THE UNIVERSITY OF CHICAGO

SOLUTION AND ANALYSIS OF THE IFNλ4 RECEPTOR COMPLEX USING CRYOEM ENABLED BY PROTEIN ENGINEERING AND MOLECULAR SIMULATION

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE PRITZKER SCHOOL OF MOLECULAR ENGINEERING

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

WILLIAM SANDERS GRUBBE

CHICAGO, ILLINOIS

JUNE 2024

“Leave chaos as is but cause it to evolve.”

– Shunji Yamanaka

“Inventions have long since reached their limit, and I see no hope for further development.”

– Julius Sextus Frontinus, Roman general and engineer, 1st century A.D.

“Pretend you’re waiting for the subway. You have to get in line. Wait your turn, after all.”

– Pablo Picasso to Joan Miró

Table of Contents

List of figures

List of tables

Acknowledgements

The work presented in this document has benefitted from hundreds of other people and is a product that builds on a field with a rich history and a bright future.

I am forever grateful to the University of Chicago, the Pritzker School of Molecular Engineering, and my advisors Juan Mendoza and Juan de Pablo for the opportunity to pursue my doctoral research. Juan Mendoza placed a great deal of trust in me and provided me with the opportunity to start a lab together. It was an honor and a privilege to work so closely together, navigating the lab from two people to ten through an exceptional time in human history. I am excited for the future of the Mendoza lab - I hope I have left it a better place. Juan de Pablo was unconditionally supportive of my research and professional endeavors, especially when I was doubting myself. Thank you for your enthusiasm to explore this new area of research together and for building a group with the resources and people to help those ideas become reality.

I am grateful for the guidance and insight provided by the other members of my committee, Jeffrey Hubbell and Aaron Esser-Kahn as well as numerous other professors who have mentored me. Every conversation with Jeff left me with insightful comments progressing this work, and Aaron was one of the reasons I joined the then-IME because of his infectious enthusiasm and curiosity. Andrew Ferguson, the late Albert Bendelac, Minglei Zhao, and others have helped me navigate this research journey and I am grateful for the passion they have put into their careers.

Many great collaborators contributed to this work. Thank you to Aileen Kauffman, who enabled great progress on these projects. Thank you to the Protein Simulation Subgroup (PSS), especially

Fabian and Walter, and for many conversations with Josh, Gustavo, Cintia, and Dan who nurtured my understanding of computational biophysics as friends and mentors. Bixia's contributions to the structural biology presented here were instrumental, and many collaborators at the National Institutes of Health contributed time and resources to this work.

Thank you to the many staff members at the University of Chicago, specifically in the Cummings Life Science Center, who do everything from maintaining common areas to ensuring that the building is functioning. I would also like to thank the Chicago Center for Teaching and Learning for their professional and financial support during my PhD. It was an honor to serve twice as a Graduate Teaching Fellow and to enrich pedagogy at this outstanding university.

To Lucy, who has been by my side throughout this adventure, this is as much yours as it is mine. At times, you were the only thing keeping me going. I will love you forever.

My family and friends are wonderful, and the balance and perspective they have provided me with was critical to my success in graduate school. Thank you to my mother who taught me how to read, my father who showed me how to work, and my sister for her unconditional love and support. Thank you to Karen and Mike, Jono, Joe, Kevin, the Lyndale 5, the late Randy Madry, and countless others who have supported and influenced me.

Finally, I feel the need to thank the many artists who inadvertently provided the soundtrack to this PhD. I spent quite a lot of time listening to their music, and the journey would not have been the same without them.

Abstract

Many cells in the immune system communicate using proteins called cytokines, which consist of distinct structural and functional families with roles in immunology, endocrinology, and cellular biology. A subset of cytokines, interferons, can be classified into three types (type I, II, and III) and have been explored for decades as therapeutics but have been plagued by issues ranging from off-target delivery to dramatic variations in patient response. The latter issue has presented the field of interferon biology with an interesting conundrum – genome-wide association studies of patients who failed to respond to interferon treatment identified a genetic region within the type III interferon locus that leads to expression of a protein called IFNλ4, implicating expression of this protein as the barrier for treatment. Since its discovery in 2013, structural and functional questions have continued to build around IFNλ4, but the inability to produce useful quantities of the protein has drastically limited experimental progress. This work presents the first method for robust expression of IFNλ4 and enables the solution of its ternary extracellular signaling complex, providing a structural basis of IFNλ4 function and lending insights into its dysfunction. Chapter 1 presents a discussion of interferons as therapeutics and a thorough discussion of therapeutic implications of type III interferons and specifically IFNλ4. Excerpts of this Chapter appear as published in *Trends in Biochemical Science* under the title "Recent and future perspectives on engineering interferons and other cytokines as therapeutics". Chapter 2 describes a method for high-yield expression and purification of IFNλ4, accompanied by select *in vitro* data exploring its extracellular function. Chapter 3 explores the use of molecular modeling and simulation to study and engineer cytokine-receptor complexes, specifically the ternary IFNλ complex, towards solving the structure of the IFNλ4 receptor complex. This Chapter appears as published in *Biophysical Journal,* titled "Molecular analysis of the type III interferon complex and its application in protein

engineering". Chapter 4 reports the structure of IFNλ4 in complex with its receptors IFNλR1 and IL10Rβ solved by cryogenic electron microscopy, as well as the structure of IFNλ3 in complex, and includes a comparison of the two related but distinct structures. Chapter 5 includes a thorough discussion of these findings and outlines future research questions for IFNλ4. In total, these results resolve the decade-long expression and purification bottleneck of IFNλ4 and provide structural and mechanistic insight into both its function and dysfunction in the human immune system. Future studies should be focused on the systemic impact of IFNλ4 expression and its role in the pathology of multiple diseases, with a specific focus on developing structure-informed treatments for patients with poor clinical outlooks due to IFNλ4 expression.

CHAPTER 1: Type III IFNs and the therapeutic ramifications of IFNλ4

Excerpts of this Chapter appear as published under the title "Recent and future perspectives on engineering interferons and other cytokines as therapeutics" in *Trends in Biochemical Sciences* by authors: Aung, T, Grubbe, W.S., Nusbaum, R.J., and Mendoza, J.L. Further information on this article can be found using the DOI: 10.1016/j.tibs.2022.09.005.

1.1 Interferons as therapeutics

Cytokines represent a family of secreted proteins that influence a broad range of physiological processes to regulate innate and adaptive immunity.[1] Over 130 known cytokines are categorized into six superfamilies that include interferons (IFNs), interleukins (ILs), chemokines, colony-stimulating factors (CSFs), tumor-necrosis factors (TNFs), and transforming growth factors (TGFs).[2–4] With their broad implications in various disease pathologies, cytokines were targeted in the first generations of immunotherapeutics for cancers and autoimmune disorders, with specific focus on IFNs due to their potent roles in the innate immune system. In 1986, recombinant IFNα received U.S. Food and Drug Administration (FDA) approval for treating hematological malignancies such as hairy cell leukemia and chronic myeloid leukemia, AIDS-related Kaposi sarcoma, melanoma, and chronic viral hepatitis B and C infections.[5] IFNγ was then approved for chronic granulomatous disease in 1990 followed by IFNβ for multiple sclerosis in 1996.[6–8] IFNs alone have continued to be explored as therapies for both chronic viral infections and cancers, with more than 20 clinical trials initiated since 2017 and continue to be explored for many indications.[9]

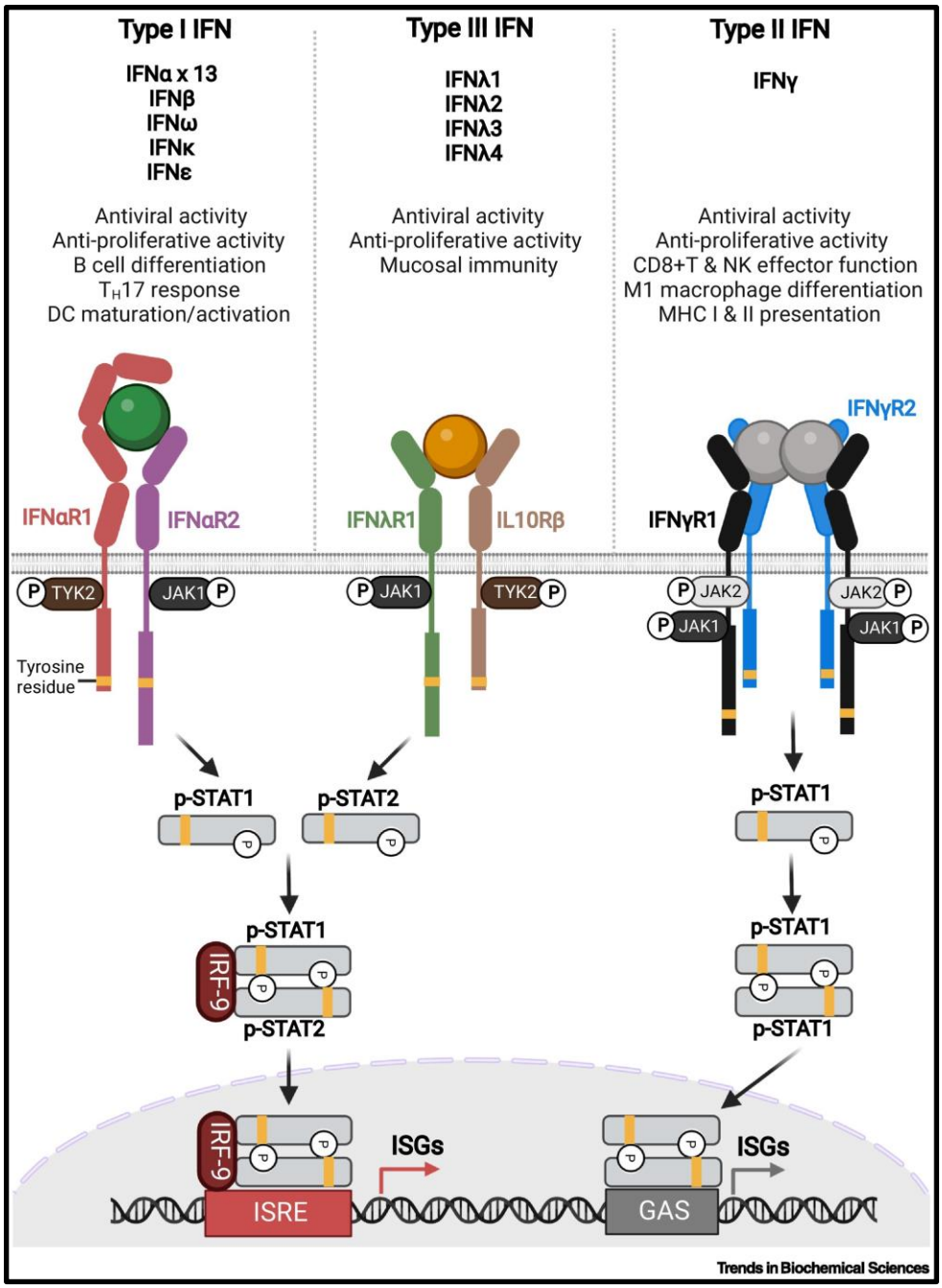


Figure 1: Interferon (IFN) signaling pathway and downstream cellular effects. Type I, II, and III IFNs signal via the intracellular JAK/STAT pathway through phosphorylation of signal transducer and activator of transcription (STAT)1 and STAT2 proteins. These STAT proteins then dimerize and, for type I and III IFNs, form a complex with IFN regulatory factor 9 (IRF-9) and translocate to the nucleus to bind to the IFN stimulated response element (ISRE) and activate transcription of IFN stimulated genes (ISGs). For type II IFNs, STAT1 forms a homodimer and binds the gamma IFN activation site (GAS) element.

IFNs can be classified into three families of proteins, referred to as type I, type II, and type III IFNs (**Figure 1**), all having distinct therapeutic applications. Most interferon monotherapies, however, are restricted by critical issues, the first of which is the biological pleiotropism inherent to cytokines with high therapeutic potential.[10] The context-dependent and highly dynamic nature of responses induced by these interferons *in vivo* contributes to suboptimal specificity and efficacy,[11] oftentimes requiring extensive engineering and design efforts to decouple the dual immunosuppressive and immunostimulatory or anti-proliferative and antiviral properties to bias signaling activity.[12,13] Second, the poor drug-like properties of these proteins impose another layer of limitations,[14,15] as the short circulating half-life of most cytokines and narrow therapeutic windows reduce their efficacy. Although high doses and frequent administrations allow therapeutic concentrations of interferons at targeted sites, this approach incurs a third issue – acute toxicity and morbidity.[16,17] IFN use is associated with flu-like symptoms and neuropsychiatric issues, as well as more severe conditions such as thrombocytopenia, leukopenia, and neutropenia; in many cases, these treatment regiments require either a dose reduction or complete cessation.[18–20] The fourth and final issue discussed here is the variability in patient response to IFN treatment, as well as many other types of immunotherapies, due to complex factors such as patient genotype.[21,22] Collectively, these challenges have represented major roadblocks for mainstream clinical applications of interferons and have incentivized countless research projects to harness and improve the natural therapeutic capabilities of these proteins.

1.2 Type III IFNs (IFNλ1-4) the IFNλ4 conundrum

Type III IFNs, also known as IFNλs, offer exciting therapeutic potential because they signal through the same pathway as type I IFNs, but expression of one the receptors in the signaling complex, IFNλR1, is limited to immune cell subsets and barrier tissues. As a result, this family of IFNs has represented an exciting group of naturally-targeted therapeutics that can serve as an alternatives to type I IFNs and avoid the negative off-target effects associated with ubiquitous expression of type I IFN receptors.[23–25] However, their signaling potency has consistently been found to be lower than type I IFNs, leading to limited clinical efficacy and introducing questions about their differential mechanism of action. Further, despite the promise of type III IFNs as therapeutics, the discovery of a new member of the type III IFN family in 2013, IFNλ4, presented the field with an unusual conundrum.

IFNλ4 was discovered as part of a genome-wide association study (GWAS) analyzing patient responses to pegylated interferon-alpha with ribavirin (pegIFN-α/RBV) for treatment of chronic hepatitis C (CHC).[21] Surprisingly, this study found that single nucleotide polymorphisms (SNPs) on chromosome 19q13.13 near the IFNλ3 gene were strongly associated with failure to clear hepatitis C virus (HCV). Specifically, these SNPs lead to a frameshift mutation resulting in IFNλ4 expression, posing a curious question: why does the expression of a protein from a canonically antiviral protein family lead to lower viral clearance? More recent studies have continued to perplex researchers due to the paradoxical observations that *in vitro* data indicates that the ability to activate the JAK/STAT pathway and induce antiviral genes is similar to its family members.[21,26] These questions become even more important when considering the global and public health implications of IFNλ4 expression, as surprising amounts of the world population are predicted to make this protein (as high as 70.7% of people of African descent, 31.2% European,

and 8% Asian).[27] As increasing numbers of studies link IFNλ4 expression to changes in disease progression and presentation,[28–32] the currently unexplained basis of function and activity for the IFNλ4 protein remains a priority for improving patient care.

While it is well understood that IFNλ4 is poorly secreted and that its expression leads to cell stress,[33] the full extracellular function of IFNλ4 and its comprehensive role in human health remains unrealized. Further, while predicted to be structurally similar to IFNλ1-3, IFNλ4 shares no more than 28% sequence similarity to the other IFNλs, with its closest structural relative being IFNλ3. This raises interesting questions regarding similarities and differences of receptor engagement between IFNλ1-4 – while this receptor promiscuity has been reported and well-studied for its low-affinity receptor IL10Rβ, similar questions about IFNλR1 have remained unanswered. A tremendous barrier in the field of IFNλ biology has been the inability to produce sufficient quantities of IFNλ4 protein to initiate these structural and functional studies, and while efforts have been made to express modified or engineered IFNλ4 protein, these studies remain largely non-quantitative or draw conclusions from impure and heavily modified protein products.[33–36] Answering these structural and functional questions must be approached by improving methods for IFNλ4 expression and is key to developing a better understanding of the type III IFN family of proteins and their role in the human immune system.

1.3 A structural approach to resolving the IFNλ4 conundrum

Towards understanding the mechanisms of IFNλ function, extensive efforts have gone into understanding the structural basis of type III IFN signaling in the hopes that it can be improved and to mechanistically explain the differences in signaling relative to type I IFNs. The first structural insights into type III IFN signaling were generated from solutions of the IFNλR1-IFNλ1 binary complex,[37] but the native affinity of IL10Rβ for the two proteins is too low to enable the solution of the native protein complex, limiting the understanding of the entire cytokine-receptor complex and requiring protein engineering efforts to overcome affinity limitations. The most successful attempt at potentiating IFNλs utilized directed evolution to affinity-mature IFNλ3 for its two receptors, IFNλR1 and IL10Rβ, yielding a version of IFNλ3 (IFNλ3-H11) with increased affinity for its receptor complex, enabling the solution of the type III IFN ternary complex.[38] IFNλ3-H11 also showed enhanced *in vitro* and *in vivo* efficacy but still fell short relative to type I IFN activity, highlighting the roles of mechanisms beyond extracellular affinity that must explain the difference in type I and type III IFN signaling. Further structural studies of the type III IFN signaling complex would continue to improve our understanding, as the ternary complexes of the four native IFNλs remain unsolved and are seemingly inaccessible because of the low affinity for IL10Rβ. These studies would also unearth key details in explaining how IFNλR1-IL10Rβ recognizes the four distinct members of the IFNλ family of proteins, including the sequence-distinct IFNλ4.

As previously discussed, thorough studies of IFNλ4 have been limited to the genetic and cellular level because of the previously described barrier imposed by IFNλ4 expression. This has left major gaps in understanding the extracellular function of the IFNλ4 protein and the structural basis of recognition by IFNλR1 and IL10Rβ. To fill these gaps in knowledge, we resolved to first

solve the expression bottleneck of IFNλ4 by creating a protocol for high-yield expression of the protein. By overcoming this challenge, we would not only provide a robust alternative to recombinant IFNλ4 expression for use in *in vitro* assays but also enable the structural studies needed to explain its peculiar *in vivo* behavior. These structural studies are further limited by the low affinity of the type III receptor complex, requiring the use of protein engineering to increase the overall affinity of the complex. Traditional methods of engineering a high-affinity ligand to enable solutions of cytokine-receptor complexes are not viable for IFNλ4, as the same issues associated with its expression lead to issues with yeast display and other comparable methods for directed evolution of IFNλ4. To initiate these structural studies, we required a two-pronged approach to engineer this protein complex, utilizing computational and experimental techniques to analyze the behavior of other type III IFN complexes with the goal of rationally engineering a ternary complex with improved affinity. Solution of the IFNλ4 receptor complex is further complicated by the behavior of the IFNλ4 protein and its well-documented tendency to precipitate out of solution and stick non-specifically to most materials, requiring adjustments to the traditional methods used to determine the structures of cytokine-receptor complexes. In total, solution of the IFNλ4 receptor complex represents a challenging but crucial discovery to advance the field.

By successfully expressing and purifying IFNλ4, extensive, quantitative experiments will be enabled so that we can better explore its comprehensive function in the human immune system. Going further, solution of the IFNλ4 receptor complex will elucidate the structural basis of IFNλ4 function and gain potential insights into the structural mechanisms of its dysfunction. The following Chapters thoroughly document our approaches to solving these critical issues in IFN biology, as well as our findings, towards resolving the IFNλ4 conundrum and providing therapeutic options for patients that express IFNλ4.

CHAPTER 2: High-yield expression of IFNλ4 and characterization of select *in vitro* functions

2.1 Introduction

Troubles associated with recombinant IFNλ4 expression and the resulting barriers it has presented for the field were major motivations for this work, specifically towards solution of the ternary IFNλ4/IFNλR1/IL10Rβ complex. Recombinant expression of IFNλ4 was first reported in the manuscript detailing discovery of IFNλ4,[21] along with notes detailing that "expression of [IFNλ4] was not detectable in cell medium, and protein was purified from the cell pellet after the cultivation of cells". This phenomena of IFNλ4 not being detectable in cell medium has since been a common occurrence, with several other published reports indicating the same findings,[33,34] leading the authors to estimate the amount of protein produced and thus used in experiments. While it is important to mention that one study reported improved expression and recovery of IFNλ4 by rationally glycoengineering the protein,[35] the protocols outlined by the authors were not repeatable in our hands, and many of the results reported in the study do not align with previously published biochemical and biophysical characterizations of type III IFNs.[38] Currently, there are two frequently utilized options available for obtaining unmodified IFNλ4: expression in bacterial culture followed by refolding,[32] or purchase from a single distributor (R&D Systems Product #9165-IF-025). The former requires approximately 10 liters of bacterial expression to acquire <100 micrograms of folded and active protein, while the latter requires $662 for 25 micrograms of ~85% pure protein; neither path enables structural studies, as milligram (mg) amounts of protein are typically required for x-ray crystallography and cryogenic electron microscopy, and both paths prevent clear financial and logistical bottlenecks to extensive studies of IFNλ4. Thus, a method enabling high-yield expression of IFNλ4 is of paramount importance to enable research progress.

This Chapter outlines methods we have developed for expression and purification of amounts of IFNλ4 required for structural studies as well as any extensive *in vitro* or *in vivo* experiments. This modular system uses baculovirus infection in insect cells, a well-reported and robust method for recombinant cytokine expression,[38–40] lending itself well to pre-existing workflows for type III IFN expression and characterization, but expression of the construct and the following downstream purification steps can be considered for other systems of protein expression. Major innovations reported in this Chapter include the design of a plasmid construct covalently linking the low-affinity receptor of IFNλ4, IL10Rβ, via a protease cleavable flexible peptide linker to enable secretion of IFNλ4, and adjustments for the purification of the IFNλ4 protein to incorporate increased concentrations of sodium chloride (NaCl) at all steps to reduce non-specific adherence during processing. With successful protein expression and purification, we are also able to perform select *in vitro* experiments including cell signaling and gene induction to assess the activity of IFNλ4 compared to IFNλ3, IFNλ4's closest known relative, and IFNω1, a type I IFN, laying the groundwork for future experimental work utilizing our methods.

2.2 Methods and Materials

Chemicals, materials, and reagents

All chemicals, materials, and reagents were ordered from either Thermo Fisher Scientific or MilliporeSigma unless otherwise noted.

Molecular cloning of genes of interest for protein expression

All genes used for protein expression in this document were either ordered from Twist Bioscience or were obtained from previously published genetic constructs.[38,40] All DNA primers used for cloning, sequencing, and all other experiments were ordered from Sigma-Aldrich unless otherwise stated. A list of all relevant genes and DNA primers used for cloning, specifically for IFNλ4, are available in the supplemental information **(Appendix A, Supplemental Table 1) (Appendix A, Supplemental Table 2)**. Polymerase chain reaction (PCR)[41] was used in all cases to amplify genes and prepare them for molecular cloning, either by isothermal assembly[42] or using restriction digest for ligation.[43] Proteins expressed in mammalian systems utilized the pVLAD plasmid construct[44] and proteins expressed in insect systems utilized the pAC plasmid construct.[45]

Production of baculovirus for protein expression using insect cells

All proteins discussed in this document were made using baculovirus infection. Virus was created using Sf9 cells (Gibco) cultured in Sf-900 II SFM media (Gibco). In brief, 400,000 cells were plated in a 12-well plastic cell culture plate to adhere for 30 minutes. Simultaneously, 800 nanograms of plasmid containing the protein of interest were mixed in a microtube with linearized defective backbone (BestBac, Expression Systems, product #91-002) at a 10:1 mass ratio left to sit for five minutes. After incubation, this DNA was added to antibiotic-free SFM

media containing 10% FuGENE HD Transfection Reagent (Promega) and left to incubate for 30 minutes. After 30 minutes, the media of the plated cells was removed and antibiotic-free SFM media was gently added to the tubes containing the DNA-FuGENE mixture, followed by delicate application of the media-DNA-FuGENE mixture to the adhered Sf9 cells. The 12-well plate was then wrapped in a moist paper towel, wrapped tightly in plastic wrap, and left to sit for seven days at 28° C.

After seven days, the liquid containing the virus (P0) was removed from the plate and 10% fetal bovine serum (FBS) (Gibco) was added to stabilize the virus for storage at 4° C. This P0 was then added to a liquid suspension of Sf9 cells at a density of 1 million cells/mL (1e6 cells/mL) at a 1:500 ratio and was left to incubate with rotation for at least five days at 28° C. The first passage (P1) of the virus was then tested for protein expression before infection of large volumes of liquid cell culture. Further passages of the virus were made as needed using the same protocol detailed above, with no more than three passages (P3) being used.

Insect cell culture for protein expression

High Five (Hi5) insect cells (Expression Systems) for protein expression were cultured in Insect Xpress liquid media (Lonza) containing 1:1000 gentamycin sulfate (Corning) with 120 RPM at 28° C. For optimal protein expression, cells were passaged daily to maintain a working density of 1e6 cells/mL. Test reactions of baculovirus were performed using viral dilutions of 1:1,000, 1:500, and 1:250 for times of either two or three days to determine conditions for expression at scale. Prescence of protein was determined using an SDS-PAGE Stain-Free protein gel (Bio-Rad). All proteins discussed in this document use insect cell expression with the exception of IFNλR1.

Mammalian cell culture for protein expression

HEK293S GnTi- cells (ATCC) were cultured in FreeStyle 293 expression medium (Gibco) with 2% FBS and 1:1,000 penicillin-streptomycin (Gibco) with 120 RPM at 37° C with 5% CO_2. Cells were maintained at a culture density between 0.5e6 cells/mL and 2e6 cells/mL. Test reactions of baculovirus were performed using viral dilutions of 1:25, 1:12, and 1:8 with 1:100 sodium butyrate for three days to determine conditions for expression at scale. For IFNλR1, cells were moved to a 30° C incubator to improve protein folding. Prescence of protein was determined using an SDS-PAGE Stain-Free protein gel (Bio-Rad).

Mammalian cell culture for functional assays

HAP1 cells were a gift from Jan Carette at Stanford University and were cultured in DMEM with high glucose and no glutamine (Gibco) with 10% FBS and 1:1,000 penicillin-streptomycin. Subculturing utilized 6-well plastic tissue culture plates and expansion utilized either T25 or T75 flasks at 37° C with 5% CO_2. Cells were split per manufacturer instructions using either 10:1 or 15:1 dilution. A549 cells were a gift from Curt Horvath at Northwestern University and were cultured using the same conditions, except with splits of either 3:1 or 8:1.

General protocol for protein harvest using insect expression

Cell culture containing the expressed proteins was first centrifuged at 2,000 RPM for 20 minutes to remove cell pellets. To remove impurities, 40:1 2 M Tris-HCl (pH = 8), 1:1,000 5 M $CaCl_2$, and 1:1,000 1 M $NiCl_2$ were added and mixed at 200 RPM at room temperature for at least 30 minutes. This mixture was then centrifuged again at 6,000 RPM for 20 minutes to remove precipitate. The final supernatant containing the protein was transferred to a new container containing a magnetic stir bar and nickel chelating resin (G-bioscience) was added for

purification to recover expressed proteins which included a 6x histidine tag (HisTag). Nickeling proceeded overnight with 200 RPM at room temperature.

The next day, protein was harvested using a vacuum Buchner funnel and washed with 20 mM HEPES + 150 mM NaCl (1X HBS, pH = 7.5) with 20 mM imidazole. Nickel was then transferred to a Poly-Prep chromatography column (Bio-Rad) and elution buffer (1X HBS (pH = 7.5) + 250 mM imidazole) was added to remove protein product from the nickel.

General protocol for protein harvest using mammalian expression

Cell supernatant containing the expressed proteins were centrifuged at 2,000 RPM for 20 minutes to remove cell pellets. Supernatant was then carefully removed and transferred to a container containing a magnetic stir bar. Nickel was added as previously described, with nickeling proceeding overnight with 200 RPM at 4° C. Protein was then harvested the next day as previously described.

Protein harvest for IFNλ4

High-yield expression of IFNλ4 is enabled by covalently linking the protein to its low-affinity receptor IL10Rβ, with sfGFP added as an additional chaperone protein to increase molecular weight and improve separation by size exclusion chromatography (SEC). After expression, protein was harvested as previously described except with 1) a higher pH of the buffer (pH = 8.3) and 2) the inclusion of 500 mM NaCl at all steps. After harvest, protein was treated with 3C protease overnight at 4° C to cleave IFNλ4 away from the sfGFP-10RB chaperone.

Protein purification using fast-protein liquid chromatography (FPLC)

Proteins were purified using an S75 Superdex column (Cytiva) equilibrated with 1X HBS (pH = 7.5) on an NGC liquid chromatography system (Bio-Rad). Elution fractions containing the proteins of interest were visualized using an SDS-PAGE Stain-Free protein gel and were then concentrated, quantified via UV-Vis spectroscopy to determine molar concentration, and frozen in 10% glycerol before storage at -80° C. For IFNλ4 and proteins containing IFNλ4, 1X HBS (pH = 8.3) + 500 mM NaCl was used as the equilibration buffer.

***In vitro* phosphorylated signal transducer and activator of transcription (pSTAT) signaling assay**

Hap1 cells were plated in a 96-well plate and treated with serial dilutions of either IFNλ4, IFNλ3, or IFNω1 for 15 minutes at 37° C. After removing the supernatant, cells were then treated with trypsin (Gibco) for 5 minutes and resuspended gently. Following resuspension, cells were added to 16% PFA (Electron Microscopy Services) to fix for 10 minutes at room temperature before adding ice-cold methanol to permeabilize cells. Samples were stored at -80° C until use. To quantify activation of pSTAT1, cells were washed 3 times with 0.5% PBSA and then stained with an anti-Y701 pSTAT1 monoclonal antibody (Cell Signaling, product #9174S) according to manufacturer instructions. Samples were analyzed using a CytoFlex S flow cytometer (Beckman Coulter). Data were analyzed in Prism (GraphPad version 9).

Quantification of gene induction by quantitative reverse transcription PCR (RT-qPCR)

Hap1 cells were plated in a 6-well format (approximately 600,000 cells) and treated with 100 nM IFNλ4, IFNλ3, or IFNω1 for 6 hours before sample mRNA was harvested using a Monarch Total RNA miniprep kit T2010 (NEB) following manufacturer instructions for

adherent cells. cDNA was then synthesized via RT-PCR according to manufacturer instructions (Applied Biosystems) using 1 microgram of RNA. qPCR was performed using a QuantStudio Real-Time PCR system (Thermo Fisher Scientific) with SYBR Green as the reporter, again according to manufacturer instructions (Applied Biosystems). Transcription quantification was normalized to expression of 18S ribosomal rRNA. Primers used for these experiments can be found in the supplemental information (**Appendix A, Supplemental Table 3**). Data were analyzed using the Relative Quantification (RQ) module available through Thermo Fisher Connect.

2.3 Results

Resolving the IFNλ4 expression bottleneck required two distinct steps: enabling secretion of the protein from cells, and successfully purifying the protein from cell supernatant. Before initiating in-depth experimentation, we attempted to express IFNλ4 according to previously published protocols for type III IFN expression[38] and also attempted previously published methods that proposed improved expression and recovery in our own hands.[35] We were unsuccessful in expressing any protein construct containing just IFNλ4, either by baculovirus or by transient transfection, and attempts to express the glycoengineered version of IFNλ4 using baculovirus and transient transfection were unsuccessful. These findings led us to the belief that, like other authors have reported,[33] the IFNλ4 protein was not secreted from cells, and that alternative methods involving molecular chaperones for protein secretion may be required.

Molecular chaperones have been utilized extensively to improve expression and purification of proteins of interest,[46] often by covalently linking the chaperone to the protein of interest via a flexible peptide linker. Traditional chaperone proteins, including SUMO and GST, did not resolve the issues associated with IFNλ4, requiring us to explore alternative approaches. Motivated by studies of other difficult-to-express cytokines,[47] we attempted baculovirus expression of IFNλ4 linked to its receptors (either IFNλR1 or IL10Rβ) in hopes that the presence of the receptor would improve expression, either through speculated low-affinity interactions or simply by biasing expression towards the receptors that are well-behaved in our hands. We also noted that IFNλ4, like IFNλ1-3, has 20 amino acids towards its N-terminus. Removal of this region has promoted expression of type III IFNs in previous studies and confirmed that these amino acids to not interact at the receptor interface,[38] leading us to investigate if this would improve IFNλ4 expression as well.

In both cases, the receptor construct was located N-terminal (upstream) to IFNλ4 to encourage expression of the chaperone. All attempts to express a IFNλR1-IFNλ4 construct failed, and expression of any construct containing the N-terminal peptide (amino acids 1-21) on IFNλ4 also failed. Excitingly, though, secretion of IFNλ4 is enabled by removing this hydrophobic region and covalently linking it to IL10Rβ via a flexible peptide linker. This linker also contains a 3C protease cut site such that treatment with the protease will cleave the IFNλ4 away from IL10Rβ and enable recovery of pure IFNλ4. We observed expression of the IL10Rβ-IFNλ4 construct in both mammalian and insect systems, but expression in mammalian systems was only successful when paired with expression of IFNλR1 and led to low total amounts of IFNλ4 (**Appendix A, Supplemental Figure 1**) robust expression of the IL10Rβ-IFNλ4 construct in insect cells, then, became the chosen method for IFNλ4 expression and was the focus of all future experiments (**Figure 2A)**.

After ensuring expression and confirming presence of protein using SDS-PAGE protein gels, we tested to see if the proteins would cleave after treatment with 3C protease and separate using FPLC, which is readily observed (**Figure 2B**). After confirming expression in small-scale tests, we attempted expression at larger volumes, leading to yields >1 mg/L of total protein and providing ample amounts of IFNλ4 to proceed with optimization of the purification steps. Using these improved expression conditions, protein harvest and elution was initially working as anticipated, but purification using fast protein liquid chromatography (FPLC), a form of size exclusion chromatography (SEC), was not. We were able to quantify the amount of protein being injected for purification using UV-Vis spectroscopy, revealing that we were injecting large amounts of product. However, protein injected over the column would not elute, leading us to

believe that it was sticking to the column. This led us to analyze our buffer conditions and adjust our standard protein purification protocols in order to make IFNλ4 less prone to sticking.

The biochemistry of IFNλ4 was something that we identified as a potential issue, as the isoelectric point (pI) of the protein is around 11.5, which is highly basic. Indeed, many of the residues predicted to be solvent exposed are positively charged, providing a potential explanation for the protein's sticky nature. To improve recovery, we elected to increase the pH of the buffers used for IFNλ4 from 7.5 to 8.3 and to add 500 mM NaCl to all buffers at all steps in hopes that the presence of excess salt and ions would reduce undesired interactions. We found that these conditions not only increased the stability of the IFNλ4 proteins in solution, preventing precipitation, but also prevented the protein from sticking to the size-exclusion column, enabling recovery using an S75 Superdex column (**Figure 2C)** – ion exchange chromatography was attempted, but the high concentration of NaCl needed to reduce sticking of IFNλ4 to the inner mechanisms of the instrument made for additional and unnecessary challenges. Decreasing the concentration of NaCl below this threshold resulted in sticking and increasing it above this point did not make an apparent difference in behavior. We explored other salts like $MgCl_2$ as well as different buffers including 3-(*N*-morpholino)propanesulfonic acid (MOPS) and Tris(hydroxymethyl)aminomethane (Tris base) but ultimately found that 1X HBS (pH = 8.3) + 500 mM worked well for our purposes. While increasing the pH of the buffer was found to slightly increase yields and brought us closer to the pI of IFNλ4, increased concentration of NaCl was the main factor enabling recovery.

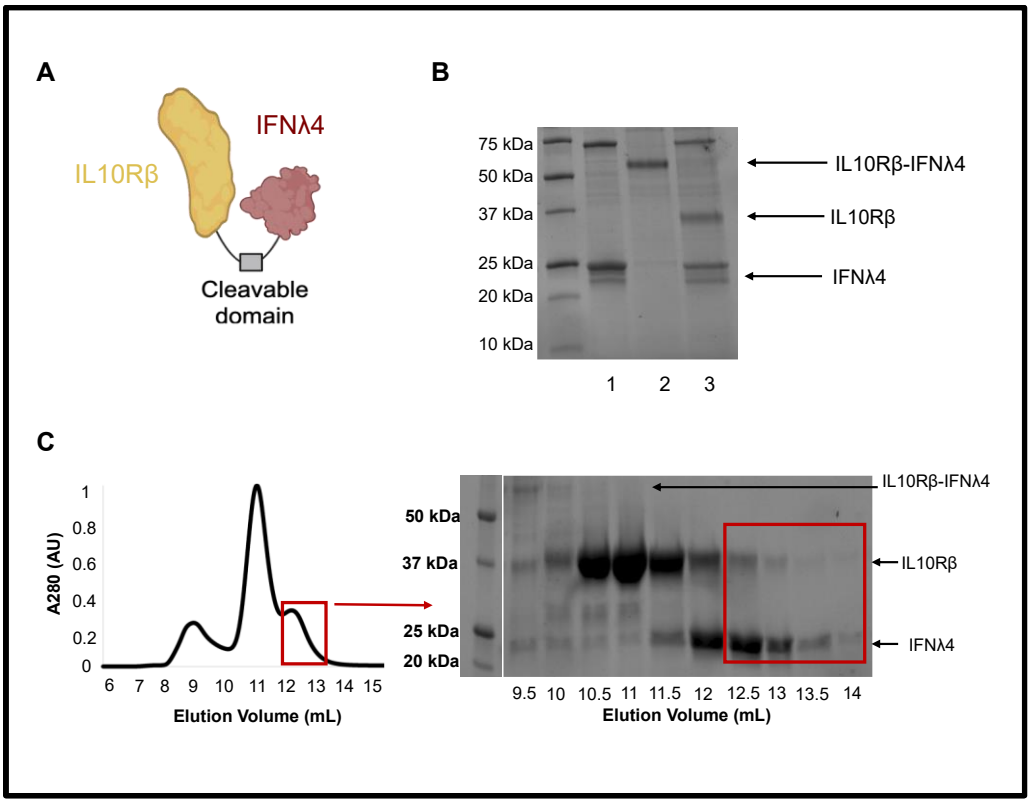


Figure 2: Expression and purification of IFNλ4 linked to IL10Rβ. A) Cartoon representing the protein construct used to enable IFNλ4 secretion. **B)** SDS-PAGE protein gel showing IL10Rβ-IFNλ4 protein samples without (Lane 2) and with (Lane 3) 3C protease treatment. Lane 1 is 3C protease alone. **C)** FPLC trace (left) and SDS-PAGE protein gel (right) demonstrating expression and purification of IFNλ4 after treatment with 1:100 3C protease. (AU = arbitrary units)

This first generation of IFNλ4 expression constructs demonstrated a vast improvement over all other IFNλ4 expression protocols. However, the molecular weights of IL10Rβ (25 kilodaltons (kDa)) and IFNλ4 (19.7 kDa) were too similar post-cleavage to enable complete separation (**Figure 2C**). We elected to continue to improve on this construct by adding superfolder green fluorescent protein (sfGFP)[48] to the N-terminus of IL10Rβ (**Figure 3A**), yielding an sfGFP-IL10Rβ-IFNλ4 construct that expresses equally as well and adds 27 kDa of mass to the IL10Rβ protein, yielding superior separation post-3C treatment (**Figure 3B**) **(Appendix A, Supplemental Figure 2)**. The small ubiquitin-related modifier (SUMO) tag was

explored, but spontaneous cleavage by endogenous proteases only complicated the initial issue. IFNλ4 expressed and recovered using this sfGFP construct yielded the best final product (~2 mg/L) and this approach was used for all future experiments, with good separation enabled by 3C treatment and FPLC and further purification optional if removing all the sfGFP-10RB is desired (a trace amount appears in the earlier fractions of IFNλ4) (**Figure 3C**).

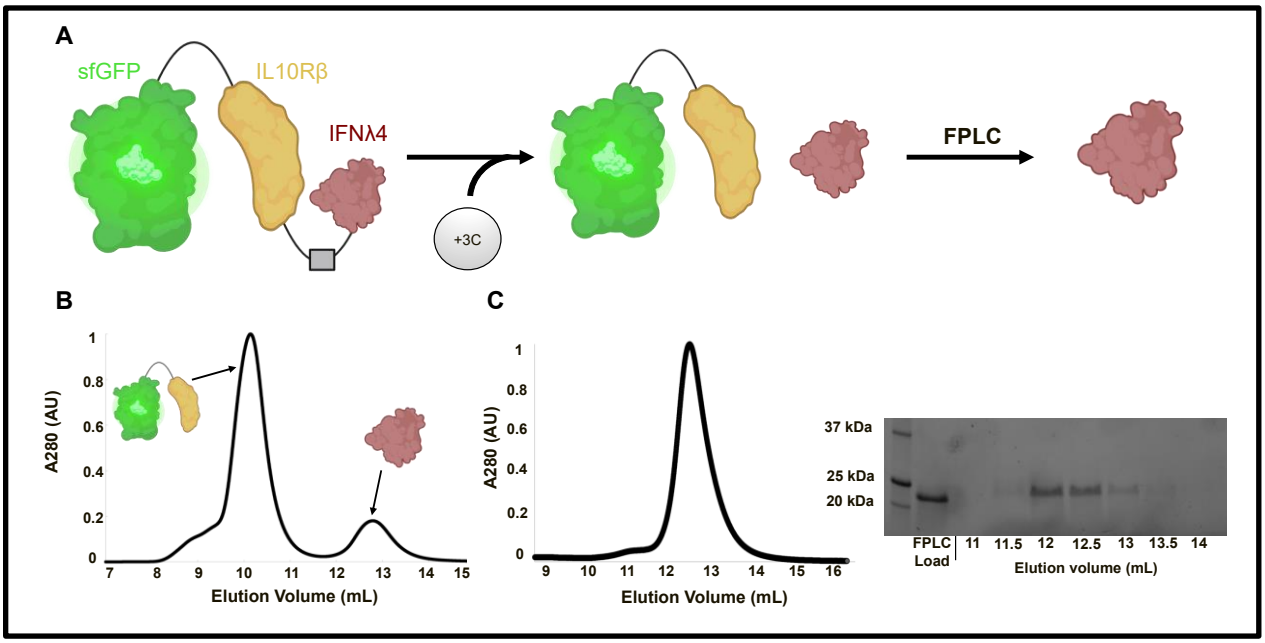


Figure 3: High-yield expression of IFNλ4 using sfGFP-IL10Rβ as a molecular chaperone. **A)** Superior expression of IFNλ4 is enabled by covalently linking it to sfGFP-IL10Rβ via a protease-cleavable site. Following treatment, IFNλ4 is harvested and purified via fast-protein liquid chromatography (FPLC). (AU = arbitrary units) **B)** FPLC trace of sfGFP-10RB-IFNλ4 after treatment with 3C protease. **C)** FPLC trace and SDS-PAGE gel of pure IFNλ4.

With the ability to make larger amounts of high-quality IFNλ4, we sought to characterize some of its *in vitro* functions and see if the differences observed in the structure of the protein complexes would be apparent in experimental results. We first measured *in vitro* phospho-STAT1 (pSTAT1) signaling and induction of interferon-stimulated genes (ISG) relative to IFNλ3 and a type I IFN, IFNω1 in Hap1 cells (**Figure 4A,B**). We confirm that treatment of cells with IFNλ4 activates the JAK/STAT pathway and potently induces the antiviral ISGs MX1 and

ISG15 but find that IFNλ4 does not elicit as potent a response as neither IFNλ3 nor IFNω1 and does not elicit a response from canonical anti-proliferative ISGs SAMD9L and APOL3. Interestingly, we observe that IFNλ4 has a lower EC50 than IFNλ3; while the EC50 IFNλ3 is 100-fold different than IFNω1, IFNλ4 is only around 20-fold, potentially indicating that cells are more sensitive to IFNλ4 relative to IFNλ3 but that extracellular IFNλ4 signaling does not elicit the same potency as IFNλ3 as measured by Emax. This trend was observed with A549 cells, as well (**Appendix A, Supplemental Figure 3),** indicating that this differential behavior may be observed across many cell types. Treatment with type I IFN continues to elicit the most potent responses from cells, in line with many previous studies.

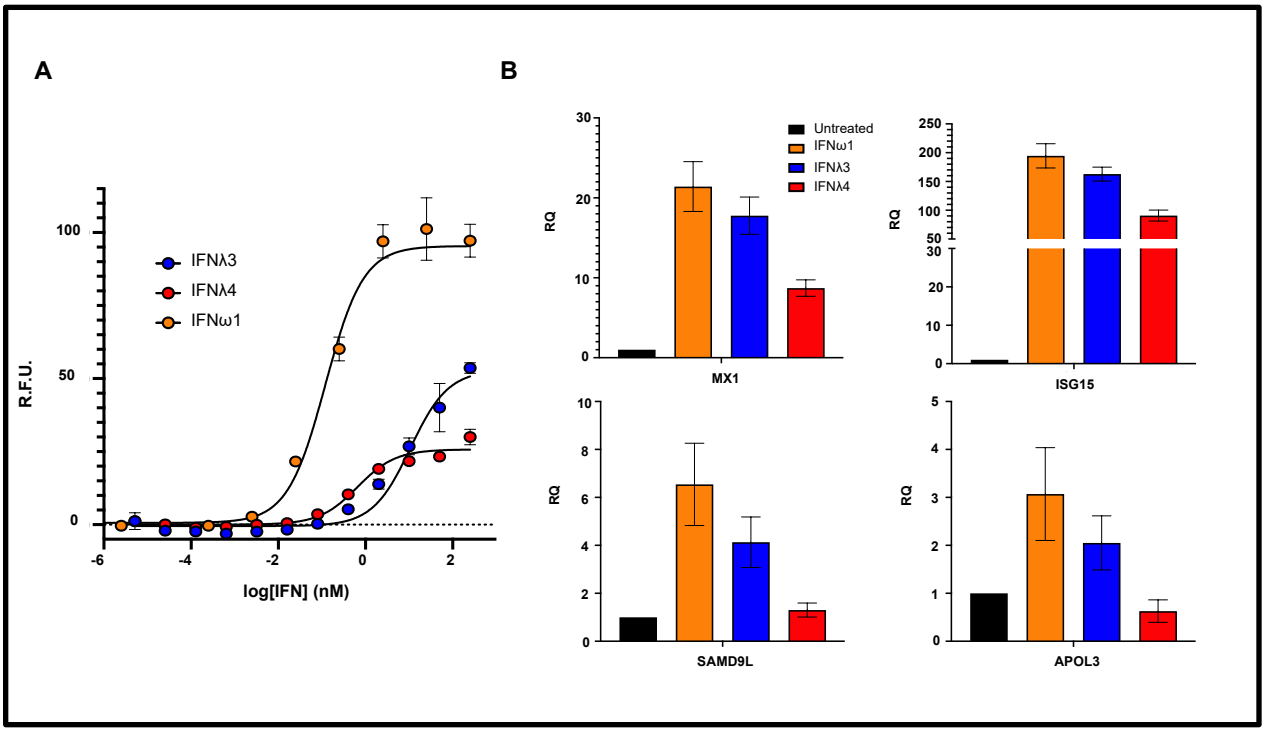


Figure 4: *in vitro* characterization of IFNλ4 compared to IFNλ3 and IFNω1. A) pSTAT1 signaling of IFNω1, IFNλ3, and IFNλ4 in Hap1 cells. Curves are fit to a first-order logistic model, with error bars representing +/- standard error of the mean (SEM) (n = 3). **B)** Relative quantification (RQ) of select genes induced by IFNω1, IFNλ3, and IFNλ4 in Hap1 cells treated with 100 nM interferon for 6 hours as measured by qPCR. MX1, ISG15 = antiviral ISGs. SAMD9L, APOL3 = anti-proliferative ISGs. Error bars represent 95% confidence intervals from three independent biological samples (N = 3).

2.4 Discussion

Using the methods outlined in this Chapter, we have developed, to our knowledge, the only high-yield expression and recovery protocol of unmodified IFNλ4 protein using eukaryotic cells. From 1 L of insect culture, we typically expect to recover ~100 micrograms of pure IFNλ4. When compared to the standard of bacterial refolding, this represents a technical improvement in yield of about 10-fold, but expression and recovery benefits from scale in a way that makes this relationship non-linear, as batches can be combined and processed together. This important development enables robust IFNλ4 expression and can produce sufficient protein for any experiments desired, including further *in vitro* and *in vivo* characterization as well as structural studies of the IFNλ4 receptor complex.

The construct utilized for IFNλ4 expression could continue to be optimized for pure IFNλ4 recovery by adjusting the linker length between sfGFP and IL10Rβ as well as IL10Rβ and IFNλ4. Initial linker design between IL10Rβ and IFNλ4 was informed by the previously mentioned strategy for receptor-linked expression of troublesome cytokines[47] to position IFNλ4 roughly 40 angstroms (Å) away from IL10Rβ (~20 amino acids) and a short linker (4 amino acids) separated sfGFP from IL10Rβ. Rigorous optimization of these parameters would likely yield sufficient purity of IFNλ4 after a single expression, leading to higher overall recovery after FPLC, but the methods and constructs presented above are sufficient for the goals of this project.

Expression in insect cells represented the best overall path for high-yield IFNλ4 expression, but as previously mentioned, we did observe expression of IL10Rβ-IFNλ4 in our mammalian system when co-infecting with IFNλR1. Enthusiasm for this approach was high at first because co-expression of these constructs could have enabled single-step purification of the IFNλ4 receptor complex after treatment with 3C – this was a key part of the strategy for previous

studies utilizing receptor-linked expression of cytokines.[47] However, we observed that the proteins do not assemble as a complex during purification (**Appendix A, Supplemental Figure 1)**, indicating that the affinity of the native IFNλ4 ternary complex is too low to enable co-elution and that engineering would indeed be required to overcome this barrier for structural studies.

The *in vitro* functional data presented here represents the first steps towards rigorous extracellular characterization and ultimately functions as proof that our protein behaves as expected in these assays. Comprehensive functional assays, including antiviral and anti-proliferative assays, extensive biophysical characterization, and more diverse panels of gene induction represent a natural next step in the future of this project. Given the biochemical nature of IFNλ4 and the problems associated with it, these exercises are non-trivial, but are limited only by the goals and enthusiasm of the IFNλ research community. The modular nature of this expression cassette also lends itself well to the rigorous investigation of the multiple variants of interest that have been identified by genetic studies of IFNλ4, with many point mutations being identified as impactful for *in vivo* function.[21,27,49–52] Until now, extracellular characterization of these proteins has been largely inaccessible, and the methods described here enable studies to understand the impact of these mutations on patients in the clinic.

2.5 Conclusion

This Chapter details our approaches to enabling high-yield expression of recombinant IFNλ4. With the IFNλ4 expression and purification bottleneck resolved, structural studies of the IFNλ4 receptor complex are now viable. These structural studies are further motivated by the interesting *in vitro* results observed for differential action of IFNλ3 and IFNλ4, with structural mechanisms potentially explaining these differences.

CHAPTER 3: Molecular analysis and simulation of the type III IFN complex and its applications in protein engineering

This Chapter appears as published under the title “Molecular analysis of the type III interferon complex and its applications in protein engineering” in *Biophysical Journal* by authors: Grubbe, W.S., Byléhn, F, Alvarado, W, de Pablo, J.J., and Mendoza, J.L. Supplemental information for this Chapter can be found in Appendix B. Further information on this manuscript can be found using the DOI: 10.1016/j.bpj.2023.09.021.

3.1 Introduction

The relationship between the structure and function of a protein is a critical paradigm in biology. However, the influence of specific interactions at each protein-protein interface on the stability and overall behavior of protein complexes is not always known. Protein complexes interact via transient, non-covalent protein-protein interactions (PPIs) that can often have a low overall affinity (millimolar to micromolar) and complicate efforts to obtain experimental structures, drastically limiting structural and behavioral knowledge of many native protein complexes. If a structural understanding of a complex is desired, researchers must often use protein engineering approaches, such as introducing mutations that enhance binding between two or more proteins via directed evolution.[53] This technique has been applied frequently in the field of cytokine and interferon biology to solve receptor-ligand complexes, leading to an improved understanding of protein complex behavior and immune function as well as helping to guide the design of improved cytokine therapeutics.[13,38,39] While this approach can be effective, it can often require years of experimental work, can be limited by how difficult a protein is to express and purify, and can result in perturbation of the true binding interfaces of a complex; thus, new approaches for understanding and guiding engineering of protein complexes like these are needed.

Advances in computation have led to powerful computational tools to study protein structure. The most robust methods in this space use a machine learning and artificial intelligence approach to predict a protein's 3-dimensional shape from its primary amino acid sequence as well as its sequence homology to other proteins, with some of these modeling tools approaching functional equivalency to experimental determination.[54–56] While structure prediction offers valuable information, the dynamic nature of proteins, PPIs, and the effects of

solvent on protein complex behavior are not fully captured by a structural model alone. Further, these approaches continue to struggle with assessing the impact of individual mutations on protein complex formation and behavior. Molecular dynamics (MD) simulations, however, provide deeper insights into the atomic behavior of proteins, leading to highly accurate and quantitative understandings of the system.[57,58] Particularly in cytokine research, MD approaches have been limited, and have focused largely on either studying individual cytokine behavior[59–61] or speculating on mechanisms of binding between a cytokine and its receptors;[62–65] to the best of our knowledge, rigorous studies of an entire extracellular cytokine-receptor complex have been limited, with none of them leveraging simulation data to inform experimental protein design and engineering. With major systems of human health, including the immune system, signaling through the formation of protein complexes, a method that accurately models and predicts dynamic molecular behavior of these multicomponent systems would expand the abilities of researchers to study and engineer biology.

In this work, we combine molecular simulations and experimental techniques to study and engineer the type III interferon (IFNλ3) complex (IFNλ3/IFNλR1/IL10Rβ). Studies of IFNλs are of great interest in the field of mucosal and innate immunity because of their potential as therapeutics due to tissue-specific antiviral and anti-proliferative activity from restricted receptor expression,[9,23,66] as well as the many structural and functional questions that surround IFNλ1-4.[21,37,67–69] We first use the published structure of the engineered type III interferon complex (PDB ID 5T5W[38]), which contains an engineered, high-affinity IFNλ3 ligand (called H11) that increases the stability of the complex by 150-fold, to create a model of the predicted wild-type complex (WT) by computationally reverting the five mutant residues to the amino acids in the wild-type protein (**Figure 5A) (Appendix B, Supplemental Figure 4)**. We then perform

simulations of these two complexes to first determine how our molecular model compares to published experimental results as well as study differences which emerge based on the presence of these five mutations. We report that our models recapitulate differences between these two structures observed experimentally while also identifying key differences in PPIs, residue strain and fluctuation, and free energy landscapes. We then use this data to inform protein engineering and design to increase interactions between the low-affinity receptor in the complex, IL10Rβ, for IFNλ3. By making a single mutation at the hot spot identified from simulations, we stabilize and improve complex formation for IFNλ1-3 as measured by both yeast surface display and size exclusion chromatography, with the latter demonstrating the complex can be stabilized sufficiently to purify the complex, a step required for future structural studies of minimally engineered wild-type IFNλ1-4 complexes. Together, this study illustrates the use of molecular simulations as a powerful tool to study and engineer protein complexes and provides a readily applicable framework for use with other protein complexes in biology and immunology that are

difficult or inaccessible to experimentally study.

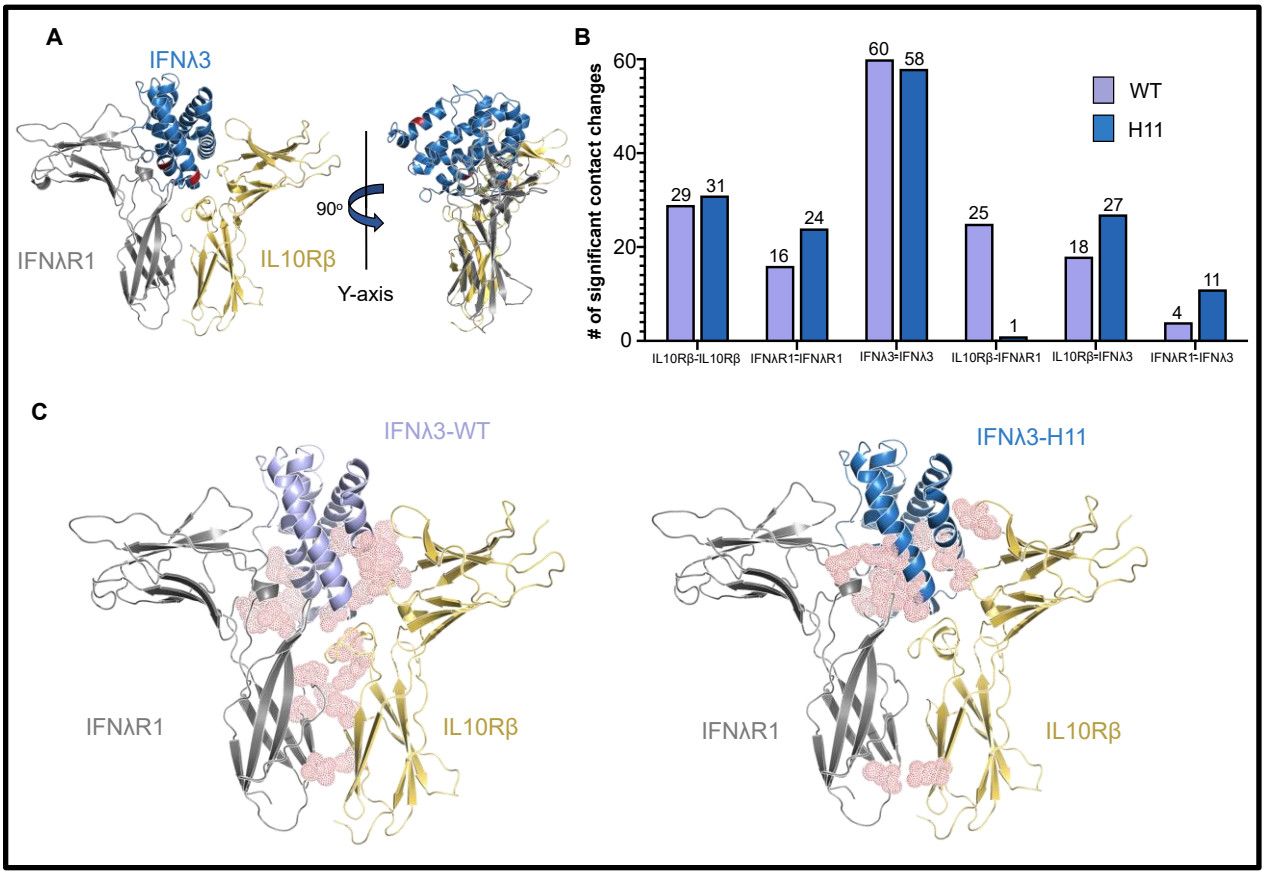


Figure 5: Molecular modeling and simulation of the IFNλ3 ternary complex reveals differences in protein-protein interactions (PPIs) between native and engineered ligands. A) An engineered IFNλ3, H11 (blue), was used to solve the structure of the type III interferon in complex with its receptors, IFNλR1(silver) and IL10Rβ (gold) (PDB ID 5T5W). Residues on the ligand mutated in the engineered protein are shown in red. Five amino acid mutations separate the wild-type IFNλ3 from the engineered H11. **B)** Residue contact patterns differ greatly between wild-type and engineered IFNλ3, leading to significant differences in protein-protein interactions throughout simulations. ($\Delta\tau > 0.25$). **C)** Hot spots of interactions (residue pairs with the highest values of $\Delta\tau$) between each set of protein chains are shown in red. Visualized interactions within protein chains can be seen in Appendix B, Supplemental Figure 9.

3.2 Materials and Methods

Molecular modeling

Simulated complexes were based on the solved protein structure of the ternary IFNλ complex (PDB ID 5T5W). For the engineered complex, structure 5T5W was completed using Chimera (UCSF)[70] and all additional molecules were removed. For the wild-type complex, the complete structure generated using Chimera was edited via PyMOL (Schrödinger, Inc.)[71] to replace the five engineered amino acids with their wild-type amino acids in the most probable rotamer conformation as determined by lack of steric clashing and minimization of predicted residue strain from the PyMOL rotamer library. A model for IL10Rβ-N147D was created similarly.

Molecular simulations

Classical MD simulations were performed in GROMACS 2020.5[72] with the AMBER ff99sb-ILDN* force field.[73,74] Production runs were performed for sufficient time to reach equilibration as measured by RMSD (**Appendix B, Supplemental Figure 5)**, with 3x100 ns replicates for the computationally modeled systems and 3x50 ns replicates for the studies starting from a known crystal structure. Approximately 35,000 TIP3P water molecules and 8 Na+ counterions were added to a dodecahedral box with a 10 Å distance from the protein complex to the box edge to ensure that the minimum image convention had been applied. The simulation protocol involved 50,000 steps of steepest descent minimization, followed by 500 ps NVT equilibration with position restraints on the protein complex with the velocity rescaling thermostat, then a 5 ns NPT equilibration with position restraints on the protein complex with the velocity rescaling thermostat and Berendsen barostat. Finally, production runs were made in the NPT ensemble for 100 ns with the Nosé-Hoover thermostat and Parrinello-Rahman barostat.

LINCS constraints were applied to bonds containing hydrogens, and long-range electrostatic interactions were calculated using the Particle Mesh Ewald method with a nonbonded cutoff of 12 Å. The strain analysis, principal component analysis (PCA), and contact analysis were performed on these trajectories.

To get a free energy profile, Gaussian-accelerated molecular dynamics (GaMD) simulations[75,76] were performed in AMBER20[77] with the AMBER ff14sb force field.[78] The converged structure from the unbiased GROMACS simulations were taken as inputs to the AMBER system and prepared using tleap. The GaMD protocol included an initial energy minimization, equilibration, and a 5 ns run in the NPT ensemble to relax the system in AMBER. Another 8 ns of conventional MD simulation were run to collect the potential statistics to calculate the GaMD boost parameters, followed by a 64 ns equilibration with the added boost potential and then three independent replicas of 100 ns GaMD production runs for both systems. The GaMD parameters used were a dual-boost scheme with one boost potential term applied to the total potential energy of the system, and the other boost potential applied to the dihedral potential energy, both of which set the reference energy to the lower bound. The average and standard deviation of the potential energies were calculated every 0.8 ns, and the upper limit of the boost potential standard deviation was set to 6.0 kcal/mol for both boost potentials. The 2D potential of mean force (PMF) for both systems was created by reweighting the combined trajectories of the replicas for each system using the PyReweighting toolkit.[79,80] After confirming sufficient sampling of the systems (**Appendix B, Supplemental Figure 6),** the PMF profiles were obtained by reweighting the system onto the RMSD of the IFNλR1 protein and the IFNλ3 protein (either engineered or wild-type), with a bin size of 1 Å and a cutoff of 500 frames.

Contact maps, contact analysis, and root-means squared fluctuation

To identify specific motifs and residues that impact the global conformation of the ternary complex, we applied contact maps and root mean squared fluctuations (RMSF) calculations using CPPTRAJ.[81] For contact maps, a cutoff distance of 10 Å was informed by previous studies[82] to capture sufficient C_α contacts. Native contacts were defined relative to an energy-minimized structure with a cutoff distance of 4 Å. RMSF was calculated with respect to a 100 ns averaged structure using the mass-weighted average over C_α, N, and C atoms and by residue.

Principal component analysis (PCA)

Principal component analysis (PCA) is a dimensionality reduction technique that can identify correlated motions in molecular dynamic trajectories and was performed as previously described[83]. Briefly, the low-dimensional subspace spanned by the first few principal components (PCs) represents the essential correlated conformational changes of the complex. PCs are evaluated by diagonalizing the correlation matrix C_{ij}, for coordinates i and j, $C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$, where $x_{1,\ldots,}x_{3N}$ are the mass-weighted atomic coordinates of the protein, averaged over all sampled trajectory structures. The highest proportion of variance within the dataset is accounted for by the eigenvectors with the largest eigenvalues and decreases sequentially while maintaining orthogonality to the first principal component. The "essential dynamics" (i.e. largest variance) of a simulation are contained within the first several PCs.

Strain analysis

The strain imposed by the mutations on the protein complex is a concept borrowed from continuum theory and applied to atomistic systems, wherein differential operators replace

derivatives.[84,85] First, a local neighborhood around each atom i is defined by a radius R, which contains n other atoms defined by $j = 1, 2, ..., n$. This is done for both a reference state and a deformed state. The instantaneous position of atom i in both the reference state and the deformed state is then denoted by $x_{0,i}$ and x_i, respectively. The deformation gradient $\boldsymbol{F}$ maps an infinitesimal segment at the reference configuration to one in the deformed configuration via the relation: $x_j - x_i = F(x_{0,j} - x_{0,i})$, and an optimized deformation gradient is found by minimizing the differences between the actual distances and the projected distances to an affine deformation:

$$min \sum_{j=1}^{n} ((x_j - x_i) - F(x_{0,j} - x_{0,i}))^2$$

The Lagrangian strain tensor is then found through $\varepsilon = 0.5(F^T F - I)$, where I is the identity matrix, and the magnitude is defined by the Frobenius norm of the strain tensor: $Tr(\varepsilon) - \frac{1}{3}(Tr(\varepsilon))^2$. In this analysis, the strain between pairwise frames of the wild-type (reference state) and engineered (deformed state) complex trajectories was calculated and averaged, only using C_α and choosing a radius r = 10 Å.

Protein expression, purification, and size exclusion chromatography

Proteins were expressed and purified as previously described[38]. The extracellular domain of IFNλR1 was expressed in HEK293 GnTi- cells (Gibco), and IFNλ3, IL10Rβ, and IL10Rβ-N147D were expressed in the Hi5 insect expression system (Lonza); both proteins were purified as previously described and stored in 10% glycerol. For size exclusion chromatography, equimolar ratios of wild-type IFNλ3, IFNλR1, and either IL10Rβ or IL10Rβ-N147D were combined and then purified by size exclusion chromatography on a Superdex S75 column (Cytiva) equilibrated with 1X HBS (20 mM HEPES, 150 mM NaCl, pH = 7.5). Elution volume

fractions were then run on an SDS-PAGE Stain-Free protein gel (Bio-Rad) and imaged to confirm the presence of individual proteins.

Yeast surface display of IFNλ1-3

Yeast surface display was performed as previously described.[38] Briefly, yeast displaying IFNλ1, IFNλ2, or IFNλ3 were first stained with an anti-Myc antibody conjugated to Alexa 647 (Cell Signaling, product #2233S) to determine protein expression and normalize fluorescence for protein display. Yeast displaying IFNλ1,2, or 3 were then incubated with 125 nM IFNλR1 and 1 μM biotinylated IL10Rβ or biotinylated IL10Rβ-N147D for 1 hour at 4ºC with gentle rotation. Phosphate buffered saline supplemented with 0.5% w/v bovine serum albumin (PBSA) was used to wash away unbound protein, and yeast were then resuspended with 0.5% PBSA + 1:100 streptavidin tetramers conjugated to Alexa 647 for 15 minutes at 4ºC with gentle rotation. Following another wash, yeast binding was assessed via flow cytometry. Fold increase was calculated by the increase of ternary complex formation relative to the wild-type IL10Rβ and normalized by expression. Statistical significance was determined using a 2-way ANOVA test (GraphPad Prism).

3.3 Results

Simulations reveal structural and behavioral differences between the engineered and wild-type IFNλ3 protein complexes

We first investigated if differences in residue contact can quantify differences in PPIs between the simulations; differences were identified by calculating the total simulation time as a percentage that a residue pair spent within contact distance (τ, r_{cutoff} = 10 Å) (**Appendix B, Supplemental Figure 7**) and then taking the difference between those pairs for both simulations to yield a value symbolizing the difference in contact time, $\Delta\tau$ (**Appendix B, Supplemental Figure 8**). A max score of 1 indicates that residues are only within 10 Å of each other throughout one simulation, and a minimum score of 0 indicates that the residue contact patterns are identical when comparing the wild-type and engineered simulations. To focus on residue pairs that experienced a more substantial difference in simulation contact, only $\Delta\tau$ values greater than 0.25 were considered significant. In total, 152 residue pairs were identified as significant using the criteria described above for both the H11 and WT complexes. When grouped into protein pairs, noticeable differences in global complex behavior are readily observable (**Figure 5B**). Interactions between residues within the proteins themselves were slightly different between simulations; total significant receptor-receptor interactions were slightly greater for the engineered complex, while total significant residue-residue interactions in the ligand were slightly greater for the wild-type. However, stark differences exist when comparing the interactions between the protein, with interactions between IL10Rβ-IFNλ3 and IFNλR1-IFNλ3 both favored for the engineered complex (27 to 18 and 11 to 4, respectively). Most noticeably, the differences in interactions between IL10Rβ and IFNλR1 are wildly different, with simulations identifying 25 significant residue pairs

between the two receptors during the wild-type simulations, but only one during the engineered simulations.

In addition to looking at PPIs, the residue-residue interactions themselves were analyzed so that potential areas or specific amino acids could be identified as may be important for complex interaction. From the 152 interactions identified, we identified the five most significant pairs (highest values of $\Delta\tau$) from each protein-protein interface (**Appendix B, Supplemental Table 4**) and visualized them on the protein structures to see where they clustered (**Figure 5C**) (**Appendix B, Supplemental Figure 9)**. These results are consistent with the engineering of H11, with significant contacts clustering at the receptor-ligand interface for the engineered simulations. Visualizing these differences also emphasizes the lack of interaction between the two receptors in the engineered complex and highlights the potential importance of IL10Rβ-IFNλR1 interactions in recognizing and interacting with the wild-type IFNλ3. Importantly, many of the key residues reported in the crystal structure of the engineered ternary complex were deemed significant by the MD simulations, including Tyr59 and Tyr82 on IL10Rβ for interaction with IFNλ3, which have been reported as contact hotspots for broad IL10Rβ interactions.[38] While these contact patterns reveal some new and previously unknown potential sites of interaction (particularly for the wild-type complex), the identification of these residues' importance in ligand recognition helps validate the simulations as reflecting the previously reported experimental results. Interestingly, while the 5 mutation sites that separate the wild-type and engineered IFNλ3 appear in this data set, their behavior between complexes is not as significant as other residues within the proteins but they may contribute to allosteric changes throughout the ligand.

The trajectory data from simulations were also used to quantify differences in strain fluctuation experienced by individual residues between the wild-type and engineered complexes.

Strain tensors were calculated and differences in total strain were converted to a B-factor and imposed on the protein structure to highlight areas of the proteins experiencing high differences in strain (**Figure 6**). Visually it is immediately clear that the regions of the complex experiencing high differences in strain are also the binding sites identified in the original crystal structure, indicating that high differences in strain may help explain interaction mechanisms of protein complexes. In a similar manner as above, when looking at the individual residues which appear as having high differences in strain, there are key residues as identified by the crystal structure. Side-chain to side-chain hydrogen bonds reported between IFNλR1 and IL10Rβ, including those formed by Gln163 on IFNλR1 and Tyr59, His128, and Thr142 on IL10Rβ, are found as residues experiencing high differences in strain, with residues flanking these spots also experiencing high strain. While differences overall were higher when comparing receptors, differences in strain experienced by residues on IFNλ3 were smaller in magnitude and mostly concentrated in the

binding sites as shown; again, the five mutations show some differences, but are not identified as the main drivers of the differences observed.

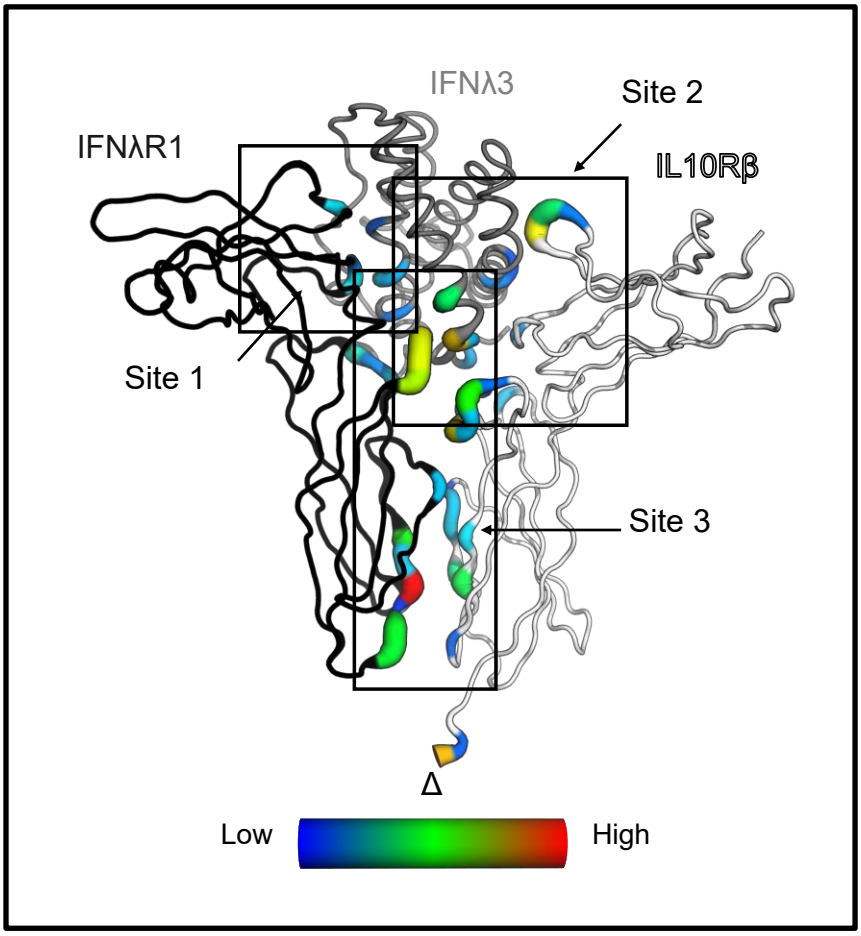


Figure 6: Residue strain fluctuations reveal sites of interaction known to drive protein complex formation. Residues experiencing high differences in strain ($\Delta\epsilon$) between the engineered and wild-type simulations align with known regions of interaction in the 5T5W structure (Site 1, IFNλR1-IFNλ3 interface: Site 2, IL10Rβ-IFNλ3 interface: Site 3, IFNλR1-IL10Rβ interface).

With the dramatic differences in residue contact seen between simulations, we then turned to see if these individual differences in contact were impacting the global behavior of the complexes. Using the simulation trajectory data, differences in global root-means squared fluctuation (RMSF) were calculated and revealed differences in RMSF near residues which differ

between the engineered and wild-type ligand (**Figure 7A**), with the region of amino acids leading up to the third mutation site experiencing the largest differences. To better visualize this difference in fluctuation, the primary modes from Principal Component 1 were visualized for both the wild-type and engineered complex (**Figure 7B**). The structural differences of these superimposed images are shown in white and reveal that the wild-type ligand has a highly flexible region, the same region identified to have large differences in RMSF, whereas the engineered ligand does not. This visualization led to questions about the behavior of this peptide region in the engineered

ligand, as that space could not be resolved in the crystal structure and seems to show a more ordered secondary structure when compared to the wild-type.

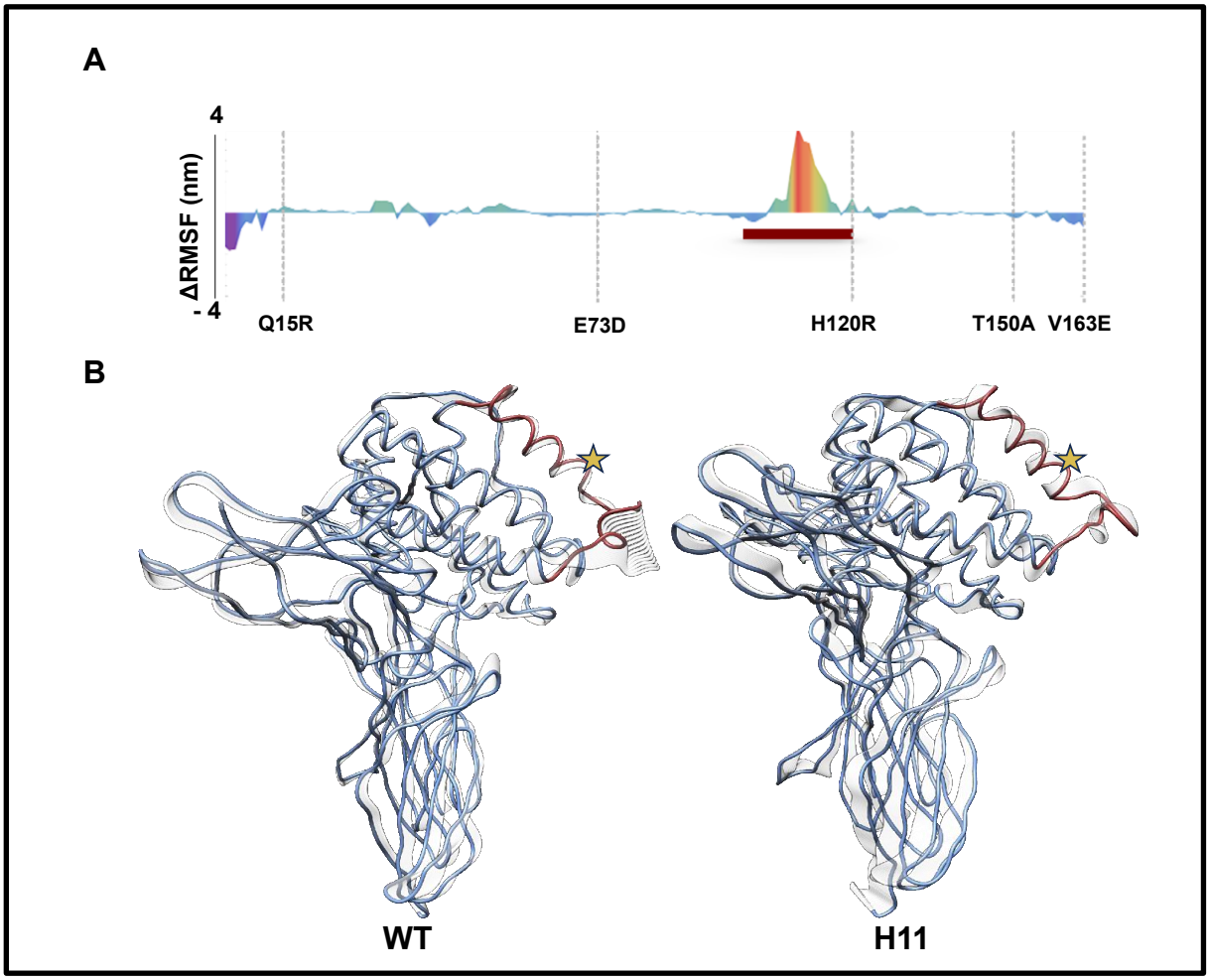


Figure 7: Mutations in H11 increase ligand stability by reducing chain fluctuations. A) Differences in root-means squared fluctuation (ΔRMSF) for the IFN*λ*3 protein between the wild type and engineered simulations are shown. The five mutations which separate the proteins are called out with grey lines, with a red bar indicating the region experiencing the largest differences**. B**) The primary modes of the first principal component, when superimposed, reveal a highly flexible region (red backbone) for the wild type IFN*λ*3. This same region remains stable within H11. (Star = location of H120 (WT) or R120(H11)).

To continue to probe these differences, we used Gaussian-accelerated molecular dynamics (GaMD) to better explore the full range of conformations for these complexes and

identify their relative changes in free energy. The resulting free-energy landscapes, as well as the corresponding coordinates and RMSDs of IFNλR1 and either the wild-type (WT) or engineered (H11) IFNλ3, highlight distinct behavioral differences between the two complexes (**Figure 8**). Throughout the WT GaMD simulations, the complex remains centered at its free energy minimum. The engineered complex, however, explores two significant conformations separated by a relatively high energy barrier. Looking particularly at the differences in ligand structure, as the H11 GaMD simulations progress from the lower-left coordinates (where it begins) to the upper-right coordinates (where it spends most of the simulation time), a secondary structure (highlighted in red) initially hinted at by the differences in RMSF is shown to be the conformation of the global free energy minima for the engineered complex. This secondary structure, again, is not observed at any point during the wild-type GaMD simulations, indicating

that its formation is unique to H11 and that the engineered complex achieves a global free energy minimum in this system when it is present.

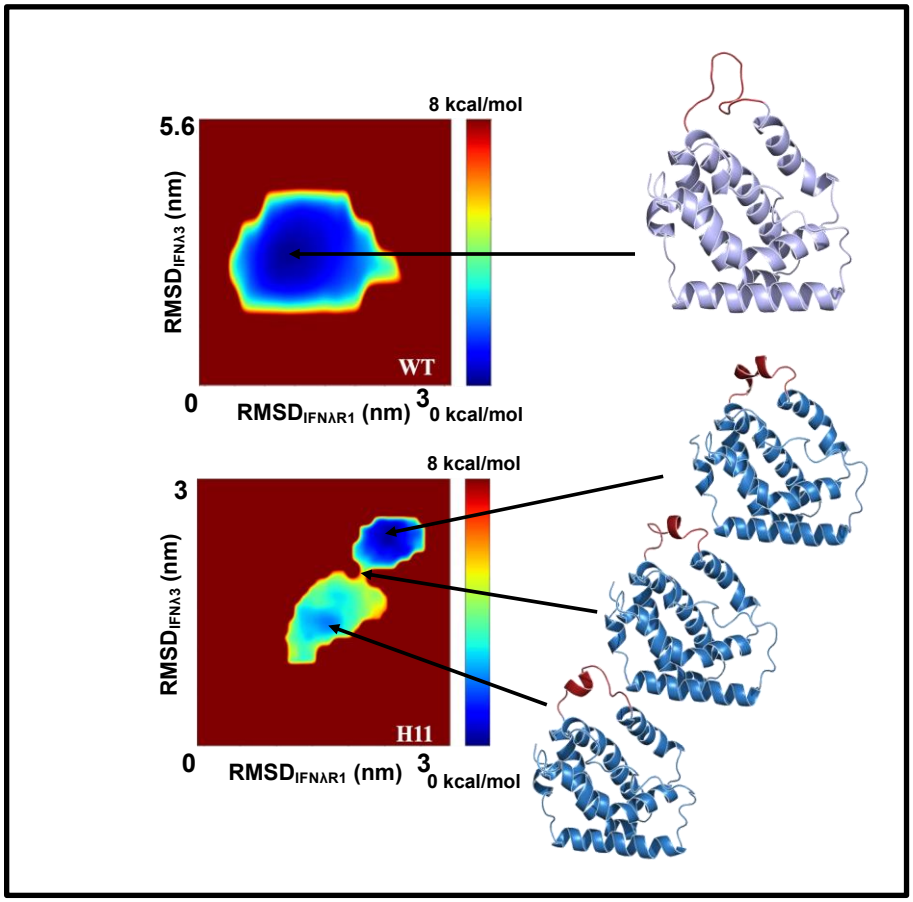


Figure 8: Gaussian-accelerated molecular dynamics simulations reveal that mutations in H11 help achieve complex stability through the formation of a secondary structure. Relative free energy minima generated from Gaussian-accelerated molecular dynamics simulations (GaMD) are plotted using the root-means squared deviation (RMSD) of IFNλ3 and IFNλR1. The engineered complex shows two distinct free energy minima, whereas the wild-type complex only shows one. As the engineered complex progresses through the simulation, the formation of a secondary structure is observed (in red).

The GaMD simulations also highlight how individual residues that differ between the wild-type and engineered ligand may lead to differences in binding affinity and are likely strong contributors to the structural differences observed during these simulations. Using the structural coordinates at the free energy minimums for the wild-type and engineered complexes, as well as

the apparent transition point for the engineered complex, we looked at how the 5 residues that differ between the two proteins were oriented at their respective free energy minimums. While three of the mutations - E73D, T150A, and V163E - do not vary greatly in their positioning or binding activity, Q15R and H120R show dramatic differences in their interactions within IFNλ3 and with IL10Rβ. As we visualize the conformational changes going from the initial structure of the engineered complex to its final free energy minima, we see that Arg15 actively rotates out from its original intramolecular interaction and down towards IL10Rβ, pulling up a loop in the receptor and initiating a large structural change through IFNλ3, leading to Arg120 rotating in towards the new secondary structure observed by GaMD (**Appendix B, Supplemental Figure 10**). The proposed existence of two free energy minimums for the engineered complex might explain the lack of resolution of this region in the crystal structure, and the formation of this secondary structure could also provide a structural explanation for the improved binding and signaling ability of H11 compared to the wild-type beyond the increased affinity for the receptors.

Design of an IL10Rβ to improve ternary complex stability for the entire type III IFN Family

After observing these distinct differences between the wild-type and engineered complexes in our simulations, we were interested to apply these insights to protein engineering and design. It was noted earlier that receptor-receptor interactions are significantly increased in the wild-type simulations, and that many residues experiencing high differences in strain occurred at that interface. Because engineering receptor-receptor interactions would also leave the ligand-receptor interactions unaltered, we chose to rationally engineer residues at this interface on the low-affinity receptor, IL10Rβ, to increase ternary complex formation. We identified the residue pair with the highest difference in contact time at the receptor-receptor interface for the wild-type simulations – proline 157 (Pro157) on IFNλR1 and asparagine 147 (Asn147) on IL10Rβ ($\Delta\tau$ = 0.5852) - and

hypothesized that rationally engineering Asn147 on IL10Rβ to aspartic acid (Asp147) to introduce a negatively charged residue at the junction of the receptor-receptor and ligand-receptor interfaces would introduce new polar interactions between IL10Rβ and the other proteins while maintaining the majority of the original amino acid structure.

We first ran simulations as previously described to gain a molecular understanding of the predicted differences between the protein complexes containing either WT IL10Rβ or IL10Rβ-N147D, this time using the model of the WT IFNλ3 complex as a template and introducing a single mutation on IL10Rβ to create a model of the mutant receptor. Contact analysis comparing the trajectories of the complexes again reveals significant differences in protein-protein interaction between the two (**Fig. 9)**. In total we found 91 total contacts revealed for IL10Rβ-N147D complex versus 63 for the wild-type IL10Rβ, with the single mutation made to IL10Rβ resulting in an almost 10-fold increase in significant contacts between the receptor and the wild-type IFNλ3. This mutation also seems to remove differences in IFNλR1 behavior between the simulations, with IFNλR1-IFNλR1 contacts being equal and low and no significant contacts between IFNλR1 and IFNλ3 in the simulations of IL10Rβ-N147D (**Figure 9A**). When visualized, these changes in protein-protein interactions reveal that hotspots of interaction cluster at the IFNλ3-IL10Rβ-N147D interface as a result of the rational mutagenesis **(Figure 9B)**. When analyzing the 20 most significant contact pairs for each simulation as well as the intra-protein interactions (**Appendix B, Supplemental Figure 11, 12; Supplemental Table 5)**, all 20 for the IL10Rβ-N147D implicate IFNλ3, either with itself (11 of 20) or with IL10Rβ (9 of 20), mostly clustering at the ligand-receptor interface. Specifically, the mutation at site 147 is identified as a significant contact for two residues on IFNλ3 – Lys10 ($\Delta\tau$ = 0.3867) and Leu12 ($\Delta\tau$ = 0.3688) – and site 148 is also identified as a significant contact for two residues on IFNλ3 – Lys10 ($\Delta\tau$ = 0.4291) and Phe9 ($\Delta\tau$

= 0.3215) – highlighting that these residue-residue interactions between the engineered site on IL10Rβ and residues 9-12 on IFNλ3 are closer to each other during simulation time for the engineered complex.

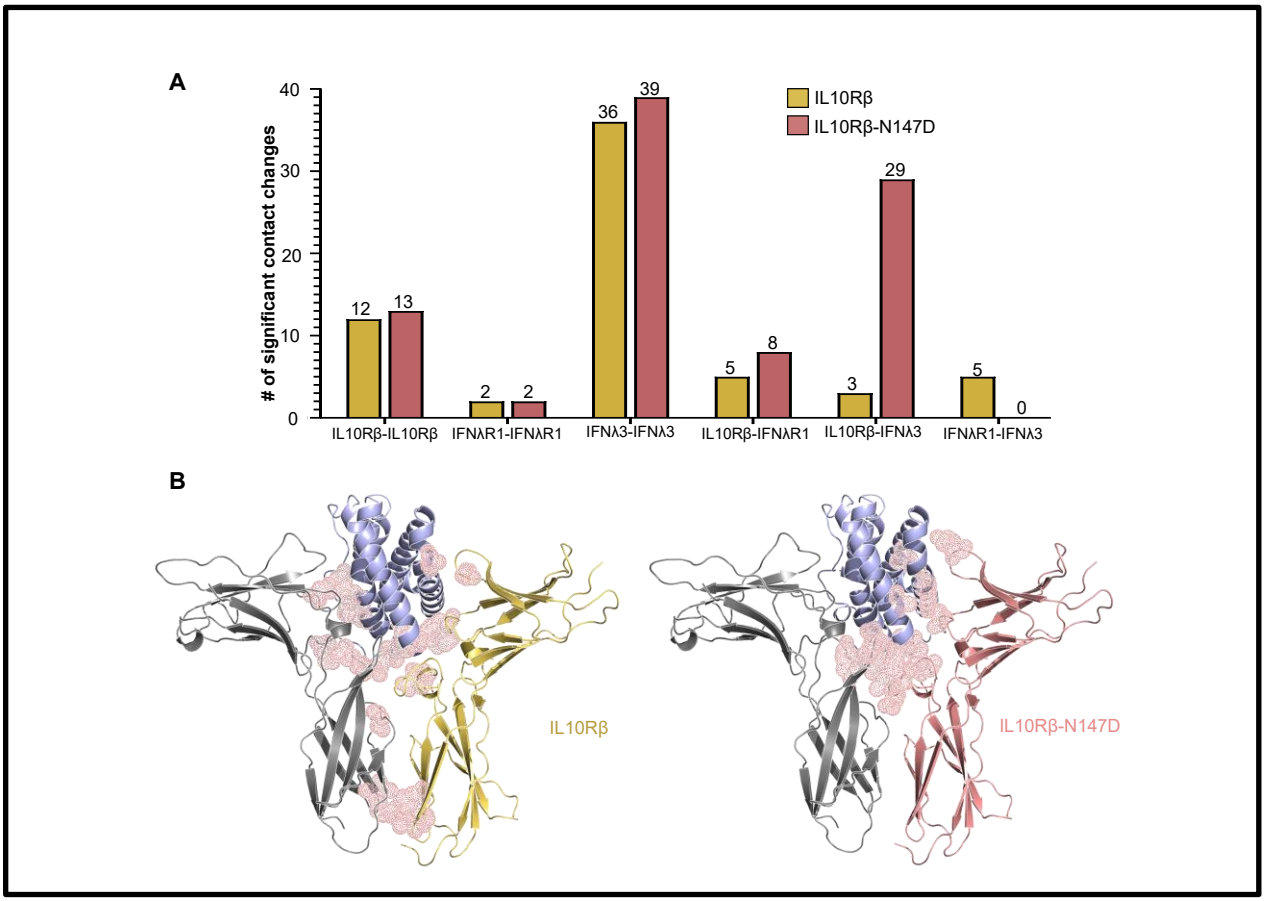


Figure 9: Mutation on IL10Rβ identified by molecular simulation leads to an increase in predicted protein-protein interactions (PPIs) between IL10Rβ and wild type IFNλ3. A) Residue contact patterns differ greatly between wild-type and engineered IL10Rβ, biasing the system toward receptor-ligand contacts. ($\Delta\tau > 0.25$). **B)** The top 5 residue-residue interactions occurring between proteins are shown in red, demonstrating a shift in protein-protein interactions occurring at the IL10Rβ-IFNλ3 interface.

To validate these simulations, we next leveraged these insights to experimentally determine if this single mutation could increase the ability to form a complex relative to the native type III interferon complex. To begin, yeast displaying IFNλ1-3 were incubated with IFNλR1 and either wild-type or engineered IL10Rβ and their relative fluorescence upon complex formation was

measured (**Figure 10A) (Appendix B, Supplemental Table 6)**. Compared to the wild-type IL10Rβ, IL10Rβ-N147D improves the formation of the ternary complex for all members of the type III interferons that display on yeast. While the difference is most pronounced for IFNλ3 (2.56-fold increase vs. wild-type), there is an improvement for binding to IFNλ2 (1.26-fold) and this mutation also appears to rescue binding of IFNλ1 (2.29-fold), indicating that this residue may be an important interaction for general formation of the type III interferons complexes. To further validate this improved binding, we compared the elution profiles of a mixture of wild-type IFNλ3, IFNλR1, and either IL10Rβ or IL10Rβ-N147D processed via fast protein liquid chromatography (FPLC) (**Figure 10B**). In keeping with previous reports, we do not observe a stable ternary complex, indicated by the co-elution of the three proteins, when the wild-type IL10Rβ is used (**Appendix B, Supplemental Figure 13)**. By contrast, the ternary complex co-elutes when IL10Rβ-N147D is used. The ability to purify a complex for structural studies using this approach enables solution of valuable experimental structures in a more direct way than traditional protein engineering techniques. Together, this computational and experimental data show that increasing PPIs within the ternary complex is the driving factor for increased binding observed both on yeast and over FPLC. The amino acid sequence of the IFNλ proteins in the region shown in simulations

to contact IL10Rβ-N147D (xFKSLSx) is also conserved in IFNλ1-3, which may explain the recognition of this receptor for the other type III interferons.

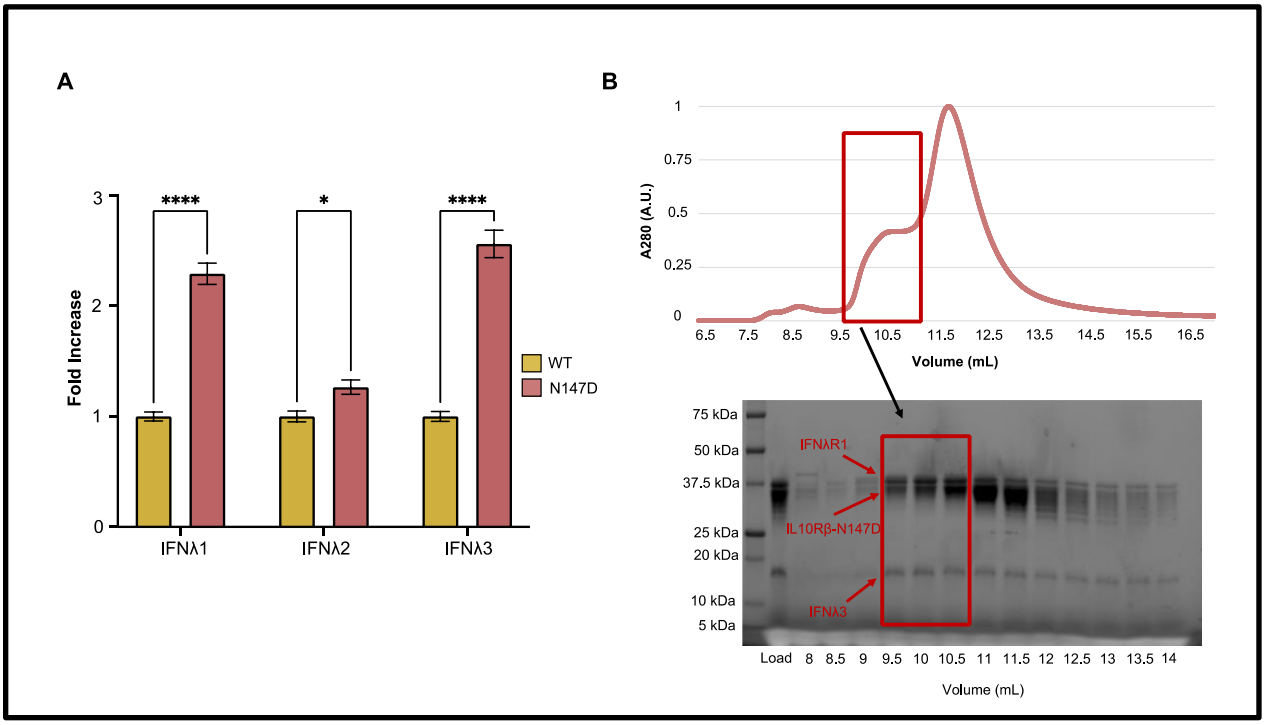


Figure 10: Computationally designed IL10Rβ shows improved binding for all type III interferons and enables purification of the wild-type IFNλ3 ternary complex. A) Fold increase in binding data for yeast displaying IFNλ1, IFNλ2, or IFNλ3 in complex with IFNλR1 and either wild-type IL10Rβ (WT) or IL10Rβ-N147D (N147D). (N = 3, plotted with SEM, statistical significance determined by 2-way ANOVA). **B)** Size exclusion chromatogram (Superdex S75 column) of the wild-type IFNλ3/IFNλR1/IL10Rβ-N147D ternary complex (top). Volume fractions are shown on the SDS-PAGE protein gel (bottom).

3.4 Discussion

With this initial molecular-level study, we replicate and confirm the experimental differences observed in the previous studies of the IFNλ complex while providing molecular details and explanations for the differences observed - in this way, molecular modeling serves to visualize the behavior of these protein complexes over time to generate new structural and behavioral data. Some of these important differences are consistent with the experimental data despite being generated *de novo*; for example, the shift from receptor-receptor residue contacts in the simulations of the wild-type to receptor-ligand interactions in the simulations of the engineered ligand, is consistent with the selection and engineering method used to affinity-mature IFNλ3 for its receptors. Observing that the areas experiencing high differences in individual residue strain also align with the identified sites of interaction for the IFNλ ternary complex and highlight that molecular modeling and simulations can be used to recreate experimental data for interferon complexes. Furthermore, this suggests that residues experiencing large differences in strain throughout simulations are potential areas of important protein-protein interactions that may be focused upon for engineering, which provides useful insights into how other immune protein complexes come together to signal. To this end, we believe that these computational approaches represent a valuable tool for studying protein complexes, specifically ones that cannot be readily obtained experimentally or when recombinant expression of the protein is difficult. These methods can also provide a more rapid and cost-effective means of assessing the impact of individual mutations on protein complex behavior when compared to methods using predicted differences in free energy that, while powerful, often do not completely align with experimental values.[86,87]

Because we cannot obtain a structure of IFNλ3 in complex with its wild-type receptors due to its low overall affinity, the structural details revealed by the simulations cannot be validated experimentally. In one possible application of this data, though, mutants generated from these simulations may help guide efforts to crystallize the wild-type complex. By engineering the amino acid residue Asn147 at the IL10Rβ-IFNλR1 interface to Asp147, we have increased the total binding of the ternary complex on yeast while minimally altering the composition of IL10Rβ and have confirmed that stable ternary complex formation is improved while demonstrating a link between simulation and experimental data. This broad recognition of this mutant receptor for all 3 IFNλs tested after only one mutation indicates that other interactions identified by simulation may further increase complex affinity, specifically at the receptor-receptor interface, and that this mutation at position 147 may be critical for general ligand and receptor stability by IL10Rβ. Further mutagenesis at different sites at the IL10Rβ-IFNλR1 interface could result in a super high-affinity IL10Rβ that can enable the determination of all four type III interferons without dramatically altering the ligand-receptor interface. Of important note is that due to its biochemistry, IFNλ4 does not display on yeast or express well in any known expression system, greatly limiting structural and functional characterization. While this goes beyond the scope of this current study, molecular simulations may hold great promise for studying IFNλ4 by elevating previous attempts to explain its behavior by molecular modeling alone.[88]

Rational mutagenesis was chosen as the method for this approach to minimally alter the composition of the residue on IL10Rβ, but this method could be extended further by performing site-saturated mutagenesis at locations identified as important by the simulations to tune or

modulate PPIs, yielding more focused protein libraries and drastically lowering the theoretical diversity to easily achievable experimental values. This more rational design of libraries can be readily applied for engineering new protein-based therapeutics informed by physical and dynamic modeling, including cytokines, interferons, or other protein binders like antibodies which exist in complex with multiple proteins and benefit from either affinity maturation or biased design for partial agonism. In the past, this approach has required extensive structural or experimental studies,[13,38,39] but a simulation-based approach like the one taken here enables rapid, accurate physical predictions for the design of improved versions of proteins that are currently too weak or too potent for therapeutic use. As molecular modeling and computational biology techniques continue to improve, the two-pronged approach to protein engineering and design outlined in this study holds great promise, and further work will be done to validate this method with other protein complexes or families such that this workflow can be broadly applied outside of immunology.

3.5 Conclusion

Our work describes molecular dynamics simulations of the type III interferon complex, yielding differences in trajectory and free energy behavior for a high-affinity and low-affinity protein complex, and its applications in protein design and engineering. To our knowledge, this study is the first to simulate the type III interferon complex, and one of the first to rigorously simulate and analyze a complete extracellular cytokine-receptor complex, producing some of the most extensive all-atom molecular dynamics simulations and analysis performed on a cytokine-receptor complex to date. This method of study also implements many powerful tools useful for studying and simulating protein complexes in general and then leveraging the data generated from simulation to guide protein design and engineering, including analyses of residue contacts, strain, fluctuation, and free energy behavior. Ultimately, this combination of simulation and experiments paints a more complete picture of the structural differences between the wild-type and engineered IFNλ3, how type III interferons interact with their receptors, and could be further applied to develop more effective IFNλ-based therapeutics using molecular dynamics simulations as a key technology.

CHAPTER 4: The structure of the IFNλ4 ternary complex determined by cryogenic electron microscopy

4.1 Introduction

In Chapter 2, a robust method for IFNλ4 expression and purification has been presented, breaking through the expression bottleneck that has limited structural studies of IFNλ4. Armed with sufficient quantities of IFNλ4 protein and an improved mechanistic understanding of the type III IFN complex detailed in Chapter 3, we moved our attention towards solving the IFNλ4 receptor complex.

Functional comparisons of IFNλ3 and IFNλ4 have become increasingly relevant, specifically in the field of hepatitis C research, due to an increasing body of work identifying their differential expression and associated SNPs as impactful for viral progression in patient populations.[89–91] While the impact of these genetic differences continues to be characterized, fewer studies have focused on the extracellular functions of these proteins, and none have investigated the impact of structure and receptor engagement on these findings. In general, and in keeping with our findings reported in Chapter 2, IFNλ3 and IFNλ4 appear to elicit similar enough functions in their extracellular signaling abilities, though the fact that they signal through the same receptor complex yet invoke different responses at all points to potential differences in receptor engagement. While extensive characterization of the extracellular functions of IFNλ4 is enabled by our improved protocol for IFNλ4 expression, a structural basis for differential IFNλ3 and IFNλ4 activity could provide mechanistic insight to the differences observed both *in vitro* and *in vivo*.

Like other protein complexes involving IL10Rβ, the native affinity of IL10Rβ for the IFNλ complex is too low to enable isolation of the ternary complex for structural studies. Previous attempts to overcome this limitation have found success by leveraging directed evolution to engineer a high-affinity ligand to enhance overall complex affinity;[13,49] however, as

previously discussed, this approach is not viable for IFNλ4 for presumably the same reasons limiting its expression. The previous IL10Rβ engineered using molecular simulations in Chapter 3 improved ternary complex stability but was not sufficient to enable complete isolation of the IFNλ3 complex, indicating that further engineering would be required for structural studies of any IFNλ complex. To improve upon this approach, we elected to perform directed evolution on IL10Rβ to not only improve the engagement of the low-affinity receptor but to also enable structure determination of both the IFNλ3 and IFNλ4 ternary complexes, allowing for direct comparisons of the two related ligands solved using the same methods and under the same conditions.

In this Chapter, we present the structure of IFNλ4 as well as IFNλ3 in complex with IFNλR1 and an engineered, high-affinity IL10Rβ termed A3, via cryogenic electron microscopy (cryoEM). By comparing these two structures, we gain mechanistic insights into the extracellular engagement of IFNλ4 by its two receptors and highlight the structurally conserved regions between IFNλ3 and IFNλ4, resolving questions about the extracellular signaling mechanism of IFNλ4. We elevate our understanding of these complexes using molecular dynamics simulations to study the dynamics of the ligands and identify a disordered and non-receptor engaging region of IFNλ4, lending a structural explanation for the IFNλ4 "sticking" that has been well-characterized and observed. In depth structural analyses of these complexes also highlight changes in the hydrophobic network of residues on IL10Rβ that engage IFNλ3 and IFNλ4 and identify distinct conformational rotation of the IL10Rβ receptor that potentially explain the *in vitro* functional differences of the two proteins presented in Chapter 2 and highlight the role of receptor geometry in differential ligand signaling and activity. In total, this Chapter presents the molecular details of IFNλ4 signaling and receptor engagement and enables structure-informed

approaches to overcome the issues associated with IFNλ4 expression in afflicted patient populations.

4.2 Methods and Materials

Chemicals, materials, and reagents

All chemicals and reagents were acquired from Thermo Fisher Scientific or Millipore Sigma unless otherwise specified.

Molecular cloning and genetic constructs for protein expression

All relevant genetic constructs used have been previously detailed in Chapter 2.

Cell lines and cell culture

Methods for cell lines and culture are explained in detail in Chapter 2. IL10Rβ-A3, IFNλ3, and all IFNλ4 protein constructs were expressed using insect Hi5 cells (Lonza). IFNλR1 was expressed in HEK293 GnTi- cells. Baculovirus containing genes of proteins of interest were transfected and grown using Sf9s cells according to previously described protocols.

Protein expression and purification

Methods for protein expression and purification are explained in detail in Chapter 2. Expression of IL10Rβ-A3, IFNλ3, and IFNλR1 were expressed as previously described using baculovirus.[40] These proteins were purified using FPLC in 1X HEPES-buffered saline (HBS, 20 mM HEPES + 150 mM NaCl, pH = 7.5) and flash-frozen with 10% glycerol to protect samples. Expression of IFNλ4 is explained in great detail in Chapter 2.

Yeast surface display and engineering of IL10Rβ

General yeast surface display protocols for quantification of protein binding and for directed evolution were performed as previously described.[40,92] Appropriate supplementary details for yeast surface display are provided in Chapter 3. To engineer IL10Rβ for higher

affinity towards IFNλR1 and IFNλs, yeast displaying IL10Rβ were first stained with an anti-Myc antibody conjugated to Alexa-647 (Cell Signaling) to confirm protein display. Details on selection and engineering are presented in the results section of this Chapter.

For yeast binding experiments, yeast displaying either IL10Rβ or IL10Rβ-A3 were stained to confirm protein display and normalize fluorescence for future data collection. Equimolar ratios of IFNλR1 and either IFNλ3 or IFNλ4 were then titrated over a range of concentrations to determine relative bindings behaviors as well as comparative on-yeast binding constants.

Electron microscopy sample preparation

The IFNλ4 complex was assembled by mixing purified IFNλ4, IFNλR1, and IL10Rβ-A3 at a 1:1:1.1 molar ratio. The IFNλ3 complex was assembled in a similar manner. The mixture was incubated on ice for 30 minutes before being loaded onto a Superdex S75 column (Cytiva) equilibrated with either 1X HBS (pH = 8.3) + 500 mM NaCl for the IFNλ4 complex or 1X HBS (pH = 7.5) for the IFNλ3 complex. Peak fractions containing 0.1 mg/mL complex were taken directly off the column and used for grid preparation; it was crucial to avoid concentrating the protein complexes, especially IFNλ4, to prevent aggregation or loss by sticking to the concentrator membrane. Complex treated with Endoglycosidase F and H was also tested to remove excess glycans on IFNλR1, however it was found that this process resulted in severely aggregated complex.

Sample vitrification was performed using a Vitrobot Mark IV (Thermo Fisher Scientific) operating at 22 °C and 100% humidity. A 3.5-µl sample was applied to holey carbon grids (UltrAuFoil 300 mesh Au 1.2/1.3) that had been glow-discharged for 30 seconds. The grids were

blotted for 4 seconds at a 'blotting force' 0 by standard Vitrobot filter paper (Ted Pella, product #47000-100) and were then plunge-frozen in liquid ethane.

CryoEM data collection

Frozen grids were sent to the Advanced Electron Microscopy Facility at the University of Chicago for data collection. The dataset was acquired as movie stacks using EPU (Thermo Fisher Scientific) installed on a Titan Krios transmission electron microscope operating at 300 kV and equipped with a BioQuantum K3 imaging filter (Gatan). Images were recorded at a nominal magnification of ×81,000 and super-resolution counting mode by image shift. The total exposure time was set to 8 seconds with 40 frames in a single stack and a total exposure of around 60 electrons $Å^{-2}$. The defocus range was set at −0.9 μm to −2.3 μm. Detailed parameters for cryoEM data collection are discussed in the supplemental information **(Appendix C, Supplemental Table 7)**.

CryoEM image processing

Raw images were imported to a CryoSPARC live session for contrast transfer function (CTF) determination, motion correction, and particle picking in order to solve the IFNλ4 **(Appendix C, Supplemental Figure 14A,B)** and IFNλ3 **(Appendix C, Supplemental Figure 15A,B)** complexes. Particles were automatically picked using two-dimensional (2D) class averages as templates, which were generated from blob picking. The extracted particles were imported to cryoSPARC for further processing. To increase the number of good particles, Topaz train and Topaz extraction were also performed based on well-resolved 2D class averages. The extracted particles were then combined with particles from blob picking. After additional 2D

classification with non-overlapping particles, contamination and poorly-aligned classes were disposed. The resulting particles were used to generate three initial models by *ab initio* reconstruction. Hetero refinement was then performed in CryoSPARC[93] using the three initial models as the starting points. The coordinates of the particles from the best class were imported into RELION 5.0 beta[94] for particle re-extraction. Contrast Transfer Function (CTF) estimation and motion correction were independently performed in RELION, with one round of both 2D and 3D classifications performed using the map generated from cryoSPARC as the initial map. RELION 5 Blush was applied throughout the process. The best class was subjected to 3D refinement, CTF refinement, Bayesian polishing, and post-processing. The final map of the IFNλ4 complex was resolved at 3.26 Å based on a standard acceptance criteria of a Fourier shell correlation (FSC) value of 0.143 (**Appendix C, Supplementary Figure 14C**). The final map of the IFNλ3 complex was resolved at 3.00 Å to identical FSC standards (**Appendix C, Supplementary Figure 15C**).

CryoEM model building, refinement, and validation

IFNλ4 model building was performed using a starting model of IFNλ4 complex predicted by AlphaFold2. Rigid body refinement was first performed in PHENIX[95] to adjust relative orientation of the chains from AlphaFold2 model, followed by real space refinement in Coot.[96] The previously solved ternary structure of IFNλ3-H11/IFNλR1/ IL10Rβ (PDB ID 5T5W) was used as the IFNλ3 starting model. The final model was refined in real space and validated using PHENIX. Molecular graphics were prepared using UCSF ChimeraX[70] and PyMOL v2.5.[71] Detailed statistics of model refinement and geometry are available in the supplemental information (**Appendix C, Supplemental Table 7)**.

Structure and sequence analysis

Structural alignments, models, and figures were generated in PyMOL. Sequence alignment of IFNλ4 and IFNλ3 was performed using Jalview.[97] Buried surface area calculations as well as analysis of structural interfaces were performed using PISA.[98]

Molecular modeling and simulations

Molecular modeling and simulations were performed as described in Chapter 3.[40] Briefly, classical MD simulations were performed in GROMACS 2020.5[72] with the AMBER ff99sb-ILDN* (Best 2009, Lindorff-Larsen 2010) force field.[73,74] Production runs were performed for sufficient time to reach equilibration as measured by RMSD (**Appendix C, Supplemental Figure 16)**, with 3x100 ns replicates for all simulations.

The unresolved regions of the IFNλ4/IFNλR1/IL10Rβ-A3 complex were first modeled as informed by cryoEM data and completed in Chimera using AlphaFold. A molecular model of the IFNλ4/IFNλR1/IL10Rβ complex was then generated from the structure of the IFNλ4/IFNλR1/IL10Rβ-A3 complex by editing the three engineered residues to their wild-type amino acids using PyMOL, with the most probable rotamer conformations determined by lack of steric clashing and minimization of predicted strain from the PyMOL rotamer library.

Residue contact analysis, root mean squared fluctuation (RMSF) calculations, and ion tracking

Residue contact analysis and root mean squared fluctuation (RMSF) calculations were performed as described in Chapter 3. Ion analysis was performed using the MDAnalysis Contact module.[99,100] Production runs were analyzed by counting each frame in which any ion was in

direct contact with a residue (within 1 Å of a residue's center of mass), expressed as a percentage calculated by dividing over the total number of frames.

4.3 Results

Engineering a high-affinity IL10Rβ for high-throughput structure determination of type III IFN complexes

We used yeast surface display to engineer a high-affinity IL10Rβ. A library of 5 x 10^7 clones was generated using error-prone PCR, and yeast displaying IL10Rβ were subjected to multiple rounds of selection **(Fig. 11A)**. The selection strategy involved first performing four rounds of selection with IFNλ3 as the ligand and decreasing concentrations of biotinylated IFNλR1 (400 nM tetramers to 100 nM monomer) as the selection pressure to intentionally select for versions of IL10Rβ with increased affinity at the receptor-receptor interface. The final two rounds of selection used refolded IFNλ4 as a ligand (500 nM, 100 nM), a gift from Ludmila Prokunina-Olsson at the National Cancer Institute (NCI), to ensure that the engineered receptor was capable of recognizing these related but sequence-dissimilar proteins. 96 clones from this final selection were screened for binding to IFNλ4 and IFNλR1, and the highest affinity clone (referred to as A3) was selected after demonstrating greatly enhanced affinity and complex formation on-yeast relative to the wild-type IL10Rβ for both IFNλ4 and IFNλ3 (**Fig. 11B)**. Sequencing of A3 showed three mutations at residues 147, 148, and 184, with residues 147 and 148 located on Loop 5 of IL10Rβ near the ternary complex junction **(Fig. 11C)**. Interestingly and excitingly, the mutation at residue 147 was the same mutation (Asn147Asp / N147D) identified by molecular dynamics simulations to inform design of a type III IFN complex with enhanced ternary complex stability via increased protein-protein interactions (PPIs).[40] Molecular dynamics simulations of the IFNλ4 ternary complex with IL10Rβ-A3 compared with a model of the complex containing the wild-type IL10Rβ reveal that the mechanism for this increased complex stability can be explained by enhanced PPIs at the receptor-receptor interface as well as the

IFNλ4-IL10Rβ interface, consistent with the method of engineering and selection **(Fig 11D,E**), with the engineered Asp147 residue being in closer proximity to both IFNλ4 and IFNλR1 relative to the wild-type complex throughout simulation time (**Appendix C, Supplemental Figure 17) (Appendix C, Supplemental Figure 18) (Appendix C, Supplemental Figure 19**). In line with the yeast data, IL10Rβ-A3 was able to form a stable ternary complex by gel filtration chromatography with both IFNλ3/IFNλR1 and IFNλ4/IFNλR1, enabling structural studies to move forward (**Appendix C, Supplemental Figure 20)**.

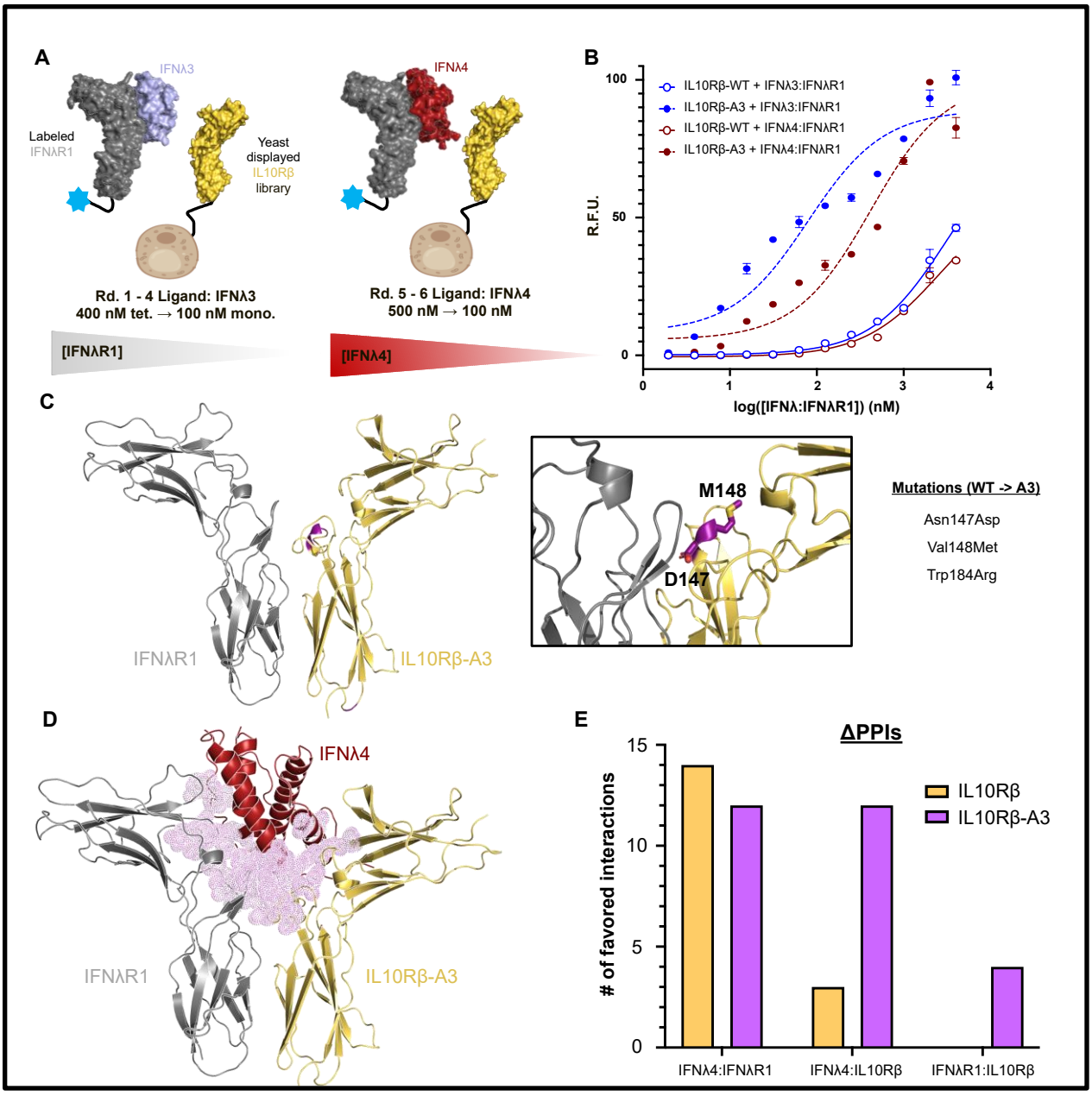


Figure 11: Engineering a high-affinity IL10Rβ for high-throughput structure determination of type III IFNs. A) Wild-type IL10Rβ was displayed on yeast and subjected to six rounds of selection to increase the overall affinity of the ternary protein complex. **B)** Yeast displaying either IL10Rβ (WT) or IL10Rβ-A3 (A3) titrated with equimolar ratios of either IFNλ3 or IFNλ4 with IFNλR1. (n = 2, data plotted with SEM). **C)** Mutations N147D and V148M are located at the interface of the three proteins (inset, in purple). W184R is locating towards the C-terminus of IL10Rβ-A3. **D)** Hotspots of interaction created by the engineered receptor IL10Rβ-A3 for the ternary IFNλ4 complex, visualized in purple. **E)** Differences in protein-protein interactions (PPIs) for the engineered IFNλ4 complex relative to a model of the IFNλ4 complex with the wild-type IL10Rβ.

Comparative structural analysis of the IFNλ4 and IFNλ3 ternary complexes

To facilitate structural determination, an N-glycan minimized version of the IL10Rβ-A3 protein was used as previously described.[38] Previous structural studies of related complexes relied on X-ray crystallography;[38,45] however, the relatively low yield of IFNλ4, its tendency to aggregate during concentration, as well as the conditions required to purify it became impassable barriers for large-scale crystallization trials. The ternary complexes of IFNλs with their receptors are around 70 kDa, making them challenging but exciting samples for use in single-particle cryoEM.[101,102] Through extensive optimization of sample preparation, we successfully determined both ternary complexes using single particle analysis to final resolutions of 3.26 Å for IFNλ4 and 3.00 Å for IFNλ3 (**Fig. 12A,E**). In both data sets, approximately 5% of particles from the initial particle picking contributed to the final reconstruction. A workflow combining CryoSPARC and RELION was adopted for data processing as described more thoroughly in the methods section of this Chapter, with the newly implemented data-driven regularization in RELION 5.0 beta proving to be crucial in reducing artifacts from preferred orientations and further enhancing the quality of the final maps (**Fig. 12B,C,F,G)**.

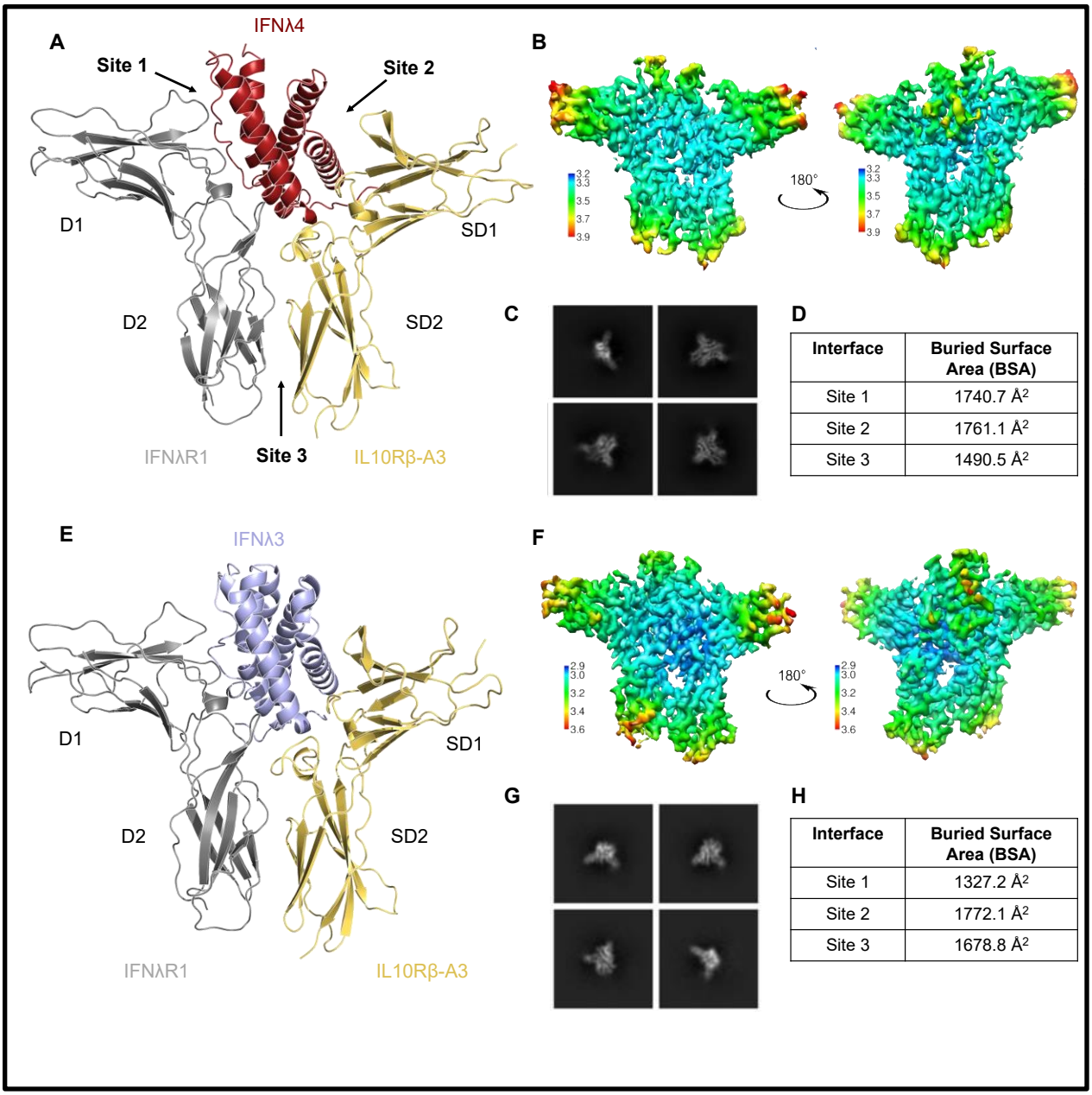


Figure 12: The structures of the IFNλ4 and IFNλ3 ternary complexes solved by cryoEM. **A)** Ternary IFNλ4/IFNλR1/IL10Rβ-A3 complex resolved to 3.26 Å. **B)** Per-residue resolution of the ternary IFNλ4 complex. **C)** 2D particle classes used in model building of the IFNλ4 ternary complex. **D)** Table highlighting the buried surface area (BSA) of protein-protein interactions (PPIs) present in the ternary IFNλ4 complex. **E)** Ternary IFNλ3/IFNλR1/IL10Rβ-A3 complex resolved to 3.00 Å. **F)** Per-residue resolution of the ternary IFNλ3 complex. **G)** 2D particle classes used in model building of the IFNλ3 ternary complex. **H)** Table highlighting the BSA of PPIs present in the ternary IFNλ3 complex.

Overall, the structures of the IFNλ4/IFNλR1/IL10Rβ complex and the IFNλ3/IFNλR1/IL10Rβ complex are similar to the previously solved ternary complex of an engineered, high-affinity IFNλ3 (H11) with IFNλR1 and IL10Rβ ($RMSD_{H11\text{-}IFN\lambda3}$ = 0.614 Å, $RMSD_{H11\text{-}IFN\lambda4}$ = 1.728 Å, $RMSD_{IFN\lambda3\text{-}IFN\lambda4}$ = 1.417 Å). The regions of contact between IFNλR1 and residues involved at the interface are largely similar for the two complexes. Both IFNλ3 and IFNλ4 make extensive contact with IFNλR1 and both interfaces highlight the ability of the receptor to engage with basic residues. In total, only 38 residues are conserved by the receptor binding domains of IFNλ3 and IFNλ4, with the greatest conservation occurring at the IFNλR1 interface on helices A and F **(Appendix C, Supplemental Figure 21)**. These chains find themselves in closer proximity to IFNλR1 on IFNλ4, likely due to the higher density of surrounding non-polar residues on IFNλ4. In turn, this leads to a noticeable difference in buried surface area (BSA), with the IFNλ4-IFNλR1 interaction higher than the IFNλ3-IFNλR1 by about 30% (**Fig. 12D,H).** IFNλ4 forms more polar contacts at this interface, with eight compared to four for IFNλ3, featuring unique involvement of the Ser185 residue on IFNλR1 forming three hydrogen bonds with residues Pro37, Arg38, and Ala41. Asp98 on IFNλR1 interacts with Arg47 of IFNλ4 and Lys21 of IFNλ3; while these two residues are not identical in their location on the ligand, the positively charged groups of the amino acids find themselves at similar locations at the interface.

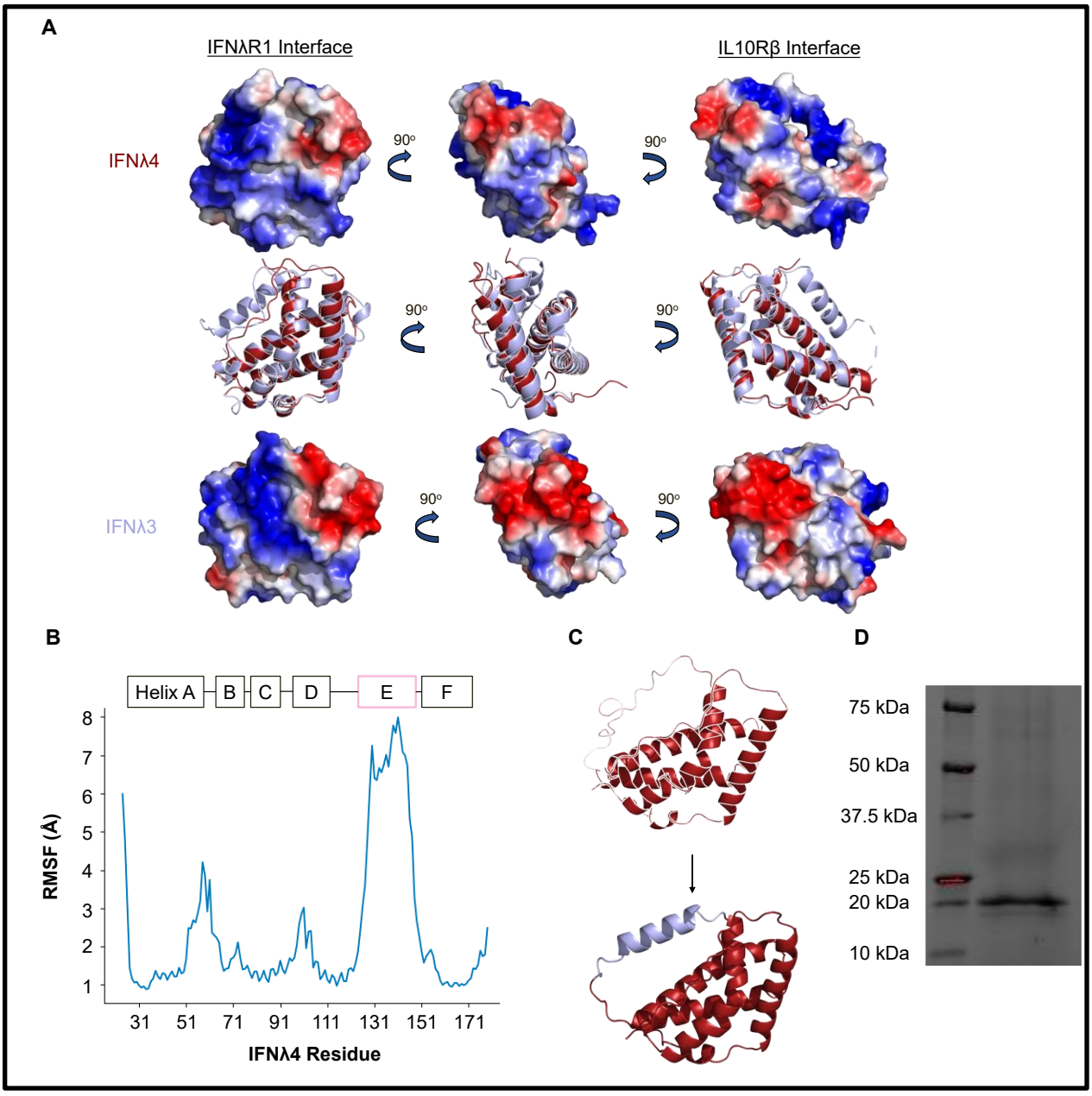


Figure 13: Structural and dynamical analysis of IFNλ3 and IFNλ4 reveals a structural basis of IFNλ4 aggregation. A) Differences in charge distribution and structure are shown for IFNλ4 (top) and IFNλ3 (bottom). White = non-polar residue, Red = acidic residue, Blue = basic residue. **B)** Root mean squared fluctuation (RMSF) of the IFNλ4 protein during molecular dynamics simulations reveal flexibility and disorder of Helix E. **C)** Design of a chimeric IFNλ4 (xIFNλ4) replacing the disordered and positively-charged Helix E (pink) with the corresponding region of IFNλ3 (blue). **D)** SDS-PAGE protein gel of xIFNλ4.

The noticeable difference between IFNλ3 and IFNλ4 in the sequence space leads to dramatic differences in surface charge of the molecules which are readily visualized by coloring non-polar, basic, and positive residues differently **(Fig. 13A**). Viewing the ligand from the front, IFNλ3 has a large acidic patch that engages with both IFNλR1 and IL10Rβ, but mostly the latter; while the acidic regions of the protein contacting IFNλR1 are mostly conserved, this same region for IFNλ4 at the IL10Rβ interface contains more basic and neutral residues, highlighting another example of how IL10Rβ recognizes diverse arrays of ligand despite not only poor sequence but also charge conservation. During structural refinement, we were unable to resolve the structure of Helix E of IFNλ4. While this region of the protein is seemingly uninvolved in receptor engagement, molecular modeling reveals that this highly basic region of IFNλ4 is predicted to lack secondary structure unlike the corresponding region of IFNλ3, breaking from the structural conservation of the two ligands. We sought to characterize the behavior of this region and used molecular dynamics simulations to observe behavioral differences between individual residues of the IFNλ4 complex (**Fig. 13B).** By calculating the root mean squared fluctuation (RMSF) of IFNλ4 throughout the simulation, the disorder and flexibility of Helix E becomes immediately clear. More interestingly, throughout the course of the simulations, the negative ions in the simulation environment congregate exclusively around this region **(Appendix C, Supplemental Figure 22A,B**). We interpret these results to show that this disordered region remains highly flexible when IFNλ4 is bound to IFNλR1 and IL10Rβ and has a proclivity to interact with negatively charged objects present in its surroundings. Because of the known tendency of IFNλ4 to induce cell stress and seemingly adhere to its surroundings during protein translation and trafficking,[33] we were curious to see if removal of this region would have an impact on protein expression. We designed a chimeric version of IFNλ4 (xIFNλ4) **(Appendix C, Supplemental**

Table 8) by replacing Helix E with the corresponding region of IFNλ3 to see if there were observable differences in protein expression **(Fig. 13C**). Excitingly, we found that removing Helix E of IFNλ4 and replacing it with a more ordered and less basic peptide rescues protein expression in insect cells **(Fig. 13D)**, implicating this region as a potential structural basis of IFNλ4 aggregation and resulting dysfunction. These charged, disordered regions are not uncommon in cytokines and have been demonstrated to have roles in cytokine sequestration via interaction with components of the extracellular matrix,[103–105] but to find one in the middle of a folded protein is a curious and unique finding deserving of more thorough investigation.

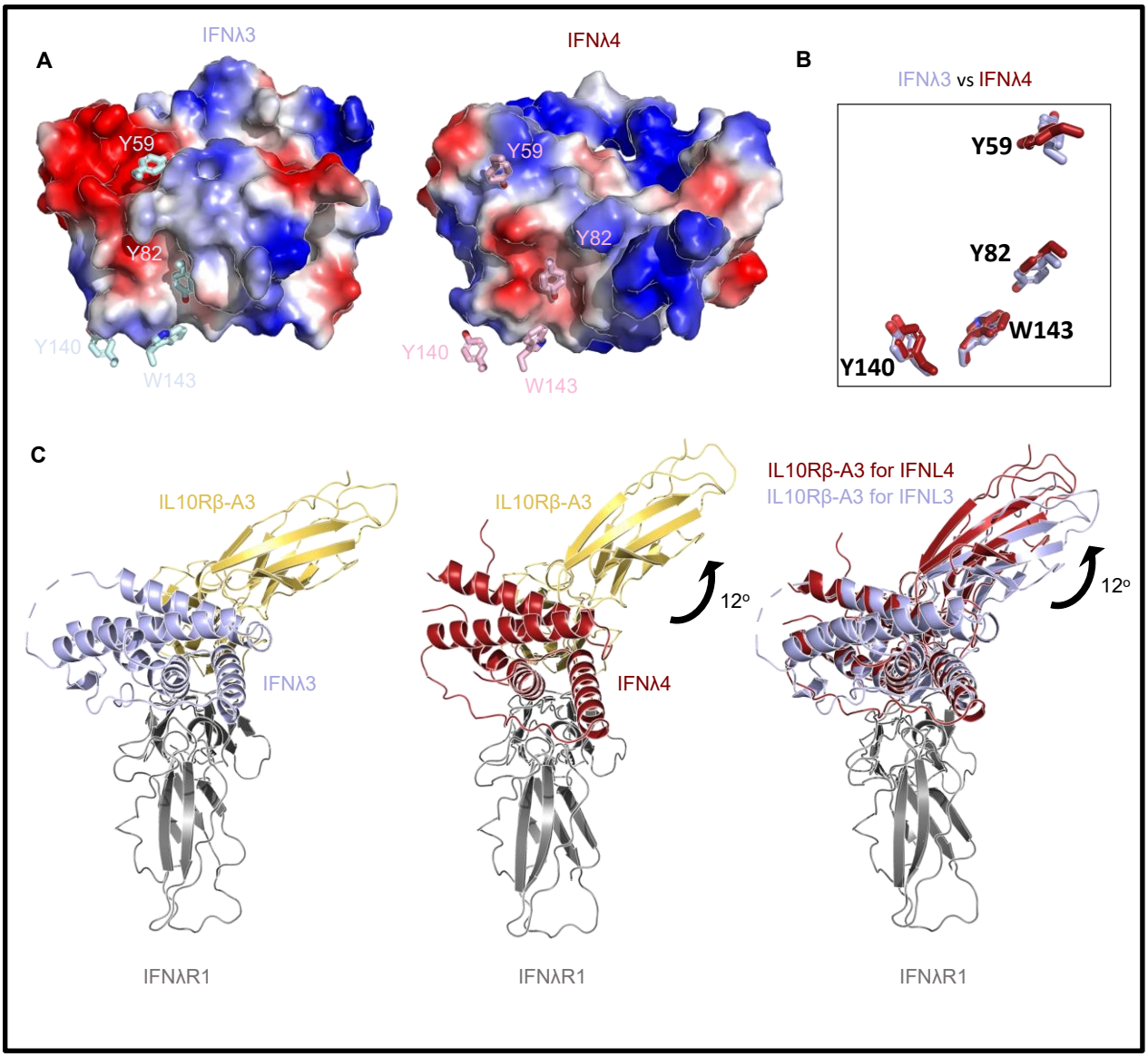


Figure 14: Distinct rotation of SD1 domain on IL10Rβ occurs when binding IFNλ4. A) IFNλ3 (left) and IFNλ4 (right) both engage the hydrophobic residue network on IL10Rβ despite differences in surface charge and amino acid composition. **B)** Exact differences in the orientations of Tyr59, Tyr82, Tyr140, and Trp143 on IL10Rβ when engaging IFNλ3 (blue) and IFNλ4 (red). **C)** A distinct rotation of the IL10Rβ receptor is observed between the IFNλ3 and IFNλ4 ternary complexes. 12-degree rotation is calculated after alignment of the IFNλR1 proteins. Protein complexes are viewed from the top.

The importance of the hydrophobic residue network on IL10Rβ has been extensively described and consists of the Tyr59 residue on Loop 2, the Tyr82 residue on Loop 3, and the Tyr140 and Trp143 residues on Loop 5.[38,69,106–108] These residues indeed form extensive contact

with IFNλ4 and IFNλ3, but in distinct and different orientations **(Fig. 14A,B**). While the Tyr140 residue remains in largely the same place, the Trp143 residue is completely flipped for IFNλ4 and IFNλ3, with its position for IFNλ4 in the exact conformation it is in the engineered IFNλ3-H11. As a result, this leads to the same observed "pinch" on Helix A of IFNλ4 formed by Tyr140 and Trp143, with Tyr140 forming a hydrogen bond with Arg38 on IFNλ4 and Trp143 forming a hydrogen bond with the backbone oxygen of Leu35. In both cases, the Trp143 is buried at the interface, however. Serendipitously, this Arg38 of IFNλ4 is the same as the one selected by directed evolution for IFNλ3-H11 (Q15R) and forms a hydrogen bond as well. The orientation of Tyr82 leads it to engage spatially similar areas of the type III IFNs and in both cases is almost completely buried at the interface (>97% of total amino acid surface area) but due to the differences in their surface charges, Tyr82 with IFNλ4 is more perpendicular to the ligand and forms a hydrogen bond with the backbone nitrogen of Glu36 instead of the multiple hydrogen bonds formed with Ser11 and Ser13 on IFNλ3. Tyr59 displays extreme differences in orientation, with significant rotation of both the main chain of IL10Rβ and the side chain of the tyrosine. Tyr59 is 76% buried at the IFNλ3-IL10Rβ interface and forms two hydrogen bonds with the ligand at Glu73 and Asp87, affirming the importance of the bond at residue 73 on IFNλ3 for recognition by IL10Rβ. In stark contrast, Tyr59 forms no polar contacts with IFNλ4 and is approximately 51% buried, indicating that this interaction is more hydrophobic in nature.

As previously discussed, the overall topology of the IFNλR1 receptor is largely the same for the IFNλ4 and IFNλ3 complexes - as such, our analysis is performed by aligning the two complexes to the IFNλR1 protein. This approach revealed the previously described details of the IL10Rβ hydrophobic network, but also revealed a significant rotation of the orientation IL10Rβ (**Fig. 14C**). This was calculated to be about 12^{o} relative to the orientation of the IL10Rβ receptor

for IFNλ3 by measuring the carbon alpha on the outer most residue of the outer most loop and calculating its position relative to the carbon alpha of Tyr82, which is in roughly the same spot for both complexes. This rotation takes places entirely in the SD1 domain of IL10Rβ and highlights the potential role of receptor geometry in both recognition of diverse ligands as well as potentially explaining differences in experimental results observed between IFNλ3 and IFNλ4. Previous studies have explored the role of receptor geometry with other cytokine-receptor complexes and speculated on connections between geometry to signaling amplitude, gene induction, and other key *in vitro* metrics of cell signaling.[109–111] Observing this behavior between these two complexes solved using identical methods and proteins (aside from the ligand) not only highlights again the ability of IL10Rβ to recognize diverse ligands but also presents interesting design insights for potentiating signaling through IL10Rβ or other receptors. Structurally this rotation could be explained by the interaction of Arg26 residue on IFNλ4 forming two hydrogen bonds with the backbone oxygens of residues Asp197 (~50% buried) and Arg198 (~80% buried) on IL10Rβ, as this residue is not conserved on IFNλ3. While Asp197 and Arg198 remain at the IFNλ3-IL10Rβ interface, they do not form hydrogen bonds and are only ~15% and ~31% buried, respectively.

4.4 Discussion

This Chapter presents the structure of IFNλ4 in complex with its two receptors, IFNλR1 and an engineered IL10Rβ, as well as the ternary complex of IFNλ3. These structure solved in parallel provide the most complete and thorough structural comparison possible for this study and lends high confidence to the differences observed in our analysis. Through this structural analysis, we identify a potential structural mechanism for intracellular IFNλ4 accumulation as well as highlight differences in receptor geometry that potentially explains the clearly observable differences in extracellular signaling ability.

The biochemical nature of IFNλ4 required significant effort to isolate sufficient quantities of ternary complex for structural studies. Indeed, these structures represent some of the smallest protein complexes resolved by single particle cryoEM to date, and this work has demonstrated the feasibility of structural study of other cytokine-receptor complexes by single particle cryoEM with small amount of protein complex. The approach to engineer a single protein to solve multiple structures is also of note, calling attention to the traditional methods in this field of engineering a single protein for a single structure. While the ternary complexes of IFNλ1 and IFNλ2 remain unresolved, their sequence similarity to IFNλ3 implies that the complexes would be highly similar. The structural insights gained from this study, most notably the rotation of the receptors to accommodate two distinctly different ligands with noticeably different extracellular geometries, can be utilized as design considerations for future protein engineering and design efforts to signaling through the type III IFN pathway as mediated by IFNλR1 and IL10Rβ, but more broadly should be considered as another variable for therapeutic design.

The role of IFNλ4 in the human immune system is still being explored, with studies presenting IFNλ4 expression as both a benefit and a detriment to human health. Expression of

IFNλ4 enabled by this study can help resolve questions associated with its *in vitro* function, but questions about its function in the immune system and the systemic impact of IFNλ4 expression remain unanswered. The clinical ramifications of this problem are complicated by the use of mouse models, which do not express IFNλ4[112] – towards this point, the cross-reactivity for human IFNλ4 is currently unknown, and could perhaps be cross-reactive with murine models in the same way that IFNλ2 and IFNλ3 are. Extensive characterization of systems-level IFNλ4 signaling could be further explored using non-human primate models, which are known to express IFNλ4, but would require extensive resources and effort. Humanized mouse models[113] would create pathways for *in vivo* IFNλ4 characterization, as well, but again may present challenges in the field of tissue and organism engineering. Previous problems identified by IFNλ4 expression have implicated endoplasmic reticulum stress and thus impairment of antigen loading for T cell recognition as the causative agent of the diminished viral clearance observed in chronic HCV patients who express IFNλ4, and the exact mechanism of how stress is induced remains unclear. We propose that the role of the disordered, highly basic Helix E of IFNλ4 discussed in our results is non-trivial and may lead to recruitment of undesired interactions during protein translation and trafficking, leading to accumulation and associated stress. This hypothesis seems especially plausible in the context of other cytokines containing similarly charged and disordered regions enabling interactions with components of the extracellular matrix (ECM). Targeting this domain with small molecules[114] may promote improved folding or stability of this chain, improving expression of the protein and alleviating these differences in antiviral clearance. Whether IFNλ4 is meant to be retained in this fashion and that cell stress is indeed a desired component of the proteins function, however, remains an important question.

4.5 Conclusion

In this Chapter, we report the structure of the ternary IFNλ4-IFNλR1-IL10Rβ protein complex at a resolution of 3.26 Å as well as the ternary IFNλ3-IFNλR1-IL10Rβ complex at a resolution of 3.00 Å as determined by cryogenic electron microscopy. Analysis of these structures reveals similarities and differences to how IFNλ4 engages with its receptors compared to IFNλ3, providing mechanistic insight into how IFNλR1 and IL10Rβ can recognize these related ligands despite sharing only 28% sequence similarity. These findings determine the structural basis of IFNλ4 function and can be leveraged to design therapeutics that correct its dysfunction in affected patient populations as well as enable further *in vitro* and *in vivo* work characterizing this protein's role in the human immune system.

CHAPTER 5: Conclusions and future work

5.1 Introduction

IFNλ4 has posed one of the most interesting conundrums in immunology since its discovery in 2013, with expression of this protein being linked to complications associated with viral clearance. While genetic studies have enabled an understanding of the cellular impact of IFNλ4 expression, extensive structural and functional characterization have been greatly limited by the inability to express and purify sufficient quantities of protein. In this work, we report a method for high-yield expression and recovery of IFNλ4, enabling previously inaccessible functional and structural studies. Using this high-yield system in combination with protein engineering to overcome affinity limitations, we report the structure of the ternary IFNλ4-IFNλR1-IL10Rβ protein complex at a resolution of 3.26 Å as well as the ternary IFNλ3-IFNλR1-IL10Rβ complex at a resolution of 3.00 Å as determined by cryogenic electron microscopy. Analysis of these structures reveals similarities and differences to how IFNλ4 engages with its receptors compared to IFNλ3, providing mechanistic insight into how IFNλR1 and IL10Rβ can recognize these related ligands despite sharing only 28% sequence similarity, and by integrating molecular simulations to the workflow of cytokine-receptor structure analysis, we identify otherwise inaccessible behaviors and dynamics of the IFNλ4 complex, specifically the behavior of the disordered peptide region and its potential functional implications on IFNλ4 signaling and expression.

While this work represents some of the most extensive work on the IFNλ4 protein to date, these findings open the door to many new paths of research by revealing several intriguing questions. The following sections document proposed future experiments for the characterization of the extracellular and systems functions of IFNλ4, proposed improvements and applications of

the methods detailed in these chapters, and relevant therapeutic strategies for alleviating the burden of IFNλ4 expression in afflicted patient populations.

5.2 Future characterization of IFNλ4

The initial and ultimate goal of this work was to enable solution of the IFNλ4 receptor complex, which required overcoming the long-standing expression bottleneck. In doing so, rigorous characterization of the IFNλ4 protein will now be available to the IFNλ4 community. The signaling and gene induction data presented in Chapter 2 highlight the differential activity of IFNλ3 and IFNλ4 despite using the same receptor pair, and these differences are further highlighted by the distinct structural configurations of their receptor complexes presented in Chapter 4. Definitive and extensive characterization of the *in vitro* function of IFNλ4 represents an obvious next step in this field, specifically in the context of hepatitis C and other viral infections, and the antiviral abilities of IFNλ4 relative to IFNλ3 and other IFNs must be determined using the quantitative methods for IFNλ4 expression detailed here. Other experiments evaluating the *in vitro* function of IFNλ4, including RNA-Seq,[115] should also be explored, but with careful consideration for the cell lines selected and the conclusions drawn from these results. It has been determined that IFNλ4 signals through IFNλR1 and IL10Rβ[21] and indeed we have confirmed that this receptor complex forms readily, but administration of extracellular IFNλ4 could lead to wide ranges of outcomes in cells that do and do not express IFNλR1. Our findings that IFNλ4 and IFNλ3 differentially activate the JAK/STAT pathway and the apparent correspondence to differences in geometry of the IL10Rβ receptor should also be more rigorously characterized in various cell types, with emphasis and attention to human hepatic cell lines that are more relevant to the chronic HCV infections that illuminated the existence of IFNλ4 in the first place, as well as other barrier interfaces like lung tissues that are infected by respiratory viruses. The comprehensive role of IFNλ4 in the human immune system and the systems-level impact of IFNλ4 expression remains unrealized, but the production of

large quantities of IFNλ4 for administration into animal models may provide some insight into this important question. Treatment of cell lines that express IFNλ4 with extracellular IFNλ4 to observe the impact on cell function would provide a more reasonable first step in answering this question.

The importance of Helix E of IFNλ4, which could not be resolved by cryoEM and is predicted to be disordered, is another key finding of this work. Preliminary results presented in Chapter 4 demonstrate that this region is highly flexible, associates with negatively charged ions throughout molecular simulations, and that removing this region from the IFNλ4 protein and replacing it with the comparable region of IFNλ3 rescues protein secretion, implicating this region as one cause of intracellular IFNλ4 accumulation and cellular retention. Previous studies documenting the accumulation of IFNλ4 have reported that the protein accumulates mostly in the endoplasmic reticulum and Golgi apparatus and that this in turn inhibits HCV peptide-specific CD8+ T cell activation.[33] Similar sets of experiments involving immunofluorescence microscopy and reporter cell lines expressing this chimeric IFNλ4 would be needed to definitively characterize this region as the causative agent of IFNλ4 accumulation; first steps can instead be taken by creating cell lines that express chimeric IFNλ4 to observe the impact on cell function and gene transcription as done in previous studies analyzing the impact of IFNλ4 expression on host cells.[29] It is important to note that any fluorescent studies tracking chimeric IFNλ4 expression should first confirm that the protein can be detected by commercially available reagents, as removal of Helix E may impact antibody binding.

Chapter 2 alludes to the importance of clinically identified point mutations of IFNλ4 and the impact that these mutations have on patient progress and presentation in the clinic. With this improved method of recombinant IFNλ4 expression, previously infeasible *in vitro* experiments to

characterize the impact of these mutations on IFNλ4 function are also now viable. The structure of the IFNλ4 complex can also be used to create highly accurate models of these proteins in complex with IFNλR1 and IL10Rβ to observe structural or dynamical differences imparted by these mutations; for these experiments, the molecular simulation and modeling protocols utilized in Chapter 3 and Chapter 4 could be particularly impactful.

5.3 Proposed improvements for methods and approaches to cytokine engineering and the solution of cytokine-receptor complexes

Improvements in structural and computational biology, most notably AlphaFold,[55] have revolutionized our abilities to predict and solve the three-dimensional shape of proteins and protein complexes. These approaches will only continue to improve as computational performance advances and as the protein database from which these programs draw their training data becomes more comprehensive. Towards that understanding, the methods and workflow detailed in Chapter 3 for the molecular modeling and simulation of protein complexes will become increasingly accurate. An attractive application of these methods is in protein engineering and design, where the physically informed behaviors of a protein complex obtained using molecular dynamics simulations can accurately quantify changes in protein-protein interactions to then inform experimental design. However, these results must be extensively validated by experimental data to be considered an impactful alternative to traditional protein engineering approaches like directed evolution[116,117] or to compete with the burgeoning field of computational protein design.[118–120] This work highlights the applications of molecular simulation for the quantification of protein-protein interactions and highlights the ability to rationally design cytokine-receptor complexes with improved stability and could be readily applied to other protein complexes, specifically low-affinity complexes that cannot be isolated by traditional methods, but could be further improved by integrating machine learning approaches to better leverage the insights obtained from simulation data, specifically in the context of free energy calculations, as many methods utilized in protein simulation struggle to replicate or predict experimental results.[86,118]

The structural studies initiated in this work led to the solutions of some of the smallest protein complexes resolved by single particle cryoEM, offering an attractive workflow for structural biology as an alternative to x-ray crystallography. No more than 500 micrograms of total protein were used in the solution of the IFNλ4 and IFNλ3 complexes, with neither sampling requiring concentration after elution from the size-exclusion column to yield an acceptable number of particles of the complex for model refinement. Contrasted with the large amounts of protein that are usually expressed and consumed for alternative methods, the methodology detailed here may be of great use to structural biologists working with difficult-to-purify proteins like IFNλ4 or for those working with complexes traditionally considered too small to resolve using cryoEM.

5.4 Therapeutic approaches for remedying cellular issues associated with IFNλ4 expression

The IFNλ4-ΔG allele that leads to the expression of the IFNλ4 protein is a gene that has demonstrated some of the strongest evolutionary selection pressure.[27] While previous studies have speculated on exact mutations or mechanisms leading to its diminished selection in European and Asian populations,[49] the exact drivers of this selection pressure remain unknown and are of vital importance for understanding the global and public health impacts of IFNλ4 expression, specifically for communities that already experience disparities in healthcare equity and access.[121,122] Many of these questions go beyond the scope of structural biology, but determining an exact mechanism by which IFNλ4 functions in the human immune system remains of the utmost importance, as it seems unlikely that this protein would be so strongly selected for in specific populations for it to only negatively impact physiology. The experiments proposed earlier in this Chapter using recombinant IFNλ4 to survey the cellular response of a diverse range of cell types may provide insights into these challenges.

Towards therapeutic strategies for improving the clinical progression of patients with diseases like chronic Hepatitis C (CHC), further analysis of the disordered region of IFNλ4 as previously discussed offers a structure-informed path for treatment. Correcting protein misfolding has been demonstrated to resolve protein function in the context of disease, both by engineering[123] and in some cases addition of small molecules.[114] Coadministration of a small molecule designed to stabilize the Helix E of IFNλ4 with a previously approved treatment for CHC[124] may offer improved viral clearance if the mechanism of IFNλ4 dysfunction intuited by our findings holds true. Genetic therapies ablating IFNλ4 expression currently seem unrealistic, especially given the lack of understanding of the systemic function of IFNλ4. At the very least, consideration of treatment avenues and clinical trial design should incorporate knowledge of

IFNλ4 and the known consequences of its expression such as resistance to IFN treatment, as patients treated with other immunotherapeutic proteins may respond similarly.

REFERENCES

1. Kany, S., Vollrath, J. T. & Relja, B. Cytokines in Inflammatory Disease. *IJMS* **20**, 6008 (2019).

2. Oppenheim, J. J. Cytokines: Past, Present, and Future. *Int J Hematol* **74**, 3–8 (2001).

3. Turner, M. D., Nedjai, B., Hurst, T. & Pennington, D. J. Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 2563–2582 (2014).

4. Ait-Oufella, H., Taleb, S., Mallat, Z. & Tedgui, A. Recent Advances on the Role of Cytokines in Atherosclerosis. *ATVB* **31**, 969–979 (2011).

5. Castro, L. S. *et al.* Interferon-Based Biopharmaceuticals: Overview on the Production, Purification, and Formulation. *Vaccines* **9**, 328 (2021).

6. Åhlin, A., Lärfars, G., Elinder, G., Palmblad, J. & Gyllenhammar, H. Gamma Interferon Treatment of Patients with Chronic Granulomatous Disease Is Associated with Augmented Production of Nitric Oxide by Polymorphonuclear Neutrophils. *Clin Diagn Lab Immunol* **6**, 420–424 (1999).

7. Klapper, J. A. *et al.* High-dose interleukin-2 for the treatment of metastatic renal cell carcinoma: A retrospective analysis of response and survival in patients treated in the surgery branch at the National Cancer Institute between 1986 and 2006. *Cancer* **113**, 293–301 (2008).

8. Filipi, M. & Jack, S. Interferons in the Treatment of Multiple Sclerosis. *International Journal of MS Care* **22**, 165–172 (2020).

9. Aung, T., Grubbe, W. S., Nusbaum, R. J. & Mendoza, J. L. Recent and future perspectives on engineering interferons and other cytokines as therapeutics. *Trends in Biochemical Sciences* **48**, 259–273 (2023).

10. Zhang, J.-M. & An, J. Cytokines, Inflammation, and Pain. *International Anesthesiology Clinics* **45**, 27–37 (2007).

11. Wang, X., Lupardus, P., LaPorte, S. L. & Garcia, K. C. Structural Biology of Shared Cytokine Receptors. *Annu. Rev. Immunol.* **27**, 29–60 (2009).

12. Silva, D.-A. *et al.* De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).

13. Saxton, R. A. *et al.* Structure-based decoupling of the pro- and anti-inflammatory functions of interleukin-10. *Science* **371**, eabc8433 (2021).

14. Khodabakhsh, F., Salimian, M., Hedayati, M. H., Ahangari Cohan, R. & Norouzian, D. Challenges and advancements in the pharmacokinetic enhancement of therapeutic proteins. *Preparative Biochemistry & Biotechnology* **51**, 519–529 (2021).

15. Leader, B., Baca, Q. J. & Golan, D. E. Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov* **7**, 21–39 (2008).

16. Lerner, D. M., Stoudemire, A. & Rosenstein, D. L. Neuropsychiatric Toxicity Associated With Cytokine Therapies. *Psychosomatics* **40**, 428–435 (1999).

17. Baldo, B. A. Side Effects of Cytokines Approved for Therapy. *Drug Saf* **37**, 921–943 (2014).

18. Sleijfer, S., Bannink, M., Gool, A. R., Kruit, W. H. J. & Stoter, G. Side Effects of Interferon-α Therapy. *Pharm World Sci* **27**, 423–431 (2005).

19. Kirkwood, J. M. *et al.* Mechanisms and Management of Toxicities Associated With High-Dose Interferon Alfa-2b Therapy. *JCO* **20**, 3703–3718 (2002).

20. Ravaud, A., Bedane, C., Geoffrois, L., Lesimple, T. & Delaunay, M. Toxicity and feasibility of adjuvant high-dose interferon alpha-2b in patients with melanoma in clinical oncologic practice. *Br J Cancer* **80**, 1767–1769 (1999).

21. Prokunina-Olsson, L. *et al.* A variant upstream of IFNL3 (IL28B) creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus. *Nat Genet* **45**, 164–171 (2013).

22. Sambi, M., Bagheri, L. & Szewczuk, M. R. Current Challenges in Cancer Immunotherapy: Multimodal Approaches to Improve Efficacy and Patient Response Rates. *Journal of Oncology* **2019**, 1–12 (2019).

23. Muir, A. J. *et al.* A randomized phase 2b study of peginterferon lambda-1a for the treatment of chronic HCV infection. *Journal of Hepatology* **61**, 1238–1246 (2014).

24. Donnelly, R. P. & Kotenko, S. V. Interferon-Lambda: A New Addition to an Old Family. *Journal of Interferon & Cytokine Research* **30**, 555–564 (2010).

25. Steen, H. C. & Gamero, A. M. Interferon-Lambda as a Potential Therapeutic Agent in Cancer Treatment. *Journal of Interferon & Cytokine Research* **30**, 597–602 (2010).

26. Onabajo, O. O., Muchmore, B. & Prokunina-Olsson, L. The IFN-λ4 Conundrum: When a Good Interferon Goes Bad. *Journal of Interferon & Cytokine Research* **39**, 636–641 (2019).

27. Fang, M. Z., Jackson, S. S. & O'Brien, T. R. IFNL4: Notable variants and associated phenotypes,. *Gene* **730**, 144289 (2020).

28. Rugwizangoga, B. *et al.* IFNL4 Genotypes Predict Clearance of RNA Viruses in Rwandan Children With Upper Respiratory Tract Infections. *Front. Cell. Infect. Microbiol.* **9**, 340 (2019).

29. Onabajo, O. O. *et al.* Expression of Interferon Lambda 4 Is Associated with Reduced Proliferation and Increased Cell Death in Human Hepatic Cells. *Journal of Interferon & Cytokine Research* **35**, 888–900 (2015).

30. Tang, W. *et al. IFNL4* -ΔG Allele Is Associated with an Interferon Signature in Tumors and Survival of African-American Men with Prostate Cancer. *Clinical Cancer Research* **24**, 5471–5481 (2018).

31. Zahid, W. *et al.* Association of Interferon Lambda 3 and 4 Gene SNPs and Their Expression with COVID-19 Disease Severity: A Cross-Sectional Study. *IDR* **Volume 16**, 6619–6628 (2023).

32. Obajemu, A. A. *et al.* IFN-λ4 Attenuates Antiviral Responses by Enhancing Negative Regulation of IFN Signaling. *The Journal of Immunology* **199**, 3808–3820 (2017).

33. Chen, Q. *et al.* Interferon lambda 4 impairs hepatitis C viral antigen presentation and attenuates T cell responses. *Nat Commun* **12**, 4882 (2021).

34. Lu, Y.-F., Goldstein, D. B., Urban, T. J. & Bradrick, S. S. Interferon-λ4 is a cell-autonomous type III interferon associated with pre-treatment hepatitis C virus burden. *Virology* **476**, 334–340 (2015).

35. Chung, J.-H. *et al.* Structure-based glycoengineering of interferon lambda 4 enhances its productivity and anti-viral potency. *Cytokine* **125**, 154833 (2020).

36. Hamming, O. J. *et al.* Interferon lambda 4 signals via the IFNλ receptor to regulate antiviral activity against HCV and coronaviruses: Recombinant IFNλ4 is antiviral. *The EMBO Journal* **32**, 3055–3065 (2013).

37. Miknis, Z. J. *et al.* Crystal Structure of Human Interferon-λ1 in Complex with Its High-Affinity Receptor Interferon-λR1. *Journal of Molecular Biology* **404**, 650–664 (2010).

38. Mendoza, J. L. *et al.* The IFN-λ-IFN-λR1-IL-10Rβ Complex Reveals Structural Features Underlying Type III IFN Functional Plasticity. *Immunity* **46**, 379–392 (2017).

39. Mendoza, J. L. *et al.* Structure of the IFNγ receptor complex guides design of biased agonists. *Nature* **567**, 56–60 (2019).

40. Grubbe, W. S., Byléhn, F., Alvarado, W., De Pablo, J. J. & Mendoza, J. L. Molecular analysis of the type III interferon complex and its applications in protein engineering. *Biophysical Journal* **122**, 4254–4263 (2023).

41. Erlich, H. A. Polymerase chain reaction. *J Clin Immunol* **9**, 437–447 (1989).

42. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009).

43. Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucl Acids Res* **18**, 6069–6074 (1990).

44. Dukkipati, A., Park, H. H., Waghray, D., Fischer, S. & Garcia, K. C. BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. *Protein Expression and Purification* **62**, 160–170 (2008).

45. Thomas, C. *et al.* Structural Linkage between Ligand Discrimination and Receptor Activation by Type I Interferons. *Cell* **146**, 621–632 (2011).

46. Chen, X. *et al.* Molecular chaperones (TrxA, SUMO, Intein, and GST) mediating expression, purification, and antimicrobial activity assays of plectasin in *Escherichia coli*. *Biotech and App Biochem* **62**, 606–614 (2015).

47. Lubkowski, J. *et al.* Crystal Structure of the Labile Complex of IL-24 with the Extracellular Domains of IL-22R1 and IL-20R2. *The Journal of Immunology* **201**, 2082–2093 (2018).

48. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79–88 (2006).

49. Bamford, C. G. G. *et al.* A polymorphic residue that attenuates the antiviral potential of interferon lambda 4 in hominid lineages. *PLoS Pathog* **14**, e1007307 (2018).

50. Bhushan, A., Ghosh, S., Bhattacharjee, S. & Chinnaswamy, S. Confounding by Single Nucleotide Polymorphism rs117648444 (P70S) Affects the Association of Interferon Lambda Locus Variants with Response to Interferon-α-Ribavirin Therapy in Patients with Chronic Genotype 3 Hepatitis C Virus Infection. *Journal of Interferon & Cytokine Research* **37**, 369–382 (2017).

51. Zhou, H. *et al.* The *IFNL4* Gene Is a Noncanonical Interferon Gene with a Unique but Evolutionarily Conserved Regulation. *J Virol* **94**, e01535-19 (2020).

52. Ansari, M. A. *et al.* Interferon lambda 4 impacts the genetic diversity of hepatitis C virus. *eLife* **8**, e42463 (2019).

53. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat Rev Genet* **16**, 379–394 (2015).

54. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).

55. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

56. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* **13**, 1265 (2022).

57. Hollingsworth, S. A. & Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **99**, 1129–1143 (2018).

58. Childers, M. C. & Daggett, V. Insights from molecular dynamics simulations for computational protein design. *Mol. Syst. Des. Eng.* **2**, 9–33 (2017).

59. Sahoo, B. R. *et al.* 3D modeling and molecular dynamics simulation of an immune-regulatory cytokine, interleukin-10, from the Indian major carp, Catla catla. *J Mol Model* **18**, 1713–1722 (2012).

60. Xu, D. *et al.* Molecular insights into the improved clinical performance of PEGylated interferon therapeutics: a molecular dynamics perspective. *RSC Adv.* **8**, 2315–2322 (2018).

61. Lilkova, E. *et al.* Molecular modeling of the effects of glycosylation on the structure and dynamics of human interferon-gamma. *J Mol Model* **25**, 127 (2019).

62. Levin, A. M. *et al.* Exploiting a natural conformational switch to engineer an interleukin-2 'superkine'. *Nature* **484**, 529–533 (2012).

63. Schillinger, O., Panwalkar, V., Strodel, B. & Dingley, A. J. Molecular Dynamics Simulations Reveal Key Roles of the Interleukin-6 Alpha Receptor in the Assembly of the Human Interleukin-6 Receptor Complex. *J. Phys. Chem. B* **121**, 8113–8122 (2017).

64. Chuartzman, S. G. *et al.* Binding of interferon reduces the force of unfolding for interferon receptor 1. *PLoS ONE* **12**, e0175413 (2017).

65. Sousa, Laurent, Quéméner, Mortier, & Questel. Mechanistic and Structural Insights on the IL-15 System through Molecular Dynamics Simulations. *Molecules* **24**, 3261 (2019).

66. Reis, G. *et al.* Early Treatment with Pegylated Interferon Lambda for Covid-19. *N Engl J Med* **388**, 518–528 (2023).

67. Prokunina-Olsson, L. *et al.* COVID-19 and emerging viral infections: The case for interferon lambda. *Journal of Experimental Medicine* **217**, e20200653 (2020).

68. Lazear, H. M., Nice, T. J. & Diamond, M. S. Interferon-λ: Immune Functions at Barrier Surfaces and Beyond. *Immunity* **43**, 15–28 (2015).

69. Gad, H. H. *et al.* Interferon-λ Is Functionally an Interferon but Structurally Related to the Interleukin-10 Family. *Journal of Biological Chemistry* **284**, 20869–20875 (2009).

70. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).

71. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).

72. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

73. Best, R. B. & Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix−Coil Transition of Polypeptides. *J. Phys. Chem. B* **113**, 9004–9015 (2009).

74. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).

75. Miao, Y., Feher, V. A. & McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **11**, 3584–3595 (2015).

76. Wang, J. & Miao, Y. Mechanistic Insights into Specific G Protein Interactions with Adenosine Receptors. *J. Phys. Chem. B* **123**, 6462–6473 (2019).

77. D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, *et al.* AMBER20.

78. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

79. Miao, Y., Bhattarai, A. & Wang, J. *Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics*. http://biorxiv.org/lookup/doi/10.1101/2020.04.20.051979 (2020) doi:10.1101/2020.04.20.051979.

80. Miao, Y. *et al.* Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **10**, 2677–2689 (2014).

81. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

82. Eckmann, J.-P., Rougemont, J. & Tlusty, T. *Colloquium* : Proteins: The physics of amorphous evolving matter. *Rev. Mod. Phys.* **91**, 031001 (2019).

83. Perez-Lemus, G. R., Menéndez, C. A., Alvarado, W., Byléhn, F. & de Pablo, J. J. Toward wide-spectrum antivirals against coronaviruses: Molecular characterization of SARS-CoV-2 NSP13 helicase inhibitors. *Sci. Adv.* **8**, eabj4526 (2022).

84. Gullett, P. M., Horstemeyer, M. F., Baskes, M. I. & Fang, H. A deformation gradient tensor and strain tensors for atomistic simulations. *Modelling Simul. Mater. Sci. Eng.* **16**, 015001 (2007).

85. Mitchell, M. R., Tlusty, T. & Leibler, S. Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings. *Proceedings of the National Academy of Sciences* **113**, E5847–E5855 (2016).

86. Bhati, A. P., Wan, S., Wright, D. W. & Coveney, P. V. Rapid, Accurate, Precise, and Reliable Relative Free Energy Prediction Using Ensemble Based Thermodynamic Integration. *J. Chem. Theory Comput.* **13**, 210–222 (2017).

87. Wang, C., Greene, D., Xiao, L., Qi, R. & Luo, R. Recent Developments and Applications of the MMPBSA Method. *Front. Mol. Biosci.* **4**, 87 (2018).

88. De, M. *et al.* Distinct molecular phenotypes involving several human diseases are induced by IFN-λ3 and IFN-λ4 in monocyte-derived macrophages. *Genes Immun* 1–12 (2022) doi:10.1038/s41435-022-00164-w.

89. O'Brien, T. R. *et al.* Comparison of functional variants in IFNL4 and IFNL3 for association with HCV clearance. *Journal of Hepatology* **63**, 1103–1110 (2015).

90. *LP 2013, 14th International Symposium on Loss Prevention and Safety Promotion in the Process Industries: 12 - 15 May 2013, Florence, Italy*. (AIDIC, Milano, 2013).

91. Sarkar, M. *et al.* Association of IFNL3 and IFNL4 polymorphisms with liver-related mortality in a multiracial cohort of HIV/HCV-coinfected women. *Journal of Viral Hepatitis* **22**, 1055–1060 (2015).

92. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat Protoc* **1**, 755–768 (2006).

93. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).

94. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **180**, 519–530 (2012).

95. Adams, P. D. *et al. PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).

96. Emsley, P. & Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).

97. Procter, J. B. *et al.* Alignment of Biological Sequences with Jalview. in *Multiple Sequence Alignment* (ed. Katoh, K.) vol. 2231 203–224 (Springer US, New York, NY, 2021).

98. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* **372**, 774–797 (2007).

99. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319–2327 (2011).

100. Gowers, R. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. in 98–105 (Austin, Texas, 2016). doi:10.25080/Majora-629e541a-00e.

101. Piehler, J., Thomas, C., Garcia, K. C. & Schreiber, G. Structural and dynamic determinants of type I interferon receptor assembly and their functional interpretation. *Immunological Reviews* **250**, 317–334 (2012).

102. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* **10**, 584–590 (2013).

103. Kemna, J. *et al.* IFNγ binding to extracellular matrix prevents fatal systemic toxicity. *Nat Immunol* **24**, 414–422 (2023).

104. Nguyen, P. T. *et al.* Microglial Remodeling of the Extracellular Matrix Promotes Synapse Plasticity. *Cell* **182**, 388-403.e15 (2020).

105. Vorobjeva, N. V. & Chernyak, B. V. NETosis: Molecular Mechanisms, Role in Physiology and Pathology. *Biochemistry Moscow* **85**, 1178–1190 (2020).

106. Logsdon, N. J. *et al.* The IL-10R2 Binding Hot Spot on IL-22 is Located on the N-terminal Helix and is Dependent on N-linked Glycosylation. *Journal of Molecular Biology* **342**, 503–514 (2004).

107. Wu, P. W. *et al.* IL-22R, IL-10R2, and IL-22BP Binding Sites Are Topologically Juxtaposed on Adjacent and Overlapping Surfaces of IL-22. *Journal of Molecular Biology* **382**, 1168–1183 (2008).

108. Yoon, S. *et al.* Structure and Mechanism of Receptor Sharing by the IL-10R2 Common Chain. *Structure* **18**, 638–648 (2010).

109. Moraga, I. *et al.* Tuning Cytokine Receptor Signaling by Re-orienting Dimer Geometry with Surrogate Ligands. *Cell* **160**, 1196–1208 (2015).

110. Mohan, K. *et al.* Topological control of cytokine receptor signaling induces differential effects in hematopoiesis. *Science* **364**, eaav7532 (2019).

111. Moraga, I. *et al.* Synthekines are surrogate cytokine and growth factor agonists that compel signaling through non-natural receptor dimers. *eLife* **6**, e22882 (2017).

112. Jacobs, S. *et al.* Species Specificity of Type III Interferon Activity and Development of a Sensitive Luciferase-Based Bioassay for Quantitation of Mouse Interferon-λ. *Journal of Interferon & Cytokine Research* **38**, 469–479 (2018).

113. Sefik, E. *et al.* A humanized mouse model of chronic COVID-19. *Nat Biotechnol* **40**, 906–920 (2022).

114. Chiti, F. & Kelly, J. W. Small molecule protein binding to correct cellular folding or stabilize the native state against misfolding and aggregation. *Current Opinion in Structural Biology* **72**, 267–278 (2022).

115. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).

116. Wang, Y. *et al.* Directed Evolution: Methodologies and Applications. *Chem. Rev.* **121**, 12384–12444 (2021).

117. Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew Chem Int Ed* **57**, 4143–4148 (2018).

118. Krishna, R. *et al.* Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* eadl2528 (2024) doi:10.1126/science.adl2528.

119. Torres, S. V. *et al.* De novo design of high-affinity binders of bioactive helical peptides. *Nature* (2023) doi:10.1038/s41586-023-06953-1.

120. Muratspahić, E. *et al.* Design and structural validation of peptide–drug conjugate ligands of the kappa-opioid receptor. *Nat Commun* **14**, 8064 (2023).

121. Spearman, C. W. & Sonderup, M. W. Health disparities in liver disease in sub-Saharan Africa. *Liver International* **35**, 2063–2071 (2015).

122. Gillispie-Bell, V. The Contrast of Color: Why the Black Community Continues to Suffer Health Disparities. *Obstetrics & Gynecology* **137**, 220–224 (2021).

123. Mendoza, J. L. *et al.* Requirements for Efficient Correction of ΔF508 CFTR Revealed by Analyses of Evolved Sequences. *Cell* **148**, 164–174 (2012).

124. Manns, M. P. & Maasoumy, B. Breakthroughs in hepatitis C research: from discovery to cure. *Nat Rev Gastroenterol Hepatol* **19**, 533–550 (2022).

Appendix A: Supplementary information for Chapter 2

Primer	Sequence
pAC-IL10Rβ-IFNλ4_fwd	CGGCGCATTCTGCCTTTGCGGGATCCATGGTAC CACCTCCCGAAAATGTCAG
pAC-IL10Rβ-IFNλ4_rev	GATGGTGGTGATGATGCGCGGCCGCGAGGCAA GGCCCAGAGTGTG
pAC-IL10Rβ-IFNλ4-linker_fwd	GTTCTGTTTCAAGGGCCTGGAGGTGGTTCTGAA GGTGGAGGATCTGAAATGCGGCCGAGTGTCTGG G
pAC-IL10Rβ-IFNλ4-linker_rev	CCTCCAGGCCCTTGAAACAGAACTTCCAAAGAT CCACCACCACCGGAGGGGACCGTTTCGTCATG
pAC-sfGFP-IL10Rβ-IFNλ4_fwd	GCGCATTCTGCCTTTGCGGGATCCATGAGCAAA GGTGAAGAACTGTTTACCG
pAC-sfGFP-IL10Rβ-IFNλ4_rev	GATGGTGGTGATGATGCGCGGCCGCGAGGCAA GGCCCAGAGTGTG
pAC-sfGFP-IL10Rβ-IFNλ4-linker_fwd	CGAATATGTGAATGCGGCAGGTATTACGGGATC AATGGTACCACCTCCCGAAAATGTCAG
pAC-sfGFP-IL10Rβ-IFNλ4-linker_rev	CGTAATACCTGCCGCATTCACATATTCG

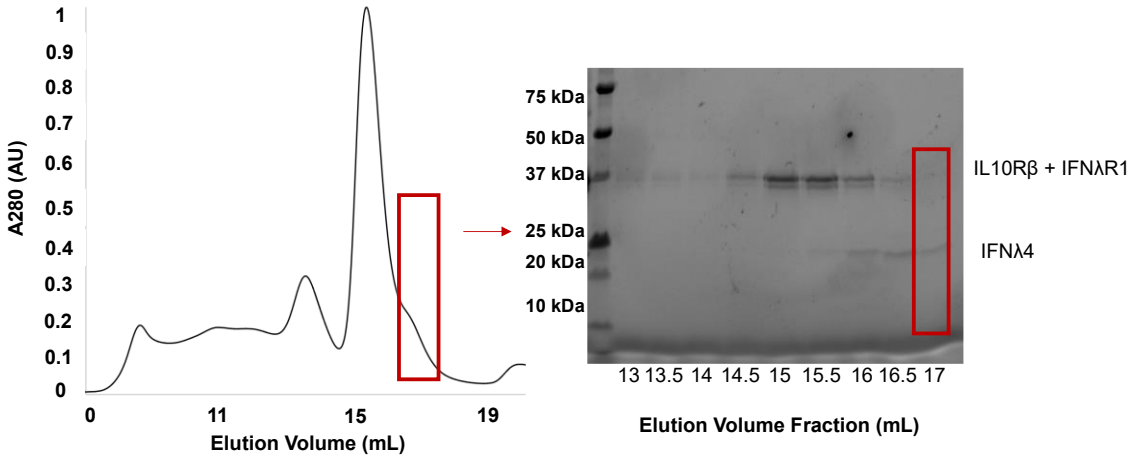
Supplemental Table 1: DNA primers used for molecular cloning of IFNλ4 constructs.

Protein	Genetic sequence
IFNλ4 (Sequence1)	GCTCCAAGAAGATGTTTGTTGTCTCACTACAGATCTTTGGAACCAAGAA CTTTGGCTGCTGCTAAGGCTTTGAGAGACAGATACGAAGAAGAAGCTTT GTCTTGGGGTCAAAGAAACTGTTCTTTCAGACCAAGAAGAGACCCACCA AGACCATCTTCTTGTGCTAGATTGAGACACGTTGCTAGAGGTATCGCTGA CGCTCAAGCTGTTTTGTCTGGTTTGCACAGATCTGAATTGTTGCCAGGTG CTGGTCCAATCTTGGAATTGTTGGCTGCTGCTGGTAGAGACGTTGCTGCT TGTTTGGAATTGGCTAGACCAGGTTCTTCTAGAAAGGTTCCAGGTGCTCA AAAGAGAAGACACAAGCCAAGAAGAGCTGACTCTCCAAGATGTAGAAA GGCTTCTGTTGTTTTCAACTTGTTGAGATTGTTGACTTGGGAATTGAGAT TGGCTGCTCACTCTGGTCCATGTTTG
IL10Rβ-IFNλ4 (Sequence2)	ATGGTACCACCTCCCGAAAATGTCAGAATGAATTCTGTTAATTTCAAGA ACATTCTACAGTGGGAGTCACCTGCTTTTGCCAAAGGGAACCTGACTTTC ACAGCTCAGTACCTAAGTTATAGGATATTCCAAGATAAATGCATGAATA CTACCTTGACGGAATGTGATTTCTCAAGTCTTTCCAAGTATGGTGACCAC ACCTTGAGAGTCAGGGCTGAATTTGCAGATGAGCATTCAGACTGGGTAA ACATCACCTTCTGTCCTGTGGATGACACCATTATTGGACCCCCTGGAATG CAAGTAGAAGTACTTGCTGATTCTTTACATATGCGTTTCTTAGCCCCTAA AATTGAGAATGAATACGAAACTTGGACTATGAAGAATGTGTATAACTCA TGGACTTATAATGTGCAATACTGGAAAAACGGTACTGATGAAAAGTTTC AAATTACTCCCCAGTATGACTTTGAGGTCCTCAGAAACCTGGAGCCATG GACAACTTATTGTGTTCAAGTTCGAGGGTTTCTTCCTGATCGGAACAAAG CTGGGGAATGGAGTGAGCCTGTCTGTGAGCAAACAACCCATGACGAAAC GGTCCCCTCCGGTGGTGGTGGATCTTTGGAAGTTCTGTTTCAAGGGCCTG GAGGTGGTTCTGAAGGTGGAGGATCTGGAGCGGCC+Sequence1
sfGFP-IL10Rβ-IFNλ4 (Sequence3)	ATGAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGG AACTGGATGGCGATGTGAACGGTCACAAATTCAGCGTGCGTGGTGAAGG TGAAGGCGATGCCACGATTGGCAAACTGACGCTGAAATTTATCTGCACC ACCGGCAAACTGCCGGTGCCGTGGCCGACGCTGGTGACCACCCTGACCT ATGGCGTTCAGTGTTTTAGTCGCTATCCGGATCACATGAAACGTCACGAT TTCTTTAAATCTGCAATGCCGGAAGGCTATGTGCAGGAACGTACGATTA GCTTTAAAGATGATGGCAAATATAAAACGCGCGCCGTTGTGAAATTTGA AGGCGATACCCTGGTGAACCGCATTGAACTGAAAGGCACGGATTTTAAA GAAGATGGCAATATCCTGGGCCATAAACTGGAATACAACTTTAATAGCC ATAATGTTTATATTACGGCGGATAAACAGAAAAATGGCATCAAAGCGAA TTTTACCGTTCGCCATAACGTTGAAGATGGCAGTGTGCAGCTGGCAGAT CATTATCAGCAGAATACCCCGATTGGTGATGGTCCGGTGCTGCTGCCGG ATAATCATTATCTGAGCACGCAGACCGTTCTGTCTAAAGATCCGAACGA AAAAGGCACGCGGGACCACATGGTTCTGCACGAATATGTGAATGCGGCA GGTATTACGGGATCA+Sequence2

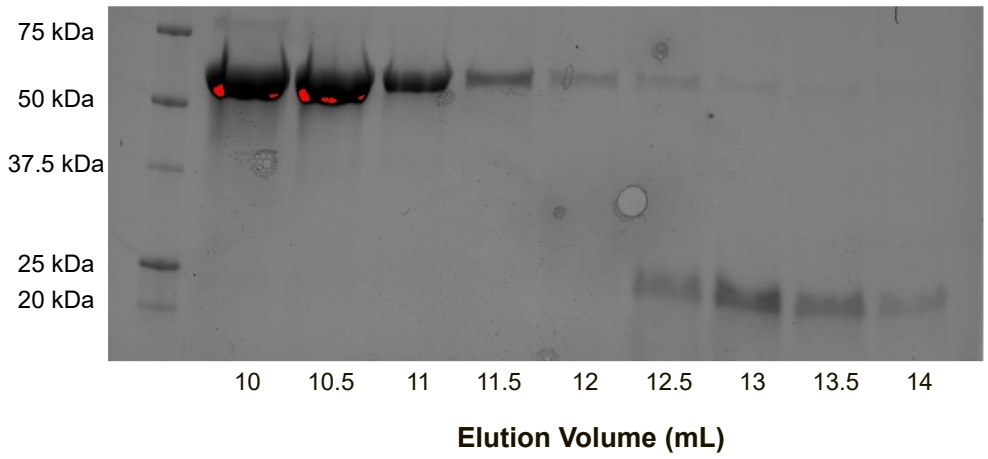
Supplemental Table 2: Genetic sequences of IFNλ4 constructs.

Primer	Sequence
ISG15_fwd	CGCAGATCACCCAGAAGATCG
ISG15_rev	TTCGTCGCATTTGTCCACCA
MX1_fwd	GTTTCCGAAGTGGACATCGCA
MX1_rev	CTGCACAGGTTGTTCTCAGC
APOL3_fwd	GGGACGAGTCTGGCCCTTA
APOL3_rev	TCAATCGGTCAATGCTGGTTG
SAMD9L_fwd	ATTCCAAGCAACGGGATGTAG
SAMD9L_rev	AGTCTCGGTTTCCTATGAGAAGT
18S_fwd	GTAACCCGTTGAACCCCATT
18S_rev	CCATCCAATCGGTAGTAGCG

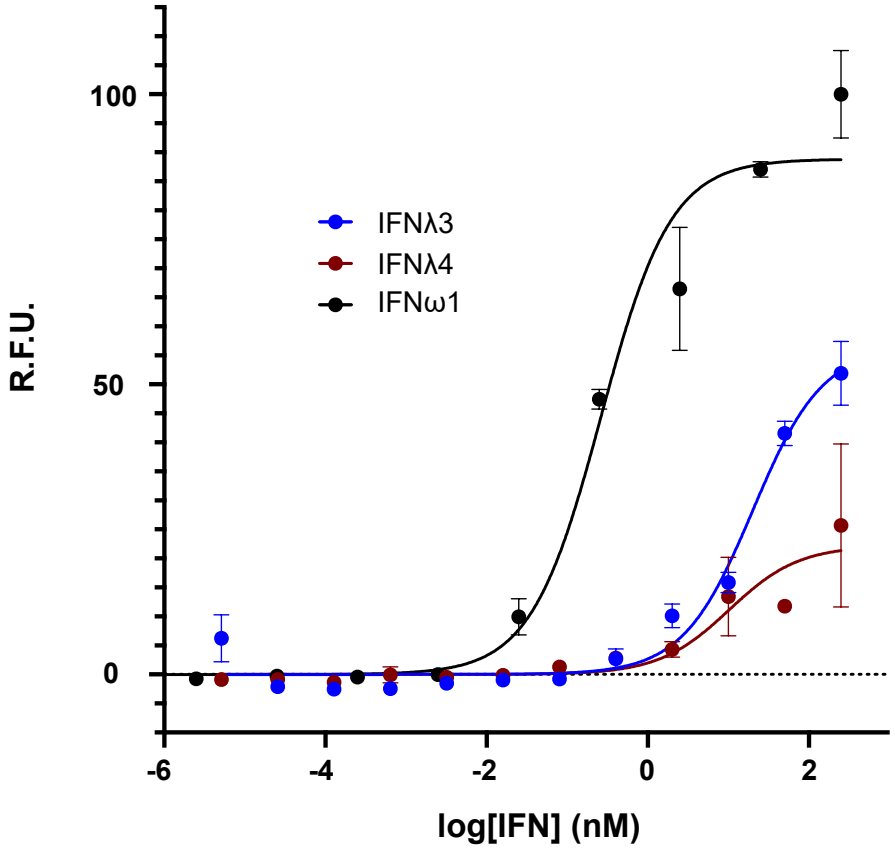
Supplemental Table 3: DNA primers used for quantification of gene induction.



Supplemental Figure 1: Co-expression of IFNλR1 and IL10Rβ-IFNλ4 enables expression in mammalian cells. FPLC trace of IFNλR1 and IL10Rβ-IFNλ4 post 1:100 3C protease treatment (left). Volume fractions containing proteins of interest were then visualized using an SDS-PAGE StainFree protein gel (Bio-Rad) (right). Areas enclosed in red square correspond to fractions containing mostly IFNλ4. (AU = arbitrary units)



Supplemental Figure 2: SDS-PAGE protein gel showing separation of sfGFP-10RB from IFNλ4 following size-exclusion chromatography. This protein gel visualizes elution fractions from the FPLC trace presented in Chapter 2, Figure 3B.

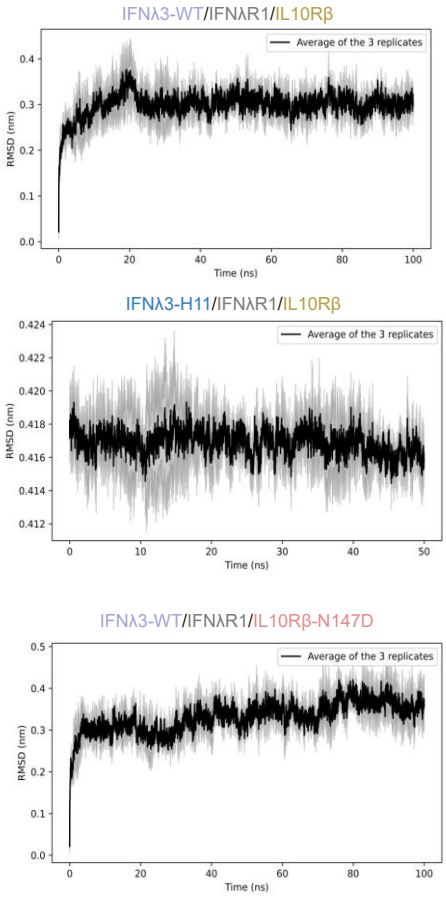


Supplemental Figure 3: Differential activity of IFNλ4 and IFNλ3 is observed in multiple adherent cell lines. pSTAT1 signaling of IFNω1, IFNλ3, and IFNλ4 in A549 cells. Curves are fit to a first-order logistic model, with error bars representing +/- standard error of the mean (SEM) (n = 3).

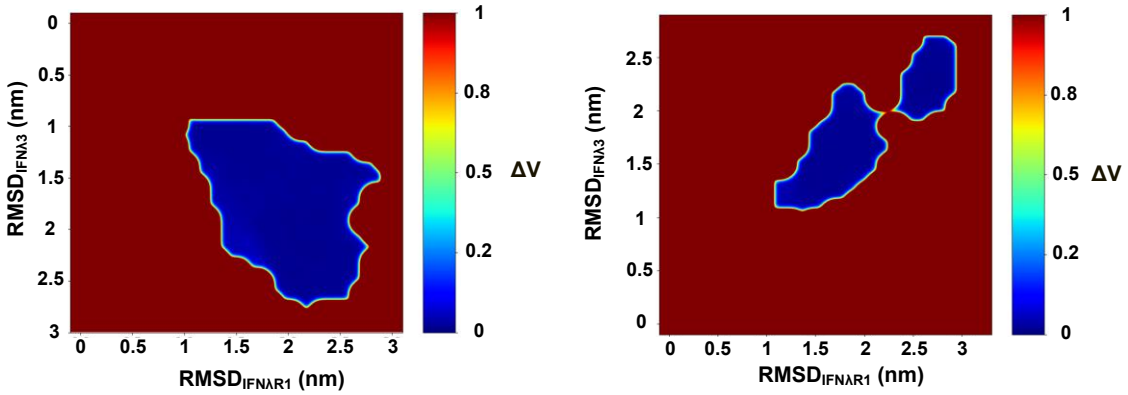
Appendix B: Supplementary information for Chapter 3

Residue	IFNλ3 – Wild Type	IFNλ3 – H11
15	Q (Gln)	R (Arg)
73	E (Glu)	D (Asp)
120	H (His)	R (Arg)
150	T (Thr)	A (Ala)
163	V (Val)	E (Glu)

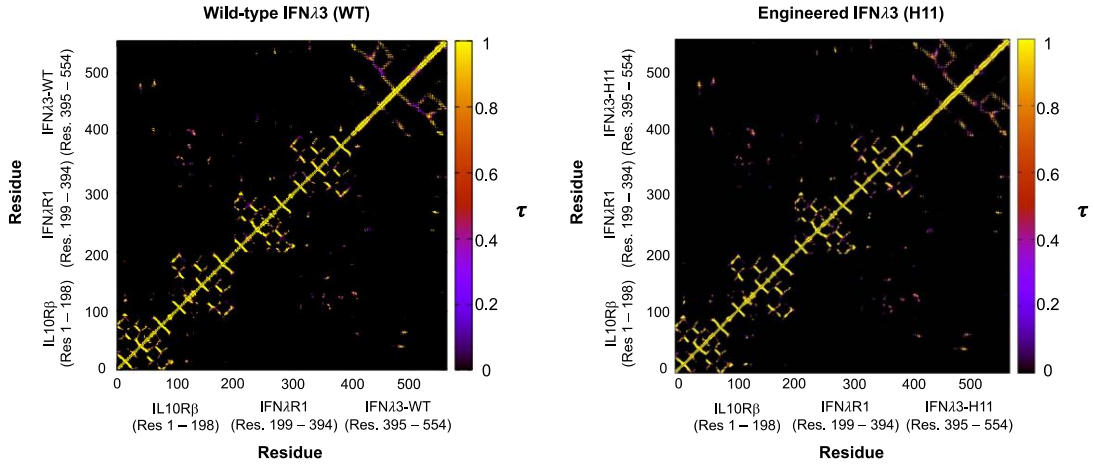
Supplemental Figure 4: Residue differences between the wild-type and engineered IFNλ3 proteins.



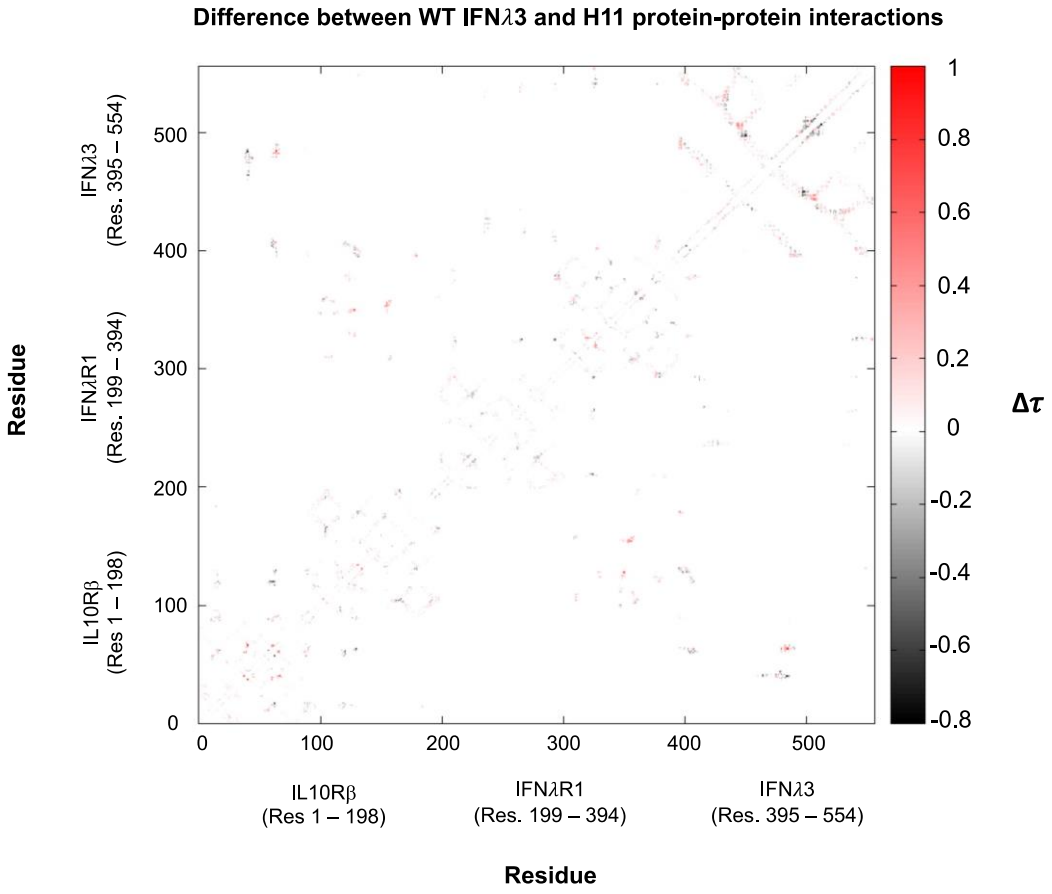
Supplemental Figure 5: RMSD analysis of the molecular dynamics production runs. RMSD analysis for the three simulated complexes are shown to demonstrate equilibration of the system.



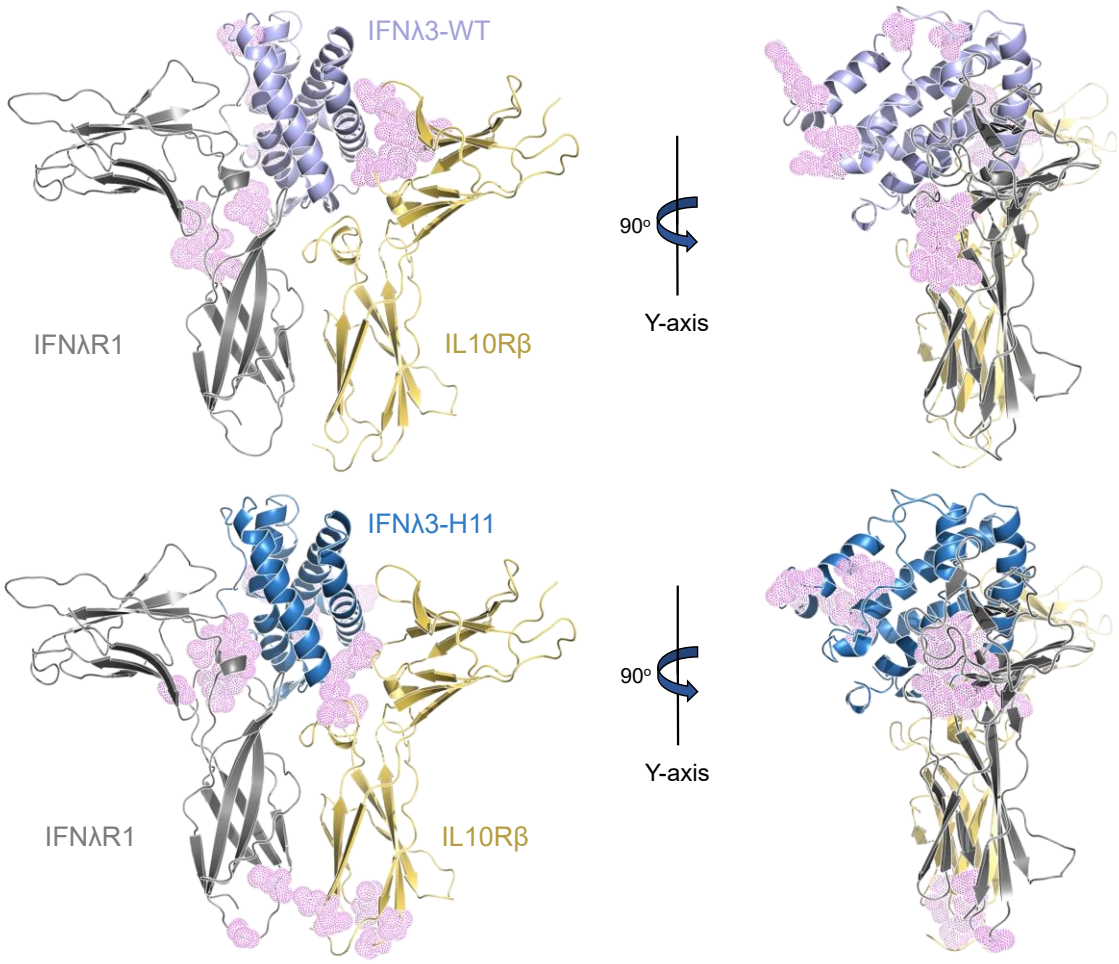
Supplemental Figure 6: Convergence of Gaussian-accelerated molecular dynamics (GaMD) simulations. Thorough exploration and convergence of the sampled regions during GaMD simulations is achieved when anharmonicity is minimized such that the boost potential (ΔV) aligns with a Gaussian distribution.



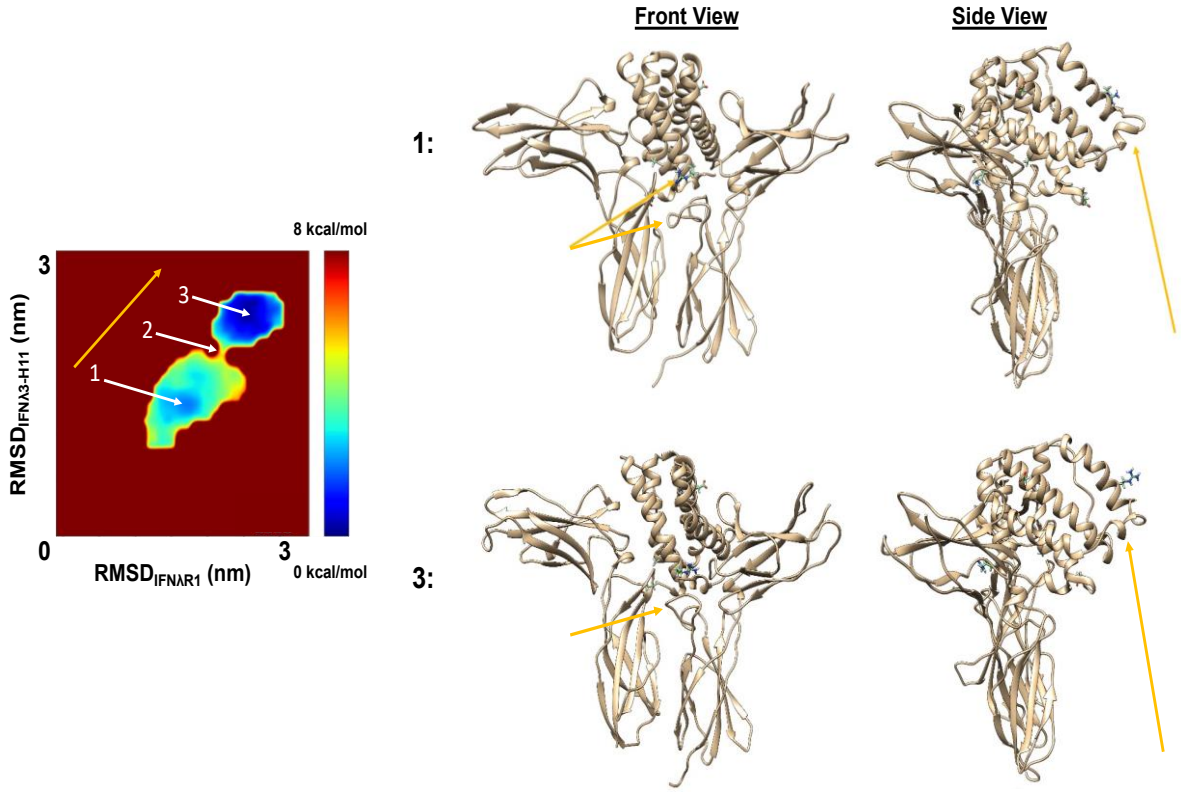
Supplemental Figure 7: Residue contact maps for the wild-type and engineered IFNλ3 complexes. Contact maps generated using trajectory data from molecular dynamics simulations of the wild-type IFNλ3 (left) and engineered IFNλ3 (right) complexes, where tau (τ) represents the percent of simulation time that the alpha carbons of two residues are within 10 angstroms of each other.



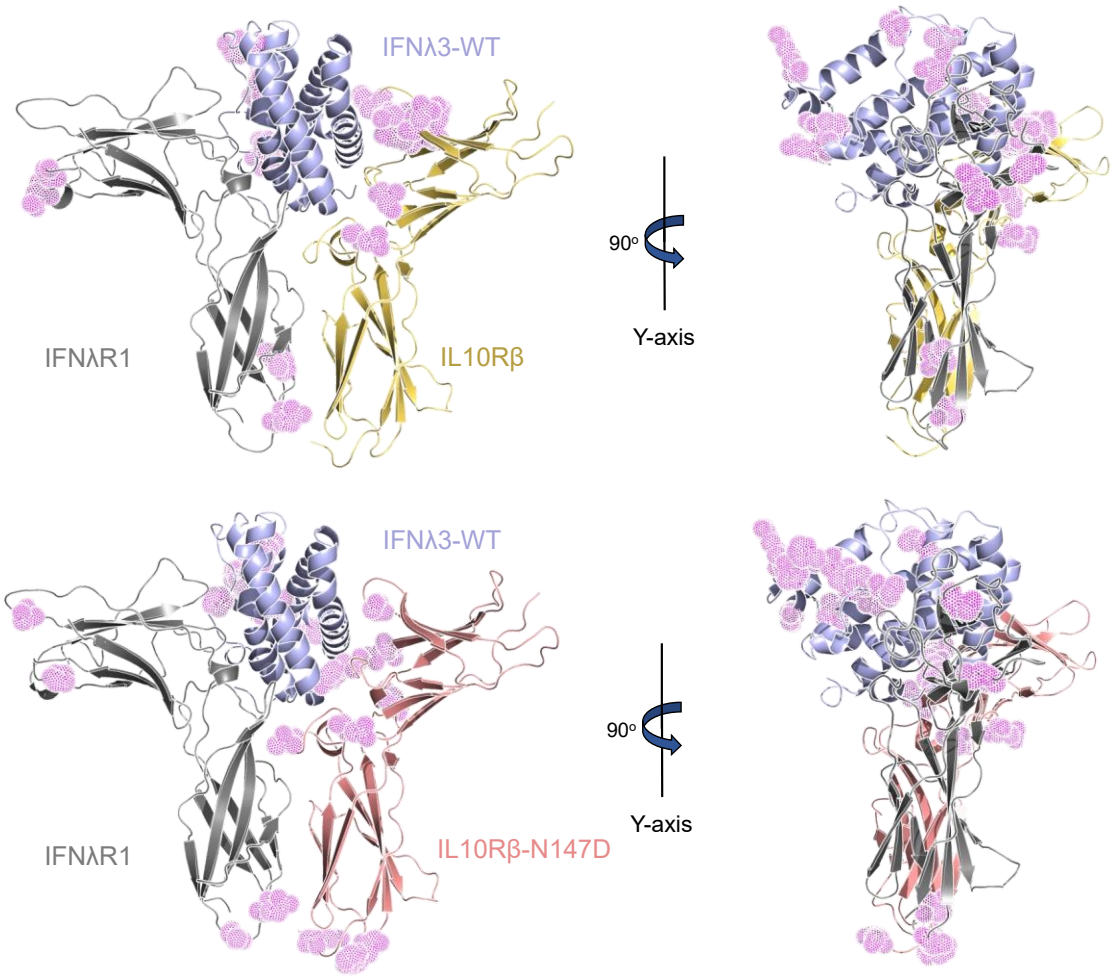
Supplemental Figure 8: Differences in residue contact maps for the wild-type and engineered IFNλ3 complexes reveal changes in protein-protein interactions. By taking the difference in contact percentage time (τ), a new contact map shows residue-residue contacts that are favored for the wild-type ($\Delta\tau > 0$, red) or engineered ($\Delta\tau < 0$, black) IFNλ3 complexes. $\Delta\tau$ with absolute values greater than 0.25 were deemed significant for analysis. All data is available by request.



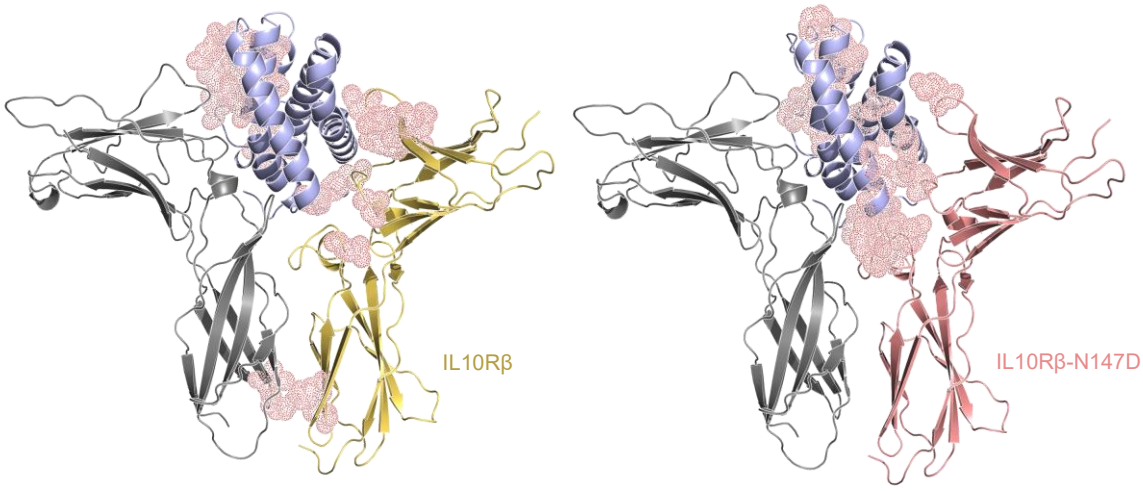
Supplemental Figure 9: Favored intra-protein contacts identified by molecular simulation for the WT and H11 ternary complexes. Significant contacts ($\Delta\tau > 0.25$) within proteins are shown in pink.



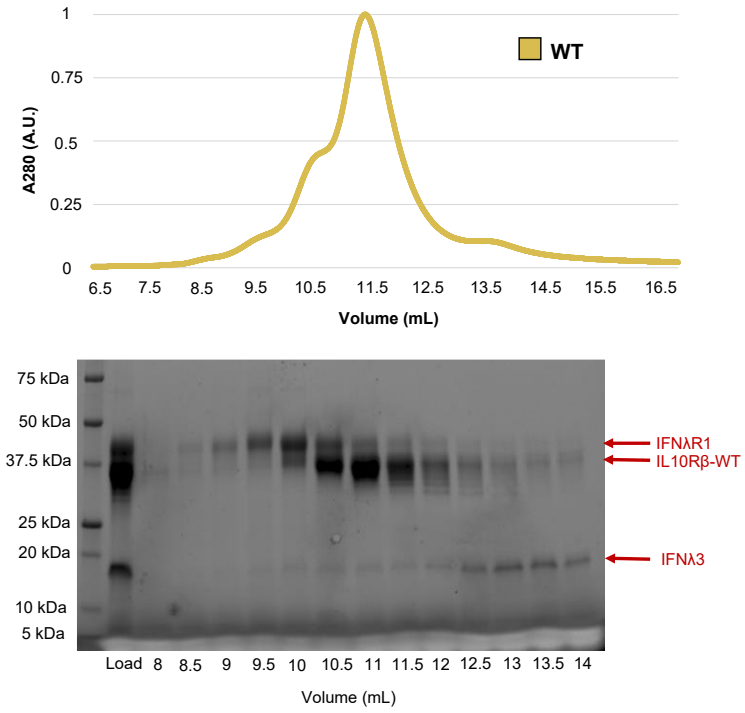
Supplemental Figure 10: Mutations in IFNλ3-H11 initiate a global conformational change. Superimposed structures of the engineered protein complex as it moves throughout 3 positions in free energy space are shown with front and side views (1. local free energy minima, 2. transition point, 3. global free energy minima – Transition occurs in direction of gold arrow on plot). As residue 15 twists (blue), a proximal loop of IL10Rβ pulls towards the ligand. This movement induces a conformational change throughout IFN*λ*3-H11 as residue 15 rotates towards IL10Rβ and leads to the formation of a novel secondary structure emerging at the distal end of the ligand (side view, gold arrow). Initial and final conformations are shown.



Supplemental Figure 11: Favored intra-protein contacts identified by molecular simulation for the WT and IL10Rβ-N147D ternary complexes. Significant contacts ($\Delta\tau > 0.25$) within proteins are shown in pink.



Supplemental Figure 12: Top 20 residue-residue interactions identified by molecular simulation for the WT and IL10Rβ-N147D ternary complexes. The 20 residue-residue pairs with the highest values of $\Delta\tau$ are shown in red.



Supplemental Figure 13: IL10Rβ-N147D enables purification of the ternary complex formed with wild-type IFNλ3 and IFNλR1. Size exclusion chromatograms (Superdex S75 column) of the wild-type IFNλ3/IFNλR1/IL10Rβ ternary complex (top). Volume fractions are shown on SDS-PAGE protein gel (bottom).

IFNλ3-WT (Δτ)

IL10Rβ:IL10Rβ	IFNλR1:IFNλR1	IFNλ3:IFNλ3	IL10Rβ:IFNλR1	IL10Rβ:IFNλ3	IFNλR1:IFNλ3
His85:Tyr59 (0.7413)	Leu134:Leu127 (0.5599)	Arg115:Glu56 (0.6797)	Asn147:Pro157 (0.5852)	Tyr82:His94 (0.8476)	Pro132:Val163 (0.3799)
Tyr59:Tyr56 (0.6875)	Pro133:Tyr125 (0.5093)	Arg115:Gln53 (0.5938)	Lys146:Pro157 (0.4741)	Tyr82:Leu93 (0.7909)	Leu134:Ser11 (0.3136)
Thr86:Tyr59 (0.6283)	Pro133:Gln126 (0.5006)	Arg115:Leu52 (0.5713)	Tyr173:Pro161 (0.4709)	Tyr82:Gln88 (0.7707)	Met131:Arg151 (0.2998)
His85:Ser80 (0.5751)	Pro133:Pro128 (0.4804)	Cys136:Ser40 (0.4922)	Val177:Thr165 (0.4521)	Gly83:Leu93 (.7413)	Pro132:Cys162 (0.296)
Leu79:Ser58 (0.5539)	Pro133:Leu127 (0.4441)	Gly116:Leu52 (0.4891)	Asp174:Pro161 (0.4505)	Ser80:His91 (0.596)	-

IFNλ3-H11 (Δτ)

IL10Rβ:IL10Rβ	IFNλR1:IFNλR1	IFNλ3:IFNλ3	IL10Rβ:IFNλR1	IL10Rβ:IFNλ3	IFNλR1:IFNλ3
Tyr82:Val148 (0.6722)	Leu72:Pro132 (0.5391)	Gly116:Trp121 (0.7803)	Glu121:Leu166 (0.3102)	Tyr59:Leu93 (0.7355)	Pro133:Ala150 (0.591)
Ser80:Glu139 (0.6339)	Leu100:Pro132 (0.5138)	Ala110:Thr114 (0.6802)	-	Ser58:Thr92 (0.7077)	Asp98:Lys21 (0.4232)
Ala124:Thr185 (0.5392)	Tyr73:Pro132 (0.4496)	Pro58:Pro106 (0.6587)	-	Ser80:Leu12 (0.6529)	Pro133:Cys155 (0.3725)
Leu123:Pro183 (0.5174)	Ser18:Met131 (0.4085)	Val59:Pro106 (0.6326)	-	Ser80:Pro14 (0.5979)	Pro133:Asn154 (0.3667)
Leu123:Leu181 (0.5131)	Leu166:Ala170 (0.3775)	Val59:Gln107 (0.6249)	-	Ser80:Leu17 (0.5473)	Phe184:Ala150 (0.3318)

Supplemental Table 4: Top 5 differences in Δτ for the IFNλ3-WT and IFNλ3-H11 complexes by protein-protein interaction. Top 5 differences in Δτ by protein-protein interaction are grouped for the wild-type (WT) and engineered (H11) complexes. Residues are written to correspond to the order of proteins as written in the columns (i.e. Protein1:Protein2 corresponds to ResidueProtein1:ResidueProtein2). Values of Δτ are included in parentheses.

IL10Rβ-WT

Protein 1	Protein 2	Residue 1	Residue 2	Δτ
IFNλ3	IFNλ3	Cys136	Ser40	0.6791
IFNλ3	IFNλ3	Arg115	Glu56	0.6372
IFNλ3	IFNλ3	Arg115	Gln53	0.5816
IFNλ3	IFNλ3	Ser140	Ser40	0.5513
IFNλ3	IFNλ3	Arg115	Leu52	0.5417
IL10Rβ	IL10Rβ	Asp64	Ser58	0.4931
IFNλ3	IFNλ3	Val141	Leu42	0.4895
IFNλ3	IFNλ3	Leu137	Ser40	0.4696
IFNλ3	IFNλ3	Ser40	Lys37	0.4571
IL10Rβ	IL10Rβ	Glu137	Ser77	0.4308
IFNλ3	IFNλ3	Trp121	Thr46	0.425
IL10Rβ	IL10Rβ	Gln63	Tyr59	0.419
IFNλ3	IFNλ3	Phe143	Phe43	0.403
IFNλR1	IL10Rβ	Glu117	Ser126	0.4002
IFNλ3	IFNλ3	Ala139	Arg41	0.392
IL10Rβ	IL10Rβ	Ser58	Gln55	0.3851
IL10Rβ	IL10Rβ	Tyr59	Tyr56	0.3827
IFNλ3	IL10Rβ	Ser13	Tyr82	0.3827
IFNλ3	IFNλ3	Ala139	Leu42	0.3806
IFNλR1	IL10Rβ	Ile118	His128	0.3771

IL10Rβ-N147D

Protein 1	Protein 2	Residue 1	Residue 2	Δτ
IFNλ3	IFNλ3	Trp121	Gly116	0.584
IFNλ3	IFNλ3	Leu137	Pro44	0.5309
IFNλ3	IL10Rβ	Ala19	Ser80	0.5158
IFNλ3	IFNλ3	Glu64	Pro44	0.5156
IFNλ3	IL10Rβ	Ser13	Tyr140	0.4907
IFNλ3	IFNλ3	Arg120	Arg115	0.4564
IFNλ3	IFNλ3	Pro108	Val59	0.4543
IFNλ3	IFNλ3	Glu132	Phe43	0.4509
IFNλ3	IFNλ3	Ala128	Phe43	0.4354
IFNλ3	IL10Rβ	Lys10	Val148	0.4291
IFNλ3	IL10Rβ	Gln88	Arg60	0.4185
IFNλ3	IFNλ3	Ala139	Asp35	0.404
IFNλ3	IL10Rβ	Leu86	Arg60	0.4023
IFNλ3	IFNλ3	Glu138	Asp35	0.4021
IFNλ3	IFNλ3	Leu137	Asp35	0.394
IFNλ3	IFNλ3	Trp121	Leu42	0.3886
IFNλ3	IL10Rβ	Glu16	Glu139	0.388
IFNλ3	IL10Rβ	Lys10	Asp147	0.3867
IFNλ3	IL10Rβ	Ser13	Tyr143	0.386
IFNλ3	IL10Rβ	Pro14	Asn138	0.3841

Supplemental Table 5: Top 20 overall differences in Δτ for the IL10Rβ-WT and IL10Rβ-N147D complexes by protein-protein interaction. Top 20 overall differences in Δτ by protein-protein interaction are grouped for the wild-type IL10Rβ-WT and engineered IL10Rβ-N147D complexes.

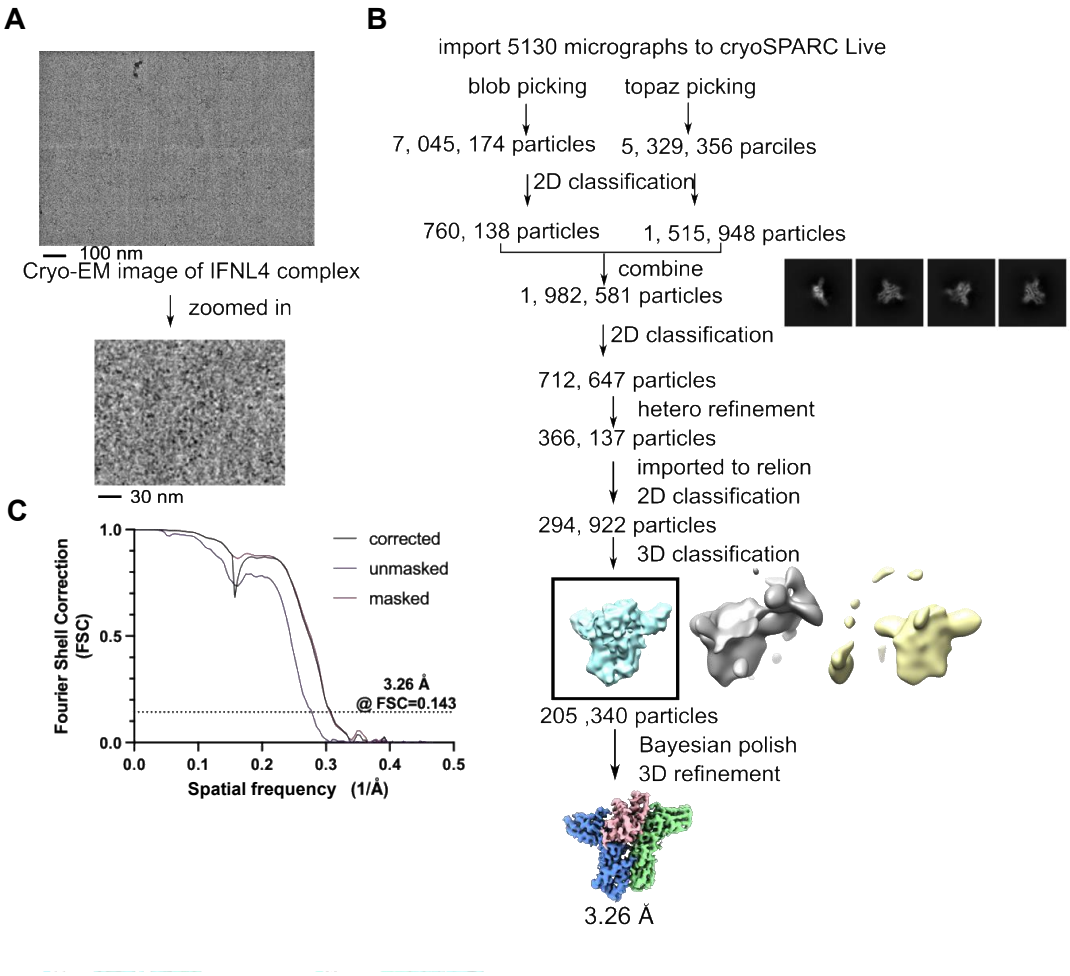
Protein displayed on yeast	Background (R.F.U.)			IL10Rβ (R.F.U.)			IL10Rβ-N147D (R.F.U.)		
IFNλ1	225	219	230	339	329	295	674	768	765
IFNλ2	194	214	202	994	1086	1178	1403	1239	1479
IFNλ3	221	286	220	1212	1092	1046	3016	2585	2985

Supplemental Table 6: Fluorescent values for ternary complex formation as measured by yeast surface display. Relative fluorescent units (RFUs) are shown in a table as measured via flow cytometry. Values were obtained as described in the methods and used to calculate fold-increases in ternary protein complex relative to the wild-type receptor.

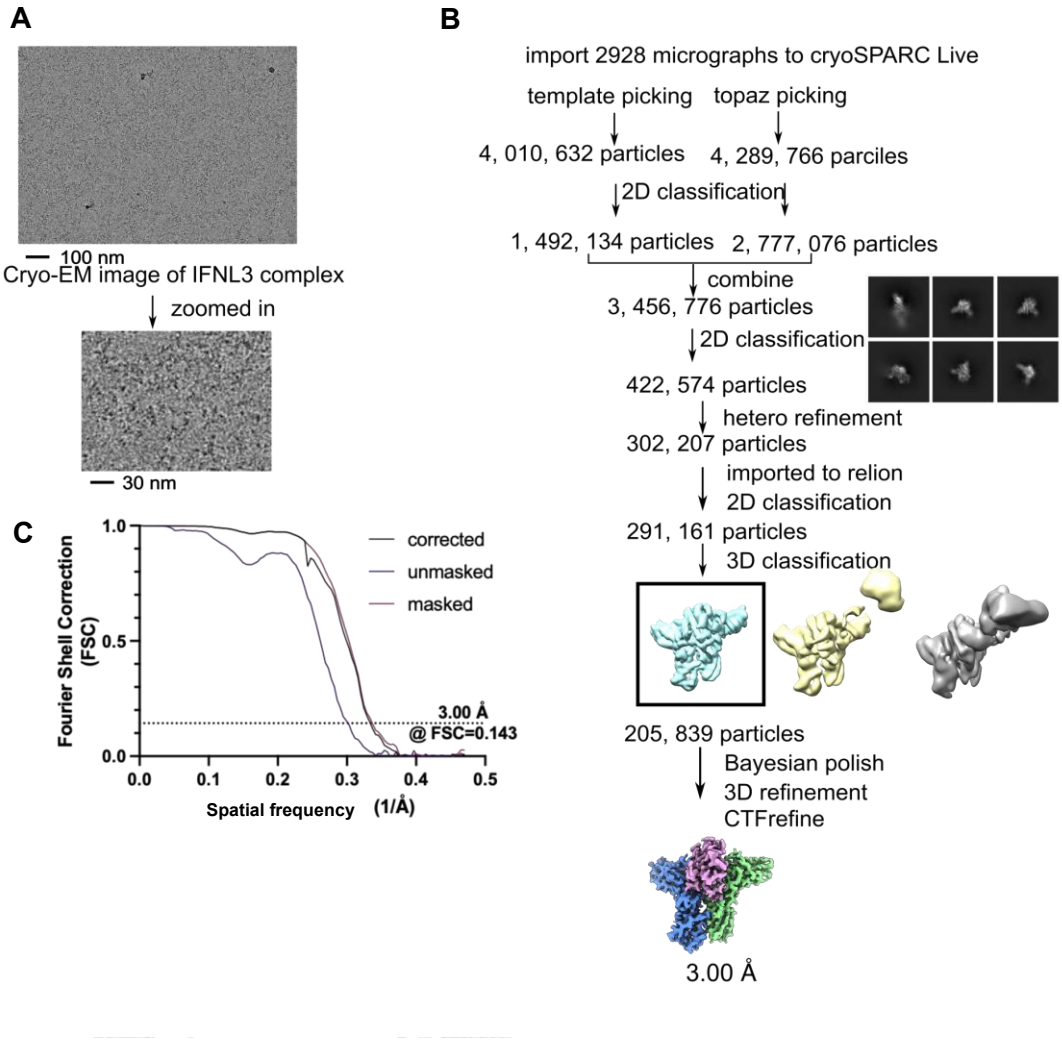
Appendix C: Supplementary information for Chapter 4

	IFNλ4	IFNλ3
Data collection and processing		
Microscope	Krios (University of Chicago)	Krios (University of Chicago)
Magnification	81,000	81,000
Voltage (kV)	300	300
Spherical aberration (mm)	2.7	2.7
Detector	K3	K3
Camera mode	Super resolution counting	Super resolution counting
Exposure rate (e^-/pixel/s)	15	15
Total exposure (e^-/Å^2)	60	60
Defocus range (μm)	-0.9 to -2.3	-0.9 to -2.3
Pixel size (Å)	1.068	1.065
Mode of data collection	Image shift	Image shift
Energy filter	20 eV slit	20 eV slit
Software for data collection	EPU	EPU
Number of micrographs	5,130	2,928
Refinement		
Model resolution (Å)	3.26	3.00
FSC threshold	0.143	0.143
Map sharpening B factor (Å^2)	-95.43	-85.22
Model composition		
Non-hydrogen atoms	4319	4413
Protein residues	529	542
Ligands	5	2
B factors (Å^2)		
Protein	57.31	59.46
Ligand	43.83	74.00
R.m.s. deviations		
Bond lengths (Å)	0.005	0.003
Bond angles (°)	0.745	0.658
Validation		
MolProbity score	1.82	1.49
Clashscore	8.90	7.91
Poor rotamers (%)	0.64	0.41
Ramachandran plot		
Favored (%)	95.01	97.75
Allowed (%)	4.99	2.25
Disallowed (%)	0	0

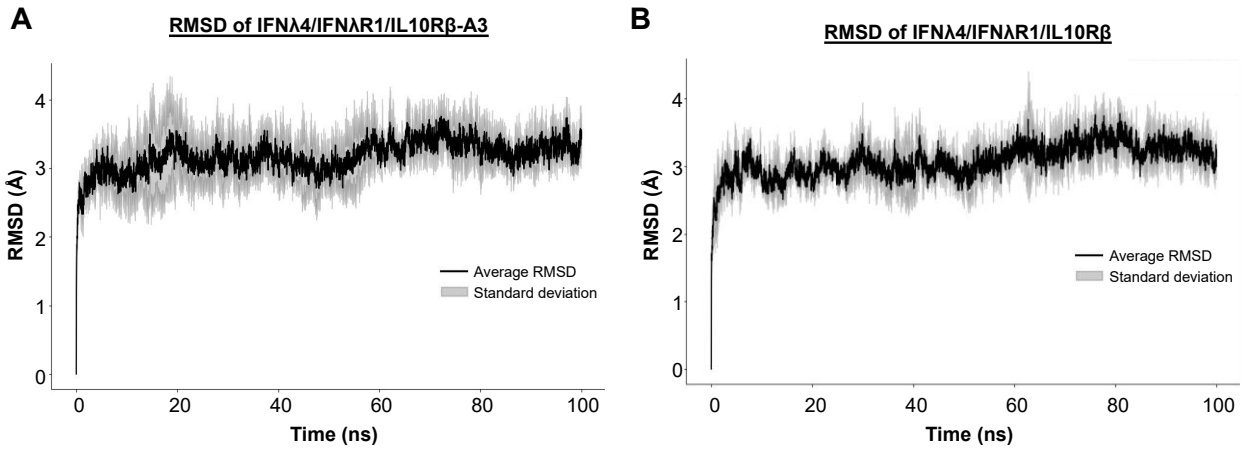
Supplemental Table 7: Information regarding data collection, processing, and refinement for cryoEM of the IFNλ4 and IFNλ3 ternary complexes.



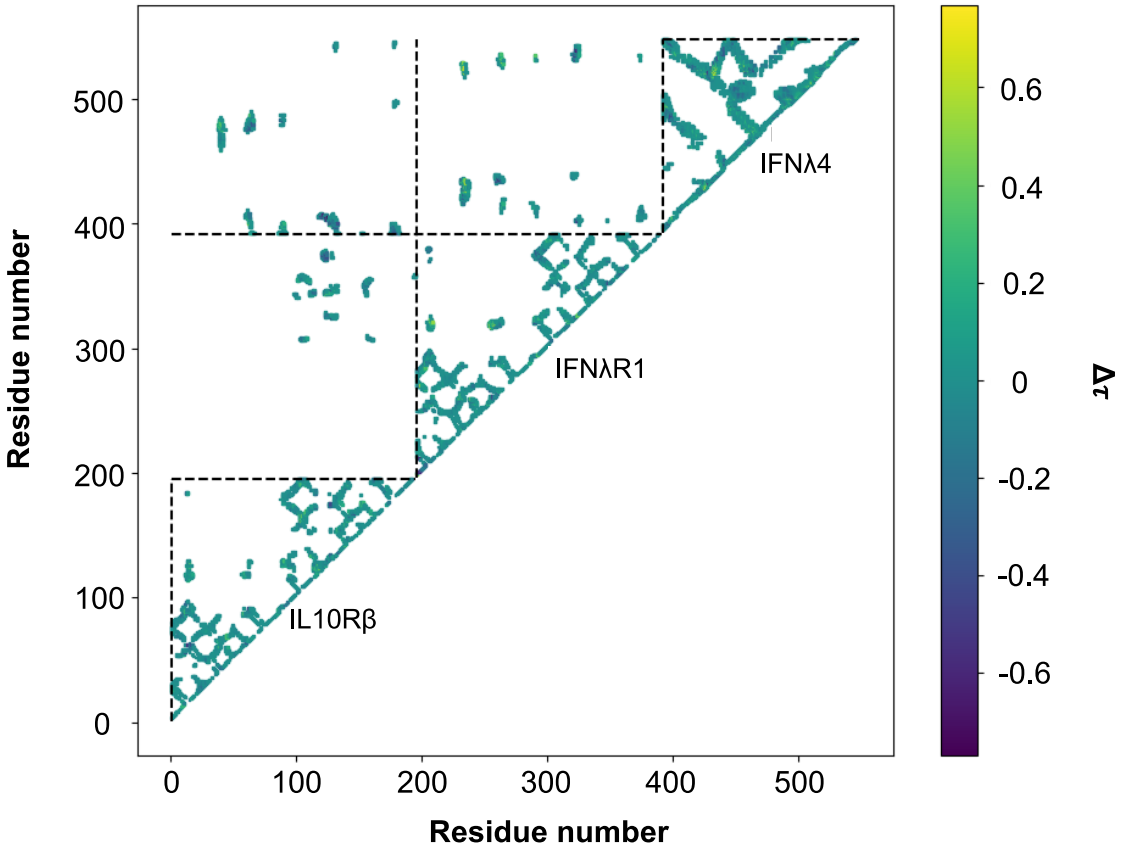
Supplemental Figure 14: Overview of the workflow, particle picking, and model refinement of the IFNλ4 receptor complex. A) Image of the IFNλ4 complex on grids. **B)** Step-by-step schematic detailing the solution of the IFNλ4 receptor complex to a final resolution of 3.26 Å. (Blue = IL10Rβ-A3, Red = IFNλ4, Green = IFNλR1) **C)** Fourier Shell Correction (FSC) versus spatial frequency for the final model of the IFNλ4 receptor complex.



Supplemental Figure 15: Overview of the workflow, particle picking, and model refinement of the IFNλ3 receptor complex. A) Image of the IFNλ3 complex on grids. **B)** Step-by-step schematic detailing the solution of the IFNλ3 receptor complex to a final resolution of 3.26 Å. (Blue = IL10Rβ-A3, Red = IFNλ3, Green = IFNλR1) **C)** Fourier Shell Correction (FSC) versus spatial frequency for the final model of the IFNλ3 receptor complex.

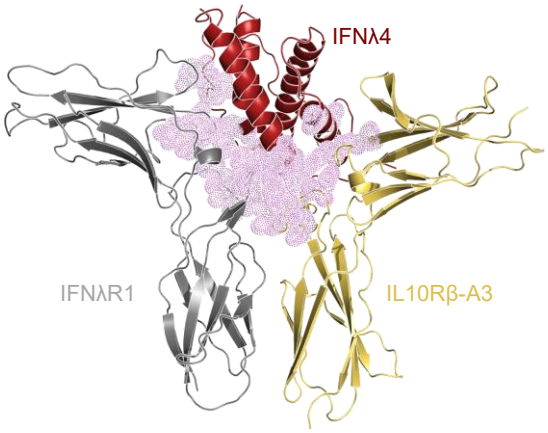


Supplemental Figure 16: RMSD analysis of the molecular dynamics production runs for the IFNλ4 complexes. A) RMSD analysis of the simulated IFNλ4 complex is shown to demonstrate equilibration of the system. **B)** RMSD analysis of the simulated IFNλ4 model containing the wild-type IL10Rβ is shown.



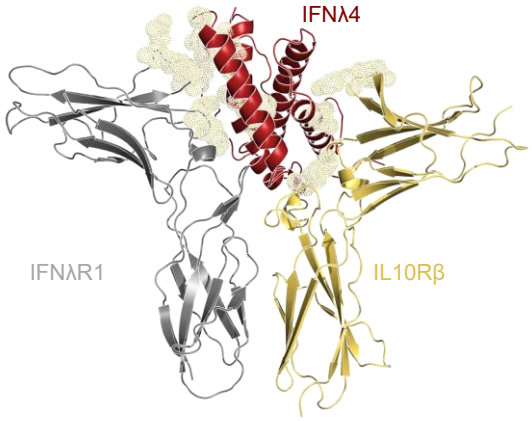
Supplemental Figure 17: Differences in residue contact maps for the wild-type and engineered IFNλ4 complexes reveal changes in protein-protein interactions. Residue contact map showing residue-residue contacts that are favored for the wild-type ($\Delta\tau > 0$, yellow) or engineered ($\Delta\tau < 0$, purple) IFNλ4 complexes.

Protein 1	Protein 2	Residue 1	Residue 2	Δ
IFNλ4	IL10Rβ	GLU36	THR144	0.7677
IFNλ4	IL10Rβ	PRO37	THR142	0.6951
IFNλ4	IL10Rβ	LEU35	TRP143	0.689
IFNλ4	IFNλR1	ARG67	ASP71	0.6874
IFNλ4	IFNλR1	ARG170	PRO133	0.6483
IFNλ4	IFNλR1	PRO66	LEU72	0.6369
IFNλ4	IL10Rβ	ARG38	TYR140	0.6225
IFNλ4	IFNλR1	LEU171	PRO133	0.5998
IFNλ4	IL10Rβ	SER34	TRP143	0.5724
IFNλ4	IL10Rβ	ARG38	ASN138	0.5423
IFNλ4	IFNλR1	ARG67	TYR73	0.5212
IFNλR1	IL10Rβ	ASP135	ASP147	0.5099
IFNλ4	IFNλR1	ARG163	PRO132	0.5012
IFNλR1	IL10Rβ	VAL186	THR142	0.4941
IFNλ4	IFNλR1	THR166	PRO133	0.4925
IFNλ4	IFNλR1	SER34	LEU134	0.4871
IFNλ4	IFNλR1	THR166	PRO132	0.4838
IFNλ4	IFNλR1	ARG33	LEU134	0.4781
IFNλR1	IL10Rβ	SER185	TYR140	0.464
IFNλ4	IFNλR1	ARG170	LEU134	0.4584
IFNλ4	IL10Rβ	GLU36	TYR140	0.4574
IFNλ4	IFNλR1	ARG65	GLN70	0.4457
IFNλR1	IL10Rβ	PHE184	TRP143	0.4437
IFNλ4	IL10Rβ	ALA113	ASP84	0.4363
IFNλ4	IL10Rβ	TYR32	ASP147	0.4339
IFNλ4	IL10Rβ	PRO37	TRP143	0.432
IFNλ4	IL10Rβ	PRO107	GLY83	0.4223
IFNλ4	IL10Rβ	PRO107	TYR82	0.4132

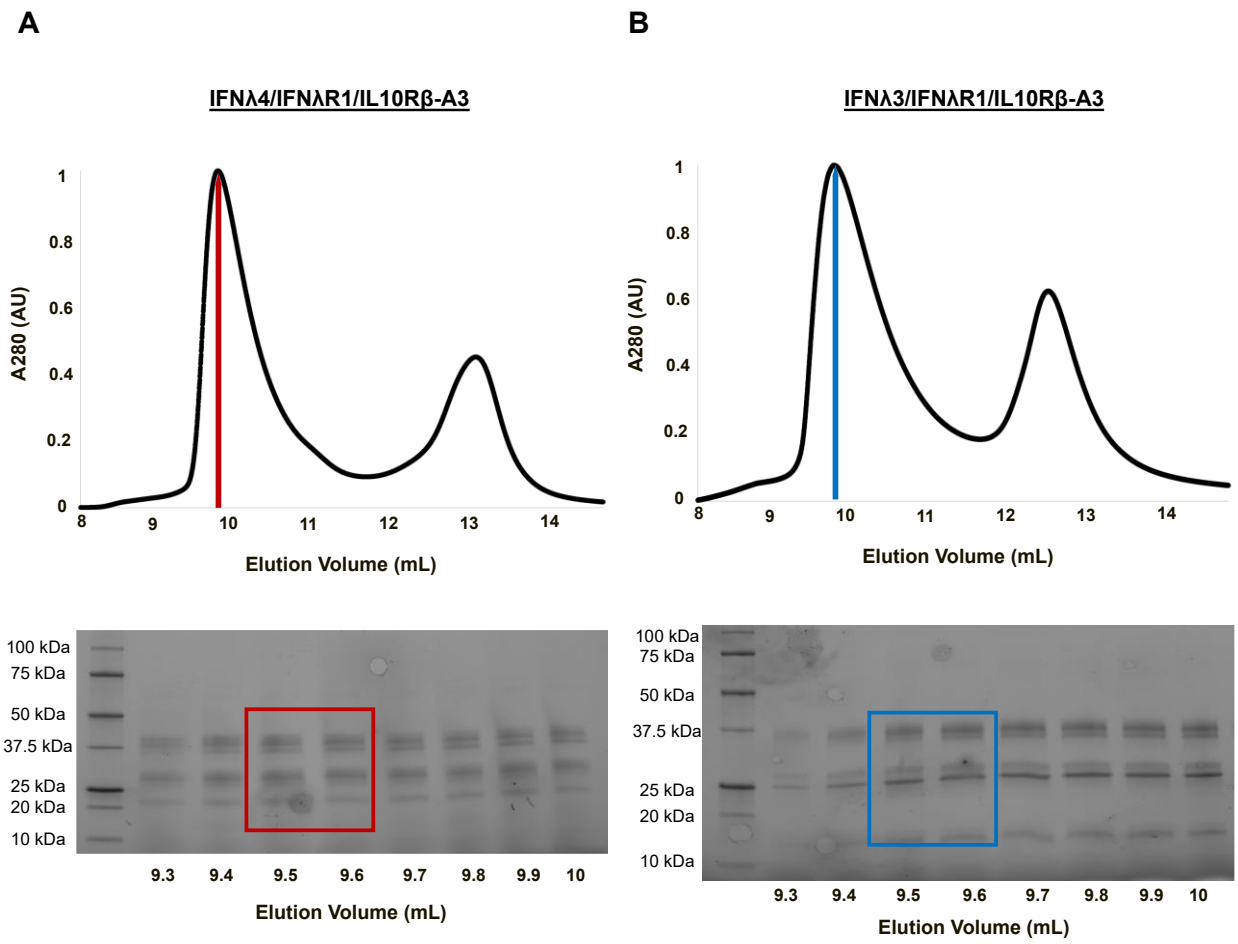


Supplemental Figure 18: Protein-protein interactions favored by the engineered IFNλ4 complex. A table detailing protein-protein interactions, specific residue-residue interactions, and the magnitudes of differences in residue contact time ($\Delta\tau$) are provided. Data represent protein-protein interactions ranking in the top 50 of overall contact differences. These values are presented on the structure (right) using purple surfaces.

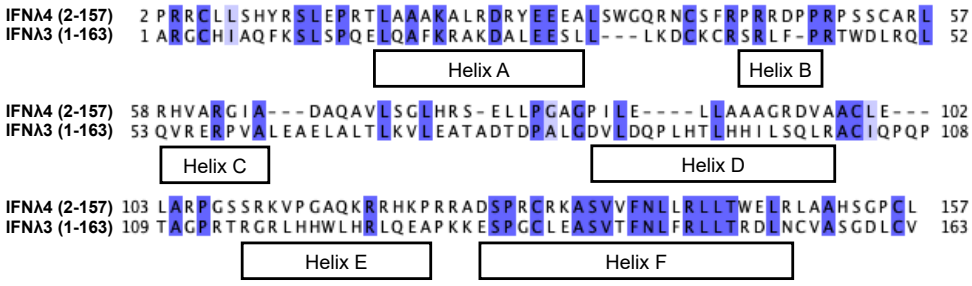
Protein 1	Protein 2	Residue 1	Residue 2	Δ
IFNλ4	IFNλR1	SER156	THR44	0.7686
IFNλ4	IFNλR1	ALA155	THR44	0.6208
IFNλ4	IFNλR1	PHE64	ARG45	0.6185
IFNλ4	IFNλR1	SER63	ARG46	0.6015
IFNλ4	IFNλR1	CYS152	THR44	0.5992
IFNλ4	IL10Rβ	HIS31	LYS81	0.5864
IFNλ4	IFNλR1	GLY58	ARG46	0.5851
IFNλ4	IFNλR1	ARG65	ARG45	0.5764
IFNλ4	IFNλR1	ALA155	PRO43	0.5693
IFNλ4	IFNλR1	LEU161	ASN74	0.548
IFNλ4	IFNλR1	SER63	ARG45	0.5166
IFNλ4	IFNλR1	ASN160	PRO43	0.4972
IFNλ4	IFNλR1	LEU164	TYR73	0.461
IFNλ4	IFNλR1	GLY58	ARG45	0.459
IFNλ4	IL10Rβ	ALA113	TYR59	0.445
IFNλ4	IFNλR1	SER156	PRO43	0.436
IFNλ4	IL10Rβ	GLU110	ARG60	0.4266



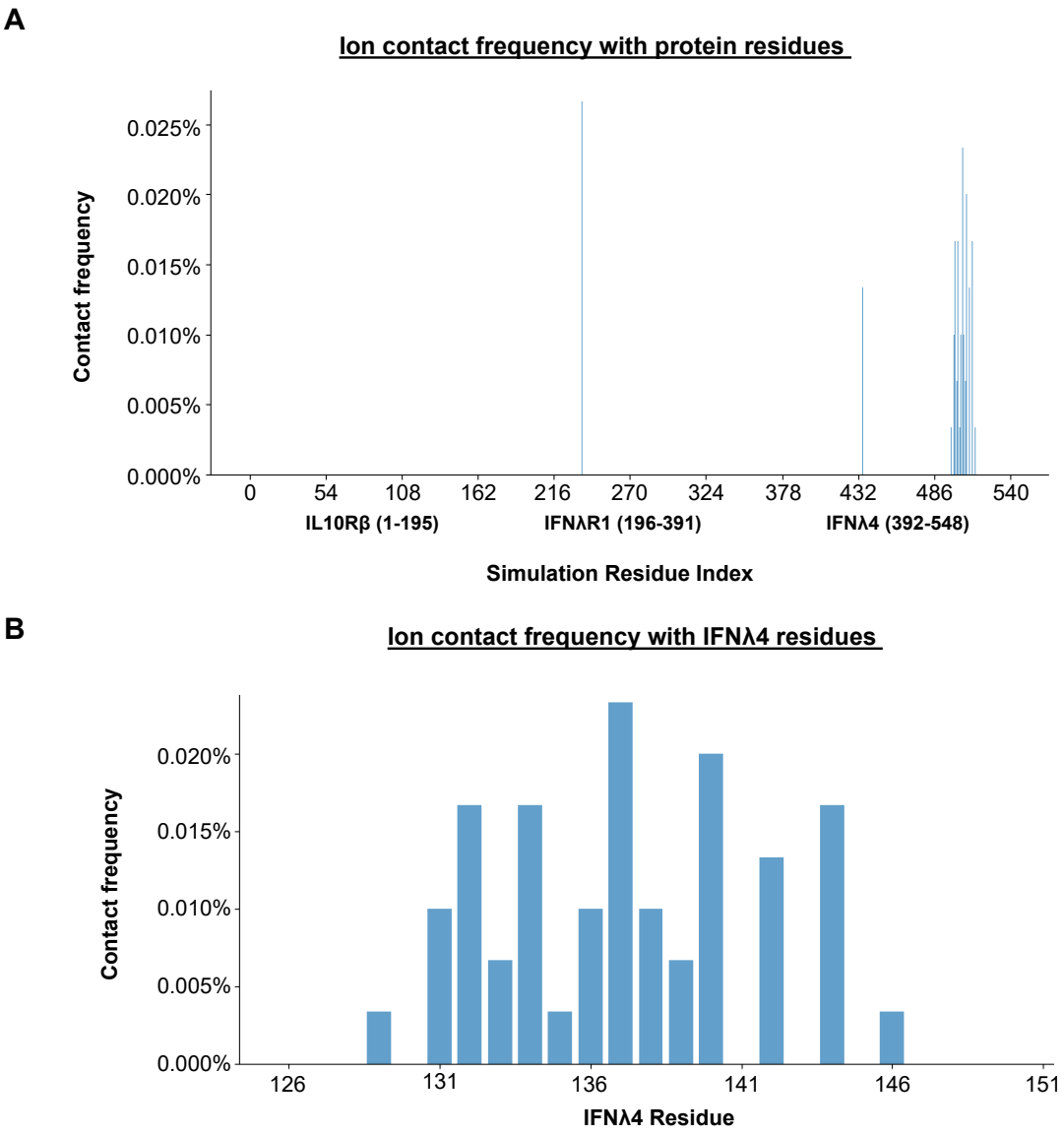
Supplemental Figure 19: Protein-protein interactions favored by a model of the wild-type IFNλ4 complex. A table detailing protein-protein interactions, specific residue-residue interactions, and the magnitudes of differences in residue contact time ($\Delta\tau$) are provided. Data represent protein-protein interactions ranking in the top 50 of overall contact differences. These values are presented on the structure (right) in gold surfaces.



Supplemental Figure 20: IL10Rβ-A3 enables purification of the ternary complexes formed with IFNλ4:IFNλR1 and IFNλ3:IFNλR1. A) Size exclusion chromatograms (Superdex S75 column) of the IFNλ4/IFNλR1/IL10Rβ-A3 ternary complex (top). Volume fractions are shown on SDS-PAGE protein gel (bottom). Red area and square indicate the fractions used for data collection. The peak at ~13 mL is unbound excess IL10Rβ-A3. **B)** Size exclusion chromatograms of the IFNλ3/IFNλR1/IL10Rβ-A3 ternary complex (top). Volume fractions are shown on SDS-PAGE protein gel (bottom). Blue area and square indicate the fractions used for data collection. The peak at ~13 mL is unbound excess IL10Rβ-A3.



Supplemental Figure 21: Amino acid sequence alignment of IFNλ4 and IFNλ3. Helix labels are placed underneath the corresponding amino acid sequences. Residues highlighted in dark blue are identical. Residues highlighted in light blue are highly similar.



Supplemental Figure 22: Negative ions associate with the disordered and positively-charged Helix E of IFNλ4 throughout molecular dynamics simulations. A) Ion contact frequency for all protein residues throughout simulation time. **B)** Ion contact frequency for residues on Helix E of IFNλ4.

Protein	Genetic sequence
xIFNλ4	GCTCCAAGAAGATGTTTGTTGTCTCACTACAGATCTTTGGAACCAAGAA CTTTGGCTGCTGCTAAGGCTTTGAGAGACAGATACGAAGAAGAAGCTTT GTCTTGGGGTCAAAGAAACTGTTCTTTCAGACCAAGAAGAGACCCACCA AGACCATCTTCTTGTGCTAGATTGAGACACGTTGCTAGAGGTATCGCTG ACGCTCAAGCTGTTTTGTCTGGTTTGCACAGATCTGAATTGTTGCCAGGT GCTGGTCCAATCTTGGAATTGTTGGCTGCTGCTGGTAGAGACGTTGCTG CTTGTTTGGAATTGGCTCCAACTGCTGGTCCAAGAACTAGAGGTAGATT GCACCACTGGTTGCACAGATTGCAAGAAGCTGCTGACTCTCCAAGATGT AGAAAGGCTTCTGTTGTTTTCAACTTGTTGAGATTGTTGACTTGGGAATT GAGATTGGCTGCTCACTCTGGTCCATGTTTG

Supplemental Table 8: Genetic sequence of the chimeric IFNλ4 (xIFNλ4). Chimeric region of the protein highlighted in blue.