CMLS, Cell. Mol. Life Sci. 54 (1998) 880–891
1420-682X/98/080880-12 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Research Article

The V(D)J recombination activating protein RAG2 consists of a six-bladed propeller and a PHD fingerlike domain, as revealed by sequence analysis

I. Callebaut* and J.-P. Mornon

Systèmes moléculaires et Biologie structurale, LMCP, CNRS UMR C7590, Universités Paris 6 et Paris 7, case 115, 4 place Jussieu, F-75252 Paris Cedex 05 (France), Fax +33 1 44 27 37 85, e-mail: Isabelle.Callebaut@lmcp.jussieu.fr

Received 6 June 1998; accepted 9 June 1998

Abstract. The RAG1 and RAG2 proteins play a crucial role in V(D)J recombination by cooperating to make specific double-stranded DNA breaks at a pair of recombination signal sequences (RSSs). However, the exact function they perform has heretofore remained elusive. Using a combination of sensitive methods of sequence analysis, we show here that the active core region of the RAG2 protein, confined to the first three quarters of its sequence, is in fact composed of a six-fold repeat of a 50-residue motif which is related to the kelch/mipp motif. This motif, which forms a four-stranded twisted antiparallel β sheet, is arranged in a circular formation like blades of a propeller or turbine. Given the known properties of the β-propeller fold in mediating protein-protein interactions, it is proposed that this six-laded propeller structure of the RAG2 active core would play a crucial role in the tight complex formed by the RAG1 and RAG2 proteins and RSSs. Moreover, the presence of a plant homeodomain finger-like motif in the last quarter of the RAG2 sequence suggests a potential interaction of this domain with chromatin components.

Key words. RAG2; V(D)J recombination; kelch motif; β-propeller; PHD motif; zinc binding; hydrophobic cluster analysis.

V(D)J recombination is the fundamental process by which the antigen receptor genes (immunoglobulins and T-cell receptors) are assembled from multiple coding segments (reviewed in refs 1–3). It is mediated through conserved recombination signal sequences (RSS) that flank coding segments and consist of conserved heptamer and nonamer sequences separated by a non-conserved spacer of either 12 or 23 bp. Efficient recombination is restricted to pairs of signals with dissimilar spacer length, obeying the '12/23' rule [4].

The process is initiated by the introduction of double-strand DNA breaks at the border between each RSS and the coding segment. The products of the cleavage are, on the one hand, blunt 5′-phosphorylated signal ends (SEs) [5] which undergo head-to-head ligation to form a precise signal joint and, on the other hand, covalently closed hairpin-coding ends (CEs) [6–8] which are joined to form an imprecise coding joint after additional and extensive processing.

* Corresponding author.

The lymphoid-specific proteins RAG-1 and RAG-2 (recombination-activating genes) [9, 10] are involved in the initial steps of the V(D)J recombination process by performing coordinated cleavage on pairs of recombination signals [11, 12]. The RAG-mediated cleavage occurs through the introduction of a nick at the heptamer/coding segment border followed by attack of the free 3′-hydroxyl on the other strand to generate a blunt 5′-phosphorylated SE and a hairpin CE [13]. This latter step occurs via a direct transesterification reaction, suggesting a mechanistic link between V(D)J recombination, transposition and retroviral integration [14]. Specific and stable cleavage-competent complexes containing only RAG-1, RAG-2 and an RSS can be formed in vitro in the presence of a divalent metal ion, with completion of both steps of cleavage requiring the presence of $Mn2^{+}$ [15]. These complexes require the conserved heptamer and nonamer motifs of the RSS as well as both the RAG1 and RAG2 proteins. Other studies on RAG complex-binding specificity have revealed that RAG1 by itself is capable of sequence-specific recognition of 12 to 23 signals via the nonamer, while localization of RAG2 to the signal is RAG1-dependent, with a more efficient recruitment at 12 signals than 23 signals [16, 17]. The DNA-bending proteins HMG1 and HMG2 stimulate RAG binding and cleavage at the 23-bp signals, suggesting that a similar array of RAG proteins may accomodate a signal with 12- or 23-bp spacers and that the much more severe bending required in the longer spacer can be stabilized by these chromatin proteins [18, 19]. The RAG complexes appear to contain three to five molecules of RAG2 for each molecule of RAG1, as shown by coimmunoprecipitation studies [20, 21]. Although RAG2 appears dispensable for sequence-specific DNA binding [16, 17], RAG1 is unable to cleave DNA in its absence and stable cleavage complexes are not formed [15, 19], raising the possibility either that an RAG2-induced conformational change of RAG1 is needed to activate an active site in RAG1 or, alternatively, that RAG2 may directly contribute to the enzymatic site by interacting with the DNA near the heptamer-coding end border [16]. The RAG proteins also appear to play a role in the later steps of V(D)J recombination for efficient joining of coding ends [22]. Moreover, they form stable postcleavage complexes between synapsed recombination signals [19].

Although substantial progress has been made in understanding the V(D)J recombination machinery in which the RAG proteins are directly involved, the exact biochemical role of these two proteins remains unclear. Mutagenesis procedures have been used in order to define the minimum core regions capable of recombination of plasmid substrates [23–26]. The C-terminal half of RAG1 as well as the N-terminal three quarters of RAG2 maintain recombination activity. These 'active cores' are necessary and sufficient to induce cleavage in vitro [13, 27] and can reconstitute the 12/23 rule in cleavage reactions in vitro [11, 12]. In contrast, the C-terminal acidic region of RAG2 as well as the N-terminal one-third of RAG1 have been found to be dispensable for the recombinase activity. This latter region, which contains a zinc-binding domain composed of a zinc RING finger (C3HC4) and a C2H2 zinc finger [28], is thought to be involved in RAG1 dimerization [29]. The RAG1 active core has otherwise been shown to be responsible for the interaction with RAG2 [30].

The C-terminal half of RAG1 has been reported to share weak sequence similarities with the yeast topoisomerase-like protein HPR1 [31], but mutations of the potential active tyrosines of RAG1 [23] which would covalently bind to the DNA phosphodiester backbone do not abolish recombination activity and thus do not support a topoisomerase-like function for RAG1 [23, 26, 32]. On the other hand, the DNA-binding activity of RAG1 has recently been located in the N-terminal region of the RAG1 active core, sharing similarities with bacterial invertases [16, 17].

In contrast to RAG1, the RAG2 protein has heretofore been reported as sharing no detectable similarity with any other protein in the data banks. In order to gain more insight into the structural and functional features of this protein, we used a combination of sensitive methods, including hydrophobic cluster analysis (HCA) ([33], [34]) and gapped-BLAST (basic local alignment search tool) [35]. HCA is a sensitive bidimensional method of sequence analysis that is able to detect significant similarities between proteins sharing low levels of sequence identity (typically below 25% identity) [34]. Its sensitivity stems from its ability to efficiently combine secondary structure detection with sequence comparison. A recent example of the effectiveness of such a stategy is the detection of a widespread BRCT (BRCA1 C-terminus) module likely to play an important role in mediating interactions between several nuclear proteins involved in cell cycle control and response to DNA damage ([36], see also ref. 37). The BRCT module is, among other things, involved in DNA double-strand break repair with which V(D)J recombination shares a number of factors. Indeed, the XRCC4 protein, whose absence causes nonhomologous end joining (NHEJ) and V(D)J recombination defects, interacts functionally with DNA ligase IV [38, 39] through the two BRCT modules that this enzyme contains in its C-terminal end [38]. The BRCT module is also found in terminal deoxynucleotidyl transferase (TdT), which is an accessory factor for generating immune diversity by mediating addition of nontemplated nucleotides into coding junctions [2].

Here, we show that the RAG2 active core is in fact composed of a sixfold internal repeat related to the kelch/mipp family. The kelch motif is a motif of approximately 50 residues, whose name comes from the *Drosophila* egg-chamber regulatory protein in which it was first identified [40, 41]. A variable number of copies of this motif are found in diverse proteins such as the mouse intracisternal A-particle (IAP)-promoted placental (mipp) protein [40], α- and β-scruin [42, 43] and in a number of proteins from poxviruses [40]. It is known from the experimental three-dimensional 3D structure of the fungal galactose oxidase [44, 45], whose catalytic unit contains seven kelch repeats [41], that this short motif corresponds to a four-stranded antiparallel β sheet which is arranged in a circular formation like flower petals or the blades of a propeller (fig. 1). β-propeller structures, also named superbarrels, can accomodate several symmetries, ranging from four (hemopexin) [46] to eight (methanol dehydrogenase) [47]. The most frequent arrangement, which would be the most energetically favourable [48], probably corresponds to the sevenfold symmetry, like that observed in the β subunit of the heterotrimeric G proteins [49–51], in the catalytic subunit of galactose oxidase [44, 45], in methylamine dehydrogenase [52] and in the recently solved structure of the regulator of chromosome condensation (RCC1) [53]. A sixfold symmetry, like that predicted here in the RAG2 active core, is observed in viral neuraminidases and bacterial sialidases [54, 55].

Although the repeats at the origin of the formation of β propellers are involved in a wide variety of functions, their obvious common denominator is their involvement in mediating interaction between proteins but also between proteins and DNA, as recently exemplified by the RCC1 structure [53]. This 'interactive' scaffold could therefore play a crucial role in the assembly and in the function of the tight complex formed by the RAG1 and RAG2 proteins with RSSs.

Materials and methods

Guidelines to the use of HCA have been published elsewhere [34]. Briefly, protein sequences are displayed on a duplicated α-helical net with amino acids indicated above. The contours of the hydrophobic residues are automatically drawn. They form clusters which mainly correspond to the internal faces of regular secondary structures. Four symbols are used in the plots, relating to the specific structural behaviour of the amino acids they represent (see inset in fig. 2).

The statistical significance of the alignments is assessed computation of Z-scores. Pairwise Z-score (pZ-scores) represent the difference between the alignment score under consideration and the mean score of a distribution computed for the alignment of sequence 1 vs. a large number (10,000) of randomly shuffled versions of sequence 2 [34]. These differences are expressed relative to the standard deviation of the random distribution (σ), with values exceeding 3 σ strongly indicative of genuine relationships for small domains (e.g. ref. 36). The Z-scores are calculated for identity, similarity (using BLOSUM-62 [56]) and HCA (% of hydrophobic

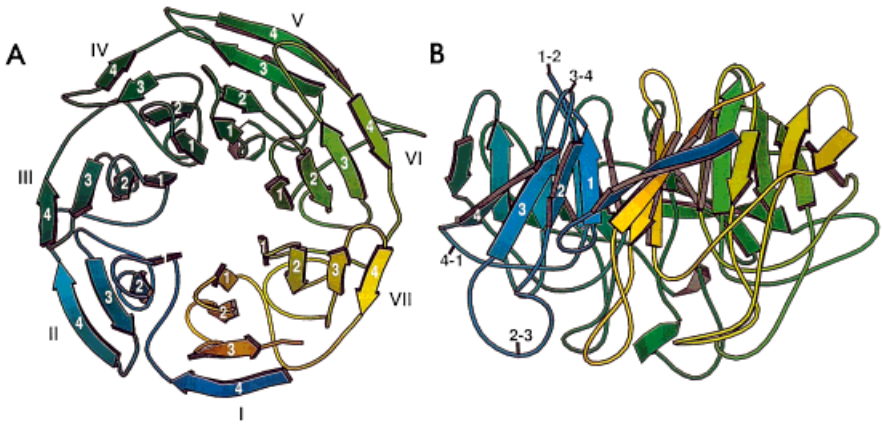


Figure 1. Molscript2 ribbon diagram of the seven-bladed propeller of *Dactylium dendroïdes* galactose oxidase (pdb identifier: 1GOF) whose sequence repeats belong to the kelch family. The β propeller is viewed along (*A*) or perpendicular to (*B*) the central shaft. β strands are labelled 1 (inner strand) to 4 (outer strand). The blades are numbered I to VII along the sequence. Loops 1–2 and 3–4 (connecting β1 to β2 and β3 to β4, respectively) protrude on one side of the propeller (*B*, top) whereas the much longer loops 2–3 and 4–1 (connecting β2 to β3 and β4 to β1, respectively) are found on the other side (*B*, bottom). Loops 4–1 connect the different β-sheet blades to each other.

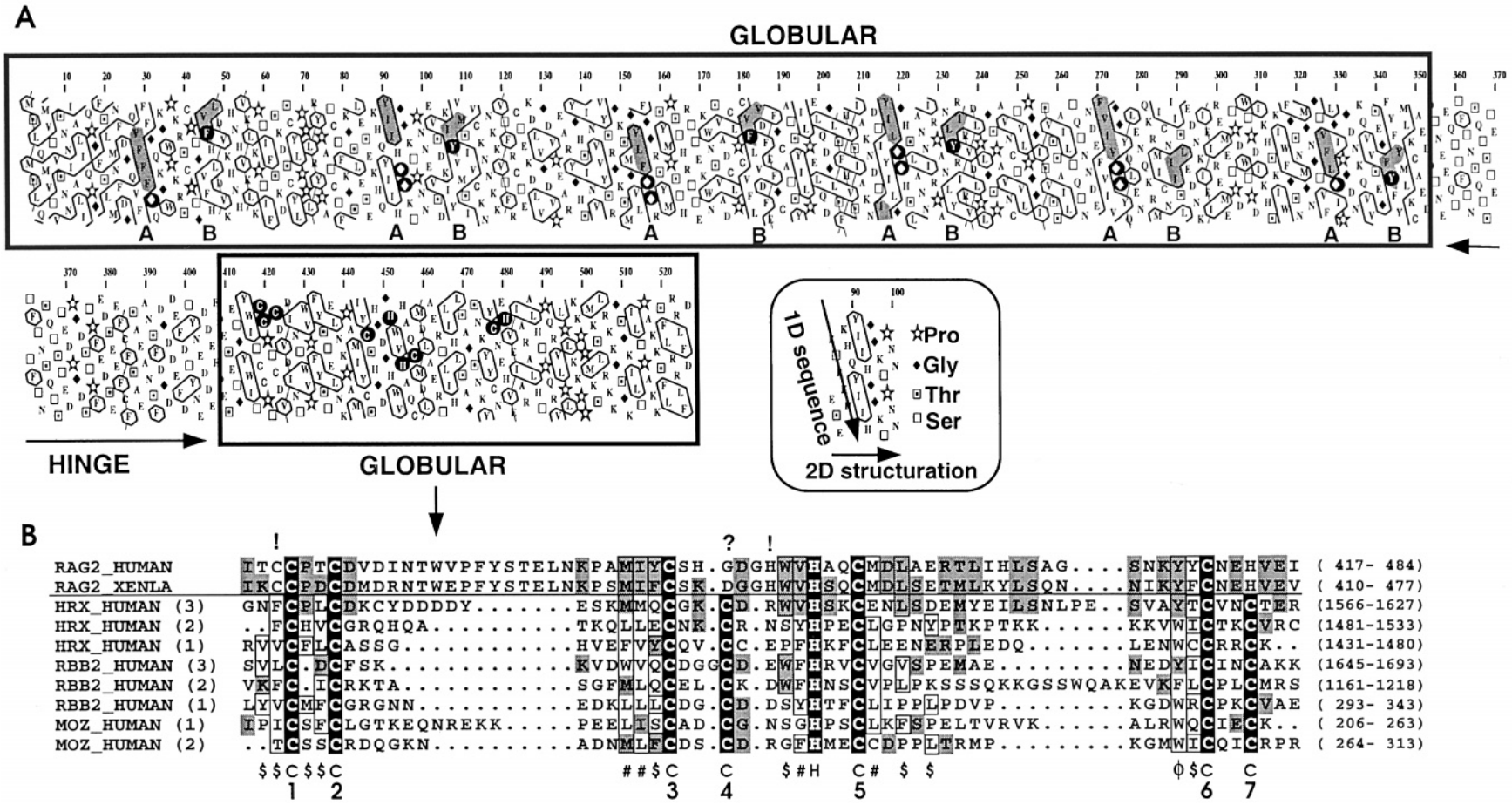


Figure 2. (*A*) HCA representation of the RAG2 sequence (RAG2_HUMAN). Globular regions are boxed, and the hinge separating them is underlined. Symbols used and the manner in which the 1D sequence and the 2D structuration are presented are indicated in the inset. Repeated motifs (A and B) are shaded with highly conserved amino acids in white on a black background. The cysteine and histidine of the C-terminal globular domain which are conserved among species are also highlighted. (*B*) 1D alignment of the RAG2 C-terminal domain with the PHD/TTC/LAP domains of ALL-1, the human homologue of the *Drosophila* trithorax protein (sw:TRX_HUMAN), the retinoblastoma-binding protein RBP2 (sw:RBB2_HUMAN) and the putative chromatin associated acetylase MOZ (sw:MOZ_HUMAN). The position of the motifs in the protein sequences is indicated. Conserved amino acids of the PHD/TTC/LAP family are boxed, and identical amino acids between the RAG2 sequences and PHD/TTC/LAP domains are shaded. The RAG2 PHD/TTC/LAP-like domain matches most of the consensus sequence described by Aasland et al. [60] and Koken et al. [61] [Φ, aromatic amino acids, # and $, highly and semiconserved hydrophobic amino acids (VILMFYW)]. The missing cysteine (?) could be functionally replaced either by the conserved cysteine or by the conserved histidine (shown with the symbol !).

matching) scores. Reliability indexes (RI) are also calculated to assess relationships: these are obtained by dividing the product of the three observed Z-scores defined above (observed Z3) by the best product of the three Z-scores obtained in the random comparison process (best random Z3).

The sensitivity of the statistical analysis for assessing distant relationships can be greatly enhanced in a 'profilelike' manner by comparing sequences to a large set of sequences representative of a family, instead of only to a single one. This procedure distinguishes positions which are highly conserved in the family from the much more variable ones, which can be substituted by a wide variety of amino acids. Scores are computed between an $n \times m$ matrix (where n and m are the number of sequences in the family and the alignment length, respectively) and the sequence supposed to belong to the family (query) (I. Callebaut et al., unpublished data). Multiple Z-scores (mZ) are calculated by comparing observed scores with the distribution of scores obtained after randomization of the query sequence.

Database screening using a multiple iterative strategy was performed with the gapped-BLAST program [35], running at the National Center of Biological Information (NCBI).

3D visualization and manipulations of 3D structures were performed using InsightII (Molecular Simulation).

Results and discussion

Identification of kelch-like repeats in the RAG2 active core. Examination of the RAG2 sequence using HCA highlighted the presence of two distinct globular domains separated by a hinge region of approximately 60 amino acids (fig. 2A). Globular regions are characterized by a typical 'texture' of hydrophobic clusters (these regions contain approximately 1/3 of strong hydrophobic amino acids) constrasting with hinge regions which only contain scattered hydrophobic amino acids. The limits of the first domain (aa 1–355) correspond to the limits of the RAG2 active core, therefore supporting the functional independence of this region relative to the protein as a whole. Further examination of the RAG2 HCA plot led to identification of a repeated motif (motif A in fig. 2A), including two consecutive glycines (shown white on a black background), which are preceded by a vertical cluster containing four contiguous hydrophobic amino acids (shaded gray), strongly predicted to statistically correspond to a β strand (at least 75%, data not shown). Slight variations on this consensus (underlined) were observed in the three first and in the last copies of the motif (A1: VFFFGQ, A2: YIIHGG, A3: GVLFGG, A6: TVFLGI). Note that amino acids such as H (A2) and T (A6) are frequently found to substitute aromatic (FY) and strong hydrophobic (VILMFYW) amino acids, respectively. Downstream from this conserved pattern, another conserved motif (motif B) can also be observed, centred on a typical cluster (shaded gray) which would also correspond to a β strand and which generally contains a conserved aromatic amino acid (shown white on a black background). These features were also found in the conserved RAG2 core sequences of different species (sequence identities with the human sequence range from 90 (rabbit) to 49% (*Xenopus*). This observation suggests that the active core of RAG2 consists of a sixfold repeated module of approximately 50 amino acids (the mean distance separating two consecutive conserved motifs).

The screening of sequence databases provided further support of this hypothesis. Indeed, we detected in several proteins short regions of sequence similarity, typically over a 50-amino acid length, in which the above-described motif A is highly conserved. These regions are found, sometimes in several places, in various proteins including the protein encoded by the gene *LZTR-1*, which is deleted in DiGeorge syndrome [57]; calicin, a major basic protein of the mammalian sperm head cytoskeleton [58]; tea1, a fission yeast protein which acts as an end marker directing the growth machinery to the cell poles [59]; and several hypothetical proteins from yeast (Swiss-prot identifier YG52_yeast), *Arabidopsis thaliana* (F21B7.27) and *C. elegans* (C55A1.11). A careful HCA-based examination of the sequences of these proteins showed that the regions matching the RAG2 repeats always correspond to internal repeats which are known to belong, in calicin and tea1p, to the 'kelch' family. Even though the E statistical values associated with each individual BLAST match are not significant in themselves, the relationship between the RAG2 repeats and the kelch motif is further supported by the following observations. First, in each case the similarity is concentrated around motif A, which, interestingly, corresponds to the most conserved positions of the kelch pattern (fig. 3). The glycines contained in this motif can satisfy ϕ and ψ angles that are not allowed for other amino acids and would be located in positions where there is no room to accommodate side chains, as observed in the galactose oxidase structure. Furthermore, the similarity region initially retained by BLAST can often be extended through HCA. The conserved aromatic amino acid of the RAG2 motif B is also found highly conserved in the kelch pattern, at a similar distance relative to motif A. The two motifs A and B participate in the two internal β strands ($\beta 2$ and $\beta 3$) of the four-stranded β sheet, as observed in the structure of galactose oxidase (figs 1 and 3). Finally, the RAG2 repeated modules and kelch repeats share approximately the same length (~ 50 residues). Other striking sequence stretches, particularly

those located in the $\beta 4-\beta 1$ loop (linking two different blades), are also found to be conserved between RAG2 repeats and some kelch repeats [e.g. PxPRYGH, in which the arginine is highly conserved in several members of the kelch family (fig. 3)]. All these observations strongly suggest that the RAG2-repeated module probably corresponds to a kelch motif. Variations relative to canonical kelch repeats are principally observed in the region including the inner and outer strands, namely $\beta 1$ and $\beta 4$, whose corresponding sequences are otherwise also less conserved in the kelch family. In particular, the tryptophan of the strand $\beta 4$ is always missing in the RAG2 kelch-like repeats.

The statistical significance of the alignment of the kelch-like repeats of RAG2 relative to each other (pairwise comparisons) and relative to well-established kelch repeats (multiple comparisons) was assessed through the calculation of Z-scores (fig. 3). Taking into account the small length of the repeats, pairwise Z-scores calculated for the comparison of the kelch-like repeats of RAG2 are not very high but indicative of a potential relationship (mean pZ-scores for identity, similarity and HCA scores are 1.5, 1.4 and 2.8, respectively) and similar to values obtained for well-established kelch repeats (e.g. for the CEF53E4 sequence, mean mZ-scores for identity, similarity and HCA scores are 2.0, 1.3 and 1.3, respectively). In contrast, the multiple Z-scores and corresponding RIs calculated between each RAG2 repeat and a set of representative sequences of the kelch family (fig. 3) are consistent with a genuine relationship at low level of sequence identity for small domains and clearly connect the RAG2-repeated sequences to the kelch family [mean mZ-scores for identity, similarity and HCA scores are 7.0 (σ 1.3), 6.6 (σ 1.0) and 4.2 (σ 0.5), respectively, with a good mean RI of 9.5].

It is worth noting that in RAG2, as in other members of the kelch family and in other β-propeller structures, the sequence repeats do not coincide with the structural repeat. Indeed, as the N-terminal strand corresponds to the last strand ($\beta 4$) of the sheet formed by the three C-terminal β strands ($\beta 1$-$\beta 2$-$\beta 3$) (as in galactose oxidase, fig. 1), the sequence repeat $\beta 4$-$\beta 1$-$\beta 2$-$\beta 3$ is shifted relative to the structural repeat $\beta 1$-$\beta 2$-$\beta 3$-$\beta 4$. The circular closure of the RAG2 propeller would be achieved, as in galactose oxidase [44, 45] or the β subunit of the heterotrimeric G proteins [49–51], by a $1+3$ combination of β strands from the N- ($\beta 4$) and C-terminal ($\beta 1$-$\beta 2$-$\beta 3$) ends. Note that different combinations are found in the kelch family (fig. 3; e.g. the $2+2$ or $3+1$ combination like that observed in RCC1 [53] and methylamine dehydrogenase [52], respectively). This arrangement, like a molecular clasp and common to many propeller structures, seems to be necessary to stabilize the circular formation.

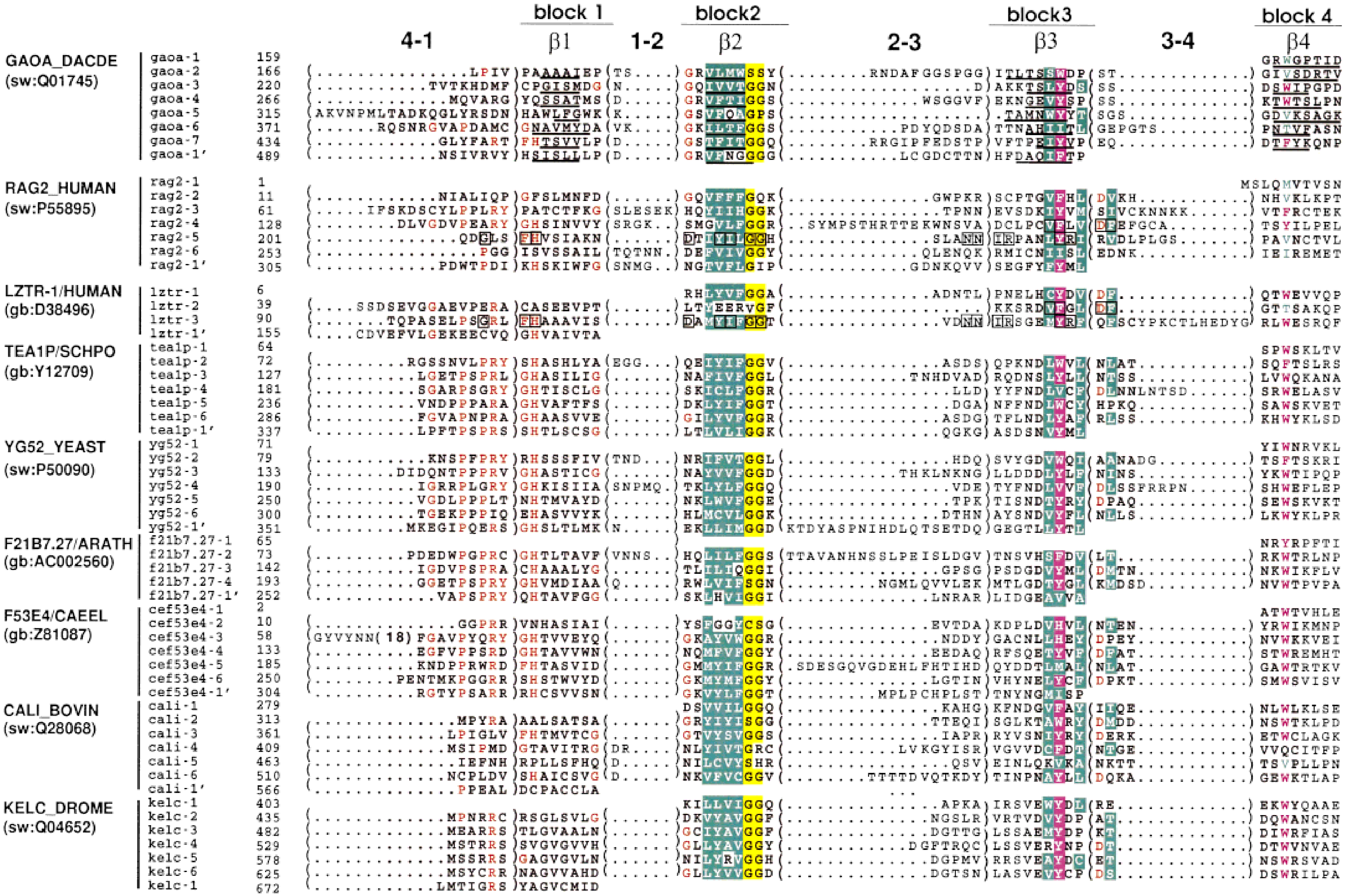
Conserved sequence motifs in different repeats forming β-propeller structures. The lack of immediate sequence similarity between the different families of β-propeller repeats has suggested that they do not share a common ancestor and that the β-propeller fold is a source of convergent evolution associated with its particular structural stability [45].

However, by searching the conserved sequence motifs (A and B) characterizing the core structure of each blade of the kelch motif ($\beta 2$ and $\beta 3$), we found striking similarities in similar positions in the WD repeat family. This finding is strengthened by comparison of the superimposed structures of the blades of galactose oxidase (kelch repeat family) and of the β subunit of the heterotrimeric protein G (WD repeat family) (fig. 4). Indeed, the consensus sequence associated with the strand $\beta 2$ in Gβ (three hydrophobic amino acids, alanine, serine, threonine or cysteine and two small amino acids, the first one often being G) is strikingly similar and has good structural correspondence to motif A of the kelch strand $\beta 2$ (four hydrophobic amino acids followed by two glycines). A conserved glycine two amino acids before the first hydrophobic residue is observed in both families. Moreover, the highly conserved aromatic amino acid of the kelch motif B is found to correspond to the highly conserved aromatic amino acid of the WD motif. These aromatic residues are preceded in the two cases by a hydrophobic amino acid. The sequences linking these two conserved structural elements differ notably in the two families, limited to tight turns in the WD family and allowing large insertions in the kelch family.

This sequence-associated structural conservation was also found to a lower extent in the recently published structure of the regulator of chromosome condensation RCC1 [53]. In this case, four strong amino acids followed by one small amino acid are associated with strand $\beta 2$, whereas an aromatic amino acid is also found at the end of strand $\beta 3$ (the WG fingerprint).

This observation contributes further data to the debate concerning the evolution of β-propeller proteins, raising the possibility that these proteins share a remote ancestor despite sequence divergence to the point where their heritage is not readily apparent, as already suggested by Bork and Doolittle [41] and Sondek and co-workers [50]. This hypothesis is further supported by theoretical structural investigations which demonstrate the absence of strong sequence constraints on this particular fold [48], a situation allowing high sequence divergence.

Identification of a PHD/TTC/LAP-like domain in the RAG2 C-terminal domain. Although the RAG2 C-terminal quarter appears not to be required for recombination, it is strongly conserved among species. It contains a very acidic region (hinge in fig. 2A) followed by a globular domain. The second globular domain of


block 1
block2
block3
block 4
4-1
β1
1-2
β2
2-3
β3
3-4
β4
GAOA_DACDE (sw:Q01745)
RAG2_HUMAN (sw:P55895)
LZTR-1/HUMAN (gb:D38496)
TEA1P/SCHPO (gb:Y12709)
YG52_YEAST (sw:P50090)
F21B7.27/ARATH (gb:AC002560)
F53E4/CAEEL (gb:Z81087)
CALI_BOVIN (sw:Q28068)
KELC_DROME (sw:Q04652)

RAG2 which ends the protein contains an unusual proportion of cysteine and histidine which are conserved among species (six cysteines, three histidines, highlighted in white in fig. 2A), suggesting that these residues could be involved in zinc binding. A search in the sequence databanks using this domain as a probe strengthened this hypothesis. Indeed, interesting matches including most of the cysteine and histidine residues of RAG2 were found with zinc-binding regions of ALL-1, the human homologue of the *Drosophila* trithorax protein, as well as with those of the putative chromatin-associated acetylase MOZ and the retinoblastoma-binding protein RBP2 (fig. 2B). The regions of similarity in these proteins correspond to C4HC3 zinc-finger domains, also named PHD (plant homeodomain) [60], TTC (trithorax consensus) [61] or LAP (leukaemia-associated-protein) [62] domains. This motif has a unique C4HC3 pattern spanning 50–80 residues and enriched by several additional conserved positions. It is clearly distinct from other zinc-binding domains such as the RING finger (C3HC4) and LIM domain (C2HC5) [60–62]. The PHD/TTC/LAP domain has been identified in a series of chromatin-associated proteins, including the Pc-G protein Polycomb-like and the transcriptional adaptors P300/CBP, and is believed to correspond to a protein-protein interaction domain [60–62].

The RAG2 sequences differ slightly from the PHD/TTC/LAP consensus, as the fourth and seventh cysteines are missing. However, the last position is always occupied by a histidine residue which can functionally substitute cysteine. On the other hand, the glycine or aspartic acid found at the position of the fourth cysteine (? in fig. 2B) suggests that another spatially proximal position could provide an appropriate amino acid for zinc binding. Two potential candidates for this role are the conserved cysteine which is found one amino acid before the first cysteine of the PHD/TTC/LAP consensus and the conserved histidine three amino acids before the consensus histidine (! in fig. 2B). Additional conserved positions of the PHD/TTC/LAP motif are well conserved in the RAG2 sequence, suggesting that this domain in RAG2 indeed belongs to this family. Multiple Z-scores and corresponding RIs further assess this relationship. Indeed, mZ-scores calculated for the identity, similarity and HCA scores between the X. *laevis* RAG2 sequence and the set of representative members of the PHD/TTC/LAP family shown in figure 3, are 8.3, 7.9 and 4.1, respectively, with an RI of 17.9. It is worth noting that the RAG2 PHD/TTC/LAP-like domain appears more closely related to the third PHD/TTC/LAP domain of the ALL-1 protein, since both domains share a highly conserved hydrophobic cluster (included between C5 and C6) that appears to be unique to the PHD/TTC/LAP family. Pairwise Z-scores, calculated between these two domains (using the X. *laevis* RAG2 sequence) – which share 33.3% sequence identity – are 8.7, 5.8 and 3.5 for the identity, similarity and HCA scores, respectively, with a reliability index of 10.2.

Functional implications. β propeller structures can be viewed as a general scaffold that is used to display variable sequences (the loops linking the β strands) around the faces of its compact and symmetrical structure. As a result, they are widely involved in mediating protein-protein interactions through the relatively flat surfaces they present on both sides of the structural core. The interactive properties of β propellers from two different sequence families, which both bind GTP-

Figure 3. HCA-deduced 1D alignment of the six RAG2 kelch-like repeats relative to representative members of the kelch family. The alignment has been constructed on the basis of the structural alignment of the seven blades of the *Dactylium dendroïdes* galactose oxidase propeller in which only the strands ($\beta 1$–$\beta 4$) are perfectly superimposable (fig. 4). The positions of these strands define four conserved blocks (blocks 1–4), linked by loops (4–1, 1–2, 2–3 and 3–4; in parentheses) which are much more variable and sometimes completely different (and therefore not alignable). The sequences of these loops are shown, however, because of the presence of conserved motifs in some kelch repeats which help the alignment procedure. Sequences have been aligned by an iterative strategy using automatic pairwise and multiple alignment procedures combined with HCA [34]. Note that the kelch repeats in the LZTR-1 sequence [57] were identified in this study, as well as the last three kelch repeats of calicin, which was previously reported to contain only three [58]. The Swiss-Prot (sw) or GenBank (gb) accession numbers are given for each protein as well as the positions of the N-terminal end of each sequence. Secondary structure positions, as deduced from the experimental structure of *D. dendroïdes* galactose oxidase (fig. 1), are shown underlined below its sequence. Residues which are highly conserved in the family are coloured (green, purple and red for strong hydrophobic (VILMFWY), aromatic (FY) and other amino acids, respectively). The position of the two highly conserved glycines is boxed yellow. Amino acids which can substitute these classes of residues in a position-specific manner are also coloured [e.g. T for strong hydrophobic amino acids, H for aromatic amino acids, small amino acids (S, A) or turn residues (e.g. N or D) for G]. An example of an initial BLAST match is shown boxed (between RAG2 and LZTR-1). As the limits of the sequence repeats differ according to the strand combination adopted by each protein to constitute the circular closure of the propeller [e.g. the sequence repeats of galactose oxidase and RAG2 beginning at strand $\beta 4$ ($1 + 3$ combination) are shifted relative to those of the kelch protein, which start at strand $\beta 2$ ($3 + 1$ combination)], we adopted the limits of the structural repeats to perform the statistical analysis. To this end, the sequence corresponding to the first structural repeat (always designated 1) was built from the combination of N-terminal and C-terminal ends [e.g. in RAG2, the first structural repeat is made from the combination of an N-terminal $\beta 4$ strand (rag2-1 in fig. 3) and the C-terminal sequence including $\beta 1$ to $\beta 3$ (rag2-1′)]. This figure was made using the ESPript Software (P. Gouet et al., unpublished data).

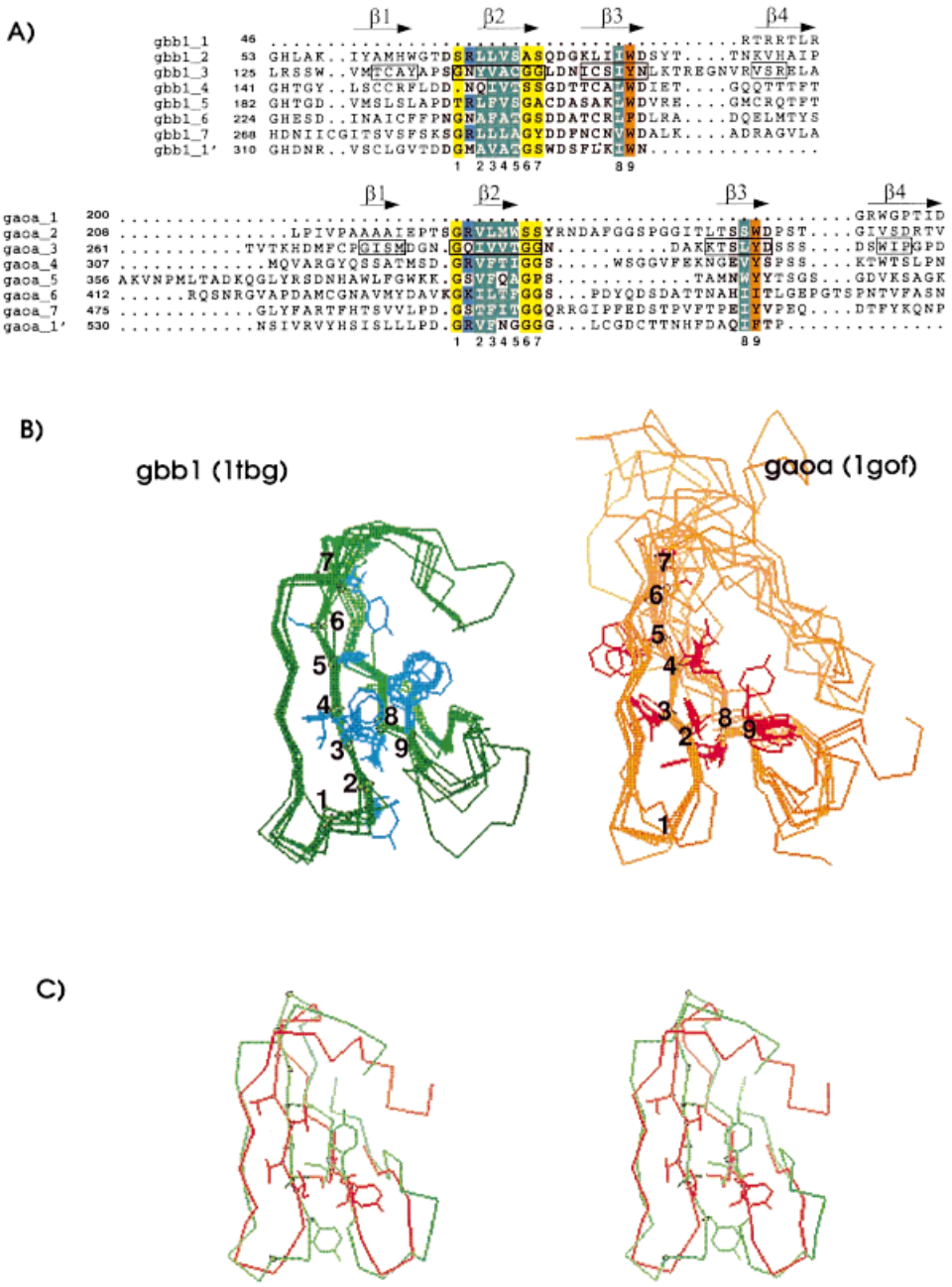


Figure 4. Sequence and 3D structure similarities between two families of repeated sequences forming β propellers (WD and kelch repeats). Primary and secondary structure alignment (*A*) and corresponding 3D superimposition (*B*) of the seven propeller blades of the β subunit of the heterotrimeric G protein (sw identifier, GBB1_HUMAN; pdb identifier, 1TBG – green) and of galactose oxidase (sw identifier, GAOA_DACDE, pdb identifier, 1GOF – orange). The conserved amino acids belonging to strands $\beta 2$ and $\beta 3$ are highlighted and numbered along the sequence. (*C*) 3D superimposition of the Cα trace of the third blades of both structures (1TBG, green; 1GOF, red). The corresponding superimposed 21 Cα belonging to $\beta 1$-$\beta 2$-$\beta 3$ and $\beta 4$ strands are boxed in (*A*) and result in an rms of 1.89 Å.

binding proteins, have been well described on the structural level. The first one, corresponding to the seven-bladed propeller of the β subunit of heterotrimeric G proteins and belonging to the WD family, makes contacts through its 2–3/1–4 side with the switch regions of the Gα subunit, whereas its other side (the 1–2/3–4

side) interacts with the γ subunit [49, 51]. The second β-propeller structure corresponds to the seven-bladed propeller of the regulator of chromosome condensation RCC1, which is a guanine-nucleotide exchange factor (GEF) for the nuclear Ras homologue Ran [53]. Although the structure of the RCC1 complex with Ran has not yet been determined, an alanine scan of invariant charged residues has provided evidence that a specific side of the propeller (the 1–2/3–4 side) constitutes the site of inter-action with the small G protein. Interestingly, the side opposite to the Ran-binding site (the 2–3/1–4 side) is thought to be involved in DNA binding. According to these observations, it is tempting to hypothesize that the β propeller of RAG2 may serve as a binding scaffold for RAG1 at one side and perhaps for DNA at the other side. Consistent with this hypothesis is the observation that the flat 2–3/1–4 surface of the WD-repeat propeller of the yeast Tup1, a protein which represses the transcription of a large number of genes, is used to contact α2, and probably various proteins bound to DNA sequences found upstream of target genes [63]. Such protein-DNA complexes could be compared with the assembly of RAG-RSS, in which RAG1 has been proposed to bind nonamer sequences distant to the cleavage site [16, 17]. Propellers of the WD family also participate in the formation of RNA-protein complexes, termed spliceo-somes, in the U4/U6 small nuclear ribonucleoprotein (snRNP) particle. Indeed, the U4/U6-associated splicing factors Prp4 and Prp3, which together play a critical role in RNA splicing, interact strongly with each other ([64], Ayadi et al., unpublished data) through the WD-repeats of Prp4 (L. Ayadi et al., unpublished data). Besides the common involvment of Prp3/Prp4 and RAG1/RAG2 complexes in nucleic acid processing, it is interesting to note that the β propeller of Prp4, lying C-terminal, is essential for splicing and cell growth, whereas the N-terminal is not [65], a situation reminiscent of the essential character of the RAG2 propeller for V(D)J recombination.

In the kelch family, to which the RAG2 sequence is related, several proteins are known or suspected to bind actin [40, 43, 58, 59, 66, 67], and several lines of evidence indicate that loops at the side of the propellers are also directly involved in this interaction. For example, the cysteine C837 of β-scruin, included in a blade-linking loop 4–1, would be directly involved in actin binding (e.g. ref. 68). However, β propellers of the kelch family are not limited to actin binding. For example, the kelch motif of the cellular host cell factor (HCF) is known to bind to the herpes simplex virus (HSV) regulatory protein VP16 [69]. This complex could be particularly relevant to the present RAG2 analysis, as the HCF β propeller has been shown to stabilize VP16-induced complex assembly with Oct-1, another cellular factor, and DNA, and to activate transcription in vivo [69]. Interestingly, a proline residue (P134) is found to be mutated to serine in a temperature-sensitive hamster cell line with cell proliferation deficiency [70]. This missense mutation, preventing association between HCF and VP16, is located in a 4–1 loop and corresponds to the second proline of the conserved motif P-x-P-R-x-G-H of the kelch family (fig. 3). This motif is well conserved in two loops linking blades in the RAG2 active core. Thus, from all these data, it is reasonable to hypothesize that the RAG2 kelch repeats could stabilize a protein-DNA complex by association with RAG1 through a β-propeller structure in which the surface formed by loops 2–3 and 1–4 could play a direct role. It is interesting to note that the interactions mediated by propellers from various families often involve the same structural interface, namely the 2–3/1–4 side, which generally includes loops of more variable lengths (as exemplified in fig. 1). A particular mutation identified in the RAG2 sequence further supports such a hypothesis. This mutation is one of of the mutations associated with the RAG sequences of patients with human severe combined immunodeficiency lacking B lymphocytes (B^-SCID) [71]. All but two of these mutations involve amino acids of the RAG1 sequence and only one, associated with a RAG1 mutation, has been found in the RAG2 active core sequence (Arg229 → Gln). This mutation was nonetheless shown to be sufficient in itself to significantly reduce V(D)J recombination activity. It is particularly interesting as it is located at the end of the exposed 2–3 loop in the fifth blade of the propeller, suggesting a functional role on the interacting surface rather than an effect on protein structure and stability.

On the other hand, β-propeller folds also seem to be well suited for totally or partly contributing active sites for enzymatic activities as exemplified in neuraminidase/sialidase [54, 55], methylamine dehydrogenase [52], methanol dehydrogenase [47] and galactose oxidase [44, 45]. Interestingly, the enzymatic sites also involve residues located on the sides of the propellers. In light of these observations, it can also be hypothesized that the β-propeller structure of RAG2 could provide a structural scaffold for displaying enzymatic activity after binding DNA near the heptamer-coding end border.

Although the active cores of RAG proteins are sufficient for recombination of plasmid substrates, recombination of the endogenous antigen receptor loci is thought to be more complex, especially as it is a carefully regulated developmental process. In this regard, the recombination machinery would be able to sense

aspects of chromatin structure. The identification, adjacent to the active core of RAG2, of a PHD-like domain, a motif known to bind chromatin components, may provide some insight for further understanding V(D)J recombination regulation.

Note added in proof. After submission of this paper, Villa and co-workers (Cell (1998) **93:** 885–896) reported on mutations in the *Rag-1* and *Rag-2* genes beared by patients with a rare autosomal recessive genetic disorder known as the Omenn syndrome. Interestingly, the two mutations which are found in the RAG2 sequence (C41W and M285R) are located on the same side of the propeller (both at the end of the 2-3 loop; see our figure 3) and are associated with a reduced interaction with RAG1. This observation provides evidence of the predicted interactive potential of the RAG2 propeller and strengthens the hypothesis that its 2-3/4-1 side may indeed directly participate to the RAG1 interaction. Our study also shows that, contrary to the Villa and coworkers hypothesis, these two mutations belong to a same domain which appears not to be related to topoisomerase II, as evidenced by further HCA analysis (data not shown). Indeed, the reported local similarity between RAG2 and topoisomerase II (41% identity on a 29 amino acid length) could not be extended to the whole considered domains and is likely due to chance. A similar conclusion was drawn for a similar local similarity ($\sim$ 50% identity on a 20 amino acid length) observed between the same sequence segment of RAG2 and the first BRCT domain of XRCC1 (ref. 36).

Acknowledgements. This work was supported by the CNRS program Physique et Chimie du Vivant. We thank J. Banroques and L. Ayadi for helpful discussions on the Prp4 protein and A. Soyer for his efficient contribution to the design of the Z-score program.

1 Gellert M. (1992) Molecular analysis of V(D)J recombination. Annu. Rev. Genet. **22:** 425–446
2 Lewis S. M. (1994) The mechanism of V(D)J joining: lessons from molecular, immunological and comparative analyses. Adv. Immunol. **56:** 27–150
3 Gellert M. (1997) Recent advances in understanding V(D)J recombination. Adv. Immunol. **64:** 39–64
4 Tonegawa S. (1983) Somatic generation of antibody diversity. Nature **302:** 575–581
5 Roth D. B., Zhu C. M. and Gellert M. (1993) Characterization of broken DNA molecules associated with V(D)J recombination. Proc. Natl. Acad. Sci. USA **90:** 10788–10792
6 Roth D. B., Menetski J. P., Nakajima P. B., Bosma M. J. and Gellert M. (1992) V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. Cell **70:** 983–991
7 Ramsden D. A. and Gellert M. (1995) Formation and resolution of double strand break intermediates in V(D)J rearrangement. Genes Dev. **9:** 2409–2420
8 Zhu C. M. and Roth D. B. (1995) Characterization of coding ends in thymocytes of scid mice: implications for the mechanism of V(D)J recombination. Immunity **2:** 101–112
9 Schatz D. G., Oettinger M. A. and Baltimore D. A. (1989) The V(D)J recombination activating gene, RAG-1. Cell **59:** 1035–1048
10 Oettinger M. A., Schatz D. G., Gorka C. and Baltimore D. (1990) RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science **248:** 1517–1523
11 Eastman Q. M., Leu T. M. J. and Schatz D. G. (1996) Initiation of V(D)J recombination in vitro obeying the 12/23 rule. Nature **380:** 85–88
12 van Gent D. C., Ramsden D. A. and Gellert M. (1996) The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell **85:** 107–113
13 McBlane J. F., van Gent D. C., Ramsden D. A., Romeo C., Cuomo C. A., Gellert M. et al. (1995) Cleavage at V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell **83:** 387–395
14 van Gent D. C., Mizuuchi K. and Gellert M. (1996) Similarities between initiation of V(D)J recombination and retroviral integration. Science **271:** 1592–1594
15 Hiom K. and Gellert M. (1997) A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage. Cell **88:** 65–72
16 Difilippantonio M. J., McMahan C. J., Eastman Q. M., Spanopoulou E. and Schatz D. G. (1996) RAG1 mediates signal sequence recognition and recruitement of RAG2 in V(D)J recombination. Cell **87:** 253–262
17 Spanopoulou E., Zaitseva F., Wang F.-H., Santagata S., Baltimore D. and Panayotou G. (1996) The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. Cell **87:** 263–276
18 van Gent D. C., Hiom K., Paull T. T. and Gellert M. (1997) Stimulation of V(D)J cleavage by high mobility group proteins. EMBO J. **16:** 2665–2670
19 Agrawal A. and Schatz D. G. (1997) RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. Cell **89:** 43–53
20 Spanopoulou E., Cortes P., Shih C., Huang C.-M., Silver D. P., Svec P. et al. (1995) Localization, interaction and RNA binding properties of the V(D)J recombination-activating proteins RAG1 and RAG2. Immunity **3:** 715–726
21 Leu T. M. J. and Schatz D. G. (1995) Rag-1 and rag-2 are components of a high molecular-weight complex, and association of rag-2 with this complex is rag-1 dependent. Mol. Cell. Biol. **15:** 5657–5670
22 Ramsden D. A., Paull T. T. and Gellert M. (1997) Cell-free V(D)J recombination. Nature **388:** 488–491
23 Sadofksy M. J., Hesse J. E., McBlane J. F. and Gellert M. (1993) Expression and V(D)J recombination activity of mutated RAG-1 proteins. Nucleic Acids Res. **21:** 5644–5650
24 Sadofsky M. J., Hesse J. E. and Gellert M. (1994) Definition of a core region of RAG-2 that is functional in V(D)J recombination. Nucleic Acids Res. **22:** 1805–1809
25 Cuomo C. A. and Ottinger M. A. (1994) Analysis of regions of RAG-2 important for V(D)J recombination. Nucleic Acids Res. **22:** 1810–1814
26 Silver D. P., Spanopoulou E., Mulligan R. C. and Baltimore D. (1993) Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. Proc. Natl. Acad. Sci. USA **90:** 6100–6104
27 van Gent D. C., McBlane J. F., Ramsden D. A., Sadofsky M. J., Hesse J. E. and Gellert M. (1995) Initiation of V(D)J recombination in a cell-free system. Cell **81:** 925–934
28 Bellon S. F., Rodgers K. K., Schatz D. G., Coleman J. E. and Steitz T. A. (1997) Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster. Nat. Struct. Biol. **4:** 586–591
29 Rodgers K. K., Bu Z., Fleming K. G., Schatz D. G., Engelman D. M. and Coleman J. E. (1996) A zinc-binding domain involved in the dimerization of RAG1. J. Mol. Biol. **260:** 70–84
30 McMahan C. J., Sadofsky M. J. and Schatz D. G. (1997) Definition of a large region of RAG1 that is important for coimmunoprecipitation of RAG2. J. Immunol. **158:** 2202–2210
31 Wang J. C., Caron P. R. and Kim R. A. (1990) The role of DNA topoisomerase in recombination and genome stability: a double-edge sword? Cell **62:** 403–406
32 Kallenbach S., Brinkmann T. and Rougeon F. (1993) RAG-1-a topoisomerase? Int. Immunol. **5:** 231–232
33 Gaboriaud C., Bissery V., Benchetrit T. and Mornon J. P. (1987) Hydrophobic cluster analysis. An efficient new way to compare and analyse amino-acid sequences. FEBS Lett. **224:** 149–155
34 Callebaut I., Labesse G., Durand P., Poupon A., Canard L., Chomilier J. et al. (1997) Deciphering protein sequence information through hydrophobic cluster analysis: current status and perspectives. Cell. Mol. Life Sci. **53:** 621–645

35 Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402

36 Callebaut I. and Mornon J.-P. (1997) From BRCA1 to RAP**1:** a widespread BRCT module closely associated to DNA repair. FEBS Lett. **400:** 25–30

37 Bork P., Hofmann K., Bucher P., Neuwald A. F., Altschul S. F. and Koonin E. V. (1997) A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. FASEB J. **11:** 68–76

38 Critchlow S. E., Bowater R. P. and Jackson S. P. (1997) Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. Curr. Biol. **7:** 588–598

39 Grawunder U., Wilm M., Wu X., Kulesza P., Wilson T. E., Mann M. et al. (1997) Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature **388:** 492–495

40 Xue F. and Cooley L. (1993) *kelch* encodes a component of intercellular bridges in *Drosophila* egg chambers. Cell **72:** 681–693

41 Bork P. and Doolittle R. F. (1994) *Drosophila kelch* motif is derived from a common enzyme fold. J. Mol. Biol. **236:** 1277–1282

42 Way M., Sanders M., Chafel M., Tu Y.-H., Knight A. and Matsudaira P. (1995) β-scruin, a homologue of the actin crosslinking protein scruin, is localized to the acrosomal vesicle of *Limulus* sperm. J. Cell Sci. **108:** 3155–3162

43 Way M., Sanders M., Garcia C., Sakai J. and Matsudaira P. (1995) Sequence and domain organization of scruin, an actin-cross-linking protein in the acrosomal process of *Limulus* sperm. J. Cell Biol. **128:** 51–60

44 Ito N., Philips S. E. V., Stevens C., Ogel Z. B., McPherson M. J., Keen J. N. et al. (1991) Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. Nature **350:** 87–90

45 Ito N., Philips S. E. V., Yadav K. D. S. and Knowles P. F. (1994) Crystal structure of a free radical enzyme, galactose oxidase. J. Mol. Biol. **238:** 794–814

46 Faber H. R., Groom C. R., Baker H. M., Morgan W. T., Smith A. and Baker E. N. (1995) 1.8 Å crystal structure of the C-terminal domain of rabbit serum haemopexin. Structure **3:** 551–559

47 Xia Z. X., Dai W. W., Xiong J. P., Hao Z. P., Davidson V. L., White S. et al. (1992) The three-dimensional structures of methanol dehydrogenase from two methylotrophic bacteria at 2.6-Å resolution. J. Biol. Chem. **267:** 22289–22297

48 Murzin A. (1992) Structural principles for the propeller assembly of β-sheets: the preference for seven-fold symmetry. Proteins **14:** 191–201

49 Lambright D. G., Sondek J., Bohm A., Skiba N. P., Hamm H. E. and Sigler P. B. (1996) The 2.0 Å crystal structure of a heterotrimeric G protein. Nature **379:** 311–319

50 Sondek J., Bohm A., Lambright D. G., Hamm H. E. and Sigler P. B. (1996) Crystal structure of a G_A protein $\beta\gamma$ dimer at 2.1 Å resolution. Nature **379:** 369–374

51 Wall M. A., Coleman D. E., Lee E., Iniguez-Lluhi J. A., Posner B. A., Gilman A. G. et al. (1995) The structure of the G protein heterotrimer $G_{i\alpha1}\beta_1\gamma_2$. Cell **83:** 1047–1058

52 Vellieux F. M. D., Huitema F., Groendijk H., Kalk K. H., Jzn J. F., Jongeigan J. A. et al. (1989) Structure of quinoprotein methylamine dehydrogenase at 2.25 Å resolution. EMBO J. **8:** 2171–2178

53 Renault L., Nassar N., Vetter I., Becker J., Klebe C., Roth M. et al. (1998) The 1.7 Å crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller. Nature **392:** 97–101

54 Varghese J. N., Laver W. G. and Colman P. M. (1983) Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. Nature **303:** 35–40

55 Crennell S. J., Garman E. F., Laver W. G., Vimr E. R. and Taylor G. L. (1993) Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. Proc. Natl. Acad. Sci. USA **90:** 9852–9856

56 Henikoff S. and Henikoff J. G. (1992) Amino acids substitution matrices from proteins blocks. Proc. Natl. Acad. Sci. USA **89:** 10915–10919

57 Kurahashi H., Akagi K., Inazawa J., Ohta T., Niikawa N., Kayatani F. et al. (1995) Isolation and characterization of a novel gene deleted in DiGeorge syndrome. Hum. Mol. Genet. **4:** 541–549

58 von Bülow M., Heid H., Hess H. and Franke W. W. (1995) Molecular nature of calicin, a major basic protein of the mammalian sperm head cytoskeleton. Exp. Cell Res. **219:** 407–413

59 Mata J. and Nurse P. (1997) tea1 and the microtubular cytoskeleton are important for generating global spatial order within the fission yeast cell. Cell **89:** 939–949

60 Aasland R., Gibson T. J. and Stewart A. F. (1995) The PHD-finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem. Sci. **20:** 56–59

61 Koken M. H. M., Saïb A. and de Thé H. (1995) A C_4HC_3 zinc finger motif. C. R. Acad. Sci. Paris, Life Sci. **318:** 733–739

62 Saha V., Chaplin T., Gregorini A., Ayton P. and Young B. D. (1995) The leukemia-associated-protein (LAP) domain, a cysteine-rich motif, is present in a wide range of proteins, including MLL, AF10, and MLLT6 proteins. Proc. Natl. Acad. Sci. USA **92:** 9737–9741

63 Komachi K. and Johnson A. D. (1997) Residues in the WD repeats of Tup1 required for interaction with α2. Mol. Cell. Biol. **17:** 6023–6028

64 Wang A., Forman-Kay J., Luo Y., Luo M., Chow Y.-H., Plumb J., et al. (1997) Identification and characterization of human genes encoding Hprp3p and Hprp4p, interacting components of the spliceosome. Hum. Mol. Genet. **6:** 2117–2126

65 Hu J., Xu Y., Schappert K., Harrington T., Wang A., Braga R. et al. (1994) Mutational analysis of the PRP4 protein of *Saccharomyces cerevisiae* suggests domain structure and snRNP interactions. Nucleic Acids Res. **22:** 1724–1734

66 Varkey J. P., Muhlrad P. J., Minniti A. N., Do B. and Ward S. (1995) The *Caenorhabditis elegans* spe-26 gene is necessary to form spermatids and encodes a protein similar to actin-associated protein kelch and scruin. Genes Dev. **9:** 1074–1086

67 Eichinger L., Bomblies L., Vandekerckhove J., Schleicher M. and Gettemans J. (1996) A novel type of protein kinase phosphorylates actin in the actin-fragmin complex. EMBO J. **15:** 5547–5556

68 Sun S., Footer M. and Matsudaira P. (1997) Modification of Cys-837 identifies an actin-binding site in the beta-propeller protein scruin. Mol. Cell. Biol. **8:** 421–430

69 Wilson A. C., Freiman R. N., Goto H., Nishimoto T. and Herr W. (1997) VP16 targets an amino-terminal domain of HCF involved in cell cycle progression. Mol. Cell. Biol. **17:** 6139–6146

70 Goto H., Motomura S., Wilson A. C., Freiman R. N., Nakabeppu Y., Fukushima K. et al. (1997) A single-point mutation in HCF causes temperature-sensitive cell-cycle arrest and disrupts VP16 function. Genes Dev. **11:** 726–737

71 Schwartz K., Gauss G. H., Ludwig L., Pannicke U., Li Z., Lindner D. et al. (1996) RAG mutations in human B cell-negative SCID. Science **274:** 97–99