How do extracellular membrane-associated proteins stand erect and signal?

Swetha Lankipalli[1,2] and Udupi A. Ramagopal[1 *]

[1]Division of Biological Sciences, Poornaprajna Institute of Scientific Research,

#4, 16th Cross, Sadashivnagar, Bangalore 560080, India

[2]Manipal Academy of Higher Education, Manipal, Karnataka, India-576104

*To whom correspondence may be addressed:

Udupi A. Ramagopal, Division of Biological Sciences, Poornaprajna Institute of Scientific Research, #4, 16th Cross, Sadashivanagar, Bangalore 560080, India;

Ph.: +91 9900810182; E-mail: ramagopal.udupi@gmail.com

Abbreviations:

IgV- Immunoglobulin V-type domain; IgC- Immunoglobulin C-type domain; EJ-linker- Extracellular juxtamembrane linker; TM domain- transmembrane domain

Keywords: Ectodomain; upright orientation, membrane tethering, protein clustering; one-dimensional array; nanoclusters, spurious signaling; non-signaling resting state;

Abstract:

Even after decades of research, a comprehensive mechanism that elucidates the underpinnings of signalling through the cell membrane is still elusive. Here, we address a simple question "how do the ectodomain of a membrane-associated protein consisting of multiple domains and connected by flexible linkers stand 'upright' on the membrane?". Our analysis based on large amount of available functional and structural data, looking for a pattern of association of these molecules in the crystal structures and with the concept that 'random things seldom repeat' lead to a surprisingly interesting and consistent observation that the weak *cis* interaction mediated ordered array of constitutively expressed signalling molecules not only support their 'upright' orientation but also bury their ligand-binding surface to avoid spurious signaling in the non-signalling resting state. With CD4, pMHCII, CD2 and TNFR1 as examples, we show that the observed association of molecules also correlate well with their functional role. Further, our model reconciles the long-standing controversies related to these molecules and appear to be generic enough to be applied to other signalling molecules

The cell membrane, the barricade of life with organized lateral heterogeneity, not only separates it's inside from the external world but is designed to accurately recognize external cues and optimally respond to it. Its organization is of interest for over a century. Membrane-associated proteins constitute 50% of the total cell-surface volume and play a critical role in the detection and integration of extracellular mechanical, physical and chemical stimuli and hence play a vital role in cell-cell recognition and communications[1]. The ectodomains of these proteins vary in size, architecture and most of them consist of multiple domains connected by linkers (Fig. 1). Further, they are membrane-tethered via either an extended extracellular juxtamembrane linker (hereafter called as EJ-linker) which is purported to be flexible or anchored through glycosylphosphatidylinositol (GPI) moieties. For these membrane proteins to produce reproducible organization and function, they are expected to be robust and at the same time pliable to respond to weak cues and transform them into reproducible signals across the membrane. For example, in the case of adhesion proteins like E-cadherin, the ~3nm transmembrane α-helix anchors an extracellular stretch of ~600 amino acids organized into five domains measuring up to ~25nm in length. Similarly, in the case of CD45, the average dynamic height of its ectodomain was measured to be ~25nm while its maximum extended height was 40nm[2]. It is intriguing to know, how these molecules connected by multiple interdomain linkers including the EJ-linker that is presumed to be flexible, support the appropriate orientation of considerably large cell surface molecules and perform an accurate function. Further, flexible linkers can increase the configurational entropy of a protein on the cell membrane, and hence it is likely that such molecules on the crowded membrane are prone to tangle with each other forming aggregates that can never respond to a specific signal. In most of these cases, the extracellular domains distal to the membrane are the

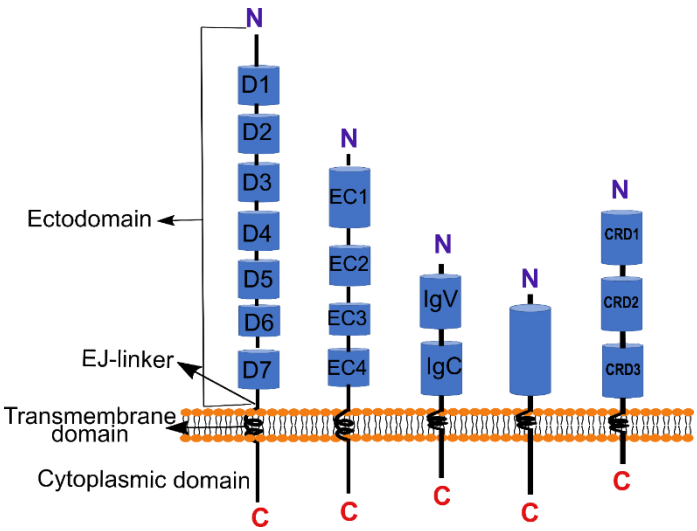

Fig.1 Representative structures of diverse type-I membrane associated proteins. The naming convention used are also diverse- they are named as D_1 to D_n (Domain 1 to Domain 'n'), EC1 to EC_n (ectodomain 1 to 'n'), CRD1 (Cysteine rich domain 1 to 'n'), where is 'n' is an integer and also sometimes they are named based on their fold such as IgV, IgC_1, IgC_2, etc. The multidomain proteins with a stalk like EJ-linker (extracellular juxtamembrane linker) are tethered to the membrane by a single pass transmembrane helix. The N terminus is at the membrane distal end on the extracellular side while the C terminus is on the cytoplasmic side of each protein.

ones that interact with their cognate partners on the opposing cell. Hence, these molecules are expected to stand 'upright' on the cell surface and position the membrane-distal domain(s) in an optimal orientation on the cell surface to ensure efficient function. Further, the elegant 'kinetic segregation' model that explains the segregation of molecules in the cell-cell contact during T-cell signaling requires that the distance spanned by each molecule from the membrane is proportional to the effective length of the molecule [3]. Here, whether these molecules are single or not, the requirement appears to be that they maintain an 'upright' orientation on the cell surface. In fact, the measurement using sophisticated microscopy techniques suggests that the distances spanned by different isoforms of CD45 from the membrane are proportional to their ectodomain size [4], meaning, these molecules stand erect on the cell surface. Although, current techniques such as super-resolution microscopy [5]and single-molecule tracking [6, 7]and allied techniques have revolutionized the way we look at the biological systems[8], in particular, the molecular organization and their dynamics on the cell surface, unfortunately, even the improved resolution does not provide a comprehensive mechanism controlling their association and function[9]. Even a fundamental question, how cell-surface proteins position themselves 'upright' on the membrane prior to their interaction with a cognate partner on the opposing cell is a mystery.

Nano, meso, and microscopic organization of membrane-associated proteins have been revealed by super-resolution microscopy and now can be quantitatively investigated[10]. It is also known that on the cell membrane, these proteins organize/reorganize depending on the functional context such as polarity, adhesion, immune, and synaptic functions [11-13]. However, what is not clear is; how do they achieve this compartmentalization and how the clustering state imparts the ability to perform the accurate and reproducible function? Models such as 'lipid rafts' [14]and 'fences and pickets' have been proposed[15] to explain the clustering phenomenon. Unfortunately, the diversity of these clusters can't be explained satisfactorily by both these models [13, 16-22]. For example, only certain membrane proteins enrich in the raft and others do not[23-25]. Similarly, the control of large sequence and architectural diversity of these membrane-associated proteins, by the intracellular cytoskeleton network alone is difficult to perceive. For example, most clusters are homotypic [26, 27] which requires a mechanism that can segregate proteins based on their identity. Recent reports propose that the ability of these proteins to cluster with each other (homo or hetero depending on the cluster) is inherent in them and is mediated by weak protein-protein interactions[16, 28, 29]. However, the physicochemical signatures governing such an association, their

dissociation, and reorganization remain unanswered[12]. As the techniques to look at these molecules are improving, we are observing more diversity and the complexity of these clusters. For example, cadherin clusters once thought to be initiated by trans (cell to cell) interactions, followed by diffusion trap and formation of trans/cis microclusters, now appears to be different in that larger *trans-cis* microclusters co-exist with lateral (*cis*) nano-clusters [30]. Similarly, in the case of endothelial receptor CD36, its interaction with multivalent ligands, such as TSP-1 (thrombospondin-1) was shown to result in clustering and consequent downstream signaling [31]. However, recent super-resolution microscopy studies reveal that CD36 nanocluster that can not only support TSP-1 binding but also enriched with downstream effector Fyn exists on the cell and only the denser and larger cluster activates Fyn[32]. A similar observation on TCR-CD3 complexes, where only denser clusters were found to be phosphorylated and associated with downstream signaling was reported[33]. Similarly, in the case of β2-integrin LFA-1(Lymphocyte function-associated antigen-1), three different avidity patterns have been observed- randomly distributed inactive molecules, ligand-independent ordered proactive nanoclusters and the macroclusters triggered by ligand interaction[34].

These observations lead to a question as to whether these less dense nanoclusters represent a basal organization and pre-reserved pool that reorganize into a denser and larger supramolecular functional unit that underpins signaling? Again, how these different organizations i.e., nanoclusters probably with 'off'-state and larger clusters with 'on' state are achieved? All these jargons, (1) how a cell surface receptor stand erect? (2) how do they organize into clusters of different sizes and architecture? (3) how these clusters perform reproducible and accurate function avoiding spurious signaling? we believe, are probably interconnected and such diversity and accuracy can be achieved only through highly organized, but weak and tuneable protein-protein interactions. If so, how do we see them when they are in action? Probably, we need to look at them at much higher resolution and speed to obtain their spatiotemporal information when they are on live cells. Although electron microscopy can provide higher resolution, imaging of live cells is difficult or almost impossible. Similarly, X-ray crystallography can provide the atomic resolution picture of these molecules, but it is difficult to work with a complete protein containing extracellular, membrane-integral and intracellular parts of these proteins. When a protein has several domains, due to difficulty in crystallization, structures of single or a few of these domains are studied. However, in some fortunate cases, structures of complete ectodomains have been resolved to atomic resolution. Since the crystal formation requires epitope(s) of association[35] on the surface of each

molecule that can support the formation of 3D lattice, in the absence of any preferred interaction, these molecules are expected to have an infinite mode of self-association. In contrast, if the ectodomains of these proteins have physiologically relevant association preference, although weak-to-be-detected in solution, we asked the question, whether any such preferred association can be observed in the crystals from the pool of available crystal structures of these receptors. Supporting this argument, super-resolution microscopy experiments detected the densities of E-cadherin clusters on the cell surface that is consistent with the densities observed in crystals[30]. Further, in many structures of ectodomains of homophilic immunoglobulin superfamily cell adhesion molecules, the molecules in the crystal arrange in a zipper-like array representing their functional role in the cell-cell junction[36]. Similarly, crystal structures of CTLA-4 in complex with B7-1 and with B7-2, two functionally equivalent molecules revealed identical 1D array consisting of alternative dimers of B7s and CTLA-4[37, 38], representing the possible arrangements of these molecules during the immunological synapse formation. In all these cases, the *trans* (between the molecules from opposing cells) interaction are strong enough to be detected in solution. Hence, it can be inferred that the relatively strong receptor: ligand (homophilic or heterophilic) interactions can survive and in most cases, override the crystal packing forces, representing the expected association *in vivo*.

Since our question was about the 'upright' orientation of these molecules on the cell, we looked for possible *cis* (between the molecules residing on the same cell) interactions and their effect on 'upright' orientation. Firstly, we analyzed several structures (see, Methodology in Supplementary information) of these cell surface receptors to understand whether the ectodomains of these molecules connected by interdomain linkers maintain similar conformation and if it is true, the possible reason for such an inter-domain stabilization? It was interesting to observe that in most cases these ectodomains are stable with minimal inter-domain flexibility and the EJ-linker is also probably not very flexible. Finally, we analyzed the crystal packing to understand their association preference in their apo form and compared their association in the presence of a cognate partner. The proteins described here are those that fit into the following criteria (1) ectodomains that are known to be monomers in solution (or weak-to-be-detectable in solution) (2) multiple crystal structures have been determined for the complete ectodomain of the same molecule either from the same organism or from different organisms. This criterion was set to make sure that the observed association in the crystal is not an accidental observation and those that repeat might have some significance (random

things seldom repeat) (3) to eliminate the effect of cognate binding partners, only apo structures were considered first and later it was compared with the complex structure containing cognate partner and (4) all those cases, where the observed association and oligomerization represented the physiologically relevant 'upright' orientation of these molecules on the cell surface. We found that in many cases, these ectodomains in the crystal associates symmetrically as if they are coming from the membrane of the same cell and appear to have oligomerization preference that is too weak-to-be-detectable in solution. Further, we mostly restricted ourselves to those molecules involved in immune function. Based on the analysis, we propose that the cryptic *cis*-oligomerization preference of rigid ectodomains together with conformational entropy constrained by membrane tethering helps them to stand 'upright' on the cell surface. More interestingly, these weak oligomers hide their ligand-binding surface to avoid spurious signaling and can further associate into higher-ordered arrays. Importantly, this analysis reconciles some of the contradicting observations pertinent to T-cell signaling mediated by MHCII-TCR-CD4 and other molecules such as CD2 and TNFs involved in immune signaling.

Results and discussion:

To understand the interdomain flexibility of multi-domain cell-surface receptors, we analyzed the crystal structures of several of these molecules present on immune cells. The structure of extracellular receptors containing two Ig domains (IgC and IgV), such as CD2, NTB-A, PD-L1, PD-L2, B7-1and so on reveal that the hydrophobic interactions between the two domains maintain a stable relative orientation [39-42]. Similarly, structures containing three extracellular domains (D1-D3) of Nectin1 (two structures), the first structure with one molecule in ASU (PDB-entry:3U83)[43] and the other with four molecules in ASU(PDB-entry: 4FMF)[44] also appear to be rigid. The superposition of complete ectodomain of Nectin-1 in two different crystal forms results in C_α RMSD of less than 2.0Å. It was also interesting to note that the structure of D1-D3 domains of Nectin-3 (PDB-entry: 4FOM) is very similar to Nectin-1. Similarly, in the case of larger molecules such as CD4, cadherins, ephrins, CD45, and so on, it has been shown that the structure of part of the protein containing two or three domains superposes well with the complete ectodomains of these proteins [4, 45-47]. Also, in most of these proteins, solution studies confirm that the structures observed in the crystals are highly similar to their conformation in solution [4, 48].

In the case of CD22 with seven Ig domains, although the structure of complete ectodomain is not available, the apo and sialic acid bound structures having D1-D3 domains are very similar.

In line with this observation, using small-angle scattering studies and negative stain electron microscopy; it has been shown that full-length CD22 ectodomain (D1-D7) adopts similar rod-like conformation with limited interdomain flexibility [49]. The above analysis suggests that these receptors containing multiple domains connected by linkers are rigid or posses' minimal inter-domain flexibility.

The observed interdomain stability of the ectodomain can't be a complete answer for their ability to stand 'upright' as they are tethered to a jelly-like membrane via an EJ-linker which is presumed to be flexible. Since the EJ-linker is located in the aqueous environment just outside the cell surface, if flexible, it is expected to contain more of hydrophilic residues. Our analysis suggests that it is not always true. For example, in some of the key molecules involved in immune function, the sequence of the EJ-linker contains close to 50% hydrophobic residues and sometimes rich in conformation constraining residues like proline, together with aromatic residues (Table-1 for more information). Such sequences are expected to be rigid and being substantially hydrophobic, their existence in the aqueous extracellular region requires burial of these hydrophobic residues. This raises the question of whether they are flexible as generally presumed or such linkers from adjacent molecules on the cell membrane can associate with each other through their hydrophobic residues, stabilizing well-defined oligomers and/or have the ability to get involved in energetically similar but different oligomerization states? Unfortunately, EJ-linker is the most neglected part of all the studies directed towards structural characterization, oligomerization and clustering. Most such studies involving cell-surface molecules, manipulations are done on the ectodomain (excluding the EJ-linker) and/or on the intracellular domain.

The possible role of EJ-linker can be observed in, for example, B7-1 and B7-2 which are functionally equivalent molecules. Upon interaction with CTLA-4, both B7-1 and B7-2 form a similar kind of 1D zipper-like organization consisting of an alternative arrangement of B7 dimer and CTLA-4 dimer mimicking the situation expected for these molecules in the immunological synapse[37, 38]. However, while B7-1 is a known dimer, B7-2 is shown to be a monomer in solution [39, 50]. Interestingly, the human B7-2 linker region ($_{226}$RLLSSPFSIELEDPQPPPDHIP$_{247}$) contains 54% hydrophobic residues of which 29% are prolines. Proline is an anomalous residue (average occurrence of around 6%[51]), and proline-rich sequences such as collagen, keratin and so on are known to associate in parallel.

Table-1 Proline rich sequences of extracellular juxtamembrane linker(EJ-linker) of selected type-I membrane associated proteins and overall percentage of hydrophobic residues

Uniprot ID	Protein name	No. of amino acids (sequence position)	EJ-linker-sequence	% of prolines in the linker sequence	total % of hydrophobic aa in the linker sequence
P42081	hB7-2	21 (227-247)	LLSSPFSIELEDPQPPPDHIP	28.50	57
P01730	hCD4	23 (374-396)	DSGQVLLESNIKVLPTWSTPVQP	13	52.10
Q9NZQ7	hPD-L1	13 (226-238)	IPELPLAHPPNER	30	61.50
Q9BQ51	hPD-L2	17 (204-220)	LASIDLQSQMEPRTHPT	11.70	41.17
P19438	hTNR1A(hTNFR1)	15 (197-211)	PQIENVKGTEDSGTT	6	33.30
P20333	hTNR1B (hTNR2)	56 (202-257)	STSPTRSMAPGAVHLPQPVSTRSQHTQPTPEPSTAPSTSFLLPMGPSPPAEGSTGD	21.40	50
P13688	hCEACAM1	15 (414-428)	VNYNALPQENGLSPG	13.30	53.30
P19022	N-cadherin	11 (715-725)	IVGAGLGTGAI	0	90.90
P12830	E-cadherin	12 (698-710)	QPVEAGLQIPAI	16.66	75
Q9NQS3	hNectin-3	51 (355-405)	ISDPPTTTTLQPTIQWHPSTADIEDLATEPKKLPFPLSTLATIKDDTIATI	13.70	47
Q15223	hNectin-1	22 (335-356)	EFPYTPSPPEHGRRAGPVPTAI	27.20	59
Q92692	hNectin-2	15 (346-360)	VRETPNTAGAGATGG	6	60
Q13740	hCD166	27 (502-528)	AISIPEHDEADEISDENREKVNDQAKL	3.70	33.33
Q15116	hPD-1	25 (146-170)	ERRAEVPTAHPSPSPRPAGQFQTLV	20.00	52.00
Q495A1	hTIGIT	18 (125-142)	EVLESSVAEHGARFQIPL	5.50	55.55
P47741	OX-40(CD134)	48 (168-214)	DRDPPATQPQETQGPPARPITVQPTEAWPRTSQGPSTRPVEVPGGRAV	22.90	52
P36941	TNR3(TNFR3)	16 (212-227)	NPLEPLPPEMSGTMLM	25	68.75
P10747	hCD28	16 (138-153)	KHLCPSPLFPGPSKPF	31.25	62.50

Unfortunately, constructs of B7-2 used in the experiments to understand the B7-2 oligomerization did not have this proline-rich sequence. Even in the FRET-based studies[52], aimed to reveal the oligomerization of B7-2 on the cell surface, the EJ-linker sequence has been modified to a sequence rich in -Gly-Ala-, thus leaving no clue on the effect of original EJ-linker on B7-2 oligomerization on the cell membrane. Similarly, the EJ-linker of TNFR2 contains 56 residues with 50% hydrophobic residues of which 12 are prolines (~22%). A recent study that provides a model for TNF superfamily signaling based on the structures of different TNF receptors does not consider the importance of the probable structure of such a long linker. However, the importance of the EJ-linker (also termed as stalk) in the case of TNFR-1 and TNFR-2 proteins has been demonstrated in a study where the exchange of the stalk regions led to the change in their responsiveness to the membrane-bound ligand[53]. These observations

suggest that the inclusion of the EJ-linker region is important in studies involving, oligomerization and signaling.

Apart from the importance of EJ-linker and limited inter-domain flexibility of ectodomains, the preferential association of ectodomains appears to be important for 'upright' orientation and robust signaling. In the following section, we present few cases, where ectodomain of these receptors are known to be monomers in solution (or more accurately, possess undetectable oligomerization potential), but consistently exist as multimers in the crystal, maintaining similar physiologically relevant association, that can provide support for 'upright' orientation and their function on the cell surface. Other than the criteria for the selection discussed before, we restricted ourselves to immune molecules as we should be able to connect the known function to observed association preference.

CD4-oligomerization: The CD4 is a key co-receptor on the T-cells that interact with peptide-loaded major histocompatibility complex-II (pMHCII) on the antigen-presenting cells (APCs) and contributes to signal transduction through its cytoplasmic tail by interacting with lymphocyte kinase (Lck) [54]. The ectodomain of hCD4 consists of four Ig domains (named D1 to D4), where the membrane distal domain (D1) is shown to interact with pMHCII[55]. While no biophysical experiments detected the oligomerization of CD4 ectodomain in solution[56], the dimeric association of CD4, involving the membrane-proximal D4 domain has been observed in three different isoforms of the complete ectodomain of CD4 crystals **(Fig. 2A)**[45]. In all the

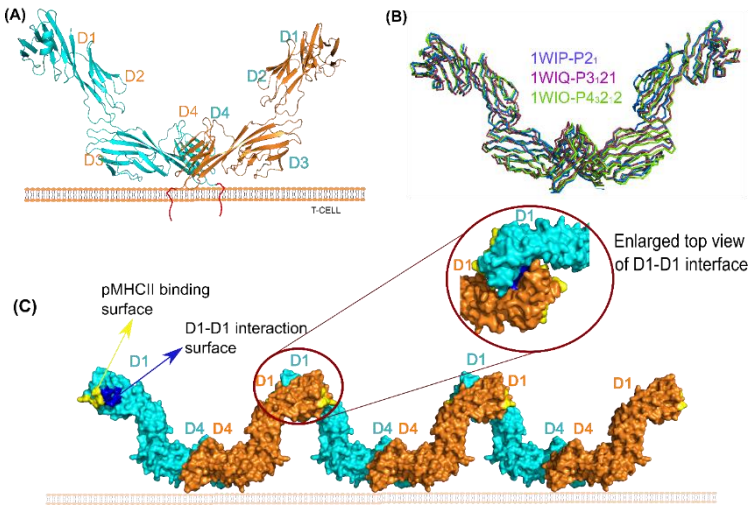

Fig.2 Structural details of ectodomain of CD4 (D1-D4) (A) Cartoon representation of dimeric CD4, the monomers coloured in cyan and orange associate via their membrane proximal D4 domains and the D4 domains interact with the membrane placing the D1-D3 domains distal to the cell surface. **(B)** Superposition of three different crystal structures of CD4 (ribbons) (PDB entries: 1WIP, 1WIQ, 1WIO) highlighting the exactly similar D4-D4 mediated dimeric association of two protomers. **(C)** packing analysis of the three different structures of CD4 (surface representation) reveals a formation of identical 1D array of CD4 homodimers due to D4-D4 interaction and D1-D1 interaction conserved in all the three structures. The D1-D1 interface (shaded in dark blue), partially overlaps with pMHCII binding site (shaded in yellow) precluding its binding to MHCII in non-signalling resting state and hence appear to avoid spurious signalling.

three structures (PDB-entry: 1WIQ, 1WIP, 1WIO), the dimer is very similar with a maximum C_α RMSD of ~1.2Å **(Fig. 2B)** and the description of the structure in the author's own words is very interesting "*the D4-D4 associated dimers are butterfly-shaped in profile, with D1-D3 as 'wings', D4 pair as 'torso' and C termini as 'legs', as if perching on the hypothetical membrane surface perpendicular to the dyad axis*", clearly depicts the dimeric ''upright'' orientation of CD4 on the cell membrane. Although analytical ultracentrifugation and light scattering studies did not support CD4 oligomerization, they were able to detect weak CD4 dimerization mediated by the D4-D4 domain by performing crosslinking studies[45]. Consistent with this observation, several mutation studies also confirmed the role of the D4 domain in CD4 dimerization[57].

Contrary to this observation, there is a considerable argument in favor of CD4 dimerization through the D1 domain[58] and independent experiments involving mutations on either D1 or D4 domains resulted in impaired dimerization[57, 59, 60]. Assuming, the interaction at both ends of the molecule (that is, D1-D1 and D4-D4) is possible, then the resulting structure should be a non-covalent polymer. Consistent with this observation, at higher concentrations (5mg/ml) analytical ultracentrifugation studies revealed a non-ideal behavior of CD4 in solution[45]. Further, packing of CD4 molecules in the crystal suggests that in two of the three reported structures of CD4 (D1-D4, 1WIO, 1WIQ), the membrane distal N-terminal domain (D1) from adjacent D4-D4 dimers interacts with each other maintaining similar interactions and forming continuous 1D array **(Fig. 2C)** of CD4 molecules. It should be noted that the relative orientation of D1-D1 dimer is retained in the third structure too (1WIP): however, probably due to crystal packing forces, the D1 domains are slightly pulled apart from each other. Further, the D1-D1 domain interaction observed in the above structures is also conserved in all the six crystal structures of CD4 containing only D1-D2 domains, **(Fig. S1)**, which is in line with mutation studies mentioned above and suggest the probable physiological role of D1-D1 interaction in CD4 function. In the crystals of complete ectodomain (PDB-entry: 1WIQ, 1WIP, 1WIO), the pairing of D4 domains together with the pairing of D1 domains form a zigzagged 1D array of CD4 molecules (similar to a portable zigzagged display panel). Further, the structures of all the apo CD4 monomers, truncated structures, and the one that is observed in the MHCII:TCR:CD4 ternary complex[61] superpose well with each other, suggesting the elongated CD4 ectodomain is rigid. Hence, it appears that the rigid structure of CD4 ectodomain together with zigzagged arrangement supports the erect orientation of CD4 molecules on the cell surface. The most interesting observation that emerged out of this analysis is that the membrane distal D1-D1

interface that appears to support the 'upright' orientation of CD4 molecules on the cell surface is also used for pMHCII binding. Hence, the formation of 1D array and consequent closing of pMHCII binding site on D1, not only helps its 'upright' orientation on the cell surface but also appear to avoid spurious interaction with the pMHCII molecules present on the opposing APCs during the screening of peptide by TCR.

New model for pMHC-TCR-CD4 mediated signalling: The pMHCII-TCR-CD4 complex is central to immunological synapse formation and acts as a first signal in the process of T-cell activation. Although signaling through pMHCII with TCR and CD4 is important, the mechanism by which such an accurate process is achieved is still enigmatic. Recognition of pMHCII present on the APCs by TCR on T-helper cells is known to trigger the interaction of co-receptor CD4 (present on T-helper cells) with pMHCII. pMHCII is a heterodimer consisting of two chains named α and β chains with the peptide-binding groove crafted from residues involving both the chains. The first report describing the two crystal structures of pMHCII (HLA-DR1 with antigenic peptide) from the human B-cell membrane describes a dimer of heterodimers (later dubbed as super-dimers) in both the crystal forms(PDB entries: 1DLH, 1AQD)[62]. The parallel association of two heterodimers brings all the four C-termini in one direction with both the peptide binding groves facing in the opposite direction that supports the presentation of bound peptides in an orientation appropriate for recognition by TCRs on juxtaposed T-helper cells **(Fig. 3A).** Here, facing of the C-terminal end of all the four molecules in one direction and consequent tethering to the membrane at four different points in proximity appears to stabilize the ''upright'' orientation of MHCII. Although, tethering of two C-termini of αβ heterodimer alone is probably sufficient, tethering at two more points might help in the optimization of peptide presentation, during the screening process. The authors also propose that the superdimer might bind to CD4 coreceptor with increased affinity and hence initiate TCR mediated signaling, although, this proposition is probably a provocative statement to derive solely based on two initial structures. Subsequent determination of several crystal structures of pMHC variants (HLA-DR1; PDB-entries: 1KG0, 4I5B, 1AQD) including HLA-DR1 in complex with superantigen *Staphylococcus aureus* enterotoxin B (PDB entry: 1SEB) and HLA-DR3 (PDB entry: 1A6A) revealed a similar association of dimer of αβ heterodimers. This super-dimer is expected to be very weak as the interface is formed by discontinuous stretches of polar and charged residues with no considerable hydrophobic interactions [62]. This might probably be the reason for the lack of any evidence of such superdimer in solution as measured by hydrodynamic radii and other techniques[63, 64]. However, chemical

crosslinking studies and immunoprecipitation of pMHCII dimers (I-E^k) from B-cell[65, 66] also supported the "superdimer mediated T-cell activation" hypothesis. Later, the structure determination of complete ectodomain of CD4 and the observation that the CD4 also exists as a dimer (discussed in the previous section on CD4), triggered interest in investigations on the possibility of dimeric CD4 and pMHCII superdimer mediated T-cell signaling[60].

Interestingly, the superdimer which was observed in most of the pMHCII variant crystal structures was not seen in any of the crystal structures of pMHCII in complex with TCR (PDB-

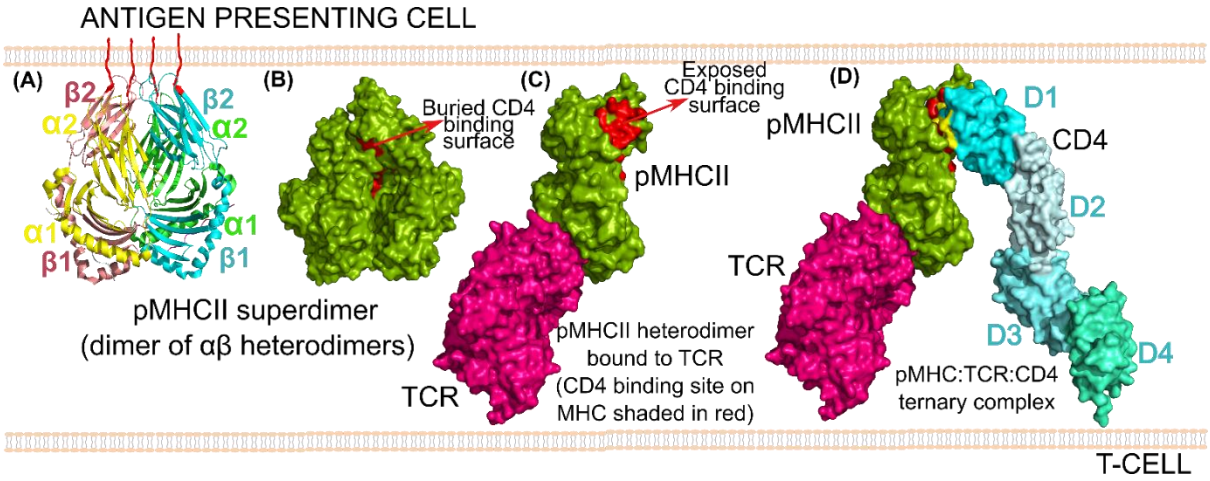

Fig.3 A new model of pMHCII:TCR:CD4 signalling: **(A)** Cartoon representation of pMHC-II superdimer observed in different crystal structures depicting dimer of heterodimer formed by an α and β chain. The C-termini of all the four chains (2α and 2β) face in the same direction as if emerging from the same cell. Thus, it is anchored at four points stabilising the upright orientation **(B)** Surface representation of the pMHCII superdimer hiding the CD4 binding site (shaded in red) in the dimer interface **(C)** TCR (magenta):pMHCII (green) complex reveals that pMHCII exists as heterodimer but not as superdimer in the complex. Note that the disassembly of superdimer into heterodimer exposes the CD4 binding site (shaded in red) on MHC. **(D)** Ternary complex structure of pMHC:TCR:CD4(cyan) (PDB entry: 3T0E) showing simultaneous interaction of TCR and monomeric CD4 with pMHCII heterodimer. CD4 interacts with MHCII via its D1 domain.

entries: 1FYT, 2ICW, 2WBJ, 2XN9, 1ZGL, IJ8H, etc.,), in the structures of pMHCII in complex with CD4(D1-D2, PDB entries: 1JL4, 3S4S, 3S5L, etc.,) and also in the ternary complex structure of pMHCII-TCR and CD4 (PDB entry: 3T0E)[61]. The structures of pMHCII in complex with CD4 (D1-D2) revealed that the CD4 binding region is in the interface formed by α_2 and β_2 domains on pMHCII which is completely hidden in the pMHCII superdimer interface[55]. The authors also propose a V-shaped model for the TCR-MHC-CD4 ternary complex which later matched with the crystal structure of the ternary complex (PDB entry: 3T0E,[61] with complete ectodomains of CD4. With the absence of superdimeric structures of pMHCII in complex with TCR and/or CD4, the earlier models which emphasized pMHCII superdimer as a functional entity lost their relevance and it is now established that TCR-pMHCII-CD4 is a complex containing a single unit of each protein.

Unfortunately, the importance of such an elegant pMHCII superdimer which was observed in many crystal-structures of pMHC and proven to be present on the cell-surface by other studies [66]was comfortably forgotten thereafter. Based on the observations made from all the currently available crystal structures of pMHCII superdimers, pMHC-TCR structures (>25 structures), pMHC-CD4(D1-D2), and the ternary complex structures of TCR-pMHC-CD4, we provide a model that reconciles all the experimental observations and highlight the importance of pMHCII superdimer. As observed in the pMHCII structures, the super-dimer might be the native state before binding to TCR which precludes CD4 binding as the CD4 binding site is buried in the superdimer interface **(Fig. 3B).** When the MHCII loaded with antigenic peptide interacts with specific TCR and with an appropriate strength, the resulting mechanoforce might dislodge the weak pMHCII superdimer, exposing the CD4 binding site **(Fig. 3C).** As mentioned above this is consistent with all the TCR:pMHCII crystal structures, where monomeric TCR is observed to interact with pMHCII heterodimer (and not the superdimer) with exposed CD4 binding site **(Fig. 3C)**. Subsequently, the binding of monomeric CD4 to pMHC-TCR complex forms the V-shaped arch **(Fig. 3D)** as observed in the ternary complex structures leading to downstream signaling initiated through Lck. Thus, it appears that pMHCII is an obligate superdimer that hides its CD4 biding site and avoids CD4 interaction prior to recognition of specific antigenic peptide on it by TCR with appropriate strength to ensure that the process of screening is accurate and intact. It should be noted that the hiding of the binding partner interface and utilization of the same interface to oligomerize and stand 'upright' on the cell surface, is a common strategy used by both CD4 (as described in the previous section) and pMHCII. This way, the superdimer of pMHCII stabilizes the 'upright' orientation for an efficient presentation of the antigenic peptide to T-cell. Therefore, the repeated observation of superdimers of pMHCII appears to be physiologically relevant, however, the presumption that pMHCII super-dimer is the signaling unit led to models that did not explain subsequent findings. Similarly, in the case of CD4, the investigations supporting both D4-D4 pairs as well as D1-D1 pair appears to be correct and together with several structures, they appear to represent the physiologically important association.

Another important observation that emerged out of this analysis is that, in the ternary complex of CD4-pMHCII-TCR, the overall structure of monomeric CD4 (D1-D4 domains) is very similar to the apo structure of CD4 (RMSD ~1.8Å) **(Fig. S2-A)** [61]. Our analysis indicates that the D4 domain which appears to interact with the membrane along its elongated surface as described in the apo structure, upon interaction with pMHCII moves away from the membrane,

orienting itself almost perpendicular to the membrane surface **(Fig. S2-B)**. Although the overall structure of CD4 is not changed, the reorientation of the D4 domain with respect to the membrane and the EJ-linker appears to provide sufficient length to reach the CD4 binding site on pMHCII. This reorientation due to interaction with pMHCII probably exerts sufficient machanoforce that dislodge the D4-D4 dimer and consequently disturbing the relative orientation of its intracellular domains. This change in the intracellular part of CD4 might also help its association with Lck. It has been proposed in the crystal structure of the ternary complex of CD4-pMHCII-TCR, the arc-shaped structure also brings the C-terminal D4 domain of CD4 closer to the TCR:CD3 assembly.

CD2 Oligomerization: CD2 is one of the first few structurally investigated cell surface molecules implicated in immune function and on the cell surface it is known to be present in close proximity to TCR:pMHC:CD4 complex[42, 67]. It is a cell adhesion molecule, expressed on T-lymphocytes. Human and rat CD2 are shown to interact with their cognate ligands CD58(LFA-3)[68] and CD48 respectively[69], all of which belong to CD2-like family, which also involve 'signaling lymphocyte activation molecules' (SLAM) consisting of nine members (SLAMF1, CD48, 2B4, Ly9, CD84(Ly9-b), NTB-A, CRACC, BLAME, CD2F)[70]. The ectodomain of CD2 consists of a membrane distal IgV followed by membrane-proximal IgC domains. Currently, there are two crystal structures of the complete ectodomain of CD2, one from human (PDB-entry: 1HNF)[42] and the other is from rat (PDB-Entry: 1HNG)[67]. Overall structures of human (hCD2) and rat CD2 (rCD2) are very similar **(Fig. S3)** with average C_α RMSD of 1.3Å, which confirms the relative

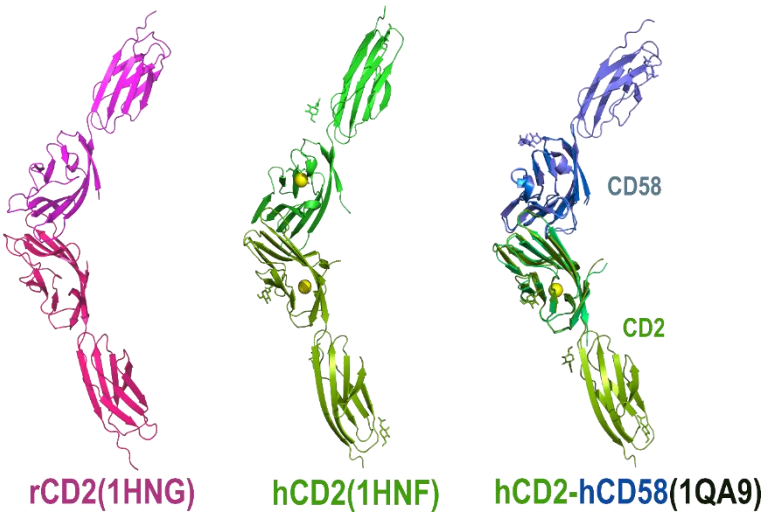

Fig.4 Comparative analysis of crystal structures of rCD2(PDB entry: 1HNG), hCD2 (PDB entry: 1HNF) and hCD2:CD58 complex (PDB entry: 1QA9). The dimeric associations observed in different crystal structures are similar and the homodimeric interactions of rCD2(magenta) and hCD2 (green) mimics the heterodimeric interaction of hCD2 with its cognate partner, CD58(blue) supporting the proposition that heterophilic interactions in CD2 family evolved from a homophilic precursor. hCD2:hCD58 structure shown here is obtained by superposition of full length hCD2 and full length CD58 on to respective proteins in CD2:CD58 containing only IgV domains (PDB entry: 1QA9)

rigid orientation of IgV and IgC domains. Interestingly, although CD2 is known to be a monomer in solution[71, 72], each molecule seen in both the structures is involved in the exactly similar dimeric association of IgV domains to that observed in canonical homophilic and heterophilic interactions involving IgV domains of immune receptors[73] which mimics the interaction of CD2 with its cognate partner CD58 **(Fig. 4)** [42].

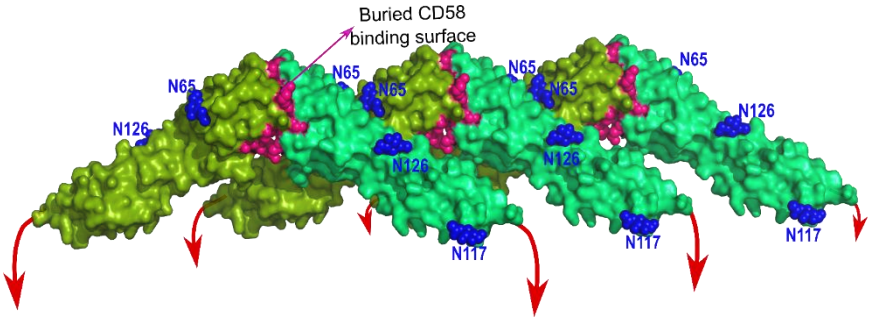

Fig.5 Surface representation of one-dimensional array of hCD2 as observed in the crystal structure (PDB entry:1HNF). Note that the dimer which appeared as '*trans'* as shown in Fig.4 can also be considered as '*cis'* and such *cis* dimers hide the CD58 binding site (shown as pink spheres) on CD2. The *'cis'* dimers further interact to form a one-dimensional array of molecules supporting each other in an upright orientation perpendicular to the membrane. The N linked glycans are shown in blue spheres. N-glycans (at N65) between the adjacent *'cis'* dimers act as fillers and the glycan at N117 which is close to the membrane might stabilise the upright orientation. The red arrows indicate EJ-linkers extending towards membrane.

Based on the observation that the genes for CD58/CD48 and CD2 lie in proximity on chromosome-1[74, 75], it has been proposed that the heterophilic interaction in this family might have evolved from a common homophilic precursor[72], suggesting that the conservation of canonical homodimeric interaction observed in both CD2 structures, is probably physiologically relevant. Further, another structure of hCD2 containing only IgV domain maintains a similar canonical front-face to front-face interaction **(Fig. S3)**, even though the structure is domain-swapped (PDB entry: 1CDC)[76]. Maintenance of similar and canonical interactions between IgV domains in all the three different crystal structures suggest that the observed association, although very weak-to-be-detectable from the solution studies, appears to be the preferred mode of association of CD2 molecules prior to its interaction with cognate partner. Careful analysis of crystal packing of hCD2 molecules reveals that these molecules form a 1D array along the crystallographic *z*-axis, which places the membrane-proximal IgC domain (hence the EJ-linker) of all the molecules of the 1D array in the same direction as if they are emanating from the same cell **(Fig. 5)**. It is interesting to note that the N-glycosylation at Asn65(in the structure-1HNF) is located at the gap created between canonical CD2 dimers of this array and appears to act as a filler **(Fig. 5).** Conservation analysis reveals that this residue is highly conserved (91%) across species **(Fig. 6).** Although the 1D array of CD2 appears to be physiological, the crystal dimensions are generally in the range of 100 microns or more, hence, such a long array can't be expected on the cell surface. Probably,

the membrane architecture (curved membrane surface) defined by the cytoskeleton and dynamics of the active cell membrane, limit the growth of such an array made of fragile interactions. In this array, the CD58 interaction surface of CD2 is completely buried in the CD2 homodimeric interface. If the array seen in the crystal is reminiscent of CD2 molecules on the cell membrane prior to its interaction with CD58, then such an association with 'closed' ligand-binding surface might suffice the condition required for avoidance of spurious interactions and also provide support for 'upright' orientation of the molecules on the cell surface. Concomitant with this observation, single-molecule tracking of few raft-associated (Lck and LAT) and non-raft associated proteins (CD2 and CD45), which are known to be important in T-cell function, suggested an unexpectedly low diffusion coefficient for a non-raft associated protein CD2 (lowest among all the molecules studied)[77]. Here the assumption was that raft-associated protein should diffuse slowly compared to non-raft associated proteins. Assuming, that the 1D array seen in the crystals of human CD2 is representative of the physiologically relevant association on the cell membrane, such an array of molecules is expected to diffuse slowly and hence appears to explain the unusual low diffusion coefficient observed for CD2.

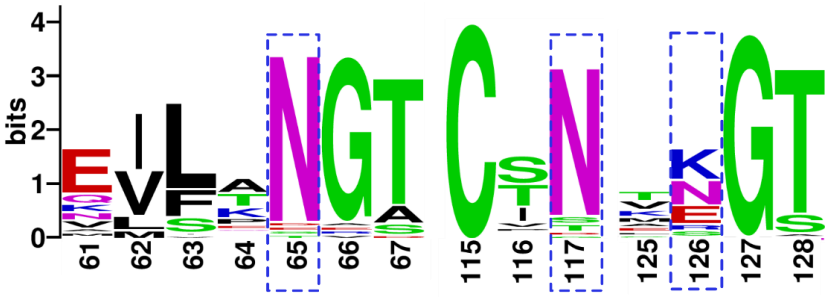

Fig.6 Sequence logo depicting the conservation of N-glycosylation sites in CD2 across species: N65 and N117 boxed in blue dashed lines (numbering according to PDB structure 1HNF) are conserved by 91% and 86.4% respectively. While the glycan at N65 might act as a filler between the adjacent proteins of the 1D-array, the glycan at N117 lies close to the membrane and might have stabilising interactions with the membrane. The third glycan at N126, also seen in the structure, but the site is only 27%

The superposition of rCD2 and hCD2 canonical dimers reveals that the two dimers are very similar **(Fig. S3).** However, while rCD2 dimer depicted a *trans* interaction with its cognate partner CD48, as explained above the highly similar hCD2 dimeric association represented *cis* interaction of the molecules on the cell surface **(Fig. S4-A).** This observation suggests that differentiating these interactions as either *cis* or *trans* appears to be inappropriate and sometimes leads to wrong interpretations. For example, in the structure of the D1-D2 fragment of CD4, it was concluded that the observed interaction between the D1 pair is *trans* as they appeared to come from two different cell surfaces (D2 domains are placed in opposite directions). However, this structure superposes well with the *cis* dimer observed in the crystal

structure of complete ectodomain (D1-D4, **Fig. S1-A**) of CD4[78]. Therefore, the canonical dimer-like association of CD2 can suffice condition required for *cis* as well as *trans* association, the same argument can be applied to other known homodimers **(Fig. S4-A)**. For example, NTB-A also called SLAMF6, is a homodimer and member of the CD2 family. The angle between the two protomers of the dimer is approximately 86°. The homomeric interaction seen in the crystal is always represented as *trans* interaction as if two molecules are coming from juxtaposed cells **(Fig. S4-B)**[41, 79], However, it is also possible that they can stand on the same cell surface as dimers using the same canonical dimeric interface as observed for human CD2. Hence, it is likely that the homomers can sometimes use their membrane distal domains for both *cis* and *trans* interactions, and hence in the absence of *trans* interaction with juxtaposed cells, they might utilize the same surface for *cis* interaction and stabilization of 'upright' orientation and avoid spurious signaling. Structures of homophilic dimers such as Nectin-1, CD166 are shown in **Fig. S5**, representing their possible orientation on the cell membrane.

1D array of TNFR1 with both parallel and anti-parallel dimers: Tumour Necrosis Factor (TNF) is a cytokine which interacts with its receptors TNFR-1 and TNFR-2. Currently, several TNF family members (29 receptors and 19 ligands in humans) have been identified and are known to play an important role in the immune response. While the ligands are type-II membrane proteins with trimeric extracellular domain and include both membrane-tethered as well as soluble versions, the receptors are type-I membrane proteins mostly membrane-bound with exception of soluble decoy receptors. Also, receptors are elongated structures with 3-4 cysteine-rich domains (CRDs) where some of them interact with multiple ligands[80]. Further, the 'ligand-induced trimerization' model suggests that the receptors get trimerized only in the presence of trimeric ligand, where the inter-digitization of monomeric receptors in the groves of trimeric ligands have been observed[81]. It has been proposed that the common trimeric architecture of TNF family ligands together with the promiscuous binding of receptors to different ligands creates the possibility of the formation of deleterious mixed trimers [82]. TNF mediated signaling being involved in a variety of processes ranging from cell proliferation to cell death induction, the formation of deleterious mixed-trimers put forth a challenge that how the receptors avoid the formation of mixed trimers to generate robust signaling? This raises the question of whether there exists any pre-signaling assembly that can segregate specific TNF receptors prior to the ligand-binding and avoid mixed trimeric assembly of ligand and receptors?

As mentioned above, TNFRs have elongated structures with 3-4 cysteine-rich domains (CRDs) which are difficult to crystallize. The crystal structures of TNFR1(PDB-1NCF)[83], reveals parallel dimeric association with predominant interactions located on the N-terminal CRD1-domain **(Fig. 7A).** Consistent with this observation, based on the mutational studies on the N-terminal domain and/or based on complete replacement/exchange of this domain with the same domain from the related receptor, it has been proposed that this domain is responsible for ligand-independent assembly of the receptor, and hence named as 'pre-ligand binding assembly domain (PLAD)'[84]. In brief, what is evident from various studies is that in the TNF family, the control of the signaling event is very complex and not mediated by mere ligand-receptor interactions. This further supports the possibility of the existence of a pre-signaling assembly of specific TNF receptors, prior to the ligand-binding and which is distinct from the ligand-induced trimer. The PLAD mediated cheek-to-cheek association of TNFR1 appears to satisfy the condition required for stabilization of the 'upright' orientation of this receptor on the cell surface. However, the existence of a long EJ-linker in TNFR1 (15 residues) raises the question that how such a long linker stabilizes the 'upright' parallel orientation of TNFR1 on the cell surface (the same argument applies to TNRF2 with 56 residue EJ-linker)?

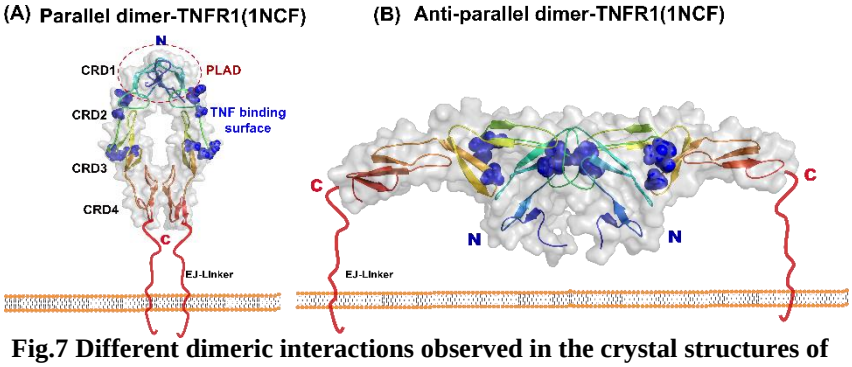

Fig.7 Different dimeric interactions observed in the crystal structures of TNFR1 (A) Parallel dimeric association of TNFR1 (PDB entry: 1NCF) interacting via the N-terminal PLAD (encircled in red dashed lines) exposes the ligand binding surface (residues shown as blue spheres). **(B)** The anti-parallel association, also observed in the same crystal structure effectively hides the ligand binding residues in the dimeric interface.

From the TNFR1 crystal structure, Naismith *et. al*, also recognize an association that can be assigned as an anti-parallel association of the TNFR1 receptors, which buries the ligand biding surface distributed over both CRD2 and CRD3 domains **(Fig. 7B).** Based on this observation, they propose that the burial of ligand-binding residues might avoid unintended signaling events[83]. The anti-parallel model also appeared correct as it explained the differential cross-linking effects of therapeutic agonistic and antagonistic antibodies[85]. However, the anti-parallel model undermines the role of PLAD observed to be critical for parallel dimer and which was proven to be physiologically important by many experimental studies. Our packing analysis reveals an interesting observation, where PLAD mediated parallel-dimers associate

with each other burying the ligand-binding site and forms a 1D array containing both parallel and anti-parallel dimers **(Fig. 8).** Such an ordered and continuous association of molecules, together with the EJ-linker might help these molecules to stand 'upright' and segregate on the cell surface. Further, the zipper like arrangement with hidden active sites can only respond to the presence of sufficient concentration of specific TNF ligands that can effectively unzip the 1D array. Again, this observation reconciles two contradicting experimental observations that (1) PLAD mediated parallel dimerization is important for function and (2) anti-parallel association is physiologically relevant. Further, the 1D zipper mediated segregation of specific kinds of TNFRs might also avoid the deleterious mixed-trimer formation of ligands and receptors. Supporting this observation, another crystal structure of TNFR family member called DCR3 maintains both parallel and anti-parallel dimeric association (PDB-3MHD) **(Fig S6)**, probably suggesting a common association preference of TNFR family members

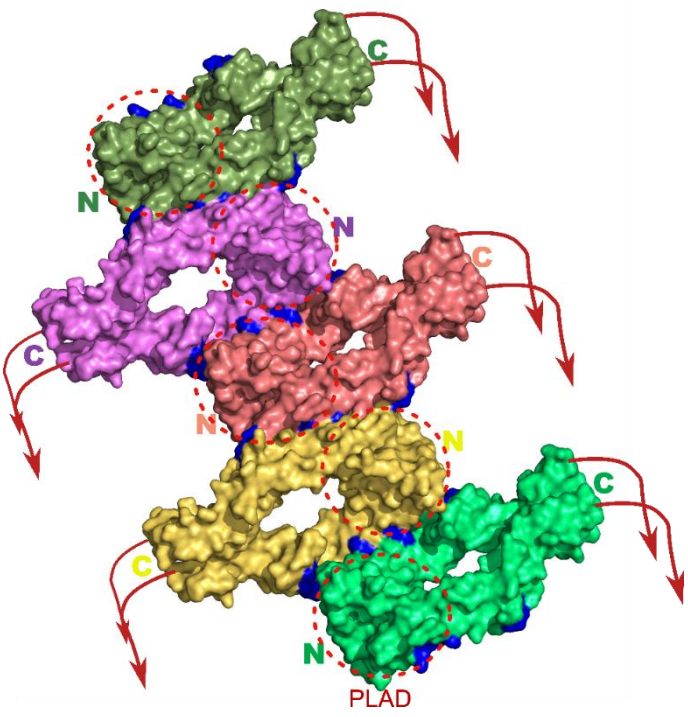

Fig.8 One dimensional array observed in the crystal structure of TNFR1 that supports both parallel and antiparallel dimeric associations. Such an array of molecules on the cell membrane stabilized by velcro of weak interactions supports the upright orientation of receptors and accounts for the importance of PLAD (encircled in red dashes) in receptor association as well as hides the ligand binding site (spheres in blue) to prevent spurious signalling and thus reconciles the opposing views.

At this stage, the question remains, how these dimeric receptors involve themselves in 'ligand-induced trimerization'. Various studies on single-pass transmembrane (TM) domains have shown that their organization does play an important role in the function of these proteins. [86-88]. The CD95 (FasR), an apoptosis-inducing death receptor and a member of tumor necrosis factor receptor (TNFR) superfamily. The NMR structures of both human and mouse TM domain revealed a highly analogous homo-trimeric organization[89]. Structure-guided and cancer-related mutations that disturb the TM trimerization and *in vivo* studies demonstrated that these mutants show impaired apoptosis induction, suggesting the importance of

trimerization of TM domains in signaling [90]. Sequence analysis of different TNFR superfamily members reveals a conserved proline in each family at a specific position on the TM domain, indicating that proline plays a critical role in the trimerization of TNFR superfamily members. Based on the extensive analysis, the authors conclude that the TM-trimerization helps in post-ligand trimeric associations of FasR with its ligands including Fas, which are again trimers, resulting in clustering of liganded FasR:FasL complexes[89]. They also propose that the pre-liganded 'upright' dimer of the ectodomain, re-associates into trimer in the presence of Fas ligand. Hence, signatures at the TM domain that supports the trimerization and those at the ectodomain that supports dimerization, being facile, appear to facilitate the required oligomeric organization depending on the pre- and post signaling states. Consonant with these observations, super-resolution microscopy data shows that the dimers and trimers of FasR co-exist on the plasma membrane[88].

Other examples: Apart from the specific examples described above, the mode of association of several cell surface receptors observed in the crystal structures mimic the condition required for stable 'upright' orientation. Our analysis also suggests that stabilization of 'upright' orientation can be mediated by any part of the ectodomain, and appears to be optimally carved for its size, architecture and functional role. For example, there are four structures of PD-L1 containing both IgV and IgC domains crystallized in different conditions (PDB entries: 4Z18, 3FN3, 5JDR, 3BIS)[40, 89, 91, 92]. It is interesting to note that in all the four cases, PD-L1 crystallizes as a parallel dimer (*cis* dimer), where the membrane-proximal C-terminal ends are close to each other **(Fig. S7-A).** Although no biophysical experiments can detect their dimeric association in solution, crosslinking studies show mostly the dimeric population[40]. However, the possible reason for such an association has never been described before. Our analysis of the structures reveals that the back-face of the IgC domain consisting of ABED strands from both the interacting molecules fuse together to form a continuous β-sheet stabilizing the expected physiologically relevant dimeric structure that can support 'upright' orientation **(Fig. S7-B).** The similar, but slightly distinct fusing of ABED strands from the IgC domain has also been observed in the structure of the full-length ectodomain of CD58.

In the crystal structure of human EphA2 receptor (Eph, a cell surface receptor protein tyrosine kinase) both the liganded and unliganded forms show *cis* contacts between ligand-binding domains (LBD- LBD), sushi domains (sushi-sushi) of adjacent proteins. Similar to the perching of the D4 domain of CD4 on to membrane, in the case of the EphA2 receptor, the membrane-proximal FN2 domain is oriented almost perpendicular to the other four domains of the

structure[46]. In a recent study, using MD simulations and mutational studies it was shown that the FN2 domain interacts with membrane along its elongated surface through positively charged lysine and arginine residues and thus stabilizes the Ephrin dimers on the cell surface **(Fig. S8)**[93]. It can be noticed that as the size of the ectodomain increases (four domains in CD4 and five domains in Ephrin), other than the oligomeric association, nature sometimes appears to utilize direct interaction of membrane-proximal domain with the plasma membrane as an additional stabilization factor. In the case of CEACAM1, SPR based binding studies, chemical cross-linking experiments and molecular electron tomography studies on CEACAM1 confirms the adaptability of ectodomain to make parallel *cis* interactions (A-dimers) as well as anti-parallel *trans* dimers (C-dimers). Liposomes with anchored CEACAM-1 molecules also have shown clusters and the degree of clustering did not change with change in lipid quantity suggesting that the clustering is influenced more by ectodomain interactions rather than by protein-lipid ratio on the membrane[48].

Adding to these possibilities, small molecules and ions in the extracellular milieu can also contribute to the specific organization of cell surface proteins. For example, in the case of cadherins, interdomain stabilization is mediated by calcium ions[94] and hence most crystallization soup, and consequently, the crystal structures are derived in the presence of Ca^{2+} ion. From these structures, it is clear that Ca^{2+} ions play a critical role in interdomain stabilization. Also, it has been observed that the presence of cations, such as Ca^{2+}/Mg^{2+} contributes to the stability of CEACAM-1 ectodomains and increases the formation of dimers and multimers[48]. The role of Ca^{2+} ions in the organization of membrane-associated proteins like Syntaxin and SNAP25 has been reported and its role in higher-order clustering has also been discussed[95]. This situation calls for the use of physiological concentration of Ca^{2+} in other experiments such as super resolution microscopy to effectively compare the results from independent experiments. Similarly, the negative charge on glycans is expected to repel from the membrane and probably help the weak dimers to stand ‘upright’[96]. Considering the hydrophilicity, flexibility, mobility of glycans in the aqueous environments, they are expected to be the best fillers that can allow adaptation to changing environments, depending on the organization/reorganization of these receptors on the cell surface. For example, in the case of symmetric dimers, such as CD2 and NTB-A any glycan attached to the membrane-proximal domain of one of the protomers of the dimer places the glycan attached to another protomer in the opposite direction, creating similar glycan-membrane interactions at the two opposite sides of the plane passing through both the protomers of the dimer and perpendicular to the

membrane **(Fig. 5)**. Hence, glycans might also play a role in the erect orientation of the molecules on the surface, however, considering their flexibility, they may not by themselves define the preferential association.

Conclusion: The consistent observation that the burying of ligand binding surface and the formation of a 1D array (as explained here for CD4, CD2, and TNFR1) in the pre-liganded state and its ability to reconcile contradicting observations suggests that the formation of an ordered array is a common phenomenon, probably utilized by many (if not all) of the cell surface molecules involved in signaling. This is important particularly in the case of constitutively expressed molecules as they might involve in spurious signaling. This kind of ordered association provides support for the stabilization of 'upright' orientation and appropriate positioning of membrane distal domains for robust signaling, effectively avoiding spurious signaling events. However, none of these proteins show such oligomerization propensity in solution, suggesting that the observed *cis* interactions are too weak-to-be-detected by biophysical studies. This weak oligomerization propensity appears to be essential as these molecules are required to organize/reorganize depending on the signaling state. Although in all the examples discussed here, the complete burial of interacting surface is observed, this may not be the criteria to avoid spurious signaling as even the partial burial and their further association into an array can probably avoid spurious signaling events. The analysis and the conclusions drawn here entirely relies on the repeatability of a pattern, i.e., the pattern of association in several crystal structures of the same or related molecules, and repeated observation that constitutively expressed cell-surface molecules often burry their ligand-binding surface and so on. To retrieve a pattern and matching them requires sufficient data. Unfortunately, crystal structure determination being a laborious process, unless the molecules are of great importance, it is difficult to find several structures of the same or related molecules. Further, it should also be noted that proteins are crystallized in the presence of a combination of various ions, precipitants, and sometimes with strongly interacting metal ions, the interaction with which the protein surface can override the natural preference of weakly interacting proteins [97]. Similarly, cloning artifacts at the interface of association is another factor that can affect their preferred association. Moreover, we need to have sufficient functional data on these molecules to efficiently connect the results of the pattern of association to their functional roles. For example, pMHCII and TCR are molecules involved in the continuous screening process; hence, the self-association of pMHCII into higher-ordered arrays may not support efficient peptide presentation. Similarly, the pMHCII recognition site on TCR should always

be accessible. Accordingly, neither the pMHCII superdimer nor the TCR gets involved in a supramolecular organization. However, in the case of pMHCII, tethering to the membrane at four different points due to superdimer formation appears to stabilize the 'upright' orientation appropriate for peptide presentation and also buries CD4 binding site to avoid spurious signaling. Similarly, TCR might stabilize itself with the association of CD3 molecules to stand 'upright' on the cell surface, without involving itself in the higher-ordered association that occludes its pMHC recognition sites. However, CD4 co-receptor forms a pre-liganded 1D array, hiding its pMHCII biding surface in the non-signaling resting state and interacts with pMHCII only upon proper recognition of antigenic peptide by TCR that opens the CD4 binding site on pMHCII. In the case of TNFR1, a constitutively expressed protein, the observed 1D array suffices the requirements for both the contradicting observations that (1) the PLAD mediated parallel association is important for function and (2) and that the antiparallel association that satisfactorily explains the differential cross-linking effects of therapeutic agonistic and antagonistic antibodies. Hence, these signaling molecules are carved with well-defined surface features that suffice their functional role. Therefore, what is generally termed as clusters of molecules on the cell surface, without any impetus to their specific order appears to be inappropriate and models such as 'lipid-rafts' and even the 'fence and pickets' may not provide an explanation that emulates specific sorting and robust signaling ability of these molecules. Although thousands of experimental work supporting these models exist with unhealthy controversies[21], in the absence of know-how on their ordered association preferences, more experiments and controversies appear to survive. It can be arguable that the observed effect of 'lipid-rafts' or 'fences and pickets' is probably a consequence of the inherent ordered association of cell surface proteins that places their relatively rigid transmembrane domain in regular and ordered fashion to which molecules such as sphingolipids and cholesterol can adhere to.

Similarly, as observed in the crystal structures of hCD2, cadherins and ephrin and so on, when these molecules are free in solution, at higher concentrations they can form large 1D or 2D arrays of the dimension of the crystal itself (~100 microns). In contrast, when they are tethered to a membrane, the dynamics of cytoskeleton supported active membrane might limit the growth of these weakly associated arrays or due to the curved membrane architecture, they might circularize by joining of two ends of the 1D array. Hence, it appears that 'fences and pickets' imagined out of cytoskeleton elements, probably can control the growth of these clusters, but do not carry any element of their functional control. Consonant with this

observation, recent studies suggest that when the cytoskeleton is disturbed by cytoskeleton inhibitor ‘latrunculin A’, cadherin clusters were found to grow to a size that is larger than generally observed [30].

Overall, it appears that the clustering (rather an ordered array) being the key element of signal processing; the rules that govern the clustering, their type, optimum strength, accuracy and reproducibility of the signal are hidden within the defined set of self organizations of cell surface molecules. Such an organization is maintained by surface complementarity of the molecules together with polar interactions that are facile enough to allow their organization/reorganization depending on the functional context (polar interactions are highly directional and are pliable in the aqueous environment). The *cis* interaction mediated arrays, in the presence of cognate-partner of higher affinity might reorganize into an array of different architecture, strength and dimension and consequently position their intracellular cytoplasmic domains in an ordered fashion, the geometry of which might support the recruitment of downstream signaling molecules Considering these observations, we would like to call our model as “Preferential Protein Interaction of Signalling Receptors/Ligands (**PPISR-Model**). In the absence of a suitable model that can explain the cell surface heterogeneity and its importance in function, we believe that this new conceptual framework solely developed based on the available data and recurrence of pattern, appears to provide a basis for future experiments for molecules not only involved in immune function but also for other receptors that are involved in the repertoire of signaling events.

Acknowledgment

UAR would like to thank DBT, India for Ramalingaswamy fellowship (2011-2015) and Vision Group on Science and Technology (VGST), grant #191, Karnataka, India. SL would like to acknowledge the fellowship from CSIR, Government of India.

Author contribution

UAR conceptualized and reviewed the manuscript. UAR and SL analysed the data and wrote the manuscript

Declaration of Interest

UAR and SL declare no conflict of interest

References:

1. Alberts B, J.A., Lewis J, et al., *Membrane Proteins*, in *Molecular Biology of the Cell*. 2008, Garland Science: NewYork.
2. Junghans, V., et al., *Dimensions and Interactions of Large T-Cell Surface Proteins.* Front Immunol, 2018. **9**: p. 2215.
3. Davis, S.J. and P.A. van der Merwe, *The kinetic-segregation model: TCR triggering and beyond.* Nat Immunol, 2006. **7**(8): p. 803-9.
4. Chang, V.T., et al., *Initiation of T cell signaling by CD45 segregation at 'close contacts'.* Nat Immunol, 2016. **17**(5): p. 574-582.
5. Truong-Quang, B.A. and P.F. Lenne, *Membrane microdomains: from seeing to understanding.* Front Plant Sci, 2014. **5**: p. 18.
6. Liu, C., et al., *Single-Molecule Tracking and Its Application in Biomolecular Binding Detection.* IEEE J Sel Top Quantum Electron, 2016. **22**(4).
7. Sibarita, J.B., *High-density single-particle tracking: quantifying molecule organization and dynamics at the nanoscale.* Histochem Cell Biol, 2014. **141**(6): p. 587-95.
8. Siegel, R.M., et al., *Measurement of molecular interactions in living cells by fluorescence resonance energy transfer between variants of the green fluorescent protein.* Sci STKE, 2000. **2000**(38): p. pl1.
9. Baaden, M., *Visualizing Biological Membrane Organization and Dynamics.* J Mol Biol, 2019. **431**(10): p. 1889-1919.
10. Garcia-Parajo, M.F., et al., *Nanoclustering as a dominant feature of plasma membrane organization.* J Cell Sci, 2014. **127**(Pt 23): p. 4995-5005.
11. Recouvreux, P. and P.F. Lenne, *Molecular clustering in the cell: from weak interactions to optimized functional architectures.* Curr Opin Cell Biol, 2016. **38**: p. 18-23.
12. Liu, W., H. Wang, and C. Xu, *Antigen Receptor Nanoclusters: Small Units with Big Functions.* Trends Immunol, 2016. **37**(10): p. 680-689.
13. Hartman, N.C. and J.T. Groves, *Signaling clusters in the cell membrane.* Curr Opin Cell Biol, 2011. **23**(4): p. 370-6.
14. Simons, K. and E. Ikonen, *Functional rafts in cell membranes.* Nature, 1997. **387**(6633): p. 569-72.
15. Kusumi, A., et al., *Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: high-speed single-molecule*

tracking of membrane molecules. Annu Rev Biophys Biomol Struct, 2005. **34**: p. 351-78.
16. Sieber, J.J., et al., *Anatomy and dynamics of a supramolecular membrane protein cluster.* Science, 2007. **317**(5841): p. 1072-6.
17. Levental, I. and S. Veatch, *The Continuing Mystery of Lipid Rafts.* J Mol Biol, 2016. **428**(24 Pt A): p. 4749-4764.
18. Sevcsik, E. and G.J. Schutz, *With or without rafts? Alternative views on cell membranes.* Bioessays, 2016. **38**(2): p. 129-39.
19. Nichols, B., *Cell biology: without a raft.* Nature, 2005. **436**(7051): p. 638-9.
20. Calder, P.C. and P. Yaqoob, *Lipid rafts--composition, characterization, and controversies.* J Nutr, 2007. **137**(3): p. 545-7.
21. Levental, I., K.R. Levental, and F.A. Heberle, *Lipid Rafts: Controversies Resolved, Mysteries Remain.* Trends Cell Biol, 2020. **30**(5): p. 341-353.
22. Bernardino de la Serna, J., et al., *There Is No Simple Model of the Plasma Membrane Organization.* Front Cell Dev Biol, 2016. **4**: p. 106.
23. Kenworthy, A.K., et al., *Dynamics of putative raft-associated proteins at the cell surface.* J Cell Biol, 2004. **165**(5): p. 735-46.
24. Rothberg, K.G., et al., *Caveolin, a protein component of caveolae membrane coats.* Cell, 1992. **68**(4): p. 673-82.
25. Skibbens, J.E., M.G. Roth, and K.S. Matlin, *Differential extractability of influenza virus hemagglutinin during intracellular transport in polarized epithelial cells and nonpolar fibroblasts.* J Cell Biol, 1989. **108**(3): p. 821-32.
26. Low, S.H., et al., *Syntaxins 3 and 4 are concentrated in separate clusters on the plasma membrane before the establishment of cell polarity.* Mol Biol Cell, 2006. **17**(2): p. 977-89.
27. Uhles, S., et al., *Isoform-specific insulin receptor signaling involves different plasma membrane domains.* J Cell Biol, 2003. **163**(6): p. 1327-37.
28. Magenau, A., et al., *Discreet and distinct clustering of five model membrane proteins revealed by single molecule localization microscopy.* Mol Membr Biol, 2015. **32**(1): p. 11-8.
29. Vyas, N., et al., *Nanoscale organization of hedgehog is essential for long-range signaling.* Cell, 2008. **133**(7): p. 1214-27.
30. Wu, Y., P. Kanchanawong, and R. Zaidel-Bar, *Actin-delimited adhesion-independent clustering of E-cadherin forms the nanoscale building blocks of adherens junctions.* Dev Cell, 2015. **32**(2): p. 139-54.
31. Dawson, D.W., et al., *CD36 mediates the In vitro inhibitory effects of thrombospondin-1 on endothelial cells.* J Cell Biol, 1997. **138**(3): p. 707-17.
32. Githaka, J.M., et al., *Ligand-induced growth and compaction of CD36 nanoclusters enriched in Fyn induces Fyn signaling.* J Cell Sci, 2016. **129**(22): p. 4175-4189.
33. Pageon, S.V., et al., *Functional role of T-cell receptor nanoclusters in signal initiation and antigen discrimination.* Proc Natl Acad Sci U S A, 2016. **113**(37): p. E5454-63.
34. Cambi, A., et al., *Organization of the integrin LFA-1 in nanoclusters regulates its activity.* Mol Biol Cell, 2006. **17**(10): p. 4270-81.
35. Price, W.N., 2nd, et al., *Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data.* Nat Biotechnol, 2009. **27**(1): p. 51-7.
36. Aricescu, A.R. and E.Y. Jones, *Immunoglobulin superfamily cell adhesion molecules: zippers and signals.* Curr Opin Cell Biol, 2007. **19**(5): p. 543-50.
37. Schwartz, J.C., et al., *Structural basis for co-stimulation by the human CTLA-4/B7-2 complex.* Nature, 2001. **410**(6828): p. 604-8.

38. Stamper, C.C., et al., *Crystal structure of the B7-1/CTLA-4 complex that inhibits human immune responses.* Nature, 2001. **410**(6828): p. 608-11.
39. Ikemizu, S., et al., *Structure and dimerization of a soluble form of B7-1.* Immunity, 2000. **12**(1): p. 51-60.
40. Chen, Y., et al., *A dimeric structure of PD-L1: functional units or evolutionary relics?* Protein Cell, 2010. **1**(2): p. 153-60.
41. Cao, E., et al., *NTB-A receptor crystal structure: insights into homophilic interactions in the signaling lymphocytic activation molecule receptor family.* Immunity, 2006. **25**(4): p. 559-70.
42. Bodian, D.L., et al., *Crystal structure of the extracellular region of the human cell adhesion molecule CD2 at 2.5 A resolution.* Structure, 1994. **2**(8): p. 755-66.
43. Zhang, N., et al., *Binding of herpes simplex virus glycoprotein D to nectin-1 exploits host cell adhesion.* Nat Commun, 2011. **2**: p. 577.
44. Harrison, O.J., et al., *Nectin ectodomain structures reveal a canonical adhesive interface.* Nat Struct Mol Biol, 2012. **19**(9): p. 906-15.
45. Wu, H., P.D. Kwong, and W.A. Hendrickson, *Dimeric association and segmental variability in the structure of human CD4.* Nature, 1997. **387**(6632): p. 527-30.
46. Himanen, J.P., et al., *Architecture of Eph receptor clusters.* Proc Natl Acad Sci U S A, 2010. **107**(24): p. 10860-5.
47. Harrison, O.J., et al., *The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins.* Structure, 2011. **19**(2): p. 244-56.
48. Klaile, E., et al., *The CEACAM1 N-terminal Ig domain mediates cis- and trans-binding and is essential for allosteric rearrangements of CEACAM1 microclusters.* J Cell Biol, 2009. **187**(4): p. 553-67.
49. Ereno-Orbea, J., et al., *Molecular basis of human CD22 function and therapeutic targeting.* Nat Commun, 2017. **8**(1): p. 764.
50. Zhang, X., et al., *Expression, refolding, purification, molecular characterization, crystallization, and preliminary X-ray analysis of the receptor binding domain of human B7-2.* Protein Expr Purif, 2002. **25**(1): p. 105-113.
51. Morgan, A.A. and E. Rubenstein, *Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome.* PLoS One, 2013. **8**(1): p. e53785.
52. Bhatia, S., et al., *Different cell surface oligomeric states of B7-1 and B7-2: implications for signaling.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15569-74.
53. Richter, C., et al., *The tumor necrosis factor receptor stalk regions define responsiveness to soluble versus membrane-bound ligand.* Mol Cell Biol, 2012. **32**(13): p. 2515-29.
54. Rudd, C.E., et al., *Molecular analysis of the interaction of p56lck with the CD4 and CD8 antigens.* Adv Exp Med Biol, 1991. **292**: p. 85-96.
55. Wang, J.H., et al., *Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10799-804.
56. Kwong, P.D., et al., *Molecular characteristics of recombinant human CD4 as deduced from polymorphic crystals.* Proc Natl Acad Sci U S A, 1990. **87**(16): p. 6423-7.
57. Moldovan, M.C., et al., *CD4 dimers constitute the functional component required for T cell activation.* J Immunol, 2002. **169**(11): p. 6261-8.
58. Briant, L., et al., *Transduction of activation signal that follows HIV-1 binding to CD4 and CD4 dimerization involves the immunoglobulin CDR3-like region in domain 1 of CD4.* J Biol Chem, 1997. **272**(31): p. 19441-50.

59. Bourgeois, R., et al., *Association between disruption of CD4 receptor dimerization and increased human immunodeficiency virus type 1 entry.* Retrovirology, 2006. **3**: p. 31.
60. Langedijk, J.P., et al., *Location of CD4 dimerization site explains critical role of CDR3-like region in HIV-1 infection and T-cell activation and implies a model for complex of coreceptor-MHC.* J Biol Chem, 1993. **268**(23): p. 16875-8.
61. Yin, Y., X.X. Wang, and R.A. Mariuzza, *Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4.* Proc Natl Acad Sci U S A, 2012. **109**(14): p. 5405-10.
62. Brown, J.H., et al., *Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1.* Nature, 1993. **364**(6432): p. 33-9.
63. Gorga, J.C., et al., *Crystallization of HLA-DR antigens.* Res Immunol, 1991. **142**(5-6): p. 401-7.
64. Stern, L.J. and D.C. Wiley, *The human class II MHC protein HLA-DR1 assembles as empty alpha beta heterodimers in the absence of antigenic peptide.* Cell, 1992. **68**(3): p. 465-77.
65. Schafer, P.H., et al., *The assembly and stability of MHC class II-(alpha beta)2 superdimers.* J Immunol, 1998. **161**(5): p. 2307-16.
66. Schafer, P.H., S.K. Pierce, and T.S. Jardetzky, *The structure of MHC class II: a role for dimer of dimers.* Semin Immunol, 1995. **7**(6): p. 389-98.
67. Jones, E.Y., et al., *Crystal structure at 2.8 A resolution of a soluble form of the cell adhesion molecule CD2.* Nature, 1992. **360**(6401): p. 232-9.
68. Selvaraj, P., et al., *The T lymphocyte glycoprotein CD2 binds the cell surface ligand LFA-3.* Nature, 1987. **326**(6111): p. 400-3.
69. Kato, K., et al., *CD48 is a counter-receptor for mouse CD2 and is involved in T cell activation.* J Exp Med, 1992. **176**(5): p. 1241-9.
70. Veillette, A., *SLAM-family receptors: immune regulators with or without SAP-family adaptors.* Cold Spring Harb Perspect Biol, 2010. **2**(3): p. a002469.
71. van der Merwe, P.A., et al., *Human cell-adhesion molecule CD2 binds CD58 (LFA-3) with a very low affinity and an extremely fast dissociation rate but does not bind CD48 or CD59.* Biochemistry, 1994. **33**(33): p. 10149-60.
72. Van der Merwe, P.A., et al., *Measuring very low affinity interactions between immunoglobulin superfamily cell-adhesion molecules.* Biochem Soc Trans, 1993. **21**(4): p. 340S.
73. Kundapura, S.V. and U.A. Ramagopal, *The CC' loop of IgV domains of the immune checkpoint receptors, plays a key role in receptor:ligand affinity modulation.* Sci Rep, 2019. **9**(1): p. 19191.
74. Staunton, D.E., et al., *Blast-1 possesses a glycosyl-phosphatidylinositol (GPI) membrane anchor, is related to LFA-3 and OX-45, and maps to chromosome 1q21-23.* J Exp Med, 1989. **169**(3): p. 1087-99.
75. Wong, Y.W., et al., *Structure, expression, and genetic linkage of the mouse BCM1 (OX45 or Blast-1) antigen. Evidence for genetic duplication giving rise to the BCM1 region on mouse chromosome 1 and the CD2/LFA3 region on mouse chromosome 3.* J Exp Med, 1990. **171**(6): p. 2115-30.
76. Murray, A.J., et al., *One sequence, two folds: a metastable structure of CD2.* Proc Natl Acad Sci U S A, 1995. **92**(16): p. 7337-41.
77. Douglass, A.D. and R.D. Vale, *Single-molecule microscopy reveals plasma membrane microdomains created by protein-protein networks that exclude or trap signaling molecules in T cells.* Cell, 2005. **121**(6): p. 937-50.
78. Ryu, S.E., et al., *Structures of an HIV and MHC binding fragment from human CD4 as refined in two crystal lattices.* Structure, 1994. **2**(1): p. 59-74.

79. Chattopadhyay, K., et al., *Sequence, structure, function, immunity: structural genomics of costimulation.* Immunol Rev, 2009. **229**(1): p. 356-86.
80. Locksley, R.M., N. Killeen, and M.J. Lenardo, *The TNF and TNF receptor superfamilies: integrating mammalian biology.* Cell, 2001. **104**(4): p. 487-501.
81. Smith, C.A., T. Farrah, and R.G. Goodwin, *The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death.* Cell, 1994. **76**(6): p. 959-62.
82. Chan, F.K., *The pre-ligand binding assembly domain: a potential target of inhibition of tumour necrosis factor receptor function.* Ann Rheum Dis, 2000. **59 Suppl 1**: p. i50-3.
83. Naismith, J.H., et al., *Crystallographic evidence for dimerization of unliganded tumor necrosis factor receptor.* J Biol Chem, 1995. **270**(22): p. 13303-7.
84. Chan, F.K., et al., *A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling.* Science, 2000. **288**(5475): p. 2351-4.
85. Vanamee, E.S. and D.L. Faustman, *Structural principles of tumor necrosis factor superfamily signaling.* Sci Signal, 2018. **11**(511).
86. Call, M.E., et al., *The structure of the zetazeta transmembrane dimer reveals features essential for its assembly with the T cell receptor.* Cell, 2006. **127**(2): p. 355-68.
87. Call, M.E., K.W. Wucherpfennig, and J.J. Chou, *The structural basis for intramembrane assembly of an activating immunoreceptor complex.* Nat Immunol, 2010. **11**(11): p. 1023-9.
88. Lau, T.L., et al., *The structure of the integrin alphaIIbbeta3 transmembrane complex explains integrin transmembrane signalling.* EMBO J, 2009. **28**(9): p. 1351-61.
89. Fu, Q., et al., *Structural Basis and Functional Role of Intramembrane Trimerization of the Fas/CD95 Death Receptor.* Mol Cell, 2016. **61**(4): p. 602-613.
90. Lee, S.H., et al., *Somatic mutations of Fas (Apo-1/CD95) gene in cutaneous squamous cell carcinoma arising from a burn scar.* J Invest Dermatol, 2000. **114**(1): p. 122-6.
91. Zhang, F., et al., *Structural basis of a novel PD-L1 nanobody for immune checkpoint blockade.* Cell Discov, 2017. **3**: p. 17004.
92. Lin, D.Y., et al., *The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors.* Proc Natl Acad Sci U S A, 2008. **105**(8): p. 3011-6.
93. Chavent, M., et al., *Structures of the EphA2 Receptor at the Membrane: Role of Lipid Interactions.* Structure, 2016. **24**(2): p. 337-47.
94. Kim, S.A., et al., *Calcium-dependent dynamics of cadherin interactions at cell-cell junctions.* Proc Natl Acad Sci U S A, 2011. **108**(24): p. 9857-62.
95. Zilly, F.E., et al., *Ca2+ induces clustering of membrane proteins in the plasma membrane via electrostatic interactions.* EMBO J, 2011. **30**(7): p. 1209-20.
96. Polley, A., et al., *Glycosylation and Lipids Working in Concert Direct CD2 Ectodomain Orientation and Presentation.* J Phys Chem Lett, 2017. **8**(5): p. 1060-1066.
97. Hegde, R.P., et al., *Can the propensity of protein crystallization be increased by using systematic screening with metals?* Protein Sci, 2017. **26**(9): p. 1704-1713.