immunoglobulin-like domain, was determined for human G-CSF and mouse G-CSFR, likely representing a crystallization artifact.

Within the intracellular domain of G-CSFR, three discrete regions, denoted as box 1, box 2, and box 3, are essential for downstream signaling (Figure 3 D). Boxes 1 and 2 play a crucial role in transmitting proliferation signals while box 3 which is located closer to the C-terminus of the receptor was shown to be crucial for enabling the differentiation of myeloid precursor cells [33]. Additionally, four tyrosine residues (Y704, Y729, Y744, and Y764), located near box 2 and box 3, play an important role in transmitting signals upon G-CSF binding [34, 35]. For example, the residues Y704 and Y744 have been identified to enable the activation of STAT3 by serving as docking sites [36]. Additionally, Y729, Y744, and Y764 are significant for differentiation, with Y764 proposed to be crucial in balancing proliferation and differentiation during G-CSF-driven neutrophil development [37].

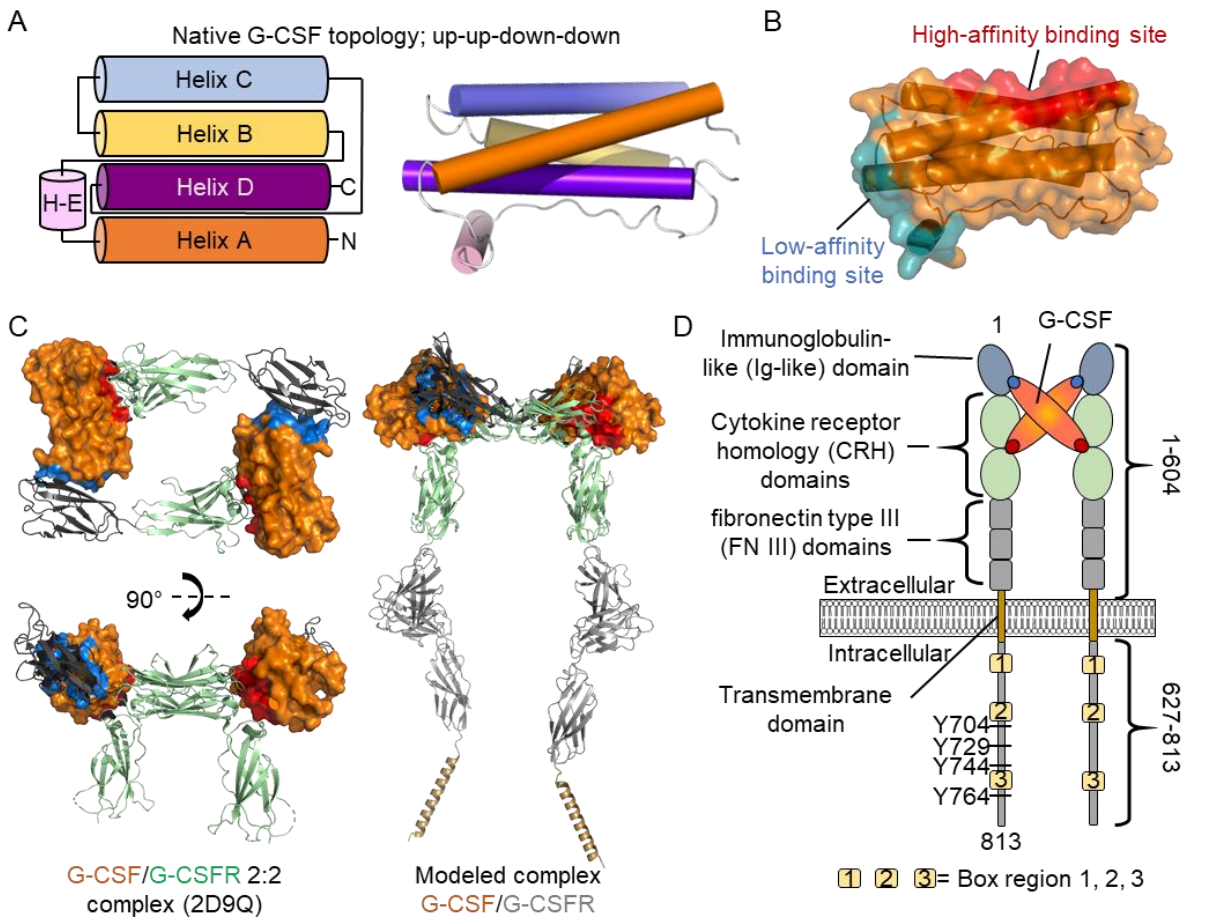



***Figure 3: (A)*** *Topology of native G-CSF.* ***(B)*** *High-affinity (red) and low-affinity (blue) binding site of G-CSF.* ***(C)*** *Structure of the 2:2 complex of G-CSF (orange) and G-CSFR (pale green) (PDB:2D9Q, [15]) and the modeled complex based on the full size AF2 model of G-CSFR (AF-Q99062-F1, [38, 39]), which was cropped for its intracellular part for representative reasons.* ***(E)*** *Schematic of the G-CSF:G-CSFR complex with corresponding colors in (B), (C) and (D).*

### *2.1.3.2 Activation pathways*

G-CSFR undergoes dimerization upon ligand binding which triggers the activation of multiple downstream pathways, including the JAK (Janus kinase)-STAT (signal transducer and activator of transcription), the PI3K (phosphoinositide 3-kinase)-AKT (protein kinase B), and the MAPK (mitogen-activated protein kinase)-ERK (extracellular signal-regulated kinase) pathway (Figure 4) [40]. While the PI3K-AKT and MAPK-ERK pathways are well-known for promoting cell survival and proliferation, the JAK-STAT pathway additionally drives the differentiation of hematopoietic cells.

The intracellular segment of G-CSFR contains membrane proximal docking sites for JAKs, which are activated by transphosphorylation once the receptor dimerizes [41]. The activated JAK2 phosphorylates conserved tyrosine residues on the intracellular

domain of G-CSFR, thereby recruiting STATs [42]. The JAKs further activate STAT3 and STAT5 by phosphorylation, enabling their dimerization and translocation into the nucleus, which ultimately upregulates the transcription of target genes [43].

The complexity of these intracellular signaling cascades is decoded into different cellular responses depending on the cell lineage and differentiation stage [40, 44, 45]. For example, STAT3 is linked to pro-differentiation activity while STAT5 is more associated with proliferation and survival [46]. Additionally, STAT3 persists for hours while STAT5 is transient, so the interplay of both might be a regulatory key in modulating proliferation versus differentiation [40].

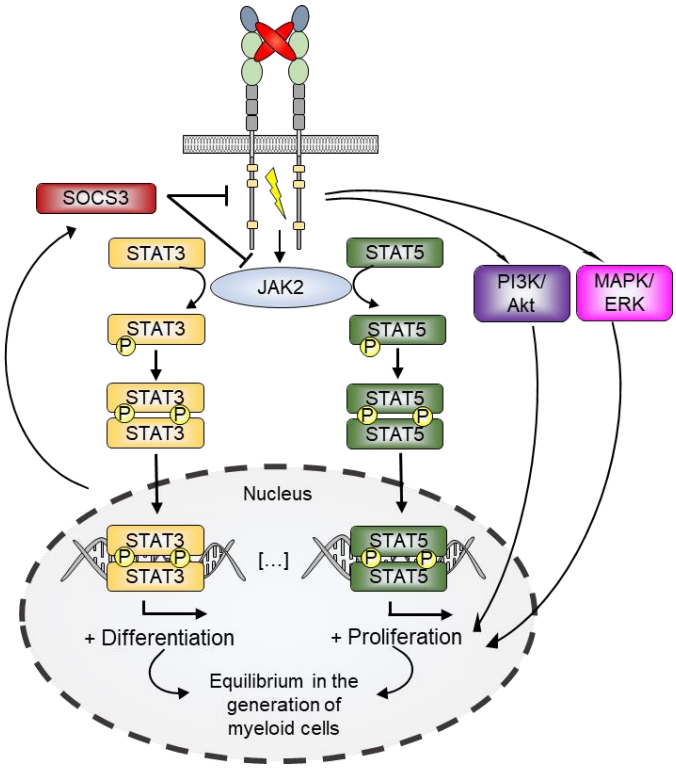



***Figure 4**: Main signaling pathways (JAK/STAT, PI3K-AKT, and MAPK-ERK) that get activated by G-CSFR. Figure is adopted from [40].*

Another difference between STAT3 and STAT5 lies in their activation mechanisms. STAT5 activation requires the presence of the conserved box 1 and box 2 motifs [47]. In contrast, STAT3 can be activated through either a tyrosine-dependent or tyrosine-independent mechanism [48]. In the tyrosine-dependent mechanism, STAT3 binds to a membrane-proximal site on G-CSFR, where JAK phosphorylates specific sites, Y704 and Y744, and subsequently phosphorylates STAT3 itself. This activation type is similar to what is observed for STAT5 and is for STAT3 mainly associated with low G-CSF concentrations [36]. The tyrosine-independent mechanism does not rely on phosphorylated tyrosine residues [49]. Instead, it involves the C-terminal region of the receptor and occurs when G-CSF is present at saturating concentrations (100 ng/mL).

In the JAK-STAT pathway, another crucial component is the family of SOCS (suppressor of cytokine signaling) proteins [50]. While STATs primarily control signal

intensity, SOCS proteins regulate signal duration by representing a negative feedback loop on STATS activation. SOCS3, for instance, is a gene that gets activated by STAT3. Once induced, SOCS3 inhibits JAK2 through two distinct mechanisms [51]. Firstly, it directly binds to JAK2, covering its active site and preventing further JAK2-mediated signaling. Secondly, SOCS3 recruits an E3-ligase, which attaches ubiquitin molecules to both JAK2 and G-CSFR targeting them to proteasomal degradation. These dual inhibitory actions prevent excessive JAK-STAT signaling and is essential for maintaining proper cellular functions.

*2.1.3.3 Clinical significance*

G-CSF is a hematopoietic cytokine that has been in clinical use for several decades as a significant immunotherapeutic agent [3, 52]. Recombinant human G-CSF (rhG-CSF) is utilized in clinical settings either in its native form or with slight modifications [53, 54] to increase neutrophil production and to mobilize hematopoietic stem cells from the bone marrow into the bloodstream [25]. This approach is used to temporally prevent infections in immune system compromised patients, e.g. after chemotherapy, and to facilitate bone marrow donations, respectively [55].

Patients with chronic forms of neutropenia, such as severe congenital neutropenia (SCN), which are characterized by a deficiency of neutrophils, can receive daily treatment with G-CSF to enhance the number of neutrophils in their blood [56]. SCN and related myeloid diseases, including acute myeloid leukemia (AML), result from an imbalance in myeloid progenitor cell proliferation and differentiation, often linked to impaired G-CSFR signaling [27, 40, 57, 58].

The proliferative activity of G-CSFR is implicated in a wide range of non-hematopoietic diseases, including various cancers [48] such as colon cancer [59] and breast cancer [60, 61]. Furthermore, G-CSF plays a critical role in inflammatory diseases [62, 63], such as rheumatoid arthritis [64-66], and exerts an impact on neuronal and vascular plasticity [67, 68] along with different types of pain [69-71].

The widespread involvement of G-CSFR in various diseases renders it a highly appealing target for the development of new treatment approaches. Thus, the central aim of my thesis is to create a diverse array of protein-based G-CSFR modulators that exhibit clinical-grade potency by utilizing protein design.

## *2.2 Protein design*

### *2.2.1 Sculpturing novel proteins*

Before delving into details about protein design, it is important to differentiate it from protein engineering. Protein engineering relies exclusively on the modification of existing proteins [72], while protein design is primarily concerned with the creation of novel protein scaffolds and identifying a sequence that stabilizes this structure [73]. Therefore, protein design is often referred to as the “inverse protein-folding problem” [74]. Protein design can be used to go beyond nature's limited repertoire [75, 76] and create novel, non-natural protein folds that potentially serve desired functions. These non-natural protein folds are often referred to as the “dark matter of protein space” [77].

On the other hand, protein engineering focuses on preserving the defined shape, stability, and dynamics of an existing protein, allowing targeted functional modifications through either empirical mutagenesis or directed evolution [78]. The profound influence of protein engineering in the past forty years has been widely recognized, as evidenced by the 2018 Nobel Prize in Chemistry awarded to Frances Arnold, George Smith, and Greg Winter [79]. Regardless of the differences, protein design and protein engineering closely intertwine, and both are important for creating new scaffolds that can reach practical relevance for applications.

Protein design can be described as three main branches: minimal, rational, and computational design [73]. Minimal design [80] utilizes fundamental chemical principles, mainly the hydrophobic effect of polar and hydrophobic amino acid residues, to direct the folding and assembly of secondary structures. Rational design [81] expands on minimal approaches by incorporating more precise sequence-to-structure relationships derived from biochemical, bioinformatics, or empirical studies. Finally, computational protein design involves the generation and evaluation of full atomistic models in order to search for a low energy sequence given a target structure [82]. Undoubtedly, computational design stands out as the most widely employed approach today and is rapidly advancing also through the use of machine learning methods [83].

For example, substantial progress in structure prediction has been made by machine learning based applications like AlphaFold2 [38]. Additionally, methods like

RFdiffusion [84] have been invented to generate novel protein scaffolds while other applications like ProteinMPNN [85] can rapidly identify sequences that are meant to stabilize a given fold. While AlphaFold2 has also been widely recognized in the field of protein design for structure determination, the long-term impact of new design tools such as RFdiffusion and ProteinMPNN remains to be seen.

### *2.2.2 Facilitating the development of tailored protein therapeutics*

Given the crucial role of proteins in life, it comes as no surprise that a significant portion of diseases can potentially be treated with proteins. Although this potential is not yet fully harnessed, protein therapeutics already play a role in almost every field of medicine [86].

Protein design has evolved into a critical aspect of novel protein therapeutics development, as it enables the creation of proteins with enhanced biophysical [87] and pharmacological properties [88, 89], diverse molecular shapes [90], and plays a crucial role in guiding the *in vitro* affinity maturation of binding proteins [91]. A wide range of novel protein therapeutics were generated like precursor molecules of vaccines [92-94], ligand binding proteins [95], various kinds of binding proteins like a SARS-CoV-2 inhibitor [96], and customized cytokine therapeutics [97]. The various approaches of cytokine design and engineering have led to the development of novel cytokine-based agents. Some of these agents have already gained FDA approval, while an increasing number of other candidates are under investigation in clinical trials [98].

The primary objective of most of these candidates was to mitigate the adverse effects of cytokines arising from their pleiotropic nature or to enhance their pharmacokinetics [99]. Numerous methodologies have been utilized, including the development of fusion proteins, such as immunocytokines. In this context, cytokines are fused with antibodies to facilitate precise cytokine delivery to specific target sites, thereby enhancing localized concentrations within the body and mitigating systemic adverse effects [6]. As this primarily reflects a protein engineering perspective rather than a design-oriented approach, in the following I aim to emphasize the potential offered by protein design through a recent example.

In 2023, Rubio *et al.* created a split version of a known cytokine mimetic, Neoleukin-2/15 [91], to reduce dose-limiting toxicities of monomeric cytokines [100]. They developed inactive Neoleukin-2/15 fragments (Neo2A and Neo2B), which regained

activity when co-localized after independently targeting specific cell surface markers, showing both cis-activation and trans-activation of immune cells (Figure 5). The split Neoleukin-2/15 fusions could offer improved antitumor efficacy and reduced toxicity.

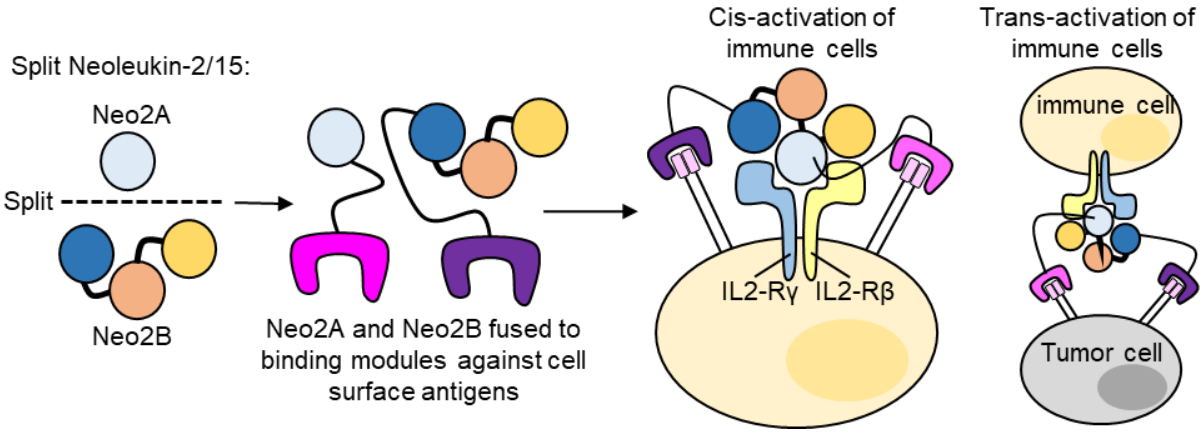



***Figure 5**: Design of a split cytokine. Figure is adopted from [100].*

Alternative design strategies focus on directly fine-tuning parameters of receptor activity through ligand design, rather than altering cytokine targeting, as explained above in the case of the split Neoleukin. Generally, three fundamental parameters are delineated for the modulation of cytokine pharmacology: receptor composition, receptor affinity, and receptor geometry (Figure 6). Tuning these parameters has resulted in numerous cytokine designs characterized by tailored properties, including super agonism, partial agonism, antagonism, or cell selectivity [10]. Governing these parameters represents a significant challenge within the realm of cytokine design.

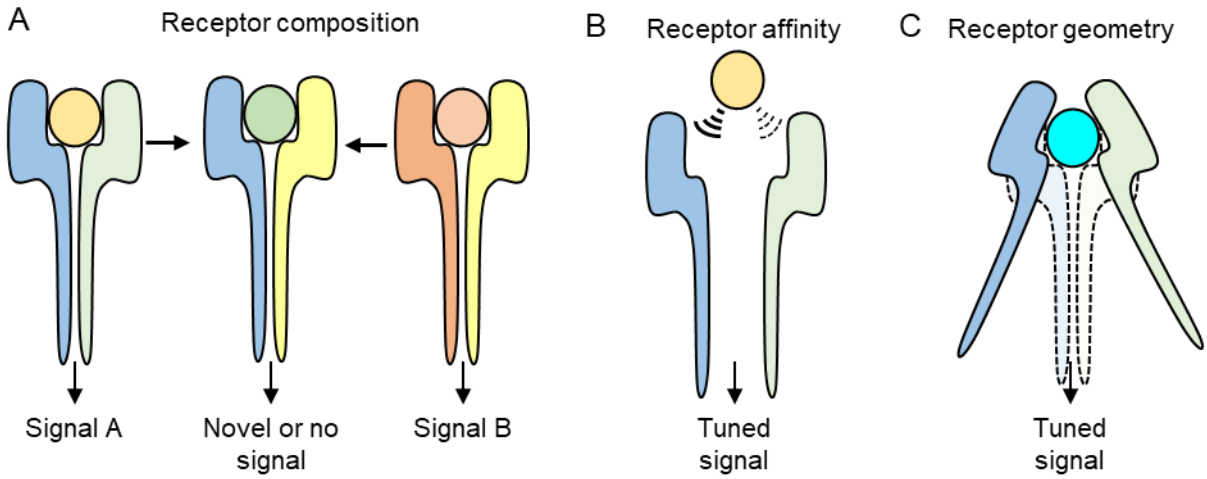



***Figure 6**: The three main parameters in cytokine pharmacology, namely receptor geometry **(A)**, receptor affinity **(B)**, and receptor composition **(C)**, that can be leveraged to modulate cytokine receptor activity. Figure is adopted from [10].*

In addition to the customization of cytokine pharmacology, another critical challenge in cytokine design lies in the creation of optimized cytokine receptor ligands possessing superior stability and cost-effective production compared to their native counterparts [87]. The increased stability of hyper-thermostable cytokine receptor modulators have several significant implications for their utility as enhanced protein therapeutics like simplified storage, administration [101], deimmunization [102] and pharmacokinetic adjustments [103].

## *2.3 Development of G-CSFR modulators*

After highlighting the influence of protein design for the generation of tailored protein therapeutics, particularly cytokines, I now shift the focus to the development of binding modules, specifically targeting G-CSFR as the intended receptor. As emphasized in Section 2.1.3, the clinical significance of G-CSFR makes it an exceptionally attractive target for the development of innovative protein-based activity modulators. To fully explore the potential of G-CSFR, a diverse range of modulators need to be created. Consequently, my work primarily focused on the generation of such protein-based modulators. For this, I selected a hyper-thermostable G-CSFR binding module that had been designed in our group before (Section 2.3.1). Based on this, I established a computationally guided optimization pipeline aimed at identifying superior binding modules for the development of a variety of G-CSFR modulators, as elaborated in Chapters 3 and 4.

### *2.3.1 Design of G-CSFR binding modules*

I will discuss two approaches, topological rescaffolding [104] and topological refactoring [87], that were developed in our group, which focus on designing binding modules targeting a specific G-CSFR epitope while featuring enhanced biophysical properties relative to G-CSF.

Topological rescaffolding [104] was employed to graft the high-affinity G-CSFR binding site from native G-CSF onto an unrelated and simplified scaffold. To identify structurally compatible scaffolds, essential residues of the high-affinity binding site were isolated and used as a geometric search query across the entire Protein Data Bank (PDB). Among the top candidates, the binding epitope was retrofitted, and subsequent molecular dynamics simulations were conducted to identify the most stable constructs. Two promising scaffolds were adopted: one derived from an

uncharacterized protein from *Bacillus halodurans* (PBD:2QUP) [105], and the other being a *de novo* designed, homodimeric, 4-helical, coiled-coil bundle (PBD:5J73) [106]. Through this systematic approach, various novel granulopoietic proteins were generated, surpassing G-CSF in terms of stability and protease resistance.

As mentioned above, another strategy to design compact binding modules is topological refactoring [87] that encompasses two primary steps as explained in Figure 7. Firstly, the protein of interest, in this case, G-CSF, undergoes topological rewiring. This process involves simplifying the protein's topology from an up-up-down-down fold, which necessitates long loops between helix A and B, and between helix C and helix D, to an up-down-up-down fold. This modification avoids the need for long linkers and replaces them with short, *de novo* designed loops. Additionally, the core of the rewired protein is optimized, along with the non-binding epitope regions on its surface. The procedure yielded *de novo* designed G-CSFR binding modules exclusively harboring the high-affinity binding region, and exhibiting exceptional characteristics: hyper-thermostability, high protease resistance, and significantly increased yield compared to natural G-CSF. In tandem fusions, they showed potent activation of G-CSFR signaling, making them highly promising for therapeutic applications.

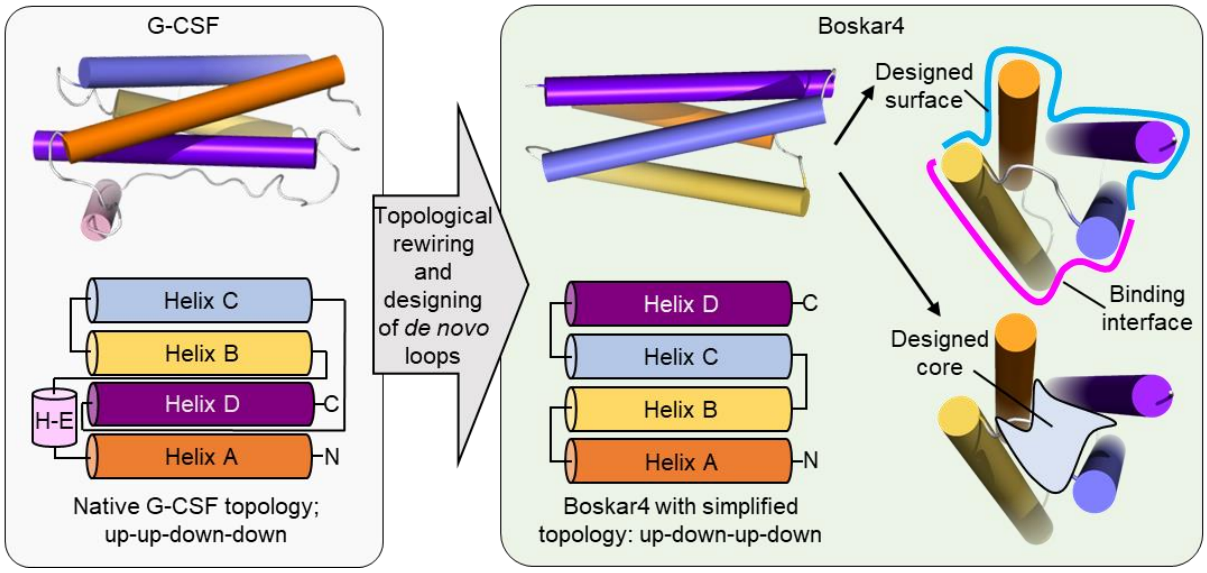



***Figure 7**: Generation of a binding module designed through a topological refactoring approach, exemplified in the design of G-CSFR binders. Figure is adopted from [87].*

Among the candidates generated through topological refactoring in this study, the design named Boskar4 emerged as the most promising. Boskar4 was selected as an

initial candidate to be enhanced by an optimization pipeline and finally to generate a broad range of G-CSFR modulators, as detailed in Chapter 3 and 4.

### *2.3.2 Generation of highly optimized G-CSFR binding modules*

While the realm of protein design holds the promise of creating innovative proteins tailored to specific functions, in many instances, the initially designed candidates necessitate further refinement regarding activity or stability to attain a quality suitable for practical applications [107]. Such refinements typically involve introducing specific mutations to the initial candidate in order to enhance its properties for example by rational engineering approaches [108]. However, the feasibility of these approaches is often constrained by the limitations of classical protein purification methods, which restrict the number of mutants that can be screened. Since my main goal was to improve the affinity of binding modules, I focused my work on high-throughput methods such as directed evolution.

#### *2.3.2.1 Directed evolution*

A widely adopted strategy for augmenting protein functions is directed evolution [78, 109, 110]. In essence, this approach involves initiating mutations in a starting protein, followed by the selection of mutants exhibiting the desired property. This cycle of mutation and selection is iteratively repeated, with each round focusing on the most promising candidates from the previous selection. This process continues until a point of saturation is reached, resulting in significant enhancements of the desired property.

Directed evolution can be applied to a diverse array of challenges, spanning from the evolution of enzyme functions [111] to the enhancement of protein stability [112] and solubility [113] as well as the improvement of binding affinity in binding proteins [114]. The latter process is often referred to as *in vitro* affinity maturation [115]. All these processes share a common characteristic – they are high-throughput, signifying their capacity to screen a wide range of mutants, spanning from thousands to trillions [116]. In brief, these protein mutants are encoded in DNA, and the collective group of mutants constitutes is called a “library”.

In order to screen through large libraries, a fundamental principle is followed: all these techniques establish a connection between genotype and phenotype [78]. Practically, biological systems are used that combine the genetic information with the specific protein of interest they encode within a single particle. The high efficiency of this

method lies in the fact that a single biological particle, such as a bacterial cell harboring a plasmid encoding the protein of interest, allows the evaluation of a single mutant.

*2.3.2.2 Considerations about screening in high-throughput approaches*

While high-throughput systems excel at screening vast quantities of candidates, certain considerations about screening limitations apply universally to all directed evolution methods that employ DNA libraries to screen across a particular sequence space. This sequence space is essentially defined by sequences that vary in a given number of positions containing a specific count of mutations. To illustrate, when one position undergoes mutation among the other 19 canonical proteinogenic amino acids, there emerge 19 potential resultant mutants. Yet, when 10 positions are mutated in a combinatorial fashion, the potential variants already reach around ten trillion, $10^{13}$.

Additionally, the degeneracy of codon triplets that code for desired amino acids plays a role. For instance, to entirely randomize one position, the smallest degenerate codon accommodating all 20 proteinogenic amino acids is NNK or NNS (where N stands for A, T, G, or C, and K represents G or T, and S = C or G) [117]. This leads to 32 distinct codons, causing the candidate pool for the earlier example (where 10 positions were mutated to include all proteinogenic amino acids) to increase from $10^{13}$ to $10^{18}$ candidates at the DNA level.

Furthermore, the produced DNA must be delivered to a corresponding expression system, which gives rise to two inherent limitations. First, there is a constraint on the number of particles that can be generated to establish the genotype-phenotype link. Second, due to the stochastic nature of this delivery process, there exists a probability that a specific candidate becomes part of the particle pool. This delivery process can be approximated by a Poisson distribution [118] and bears significant implications for experimental design [119]. It necessitates the generation of a surplus of particles compared to the potential number of candidates within the library. Essentially, this translates to screening additional particles to maximize the probability of not missing the optimal candidate. For example, if a 95% assurance of encompassing all possible variants is sought, the screening requirement increases to 20 times the total size of the library.

*2.3.2.3 Optimized library design*

A crucial phase in library design involves minimizing the count of potential library members. This aims to maximize the inclusion of sequences encoding enhanced variants and facilitate the screening of all library constituents. To address this challenge in my work, I utilized to two strategies.

Firstly, library generation can be refined to accommodate as many desired sequences as possible. A distinction can be drawn between random mutagenesis and site-directed mutagenesis (Figure 8 A). While random mutagenesis solely offers control over mutation rates by varying PCR conditions [120], site-directed mutagenesis enables the introduction of precise mutations into the target protein [121]. Consequently, the site-directed mutagenesis creates focused libraries and is more vital to my research. For further refinement tools like SwiftLib [122] can optimize degenerate codons to reduce the prevalence of undesired sequences. As such, more costly methods like gene synthesis can be employed to forge entirely focused libraries containing the precise desired variants pool.

Secondly, the sequence count can be reduced before DNA library generation using computational filters that identify optimal mutations at given positions (Figure 8 B). Upstream *in silico* mutagenesis represents such a filter, allowing the design of focused libraries with preoptimized mutations. For this purpose, I used Damietta, a protein design engine, developed in our group, that employs tensorized energy calculations to enhance design throughput and accuracy [123]. In short, Damietta enables the *in silico* optimization of the free energy change in a modeled complex consisting of a mutant binding protein and its antigen. The optimized protein sequences derived from the mutated structures are then utilized to design a corresponding library.

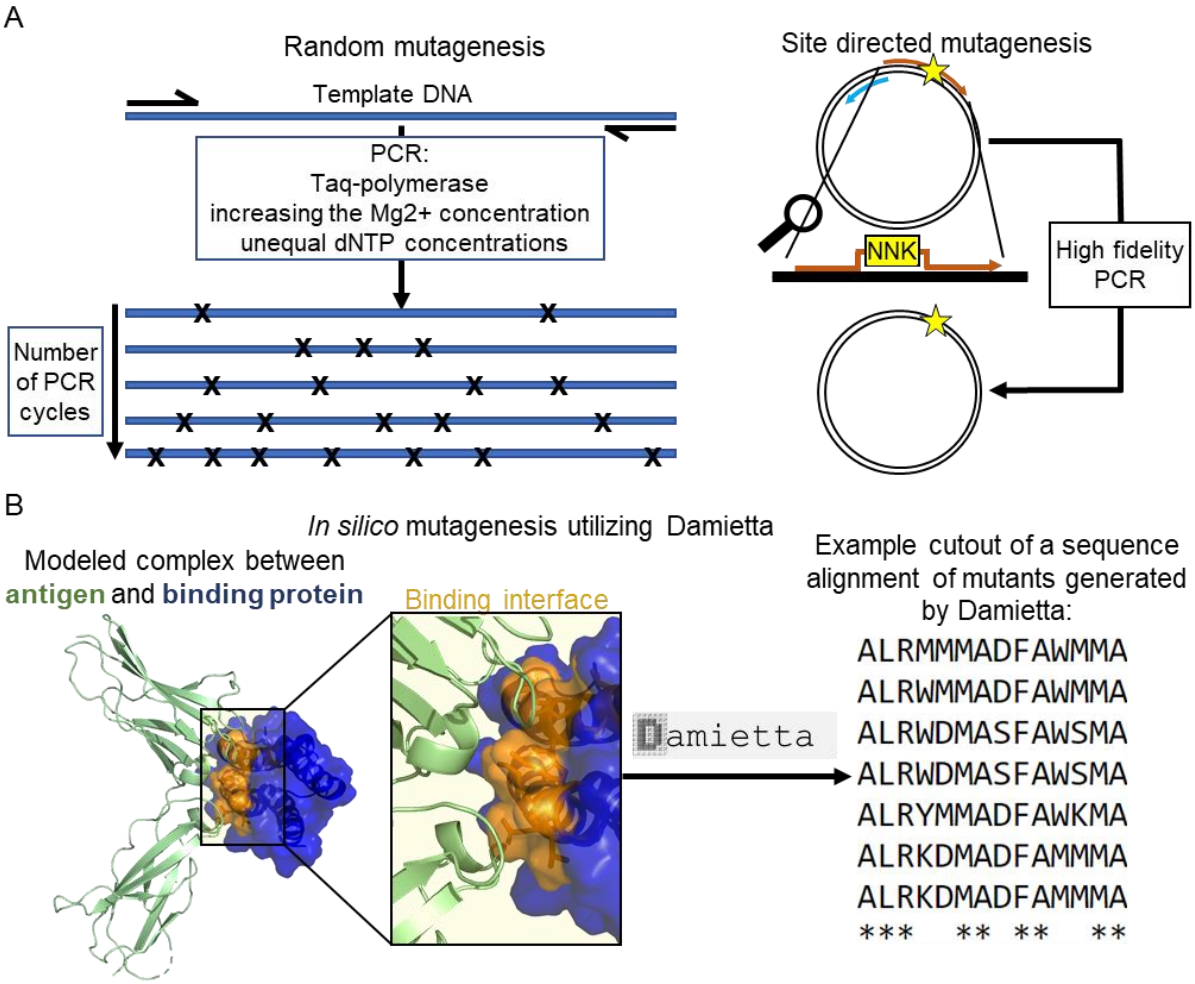



***Figure 8:*** *Distinct categories of mutagenesis.* ***(A)*** *Distinctions between random and site-directed mutagenesis in vitro at the DNA level. Here, "X" signifies a random mutation, while the yellow star denotes a site-directed mutation utilizing a codon NNK that encodes all 20 proteinogenic amino acids.* ***(B)*** *In silico mutagenesis of a modeled complex containing a binding protein (blue) and its corresponding target protein (green). Residues located at the binding interface (orange) are optimized to minimize free energy through the application of the design software Damietta, facilitating the extraction of optimized protein sequences from the mutated binding protein.*

### *2.3.2.4 Building an optimization pipeline*

In the course of advancing binding protein based therapeutics, a prevalent strategy involves the refinement of protein-target interactions through affinity maturation, as part of the development process [115]. This procedure can be executed by constructing a library comprising mutants of the engineered protein, succeeded by a high-throughput screening to identify enhanced binding candidates [91]. Various display systems have been devised [124-127] to discern superior binders, achieved

by linking the binding protein (phenotype) with its corresponding DNA sequence (genotype). Broadly speaking, display systems function by presenting the protein of interest in a manner that facilitates its interaction with its corresponding antigen.

As the focus of my research is to enhance small and stable *de novo* designed binding modules that can be expressed in *Escherichia coli* (*E. coli*), I devised an optimization pipeline as illustrated in Figure 9. This pipeline utilizes computationally guided library design, facilitated by the recently developed software Damietta [123], in conjunction with an *E. coli* display system [128].

Several different bacterial display systems have been devised over time [129]. I will focus here on a single system that utilizes intimin, an outer membrane protein (OMP) originating from enterohemorrhagic *E. coli* O157:H7 (EHEC) [128]. In brief, the binding protein, connected at the C-terminus of intimin, undergoes two steps: (i) transference through the inner membrane (IM) facilitated by the SEC complex, (ii) subsequent transport across the periplasmic space into the outer membrane (OM) aided by chaperones and the β-barrel assembly machinery (BAM) complex.

Upon successful presentation of the binding protein on the cell surface, the interaction with the corresponding antigen can be directly evaluated, often achieved by fluorescently labeling the antigen. Given that each individual cell hosts a distinct variant, fluorescent activated cell sorting (FACS) can be employed to screen through vast populations of cells [125]. FACS detects and sorts cells exhibiting high fluorescence, indicating the presence of a binding protein that demonstrates an increased affinity for the antigen. This enrichment procedure is performed multiple times until the pool of variants no longer exhibits an enhancement in overall fluorescence compared to the preceding enrichment step. Subsequently, individual variants extracted from the enriched pool undergo a thorough analysis (compare Chapter 3 and Chapter 4).

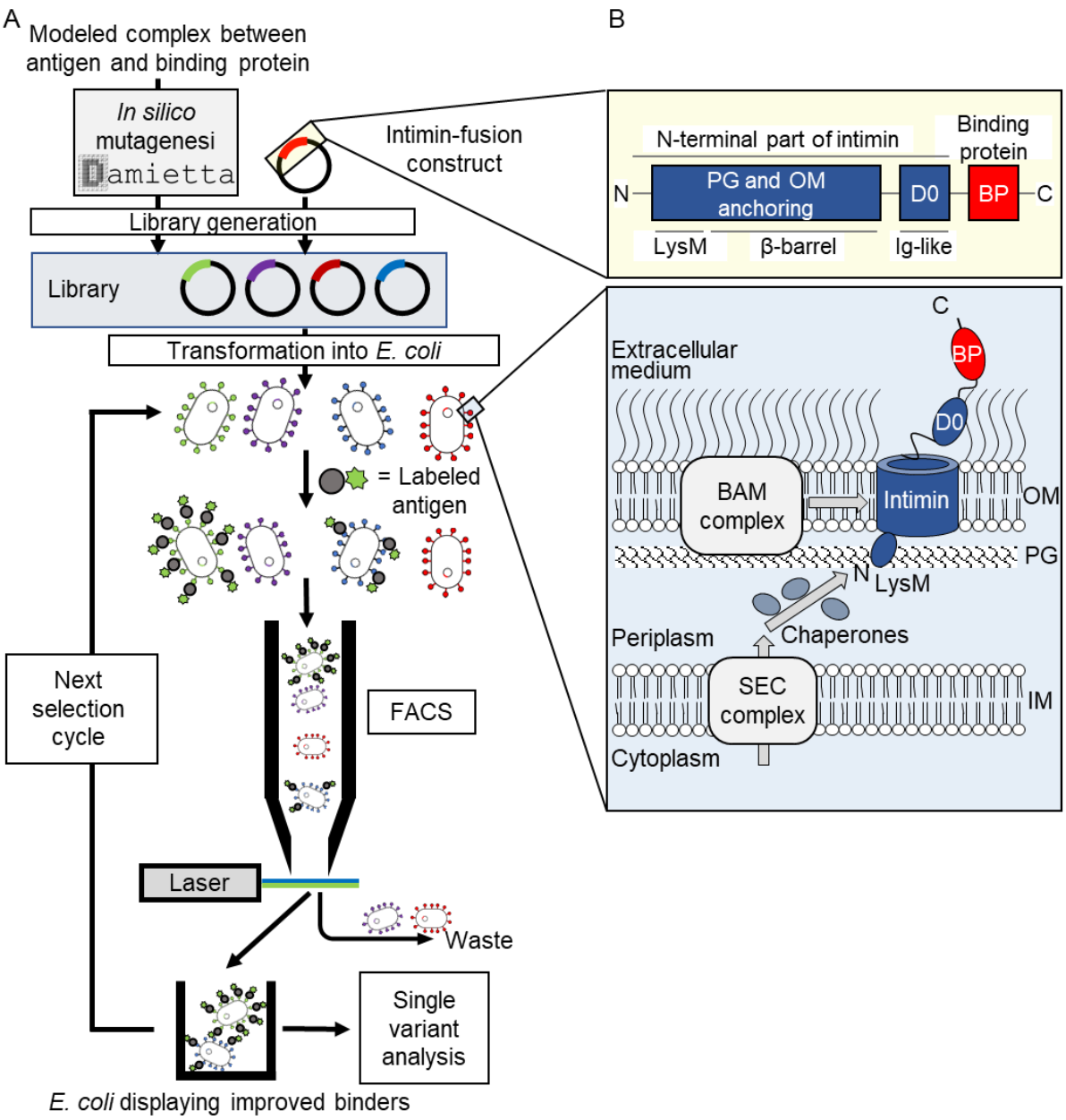



***Figure 9**: Overview on the established pipeline for optimizing designed binding modules. **(A)** The established pipeline employs computationally guided libraries, effectively facilitated by Damietta in tandem with an E. coli display system. **(B)** The secretion pathway of the intimin fusion construct utilized for the E. coli display system. IM, inner membrane; PG, peptidoglycan; OM, outer membrane; N, N-terminus; C, C-terminus; BP, binding protein. Part B) of the figure is adopted from [125].*

# 3 Tuning of granulopoietic signaling by *de novo* designed agonists


## *Abstract*

Enhancing cytokine-based therapies by systematically tuning how an agonist associates its receptor is emerging as a powerful new concept in drug discovery. Here, we report the design and characterization of agonists that tune the granulocyte-colony stimulating factor receptor (G-CSFR) activity, which is central for the proliferation and granulocytic differentiation of hematopoietic stem cells. Using design agonists, we study the impact of varying the receptor-binding affinity and dimerization geometry on receptor association, downstream signaling, and cellular response. Hence, we achieved agonists with altered signaling specificities that are hyper-thermostable, can outcompete the native ligand (G-CSF), and bias granulopoietic differentiation over triggering proliferation. Furthermore, the design agonists differentially modulate the kinetics and amplitudes of signal transduction pathways, and gene expression patterns. Unlike G-CSF, they achieve selective activation of hematopoietic functions with minimal undesired immunomodulatory effects. These findings demonstrate the potential of dissecting the complex G-CSFR signaling, and open ways for new therapeutic applications.


## *Autor contributions*

| **Author** | **Author position** | **Scientific ideas %** | **Data generation %** | **Analysis & interpretation %** | **Paper writing %** |
|---|---|---|---|---|---|
| Timo Ullrich | 1 | 20 | 60 | 50 | 20 |
| Christoph Pollmann | 2 | | 5 | 5 | |
| Maya El-Riz | 3 | | 5 | | |
| Malte Ritter | 4 | | 5 | | |
| Jérémy Haaf | 5 | | 5 | | |
| Narges Aghaallaei | 6 | | 2.5 | 5 | |
| Ivan Tesakov | 7 | | 5 | 5 | |
| Kateryna Maksymenko | 8 | | 2.5 | | |
| Valeriia Hatsovska | 9 | | 5 | | |
| Sergey Kandabarau | 10 | | | 5 | |
| Maksim Klimiankou | 11 | | | 2.5 | |

| | | | | | |
|---|---|---|---|---|---|
| Claudia Lengerke | 12 | | | | 2.5 |
| Karl Welte | 13 | 5 | | | 2.5 |
| Birte Hernandez-Alvarez | 14 | | 2.5 | | |
| Patrick Müller | 15 | 10 | | | |
| Andrei Lupas | 16 | 5 | | | |
| Jacob Piehler | 17 | 10 | | 2.5 | 10 |
| Julia Skokowa* | 18 | 25 | | 15 | 20 |
| Mohammad ElGamacy* | 19 | 25 | 2.5 | 10 | 45 |
| **Title of paper:** Tuning of granulopoietic signaling by *de novo* designed agonists | | | | | |
| **Status in publication process:** Manuscript in preparation | | | | | |

# 4 *De novo* design of G-CSFR-inhibiting proteins


## *Abstract*

Excessive cytokine signaling resulting from dysregulation of a cytokine or its receptor can be a main driver of cancer, autoimmune or hematopoietic disorders. Here we leverage protein design to create tailored cytokine receptor blockers with idealized properties. Specifically, we aimed to tackle the granulocyte-colony stimulating factor receptor (G-CSFR), a mediator of different types of leukemia and autoinflammatory diseases. By creating strictly monovalent G-CSFR binders, we identified hyper-stable, nanomolar signaling inhibitors. X-ray crystallography indicated atomic-level agreement with the experimental structure of an exemplary design. Furthermore, the most potent design blocks G-CSFR in acute myeloid leukemia cells and primary hematopoietic stem cells. Thus, the resulting designs can be used for inhibiting or homing to G-CSFR-expressing cells, and demonstrate a generalizable strategy to tackling a diverse range of cytokine receptors.


## *Autor contributions*

| **Author** | **Author position** | **Scientific ideas %** | **Data generation %** | **Analysis & interpretation %** | **Paper writing %** |
|---|---|---|---|---|---|
| Timo Ullrich | 1 | 60 | 90 | 85 | 40 |
| Olga Klimenkova | 2 | | 5 | 5 | |
| Marcus Hartmann | 3 | | 5 | 10 | |
| Julia Skokowa | 4 | 10 | | | 5 |
| Mohammad ElGamacy* | 5 | 30 | | | 55 |
| **Title of paper:** *De novo* design of G-CSFR-inhibiting proteins | | | | | |
| **Status in publication process:** Manuscript in preparation | | | | | |

# 5 Discussion

The primary objective of this research was to employ protein design techniques to engineer novel ligands that can intentionally modulate the activity of the clinically relevant G-CSFR, with the ultimate goal to fully unlock its untapped potential. To achieve this, a hyper-thermostable *de novo* designed G-CSFR binding module was optimized using *in silico* and *in vitro* high-throughput methods. This optimization process resulted in a diverse range of variants with enhanced binding affinity (compare Chapter 3 and 4). The presented studies demonstrate that the enhanced binding modules can be utilized to generate highly potent G-CSFR agonists capable of fine-tuning G-CSFR activity as well as crafting novel competitive G-CSFR antagonists with nanomolar activity. These ligands represent promising candidates for the development of innovative protein therapeutics. This is attributed not only to their ability to modulate G-CSFR activity but also to their advantageous biophysical characteristics, which streamline the processes of production, storage, deimmunization, and pharmacokinetic adjustments.

## *5.1 Tuning of the G-CSFR activity*

In light of the considerable therapeutic potential associated with cytokines, significant efforts have been undertaken to engineer and design cytokines for therapeutic applications. Over the past several decades, numerous strategies have been developed to address this challenge (compare Section 2.2.2). Some of these strategies involve the targeted delivery of cytokines to specific cell types, while others focus on modulating cytokine receptor activity. In this study, my primary focus was on fine-tuning G-CSFR activity through adjustments in receptor affinity and receptor geometry, while in the following discussion I also probe receptor composition as a promising avenue for developing tailored cytokine designs (compare Figure 6).

### *5.1.1 Receptor composition*

The majority of cytokines exhibit distinct binding affinities for both a unique receptor and a cognate receptor that is shared among multiple cytokines [130]. Thus, the formation of heteromeric complexes among cytokines are well known from natural complexes, yet the existing natural compositions only represent a fraction of the potential cytokine-receptor pairings. These potential non-native pairings hold the potential to introduce novel downstream signaling cascades and could contribute to

cellular specificity due to the selective expression of specific receptor combinations on particular cell types. Therefore, I aimed to contribute to the development of cytokines that incorporate G-CSFR into heteromeric complexes with other receptors potentially improves cellular specificity and enables novel receptor signaling.

However, a limited number of cytokine design categories like synthetic cytokines (synthekines) [131], nanobody-based cytokine surrogates [132], and cytokine chimeras [133] have been established to initiate these non-native receptor compositions. Synthekines consist of two cytokine mutants, each capable of binding only to a single receptor, interconnected by a flexible peptide linker [131]. These constructs have exhibited distinct signaling behaviors compared to the original cytokines, and have been shown to induce unique gene expression profiles. Nanobody-based cytokine surrogates utilize two nanobodies targeting certain receptors connected by flexible linkers [132]. Cytokine chimeras are single-domain cytokine designs where a binding site from one cytokine is integrated into another. Notably, a fusion protein of a cytokine chimera and the Fc domain of immunoglobulin G was observed to reduce the progression of fatty liver disease in *in vivo* experiments [133, 134], positioning it as a promising molecule for type 2 diabetes treatment.

In our group we aim to expand the repertoire of cytokine designs that combine non-native receptor compositions and create cytokines with novel functions (novokines). These novel structures are characterized by a deliberately engineered scaffold, which integrates multiple cytokine binding sites within a single domain (unpublished data). In general, one can distinguish between signaling and non-signaling receptor compositions that can activate or inhibit the corresponding receptors. This design overcomes the constraints associated with cytokine chimeras, which are restricted by the requirement for structurally compatible binding sites due to their utilization of native scaffolds. Furthermore, given that novokines consist solely of a single domain, there is no need to consider any residual activity originating from individual functional sites. This contrasts with synthekines and nanobody-based cytokine surrogates, which consist of two distinct domains connected by a flexible peptide linker. In cases where proteolysis cleaves these synthekines into their individual domains, adverse effects may arise, since they would antagonize instead of agonizing the corresponding receptor.

So far, I investigated a novokine capable of binding to both G-CSFR and Interleukin-3 receptor α (IL-3Rα) (Figure 10). Notably, the residues highlighted in red within this design were derived from bv6, an affinity-enhanced G-CSFR binding module I identified during my high-throughput optimization screens (compare Chapter 3). Surface plasmon resonance measurements demonstrated that the novokine designs incorporating these affinity-enhancing residues exhibited significantly tighter binding affinities compared to those lacking such residues (data not shown). The improved binding affinity remained consistent across various designs incorporating different sets of affinity-enhancing binding residues, thereby reinforcing the potential utility of these enhanced binding modules in the development of novel G-CSFR receptor modulators. For example, I could demonstrate the specific binding of such a novokine targeting G-CSFR and IL-3Rα on the cell surface of mammalian cells using flow cytometry (data not shown). However, it is important to note that this research remains in its early developmental phases and requires thorough investigations to uncover the broader benefits and potential applications of novokines, particularly concerning combinations of receptors that incorporate G-CSFR.

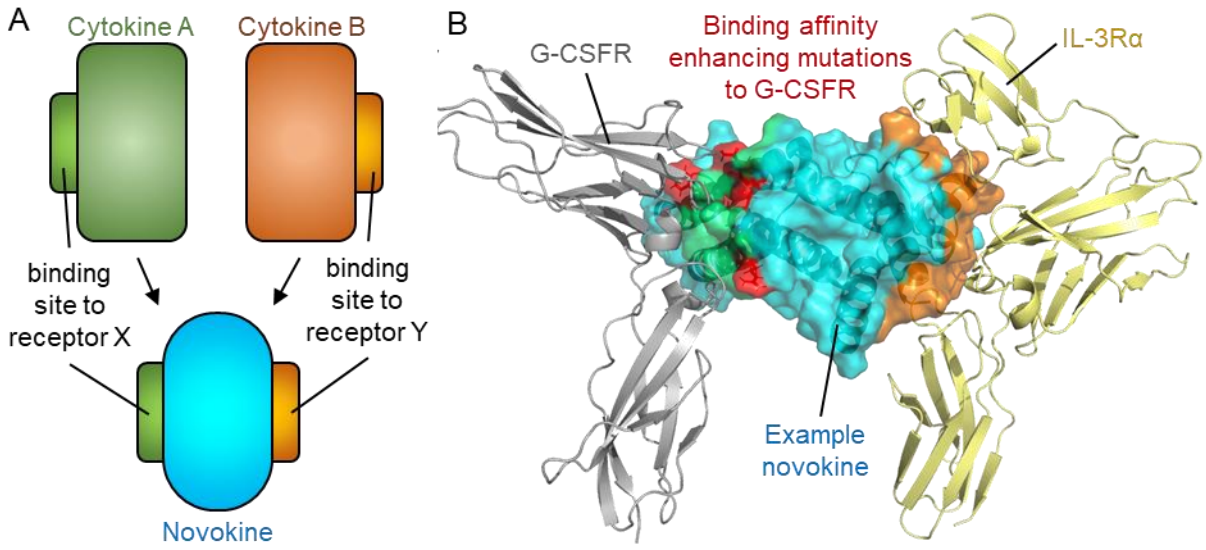



***Figure 10: (A)*** *Concept of Novokine design, aiming to combine multiple cytokine receptor binding sites on a single domain.* ***(B)*** *An example novokine (cyan) is shown which was designed to combine an IL-3Rα binding site (orange) and a G-CSFR binding site (green) carrying the affinity enhancing mutations of bv6 (red). Illustrated is a modeled complex of this example novokine with partial IL-3Rα (pale yellow) and partial G-CSFR (gray).*

### *5.1.2 Receptor affinity*

Engineering the affinity of cytokines to develop molecules with tailored properties is widely studied in diverse biological systems and contexts. In general, one can separate the affinity-engineered cytokines into five classes: super-agonists [135-137], partial agonists [138], cell selective agonists [139-141], biased agonists [136], and decoy-resistant-agonists [142, 143]. In the following, I will focus on super-agonists and biased agonists that result in enhanced signaling or preference for certain downstream signaling pathways compared to the natural agonist.

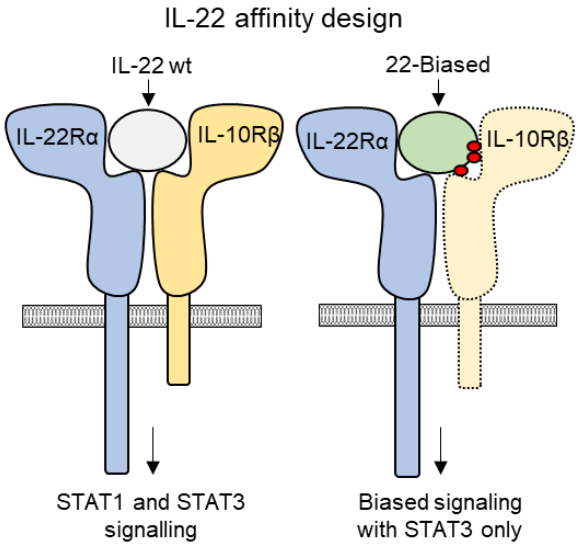



***Figure 11**: Example of cytokine design utilizing affinity tuning. Figure is adopted from [136].*

For example in 2021, Saxton *et al.* [136] presented the design of an biased interleukin-22 (IL-22) agonist that was demonstrated to be dissociated from pro-inflammatory actions, achieved through structure-based design affinity tuning [136]. Essentially, this was achieved by engineering an IL-22 super-agonist with high affinity for the IL-10 receptor beta (IL-10Rβ), enabling the determination of the ternary complex consisting of IL-22, IL-22 receptor alpha (IL-22Rα) and IL-10Rβ. Through destabilizing the IL-22–IL-10Rβ binding interface, partial agonist analogs were designed with diverse STAT bias as shown in Figure 11, leading to tissue-selective signaling for tissue protection without inducing inflammation.

The provided example shows the potential application of affinity tuning in the development of customized cytokine-based therapeutics. Hence, the first aim of my research was to design G-CSFR binding modules featuring a range of binding affinities, thereby facilitating the investigation of affinity tuning G-CSFR agonists.

Previous research had already demonstrated that combining two of these modules in tandem fusions yielded biologically active G-CSFR dimerizers [87]. Consequently, for the binding modules I generated in Chapter 3, it was expected that the enhanced affinity would lead to a higher dimerization efficiency, which would increase biological

activity by reducing the half-effective concentration ($EC_{50}$). I identified sixteen affinity-enhanced variants that outperformed the non-enhanced module in cell proliferation activity, with the most active one matching rhG-CSF, and subsequent *in vivo* experiments confirmed their efficacy in zebrafish and mice. However, I did not observe a straightforward correlation between binding affinity and proliferative activity when comparing the affinity-enhanced variants with one another.

Several factors may account for this absence of correlation. For example, the variants might possess distinct binding specificities, potentially resulting in lower effective concentrations and thus affecting observed proliferation activities. Furthermore, these variants might display modified receptor internalization, intracellular trafficking, or intracellular signaling activation patterns, all of which could lead to alterations in their activity [144, 145]. Additionally, receptor activation is influenced not only by binding affinity but also by the exact binding kinetics, described by the association and dissociation rate. Kim *et al.* demonstrated that a single-point mutation (R150Q) within the high-affinity binding site of Erythropoietin (EPO), resulted in an increase of both the association and dissociation rate to the EPO receptor (EPOR) [146]. Although this mutation led to a mild reduction in binding affinity compared to the wild-type EPO, it induced a pronounced bias in STAT5 activation, ultimately causing severe anemia in individuals harboring homozygous R150Q mutations. This example underscores the potential to create biased agonists through affinity tuning for homomeric receptor signaling complexes, such as EPOR and G-CSFR. Consequently, the lack of a clear correlation between $EC_{50}$ values and the corresponding affinity within the enhanced G-CSFR agonists may be partly attributed to variations in the binding modules' association and dissociation rates.

Taken together, in my work, I demonstrate, that the modulation of G-CSFR activity by receptor affinity tuning can be applied to increase the potency of designed agonists. Nonetheless, the receptor affinity tuning of G-CSFR was observed to be not straightforward on a detailed level and a multitude of other factors could be further investigated, to widen the knowledge about G-CSFR activation. The affinity enhanced variants I found during the affinity maturation process represent a broad spectrum of binding modules with varying binding kinetics and they could be utilized as a basis for more detailed studies.

### *5.1.3 Receptor geometry*

Curiously, despite the essential roles of cytokines, the structural elucidation of cytokine-receptor complexes revealed a subset of theoretically possible complex geometries [41], indicating an inherent limitation of natural cytokines to facilitate diverse receptor conformations. Consequently, cytokines and their receptors can be classified according to their structural characteristics [147]. The understanding of manipulating cytokine receptor activity through modulation of ligand-receptor complex geometry remains less explored in contrast to cytokine affinity engineering. Nonetheless, I designed novel agonists capable of modulating G-CSFR activity through receptor geometry manipulations for three main reasons.

Firstly, to the best of my knowledge, no studies have attempted to design proteins to intentionally manipulate the G-CSFR complex geometry in a predetermined manner, aiming to induce G-CSFR signaling beyond its native activation patterns [130, 148, 149]. However, there is one report where an agonistic antibody against G-CSFR was screened utilizing an autocrine high-throughput mammalian cell system [150]. The researchers document an unforeseen occurrence wherein this antibody induced neurogenesis within human $CD34^+$ stem cells. Consequently, they claim that the antibody agonists and the native ligand, both binding to a common receptor, could elicit diverse cellular outcomes from an initially uniform cell population.

Secondly, the use of receptor geometry, rather than binding affinity, was chosen to effectively override G-CSF activity. With affinity tuning certain downstream signaling patterns can only be achieved by diminishing the affinity to one receptor binding site, as exemplified by the biased IL-22 design and the EPO R150Q mutant (compare Section 5.1.2). In such cases, the natural ligand can outcompete the designed one due to its superior affinity for the tuned binding site. In contrast, the affinity-enhanced geometry tuner designs can outcompete the natural ligand, thereby effectively overriding the signal.

Thirdly, the development of G-CSFR modulators enabling precise activity adjustment would offer essential protein-based tools to address hematopoietic stem cell disorders associated with aberrant proliferation or differentiation [27, 40, 56, 151]. In addressing this as a pivotal challenge within my thesis, I designed ligands that I termed “orientation

rigging designs" (ori) which have been extensively elucidated in Chapter 3. These oris represent a new class of cytokine receptor geometry modulating designs.

Generally, a limited number of studies have employed the deliberate tuning of cytokine receptor activity by manipulating the complex geometry through the utilization of diabodies [152], nanobody-based surrogate cytokines [132], designed ankyrin repeat proteins (DARPins) [90]. Figure 12 shows a comparison between an example of each of these designs and an ori.

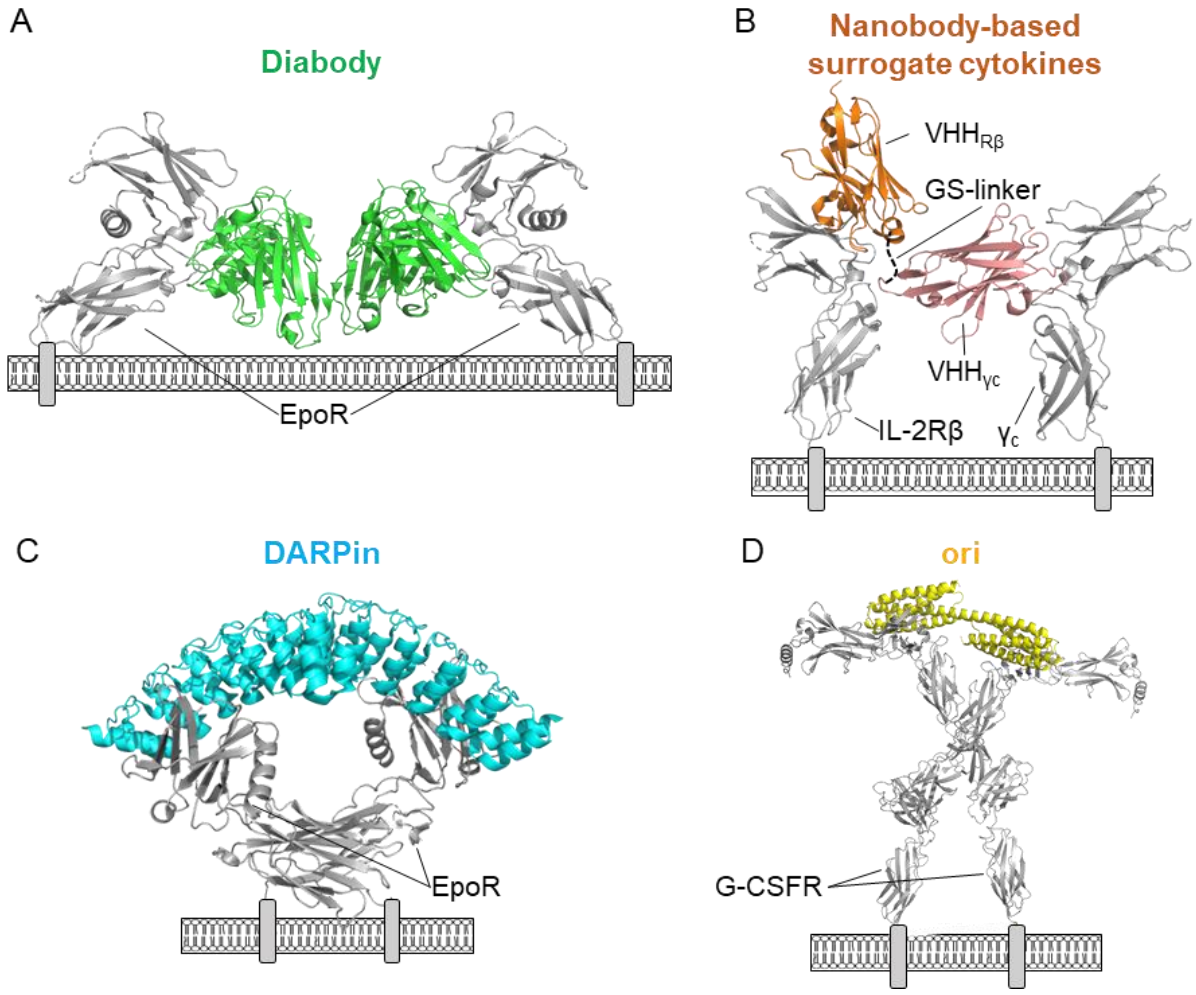



***Figure 12:*** *Receptor geometry modulating designs:* ***(A)*** *diabody (green, PDB:4Y5Y [152]),* ***(B)*** *a nanobody-based cytokine surrogate (modeled complex based on PDB:7S2S [132] (orange) and PDB:7S2R [132] (salmon)),* ***(C)*** *DARPin (cyan, PDB:6MOI [90]), and* ***(D)*** *ori (modeled complex based on PDB:2D9Q [15], AF-P19235-F1 [38, 39] and an AF2-model of ori0 (yellow)). Gray boxes: transmembrane domains of the corresponding receptors (gray), GS-linker: flexible glycine-serine-linker, other abbreviations: see text.*

Diabodies are covalently linked dimeric antibody variable fragments that were shown to enable the generation of surrogate ligands [153]. In the case of EPOR, diabodies were discovered capable of activating the receptor and inducing variable STAT levels [152]. Subsequent structural analysis of ligand-receptor complexes revealed strong differences in EPOR subunit orientations, indicating a correlation between topology and signaling potency. For simplifying surrogate cytokine design, an alternative strategy replaces diabodies with nanobodies [132], single-domain fragments derived from a VHH (variable domain of heavy chain of heavy chain-only antibodies) [154]. Practically, this streamlined approach enables methodical investigation of nearly all cytokine receptor combinations in a semi-high throughput manner. Thus far, this approach has been employed for two receptor pairs, namely Interleukin-2 receptor β (IL-2Rβ) and common γ-chain (γc), along with type I interferons receptor 1 and type I interferons receptor 2. Nevertheless, due to the use of flexible peptide linkers for nanobody linkage, precise control over the induced complex geometry remains limited.

In 2019, Mohan *et al.* created DARPins that enable the precise homodimerization of the EPOR at different distances and angles [90]. The designed structures were found to display a spectrum of agonistic behaviors, including full, partial, and biased agonism. These designs also influenced hematopoietic processes, impacting both differentiation and proliferation pathways. One drawback associated with the designs featured in this study is their relatively large size, comprising over 400 residues. Furthermore, depending on the specific design, their ability to dimerize the target receptor necessitates the formation of a homodimer. Consequently, these modules would encounter a loss of their intended functionality when faced with situations where homodimerization is not favored, e.g. due to local changes in ion concentration or pH *in vivo*. In this scenario, the individual monomers would probably function as competitive EPOR antagonists.

My ori designs encompass a minimal, rigid linker helix connecting two G-CSFR binding modules. Modulating the length of the compact linker helix induces an angular displacement between the binding sites of the individual binding modules. This approach creates compact and rigid binding proteins that generate diverse homodimeric G-CSFR complex geometries. These single-chain constructs offer enhanced structural rigidity compared to diabodies and nanobody surrogates, as well as reduced molecular size in comparison to DARPins.

On the other hand, an inherent constraint in the ori designs is the restricted set of possible angles between the binding sites of individual binding modules imposed by the length of the rigid helix linker. The five ori designs of my work (ori0, ori1, ori2, ori3, and ori4) mainly embodied the spectrum of attainable receptor geometries within this design strategy. Since these attained geometries encompass extreme configurations in terms of inter-transmembrane domain (inter-TMD) spacing, they indeed represent a substantial subset within the generally boundless realm of potential receptor geometries. Henceforth, I successfully initiated the manipulation of receptor geometry for G-CSFR, resulting in significant modifications in receptor activities as described in detail in Chapter 3. This not only highlights the overall tunability of the receptor but also provides comprehensive insights into the activity adjustments within these extreme geometrical configurations. In the following paragraphs, I will consider potential reasons for the modified signaling of G-CSFR resulting from non-native geometries characterized by altered inter-TMD spacing.

The observed variations in cellular activity attributed to different ori designs likely result from shifts in their JAK-STAT pathway signaling patterns (compare Section 2.1.3). Notably, cells treated with oris featuring high inter-TMD spacing like ori1 and ori2 exhibit reduced levels of phosphorylated STAT5 (pSTAT5), a known driver of cell proliferation, compared to those with shorter, more G-CSF-like inter-TMD spacing like ori0 (see Chapter 3). This may partially explain the clear correlation between cell proliferation activity and inter-TMD spacing during NFS-60 assays.

Additionally, the ori designs, especially ori0, showed minor proliferative activity on HSPCs but a significant capacity to differentiate primary stem cells into granulocytes. This could be the case because phosphorylated STAT3 (pSTAT3) levels were more consistent across designs compared to pSTAT5 levels, possibly due to distinct activation mechanisms. STAT5 relies on membrane-proximal box 1 and box 2 motifs, while STAT3 can be activated via tyrosine-dependent, similar to STAT5, or tyrosine-independent mechanisms [48]. The latter mechanism does not rely on phosphorylated tyrosine residues but involves a receptor region near the C-terminus and is predominantly observed at high G-CSF concentrations [49]. For designs with high inter-TMD spacing, the membrane-proximal regions may not come into sufficient spatial proximity to effectively activate the tyrosine-dependent STAT3 and STAT5 pathways. Conversely, the long and flexible intracellular part of G-CSFR may allow

sufficient spatial proximity in the C-terminal region to activate the tyrosine-independent STAT3 pathway, which might explain the observed preference for pSTAT3 of the ori designs.

In addition, pSTAT3 is primarily associated with hematopoietic cell differentiation; hence, it is postulated that pSTAT3 and pSTAT5 play a pivotal role in the balancing between proliferation and differentiation. Moreover, Y764, also located near the C-terminus, has been demonstrated to be crucial for differentiation and is similarly considered to maintain the balance between proliferation and differentiation in the context of G-CSF-driven neutrophil development [37]. Therefore, manipulating inter-TMD spacing with G-CSFR modulators holds great promise for shifting the differentiation-proliferation equilibrium, offering innovative therapeutic potential for conditions with such imbalances and the development of tailored G-CSFR agonists.

### *5.2 Design of novel small and stable G-CSFR antagonists*

As expounded in Section 2.1.3.3, G-CSFR emerges as a clinically significant receptor. The creation of novel G-CSFR antagonists holds considerable promise due to the well-established connection between G-CSFR and a multitude of hematopoietic diseases such as leukemia [40, 57], as well as non-hematopoietic disorders. This encompasses a wide range of cancers [48, 59-61], as well as contributions to inflammatory ailments [62-66], adaptability in the neurovascular system [67, 68], and different types of pain [69-71]. Despite the extensive range of connections between G-CSFR activity and various disorders, the availability of G-CSFR antagonists remains notably limited.

To the best of my knowledge, no small and stable protein-based non-antibody antagonists targeting G-CSFR have been documented in literature. In Chapter 4, I introduce a diverse array of compact, hyper-thermostable, nanomolar inhibitory agents for G-CSFR, paving the way for innovative therapeutic strategies.

The predominant approach in developing G-CSFR antagonists thus far was the creation of a neutralizing anti-G-CSFR antibody, which acts to inhibit G-CSF-induced neutrophilia by prohibiting the binding of G-CSF to its receptor [155, 156]. Importantly, it has been noted that the administration of this antibody in nonhuman primates does not result in neutropenia. This observation suggests that the antagonism of G-CSFR might be attainable without compromising basal granulopoiesis. Potential applications for G-CSFR antagonists have been investigated in mouse models for the treatment of

inflammatory arthritis [64], experimental autoimmune uveoretinitis [157], neutrophilic inflammation [158], severe asthma [158, 159] and renal Ischemia-Reperfusion Injury [160].

While antibodies are extensively employed as protein therapeutics within clinical practice [161, 162], their utility is constrained by several challenges, including their substantial size and costly production [163]. The latter contributes to their limited affordability for a significant portion of the global population [163]. In addition, the effectiveness of antibody therapeutics relies on the continuous maintenance of a cold chain storage system, and their administration usually necessitates injections. Alternative, more convenient, and secure routes, such as oral application, are not feasible for antibodies due to the high production costs and sensitivity to protease degradation. These draw backs could be overcome by the development of antibody alternatives with lower production cost and enhanced stability [164]. The designs presented in Chapter 4 possess hyper-thermostability and can be cheaply produced in bacterial expression systems. Therefore, they might be superior to antibodies for broader clinical applications. Further details on the stability aspect can be found in Section 5.3.

Given the constraints associated with antibodies, significant endeavors are directed towards the development of non-antibody binding proteins like Anticalins [165, 166], Affibodies [167], DARPins [168] and many others [163]. These non-antibody binders can be produced in *E. coli* as they lack disulfide bridges and post-translational modifications, whereas more complex proteins such as antibodies often require mammalian cell expression systems such as Chinese hamster ovary cells [169]. Due to the enhanced expression efficiency and reduced maintenance costs in comparison to mammalian cells, *E. coli*-mediated protein production leads to a substantial reduction in overall production expenses. This consequently renders the manufacturing process economically more viable than that for antibodies [170, 171]. Nevertheless, akin to antibodies, these binding modules lack inherent specificity for particular epitopes. Hence, a significant screening endeavor is necessary to create epitope-specific binders with the capacity to act as competitive inhibitors. Conversely, the non-antibody binders I designed in Chapter 4 were tailored to selectively target a specific epitope, circumventing the need of laborious screening efforts for competitive binders.

Additionally, my designs offer a general advantage due to their compact size of approximately 15 kDa which is 10-times less than typical antibodies with around 150 kDa. This reduced size enhances their tissue and tumor penetration capabilities [172-174]. Although the specific applications of the potent antagonists introduced in Chapter 4 remains to be determined, enhanced tumor penetration could be particularly valuable for treating solid tumors linked to G-CSFR activity, such as colon cancer [59] and breast cancer [60, 61]. Furthermore, the developed designs can be readily integrated into fusion constructs potentially serving as a homing tag, directing payloads towards cellular populations that exhibit G-CSFR expression.

### *5.3 Impact of protein stability on their applicability as therapeutics*

As exemplified in Chapter 3 and Chapter 4, my work focuses on the development of compact, hyper-thermostable proteins equipped with tailored functionalities, with the ultimate goal of facilitating novel applications in the field of protein therapeutics. Given the exceptional thermo-stability of all the designed ligands presented in these studies, this attribute greatly enhances their suitability for therapeutic applications.

In general, protein stability can be described by chemical and physical instability [175]. Chemical instability encompasses reactions that result in the formation or cleavage of covalent bonds, leading to the emergence of new chemical entities. Processes attributed to chemical instability include phenomena such as aspartic acid isomerization, tryptophan hydrolysis, oxidation, or proteolysis. Conversely, physical instabilities in proteins pertain to scenarios where the chemical composition remains unaltered, yet the protein undergoes changes in its physical state [176]. Processes associated with physical instability predominantly encompass denaturation, surface adsorption, aggregation, and precipitation. Often, these processes are interrelated; for instance, the adsorption of proteins, particularly those that have partially unfolded and exposed hydrophobic side chains, can trigger their aggregation.

Protein aggregation often presents the most critical aspect in the development of therapeutic proteins. Aggregates can induce immunogenic responses [177], and compromising the efficacy of therapy rendering them unsuitable for medical applications. Therefore, the control over a protein's intrinsic stability is important to minimize aggregation during processing, storage and application stages. Conformational stability was found to be being a primary factor and is crucial for

managing protein aggregation effectively [178]. Augmented conformational stability is indicated by heightened thermal stability, as denaturation, a process characterized by the disruption of the protein's native structure, is frequently instigated by elevated temperatures. Additionally, the reversibility of denaturation also holds significant importance in ensuring storage stability [179]. Considering that the ligands introduced in Chapter 3 and Chapter 4 exhibit thermostability over 110 °C, reversible folding capabilities, or both, they represent optimal candidates for mitigating the risk of protein aggregation and, consequently, addressing the previously mentioned issues.

Additionally, engineering protein therapeutics with exceptional thermostability potentially eliminates the need for cold chain logistics, which are typically essential for preserving the integrity of native protein-based therapeutics. This advancement is particularly significant in regions where comprehensive refrigeration infrastructure is not readily available. Moreover, an elevated level of conformational stability frequently imparts heightened resistance to enzymatic degradation, consequently augmenting the observed protease resistance [87, 104].

These attributes may render such proteins promising candidates for oral administration, offering a more convenient and lower-risk mode of delivery compared to conventional protein therapeutics, which are commonly administered via parenteral routes [101]. One of the main reasons why oral administration is very challenging for protein and peptide drugs are the harsh gastrointestinal (GI) milieu, with extreme pH variations and enzymatic degradation catalyzed by GI peptidases [180]. In this regard, my agonist designs that exhibit high stability are better equipped to withstand these hostile environments. Additionally, their compact size enhances the likelihood of efficient absorption from the gastrointestinal system into the bloodstream [181]. To further augment this absorption rate, the application of permeabilization enhancers could be considered [182, 183].

In general, the ability to facilitate the oral administration of G-CSFR modulators would represent a significant stride in the treatment of G-CSFR-related diseases. To tackle this challenge, an ongoing collaboration is focusing on to the investigation of oral bioavailability in mouse models, specifically targeting the most promising agonist design I developed in Chapter 3. To produce the substantial amounts (in the gram range) of the agonist design required for these assays, I have optimized a protein

purification protocol that leverages the hyper-thermostability and high yield of the design (unpublished data).

Generally, hyper-thermostability confers additional advantages, opening up novel applications that are unattainable with traditional proteins and thereby augmenting their utility in diverse therapeutic and diagnostic domains. An illustrative instance is the potential to seamlessly integrate such proteins directly into the 3D printing processes of biomaterials, as demonstrated with a hyper-thermostable IL-4 mimetic [184]. In general, 3D printing of biomaterials holds substantial promise in the creation of scaffolds for tissue engineering and regenerative medicine [185]. Considering that G-CSFR activity significantly influences granulopoiesis in the bone marrow (compare Section 2.1.2), the hyper-stable designs presented in Chapter 3 could potentially be incorporated directly into 3D printing processes for bone-like tissues, thereby supporting the production of granulocytes.

Finally, highly stable binding modules can be reengineered without compromising their functionality, unlike less stable ones derived from natural scaffolds [186]. These stable designs tolerate extensive modifications while maintaining core functionality. In contrast, less stable proteins face a stronger trade-off between functionality and stability [187]. Minimizing this trade-off enables seamless incorporation of high-stability designs into fusion constructs, adding functionality without sacrificing structural stability [188]. Furthermore, the high flexibility in reengineering is particularly crucial for the processes of deimmunization.

### *5.4 Deimmunization of designed proteins over native proteins*

The development of protein therapeutics, critically depends on conducting a thorough assessment of immunogenicity, followed by the subsequent deimmunization [189] of the corresponding constructs. To develop G-CSFR modulators suitable for clinical use, a future objective is to provide deimmunized binding modules.

Generally, the immune system plays a pivotal role safeguarding it against external threats such as infectious bacteria, viruses, fungi, and parasites. It possesses the crucial ability to detect non-self proteins, triggering an immune response that not only counters these foreign substances but also induces inflammatory reactions [190]. Therapeutic proteins can trigger such reactions if they are sensed as non-self substances, which can lead not only to the production of anti-drug antibodies (ADAs),

diminishing the therapeutic efficacy of the administered drug [191] but also pose additional risks such as hypersensitivity and anaphylactic reactions [192]. Furthermore, human protein-based drugs can induce cross-neutralizing antibodies that obstruct the drug's efficacy and interfere with the function of their native, body's own counterpart [193]. Designed protein therapeutics that exhibit only minor sequence similarities to their natural counterparts are less likely to cause such a cross-reactivity.

Also, the absence of post-translational modifications in recombinant protein drugs contributes to the risk of immunogenicity associated with proteins [194] and protein instability, which can result in degradation and aggregation (compare Section 5.3). The latter is believed to be a critical factor in the development of ADAs [195]. Consequently, highly conformationally stable proteins, like the designs I introduced in Chapter 3 and Chapter 4, may be less immunogenic, as they are potentially less susceptible to *in vivo* degradation and aggregation.

One primary pathway for recognizing non-self proteins involves their uptake by APCs, such as dendritic cells, followed by the proteolytic breakdown of these proteins into smaller peptide fragments. Subsequently, depending on the specific cell type of the APCs, the acquired peptides can be presented on the cell surface through binding to major histocompatibility complexes class I or class II (MHCI or MHCII). These presented peptides activate T cells with corresponding T cell receptors, finally initiating a cascade of immune reactions [196]. MHC ligands capable of eliciting a T cell immune response are commonly termed T cell epitopes. Typically, non-self peptide sequences harbor T cells epitopes [197]. Therefore, substantial efforts have been dedicated to the identification of T cell epitopes within protein therapeutics [198] to deimmunize the protein of interest by mutagenizing its sequence accordingly.

To account for non-self-recognition, the sequences of the most promising improved G-CSFR binding modules that I identified in high-throughput optimization screens in Chapter 3 and Chapter 4 are planned to be deimmunized. Therefore, one could use machine learning-based tools [199] to detect potential T cell epitopes, validate these selections through immunogenicity assays, and finally alter the sequence of the designs accordingly. However, alterations in the amino acid sequence harbor the risk of compromising the overall stability of the protein. In contrast to proteins that lack hyper-stability, the remarkable thermostability of my G-CSFR modulator designs

probably permits substantial alterations to their sequences without compromising their functionality.

Another challenge is the potential need for mutations near or within the binding site of the protein. Consequently, a range of diverse binding interfaces may be necessary as part of the deimmunization strategy. This can be accomplished by high-throughput approaches to identify alternative binding sites that preserve comparable functionality but are encoded within distinct sequences. Therefore, the value of this work, which has yielded a diversity of G-CSFR binding interfaces, extends beyond providing varying affinities for receptor activity tuning; it also contributes to the sequence variability within the binding site required for the deimmunization process.

### *5.5 Modulating pharmacokinetics of G-CSFR agonist designs*

In my view, the next critical steps in the development of a clinically relevant G-CSFR agonist include improving their pharmacokinetics, especially extending the elimination half-life. Broadly speaking, optimizing the pharmacokinetics of protein therapeutics is vital for their efficacy. The hydrodynamic radius is linked to a molecule's size and exerts a pronounced influence on the pharmacokinetic properties [200]. This has been demonstrated in the context of antibody and corresponding antibody fragment clearance across various size ranges [201]. In the following I will discuss existing long-acting G-CSFR agonists and propose an alternative strategy based on the enhanced G-CSFR binding modules described in Chapter 3 and Chapter 4.

Most cytokines have short circulation half-life thus need to be administered frequently to enable a therapeutic effect. Therefore, various approaches aim to generate long-acting cytokines [6, 200]. One frequently employed strategy involves conjugating the cytokine of interest with various entities such as PEG, thereby augmenting their molecular size to reduce renal clearance [202]. Another prevalent approach involves fusion with Fc-domains sourced from IgG [203] or albumin [204], which, aside from augmenting the molecule's size, facilitates interaction with the Neonatal Fc receptor. This interaction protects the protein of interest from intracellular catabolism [205].

Since rhG-CSF (e.g. Filgrastim) has a molecular weight of 19 kDa and a short circulating half-life of 3 to 4 hours [206] substantial endeavors have been directed toward the creation of a long-acting G-CSF. The clearance of rhG-CSF occurs through proteolysis, renal clearance and neutrophil-mediated clearance [207, 208]. However,

the main case for the fast clearance of Filgrastim was sought out to be renal clearance. To overcome the short half-life of rhG-CSF pegylated G-CSF (Pegfilgrastim, [209]) was developed which carries a 20 kDa PEG group increasing its molecular weight to about 39 kDa and its half-life to around 42 h. Clinical studies demonstrated that a single dose of Pegfilgrastim was effective and safe compared to multiple filgrastim injections [210].

While PEGylation enhanced the *in vivo* activity of G-CSF by prolonging the half-life time it compromised the *in vitro* activity [211], probably because the PEG molecule is an inactive moiety that sterically hinders the binding to G-CSFR. Additionally, it was shown that Pegfilgrastim induces anti-PEG antibodies that conflicts with the repeated application of Pegfilgrastim [212, 213] and therefore also for other pegylated-therapeutics. Another factor driving the development of long-acting G-CSF alternatives is the substantial cost associated with PEGylated protein therapeutics [103]. A recent illustration Pegfilgrastim alternative is Efbemalenograstim Alfa, which received its initial approval in China in 2023 for the treatment of febrile neutropenia [214]. This therapeutic agent consists of a homodimeric structure, incorporating two G-CSF-Fc molecules linked through covalent disulfide bonds within the Fc domains of the molecule.

Analogous approaches that dependent on G-CSF as the active constituent to create long-acting G-CSFR agonists are inherently limited by both G-CSF's stability and production cost. Hence, substituting G-CSF with highly optimized binding modules such as bv6 (compare Chapter 3) is a significant step towards the development of the next generation G-CSFR agonists. In preliminary work, I hypothesized that bv6's exceptional stability offers an opportunity to engineer G-CSFR agonists with variable sizes through the fusion of multiple binding modules into a single molecule chain. These daisy chains of idealized cytokine mimetics I denote as "chainkines". Previously reported tandem fusions possessed a molecular weight of approximately 29 kDa [87] a characteristic that is likely to lead to a short elimination half-life caused by renal clearance. Conversely, a chain comprising three (C3), four (C4), or five (C5) binding modules would result in a weight of 43 kDa, 58 kDa, or 72 kDa, respectively (Figure 13 A). In general, a molecular weight in the range of 60-70 kDa is considered necessary to prevent glomerular filtration and subsequent elimination via the kidneys [207].

The modularity presented in my preliminary work can be leveraged to engineer the circulating half-life, of a hypothermal-stable, high-yielding and protease resistant G-CSFR agonist. Furthermore, this approach could enable the engineering of G-CSFR ligands capable of clustering more than two receptors, like G-CSF, which is thought to form larger receptor clusters at high ligand concentrations [215, 216]. Consequently, the chainkines would be capable of emulating the effect of higher-order G-CSFR clustering but in a more controlled manner. Unlike G-CSF, chainkines would have the ability to cluster a specified number of receptors equal to or fewer than the number of binding modules within the chain. Importantly, achieving this higher-order clustering would not be limited to high ligand concentrations, as is the case with G-CSF, but would be attainable with a single chainkine molecule.

I designed three distinct chainkines as proof of concept molecules, each incorporating the enhanced G-CSFR binding module bv6 and progressively increasing the number of binding modules. These chainkines are designated as bv6_C3, bv6_C4, and bv6_C5. Thus far, I have demonstrated that these chainkines are well purifiable from *E. coli* with immobilized metal ion affinity chromatography (IMAC) followed by size exclusion chromatography (SEC). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis validated the anticipated molecular weight of the designs (Figure 13 B). They exhibited the expected circular dichroism (CD) spectrum characteristic of alpha-helical proteins, and displayed exceptional thermostability, remaining stable up to 110 °C verified by nano differential scanning fluorimetry (nanoDSF) (Figure 13 C). Moreover, surface plasmon resonance (SPR) experiments conducted against immobilized G-CSFR indicated a sub-nanomolar affinity with exceedingly low dissociation rates. This low dissociation is likely a consequence of heightened avidity effects when interacting with surfaces containing a high content of G-CSFR.

Even more promising, these chainkines exhibited cell proliferation activity comparable to that of commercial rhG-CSF, and the *in vitro* potency rather increased for designs of higher molecular weight (Figure 13 E). This strongly suggests that the utilization of functional units to augment the overall molecule size may represent a superior approach in terms of optimizing activity, stability, and cost-effectiveness compared to employing non-functional units such as PEG.

Furthermore, preliminary findings from a colony forming unit (CFU) assay involving human primary $CD34^+$ hematopoietic stem and progenitor cells obtained from healthy donors suggest that bv6_C5 can indeed promote their proliferation and differentiation (Figure 13 F). As noted with all other designs elaborated upon in Chapter 3, bv6_C5 failed to stimulate the cells in the absence of IL-3 and stem cell factor (SCF) in the growth conditions. In conjunction with the observation that G-CSF elicits the most robust overall stimulation, this observation may suggest that G-CSF activates additional signaling pathways, which aligns with the transcriptomic analysis detailed in Chapter 3.

Notably, G-CSF's capacity to independently activate cells and trigger additional pathways can be a drawback for designing therapeutics with precise activity, potentially contributing to observed adverse effects associated with G-CSF. Another interesting observation is that in this preliminary experiment the bv6_C5 exhibited a higher relative proportion of granulocytes (G) over granulocyte–macrophage (GM) CFUs than G-CSF (Figure 13 F). One interpretation of this finding could be attributed to the heightened avidity effects of chainkines, which likely enable higher local concentrations on cells expressing elevated levels of G-CSFR, thereby promoting more specific differentiation towards granulocyte CFUs.

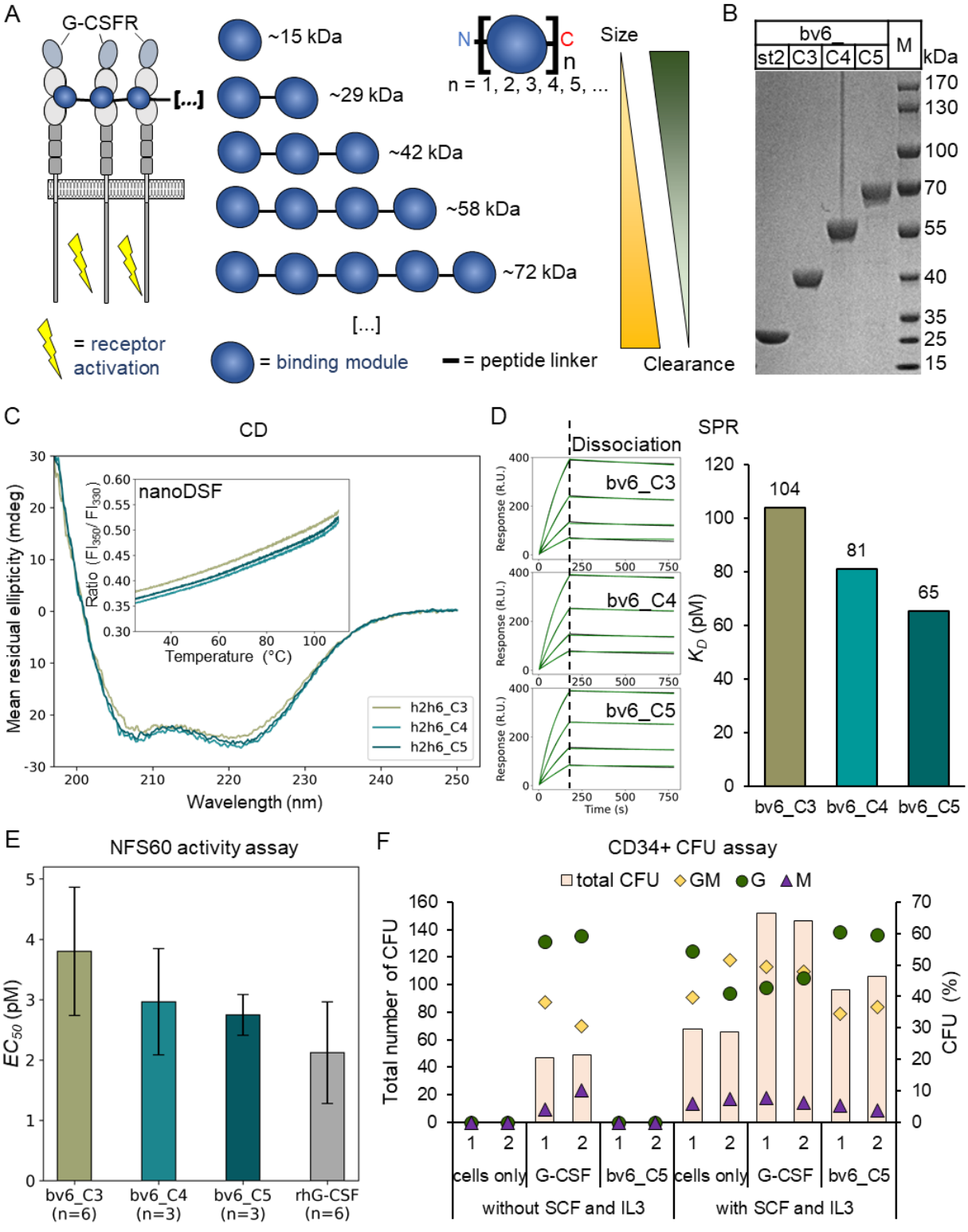



***Figure 13:*** ***(A)*** *Proposed concept of chainkines.* ***(B)*** *SDS-PAGE was performed for the chainkine designs (bv6_C3, bv6_C4, and bv6_C5) and the established short tandem fusion of bv6 (bv6_st2).* ***(C)*** *The secondary structure and stability were examined CD and nanoDSF.* ***(D)*** *SPR measurements (black line) of a two-fold dilution series starting at 10 nM were performed for each chainkine design. The apparent equilibrium dissociation constant ($K_D$) was*

*obtained by a corresponding fit (green line) and are represented in the graph at the right-hand side.* ***(E)*** *G-CSF responsive cell line (NFS-60) proliferation assays were performed to determine the half effective concentration ($EC_{50}$) of the generated chainkines in comparison to rhG-CSF. Shown is the mean and standard deviation from at least 3 independent experiments.* ***(F)*** *Preliminary data of one CFU assay of human primary $CD34^+$ hematopoietic stem and progenitor cells from healthy donors performed once in parallel duplicates for each condition. Indicated are the observed total number of colony-forming units (CFU) and the relative proportion of granulocyte–macrophage (GM), granulocytes (G), or monocytes (M) within the conditions. The cells were treated with or without 50 ng/mL SCF and 20 ng/mL IL-3. bv6_C5 underwent testing at 100 ng/mL in simultaneously with one of the biological replicates featured in Chapter 3 of the $CD34^+$ experiment. Therefore, the corresponding control data was integrated in F), specifically encompassing the conditions of untreated cells (cells only) and cells treated with 20 ng/mL G-CSF. All experiments from C) to F) were performed the same way as described in Chapter 3.*

At present, it remains an unanswered question whether the designed constructs indeed manifest the desired attributes, particularly in their capacity to enhance *in vivo* pharmacokinetics. Consequently, further investigations are needed that encompass not only evaluations of *in vivo* activity over defined time intervals and the clearance from the organism but also a thorough assessment of their overall biodistribution. Nevertheless, the successful generation of the presented chainkine G-CSFR agonists serves as a promising foundation for the future development of novel long-acting cytokine designs featuring advanced biophysical and pharmacokinetic properties.

# 6 Conclusion

The complex interplay of factors in cytokine receptor signaling, such as receptor composition, affinity, and geometry, poses a formidable challenge, yet overcoming it promises great advancements in therapeutics applications. This research represents a pioneering effort demonstrating the tunable nature of the clinically relevant G-CSFR.

In order to tackle this challenge, I first enhanced a small and stable epitope-targeted binding module using *in silico* and *in vitro* high-throughput screening methods. This optimization pipeline can be further translated to other designed binding modules. Subsequently, I illustrated the potential of these enhanced binding modules in producing G-CSFR agonists designs that reach G-CSF activity in cell-based assays. Moreover, I designed ligands that could fine-tune G-CSFR activity by adjusting the receptor geometry. These designs displayed precise control over intracellular signaling, transcriptomic activity, and primary stem cell differentiation, as well as *in vivo* activity in zebrafish and mouse models. Additionally, I demonstrated that the design of stringently monomeric binders based on the enhanced modules enabled the generation of potent G-CSFR antagonists with low nanomolar inhibitory activity.

Furthermore, aside from agonizing or antagonizing receptor activity, these designed ligands exhibit significantly heightened stability and enhanced production efficiency when compared to commercially available G-CSF and antibodies, respectively. These attributes hold substantial importance for their prospective advancement as protein therapeutics, as they have the potential not only to render them economically attractive but also to streamline their applicability concerning simplified storage, application, deimmunization, and modulation of pharmacokinetic parameters. Hence, the array of G-CSFR binding modules, each possessing differing affinities, the activity-modulating designs, and the G-CSFR antagonist designs, will serve as a cornerstone for both the systematical investigation of G-CSFR activity modulation and the development of enhanced protein therapeutics.

## References


1. Vilček, J. and M. Feldmann, *Historical review: cytokines as therapeutics and targets of therapeutics.* Trends in pharmacological sciences, 2004. **25**(4): p. 201-209.
2. Dinarello, C.A., *Historical insights into cytokines.* European journal of immunology, 2007. **37**(S1): p. S34-S45.
3. Welte, K., *et al.*, *Filgrastim (r-metHuG-CSF): The First 10 Years.* Blood The Journal of The American Society of Hematology, 1996. **88**(6): p. 1907–1929.
4. *G-CSFPEG-G-CSF Market Set to Reach at USD 5,685.40 Million by 2030, Size, Share, Demand, Future Growth and Competitors Outlook*. Tue. 26 Sep 2023]; Available from: https://www.bloomberg.com/press-releases/2023-01-11/g-csfpeg-g-csf-market-set-to-reach-at-usd-5-685-40-million-by-2030-size-share-demand-future-growth-and-competitors-outlook.
5. Rochman, Y., R. Spolski, and W.J. Leonard, *New insights into the regulation of T cells by gamma(c) family cytokines.* Nature reviews. Immunology, 2009. **9**(7): p. 480–490.
6. Pires, I.S., P.T. Hammond, and D.J. Irvine, *Engineering Strategies for Immunomodulatory Cytokine Therapies - Challenges and Clinical Progress.* Advanced therapeutics, 2021. **4**(8).
7. Cohen, S., P.E. Bigazzi, and T. Yoshida, *Similarities of T cell function in cell-mediated immunity and antibody production.* Cellular immunology, 1974. **12**(1): p. 150-159.
8. Chauhan, P., *et al.*, *A primer on cytokines.* Cytokine, 2021. **145**: p. 155458.
9. Gulati, K., *et al.*, *Cytokines and their role in health and disease: a brief overview.* Moj Immunol, 2016. **4**(2): p. 00121.
10. Saxton, R.A., C.R. Glassman, and K.C. Garcia, *Emerging principles of cytokine pharmacology and therapeutics.* Nature reviews. Drug discovery, 2022.
11. Turner, M.D., *et al.*, *Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease.* Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2014. **1843**(11): p. 2563-2582.
12. Toomer, K.H. and T.R. Malek, *Cytokine signaling in the development and homeostasis of regulatory T cells.* Cold Spring Harbor perspectives in biology, 2018. **10**(3): p. a028597.

13. Zhang, J.-M. and J. An, *Cytokines, inflammation and pain.* International anesthesiology clinics, 2007. **45**(2): p. 27.
14. Yoshimura, A., T. Naka, and M. Kubo, *SOCS proteins, cytokine signalling and immune regulation.* Nature Reviews Immunology, 2007. **7**(6): p. 454-465.
15. Tamada, T., *et al.*, *Homodimeric cross-over structure of the human granulocyte colony-stimulating factor (GCSF) receptor signaling complex.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(9): p. 3135–3140.
16. Broughton, S.E., *et al.*, *Dual mechanism of interleukin-3 receptor blockade by an anti-cancer antibody.* Cell reports, 2014. **8**(2): p. 410-419.
17. Cowland, J.B. and N. Borregaard, *Granulopoiesis and granules of human neutrophils.* Immunological reviews, 2016. **273**(1): p. 11-28.
18. Othman, A., M. Sekheri, and J.G. Filep, *Roles of neutrophil granule proteins in orchestrating inflammation and immunity.* The FEBS journal, 2022. **289**(14): p. 3932-3953.
19. Voynow, J.A. and M. Shinbashi, *Neutrophil elastase and chronic lung disease.* Biomolecules, 2021. **11**(8): p. 1065.
20. Radin, M.J. and M.L. Wellman, *Granulopoiesis.* Schalm's Veterinary Hematology, 2022: p. 323-332.
21. Yvan-Charvet, L. and L.G. Ng, *Granulopoiesis and neutrophil homeostasis: a metabolic, daily balancing act.* Trends in immunology, 2019. **40**(7): p. 598-612.
22. Lee, M., S.Y. Lee, and Y.-S. Bae, *Emerging roles of neutrophils in immune homeostasis.* BMB reports, 2022. **55**(10): p. 473.
23. Lin, A. and K. Loré, *Granulocytes: new members of the antigen-presenting cell family.* Frontiers in immunology, 2017. **8**: p. 1781.
24. Fiedler, K. and C. Brunner, *The role of transcription factors in the guidance of granulopoiesis.* American journal of blood research, 2012. **2**(1): p. 57.
25. Bendall, L.J. and K.F. Bradstock, *G-CSF: From granulopoietic stimulant to bone marrow stem cell mobilizing agent.* Cytokine & growth factor reviews, 2014. **25**(4): p. 355–367.
26. Greenbaum, A. and D. Link, *Mechanisms of G-CSF-mediated hematopoietic stem and progenitor mobilization.* Leukemia, 2011. **25**(2): p. 211-217.
27. Skokowa, J., *et al.*, *Severe congenital neutropenias.* Nature reviews. Disease primers, 2017. **3**: p. 17032.

28. Paudel, S., *et al.*, *Regulation of emergency granulopoiesis during infection.* Frontiers in immunology, 2022. **13**: p. 961601.
29. Nagai, Y., *et al.*, *Toll-like receptors on hematopoietic progenitor cells stimulate innate immune system replenishment.* Immunity, 2006. **24**(6): p. 801-812.
30. Pauksen, K., *et al.*, *Serum levels of granulocyte-colony stimulating factor (G-CSF) in bacterial and viral infections, and in atypical pneumonia.* British journal of haematology, 1994. **88**(2): p. 256-260.
31. Mine, S., *et al.*, *Thermodynamic analysis of the activation mechanism of the GCSF receptor induced by ligand binding.* Biochemistry, 2004. **43**(9): p. 2458–2464.
32. Aritomi, M., *et al.*, *Atomic structure of the GCSF-receptor complex showing a new cytokine-receptor recognition scheme.* Nature, 1999. **401**(6754): p. 713–717.
33. Fukunaga, R., E. Ishizaka-Ikeda, and S. Nagata, *Growth and differentiation signals mediated by different regions in the cytoplasmic domain of granulocyte colony-stimulating factor receptor.* Cell, 1993. **74**(6): p. 1079-1087.
34. de Koning, J.P., *et al.*, *Specific involvement of tyrosine 764 of human granulocyte colony-stimulating factor receptor in signal transduction mediated by p145/Shc/GRB2 or p90/GRB2 complexes.* 1996.
35. Dong, F., *et al.*, *Distinct cytoplasmic regions of the human granulocyte colony-stimulating factor receptor involved in induction of proliferation and maturation.* Molecular and cellular biology, 1993.
36. Shao, H., *et al.*, *Unique structural determinants for Stat3 recruitment and activation by the granulocyte colony-stimulating factor receptor at phosphotyrosine ligands 704 and 744.* The Journal of Immunology, 2006. **176**(5): p. 2933-2941.
37. de Koning, J.P., *et al.*, *Proliferation signaling and activation of Shc, p21Ras, and Myc via tyrosine 764 of human granulocyte colony-stimulating factor receptor.* Blood, The Journal of the American Society of Hematology, 1998. **91**(6): p. 1924-1933.
38. Jumper, J., *et al.*, *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.

39. Varadi, M., *et al.*, *AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models.* Nucleic acids research, 2022. **50**(D1): p. D439-D444.
40. Dwivedi, P. and K.D. Greis, *Granulocyte colony-stimulating factor receptor signaling in severe congenital neutropenia, chronic neutrophilic leukemia, and related malignancies.* Experimental hematology, 2017. **46**: p. 9–20.
41. Morris, R., N.J. Kershaw, and J.J. Babon, *The molecular details of cytokine signaling via the JAK/STAT pathway.* Protein science : a publication of the Protein Society, 2018. **27**(12): p. 1984–2009.
42. Schindler, C. and J.E. Darnell, *Transcriptional responses to polypeptide ligands: the JAK-STAT pathway.* Annual Review of Biochemistry, 1995. **64**(1): p. 621–651.
43. Darnell, J.E., *STATs and gene regulation.* Science, 1997. **277**(5332): p. 1630–1635.
44. Basu, S., A. Dunn, and A. Ward, *G-CSF: function and modes of action (Review).* International Journal of Molecular Medicine, 2002.
45. Roberts, A.W., *G-CSF: a key regulator of neutrophil production, but that's not all!* Growth factors (Chur, Switzerland), 2005. **23**(1): p. 33–41.
46. Touw, I.P. and G.-J.M. van de Geijn, *Granulocyte colony-stimulating factor and its receptor in normal myeloid cell development, leukemia and related blood cell disorders.* Frontiers in Bioscience, 2007. **12**(1): p. 800–815.
47. Dong, F., *et al.*, *Stimulation of Stat5 by granulocyte colony-stimulating factor (G-CSF) is modulated by two distinct cytoplasmic regions of the G-CSF receptor.* The Journal of Immunology, 1998. **161**(12): p. 6503-6509.
48. Park, S.D., *et al.*, *A review of granulocyte colony-stimulating factor receptor signaling and regulation with implications for cancer.* Frontiers in Oncology, 2022. **12**.
49. Ward, A.C., *et al.*, *Tyrosine-dependent and-independent mechanisms of STAT3 activation by the human granulocyte colony-stimulating factor (G-CSF) receptor are differentially utilized depending on G-CSF concentration.* Blood, The Journal of the American Society of Hematology, 1999. **93**(1): p. 113-124.
50. Kubo, M., T. Hanada, and A. Yoshimura, *Suppressors of cytokine signaling and immunity.* Nature immunology, 2003. **4**(12): p. 1169-1176.

51. Babon, J.J. and N.A. Nicola, *The biology and mechanism of action of suppressor of cytokine signaling 3.* Growth factors, 2012. **30**(4): p. 207-219.
52. Welte, K., *et al.*, *Purification and biochemical characterization of human pluripotent hematopoietic colony-stimulating factor.* Proceedings of the National Academy of Sciences of the United States of America, 1985. **82**(5): p. 1526–1530.
53. Kinstler, O.B., *et al.*, *Characterization and stability of N-terminally PEGylated rhG-CSF.* Pharmaceutical research, 1996. **13**(7): p. 996–1002.
54. Piedmonte, D.M. and M.J. Treuheit, *Formulation of Neulasta (pegfilgrastim).* Advanced drug delivery reviews, 2008. **60**(1): p. 50–58.
55. Bendall, L.J. and K.F. Bradstock, *G-CSF: From granulopoietic stimulant to bone marrow stem cell mobilizing agent.* Cytokine & growth factor reviews, 2014. **25**(4): p. 355-367.
56. Welte, K., C. Zeidler, and D.C. Dale, *Severe congenital neutropenia.* Seminars in hematology, 2006. **43**(3): p. 189–195.
57. Liongue, C. and A.C. Ward, *Granulocyte colony-stimulating factor receptor mutations in myeloid malignancy.* Frontiers in oncology, 2014. **4**: p. 93.
58. Nara, N., *et al.*, *Granulocyte colony-stimulating factor-dependent growth of an acute myeloblastic leukemia cell line.* Japanese journal of cancer research, 1990. **81**(6-7): p. 625-631.
59. Morris, K., *et al.*, *G-CSF and G-CSFR are highly expressed in human gastric and colon cancers and promote carcinoma cell proliferation and migration.* British journal of cancer, 2014. **110**(5): p. 1211-1220.
60. Liu, L., *et al.*, *The role of granulocyte colony-stimulating factor in breast cancer development: a review.* Molecular medicine reports, 2020. **21**(5): p. 2019-2029.
61. Mouchemore, K.A., R.L. Anderson, and J.A. Hamilton, *Neutrophils, G-CSF and their contribution to breast cancer metastasis.* The FEBS journal, 2018. **285**(4): p. 665-679.
62. Hamilton, J.A., A.D. Cook, and P.P. Tak, *Anti-colony-stimulating factor therapies for inflammatory and autoimmune diseases.* Nature reviews Drug discovery, 2017. **16**(1): p. 53-70.
63. Martin, K.R., *et al. G-CSF–A double edge sword in neutrophil mediated immunity*. in *Seminars in Immunology*. 2021. Elsevier.

64. Campbell, I.K., *et al.*, *Therapeutic targeting of the G-CSF receptor reduces neutrophil trafficking and joint inflammation in antibody-mediated inflammatory arthritis.* The Journal of Immunology, 2016. **197**(11): p. 4392-4402.
65. Cornish, A.L., *et al.*, *G-CSF and GM-CSF as therapeutic targets in rheumatoid arthritis.* Nature Reviews Rheumatology, 2009. **5**(10): p. 554-559.
66. Eyles, J.L., *et al.*, *A key role for G-CSF–induced neutrophil production and trafficking during inflammatory arthritis.* Blood, The Journal of the American Society of Hematology, 2008. **112**(13): p. 5193-5201.
67. Braun, B., *et al.*, *Expression of G-CSF and GM-CSF in human meningiomas correlates with increased tumor proliferation and vascularization.* Journal of neuro-oncology, 2004. **68**(2): p. 131-140.
68. Wallner, S., *et al.*, *The granulocyte-colony stimulating factor has a dual role in neuronal and vascular plasticity.* Frontiers in cell and developmental biology, 2015. **3**: p. 48.
69. Carvalho, T.T., *et al.*, *Granulocyte-colony stimulating factor (G-CSF)-induced mechanical hyperalgesia in mice: Role for peripheral TNFα, IL-1β and IL-10.* European journal of pharmacology, 2015. **749**: p. 62-72.
70. Schweizerhof, M., *et al.*, *Hematopoietic colony–stimulating factors mediate tumor-nerve interactions and bone cancer pain.* Nature medicine, 2009. **15**(7): p. 802-807.
71. Stösser, S., M. Schweizerhof, and R. Kuner, *Hematopoietic colony-stimulating factors: new players in tumor–nerve interactions.* Journal of molecular medicine, 2011. **89**(4): p. 321-329.
72. Brannigan, J.A. and A.J. Wilkinson, *Protein engineering 20 years on.* Nature Reviews Molecular Cell Biology, 2002. **3**(12): p. 964-970.
73. Woolfson, D.N., *A Brief History of De Novo Protein Design: Minimal, Rational, and Computational.* Journal of molecular biology, 2021. **433**(20): p. 167160.
74. Yue, K. and K.A. Dill, *Inverse protein folding problem: designing polymer sequences.* Proceedings of the National Academy of Sciences, 1992. **89**(9): p. 4163-4167.
75. Cossio, P., *et al.*, *Exploring the universe of protein structures beyond the Protein Data Bank.* PLoS computational biology, 2010. **6**(11): p. e1000957.

76. Woolfson, D.N., *et al.*, *De novo protein design: How do we expand into the universe of possible protein structures?* Current Opinion in Structural Biology, 2015. **33**: p. 16–26.
77. Taylor, W.R., *et al.*, *Probing the "dark matter" of protein fold space.* Structure, 2009. **17**(9): p. 1244-1252.
78. Wang, Y., *et al.*, *Directed evolution: methodologies and applications.* Chemical reviews, 2021. **121**(20): p. 12384-12444.
79. *The Nobel Prize in Chemistry 2018*. Tue. 26 Sep 2023]; Available from: https://www.nobelprize.org/prizes/chemistry/2018/summary/.
80. DeGrado, W.F., Z.R. Wasserman, and J.D. Lear, *Protein design, a minimalist approach.* Science, 1989. **243**(4891): p. 622-628.
81. Hill, R.B., *et al.*, *De novo design of helical bundles as models for understanding protein folding and function.* Accounts of chemical research, 2000. **33**(11): p. 745-754.
82. Street, A.G. and S.L. Mayo, *Computational protein design.* Structure, 1999. **7**(5): p. R105-R109.
83. Pan, X. and T. Kortemme, *Recent advances in de novo protein design: Principles, methods, and applications.* Journal of Biological Chemistry, 2021. **296**.
84. Watson, J.L., *et al.*, *De novo design of protein structure and function with RFdiffusion.* Nature, 2023: p. 1-3.
85. Dauparas, J., *et al.*, *Robust deep learning–based protein sequence design using ProteinMPNN.* Science, 2022. **378**(6615): p. 49-56.
86. Leader, B., Q.J. Baca, and D.E. Golan, *Protein therapeutics: a summary and pharmacological classification.* Nature reviews Drug discovery, 2008. **7**(1): p. 21-39.
87. Skokowa, J., *et al.*, *A topological refactoring design strategy yields highly stable granulopoietic proteins.* Nature communications, 2022. **13**(1): p. 2948.
88. Khoryati, L., *et al.*, *An IL-2 mutein engineered to promote expansion of regulatory T cells arrests ongoing autoimmunity in mice.* Science Immunology, 2020. **5**(50).
89. Sahin, D., *et al.*, *An IL-2-grafted antibody immunotherapy with potent efficacy against metastatic cancer.* Nature Communications, 2020. **11**(1): p. 6440.

90. Mohan, K., *et al.*, *Topological control of cytokine receptor signaling induces differential effects in hematopoiesis.* Science, 2019. **364**(6442).
91. Silva, D.-A., *et al.*, *De novo design of potent and selective mimics of IL-2 and IL-15.* Nature, 2019. **565**(7738): p. 186–191.
92. Campeotto, I., *et al.*, *One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen.* Proceedings of the National Academy of Sciences, 2017. **114**(5): p. 998-1002.
93. Correia, B.E., *et al.*, *Proof of principle for epitope-focused vaccine design.* Nature, 2014. **507**(7491): p. 201-206.
94. Marcandalli, J., *et al.*, *Induction of potent neutralizing antibody responses by a designed protein nanoparticle vaccine for respiratory syncytial virus.* Cell, 2019. **176**(6): p. 1420-1431. e17.
95. Polizzi, N.F., *et al.*, *De novo design of a hyperstable non-natural protein–ligand complex with sub-Å accuracy.* Nature chemistry, 2017. **9**(12): p. 1157-1164.
96. Cao, L., *et al.*, *De novo design of picomolar SARS-CoV-2 miniprotein inhibitors.* Science (New York, N.Y.), 2020. **370**(6515): p. 426–431.
97. Zheng, X., *et al.*, *The use of supercytokines, immunocytokines, engager cytokines, and other synthetic cytokines in immunotherapy.* Cellular & molecular immunology, 2022. **19**(2): p. 192-209.
98. Deckers, J., *et al.*, *Engineering cytokine therapeutics.* Nature Reviews Bioengineering, 2023. **1**(4): p. 286-303.
99. Aung, T., *et al.*, *Recent and future perspectives on engineering interferons and other cytokines as therapeutics.* Trends in Biochemical Sciences, 2023.
100. Quijano-Rubio, A., *et al.*, *A split, conditionally active mimetic of IL-2 reduces the toxicity of systemic cytokine therapy.* Nature biotechnology, 2023. **41**(4): p. 532-540.
101. Haddadzadegan, S., F. Dorkoosh, and A. Bernkop-Schnürch, *Oral delivery of therapeutic peptides and proteins: Technology landscape of lipid-based nanocarriers.* Advanced Drug Delivery Reviews, 2022. **182**: p. 114097.
102. De Groot, A., P. Knopp, and W. Martin, *De-immunization of therapeutic proteins by T-cell epitope modification.* Developments in biologicals, 2005. **122**: p. 171-194.

103. Theyab, A., *et al.*, *New insight into strategies used to develop long-acting G-CSF biologics for neutropenia therapy.* Frontiers in Oncology, 2023. **12**: p. 1026377.
104. Alvarez, B.H. and *et al.*, *Design of novel granulopoietic proteins by topological rescaffolding.* PLOS Biology, 2020.
105. Bonanno, J.B., *et al.*, *New York-Structural GenomiX Research Consortium (NYSGXRC): a large scale center for the protein structure initiative.* Journal of structural and functional genomics, 2005. **6**: p. 225-232.
106. Boyken, S.E., *et al.*, *De novo design of protein homo-oligomers with modular hydrogen-bond network–mediated specificity.* Science, 2016. **352**(6286): p. 680-687.
107. Olsen, M., B. Iverson, and G. Georgiou, *High-throughput screening of enzyme libraries.* Current opinion in biotechnology, 2000. **11**(4): p. 331-337.
108. Russell, A.J. and A.R. Fersht, *Rational modification of enzyme catalysis by engineering surface charge.* Nature, 1987. **328**(6130): p. 496-500.
109. Arnold, F.H., *Design by directed evolution.* Accounts of chemical research, 1998. **31**(3): p. 125-131.
110. Packer, M.S. and D.R. Liu, *Methods for the directed evolution of proteins.* Nature Reviews Genetics, 2015. **16**(7): p. 379-394.
111. Zeymer, C. and D. Hilvert, *Directed Evolution of Protein Catalysts.* Annual review of biochemistry, 2018. **87**: p. 131–157.
112. Eijsink, V.G., *et al.*, *Directed evolution of enzyme stability.* Biomolecular engineering, 2005. **22**(1-3): p. 21-30.
113. Wang, T., *et al.*, *Continuous directed evolution of proteins with improved soluble expression.* Nature chemical biology, 2018. **14**(10): p. 972-980.
114. Boder, E.T., K.S. Midelfort, and K.D. Wittrup, *Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity.* Proceedings of the National Academy of Sciences, 2000. **97**(20): p. 10701-10705.
115. Hawkins, R.E., S.J. Russell, and G. Winter, *Selection of phage antibodies by binding affinity.* Journal of molecular biology, 1992. **226**(3): p. 889–896.
116. Leemhuis, H., *et al.*, *New genotype–phenotype linkages for directed evolution of functional proteins.* Current Opinion in Structural Biology, 2005. **15**(4): p. 472-478.

117. Mena, M.A. and P.S. Daugherty, *Automated design of degenerate codon libraries.* Protein Engineering Design and Selection, 2005. **18**(12): p. 559-561.
118. Patrick, W.M., A.E. Firth, and J.M. Blackburn, *User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries.* Protein engineering, 2003. **16**(6): p. 451-457.
119. Bosley, A.D. and M. Ostermeier, *Mathematical expressions useful in the construction, description and evaluation of protein libraries.* Biomolecular engineering, 2005. **22**(1-3): p. 57-61.
120. McCullum, E.O., *et al.*, *Random mutagenesis by error-prone PCR.* In Vitro Mutagenesis Protocols: Third Edition, 2010: p. 103-109.
121. Chronopoulou, E.G. and N.E. Labrou, *Site-saturation mutagenesis: A powerful tool for structure-based design of combinatorial mutation libraries.* Current Protocols in Protein Science, 2011. **63**(1): p. 26.6. 1-26.6. 10.
122. Jacobs, T.M., *et al.*, *SwiftLib: rapid degenerate-codon-library optimization through dynamic programming.* Nucleic acids research, 2015. **43**(5): p. e34-e34.
123. Maksymenko, K., *et al.*, *The design of functional proteins using tensorized energy calculations.* Cell Reports Methods, 2023.
124. Galán, A., *et al.*, *Library-based display technologies: Where do we stand?* Molecular BioSystems, 2016. **12**(8): p. 2342–2358.
125. Salema, V. and L.Á. Fernández, *Escherichia coli surface display for the selection of nanobodies.* Microbial Biotechnology, 2017. **10**(6): p. 1468–1484.
126. Sidhu, S.S., *Phage display in pharmaceutical biotechnology.* Current Opinion in Biotechnology, 2000. **11**(6): p. 610–616.
127. Zahnd, C., P. Amstutz, and A. Plückthun, *Ribosome display: Selecting and evolving proteins in vitro that specifically bind to a target.* Nature Methods, 2007. **4**(3): p. 269–279.
128. Salema, V., *et al.*, *Selection of single domain antibodies from immune libraries displayed on the surface of E. coli cells with two β-domains of opposite topologies.* PloS one, 2013. **8**(9): p. e75126.
129. Löfblom, J., *Bacterial display in combinatorial protein engineering.* Biotechnology journal, 2011. **6**(9): p. 1115–1129.

130. Spangler, J.B., *et al.*, *Insights into cytokine-receptor interactions from cytokine engineering.* Annual review of immunology, 2015. **33**: p. 139–167.
131. Moraga, I., *et al.*, *Synthekines are surrogate cytokine and growth factor agonists that compel signaling through non-natural receptor dimers.* Elife, 2017. **6**: p. e22882.
132. Yen, M., *et al.*, *Facile discovery of surrogate cytokine agonists.* Cell, 2022. **185**(8): p. 1414-1430. e19.
133. Kallen, K.-J., *et al.*, *Receptor recognition sites of cytokines are organized as exchangeable modules: transfer of the leukemia inhibitory factor receptor-binding site from ciliary neurotrophic factor to interleukin-6.* Journal of Biological Chemistry, 1999. **274**(17): p. 11859-11867.
134. Findeisen, M., *et al.*, *Treatment of type 2 diabetes with the designer cytokine IC7Fc.* Nature, 2019. **574**(7776): p. 63-68.
135. Levin, A.M., *et al.*, *Exploiting a natural conformational switch to engineer an interleukin-2 'superkine'.* Nature, 2012. **484**(7395): p. 529–533.
136. Saxton, R.A., *et al.*, *The tissue protective functions of interleukin-22 can be decoupled from pro-inflammatory actions through structure-based design.* Immunity, 2021. **54**(4): p. 660-672. e9.
137. Gorby, C., *et al.*, *Engineered IL-10 variants elicit potent immunomodulatory effects at low ligand doses.* Science signaling, 2020. **13**(649): p. eabc0653.
138. Mitra, S., *et al.*, *Interleukin-2 activity can be fine tuned with engineered receptor signaling clamps.* Immunity, 2015. **42**(5): p. 826-838.
139. Ozaki, K. and W.J. Leonard, *Cytokine and cytokine receptor pleiotropy and redundancy.* Journal of Biological Chemistry, 2002. **277**(33): p. 29355-29358.
140. Peterson, L.B., *et al.*, *A long-lived IL-2 mutein that selectively activates and expands regulatory T cells as a therapy for autoimmune disease.* Journal of autoimmunity, 2018. **95**: p. 1-14.
141. Glassman, C.R., *et al.*, *Calibration of cell-intrinsic interleukin-2 response thresholds guides design of a regulatory T cell biased agonist.* Elife, 2021. **10**: p. e65777.
142. Mantovani, A., *et al.*, *Decoy receptors: a strategy to regulate inflammatory cytokines and chemokines.* Trends in immunology, 2001. **22**(6): p. 328-336.
143. Zhou, T., *et al.*, *IL-18BP is a secreted immune checkpoint and barrier to IL-18 immunotherapy.* Nature, 2020. **583**(7817): p. 609-614.

144. Gonnord, P., C.M. Blouin, and C. Lamaze. *Membrane trafficking and signaling: two sides of the same coin*. in *Seminars in cell & developmental biology*. 2012. Elsevier.
145. Cendrowski, J., A. Mamińska, and M. Miaczynska, *Endocytic regulation of cytokine receptor signaling.* Cytokine & growth factor reviews, 2016. **32**: p. 63-73.
146. Kim, A.R., *et al.*, *Functional selectivity in cytokine signaling revealed through a pathogenic EPO mutation.* Cell, 2017. **168**(6): p. 1053-1064. e15.
147. Sato, N. and A. Miyajima, *Multimeric cytokine receptors: common versus specific functions.* Current opinion in cell biology, 1994. **6**(2): p. 174-179.
148. O'Shea, J.J. and R. Plenge, *JAK and STAT signaling molecules in immunoregulation and immune-mediated disease.* Immunity, 2012. **36**(4): p. 542–550.
149. Palande, K., *et al.*, *Scratching the surface: signaling and routing dynamics of the CSF3 receptor.* Frontiers in bioscience (Landmark edition), 2013. **18**(1): p. 91–105.
150. Xie, J., *et al.*, *Autocrine signaling based selection of combinatorial antibodies that transdifferentiate human stem cells.* Proceedings of the National Academy of Sciences, 2013. **110**(20): p. 8099-8104.
151. Boxer, L.A., *How to approach neutropenia.* Hematology, 2012. **2012**(1): p. 174–182.
152. Moraga, I., *et al.*, *Tuning cytokine receptor signaling by re-orienting dimer geometry with surrogate ligands.* Cell, 2015. **160**(6): p. 1196–1208.
153. Nakano, K., *et al.*, *Effective screening method of agonistic diabodies based on autocrine growth.* Journal of immunological methods, 2009. **347**(1-2): p. 31-35.
154. Muyldermans, S., *A guide to: generation and design of nanobodies.* The FEBS journal, 2021. **288**(7): p. 2084-2102.
155. Scalzo-Inguanti, K., *et al.*, *A neutralizing anti–G-CSFR antibody blocks G-CSF–induced neutrophilia without inducing neutropenia in nonhuman primates.* Journal of leukocyte biology, 2017. **102**(2): p. 537-549.
156. Scalzo-Inguanti, K., *et al.*, *164: CSL324, a humanised anti G-CSFR antibody, can inhibit neutrophil migration while not impaction on neutrophil number or effector functions.* Cytokine, 2014. **70**(1): p. 68.

157. Goldberg, G.L., *et al.*, *G-CSF and neutrophils are nonredundant mediators of murine experimental autoimmune uveoretinitis.* The American Journal of Pathology, 2016. **186**(1): p. 172-184.
158. Wang, H., *et al.*, *G-CSFR antagonism reduces neutrophilic inflammation during pneumococcal and influenza respiratory infections without compromising clearance.* Scientific Reports, 2019. **9**(1): p. 1-12.
159. Papanicolaou, A., *et al.*, *Novel therapies for pneumonia-associated severe asthma phenotypes.* Trends in Molecular Medicine, 2020. **26**(11): p. 1047-1058.
160. McRae, J.L., *et al.*, *Blockade of the G-CSF Receptor Is Protective in a Mouse Model of Renal Ischemia–Reperfusion Injury.* The Journal of Immunology, 2020. **205**(5): p. 1433-1440.
161. Carter, P.J., *Potent antibody therapeutics by design.* Nature reviews immunology, 2006. **6**(5): p. 343-357.
162. Strebhardt, K. and A. Ullrich, *Paul Ehrlich's magic bullet concept: 100 years of progress.* Nature Reviews Cancer, 2008. **8**(6): p. 473-480.
163. Simeon, R. and Z. Chen, *In vitro-engineered non-antibody protein therapeutics.* Protein & cell, 2018. **9**(1): p. 3-14.
164. Qing, R., *et al.*, *Protein design: From the aspect of water solubility and stability.* Chemical Reviews, 2022. **122**(18): p. 14085-14179.
165. Gebauer, M. and A. Skerra, *Anticalins: small engineered binding proteins based on the lipocalin scaffold.* Methods in enzymology, 2012. **503**: p. 157-188.
166. Richter, A., E. Eggenstein, and A. Skerra, *Anticalins: exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins.* FEBS letters, 2014. **588**(2): p. 213-218.
167. Justino, C.I., A.C. Duarte, and T.A. Rocha-Santos, *Analytical applications of affibodies.* TrAC Trends in Analytical Chemistry, 2015. **65**: p. 73-82.
168. Plückthun, A., *Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy.* Annual review of pharmacology and toxicology, 2015. **55**: p. 489-511.
169. Kunert, R. and D. Reinhart, *Advances in recombinant antibody manufacturing.* Applied microbiology and biotechnology, 2016. **100**: p. 3451-3461.
170. Kelley, B. *Industrialization of mAb production technology: the bioprocessing industry at a crossroads*. in *MAbs*. 2009. Taylor & Francis.

171. Sparrow, E., *et al.*, *Therapeutic antibodies for infectious diseases.* Bulletin of the World Health Organization, 2017. **95**(3): p. 235.
172. Zahnd, C., *et al.*, *Efficient Tumor Targeting with High-Affinity Designed Ankyrin Repeat Proteins: Effects of Affinity and Molecular SizeInfluence of DARPin Affinity and Size on Tumor Targeting.* Cancer research, 2010. **70**(4): p. 1595-1605.
173. Chauhan, V.P., *et al.*, *Delivery of molecular and nanoscale medicine to tumors: transport barriers and strategies.* Annu Rev Chem Biomol Eng, 2011. **2**(1): p. 281-298.
174. Shah, D.K. and A.M. Betts. *Antibody biodistribution coefficients: inferring tissue concentrations of monoclonal antibodies based on the plasma concentrations in several preclinical species and human*. in *MAbs*. 2013. Taylor & Francis.
175. Manning, M.C., K. Patel, and R.T. Borchardt, *Stability of protein pharmaceuticals.* Pharmaceutical research, 1989. **6**: p. 903-918.
176. Manning, M.C., *et al.*, *Stability of protein pharmaceuticals: an update.* Pharmaceutical research, 2010. **27**: p. 544-575.
177. Rosenberg, A.S., *Effects of protein aggregates: an immunologic perspective.* The AAPS journal, 2006. **8**: p. E501-E507.
178. Krishnan, S., *et al.*, *Aggregation of granulocyte colony stimulating factor under physiological conditions: characterization and thermodynamic inhibition.* Biochemistry, 2002. **41**(20): p. 6422-6431.
179. Remmele, R.L., *et al.*, *Minimization of recombinant human Flt3 ligand aggregation at the T m plateau: a matter of thermal reversibility.* Biochemistry, 1999. **38**(16): p. 5241-5247.
180. Woodley, J.F., *Enzymatic barriers for GI peptide and protein delivery.* Critical reviews in therapeutic drug carrier systems, 1994. **11**(2-3): p. 61-95.
181. Carlson, T., J. Lock, and R. Carrier, *Engineering the mucus barrier.* Annual review of biomedical engineering, 2018. **20**: p. 197-220.
182. Maher, S., R.J. Mrsny, and D.J. Brayden, *Intestinal permeation enhancers for oral peptide delivery.* Advanced drug delivery reviews, 2016. **106**: p. 277-319.
183. Twarog, C., *et al.*, *Intestinal permeation enhancers for oral delivery of macromolecules: a comparison between salcaprozate sodium (SNAC) and sodium caprate (C10).* Pharmaceutics, 2019. **11**(2): p. 78.

184. Yang, H., *et al.*, *Design of cell-type-specific hyperstable IL-4 mimetics via modular de novo scaffolds.* Nature chemical biology, 2023: p. 1-11.
185. Jammalamadaka, U. and K. Tappa, *Recent advances in biomaterials for 3D printing and tissue engineering.* Journal of functional biomaterials, 2018. **9**(1): p. 22.
186. Tokuriki, N., *et al.*, *How protein stability and new functions trade off.* PLoS computational biology, 2008. **4**(2): p. e1000002.
187. Bigman, L.S. and Y. Levy, *Proteins: molecules defined by their trade-offs.* Current Opinion in Structural Biology, 2020. **60**: p. 50-56.
188. Kondo, E., H. Iioka, and K. Saito, *Tumor-homing peptide and its utility for advanced cancer medicine.* Cancer science, 2021. **112**(6): p. 2118-2125.
189. Zinsli, L.V., *et al.*, *Deimmunization of protein therapeutics–Recent advances in experimental and computational epitope prediction and deletion.* Computational and Structural Biotechnology Journal, 2021. **19**: p. 315-329.
190. Guillet, J.-G., *et al.*, *Immunological self, nonself discrimination.* Science, 1987. **235**(4791): p. 865-870.
191. Pratt, K.P., *Anti-drug antibodies: emerging approaches to predict, reduce or reverse biotherapeutic immunogenicity.* Antibodies, 2018. **7**(2): p. 19.
192. Arnold, D.F. and S.A. Misbah, *Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1, 3-galactose.* The New England journal of medicine, 2008. **358**(25): p. 2735-2736.
193. Casadevall, N., *et al.*, *Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin.* New England Journal of Medicine, 2002. **346**(7): p. 469-475.
194. Hermeling, S., *et al.*, *Structure-immunogenicity relationships of therapeutic proteins.* Pharmaceutical research, 2004. **21**: p. 897-903.
195. Ratanji, K.D., *et al.*, *Immunogenicity of therapeutic proteins: influence of aggregation.* Journal of immunotoxicology, 2014. **11**(2): p. 99-109.
196. Chaplin, D.D., *Overview of the immune response.* Journal of allergy and clinical immunology, 2010. **125**(2): p. S3-S23.
197. Joyce, S. and N. Ternette, *Know thy immune self and non-self: Proteomics informs on the expanse of self and non-self, and how and where they arise.* Proteomics, 2021. **21**(23-24): p. 2000143.

198. Peters, B., M. Nielsen, and A. Sette, *T cell epitope predictions.* Annual Review of Immunology, 2020. **38**: p. 123-145.
199. Reynisson, B., *et al.*, *NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data.* Nucleic acids research, 2020. **48**(W1): p. W449-W454.
200. Holder, P.G., *et al.*, *Engineering interferons and interleukins for cancer immunotherapy.* Advanced drug delivery reviews, 2022. **182**: p. 114112.
201. Li, Z., B.-F. Krippendorff, and D.K. Shah, *Influence of molecular size on the clearance of antibody fragments.* Pharmaceutical research, 2017. **34**: p. 2131-2141.
202. Filpula, D. and H. Zhao, *Releasable PEGylation of proteins with customized linkers.* Advanced drug delivery reviews, 2008. **60**(1): p. 29-49.
203. Suzuki, T., *et al.*, *Importance of neonatal FcR in regulating the serum half-life of therapeutic proteins containing the Fc domain of human IgG1: a comparative study of the affinity of monoclonal antibodies and Fc-fusion proteins to human neonatal FcR.* The journal of immunology, 2010. **184**(4): p. 1968-1976.
204. Sleep, D., J. Cameron, and L.R. Evans, *Albumin as a versatile platform for drug half-life extension.* Biochimica et Biophysica Acta (BBA)-General Subjects, 2013. **1830**(12): p. 5526-5534.
205. Sockolosky, J.T. and F.C. Szoka, *The neonatal Fc receptor, FcRn, as a target for drug delivery and therapy.* Advanced drug delivery reviews, 2015. **91**: p. 109-124.
206. Molineux, G., *The design and development of pegfilgrastim (PEG-rmetHuG-CSF, Neulasta®).* Current pharmaceutical design, 2004. **10**(11): p. 1235-1244.
207. Arvedson, T., J. O'Kelly, and B.-B. Yang, *Design rationale and development approach for pegfilgrastim as a long-acting granulocyte colony-stimulating factor.* BioDrugs, 2015. **29**: p. 185-198.
208. Roskos, L.K., *et al.*, *Pharmacokinetic/pharmacodynamic modeling of pegfilgrastim in healthy subjects.* The Journal of Clinical Pharmacology, 2006. **46**(7): p. 747-757.
209. Curran, M.P. and K.L. Goa, *Pegfilgrastim.* Drugs, 2002. **62**: p. 1207-1213.
210. Green, M., *et al.*, *A randomized double-blind multicenter phase III study offixed-dose single-administration pegfilgrastim versus daily filgrastim in patients*

*receiving myelosuppressive chemotherapy.* Annals of Oncology, 2003. **14**(1): p. 29-35.

211. Gaertner, H.F. and R.E. Offord, *Site-specific attachment of functionalized poly (ethylene glycol) to the amino terminus of proteins.* Bioconjugate chemistry, 1996. **7**(1): p. 38-44.
212. Elsadek, N.E., *et al.*, *Pegfilgrastim (PEG-G-CSF) Induces anti-polyethylene glycol (PEG) IgM via a T cell-dependent mechanism.* Biological and Pharmaceutical Bulletin, 2020. **43**(9): p. 1393-1397.
213. Elsadek, N.E., *et al.*, *Pegfilgrastim (PEG-G-CSF) induces anti-PEG IgM in a dose dependent manner and causes the accelerated blood clearance (ABC) phenomenon upon repeated administration in mice.* European Journal of Pharmaceutics and Biopharmaceutics, 2020. **152**: p. 56-62.
214. Blair, H.A., *Efbemalenograstim Alfa: First Approval.* Drugs, 2023: p. 1-6.
215. Hiraoka, O., H. Anaguchi, and Y. Ota, *Evidence for the ligand-induced conversion from a dimer to a tetramer of the granulocyte colony-stimulating factor receptor.* FEBS letters, 1994. **356**(2-3): p. 255-260.
216. Ward, A.C., *et al.*, *Novel point mutation in the extracellular domain of the granulocyte colony-stimulating factor (G-CSF) receptor in a case of severe congenital neutropenia hyporesponsive to G-CSF treatment.* The Journal of experimental medicine, 1999. **190**(4): p. 497-508.

## Appendix

I. Tuning of granulopoietic signaling by *de novo* designed agonists

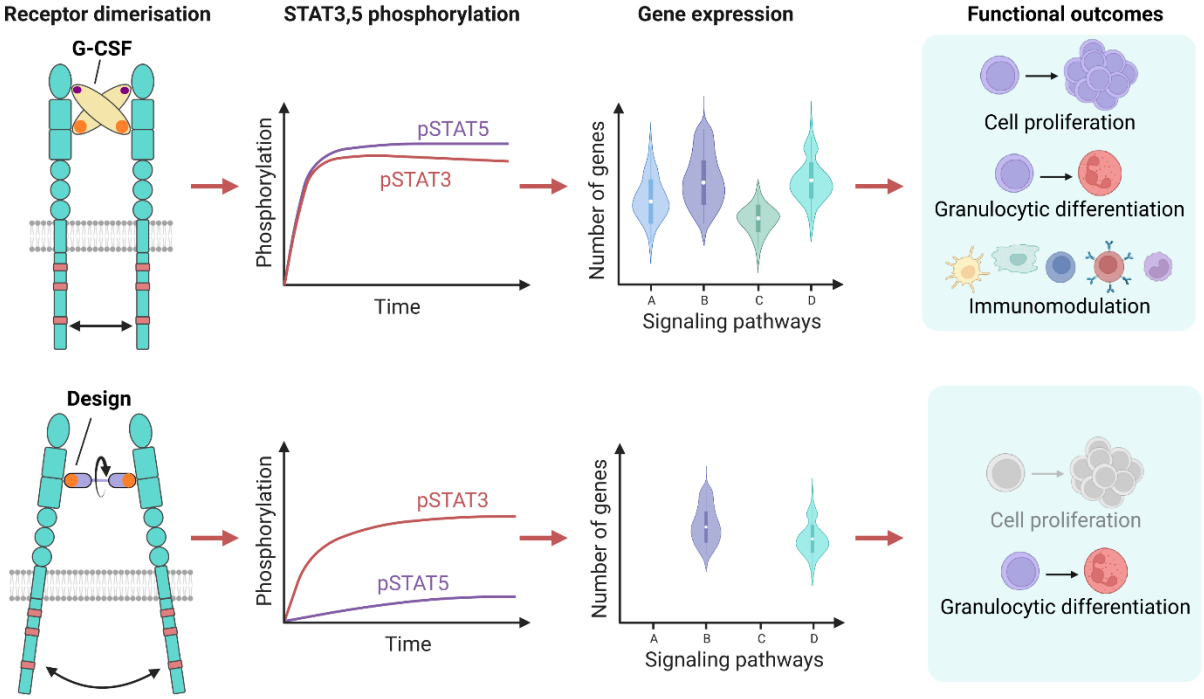
II. *De novo* design of G-CSFR-inhibiting proteins

# Tuning of granulopoietic signaling by *de novo* designed agonists

Timo Ullrich[1,5], Christoph Pollmann[2], Malte Ritter[3], Jérémy Haaf[3], Narges Aghaallaei[3,4], Ivan Tesakov[3], Maya El-Riz[2], Kateryna Maksymenko[1,5], Valeriia Hatskovska[1,3], Sergey Kandabarau[3], Maksim Klimiankou[3], Claudia Lengerke[3], Karl Welte[3,7], Birte Hernandez-Alvarez[1], Patrick Müller[5,6], Andrei Lupas[1], Jacob Piehler[2], Julia Skokowa[3,*,#], Mohammad ElGamacy[1,3,*,#]

[1] Max Planck Institute for Biology, Department Protein Evolution, 72076 Tübingen, Germany
[2] Department of Biology/Chemistry and Center for Cellular Nanoanalytics, Osnabrück University, 49076 Osnabrück, Germany
[3] Internal Medicine II, University Hospital Tübingen, 72076 Tübingen, Germany
[4] Present address: Ludwig Boltzmann Institute for Hematology and Oncology, Medical University of Vienna, 1090 Vienna, Austria
[5] Friedrich Miescher Laboratory of the Max Planck Society, 72076 Tübingen, Germany
[6] Present address: Systems Biology of Development and Disease, University of Konstanz, 78464 Konstanz, Germany
[7] University Children's Hospital Tübingen, 72076 Tübingen, Germany
*equal contribution as senior authors
#corresponding authors: julia.skokowa@med.uni-tuebingen.de, mohammad.elgamacy@med.uni-tuebingen.de

## Graphical abstract





## Abstract

Enhancing cytokine-based therapies by systematically tuning how an agonist associates its receptor is emerging as a powerful new concept in drug discovery. Here, we report the design and characterization of agonists that tune the granulocyte-colony stimulating factor receptor (G-CSFR) activity, which is central for the proliferation and granulocytic differentiation of hematopoietic stem cells. Using design agonists, we study the impact of varying the receptor-binding affinity and dimerization geometry on receptor association, downstream signaling, and cellular response. Hence, we achieved agonists with altered signaling specificities that are hyper-thermostable, can outcompete the native ligand (G-CSF), and bias granulopoietic differentiation over triggering proliferation. Furthermore, the design agonists differentially modulate the kinetics and amplitudes of signal transduction pathways, and gene expression patterns. Unlike G-CSF, they achieve selective activation of hematopoietic functions with minimal undesired immunomodulatory effects. These findings demonstrate the potential of dissecting the complex G-CSFR signaling, and open ways for new therapeutic applications.

## Introduction

Cytokines are secreted growth factors that canonically function by associating with the ectodomains of their cognate receptor subunits, which leads to the activation of downstream signaling pathways that steer the fate and function of hematopoietic or immune cells [1, 2]. Receptor activation by its native cytokine ligand triggers a complex pattern of different signaling pathways that results in different cellular responses depending on the type and stage of the target cell [3, 4]. Synthetic ligands, purpose-built to engage the cytokine receptors in non-native ways, provide ideal tools to selectively bias these signaling outcomes [5]. Such approaches have proven particularly attractive for the class I/II cytokine receptor families, which signal via non-covalently associated Janus family tyrosine kinases (JAKs). These receptors must dimerize for activating downstream signal activation, which is most prominently mediated by signal transducer and activators of transcription (STAT) proteins [6-9]. Controlling activation of class I/II cytokine receptors is of considerable interest for therapeutic intervention, yet hampered by the often broad and pleiotropic responses [10]. Decomposing these signaling outcomes can unlock previously untapped therapeutic outcomes and applications of cytokine receptor agonists [11].

Cases of differential cytokine receptor signaling by natural ligands [12] have inspired efforts to systematically bias signaling specificity by steering receptor dimerization with design agonists [11]. These ligands can be further engineered to possess optimal pharmaceutical properties, offering a means to reversible, extracellular correction of pathologic cellular behavior, or enhanced activity in physiological settings [11, 13]. For instance, engineered ligands with enhanced binding affinity to the IL-2R β-subunit and attenuated affinities to the γ-subunit induce immunomodulatory outcomes distinct from the native ligands (i.e. IL-2) while outcompeting endogenous IL-2 [14]. Another notable study reported the affinity-engineering of IL-22-mimetics that not only could induce altered intracellular signaling but also exhibited altered tissue-targeting specificity compared to the native ligand [15]. These examples and others highlight the effect of the ligand's receptor-binding affinity on various aspects of complex association and signaling [16, 17]. In addition to affinity, the receptors' association geometry can provide another layer of control over downstream signaling [18-21]. Only recently, ligand-based tuning of cytokine receptors has been demonstrated for three receptor types. Specifically, diabodies and DARPins dimerizing the erythropoietin receptor (EPOR) in diverse, non-native orientations could induce a range of graded or differential signaling through different association geometries [22, 23]. Using bispecific variable antibody fragment fusions (i.e. VHH or scFv), signal tuning or decomposition could be demonstrated for heterodimeric receptor complexes of the IL-2 receptor (IL-2Rβ:$\gamma_c$) and the type I interferon receptor (IFNAR1:IFNAR2) [24].

In this study, we aim to explore the possibility of tuning signaling via designed agonists of the granulocyte-colony stimulating factor receptor (G-CSFR), which has not been investigated to date. The granulocyte-colony stimulating factor (G-CSF) induces downstream signaling by homodimerizing its cognate receptor (G-CSFR) in a 2:2 cross-over assembly [2, 25-30]. This signaling is a crucial driver of granulopoiesis and plays diverse roles on hematopoietic stem cells, more mature myeloid progenitors, and mature immune cells, primarily acting on proliferation, differentiation, and mobilization of stem cells, but also regulating the functions of immune cells [31]. The therapeutic importance of G-CSF has thus led to the wide adoption of its recombinant human form, rhG-CSF, in the clinic for treating acquired or congenital neutropenia, in addition to the mobilization of hematopoietic stem cells for transplantation [32, 33]. The therapeutic application of rhG-CSF is however still limited since G-CSFR is also expressed in several subtypes of immune cells, osteoclasts and tumor cells. Inducing specific, e.g., granulopoietic, activity of G-CSF while suppressing its unintended action on cell proliferation and activation of immune cells, osteoclasts or cancer cells can broaden its clinical use. To achieve this, we must distinguish between signal transduction pathways that trigger differentiation over proliferation or immune cell activation. The key effectors of G-CSFR signaling are STAT3 and STAT5 proteins [34], among others. STATs are activated by tyrosine

phosphorylation (pSTAT) through Janus Kinase 2 (JAK2) upon G-CSFR dimerization whereby STAT3 phosphorylation occurs at the membrane-distal regions of the intracellular segments of the receptor [35]. At the same time, STAT5 is phosphorylated at the membrane-proximal receptor region. The sequential activation of STAT3 and STAT5 may dictate their functional outcomes where pSTAT3 mostly transduces differentiation signals to the cell while retaining some proliferative activity, and pSTAT5 predominantly induces proliferation [36-39]. The balance between STAT3 and STAT5 phosphorylation is decisive for cellular outcomes upon G-CSF stimulation [40, 41]. Thus, by dissecting the activation kinetics of STAT3 and STAT5, specific G-CSF activity might be achieved.

Leveraging these differences in STAT3/STAT5 recruitment and activation, we here aimed to systematically bias pSTAT3/pSTAT5 kinetics and amplitudes by designing G-CSFR agonists with altered receptor orientations in the signaling complex. We achieved this by focusing on highly optimized G-CSFR binding modules, which were subsequently harnessed to design G-CSFR agonists imposing diverse receptor orientations. Based on characterizing their biophysical properties and investigating their impact on receptor dimerization in the plasma membrane, in conjunction with intracellular signal activation, and cellular activation *in vitro* and *in vivo*, we highlight the potential to tune G-CSFR signaling towards differentiation, while avoiding its undesired non-hematopoietic effects.

# Results

### *Computationally-guided screening identifies higher-affinity G-CSFR binders*

G-CSF dimerizes G-CSFR in a 2:2 stoichiometry with each ligand cross-linking the two receptor subunits by binding via high and low affinity binding sites in a reciprocal manner (Fig. 1A). Previously, we have designed an artificial ligand *de novo*, Boskar4, that encodes one G-CSFR-binding site. When two copies of Boskar4 were tandemly linked by a short flexible linker (i.e. Boskar4 short tandem of 2 domains; B4_st2), it could dimerize and activate G-CSFR (Fig. 1B) [42]. G-CSF engages G-CSFR through its two high- and low-affinity binding sites; site II and site III (located on the ligand) the CRH and Ig-Like domains of the receptor, respectively. In contrast, Boskar4 encodes only the high-affinity binding site (a site II-like epitope) that engages the CRH domain through a relatively large interface surface area (Fig. 1C).

We first sought to enhance the receptor-binding affinity of Boskar4 (Fig. 1D), with the dual aim of achieving more efficient receptor dimerizers (Fig. 1E), and evaluating the impact of affinity change on receptor signaling. Hence, we needed to design additional mutations at the Boskar4:G-CSFR binding interface in order to improve its binding affinity. We thus modelled the Boskar4:G-CSFR complex (Fig. 1C) using all conformations from our previously compiled NMR ensemble (PDB: 7NY0). These models were used as input for the multi-model, single-point mutagenesis application from the Damietta protein design software [43]. We set six amino acid positions on the Boskar4 surface as mutation target sites at the binding interface with G-CSFR (Fig. S1A), which were selected based on their proximity to the G-CSFR surface, while avoiding previously described critical binding residues [44]. We included the top five mutations at each position with the joint lowest energy at the binding interface across all models (as well as wild-type identities) for library generation. Finally, we built two site-directed mutagenesis libraries based on either the designed sequences or site-saturation mutagenesis (as a random control) of the same positions (Fig. S1C, D). The control library contained on average 12 mutations per position, in contrast to an average of 6 mutations per position for the designed library. The resulting amino acid diversity of the designed library was $3.9\times10^4$ and a final diversity of $2.3\times10^5$ variants after PCR mutagenesis, in comparison to the saturation library, which encompassed $3.1\times10^6$ variants. We used a bacterial display screening system [45] in which the ligand is presented as an extracellular Intimin fusion on the surface of the bacterial particles. Upon adding fluorescently-labelled G-CSFR, we sorted the bacterial particles for receptor binding using fluorescence-activated sorting (FACS) (Materials and Methods). We ran both libraries (saturation and designed) through five cycles of FACS enrichment under the same conditions, where shifts in the fluorescence of both populations were observed across

the enrichment cycles (Fig. 2A). Forty-eight single clones were picked from either library after the third and fifth FACS cycles, amounting to a total of 192 clones. Using on-plate fluorescence- and luminescence-based assays we quantified the relative binding of these clones to the receptor under the same conditions (Fig. S2).

The sequences of the top 40 clones picked from both libraries after Sanger sequencing yielded 16 unique Boskar4 variants (Fig. S2D). A focused FACS-based analysis for these ligands' expression level and binding affinity on the surface of *Escherichia coli* (*E. coli)* (Fig. S3) showed the expression-normalized binding affinities to range from 3- to 37-fold increase, when compared to the starting Boskar4 clone. To narrow down the number of candidates for subsequent investigations, while keeping diversity, we chose six variants based on their estimated affinity, frequency in the screen, or the respective mutation composition (Fig. S2D and S3B). The most frequent variant was bv1, while the highest affinity variant was bv2. Four additional variants, bv6, bv8, bv15 and bv16 were chosen based on their middle- or lower- binding affinity profiles. Specifically, variants with polar and charged residues were included with the goal of selecting the most soluble and specific binders.

In order to create efficient G-CSFR agonists, the six selected variants were incorporated into short tandem repeats of two domain copies connected by a $G_4S_2$ sequence (Fig. S4). The generated st2 variants were readily purified using immobilized metal affinity chromatography (IMAC) from the soluble fraction of *E. coli* cell lysate, where preparative size exclusion chromatography and SDS-PAGE analysis indicated that they have the expected molecular weight of approximately 29 kDa (Fig. S5). Using surface plasmon resonance (SPR), the binding affinity and kinetic parameters of the designs to G-CSFR were obtained. The SPR results showed that all the six tested variants in their tandem form bind to G-CSFR within the sub-nanomolar affinity range, with up to 7-fold enhanced affinity compared to the starting template B4_st2 (S6 and Table 1). The observed affinity improvement was also higher when evaluated for single-domain Boskar4 variants. For instance, the single-domain form of the most biologically active variant (bv6) exhibited a $K_D$ of 14 ± 2 nM, in comparison to the single-domain Boskar4 template, with $K_D$ of 173 ± 43 nM (Fig. 2B, S7B). Notably, circular dichroism (CD) measurements of bv6 showed the expected strong alpha-helical signal, and nanoscale differential scanning fluorimetry (nanoDSF) measurements demonstrated that bv6 exhibited uncompromised thermostability when compared to Boskar4 (Fig. S7C, D).

***Affinity enhancement yields highly potent G-CSFR agonists with uncompromised thermostability***

Given the tight binding affinities, we expected our bivalent constructs to function as more efficient G-CSFR agonists compared to the not affinity-enhanced designs, which we evaluated using the G-CSF-dependent cell line NFS-60 [46, 47]. These results showed that all of the affinity-enhanced variants were more potent in inducing the proliferation of NFS-60 cells when compared to the starting template (B4_st2). Remarkably, the most active design, bv6_st2 had an $EC_{50}$ of 2.9 ±1.3 pM compared to 33 ±23.6 pM for B4_st2 and 2.6 ±1.8 pM for rhG-CSF (Fig. 2C, S8 and Table 1) while no differences in the maximum response of proliferative activity were noticed (Fig. S8E). The observed difference between the mean $EC_{50}$ values of the most active variant (bv6_st2) and rhG-CSF was statistically insignificant (*p-value* = 0.9). The observed range of affinities and activities of the studied variants showed no strong correlations between the proliferative activity ($EC_{50}$) and the binding affinity ($K_D$; $R_{pearson}$ = 0.15), the association rate constant ($k_a$; $R_{pearson}$ = 0.01), or the dissociation rate constant ($k_d$; $R_{pearson}$ = 0.12) (Fig. S9). We therefore expect that the activity of variants within this narrow range of affinities might be influenced by other factors such as their relative stabilities and binding specificity in the complex environment of cell culture-based assays, in comparison with SPR-derived affinities which were measured between reconstituted proteins.

To assess the structural stability of our designs, circular dichroism (CD) measurements were performed for all selected variants which showed the expected strong alpha-helical signal (Fig. S10A). Additionally, the thermostability was measured using nanoscale differential scanning fluorimetry

(nanoDSF) to inspect if the included mutations destabilized the proteins. Half of the tested variants (bv1_st2, bv6_st2 and bv8_st2) exhibited no detectable melting transitions up to 110 °C, while the other three (bv2_st2, bv15_st2 and bv16_st2) exhibited a melting onset at around 100 °C. No scattering (i.e. precipitation) was observed for any of the proteins over the full temperature range (Fig. S10B-H). Similar results were observed for the single-domain form of the variants, where bv6 displayed tighter receptor-binding affinity compared to the parent single-domain Boskar4, while preserving the hyper-thermostability of the latter (Fig. S7D).

These results indicate, that the affinity-driven functional enhancement did not alter the biophysical properties of the designs, which are far superior to rhG-CSF. Therefore, we further sought to test whether the design agonists exhibit similarly enhanced activity *in vivo*.

***Boskar4 variants induce neutrophil production in zebrafish embryos and mice***

We tested the effects of two optimized variants, bv6_st2 and bv8_st2 on the induction of granulopoiesis *in vivo* using two model organisms, zebrafish and mice. We used a fluorescent neutrophil reporter zebrafish transgenic line, *Tg(mpx:GFP)* [48], and C57BL/6 mice. For zebrafish experiments, equal volumes of rhG-CSF and Boskar4 variants were injected into the cardinal vein of *Tg(mpx:GFP)* larvae at 1.5 days post-fertilization (dpf). As negative control, Moevan_control (a design protein not capable of binding and activating G-CSFR, but expressed and purified under the same conditions as Boskar4 variants [49]) was injected. Interestingly, the increased number of $GFP^+$ neutrophils localized near the caudal hematopoietic tissue (tail region) was comparable between larvae treated with rhG-CSF or Boskar4 variants 24 h post-injection (Fig. 2D). The number of neutrophils was almost unchanged in the Moevan_control-injected larvae.

We additionally treated C57BL/6 Ly5.1 mice with 300 μg/kg of rhG-CSF or the two most active Boskar4 variants bv6_st2 and bv8_st2 by intraperitoneal (i.p.) injection every second day with five injections in total. One day after the fifth injection, the number of $Ly6G^+Ly6C^+$ neutrophils and $Ly6G^-Ly6C^+$ monocytes in the bone marrow of treated mice was evaluated. Indeed, treatment of mice with bv6_st2 or bv8_st2 induced the production of neutrophils and monocytes at a comparable level to rhG-CSF (Fig. 2E). These results highlight the ability of optimized Boscar4 variants to potently activate G-CSFR *in vivo*.

***Design of agonists altering the relative orientation of G-CSFR***

After identifying the most active G-CSFR binders, our next step was to build geometry-tuning agonists to evaluate the sensitivity of G-CSFR signaling to non-native receptor orientations. In order to dimerize G-CSFR subunits at bespoke angles, we sought to create rigidly juxtaposed bv6 modules, where two modules share one continuous common helix (Fig. 1E). By varying the length of an inserted helical segment, and splicing the N-terminal and C-terminal helices of two binding modules, different inter-domain rotations could be modelled (Fig. S11A, B). Using epitope-alignment and AlphaFold2 [50] we modelled the predicted full-length assemblies of resulting 2:1 complexes (Fig. 3A). The templates named ori0 through ori4 (orientation-rigging designs) were initially sculpted with poly-alanine splicing helical segments that were sequence-designed using the combinatorial sampler application of Damietta [43]. We then ran molecular dynamics simulations of the lowest energy candidates in order to validate the putative conformation for each ori design, and estimate the degree of conformational deviation from the AlphaFold2-predicted models (Fig. S11C, D). The oris were predicted to impose discrete distances across the receptor transmembrane domains (TMDs; $d_0$ = 78 ±32 Å, $d_1$= 206 ±14 Å, $d_2$ = 287 ±6 Å, $d_3$= 155 ±35 Å, and $d_4$ = 143 ±37 Å, for ori0, ori1, ori2, ori3, and ori4, respectively), whereas the 2:2 native complex (PDB: 2D9Q) model exhibited a relatively short distance of $d_g$ = 57 Å (Fig. 3A, B).

The generated designs ori0-4 were purified from *E. coli* and analyzed by size exclusion chromatography, which confirmed the expected size of around 29 kDa. With the exception of ori1, which showed a clear monomer-dimer equilibrium, all the other designs were purely monomeric in solution (Fig. S12). CD and nanoDSF measurements confirmed the helical secondary structure and the

hyperstability of the design agonists, similar to their template, bv6 (Fig. S13). In all cases, no precipitation (increased scattering) was observed for any of the designs over the full temperature course. All the designs bound to G-CSFR as expected, where their apparent binding affinities ranged from 0.44 nM to 0.73 nM (Fig. S14 and Table 1). These values are comparable with those observed for bv6_st2, which exhibits an average $K_D$ of 0.67 nM.

***The geometric modulation of the G-CSFR by design agonists differentially affects cell proliferation***

The designed geometric modulators possessed varying proliferative activities, as tested in the NFS-60 cell line. The $EC_{50}$ values of these ranged from 36 ±19.8 pM down to 2.8 ± 1.4 pM, and exhibited a reproducible pattern (Table 1 and Fig. S15), being comparable to Boskar4_st2 designs. For ori1, we observed a reduction in the maximum proliferation level ($E_{max}$) by about 20 %, while no significant differences were observed for all other designs compared to rhG-CSF (Fig. S15H). The proliferative $EC_{50}$ values of these design agonists correlated with the modelled distances across the receptor TMD for these designs ($R^2$ = 0.92; Fig. 3C). Highest potency was observed for ori0, which dimerizes G-CSFR with a very similar TMD spacing as native G-CSF ($d_0 \approx d_g$). On the other hand, the design agonists inducing the farthest distance between the TMD, ori1 and ori2, showed the lowest potency. While these results clearly underline inter-TMD spacing as a key parameter defining signaling activity, imposing these non-native geometries across two-dimensionally arranged receptors may result in significant conformational strain that reduces for some geometries the effective binding affinity to the cell surface receptor.

***Design agonists impose distinct complexes with G-CSFR and can override the native ligand***

The above results thus point towards a clear sensitivity of receptor activation towards the differential geometries of the design agonists. These differences stand out despite the fact that the designs (ori0-4) and the flexible tandem (bv6_st2) all share the same binding interface and possess very similar affinity values (Table 1). To query the properties of these complexes in a cellular environment, we explored G-CSFR assembly in the plasma membrane by live cell single-molecule fluorescence imaging. To this end, G-CSFR fused to an N-terminal ALFA-tag was expressed in HeLa cells and labeled selectively at the cell surface by applying a mixture of anti-ALFA nanobodies conjugated with Cy3B and AT643, respectively (Fig. 4A) [51]. Since it has been shown, that associated JAK2 energetically contributes to cytokine receptor complex formation, we additionally transfected JAK2 C-terminally fused to mEGFP [8]. This construct lacks the Tyrosine kinase domain (TK) to ensure that the experiments are not biased by activated downstream pathways while maintaining its contribution to dimerization. Simultaneous dual-color single molecule imaging by total internal reflection fluorescence microscopy enabled identifying individual G-CSFR dimers by co-localization and co-tracking analysis (Supplementary Movie S1, Fig. 4B and S16A). As expected, G-CSFR dimerization was negligible in the resting state, but strongly stimulated by G-CSF and all design agonists (Fig. 4C). While similar dimerization levels were observed for most design agonists, dimerization induced by ori2 and ori3 was significantly lower. Importantly, similar receptor cell surface densities were used in all experiments (Fig. S16B), thus ensuring that dimerization levels were not biased by the two-dimensional (2D) receptor concentration governing agonist-mediated crosslinking of G-CSFR into dimers (cf. Fig. 1A). Since the three-dimensional (3D) binding affinity of all design agonists is identical, the distinct differences in dimerization potencies of different design agonist can be ascribed by different two-dimensional (2D) binding affinity (cf. Fig. 4A). This can be rationalized by changes in the 2D association rate constant $k_a^{2D}$, but not the complex stability $k_d^{2D}$ due to geometrical constraints imposed by the orientation of the receptor ectodomains at the cell surface. With the formation of 1:1 G-CSFR:agonist complexes competing with G-CSFR dimerization, reduction in 2D binding affinity is accompanied by a lower maximum of the bell-shaped concentration-dimerization relationship of homomeric dimerizers [52]. Changes in dimerization efficiency can explain differences in the $EC_{50}$ values, as the overall receptor binding affinity depends on dimerization efficiency [53, 54]. Similar effects have been observed for design agonists for geometric control of the EPOR [22, 23]. However, different G-CSFR dimerization

levels induced by different design agonists did not directly correlate with their activity. Thus, ori1 showed substantially higher dimerization levels as compared to ori2, yet their NFS-60 proliferation activity was similar. Additionally, ori1 and ori0 had similar dimerization levels, but ori0 had about ten times higher proliferation activity compared to ori1 (Table 1). These results therefore confirm that biased signaling observed for design agonists is related to the altered geometry of the signaling complex rather than weakened dimerization.

We furthermore analyzed the diffusion properties of agonist-induced G-CSFR dimers as compared to unstimulated monomers. Diffusion constants as determined by mean squared displacement analysis were very similar for all G-CSFR dimers independent of the nature of the agonist, showing a decrease by ~45% compared to G-CSFR monomers (Fig. 4D). Similarly, the fraction of immobile G-CSFR increased by a factor of ~2, with ori3 showing surprisingly higher immobilization by a factor of ~3 (Fig. 4E). Receptor immobilization is very likely caused by endocytosis, and therefore differences for different design agonist may indicate changes in endocytic uptake for different design agonists [55]. Effective competition of design agonists with the native ligand was confirmed by co-tracking analysis in the presence of G-CSF and ori2 at equimolar concentrations. Under these conditions, very similarly reduced dimerization levels were observed as for ori2 alone (Fig. 4F and S16C, D). Taken together, these results corroborate that our design agonists potently induce G-CSFR dimerization at the cell surface and that biased signaling can be explained by systematically altered complex geometries.

***Different design agonists tune G-CSFR signaling amplitudes, kinetics and cellular response patterns***

To pinpoint biased G-CSFR signaling by designed agonists at the level of downstream signal propagation, we quantified the phosphorylation of STAT3 and STAT5 proteins in NFS-60 cells treated with rhG-CSF or designed agonists. Since STAT3 phosphorylation occurs at the membrane-distal regions of the intracellular segments of the receptor , while STAT5 is phosphorylated at the membrane-proximal receptor region [56], we reasoned that the relative ratio of pSTAT3/5 might be altered by different design agonists. We therefore aimed to investigate whether design proteins can affect STATs phosphorylation dependent on the engagement geometry of two G-CSFR receptor subunits. Western blot analysis after stimulating cells for 5 and 30 min confirmed increased phospho-STAT3 (pSTAT3, Tyr705) levels for bv6_st2 as compared to B4_st2, validating that our strategy increases potency by enhancing the affinity (Fig. 5A, C and S17). For the oris, however, striking differences were observed: while after 5 and 30 minutes ori0 and ori2 reached similar pSTAT3 levels and activation kinetics as bv6_st2, markedly lower pSTAT3 levels were found for ori1 for 5 min of stimulation (Fig. S17). As expected, designed proteins showed substantially lower and prolonged STAT5 phosphorylation (pSTAT5, Tyr694) as compared to rhG-CSF (Fig. 5B, D). Again, bv6_st2 and ori0 phosphorylated STAT5 more efficiently than B4_st2, ori2 or ori1, but the magnitude of STAT5 phosphorylation was markedly weaker for all variants, as compared to rhG-CSF. An increase of pSTAT5 levels from 5 min to 30 min stimulation was observed for all design agonists, while pSTAT5 levels remained largely constant upon stimulation with rhG-CSF. These results highlight that amplitudes and kinetics of STAT3 and STAT5 phosphorylation by G-CSFR can be tuned by its dimerization with altered geometry.

G-CSF induces proliferation and granulocytic differentiation of hematopoietic stem cells, and until now it is impossible to dissect these two cellular outcomes intentionally. It has been demonstrated that different magnitudes and kinetics of STAT3 vs STAT5 phosphorylation have different outcomes on hematopoietic stem cell functions, where the shift towards elevated activation of STAT3 triggers myeloid differentiation, while extensive STAT5 phosphorylation is rather responsible for cell proliferation [57]. Observing different activation kinetics of STAT3 and STAT5 by the designed agonists, we thought to achieve dissociation of proliferation from differentiation outcomes on hematopoietic stem cells. Indeed, as expected, we found that the designed proteins alone failed to trigger proliferation of hematopoietic stem and progenitor cells (HSPCs) compared to G-CSF (Fig. 5E). Proliferative activity of agonists was slightly induced in the presence of hematopoietic cytokines, IL-3

and SCF (Fig. 5F). In line with previous observations, ori0 has the strongest potential, as compared to other designs, but was still weaker, as G-CSF.

However, ori0 was as active as native G-CSF in inducing granulocytic differentiation of HSPCs when combined with IL-3 and SCF - stimulation with ori0 yielded substantially higher levels of granulocyte colony-forming units (CFU-G) in comparison to the flexibly connected bv6_st2 (Fig. 5G). G-CSF and ori0 also had a comparable effect in stimulating granulocyte-macrophages CFUs (CFU-GM) (Fig. 5H). Meanwhile, the other design agonists (ori1, ori2 and bv6_st2) did not significantly induce CFU-Gs in comparison with untreated cells. Interestingly, all the design agonists had a weak effect on the formation of monocytic colonies (CFU-M) (Fig. 5I), suggesting selective activation of granulocytic differentiation by then, when compared to native G-CSF. These results highlight the capacity of the design agonists to specifically induce granulocyte differentiation over the proliferation of HSPCs, with differential potencies.

Taken together, modeling the geometry of G-CSFR agonists allows fine-tuning of STAT3,5 phosphorylation that ultimately results in the dissection of proliferation from granulocytic differentiation of HSPCs.

***Designed G-CSFR agonists modulate intracellular transcriptional programs with hematopoietic bias***
For a more detailed picture of the intracellular response pattern initiated by different design agonists, we assessed their transcriptomic footprints. We carried out RNA-Seq of NFS-60 cells undergoing 8-h treatment with 1 nM of the selected agonists, thus ensuring saturated response for all variants. In the rhG-CSF group, we identified 1950 differentially expressed genes compared to the PBS-treated control group (log FC 1, FDR less than 0.05). The number of differentially regulated genes was lower (1292 genes) in the group treated with the most active design agonist with the shortest inter-TMD distance, ori0. This number was even lower (524 genes) in the ori2-treated group, the design agonist imposing the longest inter-TMD spacing (Fig. 6A). Stimulation of cells with bv6_st2 led to the differential expression of 1115 genes. Genes differentially regulated by the design agonists overlapped across all design agonists and were clearly subsets of rhG-CSF-regulated genes (Fig. 6B, C). The general trend revealed a gradual decrease in the expression of the regulated genes across the treatment groups, with the order of rhG-CSF > ori0 > bv6_st2 > B4_st2 > ori2 > ori1 > PBS (Fig. 6C and Fig. S18). Gene set enrichment analysis showed similar enrichment patterns, particularly in the regulation of hematopoiesis/myelopoiesis-related gene sets among rhG-CSF and the design agonists (Fig. 6D and Fig. S19). As with individual genes, the degree of regulation of the enriched gene sets was either the same as in the rhG-CSF group, or gradually reduced (Fig. S20).

A closer look at the transcription patterns identified subsets of genes to be either exclusively regulated by rhG-CSF, or by the design agonists (Supplementary data). Thus, rhG-CSF, but not design agonists regulated the expression of genes involved in immunomodulation of T lymphocytes, MDSCs, monocytes/macrophages, dendritic cells and basophils such as IL18R1, Lilrb4, Gp49a, Inhba, Ifitm6, Stfa1, Slfn2, Osm, Ly6i, Il19, Oas1g, Apol9b, Tarm1, Ifit3, CCR5, Trem1, Irg1, NLRRP12, MX2, and IL17c. These data argue for the possible specificity of the designs toward HSPCs. Chemokine receptors, CXCR2, CXCR4, CX3CR1, CCR5 were also regulated by G-CSF, but not design proteins. Additionally, VEGFR/ Kdr, transcription factors HOXA1, HOXB3, Klf3, Junb, MEIS3, Egr1, HOXC4; cell cycle regulators CDKN1A/P21 and CDKL2 and adhesion molecules AMIGO2, Adgre1, Tns1, Trio, and Pcdhga6, were affected in the rhG-CSF group only. Simultaneously, a small set of genes was regulated by the design agonists, but not by rhG-CSF, with different degrees of regulation between different designs (Fig. 6B). In line with these observations, gene ontology (GO) analysis of signaling pathways showed a gradual reduction of the number of regulated genes in the categories “signaling receptor activity”, “molecular transducer activity”, “transmembrane signaling receptor activity” with a slight overlap in “immune receptor activity”, and “cell adhesion molecule binding” (Fig. 6D).

## Discussion

Chimeras of natural cytokines or variable antibody domains have demonstrated the potential of developing new therapeutic approaches by targeting cytokine receptors [24, 58-67]. However, *de novo*-design agonists based on stable protein scaffolds can overcome several shortcomings. First, the compactness of these binding modules allows for constructing functional fusions with smaller increments of molecular weight. Second, starting from a highly stable template gives room for affinity maturation and functional enhancements, which typically erode structural stability [68]. Finally, sufficiently affine binding modules capable of outcompeting the native ligand at its binding site can impose the desired receptor induction mode, allowing it to override the natural cytokine's signaling.

As a proof-of-concept, here we have leveraged the *de novo*-designed G-CSFR binder (Boskar4) [42] to create functionally diverse agonists capable of tuning intracellular signaling and cellular responses. Our goal was to explore agonists that bind and dimerize G-CSFR with varying affinities and geometries inducing specific signaling patterns and thus distinct cellular activity. To ensure efficient dimerization even at non-native receptor orientations, we generated Boskar4 interface mutants with increased G-CSFR affinity by means of computationally-guided screening. Tandemly-linked agonists based on the affinity-optimized Boskar4 showed increased potency in inducing G-CSFR-triggered cell activation *in vitro* and *in vivo*. Despite the order-of-magnitude of enhancement in both affinity and cellular activity, the designs remained hyper-thermostable. In contrast, all previous studies aiming to improve the stability of rhG-CSF have yielded variants with only marginal improvements in biophysical properties [69-71]. This emphasizes the value of developing therapeutics starting from idealized *de novo* templates, which allows for a large room of stability erosion during functional optimization [68]. The high residual stability and the diverse binding site sequences of the described design variants are invaluable for de-risking further optimization to cloak any immune epitopes; an essential step towards clinical development.

This observed affinity-activity relationship for the agonists indicates that more stable ligand:receptor complexes lead to stronger signaling, despite their reliance on a flexible linker. Therefore, we sought to design rigid rotameric fusions (oris) that can probe the geometric determinants of G-CSFR activation. Proliferation assays along with signal transduction evaluations in NFS-60 cells have set forth the inter-TMD spacing of the resulting ori:G-CSFR complexes as a key parameter controlling the overall signaling amplitude. The analysis of the phosphorylation of the main secondary messengers (STAT3 and STAT5) revealed that both, higher affinity, and shorter inter-TMD spacing selectively increase the intracellular level of pSTAT3 while only minimally inducing pSTAT5. This might relate to the different localization and mechanisms by which STAT3 and STAT5 are phosphorylated along the intracellular segment of G-CSFR [37, 56, 72] and the effect of the inter-TMD spacing on the balance between these two processes. Broadly speaking, pSTAT3 primarily guides differentiation, while pSTAT5 mainly drives proliferation [36-39]. Corresponding results were obtained through granulocyte-colony formation assays in primary human HSPCs, where more affine and shorter-spacing designs led to more selective formation of granulocytic colonies, in comparison to G-CSF.

Single molecule fluorescence imaging confirmed the ability of most design agonists to potently dimerize G-CSFR in the plasma membrane of living cells, reaching similarly high levels as observed for G-CSF. The reduced dimerization levels observed for ori2 and ori3 indicate that manipulating the geometry of G-CSFR dimers is accompanied by a loss of 2D association kinetics. Changes in signaling activity and specificity, however, cannot be attributed to changes in dimerization. These experiments therefore clearly support the idea that design agonists assemble G-CSFR dimers with altered geometry. The very similar diffusion constants of all agonist-induced dimers, however, suggest more subtle changes in orientation as compared to the predicted structures, which should significantly reduce mobility. This could be explained by the additional strain imposed by membrane anchoring and additional interactions involving associated JAKs. Nonetheless, the distinct rise in the immobile fraction observed for ori3 points towards changes in endocytic uptake for that particular agonist. This warrants

further investigation, as endocytic trafficking of cytokine receptor signaling complexes is emerging as a complex determinant regulating potency and kinetics of signal activation and the specificity of cellular responses [55, 73, 74]. Indeed, we observe distinctly reduced STAT5 phosphorylation magnitude for the different design agonists, which could be related to altered endocytic trafficking. Therefore, manipulating trafficking kinetics by design agonists may be an additional parameter for biasing cytokine signaling.

By looking at the gene expression footprints induced by various design agonists, we could identify differential degrees of activity on downstream signal transduction pathways. Through our different design:receptor constellations, we observed varying degrees of tuning of the G-CSFR-triggered intracellular signaling activation or inhibition. Moreover, the design agonists minimally affect the expression of genes that are not regulated by rhG-CSF. Thus, we could fine-tune G-CSFR signaling without major unexpected transcriptional outcomes that are not linked to physiological G-CSFR activation. This information is important for at least two aspects – first, it gives significant insights into the signal transduction through G-CSFR using *de novo*-designed agonists. Second, it opens up an entirely new perspective in designing hematopoietic and granulopoietic proteins with specific activity and less undesired functions, when compared to native or recombinant cytokines. To date, designed IL-2 and and IL-4 mimics, that bind a subset of the natural cytokine's heteromeric receptor subunits, could achieve more selective biological responses [75, 76]. Alternatively, ligand-based tuning of homodimeric receptors, could rather be achieved through synthetic agonists with varying affinities and geometries for EPOR [22, 23] and TPOR [77, 78]. These synthetic agonists could impose different modes of receptor regulation, ranging from activation to total inhibition. Our results are first to establish the tunable modulation potential for G-CSFR, opening the way for selective modulation of its hematopoietic and non-hematopoietic functions through the described design strategy. Notably, this strategy may be extended to several other cytokine receptors to fine-tune cellular responses and thus, potentially facilitate novel therapeutic applications.

These findings demonstrate the potential for tailoring G-CSFR agonists to achieve more specific therapies for hematopoietic stem cell disorders, particularly given that G-CSFR activation by rhG-CSF can lead to diverse functional outcomes. The main function of G-CSF is to induce proliferation, migration and myeloid differentiation of hematopoietic stem cells. However, G-CSF can also protect from cell death and have immunomodulatory activity [31, 56, 79, 80]. For instance, rhG-CSF's widest use is to treat neutropenia acquired as a result of cytotoxic anti-cancer therapy [32]. In some cases, rhG-CSF treatment may inadvertently influence cancer treatment by directly activating G-CSFR-expressing tumor cells or dysregulating anti-cancer immune control by G-CSFR-expressing monocytes/macrophages, myeloid-derived suppressor cells (MDSC), T-lymphocyte subsets, dendritic cells or natural killer cells. These outcomes can negatively influence the course of treatment, and motivate the discovery of G-CSFR ligands capable of selective induction of specific granulopoietic signaling on hematopoietic stem cells without affecting proliferation or immune cell functions. The signaling footprint by our design agonists that are triggered by specific observed STAT3/STAT5 phosphorylation kinetics suggests that alteration in the dynamics of STAT phosphorylation can selectively activate hematopoietic and granulopoietic responses with only minimally inducing proliferation and immunomodulatory side effects. This study thus paves the way for the further development of potent, hyperstable, and selective sub-agonists of G-CSFR.

## Material and Methods

### *Computationally-guided library generation for affinity maturation*

Two libraries of mutants of 6 amino acid positions on the Boskar4 receptor-binding sites were generated using either site-saturation mutagenesis or computational design. The mutated positions on the Boskar4 surface were chosen based on two criteria (Fig. S1A). First, the positions in Boskar4 were determined which had the smallest average distance to G-CSFR in a modeled complex between the Boskar4

ensemble (PDB: 7NY0, [42]) and G-CSFR (PBD:2D9Q, [30]). Second, residues known to be critical for G-CSF activity were excluded from the final choice [44]. For the saturation library, high diversity degenerated codons were chosen which contained the original amino acid and other random amino acids constructing a library of a theoretical diversity (number of distinct gene variants) of $4.0\times10^6$ and an amino-acid diversity (number of distinct protein variants) of $3.2\times10^6$. Finally, the library was generated by PCR (Table S1 and S2) using one pair of primers (primer 3 and 4), each containing 3 degenerated sites to linearize an *E. coli* display system plasmid [45] encoding a fusion between the N-terminal part of Intimin and Boskar4 (pNB4, Table S3, Fig. S1B, C). The obtained PCR product was purified with the Wizard® SV Gel and PCR Clean-Up System (A928, Promega) and blunt end ligation was performed followed by electroporation of freshly made competent cells [81] to yield $3.6\times10^7$ total transformants. The *E. coli* strain DH10B $T1^R$ (C640003, Invitrogen) was used for all display experiments.

In more detail, prior to transformation, 0.4 U/µL T4 polynucleotide kinase (EK0031, Thermo Scientific™) was incubated with 5 ng/µL of linearized PCR product for 30 min at 37 °C, according to manufacturer instructions. After that, 0.25 U/µL T4 DNA ligase (EL0011, Thermo Scientific™) was added and incubated overnight at 16 °C. The ligation product was purified with the same PCR Clean-Up System mentioned above and eluted into 20 µL filtered $ddH_2O$. The complete 20 µL were transformed into electrocompetent *E. coli* as described in Tu *et al.* with the following changes [81]. In short, 50 mL of freshly grown *E. coli* exhibiting an optical density at 600 nm ($OD_{600}$) of ~ 0.6 were washed two times with 35 mL filtered $H_2O$. After washing, the bacterial pellet was resuspended with the purified ligation mix. The electroporation was performed at 1250 V and 5 ms. The cells were grown in 1 mL SOC Outgrowth Medium (B9020, NEB) for 2 h in a standard glass tube at 30 °C with 160 rpm. A 10-fold dilution series of the cells was made in SOC-medium, and they were plated on agar plates containing 34 µg/mL chloramphenicol and 2 % (w/v) D-glucose to estimate the total number of transformants. The rest of the sample was plated on five agar plates of the same type and incubated at 30 °C for about 20 h. The bacterial lawn was scraped from the plates and well mixed in 10 mL lysogeny broth (LB) medium. From 500 µL of bacterial suspension, the display plasmid containing the Boskar4 library was isolated and analyzed by Sanger sequencing with primers 19 and 20.

For the designed library, Damietta spl application [43] was used to estimate the free energy change of each single mutants of the same positions and same amino acids per position as the control library in a non-combinatorial manner. For each of the six positions, five amino acids with the lowest free energy change plus the original amino acid of Boskar4 at the corresponding position were used to define the designed, size-reduced library with a theoretical diversity of $3.89\times10^4$. The library was generated by PCR the same way as the original one but using three pairs of primers (primer 5 to 10, Table S1) such that the final library contained 15.6% of desired amino acids. The primer composition was designed by using the online tool of SwiftLib [82]. The designed library had a final theoretical diversity of $2.5\times10^5$, an amino-acid diversity of $2.3\times10^5$ and $2.7\times10^7$ total transformants were yielded.

***Screening procedure for enhanced binders***

The FACS-based screening of bacterially-displayed Boskar4 variants followed a slightly adapted procedure from that described by Salema *et al.* [45]. In brief, G-CSFR was first fluorescently labelled. This was done by incubating 100 µg of recombinant human G-CSFR (381-GR/CF, R&D Systems) in 950 µL PBS for 2 h at room temperature (RT) with Biotin-NHS (H1759, Sigma) in a 1:20 molar ratio (G-CSFR:Biotin-NHS). The reaction was stopped with 50 µL Tris (1 M pH 7.5). After 1 h of incubation on ice, the product was purified with a desalting column (Sephadex G25 PD-10, GE Healthcare), and the elution fractions were concentrated with Amicon Ultra centrifugal filter units (3 kDa, UFC800324, Merck) to a concentration of 0.24 mg/mL (3.6 µM). Aliquots of the biotinylated protein (herein referred to as BioGCSFR) were stored at -20 °C until use.

For the first sorting, DH10B $T1^R$ *E. coli* cells were transformed with the display-vector pN containing a Boskar4 library. The bacteria were freshly grown on LB agar plates containing 34 µg/mL chloramphenicol and 2% (w/v) D-glucose at 30 °C yielding ~$3.0\times10^8$ total transformants. Following the harvesting of bacteria from the agar plates, 1 mL of resuspended bacterial culture exhibiting an $OD_{600}$ of 3.3 was subjected to two successive washes, each involving 1 mL of LB medium. Subsequently, the culture was diluted to an initial $OD_{600}$ of 0.33 by adding it to 10 mL of LB medium containing 34 µg/mL of chloramphenicol. The cells were cultivated at a temperature of 30 °C and an agitation rate of 160 rpm. After a duration of 1 h, 50 µM of isopropyl β-D-1-thiogalactopyranoside (IPTG) was introduced into the culture, and the cultivation was continued for an additional 3 h under the same growth conditions. Cells equivalent to 1 mL of a cell resuspension possessing an $OD_{600}$ of 1 were collected, subjected to two washes with 1000 µL of phosphate-buffered saline (PBS), and eventually resuspended in 400 µL of PBS. 190 µL of this cell suspension was incubated with 10 nM BioGCSFR for 1 h at room temperature, followed by one wash with 1000 µL PBS. The washed cells were resuspended in 200 µL PBS containing 0.375 µL PE-Streptavidin (405203, BioLegend) and were subsequently incubated for 30 min at 4 °C. Finally, the sample was washed one more time with 1000 µL PBS and resuspended in 1000 µL PBS. FACS was performed with a BD FACSMelody™ Cell Sorter having set the PTM voltage for PE to 517 V, SSC to 426 V and a SSC threshold to 673 V. No gating was applied. All flow cytometry results were analyzed using FlowJo™ v10.1 Software (BD Life Sciences). 300 µL of the sample was mixed with 4 mL ice cold PBS and the flow rate was adjusted to end up with an event rate of ~8000 events/sec. 100,000 events were recorded per sample. The top ~0.1 % of the total population depending on a SSC-A/PE-A plot were sorted into a 1.5 mL tube provided with 200 µL LB medium and cooled to 5 °C. The sorting was continued until at least 20-fold more processed events were screened than the theoretical diversity of the correspondent library was. All sorted cells were plated on appropriate LB agar plates and incubated for about 20 h at 30 °C. Those plates were harvested and grown over night in LB medium containing 34 µg/mL chloramphenicol and 2% (w/v) D-glucose at 30 °C under static conditions. From this culture, 1 mL of resuspended bacteria with an $OD_{600}$ of 3.3 used for the next selection cycle performed in the same way as the first one.

This was followed by evaluating the relative binding of the top clones using plate-based assays. At this stage, single colonies of bacteria pools which were enriched for 3 or 5 cycles of FACS-based sorting (as described in the section above) were grown overnight in 100 µL LB (supplemented with 34 µg/mL chloramphenicol and 50 µM IPTG) in a 96-well plate sealed with Breathe-Easy ® sealing membrane (Z380059, Sigma-Aldrich) at 30 °C with 1100 rpm on a table top shaker (Thermomixer comfort, Eppendorf) covered with aluminum foil. On the next day, the cells were centrifuged at 3200 g for 5 min, the supernatant was decanted, and the cells were resuspended in 150 µL PBS. The washing was repeated the same way one more time, and the cells were resuspended in 50 µL PBS containing 10 nM BioGCSFR. After the plate was incubated for 1 h at RT, the cells were washed once with 200 µL PBS and resuspended in 50 µL PBS. To visualize the binding activity of each displayed Boskar4 variant to BioGCSFR, two separate secondary stainings, one fluorescent and one chemiluminescent, were performed as follows:

For the fluorescent staining, 25 µL of the cells were mixed with 25 µL PBS containing 0.094 µL PE-streptavidin followed by a 30 min incubation at 4 °C in the dark. After washing the cells one more time with PBS, the $OD_{600}$ and the fluorescence (excitation wavelength = 495 nm, emission wavelength = 574 nm) were measured on a plate reader (Synergy H4 Hybrid Microplate Reader, BioTek).

For the chemiluminescence stain, 25 µL of the cells were mixed with 25 µL PBS containing 1.25 µg/mL Avidin-HRP conjugate (Invitrogen, 434423) followed by a 30 min incubation at 4 °C in the dark. Afterwards the cells were washed once with 200 µL PBS and resuspended in 100 µL PBS. The $OD_{600}$ of each well was measured with the plate reader. In a separate dark plate (655097, Greiner Bio One) 12.5 µL cell suspension was added to 75 µL PBS. Then 12.5 µL ECL-substrate-solution (1705061, Bio-Rad) was pipetted quickly into each well, the plate was sealed with parafilm and mixed vigorously for

10 seconds on a vortex orbital shaker. Then the luminescence of each well was measured on the plate reader (Synergy H4 Hybrid Microplate Reader, Agilent BioTek) set to an integration time of 1 second, a gain of 135, normal read speed and a 100-millisecond delay, and a read height of 1 mm.

Both readouts were analyzed by normalizing to the corresponding cell density and calculating a z-score for each sample over all wells of the same condition. From the top 5 variants of each condition, plasmids were isolated with a plasmid preparation kit (740588, Macherey-Nagel) from a 3 mL overnight culture grown at 30 °C with 160 rpm in LB with 34 µg/mL chloramphenicol and 2% (w/v) D-glucose.

Finally, to obtain the expression-normalized binding of the top clones with unique sequences another FACS experiment was carried out. The individual bacteria clones carrying the display vector pN with Intimin fused either to Boskar4 or enhanced Boskar4 variant (bv1 to bv16) were grown over night at 30 °C with 160 rpm in 3 mL LB with the addition of 34 µg/mL chloramphenicol and 50 µM IPTG. The cells were pelleted at 4000 g for 3 min, the supernatant was discarded and the sample was resuspended in 1000 µL filtered PBS. This washing step was repeated one more time and the cells were resuspended in 1000 µL PBS. 45 µL of cells suspension were incubated for 90 min at RT in a total volume of 55 µL with 10 nM BioGCSFR and 1:200 Myc-Tag (9B11) Mouse mAb (2276, Cell Signaling Technology). After one wash with 500 µL PBS, the cells were resuspended in 200 µL of PBS containing 0.375 µL PE-Streptavidin (405203, BioLegend) and 0.5 µL of anti-mouse-IgG1 conjugated to Alexa 488 Fluor (A-21202, Invitrogen) and incubated for 30 min at 4 °C in the dark. Finally, the sample was washed one more time with 500 µL PBS, resuspended in 1000 µL PBS, and 100000 events were measured by FACS.

***Protein expression and purification***

First, a set of affinity-enhanced Boskar4 (B4) variant genes (bv1, bv2, bv6, bv8, bv15, and to bv16), which were identified from the affinity maturation screens, were used to generate short tandem fusions (st2, compare Fig. S4). In more detail, a single Boskar4 variant gene was amplified by PCR from pN separately using two different primer pairs (15 and 16, or 17 and 18; Table S1). The two obtained PCR fragments were assembled by NEBuilder® HiFi DNA Assembly Master Mix (E2621, New England BioLabs) into a pET28a backbone generated by PCR with primers 13 and 14. Additionally, the monomer of bv6 was cloned from pN to pET28a.

*E. coli* BL21(DE3) carrying a certain expression construct were grown to an $OD_{600}$ of ~0.6 in LB medium containing 40 µg/mL kanamycin at 37 °C and 160 rpm. Cells were induced with 0.5 mM IPTG and the protein expression was performed at 25 °C and 160 rpm for about 16 h. The cells were harvested at 8000 g for 30 min, the supernatant was decanted, and the pellet was lysed by sonication in lysis buffer (50 mM Tris pH8, 100 mM (for st2 designs) or 1 M NaCl (for ori designs), 20 µg/mL DNase (#A3778, ITW Reagents) and protease inhibitor cocktail (04693132001, Roche)). The lysate was centrifuged at 16000 g for 50 min, and the supernatant was used to perform Nickel Immobilized Metal Affinity Chromatography (IMAC). Finally, size exclusion chromatography (SEC) was performed with PBS on the concentrated sample with a HiLoad 16/600, Superdex 200 pg (GE28-9893-35, Merck) column or other if mentioned in the corresponding legend description. Fractions that were supposed to contain the protein of the expected size were concentrated by ultrafiltration (10 kDa, UFC9010, Millipore), and aliquots were frozen at -20 °C until use.

***Analysis of folding and thermal stability***

Circular dichroism (CD) spectra were recorded using a JASCO J-810 spectrometer with 0.3 mL samples at a protein concentration of 0.1 mg/mL in PBS buffer (pH 7.1) placed in 2 mm path length cuvettes. The spectral scans of the mean residual ellipticity of three accumulations were measured at a resolution of at least 0.5 nm over a range of 250-190 nm. Nano differential scanning fluorimetry (nanoDSF) was conducted using a Prometheus NT.48 (Nanotemper) and standard Prometheus capillaries (Nanotemper, PR-C002). A temperature ramp of 1 °C/min from 20 °C to 110 °C to 20 °C was applied for melting and

cooling, using 1 mg/mL protein samples in PBS (unless otherwise specified in the corresponding figure legend).

***Affinity determination with surface plasmon resonance***

To determine the receptor-binding affinity of Boskar4 variants, multi-cycle kinetics experiments were performed on a Biacore X100 system (GE Healthcare Life Sciences). Recombinant human G-CSFR (381-GR/CF, R&D Systems) was diluted to 50 µg/mL in 10 mM acetate buffer at pH 5.0 and immobilized on the surface of a CM5 sensor chip (GE Healthcare 29149604) using amine coupling chemistry. The protein samples were diluted in a running buffer (PBS with 0.05% v/v Tween-20). The measurements were performed at 25°C at a flow rate of 30 µL/min. Four sequential concentrations of the sample solution were used as follows: for Boskar4; 5000 nM, 2500 nM, 625 nM and 156.3 nM, for bv6; 125 nM, 63.5 nM, 31.3 nM, 15.6 nM; for all st2 variants (bv1_st2 to bv16_st2) and oris (ori0 to ori4); 20 nM, 10 nM, 5 nM and 2.5 nM). These samples were injected over the functionalized sensor chip surface for 180 s, followed by a 600 s dissociation phase with running buffer. At the end of each run, the sensor surface was regenerated with a 60 s injection of 50 mM NaOH. The reference responses and zero-concentration sensograms were subtracted from each dataset (double-referencing). Association rate constant ($k_a$), dissociation rate constant ($k_d$), and apparent equilibrium dissociation constants ($K_D$) were obtained using the Biacore X100 Evaluation Software following a 1:1 binding kinetic.

***Generation of orientation-rigging designs***

To generate preliminary constructs of the orientation-rigging designs (oris) containing a ridged helix-linker fusing two binding protein monomers, two bv6 monomers were connected N- to C-terminally (helix 4 of monomer 1 to helix 1 of monomer 2) with variable lengths of poly-alanine-stretches. The obtained sequences were modeled by AF2, and four constructs (ori1, ori2, ori3 and ori4) with an increasing number of connecting alanines were settled as a starting set of designs. The obtained AF2 structures were used as input to optimize the sequence of the helix-linker region with Damietta. Depending on the construct, up to 14 positions (compare Supplementary Methods) were mutated utilizing the few-to-many-to-few combinatorial sampler of Damietta (v0.32, [43]). Analysis of the designs' structural stability was done using a tempering molecular dynamics routine, and candidates ranked by their conformational homogeneity score, as previously described in [42]. This molecular dynamics analysis was performed for all unique variants that were generated by Damietta and the most stable variant of each construct was considered as the final design. To generate a binding protein in the same fashion that minimizes the distance of the two transmembrane domains of the bound G-CSFRs, hypothetical complexes of ridged tandem-fusions with varying poly-alanine-helix-stretches and two bound G-CSFRs were created manually. Based on the expected distance, possible steric hindrances and overall length of the ridged helix-linker, ori0 was selected and a final design was obtained as described above.

The ori expression constructs were generated the same way as explained for the short tandem fusions (compare Material and Methods section "Protein expression and purification"). In more detail, the two fragments (fr1 and fr2) encoding each one bv6 monomer and the corresponding ridged helix-linker were amplified and assembled together with the pET28a-backbone. The first fragment was generated with primer 15 plus the corresponding primer for each construct (primers 21, 23, 25, 27 or 29) and the second fragment with primer 18 plus the corresponding primer (primers 22, 24, 26, 28 or 30, Table S1).

***NFS-60 cell proliferation assays***

NFS-60 cells [46, 47] were maintained at 5% $CO_2$ and 37 °C in NFS-60-medium (RPMI 1640 medium with additional 1 mM L-glutamine, 1 mM Na-pyruvate, 10% FCS, 1% Antibiotic-Antimycotic (15240062, Gibco™) and 12.5% KMG-2: 5637-CM (Conditioned Medium from CLS Cell Lines Service)). Prior to the proliferation assay, the cells were washed 3 times with NFS-60-medium without KMG-2. The assay was performed in black 96-well plates (6005660, Perkin Elmer). For this purpose,

45.000 cells per well in a total volume of 150 µL NFS-60-medium without KMG-2 were cultured under maintenance conditions with different concentrations of designs or rhG-CSF for 48 h. Then, 30 µL of CellTiter-Blue® Reagent (G808, Promega) was added to each well and cells were further cultivated for approximately 90 min under maintenance conditions. Subsequently, the fluorescence of each well was recorded with a plate reader (Synergy H4 Hybrid Microplate Reader, Agilent BioTek) with an excitation wavelength of 560 nm and emission at 590 nm. The half-maximal effective concentration ($EC_{50}$) of each design or rhG-CSF was determined by fitting the obtained fluorescence values to their corresponding concentrations using a four-parameter sigmoidal function and the Nelder-Mead Simplex algorithm from the Python module SciPy [83]. For the analysis of the maximum response of proliferative activity ($E_{max}$), the mean of the three highest concentrations was estimated for each sample, such that all values considered for the analysis were in the activity plateau. These means were normalized to the maximum response within each experiment to account for the general variability of cell activity between the independent replicates. For statistical analysis, an ordinary one-way ANOVA followed by a Tukey HSD test was performed.

***Testing of granulopoietic activity of designed proteins in vivo***

To test the effect of the designs on zebrafish granulopoiesis, equal volumes (4 nL) of Moevan_control (4 mg/mL), rhG-CSF (2 mg/mL), Boskar4 variants, bv6_st2 (2 mg/mL) and bv8_st2 (2 mg/mL) were injected into the cardinal vein of transgenic *Tg(mpx:GFP)* [48] larvae at 1.5-day post-fertilization (dpf). Injected larvae were incubated at 28 °C. To quantify neutrophils 24 hours upon injection, larvae were positioned and orientated laterally within cavities formed in 1% agarose on a 96-well plate and then imaged using an SMZ18 Nikon fluorescence stereomicroscope. The number of GFP-expressing neutrophils was automatically determined by Imaris software using the spot detection tool. Fold change of the neutrophils was calculated by normalizing to the average of uninjected mpx:*gfp* counterparts at the same developmental stage. Zebrafish lines were maintained according to standard protocols and handled in accordance with European Union animal protection directive 2010/63/EU and approved by the local government (Tierschutzgesetz §11, Abs. 1, Nr. 1, husbandry permit 35/9185.46/Uni TÜ). All experiments described in the present study were conducted on larvae younger than 5 dpf.

For the mouse treatment experiments, B6.SJL-PtprcaPepcb/BoyCrl (Ly5.1) mice, aged between 6 to 8 weeks, were treated with intraperitoneal injections (i.p.) of rhG-CSF, or Boskar4 variants, bv6_st2 and bv8_st2. A concentration of 300 µg/kg was used for each protein injecting mice every second day for a total of five injections. Bone marrow cells of treated mice were isolated by flushing with a 22G syringe and subsequent filtering through a 45 µm cell strainer. After that, cells were counted and used for flow cytometry analyses on a FACS Canto II (BD) and using FlowJo (BD). Cells were stained with DAPI (#D9542, Sigma), anti-mouse CD45.1 PerCP (#110725, Biolegend), anti-mouse Ly6C AF488 (#128022, Biolegend), and anti-mouse Ly6G APC (#127614, Biolegend) antibody. Gates were set according to fluorescence minus one (FMO) control. Mice were maintained under pathogen-free conditions in the research animal facility of the University of Tübingen, according to German federal and state regulations (Regierungspräsidium Tübingen, M 05-20 G).

***Single molecule imaging and analysis***

For cell surface labelling, G-CSFR was N-terminally fused to the ALFA-tag. Additionally, to assess the co-expression of JAK2ΔTK it was C-terminally fused to mEGFP. Both proteins were encoded on pSems vectors including the signal sequence of Igκ (pSems-leader) [84]. HeLa cells (ACC 57, DSMZ Germany) were cultured as previously described [85]. For transient transfection, cells were incubated for 4-6 h with a mixture of 150 mM NaCl, 10 µL of 1 mg/mL polyethylenimine (PEI MAX®, Polysciences 24765) and 1500 ng (ALFAtag-G-CSFR) and 2500 ng (JAK2ΔTK-mEGFP) of the desired constructs. Labeling, washing and subsequent imaging were performed after mounting the coverslips into custom-made incubation chambers with a volume of 1 ml. Cells were equilibrated in medium with FBS but lacking phenol red supplemented with an oxygen scavenger and a redox-active photoprotectant

(0.5 mg/mL glucose oxidase (Sigma-Aldrich), 0.04 mg/mL catalase (Roche), 5% w/v glucose, 1 µM ascorbic acid and 1 µM methylviologene) to minimize photobleaching [86].

Selective cell surface receptor labeling was achieved by using anti-ALFAtag NBs, which were site-specifically labeled by maleimide chemistry via a single cysteine residue at their C-termini [86]. NBs labeled with Cy3B (DOL: 1.0) and ATTO 643 (DOL: 1.0) were added at concentrations of 3 nM each, at least 10 min before imaging. Coverslips were precoated with poly-l-lysine-graft-poly(ethylene glycol) to minimize unspecific binding of NBs and functionalized with RGD peptide for efficient cell adhesion [87].

Single-molecule imaging was carried out by dual-color total internal reflection fluorescence microscopy using an inverted microscope (IX71, Olympus) equipped with a spectral image splitter (DualView, Optical Insight) and a back-illuminated electron multiplied CCD camera (iXon DU897D, Andor Technology). Fluorophores were excited by simultaneous illumination with a 561-nm laser (CrystaLaser; approximately 32 W $cm^{-2}$) and a 642-nm laser (Omicron: approximately 22 W $cm^{-2}$). Image stacks of 150 frames were recorded for each cell at a time resolution of 32 ms per frame, with at least 10 cells recorded in each experiment. Ligands were incubated for 10 min before imaging. All imaging experiments were carried out at room temperature. Prior to image acquisition, the presence of JAK2ΔTK-mEGFP was confirmed through emission with a 488-nm laser (Sapphire LP, Coherent)

Dual-color time-lapse images were evaluated using an in-house developed MATLAB software (SLIMfast4C, https://zenodo.org/record/5712332) as previously described in detail [86]. After channel registration based on calibration with fiducial markers, molecules were localized using the multi-target tracking algorithm [88]. Immobile emitters were filtered out by spatiotemporal cluster analysis [89]. For co-tracking, frame-by-frame co-localization within a cut-off radius of 100 nm was applied followed by tracking of co-localized emitters using the utrack algorithm [90]. Molecules co-diffusing for 10 frames or more were identified as co-localized. Relative levels of co-localization were determined based on the fraction of co-localized particles [87]. Diffusion properties were determined from pooled single trajectories using mean squared displacement analysis for all trajectories with a lifetime greater than 10 frames. Diffusion constants were determined from the mean squared displacement by linear regression.

***Analysis of STAT3/5 phosphorylation by Western Blot***

NFS-60 cells were washed two times with NFS-60-medium (RPMI 1640 medium with additional 1 mM L-glutamine, 1 mM Na-pyruvate, 10% FCS, 1% Antibiotic-Antimycotic (15240062, Gibco™)) and starved at 300,000 cells/mL for 16 h at 5 % $CO_2$ and 37 °C. Subsequently, the starved cells were treated with 1 nM of rhG-CSF or the selected design for 5 min or 30 min.

Following the treatment, the cells were consistently maintained at low temperatures on ice. They were harvested using a refrigerated centrifuge operating at 4 °C at 300 g for 2 min. The cells were washed once with ice-cold PBS and the dry pellets were frozen at – 70 °C until further use. Whole-cell lysates were obtained by lysing $1\times10^6$ cells in 200 µL Laemmli buffer (30% glycerol, 6% SDS, 7.5% β-Mercaptoethanol, 0.75% Bromphenol blue in 200 nM Tris-HCL [pH 6.8]), which were subsequently heated at 95 °C for 5 min and spun down. Proteins were separated by dodecyl-sulfate polyacrylamide gel electrophoresis and transferred to Amersham™ Protran® nitrocellulose membranes or PVDF membranes (Invitrogen). The membranes were blocked with 5% non-fat dry milk-TBST (10 mM Tris-HCL [pH 8.0], 150 mM NaCl, 0.1% Tween® 20) for 1 h at room temperature and subsequently incubated with primary antibodies in 8 mL milk-TBST buffer overnight at 4 °C or for 2 h at room temperature (RT). After washing 4 times for 5 min with TBST, membranes were incubated with secondary antibodies in 8 mL milk-TBST buffer for 1 h at RT and then again washed 4 times for 5 min with TBST. The protein bands were detected using Pierce (Thermo Fisher Scientific) or Clarity (BioRad) ECL western blotting substrate kits, and imaged using Fusion FX (Vilber) or Sapphire™ (Azure Biosystems) imagers. The following antibodies and corresponding dilutions were used: primary rabbit polyclonal antibodies against: STAT5 1:500 (Cell Signaling, #9363S), phosphorylated STAT5

(Tyr694) 1:750 (Cell Signaling, #9351S), STAT3 1:200 (SantaCruz, # SC-482), alpha-tubulin 1:1000 (Cell Signaling, #2144S), primary rabbit monoclonal antibodies against: STAT5 1:1000 (Cell Signaling, #94205S), STAT3 1:1000 (Cell Signaling, #12640S), phosphorylated STAT3 (Tyr705) 1:1000 (Cell Signaling, #9145L), and GAPDH 1:1000 (Cell Signaling #2118S), and secondary horseradish peroxidase-conjugated anti-rabbit antibodies 1:3000 (Cell Signaling, #7074, or Jackson ImmunoResearch, #111-035-003). Gel images were exported for analysis as .tiff files and processed using FIJI software [91]. Following brightness/contrast adjustments, the bands of interest were isolated with the rectangle tool, and the peaks of band intensity were plotted. The peaks were individually separated using the line tool, and the areas under each peak were measured to determine the signal intensities of the immuno-detected proteins. Subsequently, these signal intensities were normalized by the corresponding signal of the housekeeping protein.

***Colony-forming unit (CFU) assay with human HSPCs***

Human $CD34^+$ cells were isolated from the bone marrow mononuclear cell fraction of two healthy donors by Ficoll density gradient centrifugation with subsequent magnetic bead separation using the Human CD34 Progenitor Cell Isolation Kit (Miltenyi Biotech Germany; #130-046-703). Isolated cells were frozen until further analysis. $CD34^+$ HSPCs at a concentration of $1 \times 10^4$ cells/mL were plated in 35 mm cell culture dishes in 1 mL Methocult H4230 medium (Stemcell Technologies) supplemented with 2% FBS, 10 µg/mL 100x Antibiotic-Antimycotic Solution (Sigma) 50 ng/mL SCF, 20 ng/mL IL-3 and 20 ng/mL rhG-CSF, or 100 ng/mL designed agonists. Cells were cultured at 37 °C and 5% $CO_2$. Colonies were counted on day 14.

***Evaluation of time-dependent effects of the designed agonists on the proliferation of CD34+ HSPCs***

$CD34^+$ HSPCs were isolated and processed as described above. Cells were incubated in poly-L-lysine–coated 96-well plates (2 × 104 cells/well) in Stemline II Hematopoietic Stem Cell Expansion medium (Sigma Aldrich; #50192) supplemented with 10% FBS, 1% penicillin/streptomycin, 1% L-glutamine and 50 ng/mL SCF, 20 ng/mL IL-3 and 10 ng/mL rhG-CSF, or designed agonists at different concentrations. using an IncuCyte S3 Live-Cell Analysis System (Essen Bio) with a 10 x objective at 37 °C and 5% $CO_2$. Cell proliferation over time was analyzed using IncuCyte S3 Software. Experiments in this study involving human samples were conducted according to Helsinki's declaration, and study approval was obtained from the Ethical Review Board of the Medical Faculty, University of Tübingen (012/2021BO2).

***RNA-sequencing analysis***

NFS-60 cells were washed two times with NFS-60-medium (RPMI 1640 medium with additional 1 mM L-glutamine, 1 mM Na-pyruvate, 10% FCS, 1% Antibiotic-Antimycotic (15240062, Gibco™)) and starved at 300,000 cells/mL for 16 h at 5% $CO_2$ and 37 °C. 1.5 million cells per sample in a T25 cell-culture flask were stimulated with 1 nM of the selected design or rhG-CSF for 8 h. Subsequently, the cells were washed twice with PBS, lysed in 350 µL RTL buffer and stored at -70 °C until RNA extraction. At least 500 ng of RNA per sample was used to prepare libraries for RNA sequencing. The RNA Integrity Number (RIN) for RNA quality assessment was measured using a 2100 Bioanalyzer (Agilent). All RNA samples showed RIN > 9,6 (max = 10), demonstrating high quality of the RNA samples. Directional mRNA library preparation (poly A enrichment) and sequencing were done by Novogene (https://www.novogene.com). The libraries were sequenced on an Illumina NovaSeq 6000 using paired-end 150 bp sequencing mode with a minimum of 6 Gb of raw data per sample. For data analysis, nf-core/rnaseq (https://nf-co.re/rnaseq, [92]) was used, an RNA sequencing analysis pipeline that takes a sample sheet and FASTQ files as input, performs quality assessment, trimming, alignment, and produces a gene count table and extensive QC report. Nf-core/rnaseq pipeline was run with the following parameters: --profile 'docker,' --genome 'GRCm38,' --aligner 'star_salmon,' pipeline version 3.10. Raw sequencing read counts were analyzed with the R package edgeR v3.42.2 [93]. 11,801 of initial 45,706 transcripts showed sufficient level of expression. Normalized log2 cpm values were used for PCA and descriptive part of the analysis. Differential expression was determined by fitting a quasi-

likelihood negative binomial generalized log-linear model to count data. Resulting candidate genes that passed thresholds of absolute log2 fold change > 1 and FDR < 0.05 were tested for gene set enrichment using the R package clusterProfiler v4.8.1 [94]. Two gene set collections were used: GO terms and MSigDB m2 mouse collection v2022.1.

***Acknowledgements***

This project has received funding from the IMPRS (TU, KM), the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No 863952 (ACE-OF-SPACE)) (PM), the M. Schickedanz Kinderkrebsstiftung (ME, JS, PM, NA), "German Universities Excellence Initiative" of the Tübingen University (JS), Deutsche Forschungsgemeinschaft (DFG; No 500215849; ME, BHA, JS, VH), (DFG; PI 405/15-2 – 326558201 and SFB; 1557, P13 – 467522186; JP). The authors would also like to thank Luis Ángel Fernández for providing the bacterial display system plasmid, and Regine Bernhard, Gabriele Hikade and Hella Kenneweg for technical assistance.

# References


1. Rochman, Y., R. Spolski, and W.J. Leonard, *New insights into the regulation of T cells by gamma(c) family cytokines.* Nature reviews. Immunology, 2009. **9**(7): p. 480–490.
2. Spangler, J.B., *et al.*, *Insights into cytokine-receptor interactions from cytokine engineering.* Annual review of immunology, 2015. **33**: p. 139–167.
3. Leonard, W.J. and J.X. Lin, *Cytokine receptor signaling pathways.* J Allergy Clin Immunol, 2000. **105**(5): p. 877-88.
4. Bezbradica, J.S. and R. Medzhitov, *Integration of cytokine and heterologous receptor signaling pathways.* Nature Immunology, 2009. **10**(4): p. 333-339.
5. Leonard, W.J. and J.X. Lin, *Strategies to therapeutically modulate cytokine action.* Nat Rev Drug Discov, 2023. **22**(10): p. 827-854.
6. Stroud, R.M. and J.A. Wells, *Mechanistic Diversity of Cytokine Receptor Signaling Across Cell Membranes.* Science's STKE, 2004. **2004**(231): p. re7.
7. Spangler, J.B., *et al.*, *Insights into Cytokine–Receptor Interactions from Cytokine Engineering.* Annual Review of Immunology, 2015. **33**(1): p. 139-167.
8. Wilmes, S., *et al.*, *Mechanism of homodimeric cytokine receptor activation and dysregulation by oncogenic mutations.* Science, 2020. **367**(6478): p. 643-652.
9. Glassman, C.R., *et al.*, *Structure of a Janus kinase cytokine receptor complex reveals the basis for dimeric activation.* Science, 2022. **376**(6589): p. 163-169.
10. Schreiber, G. and M.R. Walter, *Cytokine-receptor interactions as drug targets.* Curr Opin Chem Biol, 2010. **14**(4): p. 511-9.
11. Saxton, R.A., C.R. Glassman, and K.C. Garcia, *Emerging principles of cytokine pharmacology and therapeutics.* Nature reviews. Drug discovery, 2022.
12. Schreiber, G. and J. Piehler, *The molecular basis for functional plasticity in type I interferon signaling.* Trends Immunol, 2015. **36**(3): p. 139-149.
13. Heike, T. and T. Nakahata, *Ex vivo expansion of hematopoietic stem cells by cytokines.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2002. **1592**(3): p. 313-321.
14. Mitra, S., *et al.*, *Interleukin-2 Activity Can Be Fine Tuned with Engineered Receptor Signaling Clamps.* Immunity, 2015. **42**(5): p. 826-838.
15. Saxton, R.A., *et al.*, *The tissue protective functions of interleukin-22 can be decoupled from pro-inflammatory actions through structure-based design.* Immunity, 2021. **54**(4): p. 660-672.e9.
16. Wilmes, S., *et al.*, *Receptor dimerization dynamics as a regulatory valve for plasticity of type I interferon signaling.* Journal of Cell Biology, 2015. **209**(4): p. 579-593.
17. Moraga, I., *et al.*, *Multifarious determinants of cytokine receptor signaling specificity.* Adv Immunol, 2014. **121**: p. 1-39.
18. Syed, R.S., *et al.*, *Efficiency of signalling through cytokine receptors depends critically on receptor orientation.* Nature, 1998. **395**(6701): p. 511-516.
19. Staerk, J., *et al.*, *Orientation-specific signalling by thrombopoietin receptor dimers.* The EMBO Journal, 2011. **30**(21): p. 4398-4413.
20. Wilson, I.A. and L.K. Jolliffe, *The structure, organization, activation and plasticity of the erythropoietin receptor.* Current Opinion in Structural Biology, 1999. **9**(6): p. 696-704.
21. Ballinger, M.D. and J.A. Wells, *Will any dimer do?* Nature Structural Biology, 1998. **5**(11): p. 938-940.
22. Mohan, K., *et al.*, *Topological control of cytokine receptor signaling induces differential effects in hematopoiesis.* Science, 2019. **364**(6442).
23. Moraga, I., *et al.*, *Tuning cytokine receptor signaling by re-orienting dimer geometry with surrogate ligands.* Cell, 2015. **160**(6): p. 1196–1208.
24. Yen, M., *et al.*, *Facile discovery of surrogate cytokine agonists.* Cell, 2022. **185**(8): p. 1414-1430.e19.
25. Dwivedi, P. and K.D. Greis, *Granulocyte colony-stimulating factor receptor signaling in severe congenital neutropenia, chronic neutrophilic leukemia, and related malignancies.* Experimental hematology, 2017. **46**: p. 9–20.

26. O'Shea, J.J. and R. Plenge, *JAK and STAT signaling molecules in immunoregulation and immune-mediated disease.* Immunity, 2012. **36**(4): p. 542–550.
27. Palande, K., *et al.*, *Scratching the surface: signaling and routing dynamics of the CSF3 receptor.* Frontiers in bioscience (Landmark edition), 2013. **18**(1): p. 91–105.
28. Aritomi, M., *et al.*, *Atomic structure of the GCSF-receptor complex showing a new cytokine-receptor recognition scheme.* Nature, 1999. **401**(6754): p. 713–717.
29. Mine, S., *et al.*, *Thermodynamic analysis of the activation mechanism of the GCSF receptor induced by ligand binding.* Biochemistry, 2004. **43**(9): p. 2458–2464.
30. Tamada, T., *et al.*, *Homodimeric cross-over structure of the human granulocyte colony-stimulating factor (GCSF) receptor signaling complex.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(9): p. 3135–3140.
31. Franzke, A., *The role of G-CSF in adaptive immunity.* Cytokine & growth factor reviews, 2006. **17**(4): p. 235-244.
32. Welte, K., *et al.*, *Filgrastim (r-metHuG-CSF): The First 10 Years.* Blood The Journal of The American Society of Hematology, 1996. **88**(6): p. 1907–1929.
33. Welte, K., *et al.*, *Purification and biochemical characterization of human pluripotent hematopoietic colony-stimulating factor.* Proceedings of the National Academy of Sciences of the United States of America, 1985. **82**(5): p. 1526–1530.
34. Morris, R., N.J. Kershaw, and J.J. Babon, *The molecular details of cytokine signaling via the JAK/STAT pathway.* Protein science : a publication of the Protein Society, 2018. **27**(12): p. 1984–2009.
35. Tian, S.-S., *et al.*, *Multiple signaling pathways induced by granulocyte colony-stimulating factor involving activation of JAKs, STAT5, and/or STAT3 are required for regulation of three distinct classes of immediate early genes.* 1996.
36. Shimozaki, K., *et al.*, *Involvement of STAT3 in the granulocyte colony-stimulating factor-induced differentiation of myeloid cells.* Journal of Biological Chemistry, 1997. **272**(40): p. 25184-25189.
37. Dong, F., *et al.*, *Stimulation of Stat5 by granulocyte colony-stimulating factor (G-CSF) is modulated by two distinct cytoplasmic regions of the G-CSF receptor.* The Journal of Immunology, 1998. **161**(12): p. 6503-6509.
38. Ward, A.C., *et al.*, *Multiple signals mediate proliferation, differentiation, and survival from the granulocyte colony-stimulating factor receptor in myeloid 32D cells.* Journal of Biological Chemistry, 1999. **274**(21): p. 14956-14962.
39. Chakraborty, A. and D.J. Tweardyabc, *Stat3 and G-CSF-induced myeloid differentiation.* Leukemia & lymphoma, 1998. **30**(5-6): p. 433-442.
40. Ward, A.C., *et al.*, *Defective internalization and sustained activation of truncated granulocyte colony-stimulating factor receptor found in severe congenital neutropenia/acute myeloid leukemia.* Blood, The Journal of the American Society of Hematology, 1999. **93**(2): p. 447-458.
41. Ward, A.C., *et al.*, *Novel point mutation in the extracellular domain of the granulocyte colony-stimulating factor (G-CSF) receptor in a case of severe congenital neutropenia hyporesponsive to G-CSF treatment.* The Journal of experimental medicine, 1999. **190**(4): p. 497-508.
42. Skokowa, J., *et al.*, *A topological refactoring design strategy yields highly stable granulopoietic proteins.* Nature communications, 2022. **13**(1): p. 2948.
43. Maksymenko, K., *et al.*, *The design of functional proteins using tensorized energy calculations.* Cell Reports Methods, 2023. **3**(8): p. 100560.
44. Young, D.C., *et al.*, *Characterization of the receptor binding determinants of granulocyte colony stimulating factor.* Protein Science, 1997. **6**(6): p. 1228-1236.
45. Salema, V., *et al.*, *Selection of single domain antibodies from immune libraries displayed on the surface of E. coli cells with two β-domains of opposite topologies.* PloS one, 2013. **8**(9): p. e75126.
46. Weinstein, Y., *et al.*, *Truncation of the c-myb gene by a retroviral integration in an interleukin 3-dependent myeloid leukemia cell line.* Proceedings of the National Academy of Sciences, 1986. **83**(14): p. 5010-5014.

47. Nakoinz, I., *et al.*, *Differentiation of the IL-3-dependent NFS-60 cell line and adaption to growth in macrophage colony-stimulating factor*. Journal of immunology (Baltimore, Md.: 1950), 1990. **145**(3): p. 860-864.
48. Renshaw, S.A., *et al.*, *A transgenic zebrafish model of neutrophilic inflammation.* Blood, 2006. **108**(13): p. 3976-8.
49. Hernandez Alvarez, B., *et al.*, *Design of novel granulopoietic proteins by topological rescaffolding.* PLOS Biology, 2020. **18**(12): p. e3000919.
50. Jumper, J., *et al.*, *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
51. Tsirigotaki, A., *et al.*, *Mechanism of receptor assembly via the pleiotropic adipokine Leptin.* Nat Struct Mol Biol, 2023. **30**(4): p. 551-563.
52. Whitty, A. and C.W. Borysenko, *Small molecule cytokine mimetics.* Chem Biol, 1999. **6**(4): p. R107-18.
53. Lamken, P., *et al.*, *Ligand-induced assembling of the type I interferon receptor on supported lipid bilayers.* J Mol Biol, 2004. **341**(1): p. 303-18.
54. Whitty, A., *et al.*, *Interaction affinity between cytokine receptor components on the cell surface.* Proc Natl Acad Sci U S A, 1998. **95**(22): p. 13165-70.
55. Salavessa, L., *et al.*, *Cytokine receptor cluster size impacts its endocytosis and signaling.* Proc Natl Acad Sci U S A, 2021. **118**(37).
56. Park, S.D., *et al.*, *A review of granulocyte colony-stimulating factor receptor signaling and regulation with implications for cancer*. Frontiers in Oncology, 2022. **12**: p. 932608.
57. Touw, I.P. and G.-J.M. van de Geijn, *Granulocyte colony-stimulating factor and its receptor in normal myeloid cell development, leukemia and related blood cell disorders.* Frontiers in Bioscience, 2007. **12**(1): p. 800–815.
58. Acres, B., *et al.*, *Fusokine interleukin-2/interleukin-18, a novel potent innate and adaptive immune stimulator with decreased toxicity.* Cancer Res, 2005. **65**(20): p. 9536-46.
59. Bhatt, S., *et al.*, *Anti-CD20-interleukin-21 fusokine targets malignant B cells via direct apoptosis and NK-cell-dependent cytotoxicity.* Blood, 2017. **129**(16): p. 2246-2256.
60. Deng, J., *et al.*, *GM-CSF and IL-4 Fusion Cytokine Induces B Cell-Dependent Hematopoietic Regeneration.* Mol Ther, 2017. **25**(2): p. 416-426.
61. Deng, J., *et al.*, *GIFT4 fusokine converts leukemic B cells into immune helper cells.* J Transl Med, 2016. **14**(1): p. 106.
62. Deng, J., *et al.*, *Engineered fusokine GIFT4 licenses the ability of B cells to trigger a tumoricidal T-cell response.* Cancer Res, 2014. **74**(15): p. 4133-44.
63. Rafei, M., *et al.*, *Selective inhibition of CCR2 expressing lymphomyeloid cells in experimental autoimmune encephalomyelitis by a GM-CSF-MCP1 fusokine.* J Immunol, 2009. **182**(5): p. 2620-7.
64. Rafei, M., *et al.*, *A granulocyte-macrophage colony-stimulating factor and interleukin-15 fusokine induces a regulatory B cell population with immune suppressive properties.* Nat Med, 2009. **15**(9): p. 1038-45.
65. Rafei, M., *et al.*, *A GMCSF and IL-15 fusokine leads to paradoxical immunosuppression in vivo via asymmetrical JAK/STAT signaling through the IL-15 receptor complex.* Blood, 2007. **109**(5): p. 2234-42.
66. Van der Jeught, K., *et al.*, *Intratumoral administration of mRNA encoding a fusokine consisting of IFN-β and the ectodomain of the TGF-β receptor II potentiates antitumor immunity.* Oncotarget, 2014. **5**(20): p. 10100-13.
67. Moraga, I., *et al.*, *Synthekines are surrogate cytokine and growth factor agonists that compel signaling through non-natural receptor dimers.* eLife, 2017. **6**: p. e22882.
68. ElGamacy, M., *Accelerating therapeutic protein design.* Adv Protein Chem Struct Biol, 2022. **130**: p. 85-118.
69. Bishop, B., *et al.*, *Reengineering granulocyte colony-stimulating factor for enhanced stability.* J Biol Chem, 2001. **276**(36): p. 33465-70.
70. Shibuya, R., T. Miyafusa, and S. Honda, *Stabilization of backbone-circularized protein is attained by synergistic gains in enthalpy of folded structure and entropy of unfolded structure.* The FEBS Journal, 2020. **287**(8): p. 1554-1575.

71. Wood, V.E., *et al.*, *Protein Engineering and HDX Identify Structural Regions of G-CSF Critical to Its Stability and Aggregation.* Mol Pharm, 2022. **19**(2): p. 616-629.
72. Ward, A.C., *et al.*, *Tyrosine-dependent and-independent mechanisms of STAT3 activation by the human granulocyte colony-stimulating factor (G-CSF) receptor are differentially utilized depending on G-CSF concentration.* Blood, The Journal of the American Society of Hematology, 1999. **93**(1): p. 113-124.
73. Gonnord, P., C.M. Blouin, and C. Lamaze, *Membrane trafficking and signaling: two sides of the same coin.* Semin Cell Dev Biol, 2012. **23**(2): p. 154-64.
74. Cendrowski, J., A. Maminska, and M. Miaczynska, *Endocytic regulation of cytokine receptor signaling.* Cytokine Growth Factor Rev, 2016. **32**: p. 63-73.
75. Silva, D.A., *et al.*, *De novo design of potent and selective mimics of IL-2 and IL-15.* Nature, 2019. **565**(7738): p. 186-191.
76. Yang, H., *et al.*, *Design of cell-type-specific hyperstable IL-4 mimetics via modular de novo scaffolds.* Nat Chem Biol, 2023. **19**(9): p. 1127-1137.
77. Cui, L., *et al.*, *Tuning MPL signaling to influence hematopoietic stem cell differentiation and inhibit essential thrombocythemia progenitors.* Proceedings of the National Academy of Sciences, 2021. **118**(2): p. e2017849118.
78. Tsutsumi, N., *et al.*, *Structure of the thrombopoietin-MPL receptor complex is a blueprint for biasing hematopoiesis.* Cell.
79. Franzke, A., *et al.*, *G-CSF as immune regulator in T cells expressing the G-CSF receptor: implications for transplantation and autoimmune diseases.* Blood, 2003. **102**(2): p. 734-739.
80. Karagiannidis, I., *et al.*, *G-CSF in tumors: Aggressiveness, tumor microenvironment and immune cell regulation.* Cytokine, 2021. **142**: p. 155479.
81. Tu, Q., *et al.*, *Room temperature electrocompetent bacterial cells improve DNA transformation and recombineering efficiency.* Scientific reports, 2016. **6**(1): p. 1-8.
82. Jacobs, T.M., *et al.*, *SwiftLib: rapid degenerate-codon-library optimization through dynamic programming.* Nucleic acids research, 2015. **43**(5): p. e34-e34.
83. Virtanen, P., *et al.*, *SciPy 1.0: fundamental algorithms for scientific computing in Python.* Nature methods, 2020. **17**(3): p. 261-272.
84. Gotzke, H., *et al.*, *The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications.* Nat Commun, 2019. **10**(1): p. 4403.
85. Wilmes, S., *et al.*, *Receptor dimerization dynamics as a regulatory valve for plasticity of type I interferon signaling.* J Cell Biol, 2015. **209**(4): p. 579-93.
86. Sotolongo Bellon, J., *et al.*, *Four-color single-molecule imaging with engineered tags resolves the molecular architecture of signaling complexes in the plasma membrane.* Cell reports methods, 2022. **2**(2): p. 100165.
87. You, C., *et al.*, *Dynamic submicroscopic signaling zones revealed by pair correlation tracking and localization microscopy.* Anal Chem, 2014. **86**(17): p. 8593-602.
88. Serge, A., *et al.*, *Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes.* Nature methods, 2008. **5**(8): p. 687-94.
89. Niewidok, B., *et al.*, *Single-molecule imaging reveals dynamic biphasic partition of RNA-binding proteins in stress granules.* The Journal of cell biology, 2018. **217**(4): p. 1303-1318.
90. Jaqaman, K., *et al.*, *Robust single-particle tracking in live-cell time-lapse sequences.* Nature methods, 2008. **5**(8): p. 695-702.
91. Schindelin, J., *et al.*, *Fiji: an open-source platform for biological-image analysis.* Nature methods, 2012. **9**(7): p. 676-682.
92. Ewels, P.A., *et al.*, *The nf-core framework for community-curated bioinformatics pipelines.* Nat Biotechnol, 2020. **38**(3): p. 276-278.
93. Chen, Y., A.T. Lun, and G.K. Smyth, *From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline.* F1000Res, 2016. **5**: p. 1438.
94. Yu, G., *et al.*, *clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters.* OMICS: A Journal of Integrative Biology, 2012. **16**(5): p. 284-287.
95. Nguyen-Jackson, H.T., H. Zhang, and S.S. Watowich, *G-CSF Receptor Structure, Function, and Intracellular Signal Transduction*, in *Twenty Years of G-CSF: Clinical and Nonclinical*

*Discoveries*, G. Molineux, M. Foote, and T. Arvedson, Editors. 2012, Springer Basel: Basel. p. 83-105.

Figures

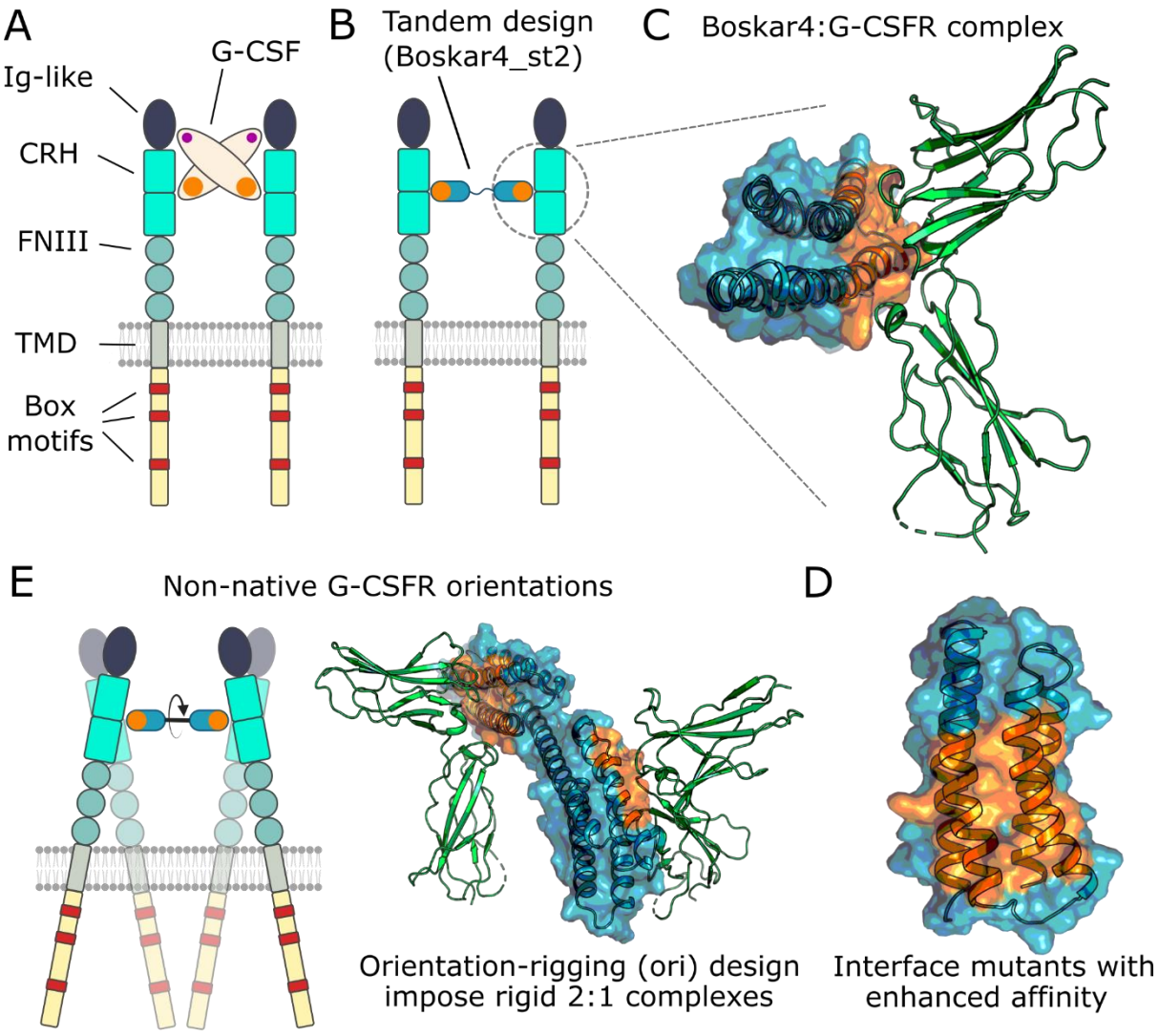



**Figure 1. Overall strategy to generate novel G-CSFR modulators with varying affinities and geometries. (A)** G-CSF (beige oval) activates G-CSFR by dimerizing the Ig-like and CRH domains (dark blue and cyan) through two distinct receptor-binding sites (orange and purple) [95]. **(B)** In contrast, the *de novo*-designed Boskar4_st2 agonist consists of two tandemly repeated Boskar4 domains, which **(C)** possess minimal architecture and encode a single high-affinity, receptor-binding site (blue domain and orange surface patch). **(D)** On the Boskar4 module, the residues forming the receptor-binding entity were diversified in order to identify affinity-enhanced variants. **(E)** The improved Boskar4 module was fused into rigid, orientation-rigging designs (oris) that can dimerize and activate G-CSFR in non-native geometries. The purple box shows a modeled structure of one example ori binding to two G-CSFR.

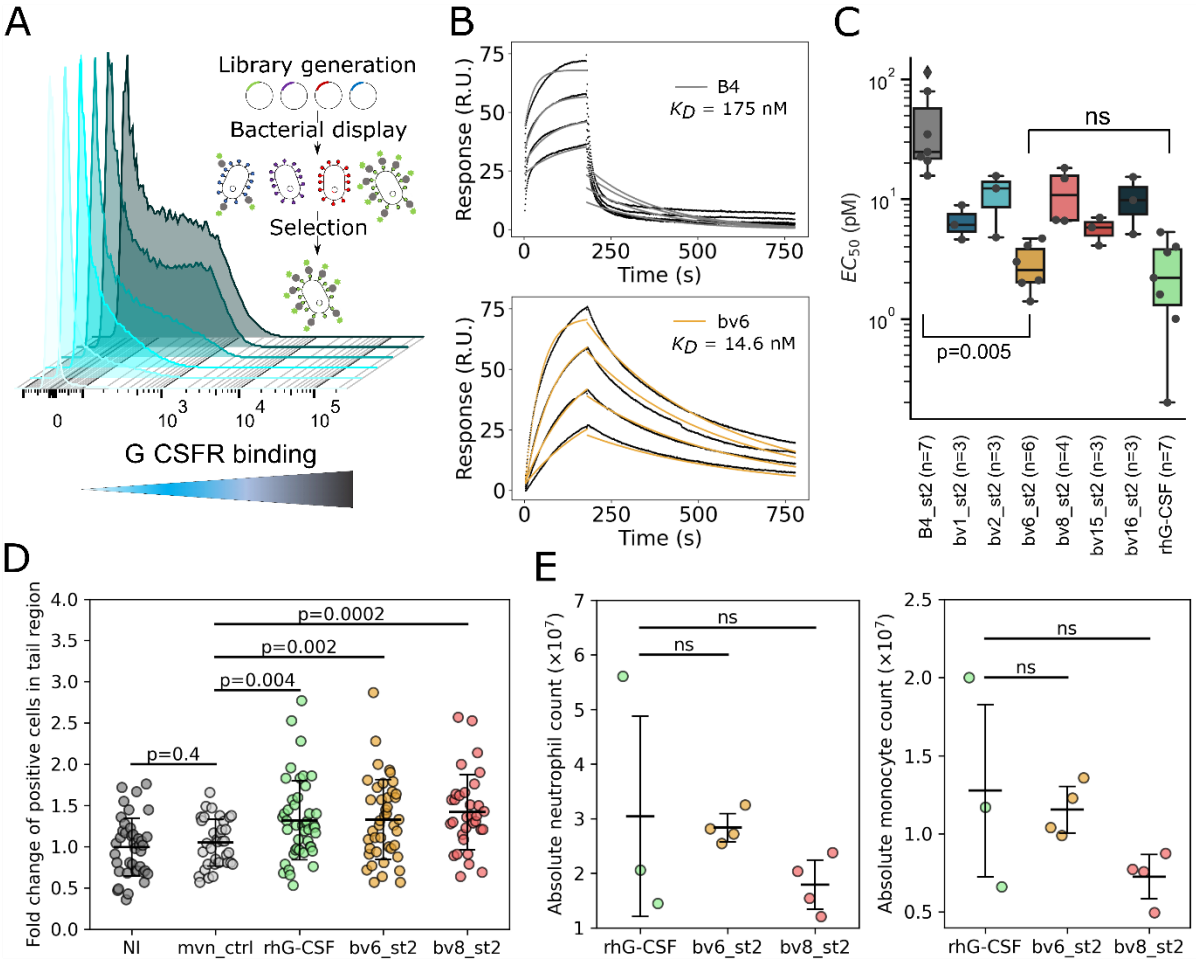



**Figure 2. Affinity-enhanced binders exhibit improved granulopoietic activity *in vitro* and *in vivo*.** **(A)** Cell sorting of bacterially-displayed Boskar4 mutants bound to fluorescently labeled G-CSFR was conducted over 5 steps of enrichment. Successive fluorescence shifts are shown as light cyan to dark gray distributions. **(B)** SPR titrations show single-domain bv6 to bind 12-fold tighter than its Boskar4 counterpart (Fig. S7D and Table 1). **(C)** Proliferative activity assays in NFS-60 cells showed the enhanced activity ($EC_{50}$) of all variants in comparison to Boskar4_st2, where the most active variant bv6_st2 was not significantly different in activity compared to the native ligand (rhG-CSF; lightgreen box). For statistical analysis a one-way ANOVA over all samples was performed followed by a Tukey HSD test. Proliferation assays were conducted in at least three ($3 \leq n \leq 7$) independent experiments per variant or rhG-CSF (Table 1 and Fig. S8). **(D)** Interval plots of the fluorescent neutrophils in transgenic *Tg:mpxGFP* zebrafish larvae that were either not injected (NI), injected with the inactive protein Moevan_control (mvn_ctrl) [49], rhG-CSF, or indicated Boskar4 variants (bv6_st2 or bv8_st2) after 24 h of treatment. Data shows mean ± standard deviation, where each circle represents one zebrafish larva (NI n=40, mvn_ctr n=35, rhG-CSF n=44, bv6_st2 n=42, bv8_st2 n=32). Statistical analysis was carried out by ordinary one-way ANOVA followed by a Tukey HSD test. **(E)** C57BL/6 Ly5.1 mice were treated (i.p.) with rhG-CSF, bv6_st2, or bv8_st2 (each circle indicates one mouse). Absolute cell counts of mouse neutrophils ($DAPI^-CD45^+Ly6G^+Ly6C^+$, left image) and monocytes ($DAPI^-CD45^+CD19^-CD3^-CD11b^+Ly6C^+Ly6G^-$, right image) in the bone marrow of treated mice are shown. Statistical analysis was carried out by ordinary one-way ANOVA with a multiple comparisons test of all experimental drugs to the G-CSF as a positive control (rhG-CSF, bv6_st2 n=4, bv8_st2 n=4; ns: not significant).

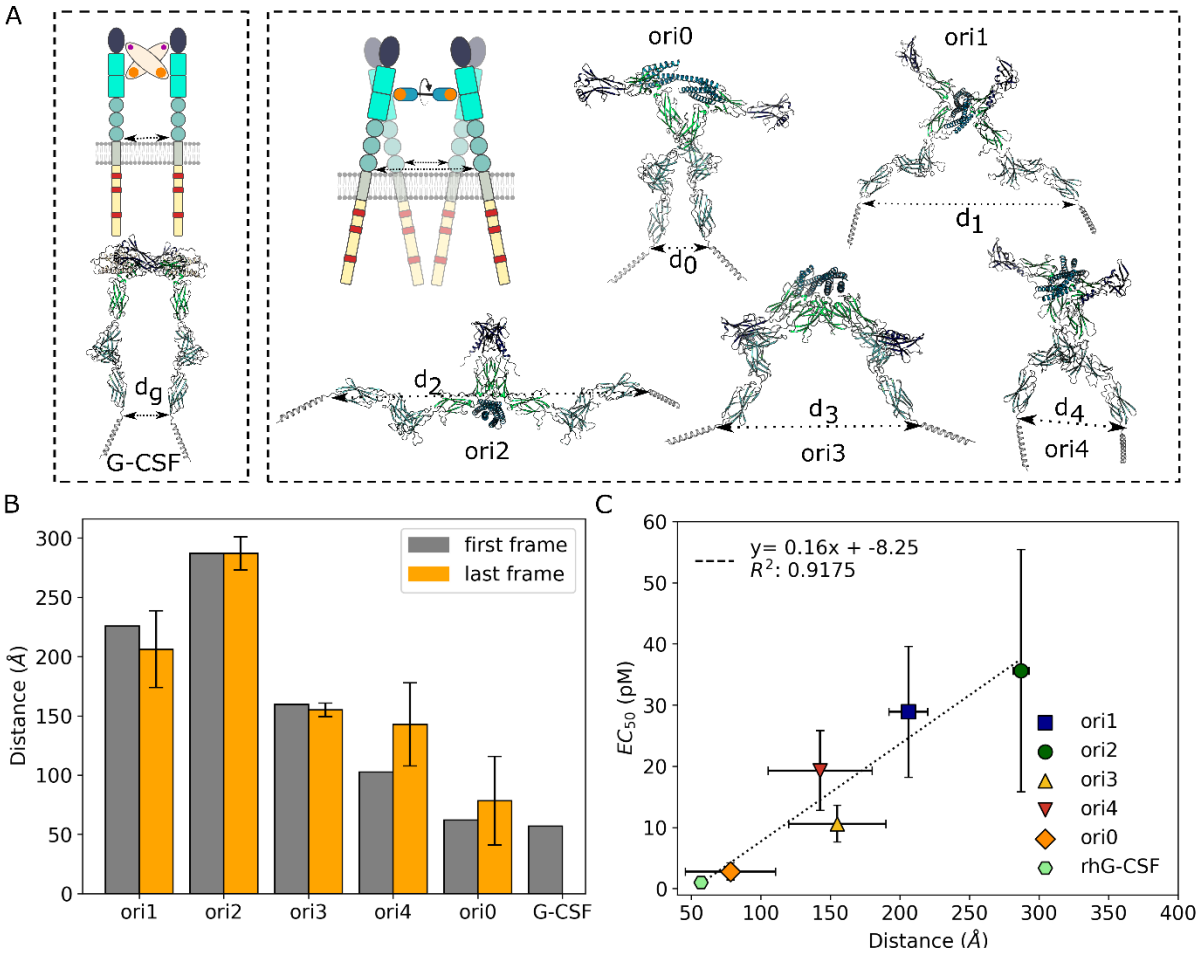



**Figure 3. Agonists designed to associate G-CSFR in non-native dimeric geometries yield distinct cell activation potencies. (A)** Unlike the native 2:2 ligand-receptor assembly induced by G-CSF (PDB: 2D9Q), we developed a range of novel ligands by rigidly connecting two copies of the affinity-enhanced bv6 module (Materials and Methods; Fig. S11) that form 1:2 ligand-receptor complexes. Modeled complexes of these orientation-rigging designs (ori0 through ori4) and G-CSFR indicate varying transmembrane domain spacings ($d_x$). Ori0 was designed to closely mimic this distance parameter to that of G-CSF (i.e., $d_0 \approx d_g$). **(B)** Molecular dynamics simulations performed on a selected set of designed ori candidates confirmed the rigidity of the design linkers. Orange bars represent the mean spacing and its standard deviation from 5 simulation replicas. **(C)** Proliferation assays in NFS-60 cells of the five selected ori designs and rhG-CSF (Lenograstim) showed the proliferative $EC_{50}$ (mean and standard deviation of at least 5 independent experiments) to strongly correlate to the modelled transmembrane domain spacing ($d_x$) for the different proteins. The dotted line represents the linear fit between the obtained $EC_{50}$ and the distances of the last frame (where the $d_g$ of G-CSF was obtained as a single value from the crystal structure (PDB:2D9Q).

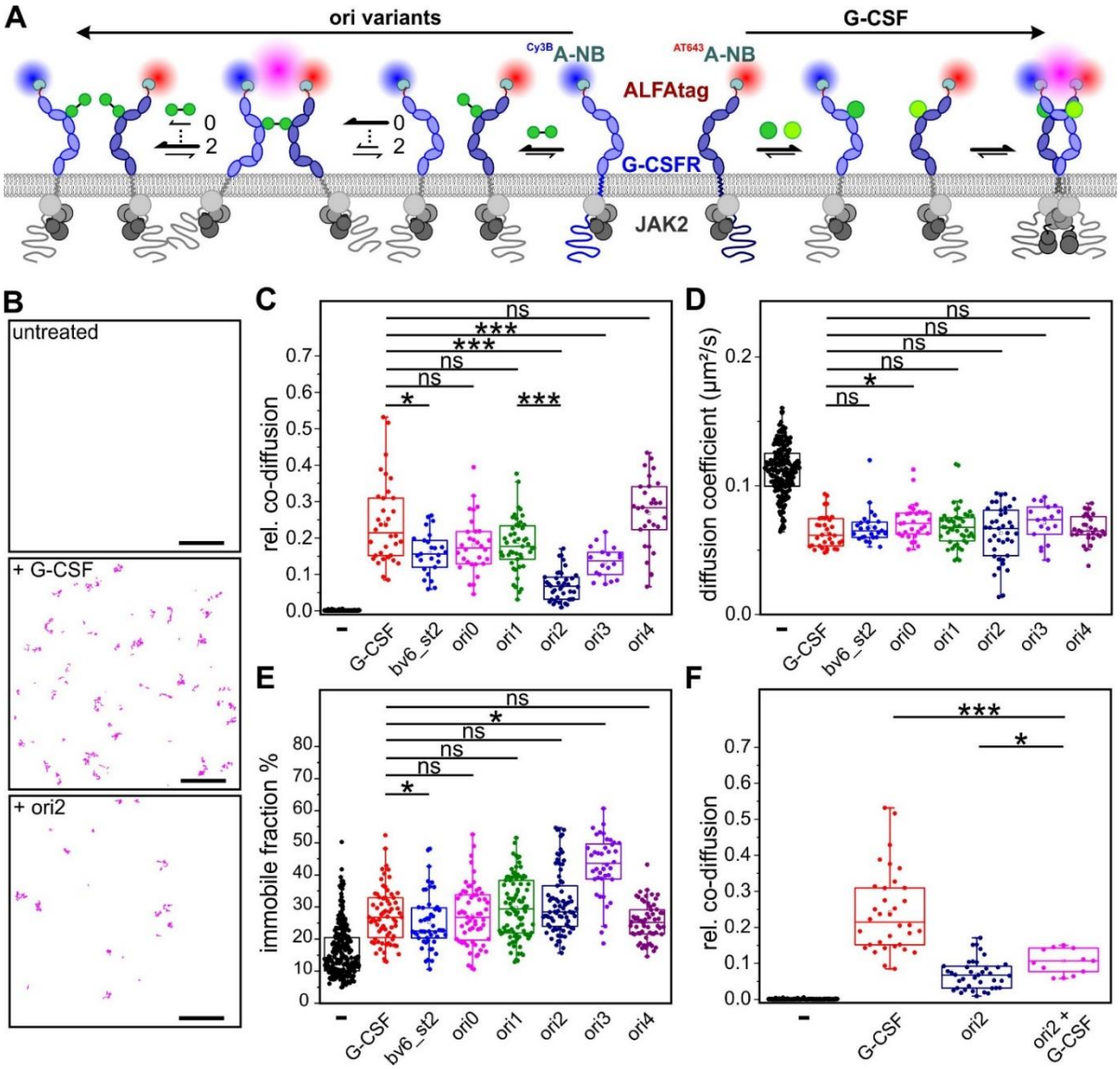



**Figure 4. The design agonists differentially dimerize G-CSFR subunits in live cells. (A, B)** Dual-color single molecule co-tracking for quantifying G-CSFR dimerization in live cells. **(A)** G-CSFR labeled via an N-terminal ALFA-tag using a mixture of nanobodies conjugated with Cy3B and AT643, respectively. The G-CSFR monomer-dimer equilibrium probed by dual-color single molecule co-tracking depends on the two-dimensional and three-dimensional binding affinities of the respective design agonist. Different dimerization principles by G-CSF (arrow to the right) and design agonists arrow to the left), as well as the consequences of altered orientation on the assembly kinetics are schematically outlined. (numbers on arrows indicate expected impact of ori0 (0) or ori2 (2) **(B)** Typical single molecule co-trajectories of G-CSFR in the absence of agonist and presence of G-CSF and ori0, respectively. A ROI from a single cell is shown. Scale bar: 5 µm. **(C)** Relative dimerization levels of G-CSFR in the absence of ligand and in the presence of G-CSF and design agonists. **(D)** Diffusion constants of G-CSFR monomers in the absence of ligand compared to G-CSFR dimers in the presence of G-CSF and design agonists identified by co-tracking analysis. **(E)** Comparison of the immobile fraction of G-CSFR in the absence of ligand and in the presence of G-CSF and design agonists. **(F)** Comparison of G-CSFR dimerization by G-CSFR by G-CSF, ori2 and an equimolar mixture of both. Box plots in C-F were obtained from analysis of multiple cells (For C-E; Numbers of cells and experiments, respectively, are 124 and 17, 34 and 3, 25 and 2, 30 and 2, 47 and 3, 39 and 3, 18 and 2, 30 and 2 for each column, going from left to right. For F; Numbers of cells and experiments, respectively, are 124 and 17, 34 and 3, 39 and 3, 10 and 1 for each column, going from left to right), with each dot representing the result from one cell. Box plots indicate data distribution of the second

and third quartiles (box), median (line), mean (square) and 1.5× interquartile range (whiskers). Statistics were performed using two sided two-sample Kolmogorov-Smirnov test (not significant: ns;*P ≤0.05***P ≤ 0.001)

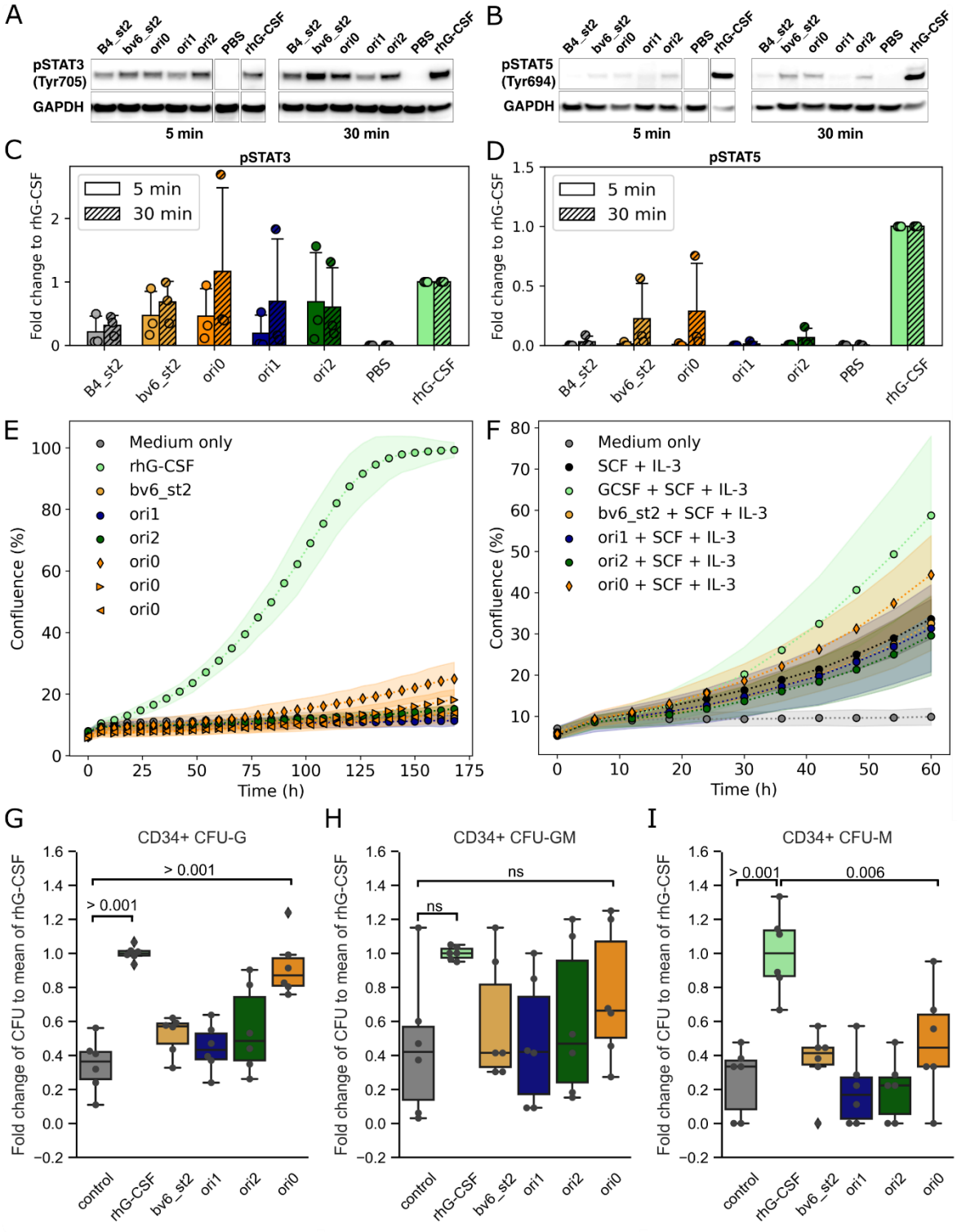



**Figure 5. The affinity and geometry of the design agonists bias intracellular signaling and primary stem cell differentiation. (A, B)** Representative western blot images of intracellular levels of phospho-STAT3 (Tyr705) and STAT3 (A) and phospho-STAT5 (Tyr694) and STAT5 (B) proteins after treatment of NFS-60 cells with 1 nM of different designs (saturating conditions) for 5 min (left pane) or 30 min (right pane). Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) staining was used as a loading control. Three biological replicates of this experiment showed the same trend (Fig. S17). **(C, D)** The pSTAT3 and pSTAT5 levels were quantified after 5 min and 30 min from those replicates and are represented in (C) and (D), respectively. Shown is the fold change to the GAPDH normalized rhG-CSF signal. **(E, F)** To probe the effect of the agonist designs (100 ng/ml) and rhG-CSF (10 ng/ml) on

healthy donor´s $CD34^+$ HSPCs, proliferation assays were performed without (E) or with the addition of 50 ng/mL SCF and 20 ng/mL IL-3 (F). Ori0 without SCF and IL-3 was tested at three concentrations 100 ng/mL (diamond), 10 ng/mL (triangle-right), and 1 ng/mL (triangle-left). Shown is the mean (points) and standard deviation (shades) of three parallel replicates. **(G, H, I)** Additionally, we performed CFU assays of healthy donor´s $CD34^+$ HSPCs incubated on semi-solid medium supplemented with corresponding cytokines and design agonists. Shown is the fold change to rhG-CSF of the quantified colony-forming units (CFU) of granulocytes (G-CFU) (G), granulocyte-macrophages (GM-CFU) (H), and macrophages (M-CFU) (I). The circles represent the obtained values for each condition of three independent experiments with two parallel replicates each. The indicated p-values were estimated by an ordinary one-way ANOVA followed by a Tukey HSD test.

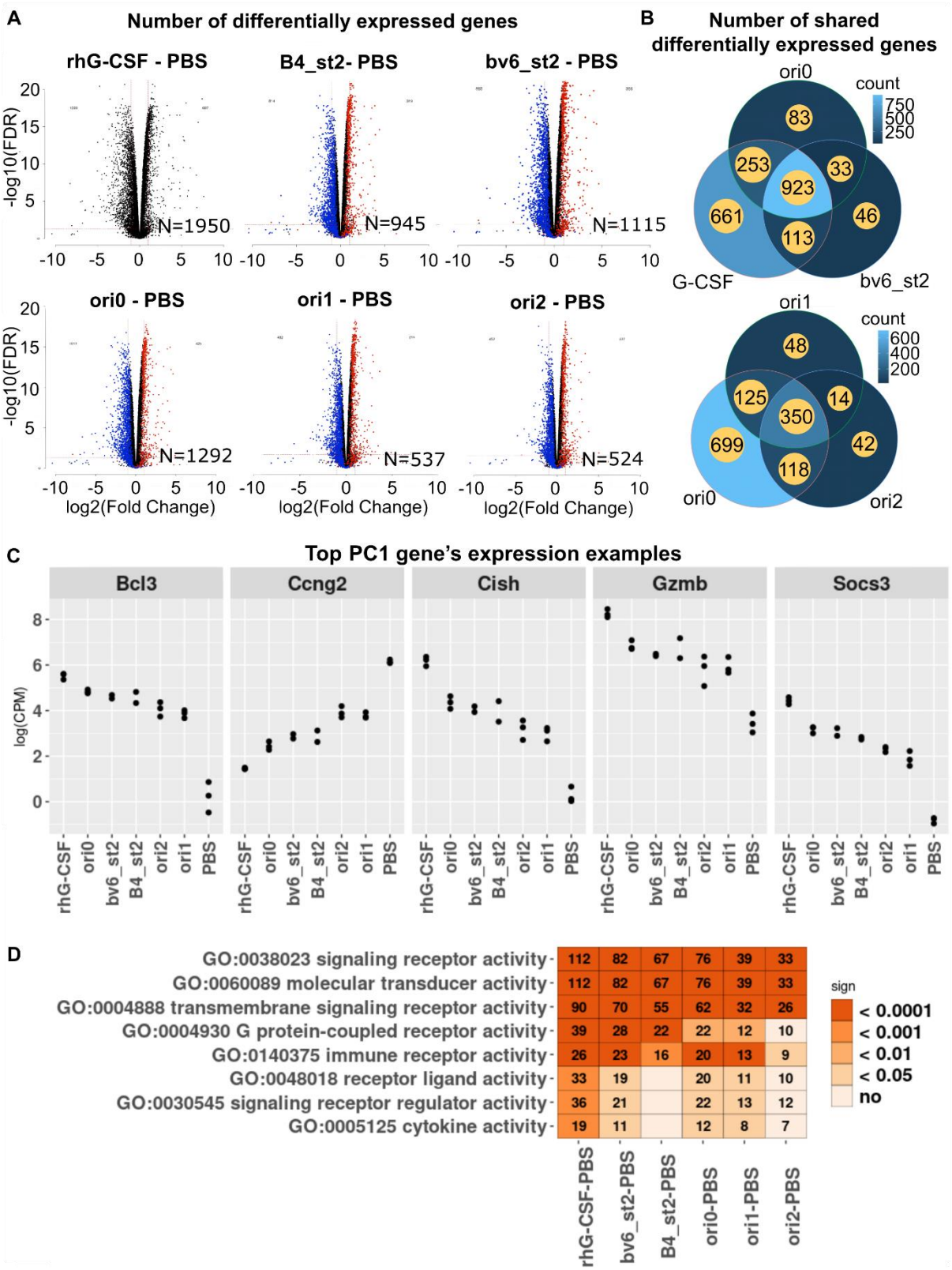



**Figure 6. The design agonists regulate the key target genes to varying degrees. (A)** The number (N) of differentially expressed genes in NFS-60 cells after treatment with rhG-CSF or different designs compared to PBS treated control. Blue (downregulated genes) and red (upregulated genes) dots in the volcano plots indicate genes that are shared between the corresponding design and rhG-CSF. **(B)** The Venn diagrams show the numbers of shared and design-specific regulated genes. **(C)** Examples of the top 20 differentially expressed hematopoiesis-related genes in the first principal component (PC1) of

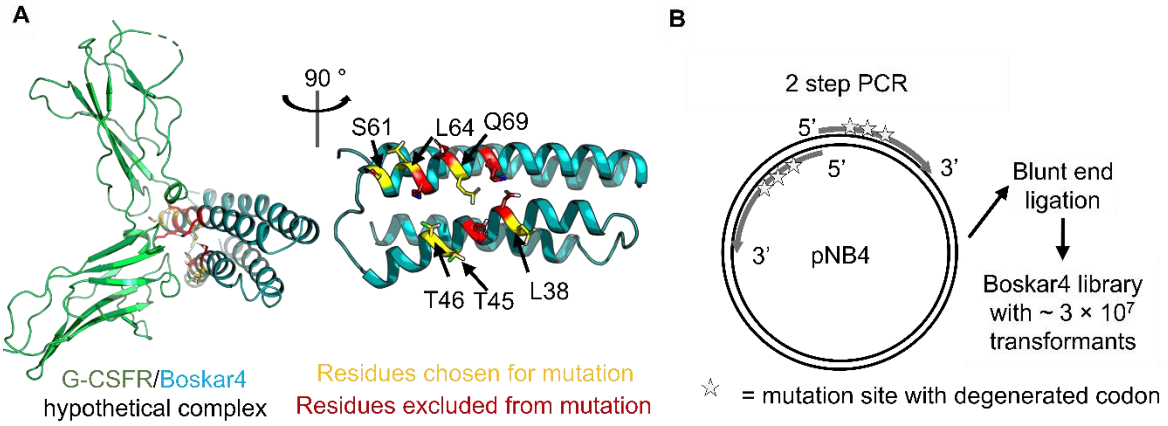
the principal component analysis (PCA). **(D)** Gene Ontology (GO) analysis of significantly regulated signaling pathways in the indicated groups. Each number indicates the number of differentially expressed genes related to the corresponding pathway.

# Tables

**Table 1. Binding affinity and activity parameters determined for G-CSFR modulators.** The association rate constant ($k_a$), the dissociation rate constant ($k_d$) and the apparent equilibrium dissociation constant ($K_D$) were obtained by surface plasmon resonance (SPR) against immobilized hG-CSF receptor (compare Fig. S6, S7, and S14). The mean and standard deviation (sd) of the half-maximal effective concentration $EC_{50}$ (pM) was obtained from NFS-60 activity assays of at least 3 biological replicates (compare Fig. S8 and S15). * The SPR parameters of rhG-CSF (self-made purified from *E. coli*) were taken from Skokowa *et al.* 2022 [42].

| | $k_a$ ($M^{-1}$ $s^{-1}$) | $k_d$ ($s^{-1}$) | Apparent $K_D$ (M) | Chi² ($RU^2$) | $EC_{50}$ (pM) mean ± sd |
|---|---|---|---|---|---|
| **Boskar4** | $(3.4 \pm 1.7) \times 10^{4}$ | $(5.4 \pm 1.4) \times 10^{-3}$ | $(1.7 \pm 0.4) \times 10^{-7}$ | 13.9 | NA |
| **bv6** | $(2.4 \pm 0.7) \times 10^{5}$ | $(3.3 \pm 0.4) \times 10^{-3}$ | $(1.4 \pm 0.2) \times 10^{-8}$ | 2.51 | NA |
| **Boskar4_st2** | $(2.3 \pm 1.0) \times 10^{6}$ | $(3.3 \pm 0.7) \times 10^{-3}$ | $(1.6 \pm 0.3) \times 10^{-9}$ | 0.678 | 33.1 ± 23.6 |
| **bv1_st2** | $(9.7 \pm 2.0) \times 10^{5}$ | $(2.2 \pm 0.2) \times 10^{-4}$ | $(2.4 \pm 0.6) \times 10^{-10}$ | 1.44 | 6.5 ± 2.2 |
| **bv2_st2** | $(3.5 \pm 1.3) \times 10^{5}$ | $(1.7 \pm 0.1) \times 10^{-4}$ | $(5.3 \pm 1.8) \times 10^{-10}$ | 0.371 | 10.9 ± 5.5 |
| **bv6_st2** | $(4.9 \pm 0.9) \times 10^{5}$ | $(5.9 \pm 0.5) \times 10^{-4}$ | $(6.4 \pm 1.0) \times 10^{-10}$ | 1.16 | 2.9 ± 1.3 |
| **bv8_st2** | $(1.8 \pm 0.4) \times 10^{6}$ | $(1.5 \pm 0.1) \times 10^{-3}$ | $(8.3 \pm 1.6) \times 10^{-10}$ | 1.36 | 11.6 ± 5.9 |
| **bv15_st2** | $(7.4 \pm 5.0) \times 10^{5}$ | $(4.0 \pm 0.1) \times 10^{-4}$ | $(6.5 \pm 2.1) \times 10^{-10}$ | 0.982 | 5.6 ± 1.5 |
| **bv16_st2** | $(8.1 \pm 0.1) \times 10^{5}$ | $(7.1 \pm 0.8) \times 10^{-4}$ | $(8.8 \pm 1.2) \times 10^{-10}$ | 0.473 | 10.1 ± 5.1 |
| **rhG-CSF (self-made from *E. coli*)*** | $(3.0 \pm 0.3) \times 10^{5}$ | $(4.9 \pm 2.8) \times 10^{-4}$ | $(1.1 \pm 1.6) \times 10^{-9}$ | 4.9 | 2.6 ± 1.8 |
| **Lenograstim (Glycosylated, commercial rhG-CSF)** | NA | NA | NA | NA | 1.0 ± 0.7 |
| **ori0** | $(6.5 \pm 6.6) \times 10^{5}$ | $(2.5 \pm 0.8) \times 10^{-4}$ | $(5.6 \pm 2.4) \times 10^{-10}$ | 1.11 | 2.8 ± 1.4 |
| **ori1** | $(1.4 \pm 2.1) \times 10^{6}$ | $(1.2 \pm 0.3) \times 10^{-4}$ | $(3.6 \pm 2.8) \times 10^{-10}$ | 0.779 | 28.9 ± 0.7 |
| **ori2** | $(1.1 \pm 0.8) \times 10^{5}$ | $(4.6 \pm 1.3) \times 10^{-5}$ | $(3.4 \pm 1.6) \times 10^{-10}$ | 2.83 | 35.6 ± 19.8 |
| **ori3** | $(8.7 \pm 5.9) \times 10^{5}$ | $(2.5 \pm 0.1) \times 10^{-4}$ | $(4.1 \pm 0.1) \times 10^{-10}$ | 2.36 | 10.6 ± 3.0 |
| **ori4** | $(1.2 \pm 0.1) \times 10^{5}$ | $(7.6 \pm 1.4) \times 10^{-5}$ | $(6.4 \pm 1.14) \times 10^{-10}$ | 1.37 | 19.3 ± 6.5 |

## Supplementary Figures

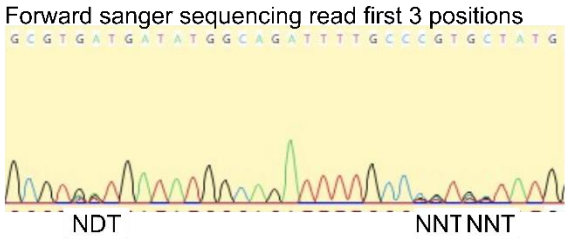



**C**

**Boskar4 saturation library**

| Position (in 7NY0) | Codons (reverse complement) | Coded amino acids |
|---|---|---|
| L38 | NDT (AHN) | F, L, I, V, Y, H, N, D, C, R, S, G |
| T45 | NNT (ANN) | F, L, I, V, S, P, T, A, Y, H, N, D, C, R, G |
| T46 | NNT (ANN) | F, L, I, V, S, P, T, A, Y, H, N, D, C, R, G |
| S61 | NDT | F, L, I, V, Y, H, N, D, C, R, S, G |
| L64 | NDT | F, L, I, V, Y, H, N, D, C, R, S, G |
| Q69 | VDG | L, M, V, Q, K, E, R, G |

| | |
|---|---|
| Theoretical diversity | 4.0E+06 |
| Amino-acid diversity | 3.1E+06 |
| Total transformants | 3.5E+07 |

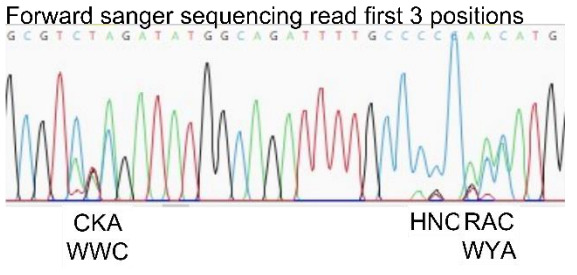



**D**

**Boskar4 library designed with Damietta**

| Position (in 7NY0) | aa lowest energy change | Codons (reverse complement) | Present |
|---|---|---|---|
| L38 | R, F, L, N, I | CKA, WWC (TMG, GWW) | F(*) I(*) L(*) N(*) R(*) Y(0) |
| T45 | Y, F, R, N, H | HNC (GND) | C(0) F(*) H(*) I(0) L(0) N(*) P(0) R(*) S(0) T(*) Y(*) |
| T46 | N, L, D, I, S | RAC, WYA (GTY, TRW) | D(*) I(*) L(*) N(*) S(*) T(*) |
| S61 | F, Y, R, D, N | NDC | C(0) D(*) F(*) G(0) H(0) I(0) L(0) N(*) R(*) S(*) V(0) Y(*) |
| L64 | Y, F, R, H, S | YNC | C(0) F(*) H(*) L(*) P(0) R(*) S(*) Y(*) |
| Q69 | M, K, L, V, I | MWA, RTG | I(*) K(*) L(*) M(*) Q(*) V(*) |

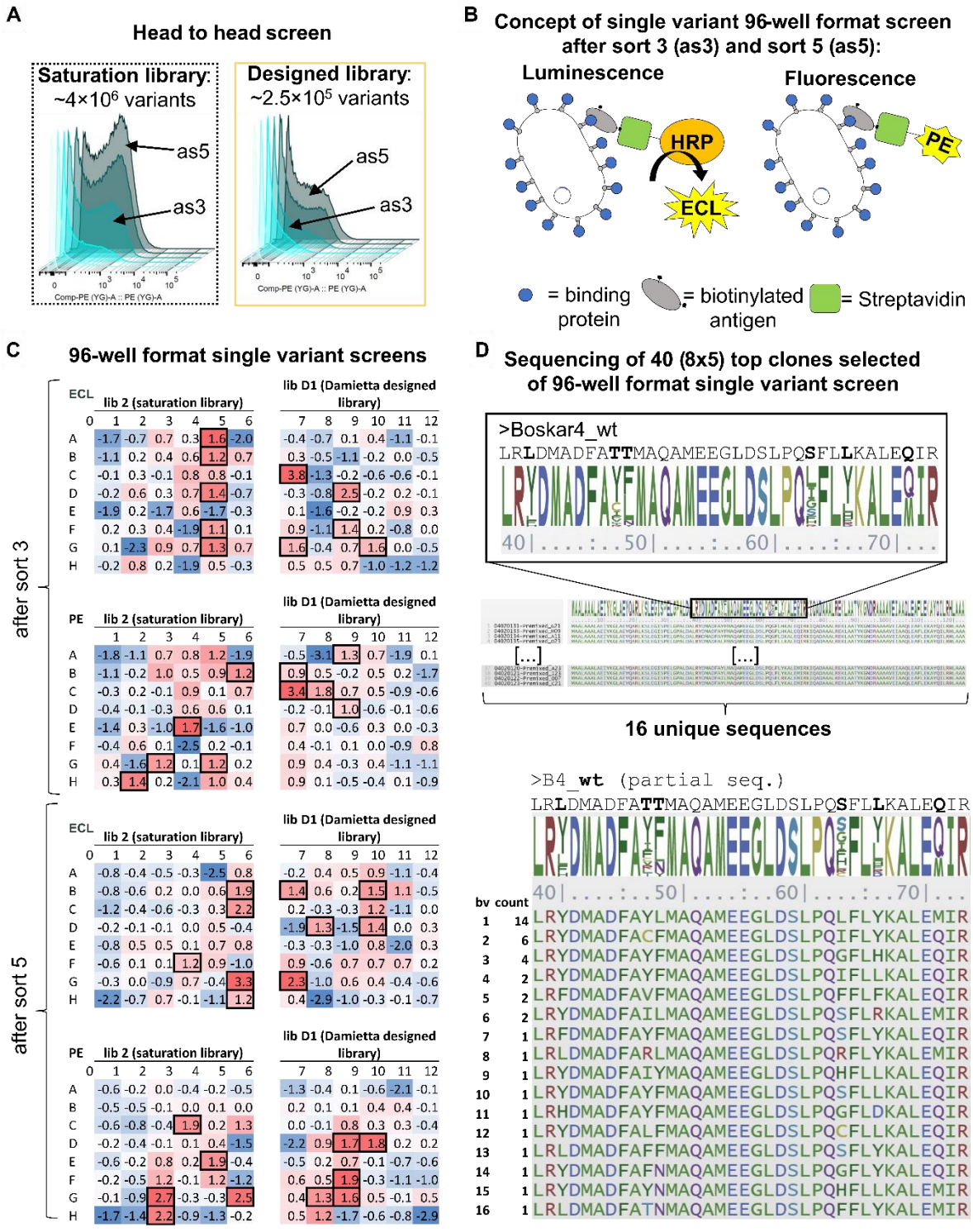
| | |
|---|---|
| Theoretical diversity | 2.5E+05 |
| Amino-acid diversity | 2.3E+05 |
| Total transformants | 2.7E+07 |




(*) = wanted aa (top 5 aa at the given positions or "wt"-aa) (over all 15.63 %)
(0) = unwanted aa (aa added because of used degeneration)

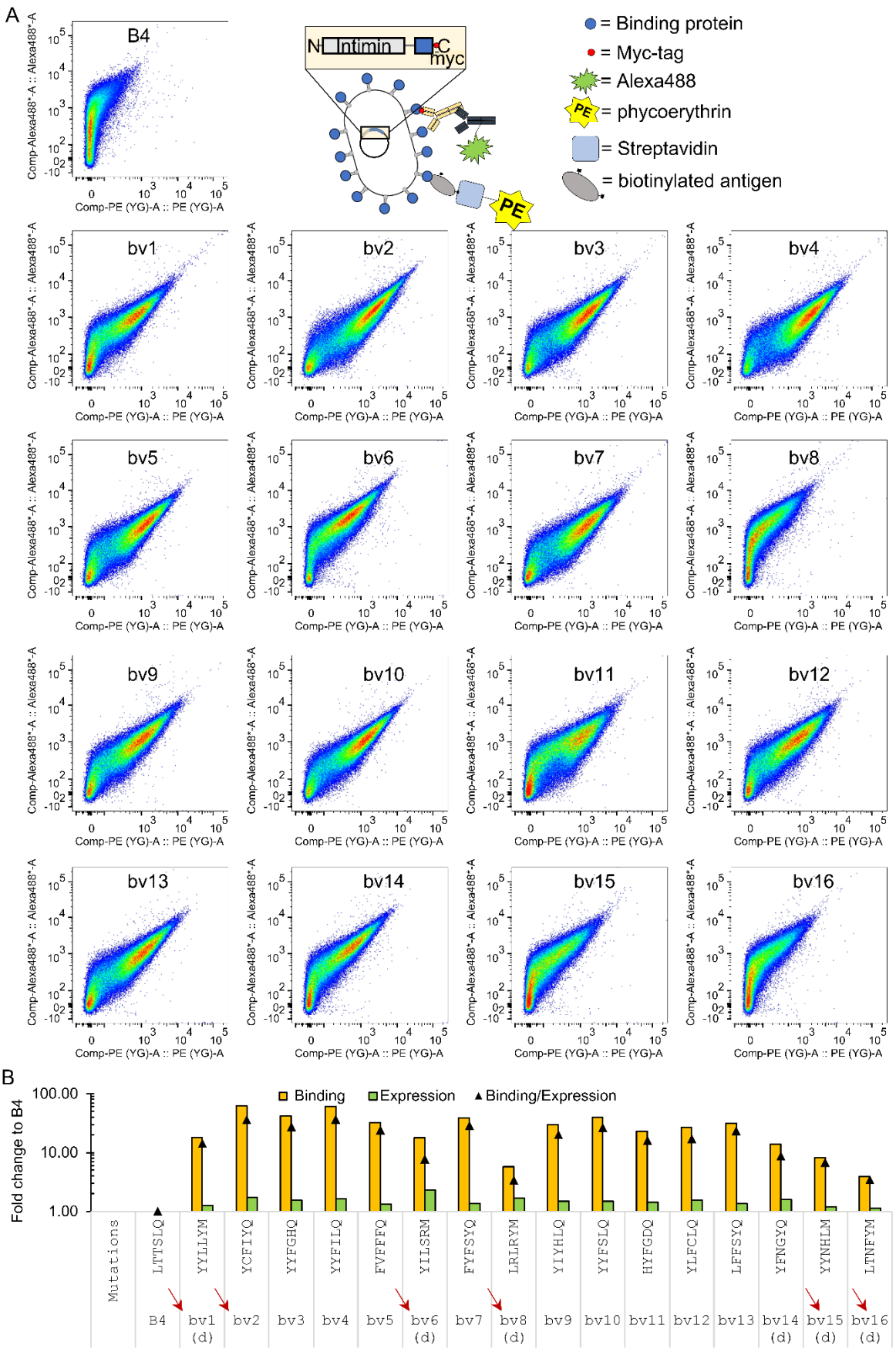
**Supplementary Figure 1. (A)** Side view on Boskar4 (cyan) in a modeled complex with G-CSFR (green) and detailed view on Boskar4 with marked residues based on 7NY0 considered for mutagenesis (yellow) based on the minimal distance to G-CSFR residues. Residues that were thought to be critical

for G-CSF activity (red) regarding to Young *et al. [44]* were excluded from mutagenesis also if they fulfilled the distance criteria. **(B)** Overview of library generation via degenerated codons and blunt end ligation. **(C)** Overview of the saturation library of Boskar4 for affinity maturation. **(D)** Overview of the Damietta *in silico* designed library of Boskar4 which was estimated from a reduced set of the same residues as the saturation library.

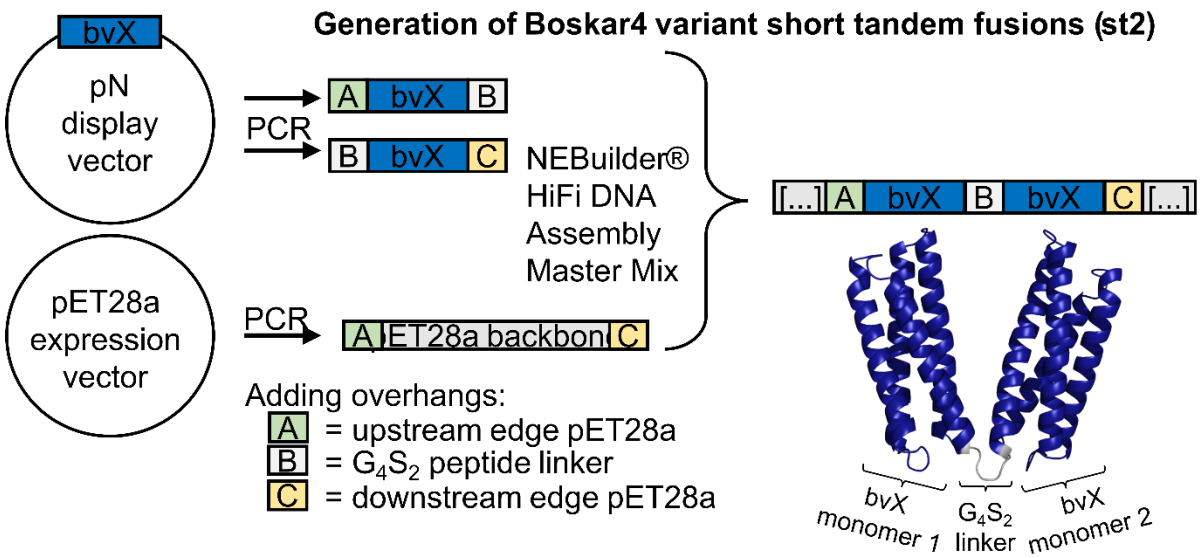



**Supplementary Figure 2. (A)** Fluorescence measured with FACS of bacteria carrying either the saturation library (dotted line) or the Damietta designed library (light orange) over 5 steps of enrichment (light cyan to dark gray) in a head-to-head screening experiment. **(B)** Concept of single variant 96-well format screen after sort 3 (as3) and sort 5 (as5). Two different readout methods were utilized to estimate the relative binding activity of each variant. Luminescence was applied by a horse radish peroxidase (HRP) avidin conjugate with enhanced chemiluminescence substrate (ECL), and fluorescence was measured by streptavidin phycoerythrin (PE) conjugate. **(C)** Results of the 96-well format single variant screens after enrichment of the library with FACS. Each value corresponds to the z-score given by a single variant (single well on the plate) to all other variants from the same condition. In total, 8 such

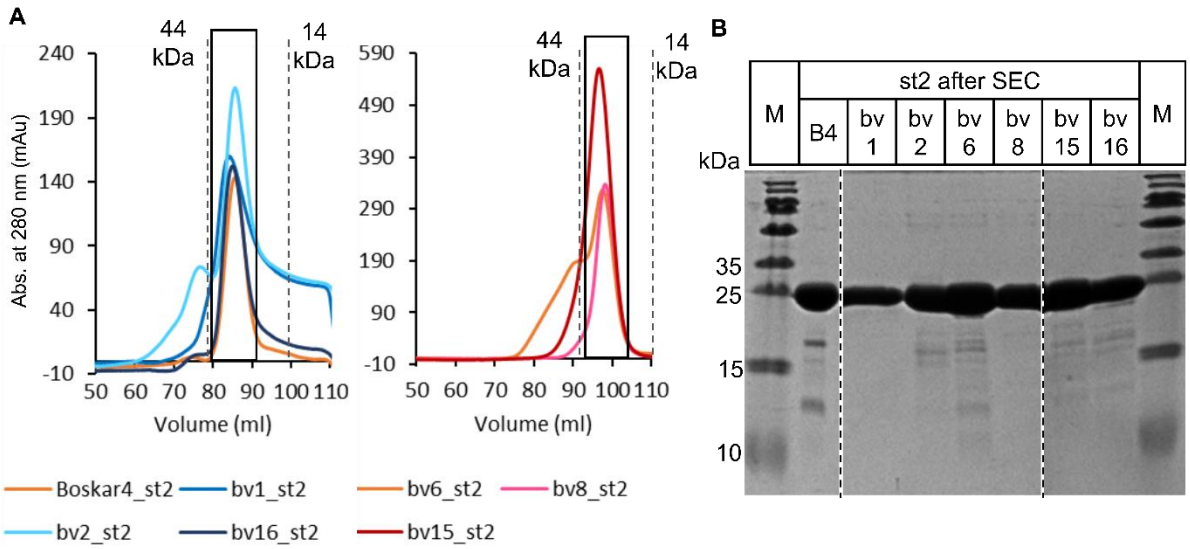
screens were made under different conditions (PE or ECL, after sort 3 or after sort 5, saturation or designed library), and top 5 clones of each condition were picked and sequenced. **(D)** Sixteen unique Boskar4 variants (bv1 to bv16) were identified among the 40 analyzed.


A
B4
N
Intimin
C
myc
= Binding protein
= Myc-tag
= Alexa488
= phycoerythrin
= Streptavidin
= biotinylated antigen
PE
bv1
bv2
bv3
bv4
bv5
bv6
bv7
bv8
bv9
bv10
bv11
bv12
bv13
bv14
bv15
bv16
Comp-Alexa488*-A :: Alexa488*-A
Comp-PE (YG)-A :: PE (YG)-A
B
Fold change to B4
100.00
10.00
1.00
Binding
Expression
Binding/Expression
Mutations
LTTSLQ
YYLLYM
YCFIYQ
YYFGHQ
YYFILQ
FVFFFQ
YILSRM
FYFSYQ
LRLRYM
YIYHLQ
YYFSLQ
HYFGDQ
YLFCLQ
LFFSYQ
YFNGYQ
YYNHLM
LTNFYM
B4
bv1 (d)
bv2
bv3
bv4
bv5
bv6 (d)
bv7
bv8 (d)
bv9
bv10
bv11
bv12
bv13
bv14 (d)
bv15 (d)
bv16 (d)

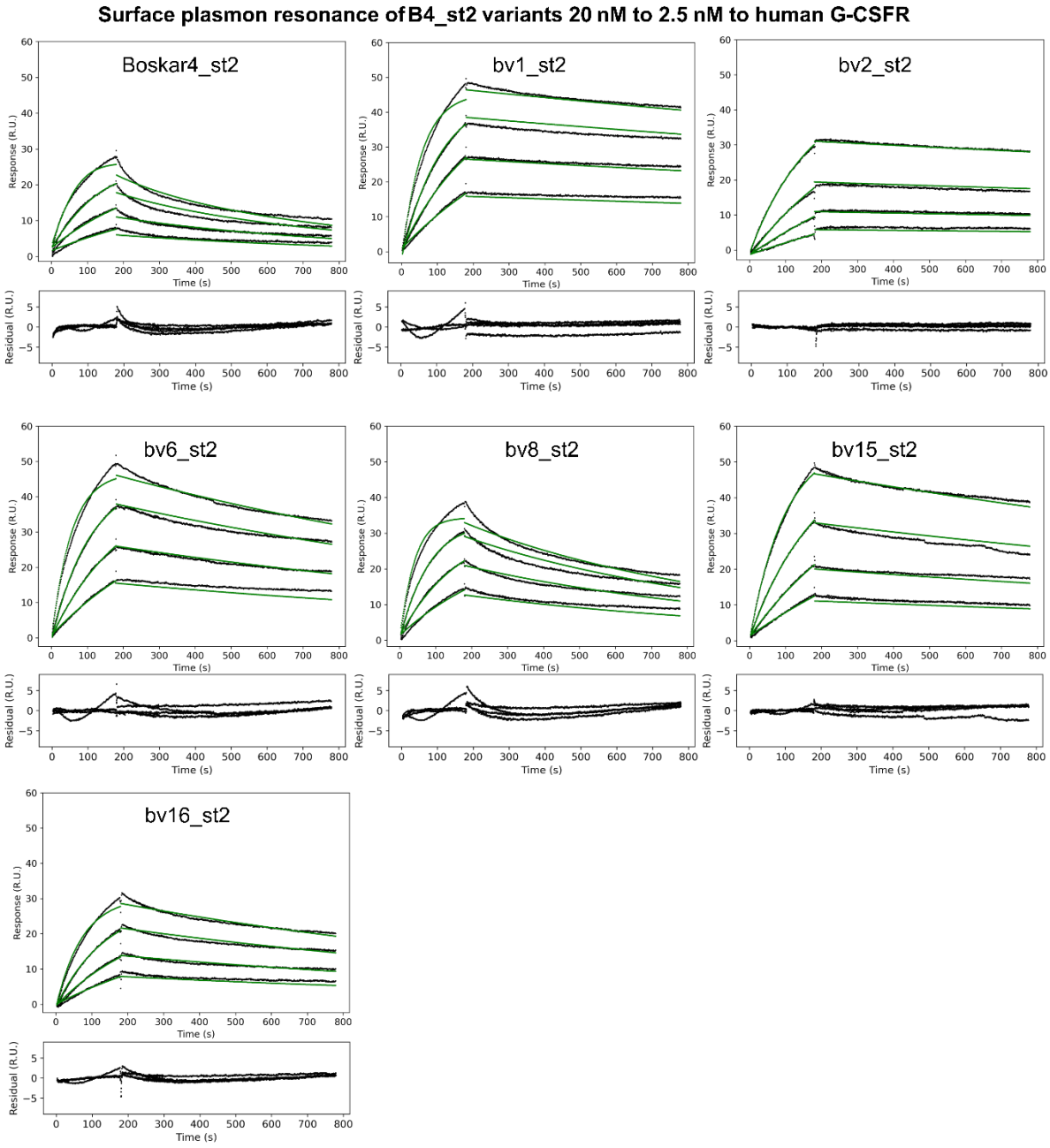
**Supplementary Figure 3. (A)** The 16 new affinity-enhanced Boskar4 variants (bv1 to bv16) and Boskar4 (B4) were analyzed on the surface of *E. coli* by FACS for their binding activity to BioGCSFR (10 nM, PE channel) and the expression level of the corresponding variant (anti-myc labeling, Alexa488 channel). **(B)** Based on the measurements in (A) the binding to G-CSFR (orange bars), expression of the binder (green bars), and expression-normalized binding to G-CSFR (black triangles) of the affinity-enhanced Boskar4 variants normalized to Boskar4 was analyzed. Sequences selected for further characterization are marked with dark red arrows. The small "d" in parentheses (d) indicates that the variant was selected from the designed library. Additionally, the corresponding mutation set of each enhanced variant is indicated in comparison to Boskar4.



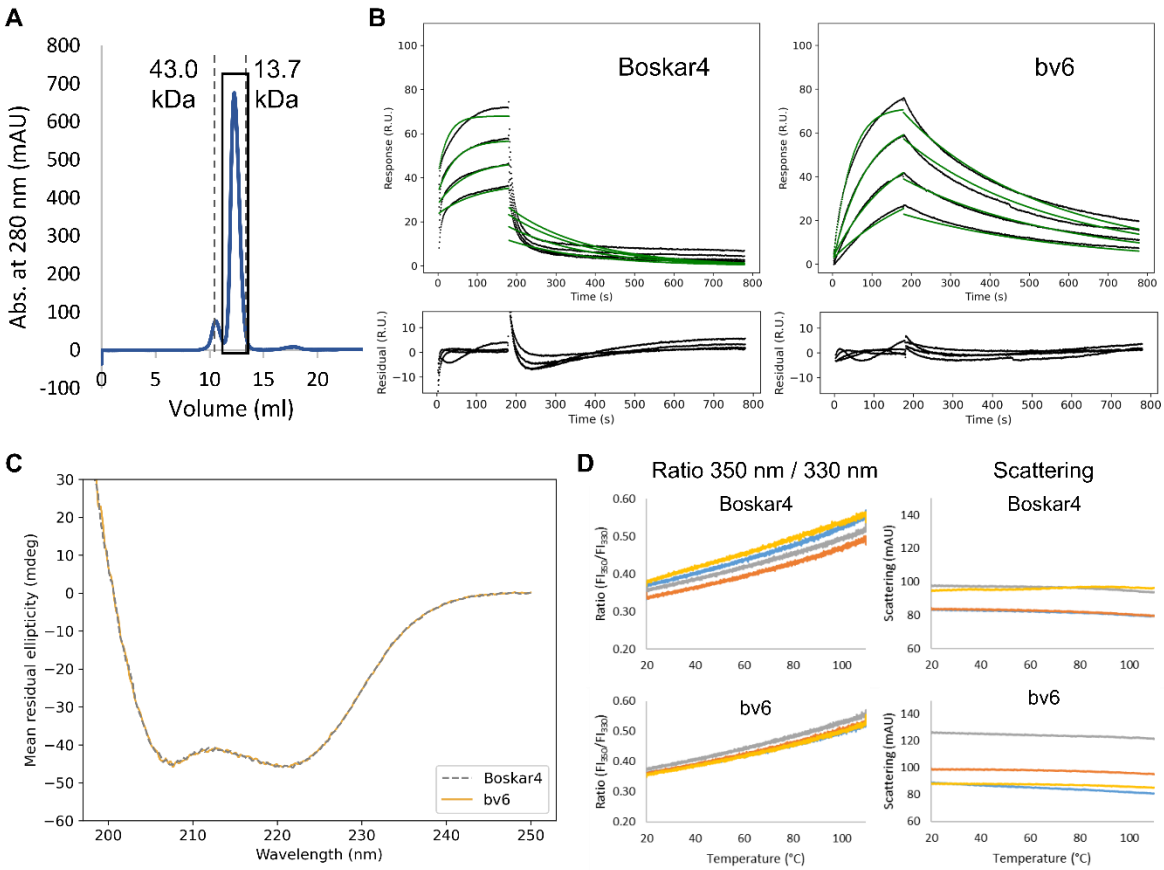

**Supplementary Figure 4.** The Boskar4 variant short tandem fusions (st2) were generated by the assembly of three fragments consisting of the expression vector, two copies of a Boskar4 variant (bvX) and an overlap including the $G_4S_2$ peptide linker. The same approach was also used to generate the rigid tandem fusions (ori0 to ori4) by encoding a rigid helix linker instead of a flexible GS-peptide-linker.



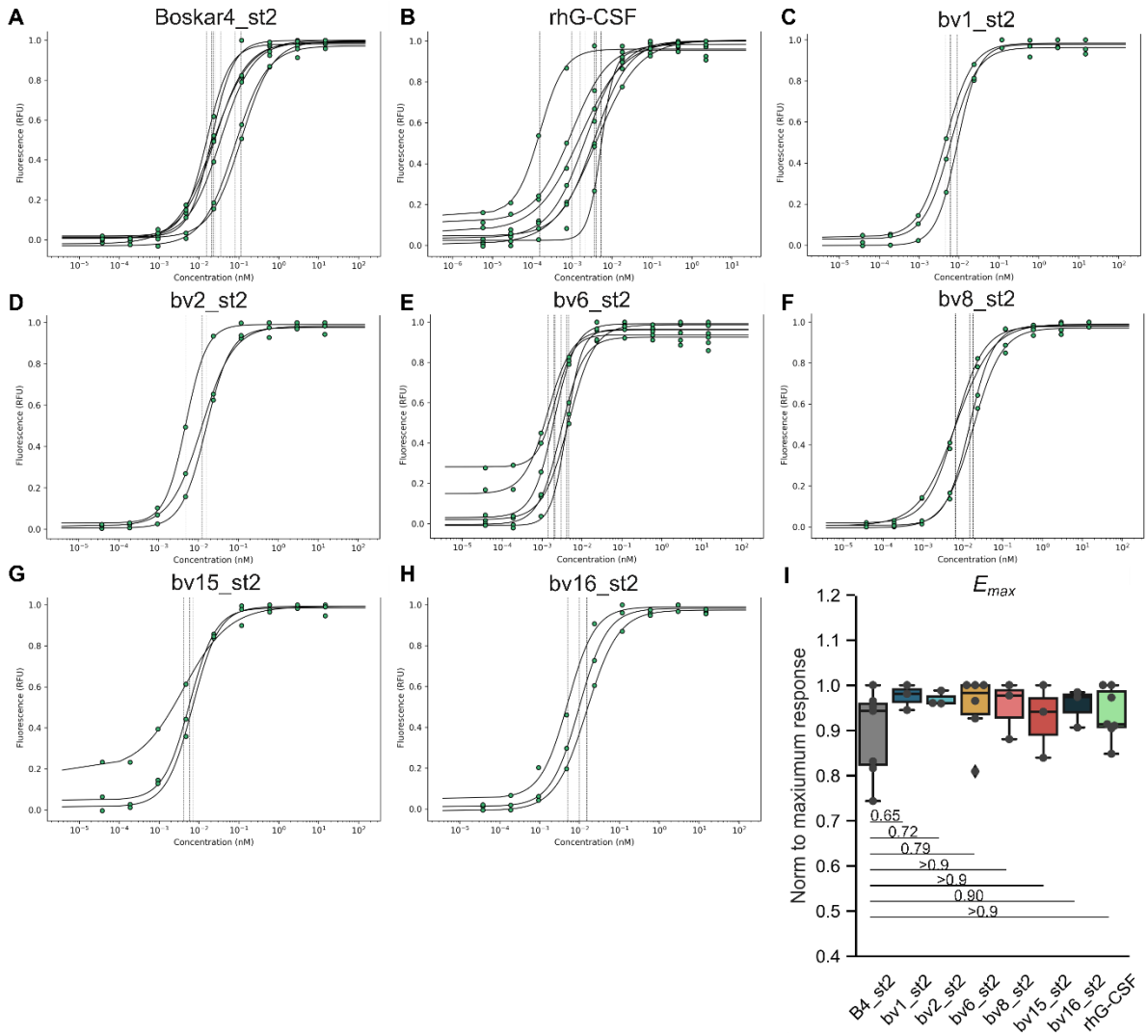

**Supplementary Figure 5. (A)** Preparative size exclusion chromatography of Boskar4 short tandem fusion (st2) variants after Nickel IMAC on two separate SEC-columns of the same type (HiLoad 16/600, Superdex 200 pg). Dotted lines mark the molecular weight standard for the corresponding column. All fractions of the main peak (black box) were collected and concentrated by ultrafiltration (10 kDa cut-off). **(B)** A representative sample of each protein and a protein molecular weight marker (M) (11852124, Thermo Fisher Scientific) was loaded on a SDS-PAGE (18% PA, 150 V, 1 h) which was subsequently stained with Coomassie. The expected molecular weight of Boskar4_st2 and variants (with not cleaved TEV-cleavage site and His-tag) is ~29 kDa.



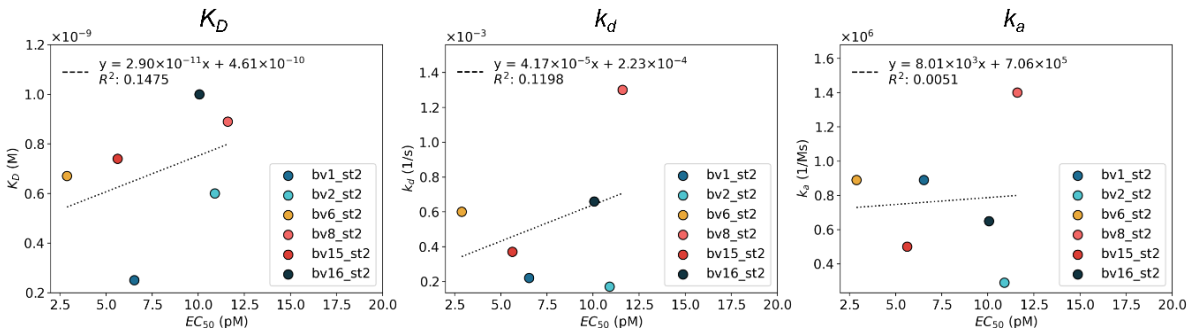

**Supplementary Figure 6.** Six of the enhanced variants and Boskar4 itself (B4) were chosen to generate short tandem fusions (st2) and measured with surface plasmon resonance (SPR) against immobilized hG-CSFR. The association rate constant ($k_a$), the dissociation rate constant ($k_d$), and the apparent equilibrium dissociation constant ($K_D$) were estimated from the corresponding fits (green lines) of a two-fold titration series from 20 nM to 2.5 nM (compare Table 1).



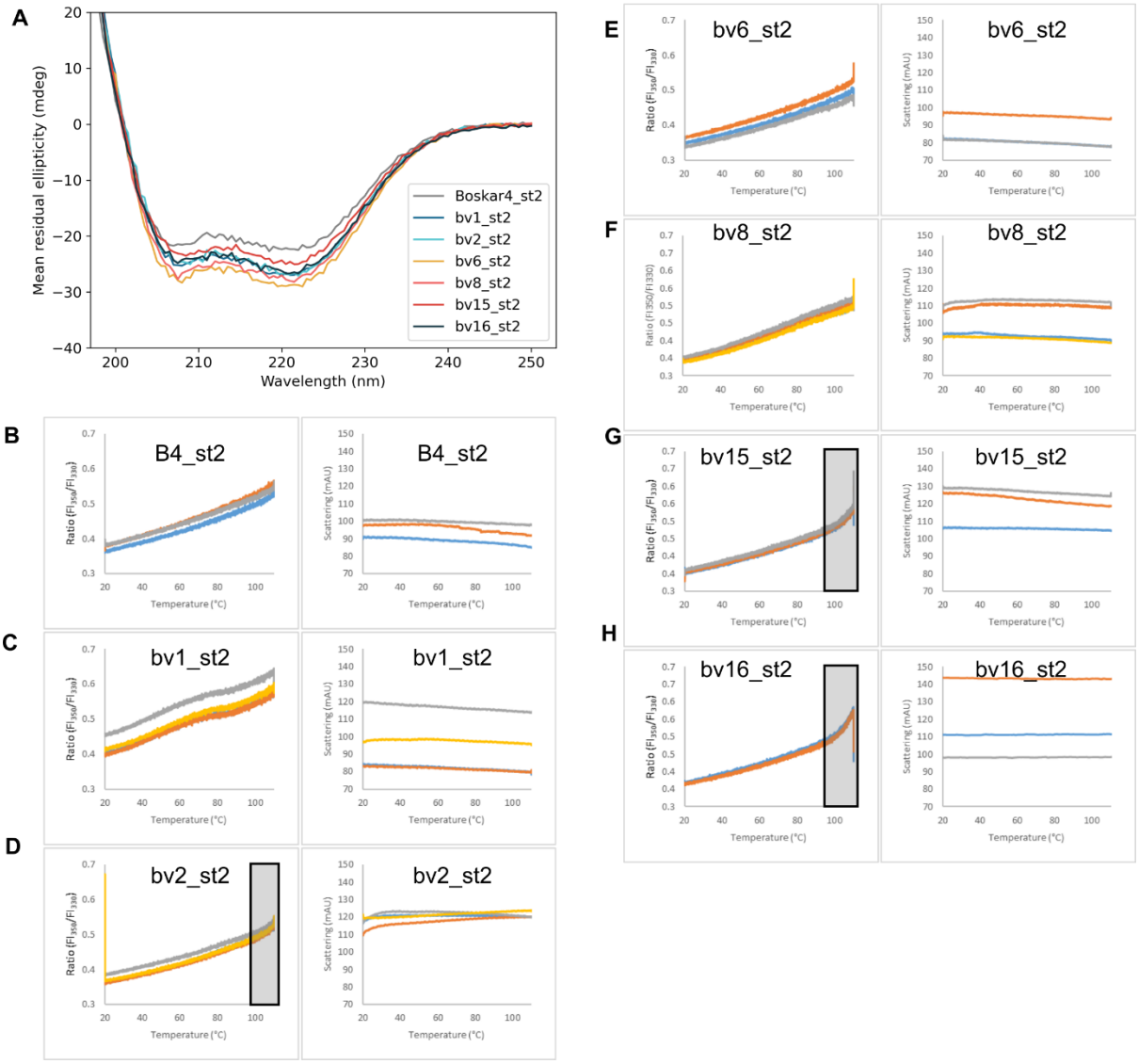

**Supplementary Figure 7.** **(A)** Preparative size exclusion chromatography of monomeric bv6 on a Superdex® 75 10/300 GL with an expected size of ~15 kDa. Dotted lines mark the molecular weight standard for the corresponding column. All fractions of the main peak (black box) were collected and concentrated by ultrafiltration (10 kDa cut-off). **(B)** Surface plasmon resonance measurement (black line) and the corresponding fit (green lines) of a two-fold dilution series of monomeric Boskar4 and the variant bv6. The highest concentration used for Boskar4 was 1250 nM and for bv6 125 nM. The obtained binding parameters are listed in Table 1. **(C)** The single-domain Boskar4 (B4; gray) and an example variant (bv6; orange) were analyzed with circular dichroism and show strong helical signal. **(D)** Thermostability measurement with nanoDSF of 1 mg/mL of monomeric Boskar4 and bv6 with 4 technical replicates, each indicating the proteins to be stable to at least 110 °C.



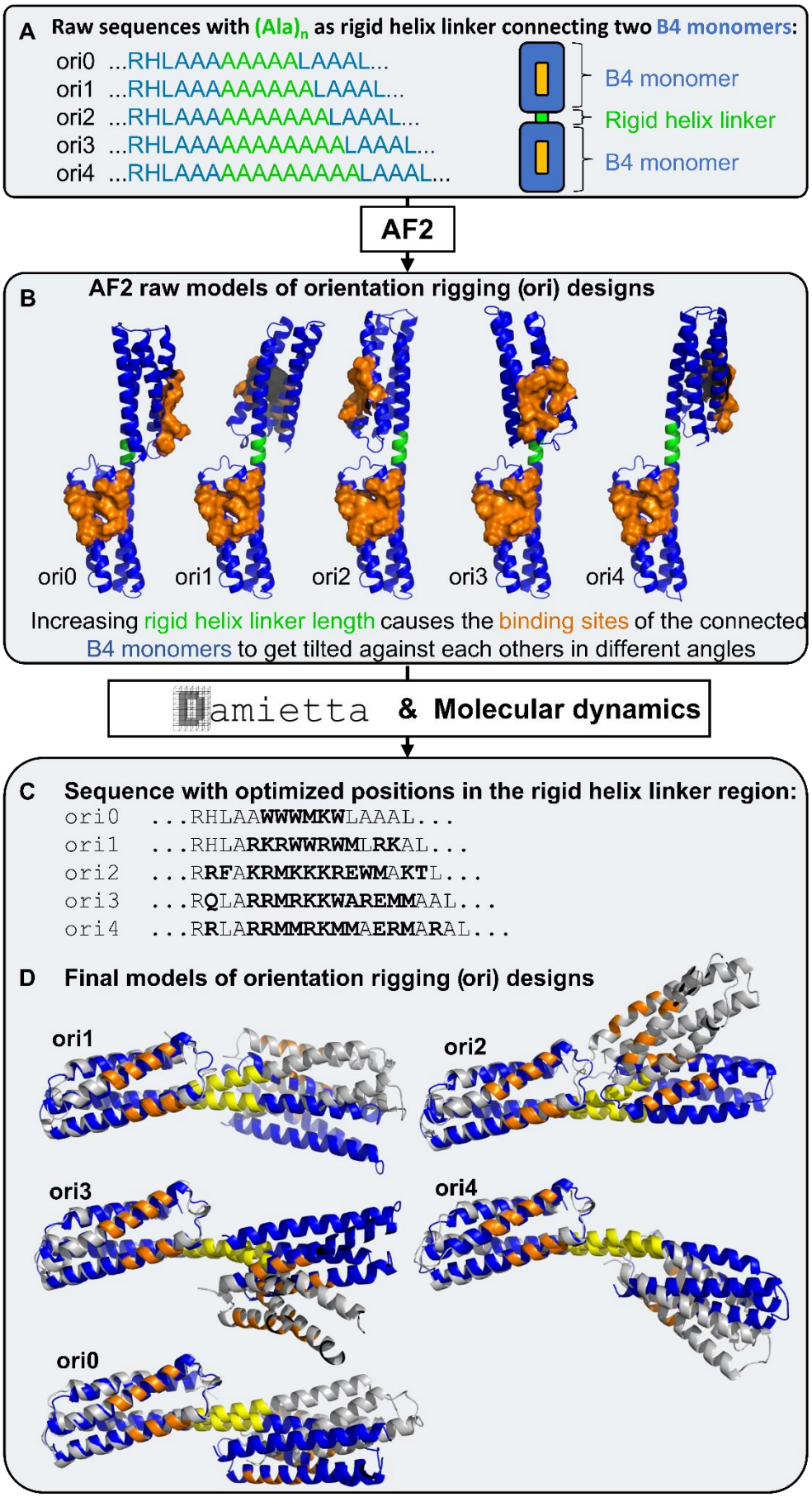

**Supplementary Figure 8. (A-H)** Dose-response NFS-60 cell proliferation assays to determine the half-maximal effective concentration ($EC_{50}$) of Boskar4_st2 variants and human recombinant G-CSF (hrG-CSF). The fits to determine the $EC_{50}$ (pM) of at least three independent experiments per G-CSFR agonist are shown (compare also Table 1 and Fig. 2C). **(I)** Maximum proliferative activity ($E_{max}$) normalized to the maximum response within each independent experiment represented in (A-H). For statistical analysis, an ordinary one-way ANOVA was performed, followed by a Tukey HSD test.



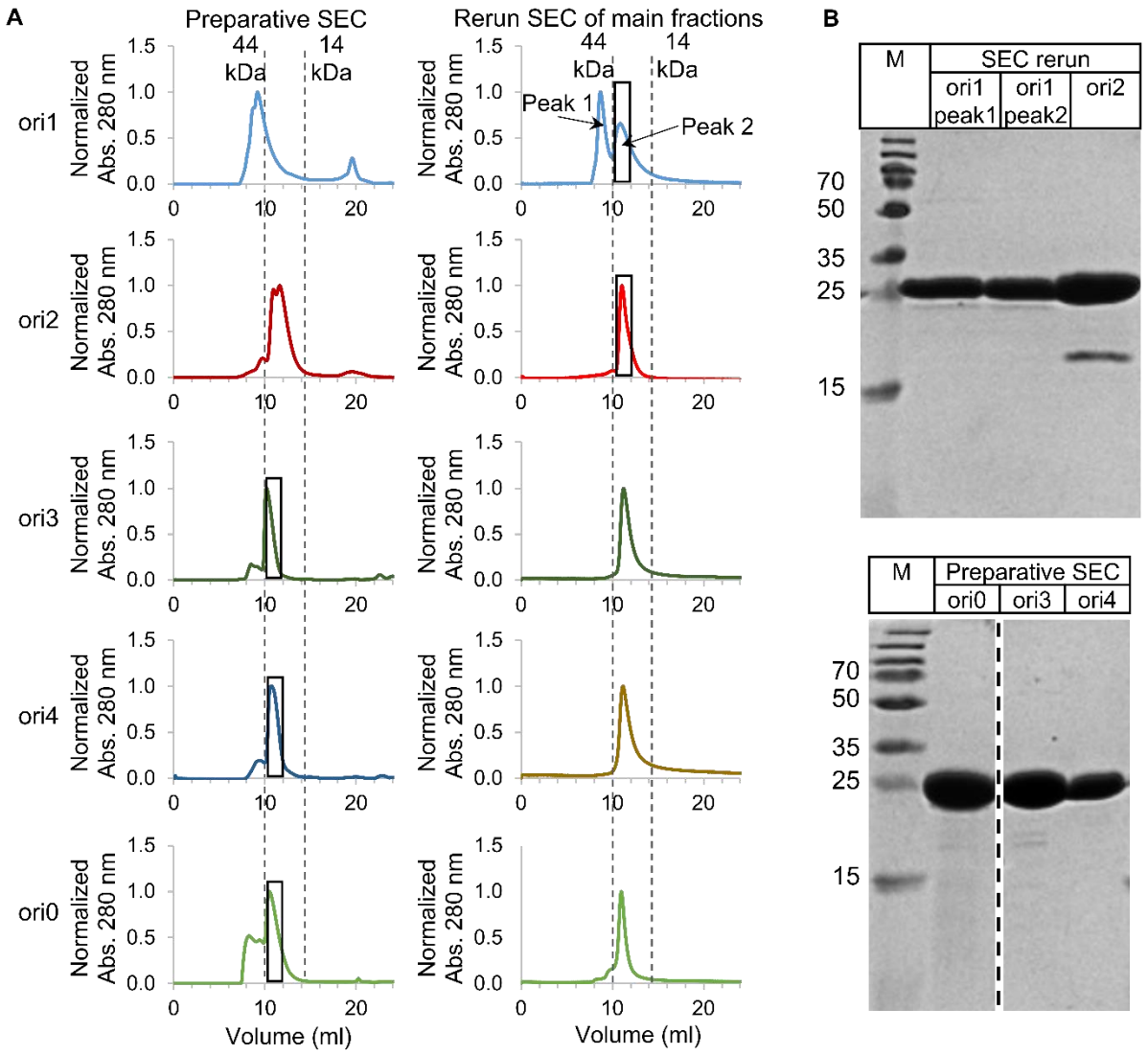

**Supplementary Figure 9.** Shown is the correlation of a linear fit between the apparent affinity ($K_D$), association rate constant ($k_a$) or dissociation rate constant ($k_d$) and the obtained NFS-60 activity ($EC_{50}$) for short tandem fusions of Boskar4 (B4) or the affinity-enhanced variants (bv1, bv2, bv6, bv8, bv15, bv6). The coefficient of determination ($R^2$) is shown in the left upper corner.



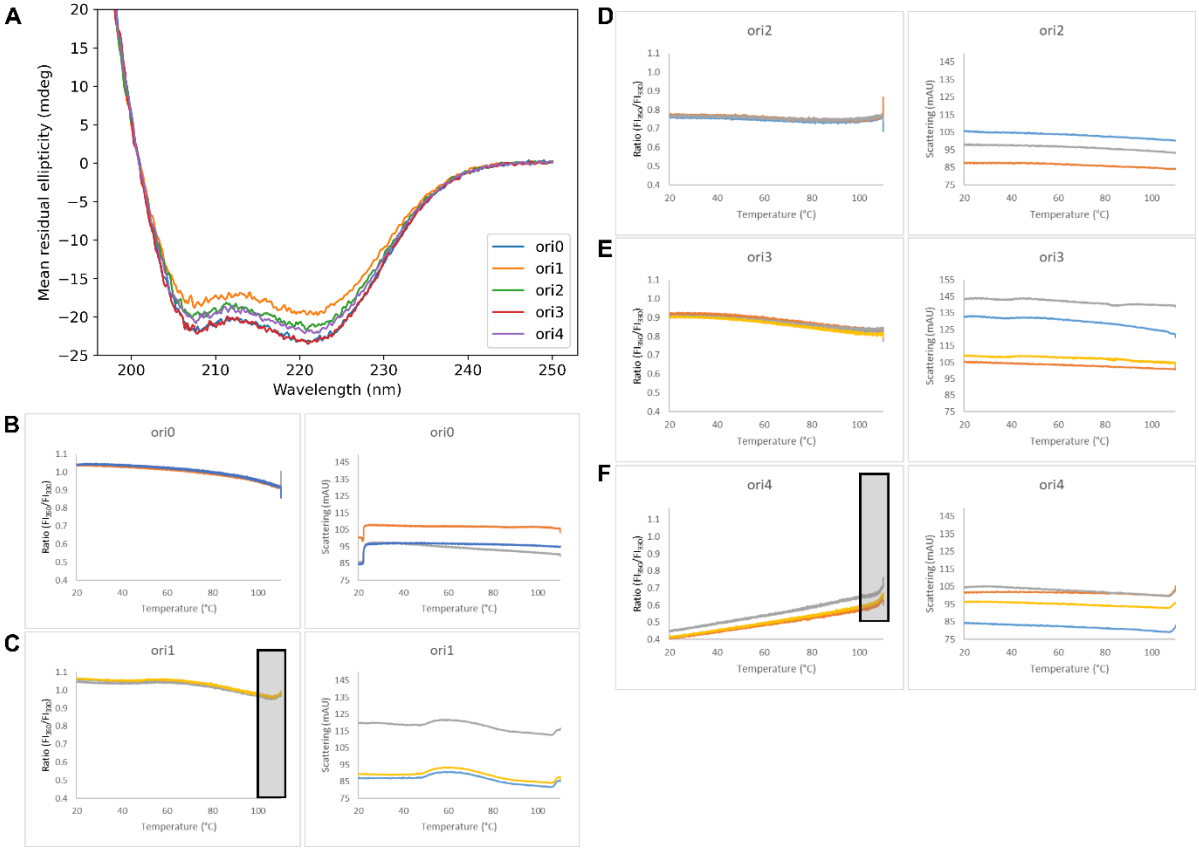

**Supplementary Figure 10. (A)** Circular dichroism (CD) measurement of Boskar4_st2 and variants (bv1 to bv16) in PBS at 0.1 mg/mL shows expected alpha-helical signal for all samples. **(B-H)** Thermostability measurement with nanoDSF of Boskar4_st2 (B4_st2) and variants (bv1_st2 to bv16_st2) with at least 3 technical replicates. All samples were diluted to 1 mg/mL in PBS, except for bv1_st2, which was used at a concentration 0.5 mg/ml. The gray boxes indicate areas in which a melting onset was observed.


A Raw sequences with (Ala)n as rigid helix linker connecting two B4 monomers:
ori0 ...RHLAAAAAAAALAAAL...
ori1 ...RHLAAAAAAAAALAAAL...
ori2 ...RHLAAAAAAAAAALAAAL...
ori3 ...RHLAAAAAAAAAAALAAAL...
ori4 ...RHLAAAAAAAAAAAALAAAL...
B4 monomer
Rigid helix linker
B4 monomer
AF2
B AF2 raw models of orientation rigging (ori) designs
ori0
ori1
ori2
ori3
ori4
Increasing rigid helix linker length causes the binding sites of the connected B4 monomers to get tilted against each others in different angles
amietta & Molecular dynamics
C Sequence with optimized positions in the rigid helix linker region:
ori0 ...RHLAAWWWMKWLAAAL...
ori1 ...RHLARKRWWRWMLRKAL...
ori2 ...RRFAKRMKKKREWMAKTL...
ori3 ...RQLARRMRKKWAREMMAAL...
ori4 ...RRLARRMMRKMMAERMARAL...
D Final models of orientation rigging (ori) designs
ori1
ori2
ori3
ori4
ori0

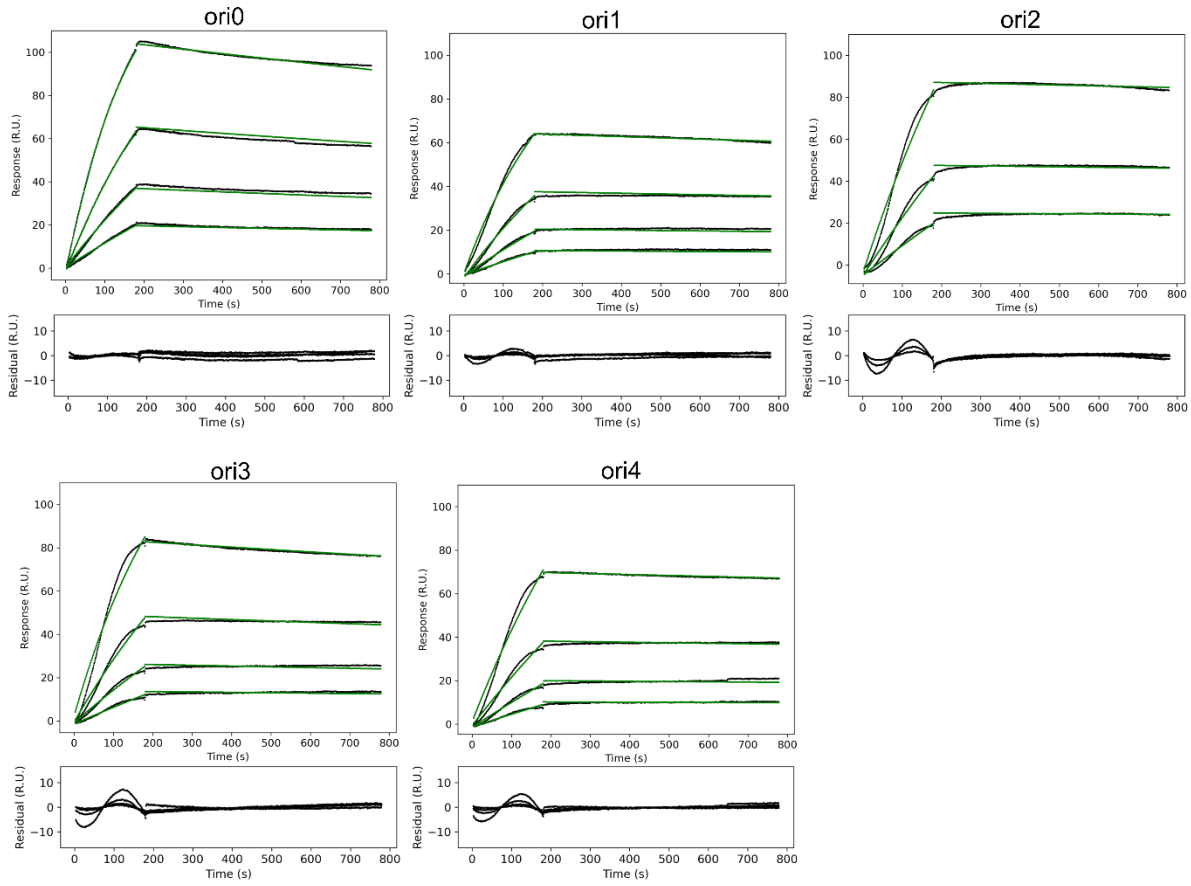
**Supplementary Figure 11.** Generation of the orientation-rigging designs (oris) tuning G-CSFR activity. **(A)** Raw sequence design. **(B)** AlphaFold2 raw models. **(C)** Selected residues (bold) in the rigid helix linker region were optimized with Damietta, and the most stable designs were identified with molecular dynamics (MD). **(D)** The MD input structure of each top hit is displayed in blue, and the last frame of the MD-simulation is indicated in gray. The linker helix area is depicted in yellow, and the G-CSFR binding sites are highlighted in orange.



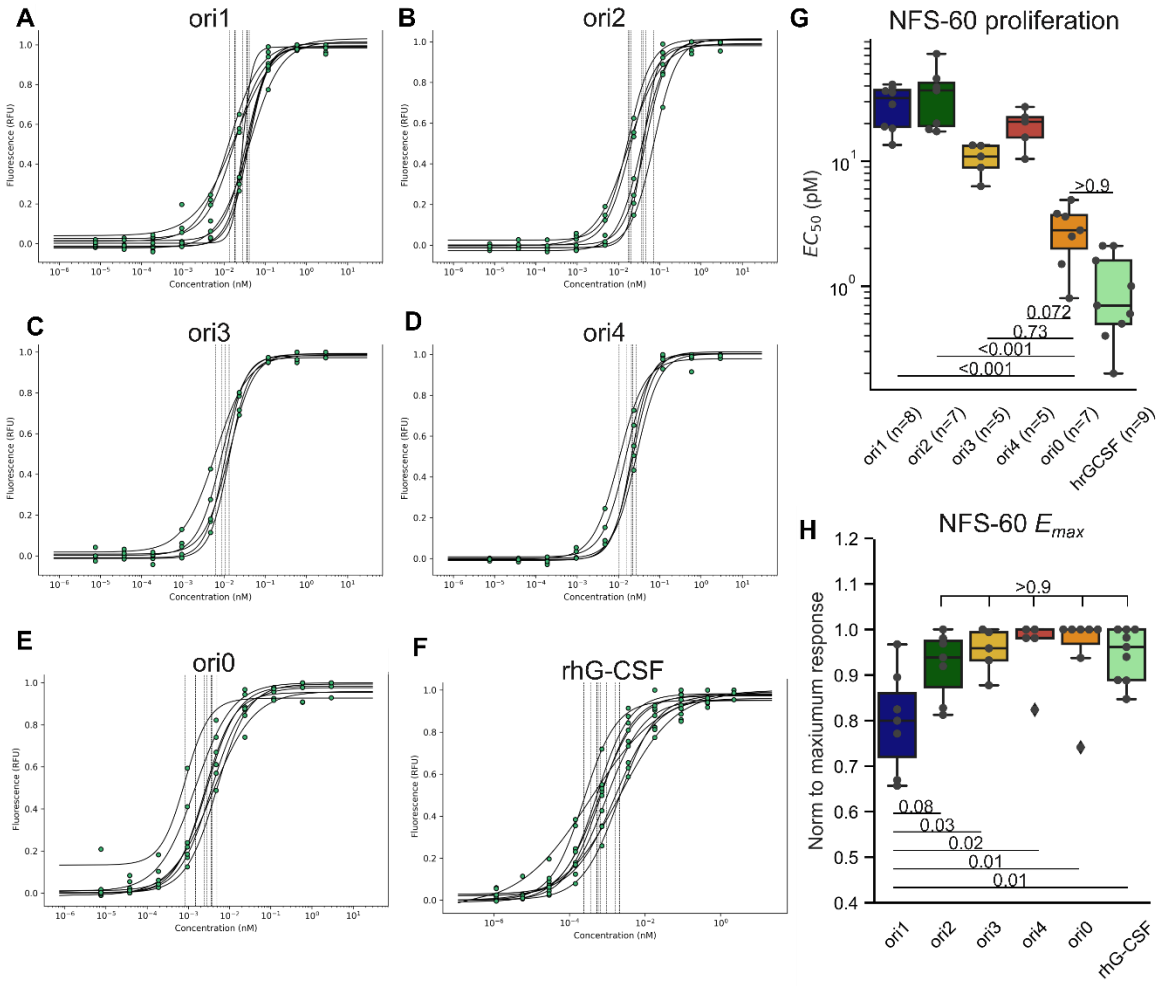

**Supplementary Figure 12. (A)** Size exclusion chromatography of ori0, ori1, ori2, ori3 and ori4 after Nickel IMAC was performed on a Superdex Increase 75 10/300 GL (Cytiva). The dotted lines mark the molecular weight standard for the corresponding column. All fractions of the main peak were collected and concentrated by ultrafiltration (10 kDa cut-off). Part of the collected fractions were rerun on the same column to validate the purity and correct molecular mass. **(B)** A representative sample of each fraction containing the protein as used for the subsequent experiments (black boxes in (A)) was loaded on a SDS-PAGE. The expected molecular weight of the oris (with uncleaved TEV-cleavage site and His-tag) is ~29 kDa. For ori2, the main fraction was collected after the SEC rerun and for ori1, peak 1 and peak 2 were collected separately. The main fractions after the preparative size exclusion of ori0, ori3 and ori4 were directly used for subsequent experiments.



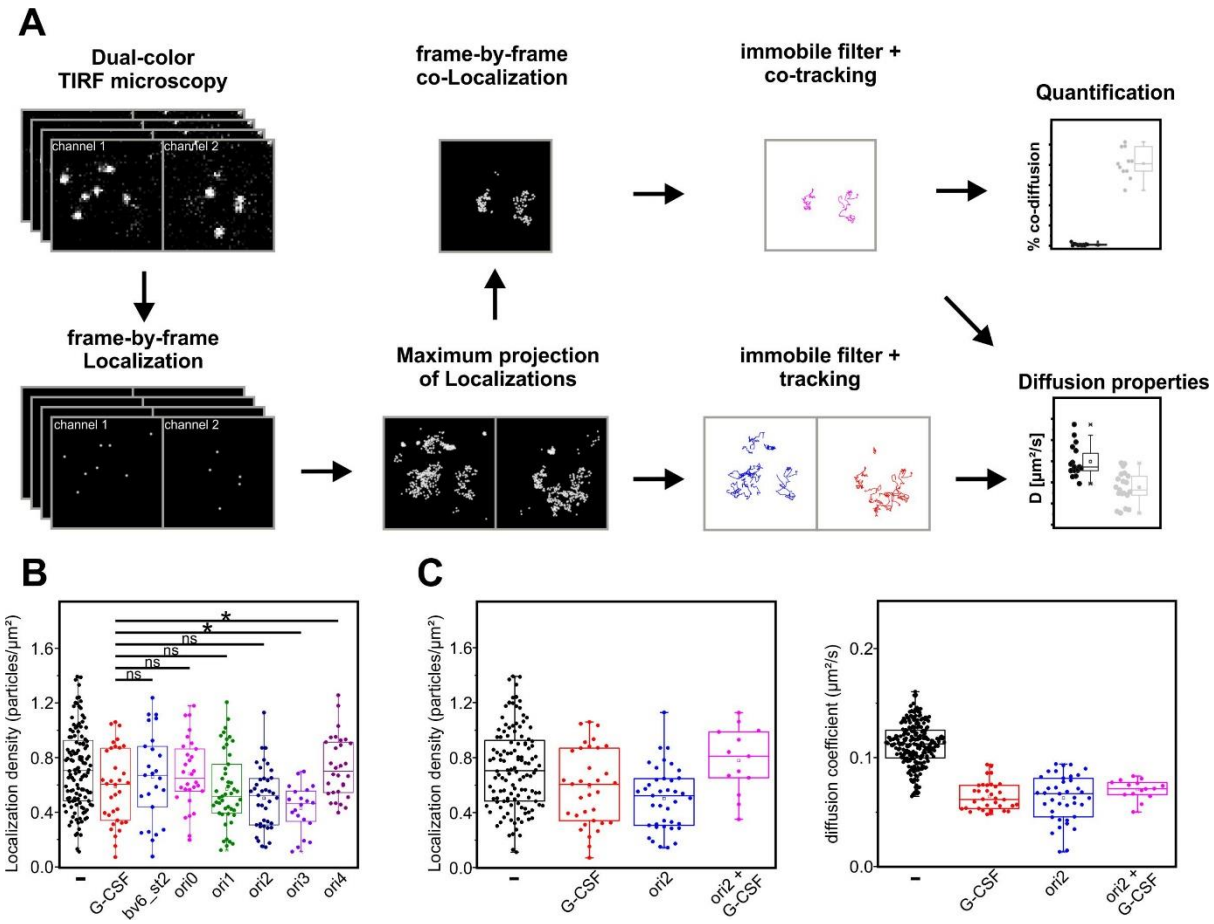

**Supplementary Figure 13: (A)** Circular dichroism (CD) measurement of ori0, ori1, ori2, ori3 and ori4 in PBS at 0.1 mg/mL showed the expected alpha-helical signal for all samples. **(B-F)** Thermostability measurement with nanoDSF of the rigid tandem fusions (ori0 to ori4) with at least 3 technical replicates. All samples were diluted to 1 mg/mL in PBS. The gray boxes indicate areas in which a melting onset was observed.



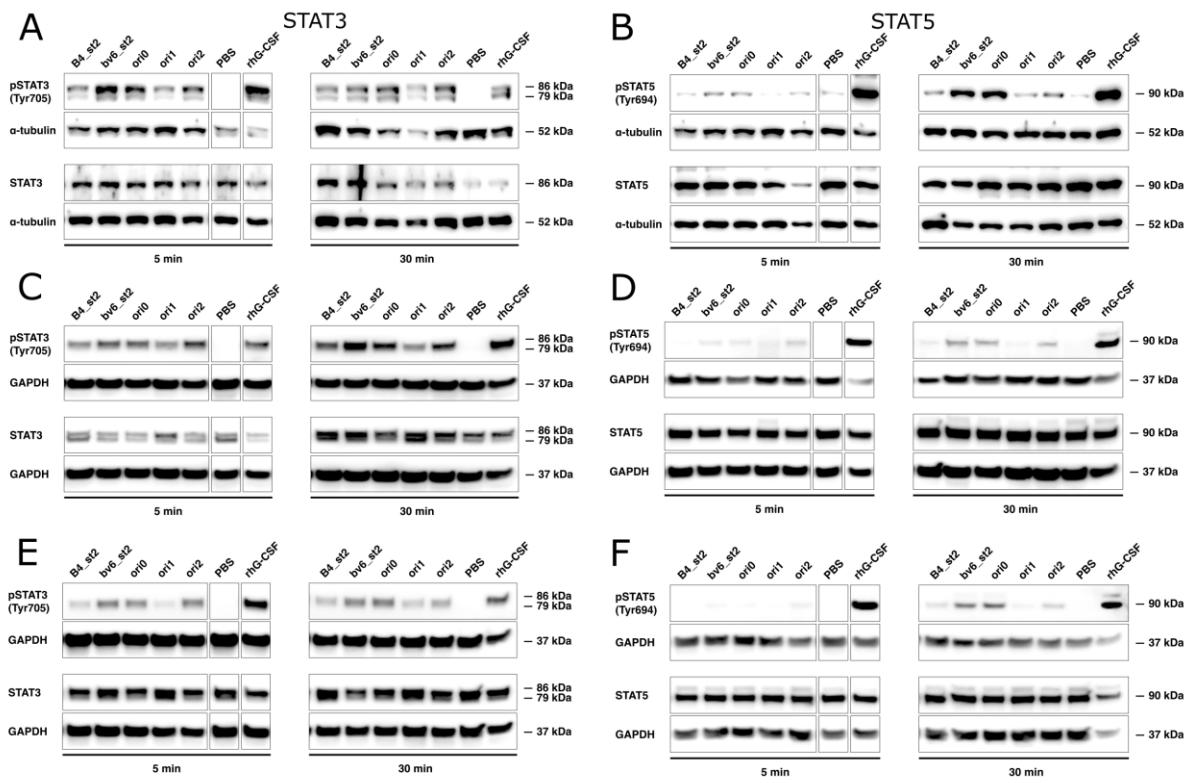

**Supplementary Figure 14.** Surface plasmon resonance (SPR) measurement of oris against immobilized hG-CSFR. The association rate constant ($k_a$), the dissociation rate constant ($k_d$), and the apparent equilibrium dissociation constant ($K_D$) were estimated from the corresponding fits (green lines) of a titration series ranging from 20 nM to 2.5 nM (compare Table 1).



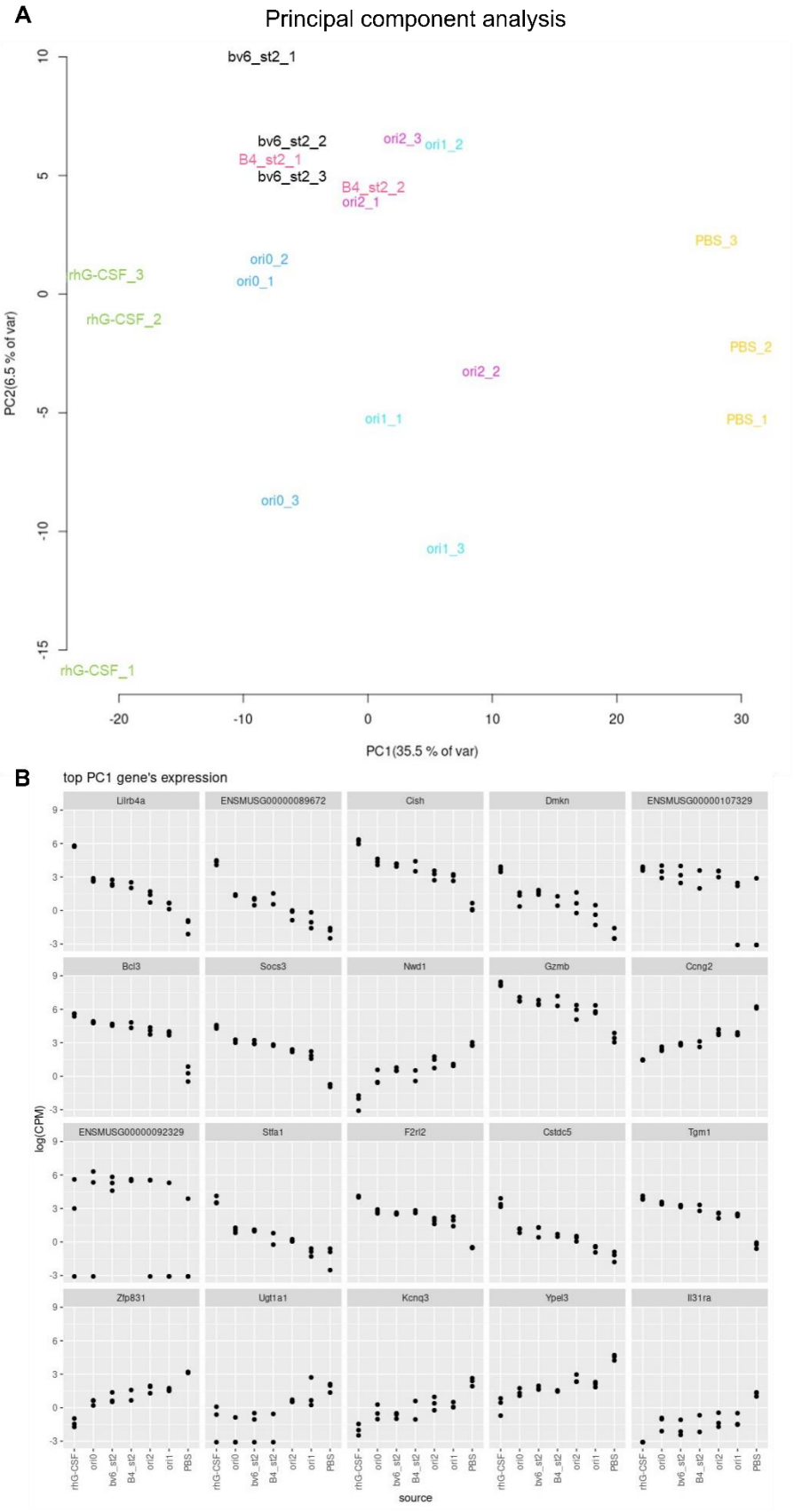

**Supplementary Figure 15. (A-F)** Dose-response NFS-60 cell proliferation assays to determine the half-maximal effective concentration ($EC_{50}$) of the oris and G-CSF (rhG-CSF, Lenograstim). Given are the fits to determine the $EC_{50}$ of at least five independent experiments per G-CSFR agonist (compare also Table 1 and Fig. 3C). **(G, H)** Box plot of the $EC_{50}$ (pM) values (G) and the maximum proliferative activity ($E_{max}$) normalized to the maximum response (H) within each independent experiment represented in (A-F). For statistical analysis in (G) and (H), ordinary one-way ANOVA, followed by a Tukey HSD test was performed.



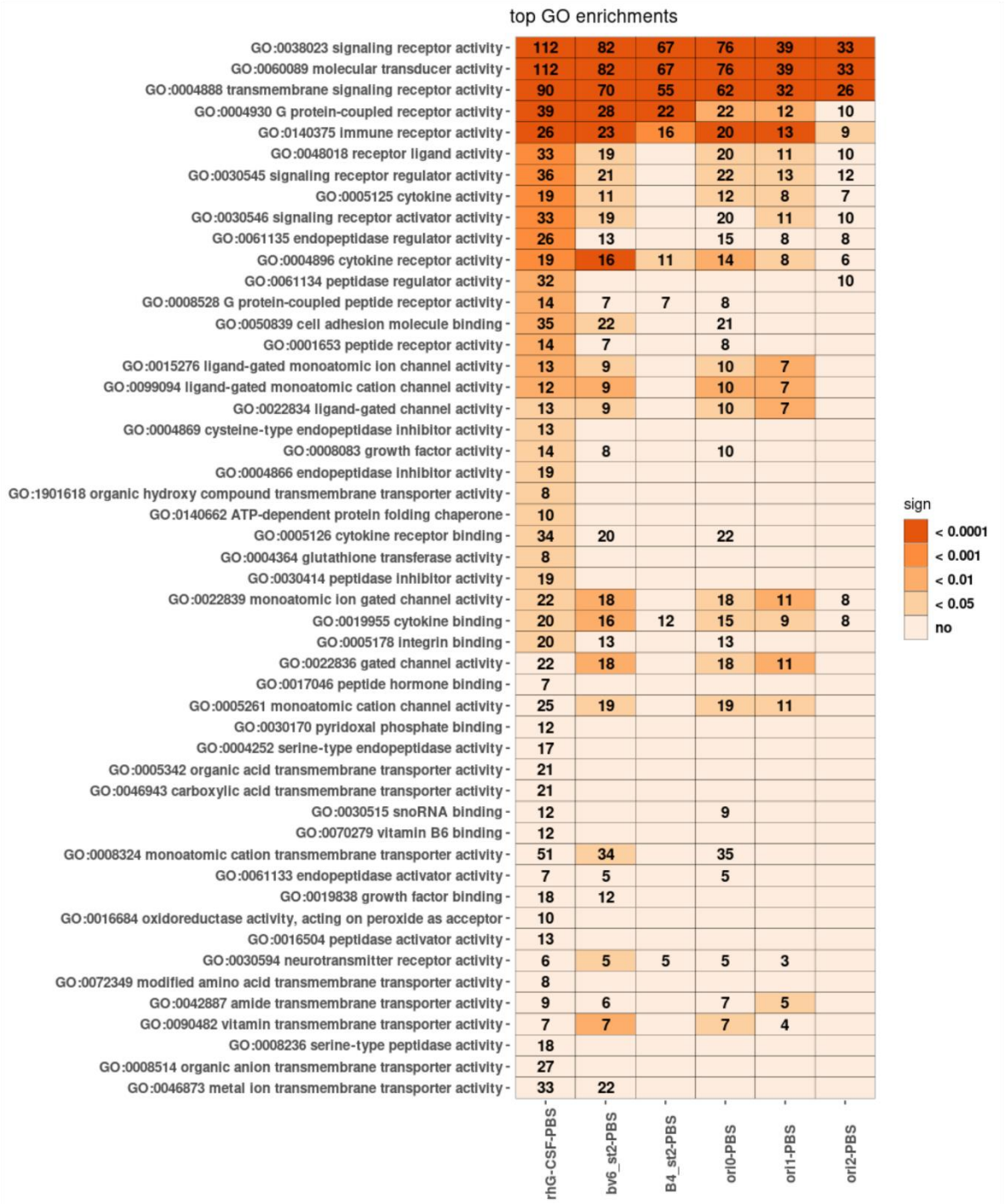

**Supplementary Figure 16.** The dimerization of G-CSFR was analyzed with single molecule imaging for G-CSF and the design agonists. **(A)** Workflow used for single molecule tracking and co-tracking analyses. **(B, C)** Receptor density in co-tracking experiments comparing different design agonists **(B)** and the competition between G-CSF and ori2 **(C)**. **(D)** Co-diffusion constants observed in G-CSF/ori2 competition co-tracking experiments.



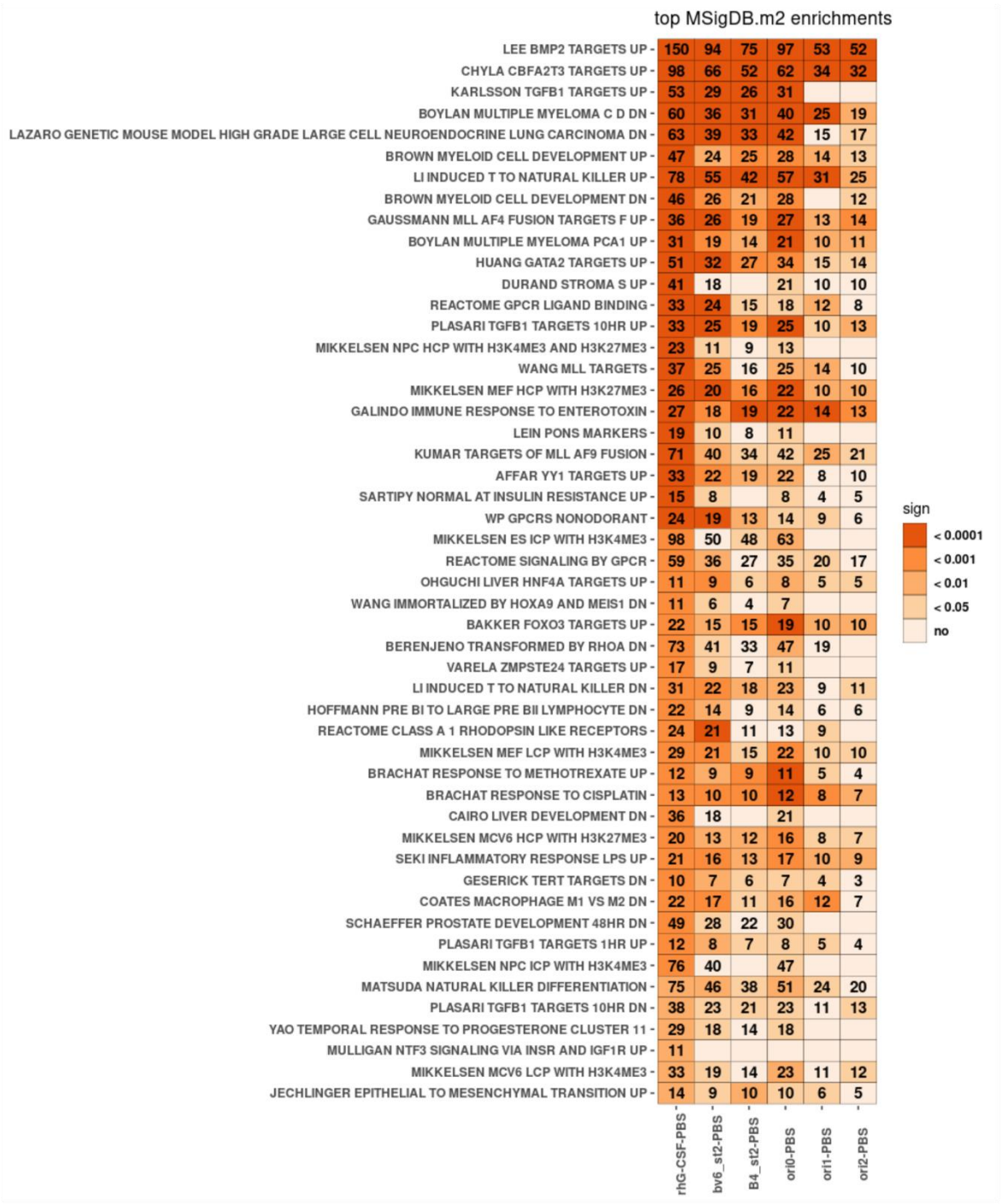

**Supplementary Figure 17. (A-F)** Western Blot analysis of extracts from NFS-60 cells treated with B4_st2, bv6_st2, ori0, ori1, ori2 designs, PBS, or rhG-CSF for 5 or 30 min, using phospho-STAT3 (Tyr705) (A, C, E) or phospho-STAT5 (Tyr694) (B, C, D) rabbit antibodies. As loading control, α-tubulin rabbit antibody or GAPDH rabbit antibody was used. A and B, C and D, E and F represent independent experiments.




**Supplementary Figure 18. (A)** Principal component analysis (PCA) of groups treated with rhG-CSF, agonist designs (B4_st2, bv6_st2, ori0, ori1, ori2) or PBS and **(B)** the top 20 differentially expressed genes in PC1 of PCA.




**Supplementary Figure 19.** Top 50 GO pathways in the groups treated with rhG-CSF or agonist designs (B4_st2, bv6_st2, ori0, ori1, ori2) compared to PBS.




**Supplementary Figure 20.** Top 50 MSigDB.m2 pathways in the groups treated with rhG-CSF or agonist designs (B4_st2, bv6_st2, ori0, ori1, ori2) compared to PBS.

## Supplementary Tables

Supplementary Table 1. Primers used in this study.

| Number | Name | Sequence |
|---|---|---|
| 1 | pN_insert_SfiI/BglI_F | gtaaggcccagccggccatggcagcactggcagcagcactggcagaaatttac |
| 2 | pN_insert_NotI_R | gtaagcggccgctgctgctgccagatgacgcaggatctg |
| 3 | B4_lib_sat_F | ggtctggatagtctgccgcagNDTtttctgNDTaaagcgctggaaVDGattcgtaaaattcaggcggacgcagcag |
| 4 | B4_lib_sat_R | ttcttccattgcctgcgccatANNANNggcaaaatctgccatatcAHNacgcagagcgtccagtgccggac |
| 5 | B4_lib_des_F1 | ggtctggatagtctgccgcagNDCtttctgYNCaaagcgctggaaMWAattcgtaaaattcaggcggacgcagcag |
| 6 | B4_lib_des_F2 | ggtctggatagtctgccgcagNDCtttctgYNCaaagcgctggaaRTGattcgtaaaattcaggcggacgcagcag |
| 7 | B4_lib_des_R1 | ttcttccattgcctgcgccatGTYGNDggcaaaatctgccatatcTMGacgcagagcgtccagtgccggac |
| 8 | B4_lib_des_R2 | ttcttccattgcctgcgccatTRWGNDggcaaaatctgccatatcTMGacgcagagcgtccagtgccggac |
| 9 | B4_lib_des_R3 | ttcttccattgcctgcgccatGTYGNDggcaaaatctgccatatcGWWacgcagagcgtccagtgccggac |
| 10 | B4_lib_des_R4 | ttcttccattgcctgcgccatTRWGNDggcaaaatctgccatatcGWWacgcagagcgtccagtgccggac |
| 11 | pET28_B4var_F | GGCAGCCATATGGCAGCACTGGCAGCAG |
| 12 | pET28_B4var_R | GTGGTGCTCGAGTTATGCTGCTGCCAGATGACGC |
| 13 | st2_backbone_F | taactcgagcaccaccaccaccaccactga |
| 14 | st2_backbone_R | atggctgccgcgcggcac |
| 15 | st2_fr1_F | AGCAGCGGCCTGGTGCCGCGCGGCAGCCATatggcagcactggcagcagcactggcagaaatttac |
| 16 | st2_fr1_R | GCTGGATCCGCCACCGCCtgctgctgccagatgacgcaggatctg |
| 17 | st2_fr2_F | GGCGGTGGCGGATCCAGCatggcagcactggcagcagcactggcagaaatttac |
| 18 | st2_fr2_R | TCAGTGGTGGTGGTGGTGGTGCTCGAGTTAtgctgctgccagatgacgcaggatctg |
| 19 | pN_Seq_insert_F | gaatggtcaagttgtcgac |

| | | |
|---|---|---|
| 20 | pN_Seq_insert_R | Gaagaaagcgaaaggagc |
| 21 | ori1_fr1_R | CATCCAGCGCCACCAGCGTTTGCGTGCCAGATGACGCAG GATCTGG |
| 22 | ori1_fr2_F | CGCTGGTGGCGCTGGATGCTGCGCAAAGCACTGGCAGAA ATTTACAAAGGTCTGGC |
| 23 | ori2_fr1_R | TTCGCGTTTTTTTTTCATGCGTTTCGCAAAGCGACGCAGG ATCTGGTACGCTTTTTCCAG |
| 24 | ori2_fr2_F | ATGAAAAAAAAAACGCGAATGGATGGCGAAAACCCTGGC AGAAATTTACAAAGGTCTGGCGGAATATC |
| 25 | ori3_fr1_R | CCATTTTTTGCGCATGCGGCGCGCCAGCTGACGCAGGATC TGGTACGCTTTTTCCAG |
| 26 | ori3_fr2_F | CGCATGCGCAAAAAATGGGCGCGCGAAATGATGGCAGCA CTGGCAGAAATTTACAAAGGTCTGG |
| 27 | ori4_fr1_R | CGCCATCATTTTGCGCATCATGCGGCGCGCCAGGCGACGC AGGATCTGGTACGCTTTTTCCAG |
| 28 | ori4_fr2_F | ATGCGCAAAATGATGGCGGAACGCATGGCGCGCGCACTG GCAGAAATTTACAAAGGTCTGGC |
| 29 | ori0_fr1_R | CCATTTCATCCACCACCATGCTGCCAGATGACGCAGGATC |
| 30 | ori0_fr2_F | TGGTGGTGGATGAAATGGCTGGCAGCAGCACTGGCAGAA ATTTAC |

Supplementary Table 2: Composition of PCR reaction mix used for all reactions with the following specifications; for all reactions 5 ng/µL template was inputted but for the ones whose products were used for assembly reactions, 0.3 ng/mL were inputted. In the case of the use of multiple primers for the library backbone PCR, all F or R primers of each construct were mixed at a total concentration of 10 µM (e.g. 2.5 µM of primer 7 to 10 each). PCR program was used for all reactions either with 60 seconds extension time for all insert fragments or 340 seconds for all backbone fragments.

| **PCR reaction mix** | Volume (µL) |
|---|---|
| dNTPs 10 mM each | 1 |
| Template 0.3 ng/µL or 5 ng/µL | 1 |
| Primer F 10 µM | 2.5 |
| Primer R 10 µM | 2.5 |
| 5x Q5® Reaction Buffer (#B9027, NEB) | 10 |
| GC enhancer | 10 |
| Q5® High-Fidelity DNA Polymerase (#M0491S, NEB) | 0.5 |
| $H_2O$ ad 50 µl | 22.5 |

| **PCR program** | Temperature (°C) | Time (s) | Cycle |
|---|---|---|---|
| Initial denaturation | 98 | 30 | 1x |
| Denaturation | 98 | 10 | 30x |
| Extension | 72 | 60 or 340 | |
| Final Extension | 72 | 300 | 1x |
| Hold | 8 | - | 1x |

Supplementary Table 3: Sequence of the fusion protein between Intimin [45] and Boskar4 coded by pNB4 (colored) and sequences of all Boskar4 variants and designs (black) described in this study

| **Name** | **Sequence** |
|---|---|
| Intimin$_{EHEC}$ (1-659)-E_tag-Boskar4-Myc_tag | MITHGCYTRTRHKHKLKKTLIMLSAGLGLFFYVNQNSFANGENYFKLGSDSKLL THDSYQNRLFYTLKTGETVADLSKSQDINLSTIWSLNKHLYSSESEMMKAAPGQ QIILPLKKLPFEYSALPLLGSAPLVAAGGVAGHTNKLTKMSPDVTKSNMTDDKA LNYAAQQAASLGSQLQSRSLNGDYAKDTALGIAGNQASSQLQAWLQHYGTAE VNLQSGNNFDGSSLDFLLPFYDSEKMLAFGQVGARYIDSRFTANLGAGQRFFLP ANMLGYNVFIDQDFSGDNTRLGIGGEYWRDYFKSSVNGYFRMSGWHESYNKK DYDERPANGFDIRFNGYLPSYPALGAKLIYEQYYGDNVALFNSDKLQSNPGAAT VGVNYTPIPLVTMGIDYRHGTGNENDLLYSMQFRYQFDKSWSQQIEPQYVNEL RTLSGSRYDLVQRNNNIILEYKKQDILSLNIPHDINGTEHSTQKIQLIVKSKYGLD RIVWDDSALRSQGGQIQHSGSQSAQDYQAILPAYVQGGSNIYKVTARAYDRNG NSSNNVQLTITVLSNGQVVDQVGVTDFTADKTSAKADNADTITYTATVKKNGV AQANVPVSFNIVSGTATLGANSAKTDANGKATVTLKSSTPGQVVVSAKTAEMT SALNASAVIFFDGAPVPYPDPLEPAQPAMAALAAALAEIYKGLAEYQARLKSLE GISPELGPALDALRLDMADFATTMAQAMEEGLDSLPQSFLLKALEQIRKIQADA AALREKLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAAAAAAEQKLIS EEDAAA |
| **bv1** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAYLMAQA MEEGLDSLPQLFLYKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ LEAFLEKAYQILRHLAAA |
| **bv2** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFACFMAQA MEEGLDSLPQIFLYKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL EAFLEKAYQILRHLAAA |
| **bv3** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAYFMAQA MEEGLDSLPQGFLHKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ LEAFLEKAYQILRHLAAA |
| **bv4** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAYFMAQA MEEGLDSLPQIFLLKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL EAFLEKAYQILRHLAAA |
| **bv5** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRFDMADFAVFMAQA MEEGLDSLPQFFLFKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL EAFLEKAYQILRHLAAA |
| **bv6** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAILMAQA MEEGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ LEAFLEKAYQILRHLAAA |
| **bv7** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRFDMADFAYFMAQA MEEGLDSLPQSFLYKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL EAFLEKAYQILRHLAAA |
| **bv8** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRLDMADFARLMAQA MEEGLDSLPQRFLYKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ LEAFLEKAYQILRHLAAA |

| | |
|---|---|
| **bv9** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAIYMAQA<br>MEEGLDSLPQHFLLKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAA |
| **bv10** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAYFMAQA<br>MEEGLDSLPQSFLLKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL<br>EAFLEKAYQILRHLAAA |
| **bv11** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRHDMADFAYFMAQA<br>MEEGLDSLPQGFLDKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAA |
| **bv12** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFALFMAQA<br>MEEGLDSLPQCFLLKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL<br>EAFLEKAYQILRHLAAA |
| **bv13** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRLDMADFAFFMAQA<br>MEEGLDSLPQSFLYKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL<br>EAFLEKAYQILRHLAAA |
| **bv14** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAFNMAQA<br>MEEGLDSLPQGFLYKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAA |
| **bv15** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAYNMAQA<br>MEEGLDSLPQHFLLKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAA |
| **bv16** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRLDMADFATNMAQA<br>MEEGLDSLPQFFLYKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAA |
| **bv1_st2** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAYLMAQA<br>MEEGLDSLPQLFLYKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAAGGGGSSMAALAAALAEIYKGLAEYQARLKSLEGISP<br>ELGPALDALRYDMADFAYLMAQAMEEGLDSLPQLFLYKALEMIRKIQADAAAL<br>REKLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |
| **bv2_st2** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFACFMAQA<br>MEEGLDSLPQIFLYKALEQIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL<br>EAFLEKAYQILRHLAAAGGGGSSMAALAAALAEIYKGLAEYQARLKSLEGISPE<br>LGPALDALRYDMADFACFMAQAMEEGLDSLPQIFLYKALEQIRKIQADAAALR<br>EKLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |
| **bv6_st2** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAILMAQA<br>MEEGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAAGGGGSSMAALAAALAEIYKGLAEYQARLKSLEGISP<br>ELGPALDALRYDMADFAILMAQAMEEGLDSLPQSFLRKALEMIRKIQADAAAL<br>REKLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |
| **bv8_st2** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRLDMADFARLMAQA<br>MEEGLDSLPQRFLYKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAAGGGGSSMAALAAALAEIYKGLAEYQARLKSLEGISP |

| | |
|---|---|
| | ELGPALDALRLDMADFARLMAQAMEEGLDSLPQRFLYKALEMIRKIQADAAAL<br>REKLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |
| **bv15_st2** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAYNMAQA<br>MEEGLDSLPQHFLLKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAAGGGGSSMAALAAALAEIYKGLAEYQARLKSLEGISP<br>ELGPALDALRYDMADFAYNMAQAMEEGLDSLPQHFLLKALEMIRKIQADAAA<br>LREKLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |
| **bv16_st2** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRLDMADFATNMAQA<br>MEEGLDSLPQFFLYKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAAGGGGSSMAALAAALAEIYKGLAEYQARLKSLEGISP<br>ELGPALDALRLDMADFATNMAQAMEEGLDSLPQFFLYKALEMIRKIQADAAAL<br>REKLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |
| **ori1** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAILMAQA<br>MEEGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLARKRWWRWMLRKALAEIYKGLAEYQARLKSLEGISPEL<br>GPALDALRYDMADFAILMAQAMEEGLDSLPQSFLRKALEMIRKIQADAAALRE<br>KLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |
| **ori2** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAILMAQA<br>MEEGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRRFAKRMKKKREWMAKTLAEIYKGLAEYQARLKSLEGISPEL<br>GPALDALRYDMADFAILMAQAMEEGLDSLPQSFLRKALEMIRKIQADAAALRE<br>KLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |
| **ori3** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAILMAQA<br>MEEGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRQLARRMRKKWAREMMAALAEIYKGLAEYQARLKSLEGISP<br>ELGPALDALRYDMADFAILMAQAMEEGLDSLPQSFLRKALEMIRKIQADAAAL<br>REKLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |
| **ori4** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAILMAQA<br>MEEGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRRLARRMMRKMMAERMARALAEIYKGLAEYQARLKSLEGIS<br>PELGPALDALRYDMADFAILMAQAMEEGLDSLPQSFLRKALEMIRKIQADAAAL<br>REKLAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |
| **ori0** | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAILMAQA<br>MEEGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQ<br>LEAFLEKAYQILRHLAAWWWMKWLAAALAEIYKGLAEYQARLKSLEGISPELG<br>PALDALRYDMADFAILMAQAMEEGLDSLPQSFLRKALEMIRKIQADAAALREK<br>LAATYKGNDRAAAAVEIAAQLEAFLEKAYQILRHLAAA |

## Supplementary Methods

***Damietta specifications for the design of orientation-rigging ligands***

Example Damietta specifications for weights, sample parameters, the mutable residues, and repackable residues for designing the rigid linker helix region of ori0 (compare Material and Methods, Design of orientation-rigging ligands):

```
# mutational targets
mut_res 124 ADEFHIKLMNQRSTVWY
mut_res 125 ADEFHIKLMNQRSTVWY
mut_res 126 ADEFHIKLMNQRSTVWY
mut_res 127 ADEFHIKLMNQRSTVWY
mut_res 128 KREQ
mut_res 129 ADEFHIKLMNQRSTVWY

# repacking targets
rpk_res 61
rpk_res 63
rpk_res 66
rpk_res 119
rpk_res 120
rpk_res 121
rpk_res 122
rpk_res 123
rpk_res 130
rpk_res 131
rpk_res 132
rpk_res 133
rpk_res 134
rpk_res 184
rpk_res 185
rpk_res 247
rpk_res 250

# sampling parameters (optional)
scramble_order        1 # default:= 0
m_mutations                 3 # default:= 3
n_paths                     7 # default:= 1
n_iters                     5 # default:= 1

# mutagenesis scoring weights (optional)
mut_max_lj                  25.0
mut_w_pp              1.0
mut_w_k               0.0
mut_w_lj              1.0
mut_w_solv                  1.0
mut_w_elec                  0.125

# repacking scoring weights (optional)
rpk_max_lj            25.0
rpk_w_pp        1.0
rpk_w_k         1.0
rpk_w_lj        1.0
rpk_w_solv            1.0
rpk_w_elec            0.125
```

# *De novo* design of G-CSFR-inhibiting proteins

Timo Ullrich[1], Olga Klimenkova[2], Marcus D. Hartmann[1,3], Julia Skokowa[2], Mohammad ElGamacy[1,2*]

[1] Max-Planck-Institute of Biology, Department Protein Evolution, Tübingen
[2] Translational oncology, Internal Medicine II, University Hospital Tübingen
[3] Interfaculty Institute of Biochemistry, University of Tübingen, Germany
* corresponding author: mohammad.elgamacy@med.uni-tuebingen.de

***Abstract***
Excessive cytokine signaling resulting from dysregulation of a cytokine or its receptor can be a main driver of cancer, autoimmune or hematopoietic disorders. Here we leverage protein design to create tailored cytokine receptor blockers with idealized properties. Specifically, we aimed to tackle the granulocyte-colony stimulating factor receptor (G-CSFR), a mediator of different types of leukemia and autoinflammatory diseases. By creating strictly monovalent G-CSFR binders, we identified hyper-stable, nanomolar signaling inhibitors. X-ray crystallography indicated atomic-level agreement with the experimental structure of an exemplary design. Furthermore, the most potent design blocks G-CSFR in acute myeloid leukemia cells and primary hematopoietic stem cells. Thus, the resulting designs can be used for inhibiting or homing to G-CSFR-expressing cells, and demonstrate a generalizable strategy to tackling a diverse range of cytokine receptors.

***Introduction***
Inhibiting cytokine signaling can be highly effective in treating various cancers, autoimmune and hematologic disorders. Highly specific protein-based therapies have been developed to block cytokine signaling to tackle a range of diseases [1]. These include cytokine-quenching antibodies [2, 3] and recombinant soluble receptors [4, 5], or alternatively, receptor-blocking antibodies [6, 7] and cytokine-based inhibitors [8, 9]. *De novo* protein design unlocks new strategies for creating purpose-tailored inhibitors with optimal structural, biophysical, and functional features [10, 11]. Here we describe a strategy for inhibiting cytokine receptor signaling based on designed receptor binders with idealized structural and biophysical properties. We use this approach with the aim of inhibiting the granulocyte-colony stimulating factor receptor (G-CSFR); a crucial driver of granulopoiesis, that when dysregulated can promote tumor progression and inflammatory diseases [12].

G-CSFR is activated through ligand-induced dimerization of the ectodomains of two receptor subunits, which leads to the cross-phosphorylation and activation of the intracellular receptor-associated kinases [13]. Physiologically, G-CSFR signaling is primarily responsible for the proliferation and differentiation of hematopoietic stem cells into neutrophils, and for mobilization of hematopoietic cells and neutrophils into peripheral blood from the bone marrow [14]. Additionally, it has pro-proliferative and pro-survival effects on neuronal, vascular, and myocardial tissues [15-18]. Conversely, G-CSFR signaling promotes several types of leukemia [13, 19, 20] and solid tumors [21-25]. It also constitutes a driver of inflammatory disorders [26, 27], including rheumatoid arthritis [28-30], further motivating the development of specific G-CSFR inhibitors.

The native ligand (G-CSF) engages and dimerizes G-CSFR through two binding sites in a crossover 2:2 receptor:ligand complex that juxtaposes the receptor subunits into a signaling-competent configuration. Through this study we aimed to design a series of bifaceted, *de novo*-designed ligands to engage G-CSFR in non-signaling configurations; outcompeting native G-CSF (Fig. 1A). We follow this up by a detailed biophysical and functional characterization to delineate the inhibitory potential of the most promising designs. We further carry out computationally-guided functional enhancement, and demonstrate the capacity of our designs to inhibit the effect of G-CSF in acute myeloid leukemia and primary hematopoietic stem cells. This approach of cytokine receptor inhibition is superior to

previously described strategies and can be readily extended to tackle other target cytokine and growth factor receptors.

## *Material and methods*

### *Modeling of bifaceted variants of G-CSFR-binding designs*

The original Boskar4 template encodes a single G-CSFR-binding site (PDB: 7NY0 [31]) that matches site II of the G-CSF (PDB: 2D9Q [32]). This binding site is mounted across the helix 2 and helix 3 (H2/3) regions of the Boskar4 starting template, and is herein referred to as the primary binding site. Boskar4 structure however allows for grafting another binding site on helices 1 and 4 (H1/H4) due to the pseudosymmetry of the 4-helix-bundle, which can constitute a secondary binding site. This additional graft of a secondary binding site on H1/H4 was performed in both parallel and antiparallel orientations with respect to the primary binding site, where an AlphaFold2 [33] model of the affinity-enhanced variant of Boskar4; bv6 [34] was used. The antiparallel grafts; named *boa* (bivalent-orientation antiparallel) were modelled using either a selection of 12 or 15 binding-site residues, denoted as boa1 and boa2, respectively (Table S1). The parallel orientation grafts, named *bop* (bivalent-orientation parallel), were modelled using either 12 or 17 binding-site residues, denoted as bop1 and bop2, respectively (Table S1). The grafting position was adjusted to minimize the backbone alignment RMSD of the re-grafted site, where the resulting models and complexes with the receptor were modelled using AlphaFold2 [33, 35]. The sequences of the final designs are provided in Table S2.

### *Protein over-expression and purification*

Synthetic genes of all designs were cloned into a pET28a downstream of N-terminal His-tag and thrombin cleavage site. Expression was performed in *Escherichia coli* (*E. coli*) BL21(DE3) cells grown in LB medium supplemented with 40 µg/ml kanamycin in a shaking flask incubator 37 °C and 160 rpm. Cells were induced at an optical density at 600 nm ($OD_{600}$) of approximately 0.6, using 0.5 mM IPTG, and the expression of the target protein was carried out at 25 °C and 160 rpm for approximately 16 hours. After 16 hours, the cells were harvested by centrifugation at 5000 g for 30 minutes. The supernatant was discarded, and the cell pellet was lysed using ultrasonic homogenization in a lysis buffer containing 50 mM Tris (pH 8), 100 mM NaCl, 20 µg/ml DNaseI (A3778, ITW Reagents), and a protease inhibitor cocktail (04693132001, Roche). The lysate was then centrifugated at 16,000 g for 30 minutes, and the resulting supernatant was loaded onto nickel immobilized metal affinity chromatography (IMAC) beads for batch purification (11912422, Macherey-Nagel). Subsequently, size exclusion chromatography (SEC) was performed on a Superdex® 75 Increase 10/300 GL (29148721, Cytiva) using PBS as the mobile phase. Fractions expected to contain the protein of the desired size were concentrated using ultrafiltration with a 10 kDa cutoff membrane (UFC8010, Amicon® Ultra) and were analyzed again with SEC under the same conditions. Aliquots of these concentrated fractions were frozen at -20 °C until further use.

### *Analysis of folding stability*

Circular dichroism (CD) spectra were obtained using a JASCO J-810 spectrometer for protein samples diluted to a concentration of 0.2 mg/ml in PBS buffer at pH 7.0. The CD spectra were recorded by measuring the residual ellipticity of three accumulations, with a resolution of 0.1 nm over a range of 250-190 nm. To determine the melting temperature with CD, the average residual ellipticity was measured at a wavelength of 222 nm over a temperature range of 20 °C to 100 °C with a temperature increase of 0.5 °C per minute. Nanoscale-differential scanning fluorimetry (nanoDSF) experiments were carried out using a Prometheus NT.48 instrument (Nanotemper Technologies) in standard Prometheus capillaries (Nanotemper, PR-C002). These measurements were made on protein samples with a concentration of 1 mg/mL in PBS. The temperature ramp for melting and cooling was set at 1 °C per minute, ramping from 25 °C to 110 °C, followed by a reverse ramp back to 25 °C.

***Binding affinity determination using surface plasmon resonance***

To evaluate the affinity of the designs to G-CSFR, multi-cycle kinetics experiments were conducted using a Biacore X100 system (GE Healthcare Life Sciences). Recombinant human G-CSFR (Glu25-Pro621, 381-GR/CF, R&D Systems) was diluted to a concentration of 50 µg/mL in a 10 mM acetate buffer at pH 5.0 and immobilized on a CM5 sensor chip (GE Healthcare 29149604) using amine coupling chemistry. The protein samples were diluted in a running buffer consisting of PBS with 0.05% v/v Tween-20. The experiments were carried out at a temperature of 25 °C and a flow rate of 30 µL/min. The sample solutions were injected sequentially over the functionalized sensor chip surface for 180 s, followed by a dissociation phase of 600 s with running buffer. Four different concentrations for each sample were measured which are indicated in the corresponding legend of the corresponding figure. After each run, the sensor surface was regenerated by injecting a 50 mM NaOH solution for 60 s. To analyze the data, the reference responses and zero-concentration sensograms were subtracted from each dataset (double-referencing). The Biacore X100 Evaluation Software following a 1:1 binding kinetic model was used to determine the association rate ($k_a$), dissociation rate ($k_d$), and equilibrium dissociation constant ($K_D$) constants. Leave-one-out cross-validation was used to determine mean and standard deviation of $k_a$, $k_d$, and $K_D$.

***NFS-60 cell proliferation assay***

The NFS-60 cells [36, 37] were cultured at 5% $CO_2$ and 37 °C in NFS-60-medium, which consisted of RPMI 1640 medium supplemented with 1 mM L-glutamine, 1 mM Na-pyruvate, 10% FCS, 1% Antibiotic-Antimycotic (15240062, Gibco™), and 12.5% KMG-2: 5637-CM (Conditioned Medium from CLS Cell Lines Service). Prior to the proliferation assay, the cells were washed three times with NFS-60-medium without KMG-2. The assay was conducted in black 96-well plates. Each well contained 45,000 cells in a total volume of 150 µL NFS-60-medium without KMG-2, cultured under maintenance conditions with different concentrations of the protein sample for 48 hours. Following that, 30 µL of CellTiter-Blue® Reagent (G808, Promega) was added to each well, and the cells were further incubated for approximately 90 minutes at 5% $CO_2$ and 37 °C. Subsequently, the fluorescence of each well was measured on a plate reader with an excitation wavelength of 560 nm and emission at 590 nm. The half-maximal effective concentration ($EC_{50}$) of each protein sample was determined by fitting the obtained fluorescence values to their corresponding concentrations using a four-parameter sigmoidal function minimized with the Nelder-Mead Simplex algorithm as implemented in the SciPy [38].

All NFS-60 cell proliferation assays were conducted as described above unless indicated otherwise. In live cell tracking assays, cells were monitored in the IncuCyte S3 Live-Cell Analysis System (Essen Bio) with a 10x objective. Cell proliferation was analyzed using IncuCyte S3 Software. For these experiments NFS-60 cells were maintained at 5% CO2 and 37 °C in NFS-60-medium (RPMI 1640 medium with additional 2 mM L-glutamine, 1 mM Na-pyruvate, 10% FCS, 1% Pen-Strep and 10 ng/mL rhG-CSF). Prior to the proliferation assay, the cells were washed three times with PBS and starved for at least 4 hours in NFS-60 medium without rhG-CSF. The assay was then performed in an L-ornithine pretreated 96-well plate, with 45,000 cells seeded per well.

***NFS-60 competitive inhibition assay***

The NFS-60 cells were cultured under the same condition as described for the NFS-60 cell proliferation assays with the following changes. After washing the cells three times with NFS-60-medium without KMG-2, a final constant concentration of 50 pg/mL of rhG-CSF (unless otherwise indicated in the figures) was applied to the cells together with varying concentrations of the designed proteins. The choice of rhG-CSF concentration mimics the reported physiological serum levels in healthy individuals [39]. Fluorescence measurements were acquired for each experimental condition and subsequently min-max normalized against control samples. These control samples consisted of cells treated with and without the constant concentration of rhG-CSF. From these values the half-maximal inhibitory

concentration ($IC_{50}$) was determined in the same way as described for the $EC_{50}$ parameter in the proliferation assays.

***Structure determination of bop1 using X-ray crystallography***

Crystallization screens were set up with a Mosquito robot (TTP Labtech) in 96-well plates at 21 °C, using 50 µL of reservoir solution and sitting drops containing 400 nL of reservoir and 400 nL of protein solution of bop1 at a concentration of 9 mg/ml. Within 14-30 days, crystals of bop1 grew in a condition 0.1 M MES pH 7.0, 20 % (w/w) PEG 8,000. Crystals were cryoprotected through the addition of 20% PEG 400, flash cooled and stored in liquid nitrogen until data collection. Diffraction data were collected at 100K on an EIGER 16M detector at beamline X10SA at the Swiss Light Source (SLS). Data were processed and scaled using XDS [40] and the structures solved using molecular replacement with MOLREP [41] and the AF2 model of bop1 as a search model. The structure was completed by cyclic refinement with REFMAC5 [42] and modelling using Coot [43]. Data collection and refinement statistics, coordinates, and structure factors are deposited in the PDB under accession code 8QUP.

***Computationally-guided affinity maturation***

In order to further enhance the binding affinity of the primary binding site of bop1, we selected 11 amino acid positions for computational optimization. The structural models were built as a complex of each of the 17 NMR frames of Boskar4 (7NY0 [31]) structurally-aligned to the G-CSF:G-CSFR complex (2D9Q [32]). Two starting template sequences of the ligand were based on the enhanced affinity variants bv6 and bv8 [34] and modelled separately into the complexes using Modeller [44]; yielding a total of 34 complexes. For each of these complexes all 11 positions were mutated *in silico* to all proteinogenic amino acids (but G, P, and C) utilizing the combinatorial sampler of Damietta (v0.32; supplementary methods) [45]. The combinatorial design simulations were performed in ten scrambled-order replicas for each modelled complex, resulting into a total of 612 unique designed amino acid sequences. The sequences of the starting variants bv6 and bv8 were added to this pool of unique sequences, resulting in a total of 614 sequences. The obtained amino acid sequences were split into two pools of contiguous fragments, the first encoding the binding protein till L55 (numbering of PDB:7NY0) and the second starting from D56 (number of PDB:7NY0) till the end of the protein sequence. The two amino acid sequence pools were used to design two oligo pools encoding for the corresponding mutants. In total oligo pool 1 covered 298 unique oligos, and oligo pool 2 included 330 unique oligos. These two oligo pools were ordered (IDT, Inc.) encoding the desired mutations on a forward and reverse primer pair (supplementary data) to generate a library with a maximum degeneracy and diversity of $9.8\times10^4$. Finally, the library was generated by PCR using the two oligo pools to linearize an *E. coli* display system plasmid [46] encoding a fusion between the N-terminal part of intimin and Boskar4 (pNB4, Table S2). The obtained PCR product was purified with the Wizard® SV Gel and PCR Clean-Up System (A928, Promega) and blunt end ligation was performed followed by electroporation of freshly made competent *E. coli* DH10BT1R cells [47] to yield $6.0\times10^6$ total transformants.

In more detail, prior to transformation in DH10BT1R cells, 0.4 U/µL T4 polynucleotide kinase was incubated with 5 ng/µL of linearize PCR product for 30 min at 37 °C, according to manufacturer instructions. After that, 0.25 U/µL T4 DNA ligase was added and incubated overnight at 16 °C. The ligation product was purified with the PCR Clean-Up System (A928, Promega) and eluted into 20 µL filtered ddH2O. The complete 20 µL were transformed into electrocompetent *E. coli* as described in Tu *et al.* [47]. Specifically, 50 mL of freshly grown *E. coli* (DH10B $T1^R$) of an $OD_{600}$ of ~0.6 was washed two times with 35 mL filtered ddH2O. After washing, the bacterial pellet was resuspended with the purified ligation mix. The electroporation was performed at 1250 V and 5 ms. The cells were grown in 1 mL SOC Outgrowth Medium (B9020, NEB) for 2 h in a standard glass tube at 30 °C with 160 rpm. A 10-fold dilution series of the cells were made in SOC-medium and they were plated on agar plates containing 34 µg/mL chloramphenicol and 2% (w/v) D-glucose to estimate the total number of transformants. The rest of the sample was plated on five agar plates of the same type and incubated at 30 °C for about 20 h. The bacterial lawn was scraped from the plates and well mixed in 10 ml LB-

medium. The plasmids were isolated from 500 µL of bacterial suspension and the obtained plasmid library was analyzed by sanger sequencing. Single Boskar4 variant genes was amplified by PCR from the pN vector [46] and ligated between NdeI and XhoI replacing Boskar4 of pET28a.

To perform cell sorting of the bacterial clones bound to G-CSFR, fluorescently-labelled rhG-CSFR was prepared according to the protocol outlined in Salema *et al.* [46]. Specifically, a total of 100 µg of recombinant human G-CSFR (381-GR/CF, R&D Systems) was subjected to a reaction with Biotin-NHS (H1759, Sigma) in a 1:20 molar ratio (G-CSFR:Biotin-NHS), within a 950 µL solution of PBS. The mixture was incubated at room temperature (RT) for a duration of 2 hours. Subsequently, the reaction was quenched by adding 50 µL of Tris (1 M pH 7.5). The resulting product underwent purification through a desalting column (Sephadex G25 PD-10, GE Healthcare), followed by concentration of elution fractions via ultrafiltration (3 kDa, UFC800324, Merck). The biotinylated rhG-CSFR (BioG-CSFR) was then aliquoted and stored at -20 °C until it was utilized for further experiments.

The screening process of the library also followed a modified protocol based on the procedure originally described in [46]. Initially, DH10BT1R *E. coli* cells were transformed using the display vector containing the protein library. The bacteria were grown on LB agar plates with 34 µg/mL chloramphenicol and 2% (w/v) D-glucose at 30 °C. After harvesting the plates, 1 mL of resuspended bacteria with an $OD_{600}$ of 3.3 was washed twice with 1 mL LB and then diluted to a starting $OD_{600}$ of 0.33 in 10 ml LB with 34 µg/mL chloramphenicol. The cells were grown at 30 °C and 160 rpm and after 1 hour, 50 µM IPTG was added to the culture, and it was further grown for 3 hours under the same conditions. Cells were collected (equal to 1 mL of a cell resuspension with an $OD_{600}$ of 1) were washed twice with 1 mL PBS and resuspended in 400 µL PBS. In the next step, 190 µL of this cell suspension was incubated with 10 nM BioG-CSFR at room temperature for 1 hour, followed by a single wash with 1 mL PBS. The washed cell pellet was resuspended in in 200 µL PBS, to which 0.375 µL PE-Streptavidin (405203, BioLegend) was added, and incubated at 4 °C for 30 minutes. Finally, the sample was washed once more with 1 mL PBS and resuspended in 1 mL PBS. The screening was performed using a BD FACSMelody™ Cell Sorter having set the PTM voltage for PE to 517 V, SSC to 426 V and an SSC threshold of 673 V. No gating was applied, and 100,000 events were measured per sample. For sorting, 300 µL of the sample was mixed with 4 mL ice-cold PBS, and the flow rate was adjusted to achieve an event rate of approximately 8000 events/sec. Approximately, the top 0.1% of the total population, based on a SSC-A/PE-A plot, was sorted into a 1.5 mL tube containing 200 µL LB medium, which was cooled to 5 °C. The sorting process continued until at least 20 times more events were screened than the complexity of the corresponding library. All sorted cells were plated on LB agar plates with 34 µg/ml chloramphenicol and 2% (w/v) D-glucose and were incubated at 30 °C for around 20 hours. The harvested plates were then grown overnight in LB medium with 34 µg/mL chloramphenicol and 2% (w/v) D-glucose at 30 °C under static conditions. From this culture, 1 mL of resuspended bacteria with an $OD_{600}$ of 3.3 was used for the next selection cycle performed in the same way, and repeated for five iterative cycles. All flow cytometry results were analyzed using FlowJo™ v10.1 Software (BD Life Sciences).

Single colonies of bacteria pools which were enriched for 3 or 5 cycles of FACS (as described above) were grown overnight in 100 µL LB (supplemented with 34 µg/mL chloramphenicol and 50 µM IPTG) in a 96-well plate sealed with Breathe-Easy ® sealing membrane (Z380059, Sigma-Aldrich) at 30 °C with 1100 rpm on a benchtop shaker (Thermomixer comfort, Eppendorf) covered with aluminum foil. At the next day the cells were centrifugated at 3200 g for 5 min, resuspended in 150 µL PBS and centrifugated again. The washing was repeated one more time and the cells got resuspend in 50 µL PBS containing 10 nM BioGCSFR. After the plate was incubated for 1 h at room temperature, the cells were washed one time with 200 µL PBS and resuspended in 50 µL PBS. 25 µL of the resuspended cells were mixed with 25 µL PBS containing 1.25 µg/mL Avidin-HRP conjugate (Invitrogen, 434423) followed by a 30 min incubation at 4 °C in the dark. Afterwards the cells were washed one time with 200 µL

PBS and resuspended in 100 µL PBS. The cell density ($OD_{600}$) of each well was measured with the plate reader. In a separate dark plate (655097, Greiner Bio One) 12.5 µL cell suspension was added to 75 µL PBS. Then 12.5 µL ECL-substrate-solution (1705061, Bio-Rad) was pipetted quickly into each well, the plate was sealed with a parafilm and mixed vigorously for 10 seconds on a vortex orbital shaker. Then the luminescence of each well was measured on the plate reader. The luminescence signal of each variant in each well was analyzed by normalizing to the corresponding cell density and calculating a Z-score for each sample over all wells of the same condition. From the top 5 variants of each condition plasmids were isolated with a plasmid preparation kit (740588, Macherey-Nagel) from a 3 ml overnight culture grown at 30 °C with 160 rpm in LB with 34 µg/mL chloramphenicol and 2% (w/v) D-glucose. Afterwards, the acquired plasmids were sequenced, and distinct Boskar4 variant (bv) genes were identified which were designated as bv21, bv22, bv23, and bv24. The Alignment consensus logos of the mutated sites were illustrated using AlignmentViewer [48].

***Kasumi-1 competitive inhibition assay***

Kasumi-1 cells [49] were cultured at 5% $CO_2$ and 37 °C in RPMI 1640 medium supplemented with 1 mM L-glutamine, 10% FCS, and 1% Antibiotic-Antimycotic (15240062, Gibco™). In a black 96-well plate, 40,000 cells were seeded in each well with a constant concentration of 400 pg/mL of rhG-CSF, alongside varying concentrations of the designed proteins. The cells were then incubated under standard conditions for 72 hours. CellTiter-Blue® Reagent was added into each well, and the resulting fluorescence was measured as described in the NFS-60 cell proliferation assay. The fluorescence values obtained were subsequently normalized to the cells treated with 400 pg/mL of rhG-CSF. Finally, the $IC_{50}$ was determined as described above. Leave-one-out cross-validation was used to calculate the mean and standard deviation of the $IC_{50}$ across the three biological replicates. This approach was chosen because the biological replicates encompassed varying numbers of treatments.

***Evaluation of dose-dependent effects of inhibitors on the proliferation of $CD34^+$ HSPCs***

Human $CD34^+$ cells were isolated from the bone marrow mononuclear cell fraction of two healthy donors by the Ficoll density gradient centrifugation with subsequent magnetic bead separation using the Human CD34 Progenitor Cell Isolation Kit (Miltenyi Biotech Germany; #130-046-703). Isolated cells were frozen till the experimentation. After thawing, $CD34^+$ cells were cultured for up to five days in Stemline II Hematopoietic Stem Cell Expansion medium (Sigma Aldrich; #50192) supplemented with 10% FBS, 1% penicillin/streptomycin, 1% L-glutamine and 20 ng/ml IL-3, 20 ng/mL IL-6, 20 ng/mL TPO, 50 ng/mL SCF, and 50 ng/mL FLT-3L. Before experimentation, cells were washed twice in ice-cold PBS and $2 \times 10^4$ cells/well were incubated in poly-L-lysine–coated 96-well plates for five days in an IncuCyte S3 Live-Cell Analysis System (Essen Bio) with a 10x objective at 37°C and 5% $CO_2$. Cells were incubated in the same Stemline II Hematopoietic Stem Cell Expansion medium supplemented with 10% FBS, 1% penicillin/streptomycin, 1% L-glutamine and 1 ng/mL of rhG-CSF or PBS in the presence or absence of designed inhibitors (i.e. bop1 or bop3) or an inert helical protein (moevan_control [50]), at concentrations ranging from 5 µg/mL to 0.1 µg/mL (as indicated in the respective figure). Cell proliferation was analysed using IncuCyte S3 Software. All experiments in this study involving human-derived HSCPs samples were conducted according to Helsinki's declaration, and study approval was obtained from the Ethical Review Board of the Medical Faculty, University of Tübingen (012/2021BO2).

***Results***

***Design of single-domain, bifaceted G-CSFR binders***

In this study we used Boskar4 as our starting design template to build G-CSFR inhibitors. Boskar4 is a *de novo* designed protein domain that is small (13 kDa), proteolytically and thermally stable, and binds to G-CSFR at the same binding site of the native ligand, G-CSF [31]. Previously we showed potent G-CSFR agonists can be made by tandemly-repeated Boskar4 fusions, which dimerize and activate the

receptor [31]. Through this work we pursue the inverse goal of creating potent G-CSFR inhibitors based on single-domain Boskar4 template. Based on our previous work to create rigid Boskar4-based dimerizers that can associate G-CSFR in non-native geometries we observed that the inter-transmembrane domain (inter-TMD) spacing of the receptor subunits to be a critical parameter affecting activity [34]. Specifically, ligands that impose 2:1 receptor:ligand complexes with longer inter-TMD spacing result in weaker activity and reduced receptor dimerization [34]. However, an additional challenge to this end relates to identifying Boskar4 variants that are exclusively monomeric in solution to avoid any unintended oligomerization, and thus unintended activation, of the receptor. Accordingly, our hypothesis to create inhibitors aimed to exploit the concept of creating a rigid, single-domain, bivalent receptor binder that associates G-CSFR in an inactive configuration (Fig. 1A). Moreover, we also initially hypothesized that the ligand self-interactions could potentially be reduced by repurposing more of the ligand's surface area for receptor binding.

The up-down four-helix-bundle of Boskar4 has a cuboid-like regularity, where only one facet of the cuboid encodes the receptor binding site (herein referred to as the *primary binding site*). Given the pseudosymmetry of the helical bundle, the opposite facet of the bundle can be used as template to graft another *secondary binding site* that can be juxtaposed in a parallel or an antiparallel orientation relative the primary binding site. Modelling the resulting receptor:ligand 2:1 complexes indicated the antiparallel configuration to result in large inter-TMD spacing (292 Å), this compares with 60 Å for the native G-CSF:G-CSFR complex [32]. Such antiparallel spacing would yield in a highly strained complex in the physiological context of membrane-embedded receptor chains (Fig. 1B). Conversely, our models predicted the parallel configuration to be signaling-competent given its relatively short inter-TMD spacing (188 Å), as compared to our modelling of receptor geometries induced by G-CSF or our previously described geometric modulators [34].

To start from a near-optimal design template, we used a structural model of bv6. The bv6 is a functionally-enhanced variant (5-point mutant) of Boskar4 that binds 12-fold tighter to G-CSFR, and functions as a more potent granulopoietic agonist when tandemly repeated in a two-domain fusion [34]. Whereas the primary binding site lies on H2 and H3 (i.e. 2$^{nd}$ and 3$^{rd}$ helix of bv6), the secondary binding site was grafted onto H1 and H4 (i.e. 1$^{st}$ and 4$^{th}$ helix of bv6) in antiparallel orientation to the primary site. We grafted two sets of residues representing a small patch (12 grafted residues) and a large patch (15 grafted residues) of the primary binding site. The final binding sites of corresponding AlphaFold2 models of each design matched the original binding site within 1.13 Å and 1.18 Å backbone atoms root mean square deviation (RMSD) for the small and large grafts, respectively. We refer to this class of designs as *boa* (bivalent-orientation antiparallel), where boa1 or boa2 refer to the designs with either the small or the large secondary binding patches, respectively.

In order to conversely test our hypothesis that the boa designs yield non-activating G-CSFR complexes, we sought to generate control designs with parallel orientation of binding sites based on bv6 as well; *bop* (bivalent-orientation parallel). Likewise, the secondary binding site was grafted from either a small patch (bop1, 12 residues) or a large patch (bop2, 17 residues) from the primary binding sites. Akin to the boa designs, the grafted binding sites of corresponding AlphaFold2 models of bop1 and bop2 were compared to the intended biding site and exhibited 1.17 Å and 1.50 Å backbone atoms RMSD. The backbone RMSD of all designs to the original binding site was on average 0.57 ± 0.08 Å for the primary binding site and 1.24 ± 0.17 Å for the secondary binding site (Fig. S1). Therefore, we expected the secondary binding site to bind G-CSFR less efficiently than the primary binding site. We hence started by experimentally characterizing these four designs (boa1, boa2, bop1 and bop2).

***The bifaceted designs are hyper-stable and exclusively monomeric in solution***

The expression of all four designs in *E. coli* resulted in a high soluble protein yield, which was readily purified to homogeneity using IMAC and preparative SEC. Analytical SEC showed the purified proteins (bop1/2 and boa1/2) to be strictly monomeric in solution (Fig. 2A, B). This soluble, monomeric property is advantageous for obtaining inhibitors that do not inadvertently aggregate or oligomerize the G-CSFR, which might lead to receptor activation. In contrast, the single-domain form of the original design, Boskar4, partitions between monomer and dimer in solution (Fig. S2B), where its isolated monomeric fraction could only weakly inhibit G-CSFR activation (Fig. S2C, D). The starting template in this study, bv6, also showed a tendency to form dimers in solution, similar to Boskar4 (Fig. 2C), which explains its residual capacity to activate G-CSFR (Fig. S2E).

CD measurements revealed strong α-helical profiles for all designs indicating their well-folded nature (Fig. 2D). Additionally, nanoDSF showed all designs to be hyper-thermostable. During the temperature ramp none of the designs showed increased scattering up to 110 °C, highlighting the colloidal stability of the proteins (Fig. S3). For boa1 and boa2, the fluorescence ratio (350 and 330 nm) was unchanged during the heating ramp, while for bop1 and bop2 it indicated a melting onset beyond 90 °C. However, the cooling ramp of bop1 and bop2 (from 110 °C to 25 °C) showed this melting signal onset to be reversible.

***The designs show strong binding and inhibition of G-CSFR***

We then set out to evaluate the binding affinities of the bifaceted designs to G-CSFR using surface plasmon resonance, which was done in a direct comparison to their starting template bv6. The results revealed a pattern whereby all the bivalent designs were more affine than the starting template, and the large-patch grafts (boa2 and bop2) bound G-CSFR tighter in comparison to the small-patch grafts (boa1 and bop1). Specifically, bv6 showed the weakest binding affinity ($K_D$ = 4.3 ± 0.4 nM) in comparison to the other designs (Fig. 2E, S4 and Table S3), which points to the putative role of the secondary binding site in reducing the apparent dissociation rate ($k_d$). Regardless of the orientation of the secondary binding site, its small-patch variants, boa1 ($K_D$ = 3.7 ± 0.8 nM) and bop1 ($K_D$ = 2.7 ± 0.4 nM), displayed weaker receptor-binding affinity than their large-patch counterparts; boa2 ($K_D$ = 1.1 ± 0.4 nM) and bop2 ($K_D$ = 0.9 ± 0.3 nM). These results highlight the symmetric regularity of the template helical bundle, whereby the orientation of the secondary of binding site was less important than the graft size in improving the interaction with the receptor.

To evaluate how the differences in geometry and affinity affect the biological activity of the different designs, we started by testing the proliferative potential of them in a G-CSF-responsive cell line (NFS-60). As expected, these experiments showed the strongest proliferation activator was the design with the large, parallel secondary binding site (i.e. bop2). Also, expectedly, the designs with the antiparallel secondary binding site induced only marginal proliferation increase at a concentration range between 5 and 10 nM concentrations. Interestingly however, the design that showed no measurable residual activity was the one with the small, parallel secondary binding site (bop1) (Fig. 2F). These results highlight that the most efficient receptor dimerizer and activator *in vitro* is indeed is bop2 ($EC_{50}$ = 28.9 ± 7.9 pM; Fig. S5A). This is in line with our previous observation from a geometry-rigging design, ori1, which dimerizes G-CSFR at a similar inter-TMD spacing (i.e. $d_{ori1}$ = 206 Å, while $d_{bop2}$ = 188 Å) [34]. Nonetheless, bop1 did not result in any receptor dimerization, and thus activity, under the same conditions.

To evaluate the inhibitory potential of our designs, we followed up by conducting competitive inhibition assays of rhG-CSF in NFS-60 cells (Materials and Methods). At high concentrations, approximately above 10 > nM, all designs showed substantial inhibition of rhG-CSF-induced proliferation. Lower

concentrations however showed different levels of design-induced proliferation by bop2, boa1, and boa2, which was not the case for bop1 (Fig. 2G, Fig. S5B). Indeed, bop1 treatment resulted in no residual activation at low concentrations, and possessed the highest inhibitory potency ($IC_{50}$ = 35.3 ± 12.9 nM; n=6). To validate the specificity of this inhibitory activity, we evaluated the ability of bop1 to outcompete higher concentrations of rhG-CSF in a concentration-dependent manner, which was true for above-physiological concentrations (100 and 200 pg/mL) of G-CSF [39] (Fig. S6). Moreover, we conducted proliferation assays of $CD34^+$ hematopoietic stem and progenitor cells (HSPCs) to assess the biological activity of bop1 on primary human cells. As expected, there was no observable proliferative activity when the cells were treated with up to 1 µg/mL bop1 (Fig. 2H). Conversely, an inhibition of HSPCs stimulated with 1 ng/mL rhG-CSF and treated with 1 µg/mL bop1 was evident (Fig. 2I). Therefore, we selected bop1 as the design candidate to further analyze it and enhance its activity.

***The structure of bop1 matches the design model at atomic accuracy***

To judge the design accuracy, we sought to solve the crystal structure of the strongest identified inhibitor, bop1. We obtained crystals diffracting to 1.3 Å, with one copy of bop1 in the asymmetric unit, which could be solved by molecular replacement using an AlphaFold2-generated design model (Table S4). Overall, the crystal structure matched the design model to atomic accuracy, with a global backbone $C^\alpha$ RMSD between the structure and the model of 1.03 Å (Fig. 3A). The comparison between the primary binding site residues of bop1 and the native receptor-binding site II of G-CSF (PDB: 2D9Q) showed only minor backbone $C^\alpha$ RMSD of 0.6 Å (Fig. 3B). The alignment was worse however when comparing the secondary binding site residues of bop1 to site II of G-CSF, which yielded and RMSD of 1.2 Å (Fig. 3C). While these results emphasize the weaker structural match, as was expected from the design model, the crystal structure showed no unexpected conformational deformities. This leads us to hypothesize the inability of bop1 to dimerize G-CSFR in cells might be attributed to weak or insufficient receptor interactions through the secondary binding site.

Taken together with the excellent agreement of the bop1 design model and its experimental structure, as well as its superior stability, solubility, and monomeric nature, we were motivated to further improve its inhibitory potency.

***Computationally-guided enhancement of bop1 inhibitory potency***

To improve the inhibitory potency of bop1, we reasoned that enhancing the affinity of only the primary binding site of bop1 can result in stronger G-CSF out-competition, without altering the monomeric nature of bop1, and thus its inability to dimerize G-CSFR. We performed combinatorial design on 11 selected amino acid positions on surface of Boskar4 variants (bv6 and bv8) as mutable and 24 repackable positions across the modelled Boskar4:G-CSFR binding interface, using the Damietta software [45] (Fig. 4A; Supplementary methods). We then created a library from the sequence profile of lowest energy designs (Fig. 4B, S7A, B; Materials and methods). The resulting library of binders was screened using a bacterial display system (Fig. 4C, S7C; Materials and methods). Five sequential cycles of fluorescence-activated cell sorting (FACS) enriched the bacterial clones displaying binders to fluorescently-labelled rhG-CSFR. After the third and fifth sorts, 48 single clones per condition were screened on plate-based assays to identify the clones with the tightest relative binding signal (Fig. S7D). From the 96 analyzed clones, the top 10 clones were sequenced. We identified four unique enhanced Boskar4 variants, named bv21, bv22, bv23 and bv24. Within these variants we observed that 7 mutated positions have reverted to the starting sequence, indicating the near-optimal sequence of the primary binding site of the bv6 and bv8 variants (Fig. 4D).

The new unique variants bv21, bv22, bv23, and bv24 encoding only the primary binding site (Table S2) were expressed, purified, and biophysically characterized. All these variants were readily produced,

well-folded and highly thermostable (Fig. S8). Receptor-binding assays performed to compare the binding affinity of Boskar4, bv6, bv8, and the new variants (i.e. bv21-24), showed all the new variants to possess higher affinity. Specifically, all the new variants possessed a dissociation constant ($K_D$) within the range of 1 nM to 2 nM, compared with 14 ± 2 nM for bv6, 26 ± 3 nM for bv8, and 173 ± 43 nM for the original Boskar4. Notably, bv22 demonstrated the highest affinity among the variants with a $K_D$ of 1.1 ± 0.2 nM (Fig. 4E, S9 and Table S3). In their existing form (i.e. encoding only the primary binding site) the new variants already exhibited stronger G-CSFR inhibition in competitive inhibition assays. For instance, bv22 had an $IC_{50}$ of 35 nM, compared to 93 nM for bv6, and 2.3 µM for Boskar4. As observed previously, all tested designs, except for bop1, displayed residual activity at non-inhibiting concentrations (Fig. S9H).

Finally, in order to combine the potent inhibitory activity of bv22 and the eliminated residual activity of bop1 we created the bop3 design, which encodes the primary binding site of bv22 and the secondary binding site of bop1. The bop3 design was readily produced and purified, well-folded and hyper-thermostable (Fig. 5A, S10A, B). SPR titrations of bop3 showed it to bind G-CSFR with a $K_D$ of 1.2 ± 0.1 nM (Fig. 5B), which is comparable to that observed for bv22 ($K_D$ = 1.3 ± 0.1 nM; Fig. S10C). As expected, the proliferation assays in NFS-60 cells indicated no proliferative activity for bop3 across a broad range of concentrations (Fig. S10D). Competitive inhibition assays in the same cells showed bop3 to exhibit 3-fold improvement in inhibitory potency compared bop1 ($IC_{50}$ = 13.4 ± 16.9 nM; Fig. 5C and S10E). We hence further evaluated the potential of bop3 to inhibit a G-CSF-dependent acute myeloid leukemia (AML) cell line, Kasumi-1 [49]. The results of this competitive inhibition assay G-CSF showed bop3 exhibit an $IC_{50}$ of 1.97 ± 0.17 nM (Fig. 5D). Finally, we sought to repeat both the proliferation and competitive inhibition assays in the healthy donor-derived primary hematopoietic ($CD34^+$) stem cells, as a more realistic readout system. As expected, no discernible proliferative activity was observed following treatment with bop3 alone at concentrations of 0.5 µg/mL or 1 µg/mL (Fig. 5E). In contrast, a pronounced inhibition of HSPCs stimulated with 1 ng/mL rhG-CSF was apparent at the same concentrations of bop3 (Fig. 5F). A moderate inhibitory effect was noticeable even at the lower concentration of 100 ng/mL bop3 (Fig. S10F). An inert protein with a similar structural (moevan_control [50]) was also purified and tested under the same conditions, and did not show any measurable activity at the highest used concentration (i.e. 1 µg/mL; Fig. 5). Furthermore, bop3 was not toxic to any of the used cells, highlighting the specificity of its G-CSFR antagonism.

***Discussion***

Inhibiting a cytokine signaling pathway is more safely achieved through the blockade of the respective receptor than by quenching its soluble cytokine ligand. This is since receptor blockers can be retrofitted with homing fusions to achieve local targeting effects, and thus avoid systemic inhibition of cytokines, which are multirole regulators [51]. However, the challenge of blocking homodimeric cytokine receptors is two-fold. The first relates to the inhibitor's capacity to outcompete the binding avidity of the bivalent ligand [52]. This can be further complicated by oncogenic mutations enhancing the dimerizing interactions between the receptor subunits [53] or its associated kinases [54]. The other challenge pertains to avoiding any formation of dimeric (or higher oligomeric) states of the inhibitor in solution, which can in turn form assemblies with signaling-compatible geometries. Such inadvertent signaling can yield a mixed biological effect (i.e. activation and inhibition) *in vitro* or *in vivo*. Therefore, the tendency of antibodies and the parent cytokine (here G-CSF) to self-associate [55-58] precludes them providing suitable templates for inhibitor design.

In this study we aimed to overcome these challenges by starting from a *de novo* G-CSFR-binding domain (Boskar4) as a template. Boskar4 was designed to encode a single receptor-binding site, but was found to partition between monomeric and dimeric forms in solution [31]. Hence, our first step was

to graft a secondary G-CSFR-binding site onto the same domain in an antiparallel orientation (i.e. *boa* designs). This was done with the three-pronged aim of: i) enhancing its effective binding affinity to the receptor (through increased reassociation likelihood), ii) minimizing the receptor activation probability by imposing an unfavorable (strained) orientation for dimerized receptor subunits (Fig. 1A), and iii) disfavoring ligand dimerization in solution (at least in the presence of the receptor) by encoding receptor-binding information on a larger fraction of its solvent exposed surface area. We additionally sought to evaluate the effect of grafting a secondary binding site in a parallel orientation, which we expected to result in signaling 2:1 receptor:ligand complexes (i.e. *bop* designs). In either case, we also sought to graft both large and small binding surfaces to evaluate their effect on binding (Fig. 1B). As expected, all the bivalent designs showed stronger binding affinity compared with the starting template, and were homogeneously monomeric in solution. As expected, receptor activation was strongest with the parallel design with a large secondary binding site (bop2), and only minimal activation was induced by the antiparallel designs (boa1 or boa2). However, our findings showed the parallel design with the small secondary binding site (bop1) to have no measurable receptor-inducing activity, and the highest inhibitory potency. By inspecting the crystal structure of bop1, its structure matched the design accurately, and its model in a 2:1 complex with the receptor indicated no possible steric incompatibility. Therefore, we assume for bop1 the receptor:ligand interactions mediated through secondary binding site are insufficient to overcome the thermodynamic cost of associating a 2:1 complex. Hence, we decided to further improve bop1 by enhancing the affinity of its primary binding site through computational redesign and experimental screening. This has led us to identify an improved variant (bop3), which bound G-CSFR approximately 10-fold tighter and was 3-fold stronger in inhibiting G-CSF-induced proliferation in NFS-60 cells. Finally, our *in vitro* assays show bop3 to be potent and specific in inhibiting G-CSF activity in an acute myeloid leukemia cell line and in primary hematopoietic stem cells.

Our approach to deploy a *de novo*-designed protein to outcompete the native ligand precisely at its high-affinity binding site on the receptor surface bears several advantages. Among the advantages, is the guaranteed maximum epitope overlap between the inhibitor and the native ligand; ensuring effective inhibition while avoiding recourse to any epitope mapping procedure. Moreover, the biophysically superior features of the *de novo*-designed template gives room for substantial functional enhancement, where the latter often comes at the cost of the protein's stability [59]. In this study, we start from our previous design Boskar4, which is thermally and proteolytically stable, and has maintained its stability after substantial redesign. Particularly, the most potent inhibitor bop3, which behaves as a strict monomer in solution, and binds G-CSFR 100-fold tighter than Boskar4, still preserves the hyper-thermostability of the parent template (Fig. 5A and S10A, B). The high yield, solubility, and stability of bop3 greatly facilitate its modification, production and downstream processing, and obviates the need for cold chain handling.

Excessive G-CSFR signaling manifests in different pathologies depending on the type of the cells involved. Several hematopoietic malignancies of myeloid origin are strongly associated with the overexpression, or acquired or inherited mutations of G-CSFR [19, 60-64]. Our results on inhibiting G-CSF activity in a leukemia cell line and in HSPCs demonstrate the therapeutic potential of our designs against myeloid malignancies. Furthermore, besides promoting leukemia, G-CSFR signaling can be responsible for the progression and metastasis of several tumors, and was shown as a potential target for treating glial [65, 66], bladder [25], colon [21], ovary [22], and breast [23, 24] cancers. Recently, some prevalent inflammatory conditions have been also tightly associated with elevated local or systemic levels of G-CSF. These include elevated G-CSF in lung tissue in the case of chronic obstructive pulmonary disease [67] or in peripheral blood in the case of rheumatoid arthritis [68], where inhibition of G-CSFR was effective in reducing neutrophilic inflammation in animal models [69]. These

all represent potential therapeutic applications where our presented designs can be advantageous with regard to ease of production, stability, and local bioavailability in comparison to antibodies [70-72]. Our designed G-CSFR binders can also be used as homing or labelling tags, to increase the local targeting specificity of functional payloads to G-CSFR-expressing cells, without activating them. Given the stability of our designed helical bundles, and unlike immunoglobulin-based binders, their folding robustness provide for minimal folding interference with other fused proteins [73]. For instance, bop2 as a compact bivalent activator of G-CSFR can be deployed as a fusion with another cytokine (i.e. as a *fusokine* [74]) to achieve co-targeting and co-activation of two or more receptors. Conversely, a monovalent binder such as bop3 can be readily used, in fusion with a different monovalent receptor binder, as a synthetic or surrogate cytokine [75, 76] to induce novel receptor associations. These latter applications can greatly overcome the otherwise pleiotropic effects or systemic toxicities of cytokine therapies.

***Acknowledgements***

This project has received funding from the IMPRS (TU), the M. Schickedanz Kinderkrebsstiftung (ME, JS), the "German Universities Excellence Initiative" of the Tübingen University (JS), Deutsche Forschungsgemeinschaft (DFG; No 500215849; ME, JS). The authors would also like to thank Luis Ángel Fernández for providing the bacterial display system plasmid, and Regine Bernhard for technical assistance.

***Conflict of interests***

TU, JS, and ME are co-inventors on a pending patent application encompassing the G-CSFR inhibitors described in this article.

***References***

1. Donnelly, R.P., H.A. Young, and A.S. Rosenberg, *An overview of cytokines and cytokine antagonists as therapeutic agents.* Ann N Y Acad Sci, 2009. **1182**: p. 1-13.
2. Marzo, M., *et al.*, *Infliximab: A Review of Its Use in the Treatment of Inflammatory Bowel Disease.* Clinical Medicine Insights: Therapeutics, 2011. **3**: p. CMT.S1978.
3. Weber, J. and S.J. Keam, *Ustekinumab.* BioDrugs, 2009. **23**(1): p. 53-61.
4. Schreiber, S., *et al.*, *Therapeutic Interleukin-6 Trans-signaling Inhibition by Olamkicept (sgp130Fc) in Patients With Active Inflammatory Bowel Disease.* Gastroenterology, 2021. **160**(7): p. 2354-2366.e11.
5. Scott, L.J., *Etanercept: a review of its use in autoimmune inflammatory diseases.* Drugs, 2014. **74**(12): p. 1379-410.
6. Jones, G., *et al.*, *Comparison of tocilizumab monotherapy versus methotrexate monotherapy in patients with moderate to severe rheumatoid arthritis: the AMBITION study.* Annals of the Rheumatic Diseases, 2010. **69**(01): p. 88.
7. Raimondo, M.G., *et al.*, *Profile of sarilumab and its potential in the treatment of rheumatoid arthritis.* Drug Design, Development and Therapy, 2017. **11**(null): p. 1593-1603.
8. Aversa, G., *et al.*, *An interleukin 4 (IL-4) mutant protein inhibits both IL-4 or IL-13-induced human immunoglobulin G4 (IgG4) and IgE synthesis and B cell proliferation: support for a common component shared by IL-4 and IL-13 receptors.* J Exp Med, 1993. **178**(6): p. 2213-8.
9. Freeman, B.D. and T.G. Buchman, *Interleukin-1 receptor antagonist as therapy for inflammatory disorders.* Expert Opin Biol Ther, 2001. **1**(2): p. 301-8.
10. Kuhlman, B. and P. Bradley, *Advances in protein structure prediction and design.* Nature Reviews Molecular Cell Biology, 2019. **20**(11): p. 681-697.
11. ElGamacy, M., *Chapter Four - Accelerating therapeutic protein design*, in *Advances in Protein Chemistry and Structural Biology*, R. Donev, Editor. 2022, Academic Press. p. 85-118.
12. Park, S.D., *et al.*, *A review of granulocyte colony-stimulating factor receptor signaling and regulation with implications for cancer.* Frontiers in Oncology, 2022. **12**.

13. Dwivedi, P. and K.D. Greis, *Granulocyte colony-stimulating factor receptor signaling in severe congenital neutropenia, chronic neutrophilic leukemia, and related malignancies.* Experimental hematology, 2017. **46**: p. 9-20.
14. Bendall, L.J. and K.F. Bradstock, *G-CSF: From granulopoietic stimulant to bone marrow stem cell mobilizing agent.* Cytokine & growth factor reviews, 2014. **25**(4): p. 355-367.
15. Schneider, A., *et al.*, *The hematopoietic factor G-CSF is a neuronal ligand that counteracts programmed cell death and drives neurogenesis.* J Clin Invest, 2005. **115**(8): p. 2083-98.
16. Wallner, S., *et al.*, *The granulocyte-colony stimulating factor has a dual role in neuronal and vascular plasticity.* Frontiers in cell and developmental biology, 2015. **3**: p. 48.
17. Bussolino, F., *et al.*, *Granulocyte- and granulocyte-macrophage-colony stimulating factors induce human endothelial cells to migrate and proliferate.* Nature, 1989. **337**(6206): p. 471-3.
18. Harada, M., *et al.*, *G-CSF prevents cardiac remodeling after myocardial infarction by activating the Jak-Stat pathway in cardiomyocytes.* Nat Med, 2005. **11**(3): p. 305-11.
19. Liongue, C. and A.C. Ward, *Granulocyte colony-stimulating factor receptor mutations in myeloid malignancy.* Frontiers in oncology, 2014. **4**: p. 93.
20. Nara, N., *et al.*, *Granulocyte colony-stimulating factor-dependent growth of an acute myeloblastic leukemia cell line.* Japanese journal of cancer research, 1990. **81**(6-7): p. 625-631.
21. Morris, K., *et al.*, *G-CSF and G-CSFR are highly expressed in human gastric and colon cancers and promote carcinoma cell proliferation and migration.* British journal of cancer, 2014. **110**(5): p. 1211-1220.
22. Kumar, J., *et al.*, *Granulocyte colony-stimulating factor receptor signalling via Janus kinase 2/signal transducer and activator of transcription 3 in ovarian cancer.* British journal of cancer, 2014. **110**(1): p. 133-145.
23. Liu, L., *et al.*, *The role of granulocyte colony-stimulating factor in breast cancer development: a review.* Molecular medicine reports, 2020. **21**(5): p. 2019-2029.
24. Mouchemore, K.A., R.L. Anderson, and J.A. Hamilton, *Neutrophils, G-CSF and their contribution to breast cancer metastasis.* The FEBS journal, 2018. **285**(4): p. 665-679.
25. Tachibana, M., *et al.*, *Granulocyte colony-stimulating factor receptor expression on human transitional cell carcinoma of the bladder.* Br J Cancer, 1997. **75**(10): p. 1489-96.
26. Hamilton, J.A., A.D. Cook, and P.P. Tak, *Anti-colony-stimulating factor therapies for inflammatory and autoimmune diseases.* Nature reviews Drug discovery, 2017. **16**(1): p. 53-70.
27. Martin, K.R., *et al. G-CSF–A double edge sword in neutrophil mediated immunity*. in *Seminars in Immunology*. 2021. Elsevier.
28. Campbell, I.K., *et al., Therapeutic targeting of the G-CSF receptor reduces neutrophil trafficking and joint inflammation in antibody-mediated inflammatory arthritis.* The Journal of Immunology, 2016. **197**(11): p. 4392-4402.
29. Cornish, A.L., *et al., G-CSF and GM-CSF as therapeutic targets in rheumatoid arthritis.* Nature Reviews Rheumatology, 2009. **5**(10): p. 554-559.
30. Eyles, J.L., *et al., A key role for G-CSF–induced neutrophil production and trafficking during inflammatory arthritis.* Blood, The Journal of the American Society of Hematology, 2008. **112**(13): p. 5193-5201.
31. Skokowa, J., *et al.*, *A topological refactoring design strategy yields highly stable granulopoietic proteins.* Nature Communications, 2022. **13**(1): p. 2948.
32. Tamada, T., *et al.*, *Homodimeric cross-over structure of the human granulocyte colony-stimulating factor (GCSF) receptor signaling complex.* Proceedings of the national academy of sciences, 2006. **103**(9): p. 3135-3140.
33. Jumper, J., *et al., Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
34. Ullrich, T., *et al., Tuning of granulopoietic signaling by de novo designed agonists.* bioRxiv, 2023: p. 2023.11.25.568662.
35. Varadi, M., *et al.*, *AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models.* Nucleic acids research, 2022. **50**(D1): p. D439-D444.

36. Nakoinz, I., *et al.*, *Differentiation of the IL-3-dependent NFS-60 cell line and adaption to growth in macrophage colony-stimulating factor.* Journal of immunology (Baltimore, Md.: 1950), 1990. **145**(3): p. 860-864.
37. Weinstein, Y., *et al.*, *Truncation of the c-myb gene by a retroviral integration in an interleukin 3-dependent myeloid leukemia cell line.* Proceedings of the National Academy of Sciences, 1986. **83**(14): p. 5010-5014.
38. Virtanen, P., *et al.*, *SciPy 1.0: fundamental algorithms for scientific computing in Python.* Nature Methods, 2020. **17**(3): p. 261-272.
39. Kawakami, M., *et al.*, *Levels of Serum Granulocyte Colony-Stimulating Factor in Patients With Infections.* Blood, 1990. **76**(10): p. 1962-1964.
40. Kabsch, W., *xds.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(2): p. 125-132.
41. Vagin, A. and A. Teplyakov, *Molecular replacement with MOLREP.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 22-5.
42. Murshudov, G.N., *et al.*, *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallographica Section D: Biological Crystallography, 2011. **67**(4): p. 355-367.
43. Emsley, P., *et al.*, *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
44. Šali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* Journal of molecular biology, 1993. **234**(3): p. 779-815.
45. Maksymenko, K., *et al.*, *The design of functional proteins using tensorized energy calculations.* Cell Reports Methods, 2023. **3**(8): p. 100560.
46. Salema, V., *et al.*, *Selection of single domain antibodies from immune libraries displayed on the surface of E. coli cells with two β-domains of opposite topologies.* PloS one, 2013. **8**(9): p. e75126.
47. Tu, Q., *et al.*, *Room temperature electrocompetent bacterial cells improve DNA transformation and recombineering efficiency.* Scientific reports, 2016. **6**(1): p. 1-8.
48. Reguant, R., *et al.*, *AlignmentViewer: sequence analysis of large protein families.* F1000Research, 2020. **9**.
49. Asou, H., *et al.*, *Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8; 21 chromosome translocation.* Blood, 1991. **77**(9): p. 2031-2036.
50. Hernandez Alvarez, B., *et al.*, *Design of novel granulopoietic proteins by topological rescaffolding.* PLOS Biology, 2020. **18**(12): p. e3000919.
51. Turner, M.D., *et al.*, *Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2014. **1843**(11): p. 2563-2582.
52. Vauquelin, G. and S.J. Charlton, *Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands.* Br J Pharmacol, 2013. **168**(8): p. 1771-85.
53. Maxson, J.E., *et al.*, *Ligand independence of the T618I mutation in the colony-stimulating factor 3 receptor (CSF3R) protein results from loss of O-linked glycosylation and increased receptor dimerization.* J Biol Chem, 2014. **289**(9): p. 5820-7.
54. Wilmes, S., *et al.*, *Mechanism of homodimeric cytokine receptor activation and dysregulation by oncogenic mutations.* Science, 2020. **367**(6478): p. 643-652.
55. Makowski, E.K., *et al.*, *Optimization of therapeutic antibodies for reduced self-association and non-specific binding via interpretable machine learning.* Nature Biomedical Engineering, 2023.
56. Raso, S.W., *et al.*, *Aggregation of granulocyte-colony stimulating factor in vitro involves a conformationally altered monomeric state.* Protein Sci, 2005. **14**(9): p. 2246-57.
57. Wood, V.E., *et al.*, *Protein Engineering and HDX Identify Structural Regions of G-CSF Critical to Its Stability and Aggregation.* Molecular Pharmaceutics, 2022. **19**(2): p. 616-629.
58. Chi, E.Y., *et al.*, *Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor.* Protein Science, 2003. **12**(5): p. 903-913.

59. Teufl, M., C.U. Zajc, and M.W. Traxlmayr, *Engineering Strategies to Overcome the Stability–Function Trade-Off in Proteins.* ACS Synthetic Biology, 2022. **11**(3): p. 1030-1039.
60. Skokowa, J., *et al.*, *Interactions among HCLS1, HAX1 and LEF-1 proteins are essential for G-CSF–triggered granulopoiesis.* Nature Medicine, 2012. **18**(10): p. 1550-1559.
61. Eisenwort, G., *et al.*, *Phenotypic characterization of leukemia-initiating stem cells in chronic myelomonocytic leukemia.* Leukemia, 2021. **35**(11): p. 3176-3187.
62. Vellenga, E., *et al.*, *The effects of GM-CSF and G-CSF in promoting growth of clonogenic cells in acute myeloblastic leukemia.* Blood, 1987. **69**(6): p. 1771–1776.
63. Griffin, J.D. and B. Löwenberg, *Clonogenic cells in acute myeloblastic leukemia.* Blood, 1986. **68**(6): p. 1185-95.
64. Young, D.C., *et al.*, *In vitro expression of colony-stimulating factor genes by human acute myeloblastic leukemia cells.* Exp Hematol, 1988. **16**(5): p. 378-82.
65. Alghamri, M.S., *et al.*, *G-CSF secreted by mutant IDH1 glioma stem cells abolishes myeloid cell immunosuppression and enhances the efficacy of immunotherapy.* Sci Adv, 2021. **7**(40): p. eabh3243.
66. Wang, J., *et al.*, *Granulocyte-colony stimulating factor promotes proliferation, migration and invasion in glioma cells.* Cancer Biol Ther, 2012. **13**(6): p. 389-400.
67. Tsantikos, E., *et al.*, *Granulocyte-CSF links destructive inflammation and comorbidities in obstructive lung disease.* J Clin Invest, 2018. **128**(6): p. 2406-2418.
68. Nakamura, H., *et al.*, *High serum and synovial fluid granulocyte colony stimulating factor (G-CSF) concentrations in patients with rheumatoid arthritis.* Clin Exp Rheumatol, 2000. **18**(6): p. 713-8.
69. Scalzo-Inguanti, K., *et al.*, *A neutralizing anti–G-CSFR antibody blocks G-CSF–induced neutrophilia without inducing neutropenia in nonhuman primates.* Journal of leukocyte biology, 2017. **102**(2): p. 537-549.
70. Simeon, R. and Z. Chen, *In vitro-engineered non-antibody protein therapeutics.* Protein & cell, 2018. **9**(1): p. 3-14.
71. Chauhan, V.P., *et al.*, *Delivery of molecular and nanoscale medicine to tumors: transport barriers and strategies.* Annu Rev Chem Biomol Eng, 2011. **2**(1): p. 281-298.
72. Shah, D.K. and A.M. Betts. *Antibody biodistribution coefficients: inferring tissue concentrations of monoclonal antibodies based on the plasma concentrations in several preclinical species and human*. in *MAbs*. 2013. Taylor & Francis.
73. ElGamacy, M. and B. Hernandez Alvarez, *Expanding the versatility of natural and de novo designed coiled coils and helical bundles.* Current Opinion in Structural Biology, 2021. **68**: p. 224-234.
74. Ng, S. and J. Galipeau, *Concise review: engineering the fusion of cytokines for the modulation of immune cellular responses in cancer and autoimmune disorders.* Stem Cells Transl Med, 2015. **4**(1): p. 66-73.
75. Moraga, I., *et al.*, *Synthekines are surrogate cytokine and growth factor agonists that compel signaling through non-natural receptor dimers.* eLife, 2017. **6**: p. e22882.
76. Yen, M., *et al.*, *Facile discovery of surrogate cytokine agonists.* Cell, 2022. **185**(8): p. 1414-1430.e19.
77. Young, D.C., *et al.*, *Characterization of the receptor binding determinants of granulocyte colony stimulating factor.* Protein Science, 1997. **6**(6): p. 1228-1236.
78. Skokowa, J., *et al.*, *A topological refactoring design strategy yields highly stable granulopoietic proteins.* Nature communications, 2022. **13**(1): p. 1-17.
79. Ullrich, T., *et al.*, *Tuning of granulopoietic signaling by de novo designed agonists.* Preprint, 2023, https://doi.org/10.1101/2023.11.25.568662.

***Figures***

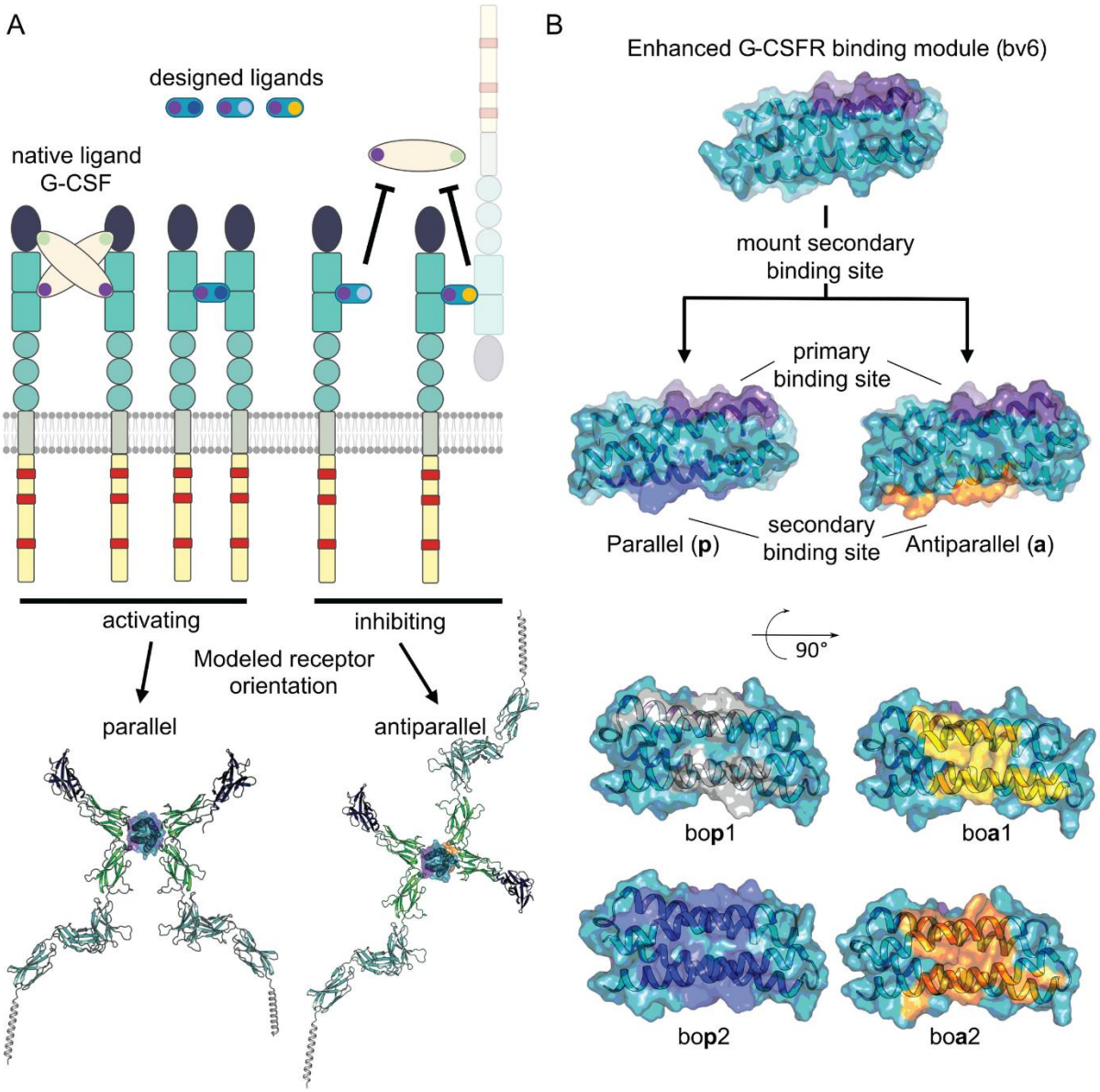



**Figure 1. Design strategy of bifaceted *de novo* ligands of G-CSFR. (A)** G-CSF associates its receptor through two binding sites, illustrated as purple and green dots, that bind the receptor's CRH and Ig-like domains, respectively, resulting in a 2:2 complex. The bifaceted ligands on the other hand encode two (primary and secondary) CRH-binding sites, where the relative orientation and affinity of the secondary binding site defines the biological outcome that can range between activation to inhibition. Under this scheme, the primary binding site (purple dot) has a strong affinity to the receptor and serves as the anchoring point of the ligand. A secondary binding site with sufficient affinity to dimerize the receptor in a parallel orientation (blue dot) in a 2:1 assembly with proximal intracellular segments is conducive of intracellular signaling. Alternatively, signaling inhibitors can be achieved through an antiparallel orientation of the secondary binding site (orange dot) that can result in a strained, unlikely receptor assembly that does not support signaling or stable receptor dimerization. Additionally, inhibitors can also be obtained in cases where the secondary binding site has insufficient affinity to dimerize the receptors, resulting in a 1:1 receptor:ligand complex. **(B)** Side and front views of the secondary binding site grafts with parallel (bop1 and bop2; gray and blue patches) or anti-parallel (boa1 and boa2; yellow and orange patches) orientation to the primary binding site (purple patch). These grafts comprised either a small (bop1 and boa1; gray and yellow patches) or a large (bop2 and boa2; blue and orange patches) number of residues.

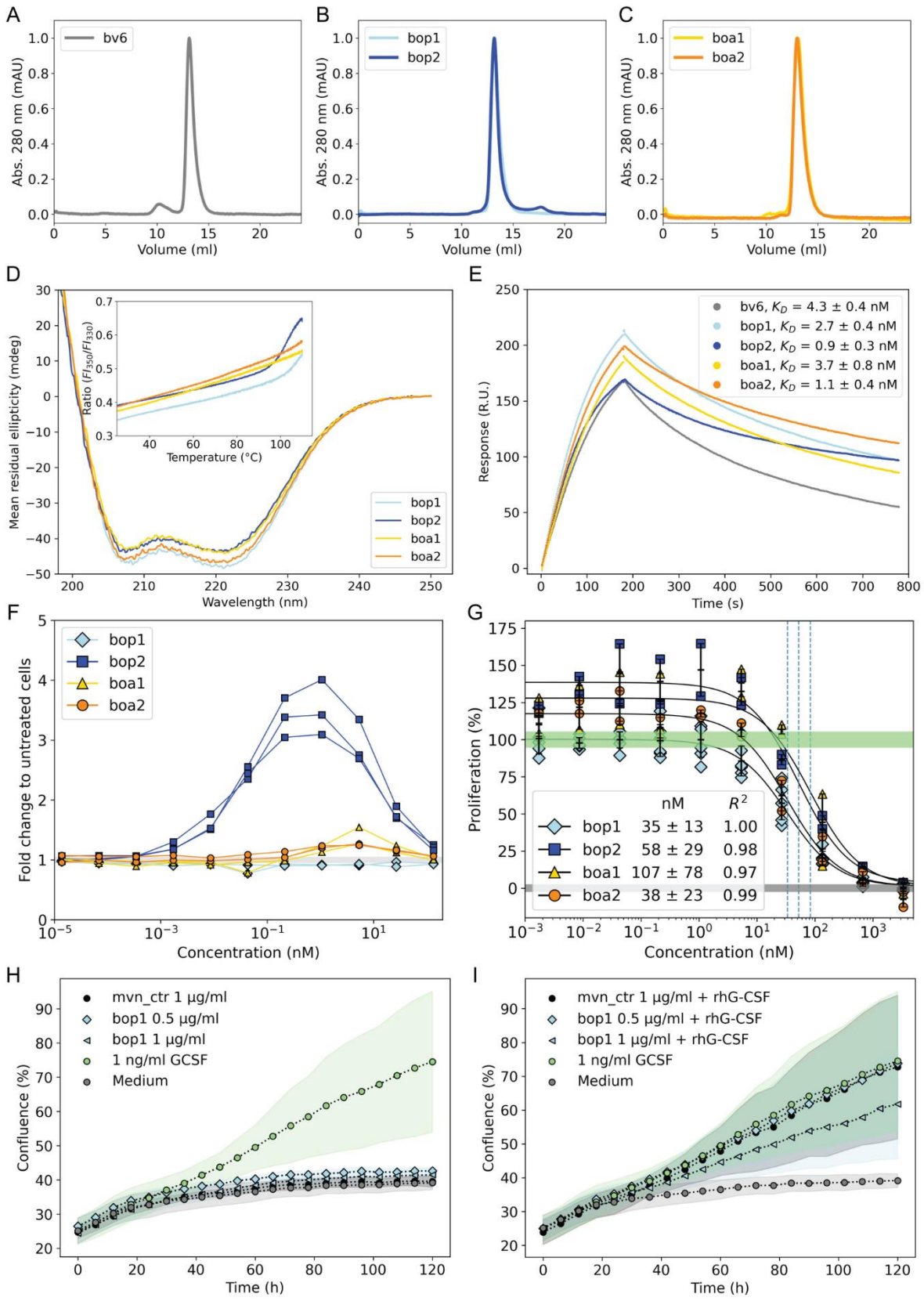



**Figure 2. Characterization of the first generation of bifaceted designs leads to identifying *bop1* as an inhibitor with no residual activity. (A, B)** Size exclusion chromatography shows the bifaceted designs to be exclusively monomeric proteins. **(C)** In contrast, the design template bv6 (an enhanced-affinity Boskar4 variant [34]) shows to be minorly dimeric. **(D)** All the designs have helical CD spectra, and are highly thermostable (inset; nanoDSF melting curves), where the bop designs are less stable (i.e. melting onset at 100 °C) in comparison to the boa designs. NanoDSF plots show the mean of ratio of the

fluorescence intensity at 350 nm and 330 nm from 4 technical replicates. **(E)** SPR sensograms showed that all the designs with the added secondary binding site bound rhG-CSFR with higher affinity in comparison to the starting template, bv6. Particularly, designs with larger binding interfaces (bop2 and boa2) bound 4-fold tighter than bv6. The binding affinity ($K_D$) values were obtained from fitting 2-fold dilution series performed under the same conditions (Fig. S4, Table S3). **(F)** The proliferative activity (fold-change to untreated cells) of the designs in NFS-60 cells indicated bop2 to be the most active, whereas boa1 and boa2 had low residual activity, and bop1 had no detectable residual activity across the entire concentration range. The proliferation results were obtained from at least two independent biological replicates. **(G)** The bop1 design was the most potent inhibiting G-CSF-induced proliferation in competitive inhibition assays of NFS-60 cells. Unlike other designs, bop1 had no residual activity at sub-inhibitory concentrations. A constant concentration of rhG-CSF (50 pg/mL; 100% proliferation) was used in all conditions, where the green and gray shades cover the mean and standard deviation of normalized cell proliferation with and without rhG-CSF, respectively. At least two biological replicas were performed for each design, where half-maximal inhibitory concentrations ($IC_{50}$) values were obtained from the corresponding fit. **(H, I)** HSCP proliferation assays using 0.5 µg/ml and 1 µg/mL of bop1 where performed on healthy donors' CD34+ HSPCs either without **(H)** or with the addition of 1 ng/mL rhG-CSF **(I)**. As a negative control, 1 µg/mL of Moevan_control (mvn_ctrl,[50]) was tested, which is a protein lacking a functional G-CSFR binding site. Shown is mean and standard deviation of two independent replicates of three technical replicates each.

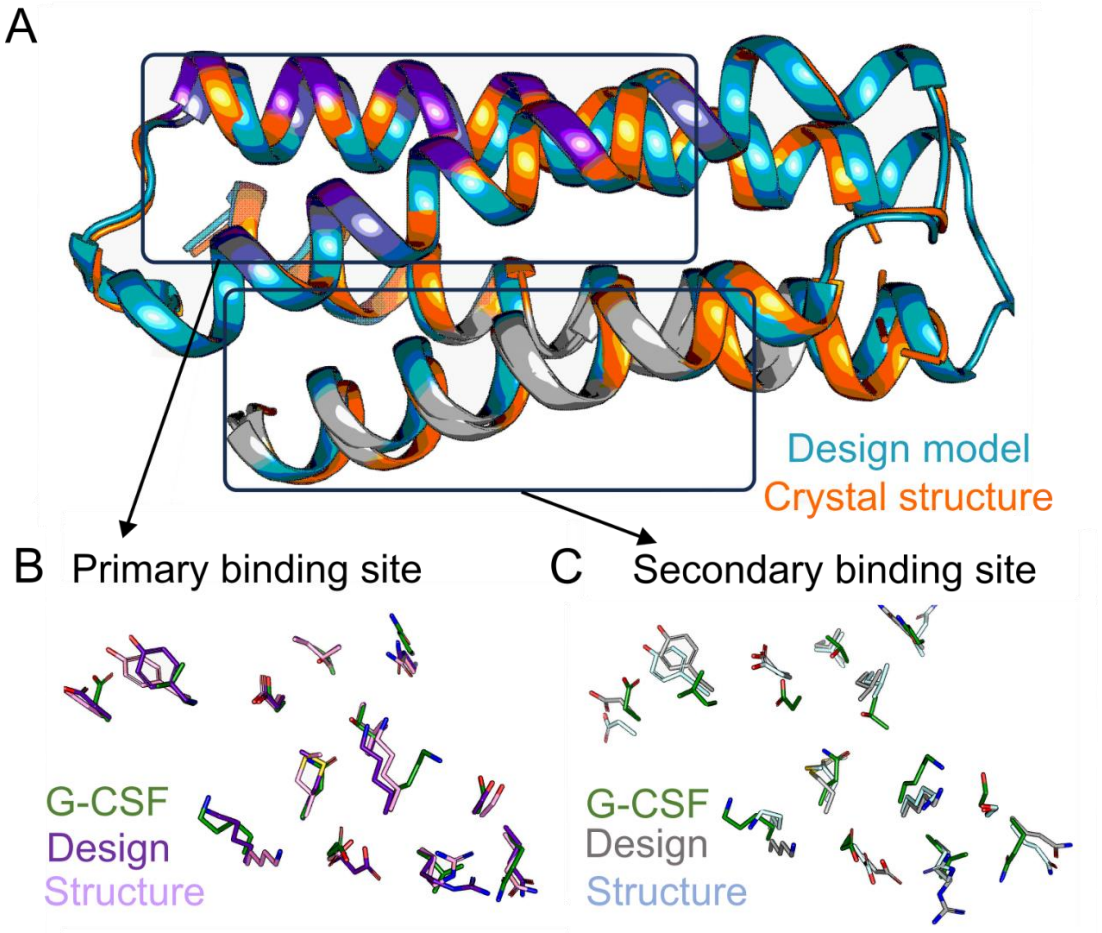



**Figure 3. Crystal structure of bop1 matches its design model at atomic resolution. (A**) Alignment of the crystal structure of bop1 (orange) with the design model (teal) shows a global backbone RMSD of 1.03 Å. Primary and secondary binding site helices are colored in purple and gray, respectively. **(B, C)** Backbone RMSD of the binding sites show the primary binding site of the design model and the crystal structure to align better to the corresponding positions in the G-CSF structure (0.6 Å; PDB: 2D9Q) than the secondary binding site (1.2 Å).

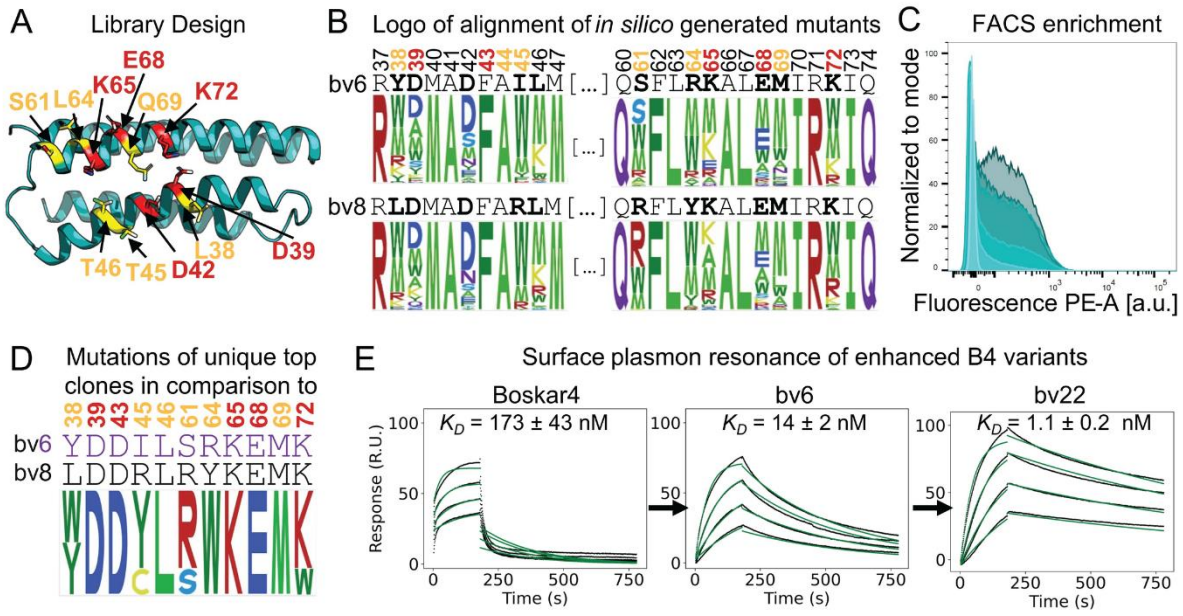



**Figure 4. Computationally-guided affinity enhancement of the primary binding site. (A)** Position on the Boskar4 structure subjected to optimization in the modelled complex with G-CSFR. In total 11 residues were optimized (orange or red). These included five positions identified to be critical for G-CSFR binding based on an alanine-scanning study of G-CSF (red, [77]). **(B)** Comparison between the mutants generated by the Damietta combinatorial sampler [45] in a logo that contains the alignment of the mutated sites. **(C)** High-throughput screening of the bacterial display library resulted in a strong FACS-enrichment of better binding clones. **(D)** Sequence logo of the top 10 selected clones. **(E)** SPR measurements of G-CSFR binding to Boskar4 (Original design), bv6 (first-generation variant), and bv22 (second-generation variant) demonstrate the enhancement in binding affinity over the first and second round of computationally-guided affinity maturation (Fig. S9 and Table S3).

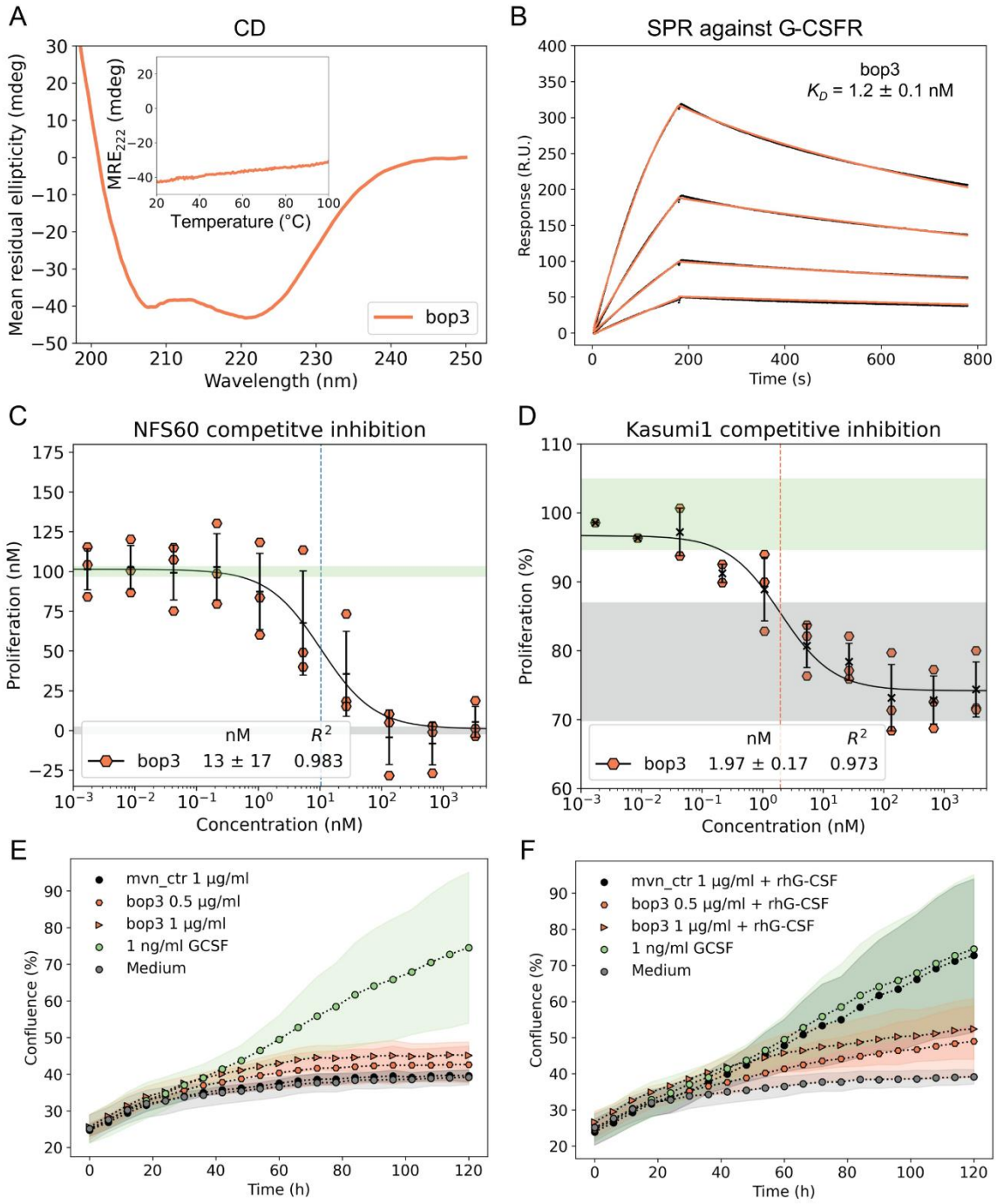



**Figure 5. Characterization of the functionally-enhanced inhibitor bop3. (A)** Circular dichroism spectrum and melting curve confirm the helical and hyper-thermostable properties of bop3. This thermal stability was also confirmed by nanoDSF (Fig. S10B). **(B)** SPR measurements also confirm a 10-fold improvement in the receptor-binding affinity. Shown is a twofold dilution series starting from 20 nM of bop3 against chip-immobilized G-CSFR. **(C)** Competitive inhibition cell assays in NFS-60 cells show bop3 to be approximately 3-fold more potent in inhibiting G-CSF-induced proliferation than bop1, with no observed residual activity (Fig. S10D). **(D)** Competitive inhibition of G-CSF-induced proliferation in the Kasumi-1 AML cell line using bop3. In both NFS-60 (C) Kasumi-1 cells (D) three independent experiments each, and the half-maximal inhibitory concentrations ($IC_{50}$ in nM) were obtained from a corresponding fit. In both assays, a constant concentration of rhG-CSF was employed, with 50 pg/mL in (C) and 400 pg/mL in (D), each corresponding to 100% proliferation. **(E, F)** CD34+ HSPCs proliferation assays using 0.5 µg/ml or 1 µg/ml of bop3 either without (E) or with the addition of 1 ng/ml rhG-CSF (F). Moevan_control (mvn_ctrl) was used as negative control. Shown is mean and standard deviation of two independent replicates, with three technical replicates per each independent replicate.

# Supplementary material

# *De novo* design of G-CSFR-inhibiting proteins

Timo Ullrich[1], Olga Klimenkova[2], Marcus D. Hartmann[1,3], Julia Skokowa[2], Mohammad ElGamacy[1,2*]

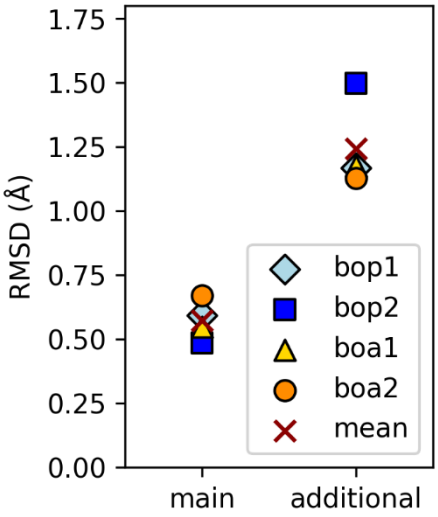
[1] Max-Planck-Institute of Biology, Department Protein Evolution, Tübingen
[2] Translational oncology, Internal Medicine II, University Hospital Tübingen
[3] Interfaculty Institute of Biochemistry, University of Tübingen, Germany
* corresponding author: mohammad.elgamacy@med.uni-tuebingen.de

**Supplementary figures 1-10.**

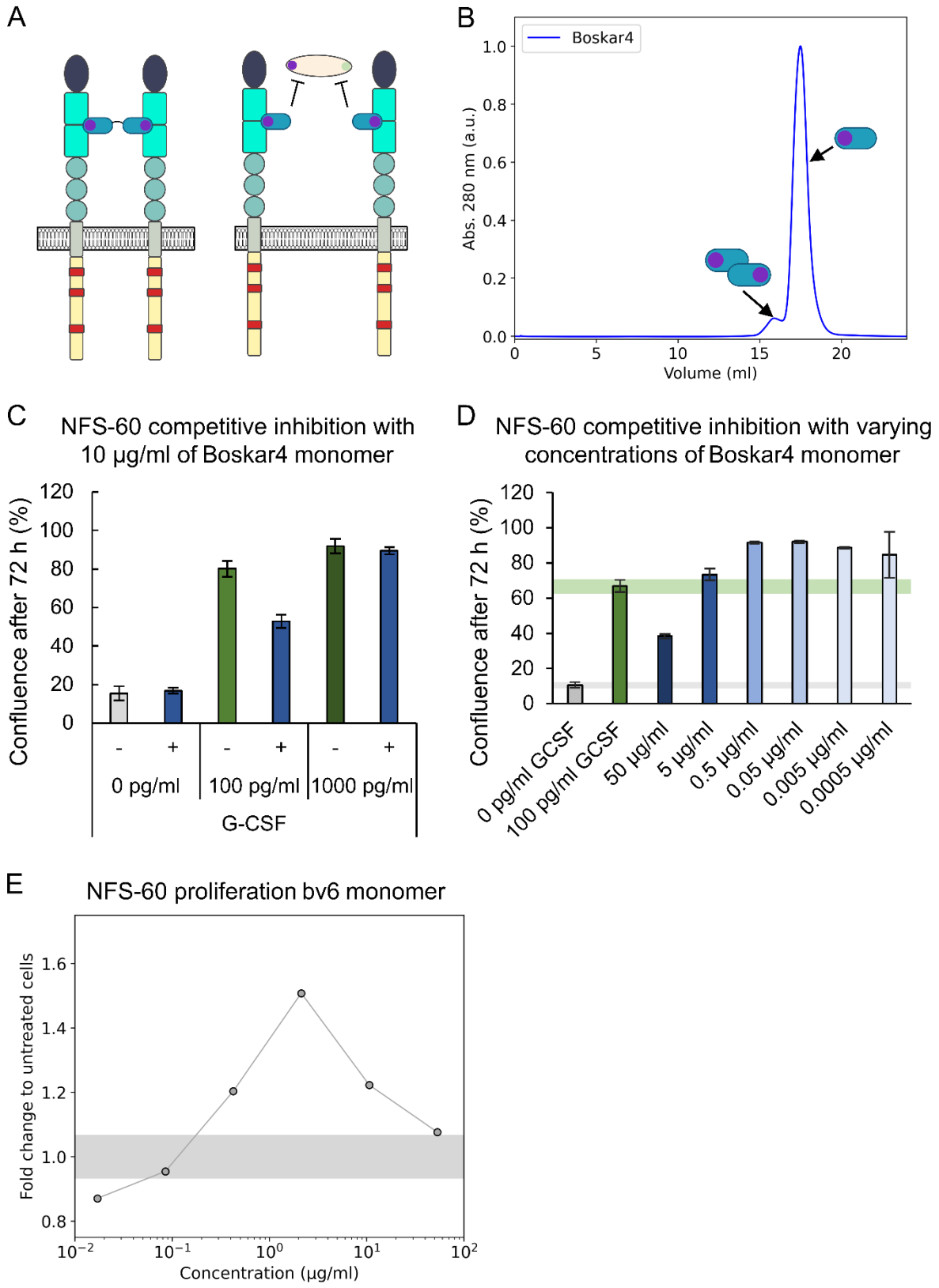
**Supplementary tables 1-4.**

**Supplementary method.**

***Supplementary Figures***

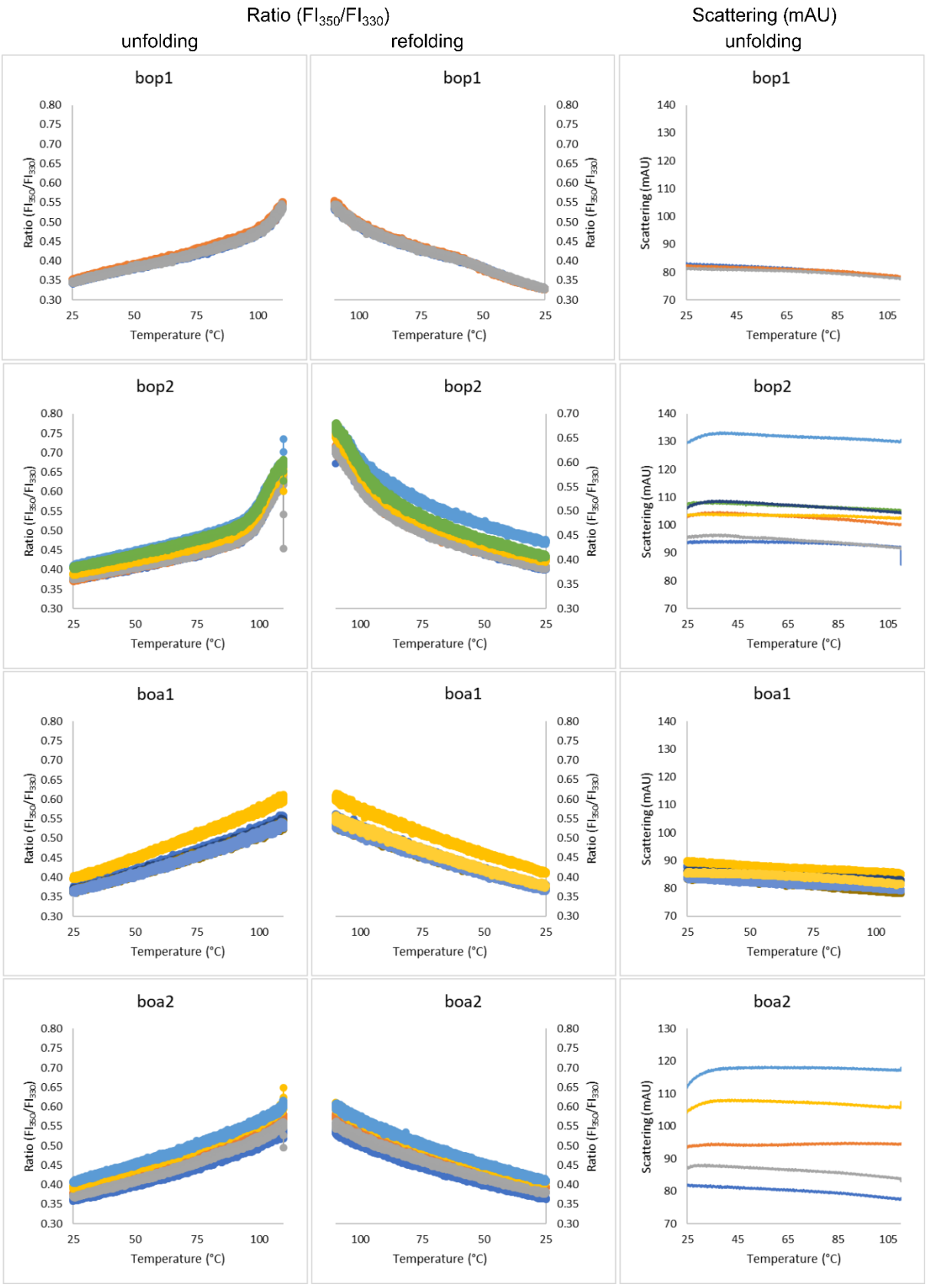



**Supplementary Figure 1. The G-CSFR-binding site in all design models aligned more poorly to the secondary binding site compared to the primary binding site.** AlphaFold2 models of the designed G-CSFR binders bop1, bop2, boa1 and boa2. The corresponding residues of the primary binding site and secondary binding site were structurally aligned to the corresponding residues of G-CSF (from PBD:2D9Q) to determine their backbone atoms RMSD.



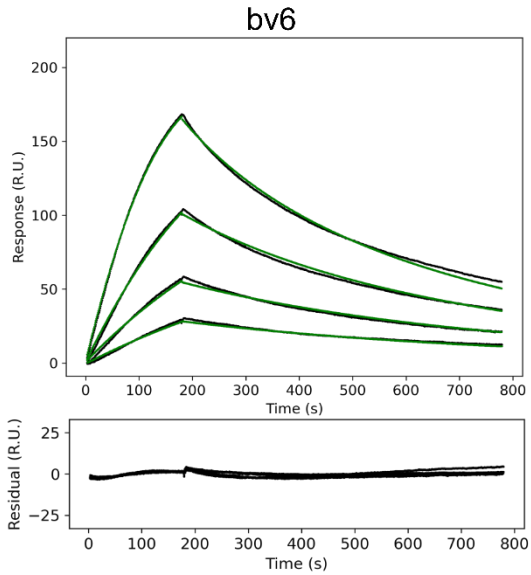

**Supplementary Figure 2. Oligomeric state of Boskar4 in solution dictates its effect on either activating or inhibiting G-CSFR. (A)** The illustration demonstrates how Boskar4 in a two-copy tandem connected by a short linker (Boskar4_st2) can dimerize the receptor subunits (left), while the monomeric fraction can inhibit G-CSF-induced receptor dimerization (right). **(B)** Analytical SEC shows the Boskar4 design to partition between a dimer and monomer in solution. Chromatography was

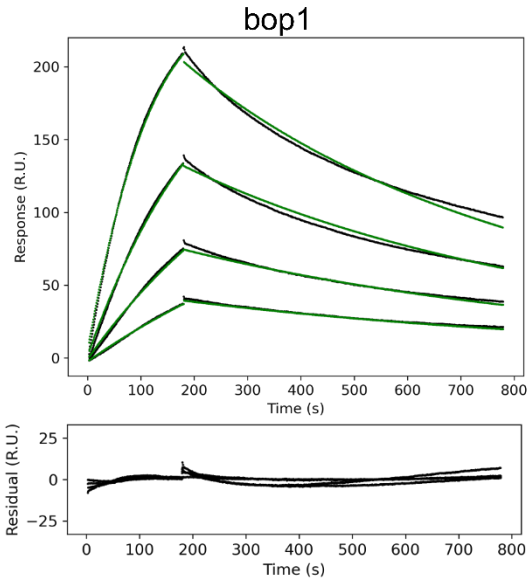
performed using a Superdex® 200 Increase 10/300 GL analytical column, where the data was obtained from [1]. **(C)** The monomeric fraction of Boskar4 could to some extent inhibit G-CSF-induced proliferation of NFS-60 cells. The figure shows the results of testing 10 µg/mL Boskar4 in a cell-based (NFS-60) competitive inhibition assay against 0 pg/mL, 100 pg/mL and 1000 pg/mL of G-CSF, where Boskar4 could outcompete 100 pg/mL G-CSF as indicated by an observed proliferation reduction. **(D)** Varying concentrations of Boskar4 titrated against 100 pg/mL G-CSF in the same cell assay explained also showed a clear reduction in proliferation at high concentration of monomeric Boskar4 (50 µg/mL). However below non-inhibiting concentrations, an increase in proliferation was clear. Confluence values shown in (C, D) where obtained using the IncuCyte S3 Live-Cell Analysis System, where the bar plots show mean and standard deviation of three technical replicates. **(E)** Proliferative activity for non-inhibiting concentrations were also observed in NFS-60 proliferation assays for the affinity enhanced Boskar4 variant bv6 (dark gray circles). Shown is the fold change to untreated cells. The proliferation level of untreated cells is indicated by a gray box which represents mean and standard deviation of twelve parallel replicates.

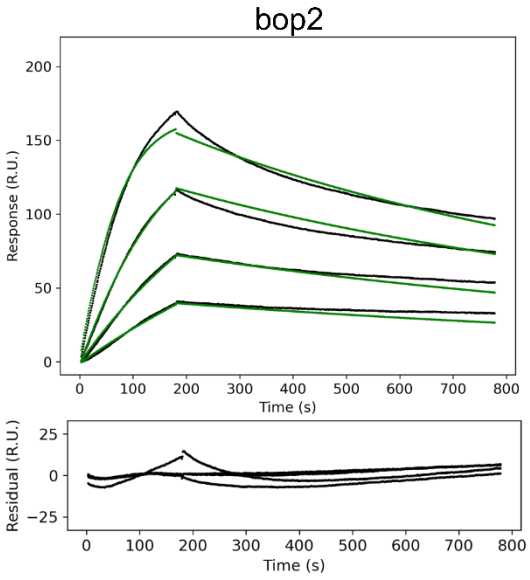


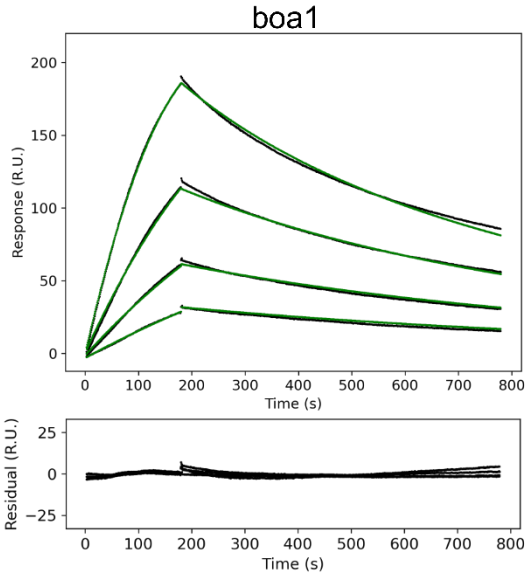

**Supplementary Figure 3. NanoDSF shows the bifaceted designs to be highly thermostable.** Heating (left column) and cooling (middle column) show no unfolding or refolding transitions for all four designs (bop1, bop2, boa1, and boa2) within the temperature range from 25 °C to 110 °C. Ratio values indicate relative fluorescence intensity at 350 nm and 330 nm. Scattering signal during the heating phase

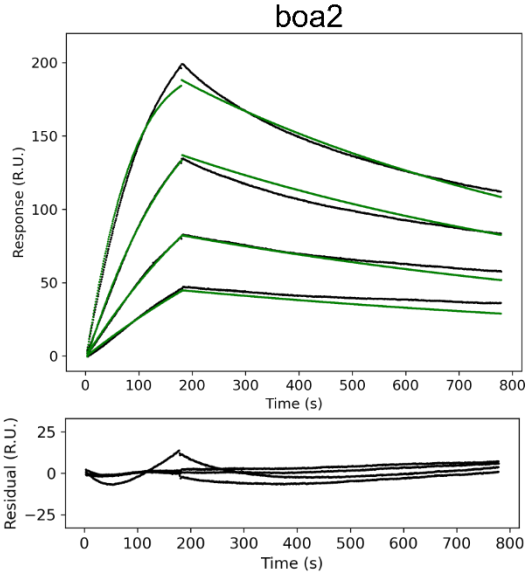
(right column) also indicate no aggregation of the designs. At least four (at most seven) technical replicates are shown for each design.


bv6
Response (R.U.)
Time (s)
Residual (R.U.)
Time (s)

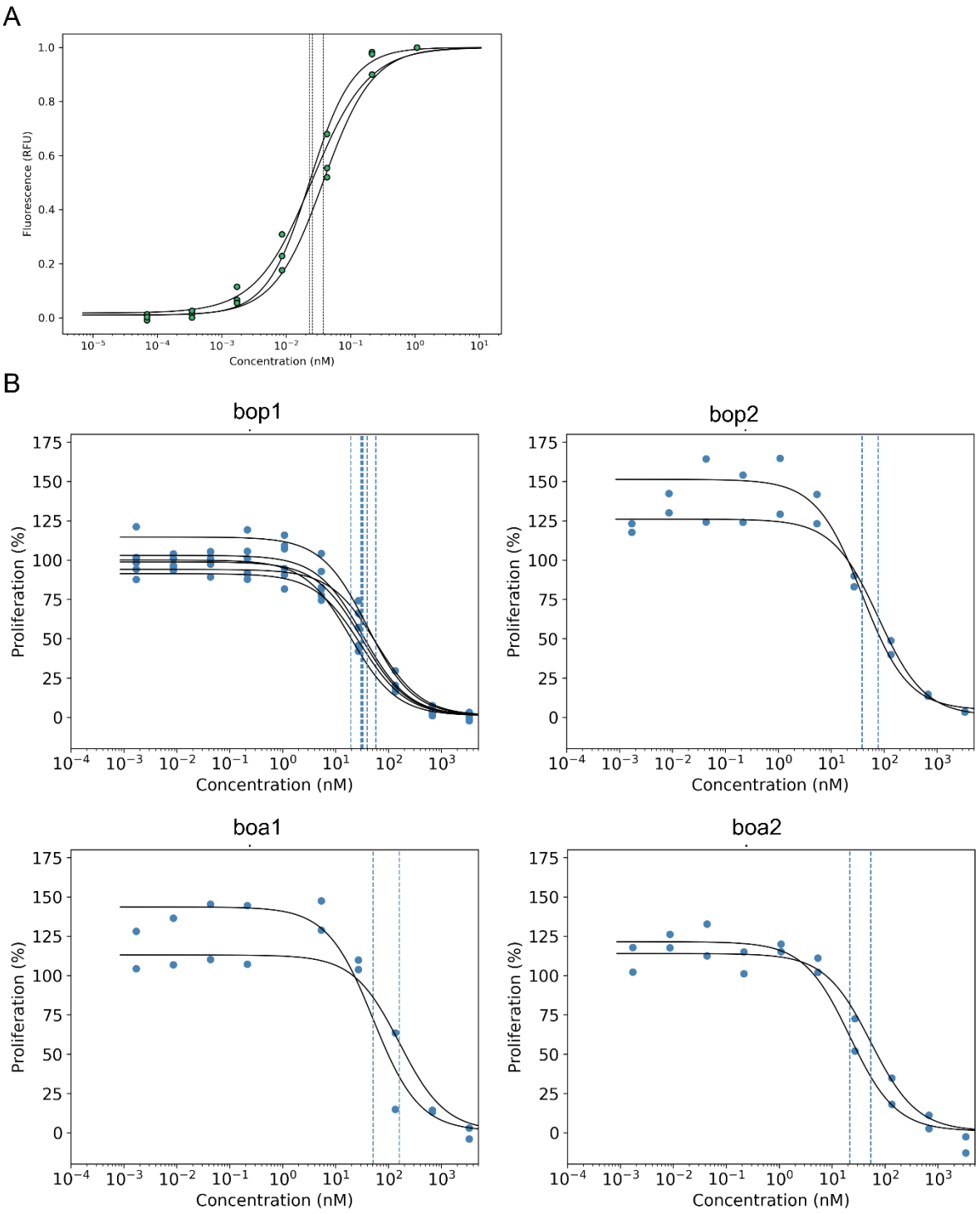

bop1
Response (R.U.)
Time (s)
Residual (R.U.)
Time (s)


bop2
Response (R.U.)
Time (s)
Residual (R.U.)
Time (s)

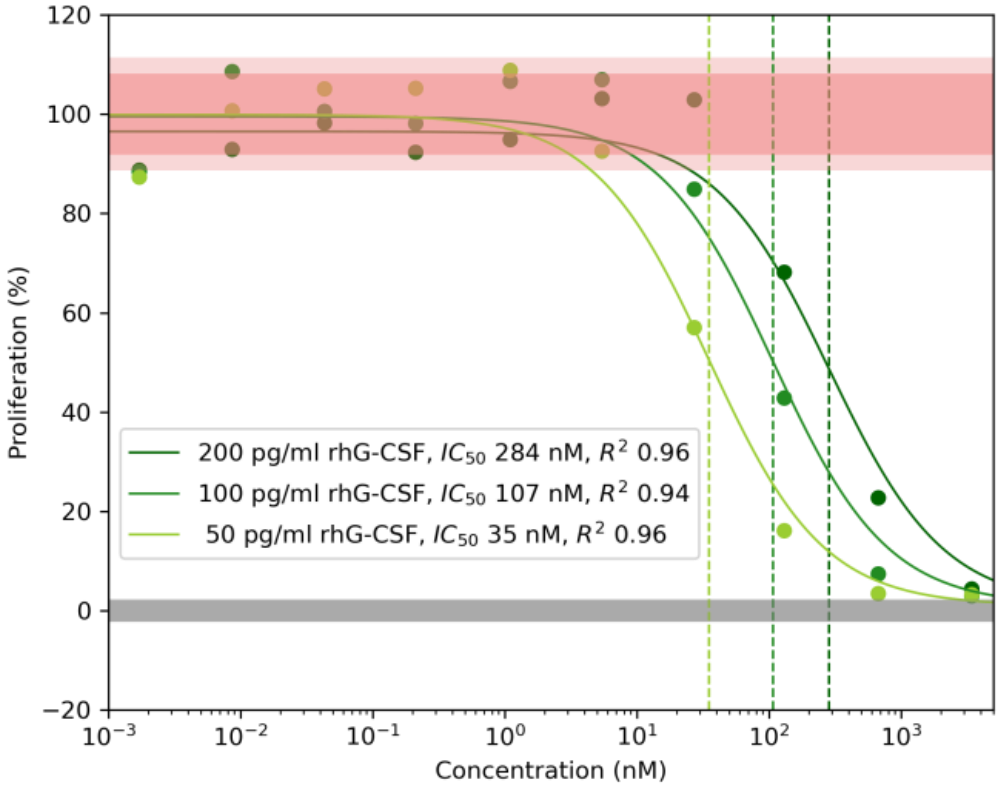

boa1
Response (R.U.)
Time (s)
Residual (R.U.)
Time (s)


boa2
Response (R.U.)
Time (s)
Residual (R.U.)
Time (s)

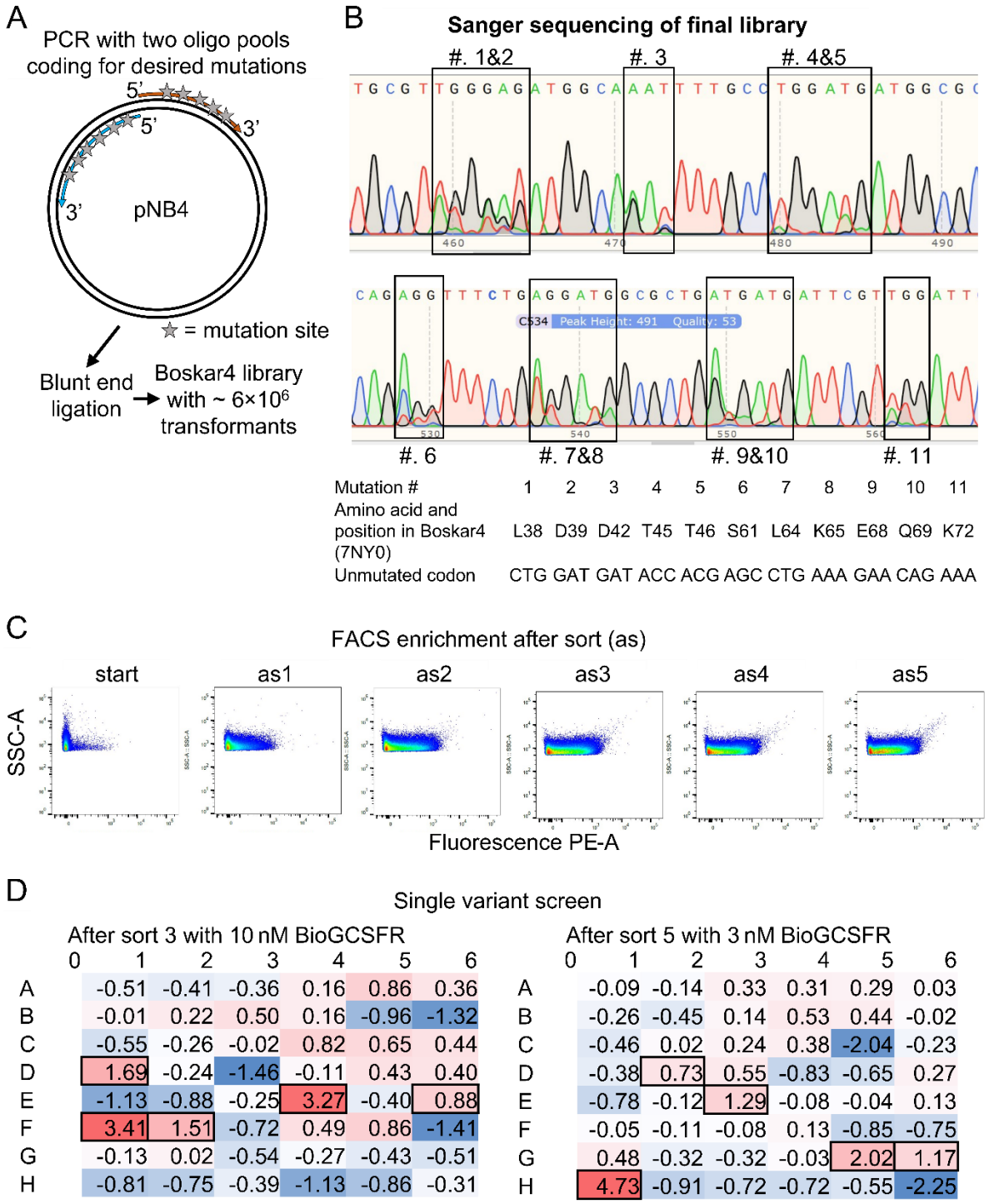
**Supplementary Figure 4. SPR titrations of the bifaceted designs (bop/boa) compared to the starting template (bv6).** SPR sensograms (black line) and the corresponding fit (green line) of a two-fold dilution series starting at 20 nM of bv6, bop1, bop2, boa1, and boa2. Binding kinetic parameters that were calculated from the displayed fits (Table S3; indicated in blue).



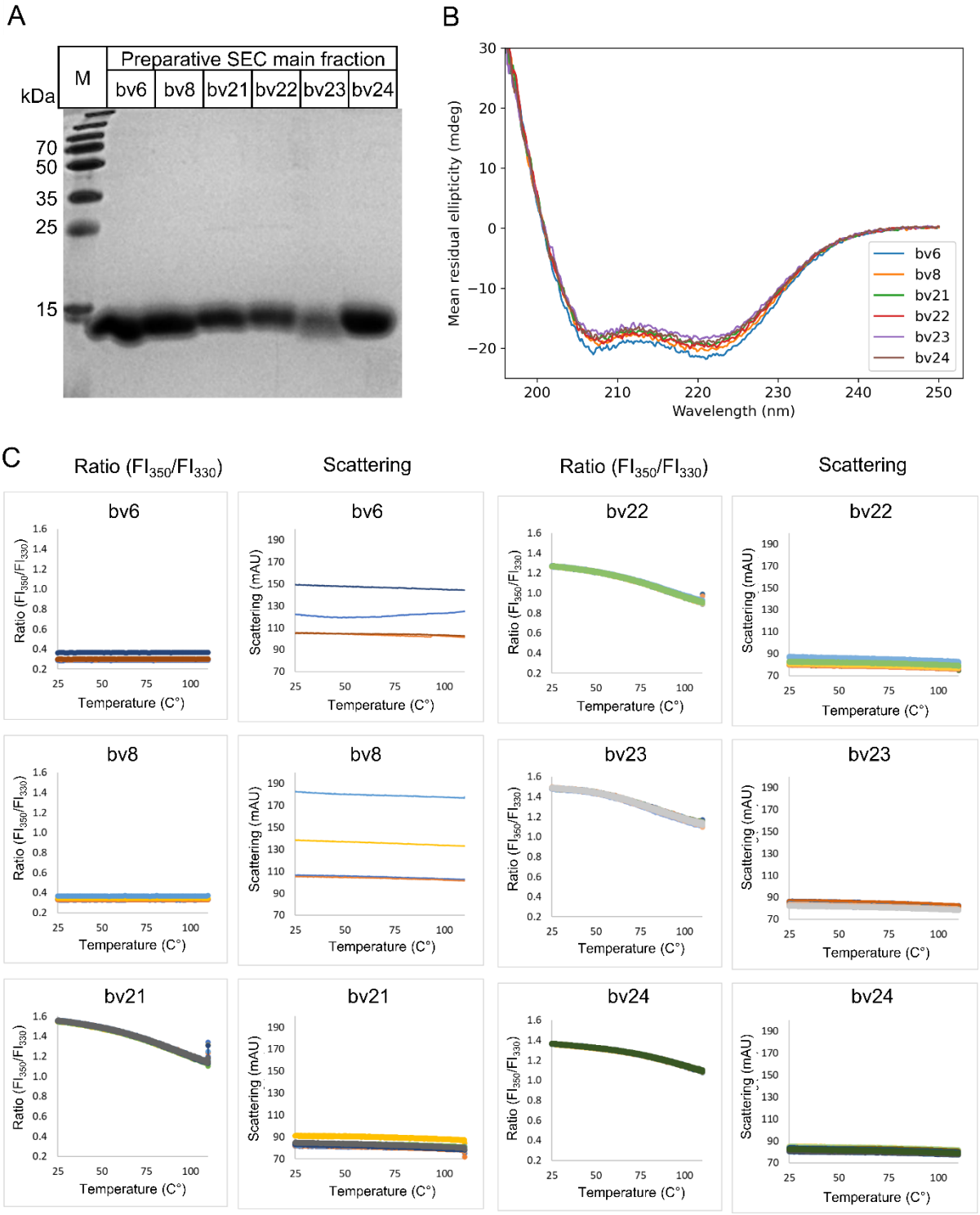

**Supplementary figure 5. The bop1 design is the most potently inhibits G-CSF activity with no residual receptor activation. (A)** The half-maximal effective concentration ($EC_{50}$) of the bop2 was determined for non-inhibiting concentrations observed in NFS-60 activity assays (cf. Fig. 2F). The fits of three independent experiments are shown. Mean and standard deviation values are provided in Table S3. **(B)** The half-maximal inhibitory concentrations ($IC_{50}$, indicated by dotted lines) were obtained from the corresponding competitive inhibition NFS-60 assay (cf. Fig. 2G) by fitting the individual independent replicates. Mean and standard deviation of the obtained $IC_{50}$ values is presented in Table S3.



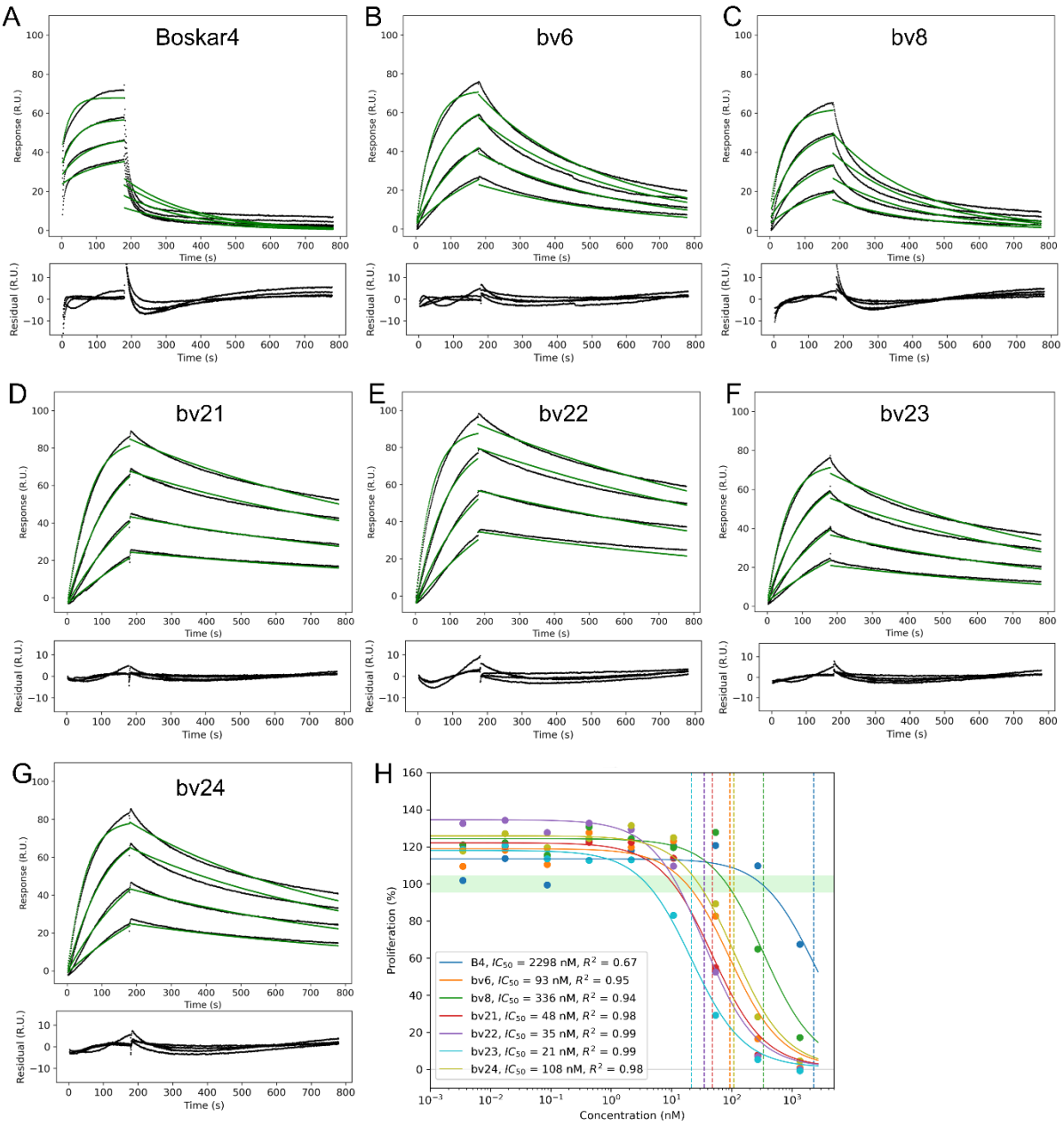

**Supplementary Figure 6. Bop1 outcompetes G-CSF in a dose-dependent manner.** Competitive inhibition assays of NFS-60 proliferation using varying concentrations of bop1 against three concentrations of of G-CSF (200, 100, 50 pg/mL). The half-maximal inhibitory concentrations ($IC_{50}$) were obtained from the corresponding fit.



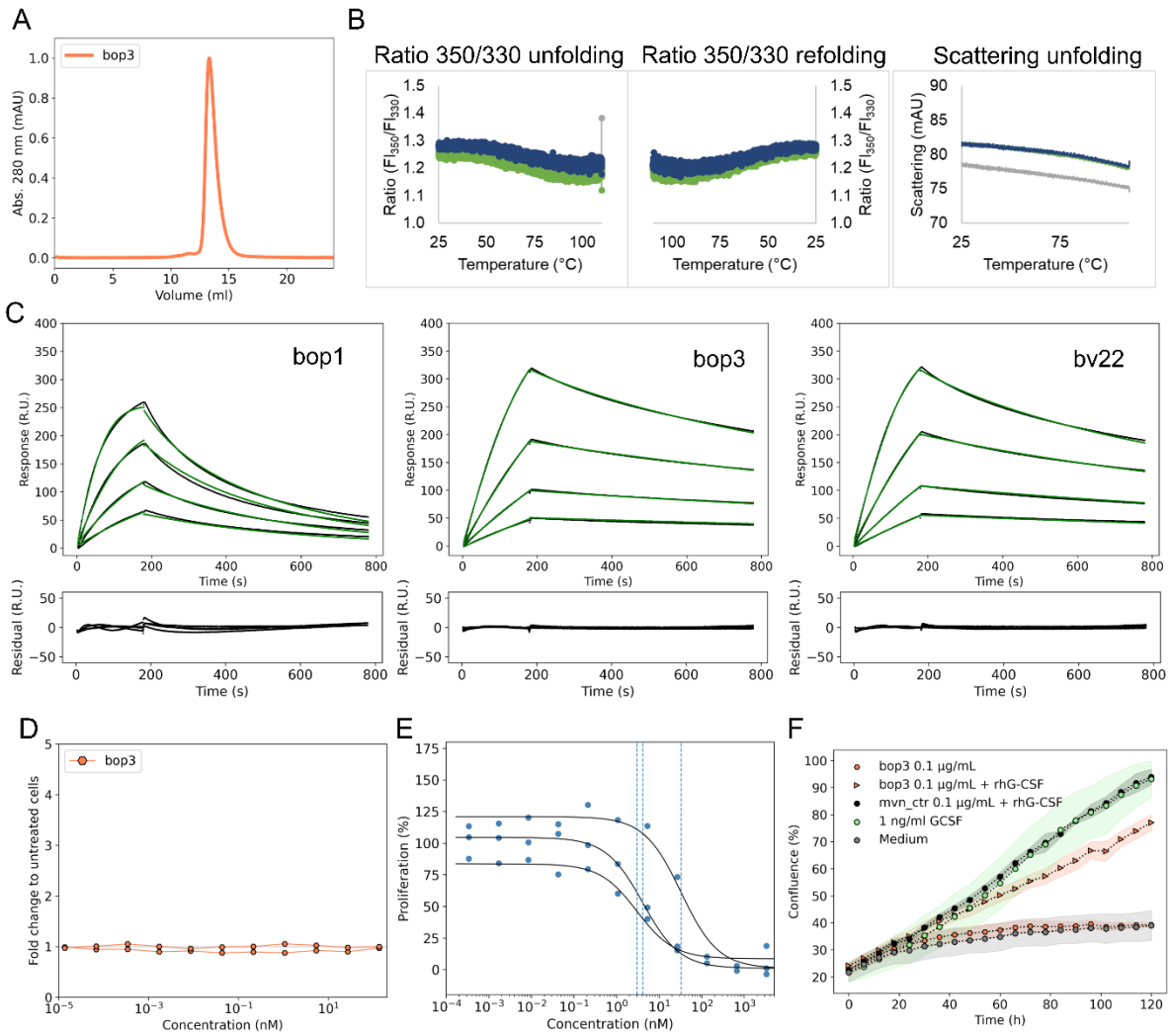

**Supplement Figure 7. Library generation and screening. (A)** Illustration of the PCR mutagenesis strategy used for library generation. **(B)** Sanger sequencing analysis of mutated sites. **(C)** FACS enrichment of binding clones after the start and sort 1 to 5 (SSC-A vs. PE-A). **(D)** Single-variant screen Z-scores to identify top clones in 96-well assay format.




**Supplement Figure 8. The enhanced-affinity variants preserve the biophysical properties of the starting designs. (A)** SDS-PAGE of main fractions of enhanced-affinity variants bv21, bv22, bv23, and bv24 after preparative size exclusion chromatography. **(B)** CD analysis of the Boskar4 variants (bv6, bv8, bv21, bv22, bv23 and bv24) show them to be helical. **(C)** NanoDSF of the variants shows them be to hyper-thermostable. Folding stability and colloidal stability are represented by ratio of fluorescence intensity at 350 nm and 330 nm and the scattering signal, respectively, for at least 4 technical replicates.




**Supplement Figure 9. The enhanced variants display stronger receptor-binding affinity and more potent inhibition of G-CSF-induced proliferation. (A-G)** SPR sensograms (black line) and the corresponding fits (green lines) of a two-fold dilution series of the Boskar4 (A) and the enhanced Boskar4 variants bv6 (B), bv8 (C), bv21 (D), bv22 (E), bv23 (F) and bv24 (G). The highest concentrations measured were 1250 nM for Boskar4 (B4), 125 nM for bv6 and bv8 and 31 nM for all other shown variants. Binding kinetic parameters derived from the displayed fits are listed in Table S3 (indicated in green). **(H)** Competitive inhibition NFS-60 assay with a constant concentration of 50 pg/ml rhG-CSF and the corresponding fit used to derive *$IC_{50}$* values for the designs (also compare Table S3).




**Supplement figure 10. Grafting the bv22 binding site as the primary binding site in bop1 yields the most potent inhibition, bop3. (A)** The bop3 is exclusively monomeric according to analytical SEC. Chromatogram shown was obtained using Cytiva Superdex 75 Increase 10/300 GL. **(B)** NanoDSF was performed in a temperature range from 25 °C to 110 °C. fluorescence intensity ratio at 350 nm and 330 nm, and the scattering signal of 4 technical replicates are shown. **(C)** SPR sensograms (black lines) and the corresponding fits (green lines) of a two-fold dilution series starting at 20 nM of bop1, bop3 and bv22. Binding kinetic parameters derived from these fits are provided in Table S3 (indicated in orange). **(D)** NFS-60 activity assay of two independent biological replicates of bop3 show no residual activity. The gray shade indicates the mean and standard deviation of the fold-change of untreated cells to the mean of untreated cells. **(E)** Competitive inhibition assays of G-CSF-induced proliferation in NFS-60 cells show bop3 to possess lower $IC_{50}$ values (indicated by dotted vertical lines). Results were obtained from a fit of three independent replicates. Mean and standard deviation of the individual fits are provided in Table S3. **(F)** 0.1 µg/ml of bop3 was tested in a healthy donor´s CD34+ HSPCs proliferation assay either without or with the addition of 1 ng/ml rhG-CSF. 0.1 µg/ml Moevan_control (mvn_ctrl, [1]), a protein lacking a functional G-CSFR binding site, was employed as negative control. Shown is mean and standard deviation of three parallel replicates.

***Supplementary Tables***

**Supplementary Table 1:** Shown are the mutations in bv6 compared to the given design (bop1, bop2, boa1, and boa2) and the corresponding residue in G-CSF (2D9Q).

| bop1 | G-CSF | bop2 | G-CSF | boa1 | G-CSF | boa2 | G-CSF |
|---|---|---|---|---|---|---|---|
| A5D | D104 | A5D | D104 | A9Q | Q11 | A9Q | Q11 |
| A9Y | L108 | A9Y | L108 | E10S | S12 | E10S | S12 |
| K13D | D112 | E10D | D109 | K13R | L15 | K13R | L15 |
| A16I | T115 | K13D | D112 | G14K | K16 | G14K | K16 |
| E17L | T116 | A16I | T115 | Y18M | Q20 | Y18M | Q20 |
| A20Q | Q119 | E17L | T116 | R21K | K23 | A20R | R22 |
| V101Q | Q11 | A20Q | Q119 | A98D | D104 | R21K | K23 |
| E102S | S12 | R21Q | Q120 | E102Y | L108 | A98D | D104 |
| A105R | L15 | K23E | E122 | Q106D | D112 | E102Y | L108 |
| Q106K | K16 | V101Q | Q11 | A109I | T115 | I103D | D109 |
| A109E | E19 | E102S | S12 | F110L | T116 | Q106D | D112 |
| F110M | Q20 | A105R | L15 | K113Q | Q119 | A109I | T115 |
| | | Q106K | K16 | | | F110L | T116 |
| | | A109E | E19 | | | K113Q | Q119 |
| | | F110M | Q20 | | | Q116E | E122 |
| | | E112R | R22 | | | | |
| | | I117D | D27 | | | | |

**Supplementary Table 2:** Sequences of constructs presented in this study.

| Construct | Sequence |
|---|---|
| bop1 | MAALDAALYEIYDGLILYQQRLKSLEGISPELGPALDALRYDMADFAILMAQAME EGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRAAAAQSIARKLEEM LEKAYQILRHLAAA |
| bop2 | MAALDAALYDIYDGLILYQQQLESLEGISPELGPALDALRYDMADFAILMAQAME EGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRAAAAQSIARKLEEM LRKAYQDLRHLAAA |
| bop3 | MAALDAALYEIYDGLILYQQRLKSLEGISPELGPALDALRYDMADFAYLMAQAM EEGLDSLPQRFLWKALEMIRKIQADAAALREKLAATYKGNDRAAAAQSIARKLEE MLEKAYQILRHLAAA |
| boa1 | MAALAAALQSIYRKLAEMQAKLKSLEGISPELGPALDALRYDMADFAILMAQAM EEGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRADAAVYIAADLEI LLEQAYQILRHLAAA |
| boa2 | MAALAAALQSIYRKLAEMQRKLKSLEGISPELGPALDALRYDMADFAILMAQAM EEGLDSLPQSFLRKALEMIRKIQADAAALREKLAATYKGNDRADAAVYDAADLEI LLEQAYEILRHLAAA |
| bv21 | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRWDMADFAYLMAQA MEEGLDSLPQRFLWKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL EAFLEKAYQILRHLAAA |
| bv22 | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAYLMAQA MEEGLDSLPQRFLWKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL EAFLEKAYQILRHLAAA |
| bv23 | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRYDMADFAYLMAQA MEEGLDSLPQSFLWKALEMIRWIQADAAALREKLAATYKGNDRAAAAVEIAAQL EAFLEKAYQILRHLAAA |
| bv24 | MAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRWDMADFACLMAQA MEEGLDSLPQRFLWKALEMIRKIQADAAALREKLAATYKGNDRAAAAVEIAAQL EAFLEKAYQILRHLAAA |
| $Intimin_{EHEC}$ (1-659)-E_tag-Boskar4-Myc_tag (display system reading frame of pNB4) | MITHGCYTRTRHKHKLKKTLIMLSAGLGLFFYVNQNSFANGENYFKLGSDSKLLT HDSYQNRLFYTLKTGETVADLSKSQDINLSTIWSLNKHLYSSESEMMKAAPGQQII LPLKKLPFEYSALPLLGSAPLVAAGGVAGHTNKLTKMSPDVTKSNMTDDKALNY AAQQAASLGSQLQSRSLNGDYAKDTALGIAGNQASSQLQAWLQHYGTAEVNLQ SGNNFDGSSLDFLLPFYDSEKMLAFGQVGARYIDSRFTANLGAGQRFFLPANMLG YNVFIDQDFSGDNTRLGIGGEYWRDYFKSSVNGYFRMSGWHESYNKKDYDERPA NGFDIRFNGYLPSYPALGAKLIYEQYYGDNVALFNSDKLQSNPGAATVGVNYTPI PLVTMGIDYRHGTGNENDLLYSMQFRYQFDKSWSQQIEPQYVNELRTLSGSRYDL VQRNNNIILEYKKQDILSLNIPHDINGTEHSTQKIQLIVKSKYGLDRIVWDDSALRS QGGQIQHSGSQSAQDYQAILPAYVQGGSNIYKVTARAYDRNGNSSNNVQLTITVL SNGQVVDQVGVTDFTADKTSAKADNADTITYTATVKKNGVAQANVPVSFNIVSG TATLGANSAKTDANGKATVTLKSSTPGQVVVSAKTAEMTSALNASAVIFFDGAP VPYPDPLEPAQPAMAALAAALAEIYKGLAEYQARLKSLEGISPELGPALDALRLD MADFATTMAQAMEEGLDSLPQSFLLKALEQIRKIQADAAALREKLAATYKGNDR AAAAAVEIAAQLEAFLEKAYQILRHLAAAAAAEQKLISEEDAAA |

**Supplementary Table 3:** The estimated dissociation rate ($k_d$), association rate ($k_a$), and dissociation constant ($K_D$) of the corresponding designs are provided the corresponding colors signify whether the samples were measured in the same experiments. SPR parameter for Boskar4 and bv6 colored in green were adopted from [2]. Furthermore, the half-maximal inhibitory concentration ($IC_{50}$) and the half-maximal effective concentration ($EC_{50}$) of NFS-60 assays are presented, along with indications of residual activity for non-inhibiting concentrations.

| Design | SPR parameter | | | | $IC_{50}$ mean ± sd (nM) | $EC_{50}$ mean ± sd (pM) | Residual activity |
|---|---|---|---|---|---|---|---|
| | $k_a$ ($M^{-1}$ $s^{-1}$) | $k_d$ ($s^{-1}$) | $K_D$ (M) | Chi² ($RU^2$) | | | |
| bv6 | $(1.2 \pm 0.4) \times 10^{6}$<br>$(2.4 \pm 0.7) \times 10^{5}$ | $(5.1 \pm 1.1) \times 10^{-3}$<br>$(3.3 \pm 0.4) \times 10^{-3}$ | $(4.3 \pm 0.4) \times 10^{-9}$<br>$(1.4 \pm 0.2) \times 10^{-8}$ | 2.33<br>2.51 | 93.1 | NA | yes |
| bop1 | $(8.8 \pm 4.0) \times 10^{5}$<br>$(2.0 \pm 0.6) \times 10^{6}$ | $(2.2 \pm 0.6) \times 10^{-3}$<br>$(6.3 \pm 0.7) \times 10^{-3}$ | $(2.7 \pm 0.4) \times 10^{-9}$<br>$(3.2 \pm 0.5) \times 10^{-9}$ | 4.76<br>13.2 | 35.3 ± 12.9 | NA | no |
| bop2 | $(1.9 \pm 1.1) \times 10^{6}$ | $(1.3 \pm 0.3) \times 10^{-3}$ | $(8.5 \pm 3.2) \times 10^{-10}$ | 12.7 | 58.2 ± 28.8 | 28.9 ± 7.9 | yes |
| boa1 | $(6.9 \pm 1.1) \times 10^{5}$ | $(2.4 \pm 0.2) \times 10^{-3}$ | $(3.7 \pm 0.8) \times 10^{-9}$ | 2.18 | 106.5 ± 78.0 | NA | yes |
| boa2 | $(1.4 \pm 0.7) \times 10^{6}$ | $(1.3 \pm 0.2) \times 10^{-3}$ | $(1.1 \pm 0.4) \times 10^{-9}$ | 12.3 | 37.8 ± 22.9 | NA | yes |
| bop3 | $(1.5 \pm 0.3) \times 10^{6}$ | $(1.8 \pm 0.3) \times 10^{-3}$ | $(1.2 \pm 0.1) \times 10^{-9}$ | 2.1 | 13.4 ± 16.9 | NA | yes |
| Boskar4 | $(3.4 \pm 1.7) \times 10^{4}$ | $(5.4 \pm 1.4) \times 10^{-3}$ | $(1.7 \pm 0.4) \times 10^{-7}$ | 13.9 | 2298.2 | NA | yes |
| bv8 | $(7.2 \pm 6.7) \times 10^{5}$ | $(1.7 \pm 1.4) \times 10^{-2}$ | $(2.6 \pm 0.3) \times 10^{-8}$ | 5.52 | 336.1 | NA | yes |
| bv21 | $(8.2 \pm 1.4) \times 10^{5}$ | $(1.1 \pm 0.1) \times 10^{-3}$ | $(1.4 \pm 0.1) \times 10^{-9}$ | 1.34 | 47.8 | NA | yes |
| bv22 | $(8.8 \pm 2.2) \times 10^{5}$<br>$(1.4 \pm 0.1) \times 10^{6}$ | $(9.2 \pm 0.7) \times 10^{-4}$<br>$(1.8 \pm 0.1) \times 10^{-3}$ | $(1.1 \pm 0.2) \times 10^{-9}$<br>$(1.3 \pm 0.1) \times 10^{-9}$ | 3.88<br>3.34 | 35.0 | NA | yes |
| bv23 | $(8.6 \pm 1.9) \times 10^{5}$ | $(1.5 \pm 0.1) \times 10^{-3}$ | $(1.8 \pm 0.2) \times 10^{-9}$ | 1.29 | 21.4 | NA | yes |
| bv24 | $(1.0 \pm 0.4) \times 10^{6}$ | $(1.7 \pm 0.3) \times 10^{-3}$ | $(1.7 \pm 0.3) \times 10^{-9}$ | 2.86 | 108.4 | NA | yes |

**Supplementary Table 4:** Crystallographic Data Collection and Refinement Statistics

| Structure | **bop1** |
|---|---|
| **Data collection** | |
| Space group | P1 |
| Cell parameters<br>a, b, c, (*Å)* | 23.26, 29.13, 37,76, 93.02,<br>92.60, 97.87 |
| Wavelength (*Å)* | 1.000 |
| Resolution limits (*Å)*[a] | 23.57 – 1.30 (1.38 – 1.30) |
| Unique reflections | 41065 (4515) |
| Completeness (%) | 85.3 (58.3) |
| Redundancy | 1.79 (1.44) |
| I/σI | 7.20 (1.00) |
| $R_{merge}$ (%) | 4.9 (53.0) |
| CC (1/2) | 99.5 (72.2) |
| **Refinement** | |
| Resolution limits (*Å)* | 23.57 – 1.30 |
| $R_{cryst}$ (%) | 22.6 |
| $R_{free}$ (%) | 26.5 |
| Mean B value (*Å$^2$)* | 28.551 |
| Ramachandran<br>preferred/allowed/outlier<br>regions (%) | 99.0/1.0/0.0 |

[a] Values in parenthesis refer to the highest-resolution shell.

***Supplementary Methods***

***Damietta input for the computationally-guided affinity maturation of bv6 and bv8***

Example Damietta specifications for weights, sample parameters, the mutable residues, and repackable residues of bv6 and bv8 for affinity maturation purposes (compare Material and methods, Computationally-guided affinity maturation).

```
# mutational targets
mut_res 38  ADEFHIKLMNQRSTVWY
mut_res 39  ADEFHIKLMNQRSTVWY
mut_res 42  ADEFHIKLMNQRSTVWY
mut_res 45  ADEFHIKLMNQRSTVWY
mut_res 46  ADEFHIKLMNQRSTVWY
mut_res 61  ADEFHIKLMNQRSTVWY
mut_res 64  ADEFHIKLMNQRSTVWY
mut_res 65  ADEFHIKLMNQRSTVWY
mut_res 68  ADEFHIKLMNQRSTVWY
mut_res 69  ADEFHIKLMNQRSTVWY
mut_res 72  ADEFHIKLMNQRSTVWY

# repacking targets
rpk_res 34
rpk_res 35
rpk_res 37
rpk_res 41
rpk_res 49
rpk_res 60
rpk_res 62
rpk_res 71
rpk_res 286
rpk_res 287
rpk_res 290
rpk_res 291
rpk_res 292
rpk_res 293
rpk_res 316
rpk_res 318
rpk_res 319
rpk_res 350
rpk_res 351
rpk_res 352
rpk_res 353
rpk_res 400
rpk_res 401
rpk_res 403

# sampling parameters (optional)
scramble_order                  1
m_mutations                     3
n_paths                         7
n_iters                         5

# mutagenesis scoring weights (optional)
mut_max_lj                      25.0
mut_w_pp                        1.0
mut_w_k                         0.0
mut_w_lj                        1.0
mut_w_solv                      1.0
mut_w_elec                      0.125

# repacking scoring weights (optional)
```

```
rpk_max_lj              25.0
rpk_w_pp                1.0
rpk_w_k                 1.0
rpk_w_lj                1.0
rpk_w_solv              1.0
rpk_w_elec              0.125
```

***References***


1. Skokowa, J., *et al.*, *A topological refactoring design strategy yields highly stable granulopoietic proteins.* Nature communications, 2022. **13**(1): p. 1-17.
2. Ullrich, T., *et al.*, *Tuning of granulopoietic signaling by de novo designed agonists.* Preprint, 2023, https://doi.org/10.1101/2023.11.25.568662.