(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 June 2024 (27.06.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/134502 A1

(51) International Patent Classification:
C12N 9/00 (2006.01)

(21) International Application Number:
PCT/IB2023/062949

(22) International Filing Date:
19 December 2023 (19.12.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
22215201.9 20 December 2022 (20.12.2022) EP

(71) Applicant: **NOVARTIS AG** [CH/CH]; Lichtstrasse 35, 4056 Basel (CH).

(72) Inventors: **MANN, Gregory**; Novartis Pharma AG, Postfach, 4002 Basel (CH). **STANGER, Frederic Valentin**; Novartis Pharma AG, Postfach, 4002 Basel (CH).

(74) Agent: **NOVARTIS AG**; Lichtstrasse 35, 4056 Basel (CH).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** ENGINEERED DOUBLE-STRAND RNA LIGASES AND USES THEREOF

(57) **Abstract:** The present disclosure relates to the field of biotechnology, in particular to engineered double-stranded RNA (dsRNA) ligases and their application in industrial biocatalysis. The present disclosure also relates to a process of producing an engineered dsRNA ligase, and to a method for producing an oligonucleotide by contacting oligonucleotide fragments with an engineered dsRNA ligase.

WO 2024/134502 A1

ENGINEERED DOUBLE-STRAND RNA LIGASES AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to, and the benefit of, EP Application No. 22215201.9, filed on December 20, 2022, the content of which is incorporated herein by reference in its entirety.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in .XML format and is hereby incorporated by reference in its entirety. Said .XML copy, created on 2 December 2023, is named PAT059445-WO-PCT_SL.xml and is 1.24MB in size.

Technical field

The present disclosure relates to the field of biotechnology, in particular to engineered double-stranded RNA (dsRNA) ligases and their application in industrial biocatalysis. The present disclosure also relates to a process of producing an engineered dsRNA ligase, and to a method for producing an oligonucleotide by contacting oligonucleotide fragments with an engineered dsRNA ligase.

Background art

Therapeutic oligonucleotides, including small interfering RNA (siRNA) and inhibitory antisense oligonucleotides (ASOs) have the potential to treat a diverse range of life-threatening diseases. In recent years there has been a significant increase in the number of approved oligonucleotide-based drugs, and a large rise in the number of therapeutic oligonucleotides under clinical investigation (Roberts, T. C., Langer, R. & Wood, M. J. A. *Nature Reviews Drug Discovery 2020 19:10* **19**, 673–694 (2020)).

In support of green synthesis initiatives throughout the pharmaceutical industry, there is a significant need for next-generation oligonucleotide synthesis methods that are both sustainable and economical at the scale required to reach wider patient populations (Mishra, M. *et al. Current Research in Green and Sustainable Chemistry* **4**, (2021)).

To this end, biocatalysis is being more frequently applied in the manufacture of active pharmaceutical ingredients (APIs) since enzymes are capable of highly selective transformation under mild reaction conditions and in aqueous media (Mann, G. & Stanger, F. V. *Chimia (Aarau)* **74**, 407–417 (2020)). The biocatalysis of short oligonucleotide fragments

offers a sustainable and economical alternative to the solid phase chemical synthesis of full-length therapeutic oligonucleotides currently used.

Shorter oligonucleotides can be synthesized more easily and with higher purities than longer oligonucleotides, simplifying downstream processing and reducing solvent waste. These short oligonucleotide fragments can then be combined using nucleic acid ligases to produce oligonucleotide products. Nucleic acid ligases have shown remarkable tolerance towards unnatural DNA/RNA containing pharmaceutically relevant chemical modifications (Kestemont, D., Herdewijn, P. & Renders, M. *Curr Protoc Chem Biol* **11**, e62 (2019); Kestemont, D. *et al. Chemical Communications* **54**, 6408–6411 (2018); and Nandakumar, J. & Shuman, S. *Molecular Cell* **16**, 211–221 (2004)), and the use of a dsRNA ligase to synthesize an siRNA product, starting from short fragments (≤ 9 nts), containing extensive chemical modification, including 2'-OMe, 2'-F modified nucleotides, phosphorothioate backbone modified nucleotides and a terminal fragment that is functionalized with a bulky N-acetyl galactosamine (GalNAc) moiety has previously been described (Mann, G. *et al. Tetrahedron Letters* **93**, 153696 (2022)).

To achieve cost-effective and sustainable industrial scale biocatalysis of oligonucleotides, enzymes exhibiting high ligase activity are required. There exists an urgent and unmet need for engineered ligase enzymes, which exhibit improved ligase activity relative to wild-type enzymes. There is also an unmet need for biocatalytic methods of producing oligonucleotides from oligonucleotide fragments.

Brief description of the drawings

Figure 1. dsRNA ligase catalyzed ligation of: (A) oligonucleotide fragments 6:9; 7:10 and 11:12 to generate oligonucleotide 2:3 (= siRNA 1); and (B) oligonucleotide fragments 6:9, 7:10 and 8:11 to generate oligonucleotide 5:3 (= siRNA 4). The sequences of oligonucleotides 2, 3 and 5-12 are provided in Table 1.

Figure 2. Comparative data showing the relative peak area % of siRNA (1) present in the reaction samples comprising different concentrations of the wild-type enzyme (SEQ ID NO: 2) and engineered enzymes (SEQ ID NOs: 288, 290 and 292) assayed under: (A) condition 1; and (B) condition 2, as described in Example 13. Enzyme concentration is provided as g/L of shake-flask powder (SFP) produced by lyophilization of frozen clarified lysate as described in the Examples.

Figure 3. (A) Comparative data showing the relative peak area % of siRNA (1) present in the reaction samples comprising different concentrations of wild-type enzyme (SEQ ID NO: 2)

and engineered enzymes (SEQ ID NOs: 288 and 632) following pre-incubation of the enzyme at 4 °C or 37 °C for 4 h. (B) Comparative data showing residual enzyme activity following pre-incubation of the SFP at 37 °C for 4 h, expressed relative to the ligation activity of the SFP pre-incubated at 4 °C for 4 h. Enzyme concentration is provided as g/L of shake-flask powder (SFP) produced by lyophilization of frozen clarified lysate as described in the Examples.

Summary of the disclosure

The present disclosure provides engineered double-stranded RNA (dsRNA) ligase polypeptides. The present disclosure also provides gene sequences of engineered polypeptides, recombinant expression vectors comprising the genes, engineered host strains and efficient methods for the production thereof, as well as reaction processes for the biocatalysis of oligonucleotides using engineered polypeptides.

The engineered double-stranded RNA (dsRNA) ligase polypeptides described herein have improved catalytic activity as compared to the wild-type dsRNA ligase from which they are derived. Through substitutions and/or deletions of amino acid residues in directed evolution processes, the engineered polypeptides provided herein were derived from a wild-type dsRNA ligase from Bacteriophage RB69. The wild-type dsRNA ligase consists of 332 amino acids and has the amino acid sequence shown in SEQ ID NO: 302 (also accessible under accession number Q7Y4V8 in UniProt).

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, and 600; wherein the engineered dsRNA ligase polypeptide: (a) has dsRNA ligase activity; and (b) does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668; wherein the engineered dsRNA ligase polypeptide: (a) has dsRNA ligase activity; and (b) does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668; wherein the engineered dsRNA ligase polypeptide: (a) has dsRNA ligase activity; and (b) does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 85% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668; wherein the engineered dsRNA ligase polypeptide: (a) has dsRNA ligase activity; and (b) does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide, which is a polypeptide of: (a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, and 600; or (b) a polypeptide having dsRNA ligase activity, which comprises an amino acid sequence having (i) at least 80% sequence identity to one of the polypeptides recited in (a), and (ii) a substitution, deletion, addition or insertion of one or more amino acid residues relative to said one amino acid sequence recited in (a); wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide, which is a polypeptide of: (a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668; or (b) a polypeptide having dsRNA ligase activity, which comprises an amino acid sequence having (i) at least 80% sequence identity to one of the polypeptides recited in (a), and (ii) a substitution, deletion, addition or insertion of one or more amino acid residues relative to said one amino acid sequence recited in (a); wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide, which is a polypeptide of: (a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534,

536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668; or (b) a polypeptide having dsRNA ligase activity, which comprises an amino acid sequence having (i) at least 80% sequence identity to one of the polypeptides recited in (a), and (ii) a substitution, deletion, addition or insertion of one or more amino acid residues relative to said one amino acid sequence recited in (a); wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668; wherein: (a) the engineered dsRNA ligase polypeptide has dsRNA ligase activity; and (b) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X6 is G or E ; X7 is Q; X15 is R, D or E; X19 is Q or D; X29 is N or L; X36 is V; X39 is A; X44 is V; X45 is V; X46 is Y; X47 is E; X49 is G; X51 is L; X53 is Y; X56 is R or A; X57 is S; X60 is T, G or P; X63 is S, Q or G; X64 is R, T, Q, F, G, or M; X66 is F or W; X67 is N; X87 is T, P, K or absent; X88 is C; X89 is T; X91 is S; X92 is D; X93 is G, C, or A; X103 is V, C, Y, or T; X105 is V; X107 is R or T; X114 is N; X122 is W; X126 is G; X129 is N; X130 is R, S or Y; X131 is R; X137 is V or C; X144 is N; X146 is R; X158 is W; X163 is G; X173 is L; X178 is R; X185 is K; X190 is Q; X196 is S or C; X216 is L or R; X221 is I; X228 is R; X230 is T; X232 is R; X235 is A, T, or G; X236 is S, L, or F; X237 is S, Q, R, L or G; X238 is F; X239 is G or R; X242 is R or M; X243 is N, S, G, or M; X244 is G or K; X251 is D or L; X252 is V; X254 is K; X255 is C; X258 is V; X269 is L; X280 is W; X284 is A; X285 is A; X293 is R; X296 is R; X301 is G, L, E, or F;

X303 is Q; X305 is G; X313 is A; X314 is A or V; X325 is R; and X328 is R; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 370, 488, 526, 578, 588, 590, and 592. In some embodiments, the polypeptide comprises an amino acid sequence of SEQ ID NO: 666. In some embodiments, the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 370, 488, 526, 578, 588, 590, 592, and 666.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 370, 488, 526, 578, 588, 590, 592, and 666; wherein: (a) the engineered dsRNA ligase polypeptide has dsRNA ligase activity; and (b) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X15 is D or E; X19 is D; X36 is V; X39 is A; X53 is Y; X185 is K; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 666; wherein: (a) the engineered dsRNA ligase polypeptide has dsRNA ligase activity; and (b) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or all 10) of the following amino acid residues: X15 is D; X39 is A; X53 is Y; X185 is K; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 370; wherein: (a) the engineered dsRNA ligase polypeptide has dsRNA ligase activity; and (b) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, or all 4) of the following amino acid residues: X36 is V; X39 is A; X218 is N; and X221 is I.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 488; wherein: (a) the engineered dsRNA ligase polypeptide has dsRNA ligase activity; and (b) the amino acid sequence of the engineered

dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, or all 3) of the following amino acid residues: X39 is A; X218 is N; and X221 is I.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 526; wherein: (a) the engineered dsRNA ligase polypeptide has dsRNA ligase activity; and (b) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, or all 4) of the following amino acid residues: X39 is A; X218 is N; X221 is I; and X255 is C.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 578; wherein: (a) the engineered dsRNA ligase polypeptide has dsRNA ligase activity; and (b) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, or all 8) of the following amino acid residues: X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 588 or 590; wherein: (a) the engineered dsRNA ligase polypeptide has dsRNA ligase activity; and (b) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, or all 9) of the following amino acid residues: X15 is D or E; X39 is A; X53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO: 592; wherein: (a) the engineered dsRNA ligase polypeptide has dsRNA ligase activity; and (b) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, or all 9) of the following amino acid residues: X19 is D; X39 is A; X53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

In some embodiments, the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666 and 668. In some embodiments, the polypeptide comprises an

amino acid sequence selected from the group consisting of SEQ ID NOs: 636, 638, 642, 646, 664, and 666.

The disclosure also provides an engineered dsRNA ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 302, which produces at least 5% more oligonucleotide product than a dsRNA ligase polypeptide comprising the amino acid sequence of SEQ ID NO: 302 under the same ligation reaction conditions, wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

In some embodiments, the ligation reaction conditions include about 1 μM to about 10 mM oligonucleotide fragment, a source of ATP, about 5 mM to about 100 mM divalent cation, and about 0.5 g/L to about 10 g/L engineered dsRNA ligase polypeptide, pH of about 4.0 to about 8.0, and temperature of about 10 °C to about 50°C. In some embodiments, the source of ATP comprises ATP, optionally a stoichiometric excess of ATP. In some embodiments, the source of ATP comprises: (a) polyphosphate kinase (PPK); (b) polyphosphate; and (c) AMP and/or ATP.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X6, X7, X15, X19, X29, X36, X39, X46, X47, X49, X51, X53, X56, X57, X60, X63, X64, X66, X67, X87, X88, X91, X93, X103, X105, X107, X114, X122, X126, X129, X130, X131, X137, X144, X146, X158, X163, X173, X178, X190, X196, X216, X218, X221, X228, X230, X232, X235, X236, X237, X238, X239, X242, X243, X244, X251, X252, X254, X255, X258, X269, X280, X284, X285, X293, X296, X301, X303, X305, X314, X325, and X328, wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X6, X7, X15, X19, X29, X36, X39, X44, X45, X46, X47, X49, X51, X53, X56, X57, X60, X63, X64, X66, X67, X87, X88, X89, X91, X92, X93, X103, X105, X107, X114, X122, X126, X129, X130, X131, X137, X144, X146, X158, X163, X173, X178, X185, X190, X196, X216, X218, X221, X228, X230, X232, X235, X236, X237, X238, X239, X242, X243, X244, X251, X252, X254, X255, X258, X269, X280, X284, X285, X293, X296, X301, X303, X305, X313, X314, X325, and X328, wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X6 is G; X7 is Q; X15 is R, D or E; X19 is Q or D; X29 is N or L; X36 is V; X39 is A; X46 is Y; X47 is E; X49 is G; X51 is L; X53 is Y; X56 is R or A; X57 is S; X60 is T, G or P; X63 is S, Q or G; X64 is R, T, Q, F, G, or M; X66 is F or W; X67 is N; X87 is T, P, K or absent; X88 is C; X91 is S; X93 is G, C, or A; X103 is V, C, Y, or T; X105 is V; X107 is R or T; X114 is N; X122 is W; X126 is G; X129 is N; X130 is R, S or Y; X131 is R; X137 is V or C; X144 is N; X146 is R; X158 is W; X163 is G; X173 is L; X178 is R; X190 is Q; X196 is S or C; X216 is L or R; X218 is N; X221 is I; X228 is R; X230 is T; X232 is R; X235 is A, T, or G; X236 is S, L, or F; X237 is S, Q, or R; X238 is F; X239 is G or R; X242 is R or M; X243 is N, S, G, or M; X244 is G or K; X251 is D or L; X252 is V; X254 is K; X255 is C; X258 is V; X269 is L; X280 is W; X284 is A; X285 is A; X293 is R; X296 is R; X301 is G, L, E, or F; X303 is Q; X305 is G; X314 is A or V; X325 is R; and X328 is R; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X6 is G or E; X7 is Q; X15 is R, D or E; X19 is Q or D; X29 is N or L; X36 is V; X39 is A; X44 is V; X45 is V; X46 is Y; X47 is E; X49 is G; X51 is L; X53 is Y; X56 is R or A; X57 is S; X60 is T, G or P; X63 is S, Q or G; X64 is R, T, Q, F, G, or M; X66 is F or W; X67 is N; X87 is T, P, K or absent; X88 is C; X89 is T; X91 is S; X92 is D; X93 is G, C, or A; X103 is V, C, Y, or T; X105 is V; X107 is R or T; X114 is N; X122 is W; X126 is G; X129 is N; X130 is R, S or Y; X131 is R; X137 is V or C; X144 is N; X146 is R; X158 is W; X163 is G; X173 is L; X178 is R; X185 is K; X190 is Q; X196 is S or C; X216 is L or R; X218 is N; X221 is I; X228 is R; X230 is T; X232 is R; X235 is A, T, or G; X236 is S, L, or F; X237 is S, Q, R, L or G; X238 is F; X239 is G or R; X242 is R or M; X243 is N, S, G, or M; X244 is G or K; X251 is D or L; X252 is V; X254 is K; X255 is C; X258 is V; X269 is L; X280 is W; X284 is A; X285 is A; X293 is R; X296 is R; X301 is G, L, E, or F; X303 is Q; X305 is G; X313 is A; X314 is A or V; X325 is R; and X328 is R; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X15, X19, X36, X39, X53, X218, X221, X237, X251, X255, and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X15 is D or E; X19 is D; X36 is V; X39 is A; X53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X15, X19, X36, X39, X53, X185, X218, X221, X237, X251, X255, and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X15 is D or E; X19 is D; X36 is V; X39 is A; X53 is Y; X185 is K; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* 2 or more, 3 or more, or all 4) amino acid residues selected from: X36, X39, X218 and X221, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, or all 4) of the following amino acid residues: X36 is V; X39 is A; X218 is N; and X221 is I.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* 2 or more, or all 3) amino acid residues selected from: X39, X218 and X221, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, or all 3) of the following amino acid residues: X39 is A; X218 is N; and X221 is I.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* 2 or more, 3 or more, or all 4) amino acid residues selected from: X39, X218, X221 and X255, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more

(*e.g.* 2 or more, 3 or more, or all 4) of the following amino acid residues: X39 is A; X218 is N; X221 is I; and X255 is C.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, or all 8) amino acid residues selected from: X39, X53, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, or all 8) of the following amino acid residues: X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, or all 9) amino acid residues selected from: X15, X39, X53, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, or all 9) of the following amino acid residues: X15 is E; X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, or all 9) amino acid residues selected from: X19, X39, X53, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, or all 9) of the following amino acid residues: X19 is D; X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8

or more, 9 or more, or all 10) amino acid residues selected from: X15, X39, X53, X185, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or all 10) of the following amino acid residues: X15 is D; X39 is A; X 53 is Y; X185 is K; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

In some embodiments, the engineered dsRNA ligase polypeptide comprises a purification tag. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298 and 300. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, and 634. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, and 634.

The disclosure also provides a polypeptide immobilized on a solid material by chemical bond or a physical adsorption method, wherein the polypeptide comprises an engineered dsRNA ligase polypeptide described herein.

The disclosure also provides a polynucleotide encoding the engineered dsRNA ligase polypeptide described herein.

In some embodiments, the polynucleotide comprises a nucleic acid sequence selected from SEQ ID NOs: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, and 599.

In some embodiments, the polynucleotide comprises a nucleic acid sequence selected from SEQ ID NOs: 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, and 667.

In some embodiments, the polynucleotide comprises a nucleic acid sequence selected from: (a) SEQ ID NOs: 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, and 667; and/or (b) SEQ ID NOs: 3, 5, 7, 9, 11, 13, 15, 17,

19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, and 633.

The disclosure also provides an expression vector comprising the polynucleotide described herein. In some embodiments, the vector comprises a plasmid, a cosmid, a bacteriophage or a viral vector.

The disclosure also provides a host cell comprising the polynucleotide described herein or the expression vector described herein. In some embodiments, the host cell is *E. coli*.

The disclosure also provides a method of preparing an engineered dsRNA ligase polypeptide, which comprises the steps of culturing the host cell described herein and obtaining an engineered dsRNA ligase polypeptide from the culture.

The disclosure also provides an engineered dsRNA ligase catalyst obtainable by culturing the host cells described herein, or according to the method described herein, wherein said engineered dsRNA ligase catalyst comprises cells or culture fluid containing the engineered dsRNA ligase polypeptides, or an article processed therewith, wherein the article refers to an extract obtained from the culture of host cell, an isolated product obtained by isolating or purifying an engineered dsRNA ligase from the extract, or an immobilized product obtained by immobilizing host cell, an extract thereof, or isolated product of the extract.

The disclosure further provides a method of producing an oligonucleotide from two or more oligonucleotide fragments, wherein the method comprises contacting: (i) two or more oligonucleotide fragments; (ii) an engineered dsRNA ligase polypeptide described herein; (iii) a source of ATP; and (iv) a divalent cation; to obtain an oligonucleotide.

In some embodiments, the source of ATP comprises ATP.

In some embodiments, the source of ATP comprises: (a) polyphosphate kinase (PPK); (b) polyphosphate; and (c) AMP and/or ATP. In some embodiments, the PPK is selected from PPK12 or ajPAP.

In some embodiments, the method is performed using a sub-stoichiometric concentration of AMP and/or ATP.

In some embodiments, the polyphosphate is a polyphosphate salt. In some embodiments, the polyphosphate salt is sodium polyphosphate (Maddrell's salt) or sodium hexametaphosphate (Graham's salt).

In some embodiments, the divalent cation cofactor is Mg^{2+} or Mn^{2+}.

In some embodiments, the method is performed with a divalent cation concentration of 5-100 mM, optionally 30-50 mM.

In some embodiments, the method further comprises a step of purifying the oligonucleotide.

The disclosure also provides use of the engineered dsRNA ligase polypeptide described herein in the production of an oligonucleotide from two or more oligonucleotide fragments.

In some embodiments, the oligonucleotide is up to 60 nucleotides in length.

In some embodiments, each of the oligonucleotide fragments are 4-16 nucleotides in length, optionally 6-9 nucleotides in length.

In some embodiments, one or more of the oligonucleotide fragment(s) comprises one or two overhangs.

In some embodiments, one or more of the oligonucleotide fragments comprises a chemical modification. In some embodiments, the chemical modification is selected from: (a) a modified backbone, optionally selected from a phosphorothioate (*e.g.* chiral phosphorothioate) or methylphosphonate internucleotide linkage; (b) a modified nucleotide, optionally selected from 2'-O-methyl (2'-OMe), 2'-flouro (2'-F), 2'-deoxy, 2'-deoxy-2'-fluoro, 2'-O-methoxyethyl (2'-O-MOE), 2'-O-aminopropyl (2'-O-AP), 2'-O-dimethylaminoethyl (2'-O-DMAOE), 2'-O-dimethylaminopropyl (2'-O-DMAP), 2'-O-dimethylaminoethyloxyethyl (2'-O-DMAEOE), 2'-O-N-methylacetamido (2'-O-NMA), locked nucleic acid (LNA), glycol nucleic acid (GNA), phosphoramidate (*e.g.* mesyl phosphoramidate), 2',3'-seco nucleotide mimic, 2'-F-arabino nucleotide, abasic nucleotide, 2'-amino modified nucleotide, 2'-alkyl-modified nucleotide, morpholino nucleotide, vinylphosphonate (*e.g.* 5' vinylphosphonate), and cyclopropyl phosphonate deoxyribonucleotide; and/or (c) conjugation to a ligand, optionally wherein the ligand comprises one or more N-Acetylgalactosamine (GalNAc) derivatives.

The disclosure also provides a composition comprising: i. the engineered dsRNA ligase polypeptide described herein; ii. a source of ATP; and iii. a divalent cation.

In some embodiments, the composition further comprises two or more oligonucleotide fragments.

The disclosure also provides a kit comprising: i. the engineered dsRNA ligase polypeptide described herein; ii. a source of ATP; iii. a divalent cation; and iv. instructions for use in a method of producing an oligonucleotide from two or more oligonucleotide fragments.

In some embodiments, the source of ATP comprises ATP.

In some embodiments, the source of ATP comprises: (a) polyphosphate kinase (PPK); (b) polyphosphate; and (c) AMP and/or ATP.

In some embodiments, the PPK is selected from PPK12 or ajPAP.

In some embodiments, the polyphosphate is a polyphosphate salt.

In some embodiments, the polyphosphate salt is sodium polyphosphate (Maddrell's salt) or sodium hexametaphosphate (Graham's salt).

In some embodiments, the divalent cation cofactor is Mg^{2+} or Mn^{2+}.

Definitions

Unless expressly defined otherwise, technical and scientific terms used in this disclosure have the meanings that are commonly understood by people skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton *et al.*, Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them below, unless specified otherwise.

As used throughout this disclosure, articles such as "a" and "an" refer to one or more than one (at least one) of the grammatical object of the article.

The term "and/or" means either "and" or "or" unless indicated otherwise.

As used herein, the term "about" typically refers to the value which immediately follows the term 'about'. For example, "about 15 or more nucleotides" typically refers to 15 or more nucleotides. In some embodiments, the term "about" embraces values which are +/- 1, 2 or 3 of the stated value. For example, "about 15 or more nucleotides" may refer to 15+/-3 nucleotides, *e.g.* 12, 13, 14, 15, 16, 17 or 18 nucleotides.

The terms “double-stranded RNA ligase” and “dsRNA ligase” are used interchangeably herein to refer to an enzyme having dsRNA ligase activity. A dsRNA ligase polypeptide may also be referred to herein as a “dsRNA ligase catalyst”.

A dsRNA ligase of the invention is an ATP-dependent nucleic acid ligase. dsRNA ligase activity as used herein typically involves the ATP-dependent formation of a covalent bond between the 3’-OH of a ribonucleotide and the 5’-PO_4 of a ribonucleotide or deoxyribonucleotide via the following steps: (1) dsRNA ligase reacts with ATP to form a covalent dsRNA ligase-AMP intermediate and release pyrophosphate; (2) AMP is transferred from the dsRNA ligase-AMP intermediate to the 5’-phosphate of a 3’ oligonucleotide fragment forming an adenylated oligonucleotide intermediate; and (3) the 3’-OH of a 5’ oligonucleotide fragment attacks the 5’ phosphate of the adenylated intermediate resulting in the formation of a phosphodiester bond and the release of AMP.

The stoichiometric concentration of cofactor is the theoretical concentration required to achieve complete ligation in a given ligation reaction. The skilled person can readily derive the stoichiometric concentration of ATP required to achieve complete ligation based on the concentration of oligonucleotide fragments and the number of ligation reactions required to produce the oligonucleotide product. For example, a ligation reaction using 1 mM substrate which requires four ligation reactions has a stoichiometric ATP concentration of 4 mM. A stoichiometric excess of ATP can help ensure that complete ligation is achieved. In some embodiments, the stoichiometric excess is at least 105% of the theoretical stoichiometric concentration of ATP required to achieve complete ligation, *e.g.* at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 160%, at least 170%, at least 180%, at least 190%, or at least 200%.

The terms "engineered dsRNA ligase", "engineered dsRNA ligase polypeptide", "improved dsRNA ligase polypeptide", and "engineered polypeptide" are used interchangeably herein.

As used herein, the term “oligonucleotide” refers to a nucleic acid, typically comprising up to 100 nucleotides. As used herein, the term “oligonucleotide product” refers to an oligonucleotide formed by the ligation of two or more oligonucleotide fragments by a dsDNA ligase described herein. Oligonucleotide products are also referred to herein simply as oligonucleotides. It will be understood that oligonucleotide products described herein comprise RNA. It will also be understood that oligonucleotide products described herein comprise a double-stranded region. In some embodiments, oligonucleotide products described herein comprise RNA and DNA. For example, a portion of the oligonucleotide

product may be double-stranded DNA, while another portion is double-stranded RNA, forming a DNA-RNA chimera.

The term "therapeutic oligonucleotide" refers to an oligonucleotide that can provide a therapeutic effect, *e.g.* by interacting with a biomolecule and/or by regulating gene expression. Therapeutic oligonucleotides include, but are not limited to, RNA interference (RNAi) agents and antisense oligonucleotides (ASO). RNAi is a post-transcriptional, targeted gene-silencing technique that uses RNAi agents to degrade messenger RNA (mRNA) containing the same sequence as the RNAi agent. ASOs are single-stranded nucleic acids that can be used to target mRNA derived from a gene of interest. ASOs can alter gene expression via a number of mechanisms including direct steric blockage of mRNA and ribonuclease H (RNase H) mediated degradation of mRNA.

RNAi agents include, as non-limiting examples, siRNAs (small interfering RNAs), dsRNAs (double-stranded RNAs), shRNAs (short hairpin RNAs) and miRNAs (micro RNAs). RNAi agents also include, as additional non-limiting examples, locked nucleic acid (LNA), Morpholino, UNA, threose nucleic acid (TNA), glycol nucleic acid (GNA), peptide nucleic acid (PNA) and fluoro-arabinonucleic acid (FANA). RNAi agents also include molecules in which one or more strands are a mixture of RNA, DNA, LNA, Morpholino, UNA (unlocked nucleic acid), TNA, GNA, and/or FANA. As a non-limiting example, one or both strands of an RNAi agent could be, for example, RNA, except that one or more RNA nucleotides is replaced by DNA, LNA, Morpholino, UNA, TNA, GNA, and/or FANA, etc. In some embodiments, one or both strands of the RNAi agent can be nicked, and both strands can be the same length, or one strand can be shorter than the other. The oligonucleotide of the invention may be any of the RNAi agents described herein.

The term "oligonucleotide fragment" herein refers to a nucleic acid that can be ligated to one or more additional oligonucleotide fragments to provide an oligonucleotide (or oligonucleotide product). Each oligonucleotide fragment corresponds to a portion of the oligonucleotide product. Oligonucleotide fragments may be referred to herein as "substrates" of the ligation reaction.

As described above, dsRNA ligase activity involves the ligation of a 5' oligonucleotide fragment to a 3' oligonucleotide fragment. In the context of oligonucleotide fragments, the prefixes 5' and 3' refer to the relative position of each oligonucleotide fragment in the oligonucleotide product after ligation, wherein the 5' oligonucleotide fragment is located upstream of the 3' oligonucleotide fragment (when the oligonucleotide product is presented in the 5' to 3' direction). As used herein, a "5' oligonucleotide fragment" typically comprises a

3' terminal ribonucleotide having a 3'-hydroxyl group. As used herein, a "3' oligonucleotide fragment" comprises a 5'-phosphate, wherein the 5' terminal nucleotide is a deoxyribonucleotide or a ribonucleotide.

It will be understood that, in some embodiments, an oligonucleotide fragment may be a 3' oligonucleotide fragment and a 5' oligonucleotide fragment (*e.g.* wherein ligation reactions occur at the 5' and 3' ends of the oligonucleotide fragment). For example, said oligonucleotide fragment may provide: (i) the 3' oligonucleotide fragment in a ligation reaction with a 5' oligonucleotide fragment; and (ii) the 5' oligonucleotide fragment in a ligation reaction with a 3' oligonucleotide fragment. For example, oligonucleotide fragment 7 in Figure 1A provides: (i) the 3' oligonucleotide fragment in a ligation reaction with 5' oligonucleotide fragment 6 and; (ii) the 5' oligonucleotide fragment in a ligation reaction with 3' oligonucleotide fragment 12 to provide oligonucleotide product 2.

A "terminal oligonucleotide fragment" herein refers to a nucleic acid that corresponds to an end (*e.g.* 5' or 3' end) portion of the oligonucleotide product. The 5' terminal oligonucleotide fragment typically provides a 5' oligonucleotide fragment for ligation to a 3' oligonucleotide fragment. The 3' terminal oligonucleotide fragment typically provides a 3' oligonucleotide fragment for ligation to a 5' oligonucleotide fragment. In some embodiments, the 5' terminal oligonucleotide is ligated directly to the 3' terminal oligonucleotide. In some embodiments, the 5' terminal oligonucleotide and the 3' terminal oligonucleotide are separated by one or more oligonucleotide fragments.

In some embodiments, oligonucleotide fragments described herein comprise RNA and DNA. For example, a portion of an oligonucleotide fragment may be double-stranded DNA, while another portion is double-stranded RNA, forming a DNA-RNA chimera.

The term "overhang" or "nucleotide overhang" herein refers to at least one unpaired nucleotide that protrudes from the end of at least one of the two strands of a double-stranded oligonucleotide. In some embodiments, when a 3'-end of one strand extends beyond the 5'-end of the other strand, or vice versa, this forms a nucleotide overhang, *e.g.,* the unpaired nucleotide(s) form the overhang. An overhang that is complementary to the overhang of a second oligonucleotide fragment may be referred to as a "sticky end". The oligonucleotide fragments described herein may have one or two sticky ends.

"Blunt" or "blunt end" means that there are no unpaired nucleotides at that end of a double-stranded oligonucleotide, *i.e.*, no nucleotide overhang. A "blunt ended" oligonucleotide or oligonucleotide fragment is an oligonucleotide that is double-stranded over its entire length, *i.e.*, no nucleotide overhang at either end of the molecule.

Double-stranded nucleic acids comprise two anti-parallel and substantially complementary nucleic acid strands which are referred to as “sense” and “antisense” strands. In the context of double-stranded RNAi agents, the “antisense strand” refers to the strand of an RNAi which includes a region that is substantially complementary to a target sequence, *e.g.* an mRNA sequence. The “sense strand” refers to the strand of an RNAi that includes a region that is substantially complementary to a region of the antisense strand. The sense and antisense strands of an RNAi agent may be referred to as the passenger and guide strands, respectively.

Sequences that are “substantially complementary” may be fully complementary or may contain one or more mismatches upon hybridization, while retaining the ability to hybridize under the conditions most relevant to their ultimate application.

"Conversion" refers to the enzymatic transformation of a substrate to the corresponding product. "Percent conversion" or "conversion" refers to the percentage of oligonucleotide fragments that is converted to oligonucleotide product within a defined period of time under specified conditions. Thus, "enzymatic activity" or "activity" of a ligase can be expressed as the "percent conversion" of oligonucleotide fragments to oligonucleotide product.

Ideally to compare the activity between ligation reactions and account for natural variation in peak intensity between injections, the % conversion to product would be calculated for each sample analyzed using the following equation:

$$\%\ Conv = \frac{\left(\frac{A260\ Product(s)}{\varepsilon_p}\right)}{\left(\frac{A260\ Product(s)}{\varepsilon_p}\right) + \left(\frac{A260\ Substrate(s)}{\varepsilon_s}\right) + \left(\frac{A260\ Intermediate(s)}{\varepsilon_i}\right)} \times 100$$

Whereby ε_p, ε_s and ε_i = the extinction coefficient of the product, substrate, and intermediate oligonucleotides respectively. In some instances, such as using the analytical method described herein, it is not possible to resolve all substrates, reaction intermediates and products. Therefore, the % conversion according to the above equation cannot be determined. However, in some instances, such as using the analytical method described herein, it is possible to resolve at least one substrate, reaction intermediate and product, such as well-defined GalNAc-containing oligonucleotides, including GalNAc containing substrate fragments (*e.g.* oligonucleotide (12) as used in the examples described herein), reaction intermediates (*e.g.* oligonucleotide (14) as demonstrated herein) and product strands (*e.g.* product oligonucleotide (2) as demonstrated herein). Therefore, a pseudo-% conversion can be calculated, denoted with arbitrary units (AU), which considers only these well resolved species according to the following equation:

$$AU = \frac{\left(\frac{A260\ (2)}{\varepsilon_{(2)}}\right)}{\left(\frac{A260\ (2)}{\varepsilon_{(2)}}\right) + \left(\frac{A260\ (12)}{\varepsilon_{(12)}}\right) + \left(\frac{A260\ (14)}{\varepsilon_{(14)}}\right)}$$

whereby $\varepsilon_{(2)}$, $\varepsilon_{(12)}$ and $\varepsilon_{(14)}$ are the extinction coefficient of oligonucleotides (2), (12) and (14) respectively. Using such a calculation an AU = 1.0 would imply that no more GalNAc-containing substrate or intermediate oligonucleotides are present in the reaction and that they have all be converted to GalNAc-containing product (2). In reality, for samples where AU = 1.0 the only other peak present in the chromatogram corresponds with the product oligonucleotide (2), and no other intermediates or starting materials can be identified. Furthermore, the ratio of the product oligonucleotides (2) and (3) are consistent with that of the authentic standard of siRNA product (1). Taken together, it can be concluded that AU = 1.0 is an approximation that is essentially equivalent to 100 % conversion.

"Improved enzyme properties" refers to an enzyme property that is better or more desirable for a specific purpose as compared to a reference dsRNA ligase such as a wild-type dsRNA ligase or another engineered dsRNA ligase under the same reaction conditions. Improved enzyme properties are exhibited by engineered dsRNA ligase polypeptides in this disclosure. The engineered dsRNA ligase polypeptides described herein exhibit increased enzyme activity (which can be expressed as a percentage of substrate conversion). Additional enzyme properties that may be improved include, but are not limited to, thermal stability, pH activity characteristics, cofactor requirements, and tolerance to inhibitors (*e.g.*, reaction component, substrate or product inhibition).

An "isolated polypeptide" refers to a polypeptide that is substantially separated from other substances with which it is naturally associated, such as proteins, lipids, and polynucleotides. The term comprises polypeptides that have been removed or purified from their naturally occurring environment or expression system (*e.g.*, in host cells or *in vitro* synthesis). Engineered dsRNA ligase polypeptides may be present in the cell, in the cell culture medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the engineered dsRNA ligase polypeptide may be an isolated polypeptide.

"Wild-type" refers to the form found in nature. For example, a wild-type polypeptide or polynucleotide sequence is a sequence that is present in an organism that can be isolated from sources in nature, and which has not been intentionally modified by manual procedures. The polypeptide sequence of the wild-type dsRNA ligase described herein is provided by

SEQ ID NO: 302. As used herein, the wild-type sequence may also comprise a purification tag and may be provided by SEQ ID NO: 2.

The terms "polynucleotide" and "nucleic acid" are used interchangeably herein.

The terms "protein", "polypeptide" and "peptide" are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (*e.g.*, glycosylation, phosphorylation, lipidation, myristoylation, ubiquitination, etc.).

"Recombinant" or "engineered" when used with reference to, for example, a cell, nucleic acid or polypeptide, refers to a material or material corresponding to the native or native form of the material, that has been modified in a manner that would not otherwise exist in nature, or is identical thereto but produced or derived from synthetic material and/or by manipulation using recombinant techniques.

The abbreviations used for the genetically encoded amino acids are conventional and are as follows:

Amino Acid	**Three-Letter**	**One-Letter Abbreviation**
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

When the three-letter abbreviations are used, unless specifically preceded by an “L” or a “D” or clear from the context in which the abbreviation is used, the amino acid may be in either the L- or D-configuration about α-carbon (Cα). For example, whereas “Ala” designates alanine without specifying the configuration about the α-carbon, “D-Ala” and “L-Ala” designate D-alanine and L-alanine, respectively.

When the one-letter abbreviations are used, upper case letters designate amino acids in the L-configuration about the α-carbon and lower-case letters designate amino acids in the D-configuration about the α-carbon. For example, “A” designates L-alanine and “a” designates D-alanine. When polypeptide sequences are presented as a string of one-letter or three-letter abbreviations (or mixtures thereof), the sequences are presented in the amino (N) to carboxy (C) direction in accordance with common convention.

The abbreviations used for the genetically encoding nucleotides are conventional and are as follows: adenosine (A); guanosine (G); cytidine (C); thymidine (T); and uridine (U). Unless specifically delineated, the abbreviated nucleotides may be either ribonucleotides or 2’-deoxyribonucleotides. The nucleotides may be specified as being either ribonucleotides or 2’-deoxyribonucleotides on an individual basis or on an aggregate basis. When nucleic acid sequences are presented as a string of one-letter abbreviations, the sequences are presented in the 5’ to 3’ direction in accordance with common convention, and the phosphodiester bonds are not indicated.

The skilled person is well aware that guanine, cytosine, adenine, and uracil may be replaced by other moieties without substantially altering the base pairing properties of an oligonucleotide comprising a nucleotide bearing such replacement moiety. For example, without limitation, a nucleotide comprising inosine as its base may base pair with nucleotides containing adenine, cytosine, or uracil. Hence, nucleotides containing uracil, guanine, or adenine may be replaced in the nucleotide sequences of oligonucleotides featured in the present disclosure by a nucleotide containing, for example, inosine. In another example, adenine and cytosine anywhere in the oligonucleotide can be replaced with guanine and uracil, respectively to form Wobble base pairing with the target mRNA.

"Amino acid difference" or "residue difference" refers to the difference in amino acid residues at a position of a polypeptide sequence relative to the amino acid residue at a corresponding position in the reference sequence. The positions of amino acid differences are generally referred to herein as "Xn", where n refers to the corresponding position in the reference sequence on which the residue differences are based. For example, "a residue

difference at position X6 as compared to SEQ ID NO: 302" refers to a difference in amino acid residue at the polypeptide position corresponding to position 6 of SEQ ID NO: 302. Thus, if the reference polypeptide of SEQ ID NO: 302 has a serine at position 6, then "a residue difference at position X2 as compared to SEQ ID NO: 302" refers to an amino acid substitution to any residue other than serine at the position of the polypeptide corresponding to position 6 of SEQ ID NO: 302.

The specific amino acid residue difference at the position may be indicated as "XnY" or “Xn is Y”, wherein "Xn" specifies the corresponding position in the reference sequence as described above, and "Y" is the single letter identifier of the residue present at that position in the engineered polypeptide. Specific amino acid differences may also be denoted by the conventional notation "AnY", where A is a single letter identifier of the residue in the reference sequence, "n" is the number of residue position in the reference sequence, and “Y” is the single letter identifier of the residue present at that position in the engineered polypeptide.

In some examples, an engineered polypeptide of this disclosure may comprise one or more amino acid residue differences relative to a reference sequence, which is indicated by a list of specific positions at which residue differences are present relative to a reference sequence. In some embodiments, more than one amino acid residue can be used in a specific residue position of an engineered polypeptide, the various amino acid residues can be listed as alternatives, *e.g.* “X19 is Q or D”.

Deletion of an amino acid may be represented by “-”, *e.g.* “an amino acid sequence comprising Xn-” indicates that the amino acid sequence contains a deletion at the position corresponding to “Xn” in the reference sequence. "Deletion" refers to the modification of a polypeptide by removing one or more amino acids from a reference polypeptide. Deletions can include the removal of one or more amino acids, two or more amino acids, five or more amino acids, ten or more amino acids, fifteen or more amino acids, or twenty or more amino acids, up to 10% of the total number of amino acids of the enzyme, or up to 20% of the total number of amino acids making up the reference enzyme while retaining the enzymatic activity of the engineered dsRNA ligase and/or retaining the improved properties of the engineered dsRNA ligase. Deletion may involve the internal portion and/or the terminal portion of the polypeptide. In various embodiments, deletions may include a contiguous segment or may be discontinuous.

In the context of the numbering for a given amino acid or polynucleotide sequence, "corresponding to," "reference to" or "relative to" refers to the numbering of the residues of a

specified reference when the given amino acid or polynucleotide sequence is compared to the reference sequence. In other words, the residue number or residue position of a given sequence is designated with respect to the reference sequence, rather than by the actual numerical position of the residue within the given amino acid or polynucleotide sequence. For example, a given amino acid sequence such as an engineered dsRNA ligase can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although there are gaps, the numbering of the residue in the given amino acid or polynucleotide sequence is made with respect to the reference sequence to which it has been aligned.

"Reference sequence" refers to a defined sequence that is used as a basis for sequence comparison. The reference sequence may be a subset of a larger sequence, for example, a full-length gene or a fragment of a polypeptide sequence. In some embodiments, a "reference sequence" is a wild-type sequence. In some embodiments, a "reference sequence" is an engineered or altered sequences.

Methods of determining percentage sequence identity are known in the art. By way of example, when assessing sequence identity, a sequence having a defined number of contiguous nucleotides or amino acids may be aligned with a nucleic acid or peptide sequence (having the same number of contiguous nucleotides or amino acids) from the corresponding portion of a nucleic acid or peptide sequence disclosed herein. The percentage sequence identity can be calculated by determining the number of positions at which either the identical nucleic acid base or amino acid residue occurs in both sequences, or a nucleic acid base or amino acid residue is aligned with a gap to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the sequence and multiplying the result by 100 to yield the percentage of sequence identity. Those skilled in the art will appreciate that there are many established algorithms available to align two sequences. The optimal alignment of sequences for comparison can be conducted, for example, by the local homology algorithm of Smith and Waterman, 1981, Adv. Appl. Math. 2: 482, by the Homology alignment algorithm of Needleman and Wunsch, 1970, J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. USA 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Package) or by visual inspection (see generally, Current Protocols in Molecular Biology, FM Ausubel *et al.* eds., Current Protocols, a Joint Venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1995 Supplement) (Ausubel)). Examples of algorithms that are suitable for

determining the percent sequence identity and percent sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.*, 1990, J. Mol. Biol. 215: 403-410 and Altschul *et al.*, 1977, Nucleic Acids Res. 3389-3402, respectively. Software for performing BLAST analysis is publicly available through the National Center for Biotechnology Information website. The algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold scores T when aligned with a word of the same length in the database sequence. T is referred to as, the neighborhood word score threshold (Altschul *et al.*, Supra). These initial neighborhood word hits serve as seeds for initiating searches to find longer HSPs that contain them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. For nucleotide sequences, the cumulative scores are calculated using the parameters M (reward score for matched pair of residues; always> 0) and N (penalty score for mismatched residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. The extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quality X from its maximum achieved value; the cumulative score goes 0 or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, the expected value (E) of 10, M = 5, N = -4, and a comparison of both strands as a default value. For amino acid sequences, the BLASTP program uses as defaults the word length (W) of 3, the expected value (E) of 10 and the BLOSUM62 scoring matrix (see Henikoff and Henikoff, 1989, Proc Natl Acad Sci USA 89: 10915). Exemplary determination of sequence alignments and %sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using the default parameters provided.

It will be appreciated that, regardless of the percent sequence identity to a reference sequence, an engineered dsRNA ligase possesses dsRNA ligase activity.

"Suitable reaction conditions" refer to those conditions (*e.g.*, enzyme loading, substrate loading, temperature, pH, etc.) in the reaction system, under which the substrate is converted to the desired product. Suitable reaction conditions can be readily identified by the person skilled in the art. Exemplary "suitable reaction conditions" are provided in the present disclosure and illustrated by examples.

Engineered dsRNA ligase polypeptides

The disclosure provides an engineered dsRNA ligase polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, and 600.

The disclosure provides an engineered dsRNA ligase polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668.

The disclosure provides an engineered dsRNA ligase polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668.

The disclosure also provides an engineered dsRNA ligase polypeptide having dsRNA ligase activity and comprising an amino acid sequence having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488,

490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, and 600, wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure also provides an engineered dsRNA ligase polypeptide having dsRNA ligase activity and comprising an amino acid sequence having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668, wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure also provides an engineered dsRNA ligase polypeptide having dsRNA ligase activity and comprising an amino acid sequence having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668, wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide, which is a polypeptide of: (a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534,

536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, and 600; or (b) a polypeptide having dsRNA ligase activity, which comprises an amino acid sequence having (i) at least 80% sequence identity to one of the polypeptides recited in (a), and (ii) a substitution, deletion, addition or insertion of one or more amino acid residues relative to said one amino acid sequence recited in (a); wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide, which is a polypeptide of: (a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668; or (b) a polypeptide having dsRNA ligase activity, which comprises an amino acid sequence having (i) at least 80% sequence identity to one of the polypeptides recited in (a), and (ii) a substitution, deletion, addition or insertion of one or more amino acid residues relative to said one amino acid sequence recited in (a); wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

The disclosure provides an engineered double-stranded RNA (dsRNA) ligase polypeptide, which is a polypeptide of: (a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668; or (b) a polypeptide having dsRNA ligase activity, which comprises an amino acid sequence having (i) at least 80% sequence identity to one of the polypeptides recited in (a), and (ii) a substitution, deletion, addition or insertion of one or more amino acid residues relative to said one amino acid sequence recited in (a); wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence having at least 85% sequence identity to an even numbered sequence identifier of SEQ ID NOs: 304-600, optionally at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to an even numbered sequence identifier of SEQ ID NOs: 304-600. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence having at least 80% sequence identity to an even numbered sequence identifier of SEQ ID NOs: 304-600, optionally at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to an even numbered sequence identifier of SEQ ID NOs: 304-600.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence having at least 80% sequence identity to an even numbered sequence identifier of SEQ ID NOs: 636-668, optionally at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to an even numbered sequence identifier of SEQ ID NOs: 636-668.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence having at least 80% sequence identity to an even numbered sequence identifier of SEQ ID NOs: 304-600 or 636-668, optionally at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to an even numbered sequence identifier of SEQ ID NOs: 304-600 or 636-668.

The engineered dsRNA ligase polypeptides represented by the even numbered sequence identifiers of SEQ ID NOs: 304 to 600 and 636 to 668 exhibit higher activity than that of SEQ ID NO: 302, as shown in the Examples. The dsRNA ligase polypeptides used in the Examples (represented by even numbered sequence identifiers of SEQ ID NOs: 4 to 300 and 602 to 634, respectively) comprise an even numbered sequence identifier of SEQ ID NOs: 304 to 600 and 636 to 668 and an N-terminal purification tag (MHHHHHHENLYFQS (SEQ ID NO: 669)). For example, SEQ ID NO: 4 comprises: (i) the N-terminal purification tag MHHHHHHENLYFQS (SEQ ID NO: 669); and (ii) SEQ ID NO: 304. dsRNA ligase

polypeptides represented by even numbered sequence identifiers of SEQ ID NOs: 304 to 600 and 636 to 668 do not comprise the N-terminal purification tag represented by SEQ ID NO: 669.

The wild-type dsRNA ligase polypeptide comprises the amino acid sequence of SEQ ID NO: 302 (also accessible under UniProt accession number Q7Y4V8). SEQ ID NO: 2 comprises: (i) the N-terminal purification tag MHHHHHHENLYFQS (SEQ ID NO: 669); and (ii) SEQ ID NO: 302. It will be readily understood that SEQ ID NOs: 2 and 302 both comprise the wild-type dsRNA ligase polypeptide sequence and so both sequences may be referred to herein as the wild-type sequence.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 370, 488, 526, 578, 588, 590 and 592. In some embodiments, the engineered dsRNA ligase polypeptide comprises the amino acid sequence of SEQ ID NO: 666. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 370, 488, 526, 578, 588, 590, 592 and 666. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 70, 188, 226, 278, 288, 290 and 292. In some embodiments, the engineered dsRNA ligase polypeptide comprises the amino acid sequence of SEQ ID NO: 632. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 70, 188, 226, 278, 288, 290, 292 and 632. SEQ ID NOs: 70, 188, 226, 278, 288, 290, 292 and 632 comprise: (i) an N-terminal purification tag MHHHHHHENLYFQS (SEQ ID NO: 669); and (ii) an amino acid sequence provided by SEQ ID NOs: 370, 488, 526, 578, 588, 590, 592 and 666, respectively.

In some embodiments, the engineered dsRNA ligase polypeptide comprises the amino acid sequence of SEQ ID NO: 370. In some embodiments, the engineered dsRNA ligase polypeptide comprises the amino acid sequence of SEQ ID NO: 488. In some embodiments, the engineered dsRNA ligase polypeptide comprises the amino acid sequence of SEQ ID NO: 526. In some embodiments, the engineered dsRNA ligase polypeptide comprises the amino acid sequence of SEQ ID NO: 578. In some embodiments, the engineered dsRNA ligase polypeptide comprises the amino acid sequence of SEQ ID NO: 588. In some embodiments, the engineered dsRNA ligase polypeptide comprises the amino acid sequence of SEQ ID NO: 590. In some embodiments, the engineered dsRNA ligase polypeptide comprises the amino

acid sequence of SEQ ID NO: 592. In some embodiments, the engineered dsRNA ligase polypeptide comprises the amino acid sequence of SEQ ID NO: 666.

The disclosure also provides an engineered dsRNA ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 302, which produces at least 5% more oligonucleotide product than a dsRNA ligase polypeptide comprising the amino acid sequence of SEQ ID NO: 302 under the same ligation reaction conditions, wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide produces at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 100% more oligonucleotide product than a dsRNA ligase polypeptide comprising the amino acid sequence of SEQ ID NO: 302 under the same ligation reaction conditions. In some embodiments, the ligation reaction conditions are as described herein.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 302, optionally at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 302, optionally at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to SEQ ID NO: 302.

In some embodiments, the ligation reaction conditions include about 1 μM to about 10 mM oligonucleotide fragment, a source of ATP, about 5 mM to about 100 mM divalent cation, and about 0.5 g/L to about 10 g/L engineered dsRNA ligase polypeptide, pH of about 4.0 to about 8.0, and temperature of about 10 °C to about 50 °C. In some embodiments, the source of ATP is a stoichiometric concentration of ATP or a stoichiometric excess of ATP. In some embodiments, the source of ATP comprises: (a) polyphosphate kinase (PPK); (b) polyphosphate; and (c) AMP and/or ATP.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* two or

more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more) amino acid residues selected from: X6, X7, X15, X19, X29, X36, X39, X46, X47, X49, X51, X53, X56, X57, X60, X63, X64, X66, X67, X87, X88, X91, X93, X103, X105, X107, X114, X122, X126, X129, X130, X131, X137, X144, X146, X158, X163, X173, X178, X190, X196, X216, X218, X221, X228, X230, X232, X235, X236, X237, X238, X239, X242, X243, X244, X251, X252, X254, X255, X258, X269, X280, X284, X285, X293, X296, X301, X303, X305, X314, X325, and X328, wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more) amino acid residues selected from: X6, X7, X15, X19, X29, X36, X39, X44, X45, X46, X47, X49, X51, X53, X56, X57, X60, X63, X64, X66, X67, X87, X88, X89, X91, X92, X93, X103, X105, X107, X114, X122, X126, X129, X130, X131, X137, X144, X146, X158, X163, X173, X178, X185, X190, X196, X216, X218, X221, X228, X230, X232, X235, X236, X237, X238, X239, X242, X243, X244, X251, X252, X254, X255, X258, X269, X280, X284, X285, X293, X296, X301, X303, X305, X313, X314, X325, and X328, wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising one or more (*e.g.* two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more) of the following amino acid residues: X6 is G ; X7 is Q; X15 is R, D or E; X19 is Q or D; X29 is N or L; X36 is V; X39 is A; X46 is Y; X47 is E; X49 is G; X51 is L; X53 is Y; X56 is R or A; X57 is S; X60 is T, G or P; X63 is S, Q or G; X64 is R, T, Q, F, G, or M; X66 is F or W; X67 is N; X87 is T, P, K or absent; X88 is C; X91 is S; X93 is G, C, or A; X103 is V, C, Y, or T; X105 is V; X107 is R or T; X114 is N; X122 is W; X126 is G; X129 is N; X130 is R, S or Y; X131 is R; X137 is V or C; X144 is N; X146 is R; X158 is W; X163 is G; X173 is L; X178 is R; X190 is Q; X196 is S or C; X216 is L or R; X218 is N; X221 is I; X228 is R; X230 is T; X232 is R; X235 is A, T, or G; X236 is S, L, or F; X237 is S, Q, or R; X238 is F; X239 is G or R; X242 is R or M; X243 is N, S, G, or M; X244 is G or K; X251 is D or L; X252 is V; X254 is K; X255 is C; X258 is V; X269 is L; X280 is W; X284 is A; X285 is A; X293 is R; X296 is R; X301 is G, L, E, or F; X303 is Q; X305 is G; X314 is A or V; X325 is R; and X328 is R; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising one or

more (*e.g.* two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more) of the following amino acid residues: X6 is G or E; X7 is Q; X15 is R, D or E; X19 is Q or D; X29 is N or L; X36 is V; X39 is A; X44 is V; X45 is V; X46 is Y; X47 is E; X49 is G; X51 is L; X53 is Y; X56 is R or A; X57 is S; X60 is T, G or P; X63 is S, Q or G; X64 is R, T, Q, F, G, or M; X66 is F or W; X67 is N; X87 is T, P, K or absent; X88 is C; X89 is T; X91 is S; X92 is D; X93 is G, C, or A; X103 is V, C, Y, or T; X105 is V; X107 is R or T; X114 is N; X122 is W; X126 is G; X129 is N; X130 is R, S or Y; X131 is R; X137 is V or C; X144 is N; X146 is R; X158 is W; X163 is G; X173 is L; X178 is R; X185 is K; X190 is Q; X196 is S or C; X216 is L or R; X218 is N; X221 is I; X228 is R; X230 is T; X232 is R; X235 is A, T, or G; X236 is S, L, or F; X237 is S, Q, R, L or G; X238 is F; X239 is G or R; X242 is R or M; X243 is N, S, G, or M; X244 is G or K; X251 is D or L; X252 is V; X254 is K; X255 is C; X258 is V; X269 is L; X280 is W; X284 is A; X285 is A; X293 is R; X296 is R; X301 is G, L, E, or F; X303 is Q; X305 is G; X313 is A; X314 is A or V; X325 is R; and X328 is R; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, or 11) amino acid residues selected from: X15, X19, X36, X39, X53, X218, X221, X237, X251, X255, and X285; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, or 11) of the following amino acid residues: X15 is D or E; X19 is D; X36 is V; X39 is A; X53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, or 12) amino acid residues selected from: X15, X19, X36, X39, X53, X185, X218, X221, X237, X251, X255, and X285; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising one or more (*e.g.* 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, or 12) of the following amino acid residues: X15 is D or E; X19 is D; X36 is V; X39 is A; X53 is Y; X185

is K; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* two or more or three or more) amino acid residues selected from: X36, X39, X218 and X221; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising one or more (*e.g.* two or more or three or more) of the following amino acid residues: X36 is V; X39 is A; X218 is N; and X221 is I; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 at amino acid residues: X36, X39, X218 and X221; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising the following amino acid residues: X36 is V; X39 is A; X218 is N; and X221 is I; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* two or more) amino acid residues selected from: X39, X218 and X221; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising one or more (*e.g.* two or more) of the following amino acid residues: X39 is A; X218 is N; and X221 is I; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 at amino acid residues: X39, X218 and X221; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising the following amino acid residues: X39 is A; X218 is N; and X221 is I; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* two or more or three or more) amino acid residues selected from: X39, X218, X221, and X255; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising one or more (*e.g.*

two or more or three or more) of the following amino acid residues: X39 is A; X218 is N; X221 is I; and X255 is C; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 at amino acid residues: X39, X218, X221, and X255; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising the following amino acid residues: X39 is A; X218 is N; X221 is I; and X255 is C; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* two or more, three or more, four or more, five or more, six or more, or seven or more) amino acid residues selected from: X39, X53, X218, X221, X237, X251, X255 and X285; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising one or more (*e.g.* two or more, three or more, four or more, five or more, six or more, or seven or more) of the following amino acid residues: X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 at amino acid residues: X39, X53, X218, X221, X237, X251, X255 and X285; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising the following amino acid residues: X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more (*e.g.* two or more, three or more, four or more, five or more, six or more, seven or more, or eight or more) amino acid residues selected from: X15, X39, X53, X218, X221, X237, X251, X255 and X285; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising one or more (*e.g.* two or more, three or more, four or more, five or more, six or more, seven or more, or eight or more) of the following amino acid residues: X15 is D or E; X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 at amino acid residues: X15, X39, X53, X218, X221, X237, X251, X255 and X285; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising the following amino acid residues: X15 is D; X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302. In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence comprising the following amino acid residues: X15 is E; X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302.

In some embodiments, the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X15, X39, X53, X185, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X15 is D; X39 is A; X 53 is Y; X185 is K; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

In some embodiments, the engineered dsRNA ligase polypeptide comprises a purification tag. Purification tags are typically appended to polypeptides so that they can be purified from their crude biological source using an affinity technique. In some embodiments, the purification tag comprises a poly-histidine tag. Poly-histidine tags bind to matrices bearing immobilized metal ions and can be used to purify polypeptides by affinity chromatography. In some embodiments, the purification tag further comprises a protease recognition site for removal of the purification tag. In some embodiments, the protease recognition site comprises a Tobacco Etch Virus (TEV) protease recognition sequence. In some embodiments, the purification tag comprises the amino acid sequence MHHHHHHENLYFQS (SEQ ID NO: 669).

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184,

186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298 and 300.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, and 634.

In some embodiments, the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, and 634.

Immobilization

The disclosure also provides a polypeptide immobilized on a solid support material by chemical bond or a physical adsorption method, wherein the polypeptide comprises an engineered dsRNA ligase polypeptide disclosed herein.

Immobilization of a polypeptide by physical absorption typically involves the polypeptide being physically adsorbed or attached onto a solid support material. Adsorption can occur through weak non-specific forces such as van der Waals, hydrophobic interactions and hydrogen bonds. Physical adsorption may be achieved by soaking the support material in a solution of the polypeptide and incubating to allow time for physical adsorption to occur. Immobilization of a polypeptide by chemical bonding typically involves the attachment of the polypeptide to the support material via a covalent bond.

In some embodiments, the dsRNA ligase polypeptide is immobilized via a spacer positioned between the dsRNA ligase polypeptide and the solid material. In some embodiments, the spacer is a peptide (*e.g.* a peptide comprising 2 or more, 3 or more, 4 or

more, 5 or more, 10 or more, 15 or more, 20 or more, 25 or more, 30 or more, 35 or more, 40 or more, 45 or more, 50 or more, 75 or more, or 100 or more amino acids).

In some embodiments, the engineered dsRNA ligase polypeptide is immobilized using affinity immobilization. In some embodiments, the engineered dsRNA ligase polypeptide is immobilized using metal affinity immobilization, *e.g.* by contacting His-tagged engineered dsRNA ligase polypeptide with immobilized metal such as nickel, zinc, cobalt, or copper.

In some embodiments, the solid support material comprises a membrane, resin, solid carrier, or other solid phase material. A solid support material can be composed of organic polymers such as polystyrene, polyethylene, polypropylene, polyfluoroethylene, polyethyleneoxy, polymethacrylate, and polyacrylamide, as well as co-polymers and grafts thereof. A solid support material can also be inorganic, such as glass, silica, controlled pore glass (CPG), reverse phase silica or metal, such as gold or platinum. The configuration of a solid support material can be in the form of beads, spheres, particles, granules, a gel, a membrane or a surface. Surfaces can be planar, substantially planar, or non-planar. Solid support materials can be porous or non-porous and can have swelling or non-swelling characteristics. A solid support material can be configured in the form of a well, depression, or other container, vessel, feature, or location. Solid support materials useful for immobilizing the dsRNA ligase polypeptide for carrying out a ligase reaction include but are not limited to beads or resins such as polymethacrylate, *e.g.*, polymethacrylates with epoxy functional groups, polymethacrylates with amino epoxy functional groups, polymethacrylates, styrene/DVB copolymer or polymethacrylates with octadecyl functional groups.

Exemplary solid supports include, but are not limited to, chitosan beads, Eupergit C, IB-150, IB-350, IB-C435, IB-A369, IB-A161, IB-A171, IBS500, IB-S861, SEPABEADS (Mitsubishi), *e.g.*, Sepabeads EC-EP, Sepabeads EC-HFA, Sepabeads EC-HG, Sepabeads EC-BU, Sepabeads EC-OD, Sepabeads EC-CM, Sepabeads EC-IDA, Sepabeads EC-EA, Sepabeads EC-HA, Sepabeads EC-QA, Sepabeads EXE, Sepabeads EXA, Dilbeads-TA, Amberzyme Oxirane, Amberlite XAD-7HP, Amberlite FPA98Cl, Amberlite IRA958Cl, Amberlite IRA67, Amberlite FPA90Cl, Amberlite FPA40Cl, Amberlite XAD18, Accurel EP100, ECR8206F/5730, ECR8206/5803, ECR8206M/5749, ReliZyme EP403, ReliZymeEP113, Lewatit VP OC 1600, Diaion WA20, Diaion WA21J, Diaion WA30, Dowex 66, Diaion HPA-25L, Lewatit VP OC 1064 MD PH, Lewatit VP OC 1163, Lifetech ECR8304F. Lifetech ECR8309F, Lifetech ECR8315F, Lifetech ECR8204F, Lifetech

ECR8285, Lifetech ECR1090M, Lifetech ECR1030M, Lifetech ECR8806M, Chromalite (MAM2/F) D6591, Chromalite MIDA/M, Chromalite MIDA/M/Fe, Chromalite MIDA/M/Co, Chromalite MIDA/M/Ni, Chromalite MIDA/M/Cu and Chromalite MIDA/M/Zn.

5

Polynucleotides, control sequences, expression vectors and host cells that can be used to produce engineered dsRNA ligase polypeptides

In another aspect, this disclosure provides polynucleotides encoding engineered polypeptides having dsRNA ligase activity described herein. The polynucleotides can be linked to one or more heterologous regulatory sequences that control gene expression to produce recombinant polynucleotides that are capable of expressing the engineered polypeptides. Expression constructs comprising a heterologous polynucleotide encoding an engineered dsRNA ligase may be introduced into a suitable host cell to express the corresponding engineered dsRNA ligase polypeptide.

As apparent to one skilled in the art, the availability of protein sequences and knowledge of codons corresponding to a variety of amino acids provide an illustration of all possible polynucleotides that encode the protein sequence of interest. The degeneracy of the genetic code, in which the same amino acid is encoded by selectable or synonymous codons, allows for the production of an extremely large number of polynucleotides, all of which encode the engineered dsRNA ligase polypeptides disclosed herein. Thus, upon determination of a particular amino acid sequence, one skilled in the art can generate any number of different polynucleotides by modifying one or more codons in a manner that does not alter the amino acid sequence of the protein. In this regard, this disclosure specifically contemplates each and every possible alteration of a polynucleotide that can be made by selecting a combination based on possible codon selections, for any of the polypeptides disclosed herein, comprising those amino acid sequences of exemplary engineered polypeptides listed in Examples 7 to 12, any of the polypeptides disclosed as even sequence identifiers of SEQ ID NOs: 304 to 600 and 636 to 668, and any of the polypeptides disclosed as even sequence identifiers of SEQ ID NOs: 4 to 300 and 602 to 634.

In various embodiments, the codons are preferably selected to accommodate the host cell in which the recombinant protein is produced. For example, codons preferred for bacteria are used to express genes in bacteria; codons preferred for yeast are used to express genes in yeast; and codons preferred for mammals are used for gene expression in mammalian cells.

In some embodiments, the disclosure provides a polynucleotide encoding an engineered dsRNA ligase polypeptide described above.

In some embodiments, the polynucleotide encodes a polypeptide comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or

at least 99% to a reference sequence that is an even numbered sequence identifier of SEQ ID NOs: 304 to 600 or 636 to 668, wherein the polypeptide has dsRNA ligase activity and exhibits higher enzyme activity than a polypeptide comprising the amino acid of SEQ ID NO: 2 and/or 302.

In some embodiments, the polynucleotide encodes an engineered dsRNA ligase polypeptide described herein and comprises a nucleic acid sequence having at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to a reference polynucleotide selected from the sequences having an odd numbered sequence identifier of SEQ ID NOs: 303 to 599 or 635 to 667, wherein: (i) the polynucleotide does not comprise SEQ ID NO: 301; and (ii) the polynucleotide does not encode a dsRNA ligase polypeptide having the amino acid sequence of SEQ ID NO: 302.

In some embodiments, the polynucleotide encodes an engineered dsRNA ligase polypeptide described herein and comprises a nucleic acid sequence having at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to a reference polynucleotide selected from the sequences having an odd numbered sequence identifier of SEQ ID NOs: 3 to 299 or 601 to 633, wherein: (i) the polynucleotide does not comprise SEQ ID NO: 1; and (ii) the polynucleotide does not encode a dsRNA ligase polypeptide having the amino acid sequence of SEQ ID NO: 2. It will be readily understood that polynucleotides having an odd numbered sequence identifier of SEQ ID NOs: 3 to 299 or 601 to 633 encode an engineered dsRNA ligase polypeptide comprising an N-terminal purification tag (SEQ ID NO: 669).

The isolated polynucleotides encoding engineered dsRNA ligase polypeptides can be manipulated to enable the expression of the engineered polypeptides in a variety of ways, which may comprise further modification of the sequences by codon optimization to improve expression, insertion into suitable expression elements with or without additional control sequences, and transformation into a host cell suitable for expression and production of the engineered polypeptides.

Depending on the expression vector, manipulation of the isolated polynucleotide prior to insertion of the isolated polynucleotide into the vector may be desirable or necessary. Techniques for modifying polynucleotides and nucleic acid sequences using recombinant

DNA methods are well known in the art. Guidance is provided below: Sambrook *et al.*, 2001, Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratory Press; and Current Protocols in Molecular Biology, Ausubel. F. Eds., Greene Pub. Associates, 1998, 2010 Year update.

The disclosure also provides an expression vector comprising the polynucleotide described herein. In some embodiments, the vector is selected from a plasmid, a cosmid, a bacteriophage, or a viral vector. Recombinant expression vectors typically comprise one or more expression regulatory regions, such as promoters and terminators, origin of replication and the like.

Polynucleotides encoding an engineered dsRNA ligase polypeptide described herein can be expressed by inserting the polynucleotide or the nucleic acid construct comprising the polynucleotide sequence into an appropriate expression vector. In generating the expression vector, the coding sequence is located in the vector such that the coding sequence is linked to a suitable control sequence for expression. The recombinant expression vector can be any vector (*e.g.*, a plasmid or virus) that can be conveniently used in recombinant DNA procedures and can result in the expression of a polynucleotide sequence. The choice of vector will generally depend on the compatibility of the vector with the host cell to be introduced into. The vector can be linear or closed circular plasmid. The expression vector may be an autonomously replicating vector, *i.e.*, a vector that exists as an extrachromosomal entity whose replication is independent of chromosomal replication, such as a plasmid, extrachromosomal element, minichromosome, or artificial chromosome. The vector may contain any tools for ensuring self-copying. Alternatively, the vector may be a vector that, when introduced into a host cell, integrates into the genome and replicates with the chromosome into which it is integrated. Moreover, a single vector or plasmid or two or more vectors or plasmids that together comprise the total DNA to be introduced into the genome of the host cell may be used.

Many expression vectors useful to the embodiments of the present disclosure are commercially available. An exemplary expression vector can be prepared by inserting a polynucleotide encoding an engineered dsRNA ligase polypeptide to plasmid pACYC-Duet-1 (Novagen), pBR322 Vector (New England Biolabs), pUC19 Vector (New England Biolabs) or pET T7 Expression Vectors (Novagen).

The disclosure also provides a host cell capable of expressing an engineered dsRNA ligase polypeptide described herein. In some embodiments, the host cell comprises the

nucleic acid molecule described herein, or the vector described herein. In some embodiments, the host cell is *Escherichia coli*.

In some embodiments, the polynucleotide encoding the polypeptide is linked to one or more control sequences for expression of polypeptides in the host cell. Host cells for expression of polypeptides encoded by the expression vectors of the present disclosure are well known in the art, including, but not limited to, bacterial cells such as *E. coli*, *Streptomyces*, and *Salmonella typhimurium*; fungal cells (*e.g.*, *Saccharomyces cerevisiae* or *Pichia pastoris*); insect cells such as Drosophila S2 and Spodoptera Sf9; animal cells such as CHO, COS, BHK, 293 and Bowes melanoma cells; and plant cells. An exemplary host cell is *E. coli* BL21 (DE3). The host cell may be wild-type or may be engineered through genomic editing. Suitable media and growth conditions for the above host cells are well known in the art.

Polynucleotides or vectors used to express polypeptides can be introduced into cells by a variety of methods known in the art. Techniques comprise, among others, electroporation, bio-particle bombardment, liposome-mediated transfection, calcium chloride transfection, and protoplast fusion. Different methods of introducing polynucleotides into cells are known to those skilled in the art.

The host cell may be used to express and isolate the polypeptide described herein.

Process of producing an engineered dsRNA ligase polypeptide

Engineered dsRNA ligase can be obtained by subjecting a polynucleotide encoding an dsRNA ligase to mutagenesis and/or directed evolution. An exemplary directional evolution technique can be found in "Biocatalysis for the Pharmaceutical Industry: Discovery, Development, and Manufacturing" (2009 John Wiley &Sons Asia (Pte) Ltd. ISBN: 978-0-470-82314-9).

When the sequence of an engineered polypeptide is known, the encoding polynucleotide may be prepared by standard solid-phase methods according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be synthesized separately and then ligated (*e.g.*, by enzymatic or chemical ligation methods or polymerase-mediated methods) to form any desired contiguous sequence. For example, the polynucleotides and oligonucleotides of the present disclosure can be prepared by chemical synthesis using, for example, the classic phosphoramidite methods described by Beaucage *et al.*, 1981, Tet Lett 22: 1859-69, or Matthes *et al.* People, 1984, EMBO J. 3: 801-05, as typically practiced in automated synthesis methods. According to the phosphoramidite

method, oligonucleotides are synthesized, purified, annealed, ligated, and cloned into a suitable vector, for example, in an automated DNA synthesizer. In addition, essentially any nucleic acid is available from any of a variety of commercial sources.

The disclosure provides a method of preparing an engineered dsRNA ligase polypeptide, which comprises the steps of culturing a host cell described herein and obtaining an engineered dsRNA ligase polypeptide from the culture. In some embodiments, the process of preparing a polypeptide further comprises isolating the polypeptide. Engineered polypeptides may be expressed in suitable cells and isolated (or recovered) from the host cell and/or culture medium using any one or more of the well-known techniques for protein purification, the techniques for protein purification include, among others, lysozyme treatment, sonication, filtration, salting out, ultracentrifugation and chromatography.

The invention also provides an engineered dsRNA ligase catalyst obtainable by culturing a host cell described herein, or from the method of preparing an engineered dsRNA ligase polypeptide described herein, wherein said engineered dsRNA ligase catalyst comprises cells or culture fluid containing the engineered dsRNA ligase polypeptides, or an article processed therewith, wherein the article refers to an extract obtained from the culture of host cells, an isolated product obtained by isolating or purifying an engineered dsRNA ligase from the extract, or an immobilized product obtained by immobilizing host cells, an extract thereof, or isolated product of the extract.

Ligation reactions

The disclosure provides a method of producing an oligonucleotide from two or more oligonucleotide fragments, wherein the method comprises contacting: (i) two or more oligonucleotide fragments; (ii) an engineered dsRNA ligase polypeptide disclosed herein; (iii) a source of ATP; and (iv) a divalent cation; to obtain an oligonucleotide.

Oligonucleotide products and fragments

Methods of the invention produce oligonucleotides by ligating two or more oligonucleotide fragments. The produced oligonucleotides (also referred to herein as "oligonucleotide products") are nucleic acids which typically comprise up to 100 nucleotides. It will be understood that oligonucleotides described herein comprise RNA. It will also be understood that oligonucleotides described herein comprise a double-stranded region.

As used herein, "oligonucleotide fragment" refers to a nucleic acid that can be ligated to one or more additional oligonucleotide fragments to provide an oligonucleotide product. Each oligonucleotide fragment corresponds to a portion of the oligonucleotide product.

In some embodiments, the oligonucleotide is a therapeutic oligonucleotide. In some embodiments, the therapeutic oligonucleotide is a small interfering RNA (siRNA) or an antisense oligonucleotide (ASO). In some embodiments, the oligonucleotide is an aptamer.

In some embodiments, the oligonucleotide comprises an overhang. In some embodiments, the oligonucleotide comprises a 3' overhang. In some embodiments, the oligonucleotide comprises a 5' overhang. In some embodiments, the overhang comprises 1, 2, 3, 4, 5, 6, 7, or 8 nucleotides. In some embodiments, the oligonucleotide comprises a blunt end. In some embodiments, the oligonucleotide comprises two blunt ends.

In some embodiments, the oligonucleotide is up to 20 nucleotides in length. In some embodiments, the oligonucleotide is up to 25, up to 30, up to 35, up to 40, up to 45, up to 50, up to 55, up to 60, up to 65, up to 70, up to 75, up to 80, up to 85, up to 90, up to 95, or up to 100 nucleotides in length. In some embodiments, the oligonucleotide is up to 60 nucleotides in length.

In some embodiments, the oligonucleotide is at least 20 nucleotides in length. In some embodiments, the oligonucleotide is at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or 100 nucleotides in length.

In some embodiments, the oligonucleotide is 10-100 nucleotides in length. In some embodiments, the oligonucleotide is 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, 10-25, 15-80, 15-70, 15-60, 15-50, 15-40, 15-30, or 15-25 nucleotides in length. In some embodiments, the oligonucleotide is 15-25 nucleotides in length.

As used herein, the two or more oligonucleotide fragments comprise one or more 3' oligonucleotide fragments and one or more 5' oligonucleotide fragments, wherein each of the one or more 3' oligonucleotide fragments comprise a 5'-phosphate group and each of the one or more 5' oligonucleotide fragments typically comprise a 3' terminal ribonucleotide having a 3'-hydroxyl group.

In some embodiments, one or more of the oligonucleotide fragments comprises one or more mismatches. In some embodiments, one or more of the oligonucleotide fragments comprise an overhang. In some embodiments, one or more of the oligonucleotide fragments comprise a 3' overhang. In some embodiments, one or more of the oligonucleotide fragments comprise a 5' overhang. In some embodiments, one or more of the oligonucleotide fragments

comprise a 3' overhang and a 5' overhang. In some embodiments, the overhang comprises 1, 2, 3, 4, 5, 6, 7, or 8 nucleotides.

In some embodiments, the two or more oligonucleotide fragments comprise a first oligonucleotide fragment having an overhang that is complementary to the overhang of a second oligonucleotide fragment. In some embodiments, the two or more oligonucleotide fragments comprise a first oligonucleotide fragment having a 3' overhang and a 5' overhang, wherein the 3' overhang is complementary to the 5' overhang of a second oligonucleotide fragment and the 5' overhang is complementary to the 3' overhang of a third oligonucleotide.

In some embodiments, one or more of the oligonucleotide fragments comprise a blunt end. In some embodiments, one or more of the oligonucleotide fragments comprise a 3' overhang and a 5' blunt end. In some embodiments, one or more of the oligonucleotide fragments comprise a 5' overhang and a 3' blunt end. In some embodiments, the 5' terminal oligonucleotide fragment comprises a 3' overhang and a 5' blunt end. In some embodiments, the 3' terminal oligonucleotide fragment comprise a 5' overhang and a 3' blunt end.

In some embodiments, two or more of the oligonucleotide fragments comprise two or more RNA oligonucleotide fragments. In some embodiments, the two or more RNA oligonucleotide fragments comprise double-stranded RNA (dsRNA) oligonucleotide fragments.

In some embodiments, one or more of the oligonucleotide fragments comprise DNA and RNA. For example, a portion of the oligonucleotide fragment may be double-stranded DNA, while another portion is double-stranded RNA, forming a DNA-RNA chimera.

In some embodiments, one or more of the oligonucleotide fragments comprise one or two strands which are RNA, or a mixture of RNA, DNA, LNA, Morpholino, UNA (unlocked nucleic acid), TNA (threose nucleic acid), GNA (glycol nucleic acid), and/or FANA (Fluoro-arabino nucleic acid), modified RNA, etc. As a non-limiting example, one or both strand(s) could be, for example, RNA except that one or more nucleotide(s) is replaced by DNA, LNA, Morpholino, UNA, TNA, GNA, and/or FANA, and/or modified RNA (*e.g.*, any modified RNA disclosed herein or known in the art, such as 2' modified RNA, including but not limited to 2'-F, 2'-OMe, 2'-O-MOE RNA, etc.).

In some embodiments, the two or more oligonucleotide fragments are the same length. In some embodiments, the two or more oligonucleotide fragments are different lengths. In some embodiments, each of the two or more oligonucleotide fragments are 3-20 nucleotides in length. In some embodiments, each of the two or more oligonucleotide fragments are 4-16 nucleotides in length. In some embodiments, each of the two or more

oligonucleotide fragments are 4-16, 4-15, 5-15, 6-15, 4-14, 4-13, 4-12, 4-11, 4-10, 4-9, 5-9, or 6-9 nucleotides in length.

In some embodiments, each of the two or more oligonucleotide fragments are at least 3 nucleotides in length. In some embodiments, each of the two or more oligonucleotide fragments are at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, or at least 15 nucleotides in length.

In some embodiments, the two or more oligonucleotide fragments comprise 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or 10 or more oligonucleotide fragments.

In some embodiments, one or more ligation reactions are required to generate the oligonucleotide product. In some embodiments, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or 10 or more ligation reactions are required to generate the oligonucleotide product.

In some embodiments, one or more of the oligonucleotide fragments and/or the oligonucleotide comprises a chemical modification. In some embodiments, one or more of the oligonucleotide fragments and/or the oligonucleotide comprises at least one modified backbone modification. In some embodiments, one or more of the oligonucleotide fragments and/or the oligonucleotide comprises at least one modified nucleotide modification. In some embodiments, one or more of the oligonucleotide fragments and/or the oligonucleotide comprises at least one sugar modification (*e.g.* at the 2'-position or 4'-position). In some embodiments, one or more of the oligonucleotide fragments and/or the oligonucleotide comprises: (i) at least one modified backbone modification; (ii) and at least one modified nucleotide modification; and/or (iii) at least one sugar modification.

In some embodiments, one or more of the oligonucleotide fragments and/or the oligonucleotide comprise a modification selected from the group consisting of: 2'-O-methyl (2'-OMe), 2'-flouro (2'-F), 2'-deoxy, 2'-deoxy-2'-fluoro, 2'-O-methoxyethyl (2'-O-MOE), 2'-O-aminopropyl (2'-O-AP), 2'-O-dimethylaminoethyl (2'-O-DMAOE), 2'-O-dimethylaminopropyl (2'-O-DMAP), 2'-O-dimethylaminoethyloxyethyl (2'-O-DMAEOE), 2'-O-N-methylacetamido (2'-O-NMA), locked nucleic acid (LNA), glycol nucleic acid (GNA), phosphoramidate (*e.g.* mesyl phosphoramidate), 2',3'-seco nucleotide mimic, 2'-F-arabino nucleotide, abasic nucleotide, 2'-amino modified nucleotide, 2'-alkyl-modified nucleotide, morpholino nucleotide, vinylphosphonate (*e.g.* 5' vinylphosphonate), and cyclopropyl phosphonate deoxyribonucleotide. In some embodiments, one or more of the oligonucleotide

fragments and/or the oligonucleotide comprises a 2'-modification selected from the group consisting of: 2'-OMe, 2'-F, and 2'-deoxy.

In some embodiments, one or more of the oligonucleotide fragments and/or the oligonucleotide comprises at least one phosphorothioate or methylphosphonate internucleotide linkage. In some embodiments, the oligonucleotide comprises at least one chiral phosphorothioate linkage.

In some embodiments, one or more of the oligonucleotide fragments and/or the oligonucleotide is conjugated to at least one ligand. The ligand may be conjugated to the sense strand, antisense strand or both strands, in any configuration *e.g.* at the 3'-end, 5'-end, non-end or a combination.

In some embodiments, the ligand comprises one or more N-Acetylgalactosamine (GalNAc) derivatives. GalNAc is an amino sugar derivative of galactose which may be used as a targeting ligand in oligonucleotides intended for targeting to the liver, where it binds to the asialoglycoprotein receptors on hepatocytes. In some embodiments, the ligand comprises one or more GalNAc derivatives conjugated through a bivalent or trivalent branched carrier. In some embodiments, the ligand is a peptide or a peptidomimetic.

In some embodiments, the ligand is conjugated to the sense strand. In some embodiments, the ligand is conjugated to the 3' end of the sense strand. In some embodiments, the ligand is conjugated to the 5' end of the sense strand. In some embodiments, the ligand is conjugated to a non-end of the sense strand.

In some embodiments, the ligand is conjugated to the antisense strand. In some embodiments, the ligand is conjugated to the 3' end of the antisense strand. In some embodiments, the ligand is conjugated to a non-end of the antisense strand.

In some embodiments, the oligonucleotide is an RNAi agent comprising at least one 2'-modified nucleotide selected from a group consisting of 2'-OMe, 2'-F, 2'-deoxy, 2'-deoxy-2'-fluoro, and 2'-O-MOE. In some embodiments, the oligonucleotide is an RNAi agent wherein the sense strand is conjugated to one or more GalNAc ligand(s). In some embodiments, one or more of the oligonucleotide fragments comprises at least one 2'-modified nucleotide selected from a group consisting of 2'-OMe, 2'-F, 2'-deoxy, 2'-deoxy-2'-fluoro, and 2'-O-MOE. In some embodiments, one or more of the oligonucleotide fragments is a dsRNA wherein the sense strand is conjugated to one or more GalNAc ligand(s).

In some embodiments, the method is performed with an oligonucleotide fragment concentration of at least 1 mM, at least 2 mM, at least 3 mM, at least 4 mM, at least 5 mM, at least 6 mM, at least 7 mM, at least 8 mM, at least 9 mM, or at least 10 mM. In some

embodiments, the method is performed with at least 1 mM, at least 2 mM, at least 3 mM, at least 4 mM, at least 5 mM, at least 6 mM, at least 7 mM, at least 8 mM, at least 9 mM, or at least 10 mM of each oligonucleotide fragment. In some embodiments, the method is performed with equimolar amounts of each of the two or more oligonucleotide fragments.

In some embodiments, the method produces at least 15 g of oligonucleotide product per litre of reaction mixture. In some embodiments, the method produces at least 16 g, at least 17 g, at least 18 g, at least 19 g, at least 20 g, at least 30 g, at least 40 g, at least 50 g, at least 60 g, at least 70 g, at least 80 g, at least 90, or at least 100 g of oligonucleotide product per litre of reaction mixture.

Engineered dsRNA ligase polypeptide

The method is performed using an engineered dsRNA ligase as described herein.

In some embodiments, the method is performed using about 1 g/L engineered dsRNA ligase polypeptide, optionally 1.1 g/L, 1.15 g/L, 1.2 g/L, 1.25 g/L, 1.3 g/L, 1.35 g/L, 1.4 g/L, 1.45 g/L, 1.5 g/L, 1.55 g/L, 1.6 g/L, 1.65 g/L, 1.7 g/L, 1.75 g/L, 1.8 g/L, 1.85 g/L, 1.9 g/L, 1.95 g/L, 2 g/L, 2.1 g/L, 2.2 g/L, 2.3 g/L, 2.4 g/L, 2.5 g/L, 2.6 g/L, 2.7 g/L, 2.8 g/L, 2.9 g/L, 3 g/L, 3.25 g/L, 3.5 g/L, 3.75 g/L, 4 g/L, 4.5 g/L or 5 g/L engineered dsRNA ligase polypeptide.

Source of ATP

The enzymatic activity of dsRNA ligase requires ATP as a cofactor. One molecule of ATP is converted to AMP per ligation reaction. The catalytic mechanism of dsRNA ligase and the role of ATP in nucleic acid ligation reactions are described above.

In some embodiments, the source of ATP is ATP. In some embodiments, the method is performed using a stoichiometric concentration of ATP. In some embodiments, the method is performed using a stoichiometric excess of ATP. The skilled person can readily determine the stoichiometric concentration of ATP required for a given ligation based on the concentration of the oligonucleotide fragments and the number of ligation reactions required to produce the oligonucleotide product.

In some embodiments, the method is performed using an ATP and/or AMP concentration of about 0.5 mM, about 1 mM, about 2 mM, about 3 mM, about 4 mM, about 5 mM, about 6 mM, about 7 mM, about 8 mM, about 9 mM, about 10 mM, about 12 mM,

about 14 mM, about 16 mM, about 18 mM, about 20 mM, about 22 mM, about 24 mM, about 26 mM, about 28mM or about 30mM.

In some embodiments, the source of ATP is an ATP regeneration system. In some embodiments, the ATP regeneration system comprises: (a) polyphosphate kinase (PPK); (b) polyphosphate; and (c) AMP and/or ATP. Advantageously, the use of an ATP regeneration system overcomes the requirement for high concentrations of ATP to achieve complete ligation. The ATP regeneration system described herein comprises PPK and polyphosphate. PPK generates ATP from AMP using polyphosphate as a phosphate donor. ATP that is converted to AMP during the ligation reaction can be regenerated to ATP by PPK and used as a cofactor in a subsequent ligation reaction. This cycling of ATP obviates the need for high ATP concentration in the starting reaction. Instead, the reaction can be performed using sub-stoichiometric concentrations of ATP, and/or using the cheaper alternative, AMP.

"Polyphosphate kinases" or "PPKs" are a family of enzymes which catalyze the formation of ATP from AMP and polyphosphate.

In some embodiments, the PPK is PPK12. In some embodiments, the PPK comprises an amino acid sequence having at least 70% sequence identity to SEQ ID NO: 670, which is the amino acid sequence of PPK12. In some embodiments, the PPK comprises an amino acid sequence having at least 75%, at least 80%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% sequence identity to SEQ ID NO: 670.

In some embodiments, the PPK comprises an amino acid sequence having at least 70% sequence identity to SEQ ID NO: 671, which is the amino acid sequence of an optimized PPK12. In some embodiments, the PPK comprises an amino acid sequence having at least 75%, at least 80%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% sequence identity to SEQ ID NO: 671.

In some embodiments, the PPK is *Acinetobacter johnsonii* polyphosphate:AMP phosphotransferase (AjPAP) (UniProt ID: Q83XD3). In some embodiments, the PPK comprises an amino acid sequence having at least 70% sequence identity to SEQ ID NO: 672, which is the amino acid sequence of AjPAP. In some embodiments, the PPK comprises an amino acid sequence having at least 75%, at least 80%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least

94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% sequence identity to SEQ ID NO: 672.

In some embodiments, the PPK is used in the form of whole cell, crude extract (*e.g.* cell free lyophilized extract or cell lysates), isolated polypeptide, or purified polypeptide. In some embodiments, the PPK polypeptide is used in an immobilized form as described herein, such as immobilized on a resin.

In some embodiments, the method is performed using about 1 g/L PPK, optionally 1.1 g/L, 1.15 g/L, 1.2 g/L, 1.25 g/L, 1.3 g/L, 1.35 g/L, 1.4 g/L, 1.45 g/L, 1.5 g/L, 1.55 g/L, 1.6 g/L, 1.65 g/L, 1.7 g/L, 1.75 g/L, 1.8 g/L, 1.85 g/L, 1.9 g/L, 1.95 g/L, 2 g/L, 2.1 g/L, 2.2 g/L, 2.3 g/L, 2.4 g/L, 2.5 g/L, 2.6 g/L, 2.7 g/L, 2.8 g/L, 2.9 g/L, 3 g/L, 3.25 g/L, 3.5 g/L, 3.75 g/L, 4 g/L, 4.5 g/L or 5 g/L PPK.

In some embodiments, the polyphosphate is a polyphosphate salt. In some embodiments, the polyphosphate salt is sodium polyphosphate (Maddrell's salt) or sodium hexametaphosphate (Graham's salt).

In some embodiments, the method is performed using a stoichiometric excess of polyphosphate. In some embodiments, the method is performed using a polyphosphate concentration of at least 5 mM, at least 10 mM, at least 15 mM, at least 20 mM, at least 25 mM, at least 30 mM, at least 35 mM, at least 40 mM, at least 45 mM, at least 50 mM, 55 mM, at least 60 mM, at least 65 mM, at least 70 mM, at least 75 mM, at least 80 mM, at least 85 mM, at least 90 mM, at least 95 mM, or at least 100 mM.

In some embodiments, wherein the method is performed in the presence of PPK and polyphosphate, the method is performed in the presence of AMP. In some embodiments, wherein the method is performed in the presence of PPK and polyphosphate, the method is performed using a sub-stoichiometric concentration of ATP and/or AMP.

Divalent cation

The enzymatic activity of dsRNA ligases requires the presence of a divalent cation. The enzymatic activity of PPKs requires the presence of a divalent cation. In some embodiments, the divalent cation comprises Mg^{2+} and/or Mn^{2+}.

In some embodiments, the method is performed with a divalent cation concentration of 5-100 mM, 10-100 mM, 15-100 mM, 20-100 mM, 30-100 mM, 5-90 mM, 5-80 mM, 5-70 mM, 5-60 mM, 5-50 mM, or 30-50 mM. In some embodiments, the method is performed with a divalent cation concentration of at least 5 mM, at least 10 mM, at least 15 mM, at least 20 mM, at least 25 mM, at least 30 mM, at least 35 mM, at least 40 mM, at least 45 mM, at

least 50 mM, 55 mM, at least 60 mM, at least 65 mM, at least 70 mM, at least 75 mM, at least 80 mM, at least 85 mM, at least 90 mM, at least 95 mM, or at least 100 mM.

In some embodiments, the method further comprises purifying the oligonucleotide product from the reaction mixture. In some embodiments, the oligonucleotide product is at least 80% pure, optionally wherein the oligonucleotide product is at least 85%, at least 90%, at least 95% pure, optionally wherein the oligonucleotide product is at least 98% pure, optionally wherein the oligonucleotide product is at least 99% pure, optionally wherein the oligonucleotide product is at least 99.5% pure, optionally wherein the oligonucleotide product is at least 99.9% pure. An oligonucleotide product that is pure does not contain oligonucleotide fragments, intermediate ligation products, or side products arising from non-specific ligation. The oligonucleotide product may be purified or isolated using any method known in the art, for example using gel extractions or using cellulose-based matrices.

The disclosure also provides an oligonucleotide produced by a method described herein. The oligonucleotide may be in any suitable buffer solution. In some embodiments, the buffer solution is selected from Tris buffer (*e.g.* Tris-HCl), phosphate buffer, HEPES, MOPS (3-(*N*-morpholino)propanesulfonic acid), and triethanolamine (TEOA) buffer. In some embodiments, the buffer solution comprises acetate, citrate, prolamine, carbonate, or phosphate, or any combination thereof. In some embodiments, the buffer solution further comprises an agent for controlling the osmolarity of the solution, such that the osmolarity is kept at a desired value, *e.g.*, at the physiologic values of the human plasma. Solutes which can be added to the buffer solution to control the osmolarity include, but are not limited to, proteins, peptides, amino acids, non-metabolized polymers, vitamins, ions, sugars, metabolites, organic acids, lipids, or salts. In some embodiments, the agent for controlling the osmolarity of the solution is a salt. In some embodiments, the agent for controlling the osmolarity of the solution is sodium chloride or potassium chloride.

Reaction conditions

As disclosed herein and exemplified in the examples, the present disclosure contemplates a range of suitable reaction conditions that may be used in the methods described herein, including but not limited to pH, temperature, buffers, substrate loadings, enzyme loading, cofactor loading, pressure, and reaction time. Additional suitable reaction conditions for ligation reactions described herein can be readily optimized by routine experimentation, *e.g.* performing the method described herein under experimental reaction

conditions of varying reagent concentration, pH, temperature, and detecting the rate of oligonucleotide product formation.

In any of the embodiments of the process disclosed herein, the reaction conditions may include a suitable pH. As noted above, the desired pH or desired pH range can be maintained by using an acid or base, a suitable buffer, or a combination of buffer and added acid or base. The pH of the reaction mixture can be controlled before and/or during the reaction. In some embodiments, suitable reaction conditions include a solution pH of about 4 to about 8, a pH of about 5 to about 8, a pH of about 6 to about 8, or a pH of about 7 to about 8. In some embodiments, the reaction conditions include a solution pH of about 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5 or 8.

In any of the embodiments of the method disclosed herein, suitable temperatures can be used for the reaction conditions, taking into consideration of, for example, the increase in reaction rate at higher temperatures, the activity of the enzyme for sufficient duration of the reaction. Accordingly, in some embodiments, suitable reaction conditions include a temperature of about 10°C to about 60°C, about 10°C to about 50°C, about 25°C to about 50°C, about 25°C to about 40°C, about 25°C to about 30°C, or about 10°C to about 30°C. In some embodiments, suitable reaction temperatures include a temperature of about 10°C, 15°C, 20°C, 25°C, 30°C, 35°C, 40°C, 45°C, 50°C, 55°C, or 60°C. In some embodiments, the temperature during the enzymatic reaction can be maintained at a certain temperature throughout the reaction. In some embodiments, the temperature during the enzymatic reaction may be adjusted over a temperature profile during the course of the reaction.

The reaction may be performed in any suitable buffer solution. In some embodiments, the buffer solution is selected from Tris buffer (*e.g.* Tris-HCl), phosphate buffer, HEPES, MOPS (3-(*N*-morpholino)propanesulfonic acid), and triethanolamine (TEOA) buffer. In some embodiments, the buffer solution comprises acetate, citrate, prolamine, carbonate, or phosphate, or any combination thereof. In some embodiments, the buffer solution is phosphate buffered saline (PBS).

In some embodiments, the reaction mixture further comprises a reducing agent, optionally DTT (Dithiothreitol).

In carrying out the ligation reactions described herein, the engineered dsRNA ligase polypeptide may be added to the reaction mixture in different formulation forms, as frozen or lyophilized whole cells (FWC or LWC) transformed with the gene encoding the engineered dsRNA ligase polypeptide and/or as cell lysate or lyophilized cell lysate of such cells, so called shake flask powder (SFP), where the cell debris was removed and/or further purified as

fermentation powder (FP). Whole cells transformed with gene(s) encoding the engineered dsRNA ligase polypeptide or cell extracts, lysates thereof, and isolated enzymes can be used in a wide variety of different forms, including solids (*e.g.*, lyophilized, spray dried, or the like) or semisolid (*e.g.*, a crude paste). The cell extract or cell lysate may be partially purified by precipitation (*e.g.*, ammonium sulfate, polyethyleneimine, heat treatment or the like), followed by desalting procedures (*e.g.*, ultrafiltration, dialysis, and the like) prior to lyophilization. Any of the enzyme preparations can be immobilized to a solid phase material (such as a resin).

In any of the embodiments of the process disclosed herein, wherein an engineered polypeptide is expressed in the form of a secreted polypeptide, a culture medium containing the secreted polypeptide can be used in the process herein.

In any of the embodiments of the process disclosed herein, the solid reactants (*e.g.*, enzymes, salts, etc.) can be provided to the reaction in a variety of different forms, including powders (*e.g.*, lyophilized, spray dried, etc.), solutions, emulsions, suspensions, and the like. The reactants can be readily lyophilized or spray-dried using methods and instrumentation known to one skilled in the art. For example, the protein solution can be frozen at -80 °C in small aliquots, and then added to the pre-chilled lyophilization chamber, followed by the application of a vacuum.

In any of the embodiments of the process disclosed herein, the order of addition of reactants is not critical. The reactants may be added together to the solvent at the same time or alternatively, some reactants may be added separately, and some may be added together at different time points.

The methods of performing a ligation reaction may comprise the further step of isolating the oligonucleotide product of the enzymatic reaction. In particular, this step is typically performed after completion of the enzymatic reaction. The oligonucleotide is in particular typically separated from one or more, in particular essentially all of the other components of the reaction mixture. For example, the oligonucleotide is typically separated from the remaining substrate, side products, and/or enzymes. Isolation of the oligonucleotide may be achieved by means and techniques known in the art, *e.g.* by separating oligonucleotides based on their size such as by gel electrophoresis and gel extractions or using cellulose-based matrices. In some embodiments, the method further comprises purifying the oligonucleotide by ultrafiltration and chromatography.

Modifications

In some embodiments, the oligonucleotide fragment(s) and/or the oligonucleotide comprises a modification, *e.g.* a chemical modification. As used herein, the term "oligonucleotide fragment(s)" means one or more oligonucleotide fragments. It will be appreciated that modifications which are present in the oligonucleotide fragment(s) are typically present in the oligonucleotide produced from said oligonucleotide fragment(s). In some embodiments, modification(s) are introduced to and/or removed from the oligonucleotide product.

In some embodiments, the oligonucleotide fragment(s) and/or oligonucleotide comprises a chemical modification. In some embodiments, the oligonucleotide fragment(s) and/or oligonucleotide comprises at least one backbone modification. In some embodiments, the oligonucleotide fragment(s) and/or oligonucleotide comprises at least one nucleotide modification. In some embodiments, the oligonucleotide fragment(s) and/or oligonucleotide comprises at least one sugar modification (*e.g.* at the 2'-position or 4'-position). In some embodiments, the oligonucleotide fragment(s) and/or oligonucleotide comprises: (i) at least one backbone modification; (ii) at least one nucleotide modification; and/or (iii) at least one sugar modification.

Modifications include, but are not limited to, end modifications of the terminal oligonucleotide fragments, *e.g.,* 5'-end modifications (phosphorylation, conjugation, inverted linkages) or 3'-end modifications (conjugation, inverted linkages, etc.); base modifications, *e.g.*, replacement with stabilizing bases, destabilizing bases, or bases that base pair with an expanded repertoire of partners, removal of bases (abasic nucleotides), or conjugated bases; sugar modifications (*e.g.*, at the 2'-position or 4'-position) or replacement of the sugar; or backbone modifications, including modification or replacement of the phosphodiester linkages.

In some embodiments, a terminal oligonucleotide fragment and/or oligonucleotide comprises a cap. The term "cap" and the like include a chemical moiety attached to the end of a double-stranded nucleotide duplex, but is used herein to exclude a chemical moiety that is a nucleotide or nucleoside. A "3' cap" is attached at the 3' end of a nucleotide or oligonucleotide and protects the molecule from degradation, *e.g.*, from nucleases, such as those in blood serum or intestinal fluid. A non-nucleotidic 3' cap is not a nucleotide and can replace a TT or UU dinucleotide at the end of a blunt-ended oligonucleotide. In some embodiments, non-nucleotidic 3' end caps are as disclosed in, for example, WO 2005/021749 and WO 2007/128477; and U.S. Pat. No. 8,097,716; U.S. Pat. No. 8,084,600; and U.S. Pat.

No. 8,344,128. A "5' cap" is attached at the 5' end of a nucleotide or oligonucleotide. A cap should not interfere (or unduly interfere) with oligonucleotide activity.

In some embodiments, the oligonucleotide fragment(s) and/or oligonucleotide comprises one or more mismatches. A mismatch is defined herein as a difference between the base sequence or length when two sequences are maximally aligned and compared. In the context of double-stranded oligonucleotides (in which two sequences are aligned antiparallel to each other) a mismatch is defined as a position wherein the base of one sequence is not complementary to the base of the other sequence. Thus, a mismatch is counted, for example, if a position in the first sequence has a particular base (e.g., A), and the corresponding position in the second sequence has a base which is not complementary to said base in the first sequence (e.g., G), when the first and second sequences are aligned antiparallel to each other. Note, however, that on a given RNA strand, a U can be replaced by T (either as RNA or, preferably, DNA, e.g., 2'-deoxy-thymidine); the replacement of a U with a T is not a mismatch as used herein, as either U or T can pair with A on the opposite strand. An RNA oligonucleotide can thus comprise one or more DNA bases, e.g., T. No mismatch is counted between a DNA portion(s) of an RNAi agent and the corresponding target mRNA if basepairing occurs (e.g., between A, G, C, or T in the DNA portion, and the corresponding U, C, G, or A, respectively in the mRNA).

A mismatch is also counted, e.g., if a position in one sequence has a base (e.g., A), and the corresponding position on the other sequence has no base (e.g., that position is an abasic nucleotide, which comprises a phosphate-sugar backbone but no base). A single-stranded nick in either sequence (or in the sense or anti-sense strand) is not counted as mismatch. Thus, as a non-limiting example, no mismatch would be counted if one sequence (in the 5'→3' orientation) comprises the sequence AG, but the complementary sequence (in the 3'→5' orientation) comprises the sequence TC with a single-stranded nick between the T and the C. A nucleotide modification in the sugar or phosphate is also not considered a mismatch. Thus, if one sequence comprises a G, and the complementary sequence comprises a modified C (e.g., 2'-modification) at the same position, no mismatch would be counted.

Thus, no mismatches are counted if modifications are made to the sugar, phosphate, or backbone of the oligonucleotide without modifying the base. Thus, in the context of double-stranded RNAi, a strand having a given sequence as an RNA would have zero mismatches from its complement sequence as a PNA; or morpholino; or LNA; or TNA; or GNA; or FANA; or a mix or chimera of RNA and DNA, TNA, GNA, FANA, Morpholino, UNA,LNA, and/or PNA, etc. No mismatch would occur between a nucleotide which is T,

and a nucleotide which is A with a 5' modification and/or a 2'-modification. The key feature of a mismatch (base replacement) is that it would not be able to base-pair with the corresponding base on the opposite strand. In addition, terminal overhangs such as "UU" or "dTdT" are not counted when counting the number of mismatches. In such cases, a mismatch is defined as a position wherein the base of one sequence does not match the base of the other sequence.

It is noted that dTdT (2'-deoxy-thymidine-5'-phosphate and 2'-deoxy-thymidine-5'-phosphate), or in some cases, TT or UU, can be added as a terminal dinucleotide cap or extension to one or both 3'-ends of the oligonucleotide, but this cap or extension is not included in the calculation of the total number of mismatches and is not considered part of the target sequence. This is because the terminal dinucleotide protects the ends from nuclease degradation but does not contribute to target specificity (Elbashir et al. 2001 Nature 411: 494-498; Elbashir et al. 2001 EMBO J. 20: 6877-6888; and Kraynack et al. 2006 RNA 12:163-176).

There are several examples in the art describing sugar, base, phosphate and backbone modifications that can be introduced into nucleic acid molecules with significant enhancement in their nuclease stability and efficacy. For example, oligonucleotides are modified to enhance stability and/or enhance biological activity by modification with nuclease resistant groups, for example, 2'-amino, 2'-C-allyl, 2'-flouro, 2'-O-methyl, 2'-O-allyl, 2'-H, nucleotide base modifications. Sugar modification of nucleic acid molecules are extensively described in the art.

Additional modifications and conjugations of oligonucleotides have been described. Soutschek *et al.* 2004 Nature 432: 173-178 presented conjugation of cholesterol to the 3'-end of the sense strand of an siRNA molecule by means of a pyrrolidine linker, thereby generating a covalent and irreversible conjugate. Chemical modifications (including conjugation with other molecules) of oligonucleotides may also be made to improve the *in vivo* pharmacokinetic retention time and efficiency.

In some embodiments, the oligonucleotide fragment(s) and/or oligonucleotide comprises a modified base. The disclosure encompasses an oligonucleotide and oligonucleotide fragments with a substitution of a single nucleotide at a given position with a modified version of the same nucleotide. Thus a nucleotide (A, G, C or U) can be replaced by a modified base selected from 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil,

dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5- oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, 2,6-diaminopurine, 5-hydroxymethyl cytosine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiothymine, 5-propynyl (—C═C—CH_3) uracil and cytosine and other alkynyl derivatives of pyrimidine bases, 6-azo uracil, cytosine and thymine, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methyladenine, 2-F-adenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine..

Additional modified variants include the addition of any other moiety (e.g., a radiolabel or other tag or conjugate) to the oligonucleotide or oligonucleotide fragment; provided that the base sequence is identical, the addition of other moieties produces a "modified variant" (with no mismatches).

In addition to these modifications and patterns (e.g., formats) for modifications, other modifications or sets of modifications of the sequences provided can be generated using common knowledge of nucleic acid modification. These various embodiments and embodiments of the oligonucleotides of the present disclosure can be used in RNA interference.

In some embodiments, the oligonucleotide and/or oligonucleotide fragment(s) comprises a modification that causes the oligonucleotide to have increased stability in a biological sample or environment (e.g., cytoplasm, interstitial fluid, blood serum, lung or intestinal lavage).

In some embodiments, the oligonucleotide and/or oligonucleotide fragment(s) comprises a modification that promotes cleavage by the RNA-induced silencing complex (*i.e.* a "RISC cleavage site"). The RISC cleavage site is the site on the target at which cleavage occurs. In some embodiments, the antisense strand comprises a RISC cleavage site. For an RNAi agent having a duplex region of 17-23 nucleotide in length, the cleavage site of the

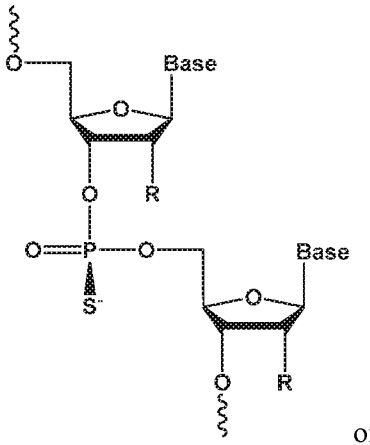
antisense strand is typically around the 10, 11 and 12 positions from the 5'-end. As used herein, the term "cleavage region" refers to a region that is located immediately adjacent to the cleavage site. In some embodiments, the cleavage region comprises three bases on either end of, and immediately adjacent to, the cleavage site. In some embodiments, the cleavage region comprises two bases on either end of, and immediately adjacent to, the cleavage site. In some embodiments, the cleavage site specifically occurs at the site bound by nucleotides 10 and 11 of the antisense strand, and the cleavage region comprises nucleotides 11, 12 and 13 of the antisense strand.

In some embodiments, the oligonucleotide fragment(s) and/or oligonucleotide comprises a modified backbone. As used herein, an unmodified backbone consists of 3' to 5' phosphodiester bonds. A modified backbone may comprise non-natural internucleoside linkages. Oligonucleotides having a modified backbone include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone.

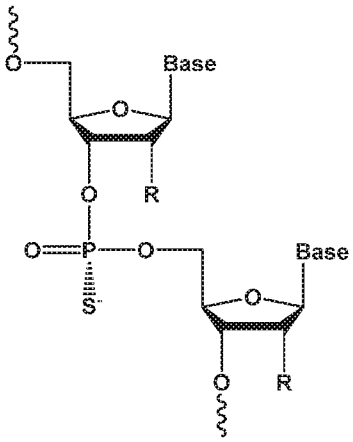
Oligonucleotide fragments comprising a modified backbone include, but are not limited to, those that do not have a phosphorus atom in the backbone. Modified backbones include, but are not limited to, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates (*e.g.* 3'-alkylene phosphonates and chiral phosphonates), phosphinates, phosphoramidates (*e.g.* mesyl phosphoramidate, 3'-amino phosphoramidate and aminoalkylphosphoramidates), thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5'-linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'.

Oligonucleotide fragments comprising a modified backbone that does not include a phosphorus atom therein may have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatoms and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts.

In some embodiments, the oligonucleotide and/or oligonucleotide fragment(s) comprises at least one phosphonate linkage, wherein the phosphonate is a modified phosphonate selected from the group consisting of: phosphorothioate (which may be an *R*p isomer or an *S*p isomer):

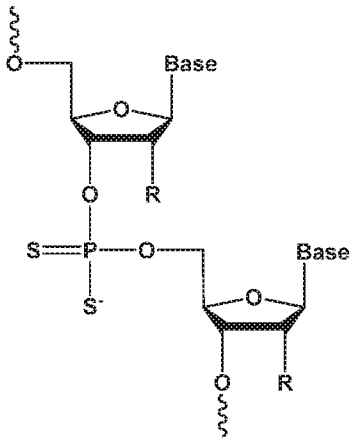


or



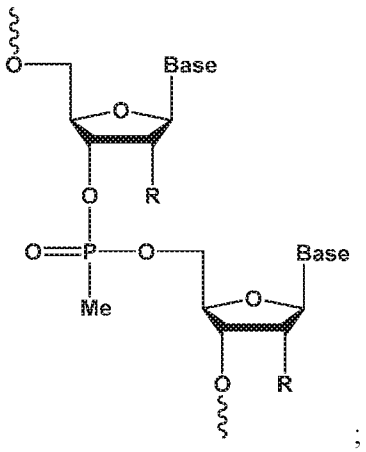
;

phosphorodithioate:



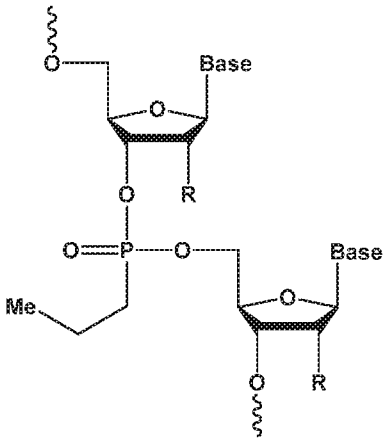
;

methylphosphonate:



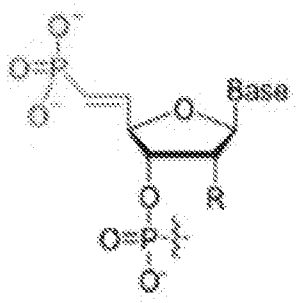
;

methoxypropylphosphonate:



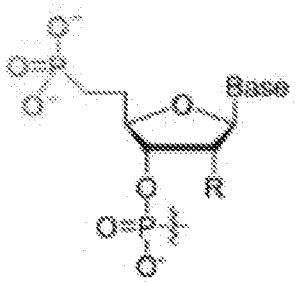
;

5'-(*E*)-vinylphosphonate:



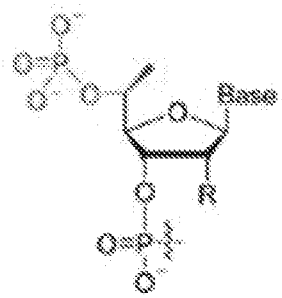
;

5’-methylphosphonate:



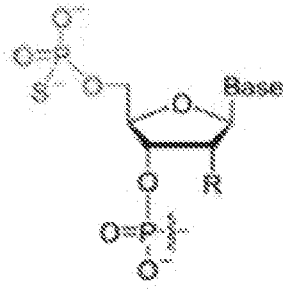
;

(*S*)-5’-C-methyl with phosphonate:



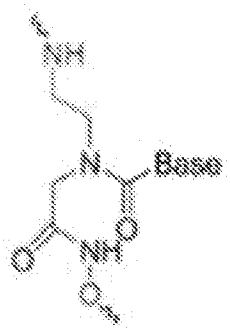
;

5’-phosphorothioate;



;

and peptide nucleic acid:



.

In some embodiments, the oligonucleotide and/or oligonucleotide fragment(s) comprises: at least one 5’-uridine-adenine-3’ (5’-ua-3’) dinucleotide, wherein the uridine is a 2’-modified nucleotide; at least one 5’-uridine-guanine-3’ (5’-ug-3’) dinucleotide, wherein the 5’-uridine is a 2’-modified nucleotide; at least one 5’-cytidine-adenine-3’ (5’-ca-3’)

dinucleotide, wherein the 5'-cytidine is a 2'-modified nucleotide; or at least one 5'-uridine-uridine-3' (5'-uu-3') dinucleotide, wherein the 5'-uridine is a 2'-modified nucleotide. These dinucleotide motifs are particularly prone to serum nuclease degradation (e.g. RNase A). Chemical modification at the 2'-position of the first pyrimidine nucleotide in the motif prevents or slows down such cleavage. This modification recipe is also known under the term 'endo light'.

In some embodiments, the oligonucleotide and/or oligonucleotide fragment(s) comprise a modified nucleobase, wherein the modified nucleobase is difluorotolyl, nitroindolyl, nitropyrrolyl, or nitroimidazolyl. In a particular embodiment, the modified nucleobase is difluorotolyl. In some embodiments, wherein the oligonucleotide and/or oligonucleotide fragment(s) is double-stranded, only one of the two strands contains a modified nucleobase. In some embodiments, wherein the oligonucleotide and/or oligonucleotide fragment(s) is double-stranded, both of the strands contain a modified nucleobase.

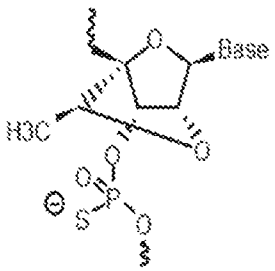
In some embodiments, the oligonucleotide fragment(s) and/or oligonucleotide comprises a modified sugar. Sugar modifications typically involve chemical modification of the sugar moiety of RNA or DNA. Sugar modifications include, but are not limited to, one of the following at the 2'-position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl can be substituted or unsubstituted C_1 to C_{10} alkyl or C2 to C_{10} alkenyl and alkynyl. Exemplary modifications include $O[(CH_2)_nO]_mCH_3$, $O(CH_2)_nOCH_3$, $O(CH_2)_nNH_2$, $O(CH_2)_nCH_3$, $O(CH_2)_nONH_2$, and $O(CH_2)_nON[(CH_2)_nCH_3)]_2$, where n and m are from 1 to about 10. Oligonucleotide fragments for use in the methods described herein may include one of the following at the 2' position: C_1 to C_{10} lower alkyl, substituted lower alkyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, SO_2CH_3, ONO_2, NO_2, N_3, NH_2, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of a therapeutic RNA, or a group for improving the pharmacodynamic properties of a therapeutic RNA. In some embodiments, the modification comprises a 2'-methoxyethoxy (also known as 2'-O-(2-methoxyethyl) or 2'-O-MOE), 2'-dimethylaminooxyethoxy (also known as 2'-DMAOE), and 2'-dimethylaminoethoxyethoxy (also known in the art as 2'-O-dimethylaminoethoxyethyl or 2'-DMAEOE). Further exemplary modifications include: 5'-Me-2'-F nucleotides, 5'-Me-2'-Ome nucleotides, 5'-Me-2'-deoxynucleotides, 2'-alkoxyalkyl; and 2'-NMA (N-methylacetamide).

Other modifications include 2'-methoxy (2'-OCH_3), 2'-aminopropoxy (2'-$OCH_2CH_2CH_2NH_2$) and 2'-fluoro (2'-F). Similar modifications can also be made at other positions on an RNA, particularly the 3' position of the sugar on the 3' terminal nucleotide or in 2'-5' linked dsRNA and the 5' position of 5' terminal nucleotide.

In some embodiments, the oligonucleotide fragment(s) and/or oligonucleotide comprises at least one modified nucleotide. In some embodiments, the modification is selected from the group consisting of: 2'-O-methyl (2'-Ome), 2'-flouro (2'-F), 2'-deoxy, 2'-deoxy-2'-fluoro, 2'-O-methoxyethyl (2'-O-MOE), 2'-O-aminopropyl (2'-O-AP), 2'-O-dimethylaminoethyl (2'-O-DMAOE), 2'-O-dimethylaminopropyl (2'-O-DMAP), 2'-O-dimethylaminoethyloxyethyl (2'-O-DMAEOE), 2'-O-N-methylacetamido (2'-O-NMA), locked nucleic acid (LNA), glycol nucleic acid (GNA), phosphoramidate (*e.g.* mesyl phosphoramidate), 2',3'-seco nucleotide mimic, 2'-F-arabino nucleotide, abasic nucleotide, 2'-amino modified nucleotide, 2'-alkyl-modified nucleotide, morpholino nucleotide, vinylphosphonate (*e.g.* 5' vinylphosphonate), and cyclopropyl phosphonate deoxyribonucleotide. In some embodiments, one or more of the oligonucleotide fragments comprises a 2'-modification selected from the group consisting of: 2'-Ome, 2'-F, and 2'-deoxy. In some embodiments, the oligonucleotide and/or oligonucleotide fragment(s) comprises one or more 3'-O-methyl nucleotide.

In some embodiments, the oligonucleotide and/or oligonucleotide fragment(s) comprises a 2'-modification selected from the group consisting of: 2'-O-methyl (2'-OMe), 2'-flouro (2'-F), 2'-deoxy, 2'-deoxy-2'-fluoro, 2'-O-methoxyethyl (2'-O-MOE), 2'-O-aminopropyl (2'-O-AP), 2'-O-dimethylaminoethyl (2'-O-DMAOE), 2'-O-dimethylaminopropyl (2'-O-DMAP), 2'-O-dimethylaminoethyloxyethyl (2'-O-DMAEOE), 2'-O-N-methylacetamido (2'-O-NMA), locked nucleic acid (LNA), phosphoramidate (*e.g.* mesyl phosphoramidate), 2',3'-seco nucleotide mimic, 2'-F-arabino nucleotide, abasic nucleotide, 2'-amino modified nucleotide, 2'-alkyl-modified nucleotide, morpholino nucleotide, vinylphosphonate (*e.g.* 5' vinylphosphonate), deoxyribonucleotide, and cyclopropyl phosphonate. In some embodiments, the oligonucleotide and/or oligonucleotide fragment(s) comprises one or more 3'-O-methyl nucleotide.

In some embodiments, the oligonucleotide and/or oligonucleotide fragment(s) comprises a bridged nucleic acid. In some embodiments, the bridged nucleic acid is locked nucleic acid. In some embodiments, the bridged nucleic acid is constrained ethyl bridged nucleic acid:



In some embodiments, all pyrimidines (uridine and cytidine) are 2' O-methyl-modified nucleosides.

In some embodiments, the sense and/or antisense strand is conjugated to one or more diagnostic compound, reporter group, cross-linking agent, nuclease-resistance conferring moiety, modified or unmodified nucleobase, lipophilic molecule, cholesterol, lipid, lectin, steroid, uvaol, hecigenin, diosgenin, terpene, triterpene, sarsasapogenin, Friedelin, epifriedelanol-derivatized lithocholic acid, vitamin, carbohydrate, dextran, pullulan, chitin, chitosan, synthetic carbohydrate, oligo lactate 15-mer, natural polymer, low- or medium-molecular weight polymer, inulin, cyclodextrin, hyaluronic acid, protein, protein-binding agent, integrin-targeting molecule, polycationic, peptide, polyamine, peptide mimic, and/or transferrin.

In some embodiments, the antisense strand comprises at least one 2'-OMe modified nucleotide. In some embodiments, the antisense strand comprises at least one 2'-F modified nucleotide. In some embodiments, the antisense strand comprises at least one 2'-deoxy modified nucleotide. In some embodiments, the antisense strand comprises at least one 2'-OMe modified nucleotide, at least one 2'-F modified nucleotide, or at least one 2'-deoxy modified nucleotide, or any combination thereof. In some embodiments, the antisense strand comprises alternating 2'-OMe and 2'-F modified nucleotides. In some embodiments, the antisense strand comprises at least one 5' vinylphosphonate. In some embodiments, the antisense strand comprises at least one chiral phosphorothioate linkage. In some embodiments, the antisense strand comprises at least one GNA. In some embodiments, the sense strand comprises at least one 2'-OMe modified nucleotide. In some embodiments, the sense strand comprises at least one 2'-F modified nucleotide. In some embodiments, the sense strand comprises at least one 2'-deoxy modified nucleotide. In some embodiments, the sense strand comprises at least one 2'-OMe modified nucleotide, at least one 2'-F modified nucleotide, or at least one 2'-deoxy modified nucleotide, or any combination thereof. In some embodiments, the sense strand comprises alternating 2'-OMe and 2'-F modified nucleotides. In some embodiments, the antisense strand and the sense strand each comprise at

least one 2'-OMe modified nucleotide. In some embodiments, the antisense strand and the sense strand each comprise at least one 2'-F modified nucleotide. In some embodiments, the antisense strand and the sense strand each comprise alternating 2'-OMe and 2'-F modified nucleotides. In some embodiments, the sense strand comprises at least one 5' vinylphosphonate. In some embodiments, the sense strand comprises at least one chiral phosphorothioate linkage. In some embodiments, the sense strand comprises at least one GNA.

In some embodiments, the sense strand comprises alternating 2'-OMe and 2'-F modified nucleotides over the full length of the sense strand. In some embodiments, the sense strand comprises alternating 2'-OMe and 2'-F modified nucleotides over part of the length of the sense strand e.g. over at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides of the sense strand.

In some embodiments, the antisense strand comprises alternating 2'-OMe and 2'-F modified nucleotides over the full length of the antisense strand. In some embodiments, the antisense strand comprises alternating 2'-OMe and 2'-F modified nucleotides over part of the length of the antisense strand e.g. over at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides of the antisense strand.

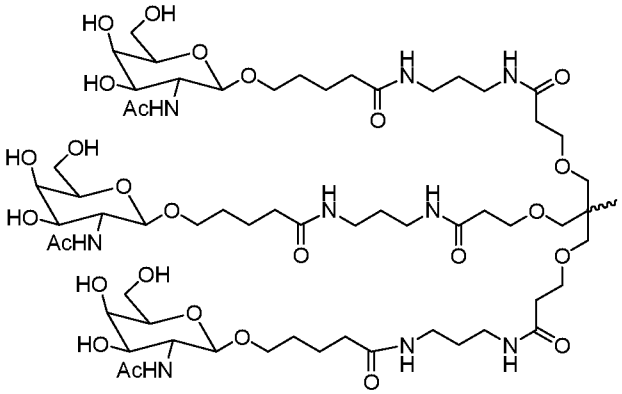
In some embodiments, the sense strand and antisense strand each comprise alternating 2'-OMe and 2'-F modified nucleotides over the full length of the sense strand and the antisense strand. In some embodiments, the sense strand and the antisense strand comprise alternating 2'-OMe and 2'-F modified nucleotides over part of the length of the sense strand and the antisense strand e.g. over at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides of the sense strand and the antisense strand.

Ligands

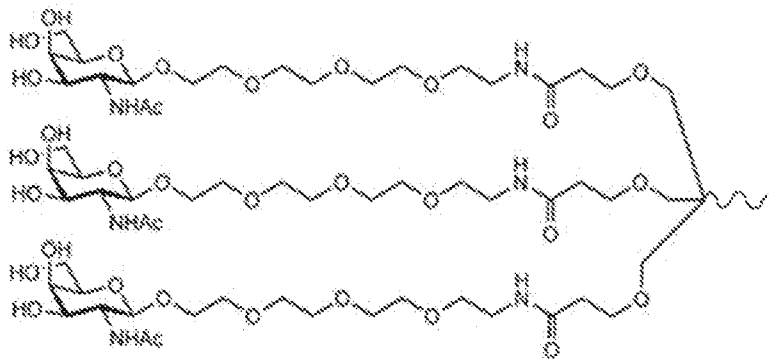
In some embodiments, one or more of the oligonucleotide fragments is conjugated to at least one ligand. In some embodiments, the oligonucleotide product is conjugated to at least one ligand. The ligand may be conjugated to the sense strand, antisense strand or both strands, in any configuration *e.g.* at the 3'-end, 5'-end, non-end or a combination.

In some embodiments, the ligand comprises one or more N-Acetylgalactosamine (GalNAc) derivatives. In some embodiments, the ligand comprises one or more GalNAc derivatives conjugated through a bivalent or trivalent branched carrier.

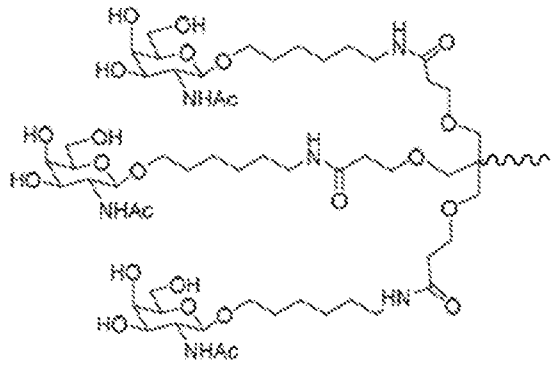
In some embodiments, the ligand is:



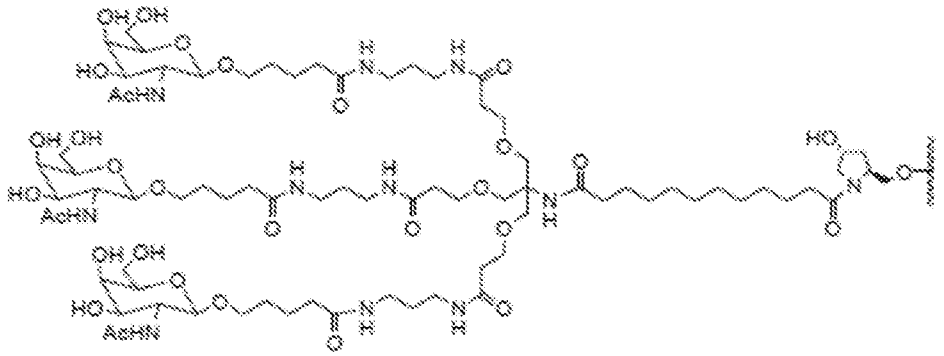
In some embodiments, the ligand is:



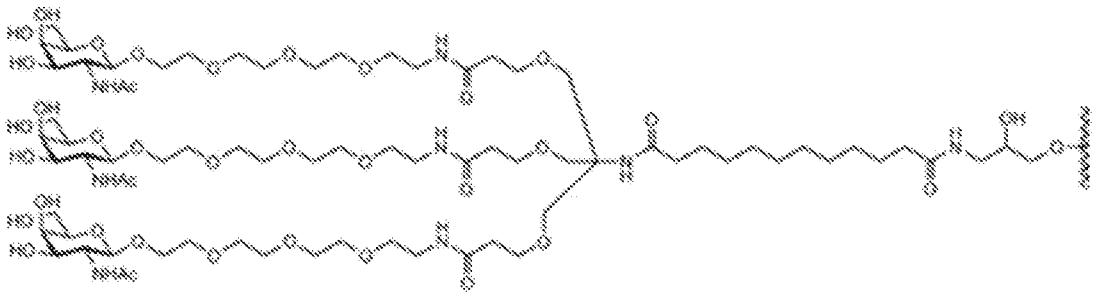
In some embodiments, the ligand is:



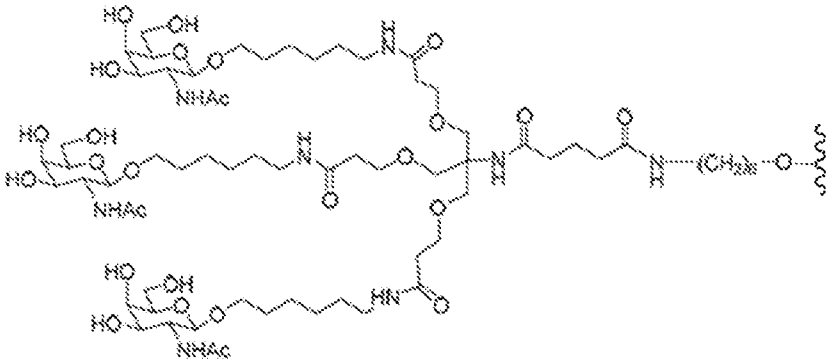
In some embodiments, the ligand is:



In some embodiments, the ligand is:



In some embodiments, the ligand is:



In some embodiments, a ligand alters the distribution, targeting or lifetime of the molecule into which it is incorporated. In some embodiments, a ligand provides an enhanced affinity for a selected target, *e.g.*, molecule, cell or cell type, compartment, receptor *e.g.*, a cellular or organ compartment, tissue, organ or region of the body, as, *e.g.*, compared to a

species absent such a ligand. Ligands providing enhanced affinity for a selected target are also termed targeting ligands.

Some ligands can have endosomolytic properties. The endosomolytic ligands promote the lysis of the endosome and/or transport of the oligonucleotide, or a composition comprising the oligonucleotide, from the endosome to the cytoplasm of the cell. The endosomolytic ligand may be a polyanionic peptide or peptidomimetic which shows pH-dependent membrane activity and fusogenicity. In some embodiments, the endosomolytic ligand assumes its active conformation at endosomal pH. The "active" conformation is that conformation in which the endosomolytic ligand promotes lysis of the endosome and/or transport of the oligonucleotide, or a composition comprising the oligonucleotide, from the endosome to the cytoplasm of the cell. Exemplary endosomolytic ligands include the GALA peptide (Subbarao et al., Biochemistry, 1987, 26: 2964-2972), the EALA peptide (Vogel et al., J. Am. Chem. Soc., 1996, 118: 1581-1586), and their derivatives (Turk et al., Biochem. Biophys. Acta, 2002, 1559: 56-68). The endosomolytic component may contain a chemical group (e.g., an amino acid) which will undergo a change in charge or protonation in response to a change in pH. The endosomolytic component may be linear or branched.

Ligands can improve transport, hybridization, and specificity properties and may also improve nuclease resistance of the resultant natural or modified oligonucleotide.

Ligands in general can include therapeutic modifiers, *e.g.*, for enhancing uptake; diagnostic compounds or reporter groups *e.g.*, for monitoring distribution; cross-linking agents; and nuclease-resistance conferring moieties. General examples include lipids, steroids, vitamins, sugars, proteins, peptides, polyamines, and peptide mimics.

Ligands can include a naturally occurring substance, such as a protein (*e.g.*, human serum albumin (HSA), low-density lipoprotein (LDL), high-density lipoprotein (HDL), or globulin); a carbohydrate (*e.g.*, a dextran, pullulan, chitin, chitosan, inulin, cyclodextrin or hyaluronic acid); or a lipid. The ligand may also be a recombinant or synthetic molecule, such as a synthetic polymer, *e.g.*, a synthetic polyamino acid, an oligonucleotide (*e.g.*, an aptamer). Examples of polyamino acids include polylysine (PLL), poly L aspartic acid, poly L-glutamic acid, styrene-maleic acid anhydride copolymer, poly(L-lactide-co-glycolied) copolymer, divinyl ether-maleic anhydride copolymer, N-(2-hydroxypropyl)methacrylamide copolymer (HMPA), polyethylene glycol (PEG), polyvinyl alcohol (PVA), polyurethane, poly(2-ethylacryllic acid), N-isopropylacrylamide polymers, or polyphosphazine. Examples of polyamines include: polyethylenimine, polylysine (PLL), spermine, spermidine, polyamine, pseudopeptide-polyamine, peptidomimetic polyamine, dendrimer polyamine,

arginine, amidine, protamine, cationic lipid, cationic porphyrin, quaternary salt of a polyamine, or an alpha helical peptide .

Ligands can also include targeting groups, *e.g.*, a cell or tissue targeting agent, *e.g.*, a lectin, glycoprotein, lipid or protein, *e.g.*, an antibody, that binds to a specified cell type. A targeting group can be a thyrotropin, melanotropin, lectin, glycoprotein, surfactant protein A, Mucin carbohydrate, multivalent lactose, multivalent galactose, N-acetyl-galactosamine, N-acetyl-gulucosamine multivalent mannose, multivalent fucose, glycosylated polyaminoacids, multivalent galactose, transferrin, bisphosphonate, polyglutamate, polyaspartate, a lipid, cholesterol, a steroid, bile acid, folate, vitamin B12, biotin, an RGD peptide, an RGD peptidomimetic or an aptamer.

Other examples of ligands include dyes, intercalating agents (*e.g.*, acridines), cross-linkers (*e.g.*, psoralene, mitomycin C), porphyrins (TPPC4, texaphyrin, Sapphyrin), polycyclic aromatic hydrocarbons (*e.g.*, phenazine, dihydrophenazine), artificial endonucleases or a chelator (*e.g.*, EDTA), lipophilic molecules, *e.g.*, cholesterol, cholic acid, adamantane acetic acid, 1-pyrene butyric acid, dihydrotestosterone, 1,3-Bis-O(hexadecyl)glycerol, geranyloxyhexyl group, hexadecylglycerol, borneol, menthol, 1,3-propanediol, heptadecyl group, palmitic acid, myristic acid,O3-(oleoyl)lithocholic acid, O3-(oleoyl)cholenic acid, dimethoxytrityl, or phenoxazine)and peptide conjugates (*e.g.*, antennapedia peptide, Tat peptide), alkylating agents, phosphate, amino, mercapto, PEG (*e.g.*, PEG-40K), MPEG, [MPEG]2, polyamino, alkyl, substituted alkyl, radiolabeled markers, enzymes, haptens (*e.g.*, biotin), transport/absorption facilitators (*e.g.*, aspirin, vitamin E, folic acid), synthetic ribonucleases (*e.g.*, imidazole, bisimidazole, histamine, imidazole clusters, acridine-imidazole conjugates, Eu3+ complexes of tetraazamacrocycles), dinitrophenyl, HRP, or AP.

Ligands can be proteins, *e.g.*, glycoproteins, or peptides, *e.g.*, molecules having a specific affinity for a co-ligand, or antibodies *e.g.*, an antibody, that binds to a specified cell type such as a cancer cell, endothelial cell, or bone cell. Ligands may also include hormones and hormone receptors. They can also include non-peptidic species, such as lipids, lectins, carbohydrates, vitamins, cofactors, multivalent lactose, multivalent galactose, N-acetyl-galactosamine, N-acetyl-gulucosamine multivalent mannose, multivalent fucose, or aptamers. The ligand can be, for example, a lipopolysaccharide, an activator of p38 MAP kinase, or an activator of NF-κB.

In some embodiments, the ligand is a lipid or lipid-based molecule. Such a lipid or lipid-based molecule preferably binds a serum protein, *e.g.*, human serum albumin (HSA).

An HSA binding ligand allows for distribution of the conjugate to a target tissue. A lipid or lipid-based ligand can (a) increase resistance to degradation of the conjugate, (b) increase targeting or transport into a target cell or cell membrane, and/or (c) can be used to adjust binding to a serum protein, *e.g.*, HSA. A lipid based ligand can be used to modulate, *e.g.*, control the binding of the conjugate to a target tissue.

In some embodiments, the ligand is a peptide or a peptidomimetic. A peptidomimetic is a molecule capable of folding into a defined three-dimensional structure similar to a natural peptide. The peptide or peptidomimetic moiety can be about 5-50 amino acids long, e.g., about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 amino acids long. A peptide or peptidomimetic can be, for example, a cell permeation peptide, cationic peptide, amphipathic peptide, or hydrophobic peptide (*e.g.*, consisting primarily of Tyr, Trp or Phe). The peptide moiety can be a dendrimer peptide, constrained peptide or crosslinked peptide. In another alternative, the peptide moiety can include a hydrophobic membrane translocation sequence (MTS). The peptide moiety can be a "delivery" peptide, which can carry large polar molecules including peptides, oligonucleotides, and protein across cell membranes. A peptide or peptidomimetic can be encoded by a random sequence of DNA, such as a peptide identified from a phage-display library, or one-bead-one-compound (OBOC) combinatorial library (Lam *et al.*, Nature, 354:82-84, 1991).

As used herein, a "peptide moiety" can range in length from about 5 amino acids to about 50 amino acids. The peptide moieties can have a structural modification, such as to increase stability or direct conformational properties. Any of the structural modifications described below can be utilized. An arginine-glycine-aspartic acid (RGD)-peptide moiety can be used to target a tumor cell, such as an endothelial tumor cell or a breast cancer tumor cell (Zitzmann *et al.*, Cancer Res., 62:5139-43, 2002). An RGD peptide can facilitate targeting of an oligonucleotide to tumors of a variety of other tissues, including the lung, kidney, spleen, or liver (Aoki *et al.,* Cancer Gene Therapy 8:783-787, 2001). The RGD peptide can be linear or cyclic, and can be modified, *e.g.*, glycosylated or methylated to facilitate targeting to specific tissues. Peptides that target markers enriched in proliferating cells can be used. For example, RGD containing peptides and peptidomimetics can target cancer cells, in particular cells that exhibit an integrin. Thus, the ligand may comprise RGD peptides, cyclic peptides containing RGD, RGD peptides that include D-amino acids, or synthetic RGD mimics.

Peptide and peptidomimetic ligands include those having naturally occurring or modified peptides, e.g., D or L peptides; α, β, or γ peptides; N-methyl peptides; azapeptides;

peptides having one or more amide, i.e., peptide, linkages replaced with one or more urea, thiourea, carbamate, or sulfonyl urea linkages; or cyclic peptides .

Ligands can be coupled to the oligonucleotide fragment(s) and/or oligonucleotide at various places, for example, 3'-end, 5'-end, and/or at an internal ("non-end") position. In some embodiments, the ligand is attached via an intervening tether, *e.g.*, a carrier described herein. The ligand or tethered ligand may be present on a monomer when the monomer is incorporated into the oligonucleotide fragment(s) and/or oligonucleotide. In some embodiments, the ligand may be incorporated via coupling to a "precursor" monomer after the "precursor" monomer has been incorporated into the oligonucleotide fragment and/or the oligonucleotide. For example, a monomer having, *e.g.*, an amino-terminated tether (*i.e.*, having no associated ligand), *e.g.*, TAP-(CH2)nNH2 may be incorporated into a growing oligonucleotide fragment. In a subsequent operation, *i.e.*, after incorporation of the precursor monomer into the oligonucleotide fragment, a ligand having an electrophilic group, *e.g.,* a pentafluorophenyl ester or aldehyde group, can subsequently be attached to the precursor monomer by coupling the electrophilic group of the ligand with the terminal nucleophilic group of the precursor monomer's tether.

In another example, a monomer having a chemical group suitable for taking part in Click Chemistry reaction may be incorporated, *e.g.,* an azide or alkyne terminated tether/linker. In a subsequent operation, *i.e.*, after incorporation of the precursor monomer into the oligonucleotide fragment(s) and/or the oligonucleotide, a ligand having complementary chemical group, *e.g.* an alkyne or azide can be attached to the precursor monomer by coupling the alkyne and the azide together.

In some embodiments, the ligand is conjugated to nucleobases, sugar moieties, or internucleosidic linkages of the oligonucleotide fragment(s) and/or oligonucleotide. Conjugation to purine nucleobases or derivatives thereof can occur at any position including, endocyclic and exocyclic atoms. In some embodiments, the 2-, 6-, 7-, or 8-positions of a purine nucleobase are attached to a conjugate moiety. Conjugation to pyrimidine nucleobases or derivatives thereof can also occur at any position. In some embodiments, the 2-, 5-, and 6-positions of a pyrimidine nucleobase can be substituted with a conjugate moiety. Conjugation to sugar moieties of nucleosides can occur at any carbon atom. Example carbon atoms of a sugar moiety that can be attached to a conjugate moiety include the 2', 3', and 5' carbon atoms. The 1' position can also be attached to a conjugate moiety, such as in an abasic residue. Internucleosidic linkages can also bear conjugate moieties. For phosphorus-containing linkages (*e.g.*, phosphodiester, phosphorothioate (*e.g.* chrial phosphorothioate),

phosphorodithiotate, phosphoroamidate, and the like), the conjugate moiety can be attached directly to the phosphorus atom or to an O, N, or S atom bound to the phosphorus atom. For amine- or amide-containing internucleosidic linkages (*e.g.*, PNA), the conjugate moiety can be attached to the nitrogen atom of the amine or amide or to an adjacent carbon atom.

In some embodiments, the ligand is conjugated to the sense strand. In some embodiments, the ligand is conjugated to the 3' end of the sense strand. In some embodiments, the ligand is conjugated to the 5' end of the sense strand. In some embodiments, the ligand is conjugated to a non-end of the sense strand.

In some embodiments, the ligand is conjugated to the antisense strand. In some embodiments, the ligand is conjugated to the 3' end of the antisense strand. In some embodiments, the ligand is conjugated to a non-end of the antisense strand.

The ligand may be attached via a carrier. The carriers include (i) at least one "backbone attachment point," preferably two "backbone attachment points" and (ii) at least one "tethering attachment point." A "backbone attachment point" as used herein refers to a functional group, *e.g.* a hydroxyl group, or generally, a bond available for, and that is suitable for incorporation of the carrier into the backbone, *e.g.*, the phosphate, or modified phosphate, *e.g.*, sulfur containing, backbone, of a nucleic acid. A "tethering attachment point" (TAP) in some embodiments refers to a constituent ring atom of the cyclic carrier, *e.g.*, a carbon atom or a heteroatom (distinct from an atom which provides a backbone attachment point), that connects a selected moiety. The moiety can be, *e.g.*, a carbohydrate, *e.g.* monosaccharide, disaccharide, trisaccharide, tetrasaccharide, oligosaccharide and polysaccharide. Optionally, the selected moiety is connected by an intervening tether to the cyclic carrier. Thus, the cyclic carrier will often include a functional group, *e.g.*, an amino group, or generally, provide a bond, that is suitable for incorporation or tethering of another chemical entity, *e.g.*, a ligand to the constituent ring.

Wherein the oligonucleotide fragment is a dsRNA, the sense and/or antisense strand may be conjugated to a ligand via a carrier, wherein the carrier can be cyclic group or acyclic group; preferably, the cyclic group is selected from pyrrolidinyl, pyrazolinyl, pyrazolidinyl, imidazolinyl, imidazolidinyl, piperidinyl, piperazinyl, [1,3]dioxolane, oxazolidinyl, isoxazolidinyl, morpholinyl, thiazolidinyl, isothiazolidinyl, quinoxalinyl, pyridazinonyl, tetrahydrofuryl and and decalin; preferably, the acyclic group is selected from serinol backbone or diethanolamine backbone.

In some embodiments, one or more oligonucleotide fragments comprise the sequence "TT", "dTdT", "dTsdT" or "UU" as a single-stranded overhang at the 3' end, also termed

herein a terminal dinucleotide or 3' terminal dinucleotide. dT is 2'-deoxy-thymidine-5'-phosphate and sdT is 2'-deoxy Thymidine 5'-phosphorothioate. Terminal dinucleotide "UU" is UU or 2'-OMe-U 2'-OMe-U, and the terminal TT and the terminal UU can be in the inverted/reverse orientation. The terminal dinucleotide (e.g., UU) is a modified variant of the dithymidine dinucleotide commonly placed as an overhang to protect the ends of siRNAs from nucleases (see, for example, Elbashir *et al.* 2001 Nature 411: 494-498; Elbashir *et al.* 2001 EMBO J. 20: 6877-6888; and Kraynack *et al.* 2006 RNA 12:163-176). A terminal dinucleotide is known from these references to enhance nuclease resistance but not contribute to target recognition.

In some embodiments, one or both terminal oligonucleotide fragments comprise a 3' end cap instead of or in addition to a terminal dinucleotide to stabilize the end from nuclease degradation provided that the 3' end cap is able to both stabilize the oligonucleotide (*e.g.*, against nucleases) and not interfere excessively with its desired activity.

Wherein the oligonucleotide fragment is a dsRNA, the sense and/or antisense strand may be conjugated to a ligand via a carrier, wherein the carrier can be cyclic group or acyclic group; preferably, the cyclic group is selected from pyrrolidinyl, pyrazolinyl, pyrazolidinyl, imidazolinyl, imidazolidinyl, piperidinyl, piperazinyl, [1,3]dioxolane, oxazolidinyl, isoxazolidinyl, morpholinyl, thiazolidinyl, isothiazolidinyl, quinoxalinyl, pyridazinonyl, tetrahydrofuryl and and decalin; preferably, the acyclic group is selected from serinol backbone or diethanolamine backbone.

Additional embodiments

Embodiment 1. An engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, and 600; wherein the engineered dsRNA ligase polypeptide:

(a) has dsRNA ligase activity; and

(b) does not the comprise the amino acid sequence of SEQ ID NO: 302.

Embodiment 2. An engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, and 668; wherein the engineered dsRNA ligase polypeptide:

(a) has dsRNA ligase activity; and

(b) does not the comprise the amino acid sequence of SEQ ID NO: 302.

Embodiment 3. The engineered dsRNA ligase polypeptide of Embodiment 1, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 370, 488, 526, 578, 588, 590, and 592.

Embodiment 4. The engineered dsRNA ligase polypeptide of Embodiment 2, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 370, 488, 526, 578, 588, 590, 592 and 666.

Embodiment 5. An engineered dsRNA ligase polypeptide comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 302, which produces at least 5% more oligonucleotide product than a dsRNA ligase polypeptide comprising the amino acid sequence of SEQ ID NO: 302 under the same ligation reaction conditions, wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

Embodiment 6. The engineered dsRNA ligase polypeptide of Embodiment 5, wherein the ligation reaction conditions include about 1 μM to about 10 mM oligonucleotide fragment, a source of ATP, about 5 mM to about 100 mM divalent cation, and about 0.5 g/L to about 10 g/L engineered dsRNA ligase polypeptide, pH of about 4.0 to about 8.0, and temperature of about 10 °C to about 50 °C.

Embodiment 7. The engineered dsRNA ligase polypeptide of Embodiment 5 or 6, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X6, X7, X15, X19, X29, X36, X39, X46, X47, X49, X51, X53, X56, X57, X60, X63, X64, X66, X67, X87, X88, X91, X93, X103, X105, X107, X114, X122, X126, X129, X130, X131, X137, X144, X146, X158, X163, X173, X178, X190, X196, X216, X218, X221, X228, X230, X232, X235, X236, X237, X238, X239, X242, X243, X244, X251, X252, X254, X255, X258, X269, X280, X284, X285, X293, X296, X301, X303, X305, X314, X325, and X328, wherein the numbering refers to SEQ ID NO: 302.

Embodiment 8. The engineered dsRNA ligase polypeptide of Embodiment 7, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X6 is G; X7 is Q; X15 is R, D or E; X19 is Q or D; X29 is N or L; X36 is V; X39 is A; X46 is Y; X47 is E; X49 is G; X51 is L; X53 is Y; X56 is R or A; X57 is S; X60 is T, G or P; X63 is S, Q or G; X64 is R, T, Q, F, G, or M; X66 is F or W; X67 is N; X87 is T, P, K or absent; X88 is C; X91 is S; X93 is G, C, or A; X103 is V, C, Y, or T; X105 is V; X107 is R or T; X114 is N; X122 is W; X126 is G; X129 is N; X130 is R, S or Y; X131 is R; X137 is V or C; X144 is N; X146 is R; X158 is W; X163 is G; X173 is L; X178 is R; X190 is Q; X196 is S or C; X216 is L or R; X218 is N; X221 is I; X228 is R; X230 is T; X232 is R; X235 is A, T, or G; X236 is S, L, or F; X237 is S, Q, or R; X238 is F; X239 is G or R; X242 is R or M; X243 is N, S, G, or M; X244 is G or K; X251 is D or L; X252 is V; X254 is K; X255 is C; X258 is V; X269 is L; X280 is W; X284 is A; X285 is A; X293 is R; X296 is R; X301 is G, L, E, or F; X303 is Q; X305 is G; X314 is A or V; X325 is R; and X328 is R; wherein the numbering refers to SEQ ID NO: 302.

Embodiment 9. The engineered dsRNA ligase polypeptide of any one of Embodiments 5-8, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises

an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X6, X7, X15, X19, X29, X36, X39, X44, X45, X46, X47, X49, X51, X53, X56, X57, X60, X63, X64, X66, X67, X87, X88, X89, X91, X92, X93, X103, X105, X107, X114, X122, X126, X129, X130, X131, X137, X144, X146, X158, X163, X173, X178, X185, X190, X196, X216, X218, X221, X228, X230, X232, X235, X236, X237, X238, X239, X242, X243, X244, X251, X252, X254, X255, X258, X269, X280, X284, X285, X293, X296, X301, X303, X305, X313, X314, X325, and X328, wherein the numbering refers to SEQ ID NO: 302.

Embodiment 10. The engineered dsRNA ligase polypeptide of Embodiment 9, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X6 is G or E; X7 is Q; X15 is R, D or E; X19 is Q or D; X29 is N or L; X36 is V; X39 is A; X44 is V; X45 is V; X46 is Y; X47 is E; X49 is G; X51 is L; X53 is Y; X56 is R or A; X57 is S; X60 is T, G or P; X63 is S, Q or G; X64 is R, T, Q, F, G, or M; X66 is F or W; X67 is N; X87 is T, P, K or absent; X88 is C; X89 is T; X91 is S; X92 is D; X93 is G, C, or A; X103 is V, C, Y, or T; X105 is V; X107 is R or T; X114 is N; X122 is W; X126 is G; X129 is N; X130 is R, S or Y; X131 is R; X137 is V or C; X144 is N; X146 is R; X158 is W; X163 is G; X173 is L; X178 is R; X185 is K; X190 is Q; X196 is S or C; X216 is L or R; X218 is N; X221 is I; X228 is R; X230 is T; X232 is R; X235 is A, T, or G; X236 is S, L, or F; X237 is S, Q, R, L or G; X238 is F; X239 is G or R; X242 is R or M; X243 is N, S, G, or M; X244 is G or K; X251 is D or L; X252 is V; X254 is K; X255 is C; X258 is V; X269 is L; X280 is W; X284 is A; X285 is A; X293 is R; X296 is R; X301 is G, L, E, or F; X303 is Q; X305 is G; X313 is A; X314 is A or V; X325 is R; and X328 is R; wherein the numbering refers to SEQ ID NO: 302.

Embodiment 11. The engineered dsRNA ligase polypeptide of any one of Embodiments 5-10, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X15, X19, X36, X39, X53, X218, X221, X237, X251, X255, and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity.

Embodiment 12. The engineered dsRNA ligase polypeptide of Embodiment 11, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more

of the following amino acid residues: X15 is D or E; X19 is D; X36 is V; X39 is A; X53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302.

Embodiment 13. The engineered dsRNA ligase polypeptide of any one of Embodiments 5-12, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X15, X19, X36, X39, X53, X185, X218, X221, X237, X251, X255, and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity.

Embodiment 14. The engineered dsRNA ligase polypeptide of Embodiment 13, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X15 is D or E; X19 is D; X36 is V; X39 is A; X53 is Y; X185 is K; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302.

Embodiment 15. The engineered dsRNA ligase polypeptide of any one of Embodiments 5-14, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X36, X39, X218 and X221, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X36 is V; X39 is A; X218 is N; and X221 is I.

Embodiment 16. The engineered dsRNA ligase polypeptide of any one of Embodiments 5-15, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X39, X218 and X221, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X39 is A; X218 is N; and X221 is I.

Embodiment 17. The engineered dsRNA ligase polypeptide of any one of Embodiments 5-16, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X39, X218, X221 and X255, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X39 is A; X218 is N; X221 is I; and X255 is C.

Embodiment 18. The engineered dsRNA ligase polypeptide of any one of Embodiments 5-17, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X39, X53, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

Embodiment 19. The engineered dsRNA ligase polypeptide of any one of Embodiments 5-18, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X15, X39, X53, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X15 is E; X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

Embodiment 20. The engineered dsRNA ligase polypeptide of any one of Embodiments 5-19, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X19, X39, X53, X218, X221, X237, X251, X255

and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X19 is D; X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

Embodiment 21. The engineered dsRNA ligase polypeptide of any one of Embodiments 5-20, wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X15, X39, X53, X185, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X15 is D; X39 is A; X 53 is Y; X185 is K; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

Embodiment 22. The engineered dsRNA ligase polypeptide of any of Embodiments 1-21, wherein the engineered dsRNA ligase polypeptide comprises a purification tag.

Embodiment 23. The engineered dsRNA ligase polypeptide of Embodiment 22, comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298 and 300.

Embodiment 24. The engineered dsRNA ligase polypeptide of Embodiment 22, comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104,

106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, and 634.

Embodiment 25. A polypeptide immobilized on a solid material by chemical bond or a physical adsorption method, wherein the polypeptide comprises an engineered dsRNA ligase polypeptide according to any one of Embodiments 1-24.

Embodiment 26. A polynucleotide encoding the engineered dsRNA ligase polypeptide of any one of Embodiments 1-24.

Embodiment 27. The polynucleotide of Embodiment 26, wherein the polynucleotide sequence is SEQ ID NO: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, or 599.

Embodiment 28. The polynucleotide of Embodiment 26, wherein the polynucleotide sequence is SEQ ID NO: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, and 667.

Embodiment 29. An expression vector comprising the polynucleotide according to any one of Embodiments 26-28.

Embodiment 30. The expression vector of Embodiment 29, which comprises a plasmid, a cosmid, a bacteriophage or a viral vector.

Embodiment 31. A host cell comprising the polynucleotide of any one of Embodiments 26-28 or the expression vector of Embodiment 29 or 30, optionally wherein the host cell is *E. coli*.

Embodiment 32. A method of preparing an engineered dsRNA ligase polypeptide, which comprises the steps of culturing the host cell according to Embodiment 31 and obtaining an engineered dsRNA ligase polypeptide from the culture.

Embodiment 33. An engineered dsRNA ligase catalyst obtainable by culturing the host cells according to Embodiment 31, or according to the method of Embodiment 32, wherein said engineered dsRNA ligase catalyst comprises cells or culture fluid containing the engineered dsRNA ligase polypeptides, or an article processed therewith, wherein the article refers to an extract obtained from the culture of host cell, an isolated product obtained by isolating or purifying an engineered dsRNA ligase from the extract, or an immobilized product obtained by immobilizing host cell, an extract thereof, or isolated product of the extract.

Embodiment 34. A method of producing an oligonucleotide from two or more oligonucleotide fragments, wherein the method comprises contacting:

(i) two or more oligonucleotide fragments;

(ii) an engineered dsRNA ligase polypeptide according to any one of Embodiments 1-24;

(iii) a source of ATP; and

(iv) a divalent cation;

to obtain an oligonucleotide.

Embodiment 35. The method of Embodiment 34, wherein the source of ATP comprises ATP.

Embodiment 36. The method of Embodiment 34 or 35, wherein the source of ATP comprises:

(a) polyphosphate kinase (PPK);

(b) polyphosphate; and

(c) AMP and/or ATP.

Embodiment 37. The method of Embodiment 36, wherein the PPK is selected from PPK12 or ajPAP.

Embodiment 38. The method of any one of Embodiments 36 or 37, wherein the method is performed using a sub-stoichiometric concentration of AMP and/or ATP.

Embodiment 39. The method of any one of Embodiments 36-38, wherein the polyphosphate is a polyphosphate salt.

Embodiment 40. The method of Embodiment 39, wherein the polyphosphate salt is sodium polyphosphate (Maddrell's salt) or sodium hexametaphosphate (Graham's salt).

Embodiment 41. The method of any one of Embodiments 34-40, wherein the divalent cation cofactor is Mg^{2+} or Mn^{2+}.

Embodiment 42. The method of any one of Embodiments 34-41, wherein the method is performed with a divalent cation concentration of 5-100 mM, optionally 30-50 mM.

Embodiment 43. The method of any one of Embodiments 34-42, further comprising a step of purifying the oligonucleotide.

Embodiment 44. Use of the engineered dsRNA ligase polypeptide according to any one of Embodiments 1-24 in the production of an oligonucleotide from two or more oligonucleotide fragments.

Embodiment 45. The method of any one of Embodiments 34-43 or the use of Embodiment 44, wherein the oligonucleotide is up to 60 nucleotides in length.

Embodiment 46. The method of any one of Embodiments 34-43 or 45 or the use of Embodiment 44 or 45, wherein each of the oligonucleotide fragments are 4-16 nucleotides in length, optionally 6-9 nucleotides in length.

Embodiment 47. The method of Embodiment 34-43, 45 or 46 or the use of any one of Embodiments 44-46, wherein one or more of the oligonucleotide fragment(s) comprises one or two overhangs.

Embodiment 48. The method of any one of Embodiments 34-43 or 45-47 or the use of any one of Embodiments 44-47, wherein one or more of the oligonucleotide fragments comprises a chemical modification.

Embodiment 49. The method or use of Embodiment 48, wherein the chemical modification is selected from:

(a) a modified backbone, optionally selected from a phosphorothioate (*e.g.* chiral phosphorothioate) or methylphosphonate internucleotide linkage;

(b) a modified nucleotide, optionally selected from 2'-O-methyl (2'-OMe), 2'-flouro (2'-F), 2'-deoxy, 2'-deoxy-2'-fluoro, 2'-O-methoxyethyl (2'-O-MOE), 2'-O-aminopropyl (2'-O-AP), 2'-O-dimethylaminoethyl (2'-O-DMAOE), 2'-O-dimethylaminopropyl (2'-O-DMAP), 2'-O-dimethylaminoethyloxyethyl (2'-O-DMAEOE), 2'-O-N-methylacetamido (2'-O-NMA), locked nucleic acid (LNA), glycol nucleic acid (GNA), phosphoramidate (*e.g.* mesyl phosphoramidate), 2',3'-seco nucleotide mimic, 2'-F-arabino nucleotide, abasic nucleotide, 2'-amino modified nucleotide, 2'-alkyl-modified nucleotide, morpholino nucleotide, vinylphosphonate (*e.g.* 5' vinylphosphonate), and cyclopropyl phosphonate deoxyribonucleotide; and/or

(c) conjugation to a ligand, optionally wherein the ligand comprises one or more N-Acetylgalactosamine (GalNAc) derivatives.

Embodiment 50. A composition comprising:

i. the engineered dsRNA ligase polypeptide according to any one of Embodiments 1-24;

ii. a source of ATP; and

iii. a divalent cation.

Embodiment 51. The composition of Embodiment 50, further comprising two or more oligonucleotide fragments.

Embodiment 52. A kit comprising:

i. the engineered dsRNA ligase polypeptide according to any one of Embodiments 1-24;

ii. a source of ATP;

iii. a divalent cation; and

iv. instructions for use in a method of producing an oligonucleotide from two or more oligonucleotide fragments.

Embodiment 53. The composition of Embodiment 50 or 51 or the kit of Embodiment 52, wherein the source of ATP comprises ATP.

Embodiment 54. The composition of any one of Embodiments 50, 51 or 53 or the kit of Embodiment 52 or 53, wherein the source of ATP comprises:

(a) polyphosphate kinase (PPK);

(b) polyphosphate; and

(c) AMP and/or ATP.

Embodiment 55. The composition or kit of Embodiment 54, wherein the PPK is selected from PPK12 or ajPAP.

Embodiment 56. The composition of any one of Embodiments 50, 51 or 53-55 or the kit of any one of Embodiments 52-55, wherein the polyphosphate is a polyphosphate salt.

Embodiment 57. The composition or kit of Embodiment 56, wherein the polyphosphate salt is sodium polyphosphate (Maddrell's salt) or sodium hexametaphosphate (Graham's salt).

Embodiment 58. The composition of any one of Embodiments 50, 51 or 53-57 or the kit of any one of Embodiments 52-57, wherein the divalent cation cofactor is Mg^{2+} or Mn^{2+}.

Different features and embodiments of the present disclosure are exemplified in the following representative examples, which are intended to be illustrative and not restrictive.

EXAMPLES

The following Examples, including experiments and results achieved, are provided for illustrative purposes only and are not to be construed as limiting the present invention.

In the Examples below, the following abbreviations apply: ppm (parts per million); M (molar); mM (millimolar), uM and μM (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μg (micrograms); L and l (liter); ml and mL (milliliter); cm (centimeters); mm (millimeters); um and μm (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); psi and PSI (pounds per square inch); °C (degrees Centigrade); RT and rt (room temperature); OD600

(Optical density at 600 nm), CAM and cam (chloramphenicol); DMSO (dimethylsulfoxide); FP (Fermentation powder); FWC (Frozen whole cells), LWC (Lyophilized whole cells), PMBS (polymyxin B sulfate); IPTG (isopropyl β-D-1-thiogalactopyranoside); LB (Lysogeny broth); TB (Terrific Broth; 12 g/L bacto-tryptone, 24 g/L yeast extract, 4 mL/L glycerol, 65 mM potassium phosphate, pH 7.0, 1 mM $MgSO_4$); TEoA (triethanolamine buffer), HEPES (HEPES zwitterionic buffer; 4-(2-hydroxyethyl)-piperazineethanesulfonic acid); SFP (shake flask powder); CDS (coding sequence); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); *E. coli* W3110 (commonly used laboratory *E. coli* strain, available from the Coli Genetic Stock Center [CGSC], New Haven, CT); HTP (high throughput); HPLC (high pressure liquid chromatography); FIOP (fold improvements over positive control); Microfluidics (Microfluidics, Corp., Westwood, MA); Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO; Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); Agilent (Agilent Technologies, Inc., Santa Clara, CA); Corning (Corning, Inc., Palo Alto, CA); Dow Corning (Dow Corning, Corp., Midland, MI); and Gene Oracle (Gene Oracle, Inc., Mountain View, CA).

The sequences of the oligonucleotides referred to in parentheses (*e.g.* "siRNA (1)" and "oligonucleotide (2)") throughout the Examples are provided in Table 1.

EXAMPLE 1

Preparation of isolated enzymes

Polynucleotides encoding the polypeptides having ligase activity, were cloned into the pCK110900 vector system (See *e.g.*, US Pat. App. No. 2006/0195947A1 FIG. 3 which is hereby incorporated by reference in its entirety) and subsequently expressed in *E. coli* W3110*fhu*A under the control of the *lac* promoter. The expression vector also contained the P15a origin of replication and the chloramphenicol (CAM) resistance gene.

E. coli W3110*fhu*A cells were transformed with the pCK110900 plasmid containing the ligase-encoding genes. Transformed cells were plated out on Lysogeny broth (LB) agar plates containing 1% glucose and 30 μg/mL CAM, and grown overnight at 37° C. Subsequently single colonies were inoculated into 25 mL of LB supplemented with 30 μg/mL CAM and 1% glucose in a 250 ml baffled shake flask. The culture was grown overnight (16-20 hours and optical density (OD_{600}) >3.8) in an incubator at 37°C, with shaking at 250 rpm. A 1 L shake flask containing 250 mL of Terrific Broth (TB) media with 30 μg/mL CAM, was inoculated with 5 mL of the grown overnight culture. The 250 mL culture was incubated at 30°C, 250 rpm, for 3 - 3.5 hours until OD_{600} reached 0.6–0.8. Expression of the ligase gene was induced

by the addition of isopropyl-β-D-thiogalactoside (IPTG) to a final concentration of 1 mM, and growth was continued for an additional 18-20 hours. Cells were harvested by transferring the culture into a centrifuge bottle, which was then centrifuged at 7,000 rpm for 5 minutes at 4°C. The supernatant was discarded, and the remaining cell pellet lysed. For lysis, the cell pellet was resuspended in 30 mL of 50 mM Tris-buffer at pH 7.5 and lysed using a LM20 MICROFLUIDIZER® processor system (Microfluidics). Cell debris was removed by centrifugation at 14,000 rpm for 30 minutes at 4°C. Ligase enzymes were then isolated from the clarified lysate using standard techniques known in the art, including immobilized metal affinity chromatography.

EXAMPLE 2

Identification of dsRNA ligase activity for the production of siRNA (1)

To identify an enzyme with dsRNA ligase activity for the production of siRNA (1) comprising of oligonucleotides (2) and (3) a collection of ligases were first screened for the production of a surrogate product, siRNA (4) comprising of oligonucleotides (3) and (5). Oligonucleotide (5) has the same sequence as oligonucleotide (2) but does not contain a 3'-GalNAc moiety. siRNAs (1) and (4) and oligonucleotides (2), (3) and (5) are depicted in Figure 1; and the sequences of oligonucleotides (2), (3) and (5) are provided in Table 1.

Screening of the isolated ligases was performed in 20 µL reaction volumes in PCR tubes, each tube containing 50 mM Tris-buffer pH 7.5, either 1 mM ATP or 1 mM NAD+, 10 mM MgCl2, 5 mM DTT and 10 µM (each) of substrate oligonucleotides (6 – 11) with 50 % (v/v) isolated ligase enzyme. Reactions were incubated in a thermocycler at 16 °C for 2 h and analyzed using standard techniques known in the art, including electrophoresis. A ligase with SEQ ID NO: 2 exhibited the highest dsRNA ligase activity towards the formation of siRNA (4). The activity of SEQ ID NO: 2 for the production of siRNA (1) was subsequently confirmed using multiple enzyme preparations including isolated enzyme (example 1), clarified lysate (example 4) and shake flask powder (SFP; example 5).

Table 1 Oligonucleotide sequences

Oligonucleotide NO:	Oligonucleotide sequence 5' to 3'
2	mC*mU*mAmGmAmCfCmUfGmUTmUmUmGmCmUmUmUmUmGmU-GalNAc **(SEQ ID NO: 673)**
3	mA*fC*mAfAfAfAmGfCmAfAmAfAmCfAmGfGmUfCmUmAmG*mA*mA **(SEQ ID NO: 674)**
5	mC*mU*mAmGmAmCfCmUfGmUTmUmUmGmCmUmUmUmUmGmU **(SEQ ID NO: 675)**
6	mC*mU*mAmGmAmCfCmU
7	pfGmUTmUmUmGmC
8	pmUmUmUmUmGmU
9	mA*fC*mAfAfAfAmGfCmA
10	pfAmAfAmCfAmGfG
11	pmUfCmUmAmG*mA*mA
12	pmUmUmUmUmGmU-GalNAc
13	mC*mU*mAmGmAmCfCmUfGmUTmUmUmGmC **(SEQ ID NO: 676)**
14	pfGmUTmUmUmGmCmUmUmUmUmGmU-GalNAc **(SEQ ID NO: 677)**
15	mA*fC*mAfAfAfAmGfCmAfAmAfAmCfAmGfG **(SEQ ID NO: 678)**
16	pfAmAfAmCfAmGfGmUfCmUmAmG*mA*mA **(SEQ ID NO: 679)**
mA = 2'-OMe adenosine; mC = 2'-OMe cytidine; mG = 2'-OMe guanosine; mU = 2'-OMe uridine; fA = 2'-fluoro adenosone; fC = 2'-fluoro cytidine; fG = 2'-fluoro guanosine; T = 2'-H thymadine * = phosphorothioate bond; p = 5' phosphate; GalNAc = *N*-acetylgalactosamine	

EXAMPLE 3

Preparation of cell pellets for high throughput (HTP) screening

Single colonies were picked in a 96-well format and grown in 190 µL LB media containing 1% glucose and 30 µg/mL CAM, at 30°C, 200 rpm, and 85% humidity. Following overnight growth, 20 µL of the grown cultures were transferred into a deep well plate containing 380 µL of TB media with 30 µg/mL CAM. The cultures were grown at 30°C, 250 rpm, with 85% humidity for approximately 2.5 hours. When the OD_{600} of the cultures reached 0.4–0.8, expression of the ligase gene was induced by the addition of IPTG to a final concentration of 1 mM. Following induction, growth continued for 18–20 hours at 30°C, 250 rpm with 85% humidity. Cells were harvested by centrifugation at 4,000 rpm and 4°C for 10 minutes; the supernatant was then discarded. The cell pellets were stored at -80°C until ready for use.

EXAMPLE 4

Lysis and preparation of clarified lysate

Prior to performing the assay, the cell pellets were thawed and resuspended in 300 µL of lysis buffer (containing 1 g/L lysozyme, 0.5 g/L PMBS and 0.1 µL/mL or 0.2U/ml of commercial DNAse (New England BioLabs, M0303L) in 50 mM Tris-buffer at pH 7.5. The

plates were agitated with medium-speed shaking for 2.5 hours on a microtiter plate shaker at room temperature. The plates were then centrifuged at 4,000 rpm for 10 minutes at 4°C, and the clarified supernatants were used in the HTP assay reaction for activity determination as described in the following examples.

EXAMPLE 5

Preparation of shake flask powder (SFP) and fermentation powder (FP)

Shake-flask procedures can be used to generate engineered dsRNA ligase polypeptide shake-flask powders (SFP), which are useful for secondary screening assays and/or use in the biocatalytic processes described herein. Shake flask powder preparation of enzymes provides a more concentrated preparation of the engineered enzyme, as compared to the cell lysate used in HTP assays.

Clarified lysate produced according to example 1 was collected, frozen at -80°C, and then lyophilized, using standard methods known in the art. Lyophilization of frozen clarified lysate provides a dry SFP comprising crude wild-type or engineered dsRNA ligase polypeptide.

EXAMPLE 6

Analytical method for activity and selectivity evaluation

Activity improvements of the engineered dsRNA ligases were analyzed by High Pressure Liquid Chromatography (HPLC) using the methods described in Table 6-1 and 6-2. HPLC methods with UV-detection were developed to analyze the formation of product oligonucleotides (2) and (3). The analytical methods aim for the shortest run time enabling good resolution of the product oligonucleotides (2) and (3). Consequently, the six substrate (6-7, 9-12) and the four intermediate (13-16) oligonucleotides could not all be well resolved from each other. However, it is possible to resolve the well-defined GalNAc-containing oligonucleotides, including the substrate oligonucleotide (12), a reaction intermediate oligonucleotide (14) and the product oligonucleotide (2). Therefore, a pseudo-% conversion can be calculated, denoted with arbitrary units (AU), which considers only these well resolved species according to the following equation:

$$AU = \frac{\left(\frac{A260\ (2)}{\varepsilon_{(2)}}\right)}{\left(\frac{A260\ (2)}{\varepsilon_{(2)}}\right) + \left(\frac{A260\ (12)}{\varepsilon_{(12)}}\right) + \left(\frac{A260\ (14)}{\varepsilon_{(14)}}\right)}$$

Whereby $\varepsilon_{(2)}$, $\varepsilon_{(12)}$ and $\varepsilon_{(14)}$ are the extinction coefficient of oligonucleotides (2), (12) and (14) respectively. Using such a calculation an AU = 1.0 would imply that no more GalNAc-containing substrate or intermediate oligonucleotides are present in the reaction and that they have all be converted to GalNAc-containing product (2). In reality, for samples where AU = 1.0 the only other peak present in the chromatogram corresponds with the product oligonucleotide (2), and no other intermediates or starting materials can be identified. Furthermore, the ratio of the product oligonucleotides (2) and (3) are consistent with that of the authentic standard of siRNA product (1). Taken together, it can be concluded that AU = 1.0 is an approximation that is essentially equivalent to 100 % conversion.

Table 6-1: HPLC method 1 used for activity determination.

Instrument	Thermo Scientific UHPLC Vanquish Horizon		
Column	Waters® Acquity UHPLC BEH C18 Length 50 mm, internal diameter 2.1 mm or equivalent column, particle size: 1.7 µm.		
Mobile Phase	A: 200 mM 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) and 10 mM triethylamine (TEA) B: Methanol		
Gradient	Time (min.)	Phase B (%)	Flowrate (mL/min)
	0	15	0.5
	0.5	15	0.5
	2.65	24	0.5
	2.8	90	0.5
	3.2	90	0.5
	3.3	15	0.5
	4.0	15	0.5
Run Time	4.0 minutes		
Compound elution	Oligonucleotide (2) RT = 2.6 min; oligonucleotide (3) RT = 2.1 min		
Column Temperature	75 °C		
Autosampler Temperature	15 °C		
Injection volume	5 µL		
Detection	UV 260 nm		

HPLC method 2 (Table 6-2) was developed from HPLC method 1 (Table 6-1) to improve the separation between the product oligonucleotide (3) and the substrate oligonucleotide (12).

Table 6-2: HPLC method 1 used for activity determination.

Instrument	Thermo Scientific UHPLC Vanquish Horizon		
Column	Waters® Acquity UHPLC BEH C18 Length 50 mm, internal diameter 2.1 mm or equivalent column, particle size: 1.7 µm.		
Mobile Phase	A: 200 mM 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) and 10 mM triethylamine (TEA) B: Methanol		
Gradient	Time (min.)	Phase B (%)	Flowrate (mL/min)
	0	16	0.5
	0.5	16	0.5
	3.5	23	0.5
	3.6	90	0.5
	3.8	90	0.5
	4.0	16	0.5
	4.7	16	0.5
Run Time	4.70 minutes		
Compound elution	Oligonucleotide (2) RT = 3.1 min; oligonucleotide (3) RT = 2.4 min		
Column Temperature	75 °C		
Autosampler Temperature	15 °C		
Injection volume	5 µL		
Detection	UV 260 nm		

The activity improvements of the engineered dsRNA ligases of Example 12 were also analyzed with RapidFire® Mass Spectrometry (RF-MS) using the method described in Table 6-3. RF-MS aims to reduce the analytical time compared to HPLC analysis. The selective detection of product oligonucleotides (2) and (3) is obtained by the specific masses of each product oligonucleotide analyzed under the multi-single ion monitoring (SIM) mode. Relative

dsRNA ligase activity is determined by comparing the sum of the MS signal of the five specific masses (given in Table 6-3) corresponding with each product oligonucleotide (2) and (3).

Table 6-3: RF-MS method used for activity determination.

Instrument	Agilent RapidFire 365 high throughput with a 6470 triple quadrupole mass spectrometer (MS-TQ)
Cartridge	RapidFire Cartridge, PLRP-S 30µm/1000Å (4PL1K, Agilent) Size : 4µL
Mobile Phase	A: Water / Acetonitrile / Triethyl amine (TEA) (95:5:0.10/ %) B: Water / Acetonitrile / TEA (5:95:0.10/ %) C: Water / Acetonitrile / TEA (5:95:0.10/ %)
MS source parameters	Gas Temperature: 325 °C Gas flow: 10 L/min Nebulizer: 28 psi Sheath Gas Temperature: 325 °C Sheath Gas Flow: 10 L/min Capillary: 4000 V (NEG) Nozzle Voltage: 1000 V (NEG)
MS detection parameters	Multi-SIM mod Polarity: Negative Masses for product oligonucleotide 2: 1233.3 / 1079 / 784.5 / 719 / 616.3 Masses for product oligonucleotide 3: 1098.3 / 854 / 768.5 / 698.5 / 591.3 Dwell Time: 20ms; Fragmentor: 110 and Cell Accelerator Voltage: 5

The methods provided herein find use in analyzing the variants produced using the present invention. However, it is not intended that present invention be limited to the methods described herein, as there are other suitable methods known in the art that are applicable to the analysis of the variants provided herein and/or produced using the methods provided herein.

EXAMPLE 7

Round 1 Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 2 for Improved Production of siRNA product (1)

The engineered polynucleotide (SEQ ID NO: 1) encoding the polypeptide with dsRNA ligase activity of SEQ ID NO: 2 was used to generate the engineered polypeptides of Table 7-1. These polypeptides displayed improved dsRNA ligase activity under the desired conditions *e.g.*, the improvement in the formation of either oligonucleotide products (2) or (3), or preferably both oligonucleotide products (2) and (3) that was produced *in situ* from the substrate oligonucleotides (6-7, 9-12) as compared to the starting polypeptide. Some polypeptides displayed improved product formation of either oligonucleotide product (2) or (3), or both oligonucleotide products (2) and (3) compared to the starting polypeptide as noted in Table 7-1. The sequences of oligonucleotides (2), (3), (6), (7) and (9-12) are provided in Table 1.

The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 2, as described below together with the analytical method described in Table 6-1. Directed evolution began with the polynucleotide set forth in SEQ ID NO: 1. Libraries of engineered polypeptides were generated using various well-known techniques (*e.g.*, saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods, described below, that measured the polypeptides' ability to produce oligonucleotide products (2) and (3).

The enzyme assays were carried out in 96-well PCR plates, in 50 µL total reaction volume per well. The reactions contained 2.5 % (v/v) of undiluted dsRNA ligase lysate, prepared as described in Example 4, 100 µM (each) substrate oligonucleotides (6-7, 9-12), 50 mM Tris-buffer at pH 7.5, 1 mM ATP, 10 mM $MgCl_2$ and 5 mM DTT. The reaction plates were heat-sealed and incubated in a thermocycler at 30 °C for 2 h.

After incubation the plates were subjected to a heat inactivation step (95 °C, 20 min) to quench the reaction and precipitate proteinaceous content of the added lysate. The plates were then centrifuged at 4,000 rpm for 5 min. Subsequently a 2 µL aliquot of the supernatant was removed from each well and added to a shallow well 96-well plate containing 98 µL of 5 mM EDTA solution (pH 7.0). The samples were analyzed via HPLC to determine the activity of the enzyme variants using the analytical method described in Table 6-1. Selected ligase variants showing greater product formation of oligonucleotides (2) and (3) relative to SEQ ID NO:2 are shown in Table 7-1.

Table 7-1. dsRNA Ligase Activity for the Production of siRNA (1) Relative to SEQ ID NO: 2					
Engineered ligase	**SEQ ID NO: (nt/aa)**	**Amino Acid Differences (Relative to SEQ ID NO: 2)[1]**	**FIOP AU[1]**	**FIOP (3)[1]**	**FIOP (2)[1]**
1	3/4	K251S	+	++	+
2	5/6	K121R	-	+	++
3	7/8	K251Q	-	++	++
4	9/10	V257N	-	++	++
5	11/12	S21Q	+	++	++
6	13/14	P252F	++	+	++
7	15/16	P258G	-	+++	+
8	17/18	G71S	-	++	+
9	19/20	N230L	+	+	++
10	21/22	Q256R	+	+	++
11	23/24	E77S	+++	++	++
12	25/26	G117V	+	++	++
13	27/28	T70R	+	+++	++
14	29/30	N230R	++	++	+
15	31/32	L74T	-	++	++
16	33/34	S20G	+	++	+
17	35/36	L74G	+	++	+
18	37/38	D78R	+++	+	++
19	39/40	L74P	-	++	+
20	41/42	E77Q	+++	++	+++
21	43/44	T53A/G81N/I119V/K121T	-	+++	+
22	45/46	G117C	++	++	++
23	47/48	I50V/T53A/A235I	++	+++	++
24	49/50	V257S	++	++	+
25	51/52	G117Y	++	++	++
26	53/54	G117T	++	++	+
27	55/56	S249A	++	++	+
28	57/58	D250S	+	+	++
29	59/60	T53A/K128N	+++	+++	+++
30	61/62	I50V/T53A	+++	+++	++
31	63/64	K33Q/T53A/A235I	+++	+++	++
32	65/66	K29R/I50V/T53A/Y80F	++	+++	+
33	67/68	T232N/A235I	+++	+++	+++
34	69/70	I50V/T53A/T232N/A235I	+	+++	+
35	71/72	I253G	+	++	++
36	73/74	T53A	+++	+++	++
37	75/76	D78T	+	++	+

38	77/78	I119V/T232N/A235I	+++	+++	++
39	79/80	T232N	+++	+++	++
40	81/82	D78Q	+++	++	+++
41	83/84	K251R	+	++	+
42	85/86	D78F	+	+++	++
43	87/88	D78G	+++	++	+++
44	89/90	E77G	-	++	+
45	91/92	D250L	+	++	++
46	93/94	S249T	++	+	+++
47	95/96	S244T	+	++	++
48	97/98	D78M	+	++	+
49	99/100	K246R	+	++	+
50	101/102	I253R	+	+	+++
51	103/104	V257G	-	+++	+
52	105/106	D250F	+	+	++

[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 2 and defined as follows:

"-" < 1.00, "+" > 1.00, "++" > 1.20, "+++" > 1.50

FIOP = fold improvement over positive control

The engineered dsRNA ligase polypeptides represented by the even numbered sequence identifiers of SEQ ID NOs: 4 to 106 comprise an even numbered sequence identifier of SEQ ID NOs: 304 to 406, respectively, and a 14 amino acid N-terminal purification tag (MHHHHHHENLYFQS (SEQ ID NO: 669)). For example, SEQ ID NO: 4 comprises: (i) the 14 amino acid N-terminal purification tag of SEQ ID NO: 669; and (ii) the dsRNA ligase polypeptide of SEQ ID NO: 304.

Throughout the Examples, the position of a given mutation is provided relative to SEQ ID NO: 2 which includes (i) the 14 amino acid N-terminal purification tag of SEQ ID NO: 669 and (ii) the wild-type dsRNA ligase polypeptide of SEQ ID NO: 302. The position of a given mutation relative to SEQ ID NO: 302 (*i.e.* the wild-type dsRNA ligase polypeptide without the purification tag) can be obtained by subtracting the 14 amino acid N-terminal purification tag from the SEQ ID NOs described in the Examples. For example, position X251 of SEQ ID NO: 2 corresponds to position X237 of SEQ ID NO: 302.

EXAMPLE 8

Round 2 Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 70 for Improved Production of siRNA product (1)

The polynucleotide from example 7 SEQ ID NO: 69 encoding the most active polypeptide with dsRNA ligase activity of SEQ ID NO: 70 was used to generate the engineered polypeptides of Table 8-1. These polypeptides displayed improved dsRNA ligase activity under the desired conditions *e.g.*, the improvement in the formation of either oligonucleotide products (2) or (3), or preferably both oligonucleotide products (2) and (3) that was produced *in situ* from the substrates oligonucleotides (6-7, 9-12) as compared to the starting polypeptide. Some polypeptides displayed improved product formation of both oligonucleotide products (2) and (3) compared to the starting polypeptide are noted in Table 8-1. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 70, as described below together with the analytical method described in Table 6-1.

Directed evolution began with the polynucleotide set forth in SEQ ID NO: 69. Libraries of engineered polypeptides were generated using various well-known techniques (*e.g.*, saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods, described below, that measured the polypeptides' ability to produce oligonucleotides (2) and (3).

The enzyme assays were carried out in 96-well PCR plates, in 50 µL total reaction volume per well. The reactions contained either 1.25 or 2.5 % (v/v) of undiluted dsRNA ligase lysate, prepared as described in Example 4, 100 µM (each) substrate oligonucleotides (6-7, 9-12), 50 mM Tris-buffer at pH 7.5, 1 mM ATP, 10 mM $MgCl_2$ and 5 mM DTT. The reaction plates were heat-sealed and incubated in a thermocycler at 30 °C for 2 h.

After incubation the plates were subjected to a heat inactivation step (95 °C, 20 min) to quench the reaction and precipitate proteinaceous content of the added lysate. The plates were then centrifuged at 4,000 rpm for 5 min. Subsequently a 2 µL aliquot of the supernatant was removed from each well and added to a shallow well 96-well plate containing 98 µL of 5 mM EDTA solution (pH 7.0). The samples were analyzed via HPLC to determine the activity of the enzyme variants using the analytical method described in Table 6-1. Selected ligase variants showing a faster product formation of oligonucleotides (2) and (3) relative to SEQ ID NO:70 are shown in Table 8-1.

Table 8-1. dsRNA Ligase Activity for the Production of siRNA (1) Relative to SEQ ID NO: 70					
Engineered ligase	**SEQ ID NO: (nt/aa)**	**Amino Acid Differences (Relative to SEQ ID NO: 70)[1]**	**FIOP AU[1]**	**FIOP (3)[1]**	**FIOP (2)[1]**
53	107/108	E307R	+	+	+
54	109/110	V50I/T70R/L74P/D78Q	+	+++	+
55	111/112	V315G	+	+	+
56	113/114	V315L	+	++	+
57	115/116	T102C	+++	+++	+
58	117/118	V43N	+	++	+
59	119/120	V315E	+	++	+
60	121/122	P339R	++	++	+
61	123/124	C272V	++	++	+
62	125/126	G117V/K121R/Q256M/P258K	+++	++	+++
63	127/128	V50I/L74T/D78T/S249G/D250F	+++	++	+++
64	129/130	V50I/T70R/D78T/Q256M/V257M/P258K	++	+++	++
65	131/132	Q256M	+++	+++	++
66	133/134	G117V/N232T	+++	+	+++
67	135/136	V50I/L74T/D78Q	+	+++	+++
68	137/138	V50I/L74T/D78Q/K121R	+	+	++
69	139/140	Q256M/P258G	+	+++	++
70	141/142	K121R/Q256M/V257G/P258K	+	++	++
71	143/144	D78G	+	++	+
72	145/146	T70R/D78Q/N232T	+	++	+
73	147/148	V50I/V257M	++	++	+
74	149/150	V50I/T70R/L74T/D78T	+	+++	+
75	151/152	G117V/D250S	+++	+	+++
76	153/154	V50I/L74T/D78Q/K121R/N232T/V257M/P258K	++	+++	+++
77	155/156	S21Q	++	++	++
78	157/158	D78T/G117V	+++	++	+++
79	159/160	V43L	++	++	+
80	161/162	V50I/K251R	+++	+++	+++
81	163/164	V50I/K251Q/I253G/V257M	++	++	++
82	165/166	A187L	+	+++	++
83	167/168	V50I/T70R/L74T/G117V/K246R/S249G/D250S	+++	+++	+++
84	169/170	D78T	+	++	+
85	171/172	G117Y	+++	+	+++
86	173/174	V50I	+++	+++	+++
87	175/176	V315F	+	++	++

88	177/178	V50I/T70R/N230R/N232T	+	+++	+
89	179/180	I210S	++	+++	+++
90	181/182	D268K	++	++	++
91	183/184	G117V	++	+	++
92	185/186	T317Q	+	+++	++
93	187/188	V50I	+++	+++	+++
94	189/190	V50I/N230R	+++	++	++
95	191/192	V50I/L74T/N232T/P258K	+	+++	++
96	193/194	S310R	+	+	+
97	195/196	V50I/N232T/K251R/V257M	++	++	++
98	197/198	E151V	+	+	+
99	199/200	V50I/T70R/E77G/D78Q/G117Y/N232T	+++	+	+
100	201/202	I342R	+	++	+
101	203/204	V204Q	+	+	+
102	205/206	V50I/L74P/D78Q	+++	+++	++
103	207/208	T70A/D143N	+	+++	++
104	209/210	I210C	+	++	++
105	211/212	V50I/T70R/D78Q/G117Y	+++	++	+
106	213/214	E151C	+	++	++
107	215/216	E158N	++	+++	+++

[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 70 and defined as follows:

"+" > 1.10, "++" > 1.80, "+++" > 2.20

The engineered dsRNA ligase polypeptides represented by the even numbered sequence identifiers of SEQ ID NOs: 108 to 216 comprise an even numbered sequence identifier of SEQ ID NOs: 408 to 516, respectively, and a 14 amino acid N-terminal purification tag (MHHHHHHENLYFQS (SEQ ID NO: 669)). For example, SEQ ID NO: 108 comprises: (i) the 14 amino acid N-terminal purification tag of SEQ ID NO: 669; and (ii) the dsRNA ligase polypeptide of SEQ ID NO: 408.

EXAMPLE 9

Round 3 Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 188 for Improved Production of siRNA product (1)

The polynucleotide from example 8 SEQ ID NO: 187 encoding the most active polypeptide with dsRNA ligase activity of SEQ ID NO: 188 was used to generate the engineered polypeptides of Table 9-1. These polypeptides displayed improved dsRNA ligase activity under the desired conditions *e.g.*, the improvement in the formation of either

oligonucleotide products (2) or (3), or preferably both oligonucleotide products (2) and (3) that was produced *in situ* from the substrates oligonucleotides (6-7, 9-12) as compared to the starting polypeptide. Some polypeptides displayed improved product formation of both oligonucleotide products (2) and (3) compared to the starting polypeptide are noted in Table 9-1. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 188, as described below together with the analytical method described in Table 6-2.

Directed evolution began with the polynucleotide set forth in SEQ ID NO: 187. Libraries of engineered polypeptides were generated using various well-known techniques (*e.g.*, saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods, described below, that measured the polypeptides' ability to produce oligonucleotides (2) and (3).

The enzyme assays were carried out in 96-well PCR plates, in 100 µL total reaction volume per well. The reactions contained 20 % (v/v) of undiluted dsRNA ligase lysate, prepared as described in Example 4, 1 mM (each) substrate oligonucleotides (6-7, 9-12), 50 mM Tris-buffer at pH 7.0, 10 mM ATP, 20 mM $MgCl_2$, 5 mM DTT and 10 % (v/v) DMSO. The reaction plates were heat-sealed and incubated in a thermocycler at 30 °C for 2 h.

After incubation the plates were subjected to a heat inactivation step (95 °C, 20 min) to quench the reaction and precipitate proteinaceous content of the added lysate. The plates were then centrifuged at 4,000 rpm for 5 min. Subsequently a 50 µL aliquot of the supernatant was removed from each well and added to a deep well 96-well plate containing 450 µL of 5 mM EDTA solution (pH 7.0). The samples were further diluted by transferring 50 µL of the diluted sample into a deep well 96-well plate containing 950 µL of 5 mM EDTA solution (pH 7.0). The samples were analyzed via HPLC to determine the activity of the enzyme variants using the analytical method described in Table 6-2. Selected ligase variants showing a faster product formation of oligonucleotides (2) and (3) relative to SEQ ID NO:188 are shown in Table 9-1.

Table 9-1. dsRNA Ligase Activity for the Production of siRNA (1) Relative to SEQ ID NO: 188

Engineered ligase	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 188)[1]	FIOP AU[1,2]	FIOP (3)[1,2]	FIOP (2)[1,2]
108	217/218	F294W	+	+++	+++
109	219/220	M298A	+	+	+
110	221/222	E328A	+	+++	+++
111	223/224	I283L	+++	+++	+++

112	225/226	V269C	+++	+++	+++
113	227/228	V43L/A187L	+	+++	+++
114	229/230	N265D	+++	+++	+++
115	231/232	N265L	+++	+++	+++
116	233/234	T102C/E158N/M177G/V315L/T317Q	+	++	++
117	235/236	V43L/V315L/T317Q	+	+++	+++
118	237/238	E328V	+	+++	++
119	239/240	V43L/V204Q/P258K/V315L/I342R	+	++	++
120	241/242	A105S	+	+	+
121	243/244	G299A	+	++	++
122	245/246	V43L/E158N/M177G	++	+++	+++

[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 188 and defined as follows:

"+" > 1.00, "++" > 1.50, "+++" > 2.00

[2]Levels of increased activity were determined from the mean of three replicates

The engineered dsRNA ligase polypeptides represented by the even numbered sequence identifiers of SEQ ID NOs: 218 to 246 comprise an even numbered sequence identifier of SEQ ID NOs: 518 to 546, respectively, and a 14 amino acid N-terminal purification tag (MHHHHHHENLYFQS (SEQ ID NO: 669)). For example, SEQ ID NO: 218 comprises: (i) the 14 amino acid N-terminal purification tag of SEQ ID NO: 669; and (ii) the dsRNA ligase polypeptide of SEQ ID NO: 518.

EXAMPLE 10

Round 4 Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 226 for Improved Production of siRNA product (1)

The polynucleotide from example 9 SEQ ID NO: 225 encoding the most active polypeptide with dsRNA ligase activity of SEQ ID NO: 226 was used to generate the engineered polypeptides of Table 10-1. These polypeptides displayed improved dsRNA ligase activity under the desired conditions *e.g.*, the improvement in the formation of either oligonucleotide products (2) or (3), or preferably both oligonucleotide products (2) and (3) that was produced *in situ* from the substrates oligonucleotides (6-7, 9-12) as compared to the starting polypeptide. Some polypeptides displayed improved product formation of both oligonucleotide products (2) and (3) compared to the starting polypeptide are noted in Table 10-1. The engineered polypeptides, having the amino acid sequences of even-numbered

sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 226, as described below together with the analytical method described in Table 6-2.

Directed evolution began with the polynucleotide set forth in SEQ ID NO: 225. Libraries of engineered polypeptides were generated using various well-known techniques (*e.g.*, saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods, described below, that measured the polypeptides' ability to produce oligonucleotides (2) and (3).

The enzyme assays were carried out in 96-well PCR plates, in 100 µL total reaction volume per well. The reactions contained 2.5 % (v/v) of undiluted dsRNA ligase lysate, prepared as described in Example 4, 1 mM (each) substrate oligonucleotides (6-7, 9-12), 50 mM Tris-buffer at pH 7.0, 10 mM ATP, 20 mM $MgCl_2$, 5 mM DTT and 10 % (v/v) DMSO. The reaction plates were heat-sealed and incubated in a thermocycler at 30 °C for 24 h.

After incubation the plates were subjected to a heat inactivation step (95 °C, 20 min) to quench the reaction and precipitate proteinaceous content of the added lysate. The plates were then centrifuged at 4,000 rpm for 5 min. Subsequently a 50 µL aliquot of the supernatant was removed from each well and added to a deep well 96-well plate containing 450 µL of 5 mM EDTA solution (pH 7.0). The samples were further diluted by transferring 50 µL of the diluted sample into a deep well 96-well plate containing 950 µL of 5 mM EDTA solution (pH 7.0). The samples were analyzed via HPLC to determine the activity of the enzyme variants using the analytical method described in Table 6-2. Selected ligase variants showing a faster product formation of oligonucleotides (2) and (3) relative to SEQ ID NO: 226 are shown in Table 10-1.

Table 10-1. dsRNA Ligase Activity for the Production of siRNA (1) Relative to SEQ ID NO: 226

Engineered ligase	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 226)[1]	FIOP AU[1,2]	FIOP (3)[1,2]	FIOP (2)[1,2]
123	247/248	Y101T	+	+	+
124	249/250	L136W	+	+	+
125	251/252	Y101P	++	++	++
126	253/254	Y101K	++	+++	+++
127	255/256	S107G/F172W	+++	+++	+++
128	257/258	K242R	+	+	++
129	259/260	E60Y	+	+	++
130	261/262	E140G	+++	+++	+++

131	263/264	S107C/F172W/L266V/S319G	+++	+++	+++
132	265/266	T192R	++	++	++
133	267/268	N144R	+++	+++	+++
134	269/270	G160R	+	++	++
135	271/272	N144S	++	++	++
136	273/274	N144Y	+	+++	+++
137	275/276	A105S	+++	+++	+++
138	277/278	A67Y/K251R/N265L/G299A	+++	+++	+++
139	279/280	N63G	++	++	++
140	281/282	T65L	+	+++	+++
[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 226 and defined as follows: "+" > 1.00, "++" > 2.00, "+++" > 4.00 [2]Levels of increased activity were determined from the mean of three replicates					

The engineered dsRNA ligase polypeptides represented by the even numbered sequence identifiers of SEQ ID NOs: 248 to 282 comprise an even numbered sequence identifier of SEQ ID NOs: 548 to 582, respectively, and a 14 amino acid N-terminal purification tag (MHHHHHHENLYFQS (SEQ ID NO: 669)). For example, SEQ ID NO: 248 comprises: (i) the 14 amino acid N-terminal purification tag of SEQ ID NO: 669; and (ii) the dsRNA ligase polypeptide of SEQ ID NO: 548.

EXAMPLE 11

Round 5 Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 278 for Improved Production of siRNA product (1)

The polynucleotide from example 10 SEQ ID NO: 277 encoding the most active polypeptide with dsRNA ligase activity of SEQ ID NO: 278 was used to generate the engineered polypeptides of Table 11-1. These polypeptides displayed improved dsRNA ligase activity under the desired conditions *e.g.*, the improvement in the formation of either oligonucleotide products (2) or (3), or preferably both oligonucleotide products (2) and (3) that was produced *in situ* from the substrates oligonucleotides (6-7, 9-12) as compared to the starting polypeptide. Some polypeptides displayed improved product formation of both oligonucleotide products (2) and (3) compared to the starting polypeptide are noted in Table 11-1. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the “backbone” amino acid sequence of SEQ ID NO: 278, as described below together with the analytical method described in Table 6-2.

Directed evolution began with the polynucleotide set forth in SEQ ID NO: 277. Libraries of engineered polypeptides were generated using various well-known techniques (*e.g.*, saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods, described below, that measured the polypeptides' ability to produce oligonucleotides (2) and (3).

The enzyme assays were carried out in 96-well PCR plates, in 100 µL total reaction volume per well. The reactions contained 10 % (v/v) of undiluted dsRNA ligase lysate, prepared as described in Example 4, 5 mM (each) substrate oligonucleotides (6-7, 9-12), 50 mM Tris-buffer at pH 7.0, 30 mM ATP, 60 mM $MgCl_2$ and 10 % (v/v) DMSO. The reaction plates were heat-sealed and incubated in a thermocycler at 30 °C for 24 h.

After incubation the plates were subjected to a heat inactivation step (95 °C, 20 min) to quench the reaction and precipitate proteinaceous content of the added lysate. The plates were then centrifuged at 4,000 rpm for 5 min. Subsequently a 50 µL aliquot of the supernatant was removed from each well and added to a deep well 96-well plate containing 950 µL of 5 mM EDTA solution (pH 7.0). The samples were further diluted by transferring 50 µL of the diluted sample into a deep well 96-well plate containing 450 µL of 5 mM EDTA solution (pH 7.0). The samples were diluted a third time by transferring 160 µL of the diluted sample into a deep well 96-well plate containing 640 µL of 5 mM EDTA solution (pH 7.0). The samples were diluted a final time by transferring 75 µL of the diluted sample into a shallow well 96-well plate containing 75 µL of 5 mM EDTA solution (pH 7.0). The samples were analyzed via HPLC to determine the activity of the enzyme variants using the analytical method described in Table 6-2. Selected ligase variants showing a faster product formation of oligonucleotides (2) and (3) relative to SEQ ID NO: 278 are shown in Table 11-1.

Table 11-1. dsRNA Ligase Activity for the Production of siRNA (1) Relative to SEQ ID NO: 278

Engineered ligase	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 278)[1]	FIOP AU[1,2]	FIOP (3)[1,2]	FIOP (2)[1,2]
141	283/284	T145R[3]	++	++	++
142	285/286	Y80W	++	++	++
143	287/288	K29D	+++	+++	+++
144	289/290	K29E[3]	+++	+++	+++
145	291/292	K33D	+++	+++	+++
146	293/294	N63G/Y101P/T192R/L265N/A299G	+	++	+

147	295/296	Y101-/A105S/S107C/L265N/A299G	+	+	+
148	297/298	S107A	++	++	++
149	299/300	R61E	++	+++	++

[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 278 and defined as follows:

"+" > 1.00, "++" > 2.00, "+++" > 4.00

[2]Levels of increased activity were determined from the mean of three replicates

[3]Levels of increased activity were determined from the mean of three replicates

The engineered dsRNA ligase polypeptides represented by the even numbered sequence identifiers of SEQ ID NOs: 284 to 300 comprise an even numbered sequence identifier of SEQ ID NOs: 584 to 600, respectively, and a 14 amino acid N-terminal purification tag (MHHHHHHENLYFQS (SEQ ID NO: 669)). For example, SEQ ID NO: 284 comprises: (i) the 14 amino acid N-terminal purification tag of SEQ ID NO: 669; and (ii) the dsRNA ligase polypeptide of SEQ ID NO: 584.

EXAMPLE 12

Round 6 Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 288 for Improved Production of siRNA product (1) and improved thermal stability

The polynucleotide from example 11 SEQ ID NO: 287 encoding the most active polypeptide with dsRNA ligase activity of SEQ ID NO: 288 was used to generate the engineered polypeptides of Table 12-1. These polypeptides displayed improved dsRNA ligase activity under the desired conditions *e.g.*, the improvement in the formation of either oligonucleotide products (2) or (3), or preferably both oligonucleotide products (2) and (3) that was produced *in situ* from the substrate oligonucleotides (6-7, 9-12) as compared to the starting polypeptide. Some polypeptides displayed improved product formation of both oligonucleotide products (2) and (3) compared to the starting polypeptide, as noted in Table 12-1. Furthermore, some polypeptides displayed improved thermal stability, resulting in higher residual activity following incubation of the dsRNA ligase solution at 30 °C for 1 h prior to setting up the reaction (Table 12-2). The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 288, as described below together with the analytical method described in Table 6-3.

Directed evolution began with the polynucleotide set forth in SEQ ID NO: 287. Libraries of engineered polypeptides were generated using various well-known techniques (*e.g.*, saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods, described below, that measured the polypeptides' ability to produce oligonucleotides (2) and (3).

The enzyme assays were carried out in 96-well PCR plates, in 100 µL total reaction volume per well. The cells were lysed according to Example 4, however 100 mM MOPS-buffer at pH 7.2 was used in place of 50 mM Tris at pH 7.5. To provide a thermal challenge and to identify hits that were more thermostable, the cell lysates were either undiluted or diluted 1:1 in 100 mM MOPS buffer, pH 7.2 and incubated at 30 °C and 4 °C respectively for 1 h. The reactions contained a final dsRNA ligase concentration of 40 % (v/v) for lysate incubated at 30 °C and 20% (v/v) for lysate incubated at 4 °C. In addition, the ligation reactions contained 3 mM (each) substrate oligonucleotides (6-7, 9-12), 100 mM MOPS-buffer at pH 7.2, 30 mM ATP, 60 mM $MgCl_2$ and 10 % (v/v) DMSO. The reaction plates were heat-sealed and incubated in a thermocycler at 30 °C for 24 h.

After incubation the plates were subjected to a heat inactivation step (95 °C, 20 min) to quench the reaction and precipitate proteinaceous content of the added lysate. The plates were then centrifuged at 4,000 rpm for 5 min. Subsequently a 50 µL aliquot of the supernatant was removed from each well and added to a deep well 96-well plate containing 950 µL of 5 mM EDTA solution (pH 7.0). The samples were further diluted by transferring 20 µL of the diluted sample into a deep well 96-well plate containing 180 µL of 5 mM EDTA solution (pH 7.0). The samples were diluted a third time by transferring 30 µL of the diluted sample into a deep well 96-well plate containing 150 µL of 5 mM EDTA solution (pH 7.0). The samples were analyzed via RF-MS to determine the activity of the enzyme variants using the analytical method described in Table 6-3. Selected ligase variants showing a faster product formation of oligonucleotides (2) and (3) relative to SEQ ID NO: 288 following pre-incubation at 4 °C are shown in Table 12-1. Selected ligase variants showing a faster product formation of oligonucleotides (2) and (3) relative to SEQ ID NO: 288 following pre-incubation at 30 °C are shown in Table 12-2.

Table 12-1. dsRNA Ligase Activity for the Production of siRNA (1) Relative to SEQ ID NO: 288 pre-incubated at 4 °C

Engineered ligase	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 288)[1]	Mean FIOP SUM (2) 5 peaks [1,2]	Mean FIOP SUM (3) 5 peaks [1,2]
150	601/602	S20E/K33D/I58V/I59V/Y80W/T145R/T192R/A299G	++	++
151	603/604	Y101T/S107G/E140G/T145R	+++	+++
152	605/606	S20E/K33D/I58V/E140G	+	+
153	607/608	K33D/I59V/T145R/T192R	+	+
154	609/610	S20E/I58V/V106D/A299G	+	++
155	611/612	R251L	+	+
156	613/614	I58V/I59V/R61E/Y101T/A103T/T145R/T192R/A299G	+	++
157	615/616	K327A	++	++
158	617/618	E140G/T145R/A299G	++	++
159	619/620	S20E/I58V/I59V/T192R	++	++
160	621/622	R251G	+	+
161	623/624	S20E/I58V/I59V	+++	+++
162	625/626	S20E/D29E/K33D/I58V/I59V/Y80W/S107G/T192R/L265N/A299G	+	+
163	627/628	I58V/Y80W/E140G/T145R/T192R/A299G	+	+
164	629/630	Y101T/A103T/E140G/T145R/A299G	+	++
165	631/632	E199K	+++	+++
166	633/634	E140G/T145R/T192R	+	++

[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 288 and defined as follows:

"+" > 3.00, "++" > 15.00, "+++" > 25.00

[2]Levels of increased activity were determined from the mean of two replicates

Table 12-2. dsRNA Ligase Activity for the Production of siRNA (1) Relative to SEQ ID NO: 288 pre-incubated at 30 °C

Engineered ligase	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 288)[1]	Mean FIOP SUM (2) 5 peaks [1,2]	Mean FIOP SUM (3) 5 peaks [1,2]
150	601/602	S20E/K33D/I58V/I59V/Y80W/T145R/T192R/A299G	+++	+++
151	603/604	Y101T/S107G/E140G/T145R	++	+++
152	605/606	S20E/K33D/I58V/E140G	++	+
153	607/608	K33D/I59V/T145R/T192R	+++	++
154	609/610	S20E/I58V/V106D/A299G	+	+
155	611/612	R251L	++	++
156	613/614	I58V/I59V/R61E/Y101T/A103T/T145R/T192R/A299G	+	++
157	615/616	K327A	+	++
158	617/618	E140G/T145R/A299G	+	+
159	619/620	S20E/I58V/I59V/T192R	++	+
160	621/622	R251G	+	+
161	623/624	S20E/I58V/I59V	++	+
162	625/626	S20E/D29E/K33D/I58V/I59V/Y80W/S107G/T192R/L265N/A299G	+	+
163	627/628	I58V/Y80W/E140G/T145R/T192R/A299G	+	+
164	629/630	Y101T/A103T/E140G/T145R/A299G	++	+++
165	631/632	E199K	+++	+++
166	633/634	E140G/T145R/T192R	+	++

[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 288 and defined as follows:

"+" > 1.50, "++" > 3.00, "+++" > 3.50

[2]Levels of increased activity were determined from the mean of two replicates

The engineered dsRNA ligase polypeptides represented by the even numbered sequence identifiers of SEQ ID NOs: 602 to 634 comprise an even numbered sequence identifier of SEQ ID NOs: 636 to 668, respectively, and a 14 amino acid N-terminal purification tag (MHHHHHHENLYFQS (SEQ ID NO: 669)). For example, SEQ ID NO: 602 comprises: (i) the 14 amino acid N-terminal purification tag of SEQ ID NO: 669; and (ii) the dsRNA ligase polypeptide of SEQ ID NO: 636.

EXAMPLE 13

Comparison of the catalytic activity of the wildtype polypeptide SEQ ID NO: 2 and the engineered polypeptides SEQ ID NO: 288, SEQ ID NO: 290 and SEQ ID NO: 292

The polynucleotides SEQ ID NO: 1 encoding for the wild-type dsRNA ligase from Bacteriophage RB69, Uniprot ID: Q7Y4V8 with the SEQ ID NO: 2 and the engineered polynucleotides SEQ ID NO: 287, SEQ ID NO: 289 and SEQ ID NO: 291 encoding for the most improved variants from example 11 with polypeptide sequences SEQ ID NO: 288, SEQ ID NO: 290 and SEQ ID NO: 292, have been used for SFP production as described in example 5.

The catalytic activity to convert the substrates oligonucleotides (6-7, 9-12) to the desired siRNA product (1) was evaluated under two reaction conditions: Condition 1 (50 mM Tris-buffer at pH 7.5, 1 mM ATP, 5 mM $MgCl_2$ and 5 mM DTT, containing either 0 g/L, 0.0020 g/L, 0.0039 g/L, 0.0078 g/L, 0.0156 g/L, 0.0313 g/L, 0.0625 g/L, 0.125 g/L, 0.25 g/L, 0.5 g/L, 1 g/L, or 2 g/L of SFP), and Condition 2 (50 mM Tris-buffer at pH 7.0, 30 mM ATP, 60 mM $MgCl_2$ and 10 % (v/v) DMSO containing either 0 g/L, 0.0049 g/L, 0.0098 g/L, 0.0195 g/L, 0.0391 g/L, 0.0781 g/L, 0.1563 g/L, 0.3125 g/L, 0.625 g/L, 1.25 g/L, 2.5 g/L, or 5 g/L of SFP). The enzyme assays were carried out in 96-well PCR plates, in 100 µL total reaction volume per well; condition 1 in reaction plate 1 and condition 2 in reaction plate 2. Both reaction plates were heat-sealed and incubated in a thermocycler at 30 °C. Reaction plate 1 was incubated for 2 h and reaction plate 2 was incubated for 24 h.

Following incubation, the plates were subjected to a heat inactivation step (95 °C, 20 min) to quench the reaction and precipitate proteinaceous content of the added SFP. The plates were then centrifuged at 4,000 rpm for 5 min. A 50 µL aliquot of the supernatant from each well of each plate was removed and subsequently diluted in 50 mM EDTA solution (pH 7.0). Reaction plate 1 was diluted 40 x and reaction plate 2 was diluted 2 400 x. The samples were analyzed via HPLC to determine the activity of the enzyme variants using the analytical method described in Table 6-2.

Comparative data in Figures 2A and 2B show the relative peak area % of siRNA (1) present in the reaction samples assayed under conditions 1 and 2 respectively. Under both conditions 1 and 2, the polypeptides SEQ ID NO: 288, SEQ ID NO: 290 and SEQ ID NO: 292 exhibit improved dsRNA ligase activity over the wild-type polypeptide SEQ ID NO: 2.

EXAMPLE 14

Comparison of the catalytic activity of the wildtype polypeptide SEQ ID NO: 2 and the engineered polypeptides SEQ ID NOs: 288 and 632

The polynucleotides SEQ ID NO: 1 encoding for the wild-type dsRNA ligase from Bacteriophage RB69, Uniprot ID: Q7Y4V8 with the SEQ ID NO: 2 and the engineered polynucleotide SEQ ID NO: 287 and SEQ ID NO: 631 encoding for the most improved variant from example 11 and example 12 with polypeptide sequences SEQ ID NO: 288 and SEQ ID NO: 632 respectively, have been used for SFP production as described in example 5.

The catalytic activity to convert the substrate oligonucleotides (6-7, 9-12) to the desired siRNA product (1) and the thermostability of the two enzymes was evaluated by incubating a stock solution of the SFP for 4 h at either 4 or 37 °C prior to setting up the following ligation reaction: 6 mM (each) substrate oligonucleotides (6-7, 9-12), 100 mM MOPS-buffer at pH 7.2, 30 mM ATP, 60 mM $MgCl_2$ and 10 % (v/v) DMSO. In addition, the ligation reactions contained, either 0 g/L, 0.156 g/L, 0.313 g/L, 0.625 g/L, 1.25 g/L, 2.5 g/L, 5 g/L, or 10 g/L of SFP. The enzyme assays were carried out in 96-well PCR plates, in 100 µL total reaction volume per well. The reaction plate was heat-sealed and incubated in a thermocycler at 30 °C for 24 h.

Following incubation, the plate was subjected to a heat inactivation step (95 °C, 20 min) to quench the reaction and precipitate proteinaceous content of the added SFP. The plate was then centrifuged at 4,000 rpm for 5 min. A 50 µL aliquot of the supernatant from each well of each plate was removed and subsequently diluted 400 x in 50 mM EDTA solution (pH 7.0). The samples were analyzed via HPLC to determine the activity of the enzyme variants using the analytical method described in Table 6-2.

Comparative data in Figure 3A shows the relative peak area % of siRNA (1) present in the reaction samples. Figure 3B shows the residual enzyme activity following pre-incubation of the SFP at 37 °C for 4 h, expressed relative to the ligation activity of the SFP pre-incubated at 4 °C for 4 h. Under all conditions, the polypeptides SEQ ID NO: 632, exhibits improved dsRNA ligase activity and thermostability over the wild-type polypeptide SEQ ID NO: 2 and engineered polypeptide SEQ ID NO: 288.

SUMMARY OF ENGINEERED dsRNA LIGASE POLYPEPTIDE SEQUENCES

Table 13 provides a summary of the nucleic acid and amino acid sequences of the wild-type and engineered dsRNA ligase sequences described herein. The purification tag used in the Examples and reference in table 13 is the N-terminal purification tag MHHHHHHENLYFQS (SEQ ID NO: 669).

Table 13.

	SEQ ID NO of Nucleic acid sequence with purification tag	SEQ ID NO of Amino acid sequence with purification tag	SEQ ID NO of Nucleic acid sequence without purification tag	SEQ ID NO of Amino acid sequence without purification tag
wild type ligase	1	2	301	302
engineered ligase 1	3	4	303	304
engineered ligase 2	5	6	305	306
engineered ligase 3	7	8	307	308
engineered ligase 4	9	10	309	310
engineered ligase 5	11	12	311	312
engineered ligase 6	13	14	313	314
engineered ligase 7	15	16	315	316
engineered ligase 8	17	18	317	318
engineered ligase 9	19	20	319	320
engineered ligase 10	21	22	321	322
engineered ligase 11	23	24	323	324
engineered ligase 12	25	26	325	326
engineered ligase 13	27	28	327	328
engineered ligase 14	29	30	329	330
engineered ligase 15	31	32	331	332
engineered ligase 16	33	34	333	334
engineered ligase 17	35	36	335	336
engineered ligase 18	37	38	337	338
engineered ligase 19	39	40	339	340
engineered ligase 20	41	42	341	342
engineered ligase 21	43	44	343	344
engineered ligase 22	45	46	345	346
engineered ligase 23	47	48	347	348
engineered ligase 24	49	50	349	350
engineered ligase 25	51	52	351	352
engineered ligase 26	53	54	353	354
engineered ligase 27	55	56	355	356
engineered ligase 28	57	58	357	358
engineered ligase 29	59	60	359	360
engineered ligase 30	61	62	361	362

	SEQ ID NO of Nucleic acid sequence with purification tag	SEQ ID NO of Amino acid sequence with purification tag	SEQ ID NO of Nucleic acid sequence without purification tag	SEQ ID NO of Amino acid sequence without purification tag
engineered ligase 31	63	64	363	364
engineered ligase 32	65	66	365	366
engineered ligase 33	67	68	367	368
engineered ligase 34	69	70	369	370
engineered ligase 35	71	72	371	372
engineered ligase 36	73	74	373	374
engineered ligase 37	75	76	375	376
engineered ligase 38	77	78	377	378
engineered ligase 39	79	80	379	380
engineered ligase 40	81	82	381	382
engineered ligase 41	83	84	383	384
engineered ligase 42	85	86	385	386
engineered ligase 43	87	88	387	388
engineered ligase 44	89	90	389	390
engineered ligase 45	91	92	391	392
engineered ligase 46	93	94	393	394
engineered ligase 47	95	96	395	396
engineered ligase 48	97	98	397	398
engineered ligase 49	99	100	399	400
engineered ligase 50	101	102	401	402
engineered ligase 51	103	104	403	404
engineered ligase 52	105	106	405	406
engineered ligase 53	107	108	407	408
engineered ligase 54	109	110	409	410
engineered ligase 55	111	112	411	412
engineered ligase 56	113	114	413	414
engineered ligase 57	115	116	415	416
engineered ligase 58	117	118	417	418
engineered ligase 59	119	120	419	420
engineered ligase 60	121	122	421	422
engineered ligase 61	123	124	423	424
engineered ligase 62	125	126	425	426
engineered ligase 63	127	128	427	428
engineered ligase 64	129	130	429	430
engineered ligase 65	131	132	431	432
engineered ligase 66	133	134	433	434
engineered ligase 67	135	136	435	436
engineered ligase 68	137	138	437	438
engineered ligase 69	139	140	439	440
engineered ligase 70	141	142	441	442
engineered ligase 71	143	144	443	444

	SEQ ID NO of Nucleic acid sequence with purification tag	SEQ ID NO of Amino acid sequence with purification tag	SEQ ID NO of Nucleic acid sequence without purification tag	SEQ ID NO of Amino acid sequence without purification tag
engineered ligase 72	145	146	445	446
engineered ligase 73	147	148	447	448
engineered ligase 74	149	150	449	450
engineered ligase 75	151	152	451	452
engineered ligase 76	153	154	453	454
engineered ligase 77	155	156	455	456
engineered ligase 78	157	158	457	458
engineered ligase 79	159	160	459	460
engineered ligase 80	161	162	461	462
engineered ligase 81	163	164	463	464
engineered ligase 82	165	166	465	466
engineered ligase 83	167	168	467	468
engineered ligase 84	169	170	469	470
engineered ligase 85	171	172	471	472
engineered ligase 86	173	174	473	474
engineered ligase 87	175	176	475	476
engineered ligase 88	177	178	477	478
engineered ligase 89	179	180	479	480
engineered ligase 90	181	182	481	482
engineered ligase 91	183	184	483	484
engineered ligase 92	185	186	485	486
engineered ligase 93	187	188	487	488
engineered ligase 94	189	190	489	490
engineered ligase 95	191	192	491	492
engineered ligase 96	193	194	493	494
engineered ligase 97	195	196	495	496
engineered ligase 98	197	198	497	498
engineered ligase 99	199	200	499	500
engineered ligase 100	201	202	501	502
engineered ligase 101	203	204	503	504
engineered ligase 102	205	206	505	506
engineered ligase 103	207	208	507	508
engineered ligase 104	209	210	509	510
engineered ligase 105	211	212	511	512
engineered ligase 106	213	214	513	514
engineered ligase 107	215	216	515	516
engineered ligase 108	217	218	517	518
engineered ligase 109	219	220	519	520
engineered ligase 110	221	222	521	522
engineered ligase 111	223	224	523	524
engineered ligase 112	225	226	525	526

	SEQ ID NO of Nucleic acid sequence with purification tag	SEQ ID NO of Amino acid sequence with purification tag	SEQ ID NO of Nucleic acid sequence without purification tag	SEQ ID NO of Amino acid sequence without purification tag
engineered ligase 113	227	228	527	528
engineered ligase 114	229	230	529	530
engineered ligase 115	231	232	531	532
engineered ligase 116	233	234	533	534
engineered ligase 117	235	236	535	536
engineered ligase 118	237	238	537	538
engineered ligase 119	239	240	539	540
engineered ligase 120	241	242	541	542
engineered ligase 121	243	244	543	544
engineered ligase 122	245	246	545	546
engineered ligase 123	247	248	547	548
engineered ligase 124	249	250	549	550
engineered ligase 125	251	252	551	552
engineered ligase 126	253	254	553	554
engineered ligase 127	255	256	555	556
engineered ligase 128	257	258	557	558
engineered ligase 129	259	260	559	560
engineered ligase 130	261	262	561	562
engineered ligase 131	263	264	563	564
engineered ligase 132	265	266	565	566
engineered ligase 133	267	268	567	568
engineered ligase 134	269	270	569	570
engineered ligase 135	271	272	571	572
engineered ligase 136	273	274	573	574
engineered ligase 137	275	276	575	576
engineered ligase 138	277	278	577	578
engineered ligase 139	279	280	579	580
engineered ligase 140	281	282	581	582
engineered ligase 141	283	284	583	584
engineered ligase 142	285	286	585	586
engineered ligase 143	287	288	587	588
engineered ligase 144	289	290	589	590
engineered ligase 145	291	292	591	592
engineered ligase 146	293	294	593	594
engineered ligase 147	295	296	595	596
engineered ligase 148	297	298	597	598
engineered ligase 149	299	300	599	600
engineered ligase 150	601	602	635	636
engineered ligase 151	603	604	637	638
engineered ligase 152	605	606	639	640
engineered ligase 153	607	608	641	642

	SEQ ID NO of Nucleic acid sequence with purification tag	SEQ ID NO of Amino acid sequence with purification tag	SEQ ID NO of Nucleic acid sequence without purification tag	SEQ ID NO of Amino acid sequence without purification tag
engineered ligase 154	609	610	643	644
engineered ligase 155	611	612	645	646
engineered ligase 156	613	614	647	648
engineered ligase 157	615	616	649	650
engineered ligase 158	617	618	651	652
engineered ligase 159	619	620	653	654
engineered ligase 160	621	622	655	656
engineered ligase 161	623	624	657	658
engineered ligase 162	625	626	659	660
engineered ligase 163	627	628	661	662
engineered ligase 164	629	630	663	664
engineered ligase 165	631	632	665	666
engineered ligase 166	633	634	667	668

CLAIMS

1. An engineered double-stranded RNA (dsRNA) ligase polypeptide comprising an amino acid sequence having at least 85% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 666, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, and 668;
wherein the engineered dsRNA ligase polypeptide:

(a) has dsRNA ligase activity; and

(b) does not the comprise the amino acid sequence of SEQ ID NO: 302.

2. The engineered dsRNA ligase polypeptide of claim 1, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 666, 370, 488, 526, 578, 588, 590, and 592.

3. An engineered dsRNA ligase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 302, which produces at least 5% more oligonucleotide product than a dsRNA ligase polypeptide comprising the amino acid sequence of SEQ ID NO: 302 under the same ligation reaction conditions, wherein the engineered dsRNA ligase polypeptide does not the comprise the amino acid sequence of SEQ ID NO: 302.

4. The engineered dsRNA ligase polypeptide of claim 3, wherein the ligation reaction conditions include about 1 μM to about 10 mM oligonucleotide fragment, a source of ATP, about 5 mM to about 100 mM divalent cation, and about 0.5 g/L to about 10 g/L engineered dsRNA ligase polypeptide, pH of about 4.0 to about 8.0, and temperature of about 10 °C to about 50 °C.

5. The engineered dsRNA ligase of any preceding claim, wherein:

(a) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X6, X7, X15, X19, X29, X36, X39, X44, X45, X46, X47, X49, X51, X53, X56, X57, X60, X63, X64, X66, X67, X87, X88, X89, X91, X92, X93, X103, X105, X107, X114, X122, X126, X129, X130, X131, X137, X144, X146, X158, X163, X173, X178, X185, X190, X196, X216, X218, X221, X228, X230, X232, X235, X236, X237, X238, X239, X242, X243, X244, X251, X252, X254, X255, X258, X269, X280, X284, X285, X293, X296, X301, X303, X305, X313, X314, X325, and X328, wherein the numbering refers to SEQ ID NO: 302; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X6 is G or E; X7 is Q; X15 is R, D or E; X19 is Q or D; X29 is N or L; X36 is V; X39 is A; X44 is V; X45 is V; X46 is Y; X47 is E; X49 is G; X51 is L; X53 is Y; X56 is R or A; X57 is S; X60 is T, G or P; X63 is S, Q or G; X64 is R, T, Q, F, G, or M; X66 is F or W; X67 is N; X87 is T, P, K or absent; X88 is C; X89 is T; X91 is S; X92 is D; X93 is G, C, or A; X103 is V, C, Y, or T; X105 is V; X107 is R or T; X114 is N; X122 is W; X126 is G; X129 is N; X130 is R, S or Y; X131 is R; X137 is V or C; X144 is N; X146 is R; X158 is W; X163 is G; X173 is L; X178 is R; X185 is K; X190 is Q; X196 is S or C; X216 is L or R; X218 is N; X221 is I; X228 is R; X230 is T; X232 is R; X235 is A, T, or G; X236 is S, L, or F; X237 is S, Q, R, L or G; X238 is F; X239 is G or R; X242 is R or M; X243 is N, S, G, or M; X244 is G or K; X251 is D or L; X252 is V; X254 is K; X255 is C; X258 is V; X269 is L; X280 is W; X284 is A; X285 is A; X293 is R; X296 is R; X301 is G, L, E, or F; X303 is Q; X305 is G; X313 is A; X314 is A or V; X325 is R; and X328 is R; wherein the numbering refers to SEQ ID NO: 302; and/or

(b) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X15, X19, X36, X39, X53, X185, X218, X221, X237, X251, X255, and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase

polypeptide comprises one or more of the following amino acid residues: X15 is D or E; X19 is D; X36 is V; X39 is A; X53 is Y; X185 is K; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; wherein the numbering refers to SEQ ID NO: 302; and/or

(c) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X36, X39, X218 and X221, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X36 is V; X39 is A; X218 is N; and X221 is I; and/or

(d) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X39, X218 and X221, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X39 is A; X218 is N; and X221 is I; and/or

(e) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X39, X218, X221 and X255, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X39 is A; X218 is N; X221 is I; and X255 is C; and/or

(f) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X39, X53, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein

the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; and/or

(g) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X15, X39, X53, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X15 E; X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; and/or

(h) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X19, X39, X53, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X19 is D; X39 is A; X 53 is Y; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A; and/or

(i) the amino acid sequence of the engineered dsRNA ligase polypeptide comprises an amino acid sequence that differs from the sequence of SEQ ID NO: 302 in one or more amino acid residues selected from: X15, X39, X53, X185, X218, X221, X237, X251, X255 and X285, wherein the numbering refers to SEQ ID NO: 302, and wherein the engineered dsRNA ligase polypeptide has dsRNA ligase activity; optionally wherein the amino acid sequence of the engineered dsRNA ligase polypeptide comprises one or more of the following amino acid residues: X15 is D; X39 is A; X 53 is Y; X185 is K; X218 is N; X221 is I; X237 is R; X251 is L; X255 is C; and X285 is A.

6. The engineered dsRNA ligase polypeptide of any of claims 1-5, wherein the engineered dsRNA ligase polypeptide comprises a purification tag; optionally wherein the engineered dsRNA ligase polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 632, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, and 634.

7. A polypeptide immobilized on a solid material by chemical bond or a physical adsorption method, wherein the polypeptide comprises an engineered dsRNA ligase polypeptide according to any one of claims 1-6.

8. A polynucleotide encoding the engineered dsRNA ligase polypeptide of any one of claims 1-6; optionally wherein:

(a) the polynucleotide comprises a nucleic acid sequence selected from SEQ ID NOs: 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, and 667; and/or

(b) the polynucleotide comprises a nucleic acid sequence selected from SEQ ID NOs: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129,

131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, and 633.

9. An expression vector comprising the polynucleotide according to claim 8; optionally wherein the expression vector comprises a plasmid, a cosmid, a bacteriophage or a viral vector.

10. A host cell comprising the polynucleotide of claim 8 or the expression vector of claim 9, optionally wherein the host cell is *E. coli*.

11. A method of preparing an engineered dsRNA ligase polypeptide, which comprises the steps of culturing the host cell according to claim 10 and obtaining an engineered dsRNA ligase polypeptide from the culture.

12. An engineered dsRNA ligase catalyst obtainable by culturing the host cells according to claim 10, or according to the method of claim 11, wherein said engineered dsRNA ligase catalyst comprises cells or culture fluid containing the engineered dsRNA ligase polypeptides, or an article processed therewith, wherein the article refers to an extract obtained from the culture of host cell, an isolated product obtained by isolating or purifying an engineered dsRNA ligase from the extract, or an immobilized product obtained by immobilizing host cell, an extract thereof, or isolated product of the extract.

13. A method of producing an oligonucleotide from two or more oligonucleotide fragments, wherein the method comprises contacting:

(i) two or more oligonucleotide fragments;

(ii) an engineered dsRNA ligase polypeptide according to any one of claims 1-6;

(iii) a source of ATP; and

(iv) a divalent cation;

to obtain an oligonucleotide;

optionally wherein:

(a) the method further comprises a step of purifying the oligonucleotide; and/or

(b) the method is performed using a sub-stoichiometric concentration of AMP and/or ATP;

and/or

(c) the method is performed with a divalent cation concentration of 5-100 mM, optionally 30-50 mM.

14. Use of the engineered dsRNA ligase polypeptide according to any one of claims 1-6 in the production of an oligonucleotide from two or more oligonucleotide fragments.

15. The method of claim 13 or the use of claim 14, wherein:

(a) the oligonucleotide is up to 60 nucleotides in length; and/or

(b) each of the oligonucleotide fragments are 4-16 nucleotides in length, optionally 6-9 nucleotides in length; and/or

(c) one or more of the oligonucleotide fragment(s) comprises one or two overhangs; and/or

(d) one or more of the oligonucleotide fragments comprises a chemical modification; optionally wherein the chemical modification is selected from:

(i) a modified backbone, optionally selected from a phosphorothioate (*e.g.* chiral phosphorothioate) or methylphosphonate internucleotide linkage;

(ii) a modified nucleotide, optionally selected from 2'-O-methyl (2'-OMe), 2'-flouro (2'-F), 2'-deoxy, 2'-deoxy-2'-fluoro, 2'-O-methoxyethyl (2'-O-MOE), 2'-O-aminopropyl (2'-O-AP), 2'-O-dimethylaminoethyl (2'-O-DMAOE), 2'-O-dimethylaminopropyl (2'-O-DMAP), 2'-O-dimethylaminoethyloxyethyl (2'-O-DMAEOE), 2'-O-N-methylacetamido (2'-O-NMA), locked nucleic acid (LNA), glycol nucleic acid (GNA), phosphoramidate (*e.g.* mesyl phosphoramidate), 2',3'-seco nucleotide mimic, 2'-F-arabino nucleotide, abasic nucleotide, 2'-amino modified nucleotide, 2'-alkyl-modified nucleotide, morpholino nucleotide, vinylphosphonate (*e.g.* 5' vinylphosphonate), and cyclopropyl phosphonate deoxyribonucleotide; and/or

(iii) conjugation to a ligand, optionally wherein the ligand comprises one or more N-Acetylgalactosamine (GalNAc) derivatives.

16. A composition comprising:

i. the engineered dsRNA ligase polypeptide according to any one of claims 1-6;

ii. a source of ATP; and

iii. a divalent cation;

optionally wherein the composition further comprises two or more oligonucleotide fragments.

17. A kit comprising:

i. the engineered dsRNA ligase polypeptide according to any one of claims 1-6;

ii. a source of ATP;

iii. a divalent cation; and

iv. instructions for use in a method of producing an oligonucleotide from two or more oligonucleotide fragments.

18. The method of claim 13 or claim 15, the composition of claim 16 or the kit of claim 17, wherein:

(A) the source of ATP comprises ATP; and/or

(B) the source of ATP comprises:

(a) polyphosphate kinase (PPK);

(b) polyphosphate; and

(c) AMP and/or ATP;

optionally wherein:

(i) the PPK is selected from PPK12 or ajPAP; and/or

(ii) the polyphosphate is a polyphosphate salt, optionally wherein the polyphosphate salt is sodium polyphosphate (Maddrell's salt) or sodium hexametaphosphate (Graham's salt); and/or

(C) the divalent cation cofactor is Mg^{2+} or Mn^{2+}.

Figure 1A

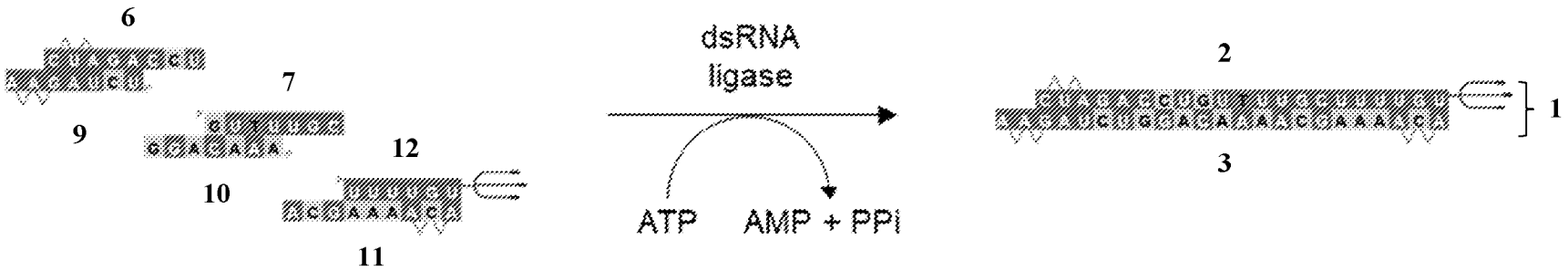


Figure 1B

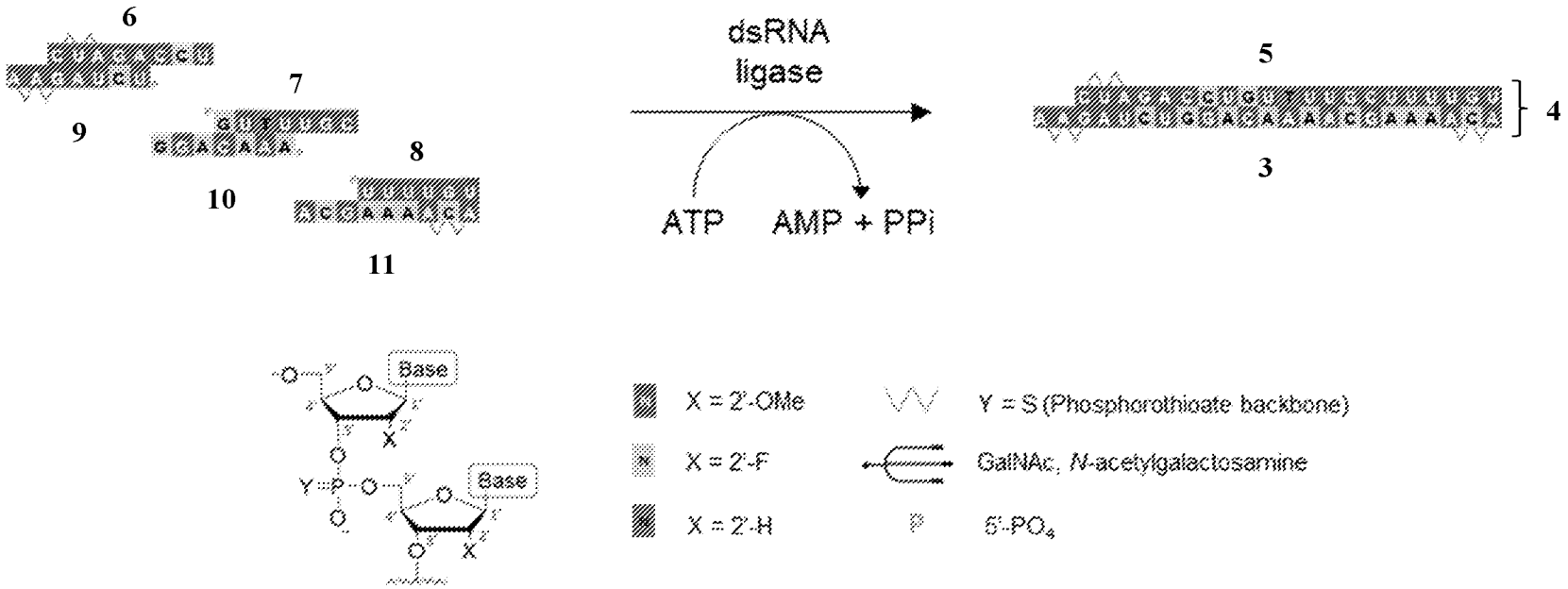


Figure 2A

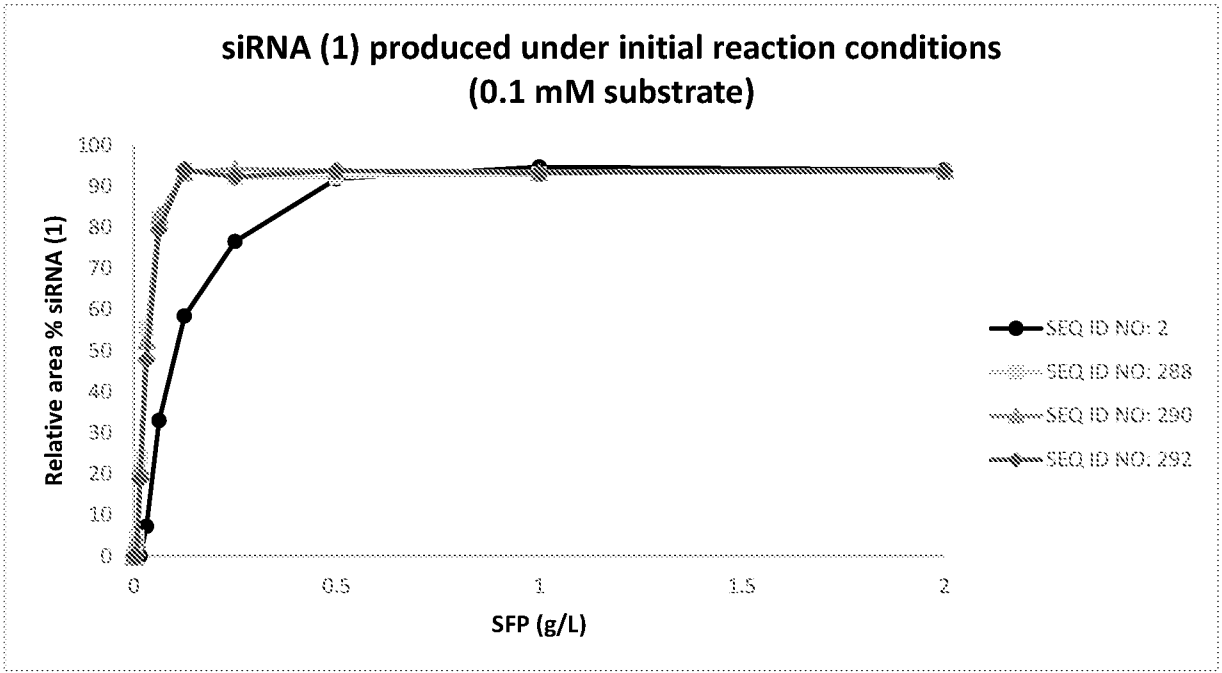


Figure 2B

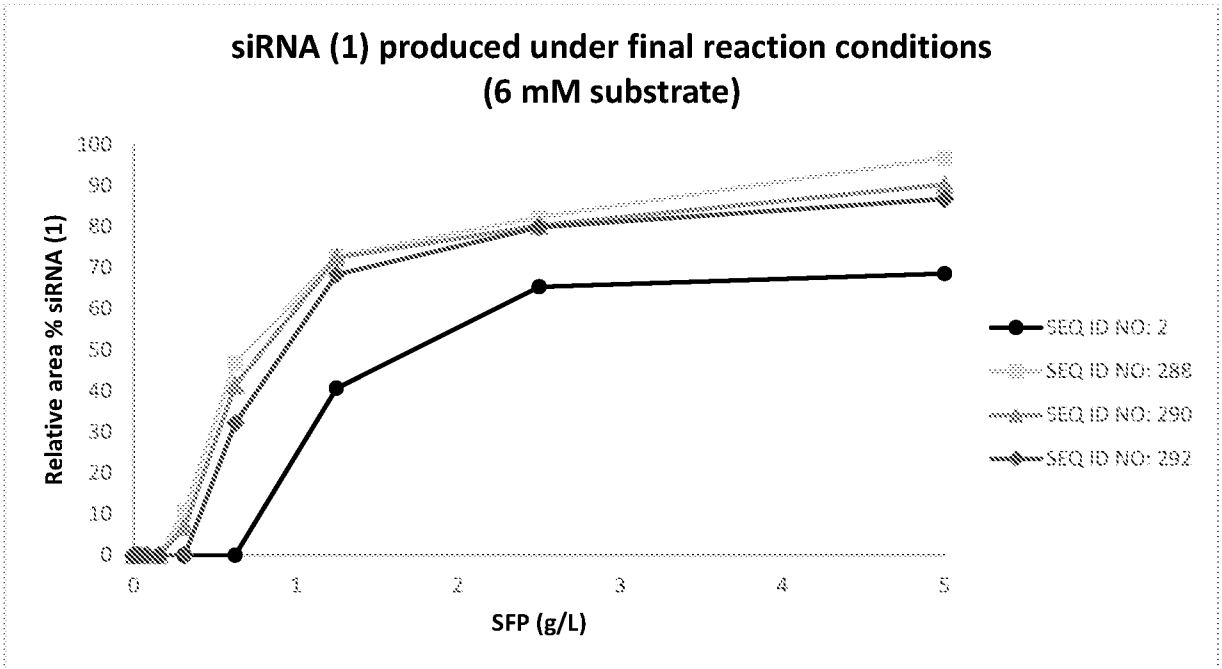


Figure 3A

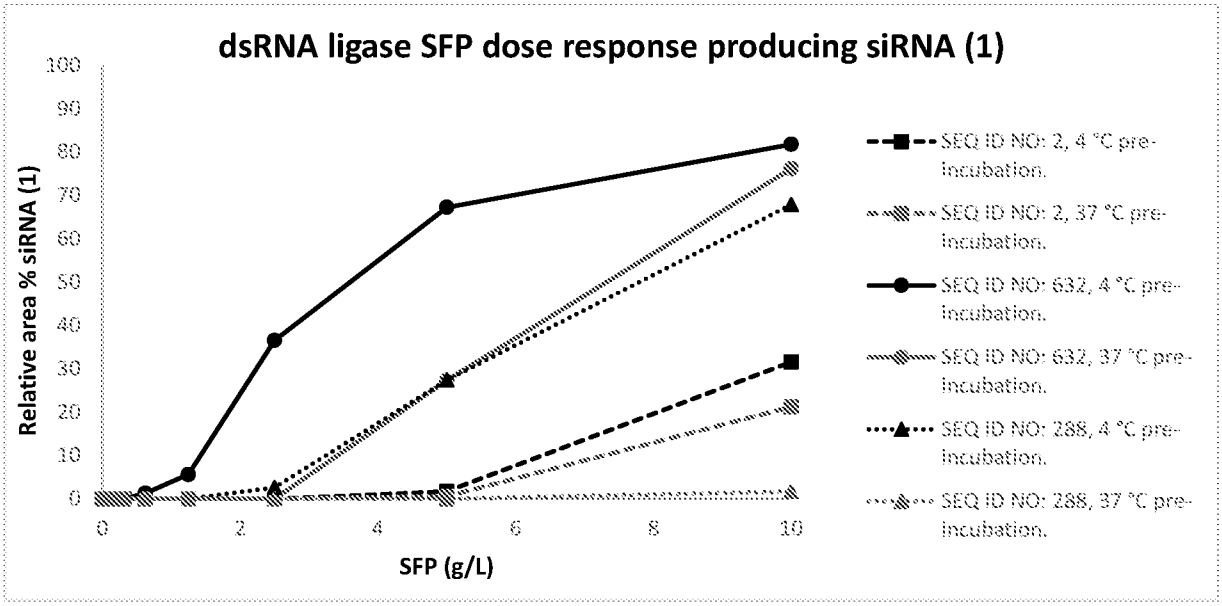
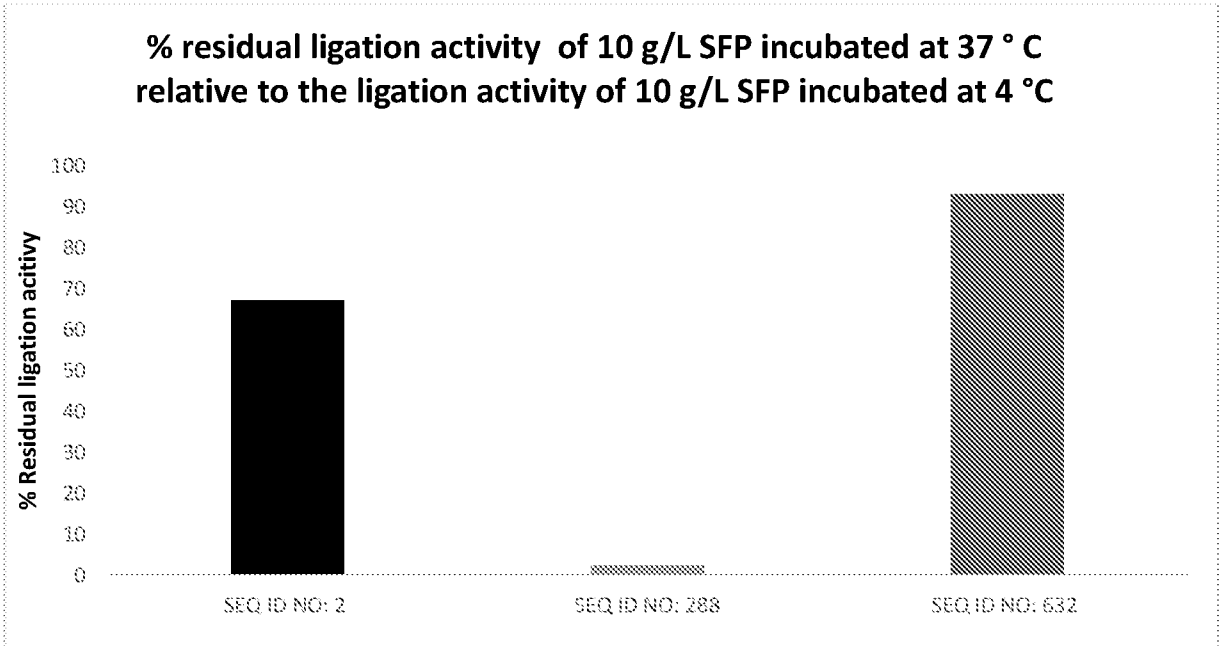


Figure 3B



INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2023/062949

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 3 885 434 A1 (AJINOMOTO KK [JP]) 29 September 2021 (2021-09-29) the whole document in particular SEQ ID NO:1; claims 1-13; paragraph [0002]; paragraphs [0053] to [0054]; -----	1,3-5, 7-18
A	YIN SHENMIN ET AL: "Structure-Function Analysis of T4 RNA Ligase 2", JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 278, no. 20, 1 May 2003 (2003-05-01), pages 17601-17608, XP093045936, US ISSN: 0021-9258, DOI: 10.1074/jbc.M300817200 the whole document ----- -/--	1-18

[x] Further documents are listed in the continuation of Box C.

[x] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 March 2024	10/04/2024

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Strobel, Andreas

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2023/062949

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed.

   b. [ ] furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

      [ ] accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. [ ] With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2023/062949

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DESAI KEVIN K. ET AL: "A tRNA splicing operon: Archease endows RtcB with dual GTP/ATP cofactor specificity and accelerates RNA ligation", NUCLEIC ACIDS RESEARCH , vol. 42, no. 6 16 January 2014 (2014-01-16), pages 3931-3942, XP093045932, GB ISSN: 0305-1048, DOI: 10.1093/nar/gkt1375 Retrieved from the Internet: URL:https://academic.oup.com/nar/article-pdf/42/6/3931/45200473/nar_42_6_3931.pdf the whole document _____	1-18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2023/062949

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 3885434 A1	29-09-2021	CN 113444698 A	28-09-2021
		EP 3885434 A1	29-09-2021
		US 2021301280 A1	30-09-2021