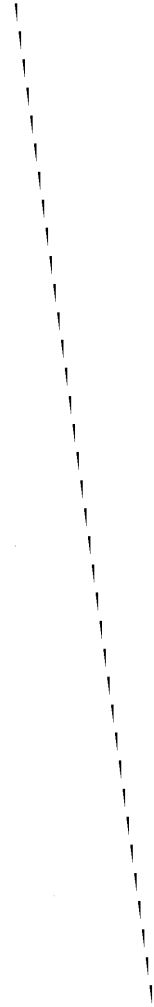


453043

60/1



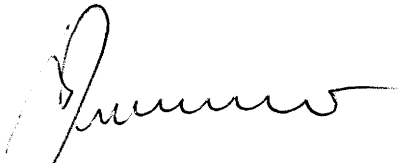
UNIVERSITA' DEGLI STUDI DI TRIESTE

Dipartimento di Biochimica, Biofisica e Chimica delle
Macromolecole

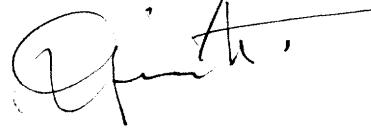
Tesi di Dottorato di Ricerca
in: BIOCHIMICA

**Titolo: Proteine associate col DNA: struttura e
funzione con riferimento all' attivazione/repressione
dell'espressione genica in sistemi normali e
neoplastici**

IL COORDINATORE (prof. B. de Bernard):



IL TUTORE (prof. V. Giancotti):



Dottoranda: Antonella Bandiera '65

cn

TRIESTE 1994

INDICE

<i>Introduzione</i>	<i>pag. 3</i>
<i>Materiali e metodi</i>	<i>pag. 30</i>
<i>Risultati e discussione</i>	<i>pag. 46</i>
<i>Conclusioni</i>	<i>pag. 67</i>
<i>Bibliografia</i>	<i>pag. 69</i>
<i><u>Appendice A</u>: elenco dei lavori pubblicati</i>		

RIASSUNTO

Il lavoro svolto durante questi quattro anni riguarda una serie di studi volti a caratterizzare alcuni componenti della cromatina in sistemi diversi. Utilizzando tecniche biochimiche ed affiancando a queste metodi immunologici e di biologia molecolare, sono state studiate proteine che hanno dimostrato di avere potenzialità regolatorie e non solamente funzioni di tipo strutturale, come si è pensato per molto tempo, nell'ambito dell'organizzazione del nucleo cellulare. Oltre allo studio di varianti istoniche e alcune forme intermedie simili correlate, sono stati portati avanti studi su sistemi biologici murini alterati in senso neoplastico, dal momento che era stato messo in evidenza precedentemente un incremento dell'espressione di alcune proteine nel fenotipo trasformato rispetto alla situazione che si riscontra in sistemi normali.

La nostra ricerca si è concentrata soprattutto in quest'ambito, allo scopo di identificare i vari componenti e caratterizzare eventualmente molecole non ancora note. Lo studio è stato quindi esteso a sistemi cellulari e tessuti di provenienza umana. Per verificare la validità del modello murino e confermare i risultati ottenuti, si è pensato di intraprendere un monitoraggio sul maggior numero di sistemi neoplastici possibile e ciò ha messo in evidenza la prospettiva di una possibile applicazione di tali metodi in campo diagnostico.

INTRODUZIONE

La più piccola unità vivente autonoma, la cellula, ha alla base della sua stessa essenza due fasi fondamentali: la proliferazione ed il differenziamento. Questi due eventi sono necessariamente di una complessità estrema dovendo essere sottoposti ad una perfetta regolazione, in modo che ogni cellula in un preciso momento, si trovi in un determinato stato adeguato alla funzione che dovrà svolgere e per il quale è stata programmata. Tutte le informazioni necessarie affinché ciò possa essere messo in pratica sono contenute nel codice genetico racchiuso nel nucleo: la lettura e la decodificazione al momento opportuno delle istruzioni determinano il destino di ciascuna cellula.

L'organizzazione del nucleo eucariotico e le interazioni dei suoi componenti sono già da molti anni estesamente ed attentamente studiati, non solo ai fini di capirne la struttura ed i meccanismi funzionali, ma anche in relazione a situazioni alterate di funzionamento che possono portare per esempio a patologie emerse prepotentemente in questo secolo, quali le neoplasie. Gli studi strutturali sul nucleo delle cellule eucariotiche hanno permesso di evidenziare la natura della cromatina e la sua organizzazione, le quali sono risultate essere estremamente simili anche in organismi evolutivamente lontani tra loro. Il materiale depositario dell'informazione genetica, il DNA, all'interno del nucleo è sempre associato a caratteristiche proteine, gli istoni, le quali sono fra le più conservate negli eucarioti. Esistono cinque tipi principali di istoni: H1 ed i "core histones" H2A, H2B, H3, H4, i quali si differenziano nella loro struttura primaria, ma sono accomunati dal fatto di avere un dominio globulare (identificato dalla zona resistente a digestione triptica) e delle estremità N-terminali flessibili particolarmente ricche in aminoacidi basici (in particolare lisina e arginina) i quali conferiscono loro carica positiva a pH fisiologico. Nonostante la loro struttura primaria fosse nota da molto tempo, l'esatta organizzazione spaziale del complesso istonico era piuttosto difficile da

determinare. Usando la diffrazione ai raggi X si riuscì ad ottenere una struttura a 7 Å di risoluzione che evidenziava le caratteristiche dell'interazione fra tali proteine e il DNA [1]. L'unità base da cui è costituita la cromatina è denominata nucleosoma che è una struttura formata da una parte centrale istonica attorno alla quale si avvolge la doppia elica che compie circa 1,75 giri impegnando 146 paia di basi (bp) (core particle) ed un altro tratto di DNA (detto "linker") che fuoriesce dalla particella [2-4]. La parte centrale è chiamata anche ottamero istonico, essendo formata da due copie di ciascuna molecola di H2A, H2B, H3 e H4. Recentemente la cristallografia a raggi X ha permesso la determinazione dell'ottamero istonico ad una risoluzione di 3.1 Å [5]. E' stata così evidenziata una struttura tripartita in cui una parte centrale formata da un tetramero (H3-H4)₂ è affiancato da due dimeri (H2A-H2B). L'elica di DNA si superavvolge sinistrorsa, contattando il tetramero centrale per circa 120 bp, e ciascuno dei due dimeri fiancheggiati per altre circa 20 bp. A questo punto, il DNA fuoriesce dalla zona di influenza degli istoni e prosegue per un breve tratto di connessione (linker) che precede la successiva strutturazione nel prossimo nucleosoma. L'interazione del nucleosoma con l'istone H1 forma un complesso chiamato cromatosoma in cui l'istone salda due giri completi di DNA corrispondenti a circa 166-168 bp attorno all'ottamero [6]. La successione dei cromatosomi dà luogo a una struttura "a filo di perle" costituisce la fibra di 10 nm [6, 7]. La zona di entrata ed uscita del DNA è quella che viene legata dall'istone H1, il quale prendendo contatto con le sue estremità flessibili sigilla in questo modo il nucleosoma bloccandolo rispetto all'elica di DNA. La presenza dell'istone H1 è stata associata alla ulteriore compattazione a cui può andare incontro il DNA [3, 6, 7] che prende l'aspetto di fibra di 30 nm, la quale è un ulteriore avvolgimento elicoidale di circa sei nucleosomi per spira. Recentemente indagini strutturali dirette hanno permesso di confermare tale livello di organizzazione e l'esatta posizione dell'istone H1 all'interno della fibra [8]. Il DNA risulta pertanto molto compattato e la maggior parte della cromatina del nucleo in interfase si trova strutturata in questo

modo [9]. In molte cellule eucariote infatti estese regioni del genoma (in particolare le aree corrispondenti alle sequenze telomeriche e centromeriche) sono impaccate nella cosiddetta eterocromatina, materiale cromosomiale particolarmente condensato che probabilmente per lo più è trascrizionalmente inattivo [7, 10]. La cromatina, a seconda della fase del ciclo cellulare in cui si trova la cellula va incontro a cambiamenti di struttura e composizione [11]. Gli studi biochimici hanno evidenziato che non più del 20% del genoma è organizzato in eucromatina o cromatina attiva in un determinato tipo cellulare in contrapposizione dal resto del materiale genomico e le differenze coinvolgono tutti i livelli di organizzazione, dal contenuto istonico dei nucleosomi ai cambiamenti sovrastrutturali. La cromatina attiva è meno compatta dell'eterocromatina e questo si riflette nella maggiore accessibilità alla doppia elica da parte di nucleasi esogene (come la DNasi I e la nucleasi micrococcica che tagliano preferenzialmente in corrispondenza di sequenze ricche in A/T) ed a sostanze chimiche che differiscono nella loro specificità per la sequenza nucleotidica o per la sua struttura. Questi sistemi sono stati usati per studiare le strutture della cromatina *in vitro* ed *in vivo*. Un determinato gene (o parte di esso) può essere sensibile a questi trattamenti in un certo tipo cellulare, mentre non lo è in un altro: la maggiore accessibilità alle nucleasi (siti ipersensibili) può rispecchiare un più libero accesso da parte delle molecole regolatrici sulla doppia elica e un diverso grado di impaccamento delle sequenze nucleotidiche nei due tipi cellulari diversi. [4, 6, 12]. Tradizionalmente sono state seguite due vie nell'intento di capire come il DNA si organizzasse in nucleosomi: le proteine istoniche venivano purificate e poi mescolate con DNA per osservare quali strutture potevano formarsi oppure venivano preparati estratti grezzi a partire da vari tipi cellulari che poi venivano addizionati del DNA. Saggiando tali sistemi con nucleasi esogene, il primo risultava costituito per lo più da complessi fra proteine e DNA della grandezza di 146 bp tipica per ottameri istonici molto compattati mentre nel secondo caso la dimensione del DNA variava tra 165 e 180 bp più simili e rappresentativi della

cromatina che si ritrova in natura [13]. La particella minima invariabile della cromatina è indicata come "core particle" corrispondente al DNA di 146 bp che esclude la regione di DNA "linker" il quale invece non ha una grandezza fissa, ma varia a seconda del tipo di cellula esaminata, per esempio da 15-20 bp nel lievito ad 80 bp negli spermatozoi del riccio di mare [6]. Ciò sta ad indicare che esistono altri fattori proteici che entrano in gioco nell'assemblaggio della cromatina, che di solito viene isolata come frammenti di materiale cromosomiale ed uno dei principali problemi pratici connessi con la sua analisi è proprio la perdita di eventuali componenti [14].

Fino a non molto tempo fa le proteine istoniche venivano considerate senza un ruolo attivo nei cambiamenti che avvengono all'interno del materiale nucleare: la loro funzione sembrava essere puramente strutturale, erano considerate semplice materiale di riempimento [3]. La stessa considerazione valeva anche per altre proteine che assieme agli istoni sono state fra le prime ad essere state scoperte, e cioè le proteine HMG (High-Mobility-Group) il cui nome fa riferimento ad una caratteristica alta mobilità elettroforetica che esse presentano quando analizzate su gel in acido acetico/urea. Anche per queste proteine relativamente abbondanti nel nucleo pareva plausibile l'ipotesi di componenti piuttosto inerti di sostegno del DNA [15, 16], specialmente dopo che si erano scoperti nel nucleo cellulare i fattori di trascrizione (*trans*-activating factors) e le sequenze regolatrici (*cis*-regulatory elements), che sembravano essere i preminenti responsabili nelle transizioni da uno stato all'altro della cromatina [17]. Inoltre la trascrizione *in vitro* che poteva essere ottenuta utilizzando estratti cellulari contenenti proteine regolatrici selezionate e DNA non legato agli istoni sembrava dare sostegno a questa ipotesi [3]. Contemporaneamente al lavoro di caratterizzazione della cromatina in vari sistemi biologici procedevano gli studi sulla regolazione genica ed in tal modo è andato delineandosi un contesto funzionale sempre più strettamente legato alla struttura. L'incrementata definizione biochimica dei complessi proteici che regolano la

trascrizione ha riportato in luce la questione del ruolo degli istoni e delle proteine cromosomiali non istoniche [17]. Di conseguenza, attualmente non solo gli istoni ma tutta la struttura nucleosomale vengono riconosciuti quali regolatori della trascrizione anche se i punti da chiarire sono sempre molti e per quanto riguarda le proteine HMG ci sono ormai le prime evidenze ed i primi modelli del loro coinvolgimento in questa attività. L'approccio genetico in particolare ha rivoluzionato la visione che si aveva sulla regolazione trascrizionale [3, 4, 6].

E' ora assodato che alcuni nucleosomi sono posizionati (nucleosome positioning), intendendo con quest'espressione il fatto che essi si trovano associati sempre con una ben determinata sequenza di DNA nella gran maggioranza delle cellule nell'ambito di uno stesso tipo cellulare [6]. In questo fenomeno probabilmente hanno un notevole peso più fattori, tipo la repulsione o attrazione che si instaura fra l'ottamero istonico ed il DNA in seguito a conformazioni specifiche dovute alla natura della sequenza nucleotidica di per se stessa, il contributo elettrostatico, le condizioni di forza ionica, la natura chimica degli ioni coinvolti, in particolare Ca^{++} e Mg^{++} e la presenza di determinati fattori proteici [4, 6, 9, 18]. E' stato dimostrato che un certo tipo di sequenze ripetute artificiali ricavate su base sperimentale, risultano più efficaci nel posizionamento dei nucleosomi rispetto le sequenze esistenti in natura note per avere tale proprietà (quali quelle che si trovano nel riccio di mare o nei geni per l'RNA ribosomiale 5S di *Xenopus laevis*) [19]. La proprietà intrinseca e non indotta di curvare il DNA (bending) dovuta alla sequenza nucleotidica si è rivelata essere anch'essa un elemento preferenziale per indurre la formazione dei nucleosomi: la flessibilità anisotropica (ovvero la tendenza a preferire un piano piuttosto che un altro nel ripiegamento che può derivare dalla disposizione delle sequenze ricche in G/C e quelle in A/T rispetto la periodicità della doppia elica) di sequenze fiancheggianti il punto di contatto fra l'ottamero la doppia elica fa sì che i solchi maggiori e minori si orientino spontaneamente rispetto la superficie proteica, indipendentemente cioè dalle interazioni di quest'ultima col

DNA [4, 6]. Proteine non istoniche con alta specificità ed affinità per particolari sequenze nucleotidiche possono letteralmente deformare la doppia elica impedendo o favorendo la formazione del nucleosoma in quell'area e questi argomenti depongono a favore di una non casuale localizzazione delle unità base della cromatina. Siccome in condizioni fisiologiche il nucleosoma è stabile, allo scopo di verificare una eventuale preferenzialità dell'ottamero istonico per sequenze specifiche sono state usate condizioni di alta forza ionica (circa dieci volte quella fisiologica) in modo da favorire così un rapido scambio DNA/proteine, successivamente è stata ripristinata nuovamente la forza ionica fisiologica per successiva diluizione. La differenza di affinità dell'ottamero per particolari sequenze così osservata, risulta essere al massimo di tre ordini di grandezza. Con molta probabilità questa relativamente modesta specificità dell'interazione DNA-ottamero può essere spiegata col fatto che in natura l'impaccamento in una struttura nucleosomale deve essere possibile per una vastissima gamma di sequenze. La presenza o meno di un nucleosoma su una determinata sequenza ne può determinare l'accessibilità o meno da parte di enzimi ed altre molecole, ma non è facile valutare esattamente quale sia il grado di posizionamento dei nucleosomi *in vivo* e conseguentemente quali conseguenze funzionali abbia [4]. Mentre i nucleosomi rappresentano un ostacolo per l'inizio della trascrizione, essi non necessariamente ne bloccano la prosecuzione, una volta che il processo è stato avviato. Vari esperimenti *in vitro* hanno messo in luce come il complesso di inizio sia in grado di avanzare lungo la doppia elica nonostante questa sia superstrutturata in nucleosomi, anche se la velocità di avanzamento può essere minore rispetto a quella che si verifica *in vivo* [3, 4, 20]. E' stato visto sperimentalmente che nelle regioni trascritte la struttura della cromatina va incontro a varie modificazioni, in particolare caratteristiche ricorrenti che vengono riportate negli studi fatti sono l'incrementata suscettibilità alla digestione con nucleasi, un cambiamento nella distribuzione dei nucleosomi, cambiamenti nel grado di torsione nel DNA, la diminuita presenza di istoni (in

particolare dell'H1 e di H2A e H2B) e una loro modificazione [4, 6, 20-23]. Nonostante la quantità di lavoro svolto e di dati disponibili, la situazione non è ancora ben chiarita; ad esempio, se da un lato la cromatina trascritta intensivamente nel caso dei geni ribosomiali appare organizzata in una forma estesa senza nucleosomi, mediamente i geni trascritti risultano mantenere una struttura nucleosomale per quanto se ne può dedurre dai saggi effettuati tramite la digestione con nucleasi. La difficoltà (anche tecnica) sta nel fatto che in quest'ultimo caso è arduo determinare se un particolare gene è trascritto ed ha allo stesso tempo una struttura nucleosomale, questa infatti potrebbe essere presente in una copia o in una parte del gene non coinvolta nel processo di trascrizione in quel determinato momento [23]. Gli studi condotti sulla struttura del materiale genetico sono stati correlati con i cambiamenti nell'attività genica ed è emerso che l'attivazione di quest'ultima è associata alla decondensazione della cromatina in corrispondenza delle zone in cui ciò si verifica. La struttura della cromatina è quindi attualmente considerata associata non solo alla potenziale repressione del processo di trascrizione, ma anche al suo innesco ed persino al suo potenziamento ed in tutto ciò hanno naturalmente parte gli istoni e le proteine cromosomiali [2-4, 6, 7, 10, 11, 17, 22, 24, 25].

Alcuni sistemi biologici sono stati estesamente studiati e caratterizzati, tanto che ora essi costituiscono dei modelli fondamentali su cui si basano le attuali teorie che considerano la cromatina come parte essenziale nei processi che hanno luogo nel nucleo cellulare. L'attività genica tessuto-specifica è stata studiata estesamente nel pollo e nel topo, confrontando l'assetto del gene per la β -globina in cellule eritroidi in cui esso è attivato o inducibile e quelle di altri tessuti, dove il gene non è attivamente espresso. In uno studio condotto su tessuti murini, è stata evidenziata una struttura nucleosomale con lo stesso posizionamento lungo l'intera lunghezza del gene eccetto che per una zona di circa 700 bp da cui quattro nucleosomi risultavano alterati o spazzati nel tessuto eritroide. Tale zona contiene delle

sequenze regolatorie coinvolte nell'induzione della trascrizione ed è delimitata da siti ipersensibili (HS, hypersensitive sites). Molto spesso tali siti coincidono con sequenze *cis*-regolatorie che possono essere legate da uno specifico fattore transattivante. E' pertanto probabile che ciò rifletta un arrangiamento nucleosomale delle sequenze regolatorie nei geni repressi che ostacola il legame di eventuali molecole in grado di attivare il gene, le quali sono presenti nei sistemi in cui il gene è appunto in uno stato di attivazione o inducibilità [6, 10].

Un altro sistema biologico che è diventato un modello per lo studio delle correlazioni fra funzionamento e struttura della cromatina è quello del LTR (long terminal repeat) del virus del tumore mammario del topo (MMTV, mouse mammary tumor virus) il quale è inducibile sotto l'azione di ormoni glucocorticoidi ed *in vivo* si trova incorporato in sei nucleosomi posizionati [26]. Tali nucleosomi prevengono l'attivazione basale del promotore del LTR impedendo il legame da parte del complesso di inizio della trascrizione in assenza di ormoni [25]. In assenza e in presenza di ormoni i nucleosomi hanno lo stesso posizionamento ma nel caso della presenza dell'ormone una regione di circa 200 bp (corrispondente grossomodo ad un nucleosoma) risulta sensibile al trattamento con nucleasi. In questa zona sono contenute le sequenze (GREs, glucocorticoid receptor responsive elements) con cui è in grado di interagire, in seguito all'attivazione, il complesso proteico ormone/recettore nonostante il nucleosoma sia posizionato. Una volta che tale legame si è formato si verifica un'alterazione della struttura nucleosomale, il fattore di trascrizione NF1 viene messo in grado di legarsi al promotore ed a sua volta permette il legame di altri fattori (fra cui quello che lega la sequenza TATA, TFIID) che portano così all'inizio della trascrizione dei geni a valle [4, 6, 25, 26]. Non è noto l'esatto meccanismo con cui la cromatina viene decondensata, ma si sa che alcune fra le proteine che entrano in gioco causano la rimozione dell'istone H1 dal DNA linker nella regione che deve essere trascritta [25, 27]. In questo caso è stato dunque indotto lo spiazzamento di un nucleosoma che altrimenti avrebbe impedito

l'inizio della trascrizione; ci sono però molti esempi di come esistano anche elementi regolatori che invece di trovarsi avvolti attorno all'ottamero istonico, sono situati nel DNA linker. Nel caso del lievito (*Saccharomyces cerevisiae*) il gene per la fosfatasi acida (*PHO5*) viene indotto se le cellule vengono fatte crescere in assenza di fosfato. E' stato dimostrato che quando il gene viene indotto, una serie (cluster) di quattro nucleosomi posizionati nella regione del promotore viene rimossa o alterata. Quando il gene è inattivo, uno dei nucleosomi copre un paio di siti di legame per i fattori transattivanti PHO2 e PHO4, mentre un secondo sito per quest'ultimo fattore si trova nel linker fra il nucleosoma e quello posizionato immediatamente a monte. L'approccio genetico ha permesso di stabilire che PHO2 e PHO4 sono essenziali per dar luogo a cromatina attiva e dati sperimentali indicano che sotto lo stimolo dell'induzione PHO4 si lega al rispettivo sito che si trova nel linker e, di concerto con PHO2, spiazza il nucleosoma rendendo disponibile il secondo sito di legame. Il ritorno allo stato inattivo cioè il ripristino della struttura nucleosomale segue il venire meno dello stimolo [2, 4, 17, 22]. In questi due casi, la struttura nucleosomale, che rappresenta un ostacolo per l'accesso alle sequenze di controllo, può essere alterata da uno stimolo in grado di renderle disponibili e che è legato alle condizioni dell'ambiente esterno alla cellula (induced displacement). Un altro tipo di controllo della trascrizione negli eucarioti è quello dell'esclusione permanente dei nucleosomi da una sequenza di controllo (persistent displacement). Tale strategia è riportata per molti geni che servono per la gestione ordinaria della cellula e quindi vengono trascritti costitutivamente (housekeeping genes) ma anche per alcuni geni inducibili per i quali è necessario un arrangiamento tale da consentire l'immediata risposta qualora le condizioni lo richiedessero [2]. Un esempio viene fornito dall'arrangiamento dei promotori dei geni *hsp* (heat shock protein) della *Drosophyla*, i quali vengono attivati in seguito all'incremento della temperatura. Alcuni di questi geni contengono siti ipersensibili alla DNasi I negli elementi a monte della regione trascritta, sia prima che dopo l'induzione dovuta a shock

termico [6]. Nelle cellule di *Drosophyla* questo sistema è in uno stato di equilibrio ("poised" state), con l'RNA polimerasi II già posizionata e pronta a trascrivere. L'inizio della trascrizione e quindi l'inscospicuo dell'attività polimerasica, è subordinato al legame del HSF (heat shock factor), il quale a sua volta è condizionato dal legame del fattore GAGA nei rispettivi siti regolatori [28]. Si ritiene che in questo caso alcuni fattori riescano a posizionarsi nei loro siti di legame sulla doppia elica in un momento in cui questa non è ancora ingaggiata nel legame con gli istoni, spiazzando così questi ultimi e favorendo l'assetto della cromatina in questo stato di "pre-attivazione". Una tale opportunità può verificarsi subito dopo la replicazione del DNA, prima che la doppia elica si strutturi in nucleosomi [2].

La regolazione può prevedere anche l'utilizzo combinato delle strategie prese in considerazione. I geni bidirezionali *GAL1/GAL10* del lievito che codificano per enzimi coinvolti nell'utilizzazione del galattosio sono attivati dalla presenza di quest'ultimo nel mezzo di coltura delle cellule. I due geni vengono trascritti in maniera divergente a partire da una regione intergenica centrale regolatrice nella quale si trova sia un sito di legame per il fattore inducibile GAL4 (il quale è in grado di rimodellare la cromatina) che le sequenze per l'attivazione di *GAL1* e *GAL10* [6]. La zona con cui interagisce GAL4 è costitutivamente occupata da un fattore che esclude la formazione di nucleosomi mentre le sequenze per l'attivazione di *GAL1* e *GAL10* sono invece contenute in tratti di DNA a struttura nucleosomale che deve venir alterata tramite l'induzione affinché siano trascritti i geni [2].

I dati sperimentali sembrano indicare che ci siano due tipi di meccanismo in grado di rimodellare la struttura della cromatina. Il primo, di "competizione dinamica" (dynamic competition) non dipende dalla replicazione del DNA per rendere disponibili gli elementi regolatori ai fattori di attivazione e ciò si verifica nei casi in cui i siti di legame per le molecole regolatrici si trovino nel DNA linker oppure correttamente posizionati nel nucleosoma e quindi sufficientemente esposti ed anche nel caso in cui entrino in gioco altri fattori in grado di legarsi in zone adiacenti e

alterare la conformazione tanto da sbloccare i siti. Il secondo meccanismo (pre-emptive competition) dipende invece dalla replicazione e interessa i geni i cui elementi regolatori sono associati con gli istoni dell'ottamero in modo tale da essere totalmente inaccessibili alle molecole regolatrici nelle cellule quiescenti. In questo caso solo in seguito alla duplicazione del DNA è possibile un cambiamento nell'assetto di queste regioni. I siti vengono esposti almeno parzialmente alla forcella di replicazione, prima che avvenga nuovamente l'assemblamento della cromatina ed è necessario che l'interazione fra regione regolatrice e rispettivo fattore sia stabile al punto di contrastare l'assemblamento in nucleosomi [17].

Il posizionamento dei nucleosomi può dunque essere condizionato da altre molecole regolatrici e diventare in questo modo un elemento di per se stesso regolatore per la trascrizione, anche in senso positivo [24]. Gli studi sul promotore di uno dei geni "heat shock" (*hsp26*), in cui è stato possibile mappare *in vivo* un nucleosoma posizionato, hanno portato all'elaborazione di un modello che spiega come la formazione di questa struttura possa consentire a due sequenze regolatrici distanti fra loro di venire a trovarsi vicine spazialmente e di contribuire all'attivazione. In questo caso la posizione esatta dei contatti fra DNA e ottamero non dipende solo dalla sequenza a cui gli istoni si legano, ma anche dalle sequenze immediatamente vicine. Queste infatti costituiscono il sito di legame per il fattore GAGA, il quale così può fungere da segnale che delimita esattamente la posizione del nucleosoma. Due sequenze che sul filamento di DNA sono situate a più di un centinaio di basi di distanza, possono venire così a trovarsi fisicamente vicine ed essere disponibili come siti di legame per l'HSTF (heat shock transcription factor), il quale è la proteina in grado di regolare il gene [25]. Quanto proposto può contribuire anche a dare una spiegazione a fatti che sono stati osservati quale il distanziamento, talora notevole, fra elementi regolatori sul DNA. E' questa una caratteristica che ricorre spesso ma fino ad ora non era mai stata interpretata in modo soddisfacente. Inoltre, sembra che tale modello abbia trovato una prima

dimostrazione sperimentale in uno studio sul promotore per il gene della vitellogenina B1 in *Xenopus*. Usando cromatina ricostituita *in vitro*, si è visto che la formazione di un nucleosoma in una ben determinata posizione a monte del sito di inizio potenziava l'aumento della trascrizione dovuta all'estrogeno. La delezione totale o parziale del tratto di DNA coinvolto nella formazione del nucleosoma provoca l'abolizione dell'effetto stimolante dell'estrogeno [25, 29].

I geni per l'rRNA 5S sono stati fra i sistemi eucariotici più studiati dal punto di vista del posizionamento proprio per l'inclinazione delle loro sequenze a strutturarsi in un arrangiamento nucleosomico ben definito. Questo sistema è stato molto utile per chiarire come i contatti tra gli istoni ed un elemento regolatorio posizionato in un nucleosoma possano essere modificati per regolare l'accessibilità ai fattori di trascrizione, indicando così il potenziale ruolo dei singoli istoni o di parti di essi e delle loro modificazioni [22, 24, 25]. L'assemblamento è un processo ordinato, in cui il tetramero (H3-H4)₂ è il primo complesso che entra in contatto con il DNA ed a cui spetta il maggior ruolo nell'organizzazione del nucleosoma riconoscendo nella doppia elica le sequenze che dirigono il posizionamento, solo in un secondo tempo si posizionano i dimeri (H2A-H2B) ed infine, a completare la struttura, l'ultimo ad essere incorporato è l'istone H1 [3, 7, 24]. Nel caso dei geni ribosomiali 5S, il tetramero posizionato non arriva ad estendere i suoi contatti con la doppia elica sino al sito di legame per il fattore di trascrizione TFIIIA che invece si trova nella zona di interazione di uno dei due dimeri (H2A-H2B). Di conseguenza, il fattore TFIIIA forma un complesso a tre assieme al tetramero ed al DNA. A questo punto sono possibili due vie che portano i geni rispettivamente ad uno stato di repressione, nel caso in cui il dimero (H2A-H2B) rimpiazza il fattore TFIIIA portando alla formazione di un nucleosoma o di attivazione se altri fattori si legano oltre a TFIIIA spiazzando il tetramero (H3-H4)₂ [22, 25]. Questo modello ha permesso di stabilire che la modificazione post-traduzionale degli istoni dell'ottamero, in particolare l'acetilazione, può influenzare in modo determinante l'accessibilità della cromatina

da parte di molecole regolatrici come TFIIIA [7, 30]. L'ottamero è una struttura proteica in cui i domini globulari dei singoli istoni interagiscono, mentre i segmenti corrispondenti alle parti N-terminali sono rivolti verso l'esterno, probabilmente in grado di interagire (non è noto come) con il DNA che si avvolge attorno alla particella. Inoltre, si è visto che tali domini non hanno grande influenza sull'assemblaggio e la stabilità dei nucleosomi [3, 6, 23]. Le sequenze primarie nel loro complesso ed in particolare le posizioni di determinati residui sono estremamente conservate in una grande varietà di organismi che vanno dal lievito all'uomo. La genetica del lievito ha contribuito a chiarire come il dominio N-terminale delle proteine istoniche con le sue possibili modificazioni sia coinvolto nell'attivazione e repressione dei geni. Studiando mutanti puntiformi e di delezione per il dominio terminale dell'istone H4, è emerso come a tale regione proteica siano attribuibili vari effetti, talora contrastanti. In particolare, è stato identificato un effetto inibitorio che si manifesta su regioni genetiche chiamate loci silenti di coniugazione, le quali regolano appunto quest'ultimo processo. Esse devono rimanere permanentemente inattive affinché nella cellula aploide di lievito venga determinato il "tipo di coniugazione" (mating type) e possa così avvenire la coniugazione con un'altra cellula appartenente all'altro tipo. Il fatto che alcune delezioni all'N-terminale dell'H4 causino la specifica derepressione dei loci silenti facendo notevolmente diminuire la capacità di coniugarsi del lievito dimostra un diretto coinvolgimento dell'istone e di una struttura nucleosomale nell'effetto repressivo [3, 6].

Da molto tempo è noto che la trascrizione ed altri processi di transizione strutturale della cromatina sono correlati alla presenza (o assenza) di modificazioni delle proteine istoniche. Un ruolo di rilievo sicuramente spetta all'acetilazione che, aggiungendo enzimaticamente gruppi acetilici sulle code istoniche in corrispondenza del gruppo aminico in ϵ di determinati residui di lisina elimina la carica positiva [3, 6, 23, 31]. I siti ipersensibili nella cromatina sono anche siti in cui è presente

acetilazione degli istoni; il gene attivo della β -globina in eritrociti di pollo risulta associato con istoni iperacetilati, al contrario di quello inattivo per l'ovalbumina. Una stretta correlazione con l'attività genica è stata evidenziata dalla carenza di questa modificazione nel cromosoma X inattivo in cellule femminili sia murine che umane, a livello dell'istone H4 e per tutti e quattro i siti in cui questo può avvenire. I loci silenti nel lievito sono pure specificamente associati con una scarsa acetilazione degli istoni dell'ottamero [7].

L'acetilazione degli istoni (H3 e H4 in particolare) è fortemente correlata con l'assemblamento della struttura nucleosomale sulla doppia elica di DNA in seguito alla replicazione: *in vivo* la sintesi del DNA è associata con quella degli istoni che vengono subito incorporati nel filamento neosintetizzato e vengono poi progressivamente deacetilati col maturare della struttura cromatinica [7]. Ci sarebbe quindi una fase in cui le proteine acetilate faciliterebbero l'accesso a molecole regolatrici che sarebbero in grado di conferire alla cromatina un determinato assetto, mentre la successiva deacetilazione potrebbe servire a "congelare" la cromatina in una condizione di pre-attivazione o repressiva a seconda dei casi [7, 30]. Molti dati sperimentali indicano che l'aumentata acetilazione degli istoni dell'ottamero è associata ad una diminuita presenza dell'istone H1 nella cromatina [7, 11]. Per molto tempo l'istone del linker è stato visto quale parte di un generico meccanismo atto ad assicurare una forte e stabile repressione tessuto specifica, sulla base dell'idea che la superstruttura della cromatina, le fibre di 10 e 30 nm in cui l'H1 è coinvolto, assicura l'inattivazione dei geni. Esperimenti *in vitro* hanno effettivamente dimostrato che l'entità della trascrizione dipendeva in maniera inversamente proporzionale alla quantità di istone H1 presente nel sistema [32]. Inoltre, sempre *in vitro*, si è visto che alcuni fattori di trascrizione caratterizzati come forti attivatori trascrizionali sia *in vitro* che *in vivo* erano in grado di contrastare la repressione mediata dall'H1 anche se questa si era già stabilita, sia su uno stampo di DNA che su nucleosomi ricostituiti [2, 17, 33, 34].

In realtà nei sistemi studiati non sempre la presenza dell'H1 si correla perfettamente con l'assenza di attività genica e ciò è stato visto sia tramite esperimenti di tipo biochimico dove l'istone risultava associato anche a geni attivi, sia utilizzando tecniche di immunolocalizzazione; si è fatta così strada l'ipotesi che l'H1 non sia assente dalle zone attive della cromatina, ma piuttosto presente in una forma modificata. Ciò causerebbe un indebolimento del suo legame col DNA che ne potrebbe facilitare la perdita selettiva durante l'estrazione [6, 32]. La principale modificazione a cui va incontro l'H1 è la fosforilazione, che avviene per aggiunta di un gruppo fosfato sui residui di serina e treonina nelle code basiche terminali. A differenza degli istoni dell'ottamero, nell'H1 il dominio globulare costituisce grossomodo la parte centrale della molecola che è quella in grado di interagire con la particella nucleosomale. Le due code amino e carbossiterminale che sporgono sono importanti per l'interazione col DNA linker e la condensazione della cromatina [11]. La fosforilazione, introducendo cariche negative nella molecola proteica influenza, indebolendolo, il suo legame con il DNA. A seconda dell'entità del fenomeno, l'interazione DNA/proteina può essere modulata e ciò viene regolato sia durante le varie fasi del ciclo cellulare che in relazione all'attività trascrizionale. Una caratteristica diffusa fra gli eucarioti è la non specifica, massiva fosforilazione a cui va incontro l'istone H1 durante la mitosi, la quale è reversibile [7, 11, 32, 35]. È stato identificato l'enzima che catalizza tale fosforilazione dell'istone H1 (growth-associated H1 kinase, poi identificata con MPF, maturation promoting factor che è il complesso formato dalla chinasi p34^{cdc2} e da una delle cicline, il cui motivo di consenso è S/TPXB, dove X è un aminoacido qualsiasi e B indica uno basico) ed è stato visto che questo è un fattore chiave per l'avanzamento della mitosi, tanto che la fosforilazione di H1 esogeno aggiunto in un estratto cellulare è diventato un metodo comunemente accettato per stabilirne la progressione del ciclo cellulare [11, 35]. L'attività fosforilativa sull'H1, ricavata da cellule mitotiche può indurre condensazione della cromatina in estratti da cellule quiescenti, mentre la sua

inibizione, con conseguente defosforilazione dell'H1 porta a parziale decondensazione [11]. La coincidenza temporale fra la condensazione del DNA e la fosforilazione dell'H1 aveva fatto inizialmente pensare a un suo possibile ruolo nell'innescare della condensazione cromosomica, ma in seguito, sono emerse evidenze per un ruolo prioritario in questo processo di altre molecole quale l'enzima topoisomerasi II [35]. Considerando inoltre che l'introduzione di cariche negative causa l'indebolimento dell'interazione fra istone e DNA ed il fatto che in alcuni determinati sistemi (i macronuclei amitotici del Protozoo *Tetrahymena*, gli stadi finali dello sviluppo degli eritrociti di pollo e il compattamento della cromatina durante la spermatogenesi del riccio di mare) uno stato condensato della cromatina è associato a defosforilazione dell'istone H1, non è facile stabilire esattamente il ruolo giocato da tale modificazione nel condensamento cromosomiale. Alcune considerazioni devono essere tenute in conto per valutare le evidenze sperimentali esposte: nei tre esempi citati prendono parte delle varianti specifiche dell'istone H1 il cui ruolo potrebbe essere specializzato nel rimodellamento della cromatina; la modificazione di alcuni siti piuttosto che altri potrebbe avere effetti differenti e specifici all'interno della stessa molecola, influenzandone il comportamento a seconda dei casi; la natura dinamica della fosforilazione e defosforilazione dell'H1 potrebbe influenzare l'accessibilità della cromatina da parte di altre proteine non istoniche che potrebbero avere un ruolo principale nei successivi rimodellamenti, anche in concerto con altre modificazioni concomitanti che possono avvenire sui componenti del sistema (ad esempio acetilazione degli istoni dell'ottamero) [11].

Durante il ciclo vitale della cellula ci sono processi che richiedono una successione di cambiamenti nella cromatina, ad esempio durante la proliferazione, ad una fase di sintesi proteica in cui il materiale genetico deve essere disponibile per fare da stampo per la trascrizione dei geni in RNA messaggero, si succede la duplicazione, in cui lo stesso materiale deve essere ricopiato quindi di nuovo disponibile ed ancora segue una fase in cui è richiesta una estrema condensazione di

questo materiale (raddoppiato in quantità) per poter essere suddiviso precisamente nelle due cellule figlie. Lo stadio successivo, il differenziamento, può essere visto come una maturazione delle cellule che hanno proliferato, le quali assumono tramite questo processo delle caratteristiche diverse che le porteranno a ricoprire differenti funzioni. A seconda del tipo di cellula, sarà necessario di volta in volta che una determinata serie di geni sia attiva o prontamente attivabile, mentre altre regioni del genoma dovranno rimanere repressi. Esistono molte distinte varianti dell'istone H1 nei vari organismi. La composizione in subtipi dell'istone H1 può variare sia in relazione allo sviluppo che da tessuto a tessuto all'interno di uno stesso organismo [7, 32]. Un meccanismo tramite cui ci può essere un controllo nella regolazione della trascrizione da parte dell'H1 è connesso proprio con l'esistenza delle sue diverse subfrazioni le quali differiscono nella loro capacità di interagire col DNA e di alterare la struttura della cromatina [32]. L'istone H5 (la forma predominante dell'istone del linker negli eritrociti aviari) può essere considerato, ad esempio, una variante estrema dell'istone H1 ed ha maggiore affinità per la cromatina rispetto quest'ultimo. [36]. La struttura del dominio globulare dell'istone H5 è stata determinata con tecniche di cristallografia a 2.5Å di risoluzione ed è ora pertanto possibile ipotizzare il comportamento nell'interazione col DNA delle varianti istoniche note confrontando le loro sequenze aminoacidiche [7, 36]. E' interessante notare come alcune varianti presenti durante i primi stadi di sviluppo di certi organismi risultino avere un dominio C-terminale più ricco in residui acidi e ciò potrebbe facilitare l'accesso ad altri fattori nella cromatina. Studi sui geni per l'rRNA 5S in *Xenopus* indicano che quando l'H1 rimpiazza la variante embrionale durante il differenziamento, i geni dell'rRNA 5S ed altri geni di classe III fino a quel momento attivi, diventano inaccessibili ai fattori di trascrizione ed alla RNA polimerasi III. Una rimozione della proteina permette la riattivazione della trascrizione mentre la sua aggiunta reprime specificamente certi geni di classe III [7]. L'eterogeneità del contenuto di istone H1 in ciascun tipo cellulare, in cui sempre risultano espresse più

varianti, suggerisce l'esistenza di un meccanismo di fine modulazione dell'assetto della cromatina a seconda delle diverse condizioni funzionali in base alle quantità relative fra i vari sottotipi. Un'altra variante ben nota dell'H1, è l'H1° che è espresso tipicamente in cellule totalmente differenziate di mammifero. Un confronto con la principale classe di isoforme dell'H1 mette in evidenza differenze nei domini N-terminale e C-terminale. L'estremità aminotermine della proteina sembra importante per il posizionamento del dominio globulare fra i punti di entrata ed uscita del DNA nel nucleosoma, per cui differenze in questa zona potrebbero indicare diversa funzionalità a questo livello. Nel dominio C-terminale si riscontra variabilità anche relativamente alla composizione aminoacidica; in questo caso le conseguenze potrebbero ripercuotersi sulla struttura di ordine superiore della cromatina, dal momento che tale parte della molecola è principalmente responsabile nell'aggregazione delle catene nucleosomali [37]. Un notevole numero di studi riguarda la caratterizzazione dei sottotipi dell' H1 in differenti tipi di cellule e di tessuti, per cui oltre all'H1°, che rappresenta un caso estremo similmente all'istone H5, sono state identificate molte altre isoforme della specie principale. Da tutti questi dati emerge il fatto che ci sono molte varianti delle quali ciascuna può presentare a sua volta microeterogeneità a livello di sequenza primaria ed a ciò si aggiunge il fatto che in ciascuna specie possono intervenire più modificazioni post-traduzionali. La situazione è quindi estremamente complessa e tutto è reso ancora più difficile dal fatto che le tecniche utilizzate tradizionalmente non riescono a dare risultati facilmente confrontabili. Nel nostro laboratorio viene portata avanti una caratterizzazione di sistemi neoplastici tramite lo studio di alterazioni della cromatina che possono verificarsi in seguito a trasformazione tumorale rispetto a sistemi biologici normali di controllo. L'istone H1 e le sue subfrazioni sono stati pertanto oggetto d'indagine in questi sistemi, allo scopo di stabilirne la composizione, la distribuzione e mettere in evidenza le eventuali differenze. E' stata fatta una precisa identificazione delle varie isoforme presenti nei sistemi studiati

sulla base dei dati disponibili in letteratura tramite l'utilizzo di tecniche tradizionali quali cromatografia ed elettroforesi, a cui è stata affiancata la spettrometria di massa. In tal modo è stato possibile caratterizzare questo tipo di molecole in assenza di dati completi e precisi sulla loro sequenza primaria. E' stata inoltre fatta un'indagine di tipo quantitativo sull'espressione di queste proteine in relazione ad altri componenti non istonici della cromatina [38].

Varie situazioni dunque possono indurre le cellule ad utilizzare forme specializzate di istoni per modificare l'organizzazione della cromatina. Un caso limite è rappresentato dal processo della spermatogenesi, durante la quale la cellula va incontro a cambiamenti particolarmente drastici che la porteranno ad essere in condizioni di poter svolgere la sua funzione. Raggiunto lo stadio finale di maturazione, il nucleo cessa la sua attività genica e il DNA deve venir altamente condensato in uno stato inerte, il quale dovrà essere revertito dopo la fecondazione. Una caratteristica comune nel passaggio da cellula diploide a cellula spermatica matura è la sostituzione dei normali componenti istonici della cromatina con proteine specifiche neosintetizzate in grado di legarsi al DNA e conferire al materiale genetico le caratteristiche necessarie. In natura si ritrova tutta una serie di modificazioni nella composizione del contenuto proteico della cromatina per cui da situazioni in cui gli istoni somatici vengono sostituiti o soltanto affiancati da quelli spermatici si passa a casi in cui c'è una completa sostituzione di proteine di un altro tipo, le protamine. La varietà delle proteine basiche spermatiche è notevole e pure la loro distribuzione fra le specie appare irregolare rispetto alla filogenesi, nonostante ciò sia in aperto contrasto con l'alto grado di conservazione mantenuto dagli istoni [39]. Sono state individuate quattro grandi categorie di proteine basiche spermatiche in base all'aumento del grado di differenza dagli istoni somatici. Nei molluschi bivalvi, in particolare, è stata riscontrata la coesistenza di istoni e di proteine spermatiche specifiche ma, a livello della cromatina non si sa molto sul tipo di struttura che risulta dall'interazione fra tali componenti ed il DNA. Nel nostro

laboratorio è stato intrapreso lo studio di tali proteine dal bivalve *Ensis minor* allo scopo di stabilirne la natura e di caratterizzare l'organizzazione della cromatina di questo sistema [40]. La conoscenza dei dettagli strutturali delle proteine spermatiche e la loro relazione con altri componenti ben noti della cromatina può infatti concorrere a determinare con maggior precisione il ruolo funzionale svolto da tutte queste molecole. Sono stati quindi individuati componenti caratteristici dagli estratti da spermatozoi maturi; sono state utilizzate tecniche di biologia molecolare al fine di ottenerne la sequenza primaria completa clonando il cDNA ed è stato fatto un confronto con la situazione riscontrata in altri organismi simili, anche in considerazione del fatto che nei molluschi bivalvi sussiste un modello di evoluzione radiale in cui gradi diversi di cambiamento risultano particolarmente utili ai fini di uno studio comparato.

Come si è visto oltre agli istoni e le loro varianti, altre componenti proteiche contribuiscono a determinare la struttura della cromatina e prendono parte ai profondi cambiamenti che vi avvengono in seguito all'innescio di processi cellulari quali proliferazione e attività genica, differenziamento, repressione [2-4, 6, 7, 9-11, 17, 21-25, 28, 41, 42]. Senza dubbio elementi che rientrano in questo quadro sono le proteine HMG, un gruppo di componenti non istonici della cromatina in grado di interagire col DNA estremamente diffusi e conservati in molti organismi eucarioti; nonostante siano conosciuti da molto tempo e si abbiano moltissimi dati su di essi, il loro ruolo non è ancora definito con certezza, [6, 7, 16, 24, 41, 42]. Innanzitutto c'è da dire che tale gruppo di proteine non è omogeneo, essendo derivato da una classificazione che si basa sulle loro caratteristiche chimiche e fisiche, quali l'estraibilità in 0,35 M NaCl o in 5% di acido perclorico, alto contenuto in aminoacidi carichi, di cui pressapoco metà acidi e metà basici, peso molecolare sotto i 30 kD, elevata mobilità elettroforetica in acido acetico/urea [16, 42]. Di fatto, gli studi che si sono susseguiti hanno portato alla distinzione di tre sottogruppi, tenendo in considerazione anche altri criteri, quali l'omologia della sequenza

primaria, caratteristiche immunologiche, localizzazione, presunte funzioni. Il primo comprende le HMG 1 e 2 che hanno peso molecolare intorno ai 25 kD e sono presenti in maniera preponderante all'interno del nucleo, rispetto le altre HMG. Queste proteine interagiscono col DNA, preferibilmente con un singolo filamento, non è stata evidenziata una particolare specificità di sequenza [42]. E' stato dimostrato che l'HMG 1 è in grado di mediare la chiusura circolare di frammenti di DNA di 66 bp, troppo poco flessibili perchè si richiudano spontaneamente e più distorti rispetto il DNA che si ripiega nel nucleosoma, pertanto ciò sta ad indicare che la proteina è in grado di indurre piegamenti nel DNA (bending) [43]. Altre evidenze hanno fatto pensare che esista una preferenza per l'interazione con strutture inusuali della doppia elica dal momento che la proteina è in grado di interagire con il DNA cruciforme *in vitro* e questo tipo di struttura (chiamata anche four-way junction) è probabilmente presente anche *in vivo*, per esempio in corrispondenza di regioni in cui avviene la ricombinazione [43, 44]. Le sequenze primarie delle proteine hanno messo in evidenza un grado molto alto di conservazione fra i vari organismi presi in esame e spesso le differenze consistono in sostituzioni conservative [42]. La struttura è tripartita con i primi due domini A e B omologi fra loro globalmente carichi positivi ed il terzo, C-terminale, di natura acida, che può interagire con l'istone H1 *in vitro* anche se, molto recentemente, studi biochimici hanno messo in dubbio la natura puramente elettrostatica dell'interazione ed hanno proposto una specificità d'interazione fra HMG 1 ed H1 [45]. In molte altre proteine capaci di interagire col DNA identificate ben più tardi, si sono riscontrate regioni aminoacidiche molto simili ai domini omologi delle HMG 1/2. E' stato così riconosciuto un "HMG box", motivo con caratteristiche peculiari in grado di legare il DNA che si ritrova in più proteine anche diverse fra loro e che presumibilmente ha una particolare struttura e funzione. L' uso della risonanza magnetica nucleare ha permesso recentemente di individuare all' interno di questo motivo tre regioni ad α -elica le quali sono orientate nello spazio in maniera

caratteristica diversa rispetto i motivi HTH (helix-turn-helix) delle proteine batteriche o dagli omeodomini eucariotici [43, 44]. Recentemente, tramite studi effettuati con proteine chimeriche è stato ricavato un modello in base al quale vengono stabilite le caratteristiche che rendono l'HMG box in grado di legarsi a sequenze specifiche sul DNA [46]. Per quanto riguarda la distribuzione intracellulare delle HMG 1/2, non è ristretta al nucleo, ma può riguardare anche il citoplasma; si è visto inoltre che è variabile da tessuto a tessuto ed a seconda del ciclo cellulare. La quantità di queste proteine è inversamente correlata con quella della variante H1° dell'istone H1, e ciò suggerisce una correlazione con la proliferazione e il differenziamento [42]. Di recente, uno studio in *Drosophyla*, ha proposto un ruolo specifico nello sviluppo del sistema nervoso della proteina HMG D, risulta infatti che quest' omologo dell'HMG 1 ha una complessa localizzazione e distribuzione intracellulare, il che rifletterebbe un ruolo ben preciso piuttosto che generico [47].

Il secondo sottogruppo comprende le HMG 14 e 17, le cui sequenze permettono una netta distinzione fra le due proteine, essendo omologhe per circa il 30%. Ciascuna delle due mostra un notevole grado di conservazione da specie a specie e l'allineamento della loro sequenza porta ad identificare regioni complessivamente molto conservate fra le quali si distinguono i primi quattro aminoacidi PKRK assolutamente conservati in entrambe. Si ritiene che i geni che le codificano derivino da un singolo gene ancestrale, in ogni caso essi si trovano su cromosomi diversi (sul 21 quello per HMG 14 e sull'1 quello per HMG 17) entrambi in regioni i cui riarrangiamenti sono associati a patologie quali la sindrome di Down e certi tipi di neoplasie [42]. Il loro peso molecolare è sotto i 10 kD e la sequenza, particolarmente ricca in lisina, ha una struttura in cui si distingue una parte verso l'N-terminale in grado di legarsi al DNA ed un C-terminale arricchito in residui acidi. Queste proteine sono tipicamente nucleari ed è stato riportato che possono andare incontro a numerose altre modificazioni post-traduzionali quali acetilazione,

poli(ADP)-ribosilazione, fosforilazione, metilazione e glicosilazione. I dati sperimentali indicano una preferenza per l'interazione con regioni della cromatina in cui i geni sono attivi, in particolare a livello del nucleosoma [15, 42]. Sono stati proposti dei modelli che localizzano la zona di interazione di queste HMG vicino ai punti di entrata/uscita del DNA nel nucleosoma, fra gli istoni dell'ottamero e la doppia elica e ciò solleva la possibilità di un loro coinvolgimento nelle transizioni conformazionali che avvengono a livello dei nucleosomi nelle zone attive [48]. Recentemente è stata dimostrata *in vitro* una capacità di strutturazione ATP-dipendente da parte delle HMG 14/17 sulla cromatina, che porta ad un arrangiamento nucleosomale regolarmente spaziato corrispondente ad una ripetizione di 160-165 bp la quale può essere in un secondo tempo incrementata per mezzo dell'aggiunta di H1 fino a valori vicini a quelli che si trovano in natura nella cromatina [49]. Il coinvolgimento di una fonte di energia esogena nei rimodellamenti del materiale genetico è un aspetto molto interessante che sta emergendo attualmente in vari studi, i quali hanno messo in evidenza il fatto che le modificazioni nella cromatina possono essere facilitate se è disponibile una fonte di energia, quale l'ATP [28]. E' stato riportato per l'HMG 17 un ruolo di stabilizzazione del nucleosoma e di potenziamento della capacità trascrizionale quando la proteina viene assemblata nella cromatina nascente durante la replicazione del DNA nel sistema dei geni per l'rRNA 5S di *Xenopus* [50]. Esperimenti *in vivo* ed *in vitro* su fibroblasti di topo hanno messo in evidenza una divergenza funzionale fra le due HMG in seguito a stimolazione mitogenica, dal momento che solo l'HMG 14 viene fosforilata transientemente ma in modo sostenuto nella regione N-terminale, vicino al dominio che si lega col DNA del nucleosoma. Nei mononucleosomi preparati dalle cellule stimulate è stato inoltre possibile identificare una chinasi in grado di fosforilare l'HMG 14 *in vitro* ed è stato dimostrato che la modificazione non è conseguente all'attivazione trascrizionale dei geni che mediano la risposta immediata a fattori di stimolazione (immediate-early

response genes), non essendo diminuita in presenza di composti che inibiscono totalmente l'induzione di tali geni [51].

Il lavoro svolto nel nostro laboratorio si inserisce in questo contesto, essendo incentrato principalmente sullo studio delle proteine appartenenti alla terza famiglia di proteine HMG che sono state caratterizzate in alcuni sistemi biologici murini ed umani disponibili. Tale gruppo di proteine è quello individuato per ultimo in ordine di tempo ed indicato con la sigla HMGI. Queste proteine sono state individuate per la prima volta nelle cellule HeLa S3 proliferanti, dove sono state contraddistinte con le sigle HMG I, Y, M ed era stato visto che erano abbondantemente fosforilate [52]. La caratterizzazione biochimica di queste proteine aveva permesso di stabilire che il loro peso molecolare era compreso fra 10 e 12 kD e la composizione era più simile alle HMG14/17 che alle HMG 1/2, essendo ricca in aminoacidi basici e priva di residui aromatici. Nel nostro laboratorio era in corso uno studio sulle alterazioni che possono essere riscontrate in sistemi trasformati in senso neoplastico. Confrontando estratti da cellule da tiroide di ratto in coltura che potevano essere trasformate per mezzo di retrovirus oncogeni, erano state identificate tre proteine (denominate allora HMG C, D ed E) la cui espressione diventava evidente solo dopo l'infezione virale. Altri gruppi di ricerca riportavano, nello stesso periodo, l'isolamento di proteine cromosomiali non istoniche, con le stesse caratteristiche biochimiche da cellule indifferenziate proliferanti e da sistemi alterati in senso neoplastico, che risultavano espresse a livelli molto bassi o addirittura assenti in tessuti normali e differenziati nell'uomo, nei roditori, nelle scimmie [53-58]. Successivamente, dati ricavati dall'esame di vari sistemi cellulari e tessuti hanno indicato che il livello di espressione di questa classe di proteine è più probabilmente correlato col grado di differenziamento del sistema più che con la proliferazione di per se stessa [59-63]. Nel nostro laboratorio, l'attenzione si era focalizzata in particolare su una delle tre proteine dal momento che, mentre due, HMG D ed E potevano essere presenti, ma a livelli nettamente inferiori, nelle cellule anche senza che queste fossero trasformate,

la terza, HMG C appariva solo in concomitanza con la trasformazione in senso neoplastico [64]. Quando sono state ottenute le sequenze primarie complete di due di queste proteine è stato possibile caratterizzare in maniera inequivocabile i vari sistemi studiati. Fino a questo momento infatti l'identificazione dei componenti degli estratti cellulari veniva fatta sulla base di separazioni elettroforetiche e cromatografiche ma, essendo le proteine molto simili fra loro e soggette a modificazioni post-traduzionali, spesso la comparazione dei sistemi biologici analizzati risultava difficile e dubbia [52-58, 64-66]. Sulla base delle sequenze nucleotidiche murina ed umana clonate dal cDNA, è stato possibile stabilire che le proteine identificate come HMG I e HMG Y sono due isoforme che differiscono nella struttura primaria a livello di una regione di 11 aminoacidi presente soltanto nella prima delle due proteine e si tratta pertanto di prodotti derivanti dallo splicing alternativo dello stesso gene a cui ora si fa riferimento come HMG I(Y). È stato inoltre constatato un alto grado di conservazione fra proteine umane e murine ed è stata evidenziata una struttura modulare, con tre domini basici adatti all'interazione con il DNA e un'estremità C-terminale ricca in residui acidi [59, 62]. Era intanto già stato osservato che queste proteine erano in grado di legarsi al DNA, dimostrando una preferenza per le zone ricche in A/T ed anche nel corso di altri studi, erano stati riportati molti siti con questa caratteristica in grado di legarsi con le HMGI quali le regioni ripetute α -satellite in cellule di scimmia; alcune regioni non tradotte al 3' di geni per le linfocine e fattori di crescita; sequenze attivatrici ed elementi di regolazione a monte di diversi geni, in particolare quelli per l'RNA ribosomiale del ratto e del topo, elementi regolatori nei promotori delle catene leggere delle immunoglobuline [59, 66-69]. Studi sul tipo di interazione che intercorre col DNA hanno poi portato all'identificazione di un motivo caratteristico per questa classe di proteine, rappresentato dal peptide -TPKRPRGRPK- per il quale è stata ipotizzata una struttura planare simile a quella di alcuni composti in grado di intercalarsi fra le basi del DNA a livello del solco minore e tale struttura è stata denominata "A·T

hook". E' probabile, visto lo scarso grado di specificità dimostrato, che i domini in grado di legare il DNA delle HMGI riconoscano una caratteristica strutturale nel solco minore delle zone ricche in A/T piuttosto che una particolare sequenza [70].

Era stato visto che varie sono le modificazioni a cui queste proteine possono andare incontro, ma quella principalmente riscontrata era la fosforilazione, che risultava correlata alle fasi del ciclo cellulare e alla proliferazione [52, 53, 64, 65, 71]. Era stato dimostrato che la casein chinasi II (CK II) era in grado di fosforilare le HMGI *in vitro* su più residui di serina, alcuni dei quali erano stati trovati effettivamente fosforilati in estratti da placenta [72, 73]. Erano inoltre stati individuati *in vitro* siti di fosforilazione sulle HMG I ed Y umane in corrispondenza di residui di treonina risultati poi modificati in effetti anche *in vivo* ed era stato evidenziato che simili sequenze erano fosforilate sull'istone H1 dalla chinasi dell'H1 associata alla crescita (growth-associated H1 kinase) [74]. Ciò ha dato un contributo ulteriore all'idea di una possibile coordinazione dei ruoli fra queste proteine e l'istone H1 con tutte le sue varianti mediata dal cambiamento nei livelli di modificazione durante le varie fasi del ciclo cellulare.

Nonostante da tutto questo insieme di dati provenisse un' evidenza per un coinvolgimento nell'attivazione della cromatina da parte delle HMG, queste proteine, di per se stesse, non erano in grado di attivare la trascrizione e non c'erano ancora indicazioni precise sulla eventuale funzione nell'ambito del sistema biologico, sia per quanto riguarda le normali condizioni di funzionamento che in presenza di alterazioni patologiche [42, 75]. A tale scopo, è stato necessario continuare la caratterizzazione in altri sistemi, utilizzare nuove tecniche, determinare le eventuali interazioni fra i vari componenti del sistema, sia *in vitro* che *in vivo*. Metodi di biologia molecolare, di biologia cellulare, quando disponibili, sono stati perciò affiancati a studi di tipo biochimico e strutturale. Nel nostro laboratorio dunque era stata continuata la caratterizzazione di sistemi neoplastici, passando all'analisi di tessuti tumorali di topo aventi diversa origine. Fra i tumori esaminati,

uno era stato indotto tramite agenti chimici, uno era di origine spontanea (Lewis lung carcinoma, LLC), un altro derivava dall'iniezione in topi nudi di cellule trasformate *in vitro* con retrovirus e l'ultimo era tessuto metastatico conseguente a tumori primari tiroidei indotti da retrovirus. In tutti i casi, era stato possibile identificare tutti e tre i membri appartenenti al gruppo HMGI. Oltre ad una relazione fra il grado di espressione delle HMG 14/17 e quello delle HMGI, l'analisi dei tumori evidenziava la costante presenza della terza proteina HMGI-C (non ancora caratterizzata in termini di sequenza primaria) in fenotipi altamente maligni, indipendentemente dalla loro origine [76]. Il lavoro svolto successivamente costituisce l'oggetto di questa tesi, in cui si riportano analisi estese a diversi sistemi volte a caratterizzare alcuni componenti della cromatina. Oltre allo studio sulle varianti dell'istone H1 e su alcune sue peculiari forme intermedie già menzionati, sono stati perseguiti gli scopi di caratterizzare completamente la sequenza primaria della proteina HMGI-C tramite il clonaggio del cDNA murino, identificarla in sistemi di origine umana, studiarne l'espressione in relazione agli altri componenti proteici, accertare il grado di conservazione della proteina fra le due specie, verificare l'attendibilità del modello murino, produrre anticorpi policlonali in grado di riconoscere le proteine ad un livello di sensibilità maggiore rispetto le tecniche tradizionali ed eseguire un lavoro di "screening" sul maggior numero di linee e tessuti tumorali per valutare la possibilità di un eventuale uso diagnostico del livello di espressione proteico.

MATERIALI E METODI

TECNICHE COMUNEMENTE USATE

COLTURE CELLULARI

Le cellule sono state cresciute in ambiente a 37°C e 5% CO₂ nel terreno di coltura opportuno (solitamente DMEM o RPMI 1640) addizionato con il 10% (volume/volume, v/v) di siero fetale bovino totale inattivato, 100 u/ml di penicillina e 0,1 mg/ml streptomina.

ESTRAZIONE DELLE PROTEINE HMG

Le proteine HMG vengono estratte con acido perclorico al 5% dai campioni di tessuto o di cellule da coltura mediante omogeneizzazione meccanica. Il materiale proteico viene fatto precipitare mediante l'aggiunta di 6-9 volumi di acetone, viene raccolto tramite centrifugazione e lasciato asciugare; il tutto viene poi ridisciolto nel solvente opportuno oppure in acqua dove viene dializzato ed infine liofilizzato. La composizione di questo estratto totale è determinata tramite analisi elettroforetica in prima dimensione di acido acetico/urea e in seconda dimensione di SDS/Tris- tricina, su gel di poliacrilamide.

ANALISI ELETTROFORETICHE DI PROTEINE

Analisi elettroforetica in I dimensione in acido acetico/urea

Le analisi elettroforetiche in acido acetico/urea sono state condotte secondo quanto già precedentemente descritto [53].

Dimensioni del gel: 14 x 20 x 0,075 cm

-Gel di separazione (running): 15% (peso/volume, w/v) acrilamide ; 0,2% (w/v) BIS (N,N'-metilenbisacrilamide); 0,2% (w/v) persolfato di ammonio; 0,5% (v/v) TEMED (N,N,N',N'-tetrametilendiammina); solvente: soluzione 0,9 M acido acetico, 4 M urea.

-Gel di caricamento (stacking): 7,5% (w/v) acrilamide; 0,2% (w/v) BIS; 0,2% (w/v) persolfato d'ammonio ; 1% (v/v) TEMED ; solvente: soluzione 0,9 M acido acetico, 4 M urea

-Tampone per la corsa: soluzione 0,9 M acido acetico.

-Corsa: le proteine sono state separate applicando una differenza di potenziale di 200 v per 15 ore.

-Analisi elettroforetica in II dimensione in SDS/Tris- tricina

Le analisi elettroforetiche in SDS/Tris/tricina sono state condotte secondo quanto già precedentemente descritto [77].

Dimensioni del gel: 14 x 20 x 0,1 cm

-Gel di separazione (running): 15% (w/v) acrilamide ; 0,2% (w/v) BIS (N,N'-metilenbisacrilamide); 0,075% (w/v) persolfato di ammonio; 0,07% (v/v) TEMED (N,N,N',N'-tetrametilendiammina) solvente: soluzione in proporzione 1:5 (v/v) di 0,1% (w/v) SDS; 0,75 M Tris HCl pH=8,8 e 0,1% (w/v) SDS; 0,5 M tricina pH=8,1

-Gel di attacco: 15% (w/v) acrilamide; 0,2% (w/v) BIS; 0,2% (w/v) persolfato di ammonio; 0,2% (v/v) TEMED; solvente: soluzione 0,1% (w/v) SDS; 0,75 M Tris HCl pH=8,8

-Tampone per la corsa: 0,1% (w/v) SDS; 0,025 M tricina, portato a pH=8,1 con Tris base

-Condizionamento: singole bande o intere strisce recuperate dal gel di acido acetico/urea, sono state condizionate per circa 30' in una soluzione 1% (w/v) SDS; 0,75 M Tris HCl a pH=8,8 prima di essere sottoposte a elettroforesi di II dimensione.

-Corsa: viene effettuata a 150 v per circa 15 ore

COLORAZIONI

-Colorazione con blu Coomassie: il gel viene colorato e contemporaneamente fissato immergendolo in una soluzione allo 0,05% (w/v) di Blue-Coomassie R-250 (Serva) sciolto in metanolo, acido acetico, acqua nelle proporzioni 5:4:1 in volume. La decolorazione: viene eseguita con una soluzione di etanolo ed acqua in proporzione 1:3 in volume.

-Colorazione argentea: effettuata secondo il metodo riportato in [78]. Il gel viene sciacquato abbondantemente in acqua e fatto equilibrare in una soluzione filtrata allo 0,1% (w/v) di TPNS (tri-isopropil-naftalensulfonato). Il gel dopo esser nuovamente sciacquato in acqua, viene colorato con una soluzione allo 0,005% (w/v) di blu di metilene; poi viene decolorato con acqua. Al gel viene aggiunta una soluzione di nitrato di argento ammoniacale preparata mescolando una soluzione 0,03 M NaOH, 0,75% (v/v) ammoniaca con una allo 0,75% (w/v) nitrato d'argento in acqua. Lo sviluppo si ottiene aggiungendo una soluzione contenente 0,05% (w/v) acido citrico, 0,075% (w/v) formaldeide. La reazione viene bloccata con una soluzione di metanolo, acido acetico, acqua in rapporto 5:4:1. Si decolora con etanolo, acido acetico, acqua in rapporto 5:1,6:13,4 resa al 5% in glicerolo.

SEPARAZIONE DELLE PROTEINE HMG

La separazione delle proteine HMG, è stata ottenuta mediante l'uso della cromatografia HPLC (High Performance Liquid Chromatography) a fase inversa. L'apparecchiatura utilizzata è un cromatografo Waters a doppia pompa (modello 510) fornito di rilevatore spettrofotometrico con lampada allo zinco; la lettura viene effettuata alla lunghezza d'onda di 229 nm. E' stata usata una colonna Bio-Rad RP-304 C4 (250 x 4,6 mm). Le proteine vengono eluite usando un gradiente ottenuto miscelando acqua resa allo 0,1% (v/v) di TFA (acido trifluoroacetico) (solvente A) e 95% (v/v) acetonitrile; 5% (v/v) acqua; 0,1% (v/v) TFA (solvente B). Le frazioni eluite

vengono raccolte, dializzate e liofilizzate oppure in alternativa evaporate sotto vuoto e poi ridissolte per effettuare le analisi successive.

QUANTIFICAZIONE DELL'ESPRESSIONE DELLE PROTEINE

La valutazione delle quantità delle proteine è stata calcolata in base alle aree dei picchi cromatografici, relativamente all'area dell'istone H1 inteso come totale delle subfrazioni. Nel caso di picchi in cui coeluiscono più componenti, la proporzione di ciascuno è stata determinata tramite lettura densitometrica dei gel di acrilamide di prima dimensione colorati con blu Coomassie, utilizzando un apparato Bio-Rad Model 1650 Transmittance/Reflectance Scanning Densitometer. Le analisi sono state ripetute più volte in modo da contenere l'errore stimato entro il 10%.

ESPRESSIONE DELLA PROTEINA HMGI-C RICOMBINANTE

La sequenza codificante per HMGI-C murina è stata inserita, rispettando il corretto modulo di lettura, in un vettore di espressione procariotico pAR3038, disponibile commercialmente col quale sono state trasformate cellule batteriche del ceppo BL21(DE 3) di *E.coli*, selezionate per la resistenza all'ampicillina. L'induzione è ottenuta mediante aggiunta di IPTG (isopropil β -D-tiogalattopiranoside) 0,1 mM finale. Dopo crescita durante la notte, i batteri vengono raccolti per centrifugazione. L'estrazione e la successiva purificazione vengono effettuate secondo le metodiche descritte.

PRODUZIONE DI ANTICORPI ANTI HMGI-C RICOMBINANTE

Sono stati inoculati in coniglio circa 300 μ g di HMGI-C ricombinante sciolti in circa un millilitro di acqua ed emulsionati in rapporto 1:1 con adiuvante completo di Freund. A distanza di circa 3 settimane ciascuno sono stati fatti vari richiami con una quantità stimata di antigene compresa tra i 200 e i 300 μ g sciolta ed emulsionata con adiuvante incompleto di Freund. Dopo 11 giorni da ciascuno dei richiami è stato fatto un salasso

da cui sono stati ricavati per centrifugazione circa 20 ml di siero per volta che è stato successivamente purificato per affinità. Il siero è stato conservato a -20°C.

CROMATOGRAFIA PER AFFINITA'

-Derivatizzazione: è stata utilizzata la resina Affi-prep 10 Affinity Chromatography Matrix Bio-Rad; condizionata in tampone acetato 10 mM a pH 4,5 rimpiazzato per tre volte dopo centrifugazione a 3000 r.p.m. per 3 minuti. Circa 200 µg di HMGI-C ricombinante sciolta in tampone 100 mM MOPS (acido 3-(N-morfolin)propan sulfonico) pH=7, è stata messa ad incubare per una notte a 4°C in agitazione. La resina derivatizzata è stata raccolta mediante centrifugazione per 10 minuti a 3000 r.p.m. E' stata aggiunta etanolamina/HCl 1 M, pH=8 per bloccare i gruppi attivi eventualmente rimasti. La resina è stata trasferita nella colonna che viene lavata con NaCl 1,5 M e successivamente con più volumi di PBS (phosphate buffered saline).

-Eluizione: la colonna viene equilibrata con PBS. Si utilizza antisiero diluito con un volume scarso di PBS e viene reso 5% (v/v) Triton X-100. Si lascia incubare con la matrice per 1-2 ore in agitazione a temperatura ambiente e poi si recupera il siero. Si lava con circa 30 volumi di PBS. Per eluire gli anticorpi si usa glicina/HCl 0,2 M, pH=2,8. L'eluato viene raccolto e subito reso 0,2 M finale con Tris/HCl 2 M, pH=8. La colonna viene poi lavata abbondantemente con PBS e conservata a 4°C, eventualmente in sodio azide.

WESTERN BLOT

Questa tecnica è stata utilizzata per trasferire le proteine da un tracciato elettroforetico su gel di poliacrilamide in acido acetico/urea, 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 Whatman 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 Bio-Rad, 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.

IMMUNORICONOSCIMENTO

Le membrane dopo trasferimento, vengono saturate in 0,05 M Tris/HCl pH=7,5, 0,5 M NaCl, 5% (w/v) latte scremato in polvere, 0,05% (v/v) Tween20, 0,2% (w/v) BSA (albumina da siero bovino) e poi incubate con gli anticorpi purificati per affinità (diluiti circa 1:30) per 1 ora a temperatura ambiente. Vengono fatti tre lavaggi con la soluzione di saturazione e poi si mette a incubare con un antisiero anti-IgG di coniglio coniugate con la perossidasi (Pierce) diluito 1:1000 per un'altra ora a temperatura ambiente. Altri tre lavaggi nelle stesse condizioni precedono lo sviluppo del colore, ottenuto, dopo un ulteriore lavaggio in 0,2 M Tris/HCl, pH=7,5, 0,5 M NaCl, trattando la membrana con 0,3% (w/v) 4-cloro-1-naftolo sciolto in metanolo freddo e mescolato 1:5 con 0,02 M Tris/HCl, pH=7,5, 0,5 M NaCl addizionato di 30% (v/v) H₂O₂. Lo sviluppo del colore viene bloccato con PBS. In alternativa è stato usato un substrato specifico disponibile commercialmente (TMB Promega).

ANALISI DI SPETTROMETRIA DI MASSA

Per la determinazioni dei pesi molecolari si è fatto ricorso alla tecnica "ion-spray". Le analisi sono state eseguite utilizzando un triplo quadrupolo Perkin Elmer Sciex API III LC/MS/MS. Da quando disponibile, è stato utilizzato lo spettrometro di massa API I Perkin Elmer Sciex presso il nostro dipartimento. In entrambi i casi, per la calibrazione dello strumento è stata utilizzata una soluzione standard di PPG (polipropilen glicole). Per la calibrazione in base agli ioni multicaricati, è stata 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 e ridisciolti in acqua/acetonitrile 80:20, 0,1% (v/v) acido formico per essere analizzati.

ESTRAZIONE DI ACIDI NUCLEICI

-Estrazione di DNA: al campione in soluzione viene aggiunto un ugual volume di fenolo equilibrato in 50 mM Tris/HCl pH=8. Le due fasi vengono miscelate e poi separate mediante centrifugazione. Si recupera la fase acquosa contenente il DNA e si procede con una successiva estrazione aggiungendo 0,5 volumi di fenolo e 0,5 volumi di cloroformio. La fase acquosa viene poi estratta con un volume di cloroformio. Il DNA viene precipitato a freddo rendendo la soluzione 0,3 M NaCl ed aggiungendo due volumi di etanolo al 95%. Si recupera per centrifugazione per 15' a 15000 g. Il precipitato viene lavato con etanolo al 70%. Il DNA viene risospeso in un volume appropriato di acqua.

-Estrazione di RNA: le cellule o tessuti vengono trattati col tampone di lisi 4 M guanidinio tiocianato, 25 mM sodio citrato, 0,1 M β -mercaptoetanololo, 0,5% (w/v) N-laurilsarcosina. L'omogenato viene reso 0,15 M tampone acetato pH=4 ed estratto con fenolo equilibrato in acqua e cloroformio. Si separa quindi la fase acquosa da quella organica mediante centrifugazione a 10.000 g per 15'. L'RNA dalla fase acquosa viene fatto precipitare a freddo aggiungendo un ugual volume di isopropanolo. Il precipitato viene raccolto mediante centrifugazione a 10.000 g per 15', risospeso in tampone di lisi e riprecipitato a freddo con etanolo. Si recupera per centrifugazione e si lava con 70% etanolo, si risospende in acqua [79]. L'RNA poliadenilato è stato preparato a partire dall' RNA totale usando colonnine di oligo(dT)cellulosa preconfezionate disponibili commercialmente (Pharmacia).

ELETTROFORESI SU GEL DI AGAROSIO

-Analisi di DNA: dimensioni del gel 10 x 7 x 0,5 cm

-Gel di separazione (running): 1% (w/v) agarosio, solvente: 1 x TAE (0,4 M Tris base, 0,2 M acido acetico, 10 mM EDTA (acido etilendiaminotetracetico) pH=8); 0,5 µg/ml etidio bromuro.

tampone per la corsa: 1 x TAE

-Analisi di RNA: dimensioni del gel 10 x 7 x 1 cm

-Gel di separazione (running): 1% (w/v) agarosio, solvente: 1 x tampone MOPS (20 mM MOPS pH=7, 1 mM EDTA pH=8, 5 mM sodio acetato), 6% (w/v) formaldeide

tampone per la corsa: 1 x MOPS pH=7, 3,7% (w/v) formaldeide.

La corsa in entrambi i casi viene effettuata a 70 v costanti.

NORTHERN BLOT

Il gel viene incubato con 50 mM NaOH per 15' in agitazione e quindi viene neutralizzato in tampone 1 M Tris/ HCl pH 7,5 per 15'. Il gel viene posto su alcuni fogli di carta 3MM Whatman. La soluzione di trasferimento è 25 mM tampone fosfato pH=6,5. Sul gel viene posta la membrana (Dupont) e sopra vengono aggiunti alcuni fogli di carta 3MM Whatman ed un blocchetto di carta assorbente dello spessore di circa 10 cm che viene mantenuto sotto una pressione costante in modo da ottenere il trasferimento per capillarità. Il trasferimento richiede circa 18 ore. Il filtro viene recuperato, lavato in tampone fosfato quindi asciugato ad una temperatura di 80°C per 2 ore. Infine si espone il filtro a radiazioni ultraviolette per 1'.

SOUTHERN BLOT

Il gel e la membrana su cui deve venir trasferito vengono incubati in 0,5 M NaOH, 1,5 M NaCl per 10'. La soluzione di trasferimento è la stessa. Il gel viene posto su alcuni fogli di carta 3MM Whatman. Sul gel viene posta la membrana (Dupont) e sopra vengono aggiunti alcuni fogli di carta da filtro ed un blocchetto di carta assorbente dello spessore di circa 10 cm che viene mantenuto sotto una pressione costante in

modo da ottenere il trasferimento per capillarità. In due ore circa il trasferimento è completo. Il filtro viene recuperato, lavato in 2 x SSC (0,3 M NaCl, 0,04 M sodio citrato, pH=7) quindi lasciato asciugare. Viene infine esposto per 1' a radiazioni ultraviolette.

REAZIONI DI SEQUENZA

Viene usato il metodo dei dideossinucleotidi [80]. La miscela fra DNA (circa 1 µg, in vettore plasmidico o fagico) e 50 pmoli di oligonucleotide d'inizio (primer) viene denaturata a 68°C rendendola 0,1 M NaOH. La soluzione si neutralizza rendendo 0,1 M HCl e la rinaturazione viene ottenuta rendendo la soluzione 40 mM Tris/HCl pH=7,4, 15 mM MgCl₂, 5 mM NaCl, 20 mM DTT (ditiotreitolo) a temperatura ambiente. Viene aggiunta la miscela di elongazione (7,5 mM dNTP, di cui tre deossinucleotidi freddi ed uno radioattivo) e 3U di DNA polimerasi T7 DNA (Sequenase, Pharmacia). Si lascia reagire per 5' a temperatura ambiente. La miscela di reazione viene suddivisa in quattro aliquote ed a ciascuna viene aggiunta la rispettiva soluzione di terminazione (80 mM dNTP, 8 mM del rispettivo ddNTP). Si lascia reagire per 10' a 37°C, e poi si aggiunge la soluzione di stop (95% formamide, 20 mM EDTA, 0,05% (w/v) blu di bromofenolo, 0,05% (w/v) xilene cianolo) in ciascuna aliquota. Prima di caricare in elettroforesi si denatura a 95°C per 3'.

ELETTROFORESI PER SEQUENZE DI DNA

Dimensioni del gel: 20 x 40 x 0,02 cm.

Ciascuno dei due vetri viene lavato e trattato in modo diverso. Il vetro sul quale si vuol fare aderire il gel viene trattato con soluzione legante di silano, mentre l'altro viene trattato con una soluzione di dimetilclorosilano ed etanolo, per facilitare il distacco del vetro dal gel.

-Gel di separazione (running): 8% (w/v) acrilamide, 50% (w/v) urea, 0,1% (w/v) BIS; 0,05% (w/v) persolfato di ammonio; 0,1% (v/v) TEMED, solvente: 1 x tampone TBE (0,09 M Tris, 45 mM acido borico, 2 mM EDTA pH=8).

Tampone per la corsa: 1 x TBE

La corsa viene condotta ad un voltaggio costante di 1800 v. Alla fine della corsa, il gel viene prima fissato per 10' in una soluzione di 10% (v/v) acido acetico, poi sciacquato in acqua, essiccato ed esposto a lastra autoradiografica in una apposita cassetta per circa 12 ore.

MARCATURA DI SONDE DI DNA

-Marcatura di frammenti di restrizione di DNA: il frammento purificato di DNA da marcare viene miscelato con 20 pmoli di esanucleotidi a sequenza casuale (random primers), denaturato mediante bollitura e poi lasciato rinaturare in ghiaccio. Viene aggiunto tampone di marcatura, rendendo la miscela 50 mM Tris/HCl pH=7, 10 mM MgSO₄, 100 mM DTT, 60 mM dATP, dGTP, dTTP, ³²P-dCTP (30 µCi) ed aggiungendo 3U di Klenow fragment; si lascia ad incubare per 3 ore a 37°C.

-Marcatura di oligonucleotidi sintetici: circa 500 ng di oligonucleotide vengono marcati per aggiunta enzimatica (tramite chinasi, T4 PNK Promega) di fosfato terminale radioattivo, usando 50 µCi di γ -³²P-ATP in tampone 40 mM Tris/HCl pH=7,5, 10 mM MgCl₂, 5 mM DTT e lasciando procedere la reazione per 1 ora a 37°C.

Le sonde così preparate vengono separate dai nucleotidi non incorporati mediante cromatografia per gel filtrazione utilizzando la resina G-50, equilibrata nel tampone TNES (10 mM Tris/HCl pH=7,4, 100 mM NaCl, 1 mM EDTA, 0,1% (w/v) SDS) ed eluite con lo stesso. In alternativa sono state usate "spun columns" S-300 HR (Pharmacia) disponibili commercialmente. L'attività specifica delle sonde ottenute è in media di 5×10^8 cpm/µg di DNA.

TRASFERIMENTO DI DNA FAGICO O BATTERICO SU FILTRO DI NYLON

Una membrana di nylon (Dupont) viene sovrapposta alla coltura di fagi o batteri su terreno agar solido e lasciata per 1' a contatto con essa. Poi viene trasferita su una soluzione denaturante 0,5 M NaOH per 3'. La membrana viene poi fatta asciugare per altri 3' e quindi viene ripetuto il trattamento con NaOH una seconda volta. Il filtro viene neutralizzato con due lavaggi successivi, ciascuno di 3', con una soluzione 1 M Tris/HCl pH 7.5. Una volta asciugato, il filtro viene esposto per 1' alla radiazione ultravioletta per legare irreversibilmente il DNA alla matrice della membrana.

IBRIDAZIONE DI ACIDI NUCLEICI SU FILTRO DI NYLON

Il filtro viene incubato a 68°C con la miscela di preibridazione composta da 1% (w/v) SDS, 1 M NaCl e DNA da spermi di salmone alla concentrazione finale di 0,1 mg/ml. Dopo circa 30', alla miscela di preibridazione viene aggiunta la sonda previamente denaturata per 5' a 100°C, ad una concentrazione finale di 5×10^5 cpm/ml, inoltre viene aggiunta la soluzione di Denhardt (0,02% (w/v) Ficoll 400, 0,02% (w/v) polivinilpirrolidone, 0,02% (w/v) BSA). Lavaggi: due volte con 2 x SSC a temperatura ambiente per 2', una volta in 2 x SSC, 1% (w/v) SDS a 65°C per 15', una volta in 0,2 x SSC, 0,1% (w/v) SDS a 65°C per 15'. La rimozione della sonda legata in modo aspecifico viene seguita con il monitor. I filtri, coperti da una pellicola di polietilene, vengono esposti a freddo ad una lastra autoradiografica in una cassetta con schermi intensificatori.

SUBCLONAGGIO IN VETTORE PLASMIDICO

I cloni in vettore fagico risultati autentici positivi, sono stati subclonati in vettore plasmidico per essere agevolmente amplificati e sequenziati. Mediante enzimi di restrizione opportuni, viene liberato l'insero corrispondente al cDNA che viene isolato dal resto tramite elettroforesi su agarosio. La banda elettroforetica costituita dal frammento viene prelevata dal gel, il quale viene sciolto a temperatura opportuna ed il

DNA viene estratto dalla soluzione mediante estrazione fenolo/cloroformio oppure tramite l'utilizzo di matrici disponibili commercialmente seguendo le istruzioni del rivenditore (Prep-A-Gene, Bio-Rad). Le estremità del frammento così recuperato vengono eventualmente trattate enzimaticamente per essere rese compatibili con quelle del vettore plasmidico Bluescript KS+, le quali invece vengono defosfatate enzimaticamente (fosfatasi alcalina bovina) all'estremità 5' per evitare la richiusura su se stesso del plasmide. Utilizzando l'inserto in una proporzione doppia rispetto al vettore, viene fatta la reazione di unione fra vettore e frammento (ligation) mediante l'enzima ligasi del fago T4 (Promega) in 30 mM Tris/HCl pH=7,8, 10 mM MgCl₂, 10 mM DTT, 0,5 mM ATP per 12 ore a 16°C.

STRATEGIE SPECIFICHE

PREPARAZIONE DELLA SONDA PER IL RICONOSCIMENTO DEL cDNA DI HMGI-C E SCREENING

E' stato progettato un oligonucleotide sulla base della sequenza aminoacidica del peptide dell' HMGI-C murina KXPQQVVQKKPAQE. Tenendo conto della degenerazione del codice e della mancata identificazione del secondo residuo, la sequenza della sonda lunga 41 bp era la seguente:

5'-AAGAAGCCCCA(A/G)CA(A/G)GTGGTGCA(A/G)AA(A/G)AA(A/G)CCTGC(T/C)CA(A/G)GA-3'

Tale sonda è stata usata per vagliare una libreria preparata a partire dall'RNA poliadenilato estratto da tessuto da Lewis lung carcinoma, usando oligonucleotidi (dT) come primers e inserendo i cDNA ottenuti in modo orientato, utilizzando due siti di restrizione diversi alle due estremità (EcoRI-HindIII) rispetto al vettore fagico λT7-T3. Sono state cresciute su piatto circa 200.000 placche fagiche in totale, in particolare, da una prima piastratura, ne sono state prelevate una decina che avevano dato un segnale positivo dopo ibridazione con la sonda e queste sono state eluite e ripiastrate una

seconda volta. Il secondo vaglio ha permesso di riconoscere solo un clone positivo che è stato isolato e subclonato in vettore plasmidico [81].

SEQUENZIAMENTO DEL CLONE CODIFICANTE PER HMGI-C

Il frammento corrispondente al cDNA derivante dal clone positivo isolato (1.7 Kb) è stato subclonato in plasmide Bluescript KS+, per essere amplificato e sequenziato. Tale inserto è risultato troppo lungo per essere sequenziato interamente utilizzando la T7 DNA polimerasi ed usando i primer di Bluescript. Pertanto si è deciso di seguire la strategia suggerita dalla casa fornitrice del vettore (Stratagene). Si tratta del sistema di "introduzione di delezioni unidirezionali", che si avvale della Esonucleasi III e della Mung Bean Nuclease le quali, usate in successione e dopo opportuni tagli con enzimi di restrizione, permettono di ottenere tutta una gamma di frammenti (derivanti da quello iniziale) ognuno recante una delezione della parte iniziale, più o meno estesa tutti inseriti nel vettore. In questo modo, dopo aver subclonato, amplificato e sequenziato tutti i frammenti ottenuti, per sovrapposizione, è stata ottenuta la sequenza completa di tutto il clone iniziale, in direzione 5'.

Questo stesso procedimento è stato ripetuto sullo stesso clone per ottenere la sequenza completa anche in direzione 3'. In questo modo, ogni base è stata letta almeno tre volte, cioè in tre sequenze diverse.

PREPARAZIONE DELLA SONDA PER IL RICONOSCIMENTO DEL cDNA DI EM6

Sulla base della sequenza aminoacidica della presunta parte C-terminale del dominio globulare di EM6 [82] sono stati preparati due oligonucleotidi degenerati, al fine di amplificare mediante PCR (polymerase chain reaction) la corrispondente regione codificante usando come stampo l'RNA messaggero. In particolare, la regione aminoacidica si estende per circa una cinquantina di residui e per progettare gli

oligonucleotidi primers sono state scelte due regioni di sette aminoacidi ciascuna spaziate da 31 residui. Sono stati così utilizzati i due segmenti di 24 bp:

EM6-5' 5'-GGATCCNGCNGGNATGAA(A/G)AA(C/T)CA-3' (corrispondente al peptide AAGMKNH)

EM6-3' 5'-AAGCTTNGGNGGNGC(C/T)TT(C/T)TTNAC-3' (corrispondente al peptide VKKAPPK)

per amplificare una regione stimata pari a 126 bp. E' stato pertanto sintetizzato un primo filamento a partire da 1 µg di RNA totale estratto da gonadi immature di *Ensis minor* quale stampo, 30 pmoli di EM6-3', 50 mM Tris/HCl pH=8,3, 75 mM KCl, 3 mM MgCl₂, 1 mM dNTP utilizzando 200U di trascrittasi inversa (M-MLV reverse transcriptase, BRL) e la reazione è stata lasciata procedere per 1 ora a 42°C ed è stata poi bloccata per congelamento. Sono state successivamente aggiunte 50 pmoli di EM6-5', 20 pmoli di EM6-3' e 2,5U di Taq polimerasi (Promega), per portare avanti un'amplificazione di 30 cicli di 1' a 94°C, 1,5' a 50°C e 1' a 72°C seguiti da un'estensione finale di 3' a 72°C. I prodotti ottenuti sono stati analizzati su gel di agarosio al 2% e una banda della lunghezza stimata è stata recuperata dal gel, purificata e subclonata nel vettore plasmidico. Uno dei subcloni così ottenuti, KA-24, è stato sequenziato ed è risultato contenere un inserto di 133 bp codificante per la parte C-terminale del dominio globulare di EM6, in accordo con i dati di sequenza aminoacidica.

PREPARAZIONE DELLA LIBRERIA cDNA E SCREENING

Usando una miscela di oligonucleotidi a sequenza casuale (random) e oligonucleotidi (dT) come primers, è stata preparata una libreria cDNA utilizzando un kit per la sintesi di cDNA in commercio (Pharmacia). Sono stati usati come stampo RNA messaggeri poliadenilati e non, ottenuti frazionando l'RNA totale tramite colonne disponibili commercialmente (Pharmacia). I cDNA ricavati sono stati selezionati in base alla lunghezza, purificati ed inseriti tramite l'uso di adattatori in corrispondenza del sito di

restrizione EcoRI nel vettore fagico λ Max1. Circa 300.000 placche sono state vagliate usando come sonda l'insero di 133 bp di KA-24, sono stati così isolati 15 cloni positivi, ciascuno dei quali è stato isolato e ripiastrato per la seconda volta. Uno degli inserti più lunghi trovati di circa 1,5 Kb è stato subclonato in vettore plasmidico Bluescript KS+ ed ha dato origine al subclone 3.1a.

SEQUENZIAMENTO DEL CLONE CODIFICANTE PER EM6

Per ottenere la sequenza completa del subclone 3.1a isolato, si è pensato di ricorrere al subclonaggio di frammenti di minori dimensioni ottenuti mediante l'uso di enzimi di restrizione. In base alle caratteristiche note della proteina, si poteva infatti ipotizzare una sequenza estremamente ripetitiva del relativo cDNA, pertanto difficilmente sovrapponibile. Un'analisi di restrizione ha messo in evidenza la capacità dell'enzima XbaI di spezzare il clone in parti di dimensioni opportune per essere interamente sequenziate. Sono stati ottenuti così i subcloni IXa, IIXv, IIIXn, IIIXt. Per determinare l'esatto ordine delle sequenze, è stato necessario generare ulteriori frammenti che confermassero la zona di giunzione fra i vari cloni. Sono stati a questo scopo usati gli enzimi di restrizione TaqI, AluI, NlaIV e tutti i frammenti ottenuti sono stati subclonati e sequenziati. L'intera sequenza è stata così dedotta, ma l'estremità 5' del clone era incompleta, mancando alcune basi codificanti e il codone di inizio trascrizione.

Tutti gli altri cloni risultati positivi al vaglio sono stati riesaminati, ma si sono rivelati tutti uguali o più corti rispetto a quello interamente sequenziato.

La libreria infine è stata vagliata con una sonda oligonucleotidica EM6-3'(99) 5'-TTGCTCGCCTTCGGCGAAGC-3' la cui sequenza corrispondeva ad una zona prossima al 5', nel tentativo di trovare cloni che si estendessero verso quell'estremità. Anche in questo caso, i cloni individuati erano più corti o uguali.

PCR-RACE

E' stato fatto ricorso a questo metodo (polymerase chain reaction-rapid amplification of cDNA ends) per ottenere la sequenza codificante completa comprendente il codone di inizio. Su uno stampo di RNA totale estratto dalle gonadi maschili immature del mollusco, trattato con DNAsi I e purificato tramite un'estrazione fenolo/cloroformio, sono stati usati esanucleotidi a sequenza casuale quali primers per effettuare una prima reazione di sintesi con la trascrittasi inversa (200U) per 45' a 42°C e per ulteriori 15' a 45°C in 50 mM Tris/HCl pH=8,3, 75 mM KCl, 3 mM MgCl₂, 1 mM dNTP. I prodotti così ottenuti sono stati purificati dai nucleotidi non incorporati rimasti in eccesso con un'estrazione fenolo/cloroformio e con precipitazioni successive con 0,5 M NaClO₄/0,5 volumi di isopropanolo e lavati ripetutamente con 70% (v/v) etanolo. E' stata aggiunta enzimaticamente una coda artificiale poli(G) all'estremità 5' mediante l'uso di desossiribosiltransferasi terminale (TdT). La reazione è stata bloccata dopo un'ora e i prodotti della trascrittasi con la coda poli(G) sintetica sono stati nuovamente estratti, precipitati e lavati con etanolo. Una volta risospesi in acqua, è stata allestita una prima amplificazione usando 50 pmoli di un primer complementare poli(C) e 50 pmoli dell'oligonucleotide (complementare alle basi 80-99 del clone 3.1a) EM6-3'(99), 10% DMSO (dimetilsulfossido), 10 mM Tris/HCl pH=8,3, 50 mM KCl, 1,5 mM MgCl₂, 0,1% Triton X-100, 1 mM dNTP, utilizzando 2,5U di Taq polimerasi (Promega). Sono stati fatti 40 cicli di 1' a 94°C, 1' a 55°C, 1' e 30" a 75°C, con un'estensione finale di 5' a 75°C. Successivamente è stato prelevato un decimo del volume di questa reazione ed usando il primer poli (C) e un altro oligonucleotide (complementare alle basi 56-75 del clone 3.1a) EM6-3'(75) più prossimo al 5' rispetto a quello precedentemente usato, è stata ripetuta l'amplificazione una seconda volta nelle stesse condizioni. Il prodotto ottenuto, della lunghezza stimata di 160 bp è stato purificato mediante passaggio su "spun column", subclonato in Bluescript e sequenziato. E' stato così possibile ottenere la sequenza completa del messaggero.

RISULTATI E DISCUSSIONE

Le analisi effettuate nel nostro laboratorio tramite elettroforesi di acido acetico/urea e cromatografia a fase inversa avevano permesso di evidenziare la presenza costante dell'istone H1, le HMG 1/2, le HMG 14/17 e le proteine del gruppo HMGI negli estratti con acido perclorico al 5% da tessuti tumorali murini [76]. In particolare, la trasformazione neoplastica risultava sempre accompagnata dalle HMGI-D e HMGI-E, identificabili con le proteine HMG I e Y descritte da altri autori [52], e da HMGI-C, la cui comparsa era stata notata in cellule trasformate o indifferenziate [53, 54, 64]. La presenza di tutti e tre i membri della famiglia HMGI era stata pertanto ritenuta peculiare del fenotipo trasformato murino [76]. Si è pensato pertanto di estendere l'analisi a tessuti umani, per vedere se anche in questo caso era possibile definire una situazione indicativa di uno stato alterato in senso neoplastico, considerando i dati preventivamente riportati da altri autori per le proteine HMG I e Y, riscontrate in cellule umane proliferanti [52] e le cui sequenze risultavano altamente conservate fra campioni umani e murini [59, 62]. In collaborazione con altri laboratori, è stata cresciuta in grandi quantità una linea cellulare di epatoma umano denominata PLC/PRF/5 e con i metodi messi a punto nel nostro laboratorio è stato possibile individuare con certezza la presenza di tutte e tre le proteine nel campione umano (fig. 5 dal riferimento bibliografico (rif.) [77]). Il campione murino di Lewis lung carcinoma (LLC) è stato da noi scelto come riferimento poiché disponibile in quantità adeguate ed inoltre era stato dimostrato che le proteine HMGI erano tutte e tre presenti in quantità maggiore rispetto agli altri campioni analizzati. Abbiamo quindi effettuato la separazione cromatografica del campione di epatoma ed analizzando in elettroforesi bidimensionale in acido acetico/urea ed in SDS, sistema che permette una buona risoluzione delle proteine HMG, le frazioni sono state riconosciute ed identificate sulla base del confronto con il campione di riferimento. Parallelamente, un estratto dalle stesse cellule cresciute in terreno nel quale il fosfato era stato sostituito con il corrispettivo composto contenente l'isotopo

radioattivo del fosforo, ^{32}P , era stato analizzato con la stessa tecnica dell'elettroforesi bidimensionale ed in autoradiografia è stato possibile distinguere le proteine fosforilate, quindi in particolare le HMGI. Il trattamento dello stesso campione con fosfatasi ha confermato la fosforilazione specifica delle proteine individuate. Abbiamo così riscontrato anche nel campione umano la presenza della proteina HMGI-C, la quale dimostrava di differire in un certo qual modo dalle altre due HMGI anche per la presenza di una piccola frazione resistente al trattamento con fosfatasi prolungato che poteva indicare un altro tipo di modificazione, probabilmente poli(ADP)ribosilazione (fig. 4 dal rif. [77]). Nel corso della stessa serie di studi, erano stati esaminati anche altri campioni umani disponibili quali le cellule linfocitarie del sangue periferico e alcune linee leucemiche. Mentre HMG I e Y erano presenti, anche se talora in quantità appena rilevabili, HMGI-C è risultata assente, mettendo in evidenza una certa atipicità per quanto riguarda la sua espressione in particolare in questo tipo di tessuti [77]. Al fine di migliorare le nostre conoscenze sul comportamento di questa proteina, era più che mai necessario venire a conoscenza della sua sequenza aminoacidica completa per poterla confrontare con quelle già note delle altre HMGI. Riscontrate infatti delle differenze nel comportamento di queste proteine, poteva senz'altro essere di aiuto un esame comparativo della loro struttura primaria. Abbiamo pensato di impiegare a questo scopo tecniche di biologia molecolare per ottenere la sequenza della proteina clonando il cDNA. Usando come materiale di partenza il tessuto derivante da LLC, è stata purificata nel nostro laboratorio una grande quantità della proteina HMGI-C usando la cromatografia a fase inversa, in modo da ottenere abbastanza materiale per ricavare tramite degradazione di Edman dei dati parziali di sequenza da cui partire. Dal momento che l'estremità N-terminale è risultata bloccata [64], la proteina è stata digerita, sulla base delle indicazioni provenienti dalle sequenze già note di HMG I(Y), con l' endoproteasi Glu-C (proteasi V8 da stafilococco) e due peptidi così ottenuti sono stati sequenziati. Uno è risultato molto simile e ben

allineabile con un tratto di sequenza di HMG I(Y), mentre l'altro presentava una zona che non era identificabile nelle sequenze fino a quel momento note (fig.2 dal rif. [81]). Sulla base di tale sequenza è stato sintetizzato un oligonucleotide degenerato lungo una quarantina di basi da usare come sonda per vagliare una libreria a cDNA in vettore fagico λ T7-T3/E-H costruita a partire dall'RNA poliadenilato estratto da LLC. 200.000 placche sono state da noi vagliate, i cloni risultati positivi sono stati ripiastrati una seconda volta e vagliati nuovamente allo stesso modo [81]. Questa volta un solo clone è risultato positivo, la cui sequenza si è dimostrata potenzialmente codificante per il peptide che era stato usato per progettare la sonda. Questo clone HMGI-C11, risultato di circa 1,7 Kb, è stato quindi subclonato in vettore plasmidico ed interamente sequenziato (vedere "Materiali e metodi"). E' stata così identificata la parte codificante per la proteina completa, delimitata dai canonici codoni di inizio traduzione e di stop. A questo punto, il cDNA è stato usato per creare una sonda radioattiva da ibridare con dei Northern blot, in cui erano stati trasferiti RNA estratti da LLC e da colture cellulari di fibroblasti di topo NIH 3T3 trasformati con alcuni oncogeni (*ras*, *fos*, *myc*, *src*), ma anche da tessuti normali di topo (cuore, polmoni, reni, cervello, fegato, etc.). I risultati di questi esperimenti sono stati due: la sonda non si ibridava mai con i campioni da tessuti normali ma solo con quelli derivanti da sistemi trasformati ed in questo caso il riconoscimento da parte della sonda avveniva in corrispondenza dell'RNA 28S, indicando che il messaggero presente in quelle determinate cellule era in effetti molto più lungo (circa 4 Kb) rispetto al cDNA riscontrato nel clone isolato dalla libreria (fig. 4 dal rif. [81]). Per questo motivo, abbiamo effettuato altri vagli sulla stessa libreria originaria, per individuare altri eventuali positivi a lunghezza maggiore. Sono stati così individuati altri 2 cloni (HMGI-C4 e HMGI-C7) di lunghezza rispettivamente 1,4 Kb e 2,3 Kb. Entrambi risultavano iniziare nella zona al 3' non tradotto e si estendevano ulteriormente in quella direzione.

La sequenza primaria completa della proteina HMGI-C ricavata dalla sequenza

nucleotidica del clone isolato ha subito rivelato caratteristiche molto interessanti (TAVOLA I). Innanzitutto sono state ritrovate interamente le sequenze dei peptidi preparati a partire dalla proteina purificata da LLC e un unico residuo aminoacidico che non era stato identificato tramite la degradazione di Edman è risultato essere triptofano, un fatto insolito per le HMGI, per altro confermato da analisi spettroscopiche di fluorescenza sulla proteina purificata. Inoltre, era già stato notato che HMGI-C murina era in grado di dimerizzare tramite la formazione di ponti disolfuro [64] ed in effetti il modulo di lettura dedotto dal cDNA ha confermato anche la presenza di un residuo di cisteina. E' stata così riscontrata un'omologia di circa il 50% con HMG I(Y), che risulta molto più alta se si confrontano i singoli domini I, II, III, IV. I primi tre sono basici e sono abbastanza simili, in particolare il II ed il III corrispondono al motivo "A-T hook" descritto da altri autori [70]; il quarto dominio corrisponde all'estremità N-terminale acida. Gli undici aminoacidi che differenziano HMG I da HMG Y sembrano mancare anche in HMGI-C, per cui la spaziatura fra i domini I e II è similmente ridotta in HMG Y e HMGI-C, d'altro canto sono maggiormente distanziati i domini III e IV di HMGI-C rispetto a ciò che si verifica nelle altre due proteine. Le differenze più significative che si riscontrano fra HMG I(Y) e HMGI-C risiedono nella parte N-terminale, in quanto su 21 residui ce ne sono 13 non conservati. Per quanto riguarda i siti di fosforilazione, altrettanto importanti ai fini di chiarire il ruolo della proteina, dei due siti di consenso per la casein chinasi II (serine 116 e 117 nella numerazione di TAVOLA I) presenti nelle proteine HMG I(Y) umane e murine, uno solo è conservato in HMGI-C murina (serina 116). Queste fosforilazioni sul IV dominio C-terminale potrebbero influire sull'interazione della proteina col DNA, in analogia a ciò che è stato dimostrato per il prodotto del proto-oncogene *c-myb* [83], ma è anche possibile che l'alterazione avvenga a livello del legame con altre proteine le quali potrebbero interagire con quella parte della molecola. A questo proposito, un successivo lavoro ha permesso di identificare esattamente i siti di fosforilazione su HMG Y estratta da LLC e

purificata mediante cromatografia. Era stato notato infatti che la proteina identificata nel nostro laboratorio come HMGI-E eluiva in due frazioni distinte che differivano appena anche nella mobilità elettroforetica in acido acetico/urea. Tramite l'utilizzo della spettrometria di massa è stato possibile stabilire che effettivamente HMGI-E corrisponde a HMG Y, le due frazioni di questa proteina sono due forme a diverso grado di fosforilazione prive della metionina iniziale e contenenti N-acetilserina all'estremità aminoterminale. E' stato inoltre possibile mappare esattamente i siti di fosforilazione che sono risultati tre (serine 113, 116 e 117 secondo la numerazione di TAVOLA I) per la frazione eluita per prima e due (serine 116 e 117) per l'altra. Non è stata rilevata la fosforilazione del motivo di consenso per la p34^{cdc2}, ma ciò può essere dovuto al fatto che le frazioni sono state ricavate da una popolazione mista di cellule tumorali in diversi stadi del ciclo cellulare, per cui solo una piccola parte poteva trovarsi nella fase in cui tale modificazione diventa rilevante [84].

Per quanto riguarda i siti per la chinasi p34^{cdc2}, la quale è correlata con le varie fasi del ciclo cellulare, sono pure due in HMG I(Y) di origine umana, ma solo uno è conservato nel corrispettivo murino (treonina 56, nella numerazione di TAVOLA I) e nell'HMGI-C murina questo stesso motivo di consenso rimane presente, sebbene la treonina sia sostituita da una serina. Questo sito può essere coinvolto in un meccanismo di regolazione collegato al ciclo cellulare, dal momento che altri autori hanno dimostrato come la sua fosforilazione specifica nell'ambito del peptide corrispondente al dominio II porti all'indebolimento dell'interazione col DNA [85]. Da tutti i dati esposti, risulta chiaro che HMGI-C è il prodotto di un gene diverso da quello che codifica per HMG I(Y), con tutto ciò che ne può conseguire in quanto a regolazione ed espressione, confermando anche da questo punto di vista la validità dei risultati ottenuti dalle analisi biochimiche effettuate sui diversi tessuti e la necessità di ulteriori indagini sulla sua espressione.

In questa direzione infatti, sono stati impostati i successivi studi portati avanti nel nostro laboratorio ed a tal fine sono state scelti due approcci diversi, uno volto a

caratterizzare in modo più dettagliato alcuni sistemi tumorali di diversa origine sulla base di un campione di riferimento e l'altro impostato sull'estensione a vasto raggio dell'analisi a campioni alterati in senso neoplastico non ancora conosciuti sotto l'aspetto dell'espressione delle HMGI.

Il tessuto da LLC è stato scelto quale modello per l'approfondimento dello studio dell'espressione delle HMGI in relazione a quella degli altri componenti proteici presenti nel sistema e presumibilmente correlati, vista la cospicua presenza di tutte e le tre proteine in questione. Avendo la possibilità di disporre di un altro tessuto tumorale murino di derivazione linfatica, è stato da noi intrapreso lo studio comparativo di questi due tessuti. Ai fini di una corretta valutazione sui dati sperimentali è stato ritenuto opportuno eseguire analisi anche su tessuti di controllo, intendendo con ciò campioni che rappresentano una controparte normale, senza alterazioni. Sono stati perciò parallelamente analizzati tessuto polmonare e tessuto da timo di topo. Tutti i tessuti sono stati estratti con acido perclorico per quantificare il contenuto in proteine HMG di ciascuno, ripetendo per due volte l'estrazione sullo stesso campione infatti si estrae completamente la parte proteica solubile dal momento che una ulteriore estrazione nelle stesse condizioni non dà più alcuna resa. Mantenendo come riferimento l'istone H1 che è presente nella cromatina in una proporzione pressochè sempre uguale rispetto al DNA e viene coestratto nelle stesse condizioni, è possibile perciò valutare eventuali differenze nella composizione relativa dei campioni, a parità dell'ammontare dell'istone [86, 38]. La caratterizzazione dei campioni è stata realizzata mediante tecniche cromatografiche ed elettroforetiche abbinate alla lettura densitometrica ed ha messo in evidenza l'incremento nell'espressione delle HMGI nei tessuti tumorali in confronto ai rispettivi tessuti di controllo. Le analisi di questo tipo sono state ripetute più volte, in modo da contenere l'errore stimato entro il 10% (vedi "Materiali e metodi"). L'espressione delle HMG 14/17 è risultata più difficilmente correlabile con uno stato alterato in senso neoplastico, dal momento che nei tessuti

di controllo prevale su quella delle HMGI, nel caso dei tessuti da linfoma la supera di poco, mentre in LLC risulta decisamente minore (fig.2 dal rif. [38]). Questi dati ci hanno indotto a pensare che non ci sia una relazione ben definita fra l'espressione delle HMG 14/17 e le HMGI, pertanto sembra plausibile ritenere che i ruoli delle due diverse classi di proteine siano distinti [38]. Mentre nel nostro laboratorio erano in corso questi esperimenti, altri gruppi di ricerca riportavano la capacità d'interazione delle HMGI con vari tipi di sequenze di attivazione ed elementi regolatori posti a monte di diversi geni, in particolare quelli codificanti per il tumor necrosis factor- β (TNF- β) [87] e per la proteina gp91-phox, necessaria per l'attività microbica nei fagociti e marker di differenziamento di questi ultimi [88]. Nonostante da tutti questi dati provenisse un' indicazione per un coinvolgimento nell'attivazione dei geni da parte delle HMGI, queste proteine, di per se stesse, non erano in grado di attivare la trascrizione e non c'erano ancora indicazioni precise sulle eventuali interazioni con gli altri fattori di trascrizione. Due importanti studi hanno aperto la strada a questo tipo di indagini dimostrando la funzione delle HMGI quali fattori di trascrizione "accessori" in grado di affiancare altre proteine per costituire dei complessi di attivazione specifici [89, 90]. E' stato dimostrato infatti che l'HMG I(Y) si lega su una sequenza ricca in A/T localizzata nel centro del sito di legame specifico per il fattore di trascrizione eterodimerico NF- κ B il quale si trova in una regione regolatrice a monte del gene umano per l' interferone- β (hINF- β), inducibile tramite infezione virale. Affinchè venga attivata la trascrizione, deve formarsi un complesso fra queste due proteine ed il DNA; la presenza di entrambe le proteine è necessaria, in quanto l'alterazione del legame dell'una o dell'altra provoca un decremento dell' induzione virale *in vivo* [89]. Successivamente lo stesso gruppo di ricerca ha dimostrato che l'HMG I(Y) è in grado di stimolare il legame di un altro fattore di trascrizione, l'ATF-2, nel suo sito specifico all'interno della stessa regione genica, la quale ha in tutto un'estensione di una cinquantina di paia di basi. ATF-2 è una proteina che si lega al DNA appartenente alla classe

"leucine-zipper" la quale anche *in vivo* è in grado di formare eterodimeri con il prodotto del protooncogene *c-jun* e in questa forma, può interagire con la sequenza specifica. Sono state dimostrate infine specifiche interazioni proteina-proteina anche in assenza del DNA: l'HMG I(Y) può interagire sia con la regione "leucine-zipper" di ATF-2, con c-Jun, con NF- κ B ed inoltre anche quest'ultimo ha la capacità di stabilire legami con ATF-2 e c-Jun. Sulla base di tali dati, si può ipotizzare che le proteine HMGI svolgano un ruolo centrale di coordinazione fra le distinte sequenze inducibili da virus, assicurando un corretto ed ordinato assetto dei vari fattori di trascrizione sul DNA, in una struttura adatta al posizionamento del complesso basale di trascrizione. E' stato pertanto proposto un modello in cui più proteine HMGI sono in grado di interagire con il DNA permettendo l'assemblamento di complessi proteici regolatori a più componenti, facilitandone i contatti [90].

L' aumentata presenza delle HMGI nei sistemi trasformati poteva quindi, con molta probabilità, essere indicativa di un'alterata struttura in alcune regioni della cromatina, per cui era senz'altro d'interesse accertare la situazione di altre molecole potenzialmente coinvolte in transizioni strutturali analoghe, quale l'istone H1. Già altri studi avevano ipotizzato un meccanismo di modulazione della struttura della cromatina tramite l'espressione differenziale delle varie isoforme esistenti di questa proteina [7, 32, 37]. In particolare, erano state messe in evidenza molte forme a diverso grado di fosforilazione dei vari subtipi dell'H1 in diversi sistemi neoplastici [38 e riferimenti ivi citati]. Il problema principale che si presentava era riuscire ad individuare esattamente la molecola di cui si parlava, dal momento che la nomenclatura era eterogenea e soggettiva in relazione al fatto che mancavano dati completi sulle sequenze primarie e le tecniche usate per l'identificazione delle singole subfrazioni (essenzialmente cromatografia ed elettroforesi in SDS) spesso si rivelavano non sufficientemente precise. Abbiamo pensato pertanto di utilizzare in aggiunta alle tecniche tradizionali, l'analisi di spettrometria di massa la quale, misurando il peso molecolare della specie si basa su di una caratteristica intrinseca e

Peak	Molecular mass of histone H1 from				Identified histone
	lung	LLC	thymus	NQ35	
I	20 730	20 727	n.d.	20 731	H1°
	<u>20 771</u>	20 770	n.d.	<u>20 772</u>	
	20 838	n.d.	n.d.	20 836	
	n.d.	22 055	22 056	22 058	
	n.d.	n.d.	22 133	n.d.	H1b H1b + 1P
	<u>22 486</u>	22 486	<u>22 487</u>	<u>22 489</u>	
	n.d.	<u>22 566</u>	22 570	n.d.	
	n.d.	n.d.	n.d.	22 583	
IV	21 259	21 263	21 266	21 265	H1c + 2P H1a
	n.d.	n.d.	n.d.	21 338	
	<u>21 694</u>	<u>21 695</u>	<u>21 697</u>	<u>21 696</u>	
	n.d.	n.d.	n.d.	21 771	
	n.d.	21 792	21 795	21 792	
V	21 216	n.d.	n.d.	n.d.	H1c + 2P H1c + 3P H1c + 5P H1a
	n.d.	21 336	21 342	21 340	
	n.d.	21 415	21 417	n.d.	
	21 457	21 455	21 459	21 458	
	21 579	21 577	21 580	21 579	
	n.d.	21 696	n.d.	n.d.	
V	<u>21 886</u>	21 886	21 888	<u>21 888</u>	H1e
	n.d.	21 968	n.d.	n.d.	H1e + 1P
	22 010	<u>22 008</u>	<u>22 011</u>	<u>22 011</u>	H1d
	n.d.	22 086	<u>22 093</u>	<u>22 091</u>	H1d + 1P
	n.d.	n.d.	22 108	n.d.	
VI	n.d.	n.d.	20 538	20 538	H1c
	20 776	20 777	20 778	20 778	
	20 854	20 854	n.d.	n.d.	
	n.d.	n.d.	20 915	n.d.	
	<u>21 177</u>	<u>21 176</u>	<u>21 178</u>	<u>21 178</u>	
	n.d.	21 198	n.d.	n.d.	
	n.d.	n.d.	21 257	n.d.	
	21 275	21 273	n.d.	n.d.	
VII	n.d.	21 176	n.d.	n.d.	H1c H1L H1e H1d
	n.d.	21 327	n.d.	n.d.	
	n.d.	21 509	n.d.	n.d.	
	n.d.	<u>21 756</u>	n.d.	n.d.	
	n.d.	21 888	n.d.	n.d.	
	n.d.	22 012	n.d.	n.d.	
VIII	n.d.	20 800	n.d.	n.d.	
	n.d.	21 097	n.d.	n.d.	
	n.d.	21 736	n.d.	n.d.	
	n.d.	22 690	n.d.	n.d.	
VIII	n.d.	22 732	n.d.	n.d.	
	n.d.	22 946	n.d.	n.d.	

tabella 1

non su di un comportamento che può essere soggetto a determinate condizioni dell'ambiente circostante. Tale approccio si è rivelato un' ottima procedura per l'identificazione delle tante subfrazioni ed ha messo in luce la reale complessità della situazione.

L'istone H1 totale estratto da ciascun campione è stato risolto tramite separazione cromatografica a fase inversa nei suoi componenti e le singole frazioni sono state analizzate in elettroforesi in SDS e tramite l'analisi di massa, per la quantificazione è stata usata la lettura densitometrica su gel (vedi "Materiali e metodi"). Tramite il confronto delle frammentarie informazioni sulla sequenza disponibili ed i dati ottenuti dalle varie analisi effettuate, in accordo con la nomenclatura preesistente [91-94] è stato assegnato il peso molecolare a ciascuna subfrazione [38]. E' stato possibile stabilire i sei subtipi predominanti dell'istone H1 nei tessuti murini, di seguito elencati secondo il loro ordine di eluizione in cromatografia a fase inversa:

H1° : 20771 Da

H1b : 22487 Da

H1a : 21696 Da

H1d : 22010 Da

H1e : 21887 Da

H1c : 21177 Da

Come si può vedere, (tabella 1), in ogni frazione cromatografica in realtà ci sono molte specie oltre a quella predominante (sottolineata), delle quali solo alcune vengono ritrovate in tutti i campioni murini analizzati. In particolare, nel tessuto LLC è stata ritrovata una specie con caratteristiche tali da poter essere considerata una variante dell'istone H1, ma che non trova corrispondenza con quelle riportate precedentemente. Questo subtipo è stato da noi chiamato H1L e probabilmente si tratta di una variante della sequenza dell'istone, la cui massa risulta di 21756 Da. L'analisi di massa ci ha permesso di stabilire quali subfrazioni sono fosforilate e quanto questa modificazione sia estesa. Nelle analisi da noi effettuate, la

subfrazione H1c risulta la più soggetta ad essere fosforilata, sono infatti state identificate le forme recanti due, tre e cinque gruppi fosfato. Anche per le subfrazioni H1b, H1d e H1e è stata evidenziata la presenza delle rispettive forme monofosforilate.

La distribuzione di queste subfrazioni all'interno di ciascun tessuto analizzato è riassunta nella fig. 6 del rif. [38]. I sottotipi H1d e H1e, che coeluiscono, costituiscono assieme circa la metà del totale di H1 presente, tranne che in LLC dove il valore è più basso e si registra la presenza di specie peculiari non presenti negli altri casi, le quali costituiscono circa il 20% del totale in questo campione. H1a risulta essere la specie più scarsamente rappresentata. Interessante ancora è seguire l'andamento della presenza di H1°, dal momento che questo istone è prevalentemente espresso in cellule ben differenziate pertanto con basso grado di proliferazione [7, 37]. E' stato riportato il valore del rapporto fra le quantità rilevate di H1° e H1b, i quali coeluiscono in cromatografia ma sono facilmente distinguibili in elettroforesi. Nei due sistemi neoplastici da noi esaminati i valori di tale rapporto sono molto diversi, ciò lascia presumere che sussistano situazioni diverse per quanto riguarda differenziamento e proliferazione. Fra i tessuti normali, il polmone presenta una situazione di prevalenza dell'H1°, in accordo con i dati riportati da altri autori, mentre nel timo questa specie risulta assente, confermando l'anomalia di questo tessuto peraltro già riportata da altri autori [94].

La fosforilazione dell' H1 è stata estesamente studiata specialmente in relazione alla proliferazione, ma sono state determinate anche molte forme fosforilate in campioni alterati in senso neoplastico [92, 94-97]. In effetti complessivamente, nel tessuto normale da polmone si osserva il livello più basso di fosforilazione, mentre la predominanza di una specie fosforilata sussiste in LLC, dove H1b è presente soprattutto nella sua forma modificata dall'aggiunta di un gruppo fosfato (tabella 1). Nei tessuti esaminati è plausibile che solo una piccola frazione delle cellule si trovi in una determinata fase della mitosi pertanto le modificazioni registrate non

necessariamente rispecchiano l'azione della chinasi p34^{cdc2}, parte del complesso già precedentemente indicato come chinasi dell'istone H1 associata alla crescita. Inoltre si deve tener presente che fra i molti potenziali siti di fosforilazione sulle isoforme dell'H1, ce ne sono vari che corrispondono a motivi di consenso per chinasi di diverso tipo, la cui azione si manifesta nell'interfase.

Varie evidenze mettono in relazione le HMGI e l'istone del DNA linker: anche per le HMGI è stata stabilita una condizione di fosforilazione durante l'interfase [52, 64, 73, 77], molti sono i dati sulla loro interazione nelle zone ricche in A/T del DNA, tramite immunolocalizzazione è stata determinata la loro presenza in corrispondenza delle bande G/Q (Giemsa/quinacrina) e C (costante eterocromatina) ricche in A/T dei cromosomi metafasici di mammifero [98].

Il coinvolgimento di queste due differenti categorie di proteine nella struttura della cromatina è stato sottolineato anche da altri studi, effettuati contemporaneamente da altri gruppi di ricerca. Una serie di analisi ha contribuito a far emergere un diverso tipo di regolazione connessa in modo più diretto alla struttura della cromatina in cui le HMGI possono avere un ruolo importante. Esistono all'interno del DNA diversi elementi che delimitano strutturalmente alcune aree della cromatina. Fra questi vi sono le regioni associate con la struttura nucleare (scaffold associated regions, SARs, o anche matrix associated regions, MARs), le quali sono ritenute le zone del DNA che vengono a contatto con l'impalcatura proteica che sostiene il nucleo. Le sequenze di tali regioni, lunghe centinaia di paia di basi, sono molto ricche in A/T (> 70%), *in vitro* sono in grado di legarsi alla matrice nucleare ed a molte proteine, fra cui l'istone H1 e la topoisomerasi II, i principali candidati nell'organizzazione delle fibre cromatiniche e dei cromosomi. Per queste regioni, la cui definizione si basa su caratteristiche biochimiche e non funzionali, è stato ipotizzato un ruolo nel controllo dell'espressione genica tramite un meccanismo di distribuzione delle proteine in grado di conferire caratteristiche di attivazione o di repressione alla cromatina. In particolare, le zone ricche in A/T potrebbero influenzare, legandolo

preferenzialmente, la distribuzione dell' H1 all'interno della cromatina e pertanto in altre regioni, ad esempio quelle ricche in G/C a monte di molti geni espressi costitutivamente (CpG islands), l'istone sarebbe presente in quantità minore o anche assente. Tale modello prevede l'esistenza di fattori in grado di competere con l'H1 per le zone ricche in A/T in maniera da permettere, nel momento in cui se ne presenti la necessità, una redistribuzione non selettiva fra le varie regioni della cromatina. Ciò può essere ottenuto *in vitro* tramite composti in grado di legarsi con alta affinità e specificità nei tratti ricchi in A/T, come l'antibiotico distamicina [99]. E' stato dimostrato *in vitro* che l' HMG I(Y) è in grado di competere con l'istone H1 per il legame sul DNA a livello di regioni di tipo SAR e di portare così ad una derepressione della trascrizione di eventuali geni posti sotto il controllo di tali elementi. E' stato messo in luce che la parte C-terminale ricca in residui acidi dell'HMG I(Y) non è coinvolta in interazioni specifiche con l'istone, dal momento che mutanti di delezione privi di tale dominio sono altrettanto capaci di esercitare l'attività di derepressione svolta dalla proteina intera [100]. Mediante frazionamento della cromatina per digestione con nucleasi esogene, è stato possibile stabilire qual'è *in vivo* la localizzazione dei vari componenti del sistema. HMG I(Y) è presente in modo preponderante nella frazione solubile della cromatina, ottenuta dopo breve digestione con nucleasi micrococcica e composta prevalentemente da mononucleosomi. La stessa risulta arricchita anche per quanto riguarda le altre proteine HMG e priva dell'istone H1, il quale è riscontrabile invece nelle altre frazioni, costituite da componenti che rimangono associati con la matrice nucleare [100]. Per quanto riguarda la relazione fra HMG e struttura nucleosomale, era stato dimostrato soltanto per le HMG 14/17 un ruolo effettivo e d'altro canto esperimenti *in vitro* sulla capacità da parte delle varie classi di HMG nell'influenzare in qualche modo l'arrangiamento nucleosomale della cromatina avevano confermato questa ipotesi, anche se la presenza delle HMGI sembrava aumentare la suscettibilità al taglio da parte delle nucleasi [49, 50]. Un altro studio ha dimostrato invece che le

HMGI sono in grado di legarsi sia *in vitro* che *in vivo* non solo al doppio filamento di DNA in corrispondenza di zone ricche in A/T, ma anche quando questo sia strutturato nella particella nucleosomale. Tramite cross-linking è stato possibile individuare la zona di interazione che è in prossimità del dimero (H2A-H2B) e dell'istone H3. La rimozione proteolitica delle estremità aminoterminali degli istoni dell'ottamero abolisce l'interazione del nucleosoma con l'HMG I(Y). Ancora una volta l'uso di mutanti di delezione ha dimostrato che neanche questa volta l'estremità carbossiterminale acida delle HMGI è coinvolta nell'interazione [101].

L'espressione delle HMGI si è rivelata fin dal primo momento diversa rispetto quella delle altre HMG, poichè connessa ad un particolare stato di proliferazione, di differenziamento o di alterazione del sistema biologico e per tale motivo, alle informazioni di tipo biochimico se ne sono aggiunte altre riguardanti la biologia cellulare ed anche la patogenesi dei sistemi studiati. Mentre erano in corso nel nostro laboratorio gli esperimenti descritti, altri dati in questo senso provenivano da altre ricerche in corso. Un clone la cui sequenza codifica per HMG I(Y) era stato isolato tramite una tecnica di ibridazione differenziale effettuata su due linee di tumore prostatico di ratto a diverso grado di malignità nel corso di un'indagine portata avanti allo scopo di individuare possibili "markers" specifici per determinare l'aggressività dei tumori prostatici. Da questo studio risulta che l'aumentata espressione di HMG I(Y) si correla con l'abilità metastatica piuttosto che col grado di proliferazione e pertanto ciò può contraddistinguere cellule in avanzato stadio tumorale [102]. Un altro studio su linee cellulari da epitelio mammari preneoplastici e tumorali di topo conferma l'aumento dell'espressione dell' HMG I(Y) in relazione al grado di trasformazione neoplastica a parità di grado di proliferazione. Un'analisi comparativa sulla struttura dei loci dell'HMG I(Y) nelle varie linee rivela che potrebbero essere implicate anomalie nell'arrangiamento dei geni [103]. Il gene murino per l'HMG I(Y) è stato localizzato sul cromosoma 17 in una zona di particolare interesse poichè vicino mappano molti geni letali nell'embrione ed

altri che sono attivati nelle cellule embrionali e carcinomatose [104]. La localizzazione del gene *Hmgi* nel topo faceva pensare che l'omologo umano di questo gene si trovasse nel cromosoma 6, come è effettivamente risultato confermato in seguito, in uno studio successivo che ha dettagliatamente messo in luce l'organizzazione del gene per l'HMG I(Y) nel genoma umano [105]. Un'altra indagine sulle traslocazioni di un sito frequentemente coinvolto in riarrangiamenti cromosomiali che si riscontrano nelle leucemie acute ha portato all'identificazione dell' omologo umano di una proteina della *Drosophyla* che differisce da questa per la presenza di un motivo "A·T hook" simile a quello delle HMGI ed altre due zone che appaiono varianti dello stesso tipo di dominio. Il riarrangiamento del gene che codifica questa proteina dà luogo a prodotti di fusione coinvolti in processi oncogenici [106, 107].

L'altro tipo di approccio che era stato intrapreso e portato avanti nel nostro laboratorio riguardava proprio l'estensione dell'analisi dell'espressione delle HMGI sul maggior numero possibile di campioni alterati in senso neoplastico. Non era semplice passare a questo tipo di studi mediante le tecniche tradizionalmente usate poichè l'estrazione ed analisi delle proteine HMGI richiedeva la disponibilità di notevoli quantità di tessuto o cellule, vista la loro relativa scarsità. Si è pensato pertanto all'utilizzo di anticorpi e tecniche di immunoriconoscimento per poter determinare agevolmente la presenza delle proteine nei campioni. L'immunizzazione di un animale ai fini di ricavare un antisiero in grado di riconoscere le HMGI non era stata intrapresa proprio perchè le quantità di proteine ricavabili dalla purificazione dei tessuti erano ben al di sotto di quelle che sarebbero state necessarie per la preparazione di un buon antigene. Si è pensato pertanto di esprimere la proteina ricombinante, inserendo in un vettore di espressione procariotico pAR3038 la sequenza nucleotidica codificante per l'HMGI-C murina e trasformando con esso un ceppo adeguato di cellule batteriche. Dai batteri raccolti per centrifugazione è stata effettuata l'estrazione con acido perclorico e

Crec/Scans 10-24

+Profile Q1SCAN

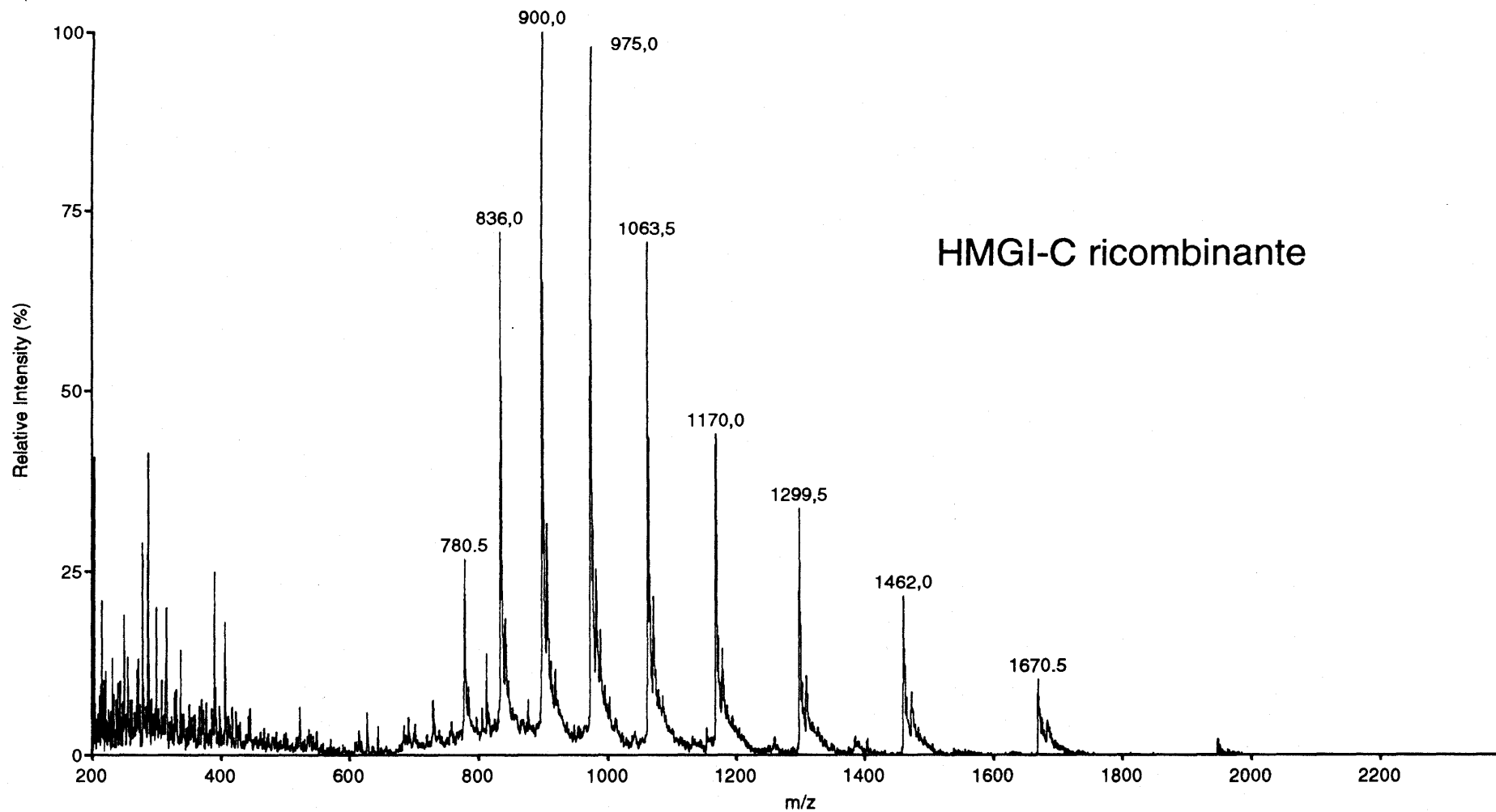
Average of scans 10 to 24 Time=2,40 min

Crec - 23/09/1994 - 5:52

Crec

9 peaks

509,667



HMGI-C ricombinante

+Profile Q1SCAN

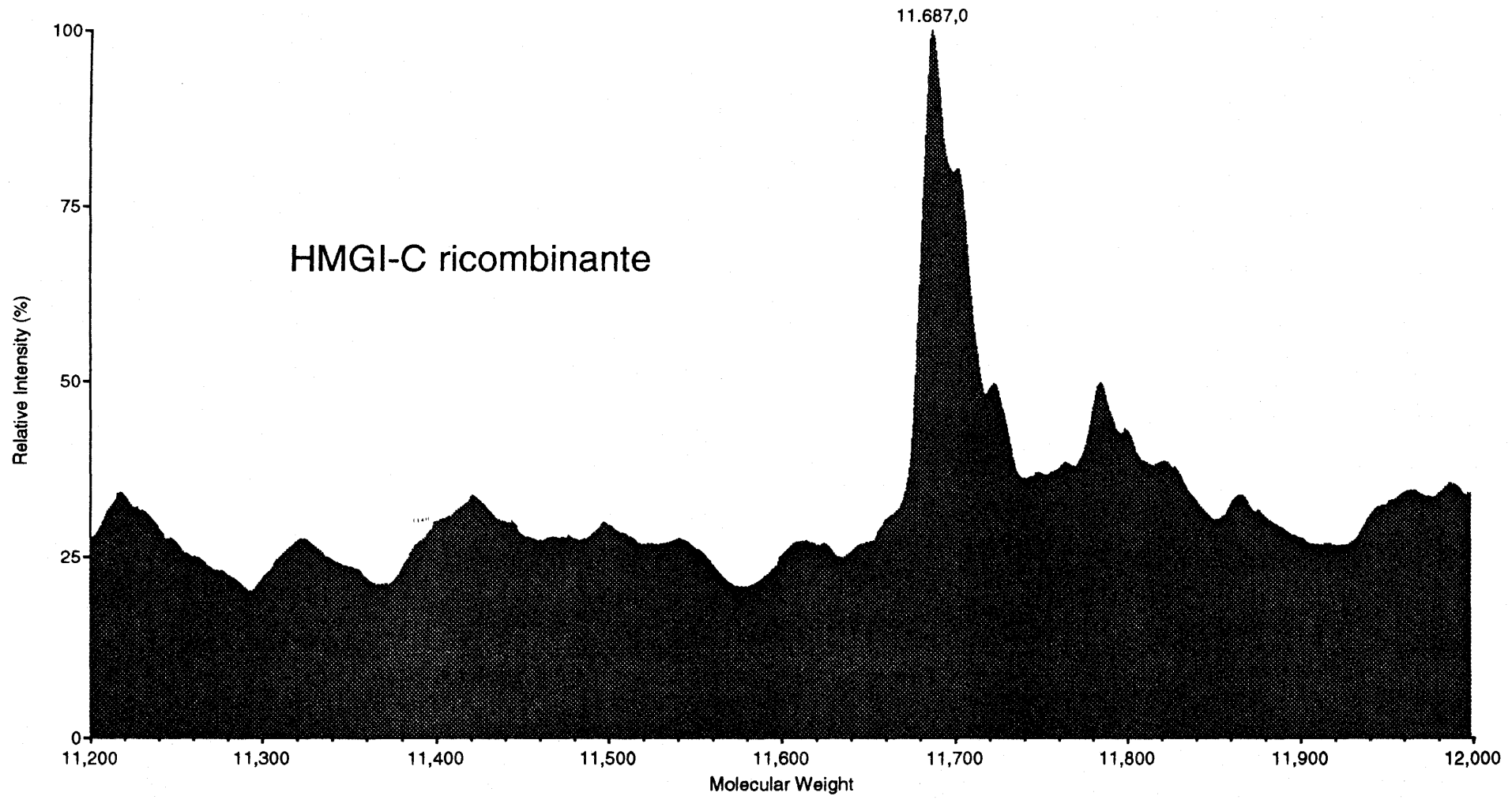
Reconstruct 6

Crec - 23/09/1994 - 5:52

Crec

1 peak

3,265,230



successivamente si è ottenuta la purificazione della proteina mediante cromatografia a fase inversa. L'analisi di spettrometria di massa sul prodotto ottenuto (TAVOLA IIa e IIb) effettivamente dimostra la presenza di un solo componente ed il peso molecolare misurato coincide perfettamente con quello teorico calcolato in base alla sequenza da cDNA che risulta di 11687 Da escludendo la metionina iniziale e considerando acetilato l'N-terminale del primo aminoacido serina. Utilizzando la proteina ricombinante come antigene è stato immunizzato un coniglio in modo da sviluppare un anticorpo policlonale. Abbiamo purificato gli anticorpi tramite cromatografia di affinità, derivatizzando una resina con la proteina ricombinante stessa. (vedi "Materiali e metodi"). E' stato messo contemporaneamente a punto un metodo di Western blot che consentisse il trasferimento delle proteine da gel di acido acetico/urea su membrana di Immobilon, dal momento che la risoluzione delle tre HMGI in gel in SDS non è ottimale e può indurre in equivoco. Per prima cosa, gli anticorpi sono stati saggiati su una miscela contenente la proteina ricombinante usata come antigene e le frazioni purificate da LLC corrispondenti a HMG I e HMG Y; poi abbiamo analizzato con la stessa tecnica l'estratto totale delle proteine ottenuto da LLC. Poichè la proteina HMGI-C presenta circa il 50% di omologia di sequenza con HMG I ed HMG Y, l'anticorpo policlonale ottenuto contro HMGI-C è in grado di riconoscere anche le altre due. E' stato infatti ottenuto un ottimo segnale in tutti i casi, nonostante la proteina ricombinante usata come antigene presenti qualche differenza con la forma che si ritrova nei campioni estratti. Le HMGI infatti sono di norma più o meno modificate, ciò si vede chiaramente nel caso di HMGI-C, dove sul tracciato totale si distinguono più bande contigue di mobilità minore rispetto quella della proteina espressa nei batteri (fig. 1B dal rif. [108]). Ciò poteva creare qualche problema, poichè la mobilità elettroforetica poteva risultare insufficiente a discriminare fra le tre proteine. Si è pensato così di preparare altri antisieri policlonali in grado di riconoscere solo HMGI-C e solo HMG I(Y). Come antigeni specifici abbiamo scelto peptidi sintetici progettati in base alle zone di

sequenza che presentavano la minor omologia. Ci si proponeva di riconoscere soltanto HMG I e HMG Y usando l'anticorpo sviluppato contro la loro estremità N-terminale e soltanto HMGI-C, utilizzando come antigene un peptide corrispondente alla zona tra il dominio III e IV, circa la stessa in base al quale era stato progettato l'oligonucleotide sintetico usato come sonda per il vaglio della libreria cDNA. Un esame incrociato dei dati ottenuti con i tre anticorpi ci avrebbe permesso di definire esattamente la situazione per tutte e tre le proteine. Anche questi antisieri sono stati purificati mediante cromatografia di affinità e si sono rivelati efficaci nel riconoscere specificamente HMG I(Y) in un caso e HMGI-C nell'altro dal totale delle proteine estratte con acido perclorico da LLC. Sulla base dei dati precedenti ottenuti [77] abbiamo esteso l'indagine ad altri campioni neoplastici umani. In particolare sono state scelte altre linee di epatoma umano (Hep G2 e Hep 3B) dove con le tecniche tradizionali, la presenza di HMGI-C era difficile da stabilire con certezza assoluta anche perchè, a differenza di ciò che era stato osservato nei campioni murini, HMGI-C migrava proprio a ridosso di HMG I in acido acetico/urea. L'uso degli anticorpi ha permesso di definire con precisione la situazione per quanto riguarda le HMGI nei sistemi esaminati, mettendo in evidenza la presenza di HMGI-C in campioni umani maligni (fig. 1B dal rif. [108]). Inoltre ci è sembrato opportuno esaminare con questa tecnica sensibile alcune linee umane di cellule emopoietiche disponibili, dal momento che con le tecniche tradizionali tale tipo di tessuti risultava contenere solo HMG I(Y), sia nei campioni murini che umani [38, 77]. Questi risultati sono stati confermati dalle analisi di immunoriconoscimento [108].

Le differenze che erano state notate nelle caratteristiche di HMGI-C nei campioni umani rispetto a quelli murini ci avevano fatto pensare a qualche cambiamento in termini di sequenza primaria. Infatti, oltre alla migrazione elettroforetica diversa, nei campioni umani esaminati non era stato notato il dimero della proteina. Per verificare il grado di conservazione fra le proteine murine ed umane in collaborazione con un altro laboratorio, è stato intrapreso il clonaggio dell'omologo

umano di HMGI-C. Il confronto fra le due sequenze primarie ha messo in evidenza quattro sostituzioni aminoacidiche, di cui una interessa effettivamente l'unico residuo di cisteina presente nella proteina murina, giustificando pertanto la mancanza del dimero nell'umano, inoltre nel C-terminale è presente un residuo ulteriore di acido glutammico. Da un esame della sequenza risulta infine che i potenziali siti di fosforilazione e le regioni di legame con il DNA rimangono invariati fra umano e murino (TAVOLA III).

In seguito ad un'altra collaborazione, sono stati ottenuti risultati molto importanti per quanto riguarda il possibile ruolo della proteina HMGI-C nel processo di trasformazione neoplastica. In un sistema biologico già ben definito di cellule di tiroide di ratto che possono essere indotte ad assumere un fenotipo neoplastico tramite infezione con retrovirus [53, 64], è stato possibile alterare tale processo per mezzo di un RNA "antisense". Infatti, introducendo in queste cellule un costrutto che riproduce la sequenza complementare del DNA codificante per la proteina HMGI-C, la successiva infezione retrovirale, nonostante gli oncogeni virali siano abbondantemente espressi, non ha portato allo sviluppo del fenotipo trasformato. L'analisi biochimica da noi effettuata su tali campioni (fig. 9 dal rif. [109]) ha evidenziato un drastico mutamento nell'espressione delle tre proteine [109] e le analisi di Northern blot hanno confermato questo quadro. In seguito a trasfezione con il costrutto "antisense" e successiva infezione con retrovirus, si osserva una grande diminuzione dell'espressione di HMGI-C e la totale assenza di HMG I(Y). Due sono le ipotesi che si possono formulare per giustificare l'assenza in queste cellule di HMG I(Y). Il blocco della sintesi di HMGI-C, prevenendo la trasformazione neoplastica, potrebbe avere l'effetto secondario di inibire la comparsa delle altre due proteine. Altrimenti, il notevole grado di identità che sussiste fra le sequenze nucleotidiche codificanti per le proteine in questione, potrebbe spiegare il fatto che il costrutto di per sé riesca a bloccare la sintesi di HMG I(Y). La proteina HMGI-C non si comporta come un tipico oncogene

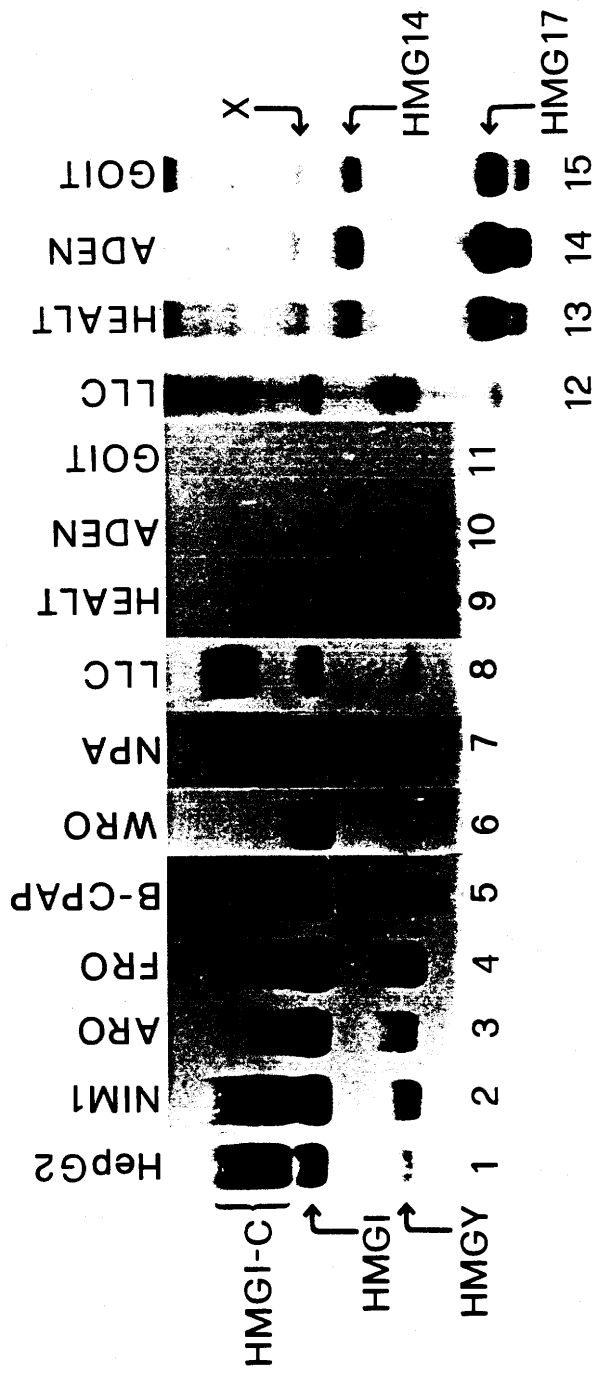


foto 1

trasformante, infatti la sua espressione indotta dalla trasfezione di un costrutto "sense" in cellule che di norma non la esprimono, risulta necessaria ma non è sufficiente per indurre il fenotipo trasformato [109].

Sulla base di tali dati è stato quindi intrapreso uno studio in collaborazione con lo stesso gruppo di ricerca, ai fini di determinare l'espressione delle HMGI nei carcinomi tiroidei umani e vedere se è definibile una correlazione col grado di malignità. Le proteine HMG I(Y) sono risultate espresse in tutte le linee cellulari di carcinoma tiroideo analizzate, mentre HMGI-C in alcuni casi non è risultata presente, e la complessità della sua regolazione necessita di ulteriori approfondimenti. Non è stata evidenziata una correlazione quantitativa fra i livelli di espressione delle HMGI e il grado di malignità. Al contrario, le analisi effettuate con varie tecniche hanno dimostrato come la presenza di queste proteine sia peculiare dei carcinomi tiroidei dal momento che non sono espresse affatto in campioni di tessuto normale o prelevato in presenza di disfunzioni tiroidee iperproliferative quali gozzo e adenoma follicolare benigno (foto 1). Inoltre in questo lavoro è stata messa a punto una tecnica di immunocitochimica su materiale paraffinato che ha permesso di estendere lo studio in maniera retrospettiva, e ciò può risultare di enorme interesse per definire l'utilità delle proteine HMGI come fattori utili ai fini di diagnosi [110].

Lo studio delle proteine basiche spermatiche del mollusco *Ensis minor* è un lavoro collaterale intrapreso parecchio tempo addietro e portato avanti nel nostro laboratorio, nell'ambito di una più generale caratterizzazione di proteine componenti la cromatina [40]. E' stato riportato che i nuclei delle cellule spermatiche mature dei molluschi bivalvi possono contenere sia istoni che proteine specifiche tipo protamine, di basso peso molecolare, quasi completamente costituite da aminoacidi basici ($\cong 80\%$) e senza particolare struttura secondaria. La non completa sostituzione degli istoni da parte delle protamine può spiegare il persistere della struttura

nucleosomale della cromatina che si osserva. In particolare è stata fatta una classificazione dei bivalvi sulla base della loro composizione riguardo le proteine spermatiche, denominate globalmente PL, protamine-like. Queste ultime, in base a criteri elettroforetici, sono state suddivise fondamentalmente in quattro categorie da I a IV. Sulla base di analisi biochimiche, si è visto che di solito alle classi I e II, appartenevano componenti istone-simili, caratterizzati dalla presenza di un dominio globulare, mentre le classi III e IV, comprendevano forme protamina simili [111].

Le proteine associate con il DNA negli spermatozoi maturi di *Ensis minor* erano state in precedenza estratte con acido perclorico, identificate e caratterizzate dal punto di vista fisico e biochimico [40]. Successivamente, i tre componenti principali dell'estratto, EM1, EM5 ed EM6 sono stati purificati e parzialmente sequenziati. È stata confermata per EM1 la sua natura di proteina spermatica specifica tipo protamina, per quanto riguarda EM5 ed EM6 è stato possibile identificarli come componenti istone H1-simili, per la presenza di un dominio globulare ben allineabile con quelli già noti per altre specie [82].

In particolare la nostra attenzione si è soffermata su EM6, la quale in elettroforesi ha l'aspetto di una miscela di isoforme ed ha un peso molecolare intorno a 40 kD. Al fine di ottenere la sequenza completa della proteina, si è pensato anche in questo caso di ricorrere a tecniche di biologia molecolare clonando il cDNA. Sono state raccolte le gonadi immature del mollusco ed è stato estratto l'RNA, allo scopo di costruire una libreria a cDNA. Mediante la tecnica della PCR ed usando l'RNA come stampo, è stata amplificata una regione corrispondente ad una cinquantina di aminoacidi, usando come "primers" due oligonucleotidi progettati sulla base della sequenza aminoacidica nota corrispondente alla fine del dominio globulare di EM6 (vedi "Materiali e metodi"). Il prodotto dell'amplificazione è stato subclonato e sequenziato, dimostrandosi effettivamente codificante per la regione aminoacidica scelta ed è stato usato quale sonda specifica per vagliare la libreria. Sono stati individuati in questo modo 15 cloni positivi, fra i quali è stato subclonato

ATTTTCATCTACAACATA ATG GCT GGA AGT AAG TCT AGA TCG AGG TCT CGT TCC AGA TCA AGG 64
 M A G S K S R S R S R S R S R
 1-EM1 →
 AGC AAG TCT CCG GCG AAG TCG GCC TCA CCG AAG AGA GCT GCT TCG CCG AAG GCG AGC AAA
 S K S P A K S A S P K R A A S P K A S K
 15
 TCT CCC AAA AGG AGC AAG TCT CCG AAA GCG GGG AAA TCT AGA AAG AGG AGT CGA TCT AGA 184
 S P K R S K S P K A G K S R K R S R S R
 35
 AGT AAA TCC AGC AAA CGA TCT GCT TCG AAA AAA CGC AGG AGC AGG AGT CCT AAG AAG AGG
 S K S S K R S A S K K R R S R S P K K R
 55
 AGC AAG TCC AAG AAA AGG TCT GCT TCT AAA AAG AGG AAG AGC AAA CCA CGA AAG AAG TCT 304
 S K S K K R S A S K K R K S K P R K K S
 75
 GCT TCC AAA AAG AGG TCT GCC TCC AAG AAG AGG AAG TCC AGG AAG AGG TCT GCT TCC AAA
 A S K K R S A S K K R K S R K R S A S K
 95
 AAG AGG AAG TCT GCC TCT AGA AAA AGG AGC ACA TCC AAG AAG AGG TCA GCT TCA AAG AAG 424
 K R K S A S R K R S T S K K R S A S K K
 115
 AGG AGC AAG TCT AGA AAG AGG TCT GCC TCT AGA AAT AAG AGT AAC AAC AAC ACC AAC AAT
 R S K S R K R S A S R N K S N N N T N N
 135
 TCC GCT AAG AAG AGG AGT CGC TCC AGG AAG AGG TCT GCC TCC AAA AAG AGG AGT CGC TCC 544
 S A K K R S R S R K R S A S K K R S R S
 155 1-EM6 → ←EM1-150
 AGG AAG AGG TCT GCC TCC AAA AAG AGG AGT CAC TCC AGG AAG AGG TCT GCC TCC AAA AAG
 R K R S A S K K R S H S R K R S A S K K
 20
 AGG AGC CAC TCC AGG AAG AGG TCT GCC TCC AAA AAG AGG AGC CAC TCC AGG AAG AGG TCT 664
 R S H S R K R S A S K K R S H S R K R S
 40
 GCC TCC AAG AAG AGG AGC AAA TCT AGG AAG AGG TCT GCC TCC AAG AAG AGG AGC AAA TCT
 A S K K R S K S R K R S A S K K R S K S
 60
 AGG AAG AGG TCT GCC TCC AAG AAG AGG AGC AAA TCT AGG AAG AGG TCT GCC TCC AAG AAG 784
 R K R S A S K K R S K S R K R S A S K K
 80
 AGG AGC AAA TCT AGG AAG AGG TCT GCC TCC AAG AAG AGG AGC AAA TCT AGG AAG AGG TCT
 R S K S R K R S A S K K R S K S R K R S
 100
 GCA TCC AAG AAG AGG AGC AAA TCT AGG AAG AGG TCT GCA TCC AAG AAG AGG AGC AAA TCT 904
 A S K K R S K S R K R S A S K K R S K S
 120
 AGG AAG AGG TCT GCA TCT AAG AAG AGG AGC AAA TCT AGG AAG AGG TCT GCA TCT AAG AAG
 R K R S A S K K R S K S R K R S A S K K
 140
 AGG AGC CAC TCC AGG AAA AGA CGT CCG TCT AAG AAG AGG AGC GAC TCC AGG AAA AGA CGT 1024
 R S H S R K R R P S K K R S D S R K R R
 160
 GCG TCT AAG AAG AGG AGC CAC TCC AGG AAA AGG TCT GCA TCT AAG AAG AGG AGC CAC TCC
 A S K K R S H S R K R S A S K K R S H S
 180
 AGG AAA AGA CGT GCA TCT AGA AGT GGG GTA ATG ACG AAA GTC ATG AAC GCG ATC GCT CAC 1144
 R K R R A S R S G V M T K V M N A I A H
 200
 TGC AAG AGT AGC AAG GGA TGT TCC GCC CAG GCA ATC AGG AAG TAT TTG GCC GCT CAC AGC
 C K S S K G C S A Q A I R K Y L A A H S
 220
 AAG CTG ACA GGT GTC TTC CTG AAC TTC CAC GTG AGG AGA GCC CTC GCA GCT GGT ATG AAG 1264
 K L T G V F L N F H V R R A L A A G M K
 240
 AAT CAC CTG CTG GCC CAT CCC AAG GGA TCA AAC AAC TTT ATC CTT GCT AAG AAG AAG GCA
 N H L L A H P K G S N N F I L A K K K A
 260
 CCC AGA CGA CGA AGG AGA GTT GCC AAG AAA GTG AAG AAG GCC CCA AAG AAG GCT AAG AAG 1384
 P R R R R R V A K K V K K A P K K A K K
 280
 ACC AAG AAG GTT GCC AAG AGG AAG AGC AAG AAG GCA AAG AAA TCA AAG AAG GCG AGG AAG
 T K K V A K R K S K K A K K S K K A R K
 300
 GCT CGC TCA CAC AAA AAG GCT GCT GCC CCA AAG AGA AGG GCA ACA CGC CGC CGC GCG TCT 1504
 A R S H K K A A A P K R R A T R R R A S
 320
 TCA CGA CGC CGC AGA GCC ACA AAG GCT CAC TAA GCTGTACATGACAGCTAAAAAACC CGCCCTT
 S R R R R A T K A H
 340 ←EM6-349
 AAGGGCCACCAAAAATAAAAAAAAAAAAAA

uno contenente un inserto di circa 1,5 Kb, ai fini di ottenerne la sequenza. Sorprendentemente, la zona al 5' è risultata codificante per la regione N-terminale di EM1 e solo dopo aver ottenuto la sequenza intera del clone, è stato possibile individuare le sequenze corrispondenti all'N-terminale e al dominio globulare di EM6, spiegando così il riconoscimento da parte della sonda. Nello stesso periodo, altri autori riferivano del clonaggio di un cDNA da *Mytilus trossolus* codificante per un potenziale precursore di proteine spermatiche che erano sempre state identificate come due prodotti non correlati, una del tipo protamina (PL-IV) ed una con un dominio globulare del tipo istone H1 (PL-II*). Ciò ha portato questi autori a formulare l'ipotesi dell'esistenza di una proteasi particolare specifica in grado di agire sul prodotto della traduzione di questo messaggero, dando luogo a due proteine distinte, anche alla luce del fatto che era stata rilevata la sequenza aminoacidica -NKSNN- posizionata fra i due prodotti estremamente inusuale per un istone o per una protamina e che quindi potrebbe fungere da consenso per la proteolisi post-traduzionale [112].

Il clone da noi trovato possedeva il segnale di poliadenilazione e la coda di poli(A) al 3', ma mancava al 5' della metionina segnale di inizio traduzione e di alcune basi codificanti per il primo aminoacido all'N-terminale di EM1. L'impiego della PCR (vedi "Materiali e metodi") ha permesso di ottenere l'intera sequenza codificante per entrambe le proteine (TAVOLA IV). Vari aspetti estremamente interessanti si possono rilevare dall'esame dei dati emersi dalla nostra ricerca, anche alla luce della situazione riscontrata dagli altri autori. Innanzitutto sono state confermate le sequenze aminoacidiche note per le porzioni N-terminali di EM1 ed EM6 e della parte C-terminale del dominio globulare di EM6; in secondo luogo, si riscontra anche in questo caso una sequenza molto particolare (-NKSNNNTNNS-) posizionata fra le due proteine, molto simile a quella già vista per *Mytilus*. Considerando i dati di massa per EM1, il valore misurato effettivamente coincide con quello calcolabile sulla base della sequenza da noi determinata via cDNA,

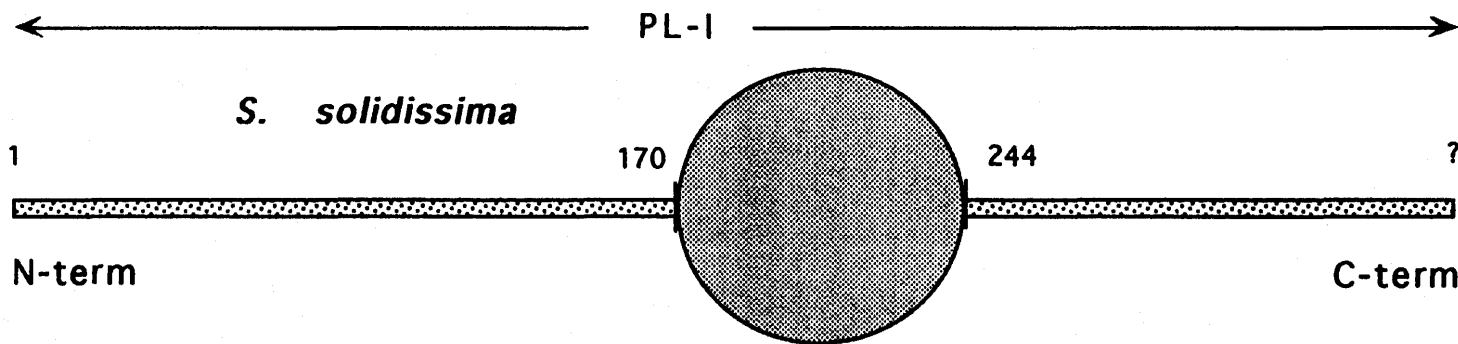
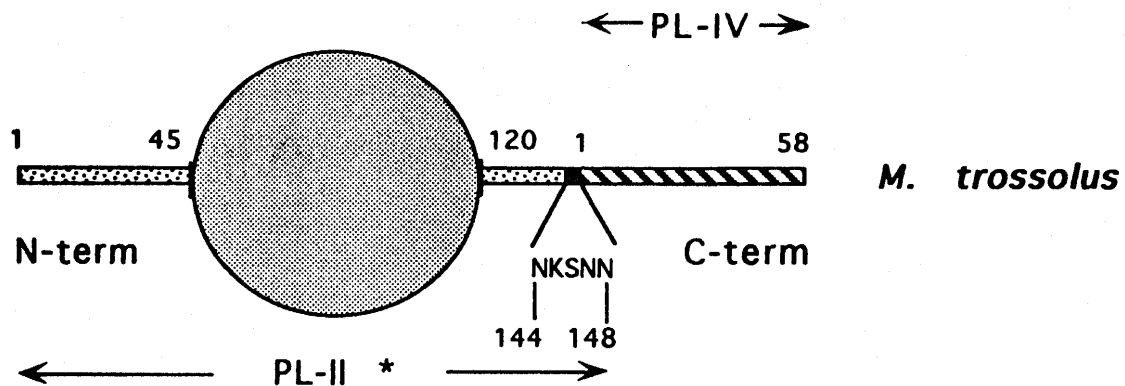
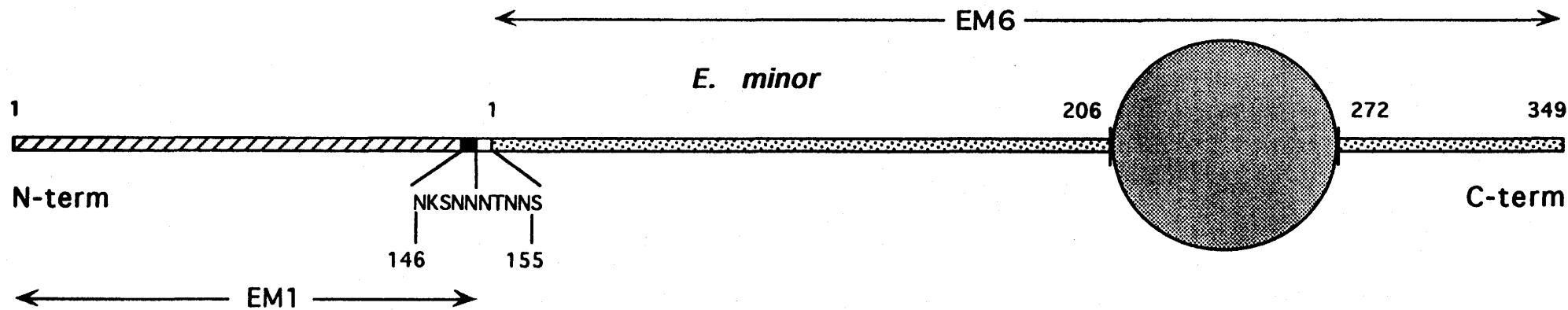


TAVOLA V

escludendo la metionina ed arrivando sino alla presunta sequenza segnale, con la perdita di un pentapeptide fra le due proteine. Il dato di massa per EM6 conferma il modulo di lettura.

Dal punto di vista del cDNA, la situazione da noi riscontrata ha delle caratteristiche particolari indubbiamente in comune con quella presente in *Mytilus*, sebbene sussistano delle differenze ad esempio nell'ordine della disposizione delle due proteine (TAVOLA V). A livello di proteine però ci sono notevoli divergenze: pur avendo in comune una struttura del tipo istone ed un certo grado di similitudine nel dominio globulare, EM6 infatti è piuttosto diversa dalla proteina PL-II*. E' invece certamente omologa alla proteina PL I da *Spissula solidissima* e ciò conferma la sua appartenenza alla categoria PL I della classificazione per le proteine spermatiche [82, 111]. EM1 si rivela molto diversa da PL IV da *Mytilus* e può collocarsi nella categoria PL III com'era stato già ipotizzato [82]. Considerando le due proteine di *Ensis minor*, si può notare come esse presentino un motivo simile, ripetuto molte volte che è riflesso anche nella sequenza nucleotidica codificante e, nel caso dell'esistenza di una proteasi specifica in grado di riconoscerlo, potrebbe essere spiegata la degradazione a cui va incontro EM6 in certe condizioni. Infine è da notare la presenza di potenziali siti di fosforilazione in EM1 che non si ritrovano nella parte N-terminale di EM6, i quali essendo motivi di consenso per la chinasi p34^{cdc2}, potrebbero modulare l'interazione fra la proteina ed il DNA nell'ambito di un meccanismo di condensamento della cromatina nello sperma maturo e successivo decondensamento che si rende necessario in seguito alla fecondazione per la fusione dei nuclei aploidi nello zigote [11]. I risultati di questa ricerca, oltre a stabilire la sequenza primaria completa di alcune di queste proteine danno un contributo all'ipotesi dell'esistenza di un meccanismo che utilizza una regolazione di tipo post-traduzionale che potrebbe essere comune e diffusa in organismi simili.

CONCLUSIONI

La distinzione fra ruolo strutturale e regolatorio dei componenti coinvolti nell'organizzazione della cromatina è diventata molto meno netta con il progredire delle ricerche in questo campo. Allo stesso tempo, da una visione di tipo "statico" della struttura della cromatina si è passati oggi a un'interpretazione "dinamica", in cui gli stessi componenti ritenuti strutturali possono venir modificati e coinvolti in altri tipi di interazioni, prendendo parte così ai rimodellamenti del materiale genetico e partecipando al cambiamento delle condizioni di attivazione/repressione dei geni. Questa visione ha cominciato ad emergere quando agli studi di tipo biochimico sono state affiancate evidenze sperimentali derivanti dalla genetica e dalla biologia cellulare e si è trovato il modo di riferire il dato biochimico ad una determinata condizione del sistema biologico. In effetti, ciò si rispecchia nel modo in cui è progredita la ricerca, iniziata parecchi anni fa, nel nostro laboratorio: le proteine da noi studiate sono tutte accomunate in quanto a caratteristiche biochimiche (il miglior esempio in questo senso è infatti la definizione delle proteine HMG), ma con l'acquisizione di nuove conoscenze sulla loro struttura primaria, espressione e relazione con fenotipi particolari, si accentuano le differenze spostando l'attenzione sul loro ruolo specifico.

Le proteine oggetto del nostro studio sono state caratterizzate dal punto di vista biochimico, sono state fatte analisi di tipo qualitativo e quantitativo e nel caso di prodotti sconosciuti, tramite metodi di biologia molecolare abbiamo determinato la sequenza primaria. Nel caso delle proteine di *Ensis minor*, nel perseguire questo scopo, è emersa un'interessante caratteristica legata all'espressione delle stesse, in quanto due prodotti distinti sono risultati codificati dallo stesso messaggero e probabilmente esiste un meccanismo post-traduzionale di regolazione in cui è coinvolta una proteasi (Bandiera, A., et al., lavoro in preparazione).

La nostra ricerca sui sistemi neoplastici ci ha portato alla conclusione che l'elevata presenza delle HMGI è peculiare del fenotipo trasformato, e l'espressione della terza proteina appartenente a questo gruppo, HMGI-C da noi individuata e studiata, pur essendo una caratteristica chiara di diversi tipi di cellule neoplastiche, risulta più ristretta. In questo senso è di fondamentale importanza la comprensione del fondamento biologico-molecolare della connessione fra la presenza di queste proteine e le alterazioni neoplastiche, a cui la nostra ricerca ha dato un significativo contributo. L'accertamento della correlazione fra espressione delle HMGI e fenotipo trasformato e l'estensione dello studio da sistemi murini a quelli umani ha dato luogo alla possibilità di approfondire l'aspetto applicativo pratico per possibili fini diagnostici. In particolare, nel sistema da noi studiato, la trasfezione con costrutti "antisense" per HMGI-C ha dimostrato che la presenza di queste proteine è necessaria per un'acquisizione del fenotipo neoplastico in seguito a infezione con retrovirus oncogeni. Lo studio eseguito su vari tipi di tessuti tiroidei alterati in senso neoplastico a diverso grado di trasformazione ha permesso di evidenziare come l'espressione delle proteine HMGI possa essere impiegata come "marker" diagnostico per questo tipo di tumori.

BIBLIOGRAFIA

1. Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., Klug, A., (1984), *Nature*, 311, 532-537..
2. Workman, J. L. & Buchman, A. R., (1993), *Trends Biochem. Sci.*, 18, 90-95.
3. Grunstein, M., (1992), *Scientific American*, 267, 40-47.
4. Kornberg, R. D. & Lorch, Y., (1992), *Ann. Rev. Cell Biol.*, 8, 563-587.
5. Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E., Moundrianakis, E. N., (1991), *Proc. Natl. Acad. Sci. USA*, 88, 10148-10152.
6. Grunstein, M., (1990), *Annu. Rev. Cell Biol.*, 6, 643-678.
7. Wolffe, A. P. & Dimitrov, S., (1993), *Crit. Rev. Euk. Gene Exp.*, 3, 167-191.
8. Graziano, V., Gerchman, S. E., Schneider, D. K., Ramakrishnan, V., (1994), *Nature*, 368, 351-354.
9. Clark, D. J. & Kimura, T., (1990), *J. Mol. Biol.*, 211, 883-896.
10. Wilson, C., Bellen, H. J., Gehring, W. J., (1990), *Annu. Rev. Cell Biol.*, 6, 679-714.
11. Roth, S. Y. & Allis, C. D., (1992), *Trends Biochem. Sci.*, 17, 93-98.
12. Bellard, M., Dretzen, G., Giangrande, A., Romain, P., (1989) *Methods Enzymol.*, 170, 317-346.
13. Drew, H. R., (1993), *J. Mol. Biol.*, 230, 824-836.
14. Kornberg, R. D., LaPointe, J. W., Lorch, Y., (1989), *Methods Enzymol.*, 170, 3-13.
15. Reeves, R. & Chang, D., (1983), *J. Biol. Chem.*, 258, 679-687.
16. Johns, E. W., (ed), (1982), *The HMG Chromosomal Proteins*, Academic Press, New York
17. Felsenfeld, G., (1992), *Nature*, 355, 219-224.
18. Hogan, M. E., Hayes, B., Wang, N. C., Austin, R. H., (1986), *Biochemistry*, 25, 5070-5082.
19. Shrader, T. E. & Crothers, D. M., (1989), *Proc. Natl. Acad. Sci. USA*, 86, 7418-7422.
20. Morse, R. H., (1992), *Trends Biochem. Sci.*, 17, 23-26.
21. van Holde, K. E., Lohr, D. E., Robert, C., (1992), *J. Biol. Chem.*, 267, 2837-2840.
22. Adams, C. C. & Workman, J. L., (1993), *Cell*, 72, 305-308.
23. Lewin, B., (1994), *Cell*, 79, 397-406.

24. Wolffe, A. P., (1994), *Curr. Opin. Genet. Dev.*, *4*, 245-254.
25. Wolffe, A. P., (1994), *Trends Biochem. Sci.*, *19*, 240-244.
26. Richard-Foy, H. & Hager, G. L., (1987), *EMBO J.*, *6*, 2321-2328
27. Bresnick, E. H., Bustin, M., Marsaud, V., Richard-Foy, H., Hager, G. L., (1992), *Nucleic Acids Res.*, *20*, 273-278.
28. van Holde, K. E., (1994), *Nature*, *367*, 512-513.
29. van Holde, K. E., (1993), *Nature*, *362*, 111-112.
30. Lee, D. Y., Jeffrey, J. H., Pruss, D., Wolffe, A. P., (1993), *Cell*, *72*, 73-84.
31. López-Rodas, G., Brosch, G., Georgieva, E. I., Sendra, R., Franco, L., Loidl, P., (1993), *FEBS Lett.*, *317*, 175-180.
32. Zlatanova, J., (1990), *Trends Biochem. Sci.*, *15*, 273-276.
33. Croston, G. E., Kerrigan, L. A., Lira, L. M., Marshak, D. R., Kadonaga, J. T., (1991), *Science*, *251*, 643-649.
34. Laybourn, P. J. & Kadonaga, J. T., (1991), *Science*, *254*, 238-245.
35. Jerzmanowski, A. & Cole, R. D., (1992), *J. Biol. Chem.*, *267*, 8514-8520.
36. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., Sweet, R. M., (1993), *Nature*, *362*, 219-223.
37. Doenecke, D., in *Architecture of Eukaryotic Genes*, Ed. G. Kahl, VCH, (1988), 123-141.
38. Giancotti, V., Bandiera, A., Ciani, L., Santoro, D., Crane-Robinson, C., Goodwin, G. H., Boiocchi, M., Dolcetti, R., Casetta, B., (1993), *Eur. J. Biochem.*, *213*, 825-832.
39. Dixon, G. H., Aiken, J. M., Jankowski, J. M., McKenzie, D. I., Moir, R., States, J. C., in *Chromosomal Proteins and Gene Expression*, Ed. G. R. Reeck, G. H. Goodwin, P. Puigdomènech, Plenum Press, (1985), 287-314.
40. Giancotti, V., Russo, E., Gasparini, M., Serrano, D., Del Piero, D., Thorne, A. W., Cary, P. D., Crane-Robinson, C., (1983), *Eur. J. Biochem.*, *136*, 509-516.
41. Reeves, R., (1992), *Curr. Opin. Cell Biol.*, *4*, 413-423.
42. Bustin, M., Lehn, D. A., Landsman, D., (1990), *Biochim. Biophys. Acta*, *1049*, 231-243.
43. Grosschedl, R., Giese, K., Pagel, J., (1994) *Trends Genet.*, *10*, 94-100.
44. Lilley, D. M. J., (1992) *Nature*, *357*, 282-283.
45. Kohlstaedt, L. A. and Cole, R. D., (1994) *Biochemistry*, *33*, 570-575.

46. Read, C. M., Cary, P. D., Preston, N. S., Lnenicek-Allen, M., Crane-Robinson, C., (1994) *EMBO J.*, *13*, 5639-5646.
47. Stroumbakis, N. D. & Tolias, P. P., (1994) *Biochim. Biophys. Acta*, *1218*, 245-249.
48. Brawley, J. V. & Martinson, H. G., (1992) *Biochemistry*, *31*, 364-370.
49. Tremethick, D. J. & Drew, H. R., (1993) *J. Biol. Chem.*, *268*, 11389-11393.
50. Crippa, M. P., Trieschmann, L., Alfonso, P. J., Bustin, M., (1993) *EMBO J.*, *12*, 3855-3864.
51. Barratt, M. J., Hazzalin, C. A., Zhelev, N., Mahadevan, L. C., (1994) *EMBO J.*, *13*, 4524-4535.
52. Lund, T., Holtlund, J., Fredriksen, M., Laland, S. G., (1983) *FEBS Lett.*, *152*, 163-167.
53. Giancotti, V., Berlingieri, M. T., Di Fiore, P. P., Fusco, A., Vecchio, G., Crane-Robinson, C., (1985) *Cancer Res.*, *45*, 6051-6057.
54. Goodwin, G. H., Cockerill, P. N., Kellam, S., Wright, C. A., (1985) *Eur. J. Biochem.*, *149*, 47-51.
55. Strauss, F. & Varshavsky, A., (1984) *Cell*, *37*, 889-901.
56. Elton, T. S. & Reeves, R., (1985) *Anal. Biochem.*, *144*, 403-413.
57. Lund, T., Holtlund, J., Laland, S. G., (1985) *FEBS Lett.*, *180*, 275-279.
58. D'Anna, J., Becker, R. R., Tobey, R. A., Gurley, L. R., (1983), *Biochim. Biophys. Acta*, *739*, 197-206.
59. Eckner, R. & Birnstiel, M. L., (1989) *Nucleic Acids Res.*, *17*, 5947-5959.
60. Vartiainen, E., Palvimo, J., Mahonen, A., Linnala-Kankkunen, A., Mäenpää, P. H., (1988) *FEBS Lett.*, *228*, 45-48.
61. Lundberg, K., Karlson, J. R., Ingebrigtsen, K., Holtlund, J., Lund, T., Laland, S. G., (1989) *Biochim. Biophys. Acta*, *1009*, 277-279.
62. Johnson, K. R., Lehen, D. A., Elton, T. S., Barr, P. J., Reeves, R., (1988) *J. Biol. Chem.*, *263*, 18338-18342.
63. Johnson, K. R., Disney, J. E., Wyatt, C. R., Reeves, R., (1990) *Exp. Cell Res.*, *187*, 69-76.
64. Giancotti, V., Pani, B., D'Andrea, P., Berlingieri, M. T., Di Fiore, P. P., Fusco, A., Vecchio, G., Philip, R., Crane-Robinson, C., Nicholas, R. H., Wright, C. A., Goodwin, G. H., (1987) *EMBO J.*, *6*, 1981-1987.
65. Lund, T., Skålhegg, B. S., Holtlund, J., Blomhoff, H. K., Laland, S. G., (1987) *Eur. J. Biochem.*, *166*, 21-26.
66. Yang-Yen, H. F. & Rothblum, L. I., (1988) *Mol. Cell. Biol.*, *8*, 3406-3414.

67. Solomon, M., Strauss, F., Varshavsky, A., (1986) Proc. Natl. Acad. Sci. USA, 83, 1276-1280.
68. Reeves, R., Elton, T. S., Nissen, M. S., Lehn, D. A., Johnson, K. R., (1987) Proc. Natl. Acad. Sci. USA, 84, 6531-6535.
69. Wegner, M., Zastrow, G., Klavinius, A., Schwender, S., Müller, F., Luksza, H., Hoppe, J., Wienberg, J., Grummt, F., (1989) Nucleic Acids Res., 17, 9909-9932
70. Reeves, R. & Nissen, M. S., (1990) J. Biol. Chem., 265, 8573-8582.
71. Elton, T. S. & Reeves, R., (1986) Anal. Biochem., 157, 53-62.
72. Palvimo, J. & Linnala-Kankkunen, A., (1989) FEBS Lett., 257, 101-104.
73. Lund, T., Dahl, K. H., Mork, E., Holtlund, J., Laland, S. G., (1987) Biochem. Biophys. Res. Commun., 171, 342-347.
74. Lund, T. & Laland, S. G., (1990) Biochem. Biophys. Res. Commun., 171, 342-347.
75. Goodwin, G. H. & Bustin, M., in *Architecture of Eukaryotic Genes*, Ed. G. Kahl, VCH, (1988), 187-205.
76. Giacotti, V., Buratti, E., Perissin, L., Zorzet, S., Balmain, A., Portella, G., Fusco, A., Goodwin, G. H., (1989) Exp. Cell Res., 184, 538-545.
77. Giacotti, V., Bandiera, A., Buratti, E., Fusco, A., Marzari, R., Coles, B., Goodwin, G. H., (1991) Eur. J. Biochem., 198, 211-216.
78. Giacotti V. & Goodwin, G. H., (1986) J. Biochem. Biophys. Methods, 12, 265-269.
79. Chomezynski, P. & Sacchi, N., (1987) Anal. Biochem., 162, 156-159.
80. Sanger, F., Nicklen, S., Coulion, A. R., (1988), Proc. Natl. Acad. Sci. USA, 74, 5463.
81. Manfioletti, G., Giacotti, V., Bandiera, A., Buratti, E., Sautière, P., Cary, P., Crane-Robinson, C., Coles, B., Goodwin, G. H., (1991) Nucl. Acids Res., 19, 6793-6797.
82. Giacotti, V., Buratti, E., Santucci, A., Neri, P., Crane-Robinson, C., (1992) Biochim. Biophys. Acta, 1119, 296-302.
83. Luscher, B., Christenson, E., Litchfield, D. W., Krebs, E. G., Eisenmann, R. N., (1990) Nature, 344, 517-522.
84. Ferranti, P., Malorni, A., Marino, G., Pucci, P., Goodwin, G. H., Manfioletti, G., Giacotti, V., (1992) J. Biol. Chem., 267, 22486-22489.
85. Reeves, R., Langan, T. A., Nissen, M. S., (1991) Proc. Natl. Acad. Sci. USA, 88, 1671-1675.
86. Thoma, F., in *Architecture of Eukaryotic Genes*, Ed. G. Kahl, VCH, (1988), 163-185.
87. Fashena, S. J., Reeves, R., Ruddle, N. H., (1992) Mol. Cell. Biol., 12, 894-903.

88. Skalnik, D. G. & Neufeld, E. J., (1992) *Biochem. Biophys. Res. Commun.*, *187*, 563-569.
89. Thanos, D. & Maniatis, T., (1992) *Cell*, *71*, 777-789.
90. Du, W., Thanos, D., Maniatis, T., (1993) *Cell*, *74*, 887-898.
91. Linder, H. & Hellinger, W., (1990) *Chromatographia*, *30*, 518-522.
92. Lennox, R. W., Oshima, R. G., Cohen, L. H., (1982) *J. Biol. Chem.*, *257*, 5183-5189.
93. Linder, H., Hellinger, W., Puschendorf, B., (1990) *Biochem. J.*, *269*, 359-363.
94. Lennox, R. W. & Cohen, L. H., (1983) *J. Biol. Chem.*, *258*, 262-268.
95. Tchouatcha-Tchouassom, J. C., Julliard, J. H., Roux, B., (1989) *Biochim. Biophys. Acta*, *1009*, 121-128.
96. Langan, T. A., (1982) *J. Biol. Chem.*, *257*, 14835-14846.
97. Aijro, K., Shibata, K., Nishikawa, Y., (1990) *J. Biol. Chem.*, *265*, 6494-6500.
98. Disney, J. E., Johnson, K. R., Magnuson, N. S., Sylvester, S. R., Reeves, R., (1989) *J. Cell Biol.*, *109*, 1975-1982.
99. Laemmli, U. K., Käs, E., Poljak, L., Adachi, Y., (1992) *Curr. Opin. Genet. Dev.*, *2*, 275-285.
100. Zhao, K., Käs, E., Gonzalez, E., Laemmli, U. K., (1993) *EMBO J.*, *12*, 3237-3247.
101. Reeves, R. and Nissen, M. S., (1993) *J. Biol. Chem.*, *268*, 21137-21146.
102. Bussemakers, M. J. G., van de Ven, W. J. M., Debruyne, F. M. J., Schalken, J. A., (1991) *Cancer Res.*, *51*, 606-611.
103. Ram, T. G., Reeves, R., Hosick, H. L., (1993) *Cancer Res.*, *53*, 2655-2660.
104. Johnson, K. R., Cook, S. A., Davisson, M. T., (1992) *Genomics*, *12*, 503-509.
105. Friedmann, M., Holth, L. T., Zoghbi, H. Y., Reeves, R., (1993) *Nucl. Acids Res.*, *21*, 4259-4267.
106. Tkachuk, D. C., Kohler, S., Cleary, M. L., (1992) *Cell*, *71*, 691-700.
107. Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. M., Canaani, E., (1992) *Cell*, *71*, 701-708.
108. Patel, U. A., Bandiera, A., Manfioletti, G., Giancotti, V., Chau, K. Y., Crane-Robinson, C., (1994) *Biochem. Biophys. Res. Commun.*, *201*, 63-70.
109. Berilingieri, M. T., Manfioletti, G., Santoro, M., Bandiera, A., Visconti, R., Giancotti, V., Fusco, A., (1995) *Mol. Cell. Biol.*, in corso di stampa.

110. Chiappetta, G., Bandiera, A., Berlingieri, M. T., Visconti, R., Manfioletti, G., Battista, S., Martinez-Tello, F., Santoro, M., Giacotti, V., Fusco, A., (1995) *Oncogene*, in corso di stampa.

111. Ausio, J., (1992) *Mol. Cell. Biochem.*, *115*, 163-172.

112. Carlos, S., Hunt, D. F., Rocchini, C., Arnott, D. P., Ausio, J., (1993) *J. Biol. Chem.*, *268*, 195-199.

Comparison of multiple forms of the high mobility group I proteins in rodent and human cells

Identification of the human high mobility group I-C protein

Vincenzo GIANCOTTI¹, Antonella BANDIERA¹, Emanuele BURATTI¹, Alfredo FUSCO², Roberto MARZARI³, Brian COLES⁴ and Graham H. GOODWIN⁵

¹ Università di Trieste, Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole, Trieste, Italy

² Centro di Endocrinologia ed Oncologia Sperimentale del C. N. R., Università di Napoli, II Facoltà di Medicina e Chirurgia, Dipartimento di Biologia e Patologia Cellulare e Molecolare, Napoli, Italy

³ Università di Trieste, Dipartimento di Biologia, Trieste, Italy

⁴ CRC Molecular Toxicology Research Group, Middlesex Hospital, London, England

⁵ Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London, England

(Received September 17/December 6, 1990) – EJB 90 1114

The class I of the high mobility group (HMG) proteins is formed by phosphoproteins which are associated with AT-rich DNA sequences in the nucleus. Three HMGI proteins have previously been described in proliferating rodent cells (HMG Y, HMG I and HMGI-C). All three proteins exhibit microheterogeneity. The microheterogeneity of mouse HMG Y has been investigated in detail and shown to be due to phosphorylation of the protein which is sensitive to alkaline-phosphatase treatment. HMG I is similarly modified. Human cells have up to now only been found to contain HMG Y and HMG I. A search for the third protein, HMGI-C, in human cells was carried out and the protein was found in a hepatoma cell line, but not in normal or transformed T-cells. This HMGI-C protein was found to be modified by phosphorylation, part of which was found to be phosphatase insensitive. An unexpected additional finding in this study was that human cells contain two HMGI7 proteins which differ in their N-terminal primary sequences.

The HMG proteins are an abundant family of non-histone proteins found in the cell nucleus of vertebrate and invertebrate organisms [1, 2]. In mammalian, avian and fish cells there are three classes of such proteins, all characterised by having a high content of basic and acidic amino acids.

The HMGI/2 class of proteins. These are highly structured proteins containing two so-called HMG boxes which are homologous to a number of transcription factors and other regulatory molecules and are involved in DNA-binding. There are multiple forms of these proteins in all cell types though the distribution of subfractions can differ between cells. The function of these proteins is not known, but there is a strong possibility that they are transcription factors, given their similar structure to ribosomal-gene-transcription factors UBF [3].

The HMGI4/17 class. This class of random-coiled proteins is found at slightly higher levels (twofold) in chromatin fractions containing transcriptionally active genomes, but their function is still not known [1, 2].

The HMGI class of proteins. Two such proteins HMG I and HMG Y, were originally described in HeLa cells by Lund et al. [4] and were later shown to be present in rodent and simian cells [5, 6]. The cDNAs for these two proteins were

first cloned by Johnson et al. [7, 8] and shown to be alternatively spliced products of one gene, HMG I differing from HMG Y by the addition of an extra 11 amino acids. In a study of rat cells transformed by oncoviruses or oncogenes [2, 9–11], some of us identified two proteins with amino acid compositions similar to HMG I and HMG Y. These two proteins were termed HMGI-E and HMGI-D. From their electrophoretic mobilities, HMGI-E probably corresponds to HMG Y and HMGI-D to HMG I identified by Lund et al. In addition, a third protein was found in rat cells, HMGI' or HMGI-C, which had not been previously described [2, 9–11]. All three HMGI proteins were found to be induced as a result of neoplastic transformation. A further complexity of HMGI proteins arose from HPLC fractionation of rodent proteins in that multiple forms were observed [11, 12]. In this report we investigate this heterogeneity in rodent and human cells. This report shows that HMGI-C is also present in human hepatoma cells but not in normal or transformed human T-cells.

The HMGI proteins bind specifically to AT-rich DNA sequences and have been shown to be associated with such sequences *in vivo* [13–17]. This has led to the suggestion that they could be chromosomal structural proteins [13]. However, recently it has been shown that HMGI proteins bind specifically to the AT-rich octamer sequence associated with a number of promoters and also to AT-rich regulatory elements of the ribosomal genes [15, 16]. The finding that they can activate transcription *in vitro* [16] and that they have an acidic C-terminal region which is phosphorylated suggests that they

Correspondence to V. Giancotti, Università di Trieste, Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole, Via A. Valerio 22, I-34127 Trieste, Italy

Abbreviations. HMG, high mobility group; PBL, peripheral blood lymphocytes.

Enzyme. Alkaline phosphatase (E.C. 3.1.3.1).

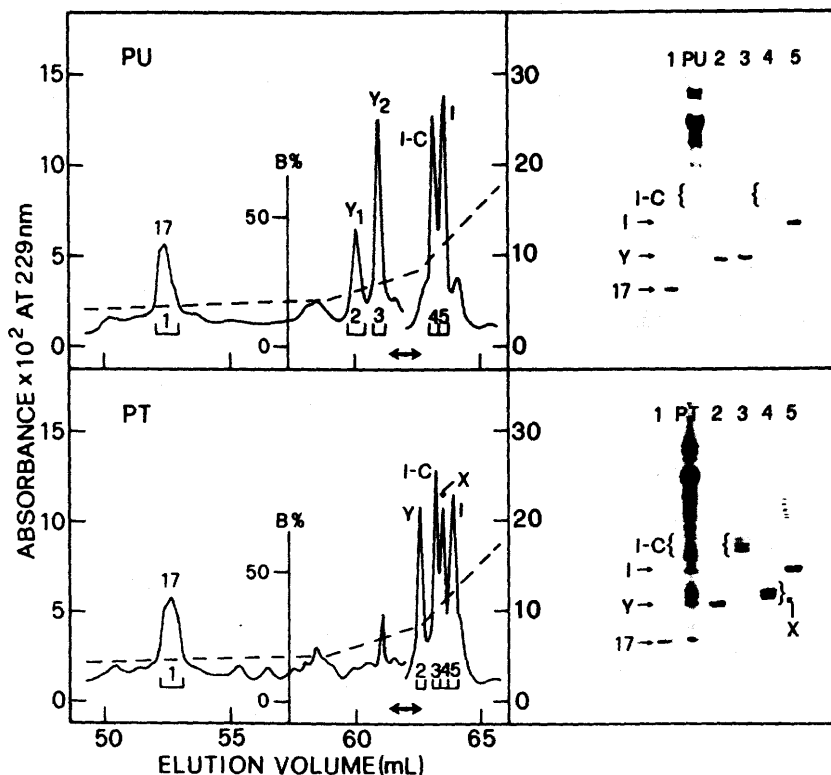


Fig. 1. Reverse-phase HPLC elution profiles of perchloric-acid-extracted proteins from Lewis-lung carcinoma, phosphatase untreated (PU) and phosphatase treated (PT). On the right of the figure the acetic acid/urea gel electrophoreses of the numbered fractions are shown. The scale in the middle of the figure and the dotted line refer to the gradient of solvents: (A) 0.1% (by vol.) trifluoroacetic acid in water, (B) water/ acetonitrile/trifluoroacetic acid (5:95:0.1, by vol.). Gels were stained with Coomassie blue. 17, HMG17; Y₁, HMG Y₁; Y₂, HMG Y₂; I-C, HMGI-C; I, HMG I

may function as transcriptional activators as described by Ptashne et al. [18].

EXPERIMENTAL PROCEDURES

Cell Culture

PLC/PRF/5 cells were grown in Dulbecco's modified Eagles medium supplemented with 10% total calf serum (GIBCO), 100 units/ml penicillin and 0.1 mg/ml streptomycin.

Human T-cell lymphoma/leukaemia cells (MOLT-4 and SUPT-1) were grown in RPMI 1640 medium supplemented with 10% heat-inactivated calf serum and antibiotics as above.

³²P-labelling was carried out as described elsewhere [9].

Peripheral blood lymphocytes, purification and phytohemagglutinin stimulation

Peripheral blood lymphocytes (PBL) were obtained from healthy donors by platelet pheresis and purified by Ficoll centrifugation. Stimulation was carried out by adding 10 µg/ml phytohemagglutinin (FLOW) for 72 h in RPMI 1640 medium.

Protein extraction

HMG and histone H1 were selectively extracted from culture cells and tumours (Lewis-lung carcinoma) with 5% (mass/vol.) perchloric acid and acetone precipitated [9–12].

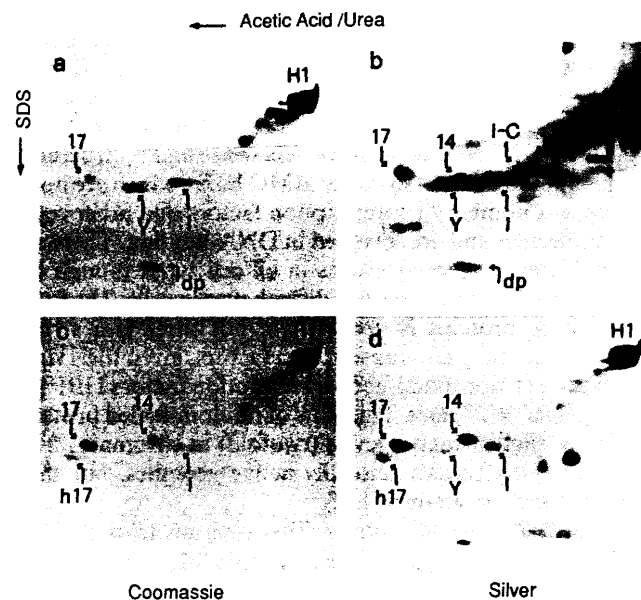


Fig. 2. Two-dimensional gel electrophoresis (acetic acid/urea and SDS) of perchloric-acid-extracted proteins from Lewis-lung carcinoma (a and b) and phytohemagglutinin-stimulated PBL (c and d). The region comprising low-molecular mass HMG proteins and histone H1 is shown. The band marked dp is an unknown degradation peptide which migrates close to HMG Y in the first acetic acid/urea dimension. Both Coomassie blue and silver staining are shown. Abbreviations as in Fig. 1; H1, histone H1; 14, HMG14

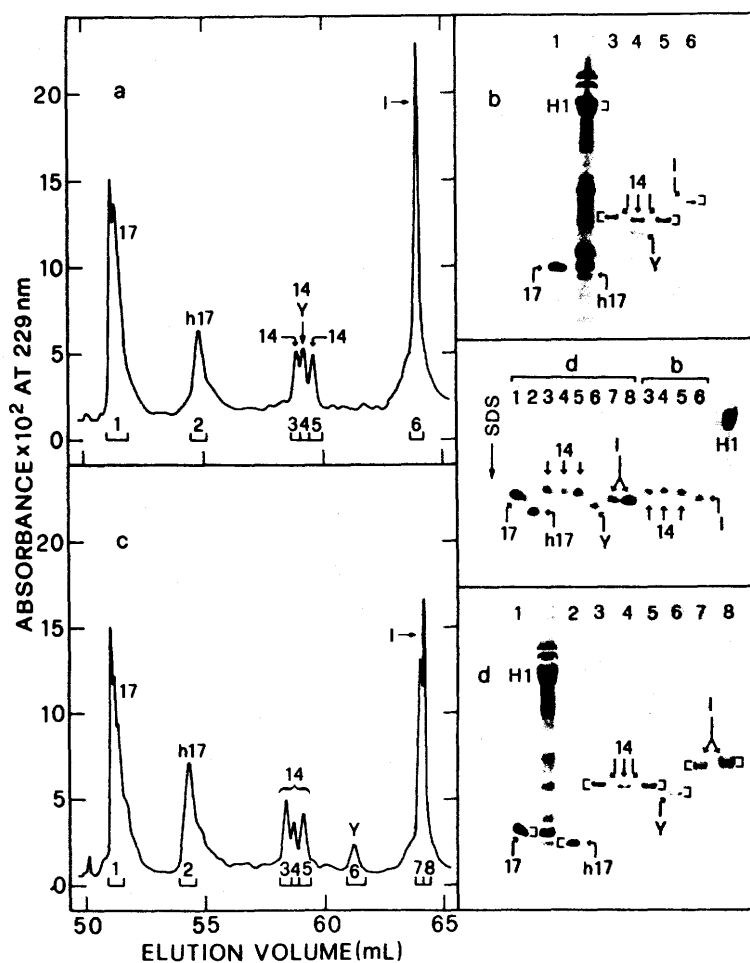


Fig. 3. Reverse-phase HPLC elution profiles of perchloric-acid-extracted proteins from quiescent (a) and phytohemagglutinin-stimulated (c) PBL. On the right of the figure the acetic acid/urea gel electrophoreses of the numbered fractions are shown (b and d). The inset between b and d shows the second dimension SDS electrophoresis of the bands excised from the acetic acid/urea gel as shown by the square brackets; a portion of the histone H1 band was used as a reference. HPLC elution as in Fig. 1. Gels were stained with Coomassie blue. Abbreviations as in Figs 1 and 2; h17, HMGl17

Polyacrylamide gel electrophoresis

First-dimension PAGE (250 mm \times 0.7 mm) was carried out essentially as described [9]. Second-dimension SDS/gel electrophoresis [250 mm \times 1 mm; 15% acrylamide (mass/vol.); acrylamide/bis-acrylamide, 75:1] was carried out using the following solvents: running gel, 0.75 M Tris/HCl, pH 8.8, containing 0.1% (mass/vol.) SDS mixed in the ratio 5:1 (by vol.) with 0.5 M Tricine, pH 8.1, containing 0.1% (mass/vol.) SDS; reservoirs, 25 mM Tricine containing 0.1% SDS (mass/vol.) brought to pH 8.1 by the use of solid Tris base. Bands excised from the first-dimension gels were soaked for 30 min in 0.75 M Tris/1% (mass/vol.) SDS before setting with fresh polyacrylamide. Separation was at a constant 6 V/cm for 15 h.

First- and second-dimension gels were fixed and stained by 0.1% (mass/vol.) Coomassie blue R250 in methanol/water/acetic acid (5:4:1, by vol.). Destaining was carried out by water/ethanol/acetic acid (12.4:6:1.6, by vol.). Silver staining was carried out essentially by the method reported previously [19]. For autoradiography, gels were destained in water/ethanol (3:1, by vol.) and vacuum dried.

High performance liquid chromatography

Reverse-phase HPLC protein fractionation was performed on a Bio-Rad RP-304 column using a Waters apparatus. Pro-

tein (1–3 mg) was eluted with a water/acetonitrile gradient containing 0.1% (by vol.) trifluoroacetic acid.

Alkaline-phosphatase treatment

Protein was dissolved in 0.2 M Tris/HCl containing 1 mM Zn^{2+} and 10 mM Mg^{2+} , pH 8.0, (30 g/l) in the presence of phenylmethylsulfonyl fluoride and treated at 37°C with alkaline phosphatase (3 units/mg protein). The reaction was allowed to proceed for 1 h, however, longer times were also used (up to 14 h). Two different enzyme samples were used (Sigma P-405 and Boehringer 567744). After phosphatase treatment, samples for electrophoretic analyses were prepared by dilution (2:1, by vol.) with a solution of 2 M acetic acid/5 M urea/30% (mass/vol.) sucrose and directly used. Samples for HPLC separation were dialysed against 1 mM HCl (membrane cut-off 3.5 kDa) and the protein recovered by lyophilization.

Protein sequence analysis

The proteins were sequenced by an automated Edman degradation using an Applied Biosystems gas-phase sequenator (470A). Final phenylthiohydantoin derivatives were analysed by reverse-phase HPLC.

RESULTS

Reverse-phase HPLC of HMG proteins from rodent cells reveals a complex mixture of proteins which in our laboratories have been characterised by a combination of amino acids analysis, two-dimensional electrophoretic analysis and protein sequencing [9–12 and unpublished data]. Fig. 1 shows elution profiles of the proteins before and after phosphatase treatment. Peak 1 of the phosphatase-untreated sample corresponds to HMG17, peak 2 and 3 (Y_1 and Y_2) according to peptide sequences are phosphorylated forms of the protein HMG Y described by Lund et al. [20]. Peak 4 corresponds to the protein HMGI-C (or HMGI') first described in transformed rodent cells [9–11]. Sequence analysis of peptides obtained from this protein shows that it is related to, but different from, HMG Y and HMG I. The full sequence of the cloned gene will be described elsewhere. A sequence study confirms that the protein of peak 5 corresponds to the protein HMG I described by Lund et al. [20–22]. All three proteins are phosphoproteins. Alkaline-phosphatase treatment of the proteins results in the two peaks of HMG Y disappearing and a single peak of HMG Y eluting further back in the gradient (Fig. 1), suggesting that there are two phosphorylated forms of HMG Y which are both converted to a dephosphorylated one. HMG I from metaphase cells is known to be phosphorylated at threonine residues 53 and 78. Moreover it was shown that growth-associated histone H1 kinase can phosphorylate threonine 78 *in vitro* [21]. The existence in an unsynchronised culture of two forms of HMG Y which probably represent different phosphorylated variants was also shown [23]. In this study, the electrophoretic analyses of the phosphorylated and dephosphorylated forms of HMG Y in acetic acid/urea gels show that with increasing phosphorylation the protein migrates faster. This anomalous migration of the phosphorylated forms is probably due to intramolecular salt-bridges between phosphate groups and basic amino acids, since addition of 2 mM guanidinium hydrochloride prevents this differential migration (data not shown). HMG I is similarly modified by phosphorylation but the phosphorylated forms are not adequately separated by reverse-phase HPLC. Peak X is a histone H1 degradation product.

Previous data from our laboratories have also shown that HMGI-C is a phosphoprotein [9, 10]. At least three forms are seen on Coomassie-blue-stained acetic acid/urea gels; these forms are still present following phosphatase treatment (Fig. 1) suggesting protein modifications other than alkaline-phosphatase-sensitive phosphorylation. HMGI-C to date has only been shown to be present in mouse and rat cells [9–12]. Lund et al. [20] have described a protein, HMG Im, in metaphase arrested HeLa cells, migrating lower than HMG I which could be the human equivalent of HMGI-C, but they demonstrated by phosphatase treatment that it is a highly phosphorylated metaphase-specific form of HMG I. Thus the status of HMGI-C in species other than rat or mouse was not certain. We undertook a comparative analysis of the sub-fractions of the HMGI proteins in normal and human PBL before and after stimulation with phytohemagglutinin. Fig. 2 shows the two-dimensional electrophoretic analysis of the HMGI proteins from phytohemagglutinin-stimulated cells (Fig. 2c and d) in comparison with HMGI proteins from Lewis-lung carcinoma (Fig. 2a and b). The PBL sample shows HMG I and HMG Y (and HMG14, HMG17 and histone H1). No band is seen in the HMGI-C position.

Reverse-phase HPLC of PBL proteins is shown in Fig. 3a and c. Acetic acid/urea gels of the eluted peaks are shown in

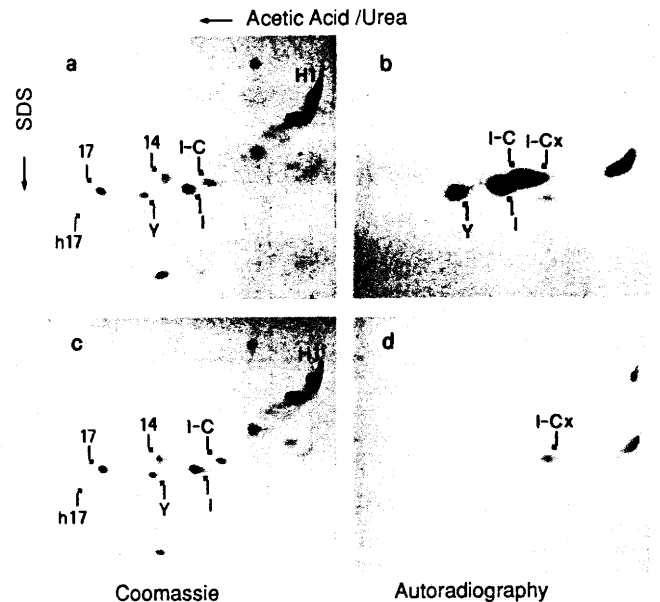


Fig. 4. Two-dimensional gel electrophoresis (acetic acid/urea and SDS) of perchloric-acid-extracted proteins from PLC/PRF/5 cells untreated (a and b) and treated (c and d) with alkaline phosphatase. The region comprising low-molecular mass HMG proteins and histone H1 is shown. Both Coomassie blue staining (a and c) and ^{32}P autoradiography (b and d) are shown. Abbreviations as in Figs 1–3; I-Cx, HMGI-Cx

Fig. 3b and d; the inset between b and d shows the second-dimension SDS analysis of the bands excised from the acetic acid/urea gels. A small amount of HMG Y in the unstimulated PBL elutes with multiple HMG14 peaks. The HMG Y protein of the stimulated PBL elutes later from the column (possibly due to a change of phosphorylation). No HMGI-C was seen in either quiescent or stimulated cells. The peak h17 eluting behind HMG17 is a variant form of HMG17 (see below). Similar analyses of human T-cell lymphoma and leukaemia cells (MOLT-4 and SUPT-1) also failed to reveal the presence of HMGI-C. However, when the human hepatoma cell line PLC/PRF/5 was analysed by electrophoresis (Fig. 4) and HPLC (Fig. 5), a HMGI-C protein was seen. Fig. 4 (a and b) shows a phosphorylated band marked I-C and I-Cx migrating just behind HMG I in the HMGI-C position. Phosphatase treatment removes most of the radioactivity in the HMGI-C position. However, the I-Cx spot remains phosphorylated (even after a 14-h phosphatase treatment); since it migrates just behind the major Coomassie-blue-stained spot I-C, it is possible that I-Cx is a phosphorylated form of HMGI-C which is not affected by alkaline-phosphatase treatment (possibly a poly(ADP)ribosylated form). The HPLC elution profile (Fig. 5a) shows two forms of HMG Y as in mouse and HMGI-C, eluting ahead of HMG I at about the same position as in the mouse HMGI-C. Multiple forms of the purified human HMGI-C are seen on the acetic acid/urea electrophoretic-gel analysis (Fig. 5b) and they migrate in the same position as the mouse HMGI-C. In order to confirm the conclusions inferred from the HPLC and acetic acid/urea gel, the bands excised from the first-dimension analysis were examined in second-dimension SDS electrophoresis (Fig. 5c).

These results, therefore, show that human hepatoma cells contain an HMG protein with similar characteristics as the rodent HMGI-C. In addition, analysis of human cells shows

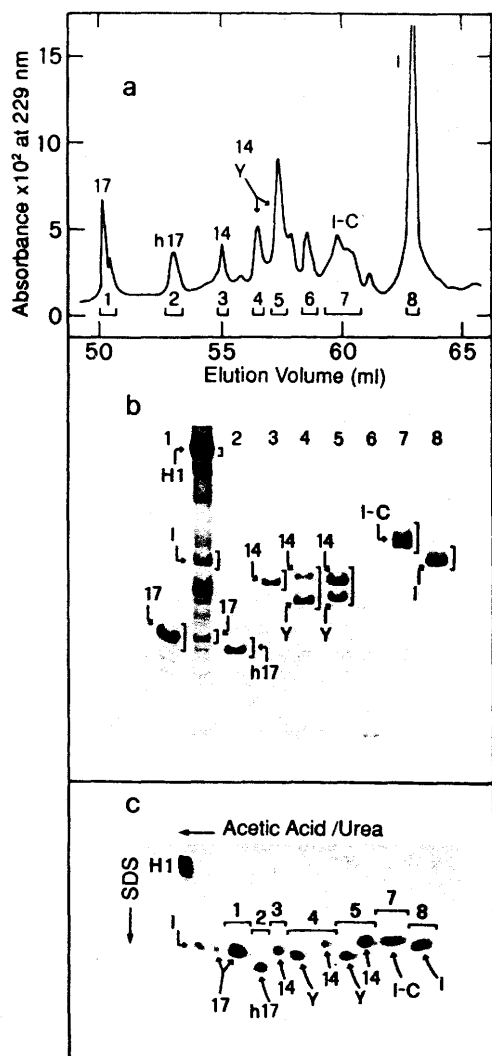


Fig. 5. Reverse-phase HPLC elution profile (a) of perchloric-acid-extracted proteins from PLC/PRF/5 cells. The acetic acid/urea gel electrophoresis of the numbered fractions is shown (b) together with the second-dimension SDS electrophoresis (c) of the bands excised from the acetic acid/urea gel as shown by the square brackets (histone H1 as reference). HPLC elution as in Fig. 1. Gels were stained with Coomassie blue. Abbreviations as in Figs 1–3

that there are two forms of HMG17. These two forms, HMG17 and h17, from the PBL cells were sequenced from the N-terminal end (Fig. 6). One of the proteins, HMG17, differs from calf thymus HMG17 [24] at positions 25 and 26, and from the human cDNA sequence [25] also at position 4. The other protein, h17, has the same N-terminal sequence as human HMG17 (this work) but differs at residue 6.

DISCUSSION

Reverse-phase HPLC of HMG1 proteins from transformed rat or mouse cells originally [9–11] showed that there are two proteins related to the HMG I/Y described by Lund et al., in HeLa cells [4]. In addition, a third protein HMG1-C (or HMG1'), which is usually present at high levels in transformed or undifferentiated cells was described [9–11]. It was also found that the various HMG1 proteins exhibited multiple peaks on chromatography [9, 12]. In this paper we

cDNA human HMG17 [25]:

1 10 20
P-K-R-N-A-E-G-D-A-K-G-D-K-A-K-V-K-D-E-P-Q-R-R-S-A-R-L-

Human HMG17 (this work):

1 10 20
P-K-R-K-A-E-G-D-A-K-G-D-K-A-K-V-K-D-E-P-Q-R-R-S-S-A-L-

h17 (this work):

1 10 20
P-K-R-K-A-K-G-D-A-K-G-D-K-A-K-V-K-D-E-P-Q-R-R-S-S-A-L-

Calf thymus HMG17 [24]:

1 10 20
P-K-R-K-A-E-G-D-A-K-G-D-K-A-K-V-K-D-E-P-Q-R-R-S-A-R-L-

Fig. 6. Comparison between HMG17 sequences

show that the heterogeneity of the HMG I and HMG Y proteins is probably due to phosphorylation. A phosphorylation site in the acidic C-terminal end of both HMG I and HMG Y has been mapped by several groups [21–23], but there may be additional phosphorylation sites since the proteins are very rich in serine and threonine residues. The C-terminal phosphorylation site is similar to the casein-kinase-II type site observed in another nuclear protein, the c-myc proto-oncogene [26]. This site is lost in the viral transforming protein v-myc and it has been postulated this may result in both deregulated DNA-binding and transcriptional activity of the viral oncogene. This may lead to deregulated growth of cells transformed with this oncogene. The phosphorylation of HMG I proteins may similarly be involved in cell growth. In this paper we show that the multiple forms of HMG Y are due to phosphorylation. One form is presumably phosphorylated at the C terminal and at an internal site, and the second form only at one site. HMG I is the same as HMG Y but has an additional 11 amino acids inserted near the N terminal [7, 8, 15]. HMG I is therefore probably phosphorylated in the same manner, though resolution of multiple forms by HPLC was not achieved in this study. In addition to the interphase phosphorylation studied here, HMG I is further phosphorylated at the metaphase by the p34 CDC2 kinase at threonine residues 53 and 78 [21]. These phosphorylated forms have not been observed in this study.

Protein HMG1-C from rodent cells can be distinguished from the HMG I/Y pair by a number of criteria. Its protein sequence is different (unpublished results) and, although it is a phosphorylated protein as HMG I/Y, a small fraction of it has phosphate groups which are phosphatase resistant. Since there is no tyrosine, it is likely that the protein is modified by poly(ADP) ribosylation. Moreover this report shows that HMG1-C is not rodent specific being present in human hepatoma cells. Rather surprisingly, it is absent in normal human and leukaemia cells.

HMG1 proteins from the mammalian and *Drosophila* cells, are known to bind AT-rich DNA sequences [13, 14, 17, 27] and have been shown to be localised to such sequences in the cell nucleus. Thus a structural role for these proteins in organising satellite and centromeric chromatin has been

implied. However, it is also known that these proteins are distributed throughout the chromosomes and may play additional roles [27]. For example, evidence has been presented that the proteins may be involved in transcription of the ribosomal genes. The fact that they are such abundant proteins ($10^5 - 10^6$ molecules/cell) would indicate that they could be involved in activating many genes, some possibly involved in cell growth.

This research was funded by grants from the *Associazione Italiana per la Ricerca sul Cancro* (AIRC), Milano, Italy; *Consiglio Nazionale delle Ricerche* (CNR), Roma, Italy; *Ministero dell'Università e della Ricerca Scientifica e Tecnologica*, Roma, Italy; *Progetto Finalizzato Biotecnologia e Biostrumentazione*, CNR, Roma, Italy; Cancer Research Campaign, U.K.

REFERENCES

1. Johns, E. W. (1982) *The HMG chromosomal proteins*, Academic Press, London.
2. Goodwin, G. & Bustin, M. (1988) in *Architecture of eukaryotic genes* (Kahl, G., ed.) pp. 187–205, VCH, Weinheim.
3. Jantzen, H., Admon, A., Bell, S. P. & Tjian, R. (1990) *Nature* **344**, 830–836.
4. Lund, T., Holtlund, J., Fredriksen, M. & Laland, S. G. (1983) *FEBS Lett.* **152**, 163–167.
5. Lund, T., Holtlund, J. & Laland, S. G. (1985) *FEBS Lett.* **180**, 275–279.
6. Strauss, F. & Varshavsky, A. (1984) *Cell* **37**, 889–901.
7. Johnson, K. R., Lehn, D. A., Elton, T. S., Barr, P. J. & Reeves, R. (1988) *J. Biol. Chem.* **263**, 18338–18342.
8. Johnson, K. R., Lehn, D. A. & Reeves, R. (1989) *Mol. Cell. Biol.* **9**, 2114–2123.
9. Giaccotti, V., Berlingieri, M. T., Di Fiore, P. P., Fusco, A., Vecchio, G. & Crane-Robinson, C. (1985) *Cancer Res.* **45**, 6051–6057.
10. Giaccotti, V., Pani, B., D'Andrea, P., Berlingieri, M. T., Di Fiore, P. P., Fusco, A., Vecchio, G., Philp, R., Crane-Robinson, C., Nicolas, R. H., Wright, C. A. & Goodwin, G. H. (1987) *EMBO J.* **6**, 1981–1987.
11. Goodwin, G. H., Cockerill, P. N., Kellam, S. & Wright, C. A. (1985) *Eur. J. Biochem.* **149**, 47–51.
12. Giaccotti, V., Buratti, E., Perissin, L., Zorzet, S., Balmain, A., Portella, G., Fusco, A. & Goodwin, G. H. (1989) *Exp. Cell Res.* **184**, 538–545.
13. Solomon, M. J., Strauss, F. & Varshavsky, A. (1986) *Proc. Natl Acad. Sci. USA* **83**, 1276–1280.
14. Wegner, M., Zastrow, G., Klavinus, A., Schwender, S., Muller, F., Luksza, H., Hoppe, J., Wienberg, J. & Grummt, F. (1989) *Nucleic Acids Res.* **17**, 9909–9932.
15. Eckner, R. & Birnstiel, M. L. (1989) *Nucleic Acids Res.* **17**, 5947–5959.
16. Yang-Yen, H. F. & Rothblum, L. I. (1988) *Mol. Cell. Biol.* **8**, 3406–3414.
17. Ashley, C. T., Pendleton, C. G., Wesley Jennings, W., Saxena, A. & Glover, C. V. C. (1989) *J. Biol. Chem.* **264**, 8394–8401.
18. Ptashne, M. (1988) *Nature* **335**, 683–689.
19. Giaccotti, V. & Goodwin, G. H. (1986) *J. Biochem. Biophys. Methods* **12**, 265–269.
20. Lund, T., Skalhogg, B. S., Holtlund, J., Blomoff, H. K. & Laland, S. G. (1987) *Eur. J. Biochem.* **166**, 21–26.
21. Lund, T. & Laland, S. G. (1990) *Biochem. Biophys. Res. Commun.* **171**, 342–347.
22. Lund, T., Dahl, K. H., Mork, E., Holtlund, J. & Laland, S. G. (1987) *Biochem. Biophys. Res. Commun.* **146**, 725–730.
23. Palvimo, J. & Linnala-Kankkunen, A. (1989) *FEBS Lett.* **257**, 101–104.
24. Walker, J. M., Hastings, J. R. B. & Johns, E. W. (1977) *Eur. J. Biochem.* **76**, 461–468.
25. Landsman, D., Soaves, N., Gonzalez, F. J. & Bustin, M. J. (1986) *J. Biol. Chem.* **261**, 7479–7484.
26. Luscher, B., Christenson, E., Litchfield, D. W., Krebs, E. G. & Eisenman, R. N. (1990) *Nature* **344**, 517–522.
27. Disney, J., Johnson, K. R., Magnuson, N. S., Sylvester, S. R. & Reeves, R. (1989) *J. Cell Biol.* **109**, 1975–1982.

cDNA cloning of the HMGI-C phosphoprotein, a nuclear protein associated with neoplastic and undifferentiated phenotypes

G.Manfioletti, V.Giancotti, A.Bandiera, E.Buratti, P.Sautière¹, P.Cary², C.Crane-Robinson², B.Coles³ and G.H.Goodwin⁴

Università di Trieste, Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole, Trieste, Italy, ¹Centre National de la Recherche Scientifique, Institute de la Recherche sur le Cancer, 59045, Lille, France, ²Biophysics Laboratories, Portsmouth Polytechnic, Portsmouth, ³CRC Molecular Toxicology Research Group, Middlesex Hospital, London and ⁴Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, UK

Received October 2, 1991; Revised and Accepted November 20, 1991

EMBL accession no. X58380

ABSTRACT

The HMGI-C protein is a nuclear phosphoprotein expressed at high levels in transformed cells. The cDNA encoding the mouse protein has been isolated and the sequence of the encoded protein shows that it is related to the HMGY and I proteins, proteins which bind in the minor groove of DNA containing stretches of A and T. The HMGI-C protein has three short highly basic domains, an acidic C-terminal domain, and potential CDC2/p34 and casein kinase II phosphorylation sites. Analysis of mRNA levels demonstrate that the HMGI-C gene is not expressed in a variety of mouse tissues but is expressed in Lewis lung carcinoma cells.

INTRODUCTION

The High Mobility Group (HMG) proteins are a class of DNA-binding proteins characterised by their high contents of basic and acid amino acids. Although originally classed together, it is now apparent that the HMG proteins can be grouped into three quite distinct families. The HMGI and 2 type proteins, which have two highly structured DNA-binding 'HMG-boxes' and an acidic C-terminal, have been found to enhance transcription *in vitro* (1,2) and may well be members of a much larger class of regulators containing HMG-boxes (3). The second family of HMG proteins are the random-coil proteins HMG14 and 17 of unknown function (4), and the third family are the HMGI-type proteins, proteins which bind in the minor groove of AT-rich sequences. So far three members of this third family, HMGI, HMGY and HMGI-C have been identified (5-10). (These proteins have also been termed HMG D (or α), E, C (or I') in the earlier literature). Most differentiated tissues and cells have very low levels of the HMGI-type proteins compared to the other HMG proteins; for example, thyroid epithelial cells grown in culture and exhibiting a differentiated phenotype have barely detectable levels of the HMGI-type proteins. However, when such cells are transformed with viral oncogenes high levels of all three HMGI-type proteins

are observed, concomitant with the cells becoming less differentiated (8,9,10). Conversely, when F9 embryocarcinoma cells are induced to differentiate with retinoic acid the expression of the HMGI-type proteins is repressed (11). Thyroid derived fibroblasts and rat-1 fibroblasts grown in culture express HMGI and Y but not HMGI-C. When transformed with viral oncogenes, HMGI-C expression is then induced (9,10). Thus viral oncogenes appear to induce HMGI-C expression, possibly as a result of inducing a less differentiated phenotype.

In the cell nucleus, the HMGI and Y proteins have been localised to AT-rich sequences associated with centromeres and telomeres where they may play a structural role (12). They have also been found associated with other regions of the genome and since HMGI/Y can activate transcription *in vitro* (13) it is possible that, when bound to such regions, they play a role in transcription processes. This is supported by the recent finding that HMGI binds to an AT-rich activator sequence upstream of the murine lymphotoxin gene (S. Fashena, R. Reeves and N. Ruddle, manuscript submitted).

The HMGI and Y proteins have been well characterised; cDNA cloning has shown that they are products of alternatively spliced mRNA species, HMGI having an additional eleven amino acids compared to HMGY (14,15). They have also been shown to be phosphorylated at multiple sites by CDC2/cyclin at metaphase and by casein kinase II in interphase (16-18). Less is known about the third protein HMGI-C except that it is a phosphoprotein with a similar size (~12000 Da) and amino acid composition to HMGI and Y (8-10). It has been found in sarcomas, carcinomas and hepatomas of human, mouse and rat origin, but its expression may be more restricted than HMGI(Y) since it is not expressed in T-cells (8-10,19,20). In this paper we describe the molecular cloning of the cDNA for mouse HMGI-C, demonstrating that HMGI-C, although similar to HMGI/Y, is the product of a different gene. The gene is expressed in Lewis lung carcinoma cells but the mRNA is undetectable in normal mouse tissues.

MATERIALS AND METHODS

Purification of HMGI-C by HPLC

HMG and histone H1 proteins were selectively extracted from mouse Lewis lung carcinoma (19) or tumours obtained from transformed rat thyroid cells (10) with 5% perchloric acid (PCA) and fractionated by reverse phase HPLC as previously described (19) using a BioRad RP-304 column (21.5×250mm) for large scale fractionation (20 mg of crude extract).

Peptide purification and characterization

Mouse HMGI-C (50–70 μ g) was digested with 25 μ g of endoproteinase Glu-C (Boehringer, Mannheim) in 25 mM ammonium carbonate, pH 7.9 at 37°C for 40 h. The peptides were separated on a Waters μ Bondapak C18 column (3.9×300 mm) with a Waters C18 pre-column insert. The elution was carried out at room temperature using a gradient from 0.1% TFA in water (solvent A) to 0.1% TFA in 95% acetonitrile, 5% water (solvent B). Two peptides were obtained which were sequenced by automated Edman degradation using an Applied Biosystems gas-phase sequencer (470A). Phenylthiohydantoin derivatives were analysed by reverse-phase HPLC. Similarly two peptides were isolated and sequenced following digestion of rat HMGI-C with AspN.

RNA preparation and Northern blot analysis

Total RNA from mouse tissues and solid tumours was extracted by disrupting the tissue in lysis buffer (4M guanidinium thiocyanate, 25 mM sodium citrate, 0.1M 2-mercaptoethanol, 0.5% N-laurylsarcosine) using a Polytron homogeniser. RNA was then prepared according to Chomczynski and Sacchi (21). Total cellular RNA (10 μ g) was used for Northern analysis on 1% agarose gels containing 6.7% formaldehyde. Gels were transferred to Genescreen plus membranes (Dupont) according to the suppliers instructions. RNA was crosslinked by exposure to UV-light. Hybridisation was performed in 1M NaCl, 1% SDS in the presence of 100 μ g/ml salmon sperm DNA at 68°C. The probe, labelled by random priming to a specific activity of 5×10^8 cpm/ μ g, was added at 5×10^5 cpm/ml hybridisation solution. For poly(A) selection, mAP paper (Amersham) was used.

cDNA library construction and screening

An oligodT primed cDNA library was made from Lewis lung carcinoma poly(A)⁺ RNA by directional cloning (22) into the lambda vector T7-T3/E-H (23). The library was screened with a degenerate oligonucleotide (41mer) designed on the basis of the mouse HMGI-C peptide sequence KXPQQVVQKKPAQE using the criteria suggested by Lathe (24). Since the second residue was not identified, it was guessed to be a lysine. The sequence of the oligonucleotide was AAGAAGCCCCA(A/G)CA(A/G)GTG GTG CA(A/G)AA(A/G)AA(A/G)CC-TGC(T/C)CA(A/G)GA. 800ng of the oligonucleotide was end-labelled with 150 μ Ci γ -³²P-ATP to a specific activity of 2×10^8 cpm/ μ g. 200,000 plaques of the cDNA library were screened with the probe in a hybridisation solution containing 6×SSC, 0.5% SDS, 100 μ g/ml salmon sperm DNA at 55°C for 36 h. The filters were then washed twice with 2×SSC at room temperature, once with 2×SSC, 0.1% SDS at 45°C, and then once with 0.5×SSC, 0.1% SDS at 37°C, all the washes being for 5 min. The positive plaques were rescreened as above. One

positive was isolated which contained the HMGI-C peptide sequence. The insert from this clone was used to screen the same library to find other positives.

DNA sequencing and sequence analysis

The cDNA was subcloned in the Bluescript KS+ plasmid (Stratagene). Nested deletions were introduced from both ends by ExoIII digestion according to the Bluescript manual. Double-stranded DNA was isolated and sequenced using the EMBL automated sequencing apparatus. The entire sequence was read on both strands and each nucleotide was read three times on average. Nucleotide and amino acid sequences analysis were carried out using the Intelligenetics sequence analysis software package.

RESULTS

Cloning of the cDNA coding for HMGI-C

HMGI-C from mouse Lewis lung carcinoma was purified to homogeneity by reverse phase chromatography. Since the N-terminal was found to be blocked the protein was digested with Glu-C endoproteinase and two peptides were sequenced (shown in Fig.2). Two peptide sequences were also obtained from rat HMGI-C. A 41mer degenerate oligonucleotide was synthesised based on the peptide sequence K[-]PQQVVQKKPAQE, taking into account codon usage in mammals and using mixed bases in the third codon position of some of the residues. This oligonucleotide was used to screen a mouse Lewis lung carcinoma cDNA library. Several positive clones were obtained but only one turned out to be a true positive in the secondary screening. The insert of this clone HMGI-C11 was 1.7 Kb long. Using the DNA of this clone as a probe two more positives were identified, HMGI-C4 and HMGI-C7, containing inserts of 1.4 and 2.3 Kb respectively (Fig.1). Clones HMGI-C4 and HMGI-C7 start 3' of the coding region but extend further downstream than the 3' end of the HMGI-C11 clone. Both have polyA tails at the 3' end but without any polyadenylation signal (data not shown).

The total length of cDNA cloned (from the three clones HMGI-C11, HMGI-C4 and HMGI-C7) is about 3 Kb which is ~1 Kb shorter than the 4 Kb mRNA seen on the Northern (see below). Since HMGI-C4 and HMGI-C7 clones do not have a polyadenylation signal but have oligodA 3' ends, it is likely that oligodT priming in the library construction occurred from internal stretches of polyA, and that the true 3' extremity of the mRNA may be missing from the library.

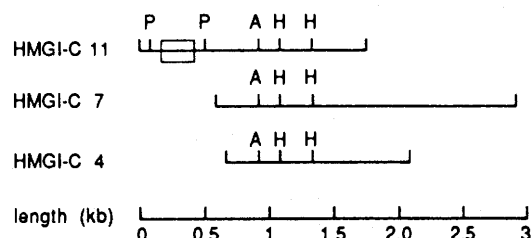


Fig. 1. Partial restriction map of three HMGI-C cDNA inserts. The boxed region indicates the amino acid coding sequence of HMGI-C. The numbers indicate the length in basepairs from the 5' end of clone HMGI-C11. Some restriction sites are shown: P, Pst I; H, Hinc II; A, Acc I.

II. It has also been shown that the CDC2 kinase phosphorylates human HMGI at the threonines 56 and 83 (Fig.3 numbering) *in vitro* (27). The threonine at position 83 of human HMGI is not conserved in murine HMGI, Y or HMGI-C. Mouse HMGI-C contains a serine residue at position 56 which could well be a site for CDC2 phosphorylation (28). From the two peptide sequences P1 and P2 obtained from rat HMGI-C (Fig.2) which span residues 54-69 and 81-102 respectively (using the numbering of Fig.3), it can be seen that the serine at residue 56 is conserved. Like mouse HMGI-C and mouse HMGI/Y the amino acid position 83 is not threonine. Also the threonine at 82 in mouse HMGI-C, which potentially could be a CDC2 phosphorylation site is not conserved in the rat HMGI-C. In conclusion, mouse and rat HMGI-C have potential sites of phosphorylation by casein kinase II and CDC2 as reported for HMGI/Y proteins.

Expression of HMGI-C mRNA in mouse tissues

The expression of HMGI-C mRNA was analysed in tissues extracted from adult Balb/c mice and compared to HMGI/Y expression. As can be seen in Figure 4, there is no detectable expression in all the tissues analysed while it is evident in the RNA sample extracted from Lewis lung carcinoma. This data on mRNA levels agrees with what is known about protein levels of HMGI-C, i.e. in normal differentiated tissues levels are very much lower than in malignant tissues. The same Northern blot was probed with the cDNA coding for HMGY which is able to recognise both HMGY and HMGI since they are very similar. The HMGI/Y mRNA expression level in Lewis lung carcinoma is considerably higher than that of HMGI-C. Furthermore a very low level of expression of HMGI/Y is detectable in some of the mouse tissues.

DISCUSSION

Cloning of the cDNA of the HMGI-C protein reveals that it is structurally homologous to the other two members of this family of proteins, HMGI and HMGY, but is expressed from a different gene. HMGI-C has three conserved basic domains and an acidic C-terminal. The domain II has a conserved motif containing a CDC2 phosphorylation site, and the C-terminal has a potential casein kinase II phosphorylation site. Although a threonine at

position 83 of human HMGI (near domain III) is a second CDC2 phosphorylation site, the mouse HMGY and mouse and rat HMGI-C do not have threonines or serines at this position. There is a threonine at residue 82 in the mouse HMGI-C, but it is not known whether phosphorylation occurs at this position, and it is not present in the rat HMGI-C. Therefore phosphorylation in this region may not be important for protein function.

The HMGI-type proteins bind in the minor groove of DNA, binding preferentially to stretches of DNA containing A and T (25,29). A peptide corresponding to the basic domain II of the HMGI-type protein will on its own bind to AT DNA and a model has been proposed in which such an extended peptide with an Asx turn at one end lies in the minor groove, stabilized by hydrogen bonds and Van der Waals contacts in the minor groove and ionic interaction between the basic side chains and DNA phosphates (25). It is probable all three basic domains, I, II and III bind in the minor groove in the same way. Thus three adjacent short stretches of AT-DNA would bind the HMGI-type proteins with high affinity. It is of interest to note that the spacing between all three basic domains are conserved in HMGY and HMGI-C. The first and second boxes of HMGI are further apart due to an eleven amino acid insertion. Similar repeats are seen in the drosophila D1 protein which also has casein kinase II phosphorylation sites (30).

Phosphorylation of the threonine in the domain II weakens the binding of this peptide to DNA (31). Thus the M-phase phosphorylation of the proteins will weaken DNA binding of the HMGI-type protein (though they still remain attached to the metaphase chromosome). Similarly phosphorylation of the C-terminal by casein kinase II may affect DNA binding as has been shown for the c-myc proto-oncogene (32). It is also possible that the highly acidic C-terminal interacts with other proteins (e.g. transcription factors).

The function of the HMGI-type proteins is not clear—they have been suggested to be chromosome structural proteins (e.g. in organising satellite chromatin (6)) or transcription factors (13,33), and it is possible that they play both roles as in the case of the RAP1 protein which is bound to telomeres as well as acting as a transcription factor (34). Further evidence for a transcriptional role comes from studies on an upstream activator sequence of the murine lymphotoxin gene which binds HMGI protein (Fashena, Reeves and Ruddle, manuscript submitted).

The fact that the levels of the HMGI-type proteins are modulated by the differentiation status of the cells suggests a regulatory role for the proteins. Levels of the proteins do not correlate with cell proliferation since some cells grown in culture (e.g. normal thyroid epithelial cells) express very low levels of all three HMGI proteins compared to the levels found in transformed cells and some immortalized cell lines such as NIH-3T3 (Giancotti *et al.*, unpublished data). One possible explanation for this difference is that in normal cells the HMGI proteins are only present at some limited stage of the cell cycle whilst in transformed cells turnover may be uncoupled from the cell cycle such that the proteins are present throughout the cycle.

Although HMGI-C is similar to HMGI(Y), the expression patterns of the proteins in different cells are different in some cases, suggesting that the two genes are under different regulatory controls. For example, HMGI and Y proteins but not HMGI-C are expressed in rat-1 fibroblasts and FRT-fibroblasts, both normal cell lines (8,10). HMGI-C is only expressed when these cells are transformed with viral oncogenes. Conversely, HMGI-C but not HMGI and Y is expressed in thyroid epithelial cells

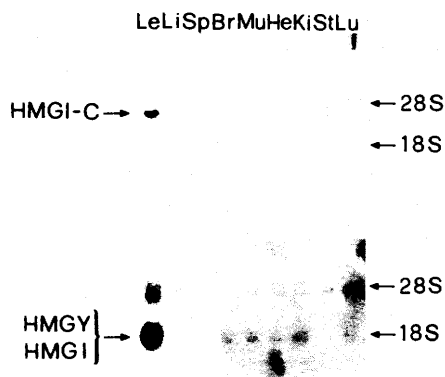


Fig. 4. Northern blot analysis of RNA extracted from Lewis lung carcinoma (Le) and various mouse tissues (Li, liver; Sp, spleen; Br, brain; Mu, muscle; He, heart; Ki, kidney; St, stomach; Lu, lung). Equal amounts (10 μ g) of RNA, estimated by ethidium bromide staining, were loaded.

infected with murine sarcoma virus containing v-Ha-ras (10). We have also found that T-cells express HMGI and Y but not HMGI-C (20). These results would suggest that HMGI-C has a different role to that of HMGI or HMGY.

ACKNOWLEDGEMENTS

We would like to thank Ray Reeves for supplying the HMGI cDNA clone and for helpful discussions, Nancy Ruddle for communicating data on the mouse lymphotoxin gene, Grace Poon for the mass spectrometric analysis of HMGI proteins, and M. Callahan for manuscript preparation. This research was funded by grants from the *Associazione Italiana per la Ricerca sul Cancro* (AIRC), Milano, Italy; *Consiglio Nazionale delle Ricerche* (CNR), Roma, Italy; *Ministero dell'Università e della Ricerca Scientifica e Tecnologica*, Roma, Italy; The Cancer Research Campaign, UK; and the European Economic Community Stimulation Action Programme.

REFERENCES

1. Watt, F. and Molloy, P.L. (1988) *Nucleic Acids Res.*, **16**, 1471-1486.
2. Singh, J. and Dixon, G.H. (1990) *Biochemistry*, **29**, 6295-6302.
3. Jantzen, H., Admon, A., Bell, S.P. and Tjian, R. (1990) *Nature* **344**, 830-836.
4. Goodwin, G. and Bustin, M. (1988) in *Architecture of eukaryotic genes* (Kahl, G., ed.), pp.187-205, VCH, Weinheim.
5. Lund, T., Holtlund, J., Fredriksen, M. and Laland, S.G. (1983) *FEBS Lett.*, **152**, 163-167.
6. Strauss, F. and Varshavsky, A. (1984) *Cell*, **37**, 889-901.
7. Lund, T., Holtlund, J. and Laland, S.G. (1985) *FEBS Lett.*, **180**, 275-279.
8. Goodwin, G.H., Cockerill, P.N., Kellam, S. and Wright, C.A. (1985) *Eur. J. Biochem.*, **149**, 47-51.
9. Giacotti, V., Berlingieri, M.T., Di Fiore, P.P., Fusco, A., Vecchio, G. and Crane-Robinson, C. (1985) *Cancer Res.*, **45**, 6051-6057.
10. Giacotti, V., Pani, B., D'Andrea, P., Berlingieri, M.T., Di Fiore, P.P., Fusco, A., Vecchio, G., Philp, R., Crane-Robinson, C., Nicolas, R.H., Wright, C.A. and Goodwin, G.H. (1987) *EMBO J.*, **6**, 1981-1987.
11. Vartiainen, E., Palvimo, J., Mahonen, A., Linnala-Kankkunen, A. and Maenpaa, P.H. (1988) *FEBS Lett.*, **228**, 45-48.
12. Disney, J., Johnson, K.R., Magnuson, N.S., Sylvester, S.R. and Reeves, R. (1989) *J. Cell. Biol.*, **109**, 1975-1982.
13. Yang-Yen, H.F. and Rothblum, L.I. (1988) *Mol. Cell. Biol.*, **8**, 3406-3414.
14. Johnson, K.R., Lehn, D.A., Elton, T.S., Barr, P.J. and Reeves, R. (1988) *J. Biol. Chem.*, **263**, 18338-18342.
15. Johnson, K.R., Lehn, D.A. and Reeves, R. (1989) *Mol. Cell. Biol.*, **9**, 2114-2123.
16. Lund, T. and Laland, S.G. (1990) *Biochem. Biophys. Res. Commun.*, **171**, 342-347.
17. Lund, T., Dahl, K.H., Mork, E., Holtlund, J. and Laland, S.G. (1987) *Biochem. Biophys. Res. Commun.*, **146**, 342-347.
18. Palvimo, J. and Linnala-Kankkunen, A. (1989) *FEBS Letts.*, **257**, 101-104.
19. Giacotti, V., Buratti, E., Perissin, L., Zorzet, S., Balmain, A., Portella, G., Fusco, A. and Goodwin, G.H. (1989) *Exp. Cell. Res.*, **184**, 538-545.
20. Giacotti, V., Bandiera, A., Buratti, E., Fusco, A., Marzari, R., Coles, B. and Goodwin, G.H. (1991) *Eur. J. Biochem.*, **198**, 211-216.
21. Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.*, **162**, 156-159.
22. Dorschers, L. and Postmes, A. (1987) *Nucleic Acids Res.*, **15**, 3629.
23. Grimaldi, G., Manfioletti, G. and Schneider, C. (1987) *Nucleic Acids Res.*, **15**, 9608.
24. Lathe, A. (1985) *J. Mol. Biol.*, **183**, 1-12.
25. Reeves, R. and Nissen, M.S. (1990) *J. Biol. Chem.*, **265**, 8573-8582.
26. Ptashne, M. (1988) *Nature*, **335**, 683-689.
27. Lund, T. and Laland, S.G. (1990) *Biochem. Biophys. Res. Comm.*, **171**, 342-347.
28. Draetta, G. (1990) *TIBS*, **15**, 378-383.
29. Solomon, M.J., Strauss, F. and Varshavsky, A. (1986) *Proc. Natl. Acad. Sci. (USA)*, **83**, 1276-1280.
30. Ashley, C.T., Pendleton, C.G., Wesley Jennings, W., Saxena, A. and Glover, C.V.C. (1989) *J. Biol. Chem.*, **264**, 8394-8401.
31. Reeves, R., Langan, T.A. and Nissen, M.S. (1991) *Proc. Natl. Acad. Sci. (USA)*, **88**, 1671-1675.
32. Luscher, B., Christenson, E., Litchfield, D.W., Krebs, E.G. and Eisenman, R.N. (1990) *Nature*, **344**, 517-522.
33. Reeves, R., Elton, T.S., Nissen, M.S., Lehn, D.A. and Johnson, K.R. (1987) *Proc. Natl. Acad. Sci. (USA)*, **84**, 6531-6535.
34. Devlin, C., Tice-Baldwin, K., Shore, K. and Arndt, K. (1991) *Mol. Cell. Biol.*, **11**, 3642-3651.

High-mobility-group (HMG) proteins and histone H1 subtypes expression in normal and tumor tissues of mouse

Vincenzo GIANCOTTI¹, Antonella BANDIERA¹, Lorenza CIANI¹, Daniela SANTORO¹, Colyn CRANE-ROBINSON², Graham H. GOODWIN³, Mauro BOIOCCHI⁴, Riccardo DOLCETTI⁴ and Bruno CASSETTA⁵

¹ Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole, Università di Trieste, Italy

² University of Portsmouth, Biophysics Laboratory, England

³ Institute of Cancer Research, Chester Beatty Laboratories, London, England

⁴ Centro di Riferimento Oncologico, Aviano (PN), Italy

⁵ Perkin-Elmer European Support Center, Munich, Germany

(Received January 4, 1993) – EJB 93 0009/2

Exhaustive extraction of mouse tissues with perchloric acid has been used together with reverse-phase HPLC and electrophoresis to quantify the amounts of chromosomal proteins HMG17, HMG14 and HMGI, relative to histone H1. Normal lung and thymus contain $\approx 3\%$ HMG17/HMG14 but only $\approx 2\%$ HMGI. In tumor tissues (Lewis lung carcinoma and lymphoma NQ35), the amount of HMG17/HMG14 is not greatly altered but HMGI levels rise considerably, reaching 10% in Lewis lung carcinoma. HMGI synthesis does not replace HMG17/HMG14 proteins, suggesting that HMGI proteins contribute to the structure of chromatin regions in a manner distinct from those of HMG17/HMG14. Ion-spray mass spectrometry has been used to determine the molecular masses of H1 subtypes from the same four mouse tissues. In addition to the six known species H1^o, H1a, H1b, H1c, H1d and H1e, a newly defined subtype of mass 21756 Da from Lewis lung carcinoma, named H1L was identified. Several phosphorylated H1 subtypes have also been defined by mass spectrometry. The combined use of reverse-phase HPLC and electrophoresis permitted quantification of these seven histone H1 subtypes in the four mouse tissues. Increased phosphorylation of H1 subtypes in tumors parallels the phosphorylation of HMGI proteins which are present in great amounts, showing that both are involved as post-translational-modified forms in the structure of the chromatin of neoplastic systems.

Histones and high-mobility-group (HMG) proteins are abundant DNA-binding molecules that affect replication and transcription [1–5]. The low-molecular-mass HMG proteins have molecular masses of approximately 10–12 kDa and form the following two subgroups: the HMG17 and HMG14 subgroups which are homologous proteins found in mammalian and avian cells [4]; the HMGI subgroup, which comprises three structurally related proteins named HMGI, HMGY and HMGI-C [6–11].

HMG17 and HMG14 proteins have been suggested to function as transcription regulating factors in active genes [4, 5] whilst HMGI proteins, preferentially expressed in rapidly dividing, malignant and/or undifferentiated cells [6–10, 12–18], could be involved in the activation of genes linked to phenotypic transformation. Correlation between the expression of the two subgroups is thus possible, particularly on the basis of the recent implication that the HMGI(Y) proteins act as regulating factors in the expression of genes such as the tumor necrosis factor β [19], the gp91-phox [20] and the human interferon β [21].

In this paper, we report quantification of the HMG17/HMG14 and HMGI proteins in normal mouse tissue and

Correspondence to V. Giancotti, Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole, Via L. Giorgieri 1, I-34127 Trieste, Italy

Abbreviations. HMG, high mobility group; LLC, Lewis lung carcinoma.

mouse tumor tissues using histone H1 as a reference. We demonstrated a decrease in the expression of the HMG17/HMG14 proteins in Lewis lung carcinoma (LLC) which is a very tumorigenic tissue, but not in lymphomas which express similar amounts of both HMG17/HMG14 and HMGI proteins. In LLC, HMGI proteins do not simply replace the HMG17/HMG14 ~~and HMGI~~ subgroup, since their expression is four-times or five-times more than the HMG17/HMG14 decrease.

In addition, we have used ion-spray mass spectrometry to completely characterize histone H1 subtypes in mouse, and thereby consolidate their nomenclature and their levels of phosphorylation. This could be of interest, since a possible role of histone H1 subtypes in transcriptionally active chromatin is under speculation [22–28] and a connection between the interaction of H1 subtypes with DNA and that of HMG proteins could be determined. In fact, the transcriptional activation of the human interferon- β gene by HMGI(Y) proteins has been interpreted as an alteration in the structure of naked DNA or chromatin which facilitates the binding of the other proteins [21]. Histones should be involved in these structural transitions. We demonstrated that the six known mouse H1 subtypes are easily identified by mass spectrometry and, at the same time, we showed the presence, in LLC, of many phosphorylated forms and a new H1 subtype characteristic of this tumor. Finally, we demon-

strated that the pathway of expression of HMGI proteins is the reverse of that of histone H1°.

MATERIALS AND METHODS

Protein extraction and purification

BD2F1 mice were used as a source of normal lung and for the growth of LLC [13]. 4-weeks-old AKR/J mice were used as a source of normal thymus and for the growth of lymphomas NQ9, NQ35 and NQ36 [29, 30]. Data for the three lymphomas were very similar, therefore only that for lymphoma NQ35 are shown. HMG proteins and histone H1 were extracted from tissues with 5% (mass/vol.) perchloric acid and precipitated by acetone [9, 12, 13]. Double extraction assured complete protein recovery. Low-molecular-mass HMG proteins were purified by reverse-phase HPLC on a Bio-Rad RP304 column using a Waters apparatus [10, 13]. Protein was eluted (0.5 ml/min) with a water/acetonitrile gradient containing 0.1% (by vol.) trifluoroacetic acid. In such experiments, histone H1 subtypes and proteins HMG1 and HMG2 eluted unresolved at the end of the gradient. Histone H1 and HMG1 and HMG2 fractions were pooled, dialyzed, lyophilized and the total protein analyzed by electrophoresis. The mixture was then rechromatographed by reverse-phase HPLC using a different gradient to separate H1 species from HMG1 and HMG2. Electrophoretic patterns for the mixture of histone H1, HMG1 and HMG2 were scanned by densitometry in order to evaluate the total amount of histone H1 for use in calculations of the expression levels of HMG proteins.

Electrophoretic analysis

Polyacrylamide (15%) gel electrophoresis (250 mm × 0.7 mm) was carried out in acetic acid/urea, essentially as described [9]. SDS/polyacrylamide-gel electrophoresis (first or second dimension) was performed as previously reported [10]. Gels were stained with Coomassie blue.

Quantification of expression

For the low-molecular-mass HMG proteins, quantification was performed by taking HPLC peak areas and dividing by the total HPLC peak area of all H1 subtypes. The proportion of H1 in the HPLC peaks comprising both H1, HMG1 and HMG2 proteins was calculated by densitometric analyses of first-dimension electrophoresis. In order to evaluate the error that affects the reported expressions, three separate extractions from LLC were carried out and nine different HPLC separations were performed. Measurement of peak areas then allowed the standard deviation to be calculated. The expression of low-molecular-mass HMG proteins were thereby estimated to be subject to an error of $\pm 10\%$. The composition of histone H1 was calculated using the HPLC peak areas of the subtypes. Densitometry of electrophoretic patterns was used to evaluate the relative contribution of each H1 component in the case of peaks containing more than one species. The error in H1 composition was calculated as for the HMG proteins using five different HPLC separations of H1 from LLC. The error in the H1 subtype composition was estimated at $\pm 8\%$.

Mass-spectrometric analysis

Determination of the molecular masses of the HMG proteins and histone H1 subtypes was carried out by ion-spray

mass-spectrometry techniques [31] using a triple quadrupole API III LC/MS/MS instrument. HPLC-purified samples (10–50 μg) were dissolved in 50–100 μl water/acetonitrile (80:20, by vol.) containing 0.1% formic acid, and injected into the ion source, at a flow rate of 2 $\mu\text{l}/\text{min}$. The spectra were scanned from m/z 600–1600 for HMG proteins and from m/z 500–1200 for histone H1. Mass-scale calibration was carried out by means of poly(propylene glycol) and checked by the multiple charged ions of a separate introduction of myoglobin (16950.5 Da). All data are reported as average masses (Da). Some histone H1 masses were also determined using a BIO-Q triple quadrupole mass spectrometer [32]. The difference in the mass values obtained for the same sample of histone H1 using the two instruments was no more than ± 2 Da.

RESULTS

Low-molecular-mass HMG proteins

HMG proteins and histone H1 were extracted with perchloric acid from two normal mouse tissue (lung and thymus), from LLC [13] and from the three lymphomas NQ9, NQ35 and NQ36 [29, 30]. Reverse-phase HPLC of the extracted protein and acetic acid/urea electrophoretic analyses of the eluted peaks (Fig. 1) allowed the identification of low-molecular-mass HMG proteins. The difference in both elution time and electrophoretic mobility between HMGY and HMG14 proteins was such that a precise identification was possible. However, the two subtypes of HMGY, Y_1 and Y_2 [10, 13, 32], could not be easily discriminated as they migrate in acetic acid/urea gels with a mobility difference of about 0.01% ($Y_2 < Y_1$). Since HMGY₁ and HMGY₂ have the same sequence but differ only in one phosphate group [32], we carried out mass-spectrometric experiments to positively identify these components. Three samples (lung, thymus and NQ35 lymphoma) show only HMGY₂ (10686.3 \pm 1.3 Da) whereas, similar to preceding data [10, 13, 32], LLC shows the two subtypes HMGY₁ (10766.4 \pm 0.8 Da) and HMGY₂ (10686.6 \pm 0.5 Da) which contain three and two phosphate groups, respectively [32]. The small peak labelled X, eluting just ahead of HMGY₂ in the thymus and NQ35 samples (Fig. 1), is not an HMGY protein.

Fig. 2 shows a quantification of the expression of low-molecular-mass HMG proteins relative to that of histone H1. All samples show predominant expression of HMG17 relative to HMG14. This result qualitatively agrees with data reported for transformed human cells [33] in which HMG17 expression is higher than HMG14. However, Fig. 2 shows that the HMG17/HMG14 ratio in all tissues studied is higher than 4.9, the maximum value reported for HL60 cells [33]. We found the highest value in the thymus (11.3) and the lowest value in normal lung (5.4) and LLC (5.7). We conclude that HMG17 is the main protein in the subgroups HMG17 and HMG14, such that the HMG17/HMG14 ratio is high even if there are considerable phenotypic changes.

We also quantified the changes in the HMGI subgroup following neoplastic transformation; Fig. 2 shows the massive increase of both HMGI and HMGY proteins in the two tumors compared to the levels found in normal tissues. The high level of expression of the HMGI subgroup in transformed systems has been previously reported in qualitative terms [12, 13] and is consistent with other data showing a decrease in HMGI and HMGY levels during differentiation of F9 mouse teratocarcinoma [16] and human K562 cells

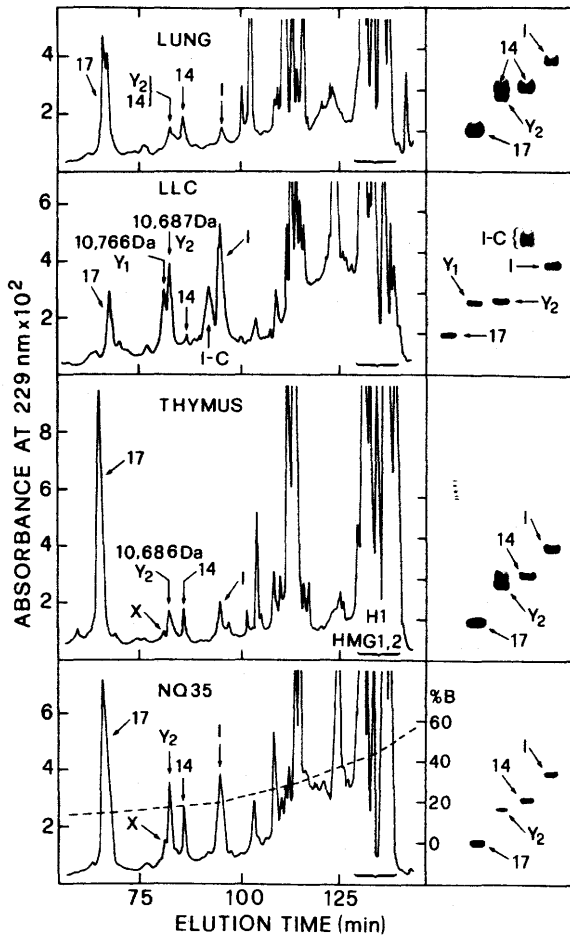


Fig. 1. Reverse-phase HPLC elution profiles of perchloric-acid-extracted mouse proteins from normal lung, LLC, thymus and NQ35 lymphoma. On the right, the acetic acid/urea gel electrophoresis of identified HMG proteins is shown. The dotted line and scale on the NQ35 elution profile gives the solvent gradient. A, 0.1% (by vol.) trifluoroacetic acid in water; B, water/acetonitrile/trifluoroacetic acid (5:95:0.1, by vol.). The relative intensity of the electrophoretic bands does not correspond to that of the HPLC peak areas. Gels were stained with Coomassie blue. 17, HMG17; 14, HMG14; I, HMG1; Y, HMGY; I-C, HMG1-C.

[18]. It is clear, therefore, that a distinctive feature of neoplastic and/or undifferentiated cells is a high level of expression of HMG1 and HMGY. These observations raise the question as to whether the increased expression of HMG1 proteins is paralleled by a simultaneous decrease in the expression of HMG17 and HMG14, i.e. do HMG1 proteins simply replace HMG17 and HMG14 proteins in the chromatin, or do HMG1 proteins activate new genes independently of HMG17 and HMG14? Fig. 2 compares the summed HMG17/HMG14 level with the summed HMG1, HMGY, HMG1-C levels for the four cell types. In the two normal tissues, the levels of HMG17 and HMG14 are well in excess of the HMG1 and, in the lymphoma both increase the levels of the HMG1 species to a considerable degree. In lung carcinoma cells, however, there is a great increase in HMG1 proteins (4–5 times), whereas the HMG17/HMG14 level falls significantly. There is thus no evidence to suggest that, when the synthesis of HMG1 proteins increases in tumor cells, this is always in replacement of HMG17/HMG14 synthesis. The data suggest that the function of the HMG1-type proteins is distinct from that of the HMG17/HMG14 type.

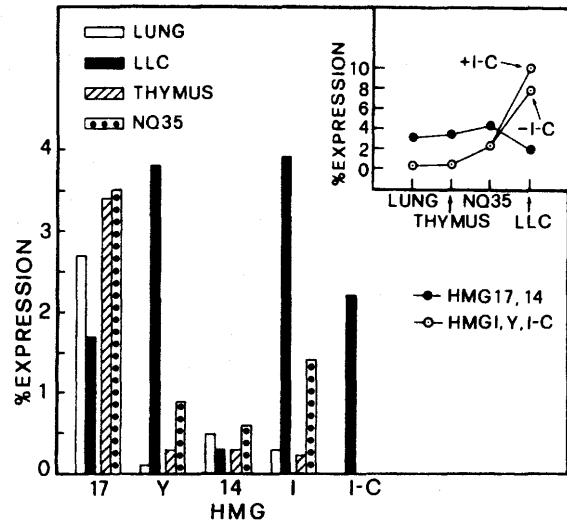


Fig. 2. Quantification ($\pm 10\%$) of the expression of the low-molecular-mass HMG proteins relative to histone H1. Quantification was using a combination of HPLC peak areas and electrophoretic data (see Materials and Methods). The inset shows the expression of the summed HMG17,14 pair (\bullet) and the total HMG1 proteins (\circ). The additional point -I-C represents the total HMG1 proteins less the contribution of HMG1-C protein.

The high levels of HMG1-C in LLC tissue is striking, particularly when contrasted with the other tumor tissue (NQ35) in which no HMG1-C was observed. Very different levels of HMG1-C are also noted in various human tumor types (Manfioletti, G., unpublished results). At present it is not possible to correlate the level of this protein with the phenotypic properties of tumors recently suggested for the other HMG1 subtypes [34].

Histone H1 subtypes

The reverse-phase HPLC experiments of Fig. 1, carried out in order to purify low-molecular-mass HMG proteins, did not completely separate histone H1 subtypes, HMG1 and HMG2. The peaks marked 'H1-HMG1,2' in Fig. 1 were therefore pooled and, after dialysis and lyophilization, a further chromatography was carried out using a different solvent gradient which gave a better separation (see Fig. 3). All samples showed six main peaks (Fig. 3; I–IV), whilst LLC showed two additional peaks (VII and VIII). The peak composition was determined using both acetic acid/urea and SDS/polyacrylamide electrophoreses; the subtypes identified are reported in Fig. 3 according to the nomenclature introduced by Lennox et al. [35–37]. More recently, this nomenclature was also assumed by Palyga [38] and Lindner et al. [39, 40]. Electrophoresis showed that peaks II and III (Fig. 3) contain several subtypes of HMG1 and HMG2 [4] but these details are not discussed further. The chromatography shown in Fig. 3 did not separate histone H1b from histone H1 $^{\circ}$ (peak I) nor histone H1d from histone H1e (peak V). Incomplete resolution of this pair of H1 histones has also been reported in the literature [40, 41]. However, subsequent electrophoretic analyses in the case of H1b, H1 $^{\circ}$, allows the identification of the two histones, since H1 $^{\circ}$ is the fastest migrating H1 subtype in both acetic acid/urea and SDS/polyacrylamide gels and H1b is the slowest migrating species on SDS/polyacrylamide gels [35–37, 39, 40].

It is more complicated to differentiate H1d from H1e since these two subtypes have very similar electrophoretic

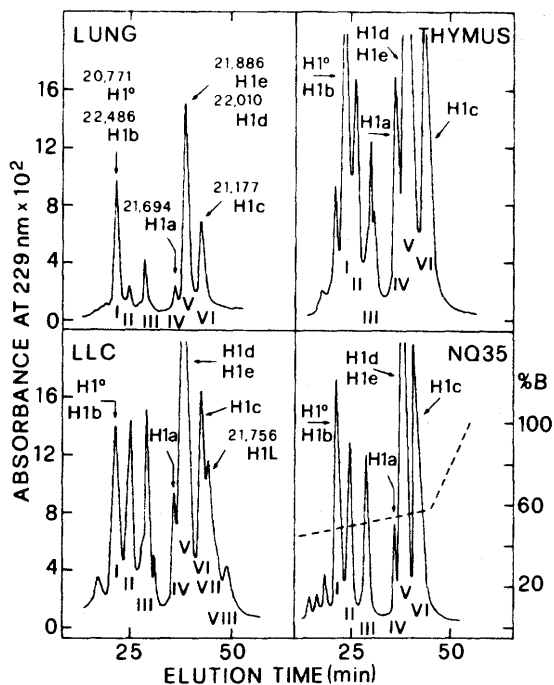


Fig. 3. Reverse-phase HPLC elution profiles of samples of histone H1 and HMG1 and HMG2 proteins obtained from the experiments of Fig. 1. The predominant H1 species in each peak is indicated. The molecular mass (Da) obtained by ion-spray mass spectrometry of the sample from lung is also given. For the LLC sample, the molecular mass of the newly identified H1L is given. The solvent gradient used is shown in the NQ35 elution profile. A, 0.1% (by vol.) trifluoroacetic acid in water; B, water/acetonitrile/trifluoroacetic acid (30:70:0.1, by vol.).

mobilities. We initially assumed their presence in peak V (which constitutes 40–50% of the total H1 histone; Fig. 3) on the basis of data given in the literature [39–41]. Ion-spray mass spectrometry was therefore used to resolve this problem and make further distinctions between H1 subtypes by analyzing all H1 peaks. Table 1 shows the molecular-mass data. The histone type identified is given in this table and also in Fig. 3 and subsequent figures. Fig. 4 is illustrative of the mass spectra obtained and shows the data for peak V in which the two H1 subtypes predominating have average masses of 21887 and 22010 Da. These two subtypes represent H1e and H1d, respectively, and it can be seen that these two species are not expressed at the same levels in all four tissues. Lennox and Cohen [36] showed by electrophoresis that, in mouse lung, H1e is the most prominent subtype, whereas the thymus has very low levels of this protein compared to H1d. Fig. 4 substantiates this finding. Among the four tissues studied, Fig. 4 shows that only lung expresses more H1e than H1d.

The mass measurements show that peak I contains two families of well-spaced molecular species, one having a mass of about 20.8 kDa and the other with a mass in the range 22.0–22.5 kDa (Table 1). The family with lower mass represents the H1^o subtype, since it is well known that H1^o is the smallest species of the H1 subtypes [35–37, 39, 40, 42]. However, all three components of the 20.8-kDa family can be considered as H1^o subtypes, since they show mass differences of less than 100 Da. The H1^o protein, therefore, consists of a family of at least three components, the predominant component having a mass of 20771 Da. A mouse H1^o sequence has been determined from cDNA cloning [42] and contains 193 amino acids with a molecular mass of

Table 1. Ion-spray molecular mass spectrometric data for histone-H1 subtypes obtained by reverse-phase HPLC as shown in Fig. 3. Underlined numbers show the predominant components found in each peak. n.d., not detected; P, phosphate group.

Peak	Molecular mass of histone H1 from				Identified histone	
	lung	LLC	thymus	NQ35		
I	20 730	20 727	n.d.	20 731	H1 ^o	
	<u>20 771</u>	20 770	n.d.	<u>20 772</u>		
	20 838	n.d.	n.d.	20 836		
	n.d.	22 055	22 056	22 058		
	n.d.	n.d.	22 133	n.d.		
	<u>22 486</u>	22 486	<u>22 487</u>	<u>22 489</u>		H1b
IV	n.d.	<u>22 566</u>	22 570	n.d.	H1b + 1P	
	n.d.	n.d.	n.d.	22 583		
	21 259	21 263	21 266	21 265		H1c + 2P
	n.d.	n.d.	n.d.	21 338		
	<u>21 694</u>	<u>21 695</u>	<u>21 697</u>	<u>21 696</u>		
	n.d.	n.d.	n.d.	21 771		H1c + 2P
n.d.	21 792	21 795	21 792			
21 216	n.d.	n.d.	n.d.	H1c + 2P		
n.d.	21 336	21 342	21 340			
n.d.	21 415	21 417	n.d.			
21 457	21 455	21 459	21 458		H1c + 3P	
21 579	21 577	21 580	21 579		H1c + 5P	
n.d.	21 696	n.d.	n.d.		H1a	
V	<u>21 886</u>	21 886	21 888	<u>21 888</u>	H1e	
	n.d.	21 968	n.d.	n.d.	H1e + 1P	
	22 010	<u>22 008</u>	<u>22 011</u>	<u>22 011</u>	H1d	
	n.d.	22 086	22 093	22 091	H1d + 1P	
	n.d.	n.d.	22 108	n.d.		
	VI	n.d.	n.d.	20 538	20 538	H1c
20 776		20 777	20 778	20 778		
20 854		20 854	n.d.	n.d.		
n.d.		n.d.	20 915	n.d.		
<u>21 177</u>		<u>21 176</u>	<u>21 178</u>	<u>21 178</u>		
n.d.		21 198	n.d.	n.d.		
n.d.	n.d.	21 257	n.d.	H1c		
21 275	21 273	n.d.	n.d.			
VII	n.d.	21 176	n.d.	n.d.	H1c	
	n.d.	21 327	n.d.	n.d.		
	n.d.	21 509	n.d.	n.d.		
	n.d.	<u>21 756</u>	n.d.	n.d.		H1L
	n.d.	21 888	n.d.	n.d.		
	n.d.	22 012	n.d.	n.d.		H1e
VIII	n.d.	20 800	n.d.	n.d.	H1d	
	n.d.	21 097	n.d.	n.d.		
	n.d.	21 736	n.d.	n.d.		
	n.d.	22 690	n.d.	n.d.		
VIII	n.d.	22 732	n.d.	n.d.	H1d	
	n.d.	22 946	n.d.	n.d.		

20715 Da, which is lower by a factor of 56 than the value of 20771 Da determined here. A characteristic of H1 histones is N-terminal acetylation [43], which would add 42 Da to give a mass of 20757 Da, but this is still far from the observed mass. The 20757 Da species does not correspond to any of the other two components presented in Table 1. The difference between an N-terminus acetylated form and the 20771-Da species is 14 Da, which is the mass of a methylene group and could thus be attributed to several possible amino acid substitutions. We conclude that either the published cDNA sequence refers to a minor component of the mouse

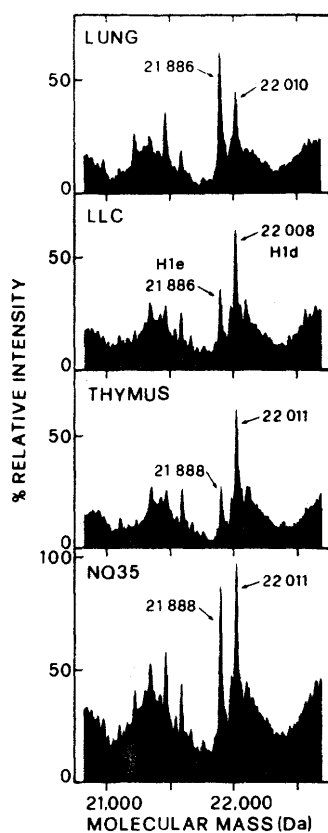


Fig. 4. Molecular-mass reconstruction from the ion-spray mass spectra of the protein recovered from the four peaks from Fig. 3.

H1^o family (that we did not detect), or that a cloning artefact led to a reading error of an amino acid, such as valine, serine or glycine in place of leucine/isoleucine, threonine or alanine.

The second family in peak I contains up to five components (Table 1), the predominant component having a mass of 22487 Da. Since it has the lowest mobility of all the H1 species in SDS/polyacrylamide gels (Fig. 5), we assume that this component is histone H1b, in agreement with published data [36]. The family of histone H1b consists of just four sequence variants (22057, 22133, 22487, 22583) since it is likely that the component of 22566 Da (seen prominently in LLC) is a mono-phosphorylated form of the 22487 Da histone H1b. It is worthwhile noting that this phosphorylated form is predominant only in LLC, where it co-exists with its unphosphorylated form. No amino-acid-sequence data are available for mouse histone H1b and, consequently, no mass-value comparisons can be made.

Peak IV contains histone H1a, as suggested by both HPLC elution profiles and electrophoretic analysis. Up to five components can be identified in this peak; the main component has an average mass 21696 Da, which we assume to be the unphosphorylated histone, H1a. No amino acid sequence is available for this subtype.

As already discussed, peak V (Table 1) contains the two subtypes H1d and H1e of masses 22010 Da and 21887 Da, respectively. There are also many other minor components in this peak, most of which represent phosphorylated H1 subtypes. Three phosphorylated species of histone H1c, which are found principally in peak VI (Table 1) have also been identified in peak V. These phosphorylated forms of histone H1c elute from a reverse-phase HPLC column before the

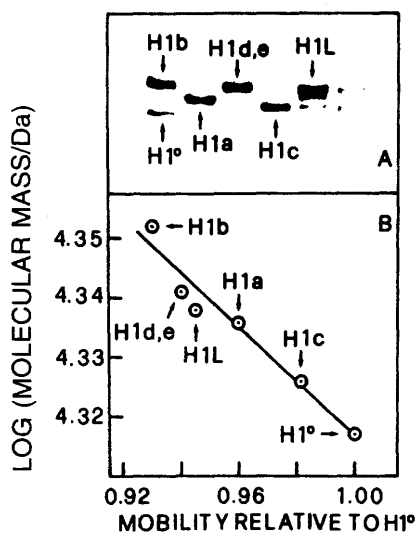


Fig. 5. SDS/polyacrylamide gel electrophoresis. (A) LLC histone H1 subtypes purified by reverse-phase HPLC as in Fig. 3. The lanes (left to right) correspond to peaks I, IV, V, VI, VII and VIII. (B) mobilities relative to the H1^o subtype are plotted against the log (molecular masses) obtained from mass spectrometry.

parent molecule, probably due to a decreased hydrophobicity. In peak V (Table 1), a phosphorylated species of histone H1e (in LLC) and a phosphorylated form of H1d (in LLC, thymus and NQ35 lymphoma, but not in lung) have been also identified. Peaks IV and V (Table 1) are not well separated by chromatography and it is possible that some components elute in both peaks. This is the case for the biphosphorylated form of the histone H1c (21338 Da in Table 1 peak IV and 21340 Da in peak V) and for histone H1a, which mainly elutes in the peak IV but is found as a minor component in peak V of LLC.

The predominant component of the family of peak VI (Table 1) is a species having a molecular mass of 21177 Da. Yang et al. [44] reported the sequence of a mouse histone named H1.1 which probably corresponds to the main component of peak VI (Table 1), which we expect to contain histone H1c. The amino acid sequence [44] (excluding the initial methionine) yields a molecular mass of 21135 Da and addition of N-terminal acetylation (42 Da) gives a molecular mass of 21177 Da, the observed mass of the main component of peak VI. The subtype sequenced by Yang et al. [44] and named H1.1 thus corresponds to H1c in the nomenclature of Lennox et al. [35, 36].

The elution profile of the LLC sample in Fig. 3 shows two extra peaks (VII and VIII) not found in the other samples. Mass determinations carried out on these peaks (Table 1) showed many components all having masses indicating that they are H1 subtypes. Among the 12 components identified, that of 21176 Da in peak VII corresponds to H1c (due probably to some overlapping of peaks VI and VII), whilst that of 21888 Da could be residual H1e from peak V. Among the 12 H1 species of peaks VII and VIII, the predominant component is a species of molecular mass 21756 Da. This histone does not appear to be the mouse equivalent of histone H1t found in rat testis [35, 39] which also elutes as the last peak in reverse-phase HPLC, since in acetic acid/urea electrophoresis, H1t migrates to the same position as H1a and in SDS/polyacrylamide electrophoresis H1t has the highest mobility together with H1^o [39]. This mouse component of 21756 Da migrates as H1b in acetic

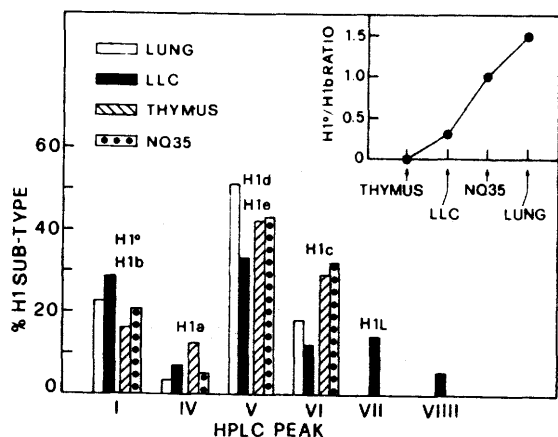


Fig. 6. Quantification ($\pm 8\%$) of the histone H1 subtypes calculated from the HPLC peak areas. The data for peaks I and V represent the summed amounts of H1° plus H1b and H1d plus H1e, respectively. The inset shows the ratio H1°/H1b obtained by densitometry of the electrophoretic patterns of peak I.

acid/urea gels (data not shown) and in SDS/polyacrylamide gels is slower than H1°, as shown in Fig. 5. This histone has not been described in previous studies of murine neoplastic cells [35, 46] and we name this component of peak VII at 21756 Da as histone H1L. Peaks VII and VIII of LLC also contain other components ranging in mass over approximately 21–23 kDa. From considerations of the hydrophobicity, these species are not expected to be phosphorylated forms of the histones that eluted earlier. In the absence of other data, we conclude that peaks VII and VIII from LLC contain newly histones having variant amino acid sequences and/or forms of the other histones modified by post-translational modifications, e.g. acetylation [43] or poly(ADP) ribosylation [47].

Peaks VI and VIII contain some species whose mass correspond to that of H1°. One of these species could correspond to the protein recently reported by Terzmanowski and Cole [48] who have found a H1 subtype having the same electrophoretic mobility as H1°.

Fig. 5A shows the SDS/polyacrylamide electrophoretic pattern obtained for the chromatographic peaks I, IV–VII and VIII from LLC and thus shows only the main components which predominate in the mass measurements. The other samples (lung, thymus and NQ35 lymphoma) show similar electrophoretic patterns, but do not show the proteins of the peaks VII and VIII which are characteristic only of LLC. The mobilities relative to that of the fastest histone (H1°) were plotted using the mass-spectrometry data of the predominant components in Table 1 (Fig. 5B). The linearity demonstrates consistency between the relative mobility and the observed molecular mass.

The total composition of the mixture of H1 subtypes has been calculated from the peak areas of Fig. 3. Fig. 6 shows H1 composition data and the predominant components in each peak are also reported. The results obtained on normal tissues (lung and thymus) are similar to those reported by Lennox and Cohen [36], but we can discriminate H1a from H1c since we used HPLC data in addition to electrophoresis as carried out by Lindner et al. [39]. The pair H1d, H1e constitute approximately 40–50% of the total histone; a lower value was found for LLC, which contains additional H1 subtypes (peaks VII and VIII) which sum to about 20% of the total. H1a is the least abundant in all samples, being only 4% in lung, the same as that obtained for rat liver by

Lindner et al. [39]. Fig. 6 also shows the H1°/H1b ratio in peak I, obtained by electrophoretic densitometric measurements. The data are consistent with the general observation that H1° is found prevalently in cells that are in terminal stages of differentiation or that have a low rate of cell division [42, 49–53]. The two neoplastic tissues studied (LLC and NQ35 lymphoma) show very different H1°/H1b ratios, suggesting that they are representative of a different stage of differentiation and/or different rates of division. In this respect, as reported by Lennox and Cohen [36], thymus, which does not express H1°, is anomalous.

DISCUSSION

There are at least three pieces of evidence that link HMGI proteins to histone H1. (a) Both groups of protein are substrates for two homologous kinases, i.e. the p34/CDC2 kinase [54, 55] and the growth-associated H1 histone kinase [49, 56, 57]. (b) HMGI protein shows preferential binding to AT-rich DNA sequences [58–62, 20, 21] and preferential binding of histone H1 to AT-rich and scaffold-attachment regions has been noted [63, 64]. (c) HMGI proteins were localized in the G/Q and C bands of metaphase chromosomes [65] and histone H1 undergoes great phosphorylation at mitosis [66, 67]. The data of Fig. 2 and Fig. 6 show that, for the two lung tissues, increased expression of HMGI proteins is accompanied by a decrease in histone H1°, whilst the amount of HMGI proteins in lung is 0.4%; it rises to 10% in LLC, but the H1° level drops from 14% to 7% of the total H1.

The data of Fig. 2 show that, in tumors (LLC and NQ35), increased HMGI synthesis is not necessarily matched by a decrease in the HMG17/HMG14 pair; high levels of HMGI and HMGY were observed in NQ35 without any decrease in the levels of HMG17/HMG14, although in LLC, where a HMG17/HMG14 decrease was noted, the three HMGI proteins together show very high levels of expression ($\approx 10\%$). Since these figures were obtained by measurements of the peak areas obtained from HPLC and electrophoretic densitometry, they represent a mass ratio rather than a molar ratio. If one assumes that the mass of an HMGI molecule is half that of an H1, this implies approximately 1 molecule HMGI protein for every 5 molecules histone H1 (summing all subtypes). The same calculation carried out for the pair HMG17/HMG14 in the case of normal lung gives 1 molecule HMG17/HMG14 for every 17 molecules histone H1. If we consider that there is 1 molecule histone H1 bound to each nucleosome [1], this means that there is, on average, 1 molecule HMGI proteins every 5 nucleosomes in LLC. The conclusion is that the chromatin structure in LLC (and to lesser extent in NQ35) should be much altered due to the increased presence of HMGI proteins which, from their primary sequence [11], are able to interact with both DNA and histones. The increased presence of HMGI proteins reported here and their involvement in the transcription of different genes [19–21] suggest a role for general activators in transcription. In light of this, a correlation with histones and particularly with both H1 molecules which are outside the nucleosome, and N-terminal tails of core histones which could develop a regulating function, should exist.

Ion-spray mass spectrometry represents a very precise method for the characterization of histone H1 subtypes and, in the absence of amino acid sequence data, molecular masses constitute an excellent procedure for identification of

H1 subtypes. We have adopted the nomenclature of [35, 36, 39, 40] which is based on electrophoretic and HPLC data. The six predominant murine H1 subtypes are, in order of elution in reverse-phase HPLC:

H1^o; 20771 Da; H1b, 22487 Da; H1a 21696 Da; H1d, 22010 Da; H1e, 21887 Da; H1c, 21177 Da.

A newly defined H1 histone was observed in LLC and named H1L; its molecular mass is 21756 Da.

The subtype H1c was identified as the species of mass 21177 Da, corresponding to the histone sequenced by Yang et al. [45] and named H1.1. This H1 histone was found to be the most phosphorylated subtype, having up to five phosphate groups (Table 1, peak V). Phosphorylation of histone H1 has been extensively studied [35, 36, 46, 48, 66, 68, 69] and up to 10 phosphorylated H1 species identified by electrophoresis in F9 murine teratocarcinoma cells [35]. Subtypes carrying 5 or 6 phosphate groups were found in Ehrlich ascites and Novikoff hepatoma cells [66]. Among H1 subtypes, H1b was reported as the most phosphorylated species by Lennox et al. [35] and by Matsukawa et al. [68]. Higurashi et al. [69] found that, in mouse lympholeukemic L5178Y cells, the most highly phosphorylated species was histone H1-II, which corresponds to the H1b described by Lennox et al. [35, 36]. In contrast, Ajiro et al. [46] reported that, in mouse neuroblastoma N18 cells, phosphorylation occurs mainly in the subtype H1c which, according to the nomenclature of Lennox et al. [35, 36], corresponds to the H1d subtype. We identified a monophosphorylated form of both H1b (LLC and thymus) and H1d (LLC, thymus and NQ35 lymphoma), but lung, as a non-dividing tissue [35], shows the lowest level of phosphorylation. LLC is the only sample (Table 1, peak I) that shows a phosphorylated species as the predominant form. This species is the H1b subtype and, in this respect, our mass spectrometric results are consistent with the literature [35, 36, 68, 69].

In the tissues studied, only a small proportion of cells could be considered as mitotic and, therefore, the H1 phosphorylation found should not be due to growth-associated histone H1 kinase [48, 54–57], but rather to interphase-acting kinases such as the cAMP-dependent or the cGMP-dependent protein kinases or to protein kinase C [70]. In the amino acid sequence of histone H1c [44], there are 5 Ser/Thr residues having the consensus R/K-X-S/T that is the presumed site for cAMP-dependent or cGMP-dependent protein kinase action, i.e. Ser35, Thr98, Thr142, Thr153, Thr180. Moreover, there are five residues having the consensus S/T-X-K/R (Ser103, Thr146, Thr153, Thr166, Ser172) which are potential sites for protein kinase C action. Residues Thr146, Thr153, and Ser172 could also be phosphorylated by growth-associated histone H1 kinase [70]. In histone H1c, there are therefore nine Ser/Thr residues in that could be phosphorylated during interphase. Only one site (Ser35) is located at the N-terminus of the molecule, none in the globular domain [71] and all others at the C-terminus of the molecule. HMGI proteins too are phosphorylated during interphase as is histone H1 [6, 7, 10, 12]. The same post-translational modification therefore concerns two groups of nuclear proteins (HMGI and H1 subtypes), both involved in the structural organization of the chromatin. HMGI binds to the minor groove of DNA, showing preferential binding to AT-rich regions [58–62, 20, 21]; H1 molecules, in contrast, bind linker DNA and some subtypes are preferentially phosphorylated. Permanent phosphorylation of HMGI proteins at the C-terminus of the molecule [32] (we did not identify unphosphorylated forms) increases the region of negative charge due

to the numerous glutamic acid residues [11] and potentiates the interaction with other positively charged proteins. Phosphorylation of HMGI proteins contributes towards the high levels at which they are found in the nucleus of neoplastic cells. At the same time, phosphorylation of some H1 subtypes makes their detachment from linker DNA easier. Both phosphorylation of HMGI proteins and high levels found should contribute to the alteration in chromatin structure of the regions which became activated during transcription.

This research was funded by grants from the *Associazione Italiana per la Ricerca sul Cancro* (AIRC), Milan, Italy, the *Consiglio Nazionale delle Ricerche* (CNR), Rome, Italy, the *Ministero della Ricerca Scientifica e Tecnologica*, Rome, Italy, the European Community Science programme and the Cancer Research Campaign, United Kingdom. We thank Dr. Pasquale Ferranti for carrying out certain mass-spectrometric measurements.

REFERENCES

- van Holde, K. E. (1988) *Chromatin* (Rich, A., ed.) Springer-Verlag New York.
- Gross, D. S. & Garrard, W. T. (1988) *Annu. Rev. Biochem.* 57, 159–197.
- Grunstein, M. (1990) *Annu. Rev. Cell Biol.* 6, 643–678.
- Johns, E. W. (1982) *The HMG chromosomal proteins* (Johns, E. W., ed.) Academic Press, New York.
- Bustin, M., Lehn, D. A. & Landsman, D. (1990) *Biochim. Biophys. Acta* 1049, 231–243.
- Lund, T., Holtlund, J. & Laland, S. G. (1983) *FEBS Lett.* 15, 163–167.
- Lund, T., Dahal, K. H., Holtlund, J. & Laland, S. G. (1987) *Biochem. Biophys. Res. Commun.* 146, 725–730.
- Johnson, K. R., Lehn, D. A. & Reeves, R. (1989) *Mol. Cell. Biol.* 9, 2114–2123.
- Giancotti, V., Berlingieri, M. T., Di Fiore, P. P., Fusco, A., Vecchio, G. & Crane-Robinson, C. (1985) *Cancer Res.* 45, 6051–6057.
- Giancotti, V., Bandiera, A., Buratti, E., Fusco, A., Marzari, R., Coles, B. & Goodwin, G. H. (1991) *Eur. J. Biochem.* 198, 211–216.
- Manfioletti, G., Giancotti, V., Bandiera, A., Buratti, E., Sautière, P., Cary, P., Crane-Robinson, C., Coles, B. & Goodwin, G. H. (1991) *Nucleic Acids Res.* 19, 6793–6797.
- Giancotti, V., Pani, B., D'Andrea, P., Berlingieri, M. T., Di Fiore, P. P., Fusco, A., Vecchio, G., Philp, R., Crane-Robinson, C., Nicolas, R. H., Wright, C. A. & Goodwin, G. H. (1987) *EMBO J.* 6, 1981–1987.
- Giancotti, V., Buratti, E., Perissin, L., Zorzet, S., Balmain, A., Portella, G., Fusco, A. & Goodwin, G. H. (1989) *Exp. Cell Res.* 184, 538–545.
- Elton, T. & Reeves, R. (1986) *Anal. Biochem.* 157, 53–62.
- Johnson, K., Lehn, D. A., Elton, T., Barr, P. & Reeves, R. (1988) *J. Biol. Chem.* 263, 18338–18342.
- Vartiainen, E., Palvimo, J., Mahonen, A., Linnala-Kankkunen, A. & Mäenpää, P. H. (1988) *FEBS Lett.* 228, 45–48.
- Eckner, R. & Birnstiel, M. L. (1989) *Nucleic Acids Res.* 17, 5947–5959.
- Johnson, K. R., Disney, J. E., Wyatt, C. R. & Reeves, R. (1990) *Exp. Cell Res.* 187, 69–76.
- Fashena, S. J., Reeves, R. & Ruddle, N. H. (1992) *Mol. Cell. Biol.* 12, 894–903.
- Skalnik, D. G. & Neufeld, E. J. (1992) *Biochem. Biophys. Res. Commun.* 187, 563–569.
- Thanos, D. & Maniatis, T. (1992) *Cell* 71, 777–789.
- Grunstein, M. (1990) *Annu. Rev. Cell Biol.* 6, 643–678.
- Zlatanova, J. S. (1990) *Trends Biochem. Sci.* 15, 273–276.
- Nacheva, G. A., Gushin, D. Y., Preobrazhenskaya, O. V., Karpov, V. L., Ebralidse, K. K. & Mirzabekov, D. A. (1989) *Cell* 58, 27–36.

25. Weintraub, H. (1984) *Cell* 38, 17-27.
26. Zlatanova, J. S. (1990) *Mol. Cell. Biochem.* 92, 1-22.
27. Cole, R. D. (1987) *Int. J. Pept. Protein Res.* 30, 433-449.
28. Churchill, M. E. A. & Travers, A. A. (1991) *Trends Biochem. Sci.* 16, 92-97.
29. Boiocchi, M., Dolcetti, R., Maestro, R., Feriotto, G., Rizzo, S., De Re, V. & Sonogo, F. (1990) *Leukemia Res.* 14, 549-558.
30. Dolcetti, R., Carbone, A., Barlati, S., Maestro, R., Rizzo, S., Sonogo, F., Feriotto, G., Bellotti, D. & Boiocchi, M. (1990) *Int. J. Cancer* 45, 928-934.
31. Covey, T. R., Bonner, R. F., Shushan, B. I. & Henion, J. D. (1988) *Rapid Commun. Mass Spectrom.* 2, 249-256.
32. Ferranti, P., Malorni, A., Marino, G., Pucci, P., Goodwin, G. H., Manfioletti, G. & Giancotti, V. (1992) *J. Biol. Chem.* 267, 22486-22489.
33. Crippa, M. P., Pash, J. M., Gerwin, B. I., Smithgall, T. E., Glazer, R. I. & Bustin, M. (1990) *Cancer Res.* 50, 2022-2026.
34. Bussemakers, M. S. G., van de Ven, W. J. M., Debruyne, F. M. J. & Schalken, J. A. (1991) *Cancer Res.* 51, 606-611.
35. Lennox, R. W., Oshima, R. G. & Cohen, L. H. (1982) *J. Biol. Chem.* 257, 5183-5189.
36. Lennox, R. W. & Cohen, L. H. (1983) *J. Biol. Chem.* 258, 262-268.
37. Lennox, R. W. (1984) *J. Biol. Chem.* 259, 669-672.
38. Palyga, J. (1990) *Int. J. Biochem.* 22, 1351-1361.
39. Lindner, H., Hellinger, W. & Puschendorf, B. (1990) *Biochem. J.* 269, 359-363.
40. Lindner, H. & Hellinger, W. (1990) *Chromatographia* 30, 518-522.
41. Tchouatcha-Tchouassom, J. C., Julliard, J. H. & Roux, B. (1989) *Biochem. Biophys. Acta* 1009, 121-128.
42. Alonso, A., Breuer, B., Bouterfa, H. & Doenecke, D. (1988) *EMBO J.* 7, 3003-3008.
43. Csordas, A. (1990) *Biochem. J.* 265, 23-38.
44. Yang, Y. S., Brown, D. T., Wellman, S. E. Sittman, D. B. (1987) *J. Biol. Chem.* 262, 17118-17125.
45. Seyedin, S. M. & Kistler, W. S. (1980) *J. Biol. Chem.* 255, 5949-5954.
46. Aijro, K., Shibata, K. Nishikawa, Y. (1990) *J. Biol. Chem.* 265, 6494-6500.
47. Bouliskas, T. (1990) *J. Biol. Chem.* 265, 14638-14647.
48. Jerzmanowski, A. & Coles, R. D. (1992) *J. Biol. Chem.* 267, 8514-8520.
49. Gjerset, R., Gorka, C., Hasthorpe, S., Lawrence, J. J. & Eisen, H. (1982) *Proc. Natl Acad. Sci. USA* 79, 2333-2337.
50. Pieler, C., Adolf, G. R. & Swetly, P. (1981) *Eur. J. Biochem.* 115, 329-333.
51. Panyim, S. & Chalkley, R. (1969) *Biochem. Biophys. Res. Commun.* 37, 1042-1049.
52. Houde, M., Shmookler Reis, R. J. & Goldstein, S. (1989) *Exp. Cell Res.* 184, 256-261.
53. Hall, J. M. & Cole, R. D. (1986) *J. Biol. Chem.* 261, 5168-5174.
54. Lund, T. & Laland, S. G. (1990) *Biochem. Biophys. Res. Commun.* 171, 342-347.
55. Reeves, R., Langan, T. A. & Nissen, M. S. (1991) *Proc. Natl Acad. Sci. USA* 88, 1671-1675.
56. Langan, T. A., Goutier, J., Lohka, M., Hollingsworth, R., Moreno, S., Nurse, P., Maller, J. & Sclafani, R. A. (1989) *Mol. Cell. Biol.* 9, 3860-3868.
57. Chambers, T. & Langan, T. A. (1990) *J. Biol. Chem.* 265, 16940-16947.
58. Solomon, M., Strauss, F. & Varshavsky, A. (1986) *Proc. Natl Acad. Sci. USA* 83, 1276-1280.
59. Strauss, F. & Varshavsky, A. (1984) *Cell* 37, 889-891.
60. Elton, T. S., Nissen, M. S. & Reeves, R. (1987) *Biochem. Biophys. Res. Commun.* 143, 260-265.
61. Reeves, R., Elton, T. S., Nissen, M. S., Lehn, D. & Johnson, K. R. (1987) *Proc. Natl Acad. Sci. USA* 84, 6531-6535.
62. Reeves, R. & Nissen, M. S. (1990) *J. Biol. Chem.* 265, 8576-8582.
63. Churchill, M. E. A. & Suzuki, M. (1989) *EMBO J.* 8, 4189-4195.
64. Izaurralde, E., Kas, E. & Laemmli, U. K. (1989) *J. Mol. Biol.* 210, 573-585.
65. Disney, J. E., Johnson, K. R., Magnuson, N. S., Sylvester, S. R. & Reeves, R. (1989) *J. Cell Biol.* 109, 1975-1982.
66. Langan, T. A. (1982) *J. Biol. Chem.* 257, 14835-14836.
67. Lennox, R. W. Cohen, L. H. (1988) in *Chromosomes and chromatin* (Aldolph, K. W., ed.) pp. 33-56, CRC Press, Boca Raton, FL.
68. Matsukava, T., Adachi, H., Kurashina, Y. & Ohba, Y. (1985) *J. Biochem.* 98, 695-704.
69. Higurashi, M., Adachi, H. & Ohba, Y. (1987) *J. Biol. Chem.* 262, 13075-13080.
70. Pearson, R. B. & Kemp, B. E. (1991) in *Methods in enzymology, protein phosphorylation* (Hunter, T. & Sefton, B. M., eds) pp. 62-81, Academic Press, New York.
71. Allan, J., Hartman, P. G., Crane-Robinson, C. & Aviles, F. X. (1980) *Nature* 288, 675-679.

EXPRESSION AND cDNA CLONING OF HUMAN HMGI-C PHOSPHOPROTEIN

Umesh A. Patel¹, Antonella Bandiera², Guidalberto Manfioletti², Vincenzo Giancotti², Kai-Yin Chau¹
and Colyn Crane-Robinson^{1*}

¹ Biophysics Laboratories, University of Portsmouth, PO1 2DT, UK

² Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole, Università di Trieste, 34127, Italy

Received March 31, 1994

SUMMARY. The HMGI family contains three members: I, Y and I-C. HMGI and HMGY are alternative splicings of the same gene and are essential transcription factors at several genetic loci. HMGI-C is transcribed from a different gene and is observed only in highly transformed cells. This work shows that human I-C is present in a more restricted range of cell types than I/Y and is absent from hemopoietic cells, as noted for mouse I-C. However, high expression in a human hepatoma line allowed the cloning of the cDNA and 812 bp of 5'-untranslated, 330 bp of coding and 58 bp of 3'-untranslated DNA were sequenced. The open reading frame showed 4 amino acid substitutions and one additional amino acid when compared to mouse I-C, none of them in the basic DNA binding motifs. © 1994 Academic Press, Inc.

The high mobility group (HMG) proteins consist of three distinct families, HMG1/2, HMG14/17 and HMGI, a common feature being their solubility in 5% perchloric acid (1). The mouse HMGI family consists of three members: HMGI and HMGY are alternatively spliced forms (2), whilst HMGI-C is closely related to I and Y in structural organisation having about 50% sequence identity and expressed from a different gene (3). Rodent HMGI, Y and I-C proteins are preferentially expressed in rapidly dividing, malignant and/or undifferentiated cells and levels of expression in cultured cells correlate with the degree of transformation (4-6). Whereas I and Y may be present in untransformed cells at very low levels, I-C only appears on transformation (5,7). Several studies have implied or directly shown that HMGI/Y are ancillary transcription factors (8-12). In particular, Thanos and Maniatis (12) showed that HMGI/Y are components of the human β -interferon (IFN- β) promoter complex and together with NF- κ B are essential for viral induction of the gene. Furthermore, HMGI/Y is capable of binding to ATF-2 and c-Jun, both bZIP proteins required for viral induction and it was proposed that HMGI/Y facilitates the formation of a multiprotein complex at the IFN- β promoter (13). The binding of HMGI/Y to promoters is mediated by three basic segments of 9 amino acids each (14) which are also present in mouse HMGI-C (3). All three HMGI proteins have a highly acidic C-terminal tail, the site of multiple

* Corresponding author. Fax: (44)-705 842053.

phosphorylations (15-17), which together with the remaining segments of the proteins must be the sites of interaction with other transcriptional factors participating in the multiprotein complexes of the combinational promoters in which the HMGI family participates. Human HMGI-C has been detected in a hepatoma cell line (18) and the present work was undertaken to define more fully the expression pattern of HMGI-C in human cells and clone the human gene as its cDNA.

MATERIALS AND METHODS

Cell culture. Human hepatoma cell lines Hep G2, Hep 3B, PLC/PRF/5 were grown in Dulbecco's modified Eagles medium supplemented with 10% heat inactivated foetal calf serum (FCS). Human erythroleukaemic K562 and lymphoblastoid leukaemic MOLT-4 cells were grown in RPMI 1640 medium supplemented with FCS.

Construction and screening of a human hepatoma cDNA library. An oligo(dT) primed and a random plus specific primed cDNA library was constructed using a cDNA synthesis kit (Pharmacia). Two separate first strands were synthesised using reverse transcriptase on human hepatoma mRNA template from PLC/PRF/5 cells. One first strand was synthesised using an oligo(dT) primer and the other synthesised using a mixture of random hexamers plus UP13, a gene specific primer (5'-GAGGATGTCTCTTCAG T-3') corresponding to a segment near the carboxy terminus of the mouse I-C sequence. The cDNAs from the two syntheses were mixed, size selected, purified on a Sephacryl S-400 spun column and then ligated with phosphorylated EcoR1/Not1 adaptors into EcoR1 digested λ max1 arms (Clontech) before *in vitro* packaging using Giga Pack II Gold (Stratagene). 8×10^5 plaques from the unamplified cDNA library were transferred to four duplicate sets of Hybond-N (Amersham) membranes. The DNA was UV crosslinked and screened for the presence of HMGI-C sequences using as probe a 224 bp fragment (clone HC29, see below), labelled by random priming to a specific activity of 3×10^8 cpm/ μ g. Membranes were hybridised in QuickHyb (Stratagene) containing the denatured labelled probe at 1×10^6 cpm/ml and 400 μ g/ml sonicated salmon sperm DNA at 65°C for 4 hr. Membranes were then washed twice with 2xSSC at room temperature, once with 2xSSC, 0.1% SDS at 65°C for 30 min and finally once with 0.2xSSC, 0.1% SDS at 65°C for 10 min and autoradiographed.

Generation of a human 224 bp DNA probe using reverse transcriptase and PCR (RT-PCR). Two degenerate oligonucleotide primers were designed on the basis of the mouse HMGI-C sequence: 5'-CGCGGGATCCCAGCA(G/A)CA(A/G)GA(G/A)CCAACCTG(T/C)GA(G/A)CC-3' (UP5) and 5'-GCGCGAATTCTC(T/C)TC(T/C)TCNGCNGA(T/C)TC(T/C)TG-3' (UP8). First strand synthesis was at 42°C for 1 hr in 20 μ l containing 1 μ g total human hepatoma RNA from PLC/PRF/5 cells, 50 pmol of UP8, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 1 mM dNTP and 200 units M-MLV reverse transcriptase (BRL). The reaction mixture was then heated at 95°C for 3 min and made up to 100 μ l containing 50 pmol of UP5, 50 pmol of UP8, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂ and 2.5 units Amplitaq DNA polymerase (Cetus). 30 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1 min plus one final cycle of extension at 72°C for 5 min were carried out in a temperature cycler (Coy, USA). The resulting 224 bp DNA fragment was ligated into BlueScript KS⁺ (Stratagene) and transformed into *E. coli* XL1-blue. The sequence of clone HC29 showed a high level of identity to the mouse HMGI-C cDNA, demonstrating that it represented a segment of the human homologue. Clone HC9 which was used as probe for Northern analysis was also prepared by the above RT-PCR method.

RT-PCR cloning of the 3' end of the coding sequence. Two antisense oligonucleotide primers (UP14 & UP15) were designed on the basis of the mouse HMGI-C 3'-untranslated sequence at 58 and 90 bases downstream of the stop codon. The sense primer represented the human HMGI-C sequence QQEPTGEP. The RT-PCR reaction was carried out as described above. The resulting 306 bp fragment from the use of UP14 was ligated into M13mp18 and sequenced. Clone HC37 gave identical sequence to HC29 in the common coding region plus an additional 36 bp sequence up to the stop codon, followed by 58 bp of 3'-untranslated sequence up to UP14.

Protein extraction and polyacrylamide gel electrophoresis. HMG proteins and histone H1 were selectively extracted from culture cells and tumours with 5% (mass/vol) perchloric acid (PCA) and acetone precipitated. First-dimension PAGE (250 mm x 0.7 mm), second-dimension SDS gel electrophoresis (250 mm x 1 mm) and silver staining were carried out as described (18,19).

Preparation and purification of anti-HMGI-C antibodies. Rabbits were injected with 200 μ g of HPLC-purified recombinant mouse HMGI-C protein, (expressed in bacteria), mixed with complete Freund's adjuvant. They were boosted with the same amount of protein in incomplete Freund's adjuvant every 3 weeks for 2 months. For affinity purification of antibodies, Affi-Prep 10 (BioRad) was coupled with 0.8 mg of HMGI-C recombinant protein. 5 ml of immune serum, diluted 2x with phosphate-buffered saline (PBS), was incubated with the affinity matrix in the presence of 0.1% Triton X-100. Elution of antibodies used 0.2 M glycine-HCl buffer (pH 2.8). Rabbit polyclonal antibodies against the peptide WQQVQKKPAQ of the mouse HMGI-C were obtained from Research Genetics (USA) and affinity purified as above.

Western blotting. First-dimension acetic acid-urea gels (without fixing or staining) were incubated for 30 min in 0.75 M Tris base, 0.1% SDS (pH \approx 11). Proteins were transferred for 40 min to a membrane (PVDF, Immobilon P) using a Trans-Blot SD Semi-Dry Transfer Cell (BioRad) in 0.075 M Tris, 0.01% SDS and stained with 0.2% Ponceau S in 3% trichloroacetic acid. Membranes for development with antibody were first saturated with 0.05 M Tris pH 7.5, 0.5 M NaCl, 5% (mass/vol) non-fat dry milk, 0.05% Tween 20, 0.2% BSA, then incubated with anti-HMGI-C affinity purified antibodies (diluted 1:30) for 2 hr at room temperature. Membranes were then washed three times with the above saturation solution and incubated with goat anti-rabbit IgG conjugated with horseradish peroxidase (Pierce), diluted 1:1000, for 1 hr at room temperature. After washing with the same buffer, they were rinsed with 0.2 M Tris pH 7.5, 0.5 M NaCl and developed with 0.3% (mass/vol) 4-chloro-1-naphthol in cold methanol mixed 1:5 with 0.02 M Tris pH 7.5, 0.5 M NaCl. Finally, 30% H₂O₂ solution was added (1:2000) and the reaction stopped by washing with PBS.

RESULTS AND DISCUSSION

Expression Patterns. Fig. 1A shows Coomassie stained gels of perchloric acid extracted proteins from 3 hepatoma cell lines (PLC/PRF/5, Hep G2 and Hep 3B) and 2 lines representative of a total of 5 human transformed hemopoietic cells (K562, MOLT-4, SUP T1, KG1 and KG1A). Protein I-C appears to be present in PLC/PRF/5 and Hep 3B cells and can just be seen in Hep G2. To make the presence of I-C in Hep G2 cells more evident, the 2-D gel was also stained with silver. In contrast, no I-C was detected in the extracts from K562 or MOLT-4 cells and this was confirmed by silver staining (data not shown). To be certain of the identity of the spots, western blotting was carried out using antibodies against recombinant mouse HMGI-C and a peptide representing an I-C specific sequence in the mouse protein,

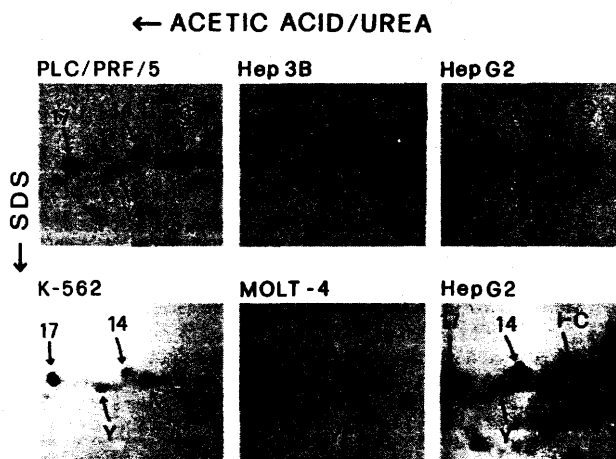


Figure 1A. Two-dimensional gel electrophoresis. Proteins were extracted with PCA from 3 human hepatoma cell lines (PLC/PRF/5, Hep 3B and Hep G2) and 2 hemopoietic cell lines, K562 and MOLT-4. The bottom right panel of Hep G2 was silver stained and the other 5 were Coomassie Blue stained.

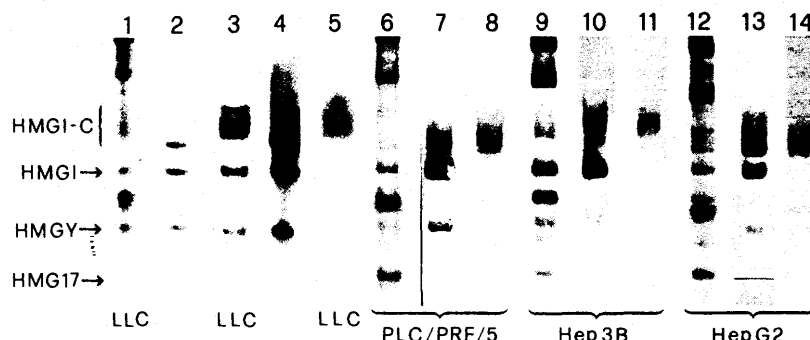


Figure 1B. Western transfers of HMG proteins. Separation was on acetic acid-urea gels, developed with anti (mouse) HMG I-C antibodies (lanes 3, 4, 7, 10, 13) or with anti-I-C peptide antibodies (lanes 5, 8, 11, 14). The Coomassie stained gels from each cell type are shown in lanes 1, 6, 9, 12. Lanes 2 (Coomassie stained) and 4 are an artificial mixture of mouse HMG I, HMG Y and HMG I-C proteins.

not present in HMG I or Y (3). Proteins from the three human hepatoma cell lines were compared with extracts from mouse Lewis lung carcinomas (LLC) and with an artificial mixture of HMG I, HMG Y and HMG I-C proteins purified by HPLC from LLC cells. Lanes 1 and 2 of Fig. 1B show Coomassie staining and lanes 3 to 5 show immunostaining of the mouse proteins. It is clear that the anti-recombinant-I-C antibodies recognise all 3 proteins (lanes 3 and 4). In contrast, the anti-I-C peptide recognises only I-C protein (lane 5) and emphasises the fact that extracted I-C protein consists of several components, probably a consequence of multiple phosphorylations. A similar situation is seen with the human proteins and the data show the presence of I-C in all 3 hepatoma lines. However, western transfers of perchloric extracts from K562, MOLT-4, SUP T1, KG1 and KG1A showed no indication of the presence of I-C protein when treated with the anti-peptide antibody (data not shown). Northern analysis of RNA from the above cell types was conducted using as probe the human HMG I-C cDNA clone HC9. Total RNA was extracted from cells in exponential growth and the data confirmed the results of western blotting: HMG I-C mRNA was detected in PLC/PRF/5, Hep G2 and Hep 3B cells but not in K562, MOLT-4, SUP T1, KG1 or KG1A cells (data not shown). Thus although high levels of I and Y are seen in all human transformed cells, I-C is expressed in a restricted range of cell types. This is similar to mouse cells for which I-C expression is also more restricted than I and Y (7).

cDNA Cloning. A cDNA library was constructed in two parts as described and these mixed before ligating into λ .max1 arms and *in vitro* packaging. The resulting cDNA library was probed with the 224 bp fragment of human I-C cDNA (clone HC29) obtained from RT-PCR by using as template another RNA sample from PLC/PRF/5 cells (see Materials and Methods). Primary and secondary screening of the library produced 8 positive clones. After restriction mapping, the two longest clones HC31 (1075 bp) and HC11 (1095 bp) were sequenced, HC31 on both strands and HC11 fully on one strand and partially on the other. This showed that both clones had resulted from priming with the specific oligonucleotide (UP13) and all the remaining 6 clones were likewise found to have been primed by the specific oligonucleotide. To obtain sequence 3' to the specific primer, RT-PCR was used with total RNA from PLC/PRF/5 cells and reverse primers selected from the 3'-untranslated sequences of mouse I-C. One of these RT-PCR amplifications (clone HC37) was successful and this was sequenced to reveal the remainder of the coding sequence and 58 nucleotides of 3'-untranslated sequence.

```

                20                40                60
TCTTGAATCTTGGGGCAGGAACTCAGAAAACCTCCAGCCCGGGCAGCGCGCTTGGTGC
                80                100               120
AAGACTCAGGAGCTAGCAGCCCGTCCCCCTCCGACTCTCCGGTCCCGCGCTGCCTGCTC
                140               160               180
CCGCCACCTTAGGAGGCGCGGTGCCACCCACTACTCTGTCTCTGCTCTCCGTGCC
                200               220               240
CGACCCTATCCCGGCGGAGTCTCCCCATCTCTCTTTGCTTTCCGACTGCCCAAGGCACTT
                260               280               300
TCAATCTCAATCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
                320               340               360
CTCTCTCTCTCGCAGGGTGGGGGAAGAGGAGGAGGAATTCTTCCCGCCCTAACATTTC
                380               400               420
AAGGGACACAATTCACCTCCAAGTCTTCCCTTTCCAAGCCGCTTCCGAAGTGTCCCGG
                440               460               480
TGCCCGCAACTCCTGATCCCAACCCCGAGAGGAGCCTCTGCCGACCTCAAAGCCTCTCTT
                500               520               540
CCTTCTCCCTCGCTFCCCCTCCCTCTTGTACTCTCCACTCCACCGCCACCTCCACCTC
                560               580               600
CGGCACCCACCCACCGCCCGCCCGCCACCGGCAGCGCCCTCCTCCTCTCCTCCTCCT
                620               640               660
CCCCCTCTCTCTTTTGGCAGCCGCTGGACGTCGGTGGTGGTGGCAGCGCGGCAG
                680               700               720
CCTAAGCAACAGCAGCCCTCGCAGCCCGCCAGCTCGCGCTCGCCCCCGCGCGTCCCCAG
                740               760               780
CCCTATCACCTCATCTCCCGAAAGGTGCTGGGCAGCTCCGGGGCGGTGAGGCCGAAGCGG
                800               820               840
CTGCAGCGGGGTAGCGCGCGCGGAGGCAGGATGAGCCACCGCGGTGAGGGCCGGGGC
                Human   M S A R G E G A G Q
                Mouse  M S A R G E G A G Q

                860                880                900
AGCGCTCCACTTCAGCCAGGGACAACCTGCCGCCCCAGCGCCTCAGAAGAGAGGACGG
P S T S A Q G Q P A A P [A] P Q K R G R G
P S T S A Q G Q P A A P [V] P Q K R G R G

                920                940                960
GCCGCCCCAGGAAGCAGCAGCAAGAACCAACCGGTGAGCCCTCTCCTAAGAGACCCAGGG
R P R K Q Q Q E P T [G] E P S P K R P R G
R P R K Q Q Q E P T [C] E P S P K R P R G

                980                1000               1020
GAAGACCCAAAGGCAGCAAAAACAAGAGTCCCTCTAAAGCAGCTCAAAGAAAGCAGAAG
R P K G S K N K S P S K A A Q K K A E [A]
R P K G S K N K S P S K A A Q K K A E [T]

                1040               1060               1080
CCACTGGAGAAAAACGGCCAAGAGGCAGACCTAGGAAATGGCCACAACAAAGTTGTTGAGA
[T] G E K R P R G R P R K W P Q Q V V Q K
[I] G E K R P R G R P R K W P Q Q V V Q K

                1100               1120               1140
AGAAGCCTGCTCAGGACGAAACTGAAGAGACATCCTCACAAGAGTCTGCCGAAGAGGACT
K P A Q E [E] T E E T S S Q E S A E E D *
K P A Q E [-] T E E T S S Q E S A E E D *

                1160               1180               1200
AGGGGGCCCAACGTTTCGATTCTACCTCAGCAGCAGTTGGATCTTTTGAAGGGAGAAGA

```

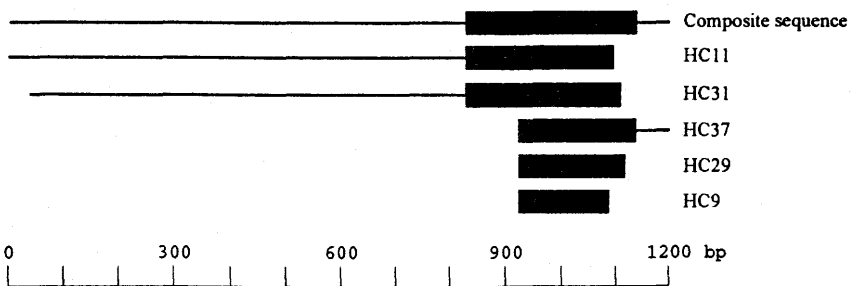
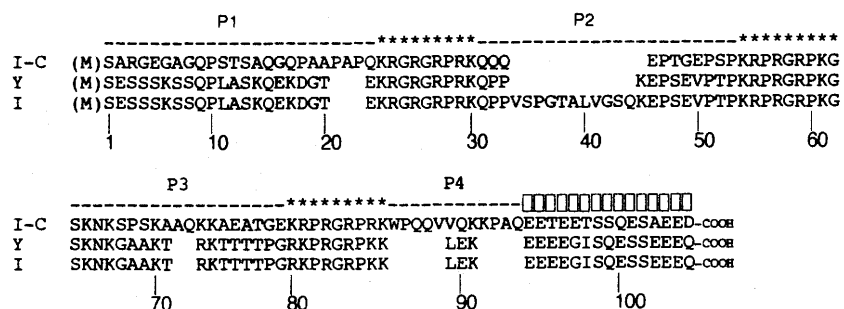


Figure 2. The composite nucleotide and deduced amino acid sequence of the human HMGI-C cDNA. Comparison between the human HMGI-C and the mouse HMGI-C amino acid sequences is given with differences shown in boxes. Shown below the sequences is an alignment of the human HMGI-C cDNA clones analysed. Filled segments represent coding regions.

The composite sequence is shown in Fig. 2. The translational start assumed corresponds to that determined for mouse HMGI-C. The next ATG is at 171 nucleotides upstream (and therefore in frame), but there is a TAA stop codon 7 triplets downstream of this ATG. The cDNA sequence obtained contains an unusually long 5'-untranslated segment, a striking feature of which is a continuous tract of (CT)₂₈ (starting at C252), interrupted by a single GT (at 279) and omission of a single C at 256. This sequence could adopt an unusual structure at either the DNA or RNA level and be responsible for some aspect of transcriptional or translational control. Since only a total of 1.2 kb has been sequenced out of an mRNA of 3.8-4.0 kb (Northern analysis, not shown), an extended 3'-untranslated segment of about 2.8 kb is also present. The 22 nucleotides located just upstream of the translational start are identical in human and mouse I-C, and in the 58 nucleotides 3'-of the stop codon there are only 4 differences between the two mRNAs.

Translation of the coding region shows, in comparison to mouse I-C, 4 changes of amino acid, 22: V to A, 40: C to G, 69: T to A, 70: I to T, and the addition of an extra glutamic acid residue at position 95. In addition to these alterations, there are 18 base changes that result in no change of amino acid. Comparison with the mouse sequence shows that the principal characteristics of HMGI-C that distinguish it from I/Y are maintained in the human protein, in particular the segment WPQQVQKKPAQ and the specific N-terminal domain of 24 amino acids before the first basic segment. The cysteine residue (C40) characteristic of mouse I-C, is however replaced by glycine in human I-C and this accords with the absence of I-C dimers seen in polyacrylamide gels of the human I family proteins (18).

The sequence obtained for human HMGI-C completes the 3 members of the human HMGI family and Fig. 3 shows them aligned for maximum homology. Human HMGI-C is a phosphoprotein but the sites of phosphorylation have not yet been determined, nor have those in mouse I-C. However, for mouse HMGI/Y, *in vivo* phosphorylation of serines 98, 101 and 102 has been established (17). *In vitro* these serines are substrates for casein kinase II (15,16,20). Both human and mouse I-C contain serine at sites



***** REPRESENT THE 3 BASIC DNA BINDING SEGMENTS, ('A-T HOOKS', Ref. 14)

----- REPRESENT THE 4 PEPTIDIC SEGMENTS, P1-P4, THAT COULD MEDIATE PROTEIN/PROTEIN INTERACTIONS

□□□□□□□□ C-TERMINAL ACIDIC TAIL

Figure 3. Human HMGI-C amino acid sequence aligned with human HMGI/Y (24) for maximum homology. The numbering is for HMGI and does not include the initial methionine which is known to be absent in all 3 mouse proteins.

corresponding to S98 and S102 of human HMGI, but S101 is replaced by glutamic acid in both mouse and human I-C and an additional serine in I-C replaces Q99. Phosphorylation due to p34/cdc2 kinase has been reported for T52 and T77 of HMGI (21-23). The potential p34/cdc2 site TPKR (residues 52-55) in I/Y is retained as SPKR in both human and mouse I-C but the site TPGR (residues 77-80) in I/Y is not present in I-C. However, another potential p34/cdc2 phosphorylation site can be found at S67 which is present in both human and mouse I-C but not in I/Y. The total potential sites of phosphorylation for both casein kinase II and p34/cdc2 are thus conserved through the three proteins, but with differences in the localization along the sequence.

Comparison of the mouse and human I-C sequences with those of I and Y shows that the DNA binding regions are fully conserved. Protein I-C can be found in several mouse, rat and human transformed cells but is not detectable either in human transformed cells of the hemopoietic system (this paper) or in mouse lymphomas (7). Protein I-C might therefore replace I/Y at specific promoters or act together with them at promoters requiring more than one copy of an HMGI family molecule.

ACKNOWLEDGMENTS

The financial support of the Wellcome Trust, the Wessex Cancer Trust and the EC Science Programme (CT91-0619) are gratefully acknowledged by U.A.P. and C.C.R. K.-Y.C. acknowledges a Studentship from Hong Kong Polytechnic. The research was also funded by grants to V.G. and G.M. from the Associazione Italiana per la Ricerca sul Cancro, the Consiglio Nazionale delle Ricerche and the Ministero dell'Università e della Ricerca Scientifica e Tecnologica (Italy). We are very grateful to Dr. Graham Goodwin (Institute of Cancer Research, London) for many helpful discussions.

REFERENCES

1. Goodwin, G.H., Sanders, C. & Johns, E.W. (1973) *Eur. J. Biochem.* **38**, 14-19.
2. Johnson, K.R., Lehn, D.A., Elton, T.S., Barr, P.J. & Reeves, R. (1988) *J. Biol. Chem.* **263**, 18338-18342.
3. Manfioletti, G., Giacotti, V., Bandiera, A., Buratti, E., Sautiere, P., Cary, P., Crane-Robinson, C., Coles, B. & Goodwin, G.H. (1991) *Nucleic Acids Res.* **90**, 6793-6797.
4. Giacotti, B.V., Berlingieri, M.T., DiFiore, P.P., Fusco, A., Vecchio, G. & Crane-Robinson, C. (1985) *Cancer Res.* **45**, 6051-6057.
5. Giacotti, V., Pani, B., D'Andrea, P., Berlingieri, M.T., Di Fiore, P.P., Fusco, A., Vecchio, G., Philp, R., Crane-Robinson, C., Nicolas, R.H., Wright, C.A. & Goodwin, G.H. (1987) *EMBO J.* **6**, 1981-1987.
6. Ram, T.G., Reeves, R. & Hosick, H.L. (1993) *Cancer Res.* **53**, 2655-2660.
7. Giacotti, V., Bandiera, A., Ciani, L., Santoro, D., Crane-Robinson, C., Goodwin, G.H., Boiocchi, M., Dolcetti, R. & Casetta, B. (1993) *Eur. J. Biochem.* **213**, 825-832.
8. Lanahan, A., Williams, J.B., Sanders, L.K. & Nathans, D. (1992) *Mol. Cell. Biol.* **12**, 3919-3929.
9. Fashena, S.J., Reeves, R. & Ruddle, N.H. (1992) *Mol. Cell. Biol.* **12**, 894-903.
10. Shalnik, D.G. & Neufeld, E.J. (1992) *Biochem. Biophys. Res. Commun.* **187**, 563-569.
11. Chuvpilo, S., Schomberg, C., Gerwig, R., Heinfling, A., Reeves, R., Grummt, F. & Serfling, E. (1993) *Nucleic Acids Res.* **21**, 5694-5704.
12. Thanos, D. & Maniatis, T. (1992) *Cell* **71**, 777-789.
13. Du, W., Thanos, D. & Maniatis, T. (1993) *Cell* **74**, 887-898.
14. Reeves, R. & Nissen, M.S. (1990) *J. Biol. Chem.* **265**, 8573-8582.

15. Lund, T., Skalhogg, B.S., Holtlund, J. Blomhoff, K.K. & Laland, S.G. (1987) *Eur. J. Biochem.* **166**, 21-26.
16. Palvimo, J. & Linnala-Kankkunen, A. (1989) *FEBS Lett.* **257**, 101-104.
17. Ferranti, P., Malorni, A., Marino, G., Pucci P., Goodwin, G. H., Manfoletti, G. & Giacotti, V. (1992) *J. Biol. Chem.* **267**, 22486-22489.
18. Giacotti, V., Bandiera, A., Buratti, E., Fusco, A., Marzari, R., Coles, B. & Goodwin, G.H. (1991) *Eur. J. Biochem.* **198**, 211-216.
19. Giacotti, V. & Goodwin, G.H. (1986) *J. Biochem. Biophys. Methods* **12**, 265-269.
20. Lund, T., Dahl, K.H., Mork, E., Holtlund, J. & Laland, S.G. (1987) *Biochem. Biophys. Res. Commun.* **146**, 725-730.
21. Lund, T. & Laland, S.G. (1990) *Biochem. Biophys. Res. Commun.* **171**, 342-347.
22. Nissen, M.S., Langan, T.A. & Reeves, R. (1991) *J. Biol. Chem.* **266**, 19945-19952.
23. Reeves, R. Langan, T.A. & Nissen, M.S. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1671-1675.
24. Eckner, R. & Birnstiel, M.L. (1989) *Nucleic Acids Res.* **17**, 5947-5959.

← O.K.

Inhibition of HMGI-C Protein Synthesis Suppresses Retrovirally Induced Neoplastic Transformation of Rat Thyroid Cells

MARIA TERESA BERLINGIERI,¹ GUIDALBERTO MANFIOLETTI,² MASSIMO SANTORO,¹ ANTONELLA BANDIERA,² ROBERTA VISCONTI,¹ VINCENZO GIANCOTTI,² AND ALFREDO FUSCO^{3*}

Centro di Endocrinologia ed Oncologia Sperimentale del Consiglio Nazionale delle Ricerche Dipartimento di Biologia Cellulare e Molecolare, Facoltà di Medicina e Chirurgia, Università degli Studi di Napoli, 80131 Naples,¹ Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole, Università degli Studi di Trieste, 34127 Trieste,² Italy and Dipartimento di Medicina Sperimentale e Clinica, Facoltà di Medicina e Chirurgia di Catanzaro, Università degli Studi di Reggio Calabria, 88100 Catanzaro,³ Italy

Received 10 May 1994/Returned for modification 22 June 1994/Accepted 13 December 1994

Elevated expression of the three high-mobility group I (HMGI) proteins (HMGI, HMGY, and HMGI-C) has previously been correlated with the presence of a highly malignant phenotype in epithelial and fibroblastic rat thyroid cells and in experimental thyroid, lung, mammary, and skin carcinomas. Northern (RNA) blot and run-on analyses demonstrated that the induction of HMGI genes in transformed thyroid cells occurs at the transcriptional level. An antisense methodology to block HMGI-C protein synthesis was then used to analyze the role of this protein in the process of thyroid cell transformation. Transfection of an antisense construct for the HMGI-C cDNA into normal thyroid cells, followed by infection with transforming myeloproliferative sarcoma virus or Kirsten murine sarcoma virus, generated cell lines that expressed significant levels of the retroviral transforming oncogenes *v-mos* or *v-ras-Ki* and removed the dependency on thyroid-stimulating hormones. However, in contrast with untransfected cells or cells transfected with the sense construct, those containing the antisense construct did not demonstrate the appearance of any malignant phenotypic markers (growth in soft agar and tumorigenicity in athymic mice). A great reduction of the HMGI-C protein levels and the absence of the HMGI(Y) proteins was observed in the HMGI-C antisense-transfected, virally infected cells. Therefore, the HMGI-C protein seems to play a key role in the transformation of these thyroid cells.

The high-mobility group I (HMGI) proteins (HMGI, HMGY, and HMGI-C) are small, highly charged polypeptides of about 100 amino acids which constitute an important component of active chromatin structure. Evidence is accumulating of their biological functions (9, 13, 33, 38, 41, 46): in vitro they bind to AT-rich DNA sequences, specifically within the minor groove (12, 34, 35, 38-41) and in particular at scaffold-associated regions, and they are associated with many highly transcribed and regulated genes derepressing their transcription in vitro (46). It has also been demonstrated that the HMGI(Y) proteins are implicated in regulating the expression of tumor necrosis factor beta in murine leukemia virus-transformed cells (13) and can cooperate with the NF- κ B factor, ATF-2 homodimer, and ATF-2/c-Jun heterodimer in the viral induction of the human beta interferon gene (9, 41).

The HMGI and HMGY proteins have been well characterized, and cloning of their cDNAs has shown that they are products of the same gene via an alternative splicing mechanism (10, 15, 26, 27). It is therefore common to designate both proteins and the corresponding gene with the abbreviation HMGI(Y). The molecular cloning of the cDNA for mouse HMGI-C has demonstrated that the HMGI-C protein, although related to HMGI(Y), is the product of a different gene (31).

A correlation between elevated expression of the three HMGI proteins and the appearance of a highly malignant phenotype in differentiated rat thyroid cells transformed with

several oncogenes has been demonstrated (21, 23). Experimental mouse carcinomas induced by dimethylbenzanthracene and rat thyroid carcinomas induced by the Kirsten murine sarcoma virus (KiMSV) exhibited high levels of expression of these three proteins (22). Conversely, all mammalian normal adult tissues so far investigated express undetectable or very low levels of HMGI proteins (19-21, 26, 31). More recently, an association has been demonstrated between elevated expression of the HMGI(Y) gene and progressive transformation of mouse mammary epithelial cells (33), and a correlation has been shown between HMGI(Y) gene overexpression and the ability of rat prostatic carcinoma cell lines to metastasize. Such observations suggest that these proteins might be considered potential tumor progression markers (7). Furthermore, the human homolog of the *Drosophila* trithorax protein (HRX), structurally altered by the 11q23 chromosomal translocations in human acute leukemia, carries an amino-terminal motif related to the DNA-binding motifs of HMGI proteins (42). The correlation between the level of HMGI proteins and the malignant phenotype makes it important to determine their role in the process of cell transformation and in particular to investigate whether their increase is a phenomenon peripherally associated with cell transformation or is in causal relation to the expression of the malignant phenotype.

To approach this question, we have transfected two differentiated rat thyroid cell lines (PC clone [C] 3 and FRTL-5 Cl 2) with an antisense construct for the HMGI-C cDNA. These cells were then infected with myeloproliferative sarcoma virus (MPSV) or KiMSV, both of which are able to drive these normal cells to the malignant phenotype (17, 18). In contrast to untransfected cells or those transfected with the sense construct, virally infected cells containing the antisense construct did not demonstrate the malignant phenotype, even though the

* Corresponding author. Mailing address: Dipartimento di Biologia Cellulare e Molecolare c/o Centro di Endocrinologia ed Oncologia Sperimentale del CNR, Facoltà di Medicina e Chirurgia, Università degli Studi di Napoli, via S. Pansini 5, 80131 Naples, Italy. Phone: 39 81 7463056. Fax: 39 81 7463037.

PC cl 3

→ O.K.

parameters of thyroid differentiation were lost and the viral oncogenes were abundantly expressed.

MATERIALS AND METHODS

Cells and viruses. PC Cl 3 rat thyroid cells are derived from 18-month-old Fischer rats (16), and FRTL-5 cells are derived from 3- to 4-week-old normal Fischer rats (1). The normal and virally infected cells were grown in Coon's modified Ham's F12 medium supplemented with 5% calf serum (Gibco Laboratories) and with a hormone mixture containing 10^{-10} M thyroid-stimulating hormone (TSH), 10 μ g of insulin per ml, 10^{-8} M hydrocortisone, 5 μ g of human transferrin per ml, 10 ng of somatostatin per ml, and 10 ng of glycyl-L-histidyl-L-lysine acetate per ml. Infection with KiMSV and MPSV was carried out as described previously (16-18). All transfections were performed by the calcium phosphate procedure as described elsewhere (24).

Assay of the transformed state. Tumorigenicity of the cell lines was tested by injecting 2×10^6 cells subcutaneously into athymic mice. Soft agar colony assays were performed as described previously (29).

RNA isolation and Northern (RNA) analysis. Total RNA was extracted by a modification of the guanidine thiocyanate method (36). Northern blots and hybridizations were carried out as previously described (37). The probes used were (i) a 680-bp *EcoRI-PstI* fragment corresponding to the 3' end of the rat thyroglobulin (TG) gene (8); (ii) the purified insert from clone TSAFB which represents residues -54 to 2780 of the rat TSH receptor gene, a generous gift of L. D. Kohn (3); (iii) a 0.7-kb *EcoRI-SacI* fragment corresponding to the 5' coding region of the rat thyroperoxidase (TPO) gene; (iv) the insert of the HiHi3 clone as a *v-ras*-Ki-specific probe (11); and (v) *v-mos* oncogene sequences as previously described (44), a generous gift of S. A. Tronick. Finally, HMGI probes were pHMGI(Y) for the HMGI(Y) gene (26) and pHMGI-C11a1 for the murine HMGI-C gene (32).

RT-PCR. RNA amplification was done by the method of Kawasaki (28). Sequences of the sense and antisense HMGI-C primers were as follows: HMGI-C(s), 5'-GCAGCAAAAACAAGAGCC-3' (sense); and HMGI-C(as), 5'-GC GAGGATGTCTCTTCAG-3' (antisense). These primers were designed on the basis of the cDNA sequence of the HMGI-C gene (nucleotides 247 to 264 and 273 to 390, respectively) (31). The expected size of the amplified product was 143 bp. To remove trace amounts of contaminating genomic DNA, RNAs were pretreated with RNase-free DNase (Boehringer Mannheim) for 30 min at 37°C as instructed by the manufacturer. Strand-specific reverse transcriptase PCR (RT-PCR) was performed in order to differentiate the expression of the sense and antisense HMGI-C constructs. Briefly, 5 μ g of total RNA was reverse transcribed by using the sense primer (for expression of the antisense construct) or the antisense primer (for expression of the sense construct), and PCR was performed by subsequently adding the antisense or sense primer, respectively. Thirty cycles of PCR were performed with a thermal cycler (Perkin-Elmer Cetus) (94°C for 30 s, 55°C for 1 min, and 72°C for 2 min). As a control for the quality of the RNA preparations used, expression of the rat glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene was monitored. Oligonucleotide primer sequences used to amplify GAPDH mRNA were GAPDH-F (5'-TTCACCAC CATGGAGAAGGCT-3' [forward]) and GAPDH-R (5'-ACAGCCTTGGCA GCACAGT-3' [reverse]), designed on the basis of the cDNA sequence (nucleotides 369 to 389 and 696 to 715, respectively) (14). To ensure no contamination of RNA samples with DNA, negative controls were obtained by performing the PCRs also on samples that were not reverse transcribed but otherwise identically processed.

Run-on assays. Preparation of nuclei, RNA elongation, and RNA isolation were performed as described previously (25) except that nuclei were frozen at -80°C before use and the labelled RNA was treated with RNase-free DNase (Boehringer Mannheim) for 30 min at 37°C. A total of 2×10^7 nuclei were used for each assay. About 4×10^6 cpm was hybridized to an excess of target DNA (10 μ g) immobilized on a nitrocellulose filter. The hybridization conditions were 50% formamide, 5 \times SSC (1 \times SSC is 0.15 M NaCl plus 0.015 sodium citrate), 5 \times Denhardt's solution, 50 mM phosphate buffer (pH 7.0), 0.1% sodium dodecyl sulfate (SDS), and 10 μ g of pGEM vector DNA per ml at 37°C for 3 days. The following plasmids were used as DNA targets: mouse β -actin (43), pTG (8), pHMGI-C11a1 (31), and pHMGI(Y) (26).

Protein extraction and Western blot (immunoblot) analysis. HMG proteins were selectively extracted by 5% (vol/vol) perchloric acid and precipitated with acetone-HCl, and acetic acid-urea electrophoretic analyses were carried out as elsewhere described (21, 23). Acetic acid-urea gels for Western analysis were obtained by loading each lane with the same amount of total protein (40 μ g). Parallel gels were run to ascertain that equal amounts of protein had in fact been loaded by staining with Coomassie blue and checking the level of histone H1, whose expression can be assumed constant in the different cell lines (20). For blotting experiments, acetic acid-urea gels were incubated in 0.75 M Tris base-0.1% SDS for 15 min and then transferred to an Immobilon-P membrane with a semidry apparatus, using 75 mM Tris base-0.01% SDS for 30 min at 10 V. Transfers were first saturated with 0.05 M Tris (pH 7.5)-0.5 M NaCl-5% (mass/vol) nonfat dried milk-0.05% Tween. 20-0.2% bovine serum albumin and then incubated with anti-HMGI-C affinity-purified antibody (diluted 1:30) for 1 h at room temperature. After incubation, the membranes were washed three times

with the saturation solution and then incubated with goat anti-rabbit immunoglobulin G (heavy plus light chain) horseradish peroxidase-conjugated antibody (1:1,000; Pierce) for 1 h at room temperature. Transfers were then washed with the same buffer, rinsed with 0.02 M Tris (pH 7.5)-0.5 M NaCl, and developed with 0.3% (mass/vol) 4-chloro-naphthol (Sigma) in cold methanol mixed 1:5 with 0.02 M Tris (pH 7.5)-0.5 M NaCl. Finally, H₂O₂ (30% solution) was added (1:2,000), and the reaction was stopped by washing with phosphate-buffered saline.

RESULTS

HMGI gene expression and regulation in transformed thyroid cells. To evaluate the role of the HMGI proteins in cell transformation, we used the thyroid cell system originally developed in our laboratories. Two rat thyroid epithelial cell lines (FRTL-5 Cl 2 and PC Cl 3) expressing thyroid differentiation markers (TG synthesis and secretion, uptake of iodide, and dependence on TSH for growth) have been infected with a wide variety of murine retroviruses carrying different oncogenes. The infection results in loss of the differentiation markers, but the fully transformed phenotype is not always obtained. There are, in fact, cell lines that do not show any of the neoplastic phenotypic markers, other cell lines that show a weak tumorigenic phenotype with the appearance of noninvasive tumors only after 4 to 5 weeks, and finally cell lines which show a highly malignant phenotype since they induce within 7 to 10 days large invasive tumors that lead eventually to animal death (3, 5, 6, 16-18). Table 1 summarizes the cell phenotypes studied.

A correlation has been demonstrated between the presence of the HMGI proteins and the expression of a highly malignant phenotype (21, 23). To investigate the possible molecular mechanism of this association, we analyzed normal and transformed cell lines for the expression of the HMGI(Y) and HMGI-C genes by Northern blot analysis. The data obtained (summarized in Table 1) show that only the highly neoplastic cells express the HMGI genes, and these results therefore confirm those obtained at the protein level for the same cell lines (23). Representative data shown in Fig. 1A demonstrate that HMGI expression is present only in the PC E1A+v-*raf*, PC E1A+v-*abl*, PC MPSV, and FRTL-5 KiMSV cell lines, which are characterized by a highly malignant phenotype, whereas no expression is detectable in normal cells and in PC v-*raf* cells, which also do not show the fully malignant phenotype. Run-on assays were performed to ascertain whether this induction was at the transcriptional level. Figure 1B clearly shows that the transcriptional activity of the HMGI-C and HMGI(Y) genes is high in PC MPSV cells; conversely, in normal PC Cl 3 cells, there is no difference from background levels. Among the control genes, we used the TG gene, which undergoes a transcriptional block in transformed thyroid cells (4). Almost identical results were obtained for normal and KiMSV-transformed FRTL-5 Cl 2 cells (data not shown). The induction of HMGI(Y) and HMGI-C gene expression therefore occurs at the transcriptional level in transformed cells. Since the rate of HMGI synthesis is known to vary during the cell cycle (32), all cell lines were harvested for RNA extraction and run-on experiments when the cells were at a confluent stage. The results of these experiments were confirmed when the quiescent state of growth was obtained by starving the cells for 3 days in absence of serum and six exogenously added growth factors (data not shown). In addition, differences in HMGI expression between normal and transformed thyroid cell lines cannot be ascribed to the different rates of cellular proliferation since in most of the cells analyzed (for instance, normal FRTL-5 Cl 2 and transformed FRTL-5 KiMSV), the rate of proliferation is almost the same between normal and

TABLE 1. Normal and transformed rat thyroid cell lines

Group	Cell Type	Oncogene	TSH dependence	HMGI(Y) and HMGI-C expression ^a
1 (non-tumorigenic cell lines) ^b	PC Cl 3		+	-
	PC Cl 3	E1A	+/-	-
	PC Cl 3	<i>v-ras</i> -Ki	-	-
	PC Cl 3	<i>v-srr</i>	-	-
	PC Cl 3	<i>v-raf</i>	-	-
	FRTL-5 Cl 2		+	-
2 (low-tumorigenic cell lines) ^c	PC Cl 3	Polyomavirus middle T	+/-	-
	PC Cl 3	<i>c-myc</i> + <i>v-ras</i> -Ha	-	-
	PC Cl 3	E1A+ <i>v-ras</i> -Ki	-	-
	PC Cl 3	E1A+ <i>v-ras</i> -Ha	-	-
3 (high-tumorigenic cell lines) ^d	PC Cl 3	<i>c-myc</i> + middle T	-	+
	PC Cl 3	E1A + middle T	-	+
	PC Cl 3	E1A + <i>v-raf</i>	-	+
	PC Cl 3	E1A + <i>v-abl</i>	-	+
	PC Cl 3	<i>v-mos</i>	-	+
	FRTL-5 Cl 2	<i>v-ras</i> -Ha	-	+
	FRTL-5 Cl 2	<i>v-ras</i> -Ki	-	+

^a Analyzed by Northern blotting.

^b No tumors appeared after injection of 2×10^6 cells into six athymic mice.

^c Tumors appeared at least 3 weeks after injection of 2×10^6 cells into athymic mice.

^d Tumors appeared not later than 10 days after injection of 2×10^6 cells into athymic mice.

transformed cells (6, 16, 17). However, to exclude the possibility that these expression levels reflect changes in the fraction of cells that are replicating rather than the transformed state, we analyzed HMGI expression in normal and transformed PC Cl 3 and FRTL-5 Cl 2 cells in the logarithmic phase of growth. As shown in Fig. 2, also in the logarithmic growth phase, the normal thyroid cells (PC Cl 3 and FRTL-5 Cl 2) and the nontumorigenic infected cell lines (PC *v-raf* and PC E1A) were negative for HMGI genes expression; in contrast, transformed PC MPSV, PC E1A+middle T, PC E1A+*v-raf*, and FRTL-5 KiMSV cells were consistently positive also in proliferating phase. Both the quiescent (cells starved for 3 days) and the logarithmic phases were confirmed with a thymidine incorporation assay (not shown).

Transfection of thyroid cells with an antisense HMGI-C construct prevents the appearance of the malignant phenotype. To understand whether the overexpression of HMGI

genes is important in the process of cell transformation or arises as a result of it, we introduced an antisense construct of the HMGI-C cDNA into normal PC Cl 3 and FRTL-5 Cl 2 cells, pursuing the aim of blocking the synthesis of this protein. Figure 3 shows the maps of the constructs used. The pMV-7-HMGI-C sense and antisense constructs were obtained by inserting the HMGI-C cDNA (31) in both sense and antisense orientations in the retroviral vector pMV-7 under the transcriptional control of the long terminal repeats of Moloney murine leukemia virus (30). This vector also carries the gene for geneticin resistance.

Both the sense and antisense constructs were then used to transfect normal PC Cl 3 and FRTL-5 Cl 2 cells. Geneticin-resistant clones were isolated, and expression of the sense (not shown) and antisense HMGI-C constructs (Fig. 4) was analyzed by strand-specific RT-PCR. Three transfectants per line were selected on the basis of their capability to express the

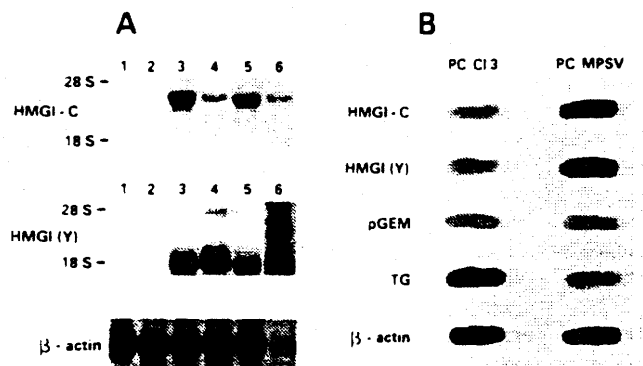


FIG. 1. (A) Northern blot hybridization of total RNA from normal and transformed thyroid cell lines with the HMGI-C (top) and HMGI(Y) (middle) probes. Ten micrograms of total RNA was loaded for each cell line. Sources of RNA were the following: lane 1, PC Cl 3; lane 2, PC *v-raf*; lane 3, PC MPSV; lane 4, PC E1A+*v-raf*; lane 5, PC E1A+*v-abl*; and lane 6, FRTL-5 KiMSV. (B) HMGI-C and HMGI(Y) transcription rates in PC Cl 3 and PC MPSV cells. Nuclear run-on assays were performed to examine transcription rates in PC Cl 3 and PC MPSV cells. Autoradiographs of a representative experiment and the relative positions of the different DNA targets are shown.

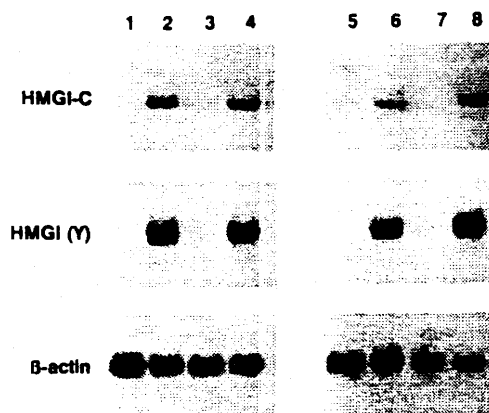


FIG. 2. Expression of the HMGI-C and HMGI(Y) genes in proliferating normal and transformed thyroid cells. Northern blot hybridization of total RNA from normal and transformed thyroid cell lines with the HMGI-C (top) and HMGI(Y) (middle) probes was performed. Ten micrograms of total RNA was loaded for each cell line. Sources of RNA were the following: lane 1, PC Cl 3; lane 2, PC MPSV; lane 3, PC E1A; lane 4, PC E1A+middle T; lane 5, PC *v-raf*; lane 6, PC E1A+*v-raf*; lane 7, FRTL-5 Cl 2; and lane 8, FRTL-5 KiMSV.

AUTHOR:
SEE QUERY
PAGE _____

d BERLINGIERI ET AL.

MOL. CELL. BIOL.

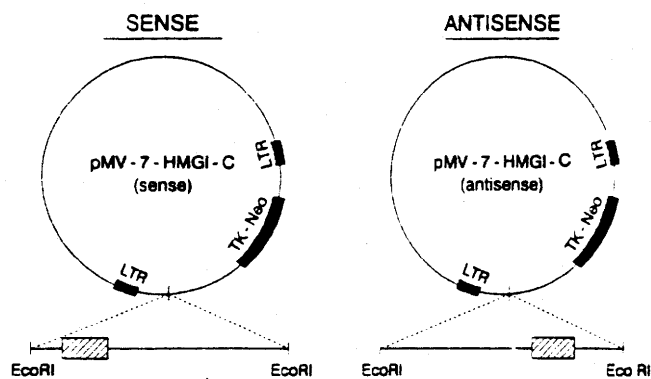


FIG. 3. Partial restriction maps of the pMV-7-HMGI-C sense and antisense constructs. The *EcoRI-EcoRI* fragment including the full-length cDNA clone of the HMGI-C gene has been inserted in the pMV-7 vector in both sense and antisense orientations. LTR, long terminal repeat; TK-Neo, thymidine kinase-neomycin.

SECRET (C)

transfected constructs (PC Cl 3 [antisense] Cl 1, Cl 2, and Cl 3; FRTL-5 [antisense] Cl 1, Cl 2, and Cl 3) and were then assayed for the expression of thyroid differentiation markers: no differences were observed between normal PC Cl 3 and

FRTL-5 Cl 2 cells and transfected cells with respect to parameters of growth and differentiation (Table 2 and data not shown).

The PC pMV-7-HMGI-C sense and antisense and the FRTL-5 pMV-7-HMGI-C sense and antisense transfectants (three per line) were then infected with MPSV and KiMSV carrying, respectively, the *v-mos* and *v-ras*-Ki oncogenes. These viruses were chosen since it has been demonstrated (Table 1) that they are able to induce the fully neoplastic phenotype in PC Cl 3 and FRTL-5 Cl 2 cells, respectively. After infection, the majority of cells showed morphological changes, and they were then selected for the capability to grow in absence of TSH. For both cell types, the TSH-independent clones carrying the pMV-7-HMGI-C antisense construct showed an increased adhesion to the plate in comparison with the analogous infected cells carrying the pMV-7-HMGI-C sense construct (data not shown). Moreover the virally infected cells carrying the pMV-7-HMGI-C antisense construct still showed an epithelial morphology, a feature completely lost in the normal or pMV-7-HMGI-C sense-transfected cells when infected with the same viruses (Fig. 5). Three independent clones and one mass population of infected cells were isolated after infection of each antisense-transfected cell clone.

The growth rate and TSH dependency were analyzed. The growth curves of PC Cl 3 and FRTL-5 Cl 2 cells expressing the sense or antisense HMGI-C construct perfectly overlapped those of untransfected cells (data not shown). Figure 6 shows that the PC pMV-7-HMGI-C antisense, MPSV-infected cells, in contrast with the normal PC Cl 3 cells, were able to grow also in the absence of the six growth factors but with a reduced growth rate in comparison with the untransfected or sense-transfected PC Cl 3 cells infected with MPSV. Likewise, the FRTL-5 pMV-7-HMGI-C antisense cells infected with KiMSV also showed loss of dependency on the six growth factors and a reduced growth rate in comparison with the same infected cells with or without the sense construct (data not shown). No differences have been observed among the different homogeneous clones; therefore these results pertain to one representative clone.

The neoplastic phenotype was evaluated in different cell clones, derived from the infection of different PC Cl 3 and FRTL-5 HMGI-C antisense-transfected clones, by the colony-forming assay in soft agar and by injecting 2×10^6 cells into athymic mice. No tumors were induced after injection of the virally infected cells carrying the pMV-7-HMGI-C antisense construct even after 4 months, nor were the same cells able to grow in soft agar. In contrast, the untransfected cells or those transfected with the sense construct, when infected, showed a high colony-forming efficiency in agar and induced subcutaneous tumors that appeared with a short latency period (1 to 2 weeks) (Table 2).

Expression of the viral oncogenes and of thyroid differentiated functions in HMGI-C sense- and antisense-transfected thyroid cells. To exclude the possibility that the prevention of thyroid cells transformation caused by the pMV-7-HMGI-C antisense construct was due to the lack of expression of viral oncogenes in the cells carrying that construct, we evaluated mRNA levels of the *v-mos* and *v-ras*-Ki oncogenes in the MPSV and KiMSV-infected cells carrying both the HMGI-C sense and antisense constructs. Northern blot analysis revealed no differences in the expression levels of *v-ras*-Ki (Fig. 7A) and *v-mos* (Fig. 7B) oncogenes between HMGI-C sense- or antisense-transfected PC Cl 3 or FRTL-5 Cl 2 cells when infected with the respective retrovirus. Expression of the typical thyroid functions, TG, TPO, and the TSH receptor, was also analyzed in the same cells by Northern blotting (Fig. 8). The TG gene

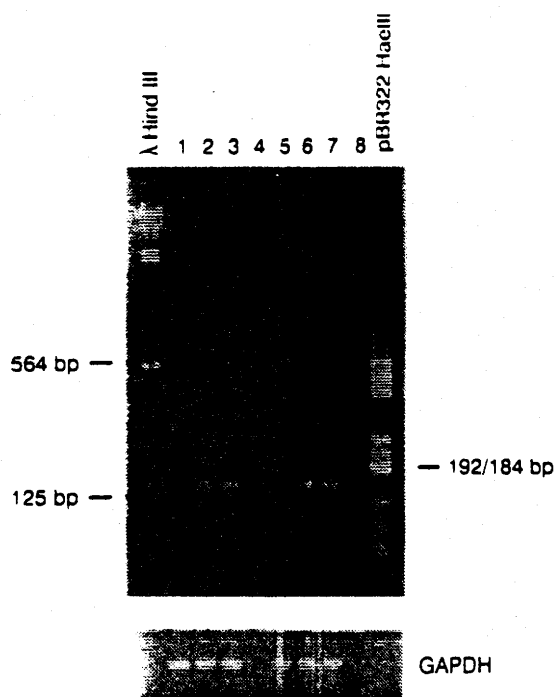


FIG. 4. Expression of the HMGI-C antisense construct in normal and virally infected thyroid cells. The pMV-7-HMGI-C antisense construct was used to transfect normal PC Cl 3 and FRTL-5 Cl 2 cells. Three geneticin-resistant transfectants per line were selected and subsequently infected with MPSV or KiMSV. Expression of the antisense HMGI-C construct in the cell lines obtained was analyzed by strand-specific RT-PCR (for details, see Materials and Methods). Bands of comparable intensity are present in duplicate samples amplified with rat GAPDH sequence-specific primers, suggesting comparable amplification of all samples. No bands are seen in non-reverse-transcribed RNAs, excluding the possibility that were generated from contaminating DNA. Lanes: 1, PC Cl 3; 2, PC pMV-7-HMGI-C antisense Cl 1; 3, PC pMV-7-HMGI-C antisense 1 MPSV Cl 1; 4, PC pMV-7-HMGI-C antisense 1 MPSV Cl 1, non-reverse transcribed; 5, FRTL-5 Cl 2; 6, FRTL-5 pMV-7-HMGI-C antisense Cl 1; 7, FRTL-5 pMV-7-HMGI-C antisense 1 KiMSV Cl 1; 8, FRTL-5 pMV-7-HMGI-C antisense 1 KiMSV Cl 1, non-reverse transcribed. Molecular weight markers are indicated.

TABLE 2. Analysis of the transformed phenotypes of HMGI-C-transfected thyroid cells

Cell line	Oncogene	Colony-forming efficiency (%)	Tumor incidence* (no. with tumors/no. tested)	Latency period
PC Cl 3		0	0/10	
PC (C-as) ^b Cl 1		0	0/10	
PC (C-as) 1 MPSV Cl 1	<i>v-mos</i>	0	0/6	
PC (C-as) 1 MPSV Cl 2	<i>v-mos</i>	0	0/4	
PC (C-as) Cl 2		0	0/4	
PC (C-as) 2 MPSV Cl 1	<i>v-mos</i>	0	0/4	
PC (C-as) Cl 3		0	0/4	
PC (C-as) 3 MPSV Cl 1	<i>v-mos</i>	0	0/4	
→ PC (C-as) (C-sense) ^f		0	0/6	
PC MPSV	<i>v-mos</i>	70	8/8	7-10 days
PC (C-sense) MPSV	<i>v-mos</i>	70	6/6	7-10 days
FRTL-5 Cl 2		0	0/10	
FRTL-5 (C-as) Cl 1		0	0/8	
FRTL-5 (C-as) 1 KiMSV Cl 1	<i>v-ras-Ki</i>	0	0/6	
FRTL-5 (C-as) 1 KiMSV Cl 3	<i>v-ras-Ki</i>	0	0/4	
FRTL-5 (C-as) Cl 2		0	0/4	
FRTL-5 (C-as) 2 KiMSV Cl 1	<i>v-ras-Ki</i>	0	0/4	
FRTL-5 (C-as) Cl 3		0	0/4	
FRTL-5 (C-as) 3 KiMSV Cl 1	<i>v-ras-Ki</i>	0	0/4	
→ FRTL-5 (C-as) (C-sense)		0	0/6	
→ FRTL-5 (C-as) KiMSV	<i>v-ras-Ki</i>	60	5/5	2 wk
→ FRTL-5 (C-as) (C-sense) KiMSV	<i>v-ras-Ki</i>	68	6/6	2 wk

* Assayed by injecting 10⁶ cells into athymic mice (4 to 6 weeks old).

^b The cells were transfected with plasmid pMV-7-HMGI-C antisense carrying the HMGI-C gene in antisense orientation.

^c The cells were transfected with plasmid pMV-7-HMGI-C sense carrying the HMGI-C gene in sense orientation.

was not expressed in any of the infected cells lines; the TSH receptor gene was expressed at very low levels in the infected cells carrying the pMV-7-HMGI-C antisense construct, and expression of the TPO gene was detectable only in the PC pMV-7-HMGI-C antisense cells infected with MPSV (Fig. 8). As already reported (17, 18), no expression of the differentiated thyroid markers was detected in PC Cl 3 or FRTL-5 Cl 2 cells infected with MPSV or KiMSV.

Analysis of the HMGI proteins in normal, transformed, and HMGI-C sense- and antisense-transfected thyroid cells. Western blot analyses of total proteins extracted from normal, transformed, and antisense-transfected cells were carried out (Fig. 9) with polyclonal antibodies raised against the recombinant HMGI-C protein (32). Since HMGI-C shows about 50% sequence identity with the other two HMGI proteins (HMGI and HMGY), the antibodies also recognized the HMGI and HMGY proteins. The antibody was tested both on total HMG extracts from mouse Lewis lung carcinoma (lane 3), which has been demonstrated to express the three HMGI proteins, and on high-pressure liquid chromatography (HPLC)-purified HMGI-C, HMGI, and HMGY proteins (lane 4).

Lanes 1 and 2 show Coomassie blue-stained proteins which were used in lanes 3 and 4 to test the antibodies. Lanes 5 to 9 show Western blot analysis of protein extracted from virally infected cells, transfected or not with the HMGI-C antisense construct. To estimate the amounts of total protein loaded, a parallel gel was run for each sample and stained with Coomassie blue (lanes 5' to 9'). Histone H1, whose expression can be assumed to be constant in the different cell lines, was used as a reference (20).

First-dimension acetic acid-urea electrophoresis, such as in lanes 5 to 9, does not allow the resolution of the variety of components which are present in a perchloric acid extract of cells. This resolution can be achieved by second-dimension SDS electrophoretic analysis (21-23). However, first-dimension acetic acid-urea electrophoresis allows very good resolu-

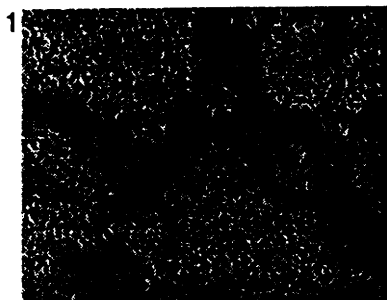
tion of the three proteins HMGI, HMGY, and HMGI-C, which can be easily identified by Western blot analysis. In the first dimension, protein HMGY migrates close to protein HMG14 (which together with protein HMG17 is present in all normal mammalian cells) and to an unknown component that is probably a degradation product, as previously reported (19).

The untransfected, virally infected cells (lanes 6 and 8) showed high levels of the three HMGI proteins in comparison with normal cells (lane 5). In contrast both PC pMV-7-HMGI-C antisense MPSV-infected and FRTL-5 pMV-7-HMGI-C antisense KiMSV-infected cells (lanes 7 and 9) revealed a great reduction in the expression of HMGI-C protein and the complete absence of HMGI and HMGY proteins. The absence of HMGI and HMGY is also evident from the Coomassie stained lanes 7' and 9'.

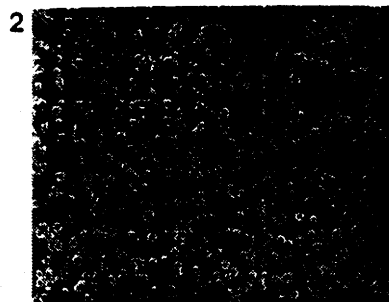
The amount of HMGI(Y)/I-C mRNA in the HMGI-C antisense-transfected, virally infected cells was also evaluated (Fig. 10). The levels of HMGI-C mRNAs (evaluated by hybridization with a riboprobe prepared from the HMGI-C cDNA) were much lower in the transfected cells than in the untransfected, virally infected cells. In accordance with Western data, no expression of the HMGI(Y) gene was observed in the antisense-transfected, virally infected cells. These results were obtained for different clones of virally infected, HMGI-C antisense-transfected thyroid cells. Two hypotheses can be envisaged to account for the absence of HMGI(Y) in these cells. Blockage of HMGI-C protein synthesis, thereby preventing rat thyroid cell transformation, may also suppress the appearance of the HMGI(Y) as a secondary consequence. Alternatively, the HMGI-C antisense construct may also be able per se to block the synthesis of the HMGI(Y) proteins, since there is a fair level of identity between the HMGI-C gene and the HMGI(Y) gene. However, the complete absence of HMGI(Y)-specific mRNA in the antisense-transfected, virally infected cells makes the second hypothesis less likely.

A

PC Cl 3



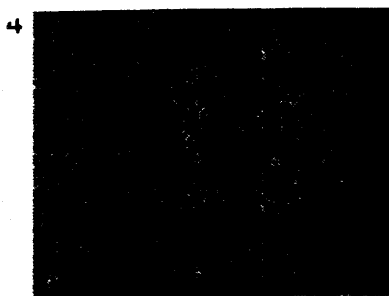
PC MPSV



PC (C-antisense)

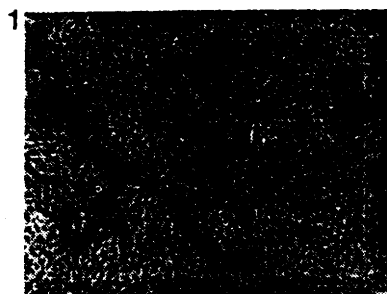


PC (C-antisense) MPSV



B

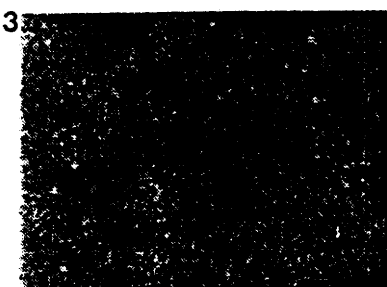
FRTL-5 Cl 2



FRTL-5 (C-sense) KiMSV



FRTL-5 (C-antisense)



FRTL-5 (C-antisense) KiMSV

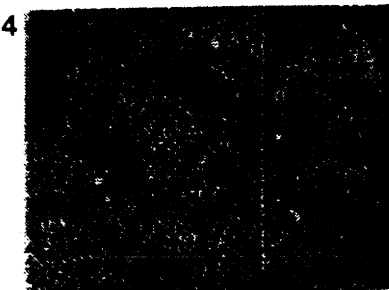


FIG. 5. Phase-contrast photomicrograph of PC Cl 3 and FRTL-5 Cl 2 cells transfected or not with the HMGI-C cDNA and then infected with MPSV and KiMSV. Magnification, $\times 150$. (A) 1, normal PC Cl 3 cells; 2, PC Cl 3 cells transfected with the pMV-7-HMGI-C sense construct and then infected with MPSV; 3, PC Cl 3 cells transfected with the pMV-7-HMGI-C antisense construct; 4, PC Cl 3 cells transfected with the pMV-7-HMGI-C antisense construct and then infected with MPSV. (B) 1, normal FRTL-5 Cl 2 cells; 2, FRTL-5 Cl 2 cells transfected with the pMV-7-HMGI-C sense construct and then infected with KiMSV; 3, FRTL-5 Cl 2 cells transfected with the pMV-7-HMGI-C antisense construct; 4, FRTL-5 Cl 2 cells transfected with the pMV-7-HMGI-C antisense construct and then infected with KiMSV.

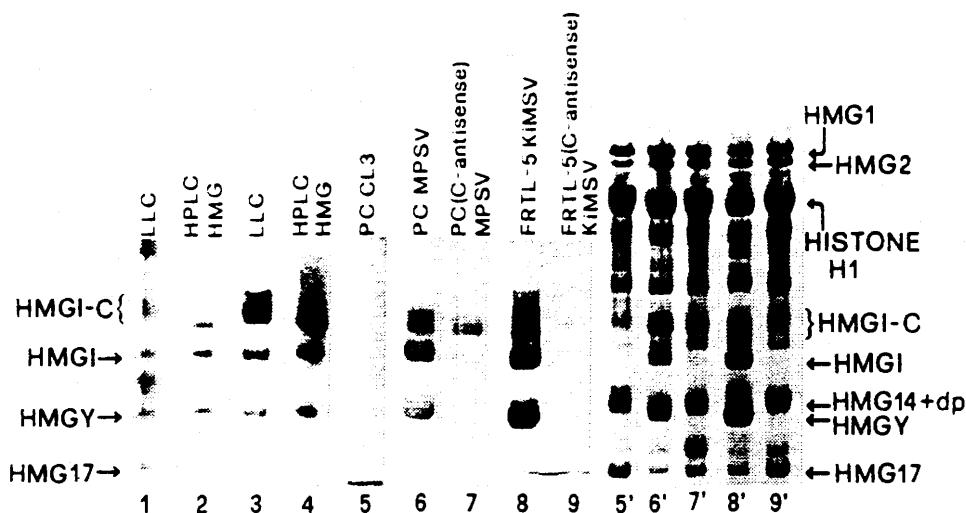


FIG. 9. Acetic acid-urea electrophoresis analyses of perchloric acid-extracted HMG proteins. Lanes: 1 and 2, polyacrylamide gels (15%) Coomassie blue stained; 3 to 9, Immobilon-P blots probed with an antibody directed against recombinant HMGI-C protein; 1 and 3, total protein from Lewis lung carcinoma (LLC); 2 and 4, mixture of the three proteins HMGI, HMGY, and HMGI-C purified by HPLC (HMGI and HMGY were from Lewis lung carcinoma, whereas HMGI-C was recombinant); 5' to 9', quantitation of the total protein loaded for each sample from lanes 5 to 9, using Coomassie blue staining and histone H1 as a reference. dp, presumed degradation product.

progress to investigate whether this phenomenon is caused by the selection of a transformed population of cells which had lost the antisense construct or its expression or if the high level of expression of the HMGI-C gene or the contemporaneous expression of the HMGI(Y) gene by the transformed cells renders the reversion only transitory. The HMGI-C gene does not behave like a classical transforming oncogene since when transfected in normal thyroid cells, it did not cause their transformation, thus suggesting that its expression is necessary but not sufficient to achieve the transformed phenotype; the activation of other factors could be required.

The important role of HMGI proteins in cell transformation does not seem to be restricted to thyroid cells, since recent data obtained in our laboratories indicate that the HMGI-C antisense constructs are also able to prevent neoplastic transformation of murine keratinocytes. The role of the HMGI proteins in human thyroid carcinomas also needs to be elucidated: preliminary data indicate that the HMGI proteins are present in human thyroid carcinoma cell lines (7a) but not in normal thyroid or nonneoplastic thyroid proliferative disorders. An

important role for HMGI proteins in human thyroid carcinomas must therefore be considered.

ACKNOWLEDGMENTS

This study was supported by the Associazione Italiana per la Ricerca sul Cancro, by Progetto Speciale AIRC Oncosoppressori, by Consiglio Nazionale delle Ricerche Progetto ACRO CNR-sottoprogetto Biologia Molecolare, and by the Ministero dell'Università e della Ricerca Scientifica e Tecnologica.

We are indebted to C. Crane-Robinson (University of Portsmouth) for carefully reading and suggesting corrections to the manuscript.

REFERENCES

1. Ambesi-Impombato, F. S., L. A. M. Parks, and H. G. Coon. 1980. Culture of hormone-dependent functional epithelial cells from rat thyroids. *Proc. Natl. Acad. Sci. USA* 77:3455-3459.
2. Angel, P., and M. Karin. 1991. The role of Jun, Fos and AP-1 complex in cell proliferation and transformation. *Biochim. Biophys. Acta* 1072:129-157.
3. Berlingieri, M. T., T. Akamizu, A. Fusco, M. Grieco, G. Colletta, A. M. Cirañci, S. Ikuyama, L. D. Kohn, and G. Vecchio. 1990. Thyrotropin receptor gene expression in oncogene-transfected rat thyroid cells: correlation between transformation, loss of thyrotropin-dependent growth and loss of thyrotropin receptor gene expression. *Biochem. Biophys. Res. Commun.* 173:172-178.
4. Berlingieri, M. T., A. M. Musti, V. E. Avvedimento, R. Di Lauro, P. P. Di Fiore, and A. Fusco. 1989. The block of thyroglobulin synthesis following transformation of rat thyroid cells is at transcriptional level and it is associated with methylation of the 5' flanking region of the gene. *Exp. Cell Res.* 183:277-283.
5. Berlingieri, M. T., G. Portella, M. Grieco, M. Santoro, and A. Fusco. 1988. Cooperation between polyomavirus middle T gene and the human *c-myc* oncogene on a rat thyroid epithelial differentiated cell line: a model of in vitro progression. *Mol. Cell. Biol.* 8:226-230.
6. Berlingieri, M. T., M. Santoro, C. Battaglia, M. Grieco, and A. Fusco. 1993. The adenovirus E1A gene blocks the differentiation of a thyroid epithelial cell line, however the neoplastic phenotype is achieved only after cooperation with other oncogenes. *Oncogene* 8:249-255.
7. Bussemakers, M. J. G., W. J. M. van de Ven, F. M. J. Debruyne, and S. A. Schalken. 1991. Identification of high mobility group protein I (Y) as potential progression marker for prostate cancer by differential hybridization analysis. *Cancer Res.* 51:606-611.
- 7a. Chiappetta, G., et al. Unpublished data.
8. Di Lauro, R., S. Obici, A. Acquaviva, and C. Alvino. 1982. Construction of recombinant plasmids containing rat thyroglobulin mRNA sequences. *Gene* 19:117-125.
9. Du, W., D. Thanos, and T. Maniatis. 1993. The high mobility group protein is required for NF- κ B-dependent virus induction of the human IFN- β gene. *Cell* 74:887-898.

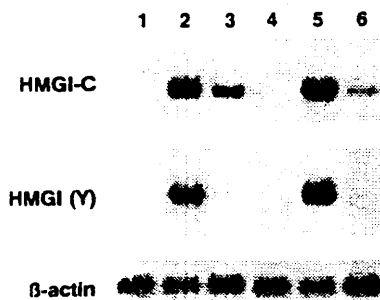


FIG. 10. Northern blot hybridization of total RNA from normal and transformed thyroid cell lines with HMGI(Y) and HMGI-C gene and β -actin probes. Ten micrograms of total RNA was loaded for each cell line. Sources of RNA were the following: lane 1, normal PC Cl 3 cells; lane 2, PC Cl 3 cells infected with MPSV; lane 3, PC Cl 3 pMV-7-HMGI-C antisense cells infected with MPSV; lane 4, normal FRTL-5 Cl 2 cells; lane 5, FRTL-5 Cl 2 cells infected with KIMSV; and lane 6, FRTL-5 Cl 2 pMV-7-HMGI-C antisense cells infected with KIMSV.

Chiappetta, G., Bandiera, A., Berlingieri, M.T., Visconti, R., Manfioletti, G., Battista, S., Francisco J. Martinez-Tello, F. J., Santoro, M., Giancotti, V., and Fusco, A. 1995. The expression of the High Mobility Group HMGI (Y) proteins correlates with the malignant progression of human thyroid neoplasias. *Oncogene* in

10. Eckner, R., and M. L. Birnstein. 1989. Cloning of cDNAs coding for human HMGI and HMGI-Y proteins: both are capable of binding to the octamer sequence motif. *Nucleic Acids Res.* 17:5947-5959.
11. Ellis, R. W., D. De Feo, T. Y. Shih, M. A. Gonda, H. A. Young, N. Tsuchida, D. R. Lowy, and E. M. Scolnick. 1981. The p21 *src* genes of Harvey and Kirsten sarcoma virus originate from divergent members of a family of normal vertebrate genes. *Nature (London)* 292:506-511.
12. Elton, T. S., M. S. Nissen, and R. Reeves. 1987. Specific A-T DNA sequence binding of RP-HPLC purified HMGI-I. *Biochem. Biophys. Res. Commun.* 143:260-265.
13. Fasheha, S. T., R. Reeves, and N. H. Ruddle. 1991. A poly (dA-dT) upstream activating sequence binds high mobility group I protein and contributes to lymphotoxin (tumor necrosis factor- β) gene regulation. *Mol. Cell. Biol.* 12:894-903.
14. Fort, P., L. Marty, M. Piechaczyk, S. El Sabrougy, C. Dani, P. Jeanteur, and J. M. Blanchard. 1985. Various rat adult tissues express only one major mRNA species from the glyceraldehyde-3-phosphate-dehydrogenase multi-gene family. *Nucleic Acids Res.* 13:1431-1442.
15. Friedmann, M., L. T. Holth, H. Y. Zoghbi, and R. Reeves. 1993. Organization, inducible-expression and chromosome localization of the human HMGI-I(Y) non-histone protein gene. *Nucleic Acids Res.* 21:4259-4267.
16. Fusco, A., M. T. Berlingieri, G. Portella, P. P. Di Fiore, M. Grieco, and G. Vecchio. 1987. One- and two-step transformations of rat thyroid epithelial cells by retroviral oncogenes. *Mol. Cell. Biol.* 7:3365-3370.
17. Fusco, A., A. Pinto, D. Tramontano, G. Tajana, G. Vecchio, and N. Tsuchida. 1981. Block in the expression of differentiation markers of rat thyroid epithelial cells by transformation with Kirsten murine sarcoma virus. *Cancer Res.* 42:618-625.
18. Fusco, A., G. Portella, P. P. Di Fiore, M. T. Berlingieri, R. Di Lauro, A. B. Schneider, and G. Vecchio. 1985. A *mos* oncogene-containing retrovirus, myeloproliferative sarcoma virus, transforms rat thyroid epithelial cells and irreversibly blocks their differentiation pattern. *J. Virol.* 56:284-292.
19. Giancotti, V., A. Bandiera, E. Buratti, A. Fusco, R. Marzari, B. Coles, and G. H. Goodwin. 1991. Analysis of multiple forms of the HMGI proteins in rodent and human cells: identification of the third protein, HMGI-C in human cells. *Eur. J. Biochem.* 198:211-216.
20. Giancotti, V., A. Bandiera, L. Ciani, D. Santoro, C. Crane-Robinson, G. H. Goodwin, M. Boiocchi, R. Dolcetti, and B. Casetta. 1993. High-mobility-group (HMG) proteins and histone H1 subtypes expression in normal and tumor tissues of mouse. *Eur. J. Biochem.* 213:825-832.
21. Giancotti, V., M. T. Berlingieri, P. P. Di Fiore, A. Fusco, G. Vecchio, and C. Crane-Robinson. 1985. Changes in nuclear proteins following transformation of rat thyroid epithelial cells by a murine sarcoma retrovirus. *Cancer Res.* 45:6051-6057.
22. Giancotti, V., E. Buratti, L. Perissin, S. Zorzet, A. Balmain, G. Portella, A. Fusco, and G. H. Goodwin. 1989. Analysis of the HMGI nuclear proteins in mouse neoplastic cells induced by different procedures. *Exp. Cell Res.* 184: 538-545.
23. Giancotti, V., B. Pani, P. D'Andrea, M. T. Berlingieri, P. P. Di Fiore, A. Fusco, G. Vecchio, R. Philp, C. Crane-Robinson, R. H. Nicolas, C. A. Wright, and G. H. Goodwin. 1987. Elevated levels of a specific class of nuclear phosphoproteins in cells transformed with *ras* and *v-mos* oncogenes and by co-transfection with *c-myc* and polyoma middle T genes. *EMBO J.* 6:1981-1987.
24. Graham, F. L., and A. J. van der Eb. 1973. A new technique for the assay of the infectivity of human adenovirus 5 DNA. *Virology* 52:456-467.
25. Greenberg, M. E., and E. B. Ziff. 1984. Stimulation of 3T3 cells induces transcription of the *c-fos* proto-oncogene. *Nature (London)* 311:433-438.
26. Johnson, K. R., D. A. Lehen, T. Elton, P. Barr, and R. Reeves. 1988. Complete murine cDNA sequence, genomic structure, and tissue expression of the high mobility group protein HMGI(Y). *J. Biol. Chem.* 263:18338-18342.
27. Johnson, K. R., D. A. Lehen, and R. Reeves. 1989. Alternative processing of mRNAs coding for human HMGI and HMGI-Y proteins. *Mol. Cell. Biol.* 9:2114-2123.
28. Kawasaki, E. S. 1990. A guide to methods and applications, p. 21-38. *In* M. A. Innis, D. H. Gelfand, J. J. Sninski, and T. J. White (ed.), *PCR protocols*. Academic Press, Inc., San Diego, Calif.
29. Macpherson, L., and I. Montagnier. 1964. Agar suspension culture for the selective assay of cells transformed by polyoma virus. *Virology* 23:291-294.
30. Maddon, P. J., A. G. Dagleish, J. S. McDougal, P. R. Clapham, R. Weiss, and R. Axel. 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* 47:333-348.
31. Manfoletti, G., V. Giancotti, A. Bandiera, E. Buratti, P. Sautière, P. Cary, C. Crane-Robinson, B. Coles, and G. H. Goodwin. 1991. cDNA cloning of the HMGI-C phosphoprotein, a nuclear protein associated with neoplastic and undifferentiated phenotypes. *Nucleic Acids Res.* 19:6793-6797.
32. Patel, U. A., A. Bandiera, G. Manfoletti, V. Giancotti, K. Chau, and C. Crane-Robinson. 1994. Expression and cDNA cloning of human high mobility group protein HMGI-C. *Biochem. Biophys. Res. Commun.* 201:63-70.
33. Ram, T. G., R. Reeves, and H. L. Hosick. 1993. Elevated high mobility group I (Y) gene expression is associated with progressive transformation of mouse mammary epithelial cells. *Cancer Res.* 53:2655-2660.
34. Reeves, R., T. S. Elton, M. S. Nissen, D. Lehen, and K. R. Johnson. 1987. Post-transcriptional gene regulation and specific binding of the non-histone protein HMGI-I by the 3' untranslated region of bovine interleukin 2 cDNA. *Proc. Natl. Acad. Sci. USA* 8:6531-6535.
35. Reeves, R., and M. S. Nissen. 1990. The AT DNA binding domain of mammalian high mobility group I chromosomal protein. A novel peptide motif for recognizing DNA structure. *J. Biol. Chem.* 265:8576-8582.
36. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor.
37. Santoro, M., R. Rosati, M. Grieco, M. T. Berlingieri, G. L. Colucci-D'Amato, V. De Francis, and A. Fusco. 1990. The *ret* proto-oncogene is consistently expressed in human pheochromocytomas and thyroid medullary carcinomas. *Oncogene* 5:1595-1598.
38. Skalik, D. G., and E. J. Neufeld. 1990. Sequence-specific binding of HMGI-I(Y) to the proximal promoter of the gp91-phox gene. *Biochem. Biophys. Res. Commun.* 187:563-569.
39. Solomon, M., F. Strauss, and A. Varshavsky. 1986. A mammalian high mobility group recognizes any stretch of six A-T base pairs in duplex DNA. *Proc. Natl. Acad. Sci. USA* 83:1276-1280.
40. Strauss, F., and A. Varshavsky. 1984. A protein binds in a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. *Cell* 37:889-891.
41. Thanos, D., and T. Maniatis. 1992. The high mobility group protein HMGI(Y) is required for NF- κ B dependent virus induction of the human IFN- β gene. *Cell* 71:777-789.
42. Tkachuk, D. C., S. Kohler, and M. L. Cleary. 1992. Involvement of a homolog of *Drosophila* Trithorax by 11q23 chromosomal translocations in acute leukemias. *Cell* 13:691-700.
43. Tokunaga, K., H. Taniguchi, K. Yoda, M. Shimizu, and S. Sakayama. 1986. Nucleotide sequence of a full-length cDNA for mouse cytoskeletal β -actin mRNA. *Nucleic Acids Res.* 14:2829.
44. Tronick, S. R., K. C. Robbins, E. Canaani, S. G. Devare, P. R. Andersen, and S. A. Aaronson. 1979. Molecular cloning of Moloney-murine sarcoma virus: arrangement of virus-related sequences within the normal mouse genome. *Proc. Natl. Acad. Sci. USA* 76:6314-6318.
45. Yang-Yen, H. F., and L. I. Rothblum. 1988. Purification and characterization of a high mobility group like DNA binding protein that stimulates rRNA synthesis in vitro. *Mol. Cell. Biol.* 8:3406-3414.
46. Zhao, K., E. Kas, E. Gonzales, and U. K. Laemmli. 1993. SAR-dependent mobilization of histone H1 by HMGI-Y in vitro: HMGI-Y is enriched in H1-depleted chromatin. *EMBO J.* 12:3237-3247.