

(12) PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 199948470 B2**
(10) Patent No. **761570**

(54) Title
Methods for generating highly diverse libraries

(51)⁶ International Patent Classification(s)
C12P 021/06 C12P 019/34
C12N 015/74

(21) Application No: 199948470 (22) Application Date: 1999 . 06 . 29

(87) WIPO No: WO00/00632

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/090970	1998 . 06 . 29	US

(43) Publication Date : 2000 . 01 . 17
(43) Publication Journal Date : 2000 . 03 . 30
(44) Accepted Journal Date : 2003 . 06 . 05

(71) Applicant(s)
Phylos, Inc.

(72) Inventor(s)
Richard Wagner; Martin C. Wright; Brent Kreider

(74) Agent/Attorney
GRIFFITH HACK,GPO Box 1285K,MELBOURNE VIC 3001

CORRECTED
VERSION*

PCT

pages 1/2-2/2, drawings, replaced by new pages 1/2-2/2; due to late transmittal by the receiving Office



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12P 21/06, 19/34, C12N 15/74		A1	(11) International Publication Number: WO 00/00632
			(43) International Publication Date: 6 January 2000 (06.01.00)
(21) International Application Number: PCT/US99/14776		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 29 June 1999 (29.06.99)		Published With international search report.	
(30) Priority Data: 60/090,970 29 June 1998 (29.06.98) US			
(71) Applicant: PHYLOS, INC. [US/US]; 128 Spring Street, Lexington, MA 02421 (US).			
(72) Inventors: WAGNER, Richard; 1007 Lowell Road, Concord, MA 01742 (US). WRIGHT, Martin, C.; 812 Memorial Drive #1105, Cambridge, MA 02139 (US). KREIDER, Brent; 4 Davis Road, Bedford, MA 01730 (US).			
(74) Agent: ELBING, Karen; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110-2214 (US).			

IP AUSTRALIA
13 DEC 2000

(54) Title: METHODS FOR GENERATING HIGHLY DIVERSE LIBRARIES

(57) Abstract

Disclosed herein is a method for generating a nucleic acid library, the method involving: (a) providing a population of single-stranded nucleic acid templates, each of the templates including a coding sequence and an operably linked promoter sequence; (b) hybridizing to the population of single-stranded nucleic acid templates a mixture of substantially complementary single-stranded nucleic acid fragments, the fragments being shorter in length than the nucleic acid template; (c) contacting each of the hybridization products of step (b) with both a DNA polymerase which lacks strand displacement activity and a DNA ligase under conditions in which the fragments act as primers for the completion of a second nucleic acid strand which is substantially complementary to the nucleic acid template; and (d) contacting the products of step (c) with RNA polymerase to generate an RNA library, the library being transcribed from the second nucleic acid strand.

* (Referred to in PCT Gazette No. 43/2000, Section II)

METHODS FOR GENERATING
HIGHLY DIVERSE LIBRARIES

5 Background of the Invention

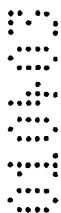
In general, this invention relates to methods for generating and altering recombinant libraries.

10 The ability to isolate a desired nucleic acid or amino acid sequence requires the availability of recombinant libraries of sufficient number and diversity that a particular species is represented in the library and can be identified by one or more screening techniques. Such libraries facilitate the isolation of useful compounds, including therapeutics, research diagnostics,
15 and agricultural reagents, as well as their coding sequences.

 Moreover, desirable libraries may be specifically designed to contain large numbers of possible variants of a single compound. This type of library may be used to
20 screen for improved versions of the compound, for example, for a compound variant having optimized therapeutic efficacy.



 For these or any other application, general approached for increasing library diversity are very
25 useful and represent an important focus of the protein design industry.



 All references, including any patents or patent applications, cited in this specification are hereby incorporated by reference. No admission is made that any
30 reference constitutes prior art. The discussion of the references states what their authors assert, and the applicants reserve the right to challenge the accuracy and pertinency of the cited documents. It will be clearly understood that, although a number of prior art
35 publications are referred to herein, this reference does not constitute an admission that any of these documents forms part of the common general knowledge in the art, in

Australia or in any other country.

For the purposes of this specification it will be clearly understood that the word "comprising" means "including but not limited to", and that the word
5 "comprises" has a corresponding meaning.

Summary of the Invention

In general, the present invention features a method
10 for generating a nucleic acid library, the method involving: (a) providing a population of single-stranded nucleic acid templates, each of the templates including a coding sequence and an operably linked promoter sequence;
(b) hybridising to the population of single-stranded
15 nucleic acid templates a mixture of substantially



-2-

complementary single-stranded nucleic acid fragments, the fragments being shorter in length than the nucleic acid template; (c) contacting each of the hybridization products of step (b) with both a DNA polymerase which lacks strand displacement activity and a DNA ligase under conditions in which the fragments act as primers for the completion of a second nucleic acid strand which is substantially complementary to the nucleic acid template; and (d) contacting the products of step (c) with RNA polymerase to generate an RNA library, the library being transcribed from the second nucleic acid strand.

In preferred embodiments, the method is used to introduce one or more mutations into the library; the mixture of substantially complementary single-stranded nucleic acid fragments is generated by cleaving a double-stranded nucleic acid molecule; the mixture of substantially complementary single-stranded nucleic acid fragments is generated by synthesis of random oligonucleotides; the single-stranded nucleic acid template is generated using an M13 phage carrying the nucleic acid, by digestion of one strand of a double-stranded nucleic acid template using gene VI exonuclease or lambda exonuclease, by capture of a biotinylated single nucleic acid strand using streptavidin, or by reverse transcription of RNA; the mixture of substantially complementary single-stranded nucleic acid fragments includes at least about 100 different species of nucleic acid fragments; step (b) is carried out using between 1 and approximately 1000 fragments per single-stranded nucleic acid template; a single strand of the product of step (c) is used as a nucleic acid template and steps (b) and (c) are repeated; steps (b) and (c) are repeated, using, in each round, the product of step (c) as the nucleic acid template; the method further involves providing one or more single-stranded nucleic acid fragments which form a homoduplex with the single-stranded nucleic acid template and carrying out step (b) in the presence of the homoduplex-forming fragments; the

-3-

promoter is a T7 promoter; the DNA polymerase is T4 DNA polymerase; the method further involves amplifying the product of step (c) prior to said contacting step (d); the method further involves the step of: (e) translating the RNA library to generate a protein library; the method further involves the step
5 of: (e) linking to the 3' terminus of the coding sequence of each of substantially all of the members of the RNA library an amino acid acceptor molecule; and the method further involves the step of: (f) translating the RNA library to generate an RNA-protein fusion library.

In a second aspect, the invention features a method for reducing
10 sequence variation in a population of nucleic acid molecules, the method involving: (a) providing a first population of single-stranded nucleic acid templates of varying sequence, each of substantially all of the templates including a coding sequence and an operably linked promoter sequence; (b) hybridizing to the members of the first population a second population of
15 substantially complementary single-stranded nucleic acid fragments, the fragments being shorter in length than the nucleic acid template and the fragments being of substantially identical sequence; (c) contacting the hybridization products of step (b) with both a DNA polymerase which lacks strand displacement activity and a DNA ligase under conditions in which the
20 fragments act as primers for the completion of a second nucleic acid strand which is substantially complementary to the nucleic acid template; and (d) contacting the products of step (c) with RNA polymerase to generate a population of RNA molecules, the population of RNA molecules being transcribed from the second nucleic acid strand and having reduced sequence
25 variation relative to the first population of single-stranded nucleic acid templates.

In preferred embodiments, the method is used to remove one or more

-4-

mutations from the first population of single-stranded nucleic acid templates; step (b) involves hybridization of the first population of single-stranded nucleic acid templates to two or more different populations of substantially complementary single-stranded nucleic acid fragments; the second population

5 of substantially complementary single-stranded nucleic acid fragments is generated by cleaving a double-stranded nucleic acid molecule; the second population of substantially complementary single-stranded nucleic acid fragments is generated by synthesis of random oligonucleotides; the single-stranded nucleic acid template is generated using an M13 phage carrying the

10 nucleic acid, by digestion of one strand of a double-stranded nucleic acid template using gene VI exonuclease or lambda exonuclease, by capture of a biotinylated single nucleic acid strand using streptavidin, or by reverse transcription of RNA; step (b) is carried out using between 1 and approximately 1000 single-stranded nucleic acid fragments per single-stranded nucleic acid

15 template; a single strand of the product of step (c) is used as a nucleic acid template and steps (b) and (c) are repeated; steps (b) and (c) are repeated, using, in each round, the product of step (c) as the nucleic acid template; the promoter is a T7 promoter; the DNA polymerase is T4 DNA polymerase; the method further involves amplifying the product of step (c) prior to said contacting step

20 (d); the method further involves the step of: (e) translating the population of RNA molecules to generate a protein library; the method further involves the step of: (e) linking to the 3' terminus of the coding sequence of each of substantially all of the members of the population of RNA molecules an amino acid acceptor molecule; and the method further involves the step of: (f)

25 translating the population of RNA molecules to generate an RNA-protein fusion library.

In a third aspect, the invention features a method for generating a

-5-

nucleic acid library, the method involving: (a) providing a population of single-stranded nucleic acid templates, each of the templates including a coding sequence; (b) providing a population of single-stranded nucleic acid molecules of varying sequence, the population of single-stranded nucleic acid templates and the population of single-stranded nucleic acid molecules of varying sequence being substantially complementary; (c) hybridizing the population of single-stranded nucleic acid templates with the population of single-stranded nucleic acid molecules of varying sequence under conditions sufficient to form duplexes; and (d) contacting the duplexes with one or more excision/repair enzymes under conditions that allow the enzymes to correct mismatched base pairs in the duplexes.

In preferred embodiments, the method further involves providing a population of single-stranded templates derived from the product of step (d) and repeating steps (c) and (d); and the steps (c) and (d) are repeated, using, in each round, a population of single-stranded templates derived from the product of step (d).

In a fourth aspect, the invention features a method for generating a nucleic acid library, the method involving: (a) providing a population of single-stranded nucleic acid templates, each of the templates including a coding sequence; (b) hybridizing to the population of single-stranded nucleic acid templates a mixture of substantially complementary single-stranded nucleic acid fragments, the fragments being shorter in length than the nucleic acid template; (c) contacting each of the hybridization products of step (b) with both a DNA polymerase which lacks strand displacement activity and a DNA ligase under conditions in which the fragments act as primers for the completion of a second nucleic acid strand which is substantially complementary to the nucleic acid template; and (d) contacting the products of step (c) with one or more

-6-

excision/repair enzymes under conditions that allow the enzymes to correct mismatched base pairs in the products.

In preferred embodiments, the method further involves providing a population of single-stranded templates derived from the product of step (d) and repeating steps (b) - (d); and steps (b) - (d) are repeated, using, in each round, a population of single-stranded templates derived from the product of step (d).

In preferred embodiments of the third and fourth aspects of the invention, the contacting with the excision/repair enzymes is carried out in vivo (for example, in a bacterial cell); the contacting with the excision/repair enzymes is carried out in vitro; the single-stranded nucleic acid template is generated using an M13 phage carrying the nucleic acid, by digestion of one strand of a double-stranded nucleic acid template using gene VI exonuclease or lambda exonuclease, by capture of a biotinylated single nucleic acid strand using streptavidin, or by reverse transcription of RNA; step (b) is carried out using between 1 and approximately 1000 single-stranded nucleic acid molecules of varying sequence or single-stranded nucleic acid fragments per single-stranded nucleic acid template; the method further involves the step of: (e) amplifying the product of step (d); each of the coding sequences is operably linked to a promoter sequence; the method further involves the step of: (e) transcribing the products of step (d) to generate an RNA library; the method further involves the step of: (f) translating the RNA library to generate a protein library; the method further involves the step of: (f) linking to the 3' terminus of the coding sequence of each of substantially all of the members of the RNA library an amino acid acceptor molecule; and the method further involves the step of: (g) translating the RNA library to generate an RNA-protein fusion library.

-7-

As used herein, by a "library" is meant at least 10^8 , preferably, at least 10^{10} , more preferably, at least 10^{12} , and, most preferably, at least 10^{14} molecules having a nucleic acid and/or an amino acid component.

By a "mixture" of nucleic acid fragments is meant at least 100,
5 preferably, at least 500, more preferably, at least 1000, and, most preferably, at least 1500 nucleic acid fragments.

By a "promoter sequence" is meant any nucleic acid sequence which provides a functional RNA polymerase binding site and which is sufficient to allow transcription of a proximal coding sequence.

10 By "substantially complementary" is meant that a nucleic acid strand possesses a sufficient number of nucleotides which are capable of forming matched Watson-Crick base pairs with a second nucleic acid strand to produce one or more regions of double-strandedness between the two nucleic acids. It will be understood that each nucleotide in a nucleic acid molecule need not
15 form a matched Watson-Crick base pair with a nucleotide in an opposing strand to be substantially complementary, and that in a "mixture of substantially complementary single-stranded nucleic acid fragments," a significant fraction of the fragments will contain one or more nucleotides which form mismatches with the "single-stranded nucleic acid template."

20 By "strand displacement activity" is meant the ability of a polymerase or its associated helicase to disrupt base pairing between two nucleic acid strands.

By "mutation" is meant any nucleotide change and includes sequence alterations that result in phenotypic differences as well as changes
25 which are silent.

By "duplex" is meant is meant a structure formed between two annealed nucleic acid strands in which sufficient sequence complementarity

-8-

exists between the strands to maintain a stable hybridization complex. A duplex may be either a "homoduplex," in which all of the nucleotides in the first strand appropriately base pair with all of the nucleotides in the second opposing strand, or a heteroduplex. By a "heteroduplex" is meant a structure
5 formed between two annealed strands of nucleic acid in which one or more nucleotides in the first strand do not or cannot appropriately base pair with one or more nucleotides in the second opposing complementary strand because of one or more mismatches. Examples of different types of heteroduplexes include those which exhibit an exchange of one or several nucleotides, and
10 insertion or deletion mutations.

By "random oligonucleotides" is meant a mixture of oligonucleotides having sequence variation at one or more nucleotide positions. Random oligonucleotides may be produced using entirely random or partially random synthetic approaches or by intentionally altering an oligonucleotide in a
15 directed fashion.

By an "amino acid acceptor molecule" is meant any molecule capable of being added to the C-terminus of a growing protein chain by the catalytic activity of the ribosomal peptidyl transferase function. Typically, such molecules contain (i) a nucleotide or nucleotide-like moiety (for example,
20 adenosine or an adenosine analog (di-methylation at the N-6 amino position is acceptable)), (ii) an amino acid or amino acid-like moiety (for example, any of the 20 D- or L-amino acids or any amino acid analog thereof (for example, O-methyl tyrosine or any of the analogs described by Ellman et al., Meth. Enzymol. 202:301, 1991)), and (iii) a linkage between the two (for example, an
25 ester, amide, or ketone linkage at the 3' position or, less preferably, the 2' position); preferably, this linkage does not significantly perturb the pucker of the ring from the natural ribonucleotide conformation. Amino acid acceptors

may also possess a nucleophile, which may be, without limitation, an amino group, a hydroxyl group, or a sulfhydryl group. In addition, amino acid acceptors may be composed of nucleotide mimetics, amino acid mimetics, or mimetics of a combined nucleotide-amino acid structure.

5 By an amino acid acceptor being linked "to the 3' terminus" of a coding sequence is meant that the amino acid acceptor molecule is positioned after the final codon of that coding sequence. This term includes, without limitation, an amino acid acceptor molecule that is positioned precisely at the 3' end of the coding sequence as well as one which is separated from the final
10 codon by intervening coding or non-coding sequence (for example, a sequence corresponding to a pause site). This term also includes constructs in which coding or non-coding sequences follow (that is, are 3' to) the amino acid acceptor molecule. In addition, this term encompasses, without limitation, an amino acid acceptor molecule that is covalently bonded (either directly or
15 indirectly through intervening nucleic acid sequence) to the coding sequence, as well as one that is joined to the coding sequence by some non-covalent means, for example, through hybridization using a second nucleic acid sequence that binds at or near the 3' end of the coding sequence and that itself is bound to an amino acid acceptor molecule.

20 By an "RNA-protein" fusion is meant any molecule that includes a ribonucleic acid covalently bonded through an amide bond to a protein. This covalent bond is resistant to cleavage by a ribosome.

By a "protein" is meant any two or more naturally occurring or modified amino acids joined by one or more peptide bonds. "Protein,"
25 "peptide," and "polypeptide" are used interchangeably herein.

By "a population of single-stranded templates of varying sequence" is meant that the nucleic acid species of the population possess sequences

which differ at one or more nucleotide positions.

By "excision/repair enzymes" is meant any combination of enzymes sufficient to replace a mismatched base pair or loop with a standard base pair (i.e., A:T or G:C).

5 For the purposes of this specification it will be clearly understood that the word "comprising" means "including but not limited to", and that the word "comprises" has a corresponding meaning.

10 It is to be understood that, if any prior art publication is referred to herein, such reference does not constitute an admission that the publication forms a part of the common general knowledge in the art, in Australia or any other country.

15 BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a schematic representation of an exemplary fragment recombination method for generating highly diverse RNA-protein fusion libraries. In this method, the fragments are derived from a double-stranded DNA molecule into which sequence variation is introduced.

20 FIGURE 2 is a schematic representation of the initial steps of a second exemplary fragment recombination method. In this method, the fragments are synthetic oligonucleotides into which sequence variation is introduced.

DETAILED DESCRIPTION

30 The present invention involves a number of novel and related methods for the random recombination of nucleic acid sequences, facilitating the generation of DNA, RNA, and protein libraries into which genetic alterations have been introduced. As described in more detail below, in one preferred embodiment, this technique is carried out *in vitro* and is used to generate traditional protein libraries or RNA-protein fusion libraries, either of which

may then be used in combination with any of a variety of methods for the selection of desired proteins or peptides (or their corresponding coding sequences) from library populations. This general approach provides a means for
5 the introduction of mutations into protein libraries in an unbiased fashion and also provides a technique by which unfavorable mutations may be removed

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
22

-11-

from a library or selected pool, or "backcrossed" out of a population of molecules during subsequent rounds of selection.

Fragment Recombination

According to one preferred method of the invention, a library is
5 generated by the production of mutant fragments and the random recombination of these fragments with an unmutated (typically, wild-type) sequence. One example of this general approach is shown in Figure 1. As indicated in this figure, mutations are first randomly introduced into an initial double-stranded DNA sequence (termed "dsDNA(init)"). This produces a
10 population of mutant double-stranded DNA sequences, which, in Figure 1, is termed "dsDNA(mut)." These mutations may be introduced by any technique, including PCR mutagenesis (which relies on the poor error-proofing mechanism of Taq polymerase), site-directed mutagenesis, or template-directed mutagenesis (for example, as described in Joyce and Inoue, Nucl. Acids Res.
15 17:171, 1989. The DNA in this mutation-containing population is subsequently fragmented using any of a variety of standard methods. For example, the DNA may be partially degraded using one or more nucleases (such as DNase I, micrococcal nuclease, restriction endonucleases, or P1 nuclease), or may be fragmented chemically using, for example, Fe•EDTA. Alternatively, mutation-
20 containing fragments may be generated by limited nucleotide consumption during polymerization (for example, during PCR amplification), or by simple physical shearing (for example, by sonication). Preferable fragment sizes range from 25-1000 base pairs and are most preferably in the range of 50-150 base pairs.

25 The DNA fragments are then heated and subsequently annealed to a full-length single-stranded DNA template which is identical to the initial DNA

-12-

in sequence and which is the non-coding (or minus) strand of that DNA. In addition, in this hybridization mixture is included a second type of fragment, sometimes referred to as a "terminator fragment" (Joyce and Inoue, Nucl. Acids Res. 17:171, 1989). This terminator fragment is complementary to the 3' end of the single-stranded template and provides a polymerization primer which binds to the template in a manner that is relatively independent of the number or nature of the randomly annealed, mutation-containing fragments.

Single-stranded templates may be generated by any standard technique, for example, by using an M13 phage carrying the DNA sequence, by digestion of the coding strand of a dsDNA(init) molecule using gene VI exonuclease (Nikiforov et al, PCR Methods Appl. 3:285, 1994) or lambda exonuclease (Higuchi and Ochman, Nucl. Acids Res 17: 5865, 1989), by capture of a biotinylated DNA strand using immobilized streptavidin (Joyce and Inoue, Nucl. Acids Res. 17:171, 1989), or by reverse transcription of RNA. To carry out the template-fragment hybridization, templates are mixed with fragments using no less than one fragment molecule per template molecule and no more than approximately 1000 fragment molecules per template molecule. A low ratio of fragments to templates produces progeny strands that closely resemble the templates, whereas a higher ratio produces progeny that more closely resemble the fragments. Hybridization conditions are determined by standard methods and are designed to allow for the formation of heteroduplexes between the template and the fragments. Exemplary hybridization techniques are described, for example, in Stemmer. U.S. Patent No. 5,605,793.

Once annealed to the template, the fragments are joined together by treating with both a DNA polymerase that lacks strand displacement activity and a DNA ligase. DNA polymerases useful for this purpose include, without limitation, T4 DNA polymerase and reconstituted DNA pol II from *E. coli* (see,

-13-

for example, Hughes et al., J. Biol. Chem. 266:4568, 1991). Any DNA ligase (for example, T4 DNA ligase) may be utilized. In this step, the DNA duplexes may be treated first with the DNA polymerase and then with the DNA ligase, or with both enzymes simultaneously, and the step may be carried out, for example, as described in Joyce and Inoue (Nucl. Acids Res. 17:711, 1989). As shown in Figure 1, this step generates a population of double-stranded DNAs (termed "dsDNA(lib)), each member of which includes one strand typically having one or more introduced mutations. Because both the mutations initially introduced and the number and nature of the fragments annealed are random, different duplexes in the population contain different mutant sequences.

An alternative to this general approach for generating a double-stranded DNA library is shown in Figure 2. By this alternative approach, single-stranded oligonucleotide fragments are synthesized which correspond to portions of the coding strand of an initial double-stranded DNA molecule. These oligonucleotide fragments preferably range from 5-2000 nucleotides, and most preferably range from 20-100 nucleotides in length and are generated, for example, using any standard technique of nucleic acid synthesis. These oligonucleotides may be synthesized with completely random or semi-random mutations by any standard technique. Preferably, such oligonucleotides include up to 3 introduced mismatches per 20 nucleotide segment and are devoid of in frame stop codons. In addition, in certain cases, it may be desirable or necessary to increase the hybridization potential of the oligonucleotide through the introduction of non-natural, affinity-enhancing base pairs, such as C-5 propyne uridine or C-5 propyne cytidine. These techniques are described, for example, in Wagner et al., Science 260:1510, 1993.

These mutation-containing oligonucleotide fragments are next annealed to single-stranded templates which, as above, are full-length strands

-14-

identical in sequence to the non-coding (or minus) strand of the initial DNA. The fragments are joined together using DNA polymerase and DNA ligase, also as described above, to create a double-stranded DNA library (dsDNA(lib)). Again, this library contains a population of duplex molecules, containing an
5 array of different coding strands having mutations which differ in number, position, and identity.

If desired, the above steps may be repeated, for either the fragment or the oligonucleotide approach, to introduce varying numbers of mutations into a DNA molecule. In particular, the mutated strands become the initial single-
10 stranded templates, and mutant fragments or oligonucleotides are annealed to those strands and polymerized and ligated.

Methods for Backcrossing

As generally described above, the methods of the invention are used to introduce mutations into an initial DNA sequence. In addition, these
15 techniques may be used to remove or reduce in frequency undesirable mutations from a DNA library. According to this approach, following fragmentation of the dsDNA(mut), oligonucleotides of wild-type sequence or specific fragments of unmutated or wild-type DNA (wtDNA) may be added to the single-stranded template together with the dsDNA(mut) fragments or
20 oligonucleotides. The fragments are strand-separated (if necessary), annealed to the full-length single-stranded template, and joined together using DNA polymerase and DNA ligase, as described above. The use of a high concentration of unmutated oligonucleotide or fragment, relative to the corresponding mutant fragment, allows for the generation of libraries in which
25 undesirable mutations are minimized or eliminated.

In addition, this approach may be used with existing mutation-

-15-

containing libraries to similarly decrease or eliminate undesirable sequences. This approach involves an initial library having mutant sequences and the annealing, polymerization, and ligation of fragments or oligonucleotides of wild-type sequence, as generally described above.

5 RNA, Protein, and RNA-Protein Libraries

In one preferred embodiment of the invention, the DNA libraries described above further include an RNA polymerase binding site, for example, for T7 or SP6 polymerase. Such binding sites are described, for example, in Milligan et al., Proc. Natl. Acad. Sci. USA 87:696, 1990. This site is
10 positioned upstream of the coding sequence at a location which allows for transcription of the sequence. Typically such sites are located at between 5-2000 base pairs upstream of the coding sequence.

Libraries containing RNA polymerase binding sites may be altered as described above. Following polymerization and ligation, the dsDNA(lib)
15 may be transcribed directly, for example, using an *in vitro* transcription system, to generate an RNA library. Alternatively, the dsDNA(lib) may be transcribed and translated directly, for example, using *in vitro* transcription and translation systems, to generate a protein library. Exemplary *in vitro* transcription systems and *in vitro* translation systems include T7 transcription systems, and rabbit
20 reticulocyte, wheat germ, yeast, and *E. coli* translation systems.

If desired, the number of copies of each RNA or protein in the library may be increased by including a strand-specific amplification step prior to transcription. For example, PCR amplification may be carried out by incorporating unique primer-binding sequences are incorporated into the
25 mutant strand during the polymerization and ligation steps. These sequences may be incorporated as either mismatches or sequence extensions at one or

-16-

both ends of the DNA, allowing amplification of the newly-synthesized strand without amplification of the template strand. Alternatively, linear amplification can be achieved by multiple cycles of annealing and extension of a single oligonucleotide primer that is complementary to the 3' end of the

5 newly-synthesized strand. Subsequent PCR and transcription steps produce a majority of RNA corresponding to mutant sequences with only a small proportion of template-derived sequences.

In one preferred approach, the above methods for introducing mutations or for backcrossing out undesirable mutations may be used to
10 produce highly diverse RNA-protein libraries. Such libraries may be constructed by ligating linkers containing a non-hydrolyzable amino acid acceptor molecule, such as puromycin, to the 3' termini of the RNAs in a library (for example, produced as described above). Exemplary techniques for generating RNA-protein fusions are described, for example, in Szostak et al.,
15 U.S.S.N. 09/007,005; and Roberts et al., Proc. Natl. Acad. Sci. USA 94:12297, 1997. Subsequent translation of these RNAs generates a library of RNA-protein fusion molecules that may subsequently be used in *in vitro* selection experiments.

In addition, if desired, RNA or RNA-protein fusion molecules, once
20 selected, may be used as templates in standard PCR reactions to obtain the corresponding coding sequence. Thus, this method provides a means for carrying out fragment recombination, molecular backcrossing, selection of proteins and/or peptides, and selection of their corresponding coding sequences, all in an *in vitro* system.

25 Excision/Repair

In addition to fragment recombination approaches, excision/repair

-17-

may also be used to alter library sequences. This approach may be used to generate DNA, RNA, and RNA-protein fusion libraries. This technique relies on the fact that the dsDNA(lib)s, produced by any of the methods described above, by their nature, contain a certain number of mismatched base pairs. To
5 generate diversity in the library sequences, these mismatches are repaired in vitro by excision/repair enzymes. This may be carried out using any excision repair system (for example, as described in Jaiswal et al., Nucl. Acids Res. 26:2184, 1998; or Fortini et al., Biochemistry 37:3575, 1998.

Alternatively, the excision/repair step may be carried out by
10 transforming a dsDNA(lib) into a bacterial or yeast strain and exploiting the bacterial or yeast repair systems in vivo. Again, this step may be carried out by transforming the library into any standard in vivo excision/repair system. Exemplary systems are described, without limitation, in Campbell et al., Mutat. Res. 211:181, 1989; Bishop and Kolodner, Mol. Cell Biol. 6:3401, 1986; Fishel
15 et al., J. Mol. Biol. 188:147, 1986; and Westmoreland et al., Genetics 145:29, 1997.

Because the above repair processes are random, this excision/repair method sometimes results in the introduction of mutations into a library sequence and at other times results in the backcrossing of wild-type sequence
20 alterations into the coding strand.

In an alternative to the above approaches, in vitro or in vivo excision/repair may also be used directly to generate diverse libraries using as a substrate a mixture of dsDNA(mut) (for example, produced as described above) and dsDNA(init) or wtDNA. In this technique, the mixture is strand-separated
25 and reannealed, and is then either incubated in vitro with excision/repair enzymes or transformed into bacteria to utilize the bacterial excision/repair system (for example, as described above). In this manner, mutations may be

-18-

randomly introduced into a sequence, and wild-type sequences may be backcrossed into dsDNA(mut) molecules.

What is claimed is:

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A method for generating a nucleic acid library, said
5 method comprising the steps of:
- (a) providing a population of single-stranded
nucleic acid templates, each of said templates comprising
a coding sequence and an operably linked promoter
sequence;
 - 10 (b) hybridizing to said population of single-
stranded nucleic acid templates a mixture of substantially
complementary single-stranded nucleic acid fragments, said
fragments being shorter in length than said nucleic acid
templates;
 - 15 (c) contacting each of the hybridization products of
step (b) with both a DNA polymerase which lacks strand
displacement activity and a DNA ligase under conditions in
which said fragments act as primers for the completion of
second nucleic acid strands which are substantially
20 complementary to said nucleic acid templates; and
 - (d) contacting the products of step (c) with RNA
polymerase to generate an RNA library, said library being
transcribed from said second nucleic acid strands.
- 25 2. A method according to claim 1, wherein said method is
used to introduce one or more mutations into said library.
3. A method for reducing sequence variation in a
population of nucleic acid molecules, said method
30 comprising:
- (a) providing a first population of single-stranded
nucleic acid templates of varying sequence, each of
substantially all of said templates comprising a coding
sequence and an operably linked promoter sequence;
 - 35 (b) hybridizing to the members of said first

population a second population of single-stranded nucleic acid fragments, said fragments being shorter in length than said nucleic acid templates and said fragments being substantially complementary to said nucleic acid templates, wherein a subpopulation of said second population comprises a substantially identical sequence and hybridizes to a region of varying sequence;

(c) contacting the hybridization products of step (b) with both a DNA polymerase which lacks strand displacement activity and a DNA ligase under conditions in which said fragments act as primers for the completion of second nucleic acid strands which are substantially complementary to said nucleic acid templates; and

(d) contacting the products of step (c) with RNA polymerase to generate a third population of RNA molecules, said third population of RNA molecules being transcribed from said second nucleic acid strands and having reduced sequence variation relative to said first population of single-stranded nucleic acid templates.

4. A method according to claim 3, wherein said method is used to remove one or more mutations from said first population of single-stranded nucleic acid templates.

5. A method according to any one of claims 1 to 4, wherein said mixture of substantially complementary single-stranded nucleic acid fragments is generated by cleaving double-stranded nucleic acid molecules or by synthesis of random oligonucleotides.

6. A method according to any one of claims 1 to 5, wherein said mixture of substantially complementary single-stranded nucleic acid fragments comprises at least about 100 different species of nucleic acid fragments.

0004

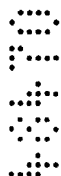
0004

7. A method according to any one of claims 1 to 6,
further comprising generating complementary strands from
said second nucleic acid strands, providing said
complementary strands as said population or said first
5 population of single-stranded nucleic acid templates, and
repeating steps (b) and (c) for one or more rounds, and
then repeating step (d).

8. A method according to any one of claims 1 to 7,
10 wherein said method further comprises providing one or
more single-stranded nucleic acid fragments which form a
homoduplex with said single-stranded nucleic acid
templates and carrying out step (b) in the presence of
said homoduplex-forming fragments.

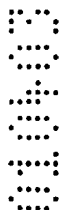
15 9. A method according to any one of claims 1 to 8,
wherein said promoter sequence is a T7 promoter.

10 10. A method according to any one of claims 1 to 9,
20 wherein said DNA polymerase is T4 DNA polymerase.



25 11. A method according to any one of claims 1 to 10,
wherein said method further comprises amplifying said
product of step (c) prior to said contacting step (d).

12. A method according to any one of claims 1 to 11,
wherein said method further comprises the step of:
(e) translating said RNA library to generate a
protein library.



30 13. A method according to any one of claims 1 to 11,
wherein said method further comprises the step of:
(e) linking an amino acid acceptor molecule to the 3'
terminus of said coding sequence of each of substantially
35 all of the members of said RNA library; and optionally

(f) translating said RNA library to generate an RNA-protein fusion library.

14. A method for generating a nucleic acid library, said
5 method comprising the steps of:

(a) providing a first population of single-stranded nucleic acid molecules wherein said molecules are templates, each of said templates comprising a coding sequence;

10 (b) providing a second population of single-stranded nucleic acid molecules of varying sequence, said first population of single-stranded nucleic acid templates and said second population of single-stranded nucleic acid molecules of varying sequence being substantially
15 complementary;

(c) hybridizing said first population of single-stranded nucleic acid templates with said second population of single-stranded nucleic acid molecules of varying sequence under conditions sufficient to form
20 duplexes; and

(d) contacting said duplexes with one or more excision/repair enzymes under conditions that allow said enzymes to correct mismatched base pairs in said duplexes.

25 15. A method according to claim 14, wherein said method further comprises, after step (d), providing a new population of single-stranded nucleic acid templates comprising a coding sequence, said templates being derived from the product of step (d), substituting said new
30 population of templates for said first population of templates, and repeating steps (b), (c), and (d) for one or more rounds.

16. A method for generating a nucleic acid library, said
35 method comprising:

(a) providing a population of single-stranded nucleic

acid templates, each of said templates comprising a coding sequence;

5 (b) hybridizing to said population of single-stranded nucleic acid templates a mixture of substantially complementary single-stranded nucleic acid fragments, said fragments being shorter in length than said nucleic acid templates;

10 (c) contacting each of the hybridization products of step (b) with both a DNA polymerase which lacks strand displacement activity and a DNA ligase under conditions in which said fragments act as primers for the completion of a second nucleic acid strand which is substantially complementary to said nucleic acid templates; and

15 (d) contacting the products of step (c) with one or more excision/repair enzymes under conditions that allow said enzymes to correct mismatched base pairs in said products.

20 17. A method according to claim 16, wherein said method further comprises the steps of: providing a new population of single-stranded nucleic acid templates comprising a coding sequence and derived from the product of step (d), substituting said new population of templates for said population of single-stranded nucleic acid templates and
25 repeating steps (b)-(d) for one or more rounds.

18. A method according to any one of claims 14 to 17, wherein said contacting with said excision/repair enzymes is carried out *in vivo* or *in vitro*.

30 19. A method according to any one of claims 14 to 18, wherein said contacting with said excision/repair enzymes is carried out in a bacterial cell.

35 20. A method according to any one of claims 1 to 19,

wherein said single-stranded nucleic acid templates are generated using an M13 phage carrying said nucleic acid templates, by digestion of one strand of double-stranded nucleic acid template using gene VI exonuclease or lambda exonuclease, by capture of biotinylated single nucleic acid strands using streptavidin, or by reverse transcription of RNA.

21. A method according to any one of claims 1 to 20, wherein step (b) is carried out using between 1 and approximately 1000 single-stranded nucleic acid molecules of varying sequence or single-stranded nucleic acid fragments per each of said single-stranded nucleic acid template.

22. A method according to any one of claims 14 to 21, wherein said method further comprises the step of:
(e) amplifying said product of step (d).

23. A method according to any one of claims 14 to 22, wherein each of said coding sequences is operably linked to a promoter sequence.

24. A method according to claim 21, wherein said method further comprises the step of:
(e) transcribing the products of step (d) to generate an RNA library.

25. A method according to any one of claims 14 to 24, wherein said method further comprises the step of:
(f) translating said RNA library to generate a protein library.

26. A method according to any one of claims 14 to 23, wherein said method further comprises the step of:

(f) linking an amino acid acceptor molecule to the 3' terminus of said coding sequence of each of substantially all of the members of said RNA library; and optionally

(g) translating said RNA library to generate an RNA-
5 protein fusion library.

27. A method according to any one of claims 1, 3, 14 or 16, substantially as herein described with reference to the examples and figures.

10

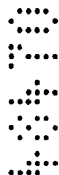
Dated this 31st day of March 2003

PHYLOS, INC.

By their Patent Attorneys

15 GRIFFITH HACK

Fellows Institute of Patent and
Trade Mark Attorneys of Australia



1/2

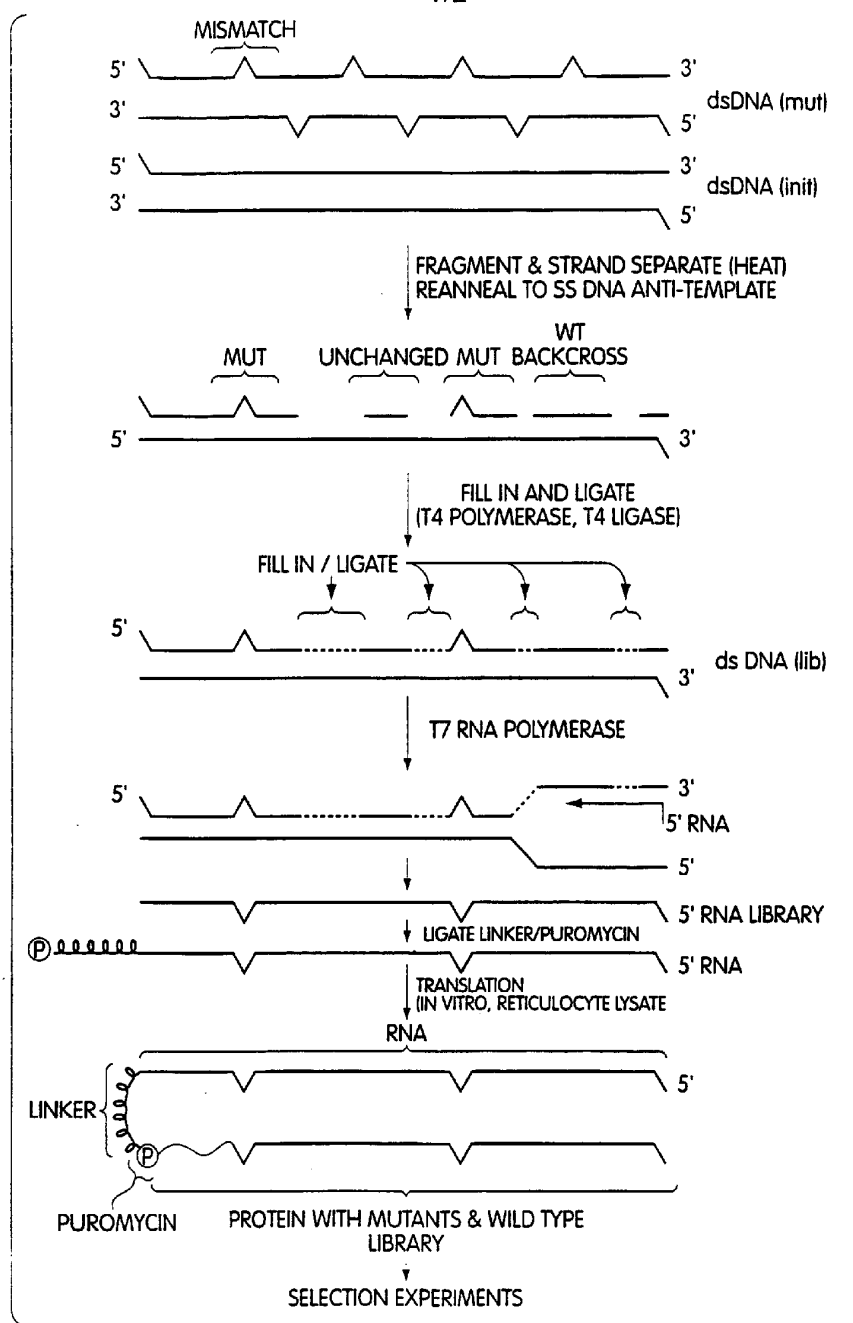


Fig.1

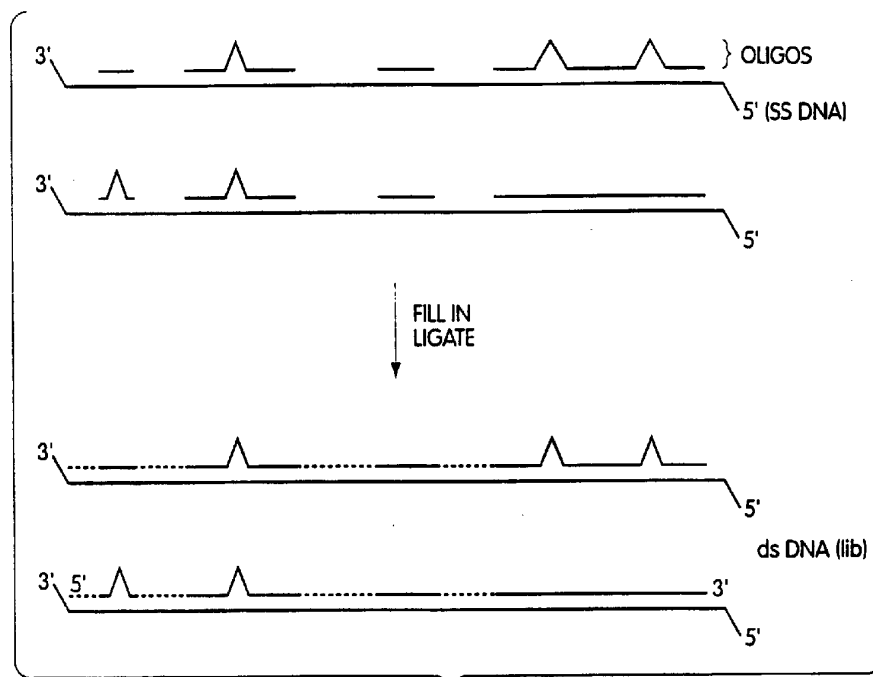


Fig. 2