



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 268 763**

51 Int. Cl.:
C12N 15/10 (2006.01)
C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **98904574 .5**
86 Fecha de presentación : **14.01.1998**
87 Número de publicación de la solicitud: **0971946**
87 Fecha de publicación de la solicitud: **19.01.2000**

54 Título: **Selección de proteínas usando fusiones de ARN-proteína.**

30 Prioridad: **21.01.1997 US 35963 P**
06.11.1997 US 64491 P

45 Fecha de publicación de la mención BOPI:
16.03.2007

45 Fecha de la publicación del folleto de la patente:
16.03.2007

73 Titular/es:
THE GENERAL HOSPITAL CORPORATION
55 Fruit Street
Boston, Massachusetts 02114, US

72 Inventor/es: **Szostak, Jack, W.;**
Roberts, Richard, W. y
Liu, Rihe

74 Agente: **Curell Suñol, Marcelino**

ES 2 268 763 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Selección de proteínas usando fusiones de ARN-proteína.

5 **Antecedentes de la invención**

La presente invención se refiere a métodos de selección de proteínas.

La invención se llevó a cabo con el apoyo gubernamental bajo la concesión F32 GM17776-01 y F32 GM17776-10 02. El Gobierno tiene determinados derechos en la invención.

Actualmente existen métodos para el aislamiento de moléculas de ARN y ADN basados en sus funciones. Por ejemplo, los experimentos de Ellington y Szostak (Nature **346**:818 (1990); y Nature **355**:850 (1992)), y de Tuerk y Gold (Science **249**:505 (1990); y J. Mol. Biol. **222**:739 (1991)), han demostrado que se pueden aislar moléculas de ácidos nucleicos muy raras (es decir, menos de 1 en 10^{13}) con propiedades deseadas a partir de conjuntos complejos de moléculas mediante rondas repetidas de selección y amplificación. Estos métodos ofrecen ventajas con respecto a las selecciones genéticas tradicionales en las que (i) se pueden identificar sistemáticamente conjuntos candidatos muy grandes ($> 10^{15}$), (ii) la viabilidad del hospedante y las condiciones *in vivo* no son preocupantes, y (iii) las selecciones se pueden llevar a cabo incluso si no existe una identificación sistemática genética *in vivo*. Además, el documento EP0962527A describe una molécula que homóloga genotipo y fenotipo y sus utilizaciones. El poder de la selección *in vitro* se ha demostrado definiendo nuevas secuencias de ARN y ADN con funciones de unión a proteínas muy específicas (véase, por ejemplo, Tuerk y Gold, Science **249**:505 (1990); Irvine *et al.*, J. Mol. Biol. **222**:739 (1991); Oliphant *et al.*, Mol. Cell Biol. **9**:2944 (1989); Blackwell *et al.*, Science **250**:1104 (1990); Pollock y Treisman, Nuc. Acids Res. **18**:6197 (1990); Thiesen y Bach, Nuc. Acids Res. **18**:3203 (1990); Bartel *et al.*, Cell **57**:529 (1991); Stormo y Yoshioka, Proc. Natl. Acad. Sci. USA **88**:5699 (1991); y Bock *et al.*, Nature **355**:564 (1992)), funciones de unión a moléculas pequeñas (Ellington y Szostak, Nature **346**:818 (1990); Ellington y Szostak, Nature **355**:850 (1992)), y funciones catalíticas (Green *et al.*, Nature **347**:406 (1990); Robertson y Joyce, Nature **344**:467 (1990); Beaudry y Joyce, Science **257**:635 (1992); Bartel y Szostak, Science **261**:1411 (1993); Lorsch y Szostak, Nature **371**:31-36 (1994); Cuenoud y Szostak, Nature **375**:611-614 (1995); Chapman y Szostak, Chemistry and Biology **2**:325-333 (1995); y Lohse y Szostak, Nature **381**:442-444 (1996)). No se ha demostrado un esquema similar para la selección y amplificación de proteínas.

Sumario de la invención

El objetivo de la presente invención es permitir que se apliquen los principios de selección *in vitro* y de evolución *in vitro* a las proteínas. La invención facilita el aislamiento de proteínas con propiedades deseadas a partir de grandes conjuntos de secuencias de aminoácidos parcial o completamente al azar. Además, la invención resuelve el problema de recuperar y amplificar la información de la secuencia proteínica uniendo covalentemente la secuencia que codifica ARNm a la molécula de proteína.

En general, el método de la invención consiste en un protocolo de transcripción/traducción *in vitro* o *in situ* que genera una proteína enlazada covalentemente al extremo 3' de su propio ARNm, es decir, una fusión de ARN-proteína. Esto se logra mediante la síntesis y la traducción *in vitro* o *in situ* de una molécula de ARNm con un aceptor de péptidos unido a su extremo 3'. Un aceptor de péptidos preferido es puromicina, un análogo nucleosídico que se añade al término C de una cadena peptídica en crecimiento, y que termina la producción. En un diseño preferido, se incluye una secuencia de ADN entre el extremo del mensaje y el aceptor de péptidos que se diseña para que el ribosoma se detenga en el extremo del marco de lectura abierto, proporcionando tiempo adicional para que el aceptor peptídico (por ejemplo, puromicina) acepte la cadena peptídica naciente antes de la hidrólisis del enlace peptidil-ARNt.

Si se desea, la fusión de ARN-proteína resultante permite rondas repetidas de selección y amplificación, debido a que la información de secuencia proteínica se puede recuperar mediante transcripción inversa y amplificación (por ejemplo, mediante amplificación por PCR, así como también mediante cualquier otra técnica de amplificación, incluyendo técnicas de amplificación a base de ARN, tales como 3SR o TSA). Después, el ácido nucleico amplificado se puede transcribir, modificar y traducir *in vitro* o *in situ* para generar fusiones de ARNm-proteína para la siguiente ronda de selección. La capacidad para llevar a cabo múltiples rondas de selección y amplificación permite el enriquecimiento y aislamiento de moléculas muy raras, por ejemplo, una molécula deseada de un conjunto de 10^{15} miembros. Esto, a su vez, permite el aislamiento de nuevas proteínas, o proteínas mejoradas, que reconocen específicamente, de forma virtual, cualquier diana, o que catalizan reacciones químicas deseadas.

En consecuencia, en un primer aspecto, la invención se refiere a un método para la selección de una proteína deseada, que implica las etapas siguientes: (a) proporcionar una población de moléculas de ARN candidatas, cada una de las cuales incluye una secuencia de iniciación de la traducción y un codón de partida enlazado funcionalmente a una secuencia codificante proteínica candidata, y cada una de las cuales está enlazada covalentemente a un aceptor peptídico en el extremo 3' de una secuencia codificante de proteína candidata; (b) traducir *in vitro* o *in situ* las secuencias codificantes de proteínas candidatas para producir una población de fusiones de ARN-proteína candidatas; y (c) seleccionar una fusión de ARN-proteína deseada, seleccionando de este modo la proteína deseada.

ES 2 268 763 T3

En un aspecto relacionado, la invención se refiere a un método para la selección de una molécula de ADN que codifica una proteína deseada, que implica las etapas siguientes: (a) proporcionar una población de moléculas de ARN candidatas, cada una de las cuales incluye una secuencia de iniciación de la traducción y un codón de partida enlazado funcionalmente a una secuencia codificante de proteína candidata, y cada una de las cuales está enlazada covalentemente a un aceptor peptídico en el extremo 3' de la secuencia codificante de proteína candidata; (b) traducir *in vitro* o *in situ* las secuencias codificantes de proteínas candidatas para producir una población de fusiones de ARN-proteína candidatas; (c) seleccionar una fusión de ARN-proteína deseada; y (d) generar, a partir de la porción de ARN de la fusión, una molécula de ADN que codifica la molécula deseada.

En otro aspecto relacionado, la invención se refiere a un método para la selección de una proteína que tiene una función alterada con relación a una proteína de referencia, que implica las etapas siguientes: (a) producir una población de moléculas de ARN candidatas a partir de una población de moldes de ADN, teniendo cada uno de los moldes de ADN candidatos una secuencia codificante de proteína candidata que difiere de la secuencia codificante de proteína de referencia, comprendiendo cada una de las moléculas de ARN una secuencia de iniciación de la traducción y un codón de partida enlazado funcionalmente a la secuencia codificante de proteína candidata, y estando cada una enlazada covalentemente a un aceptor peptídico en el extremo 3'; (b) traducir *in vitro* o *in situ* las secuencias codificantes de proteínas candidatas para producir una población de fusiones de ARN-proteína candidatas; y (c) seleccionar una fusión de ARN-proteína que tiene una función alterada, seleccionando de ese modo la proteína que tiene la función alterada.

En aún otro aspecto relacionado, la invención se refiere a un método para la selección de una molécula de ADN que codifica una proteína que tiene una función alterada con relación a una proteína de referencia, que implica las etapas siguientes: (a) producir una población de moléculas de ARN candidatas a partir de una población de moldes de ADN candidatos, teniendo cada uno de los moldes de ADN candidatos una secuencia codificante de proteína candidata que difiere de la secuencia codificante de proteína de referencia, comprendiendo cada una de las moléculas de ARN una secuencia de iniciación de la traducción y un codón de partida enlazado funcionalmente a la secuencia codificante de proteína candidata, y estando cada una enlazada covalentemente a un aceptor peptídico en el extremo 3'; (b) traducir *in vitro* o *in situ* las secuencias codificantes de proteínas candidatas para producir una población de fusiones de ARN-proteína; (c) seleccionar una fusión de ARN-proteína que tiene una función alterada; y (d) generar, a partir de la porción de ARN de la fusión, una molécula de ADN que codifica la proteína que tiene la función alterada.

En aún otro aspecto relacionado, la invención se refiere a un método para la selección de un ARN deseado, que implica las etapas siguientes: (a) proporcionar una población de moléculas de ARN candidatas, cada una de las cuales incluye una secuencia de iniciación de la traducción y un codón de partida enlazado funcionalmente a una secuencia codificante de proteína candidata, y cada una de las cuales está enlazada covalentemente a un aceptor peptídico en el extremo 3' de la secuencia codificante de proteína candidata; (b) traducir *in vitro* o *in situ* las secuencias codificantes de proteínas candidatas para producir una población de fusiones de ARN-proteínas candidatas; y (c) seleccionar una fusión de ARN-proteína deseada, seleccionando de ese modo el ARN deseado.

En realizaciones preferidas de los métodos anteriores, el aceptor peptídico es puromicina; cada una de las moléculas de ARN candidatas incluye además una secuencia de pausa, o incluye además una secuencia de ADN o de un análogo de ADN enlazada covalentemente al extremo 3' del ARN; la población de moléculas de ARN candidatas incluye al menos 10^9 , preferentemente al menos 10^{10} , más preferentemente al menos 10^{11} , 10^{12} , o 10^{13} , y más preferentemente al menos 10^{14} , moléculas de ARN diferentes; la reacción de traducción *in vitro* se lleva a cabo en un lisado preparado a partir de una célula eucariota o una porción de la misma (y, por ejemplo, se lleva a cabo en un lisado reticulocítico o lisado de germen de trigo); la reacción de traducción *in vitro* se lleva a cabo en un extracto preparado a partir de una célula procariota (por ejemplo, *E. coli*) o una porción de la misma; la etapa de selección implica unir la proteína deseada a una pareja de unión inmovilizada; la etapa de selección implica ensayar una actividad funcional de la proteína deseada; la molécula de ADN se amplifica; el método implica además repetir las etapas de los métodos de selección anteriores; el método implica además transcribir una molécula de ARN a partir de la molécula de ADN, y repetir las etapas (a) a (d); tras la etapa de traducción *in vitro*, el método implica además una etapa de incubación llevada a cabo en presencia de 50-100 mM de Mg^{2+} ; y la fusión de ARN-proteína incluye además una secuencia de ácidos nucleicos o una secuencia de análogos de ácidos nucleicos situada próxima al aceptor peptídico, lo que aumenta la flexibilidad.

En otros aspectos relacionados, la invención se refiere a una fusión de ARN-proteína seleccionada mediante cualquiera de los métodos de la invención; un ácido ribonucleico enlazado covalentemente a través de un enlace amídico a una secuencia de aminoácidos, codificándose la secuencia de aminoácidos mediante el ácido ribonucleico; y un ácido ribonucleico que incluye una secuencia de iniciación de la traducción y un codón de partida enlazado funcionalmente a una secuencia codificante de proteína candidata, estando el ácido ribonucleico enlazado covalentemente a un aceptor peptídico (por ejemplo, puromicina) en el extremo 3' de la secuencia codificante de proteína candidata.

En un segundo aspecto, la invención se refiere a un método para la selección de una proteína deseada o de un ARN deseado a través del enriquecimiento de un conjunto de secuencias. Este método implica las etapas siguientes: (a) proporcionar una población de moléculas de ARN candidatas, cada una de las cuales incluye una secuencia de iniciación de la traducción y un codón de partida enlazado funcionalmente a una secuencia codificante de proteína candidata, y cada una de las cuales está enlazada covalentemente a un aceptor peptídico en el extremo 3' de la secuencia codificante de proteína candidata; (b) traducir *in vitro* o *in situ* las secuencias codificantes de proteínas candidatas para

producir una población de fusiones de ARN-proteína candidatas; (c) poner en contacto la población de fusiones de ARN-proteína con una pareja de unión específica para la porción de ARN o para la porción proteínica de la fusión de ARN-proteína, en condiciones que separen sustancialmente los complejos de fusión de ARN-proteína-pareja de unión de los miembros no unidos de la población; (d) liberar las fusiones de ARN-proteína unidas de los complejos; y (e) poner en contacto la población de fusiones de ARN-proteína de la etapa (d) con una pareja de unión específica para la porción proteínica de la fusión de ARN-proteína deseada, en condiciones que separen sustancialmente el complejo de pareja de unión-fusión de ARN-proteína de miembros no unidos de dicha población, seleccionando de ese modo la proteína deseada y el ARN deseado.

En realizaciones preferidas, el método implica además repetir las etapas (a) a (e). Además, para estas etapas repetidas, se pueden usar las mismas parejas de unión o diferentes, en cualquier orden, para el enriquecimiento selectivo de la fusión de ARN-proteína deseada. En otra realización preferida, la etapa (d) implica el uso de una pareja de unión (por ejemplo, un anticuerpo monoclonal) específico para la porción proteínica de la fusión deseada. Esta etapa se lleva a cabo preferentemente tras la transcripción inversa de la porción de ARN de la fusión para generar un ADN que codifica la proteína deseada. Si se desea, este ADN se puede aislar y/o amplificar mediante PCR. Esta técnica de enriquecimiento se puede usar para seleccionar una proteína deseada, o se puede usar para seleccionar una proteína que tiene una función alterada con relación a una proteína de referencia.

En otras realizaciones preferidas de los métodos de enriquecimiento, el aceptor peptídico es puomicina; cada una de las moléculas de ARN candidatas incluye además una secuencia de pausa, o incluye además una secuencia de ADN o de análogo de ADN enlazada covalentemente al extremo 3' del ARN; la población de moléculas de ARN candidatas incluye al menos 10^9 , preferentemente al menos 10^{10} , más preferentemente al menos 10^{11} , 10^{12} ó 10^{13} , y más preferentemente al menos 10^{14} , moléculas diferentes de ARN; la reacción de producción *in vitro* se lleva a cabo en un lisado preparado a partir de una célula procariota o una porción de la misma (y, por ejemplo, se lleva a cabo en un lisado reticulocítico o lisado de germen de trigo); la reacción de traducción *in vitro* se lleva a cabo en un extracto preparado a partir de una célula procariota o una porción de la misma (por ejemplo, *E. coli*); se amplifica la molécula de ADN; al menos una de las parejas de unión se inmoviliza sobre un soporte sólido; tras la etapa de traducción *in vitro*, el método implica además una etapa de incubación llevada a cabo en presencia de 50-100 mM de Mg^{2+} ; y la fusión de ARN-proteína incluye además una secuencia de ácidos nucleicos o análogo de ácido nucleico situada próxima al aceptor peptídico, lo que aumenta la flexibilidad.

En un aspecto relacionado, la invención se refiere a kits para llevar a cabo cualquiera de los métodos de selección descritos en la presente memoria.

En un tercer aspecto y último, la invención se refiere a un microchip que incluye un conjunto de ácidos nucleicos monocatenarios inmovilizados, hibridándose los ácidos nucleicos con las fusiones de ARN-proteína. Preferentemente, el componente proteínico de la fusión de ARN-proteína está codificado por el ARN.

Como se utiliza en la presente memoria, el término “población” se refiere a más de una molécula (por ejemplo, más de una molécula de ARN, ADN, o molécula de fusión de ARN-proteína). Debido a que los métodos de la invención facilitan selecciones que comienzan, si se desea, con grandes números de moléculas candidatas, una “población” según la invención significa preferentemente más de 10^9 moléculas, más preferentemente más de 10^{11} , 10^{12} ó 10^{13} moléculas, y más preferentemente más de 10^{13} moléculas.

Mediante el término “seleccionar” se hace referencia sustancialmente a una molécula de otras moléculas en una población. Como se utiliza en la presente memoria, una etapa de “selección” proporciona al menos un enriquecimiento de 2 veces, preferentemente un enriquecimiento de 30 veces, más preferentemente un enriquecimiento de 100 veces, y más preferentemente un enriquecimiento de 1000 veces, de una molécula deseada con relación a moléculas no deseadas en una población tras la etapa de selección. Como se indica en la presente memoria, una etapa de selección se puede repetir cualquier número de veces, y en un enfoque dado se pueden combinar diferentes tipos de etapas de selección.

Mediante el término “proteína” se hace referencia a cualquiera de dos o más aminoácidos naturales o modificados, unidos mediante uno o más enlaces peptídicos. “Proteína” y “péptido” se usan en la presente memoria de forma intercambiable.

Mediante el término “ARN” se hace referencia a una secuencia de dos o más ribonucleótidos de origen natural o modificados, enlazados covalentemente. Un ejemplo de un ARN modificado, incluido en este término, es ARN fosforotioato.

Mediante la expresión “secuencia de iniciación de la traducción” se hace referencia a cualquier secuencia que sea capaz de proporcionar un sitio de entrada de ribosoma funcional. En sistemas bacterianos, esta región se denomina algunas veces como una secuencia de Shine-Dalgarno.

Mediante la expresión “codón de partida” se hace referencia a tres bases que señalan el comienzo de una secuencia codificante de proteínas. Generalmente, estas bases son AUG (o ATG); sin embargo, se puede sustituir por cualquier otro triplete de bases capaz de ser utilizado de esta manera.

Mediante la expresión “enlazado covalentemente” a un aceptor peptídico se hace referencia a que el aceptor peptídico está unido a una “secuencia codificante de proteínas”, ya sea directamente a través de un enlace covalente, o indirectamente a través de otra secuencia enlazada covalentemente (por ejemplo, ADN que corresponde a un sitio de pausa).

5

Mediante la expresión “aceptor peptídico” se hace referencia a cualquier molécula capaz de ser añadida al término C de una cadena proteínica creciente mediante la actividad catalítica de la función ribosómica de peptidiltransferasa. Típicamente, tales moléculas contienen (i) un resto nucleotídico o similar a nucleótido (por ejemplo, adenosina o un análogo de adenosina (la dimetilación en la posición amino N-6 es aceptable)), (ii) un resto de aminoácido o similar a aminoácido (por ejemplo, cualquiera de los 20 aminoácidos D o L, o cualquier análogo de aminoácido de los mismos (por ejemplo, O-metil tirosina o cualquiera de los análogos descritos por Ellman *et al.*, Meth. Enzymol. **202**:301, 1991), y (iii) un enlace entre los dos (por ejemplo, un enlace de éster, de amida o de cetona en la posición 3' o, menos preferentemente, en la posición 2'); preferentemente, este enlace no perturba significativamente el plegamiento del anillo de la conformación del ribonucleótido natural. Los aceptores peptídicos también pueden poseer un nucleófilo, el cual puede ser, sin limitación, un grupo amino, un grupo hidroxilo, o un grupo sulfidrilo. Además, los aceptores peptídicos pueden estar compuestos de miméticos de nucleótidos, miméticos de aminoácidos, o miméticos de la estructura combinada de nucleótido-aminoácido.

Mediante un aceptor peptídico colocado “en el extremo 3'” de una secuencia codificante de proteínas, se hace referencia a la molécula del aceptor peptídico está situada tras el codón final de esa secuencia codificante de proteínas. Este término incluye, sin limitación, una molécula de aceptor peptídico que está situada precisamente en el extremo 3' de la secuencia codificante de proteínas, así como una que está separada del codón final mediante una secuencia codificante o no codificante interventora (por ejemplo, una secuencia que corresponde a un sitio de pausa). Este término también incluye constructos en los que las secuencias codificantes o no codificantes siguen (esto es, están en 3' con respecto a) la molécula del aceptor peptídico. Además, este término comprende, sin limitación, una molécula de aceptor peptídico que está enlazada covalentemente (ya sea directa o indirectamente a través de una secuencia de ácidos nucleicos interventora) a la secuencia codificante de proteínas, así como una que está unida a la secuencia codificante de proteínas mediante algún medio no covalente, por ejemplo, a través de la hibridación usando una segunda secuencia de ácidos nucleicos que se une a o está próxima al extremo 3' de la secuencia codificante de proteínas, y que ella misma está unida a una molécula del aceptor peptídico.

Mediante la expresión “función alterada”, se hace referencia a un cambio cualitativo o cuantitativo en la función de una molécula.

Mediante la expresión “secuencia de pausa”, se hace referencia a una secuencia de ácidos nucleicos que provoca que un ribosoma ralentice o detenga su velocidad de traducción.

Por la expresión “pareja de unión”, tal como se utiliza en la presente memoria, se hace referencia a cualquier molécula que tiene una afinidad específica, covalente o no covalente, por una porción de una fusión de ARN-proteína deseada. Los ejemplos de parejas de unión incluyen, sin limitación, miembros de pares antígeno/anticuerpo, pares proteína/inhibidor, pares receptor/ligando (por ejemplo, pares receptor de superficie celular/ligando, tales como pares receptor hormonal/hormona peptídica), pares enzima/sustrato (por ejemplo, pares quinasa/sustrato), pares lectina/carbohidrato, agregados proteínicos oligoméricos o heterooligoméricos, pares proteína de unión a ADN/sitio de unión a ADN, pares ARN/proteína, y dúplex de ácidos nucleicos, heterodúplex, o hebras ligadas, así como cualquier molécula que sea capaz de formar uno o más enlaces covalentes o no covalentes (por ejemplo, enlaces de disulfuro) con cualquier porción de una fusión de ARN-proteína. Las parejas de unión incluyen, sin limitación, cualquiera de los “motivos de selección” presentados en la Figura 2.

Mediante la expresión “soporte sólido” se hace referencia, sin limitación, a cualquier columna (o material de columna), perla, tubo de ensayo, placa de microtitulación, partícula sólida (por ejemplo, agarosa o sefarosa), microchip (por ejemplo, silicio, silicio-vidrio, o chip de oro), o membrana (por ejemplo, la membrana de un liposoma o vesícula), a los que se puede unir un complejo de afinidad, ya sea directa o indirectamente (por ejemplo, a través de otros intermedios de parejas de unión tales como otros anticuerpos o Proteína A), o en los que se puede embeber el complejo de afinidad (por ejemplo, a través de un receptor o canal).

55

La invención reivindicada actualmente proporciona un número de ventajas significativas. En primer lugar, es el primer ejemplo de este tipo de esquema para la selección y amplificación de proteínas. Esta técnica supera el punto muerto creado por la necesidad de recuperar secuencias nucleotídicas que corresponden a proteínas aisladas deseadas (puesto que sólo se pueden replicar los ácidos nucleicos). En particular, muchos métodos previos que permitían el aislamiento de proteínas a partir de conjuntos parcial o totalmente aleatorizados lo hacían así a través de una etapa *in vivo*. Los métodos de esta clase incluyen tecnología de anticuerpos monoclonales (Milstein, Sci. Amer. **243**:66 (1980); y Schultz *et al.*, J. Chem. Engng. News **68**:26 (1990)), la presentación de fagos (Smith, Science **228**:1315 (1985); Parmley y Smith, Gene **73**:305 (1988); y McCafferty *et al.*, Nature **348**:552 (1990)), las fusiones de péptido-represor de lac (Cull *et al.*, Proc. Natl. Acad. Sci. USA **89**:1865 (1992)), y selecciones genéticas clásicas. A diferencia de la técnica presente, cada uno de estos métodos descansa en un enlace topológico entre la proteína y el ácido nucleico de forma que la información de la proteína es retenida y se puede recuperar en una forma de ácido nucleico legible.

65

Además, la presente invención proporciona ventajas con respecto al método de traducción retrasado (Tuerk y Gold, Science **249**:505 (1990); Irvine *et al.*, J. Mol. Biol. **222**:739 (1991); Korman *et al.*, Proc. Natl. Acad. Sci. USA **79**:1844-1848 (1982); Mattheakis *et al.*, Proc. Natl. Acad. Sci. USA **91**:9022-9026 (1994); Mattheakis *et al.*, Meth. Enzymol. **267**:195 (1996); y Hanes y Pluckthun, Proc. Natl. Acad. Sci. USA **94**:4937 (1997)), una técnica en la que la selección se realiza por alguna propiedad de una cadena proteínica naciente que está aún complejada con el ribosoma y su ARNm. A diferencia de la técnica de traducción retrasada, el presente método no descansa en mantener la integridad de un complejo ternario de ARNm:ribosoma:cadena naciente, un complejo que es muy frágil y por lo tanto muy limitante con respecto a los tipos de selecciones que son técnicamente factibles.

El presente método también proporciona ventajas con respecto al enfoque de síntesis ramificada propuesto por Brenner y Lerner (Proc. Natl. Acad. Sci. USA **89**:5381-5383 (1992)), en el que se generan fusiones de ADN-péptido, y la información genética se recupera teóricamente después de una ronda de selección. A diferencia del enfoque de síntesis ramificada, el presente método no requiere la regeneración de un péptido a partir de la porción de ADN de una fusión (que, en el enfoque de síntesis ramificada, generalmente se logra mediante rondas individuales de síntesis química). En consecuencia, el presente método permite rondas repetidas de selección usando poblaciones de moléculas candidatas. Además, a diferencia de la técnica de síntesis ramificada, que generalmente está limitada a la selección de secuencias bastante cortas, el presente método es aplicable a la selección de moléculas de proteína de longitud considerable.

En aún otra ventaja, la presente técnica de selección y de evolución dirigida puede hacer uso de librerías muy grandes y complejas de secuencias candidatas. Por el contrario, los métodos de selección de proteínas existentes, que se basan en una etapa *in vivo*, están limitados típicamente a librerías relativamente pequeñas de una complejidad en cierto modo limitada. Esta ventaja es particularmente importante cuando se seleccionan secuencias de proteínas funcionales, considerando que, por ejemplo, existen 10^{13} secuencias posibles para un péptido de sólo 10 aminoácidos de longitud. En las técnicas genéticas clásicas, los enfoques de fusión del represor de lac, y los métodos de presentación de fagos, las complejidades máximas generalmente caen dentro de órdenes de magnitud por debajo de 10^{13} miembros. Un gran tamaño de librería también proporciona una ventaja para aplicaciones de evolución dirigida, en las que el espacio de secuencias se puede explorar hasta una mayor profundidad alrededor de una secuencia de partida dada.

La presente técnica también difiere de enfoques anteriores en cuanto a la etapa de selección que depende del contexto. En muchos otros esquemas de selección, en el contexto en el que, por ejemplo, está presente una proteína expresada puede influir profundamente en la naturaleza de la librería generada. Por ejemplo, una proteína expresada puede que no se exprese apropiadamente en un sistema particular, o puede que no se presente apropiadamente (por ejemplo, sobre la superficie de una partícula fágica). Como alternativa, la expresión de una proteína puede interferir realmente con una o más etapas críticas en el ciclo de selección, por ejemplo, viabilidad o infectividad del fago, o unión al represor de lac. Estos problemas pueden dar como resultado la pérdida de moléculas funcionales, o limitaciones en la naturaleza de los procedimientos de selección que se pueden aplicar.

Finalmente, el presente método es ventajoso debido a que proporciona un control sobre el repertorio de proteínas que se pueden ensayar. En ciertas técnicas (por ejemplo, selección de anticuerpos), existe muy poco o ningún control sobre la naturaleza del conjunto de partida. En aún otras técnicas (por ejemplo, fusiones de lac y presentación de fagos), el conjunto candidato se debe de expresar en el contexto de una proteína de fusión. Por el contrario, los constructos de fusión de ARN-proteína proporcionan un control sobre la naturaleza de los conjuntos candidatos disponibles para la identificación sistemática. Además, el tamaño del conjunto candidato tiene el potencial de ser tan elevado como los conjuntos de ARN o ADN ($\sim 10^{15}$ miembros), limitado sólo por el tamaño de la reacción de traducción *in vitro* realizada. Y la estructura del conjunto candidato depende completamente del diseño experimental; se pueden identificar sistemáticamente regiones al azar en aislamiento o en el contexto de una proteína de fusión deseada, y la mayoría, si no todas, de las posibles secuencias se pueden expresar en conjuntos candidatos de fusiones de ARN-proteína.

Otras características y ventajas de la invención se pondrán de manifiesto a partir de la siguiente descripción detallada, y a partir de las reivindicaciones.

Descripción detallada

En primer lugar, se describirán brevemente los dibujos.

Breve descripción de los dibujos

Las Figuras 1A-1C son representaciones esquemáticas de etapas implicadas en la producción de fusiones de ARN-proteína. La Figura 1A ilustra un constructo de ADN de muestra para la generación de una porción de ARN de una fusión. La Figura 1B ilustra la generación de un conjugado de ARN/puromicina. Y la Figura 1C ilustra la generación de una fusión de ARN-proteína.

La Figura 2 es una representación esquemática de un protocolo de selección generalizado según la invención.

La Figura 3 es una representación esquemática de un protocolo de síntesis para moldes de traducción mínimos que contienen 3'-puromicina. La etapa (A) muestra la adición de grupos protectores a los grupos funcionales reactivos en

la puromicina (5'-OH y NH₂); a medida que se modifican, estos grupos se protegen adecuadamente para uso en la síntesis de oligonucleótidos a base de fosforamidito. La puromicina protegida se une a vidrio de poro controlado de aminohexilo (CPG), a través del grupo 2'OH, usando el protocolo estándar para la unión de ADN a través de su 3'OH (Gait, *Oligonucleotide Synthesis, A Practical Approach, The Practical Approach Series* (IRL Press, Oxford, 1984)). En la etapa (B), se sintetizó un molde de traducción mínimo (denominado "43-P"), que contenía 43 nucleótidos, usando una química estándar de ARN y ADN (Millipore, Bedford, MA), se desprotegió usando NH₄OH y TBAF, y se purificó en gel. El molde contenía 13 bases de ARN en el extremo 5', seguido de 29 bases de ADN unidas a la 3'-puromicina en su 5'OH. La secuencia de ARN contenía (i) una secuencia de consenso de Shine-Dalgarno, complementaria a las cinco bases de 16S ARNr (Stormo *et al.*, *Nucleic Acids Research* **10**:2971-2996 (1982); Shine y Dalgarno, *Proc. Natl. Acad. Sci. USA* **71**:1342-1346 (1974); y Steitz y Jakes, *Proc. Natl. Acad. Sci. USA* **72**:4734-4738 (1975)); (ii) un espaciador de cinco bases, y (iii) un único codón de partida AUG. La secuencia de ADN fue dA₂₇dCdCP, en la que "P" es puromicina.

La Figura 4 es una representación esquemática de un método preferido para la preparación de puromicina enlazada a CPG protegida.

La Figura 5 es una representación esquemática que muestra posibles modos de incorporación de metionina en un molde de la invención. Como se muestra en la reacción (A), el molde se une al ribosoma, permitiendo la formación del complejo de iniciación 70S. El ARNt fmet se une al sitio de P, y su base se empareja con el molde. La puromicina en el extremo 3' del molde entra en el sitio A de manera intramolecular, y forma un enlace de amida con la N-formilmetionina vía el centro de peptidiltransferasa, desactivando de ese modo el ARNt. La extracción con fenol/cloroformo de la reacción conduce al molde que tiene la metionina unida covalentemente. Se muestra en la reacción (B) una reacción intermolecular no deseada del molde con oligonucleótidos que contienen puromicina. Como antes, el molde mínimo estimula la formación del ribosoma 70S que contiene ARNt fmet unido al sitio de P. Esto es seguido por la entrada de un segundo molde en trans para dar una metionina unida covalentemente.

Las Figuras 6A-6H son fotografías que muestran la incorporación de ³⁵S-metionina (³⁵S met) en moldes de traducción. La Figura 6A demuestra la dependencia de la reacción con respecto al magnesio (Mg²⁺). La Figura 6B demuestra la estabilidad de bases del producto; el cambio en la movilidad mostrado en esta figura corresponde a una pérdida de la secuencia 5' de ARN de 43-P (también denominado "molde de Met"), para producir la porción de ADN-puromicina, denominada 30-P. La retención del marcador, después del tratamiento con una base, fue consistente con la formación de un enlace peptídico entre ³⁵S-metionina y la 3'-puromicina del molde. La Figura 6C demuestra la inhibición de la formación del producto en presencia de inhibidores de peptidiltransferasa. La Figura 6D demuestra la independencia de la incorporación de ³⁵S-metionina en una secuencia que codifica el molde. La Figura 6E demuestra la dependencia de la longitud del molde de ADN con respecto a la incorporación de ³⁵S-metionina. La Figura 6F ilustra la formación del producto cis frente a trans usando moldes 43-P y 25-P. La Figura 6G ilustra la formación del producto cis frente a trans usando moldes 43-P y 13-P. La Figura 6H ilustra la formación del producto cis frente a trans usando moldes 43-P y 30-P en un sistema de lisado reticulocítico.

Las Figuras 7A-7C son ilustraciones esquemáticas de constructos para ensayar la formación y selección de fusiones de péptidos. La Figura 7A muestra LP77 ("producto ligado", "77" nucleótidos de longitud) (también denominado "molde myc corto") (SEC ID n°: 1). Esta secuencia contiene el marcador del epítipo del anticuerpo monoclonal c-myc EQKLISEEDL (SEC ID n°: 2) (Evan *et al.*, *Mol. Cell Biol.* **5**:3610-3616 (1985)), flanqueado por un codón de partida en 5' y un ligador en 3'. La región de 5' contiene una secuencia bacteriana de Shine-Dalgarno, idéntica a la de 43-P. La secuencia codificante se optimizó para la traducción en sistemas bacterianos. En particular, las 5'UTR de 43-P y LP77 contenían una secuencia de Shine-Dalgarno complementaria a cinco bases de ARNr 16S (Steitz y Jakes, *Proc. Natl. Acad. Sci. USA* **72**:4734-4738 (1975)), y espaciada de forma similar a las secuencias proteínicas ribosómicas (Stormo *et al.*, *Nucleic Acids Res.* **10**:2971-2996 (1982)). La Figura 7B muestra LP154 (producto ligado, 154 nucleótidos de longitud) (también denominado "molde myc largo") (SEC ID n°: 3). Esta secuencia contiene el código para la generación del péptido usado para aislar el anticuerpo c-myc. El extremo 5' contiene una versión truncada de la secuencia en dirección 5' de TMV (denominada "TE"). Esta 5'UTR contenía una secuencia de 22 nucleótidos derivada de TMV 5'UTR que comprende dos repeticiones directas ACAAAUAC (Gallie *et al.*, *Nucl. Acids Res.* **16**:883 (1988)). La Figura 7C muestra el conjunto n° 1 (SEC ID n°: 4), una secuencia ejemplar a usar para la selección de péptidos. Los siete aminoácidos finales del péptido myc original se incluyeron en el molde para que sirvan como la región constante 3' requerida para la amplificación mediante PCR del molde. Esta secuencia no se muestra como parte del epítipo de unión al anticuerpo.

La Figura 8 es una fotografía que demuestra la síntesis de fusiones de ARN-proteína usando moldes 43-P, LP77, y LP154, y sistemas de traducción reticulocíticos ("Retic") y de germen de trigo ("Trigo"). La mitad izquierda de la figura ilustra la incorporación de ³⁵S-metionina en cada uno de los tres moldes. La mitad derecha de la figura ilustra los productos resultantes después del tratamiento con ARNasa A de cada uno de los tres moldes para eliminar la región codificante de ARN; se muestran las fusiones de ADN-proteína marcadas con ³⁵S-metionina. La porción de ADN de cada una fue idéntica al oligo 30-P. De este modo, las diferencias en la movilidad fueron proporcionales a la longitud de las regiones codificantes, consistente con la existencia de proteínas de diferente longitud en cada caso.

La Figura 9 es una fotografía que demuestra la sensibilidad a proteasas de una fusión de ARN-proteína sintetizada a partir de LP154, y analizada desnaturalizando mediante electroforesis con gel de poliacrilamida. La línea 1 contiene 30-P marcada con ³²P. Las líneas 2-4, 5-7 y 8-10 contienen los moldes de traducción marcados con ³⁵S, recuperados de

ES 2 268 763 T3

reacciones con lisados reticulocíticos, ya sea sin tratamiento, con tratamiento con ARNasa A, o con tratamiento con ARNasa A y proteinasa K, respectivamente.

La Figura 10 es una fotografía que muestra los resultados de reacciones de inmunoprecipitación usando la proteína del epítipo myc de 23 aminoácidos traducida *in vitro*. Las líneas 1 y 2 muestran los productos de traducción de la proteína del epítipo myc y los moldes de β -globina, respectivamente. Las líneas 3-5 muestran los resultados de la inmunoprecipitación del péptido del epítipo myc usando un anticuerpo monoclonal c-myc, y tampones de lavado de PBS, DB y PBSTDS, respectivamente. Las líneas 6-8 muestran las mismas reacciones de inmunoprecipitación, pero usando el producto de la traducción con β -globina.

La Figura 11 es una fotografía que demuestra la inmunoprecipitación de la fusión de ARN-proteína a partir de una reacción de traducción *in vitro*. Se indican los picomoles de molde usados en la reacción. Las líneas 1-4 muestran ARN 124 (la porción de ARN de la fusión LP154), y las líneas 5-7 muestran la fusión de ARN-proteína LP154. Después de la inmunoprecipitación usando un anticuerpo monoclonal c-myc y proteína G-sefarosa, las muestras se trataron con ARNasa A y con T4 polinucleótido quinasa, y después se cargaron en un gel de poli(acrilamida) con urea desnaturalizante, para visualizar la fusión. En las líneas 1-4, con muestras que no contienen molde o que sólo contienen la porción de ARN del molde myc largo (RNA124), no se observó ninguna fusión. En las líneas 5-7, se visualizan claramente las bandas que corresponden a la fusión. Se indica la posición de 30-P marcada con ^{32}P , y se indica en la parte superior de la figura la cantidad de molde introducido.

La Figura 12 es una gráfica que muestra la cuantificación del material de fusión obtenido a partir de una reacción de traducción *in vitro*. La intensidad de las bandas de fusión mostradas en las líneas 5-7 de la Figura 11, y la banda de 30-P (aislada de manera paralela en dT_{25} , no mostrada) se cuantificaron en placas formadoras de fosfoimágenes, y se representaron gráficamente como una función de la concentración de entrada de LP154. La 30-P modificada recuperada (eje y izquierdo) fue linealmente proporcional al molde de entrada (eje x), mientras que la fusión del ligador-péptido (eje y derecho) fue constante. A partir de este análisis, se calculó que se habían formado $\sim 10^{12}$ fusiones por ml de muestra de reacción de traducción.

La Figura 13 es una representación esquemática de tiopropilsefarosa y dT_{25} agarosa, y la capacidad de estos sus-tratos para interactuar con las fusiones de ARN-proteína de la invención.

La Figura 14 es una fotografía que muestra los resultados del aislamiento secuencial de fusiones de la invención. La línea 1 contiene 30-P marcada con ^{32}P . Las líneas 2 y 3 muestran LP154 aislada de reacciones de traducción y tratada con ARNasa A. En la línea 2, se aisló secuencialmente LP154 usando tiopropilsefarosa, seguido de dT_{25} agarosa. La línea 3 muestra el aislamiento usando sólo dT_{25} agarosa. Los resultados indicaron que el producto contenía un tiol libre, probablemente la penúltima cisteína en la secuencia codificante del epítipo myc.

Las Figuras 15A y 15B son fotografías que muestran la formación de productos de fusión usando moldes de β -globina según se ensaya mediante SDS-tricina-PAGE (electroforesis en gel de poli(acrilamida)). La Figura 15A muestra la incorporación de ^{35}S sin usar molde (línea 1), usando un molde de *syn*- β -globina (líneas 2-4), o un molde de LP- β -globina (líneas 5-7). La Figura 15B (líneas etiquetadas como la Fig. 15A) muestra el material marcado con ^{35}S , aislado mediante cromatografía de afinidad oligonucleotídica. No se aisló ningún material en ausencia de la cola de 30-P (líneas 2-4).

Las Figuras 16A-16C son diagramas y fotografías que ilustran el enriquecimiento de ADNds myc frente a ADNds del conjunto mediante selección *in vitro*. La Figura 16A es una representación esquemática del protocolo de selección. Se tradujeron *in vitro* cuatro mezclas de los moldes de myc y del conjunto, y se aislaron en dT_{25} agarosa, seguido de TP sefarosa para purificar las fusiones del molde de los moldes no modificados. Las fusiones de ARNm-péptido se transcribieron de forma inversa entonces para suprimir cualquier estructura secundaria o terciaria presente en los moldes. Se retiraron alícuotas de cada mezcla tanto antes (Figura 16B) como después (Figura 16C) de la selección por afinidad, se amplificaron mediante PCR en presencia de un cebador marcado, y se digirieron con una enzima de restricción que rompió sólo el ADN de myc. Las mezclas de entrada de los moldes fueron myc puro (línea 1), o una mezcla myc:conjunto 1:20, 1:200 o 1:2000 (líneas 2-4). El material sin seleccionar se desvió de las relaciones de entrada debido a la traducción preferente y a la transcripción inversa del molde de myc. El enriquecimiento del molde de myc durante la etapa selectiva se calculó a partir del cambio en la relación de conjunto:myc antes y después de la selección.

La Figura 17 es una fotografía que ilustra la traducción de moldes de ARN de myc. Se usaron los siguientes ligadores: líneas 1-4, $dA_{27}dCdCp$; líneas 5-8, $dA_{27}tCrCP$; y líneas 9-12, $dA_{21}C_9C_9dAdCdCP$. En cada línea, la concentración del molde de ARN fue 600 nM, y se usó ^{35}S -Met para el marcaje. Las condiciones de reacción fueron las siguientes: líneas 1, 5 y 9, 30°C durante 1 hora; líneas 2, 6 y 10, 30°C durante 2 horas; líneas 3, 7 y 11, 30°C durante 1 hora, -20°C durante 16 horas; y líneas 4, 8 y 12, 30°C durante 1 hora, -20°C durante 16 horas con 50 mM de Mg^{2+} . En esta Figura, "A" representa un péptido libre, y "B" representa la fusión de ARNm-péptido.

La Figura 18 es una fotografía que ilustra la traducción de moldes de ARN de myc marcados con ^{32}P . El ligador utilizado fue $dA_{21}C_9C_9dAdCdCP$. La traducción se realizó a 30°C durante 90 minutos, y se llevaron a cabo incubaciones a -20°C durante 2 días sin Mg^{2+} adicional. Las concentraciones de los moldes de ARNm fueron 400 nM (línea 3), 200 nM (línea 4), 100 nM (línea 5), y 100 nM (línea 6). La línea 1 muestra la fusión de ARNm-péptido marcada

ES 2 268 763 T3

con ³⁵S-Met. La línea 2 muestra ARNm marcado con ³²P. En la línea 6, la reacción se llevó a cabo en presencia de 0,5 mM de análogo de cap.

La Figura 19 es una fotografía que ilustra la traducción del molde de ARN de myc usando un lisado obtenido de Ambion (línea 1), Novagen (línea 2), y Amersham (línea 3). El ligador utilizado fue dA₂₇dCdCP. La concentración del molde fue 600 nM, y se usó ³⁵S-Met para el marcaje. Las traducciones se realizaron a 30°C durante 1 hora, y las incubaciones se llevaron a cabo a -20°C toda la noche en presencia de 50 mM de Mg²⁺.

Se describe en la presente memoria un método general para la selección de proteínas con funciones deseadas, usando fusiones en las que estas proteínas están enlazadas covalentemente a sus propios ARN mensajeros. Estas fusiones de ARN-proteína se sintetizan mediante traducción *in vitro* o *in situ* de conjuntos de ARNm que contienen un aceptor peptídico unido en sus extremos 3' (Figura 1B). En una realización preferida, después de la lectura del marco de lectura abierta del mensaje, el ribosoma se detiene cuando alcanza el sitio de pausa diseñado, y el resto del aceptor ocupa el sitio ribosómico A y acepta la cadena peptídica naciente procedente del ARNt peptídico en el sitio P, para generar la fusión de ARN-proteína (Figura 1C). El enlace covalente entre la proteína y el ARN (en forma de un enlace amídico entre el extremo 3' del ARNm y el término C de la proteína que purifica) permite que se recupere y amplifique (por ejemplo, mediante PCR) la información genética en la proteína, seguido de la selección mediante transcripción inversa del ARN. Una vez que se genera la fusión, se lleva a cabo la selección o enriquecimiento basándose en las propiedades de la fusión de ARNm-proteína, o, como alternativa, se puede llevar a cabo una transcripción inversa usando el molde de ARNm mientras está unido a la proteína, para evitar cualquier efecto del ARN monocatenario sobre la selección. Cuando se usa el constructo de ARNm-proteína, se pueden estudiar fusiones seleccionadas para determinar qué resto (la proteína, el ARN, o ambos) proporciona la función deseada.

En una realización preferida, la puomicina (que se asemeja a la tirosiladenosina) actúa como el aceptor para unir el péptido en crecimiento a su ARNm. La puomicina es un antibiótico que actúa terminando el alargamiento del péptido. Como mimético de aminoacil-ARNt, actúa como un inhibidor universal de la síntesis de proteínas uniéndose al sitio A, aceptando la cadena peptídica en crecimiento, y desprendiéndose del ribosoma (a una K_d = 10⁻⁴ M) (Traut y Monro, J. Mol. Biol. **10**:63 (1964); Smith *et al.*, J. Mol. Biol. **13**:617 (1965)). Una de las características más atractivas de la puomicina es el hecho de que forma un enlace amídico estable con la cadena peptídica en crecimiento, permitiendo de este modo fusiones más estables que los aceptores potenciales que forman enlaces ésteres inestables. En particular, la molécula de peptidil-puomicina contiene una unión amídica estable entre el péptido y la porción de O-metiltirosina de la puomicina. A su vez, la O-metiltirosina está enlazada a un enlace amídico estable al grupo amino en 3' de la porción de adenosina modificada de puomicina.

Otras elecciones posibles para aceptores incluyen estructuras semejantes a ARNt en el extremo 3' del ARNm, así como otros compuestos que actúan de manera similar a la puomicina. Tales compuestos incluyen, sin limitación, cualquier compuesto que posea un aminoácido enlazado a una adenina o a un compuesto similar a adenina, tal como los nucleótidos aminoácidos, fenilalanil-adenosina (A-Phe), tirosil-adenosina (A-Tyr), y alanil-adenosina (A-Ala), así como estructuras enlazadas mediante amida, tal como fenilalanil-3'-desoxi-3'-aminoadenosina, alanil-3'-desoxi-3'-aminoadenosina, y tirosil-3'-desoxi-3'-aminoadenosina; en cualquiera de estos compuestos, se puede utilizar cualquiera de los L-aminoácidos de origen natural, o sus análogos. Además, también se puede usar en la invención un conjugado de puomicina con estructura de 3' similar a ARNt combinado.

Se muestra en la Figura 2 un esquema de selección preferido según la invención. Las etapas implicadas en esta selección se llevan a cabo generalmente según lo siguiente.

Etapas 1

Preparación del molde de ADN

Como una etapa para generar las fusiones de ARN-proteína de la invención, se sintetiza la porción de ARN de la fusión. Esto se puede lograr mediante síntesis de ARN química directa, o, más habitualmente, se logra transcribiendo un molde de ADN bicatenario apropiado.

Tales moldes de ADN se pueden crear mediante cualquier técnica estándar (incluyendo cualquier técnica de tecnología de ADN recombinante, síntesis química, o ambas). En principio, para este fin, se puede usar cualquier método que permita la producción de uno o más moldes que contenga una secuencia conocida, aleatoria, aleatorizada, o mutagenizada. En un enfoque particular, se sintetiza un oligonucleótido (por ejemplo, que contiene bases al azar), y se amplifica (por ejemplo, mediante PCR) antes de la transcripción. También se puede usar la síntesis química para producir un casete aleatorio que se inserta entonces en el centro de una secuencia codificante de proteína conocida (véase, por ejemplo, el capítulo 8.2, Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley & Sons y Greene Publishing Company, 1994). Este último enfoque produce una densidad de mutaciones elevada alrededor de un sitio específico de interés en la proteína.

Una alternativa a la aleatorización total de una secuencia de molde de ADN es la aleatorización parcial, y un conjunto sintetizado de este modo generalmente se denomina como conjunto "dopado". Un ejemplo de esta técnica, realizada sobre una secuencia de ARN, se describe, por ejemplo, por Ekland *et al.* (Nucl. Acids Research **23**:3231 (1995)). La aleatorización parcial se puede realizar químicamente inclinando las reacciones de síntesis de tal forma

que cada mezcla de reacción de adición de bases contenga un exceso de una base y pequeñas cantidades de cada una de las otras; mediante el control cuidadoso de las concentraciones de las bases, se puede lograr mediante este enfoque una frecuencia de mutación deseada. Los conjuntos aleatorizados parcialmente también se pueden generar usando técnicas de PCR con tendencia al error, por ejemplo como se describe en Beaudry y Joyce (Science **257**:635 (1992)) y Bartel y Szostak (Science **261**:1411 (1993)).

También existen numerosos métodos para generar un constructo de ADN que comience con una secuencia conocida, y creando después un conjunto de ADN mutagenizado. Los ejemplos de tales técnicas se describen en Ausubel *et al.* (citado anteriormente, capítulo 8) y Sambrook *et al.* (Molecular Cloning: A Laboratory Manual, capítulo 15, Cold Spring Harbor Press, Nueva York, 2ª ed. (1989)). Las secuencias al azar también se pueden generar mediante la técnica de “barajado” representada en Stemmer (Nature **370**:389 (1994)).

Para optimizar un esquema de selección de la invención, también se pueden alterar las secuencias y estructuras en los extremos 5' y 3' de un molde. Preferentemente, esto se lleva a cabo en dos selecciones separadas, que implican cada una la inserción de dominios al azar en el molde, próximos al extremo apropiado, seguido de la selección. Estas selecciones pueden servir (i) para maximizar la cantidad de fusión obtenida (y de este modo para maximizar la complejidad de una librería), o (ii) para proporcionar secuencias de traducción optimizadas. Además, el método puede ser aplicable generalmente, combinada con PCR mutagénica, a la optimización de moldes de traducción tanto en las regiones codificantes como no codificantes.

Etapa 2

Generación de ARN

Como se ha señalado anteriormente, la porción de ARN de una fusión de ARN-proteína se puede sintetizar químicamente usando técnicas estándares de síntesis de oligonucleótidos. Como alternativa, y en particular si se utilizan secuencias de ARN más largas, la porción de ARN se genera mediante transcripción *in vitro* de un molde de ADN. En un enfoque preferido, se usa T7 polimerasa para generar enzimáticamente la cadena de ARN. Otras ARN polimerasas apropiadas para este uso incluyen, sin limitación, las SP6, T3 y ARN de *E.coli* polimerasas (descritas, por ejemplo, en Ausubel *et al.* (citado anteriormente, capítulo 3). Además, el ARN sintetizado puede ser, en todo o en parte, ARN modificado. En un ejemplo particular, se puede producir ARN de fosforotioato (por ejemplo, mediante transcripción con T7) usando ribonucleótidos modificados y técnicas estándares. Tal ARN modificado proporciona la ventaja de ser estable frente a nucleasas.

Etapa 3

Ligación de puromicina al molde

Seguidamente, la puromicina (o cualquier otro aceptor de péptidos apropiado) se enlaza covalentemente a la secuencia del molde. Esta etapa se puede lograr usando T4 ARN ligasa, para unir la puromicina directamente a la secuencia de ARN, o preferentemente la puromicina se puede unir por medio de una “tablilla” de ADN, usando T4 ADN ligasa o cualquier otra enzima que sea capaz de unir juntas a dos secuencias nucleotídicas (véase la Figura 1B) (véase también, por ejemplo, Ausubel *et al.*, citado anteriormente, capítulo 3, secciones 14 y 15). También se pueden usar ARNt sintetizadas para unir compuestos similares a puromicina al ARN. Por ejemplo, la fenilalanil ARNt sintetasa enlaza la fenilalanina a moléculas de fenilalanil-ARNt que contienen un grupo amino en 3', generando moléculas de ARN con extremos en 3' similares a puromicina (Fraser y Rich, Proc. Natl. Acad. Sci. USA **70**:2671 (1973)). Otros aceptores peptídicos que se pueden usar incluyen, sin limitación, cualquier compuesto que posea un aminoácido enlazado a una adenina o a un compuesto similar a adenina, tal como los nucleótidos de aminoácidos, fenilalanil-adenosina (A-Phe), tirosil-adenosina (A-Tyr), y alanil-adenosina (A-Ala), así como estructuras enlazadas mediante amida, tales como fenilalanil-3'-desoxi-3'-aminoadenosina, alanil-3'-desoxi-3'-aminoadenosina, y tirosil-3'-desoxi-3'-aminoadenosina; en cualquiera de estos compuestos, se puede utilizar cualquiera de los L-aminoácidos de origen natural, o sus análogos. En Krayevsky y Kukhanova, Progress in Nucleic Acids Research and Molecular Biology **23**:1 (1979), por ejemplo, se describe un número de aceptores de péptidos.

Etapa 4

Generación y recuperación de fusiones de ARN-proteína

Para generar fusiones de ARN-proteína, se puede utilizar cualquier sistema de traducción *in vitro* o *in situ*. Como se muestra a continuación, se prefieren sistemas eucariotas, y dos sistemas particularmente preferidos incluyen los sistemas de germen de trigo y de lisado reticulocítico. En principio, sin embargo, en la invención es útil cualquier sistema de traducción que permita la formación de una fusión de ARN-proteína, y que no degrade significativamente la porción de ARN de la fusión. Además, para reducir la degradación del ARN en cualquiera de estos sistemas, se pueden incluir en la mezcla de reacción de la traducción oligonucleótidos antisentido que bloqueen la degradación; tales oligonucleótidos se hibridan específicamente a, y cubren, secuencias dentro de la porción de ARN de la molécula que disparan la degradación (véase, por ejemplo, Hanes y Pluckthun, Proc. Natl. Acad. Sci. USA **94**:4937 (1997)).

Como se señala anteriormente, existe, para uso en la invención, cualquier número de sistemas de traducción eucariotas. Estos incluyen, sin limitación, lisados de levadura, ascitis, células tumorales (Leibowitz *et al.*, Meth. Enzymol. **194**:536 (1991)), y huevos oocíticos de *Xenopus*. Los sistemas de traducción *in vitro* útiles, procedentes de sistemas bacterianos, incluyen, sin limitación, aquellos descritos en Zubay (Ann. Rev. Genet. **7**:267 (1973)); Chen y Zubay (Meth. Enzymol. **101**:44 (1983)); y Ellman (Meth. Enzymol. **202**:301 (1991)).

Además, las reacciones de traducción se pueden llevar a cabo *in situ*. En un ejemplo particular, la traducción se puede llevar a cabo inyectando ARNm en huevos de *Xenopus*, usando técnicas estándares.

Una vez generadas, las fusiones de ARN-proteína se pueden recuperar de la mezcla de reacción de la traducción mediante cualquier técnica estándar de purificación de proteínas o de ARN. Típicamente, se utilizan técnicas de purificación de proteínas. Como se muestra a continuación, por ejemplo, la purificación de una fusión se puede facilitar mediante el uso de reactivos cromatográficos adecuados, tales como dT₂₅ agarosa o tiopropilsefarosa. Sin embargo, la purificación puede implicar también, o como alternativa, la purificación a base de la porción de ARN de la fusión; en Ausubel *et al.* (citado anteriormente, capítulo 4), por ejemplo, se describen técnicas para tal purificación.

Etapa 5

Selección de la fusión de ARN-proteína deseada

La selección de una fusión de ARN-proteína deseada se puede lograr mediante cualquier medio disponible para separar o aislar selectivamente una fusión deseada de una población de fusiones candidatas. Los ejemplos de técnicas de aislamiento incluyen, pero sin limitación, la unión selectiva, por ejemplo, a una pareja de unión que se inmoviliza directa o indirectamente en una columna, perla, membrana, u otro soporte sólido, y la inmunoprecipitación, usando un anticuerpo específico para el resto de proteína de la fusión. La primera de estas técnicas hace uso de un motivo de selección inmovilizado que consiste en cualquier tipo de molécula a la que es posible unirse. En la Figura 2 se presenta una lista de posibles moléculas del motivo de selección. La selección también se puede basar en el uso de moléculas sustrato unidas a un marcador de afinidad (por ejemplo, sustrato-biotina), que reacciona con una molécula candidata, o en cualquier otro tipo de interacción con una molécula de fusión. Además, las proteínas se pueden seleccionar basándose en su actividad catalítica, de manera análoga a la descrita por Bartel y Szostak para el aislamiento de enzimas de ARN (citado anteriormente); según esa técnica particular, las moléculas deseadas se seleccionan basándose en su capacidad para unir una molécula diana a ellas mismas, y las moléculas funcionales se aíslan entonces basándose en la presencia de esa diana. Mediante la presente invención, son posibles los esquemas de selección para aislar nuevas proteínas catalíticas o proteínas catalíticas mejoradas, usando este mismo enfoque o cualquier otra selección funcional.

Además, como se describe en la presente memoria, la selección de una fusión de ARN-proteína deseada (o su copia de ADN) puede estar facilitada mediante el enriquecimiento para esa fusión en un conjunto de moléculas candidatas. Para llevar tal enriquecimiento opcional, se pone en contacto una población de fusiones candidatas de ARN-proteína con una pareja de unión (por ejemplo, una de las parejas de unión descritas anteriormente), que es específica para la porción de ARN o para la porción de proteína de la fusión, en condiciones que separen sustancialmente el complejo de pareja de unión-fusión de los miembros no unidos en la muestra. Esta etapa se puede repetir, y la técnica incluye preferentemente al menos dos etapas de enriquecimiento secuenciales, una en la que se seleccionan las fusiones usando una pareja de unión específica para la porción de ARN, y otra en la que se seleccionan las fusiones usando una pareja de unión específica para la porción de proteína. Además, si se repiten las etapas de enriquecimiento que van dirigidas a la misma porción de la fusión (por ejemplo, la porción de proteína), preferentemente se utilizan diferentes parejas de unión. En un ejemplo particular descrito en la presente memoria, se enriquece una población de moléculas para fusiones deseadas, usando en primer lugar una pareja de unión específica para la porción de ARN de la fusión, y, después, en dos etapas secuenciales, usando dos parejas de unión diferentes, las cuales son específicas para la porción de proteína de la fusión. Nuevamente, estos complejos se pueden separar de los componentes de la muestra mediante cualquier técnica de separación estándar, incluyendo, sin limitación, la cromatografía de afinidad en columna, la centrifugación o la inmunoprecipitación.

Además, la elución de una fusión de ARN-proteína a partir de un complejo de enriquecimiento (o de selección), se puede lograr mediante un número de enfoques. Por ejemplo, como se describe en la presente memoria, se puede utilizar una etapa de elución química desnaturizante o no específica, para aislar una fusión deseada de ARN-proteína. Tal etapa facilita la liberación de los componentes del complejo entre sí, o a partir de un soporte sólido asociado, de una manera relativamente no específica, rompiendo enlaces no covalentes entre los componentes y/o entre los componentes y el soporte sólido. Como se describe en la presente memoria, un reactivo de elución química desnaturizante o no específico ejemplar es 4% de HOAc/H₂O. Otros reactivos de elución química desnaturizantes o no específicos ejemplares incluyen guanidina, urea, alta concentración de sal, detergente, o cualquier otro medio mediante el cual se pueden generalmente eliminar los aductos no covalentes. Como alternativa, se puede utilizar un enfoque de elución química específico, en el que se explota un producto químico que provoca la liberación específica de una molécula de fusión. En un ejemplo particular, si el brazo ligador de una proteína de fusión deseada contiene uno o más enlaces de disulfuro, los aptámeros de fusión unidos se pueden eluir mediante adición, por ejemplo, de DTT, dando como resultado la reducción del enlace de disulfuro y la liberación de la diana unida.

Como alternativa, la elución se puede lograr destruyendo específicamente los complejos de afinidad; tales técnicas liberan selectivamente los componentes del complejo mediante adición de un exceso de un miembro del complejo.

ES 2 268 763 T3

Por ejemplo, en una selección de unión a ATP, la elución se realiza mediante adición de ATP en exceso a la mezcla de incubación. Finalmente, se puede llevar a cabo una etapa de elución enzimática. Mediante este enfoque, la propia molécula unida, o una proteasa añadida exógenamente (u otra enzima hidrolítica apropiada), rompe y libera la diana o la enzima. En un ejemplo particular, se puede incluir un sitio de diana de proteasa en cualquiera de los componentes del complejo, y las moléculas unidas se pueden eluir mediante adición de la proteasa. Como alternativa, en una selección catalítica, se puede usar una elución como una etapa de selección para aislar moléculas capaces de liberarse (por ejemplo, de escindir-se) a partir de un soporte sólido.

Etapa 6

Generación de una copia de ADN de la secuencia de ARN usando transcriptasa inversa

Si se desea, se dispone fácilmente de una copia de ADN de una secuencia de fusión de ARN seleccionada, transcribiendo de forma inversa esa secuencia de ARN usando cualquier técnica estándar (por ejemplo, usando transcriptasa inversa Superscript). Esta etapa se puede llevar a cabo antes de la etapa de selección o de enriquecimiento (por ejemplo, como se describe en la Figura 16), o después de esa etapa. Como alternativa, el proceso de transcripción inversa se puede llevar a cabo antes del aislamiento de la fusión a partir de la mezcla de traducción *in vitro* o *in situ*.

Después, el molde de ADN se amplifica, ya sea como una secuencia bicatenaria de longitud parcial o total. Preferentemente, en esta etapa, se generan moldes de ADN de longitud total, usando oligonucleótidos apropiados y una amplificación mediante PCR.

Estas etapas, y los reactivos y técnicas para llevar a cabo estas etapas, se describe ahora con detalle usando ejemplos particulares. Estos ejemplos se proporcionan con el fin de ilustrar la invención, y no se deben de interpretar como limitantes.

Generación de moldes para las fusiones de ARN-proteína

Como se muestra en las Figuras 1A y 2, el esquema de selección de la presente invención hace uso preferentemente de moldes de ADN bicatenario que incluyen un número de elementos de diseño. El primer de estos elementos es un promotor para ser usado en conjunción con una ARN polimerasa deseada para la síntesis de ARNm. Como se muestra en la Figura 1A, y como se describe en la presente memoria, se prefiere el promotor T7, aunque se puede usar cualquier promotor capaz de dirigir la síntesis a partir de un ADN bicatenario lineal.

El segundo elemento del molde mostrado en la Figura 1A se denomina la región no traducida 5' (o 5'UTR), y corresponde al ARN en dirección 5' del sitio de partida de la traducción. Se muestra en la Figura 1A una 5'UTR preferida (denominada "TE"), que es un mutante de supresión de la región no traducida de 5' del virus del mosaico del tabaco, y, en particular, corresponde a las bases directamente en 5' del comienzo de la traducción de TMV; la secuencia de esta UTR es la siguiente: rGrGrG rArCrA rArUrU rArCrU rArUrU rUrArC rArArU rUrArC rA (estando los primeros 3 nucleótidos G insertados para aumentar la transcripción) (SEC ID n°: 5). Se puede utilizar cualquier otra 5'UTR apropiada (véase, por ejemplo, Kozak, *Microbiol. Rev.* **47**:1 (1983)).

El tercer elemento mostrado en la Figura 1A es el sitio de partida de la traducción. En general, éste es un codón AUG. Sin embargo, existen ejemplos en los que se utilizan codones distintos de AUG en secuencias codificantes de origen natural, y estos codones también se pueden usar en el esquema de selección de la invención.

El cuarto elemento en la Figura 1A es el marco de lectura abierta de la proteína (denominado ORF), que codifica la secuencia de la proteína. Este marco de lectura abierta puede codificar cualquier secuencia proteínica de origen natural, aleatoria, aleatorizada, mutagenizada, o totalmente sintética.

El quinto elemento mostrado en la Figura 1A es la región constante 3'. Esta secuencia facilita la amplificación mediante PCR de las secuencias del conjunto, y la ligación del oligonucleótido que contiene puromicina al ARNm. Si se desea, esta región puede incluir también un sitio de pausa, una secuencia que hace que el ribosoma se detenga y permita de ese modo un tiempo adicional para que un resto aceptor (por ejemplo, puromicina) acepte una cadena peptídica naciente procedente del peptidil-ARNt; este sitio de pausa se expone con más detalle más abajo.

Para desarrollar la presente metodología, se generaron inicialmente fusiones de ARN-proteína usando moldes de ARNm muy simplificados, que contienen 1-2 codones. Este enfoque se tomó por dos razones. En primer lugar, los moldes de este tamaño se podrían obtener fácilmente mediante síntesis química. Y, en segundo lugar, un marco de lectura abierta pequeño permitiría que se evaluaran fácilmente características críticas de la reacción, incluyendo la eficacia del enlace, la heterogeneidad del extremo, la dependencia del molde, y la exactitud de la traducción.

Diseño del constructo. Se usó un constructo básico para generar fusiones de ARN-proteína de ensayo. La molécula consistió en un ARNm que contiene una secuencia de Shine-Dalgarno (SD) para la iniciación de la traducción que contenía una supresión de 3 bases de la secuencia de SD a partir de la proteína ribosómica L1 y que fue complementaria con 5 bases del 16S ARNr (es decir, rGrGrA rGrGrA rCrGrA rA (SEC ID n°: 6) (Stormo *et al.*, *Nucleic Acids Research* **10**:2971-2996 (1982); Shine y Dalgarno, *Proc. Natl. Acad. Sci. USA* **71**:1342-1346 (1974); y Steitz y Jakes, *Proc. Natl. Acad. Sci. USA* **72**:4734-4738 (1975)), (ii) un codón de partida de AUG, (iii) un ligador de ADN que actúa como un

sitio de pausa (es decir, 5'-(dA)₂₇), (iv) dCdC-3', y (v) una 3'-puromicina (P). La secuencia poli dA se escogió debido a que se sabe que amolda mal el ARNt en el sitio A (Morgan *et al.*, J. Mol. Biol. **26**:477-497 (1967); Ricker y Kaji, Nucleic Acid Research **19**:6573-6578 (1991)), y se diseñó para actuar como un buen sitio de pausa. La longitud del ligador de oligo dA se escogió para expandirse una distancia de ~60-70 Å entre el sitio descodificante y el centro de transferencia peptídico del ribosoma. El dCdCP imitó el extremo de CCA de un ARNt, y se diseñó para facilitar la unión de la puromicina al sitio A del ribosoma.

Síntesis química del molde mínimo 43-P. Para sintetizar el constructo 43-P (mostrado en la Figura 3), primero se unió puromicina a un soporte sólido, de tal forma que fuese compatible con la química de síntesis de oligonucleótidos mediante fosforamidito estándar. El protocolo de síntesis para este oligo se esquematiza en la Figura 3, y se describe con más detalle a continuación. Para unir la puromicina a un soporte sólido de vidrio de poro controlado (CPG), se protegió el grupo amino con un grupo trifluoroacetilo como se describe en Applied Biosystems User Bulletin #49 para el modelo 380 de sintetizador de ADN (1988). Después, se llevó a cabo la protección del 5'-OH, usando un enfoque de DMT-Cl estándar (Gait, Oligonucleotide Synthesis a practical approach The Practical Approach Series (IRL Press, Oxford, 1984)), y la unión al aminohexil CPG a través del 2'OH se efectuó exactamente de la misma forma como se usaría el 3'OH para la unión de un desoxinucleósido (véase la Fig. 3 y Gait, citado anteriormente, p. 47). La puromicina protegida enlazada a 5'DMR-CPG es entonces adecuada para la extensión de la cadena con monómeros de fosforamidito. La síntesis del oligo transcurrió en la dirección 3' → 5' en el orden: (i) 3' puromicina, (ii) pdCpdC, (iii) ~27 unidades de dA como ligador, (iv) AUG, y (v) la secuencia de Shine-Dalgarno. Más abajo se muestra la secuencia del constructo de 43-P.

Síntesis de CPG puromicina. La síntesis de CPG puromicina protegida siguió la ruta general usada para desoxinucleósidos como se esquematiza previamente (Gait, Oligonucleotide Synthesis, A Practical Approach, The Practical Approach Series (IRL Press, Oxford, 1984)). Las desviaciones principales incluyeron la selección de un grupo bloqueante de N apropiado, la unión en el 2'OH al soporte sólido, y la reacción de ligación al soporte sólido. En el caso de esta última, la reacción se llevó a cabo a concentraciones muy bajas de nucleótido activado, puesto que este material fue significativamente más precioso que el soporte sólido. El rendimiento resultante (~20 μmoles/g de soporte) fue bastante satisfactorio, considerando las condiciones de reacción diluidas.

Síntesis de N-trifluoroacetil-puromicina. Primero se convirtieron 267 mg (0,490 mmoles) de puromicina*HCl en la base libre, disolviendo en agua, añadiendo tampón de carbonato de pH 11, y extrayendo (3 X) en cloroformo. La fase orgánica se evaporó hasta sequedad y se pesó (242 mg, 0,513 mmoles). La base libre se disolvió entonces en 11 ml de piridina seca y 11 ml de acetonitrilo seco, y se añadieron con agitación 139 μl (2,0 mmoles) de trietilamina (TEA) y 139 μl (1,0 mmoles) de anhídrido trifluoroacético (TFAA). Después se añadió TFAA a la disolución turbia en alícuotas de 20 μl, hasta que no quedase nada del material de partida, según se determina mediante cromatografía de capa fina (tlc) (93:7, cloroformo/MeOH) (un total de 280 μl). La reacción se dejó transcurrir durante una hora. En este punto, se revelaron dos bandas mediante cromatografía de capa fina, ambas de una movilidad mayor que el material de partida. El tratamiento de la reacción con NH₄OH y agua redujo el producto hasta una sola banda. La cromatografía sobre sílice (93:7, cloroformo/MeOH) produjo 293 mg (0,515 mmoles) del producto, N-TFA-Pur. En la Figura 4 se muestra esquemáticamente el producto de esta reacción.

Síntesis de N-trifluoroacetil 5'-DMT puromicina. El producto de la reacción anterior se dividió en partes alícuotas y se coevaporó 2 X con piridina seca para eliminar el agua. Se prepararon múltiples tubos para estudiar las múltiples condiciones de reacción. En una reacción a pequeña escala, se disolvieron 27,4 mg (48,2 μmoles) de N-TFA-Pur en 480 μl de piridina que contiene 0,05 eq. de DMAP y 1,4 eq. de TEA. A esta mezcla se añadieron 20,6 mg de cloruro de tritilo (60 μmoles), y la reacción se dejó transcurrir hasta su terminación, con agitación. La reacción se detuvo mediante adición de un volumen igual de agua (aproximadamente 500 μl) a la disolución. Debido a que esta reacción parece exitosa, se realizó una versión a gran escala. En particular, se disolvieron 262 mg (0,467 mmoles) de N-TFA-Pur en 2,4 ml de piridina, seguido de la adición de 1,4 eq. de TEA, 0,05 eq. de DMAP, y 1,2 eq. de cloruro de tritilo. Después de aproximadamente 2 horas, se añadieron 50 mg (0,3 eq.) adicionales de dimetoxitritilo*Cl (DMT*Cl), y la reacción se dejó transcurrir durante 20 minutos adicionales. La reacción se detuvo mediante adición de 3 ml de agua, y se coevaporó 3 X con CH₃CN. La reacción se purificó mediante 95:5 de cloroformo/MeOH en una columna de sílice (seca) de 100 ml y 2 mm de diámetro. Debido a la purificación incompleta, se utilizó una segunda columna idéntica con 97,5:2,5 de cloroformo/MeOH. El rendimiento total fue 325 mg o 0,373 mmoles (o un rendimiento de 72%). En la Figura 4 se muestra esquemáticamente el producto de esta reacción.

Síntesis de N-trifluoroacetil, 5'-DMT, 2' succinilpuromicina. En una reacción a pequeña escala, se combinaron 32 mg (37 μmoles) del producto sintetizado anteriormente con 1,2 eq. de DMAP disuelto en 350 μl de piridina. A esta disolución se añadieron 1,2 equivalentes de anhídrido succínico en 44 μl de CH₃CN seco, y se dejó agitar toda la noche. La cromatografía de capa fina reveló que quedaba poco material de partida. En una reacción a gran escala, se combinaron 292 mg (336 μmoles) del producto previo con 1,2 eq. de DMAP en 3 ml de piridina. A esto, se añadieron 403 μl de anhídrido succínico 1 M en CH₃CN seco, y la mezcla se dejó agitar toda la noche. La cromatografía de capa fina reveló nuevamente que quedaba poco material de partida. Las dos reacciones se combinaron, y se añadieron 0,2 eq. adicionales de DMAP y de succinato. El producto se coevaporó con tolueno 1 X, y se secó hasta una espuma amarilla a alto vacío. Se añadió CH₂Cl₂ (20 ml), y esta disolución se extrajo dos veces con 15 ml de ácido cítrico al 10% enfriado en hielo, y después dos veces con agua pura. El producto se secó, se redisolvió en 2 ml de CH₂Cl₂, y se precipitó mediante adición de 50 ml de hexano con agitación. Después, el producto se sometió a remolino, y se centrifugó a 600 rpm durante 10 minutos en la centrifuga clínica. La mayoría del eluyente se extrajo por eliminación,

y el resto del producto se secó, primero a bajo vacío, y después a alto vacío en un secador. El rendimiento de esta reacción fue aproximadamente 260 micromoles para un rendimiento por etapas de ~70%.

5 *Síntesis de N-trifluoroacetil 5'-DMT, 2' succinil, CPG puromicina.* El producto de la etapa previa se disolvió después con 1 ml de dioxano, seguido de 0,2 ml de dioxano/0,2 ml de piridina. A esta disolución, se añadieron 40 mg de p-nitrofenol y 140 mg de dicitohexilcarbodiimida (DCC), y la reacción se dejó transcurrir durante 2 horas. La ciclohexilurea insoluble producida por la reacción se eliminó por centrifugación, y la disolución del producto se añadió a 5 g de vidrio de poro controlado aminohexílico (CPG) suspendido en 22 ml de DMF seca, y se agitó toda la noche. La resina se lavó entonces con DMF, con metanol y con éter, y se secó. Se vio que la resina resultante contenía 10 22,6 μ moles de tritilo por g, una cantidad bastante bien dentro del intervalo aceptable para este tipo de soporte. El soporte se encapsuló entonces mediante incubación con 15 ml de piridina, 1 ml de anhídrido acético, y 60 mg de DMAP, durante 30 minutos. El material de columna resultante produjo un ensayo negativo de ninhidrina (sin color), en contraste con los resultados obtenidos antes de bloquear, en los que el material produjo una reacción de color azul oscuro intenso. En la Figura 4 se muestra esquemáticamente el producto de esta reacción.

15 *Síntesis del conjugado de ARNm-puromicina.* Como se explica anteriormente, se puede usar un oligo restringido con puromicina en cualquiera de dos formas, para generar un conjugado de ARNm-puromicina que actúa como un molde de traducción. Para marcos de lectura abierta extremadamente cortos, el oligo de puromicina típicamente se extiende químicamente con monómeros de ARN o ADN para crear un molde totalmente sintético. Cuando se desean 20 marcos de lectura abierta más grandes, el oligo de ARN o de ADN generalmente se liga al extremo 3' de un ARNm usando una tablilla de ADN y una T4 ADN ligasa como se describe por Moore y Sharp (Science 256:992 (1992)).

Traducción in vitro y ensayo de fusiones DE ARN-Proteína

25 Los moldes generados anteriormente se tradujeron *in vitro* usando sistemas de traducción *in vitro* tanto bacterianos como eucariotas, según lo siguiente.

Traducción in vitro de moldes mínimos. Se añadieron 43-P y conjugados relacionados de ARN-puromicina a varios sistemas de traducción *in vitro* diferentes, incluyendo: (i) el sistema S30 derivado de *E.coli* MRE600 (Zubay, Ann. Rev. Genet. 7:267 (1973); Collins, Gene 6:29 (1979); Chen y Zubay, Methods Enzymol. 101:44 (1983); Pratt, en Transcription and Translation: A Practical Approach, B. D. Hammes, S. J. Higgins, Eds. (IRL Press, Oxford, 1984) p. 179-209; y Ellman *et al.*, Methods Enzymol. 202:301 (1991), preparado como se describe por Ellman *et al.* (Methods Enzymol. 202:301 (1991)); (ii) la fracción ribosómica derivada de la misma cepa, preparada como se describe por Kudlicki *et al.* (Anal. Chem. 206:389 (1992)); y (iii) el sistema S30 derivado de *E.coli* BL21, preparado como se describe por Lesley *et al.* (J. Biol. Chem. 266:2632 (1991)). En cada caso, la premezcla usada fue la de Lesley *et al.* (J. Biol. Chem. 266:2632 (1991)), y las incubaciones tuvieron 30 minutos de duración.

Ensayo de la naturaleza de la fusión. El molde de 43-P se evaluó primero usando extractos de traducción de S30 a partir de *E.coli*. La Figura 5 (reacción "A") demuestra la reacción intramolecular (cis) deseada, en la que 43-P se une al ribosoma y actúa como un molde para, y como un aceptor de, fMet, al mismo tiempo. En primer lugar, se estudió la incorporación de ³⁵S-metionina y su posición en el molde, y los resultados se muestran en las Figuras 6A y 6B. Después de la extracción de la mezcla de reacción de traducción *in vitro* con fenol/cloroformo, y el análisis de los productos mediante SDS-PAGE, apareció una banda marcada con ³⁵S con la misma movilidad que el molde de 43-P. La cantidad de este material sintetizado dependió de la concentración de Mg²⁺ (Figura 6A). Parece que la concentración de Mg²⁺ 45 óptima está entre 9 y 18 mM, que fue similar al óptimo para la traducción en este sistema (Zubay, Ann. Rev. Genet. 7:267 (1973); Collins, Gene 6:29 (1979); Chen y Zubay, Methods Enzymol. 101:44 (1983); Pratt, en Transcription and Translation: A Practical Approach, B. D. Hammes, S. J. Higgins, Eds. (IRL Press, Oxford, 1984) p. 179-209; Ellman *et al.*, Methods Enzymol. 202:301 (1991); Kudlicki *et al.*, Anal. Chem. 206:389 (1992); y Lesley *et al.*, J. Biol. Chem. 266:2632 (1991)). Además, el marcador incorporado fue estable al tratamiento con NH₄OH (Figura 6B), 50 indicando que el marcador estaba localizado en la mitad 3' de la molécula (la porción de ADN estable a bases), y que estaba unido mediante un enlace estable a bases, como es de esperar para un enlace amídico entre puromicina y fMet.

Dependencia del ribosoma y del molde. Para demostrar que la reacción observada anteriormente ocurrió en el ribosoma, se estudiaron (Figura 6C) los efectos de inhibidores específicos de la función de peptidiltransferasa del ribosoma, y se examinó el efecto del cambio de la secuencia que codifica la metionina (Figura 6D). La Figura 6C demuestra claramente que la reacción se inhibió fuertemente mediante los inhibidores de peptidiltransferasa, virginamicina, gougerotina, y cloranfenicol (Monro y Vazquez, J. Mol. Biol. 28:161-165 (1967); y Vazquez y Monro, Biochemica et Biophysical Acta 142:155-173 (1967)). La Figura 6D demuestra que el cambio de una única base en el molde, de A a C, abolió la incorporación de ³⁵S-metionina a 9 mM de Mg²⁺, y la redujo enormemente a 18 mM (consistente con el hecho de que niveles elevados de Mg²⁺ permiten que se interprete mal el mensaje). Estos experimentos 60 demostraron que la reacción ocurrió en el ribosoma, de una manera dependiente del molde.

Longitud del ligador. También se estudió la dependencia de la reacción con la longitud del ligador (Figura 6E). El molde original se diseñó de forma que el ligador se extendió la distancia desde el sitio descodificante (ocupado por el AUG del molde) hasta el sitio del aceptor (ocupado por el resto de puromicina), una distancia que fue aproximadamente la misma longitud que la distancia entre el bucle anticodón y el tallo del aceptor en un ARNt, o alrededor de 60-70 Å. El primer ligador estudiado tenía 30 nucleótidos de longitud, basado en un mínimo de 3,4 Å por base (≥ 102 Å). En el intervalo entre 30 y 21 nucleótidos ($n = 27-18$; longitud $\geq 102-71$ Å), se observó poco cambio en la eficacia de

la reacción. En consecuencia, se puede variar la longitud del ligador. Aunque un ligador entre 21 y 30 nucleótidos representa una longitud preferida, también se pueden utilizar en la invención ligadores más cortos de 80 nucleótidos, y, preferentemente, más cortos de 45 nucleótidos.

5 *Reacciones intramoleculares frente a las intermoleculares.* Finalmente, se estudió si la reacción ocurrió de una manera intramolecular (Figura 5, Reacción "A"), según se desea, o intermolecularmente (Figura 5, Reacción "B"). Esto se estudió añadiendo oligonucleótidos con 3' puomicina, pero sin la secuencia de unión a ribosoma (es decir, los moldes 25-P, 13-P y 30-P) a las reacciones de traducción que contienen el molde de 43-P (Figuras 6F, 6G y 6H). Si la reacción ocurrió mediante un mecanismo intermolecular, también se deberían de marcar los oligos más cortos. 10 Como se demuestra en las Figuras 6F-H, hubo muy poca incorporación de ³⁵S-metionina en los tres oligos más cortos, indicando que la reacción ocurrió principalmente de manera intramolecular. A continuación se muestran las secuencias de 25-P (SEC ID n°: 10), 13-P (SEC ID n°: 9), y 30-P (SEC ID n°: 8).

15 *Lisado reticulocítico.* La Figura 6H demuestra que la ³⁵S-metionina se puede incorporar en el molde de 43-P usando un lisado reticulocítico de conejo (véase más abajo) para la traducción *in vitro*, además de los lisados de *E. coli* usados anteriormente. Esta reacción ocurrió principalmente en un mecanismo intramolecular, según se desea.

Síntesis y estudio de fusiones que contienen un marcador del epítipo de c-myc

20 También se generaron fusiones ejemplares que contenían, dentro de la porción proteínica, el marcador del epítipo para el anticuerpo monoclonal c-myc 9E10 (Evan *et al.*, Mol. Cell Biol. 5:3610 (1985)).

25 *Diseño de los moldes.* Se diseñaron tres moldes de marcador de epítipo iniciales (es decir, LP77, LP154, y el Conjunto n° 1), y se muestran en las Figuras 7A-C. Los dos primeros moldes contenían la secuencia EQKLISEEDL (SEC ID n°: 2) del marcador del epítipo de c-myc, y el tercer molde fue el diseño usado en la síntesis de un conjunto de selección al azar. LP77 codificó una secuencia de 12 aminoácidos, con los codones optimizados para la traducción bacteriana. LP154 y sus derivados contenían una secuencia de ARNm de 33 aminoácidos, en la que los codones se optimizaron para la traducción eucariota. La secuencia de aminoácidos codificada de MAEEQKLISEED-LLRKRREQKLKHKLEQLRNSCA (SEC ID n°: 7) correspondió al péptido original usado para aislar el anticuerpo 30 9E10. El Conjunto n° 1 contenía 27 codones de NNG/C (para generar péptidos al azar) seguido de una secuencia que corresponde a los últimos 7 aminoácidos del péptido myc (que no son parte de la secuencia del epítipo de myc). Más abajo se muestran estas secuencias.

35 *Sistemas de traducción *in vitro* reticulocítico frente a germen de trigo.* Los moldes 43-P, LP77 y LP154 se evaluaron en sistemas de traducción tanto de reticulocito de conejo como de extracto de germen de trigo (Promega, Boehringer Mannheim) (Figura 8). Las traducciones se realizaron a 30°C durante 60 minutos. Los moldes se aislaron usando dT₂₅ agarosa a 4°C. Los moldes se eluyeron a partir de la agarosa usando 15 mM de NaOH, 1 mM de EDTA, se neutralizaron con tampón de NaOAc/HOAc, se precipitaron inmediatamente con etanol (2,5-3 vol.), se lavaron (con etanol al 100%), y se secaron en un concentrador Speedvac. La Figura 8 muestra que la ³⁵S-metionina se incorporó 40 en los tres moldes, tanto en el sistema de germen de trigo como en el reticulocítico. Se observó menos degradación del molde en las reacciones de fusión del sistema reticulocítico y, en consecuencia, se prefiere este sistema para la generación de fusiones de ARN-proteína. Además, en general, se prefieren los sistemas eucariotas con respecto a los sistemas bacterianos. Debido a que las células eucariotas tienden a contener niveles más bajos de nucleasas, la vida media del ARNm generalmente es 10-100 veces más larga en estas células que en las células bacterianas. En los 45 experimentos que usan un sistema de traducción de *E. coli* particular, no se observó generación de fusiones usando un molde según el epítipo de c-myc; el marcaje del molde en diversos lugares demostró que esto es probablemente debido a la degradación tanto de las porciones de ARN como de ADN del molde.

50 Para examinar la porción peptídica de estas fusiones, las muestras se trataron con ARNasa para eliminar las secuencias codificantes. Tras este tratamiento, el producto de 43-P tuvo una movilidad casi idéntica a la del oligo 30-P marcado con ³²P, consistente con un péptido muy pequeño (quizá sólo metionina) añadido a 30-P. Para LP77, la eliminación de la secuencia codificante produjo un producto con una movilidad menor que la del oligo 30-P, consistente con la noción de que se añadió a la puomicina un péptido de 12 aminoácidos. Finalmente, para LP154, la eliminación de la secuencia codificante produjo un producto de una movilidad aún menor, consistente con una 55 secuencia de 33 aminoácidos unida al oligo de 30-P. No se observó ningún oligo en la línea de reticulocito de LP154 tratado con ARNasa, debido a un error de carga. En la Figura 9, se observa que la movilidad de este producto es la misma que la del producto generado en el extracto de germen de trigo. En resumen, estos resultados indican que los productos resistentes a ARNasa se añadieron a los extremos de los oligos de 30-P, que los tamaños de los productos fueron proporcionales a la longitud de las secuencias codificantes, y que los productos fueron bastante homogéneos 60 en tamaño. Además, aunque ambos sistemas produjeron productos de fusión similares, el sistema reticulocítico parece ser superior debido a una mayor estabilidad del molde.

65 *Sensibilidad a ARNasa A y a proteinasa K.* En la Figura 9, se estudió la sensibilidad a ARNasa A y proteinasa K usando la fusión LP154. Como se muestra en las líneas 2-4, se demostró la incorporación de ³⁵S-metionina para el molde de LP154. Cuando este producto se trató con ARNasa A, disminuyó la movilidad de la fusión, pero aún fue significativamente mayor que la del oligonucleótido de 30-P marcado con ³²P, consistente con la adición de un péptido de 33 aminoácidos al extremo 3'. Cuando este material también se trató con proteinasa K, la señal de ³⁵S desapareció completamente, nuevamente consistente con la noción de que el marcador estaba presente en un péptido en el extremo

3' del fragmento de 30-P. Se han obtenido resultados similares en experimentos equivalentes usando las fusiones de 43-P y LP77.

Para confirmar que el marcaje del molde mediante ^{35}S -Met fue una consecuencia de la traducción, y más específicamente resultó de la actividad de peptidiltransferasa del ribosoma, se examinó el efecto de diversos inhibidores sobre la reacción de marcaje. Los inhibidores específicos de peptidiltransferasa eucariota, anisomicina, gougerotina y esparsomicina (Vazquez, *Inhibitors of Protein Biosynthesis* (Springer-Verlag, Nueva York), p. 312 (1979)), así como los inhibidores de la translocación cicloheximida y emetina (Vazquez, *Inhibitors of Protein Biosynthesis* (Springer-Verlag, Nueva York), p. 312 (1979)), disminuyeron todos la formación de la fusión de ARN-péptido en un ~95% usando el molde de myc largo y un extracto de traducción de lisado de reticulocito.

Experimentos de inmunoprecipitación. En un experimento diseñado para ilustrar la eficacia de inmunoprecipitar una fusión de ARNm-péptido, se hizo el intento de inmunoprecipitar un péptido de c-myc libre generado mediante traducción *in vitro*. La Figura 10 muestra los resultados de estos experimentos evaluados en un gel peptídico de SDS-PAGE. Las líneas 1 y 2 muestran el material marcado procedente de las reacciones de traducción que contiene RNA124 (la porción de ARN de LP154) o ARNm de β -globina. Las líneas 3-8 muestran la inmunoprecipitación de estas muestras de reacción usando el anticuerpo monoclonal de c-myc 9E10, en varias condiciones de tampón diferentes (descritas más abajo). Las líneas 3-5 muestran que el péptido derivado de RNA124 fue inmunoprecipitado de forma eficaz, siendo el mejor caso la línea 4, en la que se aislaron el ~83% de los recuentos precipitables de TCA totales. Las líneas 6-8 muestran poca proteína de β -globina, indicando una purificación de >100 veces. Estos resultados indicaron que el péptido codificado por RNA124 (y por LP154) se aisló cuantitativamente mediante este protocolo de inmunoprecipitación.

Inmunoprecipitación de la fusión. Después se estudió la capacidad para inmunoprecipitar un producto de ARN-péptido quimérico, usando una reacción de traducción de LP154 y el anticuerpo monoclonal de c-myc 9E10 (Figura 11). Los productos de la traducción procedentes de una reacción reticulocítica se aislaron mediante inmunoprecipitación (como se describe en la presente memoria), y se trataron con 1 μg de ARNasa A a temperatura ambiente durante 30 minutos, para eliminar la secuencia codificante. Esto generó un 5'OH, que se marcó con ^{32}P con T4 polinucleótido quinasa, y se evaluó mediante PAGE desnaturalizante. La Figura 11 demuestra que se aisló un producto con una movilidad similar a la observada para la fusión del epítipo de c-myc con 30-P generado mediante tratamiento con ARNasa de la fusión de LP154 (véase lo anterior), pero no se obtuvo ningún producto correspondiente cuando se tradujo sólo la porción de ARN del molde (RNA124). En la Figura 12, se determinó la cantidad de proteína de fusión aislada, y se representó gráficamente frente a la cantidad de 30-P no modificado (no mostrado en esta figura). La cuantificación de la relación de ligador no modificado a fusión de ligador-péptido de myc muestra que 0,2-0,7% del mensaje de entrada se convirtió en producto de fusión. Una fracción mayor del ARN de entrada se convirtió en producto de fusión en presencia de una mayor relación de ribosoma/molde; del intervalo de concentraciones de ARNm de entrada que se estudiaron, se obtuvieron aproximadamente $0,8-1,0 \times 10^{12}$ moléculas de fusión por ml de extracto de traducción.

Además, los resultados indicaron que los péptidos unidos a la especie de ARN fueron codificados por ese ARNm, es decir, el péptido naciente no se transfirió a la puromicina de algún otro ARNm. No se observó ninguna indicación de transferencia cruzada cuando se coincubó un ligador (30-P) con el molde de myc largo en extractos de traducción en relaciones tal altas como 20:1, ni la presencia del ligador libre disminuyó significativamente la cantidad de fusión de myc larga producida. De forma similar, la cotraducción de los moldes corto y largo, 43-P y LP154, produjo sólo los productos de fusión observados cuando los moldes se tradujeron solos, y no se observaron productos de movilidad intermedia, como sería de esperar para la fusión del molde corto con el péptido de myc largo. Ambos resultados sugieren que la formación de la fusión se produjo principalmente entre un péptido naciente y un ARNm unido al mismo ribosoma.

Aislamiento secuencial. Como una confirmación adicional de la naturaleza del producto del molde de LP154 traducido *in vitro*, se examinó el comportamiento de este producto en dos tipos diferentes de medios cromatográficos. La tiopropil (TP) sefarosa permite el aislamiento de un producto que contiene una cisteína libre (por ejemplo, el producto de LP154 que tiene un resto de cisteína adyacente al término C) (Figura 13). De forma similar, dT₂₅ agarosa permite el aislamiento de moldes que contienen una secuencia poli dA (por ejemplo, 30-P) (Figura 13). La Figura 14 demuestra que el aislamiento secuencial en TP sefarosa, seguido de dT₂₅ agarosa, produjo el mismo producto que el aislamiento en dT₂₅ agarosa solo. El hecho de que el producto de la traducción *in vitro* contenía tanto un tramo poli-A y un tiol libre indica fuertemente que el producto de traducción era la fusión deseada de ARN-péptido.

Los resultados anteriores son consistentes con la capacidad para sintetizar fusiones de ARNm-péptido, y para recuperarlas intactas a partir de extractos de traducción *in vitro*. Las porciones peptídicas de las fusiones así sintetizadas parece que tienen las secuencias pretendidas, según se demuestra mediante inmunoprecipitación y aislamiento usando técnicas cromatográficas apropiadas. Según los resultados presentados anteriormente, las reacciones son intramoleculares y ocurren de una manera dependiente del molde. Finalmente, incluso con una modificación del molde de menos de 1%, el presente sistema facilita selecciones basadas en complejidades candidatas de alrededor de 10^{13} moléculas.

Selección de la recuperación del epítipo c-myc. Para seleccionar epítopos de c-myc adicionales, se generó una gran librería de moldes de traducción (por ejemplo, 10^{15} miembros), que contiene una región aleatorizada (véase la Figura 7C, y a continuación). Esta librería se usó para generar $\sim 10^{12}$ - 10^{13} fusiones (como se describe en la presente memoria), que se tratan con el anticuerpo anti-c-myc (por ejemplo, mediante inmunoprecipitación o usando un anticuerpo inmovilizado en una columna u otro soporte sólido) para enriquecer los moldes que codifican c-myc en rondas repetidas de la selección *in vitro*.

Modelos para la formación de fusiones. Sin estar ligados a una teoría particular, se propone un modelo para el mecanismo de la formación de la fusión, en el que la traducción se inicia normalmente y el alargamiento transcurre hasta el extremo del marco de lectura abierta. Cuando el ribosoma alcanza la porción de ADN del molde, la traducción empieza a caer. En este punto, el complejo se puede dividir en dos destinos: la disociación del péptido naciente, o la transferencia del péptido naciente a la puomicina en el extremo 3' del molde. La eficacia de la región de transferencia probablemente está controlada por un número de factores que influyen la estabilidad del complejo de traducción detenido, y la entrada del resto de 3'-puomicina en el sitio A del centro de peptidiltransferasa. Después de la reacción de transferencia, probablemente la fusión de ARNm-péptido permanece complejada con el ribosoma, puesto que los factores de liberación conocidos no pueden hidrolizar el enlace amídico estable entre los dominios de ARN y peptídico.

Tanto el modelo clásico para el alargamiento (Watson, Bull. Soc. Chim. Biol. **46**:1399 (1964)) como el modelo de estado intermedio (Moazed y Noller, Nature **342**:142 (1989)) requieren que el sitio A esté vacío para la entrada de puomicina en el centro de peptidiltransferasa. Para que la puomicina entre en el sitio A vacío, el ligador debe formar un bucle alrededor del exterior del ribosoma, o debe pasar directamente desde el sitio descodificante a través del sitio A hasta el centro de peptidiltransferasa. Los datos descritos en la presente memoria no distinguen claramente entre estas alternativas, debido a que el ligador más corto ensayado (21 nucleótidos) es aún suficientemente largo para pasar alrededor del exterior del ribosoma. En algunos modelos de estructura ribosómica (Frank *et al.*, Nature **376**:441 (1995)), el ARNm se enrosca a través de un canal que se extiende a cada lado del sitio descodificante, en cuyo caso se requeriría el desenroscamiento del ligador a partir del canal para permitir que la puomicina alcanzase el centro de peptidiltransferasa a través del sitio A.

Parece que la transferencia del péptido naciente a la puomicina es lenta con relación al proceso de alargamiento, según se demuestra mediante la homogeneidad y longitud del péptido unido al ligador. Si la puomicina compite efectivamente con los aminoacil ARNt durante el alargamiento, sería de esperar que las fusiones de ligador-péptido presentes en los productos de fusión tuviesen un tamaño heterogéneo. Además, no parece que el ribosoma leyese en la región del ligador, según se indica mediante la similitud en las moviidades en gel entre la fusión de Met-molde y el ligador no modificado. dA_{3n} debería de codificar (lisina)_n, lo que ciertamente disminuiría la movilidad del ligador. La velocidad lenta de desenroscamiento del ARNm puede explicar la velocidad lenta de la formación de la fusión, con relación a la velocidad de translocación. Los resultados preliminares sugieren que la cantidad de producto de fusión formado aumenta notablemente tras la incubación post-traducción prolongada a baja temperatura, quizá debido al incremento de tiempo disponible para la transferencia del péptido naciente a la puomicina.

Materiales detallados y métodos

A continuación se describen los materiales detallados y los métodos que se refieren a la traducción *in vitro* y al estudio de las fusiones de ARN-proteína, incluyendo fusiones que tienen un marcador de epítipo de myc.

Secuencias. Se usó un número de oligonucleótidos como antes para la generación de las fusiones de ARN-proteína. Estos oligonucleótidos tienen las siguientes secuencias.

NOMBRE SECUENCIA

30-P 5'AAA AAA AAA AAA AAA AAA AAA AAA AAA CCP (SEQ ID NO:8)

13-P 5'AAA AAA AAA ACC P (SEQ ID NO: 9)

25-P 5'CGC GGT TTT TAT TTT TTT TTT TCC P (SEQ ID NO: 10)

43-P 5'rGrGrA rGrGrA rCrGrA rArArU rGAA AAA AAA AAA AAA AAA AAA AAA ACC P (SEQ ID NO: 11)

43-P [CUG] 5'rGrGrA rGrGrA rCrGrA rArCrU rGAA AAA AAA AAA AAA AAA AAA AAA ACC P (SEQ ID NO: 12)

ES 2 268 763 T3

40-P 5'rGrGrA rGrGrA rCrGrA rArCrU rGAA AAA AAA AAA AAA AAA
AAA AAA ACC P (SEQ ID NO: 13)

5

37-P 5'rGrGrA rGrGrA rCrGrA rArCrU rGAA AAA AAA AAA AAA AAA
AAA ACC P (SEQ ID NO: 14)

10

34-P 5'rGrGrA rGrGrA rCrGrA rArCrU rGAA AAA AAA AAA AAA AAA
ACC P (SEQ ID NO: 15)

15

31-P 5'rGrGrA rGrGrA rCrGrA rArCrU rGAA AAA AAA AAA AAA ACC P
(SEQ ID NO: 16)

20

LP77 5'rGrGrG rArGrG rArCrG rArArA rUrGrG rArArC rArGrA rArArC
rUrGrA rUrCrU rCrUrG rArArG rArArG rArCrC rUrGrA rArC AAA AAA AAA
AAA AAA AAA AAA AAA AAA CCP (SEQ ID NO: 1)

25

30

LP154 5'rGrGrG rArCrA rArUrU rArCrU rArUrU rUrArC rArArU rUrArC rA
rArUrG rGrCrU rGrArA rGrArA rCrArG rArArA rCrUrG rArUrC rUrCrU rGrArA
rGrArA rGrArC rCrUrG rCrUrG rCrGrU rArArA rCrGrU rCrGrU rGrArA rCrArG
rCrUrG rArArA rCrArC rArArA rCrUrG rGrArA rCrArG rCrUrG rCrGrU rArArC
rUrCrU rUrGrC rGrCrU AAA AAA AAA AAA AAA AAA AAA AAA AAA CCP
(SEQ ID NO: 3)

35

40

45

LP160 5' 5'rGrGrG rArCrA rArUrU rArCrU rArUrU rUrArC rArArU rUrArC rA
rArUrG rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS
rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS
rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rNrNrS rCrArG rCrUrG
rCrGrU rArArC rUrCrU rUrGrC rGrCrU AAA AAA AAA AAA AAA AAA AAA
AAA AAA CCP (SEQ ID NO: 17)

50

55

60

Todos los oligonucleótidos están enumerados en la dirección 5' a 3'. Las bases ribonucleotídicas están indicadas mediante la minúscula "r" antes del nombre del nucleótido; P es puromicina; rN indica cantidades iguales de rA, rG, rC y rU; rS indica cantidades iguales de rG y rC; y todas las otras denominaciones de las bases indican oligonucleótidos de ADN.

65

Productos químicos. La puromicina HCl, el vidrio de poro controlado alquilamínico de cadena larga, gougerotina, cloranfenicol, virginiamicina, DMAP, cloruro de dimetiltrifilo, y el anhídrido acético se obtuvieron de Sigma Chemical

ES 2 268 763 T3

(St. Louis, MO). La piridina, dimetilformamida, tolueno, anhídrido succínico, y para-nitrofenol se obtuvieron de Fluka Chemical (Ronkonkoma, NY). El ARNm beta-globina se obtuvo de Novagen (Madison, WI). El ARN TMV se obtuvo de Boehringer Mannheim (Indianapolis, IN).

5 *Enzimas.* La proteinasa K se obtuvo de Promega (Madison, WI). La ARNasa libre de ADNasa se produjo mediante el protocolo de Sambrook *et al.* (citado anteriormente), o se adquirió de Boehringer Mannheim. La T7 polimerasa se obtuvo mediante el protocolo publicado de Grodberg y Dunn (J. Bacteriol. **170**:1245 (1988)) con las modificaciones de Zawadzki y Gross (Nucl. Acids Res. **19**:1948 (1991)). La T4 ADN ligasa se obtuvo de New England Biolabs (Beverly, MA).

10 *Cuantificación de la incorporación del radiomarcador.* Para bandas de geles radioactivos, se determinó la cantidad de radiomarcador (³⁵S o ³²P) presente en cada banda mediante cuantificación en un analizador de transferencia Betagen 603 (Betagen, Waltham, MA), o usando placas de formación de fosfoimágenes (Molecular Dynamics, Sunnyvale, CA). Para muestras líquidas y sólidas, la cantidad de radiomarcador (³⁵S o ³²P) presente se determinó mediante recuento por centelleo (Beckman, Columbia, MD).

Imágenes de geles. Las imágenes de geles se obtuvieron mediante autorradiografía (usando una película Kodak XAR), o usando placas de formación de fosfoimágenes (Molecular Dynamics).

20 *Síntesis de CPG puromicina.* Más arriba se esquematizan los protocolos detallados para la síntesis de CPG-puromicina.

Reacciones enzimáticas. En general, la preparación de ácidos nucleicos para las reacciones de quinasa, de transcripción, mediante PCR, y de traducción, usando extractos de *E. coli*, fue la misma. Cada protocolo preparativo comenzó con la extracción usando un volumen igual de 1:1 de fenol/cloroformo, seguido de la centrifugación y aislamiento de la fase acuosa. Se añadieron acetato de sodio (pH 5,2) y espermidina hasta una concentración final de 300 mM y 1 mM, respectivamente, y la muestra se precipitó mediante adición de 3 volúmenes de etanol al 100% y mediante incubación a -70°C durante 20 minutos. Las muestras se centrifugaron a >12.000 g, se retiró el sobrenadante, y los peletes se lavaron con un exceso de etanol al 95%, a 0°C. Los peletes resultantes se secaron entonces a vacío, y se resuspendieron.

30 *Oligonucleótidos.* Tanto el ADN como el ARN sintéticos se sintetizaron en un sintetizador Millipore Expedite, usando una química estándar para cada uno, según se suministra por el fabricante (Milligen, Bedford, MA). Los oligonucleótidos que contienen 3' puromicina se sintetizaron usando columnas de CPG-puromicina empaquetadas con 30-50 mg de soporte sólido (~20 μmoles de puromicina/gramo). Los oligonucleótidos que contienen una 3' biotina se sintetizaron usando 1 μmol de columnas bioteg CPG de Glen Research (Sterling, VA). Los oligonucleótidos que contienen una 5' biotina se sintetizaron mediante adición de fosforamido de bioteg (Glen Research) como la base en 5'. Los oligonucleótidos a ligar en los extremos 3' de las moléculas de ARN se fosforilaron químicamente en el extremo 5' (usando un reactivo de fosforilación químico de Glen Research) antes de la desprotección, o se fosforilaron enzimáticamente usando ATP y T4 polinucleótido quinasa (New England Biolabs) después de la desprotección. Las muestras que contienen sólo ADN (y 3' puromicina o 3' biotina) se desprotegeron mediante adición de NH₄OH al 25%, seguido de la incubación durante 12 horas a 55°C. Las muestras que contienen monómeros de ARN (por ejemplo, 43-P) se desprotegeron mediante adición de etanol (25% (v/v)) a la disolución de NH₄OH, e incubación durante 12 horas a 55°C. El 2' OH se desprotegió usando 1 M de TBAF en THF (Sigma), durante 48 horas a temperatura ambiente. El TBAF se eliminó usando una columna NAP-25 Sephadex (Farmacia, Piscataway, NJ).

Las muestras de ADN y ARN desprotegidas se purificaron entonces usando PAGE desnaturalizante, seguido de empapamiento o electroelución a partir del gel usando un Elutrap (Schleicher y Schuell, Keene, NH) y desalando usando una columna NAP-25 Sephadex o precipitación con etanol, como se describe anteriormente.

50 *Construcción de ADN de myc.* Se construyeron dos moldes de ADN que contienen el marcador del epítipo de c-myc. El primer molde se obtuvo a partir de una combinación de los oligonucleótidos 64.27 (5'-GTT CAG GTC TTC TTG AGA GAT CAG TTT CTG TTC CAT TTC GTC CTC CCT ATA GTG AGT CGT ATT A-3') (SEC ID n°: 18) y 18.109 (5'-TAA TAC GAC TCA CTA TAG) (SEC ID n°: 19). La transcripción usando este molde produjo ARN 47.1 que codificó el péptido MEQKLISEEDLN (SEC ID n°: 20). La ligación de ARN 47.1 a 30-P produjo LP77, mostrado en la Figura 7A.

60 El segundo molde se obtuvo primero como un único oligonucleótido de 99 bases de longitud, que tiene la denominación RWR 99.6, y la secuencia 5'-AGC GCA AGA GTT ACG CAG CTG TTC CAG TTT GTG TTT CAG CTG TTC ACG ACG TTT ACG CAG CAG GTC TTC TTC AGA GAT CAG TTT CTG TTC TTC AGC CAT-3' (SEC ID n°: 21). Se construyeron mediante PCR moldes de transcripción bicatenarios que contienen esta secuencia con los oligos RWR 21.103 (5'-AGC GCA AGA GTT ACG CAG CTG-3') (SEC ID n°: 22) y RWR 63.26 (5'-TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG GCT GAA GAA CAG AAA CTG-3') (SEC ID n°: 23), según los protocolos publicados (Ausubel *et al.*, citado anteriormente, capítulo 15). La transcripción usando este molde produjo un ARN denominado como RNA124 que codificó el péptido MAEEQKLISEEDLLRKRREQLKHK LEQLRNSCA (SEC ID n°: 24). Este péptido contenía la secuencia usada para producir el anticuerpo monoclonal 9E10 cuando se conjugó a una proteína portadora (Oncogene Science Technical Bulletin). El RNA124 tenía 124 nucleótidos de longitud, y la ligación de RNA124 a 30-P produjo LP154 mostrado en la Figura 7B. La secuencia de RNA124 es la

ES 2 268 763 T3

siguiente (SEC ID n°: 32): 5'-rGrGrG rArCrA rArUrU rArCrU rArUrU rUrArC rArArU rUrArC rArArUrG rGrCrU rGrArA rGrArA rCrArG rArArA rCrUrG rArUrC rUrCrU rGrArA rGrArA rGrArC rCrUrG rCrUrG rCrGrU rArArA rCrGrU rCrGrU rGrArA rCrArG rCrUrG rArArA rCrArC rArArA rCrUrG rGrArA rCrArG rCrUrG rCrGrU rArArC rUrCrU rUrGrC rGrCrU-3'.

5 *Construcción del conjunto aleatorizado.* El conjunto aleatorizado se construyó como un único oligonucleótido de 130 bases de longitud, denominado RWR 130.1. Comenzando en el extremo 3', la secuencia fue 3'-CCCTGTAAAT GATAAATGTTAATGTTAC (NNS)₂₇ GTC GAC GCA TTG AGA TAC CGA-5' (SEC ID n°: 25). N representa una posición al azar, y esta secuencia se generó según el protocolo del sintetizador estándar. S representa una mezcla igual de bases de dG y dC. La PCR se realizó con los oligonucleótidos 42.108 (5'-TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA) (SEC ID n°: 26) y 21.103 (5'-AGC GCA AGA GTT ACG CAG CTG) (SEC ID n°: 27). La transcripción de este molde produjo un ARN denominado conjunto 130.1. La ligación del conjunto 130.1 a 30-P produjo el Conjunto n° 1 (también denominado como LP160), mostrado en la Figura 7C.

15 Se realizaron siete ciclos de PCR según los protocolos publicados (Ausubel *et al.*, citado anteriormente), con las siguientes excepciones: (i) la concentración de partida de RWR 130.1 fue 30 nanomolar, (ii) cada cebador se usó a una concentración de 1,5 μM , (iii) la concentración de dNTP fue 400 μM para cada base, y (iv) se usó la Taq polimerasa (Boehringer Mannheim) a 5 unidades por 100 μl . El producto bicatenario se purificó en PAGE no desnaturizante, y se aisló mediante electroelución. La cantidad de ADN se determinó tanto mediante absorbancia de UV a 260 nm como mediante comparación de la fluorescencia con bromuro de etidio con patrones conocidos.

20 *Síntesis enzimática de ARN.* Las reacciones de transcripción a partir de ADN de PCR bicatenario y oligonucleótidos sintéticos se llevó a cabo como se describe previamente (Milligan y Uhlenbeck, *Meth. Enzymol.* **180**:51 (1989)). El ARN de longitud completa se purificó mediante PAGE desnaturizante, se electroeluyó, y se desaló como se describe anteriormente. La concentración de ARN del conjunto se estimó usando un coeficiente de extinción de 1300 O.D./ μmol ; RNA124, 1250 O.D./ μmol ; RNA 47.1, 480 O.D./ μmol . La transcripción a partir del ADN del conjunto bicatenario produjo ~ 90 nanomoles de ARN del conjunto.

30 *Síntesis enzimática de conjugados de ARN-puromicina.* La ligación de las secuencias de ARN mensajero de myc y del conjunto al oligonucleótido que contiene puromicina se realizó usando una tablilla de ADN, denominada 19.35 (5'-TTT TTT TTT TAG CGC AAG A) (SEC ID n°: 28), usando un procedimiento análogo al descrito por Moore y Sharp (*Science* 250:992 (1992)). La reacción constó de ARNm, la tablilla, y oligonucleótido que contiene puromicina (30-P, dA27dCdCP), en una relación en moles de 0,8:0,9:1,0, y 1-2,5 unidades de ADN ligasa por picomol de ARNm del conjunto. Las reacciones se realizaron durante una hora a temperatura ambiente. Para la construcción de las fusiones de ARN del conjunto, la concentración de ARNm fue $\sim 6,6$ μmolar . Después de la ligación, el conjugado de ARN-puromicina se preparó como se describe anteriormente para reacciones enzimáticas. El precipitado se resuspendió, y las fusiones de longitud completa se purificaron en PAGE desnaturizante, y se aislaron mediante electroelución como se describe anteriormente. La concentración de ARN del conjunto se estimó usando un coeficiente de extinción de 1650 O.D./ μmol , y el molde de myc de 1600 O.D./ μmol . De esta manera, se generaron 2,5 nanomoles de conjugado.

40 *Preparación de dT₂₅ estreptavidina agarosa.* Se incubó dT₂₅ que contiene una 3' biotina (sintetizada en columnas de fosforamido de bioteg (Glen Research)), a 1-10 μM , con una suspensión de estreptavidina-agarosa (agarosa al 50% en volumen, Pierce, Rockford, IL), durante 1 hora a temperatura ambiente en TE (10 mM de Tris-cloruro pH 8,2, 1 mM de EDTA), y se lavó. Después, se estimó ópticamente la capacidad de unión de la agarosa mediante la desaparición de biotina-dT₂₅ de la disolución, y/o mediante titulación de la resina con cantidades conocidas de oligonucleótido complementario.

50 *Reacciones de traducción usando extractos y ribosomas derivados de E.coli.* En general, las reacciones de traducción se realizaron con kits comprados (por ejemplo, *E.coli* S30 Extract for Linear Templates, Promega, Madison, WI). Sin embargo, también se usó *E.coli* MRE600 (obtenido del ATCC, Rockville, MD) para generar extractos de S30 preparados según los protocolos publicados (por ejemplo, Ellman *et al.*, *Meth. Enzymol.* **202**:301 (1991)), así como una fracción ribosómica preparada como se describe por Kuclicki *et al.* (*Anal. Biochem.* **206**:389 (1992)). La reacción estándar se realizó en un volumen de 50 μl con 20-40 μCi de ³⁵S-metionina como marcador. La mezcla de reacción consistió en 30% de extracto v/v, 9-18 mM de MgCl₂, 40% de premezcla menos metionina (Promega) v/v, y 5 μM de molde (por ejemplo, 43-P). Para los experimentos de coincubación, se añadieron los oligos 13-P y 25-P, a una concentración de 5 μM . Para experimentos que usen ribosomas, se añadieron 3 μl de disolución ribosómica por reacción, en lugar del lisado. Todas las reacciones se incubaron a 37°C durante 30 minutos. Los moldes se purificaron como se describe anteriormente en condiciones enzimáticas.

60 *Reacciones de traducción con germen de trigo.* Las reacciones de traducción en la Figura 8 se realizaron usando kits comprados que carecen de metionina (Promega), según las recomendaciones del fabricante. Las concentraciones del molde fueron 4 μM para 43-P, y 0,8 μM para LP77 y LP154. Las reacciones se llevaron a cabo a 25°C con 30 μCi de ³⁵S-metionina en un volumen total de 25 μl .

65 *Reacciones de traducción con reticulocitos.* Las reacciones de traducción se realizaron con kits comprados (Novagen, Madison, WI), o usando un extracto preparado según protocolos publicados (Jackson y Hunt, *Meth. Enzymol.* 96:50 (1983)). Se obtuvo sangre rica en reticulocitos a partir de Pel-Freez Biologicals (Rogers, AK). En ambos casos, las condiciones de reacción fueron aquellas recomendadas para uso con Red Nova Lysate (Novagen). Las reacciones

ES 2 268 763 T3

consistían en 100 mM de KCl, 0,5 mM de MgOAc, 2 mM de DTT, 20 mM de HEPES pH 7,6, 8 mM de fosfato de creatina, 25 μ M en cada aminoácido (con la excepción de metionina si se usó 35 S-Met), y 40% v/v de lisado. La incubación se realizó a 30°C durante 1 hora. Las concentraciones del molde dependen del experimento, pero generalmente oscilaron desde 50 nM hasta 1 μ M, con la excepción de 43-P (Figura 6H), que fue de 4 μ M.

Para la generación del conjunto aleatorizado, se llevaron a cabo 10 ml de una reacción de traducción a una concentración de molde de \sim 0,1 μ M (1,25 nanomoles de molde). Además, el molde marcado con 32 P se incluyó en la reacción para permitir la determinación de la cantidad de material presente en cada etapa del procedimiento de purificación y de selección. Después de la traducción a 30°C durante una hora, la reacción se enfrió en hielo durante 30-60 minutos.

Aislamiento de la fusión con dT₂₅ estreptavidina agarosa. Después de la incubación, la reacción de traducción se diluyó aproximadamente 150 veces en tampón de aislamiento (1,0 M de NaCl, 0,1 M de Tris cloruro pH 8,2, 10 mM de EDTA, 1 mM de DTT) que contiene más de un exceso molar de 10 X de dT₂₅-biotina-estreptavidina agarosa, cuya concentración de dT₂₅ fue \sim 10 μ M (volumen de la suspensión igual o mayor que el volumen del lisado), y se incubó con agitación a 4°C durante una hora. Después, la agarosa se eliminó de la mezcla mediante filtración (filtros MC ultralibres de Millipore), o mediante centrifugación, y se lavó con tampón de aislamiento frío, 2-4 veces. El molde se liberó entonces de la dT₂₅ estreptavidina agarosa mediante lavado repetido con alícuotas de 50-100 μ l de 15 mM de NaOH, 1 mM de EDTA. El eluyente se neutralizó inmediatamente en 3 M de NaOAc pH 5,2, 10 mM de espermidina, y se precipitó mediante etanol. Para la reacción del conjunto, la radioactividad total recuperada indicó aproximadamente que se recuperó un 50-70% del molde de entrada.

Aislamiento de la fusión con tiopropilsefarosa. Las fusiones que contienen cisteína se pueden purificar usando tiopropilsefarosa 6B, como en la Figura 13 (Farmacia). En los experimentos descritos en la presente memoria, el aislamiento se llevó a cabo directamente a partir de la reacción de traducción, o tras el aislamiento inicial de la fusión (por ejemplo, con estreptavidina agarosa). Para las muestras purificadas directamente, se usó una relación de 1:10 (v/v) de lisado a sefarosa. Para el conjunto, se usaron 0,5 ml de suspensión de sefarosa para aislar todo el material de fusión a partir de 5 ml de la mezcla de reacción. Las muestras se diluyeron en una suspensión 50:50 (v/v) de tiopropilsefarosa en 1 X TE 8.2 (10 mM de Tris-Cl, 1 mM de EDTA, pH 8,2) que contiene ARNasa libre de ADNasa (Boehringer Mannheim), y se incubó con rotación durante 1-2 horas a 4°C para terminar la reacción. Se eliminó el líquido en exceso, y la sefarosa se lavó repetidamente con tampón de aislamiento que contiene 20 mM de DTT, y se recuperó mediante centrifugación o mediante filtración. Las fusiones se eluyeron de la sefarosa usando una disolución de 25-30 mM de ditiotreitol (DTT) en 10 mM de Tris cloruro pH 8,2, 1 mM de EDTA. Después, la fusión se concentró mediante combinación de evaporación a alto vacío y precipitación con etanol como se describe anteriormente. Para la reacción del conjunto, la radioactividad total recuperada indicó que aproximadamente el 1% del molde se convirtió en la fusión.

Para ciertas aplicaciones, se añadió dT₂₅ a este eluato, y se hizo girar durante 1 hora a 4°C. La agarosa se enjuagó tres veces con tampón de aislamiento frío, se aisló vía filtración, y el material unido se eluyó como antes. Se añadió ARNt portador, y el producto de fusión se precipitó mediante etanol. La muestra se resuspendió en TE pH 8,2 que contiene ARNasa A libre de ADNasa, para eliminar la porción de ARN del molde.

Reacciones de inmunoprecipitación. Se realizaron inmunoprecipitaciones de péptidos a partir de reacciones de traducción (Figura 10) mezclando 4 μ l de reacción de traducción de reticulocitos, 2 μ l de suero de ratón normal, y 20 μ l de proteína G + A agarosa (Calbiochem, La Jolla, CA) con 200 μ l de PBS (58 mM de Na₂HPO₄, 17 mM de NaH₂PO₄, 68 mM de NaCl), tampón de dilución (10 mM de Tris cloruro pH 8,2, 140 mM de NaCl, 1% v/v de Tritón X-100), o PBSTDS (PBS + 1% de Tritón X-100, 0,5% de desoxicolato y 0,1% de SDS). Las muestras se hicieron girar entonces durante una hora a 4°C, seguido de la centrifugación a 2500 rpm durante 15 minutos. Se eliminó el eluyente, y se añadieron 10 μ l de anticuerpo monoclonal c-myc 9E10 (Calbiochem, La Jolla, CA) y 15 μ l de proteína G + A agarosa, y se hizo girar durante 2 horas a 4°C. Después, las muestras se lavaron con dos volúmenes de 1 ml de PBS, tampón de dilución, o PBSTDS. Se añadieron a la mezcla 40 μ l de tampón de carga de gel (Calbiochem Product Bulletin), y se cargaron 20 μ l en un PAGE desnaturizante como se describe por Schagger y von Jagow (Anal. Biochem. **166**:368 (1987)).

Las inmunoprecipitaciones de las fusiones (como se indican en la Figura 11) se realizaron mezclando 8 μ l de reacción de traducción de reticulocitos con 300 μ l de tampón de dilución (10 mM de Tris cloruro pH 8,2, 140 mM de NaCl, 1% v/v de Tritón X-100), 15 μ l de proteína G sefarosa (Sigma), y 10 μ l (1 μ g) de anticuerpo c-myc 9E10 (Calbiochem), seguido de la rotación durante varias horas a 4°C. Después del aislamiento, las muestras se lavaron, se trataron con ARNasa A libre de ADNasa, se marcaron con polinucleótido quinasa y con 32 P gamma ATP, y se separaron mediante urea-PAGE desnaturizante (Figura 11).

Transcripción inversa del conjunto de fusión. Se llevaron a cabo reacciones de transcripción inversa según las recomendaciones del fabricante para Superscript II, excepto que el molde, el agua y el cebador se incubaron a 70°C durante sólo dos minutos (Gibco BRL, Grand Island, NY). Para monitorizar la extensión, se incluyeron 50 μ Ci de alfa 32 P dCTP en algunas reacciones; en otras reacciones, la transcripción inversa se monitorizó usando cebadores marcados con 32 P en 5', que se prepararon usando 32 P α ATP (New England Nuclear, Boston, MA) y T4 polinucleótido quinasa (New England Biolabs, Beverly, MA).

ES 2 268 763 T3

Preparación de proteína G y anticuerpo sefarosa. Se lavaron dos alícuotas de 50 μ l de suspensión de proteína G sefarosa (50% de sólido en volumen) (Sigma) con tampón de dilución (10 mM de Tris cloruro pH 8,2, 140 mM de NaCl, 0,025% de NaN₃, 1% v/v de Tritón X-100), y se aislaron mediante centrifugación. La primera alícuota se reservó para uso como una precolumna antes de la matriz de selección. Tras la resuspensión de la segunda alícuota en tampón de dilución, se añadieron 40 μ g de anticuerpo monoclonal AB-1 c-myc (Oncogene Science), y la reacción se incubó toda la noche a 4°C con rotación. El anticuerpo sefarosa se purificó entonces mediante centrifugación durante 15 minutos a 1500-2500 rpm en una microcentrífuga, y se lavó 1-2 veces con tampón de dilución.

Selección. Tras el aislamiento de la fusión, y la síntesis de la hebra complementaria, la reacción de transcriptasa inversa completa se usó directamente en el proceso de selección. Se perfilan aquí dos protocolos. Para la ronda 1, la reacción de transcriptasa inversa se añadió directamente al anticuerpo sefarosa preparado como se describe anteriormente, y se incubó 2 horas. Para rondas subsiguientes, la reacción se incubó ~2 horas con proteína G sefarosa lavada, antes de la columna de anticuerpo, para disminuir el número de aglutinantes que interactúan con la proteína G en lugar de con el anticuerpo inmovilizado.

Para eluir el conjunto de la matriz, se pueden tomar varios enfoques. El primero es lavar la matriz de selección con ácido acético al 4%. Este procedimiento libera el péptido de la matriz. Como alternativa, se puede usar un lavado más restrictivo (por ejemplo, usando urea u otro desnaturalizante) en lugar o además del enfoque con ácido acético.

PCR de fusiones seleccionadas. Las moléculas seleccionadas se amplifican mediante PCR usando protocolos estándares como se describe anteriormente para la construcción del conjunto.

Síntesis y ensayo de fusiones de beta-globina

Para sintetizar un constructo de fusión de β -globina, se generó un ADNc de β -globina a partir de 2,5 μ g de ARNm de globina mediante transcripción inversa con 200 pmoles del cebador 18.155 (5'-GTG GTA TTT GTG AGC CAG) (SEC ID n°: 29) y transcriptasa inversa Superscript (Gibco BRL), según el protocolo del fabricante. La secuencia del cebador fue complementaria a los 18 nucleótidos de la beta-globina 5' del codón de parada. Para añadir un promotor T7, se eliminaron 20 μ l de la reacción de transcripción inversa, y se sometieron a 6 ciclos de PCR con los cebadores 18.155 y 40.54 (5'-TAA TAC GAC TCA CTA TAG GGA CAC TTG CTT TTG ACA CAA C) (SEC ID n°: 30). El ARNm de la "syn- β -globina" resultante se generó entonces mediante transcripción de desempate con T7 según Milligan y Uhlenbeck (Methods Enzymol. **180**:51 (1989)), y el gel de ARN se purificó, se electroeluyó y se desaló como se describe. Después, se generó la "LP- β -globina" a partir del constructo de syn- β -globina, mediante ligación de ese constructo a 30-P según el método de Moore y Sharp (Science **256**:992 (1992)) usando el cebador 20.262 (5'-TTTT TTT TTT T GTG GTA TTT G) (SEC ID n°: 31) como la tablilla. El producto de la reacción de ligación se purificó entonces en gel, se electroeluyó y se desaló como antes. La concentración del producto final se determinó mediante absorbancia a 260 nm.

Estos moldes de β -globina se tradujeron entonces *in vitro* como se describe en la Tabla 1 en un volumen total de 25 μ l cada uno. Se añadió Mg²⁺ a partir de una disolución madre de 25 mM. Todas las reacciones se incubaron a 30°C durante una hora, y se colocaron a -20°C toda la noche. Los CPM precipitables mediante dT₂₅ se determinaron entonces dos veces usando 6 μ l de lisado, y se promediaron restando el fondo.

TABLA 1

Reacciones de traducción con moldes de beta-globina

Reacción	Molde	Mg ²⁺ (mM)	³⁵ S Met (μ l)	TCA CPM (2 μ l)	dT ₂₅ CPM (6 μ l)
1	---	1,0	2,0 (20 μ Ci)	3312	0
2	2,5 μ g syn- β -globina	0,5	2,0 (20 μ Ci)	33860	36
3	2,5 μ g syn- β -globina	1,0	2,0 (20 μ Ci)	22470	82
4	2,5 μ g syn- β -globina	2,0	2,0 (20 μ Ci)	15696	86
5	2,5 μ g LP- β -globina	0,5	2,0 (20 μ Ci)	32712	218
6	2,5 μ g LP- β -globina	1,0	2,0 (20 μ Ci)	24226	402
7	2,5 μ g LP- β -globina	2,0	2,0 (20 μ Ci)	15074	270

Para preparar las muestras para el análisis de gel, se mezclaron 6 μ l de cada reacción de traducción con 1000 μ l de tampón de aislamiento (1 M de NaCl, 100 mM de Tris-Cl pH 8,2, 10 mM de EDTA, 0,1 mM de DTT), 1 μ l de ARNasa A (libre de ADNasa, Boehringer Mannheim), y 20 μ l de 20 μ M de dT₂₅ estreptavidina agarosa. Las muestras

se incubaron a 4°C durante una hora con rotación. Se eliminó el exceso de tampón de aislamiento, y las muestras se añadieron al filtro de MC de Millipore para eliminar cualquier tampón de aislamiento que quede. Las muestras se lavaron entonces cuatro veces con 50 μ l de H₂O, y dos veces con 50 μ l de 15 mM de NaOH, 1 mM de EDTA. La muestra (300 μ l) se neutralizó con 100 μ l de TE pH 6,8 (100 mM de Tris-Cl, 1 mM de EDTA), se añadió 1 μ l de 1 mg/ml de ARNasa A (como antes), y las muestras se incubaron a 37°C. Después se añadieron 10 μ l de 2 X de tampón de carga de SDS (125 mM de Tris-Cl pH 6,8, 2% de SDS, 2% de β -mercaptoetanol, 20% de glicerina, 0,001% de azul de bromofenol), y la muestra se liofilizó hasta sequedad y se resuspendió en 20 μ l de H₂O y 1% de β -mercaptoetanol. Las muestras se cargaron entonces sobre un gel que resuelve los péptidos, como se describe por Schagger y von Jagow (Analytical Biochemistry 166:368 (1987)), y se visualizaron mediante autorradiografía.

Los resultados de estos experimentos se muestran en las Figuras 15A y 15B. Según se indica en la Figura 15A, y 15B. Según se indica en la Figura la ³⁵S-metionina se incorporó en la parte proteínica de la syn- β -globina y en las fusiones de LP- β -globina. La proteína fue heterogénea, pero una banda fuerte mostró la movilidad esperada para el ARNm de β -globina. También, como se muestra en la Figura 15B, tras el aislamiento con dT₂₅ y la digestión con ARNasa A, no quedó ningún material marcado con ³⁵S en las líneas de syn- β -globina (Figura 15B, líneas 2-4). Por el contrario, en las líneas de LP- β -globina, se observó un producto marcado con ³⁵S de tamaño homogéneo.

Estos resultados indicaron que, como anteriormente, se aisló un producto de fusión mediante cromatografía de afinidad oligonucleotídica sólo cuando el molde contenía una 3' puomicina. Esto se confirmó mediante recuento por centelleo (véase la Tabla 1). Es de esperar que el material obtenido contenga el ligador de 30-P fusionado a alguna parte de la β -globina. El producto de fusión parece bastante homogéneo en tamaño, según se juzga mediante análisis de gel. Sin embargo, puesto que el producto mostró una movilidad muy similar a la β -globina natural (Figuras 15A y 15B, líneas de control), fue difícil determinar la longitud precisa de la porción proteínica del producto de fusión.

Optimización adicional de la formación de la fusión de ARN-proteína

Se ha encontrado que ciertos factores aumentan adicionalmente la eficacia de la formación de las fusiones de ARN-péptido. La formación de la fusión, es decir, la transferencia de la cadena peptídica naciente desde su ARNt al resto de puomicina en el extremo 3' del ARNm, es una reacción lenta que sigue a la traducción inicial, relativamente rápida, del marco de lectura abierto, para generar el péptido naciente. El grado de formación de fusión se puede potenciar sustancialmente mediante una incubación tras la traducción, en condiciones de Mg²⁺ elevado (preferentemente, en un intervalo de 50-100 mM), y/o mediante el uso de un ligador más flexible entre el ARNm y el resto de puomicina. Además, las incubaciones largas (12-48 horas) a temperaturas bajas (preferentemente, -20°C) también dan como resultado rendimientos crecientes de fusiones, con menor degradación del ARNm que la que se produce durante la incubación a 30°C. Combinando estos factores, se puede convertir hasta el 40% del ARNm de entrada a los productos de fusión de ARNm-péptido, según se muestra a continuación.

Síntesis de conjugados de ARNm-puomicina. En estos experimentos de optimización, se ligaron oligonucleótidos ligadores que contienen puomicina a los extremos 3' de los ARNm, usando T4 ADN ligasa de bacteriófago en presencia de tablillas de ADN complementarias, generalmente como se describe anteriormente. Puesto que la T4 ADN ligasa prefiere un emparejamiento de bases precisa cerca de la unión de la ligación, y los productos de la transcripción de empate con T7, T3 o SP6 ARN polimerasa a menudo son heterogéneos en sus extremos 3' (Nucleic Acids Research 15:8783 (1987)), sólo aquellos ARN que contienen el nucleótido 3'-terminal correcto se ligaron eficientemente. Cuando se usó una tablilla estándar de ADN, se ligaron aproximadamente el 40% de los productos de transcripción de empate al oligo de puomicina. La cantidad de producto de ligación se aumentó usando un exceso de ARN, pero no se aumentó usando un exceso de oligonucleótido que contiene puomicina. Sin estar atados a una teoría particular, parece que el factor limitante para la ligación fue la cantidad de ARN que fue totalmente complementaria a la región correspondiente de la tablilla de ADN.

Para permitir la ligación de esos transcritos que terminan con un nucleótido no moldeado extra en el término 3' (denominados "productos N + 1"), se usó una mezcla de la tablilla de ADN estándar con una nueva tablilla de ADN que contiene una base al azar adicional en la unión de ligación. La eficacia de la ligación aumentó en más del 70% para un molde de ARN de myc ejemplar (esto es, RNA124) en presencia de tal tablilla de ADN mixta.

Además de este enfoque de tablilla de ADN modificado, se optimizó adicionalmente la eficacia de la formación del conjugado de ARNm-puomicina teniendo en cuenta los siguientes tres factores. En primer lugar, se diseñaron o utilizaron preferentemente ARNm que carecían de términos en 3' que tienen cualquier estructura secundaria estable significativa que podría interferir con la hibridación al oligonucleótido de la tablilla. Además, debido a que la elevada concentración de sal provocó algunas veces fallo de la reacción de ligación, se incluyó preferentemente, como una etapa en el procedimiento, una desalación a conciencia de los oligonucleótidos usando una columna de NAP-25. Finalmente, debido a que la reacción de ligación fue relativamente rápida y estaba generalmente completa en los 40 minutos a temperatura ambiente, no se utilizaron generalmente períodos de incubación significativamente más largos, y a menudo dio como resultado la degradación innecesaria del ARN.

Usando las condiciones anteriores, se sintetizaron los conjugados de ARNm-puomicina según lo siguiente. La ligación de la secuencia de ARN myc (RNA124) al oligonucleótido que contiene puomicina se realizó usando una tablilla de ADN estándar (por ejemplo, 5'-TTTTTTTTTTAGCGCAAGA), o una tablilla que contiene una base al azar (N) en la unión de ligación (por ejemplo, 5'-TTTTTTTTTTNAGCGCAAGA). Las reacciones consistieron en ARNm,

ES 2 268 763 T3

la tablilla de ADN, y el oligonucleótido que contiene puromicina, en una relación molar de 1,0:1,5-2,0:1,0. En primer lugar, se calentó a 94°C durante 1 minuto una mezcla de estos componentes, y después se enfrió en hielo durante 15 minutos. Las reacciones de ligación se realizaron durante una hora a temperatura ambiente en 50 mM de Tris-HCl (pH 7,5), 10 mM de MgCl₂, 10 mM de DTT, 1 mM de ATP, 25 µg/ml de BSA, 15 µM de oligo de puromicina, 15 µM de ARNm, 22,5-30 µM de tablilla de ADN, inhibidor de ARNsin (Promega) a 1 U/µl, y 1,6 unidades de T4 ADN ligasa por picomol de oligo de puromicina. Tras la incubación, se añadió EDTA hasta una concentración final de 30 mM, y las mezclas de reacción se extrajeron con fenol/cloroformo. Los conjugados de longitud completa se purificaron mediante PAGE desnaturalizante, y se aislaron mediante electroelución.

Condiciones de traducción reticulocítica general. Además de mejorar la síntesis del conjugado de ARNm-puromicina, las reacciones de traducción también se optimizaron adicionalmente según lo siguiente. Las reacciones se realizaron en lisados de reticulocitos de conejo a partir de diferentes fuentes comerciales (Novagen, Madison, WI; Amersham, Arlington Heights, IL; Boehringer Mannheim, Indianapolis, IN; Ambion, Austin, TX; y Promega, Madison, WI). Una mezcla de reacción típica (25 µl de volumen final) constó de 20 mM de HEPES pH 7,6, 2 mM de DTT, 8 mM de fosfato de creatina, 100 mM de KCl, 0,75 mM de Mg(OAc)₂, 1 mM de ATP, 0,2 mM de GTP, 25 µM de cada aminoácido (0,7 µM de metionina si se usó ³⁵S-Met), ARNsin a 1 U/µl, y 60% (v/v) de lisado. La concentración final del molde estuvo en el intervalo de 50 nM a 800 nM. Para cada incubación, todos los componentes excepto el lisado se mezclaron cuidadosamente en hielo, y el lisado congelado se descongeló inmediatamente antes del uso. Tras la adición del lisado, la mezcla de reacción se mezcló a conciencia pipeteando suavemente, y se incubó a 30°C para comenzar la traducción. Las concentraciones óptimas de Mg²⁺ y K⁺ variaron en los intervalos de 0,25 mM-2 mM y 75 mM-200 mM, respectivamente, para diferentes ARNm, y se determinaron preferentemente en experimentos preliminares. Particularmente para ARNm pobremente traducidos, también se optimizaron algunas veces la concentración de hemina, de fosfato de creatina, de ARNt, y de aminoácidos. Generalmente se prefirió cloruro de potasio sobre el acetato de potasio para reacciones de fusión, pero algunas veces una mezcla de KCl y KOAc produjo mejores resultados.

Después de la traducción a 30°C durante 30 hasta 90 minutos, la reacción se enfrió en hielo durante 40 minutos, y se añadió Mg²⁺. La concentración final de Mg²⁺ añadida en esta etapa también se optimizó para diferentes moldes de ARNm, pero generalmente estuvo en el intervalo de 50 mM hasta 100 mM (usándose preferentemente 50 mM para conjuntos de moldes mixtos). La mezcla resultante se incubó a -20°C durante 16 hasta 48 horas. Para visualizar los productos de fusión marcados, se mezclaron 2 µl de la mezcla de reacción con 4 µl de tampón de carga, y la mezcla se calentó a 75°C durante 3 minutos. La mezcla resultante se cargó entonces sobre un gel de poli(acrilamida)-SDS con glicina al 6% (para moldes marcados radioactivamente con ³²P), o un gel de SDS-poli(acrilamida) con 8% de tricina (para moldes marcados radioactivamente con ³⁵S-Met). Como una alternativa a este enfoque, los productos de fusión también se pueden aislar usando dT₂₅ estreptavidina agarosa o tiopropilsefariosa (o ambas), generalmente como se describe en la presente memoria.

Para eliminar la porción de ARN del conjugado de ARN-ligador-puromicina-péptido, para el análisis subsiguiente mediante SDS-PAGE, se añadió una cantidad apropiada de EDTA tras la incubación post-traducciona, y la mezcla de reacción se desaló usando una columna microcon-10 (o microcon-30). Se mezclaron 2 µl de la mezcla resultante (aproximadamente 25 µl total) con 18 µl de tampón de ARNasa H (30 mM de Tris-HCl, pH 7,8, 30 mM de (NH₄)₂SO₄, 8 mM de MgCl₂, 1,5 mM de β-mercaptoetanol, y una cantidad apropiada de tablilla de ADN complementaria), y la mezcla se incubó a 4°C durante 45 minutos. Después se añadió ARNasa H, y la digestión se realizó a 37°C durante 20 minutos.

Calidad del oligo de puromicina. La calidad del oligonucleótido de puromicina también fue importante para la generación eficaz de productos de fusión. El acoplamiento de 5'-DMT, 2'-succinil, N-trifluoroacetilpuromicina con CPG no fue tan eficaz como el acoplamiento de los nucleótidos estándares. Como tal, la reacción de acoplamiento se monitorizó cuidadosamente para evitar la formación de CPG con una concentración demasiado baja de puromicina acoplada, y los aminoácidos sin reaccionar en el CPG se paralizaron completamente para evitar la síntesis subsiguiente de oligonucleótidos que carecen de una puromicina 3'-terminal. También fue importante evitar el uso de CPG que contiene partículas de malla muy fina, puesto que éstas fueron capaces de provocar problemas con la obturación de válvulas durante las etapas de síntesis automatizada subsiguientes de los nucleótidos.

Además, el oligo de puromicina sintetizado se evaluó preferentemente antes del uso a gran escala para asegurar la presencia de puromicina en el extremo 3'. En los experimentos, no se detectó fusión si se sustituye la puromicina con una desoxiadenosina que contiene un grupo amino primario en el extremo 3'. Para estudiar la presencia de grupos 3'-hidroxilo (es decir, la síntesis indeseada de oligos que carecen de una puromicina 3'-terminal), se radiomarcó en primer lugar el oligo de puromicina (por ejemplo, mediante 5'-fosforilación), y después se usó como un cebador para la extensión con desoxinucleotidiltransferasa terminal. En presencia de un resto de puromicina 3'-terminal, no se debería de observar producto de extensión.

Transcurso de tiempo de la traducción y de la incubación post-traducciona. La reacción de traducción fue relativamente rápida y se completó generalmente en 25 minutos a 30°C. Sin embargo, la reacción de fusión fue más lenta. Cuando se usó un ligador estándar (dA₂₇dCdCP) a 30°C, la síntesis de la fusión alcanzó su máximo nivel en 45 minutos adicionales. La incubación post-traducciona se podría llevar a cabo a temperaturas menores, por ejemplo a temperatura ambiente, 0°C, o -20°C. Se observó una menor degradación del molde de ARNm a -20°C, y los mejores resultados de la fusión se obtuvieron después de la incubación a -20°C durante 2 días.

ES 2 268 763 T3

El efecto de la concentración de Mg^{2+} . Una concentración elevada de Mg^{2+} en la incubación post-traduccional estimuló enormemente la formación de la fusión. Por ejemplo, para el molde de ARN de myc descrito anteriormente, se observó una estimulación de 3-4 veces en la formación de la fusión usando un ligador estándar ($dA_{27}dCdCP$) en presencia de 50 mM de Mg^{2+} durante la incubación de 16 horas a $-20^{\circ}C$ (Figura 17, compárense las líneas 3 y 4).
5 De forma similar, también se observó una formación eficaz de la fusión usando una incubación post-traduccional en presencia de una concentración de Mg^{2+} de 50-100 mM cuando las reacciones se llevaron a cabo a temperatura ambiente durante 30-45 minutos.

Longitud y secuencia del ligador. También se examinó la dependencia de la reacción de fusión con respecto a la longitud del ligador. En el intervalo entre 21 y 30 nucleótidos ($n = 18-27$), se observó poco cambio en la eficiencia de la reacción de fusión (como se describió anteriormente). Los ligadores más cortos (por ejemplo, 13 nucleótidos de longitud) dieron como resultado una menor fusión. Además, aunque los ligadores particulares de mayor longitud (esto es, de 45 nucleótidos y 54 nucleótidos) también dieron en cierto modo menores eficacias de la fusión, parece probable que también se podrían usar ligadores aún más largos para optimizar la eficacia de la reacción de fusión.
15

Con respecto a la secuencia del ligador, la sustitución de los restos de desoxirribonucleótido cerca del extremo 3' por restos de ribonucleótido no cambia significativamente la eficacia de la fusión. La secuencia de $dCdCP$ (o $rCrCP$) en el extremo 3' del ligador fue, sin embargo, importante para la formación de la fusión. La sustitución de $dCdCP$ por $dUdUP$ redujo la eficacia de la formación de la fusión significativamente.
20

Flexibilidad del ligador. También se estudió la dependencia de la reacción de fusión con respecto a la flexibilidad del ligador. En estos experimentos, se determinó que la eficacia de la fusión fue baja si se incrementaba la rigidez del ligador mediante hibridación con un oligonucleótido complementario próximo al extremo 3'. De forma similar, cuando se usó un ligador más flexible (por ejemplo, $dA_{21}C_9C_9C_9dAdCdCP$, en el que C_9 representa $HO(CH_2CH_2O)_3PO_2$), se mejoró significativamente la eficacia de la fusión. En comparación con el ligador estándar ($dA_{27}dCdCP$), el uso del ligador más flexible ($dA_{21}C_9C_9C_9dAdCdCP$) mejoró la eficacia de la fusión para RNA124 en más de 4 veces (Figura 17, compárense las líneas 1 y 9). Además, en contraste con el molde, con el ligador estándar cuya fusión post-traducción transcurrió pobremente en ausencia de una concentración elevada de Mg^{2+} (Figura 17, línea 3 y 4), el molde con el ligador flexible no requirió una concentración elevada de Mg^{2+} para producir un buen rendimiento del producto de fusión en una incubación post-traduccional prolongada a $-20^{\circ}C$ (Figura 17, compárense las líneas 11 y 12). Por lo tanto, este ligador fue muy útil si no se desearon adiciones post-traduccionales de concentraciones elevadas de Mg^{2+} . Además, el ligador flexible también produjo rendimientos óptimos de la fusión en presencia de Mg^{2+} elevada.
30

Cuantificación de la eficacia de la fusión. La eficacia de la fusión se puede expresar como la fracción de péptido traducido convertido a producto de fusión, o la fracción de molde de entrada convertido a producto de fusión. Para determinar la fracción de péptido traducido convertido a producto de fusión, se utilizó el marcaje con ^{35}S -Met del péptido traducido. En estos experimentos, cuando se usó un ligador de $dA_{27}dCdCP$ o $dA_{27}rCrCP$, alrededor del 3,5% del péptido traducido se fusionó a su ARNm después de una incubación de la traducción de 1 hora a $30^{\circ}C$. Este valor aumentó hasta 12% después de la incubación toda la noche a $-20^{\circ}C$. Cuando la incubación post-traduccional se llevó a cabo en presencia de una concentración elevada de Mg^{2+} , más del 50% del péptido traducido se fusionó al molde.
35

Para un molde con el ligador flexible, aproximadamente el 25% del péptido traducido se fusionó al molde después de 1 hora de traducción a $30^{\circ}C$. Este valor aumentó hasta alrededor de 50% tras la incubación toda la noche a $-20^{\circ}C$, y hasta más del 75% si la incubación post-traduccional se realizó en presencia de 50 mM de Mg^{2+} .
40

Para determinar el porcentaje del molde de entrada convertido a producto de fusión, las traducciones se realizaron usando molde de ARNm-ligador marcado con ^{32}P . Cuando se usó el ligador flexible y la incubación post-traduccional se realizó a $-70^{\circ}C$ sin adición de Mg^{2+} , alrededor del 20%, 40%, 40%, 35% y 20% del molde de entrada se convirtió en fusión de ARNm-péptido cuando la concentración del molde de ARN de entrada fue 800, 400, 200, 100 y 50 nM respectivamente (Figura 18). Se obtuvieron resultados similares cuando la incubación post-traduccional se realizó en presencia de 50 mM de Mg^{2+} . Los mejores resultados se lograron usando lisados obtenidos de Novagen, Amersham, o Ambion (Figura 19).
45

Las diferencias de movilidad entre las fusiones de ARNm-péptido y los ARNm, según se miden mediante SDS-PAGE, pueden ser muy pequeñas si el molde de ARNm es largo. En tales casos, el molde se puede marcar en el extremo 5' del ligador con ^{32}P . La porción de ARN larga se puede digerir entonces con ARNasa H en presencia de una tabillita de ADN complementario después de la traducción/incubación, y la eficacia de la fusión se puede determinar mediante cuantificación de la relación de ligador no modificado a fusión de ligador-péptido. En comparación con la digestión con ARNasa A, que produce 3'-P y 5'-OH, este enfoque tiene la ventaja de que no se elimina el ^{32}P en el extremo 5' del ligador.
50

Fusión intramolecular frente a intermolecular durante la incubación post-traduccional. Además de los experimentos anteriores, se estudió si la reacción de fusión que ocurría a $-20^{\circ}C$ en presencia de Mg^{2+} tenía una naturaleza intra- o intermolecular. Se coincubó ligador libre ($dA_{27}dCdCP$ o $dA_{21}C_9C_9C_9dAdCdCP$, en el que C_9 es $-O(CH_2CH_2O)_3PO_2$), con un molde que contiene un ligador de ADN, pero sin puromicina en el extremo 3', en las condiciones de traducción y de incubación post-traduccional descritas anteriormente. En estos experimentos, no se incorporó en el producto de ligador-péptido ninguna cantidad detectable (esto es, menos de 2% del nivel normal) de ^{35}S -Met, sugiriendo que la fusión post-traduccional se produjo principalmente entre el péptido naciente y el ARNm unido al mismo ribosoma.
55

ES 2 268 763 T3

Resultados de optimización. Como se ilustró anteriormente, usando el ligador flexible y/o realizando la incubación post-traduccional en presencia de una concentración elevada de Mg^{2+} , se aumentó las eficacias de la fusión hasta aproximadamente el 40% del ARNm de entrada. Estos resultados indicaron que se pudieron generar tantas como 10^{14} moléculas de fusión de ARNm-péptido por ml de mezcla de reacción de traducción *in vitro*, produciendo conjuntos de fusiones de ARNm-péptido de complejidad muy elevada para uso en experimentos de selección *in vitro*.

Enriquecimiento selectivo de fusiones de ARN-proteína

Se ha demostrado la factibilidad del uso de fusiones de ARN-péptido en experimentos de selección y de evolución enriqueciendo una fusión particular de ARN-péptido a partir de un conjunto complejo de fusiones de secuencias al azar en base al péptido codificado. En particular, se preparó una serie de mezclas en las que se combinó una pequeña cantidad de secuencia conocida (en este caso, el molde de myc largo, LP154) con cierta cantidad de conjunto de secuencia al azar (esto es, LP160). Estas mezclas se tradujeron, y los productos de la fusión de ARN-péptido se seleccionaron mediante cromatografía de afinidad de disulfuro y de oligonucleótidos como se describe en la presente memoria. Las fusiones de molde de myc se inmunoprecipitaron selectivamente con anticuerpo monoclonal anti-myc (Figura 16A). Para medir el enriquecimiento obtenido en esta etapa selectiva, se amplificaron mediante PCR, en presencia de un cebador marcado radioactivamente, alícuotas de la mezcla de fusiones de ADNc/ARNm-péptido procedentes antes y después de la inmunoprecipitación. El ADN amplificado se digirió con endonucleasa de restricción que cortó la secuencia del molde de myc, pero no el conjunto (Figuras 16B y 16C). La cuantificación de la relación de ADN cortado y sin cortar indicó que la secuencia de myc estaba enriquecida unas 20-40 veces con relación a la librería aleatoria mediante inmunoprecipitación.

Estos experimentos se llevaron a cabo según lo siguiente.

Reacciones de traducción. Las reacciones de traducción se realizaron generalmente como se describe antes. Específicamente, las reacciones se realizaron a 30°C durante una hora según las especificaciones del fabricante (Novagen) y se congelaron toda la noche a -20°C. Se obtuvieron dos versiones de seis muestras, una que contiene ^{35}S -metionina y una que contiene metionina fría añadida hasta una concentración final de 52 μM . Las reacciones 1-6 contenían las cantidades de moldes descritas en la Tabla 2. Todos los números en la Tabla 2 representan picomoles de molde por 25 μl de mezcla de reacción.

TABLA 2

Relaciones de moldes usados en selección dopada

Reacción	LP154	LP160
1	---	---
2	5	---
3	1	20
4	0,1	20
5	0,01	20
6	---	20

Preparación de dT_{25} estreptavidina agarosa. La estreptavidina agarosa (Pierce) se lavó tres veces con TE 8.2 (10 mM de Tris-Cl pH 8,2, 1 mM de EDTA), y se resuspendió como una suspensión 1:1 (v/v) en TE 8.2. Después se añadió 3' biotinil T_{25} , sintetizado usando Bioteg CPG (Glen Research), hasta la concentración final deseada (generalmente 10 ó 20 μM), y la incubación se llevó a cabo con agitación durante 1 hora. La dT_{25} estreptavidina agarosa se lavó entonces tres veces con TE 8.2, y se almacenó a 4°C hasta el uso.

Purificación de moldes a partir de reacciones de traducción. Para purificar moldes procedentes de reacciones de traducción, se eliminaron 25 μl de cada reacción y se añadieron a 7,5 ml de tampón de aislamiento (1 M de NaCl, 100 mM de Tris-Cl pH 8,2, 10 mM de EDTA, 0,1 mM de DTT), y 125 μl de 20 μM de dT_{25} estreptavidina agarosa. Esta disolución se incubó a 4°C durante una hora con rotación. Los tubos se centrifugaron, y se eliminó el eluyente. Se añadió un ml de tampón de aislamiento, la suspensión se resuspendió, y las mezclas se transfirieron a tubos de microcentrifuga de 1,5 ml. Las muestras se lavaron entonces cuatro veces con alícuotas de 1 ml de tampón de aislamiento enfriado con hielo. Después se combinaron las muestras caliente y fría procedentes de reacciones idénticas en un filtro Millipore MC de una unidad de filtro, y se eluyeron a partir de la dT_{25} agarosa lavando con 2 volúmenes de 100 μl de H_2O , 0,1 mM de DTT, y 2 volúmenes de 15 mM de NaOH, 1 mM de EDTA.

ES 2 268 763 T3

A este eluyente se añadieron 40 μ l de una suspensión al 50% de tiopropilsefaroza (Pharmacia) lavada, y la incubación se llevó a cabo a 4°C con rotación durante 1 hora. Las muestras se lavaron entonces con tres volúmenes de 1 ml de TE 8.2, y se eliminó el eluyente. Se añadió un μ l de 1 M de DTT al sólido (volumen total de aproximadamente 20-30 μ l), y la muestra se incubó durante varias horas, se eliminó, y se lavó cuatro veces con 20 μ l de H₂O (volumen total 90 μ l). El eluyente contenía 2,5 mM de tiopiridona, según se pudo comprobar mediante absorbancia por UV. Se precipitaron con etanol 50 μ l de esta muestra añadiendo 6 μ l de 3 M de NaOAc pH 5,2, 10 mM de espermina, 1 μ l de glicógeno (10 mg/ml, Boehringer Mannheim), y 170 μ l de EtOH al 100%, incubando durante 30 minutos a -70°C, y centrifugando durante 30 minutos a 13.000 rpm en una microcentrífuga.

Reacciones de transcriptasa inversa. Las reacciones de transcriptasa inversa se realizaron sobre las muestras de eluyente tanto precipitadas con etanol como la que contiene tiopiridona, según lo siguiente. Para las muestras precipitadas con etanol, se hibridaron a 70°C durante 5 minutos 30 μ l de molde resuspendido, H₂O hasta 48 μ l, y 200 picomoles de cebador 21.103 (SEC ID n°: 22), y se enfrió en hielo. A esta muestra, se añadieron 16 μ l del primer tampón de hebra (250 mM de Tris-Cl pH 8,3, 375 mM de KCl, y 15 mM de MgCl₂; disponible de Gibco BRL, Grand Island, NY), 8 μ l de 100 mM de DTT, y 4 μ l de 10 mM de NTP, y se equilibró a 42°C, y se añadieron 4 μ l de transcriptasa inversa Superscript II (Gibco BRL, Grand Island, NY). Se añadió H₂O (13 μ l) al eluyente de TP sefaroza (35 μ l), y las reacciones se realizaron como antes. Después de la incubación durante una hora, se combinaron un número igual de muestras (volumen total 160 μ l). Se reservaron 10 μ l de muestra para la PCR de cada muestra no seleccionada, y se reservaron 150 μ l de muestra para la inmunoprecipitación.

Inmunoprecipitación. Para llevar a cabo las inmunoprecipitaciones, se añadieron 170 μ l de reacción de transcriptasa inversa a 1 ml de tampón de dilución (10 mM de Tris-Cl, pH 8,2, 140 mM de NaCl, 1% v/v de Tritón X-100) y 20 μ l de conjugado de proteína G/A (Calbiochem, La Jolla, CA), y se preclaró mediante incubación a 4°C con rotación durante 1 hora. Se eliminó el eluyente, y se añadieron 20 μ l de conjugado de G/A y 20 μ l de anticuerpo monoclonal (2 μ g, 12 picomoles), y la muestra se incubó con rotación durante dos horas a 4°C. El conjugado se precipitó mediante microcentrifugación a 2500 rpm durante 5 minutos, se eliminó el eluyente, y el conjugado se lavó tres veces con alícuotas de 1 ml de tampón de dilución enfriado con hielo. La muestra se lavó entonces con 1 ml de 10 mM de Tris-Cl, pH 8,2, enfriado en hielo, 100 mM de NaCl. Los fragmentos unidos se eliminaron usando 3 volúmenes de HOAc al 4% congelado, y las muestras se liofilizaron hasta sequedad.

PCR de muestras seleccionadas y no seleccionadas. Las reacciones de PCR se llevaron a cabo añadiendo 20 μ l de NH₄OH concentrado a 10 μ l del material no seleccionado y a la totalidad del material seleccionado, e incubando durante 5 minutos cada uno a 55°C, 70°C y 90°C, para destruir cualquier ARN presente en la muestra. Las muestras se evaporaron entonces hasta sequedad usando un Speedvac. Se añadieron a cada muestra 200 μ l de mezcla de PCR (1 μ M de cebadores 21.103 y 42.108, 200 μ M de dNTP en tampón de PCR más Mg²⁺ (Boehringer Mannheim), y 2 μ l de Taq polimerasa (Boehringer Mannheim)). Se realizaron 16 ciclos de PCR sobre el número 2 de las muestras no seleccionadas, y se realizaron 19 ciclos sobre todas las otras muestras.

Las muestras se amplificaron entonces en presencia de cebador 21.103 marcado con ³²P en 5', según la Tabla 3, y se purificaron dos veces individualmente usando kits de purificación mediante PCR directa Wizard (Promega), para eliminar todo el cebador y los fragmentos más cortos.

TABLA 3

Amplificación de muestras de PCR seleccionadas y no seleccionadas

Muestra	Tipo	Volumen	Ciclos
1	No seleccionado	20 μ l	5
2	No seleccionado	5 μ l	4
3	No seleccionado	20 μ l	5
4	No seleccionado	20 μ l	5
5	No seleccionado	20 μ l	5
6	No seleccionado	20 μ l	5

ES 2 268 763 T3

TABLA 3 (continuación)

	Muestra	Tipo	Volumen	Ciclos
5	1	Seleccionado	20 μ l	5
	2	Seleccionado	5 μ l	4
10	3	Seleccionado	20 μ l	5
	4	Seleccionado	20 μ l	7
	5	Seleccionado	20 μ l	7
15	6	Seleccionado	20 μ l	7

20 *Digestiones de restricción.* El ADN marcado con 32 P, preparado a partir de cada una de las reacciones de PCR anteriores, se añadió en cantidades iguales (por cpm de muestra) a reacciones de digestión de restricción, según la Tabla 4. El volumen total de cada reacción fue 25 μ l. Se añadió a cada reacción 0,5 μ l de Alwnt (5 unidades, New England Biolabs). Las muestras se incubaron a 37°C durante 1 hora, y la enzima se inactivó por calor mediante una incubación de 20 minutos a 65°C. Las muestras se mezclaron entonces con 10 μ l de tampón de carga desnaturante (1 ml de formamida ultrapura (USB), 20 μ l de 0,5 M de EDTA, y 20 μ l de 1 M de NaOH), se calentaron hasta 90°C durante 1 minuto, se enfriaron y se cargaron en un gel de poliacrilamida desnaturante al 12% que contiene urea 8 M. Después de la electroforesis, el gel se fijó con 10% (v/v) de HOAc, 10% (v/v) de MeOH, H₂O.

TABLA 4

Condiciones de digestión de restricción con Alwnt

	Muestra	Tipo	Volumen ADN añadido a la reacción	Volumen total
35	1	No seleccionado	20 μ l	25 μ l
	2	No seleccionado	4 μ l	25 μ l
40	3	No seleccionado	20 μ l	25 μ l
	4	No seleccionado	20 μ l	25 μ l
	5	No seleccionado	4 μ l	25 μ l
45	6	No seleccionado	20 μ l	25 μ l
	1	Seleccionado	20 μ l	25 μ l
50	2	Seleccionado	8 μ l	25 μ l
	3	Seleccionado	12 μ l	25 μ l
	4	Seleccionado	12 μ l	25 μ l
55	5	Seleccionado	20 μ l	25 μ l
	6	Seleccionado	20 μ l	25 μ l

60 *Cuantificación de la digestión.* La cantidad de ADN de myc frente a ADN del conjunto presente en la muestra se cuantificó usando un formador de fosfoimágenes (Molecular Dynamics). La cantidad de material presente en cada banda se determinó como el volumen integrado de rectángulos idénticos dibujados alrededor de las bandas de gel. El cpm total presente en cada banda se calculó como el volumen menos el fondo. Se usaron tres valores de fondo: 65 (1) una media de cuadrados idénticos fuera del área en la que se produjeron los recuentos en el gel; (2) el cpm presente en la línea del conjunto sin seleccionar en la que debería de aparecer la banda de myc (en esta posición no aparece ninguna banda en el gel); y (3) un valor normalizado que reproduce el valor más próximo a los incrementos

del molde de 10 veces entre las líneas no seleccionadas. Las líneas 2, 3 y 4 de las Figuras 16B y 16C demuestran en enriquecimiento de la secuencia diana frente a la secuencia del conjunto. El enriquecimiento demostrable en la línea 3 (no seleccionado/seleccionado) produjo los valores más elevados (17, 43 y 27 veces usando los métodos 1-3, respectivamente) debido a la optimización de la relación de señal a ruido para esta muestra. Estos resultados se resumen en la Tabla 5.

TABLA 5

Enriquecimiento del molde de myc frente al conjunto

Método	Línea 2 (20)	Línea 3 (200)	Línea 4 (2000)
1	7,0	16,6	5,7
2	10,4	43	39
3	8,7	27	10,2

En un segundo conjunto de experimentos, estos mismos productos de PCR se purificaron una vez usando kits de purificación de PCR directa Wizard, y las digestiones se cuantificaron mediante el método (2) anterior. En estos experimentos, se obtuvieron resultados similares; se midieron enriquecimientos de 10,7, 38 y 12 veces, respectivamente, para muestras equivalentes a aquellas en las líneas 2, 3 y 4 anteriores.

Uso de sistemas de selección de proteínas

Los sistemas de selección de la presente invención tienen aplicaciones comerciales en cualquier área en la que se use la tecnología de proteínas para resolver problemas terapéuticos, de diagnóstico o industriales. Esta tecnología de selección es útil para mejorar o alterar proteínas existentes, así como para aislar nuevas proteínas con funciones deseadas. Estas proteínas pueden ser secuencias de origen natural, pueden ser formas alteradas de secuencias de origen natural, o pueden ser secuencias parcial o totalmente sintéticas.

Aislamiento de nuevos reactivos de unión. En una aplicación particular, la tecnología de la fusión de ARN-proteína descrita en la presente memoria es útil para el aislamiento de proteínas con propiedades de unión específicas (por ejemplo, unión a ligandos). Las proteínas que muestran interacciones de unión muy específicas se pueden usar como reactivos de reconocimiento que no son anticuerpos, permitiendo que la tecnología de la fusión de ARN-proteína burle a la tecnología tradicional de anticuerpos monoclonales. Los reactivos de tipo anticuerpo aislados mediante este método se pueden usar en cualquier área en la que se utilicen anticuerpos tradicionales, incluyendo aplicaciones de diagnóstico y terapéutico.

Mejora de los anticuerpos humanos. La presente invención también se puede usar para mejorar anticuerpos humanos o humanizados para el tratamiento de cualquiera de un número de enfermedades. En esta aplicación, se desarrollan librerías de anticuerpos y se seleccionan sistemáticamente *in vitro*, eliminando la necesidad de técnicas tales como la fusión celular o la presentación de fagos. En una aplicación importante, la invención es útil para mejorar librerías de anticuerpos de una sola cadena (Ward *et al.*, Nature **341**:544 (1989); y Goulot *et al.*, J. Mol. Biol. **213**:617 (1990)). Para esta aplicación, la región variable se puede construir a partir de una fuente humana (para minimizar posibles reacciones inmunitarias adversas del receptor), o puede contener un casete totalmente aleatorizado (para maximizar la complejidad de la librería). Para detectar sistemáticamente moléculas de anticuerpos mejoradas, se ensaya un conjunto de moléculas candidatas para determinar la unión a una molécula diana (por ejemplo, un antígeno inmovilizado como se muestra en la Figura 2). Después se aplican niveles mayores de restricción a la etapa de unión a medida que transcurre la selección desde una ronda a la siguiente. Para aumentar la restricción, se alteran las condiciones tales como el número de etapas de lavado, la concentración de competidor en exceso, las condiciones del tampón, la duración del tiempo de reacción de unión, y la elección de la matriz de inmovilización.

Los anticuerpos monocatenarios se pueden usar directamente para terapia, o indirectamente para el diseño de anticuerpos estándares. Tales anticuerpos tienen un número de aplicaciones potenciales, incluyendo el aislamiento de anticuerpos anti-autoinmunitarios, supresión inmunitaria, y en el desarrollo de vacunas para enfermedades víricas tales como SIDA.

Aislamiento de nuevos catalizadores. La presente invención también se puede usar para seleccionar nuevas proteínas catalíticas. La selección *in vitro* y la evolución se han usado previamente para el aislamiento de nuevos ARN y ADN catalíticos, y, en la presente invención, se usan para el aislamiento de nuevas enzimas proteínicas. En un ejemplo particular de este enfoque, se puede aislar indirectamente un catalizador seleccionando la unión a un análogo químico del estado de transición del catalizador. En otro ejemplo particular, el aislamiento directo se puede llevar a cabo seleccionando la formación del enlace covalente con un sustrato (por ejemplo, usando un sustrato enlazado a un marcador de afinidad), o mediante escisión (por ejemplo, seleccionando la capacidad para romper un enlace específico y de ese modo liberar miembros catalíticos de una librería a partir de un soporte sólido).

Este enfoque para el aislamiento de nuevos catalizadores tiene al menos dos ventajas importantes con respecto a la tecnología de anticuerpos catalíticos (revisada en Schultz *et al.*, J. Chem. Engng. News 68:26 (1990)). En primer lugar, en la tecnología de anticuerpos catalíticos, el conjunto inicial está limitado generalmente al plegamiento de la inmunoglobulina; por el contrario, la librería de partida de las fusiones de ARN-proteína pueden ser completamente aleatorias, pueden constar, sin limitación, de variantes de estructuras enzimáticas conocidas o estructuras proteínicas. Además, el aislamiento de anticuerpos catalíticos generalmente descansa en una selección inicial de la unión a análogos de la reacción del estado de transición, seguido de una identificación sistemática laboriosa de anticuerpos activos; nuevamente, por el contrario, la selección directa de catalizadores es posible usando un enfoque de librería de fusión de ARN-proteína, como se demostró previamente usando librerías de ARN. En un enfoque alternativo para aislar enzimas proteínicas, se pueden combinar los enfoques del análogo del estado de transición con el de selección directa.

Las enzimas obtenidas mediante este método son muy valiosas. Por ejemplo, actualmente existe una necesidad imperiosa de catalizadores industriales nuevos y eficaces que permitan que se desarrollen procedimientos químicos mejorados. Una ventaja principal de esta invención es que las elecciones se pueden llevar a cabo en condiciones arbitrarias y no están limitadas, por ejemplo, a condiciones *in vivo*. Por lo tanto, la invención facilita el aislamiento de nuevas enzimas o variantes mejoradas de enzimas existentes que pueden llevar a cabo transformaciones muy específicas (y de ese modo minimicen la formación de subproductos indeseados), a la vez que funcionen en entornos predeterminados como por ejemplo entornos de temperatura, presión, o concentración de disolvente elevados.

Una trampa de interacción in vitro. La tecnología de fusión de ARN-proteína también es útil para identificar sistemáticamente librerías de ADNc y clonar nuevos genes en base a las interacciones proteína-proteína. Mediante este método, se genera una librería de ADNc a partir de una fuente deseada (por ejemplo, mediante el método de Ausubel *et al.*, citado anteriormente, capítulo 5). A cada uno de los ADNc candidatos, se liga (por ejemplo, usando las técnicas descritas anteriormente para la generación de LP77, LP154 y LP160) un aceptor de péptidos (por ejemplo, como una cola de puromicina). Después, las fusiones de ARN-proteína se generan como se describe en la presente memoria, y entonces se estudia como se describe anteriormente la capacidad de estas fusiones (o de versiones mejoradas de las fusiones) para interactuar con moléculas particulares. Si se desea, se pueden evitar en este procedimiento codones de parada y regiones de 3'UTR (i) añadiendo ARNt supresor para permitir la lectura de las regiones de parada, (ii) eliminando el factor de liberación de la reacción de traducción mediante inmunoprecipitación, (iii) una combinación de (i) y (ii), o (iv) la eliminación de los codones de parada y 3'UTR de las secuencias de ADN.

El hecho de que la etapa de interacción tiene un lugar *in vitro* permite el control cuidadoso de la restricción de la reacción, usando condiciones iónicas, de temperatura y de competidor no específico. La alteración de moléculas pequeñas normales con análogos no hidrolizables (por ejemplo, ATP frente a ATPγS) proporciona selecciones que discriminan entre diferentes conformeros de la misma molécula. Este enfoque es útil tanto para la clonación como para la identificación funcional de muchas proteínas, puesto que la secuencia de ARN de la pareja de unión seleccionada se une covalentemente y por lo tanto se puede aislar fácilmente. Además, la técnica es útil para identificar funciones e interacciones de los ~50-100.000 genes humanos, cuyas secuencias están siendo determinadas actualmente mediante el proyecto de Genoma Humano.

Uso de fusiones de arn-proteína en un formato de microchip

Los "chips de ADN" constan de conjuntos espacialmente definidos de oligonucleótidos inmovilizados o fragmentos clonados de ADNc o ADN genómico, y tienen aplicaciones tales como la secuenciación rápida y la obtención del perfil del transcritto. Hibridando una mezcla de fusiones de ARN-proteína (por ejemplo, generadas a partir de un conjunto de ADN o ARN celular), a tal chip de ADN, es posible generar un "chip de presentación de proteína", en el que cada punto que corresponde a una secuencia inmovilizada es capaz de hibridarse a su secuencia de ARN correspondiente en el conjunto de fusiones de ARN-proteína. Mediante este enfoque, la proteína correspondiente se inmoviliza de una manera espacialmente definida debido a su enlace a su propio ARNm, y los chips que contienen conjuntos de secuencias de ADN presentan el conjunto correspondiente de proteínas. Como alternativa, los fragmentos peptídicos de estas proteínas se pueden presentar si la librería de fusión se genera a partir de fragmentos más pequeños de ADNc o ADN genómicos.

Dichas presentaciones ordenadas de proteínas y péptidos tienen muchos usos. Por ejemplo, representan herramientas poderosas para la identificación de interacciones proteína-proteína previamente desconocidas. En un formato específico, se marca de forma detectable una proteína sonda (por ejemplo, con un colorante fluorescente), y la proteína marcada se incuba con un chip de presentación de proteína. Mediante este enfoque, se determina la identidad de proteínas que son capaces de unirse a la proteína sonda, a partir de la localización de las manchas en el chip que se marcan debido a la unión de la sonda. Otra aplicación es la determinación rápida de proteínas que se modifican químicamente mediante la acción de enzimas modificadoras (por ejemplo, proteinaquinasas, aciltransferasas, y metiltransferasas). Incubando el chip de presentación de proteína con la enzima de interés y con un sustrato marcado radioactivamente, después del lavado y la autorradiografía, se pueden determinar fácilmente la localización y por tanto la identidad de aquellas proteínas que son sustratos para la enzima modificadora. Además, el uso de este enfoque con presentaciones ordenadas de pequeños péptidos permite la localización adicional de tales sitios de modificación.

La tecnología de presentación de proteínas se puede llevar a cabo usando conjuntos de ácidos nucleicos (incluyendo ARN, pero preferentemente ADN) inmovilizados sobre cualquier soporte sólido apropiado. Los soportes sólidos

ES 2 268 763 T3

ejemplares pueden estar hechos de materiales tales como vidrio (por ejemplo, placas de vidrio), silicio o vidrio de silicio (por ejemplo, microchips), u oro (por ejemplo, placas de oro). Los métodos para unir ácidos nucleicos a regiones precisas en tales superficies sólidas, por ejemplo métodos fotolitográficos, son bien conocidos en la técnica, y se pueden usar para generar soportes sólidos (tales como chips de ADN) para uso en la invención. Los métodos ejemplares para este fin incluyen, sin limitación, Schena *et al.*, *Science* **270**:467-470 (1995); Kozal *et al.*, *Nature Medicine* **2**:753-759 (1996); Cheng *et al.*, *Nucleic Acids Research* **24**:380-385 (1996); Lipshutz *et al.*, *BioTechniques* **19**:442-447 (1995); Pease *et al.*, *Proc. Natl. Acad. Sci. USA* **91**:5022-5026 (1994); Fodor *et al.*, *Nature* **364**:555-556 (1993); Pirrung *et al.*, patente U.S. nº 5.143.854; y Fodor *et al.*, documento WO 92/10092.

10

15

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

5 1. Método para la selección de un ARN proteínico deseado, o molécula de ADN, que comprende las etapas siguientes:

- 10 a) proporcionar una población de moléculas de ARN candidatas, cada una de las cuales comprende una secuencia de iniciación de la traducción y un codón de partida enlazado funcionalmente a una secuencia codificante de proteína candidata, y cada una de las cuales está enlazada covalentemente a un aceptor peptídico en el extremo 3' de dicha secuencia codificante de proteína candidata;
- 15 b) traducir *in vitro* o *in situ* dichas secuencias codificantes de proteínas candidatas, para producir una población de fusiones de candidatas ARN-proteína; y
- c) seleccionar una fusión de deseada ARN-proteína, seleccionando de ese modo dicha proteína deseada o dicho ARN deseado;

20 o generar, a partir de dicha molécula de ARN de dicha fusión, una molécula de ADN que codifica dicha proteína deseada, seleccionando de ese modo dicha molécula de ADN deseada.

2. Método para la selección de una proteína que tiene una función alterada, o una molécula de ADN que codifica una proteína que tiene una función alterada con relación a una proteína de referencia, que comprende las etapas siguientes:

- 25 a) producir una población de moléculas de ARN candidatas a partir de una población de moldes de ADN candidatos, teniendo cada uno de dichos moldes de ADN candidatos una secuencia codificante de proteína candidata que difiere de dicha secuencia codificante de proteína de referencia, comprendiendo cada una de dichas moléculas de ARN una secuencia de iniciación de la traducción y un codón de partida enlazado funcionalmente a dicha secuencia codificante de proteína candidata, y estando cada una enlazada covalentemente a un aceptor peptídico en el extremo 3';
- 30 b) traducir *in vitro* o *in situ* dichas secuencias codificantes de proteínas candidatas, para producir una población de fusiones candidatas de ARN-proteína; y
- 35 c) seleccionar una fusión de ARN-proteína que tiene una función alterada, seleccionando de ese modo dicha proteína que tiene dicha función alterada;

40 o generar a partir de dicha molécula de ARN de dicha fusión una molécula de ADN que codifica dicha proteína deseada, seleccionando de ese modo dicha molécula de ADN deseada.

3. Método según la reivindicación 1, en el que dicha proteína deseada o dicho ARN deseado en la etapa (c) se selecciona mediante

45 la puesta en contacto de dicha población de fusiones de ARN-proteína con una pareja de unión específica para la porción de ARN o para la porción proteínica de dicha fusión de ARN-proteína, en condiciones que separan sustancialmente dicho complejo de pareja de unión-fusión de ARN-proteína de los miembros no unidos de dicha población;

la liberación de dichas fusiones unidas de ARN-proteína a partir de dicho complejo; y

50 la puesta en contacto de dicha población de fusiones de ARN-proteína de la etapa (c) con una pareja de unión específica para la porción proteínica de dicha fusión deseada de ARN-proteína, en condiciones que separen sustancialmente dicho complejo de pareja de unión-fusión de ARN-proteína de los miembros no unidos de dicha población, seleccionando de ese modo la proteína deseada o el ARN deseado.

55 4. Método según la reivindicación 3, en el que dicho método comprende además repetir las etapas (a) a (c) de la reivindicación 1.

5. Método según cualquiera de las reivindicaciones 1 a 4, en el que dicho aceptor peptídico es puromicina.

60 6. Método según cualquiera de las reivindicaciones 1 a 4, en el que cada una de dichas moléculas de ARN candidatas comprende además una secuencia de pausa, o comprende además una secuencia de ADN o de análogo de ADN enlazada covalentemente al extremo 3' de dicha molécula de ARN.

65 7. Método según cualquiera de las reivindicaciones 1 a 4, en el que dicha población de moléculas de ARN candidatas comprende al menos 10^{13} moléculas diferentes de ARN.

8. Método según cualquiera de las reivindicaciones 1 a 4, en el que dicha reacción de traducción *in vitro* se lleva a cabo en un lisado preparado a partir de una célula eucariota o porción de la misma.

ES 2 268 763 T3

9. Método según la reivindicación 8, en el que dicha reacción de traducción *in vitro* se lleva a cabo en un lisado reticulocítico o en un lisado de germen de trigo.

5 10. Método según cualquiera de las reivindicaciones 1 a 9, en el que dicha reacción de traducción *in vitro* se lleva a cabo en un lisado preparado a partir de una célula bacteriana o una porción de la misma.

11. Método según cualquiera de las reivindicaciones 1 a 4, en el que dicha etapa de selección comprende unir dicha proteína deseada a una pareja de unión inmovilizada.

10 12. Método según cualquiera de las reivindicaciones 1 a 4, en el que dicha etapa de selección comprende evaluar una actividad funcional de dicha proteína deseada.

13. Método según las reivindicaciones 1 ó 2, en el que dicha molécula de ADN se amplifica.

15 14. Método según las reivindicaciones 1 ó 2, en el que dicho método comprende además repetir la etapa (a) a (c).

15. Método según las reivindicaciones 1 ó 2, en el que dicho método comprende además transcribir una molécula de ARN a partir de dicha molécula de ADN, y repetir las etapas (a) a (c).

20 16. Método según cualquiera de las reivindicaciones 1 a 4, en el que dicho ARN está enlazado covalentemente a través de un enlace de amida a dicha proteína en dicha fusión de ARN-proteína.

25 17. Método según cualquiera de las reivindicaciones 1 a 4, en el que dicho ARN está enlazado covalentemente a dicha proteína en dicha fusión de ARN-proteína, siendo dicho enlace covalente resistente a la ruptura mediante un ribosoma.

18. Método según cualquiera de las reivindicaciones 1 a 4, en el que, después de la etapa de traducción *in vitro*, se lleva a cabo una incubación en presencia de 50-100 mM de Mg^{2+} .

30 19. Método según cualquiera de las reivindicaciones 1 a 4, en el que dicha fusión de ARN-proteína comprende además un ácido nucleico o secuencia de análogo de ácido nucleico situada próxima a dicho aceptor peptídico que aumenta la flexibilidad.

35 20. Fusión de ARN-proteína seleccionada mediante cualquiera de los métodos según las reivindicaciones 1 a 4.

21. Molécula que comprende un ácido ribonucleico enlazado covalentemente a través de un enlace de amida a una proteína.

40 22. Molécula según la reivindicación 21, en la que dicha proteína está codificada por dicho ácido ribonucleico.

23. Molécula según la reivindicación 21, en la que dicho ácido ribonucleico es un ARN mensajero.

45 24. Molécula que comprende un ácido ribonucleico enlazado covalentemente a una proteína, siendo dicha proteína codificada completamente por dicho ácido ribonucleico.

25. Molécula según la reivindicación 24, en la que dicho ácido ribonucleico es ARN mensajero.

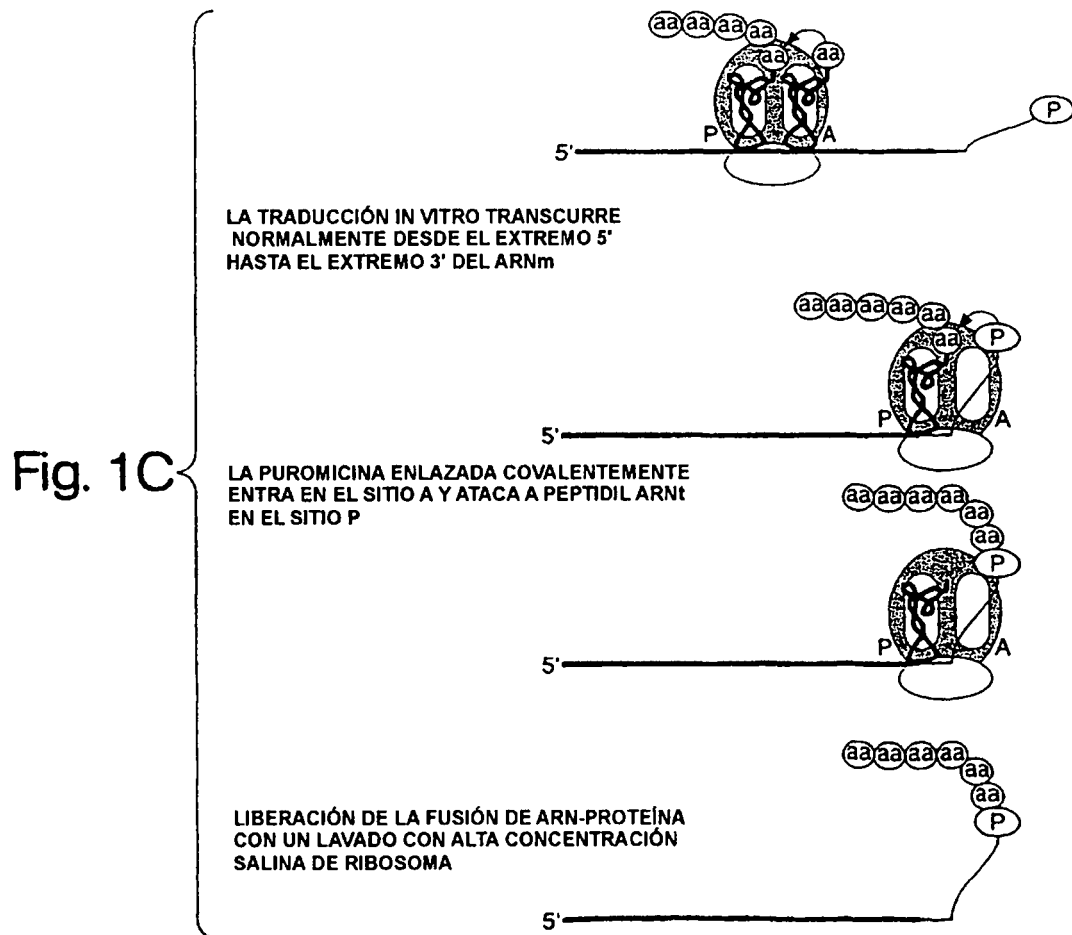
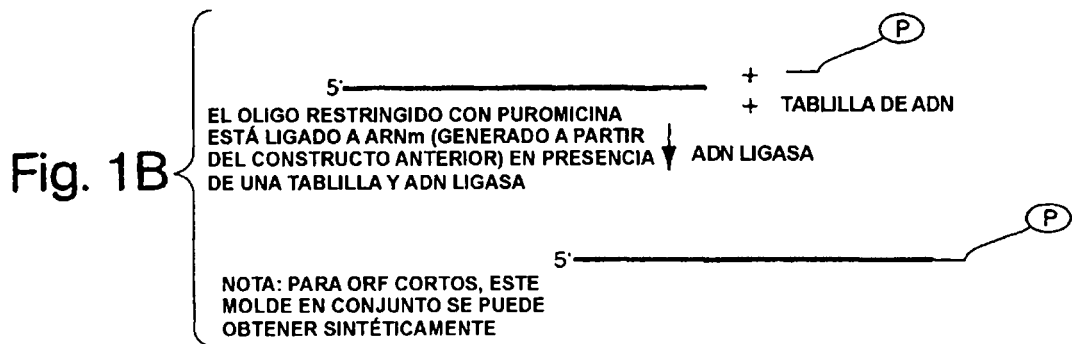
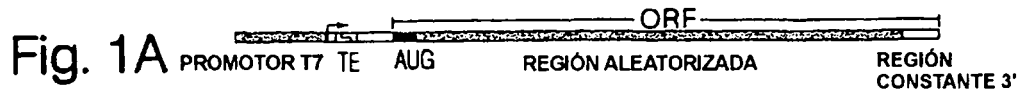
50 26. Molécula que comprende un ácido ribonucleico enlazado covalentemente a una proteína, siendo dicho enlace covalente resistente a la ruptura mediante un ribosoma.

27. Molécula según la reivindicación 26, en la que dicho ácido ribonucleico es ARN mensajero.

55 28. Ácido ribonucleico que comprende una secuencia de iniciación de la traducción y un codón de partida enlazado funcionalmente a una secuencia codificante de proteína candidata, estando dicho ácido ribonucleico enlazado covalentemente a un aceptor peptídico en el extremo 3' de dicha secuencia codificante de proteína candidata.

60

65



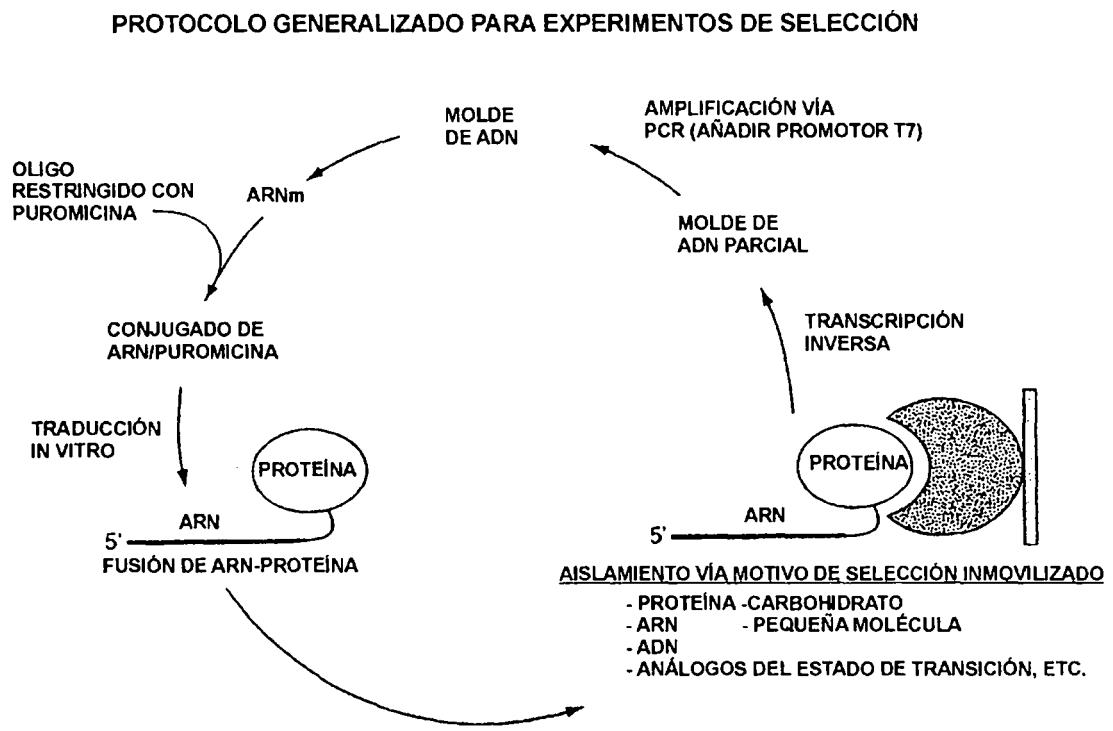


Fig. 2

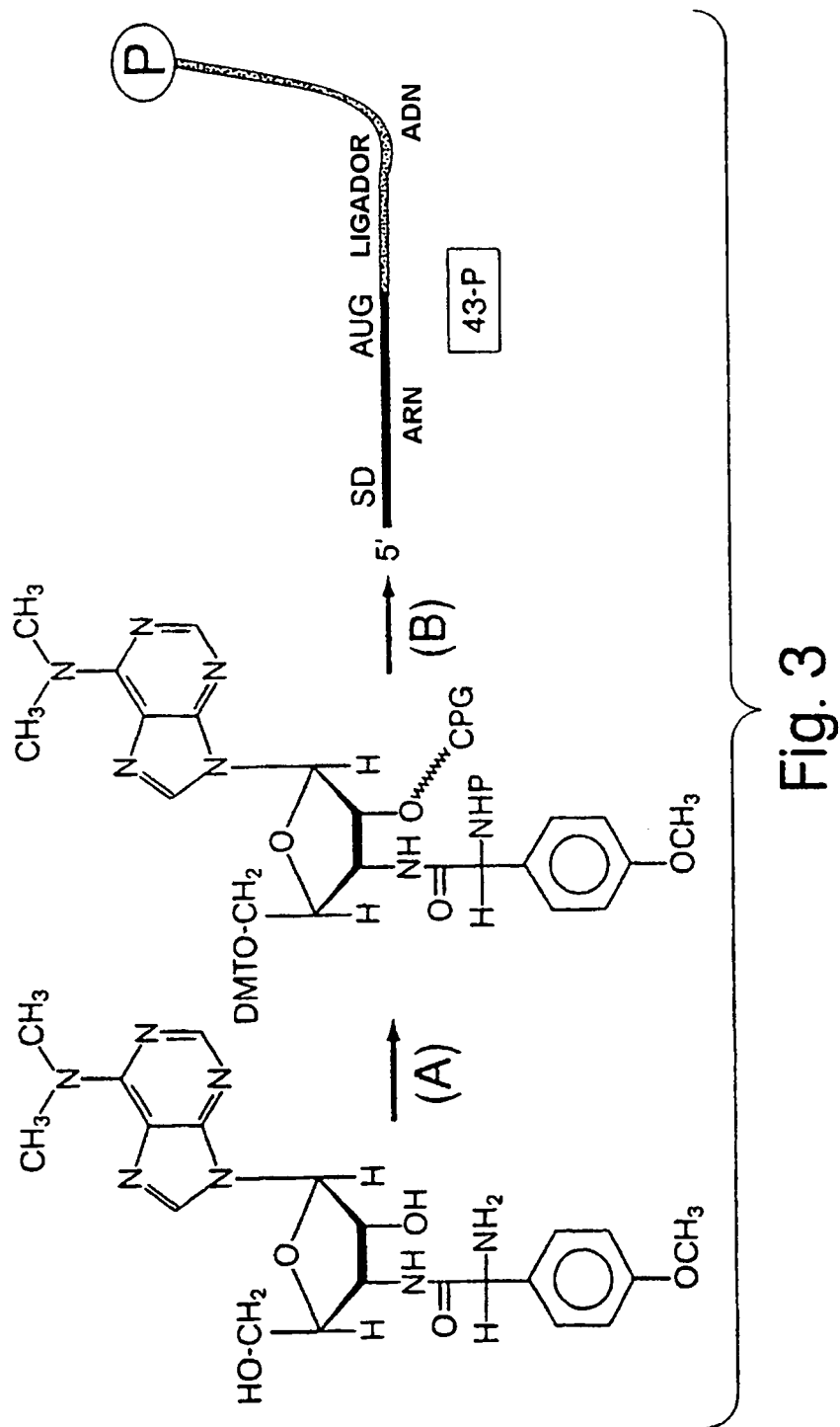


Fig. 3

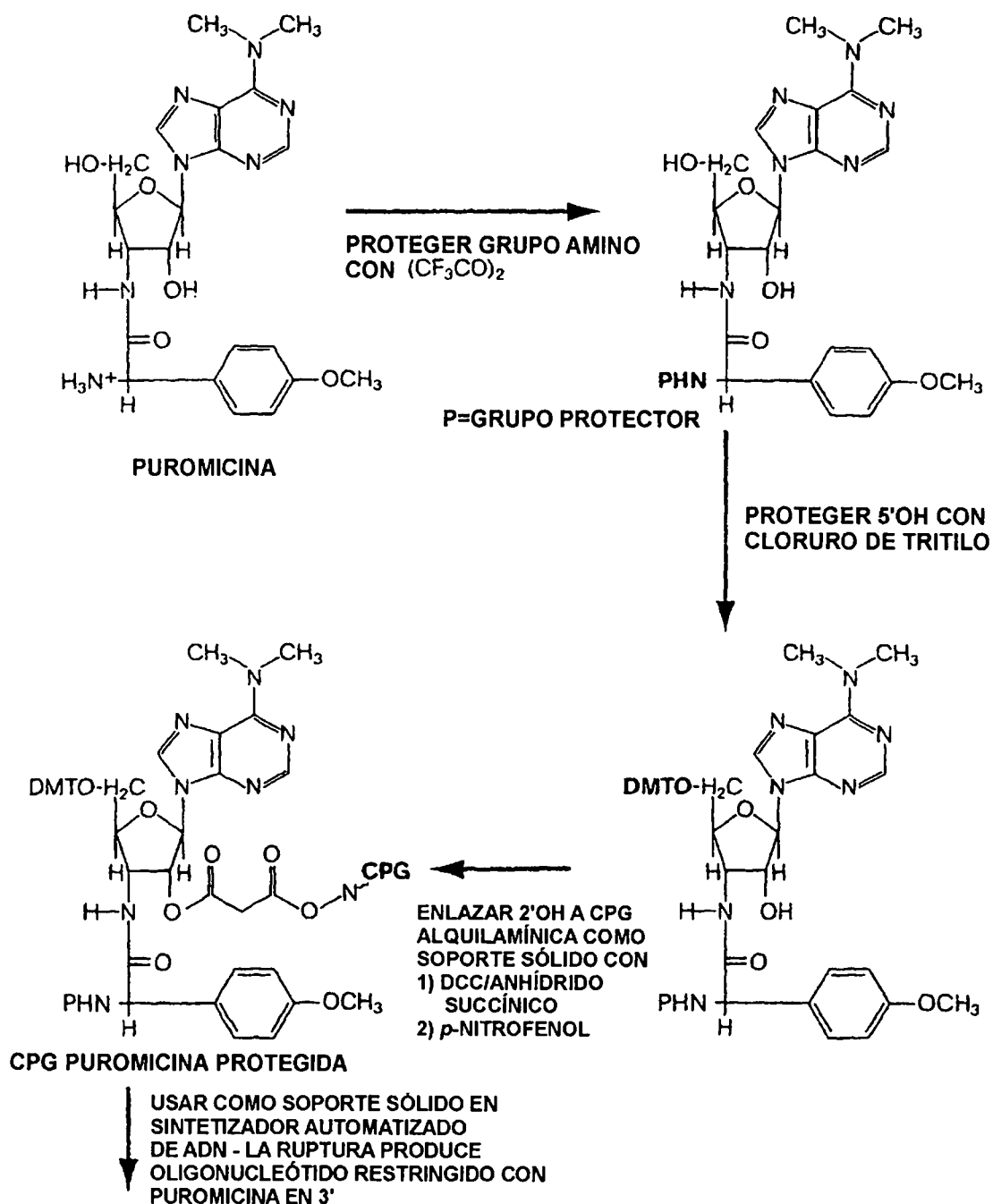


Fig. 4

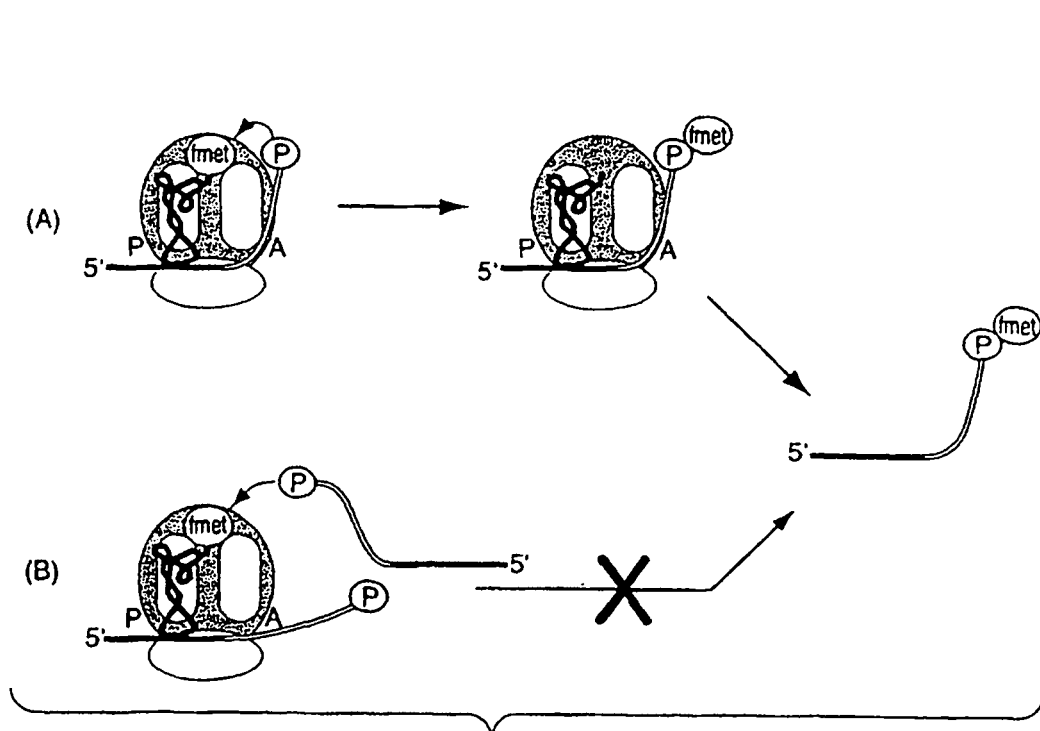


Fig. 5

LÍNEA	1	2	3	4	5	6
Mg ²⁺ [mM]	6	9	12	15	18	18

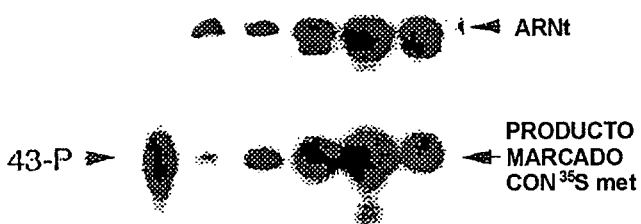


Fig. 6A

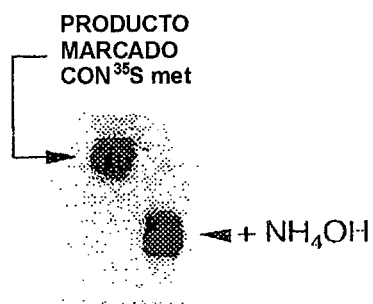


Fig. 6B



Fig. 6C

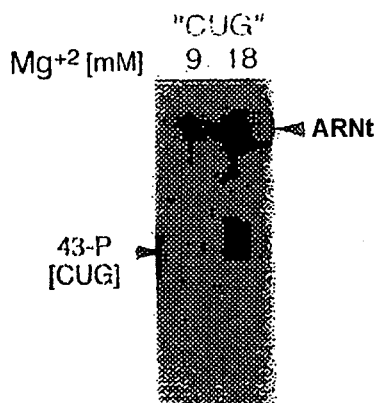


Fig. 6D

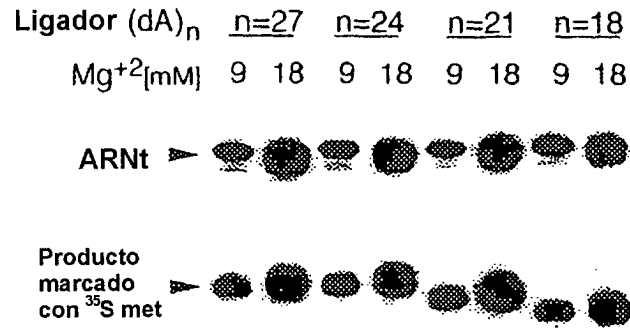


Fig. 6E

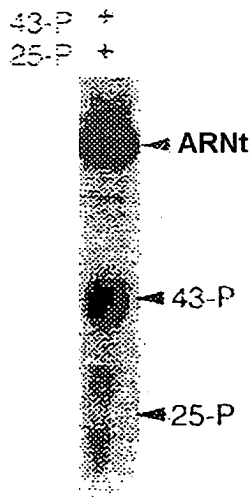


Fig. 6F

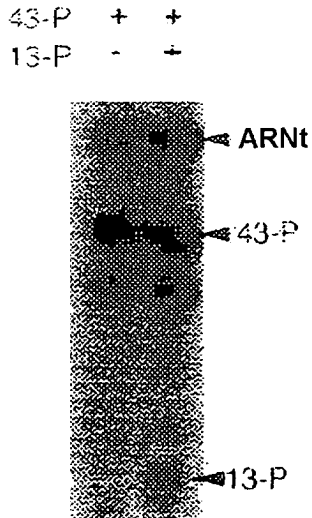


Fig. 6G

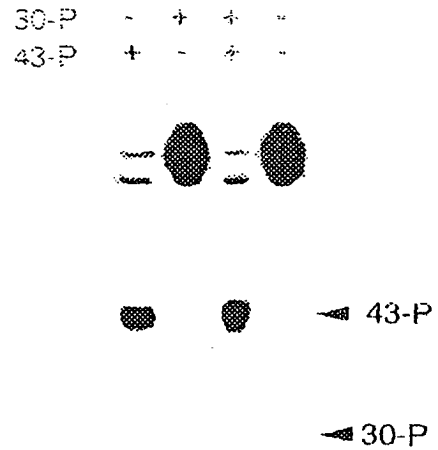


Fig. 6H

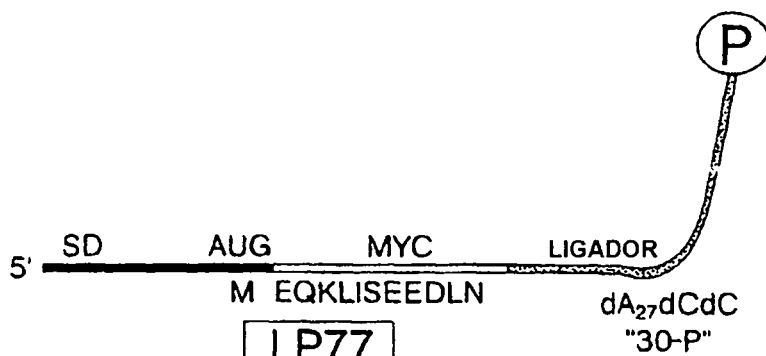


Fig. 7A

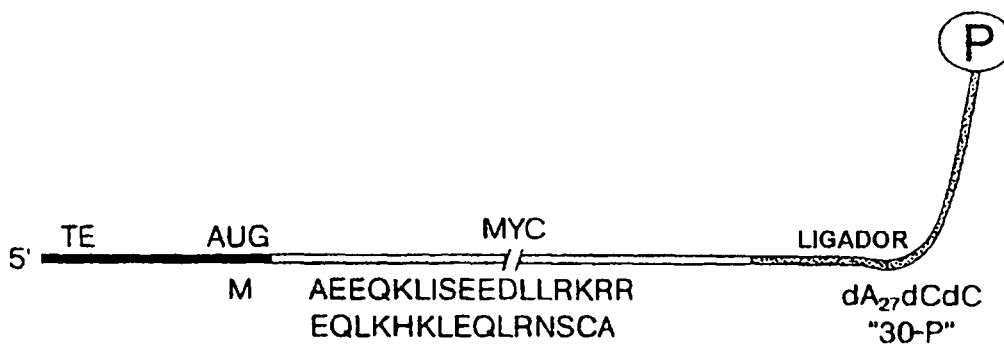
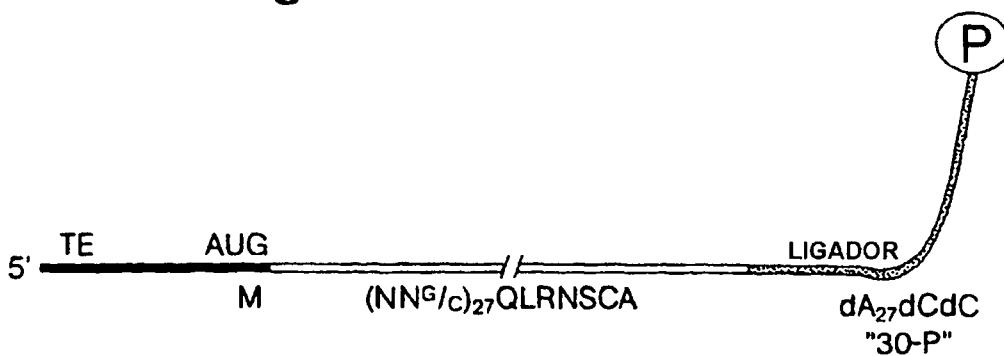


Fig. 7B



CONJUNTO n° 1
Fig. 7C

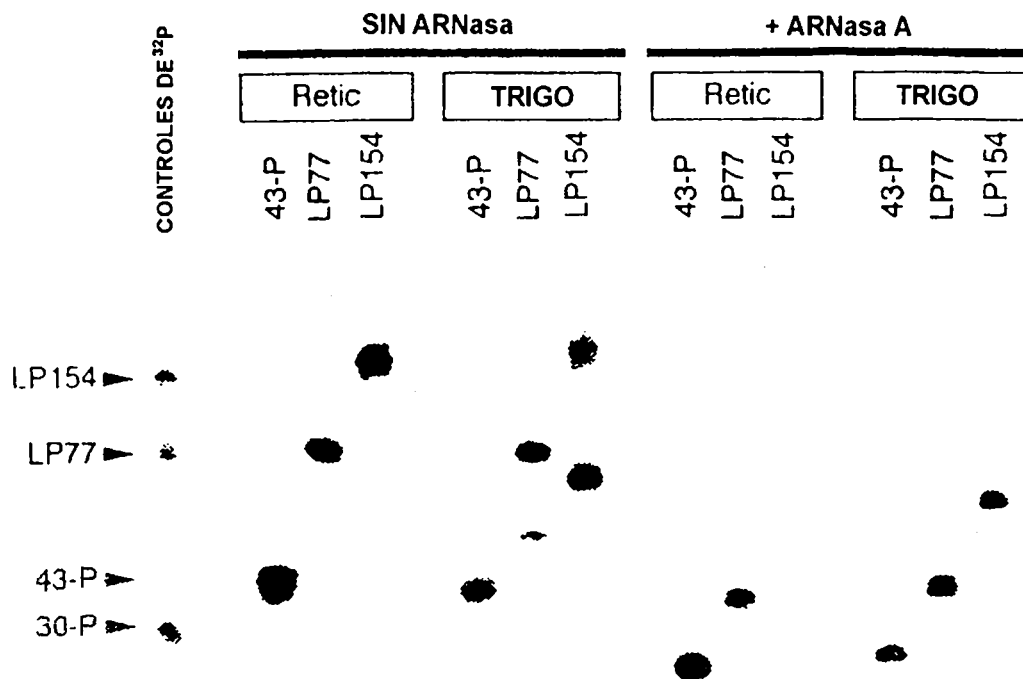


Fig. 8

EFFECTOS DE ARNasa A Y PROTEÍNA K SOBRE LA FUSIÓN

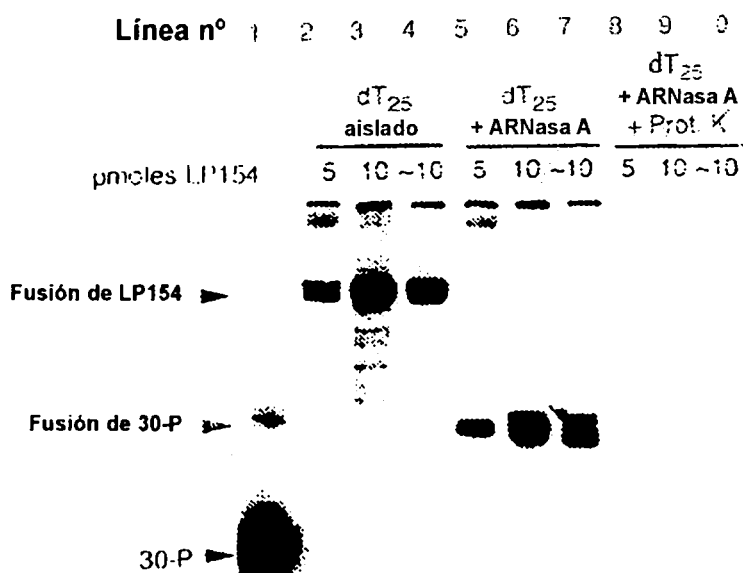


Fig. 9

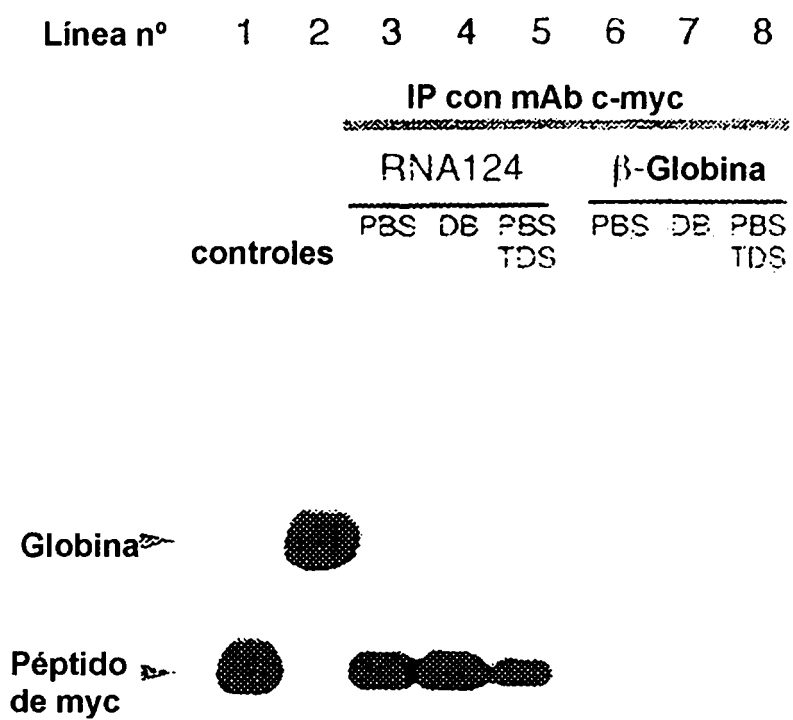


Fig. 10

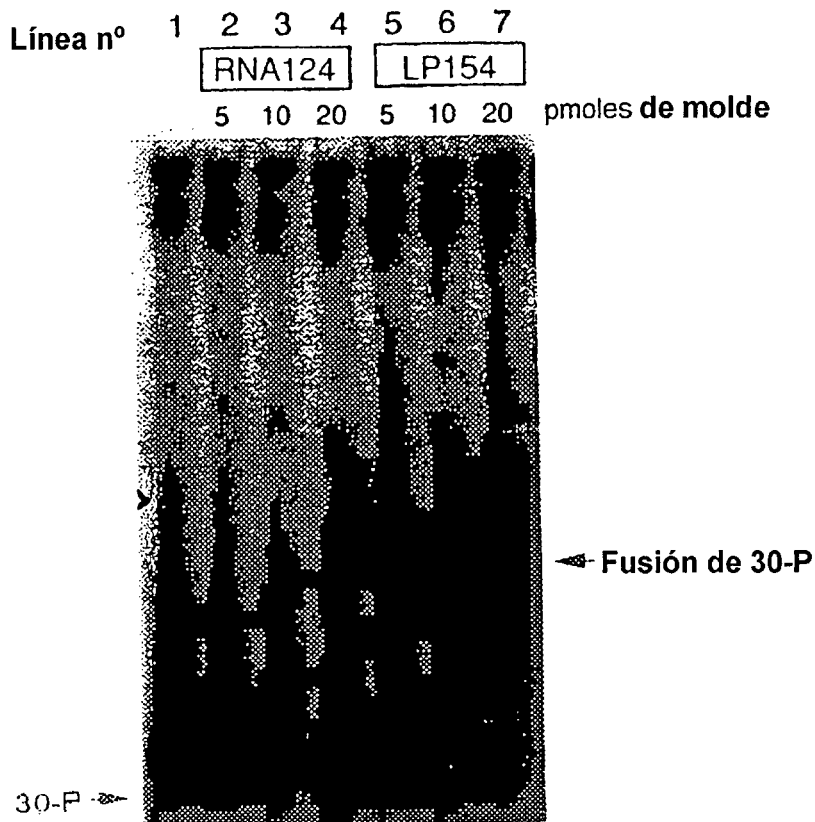


Fig. 11

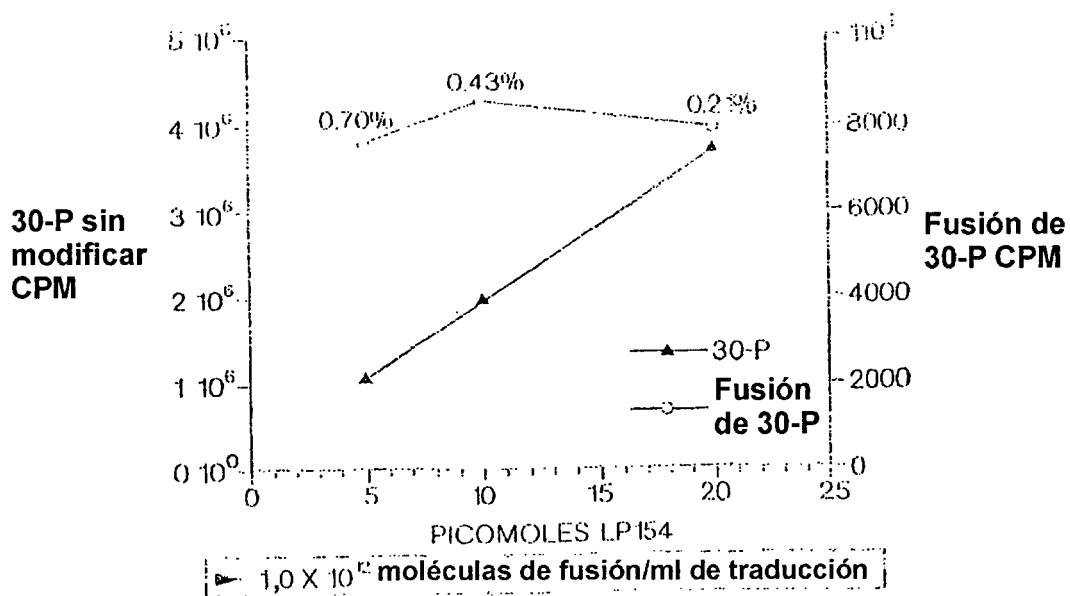


Fig. 12

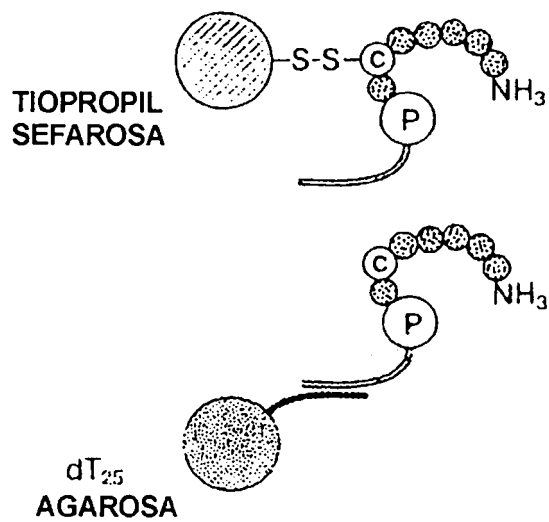


Fig. 13

Línea nº 1 2 3

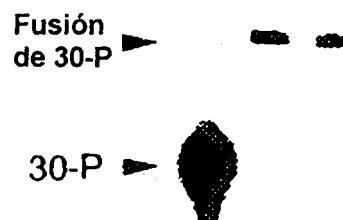


Fig. 14

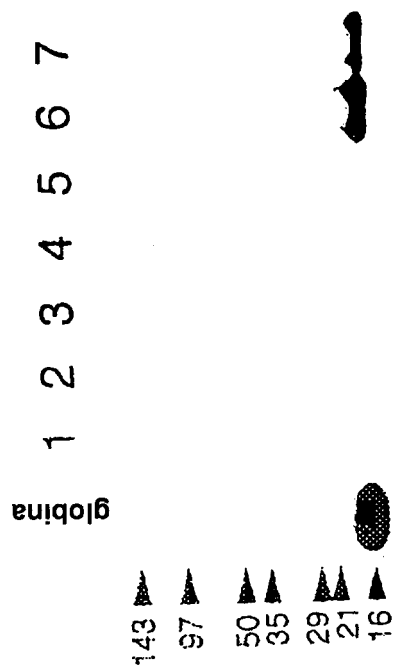


Fig. 15B

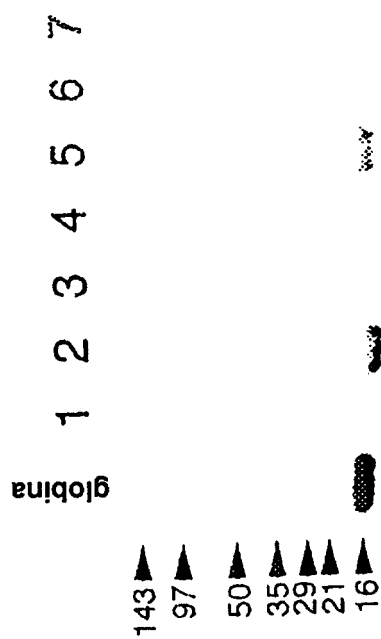


Fig. 15A

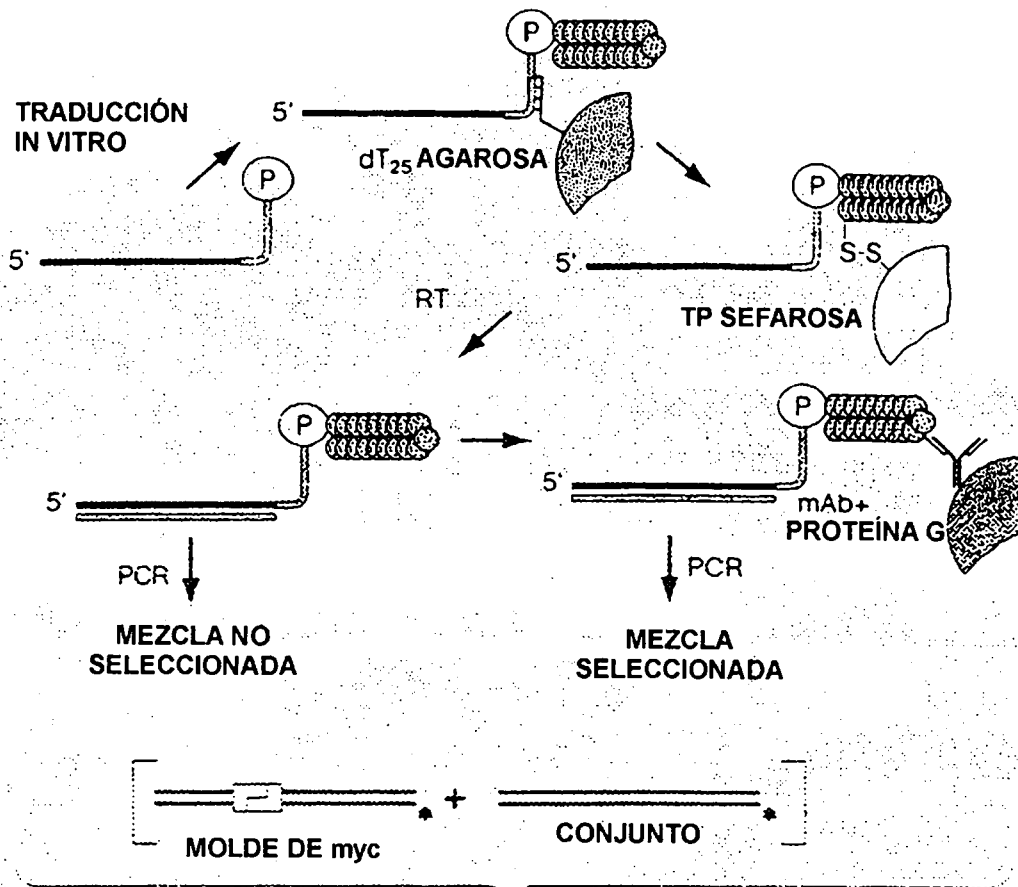


Fig. 16A

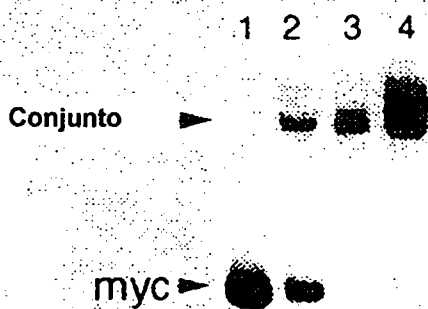


Fig. 16B

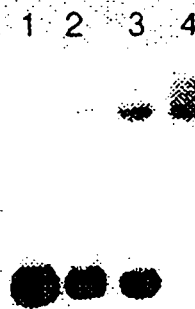


Fig. 16C

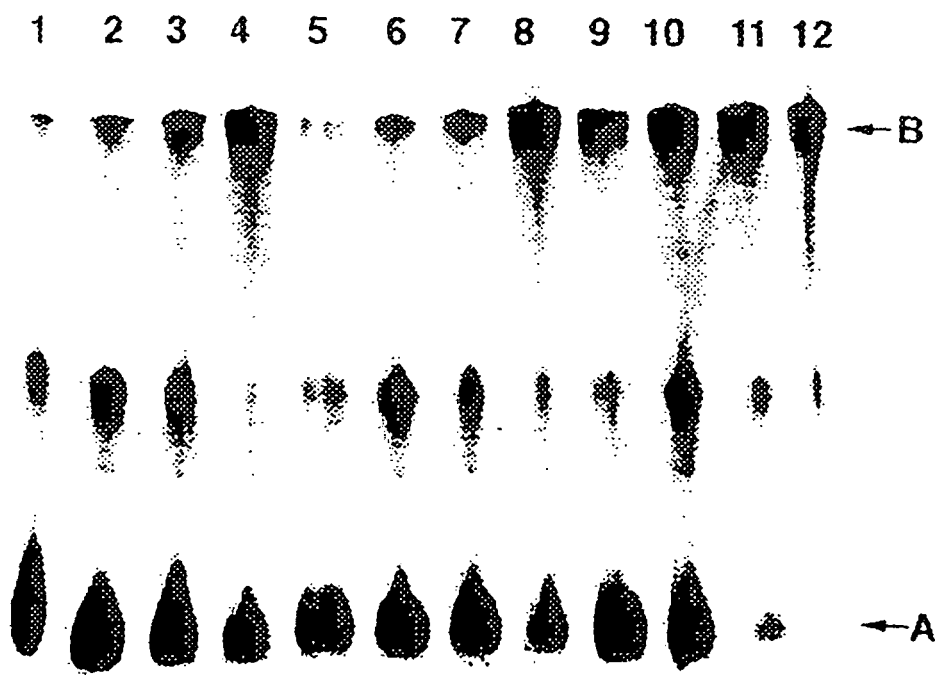


Fig. 17

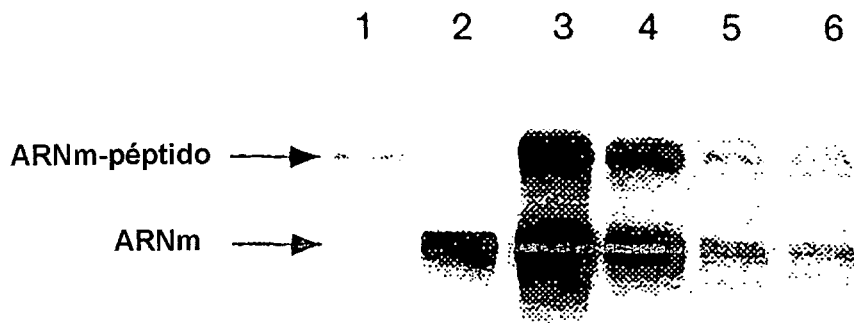


Fig. 18

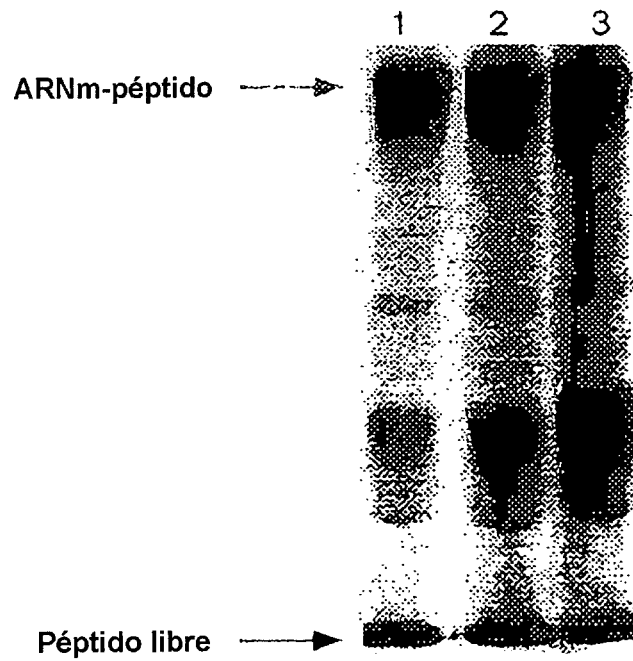


Fig. 19

ES 2 268 763 T3

LISTA DE SECUENCIAS

- <110> Szostak, Jack W.
Roberts, Richard W.
Liu, Rihe
- <120> SELECCIÓN DE PROTEÍNAS USANDO FUSIONES DE ARN-PROTEÍNA
- <130> C29610EP
- <150> 60/035,963
<151> 21/02/1997
- <150> 60/064,491
<151> 06/11/1997
- <160> 33
- <170> Windows NT, Word 97
- <210> 1
<211> 123
<212> ARN
<213> Secuencia artificial
- <220>
<223> Molde de traducción
- <400> 1
- rgrgrgrarg rgrarcrga rararurgrg rararcrag rarararcru rgrarurcru 60**
rcrurgrara rgrarargra rcrcrurgra rarcaaaaaa aaaaaaaaaa aaaaaaaaaa 120
acc 123
- <210> 2
<211> 10
<212> PRT
<213> *Homosapiens*
- <400> 2
- Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10
- <210> 3
<211> 277
<212> ARN
<213> Secuencia artificial
- <220>
<223> Molde de traducción

ES 2 268 763 T3

<210> 7
<211>34
<212> PRT
5 <213> *Homosapiens*

<400> 7
10 Met Ala Glu Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Leu Arg Lys
1 5 10 15
Arg Arg Glu Gln Lys Leu Lys His Lys Leu Glu Gln Leu Arg Asn Ser
20 25 30
15 Cys Ala

<210> 8
<211> 29
20 <212> ADN
<213> Secuencia artificial

<220>
25 <223> Molde de traducción

<400> 8
30 aaaaaaaaaa aaaaaaaaaa aaaaaaaccc 29

<210> 9
<211> 12
35 <212> ADN
<213> Secuencia artificial

<220>
40 <223> Molde de traducción

<400> 9
45 aaaaaaaaaa cc 12

<210> 10
<211> 24
50 <212> ADN
<213> Secuencia artificial

<220>
55 <223> Molde de traducción

<400> 10
60 cgcggttttt attttttt ttcc 24

<210> 11
65 <211> 55
<212> ARN
<213> Secuencia artificial

ES 2 268 763 T3

	<220>		
	<223> Molde de traducción		
5	<400> 11		
	rgrgrargrg rarcgrara rarurugaa aaaaaaaaaa aaaaaaaaaa aaacc		55
10	<210> 12		
	<211> 55		
	<212> ARN		
	<213> Secuencia artificial		
15	<220>		
	<223> Molde de traducción		
20	<400> 12		
	rgrgrargrg rarcgrara rcrurgaaa aaaaaaaaaa aaaaaaaaaa aaacc		55
25	<210> 13		
	<211> 55		
	<212> ARN		
	<213> Secuencia artificial		
30	<220>		
	<223> Molde de traducción		
35	<400> 13		
	rgrgrargrg rarcgrara rarurgaaa aaaaaaaaaa aaaaaaaaaa aaacc		55
40	<210> 14		
	<211> 49		
	<212> ARN		
	<213> Secuencia artificial		
45	<220>		
	<223> Molde de traducción		
50	<400> 14		
	rgrgrargrg rarcgrara rcrurgaaaa aaaaaaaaaa aaaaaaaacc		49
55	<210> 15		
	<211> 46		
	<212> ARN		
60	<213> Secuencia artificial		
	<220>		
	<223> Molde de traducción		
65	<400> 15		
	rgrgrargrg rarcgrara rcrurgaaaa aaaaaaaaaa aaaacc		46

ES 2 268 763 T3

<p><210> 16 <211> 43 <212> ARN 5 <213> Secuencia artificial</p> <p><220> <223> Molde de traducción 10</p> <p><400> 16</p> <p style="padding-left: 40px;">rgrgrargrg rarcgrara rcrurgaaaa aaaaaaaaaa acc</p> <p>15</p> <p><210> 17 <211> 289 <212> ARN 20 <213> Secuencia artificial</p> <p><220> <223> Molde de traducción 25</p> <p><221> característica misc. <222> (1)...(289) <223> n = A,T,C o G 30</p> <p><400> 17</p> <p>35</p> <p style="padding-left: 40px;">rgrgrgrarc rararurura rcruraruru rurarcra rururarcra rarurgrnrn 60</p> <p style="padding-left: 40px;">rsnrnrsrn rnrnrnrns rnrnrsrnrn rsnrnrsrn rnrnrnrns rnrnrsrnrn 120</p> <p style="padding-left: 40px;">rsnrnrsrn rnrnrnrns rnrnrsrnrn rsnrnrsrn rnrnrnrns rnrnrsrnrn 180</p> <p>40</p> <p style="padding-left: 40px;">rsnrnrsrn rnrnrnrns rnrnrsrnrn rsnrnrsrc rargrcrurg rcrgrurara 240</p> <p style="padding-left: 40px;">rcrurcruru rgrcrgrcru aaaaaaaaaa aaaaaaaaaa aaaaaaacc 289</p> <p>45</p> <p><210> 18 <211> 64 <212> ADN <213> <i>Homosapiens</i> 50</p> <p><400> 18</p> <p style="padding-left: 40px;">gttcaggtct tcttgagaga tcagttctg ttccatttcg tctccctat agtgagtcgt</p> <p>55</p> <p style="padding-left: 40px;">atta</p> <p><210> 19 <211> 18 <212> ADN <213> <i>Homosapiens</i> 60</p> <p><400> 19 65</p> <p style="padding-left: 40px;">taatacgact cactatag</p>	<p>43</p> <p>60</p> <p>120</p> <p>180</p> <p>240</p> <p>289</p> <p>60</p> <p>64</p> <p>18</p>
---	---

ES 2 268 763 T3

	<210> 25	
	<211> 49	
	<212> ADN	
5	<213> Secuencia artificial	
	<220>	
	<223> Cebadores para el conjunto de ARN	
10	<400> 25	
	ccctgtaat gataaatgtt aatgttacgt cgacgcattg agataccga	49
15	<210> 26	
	<211> 42	
	<212> ADN	
20	<213> Secuencia artificial	
	<220>	
	<223> Cebadores para el conjunto de ARN	
25	<400> 26	
	taatacgact cactataggg acaattacta ttacaatta ca	42
30	<210> 27	
	<211> 21	
	<212> ADN	
35	<213> Secuencia artificial	
	<220>	
	<223> Cebadores para el conjunto de ARN	
40	<400> 27	
	agcgcaagag ttacgcagct g	21
45	<210> 28	
	<211> 19	
	<212> DNA	
50	<213> Secuencia artificial	
	<220>	
	<223> Tablilla de ADN	
55	<400> 28	
	tttttttt agcgcaaga	19
60	<210> 29	
	<211> 18	
	<212> ADN	
65	<213> <i>Homosapiens</i>	

ES 2 268 763 T3

	<400> 29		
	gtggtattg tgagccag		18
5	<210> 30		
	<211> 40		
	<212> ADN		
10	<213> Fago T7		
	<400> 30		
15	taatacgaact cactataggg acactgctt ttgacacaac		40
	<210> 31		
	<211> 20		
20	<212> ADN		
	<213> Secuencia artificial		
	<220>		
25	<223> Tablilla de ADN		
	<400> 31		
30	ttttttttt gtggtattg		20
	<210> 32		
	<211> 248		
35	<212> ARN		
	<213> <i>Homosapiens</i>		
	<400> 32		
40	rgrgrgrarc rararurura rcruraruru rurarcraa rururarcra rarurgrgrc		60
	rurgrararg rararcrag rarararcru rgrarurcru rcrurgrara rgrarargra		120
45	rcrcrurgrc rurgrcrgu rarararcrg rurcrgurg rararcrag rcrurgrara		180
	rarcraara rararcrug rgrararcra rgrcruurgc rgrurararc rurcrururg		240
	rcrgrcru		248
50	<210> 33		
	<211> 19		
	<212> ADN		
55	<213> Secuencia artificial		
	<220>		
60	<223> Tablilla de ADN		
	<400> 33		
65	ttttttttt agcgcaaga		19