Structural studies on the RNA-recognition motif of NELF E, a cellular negative transcription elongation factor involved in the regulation of HIV transcription

Jampani Nageswara Rao, Liane Neumann, Sabine Wenzel, Kristian Schweimer, Paul Rösch, Birgitta M. Wöhrl

Biochemical Journal Immediate Publication. Published on 9 Aug 2006 as manuscript BJ20060421

Structural Studies on the RNA Recognition Motif of NELF E, a Cellular Negative Transcription Elongation Factor Involved in the Regulation of HIV Transcription

Jampani Nageswara Rao[#], Liane Neumann[#], Sabine Wenzel, Kristian Schweimer*, Paul Rösch, Birgitta M. Wöhrl*

Universität Bayreuth, Lehrstuhl Biopolymere, Universitätsstr. 30, D-95447 Bayreuth, Germany

* Corresponding authors, mailing address B.M. Wöhrl, K. Schweimer: Universität Bayreuth, Lehrstuhl Biopolymere, Universitätsstr. 30, D-95447 Bayreuth, Germany; Phone: +49 921 55-3540; Fax: +49 921 55-3544; E-mail: birgitta.woehrl@uni-bayreuth.de; kristian.schweimer@uni-bayreuth.de

[#] Both authors contributed equally to this work.

Biochemical Journal Immediate Publication. Published on 9 Aug 2006 as manuscript BJ20060421

The elongation of transcription of HIV RNA at the transactivation response element TAR is highly regulated by positive and negative factors. The cellular negative transcription elongation factor NELF was suggested to be involved in transcriptional regulation of human immunodeficiency virus type 1 (HIV-1) by binding to the stem of the viral TAR RNA which is synthesized by cellular RNA polymerase II at the viral long terminal repeat. NELF is a heterotetrameric protein consisting of NELF A, B, C or the splice variant D, and E. Here we determined the solution structure of the RNA recognition motif (RRM) of the RNA binding subunit NELF E and studied its interaction with the viral TAR RNA. Our data show that the separately expressed recombinant NELF E RRM owns α-helical and β-strand elements adopting a βαββαβ fold and is able to bind to TAR RNA. Fluorescence equilibrium titrations with fluorescently labeled double and single stranded oligoribonucleotides representing the TAR RNA stem imply that NELF E RRM binds to the single stranded TAR RNAs with K_D-values in the low μM range.

Short title: Structure of human NELF E RRM

Keywords: HIV termination/antitermination, TAR, NELF E, NMR structure, RNA binding protein

Abbreviations

CDK9, cyclin dependent kinase 9; DRB, 5,6-Dichloro-1β-D-ribofuranosylbenzimidazole; DSIF, DRB sensitivity inducing factor; DTT, dithiothreitol; HIV, human immunodeficiency virus; HSQC, heteronuclear single quantum coherence; IPTG, isopropyl-thiogalactoside; K_D, dissociation constant; LTR, long terminal repeat; NELF, negative elongation factor; NMR, nuclear magnetic resonance; pTEFb, positive transcription elongation factor b; RNP, ribonucleoprotein motif; RRM, RNA recognition motif; Tat, transactivator of transcription; TAR, transactivation response element;

Introduction

After the human immunodeficiency virus type 1 (HIV-1) enters the host cell the viral genomic RNA is reverse transcribed into double stranded DNA which is then integrated into the host genome. Subsequent synthesis of new viral RNA is tightly controlled by a complex interaction of viral and cellular proteins [1]. Transcription starts from the long terminal repeat (LTR) of the integrated proviral DNA by the cellular RNA polymerase II (RNA Pol II). Once the initial sequence of 60 nucleotides of the viral RNA, the so-called transactivator response (TAR) element, is synthesized, a stable RNA stem-loop structure is formed which is present at the 5' end of all viral transcripts [1]. At this step several cellular transcription factors, including the cellular negative transcription elongation factor NELF and the DRB (5,6-Dichloro-1β-D-ribofuranosylbenzimidazole) sensitivity inducing factor (DSIF), bind to the TAR element [2-5]. Elongation of transcription is possible only if the viral transactivator protein Tat recruits the cellular positive transcription elongation factor pTEFb and binds to the bulge region of TAR [3, 5, 6]. Upon phosphorylation of RNA Pol II, NELF, and DSIF by the kinase component CDK9 of pTEFb antitermination occurs, leading to productive elongation of transcription [5]. NELF consists of 4 different subunits, namely NELF A, NELF B, alternatively spliced NELF C or D, and NELF E, also called RD because of its internal repeats of the amino acids arginine (R) and aspartic acid (D) [4, 7]. NELF A exhibits sequence similarities to the viral hepatitis delta antigen HDAg which binds to and activates RNA polymerase II (RNA pol II) [8]. NELF E contains an RNA recognition motif (RRM). Furthermore, NELF E interacts with the NELF B subunit, probably *via* a leucine zipper motif. It was shown recently that NELF E binds to various RNA elements [7-9]. Most importantly, isolated NELF E binds to the HIV TAR element *in vitro* [9]. Its TAR RNA binding activity suggests NELF E to play a role in the control of HIV transcription.

Experimental procedures

Plasmid construct, expression and protein purification

A synthetically produced gene adapted to the *E. coli* codon usage harboring the RRM of NELF E and adjacent regions was cloned *via* the NdeI and BamHI restriction sites into the *E. coli* expression vector pET15b (Novagen). The soluble recombinant protein contained an N-terminal 6xHis tag (Fig. 1).

E. coli strain BL21 (DE3) (Novagen) harboring the recombinant plasmid was grown at 37 °C in LB medium containing ampicillin until an OD_{600} = 1 was reached and then induced by 1 mM IPTG. Cells were harvested 3 - 4 h after induction. For ^{15}N and ^{13}C labeling, cells were pelleted at 15 °C before induction and resuspended in ¼ volume of M9 medium [10, 11] supplemented with $^{15}NH_4Cl$ and 0.2 % ^{13}C glucose as the sole nitrogen and carbon sources, respectively [12]. After 1 h of shaking at 37 °C, induction was performed with 1 mM IPTG and cells were harvested after 3 - 4 h.

In order to purify NELF E RRM, bacterial cell pellets were lysed by sonication in 20 mM $NaPO_4$, pH 7.4, 0.5 M NaCl, 5 mM imidazole, 1 mM DTT. After centrifugation, the supernatant was loaded onto a Ni-ion affinity column (His-trap chelating, GE Healthcare) and eluted by applying an imidazole step gradient. Peak fractions containing NELF E RRM were dialyzed against 10 mM $NaPO_4$, pH 6.9, 250 mM NaCl, 1 mM DTT and further purified *via* a heparin column (GE Healthcare) by an NaCl step gradient in the same buffer with up to 1 M NaCl. The eluted fractions containing NELF E RRM were dialyzed against 10 mM $NaPO_4$, pH 6.9, 100 mM NaCl, 1 mM DTT, concentrated with Vivaspin concentrators (Vivascience, MWCO 5000 Da), aliquoted and then stored after shock freezing at – 80 °C.

in vitro transcription and purification of TAR RNA

Synthesis of TAR RNA was performed by *in vitro* transcription using T7 RNA polymerase, a T7 RNA primer oligodeoxyribonucleotide, and a template oligodeoxyribonucleotide coding for the corresponding 59 nucleotides of the TAR sequence (5' GCCUCUCUCU GGUUAGACCA GAUCUGAGCC UGGGAGCUCU CUGGCUAACU AGGGAAGGC) plus the region complementary to the T7 primer. The TAR RNA was then purified by denaturing polyacrylamide/urea gel electrophoresis as described [13].

NMR spectrometry

All NMR experiments were performed at 298 K on Bruker DRX600, Avance 700, and Avance 800 spectrometers equipped with inverse room temperature or cryogenic cooled $^{1}H/^{13}C/^{15}N$ triple-resonance probes with pulsed-field gradient capabilities. In order to obtain sequential backbone and side chain resonance assignments, standard double and triple resonance NMR experiments were recorded [14, 15] with uniformly ^{15}N- or ^{13}C-^{15}N labeled recombinant NELF E RRM at a concentration of 0.6 mM. Distance restraints for structure calculation were derived from 3D ^{13}C and

^{15}N edited NOESY HSQC [16, 17] experiments with mixing times of 120 ms. Dihedral angle restraints were derived from $^{3}J(H^{N},H^{\alpha})$ scalar coupling constants determined from the intensity ratios of cross and diagonal peaks of the HNHA spectrum [18]. A series of New MEXICO [19] experiments with different mixing times were recorded for characterizing amide proton exchange. $^{1}D(^{1}H^{N}, ^{15}N)$ residual dipolar couplings (RDCs) were determined by the IPAP method [20] using a weakly aligned sample of uniformly ^{15}N labeled NELF E RRM in a mixture of monododecyl-pentaethylenglycol-ether (C12E5), hexanol and water (molar ratio C12E5:hexanol = 0.95, 3% (wt) C12E5 / water) [21]. $\{^{1}H\}^{15}N$ NOE values were determined using the pulse sequence of Dayie and Wagner [22] with a relaxation delay of 6 s including the 3 s saturation period with 120° high power pulses for the saturated subspectrum. Chemical shift changes of NELF E RRM upon binding to TAR RNA were observed in ^{1}H-^{15}N HSQC spectra after gradually adding TAR RNA to a sample containing ^{15}N NELF E RRM.

Normalized chemical shift changes were expressed as the weighted geometric average of $^{1}H^{N}$ and ^{15}N chemical shift changes for each residue: $\Delta d_{norm} = \sqrt{(\Delta d_{1H})^2 + 0.1(\Delta d_{15N})^2}$

Normalized chemical shift changes larger than 0.04 ppm were considered significant [23]. NMR data were processed using in-house written software and analyzed with the program package NMR-View 5.2.2 [24].

Structure Calculation

Distance restraints for structure calculation were derived from ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC spectra. NOESY cross peaks were classified according to their relative intensities and converted to distance restraints with upper limits of 3.0 Å (strong), 4.0 Å (medium), 5.0 Å (weak), and 6.0 Å (very weak). For ambiguous distance restraints the r^{-6} summation over all assigned possibilities defined the upper limit [25].

The raw scalar coupling constants were multiplied with a correction factor of 1.1 to take into account the different relaxation rates of in-phase and anti-phase components [18]. Residues with scalar coupling constants below 6 Hz were restrained to dihedral angles between -80° and -40°, residues showing coupling constants above 8 Hz were restricted to dihedral angles of -160° to -80° [26]. Glycines were omitted, since they were not stereospecifically assigned and the coupling constants are likely to be affected by cross relaxation [18].

Hydrogen bonds were included in the final structure calculation if the acceptor of a slowly exchanging amide proton characterized by a missing signal in a 150 ms New MEXICO experiment [19] could be identified from the results of preceding structure calculations. Thus, a hydrogen bond was assumed if the distance of the carboxyl oxygen and the amide proton was below 2.6 Å, and the angle of the amide proton, the amide nitrogen and the carboxyl oxygen was less than 60° in all accepted structures. For each hydrogen bond the distance between the amide proton and the acceptor

was restrained to less than 2.3 Å, and the distance between the amide nitrogen and the acceptor was restrained to less than 3.3 Å [26]. All prolines were considered to adopt the trans conformation as strong HA(i)-HD(i+1) and HN(i) and HD(i+1) NOEs could be observed [27].

The structure calculations were performed with the program XPLOR 3.8.5.1 using a three-step simulated annealing protocol [28, 29] with floating assignment of prochiral groups [30]. Initial conformational space sampling was carried out for 120 ps with a time step of 3 fs at a temperature of 2000 K, followed by a cooling period of 120 ps down to 1000 K, and 60 ps cooling to 100 K, both with a time step of 2 fs. A modified conformational database potential for backbone and side chain dihedral angles was applied [31, 32]. The proline angles were modified according to Neudecker *et al.* [33]. After simulated annealing, the structures were subjected to 1000 steps of Powell minimization [34], and the final 500 steps were minimized without conformational database potential.

In a first step, 200 structures were calculated (Tab. 1) using 1926 distance, 24 hydrogen bond and 32 dihedral angle restraints. The 40 structures with the lowest total energy were then refined using 55 $^{1}D(^{1}H^{N}, ^{15}N)$ RDCs with a harmonic potential [35]. Dipolar couplings of flexible residues showing a {1H} ^{15}N NOE below 0.65 at 14.1 T were excluded from the calculations. The tensor components of the alignment were optimized with a grid search by varying the axial component D_a and the rhombicity R in steps of 0.5 and 0.1, respectively. The initial values of D_a and R were estimated from the distribution of the $^{1}D(^{1}H^{N}, ^{15}N)$ [36], and a molecular dynamics run was performed for each pair of D_a and R, yielding an axial component of 9.0 Hz, and a rhombicity of 0.4 for the energetically most favorable combination of D_a and R.

The 20 structures showing the lowest values of the target function excluding the database potential were further analyzed with X-PLOR [37], MolMol [38], and PROCHECK 3.5.4 [39, 40]. The structure coordinates were deposited in the protein data bank (pdb) under the accession code 2BZ2.

Fluorescence equilibrium titrations

Measurements were performed using a Fluorolog spectrophotometer (Horiba Jobin Yvon, Munich, Germany) in reaction buffer consisting of 10 mM $NaPO_4$, pH 7.0 and 100 mM NaCl in a volume of 2 ml. The excitation wave length was 495 nm, and the emission intensity was measured at 522 nm or 525 nm with slit widths set at 2 nm for both, excitation and emission. To analyze the double stranded TAR RNA stem, the oligoribonucleotide representing the 5' end of HIV-1 TAR, TAR(1-18) (5' GGUCUCUCUGGUUAGACC) was hybridized in reaction buffer to the complementary 5' FAM-6 labeled oligoribonucleotide (biomers.net) representing the 3' end of the HIV-1 TAR RNA, TAR(42-57) (5'FAM-6-GGCUAACUAGGGAACC). The solution was heated to 90 °C for 3 min, followed by slow cooling to room temperature. For analysis of the single strands, the corresponding 5' FAM-6 labeled oligoribonucleotides were used. 50 nM of double or single stranded substrate was titrated with increasing amounts of NELF E RRM. Values for the dissociation constant, K_D, were determined assuming a two state model, using a quadratic equation for the fitting procedure:

$$F_{obs} = F_L^0 \cdot [L_0] + \Delta F^0 \cdot \frac{(K_D + [E_0] + [L_0]) - \sqrt{(K_D + [E_0] + [L_0])^2 - 4[L_0][E_0]}}{2}$$

F_{obs}, observed fluorescence; F_L^0, starting fluorescence; L_0, concentration of the RNA; E_0, concentration of protein.

Results and Discussion

Purification of NELF E RRM

To analyze structure and function of NELF E RRM, a synthetic gene containing the corresponding DNA coding for an N-terminal 6xHis-tag followed by a thrombin site and the NELF E RRM was constructed with the codon usage adapted to *E. coli* (Fig.1). The regions adjacent to the NELF E RRM coding region had to be introduced since plasmid constructs lacking these additional DNA stretches led to insoluble protein. The DNA sequence of the corresponding gene was deposited into Genbank under the accession number DQ885937. The expressed and purified protein was stable and could be stored in 10 mM $NaPO_4$ buffer, pH 6.9, containing 100 mM NaCl, 1 mM DTT, by shock freezing aliquots at –80 °C. After thawing and addition of 10 % (v/v) D_2O, the samples were used directly for NMR studies.

Structural NMR analysis

To investigate the integrity of the structure of purified NELF E RRM we performed circular dichroism (CD) (data not shown) and one-dimensional ^{1}H NMR spectroscopy. Using standard double and triple resonance NMR techniques with isotopically labeled protein, ^{1}H, ^{13}C, and ^{15}N backbone and side chain resonances were assigned. Complete backbone and nearly complete side chain assignment was obtained for the region Ala^{35}-Arg^{109}. For residues N-formyl-Met^{1}-Arg^{34} as well as Cys^{110}-Ser^{121} several amide resonances could not be assigned because of missing signals due to conformational exchange or proton exchange with the solvent.

The ^{1}H-^{15}N HSQC spectrum of NELF E RRM shows the characteristic dispersion of a protein with an intact tertiary structure. Using standard double and triple resonance NMR techniques it was possible to assign nearly all ^{1}H, ^{13}C, and ^{15}N resonances. While for the sequence region Lys^{38}-Arg^{109} nearly all (> 95%) backbone resonances and more than 85% of the side chain resonances could be assigned, the terminal regions, especially the amino terminus, remain partially unassigned due to strong overlap or missing resonances. The {^{1}H} ^{15}N heteronuclear steady state NOE experiment at 14.1 T (Fig. 2) shows values around 0.6-0.8 for the region Lys^{38}-Arg^{109}, while outside of this region the heteronuclear NOE decreases towards the termini. {^{1}H} ^{15}N steady state NOE values cluster around 0.7, indicating the absence of pronounced motions on the picosecond to nanosecond time scale for nearly all residues in between L39 and R109. Additionally, the assigned residues from the terminal regions show chemical shifts typically found for highly flexible polypeptides. This is characteristic for a compactly folded domain within the region Lys^{38}-Arg^{109} and unstructured termini. Missing assignments of amide resonances can therefore be explained by conformational or solvent exchange. Due to the flexible character of the termini, residues N-formyl-Met^{1}-Arg^{34} and Leu^{114}–Ser^{121} were excluded from further structural determination.

During the iterative structure determination a set of 2037 experimental restraints, consisting of 1926 NOE derived distance restraints, 24 hydrogen bonds (two distance restraints for each hydrogen bond), 32 dihedral restraints and 55 RDCs, could be derived from NMR data (Table 1). The final structure calculation resulted in an ensemble of 20 structures showing no distance restraint violation larger than 0.16 Å, no violation of a dihedral restraint larger than 2.6°, and no violation of an RDC larger than 0.48 Hz. Only small deviations from the idealized covalent bond geometry were obtained (Table 1). The resulting ensemble of 20 structures shows a high coordinate precision of 0.28 Å for the heavy backbone atoms and 0.64 Å for all heavy atoms for residues Gly^{39}-Ala^{108}, corresponding to the structurally defined domain, as well as good stereochemical properties reflected by the fact that 91 % of residues are located in the most favored regions of the Ramachandran plot (Fig 3, overlay).

The solution structure of NELF E RRM exhibits a compact βαββαβ fold with a four stranded antiparallel β sheet (β_1 = Asn^{40}-Tyr^{45}, β_2 = Ile^{64}-Asp70, β_3 = Cys^{75}-Tyr^{80}, β_4 = Gln^{102}-Ile^{107}) that packs against two helices (h_1 = Pro^{51}-Phe^{61}, h_2 = Met^{83}-Leu^{93}) which are oriented approximately perpendicular to each other (interhelix angle = 114.3° ± 1.8°; Fig. 3). Numerous hydrophobic contacts involving residues Leu^{42}, Val^{44}, Leu^{54}, Ala^{57}, Phe^{58}, Val^{78}, Ala^{86}, Val^{90}, Leu^{93}, Val^{98}, Val^{105}, and Ile^{107} stabilize this packing. For example, Phe^{58} from h_1 contacts Val^{78} from the central strand β_3 and Leu^{93} from h_2 (Fig. 3C). These residues are conserved in all other known RRMs and adopt a similar conformation. Residues Thr^{96} and Gln^{97} form an additional short β-strand aligned antiparallel to β_4, thus extending the β-sheet. This is similar to other RRMs where a β-hairpin is found in the sequence region between h_2 and β_4 [41]. The highly conserved aromatic residues Tyr^{43} and Phe^{77} known to be involved in base stacking interactions with RNA in other RRMs are highlighted in Fig. 3D [41].

Identification of the binding interface

To obtain information on the binding of NELF E RRM to the HIV-1 TAR RNA we performed additional NMR experiments. The binding interface could be clearly defined by the NMR titration studies (Fig 4). Amide ($^1H^N$, ^{15}N) chemical shifts are very sensitive to local structural changes. Therefore, observation of chemical shift changes on titration of a binding partner to a ^{15}N labeled protein provides a powerful method for mapping of the binding interface. Addition of TAR RNA results in remarkable chemical shift changes for resonances located in the central β sheet (Asn^{41}, Leu^{42}, Tyr^{45}, Cys^{75}, Phe^{77}) and in the amino terminal region of h_2 (Glu^{84}, Asp^{87}; Fig. 4A). For several residues in particular in strand β_4 (Lys^{104}-Ile^{107}), signals were absent in the ^1H- ^{15}N HSQC spectra, probably due to exchange processes on the intermediate time scale, suggesting these residues to be involved in binding. This is characteristic for affinities in the low μM range. The chemical shift changes for residues located in strands β_1 and β_3 indicate the typical binding of RNA to the RRM by stacking of bases onto the two conserved aromatic residues Tyr^{43} and Phe^{77}. Fig. 4C shows a surface

representation of the protein highlighting the binding interface and the residues that exhibit chemical shift changes upon binding of TAR RNA.

The large chemical shift changes seen in the Met^{113} and Ala^{116} resonances on TAR titration imply a structural change of this region which is highly flexible in free NELF E RRM. Structural changes of the carboxyl terminus are observed, e.g., in hnRNP1 (heteronuclear ribonucleoprotein 1) RRM1 , where the corresponding region is also unstructured in the free state and adopts a 3_{10} helix upon RNA binding [42]. The additional β-strand of the NELF E RRM (Thr^{96}-Gln^{97}) does not exhibit significant chemical shift changes, indicating that this region is not involved in RNA recognition. This is in contrast to the RRM from TcUBP1, where large chemical shift changes of the resonances of the additional β-hairpin indicate interaction with RNA [43]. Unusual chemical shift changes or disappearing signals are found for several residues in the loop preceding h_2 (Glu^{81}-Glu^{84}, Asp^{87}). This region is not involved in RNA binding in complexes with known structures [42, 44-46], and the distance between the RNA contact sites and helix 2 is rather large for the induction of secondary chemical shift changes. We thus analyzed, whether a second NELF E RRM molecule is binding simultaneously to TAR.

Substrate binding

In order to obtain information on the stoichiometry of the complex, more detailed structural and dynamical characterizations, i.e. ^{15}N relaxation measurements and temperature dependent titrations, were performed. However, analysis of the data was hampered by severe line broadening due to the large molecular weight of the complex and only ambiguous data could be obtained (data not shown).

Therefore, fluorescence titrations were performed. It has been shown previously [47], that NELF E binds to the lower region of the TAR RNA stem. However, none of the RRMs analyzed so far exhibits high affinity for double stranded RNA [41]. To determine whether NELF E RRM binds to single or double stranded TAR RNA fluorescently labeled single or double stranded oligoribonuclotides representing the lower stem of TAR were used for fluorescence equilibrium titrations. The experimental data obtained with the single stranded RNA oligomers could be closely described by a two state model, and K_D-values of around 8.2 µM for TAR(1-18) and 2.6 µM for TAR(42-57), respectively, could be derived (Fig. 5A, B). Binding to the double stranded RNA oligomer, however, could not be described by such simple model (Fig. 5C). However, further experiments will be necessary to obtain the precise binding region on the TAR RNA.

In parallel to our structure determination of the RRM, the NMR structure of the RRM of Parp14 which is identical to the NELF E RRM, was determined by the RIKEN Structural Genomics / Proteomics Initiative (pdb accession code 1X5P). The average structures of both ensembles display a backbone rmsd of 1.3 Å showing a high similarity of the three-dimensional fold. The N- and C-terminal extensions of our construct, however, proved to be crucial for RNA-binding.

The presented results thus demonstrate that NELF E RRM by itself is sufficient for binding to TAR RNA. The structures of RRMs generally exhibit a βαββαβ fold. The loops between the secondary structure elements of RRMs are usually disordered and of various lengths. It is assumed that the different specificities and binding affinities of RRMs are, among other factors, modulated by variations in these loop regions [41]. Although full length NELF E can bind to various RNA elements *via* its RRM, the precise binding regions have not yet been determined [9, 47]. Thus, analysis of the NELF E RRM interaction with different single and double stranded RNAs will contribute to understand the function of NELF E in the regulation of HIV transcription in particular and eukaryotic transcription in general, and to identify new drug targets against HIV.

Acknowledgements

We would like to thank Katrin Weiß and Nadine Herz for excellent technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft (Ro617/8-4) and the Sonderforschungsbereich 466.

References

1 Taube, R., Fujinaga, K., Wimmer, J., Barboric, M. and Peterlin, B. M. (1999) Tat transactivation: A model for the regulation of eukaryotic transcriptional elongation. Virology. **264**, 245-253

2 Kao, S. Y., Calman, A. F., Luciw, P. A. and Peterlin, B. M. (1987) Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature. **330**, 489-493

3 Wada, T., Orphanides, G., Hasegawa, J., Kim, D. K., Shima, D., Yamaguchi, Y., Fukuda, A., Hisatake, K., Oh, S., Reinberg, D. and Handa, H. (2000) FACT relieves DSIF/NELF-mediated inhibition of transcriptional elongation and reveals functional differences between P-TEFb and TFIIH. Mol. Cell. **5**, 1067-1072

4 Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J. and Handa, H. (1999) NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell. **97**, 41-51

5 Ping, Y. H. and Rana, T. M. (2001) DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase II and DSIF during transcription elongation. J. Biol. Chem. **276**, 12951-12958

6 Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B. and Price, D. H. (1997) Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. Genes Dev. **11**, 2622-2632

7 Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D. K., Hasegawa, J., Omori, M., Inukai, N., Endoh, M., Yamada, T. and Handa, H. (2003) Human transcription elongation factor NELF: Identification of novel subunits and reconstitution of the functionally active complex. Mol. Cell. Biol. **23**, 1863-1873

8 Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., Wada, T., Sugimoto, S., Konarska, M. M. and Handa, H. (2001) Stimulation of RNA polymerase II elongation by hepatitis delta antigen. Science. **293**, 124-127

9 Yamaguchi, Y., Inukai, N., Narita, T., Wada, T. and Handa, H. (2002) Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol. Cell. Biol. **22**, 2918-2927

10 Meyer, O. and Schlegel, H. G. (1983) Biology of aerobic carbon monoxide-oxidizing bacteria. Annu. Rev. Microbiol. **37**, 277-310

11 Sambrook, J., Maniatis, T. and Fritsch, E. F. (1989) Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

12 Marley, J., Lu, M. and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins. J. Biomol. NMR. **20**, 71-75

13 Metzger, A. U., Schindler, T., Willbold, D., Kraft, M., Steegborn, C., Volkmann, A., Frank, R. W. and Rösch, P. (1996) Structural rearrangements on HIV-1 tat (32-72) TAR complex formation. FEBS Lett. **384**, 255-259

14 Bax, A. and Grzesiek, S. (1993) Methodological advances in protein NMR. Acc. Chem. Res. **26**, 131-138

15 Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J., Sorensen, O. W. and Griesinger, C. (1994) A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. J. Biomol. NMR. **4**, 301-306

16 Ikura, M., Kay, L. E., Tschudin, R. and Bax, A. (1990) Three-dimensional NOESY-HMQC spectroscopy of a ^{13}C-labeled compound. J. Magn. Reson. **86**, 204-209

17 Talluri, S. and Wagner, G. (1996) An optimized 3D NOESY-HSQC. J. Magn. Reson. B. **112**, 200-205

18 Vuister, W. G. and Bax, A. (1993) Quantitative J correlation: A new approach for measuring homonuclear three-bond J(HNHA) coupling constants in ^{15}N-enriched proteins. J. Am. Chem. Soc. **115**, 7772-7777

19 Koide, S., Jahnke, W. and Wright, P. E. (1995) Measurement of intrinsic exchange rates of amide protons in a ^{15}N-labeled peptide. J. Biomol. NMR. **6**, 306-312

20 Ottiger, M., Delaglio, F. and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. J. Magn. Reson. **131**, 373-378

21 Otting, G., Ruckert, M., Levitt, M. H. and Moshref, A. (2000) NMR experiments for the sign determination of homonuclear scalar and residual dipolar couplings. J. Biomol. NMR. **16**, 343-346

22 Dayie, K. T. and Wagner, G. Relaxation-rate measurements of ^{15}N-^{1}H groups with pulsed-field gradients and preservation of coherence pathways. J. Magn. Reson. **111A**, 121-126

23 Hajduk, P. J., Dinges, J., Miknis, G. F., Merlock, M., Middleton, T., Kempf, D. J., Egan, D. A., Walter, K. A., Robins, T. S., Shuker, S. B., Holzman, T. F. and Fesik, S. W. (1997) NMR-based discovery of lead inhibitors that block DNA binding of the human papillomavirus E2 protein. J. Med. Chem. **40**, 3144-3150

24 Johnson, B. A. and Blevins, R. A. (1994) NMRview: A computer program for the visualization and analysis of NMR data. J. Biomol. NMR. **4**, 603-614

25 Nilges, M. (1995) Calculation of protein structures with ambiguous distance restraints. automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. J. Mol. Biol. **245**, 645-660

26 Schweimer, K., Hoffmann, S., Bauer, F., Friedrich, U., Kardinal, C., Feller, S. M., Biesinger, B. and Sticht, H. (2002) Structural investigation of the binding of a herpesviral protein to the SH3 domain of tyrosine kinase lck. Biochemistry. **41**, 5120-5130

27 Wüthrich, K. (1986) NMR of proteins and nucleic acids., Wiley, New York

28 Nilges, M., Gronenborn, A. M., Brunger, A. T. and Clore, G. M. (1988) Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. Protein Eng. **2**, 27-38

29 Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. circumventing problems associated with folding. FEBS Lett. **239**, 129-136

30 Folmer, R. H., Hilbers, C. W., Konings, R. N. and Nilges, M. (1997) Floating stereospecific assignment revisited: Application to an 18 kDa protein and comparison with J-coupling data. J. Biomol. NMR. **9**, 245-258

31 Kuszewski, J. and Clore, G. M. (2000) Sources of and solutions to problems in the refinement of protein NMR structures against torsion angle potentials of mean force. J. Magn. Reson. **146**, 249-254

32 Neudecker, P., Sticht, H. and Rösch, P. (2001) Improving the efficiency of the gaussian conformational database potential for the refinement of protein and nucleic acid structures. J. Biomol. NMR. **21**, 373-375

33 Neudecker, P., Nerkamp, J., Eisenmann, A., Nourse, A., Lauber, T., Schweimer, K., Lehmann, K., Schwarzinger, S., Ferreira, F. and Rösch, P. (2004) Solution structure, dynamics, and hydrodynamics of the calcium-bound cross-reactive birch pollen allergen bet v 4 reveal a canonical monomeric two EF-hand assembly with a regulatory function. J. Mol. Biol. **336**, 1141-1157

34 Powell, M. J. D. (1977) Restart procedures for conjugate gradient method. Mathem. Progr. **12**, 241-254

35 Tjandra, N., Grzesiek, S. and Bax, A. (1996) Magnetic field dependence of nitrogen-proton J splittings in ^{15}N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. J. Am. Chem. Soc. **118**, 6264-6272

36 Clore, G. M., Gronenborn, A. M. and Bax, A. (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. J. Magn. Reson. **133**, 216-221

37 Bruenger, A. T. (1993) *X-PLOR version 3.1.* , Howard Hughes Medical Institute & Yale University, New Haven, CT, USA

38 Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures. J. Mol. Graph. **14**, 51-5, 29-32

39 Morris, A. L., MacArthur, M. W., Hutchinson, E. G. and Thornton, J. M. (1992) Stereochemical quality of protein structure coordinates. Proteins. **12**, 345-364

40 Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR. **8**, 477-486

41 Maris, C., Dominguez, C. and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J. **272**, 2118-2131

42 Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D. and Varani, G. (1996) Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. Nature. **380**, 646-650

43 Volpon, L., D'Orso, I., Young, C. R., Frasch, A. C. and Gehring, K. (2005) NMR structural study of TcUBP1, a single RRM domain protein from trypanosoma cruzi: Contribution of a beta hairpin to RNA binding. Biochemistry. **44**, 3708-3717

44 Deo, R. C., Bonanno, J. B., Sonenberg, N. and Burley, S. K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell. **98**, 835-845

45 Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein. Nature. **398**, 579-585

46 Wang, X. and Tanaka Hall, T. M. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein. Nat. Struct. Biol. **8**, 141-145

47 Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T. and Peterlin, B. M. (2004) Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Mol. Cell. Biol. **24**, 787-795

Figure Legends

Figure 1: Nucleotide and amino acid sequence of NELF E RRM. The construct contains a 6xHis-tag at the N-terminus, followed by a thrombin cleavage site indicated by the gray box. The amino acids defining the RRM are underlined, the amino acid sequence derived from NELF E is written in bold letters. The conserved motifs RNP2 and RNP1 containing residues $Tyr^{43}$and Phe^{77} are marked by black boxes. Amino acid numbering is indicated on the right.

Figure 2: Size of heteronuclear $\{^1H\}$ ^{15}N steady state NOE at 14.1 T along the amino acid sequence. The values in the range of 0.6 to 0.85 for residues Leu^{40} – Cys^{110} indicate a rigid protein backbone in this region.

Figure 3: Solution structure of NELF E RRM. (A) Overlay of the 20 structures (residues Ala^{35}-Met^{113}) showing the lowest values of the target function excluding the database potential. **(B)** Ribbon presentation of the NELF E RRM structure. The figure was generated using MOLMOL [38]. **(C)** Packing of Phe^{58}, Val^{78} and Leu^{93}. The amino acid residues are displayed as space-filled atoms. **(D)** The highly conserved aromatic residues Tyr^{43} and Phe^{77} derived from RNP2 and RNP1 of NELF E RRM are highlighted by balls and sticks.
The N- and C-termini are indicated.

Figure 4: Determination of the binding interface. (A) Overlay of the ^1H-^{15}N HSQC spectra of free NELF E RRM (black) and the NELF E:TAR RNA complex (red). The amino acid resonances showing significant chemical shift changes upon addition of an equimolar amount of TAR RNA are indicated by arrows. **(B)** Normalized chemical shift changes of NELF E RRM upon TAR RNA binding. The normalized chemical shift changes (weighted geometric average of $^1H^N$ and ^{15}N chemical shift changes) are shown as a function of the primary sequence. Changes larger than 0.04 (dotted line) were considered significant. Resonances disappearing after TAR binding due to extreme line broadening by chemical exchange are marked with an x. **(C)** Surface representation of NELF E RRM highlighting the binding interface. Residues with resonances showing significant chemical shift changes upon TAR binding ($0.04 < \Delta\delta = 0.1$), including the highly conserved Tyr^{43} and $Phe^{77}$of the RNP2 and RNP1sequence motifs, are shown in yellow, those with resonances showing chemical shift changes with $\Delta\delta > 0.1$ are indicated in light red, and residues whose resonances could not be detected due to extreme line broadening are presented in light green.

Figure 5: Fluorescence equilibrium titration with the TAR RNA stem. 50 nM of FAM-6 labeled **(A)** TAR(1-18), **(B)** TAR(42-57) or **(C)** double stranded TAR RNA stem was titrated with NELF E RRM in 10 mM $NaPO_4$ pH 7.0, 100 mM NaCl. The curves in (A) and (B) show the best fit to a

quadratic equation describing the binding equilibria with a K_D-values of 8.2 ± 1.0 μM and 2.6 ± 0.2 μM, respectively. No curve could be fitted to the data in (C) using a simple two state model.

Figures

Figure 1

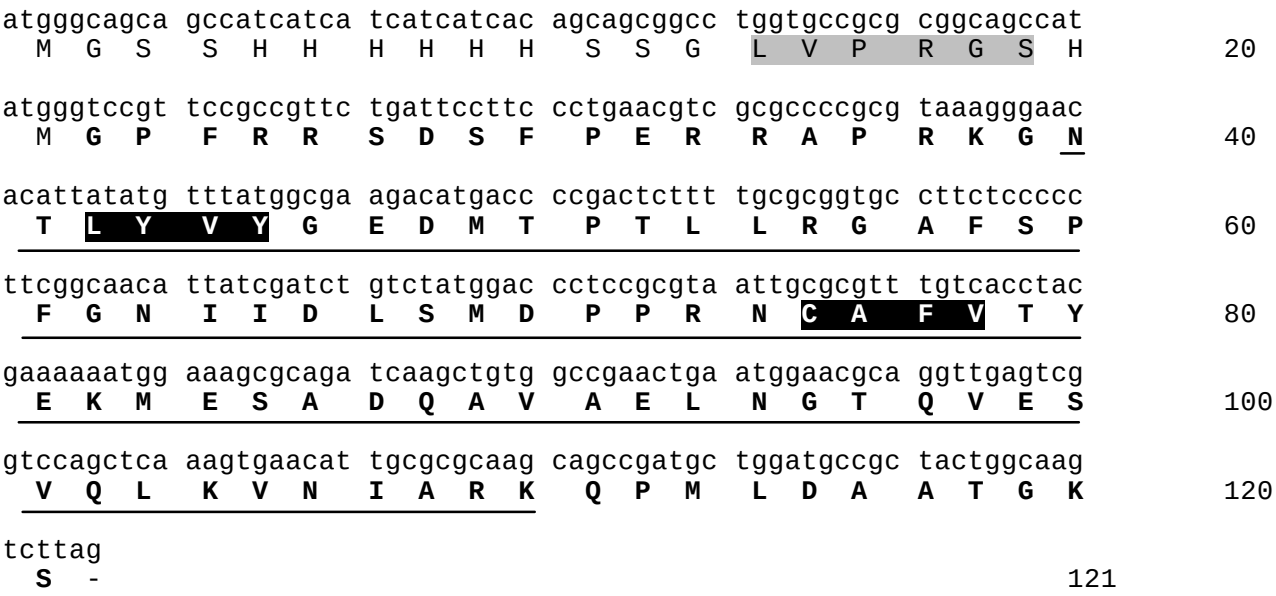


Figure 2

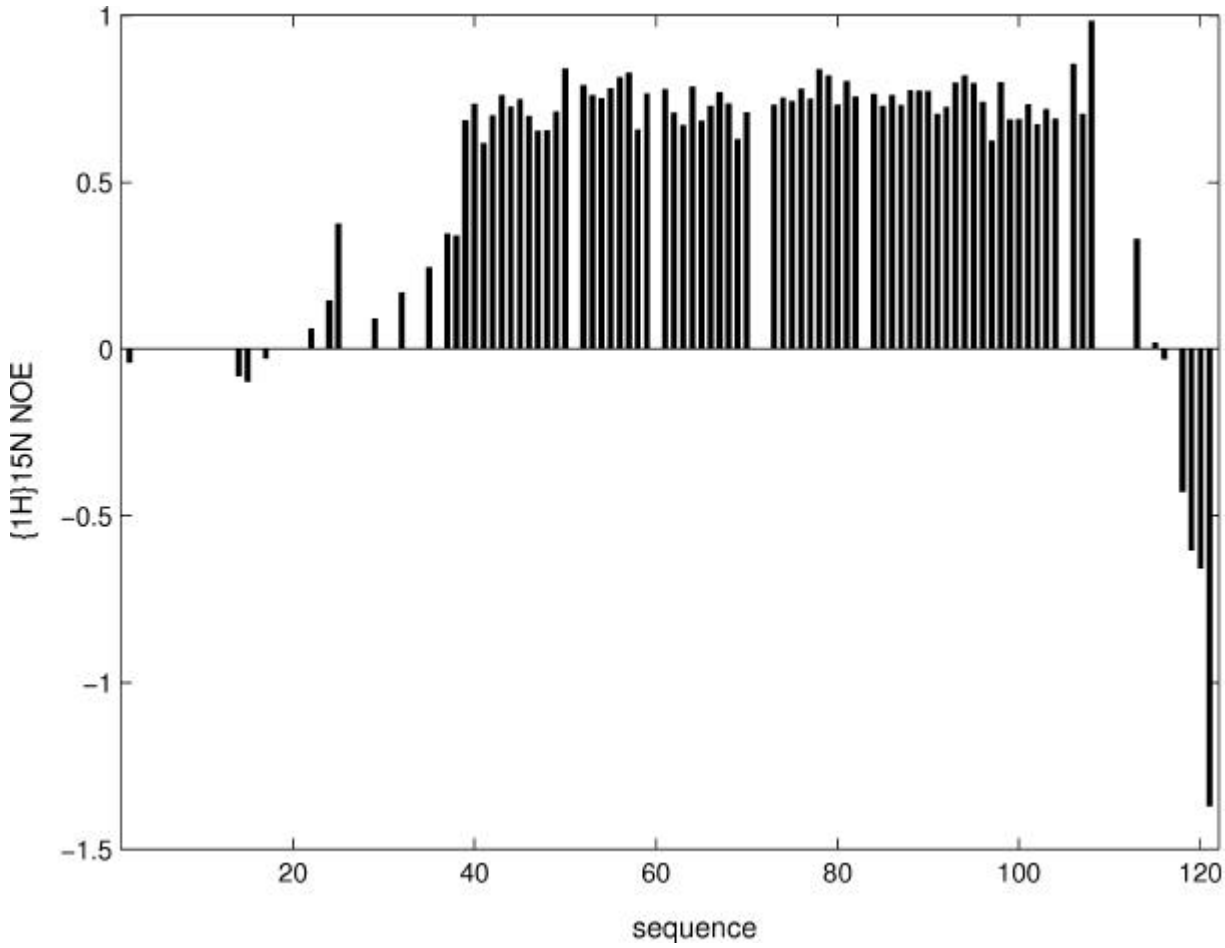


Figure 3

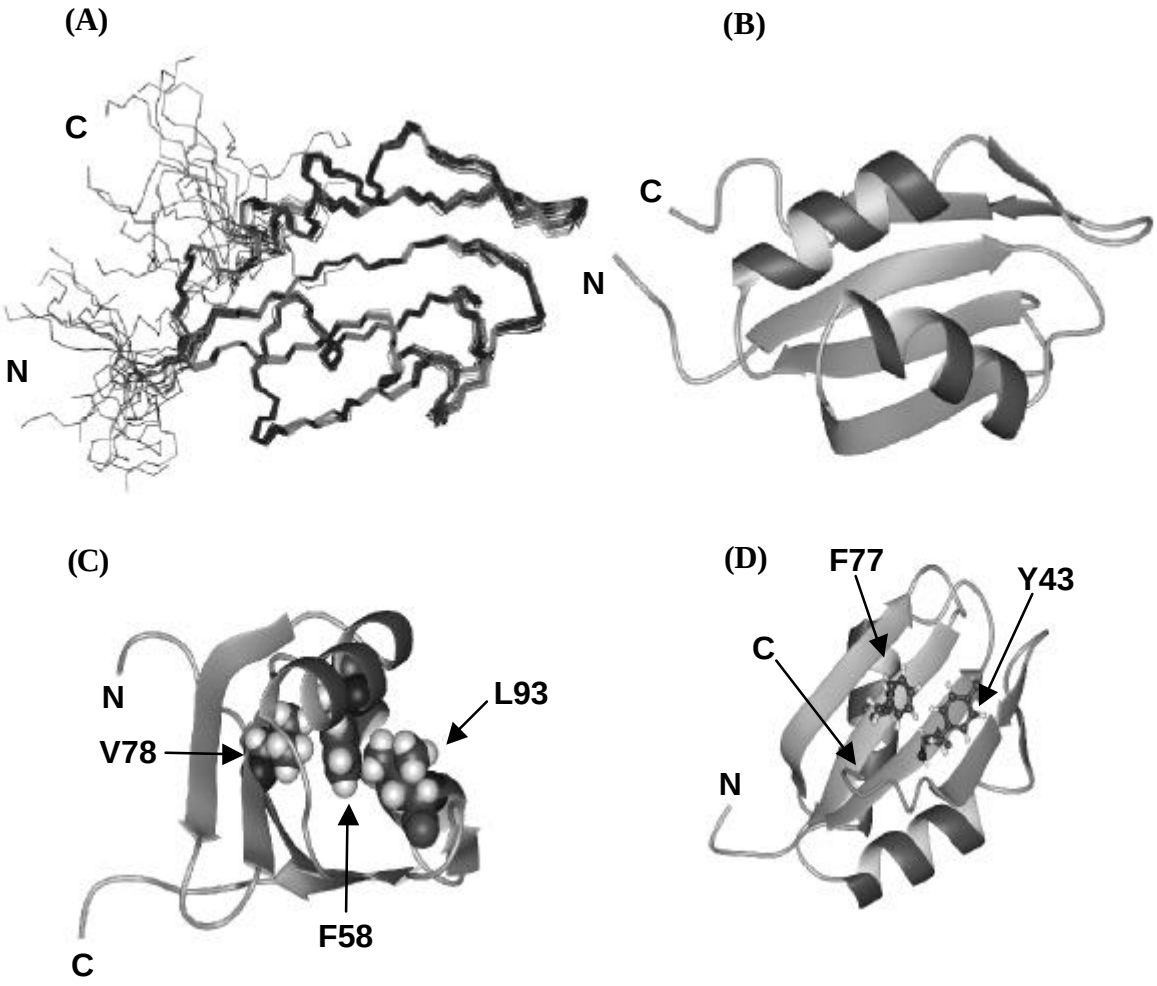


Figure 4(A)

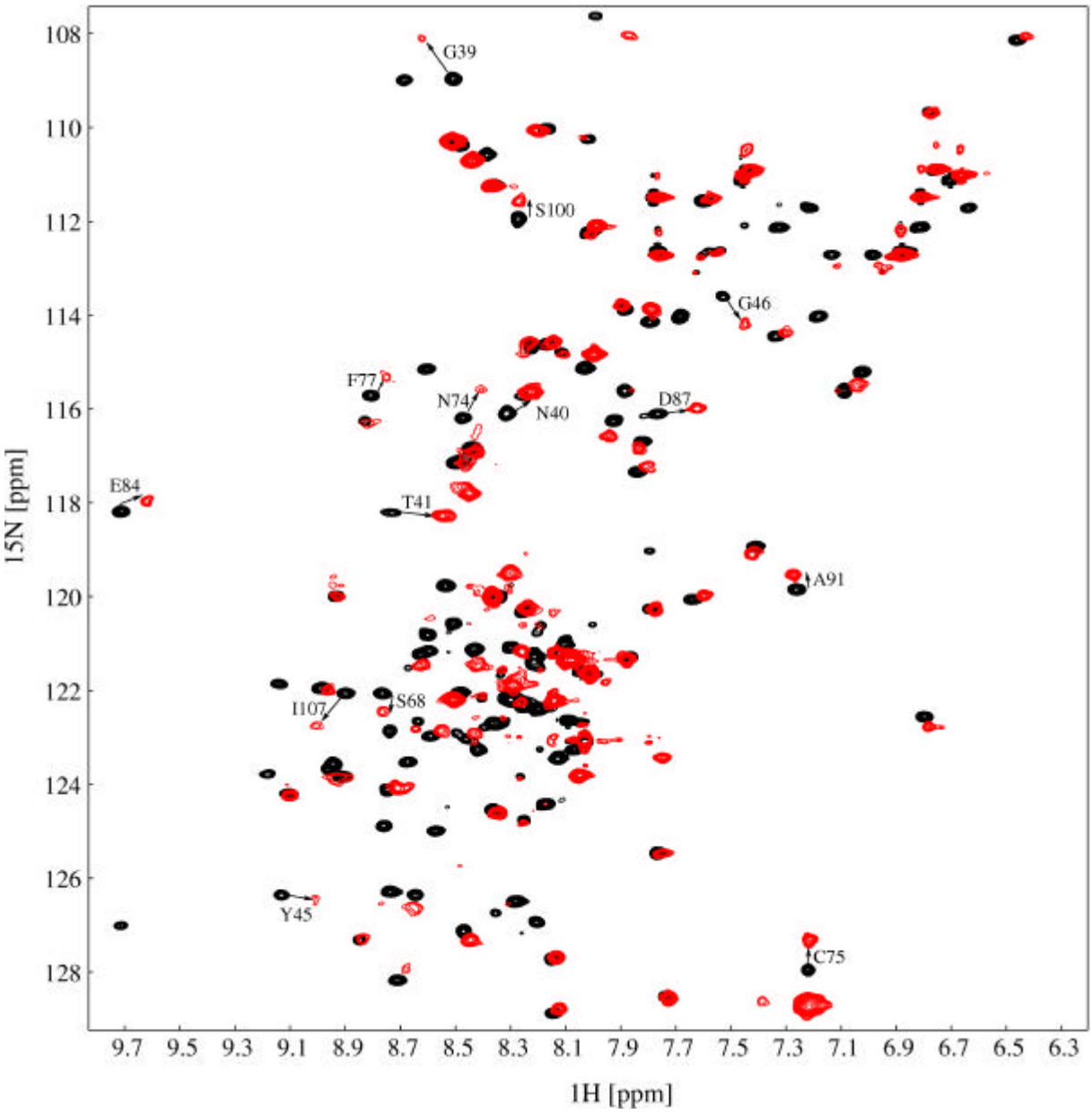
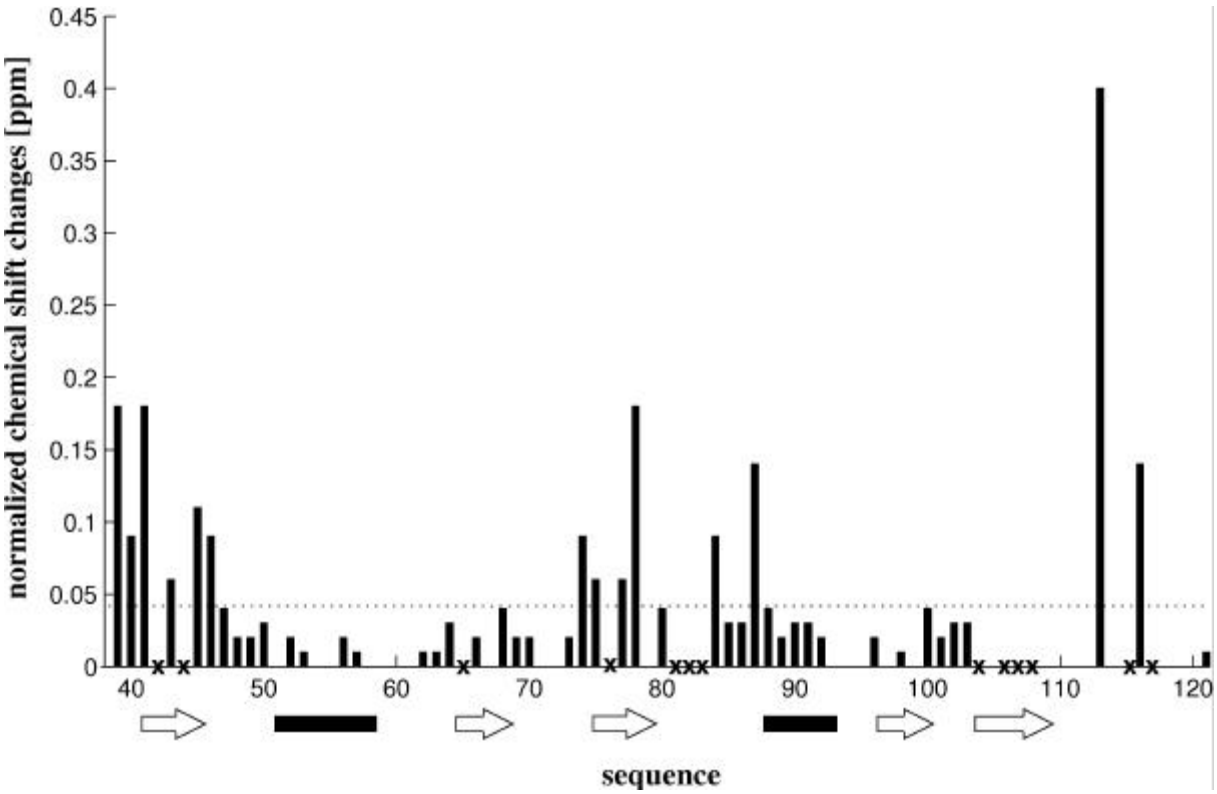


Figure 4(B)



Biochemical Journal Immediate Publication. Published on 9 Aug 2006 as manuscript BJ20060421

Figure 4(C)

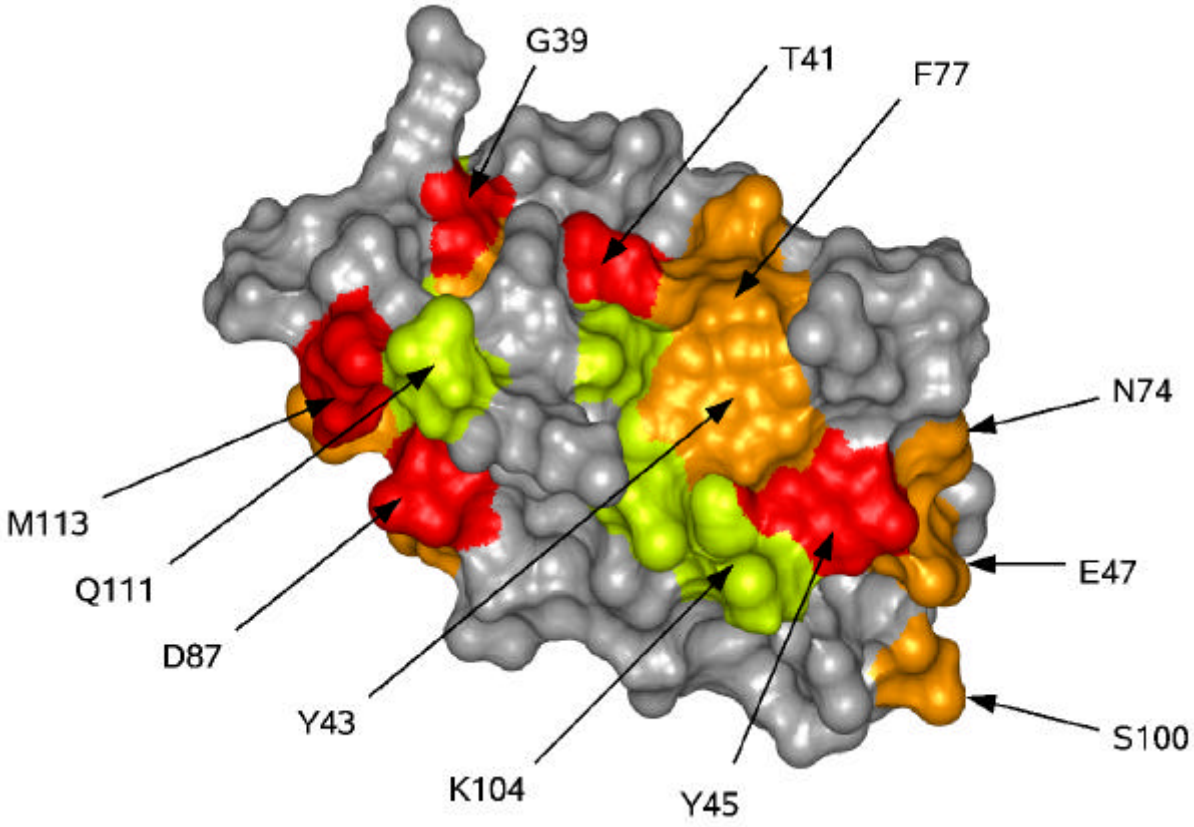


Figure 5

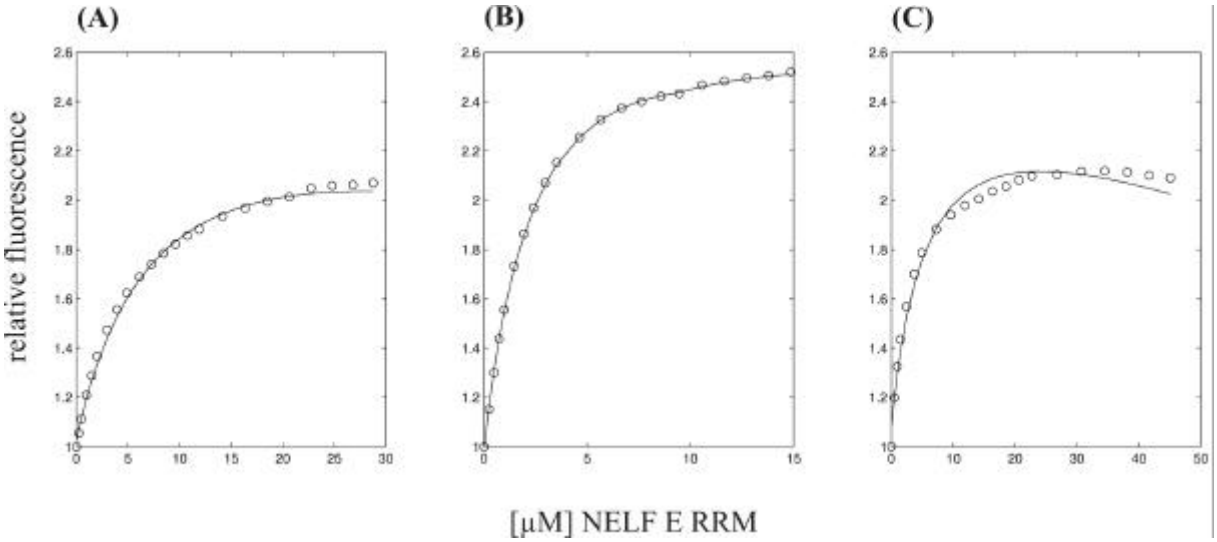


Table 1. Structural statistics

experimental restraints		
distance restraints	total	1926
	intraresidual sequential medium-range long-range	468 406 281 623
dihedral angles		32
dipolar couplings		55
hydrogen bonds (two restraints each)		24
molecular dynamics statistics		
energies (kcal/mol)		
	E_{pot}	14.1 ± 1.4
	E_{bond}	0.58 ± 0.06
	E_{angle}	6.7 ± 0.6
	E_{impr}	2.3 ± 0.2
	E_{repel}	2.4 ± 0.3
	E_{NOE}	1.3 ± 0.5
	E_{cdih}	0.03 ± 0.03
	E_{sani}	0.7 ± 0.2
RMSDs from ideal distances (Å)	bond lengths distance restraints	0.00068 ± 0.00004 0.0036 ± 0.0007
RMSDs from ideal angles(deg)	bond angles dihedral angle restraints	0.14 ± 0.06 0.15 ± 0.15
RMSDs from dipolar couplings (Hz)		0.11 ± 0.02
atomic coordinate precision (RMSD) (Å)		
backbone heavy atoms		0.28 (Gly39-Ala108)
heavy atoms		0.64 (Gly39-Ala108)
Ramachandran plot statistics		
residues in		
	most favored regions	91.0%
	allowed regions	9.0%