

<https://doi.org/10.1038/s42003-025-07830-5>

Unusual traits shape the architecture of the Ig ancestor molecule



Alejandro Urdiciain^{1,2,3,6}, Thomas Madej^{4,6}, Jiyao Wang⁴, James Song⁴, Elena Erausquin^{1,2,3},
Philippe Youkharibache⁵✉ & Jacinto López-Sagaseta^{1,2,3}✉

Understanding the ancestral Ig domain's molecular structure and tracing the evolution of Ig-like proteins are fundamental components missing from our comprehension of their evolutionary trajectory and function. We have determined high-resolution structures of two Ig-like proteins from the evolutionary most ancestral phylum, Porifera. The structures reveal N-terminal Ig-like domains with an unconventional configuration of features that set them apart from canonical Ig domains. These findings prompted us to call this novel domain as Ig "Early Variable" (EV)-set. Remarkably, the EV-sets are linked to C1-set domains. To the best of our knowledge, the C1-set has not been previously reported in non-vertebrates. The IgV and IgC1 tandems and their combination into functional Ig-like receptors are part of the adaptive immune system in higher vertebrates, which allows for highly specific immune responses. By unveiling important clues into the molecular configuration of ancestral Ig domains, these findings challenge and expand our understanding of how immunity has evolved within its current landscape.

The Ig fold is a primitive and highly conserved protein domain that has undergone a profound diversification through evolution. Primarily known for its role in the vertebrate immune system and in particular for its presence in the structure of antibodies, the Ig fold is found in a myriad of proteins involved in a wide range of biological functions, from neural development to cell signaling and adhesion^{1,2}.

The Ig-fold architecture organizes 70–110 amino acids in two beta sheets stabilized by hydrogen bonding facing each other in a beta-sandwich. Its structural properties, such as stability and adaptability, have made it a fundamental building block of many protein families across different species.

The presence of Ig-like domains in ancient lineages like sponges and the expansion observed in vertebrates shows how such a versatile structure can be reconfigured and repurposed through evolution.

The marine sponge *Geodia cydonium* provides an excellent model for understanding the molecular basis of ancient immune-like systems and cellular recognition in Metazoa. SAMs (sponge adhesion molecules) and RTK (receptor tyrosine kinases) play key roles in cell signaling, adhesion, and recognition of self and non-self, processes that likely predate the evolution of more complex immune systems. SAMs function primarily in cell adhesion and tissue fusion, and include both a short (SAMS) and long

(SAML) form, with the latter carrying an additional N-terminal region of unknown function. RTKs mediate intracellular signaling, supporting the hypothesis that these primitive forms of life represent a living window into the evolutionary development of immune recognition and signaling systems^{3–5}.

Here, by analyzing the molecular architecture of the Ig domain assemblies belonging to the marine sponge *Geodia cydonium* RTK and SAML proteins, we report novel Ig structural features that contribute important pieces to the understanding of the origins and evolution of the Ig fold.

Results and discussion

The sequence homology of the *Geodia cydonium* sponge Ig-like proteins to antibody domains was noticed early on^{4–6}, but the molecular structure remained unknown. A search for homologous structures using either VAST^{7,8} or DALI⁹ led to thousands of structural homologs to the N-terminus domain. This is expected since there are over 13,000 structures in the PDB (Protein Data Bank) containing over ca. 75,000 Ig-domains belonging to the four types of variants: V-set, I-set, C1-set, or C2-set. The overwhelming majority of which consist of antibody fragments: Fabs with IgV-IgC1 of heavy and light chains, or simply VH-VL domains.

¹Unit of Protein Crystallography and Structural Immunology, Navarrabiomed, 31008 Navarra, Spain. ²Public University of Navarra (UPNA), Pamplona, 31008 Navarra, Spain. ³Navarra University Hospital, Pamplona, 31008 Navarra, Spain. ⁴National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. ⁵Cancer Data Science Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. ⁶These authors contributed equally: Alejandro Urdiciain, Thomas Madej. ✉e-mail: philippe.youkharibache@nih.gov; jacinto.lopez.sagaseta@navarra.es

We therefore looked not just for single domain level structural homologs, but for pairs of Ig-like domains in tandem, and their consecutive domain topologies, in addition to sequence and structure homology, to determine the most relevant evolutionarily-related tandem Ig-like domains among known proteins in the PDB.

Ig-containing regions of both SAML and RTK from *Geodia cydonium*, found in their extracellular space, were successfully expressed in sf9 cells and purified to homogeneity, as indicated by the elution profiles under size-exclusion chromatography (Supplementary Fig. 1). Both SAML and RTK proteins led to thin plate-looking crystals that enabled structure determination with 1.6 and 2.1 Å resolutions, respectively (Table 1). Overall, both datasets allowed the calculation of electron density maps of enough quality to accurately build the molecular morphologies of both receptors (Fig. 1a).

SAML and RTK molecular structures are constituted by elongated assemblies of two Ig-like domains in tandem connected to one another by short and conserved Leu-Pro-Leu linkers. This configuration resembles protein domain architectures common to many cell adhesion and cell surface receptors in vertebrates.

Interestingly, some of the structural homologs with the highest sequence identity up to 30% point to human IgVL lambda (PDB 3T0X), IgVH (PDB 3N9G) domains, and 20% with Shark VNAR (Variable domain of new antigen receptor) domains (PDB 1VES) (Fig. 1c). Also, VAST structural homologs encompassing N- and C-terminus Ig-domains, in tandem with good overall (rigid) structural alignments, lead to PD-L2 (PDB 3BP5), CD200 (PDB 4BFI), Nectin-4 (PDB 4FRW), and NCR3LG1 (B7-H6) (PDB 6YJP) that all present an IgV-IgC1 extracellular chain architecture, characteristic of many vertebrate immune cell surface receptors (Fig. 1b). Precisely, the extracellular architecture is composed of two Ig-domains in tandem, the N-terminus with a topology that can be classified, preliminarily, as a V-set followed by a C-terminus domain belonging to the C1-set (Fig. 1a). The orientation of the Ig domains of SAML/RTK depicts a single polypeptide conformed by two β – sandwich barrel-like folds subtly angled in relation to each other.

To further explore the evolutionary relationships of SAML and RTK within the vast diversity of Ig-like domains, we also adopted a phylogenetic

perspective (Fig. 1d, e). The resulting trees highlight four main Ig domain types grouped into distinct clades: V-set, I-set, C1-set, and C2-set. SAML and RTK N-terminal domains cluster within the V-set group, supporting their homology to vertebrate IgV domains such as those found in antibody fragments (e.g., VH and VL domains). SAML-IgEV domain exhibits significant structural and sequence similarities to invertebrate IgV domains, such as the amphioxus IgV-C2 (PDB: 5XPW) and variable region-containing Chitin-Binding proteins (VCBP, PDB: 2FBO). The tree clearly shows that SAML (PDB 8OVQ), IgVJ-C2, and VCBP cluster with vertebrate IgV-set domains, highlighting their close evolutionary relationship (Fig. 1d).

Similarly, the C-terminal domains align with the C1-set, consistent with their structural resemblance to IgC1 domains in extracellular immune receptors, such as PD-L2 and CD200.

The placement of SAML and RTK within the tree, along with their tandem EV-set and C1-set architecture, underscores their evolutionary link to vertebrate immune surface receptors while retaining features that might represent an ancestral configuration. This phylogenetic perspective aligns with the structural and sequence homology observed between sponge Ig-like domains and vertebrate counterparts, providing further evidence for their evolutionary conservation.

Additionally, multiple sequence alignment (MSA) reveals that these invertebrate Ig-EV domains share conserved signature motifs also observed in vertebrates (Fig. 1f). Key features include the cysteine bridge between the B and F strands, a tryptophan residue on the C strand, an arginine residue on the E strand, and a highly conserved [DxGRYWC] motif on the F strand. Within this motif, the tyrosine (Y) residue is especially well conserved, where its side chain hydroxyl group forms a hydrogen bond with the backbone of nearby residues creating a structural feature called a “tyrosine corner” that contributes to the stability of the protein fold. It is found in over 90% of FnIII and Ig variable (V-type) domains.

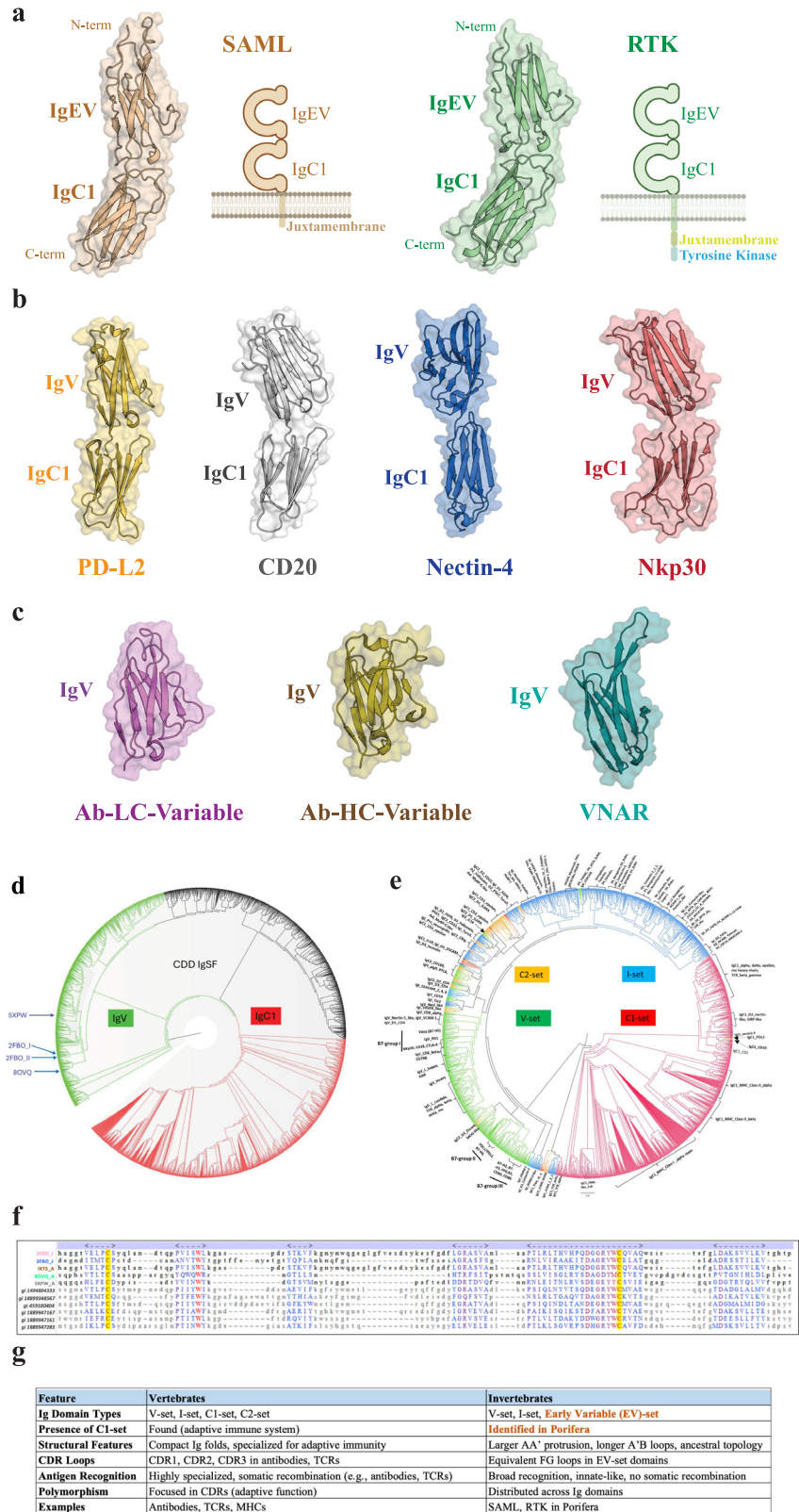
Further sequence alignments and structural comparisons across a wider range of invertebrate IgV domains are crucial for deepening our understanding of their evolutionary roles in immune recognition. However, the current limited availability of protein structures and sequences for Ig-domains in invertebrates and marine organisms constrains more

Table 1 | Data collection and refinement statistics (molecular replacement)

	SAML (PDB 8OVQ)	RTK (PDB 8QPX)
Data collection		
Space group	P12 ₁	P12 ₁
Cell dimensions		
a, b, c (Å)	33.08 41.16 77.80	40.91 32.73 92.05
α, β, γ (°)	90 90 90	90 90.8 90
Resolution (Å)	77.8–1.6 (1.64–1.61)	40.9–2.1 (2.26–2.10)
R _{merge}	0.067 (0.74)	0.035 (0.247)
I / σI	9.3 (1.5)	17.4 (5.0)
Completeness (%)	99.40 (99.82)	98.47 (98.47)
Redundancy	4.1 (3.7)	3.3 (3.4)
CC _{1/2}	0.997 (0.576)	0.999 (0.947)
Wilson B-factor	20.56	34.44
Refinement		
Resolution (Å)	77.8–1.6 (1.64–1.61)	40.9–2.1 (2.26–2.10)
No. reflections	27228 (2722)	14382 (2721)
R _{work} / R _{free}	0.169 (0.304)/0.216 (0.336)	0.222 (0.246)/0.250 (0.304)
R.m.s. deviations		
Bond lengths (Å)	0.012	0.016
Bond angles (°)	1.85	2.87

Both datasets were collected from single crystals. Values in parentheses are for the highest-resolution shell.

Fig. 1 | Structure and Ig configuration of SAML and RTK receptors. **a** The molecular structures of SAML and RTK extracellular Ig-like regions are shown with wheat and pale green colors, respectively, in cartoon and surface mode, along with a schematic representation of the corresponding full-length receptors anchored in the cell membrane. N- and C-terminal sites are indicated as well as the transmembrane and intracellular regions, which are assigned accordingly. Figure 1a contains images prepared with BioRender. **b** Display of analogous vertebrate Ig-like receptors configured by an IgV-IgC1 tandem: *PD-L2*, Programmed Cell Death 1 Ligand 2 (pale-yellow color), *CD200*, B-lymphocyte antigen (light gray color), *Nectin4*, Nectin cell adhesion molecule-4 (blue color) and *B7-H6(NCR3LGI)*, a ligand to Natural cytotoxicity triggering receptor 3 also known as NKp30 or CD337 (salmon color). **c** The structures of three different Ig-V sets are also shown for structural comparison purposes: *Ab-LC-variable*, antibody light chain variable region (violet color), *Ab-HC-variable*, antibody heavy chain variable region (olive color), *VNAR*, shark variable domain of new antigen receptor (teal color). Pro-Ser-Thr(P/S/T)-rich domain. **d** Phylogenetic analysis of these invertebrate IgV domains demonstrates clustering with conserved vertebrate IgV domains from the conserved domain database's Ig superfamily (CDD IgSF), supporting the hypothesis that IgV-type domains were broadly co-opted for immune recognition across diverse evolutionary lineages. **e** Phylogenetic tree showing the conserved structural features across vertebrate and invertebrate Ig domains, with an emphasis on the evolutionary conservation of these Ig domains: V-set (green), C1-set (red), C2-set (yellow), and I-set (blue) highlight evolutionary divergence. The tree clusters Ig domains into distinct groups based on structural and sequence conservation. **f** Multiple sequence alignment (MSA) showing sequences of immunoglobulin-like (Ig-like) domains from various proteins or species. Amino acid sequences are aligned to highlight conserved regions and motifs. Key features of the image include. Gaps in the alignment (-) indicate insertions or deletions in some sequences relative to others. The MSA provides a visual summary of conserved regions across the sequences, emphasizing key residues and motifs critical for the structure and function of Ig-like domains. **g** Comparative table showing the distribution and presence of the different Ig domain sets across vertebrates and invertebrates, specifically focused on immune system components or related aspects. The structural features primarily described in this study, i.e., the finding of a novel Early Variable set, and the presence of IgC1 set in invertebrates, are highlighted with color.



comprehensive phylogenetic analyses. With the increased accessibility of AlphaFold predictions and protein sequences from lower invertebrates, future studies may provide critical insights into the evolutionary trajectory of IgV domains.

Each of the two Ig-domains in SAML and RTK present the Cys-Cys-Trp canonical triad, highly conserved across the vast array of Ig-

domains^{12,10,11}. Thus, the morphology of these Ig-like domains presents two confronted β -pleated sheets held together by a disulfide bridge and a neighboring Trp residue that contributes to the overall compactness and stability of the Ig fold (Supplementary Fig. 2).

An additional Trp is located in the N-terminal domains of SAML and RTK, near the disulfide bridge, presumably to further balance the overall

architecture of these domains. However, this Trp is not conserved in other Ig-like proteins in vertebrates, which might indicate a less relevant structural role and subsequent loss through evolution.

The combination of a monomeric IgV-IgC1 tandem extracellular region points to an ancestral form equivalent to the immune cell surface receptors of the vertebrate immune system. There are naturally both similarities and differences between the vertebrate V-set domains found in antibodies and the SAML/RTK N-terminal domains, which resemble more to a Shark VNAR domain, but also to some human V-set domains as found in PD-L2.

The N-terminal domain

The N-terminal domain topology can be compared to Ig-domains of either V-set or I-set. Indeed, many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to the I-set type¹². Yet, most cell surface receptors in vertebrate immune cells tend to have an IgV domain designed to interact with their ligands IgV domains, as for receptors-ligands such as PD-1, PD-L1, PD-L2, CD28, and CTLA-4 of the B7:CD28 extended family, that contain one or two extracellular Ig-domains with an IgV domain at the N-terminus. V-set and I-set topologies are extremely close when considering the Ig-framework alone¹². Importantly, the most significant structural differences between the two types lie in the loops that are very short in I-set

domains as compared to their V-set counterparts (Fig. 2). These differences are particularly notorious for the short FG loop, which correspond to CDR3 in antibodies, the BC loop, CDR1 in antibodies, and a C'D loop with the absence of a C'' strand and consequently no C'C'', CDR2 in antibodies (Fig. 2 and Supplementary Fig. 3). Notably, the FG loop in SAML/RTK exhibits a significant divergence that could potentially exclude the N-terminal domain of SAML/RTK from being classified as part of the Ig-I set. The FG loop length in SAML/RTK resembles that of VNAR, falling in between the length range observed for antibody light and heavy chains CDR3 loops (Fig. 2). Thus, the relatively extended length of SAML/RTK CDR3-like loop is especially interesting in pointing to a potential functional role already present in primitive organisms.

One should note however that IgV domains in sharks, called VNAR, lack the CDR2 region too, called, in that case, HV2, still hypervariable in sequence. VNAR domains¹³ and Igl domains align in fact remarkably well as noticed previously¹⁴. This region can also be either highly flexible in well-known human cell surface receptors, such as PD-L1, or missing altogether, as in the case of PD-L2. The N-term domains of SAML/RTK show significant sequence similarities with both Igl and VNAR domains (Fig. 2).

The three share a significant analogy in their respective topologies, while all differences between Igl and VNAR, as with SAML/RTK, occur in loops' length and structure. In conclusion, SAML and RTK N-terminal domains resemble VNAR and PD-L2 in that respect.

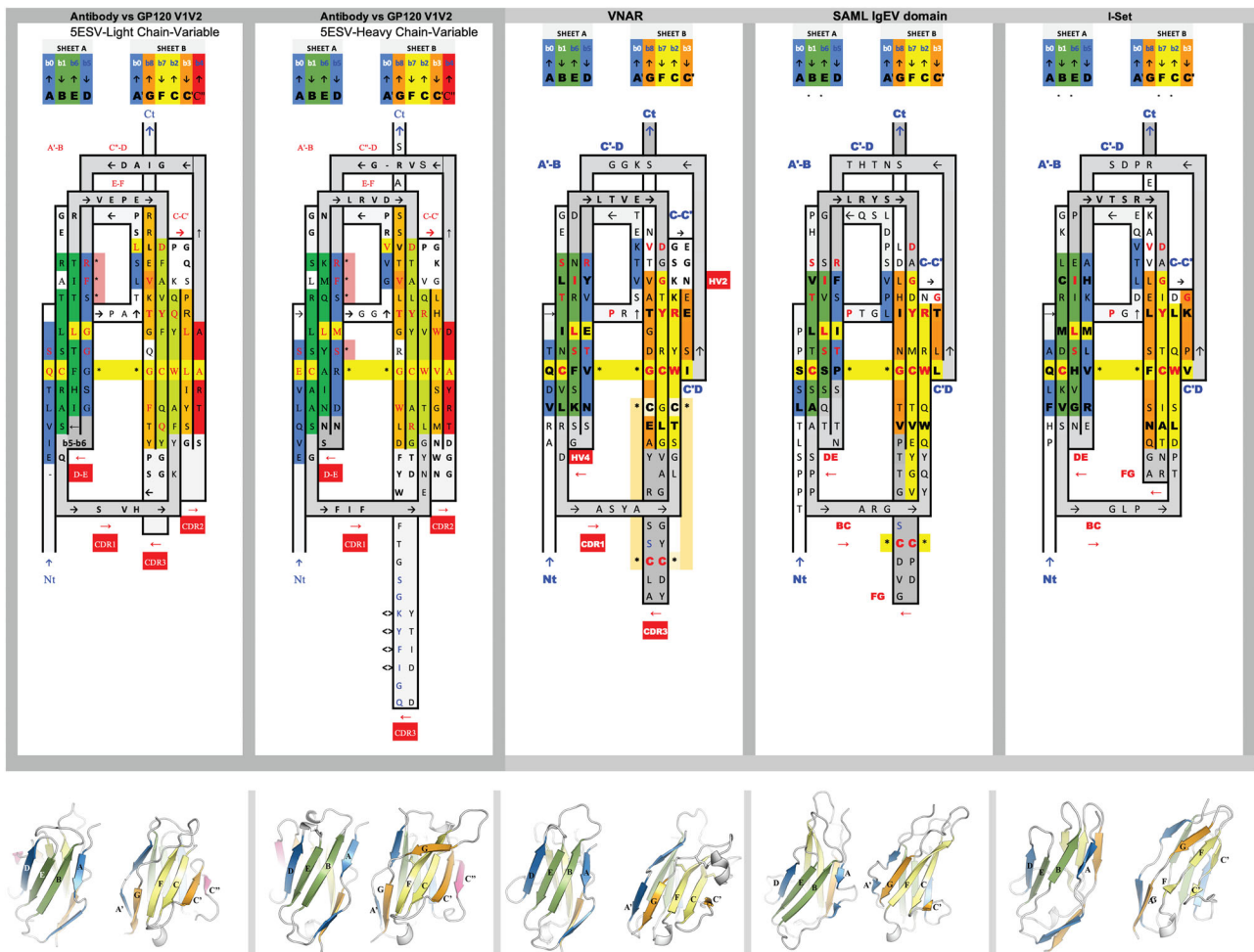


Fig. 2 | Sequence and topology 2D alignment of N-term Ig-domains. 2D sequence/topology alignments allow the integration of sequence and topological assembly into one single map and subsequent comparison of a series of diverse Ig sets. Shown are the 2D maps for a series of Ig-sets representative of the evolution of the Ig domain in the animal kingdom. Colors correspond to the *Ig Protodomain reduced rainbow*

*color spectrum*⁴¹ with four strands colors blue, green, yellow, and orange for ABCC' and DEFG and a fifth color red for the C'' strand, if present. Using this color scheme, a sheet ABED will be blue-green-green-blue; a sheet GFCC' will be orange-yellow-yellow-orange.

SAML/RTK show significant similarities on central strands B (SxTxxC), C (WxR) on the ABED sheet, E including the EF loop (SxSLxISxxLRxxx) and F (DxGxYxC) on the GFCC' sheet to both. While both VNAR and Igl domains exhibit the highly conserved pattern in strand F, the conservation on the preceding E strand and EF loop is more extensive and convincing with VNARs than with Igl. An additional sequence homology with VNAR and antibody variable domains also occurs in the D strand (RF/YxxT/S) (Fig. 2).

The presence of two cysteines structurally configured to form a disulfide bridge in the CDR3 loops occurs in natural antibodies' VH domains¹⁵. Non-canonical cysteines are also common CDR3 loops of VNAR type I-III domains. The only loop where SAML/RTK resembles an Igl is the CC' loop, that is 4-aa longer in VNAR. However, this discrepancy with VNAR suggests that SAML/RTK Ig domains should not be considered as pure VNAR type.

In addition to these observations, SAML/RTK N-terminal domains also present an AA' strand break, with A being part of the ABED sheet and A' part of the A'GFCC' sheet, as occurs in both I- and V sets. Thus, unlike the canonical IgV-set, SAML/RTK N-terminal Ig-domains do not feature a C' strand. The absence of such a strand is a known feature in the Igl-set, and therefore this feature imprints a partial Ig-I set character to SAML/RTK Ig N-terminal domains. Further, in both V-set and I-set domains, the Ig domain begins with an A strand in the ABED sheet and transitions through a structural kink into the A' strand. This A' strand interfaces with the GFCC' sheet and frequently exhibits a kink associated with a cis-Proline residue, running in parallel with the G strand to protect it from exposure. However, the AA' strand split in SAML/RTK exhibits a remarkable bulge than what is observed for Igl or IgV domains (Supplementary Fig. 3). This unique feature suggests that the AA' strand associates with functional properties^{16,17} that might have been lost through evolution, and might well constitute a further distinctive trait associated with the ancient Porifera Ig domain. Also, the A'B-loop in SAML/RTK is more extended than in IgV and Igl domains alike. In SAML, we find the A'B loop is conformed by a 4-aa (LSQP) motif, while only 2-aa patterns are detected in VNAR and the I-set, TG and EG, respectively. Additionally, we also noted that the N-terminal regions of SAML/RTK feature a proline-rich A strand, a trait commonly observed in I-set and V-set domains. On the other hand, however, these regions lack key I-set domain hallmarks: specifically, the motif where three residues follow the YxC motif in the F strand, with an Asn residue serving as a connector between the BC and FG loops¹⁸ (Fig. 2). Thus, both the AA' protrusion and extension in the A'B loop might represent unique features of ancestral Ig-like domains. Based on our observations, this domain is not an Igl from a sequence or loop structure standpoint.

Accordingly, and if considered an Ig-V similar to VNAR, it might well represent a novel domain that we accordingly name the "early variable" (EV) domain", with the major difference to most IgV domains being larger AA' protrusion, and a longer A'B loop. In addition, a short C' accompanied by a lack of C'' strand and CDR2 region, a pattern shared with VNAR domains, excludes SAML/RTK N-terminal domains from belonging to the canonical Ig-V domain family. Combined with a long CDR3 (FG) loop and the absence of a CC' loop, these findings further strengthen the concept of a novel and ancestral Ig domain.

A comparison of those regions equivalent to the CDR loops in antibodies and TCRs show a remarkable degree of structural conservation and overall positioning throughout primitive and contemporary organisms (Supplementary Fig. 4). This highlights a fine-tuned evolution of the Ig domain defined by a strong preservation of the Ig beta-sandwich architecture scaffold accompanied by structural transitions involving not only molecular diversification but the emergence of novel mechanisms, i.e., somatic hypermutation.

The C-terminal domain

The C-terminal domain, however, presents a clear C1-set topology and this is a significant finding. Effectively, while V-set domains have been found in

numerous life forms, the C1-set has only been found in vertebrates so far^{2,19,20} and is considered a hallmark of adaptive evolution in vertebrates². This represents a surprise since Ig-C1 domains have been linked to the evolution of the vertebrate immune system and is present in particular in molecules of the adaptative immune system (antibodies, TCRs and MHCs). It is found in numerous cell surface receptors of the innate immune system, such as B7 proteins such as PD-L1/L2 or CD80/86, and present especially in BCRs, antibodies, TCRs and MHC molecules involved in adaptive immunity jawed vertebrates^{19,21,22} (Fig. 1). Thus, the C1-set domains were believed to be present only in vertebrates.

Importantly, in addition to variable domains, the Ig-C1 architecture confers antibody chains the ability to dimerize. This is believed to be one of the innovations in the adaptive immune system¹⁰. Effectively, C1-set domains are found to dimerize in an antiparallel mode, using the ABED sheet in antibodies' CH1/CL domains, as well as CH3 or CH4 domains in polymeric Igs^{3,22-24}, or in TCR α/β . In shark antibodies (IgNAR), the C1-set constant domains that follow the VNAR variable domain also dimerize through their C1 and C3 domains^{13,24}. Neither SAML nor RTK C-domains have shown evidence for dimerization in either their crystal forms or in solution, as confirmed by crystallographic and size-exclusion chromatography with multi-angle static light scattering (SEC-MALS) studies (Supplementary Fig. 5).

Therefore, the presence of the C1-set domain, apart from establishing an evolutionary link to vertebrate domains, may point to a possible oligomerization *in vivo*. Alternatively, it might indicate an additional evolutionary step that confers the C1-set the potential to dimerize in more recently diverged lineages.

Immunoglobulin-like (Ig) domains are classified into four primary types: Igl, IgV, IgC1, and IgC2, each uniquely characterized by the presence or absence of additional β strands^{18,25,26}. The structural configuration and β strand composition distinguish IgV and IgC domains. IgV domains are typically composed of nine β strands, organized into two β sheets: the first sheet includes strands A, B, E, and D, while the second comprises strands G, F, C, C', and C'' (the region covering C'' is sometimes replaced by a flexible loop as in PD-1, or missing as in PD-L2, or VNAR domains found in Sharks. The A strand can be split into two segments, that are conventionally named A and A', for their association with either side of the sandwich sheets.

In contrast, IgC-set domains, which are subdivided into IgC1 and IgC2, feature a more compact structure with fewer β strands. The IgC1 variant omits the C' and C'' strands, presenting only seven β strands. IgC2 domains incorporate an extra C' strand, akin to that in the IgV domains but lack the D strand found in the back face of IgV structures. This difference alters the configuration and surface area of the β sheets between the IgC1 and IgC2 types, despite their similar strand count.

Thus, upon analysis of the canonical four main Ig sets, it is determined that SAML/RTK C-terminal domains belong to the IgC1-set, distinguished by the inclusion of a D strand and the absence of a C' strand (Fig. 3). The SAML/RTK domain features specific motifs: LxCxA on the B strand, NxxxSLxI on the E strand, CxxxS on the C strand, and YxCxV on the F strand, aligning with the hallmark characteristics of canonical IgC1-set domains. Interestingly, these findings align with the hypothesis by Pasquier and Chrétien²⁷ pointing to a primitive receptor form of monomeric nature, assembled by Ig-domains without somatic recombination and featuring adhesion molecule-like properties.

A distinctive feature of the interdomain interface of SAML and RTK is the C1 domain presenting a long DE loop forming a compact super-secondary structure with the BC loop, something quite unusual in known IgC1 domains, indeed looking quite similar to a C'E loop as observed in C2-domains straddling the two sheets of the sandwich, while stemming from the D strand. An IgC1 vs. IgC2 topology lies essentially in swapping the C' strand in IgC2 to a D strand position hence forming an IgC1 sandwich composed of GFC and ABED sheets rather than GFCC' and ABE sheets in IgC2. The prominent compact BC and DE loop of the C-terminal Ig domain (IgC1) domain forms a compact substructure that, in turn, forms a compact interface with the A'B and EF loops of the N-terminal Ig domain (IgEV),

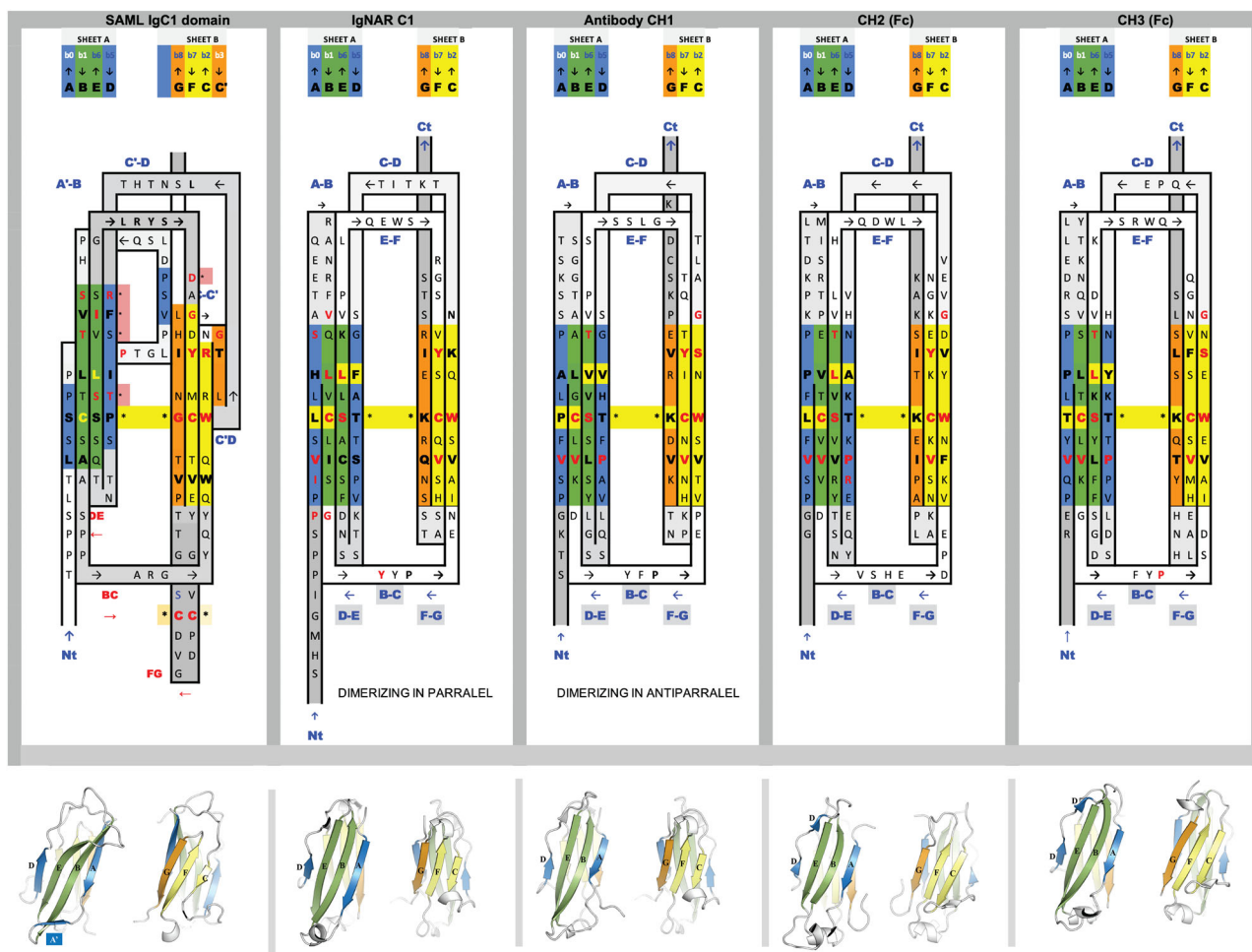


Fig. 3 | Sequence and topology 2D alignment of C-term Ig-domains. 2D topology comparison of SAML C-terminal Ig domain with IgNAR C1, and CH1, CH2, and CH3 sets.

with a buried surface of ca. 500 Å². This does confer to the IgEV-IgC1 tandem domains of SAML and RTK a fused/rigid structure.

Ig-tandemers

Rigid structural alignments can find the structural homologs of two domain structures only if their relative domain orientation is similar, and protein chains containing multiple Ig-domains exhibit very high conformational plasticity, due to flexible interdomain linkers and diverse loops involved at the domain-domain interfaces¹⁴. To identify the closest homologs among known structures, we looked for structures with two consecutive Ig-like domains in tandem, considering a rigid body alignment at the domain level, independently of their relative orientation. We call such structures Ig-tandemers. In our case, we can focus on those tandemers presenting a C-terminus domain with a C1-set topology Ig, while the N-terminus domain may present either an I-set or V-set topology. A canonical C1 topology presents a straight A strand and the presence of a D strand as a distinctive mark.

We used the NCBI structure alignment program VAST^{7,8} to compare the SAML structure against the full PDB database (as of Jan 31, 2024). The two SAML domains were run separately, as well as the full chain. We then combined the results by extracting those protein chains that aligned with the two SAML domains, in tandem. For each domain, we computed the alignment length, RMSD for that structural superposition, and the number of identical residues in the SAML/target alignment. We ended up with 18047 hits from 6581 different PDB files (Supplementary Table 1 and Supplementary Data 1). The chained Ig-domains across the PDB structures offer a diversity of tandem topologies presenting variants from IgV-IgC1,

IgV-IgC2, IgI-IgI, IgI-IgC2, etc., with diverse relative orientations (not considered explicitly).

Given the overrepresentation of antibody Fab and TCR structures in the PDB database that possess an IgV-IgC1 topology, we separated all the hits into a Fab antibody/TCR dataset and a non-antibody dataset. The former containing ca. 15,000 chains and the second with ca. 3000 chains. Sequence homology to represented Fab antibodies/TCRs averages 23.8% (max 47.7% on structurally aligned residues) for the N-terminal domain hits (Ig1) and 15.8% on the C-terminal domain (Ig2) (max 32.8%). Moreover, for the Ig1 domain, an average of 76 residues were aligned with a 2.1 Å average RMSD, and for the Ig2 domain, an average of 63 residues with a 2.0 Å average RMSD. In the non-Ab dataset, the sequence identity averages 16.4% (max 35.2%) on Ig1 and 16.6% (max 34.4%) on Ig2. For Ig1 there was an average of 71 residues aligned and a 2.3 Å average RMSD, for Ig2 an average of 59 residues aligned and 2.1 Å average RMSD. In this dataset, considering sequence, topology, and structure altogether, VAST identifies several tandem homologs, pointing to type I cell surface receptors involved in cell-cell adhesion and immunity. These proteins are all type I cell surface receptors, as are SAML and RTK. Among them CD200, PD-L2 (B7-DC, CD273), NCR3LG1(B7-H6), CD80 (B7-1), and nectin-4. CD200 (4BFI.B) is expressed in a wide range of cell types, including neurons, endothelial cells, and immune cells. It interacts with its receptor, CD200R (4BFI.A); PD-L2 (3BP5.B) is expressed on activated human T cells and regulates their function²⁸; and B7-H6 (P.C) is also part of the B7 family that consists of structurally related cell-surface protein ligands, which bind to receptors on lymphocytes that regulate immune responses. CD80 (B7.1) (118L.A) binds to CD28 and CTLA-4 receptors on T cells in forming and

stabilizing the immunological synapse. All these interacting cell surface proteins are part of an extended B7:CD28 family. Nectins are adhesion junction proteins involved in the nervous and the immune system, and also serve as virus receptors, such as the well-known poliovirus receptor (PVR or CD155), or nectin-4 (4FRW.B) as the measles virus receptor²⁹. We show a comprehensive list of structural homologs in Supplementary Data 1, including the lymphocyte activation gene 3 LAG-3 (7TZE.A), an immune checkpoint receptor and target for cancer immunotherapy^{30–32} or CD277/butyrophilin-3 (4F9L.A)³³.

Disulfide bridge in the “CDR3” loop

Given the quality of the electron density signal, we could notice a remarkable trait in the SAML Ig-EV domain. A cysteine pair in the FG loop (“CDR3”), where Cys252 presents up to three alternative conformations, of which one leads to a disulfide bond with Cys258 (Supplementary Fig. 7). Occupancies and B-factors for the three conformations were determined, respectively, to values of 0.31/23.94 (conformer 1), 0.25/22.22 (conformer 2), and 0.44/22.13 (conformer 3) through anisotropic refinement of B-factors. Non-canonical cysteines have been observed in human immunoglobulins, particularly in the CDR3 of the heavy chain (CDR-H3), as inferred from a study with ~3 billion VH sequences³⁸.

The number of non-canonical cysteines, i.e., those found outside the Ig-fold main body and which are encoded by Diversity Gene segments, in the most common human CDR-H3 can vary from 1 to 8 cysteines, and their frequency decreases as the number of cysteines increases. It has also been observed that the size of the CDR-H3 grows along with the number of non-canonical cysteines. The non-canonical disulfide bridge is more common in other species compared to humans, specifically in rooster, camel, llama, shark, and cow, and are more frequently observed in the IgM class. A comparison in Ig sequences between naive lymphocytes and memory lymphocytes revealed there is 2% fewer cysteines in CDR-H3 compared to memory lymphocytes, so it is suggested that this increase in cysteines might be linked to somatic hypermutation. Here we are talking about an adaptive immune response, since lymphocytes are involved. In the case of sponges, we would be talking about the innate immune response, as the sequences for these pattern recognition receptors are germline-encoded, meaning they are inherited and not subject to recombination or somatic hypermutation. As for humans, it has been observed that a disulfide bridge can also be formed between a cysteine found in the CDR-H3 and another cysteine found in other regions of the Ig, as in CDR-H1 and H2³⁰. In a study by Almagro and Colleagues¹⁵, the authors observed that antigen binding was diminished in a significant manner when the non-canonical CDR-H3 cysteines were mutated to alanine, which indicates the potential relevance of this class of disulfide bridges in antigen recognition. The most common sequence in which these non-canonical cysteines occur in CDR-H3 is: C-4X-C, with 4X being a 4 amino acid sequence, the most prevalent sequences being SSTS or SGGG. In the case of RTK and SAMS, the sequence is C-5X-C (CPDGVDC).

Overall, the observation of multiple conformations for a cysteine residue, with one such conformation forming a disulfide bridge with a neighboring cysteine, highlights the dynamic nature of the FG loop. We hypothesize a structural mechanism that provides this loop with the capability to undergo conformational changes that would ultimately confer these ancestral receptors a broader scope with regard to antigen recognition diversity.

Polymorphism

An important feature previously described for RTK is its polymorphic nature⁶. Polymorphism in molecules of the immune system leads to variations in their structure and functions among individuals, affecting their immune responses. One of the most well-known examples of polymorphism is within the major histocompatibility complex (MHC) genes, found in more evolved forms of life such as vertebrates, including birds, reptiles, and fish. Polymorphisms within MHC genes result in a diverse array of MHC molecules, allowing individuals to recognize and respond to a wide range of pathogens. The presence of polymorphism in proteins of Porifera, the most

ancient phylum of the Metazoa phylogenetically, along with the lack of the cellular machinery for somatic recombination and gene diversity, highlights a potential evolutionary strategy that is not exclusive to more primitive organisms. To explore a potential interplay between polymorphism and location in RTK, we located the polymorphic positions in the RTK structure (Fig. 4). However, we found that polymorphisms are not confined to particular regions. These are instead widely distributed in RTK, with equal frequencies in both N- and C-terminal Ig domains. Polymorphic sites are equally found in loops and in more rigid areas such as the largely populated β sheet strands.

Thus, no evident clues can be inferred that could attribute a particular region in these Ig domains a leading functional role in terms of antigen recognition or cell-cell adhesion mechanisms. Perhaps, the distribution of polymorphic sites in Ig molecules have evolved from a primitive, broad distribution towards a more antigen-specific and topologically concentrated configuration, as occurs in the hypervariable CDRs of the more evolved antibodies and TCRs Ig proteins. In conclusion, the polymorphism observed in proteins of ancient forms of life likely represents allelic polymorphism, and the nature of selection can be inferred by comparing mutation rates across neighboring genes.

Striking loop and PTM (N-glycosylation)

The C'D loop drew our attention due to its length and conformation, folded upon itself and packed against the N-terminal Ig-body (Supplementary Fig. 8). The presence of several solvent-exposed and polar residues points to a potential functional role, such as antigen recognition or cell-cell adhesion. Interestingly, three conserved residues, Arg205 and Arg216 in the C'D loop and Arg236 in the EF loop, create a pocket with a net positive charge (Supplementary Fig. 9). Another interesting observation in this region is the presence of sugar moiety in the form of N-glycosylation at Asn206. Nevertheless, a heavier N-glycosylation pattern is observed in the C-C' segment of RTK, with a branched sugar chain common in eukaryotic proteins (Supplementary Fig. 10). Altogether, these findings suggest potential roles associated with antigen recognition or cell-cell adhesion mechanisms³⁴.

Methods

Gene synthesis, cloning, and plasmid purification

Codon-optimized sequences for sponge RTK (with position 129 being the N-terminal native amino acid) and SAML (with position 157 being the N-terminal native amino acid) containing a N-terminal TwinStrep tag followed by human rhinovirus 3C protease site were synthesized and inserted in a pUC57-BsaI-Free vector by Gene Universal (Newark, DE, USA). Target sequences were flanked by BamHI and NotI restriction sites. The sponge coding regions were obtained by digestion with FastDigest Restriction Enzymes (Fisher Scientific) and cloned into the pAcGP67A vector using Optizyme™ T4 DNA Ligase (Thermo Fisher Scientific). The target genes were inserted downstream of the gp67 secretion signal sequence included in the pAcGP67A baculovirus transfer vector to facilitate the isolation of the recombinant protein from the supernatant. *E.coli* DH5 α (Invitrogen) were transformed with the plasmids obtained after the ligation and were incubated overnight at 37 °C. PCR-positive colonies were grown in LB Broth (Lennox) medium overnight at 37 °C in a shaking incubator. Plasmids were obtained using GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) following the manufacturer's instructions. Plasmids were purified and stored at –20 °C prior to transduction of Sf9 insect cells.

Generation of P0 and P1 baculovirus

Sequences were validated by Sanger sequencing. Then, 2×10^6 Sf9 insect cells (Gibco) were transduced with 500 ng of each transfer plasmid, 100 ng BestBac 2.0 Δ v-cath/chiA Linearized Baculovirus DNA (Chimigen), and 1.2 μ l TransIT (Mirus Bio) to produce final recombinant baculovirus. First, recombinant plasmid, BestBac and TransIT were mixed in 100 μ l Sf-900 III SFM (Thermo Fisher) and incubated for 20 min at RT. Once the cells were observed to be attached to the well, we added 1.8 ml of Sf-900 III SFM

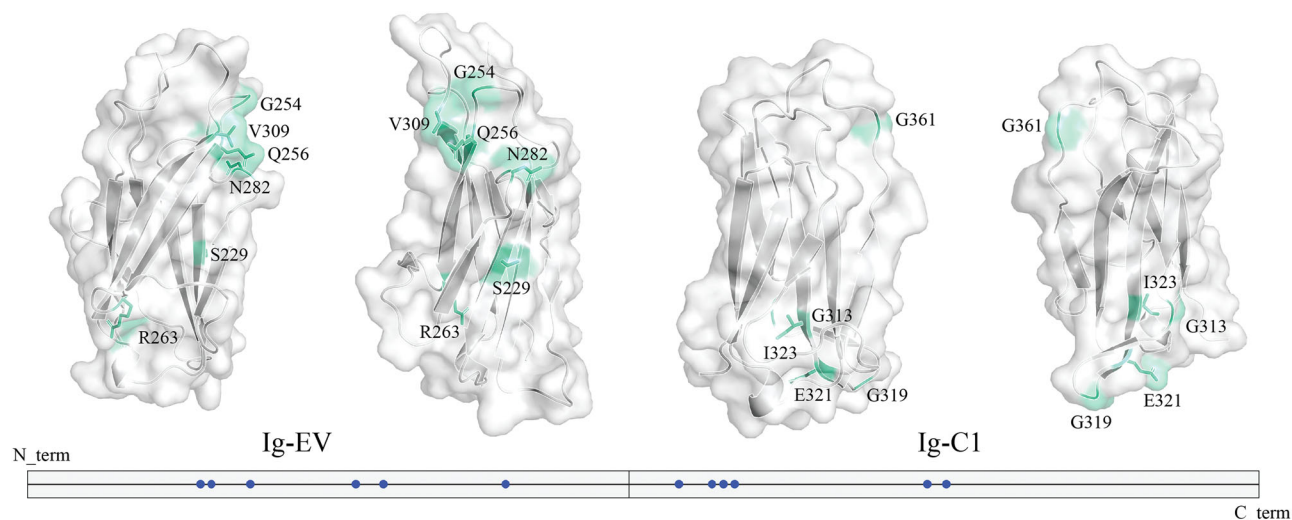


Fig. 4 | Location of polymorphic sites. A cartoon and surface representation are shown for SAML Ig-EV and RTK Ig-C1 domains, with the polymorphic sites highlighted in greencyan color. The diagram indicates the location of the polymorphic residues (blue circles) along the whole ectodomain of RTK. Those regions

that could not be built for RTK due to poor or negligible electron density signal have been replaced by their equivalents in SAML. The numbering of residues is indicated according to the PDB (8OVQ for SAML and 8QPX for RTK).

without fetal bovine serum (FBS) nor Penicillin/Streptomycin. The transduction mix was added, and the cells were incubated for 5 h at 28 °C. Then, the cells were supplemented with 100 μ l of FBS and Penicillin/Streptomycin to 0.25%. Cells were incubated for 5 days at 28 °C. After that, the initial baculovirus sample was collected from the supernatant after centrifugation (300 \times g, 10 min, 4 °C) and stored at 4 °C. In order to amplify virus titers and volume, Sf9 cells (25 ml) in Sf-900 III SFM supplemented with 10% FBS and 0.25% Penicillin/Streptomycin, at a density of 1×10^6 cells/ml were infected with 25 μ l of the initial baculovirus sample. Both cell morphology and proliferation were observed daily. Forty-eight hours after infection, the cells lost their ability to proliferate, and their overall morphology shifted to a swollen appearance. Then, cells were left for another 24 h on an orbital oscillator at 28 °C. Supernatants, containing enriched baculovirus, were collected by centrifugation (3000 rpm, 15 min, 4 °C) and stored at 4 °C.

Recombinant protein expression and purification

Sf9 insect cells at a density of 2×10^6 cells/ml in Xpress medium were infected with TwinStrep-tagged recombinant protein P1 baculoviruses at a 1:2000 ratio, and were incubated for 72 h at 28 °C in a shaker incubator. Supernatant-containing protein was collected by centrifugation (10,000 \times g, 15 min, 4 °C), and supplemented with 100 mM HEPES, 150 mM NaCl, and 1 mM EDTA, pH 7.4. Recombinant proteins were purified from the supernatant with a StrepTactin 4Flow 5 ml cartridge (Iba Lifesciences) following the manufacturer's buffers and instructions. Elution was concentrated in a 10 kDa 4 ml amicon (Merck), and TwinStrep-tagged protein was digested with 3C protease overnight at 4 °C at a 1:50 w/w ratio. 3C protease-digested protein was purified by size-exclusion chromatography using a Superdex75 10/300 GL (Cytiva) in Tris-buffered saline (TBS) pH 7.4. Peak belonging to 3C-digested protein was concentrated using an amicon of 10 kDa and 4 ml. 3C protease, which contained a 6xHis tag, was removed from the sample with Nickel NTA agarose resin (ABT). Purified Ig-like proteins were concentrated using 10 kDa Nanosep columns (Pall Corporation), aliquoted, and frozen in liquid nitrogen prior to storage at -80 °C.

SEC-MALS

Protein samples were characterized using a Superdex75 10/300 Increase column (Cytiva) coupled to a Dawn Heleos II device (Wyatt Technologies). Analyses were performed at 25 °C, in 20 mM HEPES pH 7.4, 150 mM NaCl buffer at a 0.5 mL/min flow rate. Bovine serum albumin (BSA) was used as

analytical standard. Data were processed and analysed using ASTRA 7 software (Wyatt Technologies).

Crystallization of sponge Ig-like proteins

RTK was concentrated at 4.3 μ g/ μ l and SAML at 6.5 μ g/ μ l. Each protein was screened against more than 700 crystallization conditions by the sitting-drop vapor diffusion method. RTK showed the best crystals in 0.2 M ammonium sulfate, 26% w/v PEG4000. The best crystals of SAML were obtained in 0.2 M ammonium sulfate, 30% w/v PEG4000. Crystals were cryoprotected with increasing concentrations of glycerol or ethylene glycol up to 20% in the mother liquor before being cryo-cooled in liquid nitrogen prior to diffraction data analysis.

Data collection and refinement

Crystals were diffracted at the BL13-XALOC beamline of the ALBA Synchrotron (Cerdanyola del Vallès, Barcelona, Spain). Collected datasets (wavelengths 0.97926 Å) of RTK and SAML were processed with autoPROC³⁵ and merged and scaled with AIMLESS³⁶. SAML was solved using molecular replacement with Phaser MR³⁷, using as templates individual Ig-domains from the AlphaFold predictions (AF-Q9U965-F1), which were obtained from Uniprot. The RTK structure was obtained via molecular replacement using as a template the previously solved SAML structure. A restrained refinement was carried out using phenix.refine³⁸ or refmac5³⁹. The final atomic coordinates were obtained after successive rounds of refinement and manual building in Coot⁴⁰. SAML has 98% of residues in the Ramachandran favored region, 2.0% in the allowed region, and 0.0% as outliers. RTK has 96.83% of residues in the favored region, 1.59% in the allowed region, and 1.59% as outliers.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Atomic coordinates and structure factors for SAML and RTK have been deposited in the Protein Data Bank under the accession codes 8OVQ and 8QPX, respectively.

Received: 4 October 2024; Accepted: 26 February 2025;

Published online: 20 March 2025

References

- Lesk, A. M. & Chothia, C. Evolution of proteins formed by beta-sheets. II. The core of the immunoglobulin domains. *J. Mol. Biol.* **160**, 325–342 (1982).
- Oreste, U., Ametrano, A. & Coscia, M. R. On origin and evolution of the antibody molecule. *Biology* **10**, 140 (2021).
- Mü, W. E. G., Kruse, M., Blumbach, B., Skorokhod, A. & Mü, I. M. Gene structure and function of tyrosine kinases in the marine sponge *Geodia cydonium*: autapomorphic characters of Metazoa. *Gene* **238**, 179–193 (1999).
- Müller, W. E., Blumbach, B. & Müller, I. M. Evolution of the innate and adaptive immune systems: relationships between potential immune molecules in the lowest metazoan phylum (Porifera) and those in vertebrates. *Transplantation* **68**, 1215–1227 (1999).
- Blumbach, B. et al. Cloning and expression of new receptors belonging to the immunoglobulin superfamily from the marine sponge *Geodia cydonium*. *Immunogenetics* **49**, 751–763 (1999).
- Pancer, Z. et al. Polymorphism in the immunoglobulin-like domains of the receptor tyrosine kinase from the sponge *Geodia cydonium*. *Cell Adhes. Commun.* **4**, 327–339 (1996).
- Thompson, K. E., Wang, Y., Madej, T. & Bryant, S. H. Improving protein structure similarity searches using domain boundaries based on conserved sequence information. *BMC Struct. Biol.* **9**, 33 (2009).
- Madej, T. et al. MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Res.* **42**, D297–D303 (2014).
- Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).
- Youkharibache, P. et al. The small β -barrel domain: a survey-based structural analysis. *Structure* **27**, 6–26 (2019).
- Ioerger, T. R., Du, C. & Linthicum D. S. Conservation of cys-cys trp structural triads and their geometry in the protein domains of immunoglobulin superfamily members. *Mol. Immunol.* **36**, 373–386 (1999).
- Harpaz, Y. & Chothia, C. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol.* **238**, 528–539 (1994).
- Feige, M. J. et al. The structural analysis of shark IgNAR antibodies reveals evolutionary principles of immunoglobulins. *Proc. Natl Acad. Sci. USA* **111**, 8155–8160 (2014).
- Youkharibache, P. Topological and structural plasticity of the single Ig fold and the double Ig fold present in CD19. *Biomolecules* **11**, 1290 (2021).
- Almagro, J. C. et al. Characterization of a high-affinity human antibody with a disulfide bridge in the third complementarity-determining region of the heavy chain. *J. Mol. Recognit.* **25**, 125–135 (2012).
- Richardson, J. S. & Richardson, D. C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl Acad. Sci. USA* **99**, 2754–2759 (2002).
- Benian, G. M. & Mayans, O. Titin and obscurin: giants holding hands and discovery of a new Ig domain subset. *J. Mol. Biol.* **427**, 707–714 (2015).
- Wang, J.-H. The sequence signature of an Ig-fold. *Protein Cell* **4**, 569–572 (2013).
- Buljan, M. & Bateman, A. The evolution of protein domain families. *Biochem. Soc. Trans.* **37**, 751–755 (2009).
- Schluter, S. F., Bernstein, R. M., Bernstein, H. & Marchalonis, J. J. ‘Big Bang’ emergence of the combinatorial immune system. *Dev. Comp. Immunol.* **23**, 107–111 (1999).
- Flajnik, M. F. A cold-blooded view of adaptive immunity. *Nat. Rev. Immunol.* **18**, 438–453 (2018).
- Dzik, J. M. The ancestry and cumulative evolution of immune reactions. *Acta Biochim. Pol.* **57**, 443–466 (2010).
- Kumar, N., Arthur, C. P., Ciferri, C. & Matsumoto, M. L. Structure of the human secretory immunoglobulin M core. *Structure* **29**, 564–571.e3 (2021).
- Lyu, M., Malyutin, A. G. & Stadtmueller, B. M. The structure of the teleost immunoglobulin M core provides insights on polymeric antibody evolution, assembly, and function. *Nat. Commun.* **14**, 7583 (2023).
- Williams, A. F., Davis, S. J., He, Q. & Barclay, A. N. Structural diversity in domains of the immunoglobulin superfamily. *Cold Spring Harb. Symp. Quant. Biol.* **54**, 637–647 (1989).
- Tawfeeq, C. et al. Towards a structural and functional analysis of the immunoglobulin-fold proteome. *Adv. Protein Chem. Struct. Biol.* **138**, 135–178 (2024).
- Du Pasquier, L. & Chrétien, I. Why is CTX all the RAGE? *Res. Immunol.* **147**, 261–266 (1996).
- Burke, K. P., Chaudhri, A., Freeman, G. J. & Sharpe, A. H. The B7:CD28 family and friends: unraveling coinhibitory interactions. *Immunity* **57**, 223–244 (2024).
- Mühlebach, M. D. et al. Adherens junction protein nectin-4 is the epithelial receptor for measles virus. *Nature* **480**, 530–533 (2011).
- Prabakaran, P. & Chowdhury, P. S. Landscape of non-canonical cysteines in human V(H) repertoire revealed by immunogenetic analysis. *Cell Rep.* **31**, 107831 (2020).
- Mariuzza, R. A., Shahid, S. & Karade, S. S. The immune checkpoint receptor LAG3: structure, function, and target for cancer immunotherapy. *J. Biol. Chem.* **300**, 107241 (2024).
- Ming, Q. et al. LAG3 ectodomain structure reveals functional interfaces for ligand and antibody recognition. *Nat. Immunol.* **23**, 1031–1041 (2022).
- Palakodeti, A. et al. The molecular basis for modulation of human V γ 9V δ 2 T cell responses by CD277/butyrophilin-3 (BTN3A)-specific antibodies. *J. Biol. Chem.* **287**, 32780–32790 (2012).
- Gu, J. et al. Potential roles of N-glycosylation in cell adhesion. *Glycoconj. J.* **29**, 599–607 (2012).
- Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302 (2011).
- Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
- Tawfeeq, C. et al. IgStrand: a universal residue numbering scheme for the immunoglobulin-fold (Ig-fold) to study Ig-Proteomes and Ig-Interactomes. *PLoS Comput. Biol.* <https://doi.org/10.1371/journal.pcbi.1012813> (2025). (In press).

Acknowledgements

We are grateful to Gilda Dichiara Rodríguez and Sergio Morales Hernández for their technical support. We also thank the staff of the Xaloc beamline at ALBA Synchrotron for their assistance with X-ray diffraction data collection. This work was supported by a Ramón y Cajal Grant RYC-2017-21683, Ministry of Science and Innovation, Government of Spain, to Jacinto López-Sagasetta. Alejandro Urdiciain is a recipient of a Margarita Salas contract funded by UPNA and the Ministry of Universities of Spain within the Plan of Recovery, Transformation and Resilience and the European Recovery Instrument Next Generation EU. T.M., J.S., and J.W. are supported by the

Intramural Research Program of the National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information of the USA.

Author contributions

Conceived research project: J.L.-S.; Performed experiments: A.U., J.L.-S., and P.Y.; Data analysis: A.U., J.L.-S., J.W., J.S., P.Y., and E.E.; Draft writing: J.L.-S. and P.Y.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-07830-5>.

Correspondence and requests for materials should be addressed to Philippe Youkharibache or Jacinto López-Sagaseta.

Peer review information *Communications Biology* thanks Larry J. Dishaw, Orlando Bruno Giorgetti, and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: Huan Bao and Laura Rodríguez Pérez. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025