

**Warten auf die Transkription:
Die humanen Elongationsfaktoren NELF und DSIF**

Dissertation

zur Erlangung des Doktorgrades
der Fakultät für Biologie, Chemie und Geowissenschaften
der Universität Bayreuth

vorgelegt von

Dipl.-Biochem.

Sabine Wenzel

aus Dresden

Bayreuth, 2009

Die vorliegende Arbeit wurde in dem Zeitraum von September 2006 bis November 2009 an der Universität Bayreuth unter der Leitung von Professor Paul Rösch erstellt.

Vollständiger Abdruck der von der Fakultät Biologie, Chemie und Geowissenschaften der Universität Bayreuth genehmigten Dissertation zur Erlangung des akademischen Grades Doktor der Naturwissenschaften (Dr. rer. nat.)

Erster Prüfer:	Prof. Dr. Paul Rösch
Zweiter Prüfer:	Prof. Dr. Matthias Ullmann
Dritter Prüfer:	Prof. Dr. Olaf Stemmann
Prüfungsvorsitz:	Prof. Dr. Rainer Schobert

Tag der Einreichung:	26.11.2009
Zulassung zur Promotion:	03.12.2009
Annahme der Dissertation:	21.04.2010
Auslage der Arbeit:	21.04.2010 – 04.05.2010
Kolloquium:	06.05.2010

Inhaltsverzeichnis

Zusammenfassung	iii
Summary	v
1. Einleitung.....	1
1.1 Die eukaryontische Transkription	1
1.2 Strukturelle Aspekte der RNA-Polymerase II.....	2
1.3 Regulatorische Transkriptionselemente	4
1.4 Der Transkriptionskreislauf	5
1.5 Promotorproximales Arretieren als transkriptioneller Kontrollschritt	7
1.5.1 Elongationskontrolle am proviralen HIV-1 Promotor.....	9
1.5.2 Der negative Elongationsfaktor NELF	10
1.5.3 Der DRB-sensitivity inducing factor DSIF	13
2. Zielsetzung	17
3. Zusammenfassung und Diskussion der Ergebnisse	18
3.1 Der negative Elongationsfaktor NELF	18
3.1.1 Struktur der freien und RNA-gebundenen Form von NELF E RRM	18
3.1.2 Bestimmung der Bindungsaffinität zwischen NELF E RRM und RNA	21
3.2 Der DRB-sensitivity inducing factor DSIF	25
3.2.1 Reinigung des rekombinanten hS pt4.....	25
3.2.2 Reinigung des rekombinanten hSpt4/hSpt5-NGN Komplexes	26
3.2.3 Kristallstruktur von hSpt4/hSpt5-NGN	26
3.3 Ausblick	31
4. Abkürzungsverzeichnis	33
5. Literaturverzeichnis	35
6. Publikationsliste.....	45
6.1 Einzelarbeit A.....	45
6.2 Einzelarbeit B.....	45
6.3 Einzelarbeit C	46
6.4 Einzelarbeit D	46
7. Einzelarbeit A	47
8. Einzelarbeit B	57
9. Einzelarbeit C	69
10. Einzelarbeit D	77
11. Danksagung	89
12. Erklärung	91

Zusammenfassung

Die humanen Elongationsfaktoren *negative elongation factor* (NELF) und *5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) sensitivity inducing factor* (DSIF) sind Schlüsselfaktoren beim promotorproximalen Arretieren in der frühen Elongationsphase der Transkription. Dabei interagieren beide Faktoren direkt mit RNA-Polymerase II. NELF bindet über die *RNA recognition motif* (RRM)-Domäne der Untereinheit E zusätzlich an die naszierende RNA. Eine derartige Transkriptionsblockade tritt auch während der Transkription des Genoms des humanen Immunschwächevirus 1 (HIV-1) Provirus auf. NELF bindet hierbei an die Haarnadel-Struktur der naszierenden RNA (HIV-1 TAR RNA). Ein Ziel dieser Arbeit war die strukturelle Analyse der RRM-Domäne von NELF E sowie die Untersuchung einer möglichen sequenzspezifischen Bindung von NELF E RRM an HIV-1 TAR RNA. Außerdem galt es, die Struktur der kleinen Untereinheit von DSIF, hSpt4, zu bestimmen und ihre Interaktion mit der NGN-Domäne der großen DSIF-Untereinheit hSpt5 näher zu charakterisieren.

Die Struktur der RRM-Domäne von NELF E wurde mittels magnetischer Kernresonanz (NMR) bestimmt (PDB: 2BZ2). Sie zeigt die für RRM-Domänen typische Faltungstopologie ($\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$) und faltet in ein viersträngiges β -Faltblatt mit zwei α -Helices, die annähernd senkrecht zueinander angeordnet sind und gemeinsam gegen eine Seite des Faltblattes packen. An der RNA-Bindung sind die zentralen β -Stränge mit den beiden aromatischen Resten Tyr43 und Phe77 aus dem Konsensussequenzmotiv *ribonucleoprotein domain 2* (RNP2) bzw. RNP1 involviert. Der in der freien Form unstrukturierte C-Terminus der RRM bildet in der gebundenen Form (PDB: 2JX2) eine 3_{10} -Helix aus, die höchstwahrscheinlich direkt an der RNA-Bindung beteiligt ist. RRM-Domänen gelten als klassische einzelstrangbindende Domänen. Die prognostizierte Bindestelle von NELF E ist jedoch die doppelsträngige HIV-1 TAR RNA Stammregion. Die durchgeführten ^{15}N -NMR-Titrationsexperimente und Fluoreszenzgleichgewichtstitrationen mit die HIV-1 TAR RNA umfassenden RNA Oligonukleotiden zeigten, dass NELF E RRM präferentiell einzelsträngige RNA bindet. Die ermittelten K_D -Werte lagen dabei alle im niederen mikromolekularen Bereich. NELF E RRM zeigt somit im Vergleich zu anderen RRM-Domänen eine schwächere Affinität für RNA. Eine sequenzspezifische Bindung an die HIV-1 TAR RNA konnte, auch in Versuchen mit der gesamten NELF E Untereinheit, nicht beobachtet werden. Vieles deutet somit darauf hin, dass RNA-Bindung durch NELF E sequenzunabhängig stattfindet. Der Zeitpunkt der RNA-Bindung während der Elongation der Transkription wird womöglich durch andere Faktoren bzw. Ereignisse gesteuert.

NELF übt seine inhibitorische Wirkung nur im Zusammenspiel mit dem Elongationsfaktor DSIF aus, über den es nur wenig strukturelle Informationen gibt. Circular dichroismus und

NMR-spektroskopische Untersuchungen bestätigten, dass das gereinigte rekombinante hSpt4 über eine definierte Tertiärstruktur verfügt. Die Analyse durch optische Emissionsspektrometrie mit induktiv gekoppeltem Plasma (ICP-OES) zeigte zudem, dass Zink in äquimolaren Mengen vorliegt und der prognostizierte Zinkfinger strukturiert ist. Instabilität der Probe, auch nach Variation der Puffer- und Messbedingungen, verhinderte aber eine Strukturbestimmung mittels NMR. Die Koexpression von hSpt4 mit der NusG N-terminalen Homologie Domäne von hSpt5 (hSpt5-NGN) als dessen Interaktionspartner ermöglichte schließlich die Reinigung des Heterodimers und Gewinnung hochqualitativer Proteinkristalle. Die Kristallstruktur von hSpt4/hSpt5-NGN wurde bis zu einer Auflösung von 1.55 Å gelöst (PDB: 3H7H). hSpt5-NGN besteht aus einem zentralen, viersträngigen β -Faltblatt und aus einzelnen α -Helices, die von beiden Seiten gegen das Faltblatt packen. hSpt4 verfügt über ein zwei- und ein dreisträngiges β -Faltblatt, die senkrecht zueinander angeordnet sind und sich auf der Interaktionsfläche zugewandten Seite befinden. Zusätzliche α -helikale Elemente werden durch den ausgebildeten Zinkfinger fixiert. Die Komplexstruktur ist von einem zentralen sechssträngigen intermolekularen β -Faltblatt geprägt. Die große Helix $\alpha 1$ von hSpt5-NGN packt an der Interaktionsstelle gegen das gemeinsame β -Faltblatt und stabilisiert durch hydrophobe Wechselwirkungen die Bindung. hSpt4/hSpt5-NGN zeigt sehr große Strukturähnlichkeit zur kürzlich publizierten homologen Komplexstruktur von Spt4/Spt5-NGN aus *Saccharomyces cerevisiae*. Ein für die Bindung essentieller Glutamatrest (E338) von Spt5-NGN ist auch in hSpt5-NGN konserviert (E228). Der Komplex hSpt4/hSpt5-NGN(E228Q) konnte wegen der Unlöslichkeit von hSpt5-NGN(E228Q) nicht gereinigt werden. Dies deutet an, dass die Interaktion zwischen hSpt4 und hSpt5-NGN(E228) gestört ist. Somit ist E228 von hSpt5-NGN höchstwahrscheinlich ebenso essentiell für die Wechselwirkung mit hSpt4 wie im homologen Hefekomplex. hSpt5-NGN hat auch strukturelle Ähnlichkeiten zur N-terminalen Domäne (NTD) des bakteriellen Transkriptionsfaktors NusG. *Escherichia coli* (*Ec*)NusG-NTD verfügt an der Position des Glutamatrestes über einen Glutaminrest (Q72). hSpt4-Bindung an (*Ec*)NusG-NTD bzw. (*Ec*)NusG-NTD(Q72E) wurde jedoch nicht beobachtet. Ein möglicher Grund dafür ist die unterschiedliche Ladungsverteilung an der hSpt4-Bindungsfläche. hSpt4 besitzt hingegen Ähnlichkeiten zum archaealen Transkriptionsfaktor RpoE“. Überlagerung der Strukturen von hSpt4 und RpoE“ aus *Pyrococcus furiosus* (PDB: 1RYQ) zeigten, dass beide den Zinkfinger und die senkrecht zueinander angeordneten β -Faltblätter gemein haben.

Strukturbestimmung sowie funktionelle Analyse von NELF E RRM und hSpt4/hSpt5-NGN repräsentieren somit einen weiteren Schritt auf dem Weg zur Entschlüsselung der komplexen zellulären Vorgänge während der eukaryontischen Transkription.

Summary

The human transcription factors negative elongation factor (NELF) and 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB) sensitivity inducing factor (DSIF) are key factors in transcriptional pausing during early elongation. Pausing of the transcription machinery is induced by binding of both factors to the RNA polymerase II. Additionally, NELF interacts with the nascent RNA transcript via its E subunit. A similar pausing event has been detected for the transcription of the proviral human immunodeficiency virus 1 (HIV-1) genome, in which NELF E binds to the hairpin structure of the nascent RNA transcript termed TAR RNA. The aim of this work was the structural analysis of the E subunit of NELF. An emphasis was put on the investigation of a sequence specific binding of the RNA recognition motif (RRM) domain to RNA. Additionally, the structure of the small hSpt4 subunit of DSIF was to be determined and the interaction of hSpt4 with the large DSIF subunit hSpt5 should be characterized.

The solution structure of NELF E RRM (PDB: 2BZ2) was determined by nuclear magnetic resonance (NMR). It exhibits the typical RRM folding topology ($\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$) and consists of a four stranded antiparallel β -sheet and two α -helices, that are arranged perpendicularly to each other and pack against the β -sheet. Residues located in the central β -sheet are involved in RNA binding, including the two aromatic residues Tyr43 and Phe77 of the consensus motifs ribonucleoprotein domain 2 (RNP2) and RNP1, respectively. The solution structure of the RNA-bound form of NELF E RRM (PDB: 2JX2) clearly showed, that it comprises a 3_{10} -helix at the C-terminus, which is directly involved in RNA binding. In contrast, the C-terminus of the unbound form is unstructured and highly flexible. While RRM domains are considered as truly single strand binding domains, the HIV-1 TAR RNA stem region is the proposed binding site of NELF E. ^{15}N -NMR titration experiments and fluorescence equilibrium titrations performed with short RNA oligonucleotides covering the HIV-1 TAR RNA stem region showed that NELF E RRM preferentially binds single stranded RNA. All K_D -values calculated were in the lower micromolar range, indicating a rather weak binding affinity to RNA as compared to other known RRM domains. Furthermore, no sequence specific binding of NELF E RRM as well as of the full length NELF E was observed. The obtained results indicated that NELF E binds RNA in a sequence independent manner. Furthermore, fluorescence equilibrium titration experiments with the full length NELF E subunit also failed to show a sequence specific binding to HIV-1 TAR RNA confirming that RNA binding of NELF is independent of the RNA sequence. RNA binding of NELF during transcription might therefore be regulated by other cellular factors or events.

NELF action is dependent on the presence of DSIF, of which only little structural information is available. Circular dichroism and NMR measurements verified that the purified recombinant hSpt4 possesses a defined tertiary structure. Additional inductively coupled plasma optical emission spectrometry (ICP-OES) analysis showed that zinc was present in equimolar concentrations confirming the formation of a structured zinc finger. Due to the instability of hSpt4 samples, which could not be resolved by modifying buffer or measurement parameters, structure determination by NMR was not feasible. Coexpression of hSpt4 and the NusG N-terminal homology domain of hSpt5 (hSpt5-NGN), which represents the proposed binding region of hSpt4, rendered the purification of the complex possible and yielded high quality protein crystals. The crystal structure of hSpt4/hSpt5-NGN was determined up to a resolution of 1.55 Å (PDB: 3H7H). hSpt5-NGN contains a central four-stranded antiparallel β -sheet and several α -helices which pack against the sheet. hSpt4 harbors one two- and one three-stranded antiparallel β -sheet which are arranged perpendicularly to each other. Additional α -helices are stabilized by the zinc finger. A central intermolecular six-stranded antiparallel β -sheet is formed in the hetero-dimer. A large helix (α 1) of hSpt5-NGN packs against the β -sheet at the dimer interface and stabilizes the dimerization by hydrophobic contacts. hSpt4/hSpt5-NGN shows high structural similarity to the recently published homologous Spt4/Spt5-NGN complex of *Sacharomyces cerevisiae*. A glutamate residue (E338) of Spt4/Spt5-NGN was shown to be essential for Spt4 binding. It is also conserved in hSpt4/hSpt5-NGN(E228). hSpt4/hSpt5-NGN(E228Q) could not be purified due to the insolubility of hSpt5-NGN(E228Q). This indicates that the interaction between hSpt5-NGN(E228Q) and hSpt4 is disrupted. Therefore, E228 of hSpt5-NGN seems to be essential for hSpt4 binding in the human complex as well. Additionally, hSpt5 shows structural similarities to the N-terminal domain (NTD) of the bacterial transcription factor NusG. *E. coli* (*Ec*)NusG harbors a glutamine residue (Q72) at the position corresponding to E228 of hSpt5-NGN, but does not bind to hSpt4. Also, *Ec*NusG-NTD (Q72E) shows no binding affinity for hSpt4. This may be due to differences in charge distribution at the hSpt4 binding interface of hSpt5-NGN in comparison to *Ec*NusG-NTD. hSpt4 exhibits structural similarities to the archaeal transcription factor RpoE". Comparison of hSpt4 with RpoE" from *Pyrococcus furiosus* (PDB: 1RYQ) showed, that both factors inherit the zinc finger region and the β -sheets.

Structure determination and functional analysis of NELF E RRM and hSpt4/hSpt5-NGN is another step forward to encrypt the complex cellular events during eukaryotic transcription.

1. Einleitung

1.1 Die eukaryontische Transkription

Die Transkription ist das Umschreiben der in Form von Desoxyribonukleinsäure (DNA) gespeicherten genetischen Information in Ribonukleinsäure (RNA). Dieser essentielle Prozess stellt die Grundlage der Proteinbiosynthese dar. Von entscheidender Bedeutung für die Zellerhaltung aber auch für die Zellentwicklung und -differenzierung ist die präzise Regulation des Ablaufes der Transkription, da Defekte darin zum Tod des Organismus führen können. Dabei unterscheidet sich der grundlegende Ablauf der Transkription zwischen Bakterien, Archaeen und Eukaryonten nicht wesentlich. Es ist nur die Anzahl an beteiligten Faktoren oder regulatorischen DNA Sequenzen, die einen positiven aber auch negativen Einfluss auf die Transkription haben können, in eukaryontischen Systemen um ein Vielfaches höher. Die RNA-Polymerase (RNAP) katalysiert diese DNA-abhängige RNA-Synthese. Während Bakterien und Archaeen jeweils nur eine RNAP besitzen, gibt es in Eukaryonten drei, die entsprechend mit den Ziffern I, II und III bezeichnet werden. Eukaryontische RNAPs sind Multiproteinkomplexe, die aus 14 (I), 12 (II) bzw. 17 (III) Untereinheiten aufgebaut sind und von denen jeweils zehn den RNAP Kern (*RNAP core*) bilden. Davon sind fünf Untereinheiten unter den drei eukaryontischen RNAPs austauschbar (Cramer *et al.*, 2008). Eukaryontische RNAPs synthetisieren RNA unterschiedlicher Genklassen: RNAP I ribosomale RNA (rRNA), RNAP II Boten-RNA (mRNA) und kleine nukleäre RNAs (*small nuclear RNA*, snRNA) und RNAP III tRNA sowie andere snRNAs (Cramer *et al.*, 2008; Egloff und Murphy, 2008). RNAP II ist somit für die Transkription aller proteinkodierenden Gene (ca. 20 000 beim Menschen) verantwortlich.

Im Vergleich zu Bakterien und Archaeen ist bei Eukaryonten die Transkription räumlich von der Translation getrennt. Die synthetisierte RNA wird vom Zellkern in das Cytoplasma exportiert und den Ribosomen für die Translation zugeführt. Zwischendurch unterläuft die naszierende RNA mehrere Prozessierungsschritte, die ihre Stabilität aber auch die zu translatierende Information erheblich beeinflussen. Dazu gehören mRNA-*Capping*, bei dem das 5' Ende der mRNA mit einer 7-Methylguanosin-Kappe (7mG-Kappe) versehen wird, sowie Polyadenylierung, bei der am 3' Ende ein Poly-(A) Schwanz angehängt wird. Außerdem werden naszierende mRNA-Transkripte gespleißt, d.h. es werden Introns entfernt. Gleichzeitig können durch alternatives Spleißen verschiedene Proteinisoformen erzeugt werden.

Erst in den letzten Jahren wurde das generelle Bild übernommen, dass die mRNA-Prozessierung nicht nach der Transkription, sondern kotranskriptionell stattfindet und die zuständigen Faktoren in regem Kontakt mit der Transkriptionsmaschinerie stehen.

Die Klärung des Zusammenspiels von RNAP II mit allen Transkriptionsfaktoren und mRNA-Prozessierungsfaktoren ist daher von entscheidendem Interesse, kann sie doch ein besseres Verständnis über den Ablauf zellulärer Prozesse liefern und helfen die Entstehung bestimmter tumorassoziierter Krankheiten nachzuvollziehen.

Da sich diese Arbeit mit Transkriptionsfaktoren befasst, die an der durch die RNAP II vermittelten Transkription beteiligt sind, wird im Folgenden, wenn nicht anders erwähnt, nur die von der RNAP II katalysierte Transkription behandelt.

1.2 Strukturelle Aspekte der RNA-Polymerase II

Die Untereinheiten von RNAPs, die das katalytische Zentrum der RNAP II aufbauen, sind von Bakterien bis Archaeen hochkonserviert. Während bakterielle RNAPs hauptsächlich aus den konservierten Untereinheiten bestehen, verfügen archaeale und eukaryontische RNAPs über zusätzliche, für ihre Funktion wichtige Untereinheiten (Werner, 2008). So besitzt RNAP II neben dem zehn Untereinheiten umfassenden *core*-Komplex (Rpb1-Rpb3, Rpb5-Rpb6 und Rpb8-Rpb12), die beiden Untereinheiten Rpb4 und Rpb7, die zwar nicht für die katalytische Aktivität, aber für die promotorabhängige Initiation essentiell sind (Abb.1A) (Edwards *et al.*, 1991).

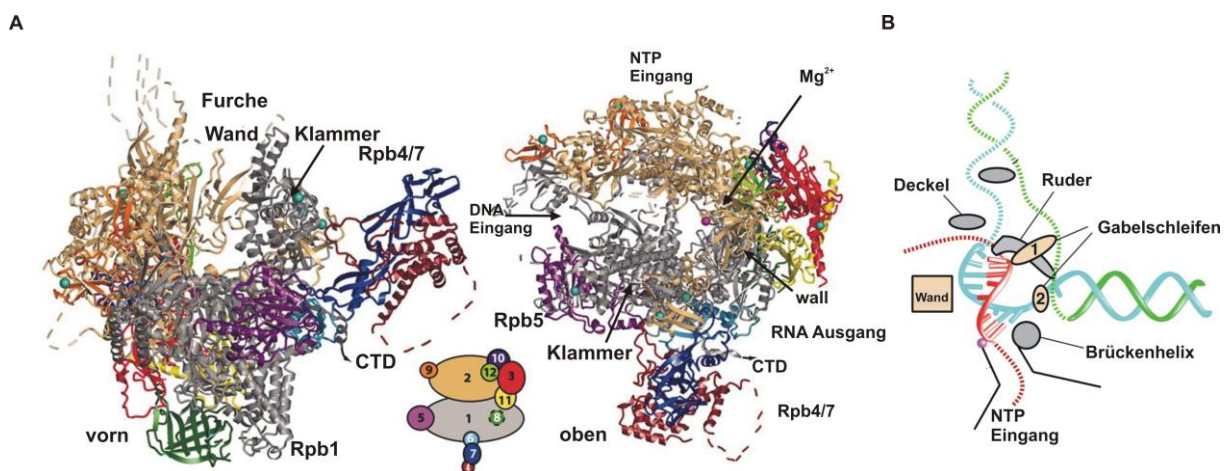


Abb. 1: RNA-Polymerase II Struktur aus *Saccharomyces cerevisiae*.

(A) Struktur der 12 Untereinheiten RNAPII in zwei verschiedenen Ansichten (vorn und oben). Rpb1 bis Rpb12 sind entsprechend der schematischen Darstellung (Mitte) gefärbt. Das Magnesiumion des aktiven Zentrums ist als lila Kugel dargestellt. Die Lage der beweglichen Klammer (*clamp*)-Domäne, der Furche (*cleft*) und Wand (*wall*)-Domäne ist angegeben; der NTP-Eintrittskanal bzw. RNA Ausgang sind ebenfalls markiert (nach: Armache *et al.*, 2005). (B) Schematische Darstellung des aktiven Zentrums von RNAP II mit gebundenen Nukleinsäuren, an der Strangtrennung bzw. -stabilisierung beteiligte RNAP II Elemente sind angegeben, RNA ist in rot, DNA in grün/blau und das Magnesiumion als lila Kugel dargestellt (nach Gnatt *et al.*, 2001)

Die zwei größten Untereinheiten Rpb1 und Rpb2 von RNAP II bilden zusammen eine positiv geladene Furche (*cleft*), in der die DNA gebunden wird. Die restlichen kleineren Untereinheiten der RNAP II befinden sich an der Oberfläche von Rpb1 und Rpb2 (Abb. 1).

Die Furche wird an einer Seite durch die Wand (*wall*)-Domäne der Rpb2 abgegrenzt und umschließt zusammen mit der Brückenhelix (*bridge helix*) von Rpb1 das aktive Zentrum (Cramer *et al.*, 2000; Cramer *et al.*, 2001). Im Allgemeinen gleicht der *RNAP core*-Komplex einer Kresschere und verfügt unter anderem über das bewegliche Klammer (*clamp*)-Element. Die Klammer-Domäne setzt sich aus Teilen von Rpb1, Rpb2 und Rpb3 zusammen und ist über flexible *switch*-Helices, die ein Öffnen und Schließen der Schere ermöglichen, mit dem Kernbereich verbunden (Shilatifard *et al.*, 2003; Hahn, 2004). Der Rpb4/7 Komplex bindet unterhalb der mobilen Klammer-Domäne und bildet einen Keil, der die freie Bewegung der Klammer-Domäne verhindert, wie es erstmalig Modelle der 12-Untereinheiten-RNAP II aus *Saccharomyces cerevisiae* basierend auf röntgenkristallographischen Daten zeigen konnten (Abb. 1A) (Armache *et al.*, 2003; Bushnell und Kornberg., 2003). Das schließt allerdings aus, dass doppelsträngige DNA je im aktiven Zentrum der RNAP II gebunden wird, wie zunächst vermutet, da die Furche im geschlossenen Zustand nur einzelsträngige DNA aufnehmen kann (Gnatt *et al.*, 2001; Cramer *et al.*, 2001). Stattdessen wird angenommen, dass der DNA-Einzelstrang nach dem Aufschmelzen an der Promotorregion in das aktive Zentrum hineinrutscht. Für den Nukleotideinbau wird bei allen Polymerasen von einem Zwei-Metall-Mechanismus ausgegangen (Steitz, 1998). Im aktiven Zentrum ist ein Magnesiumion fest über drei Aspartatreste einer Schleifen-Region der Rpb1 gebunden. Das für die enzymatische Reaktion nötige zweite Magnesium wird nur schwach gebunden und tritt offenbar mit jedem neuen Substrat-Nukleotid in das aktive Zentrum ein (Westover *et al.*, 2004b; Brückner *et al.*, 2009).

Eine Pore, die sich in direkter Nachbarschaft zum gebundenen Magnesiumion befindet und sich nach außen hin in einen Trichter öffnet, stellt den Nukleotid-Eintrittskanal dar, so dass die Nukleotide direkt zum aktiven Zentrum geleitet werden (Abb. 1B). Im Elongationskomplex zwingt die Wand-Domäne den DNA-RNA-Hybrid in eine 90° Drehung relativ zur stromaufwärts gelegenen DNA. Die Proteinschleifen Deckel (*lid*), Ruder (*rudder*) und zwei Gabelschleifen (*fork loop*), die von Rpb1 bzw. Rpb2 stammen, sorgen für die Strangtrennung des Hybrids und verhindern ein Reassoziieren beider Stränge (Abb. 1B) (Gnatt *et al.*, 2001; Westover *et al.*, 2004a). Dabei agiert die Deckelschleife als Keil, der RNA und DNA-Strang auseinander treibt und den naszierenden Strang in den Austrittskanal leitet. Die Ruder-Schleife verhindert eine Reassoziierung von RNA und DNA, indem sie direkt mit der DNA interagiert. Die Gabelschleife 1 steht in direktem Kontakt mit der RNA des DNA-RNA-Hybrids und verhindert ein zu weites Aufschmelzen des Hybridkomplexes. Die Gabelschleife 2, hilft die Reassoziierung der DNA am stromabwärtsgelegenen Ende zu verhindern (Abb. 1B) (Gnatt *et al.*, 2001; Westover *et al.*, 2004a). Die einzelnen Proteinschleifen binden nicht nur verschiedene Bereiche der Nukleinsäuren, sondern interagieren auch untereinander und bilden ein Strang-Schleifennetzwerk aus (Westover *et al.*, 2004a). Der RNA-Austrittskanal

befindet sich in der Nähe der C-terminalen Domäne (CTD) von Rpb1 (Cramer *et al.*, 2001). Die CTD ist einzigartig unter den RNAPs und spielt eine essentielle Rolle bei der Transkription. Sie besteht aus hochkonservierten Tandem-Wiederholungen der Konsensussequenz $Y_1S_2P_3T_4S_5P_6S_7$, wobei die Serinreste Ser2 und Ser5 einer Phosphorylierung und anderen reversiblen Modifikationen während der Transkription unterliegen (Buratowski, 2003). Die Anzahl der Heptapeptidwiederholungen nimmt mit steigender Komplexität des Organismus zu. So besitzt *S. cerevisiae* 26, *Drosophila melanogaster* 42 und *Homo sapiens* 52 Wiederholungen (Wu *et al.*, 2003; Palancade *et al.*, 2003). Das Phosphorylierungsmuster und die anderen Modifikationen spiegeln das Stadium der Transkription wider und geben vor, welche Faktoren binden können. Dieser „CTD-Code“ ermöglicht eine dynamische Anpassung an die entsprechende Transkriptionsphase, ohne dass die Faktoren ständig gebunden sein müssen (Buratowski, 2003). Zwar ist die CTD für die katalytische Aktivität der RNAP II entbehrlich, aber dennoch essentiell für den Ablauf und die Regulation der Transkription (Egloff und Murphy, 2008).

Mit Hilfe der Kristallstruktur des Elongationskomplexes konnte außerdem nachgewiesen werden, dass ein ca. 8 bp langer DNA/RNA-Hybrid und zusätzlich ein ca. 10 nt langer RNA-Einzelstrang im Inneren der RNAP vorliegen, so dass etwa 18 nt des naszierenden RNA Stranges nicht zugänglich sind (Gnatt *et al.*, 2001; Westover *et al.*, 2004a; Kettenberger *et al.*, 2004).

1.3 Regulatorische Transkriptionselemente

Die RNAP II ist im Allgemeinen nicht in der Lage eigenständig die promotorabhängige Transkription zu initiieren. Für eine zielgerichtete, genspezifische Transkription sind Faktoren nötig, welche die RNAP II zur Promotorregion und somit zur Transkriptionsstartstelle rekrutieren (Woychik und Hampsey, 2002). Derartige Faktoren, die die Promotorregion erkennen und die DNA entwinden können, werden als generelle Transkriptionsfaktoren (GTFs) zusammengefasst und sind essentiell für die Initiation der Transkription. Dazu gehören, benannt nach der Elutionsreihenfolge ihrer ersten Reinigung, TFIIA, TFIIB, TFIID, TFIIF und TFIIH. Der Kernpromotor (*core promoter*) ist der minimale Bereich der DNA, der für eine Rekrutierung und Positionierung der RNAP II an der Transkriptionsstartstelle nötig ist. Er umfasst in etwa die Region -40 bp bis +40 bp um die Transkriptionsstartstelle und enthält verschiedene Kombinationen an konservierten, funktionellen DNA Elementen (z.B. TATA-Box, BRE (*B recognition element*) u.a.) (Hahn, 2004). Dabei sind nie alle Elemente an einem Promotor vorhanden, sondern nur in bestimmten Kombinationen. Genomweite Promotoranalysen haben überraschenderweise gezeigt, dass die TATA Box, obwohl sie lange Zeit als das Promotorelement schlechthin galt, bei weniger als 4 % der humanen Gene vorliegt (Yang *et al.*, 2007a). Dabei ist noch weitestgehend ungeklärt, inwieweit sich der

Transkriptionsablauf im Hinblick auf die beteiligten Faktoren bei Genen mit TATA-haltigen Promotoren von jenen mit TATA-losen Promotoren unterscheidet. Weitere genspezifische Elemente sind die promotornahen DNA-Bereiche (*promoter proximal sites*), die sich stromaufwärts in unmittelbarer Nähe des Promotors befinden, und sogenannte *enhancer*, aber auch *silencer*-Regionen, die ebenfalls stromaufwärts des Promotors, aber in wesentlich größerer Entfernung (Kilobasen-Dimension), lokalisiert sind (Fuda *et al.*, 2009). Sowohl *enhancer* und *silencer*-Regionen als auch promotornahe Bereiche stellen Bindungsstellen für Faktoren (Aktivatoren bzw. Repressoren) dar, die entweder direkt oder über Koaktivatoren mit der Transkriptionsmaschinerie in Wechselwirkung treten. Somit haben derartige regulatorische DNA-Sequenzen einen Einfluss auf Zeitpunkt, Ort und Dosis einer Genexpression.

1.4 Der Transkriptionskreislauf

Zwar wurde im letzten Jahrzehnt das Verständnis der Transkription wesentlich verbessert, das genaue Zusammenspiel der zahlreichen Transkriptionsfaktoren und molekularen Abläufe wirft dennoch viele Fragen auf. Die Transkription kann grob in drei verschiedene, ineinander übergehende Phasen unterteilt werden: Initiation, Elongation und Termination. Dabei kann vor allem die Initiation in verschiedene Subphasen eingeteilt werden, welche nahtlos in die Elongation übergehen. Der genaue Ablauf der Termination, welche die Freisetzung der RNA aus der Transkriptionsmaschinerie und Regeneration der RNAP II umfasst, ist bis dato noch nicht hinreichend geklärt. Sie ist aber eng mit der Polyadenylierung der mRNA verbunden. An regulatorische DNA-Elemente gebundene Aktivatoren stimulieren über Koaktivatoren, wie z.B. Mediator, die Rekrutierung des Präinitiationskomplexes (PIC). Mediator ist ein Multiproteinkomplex, der zwischen gebundenen Aktivatoren und der Transkriptionsmaschinerie vermittelt (Kim *et al.*, 1994). Er ist ein essentieller Faktor und für die Transkription nahezu aller Gene nötig (Holstege *et al.*, 1998). Er bindet unter anderem die unphosphorylierte CTD der Rpb1 (Woychik und Hampsey, 2002), dissoziiert jedoch während der Elongation von der RNAP II (Weir und Elkins, 1993).

Aktivatorproteine rekrutieren aber auch direkt chromatinmodifizierende Faktoren zur Promotorregion (Narlikar *et al.*, 2002). Eukaryontische DNA ist nicht direkt zugänglich, da die DNA in Nukleosomen verpackt ist. Chromatinmodifikationsfaktoren katalysieren die post-translatorischen Modifikationen der Histone, was ein dynamischer Prozess ist und mit dem Transkriptionsstatus eines Gens korreliert. Diese Modifikationen beeinflussen die Wechselwirkung zwischen DNA und Histonen, stellen aber auch Bindungsstellen für andere Faktoren dar (Seet *et al.*, 2006). Zusätzlich sind Chromatinremodellierungsfaktoren in der Lage, unter ATP-Hydrolyse Histon-DNA Kontakte zu lösen und Nukleosomen entlang der

DNA zu verschieben. Dadurch wird die DNA für Transkriptionsfaktoren aber auch für die RNAP II zugänglich (Li *et al.*, 2007; Reid *et al.*, 2009) (Abb. 2 (1 und 2)).

Unter dem Begriff PIC versteht man die an der Promotorregion assemblierte Transkriptionsmaschinerie bestehend aus GTFs und RNAP II (siehe 1.3). Mediator-Bindung stabilisiert die PIC-Assemblierung (Fuda *et al.*, 2009). Die kooperative PIC-Bildung führt zur Ausbildung des geschlossenen Komplexes (Abb. 2(2)) (Martinez, 2002).

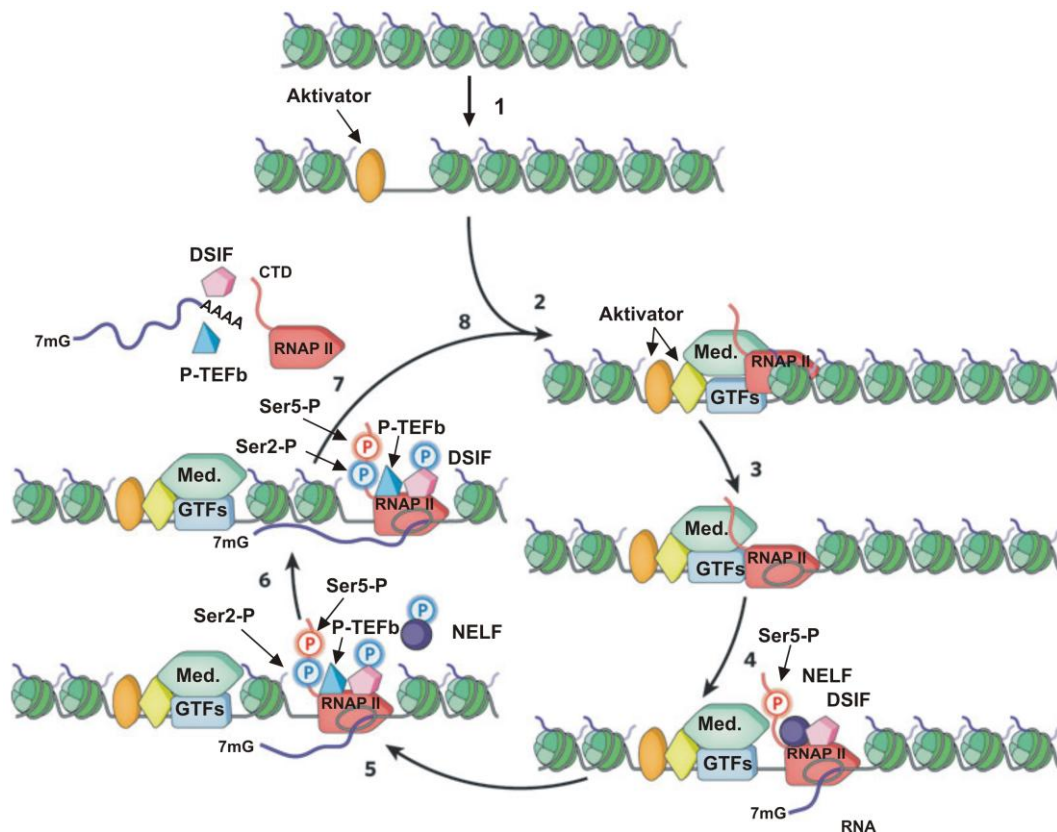


Abb. 2: Transkriptionszyklus der eukaryontischen Transkription durch RNAP II

Aktivatorproteine (orange) binden *enhancer*-Elemente der DNA und rekrutieren Chromatinremodellierungsfaktoren zu den Nucleosomen (grün) (1); weitere Aktivatoren rekrutieren Koaktivatoren (z.B. Mediator (Med.)) und die generellen Transkriptionsfaktoren (GTFs), sowie die RNAP II zur Promotorregion (2); unter ATP-Hydrolyse wird die DNA um die Transkriptionsstartstelle aufgeschmolzen und der Matrizenstrang der DNA im aktiven Zentrum platziert; dies führt zur Ausbildung des offenen Komplexes (3); die ersten Nucleotidbindungen werden geknüpft und die Kontakte zwischen RNAP II und den GTFs bzw. Mediator gelöst (*promoter clearance*); begleitet wird dies durch Ser5-Phosphorylierung der C-terminalen Domäne (CTD) der RNAP II durch TFIIF. Nach der Transkription von 20-40 nt geht der Transkriptionskomplex in eine durch DSIF und NELF induzierte promotorproximale Blockade über (4); P-TEFb phosphoryliert NELF, DSIF und Ser2 der CTD; NELF dissoziiert vom Transkriptionskomplex und DSIF wird zum Transkriptionsaktivator (5); RNAP II geht über in eine produktive Elongation (6); bei der Termination wird die RNA entlassen und polyadenyliert; die RNAP II wird für weitere Transkriptionsrunden regeneriert (7) (nach Fuda *et al.*, 2009)

Die intrinsische Helikase von TFIIF schmilzt unter ATP-Hydrolyse 11-15 bp um die Transkriptionsstartstelle auf und bildet den offenen Komplex, wodurch der Matrizenstrang im aktiven Zentrum der RNAP II platziert werden kann (Abb. 2 (3)). ATP-Hydrolyse ist für konformationelle Änderungen des PICs nötig und ist einzigartig bei RNAP II (Wang *et al.*,

1992; Saunders *et al.*, 2006). Die Knüpfung der ersten Phosphodiesterbindung stellt die Initiation dar, der sich die Phase der *promoter clearance* (Verlassen der Promotorregion) anschließt (Abb. 2 (4)). Darin wird die Bindung der RNAP II mit der Promotorregion gelockert und die Wechselwirkung mit den GTFs gelöst. *Promoter clearance* geht mit ATP-Hydrolyse durch TFIIH und Ser-5 Phosphorylierung der CTD von Rpb1 durch die CDK7 Untereinheit von TFIIH einher. Sie gilt als abgeschlossen, wenn die naszierende RNA stabil in die RNAP II integriert ist. Diese frühe Phase ist oft begleitet von der Synthese von Abbruchfragmenten, der sog. *abortive initiation*, in der nur kurze RNA-Fragmente synthetisiert werden, ohne dass die RNAP II von der DNA dissoziiert (Saunders, A. 2006;). Bevor ein stabiler Elongationskomplex ausgebildet wird, unterliegt die RNAP II dem Rückwärtslaufen (*backtracking*), was zu Fehlpaarungen des 3'OH-Endes der RNA im aktiven Zentrum der RNAP II führt. TFIIIS stimuliert die Spaltung der Phosphodiesterbindung durch Einfügen einer Haarnadel-Struktur (*hairpin*) und ermöglicht, dass ein neues transkriptionskompetentes 3'OH-Ende im aktiven Zentrum erzeugt wird und die RNAP II fortfahren kann (Kettenberger *et al.*, 2004; Kettenberger *et al.*, 2003). Ab einer Transkriptlänge von 30 nt wird von einer stabilen Elongation ausgegangen mit Bildung von Gesamtlängen mRNA. Für das Überwinden von intrinsischen Stoppstellen ist TFIIIS aber auch während der Elongation wichtig (Saunders *et al.*, 2006; Fish und Kane, 2002). In dieser frühen Elongationsphase bindet das *Capping*-Enzym an die CTD der RNAP II und fügt an der naszierenden RNA die 7mG-Kappe an. Bei einer Vielzahl an Genen kommt es im Bereich von +20 nt bis +40 nt durch Bindung des negativen Elongationsfaktors (NELF) und des *DRB sensitivity inducing factors* (DSIF) zum Halt der Transkriptionsmaschinerie (Abb. 2, (4)) (s. a. 1.5). Die Transkriptionsblockade wird durch den positiven Transkriptionselongationsfaktor b (P-TEFb) aufgehoben, der die CTD der Rpb1 am Ser2-Rest sowie NELF und DSIF phosphoryliert (Abb. 2 (5)) (Wada *et al.*, 1998b; Fujinaga *et al.*, 2004b; Yamada *et al.*, 2006). Nach Dissoziation von NELF geht der Transkriptionskomplex in eine produktive Elongation über (Abb. 2 (6)). Die phosphorylierte RNAP II CTD dient während der einzelnen Phasen als Plattform für weitere Elongationsfaktoren, die vor allem wichtig für mRNA-Capping, Spleißen und Polyadenylierung sind (Sims *et al.*, 2004). In der abschließenden Terminationsphase wird die mRNA entlassen und die RNAP II für weitere Transkriptionsrunden regeneriert (Abb. 2 (7)).

1.5 Promotorproximales Arretieren als transkriptioneller Kontrollschritt

Die Initiation ist der geschwindigkeitsbestimmende Schritt der Transkription und wird durch eine Vielzahl von Faktoren gesteuert. In den letzten Jahren hat sich jedoch immer mehr herauskristallisiert, dass der Elongationsschritt ebenso einer strengen Regulation unterliegt. Dazu gehört das promotorproximale Arretieren der RNAP II (*promoter proximal pausing*).

Promotorproximales Arretieren tritt bei einer mRNA Transkriptlänge von etwa 20-40 nt ein (Rougvie und Lis, 1988) und ist ein Postinitiationsereignis, da die Transkriptionsmaschinerie bereits Eigenschaften eines Elongationskomplexes aufweist. In diesem Stadium hat die RNAP II die Promotorregion verlassen, die an der Initiation beteiligten GTFs sind dissoziiert und die CTD ist am Ser5-Rest phosphoryliert (Koch *et al.*, 2008). Ein derartiger pausierender Transkriptionskomplex wurde erstmalig an nicht induzierten Hitzeschockgenen in *D. melanogaster* entdeckt (Rougvie und Lis, 1988).

Als die zentralen Schlüsselfaktoren des promotorproximalen Arretierens werden die drei Faktoren DSIF, NELF und P-TEFb (Yamaguchi *et al.*, 1; Yamaguchi *et al.*, 1999a) angesehen. *In vitro* Studien konnten bestätigen, dass NELF und DSIF konzertiert an die RNAP II und RNA binden und so den Transkriptionskomplex in der frühen Elongationsphase arretieren (Yamaguchi *et al.*, 1999a; Renner *et al.*, 2001). Wie die beiden Faktoren dabei die Arretierung der RNAP II bewirken, ist nicht geklärt. Obwohl für das promotorproximale Arretieren beide Faktoren nötig sind, interagieren sie nicht miteinander. Die durch DSIF/NELF verursachte Transkriptionspause wird durch P-TEFb aufgehoben, in dem es NELF, DSIF und die CTD der RNAP II phosphoryliert (Wada *et al.*, 1998b). Dadurch dissoziiert NELF vom Transkriptionskomplex. DSIF bleibt an der RNAP II gebunden und wird zum positiven Transkriptionsfaktor der stimulierend auf die weiterführende Transkription wirkt (Fujinaga *et al.*, 2004). Wie P-TEFb zum Transkriptionskomplex rekrutiert wird, ist nicht für alle Situationen geklärt. Im Fall der Hitzeschockgene in *D. melanogaster* ist der Hitzeschockfaktor (HSF) in die Rekrutierung von P-TEFb zur RNAP II involviert (Lis *et al.*, 2000).

Eine Vielzahl der in *D. melanogaster* und *H. sapiens* untersuchten Gene mit pausierender RNAP II sind für die Entwicklung und Differenzierung des Organismus wichtig. Dieser Mechanismus ermöglicht offenbar eine schnelle Geninduktion bei entsprechendem Stimulus (Wang *et al.*, 2007; Muse *et al.*, 2007; Zeitlinger *et al.*, 2007) und stellt somit einen wichtigen Regulationsmechanismus dar.

Andererseits scheint das promotorproximale Arretieren auch als Kontrollpunkt zu dienen, um die Ausbildung eines transkriptionskompetenten Elongationskomplexes und die Rekrutierung von mRNA-Prozessierungsfaktoren (z.B. *Capping*-Enzym) zur RNAP II zu ermöglichen (Sims *et al.*, 2004; Orphanides und Reinberg, 2002). Die Bindung des *Capping*-Enzyms wirkt z.B. der NELF induzierten Inhibition entgegen (Sims *et al.*, 2004; Orphanides und Reinberg, 2002).

Ein Teil der Gene in *D. melanogaster* waren jedoch bereits mit einer 7mG-Kappe versehen (Rasmussen und Lis, 1993), so dass hier die Transkriptionsblockade unabhängig von der *Capping*-Enzym-Rekrutierung zu sein scheint.

Ein weiteres gut untersuchtes Beispiel einer Blockade in der frühen Elongationsphase ist die Transkription des proviralen HIV-1 Genoms, die erst durch den viralen Transaktivator Tat (*transcriptional transactivator*) aufgehoben werden kann (s. 1.5.1).

1.5.1 Elongationskontrolle am proviralen HIV-1 Promotor

Neben den Hitzeschock-induzierten Genen bei *D. melanogaster* wurden pausierende Transkriptionskomplexe am *long terminal repeat* (LTR) des proviralen HIV-1-Genoms ausführlich studiert. Auffällig bei der Transkription des proviralen HIV-1 Genoms war die Anreicherung von kurzen mRNA-Transkriptionsfragmenten in Abwesenheit des viralen Tat-Proteins (Kao *et al.*, 1987). Ebenso wurde deutlich, dass RNAP II ohne Tat in Promotornähe zum Stillstand kommt (Kao *et al.*, 1987) und Tat hauptsächlich einen Einfluss auf die Elongation, weniger aber auf die Initiation hat (Laspia *et al.*, 1989). Tat ist essentiell für die virale Replikation. Es bindet spezifisch über die argininreiche Domäne (*arginine rich motive*, ARM) an das *cis*-regulatorische RNA Element TAR (*Tat responsive*), das sich am 5' Ende der naszierenden RNA befindet. Die Bindung von Tat hat HIV-1 TAR als ersten *RNA enhancer* identifiziert, da Tat im Gegensatz zu anderen Transkriptionsfaktoren nicht DNA sondern RNA bindet (Berkhout *et al.*, 1989; Calnan *et al.*, 1991; Cordingley *et al.*, 1990). Die 60 nt lange HIV-1 TAR RNA zeigt eine ausgeprägte Sekundärstruktur bestehend aus einer langen doppelsträngige RNA enthaltenden Stammregion, einer Ausbauchung und einer Schleifen-Region (Abb. 3 und 7) (Jones und Peterlin, 1994).

Die Tat-Bindung an die HIV-1 TAR RNA ist jedoch nicht für eine Transaktivierung der Transkription ausreichend, dazu bedarf es zusätzlich P-TEFb, der sich aus der cyclinabhängigen Kinase 7 (CDK7) und Cyclin T1 zusammensetzt (Mancebo *et al.*, 1997; Zhu *et al.*, 1997). Tat interagiert direkt mit der Cyclin T1-Untereinheit und rekrutiert P-TEFb zur Transkriptionsmaschinerie (Wei *et al.*, 1998) und bringt es in räumliche Nähe der Substrate (Abb. 3B).

Nach der Transkriptionsinitiation am induzierbaren HIV-1 Promotor in der LTR-Region des proviralen Genoms binden NELF und DSIF kooperativ an die Transkriptionsmaschinerie und initiieren das Pausieren der RNAP II (Abb. 3A). Dabei binden DSIF und NELF an die RNAP II und NELF zusätzlich über die RRM der E Untereinheit an die HIV-1 TAR RNA (Yamaguchi *et al.*, 2002). Tat bindet im Komplex mit P-TEFb an die HIV-1 TAR RNA, wobei Tat die Blasenregion und Cyclin T1 von P-TEFb die Schleife bindet (Abb. 3B) (Tauben *et al.*, 1999). Studien mit verschiedenen HIV-1 TAR RNA Konstrukten, z.B. mit einer verkürzten Stammregion oder mit einer fehlenden Schleifenregion, sowie mit der verwandten EIAV (*equine infectious anemia virus*) TAR RNA deuten darauf hin, dass die NELF E Bindung in der unteren Stammregion stattfindet (Fujinaga *et al.*, 2004).

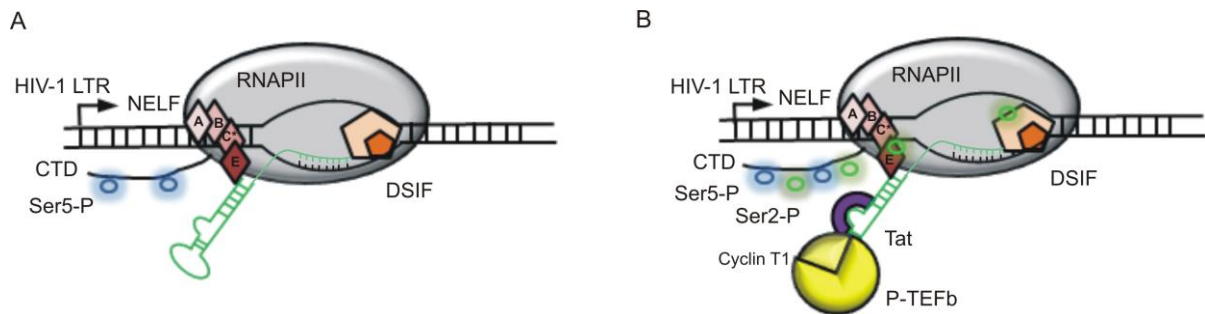


Abb. 3: Schematische Darstellung der Tat induzierten Aufhebung der Transkriptionsblockade

(A) Transkriptionsblockade am HIV-1 LTR Promotor durch NELF und DSIF; beide Elongationsfaktoren binden an die RNAP II, NELF E interagiert zusätzlich mit der HIV-1 TAR RNA; die CTD der RNAP II ist am Ser-5 Rest phosphoryliert (blaue Kreise); die vier Untereinheiten von NELF sind als Rauten dargestellt, dabei entspricht C* der Untereinheit C bzw. der Spleißvariante D; die Untereinheiten von DSIF sind als Fünfeck dargestellt; hSpt4 in orange, hSpt5 in apricot; (B) Das virale Tat-Protein rekrutiert P-TEFb zur arretierten RNAP II und beide binden an die HIV-1 TAR RNA; Tat bindet an die Ausbauchung und P-TEFb an die Schleifenregion der HIV-1 TAR RNA; P-TEFb phosphoryliert die Ser2-Reste der CTD (grüne Kreise), hSpt5 von DSIF und die E-Untereinheit von NELF

Die Bindung von Tat-P-TEFb an die TAR RNA führt zu konformationellen Änderungen innerhalb CDK9 und somit zu dessen konstitutiven Aktivierung (Chen *et al.*, 1999; Ivanov *et al.*, 1999). CDK9 phosphoryliert dann die E Untereinheit von NELF (Fujinaga *et al.*, 2004), sowie die hSpt5 Untereinheit von DSIF (Ivanov *et al.*, 2000) und den Ser2-Rest der CTD (Isel und Karn, 1999). Daraufhin dissoziiert NELF vom Transkriptionskomplex und DSIF wird zum Aktivator der Transkription (Fujinaga *et al.*, 2004; Yamada *et al.*, 2006).

Im Virion liegen keine Kopien von Tat vor, so dass es nach der Infektion einer Zelle erst gebildet werden muss, ehe es als Transkriptionsaktivator agieren kann. Studien zeigten, dass neben Tat noch andere zelluläre Faktoren P-TEFb zum Transkriptionskomplex rekrutieren können. Dazu zählt unter anderem der zelluläre Transkriptionsfaktor *nuclear factor κB* (NFκB), dessen Bindung eine basale Transkription des gesamten HIV-1 Genoms ermöglicht (Barboric *et al.*, 2001). Sobald Tat in der Zelle vorliegt, übernimmt es die P-TEFb-Rekrutierung an den HIV-1-LTR und bewirkt eine Steigerung der Transkriptionsrate und somit eine positive Rückkopplung. Zusätzlich gibt es zahlreiche Hinweise, dass Tat, neben seiner Rolle in der Elongationsphase, zusammen mit P-TEFb bereits während der Initiation eine wichtige Funktion übernimmt und mit zahlreichen Transkriptionsfaktoren und Koaktivatoren interagiert (Brady und Kashanchi, 2005).

1.5.2 Der negative Elongationsfaktor NELF

NELF wurde erstmalig als Faktor identifiziert, der neben DSIF und P-TEFb, für eine 5,6-Dichlor-1-β-D-ribofuranosylbenzimidazol (DRB)-sensitive Transkription in HeLa-Zellkernextrakten nötig ist (Yamaguchi *et al.*, 1999a) (s. 1.5.3). Weitere Analysen zeigten,

dass es ebenso an der Regulation der Transkription von HIV-1 (Ping und Rana, 2001) und am promotorproximalen Arretieren der RNAP II an nicht induzierten Hitzeschockgenen in *D. melanogaster* beteiligt ist (Wu *et al.*, 2003). Dabei hat sich herausgestellt, dass NELF nur im Zusammenspiel mit DSIF seine inhibitorische Wirkung ausübt (Yamaguchi *et al.*, 2002).

NELF besteht aus den vier Untereinheiten A (65 kDa), B (62 kDa), alternativ gespleißtes C bzw. D (60 bzw. 59 kDa) und E (46 kDa), wobei C/D und B den strukturellen Kern bilden. NELF A interagiert mit C/D, NELF E bindet an NELF B (Yamaguchi *et al.*, 1999a; Narita *et al.*, 2003). Es wurden keine homologen Proteine in *Saccharomyces cerevisiae*, *Caenorhabditis elegans* oder *Arabidopsis thaliana* gefunden (Narita *et al.*, 2003), was darauf schließen lässt, dass die Funktion von NELF in diesen Organismen anders geregelt wird. NELF A wird von *WHSC2* codiert, das im Zusammenhang mit dem Wolf-Hirschhorn-Syndrom stehen soll, eine Erbkrankheit mit Deletion im Chromosom 4p16.3, die Geistes- und Entwicklungsstörungen zur Folge hat (Yamaguchi *et al.*, 2001; Wright *et al.*, 1999). Der N-Terminus von NELF A zeigt eine schwache Homologie zum Hepatitis Delta Antigen (HDAg), dem einzigen Genprodukt des Genoms des Virusoids Hepatitis D. Der zu HDAg homologe Bereich ist entscheidend für die Interaktion von NELF mit dem RNAP II/DSIF Komplex (Narita *et al.*, 2003). Und da HDAg selbst an die Klammer-Domäne der RNAP II bindet (Yamaguchi *et al.*, 2007), wird vermutet, dass NELF A in einer ähnlichen Weise mit der RNAP II interagiert. Wahrscheinlich verdrängt HDAg NELF in einer P-TEFb unabhängigen Weise von der RNAP II und unterstützt durch die eigene Bindung die Elongation (Yamaguchi *et al.*, 2001).

NELF B wird auch als COBRA1 (*Cofactor of BRCA1*) bezeichnet, da es in Zwei-Hybrid-Tests als Interaktionspartner von BRCA1 (*breast cancer susceptibility gene 1*) identifiziert wurde (Ye *et al.*, 2001). Inaktivierung von BRCA1, Produkt eines Tumorsuppressorgens, führt zu erhöhtem Brustkrebsrisiko. Darüber hinaus kann es durch direkte Bindung den nukleären Östrogenrezeptor inhibieren. Eine Interaktion von COBRA1 im Komplex mit den anderen NELF Untereinheiten wurde auch für den Östrogenrezeptor (Aiyar *et al.*, 2004) und den Androgenrezeptor gezeigt (Sun *et al.*, 2007).

NELF C/D, das neben NELF B eine strukturelle Funktion in NELF übernimmt, ist auch unter dem Namen Trihydrophobin 1 (TH1) bekannt und ist ein negativer Regulator der A-Raf Kinase im Ras abhängigen MAPK/ERK Signaltransduktionsweg. Es bindet direkt an A-Raf und übt dadurch einen Einfluss auf den Zellzyklusablauf und die Zellproliferation aus (Yin *et al.*, 2002; Liu *et al.*, 2004). Zudem wurde es als neues Zielprotein der E6-AP (*human papilloma virus E6-associated protein*) abhängigen Polyubiquitinierung (Yang *et al.*, 2007b) identifiziert.

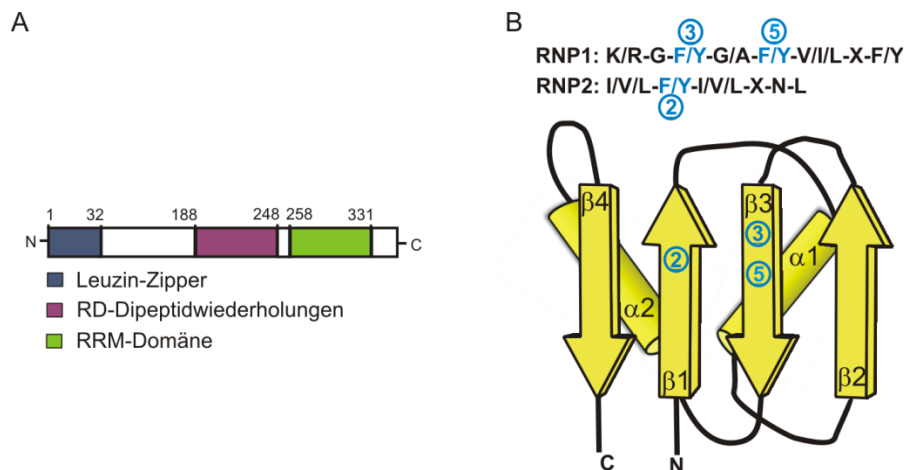


Abb. 4: Schematische Darstellung von NELF E

(A) Domänenarchitektur von NELF E; RRM: *RNA recognition motif* (B) schematische Darstellung einer RRM-Domäne; die Konsensussequenzen von RNP1 und RNP2 sind dargestellt; konservierte aromatische Aminosäuren sind blau hervorgehoben bzw. die Sequenzposition durch blaue Kreise angedeutet (nach Maris *et al.*, 2005)

Die kleinste NELF Untereinheit (E) ist 380 Aminosäuren lang und besitzt einen N-terminalen Leuzin-Zipper über den die Wechselwirkung mit NELF B erfolgt (Abb. 4A). Daran anschließend befindet sich eine RRM-Domäne (*RNA recognition motif*), die direkt an die naszierende RNA des pausierenden Transkriptionskomplexes bindet und essentiell für die inhibierende Wirkung von NELF ist (Yamaguchi *et al.*, 1999a; Narita *et al.*, 2003). RRM Domänen sind Einzelstrang-RNA-Bindedomänen, die ca. 90 Aminosäuren umfassen. Sie besitzen eine charakteristische Faltungstopologie vom Typ $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ und bilden ein viersträngiges antiparalleles Faltblatt mit zwei α -Helices, die auf einer Seite gegen das β -Faltblatt packen (Abb. 4B). Die RNA-Bindung verläuft hauptsächlich über aromatische Aminosäuren, die auf den zwei zentralen β -Strängen lokalisiert sind und zwischen die Basen der RNA interkalieren. Zusätzliche Wechselwirkungen zwischen Phosphatrückgrat und einer dritten aromatischen Aminosäure auf einem zentralen β -Strang tragen zur Stabilität der Bindung bei. Im Bereich der zentralen Faltblattstränge befinden sich konservierte Konsensussequenzen, die als *ribonucleoprotein domain 1* (RNP1) (K/R-G-F/Y-G/A-F/Y-V/I/L-X-F/Y; X = beliebige Aminosäure) und RNP2 (I/V/L-F/Y-I/V/L-X-N-L; X = beliebige Aminosäure) bezeichnet werden und aromatische Reste enthalten, die kritisch für RNA-Bindung sind (Abb. 4B). Jedoch weichen die Bindungsmechanismen bei vielen RRM-Domänen ab, da zum Teil Verbindungsregionen zusätzlich die RNA Bindung modulieren können. Außerdem können die aromatischen Reste in den zentralen Faltblattsträngen variieren (Maris *et al.*, 2005a; Clery *et al.*, 2008).

C-terminal der RRM Domäne schließen sich Arginin-Aspartat Dipeptid Wiederholungen an, deren Funktion jedoch noch unbekannt ist (Abb. 4A). Interaktionen von NELF wurden auch

mit den zellulären Faktoren TFIIS (Palangat und Bensaude, 2005) und TFIIF (Renner *et al.*, 2001) nachgewiesen. Die durch NELF verursachte Inhibition von TFIIS ist wahrscheinlich eine weitere Möglichkeit, die RNAPII länger in der pausierenden Phase zu halten, um die Ausbildung des Elongationskomplexes und die Koordinierung anderer Faktoren zu ermöglichen. Wie die Inhibition durch NELF erfolgt, ist nicht detailliert geklärt, da TFIIS ca. 100 Å vom RNA-Austrittskanal entfernt bindet. So könnte die NELF-Bindung eine konformationelle Änderung in der RNAPII induzieren, welche die Bindung von TFIIS verhindert. Ebenso könnte an die RNAP II gebundenes NELF die Distanz zwischen naszierender RNA, der prognostizierten Bindestelle von NELF E, und TFIIS überspannen. Phosphorylierung von NELF durch P-TEFb hebt TFIIS Inhibition auf und somit auch die Pausierung der RNAP II (Palangat und Bensaude, 2005). Im Gegensatz dazu wurde bei einigen menschlichen Genen ein verringertes Expressionslevel nachgewiesen, wenn kein NELF vorhanden war. In diesem Fall wird davon ausgegangen, dass die durch NELF induzierte Transkriptionsblockade dazu dient, die Promotor- als auch die promotorproximale Region nukleosomenfrei zu halten und somit zugänglich für die RNAP II zu lassen (Gilchrist *et al.*, 2008).

Darüber hinaus konnte in *in vivo* Studien mit HeLa Zellen kürzlich gezeigt werden, dass NELF in subnuklearen Foci lokalisiert ist, in denen NELF an der 3' Prozessierung von replikationsabhängigen Histon mRNAs beteiligt ist. Diese besitzen keinen Poly-A-Schwanz sondern eine spezielle Stamm-Schleife-Struktur. Derartige Foci wurden als *NELF bodies* bezeichnet (Narita *et al.*, 2007). Außerdem wurde nachgewiesen, dass NELF nicht ausschließlich kernlokalisiert ist, sondern zwischen Zellkern und Cytoplasma pendelt. Dabei können NELF A und NELF E frei pendeln, während der Import/Export von NELF B und NELF C nur im gesamten Komplex stattfindet (Yung *et al.*, 2009). Welche Bedeutung das für das Verständnis der Funktion von NELF hat, bleibt abzuwarten. Es deutet jedoch darauf hin, dass NELF nicht nur an der Transkription sondern auch an anderen wichtigen zellulären Ereignissen, wie z.B. dem RNA Export, beteiligt ist.

1.5.3 Der DRB-sensitivity inducing factor DSIF

Der in Eukaryonten konservierte und ubiquitär exprimierte Elongationsfaktor DSIF besteht aus den zwei humanen Homologen der beiden Hefepoteine Spt4 und Spt5, die als hSpt4 und hSpt5 bezeichnet werden. Die *SPT* Gene (*Suppressor of Ty*) wurden ursprünglich in Hefe dadurch identifiziert, dass die durch Translokation des Hefetransposons Ty in die Promotorregion verursachte Transkriptionsreduktion durch Mutation in *SPT*-Genen aufgehoben wird (Simchen *et al.*, 1984). Ebenso konnte anhand von Hefemutanten gezeigt werden, dass zwar Spt5, aber nicht Spt4 für das Zellwachstum essentiell ist (Swanson *et al.*, 1991; Swanson und Winston, 1992). DSIF wurde durch Aktivitätsuntersuchungen mit

humanen HeLa Zellextrakten als Faktor identifiziert, der in Gegenwart des Nucleosidanalogon DRB die Transkription inhibieren kann. *Immunodepletion* von DSIF aus dem Zellextrakt führte zu einer DRB unempfindlichen Transkription, während Zugabe von DSIF in ein DSIF-freies System die Sensitivität gegenüber DRB wiederherstellt (Wada *et al.*, 1998a).

Weitere Studien zeigten, dass DSIF nicht ausschließlich inhibitorisch auf die Transkription wirkt, sondern im Gegenteil zusammen mit anderen Faktoren RNAP II-Aktivität stimuliert. So ist DSIF auch nach der durch P-TEFb initiierten Transkriptionsaktivierung mit der RNAP II assoziiert und agiert nun als Transkriptionsaktivator (Ping und Rana, 2001; Yamada *et al.*, 2006). Außerdem ist DSIF in der Lage, frühzeitige Termination und eine Arretierung der RNAP II an intrinsischen DNA-Stopps (*pause sites*) zu unterdrücken (Bourgeois *et al.*, 2002). Außerdem wurden DSIF Interaktionen noch für andere Transkriptionsfaktoren gefunden. Dies impliziert, dass die Rolle von DSIF während der Transkription weitreichender ist, als nur das promotorproximale Arretieren. So wurden Interaktionen mit dem humanen Transkriptionsfaktor hSpt6 (Endoh *et al.*, 2004) und in *S. cerevisiae* mit dem Chromatinremodellierungsfaktor Chd1 (Simic *et al.*, 2003) nachgewiesen. Ebenso ist DSIF an der Kopplung von Transkription und mRNA-Prozessierung beteiligt, da es sowohl mit dem *Capping*-Enzym (Wen und Shatkin, 1999) als auch mit dem Paf1-Komplex in Hefe interagiert, der an der 3' Prozessierung der prä-mRNA beteiligt ist (Squazzo *et al.*, 2002). Rekrutierung des *Capping*-Enzyms durch DSIF an die Transkriptionsmaschinerie führt zur Aufhebung der von NELF induzierten Transkriptionshemmung (Sims *et al.*, 2004; Orphanides und Reinberg, 2002). Durch genomweite Promotoranalysen wurde auch eine promotorabhängige Regulation von DSIF an NFκB-Zielgenen nachgewiesen. Während bei TATA-losen Promotoren die Inhibition durch DSIF auch nach NFκB-Induktion besteht, wird bei TATA-haltigen-Promotoren die DSIF- Wechselwirkung aufgehoben und die Transkription positiv durch P-TEFb reguliert (Ainbinder *et al.*, 2004; Amir-Zilberstein *et al.*, 2007; Amir-Zilberstein *et al.*, 2008).

Die Aminosäuresequenz von hSpt4, mit 120 Aminosäuren die kleine Untereinheit von DSIF, lässt darauf schließen, dass es einen N-terminalen Zinkfinger bestehend aus vier Cysteinresten enthält (Abb. 5A) (Chiang *et al.*, 1996). Die Funktion von hSpt4 im DSIF Komplex ist noch nicht geklärt, jedoch ist bekannt, dass es an die NusG N-terminale-Homologie Domäne (NGN) von hSpt5 bindet (Ivanov *et al.*, 2000; Yamaguchi *et al.*, 1999b). Es konnte zwar gezeigt werden, dass Spt4 in Hefe als positiver Elongationsfaktor funktioniert und z.B. den negativen Effekten des Chromatinremodellierungsfaktors Isw1p (Morillon *et al.*, 2003) entgegenwirkt, jedoch wurde die Rolle von Spt5 in diesem experimentellen Ansatz nicht untersucht. So könnte Spt4 eine rein strukturelle Funktion übernehmen, in dem es Spt5

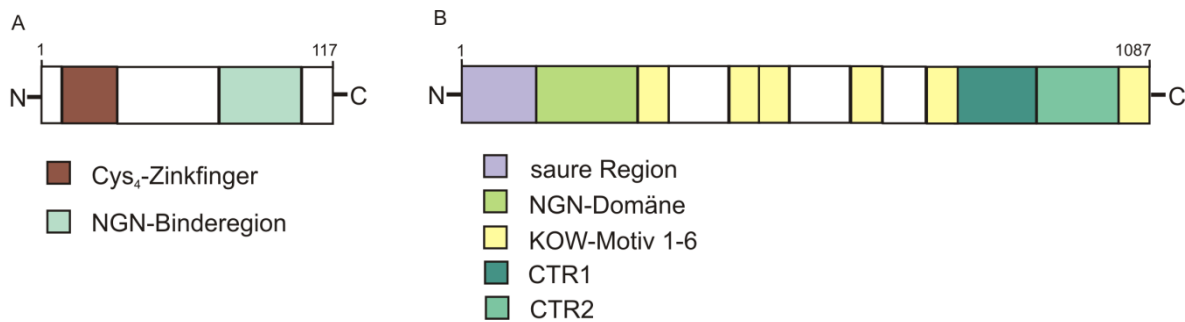


Abb. 5: Domänenstruktur von hSpt4 und hSpt5

(A) hSpt4 verfügt über einen N-terminalen Zinkfinger vom Typ Cys₄ und über eine N-terminale-NusG-Homologie Domäne (NGN)-Interaktionsregion; (B) hSpt5 besitzt N-terminal eine saure Region der sich die NGN-Domäne anschließt; hSpt4 Bindung erfolgt über die NGN-Domäne, der sich mehrere KOW-Motive anschließen (KOW: nach Kyrpides, Orphanides und Woese); zusätzlich verfügt hSpt5 über zwei C-terminale Wiederholungen (*C-terminal repeats*) (CTR1 bzw. CTR2) (nach: Wada *et al.*, 1998a und Ponting, 2002)

stabilisiert, zumal Interaktionen mit anderen Interaktionspartnern oder RNAP II nur für Spt5 beschrieben wurden.

hSpt5, mit 1087 Aminosäuren die große Untereinheit von DSIF, besitzt N-terminal eine saure Region, die die ersten 100 Aminosäuren umfasst, gefolgt von der NGN-Domäne, der Interaktionsdomäne von hSpt4 (Abb. 5B). C-terminal der NGN-Domäne befinden sich 4-6 KOW-Motive (nach Kyrpides, Orphanides und Woese, (Kyrpides *et al.*, 1996)), die ebenso charakteristisch für Proteine der *N utilization substance G* (NusG)-Proteinfamilie und für die Proteine RL24, 26 und 27 der großen ribosomalen Untereinheit sind (Kyrpides *et al.*, 1996). Die 27 Aminosäuren umfassenden KOW-Motive besitzen einen konservierten Glycinrest an elfter Position und beinhalten alternierend hydrophile und hydrophobe Abschnitte (Kyrpides *et al.*, 1996). KOW-Motive gelten, auch im Bezug zu den ribosomalen Untereinheiten, als RNA-Bindedomänen (Kyrpides *et al.*, 1996). Die Struktur von der *E. coli* NusG KOW Domäne zeigt eine hohe Ähnlichkeit zu TUDOR-Domänen (Mooney *et al.*, 2009). Zwar binden TUDOR Domänen in Eukaryonten methylierte Lysine von Histonen (Corsini und Sattler, 2007), aber auch RNA Bindung wurde nachgewiesen (Kyrpides *et al.*, 1996). Die Strukturen der drei KOW-Motive KOW2 (PDB: 2E6Z), KOW3 (PDB: 2DO3) und KOW5 (PDB: 2E70) sind bekannt, die genaue Funktion der hSpt5 KOW-Motive jedoch nicht.

C-terminal der zentralen fünf KOW-Motive von hSpt5 befinden sich Sequenzwiederholungen, die als *C-terminal repeat 1* (CTR1) bzw. CTR2 bezeichnet werden (Abb. 5B). Diese weisen zahlreiche Serin- und Threoninreste in Pentapeptidwiederholungen mit der Konsensussequenz G-S-R/Q-T-P auf (Ivanov *et al.*, 2000; Stachora *et al.*, 1997) und sind wichtig für den Übergang von DSIF aus dem inhibierenden Zustand hin zum Elongationsaktivator (Ivanov *et al.*, 2000). Wie genau die Phosphorylierung die Funktionsweise von DSIF ändert, ist nicht bekannt. So könnte eine veränderte

Wechselwirkung mit der RNAP II oder aber mit anderen Faktoren die Folge sein (Ivanov *et al.*, 2000). Ähnlich wie bei der CTD der RNAP II sollen diese Phosphorylierungsstellen wichtig für die Rekrutierung von Faktoren, wie z.B. die Peptidyl-prolyl Isomerase Pin1, sein, die gezielt an phosphoryliertes hSpt5 binden (Lavoie *et al.*, 2001). Zudem könnten diese Phosphorylierungen eine wichtige Rolle bei der mRNA Reifung spielen, da DSIF Einfluss auf *Capping* (Wen und Shatkin, 1999), 3' Prozessierung (Squazzo *et al.*, 2002) und Spleißen (Xiao *et al.*, 2005; Lindstrom *et al.*, 2003) hat. Außerdem können Argininreste im Bereich der KOW-Motive durch die Methyltransferasen PRMT5 und PRMT1 methyliert werden, was die Bindung an die RNAP II schwächen soll (Kwak *et al.*, 2003).

Die NGN-Domäne zeigt Homologien zur N-terminalen Domäne des bakteriellen Transkriptionsfaktor NusG. Dieser ist zusammen mit anderen Nus-Faktoren an der λ N abhängigen Antitermination, sowie der Rho-abhängigen Termination der Transkription beteiligt (Burns *et al.*, 1998). Ebenso wurde eine Funktion bei der Steigerung der Elongationsrate für NusG nachgewiesen (Burova *et al.*, 1995). Jedoch scheint die Rolle von NusG speziesspezifisch zu sein, da es im Gegensatz zu *E. coli* bei *Bacillus subtilis* ein Arretieren der RNAP eher stimuliert als unterdrückt. Auch ist NusG in *B. subtilis* nicht essentiell, in *E. coli* hingegen schon (Downing *et al.*, 1990; Ingham *et al.*, 1999). *E. coli* NusG ist modular aufgebaut und besitzt neben der N-terminalen Domäne noch die C-terminale Domäne, welche, ähnlich wie hSpt5, ein KOW-Motiv beinhaltet. Man nimmt an, dass die N-terminale Domäne mit den β' -*clamp* Helices der RNAP interagiert (Nickels, 2009), während die CTD an Rho, andere Nus-Faktoren oder RNA bindet (Mooney *et al.*, 2009). NusG, hochgradig konserviert in Bakterien und Archaeen, wird somit als evolutionärer Vorfahre von hSpt5 angesehen.

2. Zielsetzung

Das Wissen über die detaillierten Abläufe der Transkriptionsblockade in der frühen Elongationsphase ist immer noch unvollständig. Gerade die Entdeckung, dass die Regulation der Transkription weit häufiger auf der Elongationsebene als bei der Initiation erfolgt, hat das Bild von den zellulären Vorgängen verändert. Die Tatsache, dass vor allem bei Genen, die eine entscheidende Rolle bei Zelldifferenzierung oder Entwicklung spielen, Postinitiations-Transkriptionskomplexe in Promotornähe lokalisiert sind, gibt neue Einblicke, wie in Zellen entwicklungs- und zelltypische Signale schnell umgesetzt werden können. Wichtig ist in diesem Hinblick auch die Aufklärung der molekularen Vorgänge bei der Inhibition der Transkription des proviralen HIV-1 Genoms in der frühen Elongationsphase durch zelluläre Faktoren. Von entscheidender Rolle sind dabei die beiden humanen Elongationsfaktoren NELF und DSIF, die zusammen die RNAP II reversibel inhibieren. So sind zwar zahlreiche Faktoren bekannt, die mit DSIF und NELF interagieren, aber Informationen, wie die Interaktionen zwischen den einzelnen Faktoren und mit der RNAP II zeitlich und räumlich koordiniert werden, fehlen gänzlich. Diese Zusammenhänge könnten strukturelle Daten liefern, die für DSIF und NELF bisher fehlen. Ebenso wären damit Einblicke in die intermolekularen Wechselwirkungen möglich, was zu einem besseren Verständnis der Regulation der Transkription führt.

Daher sollten in der vorliegenden Arbeit diese beiden Faktoren strukturell charakterisiert werden, um ein besseres Verständnis über deren Funktion zu erlangen. Von besonderem Interesse war dabei die Analyse der NELF E Untereinheit, die über die RRM-Domäne mit RNA interagiert. Da die RNA-Bindung essentiell für die Blockade der Transkription des HIV-1 Genoms ist, sollte die strukturelle Analyse der RNA Bindung Hinweise auf den Mechanismus der Bindung und inhibitorischen Wirkung geben. Mittels NMR-Titrationsexperimenten und Fluoreszenzgleichgewichtstitrationen sollte untersucht werden, ob NELF E RRM an die einzel- oder doppelsträngige HIV-1 TAR RNA bindet und ob diese Interaktion sequenzspezifisch ist. Die daraus resultierenden Ergebnisse könnten neue Ansatzpunkte für eine mögliche HIV-1 Therapie liefern.

Des Weiteren galt es in dieser Arbeit den humanen Elongationsfaktor DSIF, der neben NELF an der Transkriptionsblockade in der frühen Elongationsphase beteiligt ist, strukturell zu analysieren. Dabei sollte die Struktur der kleinen DSIF Untereinheit hSpt4 bestimmt und vor allem deren Interaktion mit der großen Untereinheit hSpt5 näher untersucht werden.

3. Zusammenfassung und Diskussion der Ergebnisse

3.1 Der negative Elongationsfaktor NELF

3.1.1 Struktur der freien und RNA-gebundenen Form von NELF E RRM

Der humane negative Elongationsfaktor NELF übernimmt zusammen mit DSIF eine Schlüsselrolle in der Regulation des promotorproximalen Arretierens der Transkriptionsmaschinerie (siehe 1.5). Eine Blockade in der frühen Elongationsphase tritt auch bei der Transkription des proviralen HIV-1 Genoms auf, die erst durch das virale Tat-Protein aufgehoben werden kann (siehe 1.5.1). Ein entscheidendes Element der Inhibition der RNAP II bei der Transkription des proviralen HIV-1 Genoms spielt die naszierende RNA, die eine definierte Sekundärstruktur ausbildet. Die sogenannte TAR-RNA umfasst die ersten 60 nt der naszierenden RNA und besitzt Bindungsstellen für NELF E, Tat und CyclinT1 von P-TEFb. So bindet Tat an die Ausbauchung von TAR und Cyclin T1 die TAR Schleifenregion (siehe 1.5.1, Abb. 3B). Die Bindungsstelle von NELF E wurde bisher jedoch noch nicht charakterisiert. NELF E verfügt über eine RRM-Domäne, über die RNA-Bindung erfolgen soll (Fujinaga *et al.*, 2004). Durch Immunopräzipitationsversuche konnte eindeutig gezeigt werden, dass NELF an einen bereits existierenden RNAP II-DSIF-Komplex und an naszierende RNA binden kann (Yamaguchi *et al.*, 2002).

Die Strukturbestimmung von Proteinen mittels der magnetischen Kernresonanz (*nuclear magnetic resonance*, NMR) wird mit steigender Molekülmasse erschwert, da die damit effizientere transversale Relaxation zu Signalverbreiterung und schließlich Signalverlust führt. Begleitet wird dies von zunehmender Signalüberlagerung, die eine eindeutige Zuordnung der Signale verhindert (Clare und Gronenborn, 1998). Proteine > 30 kDa sind nur mittels Perdeuterierung oder selektiver Markierung einzelner Aminosäuren zugänglich (Gardner und Kay, 1998). Aus diesem Grund wurde nur die RRM-Domäne (13 kDa) des 43 kDa-Proteins NELF E für die Strukturbestimmung und für Interaktionsstudien mit RNA durch NMR gewählt.

Zur Strukturaufklärung und für RNA-Bindungsstudien wurde das synthetische Gen der RRM-Domäne von NELF E in *E. coli* BL21(DE3) verwendet, um isopenmarkiertes Protein zu gewinnen. Ein zunächst gewähltes Konstrukt, das nur die Aminosäurereste 258-331 umfasste, erwies sich als unlöslich. Durch Verlängerung des N- und C-Terminus um 13 bzw. 12 Reste konnte dieses Problem umgangen und lösliches Protein gewonnen werden (Einzelarbeit A, Abb. 1). Wie in den Einzelveröffentlichungen A und B bezieht sich die Nummerierung der Aminosäuren der NMR-Struktur in dieser Arbeit auf das verwendete Konstrukt und nicht auf die Sequenzposition des Gesamtproteins.

Der N-terminale Bereich M1-R34 und die C-terminalen Reste L114-S121 erwiesen sich als hochflexibel und wurden bei den Strukturberechnungen nicht berücksichtigt, so dass nur für den Bereich A35-M113 die Struktur in Lösung bestimmt wurde (PDB: 2BZ2).

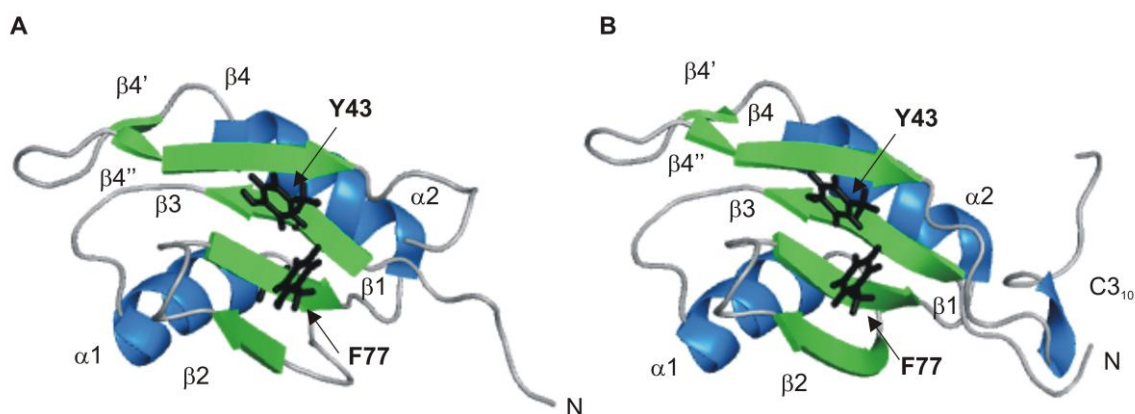


Abb. 6: Struktur des freien und gebundenen NELF E RRM

(A) Struktur von NELF E RRM im ungebundenen Zustand (PDB: 2Z2B); Sekundärstrukturelemente sind farblich dargestellt; blau: helikale Bereiche, grün: β -Faltblattbereiche, grau: unstrukturierte Bereiche; an der RNA-Bindung beteiligte, aromatische Aminosäuren der hochkonservierten Konsensusmotive RNP1 (F77) und RNP2 (Y43) sind als schwarze Stäbchen dargestellt; (B) Struktur von NELF E RRM im gebundenen Zustand (PDB: 2JX2); Legende siehe (A)

NELF E RRM zeigt die typische Faltungstopologie $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ einer RRM-Domäne (Maris *et al.*, 2005) (siehe 1.5.2, Abb. 4). Es bildet ein viersträngiges antiparalleles β -Faltblatt, das hauptsächlich hydrophobe Aminosäuren enthält. Gegen den dadurch entstehenden hydrophoben Kern packen zwei α -Helices (α_1 - α_2), die nahezu senkrecht zueinander stehen (interhelikaler Winkel: 114.3°) (Abb. 6). Die Schleifenregion zwischen α_2 und β_4 bildet zusätzlich ein kurzes zweisträngiges, antiparalleles β -Faltblatt (β_4' - β_4'') aus, das zwar nicht zur konservierten RRM-Struktur gehört, aber auch in anderen RRM Domänen verbreitet ist (Xu *et al.*, 1997; Kielkopf *et al.*, 2001).

Hinsichtlich der RNA-Bindung zeigen RRM-Domänen eine enorme Vielseitigkeit und Anpassungsfähigkeit, um eine höhere Affinität bzw. Sequenz- oder Sekundärstrukturspezifität zu erhalten. Allen gemeinsam ist jedoch, dass die RNA auf der den Helices abgewandten Seite des β -Faltblattes orientiert wird und hochkonservierte, aromatische Seitenketten der beiden zentralen β -Stränge zwischen die RNA-Basen packen (s. 1.5.2). Diese aromatischen Seitenketten sind Bestandteil der beiden konservierten RNP-Motive. Im Falle von NELF E RRM sind F77 von RNP1 und Y43 von RNP2 (Abb. 6) an entsprechender Stelle lokalisiert.

NMR-Spektroskopie eignet sich besonders für Protein-Protein oder Protein-Nukleinsäure Bindungsstudien. Dabei findet vor allem die Beobachtung der Änderung der chemischen

Verschiebung (*chemical shift perturbation*, CSP) Verwendung. Die chemische Verschiebung eines NMR-aktiven Kerns ist abhängig von der lokalen chemischen Umgebung, wobei das Lösungsmittel, Wasserstoffbrückenbindungen oder van der Waals-Wechselwirkungen sowie Bindungswinkel u.a. einen Einfluss darauf ausüben (Skinner *et al.*, 2008). Eine Bindung zwischen Protein und Interaktionspartner führt zu Veränderungen der lokalen Umgebung und somit der chemischen Verschiebungen von Resten, die unmittelbar an der Bindung beteiligt sind. Aber auch jene Reste, die durch Ligandenbindung konformationelle Änderungen erfahren, z.B. durch Umorientierung von Seitenketten, oder die durch Lösungsmittelleffekte beeinflusst werden, zeigen einen CSP Effekt (McCoy und Wyss, 2000). Markierung eines Bindungspartners, in diesem Fall NELF E RRM, mit ^{15}N als zusätzlichen NMR-aktiven Kern und Messung von mehrdimensionalen heteronuklearen Korrelationsspektren (^1H - ^{15}N -HSQC), ermöglicht die direkte Zuordnung der beteiligten Reste und so die Kartierung der Bindungsoberfläche.

Die RNA-Bindungsregion auf NELF E RRM konnte somit durch NMR-Titrationsexperimente mit der HIV-1 TAR RNA bestimmt werden (Einzelarbeit A, Abb. 4). Große Verschiebungen der Amidprotonenresonanzen wurden vor allem für Reste des zentralen β -Faltblattes beobachtet (N41, L42, Y45, C75, F77). Dies bestätigt, dass, ähnlich zu anderen RRM-Domänen, das β -Faltblatt die Bindungsfläche für die RNA bereitstellt und die hochkonservierten, aromatischen Reste von RNP1 (F77) und RNP2 (Y43) der zentralen β -Faltblattstränge zwischen die Basen interkalieren. Zusätzlich sind auch Reste in der Schleifenregion zwischen $\beta 3$ und $\alpha 2$ (E81, K82 und K83), der N-terminale Bereich von $\alpha 2$ und der C-Terminus (M113, L114, D115, A116) involviert (Einzelarbeit A, Abb. 4). Diese betroffenen Reste liegen entfernt von der RNA-Binderegion, was eine direkte RNA-Interaktion unwahrscheinlich macht. So kann davon ausgegangen werden, dass RNA-Bindung zu weitreichenden konformationellen Änderungen in NELF E RRM führt, wovon hauptsächlich die benannte Schleifenregion betroffen ist. Die Darstellung der RNA-Bindungsoberfläche (Einzelarbeit A, Abb. 5) verdeutlicht, dass sowohl die zentralen β -Faltblattstränge, als auch Reste der äußeren Bereiche maßgeblich an der RNA-Bindung beteiligt sind. NMR-Bindungsstudien von NELF E RRM wurden mit verschiedenen 9-10 nt langen RNA-Oligonukleotiden durchgeführt und zeigten, dass stets die gleichen Reste Änderungen in den chemischen Verschiebungen erfuhren. Das deutet darauf hin, dass RNA-Bindung strukturelle Veränderungen in NELF E RRM induziert, die unabhängig von der RNA-Sequenz sind.

Die Strukturbestimmung der gebundenen Form von NELF E RRM (Einzelarbeit B) zeigte, dass sich diese von der freien Form im Bereich 39-107 nur unwesentlich unterscheidet. Jedoch wird der hochflexible C-Terminus in der gebundenen Form durch hydrophobe

Wechselwirkungen stabilisiert und es bildet sich eine 3_{10} -Helix am C-Terminus aus (C3₁₀, Abb. 6B), die räumlich in der Nähe der $\beta 3$ - $\alpha 2$ -Schleife lokalisiert ist. Diese strukturelle Umorientierung unterstützt die These von einem *induced fit* Mechanismus bei RNA-Bindung und macht eine direkte RNA-Interaktion der 3_{10} -Helix wahrscheinlich. Ähnliches wurde auch bei anderen RRM-Domänen beobachtet. So zeigt die RRM-Domäne des U1A-Proteins, das ein Bestandteil des U1 *small ribonucleoprotein particle* (U1snRNP) ist, bereits in der freien Form eine α -Helix am C-Terminus. Jedoch scheint sie in der ungebundenen Form multiple Konformationen annehmen zu können, da sich die Lage der Helix in der NMR-Struktur von der Kristallstruktur unterscheidet (Rupert *et al.*, 2003). In der gebundenen Form ist diese Helix C von U1snRNP jedoch direkt in die RNA-Interaktion involviert (Allain *et al.*, 1996).

3.1.2 Bestimmung der Bindungsaffinität zwischen NELF E RRM und RNA

CSP-Messungen dienen nicht nur der Kartierung der Bindungsoberfläche eines Interaktionspartners, sondern können auch zur Bestimmung von Bindungsaffinitäten herangezogen werden. Eine Bestimmung der Dissoziationskonstante (K_D) mittels CSP eignet sich vor allem, wenn der Austausch zwischen freiem und gebundenem Zustand schnell im Vergleich zur NMR-Zeitskala stattfindet, da die Beobachtung der gemittelten chemischen Verschiebung im NMR-Spektrum ein schrittweises Verfolgen der CSP einzelner Signale während einer Serie von Titrationsexperimenten ermöglicht. Jedoch können quantitative K_D -Werte nur im Konzentrationsbereich des beobachteten Interaktionspartners bestimmt werden. Aufgrund der relativen Unempfindlichkeit der Messmethode und der daher erforderlichen hohen Konzentrationen ($\geq 10^{-5}$ M) können nur schwache Wechselwirkungen analysiert werden, die einen höheren K_D -Wert als ca. 3 μ M aufweisen. (Zuiderweg, 2002). Für die HIV-1 TAR RNA-Fragmente TAR1-10, TAR6-15 und TAR39-48 konnten K_D -Werte von 21 μ M, 35 μ M und 55 μ M ermittelt werden (Einzelarbeit B, s.a. Abb. 7). Diese liegen alle in einem ähnlichen Bereich, doch im Vergleich mit anderen RRM-Domänen zeigt NELF E RRM eine geringe Affinität für RNA. So wurden K_D -Werte im nanomolaren Bereich und niedriger beschrieben (Maris *et al.*, 2005). Die beobachteten CSPs zeigten jedoch Unterschiede im Grad der Verschiebung (Einzelarbeit B, Abb. 3 und Abbildung des Anhangs), was andeutet, dass die verschiedenen RNA-Oligonukleotide zwar in ähnlicher Weise gebunden werden, doch scheinen lokale Unterschiede an der Bindungsoberfläche die Bindungsaffinität in gewissem Maße zu beeinflussen.

Interaktionsstudien zwischen NELF E RRM und RNA wurden während der NMR Experimente ausschließlich mit kurzen, einzelsträngigen RNA Oligonukleotiden durchgeführt. Generell binden RRM-Domänen einzelsträngige RNA oder auch DNA. (Maris *et al.*, 2005). Weiterhin existieren RRM-Domänen, die RNA-Stammschleifen-Strukturen erkennen und

spezifisch binden. Jedoch findet in diesem Fall die Interaktion in der einzelsträngigen Schleifenregion und nicht in der Stammregion statt (Oubridge *et al.*, 1994; Allain *et al.*, 1996). Für die Interaktion zwischen NELF und HIV-1 TAR RNA konnte in Gelretardationsexperimenten jedoch gezeigt werden, dass NELF E die untere doppelsträngige Stammregion der HIV-1 TAR RNA bindet, da eine mutierte HIV-1 TAR RNA mit deletierter Stamm- bzw. Blasenregion noch von NELF E gebunden wurde (Fujinaga *et al.*, 2004). Auch die Fusion der HIV-1 TAR RNA Stammregion mit einer anderen Schleifen-Region beeinträchtigt die NELF E Bindung nicht. Außerdem bindet NELF E nicht die TAR RNA von EIAV (*equine infectious anemia virus*), die eine ähnliche Sekundärstruktur ausbildet wie die HIV-1 TAR RNA, aber über eine verkürzte Stammregion verfügt. Ebenso konnte ausgeschlossen werden, dass NELF an die Schleifenregion bzw. Ausbauchung der HIV-1-TAR RNA bindet, da es mit einem Komplex aus Tat, CyclinT1 und HIV-1 TAR interagiert (Fujinaga *et al.*, 2004). NELF E zeigte somit bei diesen Analysen andere Eigenschaften, als es für ein RRM-haltiges Protein erwartet worden wäre.

Deswegen wurden in dieser Arbeit Fluoreszenzgleichgewichtstitrationen mit am 5' Ende fluoreszenzmarkierten (Carboxyfluorescein, 6-FAM) RNA-Oligonukleotiden aus der HIV-1 TAR-Region durchgeführt. Es wurden der 3' und der 5' Einzelstrang sowie die doppelsträngige HIV-1 TAR RNA-Stammregion, die für die Analysen verwendet wurden, markiert (Abb. 7 und Einzelarbeit A, Abb. 5). Damit sollten die Bindungsaffinitäten von NELF E RRM an die unterschiedlichen Substrate bestimmt werden, um Hinweise zu erhalten, ob die Einzel- oder Doppelstrang-RNA präferentiell gebunden wird. Die Fluoreszenzspektroskopie ermöglicht aufgrund der hohen Sensitivität auch die Detektion von hochaffinen Interaktionen im nanomolaren Bereich. Die Verwendung fluoreszenzmarkierter RNA umgeht das Problem geringer intrinsischer Fluoreszenzintensität seitens NELF E RRM aufgrund fehlender Tryptophanreste und erhöht zudem die Sensitivität.

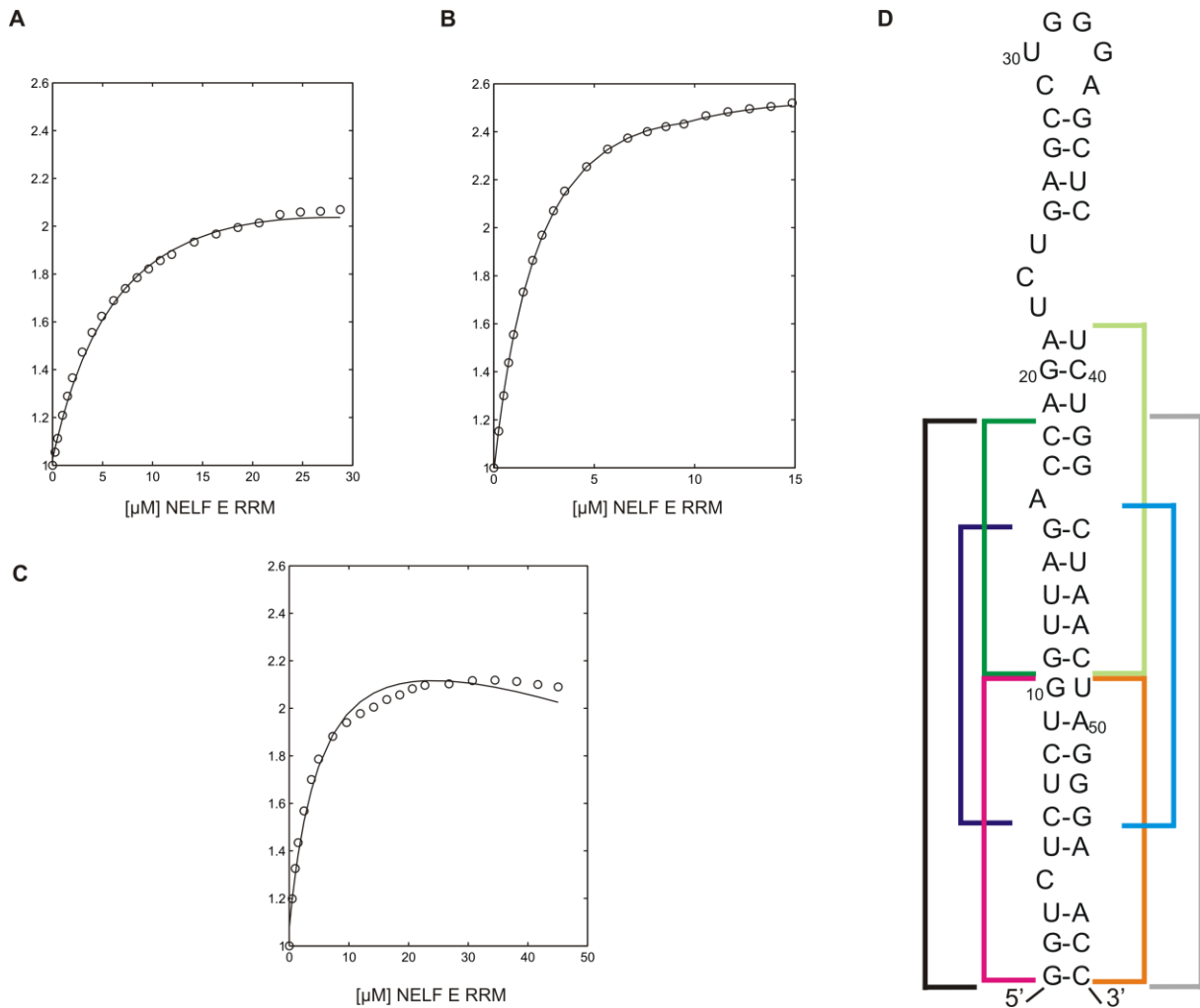


Abb. 7: Fluoreszenzgleichgewichtstitration von NELF E RRM mit 6-FAM-RNA-Oligonukleotiden
 (A)-(C) Fluoreszenztitrationskurven mit HIV-1 TAR(1-18) (A), HIV-1 TAR(42-57) (B) und Doppelstrang-HIV-1 TAR(1-18+42-57) (C) (aus Einzelarbeit A, Abb.5); (D) Sekundärstruktur von HIV-1 TAR RNA (1-57); Sequenzen der RNA Oligonukleotide, die für NMR- und Fluoreszenzgleichgewichtstitrationen verwendet wurden, sind durch Klammern gekennzeichnet; schwarz: TAR(1-18); grau: TAR(42-57); blau: TAR(6-15); rot: TAR(1-10); grün: TAR(11-18); hellgrün: TAR(39-48); hellblau: TAR(44-53); orange: TAR(49-57) (Einzelarbeit B)

NELF E RRM zeigt für beide Einzelstränge eine ähnliche Affinität mit Dissoziationskonstanten im niederen mikromolaren Bereich (TAR(1-18): $8,2 \mu\text{M}$; TAR(42-57): $2,6 \mu\text{M}$) (Abb. 7 und Einzelarbeit A). Die Bestimmung der K_D -Werte konnte unter der Annahme eines Zwei-Zustandsmodells erfolgen. Die mittels Fluoreszenzgleichgewichtstitrationen bestimmten K_D -Werte sind somit um einen Faktor 10 kleiner als jene der NMR-Titrations (Einzelarbeit B), bestätigen aber, dass die Bindung von NELF E RRM an RNA im Vergleich zu anderen RRM-Domänen eher schwach ist. Die Experimente mit der doppelsträngigen HIV-1 TAR Stammregion (Einzelarbeit A; Abb. 5C), konnten mit einem einfachen Zwei-Zustandsmodell nicht hinreichend erklärt werden.

Bindung an das doppelsträngige RNA-Substrat findet zwar statt, nur kann mit diesem Modell nicht nachvollzogen werden, wie die Bindung erfolgt. Atmen, d.h. das dynamische Öffnen und Schließen der Wasserstoffbrückenbindungen zwischen den Basen an den Doppelstrangenden, könnte zu intermediären Einzelstrangbereichen führen, an die NELF E RRM bindet und die Strangdissoziation vorantreibt. NELF-Bindung könnte ebenso partielle Einzelstrangbereiche induzieren. Außerdem zeigten die Experimente mit den HIV-1 TAR RNA-Einzelsträngen der Stammregion keine eindeutige Präferenz bzw. Sequenzspezifität. Eine Erklärung wäre, dass NELF E RRM, auch wegen seiner im Vergleich mit anderen RRM-Domänen geringen Affinität nicht nur schwach mit einem Sequenzabschnitt interagiert, sondern auch dynamisch verschiedene Bereiche bindet und am RNA-Strang entlang wandert. Ein derartiges Verhalten wurde für RRM-Domänen des *polypyrimidine tract-binding protein* (PTB) an CU-Sequenzwiederholungen beobachtet (Auweter *et al.*, 2007).

Ferner wurden Fluoreszenzgleichgewichtstitrationen von NELF E RRM mit den kürzeren 9-10 nt langen RNA Oligonukleotiden der HIV-1 TAR RNA durchgeführt, die auch bei den NMR-Titrationen Einsatz fanden, um eine mögliche sequenzspezifische Bindung näher zu untersuchen. Die hierbei ermittelten K_D -Werte für NELF E RRM sind mit denen aus den Bindungsstudien mit den 16-18 nt RNA-Oligonukleotiden vergleichbar und deuten auf eine sequenzunabhängige Bindung hin. Wie zuvor erwähnt, binden viele RRM-Domänen im nanomolaren Bereich. NELF E RRM zeigt jedoch eine weitaus geringere Affinität für RNA. Um zu untersuchen, ob nicht die RRM-Domäne allein, sondern die gesamte NELF E Untereinheit wichtig für Sequenz- oder Strukturspezifität ist, wurden Fluoreszenzgleichgewichtstitrationen mit der NELF E Untereinheit sowohl mit den kurzen als auch mit den längeren RNA-Oligonukleotiden durchgeführt. Eine sequenzspezifische Interaktion konnte aber auch hierbei nicht beobachtet werden (Daten nicht gezeigt).

3.2 Der DRB-sensitivity inducing factor DSIF

DSIF übernimmt zusammen mit NELF eine Schlüsselfunktion beim promotorproximalen Arretieren und im Speziellen bei der Transkriptionshemmung des proviralen HIV-1 Genoms (s. 1.5). Um ein detaillierteres Funktionsverständnis zu erlangen, sollte die bisher ungelöste Struktur der kleinen Untereinheit hSpt4 von DSIF bestimmt werden. Darüber hinaus stand die Charakterisierung der Interaktion von hSpt4 mit der hSpt5-NGN-Domäne, an die hSpt4 im DSIF-Komplex bindet, im Fokus dieser Arbeit.

3.2.1 Reinigung des rekombinanten hSpt4

Zur Expression in *E. coli* BL21(DE3) wurde ein synthetisches Gen, das für die kleine DSIF Untereinheit hSpt4 codiert, verwendet. hSpt4 per se erwies sich als unlöslich. Erst die Fusion mit dem Löslichkeitsanhang GB1 (*immunoglobulin binding domain B1 of streptococcal protein G1*) am N-Terminus erhöhte die Löslichkeit von hSpt4 drastisch und ermöglichte die Reinigung in entsprechenden Mengen. Zusätzlich verfügte GB1 N-terminal über einen Hexahistidinanhang und C-terminal über eine TEV (*Tobacco Etch Virus*)-Protease Schnittstelle, die eine Trennung von hSpt4 und dem Löslichkeitsanhang ermöglichte (Einzelarbeit C).

hSpt4 besitzt einen N-terminalen Zinkfinger, wobei vier Cystein-Reste ein Zinkion koordinieren (Chiang *et al.*, 1996). Durch optische Emissionsspektroskopie mit induktiv gekoppeltem Plasma (*inductively coupled plasma optical emission spectroscopy*, ICP-OES), das die quantitative Bestimmung chemischer Elemente in Proteinproben ermöglicht, konnte nachgewiesen werden, dass Zink und hSpt4 in äquimolaren Mengen vorlagen. Dies bestätigte, dass der Zinkfinger von hSpt4 strukturiert war. CD-spektroskopische Untersuchungen zeigten außerdem, dass hSpt4 Sekundärstrukturelemente besitzt (Einzelarbeit C, Abb. 4A). Basierend auf den CD-Daten wurden Sekundärstrukturanteile mit 27 % α -helikalem Anteil und 21 % β -Faltblattanteil ermittelt. Bei hSpt4 handelt es sich somit um ein $\alpha\beta$ -Typ Protein. Die gemessenen CD-Daten und ein zusätzliches gemessenes ^1H - ^{15}N -HSQC Spektrum bestätigten zudem, dass hSpt4 eine definierte Tertiärstruktur aufweist. (Einzelarbeit C, Abb. 4B).

Trotz der guten Löslichkeit und Strukturiertheit ist hSpt4 instabil. So neigten Proben mit Konzentrationen im NMR-Maßstab ($\geq 200 \mu\text{M}$) relativ schnell zur Aggregation und schließlich Präzipitation. Die Variation der Puffer und Messbedingungen ermöglichte es zwar, die Proteinprobe zu stabilisieren, doch erlaubte die Spektrenqualität der Tripelresonanzexperimente nur die Zuordnung eines Bruchteiles des Proteinerückgrates (Daten nicht gezeigt). Die Lösung der Struktur mittels NMR war daher nicht möglich. Aus

diesem Grund wurde ein zweiter Ansatz hinsichtlich der Strukturbestimmung der DSIF Komponenten gewählt.

3.2.2 Reinigung des rekombinanten hSpt4/hSpt5-NGN Komplexes

Die Instabilität von hSpt4 bei NMR-relevanten Konzentrationen führte zu erheblichen Problemen im Umgang mit den Proteinproben und verhinderte die Bestimmung der Struktur von hSpt4 in Lösung mittels NMR (s. 3.2.1). Zelluläres hSpt4 ist fester Bestandteil von DSIF und es gibt bis dato keine Hinweise, dass es auch als eigenständiger Faktor aktiv ist. Daher wurde in einem zweiten Ansatz versucht, die Interaktion zwischen hSpt4 und der NGN-Domäne von hSpt5, die die Reste 176-273 umfasst (Yamaguchi *et al.*, 1999b), im Hinblick auf eine verbesserte Stabilität von hSpt4 im Komplex näher zu charakterisieren. Hydrophobe Reste, die im Inneren des Heterodimers versteckt sind, sind höchstwahrscheinlich in der monomeren Form lösungsmittel exponiert und fördern Aggregation, was die Tendenz zur Präzipitation von hSpt4 erklären könnte.

Für die Interaktionsstudien wurde ein synthetisches Gen von hSpt5 verwendet, das für die Aminosäurereste 176-273 (NGN-Domäne) mit einem N-terminalen Hexahistidinanhang codiert. Das rekombinante hSpt5-NGN war unter Verwendung des *E. coli* BL21(DE3) Expressionssystem nicht löslich und konnte auch durch Zugabe von Detergenzien nicht in Lösung überführt werden. Erst die Koexpression von GB1-hSpt4 und 6His-hSpt5-NGN in *E. coli* BL21(DE3) ermöglichte es, geeignete Mengen an 6His-hSpt5 im Komplex mit 6His-GB1-hSpt4 in die lösliche Fraktion zu überführen und den hSpt4/hSpt5-NGN Komplex zu reinigen (Einzelarbeit D). Eine zusätzlich durchgeführte Größenausschlusschromatographie bestätigte das Vorliegen eines Heterodimers (Einzelarbeit D, Abb. 2). Die Tendenz der Proteinprobe zur Ausbildung von Proteinkristallen wurde in Kristallisationsansätzen für eine Komplexstrukturbestimmung mittels Röntgenkristallographie ausgenutzt. Durch Optimierung der Kristallisationsbedingungen konnten Kristalle von dreieckiger Form gewonnen und röntgenkristallographisch untersucht werden. Der verwendete Kristall gehörte zum orthorhombischen System der Raumgruppe $P2_12_12_1$. Unter Ausnutzung des natürlichen Vorkommens eines Zinkions im Zinkfinger von hSpt4 wurden *multiple-wavelength anomalous dispersion* (MAD)-Datensätze im Bereich der Zinkabsorptionskante aufgenommen (Einzelarbeit D).

3.2.3 Kristallstruktur von hSpt4/hSpt5-NGN

Die Kristallstruktur des Heterodimers hSpt4/hSpt5-NGN (PDB: 3H7H) konnte bis zu einer Auflösung von 1,55 Å mittels *Zink-single-wavelength anomalous dispersion* (SAD), d.h. unter Verwendung nur des *peak*-Datensatzes, gelöst werden. (Einzelarbeit D, Tab. 1).

Pro asymmetrische Einheit ist ein Heterodimer mit einem Solventgehalt von 40 % und einer zugänglichen Gesamtoberfläche von 11547 Å² vorhanden. Für die Reste 270-273 von hSpt5 wurde keine Elektronendichte detektiert, so dass die endgültige Struktur die Reste 176-269 umfasst. Für hSpt4 konnte die Struktur von 1-117 gelöst werden (Einzelarbeit D).

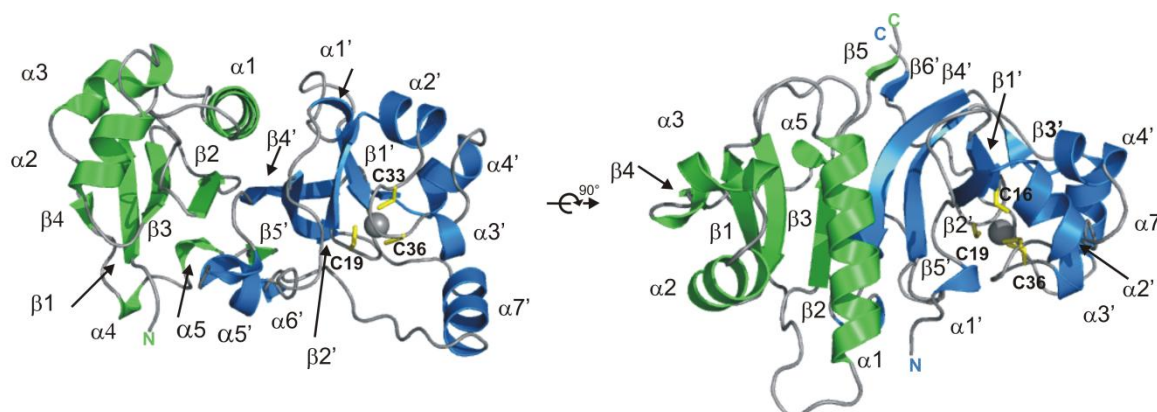


Abb. 8: Kristallstruktur des Heterodimers hSpt4/hSpt5-NGN (PDB: 3H7H)

hSpt5 (grün) besitzt ein zentrales viersträngiges β -Faltblatt und α -Helices, die auf beiden Seiten des Faltblattes angeordnet sind; hSpt4 (blau) verfügt über einen Cys₄-Zinkfinger (beteiligte Cysteinreste sind als gelbe Stäbchen hervorgehoben) und zwei β -Faltblätter, die senkrecht zu einander angeordnet sind; durch die Dimerisierung wird das zentrale β -Faltblatt von hSpt5 um zwei weitere Stränge von hSpt4 erweitert; zusätzlich packt $\alpha 1$ von hSpt5 gegen das nun sechssträngige Faltblatt und stabilisiert die Bindung; Sekundärstrukturelemente sind nummeriert; das Zinkion ist als graue Kugel dargestellt (aus: Einzelarbeit D, Abb. 3)

hSpt5-NGN besitzt die Faltungstopologie $\beta_1\alpha_2\beta_2\beta_3\alpha_2\alpha_3\beta_4\alpha_4\alpha_5\beta_6$ mit einem Faltblattanteil von 17,3 % und einem α -helikalen Anteil von 34,7 %. Es bildet ein zentrales, viersträngiges antiparalleles β -Faltblatt aus, gegen das die Helices $\alpha 1$ - $\alpha 3$ auf einer Seite und die Helices $\alpha 4$ und $\alpha 5$ auf der gegenüberliegenden Seite packen. hSpt4 hat mit 32,2 % α -helikalem Anteil und einem β -Faltblattanteil von 19,4 % einen ähnlichen Anteil an Sekundärstrukturelementen wie hSpt5-NGN, jedoch eine abweichende Faltungstopologie ($\alpha_1\beta_1\beta_2\alpha_2\alpha_3\alpha_4\beta_3\alpha_5\alpha_6\beta_4\alpha_7\beta_5$) (Abb. 8). Die auf CD-spektroskopischen Daten basierende Abschätzung der Sekundärstrukturanteile von hSpt4 (Einzelarbeit C) konnte die Kristallstruktur bestätigen. hSpt4 ist aus zwei senkrecht zueinander angeordneten zwei- bzw. dreisträngigen antiparallelen β -Faltblättern aufgebaut, die zur Interaktionsfläche hin angeordnet sind. Die helikalen Bereiche befinden sich im Dimer auf der lösungsmittelzugewandten Seite. Die vier Cysteinreste C16, C19, C33 und C36 koordinieren tetraedrisch ein Zinkion, wodurch die helikalen Bereiche $\alpha 2'$ - $\alpha 4'$ im Bereich der Faltblattstränge fixiert werden. Am N-Terminus ist eine kurze 3_{10} -Helix lokalisiert, deren Reste in hydrophobe Interaktionen mit hSpt5-NGN involviert sind und so stabilisierend auf die Dimerisierung wirken. Sowohl hSpt4 als auch hSpt5-NGN besitzt am C-Terminus einen kurzen β -Faltblattstrang ($\beta 6$ in hSpt5-NGN, $\beta 5'$ in hSpt4). Beide bilden zusammen ein paralleles β -Faltblatt aus. Dadurch sind beide C-Termini

nicht flexibel sondern in sich verankert und tragen zur Stabilisierung des Heterodimers bei. Im hSpt4/hSpt5-NGN Komplex ordnet sich der Faltblattstrang β_4' von hSpt4 antiparallel zum Faltblattstrang β_2 von hSpt5-NGN an und beide Stränge bilden die zentrale Interaktionsfläche aus. Das viersträngige antiparallele β -Faltblatt wird somit um zwei Stränge zu einem sechssträngigen zentralen β -Faltblatt des Heterodimers erweitert. Zusätzlich packt α_1 von hSpt5-NGN gegen den intermolekularen Faltblattbereich. Die Bindung beider Interaktionspartner erfolgt im Kernbereich hauptsächlich durch hydrophobe Wechselwirkungen, während am Rand der Bindungsfläche überwiegend polare Reste Bindungen eingehen (Einzelarbeit D).

Unlängst wurde die homologe Kristallstruktur von Spt4 im Komplex mit Spt5-NGN aus *S. cerevisiae* veröffentlicht (Guo *et al.*, 2008). Die Sequenzidentität zwischen Spt4 und hSpt4 sowie Spt5-NGN und hSpt5-NGN beträgt 46 % bzw. 28 %. Der hohe Grad an Konserviertheit spiegelt sich vor allem beim Vergleich beider Komplexstrukturen wider (Einzelarbeit D, Abb. S1), die eine identische Faltung einnehmen. Im Vergleich zur humanen Komplexstruktur von hSpt4/hSpt5-NGN, bei der beide Interaktionspartner eigenständige Proteine sind, wurden in der Hefestruktur Spt4 und Spt5 durch einen kurzen Peptidlinker miteinander verknüpft. Eine direkte Interaktion zwischen beiden Proteinen wird verhindert, weshalb sich in der asymmetrischen Einheit zwei Spt4/Spt5-NGN Komplexmoleküle befinden, die über Kreuz angeordnet sind, um somit eine Bindung mit dem jeweiligen Interaktionspartner zu ermöglichen.

Für die Hefestruktur wurde gezeigt, dass ein hochkonservierter Glutamat-Rest (E338) entscheidend für die Interaktion zwischen Spt4 und Spt5 ist. E228 aus Spt5 bildet Wasserstoffbrückenbindungen mit Helix α_4 sowie mit dem Serinrest S58 von Spt4. Die E338K Mutation in Spt5-NGN führte zu einem Verlust der Interaktion (Guo *et al.*, 2008). Auch in hSpt5-NGN konnten entsprechende Wasserstoffverbrückungen des konservierten Glutamatrestes (E228) mit hSpt4 nachgewiesen werden. Der Komplex hSpt4/hSpt5 konnte mit dem Aminosäureaustausch E228Q in hSpt5-NGN (hSpt4/hSpt5-NGN(E228Q)) wegen der Unlöslichkeit von hSpt5-NGN mit dem etablierten Reinigungsprotokoll nicht mehr gereinigt werden, was ein deutliches Anzeichen dafür ist, dass die Bindung beider Komplexpartner beeinträchtigt ist. So scheint der konservierte Glutamatrest auch im humanen Komplex entscheidend für die Wechselwirkungen beider Untereinheiten in DSIF zu sein (Einzelarbeit D).

Die NGN-Domäne besitzt Ähnlichkeiten zur N-terminalen Domäne (NTD) des bakteriellen Transkriptionsfaktors NusG (siehe 1.5.3). Aminosäuresequenzvergleiche mit NusG-NTD aus *E. coli* zeigten eine Sequenzidentität von 16% (Einzelarbeit D). Ein Strukturvergleich (Einzelarbeit D, Abb. 4A) bestätigt, dass eine hohe Ähnlichkeit zwischen den beiden

Strukturen besteht, jedoch sind auch deutliche Unterschiede zu erkennen. *E. coli* (*Ec*)NusG-NTD (PDB: 2K06) verfügt über eine stark verlängerte Schleifenregion zwischen β_2 und β_3 , die in hSpt5-NGN fehlt (Abb. 9). Im Gegensatz zu Spt5-NGN und hSpt5-NGN besitzt EcNusG-NTD an der konservierten Stelle des E338 bzw. E228 einen Glutaminrest (Q72).

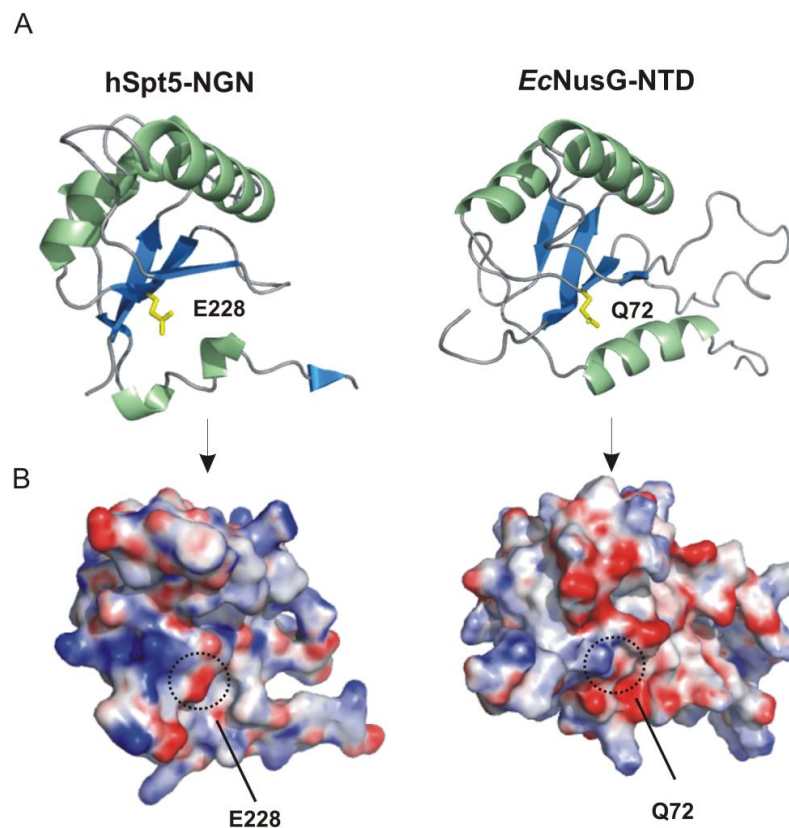


Abb. 9: Vergleich der Strukturen von hSpt5-NGN und *E. coli* NusG-NTD

(A) hSpt5-NGN und *E. coli* (*Ec*)NusG-NTD besitzen eine ähnliche Faltungstopologie; der Glutamatrest (E228) von hSpt5-NGN ist essentiell für die Interaktion mit hSpt4; EcNusG-NTD besitzt an entsprechender Position einen Glutaminrest (Q72) (B) Darstellung des elektrostatischen Oberflächenpotentials von hSpt5-NGN und EcNusG-NTD; die gezeigte Seite entspricht der Bindungsoberfläche mit hSpt4. Während in hSpt5-NGN positive und negative Reste gleichmäßig verteilt sind, überwiegen in EcNusG-NTD negativ geladene Seitenketten (aus: Einzelarbeit D, Abb. 4)

NMR-Titrationsexperimente mit isotope markiertem ^{15}N -hSpt4 zeigten, dass hSpt4, von dem es keine bakteriellen homologen Proteine gibt, und EcNusG-NTD *per se* nicht miteinander interagieren (Einzelarbeit D, Abb. S2A). Auch der Aminosäureaustausch Q72E konnte keine Interaktion zwischen hSpt4 und EcNusG-NTD induzieren (Einzelarbeit D, Abb. S2B). Grund hierfür ist sehr wahrscheinlich die negativ geladene Oberfläche von EcNusG-NTD (Abb. 9B), während in hSpt5-NGN die Bindungsoberfläche gleichmäßig aus positiven und negativen Resten aufgebaut ist (Abb. 9). Die Ladungsverteilung bei EcNusG-NTD würde die Bindung von hSpt4 auch bei vorhandenem Glutamatrest verhindern.

EcNusG-NTD interagiert direkt mit der RNAP (Mooney *et al.*, 2009). Darüber hinaus wurde in Hefe-Zwei-Hybrid Versuchen gezeigt, dass die *clamp*-Helices der β' -Untereinheit, die Rpb1 in RNAP II entspricht, an der Bindung direkt beteiligt sind (Nickels, 2009). Mutationsstudien deuten darauf hin, dass die Interaktion der RNAP über eine hydrophobe Tasche an der NusG-NTD-Oberfläche erfolgt (Mooney *et al.*, 2009). Spt5-NGN aus *S. cerevisiae* zeigt an der Spt4 abgewandten Seite eine ähnliche Verteilung hydrophober Reste wie *EcNusG*-NTD, während in hSpt5 eine derartige hydrophobe Tasche weniger stark ausgeprägt ist (Einzelarbeit D, Abb. S3). Ob nun eine Interaktion zwischen hSpt5 bzw. Spt5 mit der RNAP II in ähnlicher Weise wie bei *EcNusG*-NTD an die RNAP erfolgt, bedarf daher zusätzlicher Nachweise. In GST-*pull-down* Experimenten wurde zudem der Bereich 313-420 von hSpt5, also C-terminal der NGN-Domäne, als potentieller RNAP II-Interaktionsbereich identifiziert (Ivanov *et al.*, 2000; Yamaguchi *et al.*, 1999a).

Während in Bakterien hSpt4 fehlt, zeigt in Archaeen der Transkriptionsfaktor RpoE[“] Homologien zu hSpt4. Die Kristallstruktur von RpoE[“] aus *Pyrococcus furiosus* (PDB: 1RYQ) zeigt eine Sequenzidentität von 23% für den homologen Bereich. RpoE[“], das wesentlich kleiner (69 Reste) als hSpt4 ist, hat mit diesem die zentralen zwei, senkrecht zueinander angeordneten antiparallelen β -Faltblätter gemein. Wie hSpt4 verfügt RpoE[“] ebenso über einen aus vier Cysteinresten aufgebauten Zinkfinger. Der α -helikale Bereich ($\alpha 2$ - $\alpha 4$ in hSpt4) ist in RpoE[“] durch kurze Schleifenregionen ersetzt (Einzelarbeit D, Abb. 4)). Zwar ist über die Funktion von RpoE[“] wenig bekannt, doch deuten computergestützte *Docking*-Experimente mit der NusG-NTD Struktur aus dem Archaeon *Methanocaldococcus jannaschii* (Zhou *et al.*, 2009) und GST-*pull down* Experimente (Guo *et al.*, 2008) an, dass es im Komplex mit archaealen NusG vorliegt.

3.3 Ausblick

Mit den durchgeführten Experimenten und den bisher gewonnenen Daten konnte die Bindungsstelle von NELF E RRM auf der HIV-1 TAR RNA nicht lokalisiert werden. Des Weiteren ist NELF E RRM zwar in der Lage an die Stammregion der HIV-1 TAR RNA zu binden, nur deutet vieles darauf hin, dass die RNA nicht als Doppelstrang gebunden wird. Außerdem wurde in Chromatinimmunopräzipitationsanalysen gezeigt, dass die NELF-Elongationsblockade bei einer Transkriptlänge von 40 - 45 nt erfolgt, also bevor sich die HIV-1 TAR RNA ausgebildet hat (Zhang *et al.*, 2007). Dieses Ergebnis würde aber auch den Beobachtungen des proximalen Arretierens entsprechen. Denn sowohl für Gene von *H. sapiens* als auch für *D. melanogaster* wurde eine Transkriptionsblockade im Bereich +20-+50 von der Transkriptionsstartstelle entfernt detektiert. Ferner zeigten Permanganat-RNA-*footprints*, dass RNAP II bei NELF-assoziierten Genen in *D. melanogaster* im Bereich +30 - +50 nt stromabwärts der Transkriptionsstartstelle arretiert ist und dies unabhängig von der Transkriptsequenz ist (Lee *et al.*, 2008). Die durch die NMR- und Fluoreszenzgleichgewichtstitrationen gewonnenen Erkenntnisse, dass NELF E RRM keine Sequenzspezifität für die verwendeten RNA-Oligonukleotide zeigt, sind somit in guter Übereinstimmung mit den *in vitro* Daten aus *D. melanogaster*. So gilt es zu klären, ob die Transkriptionsblockade des proviralen HIV-1 Genoms wirklich nach dem momentan gültigen Modell stattfindet, d.h. ob eine vollständig ausgebildete HIV-1 TAR vorliegt, oder ein anderer Mechanismus zugrunde liegt. Tat interagiert noch mit anderen zellulären Transkriptionsfaktoren (z.B. Sp1) und der RNAP II, so dass TAR-unabhängige Mechanismen die Aufhebung der frühen Transkriptionsblockade initiieren könnten (Zhang *et al.*, 2007). Vor allem die Rolle von NELF in der HIV-1 TAR RNA Elongationsblockade bedarf weiterer Untersuchungen, auch im Hinblick, inwieweit sich promotornahes Arretieren bei zellulären Genen von der HIV-1 Transkriptionsblockade unterscheidet. In dieser Arbeit wurde hauptsächlich mit der RRM-Domäne von NELF E gearbeitet, es ist aber zusätzlich von Bedeutung, welche Rolle die restlichen Domänen von NELF E bzw. die verbleibenden NELF Untereinheiten bei der Transkriptionsblockade spielen. So könnte der Kontakt von NELF E mit den anderen Untereinheiten bzw. die Interaktion von NELF mit anderen Faktoren oder der RNAP II die Bindungsstelle an RNA vorgeben. Strukturbestimmung der gesamten NELF E Untereinheit bzw. der restlichen NELF Untereinheiten würde hierbei weitere wichtige Hinweise auf den Ablauf der von NELF initiierten Transkriptionsblockade geben. Die Untersuchung dieser Multiproteinkomplexe und die Aufklärung des Interaktionsnetzwerkes bzw. der zeitlichen Abläufe einzelner Schritte ist und bleibt jedoch eine Herausforderung. Durch die Struktur von hSpt4 im Komplex mit hSpt5-NGN konnte bestätigt werden, dass ein hoher Grad an Strukturähnlichkeit zum homologen Hefekomplex vorliegt und auch

Ähnlichkeiten zur bakteriellen NusG-NTD bzw. zum archaealen Transkriptionsfaktor RpoE“ vorhanden sind. Während die N-terminale Domäne von *E.coli* NusG direkt mit der RNAP interagiert (Mooney *et al.*, 2009), ist die Funktion der NGN-Domäne von hSpt5 bzw. Spt5 noch unbekannt. Ein Ansatzpunkt wäre, zu überprüfen, ob hSpt5-NGN auch an der direkten RNAP II Bindung beteiligt ist oder ob diese Wechselwirkung nur in anderen Bereichen stattfindet, wie *pull-down* Experimente zeigten (Ivanov *et al.*, 2000; Yamaguchi *et al.*, 1999b). Neben den funktionellen Analysen ist ebenfalls von großer Relevanz, die Gesamtstruktur von DSIF zu bestimmen, um ein umfassendes Bild der zellulären Funktion zu erhalten. Die NMR-Strukturen der drei KOW-Domänen KOW2 (PDB: 2E6z), KOW3 (PDB: 2DO3), und KOW5 (PDB: 2E70) von hSpt5 sind in der PDB-Datenbank hinterlegt, geben aber keinen Aufschluss über deren Funktion. Die Gesamtstruktur könnte einen besseren Einblick geben, wie DSIF zwischen RNAP II und anderen Transkriptionsfaktoren vermittelt. Für NusG aus *Aquifex aeolicus* konnte Nukleotidbindung nachgewiesen werden, jedoch ohne die beteiligte Domäne näher zu spezifizieren (Steiner *et al.*, 2002). Eine RNA- bzw. DNA-Bindung wurde für DSIF noch nicht detektiert. Sollte dies jedoch der Fall sein, würde es neue Einblicke zum einen in die Wechselwirkung von DSIF mit dem Elongationskomplex und zum anderen in das Zusammenspiel zwischen DSIF und NELF geben. Im Falle einer RNA-Bindung stellt sich die Frage, wie die zeitliche Abfolge dieser Bindung ist, da sowohl NELF als auch DSIF den gleichen naszierenden RNA-Strang binden würden.

4. Abkürzungsverzeichnis

ATP	Adenosintriphosphat
bp	Basenpaare
CD	Circulardichroismus
CDK9	Cyclinabhängige Kinase 9 (<i>cyclin dependent kinase 9</i>)
COBRA1	<i>Cofactor of BRCA 1</i>
CTD	C-terminale Domäne
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>
DNA	Desoxyribonukleinsäure (<i>deoxyribonucleic acid</i>)
DRB	5,6-Dichlor-1- β -D-ribofuranosylbenzimidazol
DSIF	<i>DRB sensitivity inducing factor</i>
<i>E. coli</i>	<i>Escherichia coli</i>
GTF	Genereller Transkriptionsfaktor
HDAg	Hepatitis Delta Antigen
HIV-1	humanes Immunschwächevirus 1 (<i>human immunodeficiency virus 1</i>)
<i>H. sapiens</i>	<i>Homo sapiens</i>
HSQC	<i>heteronuclear single quantum coherence</i>
KOW	Kyrpides, Ouzounis, Woese
LTR	<i>long terminal repeat</i>
MAD	<i>multi-wavelength anomalous dispersion</i>
NELF	negativer Elongationsfaktor
NF κ B	<i>nuclear factor factor κ B</i>
NGN	<i>NusG N-terminal homology domain</i>
NMR	magnetische Kernspinresonanz (<i>nuclear magnetic resonance</i>)
nt	Nukleotid
NusG	<i>N utilization substance G</i>
PIC	Präinitiationskomplex (<i>preinitiation complex</i>)
Pin1	Peptidyl-Prolyl-Isomerase
P-TEFb	positiver Transkriptionselongationsfaktor b
RNA	Ribonukleinsäure (<i>ribonucleic acid</i>)
RNAP II	RNA-Polymerase II
RNP	<i>ribonucleoprotein domain</i>
RRM	<i>RNA recognition motif</i>
SAD	<i>single-wavelength anomalous dispersion</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
Spt4	<i>Suppressor of Ty 4</i>

Spt5	<i>Suppressor of Ty 5</i>
TAR	<i>transactivation response</i>
Tat	<i>transcriptional transactivator</i>

5. Literaturverzeichnis

- Ainbinder, E., Amir-Zilberstein, L., Yamaguchi, Y., Handa, H. und Dikstein, R.** (2004). Elongation inhibition by DRB sensitivity-inducing factor is regulated by the A20 promoter via a novel negative element and NF- κ B. *Mol. Cell. Biol.* **24**: 2444-2454.
- Aiyar, S.E., Sun, J.L., Blair, A.L., Moskaluk, C.A., Lu, Y.Z., Ye, Q.N., Yamaguchi, Y., Mukherjee, A., Ren, D.M., Handa, H. und Li, R.** (2004). Attenuation of estrogen receptor alpha-mediated transcription through estrogen-stimulated recruitment of a negative elongation factor. *Genes Dev.* **18**: 2134-2146.
- Allain, F.H., Gubser, C.C., Howe, P.W., Nagai, K., Neuhaus, D. und Varani, G.** (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature.* **380**: 646-650.
- Amir-Zilberstein, L., Ainbinder, E., Toube, L., Yamaguchi, Y., Handa, H. und Dikstein, R.** (2007). Differential regulation of NF- κ B by elongation factors is determined by core promoter type. *Mol. Cell. Biol.* **27**: 5246-5259.
- Amir-Zilberstein, L. und Dikstein, R.** (2008). Interplay between E-box and NF- κ B in regulation of A20 gene by DRB sensitivity-inducing factor (DSIF). *J. Biol. Chem.* **283**: 1317-1323.
- Armache, K.J., Kettenberger, H. und Cramer, P.** (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 6964-6968.
- Armache, K.J., Mitterweger, S., Meinhart, A. und Cramer, P.** (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem.* **280**: 7131-7134.
- Auweter, S.D., Oberstrass, F.C. und Allain, F.H.** (2007). Solving the structure of PTB in complex with pyrimidine tracts: An NMR study of protein-RNA complexes of weak affinities. *J. Mol. Biol.* **367**: 174-186.
- Barboric, M., Nissen, R.M., Kanazawa, S., Jabrane-Ferrat, N. und Peterlin, B.M.** (2001). NF- κ B binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Mol. Cell.* **8**: 327-337.
- Berkhout, B., Silverman, R.H. und Jeang, K.T.** 1989. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell.* **59**: 273-82.
- Bourgeois, C.F., Kim, Y.K., Churcher, M.J., West, M.J. und Karn, J.** (2002). Spt5 cooperates with human immunodeficiency virus type 1 tat by preventing premature RNA release at terminator sequences. *Mol. Cell. Biol.* **22**: 1079-1093.
- Brady, J. und Kashanchi, F.** (2005). Tat gets the "green" light on transcription initiation. *Retrovirology.* **2**: 69.
- Brückner, F., Ortiz, J. und Cramer, P.** (2009). A movie of the RNA polymerase nucleotide addition cycle. *Curr. Opin. Struct. Biol.* **19**: 294-299.
- Buratowski, S.** (2003). The CTD code. *Nat. Struct. Biol.* **10**: 679-680.

- Burns, C.M., Richardson, L.V. und Richardson, J.P.** (1998). Combinatorial effects of NusA and NusG on transcription elongation and rho-dependent termination in *Escherichia coli*. *J Mol Biol.* **278**: 307-16.
- Burova, E., Hung, S., Sagitov, V., Stitt, B. und Gottesman, M.** (1995). *Escherichia coli* NusG protein stimulates transcription elongation rates in vivo and in vitro. *J. Bacteriol.* **177**: 1388-1392.
- Bushnell, D.A. und Kornberg, R.D.** (2003). Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: Implications for the initiation of transcription. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 6969-6973.
- Calnan, B.J., Biancalana, S., Hudson, D. und Frankel, A.D.** (1991). Analysis of arginine-rich peptides from the HIV tat protein reveals unusual features of RNA-protein recognition. *Genes Dev.* **5**: 201-10.
- Chen, D., Fong, Y. und Zhou, Q.** (1999). Specific interaction of tat with the human but not rodent P-TEFb complex mediates the species-specific tat activation of HIV-1 transcription. *Proc Natl Acad Sci U S A.* **96**: 2728-33.
- Chiang, P.W., Wang, S.Q., Smithivas, P., Song, W.J., Crombez, E., Akhtar, A., Im, R., Greenfield, J., Ramamoorthy, S., Van Keuren, M., Blackburn, C.C., Tsai, C.H. und Kurnit, D.M.** (1996). Isolation and characterization of the human and mouse homologues (SUPT4H and Supt4h) of the yeast SPT4 gene. *Genomics.* **34**: 368-375.
- Clery, A., Blatter, M. und Allain, F.H.** (2008). RNA recognition motifs: Boring? not quite. *Curr. Opin. Struct. Biol.* **18**: 290-298.
- Clore, G.M. und Gronenborn, A.M.** (1998). NMR structure determination of proteins and protein complexes larger than 20 kDa. *Curr. Opin. Chem. Biol.* **2**: 564-570.
- Cordingley, M.G., LaFemina, R.L., Callahan, P.L., Condra, J.H., Sardana, V.V., Graham, D.J., Nguyen, T.M., LeGrow, K., Gotlib, L. und Schlabach, A.J.** (1990). Sequence-specific interaction of tat protein and tat peptides with the transactivation-responsive sequence element of human immunodeficiency virus type 1 in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **87**: 8985-8989.
- Corsini, L. und Sattler, M.** (2007). Tudor hooks up with DNA repair. *Nat. Struct. Mol. Biol.* **14**: 98-99.
- Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brückner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.D., Lehmann, E., Leike, K., Sydow, J.F. und Vannini, A.** (2008). Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.* **37**: 337-352.
- Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R. und Kornberg, R.D.** (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science.* **288**: 640-649.
- Cramer, P., Bushnell, D.A. und Kornberg, R.D.** (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science.* **292**: 1863-1876.
- Downing, W.L., Sullivan, S.L., Gottesman, M.E. und Dennis, P.P.** (1990). Sequence and transcriptional pattern of the essential *Escherichia coli* *secE-nusG* operon. *J. Bacteriol.* **172**: 1621-1627.

- Edwards, A.M., Kane, C.M., Young, R.A. und Kornberg, R.D.** (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J. Biol. Chem.* **266**: 71-75.
- Egloff, S. und Murphy, S.** (2008). Cracking the RNA polymerase II CTD code. *Trends Genet.* **24**: 280-288.
- Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., Sato, H., Yamaguchi, Y., Mandal, S.S., Reinberg, D., Wada, T. und Handa, H.** (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Mol. Cell. Biol.* **24**: 3324-3336.
- Fish, R.N. und Kane, C.M.** (2002). Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta.* **1577**: 287-307.
- Fuda, N.J., Ardehali, M.B. und Lis, J.T.** (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature.* **461**: 186-192.
- Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T. und 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.
- Gardner, K.H. und Kay, L.E.** (1998). The use of ²H, ¹³C, ¹⁵N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **27**: 357-406.
- Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K., Collins, J.B., Li, L., Gilmour, D.S. und Adelman, K.** (2008). NELF-mediated stalling of pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev.* **22**: 1921-1933.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A. und Kornberg, R.D.** (2001). Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science.* **292**: 1876-1882.
- Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G.A., Teng, M. und Niu, L.** 2008. Core structure of the yeast Spt4-Spt5 complex: A conserved module for regulation of transcription elongation. *Structure.* **16**: 1649-1658.
- Hahn, S.** (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11**: 394-403.
- Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S. und Young, R.A.** (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell.* **95**: 717-728.
- Ingham, C.J., Dennis, J. und Furneaux, P.A.** (1999). Autogenous regulation of transcription termination factor rho and the requirement for Nus factors in *Bacillus subtilis*. *Mol. Microbiol.* **31**: 651-663.
- Isel, C. und Karn, J.** (1999). Direct evidence that HIV-1 tat stimulates RNA polymerase II carboxyl-terminal domain hyperphosphorylation during transcriptional elongation. *J. Mol. Biol.* **290**: 929-941.
- Ivanov, D., Kwak, Y.T., Guo, J. und Gaynor, R.B.** (2000). Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol. Cell. Biol.* **20**: 2970-2983.

- Jones, K.A. und Peterlin, B.M.** (1994). Control of rna initiation and elongation at the HIV-1 promoter. [review] [205 refs]. *Annu. Rev. Biochem.* **63**: 717-743.
- Kao, S.Y., Calman, A.F., Luciw, P.A. und Peterlin, B.M.** (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature.* **330**: 489-493.
- Kettenberger, H., Armache, K.J. und Cramer, P.** (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell.* **114**: 347-357.
- Kettenberger, H., Armache, K. und Cramer, P.** (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell.* **16**: 955-965.
- Kielkopf, C.L., Rodionova, N.A., Green, M.R. und Burley, S.K.** (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell.* **106**: 595-605.
- Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H. und Kornberg, R.D.** (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell.* **77**: 599-608.
- Koch, F., Jourquin, F., Ferrier, P. und Andrau, J.C.** (2008). Genome-wide RNA polymerase II: Not genes only! *Trends Biochem. Sci.* **33**: 265-273.
- Kwak, Y.T., Guo, J., Prajapati, S., Park, K.J., Surabhi, R.M., Miller, B., Gehrig, P. und Gaynor, R.B.** (2003). Methylation of SPT5 regulates its interaction with RNA polymerase II and transcriptional elongation properties. *Mol. Cell.* **11**: 1055-1066.
- Kyrpides, N.C., Woese, C.R. und Ouzounis, C.A.** (1996). KOW: A novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.* **21**: 425-426.
- Laspia, M.F., Rice, A.P. und Mathews, M.B.** (1989). HIV-1 tat protein increases transcriptional initiation and stabilizes elongation. *Cell.* **59**: 283-292.
- Lavoie, S.B., Albert, A.L., Handa, H., Vincent, M. und Bensaude, O.** (2001). The peptidyl-prolyl isomerase Pin1 interacts with hSpt5 phosphorylated by Cdk9. *J. Mol. Biol.* **312**: 675-685.
- Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F. und Gilmour, D.S.** (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. *Mol. Cell. Biol.* **28**: 3290-3300.
- Li, B., Carey, M. und Workman, J.L.** (2007). The role of chromatin during transcription. *Cell.* **128**: 707-719.
- Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., 3rd und Hartzog, G.A.** (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell. Biol.* **23**: 1368-1378.
- Lis, J.T., Mason, P., Peng, J., Price, D.H. und Werner, J.** (2000). P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev.* **14**: 792-803.
- Liu, W., Shen, X., Yang, Y., Yin, X., Xie, J., Yan, J., Jiang, J., Liu, W., Wang, H., Sun, M., Zheng, Y. und Gu, J.** (2004). Trihydrophobin 1 is a new negative regulator of A-raf kinase. *J. Biol. Chem.* **279**: 10167-10175.

- Mancebo, H.S., Lee, G., Flygare, J., Tomassini, J., Luu, P., Zhu, Y., Peng, J., Blau, C., Hazuda, D., Price, D. und Flores, O.** (1997). P-TEFb kinase is required for HIV tat transcriptional activation in vivo and in vitro. *Genes Dev.* **11**: 2633-2644.
- Maris, C., Dominguez, C. und Allain, F.H.** (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**: 2118-2131.
- Martinez, E.** (2002). Multi-protein complexes in eukaryotic gene transcription. *Plant Mol. Biol.* **50**: 925-947.
- McCoy, M.A. und Wyss, D.F.** (2000). Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations. *J. Biomol. NMR.* **18**: 189-198.
- Mooney, R.A., Schweimer, K., Rösch, P., Gottesman, M.E. und Landick, R.** (2009). Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase und regulators. *J. Mol. Biol.* **391**: 341-358.
- Morillon, A., Karabetsou, N., O'Sullivan, J., Kent, N., Proudfoot, N. und Mellor, J.** (2003). Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. *Cell.* **115**: 425-435.
- Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J. und Adelman, K.** (2007). RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39**: 1507-1511.
- 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. und 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.
- Narita, T., Yung, T.M., Yamamoto, J., Tsuboi, Y., Tanabe, H., Tanaka, K., Yamaguchi, Y. und Handa, H.** (2007). NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. *Mol. Cell.* **26**: 349-365.
- Narlikar, G.J., Fan, H.Y. und Kingston, R.E.** (2002). Cooperation between complexes that regulate chromatin structure and transcription. *Cell.* **108**: 475-487.
- Nickels, B.E.** (2009). Genetic assays to define and characterize protein-protein interactions involved in gene regulation. *Methods.* **47**: 53-62.
- Orphanides, G. und Reinberg, D.** (2002). A unified theory of gene expression. *Cell.* **108**: 439-451.
- Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. und Nagai, K.** (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature.* **372**: 432-438.
- Palancade, B. und Bensaude, O.** (2003). Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur. J. Biochem.* **270**: 3859-3870.
- Palangat, M., Renner, D.B., Price, D.H. und Landick, R.** (2005). A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 15036-15041.
- Ping, Y.H. und 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.
- Ponting, C.P.** (2002). Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* **30**: 3643-3652.
- Rasmussen, E.B. und Lis, J.T.** (1993). In vivo transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 7923-7927.
- Reid, G., Gallais, R. und Metivier, R.** (2009). Marking time: The dynamic role of chromatin and covalent modification in transcription. *Int. J. Biochem. Cell Biol.* **41**: 155-163.
- Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H. und Price, D.H.** (2001). A highly purified RNA polymerase II elongation control system. *J. Biol. Chem.* **276**: 42601-42609.
- Rougvie, A.E. und Lis, J.T.** (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell.* **54**: 795-804.
- Rupert, P.B., Xiao, H. und Ferre-D'Amare, A.R.** (2003). U1A RNA-binding domain at 1.8 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **59**: 1521-1524.
- Saunders, A., Core, L.J. und Lis, J.T.** (2006). Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol.* **7**: 557-567.
- Seet, B.T., Dikic, I., Zhou, M.M. und Pawson, T.** (2006). Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* **7**: 473-483.
- Shilatifard, A., Conaway, R.C. und Conaway, J.W.** (2003). The RNA polymerase II elongation complex. *Annu. Rev. Biochem.* **72**: 693-715.
- Simchen, G., Winston, F., Styles, C.A. und Fink, G.R.** (1984). Ty-mediated gene expression of the LYS2 and HIS4 genes of *Saccharomyces cerevisiae* is controlled by the same SPT genes. *Proc. Natl. Acad. Sci. U. S. A.* **81**: 2431-2434.
- Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A. und Arndt, K.M.** (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *EMBO J.* **22**: 1846-1856.
- Sims, R.J.,III, Belotserkovskaya, R. und Reinberg, D.** (2004). Elongation by RNA polymerase II: The short and long of it. *Genes Dev.* **18**: 2437-2468.
- Skinner, A.L. und Laurence, J.S.** (2008). High-field solution NMR spectroscopy as a tool for assessing protein interactions with small molecule ligands. *J. Pharm. Sci.* **97**: 4670-4695.
- Squazzo, S.L., Costa, P.J., Lindstrom, D.L., Kumer, K.E., Simic, R., Jennings, J.L., Link, A.J., Arndt, K.M. und Hartzog, G.A.** (2002). The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO J.* **21**: 1764-1774.
- Stachora, A.A., Schafer, R.E., Pohlmeier, M., Maier, G. und Ponstingl, H.** (1997). Human Supt5h protein, a putative modulator of chromatin structure, is reversibly phosphorylated in mitosis. *FEBS Lett.* **409**: 74-78.

- Steiner, T., Kaiser, J.T., Marinkovic, S., Huber, R. and Wahl, M.C.** (2002). Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *EMBO J.* **21**: 4641-4653.
- Steitz, T.A.** (1998). A mechanism for all polymerases. *Nature.* **391**: 231-232.
- Sun, J., Blair, A.L., Aiyar, S.E. und Li, R.** (2007). Cofactor of BRCA1 modulates androgen-dependent transcription and alternative splicing. *J. Steroid Biochem. Mol. Biol.* **107**: 131-139.
- Swanson, M.S., Malone, E.A. und Winston, F.** (1991). SPT5, an essential gene important for normal transcription in *Saccharomyces cerevisiae*, encodes an acidic nuclear protein with a carboxy-terminal repeat. *Mol. Cell. Biol.* **11**: 4286.
- Swanson, M.S. und Winston, F.** (1992). SPT4, SPT5 and SPT6 interactions: Effects on transcription and viability in *Saccharomyces cerevisiae*. *Genetics.* **132**: 325-336.
- Taube, R., Fujinaga, K., Wimmer, J., Barboric, M. und Peterlin, B.M.** (1999). Tat transactivation: A model for the regulation of eukaryotic transcriptional elongation. *Virology.* **264**: 245-253.
- Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., Buratowski, S. und Handa, H.** (1998a). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**: 343-356.
- Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D. und Handa, H.** (1998b). Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *EMBO J.* **17**: 7395-7403.
- Wang, W., Carey, M. und Gralla, J.D.** (1992). Polymerase II promoter activation: Closed complex formation and ATP-driven start site opening. *Science.* **255**: 450-453.
- Wang, X., Lee, C., Gilmour, D.S. und Gergen, J.P.** (2007). Transcription elongation controls cell fate specification in the drosophila embryo. *Genes Dev.* **21**: 1031-1036.
- Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H. und Jones, K.A.** (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell.* **92**: 451-462.
- Weir, J.P. und Elkins, K.L.** (1993). Replication-incompetent herpesvirus vector delivery of an interferon alpha gene inhibits human immunodeficiency virus replication in human monocytes. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 9140-9144.
- Wen, Y. und Shatkin, A.J.** (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.* **13**: 1774-1779.
- Werner, F.** (2008). Structural evolution of multisubunit RNA polymerases. *Trends Microbiol.* **16**: 247-250.
- Westover, K.D., Bushnell, D.A. und Kornberg, R.D.** (2004a). Structural basis of transcription: Separation of RNA from DNA by RNA polymerase II. *Science.* **303**: 1014-1016.
- Westover, K.D., Bushnell, D.A. und Kornberg, R.D.** (2004b). Structural basis of transcription: Nucleotide selection by rotation in the RNA polymerase II active center. *Cell.* **119**: 481-489.

- Woychik, N.A. und Hampsey, M.** (2002). The RNA polymerase II machinery: Structure illuminates function. *Cell*. **108**: 453-463.
- Wright, T.J., Costa, J.L., Naranjo, C., Francis-West, P. und Altherr, M.R.** (1999). Comparative analysis of a novel gene from the wolf-Hirschhorn/Pitt-rogers-danks syndrome critical region. *Genomics*. **59**: 203-212.
- Wu, C.H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H. und Gilmour, D.** (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev.* **17**: 1402-1414.
- Xiao, Y., Yang, Y.H., Burckin, T.A., Shiue, L., Hartzog, G.A. und Segal, M.R.** (2005). Analysis of a splice array experiment elucidates roles of chromatin elongation factor Spt4-5 in splicing. *PLoS Comput. Biol.* **1**: e39.
- Xu, R.M., Jokhan, L., Cheng, X., Mayeda, A. und Krainer, A.R.** (1997). Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure*. **5**: 559-570.
- Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T. und Handa, H.** (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell*. **21**: 227-237.
- Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., Wada, T., Sugimoto, S., Konarska, M.M. und Handa, H.** (2001). Stimulation of RNA polymerase II elongation by hepatitis delta antigen. *Science*. **293**: 124-127.
- Yamaguchi, Y., Inukai, N., Narita, T., Wada, T. und 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.
- Yamaguchi, Y., Mura, T., Chanarat, S., Okamoto, S. und Handa, H.** (2007). Hepatitis delta antigen binds to the clamp of RNA polymerase II and affects transcriptional fidelity. *Genes Cells*. **12**: 863-875.
- Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J. und Handa, H.** (1999a). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*. **97**: 41-51.
- Yamaguchi, Y., Wada, T., Watanabe, D., Takagi, T., Hasegawa, J. und Handa, H.** (1999b). Structure and function of the human transcription elongation factor DSIF. *J. Biol. Chem.* **274**: 8085-8092.
- Yang, C., Bolotin, E., Jiang, T., Sladek, F.M. und Martinez, E.** (2007a). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*. **389**: 52-65.
- Yang, Y., Liu, W., Zou, W., Wang, H., Zong, H., Jiang, J., Wang, Y. und Gu, J.** (2007b). Ubiquitin-dependent proteolysis of trihydrophobin 1 (TH1) by the human papilloma virus E6-associated protein (E6-AP). *J. Cell. Biochem.* **101**: 167-180.
- Ye, Q., Hu, Y.F., Zhong, H., Nye, A.C., Belmont, A.S. und Li, R.** (2001). BRCA1-induced large-scale chromatin unfolding and allele-specific effects of cancer-predisposing mutations. *J. Cell Biol.* **155**: 911-921.

- Yin, X.L., Chen, S. und Gu, J.X.** (2002). Identification of TH1 as an interaction partner of A-raf kinase. *Mol. Cell. Biochem.* **231**: 69-74.
- Yung, T.M., Narita, T., Komori, T., Yamaguchi, Y. und Handa, H.** (2009). Cellular dynamics of the negative transcription elongation factor NELF. *Exp. Cell Res.* **315**: 1693-1705.
- Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M. und Young, R.A.** (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* **39**: 1512-1516.
- Zhang, Z., Klatt, A., Gilmour, D.S. und Henderson, A.J.** (2007). Negative elongation factor NELF represses human immunodeficiency virus transcription by pausing the RNA polymerase II complex. *J. Biol. Chem.* **282**: 16981-16988.
- Zhou, H., Liu, Q., Gao, Y., Teng, M. und Niu, L.** (2009). Crystal structure of NusG N-terminal (NGN) domain from *Methanocaldococcus jannaschii* and its interaction with RpoE". *Proteins.* **76**: 787-793.
- Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M.B. und Price, D.H.** (1997). Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev.* **11**: 2622-2632.
- Zuiderweg, E.R.** (2002). Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry.* **41**: 1-7.

6. Publikationsliste

6.1 Einzelarbeit A

Jampani N. Rao*, Liane Neumann*, Sabine Wenzel, Kristian Schweimer, Paul Rösch und Birgitta M. Wöhrl (2006): Structural studies on the RNA-recognition motif of NELF E, a cellular negative transcription elongation factor involved in the regulation of HIV transcription. *Biochemical Journal* **400**(3), 449-56

*beide Autoren haben im gleichen Maße zur Arbeit beigetragen

Alle Fluoreszenzgleichgewichtstitrationen wurden von mir durchgeführt. Liane Neumann führte die Klonierung und Reinigung von NELF E RRM durch. Alle NMR-Experimente wurden durch Jampani N. Rao durchgeführt. NMR-Datenanalyse und Strukturrechnung erfolgte durch Jampani N. Rao unter Anleitung von Kristian Schweimer. Birgitta M. Wöhrl plante und betreute das Projekt. Das Manuskript wurde von Birgitta Wöhrl und Kristian Schweimer unter Mithilfe von mir, Jampani N. Rao und Paul Rösch verfasst.

6.2 Einzelarbeit B

Jampani N. Rao, Kristian Schweimer, Sabine Wenzel, Birgitta M. Wöhrl und Paul Rösch (2008): NELF-E RRM undergoes major structural changes in flexible protein regions on target RNA binding. *Biochemistry* **47**(12), 3756-61

Alle NMR-Experimente wurden von Jampani N. Rao unter Mithilfe von Kristian Schweimer durchgeführt, ebenso wie die Datenanalyse und Strukturrechnung der gebundenen NELF E RRM-Form. Vorarbeiten dazu und zusätzliche Fluoreszenzgleichgewichtstitrationen wurden von mir durchgeführt (Daten nicht aufgeführt). Birgitta M. Wöhrl plante und betreute das Projekt. Das Manuskript wurde von Paul Rösch, Kristian Schweimer und Birgitta M. Wöhrl, mit Unterstützung von mir und Jampani N. Rao, verfasst.

6.3 Einzelarbeit C

Sabine Wenzel, Kristian Schweimer, Paul Rösch und Birgitta M. Wöhrl (2008): The small hSpt4 subunit of the human transcription elongation factor DSIF is a Zn-finger protein with alpha/beta type topology. *Biophysical and Biochemical Research Communications* **370**(3):414-8

Klonierung, Reinigung und CD-spektroskopische Untersuchungen von hSpt4 wurden von mir durchgeführt. Die NMR-Experimente wurden von mir und Kristian Schweimer durchgeführt und analysiert. Birgitta M. Wöhrl plante und betreute das Projekt. Das Manuskript wurde von mir und Birgitta M. Wöhrl mit Unterstützung von Kristian Schweimer und Paul Rösch verfasst.

6.4 Einzelarbeit D

Sabine Wenzel, Berta M. Martins, Paul Rösch und Birgitta M. Wöhrl (2009): Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface. *Biochemical Journal* (**im Druck**)

Alle aufgeführten Experimente wurden von mir geplant. Die Klonierung und Reinigung des Heterodimers hSpt4/hSpt5-NGN, die NMR-Experimente und Proteinkristallisationsversuche wurden von mir durchgeführt. Die röntgenkristallographische Datenerfassung und Auswertung erfolgten durch mich und Berta M. Martins. Birgitta M. Wöhrl konzipierte und koordinierte das Projekt. Das Manuskript wurde von mir, Birgitta M. Wöhrl und Berta M. Martins mit Unterstützung von Paul Rösch verfasst.

7. Einzelarbeit A

Jampani N. Rao*, Liane Neumann*, Sabine Wenzel, Kristian Schweimer, Paul Rösch und Birgitta M. Wöhl

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

Biochemical Journal **400**(3), 449-56 (2006)

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

Jampani N. RAO¹, Liane NEUMANN¹, Sabine WENZEL, Kristian SCHWEIMER², Paul RÖSCH and Birgitta M. WÖHRL²

Universität Bayreuth, Lehrstuhl Biopolymere, Universitätsstrasse 30, D-95447 Bayreuth, Germany

The elongation of transcription of HIV RNA at the TAR (transactivation-response element) is highly regulated by positive and negative factors. The cellular negative transcription elongation factor NELF (negative elongation factor) was suggested to be involved in transcriptional regulation of HIV-1 (HIV type 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. In the present study, we determined the solution structure of the RRM (RNA-recognition motif) of the RNA-binding subunit NELF E and studied its interaction with the viral TAR RNA. Our results show that the separately expressed

recombinant NELF E RRM has α -helical and β -strand elements adopting a $\beta\alpha\beta\beta\alpha\beta$ fold and is able to bind to TAR RNA. Fluorescence equilibrium titrations with fluorescently labelled 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-micromolar range.

Key words: HIV termination/anti-termination, negative elongation factor E (NELF E), NMR structure, RNA-binding protein, transactivation-response element (TAR).

INTRODUCTION

After HIV-1 (HIV type 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 LTR (long terminal repeat) of the integrated proviral DNA by the cellular RNA Pol II (polymerase II). Once the initial sequence of 60 nucleotides of the viral RNA, the so-called TAR (transactivator-response 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 (negative elongation factor) and DSIF [DRB (5,6-dichloro-1 β -D-ribofuranosylbenzimidazole)-sensitivity-inducing factor], bind to the TAR element [2–5]. Elongation of transcription is possible only if the viral transactivator protein Tat (transactivator of transcription) recruits the cellular positive transcription elongation factor pTEFb (positive transcription elongation factor b) 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 (cyclin-dependent kinase 9) of pTEFb, anti-termination occurs, leading to productive elongation of transcription [5]. NELF consists of four 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 HDAG (hepatitis delta antigen) which binds to and activates RNA pol II [8]. NELF E contains an RRM (RNA-recognition motif). Furthermore, NELF E interacts with the NELF B subunit, probably via a leucine zipper motif. It has been shown 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 that NELF E plays a role in the control of HIV transcription.

EXPERIMENTAL

Plasmid construct, expression and protein purification

A synthetically produced gene adapted to the *Escherichia coli* codon usage harbouring 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 His₆ tag (Figure 1).

E. coli strain BL21(DE3) (Novagen) containing the recombinant plasmid was grown at 37 °C in LB (Luria–Bertani) medium containing ampicillin until a D_{600} of 1 was reached and then induced with 1 mM IPTG (isopropyl β -D-thiogalactoside). Cells were harvested 3–4 h after induction. For ¹⁵N- and ¹³C-labelling, cells were pelleted at 15 °C before induction and resuspended in

Abbreviations used: C₁₂E₅, penta(ethylene glycol)monododecyl ether; DSIF, DRB (5,6-dichloro-1 β -D-ribofuranosylbenzimidazole)-sensitivity-inducing factor; DTT, dithiothreitol; FAM, 6-carboxyfluorescein; HIV-1, HIV type 1; HSQC, heteronuclear single-quantum coherence; IPTG, isopropyl β -D-thiogalactoside; MEXICO, measurement of exchange rates in isotopically labelled compounds; NELF, negative elongation factor; NOE, nuclear Overhauser effect; Pol II, polymerase II; pTEFb, positive transcription elongation factor b; RDC, residual dipolar coupling; RMSD, root mean square deviation; RRM, RNA-recognition motif; TAR, transactivation-response element.

¹ These authors contributed equally to this work.

² Correspondence may be addressed to either of these authors (email kristian.schweimer@uni-bayreuth.de or birgitta.woehrl@uni-bayreuth.de).

Structural co-ordinates for NELF (negative elongation factor) E RNA-recognition motif have been deposited in the Protein Data Bank under accession code 2BZ2.

The nucleotide sequence data for NELF (negative elongation factor) E RNA-recognition motif have been deposited in the DDBJ, EMBL, GenBank® and GSDS Nucleotide Sequence Databases under the accession number DQ885937.

```

atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtgccgcg cggcagccat 20
M G S S H H H H H H S S G L V P R G S H
atggggcgt tcgcgcgttc tgattccttc cctgaacgtc gcgcccgcg taaagggaac 40
M G P F R R S D S F P E R R A P R K G N
acattatag tttatggcga agacatgacc cgcagctctt tgcgcggtc cttctcccc 60
T L Y V Y G E D M T P T L L R G A F S P
ttcggcaaca ttatgatct gtcctggac cctccgcgta attgcccgtt tctcaactac 80
G N I I D L S N D P P R N C A P Y T Y
gaaaaaatgg aaagcgcaga tcaagctgtg gccgaactga atggaacgca ggttgagtcg 100
E K M E S A D Q A V A E L N G T Q V E S
gtccagctca aagtgaacat tgcgcgcaag cagccgatgc tggatgccgc tactggcaag 120
V Q L K V N I A R K Q P M L D A A T G K
tcttag
S -

```

Figure 1 Nucleotide and amino acid sequence of NELF E RRM

The construct contains a His₆ tag at the N-terminus, followed by a thrombin cleavage site indicated by the grey box. The amino acids defining the RRM are underlined, the amino acid sequence derived from NELF E is written in bold letters. The conserved ribonucleoprotein motifs RNP2 and RNP1 containing residues Tyr⁴³ and Phe⁷⁷ are marked by black boxes. Amino acid numbering is indicated on the right.

1/4 vol. of M9 medium [10,11] supplemented with ¹⁵NH₄Cl and 0.2% [¹³C]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 [6 × 45 s, 200 W; Labsonic U (B. Braun Biotech)] in 20 mM sodium phosphate buffer, pH 7.4, 0.5 M NaCl, 5 mM imidazole and 1 mM DTT (dithiothreitol). After centrifugation at 19 100 g for 45 min, the supernatant was loaded on to a Ni²⁺-affinity column (His-trap chelating, GE Healthcare) and eluted by applying an imidazole step gradient. Peak fractions containing NELF E RRM were dialysed against 10 mM sodium phosphate buffer, pH 6.9, 250 mM NaCl and 1 mM DTT, and purified further on 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 dialysed against 10 mM sodium phosphate buffer, pH 6.9, 100 mM NaCl and 1 mM DTT, concentrated with Vivaspin concentrators (Vivascience, molecular-mass cut-off 5000 Da), divided into aliquots 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'-GCCUCUCUGGUUAGACCAGAUCUGAGCCUGGG-AGCUCUCUGGCUAACUAGGGAAGGC-3') plus the region complementary to the T7 primer. The TAR RNA was then purified by denaturing polyacrylamide/urea gel electrophoresis as described in [13].

NMR spectrometry

All NMR experiments were performed at 298 K on Bruker DRX600, Avance 700 and Avance 800 spectrometers equipped with standard inverted or cryogenically cooled ¹H/¹³C/¹⁵N 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 ¹⁵N- or ¹³C/¹⁵N-labelled recombinant NELF E RRM at a concentration of 0.6 mM. Distance restraints for structure calculation were derived from three-dimensional ¹³C- and ¹⁵N-edited NOESY-HSQC (heteronuclear single-quantum coherence) experiments with mixing times of 120 ms [16,17]. Dihedral angle restraints were derived from

³J(¹H^N,H^α) scalar coupling constants determined from the intensity ratios of cross and diagonal peaks of the HNHA spectrum [18]. A series of New MEXICO (measurement of exchange rates in isotopically labelled compounds) [19] experiments with different mixing times were recorded for characterizing amide proton exchange. ¹D (one-bound) (¹H^N, ¹⁵N) RDCs (residual dipolar couplings) were determined by the IPAP (in-phase/antiphase) method [20] using a weakly aligned sample of uniformly ¹⁵N-labelled NELF E RRM in a mixture of penta(ethylene glycol)-monododecyl ether (C₁₂E₅), hexanol and water [molar ratio of C₁₂E₅/hexanol = 0.95, 3% (w/v) C₁₂E₅/water] [21]. {¹H}¹⁵N NOE (nuclear Overhauser effect) 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 ¹H/¹⁵N-HSQC spectra after gradually adding TAR RNA to a sample containing ¹⁵N-labelled NELF E RRM.

Normalized chemical-shift changes were expressed as the weighted geometric average of ¹H^N- and ¹⁵N-chemical-shift changes for each residue:

$$\Delta\delta_{\text{norm}} = \sqrt{(\Delta\delta_{1\text{H}})^2 + 0.1(\Delta\delta_{15\text{N}})^2}$$

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

Structure calculation

Distance restraints for structure calculation were derived from ¹⁵N-NOESY-HSQC and ¹³C-NOESY-HSQC spectra. NOESY cross-peaks were classified according to their relative intensities and converted into distance restraints with upper limits of 3.0 Å (1 Å = 0.1 nm) (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 antiphase 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]. Glycine residues 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 between the carboxy oxygen and the amide proton was below 2.6 Å, and the angle of the amide proton, the amide nitrogen and the carboxy 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 proline residues 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

Table 1 Structural statistics

E_{pot} , overall potential energy; E_{impr} , potential energy of improper angle; E_{cdih} , potential energy for dihedral restraints; E_{sani} , potential energy for RDC restraints.

Parameter	Value
Experimental restraints	
Distance restraints	
Total	1926
Intraresidual	468
Sequential	406
Medium-range	281
Long-range	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	0.00068 ± 0.00004
Distance restraints	0.0036 ± 0.0007
RMSDs from ideal angles (°)	
Bond angles	0.14 ± 0.06
Dihedral angle restraints	0.15 ± 0.15
RMSDs from dipolar couplings (Hz)	0.11 ± 0.02
Atomic co-ordinate precision (RMSD) (Å)	
Backbone heavy atoms	0.28 (Gly ³⁹ –Ala ¹⁰⁸)
Heavy atoms	0.64 (Gly ³⁹ –Ala ¹⁰⁸)
Ramachandran plot statistics	
Residues in	
Most-favoured regions	91.0 %
Allowed regions	9.0 %

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 (Table 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 ¹D (¹H^N, ¹⁵N) RDCs with a harmonic potential [35]. Dipolar couplings of flexible residues showing a {¹H} ¹⁵N 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 ¹D (¹H^N, ¹⁵N) [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-favourable combination of D_a and R .

The 20 structures showing the lowest values of the target function excluding the database potential were analysed further with X-PLOR 3.8.5.1, MOLMOL [38] and PROCHECK 3.5.4

[39,40]. The structural co-ordinates were deposited in the Protein Data Bank (PDB) under accession code 2BZZ.

Fluorescence equilibrium titrations

Measurements were performed using a Fluorolog spectrophotometer (HORIBA Jobin Yvon) in reaction buffer consisting of 10 mM sodium phosphate, pH 7.0, and 100 mM NaCl in a volume of 2 ml. The excitation wavelength was 495 nm, and the emission intensity was measured at 522 or 525 nm with slit widths set at 2 nm for both excitation and emission. To analyse the double-stranded TAR RNA stem, the oligoribonucleotide representing the 5' end of HIV-1 TAR, TAR-(1–18) (5'-GGUCUCUGG-UUAGACC-3') was hybridized in reaction buffer to the complementary 5'-FAM (6-carboxyfluorescein)-labelled oligoribonucleotide (biomers.net) representing the 3' end of the HIV-1 TAR RNA, TAR-(42–57) (5'-FAM-GGCUAACUAGGGAACC-3'). 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-labelled oligoribonucleotides were used. Double- or single-stranded substrate (50 nM) 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:

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

where F_{obs} is observed fluorescence, F_L^0 is starting fluorescence, $[L_0]$ is concentration of the RNA, ΔF^0 is change in fluorescence, and $[E_0]$ is concentration of protein.

RESULTS AND DISCUSSION

Purification of NELF E RRM

To analyse the structure and function of NELF E RRM, a synthetic gene containing the corresponding DNA coding for an N-terminal His₆ tag followed by a thrombin site and the NELF E RRM was constructed with the codon usage adapted to *E. coli* (Figure 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 was deposited into GenBank® under the accession number DQ885937. The expressed and purified protein was stable and could be stored in 10 mM sodium phosphate buffer, pH 6.9, containing 100 mM NaCl and 1 mM DTT, by shock-freezing aliquots at –80 °C. After thawing and addition of 10 % (v/v) ²H₂O, 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 CD (results not shown) and one-dimensional ¹H-NMR spectroscopy. Using standard double- and triple-resonance NMR techniques with isotopically labelled protein, ¹H, ¹³C and ¹⁵N backbone and side-chain resonances were assigned. Complete backbone and nearly complete side-chain assignment was obtained for the region Ala³⁵–Arg¹⁰⁹. For residues *N*-formyl-Met¹–Arg³⁴, as well as Lys¹¹⁰–Ser¹²¹, several amide resonances could not be assigned because of missing signals due to conformational exchange or proton exchange with the solvent.

The ¹H/¹⁵N-HSQC spectrum of NELF E RRM shows the characteristic dispersion of a protein with an intact tertiary structure.

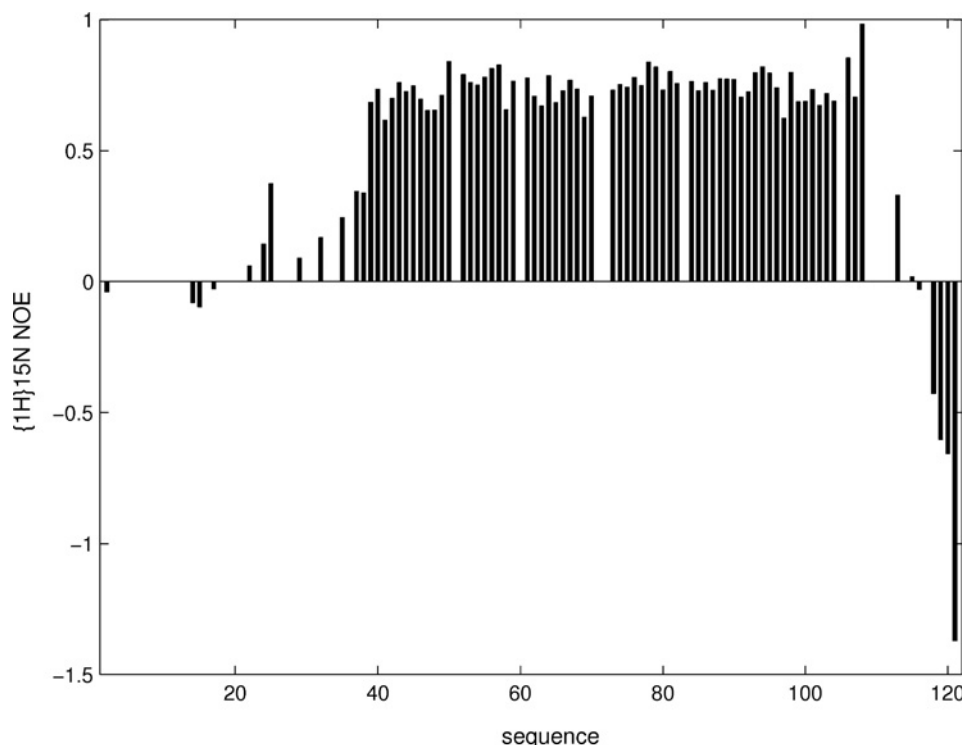


Figure 2 Size of heteronuclear $\{^1\text{H}\} \text{ }^{15}\text{N}$ steady-state NOE at 14.1 T along the amino acid sequence

The values in the range 0.6–0.85 for residues Leu⁴⁰–Cys¹¹⁰ indicate a rigid protein backbone in this region.

Using standard double- and triple-resonance NMR techniques, it was possible to assign nearly all ^1H , ^{13}C and ^{15}N resonances. While for the sequence region Lys³⁸–Arg¹⁰⁹ nearly all (> 95%) backbone resonances and more than 85% of the side-chain resonances could be assigned, the terminal regions, especially the N-terminus, remained partially unassigned owing to strong overlap or missing resonances. The $\{^1\text{H}\} \text{ }^{15}\text{N}$ -heteronuclear steady-state NOE experiment at 14.1 T (Figure 2) shows values around 0.6–0.8 for the region Lys³⁸–Arg¹⁰⁹, while outside of this region the heteronuclear NOE decreases towards the termini. $\{^1\text{H}\} \text{ }^{15}\text{N}$ 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 Gly³⁹ and Arg¹⁰⁹. 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 Lys³⁸–Arg¹⁰⁹ and unstructured termini. Missing assignments of amide resonances can therefore be explained by conformational or solvent exchange. Owing to the flexible character of the termini, residues *N*-formyl-Met¹–Arg³⁴ and Leu¹¹⁴–Ser¹²¹ 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-co-ordinate precision of 0.28 Å for the heavy backbone atoms and 0.64 Å for all heavy atoms for resi-

dues Gly³⁹–Ala¹⁰⁸, 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-favoured regions of the Ramachandran plot (Figure 3A).

The solution structure of NELF E RRM exhibits a compact $\beta\alpha\beta\beta\beta$ fold with a four-stranded antiparallel β -sheet ($\beta_1 = \text{Asn}^{40}$ –Tyr⁴⁵, $\beta_2 = \text{Ile}^{64}$ –Asp⁷⁰, $\beta_3 = \text{Cys}^{75}$ –Tyr⁸⁰, $\beta_4 = \text{Gln}^{102}$ –Ile¹⁰⁷) that packs against two helices ($h_1 = \text{Pro}^{51}$ –Phe⁶¹, $h_2 = \text{Met}^{83}$ –Leu⁹³) which are oriented approximately perpendicular to each other (interhelix angle = $114.3 \pm 1.8^\circ$; Figure 3B). Numerous hydrophobic contacts involving Leu⁴², Val⁴⁴, Leu⁵⁴, Ala⁵⁷, Phe⁵⁸, Val⁷⁸, Ala⁸⁶, Val⁹⁰, Leu⁹³, Val⁹⁸, Val¹⁰⁵ and Ile¹⁰⁷ stabilize this packing. For example, Phe⁵⁸ from h_1 contacts Val⁷⁸ from the central strand β_3 , and Leu⁹³ from h_2 (Figure 3C). These residues are conserved in all other known RRM and adopt a similar conformation. Residues Thr⁹⁶ and Gln⁹⁷ form an additional short β -strand aligned antiparallel to β_4 , thus extending the β -sheet. This is similar to other RRM where a β -hairpin is found in the sequence region between h_2 and β_4 [41]. The highly conserved aromatic residues Tyr⁴³ and Phe⁷⁷ known to be involved in base-stacking interactions with RNA in other RRM are highlighted in Figure 3(D) [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 (Figure 4). Amide ($^1\text{H}^{\text{N}}$, ^{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 -labelled protein provides a powerful method for mapping of the binding interface. Addition of TAR RNA results in remarkable

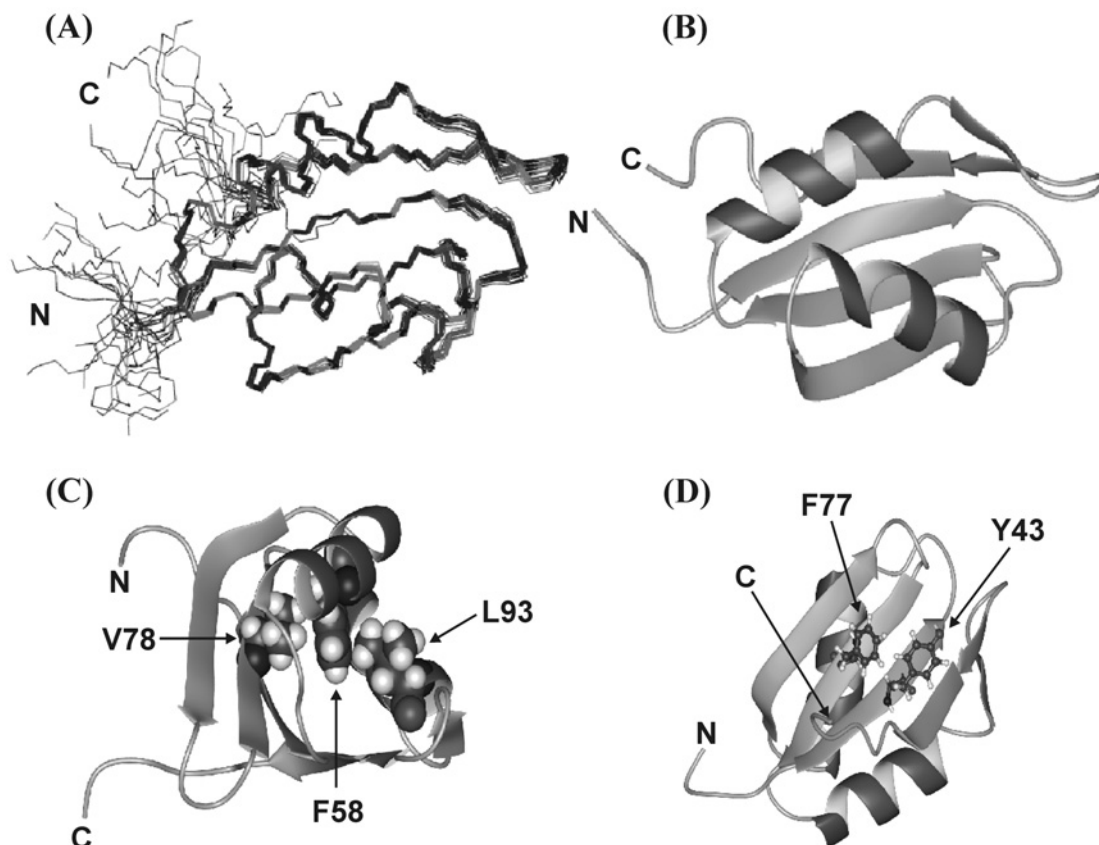


Figure 3 Solution structure of NELF E RRM

(A) Overlay of the 20 structures (residues Ala³⁵–Met¹¹³) 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 Phe⁵⁸, Val⁷⁸ and Leu⁹³. The amino acid residues are displayed as space-filled atoms. (D) The highly conserved aromatic residues Tyr⁴³ and Phe⁷⁷ derived from RNP2 and RNP1 of NELF E RRM are represented as balls and sticks. The N- and C-termini are indicated.

chemical-shift changes for resonances located in the central β -sheet (Asn⁴¹, Leu⁴², Tyr⁴⁵, Cys⁷⁵ and Phe⁷⁷) and in the N-terminal region of h₂ (Glu⁸⁴ and Asp⁸⁷; Figure 4A). For several residues in particular in strand β_4 (Lys¹⁰⁴–Ile¹⁰⁷), signals were absent in the ¹H/¹⁵N-HSQC spectra, probably due to exchange processes on the intermediate time scale, suggesting that these residues are involved in binding. This is characteristic for affinities in the low-micromolar 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 on to the two conserved aromatic residues Tyr⁴³ and Phe⁷⁷ (Figure 4B). Figure 4(C) 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 Met¹¹³ and Ala¹¹⁶ resonances on TAR titration (Figure 4B) imply a structural change of this region which is highly flexible in free NELF E RRM. Structural changes of the C-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 (Thr⁹⁶–Gln⁹⁷) does not exhibit significant chemical-shift changes, indicating that this region is not involved in RNA recognition. This is in contrast with the RRM from TcUBP1 (*Trypanosoma cruzi* U-rich-RNA-binding protein 1), 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 h₂ (Glu⁸¹–Glu⁸⁴ and Asp⁸⁷). 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 h₂ is rather large for the induction of secondary chemical-shift changes. We thus analysed 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. ¹⁵N-relaxation measurements and temperature-dependent titrations, were performed. However, analysis of the data was hampered by severe line broadening due to the large molecular mass of the complex, and only ambiguous data could be obtained (results not shown).

Therefore fluorescence titrations were performed. It has been shown previously that NELF E binds to the lower region of the TAR RNA stem [47]. However, none of the RRM's analysed 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 labelled single or double-stranded oligoribonucleotides representing the lower stem of TAR were used

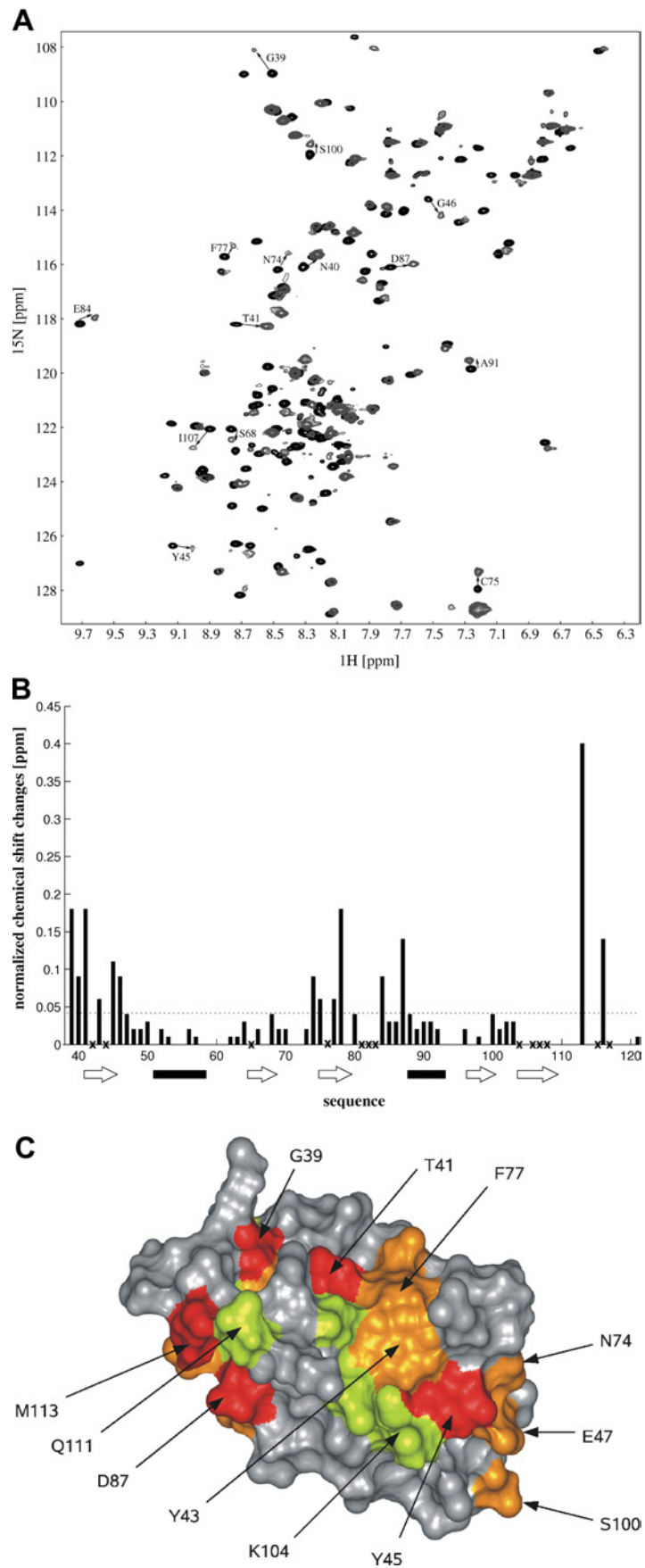


Figure 4 For legend see facing page

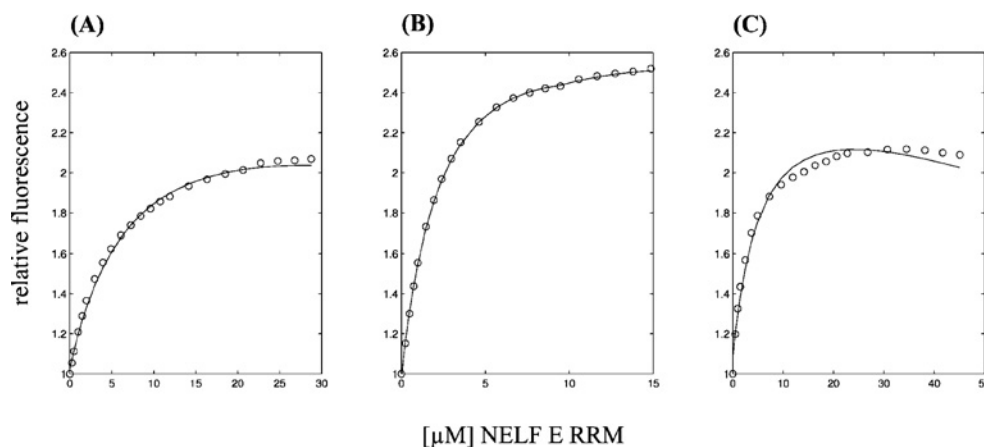


Figure 5 Fluorescence equilibrium titration with the TAR RNA stem

FAM-labelled (A) TAR-(1–18), (B) TAR-(42–57) or (C) double-stranded TAR RNA stem (50 nM) was titrated with NELF E RRM in 10 mM sodium phosphate buffer, 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 K_d values of $8.2 \pm 1.0 \mu\text{M}$ and $2.6 \pm 0.2 \mu\text{M}$ respectively. No curve could be fitted to the data in (C) using a simple two-state model.

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 approx. $8.2 \mu\text{M}$ for TAR-(1–18) and $2.6 \mu\text{M}$ for TAR-(42–57) respectively could be derived (Figures 5A and 5B). Binding to the double-stranded RNA oligomer, however, could not be described by such a simple model (Figure 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 with 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 (root mean square deviation) of 1.3 \AA 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 $\beta\alpha\beta\beta\alpha\beta$ 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 interaction of NELF E RRM with different single- and double-stranded RNAs will contribute to the understanding of 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.

We 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 TFIIF. *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. et al. (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

Figure 4 Determination of the binding interface

(A) Overlay of the $^1\text{H}/^{15}\text{N}$ -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 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 \times . (C) Surface representation of NELF E RRM highlighting the binding interface. Residues with resonances showing significant chemical-shift changes upon TAR binding ($0.04 < \Delta\delta \leq 0.1$), including the highly conserved Tyr⁴³ and Phe⁷⁷ of the RNP2 and RNP1 sequence 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.

- 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
- 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., Volkman, 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. Ser. B.* **112**, 200–205
- 18 Vuister, W. G. and Bax, A. (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond $J(\text{H}^{\text{N}}\text{H}^{\text{C}})$ 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. (1994) Relaxation-rate measurements of ^{15}N - ^1H groups with pulsed-field gradients and preservation of coherence pathways. *J. Magn. Reson. Ser. A* **111**, 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. et al. (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 Ick. *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. *Math. Program.* **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 Reference deleted
- 38 Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51–55
- 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., Rullmann, 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 TcLUBP1, a single RRM domain protein from *Trypanosoma cruzi*: contribution of a β 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

Received 20 March 2006/9 August 2006; accepted 9 August 2006

Published as BJ Immediate Publication 9 August 2006, doi:10.1042/BJ20060421

8. Einzelarbeit B

Jampani N. Rao, Kristian Schweimer, Sabine Wenzel, Birgitta M. Wöhrl und Paul Rösch
NELF-E RRM undergoes major structural changes in flexible protein regions on target RNA
binding.

Biochemistry **47**(12), 3756-61 (2008)

NELF-E RRM Undergoes Major Structural Changes in Flexible Protein Regions on Target RNA Binding[†]

Jampani Nageswara Rao, Kristian Schweimer, Sabine Wenzel, Birgitta M. Wöhrl, and Paul Rösch*

Lehrstuhl Biopolymere and Research Center for Bio-Macromolecules, Universität Bayreuth, 95440 Bayreuth, Germany

Received December 12, 2007; Revised Manuscript Received January 31, 2008

ABSTRACT: The E subunit of the human heterotetrameric negative transcription elongation factor (NELF-E) contains a canonical $\beta\alpha\beta\beta\alpha\beta$ RNA recognition motif (RRM) that binds to a wide variety of RNA sequences. These induce very similar conformational changes in the RRM as determined by nuclear magnetic resonance spectroscopy. Although the RNA binding interface of a canonical RRM is mainly located at its β -sheet surface, for NELF-E RRM large chemical shift perturbations are observed for residues in the flexible C-terminal region and the loop between β_3 and α_2 , and both regions are distant from the interface. We determined the solution structure of single-stranded transactivator responsive element (TAR) RNA-bound NELF-E RRM. This structure clearly shows that RNA binding to NELF-E RRM induces formation of a helix in the C-terminus. The RNA-bound form of NELF-E RRM is very similar to the RNA-bound form of U1A RRM, although the C-terminus of the NELF-E RRM is unstructured in the free protein, whereas it is helical in the U1A protein. Thus, RNA binding to NELF-E RRM induces a conformational change toward the U1A structure, resulting in highly similar RNA binding conformations for both proteins.

The RNA recognition motif (RRM),¹ also known as ribonucleoprotein (RNP) domain or RNA binding domain, is one of the most common RNA binding domains, and proteins containing this domain are involved in most steps of eukaryotic gene expression (1–3). RRM domains typically consist of 90 amino acids and fold into modules with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology, where the four β -strands form an antiparallel β -sheet (4). Whereas one surface of the β -sheet is involved in hydrophobic packing with the two α -helices, the other is the RNA binding surface. RRM domains contain two highly conserved ribonucleoprotein motifs, RNP1 and RNP2, located in the central β -strands. The aromatic amino acids of the RNP motifs are thought to be involved in base-stacking interactions with unpaired RNA bases (3). The RRM target RNAs are diverse in sequence and secondary structure. In general, RRMs use loop regions as well as N- and C-terminal extensions along with variations in the β -sheet surface (5) to obtain RNA binding specificity, but individual RRMs use different mechanisms for RNA recognition depending on their RNA target.

Negative transcription elongation factor (NELF) is a heterotetrameric protein of four subunits, NELF-A, NELF-B, alternatively spliced NELF-C or NELF-D, and NELF-E (6). As a general transcription factor, NELF is involved not only in the regulation of human gene expression (7, 8) but

also in the regulation of HIV-1 transcription by arresting the elongation complex (9, 10). The RNA binding activity of the NELF-E RRM is crucial for the function of NELF, and deletion of NELF-E RRM completely abolishes the function of NELF (11, 12). NELF is recruited to the elongation complex by binding to the nascent transcripts of RNA polymerase II (Rpol II) via the NELF-E RRM (11). The negative effect of NELF on HIV-1 transcription is counteracted by NELF-E phosphorylation and phosphorylation of the C-terminal domain of Rpol II by positive transcription elongation factor b (p-TEFb) which in turn is recruited to the elongation complex via Zn²⁺-dependent interaction with HIV-1 transactivator (Tat) protein (10, 13–15). NELF-E RRM exhibits a canonical topology, that is, a four-stranded antiparallel β -sheet surface packed against two α -helices, and the N- and the C-termini are unstructured in solution (16). Here we explore details of the structural changes induced into the NELF-E RRM upon RNA binding utilizing NMR spectroscopy.

MATERIALS AND METHODS

NMR Spectroscopy. ¹⁵N- and ¹³C,¹⁵N-labeled NELF-E RRM was expressed and purified as described (16). Synthetic RNA oligonucleotides purified by HPLC were purchased from biomers-GmbH, Ulm, Germany. All NMR experiments were performed at 298 or 303 K on Bruker Avance 700 and Avance 800 spectrometers equipped with cryogenic cooled ¹H/¹³C/¹⁵N triple-resonance probes with pulsed-field gradient capabilities. NMR titration experiments were performed on ¹⁵N-labeled NELF-E RRM (0.1 mM) by measuring a series of 2D ¹H–¹⁵N HSQC spectra with addition of unlabeled single-stranded TAR RNA oligonucleotides at various molar ratios, and ¹H–¹⁵N resonance shift and line width changes

[†] Financial support for this work has been provided by the Deutsche Forschungsgemeinschaft (DFG; Ro617/16-1).

* Corresponding author: phone, +49 921 553540; fax, +49 921 16490459; e-mail, roesch@unibt.de.

¹ Abbreviations: HSQC, heteronuclear single-quantum coherence spectroscopy; NELF-E, negative elongation factor subunit E; RRM, RNA recognition motif; RNP, ribonucleoprotein; TAR, transactivator responsive element; Rpol, RNA polymerase; p-TEFb, positive transcription elongation factor b.

were monitored by analysis of these spectra. Normalized chemical shift changes were expressed as the weighted geometric average of ^1H and ^{15}N chemical shift changes for each residue, and normalized chemical shift changes larger than 0.04 ppm were considered significant (31).

The dissociation constants K_d were determined from the changes of chemical shifts of ^{15}N -labeled NELF-E RRM observed in a ^1H - ^{15}N HSQC spectrum after gradual addition of the respective unlabeled binding partners. Changes of chemical shift of signals in the fast-exchange limit were fitted to a two-state model.

$$\delta_{\text{obs}} = \delta_{\text{P}} + (\delta_{\text{PL}} - \delta_{\text{P}}) \left[\frac{K_{\text{D}} + (1+r)[\text{P}]_0}{2[\text{P}]_0} - \frac{\sqrt{(K_{\text{D}} + (1+r)[\text{P}]_0)^2 - 4[\text{P}]_0^2 r}}{2[\text{P}]_0} \right] \quad (1)$$

δ_{obs} , δ_{P} , and δ_{PL} are the chemical shifts for the actual mixture, the free protein, and the completely bound protein, respectively. $[\text{P}]_0$ is the total concentration of NELF-E RRM, and r describes the NELF-E RRM/RNA ratio.

To obtain the three-dimensional structure of RNA-bound NELF-E RRM, standard double and triple resonance NMR experiments were recorded. All NMR spectra for the structure determination were recorded at 303 K with a 1:1 complex of 0.5 mM uniformly ^1H , ^{15}N - and ^{13}C -labeled NELF-E RRM and 0.5 mM unlabeled single-stranded TAR49–57 in buffer containing 10 mM sodium phosphate, pH 6.9, 100 mM NaCl, 1 mM DTT, 10% D_2O , and 0.03% of NaN_3 . Distance restraints for structure calculation were derived from three-dimensional ^{13}C - and ^{15}N -edited NOESY-HSQC experiments with mixing times of 120 ms (36, 37). NMR data were processed using in-house written software and analyzed with the program package NMRview 5.2.2 (32).

Structure Determination. Backbone resonances of NELF-E RRM in complex with TAR49–57 were mainly assigned by following the resonances during the titration. Except for the resonances involved in RNA binding, all resonances could be assigned directly since they did not show any chemical shift perturbations. The resonances exhibiting large chemical shift changes were assigned by observing the characteristic NOEs in ^{15}N NOESY-HSQC spectra. Many protein resonances in the complex gave chemical shifts and patterns of NOE cross-peaks similar to those observed for the free protein. Thus, most of the resonances were assigned by comparing the constant time ^1H - ^{13}C HSQC spectra of the free and RNA-bound NELF-E RRM. Side chain resonance assignments for residues involved in RNA binding were achieved by following the trace of side chain protons in 3D HCCH-TOCSY and identifying characteristic NOEs in ^{13}C NOESY-HSQC spectra. NOE cross-peaks identified in the 3D ^{15}N and ^{13}C NOESY-HSQC spectra in an iterative procedure as well as hydrogen bonds and $^3J_{\text{HNH}}$ scalar coupling constants were converted into distance and backbone torsion angle restraints, respectively, as described (16, 33). In total, 1929 distance restraints, 32 dihedral angles, and 24 hydrogen bonds were used for the structure calculation (Supporting Information, Table 1). All structure calculations were performed with the program X-PLOR 3.8.5.1 using the three-step simulated annealing protocol (33, 34) with floating assignment of prochiral groups (35) as described (16). The structural

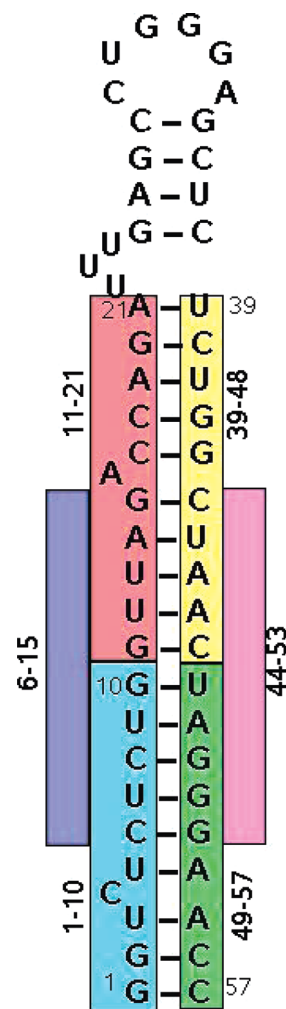


FIGURE 1: Secondary structure of HIV-1 TAR RNA. Sequences of RNA oligonucleotides used for the binding studies with NELF-E RRM are shown in different colors.

coordinates were deposited in the Protein Data Bank (PDB accession code 2jx2).

RESULTS AND DISCUSSION

RNA Binding Studies. To understand the RNA recognition process by NELF-E RRM in detail and to study the general features of RNA recognition by this protein motif, we performed NELF-E RRM-RNA binding studies by NMR using various single-stranded RNA oligonucleotides derived from the stem region of HIV-1 TAR RNA (Figure 1). Amide proton chemical shifts are usually a very sensitive probe to study side chain perturbations by direct ligand interaction or by protein conformational changes on ligand binding. Addition of unlabeled TAR1–10 to ^{15}N -labeled NELF-E RRM led to chemical shift changes in the ^1H - ^{15}N HSQC spectra for the residues located in the β -sheet surface of NELF-E RRM (Figure 2), most remarkably for L42, V44, C75, and V78 in the central β -strands but also for E81, K82, and M83 in the loop between β_3 and β_2 as well as M113, L114, D115, A116, and A117 in the C-terminus. Unambiguous assignment of the amide and nitrogen resonances was facilitated as all resonances were in the fast-exchange regime on the NMR time scale and thus could be traced easily during titration. Although NELF-E RRM binds to all TAR RNA oligonucleotides used in this study, binding resulted in

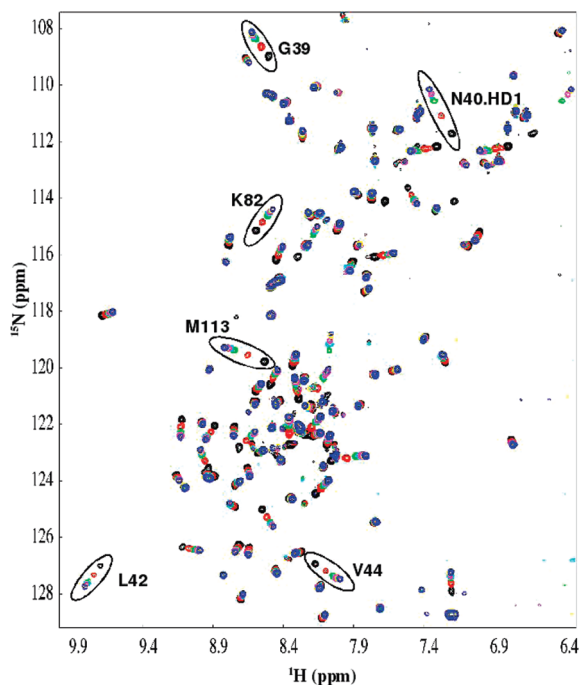


FIGURE 2: Titration of NELF-E RRM with TAR1–10. Overlay of ^1H – ^{15}N HSQC spectra recorded during the titration with different RNA/protein ratios. Key: black, 0.0; red, 0.25; green, 0.5; cyan, 0.75; magenta, 1.0; yellow, 2.0; blue, 3.0. Those amide resonances showing the strongest chemical shift changes are indicated.

different dynamics of the system as evidenced by different chemical exchange kinetics (Figure 3).

Normalized Chemical Shift Changes. In all titration experiments, regardless of the RNA oligomer added, the same set of amide resonances showed chemical shift perturbations, although with different magnitude. This clearly points to a general structural rearrangement of the NELF-E RRM on RNA binding, independent of the specific sequence of the oligomer. In particular, backbone amide protons and nitrogens located in the β -strands, the loop between β_3 and α_2 , and the C-terminal region were strongly influenced, e.g., 0.185 and 0.308 ppm in V44 in the TAR1–10 and the TAR49–57 complexes, respectively. TAR11–21 and TAR49–57 induced the largest chemical shift changes, and the final chemical shift changes in the TAR44–53 could not be detected due to severe resonance broadening. Titration of ^{15}N -labeled NELF-E RRM with TAR1–10, TAR6–15, and TAR39–48 allowed the observation of chemical shift changes for resonances under fast-exchange conditions for amino acids located in the RNA binding region of NELF-E RRM. From these data, K_d values of 21 μM , 35 μM , and 55 μM could be determined (Supporting Information). Thus, NELF-E RRM binds to all RNA oligonucleotides studied in similar fashion with K_d s in the micromolar range, but it does so with differences in detail. The ability of NELF-E RRM to bind a wide variety of single-stranded RNAs, albeit with different affinities, is not surprising, since NELF is involved in regulation of the expression of a wide variety of genes, and the nascent transcripts of Rpol II differ widely in their sequences (11).

Mapping of RNA Binding Interface. The RNA binding surface of NELF-E RRM as obtained by chemical shift mapping reveals subtle differences as compared to binding interfaces of other RRMs (5, 17–19) (Figures 4 and 5). Like other RRMs, NELF-E RRM uses the β -sheet surface and

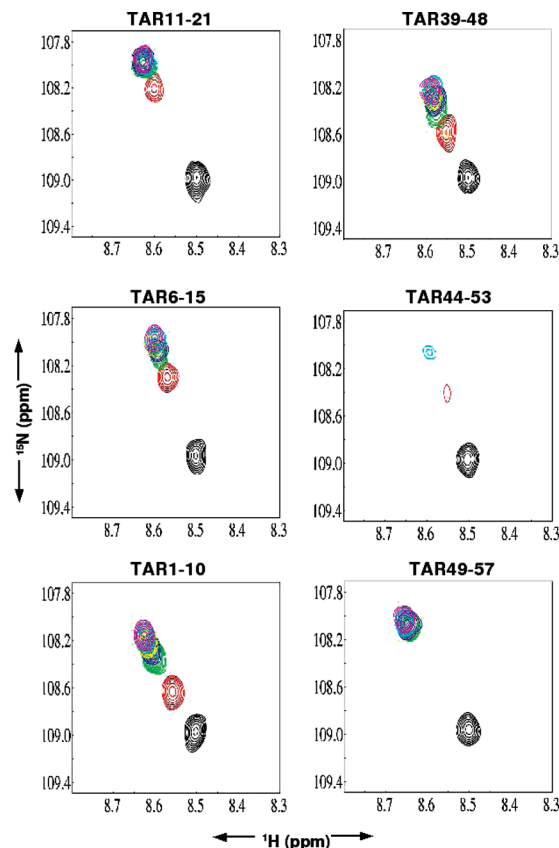


FIGURE 3: Interaction of NELF-E RRM with single-stranded TAR RNA oligonucleotides. Overlays of the G39 amide cross-peak in ^1H – ^{15}N HSQC spectra at each titration step are shown in different colors. The color codes and RNA/protein ratios are as in Figure 2.

the highly conserved RNP1 and RNP2 motifs to provide a platform for RNA recognition. Usually, if only a single RRM domain is present in a protein, this RRM uses loop regions to provide additional RNA contacts (20–22). In the U1A RRM snRNA complex, for example, loop 3 forms an extensive hydrogen bond network and provides hydrophobic contacts with RNA. NELF-E RRM, however, does not show any significant chemical shift perturbation in the loop 3 region upon addition of RNA. Thus, the detailed mode of RNA binding of NELF-E RRM is clearly different than the mode of RNA binding of the U1A RRM.

Addition of RNA to NELF-E RRM results in remarkable chemical shift changes for the resonances located in the N-terminal part of α_2 and the loop between β_2 and β_3 . The side chain resonances of N40, located close to this loop, show a significant chemical shift change after RNA addition, although the N40 side chain is located opposite to the RNA binding β -sheet surface, indicating conformational changes in regions of NELF-E RRM that are not directly involved in RNA binding. The same is true for the N-terminal part of α_2 ; i.e., the amino acid side chains are not close to the binding surface but are still perturbed on RNA binding.

Structure of RNA-Bound NELF-E RRM. Titration of NELF-E RRM with TAR49–57 induced the largest chemical shift changes in the protein and was thus studied more closely. All distance restraints observed for free NELF-E RRM (16) were also found in the TAR49–57 complex, indicating that the structures of free and bound protein were highly similar for amino acids 39–107. Forty additional

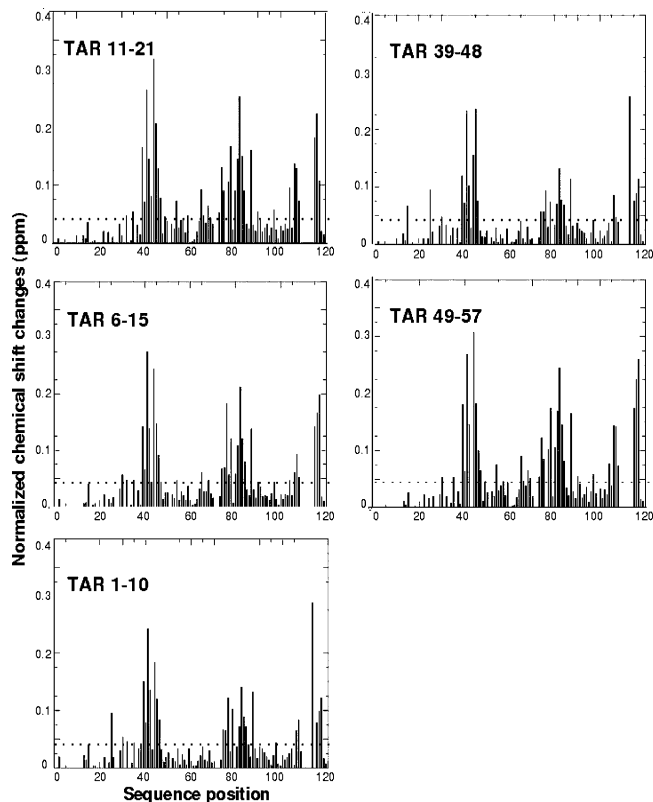


FIGURE 4: Normalized weighted chemical shift changes for NELF-E RRM upon binding to RNA oligonucleotides. Normalized chemical shift changes larger than 0.04 ppm are indicated by a dashed line.

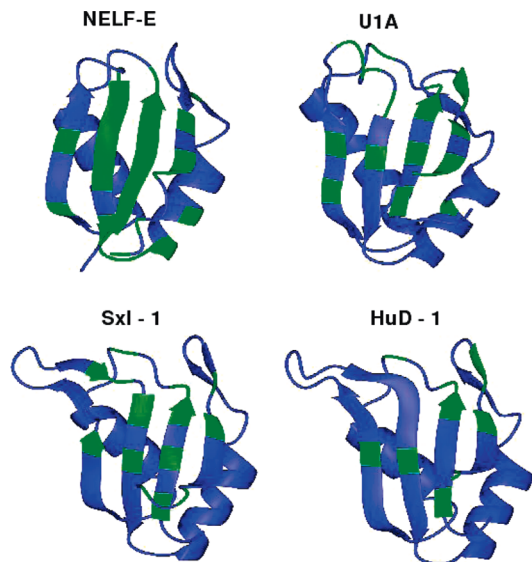


FIGURE 5: Mapping of RNA binding interface. Ribbon representation of RRM structures [NELF-E (2bz2), U1A (1urn), Sxl (1b7f), HuD (1fxl)] in the complex with RNA are known. Residues involved in RNA binding are highlighted in green. For NELF-E RRM the amide chemical shift perturbations (those residues showed normalized chemical shifts larger than 0.04 ppm upon binding to TAR49–57) were used to map the RNA binding interface.

interresidual NOEs were observed for NELF-E RRM in the complex, these additional NOEs being mostly from resonances in the C-terminal and the β_3 – α_2 loop regions. Also, significant changes were observed in the NOE pattern of M83 and Q111 on RNA binding, the bound protein showing strong NOEs (0.3 Å) between the γ -protons of M83, Q111 and the

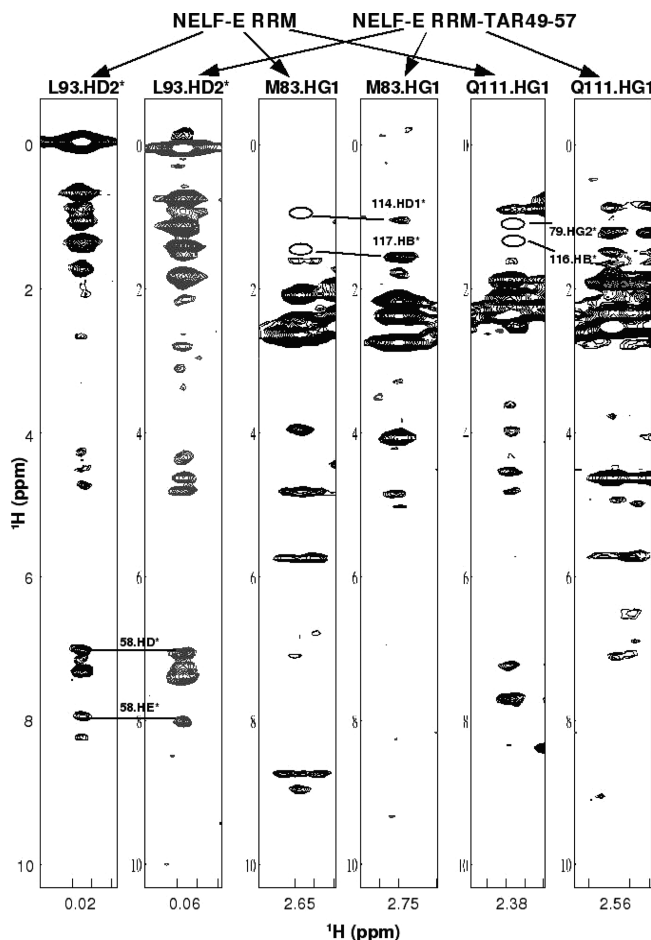


FIGURE 6: Comparison of ^{13}C NOESY-HSQC spectra of free and RNA-bound NELF-E RRM. Interresidual NOEs observed between the C-terminus and around the β_3 – α_2 loop region are labeled in the NOESY slices of the NELF-E RRM–TAR49–57 complex. The corresponding NOEs not present in the NOESY spectra of free NELF-E RRM are indicated by open circles.

methyl group protons of A117, T79 (Figure 6), indicating a conformational change of the C-terminus of the NELF-E RRM upon binding to TAR49–57.

The final structure calculation resulted in an ensemble of 20 structures showing no distance restraint violation larger than 0.2 Å, no violation of a dihedral angle restraint larger than 3°, and only small deviations from the idealized covalent bond geometry. As expected, the overall global fold of NELF-E RRM remains intact in the complex (Figure 7). The C-terminus of NELF-E RRM, however, is undergoing a major conformational change. Whereas it is highly flexible in the free state, it is stabilized by numerous hydrophobic contacts (I65, T79, M83, M113, L114, A116, A117, and T118) in the RNA-bound conformation, forming a short 3_{10} helix from M113 to D115, close to the β_3 – α_2 loop. This structure formation correlates perfectly well with the large chemical shift changes observed in this region on RNA binding. The C-terminus of NELF-E RRM is located close to the highly conserved aromatic amino acids in the RNP motif and is thus likely involved in RNA binding (Figure 7c), and the basic residues located C-terminally to β_4 (R109, K110) could be involved in electrostatic interaction with the phosphate backbone of the RNA. These structural rearrangements involving the C-terminus reinforces the concept of

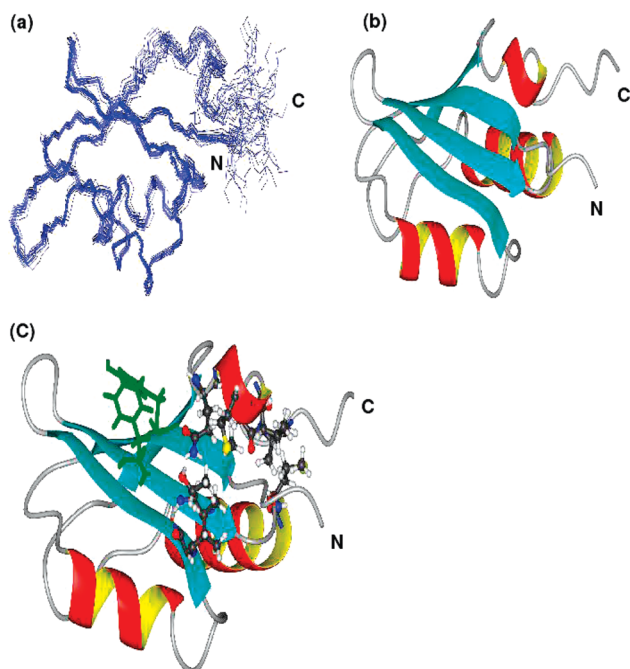


FIGURE 7: Structure of RNA-bound NELF-E RRM. (a) Overlay of the 20 accepted structures, A35–S121, showing the lowest values of the target function. (b) Schematic representation of the lowest energy structure. N- and C-termini are indicated. (c) Residues involved in hydrophobic interactions to stabilize the C-terminus of NELF-E RRM (ball and stick) and the highly conserved aromatic amino acids located in RNP1 and RNP2 (green sticks).

binding by induced fit which is observed in many other RNA binding proteins (23–26).

Structural Differences in the C-Terminal Regions of RRM. The C-terminal extensions of RRM domains are often key in RNA recognition, in several cases forming a helix in the free state that is stabilized by hydrophobic interactions with RNP motifs (4, 27, 28). The C-terminal helix of U1A, CstF64, and hnRNP-F RRMs masks the RNA binding β -sheet surface in the absence of the target (Figure 8) (29). In U1A RRM the C-terminal helix is repositioned on target binding such that the RNP motifs are accessible for the RNA and, at the same time, the C-terminal helix can also contribute to the specific target RNA hairpin recognition (20). Even though the structure of the CstF64 RRM–RNA complex is unknown, it was predicted from NMR relaxation and chemical shift data that the C-terminal helix of CstF64 RRM unfolds upon binding to RNA, rendering the RNP motifs accessible for RNA recognition (30). Quasi-RRMs of hnRNP-F, however, recognize RNA without undergoing major conformational changes in the C-terminal helix that also maintains hydrophobic contacts with the RNP motifs (28). Thus, the RNA-bound conformations of NELF-E RRM and U1A are very similar even though they are different in the free state (Figure 8).

In conclusion, a wide variety of RNA sequences binds to the NELF-E RRM and induces very similar conformational changes in the protein on binding. While the RNA binding interface is mainly located at the β -sheet surface, rearrangement of residues located in the flexible C-terminal region and the loop between β_3 and α_2 is taking place on binding. Comparison of the free vs the RNA-bound conformation of NELF-E RRM with other RRMs from the known complexes indicates the mechanism of RNA recognition to be unique

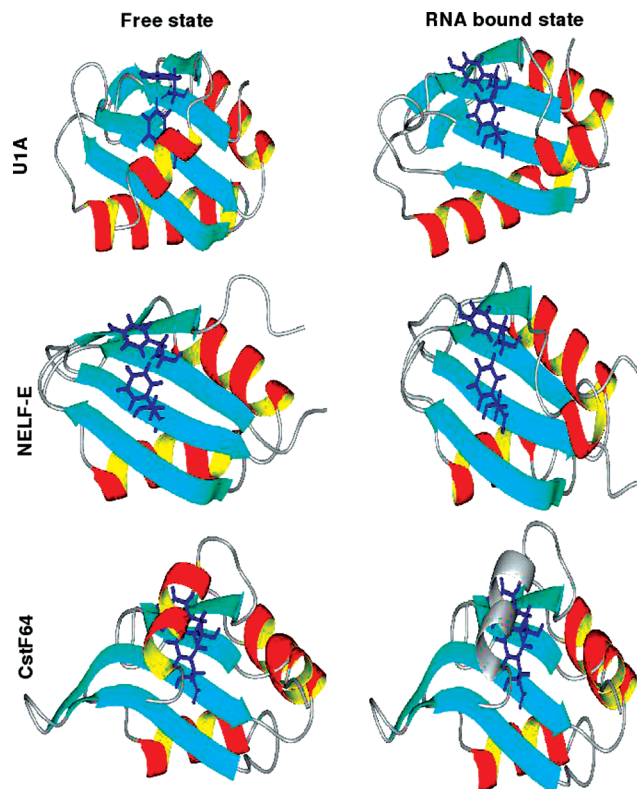


FIGURE 8: Comparison of free and RNA conformations of RRMs. The canonical RNA binding interface is indicated by showing the conserved aromatic amino acids in stick representation. The C-terminal helices are located at the solvent-exposed side of the antiparallel β -sheet surface. The unfolding of the C-terminal helix in the RNA-bound state of CstF64 RRM is represented in gray color. PDB codes for free and RNA-bound RRM are as follows: NELF-E, 2bz2 and 2jx2; U1A, 1fht and 1urn; CstF64, 1p1t.

for NELF-E, but a complete three-dimensional structure of the NELF-E RRM–RNA complex would be needed for the detailed understanding of these differences at the atomic scale.

SUPPORTING INFORMATION AVAILABLE

Structural statistics and determination of K_d values based on chemical shift changes upon RNA titration. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

- Bandziulis, R. J., Swanson, M. S., and Dreyfuss, G. (1989) RNA-binding proteins as developmental regulators. *Genes Dev.* 3, 431–437.
- Kenan, D. J., Query, C. C., and Keene, J. D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.* 16, 214–220.
- 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.
- Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* 348, 515–520.
- 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.
- 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.

7. Wu, C. H., Yamaguchi, Y., Benjamin, L. R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., and Gilmour, D. (2003) NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev.* *17*, 1402–1414.
8. Aida, M., Chen, Y., Nakajima, K., Yamaguchi, Y., Wada, T., and Handa, H. (2006) Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the junB gene. *Mol. Cell. Biol.* *26*, 6094–6104.
9. 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.
10. 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.
11. 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.
12. 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.
13. Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H., and Jones, K. A. (1998) A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* *92*, 451–462.
14. Garber, M. E., Wei, P., KewalRamani, V. N., Mayall, T. P., Herrmann, C. H., Rice, A. P., Littman, D. R., and Jones, K. A. (1998) The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev.* *12*, 3512–3527.
15. Peterlin, B. M., and Price, D. H. (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* *23*, 297–305.
16. Rao, J. N., Neumann, L., Wenzel, S., Schweimer, K., Rosch, P., and Wohrl, B. M. (2006) Structural studies on the RNA-recognition motif of NELF E, a cellular negative transcription elongation factor involved in the regulation of HIV transcription. *Biochem. J.* *400*, 449–456.
17. 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.
18. 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.
19. Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and Xu, R. M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* *13*, 1102–1115.
20. Allain, F. H., Howe, P. W., Neuhaus, D., and Varani, G. (1997) Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* *16*, 5764–5772.
21. 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.
22. Fleming, K., Ghuman, J., Yuan, X., Simpson, P., Szendroi, A., Matthews, S., and Curry, S. (2003) Solution structure and RNA interactions of the RNA recognition motif from eukaryotic translation initiation factor 4B. *Biochemistry* *42*, 8966–8975.
23. Puglisi, J. D., Chen, L., Blanchard, S., and Frankel, A. D. (1995) Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science* *270*, 1200–1203.
24. Faber, C., Scharpf, M., Becker, T., Sticht, H., and Rosch, P. (2001) The structure of the coliphage HK022 Nun protein-lambda-phage boxB RNA complex. Implications for the mechanism of transcription termination. *J. Biol. Chem.* *276*, 32064–32070.
25. Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., and Williamson, J. R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* *273*, 1547–1551.
26. Legault, P., Li, J., Mogridge, J., Kay, L. E., and Greenblatt, J. (1998) NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell* *93*, 289–299.
27. Perez Canadillas, J. M., and Varani, G. (2003) Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J.* *22*, 2821–2830.
28. Dominguez, C., and Allain, F. H. (2006) NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. *Nucleic Acids Res.* *34*, 3634–3645.
29. Avis, J. M., Allain, F. H., Howe, P. W., Varani, G., Nagai, K., and Neuhaus, D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.* *257*, 398–411.
30. Deka, P., Rajan, P. K., Perez-Canadillas, J. M., and Varani, G. (2005) Protein and RNA dynamics play key roles in determining the specific recognition of GU-rich polyadenylation regulatory elements by human Cstf-64 protein. *J. Mol. Biol.* *347*, 719–733.
31. 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.
32. 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.
33. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* *229*, 317–324.
34. 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.
35. 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.
36. Ikura, M., Kay, L. E., Tschudin, R., and Bax, A. (1990) Three-dimensional NOESY-HMQC spectroscopy of a ¹³C-labeled compound. *J. Magn. Reson.* *86*, 204–209.
37. Talluri, S., and Wagner, G. (1996) An optimized 3D NOESY-HSQC. *J. Magn. Reson., Ser. B* *112*, 200–205.

BI702429M

Supplementary data

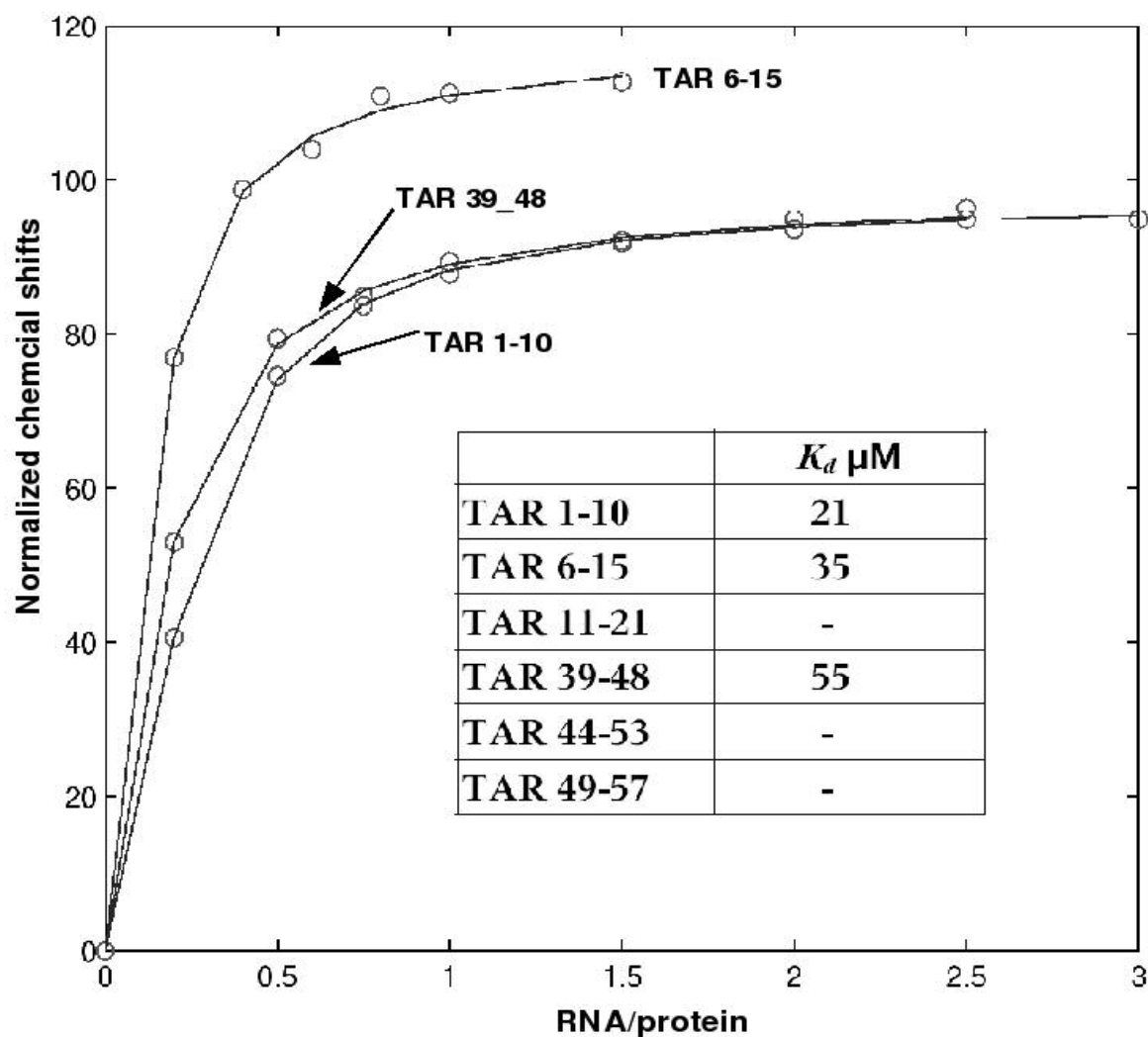
Table 1 Structural statistics. E_{pot} , overall potential energy; E_{impr} , potential energy of improper angle; E_{cdih} , potential energy for dihedral restraints; E_{sani} , potential energy for RDC restraints.

experimental restraints		
distance restraints	total	1929
	intraresidual	468
	sequential	423
	medium-range	289
	long-range	638
dihedral angles		32
dipolar couplings		55
hydrogen bonds (two restraints each)		24
molecular dynamics statistics		
energies (kcal/mol)		
	E_{pot}	12.9 ± 1.4
	E_{bond}	0.73 ± 0.06
	E_{angle}	7.9 ± 0.6

experimental restraints		
	E_{impr}	2.1 ± 0.2
	E_{repel}	0.79 ± 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	0.00074 ± 0.00004
	distance restraints	0.0035 ± 0.0007
RMSDs from ideal angles (deg)	bond angles	0.14 ± 0.06
	dihedral angle restraints	0.15 ± 0.15
RMSDs from dipolar couplings (Hz)		0.11 ± 0.02
atomic coordinate precision (RMSD) (Å)		
backbone heavy atoms		0.69 (Gly39-Thr118)
heavy atoms		1.19 (Gly39-Thr118)
Ramachandran plot statistics		
residues in		
	most favored regions	88.0%

experimental restraints		
	allowed regions	12.0%

Fig. Dissociation constants for NELF-E RRM:TAR complexes. Fitting of the curves yielded the calculated K_d values shown in the inset. Normalized chemical shift changes of G39 were used.



9. Einzelarbeit C

Sabine Wenzel, Kristian Schweimer, Paul Rösch und Birgitta M. Wöhrl

The small hSpt4 subunit of the human transcription elongation factor DSIF is a Zn-finger protein with alpha/beta type topology.

Biophysical and Biochemical Research Communications **370**(3):414-8 (2008)



Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

The small hSpt4 subunit of the human transcription elongation factor DSIF is a Zn-finger protein with α/β type topology

Sabine Wenzel, Kristian Schweimer, Paul Rösch, Birgitta M. Wöhrl *

Universität Bayreuth, Lehrstuhl für Struktur und Chemie der Biopolymere & Research Center for Biomacromolecules, 95440 Bayreuth, Germany

ARTICLE INFO

Article history:

Received 12 March 2008

Available online 26 March 2008

Keywords:

DSIF

hSpt4

Transcription elongation factor

HIV

Zn-finger protein

NMR

CD

ABSTRACT

The eukaryotic transcription elongation factor 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) sensitivity inducing factor (DSIF), is involved in regulating the processivity of RNA polymerase II. DSIF plays also a role in transcriptional activation, and in concert with the negative elongation factor NELF causes promoter proximal pausing of RNA polymerase II. Furthermore, DSIF has also been implicated in regulating the transcription of the human immunodeficiency virus proviral DNA. Human DSIF is composed of the two subunits, hSpt4 (p14) and hSpt5 (p160), corresponding to the yeast homologs Spt4 and Spt5. Here we show the purification and characterization of the small subunit, hSpt4. We were able to purify the protein in a soluble form separately from the larger hSpt5 subunit. CD and NMR spectroscopy show that the purified protein hSpt4 exhibits an α/β topology with a well defined tertiary structure. Furthermore metal analysis by ICP-OES indicates that the protein contains a functional 4-Cys Zn-finger.

© 2008 Elsevier Inc. All rights reserved.

Transcription elongation is an important step in gene expression which is highly regulated both, in prokaryotes and eukaryotes. In recent years in eukaryotes a variety of factors have been discovered that contribute to this regulatory process. One of these factors is the transcription elongation factor 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) sensitivity inducing factor (DSIF), which regulates RNA polymerase II (RNAPol II) processivity together with the negative elongation factor NELF.

DSIF is highly conserved among eukaryotes and exhibits ambivalent properties since it can function as both, a positive as well as a negative factor of elongation. Together with NELF it induces promoter proximal transcriptional pausing of RNAPol II which can only be overcome by the action of the positive transcriptional elongation factor b (pTEFb). It is assumed, that this event is critical for mRNA maturation [25].

Abbreviations: CD, circular dichroism; DSIF, 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) sensitivity inducing factor; *E. coli*, *Escherichia coli*; GB1, streptococcal immunoglobulin-binding domain of protein G; HIV, human immunodeficiency virus; ICP-OES, inductively coupled plasma optical emission spectrometry; IPTG, isopropyl-thiogalactoside; LTR, long-terminal repeat; NELF, negative elongation factor; NGN, NusG N-terminal; NMR, nuclear magnetic resonance; NTA, nitrilotriacetic acid; PMF, peptide mass fingerprint; pTEF b, positive transcription elongation factor b; RNAPol II, RNA polymerase II.

* Corresponding author. Address: Universität Bayreuth, Lehrstuhl Biopolymere, Universitätsstr. 30, D-95447 Bayreuth, Germany. Fax: +49 921 55 3544.

E-mail address: birgitta.woehr@uni-bayreuth.de (B.M. Wöhrl).

The human DSIF is a heterodimeric protein composed of the two subunits hSpt4 and hSpt5, which are homologs of the yeast Spt4 and Spt5 transcription factors. [24,29]. hSpt4 is a 14 kDa protein which harbors four conserved cysteine residues potentially forming a Zn-finger [5,15]. hSpt 5, a 160 kDa protein, is comprised of an N-terminal acidic region, a NusG N-terminal (NGN) homology domain followed by six KOW motifs and two C-terminal repeats [21,29]. The NGN domain and KOW motifs show homologies to the prokaryotic transcription factor NusG which affects transcriptional termination and antitermination. Furthermore, the NGN domain harbors the binding region for the small hSpt4 subunit. DSIF interacts directly with RNAPol II during transcription via the KOW motifs of hSpt5 [12,23,29].

DSIF has also been implicated in regulating the transcription of the proviral DNA of human immunodeficiency virus (HIV) [3,7,8,19,20]. Cell culture experiments show that addition of DRB inhibits HIV expression [6]. In the absence of the viral transactivation factor Tat only the 60 nucleotides long TAR RNA is produced by RNAPol II. DSIF and NELF bind in conjunction to the TAR-RNA/RNAPol II complex which results in transcriptional pausing. Once the cellular factor pTEFb is recruited to the TAR RNA by the viral Tat protein, the cyclin dependent kinase 9 (CDK9) subunit of pTEFb phosphorylates the C-terminal repeats of hSpt5 thus converting DSIF from a negative into a positive factor [1,2,11,16,28].

Structural data of DSIF could provide new information on complex formation and the mechanism of binding to RNAPol II. The

human Spt4 homolog was identified in 1996 [10]. So far, attempts to purify soluble hSpt4 separately have failed. However, purification of the DSIF hSpt4/hSpt5 complex was successful [30].

In this report, we show that the hSpt4 subunit can be purified separately from the large hSpt5 subunit after expression in *Escherichia coli*. Our analyses indicate that the purified hSpt4 contains a Zn-finger. CD and NMR spectroscopy demonstrate that the protein exhibits a defined three-dimensional structure with α/β topology.

Materials and methods

Plasmid construction. A synthetically produced gene adapted to the *E. coli* codon usage harboring the DSIF hSpt4 gene was cloned into the expression vector pET15b (Novagen) via its NdeI and BamHI restriction sites. The final plasmid pET15b-hSpt4 encodes an N-terminal 6 \times His-tagged fusion protein with a thrombin cleavage site.

Additionally, the DSIF hSpt4 gene was amplified by PCR with a 5' primer including an NcoI restriction site (underlined): 5'-TTC CAT GGG CGC GTT GGA AAC CGT ACC GEE AG-3' and a 3' primer with a BamHI site (underlined): 5'-TTA GCA GCC GGA TCC TTA TG-3'. The fragment was cloned into the expression vector pET-GB1a (G. Stier, EMBL, Heidelberg, Germany) via NcoI and BamHI. The expressed protein is N-terminally fused to the 56 residues long streptococcal immunoglobulin-binding domain of protein G. The 22 kDa fusion protein harbors a 6 \times His tag at the N-terminus and a TEV protease cleavage site between GB1 and hSpt4.

Protein expression. The plasmids pET15b-hSpt4 and pET-GB1-hSpt4 were transformed into the *E. coli* strain BL 21 (DE3) (Novagen). Cells were grown at 37 °C in LB medium to an optical density of 600 nm (OD₆₀₀) of 0.8. Expression of the recombinant gene was then induced by the addition of 0.1 or 1 mM isopropyl-thiogalactoside (IPTG) followed by incubation at 37 °C for 4 h or 20 °C over night. Aliquots were taken at several time points to monitor the induction level. Cells were harvested by centrifugation at 5000g for 10 min at 4 °C. For expression in minimal medium to label the protein specifically with ¹⁵N for NMR analysis, the same procedure was performed in M9 medium [22] containing 2 mM MgSO₄, 10 μ M Fe(III)citrate, 0.1 mM CaCl₂, 4 g/l glucose, trace element solution TS2 [17], MEM vitamin solution (Gibco, Invitrogen, Karlsruhe, Germany), 50 μ M ZnCl₂, 30 μ g/ml kanamycin, and 1.5 g/l (¹⁵NH₄)₂SO₄ as the sole nitrogen source.

Ni-NTA-affinity purification. Cells were resuspended in binding buffer (25 mM Tris/HCl, pH 7.4, 500 mM NaCl, 5 mM imidazole, 10 mM β -mercapto-ethanol, and 4 g/l benzamidine). After three freeze-thaw cycles, lysozyme, DNase I and one protease inhibitor cocktail tablet (Complete, EDTA-free, Roche Diagnostics GmbH, Mannheim) were added and the suspension was stirred on ice for 30 min. After sonication (6 \times 45 s, pulse 0.9, 100% amplitude), cells were centrifuged at 19,100g for 30 min at 4 °C. The pellet was resuspended in binding buffer, stirred on ice and sonicated and centrifuged again as described above. Purification of the fusion protein was performed by a step gradient using nickel ion affinity (5ml Ni-NTA column, Qiagen, Hilden, Germany) chromatography.

TEV cleavage of GB1-hSpt4. Fractions containing GB1-hSpt4 were pooled, dialyzed against 25 mM Tris/HCl, pH 7.4, 500 mM NaCl, 10 mM β -mercapto-ethanol and cleaved with 1 mg TEV protease per 10 mg fusion protein over night at 4 °C. Proper TEV cleavage was analyzed by 19% SDS-PAGE. To remove the GB1 tag, the sample was again subjected to nickel ion affinity chromatography under the conditions described above. The cleaved hSpt4 was collected in the flow through.

Ion exchange chromatography. The flow through was dialyzed in steps to remove the NaCl in a buffer containing 25 mM Tris/HCl, pH 7.4, 10 mM β -mercapto-ethanol, and decreasing concentrations of NaCl.

The dialyzed sample was loaded on a 5 ml Q XL column connected in tandem to a 5 ml SP XL column (GE Healthcare, Munich, Germany). The protein was detected in the flow through and dialyzed against 25 mM Tris/HCl, pH 7.4, 50 mM NaCl, 10 mM β -mercapto-ethanol, followed by dialysis against 25 mM Tris/HCl, pH 6.9, 50 mM NaCl, 10 mM β -mercapto-ethanol. In the third dialysis step the buffer consisted of 25 mM Tris/HCl, pH 6.9, and 10 mM β -mercapto-ethanol. The dialyzed sample was concentrated in Vivaspin concentrators MWCO 5000 (Vivascience, Stonehouse, UK).

Peptide mass fingerprint (PMF) analysis. Protein bands of ca. 1 mm \times 3 mm were excised from 19% SDS-polyacrylamide gels and the integrity and identity of hSpt4 was confirmed by peptide mass fingerprinting (ZMMK Köln, Zentrale Bioanalytik, Germany).

Metal analysis. Detection of Zn²⁺- and Ni²⁺-ions in a solution of purified hSpt4 was performed by ICP-OES with a Varian Vista-Pro radial device (BayCEER—Chemische Analytik, Universität Bayreuth, Germany).

Circular dichroism. Circular dichroism measurements were performed on a Jasco J 810 spectropolarimeter (Japan Spectroscopic, Groß-Umstadt, Germany) with a Jasco CDF 426S Peltier thermally controlled cuvette holder interfaced with a Julabo F200 water bath (Julabo Labortechnik GmbH, Seelbach, Germany). CD spectra in the far-UV region (260–190 nm) were recorded with 19 μ M protein solutions at 20 °C in a 0.1 cm cell and a bandwidth of 1 nm. The measured spectra were accumulated eight times and corrected for the solvent. The protein was dissolved in water (pH 7.4). For secondary structure predictions based on the CD data the program CDSSTR (Dichroweb) [14,27] was used.

NMR spectroscopy. NMR samples were prepared using the concentrated ¹⁵N-labeled sample in 25 mM Tris/HCl, pH 6.9, 10 mM β -mercapto-ethanol, and 10% D₂O. Final sample concentration was determined by measuring the absorption of the protein at 280 nm using the theoretically determined extinction coefficient $\epsilon = 18700 \text{ M}^{-1} \text{ cm}^{-1}$. Two-dimensional ¹H-¹⁵N HSQC NMR spectra were recorded at 298 K according to the FHSQC scheme [18] on a Bruker DRX 600 MHz NMR spectrometer equipped with pulsed field-gradient capabilities.

Results and discussion

Optimization of expression

It has been shown previously that hSpt4 can be purified in complex with hSpt5. Attempts to express and purify separate hSpt4 from *E. coli* extracts were not successful due to insolubility and aggregation of the protein. Thus, so far, hSpt4 could only be purified in complex with hSpt5 [24,30]. Therefore no structural or functional analysis of free hSpt4 has been possible.

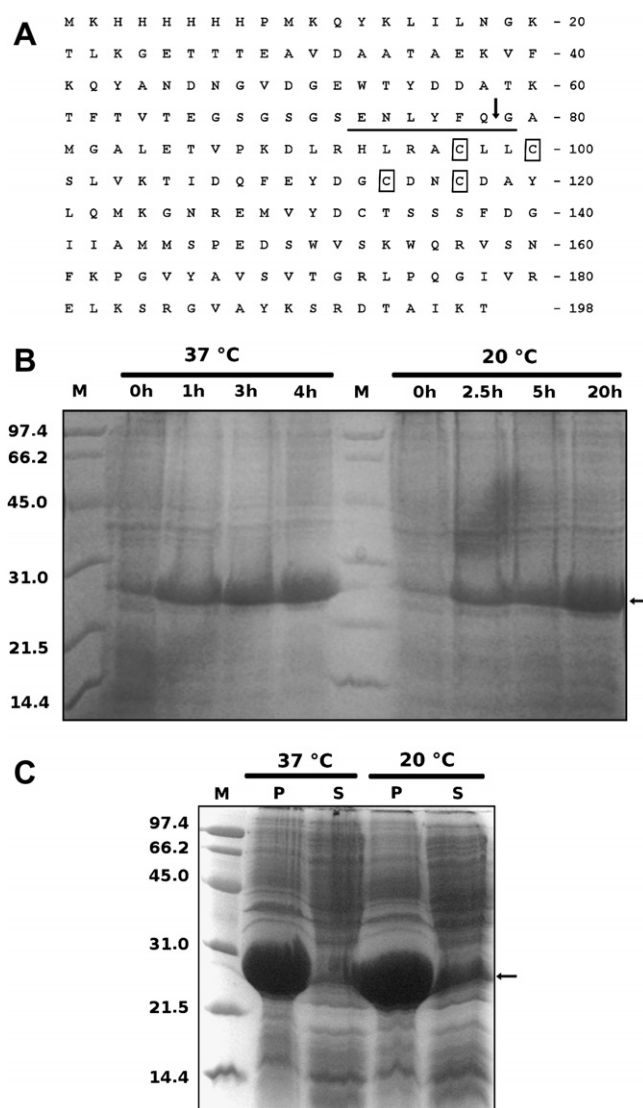


Fig. 1. Amino acid sequence and induction tests using the GB1-hSpt4 fusion protein. (A) Primary sequence of GB1-hSpt4. The TEV-protease recognition site is underlined, the cleavage site is indicated by an arrow. The four Cys forming the putative Zn-finger are highlighted by boxes. (B) Expression yields of *E. coli* BL21 (DE3) pETGB1-hSpt4 grown at 37 °C for 4 h or 20 °C for 20 h, respectively. Aliquots were taken after induction at the time points indicated and analyzed by SDS-PAGE. (C) SDS-PAGE of the soluble (S) and pelleted insoluble fractions (P) after growth at 37 or 20 °C and cell lysis by lysozyme and DNase I treatment and sonication.

To obtain free hSpt4 we first set out to express the corresponding gene without a large solubility tag, instead, a His tag was introduced at the N-terminus of the protein. Cells of *E. coli* strain BL 21 (DE3) harboring pET15b-hSpt4 were grown at either 37 or 20 °C in LB medium and hSpt4 gene expression was induced with 1 mM IPTG. However, even at 37 °C induction was rather low (data not shown).

We therefore changed our strategy and cloned the hSpt4 gene into pETGB1a. GB1, the streptococcal immunoglobulin-binding domain of protein G, has been shown to increase the solubility of recombinant proteins [4,9]. The resulting 22 kDa GB1-hSpt4 fusion protein harbors a 6×His tag at the N-terminus and a TEV protease cleavage site between GB1 and hSpt4 (Fig. 1A). The cleaved hSpt4 protein consists of 120 amino acids including 5 cysteines. The 4 cysteines in the N-terminal region were suggested to form a Zn-finger [5]. The region on hSpt4 interacting with hSpt5 has not been identified yet. However, it has been shown that the cysteines form a putative Zn-finger and the last 40 amino acid residues of the protein are dispensable for the interaction [13].

Induction of BL21(DE3)/pETGB1-hSpt4 was performed in LB medium at an $OD_{600} = 0.8$ with 0.1 mM IPTG at 37 °C for 4 h or 20 °C over night (Fig. 1B). Analysis of the induction by SDS-PAGE (Fig. 1C) shows that the recombinant protein is produced in reasonable amounts even at 20 °C. Furthermore, more protein can be detected in the supernatant of the cells grown at 20 °C. We were able to obtain sufficient protein for NMR when growing the cells on LB medium. Thus, for purifying ^{15}N -labeled protein, cells were grown over night at 20 °C in M9 medium.

Purification of GB1-hSpt4

Purification of the GB1-hSpt4 fusion protein with the His tag on GB1 was carried out using nickel ion affinity chromatography. The protein eluted at imidazole concentrations of approximately 100 mM. The corresponding fractions and the flow through were analyzed by SDS-PAGE (Fig. 2A). Lanes 6–11 show a band corresponding to a molecular mass of about 25 kDa and a double band corresponding to 16–13 kDa. Peptide mass fingerprint (PMF) analysis of the upper and the two lower protein bands indicated that the upper band represents the full length GB1-hSpt4 fusion protein, whereas the lower bands are degradation products containing C-terminal truncations of the recombinant protein. Probably, *E. coli* proteases lead to non-specific cleavages. Fractions containing GB1-hSpt4 were pooled and subjected to TEV cleavage during dialysis in TEV cleavage buffer. No precipitation of hSpt4 occurred after cleavage of the solubility-tag GB1. The fractions contaminated with the

C-terminally truncated degradation products were also used further since the GB1 truncations could be removed in the next purification step. Completeness of the cleavage was verified by analyzing aliquots on 19% SDS gels (Fig. 2B).

Purification of cleaved hSpt4

The solution harboring the cleaved off hSpt4 protein was loaded on a Ni-NTA column (Qiagen). hSpt4 can be detected in the flow through (Fig. 3A, lanes 2–6) and GB1 was eluted from the column by 250 mM imidazole (Fig. 3A, lanes 7–9). To remove the high molecular weight contaminant, the hSpt4 containing fractions

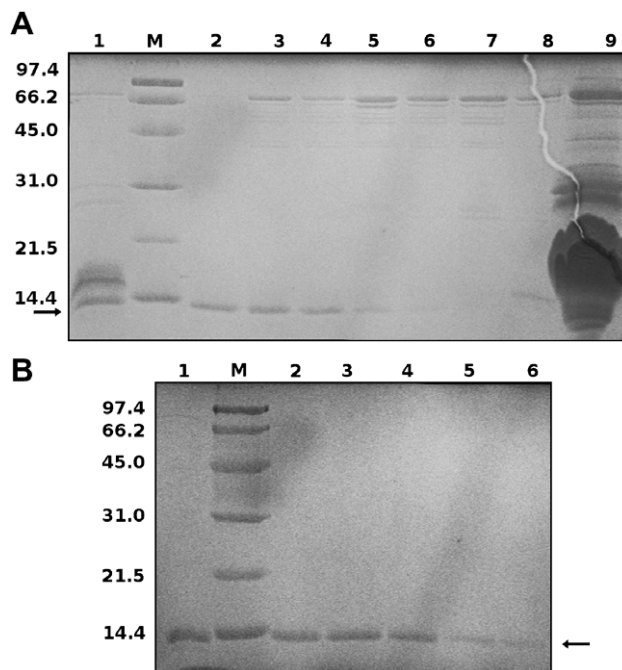


Fig. 3. Purification of hSpt4 after TEV cleavage. Aliquots were analyzed by SDS-PAGE and Coomassie staining. (A) Nickel ion affinity chromatography. Flow through (lane 1), and wash fractions (lanes 2–6) containing hSpt4; lanes 7–9, GB1 eluted with 250 mM imidazole. (B) Ion exchange chromatography using a HiTrap Q XL and a HiTrap SP XL (GE Healthcare, München, Germany) column in tandem. Lane 1, loaded sample; lanes 2–6, flow through and wash fractions containing hSpt4, indicated by an arrow; lane M, molecular weight standard, the apparent molecular weights are indicated on the left.

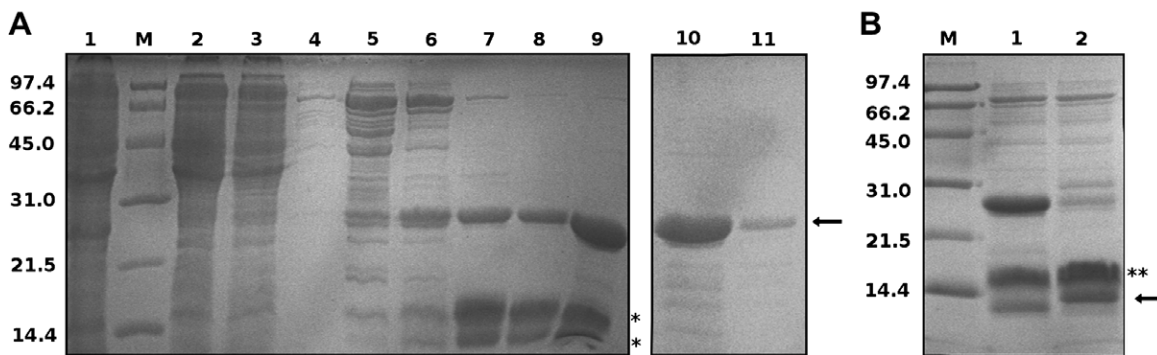


Fig. 2. Purification of GB1-hSpt4 using nickel ion affinity chromatography and TEV cleavage. Aliquots were analyzed by 19% SDS-PAGE and Coomassie staining. (A) Nickel ion affinity purification of GB1-hSpt4. Lane 1, cell lysate; lane 2, flow through; lane 3, wash fraction; elution of GB1-hSpt4 with 50 mM imidazole (lanes 4–7); 100 mM imidazole (lanes 8–9); 150 mM imidazole (lane 10); 200 mM imidazole (lane 11); lane M, molecular weight standard, the apparent molecular weights are indicated on the left. Arrow, GB1-hSpt4; *, GB1 degradation products. (B) Comparison of GB1-hSpt4 before (lane 1) and after (lane 2) TEV-protease treatment. The identity of hSpt4 (lower band lane 2, arrow) was verified by peptide mass fingerprinting. The upper band (lane 2, **) represents cleaved GB1 protein. Lane M, molecular weight standard, the apparent molecular weights are indicated on the left.

were step dialyzed against buffers containing decreasing concentrations of NaCl since we found out that much less protein precipitates during dialysis if the NaCl concentration is reduced step by step. The protein was stable in a buffer containing no NaCl and was subjected to ion exchange chromatography. An anion (Q XL HiTrap, GE Healthcare) and a cation (SP XL HiTrap, GE Healthcare) exchange column were applied in tandem. Using 50 mM NaCl in the chromatography buffer, the hSpt4 protein stayed in the flow through, whereas the contaminants bound to the ion exchangers (Fig. 3B). Thus we were able to obtain pure hSpt4. The yield from 4 l of M9 culture (28 g cells) was around 13 mg of ^{15}N -labeled hSpt4 protein which could be subjected to structural analysis.

Metal analysis

Analysis of the primary amino acid sequence of hSpt4 indicates a putative Zn-finger at the N-terminus of the protein (Fig. 1). To determine the presence of Zn^{2+} ions in hSpt4, the purified protein was subjected to inductively coupled plasma optical emission spectrometry (ICP-OES). Equimolar amounts of Zn^{2+} ions and hSpt4 were detected in the sample, indicating that the Zn-finger of the protein is functional. Since the protein was purified via nickel ion affinity chromatography we also analyzed the Ni^{2+} -ion content of the sample to determine a possible contamination by Ni^{2+} -ions or an exchange of Zn^{2+} by Ni^{2+} . However the Ni^{2+} -ion concentration was below the detection limit, implying that the four Cys bind Zn^{2+} very tightly.

CD and NMR spectroscopy

Proper folding of the cleaved and purified hSpt4 was examined first by CD analysis (Fig. 4A). Absorption minima at 207 and 222 nm indicate a defined secondary structure. Secondary structure prediction using the program CDSSTR [14,27] revealed 27% α -helix, 21% β -strands, and 30% loop proportions.

To further analyze the folding state of hSpt4, a 2D ^1H - ^{15}N -HSQC spectrum was recorded at 298 K (Fig. 4B) using uniformly ^{15}N -labeled protein. The spectrum shows well dispersed signals typical for a protein with defined tertiary structure. The number of 124 observed signals correlates well with the 132 expected ones. Missing signals probably result from residues in unstructured regions undergoing conformational or solvent exchange.

These results are reminiscent of the structure of the corresponding protein from the yeast *Saccharomyces cerevisiae* (pdb code 2EXU). Furthermore, a factor called E'' from the archaeobacterium *Pyrococcus furiosus* (pdb code 1RYQ), originally thought to be a subunit of the RNA polymerase, was shown recently to exhibit partial structural homology to yeast Spt4 [26]. However, the structure of yeast Spt4 could only be solved as a fusion protein with a peptide derived from the Spt5 subunit whereas our procedure allows the purification of the separate hSpt4 subunit of the DSIF complex, which makes the protein suitable for further functional studies and structural analyses in the absence and presence of hSpt5.

Acknowledgments

The project was funded by Grant RO617/16-1 from the Deutsche Forschungsgemeinschaft DFG and the University of Bayreuth. We thank Britta Zimmermann for excellent technical assistance.

References

- [1] M. Aida, Y. Chen, K. Nakajima, Y. Yamaguchi, T. Wada, H. Handa, Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the junB gene, *Mol. Cell. Biol.* 26 (2006) 6094–6104.
- [2] E.D. Andrulis, E. Guzman, P. Doring, J. Werner, J.T. Lis, High-resolution localization of *Drosophila* Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation, *Genes Dev.* 14 (2000) 2635–2649.
- [3] C.F. Bourgeois, Y.K. Kim, M.J. Churcher, M.J. West, J. Karn, Spt5 cooperates with human immunodeficiency virus type 1 Tat by preventing premature RNA release at terminator sequences, *Mol. Cell. Biol.* 22 (2002) 1079–1093.
- [4] Y. Cheng, D.J. Patel, An efficient system for small protein expression and refolding, *Biochem. Biophys. Res. Commun.* 317 (2004) 401–405.
- [5] P.W. Chiang, S.Q. Wang, P. Smithivas, W.J. Song, E. Crombez, A. Akhtar, R. Im, J. Greenfield, S. Ramamoorthy, M. Van Keuren, C.C. Blackburn, C.H. Tsai, D.M. Kurnit, Isolation and characterization of the human and mouse homologues (SUPT4H and Supt4h) of the yeast SPT4 gene, *Genomics* 34 (1996) 368–375.
- [6] J.W. Critchfield, J.E. Coligan, T.M. Folks, S.T. Butera, Casein kinase II is a selective target of HIV-1 transcriptional inhibitors, *Proc. Natl. Acad. Sci. USA* 94 (1997) 6110–6115.
- [7] U. Dreikhausen, K. Hiebenthal-Millow, M. Bartels, K. Resch, M. Nourbakhsh, NF-kappaB-repressing factor inhibits elongation of human immunodeficiency virus type 1 transcription by DRB sensitivity-inducing factor, *Mol. Cell. Biol.* 25 (2005) 7473–7483.
- [8] M.E. Garber, K.A. Jones, HIV-1 Tat: coping with negative elongation factors, *Curr. Opin. Immunol.* 11 (1999) 460–465.
- [9] A.M. Gronenborn, D.R. Filpula, N.Z. Essig, A. Achari, M. Whitlow, P.T. Wingfield, G.M. Clore, A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G, *Science* 253 (1991) 657–661.
- [10] G.A. Hartzog, M.A. Basrai, S.L. Ricupero-Hovasse, P. Hieter, F. Winston, Identification and analysis of a functional human homologue of the SPT4 gene of *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* 16 (1996) 2848–2856.
- [11] D. Ivanov, Y.T. Kwak, J. Guo, R.B. Gaynor, Domains in the SPT5 protein that modulate its transcriptional regulatory properties, *Mol. Cell. Biol.* 20 (2000) 2970–2983.
- [12] D. Ivanov, Y.T. Kwak, E. Nee, J. Guo, L.F. Garcia-Martinez, R.B. Gaynor, Cyclin T1 domains involved in complex formation with Tat and TAR RNA are critical for tat-activation, *J. Mol. Biol.* 288 (1999) 41–56.
- [13] D.K. Kim, N. Inukai, T. Yamada, A. Furuya, H. Sato, Y. Yamaguchi, T. Wada, H. Handa, Structure–function analysis of human Spt4: evidence that hSpt4 and

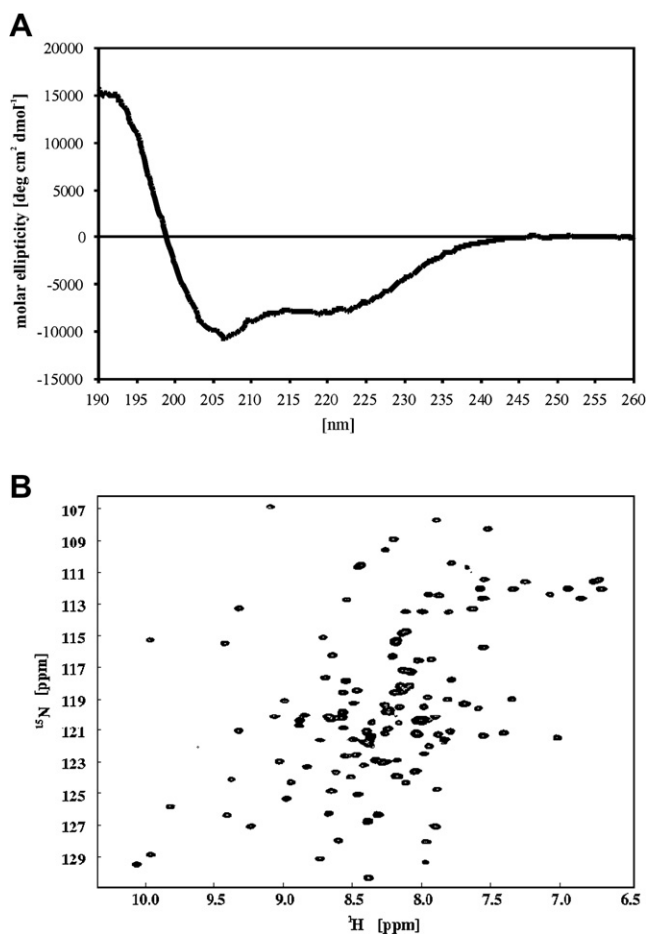


Fig. 4. Analysis of the hSpt4 secondary and tertiary structure by CD- and NMR-spectroscopy. (A) Far UV (190–260 nm) CD-spectrum of hSpt4 recorded at 293 K in water. (B) ^1H - ^{15}N -HSQC spectrum of 0.1 mM hSpt4 in 25 mM Tris/HCl, pH 6.9, 10 mM β -mercapto-ethanol, and 10% D_2O at 298 K measured with a Bruker DRX 600 MHz spectrometer (32 accumulations, 1024 data points in each dimension).

- hSpt5 exert their roles in transcriptional elongation as parts of the DSIF complex, *Genes Cells* 8 (2003) 371–378.
- [14] A. Lobley, L. Whitmore, B.A. Wallace, DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra, *Bioinformatics* 18 (2002) 211–212.
- [15] E.A. Malone, J.S. Fassler, F. Winston, Molecular and genetic characterization of SPT4, a gene important for transcription initiation in *Saccharomyces cerevisiae*, *Mol. Gen. Genet.* 237 (1993) 449–459.
- [16] P.B. Mason, K. Struhl, Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo, *Mol. Cell* 17 (2005) 831–840.
- [17] O. Meyer, H.G. Schlegel, Biology of aerobic carbon-monoxide oxidizing bacteria, *Ann. Rev. Microbiol.* 37 (1983) 277–310.
- [18] S. Mori, C. Abeygunawardana, M.O. Johnson, P.C.M. Vanzijl, Improved sensitivity of Hsqc spectra of exchanging protons at short interscan delays using a new fast Hsqc (Fhsqc) detection scheme that avoids water saturation, *J. Magn. Reson. B* 108 (1995) 94–98.
- [19] Y.H. Ping, C.Y. Chu, H. Cao, J.M. Jacque, M. Stevenson, T.M. Rana, Modulating HIV-1 replication by RNA interference directed against human transcription elongation factor SPT5, *Retrovirology* 1 (2004) 46.
- [20] Y.H. Ping, T.M. Rana, 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 (2001) 12951–12958.
- [21] C.P. Ponting, Novel domains and orthologues of eukaryotic transcription elongation factors, *Nucleic Acids Res.* 30 (2002) 3643–3652.
- [22] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning—A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1994.
- [23] R.J. Sims III, R. Belotserkovskaya, D. Reinberg, Elongation by RNA polymerase II: the short and long of it, *Genes Dev.* 18 (2004) 2437–2468.
- [24] T. Wada, T. Takagi, Y. Yamaguchi, A. Ferdous, T. Imai, S. Hirose, S. Sugimoto, K. Yano, G.A. Hartzog, F. Winston, S. Buratowski, H. Handa, DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs, *Genes Dev.* 12 (1998) 343–356.
- [25] T. Wada, T. Takagi, Y. Yamaguchi, D. Watanabe, H. Handa, Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro, *EMBO J.* 17 (1998) 7395–7403.
- [26] F. Werner, Structure and function of archaeal RNA polymerases, *Mol. Microbiol.* 65 (2007) 1395–1404.
- [27] L. Whitmore, B.A. Wallace, DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, *Nucleic Acids Res.* 32 (2004) W668–W673.
- [28] T. Yamada, Y. Yamaguchi, N. Inukai, S. Okamoto, T. Mura, H. Handa, P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation, *Mol. Cell* 21 (2006) 227–237.
- [29] Y. Yamaguchi, T. Wada, D. Watanabe, T. Takagi, J. Hasegawa, H. Handa, Structure and function of the human transcription elongation factor DSIF, *J. Biol. Chem.* 274 (1999) 8085–8092.
- [30] W. Zhu, T. Wada, S. Okabe, T. Taneda, Y. Yamaguchi, H. Handa, DSIF contributes to transcriptional activation by DNA-binding activators by preventing pausing during transcription elongation, *Nucleic Acids Res.* 35 (2007) 4064–4075.

10. Einzelarbeit D

Sabine Wenzel, Berta M. Martins, Paul Rösch und Birgitta M. Wöhrl

Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface.

Biochemical Journal (im Druck)

Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface

Sabine WENZEL*†, Berta M. MARTINS†, Paul RÖSCH*† and Birgitta M. WÖHRL*†¹

*Universität Bayreuth, Lehrstuhl für Struktur und Chemie der Biopolymere, Universitätsstr. 30, D-95447 Bayreuth, Germany, and †Research Centre for Biomacromolecules, Universitätsstr. 30, D-95447 Bayreuth, Germany

The eukaryotic transcription elongation factor DSIF [DRB (5,6-dichloro-1- β -D-ribofuranosylbenzimidazole) sensitivity-inducing factor] is composed of two subunits, hSpt4 and hSpt5, which are homologous to the yeast factors Spt4 and Spt5. DSIF is involved in regulating the processivity of RNA polymerase II and plays an essential role in transcriptional activation of eukaryotes. At several eukaryotic promoters, DSIF, together with NELF (negative elongation factor), leads to promoter-proximal pausing of RNA polymerase II. In the present paper we describe the crystal structure of hSpt4 in complex with the dimerization region of hSpt5 (amino acids 176–273) at a resolution of 1.55 Å (1 Å=0.1 nm). The heterodimer shows high structural similarity to its homologue from *Saccharomyces cerevisiae*. Furthermore, hSpt5-NGN is structurally similar to the NTD (N-terminal domain) of the bacterial transcription factor NusG. A homologue for hSpt4 has not yet been found in bacteria.

However, the archaeal transcription factor RpoE'' appears to be distantly related. Although a comparison of the NusG-NTD of *Escherichia coli* with hSpt5 revealed a similarity of the three-dimensional structures, interaction of *E. coli* NusG-NTD with hSpt4 could not be observed by NMR titration experiments. A conserved glutamate residue, which was shown to be crucial for dimerization in yeast, is also involved in the human heterodimer, but is substituted for a glutamine residue in *Escherichia coli* NusG. However, exchanging the glutamine for glutamate proved not to be sufficient to induce hSpt4 binding.

Key words: crystal structure, eukaryotic transcription, hSpt4/hSpt5, human DRB (5,6-dichloro-1- β -D-ribofuranosylbenzimidazole) sensitivity-inducing factor (DSIF), NusG N-terminal homology domain (NGN), NusG.

INTRODUCTION

The human DSIF [DRB (5,6-dichloro-1- β -D-ribofuranosylbenzimidazole) sensitivity-inducing factor] is a key player in the control of eukaryotic transcriptional elongation and has complex functions. It is involved in early elongation control by inducing pausing of RNAPol II (RNA polymerase II) in conjunction with the NELF (negative elongation factor) [1,2]. This was illustrated by promoter proximal pausing during transcription of the HIV-1 proviral genome DSIF and NELF interact directly with RNAPol II, and NELF also binds to the HIV-1 TAR RNA. Release of RNAPol II requires the viral transactivator protein Tat, which recruits the cellular P-TEFb (positive transcription elongation factor b) to the TAR RNA. In turn, P-TEFb phosphorylates the C-terminal domain of the largest RNAPol II subunit, the E subunit of NELF and the hSpt5 subunit of DSIF. Only then does NELF detach from the transcription machinery and DSIF is converted into an activator, thus stimulating elongation [3–8].

Similar promoter proximal pausing events involving NELF and DSIF have been detected at cellular genes under control of heat-shock promoters in *Drosophila melanogaster* [9–11]. Furthermore, interactions of DSIF or its yeast homologue with RNA-modifying enzymes, such as capping enzyme and cap

methyltransferase, and other cellular factors have been reported, underlining the diversity of DSIF in transcriptional regulation [12–15].

DSIF is a heterodimeric protein consisting of the 14 kDa subunit hSpt4 and the 120 kDa subunit hSpt5, which are homologous to the transcription factors Spt4 and Spt5 from the yeast *Saccharomyces cerevisiae* (Figure 1) [16]. hSpt5 is well conserved in eukaryotes and shares homology with NusG proteins of bacteria and archaea.

The bacterial transcription factor NusG fulfills various as well as species-specific functions. In *Escherichia coli*, it enhances transcriptional elongation as well as ρ -dependent termination [17,18]. In contrast, it acts as an anti-termination factor when phage λ N protein and other Nus proteins (A, B and E) are present [19,20].

hSpt5 is comprised of an N-terminal acidic region, whose function has not yet been established, followed by the NGN (NusG N-terminal homology domain) (residues 176–270) [21,22], which exhibits homology with the NTD (N-terminal domain) of NusG proteins from bacteria and archaea. The NGNs of hSpt5 and of the Spt5 proteins of *S. cerevisiae* and the fission yeast *Schizosaccharomyces pombe* have been identified as harbouring the binding region for the small subunit hSpt4 [3,16,24]. Four to six KOW (after Kyrpides, Ouzounis and Woese [26]) motifs,

Abbreviations used: DSIF, DRB (5,6-dichloro-1- β -D-ribofuranosylbenzimidazole) sensitivity-inducing factor; GB1, streptococcal immunoglobulin-binding domain of protein G; hSpt5-NGN, human Spt5(176–273); HSQC, heteronuclear single-quantum coherence; KOW motif, Kyrpides, Ouzounis and Woese motif; MAD, multi-wavelength anomalous dispersion; NELF, negative elongation factor; NGN, NusG N-terminal homology domain; Ni-NTA, Ni²⁺-nitrilotriacetate; NTD, N-terminal domain; P-TEFb, positive transcription elongation factor b; RMSD, root mean square deviation; RNAPol II, RNA polymerase II; SAD, single-wavelength anomalous dispersion; TEV, tobacco etch virus.

¹ To whom correspondence should be addressed (email birgitta.woehrl@uni-bayreuth.de).

The structural co-ordinates reported for the human transcription elongation factor DSIF [DRB (5,6-dichloro-1- β -D-ribofuranosylbenzimidazole) sensitivity-inducing factor], hSpt4/hSpt5 (176–273) will appear in the Protein Data Bank under accession code 3H7H.

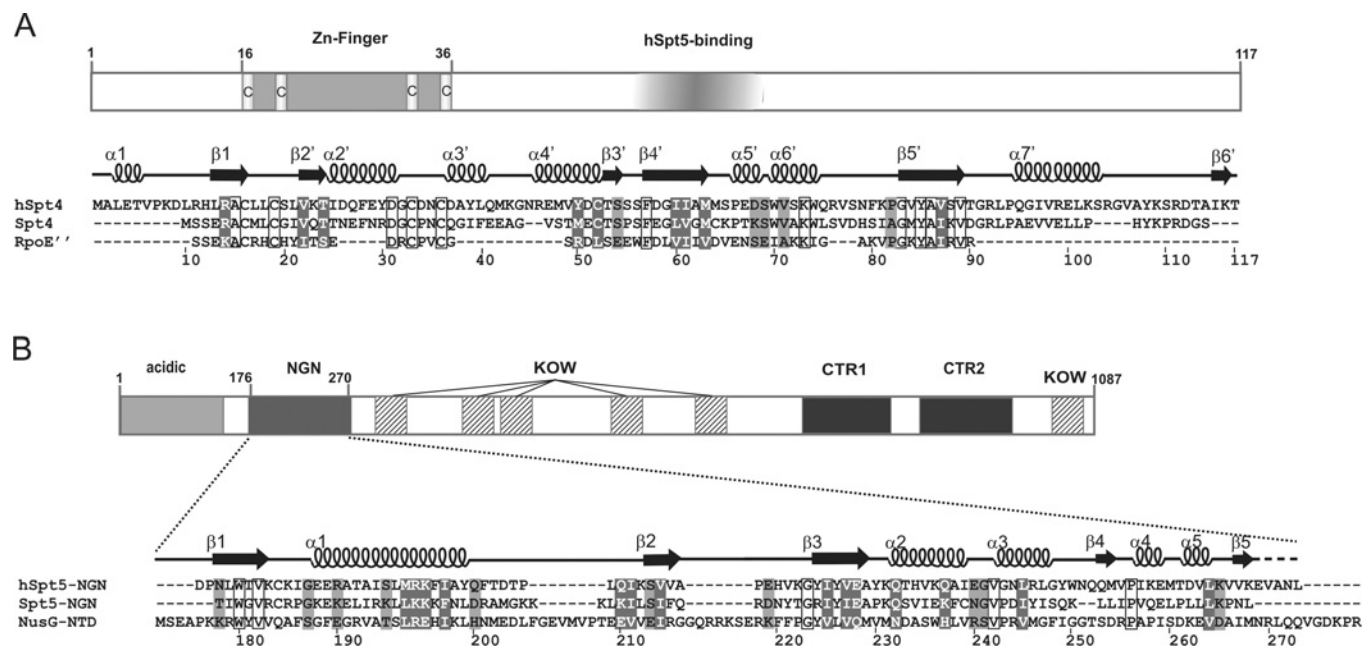


Figure 1 Schematic representation of the domain structures of hSpt4 and hSpt5

(A) Domain structure of hSpt4 and secondary structure distribution of hSpt4 in the hSpt4–hSpt5-NGN heterodimer. The amino acid sequences of hSpt4 (PDB: 3H7H), Spt4 from *S. cerevisiae* (PDB: 2EXU) and archaeal RpoE'' from *P. furiosus* (PDB: 1RYQ) were aligned. Conserved amino acids are highlighted by a black frame, conserved substitutions are represented by white letters on grey boxes and semi-conserved substitutions are highlighted by black letters on grey boxes. Secondary structure elements of complexed hSpt4 (PDB: 3H7H) are indicated on top of the alignment. Numbers underneath the alignment refer to amino acid positions of hSpt4. (B) Domain architecture of hSpt5 [21,22]. The amino acid sequences of hSpt5-NGN (residues: 176–273) (PDB: 3H7H), Spt5-NGN from *S. cerevisiae* (residues: 286–375) (PDB: 2EXU) and *E. coli* NusG-NTD (PDB: 2K06) are aligned. Labelling of conservation of amino acids is the same as in (A). The secondary structure composition of complexed hSpt5-NGN is displayed on top of the alignment. C, cysteine residues involved in zinc co-ordination; CTR 1 and 2, C-terminal repeat 1 and 2 respectively; acidic, N-terminal acidic region. Numbers underneath the alignment refer to amino acid positions of hSpt5.

which directly interact with RNAPol II during transcription, have been assigned C-terminally to the NGN in hSpt5 [21,22,25]. KOW motifs have also been identified in the NusG family proteins and the proteins RL24, 26 and 27 of the large ribosomal subunit [26]. The C-terminus of hSpt5 contains the hepta- and octapeptide repeats CTR1 and CTR2 [3,4]. Phosphorylation of threonine residues in the CTR1 region of hSpt5 as well as methylation of arginine residues near its KOW motifs affects DSIF activity [8,27].

hSpt4, the smaller subunit of DSIF, harbours an N-terminal four-cysteine zinc finger, and exhibits α/β -topology [28,29]. The presence of a bound Zn^{2+} was verified previously [29]. In contrast with hSpt5, no bacterial homologue exists for hSpt4, but it is uniformly conserved in archaea and eukaryotes. The archaeal transcription factor RpoE'', originally annotated as RNA polymerase subunit E'', is distantly related to hSpt4 and is regarded as its evolutionary ancestor [22,30].

In the present work we determined the crystal structure of the recombinant hSpt4 complexed with hSpt5-NGN [human Spt5(176–273)], representing the binding region for hSpt4, at a resolution of 1.55 Å (1 Å=0.1 nm). The structure confirms that hSpt4 is indeed an α/β -type protein with a zinc finger and that it forms a tight complex with the NGN of hSpt5. Comparisons with the crystal structure of the homologous Spt4–Spt5(286–375) complex from *S. cerevisiae* [31] verified the high structural similarity already indicated by sequence alignments. Biophysical analyses revealed that, in spite of a similar folding topology of NusG-NTD and hSpt5-NGN, NusG-NTD lacks the ability to interact with hSpt4. Moreover, we also show a structural relationship between hSpt4 and the archaeal transcription factor RpoE''.

EXPERIMENTAL

Plasmid construction

Construction of plasmid pET-GB1-hSpt4, harbouring the DNA sequences for the solubility tag GB1 (streptococcal immunoglobulin-binding domain of protein G), as well as a TEV (tobacco etch virus) protease cleavage site followed by the full-length *hSpt4* gene was described previously [29]. A synthetically produced gene adapted to the *E. coli* codon usage harbouring the NGN region (amino acids 176–273) of the DSIF *hSpt5* gene was cloned into the expression vector pET15b (Novagen) via its NdeI and BamHI restriction sites. The final plasmid pET15b-hSpt5-NGN encodes an N-terminal 6×His-tagged fusion protein with a thrombin cleavage site.

Gene expression and protein purification

The plasmids pET15b-hSpt5-NGN and pET-GB1-hSpt4 were co-transformed into the *E. coli* strain BL21 (DE3) (Novagen). Cells were grown as described previously [29] using kanamycin (30 µg/ml) and ampicillin (100 µg/ml) as antibiotics. Purification of the complex by Ni-NTA (Ni^{2+} -nitrilotriacetate)-affinity chromatography was performed as described previously for hSpt4 [29]. Fractions containing the DSIF complex were subjected to overnight TEV cleavage at 4°C, followed by a second Ni^{2+} -affinity chromatography step to remove excess cleaved hSpt4 not complexed with hSpt5. The N-terminal 6×His-tag of hSpt5 was removed by thrombin cleavage at room temperature (20°C) overnight. To remove the cleaved-off affinity tags GB1 and 6×His, the sample was applied to a Ni-NTA column. The

DSIF complex was collected in the flow-through and dialysed in steps in 25 mM Tris/HCl, pH 7.4, 10 mM 2-mercaptoethanol and decreasing concentrations of NaCl down to 50 mM.

The sample was subjected to cation-exchange chromatography (HiTrap SP XL column; GE Healthcare), dialysed against 25 mM Tris/HCl, pH 7.0, 50 mM NaCl, 10 mM 2-mercaptoethanol and concentrated in Vivaspin concentrators MWCO (molecular mass cut-off) 5000 (Sartorius, Göttingen, Germany).

The genes expressing untagged NusG-NTD and NusG-NTD(Q72E) were cloned into pET11a (Novagen) and the proteins were purified from *E. coli* extracts via heparin-affinity and size-exclusion chromatography using a Superdex 75 HR 10/30 column (GE Healthcare). Samples were dialysed and concentrated as described above for the hSpt4–hSpt5-NGN complex.

Size-exclusion chromatography

Size-exclusion chromatography to determine the molecular mass of the hSpt4–hSpt5-NGN complex was performed using a Superdex 75 HR 10/30 column (GE Healthcare). As molecular mass standards, albumin (67.0 kDa, GE Healthcare) chymotrypsinogen A (25.0 kDa, GE Healthcare) and cytochrome *c* (12.4 kDa, Sigma–Aldrich) were used. Standard proteins were dissolved in 25 mM Tris/HCl, pH 7.0, 50 mM NaCl and 10 mM 2-mercaptoethanol, which was also used for equilibration and the column run. The sample was dialysed against the same buffer.

NMR spectroscopy

Two dimensional ^{15}N - ^1H -HSQC (heteronuclear single-quantum coherence) NMR spectra were recorded on a Bruker AV 800 MHz NMR spectrometer equipped with a cryogenic probe and pulsed field-gradient capabilities at 298 K according to the FHSQC (fast HSQC) scheme [32]. For data processing and visualization, in-house software and NMRView [33] were used respectively.

Crystallization and data collection

Recombinant hSpt4–hSpt5-NGN was crystallized by the hanging-drop vapour diffusion method at 290 K by mixing 2 μl of protein solution (1.2 mg/ml in 25 mM Tris/HCl, pH 7.0, and 10 mM 2-mercaptoethanol) plus 2 μl of precipitant {50 mM Tris/HCl, pH 7.0, 0.2 M $(\text{NH}_4)_2\text{SO}_4$, 13–15 % PEG [poly(ethylene glycol)] 3350 and 10 mM 2-mercaptoethanol}. Triangular-shaped crystals grew within 1 week with dimensions of approx. $300 \times 150 \times 50 \mu\text{m}$. Crystals were harvested in 17 μl of precipitant plus 3 μl of 2*R*,3*R*-butandiol (Sigma–Aldrich) before flash-cooling in liquid nitrogen. Zinc-MAD (multi-wavelength anomalous dispersion) diffraction data were collected at 100 K at the synchrotron beam line 14.2 of BESSY (Berlin electron storage ring company for synchrotron radiation; Berlin, Germany) with a single crystal belonging to the orthorhombic system (space group $P2_12_12_1$ with cell constants of $a = 41.096 \text{ \AA}$, $b = 52.091 \text{ \AA}$ and $c = 97.099 \text{ \AA}$). The crystal diffracted to a resolution of 1.55 Å at the zinc-MAD remote wavelength (0.91841 Å). There is one heterodimer per asymmetric unit with a solvent content of 40 %. Data were processed and scaled with XDS (X-ray detector software) [34]. Data statistics are reported in Table 1.

Structure determination and refinement

Initial phases were obtained with the SHELX program [35] by SAD (single-wavelength anomalous dispersion) phasing using the wavelength 1.28308 Å corresponding to the peak of the zinc-MAD experiment. Automatic model building with ARP/wARP software

Table 1 Data collection and refinement statistics

Values in parentheses are given for the highest resolution shell (1.59–1.55 Å).

Parameter	Data collection	
Wavelength (Å)	0.91841 (remote)	1.28308 (Zn-peak)
d_{min} (Å)	19.11–1.55	19.11–1.60
Space group	$P2_12_12_1$	
Cell dimensions (Å)		
<i>a</i>	41.096	
<i>b</i>	52.091	
<i>c</i>	97.099	
Total / unique reflections	115753/30454	105462/51330
Completeness (%)	98.1 (86.9)	97.1 (94.7)
R_s^*	3.4 (23.6)	2.4 (10.8)
I/σ	22.7 (4.7)	22.1 (7.3)
Model refinement		
d_{min} (Å)	19.11–1.55	
Reflections	28930	
R -factor/ R_{free} -factor†	0.186/0.222 (0.222/0.256)	
RMSD-bond length (Å)	0.01	
RMSD-bond angles ($^\circ$)	1.277	
Average B-factor (Å^2)	14.36	
ESU‡ (Å)	0.054	
Protein atoms	1746	
Solvent atoms	196	
Metal ions (Zn^{2+})	1	

* $R_s = \frac{\sum_h \sum_i |I_i(h) - \langle I(h) \rangle|}{\sum_h \sum_i I_i(h)}$; where i are the independent observations of reflection h .

†The R_{free} -factor was calculated from a 5% subset of reflections (1523), which were removed at random before the refinement was carried out.

‡Estimated overall co-ordinate error (ESU) based on maximum likelihood.

[36] resulted in a largely complete polypeptide model. Manual model building and positioning of the Zn^{2+} atom associated with hSpt4 were performed with the program COOT [37]. Positional and temperature factor refinements were carried out with Refmac5 [38]. In the final refinement cycles, the B-factors of Zn^{2+} and the S atoms of the four co-ordinating cysteine residues were refined anisotropically and alternative conformations of side chains were included. Several side chains located at the surface of the molecules (16 from hSpt4 and 11 from hSpt5-NGN) were only partially visible in the electron density maps and were therefore modelled using one of the rotamers.

ARP/wARP and Refmac5 programs were used as implemented in the CPP4 suite [39]. The refined model contained one hSpt4 subunit (Gly¹ to Thr¹¹⁷) and one hSpt5-NGN subunit (Asp¹⁷⁶ to Glu²⁶⁹), one Zn^{2+} ion and 196 solvent molecules. The N-terminal methionine residue of hSpt4 (numbered zero in the PDB file 3H7H) and the N-terminal residue methionine of hSpt5-NGN (numbered 175 in the PDB file 3H7H) are cloning artifacts. The last four C-terminal residues of hSpt5-NGN (Val–Ala–Asn–Leu), and the next four residues (Gly–Ser–Gly–Cys; derived from the cloning strategy) were not visible in the electron density maps and are therefore not included in the refined model. The final refinement statistics are shown in Table 1.

Figures were generated by PyMOL v0.99 (DeLano Scientific; <http://www.pymol.org>) and electrostatic surface potential calculations for hSpt5-NGN, Spt5-NGN and *E. coli* NusG-NTD were performed using the APBS (adaptive Poisson–Boltzmann solver) tool [41] for the PyMOL program written by Michael Lerner.

All secondary structures shown were determined by PROCHECK [42]. ClustalW2 [43] was used for sequence alignments. The program LIGPLOT [44] was used to identify the amino acid residues involved in dimerization. The co-ordinates

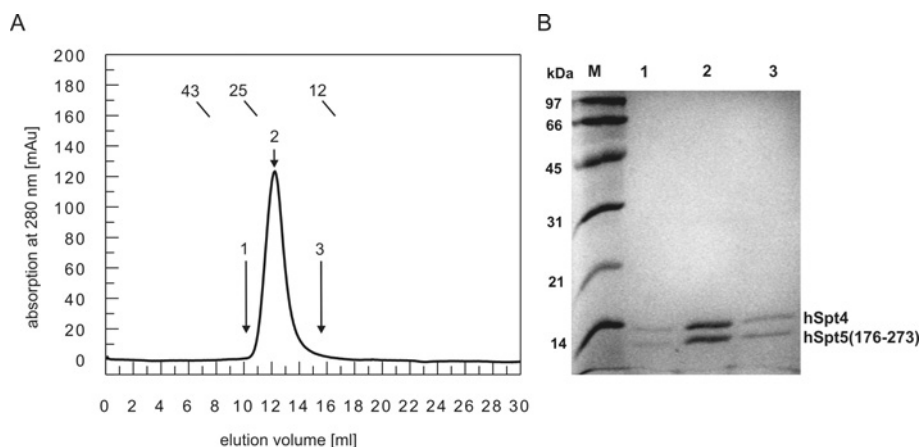


Figure 2 Analysis of purified hSpt4–hSpt5-NGN

(A) Size-exclusion chromatography using a Superdex 75 HR 10/30 column (GE Healthcare). The elution volumes and molecular masses (in kDa) of the standard proteins used for column calibration are indicated. The numbers 1–3 correspond to the lanes shown in (B). (B) SDS/PAGE (19% gel). The gel shows eluted fractions from (A) of hSpt4–hSpt5-NGN. The bands corresponding to hSpt4 and hSpt5-NGN are indicated. Numbers on the left indicate the molecular masses of standard proteins (lane M).

and structure factors were deposited in the Protein Data Bank with accession code 3H7H.

RESULTS AND DISCUSSION

To facilitate formation of the correctly folded heterodimeric hSpt4–hSpt5-NGN complex, we chose to co-express the two genes, which were encoded on separate plasmids in *E. coli*. The hSpt5 construct used (amino acids 176–273) encompasses the NGN [19,20]. Previous attempts to express the two genes, *hSpt4* and *hSpt5-NGN*, in separate strains and then combine the proteins, failed due to insolubility of hSpt5-NGN. Only hSpt4 could be purified successfully [29].

The integrity of the purified hSpt4–hSpt5-NGN complex was verified by size-exclusion chromatography (Figure 2A). Only one peak was detected, which eluted at 12.4 ml, corresponding to an apparent molecular mass of 23.0 kDa. This is in good agreement with the calculated molecular mass of the complex of 25.3 kDa. Peak fractions analysed by SDS/PAGE revealed two bands, corresponding to hSpt4 and hSpt5-NGN (Figure 2B). The identity of the proteins was verified by MS. These analyses confirm that the heterodimer observed in the crystal structure most probably reflects the physiological complex and is not an artifact from crystal packing.

Structure of the hSpt4–hSpt5-NGN complex

The crystal structure of hSpt4–hSpt5-NGN was determined to a resolution of 1.55 Å by zinc-SAD phasing. Crystal parameters, data collection and refinement statistics are summarized in Table 1. The crystal structure of the hSpt4–hSpt5-NGN complex shows a compact heterodimer with an overall surface area of 11547 Å² and an overall buried surface of 3159 Å² [45].

The hSpt5-NGN of the DSIF heterodimer is an α/β type protein with β -strand and α -helical contents of 17.3% and 34.7% respectively. It exhibits a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\alpha_3\beta_4\alpha_4\alpha_5\beta_5$ folding topology (Figure 3) with a central four-stranded antiparallel β -sheet that is sandwiched between three helices ($\alpha 1$, $\alpha 2$ and $\alpha 3$) and two short single helices ($\alpha 4$ and $\alpha 5$). Helices $\alpha 2$ and $\alpha 3$ span across the β -sheet, thereby shielding parts of the sheet from

the solvent. Furthermore, the largest helix ($\alpha 1$) of hSpt5-NGN is aligned along $\beta 2$ and both are part of the dimerization interface.

The secondary structure distribution of hSpt4 is similar to hSpt5, with approx. 32.2% α -helical and 19.4% β -strand content (Figure 1). A secondary structure prediction, based on CD data for separately purified monomeric hSpt4 [29], yielded a similar structural distribution as observed in the heterodimer. It consists of two β -sheets of three and two antiparallel β -strands ($\beta 1'$ – $\beta 3'$, $\beta 4'$ – $\beta 5'$) respectively. Both β -sheets are packed orthogonally against each other and face towards the dimer interface where $\beta 4'$ is directly involved in the interaction with hSpt5-NGN. The solvent-exposed side of hSpt4 is exclusively α -helical with three short α helices ($\alpha 2'$ – $\alpha 4'$) and $\alpha 7'$ (Figures 3A and 3B). Additionally, in the hSpt4 centre, a Zn²⁺ is tetrahedrally coordinated by four cysteine residues (Cys¹⁶, Cys¹⁹, Cys³³ and Cys³⁶) forming a zinc finger (Figure 3C). Cysteine residues involved in Zn²⁺ co-ordination are located in the loop regions between $\beta 1'$ and $\beta 2'$ (Cys¹⁶ and Cys¹⁹) and $\alpha 2'$ and $\alpha 3'$ (Cys³³ and Cys³⁶), and presumably fulfill a major structural function by arranging the large helix-loop region, from amino acids 23–54, in close proximity to the β -sheet with $\beta 1'$ and $\beta 2'$.

Furthermore, a short 3_{10} -helix ($\alpha 1'$) is located at the N-terminus and a second one ($\alpha 2'$) is between $\beta 4'$ and $\beta 5'$. In the β -strands $\beta 4'$ and $\beta 5'$, a β -bulge interruption resides on equivalent spatial positions forcing both strands to bend away from the interface region. At the C-terminus of hSpt4, the short β -strand $\beta 6'$ is involved in a two-stranded parallel β -sheet with the second strand, $\beta 5$, derived from hSpt5-NGN.

Dimerization interface

In the heterodimer, $\beta 4'$ of hSpt4 aligns antiparallely with $\beta 2$ of hSpt5-NGN, leading to a large six-stranded β -sheet that stabilizes the dimer (Figure 3B). Helix $\alpha 1$ of hSpt5-NGN packs against $\beta 2$ of hSpt5-NGN and $\beta 4'$ of hSpt4. Non-polar residues point towards the interface, thereby stabilizing the interaction by hydrophobic contacts with hSpt4 (Figure 3D). Thus the central interface comprises predominantly of hydrophobic residues. Only a few polar side chains (e.g. Gln²⁰⁴ and Glu²¹⁹ of hSpt5-NGN) at the border of the interface also contribute to dimerization.

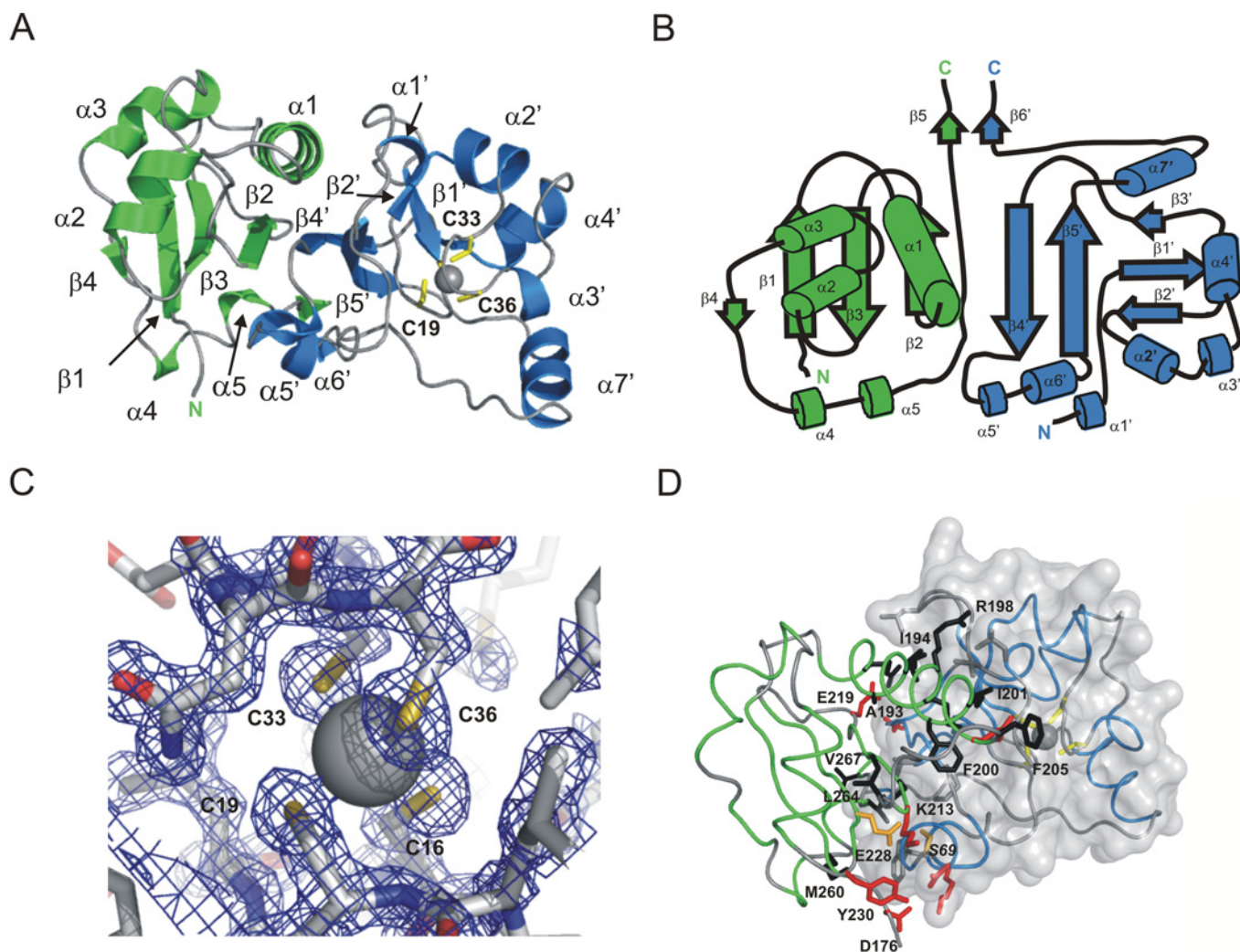


Figure 3 Structure of the hSpt4–hSpt5-NGN heterodimer

(A) Ribbon diagram of the hSpt4–hSpt5-NGN heterodimer. hSpt5-NGN is presented in green, hSpt4 in blue; three of the four cysteine residues (Cys¹⁶, Cys¹⁹, Cys³³ and Cys³⁶) co-ordinating the zinc ion (grey sphere) are shown in yellow; for clarity not all secondary structure elements are indicated. (B) Topology of the secondary structure elements in the hSpt4–hSpt5-NGN complex. Green: hSpt5-NGN; blue: hSpt4. β -strands are represented by arrows, α -helices by cylinders. (C) Electron density map. $2F_{\text{obs}} - F_{\text{cal}}$ electron density map (blue) contoured at 1.0σ of the zinc finger region of hSpt4. (D) Dimerization interface of the complex. hSpt5-NGN is depicted in green, the surface of hSpt4 is represented semi-transparently. The secondary structure elements of hSpt4 are shown in blue, unstructured regions in light grey. Residues of hSpt4 interacting with hSpt5-NGN are shown in red (polar) or dark grey (hydrophobic). Residues of hSpt5-NGN involved in hydrophobic interactions with hSpt4 are coloured in black, polar residues are coloured in red. The interacting residues of hSpt5-NGN are labelled. The conserved Glu²²⁸ of hSpt5 and its interaction partner Ser⁶⁹ (in italics) of hSpt4 are highlighted in orange.

For the *S. cerevisiae* homologues Spt4 and Spt5, it was reported that the salt bridge between the conserved Glu³³⁸ in Spt5 and Ser⁵⁸ of Spt4 is crucial for binding, and yeast mutant screens showed that mutation of Glu³³⁸ resulted in loss of dimerization [31]. In hSpt5, the corresponding conserved residue Glu²²⁸ is located at the C-terminus of $\beta 3$ and faces towards the N-terminus of $\alpha 5'$ of hSpt4. It forms interactions at hydrogen bond distance with backbone amide protons of residues Val⁷² and Trp⁷¹ and with the hydroxy group of Ser⁶⁹ of hSpt4 (Figure 3D). To analyse whether the conserved Glu²²⁸ in hSpt5 (Glu³³⁸ in Spt5 from *S. cerevisiae*) is also essential for dimerization with hSpt4, we exchanged Glu²²⁸ for a glutamine residue and co-expressed wild-type hSpt4 with mutated hSpt5-NGN-E228Q. However, the E228Q mutation obviously destroyed the structural integrity of the complex. In contrast with the wild-type protein, hSpt5-NGN-E228Q was found to be insoluble and could not be co-purified with hSpt4 (results not shown), indicating an important structural role for this residue.

hSpt4–hSpt5-NGN shares an identity with Spt4–Spt5-NGN from *S. cerevisiae* (PDB: 2EXU) of 37.5% (ClustalW2). The structures of the two complexes are remarkably conserved and superposition of the complexes clearly show identical folding topologies [RMSD (root mean square deviation): 1.36 Å over 860 backbone atoms] (see Supplementary Figure S1 at <http://www.BiochemJ.org/bj/425/bj4250373add.htm>). The two Spt4 proteins reveal a slightly less conserved tertiary structure as compared with the Spt5 proteins, although the Spt4 proteins share a higher sequence identity (Spt4 proteins: 46%; Spt5 proteins: 28%; ClustalW2). Especially, helices $\alpha 2'$ and $\alpha 4'$ of hSpt4 do not superimpose perfectly with yeast Spt4, and hSpt4 harbours 10 additional N-terminal residues. Except for the short 3_{10} -helix $\alpha 1'$, these residues show no significant secondary structural elements (see Supplementary Figure S1). However, residues Lys³ and Val⁶ of the N-terminal region of hSpt4 are involved in hydrophobic interactions with hSpt5-NGN, which is not the case in the *S. cerevisiae*

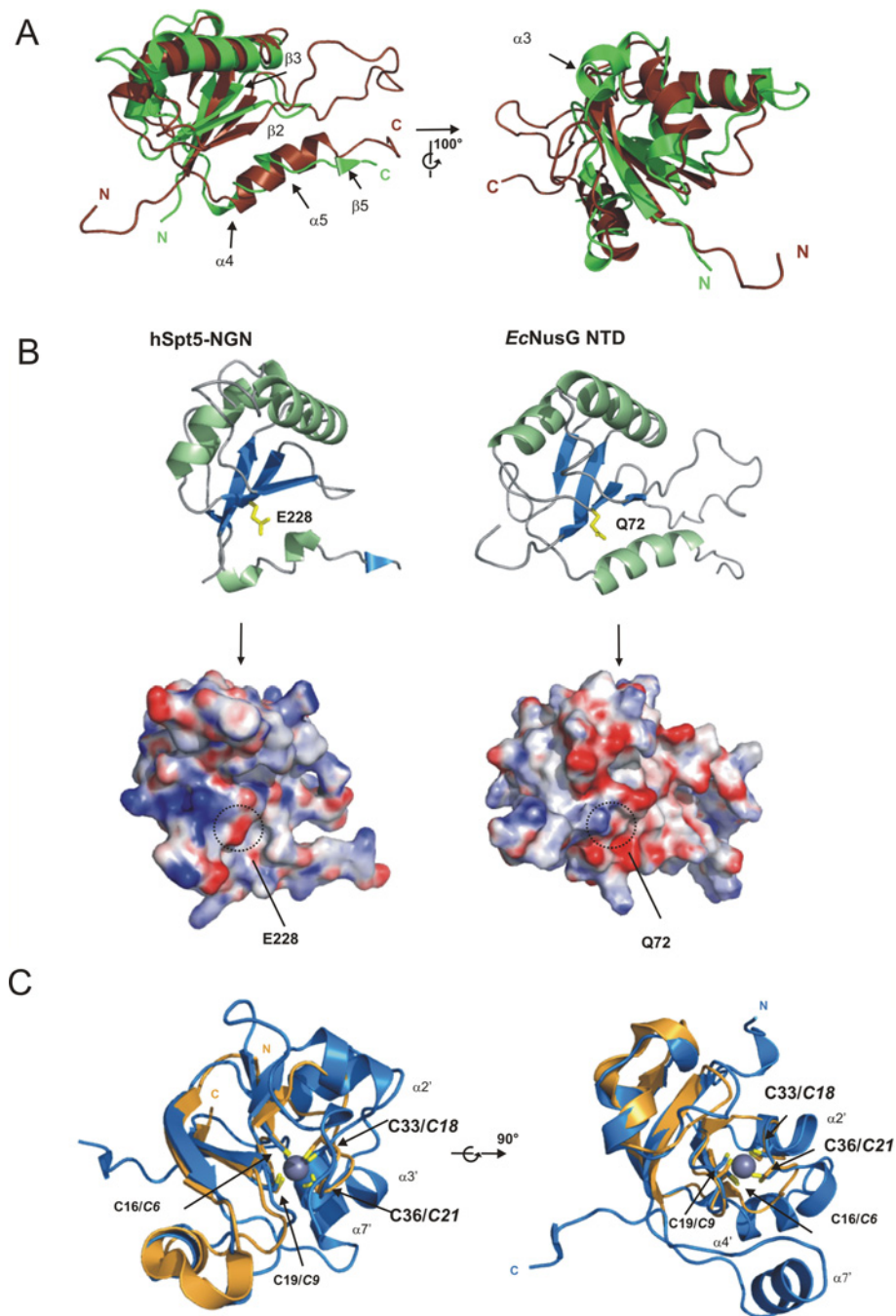


Figure 4 Comparison of hSpt4–hSpt5-NGN with *E. coli* NusG-NTD and archaeal RpoE''

(A) Superposition of hSpt5-NGN with *E. coli* NusG-NTD. hSpt5-NGN is shown in green, *E. coli* NusG-NTD (PDB: 2K06) in brown; secondary structure elements mentioned in the text are indicated. (B) Electrostatic surface potential of hSpt5-NGN with *E. coli* NusG-NTD. Structures are displayed in identical orientations. The conserved Glu²²⁸ in hSpt5-NGN and the corresponding Gln⁷² in *E. coli* NusG-NTD are shown as yellow sticks in the ribbon diagrams. The electrostatic surface potential mapped on the surfaces of hSpt5-NGN and *E. coli* NusG-NTD at a contour level of ± 20 kT is presented below; negative potentials are shown in red, positive potentials in blue and neutral potentials in white; Glu²²⁸ in hSpt5-NGN and Gln⁷² in *E. coli* NusG-NTD are indicated by a dotted circle. (C) Superposition of hSpt4 with RpoE''. Ribbon diagrams of hSpt4 (blue) and RpoE'' from *P. furiosus* (PDB: 1RYQ) (orange); cysteine residues involved in zinc co-ordination are shown as sticks with the sulphur atoms in yellow. The zinc ion of hSpt4 is displayed as a grey sphere. The zinc ion of RpoE'' is not shown, but is located at a similar position; cysteine residue numbers of RpoE'' are in italics. $\alpha 4$ is not visible on the left ribbon diagram.

complex. These interactions probably stabilize the human DSIF complex.

Moreover, in *S. cerevisiae* Spt5-NGN, helix $\alpha 3$ is replaced by a long β -strand which superimposes at its C-terminus with $\beta 4$ of the hSpt5-NGN. The 3_{10} helices $\alpha 4$ and

$\alpha 5$ in hSpt5-NGN are substituted for a single long helix enclosing both hSpt5-NGN helices in yeast Spt5-NGN (see Supplementary Figure S1). In addition, the small $\beta 5$ -strand at the C-terminus of hSpt5-NGN is missing in Spt5-NGN of *S. cerevisiae*.

Comparison of hSpt5-NGN with *E. coli* NusG-NTD

The physiological function of the NGN of DSIF has not yet been identified. In addition to the NGN, NusG proteins from bacteria and archaea also contain KOW motifs, supporting the hypothesis that hSpt5 is the eukaryotic counterpart of NusG.

Comparison of hSpt5-NGN with *E. coli* NusG-NTD (PDB: 2K06) [46] does indeed reveal structural homology [RMSD: 3.1 Å over 386 backbone atoms, sequence identity: 16% (ClustalW2)] (Figure 4A). The two proteins exhibit identical folding topologies, but also possess unique characteristics such as the α_{10} helices α_4 and α_5 in hSpt5-NGN, which are replaced by a single long C-terminal helix in *E. coli* NusG-NTD. The small β_5 of hSpt5 is lacking in *E. coli* NusG-NTD and the major β -sheet of hSpt5-NGN also does not superimpose perfectly with its *E. coli* counterpart. Additionally, discrepancies in secondary structure elements can be seen for helix α_3 of hSpt5-NGN, which is substituted by a β -strand in *E. coli* NusG-NTD. In contrast, the small β -strand (β_4) in hSpt5-NGN is missing in *E. coli* NusG-NTD. Similar structural differences can be detected between human and yeast Spt5-NGN, indicating that the structures of NTDs and NGNs are more conserved between prokaryotes and *S. cerevisiae* than between *S. cerevisiae* and humans.

Moreover, the loop between β_2 and β_3 is truncated in hSpt5-NGN (Figure 4A). Structural comparisons showed that in archaeal NusG proteins Glu³³⁸ of Spt5 is also conserved [31], whereas several bacterial NusG proteins, similar to the one from *E. coli*, harbour a glutamine residue at this position. The physiological function of the corresponding glutamate residue in archaea has not yet been identified. However, sequence comparisons with the DSIF complex from *S. cerevisiae* suggested that the conserved glutamate might play a role in the interaction of archaeal NusG with RpoE'' [31].

Since we can show that the bacterial NusG is structurally similar to hSpt5, we analysed whether the NTD domain of *E. coli* NusG interacts with hSpt4. However, the co-expression experiments and NMR titrations (see Supplementary Figure S2A at <http://www.BiochemJ.org/bj/425/bj4250373add.htm>) we performed with hSpt4 and *E. coli* NusG-NTD did not suggest any interactions.

Since *E. coli* NusG possesses a glutamine residue (Gln⁷²) at the position corresponding to the conserved Glu²²⁸ of hSpt5, we mutated Gln⁷² of NusG-NTD to a glutamate residue to make it more similar to its eukaryotic and archaeal homologues. We then tested if this amino acid exchange is sufficient for interaction with hSpt4. NMR titration experiments with ¹⁵N-labelled hSpt4 and unlabelled *E. coli* NusG-NTD-Q72E showed no significant chemical shift changes in the HSQC spectrum of hSpt4 (see Supplementary Figure S2B), implying that even with the mutated protein no interaction occurs. This result indicates that the conserved glutamate in eukaryotes and archaea cannot by itself account for the interaction, even in the context of high overall structural similarity.

Comparison of the electrostatic surfaces of hSpt5-NGN and *E. coli* wild-type NusG-NTD revealed (Figure 4B) that in *E. coli* NusG-NTD negative charges predominate in the region corresponding to the dimerization interface, whereas in hSpt5, positive and negative charges appear to be more evenly distributed. These differences could also influence the ability to interact.

It has been suggested that a hydrophobic surface on NusG mediates binding to *E. coli* RNA polymerase [46–48]. The region on NusG is located opposite to the surface which corresponds to the Spt4-binding interface of the yeast Spt5-NGN. A hydrophobic surface region is also present in hSpt5 at a similar position, although it is smaller and interrupted

by hydrophilic residues (see Supplementary Figure S3 at <http://www.BiochemJ.org/bj/425/bj4250373add.htm>)

hSpt4 and RpoE'' share core elements

hSpt4 shows distant structural homology to the archaeal RpoE'' from *Pyrococcus furiosus* (PDB: 1RYQ) (Figure 4C) [30,31]. Albeit much smaller, structure alignment with hSpt4 reveals topological similarities for the zinc finger and the corresponding hSpt5-interaction region [RMSD: 1.6 Å, over 341 backbone atoms; sequence identity 23% (ClustalW2)]. The three helical elements α_2' , α_3' and α_4' of hSpt4 are exchanged for unstructured loop regions, which are stabilized by the zinc finger formed by the four cysteine residues Cys⁶, Cys⁹, Cys¹⁸ and Cys²¹ of RpoE''. Apart from Cys¹⁸, these cysteines are arranged in the same relative orientation as the corresponding cysteine residues of hSpt4. Both proteins share the β -sheet ensemble, with one two-stranded antiparallel sheet and a second small three-stranded sheet arranged perpendicularly to each other.

Consequently, RpoE'' contains all the basic and important elements for stability and interaction with NGNs. Potential binding of the two proteins was suggested by docking experiments of RpoE'' with the crystal structure of NusG from *Methanocaldococcus jannaschii* [49] and GST pull-down experiments detected interaction of RpoE'' with NusG from *M. jannaschii* [31]. The role of RpoE'' or Spt4 proteins is still not fully understood. However, despite the structural homology between hSpt4 and RpoE'', this may not necessarily translate into similar functional roles or even a similar localization or interface with the RNA polymerase. So far, all interactions of DSIF with other transcription factors or RNA polymerases appear to be mediated via Spt5 proteins. hSpt4 might thus facilitate interaction of Spt5 with other factors or could simply stabilize the NGN of hSpt5.

AUTHOR CONTRIBUTION

Birgitta Wöhrl conceived and co-ordinated the study. Sabine Wenzel performed all the experiments and participated in data analysis. Berta Martins participated in designing the experiments and performed the structure data analysis. Sabine Wenzel, Birgitta Wöhrl and Berta Martins wrote the paper. Paul Rösch provided conceptual input and critical advice. All authors read and approved the manuscript.

ACKNOWLEDGEMENTS

We thank Britta Zimmermann for excellent technical assistance. We acknowledge the staff of beamline BL14.2 (BESSY, Berlin, Germany) for support during data collection.

FUNDING

The project was funded by grants from the Deutsche Forschungsgemeinschaft (DFG) [grant number Ro617/16–1] and the Study Program-Macromolecular Science from the Elite Network of Bavaria. B.M.M. thanks the Fonds of the Chemical Industry (FCI) for financial support.

REFERENCES

- 1 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
- 2 Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D. K., Hasegawa, J., Omori, M., Inukai, N. et al. (2003) Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Mol. Cell. Biol.* **23**, 1863–1873
- 3 Ivanov, D., Kwak, Y. T., Guo, J. and Gaynor, R. B. (2000) Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol. Cell. Biol.* **20**, 2970–2983

- 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 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
- 6 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
- 7 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
- 8 Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T. and Handa, H. (2006) P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell* **21**, 227–237
- 9 Lis, J. T., Mason, P., Peng, J., Price, D. H. and Werner, J. (2000) P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev.* **14**, 792–803
- 10 Andralis, E. D., Guzman, E., Doring, P., Werner, J. and Lis, J. T. (2000) High-resolution localization of *Drosophila* Spt5 and Spt6 at heat shock genes *in vivo*: roles in promoter proximal pausing and transcription elongation. *Genes Dev.* **14**, 2635–2649
- 11 Wu, C. H., Yamaguchi, Y., Benjamin, L. R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H. and Gilmour, D. (2003) NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev.* **17**, 1402–1414.
- 12 Lindstrom, D. L., Squazzo, S. L., Muster, N., Burckin, T. A., Wachter, K. C., Emigh, C. A., McCleery, J. A., Yates, 3rd, J. R., Hartzog and G. A. (2003) Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell. Biol.* **23**, 1368–1378
- 13 Mandal, S. S., Chu, C., Wada, T., Handa, H., Shatkin, A. J. and Reinberg, D. (2004) Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7572–7577
- 14 Sims, III, R. J., Belotserkovskaya, R. and Reinberg, D. (2004) Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* **18**, 2437–2468
- 15 Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S. and Greenblatt, J. F. (2002) RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell. Biol.* **22**, 6979–6992
- 16 Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F. et al. (1998) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356
- 17 Sullivan, S. L. and Gottesman, M. E. (1992) Requirement for *E. coli* NusG protein in factor-dependent transcription termination. *Cell* **68**, 989–994
- 18 Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E. and Nudler, E. (2008) Termination factor rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* **320**, 935–938.
- 19 DeVito, J. and Das, A. (1994) Control of transcription processivity in phage lambda: Nus factors strengthen the termination-resistant state of RNA polymerase induced by N antiterminator. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8660–8664
- 20 Torres, M., Balada, J. M., Zellars, M., Squires, C. and Squires, C. L. (2004) *In vivo* effect of NusB and NusG on rRNA transcription antitermination. *J. Bacteriol.* **186**, 1304–1310
- 21 Yamaguchi, Y., Wada, T., Watanabe, D., Takagi, T., Hasegawa, J. and Handa, H. (1999) Structure and function of the human transcription elongation factor DSIF. *J. Biol. Chem.* **274**, 8085–8092
- 22 Ponting, C. P. (2002) Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* **30**, 3643–3652
- 23 Reference deleted
- 24 Schwer, B., Schneider, S., Pei, Y., Aronova, A. and Shuman, S. (2009) Characterization of the *Schizosaccharomyces pombe* Spt5-Spt4 complex. *RNA* **15**, 1241–1250
- 25 Hartzog, G. A., Wada, T., Handa, H. and Winston, F. (1998) Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes Dev.* **12**, 357–369
- 26 Kyrpides, N. C., Woese, C. R. and Ouzounis, C. A. (1996) KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.* **21**, 425–426
- 27 Kwak, Y. T., Guo, J., Prajapati, S., Park, K. J., Surabhi, R. M., Miller, B., Gehrig, P. and Gaynor, R. B. (2003) Methylation of SPT5 regulates its interaction with RNA polymerase II and transcriptional elongation properties. *Mol. Cell* **11**, 1055–1066
- 28 Chiang, P. W., Wang, S. Q., Smithivas, P., Song, W. J., Crombez, E., Akhtar, A., Im, R., Greenfield, J., Ramamoorthy, S., Van Keuren, M. et al. (1996) Isolation and characterization of the human and mouse homologues (SUPT4H and Supt4h) of the yeast SPT4 gene. *Genomics* **34**, 368–375
- 29 Wenzel, S., Schweimer, K., Rösch, P. and Wöhrl, B. M. (2008) The small hSpt4 subunit of the human transcription elongation factor DSIF is a Zn-finger protein with α/β type topology. *Biochem. Biophys. Res. Commun.* **370**, 414–418
- 30 Werner, F. (2007) Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* **65**, 1395–1404
- 31 Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G. A., Teng, M. and Niu, L. (2008) Core structure of the yeast Spt4-Spt5 complex: A conserved module for regulation of transcription elongation. *Structure* **16**, 1649–1658
- 32 Mori, S., Abeygunawardana, C., Johnson, M. O. and van Zijl, P. C. (1995) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. B* **108**, 94–98
- 33 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
- 34 Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800
- 35 Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122
- 36 Morris, R. J., Perrakis, A. and Lamzin, V. S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244
- 37 Emsley, P. and Cowtan, K. (2004) COOT: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
- 38 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
- 39 Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
- 40 Reference deleted
- 41 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041
- 42 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291
- 43 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R. et al. (2007) Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947–2948
- 44 Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995) LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134
- 45 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
- 46 Mooney, R. A., Schweimer, K., Rösch, P., Gottesman, M. E. and Landick, R. (2009) Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J. Mol. Biol.* **391**, 341–358
- 47 Belogurov, G. A., Vassilyeva, M. N., Svetlov, V., Klyuyev, S., Grishin, N. V., Vassilyev, D. G. and Artsimovitch, I. (2007) Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol. Cell.* **26**, 117–129
- 48 Nickels, B. E. (2009) Genetic assays to define and characterize protein-protein interactions involved in gene regulation. *Methods* **47**, 53–62
- 49 Zhou, H., Liu, Q., Gao, Y., Teng, M. and Niu, L. (2009) Crystal structure of NusG N-terminal (NGN) domain from *Methanocaldococcus jannaschii* and its interaction with RpoE". *Proteins* **76**, 787–793

Received 10 September 2009/26 October 2009; accepted 27 October 2009

Published as BJ Immediate Publication 27 October 2009, doi:10.1042/BJ20091422

SUPPLEMENTARY ONLINE DATA

Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface

Sabine WENZEL*[†], Berta M. MARTINS[†], Paul RÖSCH*[†] and Birgitta M. WÖHRL*^{†1}

*Universität Bayreuth, Lehrstuhl für Struktur und Chemie der Biopolymere, Universitätsstr. 30, D-95447 Bayreuth, Germany, and [†]Research Centre for Biomacromolecules, Universitätsstr. 30, D-95447 Bayreuth, Germany

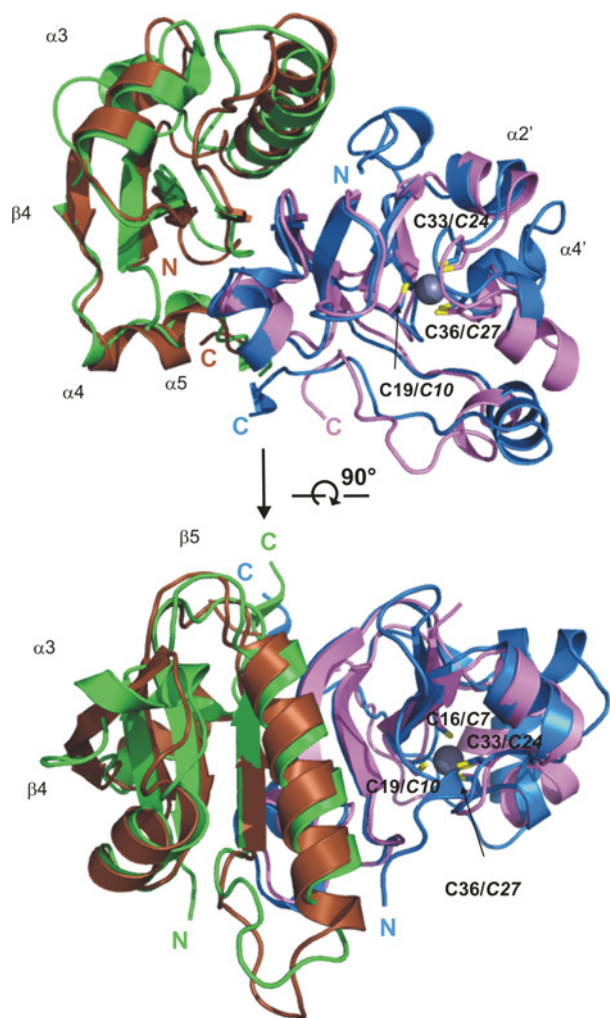


Figure S1 Comparison of human and yeast Spt4–Spt5–NGN complexes

Structural alignment of hSpt4 (blue) in complex with hSpt5–NGN (green) and Spt4 (purple) in complex with Spt5 (286–375) (brown) from *S. cerevisiae* (PDB: 2EXU) as ribbon diagrams. The cysteine residues of both proteins responsible for zinc co-ordination are highlighted as sticks with the sulphur atom coloured in yellow. For clarity, only the Zn²⁺ (grey sphere) of hSpt4 is illustrated, the Zn²⁺ of Spt4 is located in a similar position.

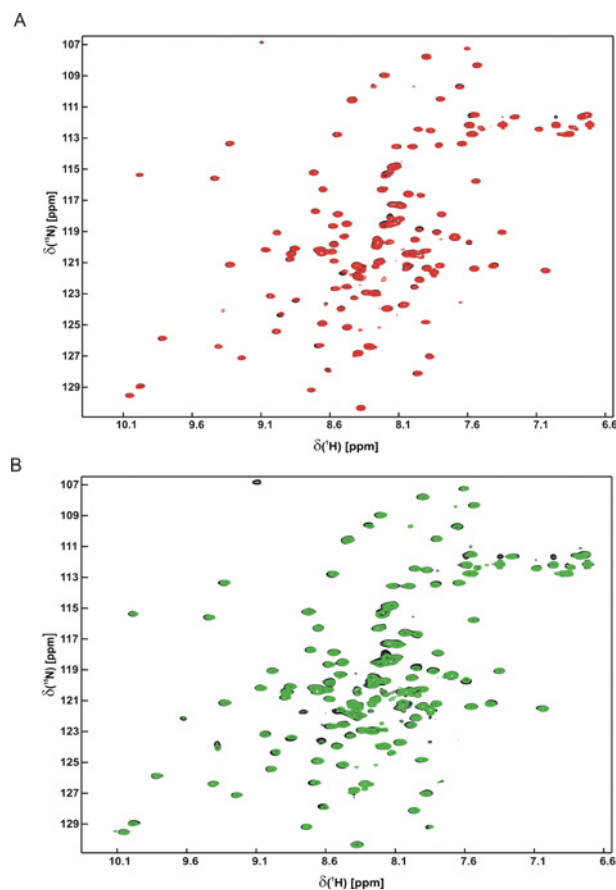


Figure S2 Interaction of hSpt4 with NusG-NTD

(A) ¹⁵N-¹H-HSQC spectra of ¹⁵N-hSpt4 before (black) and after (red) the addition of NusG-NTD. (B) ¹⁵N-¹H-HSQC spectra of ¹⁵N-hSpt4 before (black) and after (green) the addition of NusG-NTD-Q72E. The molar ratio of hSpt4 to NusG-NTD or NusG-NTD-Q72E was 1:3 for both measurements.

¹ To whom correspondence should be addressed (email birgitta.woehrl@uni-bayreuth.de).

The structural co-ordinates reported for the human transcription elongation factor DSIF [DRB (5,6-dichloro-1-β-D-ribofuranosylbenzimidazole) sensitivity-inducing factor], hSpt4/hSpt5 (176–273) will appear in the Protein Data Bank under accession code 3H7H.

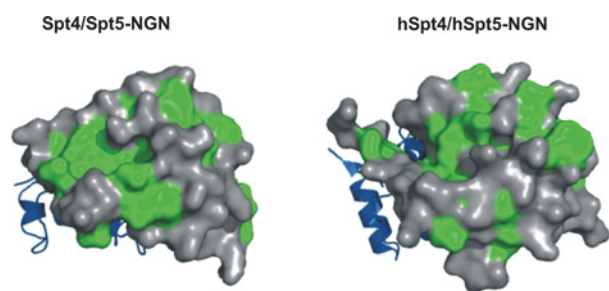


Figure S3 Comparison of the hydrophobic surfaces of the NGN domains of Spt5 and hSpt5

Spt4 and hSpt4 of the corresponding complexes are depicted as blue ribbons. The hydrophobic residues of the NGN domains of Spt5 and hSpt5 are highlighted in green.

Received 10 September 2009/26 October 2009; accepted 27 October 2009
Published as BJ Immediate Publication 27 October 2009, doi:10.1042/BJ20091422

11. Danksagung

Die vorliegende Doktorarbeit wurde zwischen 2006 und 2009 am Lehrstuhl für Struktur und Chemie der Biopolymere, Universität Bayreuth angefertigt.

An dieser Stelle möchte ich mich bei all jenen Personen bedanken, die mich während dieser Zeit in vielfältiger Weise unterstützt haben.

Mein besonderer Dank gilt Prof. Dr. Paul Rösch für die Überlassung dieses interessanten Themas. Mit seinem steten Interesse am Fortgang des Projektes und seiner Unterstützung hat er mir diese Arbeit erst ermöglicht.

Prof. Dr. Birgitta M. Wöhrl danke ich ganz herzlich für die Projektleitung und Unterstützung während aller Höhen und Tiefen in dieser Zeit. Vor allem aber danke ich ihr für die Diskussionsbereitschaft und Offenheit gegenüber neuen Ansatzpunkten.

Bei Dr. Kristian Schweimer bedanke ich mich für die immerwährende Hilfe und Ratschläge aller NMR bezogenen Fragestellungen.

Dr. Berta M. Martins danke ich vielmals für die Einweisung in die Röntgenkristallographie und die Hilfe beginnend bei der Proteinkristallisation, über die Datenmessung am BESSY in Berlin bis hin zur Datenauswertung.

Vielen Dank auch an Dr. Rao Jampani Nageswara für die Vorarbeiten an NELF E RRM und dessen Strukturbestimmung.

Ein sehr großer Dank gilt Britta Zimmermann, die mir in aufopferungsvoller Weise im Labor zugearbeitet und damit ein großes Stück zum Gelingen dieser Arbeit beigetragen hat. Ebenso danke ich Ulrike Persau, Ramona Heissmann und Andrea Hager für ihre aufmunternden Einfälle und tatkräftige Unterstützung.

Nicht vergessen möchte ich alle übrigen Kollegen, die es ermöglichten in einer konstruktiven und durchaus auch humorvollen Umgebung zu arbeiten, die ein offenes Ohr für jegliche Fragen und Probleme hatten und stets bemüht waren eine Lösung zu finden.

12. Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst und keine als die von mir angegebenen Quellen und Hilfsmittel verwendet habe.

Ferner erkläre ich, dass ich nicht anderweitig mit oder ohne Erfolg versucht habe eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

Bayreuth, den 25.11.09