

\*\*\*\*\* NOT. OF GRANT AWARD \*\*\*\*\*  
EXPLORARY/DEVELOPMENT GRANT Issue Date: 05/18/2006

Department of Health and Human Services  
National Institutes of Health  
NATIONAL HUMAN GENOME RESEARCH INSTITUTE

\*\*\*\*\*  
Grant Number: 1 R21 HG003977-01  
Principal Investigator: ROMESBERG, FLOYD E PHD  
Project Title: Evolving Novel Polymerases for Genome Sequencing

VP, OFFICE OF SPONSORED PROGRAMS  
SCRIPPS RESEARCH INSTITUTE  
10550 NORTH TORREY PINES ROAD  
LA JOLLA, CA 92037  
UNITED STATES  
Award e-mailed to: grants@scripps.edu

Budget Period: 05/19/2006 - 04/30/2007  
Project Period: 05/19/2006 - 04/30/2008

Dear Business Official:

The National Institutes of Health hereby awards a grant in the amount of \$278,850 (see  
'Award Calculation' in Section I) to SCRIPPS RESEARCH INSTITUTE in support of the above  
referenced project. This award is pursuant to the authority of 42 USC 241 42 CFR 52 and  
is subject to terms and conditions referenced below.

Acceptance of this award including the Terms and Conditions is acknowledged by the grantee  
when funds are drawn down or otherwise obtained from the grant payment system.

Award recipients are responsible for reporting inventions derived or reduced to practice in  
the performance of work under this grant. Rights to inventions vest with the grantee  
organization provided certain requirements are met and there is acknowledgement of NIH  
support. In addition, recipients must ensure that patent and license activities are  
consistent with their responsibility to make unique research resources developed under this  
award available to the scientific community, in accordance with NIH policy. For additional  
information, please visit <http://www.iedison.gov>.

If you have any questions about this award, please contact the individual(s) referenced in  
the information below.

Sincerely yours,



Diane Patterson  
Grants Management Officer  
NATIONAL HUMAN GENOME RESEARCH INSTITUTE

See additional information below

Columbia Exhibit 2118  
Illumina, Inc. v. The Trustees  
of Columbia University in the  
City of New York  
IPR2018-00291, 318, 322, 385 & 797

SECTION I - AWARD DATA - 1 R21 HG003977-01

AWARD CALCULATION (U.S. Dollars):

Federal Direct Costs	\$150,000
Federal F&A Costs	\$128,850
APPROVED BUDGET	\$278,850
TOTAL FEDERAL AWARD AMOUNT	\$278,850

Recommended future year total cost support, subject to the availability of funds and satisfactory progress of the project, is as follows.

02 \$232,375

FISCAL INFORMATION:

CFDA Number: 93.172

EIN: 1330435954A1

Document Number: RHG003977A

IC/	CAN	/	FY2006	/	FY2007
HG/	8427256	/	278,850	/	232,375

NIH ADMINISTRATIVE DATA:

PCC: X7JS / OC: 41.4A /Processed: CAMPBELLB0 060509 0257

SECTION II - PAYMENT/HOTLINE INFORMATION - 1 R21 HG003977-01

For Payment and HHS Office of Inspector General Hotline Information, see the NIH Home Page at <http://grants.nih.gov/grants/policy/awardconditions.htm>

SECTION III - TERMS AND CONDITIONS - 1 R21 HG003977-01

This award is based on the application submitted to, and as approved by, the NIH on the above-titled project and is subject to the terms and conditions incorporated either directly or by reference in the following:

- The grant program legislation and program regulation cited in this Notice of Grant Award.
- The restrictions on the expenditure of federal funds in appropriations acts, to the extent those restrictions are pertinent to the award.
- 45 CFR Part 74 or 45 CFR Part 92 as applicable.
- The NIH Grants Policy Statement, including addenda in effect as of the beginning date of the budget period.
- This award notice, INCLUDING THE TERMS AND CONDITIONS CITED BELOW.

(see NIH Home Page at <http://grants.nih.gov/grants/policy/awardconditions.htm> for certain references cited above.)

An unobligated balance may be carried over into the next budget period without Grants Management Officer prior approval.

This grant is subject to Streamlined Noncompeting Award Procedures (SNAP).

Treatment of Program Income:  
Additional Costs

#### SECTION IV - NHGRI TERMS AND CONDITIONS

1. Although the budget period start date for this award is 5/17/2006, this award includes funds for 12 months of support. Future year budget periods will cycle on 5/01/2006. Grant Progress Reports (PHS 2590) are due two months prior to this date.

2. This award is being issued according to the guidelines and requirements in RFA HG-04-002, entitled NEAR-TERM TECHNOLOGY DEVELOPMENT FOR GENOME SEQUENCING.  
URL: <http://grants2.nih.gov/grants/guide/rfa-files/RFA-HG-04-002.html>

3. Technology release: The applicant's plan for providing access to the technology and information about the technology developed under this grant support, described in the grantee letter dated April 21, 2006, shall serve as the minimum acceptable requirement for technology sharing and access.

4. Progress reports: Applicants are required to submit progress reports twice per year: one formal report at the time of the non-competing continuation and a second interim report at a time to be determined by NHGRI staff. The latter may coincide with grantee meetings or meetings of advisors to NHGRI. NHGRI will use information from reports, site visits, etc., to evaluate each grantee's progress and the success of the overall program; this information will be used to determine if funding levels should be increased or decreased for future years for each grant and for the program.

5. Grantee meetings: To accelerate progress in the field of advanced DNA sequencing technology development, grantees will be expected to participate actively and openly in at least one grantee meeting per year. Substantial information sharing will be required and is a condition of the award; failure to openly share information will be grounds for discontinuation of funding. It is understood that some information developed under the grants will be proprietary and cannot be shared immediately without damaging the commercialization potential of the technology.

NHGRI will negotiate a schedule for the grantee meetings, which will rotate among the grantee sites. NHGRI will provide additional funds as needed in support of local arrangements for these meetings. Holding these meetings at grantee sites will facilitate information-sharing and participation of a larger portion of the research staff at host institutions than would otherwise occur.

6. Timeline and Milestones: Funding for future years is contingent upon administrative review by the NHGRI staff. Funding for future years may be increased, reduced, or eliminated depending on progress towards milestones as determined by NHGRI staff. The goals and milestones to be used in the assessment of progress will be based on those presented on page 22 of the grant application and modified in a letter from the P.I., dated April 21, 2004. These goals may be further modified by NHGRI in consultation with the P.I. NHGRI appreciates that this project requires research, not just engineering; progress toward milestones will be evaluated accordingly.

7. This is a Modular Grant Award without direct cost categorical breakdowns issued in accordance with the guidelines published in the NIH Grants Policy Statement, December 2003, see

([http://grants.nih.gov/grants/policy/nihgps\\_2003/NIHGPs\\_Part12.htm#\\_Toc54600227](http://grants.nih.gov/grants/policy/nihgps_2003/NIHGPs_Part12.htm#_Toc54600227)) Recipients are required to allocate and account for costs related to this award by category within their institutional accounting system in accordance with applicable cost principles.

8. This award is included under SNAP. Future noncompeting awards will be issued without direct cost categorical breakdowns. The categorical basis is reflected in the attached spreadsheet.} Recipients are required to allocate and account for these costs by category in accordance with applicable cost principles. Reference NIH Guide, Volume 24, Number 38, October 27, 1995 and the following URL:

[http://grants.nih.gov/grants/policy/nihgps\\_2003/NIHGPs\\_Part8.htm#\\_Streamlined\\_Non-Competing\\_Award](http://grants.nih.gov/grants/policy/nihgps_2003/NIHGPs_Part8.htm#_Streamlined_Non-Competing_Award).

9. Award recipients are strongly encouraged to submit to PubMed Central (PMC), upon acceptance for publication, an electronic version of peer-reviewed, original research publications, resulting from research supported in whole or in part, with direct costs from NIH. The author's final manuscript is defined as the final version accepted for journal publication, and includes all modifications from the publishing peer review process. For additional information, please visit <http://www.nih.gov/about/publicaccess/>.

#### SECTION V - STAFF CONTACTS

INFORMATION: If you need assistance from the NHGRI during the course of this award, please contact the grants management and program staff whose contact information is listed below. These individuals work closely with one another to coordinate administrative and scientific oversight of this project. Their functions are defined as follows:

GRANTS MANAGEMENT SPECIALIST: The Grants Management Specialist serves as the focal point for receiving and responding to all questions and correspondence related to business management and policy matters, such as correspondence giving or denying any prior approval required by NIH policy or special terms of award, transfer of the grant to another institution, a change in the period of support, or any actions which commit, or may result in committing the NHGRI to a change in the amount of funding.

PROGRAM OFFICIAL: The program official is responsible for all scientific and technical matters dealing with the administration of the grant, such as the review and monitoring of scientific progress, providing advice and assistance relative to technical problems; to ensure that the scientific objectives of the research program can be pursued effectively and successfully.

PRIOR APPROVAL: Requests must be submitted in writing by the authorized business official to the Grants Management Specialist. Requests may be made via email provided they are routed through the authorized business official. The Principal Investigator countersignature is no longer required on prior approval requests, however the grantee institution must secure and retain such documentation within its own internal processes and provide it upon request. See the April 7, 2006 NIH Guide Notice:  
[<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-06-054.html>]

JEFFERY SCHLOSS, Program Official

Phone: 301-435-5538 Email: [schlossj@mail.nih.gov](mailto:schlossj@mail.nih.gov) Fax: 301-480-2770

Brian L. Campbell, Grants Specialist *hcl*

SPREADSHEET

GRANT NUMBER: 1 R21 HG003977-01

P.I.: ROMESBERG, FLOYD E

INSTITUTION: SCRIPPS RESEARCH INSTITUTE

	YEAR 01	YEAR 02
	=====	=====
TOTAL FEDERAL DC	150,000	125,000
TOTAL FEDERAL F&A	128,850	107,375
TOTAL COST	278,850	232,375

	YEAR 01	YEAR 02
	=====	=====
F&A Cost Rate 1	85.90%	85.90%
F&A Cost Base 1	150,000	125,000
F&A Costs 1	128,850	107,375

Department of Health and Human Services  
Health Services

1 R21 HG003977-01

IPF:7375802


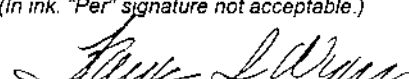
Dual: GM,RR

IRG: ZRG1 GGG-J(10) B

Received: 06/01/2005

**10027699****Application**

Do not exceed character length restrictions indicated.

1. TITLE OF PROJECT (Do not exceed 81 characters, including spaces and punctuation.) <b>Evolving Novel Polymerases for Genome Sequencing</b>					
2. RESPONSE TO SPECIFIC REQUEST FOR APPLICATIONS OR PROGRAM ANNOUNCEMENT OR SOLICITATION <input type="checkbox"/> NO <input checked="" type="checkbox"/> YES (If "Yes," state number and title) Number: PA-03-058 Title: NIH Exploratory/Developmental Research Grant Award (R21)					
3. PRINCIPAL INVESTIGATOR/PROGRAM DIRECTOR			New Investigator <input checked="" type="checkbox"/> No <input type="checkbox"/> Yes		
3a. NAME (Last, first, middle) Romesberg, Floyd E.			3b. DEGREE(S) PhD		3h. eRA Commons User Name eRA Commons User Name
3c. POSITION TITLE Assistant Professor			3d. MAILING ADDRESS (Street, city, state, zip code) The Scripps Research Institute 10550 North Torrey Pines Road La Jolla, CA 92037		
3e. DEPARTMENT, SERVICE, LABORATORY, OR EQUIVALENT Chemistry					
3f. MAJOR SUBDIVISION None					
3g. TELEPHONE AND FAX (Area code, number and extension) TEL: 858.784.7291 FAX: 858.784.7472			E-MAIL ADDRESS: floyd@scripps.edu		
4. HUMAN SUBJECTS RESEARCH <input checked="" type="checkbox"/> No <input type="checkbox"/> Yes		4b. Human Subjects Assurance No.		5. VERTEBRATE ANIMALS <input checked="" type="checkbox"/> No <input type="checkbox"/> Yes	
4a. Research Exempt <input type="checkbox"/> No <input type="checkbox"/> Yes		4c. Clinical Trial <input type="checkbox"/> No <input type="checkbox"/> Yes		4d. NIH-defined Phase III Clinical Trial <input type="checkbox"/> No <input type="checkbox"/> Yes	
If "Yes," Exemption No.		5a. If "Yes," IACUC approval Date		5b. Animal welfare assurance no.	
6. DATES OF PROPOSED PERIOD OF SUPPORT (month, day, year—MM/DD/YY) From 04/01/06 Through 03/31/08		7. COSTS REQUESTED FOR INITIAL BUDGET PERIOD 7a. Direct Costs (\$) \$150,000		8. COSTS REQUESTED FOR PROPOSED PERIOD OF SUPPORT 7b. Total Costs (\$) \$278,850 8a. Direct Costs (\$) \$275,000 8b. Total Costs (\$) \$511,225	
9. APPLICANT ORGANIZATION Name The Scripps Research Institute Address 10550 North Torrey Pines Road La Jolla, CA 92037			10. TYPE OF ORGANIZATION Public: <input type="checkbox"/> Federal <input type="checkbox"/> State <input type="checkbox"/> Local Private: <input checked="" type="checkbox"/> Private Nonprofit For-profit: <input type="checkbox"/> General <input type="checkbox"/> Small Business <input type="checkbox"/> Woman-owned <input type="checkbox"/> Socially and Economically Disadvantaged		
			11. ENTITY IDENTIFICATION NUMBER 1330435954A1 DUNS NO. 78-161-3492 Cong. District 53		
12. ADMINISTRATIVE OFFICIAL TO BE NOTIFIED IF AWARD IS MADE Name Kaye I. Wynne Title V.P., Office of Sponsored Programs Address The Scripps Research Institute 10550 North Torrey Pines Road La Jolla, CA 92037 Tel: 858.784.8653 FAX: 858.784.8037 E-Mail:			13. OFFICIAL SIGNING FOR APPLICANT ORGANIZATION Name Kaye I. Wynne Title V.P., Office of Sponsored Programs Address The Scripps Research Institute 10550 North Torrey Pines Road La Jolla, CA 92037 Tel: 858.784.8653 FAX: 858.784.8037 E-Mail:		
14. PRINCIPAL INVESTIGATOR/PROGRAM DIRECTOR ASSURANCE: I certify that the statements herein are true, complete and accurate to the best of my knowledge. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. I agree to accept responsibility for the scientific conduct of the project and to provide the required progress reports if a grant is awarded as a result of this application.			SIGNATURE OF PI/PD NAMED IN 3a. (In ink. "Per" signature not acceptable.) 		DATE 5/31/05
15. APPLICANT ORGANIZATION CERTIFICATION AND ACCEPTANCE: I certify that the statements herein are true, complete and accurate to the best of my knowledge, and accept the obligation to comply with Public Health Services terms and conditions if a grant is awarded as a result of this application. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties.			SIGNATURE OF OFFICIAL NAMED IN 13. (In ink. "Per" signature not acceptable.) 		DATE 5/31/05

DESCRIPTION: See instructions. State the application's broad, long-term objectives and specific aims, making reference to the health relatedness of the project (i.e., relevance to the mission of the agency). Describe concisely the research design and methods for achieving these goals. Describe the rationale and techniques you will use to pursue these goals.

In addition, in two or three sentences, describe in plain, lay language the relevance of this research to public health. If the application is funded, this description, as is, will become public information. Therefore, do not include proprietary/confidential information. **DO NOT EXCEED THE SPACE PROVIDED.**

There is great incentive for developing low cost technologies for sequencing genomes. While significant advances have been made in conventional sequencing methods, the reductions in cost and time required to make genome sequencing more feasible require the development of new methodologies. Indeed many new and promising methods are under development. However, these new approaches are generally limited by the inability of DNA polymerases to synthesize DNA with modified deoxynucleotide triphosphates (dNTPs). We propose to use our recently developed activity-based selection system to evolve DNA polymerases that efficiently recognize these modified dNTPs, and thus overcome one of the most significant barriers to the development of the novel sequencing methods. The selection system is based on the co-display on phage of DNA polymerase libraries and their DNA substrates, such that mutants that recognize the modified dNTPs covalently attach a biotin-dUTP, allowing for their selective isolation. In addition to the selections, various modern gene shuffling approaches are proposed to generate polymerase diversity. The Specific Aims of the proposal are:

**Specific Aim 1:** Evolve DNA polymerases that efficiently synthesize DNA using fluorophore-labeled dNTPs to enable novel synthesis-by-sequencing methodologies.

**Specific Aim 2:** Evolve polymerases that efficiently incorporate reversibly fluorophore tagged dNTPs with 3'-O-allyl protecting groups and then continue synthesis after double deprotection.

**Specific Aim 3:** Collaborate to evaluate evolved polymerases in actual sequencing applications.

The evolution of polymerases better suited for sequencing applications, and their evaluation in the actual sequencing methodologies, promises to dramatically facilitate the development of the methods and help make genome sequencing more commonplace.

PERFORMANCE SITE(S) (organization, city, state)

The Scripps Research Institute  
La Jolla, CA

Stanford University  
Palo Alto, CA

Helicos BioSciences  
Cambridge, MA

Columbia University  
New York, NY

Principal Investigator/Program Director (Last, First, Middle): Romesberg, Floyd E.

KEY PERSONNEL. See instructions. *Use continuation pages as needed* to provide the required information in the format shown below. Start with Principal Investigator. List all other key personnel in alphabetical order, last name first.

Name	eRA Commons User Name	Organization	Role on Project
Romesberg, Floyd E.	<input type="text" value="eRA Commons User Name"/>	The Scripps Res. Inst.	PI
<input type="text" value="Student Names"/>		Helicos BioSciences	Graduate Student
Ju, Jingyue		Columbia University	Collaborator
<input type="text" value="Student Names"/>		The Scripps Res. Inst.	Graduate Student
Ordoukhanian, Phillip		The Scripps Res. Inst.	Collaborator
Quake, Steve		Stanford University	Collaborator
<input type="text" value="Student Names"/>		The Scripps Res. Inst.	Graduate Student

OTHER SIGNIFICANT CONTRIBUTORS

Name	Organization	Role on Project
------	--------------	-----------------

Human Embryonic Stem Cells ☒ No ☐ Yes

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s) from the following list: <http://stemcells.nih.gov/registry/index.asp>. *Use continuation pages as needed.*

If a specific line cannot be referenced at this time, include a statement that one from the Registry will be used.

Cell Line

Disclosure Permission Statement. Applicable to SBIR/STTR Only. See SBIR/STTR instructions. ☐ Yes ☐ No

The name of the principal investigator/program director must be provided at the top of each printed page and each continuation page.

# RESEARCH GRANT

## TABLE OF CONTENTS

	Page Numbers
Face Page.....	1
Description, Performance Sites, Key Personnel, Other Significant Contributors, and Human Embryonic Stem Cells .....	2
Table of Contents.....	4
Detailed Budget for Initial Budget Period (or Modular Budget) .....	5
Budget for Entire Proposed Period of Support (not applicable with Modular Budget).....	N/A
Budgets Pertaining to Consortium/Contractual Arrangements (not applicable with Modular Budget) .....	N/A
Biographical Sketch – Principal Investigator/Program Director (Not to exceed four pages) .....	6-8
Other Biographical Sketches (Not to exceed four pages for each – See instructions) .....	9-20
Resources .....	21
Research Plan.....	22
Introduction to Revised Application (Not to exceed 3 pages) .....	
Introduction to Supplemental Application (Not to exceed one page) .....	
A. Specific Aims .....	22
B. Background and Significance.....	22
C. Preliminary Studies/Progress Report/	28
Phase I Progress Report (SBIR/STTR Phase II ONLY) .....	
D. Research Design and Methods.....	32
E. Human Subjects .....	36
Protection of Human Subjects (Required if Item 4 on the Face Page is marked "Yes") .....	
Inclusion of Women and Minorities (Required if Item 4 on the Face Page is marked "Yes" and is Clinical Research) .....	
Targeted/Planned Enrollment Table (for new and continuing clinical research studies) .....	
Inclusion of Children (Required if Item 4 on the Face Page is marked "Yes") .....	
Data and Safety Monitoring Plan (Required if Item 4 on the Face Page is marked "Yes" and a Phase I, II, or III clinical trial is proposed).....	
F. Vertebrate Animals .....	36
G. Literature Cited .....	36
H. Consortium/Contractual Arrangements.....	N/A
I. Resource Sharing.....	N/A
J. Letters of Support (e.g., Consultants) .....	
Commercialization Plan (SBIR/STTR Phase II and Fast-Track ONLY) .....	
Checklist.....	41
Appendix (Five collated sets. No page numbering necessary for Appendix.)	
Appendices NOT PERMITTED for Phase I SBIR/STTR unless specifically solicited.....	
Number of publications and manuscripts accepted for publication (not to exceed 10) .....	
Other items (list):	
Letters of collaboration from Drs. Buzby, Ju, Ordoukhanian and Quake.	

Check if  
Appendix is  
Included



**BUDGET JUSTIFICATION PAGE  
MODULAR RESEARCH GRANT APPLICATION**

	Initial Period	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	Sum Total (For Entire Project Period)
DC less Consortium F&A <i>(Item 7a, Face Page)</i>	150,000	125,000				275,000 <i>(Item 8a, Face Page)</i>
Consortium F&A						
<b>Total Direct Costs</b>	150,000	125,000				<b>\$ 275,000</b>

**Personnel**

Floyd E. Romesberg, PhD, Principal Investigator ( effort). The PI will oversee the entire project. He will review experimental protocols, supervise the graduate students and assist in analysis of experimental data and preparation of manuscripts.

Student Names Graduate Student ( effort).  Student Names will phage display polymerase libraries, and perform the selection described in Specific Aim 1.

Student Names Graduate Student ( effort).  Student Names will generate libraries and perform the selections described in Specific Aim 2.

To Be Named, Postdoctoral Fellow (100% effort). The to-be-named postdoc will assist in all selections and also kinetically characterize the evolved polymerases.

**Consortium****Fee (\$BIR/STTR Only)**

**BIOGRAPHICAL SKETCH**

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.  
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME <b>Floyd E. Romesberg</b>		POSITION TITLE <b>Assistant Professor</b>	
eRA COMMONS USER NAME <input type="text"/>			
EDUCATION/TRAINING <i>(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)</i>			
INSTITUTION AND LOCATION	DEGREE <i>(if applicable)</i>	YEAR(s)	FIELD OF STUDY
The Ohio State University, OH	B.S.	1985-1988	Chemistry
Cornell University, Ithaca, NY	M.S.	1988-1990	Organic Chemistry
Cornell University, Ithaca, NY	Ph.D.	1990-1994	Organic Chemistry
University of California, Berkeley	Postdoctoral	1994-1998	Organic Chemistry

**A. POSITIONS AND HONORS:****Positions:**

1986 - 1988 Undergraduate (Advisor: Professor Matthew S. Platz)  
The Ohio State University, Columbus, OH

1988 - 1990 Graduate Studies (Advisor: Professor David B. Collum)  
Cornell University, Ithaca, NY

1990 - 1994 Graduate Studies (Advisor: Professor David B. Collum)  
Cornell University, Ithaca, NY

1994 - 1998 Postdoctoral Research (Advisor: Professor Peter G. Schultz)  
University of California, Berkeley, CA

1998 - Pres Assistant Professor, Department of Chemistry,  
The Scripps Research Institute, La Jolla, CA

**Honors:**

NIH Postdoctoral Fellowship, 1994-1996

The MacNevin Award, The Ohio State University, 1987

Baxter Foundation Fellow, 2002-2003

Susan G. Komen Breast Cancer Foundation Award, 2003-2006

Dreyfus Teacher Scholar Award, 2003-2004

NSF CAREER Award, 2004

Discover Magazine Technology Innovation Award, 2004

**B. SELECTED PEER-REVIEWED PUBLICATIONS (in chronological order)**

1. D. L. McMinn, A. K. Ogawa, Y. Wu, J. Liu, P. G. Schultz, F. E. Romesberg. Efforts Toward Expansion of the Genetic Alphabet: DNA Polymerase Recognition of a Highly Stable, Self-Pairing Hydrophobic Base. (1999) J. Am. Chem. Soc. 121,11585.
2. A. K. Ogawa, Y. Wu, D. L. McMinn, J-Q. Liu, P. G. Schultz, and F. E. Romesberg. Efforts Toward the Expansion of the Genetic Alphabet: Information Storage and Replication with Unnatural Hydrophobic Base Pairs. (2000) J. Am. Chem. Soc. 122,3274-3287.
3. M. Berger, A. I. Ogawa, D. L. McMinn, P. G. Schultz, and F. E. Romesberg. Stable and Selective Hybridization of Oligonucleotides using Unnatural Hydrophobic Bases. (2000) Angew. Chemie. Int. Ed. 39, 2940-2942.
4. Y. Wu, A. K. Ogawa, M. Berger, D. L. McMinn, P. G. Schultz, and F. E. Romesberg. Efforts Toward the Expansion of the Genetic Alphabet: Optimization of Interbase Hydrophobic Interactions. (2000) J. Am. Chem. Soc. 122, 7621-7632.
5. M. Berger, Y. Wu, A. K. Ogawa, D. L. McMinn, P. G. Schultz, and F. E. Romesberg. Universal Bases for Hybridization, Replication, or Chain Termination. (2000) Nucleic Acids Res. 28, 2911-2914.
6. A. K. Ogawa, Y. Wu, M. Berger, P. G. Schultz, and F. E. Romesberg. Rational Design of an Unnatural Base Pair with Increased Kinetic Selectivity. (2000) J. Am. Chem. Soc. 122, 8803-8804.

7. A. K. Ogawa, O. Abou-Zied, V. Tsui, R. Jimenez, D. Case, and F. E. Romesberg. A Photo-Tautomerizable Model DNA Base Pair. (2000) *J. Am. Chem. Soc.* 122, 9917-9920.
8. J. Chin, J., Jimenez, and F. E. Romesberg. Direct Observation of Protein Vibrations by Selective Incorporation of Spectroscopically Observable Carbon-Deuterium Bonds in Cytochrome c. (2001) *J. Am. Chem. Soc.* 123, 2426-2427.
9. O. Abou-Zied, R. Jimenez, and F. E. Romesberg. Tautomerization Dynamics of a Model Base-Pair in DNA. (2001) *J. Am. Chem. Soc.* 123, 4613-4614.
10. E. Lee, Y. Wu, G. Xia, P. G. Schultz, and F. E. Romesberg. Efforts Toward Expansion of the Genetic Alphabet: Replication of DNA with Three Base Pairs. (2001) *J. Am. Chem. Soc.* 123, 7439-7440.
11. R. Jimenez, D. A. Case, and F. E. Romesberg. Flexibility of an Antibody Binding Site Measured with Photon Echo Spectroscopy. (2002) *J. Phys. Chem. B* 106, 1090-1103.
12. J. K. Chin, R. Jimenez, and F. E. Romesberg. Protein Dynamics and Cytochrome c: Correlations Between Ligand vibrations and Redox Activity. (2002) *J. Am. Chem. Soc.* 124, 1846-1847.
13. M. Berger, S. Luzzi, A. Henry, and F. E. Romesberg. F.E. Stability and Selectivity of Unnatural DNA with Five Membered Ring Nucleobase Analogs. (2002) *J. Am. Chem. Soc.* 124, 1222-1226.
14. O. K. Abou-Zied, R. Jimenez, E. K. Thompson, D. P. Millar, F. E. Romesberg. Solvent-Dependent Photoinduced Tautomerization of 2-(2'-hydroxyphenyl)benzoxazole. (2002) *J. Phys. Chem. A* 106, 3665-3672.
15. G. Xia, L. J. Chen, T. Sera, M. Fa, P. G. Schultz, F. E. Romesberg. Directed evolution of novel polymerase activities: Mutation of a DNA polymerase into an efficient RNA polymerase. (2002) *Proc. Natl. Acad. Sci.* 99, 6597-6602.
16. D. Hanway, J. K. Chin, G. Xia, G. Oshiro, E. A. Winzler, F. E. Romesberg. New Genes in the UV- and MMS-Induced DNA Damage Response in Yeast. (2002) *Proc. Natl. Acad. Sci. USA* 99, 10605-10610.
17. R. Jimenez, F. E. Romesberg, F.E. Excited State Dynamics and Heterogeneity of Folded and Unfolded States of Cytochrome c. (2002) *J. Phys. Chem. B* 106, 9172-9180.
18. C. Yu, A. A. Henry, P. G. Schultz, F. E. Romesberg. Polymerase Recognition of Unnatural Base Pairs. (2002) *Angew. Chem.* 114, 3997-4000.
19. Y. Wu, M. Fa, E. L. Tae, P. G. Schultz, F. E. Romesberg. Enzymatic Phosphorylation of Unnatural Nucleobases. (2002) *J. Am. Chem. Soc.* 124, 14626-30.
20. H. Wang, H. Zhang, O. K. Abou-Zied, C. Yu, F. E. Romesberg and M. Glasbeek. Femtosecond Fluorescence Upconversion Studies of Excited-state Proton Transfer Dynamics in 2-(2'-hydroxyphenyl)benzoxazole (HBO) in Liquid Solution and DNA. (2002) *Chem. Phys. Lett.* 367, 599-608.
21. R. Jimenez, G. Salazar, K. K. Baldrige, F. E. Romesberg. Flexibility and Molecular Recognition in the Immune System. (2003) *Proc. Natl. Acad. Sci. USA* 114, 3997-4000.
22. S. Matsuda, A. A. Henry, P. E. Schultz and F. E. Romesberg. The Effect of Minor Groove Hydrogen-Bond Acceptors and Donors on The Stability And Replication of Four Unnatural Base Pairs. (2003) *J. Am. Chem. Soc.* 125, 6134-6138.
23. A. Henry, C. Yu and F. E. Romesberg. Determinants of Unnatural Nucleobase Stability And Polymerase Recognition. (2003) *J. Am. Chem. Soc.* 125, 9638-9646.
24. F. E. Romesberg. Multidisciplinary Experimental Approaches to Characterizing Biomolecular Dynamics. (2003) *ChemBioChem* 4, 563-571.
25. A. Henry and F. E. Romesberg. Beyond A, C, G and T: Augmenting Nature's Alphabet. (2003) *Curr. Op. Chem. Biol.* 7, 727-733.
26. M. Fa, A. Radeghieri, A. A. Henry and F. E. Romesberg. Expanding The Substrate Repertoire of a DNA Polymerase By Directed Evolution. (2004) *J. Am. Chem. Soc.* 126, 1748-1754.
27. R. Jimenez, J. Yin, T. Joo and F. E. Romesberg. Protein Dynamics and the Immunological Evolution of Molecular Recognition. (2004) *Proc. Natl. Acad. Sci. USA* 101, 3803-3808.
28. L. B. Sagle, J. Zimmermann, P. E. Dawson and F. E. Romesberg. A High Resolution Probe of Protein Folding. (2004) *J. Am. Chem. Soc.* 126, 3384-3385.
29. A. Henry, R. Jimenez, D. Hanway and F. E. Romesberg. Preliminary Characterization of Light Harvesting in *E. coli* DNA Photolyase. (2004) *ChemBioChem* 4:563-571.
30. A. Henry, A. Goldbech Olsen, S. Matsuda, C. Yu., B. Geirstanger and F. E. Romesberg. Efforts To Expand The Genetic Alphabet: Identification Of A Stable And Replicable Unnatural DNA Self Pair. (2004) *J. Am. Chem. Soc.* 126(22):6923-31.
31. S. Matsuda, P. G. Schultz and F. E. Romesberg. Optimization of Interstrand Hydrophobic Packing Interactions Within Duplex DNA. (2004) *J. Am. Chem. Soc.* 126(44):14419-27.
32. B. M. O'Neill, D. Hanway, E. A. Winzler and F. E. Romesberg. Coordinated Functions of WSS1, PSY2, and TOF1 in the DNA Damage Response. (2004) *Nucleic Acids Res.* 32(22):6519-30.
33. R. T. Cirz, D. R. Andes, V. de Crecy-Legard, W. A. Craig and F. E. Romesberg. The Molecular Mechanism of Mutation Underlying the Evolution of Antibiotic Resistance. (2005) *PLoS* 3(6):e176 [epub ahead of print]

C. RESEARCH SUPPORT

Ongoing Research Support:

R01 (GM 60005) Romesberg (PI) 04/01/05 – 03/31/09

National Institutes of Health

"Expanding the Genetic Alphabet by Polymerase Evolution"

The major goals of this project are to synthesize and study unnatural DNA base pairs and to use phage display methods to evolve enzymes capable of accepting them.

N00014-03-1-0126 Romesberg (PI) 10/25/02 – 12/31/08

Office of Naval Research

"Preventing Antibiotic Resistance"

The major goals of this project are to identify the proteins of the SOS response that are required for bacteria to mutate their genome, and thus evolve resistance to antibiotics.

R01 (GM 068569) Romesberg (PI) 04/01/04 – 03/31/08

National Institutes of Health

"Pathways Controlling Genome Stability and Mutation"

The major goals of this project are genetic and biochemical approaches to understanding DNA repair in yeast.

MCB-0346967 Romesberg (PI) 03/15/04 – 03/14/09

National Science Foundation – NSF CAREER Award

"Residue Specific Probes of the Cytochrome c Folding Mechanism"

The major goal of this project is to characterize the detailed folding mechanism of cytochrome c.

BC031880 Romesberg (PI) 05/15/04 – 05/14/06

DOD BCRP IDEA Award

"Inhibition of Mutation: A Novel Approach to Preventing and Treating Cancer"

The major goal of this project is to characterize the mechanism of acquiring mutations in human cells.

COMPLETED:

None

**BIOGRAPHICAL SKETCH**

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.  
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME Philip R. Buzby		POSITION TITLE Director, Nucleic Acid Chemistry	
eRA COMMONS USER NAME			
EDUCATION/TRAINING <i>(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)</i>			
INSTITUTION AND LOCATION	DEGREE <i>(if applicable)</i>	YEAR(s)	FIELD OF STUDY
Iowa State University, Ames, IA	B.S.	1975	Biology and Biochemistry
Washington State University, Pullman, WA	Ph.D.	1981	Biochemistry

**A. POSITIONS AND HONORS.**

1980 – 1981      Nucleotide Chemist, New England Nuclear Corp., Boston, MA.  
 1981 – 1999      Manager/Scientist, E. I. DuPont, Boston, MA.  
 1999 – 2002      Director of Business Development, NEN Life Sciences Inc., Boston, MA.  
 2002 – 2004      Manager, Special Projects and Corporate IP, PerkinElmer Life and Analytical Sciences Inc. Boston, MA.  
 2004 – Pres      Director, Nucleic Acid Chemistry, Helicos BioSciences Corporation, Cambridge, MA.

**B. SELECTED PEER-REVIEWED PUBLICATIONS (IN CHRONOLOGICAL ORDER).**

Xiao, M., Phong, A., Lum, K.L., Greene, R.A., Buzby, P.R., Kwok, P.Y. Role of excess inorganic pyrophosphate in primer-extension genotyping assays. *Genome Research* 2004;14:9.

**C. RESEARCH SUPPORT.**

NONE

**BIOGRAPHICAL SKETCH**

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.  
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME Jingyue Ju		POSITION TITLE Associate Professor of Chemical Engineering, Head of DNA Sequencing and Chemical Biology at Columbia Genome Center, Columbia University	
eRA COMMONS USER NAME			
EDUCATION/TRAINING <i>(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)</i>			
INSTITUTION AND LOCATION	DEGREE <i>(if applicable)</i>	YEAR(s)	FIELD OF STUDY
Inner Mongolia University, P.R. China	B.S.	1985	Chemistry
Institute of Chemical Physics, Chinese Academy of Sciences	M.S.	1988	Organic Chemistry
University of Southern California	Ph.D.	1993	Bioorganic Chemistry
University of California, Berkeley	Postdoctoral	1993-1995	Genomics Research

**A. POSITIONS AND HONORS****Positions:**

1995 - 1998 Senior Scientist, Incyte Genomics, Inc.  
 1998 - 1999 Director, Chemistry and Assay Development, Incyte Genomics, Inc.  
 1999 - Pres Associate Professor and Head of DNA Sequencing and Chemical Biology, Department of Chemical Engineering & Columbia Genome Center, Columbia University.  
 2003 - Pres Director, Columbia-NIH Center of Excellence in Genomic Science.

**Honors:**

U.S. Biochemical Postdoctoral Fellowship, 1993  
 DOE Human Genome Distinguished Postdoctoral Fellowship, 1994-1995  
 Packard Fellowship in Science and Engineering, 2001-2006  
 Invited Attendee, National Academy of Engineering, 8th Annual Frontiers of Engineering Symposium, September, 2002.  
 Outstanding Chinese Scholar Achievement Award by the Columbia University Chinese Students and Scholars Association, 2004  
 Visiting Professor of Beijing Capital University of Medical Sciences, 2004

**Professional Activities:**

Chair, DNA Sequencing Technology Session, Human Genome Meeting, Human Genome Organization (HUGO), 1997.  
 Organizer and Chair, Symposium on Genomics and Chemical Biology, Post 15<sup>th</sup> International Conference on Phosphorus Chemistry, Beijing, China, August 6-8, 2001.  
 Organizer and Chair, "New Technology and Toxicogenomics" BioArrays-2003-New York Conference, October 1-2, 2003.  
 Organizer and Chair, "New Technologies & Genome Sequencing" BioArrays-2004-New York Conference, July 26-27, 2004.

**B. SELECTED PEER-REVIEWED PUBLICATIONS**

1. T. S. Seo, X. Bai, D. H. Kim, Q. Meng, S. Shi, H. Ruparel, Z. Li, N. Turro & J. Ju. Four-Color DNA Sequencing by Synthesis on Chip Using Photocleavable Fluorescent Nucleotide Analogues. Proceedings of the National Academy of Sciences USA 2005, 102, 5926-5931.

2. H. Ruparel, L. Bi, Z. Li, X. Bai, D. H. Kim, N. Turro & J. Ju. Design and Synthesis of a 3'-O-Allyloxy Photocleavable Fluorescent Nucleotide as a Reversible Terminator for DNA Sequencing By Synthesis. *Proceedings of the National Academy of Sciences USA* 2005, 102, 5932-5937.
3. J. R. Edwards, H. Ruparel, J. Ju. Mass Spectrometry DNA Sequencing. *Mutation Research*, 2005, 573, 3-12.
4. ACCEPTED
5. In Press
6. In Press
7. ACCEPTED
8. S. Pfeffer, A. Sewer, M. Lagos-Quintana, R. Sheridan, C. Sander, F.A. Grasser, L. F. van Dyk, C. K. Ho, S. Shuman, M. Chien, J.J. Russo, J. Ju, G. Randall, B. D. Lindenbach, C. M. Rice, V. Simon, D. D. Ho, M. Zavolan & T. Tuschl. Identification of microRNAs of the Herpesvirus Family. *Nature Methods* 2005, 2, 269-276.
9. T. Briese, G. Palacios, M. Kokoris, O. Jabado, Z. Liu, N. Renwick, V. Kapoor, I. Casas, F. Pozo, R. Limberger, P. Perez-Brena, J. Ju, W. I. Lipkin. Diagnostic system for rapid and sensitive differential detection of pathogens. *Emerging Infectious Diseases*, 2005, 11, 310-313.
10. S. Pfeffer, M. Zavolan, F.A. Grasser, M. Chien, J.J. Russo, J. Ju, B. John, A.J. Enright, D. Marks, C. Sander and T. Tuschl. Identification of Virus-Encoded microRNAs. *Science* 2004, 304, 734-736.
11. T. S. Seo, X. Bai, Z. Li, H. Ruparel, N. Turro & J. Ju. Photocleavable Fluorescent Nucleotides for DNA Sequencing on a Chip Constructed by Site-Specific Coupling Chemistry. *Proceedings of the National Academy of Sciences USA*. 2004, 101, 5488-5493.
12. S. Kim, M. Ulz, T. Nguyen, C. Li, B. Tycko, and J. Ju. Thirty Fold Multiplex Genotyping of p53 Gene Using Solid Phase Capturable Dideoxynucleotides and Mass Spectrometry. *Genomics* 2004, 83, 924-931.
13. H. Ruparel, M. E. Ulz, S. Kim, and J. Ju. Digital Detection of Genetic Mutations Using SPC-Sequencing. *Genome Research* 2004, 14, 296-300.
14. X. Bai, S. Kim, Z. Li, N. J. Turro & J. Ju. Design and Synthesis of a Photocleavable Biotinylated Nucleotide for DNA Analysis by Mass Spectrometry. *Nucleic Acids Research*. 2004, 32, 535-541
15. Z. Li, T. S. Seo, J. Ju. 1,3-Dipolar Cycloaddition of Azides with Electron-deficient Alkynes Under Mild Condition in Water. *Tetrahedron Letters* 2004, 45, 3143-3146.
16. Chien, M., Morozova, I., Shi, S., Sheng, H., Chen, J., Gomez, S.M., Asamani, G., Hill, K., Nuara, J., Feder, M., Rineer, J., Greenberg, J.J., Steshenko, V., Park, S.H., Zhao, B., Teplitskaya, E., Edwards, J.R., Pampou, S., Georgiou, A., Chou, I.- C., Iannuccilli, W., Ulz, M.E., Kim, D.H., Geringer-Sameth, A., Goldsberry, C., Morozov, P., Fischer, S.G., Segal, G., Qu, X., Rzhetsky, A., Zhang, P., Cayanis, E., De Jong, P.J., Ju, J., Kalachikov, S., Shuman, H.A., Russo, J.J. The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* 2004, 305, 1966-1968.
17. Z. Li, X. Bai, H. Ruparel, S. Kim, N. Turro & J. Ju. A Photocleavable Fluorescent Nucleotide for DNA Sequencing and Analysis. *Proceedings of the National Academy of Sciences USA*. 2003, 100, 414-419.
18. S. Kim, H. D. Ruparel, T. C. Gilliam, and J. Ju. Digital Genotyping Using Molecular Affinity and Mass Spectrometry. *Nature Reviews Genetics* 2003, 4, 1001-1008.
19. X. Bai, Z. Li, S. Jockusch, N. Turro & J. Ju. Photocleavage of a 2-Nitrobenzyl Linker Bridging a Fluorophore to the 5' End of DNA. *Proceedings of the National Academy of Sciences USA*. 2003, 100, 409-413.
20. S. Kim, S. Shi, T. Bonome, J. R. Edwards, J. Russo, and J. Ju. Multiplex Genotyping of the Human  $\beta$ 2-Adrenergic Receptor Gene Using Solid Phase Capturable Dideoxynucleotides and Mass Spectrometry. *Analytical Biochemistry*. 2003, 316, 251-258
21. T. S. Seo, C. Wang, Z. Li, H. Ruparel, and J. Ju. Site-Specific Fluorescent Labeling of DNA Using Staudinger Ligation. *Bioconjugate Chemistry*. 2003, 14, 697-701
22. S. D. Goldberg, W. Iannuccilli, T. Nguyen, J. Ju, and V. W. Cornish. Identification of Critical Residues in a Class C  $\beta$ -Lactamase Using Combinatorial Scanning Mutagenesis. *Protein Science*. 2003, 12, 1633-1645.
23. S. Kim, J. R. Edwards, L. Deng, W. Chung, and J. Ju. Solid Phase Capturable Dideoxynucleotides for Multiplex Genotyping Using Mass Spectrometry. *Nucleic Acids Research*. 2002, 30, e78 (p1-6).

24. T. S. Seo, Z. Li, H. Ruparel, and J. Ju. Click Chemistry to Construct Fluorescent Oligonucleotides for DNA Sequencing. *Journal of Organic Chemistry*. 2002, 68, 609-612.
25. A. K. Tong, & J. Ju. Single-nucleotide Polymorphism Detection by Combinatorial Fluorescence Energy Transfer Tags and Biotinylated Dideoxynucleotides. *Nucleic Acids Research*. 2002, 30, e19 (p1-7).
26. J. Ju. & C. McKenna. Synthesis of Oligodeoxyribonucleoside Phosphorothioates Using Lawesson's Reagent for the Sulfur Transfer Step. *Bioorganic & Medicinal Chemistry Letters* 2002, 12, 1643-1645.
27. J. Ju. DNA Sequencing with Solid Phase Capturable Dideoxynucleotides and Energy Transfer Primers. *Analytical Biochemistry* 2002, 309, 35-39.
28. A. K. Tong, Z. Li, J. Ju. Combinatorial Fluorescence Energy Transfer Tags: New Molecular Tools for Genomics Applications. *Journal of Quantum Electronics Special Issue for Biomedical Applications*. 2002, 38, 110-121.
29. J. Ju. Fluorescence Energy Transfer Reagents for DNA Sequencing and Analysis. In: Structure and Dynamics of Confined Spaces (NATO Science Series, High Technology). 2002, 87, 371-383. eds. J.J. Kasianowicz, M.S.Z. Kellermayer, and D.W. Deamer (Kluwer Academic Publishers, Netherlands).
30. A. K. Tong, Z. Li, G. S. Jones, J. J. Russo, J. Ju. Combinatorial Fluorescence Energy Transfer Tags for Multiplex Biological Assays. *Nature Biotechnology*. 2001, 19, 756-759.
31. J.R. Edwards, Y. Itagaki and J. Ju. DNA Sequencing Using Biotinylated Dideoxynucleotides and Mass Spectrometry. *Nucleic Acids Research*, 2001, 29, e104 (p1-6).
32. A. Tong, S. Jockusch, Z. Li, H-R. Zhu, D. Akins, N. J. Turro, and J. Ju. Triple Fluorescence Energy Transfer in Covalently Tri-Chromophore-Labeled DNA. *J. Am. Chem. Soc.*, 2001, 123, 12923-12924.
33. B.E. Slatko, L.M. Albright, S. Tabor and J. Ju. DNA Sequencing. In: Short Protocols in Molecular Biology, 4<sup>th</sup> Ed. (1999) eds. Ausubel, F.M. et al. (John Wiley & Sons, Inc, New York and London), pp. 7-25 to 7-37. (Book Chapter).

#### C. RESEARCH SUPPORT

Ongoing Research Support  
1 P50 HG002806-01 (Ju, PI)  
NIH/NHGRI

8/1/03 – 7/31/06

Genomic Approaches to Neuronal Diversity and Plasticity

The objectives of the proposed research are the development of new genomic technologies for massive parallel DNA sequencing and large-scale gene expression analysis from single living cells, and applications of these technologies to study neural functions

1 RO1 HG003582-01 (Ju, PI)  
NIH/NHGRI

9/29/04 – 9/30/07

An Integrated System for DNA Sequencing by Synthesis

The objective of the proposed research is the design and synthesis of photocleavable fluorescent nucleotides for DNA sequencing by synthesis on a solid surface.

Private Source

(Ju, PI)

10/1/01-9/30/06

Digitally Decoding DNA Using Photocleavable Mass Tags

The major goals of this project is to construct a digital DNA sequencing by synthesis system that is capable of precisely decoding DNA sequence by detecting the photocleaved mass tags

R21 HG002425 (Lipkin, PI; Ju, Co-PI)  
NIH

7/01/03-6/30/05

A Staged Strategy for Virus Identification and Discovery: Mass-Tag PCR Detection of Viral Pathogens

The goal of this project is to develop methods using the molecular weights of a library of mass tags to identify virus using mass spectrometry.

Principal Investigator/Program Director (Last, First, Middle): Romesberg, Floyd E.

NSF0088001 (Leonard, PI; Ju, Co-PI)

3/15/01-4/14/05

National Science Foundation

Curriculum in Genomic Engineering

The goal of this project is to integrate genome science and technology into undergraduate and graduate education.

Completed Research Support

BES-00-97793 (Ju, PI)

3/15/01-3/14/04

National Science Foundation

Fluorescence Imaging Chip System for Massive Parallel DNA Sequencing

Biophotonics Initiative (Ju, PI)

9/01/00-8/31/03

National Science Foundation

Combinatorial Fluorescent Energy Transfer Tags and Their Application for Multiplex Genetic Analyses

R21 HG002425 (Bestor, PI; Ju, Co-PI)

5/01/02-4/30/04

NIH/NHGRI

Methylation Landscape of the Human Genome in Health & Disease

Principal Investigator/Program Director (Last, First, Middle): Romesberg, Floyd E.

---

### **BIOGRAPHICAL SKETCH**

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.  
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

---

Excluded by Requester

**BIOGRAPHICAL SKETCH**

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.  
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME Phillip T. Ordoukhanian		POSITION TITLE Director, Nucleic Acids Core Facility	
eRA COMMONS USER NAME			
EDUCATION/TRAINING <i>(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)</i>			
INSTITUTION AND LOCATION	DEGREE <i>(if applicable)</i>	YEAR(s)	FIELD OF STUDY
University of Missouri, St. Louis	B.S.	1992	Chemistry
Washington University, St. Louis	M.A.	1994	Chemistry
Washington University, St. Louis	Ph.D.	1997	Chemistry

**A. POSITIONS AND HONORS**

- 1992-1992: Monsanto Summer Research Internship, St. Louis, Missouri  
 1992-1994: Graduate Teaching Assistantship, Washington University, St. Louis, Missouri  
 1994-1997: Graduate Assistance in Areas of National Need (GAANN) Fellowship for Chemistry, Department of Education, St. Louis, Missouri  
 1992-1997: Doctoral Student, Washington University, St. Louis, Missouri  
 1997-2000: Adjunct Fellow, NASA Specialized Center of Research and Training (NSCORT), La Jolla, California  
 1997-2001: Postdoctoral Fellow, The Scripps Research Institute, La Jolla, California  
 2001-Present: Director, Nucleic Acids Core Facility, The Scripps Research Institute, La Jolla, California

**B. SELECTED PEER-REVIEWED PUBLICATIONS (in chronological order).**

1. P. Ordoukhanian and J.-S. Taylor. Design and synthesis of a versatile photocleavable DNA building block. Application to phototriggered hybridization. *J. Am. Chem. Soc.* 1995, 117, 9570-9571.
2. P. Ordoukhanian and J.-S. Taylor. Solid phase-supported thymine dimers for the construction of dimer-containing DNA by combined chemical and enzymatic synthesis: A potentially general method for the efficient incorporation of modified nucleotides into DNA. *Nucl. Acids. Res.* 1997, 25, 3783-3786.
3. P. Ordoukhanian and G. F. Joyce. A molecular description of the evolution of resistance. *Chem. Biol.* 1999, 6(12), 881-889.
4. P. Ordoukhanian and J.-S. Taylor. Caged single and double strand breaks. *Bioconjugate Chem.* 2000, 11(1), 94-103.
5. T. L. Sheppard\*, P. Ordoukhanian\* and G. F. Joyce. A DNA enzyme with N-glycosylase activity. *Proc. Natl. Acad. Sci. USA* 2000, 97(14), 7802-7807. \*These authors contributed equally to the work.
6. P. Ordoukhanian and G. F. Joyce, J. RNA-cleaving DNA enzymes with altered regio- or enantiospecificity. *Am. Chem. Soc.* 2002, 124(42), 12499-12506.

**RESEARCH SUPPORT**

Ongoing/Completed Research Support  
 None

**BIOGRAPHICAL SKETCH**

Provide the following information for the key personnel in the order listed for Form Page 2.  
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME Stephen R. Quake		POSITION TITLE Thomas E. and Doris Everhart Professor of Applied Physics & Physics	
EDUCATION/TRAINING (Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)			
INSTITUTION AND LOCATION	DEGREE (if applicable)	YEAR(s)	FIELD OF STUDY
Stanford University, Stanford, CA	B.S.	1991	Physics
Stanford University, Stanford, CA	M.S.	1991	Mathematics
Oxford University, England	D. Phil.	1994	Physics
Stanford University, Stanford, CA	Postdoc	1994-1996	Biophysics

**A. POSITIONS AND HONORS****Positions and Employment:**

1994-1996 Post-doctoral associate, Steven Chu group, Stanford University  
 1996-1999 Assistant Professor, Dept. of Applied Physics, California Institute of Technology  
 1999-2002 Associate Professor, Dept. of Applied Physics, California Institute of Technology  
 2002-2003 Associate Professor, Dept. of Applied Physics & Physics, California Institute of Technology  
 2003-2004 Professor, Dept. of Applied Physics & Physics, California Institute of Technology  
 2004-2004 Thomas E. and Doris Everhart Professor of Applied Physics & Physics, California Institute of Technology  
 2004-present Professor, Dept. of Bioengineering, Stanford University

**Other Experience and Professional Memberships:**

1999-present Panel Member, Biological Analysis Systems Expert Panel, MITRE  
 1999, 2000 Program Sub-Committee member and session chair, CLEO/QELS  
 2000 Visiting committee for Bioengineering, RPI  
 2000 Panel member, ARES II (DoD)  
 2001-2002 Organizing Committee for the NAS Symposium for Frontiers in Science  
 2002-2003 Defense Science Study Group  
 Occasional Peer reviewer for NSF, NIH, Foundation for Fundamental Research on Matter (Netherlands), *Science*, *Physical Review Letters*, *Physical Review*, *Nature Biotechnology*, *Proceedings of the National Academy of Sciences*, *Nature Structural Biology*, and others.

**Honors and Awards:**

1991 Phi Beta Kappa (Stanford University)  
 1991 Firestone Prize (Stanford University)  
 1991 Apker Award, (American Physical Society)  
 1991 Marshall Scholar  
 1993 National Science Foundation Graduate Fellow  
 1996 Visiting scholar at "Topology and Geometry in Polymer Science", U. Minnesota  
 1997 NIH R29 "FIRST" award  
 1997 NSF "Career" award  
 1999 Participant in the N.A.E. Symposium for Frontiers in Engineering  
 1999 Packard Fellow  
 2000 Participant in the N.A.S. Symposium for Frontiers in Science  
 2001 Named one of the "top 10 innovators in emerging technology" by MIT Tech Review magazine  
 2002 Named to the TR100 "100 Young Innovators that will create the future" by MIT Tech Review magazine

**B. SELECTED PEER-REVIEWED PUBLICATIONS**

1. Thomas T. Perkins, Stephen R. Quake, Douglas E. Smith and Steven Chu, "Relaxation of a Single Polymer Chain Observed by Optical Microscopy," *Science* 264, 822 (1994).
2. Stephen R. Quake, "The Zimm Model Applied to Extended Single Polymers," *J. Chem. Phys.* 101, 4307 (1994).
3. Stephen R. Quake, "Topological Effects of Knots in Polymers," *Phys. Rev. Lett.* 73, 3317 (1994).
4. Stephen R. Quake, "Fast Monte Carlo Algorithms for Knotted Polymers," *Phys. Rev. E* 52, 1176 (1995).
5. Stephen R. Quake, Hazen Babcock and Steven Chu, "The Normal Mode Structure of Extended Single Molecules of DNA," *Nature* 388, 151 (1997).

6. J.P. Brody and S.R. Quake, "A Self-Assembled Microscopic Rotational Probe", *Appl. Phys. Lett.* 74, 144-6 (1999).
7. H.P. Chou, C. Spence, A. Scherer and S. Quake, "A Microfabricated Device for Sizing and Sorting DNA Molecules", *Proc. Nat'l Acad. Sci.* 96, 11-13 (1999).
8. J.C. Meiners and S.R. Quake, "A Direct Measurement of the Hydrodynamic Interaction Between Two Particles", *Phys. Rev. Lett.* 82, 2211 (1999).
9. John Hatfield and S.R. Quake, "Dynamic Properties of an Extended Polymer in Solution", *Phys. Rev. Lett.* 82, 3548 (1999).
10. M. Unger, E. Kartalov, C.S. Chiu, H. Lester and S. Quake, "Single Molecule Fluorescence Observed with Mercury Lamp Illumination", *Biotechniques* 27, 1008 (1999).
11. A. Fu, C. Spence, A. Scherer, F. Arnold and S. Quake, "A Microfabricated Fluorescence Activated Cell Sorter", *Nature Biotechnology* 17, 1109 (1999).
12. T.J. Yang, Guillaume Lessard and Stephen R. Quake, "An Apertureless Near Field Microscope for Fluorescence Imaging", *Applied Physics Letters* 76, 378 (2000).
13. M. Unger, H.P. Chou, T. Thorsen, A. Scherer and S. Quake, "Monolithic Microfabricated Valves and Pumps by Multilayer Soft Lithography", *Science* 288, 113 (2000).
14. J.C. Meiners and S.R. Quake, "FemtoNewton Force Spectroscopy of Single Extended DNA Molecules", *Physical Review Letters* 84, 5014 (2000).
15. C.S. Chiu, E. Kartalov, M. Unger, S. Quake and H. Lester, "Single molecule measurements calibrate green fluorescent protein surface densities on transparent beads for use with "knock-in" animals", *Journal of Neuroscience Methods* 105, 55 (2001).
16. S. Quake and A. Scherer, "From Micro to Nano Fabrication with Soft Materials", *Science* 290, 1526 (2000).
17. T. Thorsen, R. Roberts, F. Arnold, and S. Quake, "Coherent Pattern Formation in a Non-equilibrium Microfluidic Device", *Physical Review Letters* 86, 4163 (2001).
18. H.P. Chou, M. Unger, and S.R. Quake, "A Microfabricated Rotary Pump", *Biomedical Microdevices* 3, 323 (2001).
19. M. Van Dam and S. Quake, "Gene Expression Analysis with Universal n-mer Arrays", *Genome Research* 12, 145 (2002).
20. A. Fu, H. Chou, C. Spence, A. Scherer, F. Arnold and S.R. Quake, "An integrated microfabricated cell sorter", *Analytical Chemistry* 74: 2451-7 (2002).
21. J. Liu, M. Enzelberger, and S.R. Quake, "A Nanoliter Rotary Device for Polymerase Chain Reaction", *Electrophoresis*, 23: 1531-1536 (2002).
22. S. Eyal and S.R. Quake, "Velocity Independent Microfluidic Flow Cytometry", *Electrophoresis* 23: 2653-2657 (2002).
23. C.S. Chiu, K. Jensen, I. Sokolova, D. Wang, M. Li, P. Deshpande, N. Davidson, I. Mody, M.W. Quick, S.R. Quake, and H.A. Lester, "Number, Density, and Surface/Cytoplasmic Distribution of GABA Transporters at Presynaptic Structures of Knock-In Mice Carrying GABA Transporter Subtype 1-Green Fluorescent Protein Transfusions", *Journal of Neuroscience* 22: 10251-66 (2002).
24. H.G.R. Thompson, J.W. Harris, B.J. Wold, S.R. Quake and J.P. Brody, "Identification and Confirmation of a Module of Coexpressed Genes", *Genome Research* 12: 1517-1522 (2002).
25. J.P. Brody, B.A. Williams, B.J. Wold, and S.R. Quake, "Significance and Statistical Errors in the Analysis of DNA Microarray Data", *Proc. Nat'l Acad. Sci.* 99: 12975-12978 (2002).
26. T. Thorsen, S.J. Maerkl, and S.R. Quake, "Microfluidic Large Scale Integration", *Science* 298: 580-584 (2002).
27. C.L. Hansen, E. Skordalakes, J.M. Berger, and S.R. Quake, "A Robust and Scalable Microfluidic Metering Method that Allows Protein Crystal Growth by Free Interface Diffusion", *Proc. Nat'l Acad. Sci.* 99: 16531-6 (2002).
28. M.L. Adams, M. Enzelberger, S. Quake and A. Scherer, "Microfluidic integration on detector arrays for absorption and fluorescence microspectrometers", *Sensors and Actuators A* 104: 25-31 (2003).
29. E.P. Kartalov, M.A. Unger, and S.R. Quake, "Polyelectrolyte surface interface for single molecule fluorescence studies of DNA polymerase", *Biotechniques*, 34, 505-110 (2003).
30. I. Braslavsky, B. Hebert, and S.R. Quake, "Sequence Information Can Be Obtained From Single DNA Molecules", *Proc. Nat'l Acad. Sci.* 100: 3960-4 (2003).
31. A. Groisman, M. Enzelberger, and S.R. Quake, "Microfluidic Logic and Control Devices", *Science*, 300: 955-8 (2003).
32. G.R. Yi, T. Thorsen, V.N. Manoharan, M.J. Hwang, S.J. Jeon, D.J. Pine, S.R. Quake, and S.M. Yang, "Generation of Uniform Colloidal Assemblies in Soft Microfluidic Devices", *Adv. Mater.* 15: 1300-4 (2003).
33. J. Liu, C. Hansen and S.R. Quake, "Solving the World-to-Chip Problem with a Microfluidic Matrix", *Anal. Chem.* 75: 4718-23 (2003).
34. J.W. Hong and S.R. Quake, "Integrated Nanoliter Systems", *Nature Biotechnology* 21: 1179-83 (2003).
35. M.O. Reese, R.M. van Dam, A. Scherer, and S.R. Quake, "Microfabricated Fountain Pens for High-Density DNA Arrays", *Genome Research* 13: 2348-52 (2003).
36. C. Hansen and S.R. Quake, "Microfluidics in Structural Biology: Smaller, Faster... Better", *Current Opinion in Structural Biology* 13: 538-544 (2003).

**C. RESEARCH SUPPORT****Ongoing Research Support:**

DP1 OD000251 Quake (PI) 09/30/04-07/31/09

NIH Director's Pioneer Award

Biological Large Scale Integration

The major goal of this project is to be innovative.

Role: PI

Private Source

Quake (PI)

11/01/00-04/30/05

Proposal for Ultra High Resolution Physical & Biological Techniques for the Analysis of Single DNA Molecules

The major goals of this project are to develop ultrahigh resolution apertureless optical microscope for reading labeled molecules, image single transcription factors on lambda or plasmid using GFP fusion, and demonstrate spatially selective chemistry on fixed (small) DNA molecules.

Role: PI

DMI-0328162 Ferreira (PI)

0/01/03-09/30/08

NSF/University of Illinois

Center for Nano-Chemical-Electrical-Mechanical Manufacturing Systems (Nano-CEMMS)

The major goal of this project is to establish a multi-university center to develop nanoscale manufacturing processes.

Role: Co-Investigator

DMR-0351848-UAS-220995 Gomez (PI)

04/15/04-03/31/09

NSF

CSULA-Caltech Partnership for Research and Education in Materials

The major goal of this project is to develop novel polymeric materials.

Role: Co-Investigator

N66001-02-1-8929 Weiss (PI)

06/27/02-06/26/05

DARPA/Princeton

Molecular Evolution and Engineering of Single-Cell and Multi-Cellular Signal Processing Circuits

The major goal of this project is to develop engineered genetic circuits inspired by bacterial quorum sensing.

Role: Co-Investigator

1 R01 HG02644-01A1 Quake (PI)

03/05/04-04/30/09

NIH

Ultrasensitive Nanofluidic Devices for Genomic Analysis

The major goal of this project is to establish a center that will use integrated lab-on-a-chip microfluidic devices to study unculturable bacteria and stem cell biology.

Role: PI

HR0011-04-1-0032 Psaltis (PI)

04/01/04-06/24/08

DARPA

Center for Opto Fluidic Integration

The major goal of this project is to develop integrated fluid-optic devices for memory and communication.

Role: Co-Investigator

R01 HG003594-01 (Quake)

09/29/04-09/30/07

NIH

High Throughput Single Molecule DNA Sequencing

The major goal of this project is to develop next generation ultra high throughput DNA sequencing technology.

Role: PI

MCB-0425931 (Quake)

09/01/04-08/31/07

NSF

QSB: Charting the Regulatory Space of Hematopoiesis by Microfluidic Digital PCR

The major goal of this project is to use digital PCR to study transcription factor expression levels.

Role: PI

**Completed Research Support:**

W-7405-ENG-48 Eaglesham (PI) 09/03/03-09/30/04  
DOE/Lawrence Livermore National Laboratory  
Microfluidics Technology for Dried-Reagent ELISA  
The major goal of this project is to transfer technology for microfluidic ELISA for bioweapon detection.  
Role: Co-Investigator

CTS-0088649 Quake (PI) 09/01/00-08/31/0  
NSF  
XYZ on a Chip: Integrated Microfluidic Analysis System  
The major goal of this project is to build completely integrated microfluidic chips that will perform complex chemical and biological analysis.  
Role: PI

BES-0119493 Quake (PI) 08/01/01-07/31/0  
NSF  
Biophotonics: Spectroscopy of Picoliter Fluid Volumes for Biological and Chemical Analysis  
The major goal of this project is to build compact absorption spectrometers and fluorescence measurement systems with the capability to test and identify the composition of picoliters of analyte. Such devices are expected to find many applications in biology, medicine, chemical analysis and micro-reaction control systems.  
Role: PI

MDA972-00-1-0019 Esener (PI) 09/01/00-06/30/04  
DARPA  
Center for Chips with Heterogeneously Integrated Photonics (CHIPS)  
The major goal of this project is to develop integrated systems combining optics and microchips.  
Role: Co-Investigator

DAAD19-00-1-0392 Scherer (PI) 06/01/00-02/29/04  
DARPA  
Monolithic Integration of Microfluidics and Optoelectronics for Biological Analysis  
The major goal of this project is to develop microfluidic chips which can manipulate subnanoliter fluid volumes and perform biochemical assays.  
Role: Co-Investigator

11-1186-0502 Quake (PI) 07/01/03-10/31/03  
NIH/ Whitehead Institute for Biomedical Research  
Ultra high throughput single molecule DNA sequencing  
The major goal of this project is to use single molecule detection of fluorescently labelled nucleotides after DNA polymerase or reverse transcriptase incorporates labelled dNTPs into individual immobilized DNA or RNA molecules.  
Role: PI

Private Source

Quake (PI)

01/01/03-06/14/03

Private Source

To Develop Optimized Reagents for Single Molecule Sequencing

The major goal of this project is to develop optimized reagents for a practical single molecule DNA sequencing scheme.

Role: PI

5 R29 HG01642 Quake (PI) 07/01/97-06/30/02  
NIH/NHGRI

Fluorescent Photobleaching Method for Sequencing DNA

The major goal of this project is to propose a new method for large scale genome sequencing that combines microfabricated devices, single fluorophore detection and enzymatic base incorporation.

Role: PI

## BIOGRAPHICAL SKETCH

Excluded by Requester

---

## RESOURCES

---

**FACILITIES:** Specify the facilities to be used for the conduct of the proposed research. Indicate the performance sites and describe capacities, pertinent capabilities, relative proximity, and extent of availability to the project. Under "Other," identify support services such as machine shop, electronics shop, and specify the extent to which they will be available to the project. Use continuation pages if necessary.

**Laboratory:**

Dr. Romesberg's lab is located on the second floor of the CVN Building at The Scripps Research Institute and consists of 3,600 sq. ft. devoted to organic, bio-organic and molecular biology. All supplies and equipment required for these disciplines are available. The laboratory also contains a cold-room, biological prep room and equipment room.

**Clinical:**

N/A

**Animal:**

N/A

**Computer:**

The computational facility at TSRI consists of a CRAY T3E parallel supercomputer, an array of 4, 12 processor, SGI Power Challenge R-10000 systems, ethernet connections, email and various printers and tape drives; site license for MSI molecular modeling software. Macintosh PowerPC's, and an SGI Octane are available in the P.I.'s laboratory for molecular modeling, calculations and data analysis.

**Office:**

Dr. Romesberg's office is located in the laboratory described above. The lab also contains desk and office spaces for students, and research associates.

**Other:**

The department also provides Dr. Romesberg with full-time administrative assistant help.

---

**MAJOR EQUIPMENT:** List the most important equipment items already available for this project, noting the location and pertinent capabilities of each. Instruments available through normal departmental relationships include 250-, 300-, 400-, 500-, and 600-MHz NMR spectrometers. Well-equipped shops and other support activities, including small molecule X-ray crystallographic analysis, peptide sequencing, amino acid analysis, and oligonucleotide synthesis are available through the department on a fee for service spectrometer and two Perkin-Elmer SCIEX API-3 Ion Spray, and MALDI mass spectrometers which are all suitable for the analysis of peptides and proteins, and a confocal and deconvolution microscope are available in the on-site, supported facility. The department also provides Dr. Romesberg with full-time administrative assistant help.

The following equipment belongs to the Principal Investigator's research group: Applied Biosystems, Inc. DNA synthesizer; Rainin gradient HPLC; Cary 300 UV/Vis spectrophotometer; Hewlett-Packard diode array UV/Vis spectrophotometer; MJ Research PCR machine; New Brunswick incubator/shakers; Fisher incubator; Biorad electroporator, Nikon tetrad dissection microscope, Nikon TE200 epifluorescence microscope, multichannel/repeater pipettes, required synthetic equipment.

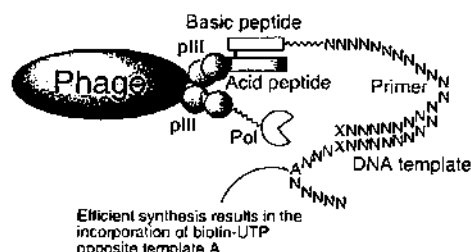
## A. Specific Aims

This R21 submission is in response to PA-03-058, 'Exploratory/Developmental Bioengineering Research Grants.' We propose the directed evolution of new DNA polymerases that will facilitate several emerging sequencing technologies. After successful completion of the proposed research we will seek additional funding to develop the most promising applications.

There is great incentive for developing low cost technologies for sequencing genomes. During the past two decades, along with computational advances, the parallelization, automation, and refinement of the Sanger method has reduced sequencing costs from tens of dollars per sequenced base to hundredths of a cent per sequenced base. However, the cost of sequencing a mammalian-sized genome is still tens of millions of dollars, and it is unlikely that further advances in instrumentation alone will achieve the 10,000-fold reduction in cost set as a goal by the NIH.

A variety of new sequencing methods currently in development have the potential to make genome sequencing commonplace. Most of these new methods rely on polymerase mediated DNA synthesis with modified deoxynucleotide triphosphate (dNTP) substrates. As a result, these DNA sequencing methods are all limited by the ability of the available polymerases to recognize modified dNTPs. Despite obvious limitations with the polymerases, and in contrast to the great effort devoted towards improvements in automation and new technique development, little effort has been directed toward using the modern methods of directed evolution to evolve enzymes that are better suited to sequencing.

We have developed an activity-based selection system to evolve polymerases with novel activities (Figure 1). As described in detail below, the selection system involves the co-display on phage of DNA polymerase libraries and an oligonucleotide substrate that primes synthesis with an attached template. If a displayed polymerase mutant recognizes the attached substrate and the provided dNTPs, then it will extend the primer, and only then covalently attach a biotin-dUTP. Biotinylated phage particles may be selectively recovered using a streptavidin solid support. We have successfully used the system to evolve several polymerases that catalyze unnatural reactions with efficiencies that rival those of natural polymerases with their native



**Figure 1.** Phage particle with attached polymerase and substrate. X represents a natural or unnatural base. For clarity, only two pIII proteins are shown. Selections are designed so that biotin incorporation is dependent on DNA synthesis with natural or modified dNTPs.

substrates [1,2]. We are now interested in using the system to evolve polymerases that will significantly improve different sequencing methodologies. This proposal focuses on evolving the Stoffel fragment of *Thermus aquaticus* polymerase I (Sf). The Klenow fragment of *E. coli* DNA Polymerase I (Kf) will serve as a backup in the event that unforeseen problems arise with Sf. The specific aims of this proposal are:

**Specific Aim 1, Year 1:** Evolve DNA polymerases that efficiently synthesize DNA using Cy5-labeled dNTPs to enable novel synthesis-by-sequencing methodologies.

**Specific Aim 2, Years 1-2:** Evolve polymerases that efficiently incorporate reversibly fluorophore tagged dNTPs with 3'-O-allyl protecting groups and then continue synthesis after fluorophore and allyl group removal.

**Specific Aim 3, Year 2:** Collaborate to evaluate evolved polymerases in actual sequencing applications.

The overall goal of the proposed research is the directed evolution of DNA polymerases to enable revolutionary sequencing methodologies. To facilitate the last specific aim, we have established collaborations with Prof. Steve Quake (Stanford), Prof. Jingyue Ju (Columbia), and Dr. Phillip Ordoukhanian, director of the Scripps sequencing facility.

## B. Background and Significance

Currently, Sanger dideoxy sequencing and capillary electrophoresis is the most well-established and popular technique employed for DNA sequencing.

In this approach, an oligonucleotide primer is hybridized to a template to be sequenced and extended by a DNA polymerase which incorporates dNTPs present in the reaction mixture. Dideoxynucleotide triphosphates (ddNTPs) are also present, and when a ddNTP is incorporated rather than a dNTP, primer elongation is terminated. The ddNTPs are labeled with unique fluorophores. When the DNA fragments are electrophoretically separated, base identity may be determined as a function of fragment length.

While the parallelization, automation and refinement of the Sanger method has reduced sequencing costs to less than one cent per base, there is a growing consensus that novel approaches to sequencing will be required for the dramatic reductions in cost and time required to make ultra-low cost sequencing possible. The majority of approaches in development today analyze sequences by directly detecting polymerase-mediated dNTP insertion. The most promising of these 'sequencing-by-synthesis' (SBS) methodologies are reviewed below, in section **B.1**.

## **B.1 Sequencing-by-Synthesis Methodologies.**

Generally, SBS involves a series of enzymatic reactions performed on either a single or a population of identical primer/templates. Each nucleotide incorporation event results in a distinct signal, allowing the observer to determine the identity of the incorporated nucleotide, and after repeated cycles, the sequence of the oligonucleotide in question. SBS techniques can be broadly classified according to their use of single or amplified molecules.

**(B.1.1) SBS and Amplified Templates.** In this approach, a target oligonucleotide is isolated and PCR amplified to produce up to millions of identical copies. The amplified copies are then physically separated and subjected to DNA polymerase mediated synthesis and analysis. The advantage of first amplifying the template is high signal intensity. This approach requires both the spatial separation of the templates and the controlled detection of dNTPs added by a polymerase.

Detection in many of the most promising SBS techniques is based on DNA polymerase-mediated synthesis of DNA using dNTPs tethered to unique fluorophores. Each dNTP is added in separate steps, followed by a washing step, and only dyes

that become covalently linked to the growing primer are detected [3]. Inefficient incorporation of the modified dNTPs inherently limits techniques using amplified templates because it results in a fraction of the primers not being extended. In subsequent rounds, the sequence of the corresponding oligonucleotides will be shifted or 'dephased' from that of the others. This results in degradation of the signal-to-noise and limits the read length. Dephasing results from the inability of available polymerases to efficiently synthesize DNA containing the modified dNTPs, which is further discussed below in section **B.2**.

**(B.1.2) SBS and Single Molecules.** When single molecules are detected, sequence dephasing is eliminated as a problem; 'dead rounds' of sequencing are followed by synthesis in later rounds without loss of information.

The challenge for SBS on single molecules is primarily one of signal intensity and detection. The Quake group has approached this problem by using confocal microscopy and fluorescence resonance energy transfer (FRET) between donor and acceptor fluorophores incorporated into the growing primer. dNTPs with an attached donor group are first incorporated. In subsequent rounds, dNTPs with an attached acceptor group are incorporated, allowing for their visualization by FRET. Between additions, the acceptor is photobleached to ensure that the signal reflects the dNTP incorporated at a given step, and not at a previous one. Unfortunately, as is the case with amplified sequences, polymerase-mediated synthesis using the required modified dNTPs is extremely inefficient.

## **B.2 Polymerase-Specific Limitations.**

The major limitation of SBS methodologies is the need to synthesize DNA using modified dNTPs [4]. These substrates are typically poorly recognized by polymerases. In the case of dNTPs with a modified nucleobase, incorporation is typically less efficient than that of a natural dNTP, and it becomes even less efficient with each modified dNTP added. (Indeed, this is generally true for polymerase-mediated synthesis of any modified DNA [5,6].) Thus, inefficient continued primer elongation limits read length [7].

The Quake group has addressed the problem using two different approaches. In the first, modified dNTPs are only incorporated with natural nucleotides regularly interspersed [3]. Attempts to

incorporate consecutive labeled dNTPs were unsuccessful. In a second approach, the modified dNTPs are used in the presence of an excess of their natural counterparts [7]. Because the majority of the primers are extended with a natural dNTP, they remain competent for extension with a modified dNTP at another position in the template. Thus, as they conclude in a recent publication, it is currently irrelevant whether tagged dNTPs or tagged ddNTPs are used, because in either case primer extension does not continue, and the signal drops exponentially with every sequenced base [7].

Another solution to this problem is dNTPs with reversibly linked dyes that allow removal of the fluorophore after incorporation. The Ju group at Columbia has reported a set of four dyes, one linked to each dNTP by a photolabile group, that were used to sequence a population of identical oligonucleotides [8]. However, only one nucleotide was added at a time and the signal appeared to fade with the number of bases sequenced (again, indicative of decreasing synthesis efficiency). The Church group used dNTPs with disulfide linked fluorophores which may be removed by reductive cleavage [9]. However, with all of these reversibly tethered fluorophores, even after deprotection, the nascent primer terminus remains modified and poorly recognized by DNA polymerases. For example, both studies required micromolar enzyme concentrations and extended incubation periods to drive extension of the 'deprotected' primer termini.

The Ju group has pioneered the development of novel dNTPs that not only bear photo-detachable nucleobase-linked fluorophores (see above), but that also bear 3'-OH protecting groups. First proposed by Metzker in 1994 [10], reversible modification of the 3'-OH allows for strict control over the synthesis beyond the modified nucleotide since the required nucleophile is not available until deprotection. When optimized, 3' modification should allow all four labeled dNTPs to be present in a given reaction, as opposed to flushing the system with solutions containing only one labeled dNTP at a time. In addition, these protecting groups would prevent multiple incorporations of the same dNTP (modified or unmodified) at homopolymeric runs. Specifically, the Ju group has developed 3'-O-allyl protecting groups that may be catalytically removed with palladium [11]. However, like all 3'-O-modified dNTPs examined, as well as the nucleobase-modified dNTPs discussed above, these 3'-protected dNTPs are very inefficient substrates

for DNA polymerases. For example, after an extensive screen of different polymerases, it was determined that exo<sup>+</sup> 9°N polymerase (at the time commercially available from New England Biolabs, but its sale has been discontinued since) was capable of incorporating 3'-O-allyl-dTTP. However, incorporation required 27 minutes and the use of micromolar concentrations of enzyme. Polymerase fidelity was not reported, but under these extraordinary enzyme concentrations it is unlikely to be suitable for sequencing applications. In addition, this polymerase was unable to extend the doubly deprotected primer terminus (nucleobase-fluorophore and 3'-O-allyl deprotected). Extension required equally extraordinary concentrations of Thermo Sequenase. Additionally, the approach was only demonstrated for the addition/extension of a single, doubly protected dNTP and it seems certain that as observed with the other approaches, synthesis will become progressively more inefficient with each nucleotide sequenced.

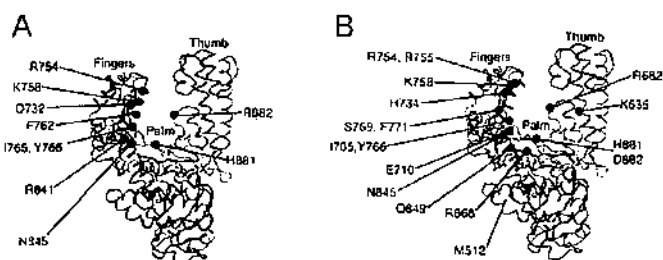
Simply put, DNA polymerases did not evolve to synthesize DNA with modified substrates, and as a result, they do so only poorly. With the powerful techniques of directed evolution, the optimization of polymerases for these applications is expected to significantly increase their practical potential.

### **B.3 Type I DNA Polymerase Structure & Function.**

Many 'Type I' (or 'A family') polymerases have played an important role in sequencing efforts, including both Taq (from which Sf is derived) and Kf [12]. Careful consideration of the structure and substrate repertoire of these polymerases is expected to facilitate their directed evolution.

**(B.3.1) Type I DNA Polymerases Structure and Mechanism.** In addition to being the polymerases of choice for sequencing, Type I DNA polymerases are an attractive starting point for directed evolution experiments due to their small size, thorough characterization [13], and structural and functional homology [14,15].

Crystal structure data for each of these type I polymerases shows a structure resembling a right-hand with 'finger,' 'palm,' and 'thumb' subdomains forming a deep substrate binding cleft (Figure 2). The duplex DNA approaches the polymerase domain from the 3'-5' exonuclease side of the cleft and binds with the primer terminus in the polymerase active site, packed against the N-



**Figure 2.** Kf residues implicated in substrate binding by crystallographic and biochemical experiments. (a) dNTP binding site includes: Arg882 [20], Asp732 [21] Arg754 [22], Lys758 [22-25], Phe762 [22], Ile765 [13,26], Tyr766 [27-29], Arg841 [30], Asn845 [30], and His881 [31]. (b) Template binding site includes: Met512 [32], Lys635 [33], Arg668 [34], Arg882 [22,35], Glu710 [34], His734 [22,36], Arg754 [22], Arg755 [36], Lys758 [23-25], Ile765 [31], Tyr766 [26,28,29,37], Ser769 [38], Phe771 [38], Arg841 [34], Asn845 [34], Gln849 [34], His881 [24,34], and Asp882 [34].

terminal end of the O  $\alpha$ -helix of the fingers domain. The side chains of Arg841 and Gln849 (Kf numbering) donate hydrogen-bonds to the N3 atom of purines and the O2 atom of pyrimidines, respectively. These H-bond acceptors in the DNA are sequence independent and are located in a widened minor groove at the primer terminus end of the duplex. The template base and incoming dNTP bind in a pocket formed by residues in the O-helix as well as residues in the loop connecting the O and O1 helices (including Phe762, Gly763, Tyr766).

The third strand of the anti-parallel  $\beta$  sheet that forms the 'palm' subdomain runs along the bottom of the active site and contains a conserved Glu residue which chelates one of the two catalytic  $Mg^{2+}$  ions. This strand then merges into the O-helix. Residues 614 and 615 (Sf numbering) are located at the junction of the sheet and O-helix and pack on the sugar ring of the incoming dNTP. The analogous residues in several homologous polymerases are thought to act as a 'steric gate' to discriminate against the incorporation of triphosphates with 2'-substituents [16-19].

Several residues appear to form important contacts with the nucleobase portion of the incoming dNTP (important for incorporation of modified dNTPs) or the primer terminus (important for continued synthesis after incorporation of the modified dNTP): Arg754 [22,23] Lys758 ([23] Phe762 [22,23] and Tyr766 [23,28].

### **(B.3.2) DNA Polymerase Recognition of Modified Substrates.**

*(B.3.2.1) DNA polymerases and sugar modified substrates.* Polymerase recognition of 3' hydroxyl

modified nucleoside analogues have been studied mainly in the context of inhibitors or permanent terminators for Sanger sequencing. Generally, polymerases tolerate small polar groups at the 3' position, such as the native hydroxyl group or close variants such as amino, fluoro, or methoxy groups.

Limited success has been achieved in the attempt to find a reversible 3'-OH protecting groups that are recognized by DNA polymerases. Metzker and coworkers systematically synthesized seven dNTP analogs with small O-linked protecting groups and analyzed their incorporation into DNA by nine different polymerases [10]. One protecting group, O-2-nitrobenzyl, was incorporated into DNA, cleaved with UV light, and the growing primer was subjected to additional rounds of synthesis. Unfortunately, the enzymatic incorporation as well as deprotection with UV light were far too inefficient for SBS techniques. Further derivatization of the O-2-nitrobenzyl group was also studied, but none of the analogues were incorporated into DNA by any polymerase examined [90]. Ju recently reported that an allyl group may be used to protect the 3' hydroxyl, however, as discussed above, this optimized dNTP modification still inhibits polymerase recognition.

DNA polymerases appear even less tolerant to substitution at the dNTP C2' position. This is likely due to the fact that DNA polymerases evolved to synthesize DNA in the presence of excess rNTPs. In one report, Smith and co-workers examined a wide variety of DNA polymerases to identify those that can synthesize DNA with 2'-fluoro-modified dNTPs. A screen of different polymerases showed that Ultima polymerase (commercially available from PE Applied Biosystems) was best able to synthesize polymers with 2'-fluoronucleotide triphosphates, but only in the presence of high concentrations of  $Mn^{2+}$ , and even then, the 99-mer full-length product was only evident after overnight reactions [39].

*(B.3.2.2) DNA Polymerases and nucleobase modified substrates.* The recognition of nucleosides bearing wholly unnatural nucleobase analogs (i.e. nucleobase analogs with non-natural H-bonding topologies or those that are predominantly hydrophobic) has been extensively examined with DNA polymerases. Although the details of these studies are not specifically relevant to the current proposal, they have revealed several interesting general trends. For example, the ability of polymerases to incorporate a modified dNTP

decreases significantly as more are added. In fact, only incorporation of the first modified dNTP (extending a primer that terminates with an *unmodified* nucleotide) is typically even moderately efficient. As discussed above, polymerase-mediated incorporation of multiple modified dNTP analogs is also severely limiting for SBS methodologies. For a more complete discussion of the literature, the reader is referred to a more comprehensive review [6].

For sequencing applications, nucleobase modification at the pyrimidine C5 position and the purine C8/deaza C7 positions are more relevant. The ability of several DNA polymerases to recognize dNTPs with these modifications has been evaluated qualitatively. For example, Vent polymerase [40-42], Taq [42,43], a modified T7 DNA polymerase (Sequenase 2.0™) [44,45], and the Pwo polymerase from *Pyrococcus woesei* [46] have all been shown to recognize C5-modified pyrimidines. However, the nature of the substituent is critical, for example rigid moieties appear to be best tolerated [40,42,43,47]. Taq [47] and Sequenase 2.0™ [44,45] recognize 7-deaza- and C8-substituted purines, respectively. However, after incorporation of a modified dNTP, the incorporation efficiency of the next dNTP (natural or unnatural) decreases dramatically [7].

## **B.4 Polymerase Evolution**

Classically, polymerase mutants with novel properties have been identified based on rational design followed by mutant-by-mutant construction and characterization. Several mutants with altered dNTP sugar recognition have been identified with such methods. For example, structural studies suggested that the F155 residue of Moloney murine leukemia virus reverse-transcriptase [48] and the E710 residue of Kf [16] are positioned to prevent ribotriphosphate (rNTP) incorporation. Indeed, characterization of the F155V and E710A mutants revealed that they both incorporate rNTPs more efficiently than their respective wild-type enzymes. Likewise, the Y639F:S641A [49-51] and Y639F [52] mutants of RNA polymerase from T7 bacteriophage utilize dNTP substrates. The Y639F mutant also exhibits reverse transcriptase activity, and the Y639F:H784A double mutant can synthesize oligomers using 2'-O-methyl or 2'-azido pyrimidine triphosphates [53].

**(B.4.1) Methods for the Directed Evolution of Polymerases.** While studies of polymerases by rationally designed mutation have allowed for the identification of certain active-site residues that facilitate recognition of dNTPs with modified sugar rings, the novel activities are, without exception, only modest (relative to natural activities). The rational approach has had even less success with more structurally perturbed dNTPs, such as those bearing fluorophore labeled nucleobases. Because of the large number of residues in the polymerase active site, and because residues located far from the active site may contribute significantly to the recognition of these substrates, it is desirable to develop techniques that more efficiently sort through large numbers of mutant polymerases. Both screens and selections based on activity have been pursued. In either method the construction of a polymerase library is first required.

**(B.4.1.1) Library Construction.** Mutations that diversify a library of proteins may be focused to a specific set of residues, or distributed throughout the entire sequence of the protein by means of automated nucleic acid synthesis, chemical agents that covalently modify nucleobases, error-prone PCR, or propagation in a bacterial mutator strain [54-56]. Alternatively, multiple genes can be combined to produce chimeric proteins by recombination-based methods like gene shuffling [57,58], staggered extension PCR [59], and random-priming recombination [60]. Recombination of homologous genes, or family shuffling, is expected to allow for the generation of more highly mutated proteins as each mutation is less likely to disrupt the tertiary structure of the encoded protein [61,62]. A particularly promising variant of family shuffling makes use of synthetic oligonucleotides that contain patches of sequence from different but related genes, as well as ends that are homologous to the gene of interest, which facilitates recombination [63].

**(B.4.1.2) Activity-Based Screening.** Polymerase libraries may be directly subjected to activity-based assays to identify those members of the library that possess a desired activity [64]. Typically, each polymerase mutant is spatially isolated, and incubated with the appropriate substrates. The products are then analyzed by conventional gel electrophoresis to determine if any of the mutants showed the desired activity. Without special technology, only hundreds of mutants may be screened.

DNA polymerase libraries may be enriched in active members prior to screening using a technique developed by Loeb and coworkers [15,65,66] that is based on the genetic complementation of an *E. coli* strain JS200 (*recA718 polA12 uvr155 trpE65 lon-11 sulA1*), which carries a temperature-sensitive mutant of DNA polymerase I, Pol I<sup>s</sup>. The encoded polymerase mutant (D186N) is functional at 30 °C, but not at 37 °C. Thus, libraries propagated in this strain at 37 °C become enriched in DNA polymerase mutants that are physiologically active.

(B.4.1.3) *Activity-Based Selections.* A selection differs from a screen in that the selection process couples persistence to activity. After sufficient enrichment, only the most fit clones remain. Thus, much larger libraries may be examined. In addition, because the selection takes place outside of the cell, greater control over the selection conditions are possible; however, genotype-phenotype association must be achieved by creative means.

There are two *in vitro* activity-based selection methods that have been used successfully for the directed evolution of polymerases. The first method, compartmentalized self-replication (CSR), combines cells that have overexpressed the polymerase of interest with an aqueous solution containing gene-specific primers, nucleoside triphosphates, divalent metal cofactor, and RNase. The cell-substrate suspension is stirred into a mineral oil-based detergent solution to generate an emulsion composed of segregated compartments 15 µm in diameter that contain all of the substrates, but only one cell [67,68]. The emulsion is heated to release the polymerase from the cell, and active clones replicate their own gene. Lastly, the amplified genes are recovered in the aqueous phase and subcloned for further rounds of selection. In this way, the polymerase library becomes enriched in its most active members.

Another technique for maintaining genotype-phenotype association *in vitro* is phage display, which is the covalent attachment of proteins of interest to the surface of filamentous bacteriophage by their expression as a fusion with a phage coat protein [69]. Both Kf and Sf have been displayed on phage as pIII fusions [1,2,70,71]. Substrates have been attached to the same phage particles via either pIII [1,2] or the more abundant pVIII coat protein [70,71]; however, only the former strategy has been used to evolve polymerase mutants. In this case, a typical phage preparation, generated by superinfection of F-pilus presenting *E. coli* cells that

harbor a phagemid encoding a polymerase mutant, yields phage particles that display 0-1 polymerase-pIII fusions, while the remainder of the pIII proteins are fused to 'acidic' peptides (Figure 1). The acidic peptide is used to attach a natural or unnatural nucleic acid substrate, which has been conjugated previously to a 'basic' peptide. The acidic and basic peptides form a stable heterodimer as well as a disulfide bond. Protected within each phage particle is the phagemid containing the gene that encodes the displayed polymerase mutant [69]. In this manner, the polymerase mutant, its DNA sequence, and its substrate are all associated with a single phage particle. Intramolecular activity results in the selective biotinylation of phage particles bearing active mutants, allowing for their isolation by means of streptavidin-coated magnetic beads. The recovered polymerase mutants may be analyzed or subjected to additional selections.

#### **(B.4.2) Polymerases Evolved to Utilize Alternative Substrates.**

(B.4.2.1) *DNA polymerases evolved into RNA polymerases.* While the evolution of a DNA polymerase into an RNA polymerase is by definition proof-of-principle, it nonetheless represented important progress in methodology development. Seven mutants of DNA polymerase I from *T. aquaticus* or *E. coli* that utilize rNTPs have been identified, four by activity-based screening and three by phage display-based directed evolution [2,66,72,73].

Taq mutants I614K and I614N:L616I were identified from a library of  $2 \times 10^5$  by means of genetic complementation followed by activity-based screening [72]. Steady-state kinetic analysis demonstrated that both mutants retained wild-type activity and could incorporate rNTP substrates 200- to 500-fold more efficiently than the parental enzyme, due primarily to a decrease in  $K_M$ . The I614N:L616I mutant synthesized RNA transcripts of up to 8 nucleotides in the presence of  $Mn^{2+}$ , with significant pausing at template dA residues, and with reduced fidelity [74]. Taq mutant A661E was isolated in a similar manner [15], and shown to incorporate rGTP opposite template dC more efficiently than wild-type by steady-state kinetic analysis; however, it also exhibited significantly compromised fidelity [73]. In a similar study, Kf mutant I709F was found to incorporate single rNTP substrates 30-100 times more efficiently than wild-type under steady-state conditions, with rCTP and

rGTP incorporated better than rATP or UTP [66]. In a transcription assay, I709F yielded a 23 nucleotide full length product, but paused at adjacent purines in the template and required high concentrations of  $Mn^{2+}$  in addition to  $Mg^{2+}$ . For a discussion of our efforts to evolve DNA polymerases into more efficient RNA polymerases using phage display, see application 1 in section C.2)

(B.4.2.2) *Polymerases evolved to utilize unnatural substrates.* The ultimate goal of directed evolution is the production of enzymes with arbitrarily tailored properties. This is particularly challenging in the case of DNA polymerases. However, three recent studies have reported success. First, we used the phage-based selection system described above to identify mutants of Sf that efficiently and selectively incorporate all 2'-O-methylribonucleoside triphosphates opposite a DNA template (see application 2 in section C.2). By using CSR to evolve mutants of Taq that efficiently extend mispaired primer termini, Holliger and coworkers isolated several mutants that utilize a wide variety of unnatural substrates [75]. For example, mutants M1 (G84A:D144G:K314R:E520G:F598L:A608V:E742G) and M4 (D58G:R74P:A109T:L245R:R343G:G370D:E520G:N583S:E694K:A743P) can replicate templates containing an abasic site, a cis-syn cyclobutane pyrimidine dimer, or the 5-nitroindole base analogue. Interestingly, although these mutants were selected based on their ability to extend mismatched primer termini, and despite their 100-400-fold greater efficiency ( $V_{max}/K_M$ ) in extending natural mismatches, M1 and M4 were shown to have an error-rate only 1.6 times that of Taq. Most recently, we evolved a polymerase to more efficiently synthesize DNA containing a wholly unnatural base pair, in fact, one wherein molecular recognition is mediated by hydrophobic forces instead of hydrogen-bonding or shape complementarity (see application 3 in section C.2).

## C. Preliminary Studies

### C.1 The Development of a Phage Display Activity-Based Polymerase Selection System

Phage display has emerged as the method of choice to evolve proteins based on affinity for a given target, but selecting for catalytic activity has proven more difficult. We have developed a modified phage selection system that allows for display of polymerase libraries and direct selection for activity under easily controlled selection conditions.

#### (C.1.1) *The activity-based phage selection system.*

In our activity-based selection method, both a DNA polymerase and its substrate are attached to the M13 phage minor coat protein, pIII (Figure 1). pIII proteins are localized to one end of the phage particle, which is expected to favor intramolecular reaction of the enzyme with its associated substrate, thus ensuring the linkage of genotype and phenotype. Selection of the desired mutants from a library of mutants depends on enzyme-catalyzed labeling of the attached substrate with biotin, which allows for the selective isolation of phage particles displaying active polymerases. There are four key aspects of this selection system: i) polymerase display on phage, ii) efficient substrate attachment, iii) polymerase library design, and iv) selection of active polymerases.

(C.1.1.1) *Polymerase display on phage.* We have successfully displayed both Sf and Kf polymerases on phage.

Each polymerase gene was cloned into a modified pFAB5.His phagemid vector, fusing the C-terminus of the polymerase to the N-terminus of the phage protein pIII via a soluble and flexible GGSGGS linker yielding the phagemids pFAB.SF and pFAB.KF (Figure 3). In addition to the phagemid encoded pol/pIII fusion, the selection system relies on X30 helper phage, which encodes a variant of the pIII protein that is N-terminally extended by fusion to an 'acidic' peptide. The acidic peptide is used to attach substrate to the phage particle proximal to the displayed polymerase (see C.1.1.2). *E. coli* XL1-Blue harboring pFAB.SF or pFAB.KF are superinfected with this X30 helper phage to generate phage particles. During packaging, the phage particles retain the polymerase encoding phagemid DNA. The phage particle coat carries an average of five pIII proteins; one of these is fused to the polymerase and four to the acidic peptide. Polymerase display was verified by phage ELISA using anti-polymerase antibodies, and activity was demonstrated by radiolabeled primer extension.

(C.1.1.2) *Substrate attachment.* DNA that is to act as the primer during the selection is synthesized

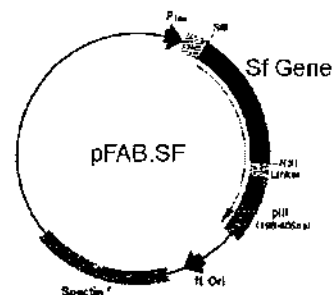


Figure 3. Map of phagemid encoding the Sf-pIII fusion

with a 5' photoprotected thiol so that it may be attached to a 'basic' peptide via a bismaleimide linker [1]. After hybridization of the basic peptide-primer conjugate to a template oligonucleotide, the basic peptide-primer/template is selectively attached to the phage by way of a high affinity leucine-zipper complex between the acidic and basic peptides. The complex is further anchored by formation of a cysteine disulfide bond between the peptides. The resulting phage particles are resuspended in 25 mM Tris-HCl pH 7.4, 140 mM NaCl, 2.5 mM KCl, 10 mM EDTA, 0.1 % BSA for subsequent selections. In this manner, the polymerase, DNA encoding the polymerase, and duplex DNA substrate are all physically associated with the same phage particle (Figure 1). Extensive model selections have been performed to optimize the technique [1,2].

(C.1.1.3) *Polymerase library design and construction.* Five focused libraries of Sf and Kf were created by two-step overlapping extension PCR [76] based on the corresponding crystal structures [77,78]. The doping ratio was set between 8% and 45%, depending on the size of the region to be mutated. Each focused library contained mutations localized to a specific region of the polymerase: two regions of the metal binding site, amino acids 596-615 (693-712, Kf residues in parentheses) and 783-786 (870-892); a portion of the template binding site, 728-734 (826-847); the duplex binding site, 568-587 (663-683); and the O-helix, 665-676 (750-771). To assess the quality of the libraries, forty-eight randomly selected clones from each of the five were sequenced. Approximately 30-50% of these clones contained frameshift mutations that are attributed to side products produced during primer synthesis (typical for oligonucleotide-based libraries). A final library for both Kf and Sf was generated by combining the five focused libraries so that each was equally represented (this library was used in Applications 1 and 3, below).

Two additional libraries of Sf (used in Application 2, below) were also created. The library IEFYNQ was randomly mutated at Ile614, Glu615, Phe667, Tyr671, Asn750, and Gln754, and the library RMQVHE was randomly mutated at Arg573, Met747, Gln754, Val783, His784, and Glu786. These residues were chosen based on their proximity to the primer terminus. Randomization was accomplished using the 1.2 kb *FseI*-*NotI* fragment of the Sf gene, which was generated by overlapping extension PCR. In the first step, four

fragments were generated from primers containing mutations. In the second step, the overlapping fragments were used as templates in conjunction with two external end primers to generate full-length polymerase gene products. In the final step, full-length PCR product was gel purified and re-amplified using two internal end primers, digested with *FseI*/*NotI*, and ligated into gel purified *FseI*/*NotI* digested pFAB.SF.

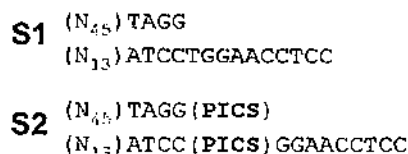
Based on the number of phage used in each selection ( $5 \times 10^{12}$ ), the size of each library (determined by random clone sequencing) and the efficiency of polymerase display (0.1%), we approximate that each clone was displayed on ten phage particles on average.

## C.2 Polymerase Evolution: Applications of the Polymerase Selection System.

### *Application 1: Evolving a DNA polymerase into an RNA polymerase.*

As an initial test of the activity-based selection system, we evolved Sf into an RNA polymerase [2]. The S1 primer/template substrate (Scheme 1) was attached to the pol/phage particles, and selections were done at 50 °C for 15 minutes in the presence of ATP, CTP, and UTP-biotin (50  $\mu$ M each). Phage biotinylation and recovery relied on the incorporation of rNTPs. After four rounds of selection, twenty Sf clones were isolated, subcloned into a modified pET-23b(+) expression vector (C-terminal His-tag), and transformed into *E. coli* BL21(DE3). Cell free extracts were prepared and heat-treated to inactivate endogenous *E. coli* proteins. Six of the clones expressed Sf, and three of these (clones SFR1, SFR2, and SFR3) could efficiently extend a DNA primer by insertion of ATP opposite dT. The plasmids corresponding to SFR1-3 were sequenced (Table 1), and the proteins were expressed and purified by nickel affinity chromatography.

The evolved polymerases were characterized using a steady-state single nucleotide kinetics assay. The details of these studies are published [2]. For purposes of this proposal, we only point out that



**Scheme 1.** Primer-template duplexes for activity based phage selections.

**Table 1.** Sequence alignment of wild-type Sf and Sf mutants isolated by activity dependent selections. Dashes indicate identical residues to wild type Sf.

	489	531	596	601	605	610	614	615
Sf-WT	Q	K	RAPIA	EKGWL	LVALD	YSQIE	I	E
SFR1	-	E	-T--T	---G-	--SV-	---TG	-	-
SFR2	-	-	-T---	-----	-----	---G	-	-
SFR3	-	-	-T---	---RQ	-----	---TG	-	-
SFM01	-	-	-----	-----	-----	-----	Y	G
SFM18	-	-	-----	-----	-----	-----	Y	G
SFM19	-	-	-----	-----	-----	-----	E	G
SFM30	-	-	-----	-----	-----	-----	E	A
SFP2	H	-	--I--	-----	-----	-----	F	-

each polymerase exhibited  $10^3$ - to  $10^4$ -fold faster rates for insertion of each rNTP relative to wild-type Sf and required only normal PCR-like conditions. In fact, one mutant (SFR3) synthesized RNA as efficiently as Sf synthesizes DNA. This is in marked contrast to the much more modest activities associated with the mutants identified by screening and described in section B.4.2.1.

Application 2: Evolving a DNA polymerase that recognizes dNTPs with modified sugar rings.

To demonstrate that the activity based phage selection system is capable of evolving a *bona fide* unnatural activity, we used it to evolve polymerase variants that efficiently synthesize polymers comprised of 2'-O-methyl ribonucleosides (OMeNTPs) [1]. These 'bio-like' polymers have properties that make them attractive for several biotechnology applications [79-85], but they have not been developed for use because no known polymerase accepts the OMeNTP substrates.

In this case we used the IEFYNQ and RMQVHE libraries which focus the diversity to regions of the protein proximal to the primer terminus (see section C.1.1.3). The selections were conducted with S1, as described above. The only difference was that in this case, biotinylation required the insertion of OMeCTP and OMeATP instead of rNTPs. After four rounds of selection, mutant Sf genes were isolated, cloned into the pET-23b(+) vector, and prepared as described above. Eighty clones were randomly chosen from each library, and more than half expressed Sf. After partial purification, protein isolated from four of the IEFYNQ library clones were identified that could efficiently extend a DNA primer by inserting OMeATP opposite dT in the template. Sequencing these clones revealed that all four had two mutations at the adjacent residues

Ile614 and Glu615, with no mutation of Phe667, Tyr671, Asn750, or Gln754 (Table 1). Clones SFM01 and SFM18 encode identical amino acid sequences, but have different DNA sequences, showing that a protein sequence was independently isolated twice and thus implying that a sufficient number of clones were assayed for activity.

Mutant proteins were purified and assayed for activity using a steady-state single nucleotide incorporation assay. The kinetic data revealed a remarkable increase in OMeNTP incorporation efficiency that was most pronounced for clone SFM19 [1]. The  $k_{cat}/K_M$  for correct OMeNTP insertion by SFM19 ranged from  $1 \times 10^6$  to  $6 \times 10^6$   $M^{-1}s^{-1}$ , even in the case of modified UTP insertion opposite dA, which has proven especially difficult for DNA polymerase mutants to catalyze due to strong discrimination against this base [72]. In contrast to SFM19, the wild-type enzyme showed no detectable activity with any of the modified triphosphates ( $k_{cat}/K_M < 10^3$   $M^{-1}s^{-1}$ ). Thus, the evolved activity of SFM19 corresponds to an increase in  $k_{cat}/K_M$  of at least 10,000-fold. In addition, the fidelity of SFM19 was virtually unchanged relative to Sf.

Application 3: Evolving a DNA polymerase that synthesizes DNA with an unnatural base pair.

My research has a long standing interest in developing unnatural base pairs with which to expand the genetic alphabet. As part of these efforts, the five focused libraries described in C.1.1.3 were employed to evolve an Sf mutant that synthesizes DNA containing an unnatural PICS self-pair (for details of this self-pair, see reference [86]). Primer/template S2 (Scheme 1) was attached to the phage. In this case, phage biotinylation first required synthesis and extension past the self-pair. The phage displayed polymerase library was subjected to four rounds of selection with dCTP (50  $\mu M$ ) and biotin-dUTP (2  $\mu M$ ), at 50 °C, for 15 min. Protein from selected phage was prepared and screened for extension of the PICS self-pair by insertion of dCTP opposite dG in the template. From 20 randomly selected clones, three were found to express protein; one of these, P2, was found to insert dCTP with significantly increased efficiency. Mutant P2 had three mutations relative to Sf; F598I, I614F, and Q489H. The unnatural activity of P2 was quantified with both steady state and pre-steady-state kinetics. Relative to the parental enzyme, P2-mediated self-pair synthesis (i.e. insertion of dPICSTP opposite dPICS in the

template) was increased by >320-fold, and synthesis of the next correct base pair (by insertion of dCTP opposite dG) was increased by >30-fold.

### **C.3 Improvements in Library Design.**

The libraries used in the three selections described above were focused to specific regions of the polymerases based on their structures. This strategy relies on the potential of a small number of mutations to confer the desired activity, but limits sequence diversity due to potential folding problems. In addition, sequence analysis suggests that enzymes evolve in nature via the accumulation of many small changes throughout the enzyme. In order to more closely mimic evolution in nature, we have pursued library generation using two related gene shuffling techniques.

(C.3.1.1) *Increased diversity through family shuffling.* Using genes that encode proteins with significant structural homology, family shuffling produces chimeric genes that may explore higher sequence diversity. We have family shuffled the Sf gene with the pol I gene from *Thermus filiformis* (81% identical at both the nucleotide and protein level), *Thermus caldophilus* (85% and 86% identical at the nucleotide and protein levels, respectively), *Thermus thermophilus* (86% and 91% identical), and *Thermus flavus* (85% and 90% identical).

Each gene (truncated to the same length as Sf, 1.6 Kb) was cloned into the phagemid pFAB5.His. PCR products were digested separately with DNaseI, then combined in equal ratios in a thermocycling reassembly step. The full-length chimeric genes were then extracted from the mixture by PCR amplification. Sequencing demonstrated that the library contained virtually 100% chimeric genes. These 'family shuffled' libraries will be used in the selection experiments described below.

(C.3.1.2) *Increased diversity through family shuffling with synthetic oligonucleotides.* Sequence diversity from homologues was also accessed through gene assembly methods, in which chimeric genes were constructed from short synthetic oligonucleotides containing mutagenic base degeneracies. The positions of the mutations as well as the identity of the encoded amino acids were determined from a multiple sequence alignment of Sf with homologues from *Thermus thermophilus*, *Thermus caldophilus*, *Thermus filiformis*, *Spirochaeta thermophila*, and *Thermomicrobium rosem.* Only variations occurring within this set of genes were represented within the assembled library. Since each

mutated residue is viable in the context of at least one parental gene, any individual mutation is unlikely to be highly disruptive. Consequently, the library of chimeric genes should exhibit high viability despite the large number of mutations. Furthermore, cumulative effects from the large number of mutations in any given library member should allow complex solutions to the biochemical requirements for unnatural activity. Such solutions are inaccessible to methods based on focused libraries and rational design.

Initial libraries were spread over Sf amino acids 440-600, which largely correspond to the polymerase 'thumb' subdomain. To limit the library to an experimentally tractable size, mutations that were both distal to the active site and predominantly solvent exposed (>25% solvent exposed surface area based on the reported structure) were eliminated as these residues are less likely to make significant contributions to function. These constraints generated a target library with variation at 40 residues, yielding approximately  $1 \times 10^{14}$  unique amino acid sequences. Included among these 40 positions are 11 residues shown to form specific contacts with the DNA backbone [78].

Construction of the library was achieved through optimization of published overlapping extension PCR assembly methods [87]. Briefly, twelve short oligonucleotides spanning both sense and antisense strands of the thumb region were chemically synthesized (Operon Technologies; 0.2  $\mu$ mole scale) with the appropriate degeneracies. Each oligonucleotide shared twenty base pairs of complementarity with its flanking oligonucleotides. PCR reactions were carried out in 50  $\mu$ l volumes containing 2.5 U Taq (NEB), 1x ThermoPol Reaction Buffer containing 2.0 mM total MgSO<sub>4</sub> (NEB), and 100 nM oligonucleotides. Annealing temperatures ranged from 43°C to 63°C and were empirically optimized to achieve maximum overlap extension. To facilitate the reactions, the library was compiled in a piecewise manner. Initially, the first and last halves of the thumb region were assembled through a series of reactions, with each successive reaction introducing a new oligonucleotide and yielding a more complete gene fragment. To ensure coverage of the  $1.2 \times 10^9$  and  $3.0 \times 10^5$  nucleotide sequences in the first half and second halves respectively, yields in excess of 1  $\mu$ g were maintained at each step. The remaining N-terminal and C-terminal portions of the Stoffel gene were PCR amplified from wild type Sf and attached to their respective

halves of the thumb region via overlap extension PCR. Finally, these halves were combined to produce full length Stoffel genes on the microgram scale, covering roughly  $1 \times 10^{12}$  sequences. Standard DNaseI gene shuffling between selection rounds will provide an opportunity to access sequences not present in this initial synthesis (see below). Full length product was gel purified, *SfiI/NotI* digested, and cloned into pFAB5.His phagemid vector, for subsequent phage display and activity based selection.

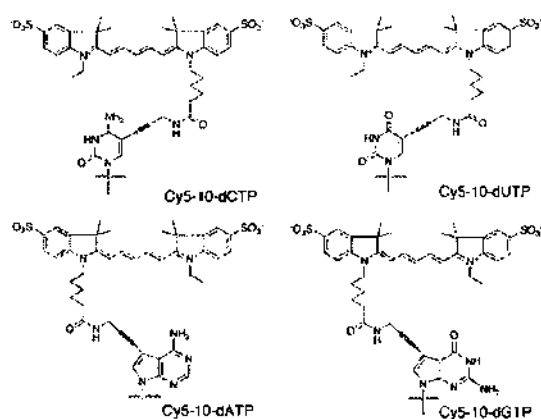


Figure 4. Cy5 labeled nucleotides

## D. Experimental Design & Methods

Both Sf and Kf are attractive for developing SBS methodologies, and we have established activity-based selection systems for each. To date, we have found Sf to be more amenable to directed evolution (see applications in section C.2), and thus, our initial selections will proceed with this polymerase. Kf will be used if unanticipated problems arise with the Sf selections.

### D.1. Specific Aim 1: Evolve Polymerases that Efficiently Synthesize DNA Using Cy5-Modified dNTPs.

Polymerase-mediated incorporation of dNTPs tethered to a fluorophore such as Cy5 (Figure 4), are an essential component of many SBS methodologies. For example, their efficient and contiguous incorporation is required for the Quake methodology described in the Background & Significance section. The first specific aim of this proposal is to evolve polymerases that more

efficiently use these analogs as substrates. This aim requires three things, i) generating polymerase libraries, ii) subjecting the libraries to a pre-selection complementation to ensure the members have retained the ability to synthesize DNA, iii) phage display and selection of desired mutants. Our proposed approach to each of these steps, is now described. To streamline the discussion, reference is made, where possible, to the Preliminary Studies section where the details of the system are discussed more thoroughly.

**(D.1.1.) Generation of Polymerase Diversity.** The five focused libraries and the IEFYNQ library (see C.1.1.3) performed very well for the evolution of polymerases that recognize dNTPs with modified sugars and nucleobases (see applications in section C.2). Thus we will use these libraries in the following selections. We will also use the family shuffled libraries (see C.3) and add a shuffling step after round three of each selection. Shuffling among the selected genes at this step will create chimeric clones that contain *multiple* beneficial mutations. Thus, after the third round of selection, the enriched genes will be isolated from the phagemid with *SfiI*, *NotI* digestion and shuffled as described above. The shuffled genes will then be subcloned into the phagemid. These 'chimeric' libraries will be subjected to two additional rounds of selection, at which point the selected polymerase genes will be isolated and characterized.

**(D.1.2.) Pre-Selection Complementation.** To ensure that the polymerase library is comprised of mutants that retain their ability to replicate natural DNA, we will first use a complementation screen,

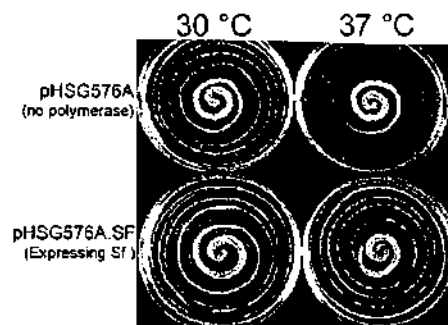
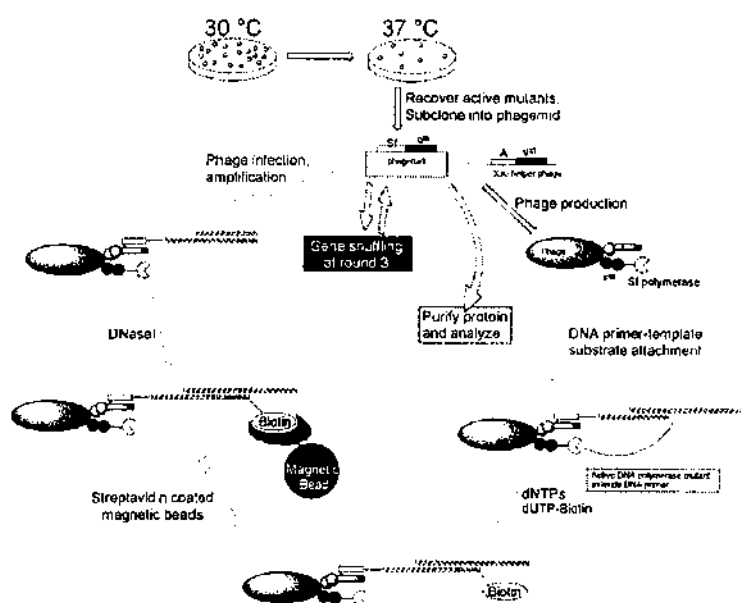


Figure 5. Spiral assay for temperature sensitive complementation. Plasmids pHSG576A and pHSG576A.SF were transformed into *E. coli* JS200. Single colonies were selected, grown overnight at 30 °C, spiral plated in duplicate to LB and grown at 30 °C or 37 °C. Successful complementation is observed in the lower right plate, where wild type Sf has allowed growth in the presence of the temperature sensitive mutation. Used with the Sf libraries, this will allow pre-selecting for only those clones that are active at 37 °C.



**Scheme 2.** General representation of the proposed selections. dNTPs will include any modified variants according to the specific selections. The template DNA will also contain modified nucleotides as described in **Specific Aim 2**.

developed by Loeb and co-workers [88,89], as shown in Figure 5. As described in section B.1.4.2, this complementation is based on transforming a mutant strain of *E. coli* whose DNA polymerase I is active at 30 °C, but not at 37 °C. Initial libraries will be subcloned into pHSG576A (chloramphenicol resistant, carrying the pol I-independent pSC100 origin of replication modified with *SfiI/NotI* restriction sites), and subsequently used to transform JS200. The transformants will be grown on chloramphenicol plates at 30 °C (permissive temperature), then plated again and grown at the non-permissive temperature (37 °C). Only those members of the library that retain their natural activity, and thus permit growth by relieving the conditional lethality of the *polA12 recA718* mutations, will survive at 37 °C. Clones surviving at 37 °C will be collected, and the corresponding polymerase genes will be subcloned into pFAB.SF, phage displayed, and prepared for selection as described in section C.1.

**(D.1.3.) Selection of Polymerases that Efficiently Incorporate Cy5-Labeled dNTPs.** In the first selection experiments, we will evolve Sf mutants that more efficiently incorporate fluorophore Cy5-10-dCTP (Figure 4, available from Perkin Elmer), and then continue primer elongation with natural dNTPs (Scheme 2). Towards this goal, primer/template assembly S3 (Scheme 3) will be

attached to the phage particles, as previously described. This template selects for polymerases that can incorporate Cy5-10-dCTP and then incorporate multiple natural dGTPs. Two rounds of selection will be run in the presence of Cy5-10-dCTP (40 µM) for 2 minutes, followed by a 5 minute incubation with dGTP, and biotin-dUTP (40 µM each). The recovered phagemids will then be used to generate new pol/S3/phage particles (now enriched for the desired activity) and subjected to a third round of selection, where the phage are incubated for 2 minutes with Cy5-10-dCTP (40 µM) as well as ddATP (400 µM) prior to incubation with dGTP and biotin-dUTP (40 µM each). At this step, excess ddATP is included to ensure that fidelity is not compromised in any of the selected mutants (misincorporation of the ddATP will preclude biotin-dUTP incorporation and phage recovery). After the third round of selection, the enriched polymerase library will be gene shuffled and subjected to two additional rounds of selection (one without ddATP and one with) to select for chimeric polymerases with multiple beneficial mutations.

After each round, the progress of the selection experiment will be followed by phage ELISA (to determine the library display level) and by phage titer to measure the input/recovered phage ratio. In other selections, we have generally found that the ratio of input to recovered phage should start to increase by the third round and begin to approach that observed in the positive control (inclusion of all requisite natural dNTPs along with biotin-dUTP). The selection will also be monitored by random clone sequence analysis. We have typically observed

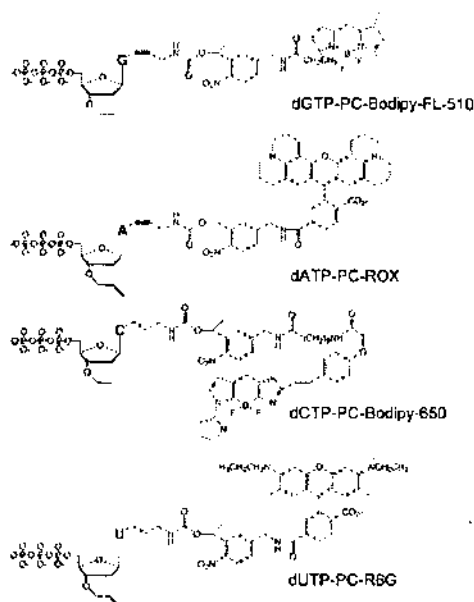
- S3** (N<sub>45</sub>) TAGG  
(N<sub>13</sub>) ATCCGCCATGGAACTCC
- S4** (N<sub>46</sub>) TAGG  
(N<sub>13</sub>) ATCCCTCAGGGAACTCC
- S5** (N<sub>21</sub>) TAGG  
(N<sub>13</sub>) ATCCCCCTAGGGAACTCC
- S6** (N<sub>25</sub>) TAGG  
(N<sub>13</sub>) ATCCGGGGCATGAACTCC
- S7** (N<sub>27</sub>) TAGGX  
(N<sub>13</sub>) ATCCGGGGCATGAACTCC
- S8** (N<sub>28</sub>) TAGGXX  
(N<sub>13</sub>) ATCCGGGGCATGAACTCC
- S9** (N<sub>29</sub>) TAGGXXX  
(N<sub>13</sub>) ATCCGGGGCATGAACTCC

**Scheme 3.** Primer template duplexes for proposed activity-based phage selections. X is doubly deprotected 3'-O-allyl-PC-dye.

sequence convergence by the fourth round in successful selections. We anticipate that phage titer and sequencing data will imply significant enrichment by the fifth round. If this is not the case, two additional rounds of selection will be performed. If enrichment is still not observed, we will gene shuffle the enriched libraries with oligonucleotides that introduce mutations throughout the 'finger' and 'palm' subdomains of Sf. If all of these selections fail, we will pursue the evolution of Kf using Kf pol/phage particles.

The selected genes will be subcloned and expressed as described in section C.2. The ability of the evolved mutants to catalyze insertion of Cy5-10-dCTP under steady state kinetic conditions will be quantitated by determining Michaelis-Menten parameters,  $k_{cat}$  and  $K_M$ , (the group has extensive experience in these kinetic assays, please see [5]). The fold increase in  $k_{cat}/K_M$  (the second order rate constant, or efficiency), as well as the fidelity (relative efficiencies of correct and incorrect insertion) relative to the parental enzyme will be used to assess the success of the selection. If required, we will perform additional selections using any substrate found not to be efficiently incorporated, or using the ddNTP corresponding to any dNTP found to compromise fidelity.

With successful evolution of polymerase mutants that satisfy the relatively simple selection conditions imposed by S3 (*i.e.* addition of one Cy5-10-dCTP followed by addition of natural dNTPs) we will progress the evolved mutants into additional rounds of diversification, including the introduction of mutations throughout the entirety of the protein, and additional rounds of selection. Sequentially, we will employ template S4 (plus Cy5-10-dATP, dGTP, ddCTP, and biotin-dUTP), S5 (plus Cy5-10-dGTP, dATP, ddCTP, and biotin-dUTP), and finally, S6 (plus Cy5-10-dCTP, dGTP, ddATP, and biotin-dUTP). The concentrations of each triphosphate and the incubation times will be analogous to those described above, unless optimization requires greater or less selection pressure. Which substrate is labeled and which are present as a natural dNTP will vary in order to avoid biases. These experiments will sequentially select for polymerases that efficiently incorporate two, three and four consecutive Cy5-modified dNTPs. The evolved polymerases will be characterized using Michaelis-Menten kinetics to determine the efficiencies and fidelities of each step of synthesis. The controlled application of



**Figure 6.** Dyes used in Specific Aim 2. Purines are linked through C8; pyrimidines are linked through C5.

increasing selection pressure, for both activity and fidelity, coupled with gene shuffling, should result in the evolution of DNA polymerases that more efficiently synthesize DNA containing the modified dNTP analogs.

## **D.2. Specific Aim 2: Evolve Polymerases that Efficiently Incorporate 3'-O-allyl-dNTP-PC-Fluorophore and then Continue Synthesis After Double Deprotection.**

As discussed in the **Background & Significance** section, the Ju group has developed interesting dNTP analogs that are doubly modified (Figure 6). Each of the four dNTPs has the same 3'-O-allyl protecting group but is also reversibly tethered to a unique fluorophore. The controlled addition of a single 3'-O-allyl-dUTP-PC-Bodipy-FL-510 to a growing primer strand has already been demonstrated [11]. However, as discussed in section B.2, polymerase recognition currently limits the practical use of these dNTP analogs for sequencing. We thus propose to evolve a polymerase that 1) incorporates 3'-O-allyl-dNTP-PC-Fluorophore analogs into a growing primer strand, and 2) continues primer elongation after double deprotection, with rates and fidelities comparable to natural synthesis.

### **(D.2.1.) Library Construction and Pre-Selection.**

The same libraries, pre-selected for natural activity

and displayed on phage in the same manner as that described for **Specific Aim 1**, will be employed for this aim.

**(D.2.2.) Selection of Polymerases that more Efficiently Incorporate 3'-O-allyl Fluorophore-Labeled Protected dNTPs.**

The first step of this aim is the evolution of a polymerase that efficiently incorporates the doubly modified dNTPs. Toward this aim, we will first select for polymerases that efficiently incorporate 3'-O-allyl-dCTP-PC-Biotin (Figure 7) into DNA. The analog is modified relative to the 3'-O-allyl-dCTP-PC-Bodipy-FL-650 substrate to allow for phage recovery. However, it is only modified at positions that should be far removed from the polymerase active site; it is linked to the nucleobase with the same linker as is the fluorophore. We thus believe that replacing the fluorophore with biotin is inconsequential. Nonetheless, we will kinetically characterize the evolved polymerases using both substrates (see below). If we find the modification to be problematic, we will synthesize analogs that attach biotin to the Bodipy fluorophore via an additional linker. Professor Ju has kindly agreed to provide us with 3'-O-allyl-dCTP-PC-Biotin, which may be synthesized by a simple modification of his 3'-O-allyl-dCTP-PC-Bodipy-FL-650 synthesis. The selection will proceed by attaching primer/template S6 (Scheme 3) to the phage which allows selection for incorporation of the biotinylated dNTP analog (Scheme 2). Two rounds of 2 minute selection will be run in the presence of 3'-O-allyl-dCTP-PC-Biotin (40  $\mu$ M). The recovered phagemids will then be subjected to a third round of selection that also includes ddATP, ddTTP, and ddGTP (400  $\mu$ M each) to select against clones with compromised fidelity. Finally, the selected polymerase genes will be gene shuffled to generate recombinants with multiple, beneficial mutations and then subjected to two additional rounds of selection (one with and one without added ddNTPs). The progress of the selection will be followed as described above. We will evaluate the ability of the evolved mutants to catalyze insertion of 3'-O-allyl-dCTP-PC-Bodipy-FL-650, as well as 3'-O-allyl-dGTP-PC-Bodipy-FL-510, 3'-O-allyl-dATP-PC-ROX, and 3'-O-allyl-dUTP-PC-BR6G (Kindly provided by Prof. Jingyue Ju) [11]. Suitable templates will be synthesized, and  $k_{cat}$  and  $K_M$  determined using standard kinetic methods. We will also characterize synthesis of all possible mispairs. Additional selections for

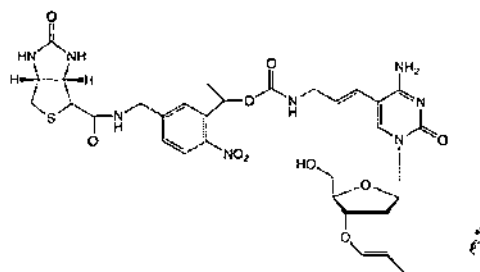
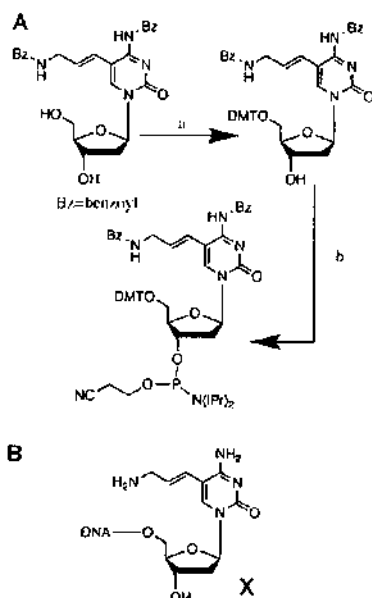


Figure 7. Biotin-PC-dCTP

efficiency or fidelity will be initiated based on these kinetic studies.

The second step of this aim is the evolution of a polymerase that efficiently adds 3'-O-allyl-dCTP-PC-Biotin to a primer terminating with doubly deprotected analogs. These selections are designed to simulate the addition of the modified dNTP after an increasing number of previous rounds of addition and deprotection. These selections require primers with 3' modified nucleotides corresponding to those produced after 3'-O-allyl-dCTP-PC-Bodipy-FL-650 incorporation and deprotection (designated X in Schemes 3 and 4). The phosphoramidite will be synthesized from the corresponding amino propynyl nucleoside as shown in Scheme 4. The required nucleoside is an intermediate in Prof. Ju's published synthesis and he has kindly agreed to provide us with the nucleoside (we have extensive experience with phosphoramidite synthesis, please see [5]). The phosphoramidite will be incorporated into oligonucleotide primers as shown in S7-S9 (Scheme 3), which will be synthesized using an Applied Biosystems Inc. 392 DNA/RNA synthesizer (already available in the lab).

The enriched libraries (from above) will be gene shuffled and displayed on phage and subjected to sequential rounds of selection with S6-S9. First primer/template S7 will be attached and the selection will include 3'-O-allyl-dCTP-PC-Biotin (40  $\mu$ M). This primer/template selects for the ability of the polymerase to add 3'-O-allyl-dCTP-PC-Biotin to a primer terminating with one doubly deprotected nucleotide analog, and is designed to simulate addition of the dNTP analog after one was already incorporated and deprotected. These selections will proceed as described above. Once a polymerase is evolved that efficiently satisfies the selection pressure associated with S7 (again determined by steady-state kinetic analysis), we will diversify the corresponding gene, and progress to selections using S8. This selection is designed to be analogous to the addition of the dNTP analog after two previous



**Scheme 4.** A. Phosphoramidite synthesis. (a) DMTrCl, pyridine, rt; (b) 2-Cyanoethyl diisopropylchlorophosphoramidite, DIEA, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C. B. Compound X, doubly deprotected 3'-O-allyl-PC-dye.

rounds of addition and deprotection. Finally, S9 will be used to select for a polymerase that efficiently adds 3'-O-allyl-dCTP-PC-Biotin to a primer terminating with three deprotected nucleotide analogs. All selections will be characterized by phage titer and random clone sequencing, and the steady-state efficiency and fidelity of all evolved polymerases will be characterized as described above.

A polymerase evolved through the entirety of the above selections (substrates S7-S9) should be capable of efficiently synthesizing DNA through successive rounds of modified dNTP addition and deprotection.

### D.3. Specific Aim 3: Collaborative Evaluation of Evolved Polymerases.

The final aim of this proposal is to collaborate with the groups that are developing the sequencing technologies. To this end, we have established three collaborations. Prof. Steve Quake (Stanford) and Dr. Philip Buzby (Helicos) will evaluate the polymerases with their microfluidic single molecule sequencing methodology. Prof. Jingyue Ju (Columbia) will evaluate the polymerases for sequencing applications using reversible fluorophore/3'-O-allyl protected-dNTPs (as well as supply several of the polymerase substrates used for Specific Aim 2). Finally, we will also evaluate potential improvements in conventional capillary

electrophoretic Sanger sequencing through a collaboration with Dr. Phillip Ordoukhanian (director of the Scripps sequencing facility). Although improvements in capillary electrophoretic Sanger sequencing are not *per se* a specific aim of this proposal, the evolved polymerases should also facilitate these methods as well. We will supply each of these collaborators with purified polymerases to evaluate at their sites. This aim allows for the practical analysis of the polymerases evolved during the course of the proposed research and will be critical to our future efforts.

### E. Human Subjects None

### F. Vertebrate Animals None

### G. Literature Cited

1. M. Fa, A. Radeghieri, A.A. Henry and F.E. Romesberg. (2004) Expanding the substrate repertoire of a DNA polymerase by directed evolution. *J. Am. Chem. Soc.* **126**, 1748-1754.
2. G. Xia, L. Chen, T. Sera, M. Fa, P.G. Schultz and F.E. Romesberg. (2002) Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci. USA* **99**, 6597-6602.
3. I. Braslavsky, B. Herbert, E. Kartalov and S.R. Quake. (2003) Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci. USA* **100**, 3960-3964.
4. J. Shendure, R.D. Mitra, C. Varma and G.M. Church. (2004) Advancing sequencing technologies: methods and goals. *Nature Rev. Genet.* **5**, 335-343.
5. A.A. Henry, A.G. Olsen, S. Matsuda, C. Yu, B.H. Geierstanger and F.E. Romesberg. (2004) Efforts to expand the genetic alphabet: identification of a replicable unnatural DNA self-pair. *J. Am. Chem. Soc.* **126**, 6923-6931.
6. A.A. Henry and F.E. Romesberg. (2003) Beyond A, C, G, and T: augmenting nature's alphabet. *Curr. Opin. Chem. Biol.* **7**, 727-733.
7. E.P. Kartalov and S.R. Quake. (2004) Microfluidic device reads up to four consecutive base pairs in DNA sequencing-by-synthesis. *Nucleic Acids Res.* **32**, 2873-2879.
8. T.S. Seo, X. Bai, D.H. Kim, Q. Meng, S. Shi, H. Ruparel, Z. Li, N.J. Turro and J. Ju. (2005) Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides. *Proc. Natl. Acad. Sci. USA* **102**, 5926-5931.

9. R.D. Mitra, J. Shendure, J. Olejnik, E.K. Olejnik and G.M. Church. (2003) Fluorescent in situ sequencing on polymerase colonies. *Anal. Biochem.* **320**, 55-65.
10. M.L. Metzker, R. Raghavachari, S. Richards, S.E. Jacutin, A. Civitello, K. Burgess and R.A. Gibbs. (1994) Termination of DNA synthesis by novel 3'-modified-deoxyribonucleoside 5'-triphosphates. *Nucleic Acids Res.* **22**, 4259-4267.
11. H. Ruparel, L. Bi, Z. Li, X. Bai, D.H. Kim, N.J. Turro and J. Ju. (2005) Design and synthesis of a 3'-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proc. Natl. Acad. Sci. USA* **102**, 5932-5937.
12. F. Sanger, S. Nicklen and A.R. Coulson. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
13. A. Kornberg and T.A. Baker. (1992) *DNA Replication*. Second ed. W. H. Freeman and Company, New York.
14. D.K. Braithwaite and J. Ito. (1993) Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **21**, 787-802.
15. M. Suzuki, D. Baskin, L. Hood and L.A. Loeb. (1996) Random mutagenesis of *Thermus aquaticus* DNA polymerase I: Concordance of immutable sites *in vivo* with the crystal structure. *Proc. Natl. Acad. Sci. USA* **93**, 9670-9675.
16. M. Astatke, N. K., N.D. Grindley and C.M. Joyce. (1998) A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl. Acad. Sci. USA* **95**, 3402-3407.
17. C.M. Joyce. (1997) Choosing the right sugar: how polymerases select a nucleotide substrate. *Proc. Natl. Acad. Sci. USA* **94**, 1619-1622.
18. J. Wang, A.K.M.A. Sattar, C.C. Wang, J.D. Karam, W.H. Konigsberg and T.A. Steitz. (1997) Crystal structure of a pol a family replication DNA polymerase from bacteriophage RB69. *Cell* **89**, 1087-1099.
19. Y. Li and G. Waksman. (2001) Crystal structures of a ddATP-, ddTTP-, ddCTP, and ddGTP-trapped ternary complex of KlenTaq1: insights into nucleotide incorporation and selectivity. *Prot. Science* **10**, 1225-1233.
20. V.N. Pandey, N.A. Kaushik, D.S. Pradhan and M.J. Modak. (1990) Template primer-dependent binding of 5'-fluorosulfonyl-benzoyldeoxyadenosine by *Escherichia coli* DNA polymerase I. Identification of arginine 682 as the binding site and its implication in catalysis. *J. Biol. Chem.* **265**, 3679-3684.
21. B.M.I. Moore, K. Li and M.B. Doughty. (1996) Deoxyadenosine-based DNA polymerase photoprobes: Design, synthesis, and characterization as inhibitors of the *Escherichia coli* DNA polymerase I Klenow fragment. *Biochem.* **35**, 11634-11641.
22. M. Astatke, N.D.F. Grindley and C.M. Joyce. (1995) Deoxynucleoside triphosphate and pyrophosphate binding sites in the catalytically competent ternary complex for the polymerase reaction catalyzed by DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **270**, 1945-1954.
23. N. Kaushik, V.N. Pandey and M.J. Modak. (1996) Significance of the O-helix residues of *Escherichia coli* DNA polymerase I in DNA synthesis: Dynamics of the dNTP binding pocket. *Biochem.* **35**, 7256-7266.
24. A. Basu and M.J. Modak. (1987) Identification and amino acid sequence of the deoxynucleoside triphosphate binding site in *Escherichia coli* DNA polymerase I. *Biochem.* **26**, 1704-1709.
25. V.N. Pandey, N. Kaushik and M.J. Modak. (1994) Role of lysine 758 of *Escherichia coli* DNA polymerase I as assessed by site-directed mutagenesis. *J. Biol. Chem.* **269**, 13259-13265.
26. L.J. Ferrin and A.S. Mildvan. (1985) Nuclear Overhauser effect studies of the conformations and binding site environments of deoxynucleoside triphosphate substrates bound to DNA polymerase I and its large fragment. *Biochem.* **24**, 6904-6913.
27. S. Tabor and C.C. Richardson. (1995) A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxynucleotides. *Proc. Natl. Acad. Sci. USA* **92**, 6339-6343.
28. S.S. Carroll, M. Cowart and S.J. Benkovic. (1991) A mutant of DNA polymerase I (Klenow fragment) with reduced fidelity. *Biochem.* **30**, 804-813.
29. J.B. Bell, K.A. Eckert, C.M. Joyce and T.A. Kunkel. (1997) Base miscoding and strand misalignment errors by mutator Klenow polymerases with amino acid substitutions at tyrosine 766 in the O helix of the fingers subdomain. *J. Biol. Chem.* **272**, 7345-7351.
30. A.H. Polesky, T.A. Steitz, N.D. Grindley and C.M. Joyce. (1990) Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli*. *J. Biol. Chem.* **265**, 14579-14591.
31. V.N. Pandey, K.R. Williams, K.L. Stone and M.J. Modak. (1987) Photoaffinity labeling of the thymidine triphosphate binding domain in *Escherichia coli* DNA polymerase I: identification of histidine-881 as the site of cross-linking. *Biochem.* **26**, 7744-7748.
32. A. Basu, K.R. Williams and M.J. Modak. (1987) Ferrate oxidation of *Escherichia coli* DNA polymerase I. Identification of a methionine residue

- that is essential for DNA binding. *J. Biol. Chem.* **262**, 9601-9607.
33. S. Basu, A. Basu and M.J. Modak. (1988) Pyridoxal 5'-phosphate mediated inactivation of *Escherichia coli* DNA polymerase I: identification of lysine-635 as an essential residue for the processive mode of DNA synthesis. *Biochem.* **27**, 6710-6716.
  34. J.R. Kiefer, C. Mao, J.C. Braman and L.S. Beese. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 304-307.
  35. V.N. Pandey, N. Kaushik, R.P. Sanzgiri, M.S. Patil, M.J. Modak and S. Barik. (1993) Site directed mutagenesis of DNA polymerase (Klenow) from *Escherichia coli*. The significance of Arg682 in catalysis. *Eur. J. Biochem.* **214**, 59-65.
  36. B.M.I. Moore, R.K. Jalluri and M.B. Doughty. (1996) DNA polymerase photoprobe 2-[(4-azidophenacyl)thio]-2'-deoxyadenosine 5'-triphosphate labels an *Escherichia coli* DNA polymerase I Klenow fragment substrate binding site. *Biochem.* **35**, 11642-11651.
  37. J. Rush and W.H. Konigsberg. (1990) Photoaffinity labeling of the Klenow fragment with 8-azido-dATP. *J. Biol. Chem.* **265**, 4821-4827.
  38. V.N. Pandey, N. Kaushik and M.J. Modak. (1994) Photoaffinity labeling of DNA template-primer binding site in *Escherichia coli* DNA polymerase I. *J. Biol. Chem.* **269**, 21828-21834.
  39. T. Ono, M. Scalf and L.M. Smith. (1997) 2'-Fluoro modified nucleic acids: polymerase-directed synthesis, properties and stability to analysis by matrix-assisted laser desorption/ionization mass spectrometry. *Nucleic Acids Res.* **25**, 4581-4588.
  40. J.A. Latham, R. Johnson and J.J. Toole. (1994) The application of a modified nucleotide in aptamer selection: novel thrombin aptamers containing 5-(1-pentynyl)-2'-deoxyuridine. *Nucleic Acids Res.* **22**, 2817-2822.
  41. T.R. Battersby, D.N. Ang, P. Burgstaller, S.C. Jurczyk, M.T. Bowser, D.D. Buchanan, R.T. Kennedy and S.A. Benner. (1999) Quantitative analysis of receptors for adenosine nucleotides obtained via in vitro selection from a library incorporating a cationic nucleotide analog. *J. Am. Chem. Soc.* **121**, 9781-9789.
  42. K. Sakthivel and C.F. Barbas, III. (1998) Expanding the potential of DNA for binding and catalysis: highly functionalized dUTP derivatives that are substrates for thermostable DNA polymerase. *Angew. Chem. Int. Ed.* **37**, 2872-2875.
  43. S.E. Lee, A. Sidorov, T. Goullain, N. Mignet, S.J. Thorpe, J.A. Brazier, M.J. Dickman, D.P. Hornby, J.A. Grasby and D.M. Williams. (2001) Enhancing the catalytic repertoire of nucleic acids: a systematic study of linker length and rigidity. *Nucleic Acids Res.* **29**, 1565-1573.
  44. D.M. Perrin, T. Garestier and C. Hélène. (1999) Expanding the catalytic repertoire of nucleic acid catalysts: simultaneous incorporation of two modified deoxyribonucleoside triphosphates bearing ammonium and imidazolyl functionalities. *Nucleosides Nucleotides* **18**, 377-391.
  45. D.M. Perrin, T. Garestier and C. Hélène. (2001) Bridging the gap between proteins and nucleic acids: a metal-independent RNaseA mimic with two protein-like functionalities. *J. Am. Chem. Soc.* **123**, 1556-1563.
  46. H.A. Held and S.A. Benner. (2002) Challenging artificial genetic systems: thymidine analogs with 5-position sulfur functionality. *Nucleic Acids Res.* **30**, 3857-3869.
  47. T. Goullain, A. Sidorov, N. Mignet, S.J. Thorpe, S.E. Lee, J.A. Grasby and D.M. Williams. (2001) Enhancing the catalytic repertoire of nucleic acids. II. Simultaneous incorporation of amino and imidazolyl functionalities by two modified triphosphates during PCR. *Nucleic Acids Res.* **29**, 1898-1905.
  48. G. Gao, M. Orlova, M.M. Geogiadis, W.A. Hendrickson and S.P. Goff. (1997) Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc. Natl. Acad. Sci. USA* **94**, 407-411.
  49. D.A. Kostyuk, S.M. Dragan, D.L. Lyakhov, V.O. Rechinsky, V.L. Tunitskaya, B.K. Chernov and S.N. Kochetkov. (1995) Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. *FEBS Lett.* **369**, 165-168.
  50. E.E. Rusakova, V.L. Tunitskaya, L.V. Memelova, S.V. Kochetkov, D.A. Kostyuk and S.N. Kochetkov. (1998) Mutant T7 RNA polymerase is capable of catalyzing DNA primer extension reaction. *FEBS Lett.* **423**, 189-192.
  51. S.O. Gudima, D.A. Kostyuk, O.I. Grishchenko, V.L. Tunitskaya, L.V. Memelova and S.N. Kochetkov. (1998) Synthesis of mixed ribo/deoxyribopolynucleotides by mutant T7 RNA polymerase. *FEBS Lett.* **439**, 302-306.
  52. R. Sousa and R. Padilla. (1995) A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J.* **14**, 4609-4621.
  53. R. Padilla and R. Sousa. (2002) A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Res.* **30**, e138.
  54. M. Zaccolo and E. Gherardi. (1999) The effect of high-frequency random mutagenesis on in vitro

- evolution: a study of TEM-1 beta-lactamase. *J. Mol. Biol.* **285**, 775-783.
55. J.P. Vartanian, M. Henry and S. Wain-Hobson. (1996) Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Res.* **24**, 2627-2631.
  56. R.C. Cadwell and G.F. Joyce. (1994) Mutagenic PCR. *PCR Methods Appl.* **3**, S136-S140.
  57. A.L. Kurtzman, S. Govindarajan, K. Vahle, J.T. Jones, V. Heinrichs and P.A. Patten. (2001) Advances in directed protein evolution by recursive genetic recombination: applications to therapeutic proteins. *Curr. Opin. Biotechnol.* **12**, 361-370.
  58. L. Giver and F.H. Arnold. (1998) Combinatorial protein design by in vitro recombination. *Curr. Opin. Chem. Biol.* **2**, 335-338.
  59. H. Zhao, L. Giver, Z. Shao, J.A. Affholter and F.H. Arnold. (1998) Molecular evolution by staggered extension process (STEPO) in vitro recombination. *Nat. Biotechnol.* **16**, 258-261.
  60. Z. Shao, H. Zhao, L. Giver and F.H. Arnold. (1998) Random-priming in vitro recombination: an effective tool for directed evolution. *Nucleic Acids Res.* **26**, 681-683.
  61. W.P.C. Stemmer. (1999) Directed evolution of enzymes and pathways by DNA shuffling. *Faseb Journal* **13**, A1431-A1431.
  62. A. Cramer, S.A. Raillard, E. Bermudez and W.P.C. Stemmer. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288-291.
  63. L.A. Castle, D.L. Siehl, R. Gorton, P.A. Patten, Y.H. Chen, S. Bertain, H.J. Cho, N. Duck, J. Wong, D.L. Liu *et al.* (2004) Discovery and directed evolution of a glyphosate tolerance gene. *Science* **304**, 1151-1154.
  64. E. Glick, J.P. Anderson and L.A. Loeb. (2002) In vitro production and screening of DNA polymerase eta mutants for catalytic diversity. *Biotechniques* **33**, 1136-+.
  65. J.B. Sweasy and L.A. Loeb. (1993) Detection and characterization of mammalian DNA polymerase beta mutants by functional complementation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **90**, 4626-4630.
  66. A. Shinkai and L.A. Loeb. (2001) In vivo mutagenesis by *Escherichia coli* DNA polymerase I. Ile709 in motif A functions in base selection. *J. Biol. Chem.* **276**, 46759-46764.
  67. D.S. Tawfik and A.D. Griffiths. (1998) Man-made cell-like compartments for molecular evolution. *Nature Biotechnology* **16**, 652-656.
  68. F.J. Ghadessy, J.L. Ong and P. Holliger. (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4552-4557.
  69. M. Russel, H. Lowman and T. Clackson. (2004) In T. Clackson and H. Lowman (eds.), *A Practical Approach*. Oxford University Press, New York, Vol. 266, pp. 1-26.
  70. J.L. Jestin, P. Kristensen and W. G. (1999) A method for the selection of catalytic activity using phage display and proximity coupling. *Angew. Chem. Intl. Ed.* **38**, 1124-1127.
  71. E. Brunet, C. Chauvin, V. Choumet and J.-L. Jestin. (2002) A novel strategy for the functional cloning of enzymes using filamentous phage display: the case of nucleotidyl transferases. *Nucleic Acids Res* **30**, 40-43.
  72. P. Patel and L. Loeb. (2000) Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J Biol Chem* **275**, 40266-40272.
  73. M. Ogawa, A. Tosaka, Y. Ito, S. Yoshida and M. Suzuki. (2001) Enhanced ribonucleotide incorporation by an O-helix mutant of *Thermus aquaticus* DNA polymerase I. *Mutation Res* **485**, 197-207.
  74. P. Patel, H. Kawate, E. Adman, M. Ashbach and L. Loeb. (2001) A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. *J Biol Chem* **276**, 5044-5051.
  75. F. Ghadessy, N. Ramsay, F. Boudsocq, D. Loakes, A. Brown, S. Iwai, A. Vaisman, R. Woodgate and P. Holliger. (2004) Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat Biotechnol* **22**, 755-759.
  76. B.K. Kay, J. Winter and J. McCafferty (eds.) (1996) *Phage Display of Peptides and Proteins: a laboratory manual*. Academic Press, San Diego.
  77. P.S. Freemont, J.M. Friedman, L.S. Beese, M.R. Sanderson and T.A. Steitz. (1988) Cocystal structure of an editing complex of Klenow fragment with DNA. *Proc. Natl. Acad. Sci. USA* **85**, 8924-8928.
  78. Y. Li, S. Korolev and G. Waksman. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7525.
  79. S. Verma and F. Ekstein. (1998) Modified oligonucleotides: synthesis and strategy for users. *Annu. Rev. Biochem.* **67**, 99-134.
  80. S.M. Freier and K.-H. Altmann. (1997) The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res* **25**, 4429-4443.
  81. B.P. Monia, L.E. A., C. Gonzalez, W.F. Lima, D. McGee, C.J. Guinasso, P. Kawasaki, D. Cook and S.M. Freier. (1993) Evaluation of 2'-modified

- oligonucleotides containing 2'-deoxy gaps as antisense inhibitors of gene-expression. *J Biol Chem* **268**, 14514-14522.
82. A.M. Kawasaki, M.D. Casper, S.M. Freier, E.A. Lesnik, M.C. Zounes, L.L. Cummins, C. Gonzalez and P.D. Cook. (1993) Antisense compounds with high-affinity and specificity for RNA targets. *J. Med. Chem.* **36**, 831-841.
  83. E.A. Lesnik, C.J. Guinosso, A.M. Kawasaki, H. Sasnor, M. Zounes, L.L. Cummins, D.J. Ecker, P.D. Cook and S.M. Freier. (1993) Oligodeoxynucleotides containing 2'-O-modified adenosine - synthesis and effects on stability of DNA-RNA duplexes. *Biochemistry* **32**, 7832-7838.
  84. R.H. Griffey, E. Lesnik, S. Freier, Y.S. Sanghvi, K. Teng, A. Kwasaki, C. Guinosso, P. Wheeler, V. Homan and P.D. Cook. (1994) In Y. S. Sanghvi and P. D. Cook (eds.), *Carbohydrate Modifications in Antisense Research*. ACS, pp. 212-224.
  85. V. Tereshko, S. Portmann, E.C. Tay, P. Martin, F. Natt, K.-H. Altmann and M. Egli. (1998) Correlating structure and stability of DNA duplexes with incorporation 2'-O-modified RNA analogues. *Biochemistry* **37**, 10626-10634.
  86. D.L. McMinn, A.K. Ogawa, Y. Wu, J. Liu, P.G. Schultz and F.E. Romesberg. (1999) Efforts toward expansion of the genetic alphabet: recognition of a highly stable, self-pairing hydrophobic base. *J. Am. Chem. Soc.* **121**, 11585-11586.
  87. J. Minshull, S. Govindarajan, T. Cox, J.E. Ness and C. Gustafsson. (2004) Engineered protein function by selective amino acid diversification. *Methods* **32**, 416-427.
  88. M. Camps and L.A. Loeb. (2003) Use of Pol I-deficient *E. coli* for functional complementation of DNA polymerase. *Methods Mol. Biol.* **230**, 11-18.
  89. J.B. Sweasy and L.A. Loeb. (1992) Mammalian DNA polymerase  $\beta$  can substitute for DNA polymerase I during DNA replication in *Escherichia coli*. *J. Biol. Chem.* **267**, 1407-1410.
  90. M. B. Welch, C. I. Martinez, A. J. Zhang, S. Jin, R. Gibbs and K. Burgess. (1999) Syntheses of nucleobases designed for combinatorial DNA sequencing. *Chem. Eur. J.* **5**, 951-960

## H. Consortium/Contractual Arrangements

None

## I. Consultants

None



Bioengineering Department  
James H. Clark Center S170  
Stanford University  
Stanford, CA 94305-5444

May 31, 2005

Professor Floyd Romesberg  
Department of Chemistry  
The Scripps Research Institute  
10550 North Torrey Pines Rd.  
La Jolla, CA 92037

Dear Floyd,

I am delighted to collaborate with your group in its effort to evolve DNA polymerases. In conjunction with our collaboration with Helicos, a biotech company I co-founded to commercialize the sequencing by synthesis methodology developed in my group, we are well equipped to evaluate the polymerases. This collaboration has great potential to significantly contribute to our efforts to develop the sequencing methodology and I look forward to the interactions between our groups.

Best Regards,

A handwritten signature in black ink that reads "Stephen Quake". The signature is fluid and cursive, with the first name "Stephen" and last name "Quake" clearly legible.

Stephen Quake  
Professor of Bioengineering  
Stanford University



May 31, 2005

Professor Floyd Romesberg  
Department of Chemistry  
The Scripps Research Institute  
10550 North Torrey Pines Rd.  
La Jolla, CA 92037

Dear Floyd,

It would be our pleasure to collaborate with you on the development of evolved DNA polymerases. Over the last year we have developed a three level test screen for evaluating polymerase and nucleotide analog performance using a series of model primer-templates: i) solution phase, ii) support bound, and iii) final evaluation on our single molecule sequencing apparatus. Almost as important as mutating the polymerases are developing and optimizing reaction buffer formulations to enhance polymerase activity towards the analogs of interest. I am sure such collaboration will help advance the development of single-molecule sequencing by synthesis. We at Helicos BioSciences Corporation look forward to working with you.

Sincerely yours,

A handwritten signature in black ink, appearing to read "P. Buzby", with a long, sweeping horizontal line extending to the right.

Philip R. Buzby, PhD  
Director, Nucleic Acid Chemistry  
Helicos BioSciences Corporation

**COLUMBIA UNIVERSITY  
COLLEGE OF PHYSICIANS & SURGEONS  
COLUMBIA GENOME CENTER**

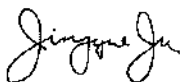
May 31, 2005

Professor Floyd Romesberg  
Department of Chemistry  
The Scripps Research Institute  
10550 North Torrey Pines Rd.  
La Jolla, CA 92037

Dear Prof. Romesberg,

This letter is written to express my interest in collaborating with your group on your efforts to evolve DNA polymerases that more efficiently recognize the modified deoxynucleotide triphosphates that we have been developing for sequencing-by-synthesis applications. The triphosphate substrates are modified with attached fluorophores at their purines/pyrimidines for spectroscopic identification as well as with 3'-O-allyl groups to control their addition. These triphosphates have the potential to revolutionize sequencing efforts coupled with further development of DNA polymerase for efficient recognition. Thus, the evolution of a polymerase that synthesizes DNA with the modified triphosphates would be of great utility. We will supply the modified triphosphates to use in your evolution experiments and then characterize the evolved polymerases in our sequencing methodologies. I look forward to a fruitful collaboration.

Sincerely,



Jingyue Ju, Ph.D.  
Associate Professor & Head of DNA Sequencing and Chemical Biology  
Columbia Genome Center and Department of Chemical Engineering  
Columbia University College of Physicians & Surgeons of  
Room 405A, Russ Berrie Medical Science Pavilion  
1150 St. Nicholas Avenue  
New York, NY 10032  
Phone: (212) 851-5172; Fax: (212) 851-5176  
Email: dj222@columbia.edu

Tuesday, May 31, 2005

Professor Floyd Romesberg  
Department of Chemistry  
The Scripps Research Institute  
10550 North Torrey Pines Rd.  
La Jolla, CA 92037

Dear Prof. Romesberg,

Thank you for your interest in collaborating with Center for Nucleic Acid Research, The Scripps Research Institute. The facility is well equipped and highly experienced with the standard methodologies employed for high-throughput fluorescently labeled DNA sequencing. We would be pleased to assist you with your proposed exploratory/development grant titled, "Evolving Novel DNA polymerases for Genome Sequencing". Specifically, we will provide any assistance or advice related to assaying or optimizing activity of any evolved DNA-directed DNA polymerases found during the course of the study.

As to the Specific Aims in the grant, I can say that having a DNA polymerase with a greater ability to incorporate modified dNTPs would greatly improve the current fluorescently labeled Sanger sequencing method and lower costs significantly. The cost reduction would be achieved in the following areas; by requiring less fluorescently labeled dNTPs in each reaction to still achieve the same signal intensities, by lowering costs associated with excess dye removal post-cycle sequencing reaction, and by increasing the signal-to-noise ratio, resulting in lowering the amount template and primer necessary. For these reasons, a DNA polymerase that had these capabilities would significantly improve and lower the costs even further for the widely used Sanger sequencing method.

Sincerely,



Phillip Ordoukhanian, Ph. D.  
Director, Center for Nucleic Acids Research  
The Scripps Research Institute  
10550 N. Torrey Pines Rd,  
La Jolla, California 92037

## CHECKLIST

## TYPE OF APPLICATION (Check all that apply.)

☒ NEW application. (This application is being submitted to the PHS for the first time.)☐ REVISION of application number:

(This application replaces a prior unfunded version of a new, competing continuation, or supplemental application.)

☐ COMPETING CONTINUATION of grant number:

(This application is to extend a funded grant beyond its current project period.)

☐ SUPPLEMENT to grant number:

(This application is for additional funds to supplement a currently funded grant.)

☐ CHANGE of principal investigator/program director.

Name of former principal investigator/program director:

☐ CHANGE of Grantee Institution. Name of former institution:☐ FOREIGN application ☐ Domestic Grant with foreign involvementList Country(ies)  
Involved:☐ SBIR Phase I ☐ SBIR Phase II: SBIR Phase I Grant No.☐ SBIR Fast Track☐ STTR Phase I ☐ STTR Phase II: STTR Phase I Grant No.☐ STTR Fast Track

## 1. PROGRAM INCOME (See instructions.)

All applications must indicate whether program income is anticipated during the period(s) for which grant support is request. If program income is anticipated, use the format below to reflect the amount and source(s).

Budget Period	Anticipated Amount	Source(s)

## 2. ASSURANCES/CERTIFICATIONS (See instructions.)

In signing the application Face Page, the authorized organizational representative agrees to comply with the following policies, assurances and/or certifications when applicable. Descriptions of individual assurances/certifications are provided in Part III. If unable to certify compliance, where applicable, provide an explanation and place it after this page.

•Human Subjects; •Research Using Human Embryonic Stem Cells•  
 •Research on Transplantation of Human Fetal Tissue •Women and  
 Minority Inclusion Policy •Inclusion of Children Policy• Vertebrate Animals•

•Debarment and Suspension; •Drug- Free Workplace (applicable to new  
 [Type 1] or revised [Type 1] applications only); •Lobbying; •Non-  
 Delinquency on Federal Debt; •Research Misconduct; •Civil Rights  
 (Form HHS 441 or HHS 690); •Handicapped Individuals (Form HHS 641  
 or HHS 690); •Sex Discrimination (Form HHS 639-A or HHS 690); •Age  
 Discrimination (Form HHS 680 or HHS 690); •Recombinant DNA  
 Research, Including Human Gene Transfer Research; •Financial Conflict  
 of Interest (except Phase I SBIR/STTR); •Smoke Free Workplace;  
 •Prohibited Research; •Select Agents  
 •STTR ONLY: Certification of Research Institution Participation.

## 3. FACILITIES AND ADMINISTRATIVE COSTS (F&amp;A)/ INDIRECT COSTS. See specific instructions.

☒ DHHS Agreement dated: 09/08/04☐ No Facilities And Administrative Costs Requested.☐ DHHS Agreement being negotiated with

Regional Office.

☐ No DHHS Agreement, but rate established with

Date

CALCULATION\* (The entire grant application, including the Checklist, will be reproduced and provided to peer reviewers as confidential information.)

a. Initial budget period:	Amount of base \$	150,000	x Rate applied	85.90 %	= F&A costs	\$	128,850
b. 02 year	Amount of base \$	125,000	x Rate applied	85.90 %	= F&A costs	\$	107,375
c. 03 year	Amount of base \$		x Rate applied		= F&A costs	\$	
d. 04 year	Amount of base \$		x Rate applied		= F&A costs	\$	
e. 05 year	Amount of base \$		x Rate applied		= F&A costs	\$	
TOTAL F&A Costs						\$	236,225

\*Check appropriate box(es):

☐ Salary and wages base☒ Modified total direct cost base☐ Other base (Explain)☐ Off-site, other special rate, or more than one rate involved (Explain)

Explanation (Attach separate sheet, if necessary.):