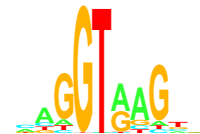# CURRICULUM VITAE
THOMAS D. SCHNEIDER, PH. D.

PERSONAL INFORMATION

Title: Research Scientist, Senior Investigator
Date of Birth: November 9, 1955
Nationality: USA
cell: 240-367-4179
work: 301-846-5581
fax: 301-846-6911
schneidt@mail.nih.gov
toms@alum.mit.edu (permanent)
http://alum.mit.edu/www/toms (permanent) [1]
https://ccr.cancer.gov/RNA-Biology-Laboratory/thomas-d-schneider
Address: National Cancer Institute
       RNA Biology Laboratory
   Frederick National Laboratory for Cancer Research
       Bldg 558, Room 5, P.O. Box B
       Frederick, MD 21702-1201

MAJOR ACCOMPLISHMENTS

- **Invented the sequence logo**,
  a widely used information-theory
  based graphic to display sequence
  alignments [13, 31]

- **Invented the Medusa™ Sequencer**,
  a DNA sequencing machine that is one molecule [80]

- **Discovered 70% molecular efficiency**,
  which defines the exact relationship between information and energy
  in biological systems [78, 79]
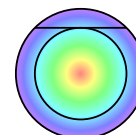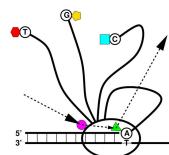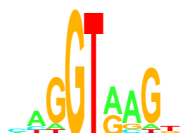
EDUCATION/TRAINING

- 1987 - Postdoctoral Fellowship (Molecular Biology), University of Colorado, Boulder, CO

- 1984 - Ph.D. (Molecular Biology), University of Colorado, Boulder, Colorado

- 1978 - B.S. (Biology), Massachusetts Institute of Technology, Cambridge, Massachusetts

---

[1]Note: web links are active on the PDF.

PROFESSIONAL EXPERIENCE

- 2016 November - present: **Tenured** position as a Senior Investigator, RNA Biology Laboratory, NIH, NCI, CCR

  Expanding the concept of **molecular efficiency** [78, 79] to all of biology.

- 2010 August - 2016 November: **Tenured** position as a Research Biologist, Gene Regulation and Chromosome Biology Laboratory, NIH, NCI, CCR

  Work on **nanotechnology** projects [80, 83, 86], **molecular efficiency** [78, 79, 85], and collaborations [81, 82, 84, 87, 88]

- 1999 June - 2010 August: **Tenured** position as a Research Biologist, Laboratory of Experimental and Computational Biology (renamed Center for Cancer Research Nanobiology Program in 2005 August), NIH, NCI, CCR

  Work to **demonstrate the application of information theory to a wide variety of molecular systems**: DNA replication [22, 33, 54, 55], transcription factors [9, 42, 36, 59, 72, 73, 76], RNA polymerases [9, 12, 65, 68, 71], splice junctions [19], (also mutations in splice junctions causing human disease [27, 37, 38, 49], including the cancer xeroderma pigmentosum [41, 51, 57, 62, 74]), RNA folding [69], ribosome binding sites [9, 10, 14, 18, 24, 53], protein structure [48] and evolution and phylogeny [28, 40, 50, 65, 68].

- December 1987 - June 1999: **Senior Staff Fellow or Investigator** (starting 11/12/95), Laboratory of Mathematical Biology (Division Basic Sciences, National Cancer Institute, National Institutes of Health, Frederick Cancer Research and Development Center, Maryland)

  Continuation of work on T7 described below and creation of a theory of molecular machines, based on Shannon's channel capacity formula [15, 16]. **Invention of the now widely used sequence logos, [13] individual information [34] and sequence walkers [35].**

- January 1987-November 1987: **Postdoctoral Research Associate** with Gary Stormo, Department of Molecular, Cellular and Developmental Biology, University of Colorado

  Conclusion of the T7 promoter experiment [12].

- 1984-1986: **Postdoctoral Research Associate** with Larry Gold, Department of Molecular, Cellular and Developmental Biology, University of Colorado

  Randomized synthetic DNA genetic engineering experiment to determine whether the excess information at bacteriophage T7 promoters is used by the polymerase [12]. The experiment was successful, and showed that the T7 polymerase does not use the excess sequence conservation. This supports the hypothesis that a second protein binds.

  Served on the GenBank Advisory Committee as a graduate student and postdoc.

- 1978-1984: **Ph.D. candidate**, Department of Molecular, Cellular and Developmental Biology, University of Colorado

  Designed and wrote a computer language called Delila for manipulating DNA sequences [2, 6], which we used to investigate the statistics of ribosome binding sites [1, 3]. We used this system to train a perceptron to identify ribosome binding sites, and so we were the first group

to apply neural networks to the problem of finding nucleic acid binding sites [4]. I designed and built plasmids for studying the quantitative relationships between sequence and function of ribosome binding sites [7, 10, 24]. My thesis was on the information content of nucleic-acid binding sites [9]. It describes the excess information (excess sequence conservation) found at bacteriophage T7 promoters. This purely mathematical calculation led me to predict the existence of a second binding protein.

- 1980: **Teaching Assistant**, Molecular Genetics, University of Colorado

  Graded tests and helped students to understand the material.

- 1979: **Teaching Assistant**, Cell and Tissue Biology, University of Colorado

  Taught two lab sections and graded tests.

- 1974-1978: **Undergraduate, Department of Biology, Massachusetts Institute of Technology**, Cambridge, Massachusetts

  Worked at the Architecture Machine in the Department of Architecture on a computer system to recognize visual scenes by parallel processing. I did this on a serial processor since no parallel processors were available at that time. This work was funded for the summer of 1977 by the MIT Undergraduate Research Opportunities Program.

- 1975 and 1976: **Computer Aide**, Space Environment Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO (summer)

  I designed and wrote programs to display meteorological data gathered in real time on a NOAA airplane flying through storm systems such as hurricanes.

HONORS

- 2016: State of Maryland Governor's Citation for contributions to the Werner H. Kirstein Student Internship Program.

- 2012: National Cancer Institute, Certificate of Appreciation for research innovation in FY2012 [83].

- 2010: National Cancer Institute, Center for Cancer Research, Federal Technology Transfer Act award for the Medusa™ Sequencing project [80].

- 2006: Federal Laboratory Consortium For Technology Transfer (FLC) Technology Recognition Award for Nanoprobes for Detection or Modification of Molecules. The award recognizes "technologies not yet commercialized, but having great potential to be commercialized". http://alum.mit.edu/www/toms/patent/nanoprobe/

- 2000: DBS Technology Transfer Award: for the Molecular Rotation Engine patent application: http://alum.mit.edu/www/toms/patent/molecularrotationengine/

- 1998: Federal Technology Transfer Act Cash Award (NCI): for "inventions, innovations, computer software, or other outstanding scientific or technological contributions of value to the United States due to commercial application or due to contributions to the missions of the NCI or the Federal government. Additionally, cash awards are provided for other exemplary activities that promote the domestic transfer of science and technology development."

- 1998: On-the-Spot Award (DHHS, PHS): "for accomplishing the development of a program that has made a dramatic improvement on the work of his group". This was for inventing the atchange program: `http://alum.mit.edu/www/toms/atchange.html`

- 1992: Frederick County Public School System's Highest honor, a Board Pin for being a mentor in the National Cancer Institute Student Intern Program for 4 years.

- 1983: Graduate Student Research and Creative Work Award, University of Colorado.

- 1980-1981: National Institutes of Health Training Grant, NIH 5 T32 GM07385-04, at the University of Colorado

- 1976: Robert A. Boit Writing Prize at MIT (2nd Place)

- 1975: Ellen King Prize for Freshman Writing at MIT (3rd Place)

- 1974: Westinghouse Science Talent Search (top 40)
  `https://student.societyforscience.org/science-talent-search-1974`

INTRAMURAL NIH RESEARCH SUPPORT

- **Z01 BC 008396** Molecular Information Theory. This is my main project since 1987 and its funding covers the other projects.
  `http://alum.mit.edu/www/toms/`

- **Z01 BC 010771** Medusa™ Sequencer.
  `http://alum.mit.edu/www/toms/patent/medusa/`

- **Z01 BC 010772** Rod-Tether Nanoprobe.
  `http://alum.mit.edu/www/toms/patent/nanoprobe/`

MEMBERSHIPS IN PROFESSIONAL ORGANIZATIONS

- Sigma XI, the Scientific Research Society

- American Association for the Advancement of Science (AAAS)

- Institute of Electrical and Electronics Engineers (IEEE): IEEE Information Theory Society and IEEE Communications Society

---

EDITORIAL BOARDS

- 1998 - 2006: member of the Editorial Board of Entropy, an International and Interdisciplinary Journal of Entropy and Information Studies ISSN 1099-4300. The first issue was published in March 1999. `http://www.mdpi.com/journal/entropy`.

- 2009 - December 2017: Editorial Board, Nucleic Acids Research

MEETINGS

- Third workshop on the national nucleic acid sequence data bank, National Institute of General Medical Sciences (NIGMS), Bethesda, Maryland, December 7, 1980

- Reviewer to evaluate the contract proposals for the establishment of a national nucleic acid sequence data bank, Bethesda, Maryland, March 30, 1982

- Advisor to NIGMS on GenBank, the national Genetic Sequence Data Bank. (Cambridge, Massachusetts, December 16, 1982; Los Alamos, New Mexico, January 13, 1984; Bethesda, Maryland, October 31, 1985; Bethesda, Maryland, November 5-6, 1987; Los Alamos, New Mexico, April 11-12, 1988; Bethesda, Maryland, November 14-15, 1988)

- Advisor to the Protein Identification Resource (August 1992)

- Consultant to the Office of Naval Research, Biological Science and Technology Program on initiating projects on molecular computation, December 20, 1994

- Schneider, T. D. Invited by Tadashi Nakano (Assistant Adjunct Professor, Department of Computer Science, Donald Bren School of Information and Computer Sciences, University of California, Irvine, tnakano@ics.uci.edu) and Tatsuya Suda (Program Director, EMT Program, CCF/CISE/NSF, suda@ics.uci.edu) to chair Session 3: Coding Theory and Channel Capacity of Biological Communications Media at the NSF funded workshop on Molecular Communication: Biological Communications Technology, February 19-21, 2008, Arlington, Virginia. I invited 7 other attendees and organized the session. The final report is available from
  `http://netresearch.ics.uci.edu/mc/nsfws08/NSFWS08_Report_080308final.pdf`
  See section 2.3.
  The slides are available at:
  `http://netresearch.ics.uci.edu/mc/nsfws08/slides/GC_finalized.pdf`
  We made these slides (pages 4-6):
  - Understanding Biology using Information and Coding Theory
  - Understanding Information and Coding Theory using Biology
  - Applying Information and Coding Theory to Bioinformatics and Bio/Nanotechnology

- Schneider, T. D. invited by Dr. John A. Niederhuber (Director, National Cancer Institute) and Anna D. Barker, Ph.D. (Deputy Director, National Cancer Institute) to attend and participate in a meeting: Integrating and Leveraging the Physical Sciences to Open a New Frontier in Oncology, sponsored by the National Cancer Institute, February 26-28, 2008, The Ritz-Carlton, Pentagon City Arlington, VA 22202.

- Dr. Niederhuber invited me again to participate in the October 2008 Think Tank "Physical Sciences-Based Frontiers in Oncology: The Coding, Decoding, Transfer, and Translation of Information in Cancer."

- Invited to Co-chair a session and present my work at the NSF Workshop on Biologically-Enabled Wireless Networks Design and Modeling July 19-20, 2011 — Hilton Arlington, Arlington VA `http://madrid.ccs.neu.edu/nsfbw11/`.

PRESENTATIONS

For the most recent and confirmed upcoming events, see . <u>Forthcoming and Previous Presentations</u>

1. Second European Molecular Biology Conference on Computer Analysis of Nucleotide Sequence Information, Schonau, West Germany, May, 1981. "The Delila System".

2. Maximum Entropy and Bayesian Methods in Applied Statistics, 6th Annual Workshop, Seattle, Washington, August 5-8, 1986. "Information Content of Binding Sites on Nucleotide Sequences".

3. Macromolecules, Genes, and Computers, Waterville Valley, N. H., August 12-17, 1986. "Directions for Genetic Sequence Data Bases".

4. December 1988, Johns Hopkins, Baltimore, MD. Host: David Draper.

5. August 17, 1990 Johns Hopkins, Baltimore, MD.

6. Supercomputing Research Center, Bowie, MD. May 24, 1991, "Theory of Molecular Machines".

7. Partnerships in Education Workshop, The Regional Education Service Agency of Appalachian Maryland, Hagerstown, MD. January 31, 1992. "NCI-FCRDC Student Intern Program: Planning for our Scientific Future."

8. CREST Statewide Conference on Science and Engineering Partnerships. University of Maryland, College Park. April 7, 1992. "National Cancer Institute Student Intern Program: Planning for our Scientific Future."

9. Physics of Computation Workshop, Dallas, Texas, October 2-4, 1992, "Use of Information Theory in Molecular Biology".

10. Neils Bohr Institute in Copenhagen, Denmark, April 20, 1993: "Information Theory of Molecular Binding Sites: Bits and Sequence Logos"

11. Department of Physical Chemistry at the Technical University of Denmark, Lyngby, Denmark, April 22, 1993: "Theory of Molecular Machines: Gumballs and Hyperspace". (5 hours of lectures, in combination with previous lecture.)

12. Frederick Community College, May 10, 1993, "Use of Information Theory in Molecular Biology".

13. The Washington Evolutionary Systems Society (WESS) Washington, D.C., November 4, 1993, "Information Theory and Molecular Recognition".

14. Washington-Baltimore Section of the Society for Industrial and Applied Mathematics (SIAM) in conjunction with the American Mathematical Society and the Mathematical Association of America, Washington, D.C., April 26, 1994, "New Approaches in Mathematical Biology: Information Theory and Molecular Machines".

15. Biomolecular Databases: Current Status, June 13-14, 1994, Biophysical Society, Bethesda, MD. "Philosophy and Definition for a Universal Genetic Sequence Database".

16. Third International *E. coli* Genome meeting, Woods Hole, Massachusetts, November 4-8, 1994, "New approaches in mathematical biology: information theory and molecular machines".

17. New England Biolabs, Beverly Mass. February 9, 1995 "New Approaches in Mathematical Biology: Information Theory and Molecular Machines".

18. Ptashne Laboratory, Harvard, Mass. February 10, 1995 "New Approaches in Mathematical Biology: Information Theory and Molecular Machines".

19. Keynote speaker for the Informatics session of the Trieste Conference on Chemical Evolution, IV: Physics of the Origin and Evolution of Life, Cyril Ponnamperuma Memorial. Trieste, Italy, September 4-8, 1995. [30] "New Approaches in Mathematical Biology: Information Theory and Molecular Machines".

20. Department of Genetics of the North Carolina State University. February 26, 1996 "New Approaches in Mathematical Biology: Information Theory and Molecular Machines".

21. Workshop on Gene Networks and Cellular Controls, Hotel duPont, Wilmington, Delaware 18-19 June 1996. Sponsored by the Office of Naval Reasearch. "Information capacity and molecular recognition in gene control".

22. Fourth Workshop on Physics and Computation: PhysComp96, Boston, Mass 22-24 November 1996, Boston University. "New approaches in mathematical biology: information theory and molecular machines"

23. Second Gordon Research Conference on "Modern Developments in Thermodynamics" February 16-21, 1997, Holiday Inn, Ventura, California. "Information theory and molecular recognition".

24. National Library of Medicine, Bethesda, MD. Tue, Sep 30, 1997. "Logos and walkers: graphical analysis of splice junctions and other binding sites, with clinical application"

25. Johns Hopkins Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD. October 1, 1997. "Logos and walkers: graphical analysis of splice junctions and other binding sites, with clinical application"

26. NIH Biotechnology Interest Group, 1997 October 14, "Logos and walkers: graphical analysis of splice junctions and other binding sites, with clinical application"

27. Organizer and speaker at the session Thermodynamics and Information Theory in Biology, 1998 American Association for the Advancement of Science (AAAS) Annual Meeting and Science Innovation Exposition Philadelphia, Pennsylvania. Monday, February 16, 1998, 3:00pm-6:00pm, Track: Emerging Science: Transforming the Next Generation Session number: 101.0. "Information Theory in Molecular Recognition: Efficiency of Molecular Machines"

28. Speaker at the meeting "After the Genome: Envisioning Biology in the Year 2010" conference IV, Jackson Hole, Wyoming, October 10-14, 1998.

29. Speaker in the Computational Sciences & Informatics Colloquium, George Mason University, "Molecular Information Theory: From Clinical Applications to Molecular Machine Efficiency" Dec. 3 , 1998.

30. Invited to give a Science Innovation Topical Lecture at the AAAS Annual Meeting and Science Innovation Exposition (1999 January 23, 1:30pm-2:15pm, Anaheim, CA). "Molecular Information Theory: from Clinical Applications to Molecular Machine Efficiency".

31. Molecular Information Theory, April 27, 1999 at the Transcription Factors Interest Group Conference, Holiday Inn Conference Center, Frederick, MD.

32. 2000 Feb 7: "Molecular Information Theory: from Clinical Applications to Molecular Machine Efficiency" at the Department of Biochemistry, University of Missouri-Columbia, Columbia MO.

33. 2000 May 20: "Evolution of Biological Information" at the Washington Evolutionary Systems Society Annual Symposium on General Evolutionary Systems, Georgetown University, Georgetown, MD.

34. "Molecular Information Theory" at the International Summer School on "DNA and Chromosomes: Physical and Biological Approaches" Institut d'Ètudes Scientifiques de Cargèse, Cargèse, Corsica, France, July 31-August 12, 2000.

35. "Molecular Information Theory: from Clinical Applications to Molecular Machine Efficiency" in a symposium on "Macromolecular Machines" at the Burnham Institute, La Jolla, USA, April 5, 2001.

36. "Flippers, Flappers and Flip-Flops in DNA Binding". George Mason University, School of Computational Sciences Bioinformatics Colloquium 2001 November 6. Host: Iosif Vaisman, http://binf.gmu.edu/vaisman/

37. "Flippers, Flappers and Flip-Flops in DNA Binding". Frederick Faculty Seminar Series, December 12, 2001.

38. "Flippers, Flappers and Flip-Flops in DNA Binding". 2002 January 17. Thursday, North Carolina State University in the Microbiology Department. 10:00-11:00 am, Stephens Room (Gardner 3533). Host: Eric Miller,
http://www.microbiology.ncsu.edu/people/faculty/Miller.html

39. "Flippers, Flappers and Flip-Flops in DNA Binding". 2002 February 7. Department of Mathematics. The Pennsylvania State University State College, Pennsylvania, Mathematics Colloquium. Host: Howard Weiss,
http://www.math.psu.edu/oldColloquium/020207.html

40. "Twenty Years of Delila and Molecular Information Theory". Altenberg-Austin Workshop in Theoretical Biology in Altenberg (Vienna), Austria, July 11-14, 2002. The 8th workshop, on BIOLOGICAL INFORMATION BEYOND METAPHOR: Causality, Explanation, and Unification.

41. Institute for Pure and Applied Mathematics (IPAM) Workshop I: Alternative Computing September 30 - October 3, 2002, UCLA Los Angeles CA. Molecular Information Theory: Molecular Efficiency and Flip-Flops. Wednesday October 2, 2002, 3:30 pm.

42. The Center for Advanced Research in Biotechnology CARB (9600 Gudelsky Drive, Rockville, Maryland 20850, contact: Dr. Harold Smith) by Tom Schneider. Flippers, Flappers and Flip-Flops in DNA Binding. 2002 December 2 Monday, 11:00 am.

43. The University of Richmond, Molecular Information Theory: from Clinical Applications to Molecular Machine Efficiency. 2003 January 27, Monday, 4:30 pm. Hosts: Karen Lewis and Peter Smallwood.

44. The Virginia Commonwealth University by Tom Schneider. Molecular Information Theory: from Clinical Applications to Binding Site Evolution. 2003 January 28, Tuesday, 1:00 pm. Host: Gail Christie

45. CANCUN Mexico 2003 17 - 21 September, 25th ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY Special Session on Communication Theory, Coding Theory and Molecular Biology

46. "Genomes 2004: International Conference on the Analysis of Microbial and Other Genomes" (http://www.tigr.org/conf/mg/) held from April 14-17, 2004, The Wellcome Trust Conference Centre Hinxton, Cambridge, United Kingdom (This conference was originally to be held at the Institut Pasteur, France, April 14 - 17th, 2004 www.pasteur.fr/gmp but it was cancelled. It was then located to the UK.)

47. The Chemical Theatre of Biological Systems" 24th - 28th May, 2004 in Bozen, North Italy. Session: "Application of Information Theory to Chemical and Biological Systems" hosted by the Beilstein-Institut.

48. Wesleyan Biology Department's Seminar Series, Middletown, Connecticut, February 3, 2005. Molecular information theory: From clinical applications to molecular machine efficiency.

49. FinBioNet 2005 Symposium, October 6-7, 2005. Ellivuori, Finland. Molecular information theory: From clinical applications to molecular machine efficiency.

50. The University of Missouri - Kansas City Department of Computer Science and Electrical Engineering (CSEE) Seminar Series by Tom Schneider. Molecular Information Theory: from Clinical Applications to Binding Site Evolution, 2005 November 11.

51. Mathematical Biosciences Institute (MBI) at the Ohio State University, Molecular Information Theory: from Clinical Applications to Binding Site Evolution, 2005 November 14-18. This is part of a workshop on Aspects of Self-Organization in Evolution organized by Chris Adami and Claus O. Wilke.

52. Molecular Information Theory: from Clinical Applications to Molecular Machine Efficiency. Frederick Faculty Seminar Series, February 15, 2006.

53. Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, Baltimore, MD. 2006 Feb 22 4:00 pm, Molecular Information Theory: Flippers, Flappers and Flip-Flops in DNA Binding,

54. School of Electrical and Computer Engineering – Cornell University Ithaca, NY 14853-6701. 2006 March 28 Molecular Information Theory: from Clinical Applications to Molecular Machine Efficiency, Host: Sergio Servetto of the Cornell Communication Networks Research Group.

55. Graduate Group in Computational and Genomic Biology, University of California, Department of Molecular and Cell Biology, Berkeley, CA 94720, Sep 19, 2006. Molecular Information Theory: From Clinical Applications to Molecular Machine Efficiency,

56. The Keck Graduate Institute, Claremont, CA, 91711, September 22, 2006. Molecular Information Theory: From Clinical Applications To Binding Site Evolution.

57. Science Unrestricted (presentation for K-12 Students, Families and Teachers), Institute for Defense Analyses, Alexandria, VA, 22311-1882, April 2, 2008. Evolution in a Nutshell

58. The Institute for Defense Analyses, Alexandria, VA, 22311-1882, May 29, 2008. Molecular Information Theory: From Clinical Applications To Binding Site Evolution.

59. NCI Frederick Faculty Seminar Series. More Molecular Information Theory. Jan 14, 2009.

60. Plenary speaker at the Workshop on Biological and Bio-Inspired Information Theory held in conjunction with the 43rd Annual Conference on Information Sciences and Systems March 18-20, 2009 The Johns Hopkins University, Baltimore, Maryland, USA

61. Science Unrestricted (presentation for K-12 Students, Families and Teachers), Institute for Defense Analyses, Alexandria, VA, 22311-1882, March 31, 2009. Evolution in a Nutshell

62. Jena, Germany June 16-21, 2009, the Jena Life Science Forum 2009: The Molecular Language of Life

63. Information theory and molecular biology, September 24, 2009, National Heart, Lung, and Blood Institute (NHLBI), Rockville, MD. Host: Bernard Brooks.

64. Information theory and molecular biology, November 13, 2009, National Heart, Lung, and Blood Institute (NHLBI), Bethesda, Maryland. Host: Mark Knepper,

65. "Information theory and molecular biology", December 10, 2009, National Library of Medicine, National Center for Biotechnology Information, Bethesda, Maryland. Host: Eugene Koonin.

66. Systems Biology Colloquium, Humboldt-Universität and Charitee University Berlin, Germany. 2010 Feb 12. Information theory and molecular biology, Host: Dr. Hanspeter Herzel of the Institute for Theoretical Biology

67. Plenary talk at the workshop Information Theory meets Biology, Feb 16 and 17 2010, Institute of Telecommunications and Applied Information Theory, Ulm University, Germany. Information theory and molecular biology, Information Theory meets Biology2010 (PDF). Host: Dr. Martin Bossert, Ulm University, Germany

68. Information theory and molecular biology, at the University of Maryland, Baltimore County Biological Sciences. 2010 April 1. BS 004, 4:00 pm, Host: Ivan Erill.

69. Science Unrestricted (presentation for K-12 Students, Families and Teachers), Institute for Defense Analyses, Alexandria, VA, 22311-1882, April 27, 2010. Evolution in a Nutshell

70. SIAM Conference on the Life Sciences (LS10 July 12-15, 2010, Pittsburgh, Pennsylvania, The David L. Lawrence Convention Center, MS69, Minisymposium: Information Theory for Bioinformatics 4:00 PM - 6:00 PM on July 15th, 4 pm. Efficiency of Molecular Machines. Organizers: Sarosh N. Fatakia (NIDDK, NIH) and Carosh Chow (NIDDK, NIH).

71. Perspectives in High Dimensions 2-6 August 2010 at Case Western Reserve University, Cleveland. 70% efficiency of bistate molecular machines explained by information theory, high dimensional geometry and evolutionary convergence, Host: Elizabeth Meckes.

72. "70% efficiency of bistate molecular machines explained by 'information theory, high dimensional geometry and evolutionary 'convergence'", at the Rutgers Department of Electrical & Computer Engineering Colloquium Series February 23, 2011 Host: Dr. Athina Petropulu and Dr. Christopher Rose. Slides from the talk

73. Science Unrestricted (presentation for K-12 Students, Families and Teachers), Institute for Defense Analyses, Alexandria, VA, 22311-1882, May 4, 2011. Evolution in a Nutshell

74. Science Unrestricted (presentation for K-12 Students, Families and Teachers), Institute for Defense Analyses, Alexandria, VA, 22311-1882, April 24, 2012. Evolution in a Nutshell

75. NCI Frederick Faculty Seminar Series.
Why Do Restriction Enzymes Prefer 4 and 6 Base DNA Sequences? Jan 11, 2012.

76. Mathematical and Statistical Models for Genetic Coding September 26th to 28th 2013, Mannheim, Germany. Why is the Genetic Code Degenerate?

77. Bits↔Biology, The Center for Bits and Atoms, MIT, May 1, 2014. Molecular Information Theory: Why is the Genetic Code Degenerate?
http://cba.mit.edu/events/14.05.BB/index.html
videos of the talk

78. Science Unrestricted (presentation for K-12 Students, Families and Teachers), Institute for Defense Analyses, Alexandria, VA, 22311-1882, May 9, 2014. Evolution in a Nutshell

79. NCI Frederick Faculty Seminar Series. Three Universal Principles of Biological States June 11, 2014.

80. Biological and Bio-Inspired Information Theory (14w5170) Three Principles of Biological States: Ecology and Cancer. 2014 Oct 29 Wednesday 09:04-10:13 at the meeting Biological and Bio-Inspired Information Theory (14w5170) at the Banff International Research Station (BIRS), Banff, Canada.
http://www.birs.ca/events/2014/5-day-workshops/14w5170/videos/watch/201410290904-Schneider.mp4

81. Three Principles of Biological States: Ecology and Cancer. 2014 Nov 21 at the , National Institute of Standards and Technology Gaithersburg, MD.

82. sDiv workshop, "sFIND" on "Functional Information: its potential for quantifying biodiversity and its relation to ecosystem functioning", 2015 September 7th to 11th, Leipzig, Germany.

83. Information Theory in Biology. Shannon Centenary `http://home.iitk.ac.in/~adrish/Shannon/`, Department of Electrical Engineering at the Indian Institute of Technology, Kanpur, India, Wednesday, October 19th, 2016 and the Department of Biological Sciences & Bioengineering `http://www.iitk.ac.in/bsbe/` Thursday, October 20th, 2016.

84. "Three Principles of Biological States: Ecology and Cancer" at the University of Missouri-Columbia, Columbia MO Life Sciences Week, in the Monsanto Auditorium at Bond Life Sciences Center. 2017 April 11, 1:15 p.m. Article about the talk:
National Cancer Institute researcher to speak at Life Sciences Week Apr 4, 2017 By Jinghong Chen, Bond Life Sciences Center.

85. Science Unrestricted (also 2015, 2016, 2017, 2018) (presentation for K-12 Students, Families and Teachers), Institute for Defense Analyses, Alexandria, VA, 22311-1882, May 7,

2019. Evolution in a Nutshell (Cancelled in 2020, 2021 and perhaps 2022 because of the pandemic.)

86. Why Do Restriction Enzymes Prefer 4 and 6 Base DNA Sequences? by Tom Schneider 2020 Jan 21 at the meeting: NSF/UMBC/UNL BiotICC WORKSHOP: BIOLOGY THROUGH INFORMATION, COMMUNICATION & CODING THEORY January 21-22, 2020, Alexandria, Virginia

    (a) Video of the talk at Vimeo: Thomas Schneider, NIH/NCI, BiotICC Talk
    (b) Video, copy Thomas Schneider, NIH/NCI, BiotICC Talk
    (c) SLIDES (PDF)
    (d) The paper: Restriction enzymes use a 24 dimensional coding space to recognize 6 base long DNA sequences

87. "Biological Information Theory (BIT) gives a natural binding site cutoff". Zoom virtual presentation 2021 Aug 16 at the National Institutes of Health (NIH), National Cancer Institute (NCI), Center for Cancer Research (CCR), Cancer Data Science Laboratory (CDSL, schedule)

    (a) Abstract and Bio
    (b) Video (at google drive)
    (c) Video (from this website)
    (d) Slides
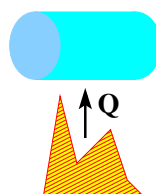
POSTER SESSIONS

(partial list)

- NIH Research Festival '90, Bethesda, MD, September 10, 1990, T. D. Schneider "Sequence Logos: A New Way to Display Consensus Sequences".

- NIH Research Festival '91, P. P. Papp, D. K. Chattoraj and T. D. Schneider Bethesda, MD, September 23, 1991, "Information analysis of a set of DNA sites that bind protein".

- Second Foresight Conference on Molecular Nanotechnology, T. D. Schneider Palo Alto, CA, November 7-9, 1991, "Theory of Molecular Machines".

- NIH Research Festival '92, T. D. Schneider Bethesda, MD, September 21, 1992, "Theory of Molecular Machines".

- ACM Workshop on Information Retrieval and Genomics Bethesda, MD, USA, May 2-4, 1994, "A Philosophy and Definition for Genetic Sequence Databases".

- NCI Division of Basic Sciences' Retreat, December 4-5, 1995, Marriott's Hunt Valley Inn, and Transcription Factor Interest Group, December 13, 1995, NIH Bethesda, "Consensus Sequences: Just Say No!".

- NCI Division of Basic Sciences' Retreat, 1996, 1997

- NIH Research Festival October 6-9, 1998. S. G. Kahn, T. D. Schneider, K. H. Kraemer (NCI) "Splice Mutations in Xeroderma Pigmentosum Group C DNA: Intron Retention and Exon Skipping"

- Comparison of algorithms for computational identification of natural and mutant splice sites. Human Genome Organization, Mutation Database Meeting, Denver 1998 Oct. 27. Peter K. Rogan and Thomas D. Schneider.

- NCI Division of Basic Sciences' Retreat, January 7-8, 1999. "Splice Mutations In Xeroderma Pigmentosum Group C DNA : Intron Retention, Exon Skipping And Internal Neoplasms", Sikandar G. Khan, Thomas D. Schneider, Kenneth H. Kraemer

- NIH Research Festival, October 3, 2001. "Influence of a single nucleotide polymorphism in an XPC splice acceptor site on alternative splicing", S. G. Khan, T. Shahlavi, V. Muniz-Medina, H. Inui, T. Ueda, C. C. Baker, T. D. Schneider and K. H. Kraemer

- Poster: Modeling Splice Site and Transcription Factor Binding Site Variation by Information Theory. P. K. Rogan, S. R. Svojanovsky, I. Hurwitz, T. D. Schneider, J. S. Leeder. American Society of Human Genetics Annual Meeting, 2002.

- NIH Research Festival, September 29, 2004, Lyakhov, Ilya G. (NCI) I.G. Lyakhov, Z. Chen, T. D. Schneider, "A Novel Bacteriophage Lambda Cro Responsive Element in the oop RNA Promoter"

- NIH Research Festival, September 29, 2004, Z. Chen and T. D. Schneider, "Comparative Analysis of T7-like Promoter Bearing Regions in Bacterial Genomes Reveals a Novel Group of Islands"

- NIH Research Festival, Oct. 19, 2005. "Take-over Pressure Explains Excess Information at T7-like Promoters," Z. Chen and T. D. Schneider

- 2012 NIH Research Festival, Oct. 9-12, 2012. "Why Do Restriction Enzymes Prefer 4 and 6 Base DNA Sequences?," T. D. Schneider and Vishnu Jejjala

- FASEB meeting poster: "Mechanism and Regulation of Prokaryotic Transcription" June 25-30, 2017, Saxtons River, VT

- Poster presentation at the online meeting "Mathematical Models in Biology: from Information Theory to Thermodynamics (20w5074)" held by the Banff International Research Station, Vancouver BC V6T 1Z4 Canada, July 28 2020.
  `http://www.birs.ca/events/2020/5-day-workshops/20w5074`
  "Restriction enzymes use a 24 dimensional coding space to recognize 6 base long DNA sequences".

REVIEWS

- Site Visit and Special Study Section to review an application by Dr. Charles Lawrence, Baylor University College of Medicine, Houston, TX. August 7-8, 1988.

- Reviewed papers for:
    - Acta Biotheoritica (Nov 2007, Sep 2009, June 2012)
    - Bioinformatics (July 1999, December 2000, July 2001, September 2001, November 2011)
    - Biopoymers (February 2001)

- BioMed Central (June 2002)
- BMC Bioinformatics (Dec 2008, Nov 2013)
- CABIOS (January 1994, March 1994, March 1995)
- Computational Statistics and Data Analysis (December 2001)
- Developmental Biology (March 2012)
- Gene (February 1992)
- FASEB Journal (January 2001)
- Entropy (March 2011)
- Genomics (April 1996)
- IEEE Transactions on Information Theory (March 2001)
- IEEE IEEE/ACM Transactions on Computational Biology and Bioinformatics (December 2005)
- Information (November 2011)
- Journal of Bacteriology (May 1993, October 1993)
- Journal of biotechnology (September 1988)
- Journal of Biosciences (November 1997)
- Journal of Molecular Biology (February 1990, April 2001)
- Journal of Molecular Graphics and Modelling (March 2011)
- Journal of Molecular Evolution (March 1997)
- Journal of Theoretical Biology (December 1991, May 1994, April 1995, June 1996, October 1997, April 2001, July 2002, October 2002, Feb 2010)
- Machine Learning (December 1993)
- Modern Pathology (July 2014)
- Molecular Carcinogenesis (May 1993)
- Nature Methods (April 2013)
- Nanotechnology (November 1991, April 1995)
- Nucleic Acids Research (December 1988, November 1989, June 1990, November 1990, September 1992, October 1992, January 1993, April 1993, July 1993, December 1994, August 1995, May 1996, March 1997, June 1997, December 1997, August 2000, November 2000, January 2001, July 2002, Aug-2002, Jan-2003, Feb-2003, May-2004, Feb-2006, Jan-2007, Jun-2007, May-2008, Jul-2008, Nov-2008, Jan-2009, Mar-2009, Sep-2009, Sep-2010, Apr-2011, Jan-2013, Mar-2013)
- Physica A (Jan 2013, Dec 2020, April 2021)
- Physical Review Letters (August 1990, Nov 2007)
- Physical Review E
  (Sep 2007, Nov 2007, Dec 2007, Feb 2009, Aug 2010, Sep 2010)
- Pharmacogenetics (January 2001)
- Proc. Natl. Acad. Sci. USA (November 1998)
- Protein Sequences & Data Analysis (May 1988, January 1993)
- Science (February 1998)

- Scientific Reports (Nature) (March 2015, September 2017, September 2021)
    - Receptors & Channels (May 1998)
    - Theoretical Biology and Medical Modelling (October 2011)
- Reviewed Grants for:
    - National Science Foundation (March 1989, September 1989, October 1990, October 1992 [two])
    - National Institute of General Medical Sciences, NIH (September 1994)
    - NASA Life Sciences Review (July 1995)
- Reviewed a book proposal for Cambridge University Press (December 2010).

PATENTS

1. **U.S. Patent 4,276,570**, 1981, Method and Apparatus for Producing an Image of a Person's Face at a Different Age. Nancy Burson and Thomas D. Schneider This method is now being used to age the images of missing children to aid in their recovery. See the October 1995 *Smithsonian* (volume 26, number 7, p. 70-80).
   `http://alum.mit.edu/www/toms/patent/face/`

2. **U.S. Patent 5,867,402**, 1999, Computational analysis of nucleic acid information defines binding sites, Thomas D. Schneider and Peter K. Rogan.
   `http://alum.mit.edu/www/toms/patent/walker/`

3. **U.S. Patent 6,774,222**, 2004, Molecular Computing Elements: Gates and Flip-Flops, Thomas D. Schneider and Paul N. Hengen; European Patent No: 1057118.
   `http://alum.mit.edu/www/toms/patent/molecularcomputing/`

4. **U.S. Patent 6,982,146**, 2006, High Speed Parallel Molecular Nucleic Acid Sequencing, Thomas D. Schneider and Denise Rubens,
   `http://alum.mit.edu/www/toms/patent/dnasequencing/`

5. **U.S. Patent 7,349,834**, 2008. Australian Patent No. 784085, 2006. European Patent 1204680 (10 September 2008). Canadian Patent 2380611, June 8, 2010. U.S. Patent 8,086,432, 2011. Molecular Motor, Thomas D. Schneider and Ilya G. Lyakhov,
   `http://alum.mit.edu/www/toms/patent/molecularrotationengine/`

6. **U.S. Patent 7,871,777**, 2011. Schneider, T. D., Lyakhov, I. G., and Needle, D.: Probe for nucleic acid sequencing and methods of use. European patent number 1960550
   `http://alum.mit.edu/www/toms/patent/medusa/`

7. **U.S. Patent 8,344,121**, 2013. Lyakhov, I. G., Schneider, T. D., and Needle, D.: Nanoprobes for detection or modification of molecules.
   `http://alum.mit.edu/www/toms/patent/nanoprobe/`

8. **U.S. Patent 8,703,734**, 2014. Lyakhov, I. G., Schneider, T. D., and Needle, D.: Nanoprobes for detection or modification of molecules.
   `http://alum.mit.edu/www/toms/patent/nanoprobe/`

9. **U.S. Patent 8,798,980**, 2014. Schneider, T. D. and Lyakhov, I. G.: Molecular motor
   http://alum.mit.edu/www/toms/patent/molecularrotationengine/

COMPUTER EXPERIENCE

- Since 1970 I have used the computer languages BASIC, FORTRAN, APL, LISP, Assembly, Algol, PL1, Pascal, PostScript, HTML, MatLab on many computer systems, mainly UNIX.

- I typeset my papers in LaTeX and maintain a web page for biologists on
  . LaTeX Style and BiBTeX Bibliography Formats for Biologists and TeX and LaTeX Resources

- Generally I build tools such as the Delila language and system of programs for manipulating nucleic acid sequences [2, 6] and making sequence logos [13] and sequence walkers [35]

- Three especially useful general tools for scientists are:
  - **atchange** - instant automation of repeated tasks
  - **nowhere** - jump to any directory in your file structure
  - **yvp** - give the Year, Volume and Page of a paper and your browser will immediately jump to the corresponding PubMed entry.

  These and other tools I have invented are available in my . toolkit

---

STUDENTS

I have brought Senior high school students into my lab by the National Cancer Institute's Werner H. Kirsten Student Intern Program (SIP)
(https://ncifrederick.cancer.gov/careers/student_programs/internships/sip/) since the beginning of the program. Publications are noted.

1. 1989-1990: R. Michael Stephens, a high school student supported by the NCI/FCRDC SIP and the NIH/FAES Mones Berman Memorial Fund [13, 19].

2. 1990-1991: Nathan D. Herman, SIP [17]

3. 1991-1992: Mark C. Shaner, SIP

4. 1991: Ian M. Blair, Montgomery Blair High School, Silver Spring, MD

5. 1992-1994: Stacy L. Bartram, SIP [36]

6. 1993-1994: Maria M. Alavanja, SIP

7. 1993-1994: Vishnu Jejjala, SIP, volunteer from University of Maryland

8. 1993-1995: Jaime A. Fenimore, SIP

9. 1994-1995: Leslie A. Strathern, SIP

10. 1994-1995: Paul A. Smith, High School student volunteer from Middletown High School

11. 1995-1996: R. Elaine Bucheimer, SIP [53]

12. 1996: Cheryl N. Johnston, SIP

13. 1995: Lisa E. Stewart, volunteer [36]

14. 1992-1997: Paul N. Hengen, Post doctoral student, Senior Staff Fellow [36, 55, 59, 61]

15. 1997-2003: Ryan K. Shultzaberger, SIP, college student, college graduate [42, 53, 71, 72, 73]

16. 1998-2002: Karen A. Lewis, SIP [52, 71, 73]

17. 1998-2009: Ilya G. Lyakhov, Postdoctoral fellow, Senior Staff Fellow [55, 59, 64, 72, 73, 75, 76, 83, 80]

18. 1999-2001: Shu Ouyang, Postdoctoral fellow, Senior Staff Fellow

19. 1999-2000: Nitasha G. Klar, SIP

20. 2000-2001: Brent M. Jewett, SIP

21. 2000-2002: Xiao (Sheldon) Ma, volunteer

22. 2001-2002: Brandon K. Cunningham, SIP, mentored by Ilya Lyakhov then TDS

23. 2001-2006: Zehua Chen, Postdoctoral fellow [65, 68, 73]

24. 2002-2009: Danielle Needle, Biologist [83, 80]

25. 2002-2003: Juliet Aiken, SIP

26. 2003-2005: Michael Y. Levashov, SIP Winner in the 2004 Spring Research Festival for his poster 'Computer Simulation of the Convergent Evolution of DNA Binding Sites as in the lambda cI/Cro Control System'. He was also a 63rd Annual Science Talent Search (STS) (2003-2004) Semifinalist.

27. 2004: Elizaveta Ershova, SIP

28. 2006-2009: Peyman Khalichi, Postdoctoral fellow

29. 2006-2009: Adam Diehl, SIP

30. 2009 Summer: David Wilson, High School Student

31. 2007 (summer) and 2008-2009: Blake Sweeney, postbac

32. 2012 (summer): Nicole Hearon, Cancer Research Intern

33. 2013-2014: Theo Nikolaitchik, SIP High School Student

34. 2013-2014: Ian Barry, volunteer

35. 2011-2018 (summers): Kevin Franco, Cancer Research Intern

SELECTED COLLABORATIONS

- 1990-1997: Dr. Dhruba Chattoraj, Peter Papp (NIH, Bethesda, MD) experimental information analysis of repA protein binding sites [22, 33]

- 1991-: Dr. Peter Rogan Department of Pediatrics, Milton S. Hershey Medical Center, Pennsylvania State University, Hershey Pa 17033; Application of sequence logos to PCR amplification for phylogenetic analysis [28]. Use of sequence logos for medical diagnosis [27, 37]

- 1991-: Dr. Gaisela Storz (NIH, Bethesda, MD) Analysis of OxyR binding sites [26, 52]

- 1992-: Dr. Kenn Rudd (University of Miami, FL) [18, 53]

- 2002: Dr. Krishnamachari Annangarachari (Chari) a Fulbright visiting scholar from Jawaharlal Nehru University, New Delhi India visited my lab for 6 months. He worked on the information analysis of p53 protein [76].

- 1998-: Dr. Michael Dean (NIH, Frederick, MD) [82, 56, 38].

- 1998-: Dr. Ken Kraemer (NIH, Bethesda, MD) on Xeroderma Pigmentosum [41, 51, 57, 62, 74].

RESEARCH INTERESTS

There must be mathematical laws that describe nucleic acid sequences and molecular interactions; my goal is to find these laws. What aspects of nucleic acids can be approached, and what mathematics should one use? The fruitful answer for me has been to apply Shannon's information theory to nucleic acid binding sites. During my Ph.D. thesis work I discovered that in many genetic systems the information in the binding site sequences on DNA or RNA to which proteins bind is just enough for the sites to be found in the genome [9]. This result is surprising because the number of sites and size of the genome are determined by history and physiology, so the amount of information in the binding sites must evolve toward the amount predicted using genome size and the number of sites. I confirmed these ideas both experimentally [12] and by using a computer simulation [50]. (You can try this model on your own computer by going to `http://alum.mit.edu/www/toms/papers/ev`.) Thus my work has three major components: theory, computer analysis and genetic engineering experiments.

Whenever one has a strong theory, anomalies are interesting. We have investigated several major ones at the lab bench because they lead to new insights into biology. One is the excess information found at bacteriophage T7 promoters [9, 12]. These sequences conserve twice as much information as the T7 polymerase requires to locate them in the presence of the bacterial genome. One possible explanation is that a second protein binds to the DNA. Alternatively the bacteriophage may be set up to overwhelm the bacterial defenses. We have found evidence supporting the latter hypothesis. In a second case, we discovered that the *E. coli* F plasmid incD region, which is responsible for correct plasmid partitioning to the daughter cells, has a three-fold excess conservation. This implies that three proteins bind there and we were able to identify three candidate binding proteins [17].

Another anomaly I found is unusually conserved bases involved in DNA replication and RNA transcription [22, 33]. Such cases can be detected by inspecting the sequence information along a binding site since the major groove of DNA can carry up to 2 bits of information while the minor groove can only support 1 bit. When the minor groove has more than 1 bit of information the DNA must not be in B form. We tested this idea in the bacteriophage P1 RepA system. Our experimental evidence suggests that the proteins are flipping bases out of the DNA to start helix melting, thereby initiating replication and transcription [54, 55].

Shannon's measure of information has the form of an average, which raises the question: for binding sites, what are the individual components that make up this average? The obvious answer is to consider it to be the average of the information for individual sequences in the set of binding sites. This immediately allows one to write down an equation that defines the individual information and this solution was proven to be unique by Dr. John Spouge [34].

To help visualize these results, we invented methods for graphically displaying a set of binding sites for the average as sequence logos [13] and for individual sequences as sequence walkers [34, 35, 36, 37]. These graphics have revealed many interesting details of a variety of binding sites and are now being used by researchers around the world. They allow rapid and quantitative visualization of genetic regions, detection of database errors, analysis of single nucleotide polymorphisms (SNPs) to distinguish polymorphisms from mutations (`http://alum.mit.edu/www/toms/g863a.html`) and quantitative genetic engineering of sequences. We have found a correlation between information measures of splice junctions and the severity of genetic diseases [37], and obtained a patent on this method [43].

For convenience, I divide my theoretical work into several levels. Level 0 is the study of genetic sequences bound by proteins or other macromolecules, briefly described above. The success of this theory suggested that other work of Shannon should also apply to molecular biology. Level 1 theory introduces the more general concept of the molecular machine which dissipates energy to make choices. From this I was able to develop the concept of a machine capacity equivalent to Shannon's channel capacity [15]. In Level 2, the Second Law of Thermodynamics is connected to the capacity theorem [16], and the limits on the functioning of Maxwell's Demon become clear [25]. Levels 3, the efficiency of molecular machines, which is often 70%, and 4, explaining the observed efficiency, are in preparation, but a short version has been published [78] and a review [79]. My

next major goal is to understand Level 5, the coding of molecular machines, by investigating the detailed structure and motions of molecules from the viewpoint of information and coding theory.

SPORTS

Skiing, Tai Chi Ch'uan, ultimate frisbee, scuba diving, mountain hiking, contra dancing, racket ball, bicycling

PUBLICATIONS

[1] L. Gold, D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo. Translational initiation in prokaryotes. *Annu. Rev. Microbiol.*, 35:365–403, 1981.

[2] T. D. Schneider, G. D. Stormo, J. S. Haemer, and L. Gold. A design for computer nucleic-acid-sequence storage, retrieval, and manipulation. *Nucleic Acids Res.*, 10:3013–3024, 1982.
https://doi.org/10.1093/nar/10.9.3013,
http://www.ncbi.nlm.nih.gov/pubmed/7099972.

[3] G. D. Stormo, T. D. Schneider, and L. M. Gold. Characterization of translational initiation sites in *E. coli*. *Nucleic Acids Res.*, 10:2971–2996, 1982. https://doi.org/10.1093/nar/10.9.2971.

[4] G. D. Stormo, T. D. Schneider, L. Gold, and A. Ehrenfeucht. Use of the 'Perceptron' algorithm to distinguish translational initiation sites in *E. coli*. *Nucleic Acids Res.*, 10:2997–3011, 1982.
https://doi.org/10.1093/nar/10.9.2997.

[5] L. Gold, M. Inman, E. Miller, D. Pribnow, T. D. Schneider, S. Shinedling, and G. Stormo. Translational regulation during bacteriophage T4 development. In B. F. C. Clark and H. U. Petersen, editors, *Gene Expression, Alfred Benzon Symposium 19*, pages 379–394, Copenhagen, 1984. Munksgaard.

[6] T. D. Schneider, G. D. Stormo, M. A. Yarus, and L. Gold. Delila system tools. *Nucleic Acids Res.*, 12:129–140, 1984.

https://doi.org/10.1093/nar/12.1Part1.129,
http://www.ncbi.nlm.nih.gov/pubmed/6694897.

[7] J. Childs, K. Villanueba, D. Barrick, T. D. Schneider, G. D. Stormo, L. Gold, M. Leitner, and M. Caruthers. Ribosome binding site sequences and function. In R. Calendar and L. Gold, editors, *Sequence Specificity in Transcription and Translation, UCLA Symposia on Molecular and Cellular Biology,* Vol. 30, pages 341–350, New York, 1985. Alan R. Liss, Inc. https://doi.org/10.1002/jcb.240290605.

[8] B. Clift, D. Haussler, R. McConnell, T. D. Schneider, and G. D. Stormo. Sequence landscapes. *Nucleic Acids Res.*, 14:141–158, 1986. https://doi.org/10.1093/nar/14.1.141.

[9] T. D. Schneider, G. D. Stormo, L. Gold, and A. Ehrenfeucht. Information content of binding sites on nucleotide sequences. *J. Mol. Biol.*, 188:415–431, 1986. https://doi.org/10.1016/0022-2836(86)90165-8, https://alum.mit.edu/www/toms/papers/schneider1986/.

[10] G. D. Stormo, T. D. Schneider, and L. Gold. Quantitative analysis of the relationship between nucleotide sequence and functional activity. *Nucleic Acids Res.*, 14:6661–6679, 1986. https://doi.org/10.1093/nar/14.16.6661.

[11] T. D. Schneider. Information and entropy of patterns in genetic switches. In G. J. Erickson and C. R. Smith, editors, *Maximum-Entropy and Bayesian Methods in Science and Engineering*, volume 2, pages 147–154, Dordrecht, The Netherlands, 1988. Kluwer Academic Publishers.

[12] T. D. Schneider and G. D. Stormo. Excess information at bacteriophage T7 genomic promoters detected by a random cloning technique. *Nucleic Acids Res.*, 17:659–674, 1989. https://doi.org/10.1093/nar/17.2.659.

[13] T. D. Schneider and R. M. Stephens. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.*, 18:6097–6100, 1990. https://doi.org/10.1093/nar/18.20.6097, https://alum.mit.edu/www/toms/papers/logopaper/.

[14] D. N. Arvidson, P. Youderian, T. D. Schneider, and G. D. Stormo. Automated kinetic assay of β-galactosidase activity. *BioTechniques*, 11(6):733–738, December 1991.

[15] T. D. Schneider. Theory of molecular machines. I. Channel capacity of molecular machines. *J. Theor. Biol.*, 148:83–123, 1991.

https://doi.org/10.1016/S0022-5193(05)80466-7,
https://alum.mit.edu/www/toms/papers/ccmm/.

[16] T. D. Schneider. Theory of molecular machines. II. Energy dissipation from molecular machines. *J. Theor. Biol.*, 148:125–137, 1991.
https://doi.org/10.1016/S0022-5193(05)80467-9,
https://alum.mit.edu/www/toms/papers/edmm/.

[17] N. D. Herman and T. D. Schneider. High information conservation implies that at least three proteins bind independently to F plasmid incD repeats. *J. Bacteriol.*, 174:3558–3560, 1992.
https://doi.org/10.1128/jb.174.11.3558-3560.1992.

[18] K. E. Rudd and T. D. Schneider. Compilation of *E. coli* ribosome binding sites. In Jeffrey H. Miller, editor, *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for* Escherichia coli *and Related Bacteria*, pages 17.19–17.45, Cold Spring Harbor, New York, 1992. Cold Spring Harbor Laboratory Press.

[19] R. M. Stephens and T. D. Schneider. Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J. Mol. Biol.*, 228:1124–1136, 1992.
https://doi.org/10.1016/0022-2836(92)90320-j,
https://alum.mit.edu/www/toms/papers/splice/.

[20] M. C. Shaner, I. M. Blair, and T. D. Schneider. Sequence logos: A powerful, yet simple, tool. In T. N. Mudge, V. Milutinovic, and L. Hunter, editors, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences, Volume 1: Architecture and Biotechnology Computing*, pages 813–821, Los Alamitos, CA, 1993. IEEE Computer Society Press.
https://alum.mit.edu/www/toms/papers/hawaii/.

[21] T. D. Schneider. Use of information theory in molecular biology. In D. J. Matzke, editor, *Workshop on Physics and Computation PhysComp '92, Proceedings of the Workshop on Physics and computation October 2-4, Dallas, Texas*, pages 102–110, Los Alamitos, CA, 1993. IEEE Computer Society Press.

[22] P. P. Papp, D. K. Chattoraj, and T. D. Schneider. Information analysis of sequences that bind the replication initiator RepA. *J. Mol. Biol.*, 233:219–230, 1993. https://doi.org/10.1006/jmbi.1993.1501
https://alum.mit.edu/www/toms/papers/helixrepa/.

[23] T. D. Schneider. Protein patterns as shown by sequence logos. In P. R.

Keller and Mary M. Keller, editors, *Visual Cues - Practical Data Visualization*, page 64, Piscataway, NJ, 1993. IEEE Press.

[24] D. Barrick, K. Villanueba, J. Childs, R. Kalil, T. D. Schneider, C. E. Lawrence, L. Gold, and G. D. Stormo. Quantitative analysis of ribosome binding sites in *E. coli. Nucleic Acids Res.*, 22:1287–1295, 1994. https://doi.org/10.1093/nar/22.7.1287.

[25] T. D. Schneider. Sequence logos, machine/channel capacity, Maxwell's demon, and molecular computers: a review of the theory of molecular machines. *Nanotechnology*, 5(1):1–18, 1994. https://doi.org/10.1088/0957-4484/5/1/001, https://alum.mit.edu/www/toms/papers/nano2/.

[26] M. B. Toledano, I. Kullik, F. Trinh, P. T. Baird, T. D. Schneider, and G. Storz. Redox-dependent shift of OxyR-DNA contacts along an extended DNA binding site: A mechanism for differential promoter selection. *Cell*, 78:897–909, 1994.

[27] P. K. Rogan and T. D. Schneider. Using information content and base frequencies to distinguish mutations from genetic polymorphisms in splice junction recognition sites. *Human Mutation*, 6:74–76, 1995. https://doi.org/10.1002/humu.1380060114, https://alum.mit.edu/www/toms/papers/colonsplice/.

[28] P. K. Rogan, J. J. Salvo, R. M. Stephens, and T. D. Schneider. Visual display of sequence conservation as an aid to taxonomic classification using PCR amplification. In Clifford A. Pickover, editor, *Visualizing Biological Information*, pages 21–32, Singapore, 1995. World Scientific.

[29] T. D. Schneider. Genetic patterns as shown by sequence logos. In C. Pickover, editor, *The Pattern Book: Fractals, Art and Nature*, pages 44–45, River Edge, NJ, 1995. World Scientific.

[30] T. D. Schneider. New approaches in mathematical biology: Information theory and molecular machines. In Julian Chela-Flores and Francois Raulin, editors, *Chemical Evolution: Physics of the Origin and Evolution of Life*, pages 313–321, Dordrecht, The Netherlands, 1996. Kluwer Academic Publishers. https://doi.org/10.1007/978-94-009-1712-5_28.

[31] T. D. Schneider. Reading of DNA sequence logos: Prediction of major groove binding by information theory. *Meth. Enzym.*, 274:445–455, 1996. https://alum.mit.edu/www/toms/papers/oxyr/, https://doi.org/10.1016/S0076-6879(96)74036-3.

[32] T. D. Schneider and D. N. Mastronarde. Fast multiple alignment of ungapped DNA sequences using information theory and a relaxation method. *Discrete Applied Mathematics*, 71:259–268, 1996. `https://alum.mit.edu/www/toms/papers/malign`, `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2785095/`, `https://doi.org/10.1016/S0166-218X(96)00068-6`.

[33] D. K. Chattoraj and T. D. Schneider. Replication control of plasmid P1 and its host chromosome: the common ground. *Prog. Nucl. Acid Res. Mol. Biol.*, 57:145–186, 1997. `http://www.sciencedirect.com/science/article/pii/S0079660308602809`, `https://doi.org/10.1016/S0079-6603(08)60280-9`.

[34] T. D. Schneider and J. Spouge. Information content of individual genetic sequences. *J. Theor. Biol.*, 189:427–441, 1997. `https://doi.org/10.1006/jtbi.1997.0540`, `https://alum.mit.edu/www/toms/papers/ri/`.

[35] T. D. Schneider. Sequence walkers: a graphical method to display how binding proteins interact with DNA or RNA sequences. *Nucleic Acids Res.*, 25:4408–4415, 1997. `https://doi.org/10.1093/nar/25.21.4408`, `https://alum.mit.edu/www/toms/papers/walker/`, erratum: NAR 26(4): 1135, 1998.

[36] P. N. Hengen, S. L. Bartram, L. E. Stewart, and T. D. Schneider. Information analysis of Fis binding sites. *Nucleic Acids Res.*, 25:4994–5002, 1997. `https://doi.org/10.1093/nar/25.24.4994`, `https://alum.mit.edu/www/toms/papers/fisinfo/`.

[37] P. K. Rogan, B. M. Faux, and T. D. Schneider. Information analysis of human splice site mutations. *Human Mutation*, 12:153–171, 1998. `https://doi.org/10.1002/(sici)1098-1004(1998)12:3%3C153::aid-humu3%3E3` Erratum in: Hum Mutat 1999;13(1):82. `https://alum.mit.edu/www/toms/papers/rfs/`.

[38] R. Allikmets, W. W. Wasserman, A. Hutchinson, P. Smallwood, J. Nathans, P. K. Rogan, T. D. Schneider, and M. Dean. Organization of the ABCR gene: analysis of promoter and splice junction sequences. *Gene*, 215:111–122, 1998. `https://alum.mit.edu/www/toms/papers/abcr/`.

[39] S. R. Matten, T. D. Schneider, S. Ringquist, and W. S. A. Brusilow. Identification of an intragenic ribosome binding site that affects expression of the uncB gene of the Escherichia coli proton-translocating ATPase (unc)

operon. *J. Bacteriol*, 180:3940–3945, 1998.
`https://jb.asm.org/content/180/15/3940.long`.

[40] P. W. Tooley, J. J. Salvo, T. D. Schneider, and P. K. Rogan. Phylogenetic inference based on information theory-based PCR amplification. *Journal of Phytopathology*, 146:427–430, 1998.
`https://doi.org/10.1111/j.1439-0434.1998.tb04776.x`.

[41] S. G. Khan, H. L. Levy, R. Legerski, E. Quackenbush, J. T. Reardon, S. Emmert, A. Sancar, L. Li, T. D. Schneider, J. E. Cleaver, and K. H. Kraemer. Xeroderma pigmentosum group C splice mutation associated with autism and hypoglycinemia. *J. Investigative Dermatology*, 111:791–796, 1998.
`http://www.nature.com/jid/journal/v111/n5/abs/5600180a.html`,
`https://doi.org/10.1046/j.1523-1747.1998.00391.x`.

[42] R. K. Shultzaberger and T. D. Schneider. Using sequence logos and information analysis of Lrp DNA binding sites to investigate discrepancies between natural selection and SELEX. *Nucleic Acids Res.*, 27:882–887, 1999. `https://alum.mit.edu/www/toms/papers/lrp/`,
`https://doi.org/10.1093/nar/27.3.882`.

[43] T. D. Schneider and P. K. Rogan. Computational analysis of nucleic acid information defines binding sites, United States Patent 5867402, 1999.
`https://alum.mit.edu/www/toms/patent/walker/`.

[44] M. Zheng, B. Doan, T. D. Schneider, and G. Storz. OxyR and SoxRS regulation of *fur*. *J. Bacteriol.*, 181:4639–4643, 1999.
`https://alum.mit.edu/www/toms/papers/oxyrfur/`.

[45] T. I. Wood, K. L. Griffith, W. P. Fawcett, K.-W. Jair, T. D. Schneider, and R. E. Wolf. Interdependence of the position and orientation of SoxS binding sites in the transcriptional activation of the class I subset of *Escherichia coli* superoxide-inducible promoters. *Mol. Microbiol.*, 34:414–430, 1999.
`https://doi.org/10.1046/j.1365-2958.1999.01598.x`.

[46] T. D. Schneider. Measuring molecular information. *J. Theor. Biol.*, 201:87–92, 1999.
`https://alum.mit.edu/www/toms/papers/ridebate/`,
`https://doi.org/10.1006/jtbi.1999.1012`.

[47] T. D. Schneider. The bottle. *Nature*, 406:351, 2000.

[48] N. Kannan, T. D. Schneider, and S. Vishveshwara. Logos for amino-acid preferences in different backbone packing density regions of protein

structural classes. *Acta Crystallogr D Biol Crystallogr*, 56:1156–1165, 2000.
`https://alum.mit.edu/www/toms/papers/Kannan.Vishveshwara2000/`.

[49] S. R. Svojanovsky, T. D. Schneider, and P. K. Rogan. Redundant designations of BRCA1 intron 11 splicing mutation; c. 4216-2A>G; IVS11-2A>G; L78833, 37698, A>G. *Human Mutation*, 16:264, 2000. `http://www3.interscience.wiley.com/cgi-bin/abstract/73001161/START`.

[50] T. D. Schneider. Evolution of biological information. *Nucleic Acids Res.*, 28:2794–2799, 2000. `https://doi.org/10.1093/nar/28.14.2794`, `https://alum.mit.edu/www/toms/papers/ev/`.

[51] S. Emmert, T. D. Schneider, S. G. Khan, and K. H. Kraemer. The human XPG gene: gene architecture, alternative splicing and single nucleotide polymorphisms. *Nucleic Acids Res.*, 29:1443–1452, 2001.

[52] M. Zheng, X. Wang, B. Doan, K. A. Lewis, T. D. Schneider, and G. Storz. Computation-Directed Identification of OxyR-DNA Binding Sites in *Escherichia coli*. *J. Bacteriol.*, 183:4571–4579, 2001. `https://doi.org/10.1128/JB.183.15.4571-4579.2001`.

[53] R. K. Shultzaberger, R. E. Bucheimer, K. E. Rudd, and T. D. Schneider. Anatomy of *Escherichia coli* Ribosome Binding Sites. *J. Mol. Biol.*, 313:215–228, 2001. `https://doi.org/10.1006/jmbi.2001.5040`, `https://alum.mit.edu/www/toms/papers/flexrbs/`.

[54] T. D. Schneider. Strong minor groove base conservation in sequence logos implies DNA distortion or base flipping during replication and transcription initiation. *Nucleic Acids Res.*, 29:4881–4891, 2001. `https://doi.org/10.1093/nar/29.23.4881`, `https://alum.mit.edu/www/toms/papers/baseflip/`.

[55] I. G. Lyakhov, P. N. Hengen, D. Rubens, and T. D. Schneider. The P1 Phage Replication Protein RepA Contacts an Otherwise Inaccessible Thymine N3 Proton by DNA Distortion or Base Flipping. *Nucleic Acids Res.*, 29:4892–4900, 2001. `https://doi.org/10.1093/nar/29.23.4892`, `https://alum.mit.edu/www/toms/papers/repan3/`.

[56] I. Arnould, L. M. Schriml, C. Prades, M. Lachtermachter-Triunfol, T. Schneider, C. Maintoux, C. Lemoine, D. Debono, C. Devaud, L. Naudin, S. Bauché, M. Annat, T. Annilo, R. Allikmets, B. Gold, P. Denèfle, M. Rosier, and M. Dean. Identifying and characterizing a five-gene cluster of ATP-binding cassette transporters mapping to human chromosome 17q24: a new subgroup within the ABCA subfamily. *GeneScreen*,

1:157–164, 2001.
`https://doi.org/10.1046/j.1466-920x.2001.00038.x.`

[57] S. G. Khan, V. Muniz-Medina, T. Shahlavi, C. C. Baker, H. Inui, T. Ueda, S. Emmert, T. D. Schneider, and K. H. Kraemer. The human *XPC* DNA repair gene: arrangement, splice site information content and influence of a single nucleotide polymorphism in a splice acceptor site on alternative splicing and function. *Nucleic Acids Res.*, 30:3624–3631, 2002.

[58] T. D. Schneider. Consensus Sequence Zen. *Applied Bioinformatics*, 1:111–119, 2002.
`https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1852464/,`
`https://alum.mit.edu/www/toms/papers/zen/.`

[59] P. N. Hengen, I. G. Lyakhov, L. E. Stewart, and T. D. Schneider. Molecular flip-flops formed by overlapping Fis sites. *Nucleic Acids Res.*, 31(22):6663–6673, 2003. `https://doi.org/10.1093/nar/gkg877.`

[60] T. D. Schneider. Some lessons for molecular biology from information theory. In Karmeshu, editor, *Entropy Measures, Maximum Entropy Principle and Emerging Applications. Special Series on Studies in Fuzziness and Soft Computing. (Festschrift Volume in honour of Professor J.N. Kapour, Jawaharlal Nehru University, India)*, volume 119, pages 229–237, New York, 2003. Springer-Verlag.
`https://alum.mit.edu/www/toms/papers/lessons2003/.`

[61] T. D. Schneider and P. N. Hengen. MOLECULAR COMPUTING ELEMENTS: GATES AND FLIP-FLOPS, United States Patent 6,774,222, European Patent 1057118, 2004, 2004. US Patent WO 99/42929, PCT/US99/03469.
`https://alum.mit.edu/www/toms/patent/molecularcomputing/.`

[62] S. G. Khan, A. Metin, E. Gozukara, H. Inui, T. Shahlavi, V. Muniz-Medina, C. C. Baker, T. Ueda, J. R. Aiken, T. D. Schneider, and K. H. Kraemer. Two essential splice lariat branchpoint sequences in one intron in a xeroderma pigmentosum DNA repair gene: mutations result in reduced XPC mRNA levels that correlate with cancer risk. *Hum Mol Genet*, 13:343–352, 2004.
`https://doi.org/10.1093/hmg/ddh026.`

[63] John M. Hancock and Marketa J. Zvelebil. *Dictionary of Bioinformatics and Computational Biology*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.
`https://alum.mit.edu/www/toms/papers/Hancock.Zvelebil2004/.`

Thomas D. Schneider contributed 50 entries to the dictionary. The web links to Tom Schneider's web site are incorrect but this has been handled on the server computer. See `https://alum.mit.edu/www/toms/`. The entries in the book originated from the online page "A Glossary for Molecular Information Theory and the Delila System", `https://alum.mit.edu/www/toms/glossary.html`.

[64] I. G. Lyakhov, T. D. Schneider, G. A. Lyakhov, and N. V. Suyazov. Orientational ordering of protein micro- and nanoparticles in a nonuniform magnetic field. *Physics of Wave Phenomena*, 13(1):1–14, 2005.

[65] Z. Chen and T. D. Schneider. Information theory based T7-like promoter models: classification of bacteriophages and differential evolution of promoters and their polymerases. *Nucleic Acids Res.*, 33:6172–6187, 2005. `https://doi.org/10.1093/nar/gki915`, `https://alum.mit.edu/www/toms/papers/t7like/`.

[66] T. D. Schneider. Claude Shannon: Biologist. *IEEE Engineering in Medicine and Biology Magazine*, 25(1):30–33, 2006. `https://alum.mit.edu/www/toms/papers/shannonbiologist/`, `https://doi.org/10.1109/MEMB.2006.1578661`.

[67] T. D. Schneider and D. Rubens. HIGH SPEED PARALLEL MOLECULAR NUCLEIC ACID SEQUENCING, 2006. US Patent 6,982,146, `https://alum.mit.edu/www/toms/patent/dnasequencing/`.

[68] Z. Chen and T. D. Schneider. Comparative analysis of tandem T7-like promoter containing regions in enterobacterial genomes reveals a novel group of genetic islands. *Nucleic Acids Res.*, 34:1133–1147, 2006. `https://doi.org/10.1093/nar/gkj511`, `https://alum.mit.edu/www/toms/papers/t7island/`.

[69] E. Bindewald, T. D. Schneider, and B. A. Shapiro. CorreLogo: An online server for 3D sequence logos of RNA and DNA alignments. *Nucleic Acids Res.*, 34:w405–w411, 2006. `https://doi.org/10.1093/nar/gkl269`, `https://alum.mit.edu/www/toms/papers/correlogo/`, `https://rnastructure.cancer.gov/correlogo/`.

[70] T. D. Schneider. Twenty Years of Delila and Molecular Information Theory: The Altenberg-Austin Workshop in Theoretical Biology Biological Information, Beyond Metaphor: Causality, Explanation, and Unification Altenberg, Austria, 11-14 July 2002. *Biological Theory*, 1(3):250–260, 2006. `https://doi.org/10.1162/biot.2006.1.3.250`,

https://link.springer.com/article/10.1162/biot.2006.1.3.250,
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2139980/,
https://alum.mit.edu/www/toms/papers/schneider2006/.

[71] R. K. Shultzaberger, Zehua Chen, Karen A. Lewis, and T. D. Schneider. Anatomy of *Escherichia coli* σ^{70} promoters. *Nucleic Acids Res.*, 35:771–788, 2007. https://doi.org/10.1093/nar/gkl956, https://alum.mit.edu/www/toms/papers/flexprom/.

[72] R. K. Shultzaberger, L. R. Roberts, I. G. Lyakhov, I. A. Sidorov, A. G. Stephen, R. J. Fisher, and T. D. Schneider. Correlation between binding rate constants and individual information of *E. coli* Fis binding sites. *Nucleic Acids Res.*, 35:5275–5283, 2007.
https://doi.org/10.1093/nar/gkm471,
https://alum.mit.edu/www/toms/papers/fisbc/.

[73] Z. Chen, K. A. Lewis, R. K. Shultzaberger, I. G. Lyakhov, M. Zheng, B. Doan, G. Storz, and T. D. Schneider. Discovery of Fur binding site clusters in *Escherichia coli* by information theory models. *Nucleic Acids Res.*, 35:6762–6777, 2007.
https://alum.mit.edu/www/toms/papers/fur/.

[74] H. Inui, K. S. Oh, C. Nadem, T. Ueda, S. G. Khan, A. Metin, E. Gozukara, S. Emmert, H. Slor, D. B. Busch, C. C. Baker, J. J. Digiovanna, D. Tamura, C. S. Seitz, A. Gratchev, W. H. Wu, K. Y. Chung, H. J. Chung, E. Azizi, R. Woodgate, T. D. Schneider, and K. H. Kraemer. Xeroderma Pigmentosum-Variant Patients from America, Europe, and Asia. *J Invest Dermatol*, 128:2055–2068, 2008.
https://alum.mit.edu/www/toms/papers/xpv/.

[75] T. D. Schneider and I. G. Lyakhov. MOLECULAR MOTOR, 2008. US Patent 7,349,834, Australian Patent 784085, European Patent 1204680, Canadian Patent 2380611, June 8, 2010.
https://alum.mit.edu/www/toms/patent/molecularrotationengine/.

[76] I. Lyakhov, K. Annangarachari, and T. D. Schneider. Discovery of Novel Tumor Suppressor p53 Response Elements Using Information Theory. *Nucleic Acids Res.*, 36:3828–3833, 2008.
https://alum.mit.edu/www/toms/papers/p53/.

[77] M. R. Hemm, B. J. Paul, T. D. Schneider, G. Storz, and K. E. Rudd. Small membrane proteins found by comparative genomics and ribosome binding site models. *Mol. Microbiol.*, 70:1487–1501, 2008.

https://doi.org/10.1111/j.1365-2958.2008.06495.x,
https://alum.mit.edu/www/toms/papers/smallproteins/.

[78] T. D. Schneider. 70% efficiency of bistate molecular machines explained by information theory, high dimensional geometry and evolutionary convergence. *Nucleic Acids Res.*, 38:5995–6006, 2010.
https://doi.org/doi:10.1093/nar/gkq389,
https://alum.mit.edu/www/toms/papers/emmgeo/.

[79] T. D. Schneider. A brief review of molecular information theory. *Nano Communication Networks*, 1:173–180, 2010.
https://doi.org/10.1016/j.nancom.2010.09.002,
https://alum.mit.edu/www/toms/papers/brmit/.

[80] Thomas D. Schneider, Ilya Lyakhov, and Danielle Needle. PROBE FOR NUCLEIC ACID SEQUENCING AND METHODS OF USE, 2010. US patent claims allowed; European patent number 1960550 granted on 2010 September 15. US patent number 7,871,777 granted on 2011 January 18.
https://alum.mit.edu/www/toms/patent/medusa/ .

[81] J. H. Jeong, H. J. Kim, K. H. Kim, M. Shin, Y. Hong, J. H. Rhee, T. D. Schneider, and H. E. Choy. An unusual feature associated with *LEE1* P1 promoters in enteropathogenic *Escherichia coli* (EPEC). *Mol. Microbiol.*, 83:612–622, 2012.
https://alum.mit.edu/www/toms/papers/leeprom/.

[82] H. Lou, H. Li, M. Yeager, K. Im, B. Gold, T. D. Schneider, J. F. Fraumeni Jr, S. J. Chanock, S. K. Anderson, and M. Dean. Promoter variants in the MSMB gene associated with prostate cancer regulate MSMB/NCOA4 fusion transcripts. *Hum. Genet.*, 131:1453–1466, 2012.
https://doi.org/10.1007/s00439-012-1182-2.

[83] Ilya Lyakhov, Thomas D. Schneider, and Danielle Needle. NANOPROBES FOR DETECTION OR MODIFICATION OF MOLECULES, 2013. US Patent 8,344,121 issued January 1, 2013
https://alum.mit.edu/www/toms/patent/nanoprobe/.

[84] Lisa A. Urry, Michael L. Cain, Steven A. Wasserman, Peter V. Minorsky, Robert B. Jackson, and Jane B. Reece. *Instructor's Review Copy for Campbell Biology in Focus*. Pearson Education Ltd, Boston, 9th edition, 2014. Sequence logos are explained on page 284. Thomas Schneider supplied the figures.

[85] T. D. Schneider. (various entries). In John M. Hancock and Marketa J.

Zvelebil, editors, *Concise Encyclopaedia of Bioinformatics and Computational Biology, 2nd Edition*, West Sussex, UK, 2014. Wiley-Blackwell.
`http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470978716.html`.

[86] Ilya Lyakhov, Thomas D. Schneider, and Danielle Needle. NANOPROBES FOR DETECTION OR MODIFICATION OF MOLECULES, 2013. US Patent 8,703,734 issued April 22, 2014
`https://alum.mit.edu/www/toms/patent/nanoprobe/`.

[87] K. M. Pluchino, D. Esposito, J. K. Moen, M. D. Hall, J. P. Madigan, S. Shukla, L. V. Procter, V. E. Wall, T. D. Schneider, I. Pringle, S. V. Ambudkar, D. R. Gill, S. C. Hyde, and M. M. Gottesman. Identification of a Cryptic Bacterial Promoter in Mouse (*mdr1a*) P-Glycoprotein cDNA. *PLoS One*, 10:e0136396, 2015.

[88] Z. Qian, A. Trostel, D. E. A. Lewis, S. J. Lee, X. He, A. M. Stringer, J. T. Wade, T. D. Schneider, T. Durfee, and S. Adhya. Genome-Wide Transcriptional Regulation and Chromosome Structural Arrangement by GalR in *E. coli*. *Front Mol Biosci*, 3:74, 2016.
`https://doi.org/10.3389/fmolb.2016.00074`.

[89] K. J. Fenstermacher, V. Achuthan, T. D. Schneider, and J. J. DeStefano. An Evolutionary/Biochemical Connection Between Promoter- and Primer-Dependent Polymerases Revealed by Selective Evolution of Ligands by Exponential Enrichment (SELEX). *J. Bacteriol.*, 200:e00579–17, 2018.
`https://doi.org/10.1128/JB.00579-17`.

[90] Z. Sun, C. Cagliero, J. Izard, Y. Chen, Y. N. Zhou, W. F. Heinz, T. D. Schneider, and D. J. Jin. Density of σ70 promoter-like sites in the intergenic regions dictates the redistribution of RNA polymerase during osmotic stress in *Escherichia coli*. *Nucleic Acids Res.*, 47:3970–3985, 2019.
`https://doi.org/10.1093/nar/gkz159`.

[91] L. C. Thomason, K. Morrill, G. Murray, C. Court, B. Shafer, T. D. Schneider, and D. L. Court. Elements in the λ Immunity Region Regulate Phage Development: Beyond the "Genetic Switch". *Mol. Microbiol.*, 112:1798–1813, 2019. `https://doi.org/10.1111/mmi.14394`.

[92] T. D. Schneider and V. Jejjala. Restriction enzymes use a 24 dimensional coding space to recognize 6 base long DNA sequences. *PLoS One*, 14:e0222419, 2019.
`https://doi.org/10.1371/journal.pone.0222419`,

https://alum.mit.edu/www/toms/papers/lattice/.

[93] T. D. Schneider. Generalizing the Isothermal Efficiency by Using Gaussian Distributions. *PLOS ONE*, pages 1–17, 2023.
https://biorxiv.org/cgi/content/short/2022.12.12.520049v1,
https://alum.mit.edu/www/toms/papers/geneff/,
https://doi.org/10.1371/journal.pone.0279758.

[94] T. D. Schneider. Information theory primer, with an appendix on logarithms. *Published on the web*, 2013, 2013.
https://doi.org/10.13140/2.1.2607.2000,
https://alum.mit.edu/www/toms/papers/primer/.

REFERENCES

- Richard J. Roberts, Ph. D., Nobel Laureate 1993
  New England Biolabs
  240 County Road
  Ipswich, MA 01938-2723
  USA
  Tel: (978) 380-7405
  Fax: (978) 380-7406
  email: roberts@neb.com

- Peter K. Rogan, Ph.D. (progan@uwo.ca)
  Adjunct Professor / Professor Emeritus
  Department of Biochemistry
  Schulich School of Medicine and Dentistry
  University of Western Ontario
  London, Ontario N6A 5C1 Canada
  T: (226) 224-3699
  http://www.cytognomix.org

  President, CytoGnomix Inc.® (info@cytognomix.com)
  60 North Centre Rd., Box 27052
  London, Ontario N5X 3X5 Canada
  http://www.cytognomix.com

- John L. Spouge, M.D., Ph.D.
  Senior Investigator
  National Library of Medicine, NIH
  38A/6N603
  Bethesda, MD 20894
  301-402-9310
  https://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/
  spouge@ncbi.nlm.nih.gov

- John S. Garavelli, Ph.D.
  Center for Bioinformatics and Computational Biology
  Delaware Biotechnology Institute
  15 Innovation Way, Suite 205
  University of Delaware
  Newark, DE 19711
  jsgarave@udel.edu
  302-831-6922
  `https://bioinformatics.udel.edu/People/Garavelli`