

# 盘点 大牛 生信课题组



*From beginner to expert*

## David Baker

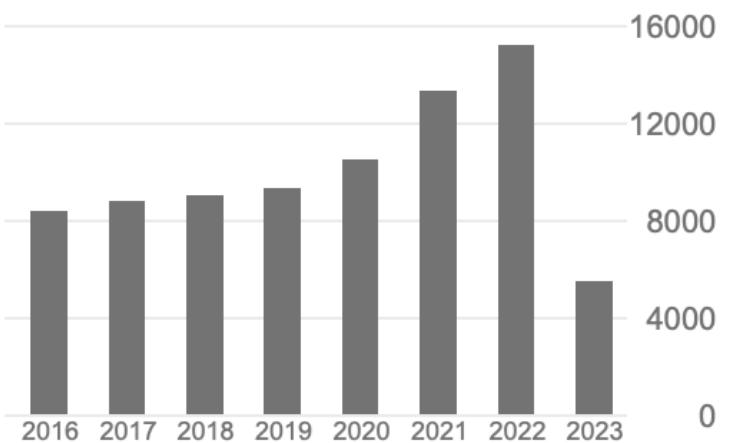


Cited by

[VIEW ALL](#)

All Since 2018

	All	Since 2018
Citations	147516	63219
h-index	204	124
i10-index	925	706



## 2. Baker lab at University of Washington



**David Baker, PhD**

Director, Institute for Protein Design

Investigator, Howard Hughes Medical Institute

Henrietta and Aubrey Davis Endowed Professor of Biochemistry, University of Washington

Adjunct Professor, Genome Sciences, Bioengineering, Chemical Engineering, Computer Science, and Physics,  
University of Washington

[Faculty Page](#)

[dabaker@uw.edu](mailto:dabaker@uw.edu)

[\(assistant\)](mailto:ipdadmin@uw.edu)

# Baker Lab

Institute for Protein Design  
University of Washington

## Welcome!

We develop protein design software and use it to create molecules that solve modern challenges in medicine, technology and sustainability. By iterating between computation and laboratory experiments, we continually improve our protein design methods.

We're always looking for new people to [join our efforts](#). Our philosophy is that the more closely interacting a research group is, the better the science and the more fun overall.

## Research Areas

deep learning

minibinders

enzymes

peptides

molecular devices

drug delivery

nanomaterials

membrane proteins

hybrid materials

structure prediction

## 人员构成：

Admin: 3

Postdoc: 54

Research scientist: 12

Graduate students: 44

Visiting scholar: 6

## 近五年文章：

CNS: 32

大子刊：4

小子刊：26

其他：11

# 研究方向一：蛋白质设计 (Protein design)

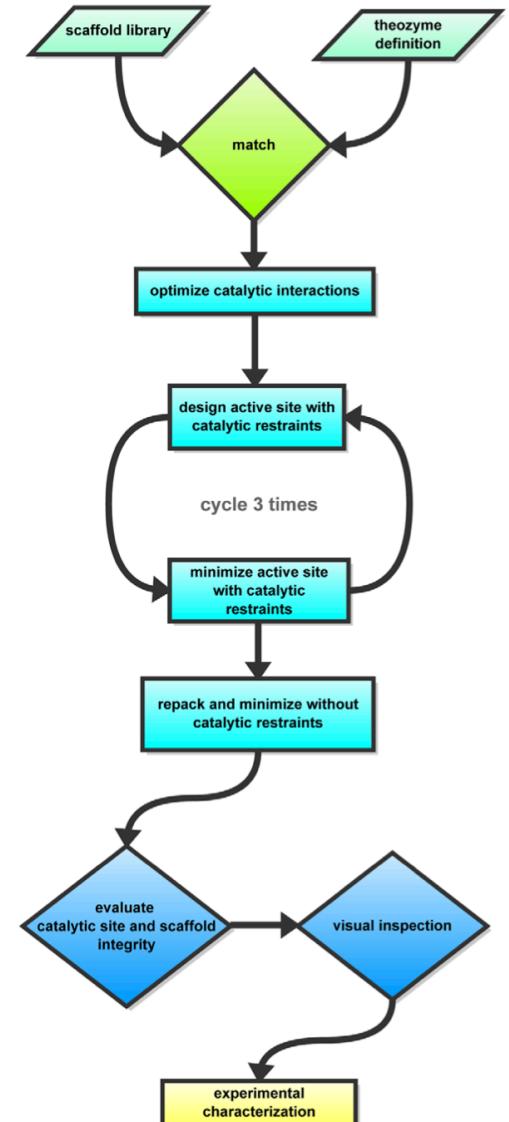
## Rosetta

*Model, design, and analyze protein structures*

The Rosetta software suite includes algorithms for computational modeling and analysis of protein structures. Rosetta has enabled notable scientific advances in computational biology, including de novo protein design, enzyme design, ligand docking, and structure prediction of biological macromolecules and macromolecular complexes.

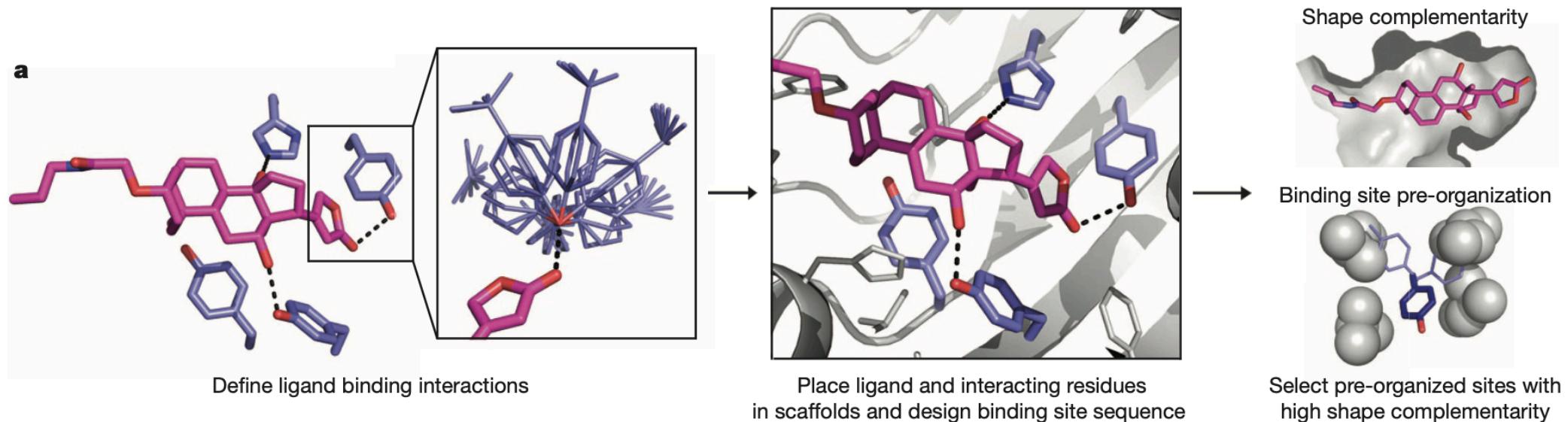
Rosetta began in the laboratory of Dr. David Baker at the University of Washington as a structure prediction tool but has since been adapted to solve common computational macromolecular problems. Development of Rosetta now happens among the members of RosettaCommons, which include government laboratories, institutes, research centers, and partner corporations.

Rosetta is available to all non-commercial users for free and to commercial users for a fee. Visit [rosettacommons.org](http://rosettacommons.org) to get started.



## Computational design of ligand-binding proteins with high affinity and selectivity

Christine E. Tinberg<sup>1\*</sup>, Sagar D. Khare<sup>1†\*</sup>, Jiayi Dou<sup>2,3</sup>, Lindsey Doyle<sup>4</sup>, Jorgen W. Nelson<sup>5</sup>, Alberto Schena<sup>6</sup>, Wojciech Jankowski<sup>7</sup>, Charalampos G. Kalodimos<sup>7</sup>, Kai Johnsson<sup>6</sup>, Barry L. Stoddard<sup>4</sup> & David Baker<sup>1,8</sup>



Article | Published: 27 January 2021

# De novo design of modular and tunable protein biosensors

Alfredo Quijano-Rubio, Hsien-Wei Yeh, Jooyoung Park, Hansol Lee, Robert A. Langan, Scott E. Boyken, Marc J. Lajoie, Longxing Cao, Cameron M. Chow, Marcos C. Miranda, Jimin Wi, Hyo Jeong Hong, Lance Stewart, Byung-Ha Oh✉ & David Baker✉

Nature 591, 482–487 (2021) | Cite this article

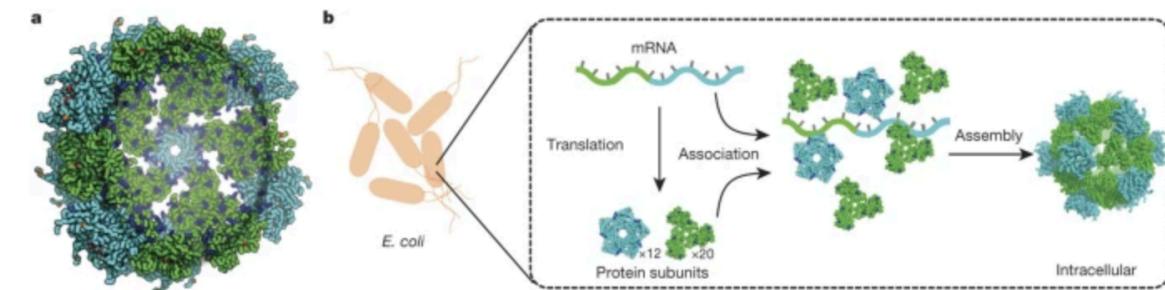
57k Accesses | 89 Citations | 378 Altmetric | Metrics

## LETTER

doi:10.1038/nature25157

### Evolution of a designed protein assembly encapsulating its own RNA genome

Gabriel L. Butterfield<sup>1,2,3\*</sup>, Marc J. Lajoie<sup>1,2\*</sup>, Heather H. Gustafson<sup>4,5</sup>, Drew L. Sellers<sup>4,5,6</sup>, Una Nattermann<sup>1,2,7</sup>, Daniel Ellis<sup>1,2,3</sup>, Jacob B. Bale<sup>1,2,3</sup>, Sharon Ke<sup>4</sup>, Garreck H. Lenz<sup>8</sup>, Angelica Yehdego<sup>9</sup>, Rashmi Ravichandran<sup>1,2</sup>, Suzie H. Pun<sup>4,5</sup>, Neil P. King<sup>1,2</sup> & David Baker<sup>1,2,10</sup>



# 研究方向二：蛋白质结构预测 (Protein structure prediction)

Science

RESEARCH ARTICLES

Cite as: M. Baek *et al.*, *Science* 10.1126/science.abj8754 (2021).

## Accurate prediction of protein structures and interactions using a three-track neural network

Minkyung Baek<sup>1,2</sup>, Frank DiMaio<sup>1,2</sup>, Ivan Anishchenko<sup>1,2</sup>, Justas Dauparas<sup>1,2</sup>, Sergey Ovchinnikov<sup>3,4</sup>, Gyu Rie Lee<sup>1,2</sup>, Jue Wang<sup>1,2</sup>, Qian Cong<sup>5,6</sup>, Lisa N. Kinch<sup>7</sup>, R. Dustin Schaeffer<sup>6</sup>, Claudia Millán<sup>8</sup>, Hahnbeom Park<sup>1,2</sup>, Carson Adams<sup>1,2</sup>, Caleb R. Glassman<sup>9,10</sup>, Andy DeGiovanni<sup>12</sup>, Jose H. Pereira<sup>12</sup>, Andria V. Rodrigues<sup>12</sup>, Alberdina A. van Dijk<sup>13</sup>, Ana C. Ebrecht<sup>13</sup>, Diederik J. Opperman<sup>14</sup>, Theo Sagmeister<sup>15</sup>, Christoph Buhlheller<sup>15,16</sup>, Tea Pavkov-Keller<sup>15,17</sup>, Manoj K. Rathinamswamy<sup>18</sup>, Udit Dalwadi<sup>19</sup>, Calvin K. Yip<sup>19</sup>, John E. Burke<sup>18</sup>, K. Christopher Garcia<sup>9,10,11,20</sup>, Nick V. Grishin<sup>6,21,7</sup>, Paul D. Adams<sup>12,22</sup>, Randy J. Read<sup>8</sup>, David Baker<sup>1,2,23\*</sup>

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. <sup>2</sup>Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. <sup>3</sup>Faculty of Arts and Sciences, Division of Science, Harvard University, Cambridge, MA 02138, USA. <sup>4</sup>John Harvard Distinguished Science Fellowship Program, Harvard University, Cambridge, MA 02138, USA. <sup>5</sup>Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX, USA.

<sup>6</sup>Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>7</sup>Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>8</sup>Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK. <sup>9</sup>Program in Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>10</sup>Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>11</sup>Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>12</sup>Molecular Biophysics & Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>13</sup>Department of Biochemistry, Focus Area Human Metabolomics, North-West University, 2531 Potchefstroom, South Africa. <sup>14</sup>Department of Biotechnology, University of the Free State, 205 Nelson Mandela Drive, Bloemfontein 9300, South Africa. <sup>15</sup>Institute of Molecular Biosciences, University of Graz, Humboldtstrasse 50, 8010 Graz, Austria. <sup>16</sup>Medical University of Graz, Graz, Austria. <sup>17</sup>BioTechMed-Graz, Graz, Austria. <sup>18</sup>Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada. <sup>19</sup>Life Sciences Institute, Department of Biochemistry and Molecular Biology, The University of British Columbia, Vancouver, BC, Canada. <sup>20</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>21</sup>Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>22</sup>Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA. <sup>23</sup>Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.

\*Corresponding author. Email: dabaker@uw.edu

DeepMind presented remarkably accurate predictions at the recent CASP14 protein structure prediction assessment conference. We explored network architectures incorporating related ideas and obtained the best performance with a three-track network in which information at the 1D sequence level, the 2D distance map level, and the 3D coordinate level is successively transformed and integrated. The three-track network produces structure predictions with accuracies approaching those of DeepMind in CASP14, enables the rapid solution of challenging X-ray crystallography and cryo-EM structure modeling problems, and provides insights into the functions of proteins of currently unknown structure. The network also enables rapid generation of accurate protein-protein complex models from sequence information alone, short circuiting traditional approaches which require modeling of individual subunits followed by docking. We make the method available to the scientific community to speed biological research.

RESEARCH ARTICLE

STRUCTURE PREDICTION

## Computed structures of core eukaryotic protein complexes

Ian R. Humphreys<sup>1,2†</sup>, Jimin Pei<sup>3,4†</sup>, Minkyung Baek<sup>1,2†</sup>, Aditya Krishnakumar<sup>1,2†</sup>, Ivan Anishchenko<sup>1,2</sup>, Sergey Ovchinnikov<sup>5,6</sup>, Jing Zhang<sup>3,4</sup>, Travis J. Ness<sup>7†</sup>, Sudeep Banjade<sup>8</sup>, Saket R. Bagde<sup>8</sup>, Viktoriya G. Stancheva<sup>9</sup>, Xiao-Han Li<sup>9</sup>, Kaixian Liu<sup>10</sup>, Zhi Zheng<sup>10,11</sup>, Daniel J. Barrero<sup>12</sup>, Upasana Roy<sup>13</sup>, Jochen Kuper<sup>14</sup>, Israel S. Fernández<sup>15</sup>, Barnabas Szakal<sup>16</sup>, Dana Branzei<sup>16,17</sup>, Josep Rizo<sup>4,18,19</sup>, Caroline Kisker<sup>14</sup>, Eric C. Greene<sup>13</sup>, Sue Biggins<sup>12</sup>, Scott Keeney<sup>10,11,20</sup>, Elizabeth A. Miller<sup>9</sup>, J. Christopher Fromme<sup>8</sup>, Tamara L. Hendrickson<sup>7</sup>, Qian Cong<sup>3,4,\*§</sup>, David Baker<sup>1,2,21,\*§</sup>

Protein-protein interactions play critical roles in biology, but the structures of many eukaryotic protein complexes are unknown, and there are likely many interactions not yet identified. We take advantage of advances in proteome-wide amino acid coevolution analysis and deep-learning-based structure modeling to systematically identify and build accurate models of core eukaryotic protein complexes within the *Saccharomyces cerevisiae* proteome. We use a combination of RoseTTAFold and AlphaFold to screen through paired multiple sequence alignments for 8.3 million pairs of yeast proteins, identify 1505 likely to interact, and build structure models for 106 previously unidentified assemblies and 806 that have not been structurally characterized. These complexes, which have as many as five subunits, play roles in almost all key processes in eukaryotic cells and provide broad insights into biological function.

**RESEARCH ARTICLE****PROTEIN DESIGN**

# Scaffolding protein functional sites using deep learning

Jue Wang<sup>1,2†</sup>, Sidney Lisanza<sup>1,2,3†</sup>, David Juergens<sup>1,2,4†</sup>, Doug Tischer<sup>1,2†</sup>, Joseph L. Watson<sup>1,2†</sup>, Karla M. Castro<sup>5</sup>, Robert Ragotte<sup>1,2</sup>, Amijai Saragovi<sup>1,2</sup>, Lukas F. Milles<sup>1,2</sup>, Minkyung Baek<sup>1,2</sup>, Ivan Anishchenko<sup>1,2</sup>, Wei Yang<sup>1,2</sup>, Derrick R. Hicks<sup>1,2</sup>, Marc Expòsit<sup>1,2,4</sup>, Thomas Schlichthaerle<sup>1,2</sup>, Jung-Ho Chun<sup>1,2,3</sup>, Justas Dauparas<sup>1,2</sup>, Nathaniel Bennett<sup>1,2,4</sup>, Basile I. M. Wicky<sup>1,2</sup>, Andrew Muenks<sup>1,2</sup>, Frank DiMaio<sup>1,2</sup>, Bruno Correia<sup>5</sup>, Sergey Ovchinnikov<sup>6,7\*</sup>, David Baker<sup>1,2,8\*</sup>

The binding and catalytic functions of proteins are generally mediated by a small number of functional residues held in place by the overall protein structure. Here, we describe deep learning approaches for scaffolding such functional sites without needing to prespecify the fold or secondary structure of the scaffold. The first approach, “constrained hallucination,” optimizes sequences such that their predicted structures contain the desired functional site. The second approach, “inpainting,” starts from the functional site and fills in additional sequence and structure to create a viable protein scaffold in a single forward pass through a specifically trained RoseTTAFold network. We use these two methods to design candidate immunogens, receptor traps, metalloproteins, enzymes, and protein-binding proteins and validate the designs using a combination of *in silico* and experimental tests.

**Science****RESEARCH ARTICLES**

Cite as: J. Dauparas *et al.*, *Science* 10.1126/science.add2187 (2022).

# Robust deep learning-based protein sequence design using ProteinMPNN

J. Dauparas<sup>1,2</sup>, I. Anishchenko<sup>1,2</sup>, N. Bennett<sup>1,2,3</sup>, H. Bai<sup>1,2,4</sup>, R. J. Ragotte<sup>1,2</sup>, L. F. Milles<sup>1,2</sup>, B. I. M. Wicky<sup>1,2</sup>, A. Courbet<sup>1,2,4</sup>, R. J. de Haas<sup>5</sup>, N. Bethel<sup>1,2,4</sup>, P. J. Y. Leung<sup>1,2,3</sup>, T. F. Huddy<sup>1,2</sup>, S. Pellock<sup>1,2</sup>, D. Tischer<sup>1,2</sup>, F. Chan<sup>1,2</sup>, B. Koepnick<sup>1,2</sup>, H. Nguyen<sup>1,2</sup>, A. Kang<sup>1,2</sup>, B. Sankaran<sup>6</sup>, A. K. Bera<sup>1,2</sup>, N. P. King<sup>1,2</sup>, D. Baker<sup>1,2,4\*</sup>

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, WA, USA. <sup>2</sup>Institute for Protein Design, University of Washington, Seattle, WA, USA. <sup>3</sup>Molecular Engineering Graduate Program, University of Washington, Seattle, WA, USA. <sup>4</sup>Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. <sup>5</sup>Department of Physical Chemistry and Soft Matter, Wageningen University and Research, Wageningen, Netherlands. <sup>6</sup>Berkeley Center for Structural Biology, Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley Laboratory, Berkeley, CA, USA.

\*Corresponding author. Email: dabaker@uw.edu

While deep learning has revolutionized protein structure prediction, almost all experimentally characterized *de novo* protein designs have been generated using physically based approaches such as Rosetta. Here we describe a deep learning-based protein sequence design method, ProteinMPNN, with outstanding performance in both *in silico* and experimental tests. On native protein backbones, ProteinMPNN has a sequence recovery of 52.4%, compared to 32.9% for Rosetta. The amino acid sequence at different positions can be coupled between single or multiple chains, enabling application to a wide range of current protein design challenges. We demonstrate the broad utility and high accuracy of ProteinMPNN using X-ray crystallography, cryoEM and functional studies by rescuing previously failed designs, made using Rosetta or AlphaFold, of protein monomers, cyclic homo-oligomers, tetrahedral nanoparticles, and target binding proteins.