



MINISTERO DELLO SVILUPPO ECONOMICO
DIREZIONE GENERALE PER LA TUTELA DELLA PROPRIETÀ INDUSTRIALE
UFFICIO ITALIANO BREVETTI E MARCHI

UIBM

DOMANDA NUMERO	101995900450740
Data Deposito	27/06/1995
Data Pubblicazione	27/12/1996

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
C	12	N		

Titolo

POLYPEPTIDE PER LA RIPARAZIONE DI INFORMAZIONE GENETICA, SEQUENZA
NUCLEOTIDICA CHE CODIFICA PER ESSO E PROCEDIMENTO PER LA SUA PREPARAZIONE.

SIB 90316

Descrizione dell'invenzione industriale dal
titolo: "POLIPEPTIDE PER LA RIPARAZIONE DI
INFORMAZIONE GENETICA, SEQUENZA NUCLEOTIDICA CHE
CODIFICA PER ESSO E PROCEDIMENTO PER LA SUA
PREPARAZIONE "

della ditta Italiana ISTITUTO DI RICERCHE DI
BIOLOGIA MOLECOLARE P. ANGELETTI S.p.A.
con sede a POMEZIA (ROMA)

.....
DESCRIZIONE

La presente invenzione cade nel settore,
della diagnosi e terapia del cancro. In
particolare, l'invenzione si riferisce a metodi
per il rilevamento di una nuova proteina, che si
lega a malaccoppiamenti (d'ora in poi indicati
anche come mismatches), denominata GTBP (Guanine
Timine Binding Protein), la quale media la
riparazione di informazione genetica, alla
sequenza nucleotidica che codifica per questa
proteina ed a procedimenti per ottenere la
suddetta proteina con tecniche di ingegneria
genetica ricombinante. Inoltre, la presente
invenzione ha per oggetto anche la rilevazione del

gene *GTBP* mutato in tessuti tumorali, e la prevenzione e diagnosi precoce di cancro umani colorettali

Come è noto, si è sinora ritenuto che, nelle cellule umane, il riconoscimento e il legame a malaccoppiamenti fossero mediati dalla proteina hMSH2. L'osservazione che cellule di cancro umani colorettali (CRC) esibiscono un fenotipo mutatore con una marcata instabilità delle sequenze microsatelliti ha suggerito che queste cellule tumorali potessero essere deficienti nella riparazione di malaccoppiamenti del DNA. Questa ipotesi è stata confermata quando si è visto che estratti da linee cellulari derivate da tumori CRC erano incapaci di riparare malaccoppiamenti in un saggio in vitro (vedi riferimenti bibliografici 1 e 2).

La fortunata scoperta di una fase di lettura aperta (open reading frame, ORF) che codifica per un polipeptide omologo della proteina MutS (3,4) ha spianato la strada per l'identificazione di una famiglia sempre più numerosa di geni MSH, che vanno dai batteri all'uomo (vedi ad esempio il riferimento bibliografico 5). Tre membri di questa famiglia, gli omologi di MutS in *S. cerevisiae*

MSH1 ed MSH2, come pure l'omologo umano hMSH2, si legano in vitro al DNA malaccoppiato (6-9). Si è stabilita una relazione tra la funzione biologica di hMSH2 ed il fenotipo dei tumori CRC quando (i) si è mostrato che il gene *hMSH2* segrega con un locus genico sul cromosoma 2p associato alla comparsa di CRC (10, 11), (ii) si è mostrato che la linea cellulare LoVo deficiente in *hMSH2* è deficiente sia nella riparazione dei malaccoppiamenti (12) che nell'attività di legame al mismatch (13) e (iii) si è visto che il genoma di questa linea cellulare presenta una marcata instabilità delle sequenze microsatelliti (14). Un fattore che si lega ai malaccoppiamenti è stato originariamente identificato nelle cellule HeLa dagli inventori (15). Si è visto che questo fattore, che è la proteina GTBP, si lega preferenzialmente a molecole di DNA eteroduplici contenenti malaccoppiamenti G/T. La purificazione di questa attività di legame al DNA mediante cromatografia di affinità a malaccoppiamenti G/T ha dato una miscela di due polipeptidi di peso molecolare apparente di 100 e 160 kDa (16), rivelando così che il complesso specifico per il mismatch è di fatto composto da due proteine. Il

costituente di 100 kDa è risultato essere hMSH2 (17). Questa scoperta implica che hMSH2 agisce come complesso con GTBP per la correzione di malaccoppiamenti base/base e per la riparazione di loop di uno o due nucleotidi. Si è visto che molti tumori sono caratterizzati da fenotipi "mutatori" con un grado ugualmente basso di instabilità dei microsatelliti (20, 21, 22, 23, 24) consistente con il ruolo funzionale del complesso hMSH2-GTBP. Prima della scoperta e della caratterizzazione di GTBP, non era stato associato a questa proteina nessun ruolo nella riparazione dell'informazione genetica e nessun difetto ereditario conseguente al suo malfunzionamento.

Bibliografia Rilevante

1. P. Modrich, *Science* **266**, 1959 (1994)
2. J. Jiricny, *Trends Genet.* **10**, 164, 1994.
3. J. P. Linton et al. *Mol. Cell. Biol.* **9**, 303(1989)
4. H. Fujii and T. Shimada, *J. Biol. Chem.* **264**, 10057 (1989).
5. L. New, K.Liu, G.E. Crouse, *Mol. Gen. Genet.* **239**, 97 (1993).
6. N.-W. Chi and R. D. Kolodner, *J. Biol. Chem.* **269**, 29984 (1994).
7. T. Prolla et al., *Science* **265**,1091(1994).
8. B. Alani, N-W Chi, R. D. Kolodner, *Genes & Development* **9**, 234 (1995).
9. R. Fishel, A. Ewel, M.K. Lescoe, *Cancer Research* **54**, 5539 (1994).
10. R. Fishel et al., *Cell* **75**,1027 (1993).
11. F.S. Leach et al., *Cell* **75**,1215 (1993).
12. A. Umar et al., *J. Biol. Chem.* **269**, 14367 (1994).
13. G. Aquilina et al., *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8905 (1994).
14. D. Shibata et al., *Nature Genet.* **6**, 273 (1994).
15. J. Jiricny et al. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8860 (1988).
16. M. Hughes and J. Jiricny, *J. Biol. Chem.* **267**, 23876 (1992).
17. F. Palombo et al. *Nature* **367**, 417 (1994).
18. J. Lingner, J. Kellernan and W. Keller, *Nature* **354**, 496 (1991).
19. Da. Costa et al., *Nature Genetics* **9**, 10 (1995).
20. L.A. Aaltonen et al., *Science* **260**, 812 (1993)
21. S.N. Thibodeau et al., *Science* **260**, 816 (1993)
22. Ionov et al., *Nature* **363**, 558 (1993)
23. R. Wooster et al., *Nature Genetics* **6**, 152 (1994)
24. A. Merlo et al., *Cancer Res.* **54**, 2098 (1994).
25. J.D. Dignam et al., *Methods Enzymol.* **101**, 382 (1983).
26. P. Gallinari et al., *J. Virol.* **68**, 3809 (1994).

Sommario dell'invenzione

Scopo della presente invenzione è quello di fornire una sequenza di 1292 amminoacidi, corrispondente al polipeptide denominato GTBP.

Altro scopo dell'invenzione è di fornire un costrutto genetico che contiene una sequenza a doppio filamento di DNA di 3992 paia di basi che codifica per il peptide di 1292 amminoacidi denominato GTBP.

Ulteriore scopo dell'invenzione è di fornire un costrutto genetico capace di esprimere un peptide di 1292 amminoacidi di massa molecolare 142 kDa denominato GTBP.

Scopo ancora ulteriore dell'invenzione è di fornire un metodo per la preparazione e l'isolamento della proteina GTBP nativa in forma pura da cellule e tessuti in coltura.

Altro scopo della presente invenzione è di fornire un metodo per la valutazione della attività di GTBP in vitro.

E' ancora uno scopo della presente invenzione fornire un metodo per la rilevazione di GTBP mutata mediante uso di anticorpi specifici diretti verso GTBP.

Altro scopo dell'invenzione è di fornire un

metodo per la rilevazione di alleli mutati di *GTBP* mediante uso di PCR (Polymerase Chain Reaction) e sequenziamento dei prodotti di amplificazione.

Altro scopo dell'invenzione è fornire sonde di DNA per la rilevazione di geni *GTBP*, di tipo selvatico o mutati in cellule umane.

E' ancora scopo dell'invenzione quello di fornire un metodo per la diagnosi e prognosi di cancro colorettali umani (CRC).

E' inoltre scopo della presente invenzione fornire un metodo per la rivelazione di predisposizione genetica ai cancro umani colorettali (CRC).

Altro scopo dell'invenzione è di fornire un metodo per lo screening su larga scala di una popolazione con la finalità di evidenziare eventuali predisposizioni genetiche ai carcinomi umani colorettali (CRC).

E' ancora scopo dell'invenzione predisporre un metodo per fornire alleli di *GTBP* di tipo selvatico a cellule che hanno perso la funzione del gene *GTBP*.

E' ulteriore scopo dell'invenzione quello di fornire un metodo per generare animali trasgenici con alleli mutati di *GTBP*.

E' infine scopo dell'invenzione fornire un metodo per saggiare l'attività di agenti terapeutici finalizzati a sopprimere i carcinomi umani coloretali (CRC).

Questi ed altri scopi vengono raggiunti con una o più delle forme di realizzazione dell'invenzione che sono descritte in seguito.

In una forma di realizzazione viene fornita la sequenza di un polipeptide di 1292 amminoacidi corrispondente alla proteina denominata GTBP.

In un'altra forma di realizzazione viene fornita una molecola di cDNA che comprende la sequenza codificante del gene GTBP.

In un'altra forma di realizzazione viene fornito un procedimento per la preparazione della proteina GTBP pura.

In un'altra forma di realizzazione della presente invenzione vengono fornite coppia di oligonucleotidi per determinare la sequenza nucleotidica del gene GTBP, oppure di regioni di DNA interne al gene GTBP, mediante PCR (Polymerase Chain Reaction) e successivo sequenziamento. La sequenza di detti primers è interna al cromosoma 2p16, dette paia di primers permettendo la sintesi del gene GTBP oppure di

parti di esso.

In un ulteriore forma di realizzazione della presente invenzione viene fornita una sonda ad acido nucleico che è complementare alla sequenza codificante del gene umano *GTBP* di tipo selvatico e che può formare malaccoppiamenti quando è appaiato con alleli mutanti di *GTBP*, rendendo in tal modo possibile la rilevazione di DNA eteroduplex mediante alterazioni nella mobilità e nel profilo elettroforetico dopo taglio enzimatico oppure chimico.

In un'altra forma di realizzazione viene insegnato un procedimento per la rilevazione della proteina *GTBP* di tipo selvatico oppure mutata negli esseri umani. Questo procedimento comprende le fasi di isolare un campione selezionato da tessuti o fluidi organici e di determinare la proteina *GTBP* di tipo selvatico oppure alterata oppure presente in qualsiasi complesso formato dall'associazione di *GTBP* con altri polipeptidi.

In un'altra forma di realizzazione della presente invenzione viene fornito un metodo per la valutazione dell'attività (i) della proteina *GTBP* di tipo selvatico oppure (ii) di peptidi derivati ottenuti mediante delezioni oppure inserzioni di

sequenze note di amminoacidi nella proteina GTBP oppure (iii) della proteina GTBP alterata per effetto di eventi mutazionali in vivo oppure (iv) di qualsiasi complesso formato dall'associazione di peptidi appena menzionati in (i), (ii), (iii) e (iv) della presente forma di realizzazione con altri polipeptidi.

In un'ulteriore forma di realizzazione viene fornito un metodo per la rilevazione del cancro negli esseri umani. Questo metodo comprende le fasi di isolare un campione selezionato da tessuti o fluidi organici; rilevare l'alterazione nel gene *GTBP* oppure nella proteina espressa (proteina *GTBP*) oppure in qualsiasi complesso formato dall'associazione di *GTBP* con altri polipeptidi, detta alterazione indicando la predisposizione a trasformazione neoplastica oppure la presenza di cancro.

In un'ancora ulteriore forma di realizzazione della presente invenzione viene fornito un metodo di diagnosi o prognosi di tessuto neoplastico comprendente le fasi di rilevare alterazioni somatiche di alleli di *GTBP* oppure loro prodotti di espressione in cancro coloretali umani (CRC), detta alterazione indicando neoplasia del tessuto.

In un'altra forma di realizzazione viene fornito un metodo per la rilevazione di predisposizione genetica a cancro coloretale umano (CRC). Questo metodo comprende le fasi di isolare un campione selezionato dal gruppo consistente di sangue, campioni biotici di tessuti, cellule di sfaldamento e qualsiasi altro campione umano generico; determinare l'alterazione nel gene *GTBP* oppure nel polipeptide espresso (proteina *GTBP*) oppure in qualsiasi complesso formato dall'associazione di *GTBP* con altri polipeptidi, detta alterazione indicando predisposizione genetica al cancro.

In un'altra forma di realizzazione della presente invenzione viene predisposto un metodo per fornire la funzione del gene *GTBP* di tipo selvatico a cellule che hanno perso detta funzione genetica per effetto di una qualsiasi mutazione nel gene *GTBP*. Questo metodo comprende le fasi di introdurre il gene *GTBP* di tipo selvatico in una cellula che ha perso la funzione di detto gene in modo che il gene *GTBP* venga poi espresso come proteina di tipo selvatico nella cellula. La proteina *GTBP* può anche essere inserita in cellule oppure somministrata ad animali per rimediare a

difetti nella funzione del gene *GTBP*. Possono anche essere usati peptidi sintetici o farmaci per mimare la funzione di *GTBP* in cellule con espressione alterata del gene *GTBP*.

In una ulteriore forma di realizzazione viene fornito un metodo per dare una porzione del gene *GTBP* di tipo selvatico ad una cellula che ha perso detto gene in modo che la detta porzione venga espressa nelle cellule e codifichi la parte della proteina *GTBP* che è richiesta per la crescita non-neoplastica di detta cellula.

E' ancora un'altra forma di realizzazione della presente invenzione quella di generare animali transgenici che portano un gene *GTBP* mutato derivato da una seconda specie oppure un gene *GTBP* mutato generato in vitro mediante tecniche di ingegneria genetica.

In un'altra forma di realizzazione della presente invenzione viene fornito un metodo per selezionare agenti terapeutici per quanto riguarda la capacità di sopprimere un fenotipo trasformato neoplasticamente. Il metodo comprende le fasi di applicare una sostanza di prova ad una cellula epiteliale in coltura che porta una mutazione del gene *GTBP* e di determinare se la sostanza sopprime

il fenotipo neoplastico della cellula.

In un'altra forma di realizzazione della presente invenzione viene fornito un metodo per saggiare agenti terapeutici per quanto riguarda l'abilità a sopprimere un fenotipo trasformato neoplasticamente. Il metodo comprende le fasi di applicare una sostanza di prova ad un animale che porta una mutazione del gene *GTBP* e di determinare se la sostanza previene la trasformazione neoplastica di determinati tessuti oppure sopprime la crescita di tumori già sviluppati.

La presente invenzione arricchisce il relativo settore tecnologico con l'insegnamento che il gene *GTBP*, un gene finora sconosciuto, codifica per la proteina *GTBP* che agisce come fattore di legame specifico a malaccoppiamenti tra basi del DNA. *GTBP* si lega preferenzialmente a DNA eteroduplice contenente malaccoppiamenti G/T come pure loops di uno oppure due nucleotidi. La purificazione di questa attività di legame a DNA rende possibile stabilire che il fattore specifico per il mismatch è un complesso composto da due proteine distinte. Il costituente più piccolo del complesso (circa 100 kDa) è la proteina *hMSH2* (17) mentre il componente più grande (circa 150 kDa) è

GTBP. La presente invenzione fornisce gli strumenti tecnici per l'identificazione e per la valutazione dell'attività di GTBP solo oppure in forma di complesso con hMSH2. Il gene *GTBP* è bersaglio di eventi mutazionali e le relative alterazioni sono associate con la genesi del tumore. Questa scoperta permette la messa a punto di saggi molto specifici per la determinazione dello status neoplastico relativo ad un particolare tessuto oppure della predisposizione al cancro delle persone. Molti tumori esibiscono fenotipi mutatori con un grado di stabilità dei microsatelliti basso nella stessa misura (20, 21, 22, 23, 24), fatto questo coerente con il ruolo funzionale del complesso hMSH2-GTBP. Si desidera ancora una volta sottolineare che, prima della scoperta e della caratterizzazione di GTBP, non era stata associata con questa proteina nessun ruolo nella riparazione di informazione genetica e nessun difetto ereditario.

Breve descrizione delle figure

Figura 1a mostra il vettore fasmidico commerciale p-Bluescript SK- (Stratagene) usato per clonare e sequenziare il cDNA di GTBP. Il frammento di DNA mostrato in SEQ ID NO: 12 è stato

clonato tra i siti *EcoRI* e *XhoI* del vettore. Figura 1b mostra il vettore commerciale pCITE 2b. L'inserto descritto inSEQ ID NO:12 è stato inserito tra i siti *EcoRI* e *XhoI* del vettore.

Amp = Gene di beta-lattamasi per la resistenza all'ampicillina

ColE1 ori = origine di replicazione derivata dal plasmide ColE1

f1 = origine di replicazione del fago F1

lacZ = alfa-peptide della beta-galattosidasi usato per la complementazione genetica

MCS = multiple cloning site, sito di clonaggio multiplo, contenente le sequenze di riconoscimento degli enzimi di restrizione indicati

T3 e T7 = sequenze promotori dai fagi T3 e T7.

Figura 2 mostra costrutti di espressione generati per clonaggio in pGEX-3x (Pharmacia Biotech) di frammenti PCR corrispondenti a residui stati: 5'CGGGATCCCCCGGAGAAGCCGACCACCAC3' e 5'CGGAATTCCTGGCCATCAACTGCGGACAT3' per i codoni da 27 a 158 di hMSH2, e 5'CGGAATTCTCAACTCGTATTCTTCTG3' e 5'CGGGATCCCCCTTGAGAGGCTACTCAGT3' per i codoni da 750 a 928 di GTBP. SEQ ID NO: 13 e 14 sono state clonate tra i siti *BamHI* e *EcoRI*. I prodotti di

espressione, in forma di polipeptidi fusi con glutathione-S-transferasi, sono stati purificati mediante cromatografia di affinità su una matrice commerciale di glutathione (Pharmacia Biotech) secondo le procedure indicate dal produttore. Le proteine di fusione pure sono state usate per l'immunizzazione di conigli femmina New Zealand White SPF mediante protocolli standard riportati nella pubblicazione *Antibodies: A Laboratory Manual* (1988) eds. Harlow and Lane, Cold Spring Harbor Laboratories Press.

Figura 3 mostra un allineamento delle sequenze amminoacidiche codice ad una lettera delle regioni C-terminali conservate delle quattro proteine di legame a malaccoppiamenti, vale a dire GTBP (*H. sapiens*), hMSH2 (*H. sapiens*), MSH2 (*S. cerevisiae*) e MutS (*E. coli*). I residui identici sono in nero, quelli conservati sono in riquadri ombreggiati. Le sequenze riportate nell'allineamento corrispondono alle registrazioni MSH2_YEAST (MSH2) e MUTS_ECLI (MutS) nella banca dati SwissProt, oppure alla regione codificante della registrazione HSU05045 (hMSH2) della banca genica (GenBank). Gli allineamenti mostrano che esiste un elevato grado di conservazione fra i tre

omologhi, con la parte C-terminale della proteina conservata in modo particolarmente elevato. GTBP può pertanto essere considerato un nuovo membro della famiglia MSH.

Figura 4 mostra l'omologia di sequenza, al livello di proteina, tra paia di membri della famiglia MSH. Sezione a mostra la matrice ottenuta dall'allineamento di GTBP (sull'ascissa) con l'omologo fungino (YEAST, numero di accesso alla banca genica GenBank Z47746, sull'ordinata); le due proteine mostrano lunghezze comparabili ed è evidente una significativa omologia per la loro intera sequenza. Sezione b mostra la matrice ottenuta dall'allineamento della proteina fungina MSH2 (MSH2_YEAST, sull'ordinata) con GTBP (sull'ascissa); anche in questo caso, le proteine mostrano differente lunghezze e la maggior parte dell'omologia è confinata alle regioni C-terminali delle due sequenze. Sezione c mostra la matrice ottenuta dall'allineamento della proteina umana hMSH2 (MSH2_HUMAN, sull'ordinata) con GTBP (sull'ascissa); le proteine mostrano differenti lunghezze e la maggior parte dell'omologia è confinata alle regioni C-terminali delle due sequenze. Sezione d mostra la matrice ottenuta

dall'allineamento della proteina umana hMSH2 (MSH2_HUMAN, sull'ordinata) con l'omologa di lievito MSH2 (MSH2_YEAST, sull'ascissa); le due proteine mostrano lunghezze comparabili e omologia significativa per tutta la loro sequenza.

Figura 5 mostra l'effetto di antisieri selettivi anti-hMSH2 ed anti-GTBP sulla formazione del complesso di legame specifico per il malaccoppiamento. Preincubazione di estratti nucleari di HeLa con entrambi gli antisieri prima dell'aggiunta della sonda di DNA eteroduplex G/T ha come effetto una diminuzione della banda specifica nel saggio di ritardo su gel, un effetto non osservato quando sono stati usati i rispettivi sieri pre-immuni. Questa figura dimostra che entrambe le proteine hMSH2 e GTBP sono presenti nel fattore che si lega ai malaccoppiamenti. Questa analisi di ritardo su gel è stata eseguita come descritto in (15), tranne per il fatto che vengono usati estratti nucleari (25). Gli antisieri sono stati aggiunti alle miscele di reazione 20 minuti prima dell'aggiunta della sonda marcata radioattivamente. La figura è un autoradiogramma di un gel nativo al 6% di poliacrilammide.

eseguito in tampone di tris-acetato (TAE) preparato secondo Maniatis et al., Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982.

Figura 6 mostra che l'attività di legame a malaccoppiamenti può essere ricostituita usando GTBP e hMSH2 ottenute usando un sistema di traduzione in vitro. Sezione a mostra la procedura seguita per generare in vitro trascritti delle sequenze che codificano per hMSH2, C1 e FLY5. La regione di DNA che codifica per hMSH2 è stata inserita in pCite-1; i moduli di lettura di C1 e FLY5 sono stati introdotti in pCite-2b (Novagen). Le reazioni di trascrizione e traduzione in vitro sono state eseguite come descritto in (26), incluso controllo negativo (cioè traduzione in assenza di mRNA; mock translation). I prodotti di traduzione marcati con ^{35}S sono stati analizzati su un gel di SDS-poliacrilammide trattato con Amplify (Amersham), seccati e autoradiografati. I saggi di ritardo su gel sono stati eseguiti come descritto nel riferimento (15). Vengono sottoposti a prova aliquote di 5 μl delle singole reazioni di traduzione in vitro; negli esperimenti di pre-miscelazione, 2,5 μl di ciascuna delle due

reazioni di traduzione vengono mescolati ed incubati per 15 minuti a temperatura ambiente prima dell'aggiunta della sonda. AMP ad una concentrazione di 5 mM è stato incluso in tutte le reazioni di legame al DNA in modo da superare l'effetto dell'ATP nei lisati di reticolociti, che previene la formazione di complessi malaccoppiamento-proteina specifica-DNA secondo il riferimento (16). Sezione a è un autoradiogramma di un gel al 7,5% di SDS-poliacrilammide denaturante il quale mostra che la traduzione di mRNA di hMSH2, GTBP (C1) e FLY5 in un sistema di lisato di reticolociti (Promega) dà origine ai polipeptidi aspettati di 113, 144 e 122 kDa rispettivamente. La sezione b mostra l'analisi di ritardo su gel che prova il legame delle proteine tradotte in vitro con l'eteroduplex G/T. La figura è un autoradiogramma di un gel al 6% di poliacrilammide nativo corso in tampone TAE.

La figura 7 mostra che l'attività di legame a malaccoppiamento è assente negli estratti cellulari privi di GTBP oppure hMSH2. L'esperimento è basato sull'analisi di due linee cellulari derivate da CRC: di cellule LoVo contengono una delezione omozigote di alleli di

hMSH2 e non esibiscono attività di legame a G/T (13), mentre nessuno degli alleli di *hMSH2* viene mutato in cellule DLD1, nonostante il fatto che anche questa linea cellulare è priva di attività di legame a G/T (19). Sezione a mostra un saggio di ritardo su gel da cui si evince che estratti di cellule LoVo (13) e DLD1 (19) non riescono a formare complessi specifici per il malaccoppiamento. Le sonde G/C e G/T sono state ottenute come precedentemente descritto (15). Le condizioni sperimentali sono quelle indicate in figura 6. La figura è un autoradiogramma di un gel al 6% di poliacrilammide nativo eseguito in tampone TAE. Sezione b mostra l'analisi di Western blot di estratti da cellule HeLa, LoVo e DLD1. Le bande dell'immunocomplesso vengono visualizzate usando un sistema anti IgG di coniglio coniugato con fosfatasi alcalina (Promega) secondo le indicazioni del produttore. Nei due elettroferogrammi di sinistra, gli antisieri anti-GTBP ed anti-*hMSH2* vengono usati da soli con l'estratto di cellule HeLa per dimostrare la loro selettività per le proteine di 160 e 100 kDa rispettivamente. Nei sentieri rimanenti, entrambi gli antisieri vengono usati insieme. Cellule di

controllo HeLa hanno rivelato sia la presenza di hMSH2 che di GTBP. Al contrario, le due linee cellulari tumorali LoVo e DLD1, derivate da CRC erano rispettivamente prive di hMSH2 e GTBP. Le quantità di hMSH2 nelle cellule DLD1 e di GTBP nelle cellule LoVo erano notevolmente più basse che nelle cellule HeLa. Dato che hMSH2 e GTBP si legano come complesso a DNA eteroduplex, la mancanza di una delle due proteine può causare instabilità del secondo componente del complesso.

Descrizione dettagliata

In vista della potenzialità e varietà dei ruoli delle proteine di legame a malaccoppiamento nella riparazione di informazione genetica e dei loro effetti sugli stati patologici, quali trasformazione e proliferazione di cellule tumorali, metastasi, e data la scarsa conoscenza delle molecole degli agenti che operano oppure modulano selettivamente l'attività di queste proteine, esiste nel settore l'esigenza di composti ed agenti con attività di effettore e modulatore come pure di metodi per identificare questi composti e questi agenti e quelli in relazione con essi. Inoltre, questi agenti possono servire come reagenti di ricerca commerciali per

il controllo della riparazione di acidi nucleici, e per altre situazioni correlate a GTBP. Nonostante il progresso nello sviluppo di un modello più definito del meccanismo molecolare che sta alla base della riparazione di acidi nucleici, si sono evoluti pochi metodi applicabili con successo alla determinazione della predisposizione al cancro e/oppure alla sua cura. L'eterodimero hMSH2/GTBP è necessario per la correzione di malappaiamenti base/base e di loops di uno oppure due nucleotidi. L'instabilità genomica in linee cellulari derivate da tumore, che sono prive di GTBP si manifesta principalmente in forma di piccole differenze (ad esempio in serie di A) piuttosto che di grandi cambi nelle ripetizioni CA, caratteristici di fenotipi associati con i quattro loci conosciuti di CRC, *hMSH2*, *hMLH2*, *hPMS1* e *hPMS2*. I cancro che esibiscono fenotipi mutatori con un basso grado di stabilità di microsatelliti (24) possono essere associati con un cattivo funzionamento di GTBP. Si è scoperto, e questa scoperta è alla base della presente invenzione, che eventi mutazionali associati con la cancerogenesi in CRC sono dovuti a difetti presenti nel gene *GTBP*.

Con la presente invenzione vengono pertanto fornite nuove composizioni che comprendono sequenze genetiche che codificano per la proteina GTBP, frammenti derivati da essa, proteine ricombinanti prodotte usando dette sequenze genomiche e metodi per l'uso di queste composizioni.

Forme di realizzazioni tipiche di esemplari sequenze amminoacidiche e nucleotidiche secondo l'invenzione sono indicate in SEQ ID NO:1 e SEQ ID NO: 12 rispettivamente. Come già visto in precedenza nelle figure ed in altre parti di questa domanda di brevetto vengono usate abbreviazioni standard per i nucleotidi e gli amminoacidi. I polipeptidi derivati da GTBP sono forme di realizzazione particolarmente preferite dell'invenzione, sebbene anche varianti riconducibili a sequenze specifiche di questi polipeptidi formano parte della presente invenzione. Nei suoi aspetti più ampi, l'invenzione (per quanto riguarda i polipeptidi per sè) include ogni polipeptide selezionato dal gruppo consistente di:

(i) qualsiasi proteina che ha una sequenza amminoacidica omologa almeno all'85% con la

sequenza amminoacidica SEQ ID NO:1, e (ii) frammenti di questa proteina che comprendono almeno 10 amminoacidi consecutivi localizzati entro la sequenza amminoacidica SEQ ID NO:1, essendo il polipeptide in questione capace di legarsi ad un anticorpo specifico per GTBP.

Per quanto riguarda gli aspetti di ingegneria genetica della presente invenzione, le sequenze codificanti specifiche indicate in SEQ ID NO:12, che corrispondono ai polipeptidi preferiti, sono esse stesse preferite.

Sono parimenti preferite sequenze di DNA ed RNA equivalenti e complementari (per quanto riguarda la definizione di questi termini si veda nel seguito). In generale, gli aspetti di ingegneria genetica della presente invenzione includono ogni molecola di DNA o RNA ricombinante che comprende una sequenza di DNA che codifica per GTBP stesso oppure per una proteina derivata da GTBP, oppure una sequenza di RNA corrispondente, oppure una sua sottosequenza comprendente almeno 10 nucleotidi. La presente invenzione mette anche a fuoco metodologie diagnostiche finalizzate al rilevamento di mancanza della funzione GTBP negli esseri umani ed alla conseguente predisposizione a

neoplasia.

Definizione dei termini

Molti termini che vengono usati correntemente nel settore dell'ingegneria genetica e della chimica delle proteine vengono qui usati con il seguente significato.

Due frammenti di acido nucleico sono "omologhi" se sono capaci di ibridizzare l'un con l'altro nelle condizioni di stringenza di ibridazione descritte in Miniatis et al, (1982), Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 320-325. Usando le seguenti condizioni di lavaggio - 2xSSC, 0,1% SDS, due volte a temperatura ambiente, 30 minuti ciascuno; 2xSSC, 0,1% SDS, una volta a 50°C, 30 minuti; poi 2xSSC, due volte a temperatura ambiente, 10 minuti ciascuna - possono essere identificate sequenze omologhe che contengono al massimo circa 25-30% di malaccoppiamenti di paia di basi. Preferibilmente, un filamento di acido nucleico omologo contiene 15-25% di malaccoppiamenti di paia di basi, più preferibilmente contiene 5-15% di malaccoppiamenti di paia di basi. Questi gradi di omologia possono essere selezionati usando condizioni di lavaggio

più stringenti per l'identificazione di cloni da librerie geniche (oppure da altre fonti di materiale genetico), come è ben noto nel settore specifico.

Due sequenze amminoacidiche sono omologhe se vi è una parziale o completa identità tra le loro sequenze. Per esempio, 85% di omologia significa che l'85% degli amminoacidi sono identici quando le due sequenze sono allineate per il massimo accoppiamento. Per ottenere un allineamento ottimale sono permessi gap, cioè interruzioni, (in ciascuna delle due sequenze che debbono essere accoppiate); lunghezze di gap di 5 residui amminoacidi o meno sono preferite, di 2 residui amminoacidi o meno sono ancora più preferite.

Alternativamente e preferibilmente, due sequenze proteiche (o sequenze polipeptidiche derivate da esse con almeno una lunghezza di 30 amminoacidi) sono omologhe, nel significato che qui si dà a questo termine, se hanno un punteggio di allineamento di più di 5 (in unità di deviazione standard) usando il programma ALIGN con la matrice dei dati di mutazione ed una penalità di gap di 6 o più grande (Dayhoff, M.O. in Atlas of protein Sequence and Structure, 1972, volume 5,

National Biomedical Research Foundation pp 101-110, e il supplemento 2 di questo volume, pp. 1-10). Le due sequenze o parti di esse sono più preferibilmente omologhe se i loro amminoacidi hanno un grado di identità maggiore o uguale a 50% quando sono allineate nel modo ottimale usando il programma ALIGN.

Un frammento di DNA è "derivato da" una sequenza di DNA che codifica per GTBP se ha la stessa, o sostanzialmente la stessa sequenza di paia di basi di una regione della sequenza codificante per la proteina GTBP.

"Sostanzialmente le stesse", se si riferisce ad attività biologiche, significa che le attività sono dello stesso tipo anche se possono differire in grado. Quando ci si riferisce a sequenze amminoacidiche. "sostanzialmente le stesse" significa che le molecole in questione hanno proprietà biologiche simili e preferibilmente hanno una omologia almeno dell'85% nelle sequenze amminoacidiche. Preferibilmente, le sequenze amminoacidiche sono identiche almeno al 90%. In altre circostanze, l'espressione "sostanzialmente le stesse" ha il significato che gli si attribuisce nel linguaggio comune.

Una proteina è "derivata da" GTBP, se ha la stessa, oppure sostanzialmente la stessa, sequenza amminoacidica di una regione della molecola proteica GTBP. Per "derivati polipeptidici" della proteina GTBP si intendono polipeptidi che differiscono per la lunghezza dalla proteina naturale e contengono cinque o più amminoacidi nello stesso ordine primario che si ritrova nella proteina ottenuta da fonti naturali. Rientrano nella definizione di "derivati da" GTBP molecole polipeptidiche che hanno sostanzialmente la stessa sequenza amminoacidica della proteina naturale, ma possiedono piccole sostituzioni amminoacidiche che non pregiudicano in modo significativo la capacità della proteina o del polipeptide ad interagire con molecole specifiche nei confronti della proteina, quali anticorpi ed acidi nucleici. I derivati includono forme glicosilate, coniugati d'aggregazione con altre molecole proteiche e coniugati covalenti con unità chimiche senza reciproca connessione. I derivati covalenti sono preparati dotando di funzionalità, con tecniche note nel settore specifico, gruppi che si trovano nella catena amminoacidica oppure nel residuo N-terminale oppure C-terminale.

Molecole specifiche nei confronti di GTBP includono polipeptidi, quali gli anticorpi, che sono specifici per la proteina o per il polipeptide che contiene una sequenza amminoacidica di GTBP quale si ritrova in natura. Per "polipeptide che si lega specificamente" si intendono i polipeptidi che si legano con la proteina GTBP e con i suoi derivati e che hanno una affinità di legame per il polipeptide bersaglio considerevolmente più alta di quella per altri polipeptidi sottoposti alla prova di legame. Si preferisce una affinità più elevata di un fattore 10, più preferibilmente di un fattore 100. L'affinità di legame per anticorpi si riferisce ad un singolo evento di legame (cioè, legame monovalente di una molecola di anticorpo). Nel caso degli anticorpi, legame specifico significa anche che il legame ha luogo nel normale sito di legame della molecola di anticorpo (all'estremità dei bracci della regione variabile).

Come già discusso in precedenza, sono previste piccole variazioni amminoacidiche rispetto alla sequenza amminoacidica naturale della proteina GTBP; in particolare, sono previste sostituzioni amminoacidiche conservative.

Sostituzioni conservative sono quelle che hanno luogo nell'ambito di una famiglia di amminoacidi che sono correlati nelle loro catena laterali. Gli amminoacidi geneticamente codificati vengono solitamente divisi in quattro famiglie: (1) acidi = aspartato, glutammato; (2) basici = lisina, arginina, istidina; (3) non-polari = alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptofano; e (4) polari senza carica = glicina, asparagina, glutammina, cistina, serina, treonina, tirosina. Fenilalanina, triptofano e tirosina sono talora classificati insieme come ammino-acidi aromatici. Per esempio, è ragionevole aspettarsi che una sostituzione isolata di una leucina con una isoleucina oppure una valina, di un aspartato con un glutammato, di una treonina con una serina, oppure una sostituzione similare di un amminoacido con un amminoacido strutturalmente correlato non avrà un effetto evidente sulle proprietà di legame della molecola risultante. Specialmente se la sostituzione non coinvolge un amminoacido ad un sito di legame coinvolto nell'interazione di GTBP, o di suoi derivati, con un anticorpo oppure con una sequenza di riconoscimento per il DNA

specifica. Se la sostituzione di un amminoacido ha effetti su un peptide funzionale, può essere facilmente determinato saggiando le proprietà di legame specifiche del derivato di polipeptide.

Isolamento di cDNA che codifica per la proteina GTBP

L'isolamento di sequenze nucleotidiche che codificano per la proteina GTBP coinvolge la creazione di una libreria di cDNA preparata da RNA messaggero maturo, in tutta la sua lunghezza, estratto da cellule o tessuti in coltura. Si è ottenuta la prova che GTBP è conservato in un ampio spettro evolutivo; in tal modo si possono isolare omologhi di GTBP dai genomi di specie filogeneticamente distanti, vale a dire dai mammiferi passando per i lieviti sino ai batteri.

Si possono costruire librerie batteriche in cellule ospiti sia eucariotiche che procariotiche. Per generare librerie genomiche adatte per l'isolamento di sequenze nucleotidiche che codificano per la proteina GTBP oppure per sue porzioni, si possono usare vettori di clonaggio largamente disponibili, quali plasmidi, cosmidi, fagi, YACs (Yeast Artificial Chromosomes) e simili. Metodi utili, per lo screening di librerie

genetiche per rilevare la presenza di sequenze nucleotidiche che codificano per la proteina GTBP, includono la preparazione di sonde oligonucleotidiche basate sulle sequenze SEQ ID NO:1 e SEQ ID NO.12. Impiegando il codice genetico standard a triplette, si possono preparare, con tecniche convenzionali di sintesi in vitro, sequenze oligonucleotidiche circa 17 paia di basi o ancora più lunghe. Le sequenze di acido nucleico risultanti possono essere successivamente marcate con radioisotopi, enzimi, biotina, sostanze fluorescenti o simili, e possono essere usate come sonde per lo screening delle librerie.

Altri interessanti metodi per l'isolamento di sequenze di acidi nucleici che codificano per la proteina GTBP includono lo screening di librerie genetiche per l'espressione della proteina GTBP, oppure di frammenti di essa, mediante anticorpi, sia policlonali che monoclonali, specifici nei confronti della proteina GTBP.

Inoltre, un metodo di selezione consigliabile per lo screening delle librerie di GTBP clonata in vettori convenzionali di espressione è basato sul legame specifico della proteina (oppure di polipeptidi contenuti in essa) a molecole di DNA

eteroduplex contenenti malaccoppiamenti G/T. Una tecnica particolarmente preferita per isolare proteine omologhe da specie o ceppi correlati richiede l'uso di oligonucleotidi (primers) degenerati basati sull'uso di sequenze amminoacidiche parziali della proteina GTBP e sulla reazione a catena della polimerasi (PCR) per amplificare i segmenti genetici tra i primers. Si può anche applicare un approccio similare per generare molecole di cDNA a doppio filamento dopo amplificazione di mRNA con appropriati primers e polimerasi. Il gene può poi essere isolato usando una specifica sonda per l'ibridazione basata sul segmento genico amplificato, che è poi analizzato per l'appropriata espressione della proteina.

La sequenza nucleotidica del materiale genetico isolato che codifica per la proteina GTBP può essere ottenuta sequenziando le sequenze nucleotidiche non-vettoriali di queste molecole ricombinanti. Si possono ottenere informazioni sulle sequenze nucleotidiche impiegando, protocolli di sequenziamento di DNA largamente usati, quali ad esempio la procedura di Maxam e Gilbert, quella di Sanger, e simili. Esempi di protocolli di sequenziamento nucleotidico adatti

si possono trovare in Berger e Kimmel, *Methods in Enzymology*, Vol 52, *Guide to Molecular Cloning Techniques*, (1987) Academic Press. Informazioni sulle sequenze nucleotidiche da diversi isolati di DNA ricombinante, provenienti sia da librerie di cDNA che genomiche, possono essere combinate in modo da fornire l'intera sequenza amminoacidica codificante per GTBP, come pure le sequenze nucleotidiche a monte ed a valle di essa.

Le sequenze nucleotidiche ottenute dal sequenziamento di cloni codificanti per la proteina GTBP da librerie genomiche possono essere sottoposte ad ulteriore analisi al fine di identificare regioni di interesse nel gene *GTBP*. Queste regioni di interesse comprendono fasi di lettura aperte addizionali, sequenze di promotori, sequenze di terminazione e simili. L'analisi dell'informazione sulle sequenze nucleotidiche viene preferibilmente eseguita mediante computer. Il software adatto per analizzare le sequenze nucleotiche per le regioni di interesse è disponibile in commercio (si veda, ad esempio, DNASIS della ditta Pharmacia Biotech). E' anche interessante usare le informazioni sulle sequenze amminoacidiche, ottenute dal sequenziamento a

partire dal residuo N-terminale della proteina GTBP purificata, quando si analizzano nuove informazioni sulla sequenza nucleotidica di questa proteina, in modo da migliorare l'accuratezza dell'analisi delle sequenze nucleotidiche.

Espressione di GTBP

Sequenze nucleotidiche isolate che codificano per la proteina GTBP possono essere usate per produrre in forma purificata la proteina GTBP oppure frammenti di essa. Questo può avvenire sia mediante la metodologia del DNA ricombinante sia mediante le tecniche di sintesi di polipeptide in vitro. I termini "purificate" e "isolate", in relazione a sequenze polipeptidiche oppure nucleotidiche, significano che la molecola indicata è presente in sostanziale assenza di altre macromolecole biologiche dello stesso tipo. Il termine "purificate", nel significato che gli si dà in questo contesto, significa che la macromolecola in questione rappresenta almeno il 95% in peso, più preferibilmente almeno il 99% in peso, ed ancora più preferibilmente almeno il 99,8% in peso di tutte le macromolecole biologiche dello stesso tipo presenti in un campione (ma possono essere presenti anche acqua, tamponi ed

altre piccole molecole, specialmente molecole che hanno un peso molecolare inferiore a 100).

Un vantaggio significativo nella produzione della proteina GBTB mediante tecniche di DNA ricombinante, piuttosto che per isolamento della proteina da sorgenti naturali, è che si possono produrre quantità equivalenti della proteina GTBP usando meno materiale di partenza di quello che sarebbe necessario per isolare la proteina GTBP da una sorgente naturale. La produzione della proteina GTBP mediante tecniche ricombinanti permette anche di isolare la proteina in assenza di alcune molecole normalmente presenti nelle cellule che producono naturalmente la proteina GTBP. E' anche evidente che si possono utilizzare tecniche di DNA ricombinante per produrre derivati polipeptidici della proteina GTBP che non si trovano in natura, quali ad esempio i varianti descritti in precedenza.

La proteina GTBP ed i derivati polipeptidici della proteina GTBP possono essere espressi mediante tecniche ricombinanti inserendo funzionalmente in un vettore la sequenza di DNA che codifica per la molecola che interessa. L'espressione "funzionalmente inserito" significa,

come lo intendono le persone esperte in questo settore, in appropriata fase di lettura ed orientamento trascrizionale. Solitamente, il gene della proteina GTBP viene inserito a valle in un promotore ed è seguito da un codone di stop, anche se, nel caso lo si desideri, può essere prodotta una proteina ibrida. In generale si usano sequenze specifiche della cellula ospite che migliorano la resa di produzione della proteina GTBP e dei suoi derivati polipeptidici, e si aggiungono al vettore di espressione appropriate sequenze di controllo, quali sequenze enhancer, sequenze di poliadenilazione e siti di legame al ribosoma.

Sono presi in considerazione due tipi di espressione fondamentali: (i) espressione in cellule di mammifero in modo da porre rimedio dell'insufficienza di GTBP in un individuo, e (ii) espressione per fornire GTBP, a fini irrilevanti rispetto all'ospite in cui avviene l'espressione, quale ad esempio la produzione di test diagnostici per la deficienza di GTBP.

Produzione di costrutti genetici per la trasformazione di cellule umane

Per l'espressione in cellule umane, viene preparato ed usato un costrutto genico per

trasformare le cellule umane.

Si sono sviluppate diverse strategie sono e stati messi a punto diversi vettori per l'espressione di proteine in cellule animali. Per esempio sono stati costruiti vettori ibridi BK-SV40. Questi vettori possono essere mantenuti in cellule umane in coltura come repliconi extracromosomiali a copie multiple di DNA a doppio filamento. Un vettore tipico consiste nel promotore di SV40 che controlla l'espressione del gene della resistenza alla neomicina (il marcatore selezionabile) e del promotore di MMTP regolato dalla sequenza enhancer di DRE che controlla l'espressione del gene clonato. In ogni caso, il costrutto estraneo include di solito segnali di iniziazione e terminazione della trascrizione e della traduzione, con i segnali di iniziazione 5' e di terminazione 3' per il gene di interesse, sebbene possa essere fornito DNA lineare all'ospite quando c'è ricombinazione per inserzione nel genoma ospite. Si può in tal modo ottenere espressione sotto il controllo del promotore nativo, sostituendo il gene difettoso con il DNA lineare che codifica per GTBP facendo uso di procedimenti cellulari, ad esempio

ricombinazione omologa. La regione di iniziazione trascrizionale che include il sito di legame per la RNA polimerasi (promotore) può essere nativa nell'ospite oppure può essere derivata da una sorgente alternativa, dovendo in quest'ultimo caso la regione essere funzionale nell'ospite. Le regioni di iniziazione della trascrizione possono includere non solo il sito di legame per la RNA polimerasi, ma anche regioni che forniscono la regolazione della trascrizione. La regione di terminazione 3' può essere derivata dallo stesso gene della regione di iniziazione trascrizionale oppure da un gene differente. Ad esempio, nel caso in cui il gene di interesse ha una regione di terminazione della trascrizione funzionale nella specie ospite, quella regione può essere mantenuta entro il gene.

Si può costruire una cassetta di espressione che include la regione di iniziazione trascrizionale, il gene della proteina GTBP sotto il controllo di regolazione trascrizionale della regione di iniziazione della trascrizione, il codone di iniziazione, la sequenza codificante del gene, con o senza introni, e i codoni di stop della traduzione, seguiti dalla regione di

terminazione trascrizionale, che include il terminatore, e può includere una sequenza segnale di poliadenilazione, ed altre sequenze associate con la terminazione trascrizionale. La direzione di trascrizione è 5'-3'. La cassetta è usualmente di circa 10 kb, frequentemente meno di 6 kb, ancora più frequentemente di meno di 5 kb.

Se il prodotto di espressione del gene deve essere localizzato in un posto diverso dal citoplasma, si costruisce il gene includendovi sequenze particolari che codificano per amminoacidi capaci di trasferire il prodotto in un sito particolare, che può essere un organello, oppure il prodotto genico può essere secreto in un ambiente esterno alla cellula. In letteratura sono descritti vari leaders di selezione, sequenze integratori di membrana e sequenze di trasferimento per indirizzare il prodotto peptidico di espressione verso un particolare sito.

Possono essere coinvolte una o più cassette, quando le cassette possono essere impiegate in tandem per l'espressione di geni indipendenti che possono esprimere prodotti indipendentemente l'uno dall'altro oppure possono essere regolate in

concomitanza, se i prodotti possono agire indipendentemente od in combinazione, ad esempio GTBP ed hMSH2.

La cassetta di espressione normalmente viene portata su un vettore che ha almeno un sistema di replicazione. Per convenienza, è usuale avere sistemi di replicazione funzionali in *E. coli*, quali quelli derivati da ColE1, pSC101, pACYC184 o simili. In questa maniera, a ciascuno stadio dopo ciascuna manipolazione, il costrutto risultante può essere clonato, sequenziato e può essere determinata la correttezza della manipolazione. In aggiunta, oppure al posto del sistema di replicazione di *E. coli*, può essere adottato un sistema di replicazione ad ampio spettro di ospiti, quali ad esempio i sistemi di replicazione dei plasmidi del gruppo di incompatibilità P1 (RK2, RP1, RP4 o R68).

In aggiunta al sistema di replicazione, è frequentemente presente almeno un marcatore che può essere utilizzato in uno o più ospiti, oppure possono essere presenti diversi marcatori per ospiti individuali. Cioè, un marcatore può essere impiegato per la selezione in un ospite procariote, mentre un altro marcatore può essere

impiegato per la selezione in un ospite eucariote. Tra i vari geni che possono essere impiegati si citano neo (resistenza alla neo-micincanamicina), cloramfenolo acetil-transfeasi (*cat*), β -lattamasi (*bla*), β -galattosidasi e così via.

I vari frammenti che comprendono i vari costrutti, le cassette di espressione, i marcatori e simili possono essere introdotti consecutivamente mediante taglio con enzimi di restrizione di un appropriato sistema di replicazione, e inserzione del particolare costrutto o frammento nella dimensione disponibili. Dopo ligazione e clonaggio il vettore può essere isolato per ulteriori manipolazioni. Tutte queste tecniche sono ampiamente descritte in letteratura e trovano particolare esemplificazione in Maniatis et al., *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1982.

Trasformazione di cellule di mammifero e terapia genica

Una volta che il vettore è stato messo a punto, può essere introdotto in cellule di mammifero. Le tecniche per trasformare le cellule di mammifero includono transfezione,

microiniezione, rilascio basato su liposomi e così via. La transfezione di cellule umane in coltura è il metodo più spesso usato e può essere ottenuto mediante protocolli standard che coinvolgono sia l'incubazione di cellule con DNA che è stato coprecipitato con fosfato di calcio sia DEAE-destrano sia elettroporazione con DNA di trasfezione purificato. In altri sistemi, per inserire DNA estraneo in cellule umane riceventi, può anche essere usato un virus geneticamente modificato, un liposoma oppure microiniezione. Una volta che il gene *GTBP* è stato introdotto nella cellula difettosa, esso può ovviare al difetto genetico, ripristinando il fenotipo normale. Questa metodologia, quando viene usata per porre rimedio ai difetti genetici di individui, va sotto il nome di terapia genica. Per attuare la terapia genica di cellule somatiche sono state messe a punto almeno due strategie che possono essere applicate anche per correggere i difetti genetici con effetti su *GTBP*: terapia genica ex vivo e in vivo. Di solito la terapia genica ex vivo comprende le seguenti operazioni:

- raccogliere le cellule dell'individuo malato.

- correggere il difetto genetico mediante trasferimento genetico
- selezionare e far crescere le cellule geneticamente corrette (riparate)
- reinserire oppure trapiantare le cellule corrette nel paziente.

Per mantenere in modo stabile e per esprimere in modo persistente il gene riparato nella cellula corretta vengono usati spesso vettori derivati da retrovirus di topo.

La terapia genica in vivo comporta l'introduzione diretta del gene riparato nelle cellule di un particolare tessuto del paziente. Il gene della proteina di tipo selvatico può essere clonato in vari virus benigni e può essere introdotto nelle cellule difettose bersaglio con una infezione in vivo.

Vettori derivati da adenovirus, herpes simplex virus e retrovirus sono eccellenti candidati per la terapia genica in vivo. Metodi e prospettive della terapia genica sono stati passati in rassegna da Mulligan (1993), Science 260:926-932.

Metodi diagnostici che fanno uso di antigeni

I metodi per rivelare analiti quali ad

esempio le proteine di legame secondo l'invenzione sono basati solitamente su saggi immunologici. Si possono condurre saggi immunologici per determinare la presenza o l'assenza di GTBP in cellule ospiti. Tali tecniche sono ben conosciute e non hanno bisogno di essere qui descritte in dettaglio. Esempi di queste tecniche comprendono tecniche di saggi immunologici sia eterogenei che omogenei. Entrambe le tecniche sono basate sulla formazione di un immunocomplesso tra la proteina di legame ed un anticorpo specifico corrispondente. I saggi eterogenei per GTBP di solito fanno uso di un anticorpo specifico monoclonale oppure policlonale, legato ad una superficie solida, come ad esempio nei saggi a sandwich. I saggi omogenei, che sono eseguiti in soluzione senza la presenza di una fase solida, possono essere usati, determinando ad esempio, la differenza nell'attività enzimatica apportata, per effetto del legame di un anticorpo libero, ad un coniugato enzima-antigene. Molti saggi adatti sono descritti in U.S. 3.817.837, U.S. 4.006.360 e U.S. 3.996.345.

Nel saggio sopra descritto il reagente sulla superficie solida viene preparato mediante

tecniche note in grado di attaccare materiale proteico ad un supporto solido, quale ad esempio granuli polimerici, dip sticks, oppure materiale da filtro. Questi metodi per attaccare prevedono l'adsorbimento non specifico della proteina al supporto oppure l'ancoraggio covalente della proteina, di solito mediante un gruppo amminico libero, ad un gruppo chimicamente reattivo sul supporto solido, quale ad esempio un gruppo carbossilico, idrossilico oppure aldeidico che è stato preventivamente attivato.

In una seconda configurazione diagnostica, nota come saggio omogeneo, l'anticorpo che si lega all'analita produce alcune modifiche nel mezzo di reazione che possono essere rivelate direttamente nel mezzo stesso. Tipi convenzionali di saggi omogenei, proposti qui di seguito, comprendono (a) spin-labeled reporters, in cui l'anticorpo che si lega all'antigene viene determinato mediante cambio in reported mobility (broadening of the spin splitting peaks), (b) reporters fluorescenti, in cui il legame viene rivelato in base alla variazione dell'emissione di fluorescenza, (c) reporters enzimatici, in cui l'anticorpo che si lega produce effetti sulle

interazioni enzima/substrato, e (d) reporters legati a liposomi, in cui il legame porta alla lisi del liposoma ed al rilascio del reporter in esso contenuto. L'adattamento di questi metodi all'antigene proteico della presente invenzione segue metodi convenzionali per la preparazione di reagenti per saggi omogenei.

In ciascuno dei saggi descritti in precedenza, il metodo di saggio comporta la reazione di estratti tissutali, ricavati da un individuo sotto osservazione, con un anticorpo e l'esame del campione per quanto riguarda la presenza dell'antigene legato. L'esame può comportare l'attacco di un anticorpo marcato anti-GTBP ad un complesso primario tra GTBP e l'anticorpo immobilizzato e la misura della quantità di reporter legato al supporto solido, come nel primo metodo, oppure può comportare l'osservazione dell'effetto dell'anticorpo di legame su un reagente di saggio omogeneo, come nel secondo metodo.

Produzione di proteine di legame specifiche

GTBP, nella sua forma nativa oppure chimicamente modificata, oppure derivati polipeptidici di essa, oppure complessi specifici con altri polipeptidi

possono essere usati per produrre anticorpi, monoclonali oppure policlonali, specifici per GTBP oppure per i suoi derivati polipeptidici, oppure per i complessi di GTBP con altri polipeptidi. Anticorpi specifici per la proteina GTBP vengono prodotti immunizzando un appropriato ospite vertebrato, ad esempio coniglio oppure topo, con proteina GTBP purificata oppure con derivati di proteina GTBP, da soli oppure in combinazione con un adiuvante convenzionale. Di solito vengono applicate due o più immunizzazioni, e qualche giorno dopo l'ultima iniezione vengono raccolti sangue oppure milza. Nel caso di antisieri policlonali, le immuno globuline possono essere precipitate, isolate e purificate con una varietà di tecniche standard, che includono purificazioni di affinità usando proteina GTBP attaccata ad una superficie solida, quale un gel oppure particelle di resina in una colonna di affinità. Nel caso di anticorpi monoclonali, le cellule di milza vengono normalmente fuse con un linfocita immortalizzato, ad esempio una linea cellulare mieloide, in condizioni controllate per la formazione di ibridomi. Gli ibridomi possono essere poi clonati in condizioni di diluizione limite e, dallo

screening dei loro sopranatanti, possono essere ricavati gli anticorpi che hanno la specificità desiderata. Le tecniche per produrre gli anticorpi sono ben note in letteratura e numerose tecniche relative si trovano nella pubblicazione *Antibodies: A Laboratory Manual* (1988) edz. Halow and Line, Cold Spring Harbor Laboratories Press, e nei brevetti U.S. 4.381.292, U.S. 4.451.970 e U.S. 4.618.577.

Applicazione di GTBP in diagnostica usando sonde genetiche

Il materiale genetico dell'invenzione può essere usato tal quale in numerosi saggi come sonda per il materiale genetico presente in un individuo. L'analita può essere una sequenza nucleotidica che ibridizza con una sonda comprendente una sequenza di almeno circa 16 nucleotidi consecutivi, solitamente da 30 a 200 nucleotidi, fino a sostanzialmente l'intera sequenza del gene. L'analita può essere RNA oppure DNA. Il campione è di solito una molecola di DNA oppure RNA estratta dal tessuto del paziente. Per determinare l'analita, nel caso in cui l'analita si ibridizza con la sonda, la sonda può contenere una un marcatore determinabile. Per l'uso come

sonda sono particolarmente preferite sequenze fino a 3000 nucleotidi consecutivi (dal nucleotide 1 al nucleotide 3000 di SEQ ID NO:12), dato che queste sequenze sembrano essere particolarmente specifiche per GTBP.

Un metodo per amplificare gli acidi nucleici bersaglio, per a successiva analisi mediante saggi di ibridazione, è noto come reazione a catena della polimerasi oppure tecnica PCR. La tecnica PCR può essere applicata per determinare sequenze dell'invenzione in campioni sospetti usando primers oligonucleotidici spazati l'uno rispetto all'altro e basati sulla sequenza genetica indicata in SEQ ID NO:12. I primers sono complementari ai filamenti opposti della molecola di DNA a doppio filamento e sono di solito separati da circa 50 fino a 450 nt o più (usualmente non più di 2000 nt). Questo metodo comporta la preparazione di primers oligonucleotidici specifici e quindi ripetuti cicli di denaturazione del DNA bersaglio, legame del primer ed estensione con una DNA polimerasi termostabile per ottenere frammenti di DNA di lunghezza attesa sulla base della spaziatura dei primers. Prodotti di estensione generati da un

primer servono come sequenze bersaglio aggiuntive per l'altro primer. Il grado di amplificazione di una sequenza bersaglio viene controllato mediante l'esecuzione di molti cicli il cui numero viene teoricamente calcolato dalla semplice formula 2^n , in cui n è il numero di cicli. Posto che l'efficienza media per ciclo va da circa 65% a 85%, 25 cicli produrranno da 0,3 a 4,8 milioni di copie della sequenza bersaglio. Il metodo PCR è descritto in molte pubblicazioni, che comprendono Saiki et al. Science (1985) 230:1350-1354; Saiki et al. Nature (1986) 324:163-166; e Scharf et al, Science (1986) 233:1076-1078. Si vedano anche U.S. 4.683.194; U.S. 4.683.195 e U.S. 4.683.202.

L'invenzione comprende un metodo di diagnosi specifico per la determinazione di GTBP, basato sull'amplificazione selettiva dei frammenti di DNA che codificano per la proteina GTBP. Questo metodo impiega un paio di primer a filamento singolo derivati da regioni non omologhe delle eliche complementari del DNA codificante del GTBP che ha la sequenza descritta in SEQ ID NO:12. Questi primer ed i frammenti da essi generati rappresentano una forma di realizzazione dell'invenzione. Il metodo segue il procedimento

per l'amplificazione di sequenze di acidi nucleici selezionate descritto in U.S. 4.683.202, già discusso in precedenza.

Mutazioni nel gene *GTBP* possono essere determinate mediante analisi con enzimi di restrizione del prodotto di amplificazione oppure mediante sequenziamento diretto. Ancora, alterazioni nella sequenza *GTBP* possono essere rivelate mediante ibridazione Southern con sonde che comprendono parzialmente o totalmente la sequenza SEQ ID NO:12.

Sonde di DNA a filamento singolo, complementari alla sequenza che codifica *GTBP* di tipo selvatico, possono anche essere ibridizzate con RNA estratto da tessuti o cellule di pazienti ed usate per determinare mutazioni nel trascritto maturo del gene *GTBP* mediante digestione enzimatica dell'eteroduplex a livello di malaccoppiamenti. Queste ed altre tecniche finalizzate ad identificare variazioni nelle sequenze geniche provenienti da *GTBP* di tipo selvatico sono riportate in modo esteso in letteratura e sono ben conosciute dalla comunità scientifica.

Saggi di legame che coinvolgono *GTBP*

La presenza di una proteina GTBP alterata può essere determinata con l'uso di saggi di legame basati sul riconoscimento specifico di malaccoppiamenti G/T ad opera di GTBP. Viene preparato un oligonucleotide di 34 nucleotidi a doppio filamento contenente malaccoppiamento G/T e lo si marca sostanzialmente come indicato da Jiricny e al. (1988) Proc: Natl: Acad. Sci. USA 85:8860-8864. Gli estratti cellulari possono essere preparati come riportato nella letteratura corrente (ad esempio si veda il riferimento bibliografico 25 ed i riferimenti in esso citati). L'estratto cellulare (1-10 microgrammi di proteine nucleari) può essere incubato con un oligonucleotide eteroduplex a temperatura ambiente per 30 minuti, per permettere a GTBP di legarsi al malaccoppiamento G/T. La miscela può essere quindi caricata su un gel preparato come riportato in figura 6. Alterazioni di GTBP nella massa oppure nell'affinità al substrato possono essere evidenziate dall'alterazione nella mobilità elettroforetica.

Depositi di materiale genetico

Ceppi di E. coli TOP 10 - trasformati con i plasmidi pBluescript SK-/C e pCite-2b che

codificano rispettivamente per la proteina GTBP dall'amminoacido 1 all'amminoacido 1292 di SEQ ID NO:1 e col plasmide pBluescript SK-/FLY5 che codifica per una proteina GTBP dall'amminoacido 116 all'amminoacido 1292 di SEQ ID NO:1 - sono stati depositati il 19 maggio 1995 presso The National Collections of Industrial Marine Bacteria Ltd (NCIMB), Aberdeen, Scotland, U.K., con i numeri di accesso NCIMB 40742, NCMB 40741 e NCIMB 40740 rispettivamente.

Esempi

Come già detto in precedenza, gli inventori hanno identificato in cellule HeLa (15) un fattore di legame a malaccoppiamenti, GTBP, che ha dimostrato di legarsi preferenzialmente ad eteroduplex contenenti malaccoppiamenti. La purificazione di questa attività di legame a DNA mediante cromatografia di affinità per malaccoppiamenti G/T ha portato ad una miscela di due proteine di peso molecolare apparente 100 e 160 kDa (16), mostrando così che il sistema specifico per i malaccoppiamenti è composto di due proteine. Il costituente del complesso da 100 kDa è hMSH2 (15), mentre il secondo componente è GTBP. Qui di seguito vengono riportati esempi

riguardanti l'identità e la funzione di GTBP.

Esempio 1

Questo esempio mostra che la sequenza della proteina GTBP, riportata in SEQ ID NO:1, contiene sette sottosequenze che corrispondono ai polipeptidi ottenuti dopo taglio proteolitico della proteina di 160 kDa, cioè di GTBP. Queste sottosequenze sono indicate come SEQ ID NO 2, SEQ ID NO:3, SEQ ID NO:4 SEQ ID NO:5, SEQ ID NO:6 SEQ ID NO:7 e SEQ ID NO:8. La proteina da 160 kDa è stata purificata come riferito in (16). Le frazioni contenenti l'attività di legame specifica per il malaccoppiamento G/T vengono caricate su un gel preparativo di poliacrilammide-SDS e le bande da 160 kDa vengono escisse, rimosse e quindi sottoposte a colorazione con Coomassie Blue. Le proteine vengono digerite in matrice di gel con tripsina (proteina da 100 kDa, Promega Corporation, UK) oppure con lys-C endopeptidasi di *Achromobacter lyticus* (proteina da 160 kDa, Wako Chemicals GmbH, Germania). I peptidi proteolitici vengono recuperati mediante estrazioni sequenziali e separati mediante HPLC in tandem su un Hewlett-Packard 1090M con diode array detection. Colonne a scambio anionico e a fase inversa di ottadecile

vengono connesse in serie, essenzialmente come descritto da H. Kawasaki e K. Suzuki, Anal.Biochem. 186, 264 (1990). Le frazioni raccolte vengono applicate direttamente ad un sequenziatore automatico applied Biosystems 477A pulsed-liquid modificato come descritto da N.F. Totty, M.D. Waterfield e J.J. Hsuan, Protein Sci, 1, 1215 (1992). Il microsequenziamento porta a sette peptidi proteolitici le cui sequenze sono state indicate come SEQ ID NO:2, SEQ id no.3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 e SEQ ID NO:8.

Esempio 2

Questo esempio mostra che regioni di DNA interne al gene *GTBP* si possono ottenere per amplificazione con primers disegnati sulla base della sequenza di peptidi derivanti dal taglio proteolitico del fattore di legame a G/T da 160 kDa (SEQ ID NO:2 fino a 8). Facendo riferimento alla strategia di Lingner et al (18), gli inventori hanno identificato una sequenza unica di DNA che codifica per gli 8 amminoacidi centrali del peptide di SEQ ID NO:6. Due primers degenerati corrispondenti alle sequenze amminoacidiche N-terminale e C-terminale dell'oligopeptide di SEQ

ID NO:6, vale a dire le sequenze di DNA
 5'GCGAATTCTAYGGNTTYAAYGC3' (SEQ ID NO:9) e
 5'GCGGATCCTAYTGDATNACYTC3' (SEQ ID NO:10),
 vengono usate per l'amplificazione con PCR su
 poli-A⁺ HeLa mRNA come descritto in (17), tranne
 che la concentrazione di MgCl₂ è 5 mM. Il
 frammento atteso di 67 bp viene eluito da un gel
 di poliacrilammide, clonato in pBluescript SK- e
 sequenziato. Due cloni contengono la sequenza
 corretta, che codifica per gli 8 residui
 amminoacidici centrali (Asn Ala Ala Arg Lys Ala
 Asn Lys) del peptide bersaglio di partenza.

Esempio 3

Questo esempio mostra che regioni di DNA
 interne al gene *GTBP* possono essere identificate
 mediante ibridazione con una sonda di DNA
 disegnata sulla base della sequenza dei peptidi
 ottenuti per taglio proteolitico del fattore di
 160 kDa che si lega a G/T. La sequenza di DNA
 indicata come SEQ ID NO:11 è stata marcata con ³²P
 mediante reazione standard di chinasi (con T4 PNK
 e γ^{32} P-ATP come descritto da Maniatis et al
 Molecular cloning: a laboratory manual, Cold
 Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 1982) per generare una sonda di DNA a doppio

filamento. La sonda marcata di seq id No: 11 viene quindi usata per lo screening di una libreria commerciale in fago lambda ottenuta utilizzando mRNA arricchito su colonna di oligo dT (HeLa S3 Uni-ZAP XR, Stratagene). Due cloni positivi vengono selezionati per l'ulteriore analisi. Il clone C1 contiene un inserto di 3992 bp corrispondenti a SEQ ID NO:12 con una fase di lettura aperta continua dal residuo amminoacidico 1 al residuo amminoacidico 1292 che codifica per un polipeptide di 1292 amminoacidi (SEQ ID NO:1) con una massa molecolare calcolata di 142 kDa; il clone FLY 5 contiene sequenze che codificano dal residuo amminoacidico 156 al residuo amminoacidico 1292.

Dato che tutti i sette peptidi ottenuti dal microsequenziamento della proteina di 160 kDa (SEQ ID NO:2 fino a SEQ ID NO 8) si trovano in SEQ ID NO:1, se ne può concludere che il clone C1 codifica per GTBP.

Esempio 4

Questo esempio mostra che la proteina GTBP può essere usata come antigene per produrre anticorpi ad elevata specificità che riconoscono GTBP ma non hMSH2. Frammenti PCR corrispondenti ai

residui amminoacidici da 27 a 158 di hMSH2 (SEQ ID NO:13) e da 750 a 928 di GTBP (SEQ ID NO:14) sono stati clonati nel vettore di espressione pGEX-3X di E. coli (Pharmacia/LKB) e le proteine ricombinanti, in forma di polipeptidi di fusione con glutathione S-transferasi, vengono indotte ed isolate come viene raccomandato dal produttore, tranne che la concentrazione finale di IPTG è 0,25 mM e le colture indotte vengono raccolte dopo 6 ore a 20°C. Le proteine di fusione vengono usate per l'immunizzazione di conigli femmine New Zealand White S.P.F. (Charles River Co.) usando protocolli standard. Due antisieri policlonali, immunoreattivi in modo specifico per GTBP e hMSH2 rispettivamente, vengono ottenuti e saggiati come riportato in Antibodies: A Laboratory Manual (1988) edz.arlow and Lane, Cold Spring Harbor Laboratories Press (vedi figure 2 e 5 per ulteriori dettagli).

Esempio 5

L'esempio seguente mostra che GTBP appartiene ad una classe di proteine per la riparazione di DNA conservate in un ampio spettro evolutivo. La figura 3 mostra l'allineamento delle sequenze amminoacidiche delle regioni C-

terminali conservate delle seguenti proteine che si legano a mismatch. GTBP (*H. sapiens*), hMSH2 (*H. sapiens*), MSH2 (*S. cerevisiae*) e MutS (*E. coli*). I residui identici sono in nero, quelli conservati in riquadri ombreggiati. Le sequenze riportate nell'allineamento corrispondono alle registrazioni MSH2_YEAST (MSH2) e MUTS_ECOLI (MutS) nella banca dati SwissProt, oppure alla regione codificante della sequenza registrata come HSU04045 (hMSH2) nella GenBank. L'allineamento è stato eseguito usando l'opzione Pileup del programma GCG. La figura è stata composta usando Prettyplot. L'allineamento rivela un alto grado di conservazione nel dominio C-terminale di tutte le proteine. GTBP può pertanto essere considerato un nuovo membro della famiglia MSH.

Tuttavia, GTBP deve essere considerato strutturalmente distinto dalle proteine MSH, dato che il dominio N-terminale di GTBP (fino a circa 1000 amminoacidi) esibisce una notevole divergenza da MSH (umano, fungino o batterico). Ciò è particolarmente evidente quando vengono confrontate l'una con l'altra le matrici di omologia di hMSH2 nei confronti di MSH2 (figura 4, sezione d) e GTBP nei confronti di hMSH2 (figura

4, sezione c) oppure GTBP nei confronti di MSH2 (figura 4 sezione b). Al contrario, esiste una chiara evidenza che GTBP è conservata su un ampio spettro evolutivo e che si possono trovare omologhi strutturali dell'intera sequenza di GTBP ad esempio in lievito (numero di accesso Z47746 di GenBank, figura 4 sezione a).

Esempio 6

Questo esempio dimostra che antisieri selettivi riconoscono un complesso composto da hMSH2 e GTBP legato a DNA malaccoppiato. La figura 5 mostra l'effetto di antisieri anti-hMSH2 e anti-GTBP sulla formazione del complesso specifico che si lega al malaccoppiamento. L'analisi di ritardo su gel viene eseguita come descritto in (15), tranne che vengono usati estratti nucleari (25). Gli antisieri vengono aggiunti alle miscele di reazione 20 minuti prima della sonda marcata radioattivamente. La figura è un autoradiogramma di un gel nativo al 6% di poliacrilammide in tampone TAE. La preincubazione di estratti nucleari di HeLa con ciascuno degli antisieri prima dell'aggiunta della sonda eteroduplex G/T ha come risultato la diminuzione della banda specifica del saggio di ritardo su gel, effetto

questo non osservato quando vengono usati i rispettivi sieri pre-immuni. Questo risultato indica che entrambe le proteine sono presenti nel fattore specifico per il malaccoppiamento. Questo risultato implica anche che estratti da cellule che mancano di una di queste proteine sono prive di attività di legame al malaccoppiamento.

Esempio 7

Questo esempio mostra che GTBP ed hMSH2 possono essere espressi separatamente in un sistema di traduzione senza cellule. Gli inventori hanno impiegato un clone hMSH2 cDNA (17) e i cloni C1 e FLY5 di GTBP come indicato in SEQ ID NO: 12. Gli ORF di C1 ed FLY5 sono stati introdotti in pCite-2b. L'ORF di hMSH2 è stato inserito in pCite-1 (Novagen). Le reazioni di trascrizione e traduzione in vitro sono state eseguite come descritto in precedenza (26) inclusa una reazione di controllo (mock translation) in assenza di RNA aggiunto. I prodotti di traduzione marcati con ³⁵S sono stati analizzati su un gel di poliacrilammide-SDS trattato con Amplify (Amersham), seccati ed autoradiografati. L'esperimento è stato condotto usando le condizioni raccomandate dal produttore. La figura

è un autoradiogramma di un gel denaturante al 7,5% di poliacrilammide-SDS. Come mostrato in figura 6 sezione b, la traduzione degli mRNA di hMSH2, GTBP (C1) e FLY5 in un sistema di lisato di reticolociti (Promega) ha dato origine a polipeptidi di 113, 144 e 122 kDa rispettivamente. In tal modo la traduzione di tutti e tre gli mRNA ha dato origine a prodotti proteici delle dimensioni attese.

Esempio 8

Questo esempio mostra che GTBP si lega a mismatch G/T quando è in complesso con hMSH2. Questo è stato ottenuto sottoponendo a prova i due polipeptidi espressi in un sistema di traduzione senza cellule per quanto riguarda la loro abilità a legare substrati contenenti mismatch. La ricostituzione dell'attività di legame a mismatch usando GTBP ed hMSH2 tradotti in vitro è mostrata in figura 6 sezione c. La figura riporta una analisi di ritardo su gel che mostra il legame delle proteine tradotte in vitro all'eteroduplex G/T. Quando le proteine GTBP ed hMSH2 sono sottoposte a prova per l'attività di legame a mismatch, si è notato che l'espressione di ciascuna delle proteine da sola non ha nessun

effetto sull'intensità della banda endogena specifica per G/T presente nei lisati ai bassi livelli. Al contrario, mescolando i prodotti di traduzione di hMSH2 e GTBP si ottiene un incremento riproducibile dell'intensità della banda specifica per il mismatch. Questo risultato è confermato usando il clone FLY5 di GTBP cDNA, che codifica per una proteina GTBP tronca (si vedano SEQ ID NO:1 e SEQ ID NO:12). Mescolando i prodotti di traduzione di hMSH2 e DI FLY5 con la sonda eteroduplex G/T si origina una nuova banda con una mobilità elettroforetica maggiore di quella del complesso endogeno, quale ci sarebbe da aspettarsi per specie con dimensioni più piccole. Questo esperimento fornisce una prova convincente che il complesso che si lega al mismatch umano è composto di hMSH2 e GTBP.

I saggi di ritardo su gel sono stati eseguiti come descritto in (15). Vengono sottoposti a prova aliquote di 5 μ l delle singole reazioni di traduzione in vitro; negli esperimenti di pre-mixing, vengono mescolati 2,5 μ l di ciascuna delle due reazioni di traduzione che vengono incubati per 15 minuti a temperatura ambiente prima dell'aggiunta della sonda. AMP 5mM viene inclusa

in tutte le reazioni di legame al DNA in modo da ovviare all'effetto di ATP nei lisati reticolociti, il quale effetto previene la formazione di complessi DNA/proteina specifica per il mismatch (16). La figura è un autoradiogramma di gel nativo al 6% di poliacrilammide eseguito in tampone TAE.

E' noto che alterazioni genetiche nei geni per la riparazione di mismatch, quali *hMSH2*, *hMLH1*, *hPMS1* e *hPMS2* (1) causano la ipermutabilità riscontrata in molte forme di cancro ereditari colorettali (CRC). Verranno di seguito riportati esempi che mostrano come differenti linee cellulari derivate da CRC, che presentano fenotipo ipermutabile, contengono alleli mutati di *GTBP* i quali vengono espressi in proteine non funzionali. Gli inventori hanno anche dimostrato che lo spettro di mutazioni trovato in queste linee cellulari è differente da quello causato dall'inattivazione di *hMSH2* oppure degli altri geni per la riparazione di mismatch. I seguenti esempi confermano il ruolo di *GTBP* nel mantenimento dell'integrità genomica umano in vivo e forniscono una spiegazione per il fenotipo

mutatore osservato in differenti CRC.

Esempio 9

Questo esempio mostra che l'attività di legame al mismatch è assente in estratti di cellule LoVo e DLD1, entrambe derivate da CRC umano. Le cellule LoVo contengono una delezione omozigote in entrambi gli alleli di *hMSH2* (13), mentre nessuno dei due alleli *hMSH2* appare mutato nella linea cellulare DLD1 (19). Gli estratti di cellule LoVo e DLD1 non sono capaci di formare complessi specifici per il mismatch come è rivelato dal saggio di ritardo su gel, mostrato in figura 7 sezione a (sono state preparate sonde come descritto in precedenza (15) e le condizioni sperimentali sono quelle indicate in figura 5). La figura è un autoradiogramma di un gel nativo al 6% di poliacrilammide eseguito in tampone TAE che mostra l'assenza di complessi specifici DNA-proteina di massa molecolare attesa in entrambi gli estratti LoVo e DLD1. In base a questo sembra evidente che la linea cellulare DLD1 deve essere priva di GTBP. Risultati di conferma sono stati anche ottenuti mediante screening diretto di estratti cellulari LoVo e DLD1 con anticorpi specifici diretti contro GTBP ed *hMSH2*. Come ci si

aspettava, l'analisi Western blot di estratti HeLa ha rivelato la presenza di quantità equivalenti di hMSH2 e GTBP. Al contrario, si è potuto dimostrare che le cellule LoVo mancano di hMSH2, mentre gli estratti di DLD1 sono completamente privi di GTBP nella sua intera lunghezza (figura 7 sezione b). E' interessante notare che le quantità di hMSH2 in estratti di DLD1 e di GTBP in estratti di LoVo sono considerevolmente più basse che negli estratti di HeLa. La spiegazione che di questo fatto danno gli inventori è che hMSH2 e GTBP sono instabili quando non sono legate in un complesso (16).

Esempio 10

La linea cellulare HCT15 derivata da CRC contiene la proteina hMSH2 nella sua intera lunghezza, ma mostra un fenotipo ipermutabile (19). Per determinare se HCT15 ha una mutazione nella sequenza che codifica per GTBP, l'RNA di questa linea cellulare è stato convertito in cDNA con esanucleotidi di innesco a sequenza casuale e trascrittasi inversa secondo protocolli standard (si veda per esempio Powell et al, New Engl. J. Med. 329, 1982, 1993). Il cDNA è stato quindi amplificato con PCR usando primers specifici per

la sequenza che codifica per GTBP. Gli oligonucleotidi usati sono stati: primers 5'-PGAGGGTTACCCCTGG-3' e 5'-ACACTGTAAGTCTGTGTACC-3' per i codoni da 32 a 458, primers 5'-PGGTGAAGGCCTGAACAGCC-3' e 5'-AAGTCCAGTCTTTTCGAGCC-3' per i codoni da 219 a 858, e primers 5'-PGAGAGGGTTGATACTTGCC-3' e 5'-AGAAGTCAACTCAAAGCTTCC-3' per i codoni da 692 a 1292 (dove P denota una sequenza del promotore T7 e un sito di legame al ribosoma per l'inizio della traduzione e i numeri dei codoni sono quelli riportati in SEQ ID NO:1 e SEQ ID NO:12). Per rivelare le mutazioni nella sequenza che codifica per GTBP, i prodotti di amplificazione sono stati dapprima trascritti e tradotti in vitro usando un kit commerciale (Promega). L'analisi dei prodotti di traduzione in un gel PAGE-SDS ha rivelato polipeptidi di GTBP tronchi in due prodotti di PCR, corrispondenti a regioni localizzate sui codoni 32-458 (estremità 5' del gene) e 692-1292 (estremità 3' del gene). Il sequenziamento di questi prodotti di PCR usando un sistema commerciale (SequiTherm Polymerase, Epicentre Technologies) ha rivelato che i troncamenti erano dovuti a mutazioni da sfasamento del modulo di

lettura (mutazioni frameshift). Vi è stata una inserzione di un paio di basi al codone 222 che ha cambiato una leucina in un codone di terminazione e una delezione/sostituzione di 5 paia di basi al codone 1103 di SEQ ID NO:1 (TTGATAGAGT diventa TTTGT vedi SEQ ID NO:12) che ha creato un nuovo codone di terminazione molte paia di basi a valle.

Esempio 11

MT1 è una linea cellulare linfoblastoide resistente all'alchilazione con una deficienza biochimica simile a quella di HCT15 (vedi Goldmacher et al. J.Biol. Chem. 261, 12462, 1986; Kat et al. Proc, Natl. Acad. Sci USA, 90, 6424, 1993). Per stabilire se MT1 ha una mutazione che riguarda GTBP, l'RNA di questa linea cellulare è stato trascritto inversamente con esanucleotidi a sequenza casuale e trascrittasi inversa ed il cDNA è stato quindi amplificato con PCR usando primer specifici per la sequenza che codifica per GTBP come riportata sopra. La trascrizione e traduzione in vitro della sequenza che codifica per GTBP derivata da MT1 non ha rivelato polipeptidi tronchi derivati da GTBP dopo analisi elettroforetica. La regione codificante di GTBP è stata quindi sequenziata e sono state trovate nel

cDNA di GTBP due mutazioni non senso. La prima era una trasversione di una A in una T al codone 1144 di SEQ ID NO:1, che ha l'effetto di una sostituzione di valina in acido aspartico. La valina al codone 1144 è localizzata nel putativo dominio di legame al DNA di GTBP, e lo stesso amminoacido si ritrova nelle posizioni omologhe in GTBP (*H. sapiens*), hMSH2 (*H. sapiens*), MSH2 (*S. cerevisiae*) e MutS (*E. coli*). Questo residuo amminoacidico altamente conservato è pertanto necessario per l'attività di GTBP. Sostituzioni non conservative di questo residuo causano drammatiche distorsioni della funzionalità di GTBP. La seconda è una transizione di una G in una A, che ha l'effetto di una sostituzione di isoleucina con valina al codone 1191 di SEQ ID NO:1.

I prodotti di amplificazione sono stati clonati nel vettore BLUESCRIPT SK- e i cloni individuali sono stati sequenziati usando protocolli convenzionali (Sequenase, USB). Si è trovato che le due mutazioni non sono associate in un singolo clone, pertanto derivano da alleli separati.

Esempio 12

Una linea cellulare tumorale, denominata 543X (dalla designazione del paziente) è stata derivata da CRC e presenta fenotipo ipermutabile ed instabilità dei microsatelliti, ma nessuna mutazione nei geni per la riparazione di mismatch finora descritti, che includono *hMSH2*, *hMLH1*, *hPMS1* e *hPMS2* (Liu et al, Nature Genetics 9, 48, 1995). Per accertare se 543X ha una mutazione che riguarda GTBP, l'RNA di questa linea cellulare è stata convertita in cDNA con esanucleotidi di innesco a sequenza casuale e trascrittasi inversa. Il cDNA è stato quindi amplificato con PCR usando primers specifici per la sequenza che codifica per GTBP come riportato nell'esempio 10. La trascrizione e la traduzione in vitro della sequenza codificante per GTBP proveniente da 543X ha rivelato un polipeptide GTBP tronco dopo analisi elettroforetica. Si è trovato che la sequenza della regione di DNA che codifica per GTBP contiene un inserto di due basi al codone 626 di SEQ ID NO:1, che ha per effetto una mutazione frameshift e un codone di terminazione a valle.

LISTA DI SEQUENZE

INFORMAZIONI GENERALI

(i) DEPOSITANTE:

ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI S.p.A.

(ii) TITOLO DELL'INVENZIONE: "POLIPEPTIDE PER LA
RIPARAZIONE DI INFORMAZIONE GENETICA, SEQUENZA
NUCLEOTIDICA CHE CODIFICA PER ESSO E PROCEDIMENTO PER
LA SUA PREPARAZIONE"

(iii) NUMERO DI SEQUENZE: 14

(iv) INDIRIZZO DI CORRISPONDENZA:

(A) DESTINATARIO: Società Italiana Brevetti

(B) VIA: Piazza di Pietra, 39

(C) CITTA': Roma

(D) PAESE: Italia

(E) CODICE POSTALE: 1-00186

(v) FORMA LEGGIBILE AL COMPUTER:

(A) TIPO DI SUPPORTO: Floppy disk 3.5" 1.44 MBYTES

(B) COMPUTER: IBM PC compatibile

(C) SISTEMA OPERATIVO: PC-DOS/MS-DOS Rev. 5.0

(D) SOFTWARE: Microsoft Word 6.0

(viii) INFORMAZIONI SULL'AGENTE

(A) NOME: DI CERBO, Mario (Dr.)

(C) RIFERIMENTO: RM/X90740 IN-DC

(ix) INFORMAZIONI PER TELECOMUNICAZIONI

(A) TELEFONO: 06/6785941

(B) TELEFAX: 06/6794692

(C) TELEX: 612287 ROPAT

(1) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 1:

(i) CARATTERISTICHE DELLA SEQUENZA

(A) LUNGHEZZA: 1292 amminoacidi

(B) TIPO: amminoacido

(C) NUMERO DI CATENE: singola

(D) CONFIGURAZIONE: lineare

(ii) TIPO DI MOLECOLA: proteina

(iii) IPOTETICA: No

- (iv) ANTISENSEN: No
- (vi) SORGENTE ORIGINALE
- (A) ORGANISMO: Homo sapiens
- (vii) SORGENTE IMMEDIATA: cDNA del clone pCITE2b-C1
- (ix) CARATTERISTICHE ULTERIORI: SEQ ID NO:1 mostra la
sequenza di 1292 amminoacidi di GTBP codificato dal
clone C1 (si veda SEQ ID NO: 12). I setti
oligopeptidi che sono stati identificati a seguito
di taglio proteolitico (si vedano SEQ ID NO: 2 fino
ad 8) sono sottolineati Il primo residuo
amminoacidico del peptide codificato da FLY5 cDNA è
Asn alla posizione 116
- (A) NOME: C1
- (C) METODO DI IDENTIFICAZIONE: sperimentale
- (xi) DESCRIZIONE SEQUENZA: SEQ ID NO: 1:

Ala	Lys	Asn	Leu	Asn	Gly	Gly	Leu	Arg	Arg	Ser	Val	Ala	Pro	Ala	Ala	1	5	10	15
Pro	Thr	Ser	Cys	Asp	Phe	Ser	Pro	Gly	Asp	Leu	Val	Trp	Ala	Lys	Met	20	25	30	
Glu	Gly	Tyr	Pro	Trp	Trp	Pro	Cys	Leu	Val	Tyr	Asn	His	Pro	Phe	Asp	35	40	45	
Gly	Thr	Phe	Ile	Arg	Glu	Lys	Gly	Lys	Ser	<u>Val</u>	<u>Arg</u>	<u>Val</u>	<u>His</u>	<u>Val</u>	<u>Gln</u>	50	55	60	
<u>Phe</u>	<u>Phe</u>	<u>Asp</u>	<u>Asp</u>	<u>Ser</u>	<u>Pro</u>	<u>Thr</u>	<u>Arg</u>	<u>Gly</u>	<u>Trp</u>	<u>Val</u>	<u>Ser</u>	<u>Lys</u>	<u>Arg</u>	<u>Leu</u>	<u>Leu</u>	65	70	75	80
Lys	Pro	Tyr	Thr	Gly	Ser	Lys	Ser	Lys	Glu	Ala	Gln	Lys	Gly	Gly	His	85	90	95	
Phe	Tyr	Ser	Ala	Lys	Pro	Glu	Ile	Leu	Arg	Ala	Met	Gln	Arg	Ala	Asp	100	105	110	
Glu	Ala	Leu	Asn	Lys	Asp	Lys	Ile	Lys	Arg	Leu	Glu	Leu	Ala	Val	Cys	115	120	125	
Asp	Glu	Pro	Ser	Glu	Pro	Glu	Glu	Glu	Glu	Glu	Met	Glu	Val	Gly	Thr	130	135	140	
Thr	Tyr	Val	Thr	Asp	Lys	Ser	Glu	Glu	Asp	Asn	Glu	Ile	Glu	Ser	Glu	145	150	155	160
Glu	Glu	Val	Gln	Pro	Lys	Thr	Gln	Gly	Ser	Arg	Arg	Ser	Ser	Arg	Gln	165	170	175	
Ile	Lys	Lys	Arg	Arg	Val	Ile	Ser	Asp	Ser	Glu	Ser	Asp	Ile	Gly	Gly	180	185	190	

Ser	Asp	Val	Glu	Phe	Lys	Pro	Asp	Thr	Lys	Glu	Glu	Gly	Ser	Ser	Asp	195	200	205
Glu	Ile	Ser	Ser	Gly	Val	Gly	Asp	Ser	Glu	Ser	Glu	Gly	Leu	Asn	Ser	210	215	220
Pro	Val	Lys	Val	Ala	Arg	Lys	Arg	Lys	Arg	Met	Val	Thr	Gly	Asn	Gly	225	230	235
Ser	Leu	Lys	Arg	Lys	Ser	Ser	Arg	Lys	Glu	Thr	Pro	Ser	Ala	Thr	Lys	245	250	255
Gln	Ala	Thr	Ser	Ile	Ser	Ser	Glu	Thr	Lys	Asn	Thr	Leu	Arg	Ala	Phe	260	265	270
Ser	Ala	Pro	Gln	Asn	Ser	Glu	Ser	Gln	Ala	His	Val	Ser	Gly	Gly	Gly	275	280	285
Asp	Asp	Ser	Ser	Arg	Pro	Thr	Val	Trp	Tyr	His	Glu	Thr	Leu	Glu	Trp	290	295	300
Leu	Lys	Glu	Glu	Lys	Arg	Arg	Asp	Glu	His	Arg	Arg	Arg	Pro	Asp	His	305	310	315
Pro	Asp	Phe	Asp	Ala	Ser	Thr	Leu	Tyr	Val	Pro	Glu	Asp	Phe	Leu	Asn	325	330	335
Ser	Cys	Thr	Pro	Gly	Met	Arg	Lys	Trp	Trp	Gln	Ile	Lys	Ser	Gln	Asn	340	345	350
Phe	Asp	Leu	Val	Ile	Cys	Tyr	Lys	Val	Gly	Lys	Phe	Tyr	Glu	Leu	Tyr	355	360	365
His	Met	Asp	Ala	Leu	Ile	Gly	Val	Ser	Glu	Leu	Gly	Leu	Val	Phe	Met	370	375	380
Lys	Gly	Asn	Trp	Ala	His	Ser	Gly	Phe	Pro	Glu	Ile	Ala	Phe	Gly	Arg	385	390	395
Tyr	Ser	Asp	Ser	Leu	Val	Gln	Lys	Gly	Tyr	Lys	Val	Ala	Arg	Val	Glu	405	410	415
Gln	Thr	Glu	Thr	Pro	Glu	Met	Met	Glu	Ala	Arg	Cys	Arg	Lys	Met	Ala	420	425	430
His	Ile	Ser	Lys	Tyr	Asp	Arg	Val	Val	Arg	Arg	Glu	Ile	Cys	Arg	Ile	435	440	445
Ile	Thr	Lys	Gly	Thr	Gln	Thr	Tyr	Ser	Val	Leu	Glu	Gly	Asp	Pro	Ser	450	455	460
Glu	Asn	Tyr	Ser	Lys	Tyr	Leu	Leu	Ser	Leu	Lys	Glu	Lys	Glu	Glu	Asp	465	470	475
Ser	Ser	Gly	His	Thr	Arg	Ala	Tyr	Gly	Val	Cys	Phe	Val	Asp	Thr	Ser	485	490	495
Leu	Gly	Lys	Phe	Phe	Ile	Gly	Gln	Phe	Ser	Asp	Asp	Arg	His	Cys	Ser	500	505	510
Arg	Phe	Arg	Thr	Leu	Val	Ala	His	Tyr	Pro	Pro	Val	Gln	Val	Leu	Phe	515	520	525
Glu	Lys	Gly	Asn	Leu	Ser	Lys	Glu	Thr	Lys	Thr	Ile	Leu	Lys	Ser	Ser	530	535	540
Leu	Ser	Cys	Ser	Leu	Gln	Glu	Gly	Leu	Ile	Pro	Gly	Ser	Gln	Phe	Trp	545	550	555
Asp	Ala	Ser	Lys	Thr	Leu	Arg	Thr	Leu	Leu	Glu	Glu	Glu	Tyr	Phe	Arg	565	570	575

Glu Lys Leu Ser Asp Gly Ile Gly Val Met Leu Pro Gln Val Leu Lys
580 585 590
Gly Met Thr Ser Glu Ser Asp Ser Ile Gly Leu Thr Pro Gly Glu Lys
595 600 605
Ser Glu Leu Ala Leu Ser Ala Leu Gly Gly Cys Val Phe Tyr Leu Lys
610 615 620
Lys Cys Leu Ile Asp Gln Glu Leu Leu Ser Met Ala Asn Phe Glu Glu
625 630 635 640
Tyr Ile Pro Leu Asp Ser Asp Thr Val Ser Thr Thr Arg Ser Gly Ala
645 650 655
Ile Phe Thr Lys Ala Tyr Gln Arg Met Val Leu Asp Ala Val Thr Leu
660 665 670
Asn Asn Leu Glu Ile Phe Leu Asn Gly Thr Asn Gly Ser Thr Glu Gly
675 680 685
Thr Leu Leu Glu Arg Val Asp Thr Cys His Thr Pro Phe Gly Lys Arg
690 695 700
Leu Leu Lys Gln Trp Leu Cys Ala Pro Leu Cys Asn His Tyr Ala Ile
705 710 715 720
Asn Asp Arg Leu Asp Ala Ile Glu Asp Leu Met Val Val Pro Asp Lys
725 730 735
Ile Ser Glu Val Val Glu Leu Leu Lys Lys Leu Pro Asp Leu Glu Arg
740 745 750
Leu Leu Ser Lys Ile His Asn Val Gly Ser Pro Leu Lys Ser Gln Asn
755 760 765
His Pro Asp Ser Arg Ala Ile Met Tyr Glu Glu Thr Thr Tyr Ser Lys
770 775 780
Lys Lys Ile Ile Asp Phe Leu Ser Ala Leu Glu Gly Phe Lys Val Met
785 790 795 800
Cys Lys Ile Ile Gly Ile Met Glu Glu Val Ala Asp Gly Phe Lys Ser
805 810 815
Lys Ile Leu Lys Gln Val Ile Ser Leu Gln Thr Lys Asn Pro Glu Gly
820 825 830
Arg Phe Pro Asp Leu Thr Val Glu Leu Asn Arg Trp Asp Thr Ala Phe
835 840 845
Asp His Glu Lys Ala Arg Lys Thr Gly Leu Ile Thr Pro Lys Ala Gly
850 855 860
Phe Asp Ser Asp Tyr Asp Gln Ala Leu Ala Asp Ile Arg Glu Asn Glu
865 870 875 Glu Asn
1045 1050 1055
Gly Lys Ala Tyr Cys Val Leu Val Thr Gly Pro Asn Met Gly Gly Lys
1060 1065 1070
Ser Thr Leu Met Arg Gln Ala Gly Leu Leu Ala Val Met Ala Gln Met
1075 1080 1085
Gly Cys Tyr Val Pro Ala Glu Val Cys Arg Leu Thr Pro Ile Asp Arg
1090 1095 1100
Val Phe Thr Arg Leu Gly Ala Ser Asp Arg Ile Met Ser Gly Glu Ser
1105 1110 1115 1120
Thr Phe Phe Val Glu Leu Ser Glu Thr Ala Ser Ile Leu Met His Ala

	1125	1130	1135
Thr Ala His Ser Leu Val Leu Val Asp Glu Leu Gly Arg Gly Thr Ala			
	1140	1145	1150
Thr Phe Asp Gly Thr Ala Ile Ala Asn Ala Val Val Lys Glu Leu Ala			
	1155	1160	1165
Glu Thr Ile Lys Cys Arg Thr Leu Phe Ser Thr His Tyr His Ser Leu			
	1170	1175	1180
Val Glu Asp Tyr Ser Gln Asn Val Ala Val Arg Leu Gly His Met Ala			
1185	1190	1195	1200
Cys Met Val Glu Asn Glu Cys Glu Asp Pro Ser Gln Glu Thr Ile Thr			
	1205	1210	1215
Phe Leu Tyr Lys Phe Ile Lys Gly Ala Cys Pro Lys Ser Tyr Gly Phe			
	1220	1225	1230
<u>Asn Ala Ala Arg Leu Ala Asn Leu Pro Glu Glu Val Ile Gln Lys Gly</u>			
	1235	1240	1245
His Arg Lys Ala Arg Glu Phe Glu Lys Met Asn Gln Ser Leu Arg Leu			
	1250	1255	1260
Phe Arg Glu Val Cys Leu Ala Ser Glu Arg Ser Thr Val Asp Ala Glu			
1265	1270	1275	1280
Ala Val His Lys Leu Leu Thr Leu Ile Lys Glu Leu			
	1285	1290	

(2) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 2:

(i) CARATTERISTICHE DELLA SEQUENZA

- (A) LUNGHEZZA: 10 amminoacidi
- (B) TIPO: amminoacido
- (C) NUMERO DI CATENE: singola
- (D) TOPOLOGIA: lineare

(ii) TIPO DI MOLECOLA: proteina

(iii) IPOTETICA: No

(iv) ANTISENSEN: No

(vi) SORGENTE ORIGINALE:

(A) ORGANISMO: Homo sapiens

(ix) CARATTERISTICHE ULTERIORI: SEQ ID NO: 2 to 8 mostrano i sette oligopeptidi derivati dal taglio proteolitico di GTBP estratto da cellule HeLa e purificato come descritto nel riferimento bibliografico (16). Il peptide corrispondente a SEQ ID NO: 6 (18 amminoacidi) è stato selezionato per disegnare due primers degenerati corrispondenti alle sequenze N-terminale e C-terminale del peptide indicate in dettaglio in SEQ ID NO: 9 e SEQ ID NO: 10

(A) NOME: FR44

(C) METODO DI IDENTIFICAZIONE: sperimentale

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 2:

Val	Arg	Val	His	Val	Gln	Phe	Phe	Asp	Asp
1				5				10	

(3) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 3:

(i) CARATTERISTICHE DELLA SEQUENZA

(A) LUNGHEZZA: 18 amminoacidi

(B) TIPO: amminoacido

(C) NUMERO DI CATENE: singola

(D) TIPOLOGIA: lineare

(ii) TIPO DI MOLECOLA: proteina

(iii) IPOTETICA: No

(iv) ANTISENSEN: No

(vi) SORGENTE ORIGINALE:

(A) ORGANISMO: Homo sapiens

(ix) CARATTERISTICHE ULTERIORI: vedere SEQ ID NO:2

(A) NOME: FR48

(C) METODO DI IDENTIFICAZIONE : sperimentale

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 3:

Lys	Leu	Pro	Asp	Leu	Glu	Arg	Leu	Leu	Ser	Lys	Ile	His	Asn	Val	Xaa
1				5					10					15	
Ser	Lys														

(4) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 4:

(i) CARATTERISTICHE DELLA SEQUENZA

(A) LUNGHEZZA: 13 amminoacidi

(B) TIPO: amminoacido

(C) NUMERO DI CATENE: singola

(D) CONFIGURAZIONE: lineare

(ii) TIPO DI MOLECOLA: proteina

(iii) IPOTETICA: No

(iv) ANTISENSEN: No

(vi) SORGENTE ORIGINALE:

(A) ORGANISMO: Homo sapiens

(ix) CARATTERISTICHE ULTERIORI: vedere SEQ ID NO 2

- (A) NOME: FR49b
 (C) METODO DI IDENTIFICAZIONE: sperimentale
 (xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO:4:

Leu Ser Arg Gly Iso Gly Val Met Leu Pro Gln Val Leu
 1 5 10

- (5) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 5:
 (i) CARATTERISTICHE DELLA SEQUENZA
 (A) LUNGHEZZA: 14 amminoacidi
 (B) TIPO: amminoacido
 (C) NUMERO DI CATENE: singola
 (D) CONFIGURAZIONE: lineare
 (ii) TIPO DI MOLECOLA: proteina
 (iii) IPOTETICA::No
 (iv) ANTISENSO: No
 (vi) SORGENTE ORIGINALE:
 (A) ORGANISMO: Homo sapiens
 (ix) CARATTERISTICHE ULTERIORI: vedere SEQ ID NO: 2
 (A) NOME: FR49c
 (C) METODO DI IDENTIFICAZIONE: sperimentale
 (xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 5:
 Thr Leu Arg Thr Leu Leu Glu Glu Glu Tyr Phe Arg Glu Lys
 1 5 10

- (6) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 6:
 (i) CARATTERISTICHE DELLA SEQUENZA
 (A) LUNGHEZZA: 18 amminoacidi
 (B) TIPO: amminoacido
 (C) NUMERO DI CATENE: singola
 (D) CONFIGURAZIONE: lineare
 (ii) TIPO DI MOLECOLA: proteina
 (iii) IPOTETICA: No
 (iv) ANTISENSO: No
 (vi) SORGENTE ORIGINALE:
 (A) ORGANISMO: HeLa cell extract
 (ix) CARATTERISTICHE ULTERIORI: vedere SEQ ID NO: 2
 (A) NOME: FR52

(C) METODO DI IDENTIFICAZIONE: sperimentale

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 6:

Ser Tyr Gly Phe Asn Ala Ala Arg Leu Ala Asn Leu Pro Glu Glu Val
1 5 10 15
Ile Gln

(7) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 7:

(i) CARATTERISTICHE DELLA SEQUENZA

(A) LUNGHEZZA: 13 amminoacidi

(B) TIPO: amminoacido

(C) NUMERO DI CATENE: singola

(D) CONFIGURAZIONE: lineare

(ii) TIPO DI MOLECOLA: proteina

(iii) IPOTETICA: No

(iv) ANTISENSO: No

(vi) SORGENTE ORIGINALE:

(A) ORGANISMO: Homo sapiens

(ix) CARATTERISTICHE ULTERIORI: vedere SEQ ID NO: 2

(A) NOME: FR59

(C) METODO DI IDENTIFICAZIONE: sperimentale

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 7:

Asn Pro Glu Gly Arg Phe Pro Asp Leu Thr Val Glu Leu

1 5 10

(8) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 8:

(i) CARATTERISTICHE DELLA SEQUENZA

(A) LUNGHEZZA: 11 amminoacidi

(B) TIPO: amminoacido

(C) NUMERO DI CATENE: singola

(D) CONFIGURAZIONE: lineare

(ii) TIPO DI MOLECOLA: proteina

(iii) IPOTETICA: No

(iv) ANTISENSO: No

(vi) SORGENTE ORIGINALE:

(A) ORGANISMO: Homo sapiens

(ix) CARATTERISTICHE ULTERIORI: vedere SEQ ID NO:2

(A) NOME: FR69

(C) METODO DI IDENTIFICAZIONE: sperimentale

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 8:

Ile Ile Asp Phe Leu Ser Ala Leu Glu Gly Phe

1

5

10

(9) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 9

(i) CARATTERISTICHE DELLA SEQUENZA

(A) LUNGHEZZA: 22 paia di basi

(B) TIPO: acido nucleico

(C) NUMERO DI CATENE: singola

(D) CONFIGURAZIONE: lineare

(ii) TIPO DI MOLECOLA: DNA sintetico

(iii) IPOTETICA: No

(iv) ANTISENSEN : No

(vii) SORGENTE IMMEDIATA: sintetizzatore di oligonucleotid

(ix) CARATTERISTICHE ULTERIORI: SEQ ID NO:9 mostra la sequenza del primer degenerato di DNA a filamento singolo dedotto dall'oligonucleotide mostrato in SEQ ID NO:6. Insieme a SEQ ID NO:10, i due primers sono usati per amplificare poly-A⁺RNA estratto da cellule HeLa. L'atteso frammento di 67 paia di basi è stato clonato in pBluescript SK⁺ (stratagene) e sequenziato con un kit commerciale basato su T7-polimerasi (Pharmacia). La sequenza di 64 paia di basi del frammento risultante ottenuto dopo sottrazione dei siti di clonaggio ingegnerizzati, viene mostrato in SEQ ID NO:11

(A) NOME: oligo 5' senso

(C) METODO DI IDENTIFICAZIONE: gel di poli-acrilammide

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 9

GCGAATTCTA YGGNTTYAAY GC

22

(10) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 10

(i) CARATTERISTICHE DELLA SEQUENZA

(A) LUNGHEZZA: 22 paia di basi

(B) TIPO: acido nucleico

(C) NUMERO DI CATENE: singola

(D) CONFIGURAZIONE: lineare

(ii) TIPO DI MOLECOLA: DNA sintetico

(iii) IPOTETICA: No

(iv) ANTISENSEN: Sì

(vii) SORGENTE IMMEDIATA: sintetizzatore di oligonucleotidi

(ix) CARATTERISTICHE ULTERIORI: SEQ ID NO: 10 mostra la sequenza del primer degenerato di DNA a filamento singolo dedotto dalla porzione C-terminale dell'oligonucleotide mostrato in SEQ ID NO:6. Insieme a SEQ ID NO:9, i due primers sono usati per amplificare poly A⁺ RNA estratto da cellule HeLa. L'atteso frammento di 67 paia di basi è stato clonato in pBluescript SK⁻ (Stratagene) e sequenziato con un kit commerciale basato su T7-polimerasi (Pharmacia). La sequenza di 54 paia di basi del frammento risultante, ottenuto dopo sottrazione dei siti di clonaggio ingegnerizzati, viene mostrato in SEQ ID NO: 11

(A) NOME: oligo 3' antisense

(C) METODO DI IDENTIFICAZIONE: gel di poli-acrilammide

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 10
GCGGATCCTC YTG DATNACY TC 22

(11) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 11

(i) CARATTERISTICHE DELLA SEQUENZA

(A) LUNGHEZZA: 54 paia di basi

(B) TIPO: acido nucleico

(C) NUMERO DI CATENE: doppio

(D) CONFIGURAZIONE: lineare

(ii) TIPO DI MOLECOLA: DNA sintetico

(iii) IPOTETICA: No

(iv) ANTISENTO: Sì

(vii) SORGENTE IMMEDIATA: Prodotto di PCR

(ix) CARATTERISTICHE ULTERIORI: SEQ ID NO: 11 mostra la sequenza a doppio filamento di DNA che codifica per l'oligo peptide di SEQ ID NO:6, dedotta sequenziando il prodotto di amplificazione clonato. Questo frammento è derivato da amplificazione PCR di HeLa cDNA, usando i primers degenerati descritti in SEQ ID NO:9 e 10. la sequenza di DNA è stata marcata all'estremità con ³²P mediante una reazione di chinasi standard (con T4 polinucleotide chinasi PNK come descritto in

Maniatis et al , Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982) per generare una sonda di DNA a doppio filamento. La sonda marcata è stata usata nello screening di una libreria genomica commerciale di cDNA nel fago lambda ottenuta utilizzando mRNA su colonna di oligo dT (HeLa S3 UNI-ZAP XR, Stratagene). Lo screening della libreria HeLa S3 UNI ZAP XR in fago lambda ha reso possibile l'identificazione di due cloni che ibridizzano con la sonda di DNA. Questi cloni sono stati chiamati C1 e FLY5

(A) NOME:

(C) METODO DI IDENTIFICAZIONE: gel di poliacrilammide

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 11

AGCTATGGCT TTAATGCAGC AAGGCTTGCT AATCTCCAG AGGAAGTTAT TCAA

54

(12) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 12

(i) CARATTERISTICHE DELLA SEQUENZA

(A) LUNGHEZZA 3980 paia di basi

(B) TIPO: acido nucleico

(C) NUMERO DI CATENE: doppio

(D) CONFIGURAZIONE: lineare

(ii) TIPO DI MOLECOLA: DNA sintetico

(iv) ANTISENSO: No

(vii) SORGENTE IMMEDIATA: cDNA del clone C1

(ix) CARATTERISTICHE ULTERIORI: SEQ ID NO: 12 mostra la sequenza di 3980 paia di basi del cDNA del clone C1. L'inserito di cDNA del clone FLY5 si estende dal nucleotide 346 al 3980 della sequenza C1 riportata in SEQ ID NO 12

(A) NOME: C1

(C) METODO DI IDENTIFICAZIONE: gel di poliacrilammide

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 12

GCGAAGAACC TCAACGGAGG GCTGCGGAGA TCGGTAGCGC CTGCTGCCCC CACCAGTTGT	60
GACTTCTCAC CAGGAGATTT GGTGGGGCC AAGATGGAGG GTTACCCCTG GTGGCCTTGT	120
CTGGTTTACA ACCACCCCTT TGATGGAACA TTCATCCGCG AGAAAGGGAA ATCAGTCCGT	180

GTTTCATGTAC	AGTTTTTTTGA	TGACAGCCCA	ACAAGGGGCT	GGGTTAGCAA	AAGGCTTTTA	240
AAGCCATATA	CAGGTTCAAA	ATCAAAGGAA	GCCCAGAAGG	GAGGTCATTT	TTACAGTGCA	300
AAGCCTGAAA	TACTGAGAGC	AATGCAACGT	GCAGATGAAG	CCTTAAATAA	AGACAAGATT	360
AAGAGGCTTG	AATTGGCAGT	TTGTGATGAG	CCCTCAGAGC	CAGAAGAGGA	AGAAGAGATG	420
GAGGTAGGCA	CAACTTACGT	AACAGATAAG	AGTGAAGAAG	ATAATGAAAT	TGAGAGTGAA	480
GAGGAAGTAC	AGCCTAAGAC	ACAAGGATCT	AGGCCGAAGTA	GCCGCCAAAT	AAAAAACGA	540
AGGGTCATAT	CAGATTCTGA	GAGTGACATT	GGTGGCTCTG	ATGTGGAATT	TAAGCCAGAC	600
ACTAAGGAGG	AAGGAAGCAG	TGATGAAATA	AGCAGTGGAG	TGGGGGATAG	TGAGAGTGAA	660
GGCCTGAACA	GCCCTGTCAA	AGTTGCTCGA	AAGCGGAAGA	GAATGGTGAC	TGGAAATGGC	720
TCTCTTAAAA	GGAAAAGCTC	TAGGAAGGAA	ACGCCCTCAG	CCACCAAACA	AGCAACTAGC	780
ATTTTCATCAG	AAACCAAGAA	TACTTTGAGA	GCTTTCTCTG	CCCCTCAAAA	TTCTGAATCC	840
CAAGCCCACG	TTAGTGGAGG	TGGTGATGAC	AGTAGTCGCC	CTACTGTTTG	GTATCATGAA	900
ACTTTAGAAT	GGCTTAAGGA	GGAAAAGAGA	AGAGATGAGC	ACAGGAGGAG	GCCTGATCAC	960
CCCGATTTTG	ATGCATCTAC	ACTCTATGTG	CCTGAGGATT	TCCTCAATTC	TTGTACTCCT	1020
GGGATGAGGA	AGTGGTGGCA	GATTAAGTCT	CAGAACTTTG	ATCTTGTCAT	CTGTTACAAG	1080
GTGGGGAAAT	TTTATGAGCT	GTACCACATG	GATGCTCTTA	TTGGAGTCAG	TGAACTGGGG	1140
CTGGTATTCA	TGAAAGGCAA	CTGGGCCCAT	TCTGGCTTTC	CTGAAATTGC	ATTTGGCCGT	1200
TATTCAGATT	CCCTGGTGCA	GAAGGGCTAT	AAAGTAGCAC	GAGTGGAACA	GACTGAGACT	1260
CCAGAAATGA	TGGAGGCACG	ATGTAGAAAG	ATGGCACATA	TATCCAAGTA	TGATAGAGTG	1320
GTGAGGAGGG	AGATCTGTAG	GATCATTACC	AAGGGTACAC	AGACTTACAG	TGTGCTGGAA	1380
GGTGATCCCT	CTGAGAACTA	CAGTAAGTAT	CTTCTTAGCC	TCAAAGAAAA	AGAGGAAGAT	1440
TCTTCTGGCC	ATACTCGTGC	ATATGGTGTG	TGCTTTGTTG	ATACTTCACT	GGGAAAGTTT	1500
TTCATAGGTC	AGTTTTTCAGA	TGATCGCCAT	TGTTTCGAGAT	TTAGGACTCT	AGTGGCACAC	1560
TATCCCCCAG	TACAAGTTTT	ATTTGAAAAA	GGAAATCTCT	CAAAGGAAAC	TAAAACAATT	1620
CTAAAGAGTT	CATTGTCCTG	TTCTCTTCAG	GAAGGTCTGA	TACCCGGCTC	CCAGTTTTTG	1680
GATGCATCCA	AACTTTGAG	AACTCTCCTT	GAGGAAGAAT	ATTTTAGGGA	AAAGCTAAGT	1740
GATGGCATTG	GGGTGATGTT	ACCCAGGTG	CTTAAAGGTA	TGACTTCAGA	GTCTGATTCC	1800
ATTGGGTTGA	CACCAGGAGA	GAAAAGTGAA	TTGGCCCTCT	CTGCTCTAGG	TGGTTGTGTC	1860
TTCTACCTCA	AAAAATGCCT	TATTGATCAG	GAGCTTTTAT	CAATGGCTAA	TTTTGAAGAA	1920
TATATTCCCT	TGGATTCTGA	CACAGTCAGC	ACTACAAGAT	CTGGTGCTAT	CTTCACCAAA	1980
GCCTATCAAC	GAATGGTGCT	AGATGCAGTG	ACATTAAACA	ACTTGAGAT	TTTTCTGAAT	2040
GGAACAAATG	GTTCTACTGA	AGGAACCCTA	CTAGAGAGGG	TTGATACTTG	CCATACTCCT	2100

TTTGGTAAGC	GGCTCCTAAA	GCAATGGCTT	TGTGCCCCAC	TCTGTAACCA	TTATGCTATT	2160
AATGATCGTC	TAGATGCCAT	AGAAGACCTC	ATGGTTGTGC	CTGACAAAAT	CTCCGAAGTT	2220
GTAGAGCTTC	TAAAGAAGCT	TCCAGATCTT	GAGAGGCTAC	TCAGTAAAAT	TCATAATGTT	2280
GGGTCTCCCC	TGAAGAGTCA	GAACCACCCA	GACAGCAGGG	CTATAATGTA	TGAAGAACT	2340
ACATACAGCA	AGAAGAAGAT	TATTGATTTT	CTTTCTGCTC	TGGAAGGATT	CAAAGTAATG	2400
TGTAAAATTA	TAGGGATCAT	GGAAGAAGTT	GCTGATGGTT	TTAAGTCTAA	AATCCTTAAG	2460
CAGGTCACTT	CTCTGCAGAC	AAAAATCCT	GAAGGTCGTT	TTCTTGATTT	GACTGTAGAA	2520
TTGAACCGAT	GGGATACAGC	CTTTGACCAT	GAAAAGGCTC	GAAAGACTGG	ACTTATTACT	2580
CCCAAAGCAG	GCTTTGACTC	TGATTATGAC	CAAGCTCTTG	CTGACATAAG	AGAAAATGAA	2640
CAGAGCCTCC	TGGAATACCT	AGAGAAACAG	CGCAACAGAA	TTGGCTGTAG	GACCATAGTC	2700
TATTGGGGGA	TTGGTAGGAA	CCGTTACCAG	CTGGAAATTC	CTGAGAATTT	CACCACTCGC	2760
AATTTGCCAG	AAGAATACGA	GTTGAAATCT	ACCAAGAAGG	GCTGTAAACG	ATACTGGACC	2820
AAAATATTG	AAAAGAAGTT	GGCTAATCTC	ATAAATGCTG	AAGAACGGAG	GGATGTATCA	2880
TTGAAGGACT	GCATGCGGCG	ACTGTTCTAT	AACTTTGATA	AAAATTACAA	GGACTGGCAG	2940
TCTGCTGTAG	AGTGTATCGC	AGTGTGGAT	GTTTTACTGT	GCCTGGCTAA	CTATAGTCGA	3000
GGGGGTGATG	GTCCTATGTG	TCGCCCAGTA	ATTCTGTTGC	CGGAAGATAC	CCCCCCTTC	3060
TTAGAGCTTA	AAGGATCAGC	CCATCCTTGC	ATTACGAAGA	CTTTTTTTGG	AGATGATTTT	3120
ATTCCTAATG	ACATTCTAAT	AGGCTGTGAG	GAAGAGGAGC	AGGAAAATGG	CAAAGCCTAT	3180
TGTGTGCTTG	TTACTGGACC	AAATATGGGG	GGCAAGTCTA	CGCTTATGAG	ACAGGCTGGC	3240
TTATTAGCTG	TAATGGCCCA	GATGGGTTGT	TACGTCCCTG	CTGAAGTGTG	CAGGCTCACA	3300
CCAATTGATA	GAGTGTTTAC	TAGACTTGGT	GCCTCAGACA	GAATAATGTC	AGGTGAAAGT	3360
ACATTTTTTG	TTGAATTAAG	TGAAACTGCC	AGCATACTCA	TGCATGCAAC	AGCACATTCT	3420
CTGGTGCTTG	TGGATGAATT	AGGAAGAGGT	ACTGCAACAT	TTGATGGGAC	GGCAATAGCA	3480
AATGCAGTTG	TTAAAGAACT	TGCTGAGACT	ATAAAATGTC	GTACATTATT	TTCAACTCAC	3540
TACCATTTCAT	TAGTAGAAGA	TTATTCTCAA	AATGTTGCTG	TGCGCCTAGG	ACATATGGCA	3600
TGCATGGTAG	AAAATGAATG	TGAAGACCCC	AGCCAGGAGA	CTATTACGTT	CCTCTATAAA	3660
TTCATTAAGG	GAGCTTGTC	TAAAAGCTAT	GGCTTTAATG	CAGCAAGGCT	TGCTAATCTC	3720
CCAGAGGAAG	TTATTCAAAA	GGGACATAGA	AAAGCAAGAG	AATTTGAGAA	GATGAATCAG	3780
TCACTACGAT	TATTTGCGGA	AGTTTGCTG	GCTAGTGAAA	GGTCAACTGT	AGATGCTGAA	3840
GCTGTCCATA	AATTGCTGAC	TTTGATTAAG	GAATTATAGA	CTGACTACAT	TGGAAGCTTT	3900
GAGTTGACTT	CTGACCAAAG	GTGGTAAATT	CAGACAACAT	TATGATCTAA	TAAACTTTAT	3960
TTTTTAAAAA	TGAAAAAAA					3980

(13) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 13

(i) CARATTERISTICHE

(A) LUNGHEZZA 394 paia di basi

(B) TIPO: acido nucleico

(C) NUMERO DI CATENE: doppio

(D) CONFIGURAZIONE: lineare

(ii) TIPO DI MOLECOLA: sintetica DNA

(iii) IPOTETICA No

(iv) ANTISENSEN: No

(vii) SORGENTE IMMEDIATA: Homo sapiens

(ix) CARATTERISTICHE ULTERIORI: SEQ ID NO: 13 mostra la sequenza di DNA a doppio filamento usata per esprimere un dominio interno di hMSH2 (corrispondente ai residui amminoacidici da 27 a 158) nel vettore di espressione pGEX-3x (si veda anche la legenda di figura 2)

(A) NOME: GST/hMSH2

(C) METODO DI IDENTIFICAZIONE: gel di poliacrilammide

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 13

GGAGAAGCCG ACCACCACAG TGC GCCTTTT CGACCGGGGC GACTTCTATA CGGCGCACGG	60
CGAGGACGCG CTGCTGGCCG CCCGGGAGGT GTTCAAGACC CAGGGGGTGA TCAAGTACAT	120
GGGGCCGGCA GGAGCAAAGA ATCTGCAGAG TGTTGTGCTT AGTAAATGA ATTTTGAATC	180
TTTTGTAAAA GATCTTCTTC TGGTTCGTCA GTATAGAGTT GAAGTTTATA AGAATAGAGC	240
TGGAAATAAG GCATCCAAGG AGAATGATTG GTATTTGGCA TATAAGGCTT CTCCTGGCAA	300
TCTCTCTCAG TTTGAAGACA TTCTCTTTGG TAACAATGAT ATGTCAGCTT CCATTGGTGT	360
TGTGGGTGTT AAAATGTCCG CAGTTGATGG CCAG	394

(14) INFORMAZIONI SULLA SEQUENZA SEQ ID NO: 14

(i) CARATTERISTICHE DELLA SEQUENZA

(A) LUNGHEZZA: 534 paia di basi

(B) TIPO: acido nucleico

(C) NUMERO DI CATENE: doppio

(D) CONFIGURAZIONE: lineare

(ii) TIPO DI MOLECOLA: DNA sintetico

(iii) IPOTETICA: No

(iv) ANTISENSEN: No

(vii) SORGENTE IMMEDIATA:

(ix) CARATTERISTICHE ULTERIORI: SEQ ID NO: 14 mostra la sequenza di DNA a doppio filamento usata per esprimere un dominio interno di GTBP (corrispondente ai residui amminoacidici da 750 a 928) nel vettore di espressione pGEX-3x (si veda anche la legenda di figura 2)

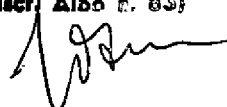
(A) NOME: GST/GTBP

(C) METODO DI IDENTIFICAZIONE: gel di poliacrilammide

(xi) DESCRIZIONE DELLA SEQUENZA: SEQ ID NO: 14

CTTGAGAGGC TACTCAGTAA AATTCATAAT GTTGGGTCTC CCCTGAAAGT CAGAACCACC	60
CAGACAGCAG GGCTATAATG TATGAAGAAA CTACATACAG CAAGAAGAAG ATTATTGATT	120
TTCTTTCTGC TCTGGAAGGA TTCAAAGTAA TGTGTAAAT TATAGGGATC ATGGAAGAAG	180
TTGCTGATGG TTTTAAGTCT AAAATCCTTA AGCAGGTCAT CTCTCTGCAG ACAAAAAATC	240
CTGAAGGTCTG TTTTCCTGAT TTGACTGTAG AATTGAACCG ATGGGATACA GCCTTTGACC	300
ATGAAAAGGC TCGAAAGACT GGACTTATTA CTCCCAAAGC AGGCTTTGAC TCTGATTATG	360
ACCAAGCTCT TGCTGACATA AGAGAAAATG AACAGAGCCT CCTGGAATAC CTAGAGAAAC	420
AGCGCAACAG AATTGGCTGT AGGACCATAG TCTATGGATT GGTAGGAACC GTTACGCAGC	480
TGGAAATTCC TGAGAATTTC ACCACTCGCA ATTTGCCAGA AGAATACGAG TTGA	534

Giulberto Tonon
(lacr. Albo n. 83)




RIVENDICAZIONI

1. Polipeptide isolato, caratterizzato dal fatto di essere scelto dal gruppo comprendente: (1) una prima sequenza corrispondente a GTBP come indicata in SEQ ID NO: 1; (2) una seconda sequenza che è una sottosequenza di detta prima sequenza ed è costituita da almeno quattro amminoacidi; oppure (3) una terza sequenza in cui almeno un amminoacido è sostituito da un diverso amminoacido.
2. Polipeptide come da rivendicazione 1, complessato con un secondo polipeptide.
3. Polipeptide complessato come da rivendicazione 2, in cui detto secondo polipeptide è hMSH2.
4. Polipeptide isolato come da rivendicazione 1 che comprende sequenze amminoacidiche tra l'amminoacido 1 e il 1292 di SEQ ID NO: 1, ad esempio le sequenze da SEQ ID NO: 2 a SEQ ID NO: 8.
5. Molecola isolata di DNA oppure di RNA, caratterizzata dal fatto di essere scelta dal gruppo comprendente: (1) una prima sequenza che codifica per GTBP come indicata in SEQ ID NO: 12;

(2) una seconda sequenza, che è una sottosequenza di detta prima sequenza ed ha una lunghezza di almeno 10 nucleotidi; (3) una terza sequenza in cui almeno un nucleotide di detta prima oppure seconda sequenza è sostituito con un diverso nucleotide; oppure (4) una quarta sequenza complementare a detta prima, seconda oppure terza sequenza;

a condizione che (i) se detta molecola è una molecola di RNA, U sostituisce T in detta sequenza di detta molecola, (ii) detta terza sequenza è almeno al 95% identica a detta prima oppure seconda sequenza, e (iii) detta seconda sequenza non è presente in hMSH2 cDNA.

6. Molecola come da rivendicazione 5, detta molecola comprendendo detta prima sequenza.

7. Molecola come da rivendicazione 5, detta molecola comprendendo detta seconda sequenza.

8. Molecola come da rivendicazione 5, detta molecola comprendendo detta terza sequenza.

9. Molecola come da rivendicazione 5, detta molecola comprendendo una sequenza di cDNA.

10. Molecola come da rivendicazione 5, detta molecola consistendo essenzialmente di DNA che codifica per GTBP.

11. Molecola come da rivendicazione 5, in cui l'RNA oppure il DNA che codifica per GTPB è quello che esiste in natura.
12. Vettore di espressione che contiene una delle molecole delle rivendicazioni da 5 a 10 .
13. Cellula trasformata con una delle molecole delle rivendicazioni da 5 a 10.
14. Cellula come da rivendicazione 13, in cui detta molecola è DNA e detto DNA è legato in associazione operativa con una sequenza di controllo capace di dirigere la replicazione e l'espressione di detto DNA.
15. Cellula come da rivendicazione 13, detta cellula essendo una cellula eucariote oppure procariote, incluse cellule animali, fungine oppure batteriche.
16. Procedimento per la produzione della proteina GTBP che comprende le operazioni di coltivare una cellula come da rivendicazione 13 in un mezzo di coltura adatto e di isolare detta proteina GTBP da detta cellula.
17. Polipeptide preparato secondo il procedimento di rivendicazione 16.
18. Metodo per identificare agenti che inibiscono o esaltano l'attività di GTBP come la si può

valutare mediante saggi in vitro di multimerizzazione oppure dimerizzazione, saggi di legame a DNA e saggi per la riparazione di mismatch.

19. Metodo per identificare agenti che modulano GTBP comprendente le fasi di: (1) eseguire una eterodimerizzazione che include un polipeptide di GTBP, hMSH2 ed un agente, e (2) determinare se l'agente modula l'eterodimerizzazione.

20. Metodo come da rivendicazione 19, in cui il saggio di eterodimerizzazione comprende una reazione di legame in vitro.

21. Preparazione di anticorpi specifici immuno-reattivi nei confronti di GTBP e sostanzialmente non immuno-reattivi nei confronti di altre proteine non correlabili a GTBP.

22. Metodo di purificazione di GTBP oppure di molecole complessanti GTBP, che comporta l'uso degli anticorpi specifici di rivendicazione 21.

23. Metodo di purificazione di GTBP oppure di molecole complessanti GTBP, basato sulla interazione specifica tra GTBP e sequenze di riconoscimento di acidi nucleici.

24. Metodo per determinare la presenza di un difetto genetico che è potenzialmente causa di

cancerogenesi negli esseri umani, che comprende l'identificazione di una mutazione di un gene *GTBP* di detto essere umano, in cui detta mutazione determina una sequenza del gene *GTBP* diversa dalla sequenza di DNA che codifica per *GTBP* umano di tipo selvatico come indicata in SEQ ID NO: 12.

25. Metodo per determinare la presenza di un difetto genetico che causa cancro negli esseri umani, che comprende l'identificazione di una mutazione di un gene *GTBP* di detto essere umano, detta mutazione comportando una sequenza del gene *GTBP* diversa dalla sequenza nucleotidica di *GTBP* umano come indicata in SEQ ID NO: 12, con l'effetto di cambiare la sequenza del prodotto proteico di detto gene *GTBP*, oppure di causare il troncamento di detto prodotto *GTBP* oppure di determinare la non traduzione oppure trascrizione di detto gene *GTBP*.

26. Metodo di diagnosi o prognosi di un tessuto neoplastico di un essere umano comprendente l'identificazione della presenza di una mutazione di un gene *GTBP* oppure del suo prodotto di espressione in detto tessuto, detta mutazione comportando una sequenza genica *GTBP* diversa dalla sequenza nucleotidica di *GTBP* umano come indicata

in SEQ ID NO: 12, e detta alterazione provocando neoplasia del tessuto.

27. Metodi come da rivendicazioni da 24 a 26, in cui dette mutazioni producono variazioni nella sequenza di un prodotto proteico di detto gene *GTBP*.

28. Metodi come da rivendicazioni da 24 a 26, in cui dette mutazioni hanno l'effetto di non far trascrivere oppure tradurre il gene *GTBP*.

29. Metodi come da rivendicazione da 24 a 26, in cui dette mutazioni creano codoni di stop in detto gene *GTBP*.

30. Metodi come da rivendicazioni da 24 a 26, in cui detti metodi comprendono amplificazione di almeno un segmento di detto gene *GTBP* mediante Polymerase Chain Reaction (PCR).

31. Metodi come da rivendicazioni da 24 a 26, in cui detti metodi comprendono l'identificazione di una variazione in un sito di restrizione per effetto di detta mutazione.

32. Metodi come da rivendicazioni da 24 a 26, in cui detti metodi comprendono analisi del poliformismo della lunghezza dei frammenti di restrizione, ibridazione allele-specifica con oligonucleotidi oppure sequenziamento del DNA.

33. Metodi come da rivendicazioni da 24 a 26, in cui detti metodi consentono di classificare detto essere umano come omozigote per detto gene *GTBP* o per detto gene *GTBP* mutato oppure eterozigote per detto gene *GTBP* o per detto gene *GTBP* mutato.

34. Metodi come da rivendicazione da 24 a 26, in cui i prodotti di espressione sono molecole di RNA.

35. Metodi come da rivendicazioni da 24 a 26, in cui la perdita della sequenza che codifica per *GTBP* di tipo selvatico è determinata mediante ibridazione Northern delle molecole di RNA estratte da cellule oppure tessuti.

36. Metodi come da rivendicazioni da 24 a 26, in cui la perdita del gene *GTBP* di tipo selvatico è determinata mediante ibridazione Southern tra una sonda di DNA contenuta nel gene *GTBP* ed un DNA genomico di detto paziente umano.

37. Metodi come da rivendicazioni da 24 a 26, in cui la perdita del gene *GTBP* di tipo selvatico è determinata identificando un mismatch tra acidi nucleici che includono (a) molecole di mRNA di detto paziente umano e (b) un acido nucleico complementare alla sequenza che codifica per *GTBP* umana di tipo selvatico, quando molecole di tipo

(a) e (b) sono ibridizzate tra loro e formano un ibrido a doppia elica.

38. Metodi come da rivendicazioni da 24 a 26, in cui la perdita del gene di tipo selvatico è determinata mediante clonaggio genico e sequenziamento del DNA clonato.

39. Metodi come da rivendicazioni da 24 a 26, in cui la perdita del gene *GTBP* di tipo selvatico è determinata mediante screening di mutazioni puntiformi, e di mutazioni per delezione oppure per inserzione.

40. Metodi come da rivendicazioni da 24 a 26, in cui i prodotti di espressione sono molecole proteiche.

41. Metodi come da rivendicazioni da 24 a 26, in cui la perdita del gene *GTBP* di tipo selvatico è determinata mediante immunoblotting, ad esempio Western blotting.

42. Metodi come da rivendicazioni da 24 a 26, in cui l'alterazione di *GTBP* di tipo selvatico è determinata mediante tecniche immunoenzimatiche ed immunocitochimiche.

43. Metodi come da rivendicazioni da 24 a 26, in cui l'alterazione del gene *GTBP* di tipo selvatico è determinata mediante interazioni di legame tra

detta proteina GTBP ed una seconda proteina cellulare.

44. Metodo come da rivendicazione 43, in cui la seconda proteina cellulare è hMSH2.

45. Procedimento per generare animali transgenici che portano alleli mutanti di *GTBP*.

46. Composizione farmaceutica utile nel trattamento di malattie dipendenti da GTBP che comprende una quantità terapeuticamente efficace di GTBP in un veicolo farmaceuticamente accettabile.

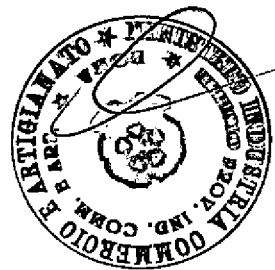
47. Metodo per fornire la funzione del gene *GTBP* di tipo selvatico a una cellula che ha GTBP alterato, detta funzione genica essendo persa a causa di una mutazione del gene *GTBP*, che comprende l'introduzione di tutto il gene *GTBP* oppure di parte di esso in una cellula che ha perso tale funzione genica, in modo tale che detto gene *GTBP*, intero o parziale, venga espresso nella cellula e codifichi per l'intera proteina GTBP oppure per parte di essa, rendendola capace di complementare il difetto genetico alla base della malattia neoplastica.

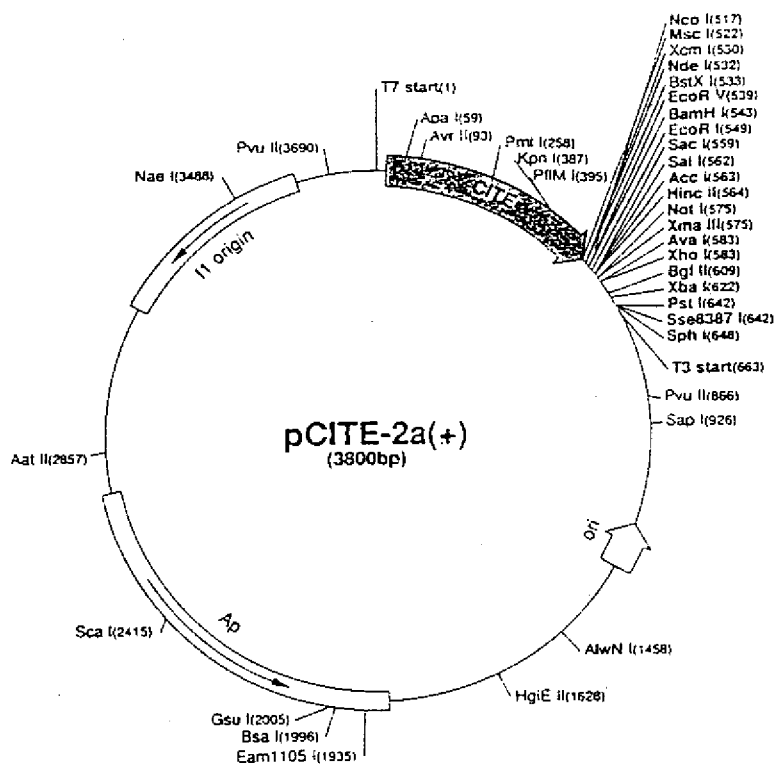
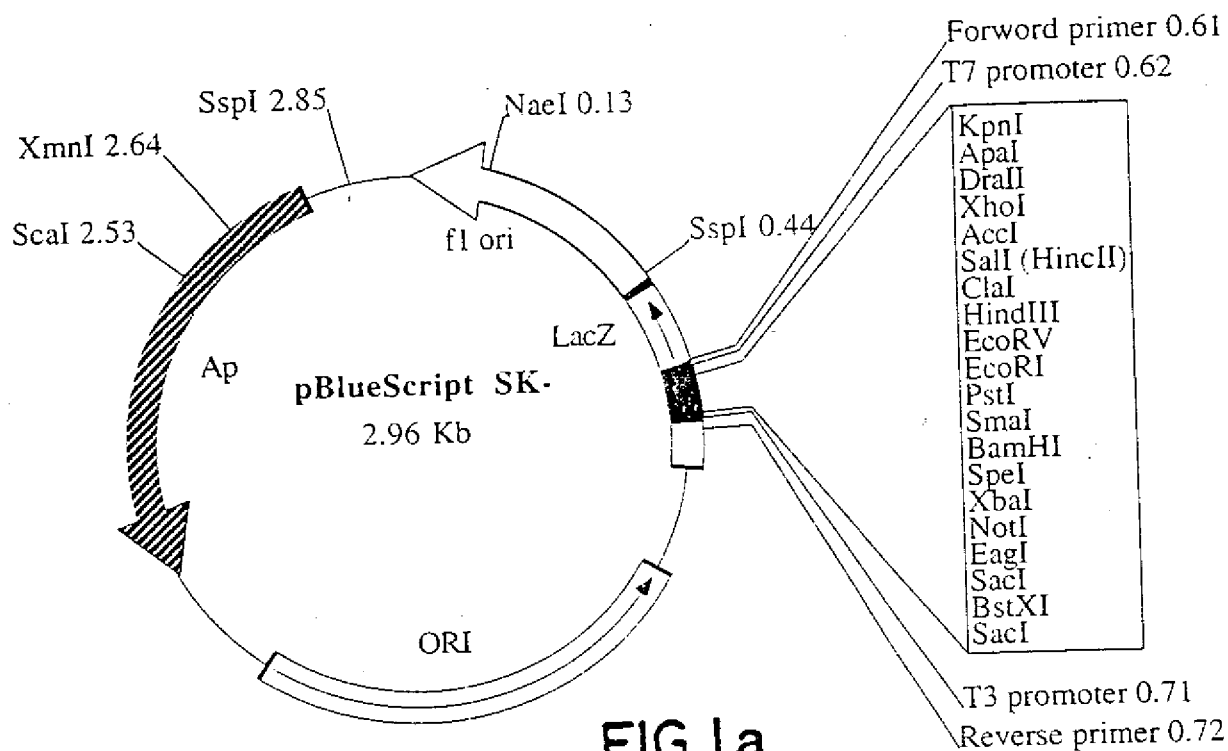
48. Metodo per fornire la funzione del gene *GTBP* di tipo selvatico ad una cella che ha GTBP

alterato, detta funzione genica essendo persa a causa di una mutazione nel gene GTBP, che comprende l'introduzione in una cellula di una molecola che mima l'effetto di GTBP da sola oppure complessata con altre molecole.

p.p. ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE
P. ANGELETTI S.p.A.

Gilberto Ronchi
(Doc. Albo n. 83)





Giulio Tonon
 (scr. Albo n. 20)

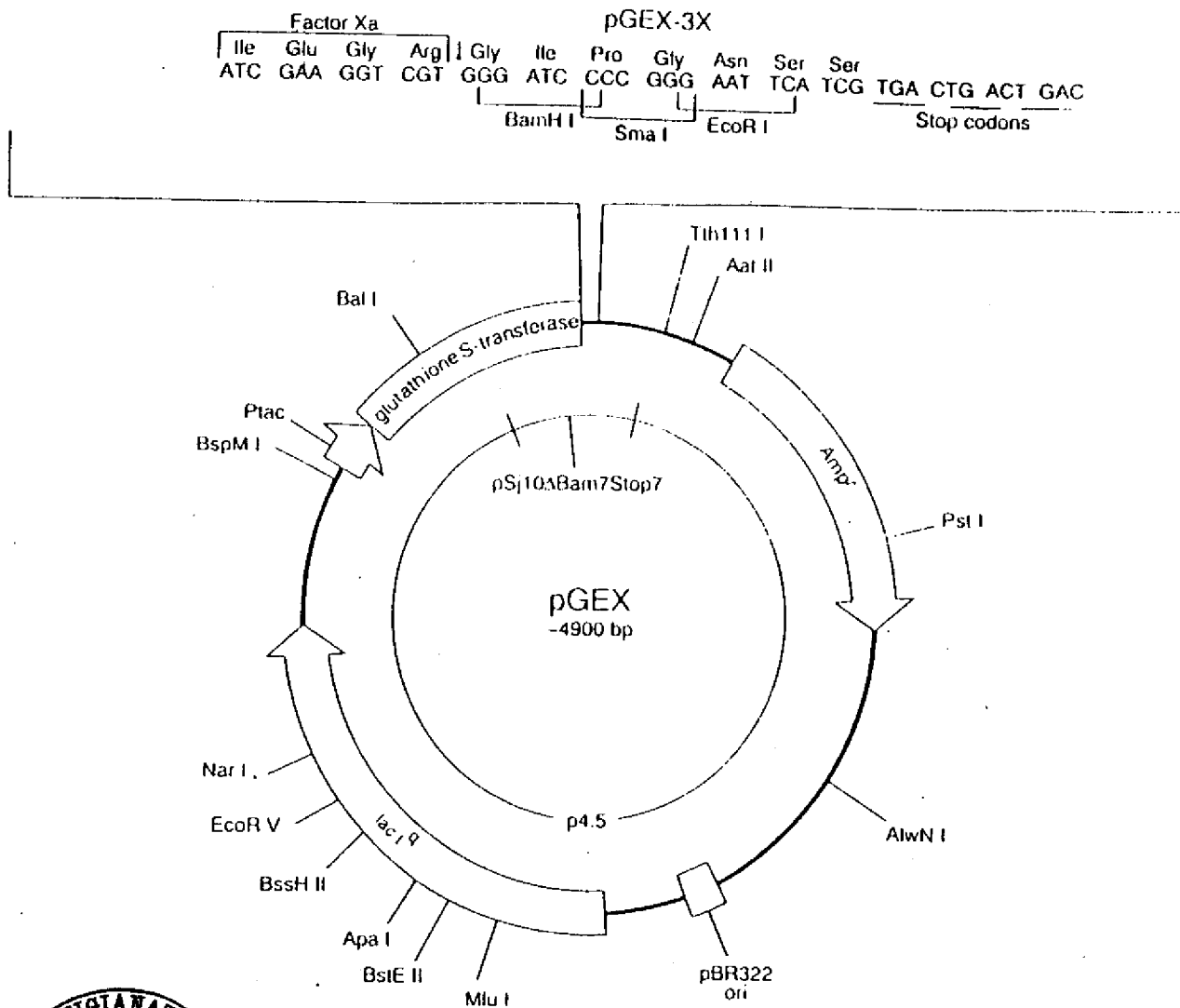


FIG 2



Gilberto Tonon
(Iscr. Albo n. 83)

RMCS A 000434

GTBP	L	V	T	G	P	N	M	G	G	K	S	T	L	M	R	Q	A	G	L	L	A	V	M	A	Q	M	G	C	Y	V	P	A	E	V	C	R	L	T	F	I	D	R	V	F	T	R	L	G	A	S	D	R	I	M	S	G	E	S	T	F	F	V	E	L	S	E	T	A	S	I	1132
bMSH2	I	I	T	G	P	N	M	G	G	K	S	T	Y	I	R	Q	T	O	V	I	V	L	M	A	Q	I	G	C	F	V	P	C	E	S	A	E	V	S	I	V	D	C	I	L	A	R	V	G	A	G	D	S	Q	L	K	G	V	S	T	F	M	A	E	M	L	E	T	A	S	I	735
MSH2	I	I	T	G	P	N	M	G	G	K	S	T	Y	I	R	Q	V	G	V	I	S	L	M	A	Q	I	G	C	F	V	P	C	E	R	A	E	I	A	I	V	D	A	I	L	C	R	V	G	A	G	D	S	Q	L	K	G	V	S	T	F	M	V	E	I	L	E	T	A	S	I	754
MUTS	I	I	T	G	P	N	M	G	G	K	S	T	Y	M	R	Q	T	A	L	I	A	L	M	A	Y	I	G	S	Y	V	P	A	Q	K	V	E	I	G	P	I	D	R	I	F	T	R	V	G	A	A	D	D	L	A	S	G	R	S	T	F	M	V	E	M	T	E	T	A	N	I	680

GTRP	L	M	H	A	T	A	H	S	L	V	L	V	D	E	L	G	R	G	T	A	T	F	D	G	T	A	I	A	N	A	V	V	K	E	L	A	E	T	I	K	C	R	T	L	F	S	T	H	Y	H	S	L	V	E	D	Y	S	Q	N	V	A	V	R	L	G	H	M	A	C	M	1202
bMSH2	L	R	S	A	T	K	D	S	L	I	I	I	D	E	L	G	R	G	T	S	T	Y	D	G	F	G	L	A	W	A	I	S	E	Y	I	A	T	K	I	G	A	F	C	M	F	A	T	H	F	H	E	L	T	A	L	A	N	Q	I	P	T	V	N	N	L	H	V	T	A	L	805
MSH2	L	K	N	A	S	K	N	S	L	I	I	V	D	E	L	G	R	G	T	S	T	Y	D	G	F	G	L	A	W	A	I	A	E	H	I	A	S	K	I	G	C	F	A	L	F	A	T	H	F	H	E	L	T	E	L	S	E	K	L	P	N	V	K	N	M	H	V	V	A	H	824
MUTS	L	H	N	A	T	E	Y	S	L	V	L	M	D	E	I	G	R	G	T	S	T	Y	D	G	L	S	L	A	W	A	C	A	E	N	L	A	N	K	I	K	A	L	T	L	F	A	T	H	Y	F	E	L	T	Q	L	P	E	K	M	E	G	V	A	N	V	H	L	D	A	L	750

FIG 3

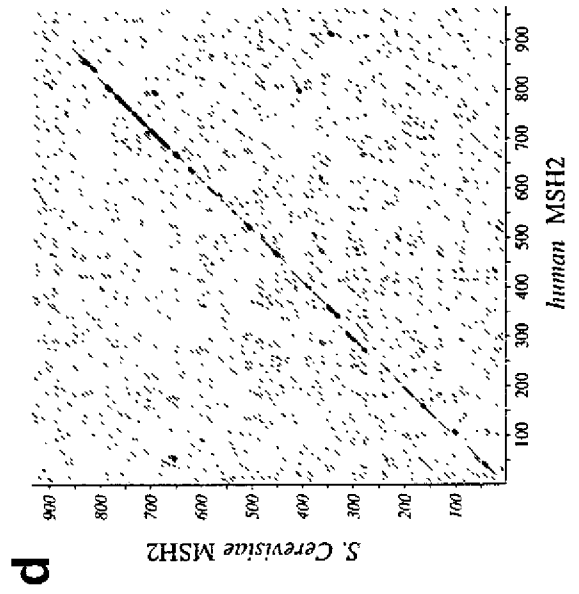
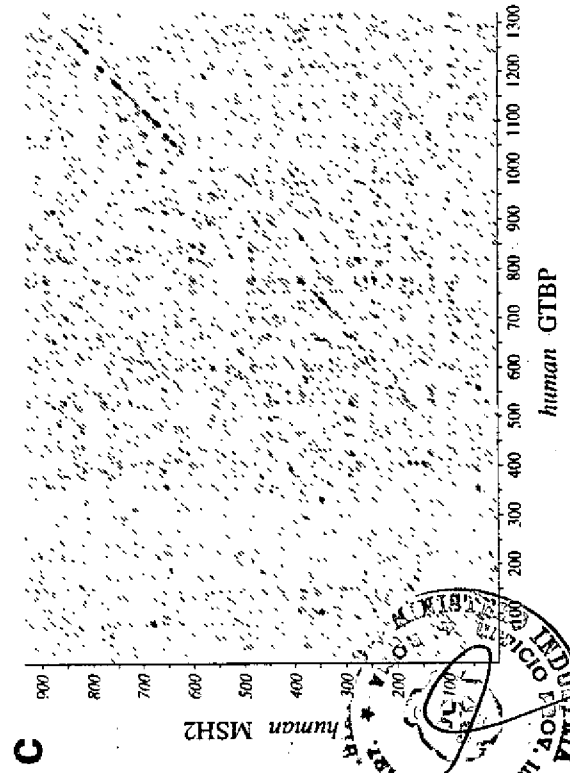
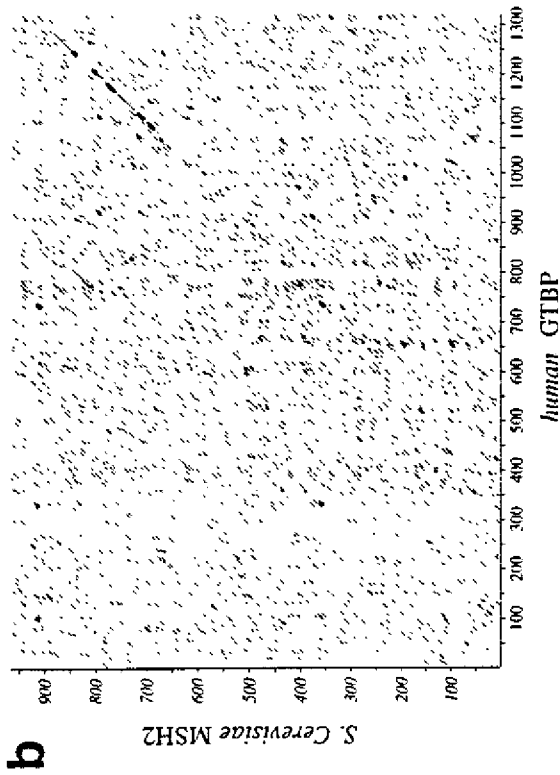
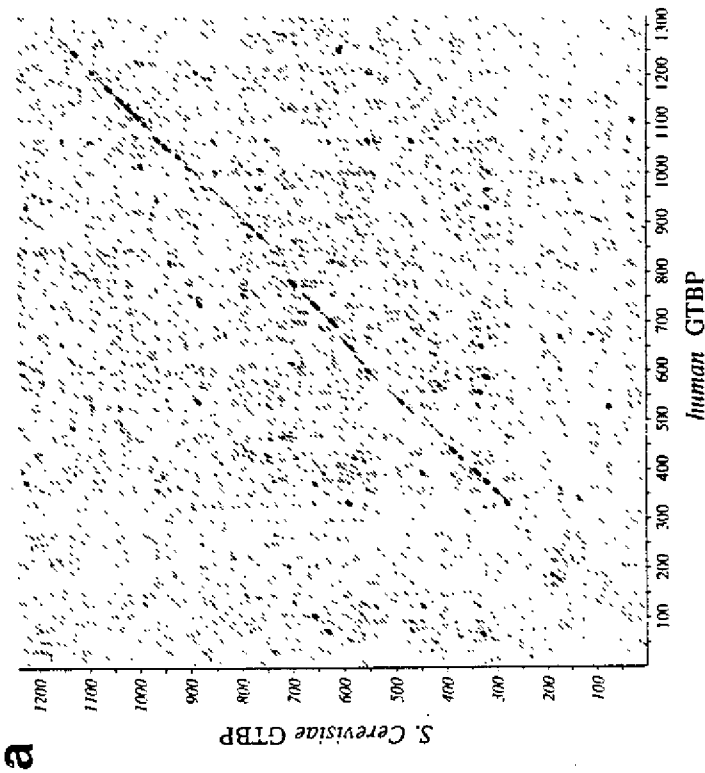


Gilberto Tomasi
(secr. Albo n. 43)

RM95 A 000434

RM95 A 000/34

FIG 4 RM95 A 000434



Alberto Tonon
(Iscr. Albo n. 93)

RM95 A 000434

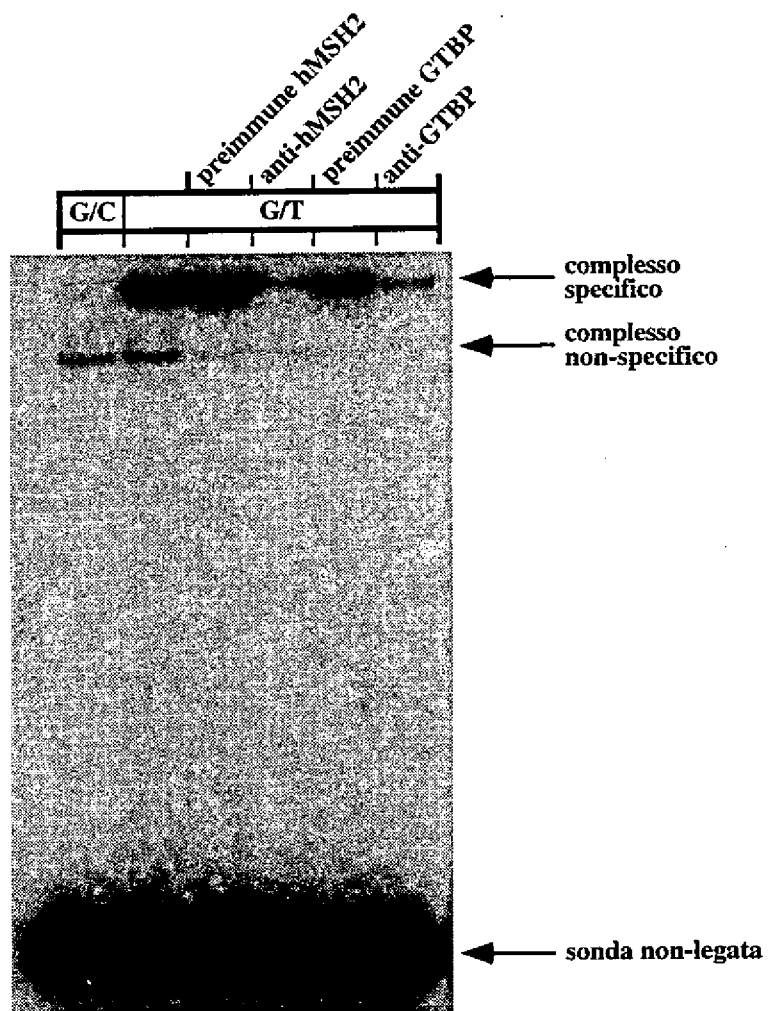


FIG 5



Alberto Tonon
Alberto Tonon
(Iscri. Albo n. 83)

Gilberto Tonon
(lec. Albo n. 83)

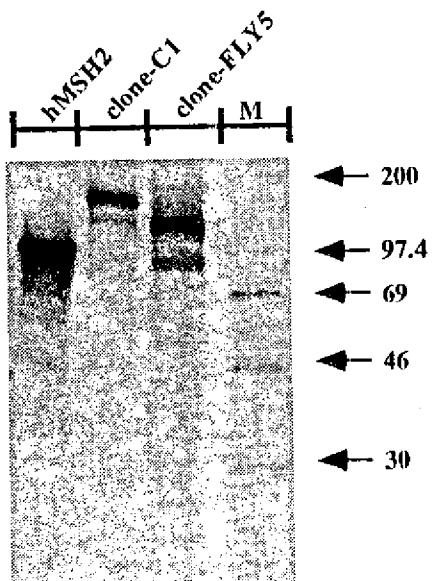


FIG 6a

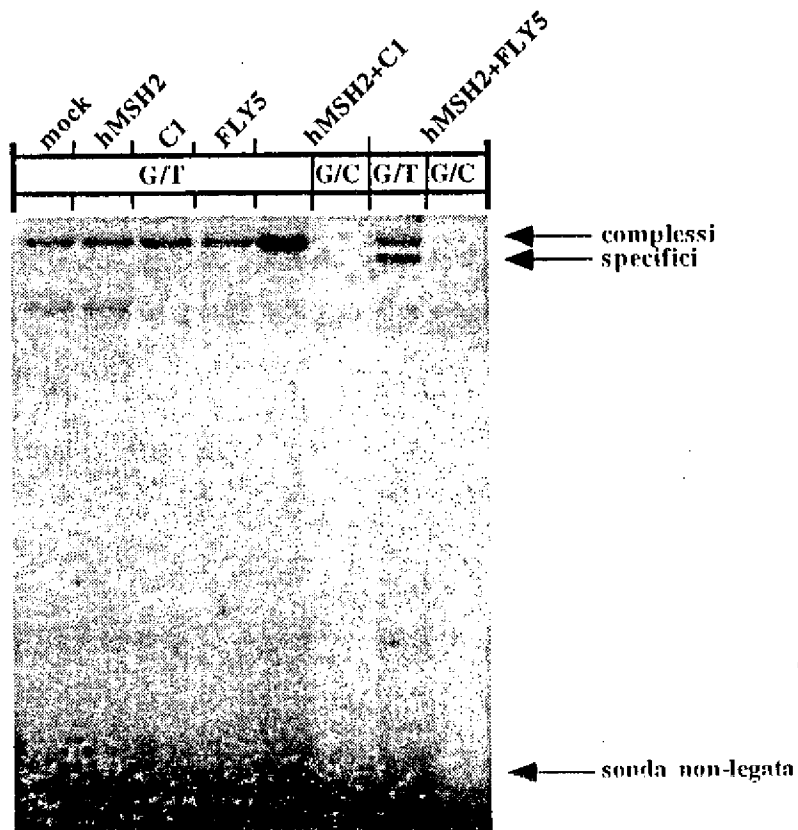


FIG 6b

RM95 A 000434

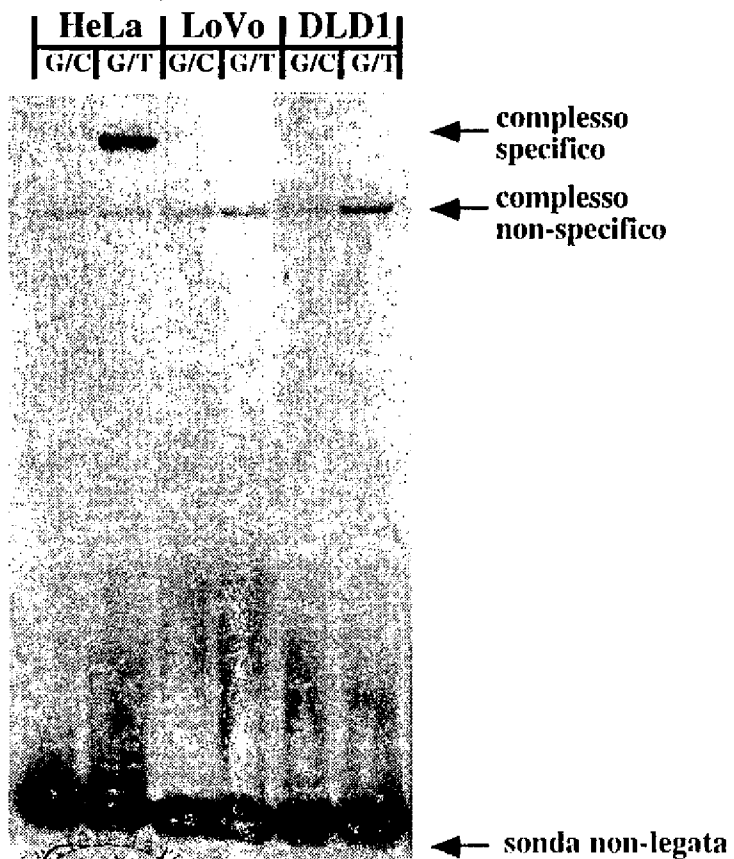


FIG 7 a

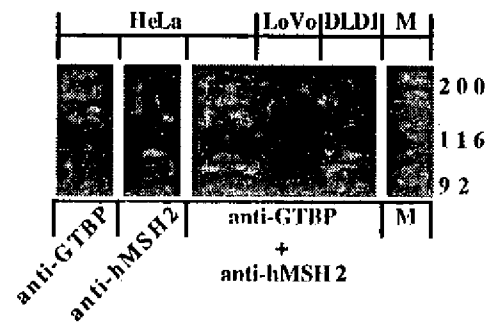
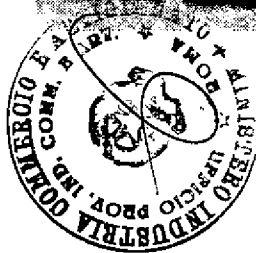


FIG 7 b



Angioletti
 Gilberto Tonon
 (Iscri. Albo n. 83)

RM95 A 000434