Family-wide analysis of integrin structures predicted by AlphaFold2

Heng Zhang[1], Daniel S. Zhu[1,3], and Jieqing Zhu[1,2] *

[1]Versiti Blood Research Institute, Milwaukee, WI, USA.

[2]Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI, USA.

[3]D.Z. is a student intern from Brookfield Central High School, Brookfield, WI, USA

*Corresponding author. Email: Jieqing.Zhu@versiti.org

Abstract

Recent advances in highly accurate protein structure prediction by AlphaFold have opened new avenues for analyzing all structures within a single protein family. In this study, we evaluated the capacity of the newly developed AlphaFold2-multimer for predicting integrin heterodimers. Integrins are heterodimeric cell surface receptors made up of a combination of 18 α and 8 β subunits, forming a family of 24 different members. Both α and β subunits contain a large extracellular domain, a short transmembrane domain, and usually a short cytoplasmic domain. Integrins play a wide range of cellular functions by recognizing diverse ligands. Structural studies in recent decades have greatly advanced our understanding of integrin biology, but high-resolution structures have only been determined for a few members of the integrin family. We studied the single-chain atomic structures of 18 α and 8 β integrins in the AlphaFold2 protein structure database. We then applied the AlphaFold2-multimer program to predict the α/β heterodimer structures of all 24 human integrins. The results show a high level of accuracy in the predicted structures for the subdomains of both α and β subunits and provide high-resolution structure information of all integrin heterodimers. Our structural analysis of the entire integrin family reveals a potentially diverse range of conformations among the 24 members and provides a useful structure database for guiding functional studies. However, our results also suggest the limitations of AlphaFold2 structure prediction and thus caution is required in the interpretation and usage of the AlphaFold2 structures.

Introduction

Integrins are cell surface receptors that recognize a variety of extracellular or cell surface ligands, allowing cells to integrate signals from both inside and outside of the cell.[1] The human integrin family consists of 24 members, which result from the combination of 18 α and 8 β subunits (**Fig. 1**). The 24 α/β integrin heterodimers are either widely distributed or specifically expressed in certain cell types, thus playing common or specialized functions in cellular responses related to cell adhesion and migration. Based on ligand or cell specificity, integrins can be classified into subfamilies of RGD (Arg-Gly-Asp) receptors, collagen receptors, laminin receptors, and leukocyte-specific receptors.[1] Integrins are key players in diseases such as thrombosis, inflammation, and cancer, making them prime targets for small molecule or antibody inhibitors.[2-4] Since their discovery in the early 1980s, understanding the structure and function of integrins has been an area of continuous research interest.[5]

The α and β subunits of integrin are composed of multiple subdomains. The α subunit contains β-propeller, thigh, calf-1, calf-2, transmembrane (TM), and cytoplasmic tail (CT) domains (**Fig. 1A-C**). The β subunit contains βI, hybrid, PSI, I-EGF 1-4, β-tail, TM, and CT domains (**Fig. 1A-C**). A subclass of α integrins have an extra αI domain inserted into the β-propeller domain (**Fig.1D-F**). For αI-less integrins, the α β-propeller and β βI domains come together to form the ligand binding site (**Fig. 1A-C**). For αI integrins, the αI domain is responsible for binding ligands (**Fig. 1D-F**). Integrin domains can also be divided into headpiece (containing head and upper legs) and lower legs (**Fig. 1C**). In the past few decades, structural studies of integrins have revealed a conformation-dependent activation and ligand binding mechanism, involving a transition among at least three conformational states.[6, 7] The bent conformation with closed headpiece represents the resting state of integrin (**Fig. 1A, D**), while the extended closed headpiece and extended open headpiece represent the intermediate and high-affinity active states, respectively (**Fig. 1B-C, E-F**). The conformational transition of integrin can be initiated by the binding of intracellular activators, including talin and kindlin, to β CT, resulting in inside-out signaling, or by the binding of extracellular ligands, resulting in outside-in signaling (**Fig. 1A-F**).[8, 9] The conformation-dependent activation model, however, was largely derived from structural studies of the highly-regulated β_2 and β_3 integrins that are primarily expressed in blood cells.[6, 7] Since structural information for most integrin members is lacking, it remains unclear whether the current model of integrin conformational change can be applied to the entire integrin family.

Since the publication of the first high-resolution crystal structure of the $\alpha_V\beta_3$ ectodomain in 2001,[10] significant efforts have been made to determine integrin structures using various methods, including crystallography, negative-stain electron microscopy (EM), nuclear magnetic resonance (NMR), and more recently, cryogenic EM (cryo-EM). However, high-resolution structure information is still limited to a few of integrin members, including $\alpha_V\beta_3$,[10-12] $\alpha_{IIb}\beta_3$,[13-21] $\alpha_5\beta_1$,[22-24] $\alpha_6\beta_1$,[25] $\alpha_X\beta_2$,[26, 27] $\alpha_L\beta_2$,[28] $\alpha_M\beta_2$,[29, 30] $\alpha_V\beta_6$,[31-33] $\alpha_V\beta_8$,[34-37] and $\alpha_4\beta_7$,[38] many of which only have the fragment structures determined to date. Also, among the 24 integrins, only the $\alpha_{IIb}\beta_3$ TM-CT heterodimer structures has been determined experimentally.[20, 39-43] The recent breakthrough in protein structure prediction

using the artificial intelligence-based AlphaFold2 program has provided a powerful tool for analyzing previously hard-to-determine protein structures with a high level of accuracy.[44] We analyzed the predicted atomic structure models of single-chain 18 α and 8 β integrins that are available in the AlphaFold2 database (**Fig. 1G**). Furthermore, using the recently developed AlphaFold2-multimer program,[45] we predicted the structures of all 24 human integrin α/β heterodimers. Our analysis of the integrin family structures revealed potential conformational diversity across its 24 members. We also identified previously unknown structural features, and created a comprehensive database of integrin structures that can guide functional and structural studies. These findings highlight the effectiveness of AlphaFold2 in predicting the structures of large, complex protein families, including integrins.

Methods

Databases and software

The single-chain structures of 18 α and 8 β integrin subunits were downloaded from the AlphaFold2 database (https://alphafold.ebi.ac.uk). Human integrin protein sequences were downloaded from the NCBI protein database (https://www.ncbi.nlm.nih.gov/protein/). Experimentally determined structures, including $\alpha_{IIb}\beta_3$ (PDB 3FCS), $\alpha_V\beta_3$ (PDB 4G1E), $\alpha_5\beta_1$ (PDB 7NXD), and $\alpha_X\beta_2$ (PDB 4NEH), were downloaded from protein databank https://www.rcsb.org. All integrin structures were analyzed in PyMOL version 2.5.4 (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

Running AlphaFold2

AlphaFold 2 Version 2.1.2 was running on the HPC Cluster at the Medical College of Wisconsin using Miniconda3 virtual environments. The AlphaFold2 downloaded reference files are located at: "/hpc/refdata/alphafold". Customized sbatch job script was submitted for structure prediction. Typically, one GPU (--gres=gpu:1) and 100 GB memory (--mem=100gb) was requested to run AlphaFold2. The maximum job running time was set to 48 h (--time=48:00:00). To run AlphaFold2-Multimer for structure prediction of integrin heterodimers, an input fasta file containing the sequences of both integrin α and β subunits was provided. The multimer prediction function was enabled with command "--model_preset=multimer". Full length or extracellular domain structures of integrin heterodimers without signal peptides were predicted with or without templates by setting the parameter of "--max_template_date=2000-05-14" or "--max_template_date=2023-01-01". For integrin $\alpha_6\beta_4$ structure prediction, the large cytoplasmic tail of β_4 was truncated after KGRDV to simplify the prediction. The top ranked models were selected for further analysis.

Comparison of AlphaFold2 predicted integrin structures

The single chain α integrin structures downloaded from AlphaFold2 database were superimposed based on the α_{IIb} calf-2 domain using the "super" command in PyMOL. The

single chain β integrin structures downloaded from AlphaFold2 database were superimposed based on the β_3 βI domain using the super command in PyMOL. The experimentally determined structures for α_{IIb} (PDB 3FCS), α_V (PDB 4G1E), α_5 (PDB 7NXD), and α_X (PDB 4NEH) were superimposed onto the predicted corresponding structures. The experimentally determined structures for β_3 (PDB 3FCS), β_1 (PDB 7NXD), and β_2 (PDB 4NEH) were superimposed onto the predicted structures accordingly. The integrin heterodimer structures predicted by AlphaFold2-multimer with or without TM-CT domains were superimposed onto the calf-2 domain of α_{IIb} in PyMOL. For structure comparison of integrin TM-CT heterodimers, the structures were superimposed based on the α_{IIb} TM domain. The aligned structures were individually oriented to position them perpendicularly to the cell membrane.

DNA constructs

The α_5 with C-terminal EGFP tag (α_5-EGFP) was a gift from Rick Horwitz (Addgene plasmid #15238; http://n2t.net/addgene:15238; RRID:Addgene_15238).[46] The α_9 with C-terminal EGFP tag (α_9-EGFP) was a gift from Dean Sheppard (Addgene plasmid #13600; http://n2t.net/addgene:13600; RRID:Addgene_13600). The α_3, α_7, and α_{10} integrins were cloned into pEGFP-N3 vector. The α_{10}-N839Q mutation was generated by QuickChange mutagenesis kit (Agilent Technologies, Inc).

Flow Cytometry Analysis of LIBS mAb binding

The HEK293T cells were grown in complete DMEM (Corning) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich). Cells were maintained in a 37°C incubator with 5% CO_2. Flow cytometry analysis of integrin expression and LIBS mAb binding were as described previously.[47] In brief, HEK293T cells were transfected with EGFP-tagged α integrin constructs plus β_1 integrin. 48 hours post-transfection, the cells were detached, washed, and resuspended in HBSGB buffer (25 mM HEPES pH 7.4, 150 mM NaCl, 2.75 mM glucose, 0.5% BSA) containing 1 mM Ca^{2+}/Mg^{2+} or 0.1 mM Ca^{2+} plus 2 mM Mn^{2+}. Cells were incubated with either 9EG7 mAb (BD Biosciences) or MAR4 (BD Biosciences) for 15 mins followed by additional 15 min incubation with Alexa Fluor 647-conjugated goat anti- rat or mouse IgG. Surface binding of mAb was measured by a BD Accuri™ C6 (BD Biosciences). Relative surface expression of extended β_1 integrins were normalized to total surface β_1 integrins and plotted with Prism 9.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The predicted structures were deposited online as supplementary materials.

Results

The AlphaFold2 structures of 18 α and 8 β human integrins

We extracted the structures of 18 human α integrins from the AlphaFold2 protein structure database. All the structures in the database were predicted based on the full-length single-chain amino acid sequence, including the signal peptide, ectodomain, TM and CT domains. To compare the ectodomain structures of all the α integrins, we superimposed the structures based on the calf-2 domain of α_{IIb}, and then the individual structures were oriented vertically to membrane normal and rotated as necessary to show the position of β-propeller domain relative to cell membrane (**Fig. 2**). The structures were grouped based on ligand or cell specificity. For all the structures, the correct folding of individual domains was successfully predicted (**Fig. 2**). The four RGD-binding α integrins all show a sharp bent conformation, with the α_{IIb} and α_V nearly identical to their crystal structures (**Fig. 2A**). However, the α_5 AlphaFold2 structure is more bent than its half-bent cryo-EM structure (**Fig. 2A)**. For the three laminin receptors, only α_6 is in a sharp bent conformation as the RGD receptors, while α_3 and α_7 are in a half-bent conformation (**Fig. 2B**). The α_4 and α_9 integrins are also in the bent conformation as the RGD receptors (**Fig. 2C**). Interestingly, all the four α integrins of collagen receptors are more extended than bent, with α_{10} exists in a nearly fully extended conformation (**Fig. 2D**). The five leukocyte-specific α integrins also display conformational diversity, with α_L and α_X more bent than α_M, α_D, and α_E (**Fig. 2E**). The AlphaFold2 structure of α_X is nearly identical to the crystal structure (**Fig. 2E**).

We compared the structures of 8 human β integrins that were predicted as single-chain structures in the AlphaFold2 protein structure database. The structures were superimposed based on the β_3 βI domain and then orientated individually to position them vertically to membrane normal. As shown in **Fig. 3**, the correct folding of individual domains including βI, PSI, hybrid, I-EGF domains, and β-tail domain (β-TD) were accurately predicted for β_1 to β_7 integrins. The β-TD of β_8 is smaller than other β integrins and its structure was incompletely predicted (**Fig. 3, β_8**). The AlphaFold2 structures of β_3, β_2, β_4, β_5, β_6, and β_7 all assembly a bent conformation as seen in the crystal structures of β_3 and β_2, while the β_1 and β_8 structures are less bent (**Fig. 3**). The half-bent conformation of β_1 AlphaFold2 structure is comparable to the cryo-EM structure (**Fig. 3 β_1**).

The domain interface where α and β integrin undergo extension

Previous structure studies revealed that the extension of α integrin happens at the interface between thigh and calf-1 domains, where located a disulfide bonded knob, denoted as genu (**Fig. 1C, F**). We did sequence alignment of all α integrins for the junction of thigh and calf-1 domains (**Fig. 4A**). The structure of α_{IIb} was used as an example to show the interface between thigh and calf-1 domains at bent conformation (**Fig. 4B**). The interfacial residues were shown in red in the sequence alignment (**Fig. 4A**) and as red sticks in the structure (**Fig. 4B**). Sequence alignment shows that the interfacial residues as well as the putative N-glycan sites are not well conserved (**Fig. 4A**). Some α integrins, such as α_V, α_8, α_4, α_9, α_{10}, and α_E have putative N-glycan sites on the interface of either thigh or calf-1. Interestingly, the laminin receptors α_3, α_6, and α_7 all have a longer

interfacial loop (region 1) on calf-1 (**Fig. 4A**). However, there are no signature sequences that appear to prefer a bent or extended conformation.

Integrin β subunit extends at the I-EGF-1 and I-EGF-2 junction (**Fig. 1C, F**). Sequence alignment of 8 human β integrin at this region showed no obvious residue conservation except the classical disulfide bonds of EGF domains (**Fig. 4C**). In the bent conformation of β_3 integrin (**Fig. 4D**), the interface between I-EGF-1 and I-EGF-2 is much smaller compared with the interface between thigh and calf-1 in α_{IIb} (**Fig. 4B**), and unlikely plays a major role in maintaining the bent structure. However, the length of C1-C2 loop in I-EGF-2 domain has been shown to regulate integrin extension.[48] A landmark disulfide bond is missing in the I-EGF-1 of β_8 (**Fig. 4C**), which may contribute at least in part to the distinct conformational regulation of β_8 integrin.

α_{10} integrin prefers an extended conformation on cell surface

Of the 18 α integrins, the AlphaFold2 structure of α_{10} presents an extended conformation (**Fig. 2D**). The extended structure was also seen for α_{10} of other species, including mouse, rat, and zebrafish (**Fig. 5A**). We also observed a conserved putative N-glycan site (N839 in human) at the thigh/calf-1 interface of α_{10} integrin (**Fig. 5A**). The N-glycans at this site may interfere with the bent conformation (**Fig. 5B**). α_{10} integrin only forms heterodimer with β_1 integrin. The AlphaFold2 structures of human, mouse, cat, and chicken β_1 are all in a half-bent conformation (**Fig. 5C**). To assess the conformation of $\alpha_{10}\beta_1$ on cell surface, we used mAb 9EG7, which recognizes the β_1 I-EGF-2 epitope that is masked in the bent conformation (**Fig. 5C**),[49] thus reporting β_1 integrin extension. α_3, α_5, α_7, and α_9 integrins were used for comparison. The α integrins were expressed as EGFP-fusion proteins along with β_1 in HEK293T cells. Flow cytometry analysis showed that, under the physiological metal ion condition (1 mM Ca^{2+}/Mg^{2+}), α_{10}-EGFP/β_1 cells bound more 9EG7 than other integrins (**Fig. 5D**). By contrast, α_7-EGFP/β_1 and α_9-EGFP/β_1 cells only bound more 9EG7 in the activating condition (0.1 mM Ca^{2+}/ 2 mM Mn^{2+}), while α_3-EGFP/β_1 had no response to Mn^{2+} (**Fig. 5D**). α_5-EGFP/β_1 also didn't response well to Mn^{2+} for 9EG7 binding compared with α_7 and α_9, but the RGD-like compound, MK-0429, greatly induced 9EG7 binding (**Fig. 5D**). α_{10}-EGFP/β_1 bound 9EG7 to a high level at both Mg^{2+} and Mn^{2+} conditions. However, mutating the putative N-glycan site (α_{10}-N839Q) didn't affect 9EG7 binding (**Fig. 5D**). These data suggest that α_{10} integrin exists in a constitutively extended conformation on cell surface.

The integrin α/β heterodimer structures predicted by AlphaFold2-multimer

The integrin structures in AlphaFold2 database are only for single-chain α and β subunits. We utilized the AlphaFold2-multimer module to predict the α/β heterodimer structures of all 24 human integrins. To avoid any potential model bias from known integrin structures, we set the template search date at the year 2000, prior to which no integrin structures had been reported. All 24 integrin structures were successfully predicted by AlphaFold2-multimer (**Fig. 6**). All the structures were superimposed onto the α_{IIb} calf-2 domain and then orientated individually to position the ectodomains vertically to membrane normal and grouped based on ligand or cell specificity (**Fig. 6**). Overall, the inter-subunit

interfaces such as that of α β-propellor and β βI domains were correctly predicted. The $\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$ ectodomain structures are very close to the crystal structures with Cα RMSD about 2 Å. All the RGD receptors assembled a bent conformation, except $\alpha_5\beta_1$ is half-bent (**Fig. 6A**). Similarly, all the laminin receptors are also bent (**Fig. 6B**). The four collogen receptors all show a half-bent structure, including $\alpha_{10}\beta_1$ (**Fig. 6C**). $\alpha_{10}\beta_1$ was suggested to be more extended based on 9EG7 mAb binding (**Fig. 5D**). Among the leukocyte-specific integrins, only $\alpha_L\beta_2$ and $\alpha_E\beta_7$ are sharp bent, while $\alpha_M\beta_2$, $\alpha_X\beta_2$, $\alpha_D\beta_2$, and $\alpha_E\beta_7$ are more extended (**Fig. 6D**). The $\alpha_4\beta_1$ and $\alpha_9\beta_1$ are in half bent conformation (**Fig. 6E**). The relative orientation of TM domains to the cell membrane are incorrectly predicted for most of the structures.

Since 9 of the integrin structures predicted by AlphaFold2-multimer show interactions between TM-CT and ectodomains (**Fig. 7A**), we asked if such artificial interactions affect the overall integrin structure prediction. We re-calculated the 9 integrin structures without TM-CT domains using AlphaFold2-multimer. The structures show essentially the same conformations as those with TM-CT domains (**Fig. 7B**), suggesting that the folding of ectodomain and TM-CT does not influence each other during the structure calculation by AlphaFold2-multimer.

Next, we investigated whether enabling the template option for AlphaFold2-multimer had any impact on the calculation of integrin structures. The $\alpha_5\beta_1$, $\alpha_{10}\beta_1$, $\alpha_V\beta_8$, and $\alpha_X\beta_2$ integrins were selected for the test. When setting the template searching date at the year of 2023, all four integrin structures calculated by AlphaFold2-multimer showed a sharp bent conformation (**Fig. 8A-D**), which closely resembled the crystal structure of bent $\alpha_{IIb}\beta_3$ (**Fig. 8A**). Notably, the AlphaFold2-multimer predicted $\alpha_5\beta_1$ structure with a template search date set at 2023 is much more bent than the one with a search date set at 2000, which closely resembles the $\alpha_5\beta_1$ cryo-EM structure (**Fig. 8A**). Similarly, the $\alpha_{10}\beta_1$-2023 is bent comparing with half bent $\alpha_{10}\beta_1$-2000 structure (**Fig. 8B**). However, both $\alpha_V\beta_8$-2023 and $\alpha_V\beta_8$-2000 structures are bent to a similar level (**Fig. 8C**). In sharp contrast, the $\alpha_X\beta_2$-2023 structure shows a conformation resembling the bent $\alpha_X\beta_2$ crystal structure, while the $\alpha_X\beta_2$-2000 structure is extended (**Fig. 8D**). These findings suggest that the inclusion of template structures can significantly affect the structure prediction outcomes of AlphaFold2-multimer.

Structures of integrin TM-CT domains

Despite the relative simplicity of the sequence and structure of integrin TM and CT domains, only the structure of the $\alpha_{IIb}\beta_3$ TM-CT heterodimer has been experimentally determined to date. Sequence alignment of the TM-CT domains from 8 β and 18 α human integrins reveals conservative features at the TM, membrane-proximal (MP), and membrane-distal (MD) regions (**Fig. 9A**). We analyzed the 24 integrin TM-CT structures calculated by AlphaFold2-multimer (**Fig. 9B**). Structure alignment reveals a high degree of structural similarity for the heterodimers at TM domain, highlighting the conserved α GXXXG motif and β small G/A residue on α/β interface (**Fig. 9B**). The conserved GFFKR motifs in the CT MP regions of 18 α integrins all adopt a reverse turn conformation, while the β CT MP regions, except for β_4 and β_8 integrins, all display an α-helical structure

extended from the TM region, with the conserved Asp residue located on α/β interface (**Fig. 9B**). The β CT Asp residue is proximal to the Arg residue in the α GFFKR motif (**Fig. 9B**), which was proposed to form a salt bridge interaction.[50] The CT MD regions of both α and β subunits exhibit diverse disordered conformations, including the conserved NPXY motif that binds talin, as shown in Figure 9B. Despite not allowing any integrin TM-CT structure templates during the AlphaFold2-multimer calculation, the predicted structure of $\alpha_{IIb}\beta_3$ TM-CT closely resembles the experimentally determined structure (**Fig. 9C**). We also generated a prediction for the $\alpha_{IIb}\beta_3$ TM-CT structure in the absence of the ectodomain, which showed a similar TM interface to that predicted with the ectodomain present (**Fig. 9D**). These results suggest that AlphaFold2-multimer is capable of accurately predicting integrin TM structures.

Discussion

Sequence alignment analysis revealed the averaged sequence identity was 30-40% among the 8 β integrins, and 20-40% among the 18 α integrins.[38] For individual integrin domains, such as the βI domain, the sequence identity can be up to over 60%.[38] Since AlphaFold2 incorporates amino acid sequence, multiple sequence alignments and homologous structures in its structure calculation,[44] the high sequence identity observed among integrin domains may facilitate the accurate prediction of integrin domain structures for most of the family members. This is demonstrated by the comparison of predicted structures with the experimental structures of $\alpha_{IIb}\beta_3$, $\alpha_V\beta_3$, and $\alpha_X\beta_2$, which show a high degree of structural similarity. Thus, the predicted integrin domain structures can be utilized with a high level of confidence.

Our sequence and structure analysis of the inter-domain interface in bent integrin conformations did not reveal any signature sequences that clearly favor either a bent or extended conformation for both α and β subunits. Despite using bent α_{IIb}, α_V, and α_X structures in the PDB as part of its training sets, AlphaFold2 was able to predict a diverse range of conformations for the 18 single-chain α structures, ranging from sharp bent to almost fully extended structures, as seen in the case of α_{10}. In contrast, the eight single-chain β structures predicted by AlphaFold2 were mostly bent. Also, AlphaFold2-multimer predicated both bent and extended conformations of α/β heterodimers. While our functional assay suggested that $\alpha_{10}\beta_1$ may adopt an extended conformation on the cell surface, consistent with the predicted single-chain α_{10} structure, the heterodimer structure of $\alpha_{10}\beta_1$ predicted by AlphaFold2-multimer does not exhibit such a conformation. Since integrins can adopt multiple conformational states, the predicted structures by AlphaFold2 may reflect such conformational diversity. We observed that the AlphaFold2-multimer prediction may be influenced by homologous structures, which could potentially introduce bias in predicting the overall conformation of integrins. AlphaFold2-multimer has demonstrated remarkable success in accurately predicting the structures of protein complexes, including those with transient interactions, multiple subunits, and large interfaces.[51, 52] Here, we showed that the AlphaFold2-multimer algorithm can successfully predict large complexes of integrin heterodimer structures.

Although AlphaFold2 has its limitations as many other protein structure prediction programs, its ability to predict integrin domain structures, inter- and intra-domain interfaces, and overall domain organization with an acceptable level of accuracy makes the predicted structures highly useful for designing both functional and structural studies. These structures can be used to analyze interesting N-glycan sites, functional mutations, antibody epitopes, and design constructs for integrin expression and purification. Additionally, the structures can be used for molecular dynamics simulations, ligand docking, and interpreting some functional data. However, caution is needed when using these predicted structures without experimental validation.

Author contributions

J.Z. designed and supervised the research. H.Z. performed the prediction using AlphaFold2-multimer and the 9EG7 binding assay. H.Z, D.Z., and J.Z. analyzed the data, prepared the figures, and wrote the manuscript.

Acknowledgement

This work was supported by the grant R01 HL131836 (to J. Zhu) from the Heart, Lung, and Blood Institute of the National Institute of Health. We thank the Research Computing Center at the Medical College of Wisconsin for providing the help and resources in running the AlphaFold2 program.

Figures and legends

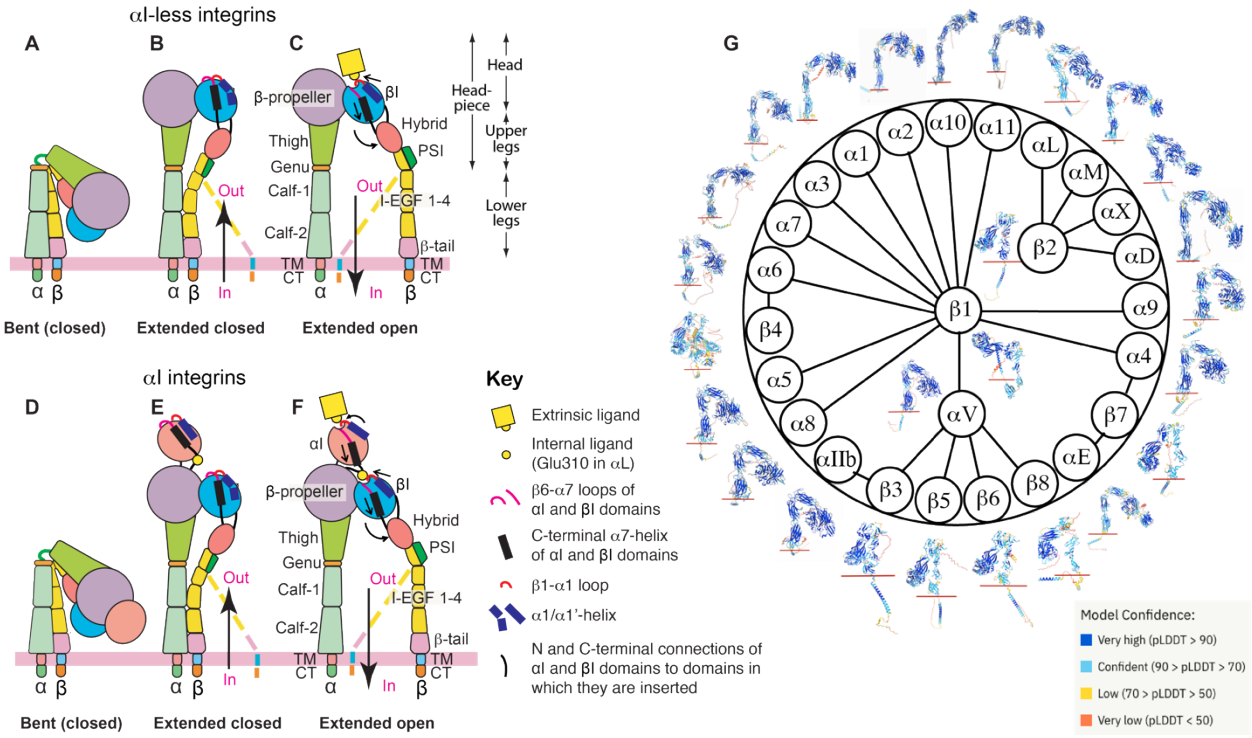

Figure 1. Integrin domain organization and structures predicted by AlphaFold2. (**A-F**) Integrin domain organization and conformational changes during activation. The αI-less (A-C) and αI-containing (D-F) integrins are shown separately. The global and local structural changes are indicated. The dashed lines indicate the alternative conformations of the β leg domains. (**G**) Integrin family and overall structures of α and β subunits predicted by AlphaFold2. The red line indicates the boundary between the extracellular and transmembrane (TM) domain. The structure images are color coded based on model confidence.

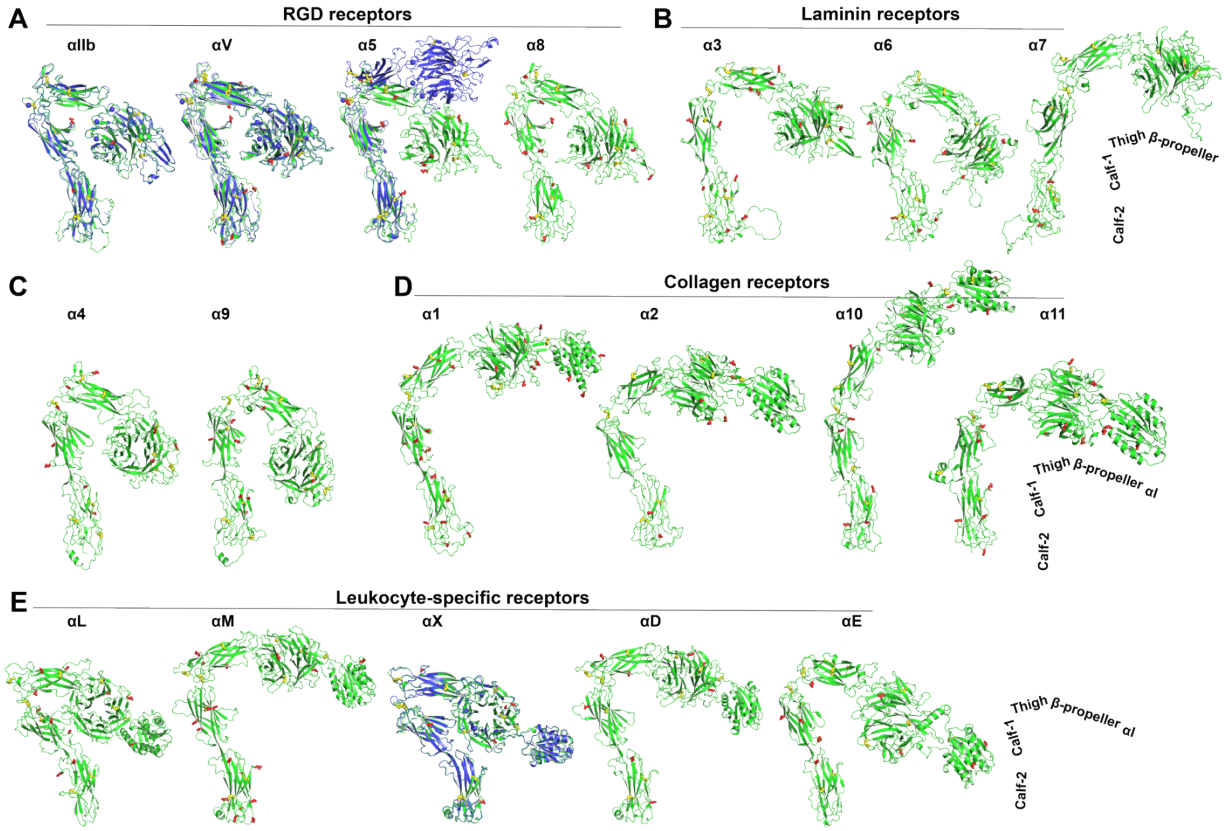

Figure 2. AlphaFold2 structures of the ectodomains of 18 human α integrins. The Alphafold2 structures are shown in green. The experimentally determined structures for α_{IIb} (PDB 3FCS), α_V (PDB 4G1E), α_5 (PDB 7NXD), and α_X (PDB 4NEH) are shown in blue. The putative N-linked glycosylation sites are shown as red sticks. Disulfide bonds are yellow sticks. The structures were aligned based on the calf-2 domain and oriented perpendicularly to cell membrane.

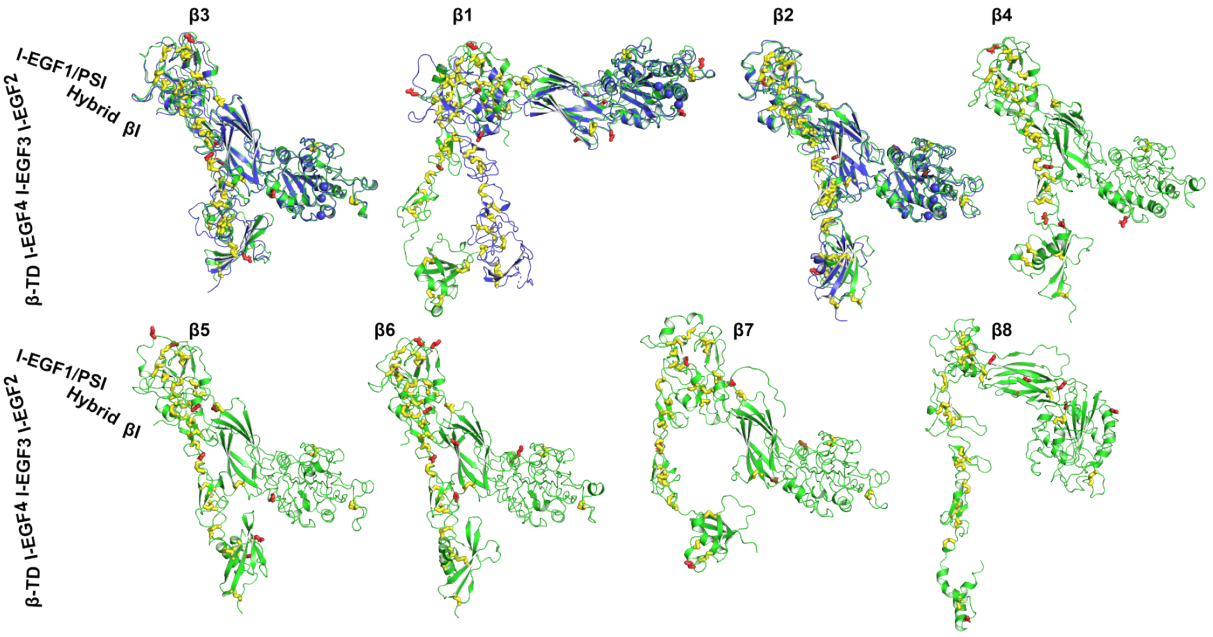

Figure 3. **AlphaFold2 structures of 8 human β integrins**. The AlphaFold2 structures are shown in green. The experimentally determined structures for β_3 (PDB 3FCS), β_1 (PDB 7NXD), and β_2 (PDB 4NEH) are shown in blue. The putative N-linked glycosylation sites are shown as red sticks. Disulfide bonds are yellow sticks. The structures were aligned based on the β_3 βI domain and oriented perpendicularly to cell membrane.

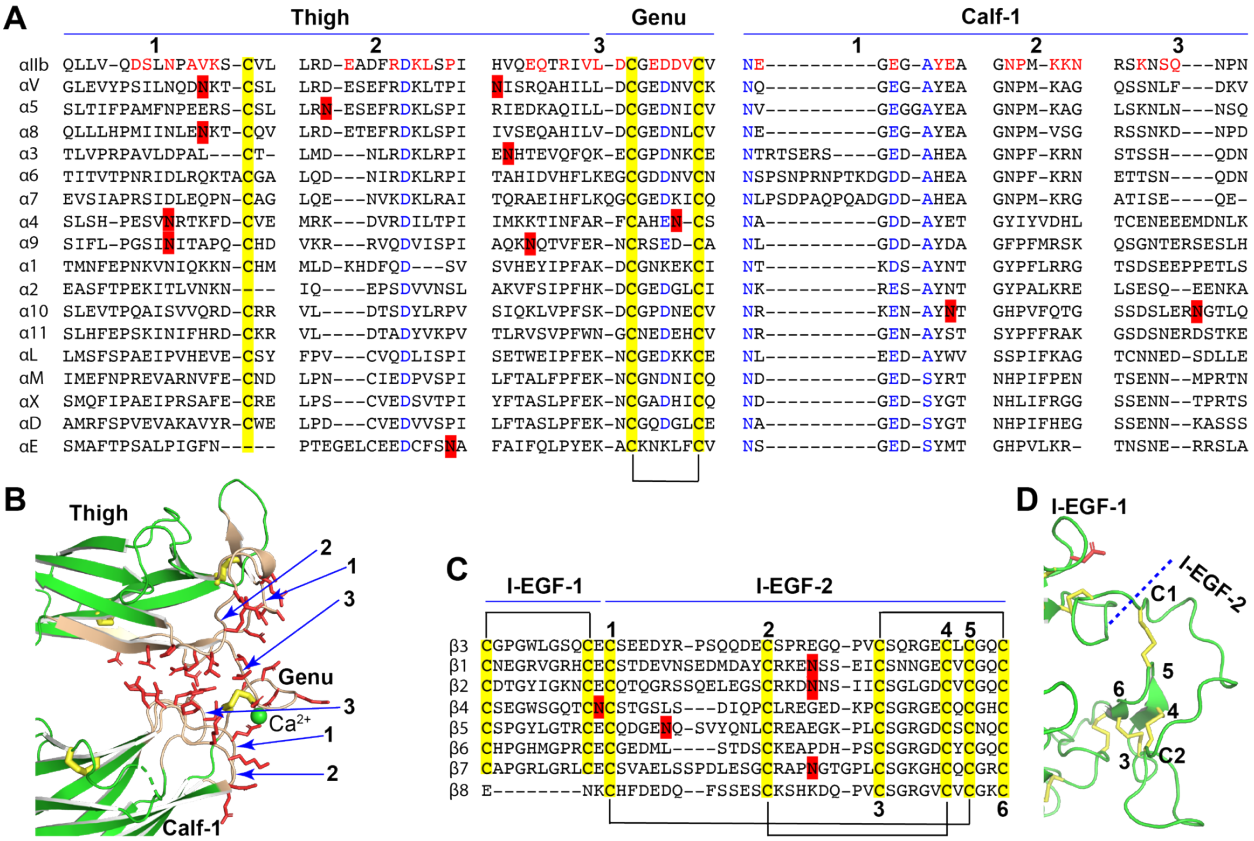

Figure 4. **Sequence and structure of the domain interface where integrin becomes extended**. (**A**) Sequence alignment of human α integrin thigh and calf-1 domain junction interface. The interfacial residues of α_{IIb} are in red. Highly conserved residues are in blue. Disulfide bonds are highlighted in yellow. The putative N-glycan sites are highlighted in red. (**B**) The interface of thigh and calf-1 of bent α_{IIb} structure. The loops at the interface are numbered in panels A and B. The interfacial residues are shown as red sticks. Disulfide bonds are yellow sticks. (**C**) Sequence alignment of human β integrin I-EGF-1 and I-EGF-2 junction. Disulfide bonds are highlighted in yellow. The putative N-glycan sites are highlighted in red. The six cysteines of I-EGF-2 are numbered. (**D**) Structure of β_3 I-EGF-1 and I-EGF-2 junction. Disulfide bonds are yellow sticks. One N-glycan site on I-EGF-1 is shown as a red stick.

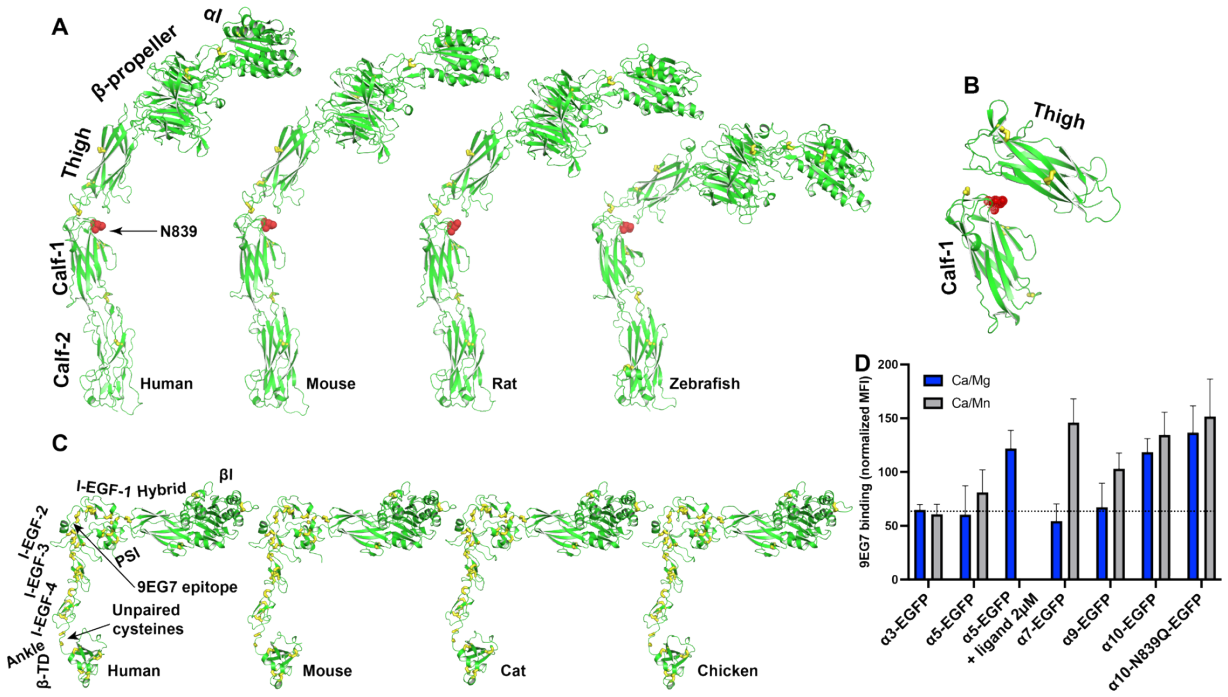

Figure 5. **AlphaFold2 structures of α_{10} and β_1 integrins from different species**. (**A**) Structures of α_{10} integrins of human, mouse, rat, and zebrafish. The structures were aligned based on calf-2 domain and oriented perpendicularly to cell membrane. The N-glycan site at the thigh/calf-1 interface is shown as a red stick. (**B**) The relative position of α_{10} thigh and calf-1 domains in the fully bent conformation, indicating the interfacial location of the N-glycan site shown as red sticks. The thigh and calf-1 domains of α_{10} were superimposed on those of bent αIIb structure. (**C**) Structures of β_1 integrins of human, mouse, cat, and chicken. The structures were aligned based on βI domain and oriented perpendicularly to cell membrane. (**D**) Conformation of β_1 integrin co-expressed with selected integrin α subunit. Human integrin α subunits with C-terminal EGFP tag were co-expressed with human β_1 in 293T cells. The binding of mAb 9EG7 or MAR4 IgG was measured by flow cytometry in the buffer containing 1 mM Ca^{2+}/Mg^{2+} or 0.1 mM Ca^{2+} plus 2 mM Mn^{2+}. The data are presented by the MFI of 9EG7 binding as a percentage of the MFI of MAR4 binding.

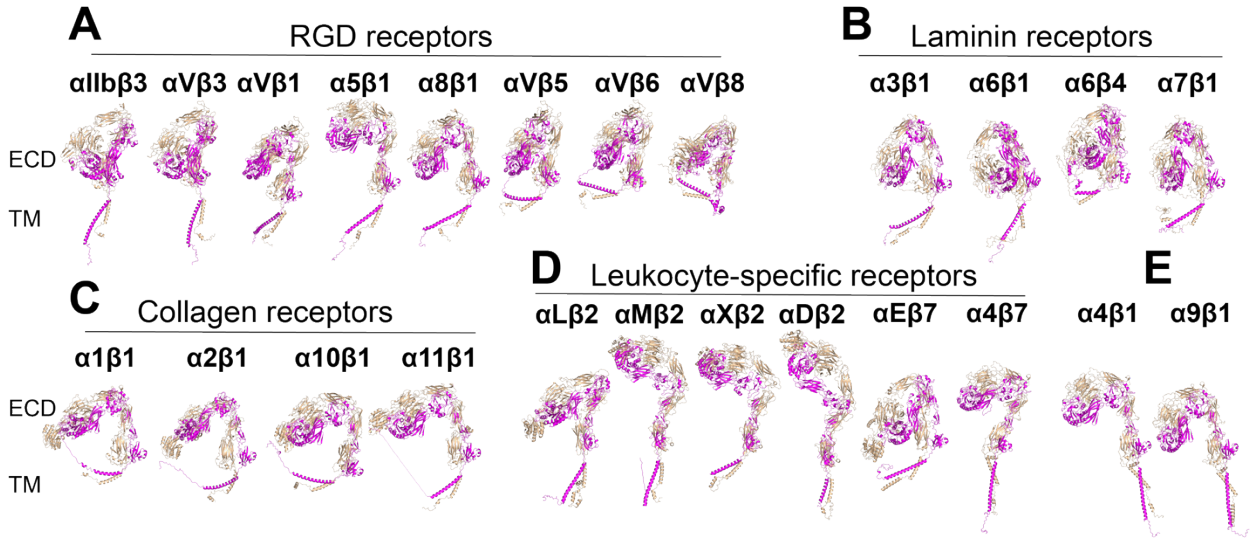

Figure 6. Structures of 24 human integrins predicted by AlphaFold2-multimer. (**A-E**) The full-length integrin structures were calculated by AlphaFold2-multimer with the max_template_date of 2000. Structures were superimposed onto the calf-2 domain of α_{IIb} subunit and shown in the same orientation. The α and β subunits are shown in wheat and magenta, respectively.

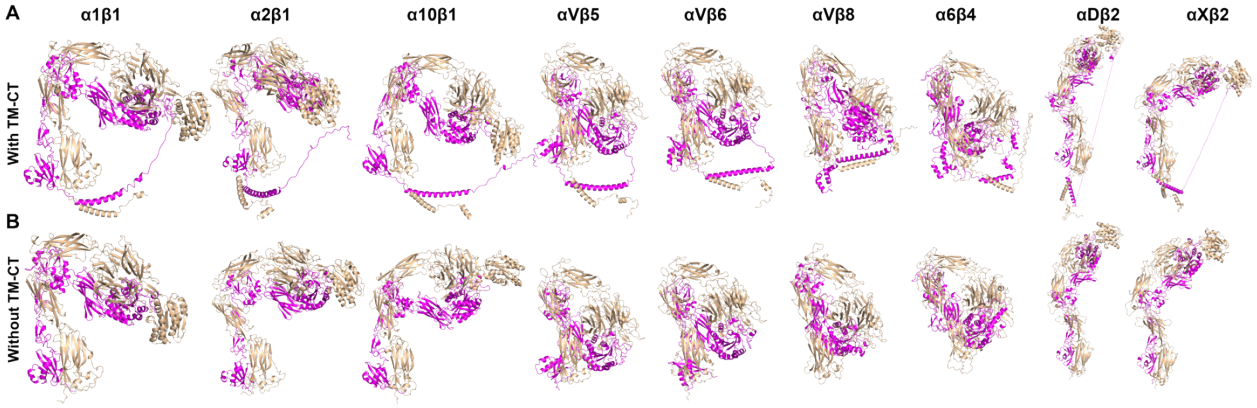

Figure 7. Comparison of structures predicted by AlphaFold2-multimer with and without transmembrane and cytoplasmic domains. (**A**) The full-length structures of some integrins calculated by Alphafold2-multimer show artificial contacts between ectodomain and TM-CT domains. (**B**) The structures in panel A were re-calculated in the absence of TM-CT domains. The structures were superimposed based on the calf-2 domain of α subunit and shown in the same orientation. The α and β subunits are shown in wheat and magenta, respectively.

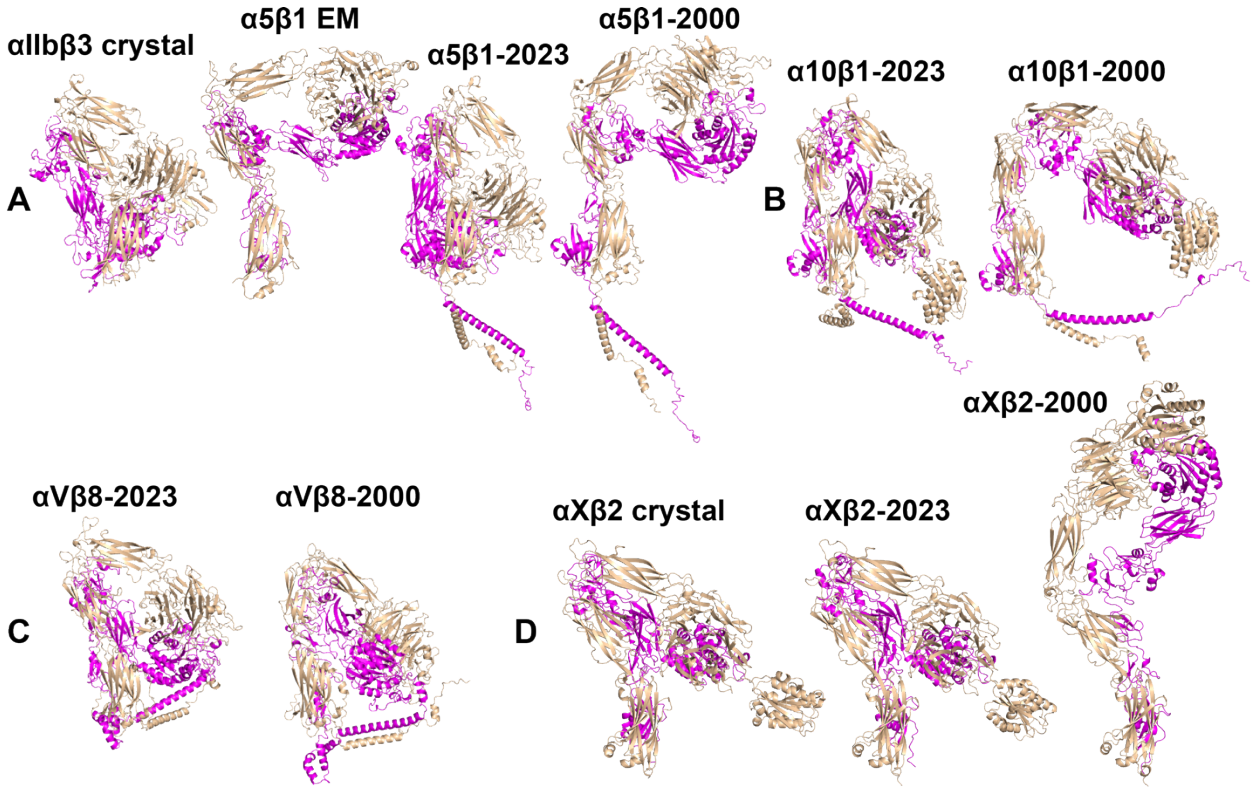

Figure 8. Comparison of selected structures predicted by AlphaFold2-multimer with and without enabling the option of using known structures as templates. (**A**) The full-length $\alpha_5\beta_1$ structures were calculated by AlphaFold2-multimer with the max_template_date of 2023 or 2000. The crystal structure of $\alpha_{IIb}\beta_3$ (PDB 3FCS) and cryo-EM structure of $\alpha_5\beta_1$ (PDB 7NXD) were shown for comparison. (**B**) AlphaFold2 structures of full-length $\alpha_{10}\beta_1$ with the max_template_date of 2023 or 2000. (**C**) AlphaFold2 structures of full-length $\alpha_V\beta_8$. (**D**) AlphaFold2 structures of ectodomain of $\alpha_X\beta_2$. The crystal structure of $\alpha_X\beta_2$ (PDB 4NEH) was shown for comparison.

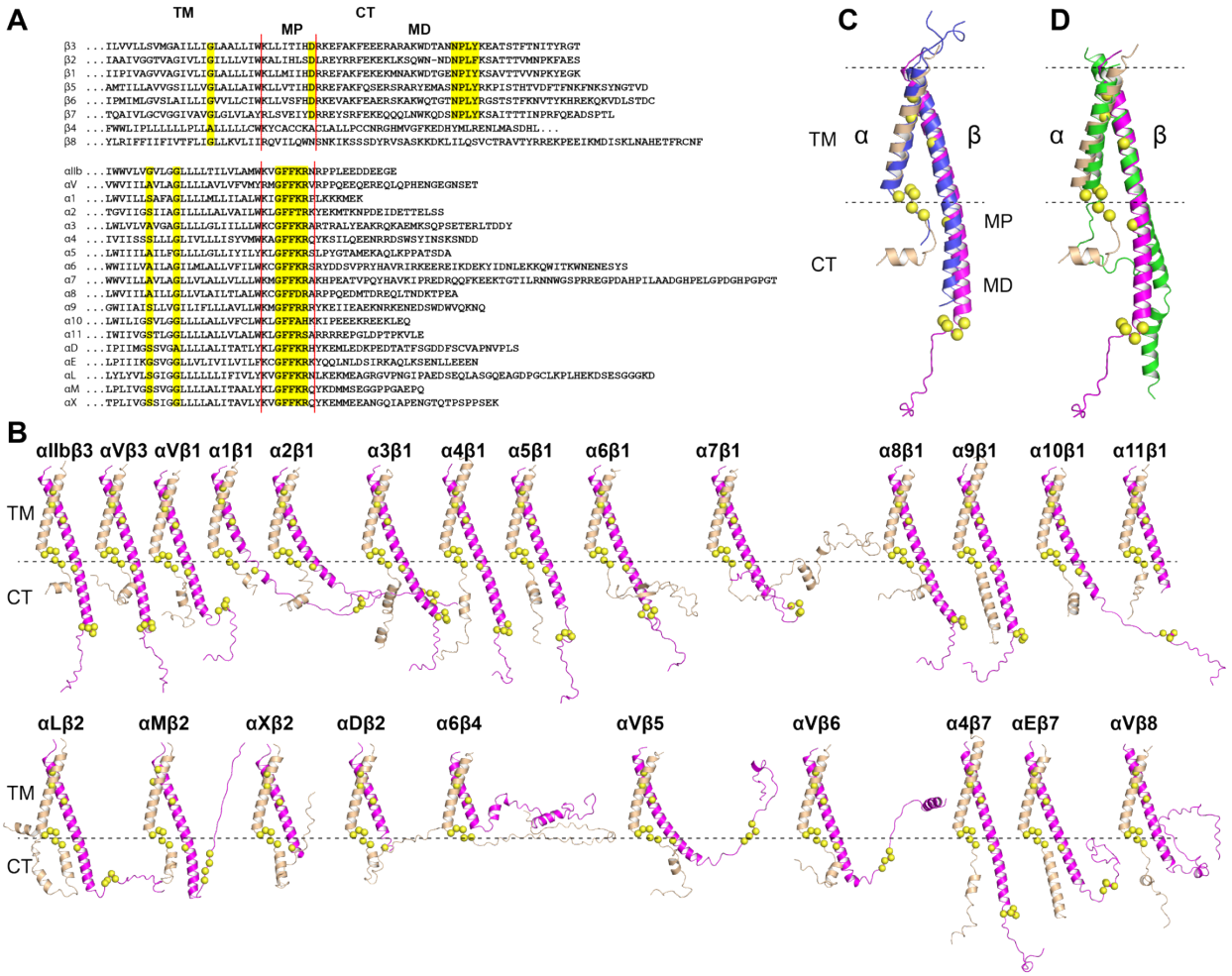

Figure 9. Structures of integrin transmembrane and cytoplasmic domains predicted by AlphaFold2-multimer. (**A**) Sequence alignment of human β and α integrin TM and CT domains. The conserved residues are highlighted in yellow. The boundaries of membrane proximal (MP) and membrane distal (MD) regions of CT domain are indicated. (**B**) AlphaFold2 structures of 24 human integrin TM-CT domains. The structures were superimposed onto the TM domain of α_{IIb} subunit and shown in the same orientation. The conserved residues highlighted in panel A are shown as yellow Cα spheres. (**C**) Superimposition of AlphaFold2-predicted $\alpha_{IIb}\beta_3$ TM-CT structure (in wheat and magenta) on the heterodimeric structure of $\alpha_{IIb}\beta_3$ TM-CT structure determined by disulfide crosslinking and Rosetta modeling (in blue). The α and β subunits are shown in wheat and magenta, respectively. (**D**) Superimposition of $\alpha_{IIb}\beta_3$ TM-CT structure predicted without (green) and with ectodomain (wheat and magenta).

References

[1] Hynes, R. O. (2002) Integrins: bi-directional, allosteric, signalling machines, *Cell 110*, 673-687.

[2] Slack, R. J., Macdonald, S. J. F., Roper, J. A., Jenkins, R. G., and Hatley, R. J. D. (2022) Emerging therapeutic opportunities for integrin inhibitors, *Nat Rev Drug Discov 21*, 60-78.

[3] Ley, K., Rivera-Nieves, J., Sandborn, W. J., and Shattil, S. (2016) Integrin-based therapeutics: biological basis, clinical use and new drugs, *Nat Rev Drug Discov 15*, 173-183.

[4] Pang, X., He, X., Qiu, Z., Zhang, H., Xie, R., Liu, Z., Gu, Y., Zhao, N., Xiang, Q., and Cui, Y. (2023) Targeting integrin pathways: mechanisms and advances in therapy, *Signal Transduct Target Ther 8*, 1.

[5] Hynes, R. O., Ruoslahti, E., and Springer, T. A. (2022) Reflections on Integrins-Past, Present, and Future: The Albert Lasker Basic Medical Research Award, *JAMA 328*, 1291-1292.

[6] Springer, T. A., and Dustin, M. L. (2012) Integrin inside-out signaling and the immunological synapse, *Current opinion in cell biology 24*, 107-115.

[7] Luo, B.-H., Carman, C. V., and Springer, T. A. (2007) Structural basis of integrin regulation and signaling, *Annu. Rev. Imm. 25*, 619-647.

[8] Calderwood, D. A., Campbell, I. D., and Critchley, D. R. (2013) Talins and kindlins: partners in integrin-mediated adhesion, *Nature reviews. Molecular cell biology 14*, 503-517.

[9] Hu, P., and Luo, B. H. (2013) Integrin bi-directional signaling across the plasma membrane, *J Cell Physiol 228*, 306-312.

[10] Xiong, J.-P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D. L., Joachimiak, A., Goodman, S. L., and Arnaout, M. A. (2001) Crystal structure of the extracellular segment of integrin $\alpha_V\beta_3$, *Science 294*, 339-345.

[11] Xiong, J. P., Mahalingham, B., Alonso, J. L., Borrelli, L. A., Rui, X., Anand, S., Hyman, B. T., Rysiok, T., Muller-Pompalla, D., Goodman, S. L., and Arnaout, M. A. (2009) Crystal structure of the complete integrin $\alpha_V\beta_3$ ectodomain plus an α/β transmembrane fragment, *J Cell Biol 186*, 589-600.

[12] Dong, X., Mi, L.-Z., Zhu, J., Wang, W., Hu, P., Luo, B. H., and Springer, T. A. (2012) $\alpha_V\beta_3$ Integrin Crystal Structures and their Functional Implications, *Biochemistry 51*, 8814-8828.

[13] Xiao, T., Takagi, J., Wang, J.-h., Coller, B. S., and Springer, T. A. (2004) Structural basis for allostery in integrins and binding of fibrinogen-mimetic therapeutics, *Nature 432*, 59-67.

[14] Zhu, J., Luo, B. H., Xiao, T., Zhang, C., Nishida, N., and Springer, T. A. (2008) Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces, *Mol. Cell 32*, 849-861.

[15] Zhu, J., Zhu, J., Negri, A., Provasi, D., Filizola, M., Coller, B. S., and Springer, T. A. (2010) The closed headpiece of integrin $\alpha_{IIb}\beta_3$ and its complex with an $\alpha_{IIb}\beta_3$-specific antagonist that does not induce opening, *Blood 116*, 5050-5059.

[16] Zhu, J., Zhu, J., and Springer, T. A. (2013) Complete integrin headpiece opening in eight steps, *J Cell Biol 201*, 1053-1068.

[17] Thinn, A. M. M., Wang, Z., Zhou, D., Zhao, Y., Curtis, B. R., and Zhu, J. (2018) Autonomous conformational regulation of β_3 integrin and the conformation-dependent property of HPA-1a alloantibodies, *Proceedings of the National Academy of Sciences of the United States of America 115*, E9105-E9114.
[18] Zhou, D., Thinn, A. M. M., Zhao, Y., Wang, Z., and Zhu, J. (2018) Structure of an extended β_3 integrin, *Blood 132*, 962-972.
[19] Nesic, D., Zhang, Y., Spasic, A., Li, J., Provasi, D., Filizola, M., Walz, T., and Coller, B. S. (2020) Cryo-Electron Microscopy Structure of the $\alpha_{IIb}\beta_3$-Abciximab Complex, *Arteriosclerosis, thrombosis, and vascular biology 40*, 624-637.
[20] Wang, Z., Huo, T., Wu, H., Moussa, Z., Sen, M., and Dalton, V. (2023) Full-length $\alpha_{IIb}\beta_3$ CryoEM structure reveals intact integrin initiate-activation intrinsic architecture, *Res Sq*.
[21] Choi, W. S., Rice, W. J., Stokes, D. L., and Coller, B. S. (2013) Three-dimensional reconstruction of intact human integrin $\alpha_{IIb}\beta_3$: new implications for activation-dependent ligand binding, *Blood 122*, 4165-4171.
[22] Xia, W., and Springer, T. A. (2014) Metal ion and ligand binding of integrin $\alpha_5\beta_1$, *Proceedings of the National Academy of Sciences of the United States of America 111*, 17863-17868.
[23] Schumacher, S., Dedden, D., Nunez, R. V., Matoba, K., Takagi, J., Biertumpfel, C., and Mizuno, N. (2021) Structural insights into integrin $\alpha_5\beta_1$ opening by fibronectin ligand, *Sci Adv 7*.
[24] Nagae, M., Re, S., Mihara, E., Nogi, T., Sugita, Y., and Takagi, J. (2012) Crystal structure of $\alpha_5\beta_1$ integrin ectodomain: Atomic details of the fibronectin receptor, *J Cell Biol 197*, 131-140.
[25] Arimori, T., Miyazaki, N., Mihara, E., Takizawa, M., Taniguchi, Y., Cabanas, C., Sekiguchi, K., and Takagi, J. (2021) Structural mechanism of laminin recognition by integrin, *Nat Commun 12*, 4012.
[26] Sen, M., Yuki, K., and Springer, T. A. (2013) An internal ligand-bound, metastable state of a leukocyte integrin, $\alpha_X\beta_2$, *J Cell Biol 203*, 629-642.
[27] Xie, C., Zhu, J., Chen, X., Mi, L., Nishida, N., and Springer, T. A. (2010) Structure of an integrin with an α I domain, complement receptor type 4, *Embo J 29*, 666-679.
[28] Sen, M., and Springer, T. A. (2016) Leukocyte integrin $\alpha_L\beta_2$ headpiece structures: The alphaI domain, the pocket for the internal ligand, and concerted movements of its loops, *Proceedings of the National Academy of Sciences of the United States of America 113*, 2940-2945.
[29] Goldsmith, J. A., DiVenere, A. M., Maynard, J. A., and McLellan, J. S. (2022) Structural basis for non-canonical integrin engagement by Bordetella adenylate cyclase toxin, *Cell reports 40*, 111196.
[30] Jensen, R. K., Pedersen, H., Lorentzen, J., Laursen, N. S., Vorup-Jensen, T., and Andersen, G. R. (2022) Structural insights into the function-modulating effects of nanobody binding to the integrin receptor $\alpha_M\beta_2$, *J Biol Chem 298*, 102168.
[31] Dong, X., Zhao, B., Iacob, R. E., Zhu, J., Koksal, A. C., Lu, C., Engen, J. R., and Springer, T. A. (2017) Force interacts with macromolecular structure in activation of TGF-beta, *Nature 542*, 55-59.

[32] Dong, X., Hudson, N. E., Lu, C., and Springer, T. A. (2014) Structural determinants of integrin β-subunit specificity for latent TGF-beta, *Nat Struct Mol Biol 21*, 1091-1096.
[33] Kotecha, A., Wang, Q., Dong, X., Ilca, S. L., Ondiviela, M., Zihe, R., Seago, J., Charleston, B., Fry, E. E., Abrescia, N. G. A., Springer, T. A., Huiskonen, J. T., and Stuart, D. I. (2017) Rules of engagement between $\alpha_V\beta_6$ integrin and foot-and-mouth disease virus, *Nat Commun 8*, 15408.
[34] Cormier, A., Campbell, M. G., Ito, S., Wu, S., Lou, J., Marks, J., Baron, J. L., Nishimura, S. L., and Cheng, Y. (2018) Cryo-EM structure of the $\alpha_V\beta_8$ integrin reveals a mechanism for stabilizing integrin extension, *Nat Struct Mol Biol*.
[35] Wang, J., Su, Y., Iacob, R. E., Engen, J. R., and Springer, T. A. (2019) General structural features that regulate integrin affinity revealed by atypical $\alpha_V\beta_8$, *Nat Commun 10*, 5481.
[36] Campbell, M. G., Cormier, A., Ito, S., Seed, R. I., Bondesson, A. J., Lou, J., Marks, J. D., Baron, J. L., Cheng, Y., and Nishimura, S. L. (2020) Cryo-EM Reveals Integrin-Mediated TGF-beta Activation without Release from Latent TGF-beta, *Cell 180*, 490-501 e416.
[37] Duan, Z., Lin, X., Wang, L., Zhen, Q., Jiang, Y., Chen, C., Yang, J., Lee, C. H., Qin, Y., Li, Y., Zhao, B., Wang, J., and Zhang, Z. (2022) Specificity of TGF-β1 signal designated by LRRC33 and integrin $\alpha_V\beta_8$, *Nat Commun 13*, 4988.
[38] Yu, Y., Zhu, J., Mi, L. Z., Walz, T., Sun, H., Chen, J.-F., and Springer, T. A. (2012) Structural specializations of $\alpha_4\beta_7$, an integrin that mediates rolling adhesion, *J Cell Biol 196*, 131-146.
[39] Yang, J., Ma, Y. Q., Page, R. C., Misra, S., Plow, E. F., and Qin, J. (2009) Structure of an integrin $\alpha_{IIb}\beta_3$ transmembrane-cytoplasmic heterocomplex provides insight into integrin activation, *Proceedings of the National Academy of Sciences of the United States of America 106*, 17729-17734.
[40] Lau, T. L., Kim, C., Ginsberg, M. H., and Ulmer, T. S. (2009) The structure of the integrin $\alpha_{IIb}\beta_3$ transmembrane complex explains integrin transmembrane signalling, *Embo J 9*, 1351-1361.
[41] Zhu, J., Luo, B. H., Barth, P., Schonbrun, J., Baker, D., and Springer, T. A. (2009) The structure of a receptor with two associating transmembrane domains on the cell surface: integrin $\alpha_{IIb}\beta_3$, *Mol. Cell 34*, 234-249.
[42] Vinogradova, O., Velyvis, A., Velyviene, A., Hu, B., Haas, T. A., Plow, E. F., and Qin, J. (2002) A structural mechanism of integrin $\alpha_{IIb}\beta_3$ "inside-out" activation as regulated by its cytoplasmic face, *Cell 110*, 587-597.
[43] Weljie, A. M., Hwang, P. M., and Vogel, H. J. (2002) Solution structures of the cytoplasmic tail complex from platelet α_{IIb}- and β_3-subunits, *Proc. Natl. Acad. Sci. USA 99*, 5878-5883.
[44] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, 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., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold, *Nature 596*, 583-589.

[45] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., and Hassabis, D. (2022) Protein complex prediction with AlphaFold-Multimer, *bioRxiv*, 2021.2010.2004.463034.
[46] Laukaitis, C. M., Webb, D. J., Donais, K., and Horwitz, A. F. (2001) Differential dynamics of α_5 integrin, paxillin, and alpha-actinin during formation and disassembly of adhesions in migrating cells, *J Cell Biol 153*, 1427-1440.
[47] Wang, Z., and Zhu, J. (2021) Measurement of Integrin Activation and Conformational Changes on the Cell Surface by Soluble Ligand and Antibody Binding Assays, *Methods Mol Biol 2217*, 3-15.
[48] Smagghe, B., Huang, P., Ban, Y.-E., Baker, D., and Springer, T. A. (2010) Modulation of integrin activation by an entropic spring in the β-knee, *J Biol Chem 285*, 32954-32966.
[49] Askari, J. A., Tynan, C. J., Webb, S. E., Martin-Fernandez, M. L., Ballestrem, C., and Humphries, M. J. (2010) Focal adhesions are sites of integrin extension, *J Cell Biol 188*, 891-903.
[50] Hughes, P. E., Diaz-Gonzalez, F., Leong, L., Wu, C., McDonald, J. A., Shattil, S. J., and Ginsberg, M. H. (1996) Breaking the integrin hinge, *J. Biol. Chem. 271*, 6571-6574.
[51] Yin, R., Feng, B. Y., Varshney, A., and Pierce, B. G. (2022) Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants, *Protein Sci 31*, e4379.
[52] Ibrahim, T., Khandare, V., Mirkin, F. G., Tumtas, Y., Bubeck, D., and Bozkurt, T. O. (2023) AlphaFold2-multimer guided high-accuracy prediction of typical and atypical ATG8-binding motifs, *PLoS Biol 21*, e3001962.