

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: BAKER, DAVID

eRA COMMONS USER NAME (credential, e.g., agency login): DABAKER

POSITION TITLE: Director

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Harvard University, Cambridge, MA	B.A.	1984	Biology
Univ of California, Berkeley, CA	Ph.D.	12/1989	Biochemistry
Univ of California, San Francisco, CA	Postdoctoral Fellow	1993	Biophysics

A. Personal Statement

My colleagues and I have developed the Rosetta computational methodology for predicting and designing macromolecular structures, interactions, and functions. We have used this methodology to predict the structures and interactions of naturally occurring biomolecules, and to design new proteins with new structures, interactions and functions. More recently, we have been developing deep learning software for protein design which is enabling highly accurate and robust design of new protein binders, catalysts and self-assembling materials.

B. Positions, Scientific Appointments, and Honors

Positions and Scientific Appointments

- 2012–Present Investigator, Howard Hughes Medical Institute
2005–Present Director, Institute for Protein Design, University of Washington, Seattle, WA
2004–Present Professor, University of Washington, Department of Biochemistry, Seattle, WA
2004–Present Adjunct Professor of Bioengineering, Computer Science, Genome Sciences, Chemical Engineering, and Physics, University of Washington, Seattle, WA
2000–2004 Associate Investigator, Howard Hughes Medical Institute
2000–2005 Associate Professor, University of Washington, Department of Biochemistry, Seattle, WA
1993–2000 Assistant Professor, University of Washington, Department of Biochemistry, Seattle, WA

Honors

2024	Nobel Prize in Chemistry, Royal Swedish Academy of Sciences
2021	Prize in Life Sciences, Breakthrough
2020	Fellow, American Institute for Medical and Biological Engineering
2018	Public Lecture, International Solvay Institutes
2018	Hans Neurath Award, Protein Society
2017	Henrietta and Aubrey Davis Endowed Professorship in Biochemistry, University of Washington
2014	Perlman Memorial Award, American Chemical Society
2012	Centenary Award, Biochemical Society
2011	Inventor of the Year Award, University of Washington
2009	Member, American Academy of Sciences
2008	Sackler Prize in Biophysics, Tel Aviv University
2007	Editorial Board, Proceedings of the National Academy of Sciences
2006	Member, National Academy of Sciences
2004	Feynman Prize, Foresight Institute
2004	Newcomb Cleveland Prize, American Academy of Sciences
2002	Overton Prize, International Society for Computational Biology
2000	Young Investigator Award, Beckman
1994	Fellowship in Science and Engineering, Packard
1994	Young Investigator Award, National Science Foundation
1993	Fellowship, Boyer Foundation

C. Contribution to Science

1. **How proteins fold.** To investigate the extent to which amino acid sequences determine protein folding rates, we used selection methods to identify very different sequences that fold to the same structures. While the stabilities of these proteins were generally less than the native protein, the folding rates were as often faster as slower, suggesting that evolution has not optimized sequences for rapid folding. Since introducing substantial sequence variation did not significantly affect protein folding rates, we sought other factors that determined the rate of folding. A series of experimental and computational analyses established that protein folding rates and mechanisms are largely determined by the topology of the native structure. A particularly notable observation was that protein folding rates are strongly correlated with the contact order of the native structure (the sequence separation between contacting residues in the protein structure), with low contact order proteins folding orders of magnitude faster than high contact order proteins.
 - a. Rocklin GJ, Chidyausiku TM, Goreshnik I, Ford A, Houlston S, Lemak A, Carter L, Ravichandran R, Mulligan VK, Chevalier A, Arrowsmith CH, Baker D. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science*. 2017 Jul 14;357(6347):168-175. PubMed Central PMCID: PMC5568797.
 - b. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. Experiment and theory highlight role of native state topology in SH3 folding. *Nat Struct Biol*. 1999 Nov;6(11):1016-24. PubMed PMID: 10542092.
 - c. Alm E, Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc Natl Acad Sci U S A*. 1999 Sep 28;96(20):11305-10. PubMed Central PMCID: PMC18029.
 - d. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol*. 1998 Apr 10;277(4):985-94. PubMed PMID: 9545386.
2. **Protein structure prediction and deep learning.** Guided by the insights gained in our studies of protein folding mechanism, we developed the Rosetta ab initio structure prediction methodology which builds protein structures by fragment assembly. The CASP blind structure prediction experiments showed that the Rosetta protein structure prediction methodology was a significant improvement over previous approaches. We developed methods for efficiently refining protein models in a physically realistic all atom potential, coupled this

with lower resolution conformational search methods, and showed that not only monomeric protein structures but also protein-protein complexes (the docking problem), higher order symmetric protein assemblies, membrane proteins, and RNA structures could be modeled accurately by searching for the lowest energy state provided the space to be searched was not too large. We then showed that the Rosetta approach could generate quite accurate models of more complex systems when provided with limited experimental data to guide conformational sampling. Rosetta supplemented with experimental data has become a powerful and widely used approach to solve macromolecular structures using sparse NMR data (CS-Rosetta), low-resolution x-ray diffraction data (MR-Rosetta), cryo-electron microscopy data, and co-evolution sequence information (Gremlin-Rosetta).

- a. Ovchinnikov S, Park H, Varghese N, Huang PS, Pavlopoulos GA, Kim DE, Kamisetty H, Kyripides NC, Baker D. Protein structure determination using metagenome sequence data. *Science*. 2017 Jan 20;355(6322):294-298. PubMed Central PMCID: PMC5493203.
 - b. DiMaio F, Terwilliger TC, Read RJ, Wlodawer A, Oberdorfer G, Wagner U, Valkov E, Alon A, Fass D, Axelrod HL, Das D, Vorobiev SM, Iwaï H, Pokkuluri PR, Baker D. Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature*. 2011 May 26;473(7348):540-3. PubMed Central PMCID: PMC3365536.
 - c. Raman S, Lange OF, Rossi P, Tyka M, Wang X, Aramini J, Liu G, Ramelot TA, Eletsky A, Szyperski T, Kennedy MA, Prestegard J, Montelione GT, Baker D. NMR structure determination for larger proteins using backbone-only data. *Science*. 2010 Feb 19;327(5968):1014-8. PubMed Central PMCID: PMC2909653.
 - d. Gray JJ, Moughon SE, Kortemme T, Schueler-Furman O, Misura KM, Morozov AV, Baker D. Protein-protein docking predictions for the CAPRI experiment. *Proteins*. 2003 Jul 1;52(1):118-22. PubMed PMID: 12784377; NIHMSID: NIHMS308621.
3. **Design of protein structure and immunogens.** The Rosetta structure prediction methodology described above searches for the lowest energy structure for a given sequence, and we realized that we could invert the process to search for the lowest energy sequence for a desired structure—the protein design problem. We demonstrated proof-of-concept for de novo protein design with the design of TOP7, a novel protein with a fold not found in nature, and later developed general principles for designing hyperstable idealized alpha-beta proteins and helical bundles. With the capability of designing stable protein structures in hand, we developed methods for stabilizing both linear and complex epitopes from pathogen proteins, and showed that these could elicit neutralizing antibodies in animals, opening up computational design approaches to developing improved vaccines. More recently, we have developed and been using a guided diffusion model called RFdiffusion to design even more complex structures and functions.
- a. Hosseinzadeh P, Bhardwaj G, Mulligan VK, Shortridge MD, Craven TW, Pardo-Avila F, Rettie SA, Kim DE, Silva DA, Ibrahim YM, Webb IK, Cort JR, Adkins JN, Varani G, Baker D. Comprehensive computational design of ordered peptide macrocycles. *Science*. 2017 Dec 15;358(6369):1461-1466. PubMed Central PMCID: PMC5860875.
 - b. Correia BE, Bates JT, Loomis RJ, Baneyx G, Carrico C, Jardine JG, Rupert P, Correnti C, Kalyuzhnii O, Vittal V, Connell MJ, Stevens E, Schroeter A, Chen M, Macpherson S, Serra AM, Adachi Y, Holmes MA, Li Y, Klevit RE, Graham BS, Wyatt RT, Baker D, Strong RK, Crowe JE Jr, Johnson PR, Schief WR. Proof of principle for epitope-focused vaccine design. *Nature*. 2014 Mar 13;507(7491):201-6. PubMed Central PMCID: PMC4260937.
 - c. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, Baker D. Principles for designing ideal protein structures. *Nature*. 2012 Nov 8;491(7423):222-7. PubMed Central PMCID: PMC3705962.
 - d. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 2003 Nov 21;302(5649):1364-8. PubMed PMID: 14631033.
4. **Design of small molecule binding, catalysis, protein interactions, and self-assembly.** A grand challenge in computational protein design is creating new binding proteins de novo for use in therapeutics and diagnostics. We have developed general methods for designing proteins which bind with high affinity/specificity to sites of interest of therapeutic importance on protein targets, both human and pathogen

(influenza, bacterial toxins, oncogenic proteins). We developed general methods for designing catalysts for arbitrary chemical reactions starting from a description of the reaction transition state geometry, and used the approach to design catalysts for a number of reactions not catalyzed by naturally occurring enzymes. We have developed an approach to computationally designing self-assembling nanomaterials and used it to design new proteins that self-assemble into regular tetrahedral, octahedral and icosahedral structures as well as two dimensional layers with near atomic-level accuracy. We are now developing these self-assembling materials for targeted delivery and vaccine applications.

- a. Yeh AH, Norn C, Kipnis Y, Tischer D, Pellock SJ, Evans D, Ma P, Lee GR, Zhang JZ, Anishchenko I, Coventry B, Cao L, Dauparas J, Halabiya S, DeWitt M, Carter L, Houk KN, Baker D. De novo design of luciferases using deep learning. *Nature*. 2023 Feb;614(7949):774-780. PubMed Central PMCID: PMC9946828.
 - b. Chevalier A, Silva DA, Rocklin GJ, Hicks DR, Vergara R, Murapa P, Bernard SM, Zhang L, Lam KH, Yao G, Bahl CD, Miyashita SI, Goreshnik I, Fuller JT, Koday MT, Jenkins CM, Colvin T, Carter L, Bohn A, Bryan CM, Fernández-Velasco DA, Stewart L, Dong M, Huang X, Jin R, Wilson IA, Fuller DH, Baker D. Massively parallel de novo protein design for targeted therapeutics. *Nature*. 2017 Oct 5;550(7674):74-79. PubMed Central PMCID: PMC5802399.
 - c. King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D. Accurate design of co-assembling multi-component protein nanomaterials. *Nature*. 2014 Jun 5;510(7503):103-8. PubMed Central PMCID: PMC4137318.
 - d. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnson K, Stoddard BL, Baker D. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*. 2013 Sep 12;501(7466):212-216. PubMed Central PMCID: PMC3898436.
5. **Involving the general public in Science.** We created a distributed computing project called Rosetta@home in which volunteers donate spare cycles on their computers to carry out protein folding trajectories. We extended Rosetta@home to the interactive multiplayer online game Foldit which allows players to guide the course of the protein structure prediction and design calculations. By relying on human intuition and 3-D problem solving skills, Foldit players have made a number of important contributions: solved the structure a retroviral protease, developed new algorithms for finding low-energy protein conformations, and designed a novel synthetic enzyme by large-scale redesign of the active site.
- a. Koepnick B, Flatten J, Husain T, Ford A, Silva DA, Bick MJ, Bauer A, Liu G, Ishida Y, Boykov A, Estep RD, Kleinfelter S, Nørgård-Solano T, Wei L, Players F, Montelione GT, DiMaio F, Popović Z, Khatib F, Cooper S, Baker D. De novo protein design by citizen scientists. *Nature*. 2019 Jun;570(7761):390-394. PubMed Central PMCID: PMC6701466.
 - b. Eiben CB, Siegel JB, Bale JB, Cooper S, Khatib F, Shen BW, Players F, Stoddard BL, Popovic Z, Baker D. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat Biotechnol*. 2012 Jan 22;30(2):190-2. PubMed Central PMCID: PMC3566767.
 - c. Khatib F, Cooper S, Tyka MD, Xu K, Makedon I, Popovic Z, Baker D, Players F. Algorithm discovery by protein folding game players. *Proc Natl Acad Sci U S A*. 2011 Nov 22;108(47):18949-53. PubMed Central PMCID: PMC3223433.
 - d. Khatib F, DiMaio F, Cooper S, Kazmierczyk M, Gilski M, Krzywda S, Zabranska H, Pichova I, Thompson J, Popović Z, Jaskolski M, Baker D. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat Struct Mol Biol*. 2011 Sep 18;18(10):1175-7. PubMed Central PMCID: PMC3705907.

Complete List of Published Work:

<https://www.ncbi.nlm.nih.gov/myncbi/david.baker.1/bibliography/public/>