Structural localization of disease-associated sequence variations in the NACHT and LRR domains of PYPAF1 and NOD2[1]

Mario Albrecht[a,*], Francisco S. Domingues[a], Stefan Schreiber[b], Thomas Lengauer[a]

[a] *Max-Planck-Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany*
[b] *First Department of Medicine, Christian-Albrechts-Universität Kiel, Schittenhelmstr. 12, 24105 Kiel, Germany*

Received 22 July 2003; accepted 2 September 2003

First published online 29 October 2003

Edited by Giovanni Cesareni

Abstract Several autoinflammatory diseases with distinct clinical manifestations have been associated with sequence variations in the gene products PYPAF1/CIAS1 and NOD2/CARD15. Both proteins belong to the PYD/CARD-containing family of apoptosis regulators and activators of pro-inflammatory caspases. To gain insight into the dysfunctional role of sequence alterations, we assembled a structure-based multiple sequence alignment of family members and related proteins. This allowed us to analyze the putative effect of the alterations on the function of nucleotide-binding (NACHT) and leucine-rich repeat (LRR) domains shared by the family members. In support of this analysis, we carefully selected template structures for the NACHT and LRR domains and mapped the genetic variations onto 3D domain models. Additionally, we propose a model of the NACHT and LRR domain complex. Our study revealed that many of the disease-associated sequence variants are located close to highly conserved sequence regions of functional relevance and are spatially adjacent in the predicted 3D structure. The implications on the domain functions such as NTP-hydrolysis or oligomerization are discussed.

Key words: PYD/CARD-containing protein; NACHT domain; Leucine-rich repeat; Structural modeling; Inflammation

1. Introduction

Several sequence variations in the gene products PYPAF1 (PYRIN-containing APAF-1-like protein) and NOD2 (nucleotide-binding oligomerization domain protein 2) have been associated with clinically distinct diseases, which are characterized pathophysiologically by similar autoinflammatory processes. The PYPAF1 and NOD2 proteins are members of a cytoplasmic protein family named CATERPILLER, which is involved in apoptosis regulation and inflammatory immune responses. Recently, several comprehensive reviews [1–11] and two online databases [12,13] have been published on this family and the related disorders.

Genetic variants of PYPAF1 are linked to chronic infantile neurological cutaneous and articular syndrome (CINCA, also known as neonatal-onset multisystem inflammatory disease NOMID), familial cold autoinflammatory syndrome (FCAS, also called familial cold urticaria FCU), and Muckle–Wells syndrome (MWS) [14–20]. Single nucleotide polymorphisms (SNPs) in the NOD2 gene predispose to Blau syndrome (BS) [21,22] and Crohn's disease (CD) [23–26], which is one of the two main types of chronic inflammatory bowel disease [27,28].

The CATERPILLER family consists of about twenty human proteins, which share a homologous domain architecture that consists of N-terminal effector-binding PYD or CARD domains, one centrally located oligomerization NACHT NTPase domain (including a yet uncharacterized C-terminal NACHT-associated domain extension NAD), and one ligand-recognition C-terminal leucine-rich repeat (LRR) domain (Fig. 1). This protein family belongs to a larger superfamily of regulators of apoptotic and stress signaling pathways, which includes another family containing the human apoptotic protease-activating factor APAF-1 and its *Caenorhabditis elegans* ortholog CED-4 [29]. This other family shows a similar domain architecture with a CARD domain followed by a NB-ARC NTPase domain (also termed AP-ATPase domain) [30,31] instead of the closely related NACHT domain [32]. Recently, the mammalian orthologs of human PYPAF1, PYPAF5, NOD1, and NOD2 were identified [33].

The PYD/PYRIN domain is named after the protein pyrin, the product of the familial Mediterranean fever gene [34,35], and is also called DAPIN or PAAD domain [36–40]. It is homologous to the CARD (caspase-activating recruitment) [41], DD (death), and DED (death-effector) domains, whose 3D structures reveal a highly conserved compact bundle of six-anti-parallel α-helices as common domain fold [42,43]. Each of the four described domain families binds to other family members through homotypic interactions in order to downstream transduce a regulatory signal on apoptosis.

The PYD domain of PYPAF1 interacts selectively with the PYD domain of the PYD-CARD adaptor protein ASC, which activates caspase-1 [44,45]. In contrast, the CARD domain of NOD2 assembles with the CARD domain of the serine-threonine kinase RICK (also called RIP2/CARDIAK) [5]. Generally, both the PYD and CARD domain families participate in the regulation of pro-inflammatory signaling pathways by the activation of NF-κB and pro-caspase-1; caspase-1 then processes pro-interleukin-1β/-18 into active cytokines. It is hypothesized that a defect in the signaling cascade

*Corresponding author.
E-mail addresses: mario.albrecht@mpi-sb.mpg.de (M. Albrecht), doming@mpi-sb.mpg.de (F.S. Domingues), s.schreiber@mucosa.de (S. Schreiber), lengauer@mpi-sb.mpg.de (T. Lengauer).

[1] Supplementary data associated with this article can be found at doi:10.1016/S0014-5793(03)01222-5

0014-5793/03/$22.00 © 2003 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)01222-5

results in an impaired innate immunity, which leads to inflammatory processes.

NACHT is an acronym for the founding family members NAIP (neuronal apoptosis inhibitor protein), CIITA (major histocompatibility complex (MHC) class II transactivator), HET-E (plant *het* gene product involved in vegetative incompatibility), and TP1 (telomerase-associated protein 1). Most of the proteins with an NB-ARC or NACHT domain mediate self-oligomerization and are predicted or shown to have ATPase activity [4], but CIITA and HET-E are GTPases [46,47]. The transactivator CIITA has an transcriptional activation domain (AD) preceding the NACHT domain and regulates the expression of MHC class II genes [48,49]. Dendritic cell-specific splice variants of CIITA contain an additional N-terminal CARD before the AD [50]. Several mutations in CIITA cause type II bare lymphocyte syndrome (BLS), a severe immunodeficiency disorder [51–57].

In contrast to the unknown function of the LRR domain in PYPAF1, the homologous LRR domains of NOD1 and NOD2 have been shown to function as specific sensors of bacterial peptidoglycan through muramyl dipeptide detection [7,58,59]. Generally, it is assumed that the LRR domain of the described proteins recognizes specific pathogen-derived molecules. This is in analogy to mammalian Toll-like receptors (TLRs) and plant disease resistance gene products (R proteins), which are all part of the surprisingly similar innate immune system of plants, animals, and man [1–11,60]. R proteins also possess a NB-ARC domain besides their extracellular C-terminal LRR domain, which they have in common with TLRs [61–63]. However, R proteins carry a non-homologous TIR (Toll and interleukin-1 receptor) or CC (coiled-coil) domain [64–70] for intracellular downstream signaling instead of a CARD or PYD domain.

The analysis of the structural localization of the sequence variants can give important insight into the potential disease mechanism on a molecular level. Because most disease-associated genetic variations are found in the NACHT domain, we assembled a manually curated multiple sequence alignment of members of the NACHT protein domain family. The alignment also includes carefully selected NTPases whose structures have been determined crystallographically and reflect specific characteristics of the NACHT domain and its secondary structure prediction. This enabled us to map the sequence variations found in the NACHT domain onto its predicted 3D structure in order to understand functional implications. In addition, we provide a hypothetical model of the NACHT and LRR domain complex and show that other sequence variants may affect the domain–domain interface. Finally, we draw conclusions on the potential dysfunctional role of the disease-associated sequence variants and suggest supportive experiments.

2. Materials and methods

All protein sequences were retrieved from the SWISS-PROT/TrEMBL (SPTrEMBL) database [71]. Their accession numbers are given in the caption of Fig. 1. The hierarchical classification and structural neighbors of protein structures downloaded from the PDB [72] were obtained from the SCOP database (release 1.63 of May 2003) [73]. The secondary structure assignments to PDB structures were taken from the DSSP database [74]. A single capital letter appended to the actual PDB identifier denotes the chosen structure chain. We employed the BLAST program [75] to search the PDB.

We used the web server PSIPRED to predict the secondary structure [76]. This server showed top performance close to a three-state accuracy of 80% according to a recent evaluation [77]. Multiple sequence alignments were constructed by means of CLUSTAL W [78] and improved manually based on the predicted secondary structure and pairwise structure superpositions as returned by the program CE (combinatorial extension) [79]. The root mean square deviations (RMSDs) with the associated sequence identities were taken from the CE superpositions. In order to obtain alternative superpositions, we employed the superposition web server ProSup [80].

We explored all standard fold recognition tools available via the online meta-server BioInfo.PL [81], which contacts a dozen other state-of-the-art prediction servers (whose names are listed on the web site http://BioInfo.PL/Meta/). Its 3D-Jury system allows for comparing and evaluating the predicted 3D models conveniently in a consensus view [82]. We also compared the 3D predictions with the results from our fold recognition server Arby [83,84]. To model the protein structure of the NACHT domain, we submitted the sequence-structure alignment to the WHAT IF modeling server [85] and applied the side chain placement program SCWRL [86] to the resulting 3D model (the side chain conformation of amino acids that are identical in both aligned sequences was preserved). Alternatively, we used the completely automatic 3D-JIGSAW modeling server [87] to obtain full-atom 3D models of the LRR domain. The sequence alignments shown in the figures were prepared in the SEAVIEW editor [88] and illustrated using the web tool ESPript [89]. The protein structure images were drawn in the Accelrys Discovery Studio ViewerLight. The online version of this article contains supplementary material, and our web site will provide additional pictures.

3. Results and discussion

3.1. Multiple sequence alignment of NACHT domains

We compiled a list of human PYD/CARD-containing NACHT-LRR proteins as follows: PYPAF1-8, DEFCAP, CLAN, NOD1/2, CIITA (synonymous protein names are listed in the caption of Fig. 1). We then assembled a multiple sequence alignment of their NACHT domains (Fig. 2 and

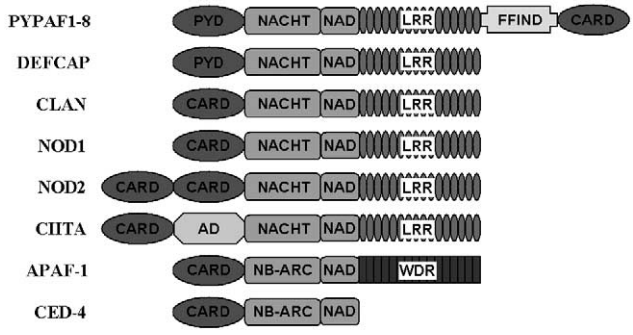

Fig. 1. Domain architectures of the human proteins PYPAF1-8, DEFCAP, CLAN, NOD1, NOD2, CIITA, APAF-1, and *C. elegans* CED-4. Note that the NACHT domain is an NTPase domain with C-terminal helical extension subdomains of an additional NACHT-associated domain NAD, and DEFCAP has additional C-terminal domains, including an intermediate FFIND ('function to find') domain. Also, the proteins actually have a slightly differing number of LRR units. In contrast to APAF-1, CED-4 lacks the WDR (WD-40 repeat) domain, which substitutes the LRR domain in APAF-1. The SPTrEMBL accession numbers and synonymous names of the selected proteins are as follows: PYPAF1/CIAS1/CRYOPYRIN/NALP3, Q96P20; PYPAF2/NALP2/NBS1/PAN1, Q9NX02; PYPAF3/NALP7/NOD12, Q8WX94; PYPAF4/NALP4/PAN2, Q96MN2; PYPAF5/NALP6/PAN3, P59044; PYPAF6/NALP11/NOD17, P59045; PYPAF7/MONARCH-1/NALP12/PAN6, P59046; PYPAF8/MATER/NALP5, P59047; DEFCAP/CARD7/NALP1/NAC, Q9C000; CLAN/CARD12/IPAF, Q9NPP4; NOD1/CARD4, Q9Y23; NOD2/CARD15/IBD1, Q9HC29; CIITA/MHC2TA, P33076; APAF-1, O14727; CED-4, P30429.

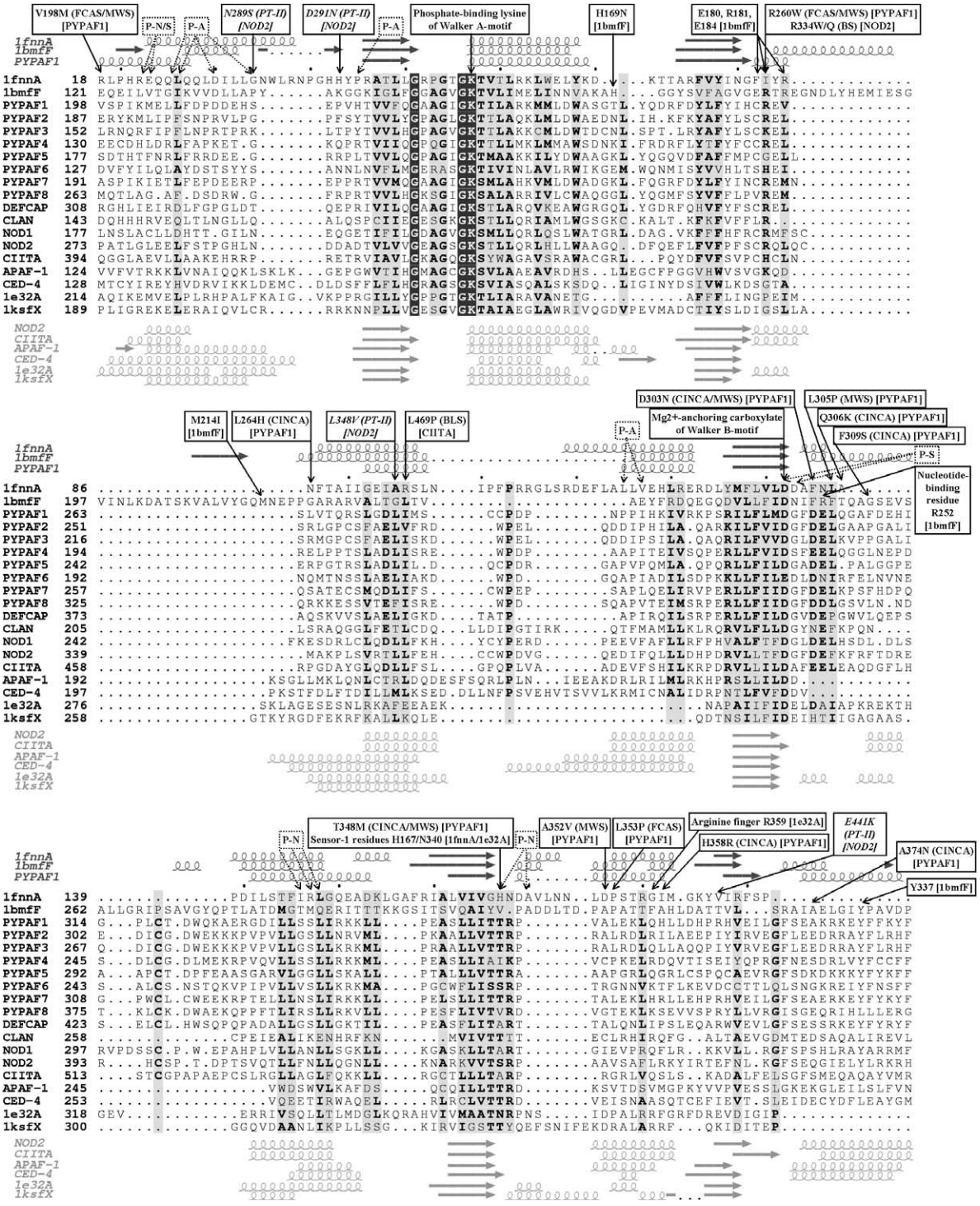

Fig. 2. Structure-based multiple sequence alignment of NACHT and NB-ARC domains including the PDB structures 1fnnA, 1bmfF, 1e32A, and 1ksfX (a multiple alignment covering the NAD extension is shown in Web Fig. A). The DSSP secondary structures and the corresponding predictions by the PSIPRED server are depicted in the upper and lower part of each alignment row (α-helices are represented by curled lines, β-strands by horizontal arrows). The alignment columns with strictly conserved residues are highlighted in dark gray boxes, those in which more than 60% of the residues are physico-chemically equivalent are shown in light gray boxes. Solid text labels denote disease-associated variants and other sequence variations. Dotted labels indicate alanine substitutions of a mutagenesis study with the close homolog CDC18 of CDC6 (1fnnA). Important residues involved in nucleotide-binding are also annotated.

Web Fig. A) and improved it manually based both on the secondary structure prediction for PYPAF1, NOD2, and CIITA and pairwise superpositions of related PDB structures, whose selection is described below. We also included APAF-1 and CED-4 with their NACHT-related NB-ARC domain and the corresponding secondary structure predictions into the alignment. The multiple alignment indicates that the NAD extension of the NACHT domain consists of three helical subdomains NAD1, NAD2, and NAD3, which are connected by linker regions of highly variable lengths. Sequence identi-

ties derived from the multiple alignment of the NTPase domains (Fig. 2) can be found in the Web Table.

3.2. Selection of modeling template

Based on the results that were returned by our in-house fold recognition server Arby and by the BioInfo.PL meta-server and its 3D-Jury system, we selected the PDB structures 1fnnA, 1e32A, and 1ksfX/1jbkA as modeling templates for the extended NACHT domain of PYPAF1 and NOD2. The ATPase domains of 1fnnA, 1e32A, and 1ksfX/1jbkA are the N-terminal domains of two consecutive ATPase domains of the cell division control protein Cdc6 from *Pyrobaculum aerophilum* [90], of the murine membrane fusion ATPase p97 (also known as vasolin-containing protein VCP) [91], and of the highly homologous *Escherichia coli* chaperone heat shock HSP100 family proteases ClpA/ClpB [92,93], respectively.

The ATPase domains of 1fnnA, 1e32A, and 1ksfX/1jbkA are contained in the same SCOP family, namely that of extended AAA-ATPases, which possess helical extension subdomains. The ATPase domains of 1fnnA, 1e32A, and 1ksfX/1jbkA superimpose very well (Web Fig. B) despite low sequence identities. In detail, using the ATPase domain of 1fnnA from R18 to P192, 1fnnA and 1e32A superimpose with an RMSD of 2.8 Å (seq. id. 19%), and 1fnnA and 1ksfX/1jbkA with an RMSD of 3.3 Å/3.1 Å (seq. id. 13%/12%). For the description of an additional manual verification of the automated template selection, see the supplementary online material. This includes a discussion of the bovine mitochondrial F_1-ATPase (PDB identifier 1bmfF) [94] and other more distantly related structural neighbors of the modeling templates.

3.3. Characteristics of NTPases and the NACHT domain

All the selected structural templates are members of the large AAA+ family of ATPases, whose sequence motifs of structural and functional relevance have been discussed extensively in the literature [95–108]. The ATPase domain of 1fnnA possesses two consecutive C-terminal helical extensions as subdomains similar to the putative NAD1 and NAD2 subdomains of the CATERPILLER family. However, it remains unclear whether they are homologous, and we refrained from aligning them based on secondary structure information only because of insufficient sequence identity.

The numerous NTPases share a structural core of a central pleated β-sheet formed by at least four parallel β-strands flanked by α-helices. The overall shape of the nucleotide-binding site near the C-termini of the β-strands is strongly preserved in all NTPase families despite minor topological differences regarding the number and sequential order of β-strands and α-helices.

Most NTPases contain two well-characterized consecutive nucleotide-binding signatures, the Walker A- and B-motifs [109] (see Fig. 2). The P-loop in the Walker A-motif (consensus sequence A/GxxxxGKT/S, x for any amino acid) wraps around the polyphosphate moiety of the nucleotide bound by a strictly conserved lysine. The P-loop is usually preceded by a β-strand and followed by an α-helix. The Walker B-motif consists of one to three conserved aspartates/glutamates at the C-terminus of a β-strand. The first strand is commonly involved in anchoring a magnesium ion, which chelates the bound nucleotide, by hydrogen bonds to Mg^{2+}-coordinating water molecules. The second (or third) aspartate/glutamate usually provides the catalytic carboxylate for NTP-hydrolysis. Another interesting motif is the sensor-1 region, which contains a conserved polar residue H/N/T/S that contacts the γ-phosphate of the bound NTP and is thought to discriminate between ATP and ADP. Particularly, the N-terminal ATPase domains of 1fnnA, 1e32A, and 1ksfX/1jbkA are involved in oligomerization of the monomers upon ATP-binding and subsequent ATP-hydrolysis, both of which induce intricate conformational changes in the 3D structures [90–93].

3.4. Comparison of structural features

We compared the highly similar secondary structure predictions of PYPAF1, NOD2, and CIITA to the known secondary structures of NTPases (Fig. 2). As expected, the secondary structures of 1fnnA, 1e32A, and 1ksfX/1jbkA match the prediction of the NACHT and NAD domains very well (while the additional α-helix and β-strand of 1bmfF seems not be contained in the NACHT domain). Furthermore, the helix preceding the Walker B-motif in 1fnnA is missing in 1e32A and 1ksfX/jbkA, but is predicted for the NACHT and NB-ARC domain. This helix, which follows a strictly conserved proline, corresponds to the C-terminal half of the aligned longer helix of 1bmfF (Fig. 2 and Web Fig. E).

Altogether, the structure 1fnnA appears to be the best template to model the overall 3D structure of the NACHT domain. Unfortunately, the relatively long loop that is predicted downstream from the Walker B-motif of the NACHT domain and that contains several disease-associated variants is much shorter in 1fnnA and only slightly longer in 1e32A and 1ksfX/1jbkA (Fig. 2). The β-strand and the α-helix at the ends of this loop in 1e32A, 1ksfX/1jbkA, and 1bmfF match the corresponding secondary structure elements in 1fnnA (Web Fig. D). Therefore, this loop may be modeled by a longer loop from another template such as 1bmfF, resulting in a hybrid, but less reliable model.

G-proteins have been previously proposed as modeling templates for the NB-ARC domain of APAF-1 and CED-4 [110,111]. However, the secondary structure elements and the topology of G-proteins do not match the identified structural features of the NACHT domain as accurately as our AAA-ATPase templates do. The supplementary online material contains an additional discussion of structural differences between the NACHT and the closely related NB-ARC domain, including a comprehensive comparison of our selected template structures to G-proteins.

3.5. Annotation of the multiple sequence alignment

We annotated the structure-based multiple sequence alignment in Fig. 2 and Web Fig. A with sequence variations and critical residues in PYPAF1 and NOD2 and the selected template structures. In addition to the known genetic variations associated with BS and CD, we included other rare variants into the multiple sequence alignment that are observed in NOD2 with CD patients and showed two prototypes of protein dysfunction in experiments [58]: either a two-fold reduction in both basal NF-κB activity and peptidoglycan-induced response (PT-II) or an unchanged basal activity, but a major impairment of the peptidoglycan-response (PT-III). We also annotated the alignment in Fig. 2 with mutagenesis data that are available for the close yeast homolog Cdc18 of our template CDC6 (1fnnA). Alanine-substitution mutations of Cdc18 result in three types of phenotypic defect in the cell

cycle as detailed in [90]: null phenotype characterized by a complete loss of function (P-N), partial function where cells arrested with a 2C DNA content (P-A), and slow progression through S phase due to a failure in the checkpoint response (P-S).

We used the multiple alignment to map the locations of the disease-associated sequence alterations contained in the NACHT and NAD domains of PYPAF1 and NOD2 onto 1fnnA and 1bmfF. In particular, we modeled the 3D structure NACHT domain for PYPAF1, NOD2, and CIITA using 1fnnA as template. To this end, we extracted pairwise sequence-structure alignments from our multiple alignment

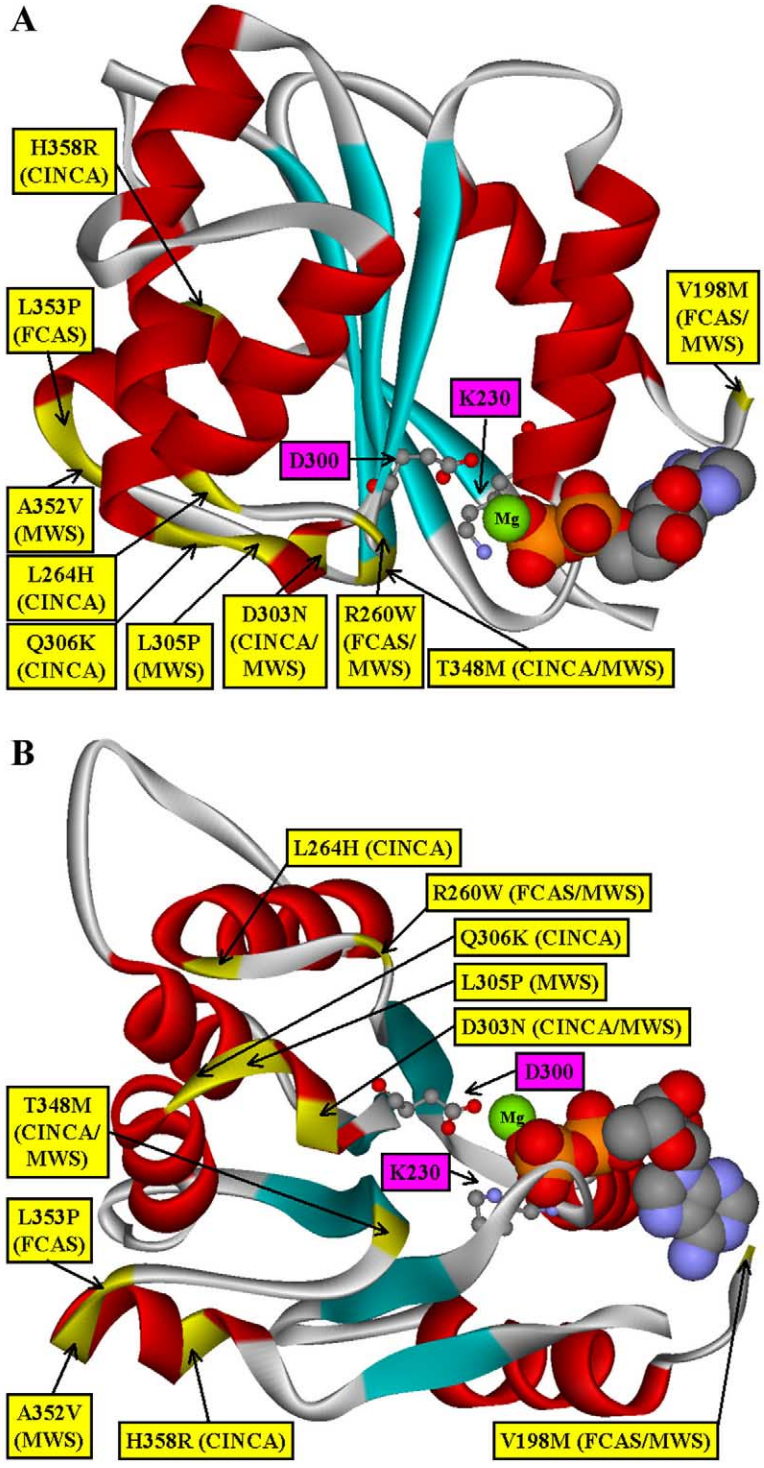

Fig. 3. Structural model of the NACHT domain of PYPAF1 (based on the PDB template 1fnnA). While α-helices are colored in red and β-strands in blue, the locations of disease-associated variants are annotated in yellow. The phosphate-binding lysine K230 and the Mg^{2+}-anchoring aspartate D300 are shown as balls-and-sticks together with the bound nucleotide. Note that a long loop in the NACHT domain after Q306K could not be modeled using 1fnnA. A: Large cluster of variants near the C-termini of the β-strands shown at the bottom of the picture. B: Rotated view of the model.

and constructed structural models based on this template. Alignment insertions and deletions in the modeled 3D structure are located predominantly in predicted loop regions and most likely do not affect the structural core. Fig. 3 shows the resulting model for PYPAF1, annotated with disease-associated variant residues. Web Fig. F provides another view by mapping the variants onto the 3D structures of 1fnnA and 1bmfF.

3.6. Localization of sequence variants in the NACHT domain

The two most interesting sequence variations associated with distinct diseases are R260W (FCAS/MWS) in PYPAF1 and R334W/Q (BS) in NOD2, which are aligned at corresponding positions (Fig. 2 and Web Fig. A). Together with the disease variant L264H of PYPAF1 (CINCA), they are located inside a loop near the NTPase active site after the C-terminus of the preceding β-strand. The three residues E180, R181, and E184 contained in this region of the F_1-ATPase structure 1bmfF are assumed to be involved in binding the nucleotide [94,112]. The proline substitution of the variant L469P (BLS) of CIITA may disrupt the predicted α-helix.

Another cluster of disease-associated SNPs found in PYPAF1 consists of D303N (CINCA/MWS), L305P (MWS), Q306K (CINCA), and F309S (CINCA). They are located close to the Walker B-motif (Figs. 2 and 3) of the nucleotide-binding site at the C-terminus of the preceding β-strand. The motif includes the Mg^{2+}-anchoring aspartate D132 of 1fnnA (in 1bmfF, the aspartate corresponds to D248 and an additional arginine R252 [94,112]). Interestingly, the variant D303N is associated with the two clinically different autoinflammatory diseases CINCA and MWS. Accordingly, it has been found with a patient who shows an overlap syndrome with clinical manifestations belonging to both diseases [113].

The variant T348M is also associated with both CINCA and MWS [20] and locates next to a highly conserved threonine at the C-terminus of another β-strand in the NACHT domain. We propose that this threonine is equivalent to the sensor-1 residues H167 in 1fnnA and N349 in 1e32A [101], which are both aligned to T348. In contrast, we could not identify a conserved arginine as a so-called sensor-2 residue in the first subdomain NAD1. This sensor-2 is found in the first helical extension subdomain of several ATPases (such as R241 of 1fnnA) and plays an essential role in oligomerization and ATPase function [105].

We could not find a conserved arginine finger either. Such a finger is typical of wheel-like hexameric or heptameric ring-forming helicases and resides as R359 in 1e32A at the N-terminus of a helix [101]. This finger plays a critical role in cooperative NTP-hydrolysis and may transduce the chemical energy of NTP-hydrolysis into conformational changes of the neighboring protomer upon self-oligomerization [101]. However, there are arginines at slightly different positions adjacent to the mapped arginine of 1e32A in the NACHT domain.

The four missense variants T348M (CINCA/MWS), A352V (MWS), L353P (FCAS) and H358R (CINCA) in PYPAF1 lie near the sensor-1 and arginine-finger regions around a helix that connects two β-strands. The variant A374N (CINCA) is located close to the nucleotide-binding site at the C-terminus of a predicted, conserved loop after the last β-strand of the central β-sheet, where Y337 of 1bmfF is known to be implicated in nucleotide-binding.

An analysis of further variants such as V198M (FCAS/MWS) of PYPAF1 and other sequence alterations contained in the NAD extension subdomains is provided as supplementary online material. However, their structural location and functional role is difficult to judge due to the lack of a reliable template structure.

3.7. Model of the NACHT and LRR domain complex

While all diseases except of CD and BLS seem to be associated solely with sequence variants in the NACHT and NAD domains, CD is associated with SNPs in the LRR domain. Also, BLS is often caused by deletions of partial LRR units, which are tandem repeats of one β-strand followed by one α-helix and exhibit the consensus motif LxxLxLxxN/CxL (x can be any amino acid, L can be substituted by V/I/F) [114]. We modeled the LRR domain for NOD2 by means of the automatic modeling server 3D-JIGSAW. The sequence-structure alignment of NOD2 with the template 1a4yA, the ribonuclease A inhibitor (RNI) [115], which also binds to angiogenin [116], is shown in Web Fig. G. This template receives the most highly significant *E*-value of 10^{-10} (seq. id. 30%) during a PDB search by BLAST with the NOD2 sequence.

It is tempting to speculate that the crystallographically determined complex of an LRR domain and an NTPase domain provides a reasonable model of the NACHT and LRR domain complex either within the same protein or between protomers after dimerization [117]. It is known that the NACHT and LRR domains of CIITA mediate self-association and bind to each other [118]. Indeed, the PDB contains such a complex, the GTPase Ran and the Ran GTPase activating protein RanGAP (PDB identifier 1k5d, chains A and C, respectively) [119]. RanGAP is contained in the same SCOP superfamily [114] as the template RNI, but the sequence similarity to the LRR domain of NOD2 is lower (seq. id. 25%). However, the two corresponding chains 1kj5dC and 1a4yA superimpose very well with an RMSD of 2.6 Å (seq. id. 17%).

Therefore, we superimposed the models of the NACHT and LRR domains of NOD2 with the structures of human Ran and RanGAP, respectively, in order to construct a model of the NACHT and LRR domain complex (Fig. 4 and Web Figs. H/I and J). However, the model complex needs to be regarded with some caution because the NACHT domain model using the template 1fnnA does not superimpose very well with RAN 1k5dA (RMSD 5.2 Å). In addition, there are alternative sequence-structure alignments of the LRR domain of NOD2 to the RNI structure because of the internal LRRs. To account for this problem, we employed the ProSup server instead of the CE program because it provides alternative superposition alignments. This enabled us to select the best superposition of the LRR domain model with RanGAP. The corresponding structure alignment has the highest number of 61 identical residues with an RMSD of 2.4 Å.

Interestingly, the GTPase Ran superimposes well with the GTPase Ras (RMSD 1.8 Å, seq. id. 26%), which has been chosen previously as modeling template for APAF-1/CED-4 (for details see the supplementary online material). Both GTPases belong to the G-protein family within the SCOP superfamily that contains our NACHT domain template 1fnnA.

3.8. Effects of sequence variants in the LRR domain

The predicted models of the NACHT and LRR domains

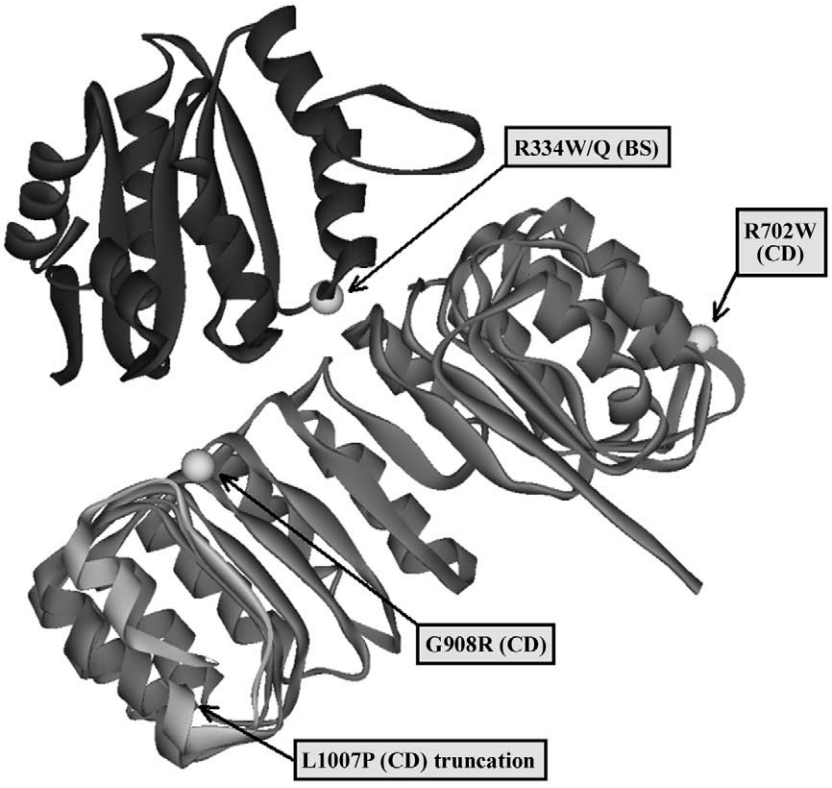

Fig. 4. Model of the complex of the NACHT (top left) and LRR (bottom right) domains of NOD2 (similar illustrations are shown in Web Figs. H/I and J). The modeled locations of the four sequence variants R334W/Q (BS), R702W (CD, SNP8), G908R (CD, SNP12), and L1007P (CD, SNP13) are annotated in light gray. The latter SNP13 leads to a frameshift and truncates the C-terminus of NOD2.

and of their complex help to understand the adverse effects of different sequence variants. For instance, the variant G908R (SNP12) associated with CD is located near the predicted interaction interface of the two domains (Fig. 4). This SNP is close to the residues D225 and N226 of RanGAP (Web Figs. I and J), which interact with K130 of Ran [119]. Therefore, SNP12 can affect the downstream signaling from the peptidoglycan-sensing LRR units to the CARD domain through the NACHT domain. This hypothesis agrees well with the experimental measurement of both a reduced basal NF-κB activity and a decreased peptidoglycan-induced response due to a constitutive defect in NOD2 function caused by SNP12 [58]. It is not yet possible to interpret the constitutive defect [58] of the remaining variant R702W (SNP8) in CD because this would require a more complete domain interaction model including the NAD extension of the NACHT domain.

The CD-associated variant L1007P (SNP13) leads to a frameshift and the insertion of a premature stop codon. This mutation results in a drastic structural change with a truncated LRR domain of NOD2 and the deletion of one C-terminal LRR unit outside of the NACHT and LRR domain interaction interface (Fig. 4 and Web Figs. H/I and J). This deletion can be expected to directly affect the peptidoglycan-sensor function of the LRR domain as it is observed experimentally without an impairment of the basal NF-κB activity [58].

The model of the NACHT and LRR domain complex can also be used to interpret the results of an experimental alanine-scanning mutagenesis study on the LRR domain of CIITA. These experiments indicate that mutations within and around the loops between the β-strand and the following α-helix of an LRR unit have a detrimental effect on CIITA function [118,120,121]. These loops are oriented towards the NACHT domain in the complex model, and some of them could be involved in domain interactions. Other disease-associated variants such as E627G (FCAS) and M662T (CINCA) of PYPAF1 may also affect those loops based on the predicted secondary structure (Web Fig. A).

3.9. *Effects of sequence variants in the NACHT domain*

In the NACHT domain, most variants associated with clinically distinct diseases form clusters in adjacent loops close to the nucleotide-binding C-termini of the β-strands that form the central β-sheet of NTPases (Fig. 3). It is striking that they are located on the same side of the protein surface and that some of them are contained in the active site such as V198M (FCAS/MWS) and T348M (CINCA/MWS) of PYPAF1. The different variants can either disturb NTP-binding and -hydrolysis by unfavorable structural changes near the active site or interfere with domain–domain interactions within the same protein or between two proteins of a dimer. Indeed, the NOD2 sequence alteration R334W/Q (BS), which corresponds to R260W (FCAS/MWS) in PYPAF1, is known to cause a constitutive dysfunction of NOD2 as revealed by a measured four-fold increase of the basal NF-κB activity [58]. This sequence alteration is located both close to the NTPase active site and near the predicted interaction interface of the NACHT and LRR domains (Figs. 3 and 4 and Web Figs. H/I and J).

Other sequence variations such as H358R (CINCA) of PYPAF1 are farther away from the nucleotide-binding site (Fig. 3). However, it is known from mutagenesis experiments

with the β-subunit of the F_1-ATPase (1bmfF) that even mutations of amino acids near residues contained in the active site can indirectly affect the conformation of those residues and thus impair the ATP-hydrolysis considerably [122,123] (M209I of *E. coli* F_1-ATPase and H211N of *Saccharomyces cerevisiae* F_1-ATPase correspond to M214I and H169N in 1bmfF, respectively, see Fig. 2). It is apparent that none of the disease-associated variants is mapped to essential residues such as the phosphate-binding lysine in the Walker A-motif or the Mg^{2+}-anchoring aspartate in the Walker B-motif. This fact indicates that NTP-function can be affected by some of the variants, but is never disabled completely.

4. Conclusions

Using structure-based sequence alignments and models of the NACHT and LRR domains and their possible complex formation, we found disease-associated sequence variants close to highly conserved alignment regions and spatially adjacent in the predicted 3D structure, pointing to very similar dysfunctional roles. In particular, several variants are observed in the loops that follow the C-termini of the central β-strands. Some of the variants in the vicinity of the NTPase active site can interfere with nucleotide-binding, -hydrolysis, and -release, albeit none has been mapped to residues that are very critical for NTP-function. Other sequence alterations may disturb an intricate signal transduction mechanism from the LRR to the PYD/CARD domain through the NACHT domain, including essential domain–domain interactions, self-oligomerization upon NTP-binding and important conformational changes induced by NTP-hydrolysis.

Our findings are consistent with published experimental observations on dysfunctional effects of sequence alterations such as a significant change of the basal NF-κB activity or a reduced response to bacterial peptidoglycan. In this context, it would be very interesting to see related experiments such as the measurement of the NTPase activity for each sequence variant. It may also be worthwhile to investigate some highly conserved residues of PYPAF1/NOD2 in further mutagenesis studies, for instance, T347/T424 of PYPAF1/NOD2 as a potential sensor-1 residue. Because the variants R260W (FCAS/MWS) of PYPAF1 and R334W/Q (BS) of NOD2 correspond to each other, it could be checked experimentally whether R260W shows a similar dysfunctional effect like R334W/Q, indicated by an increase of basal NF-κB activity. Generally, our structure-based multiple sequence alignment and the obtained 3D models can aid in the analysis of the functional relevance of further sequence variants, which are found with patients and may be associated with diseases.

Acknowledgements: Part of the research has been funded by the German Research Foundation (DFG) under contract number LE 491/14-1 (M.A.) and by the National Genome Research Network (NGFN).

References

[1] Inohara, N. and Núñez, G. (2001) Oncogene 20, 6473–6481.
[2] Girardin, S.E., Sansonetti, P.J. and Philpott, D.J. (2002) Trends Microbiol. 10, 193–199.
[3] Gumucio, D.L., Diaz, A., Schaner, P., Richards, N., Babcock, C., Schaller, M. and Cesena, T. (2002) Clin. Exp. Rheumatol. 20, S45–53.
[4] Harton, J.A., Linhoff, M.W., Zhang, J. and Ting, J.P. (2002) J. Immunol. 169, 4088–4093.
[5] Inohara, N., Ogura, Y. and Núñez, G. (2002) Curr. Opin. Microbiol. 5, 76–80.
[6] McDermott, M.F. (2002) Trends Mol. Med. 8, 550–554.
[7] Chamaillard, M., Girardin, S.E., Viala, J. and Philpott, D.J. (2003) Cell Microbiol. 5, 581–592.
[8] Hull, K.M., Shoham, N., Chae, J.J., Aksentijevich, I. and Kastner, D.L. (2003) Curr. Opin. Rheumatol. 15, 61–69.
[9] Inohara, N. and Núñez, G. (2003) Nat. Rev. Immunol. 3, 371–382.
[10] Reed, J.C. et al. (2003) Genome Res. 13, 1376–1388.
[11] Tschopp, J., Martinon, F. and Burns, K. (2003) Nat. Rev. Mol. Cell Biol. 4, 95–104.
[12] Doctor, K.S., Reed, J.C., Godzik, A. and Bourne, P.E. (2003) Cell Death Differ. 10, 621–633.
[13] Sarrauste de Menthière, C., Tèrriere, S., Pugnère, D., Ruiz, M., Demaille, J. and Touitou, I. (2003) Nucleic Acids Res. 31, 282–285.
[14] Hoffman, H.M., Mueller, J.L., Broide, D.H., Wanderer, A.A. and Kolodner, R.D. (2001) Nat. Genet. 29, 301–305.
[15] Aganna, E. et al. (2002) Arthritis Rheum. 46, 2445–2452.
[16] Aksentijevich, I. et al. (2002) Arthritis Rheum. 46, 3340–3348.
[17] Dodé, C. et al. (2002) Am. J. Hum. Genet. 70, 1498–1506.
[18] Feldmann, J., Prieur, A.M., Quartier, P., Berquin, P., Certain, S., Cortis, E., Teillac-Hamel, D., Fischer, A. and de Saint Basile, G. (2002) Am. J. Hum. Genet. 71, 198–203.
[19] Hoffman, H.M., Gregory, S.G., Mueller, J.L., Tresierras, M., Broide, D.H., Wanderer, A.A. and Kolodner, R.D. (2003) Hum. Genet. 112, 209–216.
[20] Rösen-Wolff, A., Quietzsch, J., Schröder, H., Lehmann, R., Gahr, M. and Roesler, J. (2003) Eur. J. Haematol. 71, 214–219.
[21] Miceli-Richard, C., Lesage, S., Rybojad, M., Prieur, A.M., Manouvrier-Hanu, S., Häfner, R., Chamaillard, M., Zouali, H., Thomas, G. and Hugot, J.P. (2001) Nat. Genet. 29, 19–20.
[22] Wang, X., Kuivaniemi, H., Bonavita, G., Mutkus, L., Mau, U., Blau, E., Inohara, N., Núñez, G., Tromp, G. and Williams, C.J. (2002) Arthritis Rheum. 46, 3041–3045.
[23] Hampe, J. et al. (2001) Lancet 357, 1925–1928.
[24] Hugot, J.P. et al. (2001) Nature 411, 599–603.
[25] Ogura, Y. et al. (2001) Nature 411, 603–606.
[26] Lesage, S. et al. (2002) Am. J. Hum. Genet. 70, 845–857.
[27] Bouma, G. and Strober, W. (2003) Nat. Rev. Immunol. 3, 521–533.
[28] Schölmerich, J. (2003) Eur. J. Gastroenterol. Hepatol. 15, 585–586.
[29] Adams, J.M. and Cory, S. (2002) Curr. Opin. Cell Biol. 14, 715–720.
[30] van der Biezen, E.A. and Jones, J.D. (1998) Curr. Biol. 8, R226–227.
[31] Aravind, L., Dixit, V.M. and Koonin, E.V. (1999) Trends Biochem. Sci. 24, 47–53.
[32] Koonin, E.V. and Aravind, L. (2000) Trends Biochem. Sci. 25, 223–224.
[33] Albrecht, M., Domingues, F.S., Schreiber, S. and Lengauer, T. (2003) FEBS Lett. 538, 173–177.
[34] The international FMF consortium (1997) Cell 90, 797–807.
[35] Centola, M., Aksentijevich, I. and Kastner, D.L. (1998) Hum. Mol. Genet. 7, 1581–1588.
[36] Bertin, J. and DiStefano, P.S. (2000) Cell Death Differ. 7, 1273–1274.
[37] Fairbrother, W.J., Gordon, N.C., Humke, E.W., O'Rourke, K.M., Starovasnik, M.A., Yin, J.P. and Dixit, V.M. (2001) Protein Sci. 10, 1911–1918.
[38] Martinon, F., Hofmann, K. and Tschopp, J. (2001) Curr. Biol. 11, R118–120.
[39] Pawlowski, K., Pio, F., Chu, Z., Reed, J.C. and Godzik, A. (2001) Trends Biochem. Sci. 26, 85–87.
[40] Staub, E., Dahl, E. and Rosenthal, A. (2001) Trends Biochem. Sci. 26, 83–85.
[41] Hofmann, K., Bucher, P. and Tschopp, J. (1997) Trends Biochem. Sci. 22, 155–156.
[42] Weber, C.H. and Vincenz, C. (2001) Trends Biochem. Sci. 26, 475–481.

[43] Liepinsh, E., Barbals, R., Dahl, E., Sharipo, A., Staub, E. and Otting, G. (2003) J. Mol. Biol. 332, 1155–1163.
[44] Manji, G.A. et al. (2002) J. Biol. Chem. 277, 11570–11575.
[45] Srinivasula, S.M., Poyet, J.L., Razmara, M., Datta, P., Zhang, Z. and Alnemri, E.S. (2002) J. Biol. Chem. 277, 21119–21122.
[46] Espagne, E., Balhadere, P., Begueret, J. and Turcq, B. (1997) Mol. Gen. Genet. 256, 620–627.
[47] Harton, J.A., Cressman, D.E., Chin, K.C., Der, C.J. and Ting, J.P. (1999) Science 285, 1402–1405.
[48] Accolla, R.S., De Lerma Barbaro, A., Mazza, S., Casoli, C., De Maria, A. and Tosi, G. (2001) Trends Immunol. 22, 560–563.
[49] Ting, J.P. and Trowsdale, J. (2002) Cell 109 Suppl., S21–33.
[50] Nickerson, K., Sisk, T.J., Inohara, N., Yee, C.S., Kennell, J., Cho, M.C., Yannie II, P.J., Núñez, G. and Chang, C.H. (2001) J. Biol. Chem. 276, 19089–19093.
[51] Steimle, V., Otten, L.A., Zufferey, M. and Mach, B. (1993) Cell 75, 135–146.
[52] Bontron, S., Steimle, V., Ucla, C., Eibl, M.M. and Mach, B. (1997) Hum. Genet. 99, 541–546.
[53] Cressman, D.E., Chin, K.C., Taxman, D.J. and Ting, J.P. (1999) Immunity 10, 163–171.
[54] Quan, V., Towey, M., Sacks, S. and Kelly, A.P. (1999) Immunogenetics 49, 957–963.
[55] Wiszniewski, W. et al. (2001) J. Immunol. 167, 1787–1794.
[56] Dziembowska, M. et al. (2002) Immunogenetics 53, 821–829.
[57] Patarroyo, J.C., Stuve, O., Piskurich, J.F., Hauser, S.L., Oksenberg, J.R. and Zamvil, S.S. (2002) Genes Immun. 3, 34–37.
[58] Chamaillard, M. et al. (2003) Proc. Natl. Acad. Sci. USA 100, 3455–3460.
[59] Girardin, S.E., Travassos, L.H., Hérve, M., Blanot, D., Boneca, I.G., Philpott, D.J., Sansonetti, P.J. and Mengin-Lecreulx, D. (2003) J. Biol. Chem. 278, 41702–41708.
[60] Underhill, D.M. (2003) Eur. J. Immunol. 33, 1767–1775.
[61] Armant, M.A. and Fenton, M.J. (2002) Genome Biol. 3, 3011.1–3011.6.
[62] Dunne, A. and O'Neill, L.A. (2003) Sci. STKE 171, re3.1–re3.17.
[63] Akira, S. (2003) Curr. Opin. Immunol. 15, 5–11.
[64] Ellis, J. and Jones, D. (1998) Curr. Opin. Plant. Biol. 1, 288–293.
[65] Dixon, M.S., Golstein, C., Thomas, C.M., van Der Biezen, E.A. and Jones, J.D. (2000) Proc. Natl. Acad. Sci. USA 97, 8807–8814.
[66] Shirasu, K. and Schulze-Lefert, P. (2000) Plant Mol. Biol. 44, 371–385.
[67] Young, N.D. (2000) Curr. Opin. Plant Biol. 3, 285–290.
[68] Menezes, H. and Jared, C. (2002) Comp. Biochem. Physiol. C. Toxicol. Pharmacol. 132, 1–7.
[69] Nürnberger, T. and Brunner, F. (2002) Curr. Opin. Plant. Biol. 5, 318–324.
[70] Holt, B.F., Hubert, D.A. and Dangl, J.L. (2003) Curr. Opin. Immunol. 15, 20–25.
[71] Boeckmann, B. et al. (2003) Nucleic Acids Res. 31, 365–370.
[72] Westbrook, J., Feng, Z., Chen, L., Yang, H. and Berman, H.M. (2003) Nucleic Acids Res. 31, 489–491.
[73] Lo Conte, L., Brenner, S.E., Hubbard, T.J., Chothia, C. and Murzin, A.G. (2002) Nucleic Acids Res. 30, 264–267.
[74] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[75] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[76] McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) Bioinformatics 16, 404–405.
[77] Rost, B. and Eyrich, V.A. (2001) Proteins 45 (Suppl. 5), 192–199.
[78] Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. (2003) Nucleic Acids Res. 31, 3497–3500.
[79] Shindyalov, I.N. and Bourne, P.E. (1998) Protein Eng. 11, 739–747.
[80] Lackner, P., Koppensteiner, W.A., Sippl, M.J. and Domingues, F.S. (2000) Protein Eng. 13, 745–752.
[81] Bujnicki, J.M., Elofsson, A., Fischer, D. and Rychlewski, L. (2001) Bioinformatics 17, 750–751.
[82] Ginalski, K. and Rychlewski, L. (2003) Nucleic Acids Res. 31, 3291–3292.
[83] Sommer, I., Zien, A., von Öhsen, N., Zimmer, R. and Lengauer, T. (2002) Bioinformatics 18, 802–812.
[84] von Öhsen, N., Sommer, I. and Zimmer, R. (2003) Pac. Symp. Biocomput., 252–263.
[85] Rodriguez, R., Chinea, G., Lopez, N., Pons, T. and Vriend, G. (1998) Bioinformatics 14, 523–528.
[86] Bower, M.J., Cohen, F.E. and Dunbrack Jr., R.L. (1997) J. Mol. Biol. 267, 1268–1282.
[87] Bates, P.A., Kelley, L.A., MacCallum, R.M. and Sternberg, M.J. (2001) Proteins 45 (Suppl. 5), 39–46.
[88] Galtier, N., Gouy, M. and Gautier, C. (1996) Comput. Appl. Biosci. 12, 543–548.
[89] Gouet, P., Robert, X. and Courcelle, E. (2003) Nucleic Acids Res. 31, 3320–3323.
[90] Liu, J., Smith, C.L., DeRyckere, D., DeAngelis, K., Martin, G.S. and Berger, J.M. (2000) Mol. Cell 6, 637–648.
[91] Zhang, X. et al. (2000) Mol. Cell 6, 1473–1484.
[92] Guo, F., Maurizi, M.R., Esser, L. and Xia, D. (2002) J. Biol. Chem. 277, 46743–46752.
[93] Li, J. and Sha, B. (2002) J. Mol. Biol. 318, 1127–1137.
[94] Abrahams, J.P., Leslie, A.G., Lutter, R. and Walker, J.E. (1994) Nature 370, 621–628.
[95] Bourne, H.R., Sanders, D.A. and McCormick, F. (1991) Nature 349, 117–127.
[96] Beyer, A. (1997) Protein Sci. 6, 2043–2058.
[97] Neuwald, A.F., Aravind, L., Spouge, J.L. and Koonin, E.V. (1999) Genome Res. 9, 27–43.
[98] Patel, S.S. and Picha, K.M. (2000) Annu. Rev. Biochem. 69, 651–697.
[99] Vale, R.D. (2000) J. Cell Biol. 150, F13–19.
[100] Geourjon, C., Orelle, C., Steinfels, E., Blanchet, C., Deléage, G., Di Pietro, A. and Jault, J.M. (2001) Trends Biochem. Sci. 26, 539–544.
[101] Ogura, T. and Wilkinson, A.J. (2001) Genes Cells 6, 575–597.
[102] Caruthers, J.M. and McKay, D.B. (2002) Curr. Opin. Struct. Biol. 12, 123–133.
[103] Dougan, D.A., Mogk, A., Zeth, K., Turgay, K. and Bukau, B. (2002) FEBS Lett. 529, 6–10.
[104] Leipe, D.D., Wolf, Y.I., Koonin, E.V. and Aravind, L. (2002) J. Mol. Biol. 317, 41–72.
[105] Lupas, A.N. and Martin, J. (2002) Curr. Opin. Struct. Biol. 12, 746–753.
[106] Ramakrishnan, C., Dani, V.S. and Ramasarma, T. (2002) Protein Eng. 15, 783–798.
[107] Schmitt, L. and Tampé, R. (2002) Curr. Opin. Struct. Biol. 12, 754–760.
[108] Hopfner, K.P. and Tainer, J.A. (2003) Curr. Opin. Struct. Biol. 13, 249–255.
[109] Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) EMBO J. 1, 945–951.
[110] Cardozo, T.J. and Abagyan, R. (1998) J. Mol. Model. 4, 83–93.
[111] Jaroszewski, L., Rychlewski, L., Reed, J.C. and Godzik, A. (2000) Proteins 39, 197–203.
[112] Muneyuki, E., Noji, H., Amano, T., Masaike, T. and Yoshida, M. (2000) Biochim. Biophys. Acta 1458, 467–481.
[113] Granel, B. et al. (2003) Dermatology 206, 257–259.
[114] Kobe, B. and Kajava, A.V. (2001) Curr. Opin. Struct. Biol. 11, 725–732.
[115] Kobe, B. and Deisenhofer, J. (1996) J. Mol. Biol. 264, 1028–1043.
[116] Papageorgiou, A.C., Shapiro, R. and Acharya, K.R. (1997) EMBO J. 16, 5162–5177.
[117] Aloy, P., Ceulemans, H., Stark, A. and Russell, R.B. (2003) J. Mol. Biol. 332, 989–998.
[118] Linhoff, M.W., Harton, J.A., Cressman, D.E., Martin, B.K. and Ting, J.P. (2001) Mol. Cell Biol. 21, 3001–3011.
[119] Seewald, M.J., Körner, C., Wittinghofer, A. and Vetter, I.R. (2002) Nature 415, 662–666.
[120] Hake, S.B., Masternak, K., Kammerbauer, C., Janzen, C., Reith, W. and Steimle, V. (2000) Mol. Cell Biol. 20, 7716–7725.
[121] Harton, J.A., O'Connor Jr., W., Conti, B.J., Linhoff, M.W. and Ting, J.P. (2002) Hum. Immunol. 63, 588–601.
[122] Wilke-Mounts, S., Pagan, J. and Senior, A.E. (1995) Arch. Biochem. Biophys. 324, 153–158.
[123] Schnizer, R.A. and Schuster, S.M. (1996) Arch. Biochem. Biophys. 326, 126–136.