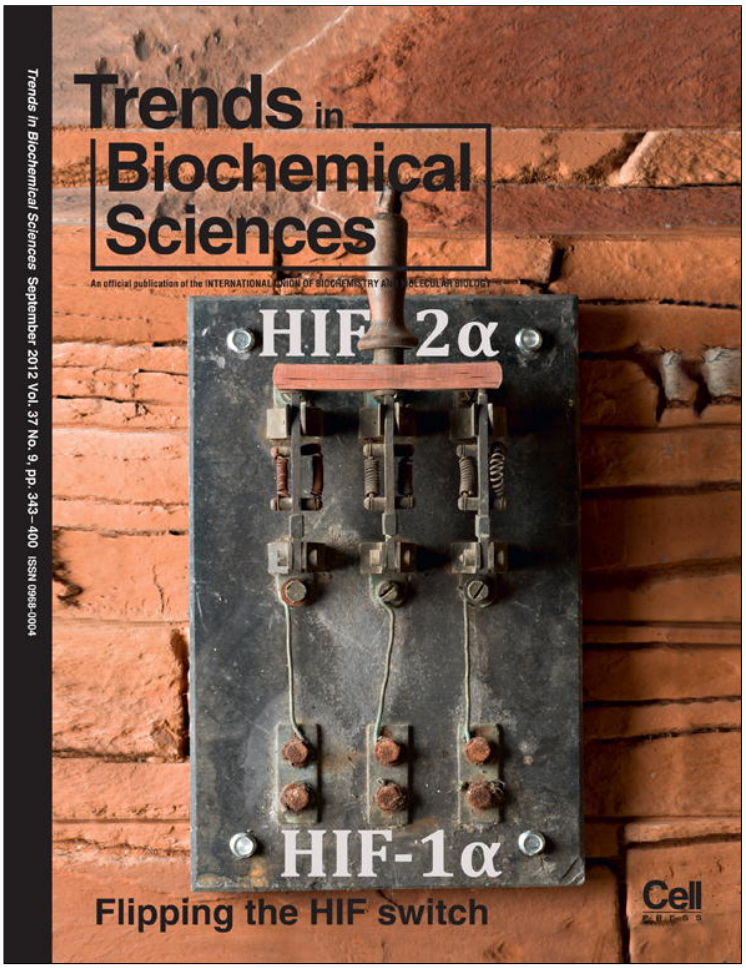

Review

Cell PRESS

An improved understanding of TNFL/TNFR interactions using structure-based classifications

Cedrik Magis[1], Almer M. van der Sloot[2], Luis Serrano[2,3] and Cedric Notredame[1]

[1] Bioinformatics and Genomics Program, Centre for Genomic Regulation (CRG) and UPF, Barcelona, Spain
[2] EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), UPF, Barcelona, Spain
[3] Institucio Catalana de Recerca i Estudis Avancats (ICREA), Pg. Lluís Companys 23, 08019 Barcelona, Spain

Tumor Necrosis Factor Ligand (TNFL)–Tumor Necrosis Factor Receptor (TNFR) interactions control key cellular processes; however, the molecular basis of the specificity of these interactions remains poorly understood. Using the T-RMSD (tree based on root mean square deviation), a newly developed structure-based sequence clustering method, we have re-analyzed the available structural data to re-interpret the interactions between TNFLs and TNFRs. This improves the classification of both TNFLs and TNFRs, such that the new groups defined here are in much stronger agreement with structural and functional features than existing schemes. Our clustering approach also identifies traces of a convergent evolutionary process for TNFLs and TNFRs, leading us to propose the co-evolution of TNFLs and the third cysteine rich domain (CRD) of large TNFRs.

TNFL and TNFR families are key players in cellular regulation

The TNFLs are an important family of cytokines involved in the regulation of key cellular processes such as differentiation, proliferation, apoptosis and cell growth [1–4]. TNFLs mediate their functions through specific interactions with members of the TNFR family. Dysregulation of these pathways has been shown to result in a wide range of pathological conditions, including cancer, autoimmune diseases, inflammation and viral infection [4–6]. This explains the importance of these genes as potential drug targets [6,7]. The diverse functions of this family are mediated through a large number of ligands and receptors: there are 18 known TNFL genes and 29 different TNFRs in humans. Although it is known that some TNFL/TNFR interactions are mutually exclusive, cross-interactions have been reported in a majority of cases [8]. A precise understanding of TNFL/TNFR interaction modes would be a precious asset for diagnostic and drug development purposes. Unfortunately, the low level of identity between functional domains (less than 30% identity in both the ligand and the receptor families) has made this goal unreachable so far. In this review, we propose a new classification system, developed through a better usage of the available structural and functional information. We show that this scheme is able to recapitulate most known functions of these two families.

The TNFL family is highly structurally conserved

In humans, 18 different TNFL genes have been identified, although more transcripts exist as a result of alternative splicing events [9–11]. All family members are type II transmembrane proteins (single pass transmembrane proteins with the N terminus on the cytoplasmic side of the membrane), although only 12 are known to be secreted upon proteolytic cleavage by metalloproteases or furin proteases [12]. These 18 genes appear to be homologous, even though the N terminal cytoplasmic and transmembrane domains are less conserved than the extracellular domain, whose identity level is close to 30%. The last 150 residues of the C terminal extracellular domain constitute the TNF homology domain (THD), characterized by a β jelly roll fold made of 10 β strands (Figure 1a), which form an inner (A, A′, C, F and H) and outer β sheet (B, B', D, E and G). Although the β strand sequences are relatively well conserved, the loops connecting them are highly variable in size and composition. This fold is conserved across the entire TNFL family, including the distantly related C1q family [13] known to be involved in immunity related pathways [14]. To interact with their cognate receptors, THDs assemble as a non-covalent trimer through a process mainly involving the inner β sheet.

The TNFR family: variable architectures with a conserved binding domain

The receptors of the TNFLs are named TNFRs. The 29 family members reported in humans are very diverse: 22 have a clear signal peptide and are classified as type I transmembrane receptors (single pass transmembrane proteins with the N terminus targeted to the extracellular side of the cell membrane); the 7 remaining members are either classified as type III transmembrane receptors (single pass transmembrane proteins with the N terminus on the extracellular side of the membrane and no signal sequence; TNFR13B, TNFR13C, TNFR17, XEDAR), attached to the membrane via a glycosylphosphatidylinositol (GPI) linker (TNFR10C), or secreted as soluble receptors (TNFR11B and TNFR6B) [9]. The only feature common to all TNFRs is the presence of relatively short

Corresponding author: Notredame, C. (cedric.notredame@crg.eu).
Keywords: TNFL; TNFR; structure-based classification; MSA; evolution.

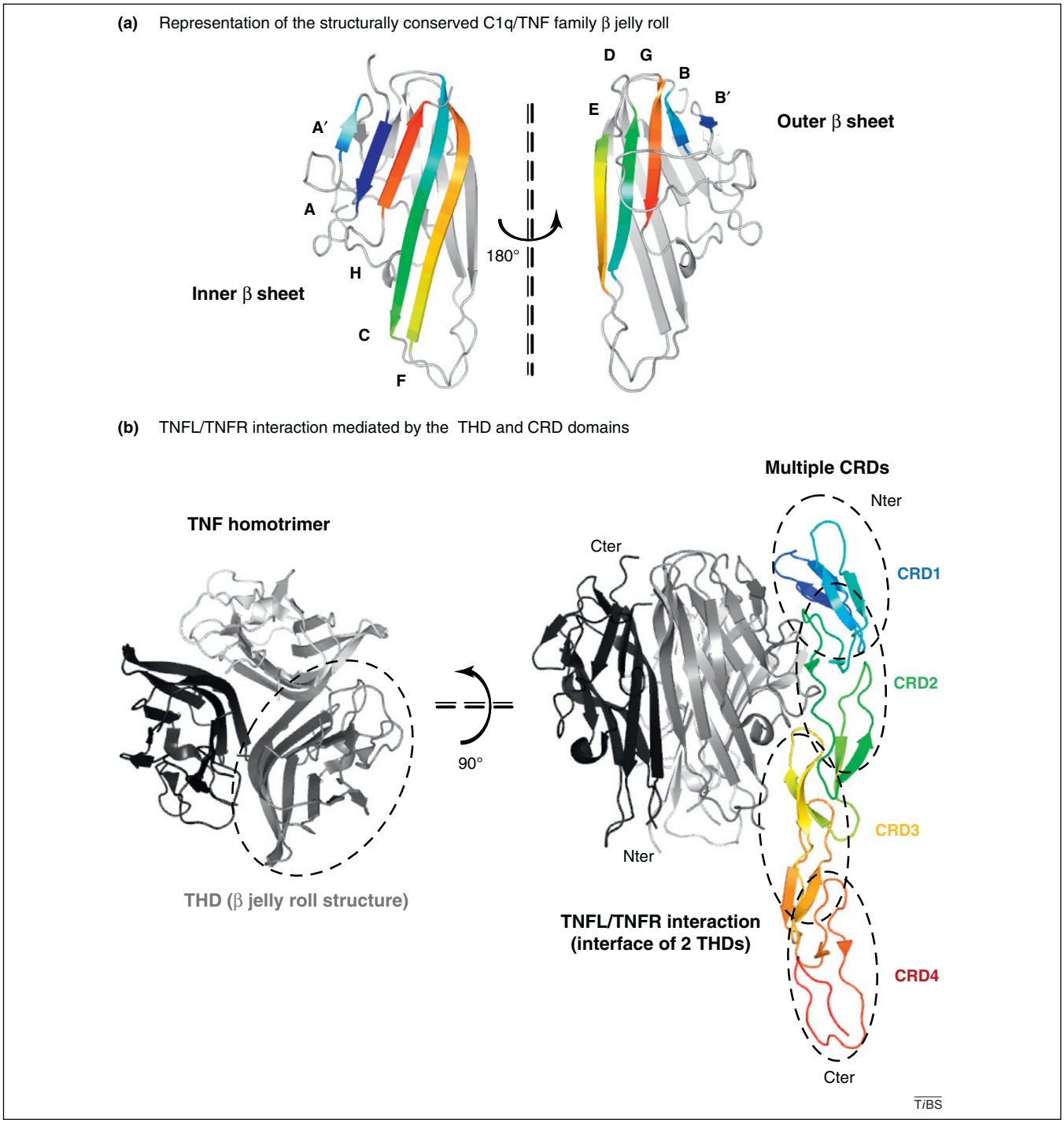

Figure 1. The Tumor Necrosis Factor Ligand (TNFL) and Tumor Necrosis Factor Receptor (TNFR). **(a)** Ribbon diagram representation of the conserved TNF homology domain (THD)/C1q β jelly roll fold, illustrated by TNFL10 (PDB ID: 1D4 V [24]). The five β strands forming the inner sheet (A, A′, C, H and F) are shown on the left, and the outer β sheet (B, B′, D, E and G) is shown on the right; the rainbow coloration indicates the front of the model. **(b)** Ribbon diagram representation of the TNFL/TNFR interaction (TNF2/TNFR2 crystallographic structure, PDB ID: 3ALQ [26]). The non-covalent TNFL trimer is shown on the left with one of the THD units delimited by a black broken circle. Interaction is shown on the right. Starting from the N terminus, the cysteine rich domains (CRDs) of the TNFR are labeled in blue (CRD1), green (CRD2), yellow (CRD3) and red (CRD4). Each unit is circled with a broken line. Visualization is done with Pymol (Version 1.1r1, Schrödinger, LLC).

(30–40 residues) CRDs located in the ectodomain, which are involved in the THD interaction [15]. TNFRs often have multiple CRDs (between one and six), with three or four copies being the most frequent configurations. It is common practice to name CRDs with respect to their position starting from the receptor N terminus (i.e., CRD1 for the N terminal CRD, CRD2 for the next one, and so on). As suggested by their name, CRDs are enriched in cysteine, a consequence of the high density disulfide bridge network that constrains their structure. Most CRDs possess three disulfide bridges, although they can have between one and four. Slight variations in cysteine pattern and connectivity have been used to derive the existing classifications [9,12,15,16], even though it

remains unclear whether these differences have functional implications.

The TNFL/TNFR interaction is a highly cooperative process

The TNFL/TNFR interaction requires the cooperation of three units of both receptors and ligands. Each individual binding interface is formed by two of the three THD units, associated as non-covalent trimer; this allows each THD trimer to bind up to three TNFRs as illustrated in Figure 1b. Assembly of these 3:3 complexes appears to be compulsory for signal transduction to occur [4], and detailed analyses have shown that many TNFRs are pre-clustered as dimers at the cell surface [17]. This process, known as pre-ligand assembly, occurs in the absence of the cognate ligand. The formation of a pre-complex appears to be a common process for TNFRs containing multiple CRDs. It has so far been experimentally validated for several receptors: TNFR1, TNFR2, TNFR5, TNFR6, TNFR10A, TNFR10B, TNFR10D [18,19], viral TNFR [20] and non-TNF family receptors such as the interleukin receptor [21]. This clustering occurs through interactions between the pre-ligand assembly domain (PLAD) [18,19,22], a specific domain that is involved in oligomerization but does not directly contribute to ligand binding. In the particular case of the TNFRs, the pre-ligand assembly activity is carried out by the N terminal domain (CRD1). In addition to increasing the local concentration of the receptors, PLADs also contribute to the stability of the nearby domain (CRD2) [23], which establishes direct contacts with the ligand [24–29]. Pre-complex formation mediated by PLADs is not, however, a universal requirement and no such mechanism has been reported in TNFRs that have only one (TNFR12, TNFR13C and TNFR17) or two CRDs. In TNFR13B (two CRDs), for instance, the second domain alone is sufficient for ligand-induced cell signaling [30].

TNFL/TNFR interactions can be mutually exclusive, even though in a majority of cases some cross-interactions have been reported [8]. When studying the first TNF1/TNFR1 [25] and TNF10/TNFR10B [31,32] crystallographic complexes, the authors reported specific patches on the receptor side, involving the second and the third CRDs (CRD2 and CRD3). These patches appear to play a role in modulating binding affinity (CRD2) and specificity (CRD3) [9,25,27,31,32]. So far, predicting specific interactions on the basis of structural information alone has remained impossible and most analysis efforts have focused on deriving an exhaustive classification scheme [9,12,15,16]. None of these classifications, however, manages to recapitulate all known functional similarities. The results we present here show how fine grain structure-based classifications can be used to derive a more informative classification scheme and improve the description of the interaction between THDs and their cognate receptors. Structure-based classifications are usually more accurate than sequence-based classifications, mostly because of the strongest resilience of structural signal over the underlying sequence variations.

T-RMSD structural classification strategy

Structure-based classifications were derived by processing T-Coffee structure-based multiple sequence alignments (MSAs) with the T-RMSD structure-based classification algorithm. To produce structure-based MSAs, we have used the 3D-Coffee mode [33,34]. Structure-based MSAs are known to be more accurate than their sequence-based counterparts, and the main goal of this initial step was to determine structurally equivalent residues (i.e., to put residues playing the same role in their respective structures in the same alignment column). The MSA model and the associated structures were then used as an input for the T-RMSD method. The T-RMSD uses a comparison of 3D intramolecular distances (similar to the Dali algorithm [35]), across the aligned sequences to estimate a structure-based clustering that groups the sequences on the basis of their structural similarity. Because structural features are known to be slow evolving, it is expected that structure-based comparisons will be more informative than their sequence-based counterparts. In the T-RMSD process, differences of distances are only considered between ungapped columns. For each such column, the differences are aggregated into a distance matrix (i.e., a matrix estimating the structural similarity between every pair of sequence). One such matrix has been produced for each column, each matrix is then resolved into a neighbor-joining tree and the resulting collection is combined into a unique consensus cluster. In this final cluster, every split (node) has a number indicating the fraction of columns (in the original MSA) that support it. This measure can be used as a reliability indicator for the definition of categories.

Structural classification of the THDs

We first applied the T-RMSD clustering strategy to the classification of the THD and C1q families. An initial dataset was compiled by combining all of the THD and C1q sequences reported in Pfam, SMART and Prosite [36–38] (Table S1 in the supplementary material online), which was complicated by inconsistencies across domain databases, and incompatibility between the structural and sequence definition of THDs. For instance, the Pfam model does not include several β strands (A, A′ and B′), whereas the SMART and Prosite models incorporate all of the structural elements but fail to recognize two experimentally characterized THDs (i.e., TNF4 and TNF18 [29,39]). Furthermore, all the datasets have clear boundary inconsistencies compared to manually curated datasets. Considering the functional importance of the N and C terminal regions, this issue is crucial for accurate classification. All selected sequences were manually trimmed, to guarantee structural integrity (Table 1), and a structure-based MSA was estimated with T-Coffee [34,40]. As expected, all structural features appear to be conserved (Figure 2a). Our MSA model also suggests two standalone structural elements to be conserved in almost every structure (except for TNF18). We named these two elements ‘a’ and ‘g’, located between the A–A′ and G–H β strands, respectively (Figure 2a). Although ‘a’ and ‘g’ are not part of the jelly roll fold, they most likely play a role in its packing and might also be involved in shaping interaction specificity with the cognate ligand.

This MSA was used to derive the structure-based clustering [41] (Figure 2b). The resulting classification clearly

Table 1. Nomenclature and structural information for proteins comprising the THD and CRD datasets

Names			Structural details				
Current name	**HGNC name**	**Protein name**	**PDB**	**Res. Å**	**Chain**	**Species**	**Refs**
Human TNFL members with available structure							
TNF1	LTA	LT-α	1TNR	2.85	A	*Homo sapiens*	[25]
TNF2	TNF	TNF-α	1TNF	2.60	A	*Homo sapiens*	[53]
TNF4	TNFSF4	OX40L	2HEV	2.41	F	*Homo sapiens*	[29]
TNF5	CD40LG	CD40L	1ALY	2.00	A	*Homo sapiens*	[54]
TNF10	TNFSF10	TRAIL	1D4V	2.20	B	*Homo sapiens*	[24]
TNF11	TNFSF11	RANKL	1S55	1.90	A	*Mus musculus*	n/a
TNF13	TNFSF13	APRIL	1XU1	1.90	A	*Mus musculus*	[30]
TNF13B	TNFSF13B	BAFF	1OQE	2.50	A	*Homo sapiens*	[55]
TNF15	TNFSF15	VEGI	2RE9	2.10	A	*Homo sapiens*	[56]
TNF18	TNFSF18	GITRL	2Q1M	2.30	A	*Homo sapiens*	[39]
EDA-A1	EDA	EDA variant 1	1RJ7	2.30	A	*Homo sapiens*	[11]
EDA-A2	EDA	EDA variant 2	1RJ8	2.20	A	*Homo sapiens*	[11]
Human C1q members with available structure							
ADIPO	ADIPOQ	ACRP30	1C3H	2.10	A	*Mus musculus*	n/a
C1QA	C1QA	Complement C1q subunit A	1PK6	1.85	A	*Homo sapiens*	[57]
C1QB	C1QB	Complement C1q subunit B	1PK6	1.85	B	*Homo sapiens*	[57]
C1QC	C1QC	Complement C1q subunit C	1PK6	1.85	C	*Homo sapiens*	[57]
CO8A1	COL8A1	Collagen α-1 VIII	1O91	1.85	A	*Mus musculus*	[58]
COOA1	COL10A1	Collagen α-1 X	1GR3	2.00	A	*Homo sapiens*	[59]
Human TNFR members with available structure							
TNFR1	TNFRSF1A	TNF-R1	1EXT	1.85	A	*Homo sapiens*	[60]
TNFR2	TNFRSF1B	TNF-R2	3ALQ	3.00	R	*Homo sapiens*	[26]
TNFR4	TNFRSF4	OX40	2HEY	2.00	T	*Homo sapiens*	[29]
TNFR5	CD40	CD40	3QD6	3.50	R	*Homo sapiens*	[46]
TNFR6B	TNFRSF6B	DcR3	3K51	2.45	B	*Homo sapiens*	[27]
TNFR10B	TNFRSF10B	DR5	1D4V	2.20	A	*Homo sapiens*	[23]
TNFR11A	TNFRSF11A	RANK	3ME4	2.01	A	*Mus musculus*	[28]
TNFR12	TNFRSF12A	TWEAK-R	2RPJ	NMR	A	*Homo sapiens*	[61]
TNFR13B	TNFSF13B	TACI	1XU1	1.90	T	*Homo sapiens*	[30]
TNFR13C	TNFRSF13C	BAFF-R	1OQE	2.50	R	*Homo sapiens*	[55]
TNFR14	TNFRSF14	HVEM	1JMA	2.65	B	*Homo sapiens*	[62]
TNFR16	NGFR	NGFR	1SG1	2.40	X	*Rattus norvegicus*	[63]
TNFR17	TNFRSF17	BCMA	1XU2	2.35	T	*Homo sapiens*	[30]
TNFR21	TNFRSF21	DR6	3QO4	2.20	A	*Homo sapiens*	[45]
Other TNFR members with available structure							
CRME	CRME	CRME	2UWI	2.00	A	*Vaccinia virus*	[64]

Current name: name used in this review for each family member; HGNC name: Human Gene Nomenclature Consortium (HGNC) name; Protein name: common name; PDB: PDB identifier of the solved structure; Res Å: structure resolution in Å; Chain: letter indicating the PDB protein chain identifier; Species: the species of the crystallized protein; Refs: bibliographic references for the PDB file used.

sets apart the THD and C1q families, with the corresponding node (branch splitting the two groups) highly supported. TNF4 and TNF18, which SMART and Prosite models fail to recognize as THDs, appear to form a separated subgroup (variable THD, V-THD) halfway between the THD and C1q domain groups. Indeed, further structural inspection shows that these two TNFs have shorter β sheets and are slightly less structurally similar to the other family members. The next well-defined subgroup (small THD, S-THD) in this classification scheme is made of TNF13, TNF13B and EDA (splice variants A1 and A2 [11]) whose main characteristic is shorter β sheets, resulting in a reduced interface. The rest of the THDs define the last subgroup (large THD, L-THD), in which longer C/D and E/F β sheets induce an elongated shape and a wider interface. It is interesting to note that in the case of THDs the structure-based classification differs very little (one misclassified sequence only) from its sequence-based counterpart (Figure S1 in the supplementary material online [42]). Using the methodology described in [41], one can easily turn each THD type into hidden Markov models (HMMs) [43] to classify unannotated sequences with an unknown structure. We used this approach to annotate all the known human THDs by labeling each of them with the subgroup HMM yielding the most significant E-value (Table S2 in the supplementary material online). The final annotation thus obtained corresponds perfectly to the previous division of TNFLs into three classes: conventional ligands (L-THD), EF-disulfide ligands (S-THD) and the divergent ligands (V-THD) [29,44]. This classification is also supported by experimental evidence, most notably the observation that in all experimentally validated interactions between TNFLs and TNFRs [45–49], cross-interactions always occur either within the L-THD or the within S-THD subgroup defined here.

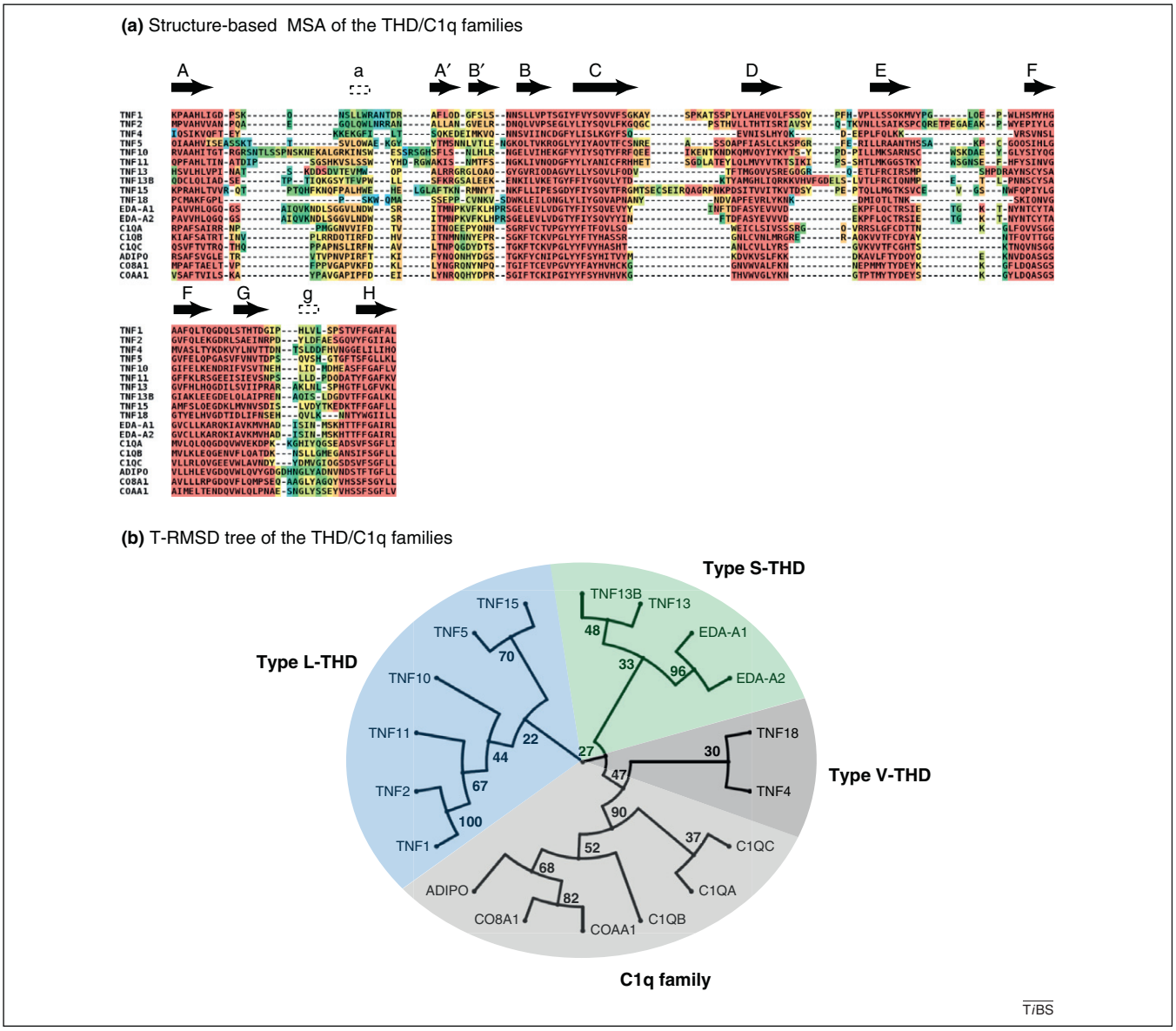

Figure 2. Structure-based analysis of the TNF homology domain (THD)/C1q families. **(a)** T-Coffee structure-based multiple sequence alignment (MSA) of the THD/C1q families. The color scheme is the T-Coffee index for reliability (ranging from red for high confidence to blue for low confidence). Conserved secondary structure elements are represented above the MSA: arrows represent β strands forming the β jelly roll fold (from A to H) and dotted boxes represent non-canonical conserved structural elements (a and g). **(b)** T-RMSD (tree based on root mean square deviation) structure-based clustering of the THD/C1q families. Blue: large interface group (L-THD), light green: small interface group (S-THD), gray: variable interface group (V-THD) and light gray: C1q family. The support values on each node indicates the support (on a 0–100) scale as estimated by T-RMSD. Cluster pictures were generated using PhyloWidget [52].

CRD-based structural classification of the TNFR domains

TNFRs have a much more complex domain organization than their ligands; at first glance, receptors can be divided into two main groups: the large receptors (three CRDs or more) and the short ones (one or two CRDs). Classification attempts, however, can easily be confounded by the CRD repeat pattern. Because CRDs are the main determinant of the ligand-binding specificity of the receptor, classification attempts have naturally focused on them. Comparing CRDs is a fairly complex task owing to their short size (30–40 residues) and low sequence identity level (<30% on average, mainly resulting from cysteine conservation). To assemble a complete dataset (Table S1 in the supplementary material online), we gathered all domains annotated as CRDs in Pfam, SMART and Prosite. Despite highly conserved cysteine patterns (an important feature for their homology detection), CRDs appear to be structurally very diverse. We defined domain boundaries using PDB structural information and compiled a dataset of 41 CRDs with a known structure (Table 1). This dataset was structurally aligned using T-Coffee (Figure 3a). Aside from the cysteine residues involved in the formation of disulfide bridges, most positions appear to be poorly conserved.

We then derived, from this MSA, a structure-based clustering. Our classification (Figure 3b) suggests the existence of at least three well-defined subgroups. The first one, labeled as type S (for small), contains CRDs occurring in small receptors. Structurally speaking, these are the less conserved CRDs and their only major constraints

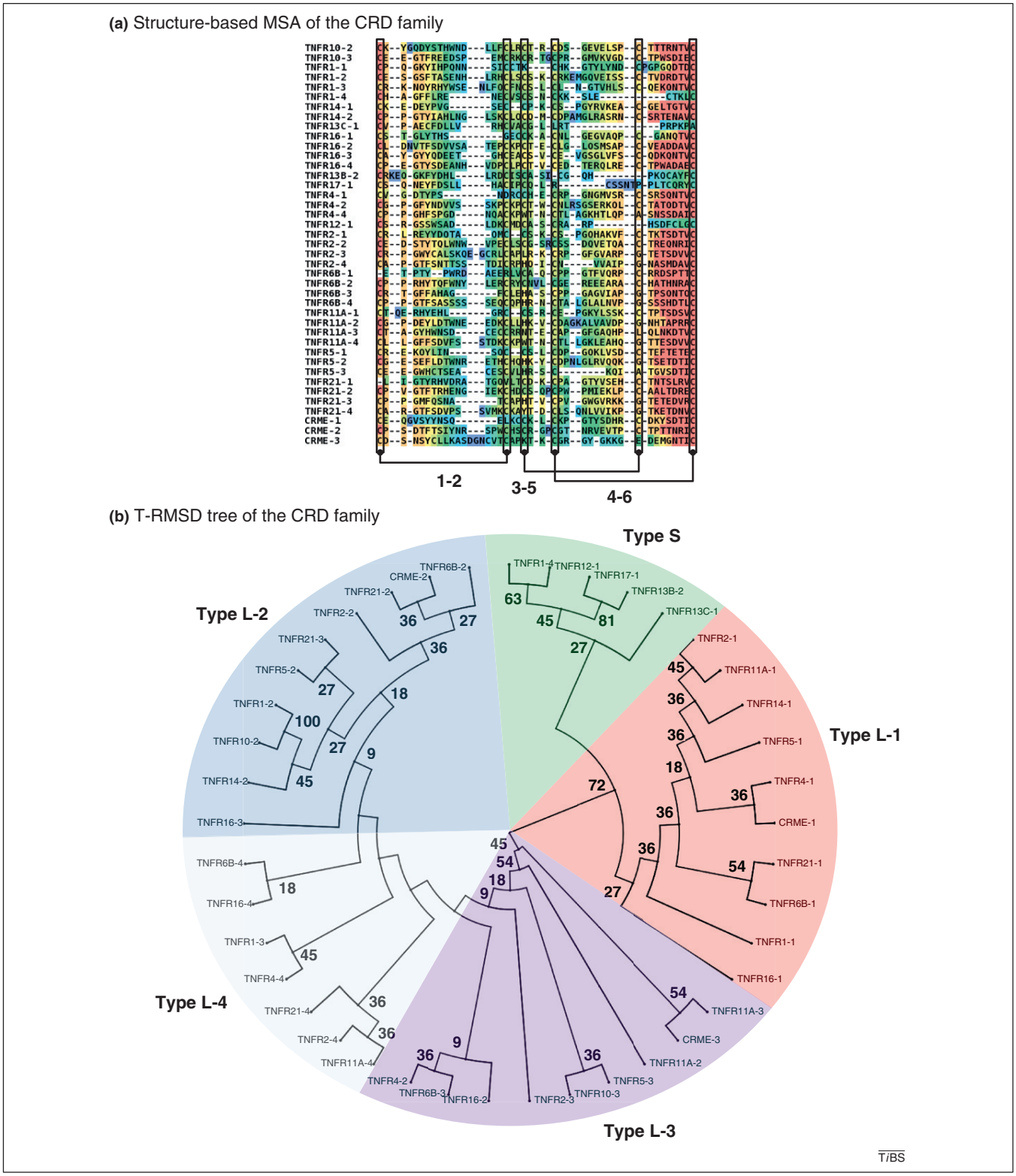

Figure 3. Structure-based analysis of the cysteine rich domain (CRD) family. **(a)** T-Coffee structure-based multiple sequence alignment (MSA) of the CRD family. Color scheme indicates reliability, as described in Figure 2. Black boxes indicate conserved cysteine residues and brackets under the MSA demarcate the disulfide bridges. **(b)** T-RMSD structure-based clustering of the CRD family. Green: type S; red: type L-1; blue: type L-2; purple: type L-3; light blue: type L-4. Node support is the same as described in Figure 2.

appear to be the 1–2 disulfide bridge and the maintenance of a β hairpin that has been shown in several cases (TNFR13B, TNFR13C and TNFR17) to interact directly with the cognate ligands [48]. The second type, labeled as L-1, contains CRDs occurring at the N terminal position of large receptors (CRD1), often associated with pre-ligand assembly activity. In this group, the main feature is the strictly conserved loop between cysteines 4 and 5

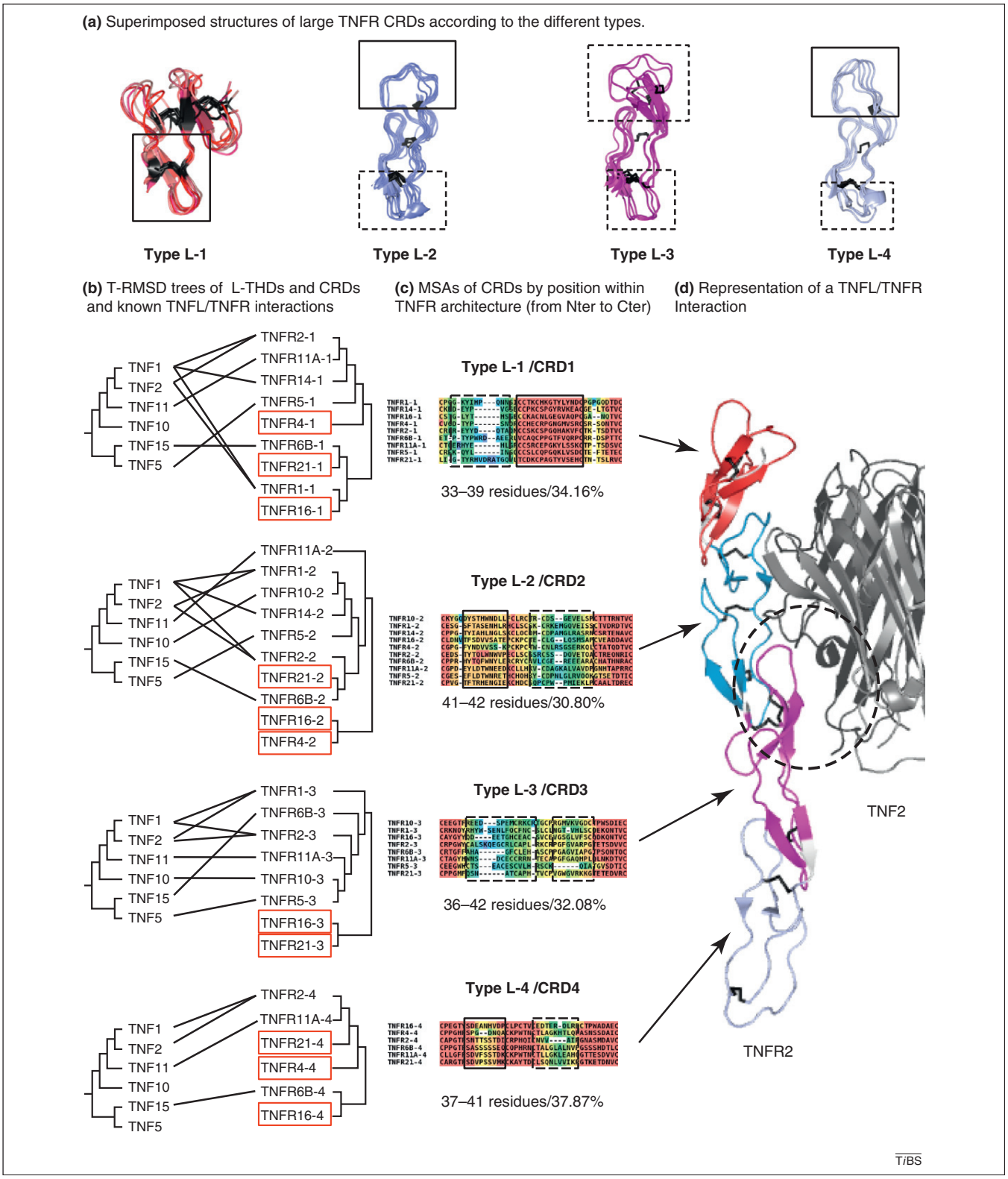

Figure 4. Comparison of large TNF homology domain (L-THD) and cysteine rich domain (CRD) structure-based clustering using interaction data. **(a)** Superposition of the CRDs classified as either L-1, L-2, L-3 or L-4. Unbroken line boxes indicate conserved elements; broken line boxes indicate variable elements. Black sticks depict disulfide bridges. **(b)** T-RMSD (tree based on root mean square deviation) structure-based clustering of the L-THDs compared with L-1 (first), L-2 (second), L-3 (third) or L-4 CRDs. Experimentally observed interactions between the THD and the Tumor Necrosis Factor Receptor (TNFR) containing the corresponding CRD are shown with black lines. Red boxes indicate CRDs belonging to a TNFR with unknown or non L-THDs ligands. **(c)** T-Coffee multiple sequence alignment (MSA) of the four CRD types L-1 (top), L-2 (second), L-3 (third) and L-4 (bottom), using the same color scheme as Figure 2. Broken boxes indicate variable regions; solid lines indicate the non-variable. **(d)** Tumor Necrosis Factor Ligand (TNFL)/TNFR interaction (TNF2/TNFR2, PDB ID: 3ALQ [26]). THDs are colored in black and gray. TNFR is colored according to CRD position, starting from the N terminus, red: CRD1; blue: CRD2; purple: CRD3; light blue: CRD4. The broken circle indicates the interaction interface. All cluster images were generated using PhyloWidget [52] and structure representations with PyMOL (Version 1.1r1, Schrödinger, LLC).

(Figure 4a,c), which is consistent with its known stabilizing effect on the interaction between the PLAD and its immediate neighbor (CRD2). Conversely, the equivalent region varies a lot within the other CRD groups. The third largest group contains core CRDs occurring at various non N terminal positions of the large receptors (CRD2, 3 and 4). This large group can be further divided into three less well-defined subgroups, labeled L-2, L-3 and L-4, respectively, in which CRDs appear to loosely cluster according to their position within the ectodomain architecture. By this

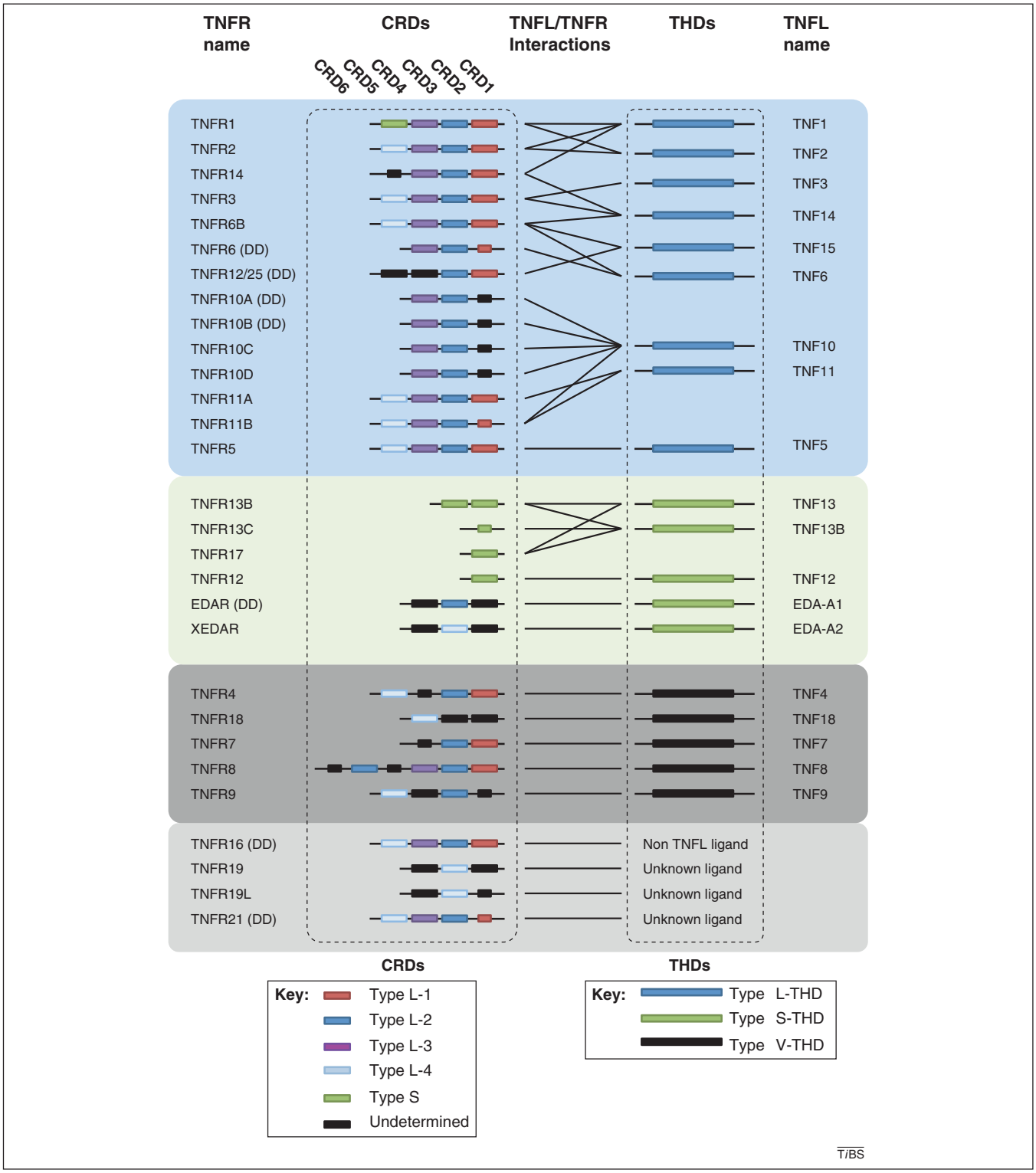

Figure 5. New structural Tumor Necrosis Factor Ligand/Tumor Necrosis Factor Receptor (TNFL/TNFR) annotation. All human TNFRs (left), TNFLs (right) and their known interactions (black lines) are displayed. TNFRs are represented with boxes corresponding to their cysteine rich domain (CRD) architecture. CRDs are single boxes colored according to the new structure-based classification (type L-1: red; type L-2: blue; type L-3: purple; type L-4: light blue; type S: green; undetermined: black). Small boxes indicate truncated domains. TNF homology domains (THDs) are shown as single boxes and colored according to the new structure-based classification (large THD, L-THD: blue; small THD, S-THD: green; variable THD, V-THD: gray). Each member is named according to Table 1; all TNFRs known to be involved in apoptosis are indicated with a (DD).

criterion, L-2 is the best-defined type as it mostly corresponds to CRD2s located immediately next to the PLAD. Type L-2 CRDs are structurally homogeneous, as reflected by the tight cluster in Figure 3b, a finding that is in good agreement with their position, sandwiched between the first and the third CRD. This very special situation (most likely to result in some highly constrained evolution) is materialized at the sequence level by the lack of any insertions or deletions in the loops directly in contact with the PLAD. The two other types (L-3 and L-4) correspond mostly to CRD3 and CRD4, respectively, although in this case the correspondence between the structure-based labeling and the ectodomain architecture is weaker. Structurally speaking, L-3 is the most variable type, hence its tendency to be scattered across the clustering. This observation is in agreement with the highly variable pattern of insertions and deletions between cysteines 1–2 and 4–5 within this domain (Figure 4a, c). Interestingly, structural data indicate that these highly variable regions interact with similarly variable THD loops (located between β strands C/D and E/F). This suggests some role in specificity control, as previously proposed [25,31,32]. One of the most striking observations of this classification is that within each type, cysteine patterns are far from being the most conserved feature among core CRDs (types L2, L-3 and L-4). Furthermore, our analyses indicate that in the optional second bridge (cysteines 3–5) the third cysteine is often replaced by aromatic or hydrophobic residues (H >> W > Y, L) and the fifth cysteine can be substituted with small residues (A or G). It is the first time an automated classification has been reported to be in agreement with receptor architecture. By contrast, this new classification is in slight disagreement with previous approaches [9,12,15,16] that explicitly relied on the comparison of disulfide bridge networks to define the classes.

Using the HMM methodology described earlier, we systematically annotated all of the human TNFR family members without known structures (Figure 5) (Table S3 in the supplementary material online). The results are in broad agreement with our classification scheme, with 78% of core CRDs identified as type L-2, L-3 or L-4 (19.5% corresponding to unidentified and truncated domains) and 84% of CRD2s recognized as type L-2. Likewise, small receptors are always recognized as type S. Since we first published this classification five TNFR structures have been released [26–28,45,46], corresponding to 19 new CRD structures; if one excludes TNFR21-1, which is truncated, 16 of these new CRDs would have been correctly classified using our initial approach, and only two (TNFR5-3 and TNFR11A-3, corresponding to CRD3s) would have been misclassified as type L-1.

Implications for TNFL/TNFR interactions

CRD and THD structural classifications can easily be compared in light of the known ligand–receptor interactions. We therefore estimated individual T-RMSD structure-based clusters for types L-1, L-2, L-3 and L-4 CRDs. The resulting classifications were then compared with the THD classifications (Figure 4b). When comparing a CRD and a THD clustering, one would intuitively expect closely related CRDs to interact with equally close THDs. Such a consistent pattern of interaction should result in a small number of interaction arrows crossing one another (as shown in Figure 4b). However, no such pattern can be seen for types L-1 (CRD1), L-2 (CRD2) or L-4 (CRD4) CRDs. For these domains, interaction lines cross one another in what appears to be a near-random pattern. By contrast, type L-3 CRDs (CRD3) show near perfect agreement, with most closely related CRDs interacting (or cross-interacting) with equally close THDs.

This observation is very interesting because it accurately reflects the known structural characteristics of TNFL/TNFR interactions (Figure 4d). In the mature complex, most structural data suggest that CRD1 is not directly in contact with the ligand. The available structural complexes also indicate close interactions between L-THD and CRD2/CRD3 (labeled here as types L-2 and L-3) as illustrated by Figure 4d. CRD2s are the most structurally constrained (Figure 4a, c), and 3D structures suggest that they interact with equally conserved regions of THDs; this makes CRD2 domains very likely to be involved in the modulation of binding affinity as previously reported in the analysis of crystallographic structures [24,31,32]. By contrast, CRD3s are the least structurally conserved of the core CRDs, and structural data show a tight interaction between their most variable regions and the equally variable regions of the THD (Figure 4d); this makes CRD3 an ideal candidate for binding specificity control. This structural variability also explains why this domain is harder to classify as a homogeneous group than its neighbors. Interestingly, the diversity of CRD3 domains, which appears to be random when considering the domain alone, clearly makes sense when taking into account the THD and CRD3 structure-based clusters. Indeed, variations within this CRD probably reflect some tight evolutionary calibration that has occurred to evolve and maintain interaction specificity. In this context, the most likely explanation is convergent evolution with some selection of receptors (CRD3) and ligands (THD) able to interact.

Concluding remarks

Our extensive review of the available structural data for both the TNFL and TNFR families suggests that in-depth structural analysis can yield very informative clues on the dynamics of ligand–receptor interactions. Our approach is almost entirely based on the compilation of structure-based clusters using T-RMSD, a newly described structural clustering method. Based on these comparisons, we show that TNFLs can be categorized into three groups that are mostly distinguished by their TNFL/TNFR interaction interface. For TNFRs, we also identified three distinct groups of CRDs, the largest one of which (core CRDs) could be further subdivided in three distinct subgroups whose structural characteristics tend to reflect the receptor ectodomain architecture.

Our approach raises the important question as to whether or not structure-based clusterings, similar to those delivered by the T-RMSD method, are evolutionary informative and might be used to support phylogenetic analysis [42,50]. In order to be evolutionary informative, a classification scheme must reflect, through its hierarchy, the true history of how diversity emerged from a unique

Box 1. Outstanding questions

- Can TNFL/TNFR interaction specificity be predicted on the basis of structural comparisons, or on the basis of sequence-based comparisons?
- What are the evolutionary relationships between receptors and ligands?
- How are receptors selected by evolution?
- What are the evolutionary trends against (or in favor of) promiscuity?

common ancestor. Convergent evolution is the most confounding phenomenon when doing evolutionary analysis as it tends to bring close together (on a tree) species or genes that might have diverged earlier than implied by the tree. The CRD clusters provide an interesting insight into this question. Notably, the global clustering brings together the domains in a way that almost perfectly reflects ectodomain architecture. This result is achieved by using only the variation of intramolecular distances estimated within individual CRDs. Importantly, these two layers of information (architecture and intramolecular distances within each CRD unit) are independent; their agreement, therefore, is highly informative. It clearly indicates that the distance-based clustering does not produce a random classification. Because these receptors probably arose through gene duplication [51], such an observation would suggest a tree driven by divergent evolution, and therefore phylogenetically informative. This hypothesis, however, is not well supported by the comparison of the individual CRD clusters (types L-1, L-2, L-3 and L-4). Indeed, if the TNFRs containing these domains had all descended from a unique common ancestor, one would expect the domains to have similar histories (i.e., the history of the gene itself) and therefore congruent clustering. This is clearly not the case. Convergent evolution is a realistic alternative model. This hypothesis is especially well supported in the case of the CRD3, for which closely related CRDs tend to interact with equally closely related THDs (Figure 4b). Such an agreement suggests an active process of convergent evolution, either by the CRD, the THD, or both. Convergence of THDs cannot be ruled out, but it is poorly supported by the strong agreement between the sequence-based phylogenetic tree (Figure S1 in the supplementary material online [42]) and the structure-based clustering (Figure 2b). This makes CRD3 the component of the system most likely to be undergoing a convergent evolutionary process. Under this scenario, novel THD ligands might appear through random drift, whereas TNFRs are selected for their capacity to capture and interpret these newly emerged signals. Importantly, the only reason we managed to catch the CRD3 red-handed in evolutionary convergence is because experimental interaction data are available. It is therefore impossible to rule out the possibility that the other domains might be evolving under equally strong constraints, even though the factor driving their evolution is unknown to us (Box 1).

Acknowledgments

C.N. and C.M. are funded by the following grants: BFU2008-00419 and BFU2011-28575 from the Spanish Ministry of Science; LEISHDRUG-223414 and Quantomics-222664, both financed by the 7th Framework Program of the European Commission; and SGR-951, from the Catalan Government. Computational resources are provided by the Centre for Genomic Regulation (CRG) of Barcelona. This work was also supported by the EU (PROSPECTS, grant agreement number HEALTH-F4-2008-201648, to L.S.). A.M.S. was partially supported by a Juan de la Cierva fellowship from the Spanish Ministry of Science and Education. The authors would like to thank the referees for useful comments and suggestions.

References

1 Gaur, U. and Aggarwal, B.B. (2003) Regulation of proliferation, survival and apoptosis by members of the TNF superfamily. *Biochem. Pharmacol.* 66, 1403–1408
2 Locksley, R.M. *et al.* (2001) The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell* 104, 487–501
3 Aggarwal, B.B. (2003) Signalling pathways of the TNF superfamily: a double-edged sword. *Nat. Rev. Immunol.* 3, 745–756
4 Hehlgans, T. and Pfeffer, K. (2005) The intriguing biology of the tumour necrosis factor/tumour necrosis factor receptor superfamily: players, rules and the games. *Immunology* 115, 1–20
5 Clark, I.A. (2007) How TNF was recognized as a key mechanism of disease. *Cytokine Growth Factor Rev.* 18, 335–343
6 Tansey, M.G. and Szymkowski, D.E. (2009) The TNF superfamily in 2009: new pathways, new indications, and new drugs. *Drug Discov. Today* 14, 1082–1088
7 Grewal, I.S. (2009) Overview of TNF superfamily: a chest full of potential therapeutic targets. *Adv. Exp. Med. Biol.* 647, 1–7
8 Bossen, C. *et al.* (2006) Interactions of tumor necrosis factor (TNF) and TNF receptor family members in the mouse and human. *J. Biol. Chem.* 281, 13964–13971
9 Zhang, G. (2004) Tumor necrosis factor family ligand-receptor binding. *Curr. Opin. Struct. Biol.* 14, 154–160
10 Pradet-Balade, B. *et al.* (2002) An endogenous hybrid mRNA encodes TWE-PRIL, a functional cell surface TWEAK-APRIL fusion protein. *EMBO J.* 21, 5711–5720
11 Hymowitz, S.G. *et al.* (2003) The crystal structures of EDA-A1 and EDA-A2: splice variants with distinct receptor specificity. *Structure* 11, 1513–1520
12 Bodmer, J.L. *et al.* (2002) The molecular architecture of the TNF superfamily. *Trends Biochem. Sci.* 27, 19–26
13 Kishore, U. *et al.* (2004) C1q and tumor necrosis factor superfamily: modularity and versatility. *Trends Immunol.* 25, 551–561
14 Ghebrehiwet, B. *et al.* (2012) The C1q family of proteins: insights into the emerging non-traditional functions. *Front. Immunol.* 3, 52
15 Bazan, J.F. (1993) Emerging families of cytokines and receptors. *Curr. Biol.* 3, 603–606
16 Naismith, J.H. and Sprang, S.R. (1998) Modularity in the TNF-receptor family. *Trends Biochem. Sci.* 23, 74–79
17 Chan, F.K. *et al.* (2000) A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling. *Science* 288, 2351–2354
18 Clancy, L. *et al.* (2005) Preligand assembly domain-mediated ligand-independent association between TRAIL receptor 4 (TR4) and TR2 regulates TRAIL-induced apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18099–18104
19 Chan, F.K. (2007) Three is better than one: pre-ligand receptor assembly in the regulation of TNF receptor signaling. *Cytokine* 37, 101–107
20 Sedger, L.M. *et al.* (2006) Poxvirus tumor necrosis factor receptor (TNFR)-like T2 proteins contain a conserved preligand assembly domain that inhibits cellular TNFR1-induced cell death. *J. Virol.* 80, 9300–9309
21 Kramer, J.M. *et al.* (2007) Cutting edge: identification of a pre-ligand assembly domain (PLAD) and ligand binding site in the IL-17 receptor. *J. Immunol.* 179, 6379–6383
22 Chan, F.K. (2000) The pre-ligand binding assembly domain: a potential target of inhibition of tumour necrosis factor receptor function. *Ann. Rheum. Dis.* 59 (Suppl. 1), i50–i53
23 Branschadel, M. *et al.* (2010) Dual function of cysteine rich domain (CRD) 1 of TNF receptor type 1: conformational stabilization of CRD2 and control of receptor responsiveness. *Cell. Signal.* 22, 404–414
24 Mongkolsapaya, J. *et al.* (1999) Structure of the TRAIL–DR5 complex reveals mechanisms conferring specificity in apoptotic initiation. *Nat. Struct. Biol.* 6, 1048–1053

25 Banner, D.W. *et al.* (1993) Crystal structure of the soluble human 55 kd TNF receptor–human TNF beta complex: implications for TNF receptor activation. *Cell* 73, 431–445
26 Mukai, Y. *et al.* (2010) Solution of the structure of the TNF–TNFR2 complex. *Sci. Signal.* 3, ra83
27 Zhan, C. *et al.* (2011) Decoy strategies: the structure of TL1A:DcR3 complex. *Structure* 19, 162–171
28 Liu, C. *et al.* (2010) Structural and functional insights of RANKL–RANK interaction and signaling. *J. Immunol.* 184, 6910–6919
29 Compaan, D.M. and Hymowitz, S.G. (2006) The crystal structure of the costimulatory OX40–OX40L complex. *Structure* 14, 1321–1330
30 Hymowitz, S.G. *et al.* (2005) Structures of APRIL–receptor complexes: like BCMA, TACI employs only a single cysteine-rich domain for high affinity ligand binding. *J. Biol. Chem.* 280, 7218–7227
31 Hymowitz, S.G. *et al.* (1999) Triggering cell death: the crystal structure of Apo2L/TRAIL in a complex with death receptor 5. *Mol. Cell* 4, 563–571
32 Cha, S.S. *et al.* (2000) Crystal structure of TRAIL–DR5 complex identifies a critical role of the unique frame insertion in conferring recognition specificity. *J. Biol. Chem.* 275, 31171–31177
33 Notredame, C. *et al.* (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217
34 Taly, J.F. *et al.* (2011) Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat. Protoc.* 6, 1669–1682
35 Holm, L. *et al.* (2006) Using Dali for structural comparison of proteins. *Curr. Protoc. Bioinform.* http://dx.doi.org/10.1002/0471250953.bi0505s14 Chapter 5, Unit 5.5
36 Punta, M. *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res.* 40, D290–D301
37 Letunic, I. *et al.* (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 40, D302–D305
38 Sigrist, C.J. *et al.* (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* 38, D161–D166
39 Chattopadhyay, K. *et al.* (2007) Assembly and structural properties of glucocorticoid-induced TNF receptor ligand: implications for function. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19452–19457
40 Armougom, F. *et al.* (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* 34, W604–W608
41 Magis, C. *et al.* (2010) T-RMSD: a fine-grained, structure-based classification method and its application to the functional characterization of TNF receptors. *J. Mol. Biol.* 400, 605–617
42 Glenney, G.W. and Wiens, G.D. (2007) Early diversification of the TNF superfamily in teleosts: genomic characterization and expression analysis. *J. Immunol.* 178, 7955–7973
43 Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics* 14, 755–763
44 Won, E.Y. *et al.* (2010) The structure of the trimer of human 4-1BB ligand is unique among members of the tumor necrosis factor superfamily. *J. Biol. Chem.* 285, 9202–9210
45 Kuester, M. *et al.* (2011) The crystal structure of death receptor 6 (DR6): a potential receptor of the amyloid precursor protein (APP). *J. Mol. Biol.* 409, 189–201
46 An, H.J. *et al.* (2011) Crystallographic and mutational analysis of the CD40–CD154 complex and its implications for receptor activation. *J. Biol. Chem.* 286, 11226–11235
47 Mackay, F. and Schneider, P. (2009) Cracking the BAFF code. *Nat. Rev. Immunol.* 9, 491–502
48 Bossen, C. and Schneider, P. (2006) BAFF, APRIL and their receptors: structure, function and signaling. *Semin. Immunol.* 18, 263–275
49 Tur, V. *et al.* (2008) DR4-selective tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) variants obtained by structure-based design. *J. Biol. Chem.* 283, 20560–20568
50 Wiens, G.D. and Glenney, G.W. (2011) Origin and evolution of TNF and TNF receptor superfamilies. *Dev. Comp. Immunol.* 35, 1324–1335
51 Collette, Y. *et al.* (2003) A co-evolution perspective of the TNFSF and TNFRSF families in the immune system. *Trends Immunol.* 24, 387–394
52 Jordan, G.E. and Piel, W.H. (2008) PhyloWidget: web-based visualizations for the tree of life. *Bioinformatics* 24, 1641–1642
53 Eck, M.J. and Sprang, S.R. (1989) The structure of tumor necrosis factor-alpha at 2.6 A resolution Implications for receptor binding. *J. Biol. Chem.* 264, 17595–17605
54 Karpusas, M. *et al.* (1995) 2 A crystal structure of an extracellular fragment of human CD40 ligand. *Structure* 3, 1031–1039
55 Liu, Y. *et al.* (2003) Ligand–receptor binding revealed by the TNF family member TALL-1. *Nature* 423, 49–56
56 Jin, T. *et al.* (2007) X-ray crystal structure of TNF ligand family member TL1A at 2.1A. *Biochem. Biophys. Res. Commun.* 364, 1–6
57 Gaboriaud, C. *et al.* (2003) The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. *J. Biol. Chem.* 278, 46974–46982
58 Kvansakul, M. *et al.* (2003) Crystal structure of the collagen alpha1(VIII) NC1 trimer. *Matrix Biol.* 22, 145–152
59 Bogin, O. *et al.* (2002) Insight into Schmid metaphyseal chondrodysplasia from the crystal structure of the collagen X NC1 domain trimer. *Structure* 10, 165–173
60 Naismith, J.H. *et al.* (1996) Structures of the extracellular domain of the type I tumor necrosis factor receptor. *Structure* 4, 1251–1262
61 He, F. *et al.* (2009) Solution structure of the cysteine-rich domain in Fn14, a member of the tumor necrosis factor receptor superfamily. *Protein Sci.* 18, 650–656
62 Carfi, A. *et al.* (2001) Herpes simplex virus glycoprotein D bound to the human receptor HveA. *Mol. Cell* 8, 169–179
63 He, X.L. and Garcia, K.C. (2004) Structure of nerve growth factor complexed with the shared neurotrophin receptor p75. *Science* 304, 870–875
64 Graham, S.C. *et al.* (2007) Structure of CrmE, a virus-encoded tumour necrosis factor receptor. *J. Mol. Biol.* 372, 660–671