Supporting Information

Insights into the NF-kappaB-DNA interaction through NMR spectroscopy

Tahseen Raza¹, Nitin Dhaka¹, David Joseph^{2#}, Prikshat Dadhwal^{1##}, Veera Mohana Rao Kakita³, Hanudatta S. Atreya^{2§} and Sulakshana P. Mukherjee^{1*}

- 1. Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand 247667, India
- 2. NMR Research Centre, Indian Institute of Science Bengaluru, Karnataka 560012, India
- UM-DAE-Centre for Excellence in Basic Sciences, University of Mumbai, Kalina Campus, Mumbai, Maharashtra 400098, India
 *sulakshana mukhorioa@ht iitr ac in

<u>*sulakshana.mukherjee@bt.iitr.ac.in</u>

		A				L	41			В	_	C			H	I	C		E	
						-										7				
p50	39 GI	PYLQII	LEQPK	QRGF	RFRY	/CEGP	SHGGL	PGASSI	EKNKKS	YPQVKIC	NYVGPAR	VIVQ	LVTNGKI	IIHLHA	ISLVGK	HC-	EDGVCT	VTAGPKI	MV-VGFA13	85
p52	37 G	PYLVIV	/EQPK(QRGF	RFRY	SCEGP	SHGGL	PGASSI	EKGRKT	YPTVKIC	NYEGPAR	IEVD	LVTHSDI	PRAHA	ISLVGK	QCS	ELGVCA	VSVGPKI	MT-AQFN 13	34
c-Rel	8-	PYVEII	EQPRO	DRGM	RFRY	CEGR	SAGSI	PGERS	TDNNRT	YPSVQIN	INYYGKG	IRIT	LVTKNDI	YKPHP	IDLVGK	DC-	RDPYYE	AEFGPEF	RP-LFFQ10	2
RelA	102 P	PIVELL	FORK	DRGM	DEDAL	CEGR	CACCT	TCESS	TEASET	OPATEL		VELS	LVIKDPI		ISLUCK		TDCVCP		SI-HSPUII	11
KEID	102 1	E	T	ZKGM	a	SCEGR	DAGOI	LGESS	EASKI	VENIEL	UCOO LINE	C		nevner	15LVGK	E.	IDGVCR	C	J	1
			L ₂		u1		_					0	~2			F			\mathbf{L}_3	
	100					-	TROUBL													
p50	136 N	LGILHV	TKKK	MET	LEAR	TEAC	INGIN	PGLLVI	HSDLAI.	LQAEGGG		EREIIR	OFAVOU	KEMDL	TUDID	FTA	FLPDST	COPOL DI	EPVVSDAZ3	6
p52 c-Rel	104 N	LGTRCV	TERM	JKGA	TTLR	SAC-	INDEN	VCF			<mark></mark> E4		-OOLUDI	FDCDL	JUVBCV	FMF	FLPDED	CNETTAL	PPTVSNP17	14
RelA	115 N	LGIOCY	KKRDI	LEOA	ISOR	OTN-	NNPFH	VPI					-EEOR	GDYDL	AVRLC	FOV	TVRDPA	G-RPLLI	TPVLSHP 18	32
RelB	202 N	LGIOCV	RKKE	IEAA	IERK	QLG-	IDPYN	AG					SLKNH	QEVDM	WVRIC	FQA	SYRD00	G-HLHRN	DPILSEP 26	59
				\$		я,		b	Т	c		c'	T	_	e			f		
					.	a'		. b	L ₄	с		¢'	L	5	e			f	• ••••	
p50	237 1	YDSKAP	PNASNI	LKI-	VRMDE	a'	VTGGE	b EIYLL			TYEEEENG	C'	GDFSPT	5 VHROF	e IVFKT	PKY	KDVNIT	f KPASVE	OLRRKSD 33	86
p50 p52	237 I 217 II	YDSKAF	PNASNI	LKI- LKI-	VRMDF		VTGGE	b EIYLLO			TYEEEEN	C' GVWEGF	GDFSPT GDFSPT	5 OVHRQF7 OVHKQY7	e AIVFKT	PKY	KDVNIT HKMKIE	f KPASVFV RPVTVFI	QLRRKSD 33	36
p50 p52 c-Rel	237 I 217 I 175 I	YDSKAF HDSKSF YDNRAF	PNASNI PGASNI PNTAEI	LKI- LKI- LRIL	VRMDF	ATAGC TAGS	VTGGE VRGGD VRGGD	b EEIYLLO DEVYLLO DEIFLLO		C DDIQIRI DDIEVRI DDIEVRI	TYEEEENG TYEDDENG	C' GVWEGF WQAF DWEAR	GDFSPTI GDFSPTI GVFSQAL	OVHRQF7 OVHKQY7 OVHKQY7	e AIVFKT AIVFRT	PKY PPY PPY	KDVNIT HKMKIE CK-AIL	f KPASVFV RPVTVFI EPVTVKN	QLRRKSD 33 QLKRKRG 31 QLRRPSD 26	4
p50 p52 c-Rel RelA	237 I 217 I 175 I 183 I	YDSKAF HDSKSF YDNRAF FDNRAF	PNASNI PGASNI PNTAEI PNTAEI	LKI- LKI- LRIL LKI-	VRMDF SRMDF ARVNF	ATAGC TAGS CNCGS RNSGS	VTGGE VRGGD VRGGD CLGGD	b EIYLLA DEIYLLA DEIFLLA	L4 CDKVQK CDKVQK CDKVQK CDKVQK	DDIQIRI DDIQIRI DDIEVRI DDIEVRI EDIEVRI	TYEEEENG TYEDDENG TYENT	GVWEGF	GDFSPTI GDFSPTI GVFSQAI GSFSQAI	5 DVHRQFI DVHRQVI DVHRQVI DVHRQVI	e AIVFKT AIVFRT AIVFRT	PKY PPY PPY	KDVNIT HKMKIE CK-AIL ADPSLQ	f RPASVFV RPVTVFI EPVTVKN APVRVSN	QLRRKSD 33 QLKRKRG 31 QLRRPSD 26 QLRRPSD 27	4
p50 p52 c-Rel RelA RelB	237 I 217 I 175 I 183 I 270 V	YDSKAF HDSKSF YDNRAF FDNRAF YDKKST	PNASNI PGASNI PNTAEI PNTAEI PNTAEI	LKI- LKI- LRIL LKI- LRI-	VRMDF SRMDF ARVNF CRVNF CRINF	RTAGC TAGS KNCGS KNSGS KESGP	VTGGE VRGGD VRGGD CLGGD CTGGE	b EIYLLA DEVYLLA DEIFLLA DEIFLLA	L4 CDKVQKI CDKVQKI CDKVQKI CDKVQKI	C DDIQIRI DDIEVRI DDIEVRI EDIEVVI EDIEVVI	TYEEEENO TYEDDENO TGP	C' GVWEGF WQAF DWEAR GWEAR	GDFSPTI GDFSPTI GVFSQAI GSFSQAI ADFSQAI	5 DVHRQF7 DVHRQY7 DVHRQV7 DVHRQV7 DVHRQV7	e AIVFKT AIVFRT AIVFRT AIVFRT	PKY PPY PPY PPY	KDVNIT HKMKIE CK-AIL ADPSLQ EDLEIS	f KPASVFV RPVTVFI EPVTVKP APVRVSN EPVTVN	QLRRKSD 33 QLKRKRG 31 QLRRPSD 26 QLRRPSD 27 FLQRLTD 36	36 4 59 77 54
p50 p52 c-Rel RelA RelB	237 I 217 I 175 I 183 I 270 V	YDSKAF HDSKSF YDNRAF FDNRAF YDKKST	PNASNI PGASNI PNTAEI PNTAEI INTSEI g	LKI - LKI - LRIL LKI - LRI -	VRMDF SRMDF ARVNF CRVNF	A RTAGC TAGS KNCGS RNSGS KESGP	VTGGE VRGGD VRGGD CLGGD CTGGE	b EEIYLLA DEIFLLA DEIFLLA CELYLLA	L4 CDKVQKI CDKVQKI CDKVQKI CDKVQKI CDKVQKI	DDIQIRI DDIEVRI DDIEVRI EDIEVRI EDIEVVI EDISVVI	TYEEEEN TYEDDENG TGP	C' GVWEGF WQAF DWEAR GWEAR SWEGR	GDFSPTI GDFSPTI GVFSQAI GSFSQAI ADFSQAI	5 OVHRQF7 OVHRQY7 OVHRQV7 OVHRQV7 OVHRQ17	e AIVFKT AIVFRT AIVFRT AIVFRT	PKY PPY PPY PPY	KDVNIT HKMKIE CK-AIL ADPSLQ EDLEIS	f RPASVFV RPVTVFI EPVTVKN APVRVSN EPVTVNV	QLRRKSD 33 QLKRKRG 31 QLRRPSD 26 QLRRPSD 27 FLQRLTD 36	36 4 59 77 54
p50 p52 c-Rel RelA RelB	237 I 217 II 175 I 183 II 270 V	YDSKAF HDSKSF YDNRAF FDNRAF YDKKST	PNASNI PGASNI PNTAEI PNTAEI PNTAEI S	LKI- LKI- LRIL LKI- LRI-	VRMDF SRMDF ARVNF CRVNF CRINF	A RTAGC TAGS CNCGS RNSGS RNSGS RNSGS	VTGGE VRGGD VRGGD CLGGD CTGGE	b DEVYLLA DEVYLLA DEIFLLA DEIFLLA		DDIQIRI DDIQIRI DDIEVRI DDIEVRI EDIEVVI EDISVVI	TYEEEENO TYEDDENO TGP	C' GVWEGF WQAF DWEAR GWEAR SWEGR	GDFSPT GDFSPT GVFSQAI GSFSQAI ADFSQAI	5 OVHRQF OVHKQY OVHRQV OVHRQV OVHRQI	e AIVFKT AIVFRT AIVFRT AIVFRT	PKY PPY PPY PPY	KDVNIT HKMKIE CK-AIL ADPSLQ EDLEIS	f RPASVFV RPVTVFI EPVTVK9 APVRVS9 EPVTVNV	QLRRKSD 33 QLKRKRG 31 QLRPSD 26 QLRPSD 27 FLQRLTD 36	36 4 59 77 54
p50 p52 c-Rel RelA RelB	237 II 217 II 175 II 270 V	YDSKAF HDSKSF YDNRAF FDNRAF YDKKST	PNASNI PGASNI PNTAEI PNTAEI SNTSEI g	LKI- LKI- LRIL LKI- LRI-	VRMDF SRMDF ARVNF CRVNF CRINF	ATAGC TAGS TAGS KNCGS RNSGS KESGP	VTGGE VRGGD VRGGD CLGGD CTGGE ROK	b EIYLLA DEIFLLA DEIFLLA ELYLLA	L4 CDKVQKI CDKVQKI CDKVQKI CDKVQKI CDKVQKI	DDIQIRI DDIEVRI DDIEVRI EDIEVRI EDIEVRI EDIEVRI EDISVVI	TYEEEENC TYEDDENC TYENT TGP	GVWEGF WQAF -DWEAR -GWEAR	GDFSPT GDFSPT GVFSQAI GSFSQAI ADFSQAI	5 DVHRQFJ DVHRQVJ DVHRQVJ DVHRQVJ DVHRQIJ	e AIVFKT AIVFKT AIVFKT AIVFKT	PPY PPY PPY PPY	KDVNIT HKMKIE CK-AIL ADPSLQ EDLEIS	f RPVTVFI EPVTVK APVRVSI EPVTVN	QLRRKSD 33 QLKRKRG 33 QLRPSD 26 QLRPSD 27 FLQRLTD 36	36 4 59 77 54
p50 p52 c-Rel RelA RelB p50 p52	237 II 217 II 175 II 183 II 270 V 337 II 315 G	YDSKAF HDSKSF YDNRAF FDNRAF YDKKST ETSEPF DVSDSF	PNASNI PGASNI PNTAEI PNTAEI STAEI RNTSEI g	LKI - LKI - LRI L LRI - LRI -	VRMDF SRMDF ARVNF CRVNF CRINF CRINF	ATAGC TAGS CTAGS CALC CALC CALC CALC CALC CALC CALC CAL	VTGGE VRGGD CLGGD CTGGE RQK RRK	b EIYLLA DEIFLLA DEIFLLA ZELYLLA 363 341	L4 CDKVQKI CDKVQKI CDKVQKI CDKVQKI CDKVQKI	C DDIQIRI DDIEVRI DDIEVRI EDIEVVI EDISVVI	YEEENO YEDDENO VLN TGP STA	C' GVWEGF WQAF DWEAR GWEAR SWEGR	GDFSPTI GDFSPTI GVFSQAI GSFSQAI ADFSQAI	s VHRQFJ VHRQYJ VHRQVJ VHRQUJ	e AIVFKT AIVFRT AIVFRT AIVFRT	PKY PPY PPY PPY	KDVNIT HKMKIE CK-AIL ADPSLQ EDLEIS	f RPVTVFI EPVTVKA APVRVSI EPVTVN	QLRRKSD 33 QLKRKRG 33 QLRPSD 26 QLRPSD 27 GLRPSD 27 FLQRLTD 36	36 4 59 77 54
p50 p52 c-Rel RelA RelB p50 p52 c-Rel	237 I 217 II 175 I 183 II 270 V 337 I 315 G 270 Q	YDSKAF HDSKSF YDNRAF FDNRAF YDKKST ETSEPF DVSDSF EVSES	PNASNI PGASNI PNTAEI PNTAEI S S S S S S S S S S S S S S S S S S S	LKI - LKI - LRIL LRI - LRI -	VRMDE SRMDE ARVNE CRVNE CRINE KDKEE EDKEE KDAY	ATAGC ATAGS ANCGS ANSS ANSS	VTGGE VRGGD CLGGD CTGGE RQK RRK KQK	b EIYLLA DEVYLLA DEIFLLA SELYLLA 363 341 296	L4 CDKVQKI CDKVQKI CDKVQKI CDKVQKI CDKVQKI	C DDIQIRI DDIEVRI DDIEVRI EDIEVVI EDIEVVI EDISVVI	YEEEEN VLN TGP STA	C' GVWEGF 	GDFSPT GDFSPT GVFSQAI GSFSQAI ADFSQAI	s vyhrory vyhrovy vyhrovy vyhrovi vyhroi	AIVFKT AIVFKT AIVFKT AIVFKT	PKY PPY PPY PPY	KDVNIT HKMKIE CK-AIL ADPSLQ EDLEIS	f RPVTVFI EPVTVKN APVRVSN EPVTVNV	QLRRKSD 33 QLKRKRG 31 QLRRPSD 26 QLRRPSD 27 GLRPSD 27 FLQRLTD 36	36 4 59 77 54
p50 p52 c-Rel RelA RelB p50 p52 c-Rel RelA	237 I 217 II 175 I 183 II 270 V 337 II 315 G 270 Q 278 R	YDSKAF HDSKSF YDNRAF FDNRAF YDKKST ETSEPF DVSDSF EVSESN ELSEPF	QASNI PNTAEI PNTAEI S CPFLYX QFTYX MDFRYI MEFQYI	LKI - LKI - LRIL LRI - LRI - LRI - LRI -	VRMDF SRMDF SRMDF CRVNF CRINF CRINF KDKEF EDKEF KDAY7 DDRHF	a' RTAGC TAGS RNS RNS RNS RNS RNS RNS RNS RNS RNS RN	VTGGE VRGGD CLGGD CTGGE RQK RRK KQK RKR	b EIYLLA DEVYLLA DEIFLLA SELYLLA 363 341 296 304	L ₄ CDKVQKI CDKVQKI CDKVQKI CDKVQKI	C DDIQIRI DDIEVRI DDIEVRI EDIEVVI EDIEVVI EDISVVI	YEEEENC YEDDENC VLN TGP STA	C' GVWEGF 	GDFSPT GDFSPT GVFSQAI GSFSQAI ADFSQAI	5 DVHRQFJ DVHRQVJ DVHRQVJ DVHRQVJ DVHRQIJ	AIVFKT AIVFKT AIVFKT AIVFKT	PKY PPY PPY PPY	KDVNIT HKMKIE CK-AIL ADPSLQ EDLEIS	f RPVTVFI EPVTVKN APVRVSN EPVTVN	OLRRKSD 33 OLKRKRG 31 OLRRPSD 26 OLRRPSD 27 FLQRLTD 36	36 4 59 77 54

Figure S1. The NF- κ B members share high sequence identity except for an additional loop present in p50 and p52 subunits. Sequence alignment of NF- κ B members (mouse) using Jalview Version 2 software ¹. β -strand and α -helix are denoted by black arrow and grey horizontal cylinders, respectively, over the sequence. The DNA-contacting, dimer interfacial and both are marked with red, blue and green colored squares, respectively. The nomenclature of the secondary structure elements and loops are same as by Ghosh *et al*².



Figure S2. Specific labeling/unlabeling of amino acids improved the resonance assignment of p50-NTD. A. 2D [¹⁵N-¹H]-HSQC spectrum of the uniformly-¹⁵N-labeled (black) N-terminal DNA binding domain of p50 subunit of NF-κB (p50NTD) at 298 K superimposed on ¹⁵N-Lys-labeled (pink) and ¹⁵N-Ile-labeled (green) p50NTD. **B.** 2D [¹⁵N-¹H]-HSQC spectrum of the u-[¹⁵N-labeled] p50NTD (black) at 298 K superimposed by u-[¹⁵N-labeled] p50NTD (black) at 298 K superimposed by u-[¹⁵N-labeled] p50NTD (golden yellow). **C.** 2D [¹⁵N-¹H]-HSQC spectrum of the N-terminal DNA binding domain of p50 subunit of NF-κB at 298 K with assignments.



Figure S3. Secondary structure for DNA free form of (A) p50-NTD and (B) RelA-NTD predicted using TALOS-N. α -helix shown with negative and β -strands with positive probability values. Loop regions are given by values between -0.3 to 0.3. The secondary structure as observed in the DNA bound form of p50-RHR (PDB id 1NFK) is illustrated as cartoon over the top of the prediction with β -strands depicted as solid black arrows and helical turns as grey cylinders.



Figure S4. Characteristic imino regions (enclosed in the dashed grey box) of the 1D ¹H spectrum of unlabeled duplex κ B DNA in free form (black), protein bound in 1:1 stoichiometric protein:full κ B DNA ratio (red) and protein bound in 1:1.5 ratio (blue). Formation of complex with the protein leads to broadening of the imino peaks as seen in the red spectrum. DNA was added till excess of the free DNA was observed (blue spectrum). This ensured complete binding of the protein to the DNA.



Figure S5. [¹⁵N,¹H]-HSQC spectra of DNA bound and free p50-NTD is markedly distinct. Superimposed [¹⁵N,¹H]-HSQC spectra of p50-NTD in free (black) and [¹⁵N,¹H]-TROSY spectra of κ B DNA bound form (red). The DNA bound form has markedly distinct chemical shift signatures.



Figure S6. 2D [¹⁵N-¹H]-TROSY spectrum of the rel-homology region of the 73.1 kDa of p50 subunit of NF-κB at 298 K. The assignments are depicted as one letter amino acid code followed by the sequence number in the full-length protein only for well-dispersed peak. The p50 in this study (residues 39-363) had 13 extra residues in the C-terminal end as compared to p50-DD fragment (245-350) which remained unassigned. Peaks marked with red asterisk (*) remained unassigned. The portion of TROSY spectrum enclosed in a red box is zoomed in the respective insets connected with the marked boxes.



Figure S7. Cartoon representation of p50 homodimer (PDB id 1NFK) highlighting the residues that undergo significantly high CSP (blue sphere) upon binding to full-κB DNA.

Supplementary Tables

Protein sample	Labeling Strategy
p50-NTD	u-[¹⁵ N]
p50-NTD	u-[¹⁵ N, ¹³ C] partially deuterated
p50-NTD (¹⁵ N-Lys)	[¹⁵ N-Lys]
p50-NTD (¹⁵ N-Ile)	[¹⁵ N-Ile]
p50-NTD (Arg unlabeled)	$[^{15}N, ^{13}C] + Arg-unlabeled$
p50-NTD (Ile, Leu unlabeled)	$[^{15}N, ^{13}C]$ + Ile, Leu-unlabeled
p50-NTD (Lys unlabeled)	$[^{15}N, ^{13}C] + Lys-unlabeled$
p50-DD	u-[¹⁵ N, ¹³ C, ² H]
p50RHR	u-[¹⁵ N, ¹³ C, ² H]

Table S1. Protein sample description

Supplementary References

Waterhouse, A. M.; Procter, J. B.; Martin, D. M.; Clamp, M.; Barton, G. J., Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009, *25* (9), 1189-91.
Ghosh, G.; van Duyne, G.; Ghosh, S.; Sigler, P. B., Structure of NF-kappa B p50 homodimer bound to a kappa B site. *Nature* 1995, *373* (6512), 303-10.