Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 850–860
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

The N-homologue LRR domain adopts a folding which explains the TMV-Cg-induced HR-like response in sensitive tobacco plants

Claudia Stange [a,1], José Tomás Matus [a,1], Calixto Domínguez [b], Tomás Perez-Acle [b], Patricio Arce-Johnson [a,*]

[a] *Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, P.O. Box 114-D, Santiago, Chile*
[b] *Centro de Genómica y Bioinformática (CGB), Pontificia Universidad Católica de Chile, P.O. Box 114-D, Santiago, Chile*

Received 6 May 2005; received in revised form 29 May 2007; accepted 29 May 2007
Available online 2 June 2007

Abstract

Following leaf infection with the tobacco mosaic virus (TMV), *Nicotiana* species that carry the disease resistance *N* gene develop a hypersensitive response (HR) that blocks the systemic movement of the virus. TMV-sensitive tobacco plants that lack the *N* gene develop classical disease symptoms following infection with most of the tobamoviruses. However, upon infection with TMV-Cg, these plants display a HR-like response that is unable to limit viral spread. We previously identified the NH gene in sensitive plants; this gene is homologous to the resistance N gene and both belong to the TIR/NBS/LRR family. Isolation and analysis of the *NH* transcript enabled the prediction of the amino acid sequence in which we detected a leucine-rich repeat domain, proposed to be involved in pathogen recognition. This domain is found in four of five classes of pathogen resistant proteins, in which sequence and structural changes may generate different specificities. In order to study the possible functional role of the LRR domain in the HR-like response, we developed a comparative three-dimensional model for the NH and N gene products, by means of functional and structural domains recognition, secondary structure prediction, domain assignment through profile Hidden Markov Models (HMM) and molecular dynamics (MD) simulations. Based on our results we postulate that the NH protein could adopt a LRR fold with a functional role in the HR-like response. Our two reliable LRR three-dimensional models (N-LRR, NH-LRR) can be used as structural frameworks for future experiments in which the structure–function relationships regarding the protein–protein interaction process may be revealed. Evolutionary aspects of the *N* and *NH* genes in *Nicotiana* species are also discussed.

Keywords: *NH* gene; HR-like; TIR/NBS/LRR receptor; LRR; Comparative modelling; Hypersensitive response; TMV

1. Introduction

Plants continuously develop different defence mechanisms in order to challenge pathogen attacks, as they are enforced to adjust this machinery into an adaptive response. This direct relationship between plants and their natural enemies is one of the key processes responsible for antagonist co-evolution and the diversity of organisms [1]. Plant resistance reactions evolve in order to generate new ranges of pathogen resistance by producing broad spectra of specific disease receptors for cognate gene products of pathogenic virus, bacteria, fungi, and nematodes [2,3]. The structures of these receptor proteins, and consequently their specificity, have been modified during the course of evolution, thereby conferring new defence mechanisms in response to the appearance of new pathogenic races [4–6].

In the last 10 years, more than 20 resistance genes (*R* genes) have been identified and classified in several plant species [2,3,7]. On the basis of their structure and functional domains, these have been grouped into five families. With the exception of the tomato *Pto* gene, almost all *R* genes possess a leucine-rich repeat (LRR) motif domain [8]. Proteins containing LRR domains are widespread and functionally diverse, with over

Abbreviations: HR, hypersensitive response; SHR, systemic HR; *NH*, homologous to *N*; N-LRR, leucine-rich repeat domain on N gene product; NH-LRR, leucine-rich repeat domain on NH gene product; NBS, nucleotide-binding site; TIR, Toll–interleukin-1 (IL-1) receptor domain; TMV, tobacco mosaic virus

* Corresponding author. Tel.: +56 2 6862897; fax: +56 2 2225515.
E-mail addresses: jtmatus@puc.cl (J.T. Matus), parce@bio.puc.cl (P. Arce-Johnson).
[1] Fax: +56 2 2225515.

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.05.006

2000 identified to date in viruses, bacteria, archaea and eukaryotes. They are especially prevalent in certain plant genomes, including the plant model *Arabidopsis thaliana*, which contains over 150 identified nucleotide-binding site (NBS)–LRR encoding genes [9]. In tobacco, one of the most studied pathogen receptors is the N gene, a member of the Toll–interleukin-1 receptor domain (TIR)/NBS/LRR class of plant disease resistance genes [8,10]. Plants bearing this gene are resistant to the viral pathogen tobacco mosaic virus (TMV). Once pathogens are recognized by the N receptor, a hypersensitive response (HR) occurs [11], characterized by localized cell death around the site of infection to limit pathogen spread [59]. In contrast, plants that lack the *N* gene are susceptible to infection by the majority of tobamoviruses and develop systemic mosaic symptoms with apical leaf deformation.

Nicotiana tabacum Xanthi^{nn} is a tobamovirus-susceptible cultivar. Nevertheless, Xanthi^{nn} plants infected with TMV-Cg, a tobamovirus isolated from cruciferous plants [12], display an unexpected local and systemic response [13]. The local response, termed the HR-like response (based on its resemblance to the HR), is characterized by the appearance of local necrotic lesions but not by the subsequent arrest of virus spread. In a previous study, we isolated and characterized a novel *N* homologous gene (*NH*) from Xanthi^{nn}-susceptible plants (GenBank accession AY535010). *NH* has a coding sequence of 5028 bp and possesses the TIR, NBS and LRR domains, and shares more than 80% nucleotide identity with the *N* gene. Our experimental evidence [13,14] led us to presume that the LRR domain of the *NH* gene may be associated with the HR-like response, as it has been postulated that these domains in plants are involved in pathogen avirulence (avr) ligand recognition [2,5,6,15,16].

From the structural point of view, all LRR domains are composed of a beta–alpha unit connected by a loop, adopting an arc-shaped structure termed a ''horseshoe''. Arrangements of parallel β-sheets are present at the concave face of the ''horseshoe'', while on the convex side a variety of secondary structures co-exist, such as α-helix, 3_{10}-helix, β-turn and pII [17]. The number of LRR repeats can range from 2 to 45, and the length of each repeat can vary from 20 to 30 residues. Each LRR repeat can be separated into a variable and a highly conserved segment described by a LxxLxLxxN/CxL pattern [18], which represents the inside and surrounding regions of the β-strand.

The elongated and curved shape of the LRR structure creates an extended and concave surface that is topographically suited for molecular docking and provides an interface for protein–protein interactions [19–21]. This structural framework is assumed to be responsible for a variety of biological processes including: hormone–receptor interactions, enzyme inhibition, cell adhesion, cellular trafficking and the immune response [21].

In order to understand the role of the *NH* gene in the HR-like response, we correlated the presence of the *N* and *NH* genes in *Nicotiana* species with HR markers. To study the appropriateness of the LRR fold as a structure–function framework in the HR-like response, we developed comparative three-dimensional models for the LRR domains of both gene products (N-LRR and NH-LRR, respectively), by means of cross-matched bioinformatics evidence arising from secondary structure prediction, functional and structural domain recognition, domain assignment through profile Hidden Markov Models (HMM) and molecular dynamics (MD) simulations.

2. Materials and methods

2.1. Plant materials and TMV strains

The TMV-U1-susceptible species *N. tabacum* L. cv. Xanthi^{nn}, *N. tabacum* L. cv. Petite Havana SR1, *N. sylvestris* Speg and Comes, *N. benthamiana* Domin, and the TMV-U1-resistant species *N. tabacum* L. cv. Xanthi^{NN}, *N. rustica* L., *N. tabacum* L. cv. Samsun, *N. glauca* Graham and *N. glutinosa* L. were used in this study. All plants were grown in a virus-free greenhouse under a 16/8 h (light/dark) photoperiod at 25–27 °C. Both virus strains, TMV-U1 and TMV-Cg, were propagated in TMV-U1-susceptible tobacco Xanthi^{nn} plants, purified as described by Bruening et al. [22] and stored at 4 °C at a concentration of 3 μg μl^{-1}.

2.2. TMV infection symptoms, callose deposition and cell death evaluation

Nicotiana plants (8-week old) were inoculated with 100 μl of TMV-U1 or TMV-Cg (10 ng μl^{-1} in 20 mM sodium phosphate buffer, pH 7). Local and apical leaf symptoms were evaluated from the second day of infection onwards.

The degree of callose deposition in the localized sites of infection (lesions), was determined using the method of Dietrich et al. [23]. At 2–3 days post-infection (dpi), leaf discs (2 cm in diameter) were removed from local sites of infection in the leaves, and after 10–12 dpi, discs were removed from apical leaves. These were first clarified in lactophenol (phenol:lactic acid:glicerol, 1:1:1) and then boiled for 2 min, followed by a wash in 50% ethanol and water. The bleached discs were then left for 1 h in 0.01% aniline blue dissolved in 0.15 M K_2HPO_4 with shaking and finally washed twice in water. The discs were then observed under an epifluorescence microscope (Olympus M081) at 365 nm. Aniline blue stain reacts with callose to generate a yellow compound, which under a 365 nm UV filter emits a blue fluorescence.

To evaluate cell death at localized sites of infection, we used a modified version of Baker and Mock [24] Evans blue staining method. Leaf discs were shaken for 30 min at 80 rpm in 1 ml of 0.25% Evans blue solution, followed by two washes with water. Blue staining associated with cell death in local and systemic lesions were observed in the treated discs under an optical microscope.

2.3. RT-PCR amplification of the NH cDNA

We amplified NH-specific cDNA by stepwise-RT-PCR to predict an amino acid sequence of the NH protein. This

sequence was subsequently used to develop the comparative models of the LRR domains of the N and NH gene products. Poly-adenylated RNA from *Nicotiana* Xanthi^{nn} was extracted using the PolyAtract® mRNA Isolation Systems kit (Promega, Madison, WI) based on biotin–streptavidine system. Complementary DNA (cDNA) was obtained using the Improm-II™ Reverse Transcriptase kit (Promega). PCR amplification was carried out using specific *NH* gene primers. Samples were analyzed on agarose gels, and the amplified products were cloned in pGEMT (Promega). The clones were sequenced with universal M13 forward, universal M13 reverse and NH-specific primers using the Fluorescence ABN Automated Laser Sequencer (Applied Biosystems, Foster City, CA). The mRNA sequence was translated into an amino acid sequence.

2.4. Comparative modelling of the N- and NH-LRR domain

LRR domains were annotated over the N and NH amino acid sequences by searching the PFAM database [25], and selecting hits with lower values than the default gathering threshold (*E*-value = 1.0). The N-LRR and NH-LRR amino acid sequences share 76.5% identical residues and 81% positive residues as aligned by ClustalW 1.8 [26] and using the weight matrix BLOSUM62.

In order to choose an appropriate protein template to develop a three-dimensional model for the LRR domains, we carried out a BLASTp search [27] against proteins whose known tertiary structures have been deposited in the Protein Data Bank (PDB) [28]. The first hit of BLASTp resulted in the plant defence polygalacturonase inhibiting protein (Pgip, PDB code: 1OGQ_A) [29]. This protein of 313 residues contains 10 LRRs with 19.7% and 18% sequence identity with N and NH, respectively. Due to the small sequence coverage of this hit in comparison with the N-LRR and NH-LRR sequence lengths, we decided to search for a more appropriate template by submitting target sequences to the 3D-PSSM [30] fold recognition server (threading). The best-scored structure protein corresponded to Internalin from *Listeria monocytogenes* (PDB code: 1O6S_A) [31]. An *E*-value of 7.7×10^{-3} resulted in more than 95% certainty [30] as a measure of confidence for threading prediction. Despite this confidence, and due to the low sequence similarity between N-LRR, NH-LRR and the threading template (16% and 18%, respectively), a closer inspection of the alignment was performed on each LRR tandem – on both gene products – for which the PFAM database search did not detected a canonical LRR pattern. These PFAM-unrecognized LRR tandems were multiple-aligned using ClustalW 1.8 in order to create a Hidden Markov Model profile (HMM) applying HMMER2.3 [32]. We identified this HMM as a LRR-like pattern. LRR tandems identified on the N and NH sequences were annotated (Fig. 2A), using the boundaries defined as the LRR signature (Fig. 2B). The proposed alignment resulting from the threading prediction (3D-PSSM) was manually fine-tuned using as assignment rule the localization of the LRR pattern (LxxLxLxxN/CxL) in accordance to the secondary structure predicted by NPS@ server [33]. The evidence used for this structural constraints was the broadly reported agreement of the LRR pattern and the β backbone conformation mainly located in the first six residues of the LRR pattern [17,34], where the "L" positions are buried into the protein core structure and the alternate polar and charged residues are solvent exposed. The adjusted alignment was used as a starting point for the development of 15 comparative molecular models resulting from MODELLER [35]. Those models that did not score at least 64.3 and 69.4 for N and NH gene products, respectively, were rejected according to the performed PROFILES-3D evaluation [36]. Accepted models were ranked by the PDF values (probability density function, computed by MODELLER), selecting one top scored model for both N-LRR and NH-LRR, in order to conduct further analysis.

2.5. Molecular dynamics (MD) simulations

To verify the stability of the selected models, we conducted molecular dynamics (MD) simulations using GROMACS 3.1 with the GROMACS forcefield [37,60]. The leap-frog algorithm was applied for integrating Newton's motion equations under a water-solvated periodic boundary condition (PBC). Every model structure was refined by means of 1000 cycles of steepest descent energy minimization to a convergence criterion of 100 kJ mol^{-1} nm^{-1}. The system was then equilibrated for 100 ps at 300 K and 1 atm, and a Berendsen temperature–pressure coupling (NPT ensemble) was applied. A full MD simulation of 1.0 ns was performed, collecting trajectories every 1 ps. The backbone Cα atom displacement from the average structure was calculated for each residue throughout the simulation by means of the standard deviation of RMS (root mean square fluctuation RMSF). In order to evaluate the conformation of the Φ and ψ angles of the backbone atoms following simulation, we used the Ramachandran plot server for RAMPAGE [38]. Further physicochemical evaluations for each model were performed using PROFILES-3D [36].

2.6. Electrostatic potential distribution

To observe and compare the charge distribution on the surface of each protein model, we used DELPHI within Insight II (Accelrys Inc., San Diego, CA), which implements the finite-difference Poisson–Boltzmann method [39,40]. We used standard protein atomic charges, an ionic strength of 50 mM and internal and external dielectric constants of 2 and 80, respectively.

3. Results

3.1. The HR-like response develops in TMV-susceptible tobacco species infected with TMV-Cg

When *Nicotiana* TMV-resistant plants (carrying the *N* gene) were infected with TMV-U1, the HR was induced in the leaf at the site of inoculation (Fig. 1A). TMV-susceptible species do not carry the *N* gene [14] and are unable to develop the HR against TMV-U1 or most of the other tobamoviruses. When

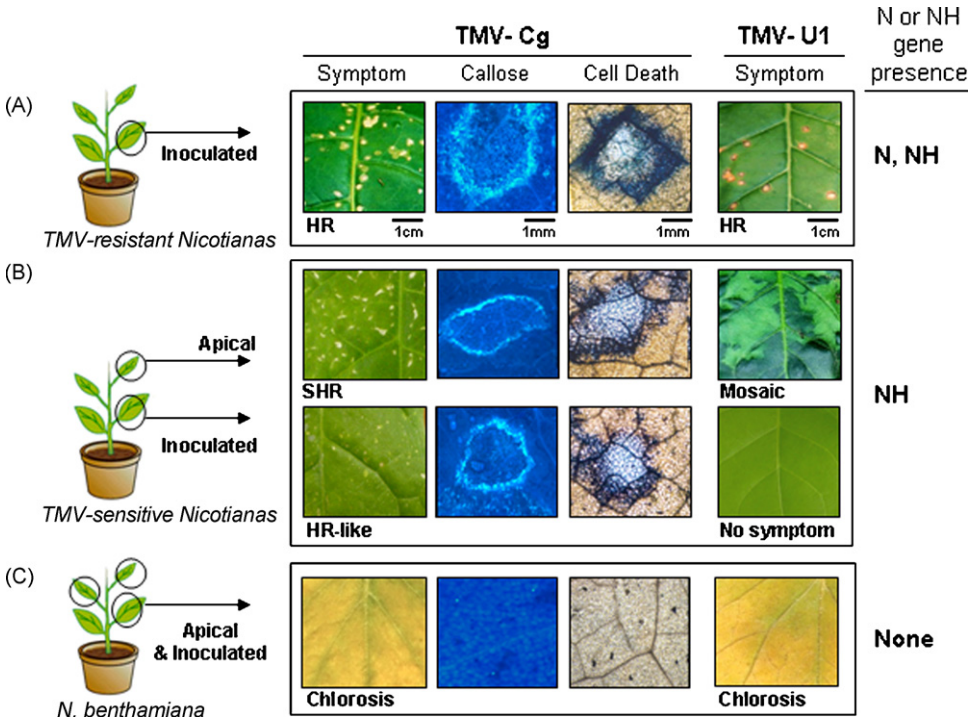

Fig. 1. Analysis of local and apical symptoms, callose deposition and cell death, in *Nicotiana* species infected with TMV-U1 or TMV-Cg at 25–27 °C. The presence of N and NH genes in these plants is indicated, according to Stange et al. [14]. A basal leaf (local) was inoculated with 10 ng μl^{-1} of TMV-U1 or TMV-Cg. Local symptoms were evaluated at 3 days post-inoculation (dpi), and apical symptoms were evaluated at 15 dpi. Callose deposition analysis in necrotic lesions was evaluated using aniline blue followed by UV light microscope analysis. Cell death at infection sites was analyzed using Evans blue staining method. (A) TMV-resistant *Nicotiana* plants (*N. tabacum* Xanthi^{NN}, *N. rustica*, *N. tabacum* Samsun, and *N. glauca*). (B) TMV-sensitive *Nicotiana* plants (*N. tabacum* Xanthi^{nn}, *N. tabacum* cv. Petite Havana SR1, *N. sylvestris*). (C) Whole plant symptoms (all leaves) in *N. benthamiana*.

these plants were infected with TMV-U1, no necrotic lesions or other HR symptoms of resistance appeared. Instead, the virus spread systemically through the plant, resulting in the appearance of mosaic disease symptoms accompanied by apical leaf deformation (Fig. 1B) [41]. When susceptible plants, for which the presence of the NH gene had been detected, were infected with TMV-Cg, they showed HR-like and SHR responses, which were accompanied by callose deposition surrounding each lesion (Fig. 1B). Localized cell death, which is indicative of the HR response, was also correlated with the onset of the HR-like and SHR responses in *Nicotiana*-susceptible plants (Fig. 1B). *N. benthamiana* was the only susceptible plant that did not develop HR-like and SHR responses following infection with TMV-Cg and TMV-U1. Instead of these symptoms they developed chlorosis in the infected leaves (Fig. 1C). This lack of HR-like and SHR responses and the appearance of chlorosis in *N. benthamiana*, correlated with the absence of both N and NH genes [14] and with the absence of callose deposit and cell death in the inoculated and apical leaves (Fig. 1C).

3.2. Three-dimensional modelling of the N-LRR and NH-LRR domains

The cDNA sequence of the *NH* gene was obtained by reverse transcription (RT)-PCR. Our results confirm those obtained from previous genomic sequencing, northern blot and RT-PCR semi-quantitative assays [14], indicating that the *NH* gene does not produce alternative transcripts. Our sequence comparison of the *NH* gene, carried out with ClustalW 1.8, revealed that the cDNA of the *NH* gene shares 82.8% nucleotide identity with the *N* gene. Exon 4, which is the primary determinant of the LRR domain, is shorter in the *NH* gene than in the *N* gene, and exon 5 is absent in *NH* (data not shown) in accordance with our previous results [14].

Beyond the limits of the PFAM search, we discovered a LRR-like pattern represented by a HMM built from the aligned segments that were not identified by the PFAM available LRR pattern (black boxes in Fig. 2A). Although these segments showed apparent sequence divergence, our HMM was also able to include the canonical PFAM pattern on the transition probability matrix of both gene products.

The alignment of N-LRR and NH-LRR against the protein template (1O6S_A) is shown, by the extent of the blocks of its respective LRR modules according to PFAM and the LRR-like HMM (Fig. 2B). Compatibility between the amino acid sequences and the amino acid side chains environment was evaluated using PROFILES-3D. The overall self-compatibility scores of N-LRR and NH-LRR models must be higher than the threshold of 64.3 and 69.4, respectively. Selected models scored 125.0 and 132.4 for N-LRR and NH-LRR, respectively. Those models produced from other alignment sources were rejected since their self-compatibility scores did not reach the above-suggested threshold.

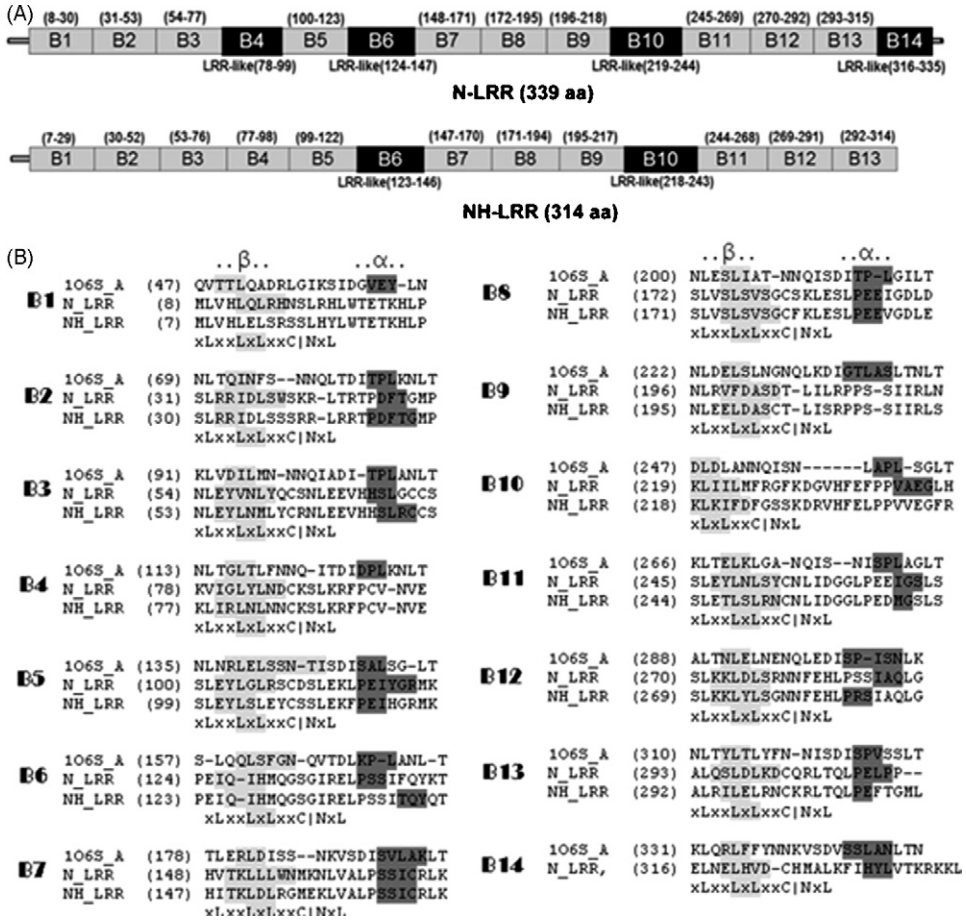

Fig. 2. Multiple alignment for comparative modelling of LRR domain amino acid sequences of the N and NH proteins. (A) Each LRR pattern is determined for the N (top) and NH (bottom) sequences by means of a PFAM database query (gray boxes) and a HMM of LRR-like structures (black boxes). (B) The selected alignment for N and NH modelling is shown by LRR blocks defined on template (1O6S_A), N and NH proteins. Alignment positions are highlighted according to secondary structure prediction: β-strand, gray background and black characters; alpha/helical region, black background and white character. The conserved xLxxLxLxxC/NxL pattern is shown below each LRR block.

3.3. Model assessment and molecular dynamics simulation

Developed models remained stable over the entire MD simulation time range. Data were collected each 1 ps, and the Cα RMSF was computed along each trajectory (Supplementary Figure A and B). The N-LRR model reached a maximum value of approximately 1.6 Å, with an average value of 0.7 Å. The NH model never exceeded the unique extreme values of 2.5 Å, with an average value of 1.1 Å. Although the Cα displacement along the MD trajectory was maintained close to the average value of 1 Å, we observed certain maximum RMSF values corresponding to residues located in the LRR variable segments. These results are in accordance with the spatial distribution of misplaced residues detected by the Ramachandran's plot analysis.

Ramachandran's plot analysis (Fig. 3B) shows that the overall backbone conformation, in comparison with the protein template, was reliable for our developed models. Nevertheless, we detected some residues in conflict with their acceptable backbone conformations; 21 residues (6.2%) for the N-LRR model and 16 residues (5.1%) for the NH-LRR. All of these residues appear located in the connecting loops and within the variable helical structural segments and none appear located in the conserved segments in the concave surface of the ''horseshoe'' structure.

Electrostatic potential calculations for the concave faces of the N-LRR and NH-LRR models showed remarkable changes in electrostatic potential distribution over the key region for protein–protein interaction (Fig. 4). Specifically, the NH-LRR model contains a negative cluster in the region of Asp152, whereas this region corresponds to Leu153 in the N-LRR model.

4. Discussion

The HR-like response, displayed in susceptible tobacco plants following TMV-Cg infection, is associated with callose deposition, cellular death (Fig. 1) and *PR1* expression [14], all of which are typical markers of HR. Despite the deposition of callose, which is a glucose polymer (1 → 3β-D-glucan chains) synthesized during HR as a physical barrier that restricts virus systemic movement [23], the HR-like response was unable to restrict systemic virus spread. This HR-like response was previously correlated with the presence of the NH gene [14].

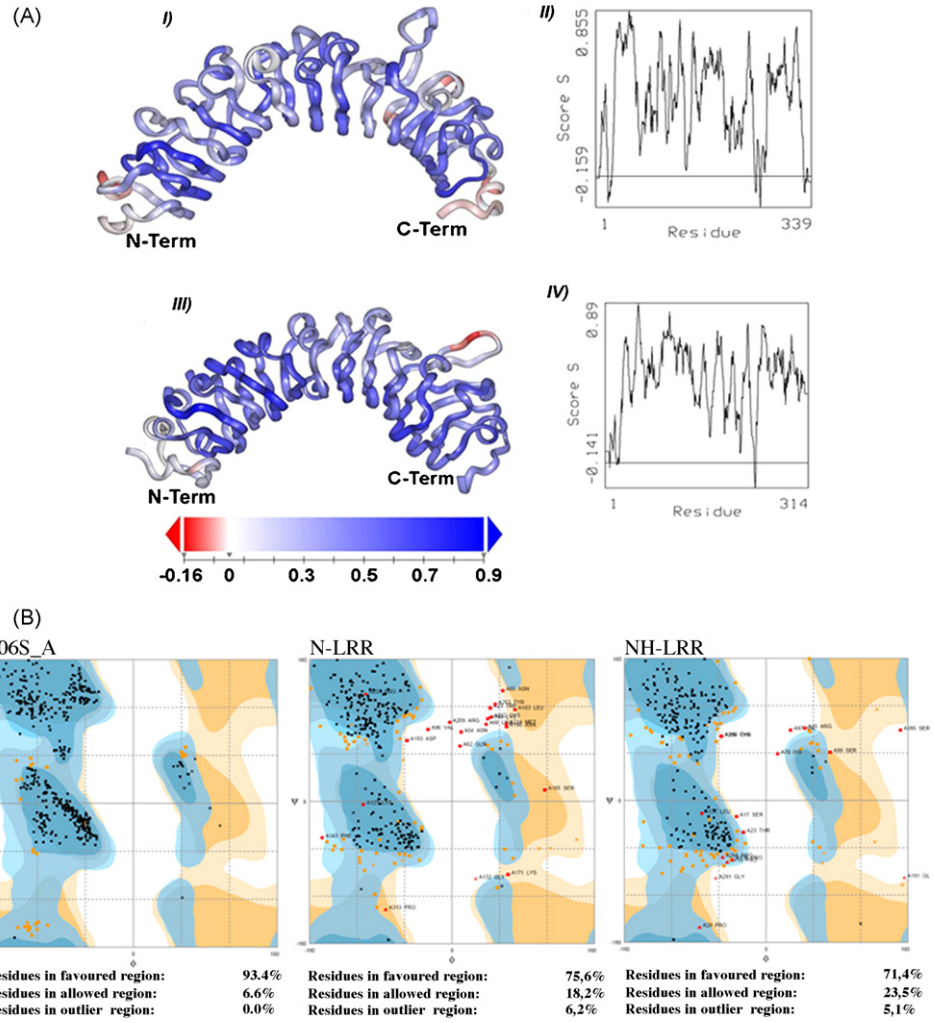

Fig. 3. Model evaluation of the LRR domain of the N and NH proteins. (A) Results of PROFILE-3D for N and NH after 1 ns of MD simulation. The colour scale specifies quality evaluation of the protein models NH (I) and N (III), with the more positive colour (blue) indicating better quality. The graphs for NH (II) and N (IV) model structures show values of $S < 0$, which indicate probable misfolded regions, and $S > 0$, which indicate putative valid regions. (B) Ramachandran plots for template and modelled structures. The main-chain dihedral angles (Φ, Ψ) for model structures NH and N in which the red sign represents residues in the outlier regions. The residues outside of acceptable regions are from variable regions.

Given the high sequence similarity between N-LRR and NH-LRR (76.5%) we expected good structural and functional correlation a priori. However, the models of N-LRR and NH-LRR suggest differential ligand recognition properties. Most importantly, we found several differences in the exposed residues in the area corresponding to the LxxLxL sequence of the LRR consensus motif (Fig. 5), a region shown to be involved in ligand recognition in other protein–protein complexes, including the ribonuclease inhibitor with RNAase A [19], internalin A with mE-caderin [31], and in the complex of RanGap with Ran [42].

There are five non-conservative substitutions in the N-LRR to NH-LRR sequence (Fig. 5B): Gly81 → Arg80; Leu153 → Asp152; Arg198 → Glu197; Val199 → Glu198 and Ile221 → Lys220, and these changes result in exposure of two contiguous negatively charged residues in NH-LRR: Glu197 and Glu198. Furthermore, the electrostatic potential distribution for the concave surface of NH-LRR showed a cluster of negative charges located in the middle of the surface around Asp152. Since this negative cluster was absent in the N-LRR surface, this region could have a key role in the divergence in ligand recognition properties between the N and NH receptors. This is in agreement with previous reports in plant systems regarding properties such as polygalacturonase recognition specificity, which is driven by negatively charged residues located in the concave surface of the LRR domain of the polygalacturonase-inhibiting protein [29].

The development of a HMM, using aligned segments that were not identified in the PFAM search allowed definition of an LRR-like pattern in both the N and NH gene products. In this LRR-like pattern, in accordance with the canonical PFAM LRR pattern, isoleucine (I) residues are over-represented relative to leucine (L) (data not shown). This is consistent with amino acid substitution in the LxxLxLxxN/CxL profiles, where ‘‘L’’ indicates Leu, Ile or Val, ‘‘N’’ represents Asn, Thr, Ser or Cys, and ‘‘C’’ indicates Cys or Ser [43]. Based on this HMM LRR-like pattern, we were able to define an extra C-terminal LRR domain in N-LRR (N-LRR14) compared to NH-LRR (Fig. 5B).

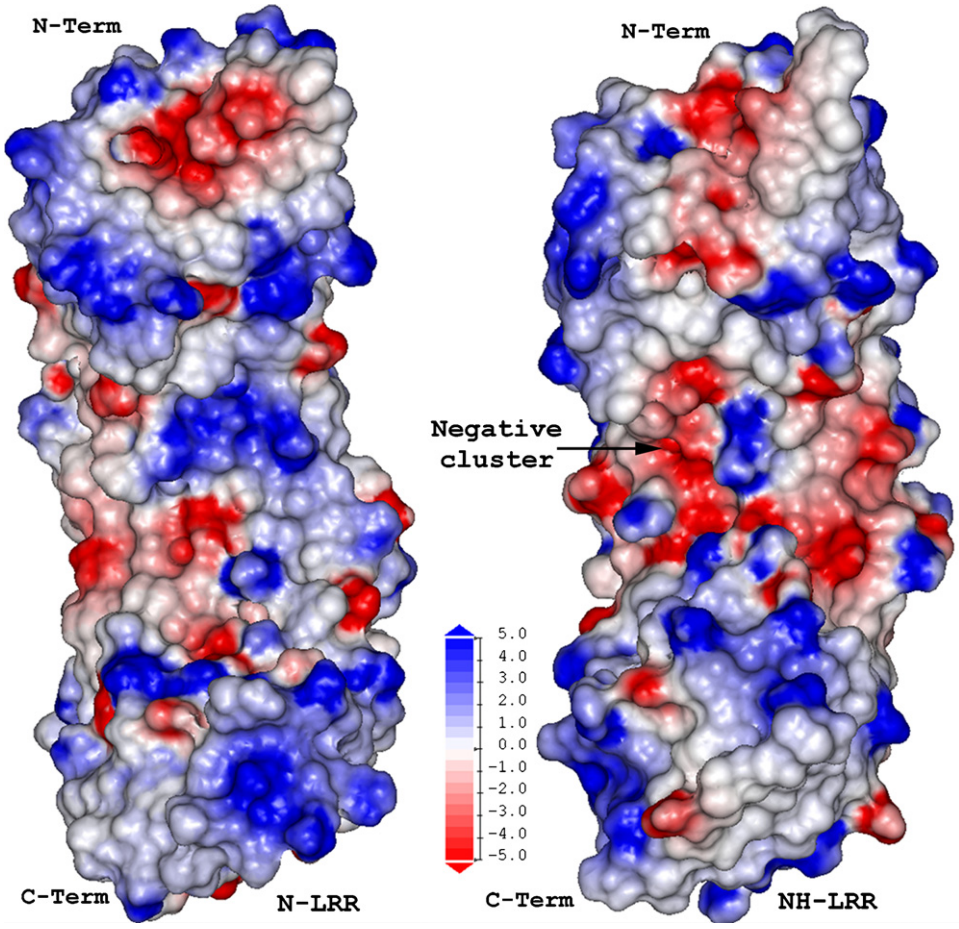

Fig. 4. Electrostatic potential for the N-LRR and NH-LRR models rendered at the concave surface. Surface residues appear coloured according to the electrostatics potential spectrum, being blue for positive potential and red for negative. Intermediate values are scaled from blue to red. As can be noted, a remarkable cluster of negatively charged residues can be distinguished (right panel, black arrow) at the inner concave surface of NH-LRR, region that has been previously implicated in ligand recognition specificity.

Further, structural analysis of N-LRR14 supports its inclusion as an LRR structural fold. This extra LRR increases the total available area for protein–protein interaction at the concave face of the LRR domain, as has been reported by Schubert et al. [31], and also could be a feasible explanation for the defence phenotype differences between NH and N genes in tobacco species.

There is no actual evidence for intracellular plant-specific subfamily NBS–LRR structures, or for variation in secondary structure architecture in each subfamily. Moreover, the N-LRR and NH-LRR tandems lengths are mostly longer than their template (1O6S_A). This produced open sequence gaps (Fig. 2B), increasing the number of misfolded residues. As seen in the Ramachandran plot in Fig. 3B, some residues are located out of the optimal conformational regions; however, these residues are located in the variable segment of the LRR pattern (data not shown), outside the first six conserved residues (LxxLxL) of the β-strands. The localization of these six conserved residues, despite the low global sequence similarity between targets and template, suggests that some residues are in inviolable positions, leading to the adoption of a ‘‘horseshoe’’ fold in the LRR domain [17]. These residues also had low RMSF values in MD simulations, which support the structural integrity of the models; whereas higher RMSF values were obtained for residues located in the variable helical segments (Supplementary Figure).

As TMV-susceptible tobacco species have the transduction machinery to display the HR-like response, we propose that these plants are able to synthesize a disease receptor; the NH protein, which can recognize a TMV-Cg elicitor but not TMV-U1. This incomplete defence response could be associated with an inefficient recognition of the TMV-Cg elicitor by the LRR domain of the NH receptor. Experiments carried out in our group have demonstrated that the TMV-Cg coat protein (CP-Cg) is the elicitor (avr-ligand) of the HR-like response [44] and not the helicase region from the 126 kDa replicase protein as was determined for TMV-U1 [45,46]. Several tobamovirus coat proteins have been identified as effectors of HR. As an example, *Tomato Mosaic Virus* (ToMV) coat protein has been identified as the elicitor of N′ (*N prime*) resistance gene in *N. sylvestris* plants [47]. TMV-Cg coat proteins are continuously forming polymerized structures in order to assemble a complete virion. Therefore, in the HR-like response, this event, associated with a weak interaction between NH and TMV-Cg coat protein, would imply that the interaction between the two proteins is not strong enough to restrict the virus to the localized site of inoculation in

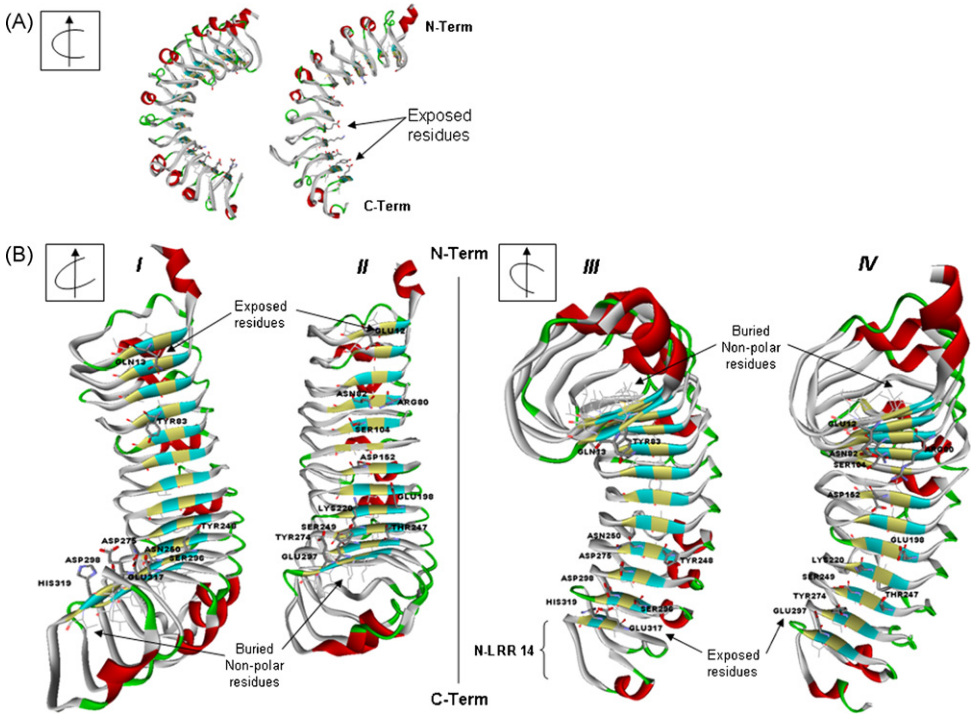

Fig. 5. Comparative models of N-LRR and NH-LRR showing secondary structure (Kabs & Sander) render. (A) Side view of the N-LRR (left) and NH-LRR (right) models adopting the classical ''horseshoe'' structure. (B) Front view of the concave face of N-LRR (I and III) and NH-LRR (II and IV) composed by the characteristic β-strands (cyan) which are composed for the first six residues of the LRR conserved segment (LxxLxL). The fourth and sixth buried non-polar residues (L) appear highlighted in yellow alternating with exposed residues. This sequence–structure linkage provides the main constraint on the input alignments used to produce the comparative models. At the concave ligand recognition interface we detected different polar and charged exposed residues between N-LRR (I and III: Gln13, Tyr83, Tyr248, Asn250, Asp275, Ser296, Asp298, Glu317, His319) and NH-LRR models (II and IV: Glu12, Arg80, Asn82, Ser104, Asp152, Glu198, Lys220, Thr247, Ser249, Tyr274, Glu297), these are labelled in black characters and displayed as sticks. The additional LRR of N (N-LRR14) with respect to the NH structure can be seen in panel B(III).

the leaf. As a consequence, TMV-Cg spreads systemically resulting in the subsequent secondary recognition response in the apical leaves.

A more complex situation is presented when we consider the alternative splicing of the N gene. This event generates a truncated N protein (Ntr) without an LRR domain. This Ntr protein is generated by the alternative splicing of the third intron of the N gene, a process which is induced by the presence of TMV [48,49]. The third intron from the NH gene has a 36.8% nucleotide identity with respect to the same region of the N gene, and there are no consensus alternative splicing motifs in the NH gene. In addition, RT-PCR semi-quantitative analyses of the NH transcript carried out in presence of dCTP [α^{32}P] demonstrated that no alternative transcript is generated [14]. The importance of the Ntr spliced variant in the HR response is that it is necessary for homo-dimerization with the TIR domain of the N protein and for hetero-dimerization with factors involved in the signal transduction, as has been reported for animal TIR domains which associate with MyD88 [50,51]. The HR-like response of sensitive plants resembles the incomplete resistance event that occurs in transgenic plants that are unable to synthesize the Ntr protein [49]. Ten days following infection, these plants develop severe infection symptoms characterized by the appearance of discontinuous necrotic stripes in the middle and upper leaves (SHR). In addition, Peart et al. [52] recently described a novel N-resistance component (NRG1) that encodes a coiled coil-NBS–LRR type R protein. NRG1 could be recruited and function in downstream signalling pathways directed by the N receptor, together with the diverse disease resistance cofactors such as EDS1 [53] and SGT1 [54]. The early event of elicitor-induced oligomerization and stabilization of the N protein, mediated by the TIR domain and possibly the NBS domain has also being associated with an effective resistance against TMV [55]. Taken together, a model that explains the appearance of the HR-like response in TMV-sensitive tobacco plants is shown in Fig. 6, based on the reduced affinity of the NH-LRR domain for the TMV-Cg coat protein, and the absence of a spliced variant necessary for a rapid and intense response.

The repetitive structure of the LRR motif may favour disease and defence co-evolution, since it can change more rapidly, thereby generating new variants with few changes in its residues [34]. This kind of positive selection includes changes both inside and outside the LRR region as well as changes in the length of the LRR. For extracellular-plant LRR it has been reported that at least one single amino acid mutation can change recognition specificities, for instance, PvPGIP1 mutated at position 224 with Gln224 of PvPGIP2, acquiring the ability to recognize polygalacturonase from F. moniliforme [56]. In other plant disease systems, small differences in the LRR domain induce species-specific recognition. For example, the L6 and L11 proteins, which are identical with respect to their TIR and

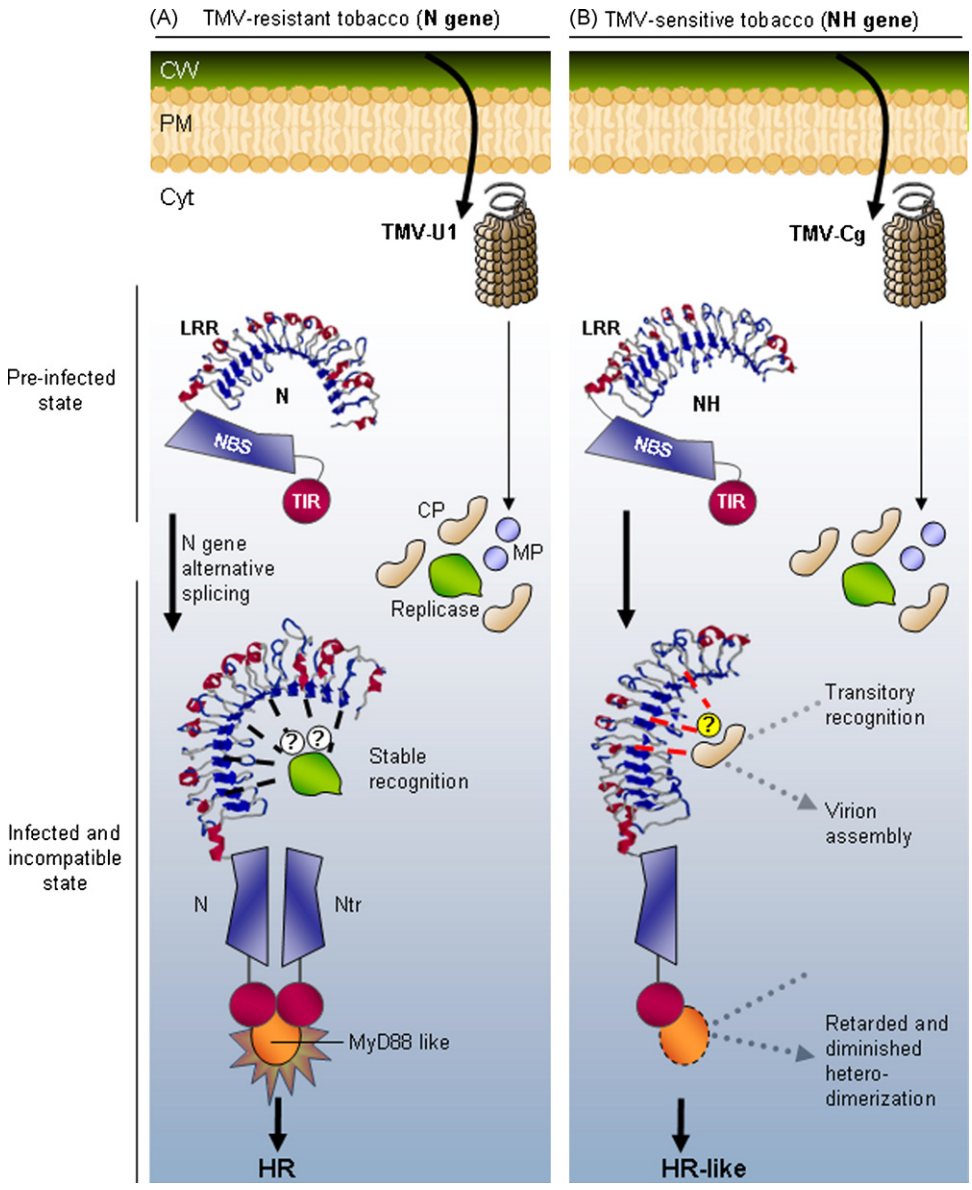

Fig. 6. A model for recognition functions of the LRR domain from N and NH putative proteins. (A) During a pre-infected stage, the N receptor is inactive. Once infection takes place, the viral replicase is synthesized in the cell. This elicitor interacts with the LRR domain from the N receptor, triggering an initial defence response, which allows the induction of the alternative splicing with the consequent synthesis of the Ntr protein. This factor could homo-dimerize with N and hetero-dimerize with adapter proteins (e.g. NRG1 or MyD88-like) or other cofactors (described in the text), through the TIR domain to induce a complete defence response (HR). (B) In the case of tobacco-sensitive plants carrying the NH gene, the elicitor is the TMV-Cg coat protein, with a much less stable interaction with the NH-LRR domain. In addition, no spliced variant is produced within the NH gene, so a retarded and diminished transduction signalling occurs. Neither positive nor negative host regulators of HR and HR-like are included. CW: cell wall; PM: plasmatic membrane; Cyt: cytoplasm.

NBS regions, differ by 33 amino acid substitutions in the LRR [57], indicating that the differences between L6 and L11 resistance specificities are caused by differences in their LRR regions [57]. *In vitro* exchanges between alleles and analyses of transgenic plants into which the resulting hybrid *L* genes have been introduced also indicate the importance of LRR variation in specificity differences. Thus, length modification of the LRR domains appears to be an important contributor to *R*-gene diversification [57].

As the N gene was introduced from only one *Nicotiana* specie (*N. glutinosa*) into sensitive *N. tabacum* plants [58], and that the complete *NH* gene is present not only in susceptible tobaccos, but also in resistant ones [14], we also propose that the NH gene is previous to the appearance of the N gene. Over time, this gene may have generated a spliced variant and a more efficient recognition LRR domain, obtaining a complete and rapid response upon TMV infection.

Despite the sequence variations and the structural differences that become apparent from the comparative analysis of the N-LRR and NH-LRR, a common sequence and structure–function framework is maintained throughout the evolution of the plant defence resistance genes. The variations on the sequence length and composition produce new ranges of specificity, thus increasing the plant resistance gene repertoire in an adaptive way that resembles the clonal expansion of the MHC genes of the mammalian immune system.

Acknowledgements

This work was supported by FONDECYT 1040789 and by the doctoral resources FONDECYT 2000078. We would like to thank Angel Gonzalez from the CGB for his encouragement and valuable advice on molecular modelling.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2007.05.006.

References

[1] T. Endo, K. Ikeo, T. Gogobori, Large-scale search for genes on which positive selection may operate, Mol. Biol. Evol. 13 (1996) 685–690.
[2] A.F. Bent, Plant disease resistance genes: function meets structure, Plant Cell 8 (1996) 1757–1771.
[3] B. Baker, P. Zambryski, B. Staskawicz, S. Dinesh-Kumar, Signaling in plant–microbe interactions, Science 276 (1997) 726–733.
[4] I.R. Crute, D.A.C. Pink, The genetics and utilization of pathogen resistance in plants, Plant Cell 8 (1996) 1747–1755.
[5] J. Ellis, P. Dodds, T. Pryor, Structure, function and evolution of plant disease resistance genes, Curr. Opin. Plant Biol. 3 (2000) 278–284.
[6] J. Ellis, P. Dodds, T. Pryor, The generation of plant disease resistance gene specificities, Trends Plant Sci. 5 (2000) 373–379.
[7] Y. Belkhadir, R. Subramaniam, J.L. Dangl, Plant disease resistance protein signaling: NBS–LRR proteins and their partners, Curr. Opin. Plant Biol. 7 (4) (2004) 391–399.
[8] K.E. Hammond-Kosak, J.D.G. Jones, Plant disease resistance genes, Annu. Rev. Plant Physiol. Plant Mol. Biol. 48 (1997) 575–607.
[9] B.C. Meyers, A. Kozik, A. Griego, H. Kuang, R.W. Michelmore, Plant Cell 15 (4) (2003) 809–834.
[10] S. Whitham, S.P. Dinesh-Kumar, D. Choi, R. Hehl, C. Corr, B. Baker, The product of the Tobacco Mosaic Virus resistance gene N: similarity to Toll and the interleukin-1 receptor, Cell 78 (1994) 1101–1115.
[11] G. Samuel, Some experiments on inoculating methods with plant viruses, and on local lesions, Ann. Appl. Biol. 18 (1931) 494–507.
[12] T. Yamanaka, H. Komatani, T. Meshi, S. Naito, M. Ishikawa, T. Ohno, Complete nucleotide sequence of the genomic RNA of tobacco mosaic virus strain Cg, Virus Genes 16 (1998) 173–176.
[13] P. Arce-Johnson, C. Medina, H.S. Padgett, W. Huanca, C. Espinoza, Analysis of local and systemic spread of the crucifer-infecting TMV-Cg virus in tobacco and several *Arabidopsis thaliana* ecotypes, Funct. Plant Biol. 30 (2003) 401–408.
[14] C. Stange, J.T. Matus, A. Elorza, P. Arce-Johnson, Identification and characterization of a novel TMV resistance N gene homologue in *Nicotiana tabacum* plants, Funct. Plant Biol. 31 (2004) 149–158.
[15] D.A. Jones, J.D.G. Jones, The role of Leucine Reach Repeat proteins in plant defenses, Adv. Bot. Res. 24 (1997) 90–167.
[16] S.P. Dinesh-Kumar, T. Wai-Hong, B. Baker, Structure–function analysis of the tobacco mosaic virus resistance gene *N*, Proc. Natl. Acad. Sci. U.S.A. 97 (26) (2000) 14789–14794.
[17] A.V. Kajava, B. Kobe, Assessment of the ability to model proteins with leucine-rich repeats in light of the latest structural information, Protein Sci. 11 (6) (2002) 1082–1090.
[18] E.L.L. Sonnhammer, S.R. Eddy, R. Durbin, Pfam: a comprehensive database of protein domain families based on seed alignments, Proteins 28 (1997) 405–420.
[19] B. Kobe, J. Deisenhofer, A structural basis of the interaction between leucine-rich repeats and protein ligands, Nature 374 (1995) 183–186.
[20] M. Marino, L. Braun, P. Cossart, P. Ghosh, A framework for interpreting the leucine-rich repeats of the *Listeria internalins*, Proc. Natl. Acad. Sci. U.S.A. 97 (16) (2000) 8784–8788.
[21] B. Kobe, A.V. Kajava, The leucine-rich repeat as a protein recognition motif, Curr. Opin. Plant Biol. 11 (2001) 725–732.
[22] G. Bruening, R.N. Beachy, R. Scalla, M. Zaitlin, *In vitro* and *in vivo* translation of the ribonucleic acids of a cowpea strain of tobacco mosaic virus, Virology 71 (2) (1976) 498–517.
[23] R.A. Dietrich, T.P. Delaney, S.J. Uknes, E.R. Ward, J.A. Ryal, J.L. Dangl, Arabidopsis mutants simulating disease resistance response, Cell 77 (1994) 565–577.
[24] C.J. Baker, N. Mock, An improved method for monitoring cell death in cell suspension and leaf disc assays using evans blue, Plant Cell Tiss. Organ Cult. 39 (1994) 7–12.
[25] A. Bateman, L. Colin, R. Durbin, R. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. Sonnhammer, D. Studholme, C. Yeats, S. Eddy, The Pfam protein families database, Nucl. Acids Res. Database Issue 32 (2004) D138–D141.
[26] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucl. Acids Res. 22 (1994) 4673–4680.
[27] S. Altschul, T. Madden, A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search Programs, Nucl. Acids Res. 25 (1997) 3389–3402.
[28] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucl. Acids Res. 28 (2000) 235–242.
[29] A. Di Matteo, L. Federici, B. Mattei, G. Salvi, K.A. Johnson, C. Savino, G. De Lorenzo, D. Tsernoglou, F. Cervone, The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 10124–10128.
[30] L. Kelley, R. MacCallum, M. Sternberg, Enhanced genome annotation using structural profiles in the program 3D-PSSM, J. Mol. Biol. 299 (2000) 501–522.
[31] W.D. Schubert, C. Urbanke, T. Ziehm, V. Beier, M.P. Machner, E. Domann, J. Wehland, T. Chakraborty, D.W. Heinz, Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor mE-cadherin, Cell 111 (6) (2002) 825–836.
[32] S.R. Eddy, Profile hidden Markov models, Bioinformatics 14 (1998) 755–763.
[33] C. Combet, C. Blanchet, C. Geourjon, G. Deléage, NPS@: network protein sequence analysis, TIBS 291 (2000) 147–150.
[34] M. Mondragón-Palomino, B.C. Meyers, R.W. Michelmore, B.S. Gaut, Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*, Genome Res. 12 (2002) 1305–1315.
[35] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[36] R. Luthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, Nature 356 (1992) 83–85.
[37] H. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56.
[38] S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by Cα geometry: ϕ, ψ and Cβ deviation, Proteins Struct. Funct. Genet. 50 (2002) 437–450.
[39] A. Nicholls, K.A. Sharp, B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, Proteins Struct. Funct. Genet. 11 (1991) 281–296.
[40] B. Honig, A. Nicholls, Classical electrostatics in biology and chemistry, Science 268 (1995) 1144–1149.
[41] W.O. Dawson, Tobamovirus–plant interactions, Virology 186 (1992) 359–367.
[42] M.J. Seewald, C. Korner, A. Wittinghofer, I.R. Vetter, RanGAP mediates GTP hydrolysis without an arginine finger, Nature 415 (2002) 662–666.
[43] P. Enkhbayar, M. Kamiya, M. Osaki, T. Matsumoto, N. Matsushima, Structural principles of leucine-rich repeat (LRR) proteins, Proteins 54 (3) (2004) 394–403.

[44] N. Ehrenfeld, P. Cañón, C. Stange, C. Medina, P. Arce-Johnson, Tobamovirus coat protein CPCg induces an HR-like response in sensitive tobacco plants, Mol. Cell 19 (3) (2005) 418–427.
[45] H.S. Padgett, Y. Watanabe, R.N. Beachy, Identification of the TMV replicase sequence that activates the N gene-mediated hypersensitive response, Mol. Plant Microbe Interact. 10 (6) (1997) 709–715.
[46] F.L. Erickson, S. Holzberg, A. Calderón-Urrea, V. Handley, M. Axtell, C. Corr, B. Baker, The helicase domain of the replicase proteins induces the N-mediated defence response in tobacco, Plant J. 18 (1) (1999) 67–75.
[47] U.M. Pfitzner, A.J. Pfitzner, Expression of a viral avirulence gene in transgenic plants is sufficient to induce the hypersensitive defense reaction, Mol. Plant Microbe Interact. 5 (4) (1992) 318–321.
[48] M. Levy, O. Edelbaum, I. Sela, Tobacco Mosaic Virus regulates the expression of it own resistance gene *N*, Plant Physiol. 135 (2004) 1–6.
[49] S.P. Dinesh-Kumar, B. Baker, Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 1908–1913.
[50] Y. Xu, X. Tao, B. Shen, T. Horng, R. Medzhitov, J.L. Manley, L. Tong, Structural basis for signal transduction by the Toll/interleukin-1 receptor domains, Nature 408 (6808) (2000) 111–115.
[51] L. O'Neill, The Toll/interleukin-1 receptor domain: a molecular switch for inflammation and host defence, Biochem. Soc. Trans. 28 (2000) 557–563.
[52] J.R. Peart, P. Mestre, R. Lu, I. Malcuit, D.C. Baulcombe, NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus, Curr. Biol. 15 (10) (2005) 968–973.
[53] J.R. Peart, G. Cook, B.J. Feys, J.E. Parker, D.C. Baulcombe, An EDS1 orthologue is required for N-mediated resistance against tobacco mosaic virus, Plant J. 29 (2002) 569–579.
[54] J.R. Peart, R. Lu, A. Sadanandom, I. Malcuit, P. Moffett, D.C. Brice, L. Schauser, D. Jaggard, S. Xiao, M. Coleman, M. Jonathan, D.G. Jones, K. Shirasu, D. Baulcombe, Ubiquitin ligase-associated protein SGT1 is required for host and nonhost disease resistance in plants, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 10865–10869.
[55] P. Mestre, D. Baulcombe, Elicitor-mediated oligomerization of the tobacco N disease resistance protein, Plant Cell 18 (2006) 491–501.
[56] F. Leckie, B. Mattei, C. Capodicasa, A. Hemmings, L. Nuss, B. Aracri, G. De Lorenzo, F. Cervone, The specificity of polygalacturonase-inhibiting protein (PGIP): a single amino acid substitution in the solvent-exposed beta-strand/beta-turn region of the leucine-rich repeats (LRRs) confers a new recognition capability, EMBO J. 18 (1999) 2352–2363.
[57] J. Ellis, G.J. Lawrence, J.E. Luck, P.N. Dodds, Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity, Plant Cell 11 (3) (1999) 495–506.
[58] F.O. Holmes, Inheritance of resistance to tobacco-mosaic disease in tobacco, Phytophatology 28 (1938) 553–561.
[59] M.C. Heath, Hypersensitive response-related death, Plant Mol. Biol. 44 (2000) 321–334.
[60] E. Lindahl, B. Hess, D. van der Spoel, GROMACS 3.0: a package for molecular simulation and trajectory analysis, J. Mol. Mod. 7 (2001) 306–317.