Computational Methods for Designing and Probing Intermolecular Interactions

by
Sophia Tan

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
William DeGrado
582727E949C7441...

William DeGrado
Chair

DocuSigned by:
Tanja Kortemme

Tanja Kortemme

DocuSigned by:
Matthew Jacobson

Matthew Jacobson

DocuSigned by:
Michael Keiser
4DF1BD06D670465...

Michael Keiser

Committee Members

ACKNOWLEDGEMENTS

I've had the incredible fortune of meeting and working with many extraordinary individuals at UCSF. First and foremost is my advisor, Bill DeGrado. During my graduate career, I faced a devastating family situation, and Bill could not have been more supportive. I'm eternally grateful for his mentorship, and I'm so lucky to have worked in an environment where he nurtured and inspired such exciting science.

I would also like to thank past and present lab members, particularly Katie Hatstat, Nina Hentzen, Hyunil Jo, Rian Kormos, Huong Kratochvil, Sam Mann, Alison Maxwell, Marco Mravic, Nick Polizzi, Lee Schnaider, and especially Hyun Jun Yang.

I'm also extremely thankful to several faculty members, starting with my thesis committee members: Tanja Kortemme, Matt Jacobson, and Mike Keiser. It still seems unreal to me that these scientific experts I greatly admire sat on my committee, and I'm grateful that they've always had my best interests at heart. I would also like to thank Xiaokun Shu, Michael Grabe, and Joel Bennett (UPenn), who I'll always remember for not only our integrin collaboration, but also his exceptional kindness in connecting me with a medical expert as my family crisis was unfolding.

I must also thank my initial scientific mentors: my summer undergraduate internship advisor, Neil P. Shah (who inspired me to change my career path from pharmacy to research) and my undergraduate advisor, Sagar Khare (who, along with my graduate mentor Nancy Hernández, essentially taught me how to look at proteins in the first place).

Finally, my graduate career is defined by the love and encouragement I've received from my friends and family. I'm deeply grateful to Precita House, my 3-in-1 housemates, classmates, and closest friends: Andrew Natale, Keely Oltion, Matvei Khoroshkin, Yessica Gomez, and Matan Appelbaum. Thank you especially to Andrew for unwavering emotional support, technical support, and everything in between. I would also like to thank Alan Cheng, who allowed me to cheerily tag along on several adventures, both scientific and non-scientific. Most of all, I am overwhelmingly grateful to my family. Thank you to my mom, my sister, my extended family, and most importantly, my Matan, to whom I owe everything.

CONTRIBUTIONS

Chapter 2 incorporates material from a previously-published journal article:

Sophia Tan, Karen P. Fong, Nicholas F. Polizzi, Alex Sternisha, Joanna S. G. Slusky, Kyungchul Yoon, William F. DeGrado, and Joel S. Bennett. "Modulating integrin αIIbβ3 activity through mutagenesis of allosterically regulated intersubunit contacts." *Biochemistry* 58.30 (2019): 3251-3259.

Computational Methods for Designing and Probing Intermolecular Interactions

Sophia Tan

ABSTRACT

This dissertation presents methods developed for two biological systems. In the first chapter, we describe computational methods for designing a *de novo* kinase reporter. Our reporter operates as a switch, undergoing a change in oligomeric state in response to kinase activation. We designed the reporter to induce liquid-liquid phase separation upon substrate phosphorylation, making it the first example of a *de novo* designed protein switch capable of forming biomolecular condensates. In the second chapter, we introduce a simple knowledge-based metric to assess the strengths of residue-residue interactions in a protein complex interface. This structural bioinformatics-based method precludes the need for modeling mutation-induced structural and energetic effects. We tested this method on an integrin interface for which current computational alanine scanning methods were not able to accurately rank hot spot residues according to the degree they destabilize the integrin complex. Encouragingly, we found that our metric could more accurately differentiate the subtle energetics of hot spot residues that regulate integrin interfacial stability. Together, these projects were developed to test the limits of current design methods and discover new design rules for protein binding.

Table of contents

List of figures

List of tables

List of equations

Chapter 1:

Computational *de novo* design of phosphorylation-responsive switches

INTRODUCTION

Signal transduction reporters are essential for studying cell states and signaling dynamics across all fields of biological studies. Recently, Zhang & Shu et al. developed a reporter that induces phase separation in response to kinase activation (Zhang et al. 2018). The reporter, termed SPARK (separation of phases-based activity reporter of kinase), was demonstrated to achieve high fluorescence and brightness; this is attributed to liquid-liquid phase separation (LLPS), which concentrates the EGFP-containing reporters into dense liquid droplets. We were interested in developing a computational pipeline to generalize reporter design for other kinases, and even other post-translational modifications, so we began working with the authors to develop the first *de novo* designed cell signaling reporter that induces LLPS. The work presented here details a computational method for designing a proof-of-concept system that responds to activation of Protein Kinase A (PKA).

De novo protein design is an ever-evolving field in which computational methods of protein modeling and sampling are used to develop new proteins that do not exist in nature (Huang et al. 2016, Korendovych & DeGrado 2020, Pan & Kortemme 2021). Such methods rely on our understanding of biophysical first principles, the fundamental

physical and chemical principles that underlie protein folding and function. In designing proteins *de novo*, we task ourselves with the challenge of designing non-natural proteins from scratch, because it tests our computational representations, assumptions, and simplifications of those biophysical first principles. In *de novo* protein design, we're constantly learning what the minimal requirements are for modeling protein structures and properties, and constantly updating where those representations fall short.

In addition to testing our understanding of theoretical biophysical first principles, *de novo* protein design also enables many practical applications, including catalysis and regulation of therapeutic targets. For *in vivo* applications, *de novo* protein design allows us the luxury of manipulating biological environments without significant off-target effects; i.e., we can design them to be orthogonal to natural biological pathways. For this reason, *de novo* protein design is an attractive method to design cellular reporters, because they can sense biological and chemical activity without perturbing cellular behavior.

In particular, *de novo* protein design of protein switches presents great challenges to theoretical modeling. The field is relatively mature in representing static protein structures, but we fall short in accurately modeling multiple conformational states. Additionally, it is still difficult to calculate the fine-tuned energetics necessary for designing switch properties such as kinetics, dynamics, and specificity (Alberstein et al. 2022). This project uses *de novo* protein design to develop a peptide switch that responds to a particular signal (kinase activation) and generates a specific response (change in oligomeric state), with the aforementioned advantage that it may probe cellular activity without perturbing natural pathways.

RESULTS

De novo *coiled coils as scaffolds for protein switches*

We have developed two sets of switches that respond to Protein Kinase A, whereby the switches adopt a monomeric, unstructured random coil conformation while the cell is at a baseline state, but self-assemble into an oligomeric coiled-coil conformation in a kinase-activated state. We designed both sets of switches as coiled-coils because they are one of the best-understood structures. Coiled-coils have been extensively characterized, and we can use well-established design rules to manipulate and tune stability and phospho-sensitivity. There are several reasons to design coiled-coils: (1) their parameterization is simple and defined by only a few parameters, making it straightforward to generate and sample around idealized helical bundle topologies, (2) they're easy to thermostabilize because their packing rules are well-defined, and (3) they're "designable" in that multiple diverse sequences can fold into the same structure (Grigoryan & DeGrado 2011, Szczepaniak et al. 2014).

The dynamics of coiled-coil assembly allow us to target two defined "on" and "off" states in the switch mechanism. Fundamentally, peptide coiled-coil assembly is thought to visit 3 distinct states (Fig. 1.1). An unstructured, monomeric random coil may transition to a structured, helical monomer, which in turn drives cooperative association of the helical monomers into a coiled-coil (Boice et al. 1996). Thus, in designing the peptide to be more stable in its alpha-helical conformation, we drive the equilibrium toward coiled-coil assembly.

Incidentally, phosphoserines on the N-terminus of helical peptides stabilize its helical conformation (Smart & McCammon 1999, Andrew et al. 2002). We structurally rationalized this in a previous study (Naudin et al. 2021), in which we surveyed the Protein Data Bank (PDB) (Berman et al. 2000) for phosphoserines at N-termini of helices. We found that the majority of these phosphoserines adopt an N-capping rotamer, which allows their phosphate groups to hydrogen bond to unpaired backbone amides also residing on the N-termini, as illustrated in Figure 1.2.

Based on the increased helical stability imparted by N-terminal phosphoserines, we used serine phosphorylation as a trigger; coiled-coil assembly should only occur if the peptide is phosphorylated by a kinase. We designed these switches to be kinase-responsive by grafting a kinase substrate sequence onto the N-termini of the peptides. Conceptually, kinase phosphorylation of the N-terminal serine stabilizes the helical conformation, biasing the equilibrium toward the oligomeric state, which we coupled to a LLPS readout.

Designing LLPS-inducing phosphoswitches from Lac repressor

We based our first set of designs on the Lac repressor tetramerization domain, a short 20-residue segment that weakly associates into a symmetric antiparallel 4-helix bundle (Fairman et al. 1995). A previous study in our group exploited the marginal stability of this domain to switch between an unphosphorylated, unstructured monomer and a phosphorylated, ordered coiled-coil. As described above, they incorporated the substrate motif (RRXS) (Kreegipuu et al. 1998) of Protein Kinase A onto the N-terminus of the Lac repressor tetramerization domain (Signarvic & DeGrado 2003). *In vitro*

experiments of their designs demonstrated an increase of 4.6 kcal/mol in stability of the phosphorylated over the unphosphorylated state.

In this current study, we improved on robustness of those initial first-geenration designs, thermostabilizing them for use as a kinase reporter in more unpredictable *in vivo* cellular environments. We extended the peptide by a helical turn to (1) provide more surface for intermolecular stabilization, and (2) impart increased helical self-stabilization, thereby driving the equilibrium from a helical monomer toward a tetramer. While only a modest change, the 4-fold symmetric nature of the structure ensures that changes in stability are amplified. We computationally designed the entire sequence of the peptide using Rosetta, a protein modeling and design software suite (Leaver-Fay et al. 2011, Fleishman et al. 2011). Sequences and models of the Rosetta-designed structures that we experimentally tested are found in Table 1.1 and Fig. 1.3, respectively.

Designing staggered, fibril-forming phosphoswitches

For our second set of switches, we wanted to design peptides that self-assemble to form long fibrils by way of self-propagating through symmetric, repeated interactions. These peptides contain 5 helical heptads, staggered in an antiparallel 4-helix bundle arrangement such that two heptads on each terminus overlap with adjacent helices. The length of the overlaps can be tuned to make the fibril more stable or less stable, depending on whether the switches are not responsive to phosphorylation, or if they fibrillize constitutively in the absence of phosphorylation. We again grafted the PKA substrate sequence onto the peptide N-terminus, taking care to ensure that the

phosphoserine is slotted in the junction between two peptides, positioned in a way that it can form intermolecular contacts with adjacent peptides, further stabilizing this topology (Fig. 1.4).

First, we generated 1,050 long, idealized antiparallel 4-helix bundles using the CCCP program (Grigoryan & DeGrado 2011), sampling around parameters found in natural 4-helix bundles. From those scaffolds, we searched for the most "designable" (i.e., that many different sequences could potentially fold into those structures). To find those designable scaffolds, we used MASTER (Zhou & Grigoryan 2015) to search for antiparallel 4-helix bundles in the protein data bank (PDB) (Berman et al. 2000) whose geometries resembled the geometries of the scaffolds.

Next, we grafted 51 phosphoserine-bearing N-terminal helical motifs onto the 35 scaffolds that had the most hits in MASTER, and we incorporated the PKA substrate motif. After selecting for phosphoserines that made at least 2 hydrogen bonds to backbone amides on the N-terminus (excluding hydrogen bonds to the backbone of its own residue) and filtering out structures whose N-terminal phosphoserine motifs clashed with adjacent helices in the bundle, we were left with 782 structures. We then designed the rest of the sequence using ProteinMPNN (Dauparas et al. 2022). The scaffolds generated by CCCP were idealized, poly-glycine structures, so to sample sufficient sequence space, we chose 6 sampling temperatures, generating 50 sequences per sampling temperature per scaffold. In total, ProteinMPNN produced 782*6*50=234,600 designed sequences. AlphaFold2 (Jumper et al. 2021) was then used to predict their structures, and we used the AlphaFold2 results to determine which ProteinMPNN-designed sequences are actually capable of assembling into our target

topology. An overwhelming majority of the designs were predicted to fold into 4-helix bundles that were perfectly aligned perpendicular to the helical axis, meaning that their termini lined up and didn't stagger. Therefore, this filtering step was extremely helpful in reducing the 234,600 designed sequences down to 135. The final 6 sequences we chose to test are listed in Table 1.1.

Experimental characterization

Only the Lac repressor-based designs have been tested to date, though experiments on the staggered coiled-coil designs are ongoing. The tested phosphoswitch sequences were synthesized with both phosphorylated and unphosphorylated variants. Initial CD spectroscopy indicated that 4 of the 6 peptides are more stably helical in the phosphorylated variant, compared to the unphosphorylated variant, as determined by ellipticity at 222 nm measured at pH 7.5 at room temperature. Size-exclusion chromatography (SEC) further indicated that some phosphorylated peptides elute at higher molecular weights compared to their unphosphorylated equivalent peptides, which suggests that they assemble into oligomers. However, the exact stoichiometric composition cannot be determined without methods such as analytical ultracentrifugation (AUC). Out of 6 peptides, 3 of their SEC traces indicated that their phosphorylated variants formed oligomers while their unphoshorylated variants did not, so we advanced those 3 sequences to be tested in cells.

All *in vivo* characterization was performed by the Shu lab (Department of Pharmaceutical Chemistry, UCSF). The full reporter consists of (1) our designed phoshpo-responsive peptides, (2) a previously-characterized homo-oligomeric tag that

forms homotetramers orthogonally to our phoshpo-responsive designs (Zhang et al. 2018), and (3) EGFP (Fig. 1.5). Each component is separated by a flexible linker, so that they can freely diffuse and cross-link with each other, forming the multivalent, high-order interactions necessary to condense into a phase-separated readout (i.e., GFP-dense liquid droplet formation). Initial in-cell experiments show that the reporter forms liquid droplets (well-defined puncta with high fluorescence signal) when PKA is activated with isoprenaline, but not before the isoprenaline treatment. Though we are careful not to make conclusions from these *in vitro* and *in vivo* data, we're encouraged by these preliminary results.

METHODS

Thermostabilizing Lac repressor-based, LLPS-inducing phosphoswitches.

We increased stability of the first-generation peptides in the "on" state by extending the monomer by a helical turn (4 residues) to increase helical self-stability (therefore driving the equilibrium toward tetramer formation), and also to provide surface for intermolecular interactions. To build the model of the Lac repressor-based design, we used a crystallographically-solved structure of the lac repressor (PDB accession code 1TLF) (Friedman et al. 1995) and isolated the 4-helix bundle comprising the tetramerization domain (residues 336-356 of each chain). Using RosettaScripts (Fleishman et al. 2011), we stitched the sequence of the first-generation phosphoswitch (Signarvic & DeGrado 2003) onto the 1TLF structure, and extended the C-terminus by 4 residues. We fixed the identities of the PKA substrate motif, restricted interior-facing residues (a, d positions in helical heptad patterning) to hydrophobic amino acids, and

disallowed Cys, Pro, and Ser (other than the kinase-reactive Ser) from being sampled. The RosettaScripts protocol began with applying a “FavorSymmetricSequence” mover. We chose not to enforce cartesian coordinate symmetry, because there’s no requirement of symmetry in the physical wrld. To enforce sequence symmetry between the four monomers, we applied this FavorSymmetricSequence mover with an arbitrarily high penalty (50,000) to significantly bias the designs toward symmetric sequences.

Next, we applied coordinate constraints (with a standard deviation of 1.5Å) to the scaffold C-alpha atoms and performed 10 rounds of FastDesign. Then, we removed the C-alpha constraints and performed 10 rounds of FastRelax. Amino acid sub-rotamers were sampled at the ex1 and ex2aro chi levels, and the InterfaceRelax2019 relax protocol (Maguire et al. 2020) was specified within the FastDesign and FastRelax movers. Rosetta generated 100 designs, and we selected the final 6 to design based on overall Rosetta score, number of phosphoserine hydrogen bonds to the N-terminal backbone amides, and backbone RMSD to the starting structure. Beause we performed unrestrained FastRelax in the last step of our RosettaScripts protocol, low-quality designs “unraveled” in the sense that strained phosposerines popped out of their N-capping conformations, backbone geometries deviated away from their idealized starting structures, etc. Therefore, filtering on the number of phosphoserine hydrogen bonds and backbone RMSD allowed us to distinguish the most robust designs. The final sequences that were advanced to experimental characterization are listed in Table 1.1.

Designing staggered, fibril-forming phosphoswitches

Whereas the Lac repressor-based designs induce LLPS when incorporated into the SPARK reporter, this second set of designs self-assemble into staggered fibrils. To achieve this topology, we sketched out the structure displayed in Figure 1.4. The phosphoswitch contains 5 helical heptads, labeled 1-5, and are staggered in a way such that two heptads on both termini of the switch overlap with adjacent helices. Long, idealized antiparallel 4-helix bundles were generated using CCCP, which is available on a public web server (Grigoryan & DeGrado 2011). The parameters were sampled around 4-helix antiparallel helical bundles found in nature: 3 values (6.5Å, 7.0Å, 7.5Å) were sampled for the superhelical radius parameter, 7 values (-4.4°, -4.1°, -3.8°, -3.5°, -3.2°, -2.9°, and -2.6°) were sampled for the superhelical frequency parameter, 10 values (0° to 360° with 36° deg intervals) were sampled for the chain-wise alpha-helical phase parameter, and 5 values (-2.5Å, -1.25Å, 0Å, 1.25Å, 2.5Å) were sampled for the Z offset chain-wise parameter. In total, we generated 3*7*10*5 = 1,050 scaffolds.

Next, to determine the most viable scaffolds, we used MASTER (Zhou & Grigoryan 2015) to survey the non-redundant PDB. We extracted a layer of 14 residues from the CCCP-generated coiled-coils (14 residues on each chain) and searched for matches in MASTER that met a 1Å RMSD criterion. Scaffolds that had at least 30 matches were determined to be sufficiently designable, and we proceeded with the 35 scaffolds that passed that filter.

To achieve the staggered effect, we removed every 6th helical heptad. Because the chains are antiparallel, the 6th heptad voids are separated by two heptads where

there is complete overlap between the 4 chains. Then, we grafted the 51 N-terminal phosphoserine-bearing motifs that we curated in our previous work (Naudin et al. 2021) onto the N-termini of each isolated chain in our scaffold, sampling phosphoserines at the Ncap, N1, and N2 helical positions.

We stitched on the RRXS substrate motif of PKA, selected structures whose phosphoserines made at least two hydrogen bonds to its upstream or downstream backbone amides, selected for structures whose phosphoserine's distal oxygens are close enough to an adjacent helix to potentially make an interhelical interaction (within 8Å), and eliminated structures where the phosphoserine clashed with the backbones of adjacent helices. This process whittled 35 scaffolds * 51 phosphoserine motifs * 3 N-terminal helical positions = 5,355 scaffolds down to 782 structures.

The 782 structures were then designed using ProteinMPNN (Dauparas et al. 2022). ProteinMPNN doesn't recognize post-translationally modified amino acids, so to represent phosphoserine, we used glutamate, which is similarly charged and isosteric to phosphoserine. Sequences were symmetrized between chains by linking equivalent positions using a provided helper function that allows users to define homo-oligomers. Additionally, an arbitrarily high bias was placed on the substrate motif sequence to prevent mutation of those positions, and cysteines were completely omitted from sampling. We had to scan temperatures much higher than conventional sampling temperatures, because ProteinMPNN, in independent runs, designed long stretches of poly-alanine to practically the whole sequence, and we had to coax them out of their poly-alanine energy minima. We tried sampling temperatures of 0.25, 0.5, 0.75, 1, 2, and 5. ProteinMPNN generated 50 sequences for each sampling temperature, resulting

in a total of 300 sequences per scaffold, for a grand total of 300 * 782 = 234,600 sequences.

To determine the most viable sequences, we ranked the sequences based on global score, percentage of alanines (< 30%), percentage of hydrophobic residues (30-70%), and net charge (within -4 and 4, inclusive). We then took the top 1,000 sequences and used AlphaFold2 (Jumper et al. 2021) to predict their structures, again using glutamate to represent phosphoserine. We evaluated 4-chain predictions of these sequences, because we observed that when we tried to predict higher-order oligomers, the structures optimized for packing and folded our sequences into globular proteins, rather than extended, fibrillar coiled-coils. We then filtered the predicted structures to select for structures in which at least 2 of the 4 glutamate-representing phosphoserines form intermolecular salt bridges with an adjacent monomer, and selected for structures that were predicted to form anti-parallel, rather than parallel, coiled-coils. Lastly, we selected for structures that formed a staggered topology. The overwhelming majority of results were structures where packing was maximized; very few were predicted to stagger. Therefore, the AlphaFold2 predictions had significant distinguishing power in isolating the best designs. The final sequences and models to be experimentally tested are listed in Table 1.1.

DISCUSSION

De novo coiled-coils are robust scaffolds for designing protein switches. Several groups have successfully designed *de novo* protein switches, including switches that respond to pH (Lizatovic et al. 2016, Boyken et al. 2019) and protein ligands (Langan et

al. 2019, Lajoie et al. 2020). Our own group designed *de novo* switches that respond to zinc (Joh et al. 2014) and kinase activation (Signarvic & DeGrado 2003), from which these second-generation designs are built.

This current study relies on design principles set forth in Signarvic & DeGrado's previous work, but whereas the first-generation peptides were rationally designed, we now have the advantage of leveraging computational methods to improve interhelical interactions and improve sensitivity to phosphorylation. When expressed in the SPARK reporter, these second-generation phosphoswitches form liquid droplets in cells in response to PKA activation, marking these as the first examples of *de novo* designed proteins that induce LLPS. We're excited to ultimately apply this computational pipeline to design highly specific and orthogonal switches for other kinases, so that we may develop a multi-color SPARK system to track spatiotemporal activity of multiple signaling pathways in complex networks. Additionally, this work sets a framework for designing switches for other post-translational modifications. Looking even further, one can envision that *de novo* designed LLPS-inducing protein switches can be designed to facilitate drug delivery, study biomolecular condensate properties, and probe or regulate signaling pathways.

This is an extraordinarily exciting era for structural biology and protein design. In this age of the AI (artificial intelligence) protein revolution, computational modeling and design methods are being developed and released at breakneck speeds (Baek & Baker 2022, Eisenstein 2023). We're optimistic that new methods will allow us to better represent concepts that are still challenging to model in design of protein switches, such as modeling multiple stable states and intermediate states necessary for switch

behavior (Alberstein et al. 2022). We can certainly expect that new AI protein design methods will allow us to impart properties that make “bespoke” protein switches robust for in-cell applications. The protein design field will soon be well-positioned to functionalize these switches and design them to precisely control and tune biological behavior with high sophistication.

REFERENCES

Alberstein, R. G., Guo, A. B., & Kortemme, T. (2022). Design principles of protein switches. *Current Opinion in Structural Biology*, *72*, 71–78. https://doi.org/10.1016/j.sbi.2021.08.004

Andrew, C. D., Warwicker, J., Jones, G. R., & Doig, A. J. (2002). Effect of Phosphorylation on α-Helix Stability as a Function of Position. *Biochemistry*, *41*(6), 1897–1905. https://doi.org/10.1021/bi0113216

Baek, M., & Baker, D. (2022). Deep learning and protein structure modeling. *Nature Methods*, *19*(1), Article 1. https://doi.org/10.1038/s41592-021-01360-8

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000a). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000b). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Boyken, S. E., Benhaim, M. A., Busch, F., Jia, M., Bick, M. J., Choi, H., Klima, J. C., Chen, Z., Walkey, C., Mileant, A., Sahasrabuddhe, A., Wei, K. Y., Hodge, E. A., Byron, S., Quijano-Rubio, A., Sankaran, B., King, N. P., Lippincott-Schwartz, J., Wysocki, V. H., … Baker, D. (2019). De novo design of tunable, pH-driven conformational changes. *Science (New York, N.Y.)*, *364*(6441), 658–664. https://doi.org/10.1126/science.aav7897

Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Courbet, A., de Haas, R. J., Bethel, N., Leung, P. J. Y., Huddy, T. F., Pellock, S., Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., … Baker, D. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. *Science (New York, N.Y.)*, *378*(6615), 49–56. https://doi.org/10.1126/science.add2187

Eisenstein, M. (2023). AI-enhanced protein design makes proteins that have never existed. *Nature Biotechnology*, 1–3. https://doi.org/10.1038/s41587-023-01705-y

Essentials of de novo protein design: Methods and applications—Marcos—2018—WIREs Computational Molecular Science—Wiley Online Library. (n.d.). Retrieved March 2, 2023, from https://wires.onlinelibrary.wiley.com/doi/10.1002/wcms.1374

Fairman, R., Chao, H.-G., Mueller, L., Lavoie, T. B., Shen, L., Novotny, J., & Matsueda, G. R. (1995). Characterization of a new four-chain coiled-coil: Influence of chain length on stability. *Protein Science*, *4*(8), 1457–1469. https://doi.org/10.1002/pro.5560040803

Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E.-M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J., & Baker, D. (2011). RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLOS ONE*, *6*(6), e20161. https://doi.org/10.1371/journal.pone.0020161

Friedman, A. M., Fischmann, T. O., & Steitz, T. A. (1995). Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science (New York, N.Y.)*, *268*(5218), 1721–1727. https://doi.org/10.1126/science.7792597

Grigoryan, G., & DeGrado, W. F. (2011). Probing Designability via a Generalized Model of Helical Bundle Geometry. *Journal of Molecular Biology*, *405*(4), 1079–1100. https://doi.org/10.1016/j.jmb.2010.08.058

Huang, P.-S., Boyken, S. E., & Baker, D. (2016). The coming of age of de novo protein design. *Nature*, *537*(7620), Article 7620. https://doi.org/10.1038/nature19946

Joh, N. H., Wang, T., Bhate, M. P., Acharya, R., Wu, Y., Grabe, M., Hong, M., Grigoryan, G., & DeGrado, W. F. (2014). De novo design of a transmembrane Zn2+-transporting four-helix bundle. *Science (New York, N.Y.)*, *346*(6216), 1520–1524. https://doi.org/10.1126/science.1261172

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), Article 7873. https://doi.org/10.1038/s41586-021-03819-2

Korendovych, I. V., & DeGrado, W. F. (2020). De novo protein design, a retrospective. *Quarterly Reviews of Biophysics*, *53*, e3. https://doi.org/10.1017/S0033583519000131

Kreegipuu, A., Blom, N., Brunak, S., & Järv, J. (1998). Statistical analysis of protein kinase specificity determinants. *FEBS Letters*, *430*(1–2), 45–50. https://doi.org/10.1016/s0014-5793(98)00503-1

Lajoie, M. J., Boyken, S. E., Salter, A. I., Bruffey, J., Rajan, A., Langan, R. A., Olshefsky, A., Muhunthan, V., Bick, M. J., Gewe, M., Quijano-Rubio, A., Johnson, J., Lenz, G., Nguyen, A., Pun, S., Correnti, C. E., Riddell, S. R., & Baker, D. (2020). Designed protein logic to target cells with precise combinations of surface antigens. *Science (New York, N.Y.)*, *369*(6511), 1637–1643. https://doi.org/10.1126/science.aba6527

Langan, R. A., Boyken, S. E., Ng, A. H., Samson, J. A., Dods, G., Westbrook, A. M., Nguyen, T. H., Lajoie, M. J., Chen, Z., Berger, S., Mulligan, V. K., Dueber, J. E., Novak, W. R. P., El-Samad, H., & Baker, D. (2019). De novo design of bioactive protein switches. *Nature*, *572*(7768), Article 7768. https://doi.org/10.1038/s41586-019-1432-8

Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y.-E. A., Fleishman, S. J., Corn, J. E., Kim, D. E., … Bradley, P. (2011). ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods in Enzymology*, *487*, 545–574. https://doi.org/10.1016/B978-0-12-381270-4.00019-6

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D. S., Banani, S. F., Russo, P. S., Jiang, Q.-X., Nixon,

B. T., & Rosen, M. K. (2012). Phase Transitions in the Assembly of Multi-Valent Signaling Proteins. *Nature*, *483*(7389), 336–340. https://doi.org/10.1038/nature10879

Linghu, C., Johnson, S. L., Valdes, P. A., Shemesh, O. A., Park, W. M., Park, D., Piatkevich, K. D., Wassie, A. T., Liu, Y., An, B., Barnes, S. A., Celiker, O. T., Yao, C.-C., Yu, C.-C. (Jay), Wang, R., Adamala, K. P., Bear, M. F., Keating, A. E., & Boyden, E. S. (2020). Spatial Multiplexing of Fluorescent Reporters for Imaging Signaling Network Dynamics. *Cell*, *183*(6), 1682-1698.e24. https://doi.org/10.1016/j.cell.2020.10.035

Maguire, J. B., Haddox, H. K., Strickland, D., Halabiya, S. F., Coventry, B., Griffin, J. R., Pulavarti, S. V. S. R. K., Cummins, M., Thieker, D. F., Klavins, E., Szyperski, T., DiMaio, F., Baker, D., & Kuhlman, B. (2021). Perturbing the energy landscape for improved packing during computational protein design. *Proteins*, *89*(4), 436–449. https://doi.org/10.1002/prot.26030

Pan, X., & Kortemme, T. (2021). Recent advances in de novo protein design: Principles, methods, and applications. *Journal of Biological Chemistry*, *296*, 100558. https://doi.org/10.1016/j.jbc.2021.100558

Serganova, I., & Blasberg, R. G. (2019). Molecular Imaging with Reporter Genes: Has Its Promise Been Delivered? *Journal of Nuclear Medicine*, *60*(12), 1665–1681. https://doi.org/10.2967/jnumed.118.220004

Smart, J. L., & McCammon, J. A. (1999). Phosphorylation stabilizes the N-termini of α-helices. *Biopolymers*, *49*(3), 225–233. https://doi.org/10.1002/(SICI)1097-0282(199903)49:3<225::AID-BIP4>3.0.CO;2-B

Szczepaniak, K., Lach, G., Bujnicki, J. M., & Dunin-Horkawicz, S. (2014). Designability landscape reveals sequence features that define axial helix rotation in four-helical homo-oligomeric antiparallel coiled-coil structures. *Journal of Structural Biology*, *188*(2), 123–133. https://doi.org/10.1016/j.jsb.2014.09.007

Thermodynamic Analysis of a Designed Three-Stranded Coiled Coil | Biochemistry. (n.d.). Retrieved March 2, 2023, from https://pubs.acs.org/doi/abs/10.1021/bi961831d

Thomas, F., Dawson, W. M., Lang, E. J. M., Burton, A. J., Bartlett, G. J., Rhys, G. G., Mulholland, A. J., & Woolfson, D. N. (2018). De Novo-Designed α-Helical Barrels as Receptors for Small Molecules. *ACS Synthetic Biology*, *7*(7), 1808–1816. https://doi.org/10.1021/acssynbio.8b00225

Woolfson, D. N. (2021). A Brief History of De Novo Protein Design: Minimal, Rational, and Computational. *Journal of Molecular Biology*, *433*(20), 167160. https://doi.org/10.1016/j.jmb.2021.167160

Zhang, Q., Huang, H., Zhang, L., Wu, R., Chung, C.-I., Zhang, S.-Q., Torra, J., Schepis, A., Coughlin, S. R., Kornberg, T. B., & Shu, X. (2018). Visualizing Dynamics of Cell Signaling In Vivo with a Phase Separation-Based Kinase Reporter. *Molecular Cell*, *69*(2), 334-346.e4. https://doi.org/10.1016/j.molcel.2017.12.008

Zhou, J., & Grigoryan, G. (2015). Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein Science: A Publication of the Protein Society*, *24*(4), 508–524. https://doi.org/10.1002/pro.2610

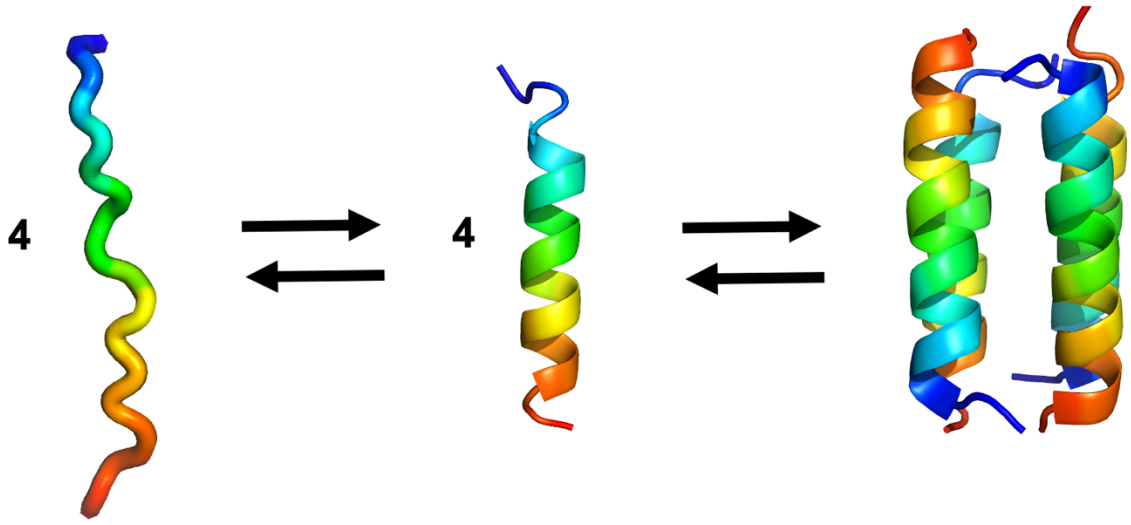

Figure 1.1. Schematic illustrating monomer-tetramer equilibrium

The designed peptides may access 3 distinct states: unstructured monomer, helical monomer, and coiled-coil tetramer. Stabilization of the alpha-helical monomeric intermediate biases the equilibrium toward coiled-coil formation, whereas destabilization of the alpha-helical intermediate biases the equilibrium toward disorder.

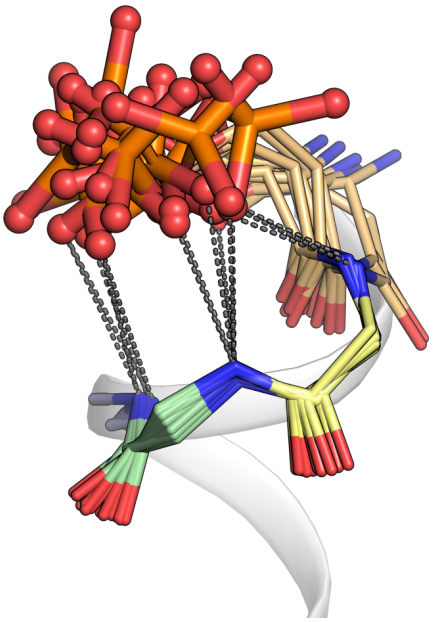

Figure 1.2 Phosphoserines capping N-termini of alpha-helices

This figure collates some N-terminal helical phosphoserines taken from crystallographically-solved structures in the PDB. These representative N-terminal motifs demonstrate that N-capping phosphoserines stabilize alpha-helices by hydrogen bonding to the exposed, unpaired backbone amides that are also residing on the N-terminus.

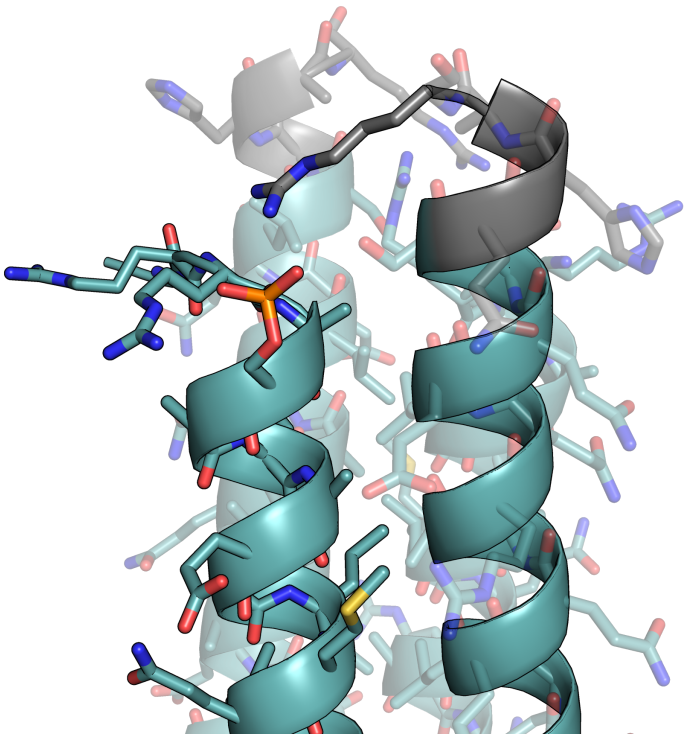

Figure 1.3 Model of a Lac repressor-based phosphoswitch design

An example phosphoswitch design is shown here. The N-terminus is PKA phosphorylation-competent because of the RRXSL substrate motif. The greyed residues at the C-termini are part of the C-terminal extension and are not present on WT Lac repressor. Without this extension, the phoshoserine cannot form an intermolecular salt bridge; though there is an “in-plane” (perpendicular to superhelical axis) residue on the adjacent helix that’s close to the phosphoserine, that adjacent helix is too close to form a favorable salt bridge, because the arginine or lysine would have been strained. In the design pictured here, that “in-plane” residue incompetent for interhelical salt bridge formation is a serine.

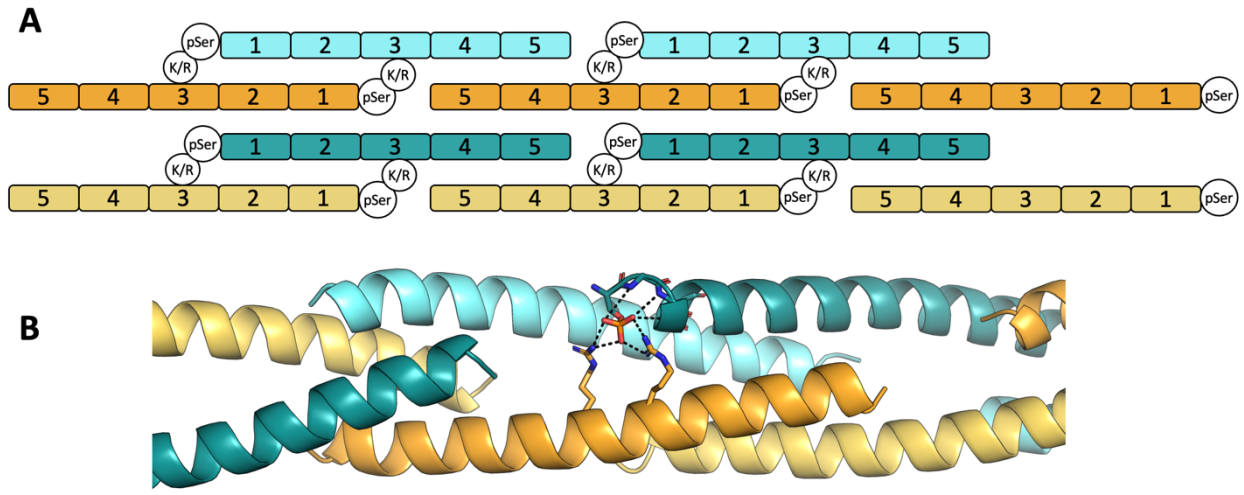

Figure 1.4 Design schematic of staggered, fibrillar coiled-coil

(A) Our designed peptide comprising 5 helical heptads is symmetrically arranged such that a phosphoserine at the N-terminus can salt bridge with a basic residue on an adjacent helix. **(B)** This schematic is illustrated in cartoon representation to better visualize the junctions formed from this staggered effect; the length of the overlap between helices can be tuned such that it's stable enough to assemble in the presence of the phosphate group, yet remain disassociated in the absence of the phosphate group. A phosphoserine at the N-terminus of the peptide self-stabilizes its helical conformation by hydrogen bonding to backbone amides of its own unit, and additionally stabilizes the complex as a whole through interactions with arginines in the middle of an adjacent peptide. Because of the staggered arrangement, this assembly is expected to self-propagate and fibrillize.

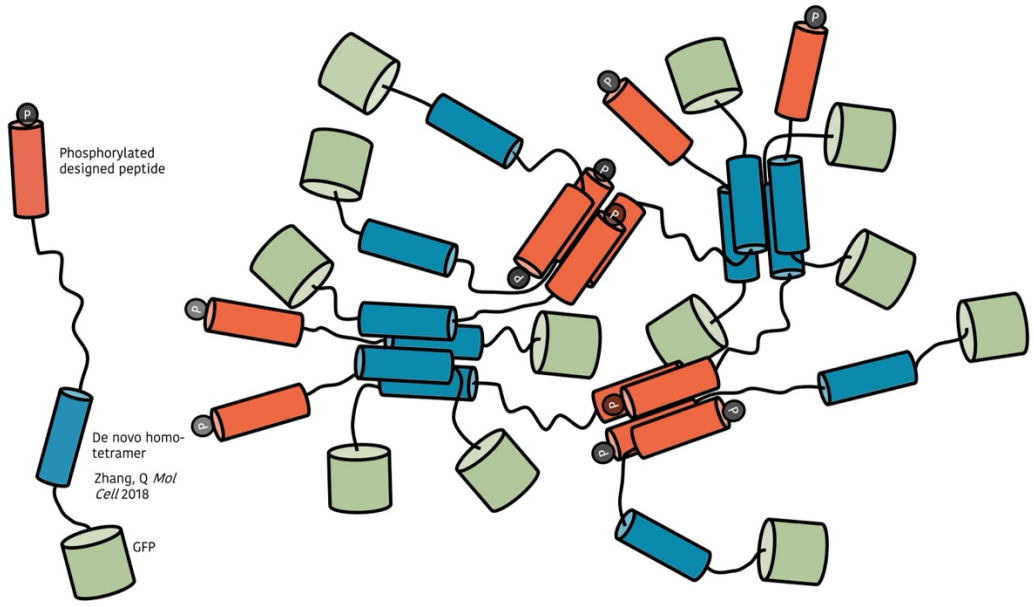

Figure 1.5 Schematic of LLPS-inducing phoshoswitch reporter

In the unphosphorylated state of the reporter, only the *de novo* homo-tetramer (blue) assembles, which isn't sufficient for LLPS. However, in the phosphorylated state, the designed peptides (orange) assemble, and the system forms highly multivalent interactions that cause phase separation.

Table 1.1. Phosphoswitch sequences experimentally characterized.

For *in vitro* synthesis, all N-termini and C-termini were acetylated and amidated, respectively. For *in vivo* expression, the designs are preceded by 3 residues: "MGG". The phosphorylated serine is bolded.

WT Lac repressor tetramerization domain (not tested)	ALADSLMQLARQVSRLE
First-generation switch (Signarvic et al. 2003)	RRR**S**ALAEALMQLARQVSRLA
LLPS switch 1	RRR**S**ALAEALMQLARQLEQLSQHRK
LLPS switch 2	RRQ**S**RLAEYLSRLAHQFQQFSSQKR
LLPS switch 3	RRQ**S**YQLAEALMQLARQVSRQSQHRR
LLPS switch 4	RRQ**S**YRLAEYLQHLAQYFWQKSQSRR
Fibril switch 1	RRR**S**ILAKVKQAIEELEQRLQERQQDVENLEALTKRMR
Fibril switch 2	RRR**S**LQVLKNLTQRAEAARQREEVLHQELDRARQG
Fibril switch 3	RRR**S**LEALLALLARLQEFAEREARLRALLEEARRL
Fibril switch 4	RRR**S**LEFLKLLEQLAAKAKERRQRMERLQALLQEALRL

Chapter 2:

Computational analysis of hot spot positions in an integrin interface

INTRODUCTION

The integrin αIIbβ3 resides on the platelet surface in a regulated and finely-tuned equilibrium between resting low affinity and active high affinity conformations that can be perturbed by bidirectional signal transduction (Li et al. 2005). Crystal structures of resting αIIbβ3 have revealed that its extracellular domain has a bent conformation with its nodular ligand-binding headpiece oriented toward the cell surface (Xiong et al. 2001, Xiao et al. 2004) and with contacts between the β3 and αIIb stalks forming a clasp that maintains its inactive state (Fig. 2.1) (Zhu et al 2008). Following platelet stimulation, αIIbβ3 undergoes a global rearrangement (Takagi et al. 2002) in which its ligand-binding headpiece turns away from the cell surface and its transmembrane (TM) and stalk domains separate. This causes the integrin to shift to a fully extended conformation (Luo et al. 2007), exposing its ligand binding site. *In vivo*, intracellular signals shift αIIbβ3 towards its high affinity ligand-binding conformation ("inside-out signaling") (Shattil & Newman 2004). However, αIIbβ3 can also be experimentally shifted to its active conformation by replacing individual amino acids in its TM or membrane-proximal extracellular stalk domains, although the magnitude of the shift varies from replacement to replacement.

In this study, we characterized the αIIbβ3 interface by introducing alanine replacements and measuring their effects on constitutive integrin activation. Protein-protein interfaces, such as the interface between the αIIb and β3 stalks, are usually large complementary surfaces with many intermolecular contacts (Bogan & Thorn 1998). Nonetheless, as Clackson and Wells pointed out, a limited number of complementary side chain interactions in protein-protein interfaces, termed “hot spots”, are disproportionately responsible for the strength of binding (Clackson & Wells 1995). Previously, our group used the Rosetta alanine scanning algorithm to identify hot spots in the β3 stalk and found that replacing them with alanine was sufficient to activate both αIIbβ3 and αvβ3 (Zhu et al. 2010). However, while the alanine scanning algorithm successfully predicted mutations that destabilized the stalk interface, it did not correctly rank their functional effects, likely because the energetic effects of the alanine substitutions were small and within the margin of error reported for this method (Kortemme et al. 2004).

Here, we sought to identify hot spots in the αIIb stalk that are complementary to those we previously identified in β3. First, we scanned the αIIb Calf-1 and Calf-2 domains using the Rosetta-based alanine scanning algorithm, and subsequently using the recently described Rosetta flex ddG protocol (Barlow et al. 2018). The latter enables more accurate ddG predictions by generating ensembles of models sampling different backbone conformations, whereas the former does not. However, neither method was able to accurately rank the effects of hot spot mutations for this specific system. We wondered if the effects of alanine mutation could be better captured using an approach based on mining the protein data bank (PDB) for all sidechain interactions.

This method is based on the premise that the detailed geometries of the most stabilizing inter-subunit sidechain-sidechain interactions would be over-represented in the PDB, and that this mining would be able to better pick up on multi-body, cooperative interactions between two binding interfaces.

We used the results of this analysis to compute an interaction geometry score that was better able to rationalize the nature of destabilizing mutations in the stalk interface than the flex ddG protocol. Further, we found that specific stalk domain hot spots are responsible for maintaining the inactive state of αIIbβ3. Because the stalks are present in an extracellular location, these results suggest that stabilizing the stalk heterodimers may be a way to allosterically attenuate αIIbβ3 function.

RESULTS

Computational alanine scanning to identify mutation-sensitive residues

Crystal structures of the ectodomain of inactive αIIbβ3 have revealed a large interface between the Calf-1 and Calf-2 domains of the αIIb stalk and the EGF-3, EGF-4, and βTD domains of the distal β3 stalk (Zhu et al. 2008, Zang & Springer 2001) (Fig. 2.1). Previously, we used the Robetta alanine scanning algorithm (Kortemme et al. 2002, Kortemme et al. 2004), to predict destabilizing alanine replacements in the β3 stalk and found that introducing these replacements into full-length αIIbβ3 caused constitutive αIIbβ3 activation (Donald et al. 2010). Using the same method to determine hot spots in the αIIb stalk, we identified 12 alanine replacements with predicted ddG's ranging from 0.1 to 1.8 kcal/mol (Table 2.1). To determine which replacements destabilized the

interface of the stalk sufficiently to cause αIIbβ3 activation, our collaborators (Bennett lab, Division of Hematology-Oncology, UPenn) introduced 10 of the replacements into full-length αIIb by site-directed mutagenesis, stably co-expressed the mutants with wild-type (WT) β3 in CHO cells, and measured both constitutive and dithiothreitol-induced fibrinogen binding to sub-clones selected by fluorescence-activated cell sorting analysis for comparable expression of αIIbβ3. To ensure that results were not unique to a particular sub-clone, fibrinogen binding measurements were performed using 2-6 different sub-clones. To normalize the activity of the various αIIbβ3 mutants, they calculated an αIIbβ3 activation index, the ratio of constitutive fibrinogen binding to αIIbβ3 to maximal fibrinogen binding induced by dithiothreitol. The R751A mutant did not express, implying that R751 may be important for either correct αIIb folding or for correct αIIbβ3 assembly. Each of the other mutants expressed to a comparable extent and caused a variable degree of constitutive αIIbβ3 activation, with αIIbβ3 activation indices ranging from 0.83±0.12 for V760A to 0.17±0.02 for H787A (Table 2.1).

Positioning molecular contacts of hot spot and neutral residues

We then mapped the hot spots we identified in αIIb, and those we previously identified in β3 (Donald et al. 2010), onto the model of the stalk heterodimer shown in Fig. 2.2. Hot spot residues whose alanine mutants promoted αIIbβ3 activation were found to lie along a discontinuous strip running through the stalk interface. In the assembled stalk heterodimer, residues having high activation indices when replaced by alanine (i.e., ≥ 0.4) were flanked by hot spot residues whose alanine mutants activate αIIbβ3 to a lesser extent. In contrast, β3 residues D552 and H626, chosen as negative

controls because they do not make inter-subunit contacts with αIIb, were predicted to have no effect on αIIbβ3 heterodimer stability and did not cause constitutive αIIbβ3 activation (Table 2.1).

Comparing computational alanine scanning results to experimental results reveals poor correlation

As we previously observed for the β3 stalk (Donald et al. 2010), the Robetta alanine scanning algorithm did not correctly rank the functional importance of the hot spots it predicted in the αIIb stalk. A recently-reported flex ddG algorithm, implemented in Rosetta, was demonstrated to more accurately calculate ΔΔGs (Barlow et al. 2018). We repeated the computational alanine scanning using flex ddG. To comprehensively examine the whole stalk interface, we extended our analysis to include the β*3* mutants characterized in our previous work (Donald et al. 2010). When we plotted the apparent free energy of fibrinogen binding (ΔG_{app}) to mutant αIIbβ3 versus the ΔΔGs predicted by flex ddG, we again found only a weak correlation (Fig. 2.3A, R^2=0.002), likely because the energetic differences between the integrin mutants are within a very small range. The largest and smallest activation indices for the αIIbβ3 mutants differ by only a factor of 4.8, corresponding to an energetic change of only 1-2 kcal/mol, close to the expected flex ddG error of ± 0.96 kcal/mol (Barlow et al. 2018).

Analyzing interaction geometry across the integrin interface qualitatively

To more accurately assess the energetic contribution of individual αIIb and β3 residues to the stability of resting αIIbβ3, we investigated whether hot spot residues

impart stability in a predictable manner that could be ascertained by evaluating their interaction geometry. For each αIIb or β3 residue whose alanine mutant was experimentally characterized, we identified the complementary residues with which it interacted and represented the interaction as single-residue fragment pairs so that we could query the non-redundant PDB for pairs of fragments that interact in the same geometry as in the αIIbβ3 crystal structure. The non-redundant dataset was then searched for geometric matches to WT αIIbβ3, which we defined as residue pairs whose fragments had low root mean squared deviation (RMSD ≤ 0.5Å) with the interacting fragment pairs in the αIIbβ3 crystal structure.

As anticipated, hot spot residues with high activation indices interacted with complementary subunit residues in geometries that are highly represented in the non-redundant PDB. Details of these interactions are shown in Fig. 2.4, Table 2.2, and Table 2.3. For example, the hot spot residues whose alanine mutants have the highest activation indices, β3 T603 and αIIb V760, have numerous geometric matches. T603, whose activation index is 0.86, has a total of 385 geometric matches with 6 fragments on 4 αIIb residues. V760, whose activation index is 0.83, has 2940 geometric matches with 1 residue on β3. By contrast, αIIb residue I673, whose activation index is 0.23, has only 54 geometric matches. Mutants that do not cause integrin activation relative to WT (i.e., β3 D552A and β3 H626A) do not make any inter-subunit interactions and have 0 geometric matches in the PDB.

Analyzing interaction geometry across the integrin interface quantitatively

Based on our findings that highly represented interaction geometries are present at the αIIbβ3 stalk interface, we developed a simple scoring function, $Geom(h)$, to score the geometric interaction propensities of hot spot residue h (Equation 2.2). The $Geom(h)$ scores for highly activating β3 T603 and αIIb V760 were -3.75 and -2.91, respectively, while those for less activating αIIb residues I673 and N753 were 0.16 and -1.87. When we examined the quantitative agreement between the $Geom(h)$ scores and the apparent binding energies to fibrinogen, we found the correlation coefficient to be 0.796 (Fig. 2.3B).

Nonetheless, there were two notable discrepancies between the computational alanine scanning and the structural bioinformatics results. First, αIIb S758 makes no direct inter-chain contacts, but its alanine replacement has a high activation index of 0.64. However, the αIIbβ3 crystal structure is not well-resolved in this region, as evidenced by the sidechains of Q954/L956 of αIIb and K612/K658 of β3 not being represented in the electron density maps. Moreover, there is unassigned density corresponding to two volumes that could potentially be in contact with S758, suggesting that there may be solvent-mediated inter-chain interactions that are not accounted for by the bioinformatics method (Fig. 2.4).

The second exception is β3 E534. Previously, using Robetta alanine scanning, we predicted that E534A was moderately destabilizing with a ΔΔG of 0.54 kcal/mol (Donald et al. 2010). Based on the crystal structure of inactive αIIbβ3, β3 E534A does not disrupt an interaction across the αIIbβ3 stalk interface. Rather, it disrupts a hydrogen

bond between the β3 EGF-3 domain and the αIIb β-propeller located in the αIIb ectodomain (Fig. 2.5), causing constitutive αIIbβ3 activation with an activation index of 0.76±0.07. The structural bioinformatics analysis revealed that E534 interacts with the β-propeller residue R402 with a $Geom(h)$ score of -1.19, an intermediate degree of geometric favorability relative to the stalk mutants. However, E534A is more activating than its geometry score would suggest. Thus, while these results confirm that the structural bioinformatics method can describe favorable interaction geometries across different regions in multi-domain proteins, it also cautions that without carefully training the model on a large database of inter-domain interactions, we can only rank the favorability of those interactions within the same interface. E534 is also noteworthy because Zang and Springer had previously reported that mutating β2 residue Q535, analogous to β3 E534, as well as β2 V526, both located in the β2 EGF-3 domain, caused the activation of the leukocyte integrin αxβ2 (CDIIc/CD18) and postulated that an interaction between these residues and unidentified residues in αx restrained αxβ2 in its inactive state (Zang & Springer 2001). Our results suggest that the inactive state of αxβ2 is stabilized through an interaction similar to β3 E534–αIIb R402 and further supports the bent conformation as a biologically-relevant state for this class of integrins.

METHODS

Computational alanine scanning in Robetta and Rosetta

We used chains A (residues 599-959) and B (residues 483-690) from the αIIbβ3 crystal structure (PDBID 3FCS) (Zhu et al. 2008) for all our computational analyses.

We initially used Rosetta interface alanine-scanning, hosted on the Robetta server, to predict hot spots in the αIIb stalk (Kortemme et al. 2004, Kortemme et al. 2002). After completing the experimental aspects of this work, we repeated the computational alanine scanning mutagenesis using flex ddG, a recently-developed method, built within the Rosetta macromolecular modeling suite, which provides more accurate ΔΔG predictions by generating ensembles of models that sample over different backbone and sidechain conformations (Barlow et al. 2018). Essentially, ensembles of wild-type (WT) and alanine-mutant models were generated by sampling 35,000 backrub backbone perturbation ("backrub") and minimization cycles. ΔΔGs between WT and alanine mutants were calculated using Rosetta energy function terms that were re-weighted to better fit experimental ΔΔGs reported in the ZEMu protein-protein interaction benchmark set (Barlow et al. 2018, Dourado et al. 2014).

In vivo characterization of alanine mutants

All experimental work was performed by our collaborators (Bennett lab, Division of Hematology-Oncology, UPenn). Their protocol is detailed in our published work (Tan et al. 2019). The αIIbβ3 activation index was calculated from Equation 2.1.

Because we extended our analysis to include the β*3* stalk domain mutants we characterized in our previous work (Donald et al. 2010), to maintain consistency, we converted the fibrinogen binding data to the apparent free energy of fibrinogen binding (ΔG_{app}) as defined by Equation 2.2.

Curating a database to query interaction motifs

To rationalize the strengths of αIIbβ3 mutant stalk interactions, we developed a structural bioinformatics analysis method based on the hypothesis that disrupting αIIb and β3 stalk interactions that are over-represented in the Protein Data Bank (PDB) (Berman et al. 2000) would destabilize the resting αIIbβ3 heterodimer and hence activate αIIbβ3. We curated a dataset of crystallographic structures from the PDB (accessed February 7, 2018) with ≤ 30% sequence identity, ≤ 2Å resolution, R value of ≤0.3, and MolProbity (Chen et al. 2010) score of <2 (which evaluates clashing and rotamer/phi/psi geometry), resulting in a database of 8,415 structures. Biological assemblies were then reconstructed using ProDy (Bakan et al. 2011) to maintain interactions across monomers of the same protein complex.

Fragmenting hot spot interactions for database searching

Next, we discretized putative hot spot interactions into fragments to query our non-redundant PDB dataset for the same interaction geometry between fragments. For each putative hot spot residue with experimentally-derived functional data, we determined its opposite-subunit interacting fragments using the program Probe (Word et al. 1999), first minimizing the αIIbβ3 crystal structure using Rosetta's minimize_ppi application (Leavery-Fay et al. 2011, Bazzoli et al. 2015). Each inter-subunit contact Probe classified as a hydrogen bond, close contact, or strong atomic overlap was discretized into fragments belonging to a single residue that could fall within a plane. Fragments from sidechains with tetrahedral geometry were limited to 3 atoms if the fragment contains an sp^3-hybridized carbon, while fragments could contain ≥ 3 atoms if one or

more atom was sp^2-hybridized or aromatic. Fragments from a single residue could also arise from backbone atoms ([N, CA, C] or [CA, C, O]). For example, aspartic acid fragments could be [N, CA, C], [CA, C, O], [N, CA, CB], [C, CA, CB], [CA, CB, CG], or [CB, CG, OD1, OD2]. However, we evaluated the interaction geometry for backbone fragments only if they originated from a complementary subunit residue rather than a putative hot spot residue, since interactions from backbone fragments of a hot spot residue would not be eliminated upon mutating the hot spot residue to alanine.

Database searching for geometric matches of intermolecular interactions

To determine the favorability of the inter-subunit interactions, we searched the non-redundant protein dataset for residue pairs that had fragments interacting in the same geometry as in the wild-type crystal structure (3FCS). For each interaction between a hot spot residue AA_h and its complementary residue, AA_i, we approximated the energetic contribution imparted by forming an interaction in that specific geometry as $Geom(h)$ (Equation 2.3). To account for residue pairs that were in contact simply because of sequence proximity, and not necessarily because of favorable interactions, we only included residue pairs on the same chain if they were separated by at least 10 residues. Calculations were performed using the Python package NumPy (van der Walt et al. 2011) and plots were created using the Python package Matplotlib (Hunter 2007).

DISCUSSION

Platelets circulate in a milieu that is rich in fibrinogen, the principal ligand for the integrin αIIbβ3. Because fibrinogen binding to activated αIIbβ3 causes platelet

aggregation, αIIbβ3 on circulating platelets is held in an inactive state by an intramolecular clasp composed of portions of its cytosolic, TM, and extracellular stalk domains to prevent the formation of intravascular platelet aggregates (Hynes 2002, Vinogradova et al. 2002, Bennett 2005). At sites of vascular injury where rapid platelet aggregation is required to stop bleeding, platelet stimulation causes disruption of the clasp, followed by a global αIIbβ3 rearrangement during which its ectodomain extends and exposes its fibrinogen binding site (Takagi et al. 2002).

In crystal structures (Xiong et al. 2001, Xiao et al. 2004) and electron microscope images (Takagi et al. 2002, Eng 3t al. 2011) of the extracellular domain of inactive αIIbβ3, the lower leg of β3 is in proximity to both the lower α leg and the integrin headpiece. One proposed trigger for the global rearrangement is the release of inter-subunit contacts located in the stalk domain, thereby allowing the αIIb and β3 components of the stalk to separate and the ectodomain to extend (Beglova et al. 2002, Wang et al. 2010). Lending support for this mechanism, three families with the inherited bleeding disorder Glanzmann thrombasthenia have been reported in whom deletion of β3 residues D647-E686 (Bury et al. 2016) or β3 residues D621-E660 (Kashiwagi et al. 2013), containing the predicted β3 hot spots K658 and V644, caused constitutive αIIbβ3 activation. Previously, Kamata et al. also reported that swapping the Calf-2 domains of αIIb and αv enhanced Mn^{2+}-induced fibrinogen binding to αIIbβ3, but suppressed Mn^{2+}-induced fibrinogen binding to αvβ3, suggesting that the interface between the α subunit Calf-2 domain and the β3 EGF-4 and βTD domains regulates Mn^{2+}-induced ligand binding to αIIbβ3 and αvβ3 (Kamata et al. 2005). Consistent with this conclusion, Mn^{2+}-

induced fibrinogen binding to αIIbβ3 was suppressed by an artificial disulfide bridge between the αIIb Calf-2 and the β3 βTD domains. However, neither the Calf-2 domain swaps, nor subsequent residue interchanges, caused integrin activation in the absence of Mn^{2+}. Mn^{2+} by itself is a weak integrin activator and neither the αIIb Calf-2 nor the β3 EGF-4-βTD domains bind cations. Thus, it is likely that perturbations in the interface between the α and β subunit stalks introduced by the Calf-2 domain swaps potentiated an effect of Mn^{2+} elsewhere in the αIIbβ3 and αvβ3 molecules. It has also been proposed that in bent integrins, the CD loop of the β3 βTD domain contacts the β3 βA domain F/α_7 loop, acting as a "deadbolt" to prevent the allosteric movement of the α7 helix that initiates opening of the integrin headpiece (Xiong et al. 2003). However, neither deleting nor mutating the CD loop perturbs ligand binding to αIIbβ3 (et al. 2007), making it doubtful that these CD loop interactions regulate integrin function.

To identify hot spot interactions in the αIIb and β3 stalks that regulate αIIbβ3 function, we used the computational alanine scanning algorithm hosted on Robetta (Kortemme et al. 2002, Kortemme et al. 2004) to predict interacting hot spots in the stalk heterodimer, initially identifying 9 alanine replacements in β3 and then 12 alanine replacements in αIIb with predicted ΔΔG's ranging from 0.06-2.89 kcal/mol. Hot spots have been variably defined as residues whose replacement by alanine destabilizes a protein-protein interface by ΔΔG's ranging from >1.0 to >4.0 kcal/mol (Bogan et al. 1998, Clackson et al. 1995, Kortemme et al. 2004, Moreira et al. 2007). It is noteworthy that only 8 of 21 alanine replacements had a predicted ΔΔG > 1.0 kcal/mol (Table 2.1),

a proposed threshold for a destabilizing alanine replacement (Kortemme et al. 2002, Kortemme et al. 2004).

A newer computational alanine scanning protocol, flex ddG, generates ensembles of structures using the “backrub” protocol in Rosetta, thereby accounting for mutation-induced local side chain and backbone conformational changes. Flex ddG has been found to outperform other existing computational methods that also sample conformational space, particularly for small to large mutations (Barlow et al. 2018). Although successful for a diverse benchmark, there was only a weak correlation between flex ddG predictions and the energy of fibrinogen binding for this specific αIIbβ3 integrin system (Fig. 2.3A) because the effects we measured were small and within the margin of error of flex ddG.

To better understand the structural basis for activation, we developed a structural bioinformatics approach to analyze functionally important contacts in both the αIIb and β3 stalks. This approach was based on the hypothesis that functional groups have preferences in the relative position, orientation, and angles at which they interact, and these preferences are reflected in the PDB. Analyses of sidechain interactions in the PDB have shown that propensities of these interactions deviate from the distributions expected from random packing, implying that sidechain interactions are guided by directional preferences (Singh & Thornton 1990, Mitchell et al. 1997, Chakrabarti & Bhattacharyya 2007). Taking this into account, we approximated the energetic contribution of a hot spot residue by identifying its interacting fragment pairs and querying the PDB for the prevalence of those interaction geometries. We found a strong

correlation between our interaction geometry term and the binding energy of fibrinogen to αIIbβ3 (R^2=0.81) (Fig. 2.3B). Thus, this approach shows promise for analyzing putative hot spots identified by alanine scanning. However, a much larger dataset would need to be examined and the choice of molecular interaction fragments would likely need to be optimized to generalize this method over a wide range of molecular interactions. Once benchmarked, this knowledge-based method could accelerate interface assessment and design because it precludes the need to directly model mutation-induced structural changes, as well as the need to directly calculate energetic effects.

Another advantage of mining the PDB for favorable residue-residue interaction geometries is that it implicitly captures multi-body interactions at a binding interface, circumventing the need to explicitly account for cooperativity and polarization. Since the time this work was published, our group introduced a method to design small molecule binding sites using brute-force sampling of these PDB-mined residue-residue interactions (Polizzi & DeGrado 2020). The advent of AI in protein modeling and design has revolutionized the way we think of sampling, and we look forward to exploring how to best incorporate the database and interaction geometry metric into AI pipelines for more efficient sampling, simplifying interface evaluation and design on a significantly larger scale.

REFERENCES

Bakan, A., Meireles, L. M., & Bahar, I. (2011). ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics (Oxford, England)*, *27*(11), 1575–1577. https://doi.org/10.1093/bioinformatics/btr168

Barlow, K. A., Ó Conchúir, S., Thompson, S., Suresh, P., Lucas, J. E., Heinonen, M., & Kortemme, T. (2018). Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation. *The Journal of Physical Chemistry. B*, *122*(21), 5389–5399. https://doi.org/10.1021/acs.jpcb.7b11367

Bazzoli, A., Kelow, S. P., & Karanicolas, J. (2015). Enhancements to the Rosetta Energy Function Enable Improved Identification of Small Molecules that Inhibit Protein-Protein Interactions. *PloS One*, *10*(10), e0140359. https://doi.org/10.1371/journal.pone.0140359

Beglova, N., Blacklow, S. C., Takagi, J., & Springer, T. A. (2002). Cysteine-rich module structure reveals a fulcrum for integrin rearrangement upon activation. *Nature Structural Biology*, *9*(4), 282–287. https://doi.org/10.1038/nsb779

Bennett, J. S. (2005). Structure and function of the platelet integrin alphaIIbbeta3. *The Journal of Clinical Investigation*, *115*(12), 3363–3369. https://doi.org/10.1172/JCI26989

Bennett, J. S., Hoxie, J. A., Leitman, S. F., Vilaire, G., & Cines, D. B. (1983). Inhibition of fibrinogen binding to stimulated human platelets by a monoclonal antibody.

Proceedings of the National Academy of Sciences of the United States of America, *80*(9), 2417–2421. https://doi.org/10.1073/pnas.80.9.2417

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Bogan, A. A., & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, *280*(1), 1–9. https://doi.org/10.1006/jmbi.1998.1843

Bury, L., Falcinelli, E., Chiasserini, D., Springer, T. A., Italiano, J. E., & Gresele, P. (2016). Cytoskeletal perturbation leads to platelet dysfunction and thrombocytopenia in variant forms of Glanzmann thrombasthenia. *Haematologica*, *101*(1), 46–56. https://doi.org/10.3324/haematol.2015.130849

Chakrabarti, P., & Bhattacharyya, R. (2007). Geometry of nonbonded interactions involving planar groups in proteins. *Progress in Biophysics and Molecular Biology*, *95*(1–3), 83–137. https://doi.org/10.1016/j.pbiomolbio.2007.03.016

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 1), 12–21. https://doi.org/10.1107/S0907444909042073

Clackson, T., & Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science (New York, N.Y.)*, *267*(5196), 383–386. https://doi.org/10.1126/science.7529940

CSDL | IEEE Computer Society. (n.d.). Retrieved March 12, 2023, from https://www.computer.org/csdl/magazine/cs/2007/03/c3090/13rRUwbJD0A

Donald, J. E., Zhu, H., Litvinov, R. I., DeGrado, W. F., & Bennett, J. S. (2010). Identification of Interacting Hot Spots in the β3 Integrin Stalk Using Comprehensive Interface Design. *The Journal of Biological Chemistry*, *285*(49), 38658–38665. https://doi.org/10.1074/jbc.M110.170670

Dourado, D. F. A. R., & Flores, S. C. (2014). A multiscale approach to predicting affinity changes in protein-protein interfaces. *Proteins*, *82*(10), 2681–2690. https://doi.org/10.1002/prot.24634

Eng, E. T., Smagghe, B. J., Walz, T., & Springer, T. A. (2011). Intact alphaIIbbeta3 integrin is extended after activation as measured by solution X-ray scattering and electron microscopy. *The Journal of Biological Chemistry*, *286*(40), 35218–35226. https://doi.org/10.1074/jbc.M111.275107

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(03), 90–95. https://doi.org/10.1109/MCSE.2007.55

Hynes, R. O. (2002). Integrins: Bidirectional, allosteric signaling machines. *Cell*, *110*(6), 673–687. https://doi.org/10.1016/s0092-8674(02)00971-6

Kamata, T., Handa, M., Sato, Y., Ikeda, Y., & Aiso, S. (2005). Membrane-proximal {alpha}/{beta} stalk interactions differentially regulate integrin activation. *The Journal of Biological Chemistry*, *280*(26), 24775–24783. https://doi.org/10.1074/jbc.M409548200

Kashiwagi, H., Kunishima, S., Kiyomizu, K., Amano, Y., Shimada, H., Morishita, M., Kanakura, Y., & Tomiyama, Y. (2013). Demonstration of novel gain-of-function mutations of αIIbβ3: Association with macrothrombocytopenia and glanzmann thrombasthenia-like phenotype. *Molecular Genetics & Genomic Medicine*, *1*(2), 77–86. https://doi.org/10.1002/mgg3.9

Kortemme, T., & Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(22), 14116–14121. https://doi.org/10.1073/pnas.202485799

Kortemme, T., Kim, D. E., & Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. *Science's STKE: Signal Transduction Knowledge Environment*, *2004*(219), pl2. https://doi.org/10.1126/stke.2192004pl2

Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y.-E. A., Fleishman, S. J., Corn, J. E., Kim, D. E., … Bradley, P. (2011). ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods in Enzymology*, *487*, 545–574. https://doi.org/10.1016/B978-0-12-381270-4.00019-6

Luo, B.-H., Carman, C. V., & Springer, T. A. (2007). Structural Basis of Integrin Regulation and Signaling. *Annual Review of Immunology*, *25*, 619–647. https://doi.org/10.1146/annurev.immunol.25.022106.141618

Mitchell, J. B., Laskowski, R. A., & Thornton, J. M. (1997). Non-randomness in side-chain packing: The distribution of interplanar angles. *Proteins*, *29*(3), 370–380. https://doi.org/10.1002/(sici)1097-0134(199711)29:3<370::aid-prot10>3.0.co;2-k

Moreira, I. S., Fernandes, P. A., & Ramos, M. J. (2007). Hot spots—A review of the protein-protein interface determinant amino-acid residues. *Proteins*, *68*(4), 803–812. https://doi.org/10.1002/prot.21396

Shattil, S. J., & Newman, P. J. (2004). Integrins: Dynamic scaffolds for adhesion and signaling in platelets. *Blood*, *104*(6), 1606–1615. https://doi.org/10.1182/blood-2004-04-1257

Singh, J., & Thornton, J. M. (1990). SIRIUS. An automated method for the analysis of the preferred packing arrangements between protein groups. *Journal of Molecular Biology*, *211*(3), 595–615. https://doi.org/10.1016/0022-2836(90)90268-Q

Takagi, J., Petre, B. M., Walz, T., & Springer, T. A. (2002). Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. *Cell*, *110*(5), 599–511. https://doi.org/10.1016/s0092-8674(02)00935-2

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Vinogradova, O., Velyvis, A., Velyviene, A., Hu, B., Haas, T., Plow, E., & Qin, J. (2002). A structural mechanism of integrin alpha(IIb)beta(3) “inside-out” activation as

regulated by its cytoplasmic face. *Cell*, *110*(5), 587–597. https://doi.org/10.1016/s0092-8674(02)00906-6

Wang, W., Fu, G., & Luo, B.-H. (2010). Dissociation of the α-subunit Calf-2 domain and the β-subunit I-EGF4 domain in integrin activation and signaling. *Biochemistry*, *49*(47), 10158–10165. https://doi.org/10.1021/bi101462h

Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., & Richardson, D. C. (1999). Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms. *Journal of Molecular Biology*, *285*(4), 1711–1733. https://doi.org/10.1006/jmbi.1998.2400

Xiao, T., Takagi, J., Coller, B. S., Wang, J.-H., & Springer, T. A. (2004). Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. *Nature*, *432*(7013), 59–67. https://doi.org/10.1038/nature02976

Xiong, J.-P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D. L., Joachimiak, A., Goodman, S. L., & Arnaout, M. A. (2001). Crystal Structure of the Extracellular Segment of Integrin αVβ3. *Science (New York, N.Y.)*, *294*(5541), 339–345. https://doi.org/10.1126/science.1064535

Xiong, J.-P., Stehle, T., Goodman, S. L., & Arnaout, M. A. (2003). New insights into the structural basis of integrin activation. *Blood*, *102*(4), 1155–1159. https://doi.org/10.1182/blood-2003-01-0334

Zang, Q., & Springer, T. A. (2001). Amino acid residues in the PSI domain and cysteine-rich repeats of the integrin beta2 subunit that restrain activation of the integrin

alpha(X)beta(2). *The Journal of Biological Chemistry*, *276*(10), 6922–6929. https://doi.org/10.1074/jbc.M005868200

Zhu, J., Boylan, B., Luo, B.-H., Newman, P. J., & Springer, T. A. (2007). Tests of the extension and deadbolt models of integrin activation. *The Journal of Biological Chemistry*, *282*(16), 11914–11920. https://doi.org/10.1074/jbc.M700249200

Zhu, J., Luo, B.-H., Xiao, T., Zhang, C., Nishida, N., & Springer, T. A. (2008). Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces. *Molecular Cell*, *32*(6), 849–861. https://doi.org/10.1016/j.molcel.2008.11.018

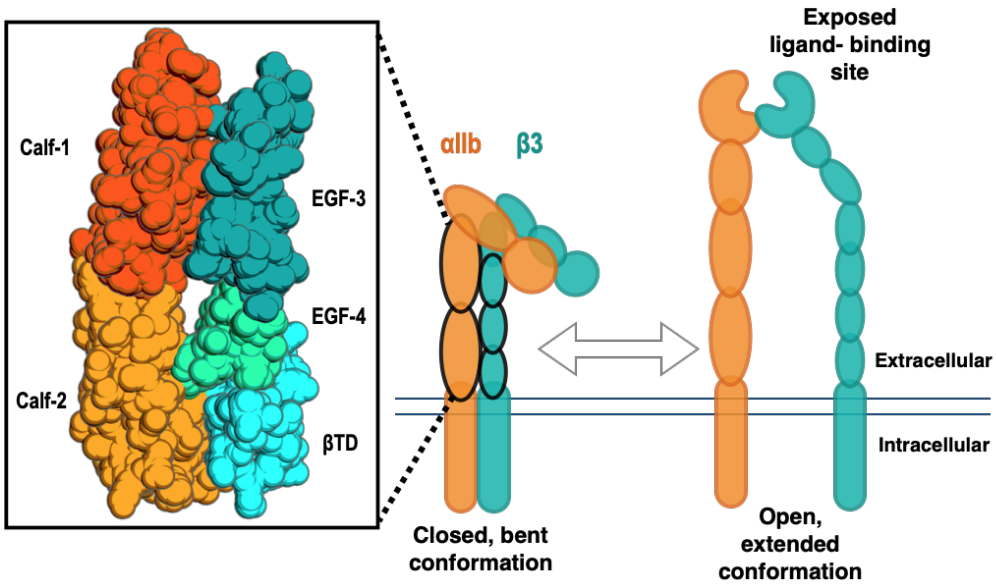

Figure 2.1. Model of conformational states in the αIIbβ3 integrin.

αIIbβ3 undergoes a global conformational shift between its bent inactive and its extended ligand-binding states. The inset is a space-filling model of the distal αIIbβ3 stalk domains, encompassing αIIb residues 599-959 and β3 residues 483-691. This model is derived from the X-ray crystal structure deposited as PDB accession code 3FCS (Zhu et al. 2008).

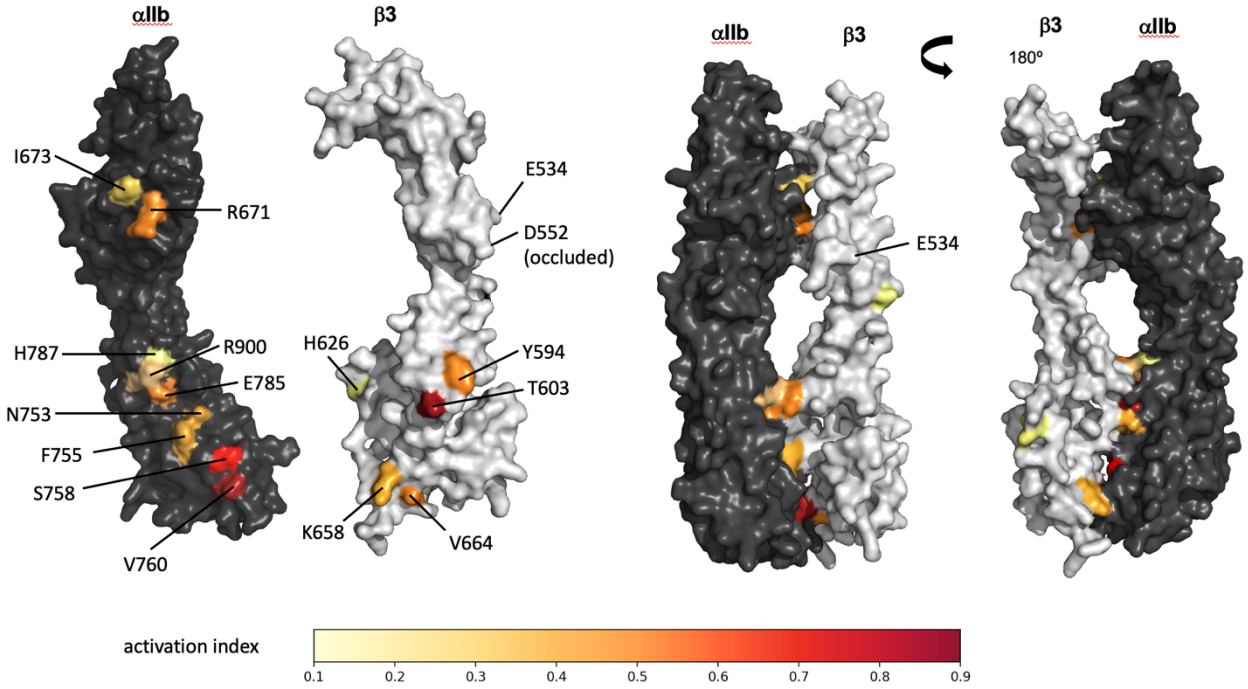

Figure 2.2. Mapping functional hot spots onto the αIIbβ3 structure.

The location of residues in the distal αIIb and β3 stalk domains whose alanine replacements caused constitutive αIIbβ3 activation are mapped onto the structures for these domains. The activation indices of their alanine replacements (Table 2.1) are color-coded according to the heat map below the models.

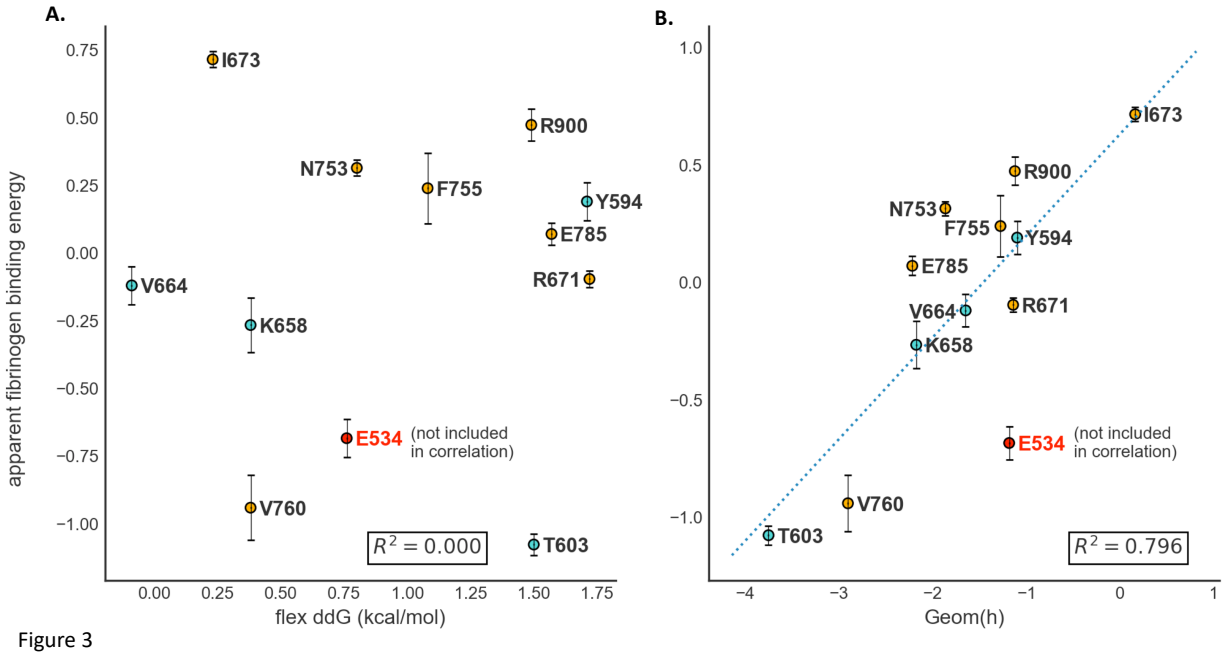

Figure 3

Figure 2.3. Correlations between experimental and computational measurements.

The apparent fibrinogen binding energies of the activating alanine replacements in the αIIb and β3 stalks are measured against **(A)** the corresponding ddGs predicted by flex ddG or **(B)** the corresponding Geom(h) scores calculated from the structural bioinformatics analysis.

The ddG values predicted by flex ddG are shown in Table 2.1, and corresponding Geom(h) scores are shown in Tables 2.2 and 2.3. Changes in the apparent energy of fibrinogen binding resulting from the scanning mutagenesis of the αIIbβ3 stalks with alanine replacements were calculated using Equation 2.3. β3 residue E534A was excluded from the correlation because it is not located in the stalk domain interface.

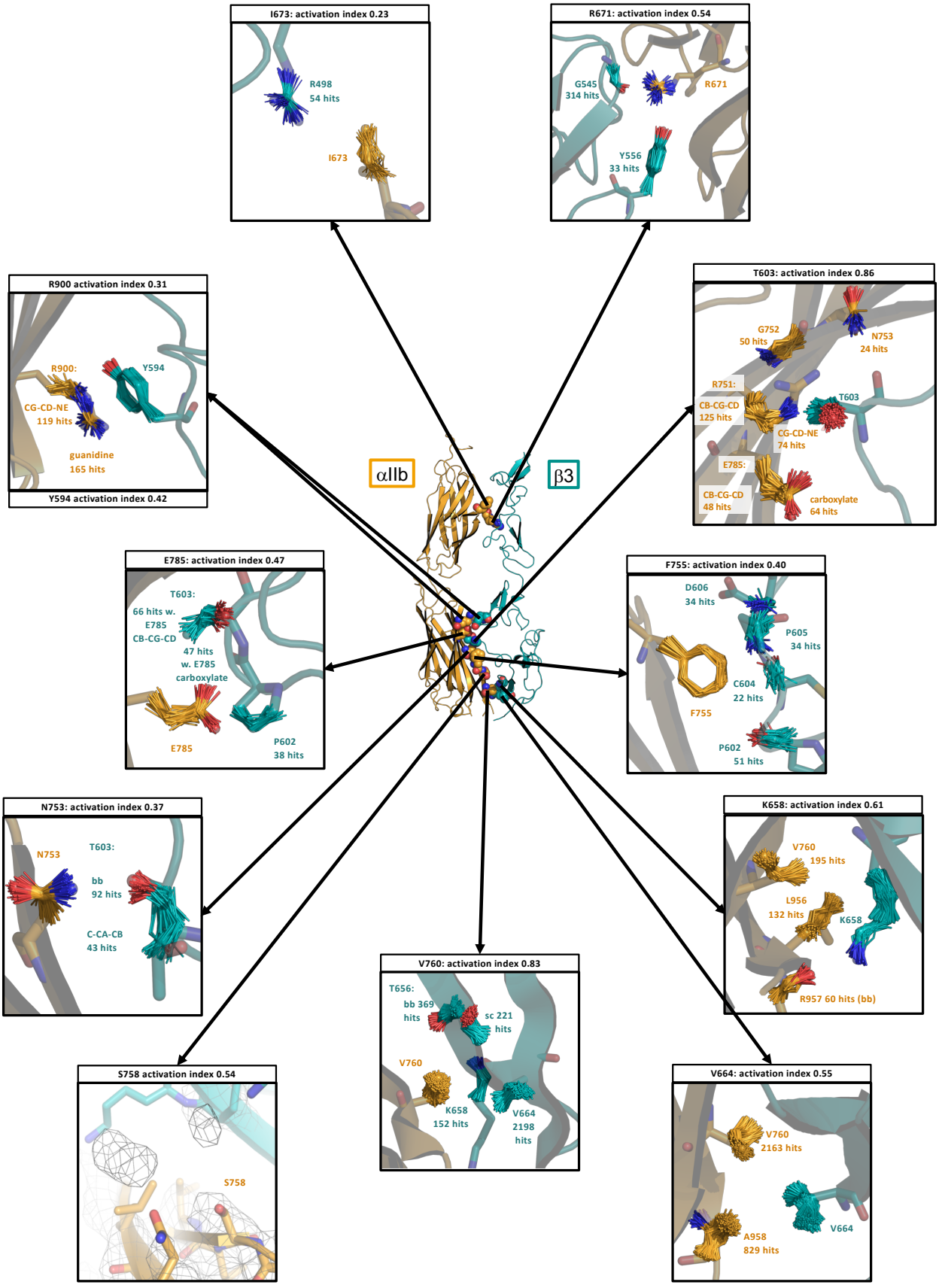
I673: activation index 0.23
R498
54 hits
I673
R671: activation index 0.54
G545
314 hits
R671
Y556
33 hits
R900 activation index 0.31
R900:
CG-CD-NE
119 hits
Y594
guanidine
165 hits
Y594 activation index 0.42
T603: activation index 0.86
G752
50 hits
N753
24 hits
R751:
CB-CG-CD
125 hits
T603
CG-CD-NE
74 hits
E785:
CB-CG-CD
48 hits
carboxylate
64 hits
αIIb
β3
E785: activation index 0.47
T603:
66 hits w.
E785
CB-CG-CD
47 hits
w. E785
carboxylate
E785
P602
38 hits
F755: activation index 0.40
D606
34 hits
P605
34 hits
C604
22 hits
F755
P602
51 hits
N753: activation index 0.37
N753
T603:
bb
92 hits
C-CA-CB
43 hits
K658: activation index 0.61
V760
195 hits
L956
132 hits
K658
R957 60 hits (bb)
V760: activation index 0.83
T656:
bb 369
hits
sc 221
hits
V760
K658
152 hits
V664
2198
hits
S758 activation index 0.54
S758
V664: activation index 0.55
V760
2163 hits
V664
A958
829 hits

Figure 2.4. Interaction geometry analysis on interfacial residues.

This figure illustrates the favorability of interaction geometry between αIIb and β3 stalk domains determined from knowledge-based structural bioinformatics.

The αIIb stalk is shown in orange and the β3 stalk in cyan. Activation index values were derived from Table 2.1. αIIb or β3 residues and their complementary interacting residues, as well as the number of geometric matches in the PDB, are shown in Tables 2.2 and 2.3.

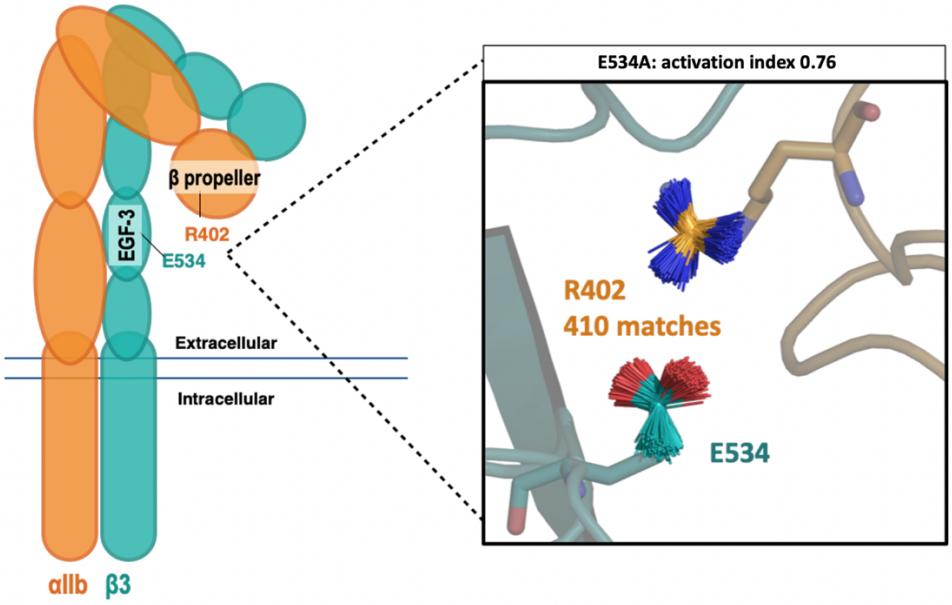

Figure 2.5. E534 on β3 interacting with β-propeller residue R402.

Replacing the β3 residue E534 with alanine causes robust αIIbβ3 activation with an A.I. of 0.76 (Table 2.1). However, E534 is not on the stalk domain interface, but it is in direct contact with αIIb residue R402 located in the αIIb β-propeller domain.

Table 2.1. Comparison of integrin αIIbβ3 stalk domain hot spots and different methods of calculating effect of mutation

Mutation	*Subunit*	*Robetta Alanine Scanning (kcal/mol)*	*αIIbβ3 Activation Index*	*Rosetta flex ddG (kcal/mol)*
F669A	αIIb	0.88	-	0.65
R671A	αIIb	1.53	0.54±0.03	1.72
I673A	αIIb	0.12	0.23±0.03	0.23
N691A	αIIb	0.62	-	2.20
R751A	αIIb	1.49	-	0.65
N753A	αIIb	1.32	0.37±0.03	0.80
F755A	αIIb	1.76	0.40±0.13	1.08
S758A	αIIb	0.56	0.64±0.10	-0.04
V760A	αIIb	0.71	0.83±0.12	0.38
E785A	αIIb	1.20	0.47±0.04	1.57
H787A	αIIb	0.06	0.17±0.02	-0.06
R900A	αIIb	0.54	0.31±0.06	1.49
Q497A	β3	1.98	-	1.86
E534A	β3	0.54	0.76±0.07	0.76
D552A	β3	0	0.17±0.04	0.03
Y556A	β3	0.40	-	0.05
Y594A	β3	0.60	0.42±0.07	1.71
T603A	β3	2.89	0.85±0.04	1.50
D606A	β3	0.71	-	0.42
T609A	β3	1.42	-	0.29
H626A	β3	0	0.19±0.05	0
K658A	β3	0.63	0.61±0.10	0.38
V664A	β3	0.34	0.55±0.02	-0.09

Table 2.2. Structural bioinformatics analysis parameters for hot spots in the αIIb stalk domain. These notations refer to parameters in Equation 2.3.

* $M_{[frag_{Aiib},\ frag_{\beta 3}]}$

** $N_{[AA(frag_{Aiib}),\ AA(frag_{\beta 3})]}$

† $f_{[AA(frag_{Aiib})]} f_{[AA(frag_{\beta 3})]}$

Hotspot residue (h)	Sub-unit	Opposite subunit interacting residues (i)	FG of hotspot residue (FG_h)	FG of interacting residue (FG_i)	# geometric matches to database *	# residue pairs of specified amino acids in database **	Joint frequency of specified residue pair in database †	Geom(h)
R671	aIIb	G545	NE, CZ, NH1 NH2	CA, C, O	314	13585	$\frac{267758}{4879281} \cdot \frac{181792}{4879281}$	-1.14
		Y556	NE, CZ, NH1 NH2	CG, CD1, CD2, CE1, CE2, CZ, OH	33	15870	$\frac{267758}{4879281} \cdot \frac{304073}{4879281}$	
I673	aIIb	R498	CB, CG1, CD1	NE, CZ, NH1, NH2	54	14131	$\frac{444020}{4879281} \cdot \frac{267758}{4879281}$	0.16
N753	aIIb	T603	CB, CG, OD1, ND2	CA, C, O	92	9213	$\frac{151322}{4879281} \cdot \frac{222702}{4879281}$	-1.87
		T603	CB, CG, OD1, ND2	C, CA, CB	43	9213	$\frac{151322}{4879281} \cdot \frac{222702}{4879281}$	
F755	aIIb	P602	CG, CD1, CD2, CE1, CE2, CZ	CA, C, O	51	13529	$\frac{371463}{4879281} \cdot \frac{183440}{4879281}$	-1.28
		C604	CG, CD1, CD2, CE1, CE2, CZ	CA, C, O	22	5859	$\frac{371463}{4879281} \cdot \frac{70567}{4879281}$	
		P605	CG, CD1, CD2, CE1, CE2, CZ	N, CA, C	44	13529	$\frac{371463}{4879281} \cdot \frac{183440}{4879281}$	
		D606	CG, CD1, CD2, CE1, CE2, CZ	N, CA, CB	34	7984	$\frac{371463}{4879281} \cdot \frac{161299}{4879281}$	
V760	aIIb	T656	CB, CG1, CG2	CB, OG1, CG2	369	20646	$\frac{469468}{4879281} \cdot \frac{222702}{4879281}$	-2.91
		T656	CB, CG1, CG2	CA, C, O	221	20646	$\frac{469468}{4879281} \cdot \frac{222702}{4879281}$	
		K658	CB, CG1, CG2	N, CA, CB	152	12247	$\frac{469468}{4879281} \cdot \frac{176327}{4879281}$	
		V664	CB, CG1, CG2	CB, CG1, CG2	2198	62629	$\frac{469468}{4879281} \cdot \frac{469468}{4879281}$	
E785	aIIb	P602	CG, CD, OE1, OE2	CB, CG, CD	38	6875	$\frac{186367}{4879281} \cdot \frac{183440}{4879281}$	-2.22
		T603	CG, CD, OE1, OE2	CB, OG1, CG2	47	9652	$\frac{186367}{4879281} \cdot \frac{222702}{4879281}$	
		T603	CB, CG, CD	CB, OG1, CG2	66	9652	$\frac{186367}{4879281} \cdot \frac{222702}{4879281}$	
R900	aIIb	Y594	NE, CZ, NH1, NH2	CG, CD1, CD2, CE1, CE2, CZ, OH	165	15870	$\frac{267758}{4879281} \cdot \frac{304073}{4879281}$	-1.12
		Y594	CG, CD, NE	CG, CD1, CD2, CE1, CE2, CZ, OH	119	15870	$\frac{267758}{4879281} \cdot \frac{304073}{4879281}$	

Table 2.3. Structural bioinformatics analysis parameters for hot spots in the β3 stalk domain. These notations refer to parameters in Equation 2.3.

* $M_{[frag_{Aiib},\ frag_{\beta 3}]}$

** $N_{[AA(frag_{Aiib}),\ AA(frag_{\beta 3})]}$

† $f_{[AA(frag_{Aiib})]} f_{[AA(frag_{\beta 3})]}$

Hotspot residue (h)	Sub-unit	Opposite subunit interacting residues (i)	FG of hotspot residue (FG_h)	FG of interacting residue (FG_i)	# geometric matches to database *	# residue pairs of specified amino acids in database **	Joint frequency of specified residue pair in database †	Geom(h)
E534	b3	R401	CG, CD, OE1, OE2	NE, CZ, NH1, NH2	410	26488	$\frac{186367}{4879281} \cdot \frac{267758}{4879281}$	-1.19
Y594	b3	R900	CG, CD1, CD2, CE1, CE2, CZ, OH	NE, CZ, NH1, NH2	162	15777	$\frac{304073}{4879281} \cdot \frac{267758}{4879281}$	-1.10
		R900	CG, CD1, CD2, CE1, CE2, CZ, OH	CG, CD, NE	115	15777	$\frac{304073}{4879281} \cdot \frac{267758}{4879281}$	
T603	b3	R751	CB, OG1, CG2	CB, CG, CD	125	11691	$\frac{222702}{4879281} \cdot \frac{267758}{4879281}$	-3.75
		R751	CB, OG1, CG2	CG, CD, NE	74	11691	$\frac{222702}{4879281} \cdot \frac{267758}{4879281}$	
		G752	CB, OG1, CG2	N, CA, C	50	10398	$\frac{222702}{4879281} \cdot \frac{181792}{4879281}$	
		N753	CB, OG1, CG2	CB, CG, OD1, ND2	24	8780	$\frac{222702}{4879281} \cdot \frac{151322}{4879281}$	
		E785	CB, OG1, CG2	CB, CG, CD	48	10312	$\frac{222702}{4879281} \cdot \frac{186367}{4879281}$	
		E785	CB, OG1, CG2	CG, CD, OE1, OE2	64	10312	$\frac{222702}{4879281} \cdot \frac{186367}{4879281}$	
K658	b3	V760	CB, CG, CD	CB, CG1, CG2	195	11477	$\frac{176327}{4879281} \cdot \frac{469468}{4879281}$	-2.18
		L956	CB, CG, CD	CB, CG, CD1	132	16696	$\frac{176327}{4879281} \cdot \frac{697963}{4879281}$	
		R957	CD, CE, NZ	CA, C, O	60	5752	$\frac{176327}{4879281} \cdot \frac{267758}{4879281}$	
V664	b3	V760	CB, CG1, CG2	CB, CG1, CG2	2163	62629	$\frac{469468}{4879281} \cdot \frac{469468}{4879281}$	-1.65
		A958	CB, CG1, CG2	N, CA, CB	829	35441	$\frac{469468}{4879281} \cdot \frac{273846}{4879281}$	

Equation 2.1. Calculation of the integrin activation index.

$$AI = (FB_c - FB_{c+EDTA}) / (FB_{DTT} - FB_{DTT+EDTA}),$$

where FB_c represents fibrinogen binding to αIIbβ3 in the absence of an activating agent; FB_{DTT}, fibrinogen binding to αIIbβ3 induced by 5 mM DTT; FB_{c+EDTA}, constitutive fibrinogen binding to αIIbβ3 in the presence of 2 mM EDTA; and $FB_{DTT+EDTA}$, fibrinogen binding to αIIbβ3 induced by 5 mM DTT in the presence of 2 mM EDTA.

Equation 2.2. Calculation of the apparent free energy of fibrinogen binding.

$$\Delta G_{app} = -RT \ln \left(\frac{activation\ index}{1 - activation\ index} \right)$$

Equation 2.3. Calculation of the interaction geometry metric

$$Geom(h) = \sum_{frag_{Aiib}, frag_{\beta 3}} -RTln \left(\frac{M_{[frag_{Aiib}, frag_{\beta 3}]_{integrin}} / N_{[AA(frag_{Aiib}), AA(frag_{\beta 3})]}}{f_{[AA(frag_{Aiib})]} f_{[AA(frag_{\beta 3})]}} \right)$$

Interaction geometry Geom(h) was defined as the inverse Boltzmann of the observed fraction of database AA_h (hot spot amino acid) interactions with AA_i (interacting amino acid) occurring in the same geometry as in the αIIbβ3 crystal structure, normalized by the expected fraction of residue AA_h coming into contact with AA_i if there were no geometric preference for the interaction. The observed fraction of AA_h interactions with AA_i that were geometric matches to the αIIbβ3 crystal structure was defined as the number of database residue pairs whose fragments could be superimposed onto the interacting fragments in the αIIbβ3 crystal structure with 0.5 Å root mean squared deviation ($M_{[frag_{Aiib}, frag_{\beta 3}]_{integrin}}$), normalized by the total number of $AA_h : AA_i$ residue pairs in the database ($N_{[AA(frag_{Aiib}), AA(frag_{\beta 3})]}$). The expected frequency of this pairwise interaction, assuming no preference for geometry, was defined as the product of the independent frequencies of each amino acid occurring in the database (the denominator).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Sophia Tan
5399D7DE0F16433...
Author Signature

3/20/2023
Date