

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole

Tesi di Dottorato di Ricerca in Biochimica

XII Ciclo di dottorato

**Caratterizzazione di proteine nucleari interagenti con il DNA
ripetitivo telomerico dei vertebrati**

Dottoranda: ELEONORA MARSICH

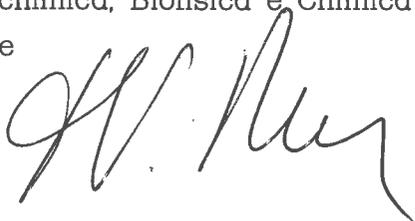
Tutore: Prof. G. MANZINI



Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole
Università di Trieste

Coordinatore: Prof. F. VITTUR

Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole
Università di Trieste





1. INTRODUZIONE	Pag.
1.1 I telomeri dei cromosomi eucariotici: funzione e struttura	1
1.2 Il problema della replicazione delle estremità del DNA cromosomico: la telomerasi	6
1.3 I telomeri e la telomerasi nell'invecchiamento e nella cancerogenesi	12
1.4 Proteine associate ai telomeri	18
2. MATERIALI E METODI	
2.1 Marcatura di sonde di DNA	28
2.2 Colture cellulari	28
2.3 Preparazione degli estratti proteici	29
2.4 Cromatografia di affinità	31
2.4.1 <i>Preparazione della resina di affinità</i>	31
2.4.2 <i>Purificazione dell'estratto proteico</i>	31
2.5 Cromatografia a fase inversa	32
2.6 Tecniche elettroforetiche	33

2.6.1	<i>Elettroforesi denaturante di proteine</i>	33
2.6.2	<i>Recupero delle proteine da gel di poliacrilammide</i>	34
2.6.3	<i>Saggi di ritardo della mobilità elettroforetica (EMSA)</i>	35
2.6.4	<i>U.V. crosslinking</i>	36
2.6.5	<i>U.V. crosslinking in situ</i>	36
2.6.6	<i>Super-shift</i>	37
2.6.7	<i>Western blot</i>	37
2.6.8	<i>South-western</i>	38
2.7	Spettrometria di massa "Ion spray"	39
2.8	Spettrometria di massa MALDI-TOFF	39
2.8.1	<i>Digestione su gel e purificazione dei peptidi</i>	39
2.8.2	<i>Spettrometria di massa</i>	40
2.8.3	<i>Identificazione delle proteine</i>	41
2.9	Preparazione del DNA plasmidico	42
2.9.1	<i>Preparazioni in piccola scala</i>	42
2.9.2	<i>Preparazione su larga scala</i>	42
2.10	Clonaggio del gene di SSDBF	43
2.11	Preparazione delle cellule E. Coli (ceppo BL 21) competenti	45
2.12	Trasformazione dei batteri BL 21 con i vettori di espressione ed induzione dell'espressione proteica	46

3. RISULTATI E DISCUSSIONE

3.1	Identificazione di una proteina che lega ad alta specificità il motivo telomerico d(CCCTAA) _n a filo singolo	47
3.2	Determinazione dei pesi molecolari dei complessi proteina-DNA	50
3.3	Presenza della proteina in estratti nucleari da diverse origini	51
3.4	Effetti della modulazione del motivo telomerico sul legame della proteina	52
3.5	Isolamento delle proteine allo studio	56

3.6	Identificazione delle proteine isolate	59
3.7	Espressione di ssDBF ricombinante	64
3.8	CBNP α e ssDBF: una proteina, pi \dot{u} funzioni?	67
3.9	Le proteine hn RNP di tipo A/B	70
3.10	Prospettive future nella caratterizzazione di ssDBF	72
3.11	Purificazione della proteina da cellule K562	75
3.12	Identificazione delle proteine isolate da cellule K562	77
3.13	Caratterizzazione delle proteine isolate da cellule K562: loro ruolo e possibile relazione funzionale con CBNP α	82
4.	CONCLUSIONI	88
5.	BIBLIOGRAFIA	91

1.1 I telomeri dei cromosomi eucariotici: funzione e struttura

I telomeri sono strutture deossiribonucleotidiche altamente specializzate che costituiscono le porzioni terminali di ciascuna estremità dei cromosomi eucariotici.

Fu proposta per la prima volta negli anni '30 da Hermann Muller e da Barbara Mc-Clintock l'ipotesi che le regioni terminali dei cromosomi lineari degli eucarioti avessero delle caratteristiche che le rendevano diverse rispetto al resto del cromosoma. Muller, lavorando sulla *Drosophila*, realizzò che, dopo irraggiamento con i raggi X, si poteva osservare nel nucleo una grande varietà di cromosomi riarrangiati, ma solo raramente si trovavano cromosomi con delezioni o inversioni terminali. Questo suggerì che i cromosomi privi delle loro estremità naturali fossero instabili e venissero persi alla prima divisione cellulare. Muller coniò quindi il termine «telomeri» per definire le estremità dei cromosomi lineari. Negli stessi anni Barbara McClintock osservò che in *Zea mays* i cromosomi rotti, le cui estremità non sono quindi costituite dai telomeri, tendono a fondersi l'un l'altro formando cromosomi dicentrici. Questi cromosomi sono instabili: infatti,

durante l'anafase, il cromosoma può rompersi ed andare incontro a successivi riarrangiamenti. I primi dati citologici indicavano quindi che i telomeri sono essenziali per la stabilità dei cromosomi e le loro caratteristiche permettono alle cellule di differenziare le estremità naturali dei cromosomi da rotture a doppia elica.

La precisa costituzione dei telomeri non venne determinata fino al 1985, anno in cui Elizabeth Blackburn riuscì a caratterizzare il DNA telomerico del protozoo ciliato *Tetrahymena termophila*, il quale risultò essere costituito da una sequenza nucleotidica semplice ed estremamente breve (TTGGGG), ripetuta decine di volte (Greider *et al.*, 1985). Il filamento ricco in G corre nella direzione 5'→3' e si estende per un breve tratto a singola elica, circa 12-18 nucleotidi, oltre il filamento complementare ricco in C. I ciliati sono stati gli organismi d'elezione per lo studio dei telomeri in quanto il loro macronucleo contiene milioni di molecole di DNA indipendenti e quindi milioni di telomeri. Solo successivamente fu possibile il clonaggio dei telomeri in altri eucarioti ed esperimenti di ibridazione incrociata tra i telomeri di organismi diversi hanno dimostrato un'elevata conservazione della sequenza e della struttura dei telomeri. Il DNA telomerico è composto nella stragande maggioranza dei casi da una breve sequenza nucleotidica, ripetuta in tandem da decine a migliaia di volte secondo l'organismo e caratterizzata da blocchi di guanosine consecutive sul filamento terminale al 3' (Figura 1). Come in *Tetrahymena*, questo protrude rispetto al complementare per un tratto di lunghezza variabile nelle diverse specie (McElligot *et al.*, 1997). In tutti i vertebrati, incluso l'uomo, il motivo è: TTAGGG, ripetuto un elevato numero di volte (Tabella1).

Sequenza del motivo	Specie
TTTTGGGG	<i>Euplotes, Oxytricha</i>
TTGGGG	<i>Glaucoma, Tetrahymena</i>
TTGGGG e TTTGGG	<i>Paramecium tetraurelia</i>
TTAGGG e TTCAGGG	<i>Plasmodium sp.</i>
TTAGGG	<i>Trypanosoma brucei</i>
TAGGG	<i>Giardia lamblia</i>
(TG) ₁₋₆ TG ₂₋₃	<i>Saccharomyces cerevisiae</i>
TTACAG ₁₋₈	<i>Saccharomyces pombe</i>
TTAGGGGG	<i>Cryptococcus neoformans</i>
TTAGGG	<i>Aspergillus nidulans</i>
TTAGGG	<i>Podospora anserina</i>
TTAGGG	<i>Fusarium oxysporum</i>
TTAGGG	<i>Didymium iridis</i>
TTAGGG	<i>Physarum polycephalum</i>
AG ₁₋₈	<i>Dictyostelium discoideum</i>
TTAGGC	<i>Ascaris lumbricoides</i>
TTGCA	<i>Parascaris univalens</i>
TTAGG	<i>Bombix mori</i> e altri insetti
TTAGGG	<i>Homo</i> e vertebrati
TTTAGGG	<i>Arabidopsis thaliana</i>
TT(T/A)AGGG	<i>Lycopersicon esculentum</i>
TTTTAGGG	<i>Chlamydomonas reinhardtii</i>
ACGGATGTCTAACTTCTTGGTG	<i>Candida albicans</i>
HeT-A, TART (retroposons)	<i>Drosophila melanogaster</i>

TABELLA 1: Motivi del DNA telomerico in organismi eucarioti

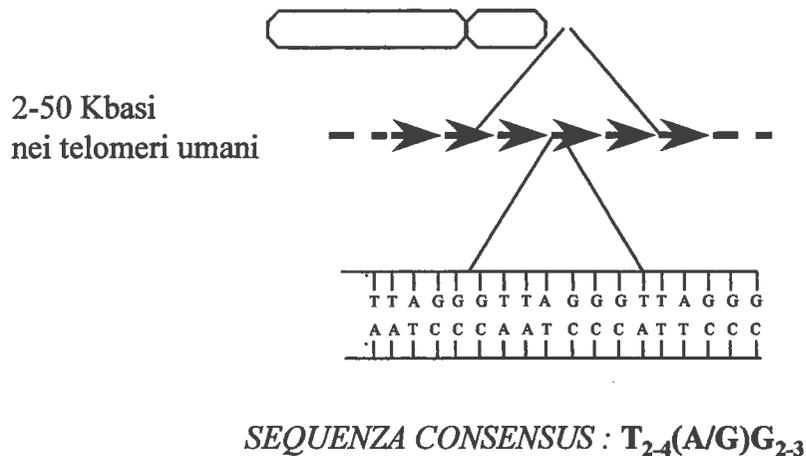


Figura 1: Struttura dei telomeri

Vi sono alcune eccezioni a questa regola, in alcuni lieviti ad esempio le ripetizioni sono più lunghe o degenerate, ma l'eccezione più eclatante riguarda i ditteri: nella *Drosophila*, il primo organismo che ha portato ad identificare i telomeri, questi non sono costituiti da corte sequenze ripetute, ma da più copie di elementi trasponibili che hanno una specifica affinità per le estremità cromosomiche (Pardue, 1996).

L'estensione dell'intera ripetizione telomerica è notevolmente diversa tra le varie specie. Per ciascuna specie si può definire una caratteristica lunghezza media dei telomeri (ad esempio, poche centinaia di coppie di basi nei ciliati e nei lieviti, alcune chilobasi nell'uomo ed anche decine di kilobasi nel topo), (Kipling *et al.*, 1990), tuttavia il numero di ripetizioni varia da cellula a cellula di uno stesso organismo, così come da cromosoma a cromosoma di una stessa cellula .

Recenti osservazioni (Griffith *et al.*, 1999) fanno supporre che nei mammiferi i cromosomi anziché terminare come molecole di DNA lineare, come ritenuto finora, formino delle strutture a cappio, chiamate t-loops (telomere loops). Il

filamento ricco in G sporgente al 3', che nei mammiferi ha sequenza (TTAGGG)_n, viene ripiegato all'indietro ed inserito all'interno del doppio filamento appaiandosi alle ripetizioni complementari CCCTAA interne, scalzando il filamento polipurinico presente nel duplex (Figura 2)

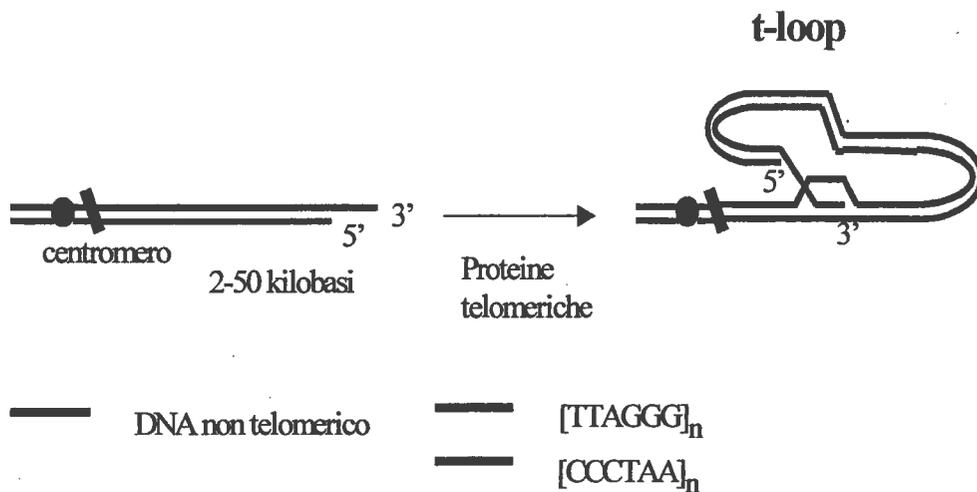


Figura 2: Struttura del t-loop

Frazioni arricchite di DNA telomerico isolate da cellule HeLa mostrano una frazione di t-loops che va da un minimo del 15% ad un massimo del 40 % del totale delle molecole. Anche il DNA telomerico isolato da cellule di fegato di topo manifesta un'elevata frazione di t-loops. Gli anelli contengono solo ripetizioni telomeriche, sono molto grandi e spesso comprendono fino a 10–20 kb di DNA. Si è visto che cellule note per possedere telomeri molto lunghi, formano anelli più grandi di quelle recanti telomeri più corti, il che implica la presenza di un

meccanismo che regoli la dimensione dell'anello. Si suppone che la lunghezza minima di DNA necessaria per la formazione del laccio sia specie specifica così come lo sono la lunghezza del DNA telomerico a doppio filamento ricco in G sporgente al 3'.

La scoperta dei t-loops suggerisce che la protezione dei telomeri dall'azione di endonucleasi sia dovuta al sequestro dell'estremità potenzialmente vulnerabile ricca di G sporgente al 3' all'interno del duplex di DNA, piuttosto che al legame tenace su di esso di proteine specifiche come ipotizzato finora. Inoltre gli anelli di DNA telomerico possono essere la spiegazione di come la cellula sia in grado di distinguere l'estremità del cromosoma da una lesione al DNA. Si ritiene che il macchinario enzimatico che riconosce le lesioni al DNA, necessiti dell'accesso all'estremità del DNA danneggiato. Il sequestro della coda sporgente al 3' all'interno del loop impedirebbe un'inopportuna attivazione dei meccanismi di riparo i quali possono essere i responsabili anche della degradazione del filamento ricco in G e della fusione tra le estremità dei cromosomi (Karlsender *et al.*, 1999). In questo modo, per stabilizzare l'estremità dei cromosomi sarebbe sufficiente impedirne il riconoscimento da parte dei meccanismi di riparo.

La perdita di funzionalità dei telomeri, osservata nelle cellule prive dell'enzima preposto alla sintesi dei telomeri stessi, la telomerasi, può essere dovuta proprio alla distruzione dei t-loops. Nei vari organismi in cui è stato osservato l'accorciamento dei telomeri, si è visto che questi perdono la capacità di proteggere efficacemente i cromosomi dalla fusione prima che l'intera sequenza telomerica vada perduta. La spiegazione più plausibile è che, affinché si formi l'anello, sia necessaria una lunghezza minima di DNA. Una volta che i telomeri sono diventati troppo corti, diventano incapaci di formare l'anello e quindi di proteggere le estremità cromosomiche.

Nei ciliati comunque la lunghezza del DNA telomerico è in genere troppo modesta per consentire la chiusura ad anello senza dispendio di energia. Ad

esempio i telomeri di *Oxytricha* consistono di solo 20 bp di DNA a doppio filamento contenenti la sequenza TTTTGGGG/CCCCAAAA, troppo poco per formare loops telomerici. E' possibile che proteine leganti le estremità cromosomiche libere sostituiscano ai fini della protezione del telomero, la struttura ad anello di tutte le specie dotate di telomeri più lunghi.

Tuttavia bisogna essere cauti nell'affermare che i ciliati risolvano un problema biologico di importanza così fondamentale come la protezione delle estremità cromosomiche in modo diverso dagli altri eucarioti, dal momento che molte strutture altamente conservate, come le sequenze telomeriche e l'RNA telomerasico, sono state scoperte inizialmente proprio nei ciliati.

1.2 Il problema della replicazione delle estremità del DNA

cromosomico: la telomerasi

Come già accennato, i telomeri sono essenziali per la stabilità cromosomica. Un altro ruolo fondamentale da essi svolto è quello di permettere la replicazione delle estremità 3' delle molecole lineari del DNA cromosomico.

Infatti le DNA polimerasi non sono in grado di copiare i cromosomi lineari per tutta la loro lunghezza fino all'estremità (Figura 3). Il meccanismo di replicazione classico lascia un piccolo segmento non copiato sull'estremità del filamento ritardato. In teoria, se le cellule non avessero modo di compensare questa incompletezza, i cromosomi si accorcerebbero ad ogni divisione cellulare. Questa continua erosione finirebbe per eliminare i telomeri e dopo di essi porzioni di DNA codificante, nell'arco di qualche generazione (Watson *et. al.*, 1972)

Il problema del mantenimento dell'integrità del proprio corredo cromosomico, è stato affrontato dai vari organismi con due strategie chiave. Gli organismi

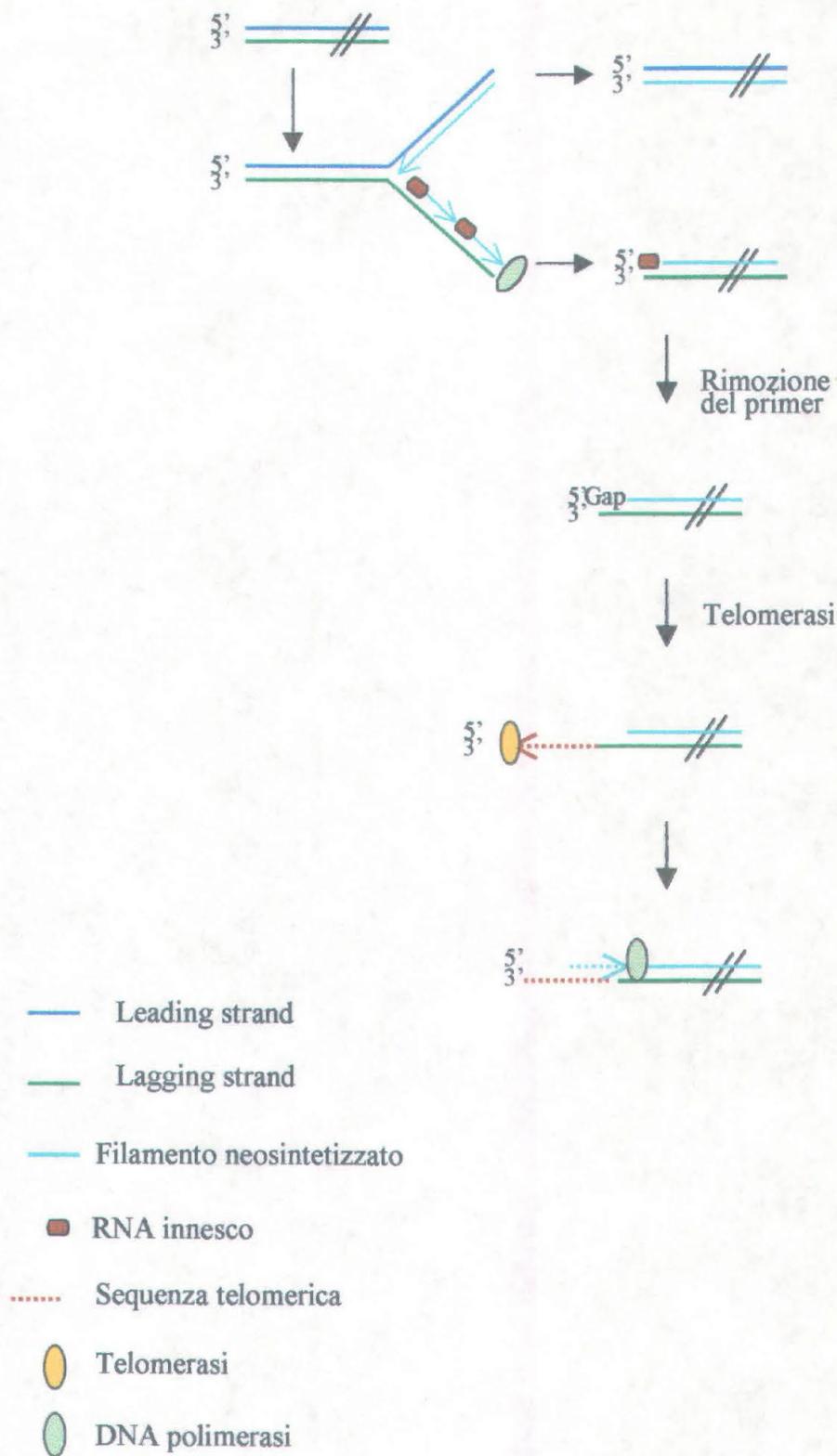


Figura 3: La replicazione dei telomeri

unicellulari procarioti, come i batteri ed anche gli eucarioti per quanto riguarda il loro DNA extracromosomico, mitocondriale e cloroplastico, hanno aggirato il problema eliminando i terminali mediante la ciclizzazione del DNA stesso.

Per quanto riguarda i cromosomi lineari eucariotici, la soluzione si è dimostrata l'evoluzione di un enzima apposito, la telomerasi, che è in grado di aggiungere ripetizioni telomeriche all'estremità 3' della molecola di DNA e contrastare così l'incompleta replicazione da parte della DNA polimerasi.

L'attività telomerasica è stata identificata per la prima volta nel ciliato *Tetrahymena* (Greider e Blackburn, 1985) e successivamente in un ampio range di eucarioti che va da organismi unicellulari quali i ciliati *Oxytricha* e *Euplotes*, a piante ed animali superiori, incluso l'uomo (Zahler *et al.*, 1989; Shippen- Lentz, 1989; Morin, 1989).

La telomerasi è una ribonucleoproteina DNA-polimerasica RNA-dipendente. La molecola di RNA contiene una sequenza complementare all'elica ricca in G che funziona da stampo per la sintesi delle ripetizioni telomeriche. Nella componente proteica risiede invece l'attività catalitica che sintetizza il DNA telomerico copiando lo stampo di RNA. La telomerasi può quindi essere classificata come una trascrittasi inversa il cui stampo ad RNA è parte integrante dell'enzima. In figura 4 è schematicamente illustrato il modello per il funzionamento dell'enzima. La telomerasi utilizza l'estremità 3' del DNA telomerico come innesco e sintetizza un'unità ripetuta copiando la sequenza stampo contenuta nell'RNA: a questo punto il complesso enzimatico si riposiziona rispetto alla molecola di DNA ed è pronto ad iniziare un nuovo ciclo di sintesi. La telomerasi è in grado di allungare solo il filamento ricco in G; per ottenere il completo allungamento del telomero è quindi necessario che il filamento ricco in C venga sintetizzato successivamente. Si ritiene che la sintesi del filamento complementare a blocchi di citosine, avvenga ad opera di una

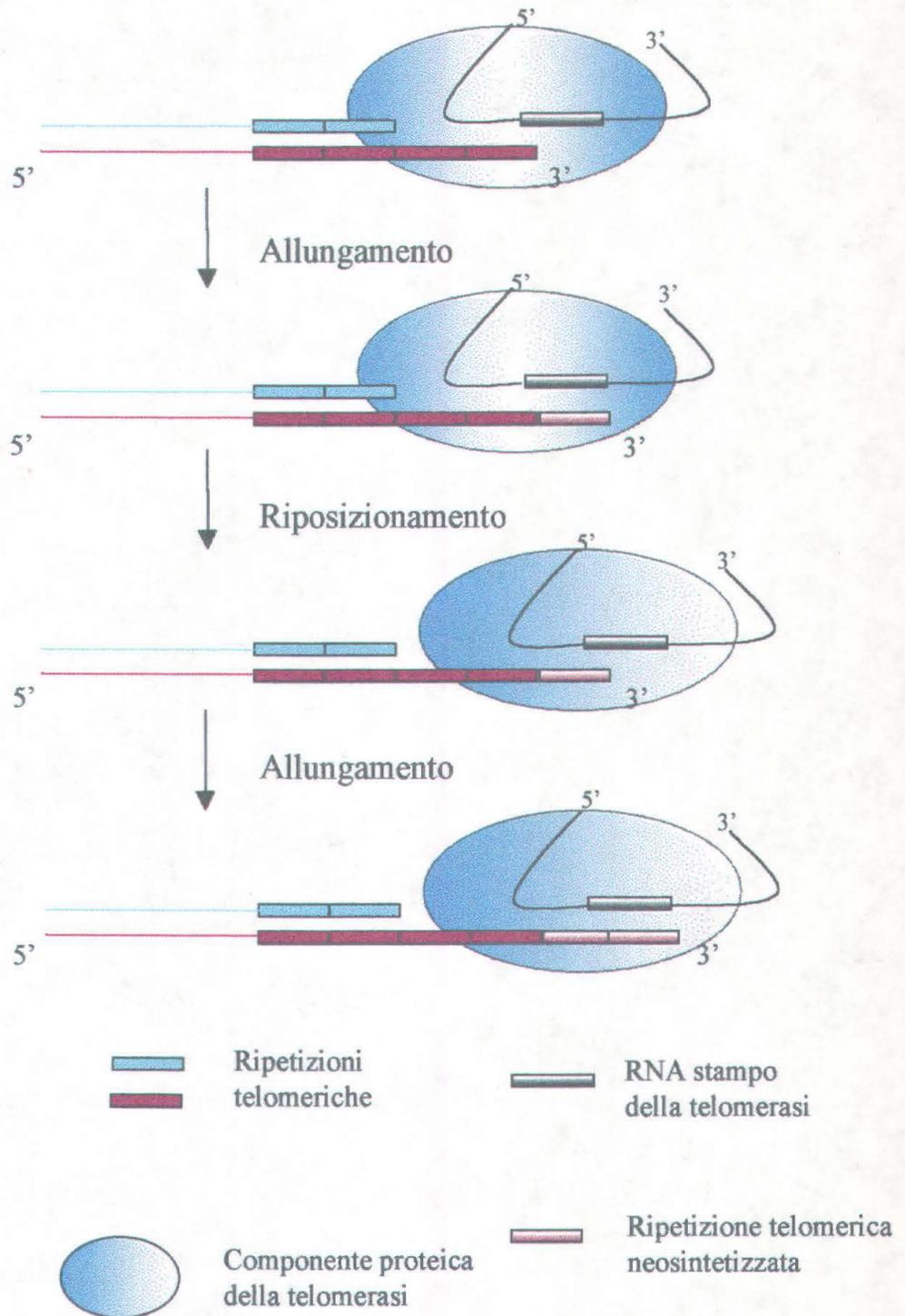


Figura 4: Modello di allungamento della telomerasi

primasi-polimerasi α con un meccanismo di tipo semiconservativo, utilizzando il filamento a guanine come stampo (Shampay *et al.*, 1984)

La telomerasi presenta solo una generica affinità per le sequenze ricche in G, a differenza di molte proteine leganti il DNA o l'RNA le quali riconoscono il loro bersaglio in modo altamente sequenza specifico (Morin, 1991; Lee & Blackburn, 1993). Questa preferenza per le sequenze telomeriche od oligonucleotidi sintetici riproducenti il motivo telomerico, senza specificità per un'unica sequenza, suggerisce che la telomerasi riconosca una struttura secondaria comune presente in tutti i telomeri e che il meccanismo di allungamento sia altamente conservato. Oligonucleotidi sintetici, riproducenti motivi telomerici, hanno la proprietà strutturale di assumere in soluzione ripiegamenti intra-ed intermolecolari stabilizzati da appaiamenti G:G, non di tipo Watson-Crick (Henderson *et al.*, 1987). In particolare, ripiegamenti a forcina con legami idrogeno G:G, sono stati proposti quali possibili candidati per strutture riconosciute dalla telomerasi (Zahler *et al.*, 1991).

I geni codificanti la componente ad RNA sono i primi ad essere stati clonati in diverse specie tra cui l'uomo (Autexier *et al.*, 1994; Singer *et al.*, 1994; Feng *et al.*, 1995). La dimensione dell' RNA della telomerasi varia da un valore minimo di 160-190 nucleotidi nei ciliati a più di 1300 basi nei lieviti. Tecniche convenzionali di ibridazione a bassa stringenza di librerie a c-DNA di diversi tipi cellulari umani, hanno permesso di caratterizzare l'RNA telomerasico umano (Feng *et al.*, 1995). L'RNA trascritto maturo, indicato con la sigla hTR, è lungo 450 nucleotidi, dei quali costituiscono la stampo per il DNA telomerico solo i nucleotidi 5'-CUAACCCUAAC-3' dal 46 al 56. La comparazione con l'omologo murino ha evidenziato regioni conservate solo in corrispondenza dello stampo.

La comparazione delle sequenze degli RNA da diverse specie ha dimostrato che non vi è alcuna evidente somiglianza tra gli RNA, né per dimensioni né per la sequenza, il che porta a concludere che la sequenza primaria dell'RNA

telomerasico, a parte il breve stampo, non è conservata. Al contrario il confronto tra gli RNA telomerasici dei ciliati, organismi tra loro strettamente affini e per i quali sono note una trentina di sequenze, ha evidenziato un'alta conservazione della struttura secondaria. Si ritiene che tale struttura nei ciliati sia costituita da una serie di steli, di anse e di pseudonodi, il che suggerisce l'esistenza di un ripiegamento tridimensionale anch'esso ben conservato. La struttura terziaria teorizzata non è ancora stata confermata da studi strutturali. Si ritiene che tali strutture abbiano non solo il ruolo di recare lo stampo e di renderlo accessibile alla subunità catalitica, ma anche di fungere da impalcatura per il legame di specifiche proteine accessorie indispensabili per il corretto funzionamento della telomerasi. Si è visto, infatti, che tale RNA è complessato con più componenti proteiche diverse, compresa la subunità contenente l'attività catalitica della telomerasi.

La conservazione della struttura secondaria e probabilmente anche della terziaria, unitamente alla capacità della telomerasi di sfruttare inneschi di specie molto diverse, suggerisce che il meccanismo di allungamento operato dall'enzima si sia mantenuto in tutti gli eucarioti.

Le componenti proteiche della telomerasi sono state clonate solo recentemente. In lievito il clonaggio del gene per la subunità catalitica della telomerasi (Lingner *et al.*, 1997) rivelò, come ci si aspettava dal supposto meccanismo di funzionamento dell'enzima, una parziale omologia di sequenza con le trascrittasi inverse virali. Per questo la subunità catalitica della telomerasi venne chiamata TERT (Telomerase Reverse Transcriptase). La sua identificazione ha consentito non solo di aprire la strada nel definire la composizione del complesso telomerasico, ma ha anche facilitato la manipolazione dell'enzima sia *in vivo* che *in vitro*, consentendo così di ampliare le conoscenze sulle modalità di sintesi del DNA telomeric

La subunità catalitica (TERT) venne individuata usando approcci complementari diversi. La purificazione eseguita da Cech portò all'isolamento dal

ciliato *Euplotes aediculatus* (Lingner *et al.*, 1996) di un grosso complesso RNA-proteine dotato di attività telomerastica. Il complesso conteneva una subunità a RNA di 66 kDa codificante per il repeat telomerico, una proteina di 123 kDa (p123) il cui gene venne clonato con la tecnica dell'antisense, ed una seconda proteina di 43 kDa.

L'analisi compiuta dal gruppo della Lunblad su ceppo di *S.Cerevisiae* mutante, il quale mostrava un accorciamento dei telomeri, portò all'identificazione di un numero di geni detti EST (Ever Shorter Telomere) (Lendvay *et al.*, 1996). Il prodotto del gene EST2 (Est2p), si dimostrò, essere l'omologo della subunità 123 dei ciliati (Linger *et al.*, 1997). Più recentemente sono stati identificati gli omologhi della p123 dei ciliati (Collins *et al.*, 1998; Bryan *et al.*, 1998) e della Est2p in lievito, anche nel topo e nell'uomo (Kilian *et al.*, 1997). Recenti analisi filogenetiche (Nakamura *et al.*, 1998), suggeriscono che gli elementi genetici più vicini alle TERT siano le trascrittasi inverse non -LTR (Long Terminal Repeat).

La prova diretta che la TERT contiene l'attività polimerasica della telomerasi, venne da esperimenti di mutagenesi sito specifica dei residui di aspartato che sono altamente conservati nelle trascrittasi inverse e nelle TERT. Questi residui sono noti costituire la tasca catalitica delle trascrittasi inverse. La loro sostituzione nella TERT di *S.cerevisiae* e in quella umana provoca l'abolizione dell'attività telomerastica. Ulteriori prove che la TERT rappresenta la subunità catalitica della telomerasi, derivano da esperimenti in cui la subunità ad RNA (detta TER) e la TERT umana trascritta e tradotta *in vitro*, si sono dimostrate sufficienti per ricostituire l'attività telomerastica (Weinrich *et al.*, 1997).

Al giorno d'oggi sono stati clonati i geni codificanti per sette TERT diverse. Codificano tutti per grosse proteine di peso molecolare compreso tra 103 kDa e 143 kDa (Bryan *et al.*, 1998). Le trascrittasi inverse contengono nella loro sequenza un insieme di sette motivi caratteristici i quali sono stati riscontrati in tutte le TERT identificate finora. Il confronto delle sequenze mostra una elevata

omologia tra i sette motivi di TERT diverse e questa omologia diventa ancora più elevata tra gli organismi evolutivamente vicini, come il topo e l'uomo, oppure all'interno della famiglia dei ciliati (Bryan *et al.*, 1998)

La componente catalitica e l'RNA della telomerasi non sembrano essere le uniche molecole che compongono il complesso ribonucleoproteico cui è associata attività telomerasica. Recentemente sono stati clonati in *Tetrahymena* i geni che codificano per due proteine, p80 e p95, che si associano all'RNA telomerasico (Greider *et al.*, 1995, Gandhi *et al.*, 1998). Queste proteine probabilmente costituiscono due subunità funzionali della telomerasi. La subunità p95 interagisce con i *primers* telomerici mentre studi *in vitro*, suggeriscono che p80 sia la subunità che porta l'RNA stampo. Mentre fino ad ora non è stato identificato in alcun altro organismo un gene omologo a p95, proteine con un'omologia significativa a p80 sono state trovate nell'uomo, nel ratto e nel topo (Nakayama *et al.*, 1997; Meyerson *et al.*, 1997). Anche in questi vertebrati le proteine omologhe a p80 sono associate con l'RNA per la telomerasi. Nel lievito non è stata trovata alcuna proteina omologa a p80; è probabile che in questo organismo si siano evolute altre proteine che interagiscono con la telomerasi.

Negli organismi eucarioti unicellulari, lieviti e protozoi, la telomerasi è espressa costitutivamente; mentre negli organismi superiori la sua espressione è regolata durante lo sviluppo. Nell'uomo vi è attività telomerasica nei tessuti fetali, mentre appena dopo la nascita essa non è riscontrabile nella maggior parte delle cellule somatiche (Haerley, 1990; Alsopp *et al.*, 1992), con conseguente progressivo accorciamento dei telomeri ad ogni replicazione. Bassi livelli di attività enzimatica sono stati descritti nelle cellule staminali del sangue e nella mucosa intestinale mentre solo nelle cellule germinali alti livelli di enzima sono presenti durante tutta la vita. L'osservazione che, al contrario delle cellule somatiche normali, le cellule tumorali e le linee cellulari immortalizzate esprimono in genere la telomerasi ha ulteriormente aumentato l'interesse nei

confronti dei telomeri e della telomerasi. Come verrà discusso in seguito, lo spegnimento dell'attività telomerasica nelle cellule somatiche normali e la sua riattivazione nelle cellule tumorali sembrano svolgere un ruolo importante nei processi di senescenza e di cancerogenesi.

1.3 I telomeri e la telomerasi nell'invecchiamento e nella cancerogenesi

Attraverso la sovraespressione di una forma inattiva della componente ad RNA della telomerasi in *Tetrahymena* o attraverso l'introduzione di mutazioni del gene che la codifica in diverse specie di lievito (Lustig *et al.*, 1990; Krauskopf *et al.*, 1996), è stata abolita l'attività della telomerasi ed è stato dimostrato che in queste condizioni i telomeri si accorciano nei successivi cicli cellulari, finché le cellule muoiono. La morte sembra sopraggiungere quando i telomeri si sono accorciati al di sotto di una lunghezza soglia. Nei ciliati e nei lieviti, quindi, la telomerasi è necessaria perché essi possano propagarsi indefinitamente. Negli organismi superiori, dove invece la telomerasi è inattivata nelle cellule differenziate, è stato proposto un modello che collega i telomeri e la telomerasi con la senescenza cellulare da un lato e con la trasformazione cellulare e la cancerogenesi dall'altro.

La maggior parte delle cellule somatiche dei mammiferi, prive di attività telomerasica, presenta una vita limitata in cultura. Infatti dopo un numero finito di divisioni, in cui i telomeri si accorciano progressivamente, queste cellule non sono più in grado di moltiplicarsi ed entrano in una fase detta di senescenza cellulare o senescenza replicativa, durante la quale non rispondono più ad alcuno stimolo replicativo. Il massimo numero di divisioni che le cellule somatiche possono compiere in coltura dipende sia dall'età del donatore sia dalla longevità della

specie da cui derivano (Harley *et al.*, 1995). La senescenza replicativa non sembra essere limitata alle cellule coltivate *in vitro*, ma sembra interessare anche le cellule dell'organismo ed è stato proposto un suo ruolo nell'invecchiamento dell'organismo (Campisi, 1997).

In vitro un prolungamento della capacità replicativa può essere indotto se le cellule vengono trattate con carcinogeni o transfettate con virus oncogeni (Klingelhuntz *et al.*, 1994; Counter *et al.*, 1994). Normali cellule somatiche umane, transfettate con virus oncogeni, sono in grado di protrarre le divisioni cellulari oltre il normale punto di senescenza, fino al raggiungimento di uno stato, definito di «crisi», in cui i telomeri raggiungono una lunghezza minima. Tali cellule non sono immortali, non vi è attività telomerasica ed i telomeri continuano ad accorciarsi con la divisione cellulare, mentre aumenta progressivamente l'instabilità cromosomica, analogamente a quanto si osserva per le cellule non immortalizzate. Dopo un limitato numero di divisioni, la maggior parte delle cellule muore a causa dell'accumularsi di aberrazioni cromosomiche. Le rare cellule che superano lo stato di «crisi», sono immortali, tipicamente aneuploidi, esprimono l'attività telomerasica e presentano telomeri a lunghezza stabile, sia pure generalmente ridotta (Allsopp *et al.*, 1992).

In questa prospettiva, l'accorciamento dei telomeri è stato paragonato ad un orologio che conta il numero di divisioni che la cellula compie e, quando ne è stato compiuto il numero massimo, ne blocca la moltiplicazione segnalando l'ingresso nella fase di senescenza (Harley *et al.*, 1995). Tra le evidenze in favore di questa ipotesi si possono ricordare gli esperimenti condotti su fibroblasti umani in cultura primaria che hanno dimostrato come il numero di divisioni a cui può andare incontro una linea di cellule somatiche in cultura, è correlato con la lunghezza dei suoi telomeri. Le cellule con telomeri più lunghi si dividono più volte di quelle con telomeri più corti (Allsopp *et al.*, 1992). Inoltre, inducendo l'espressione della telomerasi in tre diversi tipi di cellule somatiche umane,

introducendovi il gene per la subunità catalitica dell'enzima, è stato dimostrato che l'espressione della telomerasi, non solo induceva un allungamento delle estremità telomeriche, ma determinava anche un'alterazione delle capacità replicative. Mentre le cellule parentali entravano in senescenza dopo un numero definito di divisioni, le cellule in cui era stata indotta l'espressione della telomerasi, continuavano a dividersi oltre questo limite. (Bodnar, 1998, Vaziri *et al.*, 1998). Questi risultati indicano che esiste un effettivo ruolo dell'accorciamento dei telomeri nella senescenza cellulare *in vivo*. Comunque ottenere prove dirette del ruolo della lunghezza dei telomeri in colture primarie umane è molto difficile perché in ogni cellula ci sono 92 estremità cromosomiche, ognuna delle quali può variare in lunghezza. E' questo il motivo principale per cui ci sono ancora poche evidenze sperimentali che dimostrino il ruolo causale dei telomeri nella senescenza cellulare, sebbene la cosa sia ipotizzata da tempo. Nello stesso tempo è ancora da determinare quale ruolo, se ce n'è uno, questo processo svolga nel complesso fenomeno che è l'invecchiamento dell'organismo.

Sono stati proposti vari modelli per spiegare il meccanismo con cui l'accorciamento dei telomeri potrebbe segnalare l'ingresso nella senescenza e bloccare la divisione cellulare. Considerando la notevole eterogeneità nella lunghezza dei telomeri nell'ambito di una stessa cellula, è stato proposto che il segnale parta da uno o pochi cromosomi i cui telomeri avrebbero raggiunto una lunghezza al di sotto di quella necessaria per una corretta funzionalità. Queste estremità cromosomiche verrebbero riconosciute dalla cellula come rotture a doppia elica e attiverrebbero così i meccanismi di controllo del ciclo cellulare che bloccano la moltiplicazione cellulare se sono presenti danni. Alternativamente, l'accorciamento dei telomeri potrebbe determinare un cambiamento nel *pattern* di espressione di geni subtelomerici importanti per il controllo della proliferazione cellulare (Goldstein *et al.*, 1990; Allsop *et al.*, 1992) . A supporto della prima ipotesi vi è l'osservazione che tra gli agenti che trasformano le cellule umane, e

determinano quindi il superamento della senescenza, vi è l'antigene T di SV40. Questa molecola lega ed inattiva p53, uno dei principali fattori coinvolti nella regolazione del ciclo cellulare come risposta alla presenza di danni nella molecola di DNA (Lane *et al.*, 1992; de Lange, 1994).

D'altra parte esistono anche evidenze di segno opposto all'ipotesi che la lunghezza dei telomeri costituisca un segnale per l'entrata in senescenza. Infatti, recentemente è stato dimostrato che le cellule umane dell'epitelio urinario coltivate *in vitro* presentano, a differenza delle cellule *in vivo*, attività telomerasica e telomeri stabili in lunghezza. Pur presentando queste caratteristiche, entrano in senescenza, il che suggerisce che tali cellule non usino l'accorciamento dei telomeri come segnale principale dell'entrata in senescenza (Belair *et al.* 1997). Vi sono numerosi altri esempi di cellule che entrano in senescenza pur non presentando un accorciamento dei telomeri (Reddel *et al.*, 1998) Può darsi che l'accorciamento dei telomeri sia solo uno dei meccanismi che possono indurre la senescenza nelle cellule umane.

In Figura 5 è schematicamente rappresentato il modello proposto per il ruolo dei telomeri e della telomerasi nella senescenza e nell'immortalizzazione cellulare

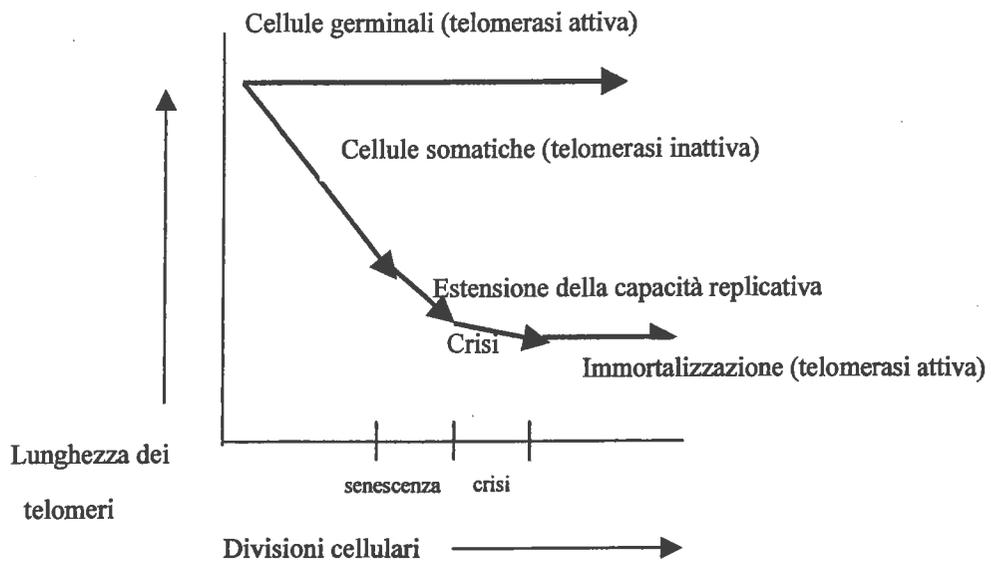


Figura 5: Modello per il ruolo dei telomeri e della telomerasi nella senescenza e nella immortalizzazione cellulare

Come precedentemente osservato, l'attività telomerasica sembra essere un elemento significativo nell'immortalizzazione cellulare, suggerendo che vengano selezionate le cellule che la esprimono o presentano una sua riattivazione. I dati ottenuti nelle cellule immortalizzate *in vitro* hanno fatto ipotizzare che la riattivazione della telomerasi possa avere un ruolo importante anche nella tumorigenesi. A supporto di questa ipotesi è stato dimostrato che in più del 85% dei tumori primari si può evidenziare attività telomerasica ed inoltre in molti tumori i telomeri sono più corti che nelle corrispondenti cellule normali (Bacchetti *et al.*, 1995; Kim, 1994; Hiyama *et al.*, 1995; Shay *et al.*, 1995). E' possibile che la telomerasi venga ad essere riattivata in cellule premaligne e la sua riattivazione consentirebbe a cellule, che già hanno perso il controllo della proliferazione a causa di mutazioni in oncogeni o geni oncosoppressori, di dividersi oltre il limite

imposto dall'accorciamento dei telomeri, permettendo così la crescita indefinita del tumore. In specie, come quella umana, a vita media elevata, la repressione della telomerasi in tessuti somatici, costituirebbe un meccanismo per ridurre la probabilità di insorgenza di cancro e la sua riattivazione sarebbe cruciale per la proliferazione tumorale.

La riattivazione della telomerasi potrebbe quindi essere un importante marcatore diagnostico, essendo il cambiamento biochimico che più comunemente si osserva nelle cellule di tutti i tipi di tumori ed in taluni casi può avere anche valore prognostico. E' però da tener presente che, sebbene la relazione tra attivazione della telomerasi e tumori sia molto stretta, non è ancora stato dimostrato un ruolo causale dell'attivazione dell'enzima nella tumorigenesi.

Inoltre, se la telomerasi è richiesta per la crescita della maggior parte delle cellule tumorali, come suggerito dalla selezione delle cellule telomerasi-positivo sia nell'immortalizzazione cellulare sia nei tumori, l'inibizione della telomerasi potrebbe essere alla base di una terapia antitumorale. In assenza di attività telomerasica, i telomeri dovrebbero accorciarsi finché, raggiunta una lunghezza critica, la cellula muore. Questa terapia dovrebbe essere altamente specifica, dato che nella maggior parte delle cellule normali l'enzima non è presente. A tal fine risultano indispensabili maggiori conoscenze relative ai meccanismi ed ai fattori che coinvolti nell'espressione e nella regolazione dell'attività dell'enzima.

Nel 1997, studi iniziali condotti su topi *knockout*, omozigoti difettivi per il gene dell'RNA telomerasi (Blasco *et al.*, 1997) avevano posto un autorevole dubbio sull'effettivo coinvolgimento dei telomeri nella cancerogenesi e nell'invecchiamento cellulare. Per almeno sei generazioni i topi *knockout* non avevano dimostrato in apparenza di risentire della mancata espressione della telomerasi.. Inoltre cellule deficienti per la telomerasi, erano in grado di diventare immortali una volta trasformate con oncogeni virali e pur finendo con il perdere le ripetizioni telomeriche e col generare anomalie ai cromosomi, non perdevano la

capacità di indurre tumori in topi nudi, indicando chiaramente che l'attività telomerasica non era indispensabile nella formazione dei tumori murini. Bisogna tuttavia tener presente che la specie di topi utilizzata è caratterizzata da telomeri particolarmente lunghi (da 10 a 80 kb) ed è possibile quindi che le cellule mTR- non esauriscano la loro scorta di sequenze telomeriche nell'arco delle divisioni richieste per formare il tumore. Non è escluso inoltre che nelle cellule tumorali mRT- si siano attivati meccanismi per il mantenimento di telomeri alternativi alla telomerasi. Successivamente però, dopo la sesta generazione, si sono cominciati ad osservare effetti consistenti nei topi *knockout*. L'accorciamento dei telomeri e l'associata instabilità cromosomica in questi topi sembrano essere legate ad una riduzione della durata media di vita e della capacità di resistere a condizioni di stress ed ad una maggiore incidenza di tumori maligni rispetto ai topi *wild-type*, (DePinho RA *et al.*, 1999).

Inoltre si è osservata in essi una diminuzione della fertilità sia nei maschi che nelle femmine ed un incremento della mortalità degli embrioni causata da mancata chiusura del tubo neurale. I topi che sopravvivono alla nascita, pur presentando difetti nei meccanismi emopoietici ed immunitari, sopravvivono fino all'età adulta anche se muoiono precocemente. Non si osservano invece modificazioni nello spettro dei caratteri patofisiologici che si manifestano con l'invecchiamento (Blasco *et al.*, 1999; Blasco *et al.*, 2000).

1.4 Proteine associate ai telomeri

Per costituire un telomero funzionante, non è sufficiente il DNA telomeric, ma è necessaria la sua interazione con proteine specifiche. L'insieme dei dati ottenuti da studi genetici, biochimici e di immunolocalizzazione da diverse specie, mostra che il DNA telomeric a doppia elica risulta complessato con una serie di

proteine diverse che lo impaccano in una struttura cromatinica specializzata (Brun *et al.*, 1997). Queste proteine a loro volta reclutano sul telomero altri fattori formando dei complessi multiproteici. Una panoramica delle principali proteine finora individuate come interagenti con sequenze telomeriche da diverse specie è riportata in tabella II.

Tra le diverse funzioni proposte per le proteine associate ai telomeri, la più importante sembra essere quella di modulare l'attività telomerasica. La lunghezza dei telomeri è mantenuta da un preciso bilanciamento fra processi antagonisti che portano: all'allungamento (telomerasi, sintesi del filamento ricco in C, ricombinazione), alla stabilizzazione (proteine leganti i telomeri, struttura cromatinica telomerica) ed all'accorciamento dei telomeri (replicazione incompleta, ricombinazione, attività enzimatiche degradanti i telomeri). Evidenze da vari organismi suggeriscono che diversi fattori concorrono nel regolare la lunghezza dei telomeri, quali proteine che si legano ai telomeri, proteine che portano alla formazione di *caps* cromosomici, enzimi coinvolti nella replicazione del DNA e la telomerasi.

Mediante saggi *in vitro* è stato dimostrato che la telomerasi non è in grado ad esempio di determinare l'estensione del tratto da sintetizzare per mantenere la lunghezza dei telomeri nell'ambito caratteristico della specie. Questo significa che *in vivo* debbono esserci dei fattori ausiliari (van Steensel *et al.*, 1997; Marcand *et al.*, 1997) che guidano la telomerasi facendo sì che essa aggiunga il corretto numero di ripetizioni.

In *S. cerevisiae*, la proteina candidata allo svolgimento di questo ruolo è RAP1 (Repressor/Activator Protein 1) (Shore & Nasmith, 1987). Si ritiene che essa sia un regolatore negativo, in quanto mutazioni che alterano il suo dominio di legame con il DNA o il suo livello di espressione, generano un allungamento incontrollato dei telomeri (Kyrion *et al.*, 1993)

Proteina	Organismo	Duplex	Filo singolo al 3'	Filo singolo al 5'
RAP1	<i>Saccharomyces</i>	+	+	-
Cdc 13p	<i>Saccharomyces</i>	-	+	-
TBF1	<i>Saccharomyces</i>	+	-	-
α - β	<i>Oxytricha</i>	-	+	-
β	<i>Oxytricha</i>	-	+	-
TBP	<i>Euplotes</i>	-	+	-
TGP	<i>Tetrahymena</i>	-	+	-
TEP	<i>Tetrahymena</i>	-	+	-
XTEF	<i>Xenopus</i>	-	+	-
TBP	<i>Mus</i>	-	+	-
Myo D	<i>Mus</i>	-	+	-
TRF1	<i>Mus, Homo</i>	+	-	-
TRF2	<i>Homo</i>	+	-	-
ST1	<i>Trypanosoma</i>	-	-	+
qTBP42	<i>Mus</i>	-	+	+
hnRNP A1/UP1	<i>Mus, Homo</i>	-	+	-
HnRNP K	<i>Homo</i>	-	-	+
ASF/SF2	<i>Homo</i>	-	-	+

TABELLA II: Proteine nucleari in grado di legare il DNA telomerico

RAP1 si lega lungo tutto il DNA telomerico a doppia elica (circa 300 bp) con una periodicità di circa 18 coppie di basi. La cellula sembra in grado di «contare» il numero di molecole di RAP1 legate al DNA telomerico ed in base a questo, inviare alla telomerasi il segnale di aggiungere o meno ulteriori ripetizioni (Marcand *et al.*, 1997). Si ritiene che il segnale sia di tipo strutturale: la presenza sui telomeri del numero corretto di molecole di RAP1 renderebbe l'estremità terminale del cromosoma inaccessibile alla telomerasi. In seguito al graduale accorciamento dei telomeri ad ogni divisione cellulare, le molecole di RAP1 legate scenderebbero al di sotto del valore soglia e questo renderebbe i telomeri nuovamente accessibili alla telomerasi.

Qualunque sia il reale meccanismo d'azione, RAP1 non sembra agire da sola in questo sistema. Sono note ormai diverse proteine in grado di interagire con la regione carbossi-terminale di Rap1, quali il fattore Rifp 1 e le proteine coinvolte nella formazione di cromatina trascrizionalmente silente: Sir3p, e Sir 4p (Moretti *et al.*, 1994); queste ultime stabilizzano il complesso repressivo formato dai lunghi schieramenti di Rap 1 associate ai telomeri che agiscono da elementi repressori nei confronti della telomerasi (Bourns *et al.*, 1998).

Insieme a Rap1 sono state descritte in lievito (Liu *et al.*, 1991) due altre proteine, TBF α e TBF β , leganti le sequenze telomeriche duplex d(TG₁₋₃); di particolare interesse sembra essere TBF α perché, oltre ad interagire specificamente con sequenze telomeriche di lievito, ha forte affinità anche per sequenze telomeriche non specie-specifiche che *in vitro* hanno mostrato di essere efficaci come *primers* per la telomerasi di lievito. Questo potrebbe suggerire un particolare ruolo di TBF α nel processo di allungamento dei telomeri *in vivo*.

Nell'uomo sono state identificate due proteine associate al DNA telomerico in forma di duplex: TRF1 (Chong *et al.*, 1995) e TRF2 (Broccoli *et al.*, 1997; van Steensel *et al.*, 1998). L'espressione di TRF1 (Telomere Repeat binding Factor) è ubiquitaria nei diversi tessuti ed un'attività simile è presente anche in altri

vertebrati, come il topo (Broccoli *et al.*, 1997). TRF1 è coinvolta nella regolazione della lunghezza dei telomeri nelle cellule umane agendo come repressore dell'attività telomerasica, in quanto la sua sovraespressione in linee cellulari immortalizzate esprimenti la telomerasi, provoca un graduale accorciamento dei telomeri (van Steensel e de Lange, 1997). La localizzazione nucleare e cromosomica *in vivo* di TRF1 è stata determinata attraverso immunofluorescenza; utilizzando proteine marcate con un epitopo ed espresse transientemente in cellule HeLa, TRF1 mostra legarsi all'estremità dei cromosomi sia in interfase che in metafase (Chong *et al.*, 1995)

Un indizio di come la TRF1 possa modulare l'attività della telomerasi venne fornito dall'identificazione della tankyrasi, una proteina che si lega alla TRF1. La tankirasi contiene un dominio omologo all'enzima poliADP polimerasi (PARP), che catalizza l'addizione di gruppi ADP ribosilici alle proteine. La tankirasi si è dimostrata in grado di ribosilare se stessa e la TRF1 *in vitro* e la ribosilazione della TRF1 inibisce il suo legame con il DNA telomerico (Smith *et al.*, 1998). In questo modo il distacco della TRF1 dai telomeri, ad opera della tankirasi, li renderebbe accessibili alla telomerasi. Comunque non si è ancora riusciti a dimostrare che la TRF1 *in vivo*, costituisca effettivamente il substrato della tankirasi.

TRF2 mostra *in vitro* la stessa specificità di legame di TRF1, tuttavia sembra svolgere un ruolo diverso. Infatti il legame di TRF2 sembra essere necessario per proteggere le estremità cromosomiche dalla fusione e per mantenere quindi l'individualità dei singoli cromosomi. Se viene impedito il legame di TRF2 endogena ai telomeri attraverso la sovraespressione di una proteina TRF2 mutata, si osserva la formazione di associazioni telomeriche, anomalie in cui due cromosomi sono uniti attraverso le loro estremità a formare un cromosoma dicentrico (van Steensel, 1998). La permanenza di lunghe ripetizioni di DNA telomerico al sito di fusione tra i due cromosomi indica che esse non sono di per

sé sufficienti per il mantenimento dell'integrità dei telomeri. Tuttavia sia i cromosomi fusi che non, perdono il filo singolo ricco in G sporgente al 3'. Questo suggerisce che la TRF2 impedisce la fusione blunt-end fra i cromosomi mantenendo la coda sporgente al 3'. Curiosamente, l'espressione della TRF2 mutata provoca un arresto della crescita con le caratteristiche tipiche della senescenza, pur non essendoci un accorciamento dei telomeri (van Steensel *et al.*, 1998), facendo pensare che la TRF2 possa far parte dei segnali che inducono la senescenza. Da osservare inoltre il possibile ruolo svolto da TRF2 nella formazione dei t-loops dei telomeri, ruolo che in parte renderebbe ragione di molte delle sue caratteristiche osservate *in vitro* ed *in vivo*. La TRF2 si legherebbe al DNA nel sito di giunzione dell'anello, dove il filamento ricco in G invade e scalza via dal duplex il filamento ricco in G presente in esso (Figura 6).

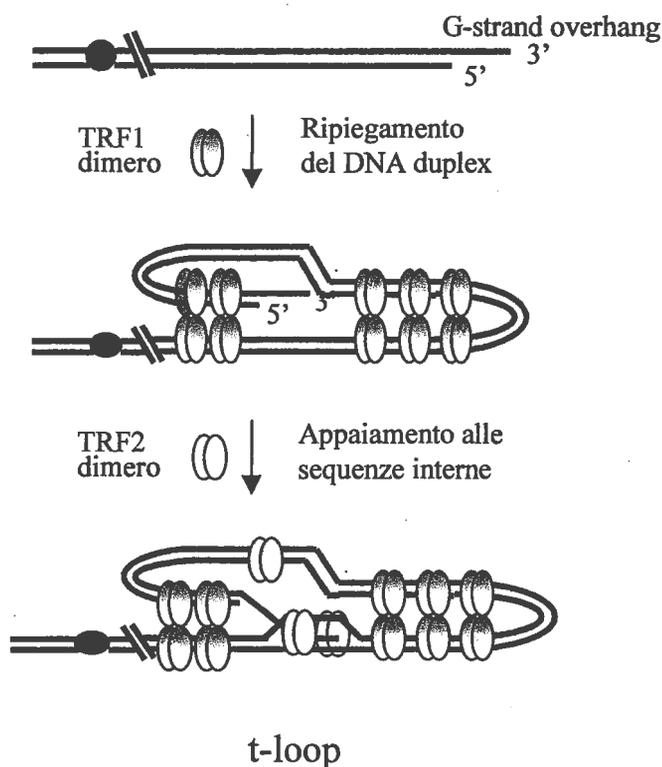


Figura 6: Formazione del t-loop

Telomeri artificiali sono stati incubati con la proteina TRF2 ed analizzati al microscopio elettronico per quanto riguarda la struttura a laccio. Inaspettatamente si è dimostrato che circa il 17-20% delle molecole forma dei larghi anelli, il cui sviluppo dipende dalla presenza del filamento ricco in G sporgente al 3' del telomero artificiale. L'omissione della TRF2 dalla reazione di legame *in vitro*, porta ad una sostanziale riduzione del numero di molecole evidenziabili sotto forma di anello, suggerendo che la proteina sia coinvolta specificamente nella formazione o nella stabilizzazione dei t-loops (Griffith *et al.*, 1999).

E' interessante notare che, sebbene non vi sia una forte omologia di sequenza nè tra le due proteine umane, né tra di esse e Rap1, i domini di legame al DNA sono tra loro strutturalmente omologhi ed in particolare sono omologhi al dominio di legame del proto-oncogene *c-Myb* (Smith *et al.*, 1997)

La ricerca di proteine in grado di legare il DNA telomerico non si limita a proteine che interagiscono con la sequenza telomerica in forma di duplex; una seconda classe di proteine riconosce il filo singolo ricco in G sporgente al 3'.

Tra le proteine in grado di legare specificamente la coda a singolo filamento, la più studiata è la proteina eterodimera α/β del ciliato *Oxytricha nova*. E' costituita da una subunità α di 41 kDa e da una subunità β di 56 kDa. Il complesso si lega tenacemente, ma non covalentemente alla coda a filo singolo d(T₄G₄T₄G₄) di ciascuna molecola di DNA macronucleare, formando una struttura detta telosoma (Gottschling *et al.*, 1986). Tale struttura, *in vitro*, si è dimostrata in grado di proteggere l'estremità del cromosoma dalla digestione con la nucleasi Bal 31 (Price & Cech, 1987) e *in vivo* di modulare l'accessibilità dei telomeri alla telomerasi e di proteggere le estremità cromosomiche vulnerabili alle nucleasi endogene (Hicke *et al.*, 1995).

Proteine in grado di legare il filamento ricco in G, ma incapaci di legare sia il duplex sia il filamento ricco in C, sono state caratterizzate anche in altri protozoi. Ad esempio in *Euplotes crassus* è stato individuato il fattore TBP (Wang *et al.*, 1992), un monomero di 51 kDa, con caratteristiche molto simili all'eterodimero trovato in *Oxytricha*, e con una sequenza amminoacidica identica per il 55% a quella della subunità α di *Oxytricha*.

Tra gli eucarioti superiori, in *Xenopus laevis* è descritto un fattore indicato come XTEF (Xenopus Telomere End Factor) in grado di interagire con la sequenza (TTAGGG)₂ quando si trova come filamento singolo (Cardenas *et al.*, 1993). Questo fattore presenta una attività molto spiccata in estratti da oociti e da ovari, mentre risulta difficilmente rilevabile nelle cellule somatiche.

Nell'ambito dei mammiferi di rilevante interesse è il presunto ruolo svolto dalla hnRNP A1 nella biogenesi dei telomeri. In letteratura viene descritta come probabile fattore di splicing specifico per la sequenza UAGGGU del sito di splicing al 3' degli m-RNA (Burd *et al.*, 1994), e come legante il DNA telomerico a filamento singolo a blocchi di guanine (Ishikawa *et al.*, 1993). Recentemente (LaBranche *et al.*, 1998) è stato dimostrato che cellule di topo deficienti per l'espressione del gene della hnRNPA1, mostrano telomeri significativamente più corti di quelli di cellule che esprimono livelli normali della hnRNPA1. Il ripristino nelle cellule dell'espressione di A1 e di un suo frammento, UP1, privo della regione carbossiterminale Gly-rich, manifestano un drammatico aumento della lunghezza dei telomeri, fino a divenire comparabile a quella dei telomeri in cellule normali. E' probabile che A1, quando si trova come frammento UP1, interagisca con il DNA telomerico a blocchi di guanine e sia in grado di stimolare e regolare l'attività dell'enzima. Sebbene alcune osservazioni dimostrino la presenza di piccole quantità di UP1 nelle cellule (Riva *et al.*, 1983; Valentini *et al.*, 1985; Pandolfo *et al.*, 1985), l'esistenza di UP1 *in vivo* rimane ancora controversa, data la possibilità che la proteolisi di A1 possa avvenire durante la fase di estrazione o

purificazione della proteina. Comunque è possibile che il processamento di A1 in UP1 rappresenti una fase critica nel meccanismo di modulazione dell'attività telomerasica.

Gran parte dell'interesse e delle ricerche volte ad identificare proteine telomeriche interagenti non con il canonico duplex Watson-Crick, ma con sequenze telomeriche a filamento singolo, si sono incentrati sul filamento a blocchi di guanine, essenzialmente per due motivi: sia negli eucarioti unicellulari sia nei vertebrati superiori, l'estremità telomerica al 3', ovvero quella a guanine, sporge per un certo numero di ripetizioni come coda singolo filamento da entrambe le terminazioni dei cromosomi (McElligot *et al.*, 1997); sequenze a filamento singolo caratterizzate dalla ripetizione di due o più blocchi di guanine, quali quelle telomeriche, sono in grado di assumere *in vitro* quadruplex intra-ed intermolecolari (Williamson *et al.*, 1989) che potrebbero costituire dei «*targets*» strutturali per il riconoscimento da parte di fattori proteici specifici (Williamson, 1994)

Solo recentemente si è iniziato a cercare proteine in grado di legare il filamento complementare a blocchi di citosine, in particolare dopo che studi condotti su *Saccharomyces cerevisiae* (Wellinger *et al.*, 1996) e su cellule umane (Makarov *et al.*, 1997) suggeriscono che la formazione della coda telomerica al 3' si formi per parziale accorciamento del filamento a citosine, mediato da componenti proteiche interagenti con esso.

Qualche anno fa è stata osservata in *Trypanosoma brucei* una proteina di 39 kDa detta ST-1 in grado di legare con alta affinità i motivi a citosine a filamento singolo (Eid *et al.*, 1995), ma tuttora non è stata ancora caratterizzata a livello molecolare.

Successivamente, da nuclei di epatociti di ratto, è stata isolata una proteina siglata come qTBP42 (Sarig *et al.*, 1997) la quale sembra legare fortemente entrambe le sequenze telomeriche d(TTAGGG)_n e d(CCCTAA)_n quando presenti

sotto forma di filamento singolo, ma non quando si appaiano nel duplex. Il successivo sequenziamento (Sarig *et al.*, 1997), anche se solo parziale, ha permesso di identificarla come l'omologo di ratto di un fattore di trascrizione da cellule muscolari di topo, il CA_rG (Cytosine Adenine rich Guanine) Binding Factor o CBF-A (Kamada *et al.*, 1992), appartenente come sequenza ad una funzionalmente eterogenea classe di proteine note come hnRNP di tipo A/B.

Molto recentemente è stato dimostrato che anche la proteina hnRNP K ed il fattore di *splicing* ASF/SF2 umani sono capaci di riconoscere il filo singolo ricco di citosine (Lacroix *et al.*, 2000).

L'hnRNP K è un'abbondante proteina, descritta come legare tenacemente DNA a filo singolo a sequenza poli-C (Matunis *et al.*, 1992), nonché come fattore di trascrizione interagente con la sequenza regolativa «TC-element» localizzata a monte del gene *c-myc* (Garcia-Basset *et al.*, 1999). ASF/SF2 è membro di una famiglia molto conservata di fattori di *splicing* noti come proteine SR (Krainer *et al.*, 1990), essenziali per lo *splicing in vitro* ed interagenti a livello del sito di *splicing* al 3' del trascritto primario. In particolare ASF/SF2 sembra coinvolta in alcuni aspetti della modulazione dello *splicing* alternativo secondo meccanismi di tipo concentrazione-dipendente (Mayede *et al.*, 1992), antagonisti a quelli della hnRNPA1.

Da quanto detto, è evidente che hanno fondamentale importanza i fattori proteici in grado di interagire con i telomeri e di regolarne la lunghezza, la struttura e di conseguenza la loro funzione. Però, per quanto riguarda proteine descritte come interagenti con il motivo telomerico ricco in C, attualmente non si hanno forti evidenze sperimentali di un loro effettivo ruolo funzionale *in vivo* nell'architettura o nella gestione dei telomeri.

Un possibile ruolo potrebbe essere quello di sequestrare attraverso interazione DNA-proteina il filamento ricco in C al duplex telomerico, rendendo così

accessibile il filamento a guanine, ad esso complementare, alla telomerasi per la sintesi di nuovi repeats telomerici (Lacroix *et al.*, 2000).

Un altro possibile ruolo, come già accennato, può essere proposto sulla base dei risultati ottenuti dagli studi condotti sul lievito *S. cerevisiae* (Wellinger *et al.*, 1996). Questi esperimenti suggeriscono un nuovo aspetto della gestione dei telomeri: la degradazione modulata dal ciclo cellulare del filamento a blocchi di citosine del duplex telomerico a formare la coda a filo singolo a blocchi di guanine, che caratterizza l'estremità di tutti i cromosomi eucariotici (Wright *et al.*, 1997).

Se però tali fattori proteici descritti *in vitro* come legare il filamento telomerico pirimidinico, svolgano effettivamente *in vivo* una qualche funzione correlata ai telomeri, rimane ancora un quesito aperto.

2.1 Marcatura di sonde di DNA

Circa 500 ng di oligonucleotide sintetico vengono marcati per aggiunta enzimatica (tramite T4 polinucleotide chinasi, Promega) di fosfato terminale radioattivo, usando 50 μCi di $\gamma\text{-}^{32}\text{P}\text{-ATP}$ in tampone 40 mM Tris/HCl pH=7.5, 10 mM MgCl_2 , 5 mM DTT e lasciando procedere la reazione per 1 ora a 37°C.

Le sonde così preparate vengono separate dall'ATP non reagito mediante cromatografia per gel filtrazione utilizzando la resina G-50 oppure, in alternativa, sono state usate «spin columns» S-300 HR (Pharmacia) disponibili commercialmente. L'attività specifica delle sonde ottenute è in media di 5×10^8 cpm/ μg di DNA.

2.2 Colture cellulari

Le cellule sono state fatte crescere in ambiente a 37°C e 5% CO_2 nel terreno di cultura opportuno (RPMI 1640 o DMEM) addizionato con il 10% (v/v) di siero fetale bovino totale inattivato, 100 U/ μl di penicillina e 100 $\mu\text{g}/\text{ml}$ di streptomicina.

Le cellule BM2 di pollo (Moscovici *et al* 1982) sono state invece cresciute in terreno HEPES-RPMI addizionato di siero di pollo al 5% v/v, siero fetale di vitello al 10%, triptosio fosfato broth al 10%, L-glutammina 4 mM, 100 u/ml di penicillina e 100 µg/ml di streptomicina.

2.3 Preparazione degli estratti proteici

a) Estratti di proteine nucleari sono stati ottenuti da diverse linee cellulari umane, di criceto, di ratto e di pollo e da eritrociti di pollo, seguendo un protocollo modificato della procedura di Dignam (1983)

Le cellule, raccolte per centrifugazione, vengono risospese in 5 volumi di PBS (tampono fosfato 10 mM pH=7.4, NaCl 0.15 M, KCl 0.02 M) e centrifugate a 3000 r.p.m. per 10 minuti. Il pellet viene risospeso, prima in 5 volumi di tampone ipoosmotico (10 mM HEPES pH=7.9, MgCl₂ 1.5 mM, KCl 10 mM, PMSF 0.2 mM, DTT 0.5 mM) e successivamente, dopo centrifugazione, in 3 volumi dello stesso tampone, nel quale le cellule vengono incubate per 10' in ghiaccio. Si procede quindi all'omogenizzazione della sospensione cellulare in un omogenizzatore Dounce B ed alla sua centrifugazione per 15' a 4000 r.p.m. allo scopo di sedimentare i nuclei. Ai nuclei, risospesi in tampone a bassa forza ionica (2mM HEPES pH=7.9, 25% glicerolo, MgCl₂ 1.5 mM, KCl 0.02 M, EDTA 0.2M, PMSF 0.2 mM, DTT 0.5mM) si aggiunge goccia a goccia un ugual volume di tampone ad alta forza ionica (20 mM HEPES, MgCl₂ 1.5 mM, KCl 1.2 M, EDTA 0.2 mM, PMSF 0.2 mM, DTT 0.5 mM) che, agitando gentilmente per circa 30 minuti, favorirà il rilascio delle proteine solubili.

Il surnatante, ottenuto dopo centrifugazione a 14500 r.p.m., viene infine dializzato contro tampone HEPES 20 mM pH=7.9, 20% glicerolo, KCl 100 mM, EDTA 0.2

mM, PMSF 0.2 mM, DTT 0.5 mM. e l'estratto, dopo determinazione della sua concentrazione proteica, suddiviso in aliquote e conservato a -80°C .

Estratti di proteine nucleari sono stati anche ottenuti da tessuti di fegato di ratto, di maiale e di vitello; in questo caso i tessuti sono stati tagliati in piccoli pezzi, omogeneizzati con pistola Turrax in tampone NPBi (Tris 10 mM pH=7.4, MgCl_2 2 mM, NaCl 10 mM, Antipain 2 $\mu\text{g/ml}$, Aprotin 2 $\mu\text{g/ml}$, Leupeptin 2 $\mu\text{g/ml}$, EGTA 50 mM, PMSF 0.5 mM) e filtrati su garza chirurgica. L'omogenato viene poi centrifugato a 14000 r.p.m. per 10' ed il pellet, risospeso in 5 volumi di PBS, processato come già descritto per gli estratti nucleari da cellule.

b) Estratti proteici totali: le cellule (eritrociti di pollo da sangue intero e cellule umane K562) sono state raccolte mediante centrifugazione a 400 g per 15 minuti a 5°C , lavate 2 volte in tampone PBS freddo e, dopo centrifugazione ed eliminazione della soluzione di lavaggio, conservate a -80°C in aliquote.

Per l'estrazione delle proteine totali, ogni pellet è stato rapidamente risospeso in 4 volumi di tampone di lisi (Hepes 10 mM pH=7.9, NaCl 400 mM, EGTA 0.1 mM, DTT 0.5 mM, PMSF 0.5 mM, 5% v/v glicerolo). Dopo la lisi, le cellule sono state centrifugate a 8000 g per 20' ed il surnatante, contenente l'estratto proteico, è stato recuperato e portato ad una concentrazione finale di glicerolo del 20% e conservati a -20°C per alcuni giorni.

2.4 Cromatografia di affinità

2.4.1 Preparazione della resina di affinità

La matrice di affinità è stata preparata usando un supporto commerciale (Affi-prep 10, Bio-Rad) derivatizzato con avidina purificata. A tale scopo la matrice è stata lavata per 3 volte con 10 volumi di tampone acetato 10 mM pH=4.5, e quindi incubata per 16 ore a 4°C in agitazione, con una soluzione di avidina 1 mg/ml (Fluka) in tampone MOPS 100 mM pH=7.0.

La resina è stata quindi bloccata a temperatura ambiente per un'ora con 1 volume di etanolamina-HCl 1M pH=8.0 e lavata prima con 10 volumi di NaCl 1.5 M e successivamente con PBS dove viene conservata a 4°C per la successiva funzionalizzazione con l'oligonucleotide HTC6 biotinilato al 3'.

Per la funzionalizzazione, la resina è stata condizionata in tampone Tris-HCl 50 mM pH=8 ed EDTA 1 mM, ed incubata per alcune ore in tampone Tris-HCl 50 mM pH= 8, NaCl 1 M, con la quantità opportuna di oligonucleotide calcolata in funzione della capacità della resina. Tipicamente 100 µl di matrice sono in grado di legare 20µg di oligonucleotide HTC6.

La resina così funzionalizzata, è stata lavata in tampone Tris-HCl 50 mM pH=8, EDTA 1 mM e conservata a 4°C nello stesso tampone fino al momento dell'utilizzo.

2.4.2 Purificazione dell'estratto proteico

Per la purificazione delle proteine dagli estratti proteici totali, è stata effettuata una cromatografia di affinità *in batch*. Gli estratti (10-20 mg/ml) sono stati diluiti con tampone Tris/HCl 50 mM pH=8, EDTA 5 mM, 20% v/v glicerolo, DTT e PMSF 0.5 mM, fino ad ottenere una concentrazione finale di NaCl di 200 mM.

In genere 40 ml di estratto proteico diluito sono stati incubati per una notte in agitazione a 4°C con 400 µl della resina di affinità funzionalizzata.

Al termine dell'incubazione il surnatante viene rimosso dalla resina e viene saggiata in esso, tramite saggi EMSA, la presenza della proteina d'interesse, valutando così l'efficienza di *binding* della proteina stessa alla resina. A sua volta la matrice è stata brevemente lavata 3 volte con 1-2 ml dello stesso tampone di incubazione e quindi eluita con tampone Hepes 10 mM pH=7.9, EGTA 0.1 mM, DTT e PMSF 0.5 mM, NaCl 1 M, per un'ora a temperatura ambiente.

Gli eluiti sono stati quindi analizzati in elettroforesi su gel denaturante in SDS, colorato con BlueCoomassie per evidenziare le proteine in esso contenute, ed in seguito saggiati in EMSA per evidenziare l'attività di legame della proteina cercata.

2.5 Cromatografia a fase inversa

Per separare le proteine eluite da cromatografia di affinità, è stata effettuata una cromatografia HPLC (High Performance Liquid Chromatography) a fase inversa. L'apparecchiatura utilizzata è un cromatografo Jasco 8880-02 a doppia pompa fornito di rilevatore spettrofotometrico con lampada allo zinco. È stata utilizzata una colonna cromatografica Jupiter (150 mm x 4.6 mm impaccata con una matrice in fase inversa C18) della Phenomenex.

Le proteine sono state eluite a 0.5 ml/min. con un gradiente lineare di acetonitrile/acqua (da 0% a 95% v/v di acetonitrile in 40 minuti) più 0.1% v/v di acido trifluoroacetico.

L'eluizione delle proteine dalla colonna è stata monitorata mediante lettura spettrofotometrica alla lunghezza d'onda di 229 nm e le frazioni corrispondenti ai picchi, raccolte manualmente, liofilizzate e risospese nel tampone opportuno.

2.6 Tecniche elettroforetiche

L'apparecchiatura elettroforetica impiegata è costituita da una cella CBS Scientific Co., collegata ad un termocriostato Haake F3 ed alimentata da un Power Supply PS 500X della Hofer.

Sono stati usati gel a diversa concentrazione di poliacrilammide, ottenuti dalla polimerizzazione dei due monomeri, acrilammide e N-N'-metilenbisacrilammide, nel rapporto 29:1, dopo l'aggiunta di persolfato di ammonio e di TEMED come catalizzatore. Lo spessore dei gel usati variava da 0.5 a 1 mm, sia di tipo nativo che di tipo denaturante.

2.6.1 Elettroforesi denaturante di proteine

Nell'analisi elettroforetica le proteine sono state separate su di un gel di poliacrilammide denaturante in SDS con tampone Tris/Tricina o Tris /Glicina.

Nell'elettroforesi in tampone Tris/Tricina, è stato utilizzato un gel di separazione (*running*) al 8% (w/v) di acrilammide, 0.2% (w/v) di Bis (N-N'-metilenbisacrilammide) in tampone Tris HCl 0.6 M, tricina 0.09 M pH=8.5, SDS 0.1%. La corsa elettroforetica, effettuata in tampone 0.1% SDS, tricina 0.025M, portato a pH=8.1 con Tris solido, è stata condotta a 165 Volt per 4-5 ore.

Nell'elettroforesi Tri-glicina, il "gel di compattazione" (*stacking gel*) è stato un gel di poliacrilammide [acrilammide : N,N'-metilen-bis-acrilammide = 30 : 0.8 (w/w)] al 3.5% (w/v) in acrilammide, Tris-HCl 125 mM pH 6.8, SDS 0.1% (w/v). Il "gel di separazione" (*running gel*) è costituito da poliacrilammide (a varie percentuali), Tris-HCl 375 mM pH 8.8, SDS 0.1% (p/v).

I campioni proteici sono stati incubati con la soluzione di caricamento Laemmli e bolliti per 3 minuti. La corsa elettroforetica è stata condotta in *running buffer* per

SDS-PAGE (Tris 25 mM, Glicina 192 mM SDS 0.1% (p/v)), inizialmente ad un amperaggio di circa 0.5 mA per cm² di gel, ed in seguito, durante la fase di separazione, ad un amperaggio di circa 1 mA per cm² di gel.

Dopo la corsa, il gel è stato colorato e contemporaneamente fissato immergendolo in una soluzione allo 0.05% (w/v) di Blue Coomassie R-250 (Serva) sciolto in metanolo, acido acetico e acqua nelle proporzioni 5:4:1 in volume. Successivamente, è stata eseguita la decolorazione con una soluzione di etanolo ed acqua in proporzioni 1:3 in volume.

2.6.2 Recupero delle proteine da gel di poliacrilammide

Al termine della corsa elettroforetica effettuata su gel denaturante SDS-Tris/Tricina o Tris/Glicina, le bande proteiche, visualizzate dopo colorazione con Blue Coomassie, sono state estratte dal gel per elettroeluizione o precipitazione con acetone.

Nel primo caso, dopo separazione elettroforetica, le bande sono state tagliate dal gel, triturate e poste in tubi di dialisi con 2 ml di tampone (0.1% SDS, 0.025 M tricina pH=8.1 con Tris solido) ed elettroeluite a 6 mA, 20 Volt per 5 ore.

I campioni sono quindi stati dializzati contro acqua per 48 ore a 5°C, liofilizzati e risospesi in tampone per i saggi EMSA (Tris-HCl 50 mM pH=8, KCl 50 mM, EDTA 5 mM, glicerolo 25%).

L'estrazione da gel attraverso precipitazione con acetone è stata effettuata tagliando le bande colorate in Blue Coomassie da gel denaturante (SDS PAGE, 9% acrilammide), triturandole e ed incubandole in agitazione tutta la notte in 0.5 ml di tampone Tris-HCl pH= 7.9, 0.1% SDS, 0.1 mg/ml BSA, DTT e PMSF 1 mM, EDTA 0.2 mM, glicerolo 2.5%.

Al termine dell'incubazione, le proteine diffuse dal gel sono state precipitate in 4 volumi di acetone a -20°C per alcune ore, lavate con metanolo freddo e liofilizzate. Ciascun pellet è stato sciolto in $10\ \mu\text{l}$ di urea 8M e successivamente diluito con $125\ \mu\text{l}$ di Tris-HCl $50\ \text{mM}$ $\text{pH}=7.6$, KCl $50\ \text{mM}$, EDTA $5\ \text{mM}$. Dopo 18 ore di rinaturazione a 4°C , le frazioni proteiche sono state utilizzate per saggi EMSA.

2.6.3 Saggi di ritardo della mobilità elettroforetica (EMSA)

Sono stati condotti diversi esperimenti di EMSA (Electrophoretic Mobility Shift Assay) al fine di evidenziare l'eventuale formazione di complessi oligonucleotide-proteina. Questa tecnica permette infatti di evidenziare la formazione di tali complessi in quanto essi daranno origine ad una o più bande a mobilità ridotta rispetto a quella dell'oligonucleotide libero in soluzione.

In questo tipo di elettroforesi, condotta in condizioni native, è stato impiegato un gel di poliacrilammide al 10% o 8% , in tampone Tris-HCl $50\ \text{mM}$ $\text{pH}=8$, KCl $50\ \text{mM}$ e EDTA 0.1M oppure, un gel di poliacrilamide all' 8% in tampone TBE (Tris-borato $0.045\ \text{M}$ $\text{pH}=8$, EDTA $0.001\ \text{M}$).

I campioni sono stati preparati preincubando per 1 ora a temperatura ambiente, $4\ \mu\text{g}$ di estratto proteico totale oppure $0.01\ \mu\text{g}$ delle proteine purificate, con un opportuno eccesso di competitore freddo (DNA o RNA), in un volume totale di $20\ \mu\text{l}$ di miscela resa Tris-HCl $50\ \text{mM}$ $\text{pH}=8$, KCl $50\ \text{mM}$, $0.1\ \text{mM}$ EDTA, glicerolo 25% . Al termine, alla miscela di incubazione sono stati aggiunti $0.2\ \text{pmol}$ di oligonucleotide sonda marcato ed il tutto è stato incubato overnight a 4°C oppure per 2 ore a temperatura ambiente.

La corsa elettroforetica, termostata a 15°C , è stata condotta con lo stesso tampone di incubazione oppure con tampone TBE a $10\ \text{V/cm}$, per un paio d'ore, preceduta da una precorsa di 30 minuti a 15°C . Al termine, il gel è stato essiccato in

Dryer su carta da filtro 3MM della Whatman e posto in esposizione su lastra autoradiografica per periodi di tempo variabili da un'ora ad alcuni giorni.

2.6.4 *U.V. crosslinking*

I campioni sono stati preparati in modo analogo a quelli per gli EMSA. Dopo l'incubazione sono stati sottoposti ad irraggiamento con lampada UV da 50 Watt per 60-90 minuti, in ghiaccio ad alla distanza di circa 10 cm dalla fonte di luce. Prima del caricamento, vengono aggiunti 3 μ l di Sample buffer (Tris-HCl 125 mM pH=6.8, SDS 2%, β -mercaptoetanololo 2%, glicerolo 20%, Blu di Bromofenolo 0.001%) ed i campioni vengono riscaldati a 100°C per 90".

La corsa viene effettuata in SDS-PAGE Tris/Glicina su gel discontinuo, a 20 V/cm per 2 ore.

In parallelo, sullo stesso gel, vengono caricati markers proteici noti, allo scopo di ottenere una stima approssimata dei pesi molecolari dei fattori nucleari a cui le bande si riferiscono. Al termine dell'elettroforesi, il gel viene colorato in Blue Coomassie e successivamente essiccato in dryer ed autoradiografato.

2.6.5 *U.V. crosslinking in situ*

Dopo l'esecuzione di un saggio di ritardo della mobilità elettroforetica come descritto sopra, il gel umido è stato autoradiografato a 5°C e successivamente irradiato con una lampada UV (300 nm, 50 Watt) per 6 minuti in modo da consentire la formazione di legami covalenti tra le proteine e gli oligonucleotidi ad esse legati.

Le bande ritardate, visualizzate dopo lo sviluppo della lastra autoradiografica, sono state tagliate dal gel e le proteine in esse contenute denaturate mediante

bollitura per 5 minuti in tampone Tris-HCl 125 mM pH=6.8, SDS 2%, β -mercaptoetanolo 2%.

Markers proteici di peso molecolare noto, separati elettroforeticamente, sono stati tagliati dal gel e denaturati in parallelo con i campioni.

In seguito, le porzioni di gel tagliate contenenti i markers di peso molecolare ed i complessi irradiati, sono state poste nei pozzetti di un gel di poliacrilammide SDS Tris/Tricina all'8%: La corsa elettroforetica è stata effettuata nel tampone di corsa per 2 ore a 20 V/cm. Al termine, il gel è stato colorato in Blue Coomassie e poi asciugato ed autoradiografato.

2.6.6 *Super-shift*

In ciascun saggio di super-shift, 10 μ g dell'estratto totale oppure 0.2 μ g delle proteine eluite dalla cromatografia di affinità, sono state preincubate un'ora in 20 μ l totali di tampone A (Tris-HCl 50 mM pH=8, KCl 50 mM, EDTA 5 mM) in presenza di un eccesso di 250 volte di DNA di *E.coli* denaturato, con differenti diluizioni di anticorpo. L'anticorpo policlonale da siero di coniglio diretto contro la proteina ssDBF è stato gentilmente donato dal dottor Smidt dell'Università di Groninga.

Al termine dell'incubazione sono stati aggiunti 0.2 pmol dell'oligonucleotide marcato HTC4 e dopo un'ora a temperatura ambiente i campioni sono stati caricati in un gel di poliacrilammide non denaturante al 6% in tampone A.

2.6.7 *Western blot*

Questa tecnica è stata utilizzata per trasferire le proteine da un tracciato elettroforetico su gel di poliacrilammide in SDS Tris/Glicina o Tris/Tricina, a una

membrana di Immobilon. Il gel viene condizionato per circa 30 minuti in 0.75 M Tris base, 0.1% (w/v) SDS. Contemporaneamente, la membrana viene equilibrata con la soluzione di Tris/SDS diluita 1:10. Con la stessa vengono imbibiti i fogli di carta 3MM a contatto con la membrana. La carta a contatto con il gel viene bagnata con la soluzione concentrata, mentre le due carte ancora sovrapposte, con quella diluita 1:10. Il trasferimento viene fatto mediante un apparato Trans-blot Semi-Dry Transfer Cell BioRad, ad un voltaggio di 15 V per 30'. Dopo il trasferimento, la membrana viene colorata con 0.2% (w/v) Ponceau S in 3% (v/v) acido tricloroacetico, 3% (v/v) acido solfosalicilico.

Per l'immunoriconoscimento, le membrane vengono saturate in Tris-HCl 0.05 M, NaCl 0.5 M, latte scremato in polvere 5% (w/v), Tween 20 0.05% (v/v), BSA 0.2% (w/v) e poi incubate con l'anticorpo policlonale anti-ssDBF (diluito circa 1:1000), per un'ora a temperatura ambiente. Vengono fatti 3 lavaggi con la soluzione di saturazione e poi si metta ad incubare con un antisiero anti-IgG di coniglio coniugato con la perossidasi (Amersham Pharmacia Biotech) diluito 1:1000, per un'altra ora a temperatura ambiente. Altri 3 lavaggi nelle stesse condizioni precedono lo sviluppo della reazione di chemiluminescenza secondo il protocollo fornito dalla ditta fornitrice del kit ECL (Amersham Pharmacia Biotech)

2.6.8 *South-western*

Aliquote di 5-10 μ g di proteine totali o di 50-250 ng di proteine purificate, sono state separate elettroforeticamente su gel SDS Tris/Glicina o Tris/Tricina all'8% per 12 ore a 10 V/cm.

Al termine le proteine sono state trasferite su filtro di nitrocellulosa per elettroblotting come già descritto sopra. La membrana è stata quindi immersa in una soluzione 6M di guanidinio idrocloruro e le proteine rinaturate a 4°C attraverso

incubazioni successive della membrana in soluzioni a concentrazioni decrescenti di guanidinio idrocloruro (6, 4, 3, 2, 1, 0.5, 0.25 e 0 M). Il filtro è stato saturato per 1 ora con una soluzione al 5% latte in polvere e quindi incubato overnight a 4° nel tampone di *binding* (Tris-HCl 50 mM pH=8, KCl 50 mM, EDTA 5 mM, DTT 1mM), con circa 20 pmoli di oligonucleotide sonda marcato ed un eccesso di 100 volte (w/w) di DNA di *E.coli* denaturato. Dopo tre lavaggi nello stesso tampone, sono state visualizzate le bande di proteine riconosciute dalla sonda attraverso autoradiografia su lastra fotografica.

2.7 Spettrometria di massa «ion spray»

Per la determinazione dei pesi molecolari delle proteine si è fatto ricorso alla tecnica «ion-spray». Le analisi sono state eseguite utilizzando lo spettrometro di massa API Perkin Elmer Sciex presso il nostro dipartimento. Per la calibrazione dello strumento è stata utilizzata una soluzione standard di PPG (polipropilenglicole). Per la calibrazione in base agli ioni multicaricati, è stata invece usata una soluzione di mioglobina (16950,5 Da) in acqua/acetonitrile 1:1 e 0.1% (v/v) acido formico. I campioni purificati da cromatografia, sono stati portati a secco, ridisciolti in acqua/acetonitrile 80:20, 0.1% (v/v) acido formico per essere analizzati.

2.8 Spettrometria di massa MALDI-TOFF

2.8.1 Digestione su gel e purificazione dei peptidi

Le bande elettroforetiche delle proteine tagliate dal gel SDS-PAGE, sono state triturate e lavate due volte con acetonitrile e ammonio bicarbonato 0.1M. Sono

state quindi ridotte su gel con 10 mM DTT in ammonio bicarbonato 0.1 M (45 minuti a 55° C) e S-alchilate con 55 mM iodoacetammide in 0.1 M ammonio bicarbonato (30 minuti a 25° C e al buio). Dopo numerosi lavaggi con ammonio bicarbonato 0.1 M, i frammenti di gel sono stati essiccati e le proteine digerite in una soluzione di 12.5 ng/ μ l di tripsina in ammonio bicarbonato 0.1 M. Dopo 2 ore di incubazione a 5°C, al campione viene aggiunto un uguale volume di soluzione di digestione ed incubato *overnight* a 37°C. Quindi, i frammenti di gel rimossi dalla miscela di digestione vengono sonicati per 2 volte in 50% acetonitrile, 0.1M ammonio bicarbonato ed i campioni vengono quindi liofilizzati.

I peptidi ottenuti dalla digestione sono stati separati per cromatografia in fase inversa-HPLC su una colonna Vydac C18 214TP52 (250x2.1mm), 5 μ m, 300 Å *pore size* (The Separation Group, USA), usando un gradiente lineare dal 5% al 60% di acetonitrile in 0.1 % TFA in 65 minuti, ad un flusso di 0.2 ml/min. Le singole frazioni sono state raccolte manualmente e portate a secco in una centrifuga Speedvac (Savant).

2.8.2 Spettrometria di massa

Gli spettri di massa sono stati ottenuti usando uno spettrometro Voyager DE MALDI-TOF (Perkin Elmer-Persaptive Biosystem, Norwalk, USA); una miscela costituita da acido α -ciano-4-idrossi-cinnamico e insulina bovina, è stata posta sul supporto porta campioni dello strumento e quindi portata a secco. La calibrazione è stata effettuata utilizzando ioni molecolari da insulina bovina (5,734.59 m/z) e di acido α -ciano-4-idrossi-cinnamico, come standard interni. I dati sono stati analizzati usando un software fornito dal produttore dello strumento e sono riportati come massa media degli ioni molecolari. Gli spettri MALDI MSMS sono stati registrati usando uno strumento Autospec OA-TOF (Micromass, Manchester, UK). In questo

caso l'acido 2,5-diidrossi benzoico è stato usato come supporto. Gli ioni molecolari vengono fatti collidere con l'argon in una cella di collisione mantenuta ad 800eV. Le masse dei frammenti sono state misurate usando un analizzatore TOF montato ortogonalmente e i dati ottenuti sono stati analizzati usando il software OPUS fornito assieme allo strumento.

2.8.3 Identificazione delle proteine

Sono stati utilizzati due software, PROWL e PeptideSearch, per identificare le proteine a partire da banche dati di sequenze proteiche non ridondanti curate ed aggiornate dall'European Molecular Biology Laboratory (EMBL) e dal National Institute of Health (NIH). I parametri utilizzati per il confronto dei valori di massa ottenuti sono stati la massa molecolare apparente delle proteine, la specie di appartenenza, eventuali errori di massa e possibilità di tagli incompleti dei frammenti, presenza di modificazioni post-traduzionali ed isoforme. Un certo numero di possibili candidati che hanno mostrato uno *score* molto elevato nell'analisi di confronto tra i peptidi sono stati confrontati in base al loro peso molecolare usando i dati sperimentali dei valori di massa molecolare delle proteine allo studio calcolati in SDS-PAGE. Questo parametro è stato usato considerando un filtro ad alta tolleranza ($\Delta MW = \pm 15\% Mw$) al fine di escludere falsi positivi dai potenziali candidati.

2.9 Preparazione del DNA plasmidico

2.9.1 Preparazioni in piccola scala

I batteri trasformati sono stati lisati con il metodo della lisi alcalina. 1.5-2 ml di coltura batterica sono stati centrifugati per 10 minuti a 8000 rpm. Il pellet è stato risospeso in 100 µl di soluzione P1 (Tris-HCl 50 mM pH 8.0, EDTA 10 mM, RNasi A 100 µg/ml) e successivamente incubato in ghiaccio per 5 minuti con 200 µl di soluzione P2 (NaOH 200 mM, SDS 1%) Sono stati aggiunti 150 µl di soluzione P3 (CH₃COOK 2.55 M pH 4.8), incubati nuovamente in ghiaccio per 5 minutie centrifugati per 10 minuti a 1300 rpm.

Il DNA è stato estratto con 0.5 volumi di fenolo equilibrato in Tris-HCl 50 mM più 0.5 vol. di cloroformio-alcool isoamidico (24:1). E' stato precipitato con 0.1 vol. di CH₃COONa 3 M e 2.5 vol. di etanolo assoluto a meno -80°C per 20'. Si è proceduto quindi alla centrifugazione a 4 °C per 20 minuti (centrifuga B-Braun-3K 30, rot. 12154), il precipitato è stato lavato con etanolo al 70% per rimuovere i sali, asciugato ed infine risospeso in acqua o tampone TE.

2.9.2 Preparazione su larga scala

I batteri contenenti i plasmidi sono stati lisati con il metodo della lisi alcalina come precedentemente descritto. Il DNA plasmidico è stato successivamente purificato con colonnine commerciali Nucleobond AX (Macherey-Nagel). Queste hanno la capacità di separare il DNA plasmidico dal DNA batterico secondo il principio della cromatografia d'affinità. La sospensione è stata caricata sulla colonnina e la frazione plasmidica recuperata mediante eluizione con gradiente di pH.

2.10 Clonaggio del gene di ssDBF

Il gene per ssDBF è stato subclonato all'interno del vettore di espressione batterica pT7.7 a partire dal vettore pGem7zf(+), gentilmente messo a disposizione dal Dott. Smidt dell'Università di Groninga.

Il gene ssDBF è stato amplificato dal vettore pGem, attraverso la tecnica della PCR. Gli oligonucleotidi utilizzati come *primers* nella reazione di amplificazione sono:

NdeFw 92: 5'-CATATGGCTGAGAGCTCCGGGGAGCA-3'

3'-ssDBF: 5'-TAAGCTTGATCAATATGGCTTGTAGTTATTCTG-3'

Le basi sottolineate rappresentano i siti di restrizione dell'enzima Nde I e Hind III usati per la creazione delle estremità coesive necessarie nelle operazioni di clonaggio. Per ogni campione da amplificare si è preparata una miscela di reazione di 100 µl contenente:

- 10 µl di tampone 10X (KCl 500 mM, Tris/HCl 100 mM, MgCl₂ 15 mM, pH 8.3)
- 4 µl di una miscela di deossinucleotidi (5 mM finale ciascuno)
- 2.5 µl di Nde FW 92 (25 picomoli);
- 2.5 µl di 3'-ssDBF (25 picomoli);
- 30 µl di DNA stampo ssDBF (100 ng);
- 2 µl di DMSO;
- 49 µl di H₂O.

Si sono ripetuti 30 cicli così impostati:

30 secondi a 95°C (denaturazione)

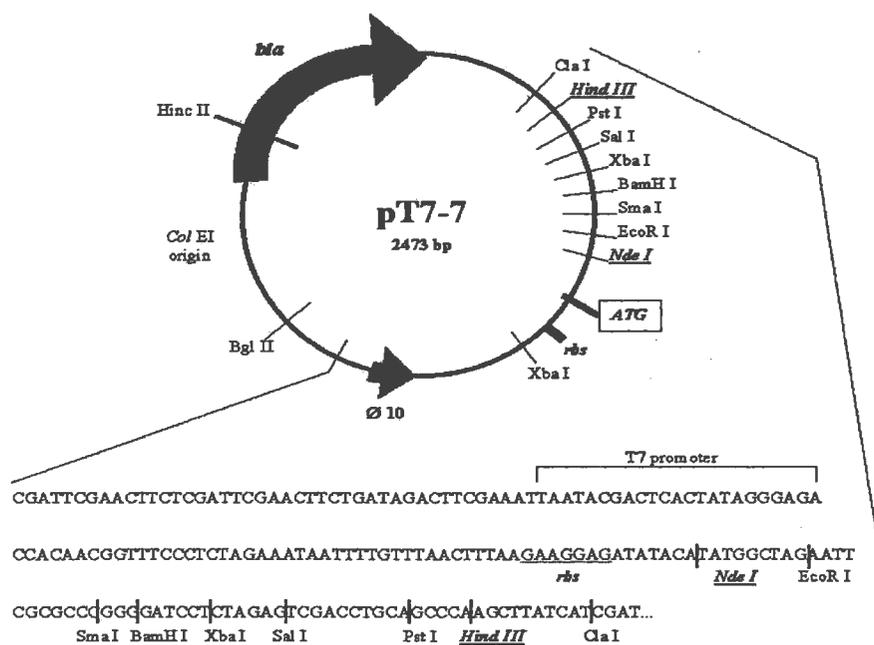
30 secondi a 56 °C (appaiamento dei primers sullo stampo)

90 secondi a 72°C (estensione dei primers)

Al termine è seguita un'estensione finale di 8 minuti a 72°C.

Dopo trattamento del frammento amplificato e del vettore pT7.7 con gli opportuni enzimi di restrizione e loro purificazione, si è allestita la reazione di ligazione con una quantità di vettore e di inserto in rapporto 1:3 in un volume finale di 20 µl contenete 1 µl dell'enzima ligasi del fago T4 (Promega) ed il tampone fornito dalla ditta; si è incubato per 12 ore a 16°C.

La miscela di ligasi è stata quindi utilizzata per trasformare le cellule E.coli Top10 (Promega) rese competenti col metodo dello shock termino come descritto da Nishimura (1990)



Mappa del vettore di espressione batterica pT7.7 utilizzato per il subclonaggio di parte del cDNA di ssDBF. In corsivo sottolineato sono rappresentati i siti di restrizione utilizzati per il subclonaggio.

2.11 Preparazione delle cellule di *E.coli* (ceppo BL21) competenti

La linea BL21 (DE 3) è un ceppo della linea K12 de *E. coli*, nel cui cromosoma è stato inserito il gene codificante per la T7-RNA polimerasi virale, posto sotto il controllo del promotore Lac UV 5, inducibile dall'IPTG. (Isopropil- β -D-tiogalattopiranoside). Il trattamento di cellule BL21 con IPTG determina l'attivazione dell'espressione del gene della T7-RNA polimerasi. In cellule BL21 trasformate con un plasmide del tipo pT7.7, la T7-RNA polimerasi attiva la trascrizione a partire dal suo promotore ϕ -10, presente nel plasmide a monte della sequenza *polilinker* di pT7.7 e quindi la sintesi della proteina il cui gene è stato clonato all'interno del plasmide (Studer *et al.*, 1991).

Per rendere competenti le cellule BL21 di *E. coli*, è stato utilizzato il metodo descritto da Nishimura *et al.* (1990). 0.5 ml di una coltura *over night* di batteri BL21 sono stati inoculati in 50 ml di medium A (LB broth addizionato di 10 mM MgSO₄·7H₂O e glucosio 0.2%) e cresciuti fino ad una fase di crescita logaritmica di 0.4 OD. Successivamente, i batteri sono stati mantenuti in ghiaccio per 10 minuti e poi recuperati per centrifugazione a 2000 g per 10 minuti a 4 °C. 2.5 ml di soluzione B (Glicerolo 30% (v/v), PEG-7500 12% (p/v), MgSO₄ 12 mM pH 7.0) sono stati aggiunti ai batteri risospesi lentamente in 0.5 ml di medium A, precedentemente raffreddato in ghiaccio. I batteri, resi competenti, sono stati suddivisi in aliquote da 0.1 ml in tubi *Eppendorf* e conservati a -80°C.

2.12 Trasformazione dei batteri BL21 con i vettori di espressione ed induzione dell'espressione proteica

100 μ l di batteri BL21, resi precedentemente competenti, sono stati addizionati 100 ng di DNA plasmidico, incubati per 30 minuti in ghiaccio e successivamente per

60 secondi a 42°C. In seguito, i batteri sono stati inoculati in 1 ml di terreno SOB, cresciuti per 1 ora a 37°C. Successivamente, sono stati trasferiti in 20 ml di brodo LB contenente ampicillina 0.1 mg/ml e cresciuti *over night* a 37°C, in agitazione.

4 ml della coltura *over night* di batteri BL21 trasformati sono stati quindi inoculati in 250 ml di brodo LB contenente ampicillina 0.1 mg/ml e glucosio 20 mM e cresciuti fino ad una O.D_{600 nm} di circa 0.6-0.8. A questo punto, è stato aggiunto IPTG alla concentrazione finale di 10 mM ed i batteri sono stati incubati per ulteriori cinque ore circa a 30°C, in agitazione, per permettere l'espressione della proteina, il cui gene è stato clonato all'interno del vettore pT7.7.

2.13 Estrazione di ssDBF ricombinante

I batteri BL21, dopo induzione con l'IPTG, sono stati centrifugati a 2000 g, per 20 minuti, a 4 °C ed il preparato è stato resospeso in una quantità di tampone di lisi (Tris-HCl 20 mM pH= 8, NaCl 250 mM, EDTA 1 mM, TWEEN-20 0.25%, PMSF e β Mercaptoetanololo 0.1 mM), pari a 10 ml per ogni grammo di precipitato; quindi i batteri sono stati lisati per sonicazione (3 strokes per 30 secondi intervallati da 30 secondi in ghiaccio). Il lisato viene centrifugato a 10000 rpm per 10 minuti a 4°C ed il supernatante recuperato. 20 μ l di surnatante sono stati usati per saggiare la presenza e l'attività della proteina ricombinante in SDS PAGE.

3.1 Identificazione di una proteina che lega ad alta specificità il motivo telomerico d(CCCTAA)_n a filo singolo

Da quanto riportato nell'introduzione, appare chiaro che questo è certamente un momento di grande fermento per la ricerca sui telomeri e sulla telomerasi ed è probabile che in futuro molto prossimo verrà fatta luce su punti ancora oscuri come ad esempio quali e con quali meccanismi specifici fattori proteici siano in grado di regolare l'attività della telomerasi per mantenere i telomeri ad una lunghezza costante o di indurre la sua inattivazione nelle cellule somatiche normali o, al contrario la sua riattivazione durante l'immortalizzazione cellulare o nella tumorigenesi umana

Sono state trovate ed in parte identificate in gruppi sistematici molto distanti, alcune proteine capaci di legarsi in modo specifico al filamento telomerico a blocchi di guanine ed al suo filamento complementare (Prince *et al.*, 1992; McKay *et al.*, 1992; Cardenas *et al.*, 1993), ma la loro funzione nell'organizzazione e gestione dei telomeri appare ancora oggi molto incerta e discussa. E' evidente che è diventato d'estremo interesse giungere alla completa caratterizzazione funzionale

e strutturale di tali proteine per elucidarne il reale ruolo *in vivo*, ma nello stesso tempo è importante continuare lo studio relativo all'identificazione nei diversi organismi di nuove componenti proteiche che interagiscono con sequenze a motivi telomerici allo scopo di giungere ad un quadro quanto più ampio e completo delle loro diffusione, natura e funzione.

In particolare, mentre sono state compiute molte indagini riguardo ai fattori proteici in grado di legare i motivi telomerici ripetuti a doppio filamento ed il motivo a singolo filamento a blocchi di guanosine, solo recentemente alcuni autori hanno posto l'attenzione ad eventuali proteine nucleari capaci di interagire specificamente con l'altro filamento singolo, quello terminante al 5' e caratterizzato dalla presenza di blocchi di citosine, anche se esso non appare essere stabilmente presente in questa forma nel cromosoma, in quanto si trova associato al suo filamento complementare in un canonico duplex.

Il lavoro svolto durante il dottorato di ricerca ha riguardato una serie di studi volti a caratterizzare nuove proteine nucleari interagenti con le sequenze telomeriche, in particolare quella a motivo pirimidinico che caratterizza i telomeri di tutti i vertebrati, uomo compreso. Utilizzando tecniche di biochimica e di biologia molecolare si è giunti all'identificazione e caratterizzazione sia in cellule aviarie sia in cellule umane, di proteine nucleari che legano il DNA a filo singolo a motivo telomerico ripetuto $d(CCCTAA)_n$, una delle quali con alta specificità di sequenza.

Per quanto riguarda la sequenza bersaglio alla base della ricerca condotta, si è scelto l'oligonucleotide da noi denominato HTC4 (Human Telomeric Cytosine-block), a quattro blocchi di citosine e riprodotto il motivo telomerico CCCTAA dei vertebrati; con HTG4 è stato invece indicato il suo complementare a motivo TTAGGG (vedere tabella III).

Saggi di ritardo della mobilità elettroforetica condotti in condizioni native, dopo incubazione dell'oligonucleotide HTC4 marcato con un estratto nucleare proteico

Nome	Sequenza (5'→3')	Specificità di <i>binding</i>
HTC4	CCCTAACCCCTAACCCCTAACCCCT	+
HTC3	CCCTAACCCCTAACCT	-
HTC2	CCCTAACCCCT	-
HTA4	TAACCCTAACCCCTAACCCCTAA	+/-
HTR2	ACTGGCCCTAACCCCTAATGACT	+/-
ATC4	CCCTAAACCCTAAACCCTAAACCCT	+
BTC4	CCTAACCTAACCTAACCT	-
TTC4	CCCCAACCCCAACCCCAACCCC	+
PTC4	GCCTAAGCCTAAGCCTAAGCCT	-*
φTC4	GCCTAAGCCTAAGCCTAAGCCT	+
HTC6	CCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAA	+
Kras	GCTCCCTCCCTCCCTCCTTCCCTCCCTCCC	-*
HTG4	AGGGTTAGGGTTAGGGTTAGGG	-*
CarG	CCATTTCTAATTAGGTAAAAG	-*
C22	CCCCCCCCCCCCCCCCCCCC	-

* Banda ritardata con differente mobilità e specificità

TABELLA III: Sequenze ed attività di *binding* degli oligonucleotidi utilizzati nello studio

da cellule umane HeLa, hanno messo in evidenza, con la presenza di due bande ritardate, la formazione di due complessi proteina-DNA (figura 7, corsia 4). Entrambi i complessi persistono alla competizione da parte di oligonucleotidi a sequenza casuale, di RNA totale e di DNA di *E. coli* denaturato ad una concentrazione 100 volte superiore (w/w) a quella dell'oligonucleotide sonda. Solo ad eccessi superiori a 100 volte di DNA competitore, la banda ritardata più lenta mostra un cospicuo effetto di competizione, mentre quella più veloce rimane pressochè inalterata anche a concentrazioni 700-1000 volte superiori di competitore (Marsich *et al.*, 1996).

Nel caso dell'oligonucleotide marcato HTG4, viene evidenziato un solo complesso, a mobilità e specificità ben distinte da quelle esibite dalla sonda HTC4, suo complementare, mentre alcun complesso stabile sembra formarsi con il duplex telomerico HTC4:HTG4 (figura 7, corsie 2 e 3).

Sono stati quindi valutati gli effetti della durata e della temperatura d'incubazione sulla formazione del complesso specifico che si forma con la sonda telomerica HTC4; i risultati sono riportati in figura 9 e 10. L'oligonucleotide marcato è stato incubato con l'estratto proteico in presenza di un eccesso di 500 volte di DNA competitore per sopprimere la banda ritardata meno specifica. Si è osservato che 30 minuti d'incubazione sono appena sufficienti a generare una debole banda ritardata a 25°C, ma non a 5°C, in accordo con una cinetica di associazione molto lenta a bassa temperatura. Dopo 3 ore viene esaurita la massima capacità di *binding* e non si osservano sostanziali differenze nell'intensità della banda relativa al complesso sia che l'incubazione sia stata effettuata a 5°C che a 25°C (figura 9, corsie dalla 2 alla 9). Per saggiare l'influenza della temperatura, l'estratto proteico è stato riscaldato a 50°C ed a 75°C prima e dopo l'incubazione a 25°C con la sonda telomerica (figura 10). In entrambi i casi, 50°C sono sufficienti a determinare un cospicuo indebolimento nell'intensità della banda elettroforetica

(figura 10, corsie 2 e 4), mentre 75°C causano l'inattivazione irreversibile della proteina come evidenziato dalla scomparsa del suo complesso con l'oligonucleotide sonda (figura 10, corsie 1 e 3).

E' inoltre interessante osservare che la forza ionica non ha un sensibile effetto destabilizzante sulla formazione del complesso specifico, ad indicazione che la componente elettrostatica dell'energia di formazione di tale complesso ha un ruolo solo marginale: concentrazioni fino a 200 mM di KCl non influenzano minimamente la sua stabilità (figura 10, mentre solo a concentrazioni di KCl superiori a 800 mM (dati non riportati), si assiste ad un rilevante indebolimento dell'intensità della banda ritardata specifica.

3.2 Determinazione dei pesi molecolari dei complessi proteine-DNA

Per quanto riguarda la o le componenti che legano HTC4, non avendo ottenuto un risultato attendibile per una loro caratterizzazione in termini di peso molecolare mediante South-Western, si è deciso di ricorrere alla tecnica dell'UV-crosslinking.

Con tale tecnica, l'irraggiamento con luce UV della miscela d'incubazione dell'estratto proteico con l'oligonucleotide marcato, produce la formazione di legami covalenti tra l'oligonucleotide e le componenti proteiche che lo complessano. Ciò permette che la o le componenti che hanno legato la sonda, possano essere evidenziate autoradiograficamente, previa separazione in elettroforesi denaturante SDS. Un presupposto per poter utilizzare tale metodo per stimare i pesi molecolari delle proteine è che il peso molecolare dell'oligonucleotide legato sia modesto rispetto a quello della componente proteica (nel nostro caso gli oligonucleotidi presentano un peso molecolare approssimativo di 7 kDa, contro un peso molecolare medio delle proteine di 50 kDa).

Con l'oligonucleotide sonda HTC4, i due complessi evidenziati in EMSA si



Figura 7

Saggio EMSA dopo incubazione dell'estratto nucleare da cellule HeLa (circa 0.01 µg), con 0.04 pmol della sonda marcata HTG4 (corsia 2), HTC4 (corsia 4) e con il duplex marcato HTC4:HTG4 (corsia 3). Corsie 1 e 5, HTG4 e HTC4 senza estratto nucleare. Le frecce indicano la posizione dei due complessi specifici per l'oligonucleotide sonda HTC4.

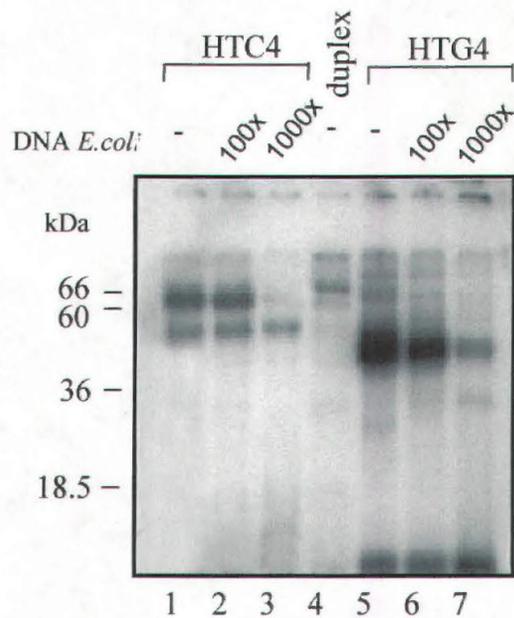
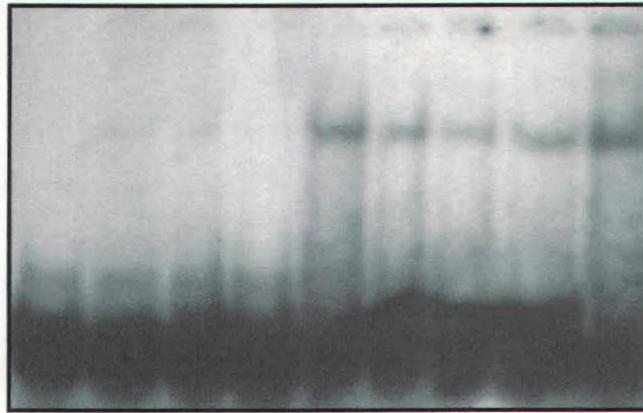


Figura 8

*SDS-PAGE al 12% dopo UV crosslinking delle miscele d'incubazione dell'estratto da cellule HeLa con gli oligonucleotidi marcati HTC4 e HTG4 in assenza di competitore (corsie 1 e 5, rispettivamente) ed in presenza degli eccessi indicati in figura di DNA di *E.coli* (corsie 2 e 3 e corsie 6 e 7), e con il duplex marcato HTC4:HTG4 in assenza di competitore (corsia 4). A sinistra della figura sono riportate le posizioni dei markers di peso molecolare.*

Temperatura (C)	25	5	25	5	25	5	25	5	25
Tempo d'incubazione (ore)	0	0.1	0.1	0.5	0.5	3	3	18	18

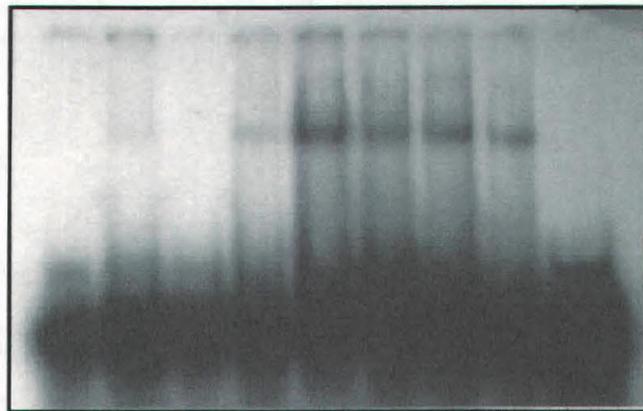


1 2 3 4 5 6 7 8 9

Figura 9

Saggio EMSA delle miscele composte dall'estratto proteico da cellule HeLa e dalla sonda HTC4, incubate a 5 o 25 gradi C per 5, 30, 180 minuti e per 18 ore, in presenza di DNA denaturato di E.coli in eccesso di 500 volte (w/w).

Temperatura (C)	75	50	75	50	-	-	-	-
Prima o Dopo incubazione	D	D	P	P	-	-	-	-
conc. KCl (mM)	50	50	50	50	200	50	10	0



1 2 3 4 5 6 7 8 9

Figura 10

- Saggio EMSA dopo incubazione a 25 gradi C dell'estratto nucleare con la sonda HTC4 :*
- corsie 1 e 2, estratto nucleare riscaldato a 75 e 50 gradi C rispettivamente, Dopo incubazione con la sonda
 - corsie 3 e 4, estratto nucleare riscaldato a 75 e 50 gradi C rispettivamente, Prima dell'incubazione con la sonda
 - corsie dalla 5 alla 8, incubazione in presenza di differenti concentrazioni di KCl
 - corsia 9, oligonucleotide senza estratto

riproducono anche negli esperimenti di UV-crosslinking, dove i pesi molecolari apparenti, determinati sulla base di quelli di markers noti, risultano essere, per la componente più lenta, di circa 65-70 kDa, e per la più veloce, 50-52 kDa (figura 8, corsie 1, 2 e 3).

Il comportamento quantitativamente e qualitativamente diverso delle due componenti nei confronti della competizione, che si riproduce sia in EMSA che in UV-crosslinking, ci ha permesso di identificare nella proteina a peso molecolare apparente di 50-52 kDa, quella responsabile del complesso EMSA più specifico, nella componente di 65-70 kDa, quella responsabile del complesso a specificità ridotta e minore mobilità in EMSA.

I prodotti di UV crosslinking di HTG4 con l'estratto, rivelano invece una banda molto intensa, che subisce un moderato effetto di competizione dal DNA di *E. coli* in eccesso di 100 volte (w/w) ed uno molto marcato dall'eccesso di 1000 volte (figura 8, corsie 5, 6 e 7). Il peso molecolare apparente di questo complesso è di 40- 42 kDa. Compaiono inoltre alcune bande molto più deboli, che però subiscono un più o meno marcato effetto di competizione.

Inoltre, sempre in linea con i risultati EMSA, appare chiaro (figura 8, corsia 4) che il duplex telomerico HTC4:HTG4 non viene legato dalle componenti sensibili ai filamenti singoli; compaiono tuttavia alcuni deboli segnali, il più cospicuo dei quali ha un peso molecolare apparente di circa 70 kDa.

3.3 Presenza della proteina in estratti nucleari da diverse origini

La ripetizione telomerica (CCCTAA)_n caratterizza le estremità cromosomiche di tutte le specie di vertebrati finora considerate quindi, se la proteina da noi descritta in HeLa (Marsich *et al.*, 1996) è in qualche modo coinvolta in aspetti della funzione telomerica, suoi omologhi potrebbero essere presenti in altre specie di

vertebrati, sebbene non necessariamente espressa in tutti i tessuti. Per verificare ciò, sono stati allestiti una serie di saggi EMSA dopo incubazione dell'oligonucleotide marcato HTC4 con estratti nucleari di diverse origini.

L'attività specifica di *binding* alla sequenza telomerica allo studio si è mostrata ampiamente diffusa negli estratti cellulari di tutti i tipi di tessuto analizzati per una data specie (fegato, timo, linee cellulari linfocitarie, mieloidi, eritroidi e da tiroide), nonché in tutte le fonti di diversa origine specie specifica (linee cellulari di umane e criceto, tessuti e linee cellulari di ratto, di maiale, di vitello e di pollo), mostrando in tutti i casi bande ritardate con comportamento alla competizione, mobilità e specificità per la sequenza telomerica a filamento singolo, strettamente omologhi a quelli evidenziati nel caso dell'estratto da cellule HeLa (Marsich *et al.*, 1998)

3.4 Effetti della modulazione nel motivo telomerico sul legame della proteina

Di notevole interesse si è mostrato lo studio riguardante il grado di specificità dell'interazione della proteina nucleare in questione con l'oligonucleotide sonda ovvero, l'influenza di modulazioni di motivi telomerici sulla formazione del complesso.

Nella tabella III sono riportati alcuni tra gli oligonucleotidi utilizzati per saggi di competizione con HTC4 o di *binding* con la proteina e la loro capacità di formare complessi stabili da noi evidenziabili in saggi di ritardo della mobilità elettroforetica.

Se si escludono oligonucleotidi a sequenza random od a sequenza non telomerico-simile, non riportati nella tabella, gli oligonucleotidi che sembrano non essere riconosciuti dalla componente in esame sono quelli che contengono blocchi di due citosine, come BTC4, riprodotte il motivo telomerico da *Bombix mori* (figura 11, corsia 4 e 7), PTC4, sequenza telomerica di *Parascaris*, o

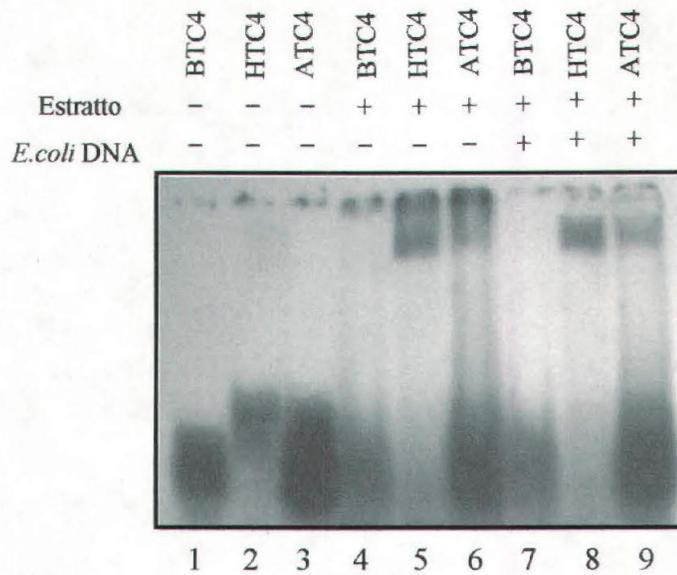


Figura 11

Saggio EMSA dopo incubazione degli oligonucleotidi marcati con l'estratto proteico:

- corsie dalla 1 alla 3, oligonucleotidi senza estratto
- corsie dalla 4 alla 6, oligonucleotidi BTC4 (corsia 4), HTC4 (corsia 5) e ATC4 (corsia 6) in assenza di competitore
- corsie dalla 7 alla 9, oligonucleotidi BTC4 (corsia 7), HTC4 (corsia 8) e ATC4 (corsia 9), in presenza di un eccesso di 500 volte (w/w) di DNA di *E.coli*)

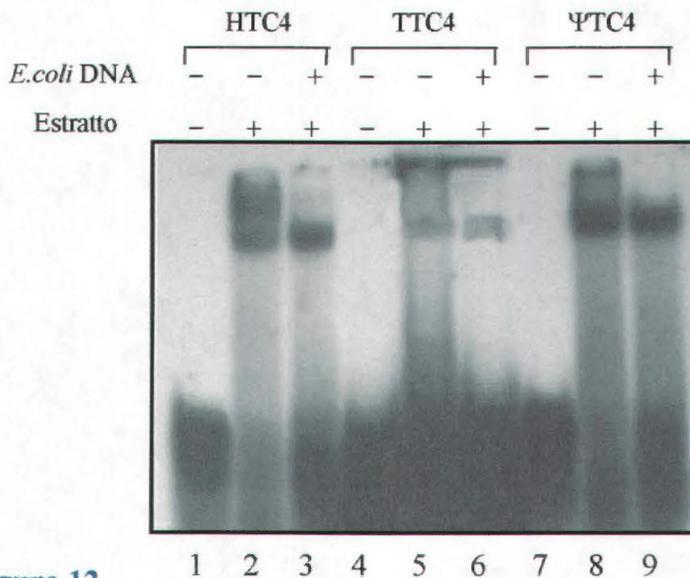


Figura 12

Saggio EMSA dopo incubazione dell'estratto proteico di cellule HeLa con le sonde oligonucleotidiche marcate qui sotto elencate:

- corsie 2 e 3, HTC4 in assenza ed in presenza di un eccesso di 500 volte (w/w) di DNA di *E.coli*
- corsie 5 e 6, TTC4 in assenza ed in presenza di un eccesso di 500 volte (w/w) di DNA di *E.coli*
- corsie 8 e 9, ϕTC4 in assenza ed in presenza di un eccesso di 500 volte (w/w) di DNA di *E.coli*
- corsie 1, 4, 7, sonde oligonucleotidiche in assenza di estratto proteico

l'oligonucleotide CArG, sequenza regolativa di alcuni geni espressi in cellule muscolari (dati non riportati). Allo stesso modo, anche oligonucleotidi che presentano blocchi di citosine non opportunamente spaziati, per esempio l'oligonucleotide kRas, non sembrano venire legati abbastanza stabilmente dalla componente proteica in esame. Come nel caso dello stesso kRas, alcuni oligonucleotidi hanno mostrato in EMSA la comparsa di una banda ritardata, più o meno specifica, ma che presentava una mobilità ben distinta da quella relativa al complesso specifico allo studio. La mancanza di reciproca competizione tra la sonda HTC4 e suddetti oligonucleotidi nella formazione dei complessi, ha mostrato chiaramente che le componenti proteiche responsabili dei relativi complessi dovevano essere diverse e distinte.

Solo con oligonucleotidi diversi caratterizzati dal possedere quattro blocchi di almeno tre citosine, come ATC4 (sequenza rappresentativa del motivo telomerico di *Arabidopsis*) (corsie 3 e 6 di figura 11), TTC4 (riproducente la sequenza telomerica del ciliato *Tetrahymena*) (figura 12, corsie 4-6) e ϕ TC4 (basato sullo spostamento della timina del motivo di vertebrato dal 3' al 5' dei blocchi di citosine) (figura 12, corsie 7-9), tutti gli estratti di proteine nucleari sono in grado di formare specifici complessi, non competuti anche da eccessi di 250-500 volte (w/w) di competitore aspecifico (DNA di *E. coli* denaturato). Comunque, saggi di competizione crociata mostrano che l'ordine di affinità dei diversi oligonucleotidi riconosciuti specificatamente dalla proteina allo studio, è tale che la più alta è quella esibita verso il motivo telomerico di vertebrati, riprodotto dall'oligonucleotide HTC4 (Marsich *et al.*, 1998).

Quanto alla dipendenza dell'interazione dal numero di ripetizioni nell'oligonucleotide sonda, queste sembrano influenzare molto la stabilità del complesso. Saggi EMSA sono stati condotti utilizzando quali oligonucleotidi sonda quelli da noi denominati HTC3 e HTC2, rispettivamente a tre e due blocchi di citosine (vedere tabella III). La totale assenza di bande ritardate (figura 13,

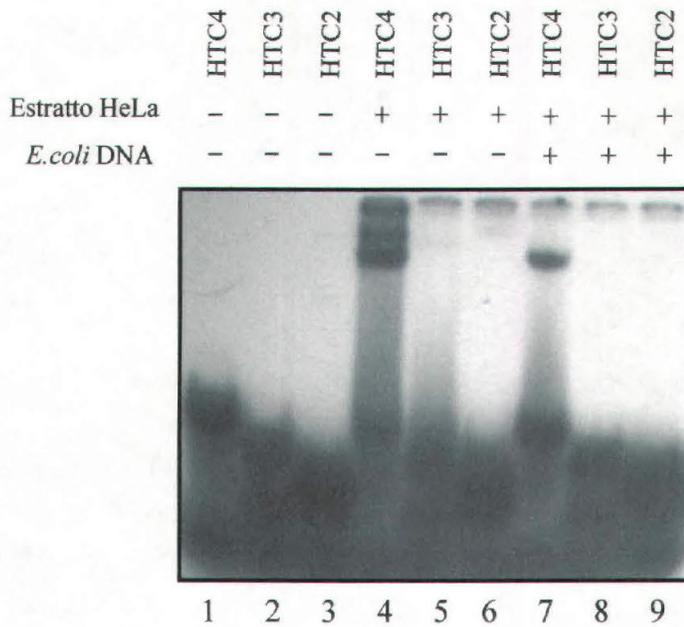


Figura 13

Saggio EMSA dopo incubazione degli oligonucleotidi marcati HTC4, HTC3 e HTC2 con l'estratto nucleare da cellule HeLa:

- corsie 1, 2 e 3, oligonucleotidi senza estratto
- corsie 4, 5 e 6, oligonucleotidi incubati con l'estratto proteico
- corsie 7, 8 e 9, oligonucleotidi incubati con l'estratto in presenza di un eccesso di competitore (DNA di *E.coli* in eccesso di 500 volte)



Figura 14

SDS PAGE dopo UV crosslinking delle miscele d'incubazione relative alle corsie dalla 4 alla 9 di figura 11

corsia 5 e 6) dopo incubazione dell'estratto proteico da cellule HeLa con questi due oligonucleotidi, indica la loro incapacità di formare complessi sufficientemente stabili in grado di resistere per la durata della corsa elettroforetica. Il medesimo risultato, qui non riportato in figura, è stato osservato anche dopo incubazione con l'estratto di proteine nucleari di eritrociti di pollo e da fegato di ratto.

Tuttavia, esperimenti in parallelo di UV-crosslinking, mostrano la formazione di due addotti anche per HTC3 e HTC2 (figura 14, corsie 2 e 3), indicando che un complesso specifico relativamente stabile si può formare a seguito di UV-crosslinking, ma d'altra parte la sua cinetica di dissociazione è troppo veloce per permettere la sua sopravvivenza durante un esperimento EMSA.

L'incubazione di HTC6 marcato con estratti proteici ha fornito risultati di specificità ed affinità paralleli a quelli di HTC4 nel caso di tutti gli estratti nucleari analizzati.

Si era sospettato che la minore stabilità del complesso nel caso degli oligonucleotidi HTC2 e HTC3 fosse dovuta, piuttosto che al numero di ripetizioni, alla loro minore lunghezza e conseguentemente ad un numero inferiore di interazioni elettrostatiche. Per di più ci si chiedeva se gli elementi più importanti nel determinare la specificità di legame fossero i blocchi CCC o quelli TAA.

Sono stati quindi presi in considerazione gli oligonucleotidi HTR2 e HTA4 (tabella III); il primo contiene due ripetizioni telomeriche fiancheggiate da due sequenze random in modo da possedere lunghezza uguale a quella di HTC4, il secondo è caratterizzato da quattro blocchi TAA uniti da tre blocchi CCC. In assenza di DNA competitore (figura 15, corsie 5 e 8), entrambi esibiscono una lunga strisciata che viene totalmente soppressa in presenza del competitore (corsie 6 e 9), indicando la formazione di complessi molto poco specifici e stabili che vanno dissociandosi durante la corsa elettroforetica. Tali risultati indicano chiaramente che sono necessari almeno quattro blocchi di citosine, opportunamente spazati, per ottenere il completo sviluppo del potenziale di

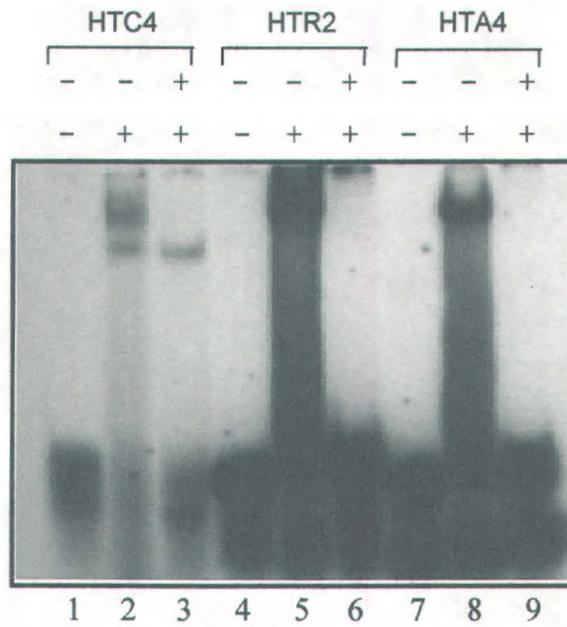


Figura 15

Saggio EMSA dopo incubazione degli oligonucleotidi marcati HTC4, HTR2 e HTA4 con l'estratto nucleare da cellule HeLa;

- corsie 1, 4 e 7, oligonucleotidi in assenza di estratto proteico

- corsie 2, 5 e 8, oligonucleotidi incubati con l'estratto proteico

- corsie 3, 6 e 9, oligonucleotidi incubati con l'estratto in presenza di DNA di E.coli (250 volte w/w)

legame della nostra componente o componenti proteiche allo studio. Si è così anche esclusa la possibilità che a conferire specificità all'interazione HTC4 e proteina fossero i quattro blocchi TAA, piuttosto che i blocchi di citosine.

Il primo punto, in particolare potrebbe suggerire un possibile ruolo, nell'interazione proteine-oligonucleotide, della struttura non canonica ad i-DNA: sono infatti necessari almeno 4 blocchi di citosine per il ripiegamento intramolecolare dell'i-DNA (Leroy *et al.*, 1994; Manzini *et al.*, 1994). Comunque, l'osservazione che l'oligomero kRas, nonostante sia capace di assumere la struttura intramolecolare ad i-DNA (Manzini *et al.*, 1994) e costituito dalla ripetizione di un motivo non molto diverso da quello telomerico, (CCCT)₃CCTT(CCCT)₃, non mostra chiaramente interazione con la proteina allo studio, sembra confutare tale ipotesi. Inoltre è contraria a questa ipotesi anche l'osservazione che il complesso si forma ugualmente a pH 8, dove la struttura i-DNA non è stabile, come a pH 6, dove lo è (dati non riportati).

Da quanto detto, è evidente che la specificità di sequenza della proteina mostra avere una notevole stringenza, richiedendo per la formazione di interazioni stabili, 4 o più blocchi di almeno 3 citosine regolarmente spaziati. E' indicativo il fatto ad esempio che una sequenza come ϕ TC4, analoga a quella di HTC4 ad eccezione dello spostamento della timina del motivo di vertebrato dal 3' al 5' dei blocchi di citosina, riduca di un certo grado l'affinità di *binding* se confrontata a quella di HTC4, mentre una sequenza estremamente simile in composizione di basi, come la sequenza kRas, non venga addirittura riconosciuta.

3.5 Isolamento delle proteine allo studio

Sulla base dei dati di specificità di *binding* all'oligonucleotide sonda della o delle componenti proteiche nei diversi estratti cellulari analizzati, si è passati alla purificazione di tali componenti utilizzando gli eritrociti di pollo quale sorgente più conveniente per l'ottenimento in forma pura delle proteine. A questo scopo è stata predisposta una procedura cromatografica *in batch*, con resina avidinata, funzionalizzata con l'oligonucleotide a sequenza telomerica ripetuta d(CCCTAA)₆, biotinilato al 3'. A differenza di ciò che viene fatto negli usuali saggi EMSA con l'estratto totale, non si è aggiunto competitore all'estratto incubato con la resina, ma ci si è riservati ad un passo successivo il saggio di specificità delle proteine così purificate.

Le proteine eluite sono state quindi analizzate su un gel di poliacrilammide denaturante in SDS. La figura 16 mostra il profilo elettroforetico risultante. Con il Blue Coomassie si evidenziano essenzialmente tre bande, con pesi molecolari apparenti di 40, 45 e 55 kDa. Queste tre componenti sono state designate CBNP α , CBNP β e CBNP γ , ovvero C-block Binding Nuclear Protein. Le corsie 4, 5 e 6 si riferiscono a tre diverse preparazioni, la corsia 2 mostra il profilo dell'estratto totale, mentre la corsia 3 riporta il profilo relativo all'eluato dopo incubazione dell'estratto proteico con la resina non funzionalizzata.

L'attività del *pool* delle proteine recuperate dalla cromatografia è stata valutata mediante saggi di *band-shift* elettroforetico. Importante osservare che per il recupero dalle proteine purificate dell'attività specifica dimostrata dall'estratto totale, si sono rilevate essenziali le condizioni riducenti nel mezzo d'incubazione con l'oligonucleotide sonda. La figura 17 riporta l'attività di *binding* alla sonda durante i vari stadi della procedura di purificazione. Le corsie 1 e 2 si riferiscono all'attività del materiale di partenza, rispettivamente in assenza ed in presenza di DNA denaturato di *E. coli* in eccesso. Dopo l'incubazione dell'estratto con la

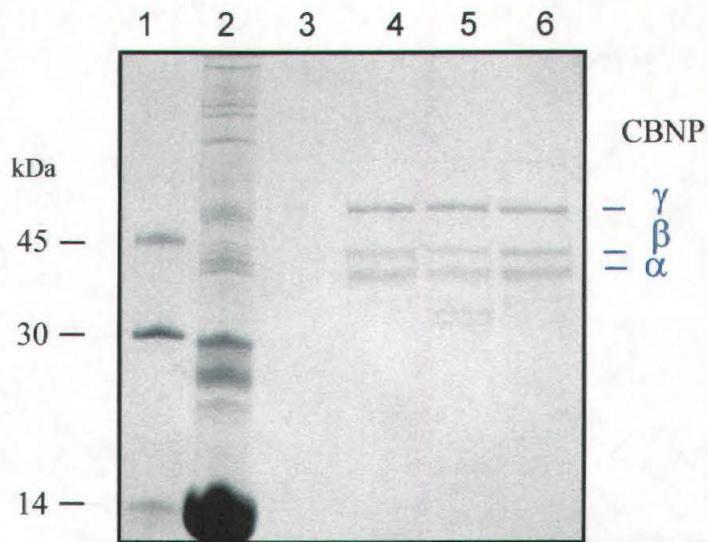


Figura 16

SDS PAGE all'8% dei campioni recuperati dalla cromatografia di affinità:

- corsia 1, markers di peso molecolare, ovalbumina (45 kDa), anidrasi carbonica (30 kDa), α lattalbumina (14 kDa);
- corsia 2, 10 μ g di estratto totale di eritrociti di pollo;
- corsia 3, componenti eluite da resina non funzionalizzata;
- corsie 4, 5, 6, tre differenti campioni proteici recuperati dalla cromatografia di affinità.

Le tre componenti eluite sono indicate al margine destro della figura come CBNP α , β , γ

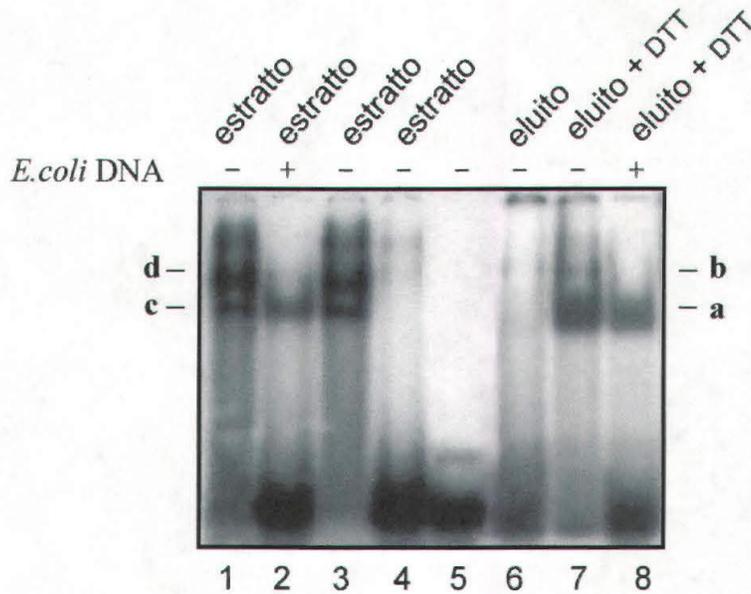


Figura 17

- corsie 1 e 2, estratto totale da eritrociti di pollo prima della purificazione con la resina funzionalizzata, incubato con la sonda HTC4, rispettivamente in assenza ed in presenza di un eccesso di 250 volte (w/w) di DNA di *E. coli*;
- corsie 3 e 4, attività di legame dell'estratto nucleare totale dopo incubazione con la resina non funzionalizzata e funzionalizzata con l'oligonucleotide HTC6
- corsia 5, sonda marcata HTC4;
- corsie da 6 a 8, attività della frazione proteica eluita dalla resina funzionalizzata, in assenza di ditiotreitolo (corsia 6) e in presenza di una concentrazione 1 mM di ditiotreitolo, (corsia 7 in assenza di competitore, corsia 8 in presenza di DNA di *E. coli* in eccesso di 250 volte w/w)

resina non funzionalizzata l'attività rimane nel surnatante (corsia 3), mentre viene rimossa dopo incubazione con la resina funzionalizzata (corsia 4). I campioni eluiti dalla resina funzionalizzata esibiscono la stessa attività dell'estratto, sia in assenza, sia in presenza del competitore (corsie 7 e 8). La presenza del DTT è indispensabile per il pieno recupero dell'attività, in quanto, la sua assenza, determina la scomparsa del segnale (corsia 6), oppure talvolta la comparsa di un segnale strisciato a mobilità e specificità notevolmente ridotte. Tuttavia tale complesso strisciato, come evidenziato da esperimenti di UV crosslinking *in situ*, dimostra di contenere la componente proteica da noi descritta. Numerose evidenze sperimentali, suggeriscono che è necessario che tale proteine si trovi in forma ridotta, e non complessata con altre componenti proteiche, per esplicare la sua attività di *binding* specifica ad HTC4. Potrebbe essere stimolante indagare se il potenziale redox intracellulare possa modulare la sua attività biologica. Saggi quantitativi in presenza di concentrazioni crescenti di riducente, hanno evidenziato che per il pieno recupero dell'attività è richiesta almeno una concentrazione 0.5 mM di DTT.

Per stabilire accuratamente la relazione tra le attività presenti nell'estratto totale e le proteine presenti nei campioni recuperati dalla cromatografia di affinità, la specificità di riconoscimento dell'oligonucleotide sonda HTC4 da parte di queste ultime è stata saggiata tramite esperimenti EMSA competitivi. Ad esempio la formazione del complesso specifico proteina-HTC4, in presenza di un eccesso di competitore aspecifico, è stata messa alla prova con concentrazioni 10 volte maggiori di vari oligonucleotidi aventi varianti di sequenze di tipo telomerico. Il risultato (figura 18) risulta in perfetto accordo con le affinità relative dimostrate dall'estratto proteico nucleare da cellule HeLa (Marsich *et al.*, 1998), nonostante la differenza di specie tra gli estratti di partenza (Tabella III).

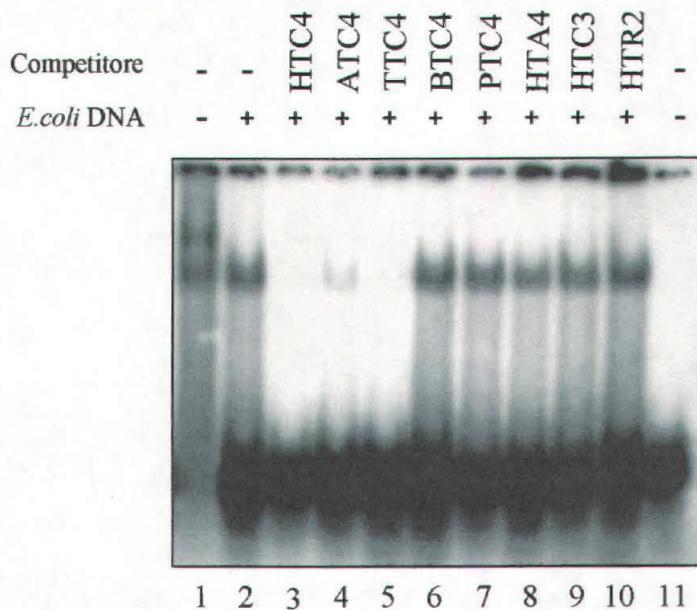


Figura 18

Saggio EMSA di competizione con differenti oligonucleotidi di tipo telomerico

La sonda marcata HTC4 (0.2 pmol) è stata incubata con 100 ng delle tre componenti proteiche recuperate dalla resina di affinità:

- corsia 1 in assenza di competitori;
- corsie dalla 2 alla 10, in presenza di un eccesso di 250 volte (w/w) di DNA denaturato di *E.coli* e di un eccesso di 10 volte di oligonucleotidi di tipo telomerico non marcati;
- corsia 11, sonda marcata HTC4 in assenza di proteina

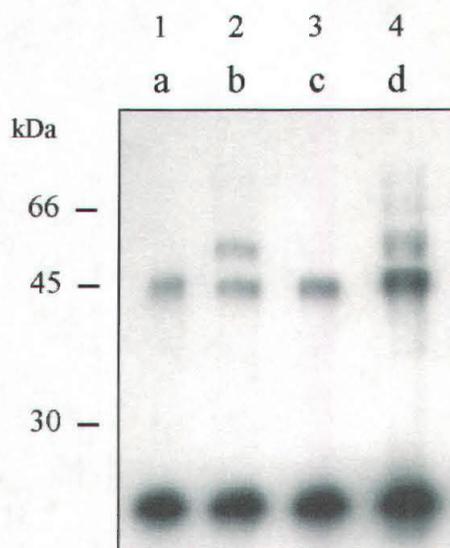


Figura 19

Analisi in SDS PAGE all'8% dopo UV crosslinking delle bande ritardate evidenziate nel saggio EMSA di Figura 15:

- corsia 1, addotti covalenti dopo trattamento con UV dei complessi evidenziati nella banda a in corsia 8 della Figura 15;
- corsia 2, addotti covalenti dei complessi evidenziati nella banda b in corsia 7;
- corsia 3, addotti covalenti dei complessi evidenziati nella banda c nella corsia 2;
- corsia 4, addotti covalenti dei complessi evidenziati nella banda d in corsia 1;

Poiché la cromatografia è stata condotta in assenza di competitore, è stato importante stabilire quale delle tre componenti da questa ottenute, fosse responsabile della banda EMSA a mobilità maggiore, ossia la più specifica.

Una prima evidenza è stata ottenuta dal risultato di un esperimento di UV crosslinking *in situ*. Le bande EMSA veloci a e c delle corsie 8 e 2 di figura 17, sono state escisse dal gel, irradiate con luce UV e nuovamente sottoposte ad elettroforesi denaturante. Gli addotti proteici così evidenziati, manifestano un unico segnale la cui mobilità è solo lievemente inferiore a quella di CBNP α di per sé stessa, rilevata in Blue Coomassie (figura 19, corsie 1 e 3). Questo dato era indicativo del fatto che CBNP α potesse essere la proteina responsabile del segnale EMSA specifico, in quanto la leggera differenza in mobilità è dovuta alla presenza o meno dell'oligonucleotide HTC4 covalentemente legato. La stessa procedura applicata ai complessi EMSA a bassa mobilità, cioè b e d delle corsie 1 e 7 della figura 17, evidenzia due segnali dopo SDS-PAGE, il più veloce corrispondente al precedente, mentre l'altro è leggermente più lento di CBNP γ (corsie 2 e 4 della figura 19). Questo in modo analogo dimostrerebbe che la banda EMSA più lenta contiene, oltre a CBNP α , anche CBNP γ .

A ulteriore conferma dell'identità della proteina responsabile del complesso più specifico, le bande corrispondenti a CBNP α , CBNP β e CBNP γ sono state escisse separatamente dal gel, elettroeluite, liofilizzate e risospese nel tampone di incubazione usato per i saggi EMSA. Per ogni componente è stato allestito un saggio EMSA in presenza di DTT 1 mM e di DNA di *E. coli*. CBNP α recupera pienamente l'attività di legame dopo elettroeluizione (figura 20, corsia 3) e questa risulta in perfetto accordo con quella dell'estratto totale (corsia 1 della stessa figura) e con quella delle tre componenti principali recuperate dalla cromatografia di affinità ed elettroeluite in *pool* (corsia 6). Al contrario, CBNP β e CBNP γ dopo

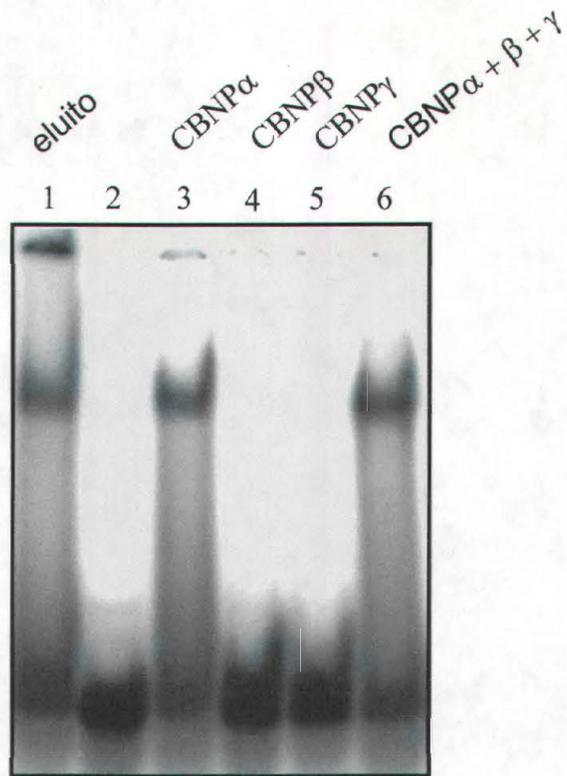


Figura 20

Saggio EMSA delle componenti proteiche CBNP α , CBNP β e CBNP γ , dopo elettroeluizione delle bande escisse dal gel SDS PAGE:

- corsia 1, totale delle proteine eluite dalla resina di affinità;
- corsia 2, oligonucleotide HTC4 marcato;
- corsia 3, CBNP α ;
- corsia 4, CBNP β ;
- corsia 5, CBNP γ ;

- corsia 6, le tre componenti proteiche elettroeluite assieme

L'incubazione delle proteine con la sonda è stata effettuata in presenza di un eccesso di 500 volte w/w di competitore aspecifico

elettroeluzione non manifestano alcun segnale EMSA (corsie 4 e 5) anche in assenza di un eccesso di competitore aspecifico.

3.6 Identificazione delle proteine isolate

A questo punto si è cercato di identificare e caratterizzare la sequenza primaria di queste proteine. I tentativi di determinare la natura di CBNP α , CBNP β e CBNP γ , mediante sequenziamento diretto dei campioni sono inizialmente falliti per la presenza di un blocco all'estremità N-terminale. Tuttavia nel caso di CBNP α si è riusciti ad ottenere la sequenza di 12 residui amminoacidici di un frammento interno derivante da digestione endoproteolitica. Questo risultato non si è rilevato particolarmente significativo in quanto la sequenza individuata, MFVGGLSWDTSK, corrisponde al dominio di legame all'RNA (motivo noto come RNP1), presente in molte proteine dell'RNA eterogeneo nucleare (hnRNP), sebbene abbia contribuito a restringere evidentemente il campo di ricerca.

Per ottenere ulteriori informazioni sulla natura di queste proteine si è fatto ricorso all'analisi della loro sequenza peptidica attraverso la tecnica di spettrometria di massa MALDI-TOF, condotta presso il laboratorio del CNR di Napoli I.A.B.B.A.M. dal dott. Andrea Scaloni.

Ciascuna delle tre proteine è stata digerita con tripsina ed i peptidi ottenuti sono stati separati per RP-HPLC e sottoposti ad analisi MALDI-TOF. I valori di massa dei peptidi sono stati quindi utilizzati per identificare la natura delle specie proteiche, dall'analisi di banche dati proteiche e nucleotidiche (vedi Materiali e Metodi) grazie all'ausilio di specifici softwares (tabella IV).

Buoni risultati sono stati ottenuti nel caso di CBNP α e CBNP β , per le quali i valori di massa di rispettivamente 22 e 18 peptidi risultano perfettamente

CBNP α			CBNP β			CBNP γ
Massa	ssDBF	CRP1	Massa	ssDBF	CRP1	Massa
7363.5	Ac2-84	n.a.	7363.8	Ac(2-84)	n.a.	1098.7
8108.0	Ac2-91	n.a.	8107.8	Ac(2-91)	n.a.	1195.3
1328.5	92-103	75-86	1328.4	92-103	75-86	1505.3
1344.3	92-103ox	75-86ox	1344.5	92-103ox	75-86ox	1519.7
1456.7	92-104	75-87	1456.8	92-104	75-87	1554.4
1472.9	92-104ox	75-87ox	1472.7	92-104ox	75-87ox	1596.1
1159.0	104-112	87-95	1159.2	104-112	87-95	1711.7
1169.9	113-122CAM	96-105CAM	1169.6	113- 122CAM	96-105CAM	1818.9
2184.8	113-131CAM	96-114CAM	1656.3	132-146	n.a.	1834.8
2200.7	113- 131oxCAM	96- 114oxCAM	1342.7	140-151	123-134	1966.7
3095.2	113-139CAM	n.a.	2072.5	171-189	154-172	2593.8
3110.9	113- 139oxCAM	n.a.	1505.1	176-189	159-172	2614.7
1656.0	132-146	n.a.	2485.2	190-210	173-193	2944.8
2110.6	152-169	135-152	2501.1	190-210ox	173-193ox	3084.7
1505.1	176-189	159-172	2215.7	192-210	175-193	3101.1
1774.1	176-191	159-174	2231.4	192-210ox	175-193ox	3832.5
2485.2	190-210	173-193	1941.6	215-230	n.a.	3895.3
2501.1	190-210ox	173-193ox	1785.3	216-230	n.a.	3910.9
2215.7	192-210	175-193	1535.2	235- 247CAM	218- 230CAM	4642.0
2231.4	192-210ox	175-193ox	1571.9	253-266	236-249	
1941.6	215-230	n.a.				
1785.3	216-230	n.a.				
1571.9	253-266	236-249				
1334.8	292-302	275-285				

Tabella IV: Analisi di spettrometria di massa MALDI-TOFF dei frammenti peptidici ottenuti da digestione delle componenti CBNP α , CBNP β e CBNP γ . Per CBNP α e CBNP β sono indicate le due proteine (ssDBF e CRP1) che hanno mostrato il *match* migliore con i valori di massa calcolati per ciascun peptide.

confrontabili a quelli aspettati nel caso di una proteina di pollo, denominata ssDBF, ovvero single-stranded D box Binding Factor. SsDBF è una proteina di 302 amminoacidi, individuata nel 1995 tramite vaglio di una libreria d'espressione a c-DNA e proposta quale fattore di trascrizione epatospecifico per il gene apoVLDLII (Smidt *et al.*, 1995). Nello stesso articolo, gli autori riportano il sequenziamento di un altro clone, che è risultato differente dal precedente per la presenza di un esone aggiuntivo al carbossi-terminale, e codificante probabilmente per una isoforma della stessa proteina.

Dopo ssDBF, la proteina che mostra un match migliore è CRP1, con 17 e 15 peptidi rispettivamente in comune con CBNP α e CBNP β . CRP1 (Chicken RibonucleoProtein 1) è una proteina sommariamente descritta esclusivamente per sequenziamento del suo c-DNA dopo vaglio di una genoteca (Cvekl *et al.*, 1995).

Altre proteine aviarie hanno fornito un match inferiore a 6 peptidi, e non sono state quindi prese in considerazione.

Le sequenze di ssDBF e CRP1 (figura 21) permettono di ascriverle entrambe alla famiglia delle hnRNP di tipo A/B e sono identiche ad eccezione della regione N-terminale di circa 80 residui amminoacidici e degli amminoacidi in posizione 136 e 221 nella regione C-terminale.

Allo scopo di discriminare CBNP α e CBNP β da ssDBF o CRP1, alcuni dei peptidi ottenuti sono stati sequenziati.

I peptidi di massa corrispondente a 1328.5, 1571.9 e 1774.1 che in figura 21 vediamo sottolineati in blu, hanno mostrato avere una sequenza amminoacidica comune ad entrambe le proteine. Determinanti sono stati i frammenti di massa 1656.0 e 1785.3 che in figura vediamo sottolineati in rosso. La presenza dell'isoleucina in posizione 136 e della treonina in posizione 221 permettono con buona certezza di identificare in CBNP α e CBNP β due possibili isoforme di ssDBF.

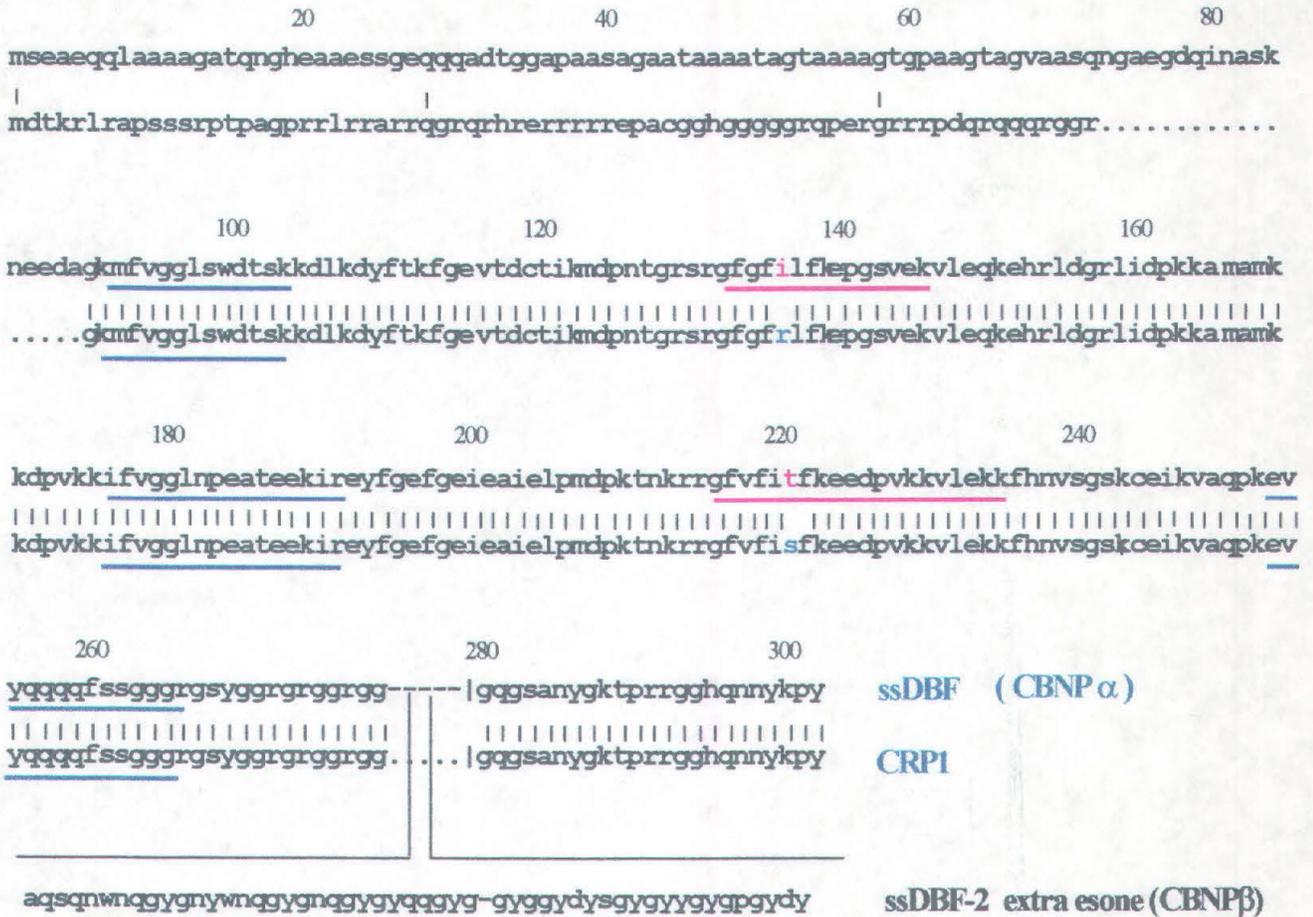


Figura 21: Confronto tra le sequenze delle proteine ssDBF e CRP1. In rosso e blu sono sottolineate le sequenze amminoacidiche dei peptidi di CBNP α che sono stati sequenziati

Da osservare inoltre che, CBNP α e CBNP β mostrano due peptidi di massa 8108.0 e 7363.5, rispettivamente, che sono ascrivibili ai peptidi N-terminali di ssDBF, mancanti del primo residuo di metionina e con il secondo residuo amminoacidico acetilato (vedere tabella IV).

A questo punto la possibile relazione strutturale tra CBNP α e CBNP β è stata desunta dalla determinazione della loro massa molecolare attraverso spettrometria di massa “*ion-spray*”.

Le tre proteine recuperate dalla resina di affinità sono state separate tramite cromatografia HPLC a fase inversa ed i campioni, come descritto nei “Materiali e Metodi”, sono stati sottoposti ad analisi di spettrometria di massa. Dalle frazioni contenenti CBNP γ l’analisi ISMS ha posto in evidenza solamente la presenza di frammenti di peso molecolare attorno ai 15 kDa, ma non è stato possibile individuare componenti di dimensioni maggiori come ci si sarebbe aspettato a giudicare dalla mobilità elettroforetica in SDS-PAGE di CBNP γ .

Al contrario dalla frazione cromatografica contenente CBNP α e CBNP β sono stati ottenuti spettri ISMS ben risolti, da cui è emersa evidente la presenza di una specie predominante la cui massa molecolare è risultata essere 31915 ± 2 Da e una specie meno abbondante di massa 37718 ± 4 Da. L’analisi ISMS e le relative deconvoluzioni sono riportate in Figura 22. Tali masse molecolari sono sensibilmente inferiori a quelle stimate dalla mobilità elettroforetica in SDS-PAGE, entrambe per circa 8 kDa. E’ questa una caratteristica riscontrata altre volte spesso nel caso di proteine basiche, quali gli istoni, per cui è possibile che la migrazione anomala di queste proteine sia legata ad una loro peculiare composizione amminoacidica.

Rilevante è l’osservazione che la differenza di massa molecolare tra CBNP α e CBNP β è di 5803 ± 4 Da, che corrisponde perfettamente al valore di massa

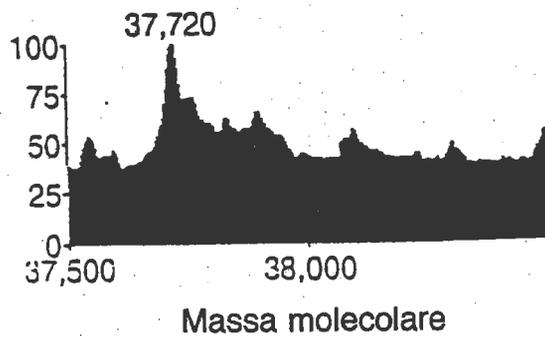
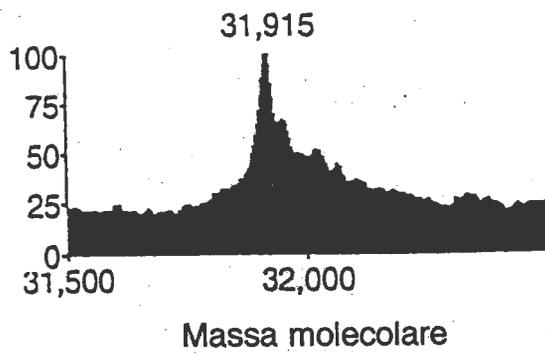
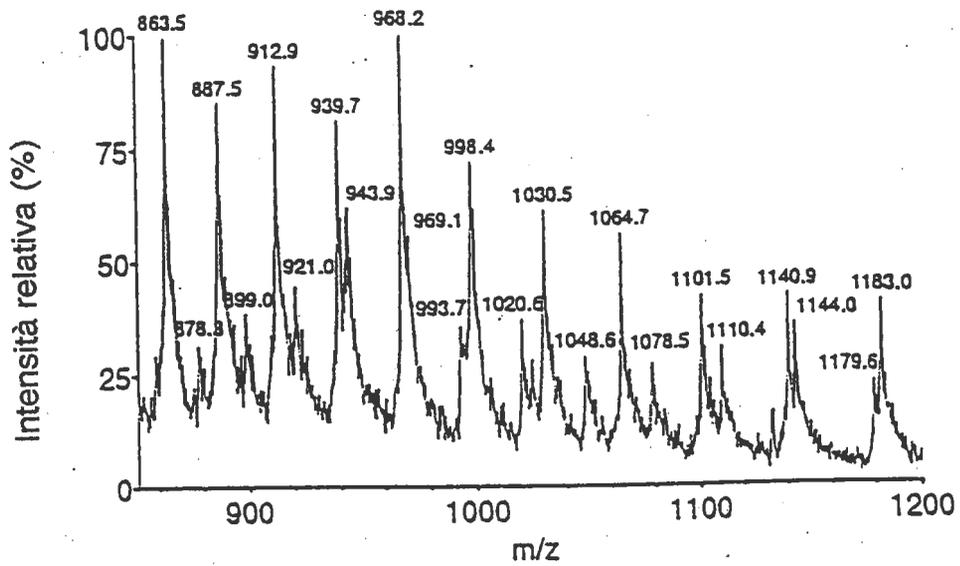


Figura 22: Analisi di spettrometria di massa Ion Spray (ISMS) della frazione cromatografica contenete CBNP α e CBNP e relative deconvoluzioni delle serie individuate

molecolare teorica calcolato per l'esone aggiuntivo dell'isoforma di ssDBF relativa al clone già descritto in letteratura (Smidt *et al*, 1995).

Un'ulteriore conferma all'identità di CBNP α e CBNP β è stata ottenuta grazie all'anticorpo policlonale anti-ssDBF gentilmente donato dal dott. Smidt dell'Università di Groninga che per primo ha identificato questa proteina. Saggi Western Blot (vedere "Materiali e metodi") condotti sia su estratti totale che sulle singole componenti purificate per affinità, hanno evidenziato chiaramente che l'anticorpo anti ssDBF policlonale da antisiero di coniglio, è in grado di riconoscere sia CBNP α che CBNP β ma non CBNP γ , a dimostrazione della diversa natura di quest'ultima proteina (Figura 23)

Analogamente saggi EMSA condotti incubando la componente CBNP α purificata o l'estratto totale con l'anticorpo, generano la chiara comparsa di segnali di *supershift* del complesso specifico da noi osservato, indicativi dell'esistenza di un'interazione tra l'anticorpo ed il complesso allo studio CBNP α -oligonucleotide sonda HTC4 (Figura24).

A questo punto i dati a nostra disposizione ci hanno permesso con buona certezza di identificare in CBNP α e CBNP β o due isoforme di ssDBF che differiscono per un extra-esone vicino all'N-terminale, o due proteine strettamente omologhe a ssDBF.

Per quanto riguarda CBNP γ , i dati preliminari ci avevano immediatamente fornito l'indicazione che dovesse trattarsi di una proteina ben distinta da CBNP α e β . Infatti, tentativi condotti in parallelo a quelli di CBNP α e CBNP β , per caratterizzarla attraverso MALDI-TOFF, non hanno fornito risultati soddisfacenti in quanto la ricerca in banche dati non ha rilevato l'esistenza di proteine ad alta omologia di sequenza con CBNP γ .

Per caratterizzarla si è quindi proceduto al sequenziamento diretto di alcuni dei peptidi ottenuti da digestione triptica della proteina. La sequenza di alcuni di

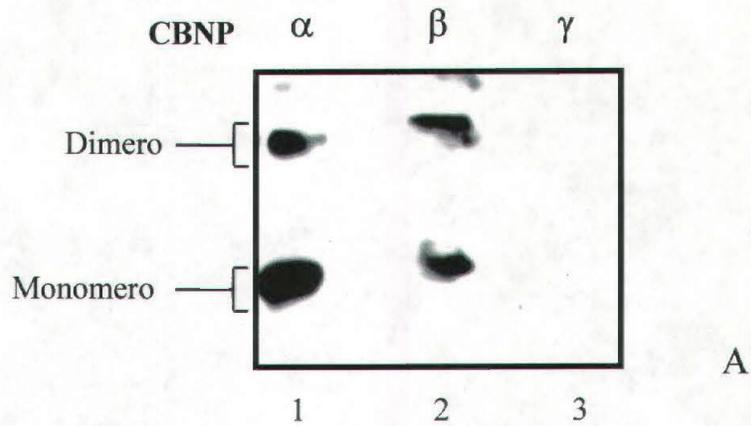


Figura 23:

Saggio Western blot delle singole componenti da eritrociti di pollo purificate per affinità. L'elettroforesi SDS PAGE è stata eseguita in condizioni non riducenti allo scopo di evidenziare la formazione di dimeri covalenti delle proteine.

- corsia 1, CBNP α ;
- corsia 2, CBNP β ;
- corsia 3, CBNP γ

Le bande sono state evidenziate usando un anticorpo policlonale anti ssDBF

Siero anti ssSDBF (μ l)	-	-	5	2	-	-
Siero pre-immune (μ l)	-	-	-	-	5	2

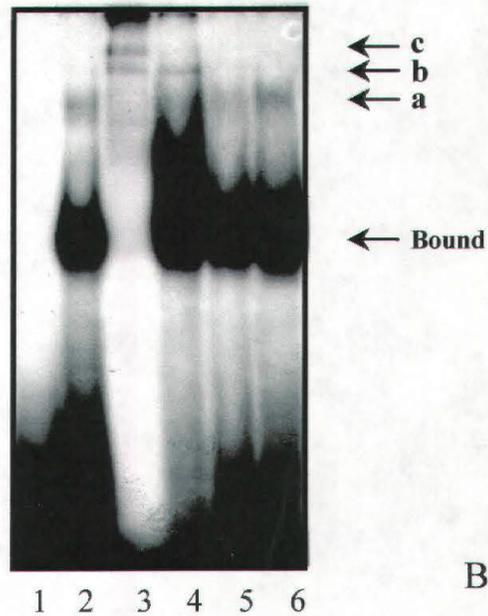


Figura 24:

Saggio di supershift per il complesso specifico tra CBNP α e la sonda HTC4.

I campioni contenenti CBNP α e HTC4 sono stati incubati in presenza (corsia 2), in assenza (corsie 3 e 4) di anticorpo policlonale da coniglio anti-ssDBF e, come controllo, in presenza di siero pre-immune (corsie 5 e 6).

Le frecce a lato della figura indicano la posizione delle bande relative al supershift e quella del complesso ritardato.

In corsia 1 è stato caricata la sonda HTC4 da sola.

questi, riportata in figura 25, è identica a due regioni conservate in diverse specie della proteina nucleare eterogea di tipo K (Matunis *et al.*, 1992)). In particolare in figura è riportata la posizione delle due regioni nella hnRNP K umana, mentre non è nota l'omologa hnRNP K di pollo.

Per l'hnRNP K umana è stato osservato un ruolo di attivatore trascrizionale del gene *c-myc*, attraverso la sua interazione con l'elemento polipirimidinico presente nel promotore di questo gene (Michelotti *et al.*, 1995). Nel caso del gene *agp* sembrerebbe invece assumere un ruolo di repressore, forse prevenendo l'interazione dell'attivatore C/EBP b con la sua sequenza target (Miau *et al.*, 1998)). Comunque l'ipotesi che CBNP γ possa essere effettivamente l'hnRNP K di pollo, è suffragata da quanto riportato da un recentissimo articolo in cui è stata descritta la capacità della hnRNP K umana di interagire con il filamento telomerico a blocchi di citosine dei vertebrati (Lacroix *et al.*, 2000).

Analizzando i dati fin qui ottenuti mediante le tecniche EMSA, di UV crosslinking *in situ* e cromatografia in fase inversa, emergono alcune rilevanti osservazioni sul comportamento delle CBNP da noi isolate.

La prima considerazione riguarda la forte tendenza all'aggregazione che tali componenti hanno mostrato nelle diverse fasi della purificazione. Ad esempio, sebbene CBNP γ venga covalentemente legata alla sonda nucleotidica dopo irraggiamento UV, essa viene rinvenuta nella banda EMSA più lenta e meno specifica solo in copresenza di CBNP α (vedi figura 19) mentre, singolarmente, una volta isolata dalle altre componenti, non dà segnali EMSA (vedi figura 20). Inoltre, nel tentativo di separare le CBNP tramite cromatografia a fase inversa, CBNP α e CBNP β coeluiscono sempre in un unico picco, qualsiasi siano le condizioni sperimentali scelte, mentre CBNP γ viene eluita sempre accompagnata da quantità minori delle altre due. Si è pure osservata la comparsa di piccole quantità di specie

CBNP γ

Massa

1098.7 gsdfdcelr

1195.3 nlplppppppr

1505.3

1519.7

1554.4

1596.1

1711.7

1818.9

1834.8

1966.7

2593.8

2614.7

2944.8

3084.7

3101.1

3832.5

3895.3

3910.9

4642.0

meteqpeetfpntetn²⁰gefgkrpaedmeeeq⁴⁰dfkrsrntdemvelrillqsknagav
ig⁶⁰kggknikalrtdyna⁸⁰svspds¹⁰⁰gperils¹²⁰isadietigeilkkiptleeglql
psptatsqlplesdavecln¹⁴⁰yqhykgsdfdc¹⁶⁰elrl¹⁸⁰l²⁰⁰h²²⁰qslaggiigvkgakikelr
entq²⁴⁰ttik²⁶⁰lfqeccphst²⁸⁰dr³⁰⁰v³²⁰l³⁴⁰iggk³⁶⁰pd³⁸⁰r⁴⁰⁰v⁴²⁰vec⁴⁴⁰ik⁴⁶⁰il⁴⁸⁰dlisespikgragpydpn
fydetydyggftmmfddrrgrpv²⁴⁰g²⁶⁰fp²⁸⁰mr³⁰⁰gr³²⁰gg³⁴⁰fdr³⁶⁰m³⁸⁰pp⁴⁰⁰g⁴²⁰g⁴⁴⁰rp⁴⁶⁰m⁴⁸⁰pp⁵⁰⁰s⁵²⁰rr⁵⁴⁰dy⁵⁶⁰dd⁵⁸⁰ms⁶⁰⁰p
rrgppppppg³⁰⁰g³²⁰g³⁴⁰gs³⁶⁰ram³⁸⁰l⁴⁰⁰pl⁴²⁰pp⁴⁴⁰pp⁴⁶⁰pr⁴⁸⁰gg⁵⁰⁰dl⁵²⁰may⁵⁴⁰dr⁵⁶⁰g⁵⁸⁰rp⁶⁰⁰gd⁶²⁰ry⁶⁴⁰d⁶⁶⁰gm⁶⁸⁰v⁷⁰⁰g⁷²⁰fs⁷⁴⁰ad
etwdsaidtwspse³⁶⁰w³⁸⁰q⁴⁰⁰mayep⁴²⁰q⁴⁴⁰gg⁴⁶⁰sg⁴⁸⁰yd⁵⁰⁰sy⁵²⁰ag⁵⁴⁰gr⁵⁶⁰gs⁵⁸⁰yg⁶⁰⁰dl⁶²⁰gg⁶⁴⁰pi⁶⁶⁰itt⁶⁸⁰qv⁷⁰⁰tip⁷²⁰kd⁷⁴⁰la
g⁴⁰⁰si⁴²⁰ig⁴⁴⁰k⁴⁶⁰gg⁴⁸⁰q⁵⁰⁰rik⁵²⁰q⁵⁴⁰ir⁵⁶⁰h⁵⁸⁰es⁶⁰⁰gas⁶²⁰ik⁶⁴⁰ide⁶⁶⁰ple⁶⁸⁰gs⁷⁰⁰ed⁷²⁰ri⁷⁴⁰it⁷⁶⁰it⁷⁸⁰gt⁸⁰⁰qd⁸²⁰qi⁸⁴⁰q⁸⁶⁰na⁸⁸⁰q⁹⁰⁰yll⁹²⁰q⁹⁴⁰ns⁹⁶⁰vk
qy⁴⁶⁰ad⁴⁸⁰veg⁵⁰⁰f

hnRNP K umana

Figura 25: Valori di massa e sequenze di due dei peptidi ottenuti dalla digestione triptica di CBNP γ . E' inoltre riportata la sequenza della hnRNP K umana, con sottolineate le sequenze relative ai due peptidi identificati anche in CBNP γ

a massa molecolare doppia che appaiono nelle elettroforesi SDS-PAGE in condizioni non riducenti, suggerendo la formazione di omodimeri covalenti (vedi figura 23). Questo schema di comportamento può essere attribuito alla tendenza che caratterizza le hnRNP ad aggregare in complessi omotipici o in associazione con altri tipi di hnRNP. Questa proprietà sembra dipendere esclusivamente dal dominio N-terminale ricco in glicine delle hnRNP, il quale presenta un periodico susseguirsi di residui aromatici interspersi tra glicine e aminoacidi polari o carichi positivamente, a formare una catena anfipatica e flessibile. L'associazione tra le proteine risulterebbe quindi dall'interazione idrofobica tra due catene di due distinte hnRNP (Cartegni *et al.*, 1996).

La seconda osservazione riguarda in particolare la componente da noi denominata CBNP β che, catturata per affinità all'oligonucleotide a sequenza telomerica, non fornisce però un addotto covalente con la sonda HTC4 in *UV crosslinking* e, una volta ottenuta pura per elettroeluzione, non dà luogo a segnali EMSA. Tale fatto, come già osservato, potrebbe suggerire che essa complessi con altre proteine, probabilmente con CBNP α , ma non leghi direttamente il DNA telomerico. Se ne deduce che la presenza dell'esone aggiuntivo in CBNP β , unico elemento che la differenzia da CBNP α , potrebbe sopprimere la capacità di legame con il DNA ma non quella di interagire con CBNP α stessa ed eventualmente altre hnRNP.

3.7 Espressione di ssDBF ricombinante

Il c-DNA per la ssDBF ci è stato gentilmente messo a disposizione dall Dott. Smidt dell'Università di Groninga, clonato all'interno del vettore pGem7zf(+). Non trattandosi di un vettore d'espressione, si è reso necessario il subclonaggio del c-

DNA all'interno del vettore pT7.7 usato per l'espressione della proteina ricombinante in *E. coli*.

Il c-DNA è stato amplificato dal vettore pGem con la tecnica della PCR utilizzando i due primers indicati nei "Materiali e metodi". La scelta dell'utilizzo del primer al 5', NdeFw 92, che appaia nella posizione corrispondente all'amminoacido 22 di ssDBF, si è resa necessaria per il fatto che i primi 21 residui della proteina sono codificati da codoni ricchi in basi GC che costituiscono una regione difficilmente amplificabile per PCR a causa di una forte tendenza a formare strutture secondarie termodinamicamente molto stabili. In figura 26 è riportata la sequenza nucleotidica di ssDBF e la corrispondente sequenza amminoacidica; in colore è evidenziata la porzione di c-DNA amplificata e clonata nel vettore pT7.7.

Dopo clonaggio nel vettore dell'inserto amplificato, il costrutto è stato utilizzato per trasformare cellule di *E. coli* TOP10 competenti e l'inserto è stato sequenziato per verificare il rispetto della cornice di lettura. La proteina ricombinante risulta essere costituita da 281 residui appartenenti alla proteina ssDBF più un residuo addizionale di metionina derivante dal clonaggio, ha un peso teorico calcolato di 29925 Da ed un punto isoelettrico di 9.08.

A questo punto si è passati all'espressione su larga scala della proteina ricombinante in cellule BL21 di *E. coli* ed alla sua estrazione come descritto in "Materiali e metodi". L'estratto batterico, ottenuto prima e dopo l'induzione con IPTG, è stato analizzato per la presenza della proteina ricombinante ed il profilo elettroforetico in SDS PAGE di tali estratti ha evidenziato l'avvenuta espressione di ssDBF (Figura 27A).

Per conferma dell'identità della proteina è stata effettuata un'analisi di immunoreattività attraverso un saggio *Western Blot* sugli estratti dei batteri non indotti ed indotti. A tale scopo è stato utilizzato l'anticorpo policlonale di coniglio già utilizzato per i saggi di immunoreattività sulle componenti purificate per affinità a partire dagli eritrociti di pollo. Come si può vedere in figura 27B, l'anticorpo è

1 gaattccgcctctgcgggcggggcagcatgtccgaagcggagcagcagttggcggccgc
M S E A E Q Q L A A A 11
61 cgccggcgccaccagaatggacacgaagcggctgagagctccggggagcagcagcaggc
A G A T Q N G H E A A E S S G E Q Q Q A 31
121 cgacaccggggggcccgggccttcggcggggcggcgacagcggcggcagcgacagc
D T G G A P A A S A G A A T A A A A T A 51
181 gggaaacggcggcggcggcgggagccggccctgcccggggagcggcgggggtggcggccag
G T A A A A G A G P A A G T A G V A A S 71
241 ccagaacggggcgaaggcgaccagatcaacgccagcaagaacgaggaggacgoggggaa
Q N G A E G D Q I N A S K N E E D A G K 91
301 aatgttttgggtggcctcagttgggatacaagcaaaaaagacttgaaagactacttcac
M F V G G L S W D T S K K D L K D Y F T 111
361 caagtttggtaggtaactgactgtaacgataaagatggaccctaacacgggaagatccag
K F G E V T D C T I K M D P N T G R S R 131
421 aggttttgatttatactcttcaaagaacctgggagtggtgaaaaggttctggaacagaa
G F G F I L F K E P G S V E K V L E Q K 151
481 agaacacaggctagatggacgactaattgacoccaagaaggccatggcaatgaaaaagga
E H R L D G R L I D P K K A M A M K K D 171
541 tccagtgaagaaaatatttgggtggacttaaccccgaagccacagaagagaaaatcag
P V K K I F V G G L N P E A T E E K I R 191
601 ggaatactttggcgagtttggagagattgaagcaattgaacttccaatggatccaaagac
E Y F G E F G E I E A I E L P M D P K T 211
661 caacaaaaggagggggttgggttcatcactttcaaggaagaagatccagtgaagaaagt
N K R R G F V F I T F K E E D P V K K V 231
721 ttggagaagaaattccataacgtcagtggaagcaagtgogagataaaggtagcacagcc
L E K K F H N V S G S K C E I K V A Q P 251
781 aaaagaagtataaccagcagcaacagttcagtagtggtggaggaagaggttagctatggagg
K E V Y Q Q Q Q F S S G G G R G S Y G G 271
841 aagaggcagaggtggaagaggcgggtggtaaggcagtgcaaatatgggaagacaccaag
R G R G G R G G G Q G S A N Y G K T P R 291
901 gcgtgggtggccatcagaataactacaagccatattgatcaaaacttattcaggcctggcag
R G G H Q N N Y K P Y - 302
961 tggcacagttgatgcctacaagcactggatattccacaagttgaagacgggaatactgac
1021 ttgtttaoctgaagataataggacttogggggtatttctcagagaaaaatataaatttta
1081 atataatttcataagctttggagggttagcttgtctttagcttcagtgatatcaaaacttc
1141 tttgagtaggagctagatagaaagctgttttagtttttgccagcatalgtaaaacaagact
1201 tttctatctgcattgtagattgtttgtggacacttgccagattcttgtgctttctgcagtc
1261 attctggggactgaactctcattgatatogatagaggttccatgaacaggacatctgcag
1321 gataccaagtgtagaagcccacgggtgtgggtttggagtagtgaactgcttttagttggaga
1381 tgactgggggagatgggaagagcaactgtacactaaggggaacaattacagatcttcttcc
1441 cagtgacatcaatgagagatttgccttgcacctccagagaagaattagctttaatttgg
1501 gacagcaaaaggggtttttcaaaagcattaaaagccagtgaggagctctttcagcagctcct
1561 ttcgcttttaaaaacctcacacccaaaagtggttagctaacgtcatgaaatggaagtaaat
1621 aaaatgcattttcattgcaaa
1681 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaggaattc

Figura 26: Sequenza nucleotidica del c-DNA di ssDBF e corrispondente sequenza amminoacidica. In viola è evidenziata la porzione del c-DNA che è stata subclonata nel vettore pT7.7 ed in blu la corrispondente sequenza amminoacidica da esso codificata.

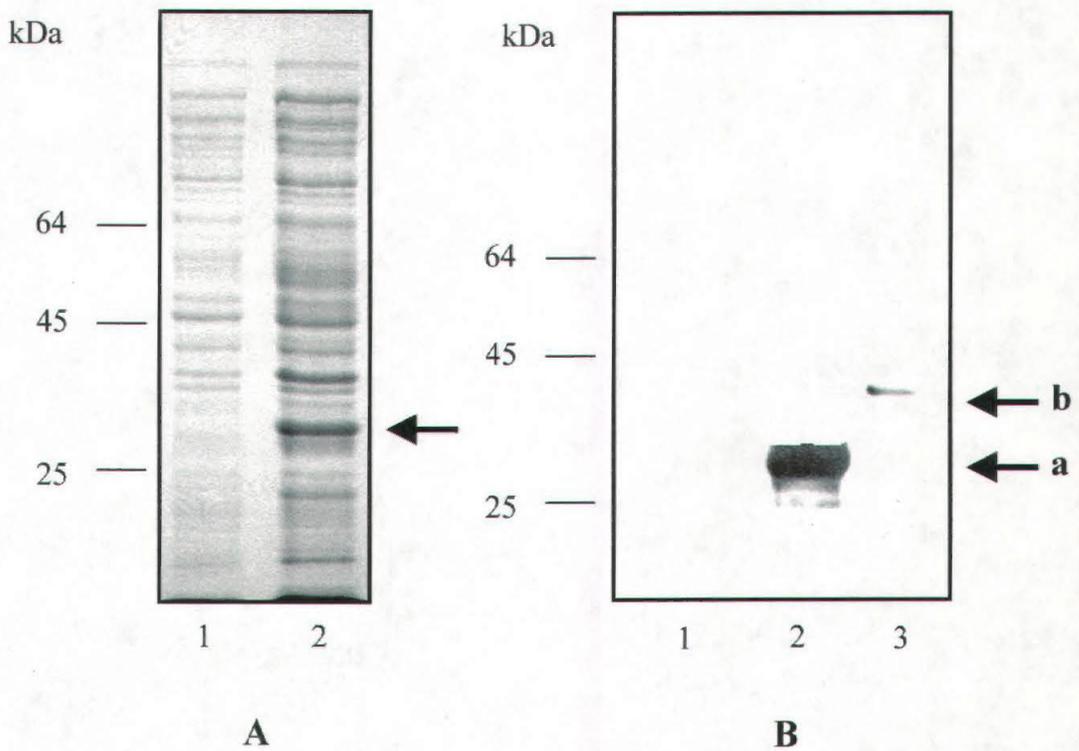


Figura 27:

- A)** *Gel di SDS PAGE al 12%. Il lisato batterico di cellule indotte e non con IPTG è stato analizzato su gel in condizioni denaturanti e riducenti.*
- corsia 1, lisato batterico di BL21 non indotto;
 - corsia 2, lisato batterico di BL21 dopo induzione con IPTG per 5 ore;
- La freccia indica la posizione della proteina ricombinante*
- B)** *Analisi in Western blot dell'espressione di ssDBF ricombinante.*
- corsia 1, lisato batterico di BL21 non indotto;
 - corsia 2, lisato batterico di BL21 dopo induzione;
 - corsia 3, estratto nucleare di cellule BM2 di pollo.
- Per la rilevazione della proteina è stato utilizzato l'anticorpo policlonale specifico anti-ssDBF.*
- Con a viene indicata la posizione della proteina ricombinante, con b la posizione della sua forma endogena in cellule BM2*

in grado di riconoscere la banda di espressione di ssDBF ricombinante solo nel lisato batterico dopo induzione con IPTG. Il peso molecolare apparente di tale banda corrisponde esattamente al peso della banda d'induzione indicata nella figura precedente. Come controllo della specificità dell'anticorpo, nella corsia 3 del gel è stato caricato un estratto nucleare di cellule di una linea mieloide di pollo (BM2) che, come da noi accertato, esprime costitutivamente la proteina ssDBF. Si può osservare come in entrambi i casi, cioè sia nel caso di ssDBF ricombinante sia nel caso di ssDBF endogena, il peso molecolare apparente in SDS PAGE risulta essere inferiore a quello teorico. Questo dato conferma quanto già osservato nel caso di ssDBF purificata dagli eritrociti e può essere ascritto alle caratteristiche fisico-chimiche della proteina stessa, come ad esempio il suo pI piuttosto basico.

Per accertare la funzionalità di ssDBF nella sua forma ricombinante ma soprattutto la specifica attività di *binding* alla sequenza telomerica, si è provveduto a saggiarne l'attività in EMSA utilizzando come sonda HTC4 marcato radioattivamente. Come risulta chiaro dalla figura 28, la proteina relativa al complesso che si forma solo nel caso dell'estratto da batteri indotti (corsie 2 e 3), ascrivibile presumibilmente a ssDBF ricombinante, mostra una notevole affinità per la sonda telomerica, non subendo alcun effetto di competizione da parte di un notevole eccesso molare (500 volte) di DNA di *E. coli* a filamento singolo. Si può osservare inoltre la formazione di un complesso a minore mobilità elettroforetica poco specifico, presente anche nei batteri non indotti ed efficacemente competuto dall'eccesso di DNA aspecifico (corsie 4 e 5).

Sembra quindi che la delezione dei primi 21 residui amminoacidici all'N-terminale della proteina ssDBF, così come l'assenza nella sua forma ricombinante delle modificazioni post-traduzionali presenti invece nella forma endogena, non compromettano drammaticamente la sua affinità di legame verso il DNA telomerico. Attualmente comunque non si hanno a disposizione dati sperimentali relativi all'effettivo grado di specificità dell'interazione tra ssDBF ricombinante e la

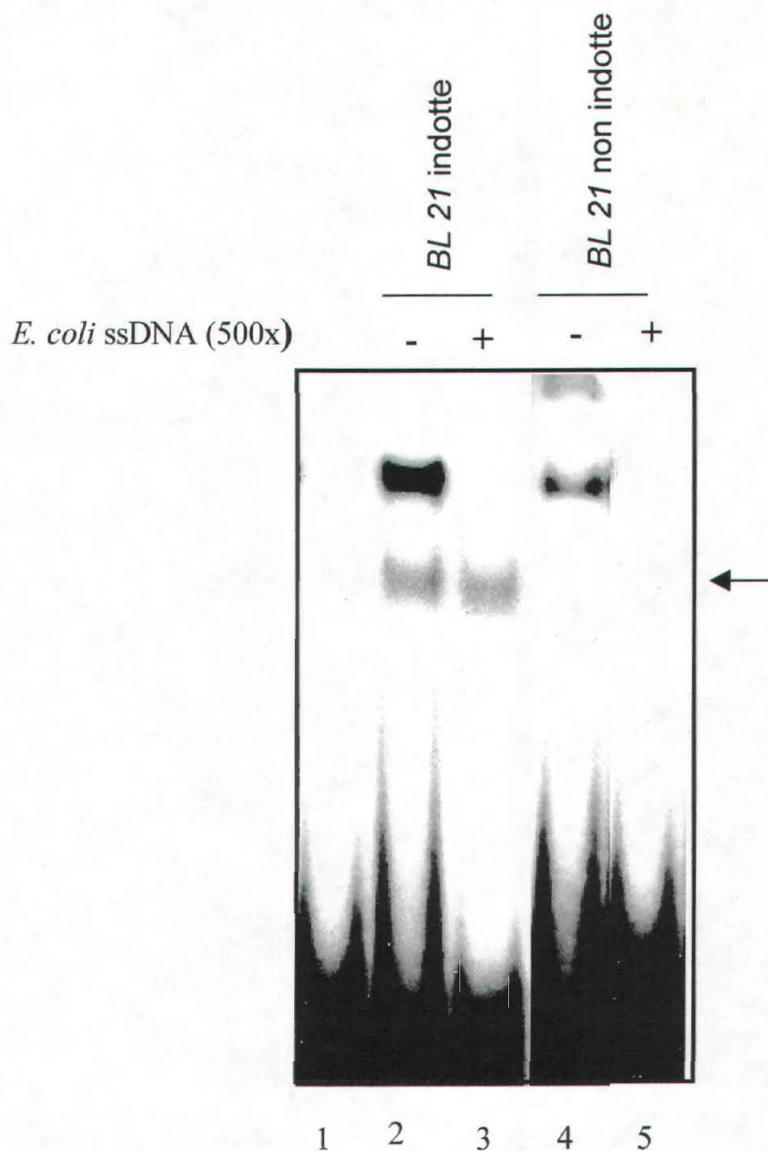


Figura 28:

- Saggio EMSA dell'attività di legame della proteina ricombinante ssDBF con la sonda HTC4. 20 microlitri di lisato batterico sono stati incubati con 0.2 picomoli di sonda:*
- corsia 1, oligonucleotide HTC4 marcato;
 - corsia 2, lisato batterico di cellule indotte con IPTG, in assenza di DNA competitore;
 - corsia 3, lisato batterico di cellule indotte, in presenza di un eccesso di DNA di *E.coli* a filamento singolo;
 - corsia 4, lisato batterico di cellule non indotte in assenza di DNA competitore;
 - corsia 5, lisato batterico di cellule non indotte in presenza di DNA competitore ,

sonda telomerica, utili a valutare in quale e quanta misura modifiche post-traduzionali o possibili isoforme della proteina possano influenzare tale specificità.

3.8 CBNP α e ssDBF: una proteina, più funzioni ?

L'identificazione di CBNP α con la proteina ssDBF, ascrivibile per la sua sequenza alle hnRNP di tipo A/B, solleva vari ordini di considerazioni.

Una argomentazione riguarda il fatto che una proteina classificata in un gruppo, le hnRNP, identificato e caratterizzato sulla base della capacità di legare l'RNA eterogeneo nucleare, leghi in realtà, ed in modo notevolmente sequenza-specifico, un DNA. In realtà il quadro formato da questa grande famiglia di proteine è molto più complesso e variegato. L'iniziale definizione di natura operativa di hnRNP (heterogeneous nuclear RiboNucleoProtein, o più esattamente heterogeneous nuclear RiboNucleoParticle protein), era sorta sulla base della loro capacità di venire covalentemente legate tramite agenti di crosslinking all'RNA eterogeneo nucleare. Sulla base della massa molecolare apparente in SDS PAGE le hnRNP sono state ulteriormente ripartite in numerosi sottogruppi chiamati inizialmente con le lettere dell'alfabeto, a cominciare dalle hnRNP A per le più piccole (a più alta mobilità elettroforetica), fino alle hnRNP U per le più grandi (a minore mobilità elettroforetica) (tabella V). Come si può osservare, nonostante le hnRNP attualmente note siano molto numerose e ascrivibili a una ventina di tipi diversi almeno, la loro caratterizzazione funzionale è nella maggior parte dei casi parziale, ipotetica o assente. In molti casi la conoscenza deriva dal sequenziamento parziale o totale di un cDNA. Inizialmente si era ipotizzato che la loro principale funzione fosse strutturale ed organizzativa, ad esempio nell'impaccamento dell'hnRNA nascente in strutture globulari ribonucleoproteiche la cui funzione era quella di

Principali hnRNP	Domini di legame all' RNA	Funzioni proposte
A1/A1	2X RBD, RGG	Splicing (repressore esonico) Trasporto dell'mRNA Biogenesi dei telomeri
A2/B1	2X RBD, RGG	Splicing Veicolamento dell'mRNA
C1/C2	1X RBD	Impaccamento dei trascritti Splicing Stabilità dell'mRNA Ritenzione nucleare
D1/D2 (AUF p42,p45)	2X RBD, RGG	Trascrizione Turnover dell'mRNA
E1/E2 (PCBP-1, α CP1) (PCBP-2, α CP2)	3X KH 3X KH	Silenziamento traduzionale Stabilità dell'mRNA
F	3X RBD	Splicing (enhancer intronico) Interazione con il complesso proteico che si lega al cap
G	1X RBD, RGG	
H	3X RBD	Splicing (enhancer intronico)
H'(DSEF-1)	3X RBD	Poliadenilazione
I (PTB)	4X RBD	Splicing (repressore/enhancer intronico) Poliadenilazione
K	3X KH, RGG	Trascrizione Silenziamento traduzionale
L	4X RBD	Stabilità dell'mRNA Esportazione del mRNA
M	3X RBD	Splicing Risposta a shock termico
P2 (TLS/FUS)	1X RBD, RGG	Oncogenesi
R	3X RBD, RGG	
U	1X RGG	Ritenzione nucleare
HAP/SAF-B	1X RBD	Trascrizione Risposta a shock termico
hnRNP secondarie		
A0	2X RBD, RGG	Splicing
CUG-BP (hNab50)	3X RBD	Splicing (enhancer intronico) Distrofia miotonica
HuR	3X RBD	Stabilità dell'mRNA Trasporto dell'mRNA

TABELLA V: hnRNPs e loro funzioni

proteggere l'RNA dalla degradazione da parte delle nucleasi e di garantire che i processi post-trascrizionali avvenissero correttamente. Recentemente è stato suggerito un ruolo più dinamico e diversificato per le hnRNP, che possono quindi essere viste come un sottoinsieme dei fattori che processano il pre-mRNA (Dreyfuss *et al.*, 1993; Swanson, 1995). Tutte le hnRNP hanno una generica affinità per l'RNA a singolo filamento. Ci sono però numerose evidenze sperimentali che indicano che molte di queste possiedono una specificità di legame per determinate sequenze. Ad esempio hnRNP P preferisce il poly(A); hnRNP E, F, H ed M legano il poly(G), hnRNP J e K legano il poly(C) e hnRNP C ed M legano il poly(U) (Dreyfuss *et al.*, 1993). Queste proteine riconoscono particolari regioni nel pre-mRNA e, attraverso specifiche interazioni, contribuiscono ad indirizzare i fattori di processamento verso determinati siti (Matunis *et al.*, 1993). Altre ancora sembrerebbero coinvolte nel trasporto dell'RNA maturo al citoplasma (Visa *et al.*, 1996; Michael *et al.*, 1995) o nella modulazione dello splicing alternativo (Matunis *et al.*, 1993). L'attribuzione di questa ampia gamma di ruoli alle hnRNP, fa pensare che i complessi da esse formati siano strutture estremamente dinamiche, soggette a numerose transizioni strutturali che servono durante la biogenesi, il processamento e il trasporto dell'mRNA (Krecic & Swanson, 1999).

Ma ciò che interessa particolarmente ai fini del nostro studio è che parecchie di queste proteine sono state osservate legare preferenzialmente il DNA a filo singolo, molto spesso in maniera altamente sequenza specifica (Krecic *et al.*, 1999), facendo pensare a funzioni diverse dal processamento dell'RNA trascritto primario. Per esempio, ricordiamo il ruolo svolto dall'ormai ampiamente citata hnRNP K quale fattore di trascrizione (Michelotti *et al.*, 1995; Miao *et al.*, 1998), oppure l'hnRNP A1 e l'hnRNP A2/B1, per le quali è stata dimostrato un ruolo nella gestione dei telomeri per la prima (Dallaire *et al.*, 2000) e la capacità di legare *in vitro* il motivo telomerico ripetuto a blocchi di guanine (TTAGGG)_n per la

seconda (McKay&Cooke, 1992). In questa prospettiva ssDBF alias CBNP α verrebbe a costituire un ulteriore caso di interazione specifica tra una hnRNP ed il DNA telomerico, in questo caso però a sequenza polipirimidinica.

La seconda, ovvia argomentazione, si limita strettamente al confronto tra l'attività di CBNP α , descritta nel nostro laboratorio, e quella di ssDBF, descritta da Smidt *et al* (1995).

Le due attività non appaiono strettamente correlate, così come i possibili ruoli funzionali che ne conseguirebbero.

Dal momento che univoche evidenze in tal senso sia *in vivo* che in culture cellulari non sono state finora raggiunte, non è ancora possibile dare un'interpretazione certa in merito. Va tuttavia notato che ssDBF è stata identificata tramite vaglio di una genoteca di c-DNA da fegato, mentre CBNP α è stata isolata con procedure di tipo biochimico da eritrociti. Potrebbe essere che questa proteina abbia un duplice (o addirittura multiplo) ruolo, modulato dal tipo di tessuto in cui è espressa, nonché dall'abbondanza relativa delle due isoforme e da modifiche post-traduzionali eventualmente differenziate. Evidenze sui dati di spettrometria di massa di CBNP α , indicano infatti chiaramente nella sua sequenza amminoacidica la presenza di molti siti di fosforilazione e metilazione.

A sostegno di questa ipotesi bisogna ricordare che la complessità della famiglia delle hnRNP è notevolmente accresciuta proprio da modifiche post-trascrizionali e post-traduzionali a partire dallo stesso trascritto primario. Molte hnRNP risultano da *splicing* alternativo a partire da un unico precursore (Buvoli *et al.*, 1990; Burd *et al.*, 1989; Biamonti *et al.*, 1994), ad esempio un'isoforma della hnRNP A1, chiamata A1^B, ha un'inserzione di 52 amminoacidi nel dominio ausiliario ricco in glicine (Buvoli *et al.*, 1990); hnRNP A2 e B1 sono prodotti di uno *splicing* alternativo dello stesso trascritto; hnRNP C1 e C2 sono identiche, eccetto per un'inserzione di 13 amminoacidi nella C2 (Burd *et al.*, 1989; Biamonti *et al.*,

1994). Della hnRNP D0 sono note quattro isoforme, ciascuna caratterizzata dalla presenza o assenza di uno o due esoni aggiuntivi (Wagner et al., 1998).

Sono stati inoltre descritti tre tipi di modificazioni post-traduzionali: fosforilazioni di serine e di treonine, metilazione di arginine e glicosilazione, che hanno mostrato avere una notevole influenza nella modulazione della funzione della proteina. Grazie alle numerose isoforme (generate da splicing alternativo, *editing* e/o da modifiche post-traduzionali) la cellula possiede un vasto repertorio di fattori proteici ciascuno con un'attività specifica talvolta tessuto specifica. Ad esempio, l'hnRNP A1, la più studiata di questo gruppo, può costituire un caso paradigmatico, essendo stati evidenziati per essa diversi ruoli funzionali. Si ritiene che essa intervenga in vari fenomeni di splicing alternativo (Mayeda *et al.*, 1998), nel processo di trasporto dell'mRNA (Nakielny & Dreyfuss, 1997) e nella biogenesi del telomero (particolarmente importante nel contesto della nostra ricerca).

3.9 Le proteine hnRNP di tipo A/B

La ricerca in banche dati proteiche (Swissprot e PIR) e nucleotidiche (GenBank e EMBL) di proteine aventi similarità significative con la sequenza amminoacidica di ssDBF (alias CBNP α) ha permesso di individuare un insieme di proteine e/o cDNA strettamente correlati. Le sequenze amminoacidiche relative sono riportate in Tabella VI. Una è la già citata CRP1 da pollo (Cvekl *et al.*, 1995), che differisce totalmente da ssDBF nella porzione N-terminale soltanto a causa di due mutazioni di tipo *frameshift* nel suo c-DNA, oltre che di due sostituzioni nel resto della sequenza. Tutte le altre riguardano proteine o relativi c-DNA provenienti da uomo, topo, ratto e gatto, anche queste classificabili nel gruppo delle hnRNP di tipo A/B.



Tabella VI: Comparazione delle sequenze di cinque hnRNP di tipo A/B correlate.

- | posizione dell'esone aggiuntivo
- § sostituisce g all'inizio dell'ultimo esone nelle forme contenenti l'esone aggiuntivo;
- ? residuo non noto
- * Leverrier, S. *et al.*, non pubblicato, Gen Bank, numero di accesso AJ238855
- **GenBank, numero di accesso AJ238854
- # 72 residui all'estremità N terminale di CRP1 (Cvekl *et al.*, 1995); i residui dal 73 al 285 sono identici ai residui dal 90 al 302 di ssDBF tranne per la sostituzione di una isoleucina con una arginina in posizione 136 di ssDBF e di una treonina con una serina in posizione 221 di ssDBF.
- 73 residui all'estremità N terminale di PRM10 (GenBank accession n. AF108653), dal residuo 74 è identica ad hnRNP 38 di ratto a partire dal residuo 61.

ssDBF	pollo		no consensus
CBF-A	topo		
A/B 38	ratto (*)		
DBP40	gatto		
A/B type umano			sostituzione

ssDBF-2, extra esone		delezione
A/B 40 rat, extra esone(**)		
ABBP1 umana extra esone		

mdtkrlrapgsssrptpagprlrrarrqgrqhrerrrrrrepacgghgggrqpergrrrpdqrqqqrggr # CRP1

mtwlsdqrlrvkcrlscwvrlfyfwccweageveavfrcssselpnwscdyrlkvpnqknggrgsrpsgnqng □ PRM10

Le loro sequenze risultano chiaramente correlate in tutta la loro estensione, con l'eccezione della porzione N-terminale di PRM10, una sequenza dedotta da c-DNA di ratto e riportata unicamente in banca dati (GenBank accession n. AF108653), di cui null'altro è noto.

Dall'esame di queste sequenze, emergono essenzialmente tre osservazioni:

a) c'è una somiglianza molto stretta nella porzione C-terminale di queste proteine, che riguarda circa l'80% dell'intera sequenza, mentre circa 60 residui N-terminali delle sequenze da mammifero risultano pure notevolmente correlati tra loro, ma in misura decisamente minore, seppure ancora riconoscibile, con la sequenza ssDBF. Nell'insieme la sequenza di ssDBF è quella che più diverge dalle altre riportate in tabella, coerentemente con la maggiore distanza filogenetica che dalle altre la separa.

b) nei casi della sequenza umana e di ratto, vengono riportate due isoforme che differiscono l'una dall'altra per l'assenza o la presenza di un esone aggiuntivo chiaramente omologo a quello osservato nel secondo clone di ssDBF (ovvero CBNP β), anche per quanto riguarda la sua posizione. Comunque l'esame delle banche dati rileva che simili sequenze si trovano anche in altre hnRNP, meno strettamente correlate con la sequenza di ssDBF.

c) per alcune delle proteine in oggetto è stato proposto un ruolo funzionale: per ssDBF, come già riportato sopra, quello di fattore di trascrizione epatospecifico (Smidt *et al.*, 1995), per la proteina murina CBF-A (e verosimilmente per la hnRNP 38 di ratto, che le è altamente omologa), quello di fattore di trascrizione specifico del tessuto muscolare (Kamada *et al.*, 1992), nonché regolatore trascrizionale per il gene Ha-ras (Mikheev *et al.*, 2000) e per la ABBP1 (ApoBec-1-Binding Protein 1) umana quello di fattore di editing (Lau *et al.*, 1997), mentre per la proteina DBP40 è stata solo osservata la sua capacità di interagire con la sequenza 5'-terminale del virus della panleucopenia felina (Wang *et al.*, 1999). Non si sa nulla a tutt'oggi della funzione della hnRNP umana di tipo A/B (Khan *et*

al., 1991), corrispettiva di ABBP1 senza l'esone aggiuntivo, né di CRP1 e PRM 10. Appare chiaro quindi come molto spesso alla stretta omologia di sequenza non corrisponda un'analogia di funzione.

Le differenze rilevabili nelle sequenze di queste proteine, come le differenze nelle loro funzioni, qualora confermate, potrebbero essere la conseguenza di una diversificazione ascrivibile da un lato alle "normali" mutazioni, sovrapposte eventualmente ad episodi di duplicazione genica e ricombinazione, susseguitesesi nel corso della filogenesi, dall'altro a modulazioni della loro attività attraverso lo sviluppo di isoforme, talvolta tessuto-specifiche, dovute a fenomeni di splicing alternativo, *editing* o modifiche post-traduzionali. In questa prospettiva andrebbe tenuta in considerazione la possibilità che la decina di casi relativi alle cinque specie riportate nelle banche dati costituiscano solo un campione limitato, e casuale, di una famiglia più numerosa di hnRNP ascrivibili al tipo A/B, in attesa di essere completamente scoperte e caratterizzate.

3.10 Prospettive future nella caratterizzazione di ssDBF

Sulla base dei risultati finora ottenuti le prospettive di approfondimento di questa linea di ricerca sono molteplici. Infatti rimangono ancora aperti vari problemi. Soprattutto la possibilità di produrre la proteina in forma ricombinante, su larga scala e con un elevato grado di purezza, consentirà di affrontare e chiarire numerosi aspetti sia funzionali sia strettamente strutturali ancora oscuri.

Per quanto riguarda il ruolo funzionale un chiarimento a breve termine potrà essere ottenuto mediante l'utilizzo di anticorpi specifici monoclonali con i quali identificare la localizzazione cellulare di queste proteine, mentre informazioni più approfondite potrebbero venire da esperimenti di transfezione con vettori di espressione recanti il gene allo studio o sue varianti mutate.

Di particolare interesse è anche l'aspetto strutturale ovvero, qual è la struttura terziaria di ssDBF alias CBNP α e quali sono i determinanti molecolari della sua specificità nel riconoscimento di stringhe di citosine? I dati sperimentali a nostra disposizione suggeriscono che la modalità di interazione della proteina con il DNA ripetitivo a motivo telomerico è a stechiometria piuttosto complessa, come evidenziato da saggi combinati di EMSA e di UV-crosslinking ed in considerazione dell'osservazione che l'assenza di condizioni riducenti abolisce l'interazione specifica con il DNA.

Particolarmente utili quali modelli negli studi strutturali potranno essere le strutture tridimensionali risolte attraverso cristallografia a raggi X o NMR dei domini RNA *binding* di alcune hnRNP di tipo A/B (Nagai *et al.*, 1992; Wittekind *et al.*, 1992). Tali domini, denominati RBD o RRM o anche RNP-MS, consistono di una sequenza di 90 amminoacidi contenente due corte sequenze altamente conservate, un octapeptide (RNP1) ed un esapeptide (RNP2) separati da una trentina di amminoacidi, più altri residui conservati dispersi (Dreyfuss *et al.*, 1993). Per quanto riguarda la struttura terziaria vi sono sei elementi di struttura secondaria (β 1, α 1, β 2, β 3, α 2, β 4) che a loro volta originano una struttura terziaria composta da un foglietto β a quattro filamenti antiparalleli sovrapposti a due α eliche perpendicolari tra loro (Nagai *et al.*, 1990; Wittekind *et al.*, 1992) (Figura 29). In questa struttura le sequenze RNP1 ed RNP2 sono localizzate rispettivamente nei filamenti centrali β 1 e β 3 perciò si pensa che siano esse ad effettuare il contatto diretto con l'RNA (Merrill *et al.*, 1988). La specificità di legame del dominio RBD sembra risiedere principalmente nella regione variabile del loop che connette i foglietti β e nelle porzioni terminali (Gorlach *et al.*, 1992).

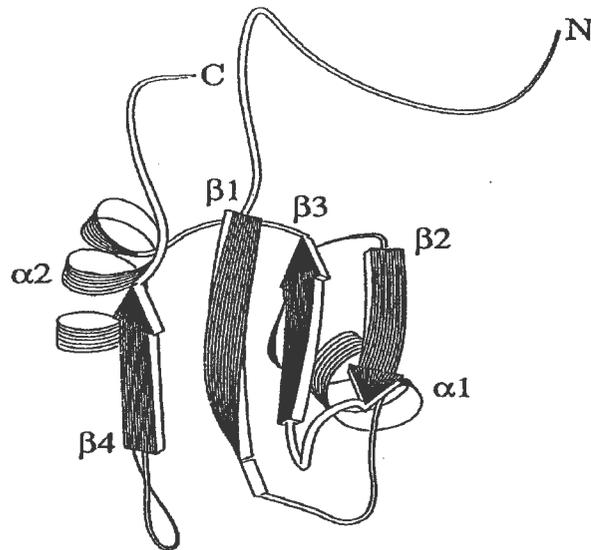


Figura 29: Struttura terziaria dell'RNA binding domain di tipo RBD. Sono rappresentati i quattro foglietti β ad orientazione antiparallela, sovrapposti a due α -eliche perpendicolari tra loro. N e C indicano rispettivamente le estremità N-terminale e C-terminale del dominio.

Recentemente è stata risolta tramite cristallografia a raggi X (Ding *et al.*, 1999) la struttura del frammento UP1 dell'hnRNP A1, comprendente i due domini RNP in tandem, non solo come proteina in sé (Shamoo *et al.*, 1997) ma anche complessato con il DNA a filo singolo a sequenza telomerica $d(\text{TTAGGG})_2$, quella complementare al filamento all'esame nella nostra ricerca. Questo studio riveste particolare importanza dal momento che è stato dimostrato il coinvolgimento di hnRNP A1 nella biogenesi dei telomeri (LaBranche *et al.*, 1998). E' plausibile che anche CBNP α /ssDBF, per quanto riguarda almeno la porzione relativa ai due

domini RNP, assuma in linea di massima la stessa struttura terziaria. Uno studio di “*molecular modelling*” prendendo tale struttura come riferimento, potrà auspicabilmente permettere di capire quali delle differenze per sostituzione amminoacidica tra UP1 e CBNP α , siano rilevanti per la discriminazione delle sequenze nucleotidiche legate.

D'altra parte, la produzione della CBNP α in forma ricombinante potrà consentire di disporre in quantità sufficienti all'ottenimento di cristalli singoli della stessa, sola o possibilmente complessata con il DNA telomerico, per l'analisi diffrattometrica ed eventualmente per lo studio tramite spettroscopia NMR ai fini della risoluzione diretta della sua reale struttura terziaria.

3.11 Purificazione della proteina da cellule K562

In questa prospettiva ci è sembrato di estremo interesse passare alla identificazione e caratterizzazione nell'uomo dell'analogo funzionale di ssDBF, alias CBNP α , a partire da estratti proteici di cellule K562, che avevano mostrato esprimere una componente ad attività specifica di *binding* alla sonda telomerica.

A partire da estratti nucleari totali di cellule K562, la cromatografia di affinità condotta nelle medesime condizioni scelte per la purificazione da eritrociti di pollo, permette di recuperare almeno otto componenti proteiche principali, evidenziabili in Blue Coomassie. In figura 30 A è riportato un esempio di elettroforesi SDS-PAGE relativa a quattro diverse preparazioni di proteine purificate da estratti di cellule K562 (corsie 5, 6, 7 e 8), a confronto con un tipico profilo elettroforetico delle tre componenti da noi purificate a partire dagli eritrociti di pollo (corsia 4 della stessa figura). Anche in questo caso la cromatografia è stata condotta in

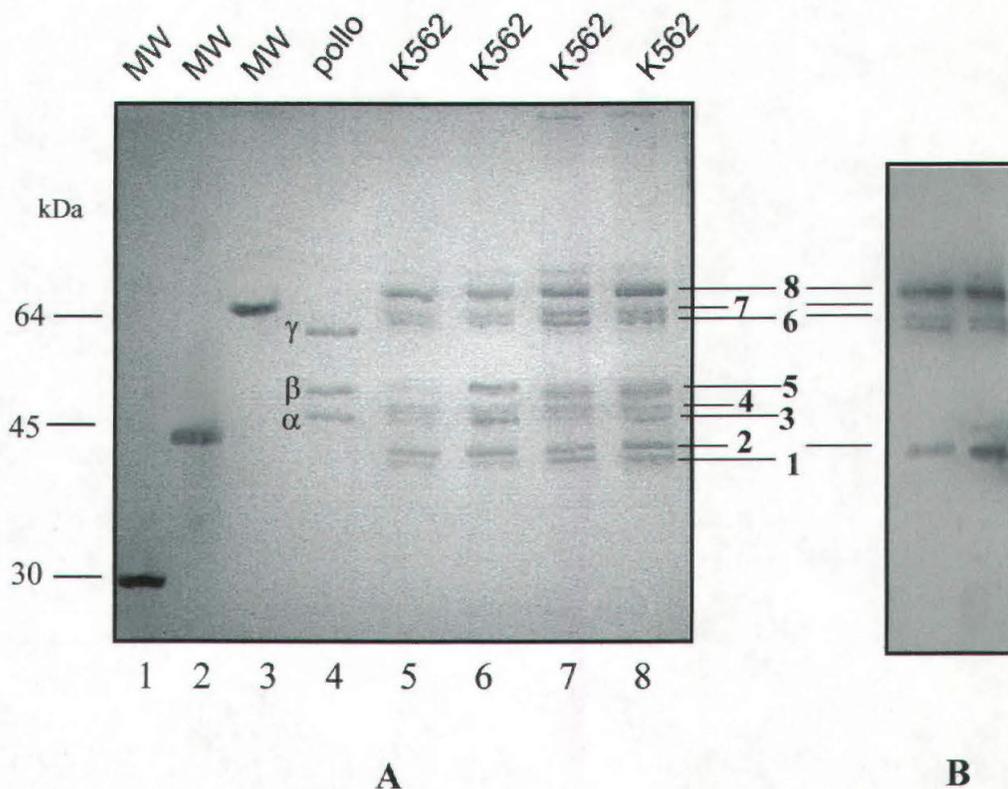


Figura 30

- (A)** *SDS PAGE Tris-Glicina all'8% dei campioni recuperati da cromatografia di affinità:*
- corsie dalla 1 alla 3, markers di peso molecolare;
 - corsia 4, componenti purificate da eritrociti di pollo;
 - corsie dalla 5 alla 8, quattro differenti campioni recuperati da cromatografia di affinità a partire da estratti nucleari di cellule K562
- (B)** *South-Western di 2 campioni recuperati dalla cromatografia di affinità di estratti proteici totali di cellule K562.*
Dopo separazione su gel SDS PAGE all'8%, i campioni sono stati trasferiti su membrana di nitrocellulosa ed incubati con la sonda oligonucleotidica HTC4.

assenza del competitore, utile ad evitare interazioni aspecifiche di proteine con l'oligonucleotide telomerico.

I saggi South Western (figura 30 B) condotti in presenza di un modesto eccesso di competitore aspecifico, come descritto nei materiali e metodi, hanno evidenziato attività di *binding* alla sonda HTC4 di almeno 4 delle 8 componenti isolate che, per confronto con il profilo elettroforetico di figura 30 A, risulterebbero coincidere con le bande numero 2, 6, 7 e 8. L'assenza di segnale per le altre componenti può essere ascrivibile ad una loro mancata rinaturazione, o al fatto che possono complessare con la sonda solo assieme a qualcuna delle altre componenti, o che non leghino direttamente la sonda ma una delle altre proteine.

L'insieme delle proteine eluite dalla resina manifestano in EMSA il segnale specifico di *binding* alla sonda telomerica manifestato dall'estratto totale di partenza (figura 31 corsia 1). Tale complesso inoltre mostra verso l'oligonucleotide sonda HTC4 la stessa affinità di legame esibita dall'estratto proteico totale di partenza e dagli estratti cellulari di diversa origine specie-specifica. La figura 31, è relativa ad un saggio EMSA condotto incubando il *pool* di proteine recuperata dalla cromatografia, con la sonda telomerica HTC4, in presenza di DNA competitore aspecifico e di un eccesso di 10 volte di una serie di oligonucleotidi a sequenza telomerica o telomerico-simile già saggiati per l'affinità di *binding* alla proteina dagli estratti totali e i cui risultati sono riassunti in tabella III. Appare chiaro che l'affinità della componente isolata per affinità rispecchia perfettamente quella esibita dalla proteina presente nell'estratto totale: il suo complesso con la sonda, oltre che ovviamente da HTC4 freddo (corsia 3), viene efficacemente indebolito solo dall'oligonucleotide TTC4, per nulla o quasi dagli altri (corsie dalla 4 alla 10). Da osservare che analogamente a quanto osservato in pollo, anche in questo caso il ripristino dell'attività specifica richiede un ambiente riducente.

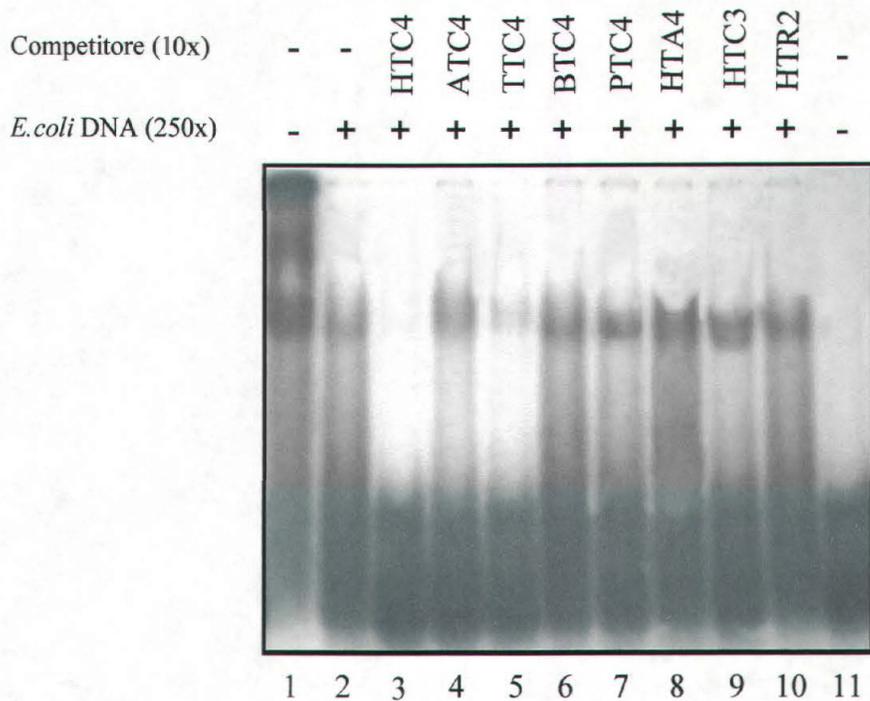


Figura 31

Saggio EMSA di competizione con differenti oligonucleotidi di tipo telomerico, le cui sequenze sono riportate in tabella III.

La sonda marcata HTC4 è stata incubata con il pool di proteine da cellule umane K562 recuperate dalla resina di affinità:

- corsia 1, incubazione effettuata in assenza di competitore
- corsia 2, incubazione in presenza di un eccesso di DNA di *E.coli* denaturato (250X)
- corsia dalla 3 alla 10, incubazione in presenza di un eccesso di 250 volte (w/w) di DNA denaturato di *E.coli* e di un eccesso di 10 volte di oligonucleotidi non marcati
- corsia 11, sonda marcata HTC4 in assenza di proteine

3.12 Identificazione delle proteine isolate da cellule K562.

Le proteine costituenti le singole bande cromatografiche sono state analizzate tramite MALDI-TOF MS presso il Laboratorio di Spettrometria di Massa dell'I.A.B.B.A.M. del C.N.R. di Napoli. La Tabella VIII riporta le masse molecolari dei vari frammenti ottenuti per le componenti isolate, con l'eccezione della componente 4, che non è stato possibile analizzare per difficoltà tecniche. La congruità delle masse molecolari ottenute sperimentalmente con le masse molecolari teoriche calcolabili per peptidi delle proteine note e riportate nelle banche dati ha consentito di identificare, con altissima probabilità, tutte e sette le componenti. Con l'eccezione della banda 5, che è risultata essere costituita da β -actina, tutte le altre bande sono state identificate con varie hnRNP. In particolare le bande 1 e 3 sono date dalla proteina JKTBP (De Maria *et al.*, 1996) e dalla proteina AUF1 (Demsey *et al.*, 1998) rispettivamente, entrambe ascrivibili alle hnRNP del gruppo D in senso lato. Le bande 2, e 8 sono invece costituite dalle proteine E1 (Ostareck *et al.*, 1997) e K (Michelotti *et al.*, 1996) rispettivamente, caratterizzate da parziale omologia di sequenza tra loro, ma non con le due precedenti (vedere tabella X). Le bande 6 e 7 infine sono costituite da due isoforme della hnRNP I (Singh *et al.*, 1995).

Poiché è noto che la hnRNP K, come alcune altre della stessa famiglia, interagisce soprattutto con sequenze molto ricche in citosine, si è ritenuto di verificare se il *pool* di proteine recuperate dalla cromatografia di affinità, basi tale preferenza solo sulla presenza di stringhe di citosine sul nostro DNA sonda. È stato effettuato un esperimento EMSA a competizione incrociata tra HTC4 e C22. L'incubazione è stata effettuata in presenza costante di competitore aspecifico in eccesso di 500 volte, per rimuovere la componente aspecifica dell'interazione (figura 32). Da tutto ciò si è osservato che a parità di sonda e proteine incubate, il segnale EMSA è decisamente più intenso nel caso di HTC4 (corsia 2) che di C22

	Banda 1	Banda 2	Banda 3	Banda 5	Banda 6	Banda 7	Banda 8
	1835.1 (156-170)	1961.3 (161-177)	1158.2 (116-125)	976.9 (19-28)	1432.4 (123-134)	1432.3 (123-134)	1099.2 (140-148)
	1895.0 (52-68)	2177.9 (326-346)	1335.4 (66-77)	1036.6 (327-335)	1898.8 (239-254)	1898.7 (239-254)	1195.6 (306-316)
	2184.3 (71-90)	2476.6 (178-200)	1484.5 (78-90)	1133.2 (197-206)	2040.2 (349-366)	2040.2 (349-366)	1225.7 (412-422)
	2203.9 (132-150)	2660.2 (47-70)	1489.3 (163-176)	1199.6 (51-61)	2113.7 (65-84)	2113.2 (65-84)	1341.5 (208-219)
	2274.2 (193-213)	3203.1 (269-297)	1565.9 (240-251)	1355.2 (51-62)	2244.4 (219-238)	2244.6 (219-238)	1520.2 (149-163)
	2428.1 (69-90)	3380.9 (71-101)	1758.8 (163-178)	1517.8 (360-372)	2267.4 (349-368)	2267.7 (349-368)	1550.3 (180-191)
	2448.9 (31-51)	3549.7 (47-78)	1860.3 (203-217)	1549.8 (313-326)	2276.4 (326-348)	2276.6 (326-348)	1555.1 (208-221)
	3234.9 (24-51)	3615.7 (145-177)	1887.4 (162-178)	1624.9 (178-191)	2489.8 (186-206)	2489.5 (186-206)	1736.5 (22-35)
	3892.6 (116-150)		1921.1 (99-115)	1630.5 (197-210)	2759.7 (123-146)	2759.4 (123-146)	1781.9 (70-86)
	4393.2 (116-154)		2016.1 (202-217)	1792.1 (239-254)	2996.4 (445-471)	2996.6 (445-471)	1792.1 (317-332)
	4657.5 (109-150)		2082.8 (235-251)	1955.6 (96-113)	3026.9 (65-92)	3026.7 (65-92)	1918.2 (378-396)
			2162.5 (179-197)	2095.1 (239-256)	3118.3 (411-437)	3118.2 (411-437)	1979.5 (22-37)
MH⁺			2171.8 (201-217)	2237.3 (178-196)	3135.9 (186-212)	3208.2 (95-122)	2070.1 (202-219)
(peptide)			2285.9 (71-90)	2259.9 (292-312)	3207.9 (95-122)	3393.6 (93-122)	2121.8 (67-86)
			2291.4 (118-137)	2388.4 (291-312)	3393.6 (93-122)	3493.1 (493-523)	2283.9 (202-221)
			2361.9 (203-221)	2550.9 (216-238)	3492.5 (493-523)	3620.9 (492-523)	2330.5 (306-326)
			2431.2 (177-197)	2731.9 (336-359)	3621.2 (492-523)	3683.1 (147-185)	2559.1 (304-326)
			2535.1 (116-137)	3185.3 (148-177)	3683.3 (147-185)	3892.8 (490-523)	2591.3 (434-456)
			2662.2 (179-201)	3233.6 (257-284)	3893.0 (490-523)	4085.1 (445-482)	2599.7 (140-163)
			2774.9 (177-200)	3537.1 (255-284)	4085.1 (445-482)	4123.0 (219-254)	2708.9 (410-433)
			2918.1 (91-115)	3663.1 (207-238)	4298.3 (326-366)	4417.3 (445-485)	2855.2 (140-166)
			3288.0 (109-137)	4052.0 (85-118)	4417.2 (445-485)	5061.7 (274-325)	3060.2 (222-246)
			3345.6 (235-263)		5062.0 (274-325)		3216.4 (222-247)
			4245.4 (163-200)				3832.4 (423-456)
			4401.8 (163-201)				4058.2 (104-139)
							4642.1 (415-456)
Proteina	JKTBP	hnRNP E1	hnRNP AUF1	β actin	hnRNP I	hnRNP I	hnRNP K

Tabella VII: Identificazione delle specie proteiche purificate dalle cellule K562. Per ciascuna proteina è riportato il numero della banda elettroforetica corrispondente all' SDS PAGE di Figura 31A, i valori di massa dei peptidi ottenuti e tra parentesi la posizione di tali peptidi nella proteina.

(corsia 7); inoltre HTC4 freddo è molto più efficiente a spiazzare la sonda C22 (corsie 8, 9 e 10) di quanto C22 freddo riesca a competere con la sonda HTC4 (corsie 3 e 5). Ciò dimostra che tra le proteine recuperate da affinità non ci sono solo “generiche” *C-rich nucleic acid binding proteins*.

Quindi, per stabilire accuratamente la relazione tra le attività presenti nell'estratto totale, ovvero l'attività più specifica, e le componenti purificate, le singole bande elettroforetiche di un gel Tris-Glicina sono state escisse separatamente ed estratte dal gel attraverso precipitazione in acetone. Per ciascuna componente, ad eccezione della numero 1, data la quantità esigua con cui veniva recuperata dalla resina di affinità, è stato allestito un saggio EMSA con l'oligonucleotide HTC4 marcato, rispettivamente in assenza od in presenza di un eccesso di 500 volte (w/w) di DNA di *E. coli* denaturato (figura 33). Com'era probabile attendersi per il fatto che le proteine sono state purificate per interazione con l'oligonucleotide telomerico legato alla resina, tutte le componenti proteiche, ad eccezione di quella corrispondente alla banda elettroforetica 5, manifestano in assenza di competizione un seppur debole segnale EMSA, ad indicazione dell'esistenza di un'interazione abbastanza stabile tra la proteina e la sonda oligonucleotidica. Al contrario, in presenza del competitore, solo due delle otto proteine sembrano in grado di formare complessi specifici con l'oligonucleotide HTC4 (corsia 2 e 4). Precisamente la componente relativa alla banda elettroforetica 2, identificata con la hnRNP E1, è responsabile delle comparsa di un segnale EMSA a mobilità paragonabile a quella del complesso specifico manifestato dall'estratto totale e dall'eluato cromatografico. Meno intensa e percettibilmente più lenta appare invece il complesso relativo alla componente numero 3 ovvero la componente costituita dalla AUF-1. (corsie 4 e 5 di figura 33).

Dall'analisi delle sequenze è emerso inaspettatamente che nessuna delle componenti da noi isolate dalla cromatografia degli estratti da cellule umane ed identificate come hnRNP mostra una elevata omologia di sequenza con ssDBF

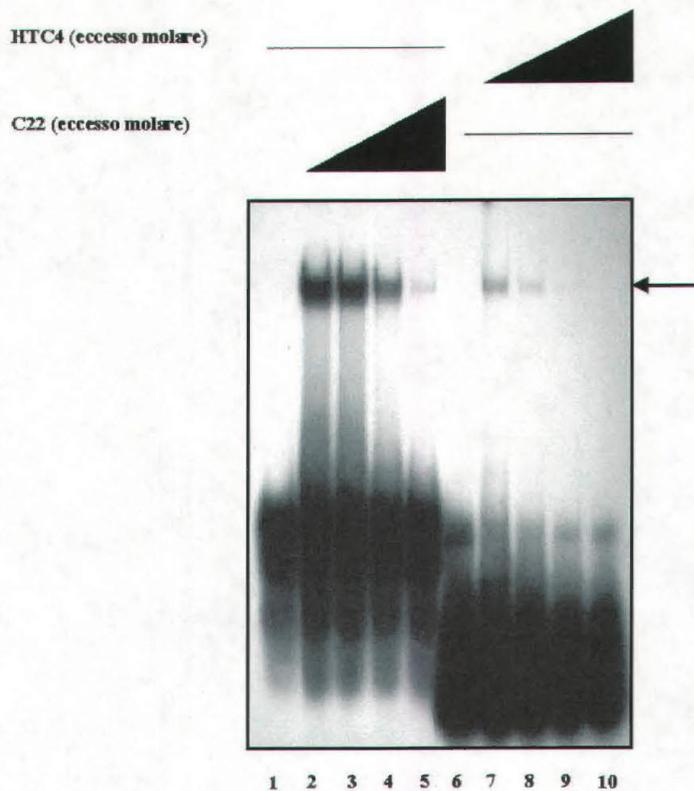


Figura 32

Saggio EMSA per la valutazione dell'attività specifica del pool di proteine eluite dalla cromatografia di affinità verso l'oligonucleotide HTC4. 0.2 picomoli di sonde HTC4 (corsia 1-5) e C22 (corsia 6-10), sono stati incubati con 0.1 microgrammi di proteine, in eccessi molari (0, 5, 50, 500 volte) di oligonucleotidi C22 (corsie 2-5), e HTC4 (corsie 7-10) rispettivamente. In tutti i campioni il saggio è stato effettuato in presenza di un eccesso di 500 volte di DNA competitore

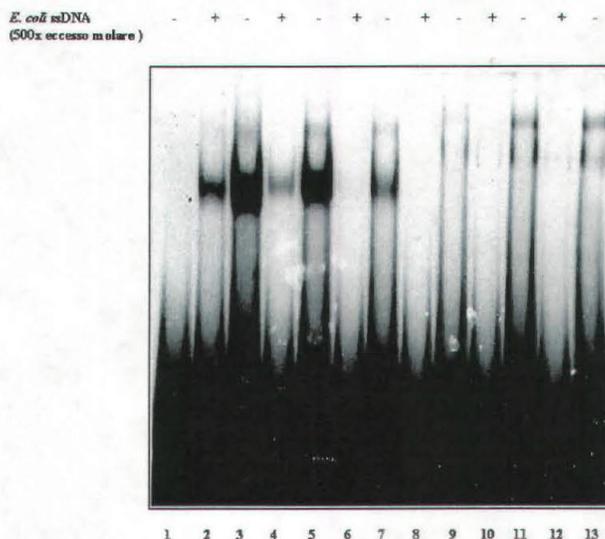


Figura 33

Attività delle proteine elettroeluite da gel. La sonda marcata HTC4 è stata incubata con 5 microlitri delle componenti proteiche ottenute dalle singole bande elettroforetiche: corsia 1, componente banda 5; corsie 2-3, componente banda 2; corsie 4-5, componente banda 3; corsie 6-7, componente banda 4; corsie 8-9, componente banda 6; corsie 10-11, componente banda 7; corsie 12-13, componente banda 8

isolata da eritrociti di pollo ed identificata come responsabile del *binding* ad alta affinità verso la sonda telomerica.

In particolare, tralasciando l'hnRNP K, le due hnRNP I, la AUF1 e la JKTBP, che come abbiamo visto sono responsabili della formazione di complessi poco specifici o comunque con una mobilità EMSA distinta chiaramente da quella manifestata dal complesso specifico dell'estratto *in toto*, solo l'hnRNP E1 sembra in grado di generare un segnale EMSA identificabile per omologia di mobilità e specificità come quello corrispondente alla proteina specifica allo studio (dati non riportati).

L'hnRNP E1, pur essendo ascrivibile all'eterogeneo gruppo delle hnRNP, in sequenza e in funzioni sembra differire molto sia da ssDBF che dalla sua presumibile omologa umana, l'hnRNP di tipo A/B (Khan *et al.*, 1991; Figura X). In letteratura, classificata genericamente come una *Poly-C Binding protein*, è nota anche come proteina α CP-1 (Buckanovic *et al.*, 1993) o PCBP-1 (Musco *et al.*, 1996) ed è altamente correlata sia in sequenza che in struttura all' hnRNP E2 / α CP-2 / PBBP-2 (Weng *et al.*, 1995). Entrambe queste proteine appartengono alle *RNA binding proteins* di tipo K (Buckanovich *et al.*, 1993), caratterizzate cioè da particolari domini di legame all'RNA, chiamati domini KH (K Homology), ben distinti da quelli che caratterizzano le regioni RNA binding delle hnRNP di tipo A/B, a cui ssDBF appartiene. E' stato dimostrato che sia l'hnRNP E1 che l'hnRNP E2 sono implicate nella regolazione dell'espressione genica a vari livelli come processamento, trasporto, stabilità e traduzione dell'm-RNA, probabilmente agiscono facendo da ponte per facilitare l'associazione di un sistema regolativo multicomponente. In particolare per quanto riguarda E1, essa fa parte del complesso ribonucleoproteico che si associa alla regione 3' non tradotta (3'-UTR) degli m-RNA a lunga vita (Wang *et al.*, 1993) ed è coinvolta nella stabilizzazione degli m-RNA dell' α -globina umana, del collagene e della tirosina-idrossilasi di

ratto (Kiledjian *et al.*, 1995; Stefanovic *et al.*, 1996; Czyzyk-Krzeska *et al.*, 1996) e risulta implicata nel controllo della traduzione dell'm-RNA per la 15-lipoossigenasi (Ostareck *et al.*, 1997) e di alcuni RNA virali (Collier *et al.*, 1998; Graff *et al.*, 1998).

Gli aspetti funzionali reali o ipotetici di E1 verranno però approfonditi in seguito.

Al fine di saggiare effetti sinergici tra le componenti nell'interazione con HTC4, sono stati effettuati esperimenti EMSA incubando contemporaneamente con la sonda due o più delle componenti, centrando l'attenzione soprattutto sulla componente 2 ovvero hnRNP E1, cioè la più specifica. Un'osservazione che potrebbe rivestire un certo interesse è che la compresenza della componente 5 (β -actina), che di per sé non lega la sonda, indebolisce l'efficacia di interazione della componente 2. Non si osservano invece effetti sinergici rilevanti nel caso della compresenza della componente 2 con le componenti 6, 7 e 8. Anche incubando, in diverse combinazioni, le altre componenti non si osservano effetti sinergici. Il segnale EMSA dato dalle singole proteine incubate con le altre, non rivela un'alterazione né della mobilità elettroforetica, né dell'intensità rispetto a quello delle stesse proteine incubate singolarmente (dati non riportati).

Siccome la proteina isolata da cellule K562 non è risultata essere l'omologa di ssDBF, è sembrato quanto meno strano, che se essa di fatto non rappresentasse la sua analogo funzionale, manifestasse però la stessa specificità ed affinità di sequenza di ssDBF. Abbiamo quindi valutato se e in quale misura la componente elettrostatica influenzasse il *binding* di ssDBF con la sonda telomerica e con gli oligonucleotidi utilizzati come competitori, allo scopo di discriminare nel legame il contributo dato dalle componenti elettrostatiche e quello dato dalle specifiche interazioni covalenti. Nei saggi di specificità condotti con la componente numero 2, cioè hnRNP E1, tutti gli oligonucleotidi utilizzati o come sonda o come competitori freddi, hanno la stessa lunghezza e quindi la stessa carica, è plausibile

quindi ritenere che la componente elettrostatica della loro interazione con le proteine in esame sia sostanzialmente sempre la stessa. Quindi le differenze di affinità osservate devono essere ascritte alla componente non elettrostatica dell'interazione. Pertanto un'analisi della dipendenza dell'interazione dalla forza ionica, a cui la parte elettrostatica è sensibile e l'altra no, dovrebbe consentire di discriminare le due componenti. La Figura 34A, mostra il risultato di un saggio EMSA effettuato su miscele tra la sonda HTC4 e la componente 2 (quest'ultima in difetto stechiometrico) incubate a valori crescenti di forza ionica, in eccesso del competitore aspecifico. Si osserva che, in effetti, l'intensità della banda EMSA in queste condizioni è largamente insensibile alla forza ionica. La Figura 34B, presenta un analogo esperimento effettuato in assenza del competitore aspecifico. Si può osservare che la banda EMSA è molto più intensa a bassa forza ionica e che, aumentando questa, l'intensità diminuisce residuando una quantità paragonabile a quella di Figura 34A. Il segnale EMSA della corsia 4 di Figura 34B, appare più debole di quello di corsia 5 di Figura 34A, solo per il minor tempo di esposizione autoradiografica. Questa differenza di comportamento indotta dalla presenza o meno del competitore aspecifico, non è di immediata razionalizzazione. La ragione più plausibile per tale comportamento sta nell'assunzione che la componente 2 sia in realtà eterogenea, costituita da una parte più abbondante in grado di interagire con la sonda in modo aspecifico, quindi sensibile alla forza ionica ed all'azione del DNA competitore, ed una parte meno abbondante, responsabile dell'interazione specifica e largamente insensibile alla forza ionica ed al competitore. La causa dell'eterogeneità potrebbe essere intrinseca al materiale di partenza (due isoforme non risolte dall'elettroforesi) o dovuta alla procedura di isolamento (rinaturazione incompleta dopo il recupero dal gel di elettroforesi in SDS). In un ulteriore esperimento EMSA si osserva che, in presenza del competitore aspecifico, la banda EMSA mostrata dalla sonda HTC4 incubata con la componente 2, non risente dell'aumento di forza ionica (conferma indipendente

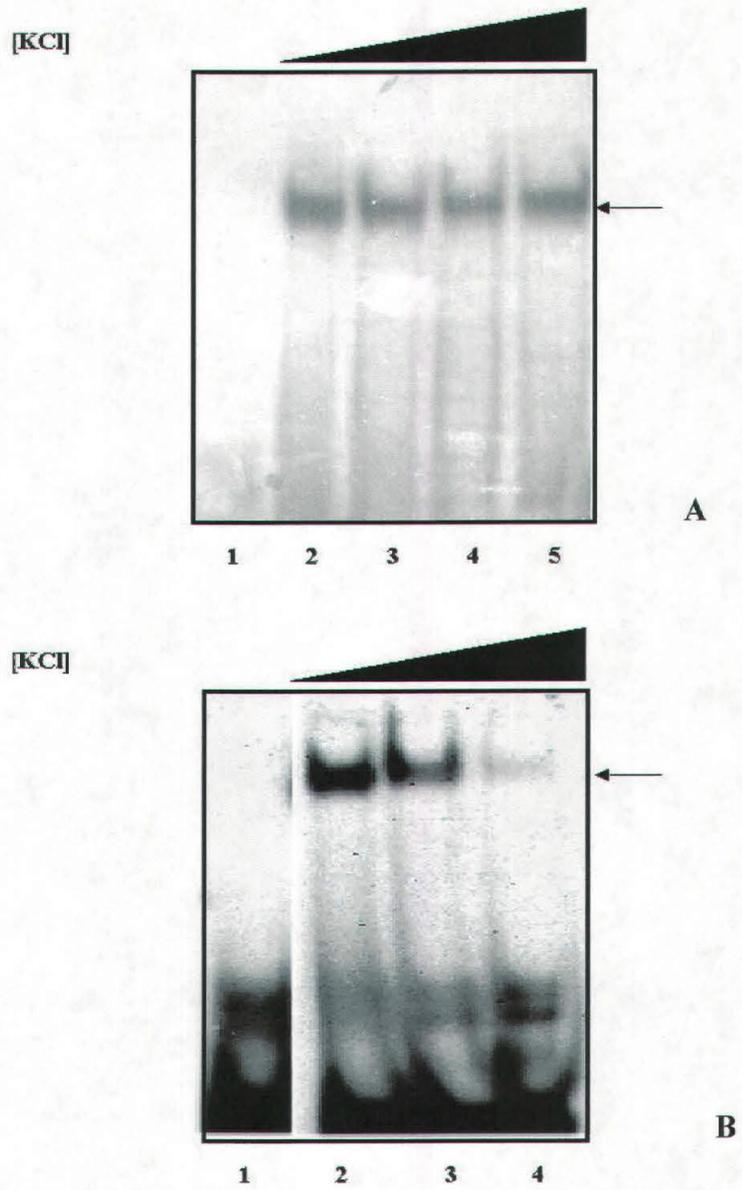


Figura 34:

Effetto della forza ionica sull'attività di legame della componente numero 2, in figura A in presenza ed in figura B in assenza di DNA competitore (500 volte w/w in eccesso). La sonda marcata è stata incubata con la componente 2 in presenza di concentrazioni crescenti di KCl: corsia 1, HTC4 marcato; corsia 2, 75 mM KCl; corsia 3, 600 mM KCl; corsia 4, 800 mM KCl; corsia 5, 1 M KCl

del risultato di Figura 34), mentre quella prodotta dalla sonda C22 ne risente, a riprova della minore specificità di tale interazione (dati non riportati).

3.13 Caratterizzazione delle proteine isolate da cellule K562: loro ruolo e possibile relazione funzionale con CBNP α

Come già detto, tra le componenti identificate e saggiate riguardo alla loro specificità di interazione con il motivo ripetuto d(CCCTAA)_n, la proteina più interessante appare essere la hnRNP E1. Questa proteina è stata inizialmente identificata da Aasheim *et al.* (1994) tramite *screening* di una genoteca a cDNA da linfociti umani come capace di legare sequenze oligo(dC) e provvisoriamente denominata *sub2.3 protein*. Tale proteina è stata successivamente caratterizzata da Leffers *et al.* (1995), assieme ad una forma altamente omologa (80%), e le due sono state denominate PCBP-1 e PCBP-2 (*Poly(C) Binding Protein*), a causa della loro preferenza nell'interazione con stringhe di citosine. Contemporaneamente Kiledjian *et al.* (1995) hanno isolato due proteine che si legano all'mRNA dell' α -globina, denominandole α CP-1 e α CP-2, le quali sono risultate coincidere con PCBP-1 e PCBP-2. Tali proteine dimostrano di legarsi, assieme ad altre, ad un elemento di sequenza, ricco in citosine, al 3' non tradotto di questo mRNA e di stabilizzarlo, permettendone l'accumulo durante il differenziamento della linea eritroide. Successivamente Ostareck *et al.* (1997) hanno identificato nelle hnRNP K ed E1 di coniglio due proteine regolatrici della traduzione dell'mRNA della 15-lipossigenasi durante la maturazione dei reticolociti, le quali hanno attività silenziatrice legandosi ad un elemento, denominato DICE (*Differentiation Control Element*), ricco in citosine al 3' non tradotto del messaggero. La hnRNP E1 è risultata essere l'omologo in coniglio di PCBP-1, *alias* α CP-1, *alias* hnRNP E1

umano. Infine Makeyev *et al.* (1999) hanno caratterizzato i geni degli omologhi murini, trovando che per l'omologo di α CP-2 esiste una forma genomica normale, composta da numerosi esoni, mentre per l'omologo di α CP-1 il gene consiste di una copia priva di introni, localizzata in un cromosoma distinto, presumibilmente derivata da un evento di retrotrasposizione, ma con un risultato funzionale. Tanto l'omologo di coniglio che quello di topo differiscono dalla PCBP-1, *alias* hnRNP E1, umana per un unico residuo amminoacidico su 301 (Ala198Val), dimostrando un grado di conservazione evolutiva estremamente elevato. Mentre l'elemento ricco in citosine nel messaggero dell' α -globina dimostra una somiglianza piuttosto superficiale con la ripetizione telomerica, l'elemento DICE presenta invece una somiglianza abbastanza spiccata con questa. Non sorprende quindi che questa proteina interagisca fortemente con entrambi, soprattutto con il secondo. Per quanto riguarda il reale ruolo biologico di hnRNP E1 ancora non è possibile dire molto. Certamente ci si può attendere che questa, come altre proteine di questa famiglia, assuma più ruoli funzionali, diversificati a seconda del livello di espressione, del tipo di tessuto e delle eventuali differenze nelle modifiche post-traduzionali.

Queste proteine condividono con la hnRNP K il fatto di contenere tre domini KH. Questa ultima proteina è tra quelle isolate per cromatografia di affinità alla ripetizione telomerica e identificate nel corso della ricerca presentata in questa tesi. Molto recentemente (Lacroix *et al.*, 2000) è stato dimostrato che essa può interagire con il motivo telomerico ripetuto d(CCCTAA)_n.

Fenomeni di *splicing* alternativo, comuni tra le hnRNP, danno origine a quattro tipi di isoforme strettamente correlate tra loro, le quali differiscono per l'inserzione di 5-20 residui amminoacidici e presentano diverse proprietà di legame con gli acidi nucleici. Una di esse, come già descritto, attiva la trascrizione del gene *myc in vitro* ed interagisce, sia *in vitro* che *in vivo*, con la proteina TBP (TATA-Binding-

Protein), che lega il TATA-box nel complesso iniziale di trascrizione (Michelotti *et al.*, 1996). La hnRNP K contiene inoltre un segnale di localizzazione nucleare basico bipartito (NLS) (Michael *et al.*, 1995) e un particolare dominio nucleare di *shuttling* (KNS), localizzato tra il motivo ricco in prolina e il terzo dominio KH (Michael *et al.*, 1997). Esso conferisce la capacità di trasporto bidirezionale attraverso la membrana nucleare. Queste caratteristiche indicano quindi che la hnRNP K non è coinvolta soltanto nel processamento dell'RNA ma anche nel trasporto nucleo-citoplasmatico e verosimilmente nel controllo della trascrizione di alcuni geni.

Per quanto riguarda la specificità nel riconoscimento di sequenze polinucleotidiche la hnRNP K potrebbe essere considerata tra quelle che preferiscono legare sequenze ad alto contenuto di citosine. Poiché DNA competitore aspecifico, quale quello di *E. coli* denaturato, risulta abbastanza efficace nel competere con quello telomerico nel complessare con questa proteina (Figura 33, corsie 12 e 13) è plausibile ritenere che la sua interazione con il motivo telomerico non rivesta un significato funzionale, ma sia solo una conseguenza tecnica della procedura cromatografica di isolamento.

Le altre due proteine identificate nel corso di questo lavoro sono la proteina denominata JKTBP (*JKT Binding Protein*) e la AUF-1 (*AU-rich binding Factor*). Entrambe sono ascrivibili alle hnRNP. In particolare AUF-1 fa parte delle hnRNP del gruppo D, ed è stata denominata anche hnRNP D0. JKTBP presenta una notevole omologia di sequenza con AUF-1 (circa il 70%), ed entrambe hanno un'omologia di sequenza simile alle hnRNP del tipo A/B). Non mostrano per contro omologie significative con le hnRNP discusse sopra. La hnRNP D0, *alias* AUF-1, è una proteina presente nella maggior parte dei tipi cellulari, è codificata da un unico gene, che però dà origine a quattro isoforme (A,B,C,D), per fenomeni di *splicing* alternativo. Le proteine umana e murina sono omologhe al 97%, dimostrando anche in questo caso una funzione evolutivamente molto conservata.

Oltrechè legare gli elementi AU-rich presenti negli mRNA di vari geni, per alcune isoforme sono state evidenziate interazioni diverse: con un motivo presente nel promotore del gene CR2 (Tolnay *et al.*, 1997), con altre componenti a livello della *switch region* dei geni delle catene pesanti delle immunoglobuline (Dempsey *et al.*, 1998), ed anche con il segnale di *splicing* al 3' nonché la ripetizione telomerica d(TTAGGG)_n (Ishikawa *et al.*, 1993). Nel nostro laboratorio si è anche osservato che questa proteina è in grado di legare anche il filamento telomerico complementare, ma senza particolare specificità (Figura 33). A questo proposito è forse rilevante l'osservazione che, tramite il saggio del doppio ibrido, è stata dimostrata un'interazione diretta tra hnRNP E1 ed AUF-1 (Kiledjian *et al.*, 1997). Considerando la specificità di AUF-1 nei confronti della ripetizione d(TTAGGG)_n e quella di E1 per la ripetizione complementare d(CCCTAA)_n, riportata in questa tesi, sarebbe interessante verificare se le due proteine possano agire sinergicamente al fine di separare i due filamenti telomerici complementari appaiati. Saggi EMSA, in cui le hnRNP E1 e AUF-1 sono state coincubate con la sonda HTC4 non mostra rilevanti effetti sinergici in tale caso, nel senso che non compare alcun segnale EMSA aggiuntivo rispetto a quelli dati dalle singole componenti, né viene soppresso (dati non riportati). Tuttavia sarà interessante effettuare analoghi esperimenti con entrambe le sequenze telomeriche ed in modo quantitativo.

Una considerazione *a latere*, tra quelle presentate in questa tesi, riguarda l'isolamento tramite cromatografia di affinità della β -actina. Questa proteina ovviamente non interagisce direttamente con il DNA, ed i saggi EMSA effettuati lo conferma. Tuttavia il suo isolamento per affinità dimostra che sicuramente interagisce con almeno una delle hnRNP ottenute. Ciò potrebbe suggerire che almeno una di queste possa svolgere anche un ruolo di collegamento strutturale tra acido nucleico e scheletro nucleare.

Quanto all'omologia di sequenza della hnRNP E1 e AUF-1 con CBNP α , risultata coincidere con la proteina ssDBF (Smidt *et al.*, 1995) e che appartiene per sequenza alle hnRNP di tipo A/B, questa risulta essere di circa il 40 % per la AUF1 (Tabella VIII), mentre è non significativa per E1. E' riportata in banca dati una proteina umana con alta omologia per ssDBF, e genericamente denominata dagli autori *hnRNP type A/B* (Khan *et al.*, 1991), la cui funzione è incognita, ma che sembra prediligere sequenze ricche in G e U. Data la stretta omologia con ssDBF ci si sarebbe attesi di trovarla tra le proteine umane isolate nel corso di questo lavoro. Così non è stato. Questo fatto potrebbe essere attribuito all'espressione di isoforme diverse per modifiche post-traduzionali che ne alterino le specificità di interazione oppure a livelli di espressione molto bassi o nulli nelle cellule usate (K562). In tal caso, è probabile che la quantità di proteina eventualmente recuperata per affinità sia troppo poca per essere rilevata su gel denaturante colorato in Blue Coomassie. La scelta di eritrociti di pollo nella prima fase di purificazione della proteina è stata di natura pratica in quanto rappresentavano una fonte abbondante e facilmente reperibile ma gli eritrociti rappresentano un tipo cellulare molto specializzato, con probabilmente un *pattern* di espressione di proteine molto diverso dagli altri tipi cellulari. Se è vero che in qualche modo ssDBF, o qualche sua isoforma, è effettivamente coinvolta in qualche aspetto della funzione e gestioni dei telomeri, allora hnRNP E1 non rappresenterebbe l'analogo funzionale di ssDBF, ma una proteina ad esse non correlata, che più o meno fortuitamente lega la sequenza telomerica con una certa specificità.

Al contrario, se in cellule umane l'omologo di ssDBF non è in grado di interagire con la sequenza telomerica, allora dovremmo aspettarci di trovare in tali cellule una proteina che costituisca l'analogo funzionale di ssDBF, e potrebbe trattarsi proprio dell'hnRNP E1. Al momento il quadro relativo agli aspetti funzionali delle hnRNP, come si vede, è ancora molto nebuloso, ma ciò che appare invece abbastanza chiaro è che questa famiglia di proteina nel corso

dell'evoluzione, ha assunto una molteplicità di ruoli notevolmente diversificati anche per una stessa proteina nell'ambito della cellula. Non è da escludere che nel corso dell'evoluzione, in specie evolutivamente molto distanti come i mammiferi e gli uccelli, due diverse hnRNP filogeneticamente non correlate si siano specializzate nel medesimo ruolo funzionale, sia questo legato o no ai telomeri.

Il lavoro fin qui svolto costituisce quindi un punto di partenza per ulteriori approfondimenti in varie direzioni. Infatti una piena comprensione delle funzioni di hnRNP E1 e di ssDBF, sarà possibile soltanto dopo ulteriori approfondimenti sperimentali. Il clone contenente il cDNA della hnRNP E1 è già a disposizione del laboratorio ed il lavoro per ottenere la proteina ricombinante è in atto. La disponibilità di quantità significative della proteina ricombinante potrà permettere, tra l'altro: a) uno studio quantitativo delle sue proprietà di interazione, b) uno studio della localizzazione cellulare non appena ottenuti anticorpi specifici, c) subclonando il cDNA in un vettore di espressione eucariotico, uno studio degli effetti di iperespressione in cellule in coltura, d) auspicabilmente studi di tipo strutturale tramite diffrattometria a raggi X, qualora risulti possibile cristallizzarla. Tutti questi dati sperimentali, associati a quelli ottenuti dalla forma ricombinante di ssDBF, contribuiranno in tempi brevi a chiarire molti aspetti funzionali ma soprattutto la relazione di funzione che sussiste tra le due proteine.

All'inizio del lavoro programmato per questa tesi di dottorato ci si proponeva di isolare ed identificare proteine nucleari capaci di interagire con il filamento al 5' del DNA ripetitivo telomerico. Uno dei presupposti era che, mentre vari gruppi di ricerca avevano iniziato indagini piuttosto intense su fattori proteici interagenti con le ripetizioni telomeriche nella forma del normale duplex W.-C. e sul filamento singolo terminante al 3', ricco in guanine, quasi nulla era riportato in letteratura su proteine interagenti col filo a blocchi di citosine. L'altro presupposto era dato dal fatto in questo laboratorio (Manzini *et al.*, 1994) si era osservato che la ripetizione a blocchi di citosine propria del DNA telomerico era in grado di assumere di per sé una struttura insolita da poco scoperta, denominata i-DNA (Gehring *et al.*, 1993), sollevando la curiosità di scoprire se qualche proteina fosse in grado di riconoscere tale struttura. Si osservò rapidamente che nell'estratto di proteine nucleari da cellule HeLa c'era almeno una componente capace di riconoscere tale sequenza con notevole specificità (Marsich *et al.*, 1996), anche se il fatto che l'interazione risultasse indipendente dal pH tra 6 e 8 faceva propendere per un coinvolgimento della sequenza di per sé; non a causa della struttura i, che è modulata dal pH, essendo stabile a pH 6, ma non a pH superiore a 6.5. La procedura di isolamento e di identificazione, attuata inizialmente a partire da

eritrociti di pollo, e successivamente dalla linea cellulare umana K562, ha rivelato che le componenti che manifestavano la maggiore specificità nei confronti della sequenza telomerica sono, in entrambi i casi, due proteine delle hnRNP. Tuttavia ha costituito, in una certa misura, una sorpresa il fatto che le due proteine, rispettivamente ascrivibili l'una alla famiglia di tipo A/B, l'altra alla famiglia di tipo E, non sono omologhe. Anche perché è già nota una proteina umana con alta omologia a quella aviaria, la cui funzione però è ancora sconosciuta (Khan *et al.*, 1991). La prima, da eritrociti di pollo, è caratterizzata dalla presenza di quattro domini, di cui i due centrali, noti come RBD (RNA Binding Domain) sono abbastanza ben caratterizzati strutturalmente. In particolare anche la hnRNP A1 possiede tali domini in tandem: essi sono stati risolti nella loro struttura terziaria tramite diffrazione di raggi X (Shamoo *et al.*, 1997), anche complessati con la ripetizione telomerica a blocchi di guanine (Ding *et al.*, 1999). E' interessante osservare che questa proteina è al momento l'unica hnRNP di cui sia stato dimostrato un effettivo coinvolgimento nella biogenesi dei telomeri (LaBranche *et al.*, 1999). La seconda, da cellule umane K562, è invece caratterizzata dalla presenza, oltre di domini ausiliari, di tre domini KH, la cui struttura terziaria è comunque simile a quella degli RBD. Tuttavia questa osservazione potrebbe non sorprendere troppo, alla luce del fatto che le hnRNP sono spesso proteine estremamente versatili: un approfondimento su quelle che sono attualmente le funzioni note di queste proteine ha rivelato che molto spesso sono coinvolte in più di una funzione, accomunate essenzialmente dal fatto di interagire con acidi nucleici a filo singolo. Su questa base è plausibile pensare che l'evoluzione le abbia condotte a volte a vicariare tra loro questa o quella funzione, modulata anche da diverse modificazioni post-traduzionali ed interazioni con altre componenti nei diversi contesti e compartimenti cellulari.

A questo punto sarà interessante da un lato, identificare *in vivo* la localizzazione di tali proteine, e nello stesso tempo valutare in sistemi cellulari un loro coinvolgimento più o meno diretto con aspetti funzionali o strutturali dei telomeri.

Aasheim H.C., Loukianova T., Deggerdal A., Smeland E.B. (1994) "Tissue specific expression and cDNA structure of a human transcript encoding a nucleic acid binding oligo (dC) protein related to the pre-mRNA binding protein K", *Nucleic Acids Res.*, **22**: 959-964.

Allsop R.C., Vanziri H., Patterson C., Goldstein S., Younglai E.V., Futcher A.B., Greider C.W., Harley C.B. (1992) "Telomere length predicts the replicative capacity of human fibroblasts", *Proc. Natl. Acad. Sci. USA*, **89**: 10114-10118.

Autexier C., Greider C.W. (1994) "Boundary elements of the *Tetrahymena* telomerase RNA template and alignment domains", *Genes Dev.*, **8**:563-575.

Baber J.L., Libutti D., Levens D., Tjandra N. (1999) "High precision solution structure of the C-terminal KH-domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor", *J. Mol. Biol.*, **289**: 949-962.

Biamonti G., Riva S. (1994) "New insights into the auxiliary domain of eukaryotic RNA-binding proteins", *FEBS Lett*, **340**: 1-8.

Biamonti G., Ruggiu M., Saccone S., Della Valle G., Riva S. (1994) “Two homologous genes, originated by duplication encode the human hnRNP proteins A2 and A1”, *Nucleic Acid Res.*, **22**: 1996-2002.

Biessmann H., Mason J.M., Ferry K., d’Hulst M., Valgeirsdottir K., Traverse K.L., Pardue M.L. (1990) “Addition of telomere-associated HeT DNA sequences "heals" broken chromosome ends in *Drosophila*”, *Cell*, **61**: 663-673.

Blasco M.A., Herrera E., Martinez A.C., (2000), “Impaired germinal center reaction in mice with short telomeres”, *EMBO J.*, **19**: 472-481

Blasco M.A., Herrera E., Samper E., Martin-Caballero J., Flores J.M., Lee H.W., (1999), “Disease states associated with telomerase deficiency appear earlier in mice with short telomeres”, *EMBO J.*, **18**: 2950-2960

Blasco M.A., Lee H.-W., Hande M.P., Samper E., Lansdorp P.M., DePinho R., Greider C.W., (1997), “Telomere shortening and tumor formation by mouse cells lacking telomerase RNA”, *Cell*, **91**:25-34

Bourns B.D., Alexander M.K., Smith A.M., Zakian V. A. (1998) “Sir proteins, Rif proteins, and cdc13p bind *Saccharomyces* telomeres *in vivo*”, *Mol. Cell. Biol.*, **18**: 5600-5608..

Broccoli D., Smogorzewska A., Chong L., de Lange T. (1997) “Human telomeres contain two distinct Myb-related proteins TRF1 and TRF2”, *Nat. Genet.*, **17**: 231-235.

Brun C., Marchand S., Gilson E., (1997), “Proteins that bind to double-stranded regions of telomeric DNA”, *Genes Dev.*, **7**: 885-894

Bryan T.M., Englezou A., Gupta J., Bacchetti S., Reddel R.R. (1995) "Telomere elongation in immortal human cells without detectable telomerase activity", *EMBO J.*, **14**: 4240-4248.

Bryan T.M., Sperger J.M., Chapman K.B., Cech T. R., (1988) "Telomerase reverse transcriptase genes identified in *Tetrahymena thermophila* and *Oxytricha trifallax*", *Proc. Natl. Acad. Sci. USA*, **95**: 8479-8484

Buckanovich R.J., Posner J.B., Darnell R.B. (1993) "Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system", *Neuron*, **11**: 657-672.

Burd C.G., Swanson M.S., Görlach M., Dreyfuss G. (1989) "Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1 and C1 proteins: a diversity of RNA binding proteins is generated by small peptide inserts", *Proc. Natl. Acad. Sci. USA*, **86**: 9788-9792.

Buvoli M., Cobianchi F., Bestagno M.G., Mangiarotti A., Bassi M.T., Biamonti G., Riva S. (1990) "Alternative splicing in the human gene for the core protein A1 generates another hnRNP protein", *EMBO J.*, **9**: 1229-1235.

Cardenas M.E., Bianchi A., de Lange T., (1993), "*Xenopus* egg factor with DNA-binding properties characteristic of terminus-specific telomeric proteins", *Genes Dev.*, **7**: 885-894

Cartegni L., Maconi M., Morandi E., Cobianchi F., Riva S., Biamonti G. (1996) "HnRNP A1 selectively interacts through its Gly-rich domain with a different RNA-binding proteins", *J. Mol. Biol.*, **259**: 337-348.

Collier B., Goobar-Larsson L., sokolowsky M., Schwartz S., (1998), "Translation inhibition in vitro of human papillomavirus type m-RNA mediated

through interaction with heterogeneous ribonucleoprotein K and poly(rC)-binding proteins 1 and 2”, *J. Biol. Mol.*, **278**: 22648-22656

Chong L., van Steensel B., Broccoli D., Erdjument-Bromage H., Hanish J., Tempst P., de Lange T. (1995) “A human telomeric protein”, *Science*, **270**: 1663-1667.

Counter C. M., Avilion A.A., LeFeuvre C.E., Stewart N.G.,m Greider C.W., Harley C.B., Bacchetti S., (1992), “Stabilization of short telomeres and telomerase activity accompany immortalization of Epstein-Barr virus transformed human B-lymphocytes”, *J. Virol.*, **91**: 2900-2904

Cvekl A., McDermott J.B., Piatigorsky J., (1995), “c-DNA encoding a chicken protein (CRP1) with homology to hnRNP type A/B”, *Biochim. Biophys. Acta*, **1261**: 290-292

Czyzyk-Krseska M.F., Beresh J.E., (1996), “ Characterization of the hipoxia-inducible protein binding site within the 3’ untraslated tyrosine region of the hydroxylase m-RNA”, *J. Biol. Chem.*, **271**: 3293-3299

Dallaire F., Dupuis S., Fiset S., Chabot B. (2000) “Heterogeneous nuclear ribonucleoprotein A1 and UP1 protect mammalian telomeric repeats and modulate telomere replication in vitro”, *J. Biol. Chem.*, **275**: 14509-14516.

DeMaria C.T., Sun Y., Long L., Wagner B.J., Brewer G. (1997) “Structural determinants in AUF1 required for high affinity binding to A+U-rich elements”, *J. Biol. Chem.*, **272**: 27635-27643.

Dempsey L.A., Li M., DePace A., Bray-Ward P., Maizels N. (1998) “The human HNRPD locus maps to 4q21 and encodes a highly conserved protein”, *Genomics*, **49**: 378-384.

DePinho R.A., Rudolph K.L., Chang S., Lee H.W., Blasco M., Gottlieb G.J., Greider C., (1999), "Longevity, stress response, and cancer in aging telomerase-deficient mice", *Cell*, **96**: 701-712

Ding J., Hayashi M.K., Zhang Y., Manchel L., Krainer A.R., Xu R.-M., (1999), "Crystal structure of two RRM domains of hnRNP A1 (UP1) complexed with single stranded telomere DNA", *Genes Dev.*, **13**: 1102-1115

Dreyfuss G., Matunis M.J., Piñol-Roma S., Burd C.G. (1993) "HnRNP proteins and the biogenesis of mRNA", *Annu. Rev. Biochem.*, **62**: 289-321.

Eid J.E., Sollner-Webb B. (1995) "ST-1, a 39 kilodalton protein in *Trypanosoma brucei*, exhibits a dual affinity for the duplex form of the 29 base pair subtelomeric repeats and its C-rich strand", *Mol. Cell. Biol.*, **15**: 389-397.

Fang G., Cech T.R. ed. (1995), *Telomeres*. Cold Spring Harbor, NY: Cold Spring Harbor Lab. Press.

Fang G., Gray J.T., Cech T.R. (1993) "*Oxytricha* telomere-binding protein: separable DNA-binding and dimerization domains of a subunit", *Genes Dev.*, **7**: 870-882.

Feng J., Funk W.D., Wang S. S., Weinrich S. L., Avilín A. A., Chiu C. P., Adams R. R., Chang E., Allsopp R. C., Jinghua Y., West M. D., Harley C. B., Villepoteau B., (1995) "The RNA component of human telomerase", *Science*, **269**: 1236-1241

Garcia-Basset I., Ortiz M., Pagans S., Romero A., Canals F., Avilés F.X., Azorín F., (1999), *Nucleic Acids Res.*, **27**: 3267-3275

Gehring K., Leroy J.-L., Guéron M., (1993), "A tetrameric DNA structure with protonated cytosine-cytosine base pairs", *Nature*, **363**: 561-565

Goldstein S. (1990) "Replicative senescence: the human fibroblast comes of age", *Science*, **249**: 1129-1133.

Görlach M., Wittekind R.A., Beckman L., Mueller L., Dreyfuss G. (1992) "Interactions of the RNA-binding domain of the hnRNP C proteins with RNA", *EMBO J.*, **11**: 3289-3295.

Gottschling D.E., Zakia V.A., (1986), "Telomere proteins: specific recognition and protection of the natural termini of *Oxitycha* macronuclear DNA" *Cell*, **47**: 195-205

Graff J., Cha J., Blyn L.B., Ehrenfelde E., (1998), "Interaction of poly(rC) binding protein 2 with the 5' noncoding region of hepatitis A virus RNA and its effects on traslation", *J. Virol.*, **72**: 9668-9675

Greider C. W., (1996) "Telomere lenght regulation", *Annu. Rev. Biochem.*, **65**: 337-365

Greider C.W., Blackburn E.H., (1985) "Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts", *Cell*, **43**: 405-413

Griffith J.D., Comeau L., Rosenfield S., Stansel R.M., Bianchi A., Moss H., de Lange T. (1999) "Mammalian telomeres end in a large duplex loop", *Cell*, **97**: 503-514.

Harley C.B., Futcher A.B., Greider C.W. (1990) "Telomeres shorten during ageing of human fibroblast", *Nature*, **345**: 458-460.

Harley C.B., Kim N.W., Prowse K.R., Weinrich S.L., Hirsch K.S., West M.D., Bacchetti S., Hirte H.W., Counter C.M., Greider C.W. (1994) "Telomerase, cell immortality and cancer", *Cold Spring Harbor. Symp. Quant. Biol.*, **59**: 307-315.

- Harley C.B., Vaziri H., Counter C.M., Allsopp R.C.** (1992) "The telomere hypothesis of cellular aging", *Exp. Geront.*, **27**: 375-382.
- Hicke B.J., Rempel R., Mahler J., Swank R.A., Hamaguchi J.R.** (1995) "Phosphorylation of the *Oxytricha* telomere protein: possible cell cycle regulation", *Nucleic Acids Res.*, **23**: 1887-1893.
- Ishikawa F., Matunis M.J., Dreyfuss G., Cech T.R.** (1993) "Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)_n", *Mol. Cell. Biol.*, **13**: 4301-4310.
- Kamada S., Miwa T.** (1992) "Protein binding to CarG box motifs and to single stranded DNA function as a transcriptional repressor", *Gene*, **119**: 229-236.
- Karlseder J., Broccoli D., Dai Y., Hardy S., de Lange T.** (1999) "p53-and ATM-dependent apoptosis induced by telomeres lacking TRF 2", *Science*, **283**: 1321-1325.
- Kiledjian M., DeMaria C.T., Brewer G., Novik K.** (1997) "Identification of AUF1 (heterogeneous nuclear ribonucleoprotein D) as a component of the alpha-globin mRNA stability complex", *Mol. Cell. Biol.*, **17**: 4870-4876.
- Kiledjian M., Wang X., Liebhaber S.A.** (1997) "Identification of two KH domain proteins in the alpha-globin mRNP stability complex", *Cell*, **89**: 597-606.
- Kim N.W., Piatyszek M.A., Prowse K.R., Harley C.B., West M.D., Ho P.L., Coviello G.M., Wright W.E., Weinrich S.L., Shay J.W.** (1994) "Specific association of human telomerase activity with immortal cells and cancer", *Science*, **266**: 2011-2014.

Kirion G., Boakye K., Lustig A., (1993), "C-terminal truncation of Rap1 results in the deregulation of telomere size, stability and function in *S.cerevisiae*", *Mol Cell. Biol.*, **12**: 5159-5173

Klingelhuntz A. J., Barber S., Smith P.P., Dyer K., McDougall J. K., (1994), "Restoration of telomeres in human papillomavirus-immortalized human anogenital epithelial cells", *Mol. Cell. Biol.*, **14**: 961-969

Krainer A.R., Conway G.C., Kozak D., (1995), *Genes Dev.*, **4**: 1158-1171

Krecic A.M., Swanson M.S. (1999) "hnRNP complexes: composition, structure and function", *Cur. Opin. Cell Biol.*, **11**: 363-371.

LaBranche H., Dupuis S., Bed-David Y., Bani M.R., Wellinger R.J., Chabot B. (1998) "Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase", *Nat. Genet.*, **19**: 199-202.

Lacroix L., Lienard H., Labourier E., Djavaheri-Mergny M., Lacoste J., Leffers H., Tazi J., Helene C., Mergny J.L. (2000) "Identification of two human nuclear proteins that recognise the cytosin-rich strand of human telomeres in vitro", *Nucleic Acids Res.*, **28**: 1564-1575.

Lane D.P., (1992), "Cancer, p53, guardian of the genome", *Nature*, **358**: 15-16

Leffers H., Dejgaard K., Celis J.E. (1995) "Characterisation of two major cellular poly(rC)-binding human proteins, each containing three K-homologous (KH) domains", *Eur. J. Biochem*, **230**:447-453.

Leroy J.L., Gueron M., Mergny J.L., Helene C. (1994) "Intramolecular folding of a fragment of the cytosine-rich strand of a telomeric DNA into an i-motif", *Nucleic Acids Res.*, **22**: 1600-1606.

Lingner J., Hughes T. R., Shevchenko A., Lundblad V., Cech T. R., (1997), *Science*, **276**: 561-565

Liu Q., Dreyfuss G. (1995) "In vivo and in vitro arginine methylation of RNA binding proteins", *Mol. Cell. Biol.*, **15**: 2800-2808.

Lustig A. J., Kurtz S., Shore D., (1990) "Involvement of silencer and USA binding protein Rap1 in regulation of telomeres length", *Science*, **250**: 659-663

Makamura T. M., Cech T. R., (1998), *Cell*, **92**: 587-590

Makarov V.L., Hirose Y., Langmore J.P., (1997), *Cell*, **88**: 657-666

Makeyev A.V., Chkheidze A.N., Liebhaber S.A. (1999) "A set of highly conserved RNA-binding proteins, α CP-1 and α CP-2, implicated in mRNA stabilization, are coexpressed from an intronless gene and its intron-containing paralog", *J. Biol. Chem.*, **274**: 24849-24857.

Manzini G., Xodo L.E., Yathindra N. (1994) "Evidence for intramolecular folded i-DNA structures in biologically relevant CCC-repeat sequences", *Nucleic Acids Res.*, **22**: 4634-4640.

Marcand S., Gilson E., Shore D. (1997) "A protein-counting mechanism for telomere length regulation in yeast", *Science*, **275**: 986-990.

Marcand S., Gilson E., Shore D., (1997), "A protein-counting mechanism for telomere length regulation in yeast", *Science*, **275**: 986-990

Marsich E., Bandiera A., Tell G., Scaloni A., Manzini G. "A chicken hnRNP, Type A/B, recognises the single-strand d(CCCTAA)_n telomeric repeated motif", in corso di stampa su *Eur. J. Biochem.*

Marsich E., Piccini A., Xodo L.E., Manzini G. (1996) "Evidence for a HeLa protein that binds specifically to the single-stranded d(CCCTAA)_n", *Nucleic Acids Res.*, **24**: 4029-4033

Marsich E., Xodo L.E., Manzini G. (1998) "Widespread presence in mammals and high binding specificity of a nuclear protein that recognises the single-stranded telomeric motif (CCCTAA)_n", *Eur. J. Biochem.*, **258**: 93-99.

Matunis E.L., Matunis M.J., Dreyfuss G. (1993) "Association of individual hnRNP proteins and snRNP with nascent transcripts, *J. Cell Biol.*, **121**: 219-228.

Matunis M.J., Michael W.M., Dreyfuss G. (1992) "Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein", *Mol. Cell. Biol.*, **12**: 164-171.

Mayeda A., Munroe S.H., Xu R.M., Kreiner A.R. (1998) "Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1", *RNA*, **4**: 1111-1123.

McKay S.J., Cooke H. (1992) "HnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGG_n", *Nucleic Acids Res.*, **20**: 6461-6464.

Merrill B.M., Stone K.L., Cobianchi F., Wilson S.M., Williams K.R. (1988) "Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic-acid binding pocket in the A1 heterogeneous nuclear ribonucleoprotein", *J. Biol. Chem.*, **263**: 3307-3313.

Meyerson M., Counter C. M., Eaton E. N., Ellisen L. W., Steiner P., Caddle S. D., Ziaugra L., Davidoff M. J., Liu Q. et al. (1997), "hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and immortalization" *Cell*, **90**: 785-795

Miau L.H., Chang C.J., Shen B.J., Tsai W.H., Lee S.C. (1998) "Identification of heterogeneous nuclear ribonucleoprotein K (hnRNP K) as a repressor of C/EBP β -mediated gene activation", *J. Biol. Chem.*, **273**: 10784-10791.

Michael W.M., Choi M., Dreyfuss G. (1995) "A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway", *Cell*, **83**: 415-422.

Michael W.M., Eder P.S., Dreyfuss G. (1997) "The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein", *EMBO J.*, **16**: 3587-3598.

Michelotti E.F., Michelotti G.A., Aronsohn A.I., Levens D. (1996) "Heterogeneous nuclear ribonucleoprotein K is a transcription factor", *Mol. Cell. Biol.*, **16**: 2350-2360.

Michelotti E.F., Tomonaga K., Krutzsch H., Levens D. (1995) "Cellular nucleic acids binding protein regulates the CT element of the human C-myc protooncogene", *J. Biol. Chem.*, **270**: 9494-9499.

Mikheev A., Micheev A.S., Zhang V., Aebersold R., Zarbl H. (2000), "Cloning of an Apobec-1 binding protein that also interacts with apolipoprotein B m-RNA and evidence for its involvement in RNA editing", *J. Biol. Chem.*, **272**: 1452-1455

Moretti P., Freeman K., Coodly L., Shore V. (1994), "Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein Rap1", *Genes Dev.*, **8**: 2256-2269

Morin G.B. (1989), "The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats", *Cell*, **59**: 521-529

Musco G., Stier G., Joseph C., Castiglione Morelli M.A., Nilges M., Gibson TJ, Pastore A. (1996) "Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome", *Cell*, **19**: 237-245.

Nagai K., Oubridge C., Jessen T.M., Li J., Evans P.R. (1990) "Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A", *Nature*, **346**: 515-520.

Nakamaia T. M., Morin G. B., Chapman K. B., Weinrich S. L., Andrews W. H., Linger J., Haeley C. B., Cech T. R., (1997), *Science*, **277**: 955-959

Nakielny S., Dreyfuss G.(1997) "Nuclear export of proteins and RNAs", *Curr. Opin. Cell Biol.*, **9**: 420-429.

Ostareck D.H., Ostareck-Lederer A., Wilm M., Thiele B.J., Mann M., Hentze M.W. (1994) "mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end", *Nature*, **368**: 867-871.

Ostareck-Lederer A., Ostareck D.H., Hentze M.W. (1998) "Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2", *TIBS*, **23**: 409-411.

Price C.M. (1990) "Telomere structure in *Euplotes crassus*: characterization of DNA-protein interactions and isolation of a telomere-binding protein", *Mol. Cell. Biol.*, **10**: 3421-3431.

Price C.M., Cech T.R., (1987), "Reversing time: origin of telomerase", *Cell*, **92**: 587-590

Reddel R.R. (1998) "A reassessment of the telomere hypothesis of senescence", *Bioessay*, **20**: 977-984.

Sarig G., Weisman-Shomer P., Fry M. (1997) "Telomeric and tetraplex DNA binding properties of qTBP42: a homologue of the CarG box binding protein CBF-A", *Biochem. Biophys. Res. Com.*, **237**: 617-623.

Shamoo J., Krueger V., Rice L.M., Williams K.R., Steitz T.A., (1997), "Crystal structure of two RNA binding domains of human hnRNP A1 at 1,75 Å resolution", *Nat. Struct. Biol.*, **4**: 215-222

Shampay J., Szostak J. W., Blackburn E.H., (1984) "DNA sequences of telomeres maintained in yeast", *Nature*, **310**: 154-157

Shay J.W., Bacchetti S. (1997) "A survey of telomerase activity in human cancer", *Eur. J. Cancer*, **33**: 787-791.

Shoppen-Lentz D., Blackburn E.H., (1989), "Telomere terminal transferase activity from *Euplotes crassus* adds large numbers of TTTTGGGG repeats onto telomeric primers" *Mol. Cell. Biol.*, **9**: 2761-2764

Singh R., Valcarcel J., Green M.R., (1995), "Distinct binding specificities and functions of higher polypyrimidine tract-binding protein", *Science*, **269**: 1173-1176

Smidt M.P., Russchen B., Snippe L., Wijnholds J., AB G. (1995) "Cloning and characterisation of a nuclear, site specific ssDNA binding protein", *Nucleic Acid Res.*, **13**: 2389-2395.

Smith S., Giriat L., Schmitt A., de Lange T. (1998) "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres", *Science*, **282**: 1484-1487.

Stefanovic, B., Hellerbrand C., Holcic M., Briedl M., Liebhaber S.A., Brenner D.A., (1997), "Postranscriptional regulation of collagen $\alpha 1$ m-RNA in hepatic stellate cells", *Mol Cell. Biol.*, **17**: 5201-5209

Swanson M.S. (1995) "Functions of nuclear pre-mRNA/mRNA binding proteins", in pre-mRNA Processing, ed. A.I. Lamond, pp. 17-33.

Tolnay M., Baranji L., Tsokos G.C. (2000) "Heterogeneous nuclear ribonucleoprotein D0 contains transactivator and DNA-binding domains", *Biochem. J.*, **348**: 151-158.

van Steensel B., de Lange T. (1997) "Control of the telomere length by the human telomeric protein TRF1", *Nature*, **385**: 740-743.

van Steensel B., Smogorzewska A., de Lange T. (1998) "TRF2 protects human telomeres from end-to-end fusions", *Cell*, **92**: 401-413.

Visa N., (1996), "A pre-mRNA-binding protein accompanies the RNA from the gene through the nuclear pore and into polysomes", *Cell*, **84**: 253-364

Wagner B.J., DeMaria C.T., Sun Y., Wilson G.M., Brewer G.(1998) "Structure and genomic organization of the human AUF1 gene: alternative pre-mRNA splicing generates four protein isoforms", *Genomics*, **48**: 195-202.

Wang W., Skopp R., Scofield M., Price C. (1992) "*Euplotes crassus* has genes encoding telomere-binding proteins and telomere-binding proteins homologs", *Nucleic Acids Res.*, **20**: 6621-6629.

Watson J.D. (1972) "Origin of concatameric T7 DNA", *Nature New Biol.*, **239**: 197-201.

Weinrich S. L., Pruzan R., Ma L., Oulette M., Tesmer V. M., Holt S. E., Bodnar A. G., Kim N. W., trager J. B. et al., (1997) "Reconstitution of human telomerase with template RNA component hTR and the catalytic protein subunit hTRT", *Nat. Genet.*, **17**: 498-502

Wellinger R.J., Ethier K., Labrecque P., Zakian V.A., (1996), *Cell*, 85: 423-433

Weng Z., Thomas S.M., Rickles R.J., Taylor J.A., Braver A.W., Seidel-Dugan C., Michael W.M., Dreyfuss G., Brugge J.S. (1994) "Identification of Src, Fyn and Lyn SH3-binding proteins: implications for a function of SH3 domains", *Mol. Cell. Biol.*, 14: 4509-4521.

Williamson J.R., Raghuraman M.K., Cech T.R. (1989) "Monovalent cation induced structure of telomeric DNA: the G-quartet model", *Cell*, 59: 871-880.

Wittekind M., Görlach M., Friedrichs M., Dreyfuss G. (1992) "¹H, ¹³C, ¹⁵N NMR assignments and global folding pattern of the RNA-binding domain of the human hnRNP C proteins", *Biochem.*, 31: 6254-6265.

Wright W.E., Tesmer V.M., Huffman K.E., Levene S.D., Shay J.W. (1997) "Normal human chromosomes have long G-rich telomeric overhangs at one end", *Genes Dev.*, 11: 2801-2809.

Zahler A.M., Prescott D.M., (1989) "DNA primase and the replication of telomeres in *Oxytricha nova*", *Nucleic Acids Res.*, 17: 6299-6317

Zahler A.M., Williamson J.R., Cech T.R., Prescott D.M. (1991) "Inhibition of telomerase by G-quartet Dna structures", *Nature*, 350: 718-729.

Evidence for a HeLa nuclear protein that binds specifically to the single-stranded d(CCCTAA)_n telomeric motif

Eleonora Marsich, Antonella Piccini¹, Luigi E. Xodo and Giorgio Manzini*

Department of Biochemistry, Biophysics and Macromolecular Chemistry, University of Trieste, Via Giorgieri 1, I-34127 Trieste, Italy and ¹ICGEB, Area Science Park, Padriciano 99, I-34012, Trieste, Italy

Received June 6, 1996; Revised and Accepted September 3, 1996

ABSTRACT

In recent years several telomere binding proteins from eukaryotic organisms have been identified that are able to recognise specifically the duplex telomeric DNA repeat or the G-rich 3'-ending single strand. In this paper we present experimental evidence that HeLa nuclear extracts contain a protein that binds with high specificity to the single-stranded complementary d(CCCTAA)_n repeat. Electrophoretic mobility shift assays show that the oligonucleotide d(CCCTAACCC-TAACCCCTAACCCCT) forms a stable complex with this protein in the presence of up to 1000-fold excesses of single-stranded DNA and RNA competitors, but is prevented from doing so in the presence of its complementary strand. SDS-PAGE experiments after UV cross-linking of the complex provide an estimate of 50 kDa for the molecular weight of this protein.

INTRODUCTION

Interest in the functional role as well as the structural aspects of telomeric DNA repeats in eukaryotic chromosomes has grown very rapidly in recent years. From the functional point of view, the quest for a deeper understanding of the role of telomeric DNA was boosted by the discovery of its peculiar replication mechanism through the telomerase enzyme (1,2) and of the tissue type dependent variation in telomerase activity in normal as well as tumour tissues (3-5). From the structural point of view, questions have been raised whether the unusual secondary structures that each of the complementary telomeric strands is capable of adopting *in vitro* may be involved also in telomere functioning *in vivo*. In particular since the discovery of a number of inter- and intramolecular structures of short guanine runs, like those in the 3'-ending strand of telomeric DNA, which are based on planar guanine tetrads arranged in cycle through Hoogsteen hydrogen bonds (6-9). In addition also C-block repeat sequences, of which the other strand of the telomere repeat is a representative, have been shown to be able to adopt the self-intercalated tetra-stranded structure i-DNA (10,11), although a slightly acidic pH value is required in this case.

Along with the intense experimental work to better characterise the telomerase enzymes both in unicellular organisms, such as ciliates, and in higher vertebrates, several research groups have directed their efforts to discover and characterise other nuclear proteins capable of binding specifically to telomeric DNA repeat sequences. Proteins have been identified that specifically recognise double-stranded telomeric repeats, such as Rap1p in yeast (12,13), PPT from *Physarum* (14) and TRF in mammalian cells (15,16); as well as proteins that bind to the single-stranded guanine-rich 3'-ending motif, such as $\alpha\beta$ protein from *Oxytricha* (17), TBP from *Euplotes* (18), TEP and TGP from *Tetrahymena* (19,20), GBP from *Chlamydomonas* (21), XTEF from *Xenopus* (22), an avian factor (23), MyoD from mammalian cells (24) and A2/B1 from HeLa (25). Some of these have been shown to bind to strands arranged in the G-DNA tetradic structure (20,24) or even to promote its formation *in vitro* (12,13).

In recent years evidence has accumulated which indicates the presence in HeLa cells of a specific binding activity toward telomeric-type DNA sequences. Ishikawa *et al.* (25) have found nuclear proteins able to bind to the single stranded (TTAGGG)_n motif and identified them as components of the hnRNP complex. Although a role of these components in telomere functioning can be envisaged, it appears that the primary reason for this specific binding stems from the rather strict similarity of the telomeric repeat sequence with the 3' splice site consensus of hnRNA. Similarly TRF, a protein binding specifically to the double-stranded telomeric repeats, has been discovered by Zhong *et al.* (15), and subsequently cloned and shown to be physically linked to the telomeres of metaphase chromosomes (16).

We report in this paper experimental evidence of the presence in HeLa nuclear extracts of a specific binding activity directed to the complementary C-rich telomeric sequence when present as a single strand.

MATERIALS AND METHODS

Samples

The HeLa nuclear protein extract was obtained according to the method of Dignam (26), and the concentration of the stock solution was determined by the Lowry assay.

* To whom correspondence should be addressed

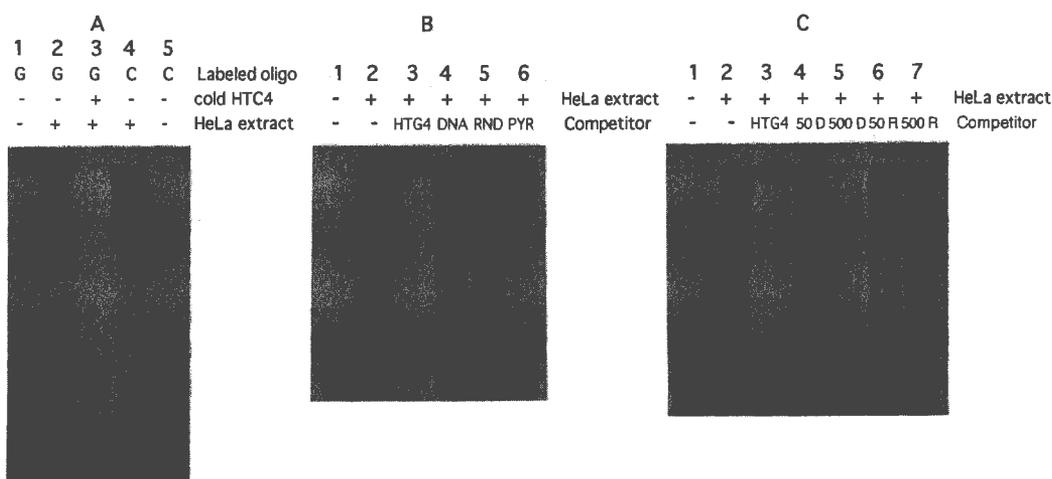


Figure 1. Electrophoretic mobility shift assays. (A) Lane 1, labeled HTG4 (G); lane 2, labeled HTG4 incubated with HeLa nuclear extract; lane 3, labeled HTG4 + excess HTC4 incubated with HeLa nuclear extract; lane 4, labeled HTC4 (C) incubated with HeLa nuclear extract; lane 5, labeled HTC4. (B) Labeled HTC4: In the absence (-) or presence (+) of: HeLa nuclear extract, and of cold HTG4 or 100-fold excess of competitors: denatured *E. coli* DNA, RND oligonucleotide, PYR oligonucleotide. (C) Labeled HTC4: In the absence (-) or presence (+) of: HeLa nuclear extract, and cold HTG4 or 50- and 500-fold excess of competitors: denatured *E. coli* DNA, total yeast RNA.

The following oligonucleotides were synthesized by an Applied Biosystem apparatus through phosphoramidite chemistry and purified according to standard methods: d(CCCTAACCTAAC-CCTAACCT) (HTC4); d(AGGGTTAGGGTTAGGGTTAGGG) (HTG4); d(AACCCCTGCATTGAACTCCA) (RND); d(CTTTC-TTCCCTTCCCTTTC) (PYR).

The oligonucleotides used for electrophoretic experiments were radiolabeled with [γ - 32 P]ATP and T4 polynucleotide kinase according to standard procedures (27): generally 2–4 pmol of each DNA strand were labeled with 10 μ Ci of high activity ATP.

Electrophoretic mobility shift assays

Aliquots of 0.1 pmol of each oligonucleotide were incubated in 10 μ l volumes (10 mM Tris buffer, pH 8, 50 mM KCl, 0.1 mM EDTA) with 4 μ g of nuclear proteins at 5 or 25°C for the prescribed time, alone or in the presence of an excess of cold complementary strand or competitor DNA (and RNA). In the case of the cold complementary strand, the duplex was allowed to anneal for several hours at room temperature prior to its addition to the incubation mixture. In the case of competitors, the radiolabeled telomeric strand was added last in the incubation mixture, allowing time for possible interactions between competitor and protein extract to occur. After incubation, the samples were loaded onto a non denaturing polyacrylamide gel [10% acrylamide (bisacrylamide 1:30); same buffer as for incubation] and run at 10 V/cm at about 15°C (running time 75 min, corresponding to ~5 cm migration for the unbound oligonucleotide). Finally the gel was dried and autoradiographed.

UV cross-linking experiments

Incubation mixtures, in parallel with those used in gel retardation assays, were prepared and irradiated with a UV lamp (50 W) for 60 min. Then the samples were loaded onto a standard SDS-PAGE and run. The samples were then run on a standard SDS-PAGE with molecular weight markers (HMW calibration kit, Pharmacia), which were subsequently submitted to blue Coomassie staining,

to allow an estimate of the molecular weights of the cross-linked complexes. After the run, the gels were dried and autoradiographed.

RESULTS

In search of a protein within HeLa nuclear extract capable of specifically recognizing the single-stranded telomeric repeat, we have conducted a number of electrophoretic mobility shift experiments. Figure 1A shows that, after incubation of labeled HTC4 with the protein extract, two well resolved retarded bands are observed (lane 4), whose mobilities are slightly higher and slightly lower than that of the band observed in the case of HTG4 (lane 2), respectively. This one very likely corresponds to the complex already observed by Ishikawa *et al.* (25). Incubation of labeled HTG4 with an excess of its complementary HTC4, that converts the radiolabeled probe in the standard double strand DNA, before the addition of the protein extract, substantially suppresses every retarded band: no double-stranded DNA binding protein is detected in this experiment (lane 3).

To evaluate the specificity of this effect, a number of competition experiments have been done. Figure 1B and C illustrates the results. Figure 1B shows that the addition of 100-fold excesses of denatured *Escherichia coli* DNA (lane 4), of the RND 20mer (lane 5), or of the PYR 18mer (lane 6), to the protein extract before labeled HTC4, fail to show any competing effect. It is worth noting that the sequence RND is very similar to HTC4 in base composition, resembling a scrambled HTC4. In Figure 1C the competing effects of 50- and 500-fold excesses of denatured *E. coli* DNA and yeast total RNA are shown: at 50-fold excesses (lane 4, DNA, and lane 6, RNA) no substantial displacement of labeled HTC4 from the protein complexes is found. However 500-fold excesses of *E. coli* DNA (lane 5) and yeast RNA (lane 7) induce an almost complete disappearance of the slowest band in the first case and its attenuation in the second, but the fastest band is not affected at all even at these large excesses of competitors. Once more, incubation of labeled HTC4 with its complementary strand prior to protein addition suppresses any band retardation effect (lane 3): in these experiments however, radioactivity is

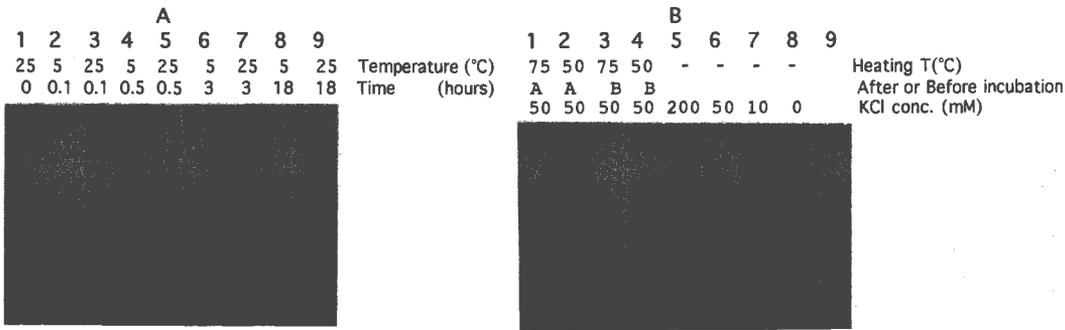


Figure 2. (A) EMSA of mixtures of HeLa extract and HTC4 incubated at 5 or 25°C for 5, 30, 180 and 1080 min, in the presence of a 500-fold excess of denatured *E.coli* DNA. (B) EMSA of mixtures of HeLa extract and HTC4 incubated for 3 h at 25°C in the presence of a 500-fold excess of denatured *E.coli* DNA, and 10 mM Tris, pH 8: at various concentrations of KCl (lanes 5–8), HeLa extract heated at 50°C (lanes 2 and 4) or at 75°C (lanes 1 and 3), before (lanes 3 and 4) or after (lanes 1 and 2) incubation with HTC4; no HeLa extract (lane 9).

observed in the well: apparently high molecular weight aggregates form between the duplex and some of the protein components.

The effects of the duration and temperature of incubation have been investigated and the results are presented in Figure 2A. The labelled oligos were incubated in the presence of 500-fold excesses of denatured *E.coli* DNA to suppress the slower, less specific, retarded band. After 5 min incubation at either 25 or 5°C a barely perceptible complex was seen; 30 min were sufficient at 25°C, but not at 5°C, to produce a strong retarded band, in accordance with a typical behaviour of slower kinetics at lower temperatures; finally, no significant differences were found between 25 and 5°C at incubation times of ≥ 3 h. The effect of temperature on DNA binding has been further investigated by heating the protein extract at 50 and 75°C, before and after incubation at 25°C with HTC4. The results of Figure 2B, lanes 1–4, show that, independent of the temporal order of heating and incubation, heating at 50°C weakens the retarded bands, whereas heating at 75°C suppresses them completely. Clearly, at 75°C the protein denatures, irrespective of the presence of HTC4, and does not renature on cooling back to room temperature. Since some complex can survive, or can form after heating at 50°C, the denaturation temperatures of the protein and of its complex with HTC4 cannot be far from this temperature. In lanes 5–8 of Figure 2B the effect of salt concentration on complex formation is presented: varying the concentration of KCl in the incubation mixture from 0 to 200 mM is almost irrelevant to the intensity of the retarded band. This observation indicates that the electrostatic component of the energy of complex formation has only a marginal role, if any.

To reach at least an approximate value of the molecular weights of the nuclear factors that produce the retarded bands, a number of SDS–PAGE gels were run in parallel to the band shift assays, after UV irradiation of the incubation mixtures. Figure 3 presents the UV cross-linking effect on aliquots of the same incubation mixtures whose band shift results are shown in Figure 1B. Lane 2 (labelled HTC4 plus protein extract) exhibits a more intense slower band and a doublet of faster bands. Lanes 4, 5 and 6 show that 100-fold excesses of denatured *E.coli*, RND 20mer and PYR 18mer do not produce any major competing effect, in agreement with the results of Figure 1B. No competition effect has been produced in an analogous cross-linking experiment in the presence of up to 500-fold excess of total yeast RNA (data not shown).

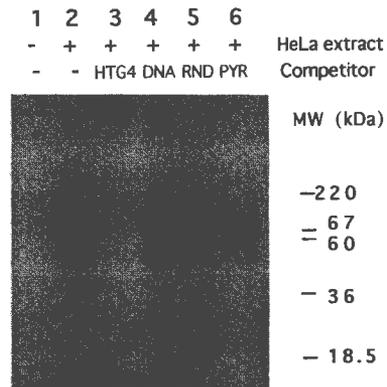


Figure 3. SDS–PAGE of HeLa nuclear extract after UV cross-linking in the presence of labeled HTC4, in the absence (–) or presence (+) of: HeLa nuclear extract, and of cold HTG4 or 100-fold excess of competitors: denatured *E.coli* DNA, RND oligonucleotide, PYR oligonucleotide. On the right, protein molecular weight markers obtained from an adjacent lane on the same gel, cut and stained with Coomassie blue.

Comparison with the mobilities of the protein MW markers run in adjacent lanes of the same gel, and stained with Coomassie blue, suggests apparent molecular weights of 70 kDa for the slow component, and 50–55 kDa for the faster doublet, which however is not always resolved. Clearly these data are rough estimates, since the cross-linked adduct comprises a 22mer oligonucleotide, which adds 7–8 kDa to the protein, as well as some 20 negative charges.

Figure 4 shows analogous results obtained with both HTC4 and HTG4, in the absence and in the presence of competitor *E.coli* DNA. The faster doublet of HTC4 (lane 1, absence of competitor) is not well resolved, as in Figure 3, and the slower one is again the stronger; anyway it can be seen that 100-fold competitor has no effect (lane 2) on either, whereas 1000-fold competitor almost suppresses the slower component, leaving the faster intact.

Lanes 5, 6 and 7 show the corresponding results obtained in the case of the labeled complementary strand HTG4: apart from some minor signals, an intense doublet is present, which is weakly but clearly competed at 100-fold excess of *E.coli* DNA, and strongly competed if not abolished at 1000-fold excess. The apparent

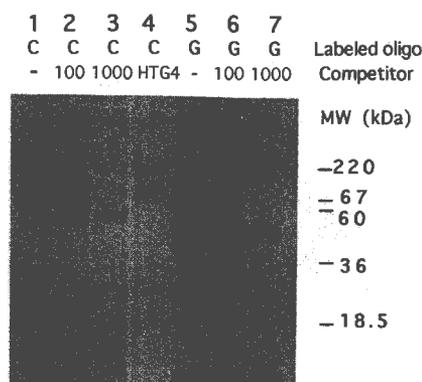


Figure 4. SDS-PAGE of HeLa nuclear extract after UV cross-linking in the presence of labeled HTC4 (C) or HTG4 (G), in the absence (-) or presence of: 100- and 1000-fold excess of denatured *E. coli* DNA. In the central lane labeled HTC4 was preincubated with HTG4, before extract addition. On the right, protein molecular weight markers obtained from an adjacent lane on the same gel, cut and stained with Coomassie blue. The full triangle marks a tiny band not removed by 1000-fold excess denatured *E. coli* DNA (see text).

molecular weight of the doublet results in the range of 40–45 kDa, are in accord with the molecular weights observed by Ishikawa *et al.* (25) for their major fraction B components, once the presence of a 22mer oligonucleotide in the complex is accounted for. Most of the minor signals are suppressed by 1000-fold excess *E. coli* DNA, but not that of ~30 kDa apparent MW. Could this be a candidate for a single stranded (TTAGGG)_n binding protein more sequence-specific than the hnRNPs quoted above?

Irradiating the mixture of the nuclear extract with the duplex telomeric repeat (labeled HTC4 preincubated with HTG4) suppresses the bands observed with the single strand, producing however a different set of very faint bands (lane 4 of Fig. 4), the most conspicuous of which has an apparent molecular weight of about 70 kDa, very near to that of TRF (16), once the cross-linked DNA fragment is taken into account.

To further support the specificity of the observed interaction, analogous band shift and UV cross-linking experiments have been performed with labeled RND 20mer: no retarded or cross-linked specific band has been revealed (data not shown).

DISCUSSION

In a typical gel retardation assay about 0.1–0.2 pmol of labeled HTC4, incubated with 4 µg of the HeLa protein extract, were run in each lane. Since at least half of the radioactivity is present in the retarded bands, the amount of bound protein can be estimated to be of the same order of magnitude of the oligonucleotide. With the molecular weight estimates from the UV cross-linking experiments and assuming a 1:1 stoichiometry for the DNA–protein complex, this means a few ng of bound proteins out of 4 µg, or 0.1% of the nuclear extract, which however does not contain histones. Less than half of this amount is attributable to the faster component, i. e. that not displaced even by a 1000-fold denatured *E. coli* DNA. If our nuclear extract contains some 10% of the total nuclear proteins, 0.05% of this amount corresponds roughly to few femtograms of protein per cell, not much higher than the amount of telomeric DNA (28).

From the UV-cross-linking experiments of Figure 4 a much higher amount of radioactivity is found in the HTG4–protein

complexes than in the HTC4–protein complexes. This observation could suggest that the faster component that binds HTC4 is at least one order of magnitude less abundant than those recognising HTG4 [i.e. the hnRNPs observed by Ishikawa *et al.* (25)], if the efficiency of UV cross-linking were the same for the two systems. However this assumption cannot be taken for granted.

As to the specificity of the interaction, some semiquantitative considerations can be based on the *E. coli* DNA competition experiments. The vertebrate telomeric motif TTAGGG and its complementary are not overrepresented in the *E. coli* genome: a brief search for the presence of the representative nonamers GGGTTAGGG and CCCTAACCC in the *E. coli* sequences in GenBank has found these strings at a frequency very near to that statistically expected, i.e. <1 every 100 000 nt. If the hnRNP components binding to HTG4 owe this activity to the fact that its repeat conforms to the 3' splicing consensus, ...YAG/G..., the statistical expectancy of this consensus in *E. coli* DNA could be roughly estimated to be 2 in every 2444 nt, i.e. 1 every 50–100 nt. This agrees with the observation that a 100-fold and a 1000-fold excess of denatured *E. coli* DNA approximately remove half and 90% of the protein bound radioactivity, respectively (lanes 6 and 7 of Fig. 4). Almost the same argument holds for the slower component of the HTC4 binding proteins: its limited specificity for HTC4 is likely fortuitous. In contrast, the faster component is not competed even by a 1000-fold excess of *E. coli* DNA, suggesting that its specificity should be considerably higher than that for a full, and exact, telomeric repeat. Indeed the statistical expectancy for the hexameric CCCTAA sequence, or for each of its cyclic permutations, can be estimated to be 2 every 4⁶ nt, or ~1 every 2000 nt: if one hexameric stretch was sufficient for full recognition by the protein, a 1000-fold excess of *E. coli* DNA would have produced a remarkable competing effect.

It has been shown that (CCCTAA)₄, in analogy to similar C-block repeats, is able to adopt an unusual secondary structure, the i-DNA (10,11) at pH values of ≤ 6. The specific complex of HTC4 with the nuclear protein described here forms at pH values of 7–8, at which HTC4 does not adopt the i-DNA structure by itself, apart from the very tiny amounts expected on the basis of the pH-dependent equilibrium between the structured and unstructured forms. It is very likely that this protein simply binds the unstructured strand and i-DNA is not involved. However at the moment one cannot rule out the possibility that the protein binds the i-DNA structure, paying for the free energy cost of stabilising it at neutral or slightly basic pH through the interaction.

Clearly the mobilities of the shifted bands in the gel retardation experiments cannot provide meaningful information about the native forms of these DNA binding proteins, and the SDS-PAGES of the cross-linked complexes provide only approximate information on the molecular weights at the subunit level. The fact that two bands are found in the mobility shift assays with the faster being much more sequence-specific, and that the same pattern is observed in the UV-cross-linking experiments, suggests that two different proteins capable of binding HTC4 exist in nondenaturing conditions. If the slower shifted band were due to the binding of two units of the same protein to one HTC4 molecule, *E. coli* DNA would not compete with it, as it does not compete with the faster band.

In conclusion, it can be argued that, if the protein responsible for the slow retarded band exhibits a rather moderate and probably fortuitous specificity in its interaction with the CCCTAA motif, the much higher specificity of the faster component indicates it as

a promising candidate to join the already identified telomere binding proteins. Work is in progress to isolate and characterise this protein.

ACKNOWLEDGEMENTS

This work has been supported by grants from the Italian MURST and from the Italian CNR (National Research Council). *Escherichia coli* DNA sequence analysis has been performed thanks to Intelligenetics software and the GenBank.

REFERENCES

- 1 Greider, C.W. and Blackburn, E.H. (1985) *Cell* **43**, 405–416.
- 2 Morin, G.B. (1989) *Cell* **59**, 521–529.
- 3 Counter, C.M., Hirte, H.W., Bachetti, S., and Harley, C.B. (1994) *Proc. Natl Acad. Sci. USA* **91**, 2900–2904.
- 4 Kim, N.W., Prowse, K.R., Ho, P.L.C., Coviello, G.M., Harley, C.B., Piatyszek, M.A., Shay, J.W., Wright, W.E. and Weinrich, S.L. (1994) *Science* **266**, 2011–2015.
- 5 Broccoli, D., Young, J.W. and de Lange, T. (1995) *Proc. Natl Acad. Sci. USA* **92**, 9082–9086.
- 6 Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) *Cell* **59**, 871–880.
- 7 Sundquist, W.I. and Klug, A. (1989) *Nature* **342**, 825–829.
- 8 Kang, C.H., Zhang, X., Ratliff, R., Moyzis, R. and Rich, A. (1992) *Nature* **356**, 126–131.
- 9 Schultze, P., Smith, F.W. and Feigon, J. (1994) *Structure* **2**, 221–233.
- 10 Leroy J.L., Guéron, M., Mergny J.L. and Hélène, C. (1994) *Nucleic Acids Res.* **22**, 1600–1606.
- 11 Manzini, G., Yathindra, N. and Xodo, L.E. (1994) *Nucleic Acids Res.* **22**, 3634–3640.
- 12 Giraldo, R. and Rhodes, D. (1994) *EMBO J.* **13**, 2411–2420.
- 13 Giraldo, R., Suzuki, M., Chapman, L. and Rhodes, D. (1994) *Proc. Natl Acad. Sci. USA* **91**, 7658–7662.
- 14 Coren, J.S., Epstein, E.M. and Vogt, V.M. (1991) *Mol. Cell. Biol.* **11**, 2282–2290.
- 15 Zhong, Z., Shiue, L., Kaplan, S. and de Lange, T. (1992) *Mol. Cell. Biol.* **12**, 4834–4843.
- 16 Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanisch, J., Tempst, P. and de Lange, T. (1995) *Science* **270**, 1663–1667.
- 17 Raghuraman, M.K. and Cech, T.R. (1989) *Cell* **59**, 719–728.
- 18 Price, C.M. (1990) *Mol. Cell. Biol.* **10**, 3421–3431.
- 19 Sheng, H., Hou, Z., Schierer, T., Dobbs, D.L. and Henderson, E. (1994) *Mol. Cell. Biol.* **15**, 1144–1153.
- 20 Schierer, T. and Henderson, E. (1994) *Biochemistry* **33**, 2240–2246.
- 21 Petracek, M.E., Konkol, L.M.C., Kable, M.L. and Berman, J. (1994) *EMBO J.* **13**, 3648–3658.
- 22 Cárdenas, M.E., Bianchi, A. and de Lange, T. (1993) *Genes Dev.* **7**, 883–894.
- 23 Gualberto, A., Patrick, R.M. and Walsh, K. (1992) *Genes Dev.* **6**, 815–824.
- 24 Weisman-Shomer, P. and Fry, M. (1993) *J. Biol. Chem.* **268**, 3306–3312.
- 25 Ishikawa, F., Matunis, M.J., Dreyfuss, G. and Cech, T.R. (1993) *Mol. Cell. Biol.* **13**, 4301–4310.
- 26 Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
- 27 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 28 Tommerup, H., Dousmanis, A. and de Lange, T. (1994) *Mol. Cell. Biol.* **14**, 5777–5785.

Table 1. Oligodeoxynucleotides used in this study. Binding specificity was determined by EMSA. +, specific retarded band present; +/-, smeared band; - no retarded band.

Name	No. of nucleotides	Sequence (5'→3')	Specific binding
HTC4	22	CCCTAACCCCTAACCCCTAACCCCT	+
HTC3	16	CCCTAACCCCTAACCCCT	-
HTC2	10	CCCTAACCCCT	-
HTA4	21	TAACCCTAACCCCTAACCCCTAA	+/-
HTR2	22	ACTGGCCCTAACCCCTAATGACT	+/-
ATC4	25	CCCTAAACCCCTAAACCCCTAAACCCCT	+
BTC4	18	CCTAACCTAACCTAACCT	-
TTC4	22	CCCCAACCCCAACCCCAACCC	+
PTC4	22	GCCTAAGCCCTAAGCCCTAAGCCCT	- ^a
ΨTC4	22	TCCCAATCCCAATCCCAATCCC	+
HTC6	36	CCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAA ^b	+
kRas	30	GCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC	- ^a
HTG4	22	AGGGTTAGGGTTAGGGTTAGGG	- ^a
CArG	22	CCATTCCTAATTAGGTAAAAG	- ^a
C22	22	CCCCCCCCCCCCCCCCCCCC	-

^a Retarded band with different mobility and specificity.

^b This oligonucleotide was biotinylated at the 3'-end for protein purification purposes.

The oligodeoxynucleotides (Table 1) were synthesised by an Applied Biosystem apparatus by phosphoramidite chemistry and purified by chromatography, or directly purchased from Pharmacia. The oligonucleotides were labelled with [γ -³²P]ATP and T4 polynucleotide kinase according to standard procedures [20]. Generally 2 pmol of each oligomer was labelled with 10 μ Ci of high-activity ATP. We observed a reproducible low efficiency of labelling by kinase T4 in the cases of oligonucleotides with cytidine as the 5' starting nucleoside.

Electrophoretic mobility shift assays. Aliquots of 0.1 pmol of each oligonucleotide tested were incubated in 20- μ l volumes (50 mM Tris, pH 8, 50 mM KCl, 1 mM EDTA) with 2 to 10 μ g nuclear proteins at 5°C overnight, in the absence or in the presence of an excess of cold competitor, usually denatured *Escherichia coli* DNA, and in some cases, an equimolar amount of one of the other oligonucleotides listed in Table 1. In the presence of competitors the radiolabelled telomeric strand was added last to the incubation mixture, allowing time for possible interactions between competitor and protein extract to occur. After incubation the samples were loaded onto a non-denaturing polyacrylamide gel (usually 10% acrylamide, bisacrylamide 1:30; same buffer as for incubation) and run at 10 V/cm and at about 15°C (running time 75 min, corresponding to about 5 cm migration for the unbound oligonucleotide). Finally the gel was dried and autoradiographed.

Ultraviolet-cross-linking experiments. Incubation mixtures, prepared as those employed for electrophoretic mobility shift assays (EMSAs), were irradiated with an ultraviolet lamp (50 W) for 45 min. Then the samples were subjected to SDS/PAGE. Molecular mass markers (HMW calibration kit, Pharmacia) were loaded in adjacent lanes and subsequently submitted to blue Coomassie staining in order to allow an estimate of the molecular masses of the cross-linked complexes. After electrophoresis, the gel portions with the lanes containing the labelled oligonucleotide/protein complexes were dried and autoradiographed.

Gel permeation chromatography. The retention volumes of the nuclear proteins from the HL60 and rat liver extracts, ultraviolet cross-linked with labelled oligonucleotide HTC4 in the presence of a 500-fold excess of denatured *E. coli* DNA to suppress unspecific binding, were determined. Small aliquots of the samples were eluted through a gel permeation chromatogra-

phy (GPC) column (1.2 cm \times 60 cm, Sepharose CL6B, Pharmacia) at 25°C and at a rate of 0.24 ml/min. The elution buffer consisted of 10 mM Tris, pH 8, 10 mM KCl, 1 mM EDTA. Fractions of 0.72 ml were collected and monitored for radioactivity or, in the case of the reference proteins, for ultraviolet absorption at 280 nm. The excluded volume (19 ml) was determined in a separate elution of high molecular mass DextranBlue.

RESULTS

Presence of the protein in nuclear extracts from different sources. The (CCCTAA)_n repeat appears to be present in the chromosomal ends of every vertebrate species so far tested. Hence, if the HeLa protein described in this laboratory [18] is involved at some stage in the telomere functioning, its homologues can be expected to be widespread in vertebrate species, although not necessarily expressed in all tissues. In order to test this hypothesis, we performed a series of EMSA experiments after incubation of the labelled single-stranded oligonucleotide HTC4 (the sequence is shown in Table 1) with nuclear protein extracts from several sources. We consistently found a retarded band with the same mobility and binding specificity as that observed with HeLa extracts. Fig. 1 shows the EMSA profiles obtained by incubating HTC4 with extracts from human HL60 and CEM cells, hamster CHO cells, rat liver, pig liver, and chicken erythrocytes. In every case more than one retarded band, sometimes not well resolved, was detected in the absence of single-stranded competitor DNA. In the presence of a 500-fold excess of denatured *E. coli* DNA, only the highest mobility band persisted, and sometimes was even stronger. In preliminary EMSAs with extracts from calf liver the same behaviour was found, although the specific retarded band appeared much fainter (data not shown). Since a major feature of HTC4 is its high C content, the formation of its complexes with HeLa and rat liver extracts was challenged also with up to 500-fold excess of the homo-oligomer C22 (Table 1). No competition effect was observed, in this case even towards the lower mobility bands, which are displaced by denatured *E. coli* DNA (data not shown).

Influence of the number of repeats on protein binding. We have already shown that this protein recognises the single-stranded telomeric sequence HTC4 and binds to it even in the

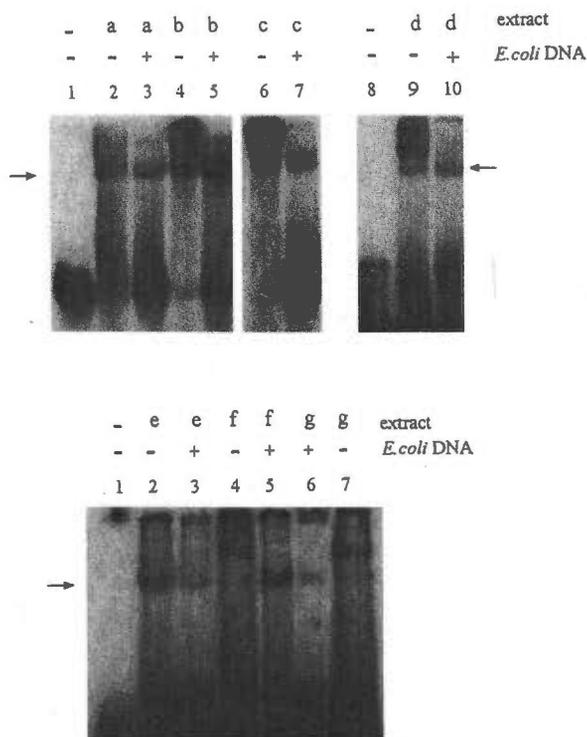


Fig. 1. Electrophoretic mobility shift assays after incubation of HTC4 with various nuclear extracts. HL60 (a), chicken erythrocytes (b), rat liver (c), pig liver (d), HeLa (e), CHO (f), CEM (g). 0.1 pmol labelled HTC4 was incubated with 5 μ g nuclear extracts from cultured cells, 10 μ g liver extracts and 70 μ g chicken erythrocyte extract (a larger amount of this extract was used since the extraction procedure did not completely remove haemoglobin). The denatured *E. coli* competitor DNA was present in a 500-fold excess (mass/mass). In this figure and in Fig. 2 the arrow indicates the specific complex.

presence of a large excess of 'random' sequence single-stranded DNAs [18]. To better delineate the sequence requirements of this protein, a number of EMSAs and ultraviolet-cross-linking experiments were done after incubation of 32 P-labelled oligonucleotides consisting of various numbers of the telomeric motif with some of the protein extracts (HeLa, rat liver and chicken erythrocytes). Fig. 2A and B shows the EMSA results after incubating the oligomers HTC2, HTC3 and HTC4 (Table 1) with HeLa and rat liver extracts, respectively. The absence of any retarded band indicates that oligonucleotides containing only two or three CCC blocks are not able to form a complex with the protein which is sufficiently stable to survive electrophoresis. The same result has been observed in the case of chicken erythrocyte protein extracts. However, a parallel ultraviolet-cross-linking experiment with HeLa extract showed the presence of the complex also for HTC2 and HTC3 (Fig. 3). This suggests that a complex is formed also in the presence of two or three blocks but can be detected only if it is stabilised by ultraviolet cross-linking, as its dissociation kinetics is too fast to allow its survival during EMSA. Incubation of labelled HTC6 with HeLa and rat liver extracts yielded results similar to those obtained with HTC4 (data not shown).

However, it could be argued that the lower stability of the complex formed by HTC2 and HTC3 was not due to the number of repeats, but rather to the shorter length of these oligomers, and thus to the absence of sequence-unspecific, but nevertheless stabilising, interactions. Moreover, it remains to be determined whether the most important specificity determinants are the CCC blocks or the TAA moieties. Fig. 4 shows the EMSA results obtained with the oligonucleotides HTR2 and HTA4 (Table 1).

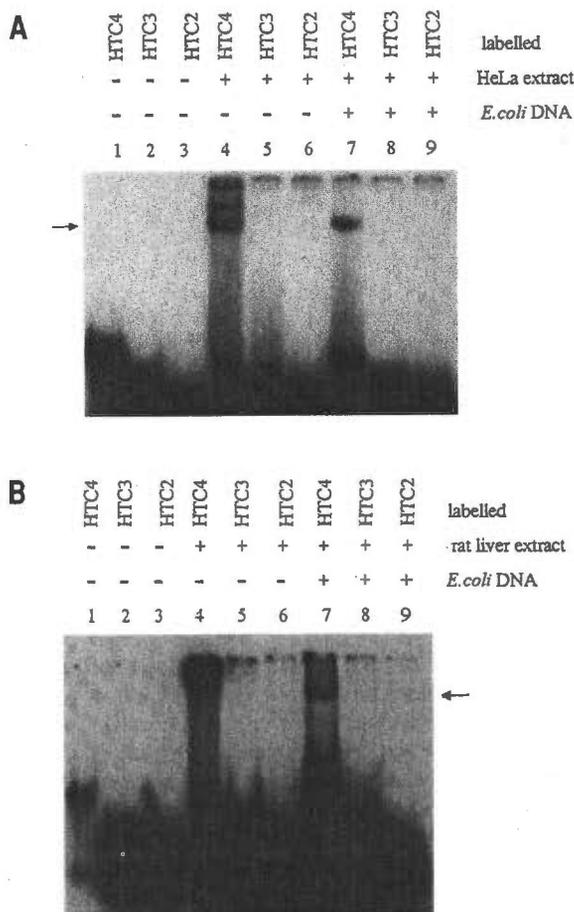


Fig. 2. Electrophoretic mobility shift assays after incubation of labelled HTC4, HTC3 and HTC2 with nuclear extracts. (A) HeLa extract. (B) Rat liver extract. Lanes 1, 4, 7: HTC4; lanes 2, 5, 8: HTC3; lanes 3, 6, 9: HTC2; lanes 1, 2, 3: oligonucleotides alone; lanes 4, 5, 6: oligonucleotides incubated with nuclear extract; lanes 7, 8, 9: oligonucleotides incubated in the presence of a 500-fold excess of denatured *E. coli* DNA.

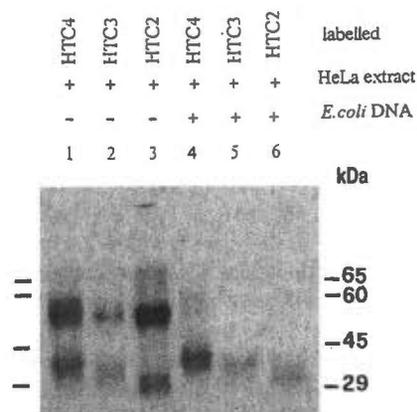


Fig. 3. Autoradiography of an SDS/polyacrylamide gel after electrophoresis and ultraviolet cross-linking. The incubation mixtures relative to lanes 4 to 9 in Fig. 2A are presented.

The former contains the sequence of HTC2 flanked by random sequences so that it is the same length as HTC4. HTA4 consists of four TAA blocks linked by three CCC blocks. In the absence of competitor DNA, both show a long smeared band, but in its presence this band is completely absent. This indicates the formation of complexes whose stabilities are higher than those

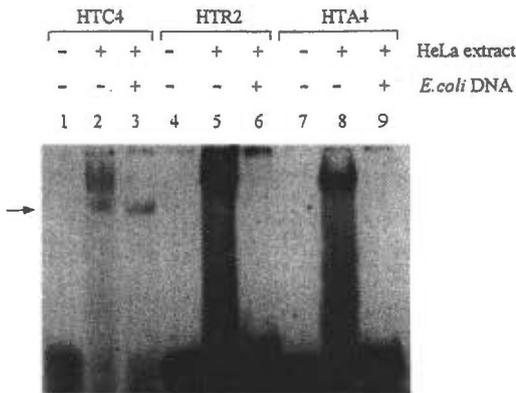


Fig. 4. Electrophoretic mobility shift assays after incubation of labelled HTC4, HTR2 and HTA4 with HeLa extract. Lanes 1–3: HTC4; lanes 4–6: HTR2; lanes 7–9: HTA4; lanes 1, 4, 7: oligonucleotides alone; lanes 2, 5, 8: oligonucleotides incubated with nuclear extract; lanes 3, 6, 9: oligonucleotides incubated in the presence of a 500-fold excess of denatured *E. coli* DNA.

formed by HTC2 and HTC3, (whose bands are seen only after ultraviolet cross-linking), but lower than that formed by HTC4, as they appear to dissociate during electrophoresis in the absence of competitor DNA and disappear in its presence.

Influence of the motif sequence on protein binding. Although the telomeric repeated motif has been highly conserved during the phylogenesis of eukaryotes, minor variations are seen among different phyla [1]. With the aim of exploring the dependence of binding specificity of this nuclear protein on the sequence of the telomeric motif, we have performed a number of EMSAs incubating some nuclear protein extracts (HeLa, HL60, rat liver and chicken erythrocytes) with various oligonucleotides, all containing four blocks of cytosines and with sequences representative of telomeric motifs from several phyla. Those found in *Arabidopsis* (CCCTAAA), *Bombyx* (CCTAA), *Parascaris* (GCCTAA), and *Tetrahymena* (CCCCAA), along with a similar one not found in telomeres (TCCCAA), were chosen for the synthesis of oligonucleotides ATC4, BTC4, TTC4, PTC4 and Ψ TC4, respectively (Table 1). The nuclear proteins in the extracts tested formed specific complexes detectable as retarded bands with the *Arabidopsis*, *Tetrahymena* and TCCCAA motif oligomers, whereas they consistently appeared to be unable to produce any retarded band with the *Bombyx* motif oligomer. A retarded band of lower mobility was found with the *Parascaris* motif oligomer, which was abolished by the standard *E. coli* DNA competitor (data not shown). Figure 5 exemplifies the results in the case of the HeLa extract.

To establish the affinity order of these oligonucleotides for the HeLa protein a number of competitive EMSAs (Fig. 6B), including an auto-competition experiment with HTC4 (Fig. 6A), were performed. In the incubation mixture containing the HeLa extract and a 250-fold excess of single-stranded *E. coli* DNA to suppress unspecific binding, each labelled oligonucleotide was challenged with an equimolar amount of the unlabelled telomeric competitor (Fig. 6B). It can be noted that: (a) the retarded band of labelled HTC4 persists in the presence of each of the other telomeric oligonucleotides (lane 4: ATC4; lane 5: TTC4; lane 6: Ψ TC4), although slightly weakened by TTC4; (b) the retarded band of labelled ATC4 persists in the presence of cold Ψ TC4 (lane 12), but is nearly suppressed in the presence of TTC4 (lane 11); (c) the retarded band of labelled Ψ TC4 is weakened in the presence of TTC4 (lane 16); (d) the retarded band of labelled ATC4 is, as expected, markedly reduced in the presence of unlabelled HTC4 (lane 7). From these data the order

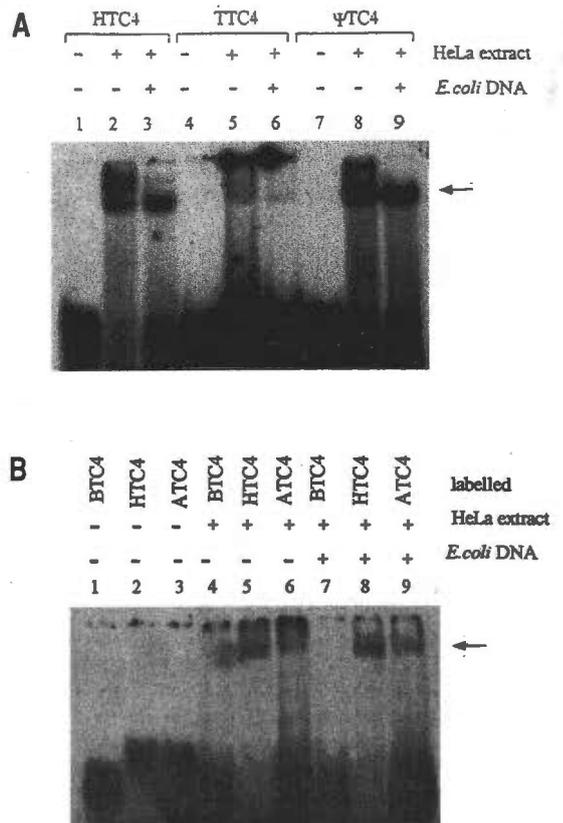


Fig. 5. Electrophoretic mobility shift assays after incubation of labelled oligodeoxynucleotides with HeLa extract. (A) Lanes 1–3: HTC4; lanes 4–6: TTC4; lanes 7–9: Ψ TC4; lanes 1, 4, 7: oligonucleotides alone; lanes 2, 5, 8: oligonucleotides incubated with nuclear extract; lanes 3, 6, 9: oligonucleotides incubated in the presence of a 500-fold excess of denatured *E. coli* DNA. (B) Lanes 1, 4, 7: BTC4; lanes 2, 5, 8: HTC4; lanes 3, 6, 9: ATC4; lanes 1–3: oligonucleotides alone; lanes 4–6: oligonucleotides incubated with nuclear extract; lanes 7–9: oligonucleotides incubated in the presence of a 500-fold excess of denatured *E. coli* DNA.

of decreasing affinity of the HeLa protein for the various oligonucleotides is $\text{HTC4} \approx \text{TTC4} > \text{ATC4} > \Psi\text{TC4}$.

The EMSA performed by incubating the HeLa and rat liver nuclear extracts with a CCC-block sequence of unrelated origin, i.e. the element (CCCT)₃CCTT(CCCT)₃, found in the murine *Ki-ras* promoter, has shown the presence of a retarded band, which is not competed by an excess of DNA, but whose mobility is dramatically lower than that observed with the other oligomers. That the protein components responsible for the retarded bands with HTC4 and *Ki-ras* are not the same, is clear from the absence of reciprocal interference; neither oligonucleotide is able to displace the other from its protein complex (Fig. 7).

Molecular mass of the bound protein in non-denaturing conditions. The molecular mass of the protein from the HeLa nuclear extract was estimated to be about 45 kDa [18] from the electrophoretic mobility following SDS/PAGE of its ultraviolet-cross-linked complex with labelled HTC4, i.e. in denaturing conditions. Similarities in the behaviour of the complex between EMSA and SDS/PAGE experiments suggested that also in non-denaturing conditions the complex was likely to be composed of one polypeptide and one oligonucleotide, i.e. the protein is probably monomeric. To corroborate this hypothesis the molecular mass of the protein was examined under non-denaturing conditions by GPC. The retention volume of the ultraviolet-cross-linked complex consisting of labelled HTC4 and the proteins from HL60 cells and rat liver in a GPC column was determined,

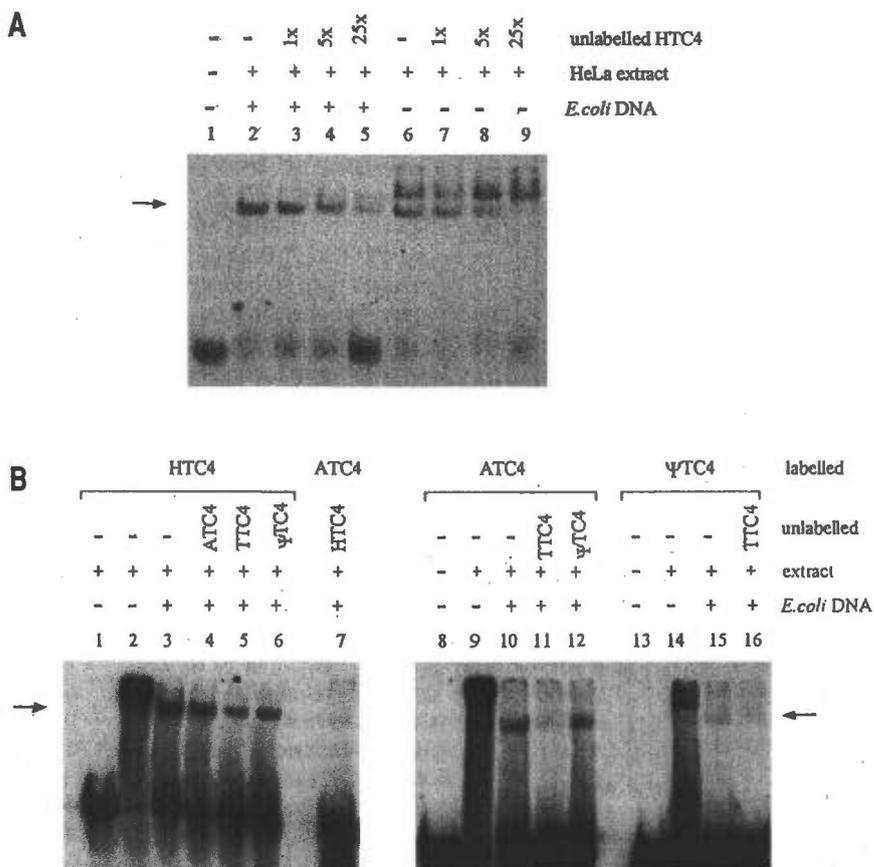


Fig. 6. Competitive electrophoretic mobility shift assays between various telomeric oligonucleotides. (A) Competition between labelled HTC4 and increasing amounts of unlabelled HTC4. Lanes 2–5 and lanes 6–9: in the presence and absence of a 250-fold excess of single-stranded *E. coli* DNA, respectively. (B) Electrophoretic run after incubation of equimolar amounts of one labelled and one non-labelled telomeric-type oligodeoxynucleotide with HeLa extract, in the presence of a 250-fold excess of denatured *E. coli* DNA as a non-specific competitor.

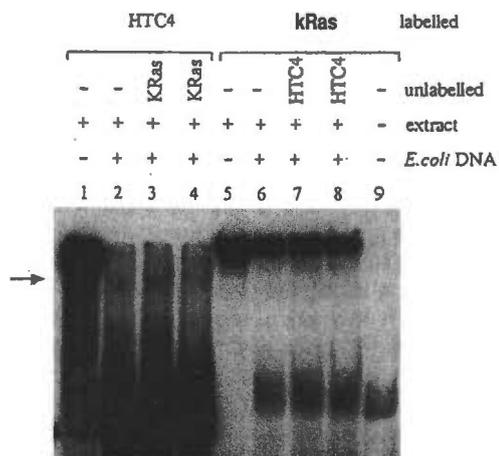


Fig. 7. Electrophoretic mobility shift assay after incubation of labelled HTC4 or labelled kRas with HeLa protein extract. EMSA was performed in the absence and presence of a 250-fold excess of single-stranded *E. coli* DNA, as a non-specific competitor, and in the absence or in the presence (equimolar and tenfold excess) of non-labelled kRas (lanes 3–5) or HTC4 (lanes 7–9).

while those of bovine serum albumin (67 kDa) and egg albumin (45 kDa) peaked at 34.5 ml and 41 ml, respectively. This result points to a molecular mass within the range 45–50 kDa for the HTC4-protein complex, which broadly agrees with the SDS/PAGE estimate, and to nearly 40 kDa for the protein itself, after subtracting the mass of the bound HTC4 (about 8 kDa). Thus, the hypothesis of the monomeric nature of the protein is supported by these data.

DISCUSSION

The sequence requirements of this nuclear protein for its binding activity appear to be markedly stringent. As for the influence of the number of CCC repeats on binding, the data clearly indicate that at least four blocks of three to four cytosines properly spaced are required for the full development of the binding potential. This would suggest a possible role of the non-canonical structure *i*-DNA in the interaction. Indeed, at least four C-blocks are necessary for the intramolecularly folded form of *i*-DNA [12, 13]. However, it was found that two or three CCC blocks are also bound, albeit rather poorly, by the protein, which does not support this hypothesis. Nevertheless, the possibility of an intermolecular form of *i*-DNA cannot be ruled out at the moment. The observation that the oligomer kRas, which is able to assume the intramolecular *i*-DNA form [13], produces a specific retarded band in the EMSA, although with a dramatically lower mobility than the other oligonucleotides tested (probably due to interaction with a different protein), also does not support the *i*-DNA hypothesis. However, work is in progress in our labora-

along with the retention volumes of the reference proteins bovine serum albumin and egg albumin. The elution profiles of the reference proteins were monitored by ultraviolet absorption, and those of the HTC4-protein complexes were monitored by radioactivity (Fig. 8). The profiles of the complexes peaked at 40 ml,

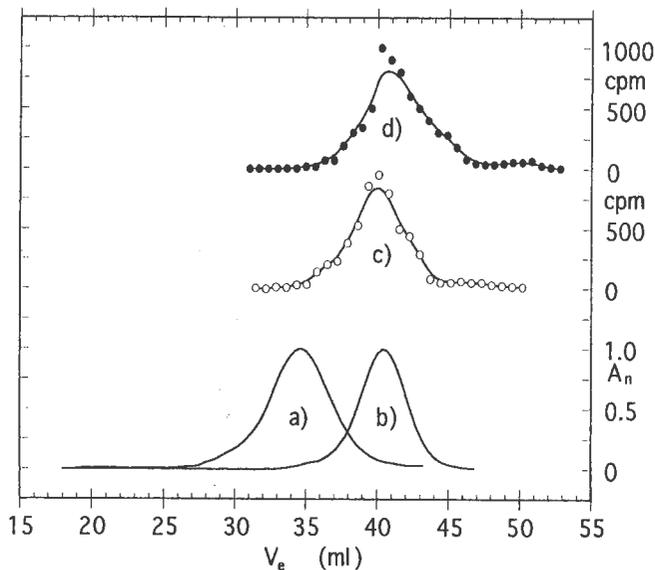


Fig. 8. GPC elution profiles. Elution profile peaks of bovine serum albumin (a), egg albumin (b; ordinate, normalised ultraviolet absorption) and the ultraviolet-cross-linked products of the incubation mixtures of labelled HTC4 with HL60 nuclear protein extract (c) and rat liver nuclear extract (d; ordinate: cpm of eluted fractions) in the presence of a 100-fold excess of single-stranded *E. coli* DNA as competitor.

tory to quantify the propensity of those oligomers to assume the i-DNA conformation, using optical and chiroptical techniques.

With respect to the dependence of complex formation on the telomeric motif, the oligomers from *Tetrahymena* and *Arabidopsis* display a behaviour close to that of the vertebrate oligomer. In contrast, BTC4 does not give rise to any retarded band and PTC4 migrates as a retarded band with a different mobility and lower specificity. Hence, blocks of two cytosines are apparently not sufficient to generate the complex formed by all of the other telomeric oligonucleotides. Interestingly, the oligomer based on the motif TCCCAA also appears to bind with similar strength. However, up to now this motif has not been found in telomeres in nature, maybe because it could interfere with the transcription machinery due to its strong resemblance to the CAAT box of eukaryotic promoters. Ψ TC4 exhibits the lowest affinity among the sequences tested in the EMSAs (Fig. 6). It is worth noting that the highest affinity exhibited by the HeLa protein is that for its species-specific motif.

Having observed the specific DNA binding activity in HeLa nuclear protein extracts, we looked for the presence of an analogous activity in a number of cell lines and tissues from different species. A binding activity with essentially the same characteristics in terms of electrophoretic mobility shift, apparent molecular mass and specificity for single-stranded telomeric sequences was observed in each case. So far our investigation has been limited mainly to mammalian sources, although including very different orders: primates (human), rodents (rat, hamster), and artiodactyla (pig). This protein-binding activity was always observed, although the shifted band signals showed different intensities. Moreover, we also observed it in the only non-mammalian source tested up to now, i.e. chicken erythrocytes, in which case the binding was even stronger. These findings suggest that the protein responsible for this binding activity is likely to be a rather strongly conserved feature in vertebrates, which share the telomeric motif CCCTAA. Work is in progress to extend the survey to species of other vertebrate classes and orders. It will also be of interest to search for differences in the level of expression of this protein in various tissues of the same organism and

to check for a possible correlation with telomerase activity and other aspects of telomere functioning.

Little attention was paid to a more slowly migrating band in an EMSA previously described [18]. Its lower specificity was thought simply be due to fortuitous similarities between the telomeric motif and other sequences in cellular DNA or RNA. However, the fact that it is observed in parallel with the more specific band in all of the sources considered, and, more significantly, that it is lacking in the absence of the higher mobility species (i.e. in the case of BTC4, HTC3 and HTC2), could suggest that this protein may be not completely unrelated to the protein responsible for the binding activity described here and may even be a precursor thereof.

Recently Sarig et al. [15] described a nuclear protein from rat liver, qTBP42, that is able to bind specifically to each single strand, the C-block and the G-block of telomeric DNA, with the latter also in the tetraplex structure. Subsequently it was shown by the same laboratory that this protein is a strict homologue of the CARG-box-binding factor CBF-A [16], a mouse nuclear factor that recognises specifically this motif which is present in many promoters of genes expressed in muscle cells. Since we observed the C-block binding activity also in rat liver extract, we suspected that qTBP42 might be responsible for the binding. A comparison of the molecular masses strengthen this hypothesis: a native size of 53.5 kDa by GPC and subunit size of 42 kDa by Southwestern blotting for qTBP42 [15] is very similar to the 45–50 kDa from GPC (native conditions) and about 50 kDa from SDS/PAGE for the complex of our protein with HTC4 (the oligomer contributes some 8 kDa to the complex). However, we performed EMSAs after incubation of labelled HTG4, i.e. the strand complementary to HTC4, with rat liver extract and observed a retarded band with a slightly higher mobility than that seen in the case of HTC4, analogous to what was observed with HeLa extract [18]. The electrophoretic mobilities of the specific retarded bands observed in the EMSAs with HTC4, ATC4, TTC4 and Ψ TC4 are the same. It seems unlikely that oligomers of equal length would produce bands of noticeably different mobility, unless major structural changes in the complex occur. More evidence against the identity of the protein as qTBP42 is that, similar to Sarig et al. [15], we detected a protein of about 40 kDa upon Southwestern blotting of the extract following SDS/PAGE with labelled HTG4 as probe. However, we did not detect any band using labelled HTC4 as probe. This may be due to the non-recovery of C-block-binding activity after SDS denaturation of the protein.

Finally, since qTBP42 binds with almost the same affinity to oligonucleotides containing the CARG motif and to oligonucleotides constituted by telomeric repeats [16], we performed a number of EMSAs to check whether there is competition between oligonucleotides representative of the two groups for the same nuclear protein component. A tenfold excess of non-labelled CARG oligonucleotide was unable to impair the formation of the protein/HTC4 complex and *vice versa* (data not shown). In addition, the single-stranded *E. coli* DNA competitor completely abolished binding to the CARG oligonucleotide at a concentration that does not affect complex formation with HTC4. Finally, the mobilities of the protein complexes with the two oligonucleotides are slightly but clearly different. These data show that our protein is not qTBP42.

Characterisation of this protein will be possible only after its isolation in a sufficiently pure form and its sequencing, or after identification and sequencing of the gene which encodes it. Work is in progress in our laboratory to achieve this goal.

This work was supported by grants from the Italian MURST and CNR.

REFERENCES

1. König, P. & Rhodes, D. (1997) Recognition of telomeric DNA, *Trends Biochem. Sci.* **22**, 43–47.
2. Greider, C. W. (1996) Telomere length regulation, *Annu. Rev. Biochem.* **65**, 337–365.
3. Lundblad, V. & Wright, W. E. (1996) Telomeres and telomerase: a simple picture becomes complex, *Cell* **87**, 369–375.
4. Shore, D. (1997) Telomerase and telomere-binding proteins: controlling the endgame, *Trends Biochem. Sci.* **22**, 233–235.
5. Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P. & de Lange, T. (1995) A human telomeric protein, *Science* **270**, 1663–1667.
6. Fang, G. W. & Cech, T. R. (1993) *Oxytricha* telomeric-binding protein: DNA-dependent dimerization of the alpha and beta subunits, *Proc. Natl Acad. Sci. USA* **90**, 6056–6060.
7. Henderson, E. R. & Blackburn, E. H. (1989) An overhanging 3' terminus is a conserved feature of telomeres, *Mol. Cell. Biol.* **9**, 345–348.
8. Makarov, V. L., Hirose, Y. & Langmore, J. P. (1997) Long G tails at both ends of human chromosomes suggest a strand degradation mechanism of telomere shortening, *Cell* **88**, 657–666.
9. Williamson, J. R. (1994) G-quartet structures in telomeric DNA, *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703–730.
10. Rhodes, D. & Giraldo, R. (1995) Telomere structure and function, *Curr. Opin. Struct. Biol.* **5**, 311–322.
11. Gehring, K., Leroy, J. L. & Gueron, M. (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs, *Nature* **363**, 561–565.
12. Leroy, J. L., Mergny, J. L., Gueron, M. & Helene, C. (1994) Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif, *Nucleic Acids Res.* **22**, 1600–1606.
13. Manzini, G., Yathindra, N. & Xodo, L. E. (1994) Evidence for intramolecularly folded i-DNA structures in biologically relevant CCC-repeat sequences, *Nucleic Acids Res.* **22**, 4634–4640.
14. Eid, J. E. & Sollner-Webb, B. (1995) ST-1, a 39-kilodalton protein in *Trypanosoma brucei*, exhibits a dual affinity for the duplex form of the 29-base pair subtelomeric repeat and its C-rich strand, *Mol. Cell. Biol.* **15**, 389–397.
15. Sarig, G., Weisman-Shomer, P., Erlitzki, R. & Fry, M. (1997) Purification and characterization of qTBP42, a new single-stranded and quadruplex telomeric DNA-binding protein from rat hepatocytes, *J. Biol. Chem.* **272**, 4474–4482.
16. Sarig, G., Weisman-Shomer, P., & Fry, M. (1997) Telomeric and tetraplex DNA binding properties of qTBP42: a homologue of the CARG box binding protein CBF-A, *Biochem. Biophys. Res. Commun.* **237**, 617–623.
17. Kamada, S. & Miwa, T. (1992) Protein binding to CARG box motifs and to single stranded DNA functions as a transcriptional repressor, *Gene* **119**, 229–236.
18. Marsich, E., Piccini, A., Xodo, L. E. & Manzini, G. (1996) Evidence of a HeLa nuclear protein that binds specifically to the single-stranded d(CCCCTAA)_n telomeric motif, *Nucleic Acids Res.* **24**, 4029–4033.
19. Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei, *Nucleic Acids Res.* **11**, 1475–1489.
20. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

A chicken hnRNP of the A/B family recognizes the single-stranded d(CCCTAA)_n telomeric repeated motif

Eleonora Marsich¹, Antonella Bandiera¹, Gianluca Tell¹, Andrea Scaloni² and Giorgio Manzini¹

¹Department of Biochemistry, Biophysics, and Macromolecular Chemistry, University of Trieste, Italy;

²I.A.B.A.M., Centro Internazionale Servizi di Spettrometria di Massa, CNR, Napoli, Italy

With the aim of identifying proteins able to interact with the C-rich single-stranded telomeric repeated motif, three nuclear polypeptides, CBNP α , CBNP β and CBNP γ , with apparent mobilities in SDS/PAGE of 38, 44 and 55 kDa, respectively, were isolated from mature chicken erythrocytes by affinity chromatography. *In situ* UV-cross-linking experiments demonstrated that CBNP α and CBNP γ interact directly with the telomeric d(CCCTAA)_n repeat, whereas CBNP β does not. Moreover, they provided information on the protein components responsible for each electrophoretic mobility-shift assay signal. Ion spray and matrix-assisted laser desorption ionization MS allowed us to identify CBNP α with single-stranded D-box-binding factor (ssDBF), a protein previously characterized as a transcription factor belonging to the A/B family of heterogeneous nuclear ribonucleoproteins, and CBNP β with an isoform of this protein containing an extra exon. Similarly, CBNP γ was shown to be probably the chicken homolog of hnRNP K, a ribonuclear protein able to bind to polyC oligonucleotides. The relation of CBNP α (i.e. ssDBF), CBNP β and CBNP γ to a number of similar proteins in the protein and nucleotide sequence databank is discussed. A rather diversified spectrum of functional roles has been assigned to some of these proteins despite the strong sequence homology among them.

Keywords: heterogeneous ribonucleoproteins (hnRNPs); nuclear proteins; ssDNA recognition; telomeric C-rich motif.

Since the discovery of the special features of repetitive DNA which constitutes the majority of telomeres of eukaryotic organisms, and of its specific replication machinery, considerable attention has also been directed to identifying and characterizing nuclear proteins that specifically bind to this DNA, in the normal Watson–Crick duplex form, as well as to the protruding single-stranded G-rich 3' overhang at the telomere terminus. Many proteins such as TRF1 and TRF2 that bind to duplex telomeric DNA in mammalian cells [1] and Rap1p in yeast [2] have been reported to be involved in telomere length maintenance and regulation of telomerase activity. In unicellular organisms, several proteins, such as $\alpha\beta$ protein from *Oxytricha* [3], TBP from *Euplotes* [4], TEP and TGP from *Tetrahymena* [5,6], GBP from *Chlamydomonas* [7], which are known to interact with the single-stranded 3'-ending motif of the telomeres, have been characterized. Another group of proteins, ST-2 from *Trypanosoma* [8], qTBP42 from rat [9], human replication factor C [10], and murine STBP [11] and

A1/UP1 [12], also bind to the single-stranded G-rich telomeric motif, although their function has not been fully ascertained. The last of these, however, is the first ssDNA-binding protein shown to be directly involved in mammalian telomere biogenesis, suggesting a possible mechanism by which telomere length can be modulated [13]. Much less attention has been given to identifying nuclear components able to recognize the complementary single-stranded C-rich DNA repeat. Interestingly, a protein from *Trypanosoma*, ST-1, binds to the telomeric double-stranded repeat as well as to its single-stranded C-rich component [14]. In vertebrates, the nuclear protein from rat hepatocytes, qTBP42, has been shown to recognize each of the single-stranded forms of the telomeric repeat [9]. Several proteins that interact with polypyrimidine ssDNA have been described. They include NOGA4 in mouse [15] and rat [16] and the human hnRNP K protein, which binds to the single-stranded CT element of the *c-myc* gene promoter [17]. While this article was being completed, it came to our attention that hnRNP K and the splicing factor ASF/SF2 from HeLa nuclear extracts are able to recognize the single-stranded C-rich telomeric repeat [18]. These authors pointed out that other so far unidentified nuclear proteins could share this property. In this context it may be relevant to mention that earlier reports from this laboratory showed that proteins present in nuclear extracts from several vertebrate sources bind to the single-stranded telomeric repeat motif (CCCTAA)_n, one of them exhibiting high sequence specificity [19,20]. This protein component does not recognize the complementary d(TTAGGG)_n nor the telomeric duplex. We report here on the isolation and molecular identification of these nuclear proteins from extracts of mature chicken erythrocytes.

Correspondence to G. Manzini, Department of Biochemistry, Biophysics and Macromolecular Chemistry, University of Trieste, Via L. Giorgieri 1, 34127 Trieste, Italy. Fax: + 39 040 6763691, Tel.: + 39 040 6763677, E-mail: manzini@bbcm.univ.trieste.it
Abbreviations: CBNP, C-block-binding nuclear protein; hnRNP, heterogeneous nuclear ribonucleoprotein; EMSA, electrophoretic mobility shift assay; ISMS, ion-spray mass spectrometry; MALDIMS, matrix assisted laser desorption induced mass spectrometry; MSMS, tandem mass spectrometry; ssDBF, single-stranded D-box-binding factor. *Enzymes:* endoproteinase Asp-N (EC 3.4.24.33); endoproteinase Glu-C (EC 3.4.21.19).

(Received 14 July 2000; revised 26 October 2000; accepted 30 October 2000)

performed in batch: extracts (10–20 mg·mL⁻¹) diluted with 50 mM Tris/HCl, pH 8, containing 5 mM EDTA, 20% (v/v) glycerol, 0.5 mM dithiothreitol and 0.5 mM phenylmethanesulfonyl fluoride, to a final concentration of 200 mM NaCl. Generally, 40 mL diluted extract was incubated with 400 µL affinity resin overnight at 4 °C on a rotor wheel. After incubation, the resin was washed several times with 50 mM Tris/HCl, pH 8, containing 5 mM EDTA and 200 mM NaCl, and the bound proteins were eluted with 10 mM Hepes, pH 7.9, containing 0.1 mM EGTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethanesulfonyl fluoride and 1 M NaCl for 1 h at room temperature.

Reverse-phase chromatography

Proteins were fractionated by RP-HPLC on a Jupiter C₁₈ column (150 × 4.6 mm; 5 µm; 300 Å pore size; Phenomenex, USA) by using a linear gradient from 0% to 100% of acetonitrile in 0.1% trifluoroacetic acid over 40 min, at a flow rate of 0.5 mL·min⁻¹. Individual components were collected manually and lyophilized.

In-gel digestion and peptide purification

Proteins from SDS/PAGE were excised from the gel, triturated and washed with acetonitrile and 0.1 M NH₄HCO₃ (twice each one). The proteins were in-gel reduced with 10 mM dithiothreitol in 0.1 M NH₄HCO₃ (45 min, at 55 °C) and S-alkylated with 55 mM iodoacetamide in 0.1 M NH₄HCO₃ (30 min at 25 °C and in the dark). After extensive washing with 0.1 M NH₄HCO₃, gel particles were dried and shrunk with a digestion solution of 12.5 ng·µL⁻¹ trypsin in 0.1 M NH₄HCO₃. After 2 h of incubation at 5 °C, an identical aliquot of digestion solution was added to the samples which were incubated overnight at 37 °C. Digestion solution was completely removed, and the gel pieces were sonicated with 50% acetonitrile in 0.1 M NH₄HCO₃ (twice). The recovered solutions were mixed together and lyophilized. Peptide mixtures were fractionated by RP-HPLC on a Vydac C₁₈ column 214TP52 (250 × 2.1 mm; 5 µm; 300 Å pore size; The Separation Group, USA) by using a linear gradient from 5% to 60% of acetonitrile in 0.1% trifluoroacetic acid over 65 min, at a flow rate of 0.2 mL·min⁻¹. Individual components were collected manually and dried in a Speed-vac centrifuge (Savant). Similar experiments were performed by using endoproteases Asp-N and Glu-C. Digestions with Asp-N were carried out at 25 °C in 50 mM Tris/HCl, pH 7.5, for 16–20 h. Two independent digestions were set up with endoprotease Glu-C in 25 mM ammonium acetate, pH 4, and in 25 mM NH₄HCO₃, pH 7.9, carrying out the reactions at 25 °C for 16–20 h.

Mass spectrometry

Intact proteins were subjected to ion-spray MS (ISMS) analysis, using an API/SCIEX 100 ion-spray mass spectrometer. Samples were dissolved in acetonitrile/water (1 : 1) plus 0.1% (v/v) formic acid and then injected at a flow rate of 2 µL·min⁻¹. The quadrupole was scanned in the range *m/z* 850–1200 and the spectra were acquired and elaborated using a manufacturer's program. Mass scale calibration was carried out using the multiple-charged ions of a separate introduction of myoglobin. All data are shown as average masses.

Matrix-assisted laser desorption ionization (MALDI) mass spectra were recorded by using a Voyager DE MALDI-TOF spectrometer (Perkin-Elmer-PerSeptive Biosystem, Norwalk,

CT, USA); a mixture of analyte solution, α-cyano-4-hydroxycinnamic acid and bovine insulin were applied to the sample plate and then dried. Mass calibration was performed using the molecular ions from bovine insulin (5734.59 *m/z*) and the α-cyano-4-hydroxycinnamic acid (379.06 *m/z*) as internal standards. Raw data were analysed by using a software program provided by the manufacturer and are reported as average masses. MALDI MSMS spectra were recorded using an Autospec OA-TOF instrument (Micromass, Manchester, UK). In this case 2,5-dihydroxybenzoic acid was used as a matrix. The molecular ions collided with argon in a collision cell floated at 800 eV. The masses of fragment ions were measured using an orthogonally mounted TOF analyser. Data analysis was carried out by using the OPUS software supplied with the instrument.

Protein identification

Two software packages, PROWL and PEPTIDESEARCH, were used to identify protein bands from independent non-redundant protein sequence databases which are maintained and updated daily at the European Molecular Biology Laboratory (EMBL) and the National Institute of Health (NIH). They were both used by selection of protein molecular-mass filter, adjustable incomplete cleavage, taxonomic choice of the organism under investigation, peptide mass error, mass changes due to protein modifications (e.g. cysteine alkylation), and isoforms or post-translational modifications. A number of top candidates with high scores from the peptide-matching analysis were further evaluated by comparison with their calculated molecular mass using the experimental values obtained from SDS/PAGE. This parameter was used as a filter with large tolerance (change in molecular mass = ± 15% of the molecular mass) in order to exclude false-positive candidates from the output lists.

Western-blot analysis

Bands from SDS/PAGE analysis of affinity-purified material were excised from the gel and subjected to further SDS/PAGE in a 10% gel. Proteins were then transferred to nitrocellulose membranes (Schleicher & Schuell, Keene, NH, USA). These were saturated by incubation at 4 °C overnight with 10% non-fat dry milk in NaCl/P_i/0.1% Tween 20 and then incubated with the rabbit polyclonal anti-[single-stranded D-box-binding factor (ssDBF)] serum for 60 min at room temperature. The rabbit polyclonal anti-ssDBF serum was used at a dilution of 1 : 1000 in NaCl/P_i/0.1% Tween 20. After three washes with NaCl/P_i/0.1% Tween 20, they were incubated with an anti-rabbit immunoglobulin coupled to peroxidase (Sigma Chemical Co, St Louis, MO, USA). After 60 min of incubation at room temperature, the membranes were washed several times with NaCl/P_i/0.1% Tween 20 and the blot was developed using the ECL chemiluminescence method (Amersham Pharmacia Biotech).

RESULTS

We showed previously in EMSA experiments the presence of proteins that bind to ssDNA with the telomeric repeated motif d(CCCTAA)_n, but not its cDNA or the duplex, in nuclear extracts from several sources [19,20]. One of these displayed a remarkably high specificity of binding to this sequence. We have now succeeded in isolating these proteins from mature chicken erythrocytes using affinity chromatography.

Q1

Q2

Q3

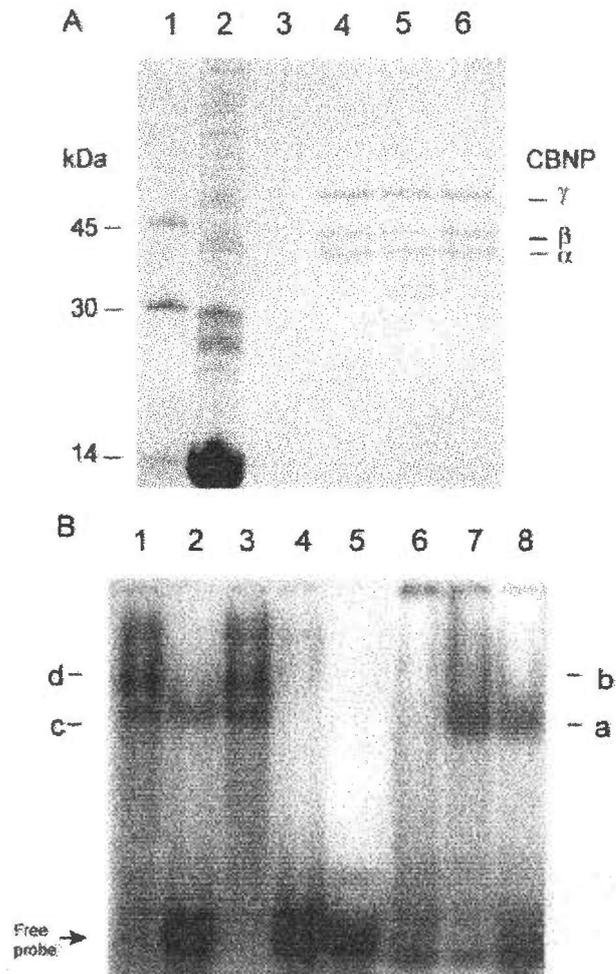


Fig. 1. (A) SDS/PAGE analysis of affinity chromatography recovered components stained with Coomassie Blue and (B) EMSA profiles to track the binding activity toward the labeled probe HTC4 during the isolation procedure. (A) lane 1, molecular-mass markers; lane 2, total extract (10 μ g) from chicken erythrocytes; lane 3, sample recovered from non-functionalized resin; lanes 4–6, three different affinity recovered samples. The three stained components are indicated in the right margin as CBNP α , β , and γ . (B) lanes 1 and 2, whole extract from chicken erythrocytes in the absence and presence, respectively, of 250-fold excess (w/w) denatured *E. coli* competitor DNA before incubation with affinity resin. Lanes 3 and 4, whole extract after incubation with non-functionalized and HTC6-functionalized resin, respectively. Lanes 5–8, activity of protein fraction recovered from HTC6-functionalized resin in the absence (lane 5) and presence of 1 mM dithiothreitol [lane 6 in the absence of competitor, lane 7, in the presence of 250-fold excess (w/w) denatured *E. coli* competitor DNA]. Lane 9, labeled HTC4 alone.

Three sample cases of SDS/PAGE profiles of the proteins recovered by affinity chromatography are shown in Fig. 1A (lanes 4, 5 and 6), along with the profile of the starting material, i.e. total protein extract (lane 2). Lane 3 shows that no stainable protein is retained by the non-functionalized resin. Essentially three components, with apparent molecular masses of 38, 44 and 55 kDa, were isolated. They have been designated C-block-binding nuclear proteins, CBNP α , CBNP β and CBNP γ .

The EMSA profiles of lanes 1–4 of Fig. 1B track the binding activity with the probe HTC4 during the isolation procedure. Lanes 1 and 2 display the activity of the starting material, in the

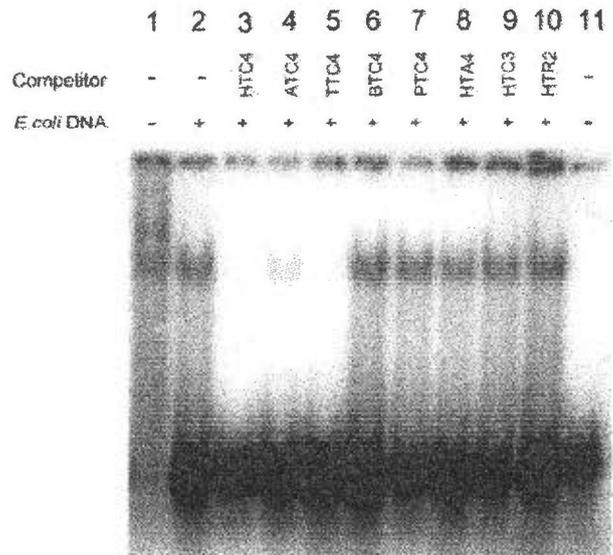


Fig. 2. Competitive EMSA with different telomeric type oligonucleotides. Labeled HTC4 (0.2 pmol) was incubated with 0.1 μ g affinity recovered proteins in the absence of competitors (lane 1) or in the presence of 250-fold excess (w/w) denatured *E. coli* DNA and 10-fold excess of cold telomeric type oligonucleotide competitors (lanes 2–10); lane 11 labeled HTC4 alone.

absence and presence of excess denatured *E. coli* DNA, respectively. After incubation of the extract with the non-functionalized resin, the supernatant retains the binding activity (lane 3), but loses it after incubation with the HTC6-functionalized resin (lane 4). The protein fractions recovered from the functionalized resin, the SDS/PAGE profile of which is shown in Fig. 1A, display the same activity as the whole extract, in both the absence and presence of excess denatured *E. coli* DNA (lanes 7 and 8, respectively). It is important to point out that the presence of dithiothreitol in the incubation buffer was critical for full recovery of DNA-binding activity, as in its absence, no EMSA signal (lane 6) or a smeared signal with low mobility was observed (data not shown).

To test the identity of the DNA-binding activity of the fraction recovered from affinity chromatography with that of the whole extract, a competition EMSA was set up. The formation of the specific protein–HTC4 complex (in the presence of denatured *E. coli* DNA) was challenged with a 10-fold excess of telomeric-related oligonucleotides (Table 1). The result (Fig. 2) matches perfectly with that already seen for whole nuclear extract from HeLa cells [20], despite the different sources. In particular, labeled HTC4 was displaced by unlabeled HTC4 (lane 3) and TTC4 (lane 5) with almost the same efficacy, and slightly less by ATC4 (lane 4), the other sequences being almost completely ineffective (lanes 6–10). *In situ* UV-cross-linking experiments were carried out to establish which of the protein components isolated by affinity chromatography is responsible for each of the EMSA bands, in particular the protein corresponding to the faster and more sequence-specific band. The faster EMSA bands (a and c in lanes 8 and 2 of Fig. 1B) were excised, UV-irradiated, and subjected to SDS/PAGE. In both cases, the protein–probe adduct shows only one signal, the mobility of which is slightly lower than that of CBNP α (lanes 1 and 3 in Fig. 3A) because it is UV-cross-linked with HTC4 which has a molecular mass of \approx 7 kDa. A similar procedure was used to characterize the

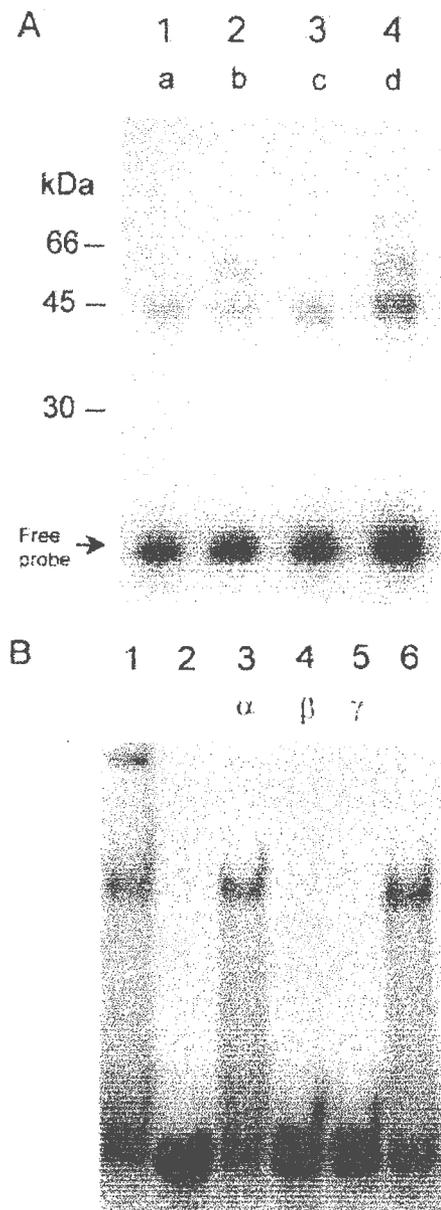


Fig. 3. (A) SDS/PAGE analysis of *in situ* UV-cross-linking of shifted bands from EMSA gel of Fig. 1B, and (B) EMSA of protein components CBNP α , CBNP β , and CBNP γ after electroelution of the excised bands from SDS/polyacrylamide gel. (A) Lane 1, UV-cross-linked product of band a in lane 8 of Fig. 1B; lane 2, UV-cross-linked product of band b in lane 7 of Fig. 1B; lane 3, UV-cross-linked product of band c in lane 2 of Fig. 1B; lane 4, UV-cross-linked product of band d in lane 1 of Fig. 1B. (B) Lane 1, whole affinity recovered sample; lane 2, labeled HTC4 probe alone; lane 3, CBNP α ; lane 4, CBNP β ; lane 5, CBNP γ ; lane 6, the three components electroeluted together.

complexes of slower EMSA bands (b and d of lanes 7 and 1, respectively, of Fig. 1B), in this case two SDS/PAGE signals being obtained, the faster corresponding to the previous one and the other with a mobility slightly lower than that of CBNP γ (lanes 2 and 4 of Fig. 3A). This result suggests that the UV-cross-linkable proteins are CBNP α only in the fast EMSA band, and both CBNP α and CBNP γ in the slow one. To confirm the identity of the proteins responsible for the specific complex

(fast band), the three components CBNP α , CBNP β , and CBNP γ were excised separately from the SDS/polyacrylamide gel, electroeluted, freeze-dried, and resuspended in EMSA binding buffer. EMSAs were performed for each of them in the presence of denatured *E. coli* DNA and 1 mM dithiothreitol. It can be seen from Fig. 3B that CBNP α recovered specific DNA-binding activity fully after electroelution (lane 3), consistently with that of the whole fraction obtained from affinity chromatography (lane 1), and after electroelution in the pool (lane 6). On the other hand, CBNP β and CBNP γ , after separation, did not display any specific EMSA signal (lanes 4 and 5), even in the absence of excess aspecific competitor (data not shown).

Attempts to determine the nature of CBNP α , CBNP β , and CBNP γ by direct amino-acid sequencing after electroblotting on poly(vinylidene difluoride) membranes failed because of the presence of an N-terminal blocking group. However, one of the digestion fragments of CBNP α yielded a dodecapeptide MFVGGGLSWDTSK. This result was not very instructive, because this sequence corresponds to a conserved RNA-binding motif (RNP1), found in many hnRNPs. Therefore, all the protein species were digested with trypsin *in situ* as described in Materials and methods. The peptide mixtures were extracted and resolved by narrow-bore RP-HPLC for further analysis by MALDI MS. Besides exhibiting similar chromatographic profiles, CBNP α and CBNP β yielded many peptides sharing the same mass values (Table 2), suggesting a strong structural relationship between them. These findings were confirmed by independent experiments with endoproteases Asp-N and Glu-C. In fact, CBNP α and CBNP β showed very similar peptide electrophoretic profiles (not shown). In contrast, CBNP γ gave unrelated peptides in all cases, when digested with trypsin (Table 2) as well as with endoproteases Asp-N and Glu-C (not shown). The molecular-mass values obtained from each of these digestions were used to search protein and nucleotide databanks to identify the nature of each molecular species. Excellent results were obtained in the cases of CBNP α and CBNP β , for which 22 out of 24 and 18 out of 20 peptides matched closely those expected for avian ssDBF. This is a 302-residue-long protein and has already been characterized as a liver nuclear factor involved in the transcription of the apoVLDL II gene [21]. The same paper reports the isolation of a further cDNA clone of ssDBF, containing an extra exon, and pointing to the existence of an isoform of this protein. The next best match was obtained for the avian protein CRP1 [22], for which 17 out of 24 and 15 out of 20 peptide matches were observed. Other avian proteins showed fewer than six matches and were not considered. The sequences of ssDBF and CRP1 clearly show that they can be ascribed to the A/B family of hnRNPs. They are identical, except for residues 136 and 221 in the C-terminal region of about 80 residues is apparently unrelated.

To clarify further the identity of CBNP α and CBNP β , and in particular to discriminate between ssDBF and CRP1, several peptides were subjected to MSMS experiments. In addition to the peptides for which MH⁺ signals were observed at *m/z* 1328.5, 1571.9 and 1774.1, corresponding to the expected amino-acid sequences of MFVGGGLSWDTSK, EVYQQQFSSGGGR and IFVGGGLNPEATEEKIR, common to both proteins, those detected for MH⁺ at *m/z* 1656.0 and 1785.3 yielded fragments with the sequences GFGFILFKEP-GSVEK and GFVFTTFKEEDPVKK, respectively (Fig. 4). The presence of Ile136 and Thr221 shows that CBNP α and CBNP β are indeed the two isoforms of ssDBF and not of CRP1. In addition, CBNP α and CBNP β showed two clear signals at *m/z*

Table 2. MS analysis of incognite proteins. CBNP α and CBNP β were assigned to two avian ssDBF forms by PROWL search and MSMS experiments. NA, not assigned.

CBNP α			CBNP β			CBNP γ
Mass	ssDBF	CRP1	Mass	ssDBF	CRP1	Mass
7363.5	Ac2-84	NA	7363.8	Ac(2-84)	NA	1098.7
8108.0	Ac2-91	NA	8107.8	Ac(2-91)	NA	1195.3
1328.5	92-103	75-86	1328.4	92-103	75-86	1505.3
1344.3	92-103ox	75-86ox	1344.5	92-103ox	75-86ox	1519.7
1456.7	92-104	75-87	1456.8	92-104	75-87	1554.4
1472.9	92-104ox	75-87ox	1472.7	92-104ox	75-87ox	1596.1
1159.0	104-112	87-95	1159.2	104-112	87-95	1711.7
1169.9	113-122CAM	96-105CAM	1169.6	113-122CAM	96-105CAM	1818.9
2184.8	113-131CAM	96-114CAM	1656.3	132-146	NA	1834.8
2200.7	113-131oxCAM	96-114oxCAM	1342.7	140-151	123-134	1966.7
3095.2	113-139CAM	NA	2072.5	171-189	154-172	2593.8
3110.9	113-139oxCAM	NA	1505.1	176-189	159-172	2614.7
1656.0	132-146	NA	2485.2	190-210	173-193	2944.8
2110.6	152-169	135-152	2501.1	190-210ox	173-193ox	3084.7
1505.1	176-189	159-172	2215.7	192-210	175-193	3101.1
1774.1	176-191	159-174	2231.4	192-210ox	175-193ox	3832.5
2485.2	190-210	173-193	1941.6	215-230	NA	3895.3
2501.1	190-210ox	173-193ox	1785.3	216-230	NA	3910.9
2215.7	192-210	175-193	1535.2	235-247CAM	218-230CAM	4642.0
2231.4	192-210ox	175-193ox	1571.9	253-266	236-249	
1941.6	215-230	NA				
1785.3	216-230	NA				
1571.9	253-266	236-249				
1334.8	292-302	275-285				

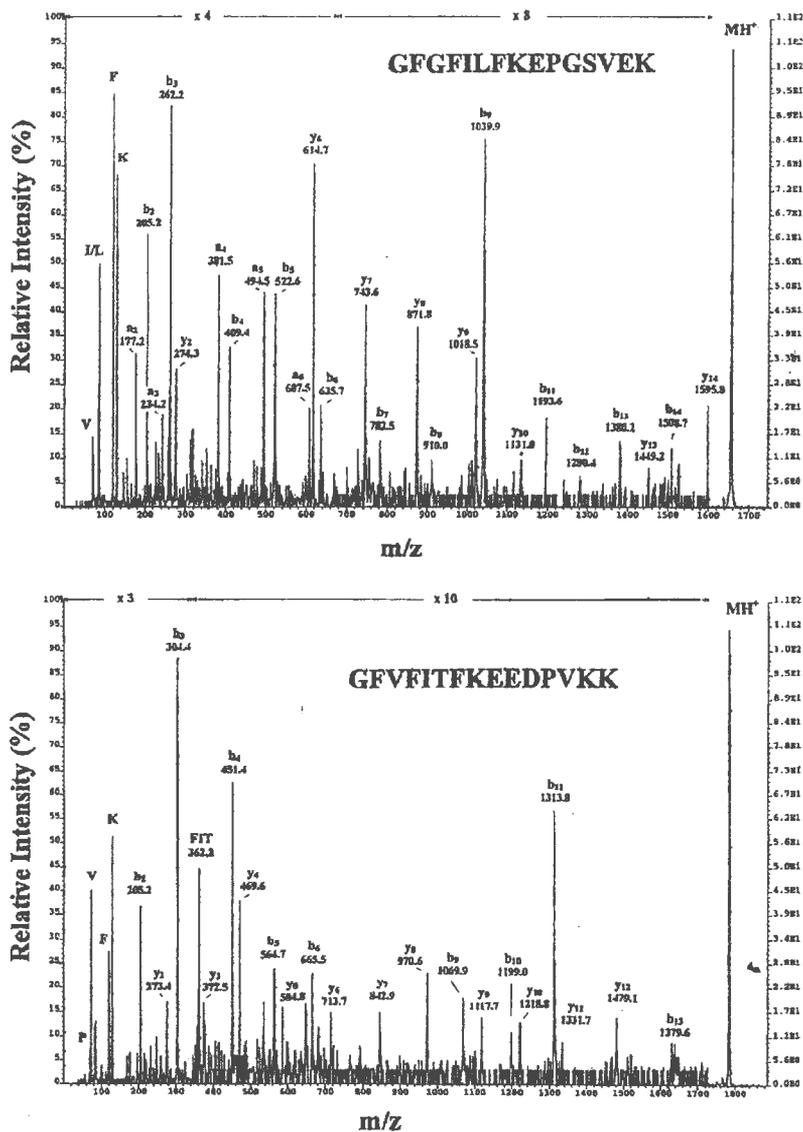
8108.0 and 7363.5 (Table 2) which were tentatively assigned to N-terminal peptides of ssDBF, with the first encoded residue removed and the second one acetylated as the result of co/post-translational processing as previously observed in other hnRNPs.

Further evidence supporting the structural relationship between CBNP α and CBNP β has come from ISMS. The protein mixture recovered from the functionalized resin was processed by RP-HPLC to obtain each component in sufficient amounts for MS analysis. SDS/PAGE revealed that the acetonitrile gradient separated CBNP γ from the other two components, but not CBNP β from CBNP α . This is probably due to their similar hydrophobic profiles. This procedure led to the irreversible loss of the DNA-binding activity. When CBNP γ was analysed by ISMS, it did not give any more signals than those corresponding to fragments of about 15 kDa. The spectrum for the CBNP α /CBNP β sample was clearly consistent with the presence of two species with molecular masses of 31915 ± 2 Da and 37718 ± 4 Da, respectively, the first being more abundant. The absolute values of their molecular masses were consistently lower than those apparent from SDS/PAGE analysis, but their mass differences, as well as their relative abundance, were in accord with the electrophoretic data. This provided further evidence that these two proteins share several physicochemical properties. Besides the coincidences detected by MALDIMS and MSMS experiments in nearly all peptide segments, the mass difference (5803 ± 4 Da) between CBNP β and CBNP α , determined by ISMS, matches perfectly the value calculated for the extra exon observed in the variant cDNA isolated from the chicken library [21]. Thus, these two proteins can be identified as the two isoforms of ssDBF, differing from each other by the occurrence of an extra exon near the

C-terminus. Both the ISMS molecular masses of CBNP α and CBNP β were about 143 Da higher than the values calculated from their sequences. However, it should be noted that they both contain at least two methionine residues in the extensively oxidized form (as determined by MALDIMS mapping) as well as putative methylation sites. Future *in vivo* and *in vitro* studies will elucidate post-translational processing of CBNP α and CBNP β and its relationship with their biological activity.

Parallel attempts to search protein and nucleotide databanks using the mass values determined by MALDIMS for the peptides generated from the *in situ* digestion of CBNP γ (Table 2) were unsuccessful. This suggests that no protein sequence with high similarity to CBNP γ is present in the current databanks. To ascertain the nature of this protein, the purified peptides showing MH⁺ at *m/z* 1098.7 and 1195.3 were subjected to MSMS experiments as previously described. As reported for CBNP α , MSMS analysis of each species yielded fragment ions that allowed us to reconstruct their sequence as GSDFDCELR and NLPLPPPPPPR, respectively. Databank searching revealed that both peptides were identical to two regions of hnRNP K, and were conserved in various species. The corresponding sequence from chicken is still not known. This protein, originally identified as a component of the hnRNP particles, binds to polyC as a consequence of its degree of phosphorylation [23].

Finally, to substantiate further the identity of CBNP α with ssDBF, immunological analysis was performed. With the polyclonal anti-ssDBF serum kindly provided by Dr Smidt [21], its ability to recognize the single bands excised from the SDS/polyacrylamide gel of Fig. 1A and the rerun in SDS/PAGE was tested. As can be seen in Fig. 5A, this serum recognized both the bands of CBNP α and CBNP β , supporting their identity



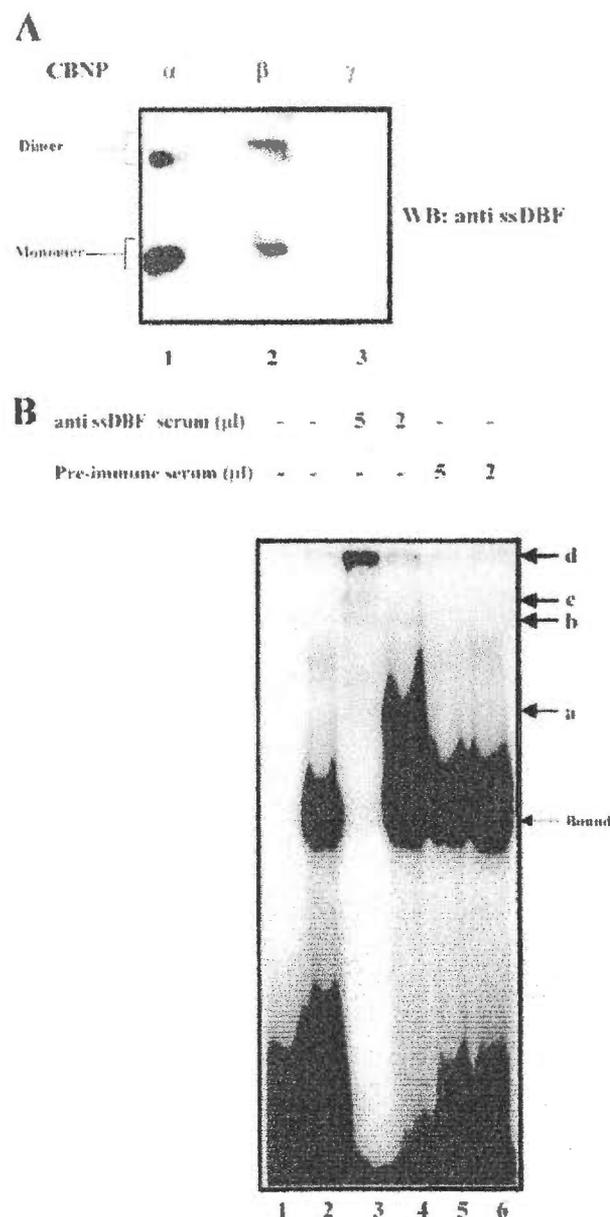


Fig. 5. (A) Western-blot analysis of the single bands excised from SDS/polyacrylamide gel of Fig. 1A and rerun on a 10% gel, and (B) supershift analysis of the specific complex between CBNP α and 32 P-labeled HTC4. (A) Lane 1, CBNP α ; lane 2, CBNP β ; lane 3, CBNP γ . The bands were assayed by using the specific rabbit polyclonal anti-ssDBF serum provided by Dr Smidt [21] and developed by using ECL (Amersham Pharmacia Biotech). (B) Samples containing CBNP α and the HTC4 probe were incubated in the absence (lane 2), in the presence of the specific polyclonal rabbit anti-ssDBF serum (lanes 3 and 4) and, as controls, in the presence of the corresponding amounts of preimmune rabbit serum (lanes 5 and 6) for 2 h at room temperature. Then, the samples were analysed by EMSA. The arrows marked a, b, c show the supershifted bands and the d band corresponds to insoluble material precipitated into the well, which, however, is mostly present in the case of the specific serum anti-ssDBF. Lane 1 contained labeled HTC4 alone

the presence of small amounts of species with twice the molecular mass in non-reducing SDS/PAGE (Fig. 5A). The observation that CBNP β does not UV-cross-link with HTC4, although it is bound by the affinity matrix, suggests that it

forms a complex with CBNP α , but does not interact directly with the telomeric repeats. As the difference between CBNP β and CBNP α is the presence of the extra exon, it can be inferred that this abolishes the DNA-binding activity but not the interaction with other hnRNPs. As far as CBNP γ is concerned, the MS data point to its identity with the chicken homolog of hnRNP K. This is in line with the recent work of Lacroix *et al.* [18], who demonstrated that human hnRNP K binds to the single-stranded telomeric C-block motif, although the SDS/PAGE mobility of CBNP γ appears to be slightly higher. However, it should be noted that the *Xenopus* homolog, for instance, has a much lower molecular mass than that of human hnRNP K [25]. The observation that the slower EMSA band contains CBNP α besides CBNP γ suggests that the two proteins interact, although homotypic protein dimers cannot be excluded. In any case, these complexes are less sequence-specific in their binding to the telomeric repeat probe d(CCCTAA)_n.

The presence of a reducing agent such as dithiothreitol appears to be required for the specific DNA-binding activity. In its absence, EMSA signals, when found, have low mobility and also lower sequence specificity. Nonetheless, they contain CBNP α as shown by *in situ* UV-cross-linking (data not shown). CBNP α therefore seems to need to be in the reduced state, as well as separated from CBNP γ , to be able to specifically interact with the C-rich telomeric repeat.

The identification of CBNP α with the type A/B hnRNP ssDBF raises two issues. The first concerns its possible function in relation to the activity of CBNP α , described here and in two preceding reports [19,20], and that of ssDBF, which has been characterized by Smidt *et al.* [21] as a transcription factor. Strong evidence from studies with cultured cells *in vivo* is required to be able to assign confidently a role to this protein. However, it should be noted that ssDBF was identified during screening of a liver cDNA library, whereas CBNP α was isolated from erythrocytes. It could well be that this protein has more than one role which is tissue dependent, and its activity could be regulated by both the relative amounts of the two isoforms (CBNP α and CBNP β , in our notation) and different post-translational modifications.

The second issue is more general and involves how the proteins studied in this work can be related to all similar proteins already known in other species. BLAST analysis [26] carried out using the protein and nucleotide databanks has revealed a set of proteins and/or cDNAs related to CBNP α /ssDBF (Fig. 6). Besides CRP1 from chicken [22], which differs completely from ssDBF in its N-terminal segment as the result of two frameshift mutations in its DNA coding sequence, the other proteins from mouse, rat, cat, and human cDNAs are homologous along their whole lengths, spanning 280–350 residues. The only other exception to this is PRM10 from rat (not shown; GenBank accession No. AF108653), the N-terminal region of which differs completely (also in its DNA coding sequence) from the others. Inspection of these sequences shows the following.

(a) The homology is very strong in the C-terminal portion of all these proteins, covering about 80% of the whole sequence. The 60–70 N-terminal residues of the mammalian sequences (except for PRM10) also display a remarkable similarity among them, to a lesser extent with the chicken protein ssDBF, and, obviously, none at all with CRP1.

(b) In the cases of rat and human sequences, two variants, which differ in the occurrence of an exon, are reported. The proteins with the exon are homologous to CBNP β (i.e. the second clone of ssDBF) in sequence and position of the exon.

REFERENCES

- Broccoli, D., Smogorzewska, A., de Chong, L. & Lange, T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.* **17**, 231–235.
- Conrad, M.N., Wright, J.H., Wolf, A.J. & Zakian, V.A. (1990) RAP1 protein interacts with yeast telomeres *in vivo*: overproduction alters telomere structure and decreases chromosome stability. *Cell* **63**, 739–750.
- Raghuraman, M.K. & Cech, T.R. (1989) Assembly and self-association of *Oxytricha* telomeric nucleoprotein complexes. *Cell* **59**, 719–728.
- Price, C.M. (1990) Telomere structure in *Euplotes crassus*: characterization of DNA–protein interactions and isolation of a telomere-binding protein. *Mol. Cell. Biol.* **10**, 3421–3431.
- Schierer, T. & Henderson, E. (1994) A protein from *Tetrahymena thermophila* that specifically binds parallel-stranded G4-DNA. *Biochemistry* **33**, 2240–2246.
- Sheng, H., Hou, Z., Schierer, T., Dobbs, D.L. & Henderson, E. (1995) Identification and characterization of a putative telomere end-binding protein from *Tetrahymena thermophila*. *Mol. Cell. Biol.* **15**, 1144–1153.
- Petracek, M.E., Konkel, L.M., Kable, M.L. & Berman, J. (1994) A *Chlamydomonas* protein that binds single-stranded G-strand telomere DNA. *EMBO J.* **13**, 3648–3658.
- Eid, J.E. & Sollner-Webb, B. (1997) ST-2, a telomere and subtelomere duplex and G-strand binding protein activity in *Trypanosoma brucei*. *J. Biol. Chem.* **272**, 14927–14936.
- Sarig, G., Weisman-Shomer, P., Erlitzki, R. & Fry, M. (1997) Purification and characterization of qTBP42, a new single-stranded and quadruplex telomeric DNA-binding protein from rat hepatocytes. *J. Biol. Chem.* **272**, 4474–4482.
- Uchiyama, F., Ohta, T. & Tanuma, S. (1996) Replication factor C recognizes 5'-phosphate ends of telomeres. *Biochem. Biophys. Res. Commun.* **229**, 310–315.
- McKay, S.J. & Cooke, H. (1992) hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGG_n. *Nucleic Acids Res.* **20**, 6461–6164.
- Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R. & 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.
- LaBranche, H., Dupuis, S., Ben-David, Y., Bani, M.R., Wellinger, R.J. & Chabot, B. (1998) Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat. Genet.* **19**, 199–202.
- Eid, J.E. & Sollner-Webb, B. (1995) ST-1, a 39-kilodalton protein in *Trypanosoma brucei*, exhibits a dual affinity for the duplex form of the 29-base-pair subtelomeric repeat and its C-rich strand. *Mol. Cell. Biol.* **15**, 389–397.
- Garcia-Bassets, I., Ortiz-Lombardia, M., Pagans, S., Romero, A., Canals, F., Aviles, F.X. & Azorin, F. (1999) The identification of nuclear proteins that bind the homopyrimidine strand of d(GA_nTC)_n DNA sequences, but not the homopurine strand. *Nucleic Acids Res.* **27**, 3267–3275.
- Ito, K., Sato, K. & Endo, H. (1994) Cloning and characterization of a single-stranded DNA binding protein that specifically recognizes deoxycytidine stretch. *Nucleic Acids Res.* **22**, 53–58.
- Tomonaga, T. & Levens, D. (1995) Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *J. Biol. Chem.* **270**, 4875–4881.
- Lacroix, L., Lienard, H., Labourier, E., Djavaheri-Mergny, M., Lacoste, J., Leffers, H., Tazi, J., Helene, C. & Mergny, J.L. (2000) Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres *in vitro*. *Nucleic Acids Res.* **28**, 1564–1575.
- Marsich, E., Piccini, A., Xodo, L.E. & Manzini, G. (1996) Evidence for a HeLa nuclear protein that binds specifically to the single-stranded d(CCCTAA)_n telomeric motif. *Nucleic Acids Res.* **24**, 4029–4033.
- Marsich, E., Xodo, L.E. & Manzini, G. (1998) Widespread presence in mammals and high binding specificity of a nuclear protein that recognises the single-stranded telomeric motif (CCCTAA)_n. *Eur. J. Biochem.* **258**, 93–99.
- Smidt, M.P., Russchen, B., Snippe, L., Wijnholds, J. & Ab, G. (1995) Cloning and characterisation of a nuclear, site specific ssDNA binding protein. *Nucleic Acids Res.* **23**, 2389–2395.
- Cvekl, A., McDermott, J.B. & Piatigorsky, J. (1995) cDNA encoding a chicken protein (CRP1) with homology to hnRNP type A/B. *Biochim. Biophys. Acta* **1261**, 290–292.
- Dreyfuss, G., Matunis, M.J., Pinol-Roma, S. & Burd, C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**, 289–321.
- Cartegni, L., Maconi, M., Morandi, E., Cobiainchi, F., Riva, S. & Biamonti, G. (1996) hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J. Mol. Biol.* **259**, 337–348.
- Siomi, H., Matunis, M.J., Michael, W.M. & Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.* **21**, 1193–1198.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
- Kamada, S. & Miwa, T. (1992) A protein binding to CARG box motifs and to single-stranded DNA functions as a transcriptional repressor. *Gene* **119**, 229–236.
- Lau, P.P., Zhu, H.J., Nakamura, M. & Chan, L. (1997) Cloning of an Apobec-1-binding protein that also interacts with apolipoprotein B mRNA and evidence for its involvement in RNA editing. *J. Biol. Chem.* **272**, 1452–1455.
- Wang, D. & Parrish, C.R. (1999) A heterogeneous nuclear ribonucleoprotein A/B-related protein binds to single-stranded DNA near the 5' end or within the genome of feline parvovirus and can modify virus replication. *J. Virol.* **73**, 7761–7768.
- Khan, F.A., Jaiswal, A.K. & Szer, W. (1991) Cloning and sequence analysis of a human type A/B hnRNP protein. *FEBS Lett.* **290**, 159–161.