

# Structural modeling of the N-terminal signal-receiving domain of I $\kappa$ B $\alpha$

Samira Yazdi<sup>1</sup>, Serdar Durdagi<sup>1†</sup>, Michael Naumann<sup>2</sup> and Matthias Stein<sup>1\*</sup>

<sup>1</sup> Molecular Simulations and Design Group, Max Planck Institute for Dynamics and Complex Technical Systems, Magdeburg, Germany, <sup>2</sup> Medical Faculty, Institute of Experimental Internal Medicine, Otto von Guericke University, Magdeburg, Germany

## OPEN ACCESS

### Edited by:

Rommie E. Amaro,  
University of California, San Diego,  
USA

### Reviewed by:

Andrzej Stasiak,  
University of Lausanne, Switzerland  
Paolo De Los Rios,  
Ecole Polytechnique Fédérale de  
Lausanne, Switzerland

### \*Correspondence:

Matthias Stein,  
Molecular Simulations and Design  
Group, Max Planck Institute for  
Dynamics of Complex Technical  
Systems, Sandtorstrasse 1, 39106  
Magdeburg, Germany  
matthias.stein@  
mpi-magdeburg.mpg.de

### †Present Address:

Serdar Durdagi,  
Department of Biophysics, School of  
Medicine, Bahcesehir University,  
Istanbul, Turkey

### Specialty section:

This article was submitted to  
Biophysics,  
a section of the journal  
*Frontiers in Molecular Biosciences*

Received: 23 February 2015

Accepted: 03 June 2015

Published: 23 June 2015

### Citation:

Yazdi S, Durdagi S, Naumann M and  
Stein M (2015) Structural modeling of  
the N-terminal signal-receiving  
domain of I $\kappa$ B $\alpha$ .  
*Front. Mol. Biosci.* 2:32.  
doi: 10.3389/fmbo.2015.00032

The transcription factor nuclear factor- $\kappa$ B (NF- $\kappa$ B) exerts essential roles in many biological processes including cell growth, apoptosis and innate and adaptive immunity. The NF- $\kappa$ B inhibitor (I $\kappa$ B $\alpha$ ) retains NF- $\kappa$ B in the cytoplasm and thus inhibits nuclear localization of NF- $\kappa$ B and its association with DNA. Recent protein crystal structures of the C-terminal part of I $\kappa$ B $\alpha$  in complex with NF- $\kappa$ B provided insights into the protein-protein interactions but could not reveal structural details about the N-terminal signal receiving domain (SRD). The SRD of I $\kappa$ B $\alpha$  contains a degron, formed following phosphorylation by I $\kappa$ B kinases (IKK). In current protein X-ray structures, however, the SRD is not resolved and assumed to be disordered. Here, we combined secondary structure annotation and domain threading followed by long molecular dynamics (MD) simulations and showed that the SRD possesses well-defined secondary structure elements. We show that the SRD contains 3 additional stable  $\alpha$ -helices supplementing the six ARDs present in crystallized I $\kappa$ B $\alpha$ . The I $\kappa$ B $\alpha$ /NF- $\kappa$ B protein-protein complex remained intact and stable during the entire simulations. Also in solution, free I $\kappa$ B $\alpha$  retains its structural integrity. Differences in structural topology and dynamics were observed by comparing the structures of NF- $\kappa$ B free and NF- $\kappa$ B bound I $\kappa$ B $\alpha$ -complex. This study paves the way for investigating the signaling properties of the SRD in the I $\kappa$ B $\alpha$  degron. A detailed atomic scale understanding of molecular mechanism of NF- $\kappa$ B activation, regulation and the protein-protein interactions may assist to design and develop novel chronic inflammation modulators.

**Keywords:** signal transduction, NF- $\kappa$ B, I $\kappa$ B $\alpha$ , secondary structure prediction, N-terminal extension, molecular dynamics simulation, protein-protein complex refinement, signal receiving domain

## Introduction

### NF- $\kappa$ B Signaling and Its Inhibitor I $\kappa$ B $\alpha$

NF- $\kappa$ B plays a crucial role in mediating responses to various types of external stimuli, thus it is a key element in multiple physiological and pathological processes (Oeckinghaus and Ghosh, 2009). Defective NF- $\kappa$ B activity may lead to very serious health problems such as cancer and chronic inflammatory diseases (i.e., arthritis and Crohn's disease; for reviews see Bouma and Strober, 2003; Schreiber et al., 2005; Viatour et al., 2005). The NF- $\kappa$ B protein is bound by I $\kappa$ B $\alpha$  in unstimulated cells, keeping it inactive and retaining it in the cytoplasm and thus inhibiting nuclear localization of NF- $\kappa$ B and its association with DNA. Since NF- $\kappa$ B binds to a specific DNA motif in the nucleus and regulates transcription of target genes, the inhibition of NF- $\kappa$ B can be a therapeutic target for

the prevention or treatment of undesired biologic responses caused by the uncontrolled activation of NF-κB.

Upon activation, the IKK $\beta$  kinase phosphorylates IκB $\alpha$  at specific amino acid positions (i.e., Ser32 and Ser36) (Viatour et al., 2005). This site-specific phosphorylation of IκB $\alpha$  is a prerequisite for its ubiquitination by a specific E3 ubiquitin ligase the SKP1-CULLIN1-F-box (SCF) E3 ligase SCF(β-TrCP). SCF(β-TrCP)-mediated IκB $\alpha$  ubiquitination and degradation is very efficient and resulting in complete degradation of IκB $\alpha$  within a few seconds of cell stimulation (Suzuki et al., 1999; Neumann and Naumann, 2007).

IκB $\alpha$  in complex with NF-κB is highly stable and has an intracellular half-life of several hours (Hatada et al., 1992; Jaffray et al., 1995), while the free IκB $\alpha$  has a half-life of less than 10 min (Jaffray et al., 1995; Sun et al., 1996) and all efforts to crystallize IκB $\alpha$  in its unbound state have been unsuccessful so far (Huxford et al., 1999). Once bound to NF-κB, IκB $\alpha$  is only degraded if it is first phosphorylated, then ubiquitinated, and finally degraded by the proteasome (see Figure 1).

The complexity and versatility of the downstream signaling network is controlled, among others, by NF-κB-specific inhibitor proteins, namely IκBs (Schuster et al., 2013). IκBs are the critical regulators of NF-κB activity. They contain a signal receiving domain (SRD), six to seven ankyrin repeat units (Dyson and Komives, 2012) and a largely unstructured PEST [enriched in amino acids proline (P), glutamate (E), serine (S) and threonine (T)] domain at the C-terminus (Figure 2). The C-terminal PEST domain is also the site of post-translational modifications due to the casein kinase II (CK2) phosphorylation at positions 283, 288, 291, 293, and 299 (Cuff et al., 1998; Palopoli et al., 2009).

A detailed understanding of molecular mechanism of NF-κB activation, regulation and the protein-protein communication with partners may assist to design and develop novel chronic inflammation modulators as well as anti-cancer drugs. The insight gained from structural biology of NF-κB and IκB $\alpha$  proteins and its implications for the signaling process control have been reviewed extensively by i.e., (Moorthy et al., 2006; Ferreiro and Komives, 2010; Huxford et al., 2011; Ghosh et al., 2012).

The SRD of IκB $\alpha$  is the central signal receiving and transmitting domain when activating NF-κB. It contains sites for post-translational modifications [phosphorylation by kinases IKK $\alpha$  and IKK $\beta$  (Huxford et al., 1998; Moorthy et al., 2006) at Ser32 and Ser36; and Lys21 and Lys22 as the sites for subsequent ubiquitination by SCF(β-TrCP), respectively (Jacobs and Harrison, 1998; Cervantes et al., 2009)]. The SRD was always assumed to be unstructured or highly disordered based on the failed attempts to crystallize full-length IκB $\alpha$  in complex with NF-κB. The instability of free IκB $\alpha$  in solution and the absence of significant SRD contributions to the interaction energy of the protein-protein complex of IκB $\alpha$ /NF-κB lead to the hypothesis of the SRD not being critical for this complex formation. Detailed knowledge of the NF-κB/IκB $\alpha$  interaction comes from protein crystallography (Huxford et al., 1998) and high resolution NMR experiments (Schuster et al., 2013). However, these results do not include

any structural information about the SRD (residues 1–72) of IκB $\alpha$ . Previous investigations by molecular dynamics (MD) simulations of NF-κB/IκB $\alpha$  focused on the amide proton/deuterium exchange kinetics of four central ankyrin repeat units of co-crystallized IκB $\alpha$  by accelerated molecular dynamics (aMD) simulations (Cervantes et al., 2009), a truncated free-IκB $\alpha$  (Ferreiro et al., 2007) and the structure of a free, doubly phosphorylated 24 amino acid peptide of the SRD (Pons et al., 2007).

The concept of conservation of secondary structure elements (SSEs) (Rost, 2001) in families can be used to identify proteins only distantly related in sequence, which may, however, still share a higher degree of conservation of SSEs. Recent approaches have demonstrated that the use of multiple tools of secondary structure prediction and the use of a “consensus” of methods yields more reliable results than single algorithms (Cuff et al., 1998; Palopoli et al., 2009).

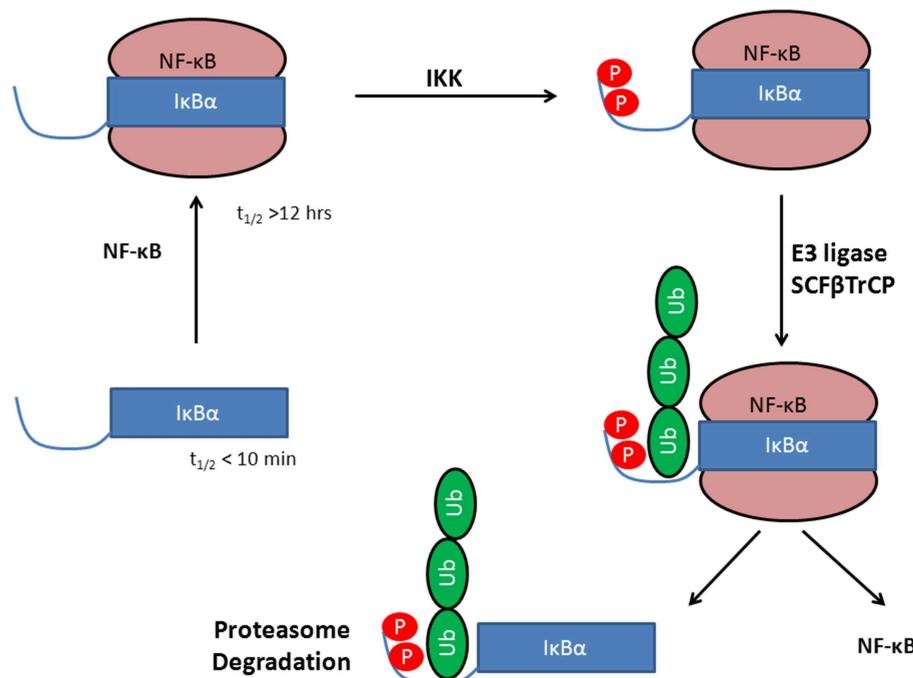
## Ankyrin Repeat Units as Interaction Modules Domain

The crystal structure of IκB $\alpha$  illustrates how the six repeating ankyrin domain assumes the shape of an arched cylinder assembled on top of the interface of the NF-κB heterodimer. Every repeat unit in IκB $\alpha$  is composed of two  $\alpha$ -helices connected to each following repeat with a loop of varying size and a  $\beta$ -hairpin turn containing short  $\beta$ -strands. However, repeats 1, 3, and 4 deviate from the canonical 33 amino acid repeat unit. These repeats are longer than the repeat units in the ankyrin consensus sequence, with the insertions contained in the loop sections, as these regions are those with the lowest sequence similarity among all ankyrin repeat proteins. Lack of homology is also observed in the sixth and last repeat unit, where the dissimilarity falls in after the second helix clearing the last 11 residues of any secondary structural elements (Huxford et al., 1998).

Free IκB $\alpha$  (67–317) was characterized by circular dichroism (CD) spectroscopy, 8-anilino-1-naphthalenesulphonic acid (ANS) binding, differential scanning calorimetry (DSC), and amide hydrogen/deuterium exchange experiments (Croy et al., 2004). The CD spectrum of free IκB $\alpha$  is nearly identical to the CD spectrum of the IκB $\alpha$ /NF-κB complex but it shows significant ANS binding and rapid amide exchange over much of the protein. These findings suggest that the secondary structure of IκB $\alpha$  is formed but the tertiary structure may not be compact. The  $\beta$ -hairpins of AR2 and AR3 were remarkably resistant to exchange, whereas AR5 and AR6 exchanged completely within the first minute in free IκB $\alpha$ . When bound to NF-κB, the  $\beta$ -hairpins of AR5 and AR6 showed dramatically less exchange in the bound state (Truhlar et al., 2006).

## The SRD in IκB $\alpha$ Protein Crystal Structures

From the structure of IκB $\alpha$  in complex with NF-κB, a valuable level of insight was rendered into the regulation of NF-κB by IκB $\alpha$  and the nature of their association (Huxford et al., 1998). Each unit of the complex was partially truncated leading to a missing IκB $\alpha$  N-terminal segment comprising ~70 residues. This



**FIGURE 1 | Stabilization of IκB $\alpha$  upon protein-protein complex formation with the transcription factor NF-κB.** Phosphorylation by IKK leads to a degron, recognized by SCF( $\beta$ -TrCP) and subsequent IκB $\alpha$  ubiquitination by E3. This activates NF-κB whereas IκB $\alpha$  is degraded in the 26S proteasome.

N-terminal SRD receives the phosphorylation and ubiquitination signals and targets the protein to the proteasome for degradation (Traenckner and Baeuerle, 1995), but has no measurable effect on binding of IκB $\alpha$  to NF-κB (Huxford et al., 1999). While SRD plays a crucial part in activation of NF-κB, it has not been found to be engaged in enabling the complex formation of IκB $\alpha$ /NF-κB (Hatada et al., 1992; Jaffray et al., 1995; Sun et al., 1996). Protein crystallization and structure determination were unsuccessful for free IκB $\alpha$  due to its short lifetime and degradation within minutes. This led to the suggestion of conformational disorder in the free protein (Cervantes et al., 2009). For IκB $\alpha$  in complex with NF-κB, however, there are two protein crystal structures available (PDB IDs 1IKN and 1NFI). The truncated IκB $\alpha$  sequences in 1IKN (residues 73–292) (Huxford et al., 1998) and 1NFI (residues 71–280) (Jacobs and Harrison, 1998), however, did not reveal information about possible secondary structure elements in the SRD.

In this study, we present for the first time, the structural elements of the full length SRD of IκB $\alpha$  in complex with NF-κB and in free IκB $\alpha$ . We clearly show that the SRD displays well-defined secondary structure elements and cannot be considered as “unstructured.” In contrast, it contains three  $\alpha$ -helical regions which are stable during molecular dynamics simulations. Also, in free IκB $\alpha$  the SRD is structured albeit displaying a larger degree of flexibility and larger fluctuations.

This represents the first step in an approach to model the signal transduction cascade of the NF-κB/IκB $\alpha$  complex from IKK phosphorylation to degradation.

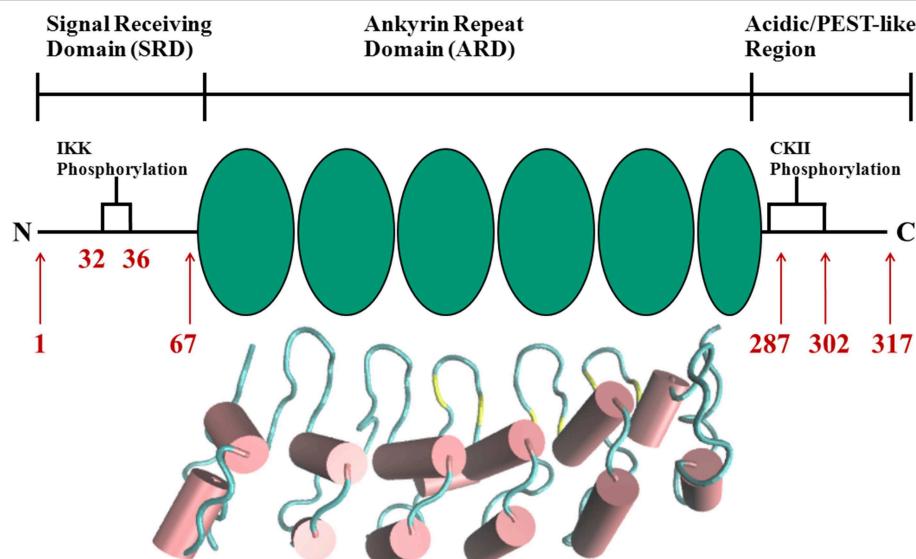
## Materials and Methods

### Structural Modeling

The secondary structure prediction of the full IκB $\alpha$  sequence was performed with SYMPRED<sup>1</sup> which builds upon results from PROF (Rost and Sander, 1994), SSPRO (Pollastri et al., 2002), YASPIN (Lin et al., 2005), and PSIPRED (Jones, 1999). In addition, JPRED3 (Leman et al., 2013), JUFO (Leman et al., 2013), NetSurfP (Petersen et al., 2009), PORTER (Pollastri and McLysaght, 2005), PredictProtein (Rost et al., 2004), and ScratchProteinPredictor (Cheng et al., 2005) were also used. All secondary structure prediction algorithms correctly identified and positioned the six ankyrin repeat units in the crystal structure (PDB ID 1IKN) in addition to four additional  $\alpha$ -helical regions in the N-terminal SRD, which is not resolved in the protein crystal structure. The consensus of predicted secondary structure elements was used for structural modeling of the SRD.

To identify a suitable structural template for modeling the SRD of IκB $\alpha$ , we used pDomThreader (Lobley et al., 2009), a profile based recognition fold method incorporating domain superfamily discrimination, which distinguished 46 probable structural templates. We chose the fourth ranked template, 1N11, as a suitable template basing our decision on a top alignment score, the degree of coverage and the structural alignment of 286 out of 317 residues in IκB $\alpha$ . pDomTHREADER (Lobley et al., 2009) makes use of the CATH database of annotation of protein

<sup>1</sup>Optimally Segmented Consensus Secondary Structure Prediction. SYMPRED is available online at: [www.ibi.vu.nl/programs/sympredwww/](http://www.ibi.vu.nl/programs/sympredwww/).



**FIGURE 2 | Schematic representation of protein domains in IkB $\alpha$ .** The N-terminal SRD is the site of phosphorylation by IKK and subsequent ubiquitination by the SCF ( $\beta$ -TrCP) E3 ligase, six ankyrin repeat units make up the central ARD domain. The ankyrin repeat units in the protein crystal structure of IkB $\alpha$  in complex with NF- $\kappa$ B (pdb entry 1IKN) are shown as a

cartoon representation. Indicated for the C-terminal PEST-like region are the CKII sites of phosphorylation. The diversity of binding sites, the great variability of  $\kappa$ B-sites in the DNA motif and the existence of suppressive and inductive NF- $\kappa$ B dimers lead to a complexity and versatility of the downstream signaling network.

structural superfamilies from the PDB (Sillitoe et al., 2013). It is an implementation of GenTHREADER, a method which predicts protein fold from sequence by integrating profile-profile alignments, secondary structure gap penalties and both classic pair and solvation potentials employing an optimized regression SVM model. pDomTHREADER is thus able to discriminate between different structural superfamilies from the protein sequences and to detect distant homology to proteins of known structure.

The structural model was generated with Prime (Jacobson et al., 2002, 2004). A manually constructed sequence alignment of our templates 1IKN and 1N11 was used. A two-template composite model was thereby constructed; residues 73–292 were based on the crystallized IkB $\alpha$  protein 1IKN while the secondary structure elements of ankyrin protein 1N11 served as basis for residues 1–98. The Build process involves coordination of the copying of the backbone atoms for aligned regions and side chains of conserved residues, building insertions and deletions in the alignment, optimization of side chains not found in template and energy minimization of those residues not derived from the templates. The Prime Build process applies the OPLS\_2005 all-atom force field for energy scoring and the Surface Generalized Born (SGB) continuum solvation model for treating solvation energies and effects. Additionally it utilizes the residue-specific side-chain rotamer and backbone dihedral libraries, derived from the non-redundant data sets extracted from the PDB.

### System Assembly and Protocol for MD Simulations

The MD simulations were carried out using Gromacs 4.5 (van der Spoel et al., 2005; Pronk et al., 2013) employing the GROMOS96

43a1 force field (Scott et al., 1999). The all-atom structural model of IkB $\alpha$  bound to the X-ray crystallographic structure of NF- $\kappa$ B included 6945 atoms in total. The protein complex was immersed in a rectangular box of dimensions  $78 \times 89 \times 145 \text{ \AA}^3$  solvated with 29686 SPC water molecules together with  $117 \text{ Na}^+$  and  $91 \text{ Cl}^-$  ions in order to neutralize the net system charge. The structural model of the free IkB $\alpha$  was immersed in a slightly smaller rectangular box of dimensions  $66 \times 68 \times 114 \text{ \AA}^3$  solvated with 15757 SPC water molecules together with  $70 \text{ Na}^+$  and  $47 \text{ Cl}^-$  ions, in total containing 50270 atoms. The LINCS algorithm (Hess et al., 1997) was applied for constraining bond lengths. Electrostatic interactions were calculated every step with the Particle-Mesh Ewald algorithm (Essmann et al., 1995). Neighbor lists were saved and reused for five steps. The simulations were performed at constant pressure of 1.0 bar with Parrinello-Rahman pressure coupling and the isotropic pressure scaling, time constant of 1.0 ps, and a system compressibility of  $4.5 \text{e-}5 \text{ bar}^{-1}$ . The temperature of the system was coupled to 300 K using the velocity-rescaling algorithm with a time constant of 0.1 ps. Newton's equations of motion were integrated using the leap-frog algorithm with a 2 fs time step.

The solvated system was first minimized with the steepest descent algorithm until a maximum force of  $< 100.0 \text{ kJ/mol}$  was reached. Equilibration of the system was initiated by 10000 steps of position-restrained MD by relaxing the solvent and keeping the non-hydrogen atoms of the system fixed. With the system relatively free of strain an NVT equilibration phase followed by an NPT phase of 10000 steps each was then carried out. Coordinates were saved every 2 ps for analysis and the production phase of the simulation ran for a total of 200 ns. Fan and Mark have shown that molecular dynamics simulation in explicit water

are able to refine homology-based protein structures within a short period of simulation (Fan and Mark, 2004). For small to medium-sized proteins (50–100 amino acids), the first 1–5 ns were able to remove initial distortions and only in few cases simulations of > 100 ns were necessary to obtain a significant reduction of RMSD. We took this as a lower threshold and added a factor of two considering the complexity of the system. Three independent replicates of our system were simulated for 600 ns in total, each starting with different initial velocities. Simulating independent replicates is a rather cost-effective way to sample conformational space (Elofsson and Nilsson, 1993).

## Results and Discussion

### Structural Elements in the Signal-receiving Domain (SRD)

In order to better understand the effects of phosphorylation and the mechanisms, which govern recognition of phosphorylated I $\kappa$ B $\alpha$ , and consequentially initiate ubiquitination, one requires a structural model of the complete N-terminal protein SRD. A BLASTp search of the first 72 residues of I $\kappa$ B $\alpha$  did not yield any significant sequence similarity with other known proteins.

A consensus-based secondary structure annotation of the full length I $\kappa$ B $\alpha$  sequence with SYMPRED was performed which builds upon results from PROF (Rost and Sander, 1994), SSPRO (Pollastri et al., 2002), YASPIN (Lin et al., 2005), and PSIPRED (Jones, 1999) (see Figure 3).

In addition, JPRED3 (Leman et al., 2013), JUFO (Leman et al., 2013), NetSurfP (Petersen et al., 2009), PORTER (Pollastri and McLysaght, 2005), PredictProtein (Rost et al., 2004), and ScratchProteinPredictor (Cheng et al., 2005) were also employed and give close to identical results (see Figure 1 of the Supplementary Material).

All six ankyrin repeat units in the crystal structure (PDB ID: 1IKN (Huxford et al., 1998)) are recovered, correctly annotated and positioned. In addition four  $\alpha$ -helical regions were detected in the N-terminal SRD, which is not resolved in the protein crystal structure (see Figure 3).

This indicates that the SRD region may contain secondary structured subregions with a high  $\alpha$ -helical content (residues 11–14, 21–29, 44–50, 56–62) not covered in any of the available I $\kappa$ B $\alpha$  crystal structures and not investigated in any of the NMR studies of free or complexed I $\kappa$ B $\alpha$ . The position of these  $\alpha$ -helices is not fixed with respect to each other and may obstruct protein crystallization of full length I $\kappa$ B $\alpha$ .

A detailed residue-based secondary structure prediction with confidence score can be found in the Supplementary Material, Figure 2. This initial finding prompted the generation of a full-length I $\kappa$ B $\alpha$  model including the SRD and the investigation of its spatial and temporal integrity and stability.

Due to the absence of sequence similarity of the SRD region to any structurally resolved protein in the PDB (~12%) sequence-based comparative modeling is not a feasible approach here. As an alternative, the choice of template was based on identification of a remote structurally related protein template with a similar secondary structure fold. The conservation of

secondary structure elements (SSEs) in protein superfamilies can guide the design of a structural model. Even when the structure of only a single member of a superfamily is known the conservation of SSEs can be used to predict the structure of other superfamily members (Mizuguchi and Blundell, 2000; Geourjon et al., 2001). Such information is useful when modeling the structure of other members of a superfamily or identifying structurally and functionally important positions in the fold. An efficient template detection allows the structural modeling to be extended even in the twilight zone of 10–30% sequence identity (Geourjon et al., 2001).

pDomTHREADER (Lobley et al., 2009) identifies 46 possible structural templates with reliable secondary structural similarity. Based on a top alignment score, the degree of coverage and the structural alignment of 286 out of 317 residues in I $\kappa$ B $\alpha$ , we chose one of the top ranked structures (1N11) as a template for modeling I $\kappa$ B $\alpha$  (for a full list of templates see Supplementary Material, Figure 3).

As an alternative approach, a combination between comparative modeling and *de novo* protein structure prediction was performed using Robetta. For proteins with detected PDB homologs, comparative models are built based on templates that are found and aligned with incorporated versions of HHSEARCH/HHpred, RaptorX, and Sparks-X. Protein domains with no close PDB homologs are generated with the Rosetta *de novo* protocol (Simons et al., 1997; Bradley et al., 2005). A structure prediction carried out by Robetta (Kim et al., 2004) for the full I $\kappa$ B $\alpha$  sequence also yielded 1N11 as the top-ranked template of choice for the generation of its structural models.

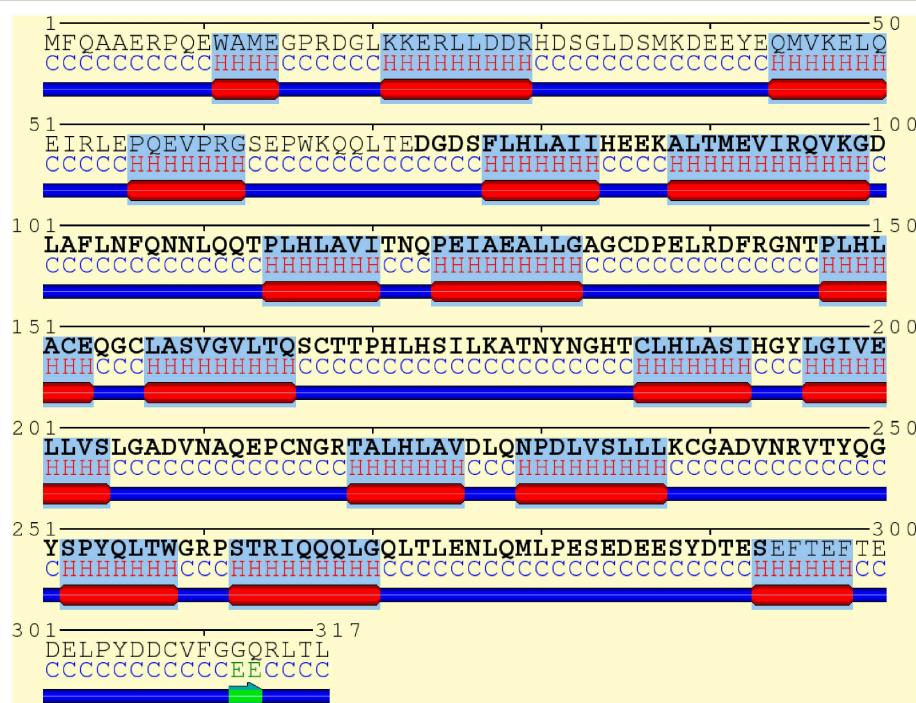
**Figure 4** shows the alignment of secondary structure elements of I $\kappa$ B $\alpha$  and 1N11 in the SRD region.

Despite an overall low primary sequence identity of only 23%, the alignment of secondary structural elements is striking. 1N11 is the crystal structure of a 12 ankyrin repeat units stack from the human ankyrinR. AnkyrinR belongs to a family of adaptor proteins that mediate anchoring between integral membrane proteins and the spectrin-actin cytoskeleton. The membrane-binding domain of ankyrins contains 24 ankyrin repeats of which the crystal structure of the human ankyrinR maps the D34 region. This region, which consists of repeats 13–24, is stacked contiguously in the shape of a left-handed superhelix (Michaely et al., 2002).

A composite model from crystallized I $\kappa$ B $\alpha$  (67–317) 1IKN and ankyrinR 1N11 PDB structures was generated. Residues at positions 73–292 were taken from the crystallized I $\kappa$ B $\alpha$  protein (PDB ID: 1IKN) and for residues 1–98 SSEs of the SRD were taken from the X-ray structure 1N11. For an overlapping stretch of residues 73–98, two  $\alpha$ -helices forming one ankyrin repeat in the 1N11 template was taken to remove any possible artifacts from truncated sequence crystallization.

### Structural Refinement by Molecular Dynamics Simulations

The protein-protein complex model was used as a starting configuration for subsequent MD refinement. The stability of the suggested secondary structural elements in the SRD and the dynamics of possible rearrangements were investigated.



**FIGURE 3 |** Consensus secondary structure annotation of full length IκB $\alpha$  (residues 1–317). Truncated, as crystallized IκB $\alpha$  from 1IKN (residues 73–293) is shown in bold letters. The six ankyrin repeat units of the ARD are

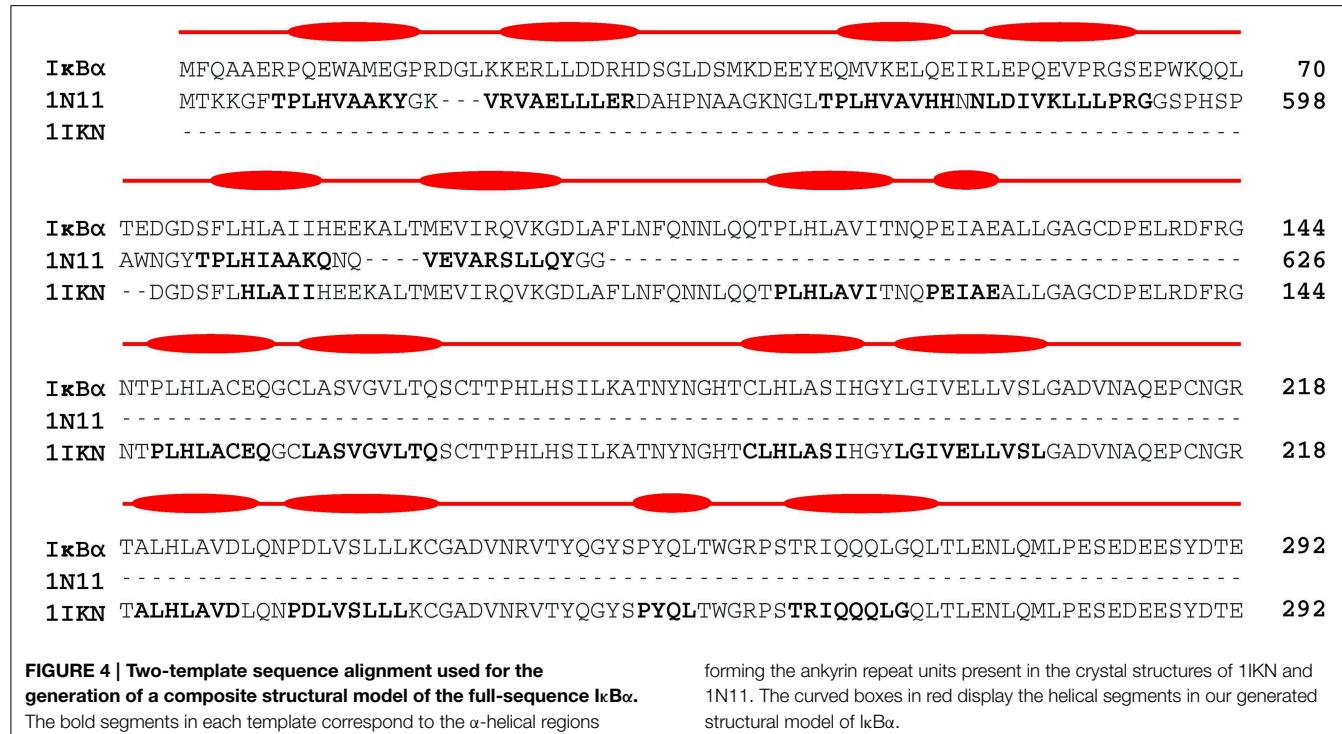
recovered, correctly annotated and positioned. Two additional  $\alpha$ -helix-loop- $\alpha$ -helix regions were detected in the N-terminal SRD. The C-terminal PEST domain displays less structural features.

In order to achieve a reliable full-sequence structural model, we performed three independent MD simulations of IκB $\alpha$  in complex with NF-κB for 200 ns each in a neutralized solvent box of about 30000 explicit water molecules. Thus, a total production simulation time of 600 ns was achieved. After energy minimization, a stepwise relaxation of the simulation setup and careful equilibration first in an NVT and then in an NPT ensemble, the general behavior of all simulation runs reveals well-behaved and stable systems. This is reflected in the conservation of total energy and temperature of the entire system (Supplemental Material, Figure 4), which is kept at a constant room temperature of 300 K (Supplemental Material, Figure 5) throughout the whole 200 ns simulation runs.

The structural stability of the IκB $\alpha$ /NF-κB complex is also monitored by calculating the root mean square displacement (RMSD) from the starting protein-protein complex structure (Supplemental Material, Figure 6). The RMSD increases sharply to 3.5–4.5 Å for the three replicate runs during the first 100 ns of the simulations, and settles at roughly 4.5–5.5 Å for the last 100 ns, indicating a well-structured stable complex. The results discussed herewith are the average findings of the three replicate runs unless otherwise stated.

In order to investigate the secondary structural profile of our IκB $\alpha$  initial model and possible structural re-arrangements, we have sectioned our 200 ns simulation into two equal parts. This provides a comparison of results at the beginning and end of the production run periods.

To better understand the inherent flexibility of our protein, the root mean square fluctuations (RMSF) of the backbone C $\alpha$  atoms of IκB $\alpha$  around the average structure were calculated (Figure 5A). The SRD N-terminal segment comprising ~75 residues clearly stands out as the most flexible region, particularly in the initial 100 ns of the simulations. Although not as expressed as in the initial 100 ns of simulations, the flexibility in the subsequent 100 ns run region is still comparatively high. We encounter the two most flexible helical regions of the whole protein, namely helices one and two, also in this region. This result indeed explains the difficulty to crystallize the SRD region. Instead an N-terminal truncation of IκB $\alpha$  was necessary to obtain protein single crystals (Cervantes et al., 2009). We see, in general, the retention of all crystallographically resolved six ankyrin repeat units in IκB $\alpha$  during the entire simulation runs (Figure 5A). While the peaks mark the hairpin loop segments connecting the  $\alpha$ -helices in each ankyrin repeat unit, the troughs of the RMSF plot correspond to helical regions. This result shows that while the helical regions are stable and not so flexible, greater flexibility is observed in the  $\beta$ -loop segments. This is in agreement with the amide  $^1\text{H}$ / $^2\text{H}$  exchange experiments followed by MALDI-TOF mass spectrometry (MS) in bound and free IκB $\alpha$  (Croy et al., 2004). The  $\beta$ -hairpins of some ankyrin repeats readily exchange amide protons for deuterons (1st, 5th, and 6th ankyrin units) whereas other units (Bouma and Strober, 2003; Schreiber et al., 2005; Viatour et al., 2005) are less solvent accessible. In particular, ankyrin repeat unit 1 remains highly solvent accessible



even in the complex. The solvent accessibility of the  $\beta$ -hairpin in ankyrin repeat unit 1 (AR 1) decreases slightly upon NF- $\kappa$ B binding (Truhlar et al., 2006).

**Figure 5B** gives the probability distributions of helical formations in the SRD of IkB $\alpha$ . Together with the RMSF of **Figure 5A**, we obtain a consistent picture of stable vs. flexible subregions in the SRD.

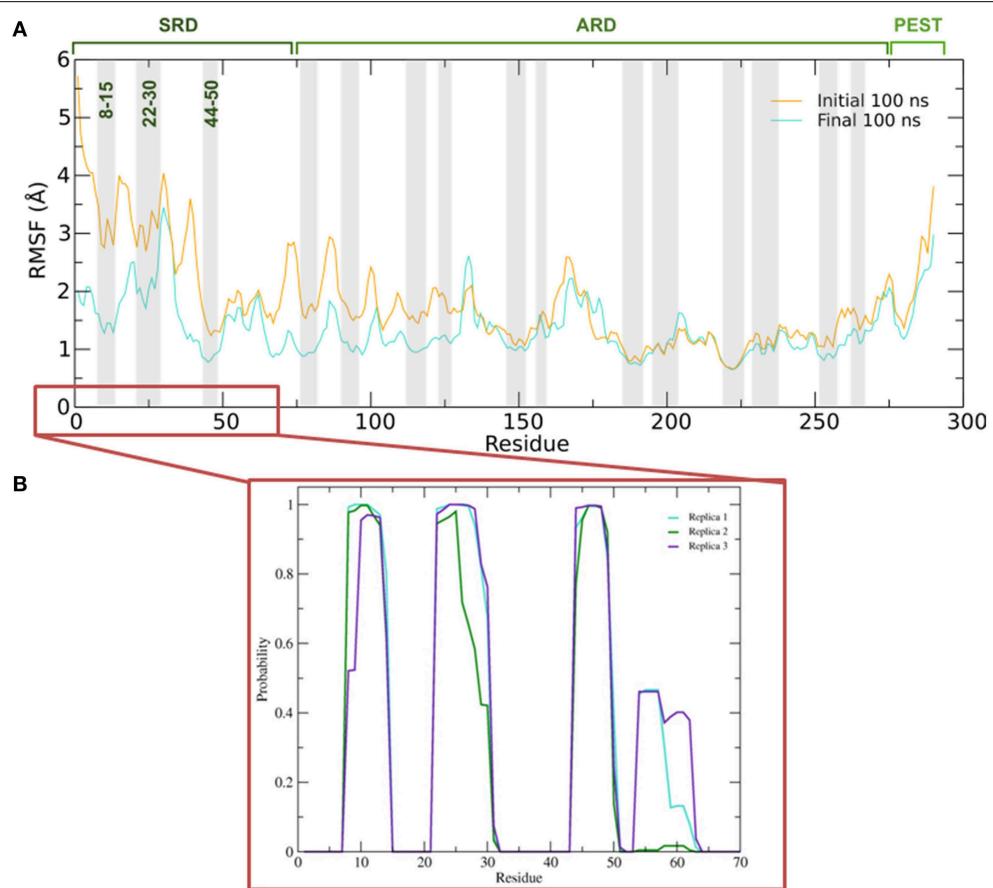
Residues 31–37 in the SRD immediately adjacent to the second  $\alpha$ -helix in the N-terminal region represent the most flexible part of IkB $\alpha$ , in the case of disregarding residues beyond 275 (**Figure 5A**). It is natural to discard residues beyond 275 from the comparison as they form a long loop and constitute a rather disordered region void of any tertiary structure. We do see the conservation of three  $\alpha$ -helices, residues between 8–15, 22–30, and 44–50 within the SRD region. These values are in agreement with the predicted secondary structure models, which identified the three  $\alpha$ -helices to lie between residues 10–13, 22–29, and 44–50. The last two  $\alpha$ -helices align perfectly, while SYMPRED predicts a somewhat shorter  $\alpha$ -helix compared to that observed in the refined structure. Furthermore, the fourth  $\alpha$ -helical element, which was positioned from residues 54–63 from the 1N11 template, no longer adopts an  $\alpha$ -helical shape but acquires instead a less ordered loop conformation (see **Figures 5B, 6**, below). Here, obviously our refinement by MD simulations is sufficient to remove the ambiguous assignment of secondary structure elements and provide a more stable conformation of this stretch of ten amino acid residues in length. All other secondary structure elements are retained during the MD simulations. This gives us confidence in the reliability of our protein-protein complex model and the existence of well-defined

forming the ankyrin repeat units present in the crystal structures of 1IKN and 1N11. The curved boxes in red display the helical segments in our generated structural model of IkB $\alpha$ .

secondary structural elements in the SRD of IkB $\alpha$  when it is in complex with NF- $\kappa$ B.

The time-evolution of secondary structure elements in the N-terminal SRD during the MD refinement is then analyzed in detail. The DSSP-annotated SSEs of the first 70 amino acid residues in IkB $\alpha$  for each of the replica systems is plotted in order to analyze the SSEs of the first 70 amino acid residues during the MD trajectory frames (see **Figure 6**). The first three  $\alpha$ -helices, residues between 8–15, 22–30, and 44–50 retain their  $\alpha$ -helical structure (blue regions) during the initial 100 ns MD simulations in all three system replicas (**Figure 6A**). They are followed by a recurring  $\beta$ -sheet turn  $\beta$ -sheet formation (green-yellow-green). This region is followed by an unstable  $\alpha$ -helix that is formed between residues 52–62. This short helix is observed only in two of the replicas (top and bottom). This segment mainly adopts the turn/bend secondary structure in the third replica. The structural stability is observed for the first three  $\alpha$ -helices throughout the entire simulations during the final 100 ns of the simulation runs. (**Figure 6B**) The temporarily formed fourth  $\alpha$ -helix, however, observed in the first 100 ns, is no longer formed and the sequence instead remains variable in its secondary structure. During most of the production runs, it takes a turn-like secondary structure (yellow) or bend (green) with short interludes of stretches of  $\beta$ -helices (gray) and  $\pi$ -helices (purple).

In **Figure 7**, we summarize our results from secondary structure prediction, initial model generation and secondary structure elements of the full-length IkB $\alpha$  obtained after MD refinement. Four helical stretches were detected from consensus SSE prediction and thus also represented the starting SRD model (top line, up to residue 70). After MD refinement, three



**FIGURE 5 | (A)** Average root mean square fluctuations (RMSF) of the backbone of IκB $\alpha$  for the initial and final 100 ns of the simulation. Shaded areas depict  $\alpha$ -helical regions at the end of the three

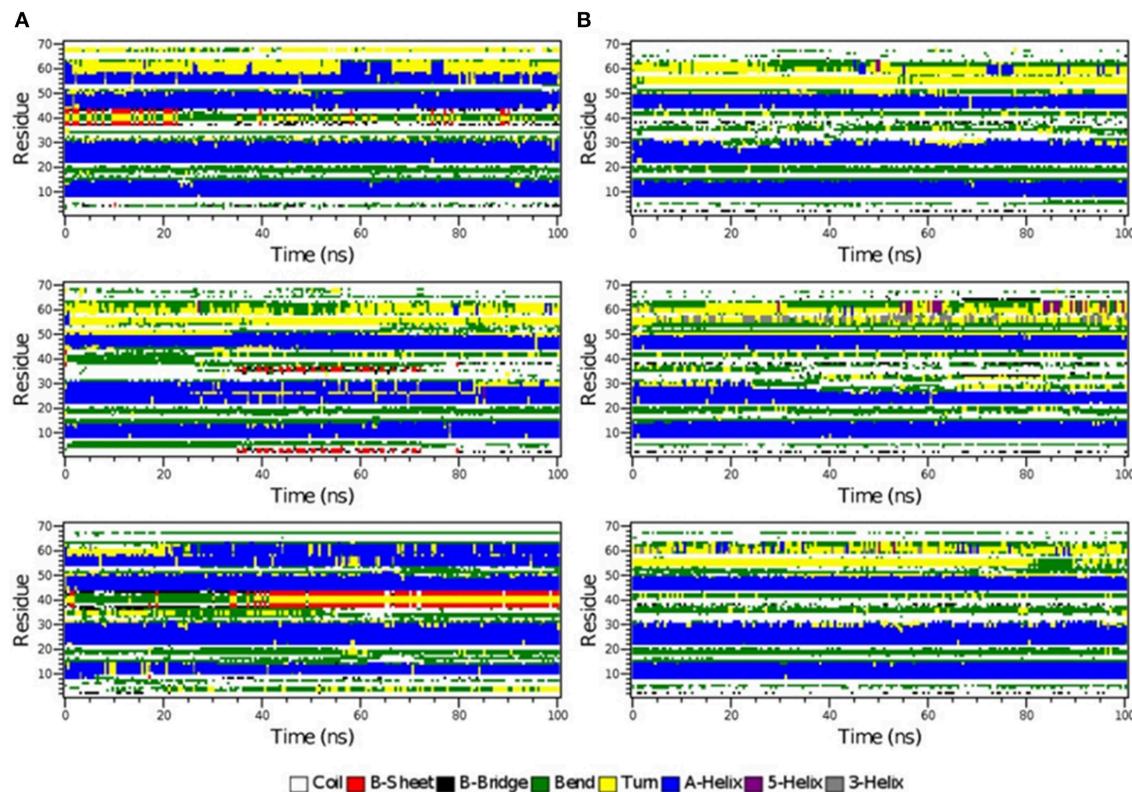
independent 200 ns simulation periods. **(B)** Probability distribution of  $\alpha$ -helix formation of the first 70 residues of the SRD of IκB $\alpha$  in complex with NF-κB.

helical stretches are structurally retained and the fourth one was not stable and adopts a disordered conformation. The ankyrin repeats of the ARD are structurally stable during the MD simulations of the protein-protein complex and well-positioned with respect to the crystal structure.

In Figure 8, we present our refined structural model of the IκB $\alpha$ -NF-κB complex (blue) portrayed together with the initial structural model (purple). The refined representative structure is depicting the last frame of a system replicate that has the lowest RMSD with respect to the average structure. This model reveals three helical structures in the previously not resolved SRD unit in addition to the six ankyrin repeats in the ANK protein domain. While the inner helix is nine residues long and extends from positions 22–30, the initial helix in the first pair of helices is eight residues long, spanning from positions 8–15 in the IκB $\alpha$ . The  $\alpha$ -helix pair is followed by a 13-residue long loop, joining this element with the consecutive  $\alpha$ -helix of seven residues long covering positions 44–50. The lengthy loop linking the third helix to the subsequent ankyrin repeat domain comprises 26 residues, and connects the unresolved N-terminal segment of IκB $\alpha$  to the crystallized ankyrin repeat domain of this protein. The structural

superpositioning of the initial and refined models of IκB $\alpha$  bound to its partner, NF-κB, reveals an ANK domain that is partly rigid and well-structured. Ankyrin repeats 4–6 remained intact and display greater stability when bound to NF-κB, while ankyrin repeats 1–3 show increased flexibility. This is in agreement with the analysis of residual dipolar coupling (RDC) of free and bound IκB $\alpha$  which showed that helix two from ankyrin repeat 3 differed most in the free and bound forms (Cervantes et al., 2009).

In particular, ankyrin repeat 1 shows the greatest displacement, which together with the SRD segment move away from NF-κB and deviate the most from the initial structure. This is in agreement with experimental studies which could show that the SRD does not contribute to the overall NF-κB binding affinity to IκB $\alpha$  (Malek et al., 1998). Also, NMR studies of IκB $\alpha$  in complex with its binding partner, NF-κB, show a more flexible ankyrin 1–4 domain in comparison to rather rigid ankyrin repeats 5–6 (Sue et al., 2008). An earlier amide H/D exchange study (Truhlar et al., 2006) indicated that when in complex with NF-κB, ankyrin repeats five and six-fold into compact domains upon binding to NF-κB. Along with ankyrin repeats 5 and 6, ankyrin repeat 1 is another region seen to



**FIGURE 6 |** The secondary structure elements of the first 70 N-terminal residues of IkB $\alpha$  in complex with NF- $\kappa$ B as calculated by DSSP for the three system replicas during the initial 100 ns (A) and final 100 ns (B) of the simulation.

display greater conformational flexibility as observed here in the refined structure of IkB $\alpha$ . The RMSFs of amino acid residues mapped onto the C $\alpha$ -backbone atoms of IkB $\alpha$  can be seen in the Supplementary Material, Figure 7.

### Conformational Change Induced in IkB $\alpha$ in Its Bound Form to NF- $\kappa$ B

Thus, so far we have looked at the structural elements in IkB $\alpha$  only. In the crystal structures and in our simulations however, IkB $\alpha$  is in complex with NF- $\kappa$ B (the RelA/p50 heterodimer) and for this reason it is imperative to look at the conformation of IkB $\alpha$  in relation to its binding partner, NF- $\kappa$ B, and see how the nature of this association was affected. The protein-protein surface area of interaction is larger than 4000 Å<sup>2</sup> and all six ankyrin repeat units are involved in forming a non-contiguous contact surface. We discuss here in particular electrostatic and hydrogen bonding interactions between IkB $\alpha$  and RelA/p50. The hydrogen bonds that are discussed here remain intact for longer than 10% of the simulation time and occur in at least two of the replicate simulations.

### The IkB $\alpha$ /RelA Interface

IkB $\alpha$  binds to RelA by forming a number of hydrogen bonds between different regions of each protein (Table 1). Several residues situated on ankyrin repeats 5 and 6 form hydrogen bonds with residues located on both the RelA

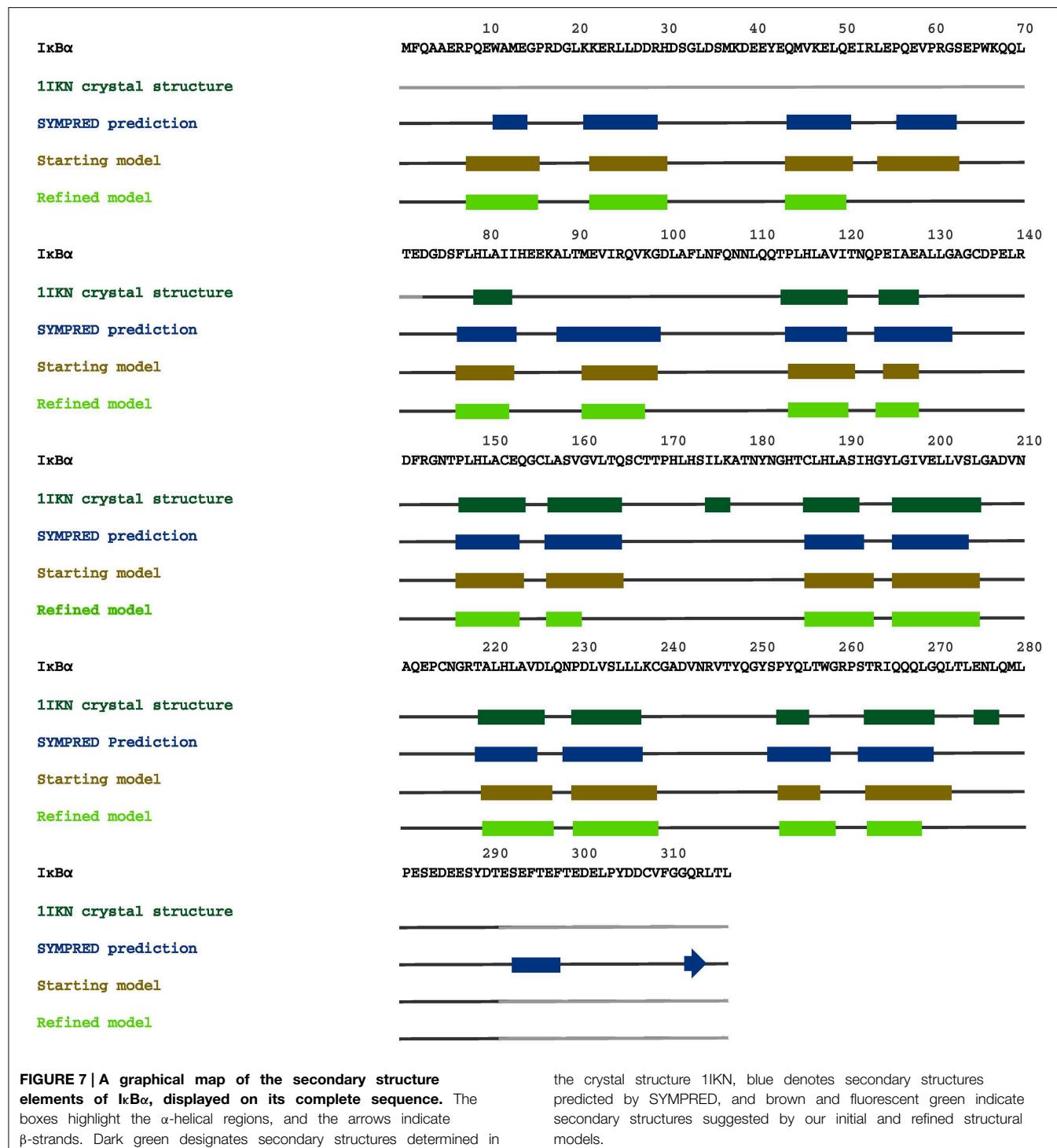
dimerization subunit and the RelA amino-terminal. The IkB $\alpha$  carboxy-terminal residues are in close contact with regions on the amino-terminal and dimerization subunit of RelA and form several hydrogen bonds.

The other major source of stabilization is via electrostatic interactions from the salt bridge interactions between the carboxy-terminal of IkB $\alpha$  and different regions of RelA (Table 2). The ARD region of IkB $\alpha$  contributes to the IkB $\alpha$ /RelA stabilization by forming salt bridges between Asp226 and Arg218 on ANK5 and between Arg253 and Asp243 as well as Glu211 on the dimerization component. In addition, Arg264 on ANK6 interacts with Glu22 on the amino-terminal of RelA. Specifically the interaction between Arg218 and Asp243 is also observed to form in the crystal structure of IkB $\alpha$  (Huxford et al., 1998).

The elongated and relatively flexible 13 residue carboxy-terminal of RelA, known as the NLS polypeptide, extends across ankyrin repeats 1–3 and makes several contacts with residues present on the loops and helical regions of these ankyrin repeats, forming both hydrogen bonds and salt bridges.

### The IkB $\alpha$ /p50 Interface

A number of residues on ankyrin repeats 4–6 interact with the dimerization domain on p50 by forming hydrogen bonds. Among these interactions, Tyr181 has previously been shown to be a key player in the interaction between NF- $\kappa$ B and IkB $\alpha$

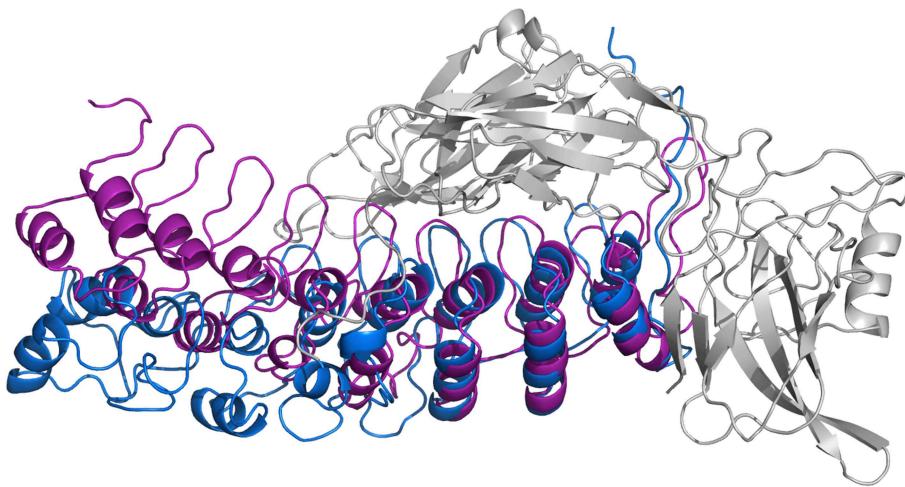


**FIGURE 7 |** A graphical map of the secondary structure elements of IkB $\alpha$ , displayed on its complete sequence. The boxes highlight the  $\alpha$ -helical regions, and the arrows indicate  $\beta$ -strands. Dark green designates secondary structures determined in

the crystal structure 1IKN, blue denotes secondary structures predicted by SYMPRED, and brown and fluorescent green indicate secondary structures suggested by our initial and refined structural models.

(Huxford et al., 1998). Eminently, residues Cys215, Tyr248, and Arg252 on the p50 subunit are among those reported to form interactions in the crystal structure of IkB $\alpha$ . The amino acid residues Lys352-Asp353 located on the carboxy-terminal of p50 engage in additional hydrogen bond interaction with the residues Asp73, Gln107, and Asn109 situated on ankyrin repeats 1 and 2. The interaction between IkB $\alpha$  and p50 is further stabilized by

electrostatic interactions. The carboxy-terminal PEST sequence residues Glu286-Glu287, Asp290, GLU292 in IkB $\alpha$  take part in forming salt bridges with the residues Lys249, Lys272, Arg305 on the amino-terminal and the interconnecting loops on the “top” of the p50 subunit. Ankyrin repeats 1–3 and the SRD in IkB $\alpha$  and the carboxy-terminal and an interconnecting loop at the “bottom” of p50 participate in another set of salt bridge network



**FIGURE 8 | Ribbon diagrams of the three-dimensional initial structure (purple) and the refined structure after a 200 ns MD simulation (blue) of IκB $\alpha$ .**  
The structures are shown in comparison by superimposing IκB $\alpha$ 's binding partner NF-κB (gray).

**TABLE 1 | Hydrogen bond contacts between IκB $\alpha$  and the p50/RelA subunits of NF-κB.**

IκB $\alpha$	RelA	IκB $\alpha$	p50	RelA	p50
<b>HYDROGEN BONDS</b>					
GLY155	ARG297	ASP73	LYS352	ARG198	HIS304
LEU157	<b>ARG297</b>	GLN107	LYS352	<b>ASN200</b>	<b>ASP254</b>
ASN216	<b>ASP243</b>	ASN109	ASP353	<b>ASP243</b>	<b>ARG252</b>
ASP226	SER238	ASN109	LYS352	<b>HIS245</b>	CYS270
THR247	ASP243	<b>TYR181</b>	THR256		
GLN249	ASP243	<b>ASN182</b>	THR256		
<b>TRP258</b>	GLN26	GLY183	THR256		
<b>GLY259</b>	GLN241	CYS215	MET253		
GLN266	ILE24	<b>TYR248</b>	LYS343		
GLN267	<b>GLU22</b>	<b>TYR248</b>	GLU341		
GLN271	LEU179	GLN249	<b>ARG252</b>		
GLN271	VAL21				
LEU280	GLN29				
SER283	GLU222				
ASP285	GLN247				
GLU286	GLN247				
GLU286	THR191				
SER288	GLN247				

Residues found in interactions in the crystal structure 1IKN are shown in bold. All Bonds are present for more than 10% of the total simulation time in at least 2 of the replicate simulations.

involving residues Glu41, Glu72-Asp73, Glu138, and Lys323, Lys352, Lys354, on respective chain. Notably, with one single exception, the acidic residues are contributed by IκB $\alpha$ , whereas the basic residues are to be found on the p50 subunit.

### The RelA/p50 Interface

The dimerization interface takes part in several hydrogen bonds formed by 8 residues including an Asp254(p50)/Asn200(RelA)

**TABLE 2 | Salt bridge formations between IκB $\alpha$  and the p50/RelA subunits of NF-κB.**

IκB $\alpha$	RelA	IκB $\alpha$	p50	RelA	p50
<b>SALT BRIDGES</b>					
GLU85	ARG302	GLU41	LYS354	ARG198	<b>ASP302</b>
GLU85	LYS301	GLU72	LYS352	ARG198	GLU265
GLU86	ARG302	GLU72	LYS354	ARG201	<b>ASP254</b>
GLU125	ARG302	ASP73	LYS354	ARG201	GLU265
ARG143	ASP294	GLU138	LYS323	<b>GLU211</b>	<b>ARG252</b>
GLU153	ARG295	ARG143	GLU350	<b>ASP217</b>	<b>ARG305</b>
GLU153	<b>ARG297</b>	GLU286	ARG305	<b>ASP243</b>	<b>ARG252</b>
<b>ARG218</b>	<b>ASP243</b>	GLU286	LYS272	<b>ARG246</b>	<b>ASP271</b>
ARG218	GLU211	GLU287	LYS272		
ASP226	ARG253	ASP290	LYS272		
ARG264	<b>GLU22</b>	GLU292	<b>LYS249</b>		
GLU282	ARG30	GLU292	LYS272		
GLU282	<b>ARG158</b>				
GLU282	LYS79				
GLU284	ARG246				
GLU284	LYS79				
ASP285	LYS218				
GLU286	LYS218				
GLU287	ARG246				
GLU287	LYS218				
ASP290	LYS221				
GLU292	ARG246				

Residues found in interactions in the crystal structure 1IKN are shown in bold. All salt bridges occur in at least 2 of the replicate simulations throughout the whole simulation.

hydrogen bond. This hydrogen bond can also be found in the crystal structure and is considered one of the most critical interactions in discriminating subunit dimerization specificity among NF-κB dimers (Huang et al., 1997; Chen et al., 1998; Huxford et al., 1998). The other hydrogen bonds include His304(p50)/Arg198(RelA), Arg252(p50)/Asp243(RelA), Cys

270(p50)/His245(RelA). The RelA/p50 dimer interface is additionally stabilized by electrostatic interactions. Several residues form salt bridges between the two subunits. Two of these include salt bridges that are also reported for the crystal structure namely residues Asp217 and Asp271 on the p50 subunit and Arg305 and Arg246 on the RelA component, respectively (Huxford et al., 1998).

### Free IkB $\alpha$ vs. Bound IkB $\alpha$

All the simulations discussed above were describing the stable, long-living complex of IkB $\alpha$  with its binding partner NF- $\kappa$ B, as revealed in their crystal structures. All efforts to crystallize IkB $\alpha$  in its unbound state have been unsuccessful (Croy et al., 2004). For this reason, additional simulations of full-length free IkB $\alpha$  in solution were performed and compared with the more stable NF- $\kappa$ B-bound state.

The simulation setups followed the same procedure as for the bound IkB $\alpha$ , resulting in three replicated systems of 200 ns each. Conservation of total energy and temperature of the three simulations (Supplementary Material, Figure 8) points to systems that have reached a stable state. In contrast to the bound IkB $\alpha$ , the RMSF of the free state of IkB $\alpha$  remains on average  $\sim$ 1 Å higher compared to its complexed state. This points to a higher degree of flexibility of free IkB $\alpha$  compared to its complexed state. The RMSF of the backbone of the protein around the average structure in the modeled SRD remains the most flexible domain throughout the protein in addition to the unstructured C-terminal region (see Figure 9A top). The probability distributions of the helical propensity in the SRD of the bound IkBa reveal (Figure 9B) the first three helical segments to be stable throughout the whole simulation. In free IkB $\alpha$ , although the first three helical segments are present in all three replicate simulations, we observe different probabilities across the different replicate simulations. Bound IkB $\alpha$  displays a narrower distribution of probabilities of helical regions and this indicates to a stabilization of the SRD upon complexation with NF- $\kappa$ B. The fourth initially assigned helix in the SRD varies in both length and probability in both the bound and free forms of IkB $\alpha$ , indicating that this fourth helix is not well-defined and not stable during MD refinement (see above).

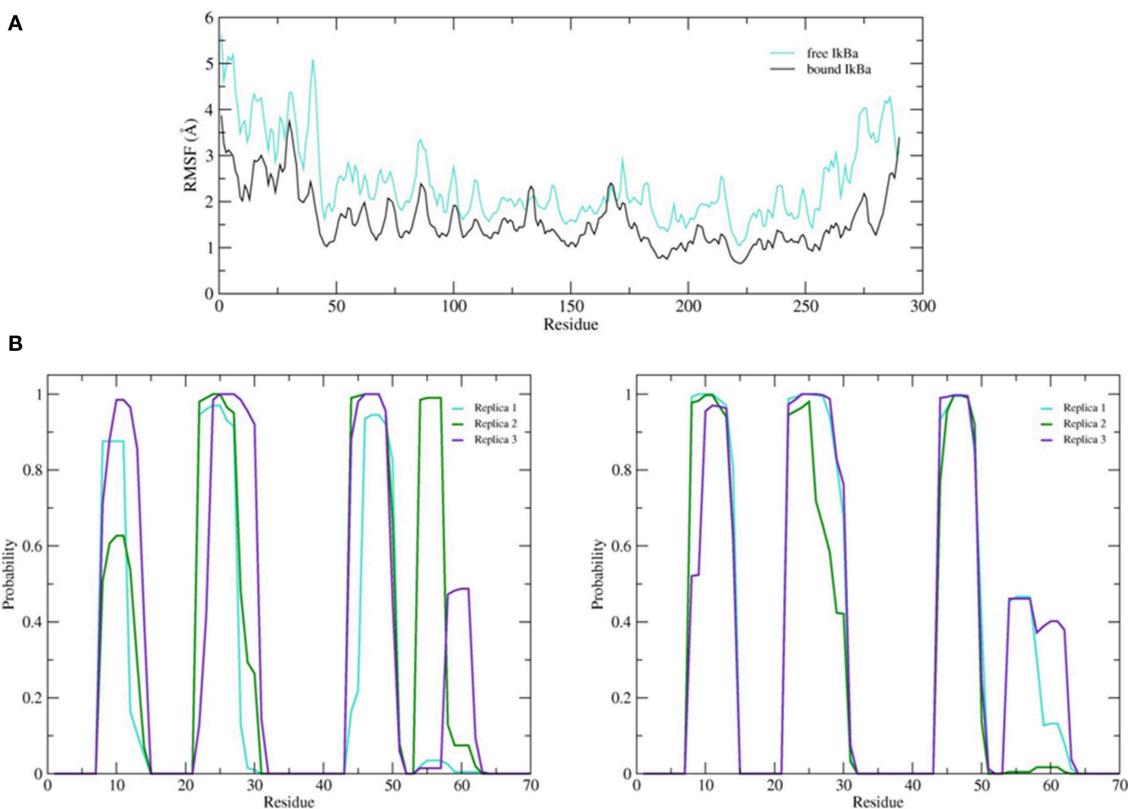
The secondary structure evolution of the first 70 amino acid residues in the SRD of the free IkB $\alpha$  (see Figure 10A) reveals greater differences in the SRD in terms of secondary structure element evolution in comparison to the bound IkB $\alpha$ . The first helix in the free IkB $\alpha$  is considerably shorter than its counterpart in the bound IkB $\alpha$ . During the first 100 ns of the simulations, this helix can be clearly distinguished whereas it is only present in two of the replicate runs in the final simulation period. The second and third helices remain intact throughout the entire 200 ns simulations in all three replicate runs, which is very similar to the pattern seen in the bound IkB $\alpha$  simulations. In contrast to the bound IkB $\alpha$ , here we observe the formation of a 4 residue long fourth helix in two of the replicate runs; in one of the simulations this helix is present during the entire simulation, whereas in the other run it appears in the last 100 ns of the simulations with irregular intervals. In a previous study (Pons

et al., 2007), the conformations of a short 24 amino acid peptide (residues 21–44) of the doubly phosphorylated free IkB $\alpha$  were characterized by NMR spectroscopy and MD simulations and compared to its  $\beta$ -TrCP bound state using saturation transfer difference NMR. The conformational observation agreed on the presence of a bend between residues 30 and 36 in both states of the phosphorylated peptide, a trend which we also observe throughout our simulations of the free and NF- $\kappa$ B bound states of IkB $\alpha$ . While the N-terminal of amino acids 30 to 36 is preceded by a short  $\alpha$ -helix and the C-terminal succeeded by a region of  $\beta$ -sheet-turn- $\beta$ -sheet flanked by bends in the free and bound states of IkB $\alpha$  in this study, Pons et al. observed disordered N- and C-terminal segments in the free IkB $\alpha$  vs. the adoption of turns in the bound state IkB $\alpha$ . This difference in results can be rationalized from the truncation of the peptide which could have influenced the conformational integrity of the N- and C-terminals, an effect which would not be detectable in our structural models of the full-length IkB $\alpha$ .

**Figure 10B** shows the interatomic distance matrices depicting the smallest distance between residue pairs in the SRD of IkBa for both free (top) and complexed IkB $\alpha$  (bottom). The distance matrices of all three replicates are very similar and there are no large differences in interatomic distances upon NF- $\kappa$ B binding. The red and yellow colors indicate shorter distances between the residues and are more detectable for regions where helical segments are present in the SRD. In both the unbound and free forms of IkB $\alpha$ , the fourth segment is less apparent across the replicates.

There are, however, also apparent stretches of amino acids which display a higher degree of flexibility upon NF- $\kappa$ B complexation (see Figure 9A; top). The residues around positions 133 and 167 become more flexible upon protein-protein complex formation. These positions correspond to loop regions following the outer helices in AR2 and AR3. This was also found by analyzing residual dipolar coupling (RDC) of ARs 1–4 (Cervantes et al., 2009).

Another interesting comparison between the free and bound IkB $\alpha$  structures is the solvent accessible surface area (SASA) or the relative solvent accessible area (RSA) of the phosphorylation and ubiquitination sites located on the SRD (Table 3). These sites (Ser32 and Ser36 for phosphorylation and Lys21 or Lys22 for ubiquitination) ought to become accessible by the kinase IKK and the E3 ligase, respectively, in the complexed form of IkB $\alpha$ . The RSA is computed by the SASA of the residue normalized by the accessible surface area of that residue in its extended tri-peptide (Gly-X-Gly) conformation. By setting a threshold of < 20% for buried residues, Ser32 and Ser36 are both surface-exposed in the bound IkB $\alpha$ , while in its free state only Ser32 lies above the threshold. SER36 in the free state has an RSA of 9.3%, which is considerably lower than the threshold and can be considered to be a buried residue. As regards to the ubiquitination sites, Lys21 stays well-buried in both the free and bound states of IkB $\alpha$ . However, Lys22 with an RSA of well over 60% in both states of IkB $\alpha$  remains surface exposed. Thus, in the bound-form of IkB $\alpha$  the phosphorylation Ser32 and Ser36 sites are accessible by the IKK and we suggest Lys22 to be the putative site of ubiquitination.



**FIGURE 9 | (A)** Average root mean square fluctuations (RMSF) of the backbone of the free IkB $\alpha$  (cyan) in comparison to the one in complex with NF-κB (black). **(B)** Probability distributions of  $\alpha$ -helix formation of the first 70 residues of the SRD. Left: free IkB $\alpha$ . Right: IkB $\alpha$  in complex with NF-κB.

## Conclusion and Outlook

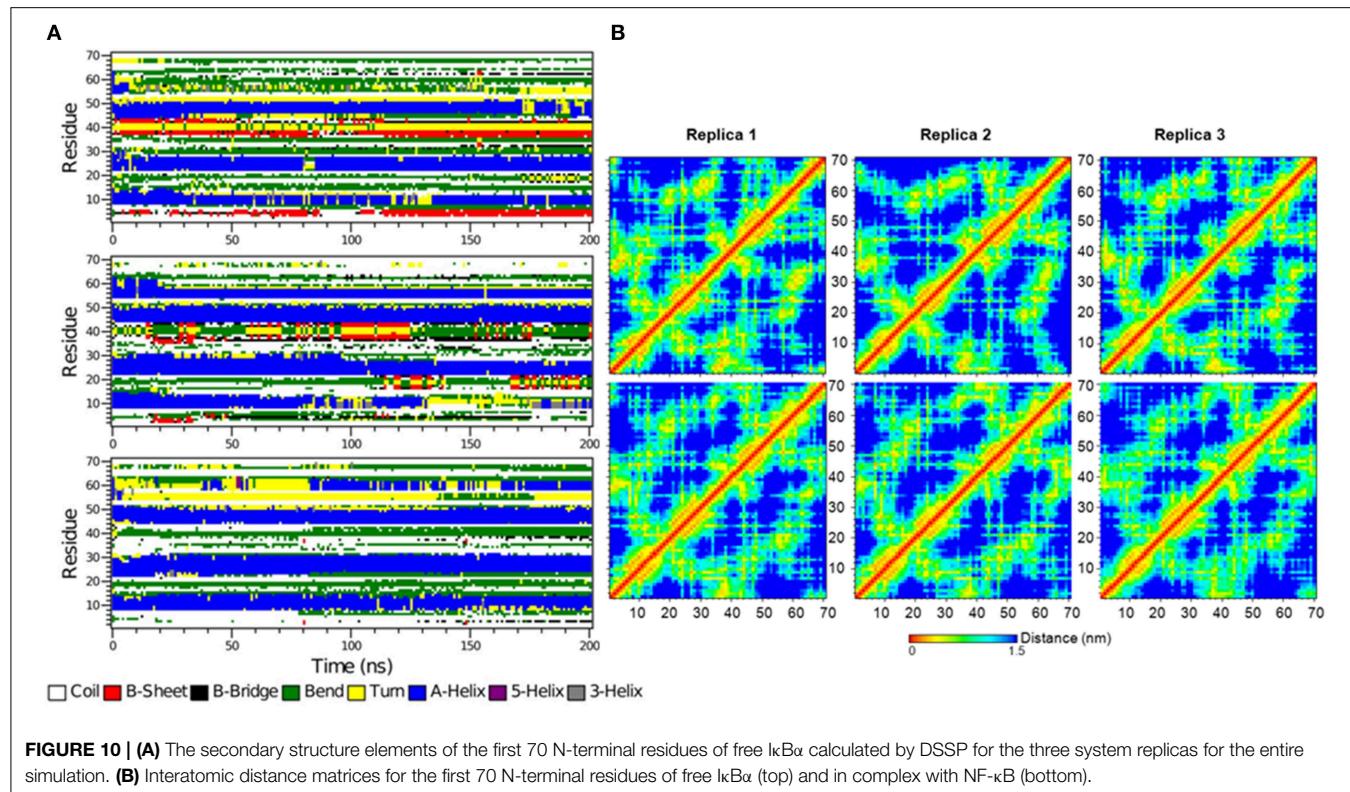
The lack of crystallographic information about the SRD of IkB $\alpha$  has led to the speculation of a disordered N-terminal extension that could not be crystallized. Furthermore, the SRD was shown not to be a major contributor to the IkB $\alpha$ /NF-κB binding affinity (Hatada et al., 1992; Jaffray et al., 1995; Sun et al., 1996). This particular region, however, contains the two highly conserved serine residues, 32 and 36, which are the sites of phosphorylation by IKKs and involved in the regulation of IkB $\alpha$ . After two-fold phosphorylation, the IkB $\alpha$ /NF-κB becomes poly-ubiquitinated at Lysine residues 21 and 22, the protein-protein complex releases NF-κB and IkB $\alpha$  to be degraded *in vivo*. Thus, investigation of dynamical and structural properties of this domain is very important for understanding of the post-translational modifications and signaling properties of this domain.

Previous structural investigations (Mizuguchi and Blundell, 2000; Ferreiro and Komives, 2010) have used bioinformatics tools like PONDR (Geourjon et al., 2001) and IUPRED (Simons et al., 1997) to annotate the potential disorder and flexibility of the ankyrin units of the ARD. For IkB $\alpha$  (67–317) the  $\beta$ -hairpin loops of each ankyrin repeat displayed a greater degree of disorder than the  $\alpha$ -helical regions. Furthermore, ankyrin repeat units 2, 3, and 4 were more structured than units

1, 5, and 6. In particular the C-terminal part of the ARD and the PEST domain were considered as being intrinsically disordered. For the SRD, however, the results were not unambiguous.

The lack of sequence identity to any known three-dimensional protein structure obstructed the comparative modeling approach. A predicted well-ordered secondary structure profile of the SRD, however, allowed the assignment of  $\alpha$ -helical structural elements in this region. Protein domain threading suggested the poly-ankyrin human ankyrinR as a suitable structural but not sequence-based model. From a secondary structure alignment, a structural model for the SRD of IkB $\alpha$  in complex with NF-κB was generated and refined by multiple-template MD simulations. In the final model, the SRD region was shown to contain three stable  $\alpha$ -helices. The MD simulations resolved ambiguities of secondary structure elements for residues 54–63 which were  $\alpha$ -helical in the template but rather occupied a loop conformation after MD refinement.

The structural stability of the model has been validated through long MD simulations. For the protein domain annotation of IkB $\alpha$ , our results clearly display stable helical conformations in the N-terminal SRD. Interestingly, the amino acid sequence composition of the SRD is in good agreement with the consensus sequence for typical ankyrin repeat domains



**FIGURE 10 | (A)** The secondary structure elements of the first 70 N-terminal residues of free IκB $\alpha$  calculated by DSSP for the three system replicas for the entire simulation. **(B)** Interatomic distance matrices for the first 70 N-terminal residues of free IκB $\alpha$  (top) and in complex with NF-κB (bottom).

**TABLE 3 | Solvent accessible surface area (SASA) and relative surface area (RSA) of the free and bound IκB $\alpha$ .**

	SASA (Å)		RSA (%)	
	Bound IκB $\alpha$	Free IκB $\alpha$	Bound IκB $\alpha$	Free IκB $\alpha$
SER32	61.2 ± 18.5	43.4 ± 15.7	50.2 ± 15.2	35.6 ± 12.8
SER36	50.6 ± 14.0	11.3 ± 9.1	41.5 ± 11.5	9.3 ± 7.4
LYS21	18.5 ± 4.8	39.2 ± 16.4	8.8 ± 2.3	18.6 ± 7.8
LYS22	135.6 ± 6.0	127.5 ± 17.9	64.3 ± 2.8	60.4 ± 8.5

The accessible surface areas of serine and lysine are 122 Å and 211 Å, respectively as calculated by Miller et al. (1987). The SASA values are the averages of the three replicate simulations over the 200 ns of total simulation time.

(Mosavi et al., 2002, 2004). This provides additional support for the stability of secondary structural elements in the SRD. The C-terminal PEST domain, however, displays large atomic fluctuations and a high degree of flexibility, which make the reliable assignment of any secondary structural element impossible.

When compared to free IκB $\alpha$  in solution, the complexed IκB $\alpha$  displays a significantly reduced degree of intrinsic flexibility and disorder. In particular, the SRD and the PEST domains show significantly reduced flexibility upon NF-κB binding. For the central ARD, the picture is less clear. Different ARs were shown to possess different solvent accessibilities by H/D exchange and MS (Croy et al., 2004). This was interpreted as an increased structural flexibility for ARs 1, 5, and 6 but

retaining all of the SSEs at the same time. According to our results, however, the degree of solvent accessibility is not determined by helical flexibility of ARs 1, 5, 6 but rather by their adjacent hairpin/loop regions. This is in agreement with the first four ankyrin repeat units exhibiting little change in solvent accessibility upon NF-κB binding (Truhlar et al., 2006) whereas ARs 5 and 6 undergoing a coupled folding and binding process. When we compared the free IκB $\alpha$  with the IκB $\alpha$  bound to NF-κB the order parameters from NMR and aMD compared well and showed not drastic structural rearrangement upon complexation (Cervantes et al., 2009). However, for IκB $\alpha$  the order parameters were generally lower for variable loop regions parameters when in complex with NF-κB than for free IκB $\alpha$ . Our simulations can rationalize this finding in terms of an increase in intrinsic flexibility of the loop regions upon complexation and thus it is in good agreement with the hypothesis of a degree of protein “fuzziness” in the IκB $\alpha$ /NF-κB complex (Komives, 2012).

To experimentally address the structural composition of the IκB $\alpha$  N-terminal SRD, it is desirable to perform NMR studies using a recombinant full-sequence IκB $\alpha$  in free solution and/or in complex with NF-κB. The overexpression of human IκB $\alpha$  in sufficient yield is currently being performed in our laboratory.

The complete structural model of IκB $\alpha$  in complex with NF-κB has now been prepared and will be used as starting structures for upcoming multi-scale investigations regarding the structural basis for IκB $\alpha$  signaling after

phosphorylation and ubiquitination at the molecular level.

## Acknowledgments

We thank the Max Planck Society for the Advancement of Science and the “Research Center for Dynamic Systems (CDS): Biosystems Engineering” funded by the Excellence

Initiative of the State of Saxony-Anhalt for financial support.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolt.2015.00032>

## References

- Bouma, G., and Strober, W. (2003). The immunological and genetic basis of inflammatory bowel disease. *Nat. Rev. Immunol.* 37, 521–533. doi: 10.1038/nri1132
- Bradley, P., Misura, K. M. S., and Baker, D. (2005). Toward high-resolution *de novo* structure prediction for small proteins. *Science* 309, 1868–1871. doi: 10.1126/science.1113801
- Cervantes, C. F., Markwick, P. R. L., Sue, S. C., McCammon, J. A., Dyson, H. J., and Komives, E. A. (2009). Functional dynamics of the folded ankyrin repeats of IκB $\alpha$  revealed by nuclear magnetic resonance. *Biochemistry* 48, 8023–8031. doi: 10.1021/bi900712r
- Chen, F. E., Huang, D. B., Chen, Y. Q., and Ghosh, G. (1998). Crystal structure of p50/p65 heterodimer of transcription factor NF-κB bound to DNA. *Nature* 391, 410–413. doi: 10.1038/34356
- Cheng, J., Randall, A. Z., Sweredoski, M. J., and Baldi, P. (2005). SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res.* 33, W72–W76. doi: 10.1093/nar/gki396
- Croy, C. H., Bergqvist, S., Huxford, T., Ghosh, G., and Komives, E. A. (2004). Biophysical characterization of the free IκB $\alpha$  ankyrin repeat domain in solution. *Protein Sci.* 13, 1767–1777. doi: 10.1110/ps.04731004
- Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* 14, 892–893. doi: 10.1093/bioinformatics/14.10.892
- Dyson, H. J., and Komives, E. A. (2012). Role of disorder in IκB–NFκB interaction. *IUBMB Life* 64, 499–505. doi: 10.1002/iub.1044
- Elofsson, A., and Nilsson, L. (1993). How consistent are molecular-dynamics simulations – comparing structure and dynamics in reduced and oxidized escherichia-coli thioredoxin. *J. Mol. Biol.* 233, 766–780. doi: 10.1006/jmbi.1993.1551
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh ewald method. *J. Chem. Phys.* 103, 8577–8593. doi: 10.1063/1.470117
- Fan, H., and Mark, A. E. (2004). Refinement of homology-based protein structures by molecular dynamics simulation techniques. *Protein Sci.* 13, 211–220. doi: 10.1110/ps.03381404
- Ferreiro, D. U., Cervantes, C. F., Truhlar, S. M., Cho, S. S., Wolynes, P. G., and Komives, E. A. (2007). Stabilizing IκB $\alpha$  by “consensus” design. *J. Mol. Biol.* 365, 1201–1216. doi: 10.1016/j.jmb.2006.11.044
- Ferreiro, D. U., and Komives, E. A. (2010). Molecular mechanisms of system control of NF-κB signaling by IκB $\alpha$ . *Biochemistry* 49, 1560–1567. doi: 10.1021/bi901948j
- Geourjon, C., Combet, C., Blanchet, C., and Deléage, G. (2001). Identification of related proteins with weak sequence identity using secondary structure information. *Protein Sci.* 10, 788–797. doi: 10.1110/ps.30001
- Ghosh, G., Wang, V. Y.-F., Huang, D.-B., and Fusco, A. (2012). NF-κB regulation: lessons from structures. *Immunol. Rev.* 246, 36–58. doi: 10.1111/j.1600-065X.2012.01097.x
- Hatada, E. N., Nieters, A., Wulczyn, F. G., Naumann, M., Meyer, R., Nucifora, G., et al. (1992). The ankyrin repeat domains of the NF-κB precursor p105 and the protooncogene bcl-3 act as specific inhibitors of NF-κB DNA binding. *Proc. Natl. Acad. Sci. U.S.A.* 89, 2489–2493. doi: 10.1073/pnas.89.6.2489
- Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472.
- Huang, D. B., Huxford, T., Chen, Y. Q., and Ghosh, G. (1997). The role of DNA in the mechanism of NFκB dimer formation: crystal structures of the dimerization domains of the p50 and p65 subunits. *Structure* 5, 1427–1436. doi: 10.1016/S0969-2126(97)00293-1
- Huxford, T., Hoffmann, A., and Ghosh, G. (2011). Understanding the logic of IκB:NF-κB regulation in structural terms. *Curr. Top. Microbiol. Immunol.* 349, 1–24. doi: 10.1007/82\_2010\_99
- Huxford, T., Huang, D. B., Malek, S., and Ghosh, G. (1998). The crystal structure of the IκB $\alpha$ /NF-κB complex reveals mechanisms of NF-κB inactivation. *Cell* 95, 759–770. doi: 10.1016/S0092-8674(00)81699-2
- Huxford, T., Malek, S., and Ghosh, G. (1999). Structure and mechanism in NF-κB/IκB signaling. *Cold Spring Harb. Symp. Quant. Biol.* 64, 533–540. doi: 10.1101/sqb.1999.64.533
- Jacobs, M. D., and Harrison, S. C. (1998). Structure of an IκB $\alpha$ /NF-κB complex. *Cell* 95, 749–758. doi: 10.1016/S0092-8674(00)81698-0
- Jacobson, M. P., Friesner, R. A., Xiang, Z., and Honig, B. (2002). On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* 320, 597–608. doi: 10.1016/S0022-2836(02)00470-9
- Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E., et al. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins* 55, 351–367. doi: 10.1002/prot.10163
- Jaffray, E., Wood, K. M., and Hay, R. T. (1995). Domain organization of IκB $\alpha$  and sites of interaction with NF-κB p65. *Mol. Cell. Biol.* 15, 2166–2172.
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202. doi: 10.1006/jmbi.1999.3091
- Kim, D. E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* 32, W526–W531. doi: 10.1093/nar/gkh468
- Komives, E. (2012). “Consequences of fuzziness in the NFκB/IκB $\alpha$  interaction,” in *Fuzziness - Structural Disorder in Protein Complexes*, eds M. Fuxreiter and P. Tompa (Austin, TX: Landes Bioscience), 74–85.
- Leman, J. K., Mueller, R., Karakas, M., Woetzel, N., and Meiler, J. (2013). Simultaneous prediction of protein secondary structure and transmembrane spans. *Proteins* 81, 1127–1140. doi: 10.1002/prot.24258
- Lin, K., Simossis, V. A., Taylor, W. R., and Heringa, J. (2005). A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics* 21, 152–159. doi: 10.1093/bioinformatics/bth487
- Lobley, A., Sadowski, M. I., and Jones, D. T. (2009). pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics* 25, 1761–1767. doi: 10.1093/bioinformatics/btp302
- Malek, S., Huxford, T., and Ghosh, G. (1998). IκB $\alpha$  Functions through direct contacts with the nuclear localization signals and the DNA binding sequences of NF-κB. *J. Biol. Chem.* 273, 25427–25435. doi: 10.1074/jbc.273.39.25427
- Michaely, P., Tomchick, D. R., Machius, M., and Anderson, R. (2002). G. Crystal structure of a 12 ANK repeat stack from human ankyrinR. *EMBO J.* 21, 6387–6396. doi: 10.1093/emboj/cdf651
- Miller, S., Janin, J., Lesk, A. M., and Chothia, C. (1987). Interior and Surface of Monomeric Proteins. *J. Mol. Biol.* 196, 641–656. doi: 10.1016/0022-2836(87)90038-6

- Mizuguchi, K., and Blundell, T. (2000). Analysis of conservation and substitutions of secondary structure elements within protein superfamilies. *Bioinformatics* 16, 1111–1119. doi: 10.1093/bioinformatics/16.12.1111
- Moorthy, A. K., Huxford, T., and Ghosh, G. (2006). “Structural aspects of NF- $\kappa$ B and IkB proteins,” in *Handbook of Transcription Factor NF- $\kappa$ B*, ed S. Ghosh (Boca Raton, FL: CRC Press), 9–24.
- Mosavi, L. K., Cammett, T. J., Desrosiers, D. C., and Peng, Z.-Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* 13, 1435–1448. doi: 10.1110/ps.03554604
- Mosavi, L. K., Minor, D. L., and Peng, Z.-Y. (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16029–16034. doi: 10.1073/pnas.252537899
- Neumann, M., and Naumann, M. (2007). Beyond I  $\kappa$ Bs: alternative regulation of NF- $\kappa$ B activity. *FASEB J.* 21, 2642–2654. doi: 10.1096/fj.06-7615rev
- Oeckinghaus, A., and Ghosh, S. (2009). The NF- $\kappa$ B family of transcription factors and its regulation. *Cold Spring Harb. Perspect. Biol.* 1:a000034. doi: 10.1101/cshperspect.a000034
- Palopoli, L., Rombo, S. E., Terracina, G., Tradigo, G., and Veltre, P. (2009). Improving protein secondary structure predictions by prediction fusion. *Inf. Fus.* 10, 217–232. doi: 10.1016/j.inffus.2008.11.004
- Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M., and Lundegaard, C. (2009). A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct. Biol.* 9:51. doi: 10.1186/1472-6807-9-51
- Pollastri, G., and McLysaght, A. (2005). Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* 21, 1719–1720. doi: 10.1093/bioinformatics/bti203
- Pollastri, G., Przybylski, D., Rost, B., and Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* 47, 228–235. doi: 10.1002/prot.10082
- Pons, J., Evrard-Todeschi, N., Bertho, G., Gharbi-Benarous, J., Sonois, V., Benarous, R., et al. (2007). Structural studies on 24P-IkB $\alpha$  peptide derived from a human IkB $\alpha$  protein related to the inhibition of the activity of the transcription factor NF- $\kappa$ B. *Biochemistry* 46, 2958–2972. doi: 10.1021/bi061401f
- Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854. doi: 10.1093/bioinformatics/btt055
- Rost, B. (2001). Review: protein secondary structure prediction continues to rise. *J. Struct. Biol.* 134, 204–218. doi: 10.1006/jsb.2001.4336
- Rost, B., and Sander, C. (1994). Conservation and prediction of solvent accessibility in protein families. *Proteins* 20, 216–226. doi: 10.1002/prot.340200303
- Rost, B., Yachdav, G., and Liu, J. (2004). The predictprotein server. *Nucleic Acids Res.* 32(Web Server issue), W321–W326. doi: 10.1093/nar/gkh377
- Schreiber, S., Rosenstiel, P., Albrecht, M., Hampe, J., and Krawczak, M. (2005). Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat. Rev. Genet.* 6, 376–388. doi: 10.1038/nrg1607
- Schuster, M., Annemann, M., Plaza-Sirvent, C., and Schmitz, I. (2013). Atypical IkB proteins - nuclear modulators of NF- $\kappa$ B signaling. *Cell Commun. Signal.* 11:23. doi: 10.1186/1478-811X-11-23
- Scott, W. R. P., Hunenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennem, J., et al. (1999). The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* 103, 3596–3607. doi: 10.1021/jp984217f
- Sillitoe, I., Cuff, A. L., Dessimailly, B. H., Dawson, N. L., Furnham, N., Lee, D., et al. (2013). New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Res.* 41(Database issue), D490–D498. doi: 10.1093/nar/gks1211
- Simons, K. T., Kooperberg, C., Huang, E., and Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* 268, 209–225. doi: 10.1006/jmbi.1997.0959
- Sue, S. C., Cervantes, C., Komives, E. A., and Dyson, H. (2008). J. Transfer of flexibility between ankyrin repeats in I  $\kappa$ B $\alpha$  upon formation of the NF- $\kappa$ B complex. *J. Mol. Biol.* 380, 917–931. doi: 10.1016/j.jmb.2008.05.048
- Sun, S. C., Elwood, J., and Greene, W. C. (1996). Both amino- and carboxyl-terminal sequences within I  $\kappa$ B $\alpha$  regulate its inducible degradation. *Mol. Cell. Biol.* 16, 1058–1065.
- Suzuki, H., Chiba, T., Kobayashi, M., Takeuchi, M., Furuichi, K., and Tanaka, K. (1999). *In vivo* and *in vitro* recruitment of an IkB $\alpha$ -ubiquitin ligase to IkB $\alpha$  phosphorylated by IKK, leading to ubiquitination. *Biochem. Biophys. Res. Commun.* 2561, 121–126. doi: 10.1006/bbrc.1999.0296
- Traenckner, E. B., and Baeruerle, P. A. (1995). Appearance of apparently ubiquitin-conjugated IkB $\alpha$  during its phosphorylation-induced degradation in intact cells. *J. Cell Sci. Suppl.* 19, 79–84. doi: 10.1242/jcs.1995.Supplement\_19.11
- Truhlar, S. M., Torpey, J. W., and Komives, E. A. (2006). Regions of IkB $\alpha$  that are critical for its inhibition of NF- $\kappa$ B. DNA interaction fold upon binding to NF- $\kappa$ B. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18951–18956. doi: 10.1073/pnas.0605794103
- van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). C. Gromacs: fast, Flexible, and Free. *J. Comput. Chem.* 26, 1701–1718. doi: 10.1002/jcc.20291
- Viatour, P., Merville, M. P., Bouras, V., and Chariot, A. (2005). Phosphorylation of NF- $\kappa$ B and IkB proteins: implications in cancer and inflammation. *Trends Biochem. Sci.* 30, 43–52. doi: 10.1016/j.tibs.2004.11.009

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Yazdi, Durdag, Naumann and Stein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.