

EXPRESIÓN RECOMBINANTE DEL FACTOR DE CRECIMIENTO DERIVADO DE MIELOIDES

La presente invención se refiere en general al campo de la expresión de genes recombinantes en células huésped. En particular, la invención se refiere a una proteína humana recombinante del factor de crecimiento derivado de mieloides (MYDGF) que presenta un grado mínimo de degradación tras su expresión en una célula huésped. La proteína recombinante es, por lo tanto, muy adecuada para uso médico, en particular para tratar daños en el tejido cardíaco y prevenir la muerte celular en el tejido miocárdico. La invención también proporciona un ácido nucleico que codifica la proteína recombinante y una célula huésped que expresa la proteína recombinante. La invención también proporciona un método para producir la proteína recombinante en una célula huésped.

ANTECEDENTES DE LA INVENCION

El infarto agudo de miocardio (IM) sigue siendo una de las principales causas de morbilidad y mortalidad en todo el mundo. El IM agudo se produce por la oclusión trombótica de una arteria coronaria, lo que provoca la muerte celular progresiva del tejido no perfundido. Esto desencadena una respuesta inflamatoria que conduce a la formación de cicatrices y a la pérdida de tejido viable. Una alteración grave de la arquitectura tisular en el ventrículo izquierdo puede causar dilatación de la cámara, disfunción contráctil e insuficiencia cardíaca. Una proteína denominada factor de crecimiento derivado de los mieloides (MYDGF) ha demostrado mejorar la reparación tisular y la función cardíaca en modelos roedores de IM. En comparación con los ratones de tipo salvaje, los ratones deficientes en MYDGF desarrollan cicatrices de infarto más grandes y disfunciones contráctiles más graves. Se descubrió que el tratamiento con MYDGF recombinante es capaz de proteger a los cardiomiocitos de la muerte celular y reparar el corazón tras un IM agudo. El desarrollo de una terapia basada en proteínas sería un enfoque prometedor para la reparación cardíaca y potencialmente también para la reparación isquémica en otros tejidos (Ebenhoch et al., 2019; Polten et al., 2019; Botnov et al., 2018, Korf-Klingebiel et al., 2015, WO 2014/111458).

En la actualidad, se producen pequeñas cantidades de MYDGF humano recombinante mediante expresión en sistemas de expresión de humanos o de mamíferos, como las células HEK-293T o CHO. Sin embargo, la producción a mayor escala de

MYDGF humano recombinante (rhMYDGF) mediante sistemas de expresión en células de mamífero está asociada a unos costos elevados que hacen que la producción resulte poco atractiva desde el punto de vista económico. Se ha intentado producir rhMYDGF en sistemas de expresión bacteriana. Zhao et al., 2020 describen la expresión de rhMYDGF soluble con una etiqueta His C-terminal en *E. coli*. La proteína marcada difiere de la proteína humana madura de tipo salvaje en 9 aminoácidos adicionales, por lo que tiene un peso molecular significativamente mayor. Aunque los autores especulan en esa publicación que el sistema de expresión podría utilizarse para producir rhMYDGF para uso clínico, el His-Tag plantearía un riesgo significativo de antigenicidad cuando se administrara a un paciente humano.

También se sabe que la expresión heteróloga de rhMYDGF está asociada a problemas considerables de degradación. Tras la expresión recombinante de la proteína, uno o más aminoácidos situados en el extremo N-terminal se degradan, lo que da lugar a fragmentos de proteína que tienen un peso molecular (PM) más pequeño en comparación con la proteína de longitud completa. Por ejemplo, Zhao et al., 2020 describen que el producto de expresión final comprende no solo la proteína rhMYDGF de longitud completa que tiene un MW de 17032 Da, sino también un producto de degradación con un MW de aproximadamente 16900 Da. La relación entre la proteína diana y la proteína degradada en Zhao et al., 2020 es de aproximadamente 10:1, como puede deducirse de los datos de cromatografía líquida de alto rendimiento-espectrometría de masas (HPLC-MS) mostrados en la Fig. 2 de la publicación de Zhao. Por lo tanto, la impureza por el producto de degradación es bastante significativa y no es aceptable para una proteína destinada a uso médico sistémico.

En vista del estado de la técnica expuesto con anterioridad, es objeto de la presente invención proporcionar una proteína recombinante con actividad MYDGF que

- (i) no presenta degradación o presenta una degradación mínima tras su expresión en un sistema de expresión heteróloga;
- (ii) puede producirse con un grado mínimo de modificaciones postraduccionales potencialmente adversas derivadas del proceso, como la carbamoilación y la gluconoilación, que pueden ser perjudiciales para el uso clínico previsto de la proteína;
- (iii) tiene la estructura secundaria y terciaria de la proteína MYDGF humana nativa;
- (iv) tiene una elevada actividad biológica;

- (v) se asocia con un bajo riesgo de epítomos antigénicos derivados de secuencias primarias distintas del MYDGF humano, creando así un riesgo mínimo de anticuerpos antifármaco;
- (vi) puede producirse en un sistema de expresión procariota para proporcionar un producto no glicosilado;
- (vii) puede producirse en una cantidad superior a 0,5 g de proteína por 100 g de células y es escalable a más de 100 g, preferentemente más de 200 g, más preferentemente más de 300 g de proteína por lote.

No todos los objetivos se alcanzarán con todas las formas de realización de la invención. El alcance de la invención se define en las reivindicaciones. Sin embargo, se prefiere cumplir 2, 3, 4, 5 o 6 de los objetivos de la invención mencionados con anterioridad.

SÍNTESIS DE LA INVENCION

En un primer aspecto, la invención se refiere a un método para la expresión recombinante de una proteína MYDGF en una célula huésped.

En un segundo aspecto, la invención se refiere a una composición obtenible a partir del método del primer aspecto de la invención.

En un tercer aspecto, la invención se refiere al uso de una composición del segundo aspecto de la invención para la preparación de una composición farmacéutica.

En un cuarto aspecto, la invención se refiere a una composición farmacéutica que comprende una composición del segundo aspecto de la invención.

En un quinto aspecto, la invención se refiere a una composición farmacéutica del cuarto aspecto de la invención para su uso como medicamento.

En un sexto aspecto, la invención se refiere a una composición del segundo aspecto de la invención o una composición farmacéutica del cuarto aspecto de la invención para su uso en un método de (i) tratamiento o prevención de una enfermedad o condición seleccionada del grupo que consiste en lesión, herida, isquemia, lesión por reperfusión, traumatismo, sobrecarga mecánica, intoxicación, cirugía, miocardiopatía primaria o adquirida, disfunción contráctil posisquémica, infarto de miocardio, preferentemente infarto agudo de miocardio, angina de pecho, fallo cardíaco, inflamación del corazón, insuficiencia cardíaca, hipertrofia y fibrosis; (ii) promover o mejorar la regeneración del

tejido cardíaco, la proliferación de cardiomiocitos, la neovascularización, la función cardíaca o la función sistólica del ventrículo izquierdo tras un infarto de miocardio; (iii) proteger a los cardiomiocitos de la muerte, por ejemplo, por apoptosis o necrosis; o (iv) reducir el tamaño del infarto tras un infarto de miocardio, preferentemente un infarto agudo de miocardio.

En un séptimo aspecto, la invención se refiere a una proteína que tiene la secuencia de aminoácidos de la SEQ ID NO: 1 o SEQ ID NO: 2, y preferentemente SEQ ID NO: 1.

En un octavo aspecto, la invención se refiere a un ácido nucleico que codifica una proteína del séptimo aspecto de la invención.

En un noveno aspecto, la invención se refiere a un vector que comprende el ácido nucleico del octavo aspecto de la invención.

En un décimo aspecto, la invención se refiere a una célula huésped que comprende una proteína del séptimo aspecto de la invención, un ácido nucleico del octavo aspecto de la invención o un vector del noveno aspecto de la invención.

En un decimoprimer aspecto, la invención se refiere a una composición farmacéutica que comprende la proteína del séptimo aspecto de la invención.

En un decimosegundo aspecto, la invención se refiere a una proteína del séptimo aspecto de la invención o a una composición farmacéutica del decimoprimer aspecto de la invención para su uso como medicamento.

En un decimotercer aspecto, la invención se refiere a una proteína del séptimo aspecto de la invención o una composición farmacéutica del decimoprimer aspecto de la invención para su uso en un método de (i) tratamiento o prevención de una enfermedad o condición seleccionada del grupo que consiste en lesión, herida, isquemia, lesión por reperfusión, traumatismo, sobrecarga mecánica, intoxicación, cirugía, miocardiopatía primaria o adquirida, disfunción contráctil posisquémica, infarto de miocardio, preferentemente infarto agudo de miocardio, angina de pecho, fallo cardíaco, inflamación del corazón, insuficiencia cardíaca, hipertrofia y fibrosis; (ii) promover o mejorar la regeneración del tejido cardíaco, la proliferación de cardiomiocitos, la neovascularización, la función cardíaca o la función sistólica del ventrículo izquierdo tras un infarto de miocardio; (iii) proteger a los cardiomiocitos de la muerte, por ejemplo, por apoptosis o necrosis; o (iv) disminuir el tamaño del infarto tras un infarto de miocardio, preferentemente un infarto agudo de miocardio.

En un decimocuarto aspecto, la invención se refiere a un método para la expresión recombinante de una proteína del séptimo aspecto de la invención en una célula huésped.

En un decimoquinto aspecto, la invención se refiere a una composición obtenible a partir del método del decimocuarto aspecto de la invención.

Finalmente, en un decimosexto aspecto, la invención se refiere al uso de una célula huésped del décimo aspecto de la invención para la expresión recombinante de una proteína MYDGF.

DESCRIPCIÓN DETALLADA DE LA INVENCION

La presente invención proporciona proteínas que tienen actividad MYDGF y ejercen un grado mínimo de degradación y modificaciones postraduccionales derivadas del proceso tras la expresión recombinante en una célula huésped. La invención también proporciona un método para producir estas proteínas en grandes cantidades en un sistema de expresión celular. El método requiere esfuerzos de purificación significativamente reducidos para proporcionar una composición proteica homogénea que sea adecuada para ser formulada en un producto farmacéutico. Si en lo sucesivo se hace referencia a la SEQ ID NO: 1 o a la SEQ ID NO: 2, debe entenderse que la SEQ ID NO: 1 es la alternativa preferida.

Las proteínas reveladas en el presente documento en el contexto de la presente invención se representan en la SEQ ID NO: 1 y SEQ ID NO: 2. Ambas proteínas constan de 143 aminoácidos y comprenden la secuencia completa de aminoácidos de la proteína MYDGF humana madura. La proteína MYDGF humana nativa se expresa como una proteína precursora con un péptido señal N-terminal de 31 aminoácidos y una secuencia de retención en el retículo endoplasmático (RE) C-terminal tipo KDEL. La secuencia de la proteína precursora de MYDGF de 173 aminoácidos se establece en el presente documento como SEQ ID NO: 5. Tras la escisión del péptido señal N-terminal, se libera el MYDGF maduro. La secuencia de la proteína MYDGF humana madura consta de 142 aminoácidos y se presenta aquí como SEQ ID NO: 6. Las proteínas de la invención difieren de la MYDGF madura solo en un único aminoácido que se ha añadido a su N-terminal. Por lo tanto, estas proteínas pueden considerarse variantes recombinantes de la proteína MYDGF humana nativa. La proteína SEQ ID NO: 1 contiene un residuo adicional de alanina en su N-terminal, que no está presente en la proteína MYDGF humana madura. Esta variante se

denomina “[+A]” o “la variante [+A]” en lo sucesivo. La proteína de la SEQ ID NO: 2 difiere de la proteína MYDGF humana madura por un residuo de serina adicional en su N-terminal. Esta variante se denomina en lo sucesivo “[+S]” o “la variante [+S]”.

Las proteínas se asocian con un riesgo particularmente bajo de comprender epítomos antigénicos que no se derivan del MYDGF humano. En consecuencia, las proteínas de la invención presentan un riesgo mínimo de generar anticuerpos antifármaco. Las diferencias entre las proteínas de la invención y la proteína MYDGF humana nativa residen en un único aminoácido, lo que significa que la región añadida a la proteína nativa es demasiado pequeña para dar lugar a nuevos epítomos. La ausencia de anticuerpos antifármaco hace que las proteínas de la invención sean muy adecuadas para ser utilizadas con fines terapéuticos. De preferencia, la administración de las proteínas a seres humanos solo generará un nivel mínimo de anticuerpos antifármaco o anticuerpos dirigidos contra el MYDGF endógeno, y más preferentemente no generará ninguno.

Como puede verse en los Ejemplos siguientes, las proteínas de la invención presentan un bajo grado de modificaciones químicas y postraduccionales tras su expresión en un sistema de expresión celular, es decir, en un sistema de expresión que utiliza células eucariotas o procariotas para la producción recombinante de la proteína. La reducción efectiva de las modificaciones químicas y postraduccionales durante la producción de proteínas para aplicaciones farmacéuticas es de importancia fundamental. En particular, las proteínas de la invención se caracterizan por un bajo grado de carbamoilación y gluconoilación. Tal como se utiliza en el presente documento, la carbamoilación es una reacción no enzimática en donde se añade una fracción de carbamoilo a una proteína, péptido o aminoácido. Tras la expresión en un sistema de expresión celular, el aislamiento de las proteínas a partir de cuerpos de inclusión y el replegamiento de la proteína, la carbamoilación se produce preferentemente en menos del 6,0 % (p/p) de la proteína total de SEQ ID NO: 1, más preferentemente en menos del 5,5 % (p/p), menos del 5,0 % (p/p), menos del 4,5 % (p/p), o menos del 4,0 % (p/p), de la proteína total que tiene la secuencia de aminoácidos de la SEQ ID NO: 1. Del mismo modo, la carbamoilación se produce preferentemente en menos del 6,0 % (p/p) de la proproteína total de SEQ ID NO: 2, más preferentemente en menos del 5,5 % (p/p), menos del 5,0 % (p/p), menos del 4,5 % (p/p), o menos del 4,0 % (p/p), de la proteína total que tiene la secuencia de aminoácidos de la SEQ ID NO: 2 tras la expresión de la proteína en un sistema de expresión celular, el

aislamiento de las proteínas a partir de cuerpos de inclusión y el replegamiento de la proteína. En general, un nivel de carbamoilación inferior al 6,0 % (p/p) es aceptable y no supone un riesgo para las aplicaciones farmacéuticas.

Además, las proteínas reveladas en el presente documento se caracterizan por un bajo grado de gluconoilación. La gluconoilación de proteínas expresadas recombinantemente se observa regularmente en células huésped bacterianas, como en células de *E. coli* BL21(DE3). Esta modificación resulta de la formación de 6-fosfogluconolactona (6-PGLac), un compuesto que es producido por la enzima glucosa-6-fosfato deshidrogenasa. Con las proteínas de la invención, la gluconoilación se produce preferentemente después de la expresión en un sistema de expresión celular, el aislamiento de la proteína a partir de cuerpos de inclusión y el replegamiento de la proteína solo en menos del 4,0 % (p/p) de la proteína total de SEQ ID NO: 1, más preferentemente en menos del 3,5 % (p/p), menos del 3,0 % (p/p), menos del 2,5 % (p/p), o menos del 2,0 % (p/p), de la proteína total de SEQ ID NO: 1. La gluconoilación se produce preferentemente después de la expresión en un sistema de expresión celular, el aislamiento de la proteína a partir de cuerpos de inclusión y el replegamiento de la proteína. Del mismo modo, la carbamoilación se produce preferentemente en menos del 4,0 % (p/p) de la proteína total de SEQ ID NO: 2, más preferentemente en menos del 3,5 % (p/p), menos del 3,0 % (p/p), menos del 2,5 % (p/p), o menos del 2,0 % (p/p), de la proteína total de SEQ ID NO: 2. La carbamoilación se produce preferentemente en menos del 3,5 % (p/p), menos del 3,0 % (p/p), menos del 2,5 % (p/p), o menos del 2,0 % (p/p), de la proteína total de SEQ ID NO: 2. En general, un nivel de gluconoilación inferior al 4 % es aceptable y no supone un riesgo para las aplicaciones farmacéuticas.

Otra forma de realización de la invención se refiere a una proteína que tiene la secuencia de aminoácidos de la SEQ ID NO: 1 y además tiene un espectro en la espectroscopia de resonancia magnética nuclear bidimensional (2D-RMN) que es esencialmente idéntico al que se muestra en la Tabla 1 siguiente. En particular, la proteína tiene un espectro de RMN que comprende al menos 2 de los picos ^1H y/o ^{15}N , y preferentemente al menos 4, al menos 6, al menos 8, al menos 10, al menos 12, al menos 14, al menos 16, al menos 18, al menos 20, al menos 22, al menos 24, al menos 26 al menos 28, al menos 30, al menos 32, al menos 34, al menos 36, al menos 38, al menos 40, al menos 42, al menos 44, al menos 46, al menos 48, al menos 50, al menos 52, al menos 54, al menos

56, al menos 58, al menos 60, al menos 62, al menos 64, al menos 66, al menos 68, al menos 70, al menos 72, al menos 74, al menos 76, al menos 78, al menos 80, al menos 82, al menos 84, al menos 86, al menos 88, al menos 90, al menos 92, al menos 94, al menos 96, al menos 98, al menos 100, al menos 102, al menos 104, al menos 106, al menos 108, al menos 110, al menos 112, al menos 114, al menos 116, al menos 118, al menos 120, al menos 122, al menos 124, al menos 126, al menos 128, al menos 130, al menos 132, o al menos 134 picos de los picos ^1H y/o ^{15}N de la Tabla 1 al analizar una muestra de 8,5 mg/ml de la proteína MYDGF respectiva en tampón fosfato sódico 50 mM a pH 7,4 que contiene cloruro de sodio 50 mM y D_2O al 9 % (v/v).

Más preferentemente, la proteína tiene un espectro de RMN que comprende o consiste en los 136 picos de ^1H y/o ^{15}N establecidos en la Tabla 1 cuando se analiza una muestra de 8,5 mg/ml de la proteína MYDGF respectiva en tampón fosfato sódico 50 mM a pH 7,4 que contiene cloruro de sodio 50 mM y 9 % (v/v) de D_2O . Así, la proteína tiene la estructura secundaria y terciaria de la proteína MYDGF humana nativa.

En otra forma de realización más, la presente invención se refiere a una proteína que está plegada de tal manera que más del 70 %, y preferentemente más del 80 %, más del 90 %, o más del 95 %, de los picos ^1H y/o ^{15}N en el mapa 2D–RMN resultan en valores de desviación de desplazamiento químico combinado (CCSD) por debajo de 0,01 ppm cuando se comparan con los picos correspondientes en la Tabla 1. La CCSD se calcula de acuerdo con la siguiente fórmula (Brinson et al. 2019):

$$\text{CCSD} = \sqrt{\frac{(\delta_H - \delta_{H\text{ref}})^2 + \left(\frac{\delta_N - \delta_{N\text{ref}}}{10}\right)^2}{2}}$$

en donde δ_H y δ_N representan los desplazamientos químicos ^1H y ^{15}N de un pico cruzado dado, respectivamente, y $\delta_{H\text{ref}}$ y $\delta_{N\text{ref}}$ representan los desplazamientos químicos ^1H y ^{15}N de referencia para el mismo pico cruzado. Si más del 70 %, y preferentemente más del 80 %, más del 90 %, o más del 95 %, de los picos ^1H y/o ^{15}N en el mapa 2D–RMN conducen a valores CCSD por debajo de 0,01 ppm, el plegamiento de la proteína es muy similar al observado para la variante [+A] en la Tabla 1. Específicamente, si más del 90 %, o más del 95 % de los picos ^1H y/o ^{15}N en el mapa 2D–RMN ejercen valores CCSD por debajo de

0,01 ppm, el plegamiento de la proteína es casi idéntico al observado para la variante [+A] en la Tabla 1.

Tabla 1: Espectro RMN [+A]

N.º de pico	Desplazamiento químico ^1H (ppm)	Desplazamiento químico ^{15}N (ppm)
1	10,37	129,76
2	10,14	129,36
3	9,60	129,47
4	9,63	128,50
5	9,49	128,58
6	9,13	129,31
7	8,87	129,18
8	9,04	128,69
9	9,02	128,32
10	9,48	126,84
11	9,20	126,47
12	8,95	126,81
13	8,79	127,05
14	8,48	127,45
15	8,14	128,38
16	7,84	129,41
17	8,81	125,63
18	8,70	125,75
19	8,25	126,85
20	9,20	125,74
21	9,30	124,99
22	9,03	124,81
23	8,86	124,98
24	8,81	124,67
25	8,20	125,80
26	7,88	125,80
27	8,44	124,83
28	8,36	124,50
29	7,72	124,72
30	7,60	123,64
31	7,63	122,68

32	7,42	122,77
33	7,91	122,40
34	8,38	122,53
35	8,37	123,27
36	8,50	123,58
37	8,60	123,33
38	8,65	123,98
39	8,77	123,34
40	8,72	122,90
41	8,89	122,88
42	9,00	121,91
43	8,27	121,54
44	9,07	122,08
45	9,12	121,13
46	9,39	121,60
47	9,33	120,96
48	9,41	120,69
49	9,03	120,73
50	8,37	120,75
51	8,49	120,29
55	9,08	119,50
56	8,82	119,49
57	8,77	119,60
58	8,45	119,33
59	7,99	119,50
60	8,06	120,13
61	8,16	120,20
62	8,30	119,98
63	8,30	119,68
64	8,30	118,69
65	7,66	121,40
66	7,57	121,52
67	7,22	121,78
68	7,17	121,00
69	7,47	119,60
70	7,34	119,67
71	7,26	119,24
72	7,17	119,38

73	7,33	118,42
74	6,68	120,50
75	6,15	117,12
76	7,71	117,75
77	8,02	117,15
78	8,09	117,30
79	8,18	117,90
80	8,19	117,22
81	8,42	118,42
82	8,52	117,35
83	8,56	116,70
84	8,26	116,36
85	8,67	117,13
86	8,70	116,53
87	8,77	117,66
88	8,77	116,10
89	8,28	115,50
90	8,22	115,01
91	7,91	115,17
92	7,88	115,81
93	7,66	115,39
94	7,65	113,72
95	8,62	114,63
96	8,59	113,55
97	9,94	115,99
98	9,26	118,39
99	9,08	118,84
100	8,73	116,99
101	7,23	115,12
102	6,64	115,15
103	6,67	113,72
104	7,53	112,93
105	6,82	112,88
106	6,68	112,93
107	7,38	112,88
108	7,36	112,79
109	7,77	111,93
110	7,97	112,54

111	8,40	112,70
112	8,33	111,68
113	8,48	111,24
114	7,36	111,36
115	6,72	111,34
116	7,29	110,79
117	6,58	110,76
118	6,77	109,87
119	7,99	108,82
120	8,84	110,91
121	8,65	109,37
122	8,58	109,47
123	8,02	107,16
124	7,70	106,72
125	7,73	102,74
126	10,67	128,23
127	9,77	127,55
128	9,22	130,73
129	6,25	123,98
130	10,09	132,99
131	9,45	131,92
132	9,02	132,15
133	9,11	123,54
134	9,96	122,02
135	6,21	109,88
136	6,58	111,93

Otra forma de realización de la invención es una composición que comprende una proteína como se describió con anterioridad, preferentemente una composición obtenible a partir de la expresión recombinante en un sistema de expresión bacteriano, como los métodos descritos con más detalle a continuación.

La composición de la invención puede comprender una proteína que tiene la secuencia de aminoácidos de la SEQ ID NO: 1 junto con variantes de la misma que son más cortas en longitud y ejercen un 100 % de identidad de secuencia en toda su longitud con la secuencia de aminoácidos de la SEQ ID NO: 1, en donde la longitud es de al menos 100 aminoácidos sin que se permitan huecos en la alineación. Dentro de dicha composición,

la relación de la señal obtenida a partir de la proteína según la SEQ ID NO: 1 y la suma de las señales obtenidas a partir de dichas variantes más cortas, determinada por espectrometría de masas por cromatografía líquida (LCMS) según Tolonen et al (2011), es superior a 20, y preferentemente superior a 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, y más preferentemente superior a 450. Como puede observarse en la Tabla 15, la relación de la señal obtenida a partir de la proteína según la SEQ ID NO: 1 y las señales obtenidas a partir de variantes más cortas resultó ser 466, como se determinó por LCMS según Tolonen et al (2011). Al calcular la relación entre la señal obtenida a partir de la proteína según la SEQ ID NO: 1 y la suma de las señales obtenidas a partir de dichas variantes más cortas, se excluyen las proteínas carbamoyladas o gluconoyladas. El porcentaje para el cálculo de la relación se basa en la suma de las intensidades de pico de la proteína MYDGF no modificada, así como de las especies de modificación postraduccional (PTM) anotadas de MYDGF en un espectro de masas intacto deconvolucionado de la proteína MYDGF en la composición. Los requisitos mínimos para la instrumentación de espectrometría de masas y el procesamiento de datos se describen en el Ejemplo 4 a continuación.

La composición de la invención también puede comprender una proteína que tiene la secuencia de aminoácidos de la SEQ ID NO: 2 junto con variantes de la misma que son más cortas en longitud y ejercen un 100 % de identidad de secuencia en toda su longitud con la secuencia de aminoácidos de la SEQ ID NO: 2, en donde la longitud es de al menos 100 aminoácidos sin que se permitan huecos en la alineación. Dentro de dicha composición, la relación entre la señal obtenida a partir de la proteína según la SEQ ID NO: 2 y la suma de señales obtenidas a partir de dichas variantes más cortas, determinada por LCMS según Tolonen et al (2011), es superior a 20, y preferentemente superior a 50, 75, 100, 125, 150, y más preferentemente superior a 175 o 180. Como puede observarse en la Tabla 15 a continuación, la relación de la señal obtenida a partir de la proteína según la SEQ ID NO: 2 y las señales obtenidas a partir de variantes más cortas resultó ser 186, según se determinó por LCMS según Tolonen et al (2011). Al calcular la relación entre la señal obtenida de la proteína según la SEQ ID NO: 2 y la suma de las señales obtenidas de dichas variantes más cortas, se excluyen las proteínas carbamoyladas o gluconoyladas. El porcentaje para el cálculo de la relación se basa en la suma de las intensidades de pico de la proteína MYDGF no modificada, así como de las especies de modificación postraduccional (PTM) anotadas

de MYDGF en un espectro de masas intacto deconvolucionado de la proteína MYDGF en la composición. Los requisitos mínimos para la instrumentación de espectrometría de masas y el procesamiento de datos se describen en el Ejemplo 4 a continuación.

En una forma de realización, menos del 8 %, y preferentemente menos del 7 %, menos del 6 % o menos del 5 % de las proteínas MYDGF en la composición de la invención están carbamoiladas, en donde el porcentaje se basa en la suma de las intensidades de pico de la proteína MYDGF no modificada así como las especies de modificación postraducciona (PTM) anotadas de MYDGF en un espectro de masas intacto deconvolucionado de la proteína MYDGF en la composición. Los requisitos mínimos para la instrumentación de espectrometría de masas y el procesamiento de datos se describen en el Ejemplo 4 a continuación.

En otra forma de realización, menos del 6 %, y preferentemente menos del 5 %, menos del 4 % o menos del 3 % de las proteínas MYDGF en la composición de la invención están gluconiladas, en donde el porcentaje se basa en la suma de las intensidades de pico de la proteína MYDGF no modificada así como las especies PTM anotadas de MYDGF en un espectro de masa intacto deconvolucionado de la proteína MYDGF en la composición. Los requisitos mínimos para la instrumentación de espectrometría de masas y el procesamiento de datos se describen en el Ejemplo 4 a continuación.

De acuerdo con otro objeto de la invención, la composición de la invención comprende preferentemente una cantidad baja de ADN que se deriva de la célula huésped que se utilizó para la producción de la proteína MYDGF. De preferencia, la composición de la invención comprende menos de 20 pg/mg, preferentemente menos de 15 pg/mg, más preferentemente menos de 10 pg/mg, y de mayor preferencia aún menos de 5 pg de ADN de la célula huésped por mg de la composición, como menos de 3 pg, menos de 2 pg o menos de 1 pg de ADN de la célula huésped por mg de la composición. De preferencia, la presencia de ADN de célula huésped en la composición se determina mediante reacción en cadena de la polimerasa cuantitativa (qPCR), tal como qPCR en tiempo real.

De acuerdo con otro objeto de la invención, la composición de la invención comprende preferentemente también solo una cantidad baja de endotoxina bacteriana que resulta de la producción de la proteína MYDGF en las células huésped bacterianas. Específicamente, se prefiere que la composición comprenda menos de 0,2 UE por mg de la composición, y preferentemente menos de 0,1 UE, menos de 0,09 UE o 0,08 UE de

endotoxina bacteriana por mg de la composición. Los métodos adecuados para detectar la presencia de endotoxina bacteriana incluyen el método cromogénico cinético descrito en la actual Farmacopea de los Estados Unidos (USP–NF 2021, edición 2, Capítulo 85), la Farmacopea Europea (10ª edición 2021, 10.5, Capítulo 2.6.14) y la Farmacopea Japonesa, Suplemento II, JP 17ª edición, 4.01).

De acuerdo con otro objeto de la invención, se prefiere que menos del 8 % (p/p), y preferentemente menos del 7 % (p/p), menos del 6 % (p/p) o menos del 5 % (p/p) de las proteínas de la composición de la invención estén carbamoiladas. Del mismo modo, se prefiere que menos del 6 % (p/p), y preferentemente menos del 5 % (p/p), menos del 4 % (p/p) o menos del 3 % (p/p), de las proteínas de la composición de la invención estén gluconoiladas. Se prefiere en particular que la composición de la invención comprenda cantidades detectables de proteínas carbamoiladas, en donde la cantidad de proteínas carbamoiladas en la composición es sin embargo inferior al 5 % (p/p). Del mismo modo, es particularmente preferible que la composición de la invención comprenda cantidades detectables de proteínas gluconoiladas, siendo sin embargo la cantidad de proteínas gluconoiladas en la composición inferior al 5 % (p/p).

La composición de la invención puede comprender urea que resulta de la solubilización del cuerpo de inclusión y/o de la etapa de replegamiento.

La composición de la invención comprende preferentemente ninguna cantidad de agregados proteicos o solo una baja cantidad de ellos que puedan resultar de la agregación de moléculas proteicas, tales como di-, tri- u oligómeros de la variante de MYDGF descrita con anterioridad. De preferencia, la composición comprende la proteína MYDGF de la invención, es decir, la proteína de la SEQ ID NO: 1 o SEQ ID NO: 2, predominantemente como un monómero. Más preferentemente, el contenido de monómero de la proteína en la composición es del 95 % (p/p) o más, y aún más preferentemente, del 96 % (p/p) o más, 97 % (p/p) o más, 98 % (p/p) o más, o 99 % (p/p) o más. Dicho de otro modo, la cantidad de agregados de la proteína en la composición es de aproximadamente 5 % (p/p) o menos, y aún más preferentemente 4 % (p/p) o menos, 3 % (p/p) o menos, 2 % (p/p) o menos, o 1 % (p/p) o menos, o no es detectable en lo absoluto. La cantidad de monómeros y agregados proteicos se mide preferentemente por cromatografía de exclusión por tamaño (SEC), y más preferentemente por cromatografía líquida de alto rendimiento por exclusión por tamaño (SEC HPLC).

La composición de la invención comprende preferentemente una proteína de 143 aminoácidos que tiene la secuencia de aminoácidos de la SEQ ID NO: 1 o SEQ ID NO: 2. La composición comprende preferentemente menos del 5 % (p/p), menos del 4 % (p/p), menos del 3 % (p/p), menos del 2 % (p/p) y preferentemente menos del 1 % (p/p) de moléculas de proteína que son más cortas que 143 aminoácidos, como se mide por LCMS.

La composición de la invención comprende preferentemente menos del 5% (p/p), menos del 4% (p/p), menos del 3% (p/p), menos del 2% (p/p) y más preferentemente menos del 1% (p/p) de moléculas de proteína que difieren de la secuencia de aminoácidos de SEQ ID NO: 1 por (i) la supresión de 1-4 aminoácidos en el N-terminal de SEQ ID NO: 1, o ii) la adición de un único aminoácido en el N-terminal de SEQ ID NO: 1, basándose en el peso global de todas las proteínas no gluconoiladas y no carbamoiladas de dicha composición que comprenden al menos 100 aminoácidos contiguos de la secuencia de SEQ ID NO: 1, y determinado por LCMS tras dimetilación reductora (etiquetado isotópico estable con dimetilo, SIDL) según Tolonen et al. (2011). La composición mostrada en la Tabla 15 que comprende la variante +A (SEQ ID NO: 1) muestra un 0,2% de proteínas según (i) o (ii).

La composición de la invención comprende preferentemente menos del 5% (p/p), menos del 4% (p/p), menos del 3% (p/p), menos del 2% (p/p) y más preferentemente menos del 1% (p/p) de moléculas de proteína que difieren de la secuencia de aminoácidos de SEQ ID NO: 1 por la delección de 1-4 aminoácidos en el N-terminal de SEQ ID NO: 1, basándose en el peso global de todas las proteínas no gluconoiladas y no carbamoiladas de dicha composición que comprenden al menos 100 aminoácidos contiguos de la secuencia de SEQ ID NO: 1, y determinado por LCMS tras dimetilación reductora (stable isotope dimethyl labelling, SIDL) según Tolonen et al. (2011). La composición mostrada en la Tabla 15 que comprende la variante +A (SEQ ID NO: 1) muestra un 0,2% (p/p) de tales delecciones proteicas.

La composición de la invención comprende preferentemente menos del 5% (p/p), menos del 4% (p/p), menos del 3% (p/p), menos del 2% (p/p) y más preferentemente menos del 1% (p/p) de moléculas de proteína que difieren de la secuencia de aminoácidos de SEQ ID NO: 2 por (i) la delección de 1-4 aminoácidos en el N-terminal de SEQ ID NO: 2, o ii) la adición de un único aminoácido en el N-terminal de SEQ ID NO: 2, basándose en el peso global de todas las proteínas no gluconoiladas y no carbamoiladas de dicha composición que comprenden al menos 100 aminoácidos contiguos de la secuencia de SEQ ID NO: 2, y

determinado por LCMS tras dimetilación reductora (etiquetado isotópico estable con dimetilo, SIDL) según Tolonen et al. (2011). La composición mostrada en la Tabla 15 que comprende la variante +S (SEQ ID NO: 2) muestra un 3,1% (p/p) de proteínas según (i) o (ii).

La composición de la invención comprende preferentemente menos del 5% (p/p), menos del 4% (p/p), menos del 3% (p/p), menos del 2% (p/p) y más preferentemente menos del 1% (p/p) de moléculas de proteína que difieren de la secuencia de aminoácidos de SEQ ID NO: 2 por la delección de 1-4 aminoácidos en el N-terminal de SEQ ID NO: 2, basándose en el peso global de todas las proteínas no gluconoiladas y no carbamoiladas de dicha composición que comprenden al menos 100 aminoácidos contiguos de la secuencia de SEQ ID NO: 2, y determinado por LCMS tras dimetilación reductora (stable isotope dimethyl labelling, SIDL) según Tolonen et al. (2011). La composición mostrada en la Tabla 15 que comprende la variante +S (SEQ ID NO: 2) muestra un 0,5% (p/p) de tales delecciones proteicas.

La composición de la invención comprende preferentemente más del 95% (p/p), más del 96% (p/p), más del 97% (p/p), más del 98% (p/p), y más preferentemente más del 99% (p/p) de moléculas de proteína que tienen una longitud de 143 aminoácidos y consisten en la secuencia de aminoácidos de SEQ ID NO: 1, basándose en el peso global de todas las proteínas no gluconoiladas y no carbamoiladas de dicha composición que comprenden al menos 100 aminoácidos contiguos de la secuencia de SEQ ID NO: 1, y determinado por LCMS tras dimetilación reductora (etiquetado isotópico estable con dimetilo, SIDL) según Tolonen et al. (2011).

La composición de la invención comprende preferentemente más del 95% (p/p), más del 96% (p/p), más del 97% (p/p), más del 98% (p/p), y más preferentemente más del 99% (p/p) de moléculas proteicas que tienen una longitud de 143 aminoácidos y consisten en la secuencia de aminoácidos de SEQ ID NO: 2, basándose en el peso global de todas las proteínas no gluconoiladas y no carbamoiladas de dicha composición que comprenden al menos 100 aminoácidos contiguos de la secuencia de SEQ ID NO: 2, y determinado por LCMS tras dimetilación reductora (etiquetado isotópico estable con dimetilo, SIDL) según Tolonen et al. (2011).

La composición de la invención comprende preferentemente más del 90% (p/p), más del 91% (p/p), más del 92% (p/p), más del 93% (p/p), más del 94% (p/p), y más

preferentemente más del 95% (p/p) de moléculas de proteína que tienen una longitud de 143 aminoácidos y consisten en la secuencia de aminoácidos de SEQ ID NO: 1, basado en el peso total de todas las proteínas en dicha composición que comprenden al menos 100 aminoácidos contiguos de la secuencia SEQ ID NO:1, y medido por espectrometría de masas de cromatografía líquida (LCMS), donde las proteínas gluconoiladas, proteínas carbamoiladas, proteínas deshidratadas y aductos de Na⁺ se ignoran para el cálculo del porcentaje.

La composición de la invención comprende preferentemente más del 90% (p/p), más del 91% (p/p), más del 92% (p/p), más del 93% (p/p), más del 94% (p/p), y más preferentemente más del 95% (p/p) de moléculas proteicas que tienen una longitud de 143 aminoácidos y consisten en la secuencia de aminoácidos de SEQ ID NO: 2, basándose en el peso total de todas las proteínas de dicha composición que contengan al menos 100 aminoácidos de la secuencia SEQ ID NO: 2, y medidas por espectrometría de masas de cromatografía líquida (LCMS), ignorándose para el cálculo del porcentaje las proteínas gluconoiladas, carbamoiladas, deshidratadas y los aductos de Na⁺.

Las proteínas de la invención comparten al menos una actividad biológica de la proteína MYDGF humana madura natural que las hace útiles para su aplicación terapéutica. Por lo tanto, la invención también se refiere a una composición como se describió con anterioridad para su uso como medicamento, en particular para los usos conocidos para MYDGF, véanse, por ejemplo, los documentos WO 2014/111458 y WO 2021/148411.

De preferencia, la proteína es activa en (i) el tratamiento o la prevención de una enfermedad o condición seleccionada del grupo que consiste en lesión, herida, isquemia, lesión por reperfusión, traumatismo, sobrecarga mecánica, intoxicación, cirugía, cardiomiopatía primaria o adquirida, disfunción contráctil posisquémica, infarto de miocardio, preferentemente infarto agudo de miocardio, angina de pecho, fallo cardíaco, inflamación del corazón, insuficiencia cardíaca, hipertrofia y fibrosis; (ii) promover o mejorar la regeneración del tejido cardíaco, la proliferación de cardiomiocitos, la neovascularización, la función cardíaca o la función sistólica del ventrículo izquierdo tras un infarto de miocardio, por ejemplo, por apoptosis o necrosis; o (iv) disminuir el tamaño del infarto tras un infarto de miocardio, preferentemente un infarto agudo de miocardio. Véase el documento WO 2014/111458.

La miocardiopatía por tratar puede ser una miocardiopatía hereditaria o una miocardiopatía causada por mutaciones espontáneas. La miocardiopatía a tratar también puede ser una miocardiopatía adquirida, preferentemente una miocardiopatía isquémica causada por aterosclerosis u otras enfermedades de las arterias coronarias, una miocardiopatía causada por infección o intoxicación del miocardio, una cardiopatía hipertensiva causada por hipertensión arterial pulmonar y/o hipertensión arterial o enfermedades de las válvulas cardíacas. La miocardiopatía es preferentemente una miocardiopatía seleccionada del grupo formado por miocardiopatía hipertrófica (HCM o HOCM), miocardiopatía arritmogénica del ventrículo derecho (ARVC), miopatía mitocondrial ventricular aislada no compactada, miocardiopatía dilatada (DCM), miocardiopatía restrictiva (RCM), miocardiopatía de Takotsubo, endocarditis de Loeffler, miocardiopatía diabética, miocardiopatía alcohólica o miocardiopatía asociada a la obesidad.

La insuficiencia cardíaca que debe tratarse es preferentemente una insuficiencia cardíaca crónica. La insuficiencia cardíaca o insuficiencia cardíaca crónica puede ser insuficiencia cardíaca con fracción de eyección preservada (HFpEF), insuficiencia cardíaca con fracción de eyección reducida (HFrEF) o insuficiencia cardíaca con fracción de eyección media (HFmrEF). Véase el documento WO 2021/148411.

Se prefiere en particular que la proteína descrita con anterioridad tenga al menos parte de la actividad de la proteína MYDGF madura humana natural para aumentar la proliferación de células endoteliales de arterias coronarias o de células endoteliales de arterias coronarias. En particular, se prefiere que la proteína tenga al menos el 50 % de la actividad de la proteína MYDGF madura humana natural para aumentar la proliferación de células endoteliales de arterias coronarias, y preferentemente al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 90 %, al menos el 95 % o el 100 % de dicha actividad. La proteína también puede ejercer una actividad de mejora de la proliferación de células endoteliales de arterias coronarias superior a la actividad de la proteína MYDGF humana madura natural, como una actividad de al menos el 110 %, al menos 120 % al menos 130 % al menos 90 %, al menos 140 %, al menos 150 % al menos 160 % al menos 180 %, al menos 190 %, al menos 200 %, al menos 250 %, al menos 300 %, al menos 350 %, al menos 400 %, al menos 450 %, o al menos 500 % en relación con la proteína MYDGF humana madura. De preferencia, la actividad para aumentar la proliferación de células

endoteliales de arterias coronarias se determina como se describe en el ensayo de potencia del Ejemplo 6 a continuación. La actividad de la proteína se calcula mediante la fórmula: Actividad [%] = EC_{50} de MYDGF madura / EC_{50} de la proteína de ensayo) x 100.

En otra forma de realización, se prefiere que la proteína tenga al menos el 50 % de la actividad de la variante [+G]–HEK en la mejora de la proliferación de células endoteliales de arteria coronaria, en donde la variante [+G]–HEK se ha fabricado como se describe en Polten et al. (2019) y Ebenhoch et al. (2019). Se prefiere en particular que la proteína tenga al menos el 60 %, al menos el 70 % al menos el 80 % al menos el 90 %, al menos el 95 % o el 100 % de dicha actividad. La proteína también puede ejercer una actividad en la potenciación de la proliferación de células endoteliales de arteria coronaria que es superior a la actividad de la variante [+G]–HEK, tal como una actividad de al menos 110 %, al menos 120 % al menos 130 % al menos 90 %, al menos del 140 %, al menos del 150 %, al menos del 160 %, al menos del 180 %, al menos del 190 %, al menos del 200 %, al menos del 250 %, al menos del 300 %, al menos del 350 %, al menos del 400 %, al menos del 450 %, o al menos del 500 % en relación con la variante [+G]–HEK. De preferencia, la actividad en la mejora de la proliferación de células endoteliales de arteria coronaria se determina como se describe en el ensayo de potencia del Ejemplo 6 a continuación. La actividad de la proteína se calcula mediante la fórmula: Actividad [%] = EC_{50} de [+G]–HEK) / EC_{50} de la proteína de ensayo) x 100.

En otra forma de realización, se prefiere que la proteína descrita con anterioridad al menos el 50 % de la actividad de la proteína MYDGF madura humana natural para aumentar la proliferación de cardiomiocitos, y preferentemente al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 90 %, al menos el 95 % o el 100 % de dicha actividad. La proteína también puede ejercer una actividad en la potenciación de la proliferación de cardiomiocitos que sea superior a la actividad de la proteína MYDGF madura humana natural, tal como una actividad de al menos el 110 %, al menos 120 % al menos 130 % al menos 90 %, al menos 140 %, al menos 150 % al menos 160 % al menos 180 %, al menos 190 %, al menos 200 %, al menos 250 %, al menos 300 %, al menos 350 %, al menos 400 %, al menos 450 %, o al menos 500 % en relación con la proteína MYDGF humana madura. De preferencia, la actividad en el aumento de la proliferación de cardiomiocitos se determina como se describe en el ensayo de potencia del Ejemplo 7 a continuación. La

actividad de la proteína se calcula mediante la fórmula: Actividad [%] = EC_{50} de MYDGF madura / EC_{50} de la proteína de ensayo) x 100.

En otra forma de realización, se prefiere que la proteína tenga al menos el 50 % de la actividad de la variante [+G]–HEK en la mejora de la proliferación de cardiomiocitos, en donde la variante [+G]–HEK se ha fabricado como se describe en Polten et al. (2019) y Ebenhoch et al. (2019). Se prefiere en particular que la proteína tenga al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 90 %, al menos el 95 % o el 100 % de dicha actividad. La proteína también puede ejercer una actividad en la mejora de la proliferación de cardiomiocitos que es superior a la actividad de la variante [+G]–HEK, tal como una actividad de al menos 110 %, al menos 120 %, al menos 130 %, al menos 140 %, al menos 150 %, al menos 160 %, al menos 180 %, al menos 190 %, al menos 200 %, al menos 250 %, al menos 300 %, al menos 350 %, al menos 400 %, al menos 450 %, o al menos 500 % en relación con la variante [+G]–HEK. De preferencia, la actividad para aumentar la proliferación de cardiomiocitos se determina como se describe en el ensayo de potencia del Ejemplo 7 a continuación. La actividad de la proteína se calcula mediante la fórmula: Actividad [%] = EC_{50} de [+G]–HEK) / EC_{50} de la proteína de ensayo) x 100.

En otra forma de realización, se prefiere que la proteína descrita con anterioridad aumente la proliferación de células endoteliales de la arteria coronaria con una EC_{50} de menos de 100 ng/ml, cuando se mide en el ensayo de potencia del Ejemplo 6 descrito a continuación. De preferencia, la proteína potencia la proliferación de células endoteliales de arteria coronaria con una EC_{50} de menos de 95 ng/ml, menos de 90 ng/ml, menos de 85 ng/ml, menos de 80 ng/ml, menos de 75 ng/ml, menos de 70 ng/ml, menos de 65 ng/ml, o menos de 60 ng/ml, cuando se mide en el ensayo de potencia del Ejemplo 6.

En otra forma de realización, se prefiere que la proteína potencie la proliferación de cardiomiocitos con una EC_{50} de menos de 100 ng/ml cuando se mide en la potencia según el Ejemplo 7 descrito a continuación. De preferencia, la proteína potencia la proliferación de células endoteliales de la arteria coronaria con una EC_{50} de menos de 95 ng/ml, menos de 90 ng/ml, menos de 85 ng/ml, menos de 80 ng/ml, menos de 75 ng/ml, menos de 70 ng/ml, menos de 65 ng/ml, menos de 60 ng/ml, menos de 55 ng/ml, menos de 50 ng/ml, menos de 45 ng/ml, menos de 40 ng/ml, menos de 35 ng/ml, menos de 30 ng/ml, o menos de 25 ng/ml, cuando se mide en el ensayo de potencia del Ejemplo 7.

La invención también se refiere a un ácido nucleico que codifica una proteína que tras la maduración da lugar a una proteína como la descrita con anterioridad, es decir, una proteína que tiene la secuencia de la SEQ ID NO: 1 o SEQ ID NO: 2. El ácido nucleico puede ser ADN o ARN. Sin embargo, se prefiere que el ácido nucleico sea una molécula de ADN.

La invención también se refiere a un vector o plásmido que comprende un ácido nucleico que codifica una proteína que al madurar da lugar a una de las proteínas MYDGF de la invención. De preferencia, el vector será un vector de expresión que permite la expresión de una proteína que madura en la proteína de la SEQ ID NO: 1 o SEQ ID NO: 2 en una célula procariota o eucariota. Se prefiere en particular que el vector sea un vector de expresión procariota, es decir, un vector que permita la expresión de la proteína recombinante en un entorno celular procariota. Aún más preferentemente, el vector es un vector de expresión bacteriano, es decir, un vector que permite la expresión de proteínas recombinantes en una célula bacteriana. El vector comprenderá preferentemente un origen de replicación, un promotor, un polienlazador para clonación, un terminador de transcripción y un gen que permita la selección como, por ejemplo, un gen que codifique una proteína que confiera resistencia a los antibióticos. Se ha descrito un gran número de vectores de expresión para *E. coli* y otros huéspedes bacterianos. Algunos ejemplos de vectores adecuados para la expresión de proteínas en células *E. coli* son, por ejemplo, los vectores de la serie pBluescript, la serie pUC, la serie pQE o la serie pET. El vector comprende preferentemente un sistema promotor inducible capaz de iniciar la expresión tras la adición de un compuesto inductor.

En una forma de realización preferida, el vector que alberga el ácido nucleico que codifica la proteína MYDGF descrita con anterioridad es un vector del tipo pET. Estos vectores comprenden típicamente un origen de replicación, un promotor T7 que es específico de la ARN polimerasa T7, un operador lac para unir la proteína represora lacI, un polienlazador para clonar la secuencia de ácido nucleico que codifica la proteína por expresar, una secuencia de terminación de la transcripción, un gen de resistencia a la ampicilina o kanamicina y un gen lacI que codifica la proteína represora lac. En ausencia de isopropil- β -D-tiogalactopiranosido (IPTG) o lactosa, la proteína represora se une al operador lac, inhibiendo así el promotor T7 y bloqueando la expresión de la proteína diana. La unión de IPTG o lactosa a la proteína represora lac provoca un cambio conformacional

que causa el desprendimiento de la proteína del operador y la inducción de la expresión a partir del promotor T7. Los vectores pET adecuados para su uso en los métodos de la presente invención comprenden, pero no se limitan a pET21a(+), pET24a(+), pET28a(+), pET29a(+), pET30a(+), pET41a(+), pET44a(+), pET21b(+), pET24b(+), pET26b(+), pET28b(+), pET29b(+), pET30b(+), pET42b(+) y pET44b(+). Se prefieren especialmente los vectores basados en el esqueleto pET-26b(+). Otros ejemplos de vectores adecuados se describen, por ejemplo, en "Cloning Vectors" (Pouwels et al. (eds.) Elsevier, Amsterdam New York Oxford, 1985).

El vector de expresión puede transformarse en la célula huésped eucariota o procariota por cualquier método adecuado. Por ejemplo, un vector de expresión para su uso en *E. coli* puede introducirse en la célula huésped, por ejemplo, por electroporación o por métodos químicos, como la transformación mediada por fosfato de calcio, tal como se describe en Maniatis et al. 1982, *Molecular Cloning, A laboratory Manual*, Cold Spring Harbor Laboratory.

La invención también se refiere a una célula huésped que comprende una proteína, un ácido nucleico o un vector como se ha descrito con anterioridad. La célula huésped puede ser una célula eucariota o procariota, pero se prefiere especialmente que sea una célula huésped procariota, como una célula bacteriana. Aunque el tipo de célula bacteriana no está particularmente limitado, se prefiere que la célula huésped sea una célula *Escherichia coli*, como una célula *Escherichia coli* BL21.

La invención también se refiere a una composición descrita con anterioridad, es decir, una composición que comprende una proteína descrita con anterioridad, preferentemente una composición obtenible por un método de expresión recombinante en un sistema de expresión celular, como los métodos descritos con más detalle a continuación, para la preparación de una composición farmacéutica.

La invención también se refiere a una composición farmacéutica que comprende la proteína descrita con anterioridad o una composición descrita con anterioridad. La composición puede comprender, además de la proteína, un portador farmacéuticamente aceptable y otros excipientes que se utilizan habitualmente para la formulación de composiciones farmacéuticas. En general, la composición farmacéutica puede formularse para diferentes vías de administración. Se prefiere que la composición de la invención esté formulada para la administración parenteral, por ejemplo, por vía intravenosa, intraarterial,

intracoronaria o intravenosa. Las composiciones adecuadas para inyección o infusión pueden incluir soluciones o dispersiones y polvos para la preparación extemporánea de dichas soluciones o dispersiones inyectables. La composición inyectable debe ser estéril y estable en las condiciones de fabricación y almacenamiento. De preferencia, las composiciones para inyección o infusión también incluyen un conservante, como clorobutanol, fenol, ácido ascórbico, timerosal y similares. Para la administración intravenosa, los portadores adecuados pueden ser solución salina fisiológica, agua bacteriostática, Cremophor EL™ (BASF) o solución salina tamponada con fosfato (PBS). Las soluciones estériles para inyección o infusión pueden prepararse incorporando la proteína MYDGF en la cantidad requerida en un disolvente apropiado seguido de esterilización por filtración.

Aparte de la proteína MYDGF, la composición farmacéutica de la invención puede comprender además agentes activos adicionales que sean eficaces en (i) tratar o prevenir una enfermedad seleccionada del grupo que consiste en lesión, herida, isquemia, lesión por reperfusión, traumatismo, sobrecarga mecánica, intoxicación, cirugía, miocardiopatía primaria o adquirida, disfunción contráctil posisquémica, infarto de miocardio, preferentemente infarto agudo de miocardio, angina de pecho, fallo cardíaco, inflamación del corazón, insuficiencia cardíaca, hipertrofia y fibrosis; (ii) promover o mejorar la regeneración del tejido cardíaco, la proliferación de cardiomiocitos, la neovascularización, la función cardíaca o la función sistólica del ventrículo izquierdo tras un infarto de miocardio; (iii) proteger a los cardiomiocitos de la muerte, por ejemplo, por apoptosis o necrosis; o (iv) disminuir el tamaño del infarto tras un infarto de miocardio, preferentemente un infarto agudo de miocardio, en un sujeto que lo necesite.

Por ejemplo, la composición farmacéutica puede comprender uno o más inhibidores de la enzima convertidora de angiotensina (ECA), como benazepril, zofenopril, perindopril, trandolapril, captopril, enalapril, lisinopril y ramipril. La composición farmacéutica también puede contener uno o más diuréticos, como clorotiazida, hidroclorotiazida, bendroflumetiazida, espironolactona, clortalidona, metilclorotiazida, politiazida, triamtereno, furosemida, ácido etacrínico, metolazona, bumetanida, indapamida, amilorida, acetazolamida, torsemida y eplerenona. La composición farmacéutica también puede

contener uno o más betabloqueantes, como acebutolol, atenolol, betaxolol, bisoprolol, carvedilol, celiprolol, esmolol, metoprolol, nebivolol, propranolol, sotalol y/o timolol.

Si la proteína se utiliza en combinación con cualquiera de los agentes activos adicionales mencionados, por ejemplo, con un inhibidor de la ECA, un diurético y/o un betabloqueante, los dos agentes activos también pueden administrarse por separado, es decir, en forma de composiciones farmacéuticas separadas, una que contenga la proteína MYDGF y otra que contenga el agente activo adicional. Las composiciones separadas pueden administrarse simultáneamente, es decir, al mismo tiempo en dos lugares de administración distintos, o pueden administrarse secuencialmente (en cualquier orden) en el mismo lugar o en diferentes lugares de administración.

La invención también se refiere a una proteína o una composición farmacéutica como la descrita con anterioridad para su uso como medicamento. Más específicamente, la proteína o composición farmacéutica es adecuada para ser utilizada como medicamento para (i) tratar o prevenir una enfermedad o condición seleccionada del grupo que consiste en lesión, herida, isquemia, lesión por reperfusión, trauma, sobrecarga mecánica, intoxicación, cirugía, miocardiopatía primaria o adquirida, disfunción contráctil posisquémica, infarto de miocardio, preferentemente infarto agudo de miocardio, angina de pecho, fallo cardíaco, inflamación del corazón, insuficiencia cardíaca, hipertrofia y fibrosis; (ii) promover o mejorar la regeneración del tejido cardíaco, la proliferación de cardiomiocitos, la neovascularización, la función cardíaca o la función sistólica del ventrículo izquierdo tras un infarto de miocardio; proteger al cardiomiocito de la muerte, por ejemplo, por apoptosis o necrosis; o (iv) reducir el tamaño del infarto tras un infarto de miocardio, preferentemente un infarto agudo de miocardio.

La invención también se refiere a un método para (i) tratar o prevenir una enfermedad o condición seleccionada del grupo que consiste en lesión, herida, isquemia, lesión por reperfusión, traumatismo, sobrecarga mecánica, intoxicación, cirugía, cardiomiopatía primaria o adquirida, disfunción contráctil posisquémica, infarto de miocardio, preferentemente infarto agudo de miocardio, angina de pecho, fallo cardíaco, inflamación del corazón, insuficiencia cardíaca, hipertrofia y fibrosis; (ii) promover o mejorar la regeneración del tejido cardíaco, la proliferación de cardiomiocitos, la neovascularización, la función cardíaca o la función sistólica del ventrículo izquierdo tras un infarto de miocardio; (iii) proteger a los cardiomiocitos de la muerte, por ejemplo, por

apoptosis o necrosis, o (iv) disminuir el tamaño del infarto tras un infarto de miocardio, preferentemente un infarto agudo de miocardio, en un sujeto que lo necesite, en donde dicho método comprende la administración de una cantidad eficaz de una composición farmacéutica como se describió con anterioridad que comprenda las proteínas de la SEQ ID NO: 1 o SEQ ID NO: 2.

La invención también proporciona un método para la expresión recombinante de una proteína MYDGF en una célula huésped, en donde dicho método comprende las siguientes etapas:

- (a) proporcionar una célula huésped como la descrita con anterioridad, preferentemente una célula huésped que comprenda un ácido nucleico que codifique una proteína que, tras la maduración, consista en 143 aminoácidos que tengan la secuencia de aminoácidos de la SEQ ID NO: 1 o SEQ ID NO: 2;
- (b) cultivar la célula huésped en condiciones que permitan la expresión de la proteína MYDGF;
- (c) aislar cuerpos de inclusión que contengan la proteína MYDGF de la célula huésped; y
- (d) solubilizar los cuerpos de inclusión y volver a plegar la proteína MYDGF.

El método anterior de la invención se refiere a la producción de las proteínas establecidas en la SEQ ID NO: 1 y SEQ ID NO: 2. El método hace uso de una célula huésped como se describe más arriba, preferentemente una célula huésped procariota que comprende un ácido nucleico que codifica una proteína que al madurar da lugar a una de las proteínas de la invención. La célula huésped bacteriana utilizada en el método de la invención contiene preferentemente un ácido nucleico que codifica una proteína que al madurar da lugar a la proteína de la SEQ ID NO: 1 o un ácido nucleico que codifica una proteína que al madurar da lugar a la proteína de la SEQ ID NO: 2 que se inserta en un vector de expresión que permite la expresión de la proteína recombinante en la célula huésped. La célula huésped puede ser cualquier tipo de célula eucariota o procariota que sea adecuada para ser utilizada para la expresión de proteínas exógenas recombinantes, es decir, proteínas que no son producidas naturalmente por la célula huésped. De preferencia, la célula huésped es una célula procariota, como una célula bacteriana. Más preferentemente, la célula es una célula bacteriana que pertenece al género *Escherichia*, y aún más preferentemente a la especie *E. coli*. Se prefiere el uso de la cepa BL21 de *E. coli* o una cepa derivada de la misma.

De preferencia, la etapa (a) del método anterior comprende (i) proporcionar una célula huésped que comprenda un ácido nucleico que contenga un marco de lectura abierto, flanqueado por un codón de inicio y un codón de parada, de acuerdo con la secuencia de la SEQ ID NO: 11 o SEQ ID NO: 12 unido de forma operable a un promotor; o (ii) proporcionar una célula huésped que comprenda un ácido nucleico que codifique una proteína que antes de la maduración consista en 144 aminoácidos que tengan la secuencia de aminoácidos de la SEQ ID NO: 15 o SEQ ID NO: 16.

Las secuencias de aminoácidos de la SEQ ID NO: 15 o SEQ:16 son proteínas MYDGF que incluyen el residuo de metionina N-terminal derivado del codón de inicio. Estas proteínas se someten a un proceso de maduración que elimina la metionina N-terminal. La maduración se efectúa preferentemente mediante una o más aminopeptidasas derivadas de la célula huésped, más preferentemente una o más metionina aminopeptidasas derivadas de la célula huésped. Por ejemplo, la una o más aminopeptidasas pueden ser producidas por la célula huésped bacteriana que se utiliza para la expresión recombinante, por ejemplo E. coli. Se prefiere en particular que la etapa (a) comprenda proporcionar una célula huésped que comprenda un ácido nucleico de SEQ ID NO: 7 o SEQ ID NO: 8, que son vectores de expresión preferentemente circulares, enrollados o superenrollados.

En la etapa (b) del método, la célula huésped bacteriana se cultiva en condiciones que permiten la expresión de la proteína en la célula huésped bacteriana. Las condiciones que permiten la expresión de la proteína MYDGF dependerán de la célula huésped procariota y del vector de expresión utilizado en el proceso. Un experto puede seleccionar y aplicar fácilmente las condiciones adecuadas. Por ejemplo, si se utiliza un sistema de expresión bacteriana inducible, como un vector pET, las condiciones que permiten la expresión de la proteína diana incluirán normalmente una temperatura de cultivo de entre 20–42 °C, preferentemente 30–40 °C, y más preferentemente 35–38 °C. El medio de cultivo tendrá normalmente un pH de entre 6,5 y 9,0, más típicamente de 7,0 a 8,0, y preferentemente de 7,5. La fermentación del cultivo puede continuarse durante un período de tiempo comprendido entre varias horas y varios días. Las células pueden cultivarse en un proceso por lotes o por lotes alimentados. Por ejemplo, si el cultivo se realiza como un proceso discontinuo, el tiempo de cultivo oscila normalmente entre unas 12 horas y unas 36 horas. Si se utiliza un proceso continuo, los tiempos de fermentación pueden ser de hasta 21 días o más. En la Encyclopaedia of Bioprocess Technology se describen métodos

adecuados para el cultivo de células huésped bacterianas: Fermentation, Biocatalysis, and Bioseparation, Volumes 1–5, Flickinger, M.C., Drew, S.W. (eds.), 1999 John Wiley & Sons. De preferencia, la proteína se expresa en la célula huésped utilizando un sistema de expresión inducible, por ejemplo, un sistema que permite iniciar la expresión de la proteína mediante la adición de un compuesto inductor como IPTG o lactosa al medio de cultivo.

Las proteínas así expresadas se acumulan en la célula huésped en forma insoluble en los llamados cuerpos de inclusión. Esto significa que las proteínas expresadas se acumulan intracelularmente y se depositan en forma de agregados insolubles de proteínas inactivas y mal plegadas.

En la etapa (c) del método de la invención, los cuerpos de inclusión que contienen la proteína MYDGF insoluble se aíslan de la célula huésped. Para ello, las células huésped bacterianas se cosechan después del cultivo y se desorganizan, por ejemplo, mediante homogeneización a alta presión u otros procedimientos de lisis celular comúnmente conocidos. Los cuerpos de inclusión pueden aislarse de los lisados por diferentes métodos, por ejemplo, por centrifugación tubular, como la centrifugación tubular de alta velocidad. Los métodos para el aislamiento de cuerpos de inclusión a partir de células bacterianas son comúnmente conocidos y descritos, por ejemplo, en Peternel & Komel (2010) y Eggenreich et al. (2020).

En la etapa (d) del método de la invención, la proteína MYDGF en los cuerpos de inclusión aislados obtenidos de la etapa (c) se solubiliza y se vuelve a plegar. Los métodos para solubilizar proteínas son conocidos e incluyen, por ejemplo, la incubación de los cuerpos de inclusión en presencia de urea, clorhidrato de guanidina (GuHCl) y/o DTT, seguida de filtración. Se prefiere que el replegamiento de las proteínas se realice en presencia de urea. La filtración puede incluir una o más de las siguientes técnicas: filtración profunda, ultrafiltración y/o diafiltración. Los métodos para replegar proteínas son igualmente conocidos e incluyen, por ejemplo, la incubación de proteínas solubilizadas a partir de cuerpos de inclusión en presencia de urea, CaCl_2 y/o cistamina. Diferentes fabricantes comercializan kits para solubilizar y volver a plegar proteínas a partir de cuerpos de inclusión.

El método también puede comprender una etapa (e) en donde se obtiene una proteína MYDGF replegada de 143 aminoácidos que tiene la secuencia de aminoácidos de

la SEQ ID NO: 1 o SEQ ID NO: 2. Esta etapa sigue a la etapa (d) del método anterior. Esta etapa sigue a la etapa (d) del método anterior.

De preferencia, el método de la invención comprende la etapa adicional (f) en donde la proteína MYDGF solubilizada y replegada obtenida de la etapa (e) se purifica aún más. La purificación de la proteína puede realizarse de acuerdo con métodos rutinarios y puede incluir uno o más de ultrafiltración, diafiltración, cromatografía de interacción hidrofóbica y/o cromatografía de intercambio iónico aniónico.

De preferencia, la purificación en la etapa (f) se efectúa mediante cromatografía de intercambio aniónico o cromatografía de interacción hidrofóbica. Estos tipos de cromatografía pueden realizarse poniendo en contacto la proteína MYDGF con el material de resina de cromatografía en condiciones que permitan la adsorción de la proteína MYDGF a la resina. A continuación, el material de resina se lava opcionalmente para eliminar impurezas, como material no proteínico o proteínas distintas de la MYDGF. En una última etapa, la proteína MYDGF se eluye de la resina.

Más preferentemente, la purificación en la etapa (f) se efectúa mediante cromatografía de intercambio aniónico. Si se utiliza una cromatografía de intercambio aniónico, la adsorción de la proteína MYDGF a la resina de cromatografía de intercambio aniónico se realiza preferentemente en condiciones de baja fuerza iónica, por ejemplo, a una conductividad inferior a 3 mS/cm, inferior a 2 mS/cm, inferior a 1,5 mS/cm o inferior a < 1 mS/cm. La elución puede lograrse aumentando la concentración de sal y/o disminuyendo el pH de la fase líquida, es decir, la fase móvil.

El método de la invención permite producir la proteína de la SEQ ID NO: 1 o SEQ ID NO: 2 en cantidades particularmente elevadas. Como se muestra en los Ejemplos 2 y 3, las proteínas pueden producirse con el método de la invención a una productividad de más de 0,4 g de proteína por 100 g de células, y preferentemente más de 0,5 g de proteína por 100 g de células, más de 0,6 g de proteína por 100 g de células, más de 0,7 g de proteína por 100 g de células, más de 0,8 g de proteína por 100 g de células, más de 0,9 g de proteína por 100 g de células, y más preferentemente más de 1,0 g de proteína por 100 g de células. Dicho de otro modo, el método de la invención permite producir la proteína de la SEQ ID NO: 1 o SEQ ID NO: 2 en una cantidad de más de 100 g de proteína por lote, y preferentemente más de 150 g de proteína por lote, más de 200 g de proteína por lote, más

de 250 g de proteína por lote, más de 300 g de proteína por lote, y más preferentemente más de 350 g o 400 g de proteína por lote (véase el Ejemplo 3).

La invención también se refiere al uso de una célula huésped bacteriana como se describe en otra parte de este documento para la expresión recombinante de una proteína MYDGF. La célula huésped es una célula procariota o eucariota que comprende un ácido nucleico, plásmido o vector que codifica una proteína descrita en el presente documento con anterioridad.

Otra forma de realización de la invención es una proteína según una de las formas de realización mencionadas con anterioridad que se puede obtener por expresión heteróloga en bacterias, y preferentemente por un método descrito con anterioridad. Otra forma de realización de la invención es una proteína según una de las formas de realización mencionadas con anterioridad que se puede obtener por producción en forma de cuerpos de inclusión y replegamiento. Otra forma de realización de la invención es una composición que comprende una proteína MYDGF, en donde dicha composición se puede obtener por expresión heteróloga en bacterias, y preferentemente por un método descrito con anterioridad, en donde dicha composición comprende menos del 1 % (p/p) de moléculas de proteína que son más cortas que 143 aminoácidos.

En otro aspecto más, la invención proporciona un método para producir una proteína MYDGF en un sistema de expresión basado en células, que comprende

- (a) proporcionar una célula huésped, preferentemente una célula huésped, que comprenda un ácido nucleico que codifique una proteína (que después de la maduración consista) de SEQ ID NO: 1 o SEQ ID NO: 2;
- (b) cultivar la célula huésped en condiciones que permitan la expresión de la proteína;
- (c) aislar los cuerpos de inclusión que contienen la proteína MYDGF de la célula huésped;
- y
- (d) solubilizar los cuerpos de inclusión y volver a plegar la proteína MYDGF.

Las distintas etapas del método se han descrito con anterioridad. Una vez más, se prefiere que el replegamiento de las proteínas se realice en presencia de urea. Por consiguiente, el replegamiento de la proteína en la etapa (d) comprende preferentemente la incubación de la proteína en presencia de urea.

El método también puede comprender una etapa (e) en donde se obtiene una proteína MYDGF replegada de 143 aminoácidos que tiene la secuencia de aminoácidos de

la SEQ ID NO: 1 o SEQ ID NO: 2. Esta etapa sigue a la etapa (d) del método anterior. Esta etapa sigue a la etapa (d) del método anterior.

El método anterior también puede comprender una etapa adicional (f) en donde se purifica la proteína MYDGF expresada por la célula huésped y aislada de los cuerpos de inclusión. Esta etapa puede comprender métodos que se utilizan habitualmente en el campo de la purificación de proteínas, como la ultrafiltración, la diafiltración y/o la cromatografía de intercambio iónico aniónico.

Como se describe en otras partes del presente documento, la célula huésped puede ser cualquier célula huésped que sea adecuada para ser utilizada en la producción de proteínas recombinantes. La célula huésped puede ser eucariota o procariota. Se prefiere el uso de una célula huésped procariota. Se prefiere aún más que la célula huésped sea una célula bacteriana, como una célula *Escherichia coli*. Se prefiere especialmente el uso de células de *Escherichia coli* de la cepa BL21 o una cepa derivada de la misma.

El ácido nucleico contenido por la célula huésped que codifica una proteína que al madurar da lugar a la proteína de la SEQ ID NO: 1 o SEQ ID NO: 2 puede ser una molécula de ADN o ARN, y preferentemente una molécula de ADN. El ácido nucleico puede estar contenido en un vector, como un vector de expresión eucariota o procariota. De preferencia, el vector es un vector de expresión procariota como se describe en otra parte del presente documento.

La invención también se refiere a una composición obtenible a partir del método anterior, en donde dicha composición comprende una proteína de 143 aminoácidos que tiene la secuencia de aminoácidos de la SEQ ID NO: 1 o SEQ ID NO: 2. La composición comprende preferentemente menos del 1 % (p/p) de moléculas de proteína que son más cortas que 143 aminoácidos, medidos por LCMS.

La composición no comprenderá ninguna o solo una pequeña cantidad de proteínas carbamoyladas. En una forma de realización preferente, la composición comprende cantidades detectables de proteínas carbamoyladas. Por ejemplo, en una forma de realización, la cantidad de proteínas carbamoyladas es del 0,01 % (p/p), 0,02 % (p/p), 0,05 % (p/p) o 0,1 % (p/p). Al mismo tiempo, la cantidad global de proteínas carbamoyladas se limitará a menos del 8 % (p/p), y preferentemente a menos del 7 % (p/p), menos del 6 % (p/p) o menos del 5 % (p/p) de las proteínas de dicha composición. En otra forma de

realización preferida, la composición no contiene cantidades detectables de proteínas carbamoiladas.

Del mismo modo, la composición no comprenderá ninguna proteína gluconoilada o solo una pequeña cantidad. En una forma de realización preferida, la composición comprende cantidades detectables de proteínas gluconoiladas. Por ejemplo, en una forma de realización, la cantidad de proteínas gluconoiladas es del 0,01 % (p/p), 0,02 % (p/p), 0,05 % (p/p) o 0,1 % (p/p). Al mismo tiempo, la cantidad global de proteínas glucoiladas se limitará a menos del 6 % (p/p), menos del 5 % (p/p), menos del 4 % (p/p) o menos del 3 % (p/p) de las proteínas de dicha composición. En otra forma de realización preferida, la composición no contiene cantidades detectables de proteínas gluconoiladas.

En una forma de realización particularmente preferida, la composición comprende cantidades detectables de proteínas carbamoiladas, en donde sin embargo menos del 7 % (p/p) o menos del 5 % (p/p) de las proteínas de la composición están carbamoiladas. Por ejemplo, la composición puede comprender al menos 0,05 % (p/p) o al menos 0,1 % (p/p) de proteínas carbamoiladas, en donde sin embargo menos del 7 % (p/p) de las proteínas de la composición están carbamoiladas. En otra forma de realización particularmente preferida, la composición puede comprender al menos 0,05 % (p/p) o al menos 0,1 % (p/p) de proteínas carbamoiladas, aunque menos del 5 % (p/p) de las proteínas de la composición están carbamoiladas.

Del mismo modo, en otra forma de realización particularmente preferida, la composición comprende cantidades detectables de proteínas gluconoiladas, en donde sin embargo menos del 5 % (p/p) o menos del 3 % (p/p) de las proteínas de la composición están gluconoiladas. Por ejemplo, la composición puede comprender al menos 0,05 % (p/p) o al menos 0,1 % (p/p) de proteínas gluconoiladas, aunque menos del 5 % (p/p) de las proteínas de la composición están gluconoiladas. En otra forma de realización particularmente preferida, la composición puede comprender al menos 0,05 % (p/p) o al menos 0,1 % (p/p) de proteínas glucoiladas, aunque menos del 3 % (p/p) de las proteínas de la composición están glucoiladas.

En una forma de realización, la composición comprende una proteína que tiene la secuencia de aminoácidos de la SEQ ID NO: 1 junto con variantes de la misma que son más cortas en longitud y ejercen un 100 % de identidad de secuencia en toda su longitud con la secuencia de aminoácidos de la SEQ ID NO: 1, en donde la longitud es de al menos

100 aminoácidos sin que se permitan huecos en la alineación. Dentro de dicha composición, la relación de la señal obtenida a partir de la proteína según la SEQ ID NO: 1 y la suma de señales obtenidas a partir de dichas variantes más cortas, según se determina por LCMS tras dimetilación reductora (etiquetado de dimetil isótopos estables, SIDL) según Tolonen et al (2011), es superior a 20, y preferentemente superior a 30, 40, 50, 60, 70, 75, 80, 85, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, y más preferentemente superior a 450. Como puede observarse en la Tabla 15 a continuación, la relación de la señal obtenida a partir de la proteína según la SEQ ID NO: 1 y las señales obtenidas a partir de variantes más cortas resultó ser 466, según se determinó por LCMS según Tolonen et al (2011).

En una forma de realización, la composición comprende una proteína que tiene la secuencia de aminoácidos de la SEQ ID NO: 2 junto con variantes de la misma que son más cortas en longitud y ejercen un 100 % de identidad de secuencia en toda su longitud con la secuencia de aminoácidos de la SEQ ID NO: 2, en donde la longitud es de al menos 100 aminoácidos sin que se permitan huecos en la alineación. Dentro de dicha composición, la relación de la señal obtenida a partir de la proteína según la SEQ ID NO: 2 y la suma de señales obtenidas a partir de dichas variantes más cortas, según se determina por LCMS tras dimetilación reductora (etiquetado de dimetil isótopos estables, SIDL) según Tolonen et al (2011), es superior a 20, y preferentemente superior a 50, 75, 100, 125, 150, y más preferentemente superior a 175 o 180. Como puede observarse en la Tabla 15 a continuación, la relación de la señal obtenida a partir de la proteína según la SEQ ID NO: 2 y las señales obtenidas a partir de variantes más cortas resultó ser 186, según se determinó mediante LCMS según Tolonen et al (2011).

De preferencia, la composición comprende una proteína que está plegada de tal manera que más del 70 %, y preferentemente más del 80 %, más del 90 %, o más del 95 %, de los picos ^1H y/o ^{15}N en el mapa de espectroscopia de resonancia magnética nuclear bidimensional (2D-RMN) resultan en valores de desviación de desplazamiento químico combinado (CCSD) inferiores a 0,01 ppm cuando se comparan con los picos correspondientes de la Tabla 1.

La composición anterior contiene una cantidad mínima de impurezas o modificaciones postraduccionales, por lo que puede utilizarse para la preparación de una

composición farmacéutica. La composición farmacéutica puede prepararse como se describe en el presente documento.

La composición o composición farmacéutica anterior es preferentemente para su uso como medicamento, y puede aplicarse como se ha descrito con anterioridad, por ejemplo, para tratar o prevenir la insuficiencia cardíaca, tratar la cardiomiopatía, promover la regeneración del tejido cardíaco, promover la proliferación de cardiomiocitos, promover la neovascularización, promover la función cardíaca, disminuir el tamaño del infarto, tratar o prevenir la fibrosis, tratar o prevenir la hipertrofia, o tratar o prevenir la insuficiencia cardíaca en un sujeto que lo necesite.

BREVE DESCRIPCIÓN DE LAS FIGURAS

Fig. 1 muestra el espectro de RMN de la variante de MYDGF [+A] descrita en el Ejemplo 5.

Fig. 2 muestra el espectro de RMN de la variante de MYDGF [+G–HEK] descrita en el Ejemplo 5.

Fig. 3 muestra una superposición de ambos espectros según las Figuras 1 y 2.

Fig. 4 muestra las curvas dosis–respuesta de diferentes variantes de MYDGF en el ensayo de migración de HCAEC. Los datos de recuperación se obtuvieron a partir de células que habían sido tratadas con vehículo (control), VEGFA (50 ng/mL), diferentes concentraciones de [+G]–HEK (referencia) o diferentes concentraciones de la variante de MYDGF [+A] lote V301, respectivamente. Se muestran los ajustes de la curva 4–PL para [+G]–HEK y el lote [+A] respectivo.

Fig. 5 muestra las curvas dosis–respuesta de diferentes variantes de MYDGF en el ensayo de migración de HCAEC. Los datos de recuperación proceden de células tratadas con vehículo (control), VEGFA (50 ng/mL), diferentes concentraciones de [+G]–HEK (referencia) o diferentes concentraciones de la variante [+A] del lote V302 de MYDGF, respectivamente. Se muestran los ajustes de la curva 4–PL para [+G]–HEK y el lote [+A] respectivo.

Fig. 6 muestra las curvas dosis–respuesta de diferentes variantes de MYDGF en el ensayo de migración de HCAEC. Los datos de recuperación se obtuvieron a partir de células tratadas con vehículo (control), VEGFA (50 ng/mL), diferentes concentraciones de [+G]–HEK (referencia) o diferentes concentraciones de la variante de MYDGF [+A] lote

V303, respectivamente. Se muestran los ajustes de la curva 4-PL para [+G]-HEK y el lote [+A] respectivo.

Fig. 7 muestra una comparación de diferentes variantes y lotes de MYDGF en ensayos de isquemia/reperfusión. La actividad metabólica se determinó para células que habían sido estimuladas con vehículo (control), IGF-1 (50 ng/mL), diferentes concentraciones de [+G]-HEK (referencia) o diferentes concentraciones de MYDGF (lotes V301, V302 y V303), respectivamente. Se muestran los ajustes de las curvas 4-PL para [+G]-HEK y el respectivo lote [+A].

Fig. 8 muestra la función cardíaca en ratones FVB/N evaluada mediante ecocardiografía. Área telediastólica del ventrículo izquierdo (LVEDA), área telesistólica del ventrículo izquierdo (LVESA) y cambio fraccional de área (FAC) evaluados mediante ecocardiografía transtorácica el día 6 (A) y el día 28 (B) tras cirugía simulada o I/R en ratones FVB/N. Se indican los tratamientos y el número de animales (en las columnas). Se indican los tratamientos y el número de animales (en las columnas). MYDGF designa la proteína humana, Mydgm la proteína murina.

Fig. 9 muestra (A) el tamaño de la cicatriz del infarto el día 28 tras el IM en ratones FVB/N. Se muestran imágenes de ejemplo y datos sumarios. El tamaño de la cicatriz se evaluó mediante la tinción de tricrómico de Masson. Se indican los tratamientos y el número de animales (en las columnas). MYDGF designa la proteína humana, Mydgm la proteína murina. (B) Densidad capilar el día 28 después del IM en ratones FVB/N. Se muestran imágenes de ejemplo y un resumen de los datos. Se muestran imágenes de ejemplo y un resumen de los datos. La densidad capilar se evaluó mediante tinción fluorescente IB4/WGA. Se indican los tratamientos. Se utilizaron 6 ratones por grupo. MYDGF designa la proteína humana, Mydgm la proteína murina.

EJEMPLOS

La invención se ilustrará mediante los siguientes Ejemplos, que se dan únicamente a modo de ejemplo. Específicamente, los Ejemplos describen la generación de la cepa de producción y la expresión heteróloga y purificación de las variantes de MYDGF.

Como se describe con más detalle en los siguientes Ejemplos, el proceso de producción de MYDGF humano se desarrolló primero a escala de 5 L utilizando un banco de células de investigación (RCB), y luego se verificó mediante series de consolidación a

escala de 20 L utilizando un banco de células de trabajo (WCB) conforme a las GMP. Por último, la expresión se transfirió a una instalación de Buenas Prácticas de Fabricación (cGMP) a escala de 200 L, lo que dio lugar a rendimientos por lotes de 16–18 kg de cuerpos de inclusión (IB) húmedos. El proceso posterior para la purificación de la proteína MYDGF a partir de los cuerpos de inclusión se desarrolló primero a escala de laboratorio, luego se verificó mediante series de consolidación a escala piloto utilizando cuerpos de inclusión de una alícuota de fermentación de 10 L y finalmente se transfirió a una instalación cGMP donde un lote posterior comienza con 10 kg de MI húmedos que representan una alícuota de fermentación de aproximadamente 110–125 L.

Se fabricaron lotes no clínicos y clínicos de MYDGF a escala de 200 L. Se realizaron varios lotes en las instalaciones GMP. Los lotes dieron lugar a altos rendimientos de 330–355 g de MYDGF a partir de una alícuota de fermentación de 125 L. Esto refleja un rendimiento de hasta 2,84 g/L de fermentación. Esto refleja un rendimiento de hasta 2,84 g/L de fermentación. El MYDGF producido por este proceso cumplía todos los requisitos de calidad necesarios para su uso en estudios toxicológicos y clínicos. El contenido de monómero medido por cromatografía de exclusión por tamaño HP fue rutinariamente superior al 99 %, con impurezas de alto peso molecular (agregados) inferiores al 1 % e impurezas de bajo peso molecular (fragmentos) inferiores al 0,1 %. El contenido de endotoxinas fue inferior a 0,03 UE/mg de proteína. El contenido de ADN de la célula huésped era ≤ 3 pg/mg de proteína.

Ejemplo 1: Preparación de vectores y transfección

Para la producción de un banco de células, se utilizó un derivado de la cepa BL21 (DE3) de *Escherichia coli* que se había modificado para que no produjera fagos. La cepa se transformó con uno de los vectores establecidos en la SEQ ID NO: 7–10 portador del gen que codifica la variante respectiva de MYDGF. Los genes de las respectivas variantes se optimizaron mediante codones para obtener altas tasas de expresión en *E. coli* y fueron sintetizados por ATUM (Newark, California, EE.UU.). Se produjeron plásmidos que codifican las siguientes variantes de MYDGF:

- Variante [+A] en donde el residuo V N-terminal en la posición +1 del MYDGF humano maduro está precedido por un residuo A (secuencia aa establecida en la SEQ ID NO: 1:

AVSEPTTVAFDVRPGGVHSFSHNVGPGDKYTCMFTYASQGGTNEQWQMS
 LGTSEDHQHFTCTIWRPQGKSYLYFTQFKAIEVRGAEIEYAMAYSKAAFERE
 SDVPLKTEEFVTKTAVAHPRGAFKAELSKLVIVAKASRTEL). El vector de
 expresión que codifica esta variante se expone en SEQ ID NO: 7.

- Variante [+S] en donde el residuo V N-terminal en la posición +1 del MYDGF humano maduro está precedido por un residuo S (secuencia aa establecida en la SEQ ID NO: 2:

SVSEPTTVAFDVRPGGVHSFSHNVGPGDKYTCMFTYASQGGTNEQWQMSL
 GTSEDHQHFTCTIWRPQGKSYLYFTQFKAIEVRGAEIEYAMAYSKAAFERES
 DVPLKTEEFVTKTAVAHPRGAFKAELSKLVIVAKASRTEL). El vector de
 expresión que codifica esta variante se expone en SEQ ID NO: 8).

- Variante [+G] en donde el residuo V N-terminal en la posición +1 del MYDGF humano maduro está precedido por un residuo G (secuencia aa establecida en la SEQ ID NO: 3:

GVSEPTTVAFDVRPGGVHSFSHNVGPGDKYTCMFTYASQGGTNEQWQM
 SLGTSEDHQHFTCTIWRPQGKSYLYFTQFKAIEVRGAEIEYAMAYSKAAFER
 ESDVPLKTEEFVTKTAVAHPRGAFKAELSKLVIVAKASRTEL. El vector de
 expresión que codifica esta variante se expone en SEQ ID NO: 9.

- Variante [-V] en donde falta el residuo V N-terminal en la posición +1 del MYDGF humano maduro (secuencia aa establecida en la SEQ ID NO: 4:

SEPTTVAFDVRPGGVHSFSHNVGPGDKYTCMFTYASQGGTNEQWQMSL
 GTSEDHQHFTCTIWRPQGKSYLYFTQFKAIEVRGAEIEYA-
 MAYSKAAFERESDVPLKTEEFVTKTAVAHPRGAFKAELSKLVIVAKASRT
 EL. El vector de expresión que codifica esta variante se expone en SEQ ID NO: 10.

En caso de discrepancia entre cualquiera de las secuencias enumeradas anteriormente y las secuencias establecidas en el listado de secuencias adjunto, prevalecerá el párrafo anterior. Los guiones que eventualmente aparezcan dentro de la secuencia son el resultado del truncamiento debido al procesamiento del texto y deben ignorarse.

Para preparar la cepa de expresión, se transformaron células E. coli con los plásmidos vecinos descritos con anterioridad mediante electroporación utilizando un sistema de electroporación Gene Pulser Xcell™ (BioRad). La proteína se expresó en la

célula E. coli en forma de cuerpos de inclusión (IB) que se acumularon en el citoplasma, como se describe en los siguientes ejemplos.

Ejemplo 2: Expresión de variantes de MYDGF

La expresión y purificación de las variantes de MYDGF se realizó mediante el siguiente esquema general:

1. Fermentación
2. Preparación de IB
3. Solubilización/Replegamiento
4. Ultrafiltración/Diafiltración
5. Cromatografía de intercambio aniónico
6. De preferencia: Cromatografía de interacción hidrofóbica
7. De preferencia: Concentración y formulación

Ejemplo 2.1: Fermentación

Se descongeló a temperatura ambiente un vial del banco de células de la cepa de producción obtenida en el Ejemplo 1. Un precultivo (PC) consistió en dos matraces agitados de 1 L con 300 mL de medio de cultivo semilla por matraz. La composición del medio de siembra se muestra en la Tabla 2. Todos los tampones y medios se prepararon con agua de ósmosis inversa (RO) y se esterilizaron antes de su uso utilizando dispositivos de nanofiltración o esterilización por calor.

Tabla 2: Medio de cultivo de semillas

Componente		Cantidad por litro	Unidad
Solución A (Medio Básico)			
1	Dihidrógeno fosfato potásico	3	g
2	Dihidrógeno fosfato potásico	4,58	g
3	Extracto de levadura Bacto	0,5	g

4	Citrato de Sodio x 2H ₂ O	1	g
5	Solución de oligoelementos	0,2	mL
		0,22	g
6	Sulfato de magnesio x 7 H ₂ O	0,4	g
7	Sulfato de amonio	1,8	g
8	Cloruro de amonio	1,48	g
9	Agua de ósmosis inversa, Ad	900	mL
		906,93	g
10	Antiespumante (si es necesario)		mL
Solución B (Adiciones estériles)			
11	Cloruro de calcio 2H ₂ O	0,08	g
12	Glucosa anhidra	10	g
13	Agua de ósmosis inversa, ad	100	mL
		103,74	g
14	Antibiótico (si es necesario)		mL

Se añadieron antiespumantes y antibióticos según fuera necesario. Se añadió kanamicina como antibiótico (a una concentración final de 50 µg/mL). Cada matraz agitado se inoculó con 100 µL de la cepa de producción. Los cultivos se cultivaron durante aproximadamente 9,4 h, con el objetivo de alcanzar una DO (densidad óptica a 550 nm) de $1,75 \pm 0,5$.

El cultivo principal (MC) se realizó en un biorreactor bruto de acero inoxidable de 20 L que contenía 10 L de medio discontinuo. La composición del medio de cultivo se muestra en la Tabla 3.

Tabla 3: Lote Medio

Componente		Cantidad por litro	Unidad
Solución A (Medio Básico)			
1	Dihidrógeno fosfato potásico	3	g
2	Fosfato ácido di potásico	4,58	g
3	Citrato de Sodio x 2 H ₂ O	1	g
4	Solución de oligoelementos	0,2	mL
		0,22	g
5	Sulfato de magnesio x 7 H ₂ O	0,4	g
6	Sulfato de amonio	1,8	g
7	Cloruro de amonio	1,48	g
8	Agua de ósmosis inversa, ad.	900	mL
		906,93	g
9	Polipropilenglicol 2000	1	mL
		1	g
Solución B (Adiciones estériles)			
10	Cloruro de calcio x 2 H ₂ O	0,08	g
11	Glucosa anhidra	10	g
12	Agua de ósmosis inversa, ad.	100	mL
		103,74	g

El medio por lotes se inoculó con 100 mL de caldo celular procedente del precultivo. En la fase discontinua y en la fase de alimentación exponencial, los parámetros del proceso

de fermentación se mantuvieron constantes a 33,5 °C, pH 6,8, 1,0 bar de presión de cabeza y un valor de consigna de OD del 20 %. Se inició una alimentación exponencial (concentración de 600 g/L de glucosa; $\mu = 0,25 \text{ h}^{-1}$) después de que se observara el agotamiento de la fuente de carbono a través de un pico de oxígeno disuelto (OD). La composición del medio de alimentación se muestra en la Tabla 4.

Tabla 4: Medio de alimentación

Componente		Cantidad por litro	Unidad
Solución A (Medio Básico)			
1	Glucosa anhidra	600	g
2	Cloruro de calcio x 2 H ₂ O	5,28	g
3	Agua de ósmosis inversa, ad.	748,99	mL
		962	g
Solución B (Adiciones estériles)			
4	Dihidrógeno fosfato potásico	7,8	g
5	Fosfato ácido di potásico	11,9	g
6	Citrato de Sodio x 2 H ₂ O	66	g
7	Sulfato de magnesio x 7 H ₂ O	26,4	g
8	Solución de oligoelementos	13,2	mL
		14,65	g
9	Agua de ósmosis inversa, ad.	238,88	mL
		292,03	g
10	NaOH al 10	14,4	mL

		15,97	g
--	--	-------	---

Tras 9 h de alimentación exponencial (de 60,48 a 573,78 g/h), la velocidad de alimentación se mantuvo constante a una velocidad de 573,78 g/h durante el resto de la fermentación (11,5 h). Se inició una rampa de temperatura de 60 min (de 33,5 °C a 30,0 °C) 11,5 h después del inicio de la alimentación exponencial, y las rampas se completaron directamente antes de llevar a cabo la inducción con IPTG. El cultivo se indujo mediante un bolo de IPTG, 12,5 h después del inicio de la alimentación. El MC se terminó 20,5 h después del inicio de la alimentación. Al final del proceso de cultivo, el caldo de cultivo se enfrió inmediatamente a <12 °C, se diluyó con agua de ósmosis inversa (RO) hasta un peso celular húmedo (WCW) objetivo del 15 % y la masa celular bacteriana se separó del sobrenadante mediante centrifugación con una centrifugadora CEPA. La biomasa se cosechó y, junto con el sobrenadante, se transfirió al procesamiento posterior.

La cuantificación de los productos se realizó mediante el sistema LabChip GXII® (Perkin Elmer), que constituye una alternativa automatizada de alto rendimiento a los métodos tradicionales de SDS–PAGE y cuantificación de proteínas. La preparación de las muestras se realizó con un sistema de manipulación de líquidos (Tecan Freedom EVO 150). Para la cuantificación de productos a partir de muestras de fermentación, se facilitó la disrupción celular analítica de las muestras de fermentación mediante lisis celular enzimática. Se diluyeron 90 µL de suspensión de fermentación en una proporción 9:10 (v/v) con tampón de disrupción celular (Lysonase™ (Merck) en reactivo de lisis celular FastBreak™ (Promega) con 32 µL de Lysonase por 1 mL de reactivo FastBreak™). Para la determinación del producto total (fracción soluble e insoluble), las muestras se mezclaron antes de cada etapa de pipeteo. Por último, las muestras se diluyeron en el tampón de muestra específico del sistema. Para minimizar la cantidad de tampón de muestra necesaria, todas las etapas de dilución se realizaron en PBS u otro tampón de formulación.

Para la dilución final, se diluyeron 8 µl de muestra (de las diluciones PBS) o las muestras de la curva estándar en 28 µl de tampón de muestra no reductor en una placa de 96 pocillos (placa Eppendorf twin.tec PCR 95100401). Para las condiciones reductoras, se utilizaron 28 µl de tampón de muestra reductor (con 35 mM de DTT). La placa se selló con papel de aluminio (Eppendorf PCR foil 0030127790), se centrifugó brevemente (30 s a 25

g) y se desnaturalizó durante 10 min a 70 °C. Tras la desnaturalización, la placa se centrifugó durante 5 min a 2200 g para centrifugar el líquido evaporado. Tras la centrifugación, se retiró la lámina y se diluyó con 140 µl de agua desionizada. La placa de 96 pocillos se selló con una lámina (Eppendorf twin.tec PCR plate 95100401) y la placa se centrifugó durante 10 min a 2200 g para sedimentar cualquier posible agregado que pudiera causar un mal funcionamiento del análisis del LabChip. Tras la centrifugación, la placa se analizó en el LabChip GXII con el ajuste “HT Protein Express 100 High Sensitivity”. La preparación del LabChip se llevó a cabo de acuerdo con la guía de fabricación. La curva estándar se preparó diluyendo el material de referencia. Como material de referencia, se utilizó la variante [+G] que se había producido en células HEK 293–6E como se describe en Polten (2019), véase la página 1303, 1ª columna y Figura S1, y Ebenhoch (2019), véase la página 8 col. 1.

La cuantificación se realizó en el intervalo de 1 mg/mL a 0,1 mg/mL mediante un ajuste lineal. Las condiciones reductoras y no reductoras no cambiaron el área integral para la cuantificación, sin embargo se observó un cambio en el tiempo de ejecución que no influyó en la cuantificación.

Ejemplo 2.2: Purificación y análisis de cuerpos de inclusión

La biomasa congelada de *E. coli* obtenida en el Ejemplo 2.1 se resuspendió 1:5 (p/v) con el tampón de preparación IB 1 (1 M de urea, 50 mM de Tris, 0,1 % (v/v) de polisorbato 20, pH 7,5). Tras 15 minutos de resuspensión con un ultraturrax, las células de *E. coli* se disgregaron mediante homogeneización a alta presión utilizando 3 pasadas a 650 – 700 bar. Los cuerpos de inclusión (BI) densos y pesados, así como los fragmentos celulares grandes, se separaron mediante centrifugación tubular de alta velocidad utilizando un rotor GLE (CEPA). El caudal de alimentación fue de 55 mL/min y la velocidad de centrifugación de 24.500 g. El tubo tenía un diámetro interior de 3,2 mm. El pellet recuperado se lavó dos veces con agua HQ. En todos los casos, los pellets se diluyeron 1:5 (p/v) y se volvieron a suspender utilizando un ultraturrax. Tras las etapas con agua HQ, el pellet contiene principalmente IB.

Ejemplo 2.3: Solubilización, replegamiento y purificación de proteínas

Las IB congeladas obtenidas del Ejemplo 2.3 se solubilizaron a temperatura ambiente en tampón de solubilización (8 M de urea, 0,14 M de GuHCl, 6 mM de DTT, 50 mM de Tris, pH 8). La mezcla se agitó primero con un ultraturrax durante 10 minutos y después con un mezclador de hélice durante 180 minutos. La concentración objetivo durante la solubilización fue de 5 mg/mL con un volumen objetivo de 100 mL. Posteriormente, el pool de solubilización se filtró sobre un filtro de profundidad CUNO (filtro E16E01A90ZB08A, 0,1 – 0,6 μ m, 3M Deutschland GmbH, Neuss, Alemania). Los filtros se preequilibraron con agua para inyección (WFI) y tampón de solubilización. A continuación, se cargó directamente el pool de solubilización. El filtrado se recogió mediante monitorización UV utilizando un sistema ÄKTA. El solubilizado del cuerpo de inclusión se diluyó con tampón de refold 1:5 (4 M de urea, 0,3125 M de Tris, 12,5 mM de CaCl_2 , 3,75 mM de cistamina, pH 7). El pool de replegamiento recuperado se agitó durante toda la noche. Al día siguiente, se filtró con un filtro de profundidad CUNO (filtro E16E01A60ZB05A, 3M Deutschland GmbH, Neuss, Alemania).

Tras la filtración con el filtro de profundidad, el filtrado se sometió a ultrafiltración/diafiltración (UFDF) para realizar un intercambio tampón. La UFDF se realizó con una membrana Pellicon 3 (88 cm^2 , Ultracell 3 kDa, malla tipo C) utilizando un tampón de diafiltración (20 mM de Tris, pH 9). Se utilizó un factor de concentración de 2 y un factor de diafiltración de 5.

Tras la UFDF, el filtrado se sometió a cromatografía de intercambio iónico (IEX). Se utilizó una columna YMC Biopro IEX de 75 μ m con un diámetro de columna de 1 cm, una altura de lecho de 9 cm y un volumen de columna de 7,5 mL. La columna se equilibró primero durante 5 minutos con 3 volúmenes de columna (CV) de tampón de equilibrio 1 (20 mM Hepes, 1 M NaCl pH 7) y posteriormente durante 5 minutos con 5 CV de tampón de equilibrio 2 (20 mM de Tris pH 9). Después de cargar el filtrado, éste se lavó durante 5 min con 5 CV 20 mM de Tris pH 9. La proteína se eluyó durante 5 min con 5 CV de tampón de elución 1 (20 mM Hepes, 1 M NaCl, pH 7) seguido de 5 min con 10 CV de tampón de elución 2 (20 mM Hepes, pH 7) utilizando un gradiente (0 % a 100 % 20 mM Hepes, 1 M NaCl pH 7). Por último, la columna se decapó durante 5 min con 5 CV con 1 M HCl.

Ejemplo 2.4: Rendimiento de proteínas y homogeneidad del producto

La purificación se realizó con las 4 variantes al menos una vez. Además, se realizó un segundo experimento de purificación con las variantes [+A] y [+S]. La purificación dio como resultado un alto rendimiento y una elevada pureza para cada una de las cuatro variantes. En particular, el rendimiento global del proceso tras la purificación y el replegamiento de la variante [+A] fue de 2,4 g/L en el primer lote y de 5,3 g/L en el segundo.

Se realizó una cromatografía analítica de exclusión por tamaño de alto rendimiento para comprobar la pureza. La variante [+A] purificada mostró una elevada pureza del 99,75 % de pico principal, 0,25 % de impurezas de bajo peso molecular y 0,0 % de niveles de agregados. También se alcanzaron altos niveles de pureza con la variante [–V] (99,64 % de pico principal, 0,0 % de impurezas de bajo peso molecular, 0,04 % de agregados) y la variante [+S] (99,73 % de pico principal, 0,2 % de impurezas de bajo peso molecular, 0,05 % de agregados). En cambio, el producto de la variante [+G] fue menos homogéneo cuando se examinó con cromatografía de exclusión por tamaño de alto rendimiento, mostrando un 60,36 % de pureza del pico principal y un 39,64 % de agregados.

Los rendimientos del proceso de purificación a escala de laboratorio para todas las variantes N-terminal se resumen en la siguiente tabla:

Tabla 5a: Resumen de la producción a escala de laboratorio de variantes N-terminales de MYDGF. La cantidad de MYDGF en las diferentes etapas del proceso se indica en mg de MYDGF

		Variante de MYDGF			
		[+A] variante	[+G] variante	[–V] variante	[+S] variante
Densidad celular al final de la fermentación	Densidad óptica (DO 550 nm)	326	332	348	331
Peso de células húmedas al final de la fermentación	(cantidad de células húmedas g por L de volumen de fermentación)	330,17	310,42	337,58	311,65

Título del producto al final de la fermentación	(cantidad MYDGF: mg por L de volumen de fermentación)	27,1	23,1	26,8	24,2
---	---	------	------	------	------

El proceso de fermentación por lotes alimentados aplicado a las cuatro variantes dio lugar a densidades celulares muy elevadas al final de la fermentación (DO a 550 nm de 326–348; masa celular húmeda de 310,42 – 337,58 g/L). Se alcanzaron títulos volumétricos muy elevados para las variantes recombinantes de MYDGF (23,1–27,1 g/L de fermentación).

Zhao et al. (2020) han informado de la fermentación de una proteína de fusión MYDGF–6His en *E. coli* seguida de su extracción y purificación a partir de lisado celular de *E. coli*.

Tabla 5b: Características del proceso según Zhao et al. (2020) en comparación con el proceso según la invención

	Zhao et al. 2019	Ejemplo 2	Ejemplo 3	Comentarios
productividad global (relacionada con la cantidad de masa celular)	216 mg de (+Met)–MYDGF–6His purificado por 100 g de células húmedas	788 mg de (+Ala)–MYDGF purificado por 100 g de células húmedas	825–888 mg de sustancia activa purificada (+Ala)–MYDGF por 100 g de células húmedas	El proceso de la invención proporciona una productividad 4 veces mayor en relación con la masa celular. Se ha demostrado que esta productividad es escalable.
Productividad global (relacionada con el volumen de fermentación)	Se desconoce. Sin embargo, basándose en la hipótesis de un máximo de 10 g de masa celular por litro de volumen de fermentación, se podría suponer una productividad de un máximo de 21,6 mg (+Met)–MYDGF–6His por litro de caldo de fermentación.	Productividad global de 2,6 g de (+Ala)–MYDGF purificado por litro de fermentación.	Productividad global de 2,6–2,8 g de (+Ala)–MYDGF purificado y formulado por litro de fermentación.	Se supone que el proceso de la invención es > 100 veces más productivo que el proceso de Zhao

Tabla 5c: Resumen de la purificación de rhMYDGF según lo divulgado en Zhao et al. (2020)

cf.: tabla S2 en Zhao et al. (2020)	Células húmedas (g)	Volumen (ml)	Concentración de proteínas (mg/ml)	Proteínas totales (mg)	Pureza (%)	Rendimiento (mg/100 g de bacterias húmedas)
Solubilización de proteínas	100	1000	0,82	820	46 %	377,2
Columna de quelación de níquel		328	0,93	305,04	80 %	244,032
Cromatografía en columna de filtración de gel		150	1,5	225	96 %	216

* Véase Zhao et al. (2020), en la página 1192, sección 3.1, líneas 18-20, que hace referencia a una pureza superior al 95% con referencia a la Fig. 1C. Véase también la Fig. 2A.

Los datos anteriores son la media de tres experimentos independientes. La proteína se cuantificó por el método BCA. La cantidad de proteínas diana se estimó mediante análisis de densitometría de la banda de proteínas en geles SDS–PAGE. Proteína total = concentración de proteína (mg/mL) × volumen (mL). Rendimiento = proteína total (mg) × pureza (%).

Ejemplo 3: Proceso avanzado de fabricación de MYDGF

Basándose en el Ejemplo 2, se siguió desarrollando el proceso de fabricación para la variante [+A]. El proceso de fermentación para MYDGF se desarrolló primero a escala de 5 L utilizando el Banco de Células de Investigación (RCB), luego se verificó mediante corridas de consolidación a escala de 20 L utilizando el Banco de Células de Trabajo GMP (WCB) y finalmente se transfirió a escala de 200 L. Un lote típico de fermentación de 200 L produce entre 16 y 18 kg de IB húmedos.

El proceso posterior para la purificación de la sustancia activa MYDGF a partir de los cuerpos de inclusión intracelulares se desarrolló primero a escala de laboratorio, luego se verificó mediante pruebas de consolidación a escala piloto utilizando cuerpos de inclusión de una alícuota de fermentación de 10 L y, finalmente, se transfirió a una instalación cGMP donde un lote posterior comienza con 10 kg de IB húmedos que representan una alícuota de fermentación de unos 110–125 L.

Tabla 6a: Descripción del proceso de fabricación de la sustancia farmacéutica CMC1a (parte anterior del proceso)

Etapa del proceso	Descripción de la etapa del proceso	Función primaria de la etapa del proceso
Banco de Células	Fabricación de Master Maestro (MCB) y bancos de células de trabajo (WCB). El MCB se derivado del Banco de Células de Investigación (RCB).	
Cultivo de semillas 1	El cultivo de semillas (SC) se realiza en matraces agitados que contienen 310 ml de medio esterilizado. El medio de cultivo de semillas esterilizado se inocula con 100 µl de WCB. A continuación, se cultiva el cultivo de semillas y se termina en función de la DO550 en un matraz de referencia de DO. El SC se transfiere al cultivo principal.	generar suficiente inóculo para el cultivo principal
Cultivo principal	El cultivo principal se lleva a cabo como una fermentación por lotes alimentados.	ampliar el cultivo de semillas y

	La formación del producto se induce mediante la adición en bolo de una solución de IPTG.	expresión de la proteína diana
Cosecha de biomasa	El caldo de fermentación se enfría y la biomasa se recoge por centrifugación utilizando una centrifugadora de cuba tubular.	recuperación y concentración de células huésped
Resuspensión de biomasa	La biomasa se resuspende y se diluye con un tampón de resuspensión.	Resuspensión de la biomasa cosechada
Homogeneización	La biomasa resuspendida se homogeneiza en tres pasadas en un homogeneizador de alta presión a 650 bares. El lisado se recoge en un tanque móvil refrigerado.	Disrupción de las células de E. coli
Cosecha de IB 1	Recolección de cuerpos de inclusión (= Centrifugación del homogeneizado) se realiza con una centrifugadora de cuba tubular.	Recogida de cuerpos de inclusión del homogeneizado
Lavado de IB 1	Las IB 1 se recogen tras centrifugar el homogeneizado y se diluyen 1:5 (p/v) en agua purificada.	Lavado de los cuerpos de inclusión 1 para eliminar restos celulares e impurezas
Cosecha de IB 2	La recolección de los cuerpos de inclusión (= centrifugación del homogeneizado) se realiza con una centrifugadora de cubeta tubular.	Recogida de cuerpos de inclusión 2 de cuerpos de inclusión 1
Lavado de IB 2	Las IB 1 se recogen tras centrifugar el homogeneizado y se diluyen 1:5 (p/v) en agua purificada.	Lavado de los cuerpos de inclusión 2 para eliminar restos celulares e

		impurezas
Cosecha de cuerpos de inclusión final	La recolección de los cuerpos de inclusión (= centrifugación del homogeneizado) se realiza con una centrifugadora de cubeta tubular.	Cosecha final cuerpos de inclusión
Embalaje de IB final	El sedimento final de IB se envasa en alícuotas de 500 g de IB y se almacena a -20 °C hasta su posterior procesamiento.	Embalaje de la inclusión final en contenedores desechables desechables

Tabla 6b: Descripción del proceso de fabricación de sustancias medicamentosas CMC1a (parte posterior)

Etapas del proceso	Descripción de la etapa del proceso	Función primaria de la etapa del proceso
Solubilización	Los cuerpos de inclusión congelados se solubilizan a una concentración diana de 32 g IB/L utilizando un tampón que contiene una alta concentración de caotrops y condiciones reductoras.	Solubilización de cuerpos de inclusión congelados
Filtración de profundidad 1	La solución IB se filtra a través de un filtro de profundidad.	Aclaración de la solución IB Reduce partículas
Replegamiento	La solución de IB solubilizada y filtrada se repliega añadiendo tampón de replegado en una proporción de 1:4 y subsiguiente incubación bajo agitación constante.	Formación de proteína similar a la nativa +A variante de MYDGF
UFDF1	El replegamiento se concentra primero con un factor de 1,5, seguido	Reducción del volumen del proceso

	de una diafiltración continua con al menos 5 diavolúmenes.	Eliminación de caotrops y replegamiento de aditivos. Disminución de la conductividad
Filtración de profundidad 2	El pool de UFDF1 se filtra a través de un filtro de profundidad.	Aclaración de la solución IB Agotamiento de impurezas relacionadas con el huésped y agregados de productos
Cromatografía AEX	El conjunto de UFDF1 filtrado se carga en una columna de cromatografía de intercambio aniónico. Las impurezas no unidas se eliminan lavando con tampón bajo en sal. El producto se recupera por elución en gradiente.	Reducción de volumen Agotamiento de impurezas relacionadas con el proceso y el producto
Cromatografía HIC	La mezcla de AEX se diluye primero con un tampón que contiene una alta concentración de sal. La carga de HIC resultante se carga luego en una columna de cromatografía de interacción hidrofóbica. Las impurezas no unidas se eliminan lavando con una solución amortiguadora con alto contenido de sal. El producto se recupera por elución en gradiente.	Agotamiento de impurezas relacionadas con el proceso y el producto
UFDF 2	El conjunto de HIC primero se concentra a una concentración objetivo de ~20 g/l usando una	Concentración del producto Agotamiento de impurezas relacionadas con el proceso de bajo peso molecular.

	membrana con un corte nominal de 5 kDa. Luego, el retenido se diafiltra con una solución de cloruro de sodio (5 diavolúmenes) seguido de 6 diavolúmenes de Tris-HCl 20 mM, pH 8,5. El retenido luego se concentra a una concentración >70 g/L y se captura del sistema.	
Formulación	Se agrega tampón de formulación de 5 veces al diafiltrado 2 de UFDF 2.	Ajustar la concentración de excipiente y llegar a 50 ± 5 g/L, pH $8,5 \pm 0,2$
Filtración DS y alicuotación	Filtración mediante un colector preesterilizado equipado con un filtro de $0,2 \mu\text{m}$; el fármaco a granel se filtra en una bolsa para su homogeneización; alícuota de DS en bolsas de 6 L; DS se congela a -60°C a $\geq -80^\circ\text{C}$ se almacena a $-40 \pm 5^\circ\text{C}$	Eliminación de carga biológica Embalaje y almacenamiento

Se realizaron varios lotes en condiciones GMP. Los lotes dieron lugar a altos rendimientos de 330–355 g de MYDGF a partir de una alícuota de fermentación de 125 L. Esto refleja un rendimiento global del proceso de hasta 2,84 g/L de fermentación. Esto refleja un rendimiento global del proceso de hasta 2,84 g/L de fermentación. La sustancia farmacológica MYDGF producida mediante este proceso cumplía todos los requisitos de calidad necesarios para su uso en estudios toxicológicos y clínicos.

El contenido de monómeros medido por cromatografía de exclusión por tamaño HP fue habitualmente superior al 99 %, con impurezas de alto peso molecular (agregados)

inferiores al 1 % e impurezas de bajo peso molecular (fragmentos) inferiores al 0,1 %. El contenido de endotoxinas fue inferior a 0,03 EU por mg de proteína MYDGF. El contenido de ADN de la célula huésped era ≤ 3 pg/mg de proteína.

Ejemplo 4: Análisis del peso molecular por LCMS y análisis avanzado del peso molecular por LCMS tras modificación química según Tolonen et al. (“aLCMS”)

Muestras del producto plegado y purificado obtenido del Ejemplo 2 se sometieron a análisis de espectrometría de masas por cromatografía de líquidos (LCMS). Se utilizó cromatografía líquida/espectrometría de masas por ionización de electroespray (LC–ESI–MS) para realizar un análisis del peso molecular intacto (no reducido) de las construcciones MYDGF para (1) verificar la secuencia mediante la conformidad del peso molecular observado con los valores predichos para cada secuencia, y (2) capturar un perfil global de las modificaciones postraduccionales netas (PTM) en cada proteína. Se utilizó un UPLC Agilent 1290 con una columna de fase inversa C3 POROS de 1,0 mm por 30 mm para desalar e introducir las muestras (0,5µg/inyección) en el espectrómetro de masas. Se utilizó un gradiente binario de tres minutos consistente en la fase móvil A (98,9 % de agua, 1 % de acetonitrilo, 0,1 % de ácido fórmico y 2 mM de acetato de amonio) y la fase móvil B (70 % de isopropanol, 20 % de acetonitrilo, 9,9 % de agua y 0,1 % de ácido fórmico) que aumentaba del 5 % al 80 % de la fase móvil B a 150 µl/min para atrapar, desalar y eluir la proteína de la columna. Los datos espectrales de masas del material eluido se adquirieron utilizando un MS de tiempo de vuelo (TOF) Agilent 6224, que luego se procesó (deconvolucionado) utilizando el algoritmo de entropía máxima dentro del software de análisis Mass Hunter (Agilent). Los datos obtenidos por este método se denominan aquí “datos LCMS de MW intacta” o datos “medidos por espectrometría de masas de cromatografía líquida (LCMS)”.

Para la confirmación de la secuencia a nivel de péptido y el análisis de la modificación postranscripcional (PTM) específica del sitio, se digirieron alícuotas de cada muestra utilizando tripsina y quimotripsina separadamente para lograr una cobertura completa de la secuencia. Se desalaron 100 µg de cada muestra y se concentraron mediante precipitación con acetona y precipitación centrífuga del material precipitado. Cada gránulo de proteína se volvió a solubilizar, desnaturizó y redujo en 10 µl de tampón de desnaturalización/reducción (5 % p/v de desoxicolato sódico [SDC], 10 mM de ditiotreitól

[DTT], 20 mM de bicarbonato amónico) y se incubó a 70 °C durante 2 minutos, seguido de una dilución diez veces mayor con 20 mM de bicarbonato amónico y 2 mM de metionina. A continuación, las moléculas reducidas/desnaturalizadas se dividieron en dos viales (50 µg cada uno), a los que se añadieron tripsina y quimotripsina por separado en una proporción de 1:10 enzima–sustrato a cada tubo, y las muestras se incubaron durante 10 minutos a 37 °C. La reacción se apagó con la adición de ácido trifluoroacético al 10 % v/v, lo que dio lugar a una concentración final del 1 % v/v de dicho reactivo. La breve etapa de digestión (10 minutos) evitó la necesidad de una etapa de alquilación, que se utiliza habitualmente en el mapeo de péptidos. El desoxicolato sódico precipitado se eliminó mediante centrifugación a 16.000 x g, y el sobrenadante que contenía péptidos se recuperó y se transfirió a viales de automuestreador que se almacenaron inmediatamente a –80 °C hasta su análisis. Los datos obtenidos por este método se denominan en el presente documento “datos LCMS de mapeo peptídico”.

aLCMS: Además, como se ha demostrado que los primeros cuatro residuos N–terminales de los diversos constructos MYGDF sufren fragmentación durante el electrospray (tanto a nivel intacto como peptídico), se realizó dimetilación reductora (también conocida como etiquetado de dimetil isótopo estable (SIDL)) de acuerdo con Tolonen et al. (2019), en alícuotas de los digestos peptídicos para delinear entre truncamientos N–terminales derivados de la muestra frente a los derivados del electrospray. Los datos obtenidos por este método se denominan en el presente documento “LCMS tras dimetilación reductora (marcado isotópico estable con dimetilo, SIDL)”. Brevemente, 50 µg de péptidos de cada digestión se inmovilizaron en cartuchos SPE Waters Oasis separados utilizando un colector de vacío. Los medios de SPE y los péptidos se acondicionaron a pH 5,5 con tampón citrato (90 mM de ácido cítrico, 230 mM de fosfato sódico divalente), y se pasaron 10 ml de formaldehído al 0,8 % v/v (en tampón citrato) y cianoborohidruro sódico 120 mM sobre los péptidos unidos durante 10 minutos. A continuación, se eliminaron los reactivos lavando con 10 volúmenes de columna de ácido fórmico al 0,1 % en agua, y se eluyeron con 10 volúmenes de acetonitrilo al 50 %, ácido fórmico al 0,1 %. Los péptidos marcados se recogieron en un tubo de microcentrífuga de baja retención y se llevaron a sequedad en una centrífuga de vacío. Los péptidos secos se reconstituyeron en 50 µl de TFA al 0,1 % y se transfirieron a un vial de automuestreador para el análisis LC–MS/MS.

El análisis LC–MS/MS (espectrometría de masas en tándem) se realizó utilizando un sistema Vanquish UHPLC interconectado con un Lumos Fusion Orbitrap (ThermoFisher) que funcionaba bajo el control del software Xcalibur 4.1.31.9 (ThermoFisher). Se cargaron 0,5 µg de cada digestión peptídica en una columna de fase inversa C18 CSH Acquity UPLC de 2,1 mm x 150 mm (1,7 µm de partícula, Waters Corp.), y se separaron utilizando un gradiente binario como sigue: (fase móvil A = 0,1 % de ácido difluoroacético (DFA) en agua) 0,5 % a 40 % de fase móvil B (99,9 % de acetonitrilo, 0,1 % de DFA) a un caudal de 200 µl/min y una temperatura de columna de 50 °C. Se utilizó un flujo de trabajo de MS de adquisición dependiente de datos (DDA) top–4 para analizar el eluido de LC. Los espectros MS de barrido completo se adquirieron con una resolución de 120.000 (FWHM) a 200 m/z, y los espectros MS/MS HCD (disociación colisional de alta energía) y EThcD (disociación por transferencia de electrones con energía HCD suplementaria) se adquirieron de forma dependiente del estado de carga con una resolución de 15.000 en el analizador Orbitrap. Los archivos .RAW resultantes de cada análisis LC–MS/MS se procesaron con los programas Byonic y Byos de Protein Metrics Inc. (PMI) para identificar y cuantificar las PTM. El análisis manual de varios espectros se realizó utilizando la función QualBrowser del software Xcalibur.

Tabla 7: Variantes de MYDGF examinadas por aLCMS o LCMS:

N.º	Tipo de MYDGF (...)	Sistema huésped	Tipo de expresión
1	+S ... variante	E. coli	Insoluble
2	–V... variante	E. coli	Insoluble
3	+A... variante	E. coli	Insoluble
4	+G ...variante	E. coli	Insoluble
5	wt	CHO	Soluble

Tabla 8: Datos LCMS MW intactos para la variante +S de MYDGF

Muestra	Categoría de PTM ⁽¹⁾	Descripción	Δ MW (Da) predicho	MW predicho(Da)	MW observado(Da)	Porcentaje relativo
+S	Longitud total	no modificado	0	15.920	15.920	80,9
	Adición N-term	+M	131	16.051	16.051	3,6
	Truncaciones N-term	-SVSE	-402	15.518	15.518	1,5
		-SVS	-258	15.646	15.647	0,2
		-SV	-186	15.733	15.734	1,6
		-S	-87	15.832	15.833	0,6
	PTM adicionales	Deshidratación	-18	15.902	15.902	1,4
		Na+	22	15.942	15.941	3,7
		Carbamoilación	43	15.963	15.962	4,5
		gluconoilación	178	16.098	16.098	2,0

(PTM = modificaciones postraduccionales). El aducto de Na⁺ en el análisis intacto es un artefacto común, no un atributo molecular. El término “N-term” en el mapa peptídico se refiere al grupo amino del N-terminal. La cobertura de la secuencia es del 100 %. N.D.= PTM < límite inferior de detección

Tabla 9: Datos LCMS de mapeo peptídico para la variante +S MYDGF

Modificación potencial		Región de unión		Región no de unión
	Residuo	Modificación en porcentaje	Residuo	Modificación en porcentaje
Variantes de secuencias	n.a.	n.a.	N-term no modificado	84,6*
			N-term +M	3,6*
			N-term -SVSE	1,5*
			N-term -SVS	0,2*
			N-term -SV	1,6*
			N-term -S	0,6*
Oxidación (16)	Y70, Y72	N.D.	M34	0,5
			M49	0
			M91	1,3
			W47	N.D.
			W65	N.D.
Isomerización (0)	n.a.	n.a.	D11	N.D.
			D29	
			D56	
			D103	
Gluconoilación (178)	K70	N.D.	N-term	1,4
	K97		K28	N.D.
	K107		K77	N.D.

	K126		K113	N.D.
	K131		K135	N.D.
Carbamoylación (43)	K70	N.D.	N-term	3,1
	K97		K28	N.D.
	K107		K77	N.D.
	K126		K113	N.D.
	K131		K135	N.D.
Acetilación (42)	K70	N.D.	N-term	2,4
	K97		K28	N.D.
	K107		K77	N.D.
	K126		K113	N.D.
	K131		K135	N.D.
Deshidratación (succinimida) (-18)	n.a.	n.a.	D29	0
			D56	4,2
Metilación/glicación/glicosilación	K70	N.D.	N-term	N.D.
	K97		K28	N.D.
	K107		K77	N.D.
	K126		K113	N.D.
	K131		K135	N.D.

*: determinado a partir del análisis de Mw intacto. El aducto Na⁺ añadido al valor original

Tabla 10: Datos LCMS MW intactos para la variante –V MYDGF

Muestra	Categoría PTM	Descripción	Δ MW (Da) predicho	MW predicho (Da)	MW observado (Da)	Porcentaje relativo
–V	Longitud total	no modificado	0	15.733	15.734	74,0
	Truncación N-term	– SE	–216	15.517	15.518	3,7
	PTM adicionales	Deshidratación	–18	15.715	15.716	1,9
		Na ⁺	22	15.755	15.754	4,4
		Carbamoilación	43	15.776	15.776	12,1
		gluconoilación	178	15.911	15.912	3,9

Tabla 11: Datos LCMS de mapeo peptídico para la variante –V MYDGF

Modificación potencial	Región de unión		Región de no unión	
	Residuo	Modificación en porcentaje	Residuo	Modificación en porcentaje
Variantes de secuencias	n.a.	n.a.	N-term no modificado	78,4*
			N-term – SE	3,7*
Oxidación (16)	Y70, Y72	N.D.	M32	2,5
			M47	3,9
			M89	N.D.
			W45	1,2
			W63	N.D.
Isomerización (0)	n.a.	n.a.	D9	N.D.
			D27	
			D54	
			D101	
Gluconoilación (178)	K68	N.D.	Prot. N-term	6,7
	K93		K28	N.D.
	K105		K77	N.D.
	K124		K113	N.D.
	K129		K135	N.D.
Carbamoilación (43)	K68	N.D.	Prot. N-term	9,5
	K93		K28	N.D.
	K105		K77	N.D.
	K124		K113	N.D.
	K129		K135	N.D.
Deshidratación (succinimida) (–18)	n.a.	n.a.	D27	3,1
			D54	0,1
	K68	N.D.	Prot. N-term	

Acetilación / metilación / glicación / glicosilación	K93		K28	N.D.
	K105		K77	
	K124		K113	
	K129		K135	

*determinado a partir del análisis de MW intacto; aducto de Na⁺ añadido al valor original.

Tabla 12: Datos LCMS de MW intacto para la variante +A de MYDGF

Muestra	Categoría de PTM	Descripción	Δ MW (Da) predicho	MW predicho (Da)	MW observado (Da)	Porcentaje relativo
+A	Longitud total	no modificado	0	15.904	15.904	83,1
	Truncaciones N-term	– AVSE	–387	15.517	15.518	1,5
		– AVS	–258	15.646	15.647	0,1
		– AV	–171	15.733	15.733	0,5
		– A	–71	15.832	15.833	0,4
	PTM adicionales	Deshidratación	–18	15.886	15.886	1,7
		Na ⁺	22	15.926	15.924	5,6
		Carbamoilación	43	15.947	15.946	4,6
		Gluconoilación	178	16.082	16.082	2,5

Tabla 13: Datos LCMS de mapeo peptídico para la variante +A de MYDGF

Modificación potencial	Región de unión		Región de no unión	
	Residuo	Modificación en porcentaje	Residuo**	Modificación en porcentaje
Variantes de secuencias	n.a.	n.a.	N-term no modificado	88,7*
			N-term – AVSE	1,5*
			N-term – AVS	0,1*
			N-term – AV	0,5*
			N-term – A	0,4*
Oxidación (16)	Y72, Y74	N.D.	M33	0,3
			M48	0,3
			M91	1,4
			W47	1,4
			W65	N.D.
Isomerización (0)	n.a.	n.a.	D11	N.D.
			D29	
			D56	
			D103	
Gluconoilación (178)	K70	N.D.	Prot. N-term	1,9
	K95		K30	N.D.
	K107		K79	N.D.
	K126		K115	N.D.
	K131		K137	N.D.
Carbamoilación (43)	K70	N.D.	Prot. N-term	2,1
	K95		K30	N.D.
	K107		K79	N.D.
	K126		K115	N.D.

	K131		K137	N.D.
Deshidratación (succinimida)	n.a.	n.a.	D29	2,8
(-18)			D56	3,3
Acetilación	K70	N.D.	Prot. N-term	N.D.
/metilación/glicación	K95		K30	
/glicosilación	K107		K79	
	K126		K115	
	K131		K137	

*forma determinada análisis de masa intacta Aducto Na⁺ añadido al valor original

**No se observó metionina N-terminal.

Tabla 14: Datos LCMS de MW intacto para la variante +G de MYDGF

Muestra	Categoría de PTM	Descripción	Δ MW (Da) predicho	MW predicho (Da)	MW observado (Da)	Porcentaje relativo
+G	Longitud total	sin modificar	0	15.889	15.890	37,1
	PTM adicionales	Deshidratación	-18	15.871	15.872	0,8
		Na	22	15.911	15.910	2,3
		Carbamoilación	43	15.932	15.932	4,2
	Adición N-term	+ Met	131	16.021	16.021	44,1
		Deshidratación	-18	16.003	16.003	1,0
		Na+	22	16.043	16.040	3,4
		Carbamoilación				
		Carbamoilación	43	16.064	16.067	7,1

Tabla 15: Datos aLCMS (LCMS combinada de MW intacto y LCMS a nivel de péptido con tapa capuchón de dimetilo tras tripsinación como se ha descrito*:

R&D Systems, rMYDGF (humano, wt, no marcado) Cat. N.º 10231			(+A) MYDGF variante			(+S) MYDGF variante		
Modificaci ón del término N	Análisi s de proteín as comple tas Porcen taje	Porcenta je de escisión del término N verificad o mediant e mapeo peptídico y metodol ogía aLCMS		Análi sis de proteí nas comp letas Porcen taje	Porcenta je de escisión del término N verificad o mediant e mapeo peptídico y metodol ogía aLCMS		Análisis de proteínas complete s Porcenta je	Porcenta je de escisión del término N verificad o mediant e mapeo peptídico y metodol ogía aLCMS
no modificado	100,0	99,9	no modifica do	93,4	93,2	no modific ado	93,8	93,3
-V		0,0	-A		0,2	-S		0,5
-VS		0,1	-AV		0,0	-SV		0,0
-VSE		0,0	-AVS		0,0	-SVS		0,0
-VSEP		0,0	-AVSE		0,0	-SVSE		0,0
Carbamoi lación	0		Carbamoi lación	4,3		Carbam oilación	4,2	

Gluconoila ción	0		Gluconoi lación	2,3		Glucono ilación	1,9	
--------------------	---	--	--------------------	-----	--	--------------------	-----	--

*No se observó metionina N-terminal en la variante +A; sin embargo, se observó metionina N-terminal en la variante +S en un 2,5%.

Ejemplo 5: Resolución estructural

Se fabricó la variante +G MYDGF (HEK) que tiene una secuencia según la SEQ ID NO: 3, tal como se describe en Polten et al. (2019), y en la página 1303, 1ª columna y Figura S1, Ebenhoch et al. (2019), página 8 col. 1. La variante +A MYDGF se preparó de acuerdo con el Ejemplo 2.

Los espectros bidimensionales de RMN $^1\text{H}/^{15}\text{N}$ HSQC se recogieron en un espectrómetro Bruker Avance III de 800 MHz equipado con una criosonda TCI de gradiente z de 5 mm en tubos de 2,5 mm a 310 K. Los espectros se registraron utilizando el programa de impulsos hsqcftp3gpplhwg (Bodenhausen & Ruben 1980; Piotto et al. 1992; Sklenar et al. 1993; Mori et al. 1995) del catálogo Bruker con 48 puntos complejos en la dimensión indirecta, 1024 barridos, un retardo entre barridos de 1 s, lo que resulta en un tiempo experimental total de 30 h. El programa de impulsos hsqcftp3gpplhwg describe un espectro de correlación H-1/X 2D sensible a la fase mediante una transferencia de doble inepto, que utiliza el canal f3 y emplea el desacoplamiento durante la adquisición, así como impulsos flip-back y una secuencia watergate para la supresión de agua. Las muestras de RMN contenían 8,5 mg/ml de la proteína MYDGF respectiva en tampón fosfato sódico 50 mM a pH 7,4 que contenía 50 mM de cloruro de sodio y 9 % (v/v) de D₂O. El procesamiento y el análisis se realizaron con Topspin 3.5 (Bruker BioSpin)

Los desplazamientos químicos ^1H y ^{15}N de los picos cruzados observados en los espectros 2D $^1\text{H}/^{15}\text{N}$ HSQC RMN muestran claramente que MYDGF es una proteína plegada para ambas variantes, la variante +G (HEK) y la variante +A. La dispersión de los picos cruzados y sus desplazamientos químicos son significativamente mayores y muy diferentes a los desplazamientos químicos aleatorios determinados experimentalmente (Wishart et al. 1995). En esta publicación se ejemplifica una comparación de espectros 2D $^1\text{H}/^{15}\text{N}$ HSQC RMN que comparan proteínas plegadas y desplegadas (Dyson & Wright, 1995). Las diferencias entre la bobina aleatoria y los desplazamientos químicos observados

se utilizan con frecuencia en los cálculos de estructura de RMN como restricciones para las estructuras secundarias (hélices α o láminas β), (Shen & Bax, 2015).

Tabla 16: Variantes examinadas por RMN

N.º	Tipo de MYDGF (...)	Sistema huésped	Tipo de expresión
1	+G ...variante (HEK)	HEK	Soluble
3	+A... variante	E. coli	Insoluble

Tabla 17: Desplazamiento químico ^1H , ^{15}N

	Desplaza-	Desplaza-	22	9,03	124,81
	miento	miento	23	8,86	124,98
	químico ^1H	químico ^{15}N	24	8,81	124,67
Pico N.º	(ppm)	(ppm)	25	8,20	125,80
1	10,37	129,76	26	7,88	125,80
2	10,14	129,36	27	8,44	124,83
3	9,60	129,47	28	8,36	124,50
4	9,63	128,50	29	7,72	124,72
5	9,49	128,58	30	7,60	123,64
6	9,13	129,31	31	7,63	122,68
7	8,87	129,18	32	7,42	122,77
8	9,04	128,69	33	7,91	122,40
9	9,02	128,32	34	8,38	122,53
10	9,48	126,84	35	8,37	123,27
11	9,20	126,47	36	8,50	123,58
12	8,95	126,81	37	8,60	123,33
13	8,79	127,05	38	8,65	123,98
14	8,48	127,45	39	8,77	123,34
15	8,14	128,38	40	8,72	122,90
16	7,84	129,41	41	8,89	122,88
17	8,81	125,63	42	9,00	121,91
18	8,70	125,75	43	8,27	121,54
19	8,25	126,85	44	9,07	122,08
20	9,20	125,74	45	9,12	121,13
21	9,30	124,99	46	9,39	121,60

47	9,33	120,96	75	6,15	117,12
48	9,41	120,69	76	7,71	117,75
49	9,03	120,73	77	8,02	117,15
50	8,37	120,75	78	8,09	117,30
51	8,49	120,29	79	8,18	117,90
55	9,08	119,50	80	8,19	117,22
56	8,82	119,49	81	8,42	118,42
57	8,77	119,60	82	8,52	117,35
58	8,45	119,33	83	8,56	116,70
59	7,99	119,50	84	8,26	116,36
60	8,06	120,13	85	8,67	117,13
61	8,16	120,20	86	8,70	116,53
62	8,30	119,98	87	8,77	117,66
63	8,30	119,68	88	8,77	116,10
64	8,30	118,69	89	8,28	115,50
65	7,66	121,40	90	8,22	115,01
66	7,57	121,52	91	7,91	115,17
67	7,22	121,78	92	7,88	115,81
68	7,17	121,00	93	7,66	115,39
69	7,47	119,60	94	7,65	113,72
70	7,34	119,67	95	8,62	114,63
71	7,26	119,24	96	8,59	113,55
72	7,17	119,38	97	9,94	115,99
73	7,33	118,42	98	9,26	118,39
74	6,68	120,50	99	9,08	118,84

100	8,73	116,99	119	7,99	108,82
101	7,23	115,12	120	8,84	110,91
102	6,64	115,15	121	8,65	109,37
103	6,67	113,72	122	8,58	109,47
104	7,53	112,93	123	8,02	107,16
105	6,82	112,88	124	7,70	106,72
106	6,68	112,93	125	7,73	102,74
107	7,38	112,88	126	10,67	128,23
108	7,36	112,79	127	9,77	127,55
109	7,77	111,93	128	9,22	130,73
110	7,97	112,54	129	6,25	123,98
111	8,40	112,70	130	10,09	132,99
112	8,33	111,68	131	9,45	131,92
113	8,48	111,24	132	9,02	132,15
114	7,36	111,36	133	9,11	123,54
115	6,72	111,34	134	9,96	122,02
116	7,29	110,79	135	6,21	109,88
117	6,58	110,76	136	6,58	111,93
118	6,77	109,87			

Los espectros 2D $^1\text{H}/^{15}\text{N}$ HSQC RMN de la variante +G MYDGF (HEK), que resultó ser activa, véase Ebenhoch et al. (2019), y la variante +A MYDGF son prácticamente idénticos. Esto demuestra que +A MYDGF tiene la estructura terciaria correcta de MYDGF.

Ejemplo 6: Ensayo de potencia en células endoteliales de arterias coronarias humanas

Para determinar las potencias relativas de la variante [+A] en relación con la proteína HEK MYDGF +G como referencia, se realizó un ensayo de potencia en células endoteliales de arterias coronarias humanas (HCAEC). La variante +A se fabricó según el Ejemplo 2. Se examinaron tres lotes que se designaron V301, V302 y V303. La variante +G (SEQ ID NO: 3) se produjo en células HEK como se describe en Polten et al. (2019), página 1303, 1ª columna y Figura S1, y Ebenhoch et al. (2019), página 8 col. 1, y se utilizó como punto de referencia de actividad interna.

Tabla 18: Variantes examinadas en el ensayo HCAEC

N.º	Tipo de MYDGF (...)	Sistema huésped	Tipo de expresión
1 (V301)	+A...variante	E. coli	Insoluble
2 (V302)	+A...variante	E. coli	Insoluble
3 (V303)	+A...variante	E. coli	Insoluble
4 (referencia)	+G...variante(HEK)	HEK	Soluble

Las HCAEC se sembraron a una densidad de 55.000–60.000 células por pocillo en una placa de 24 pocillos en medio EGM–2 (Lonza) que contenía un 10 % de suero fetal de ternera (FCS) en un volumen total de 1 ml por pocillo. 24 horas después de la siembra (las células deben ser confluentes), el medio de cada pocillo se sustituyó por 1 ml de medio DCMB131 (Life Technologies) con un 2 % de FCS y se incubó durante 3–4 horas. Tras la incubación, se rasó la monocapa celular con una punta de pipeta (200 µl) en cada pocillo. La punta se utilizó verticalmente para garantizar que el arañazo fuera lo suficientemente grande. A continuación, las células se lavaron una vez con medio DCMB que contenía un 2 % de FCS. Posteriormente, se añadió 1 mL de medio fresco (DCMB con un 2 % de FCS) a cada pocillo. A continuación, se cultivaron las células en ausencia (control) o presencia de la proteína MYDGF a diferentes concentraciones, conteniendo cada pocillo una concentración inicial y una dilución en serie. Cada proteína se probó en las siguientes concentraciones: 13,3, 19,7, 29,6, 44,4, 66,6 y 100 ng/mL. El VEGFA humano (50 ng/mL) sirvió de control positivo. Los lotes de MYDGF V301, V302, V303 se ensayaron en duplicidad y frente a frente con la referencia. Inmediatamente después del tratamiento a

T=0 h, se tomó una imagen de todos los pocillos utilizando un microscopio (Zeiss Axio Observer Z1 con 50 aumentos, 5 aumentos objetivos, con ajuste de contraste de fase). Las imágenes se tomaron del centro de los pocillos, ya que allí se observaban los contrastes óptimos. A continuación, las placas se incubaron a 37 °C. Tras un tiempo de incubación de 16 horas (T=16h), se volvieron a tomar imágenes de cada pocillo como se ha descrito con anterioridad.

Para determinar la actividad, se calculó la recuperación en el ensayo midiendo el área libre de células utilizando el software axiovision o ImageJ en imágenes a las 0h y en imágenes a las 16h, respectivamente. La recuperación (%) se calculó como $[(\text{área libre de células a las 0h} - \text{área libre de células a las 16h}) / \text{área libre de células a las 0h}] \times 100$. Se resumieron dos experimentos para realizar ajustes de curvas logísticas de 4 parámetros (4-PL) y calcular los valores de EC₅₀ (software GraphPad Prism, versión 9.1.0). Los valores EC₅₀ de los diferentes lotes y de la referencia se aplicaron para calcular la potencia relativa en comparación con la referencia: $\text{Potencia [\%]} = \text{EC}_{50} \text{ de [+G]-HEK} / \text{EC}_{50} \text{ del lote de ensayo}) \times 100$.

Tabla 19: Valores EC₅₀ y de potencia de diferentes lotes de la variante [+A] determinados en el ensayo HCAEC.

Fig. 4	Muestra	EC₅₀ (ng/mL)	Potencia (%)
#1	+G HEK	46,2	100
	V301	26,1	177
#2	+G HEK	50,9	100
	V301	53,7	95
#3	+G HEK	47,9	100
	V301	35,2	136
Fig. 5	Muestra	EC₅₀ (ng/mL)	Potencia (%)
#1	+G HEK	46,2	100
	V302	40,6	114
#2	+G HEK	50,9	100
	V302	46,2	110
#3	+G HEK	47,9	100
	V302	46,3	103
Fig. 6	Muestra	EC₅₀ (ng/mL)	Potencia (%)
#1	+G HEK	58,1	100
	V303	58,5	99
#2	+G HEK	43,8	100
	V303	39,5	111
#3	+G HEK	33,9	100
	V303	41,9	81

Resultados: Como puede verse en la Tabla 18, los tres lotes de producción de la variante +A de MYDGF (V301, V302, V303) muestran actividades biológicas comparables a la referencia. En HCAEC, se determinaron las siguientes potencias relativas comparadas con la referencia (fijada al 100 %): V301 (177, 95 y 136 %), V302 (114, 110 y 103 %), V303 (99, 111 y 81 %). Los experimentos de migración celular agrupados dieron como resultado

unos valores de EC₅₀ de 37,4, 42,1 y 47,8 ng/mL para los lotes V301, V302 y V303, respectivamente. Al agrupar los datos de todos los lotes (V301 & V302, V303) y realizar un ajuste de curvas, se calculó un valor EC₅₀ de 41,1 ng/mL. El valor EC₅₀ para la referencia correspondiente calculado del mismo modo fue de 43,0 ng/mL.

Ejemplo 7: Ensayo de potencia en cardiomiocitos de rata neonatal

Se sembraron cardiomiocitos de rata neonatal (NRCM) en placas de 96 pocillos y se sometieron a isquemia/reperfusión (I/R) simulada en ausencia (control) o presencia de diferentes lotes de MYDGF. La I/R se simuló según lo descrito con anterioridad por Korf–Klingebiel et al., (2015). La proteína de referencia se ensayó en comparaciones frente a frente. Cada proteína se ensayó en las siguientes seis concentraciones: 13,3, 19,7, 29,6, 44,4, 66,6 y 100 ng/mL. El IGF–1 de ratón (50 ng/mL) sirvió como control positivo. La actividad de los metabolitos se evaluó mediante el ensayo MTS (Promega). Los lotes de MYDGF V301, V302 y V303 se probaron en 3–4 réplicas técnicas en 2–3 experimentos. Los experimentos se resumieron para realizar ajustes de curvas 4–PL y calcular los valores EC₅₀. Los valores EC₅₀ de los diferentes lotes y de la referencia se aplicaron para calcular la potencia relativa con respecto a la referencia (véase más arriba).

Tabla 20: Valores de potencia de diferentes lotes de la variante [+A] determinados en el ensayo neonatal de cardiomiocitos de rata

Fig. 7	Muestra	EC₅₀ (ng/mL)	Potencia
#1	+G HEK	24,0	100
	V301	22,1	108
#2	+G HEK	65,6	100
	V302	15,2	431
#3	+G HEK	18,6	100
	V303	21,9	85

Resultados: En cardiomiocitos de rata neonatal, se determinaron las siguientes potencias relativas en comparación con la referencia (fijada en el 100 %): V301 (108 %), V302 (431 %), V303 (85 %). Los valores de potencia de los lotes V301 y V303 de MYDGF indicaron una actividad biológica comparable a la de referencia. Aunque la actividad metabólica del lote V302 fue similar a la de referencia (gráficos de barras de la Figura 7), el ajuste no ideal de la curva dio como resultado una potencia 4 veces superior. Al agrupar los datos de todos los lotes (V301 & V302 & V303) y realizar un ajuste de curvas, se calculó un valor EC₅₀ de 22,2 ng/mL. El valor EC₅₀ para la referencia correspondiente calculado de la misma manera fue de 24,5 ng/mL.

Ejemplo 8: Ensayo de infarto de miocardio en ratón

Para comparar la eficacia de los tratamientos con MYDGF humano y murino en la curación del infarto de miocardio (IM) en ratones, se utilizó un modelo de infarto de miocardio en ratones. Los ratones FVB/N fueron sometidos a cirugía simulada o verum (isquemia/reperfusión), tratados con MYDGF recombinante humano o murino y sometidos a seguimiento durante 28 días. Las proteínas MYDGF humana y murina utilizadas en el ensayo se describen en la Tabla 20.

Tabla 21: Variantes examinadas en el ensayo de infarto de miocardio en ratón

N.º	Tipo de MYDGF (...)	Sistema huésped	Tipo de expresión
1	Mydgm-His variante murino	HEK	Soluble
2	+G...variante (HEK)	HEK	Soluble

Se fabricó la variante +G MYDGF (HEK) que tiene una secuencia según la SEQ ID NO: 3, esencialmente como se describe en Polten et al. (2019), y en la página 1303, 1ª columna y Figura S1, Ebenhoch et al. (2019), página 8 col. 1. La variante murina MYDGM-His se preparó según Korf-Klingebiel et al. (2021), Suppl. Material.

Los ratones FVB/N propensos a la insuficiencia cardíaca fueron sometidos a cirugía simulada (toracotomía sin isquemia/reperfusión; sin I/R) o I/R (isquemia/reperfusión). Los ratones fueron tratados con MYDGF humano o murino (10 µg en bolo + 10 µg/d en bomba durante 7 días, modelo 1007D) o solo diluyente (placebo). Se realizaron ecocardiografías seriadas (los días 6 y 28) y los ratones fueron sometidos a seguimiento durante 28 días. Al final del experimento, se recogieron los corazones y se determinó el tamaño de la cicatriz mediante la tinción tricrómica de Masson y la densidad capilar mediante la tinción fluorescente IB4/WGA. La significación estadística se evaluó mediante ANOVA unidireccional con la prueba post hoc de comparación múltiple de Dunnett. *P<0,05, **P<0,01, ***P<0,001 frente a simulacro (frente a todos los grupos de RV para FAC). ##P<0,01, P<0,001 frente a placebo.

Resultados: Se observó que la proteinoterapia con MYDGF humano (“MYDGF”) mejoraba la función cardíaca (cambio fraccional de área, FAC) respecto al tratamiento con placebo en un 17,5 % el día 6 y un 16,7 % el día 28. El MYDGF murino (“Mydgf”) aumentó la FAC con respecto al tratamiento con placebo en un 16,1 % el día 6 y un 19,0 % el día 28 (véase la figura 8). Como puede verse en la Figura 9, la cicatriz del infarto se redujo con ambas proteínas MYDGF recombinantes (16,5 % con MYDGF humano; 12,4 % con MYDGF murino frente a placebo). Además, el tratamiento con la proteína MYDGF aumentó la densidad capilar en la zona del borde del infarto tras el IM en un 21,8 % con respecto al tratamiento con placebo para la proteína humana y en un 19,1 % con respecto al tratamiento con placebo para la proteína murina (véase la figura 9). Así pues, ambos tratamientos mejoraron significativamente la curación del IM, evaluada por la mejora funcional cardíaca, la reducción del tamaño de la cicatriz y el aumento de la densidad capilar en la zona del borde del infarto.

BIBLIOGRAFÍA

1. Botnov, V. et al. (2018), J Biol Chem, 293(34), 13166–13175.
2. Ebenhoch, R. et al. (2019), Nat Commun 10, 5379.
3. Korf-Klingebiel, M. et al. (2015), Nat Med 21(2): 140–9.
4. Korf-Klingebiel, M. et al. (2021), Circulation.144(15): 1227–40.

5. Polten, F. et al. (2019), *Anal Chem*, 91, 1302–1308.
6. Tolonen, A.C. et al. (2011), *Mol Systems Biol* 7(1).
7. Zhao, L. et al. (2020), *J Cell Mol Med*, 24 (2): 1189–1199.
8. Frottin, F. et al. (2006), *The Proteomics of N-terminal Methionine Cleavage, Molecular & Cellular Proteomics*, Volume 5, Issue 12, 2336–2349.
9. Bodenhausen G. & Ruben D.J. (1980), *Chem. Phys. Lett.* 69, 185.
10. Piotto, M. et al. (1992), *J Biomol RMN*, 2, 661–666.
11. Sklenar, V. et al. (1993), *J Magn Reson, Series A* 102, 241–245.
12. Mori, S. et al. (1995), *J Magn Reson B* 108, 94–98.
13. Wishart, D. S. et al. (1995), *J Biomol RMN*, 5, 67–81.
14. Dyson, H. J. & Wright, P. E. (1998), *Nat Struct Biol*, 5, 499–503.
15. Shen Y. & Bax A. (2015), *Methods Mol Biol*, 1260:17–32.
16. CN 111544572 A
17. EP 2 918 676 B1
18. US 5,744,328 B
19. US 7,851,433 B
20. US 2015/0291683 A1
21. WO 2011/154349 A2
22. Peternel, S. & Komel, R. (2010), *Microbial Cell Factories*, 9:66.
23. Eggenreich, B. et al. (2020), *Journal of Biotechnology* 324S, 100022.
24. Brinson, R. G. et al. (2019), *mAbs*, 11, 94–105.
25. WO 2021/148411
26. WO 2014111458