

**Burkhard Rost**  
**Columbia University Bioinformatics Center**  
**Department of Biochemistry and Molecular Biophysics**  
**650 West, 168<sup>th</sup> Street, BB217**  
**New York, N.Y. 10032**  
**USA**



**Email** [rost@columbia.edu](mailto:rost@columbia.edu)  
**Web** <http://cubic.bioc.columbia.edu>

**Tel** +1 (212) 305-4018  
**Fax** +1 (212) 305-7932

**Document:**

**CV Burkhard Rost**

**Columbia University**

**Contents:**

- Tabulated curriculum vitae
- Grants
- List of publications

## Tabulated curriculum vitae

**Date of preparation:** April 1, 2004  
**Name:** Burkhard Rost  
**Gender:** Male  
**Citizenship:** German  
**Birthplace:** Northeim, Lower-Saxony, Germany  
**Current position:** Associate Professor  
**Address:** Columbia University  
 Department of Biochemistry and Molecular Biophysics  
 650 West 168<sup>th</sup> Street, BB217  
 New York, N.Y. 10032  
 +1-212-305-4018 (fax: 7932), rost@columbia.edu

### Education

1971-1980 High school, Herzberg, Lower-Saxony, Germany  
 1980-1982 Obligatory service in German air force  
 1982-1985 Study of physics at Justus-Liebig-University Gießen, Hessa, Germany  
 10/1984 **Bachelor** in physics  
 1985-1988 Study of physics at Ruprecht-Karls-University Heidelberg, Germany  
 1985-1988 Study of philosophy, history and psychology at Heidelberg University  
 1986-1988 Master thesis at the Institute of Theoretical Physics, Heidelberg Univ.  
*Topic:* 'Learning algorithms for spin-glass-like neural networks'  
*Sponsor:* Prof. Dr. Heinz Horner, Heidelberg  
 11/1988 **Bachelor** in philosophy and history, Heidelberg University  
 12/1988 **Master** in physics, Heidelberg University  
 12/88-6/90 Grant of the 'Stiftung Volkswagenwerk', Inst. Theor. Physics, Heidelberg  
*Topic* 'Theoretical analysis of the possibilities of seismic and acoustical sensor networks to verify arms control treaties for aircraft'  
 7-10/89 Research project in the USA (Princeton, Washington DC, MIT)  
 1990-1993 Completing studies in physics, Heidelberg Univ. and EMBL, Germany  
 7/1993 Writing PhD thesis  
*Topic:* 'Neural networks and evolution - prediction of protein secondary structure'  
*Sponsor:* none  $\diamond$   
 7/1994 **PhD** in physics: Viva voce for Dr. rer. nat. (Doctor rerum naturarum) at Inst. for Theor. Physics, Ruprecht-Karl University Heidelberg, Germany

### Professional Positions

1986-1988 Assistant at Institute for Theoretical Physics, Heidelberg Univ., Germany  
 12/88-6/90 Research fellow at Inst. for Theoretical Physics, Heidelberg Univ.  
 7-10/1989 Visitor at Institutes in the USA (Princeton, Union of Concerned Scientists, MIT)  
 7/90- 1992 Visitor at EMBL Heidelberg, Germany  
 7/93-12/94 Research fellow at EMBL Heidelberg, Germany  
 1/95-12/95 Research fellow at EBI Hinxton, Cambridge, England

$\diamond$  Over 600 years, Heidelberg University allowed scholars to present a doctoral thesis on any subject of research independently of its presence in any of the institutes of the university and independently of any sponsor. Apparently, I was the first who exercised this right over at least the last 60 years.

1/96- 4/98 Research fellow at EMBL Heidelberg, Germany  
 5/98-11/98 Researcher at LION Heidelberg, Germany  
 12/98- 5/00 Assistant Professor at Columbia University, Department of Biochemistry  
 and Molecular Biophysics, New York, USA  
 since 6/00 Associate Professor at Columbia University

### Professional Organizations and Societies

1995-now Program Committee of ISMB meetings (Intelligent Systems for Molecular  
 Biology)  
 1996-now Member of ISCB (International Society for Computational Biology)  
 2002-now Board of Directors ISCB (International Society for Computational Biology)  
 2002-now Member of NYAS (New York Academy of Sciences)

### Professional Experiences

1992-now Over 250 referee reports for peer-reviewed journals (including Nature,  
 Science, Cell, EMBO J)  
 1995-now Over 40 referee reports on grants from individuals and institutes (countries:  
 Austria, Canada, Denmark, England, Israel, Italy, Norway, Netherlands,  
 Spain, Sweden, Switzerland, USA)  
 2001-2002 Ad-hoc panels for NIH grants  
 2001-now Editor Journal of Medical Informatics  
 2002-now Organisation of New York Computational Biology Society (NYCBS) in  
 collaboration with Terry Gaasterland (Rockefeller University)  
 1999-now SAB for company LION Biosciences, Heidelberg/Cambridge  
 UK/Cambridge USA/Cleveland/San Diego

### Invited Talks (92 invited talks in 16 countries)

12/1988 London, England: Workshop on Arms Control  
 9/1989 Washington, DC, USA: Union of Concerned Scientists  
 10/1989 Princeton, USA: Institute for Advanced Studies  
 5/1990 Prague, CSFR: Workshop on Arms Control  
 9/1990 Vienna, Austria: Conference on Arms Control  
 10/1990 Mosbach, Germany: Conference on Arms Control  
 6/1992 Elba, Italy: Conference on Neural Networks  
 7/1992 Torino, Italy: Conference on Protein Structure Prediction  
 11/1992 Nijmegen, Netherlands: EU Conference on Sequence Analysis  
 6/1993 Rennes, France: Workshop on Secondary Structure Prediction  
 11/1993 Copenhagen, DK: Conf. on 'Protein Structure by Distance Analysis'  
 1/1994 Maui, Hawaii, USA: HICCS Conference  
 3/1994 Bologna, Italy: Conference on Protein Structures  
 4/1994 Copenhagen, Denmark: CBS Mini-symposium on structure prediction  
 11/1994 Bielefeld, Germany: University  
 11/1994 Copenhagen, DK: Conf. on 'Protein Structure by Distance Analysis'  
 12/1994 Asilomar, USA: Conf. on 'Critical Assessment of Structure Prediction'  
 6/1995 Cambridge, England: ISMB'95 (Tutorial)

- 6/1995 Heidelberg, Germany: University  
10/1995 Bressanone, Italy: Workshop on Protein Structures  
10/1995 IRBM Rome, Italy: Workshop 'Frontiers of protein structure prediction'  
11/1995 Paris, France: Workshop on Protein Structures and Drug Design
- 6/1996 St. Louis, USA: ISMB'96  
8/1996 Madrid, Spain: CNB  
10/1996 Heidelberg, Germany: University  
11/1996 Copenhagen, DK: Conf. on 'Protein Structure by Distance Analysis'  
12/1996 New York, USA: Columbia University  
12/1996 Asilomar, USA: University of San Francisco
- 1/1997 Univ. Jerusalem, Israel: UNESCO workshop on sequence analysis  
2/1997 Basel, Switzerland: CIBA-Geigy  
2/1997 EBI Hinxton, England: Workshop on protein structure prediction  
3/1997 Copenhagen, Denmark: CBS University  
3/1997 New York, USA: Columbia University  
6/1997 Berlin, Germany: Inst. for Theor. Biol.  
6/1997 Chalkidiki, Greece: ISMB'97 (Tutorial)  
7/1997 EBI Hinxton, England: EMBO Workshop on protein sequence analysis  
7/1997 San Sebastian, Spain: Workshop 'Proteins: integration of life's function'  
7/1997 Madrid, Spain: CNB  
9/1997 Skövde, Sweden: Conf. 'Bio-Computing and Emergent Computation'  
9/1997 Vienna, Austria: 'Intern. Conference on Molecular Structural Biology'  
10/1997 IRBM Rome, Italy: Workshop 'Frontiers of protein structure prediction'  
11/1997 Stockholm, Sweden: Pharmacia & Upjohn  
11/1997 Stockholm, Sweden: Karolinska Inst.  
12/1997 Paris, France: Pasteur Inst.  
12/1997 Toulouse, France: INRA Inst.  
12/1997 Toulouse, France: Elf Sanofi  
12/1997 Basel, CH: Symposium 'Bioinformatics: from Exp. to Biol. Knowledge'
- 3/1998 Marseille, France: Cancer Inst., INSERM/19CNRS  
5/1998 Tallberg, Sweden: Conf. 'Annual meeting of Swedish Structural Biology'  
5/1998 Bad Honnef, Germany: Conf. 'Scientific Applications of Neural Nets'  
6/1998 New York, USA: Genome Centre, Workshop  
10/1998 Cambridge, England: Newton Inst. Conf. 'Biomolecular Function and Evolution in the Context of the Genome Project'  
10/1998 EBI, England: Conf. 'Methods for protein structure'
- 7/1999 Havana, Cuba: CIGB EMBO Workshop 'Bioinformatics'  
8/1999 Cambridge, England: EBI Workshop. 'Protein motifs and families in'  
9/1999 Graz, Austria: International Meeting of Austrian Society for Genetics.  
10/1999 Bologna, Italy: Workshop 'Protein sequence analysis in the genome era'  
11/1999 Atlanta, USA: Conf. 'In silico biology: sequence & structure & function'
- 2/2000 Boston, USA: SGI Workshop 'New Technologies for Discovery Research'  
2/2000 New York Structural Biology Society  
8/2000 Troy RPI, USA: Bioinformatics Workshop  
8/2000 Madrid, Spain: Human Genome Workshop  
8/2000 New York University  
10/2000 Heidelberg, Germany: LION's Bioinformatics 2000  
12/2000 Asilomar, USA: Invited session leader and two presentations at the fourth meeting on critical assessment of structure prediction (CASP4)

- 3/2001 Madrid, Spain: Workshop on structural genomics  
 3/2001 CNB Madrid, Spain  
 4/2001 New York, Rockefeller University  
 5/2001 New York, New York Structural Biology Society  
 5/2001 New York, Columbia University, Dept. Biology  
 7/2001 Technical University Copenhagen, Denmark  
 7/2001 Heidelberg University, Germany  
 9/2001 Il Ciocco, Italy: NATO summer school on bioinformatics (Talk, Teacher)  
 9/2001 IRBM Rome, Italy
- 1/2002 Keystone Symposium on Structural Genomics, USA  
 4/2002 Madrid, Spain: Meeting on 'Bioinformatics and Computational Biology'  
 4/2002 Basel, Switzerland: Talk at Biocenter Basel  
 5/2002 Erice, Italy: Talk on predicting protein function at summer school on structural biology  
 6/2002 San Diego, USA: 'Eleventh Annual Bioinformatics and Genome Research'  
 6/2002 Rutgers University, USA: 'Northeast Structural Genomics Consortium Annual Symposium'  
 7/2002 New York Computational Biology Society, New York Academy of Sciences  
 8/2002 Edmonton, Canada: ISMB'2002  
 8/2002 Habanna, Cuba: 'Bioinformatica-Habana'  
 10/2002 Lausanne, Switzerland, University  
 11/2002 Toronto, Canada Institute for Proteomics and Bioinformatics
- 2/2003 New York, Fordham University  
 2/2003 CABM, Rutgers University, Piscataway, USA  
 2/2003 New York, Manhattan College  
 5/2003 New York, City College  
 10/2003 Bethesda, NIH: Keynote at Workshop on 'Comparative modelling'  
 11/2003 Bethesda, NIH: Workshop on 'Target Selection in Structural Genomics'  
 11/2003 Rutgers University, Piscataway, NJ
- 02/2004 Gordon Conference, Ventura CA: Structural, Functional & Evolutionary Genomics

### Organisation of International Scientific Meetings

- 9/1990 Co-organising 'The Second Workshop on Verification of Arms Reductions', 3-5 Sep. 1990, Vienna, Austria  
 2/1997 Organising the workshop 'Methods for protein structure prediction - progress and limitations' for Pharmaceutical Companies, 10-11 Feb., 1997, EBI Hinxton-Cambridge, England  
 8/1999 Organising the workshop 'Protein motifs and families in practice: Is protein function carved into sequence?' Aug 12-13, 1999, EBI Hinxton-Cambridge, England  
 10/1999 Program committee of the 2<sup>nd</sup> Bologna summer school on biotechnology 'Protein sequence analysis in the genome era', Bologna, Italy  
 12/2000 Co-organising CAFASP sessions at the CASP meeting in Asilomar, CA  
 03/2001 Co-organising bioinformatics competition of LION Biosciences, Boston  
 03/2001 Co-organising Juan March workshop on structural genomics in Madrid, Spain

- 02/2002 Program committee of the 3<sup>rd</sup> "International Meeting on Membrane proteins in Bologna, Italy
- 06/2002 Program committee meeting on computational biology Manchester England
- 08/2002 Co-organising 10<sup>th</sup> International Meeting on 'Intelligent Systems in Molecular Biology' in Edmonton Canada (1500 participants; most important meeting in bioinformatics)
- 06/2003 Scientific Advisory Board, Editor, and Program Committee for 11<sup>th</sup> ISMB 2003 in Brisbane, Australia
- 11/2003 Co-organising NIH workshop on 'Target selection for structural genomics'
- 06/2003 Scientific Advisory Board, Editor, and Program Committee for 12<sup>th</sup> ISMB 2004 in Glasgow, Scotland (over 2000 participants expected)
- 12/2004 Co-organising meeting for the Critical Assessment of protein Structure Prediction (CASP6) in Italy (most important meeting in the field of structure prediction)
- 1995-2004 Program Committee for ISMB meeting (Intelligent Systems for Molecular Biology; central meeting in computational biology/bioinformatics)

### Departmental and University Committees

- 06/2003-now Qualifying exam Dept. Biochemistry & Molecular Biophysics, Columbia

### Teaching Experiences

#### *Courses taught at Columbia*

- 1999-2004 Supervision of graduate students and co-organizing class-room work on bioinformatics in course 'Biophysics' taught by Ann McDermott, Chemistry, Columbia (4 courses: 1999, 2001, 2002, 2003)
- 2000-2003 Teaching at Medical Faculty, Columbia; course 'Eukaryotes I' (3 courses: 2000, 2001, 2002, 2003)

#### *Courses taught at meetings/summer-schools*

- 6/1995 Tutorial at ISMB'95 (International Conference on Intelligent Systems for Molecular Biology) in Cambridge, England
- 10/1995 Supervision of post-docs at the workshop 'Frontiers of protein structure prediction' at the IRBM in Rome, Italy
- 3/1997 Supervision of post-docs at a workshop at the CBS in Copenhagen (Denmark)
- 6/1997 Tutorial at ISMB'97 (International Conference on Intelligent Systems for Molecular Biology) in Chalkidiki, Greece
- 7/1997 Supervision of pre- and post-docs at the workshop (EMBO-course) 'Genome sequence analysis' at the EBI in Hinxton, England
- 7/1997 Teacher at the summer school 'Proteins: integration of life's function' in San Sebastian, Spain
- 10/1997 Supervision of post-docs at the 2nd workshop 'Frontiers of protein structure prediction' at the IRBM in Rome, Italy
- 7/1999 Supervision of pre- and post-docs at the workshop 'Bioinformatics' at the CIB in Havana, Cuba
- 10/1997 Supervision of post-docs at the workshop 'Protein sequence analysis in the genome era' at the University of Bologna, Italy
- 9/2001 Teacher at a NATO summer school on bioinformatics in Il Ciocco, Italy
- 8/2002 Supervision of pre- and post-docs at the workshop 'Bioinformatics' at the CIB in Havana, Cuba

*Other courses taught*

|           |  |
|-----------|--|
| 1986-1988 | Organisation of seminars in Theor. Physics, Heidelberg Univ., Germany                                  |
| 1989-1990 | Lecture series on Arms control (I), Heidelberg University  |
| 1990      | Lecture series on Arms control (II), Heidelberg Univ.  |
| 1986-1990 | Courses on Physics and Arms control in extramural schools (Heidelberg, Mannheim, Frankfurt, Darmstadt) |
| 1991-1995 | Lectures, and supervision of courses for PhD students at EMBL Heidelberg                               |
| 1992-1996 | Course lectures on Protein structure prediction at Heidelberg Univ.                                    |

*PhD students who graduated in my group*

| <i>Name</i> | <i>Graduation</i> | <i>Current Affiliation</i> |
|-------------|-------------------|----------------------------|
| Yanay Ofran | 02/2004           | Columbia/Biochemistry      |
| Jinfeng Liu | 02/2004           | Columbia/Biochemistry      |

*PhD students in my group*

| <i>Name</i>        | <i>University/Department</i> | <i>Period</i> |
|--------------------|------------------------------|---------------|
| Rajesh Nair        | Columbia/Physics             | 1999-2003     |
| Dariusz Przybylski | Columbia/Physics             | 1999-2003     |
| Henry Bigelow      | Columbia/Biochemistry        | 2000-2005     |
| Andrew Kernytsky   | Columbia/Biochemistry        | 2001-2006     |
| Sven Mika          | Columbia/Biochemistry        | 2002-2006     |
| Yana Bromberg      | Columbia/Medical Inform.     | 2003-2007     |
| Avner Schlessinger | Columbia/Biochemistry        | 2003-2007     |

*Thesis committee*

| <i>Name</i>        | <i>University/Department</i> | <i>Date</i> | <i>Group</i>      |
|--------------------|------------------------------|-------------|-------------------|
| Volker Eyrich      | Columbia/Chemistry           | 07/2001     | Rich Friesner     |
| Yun Zhang          | Columbia/Biology             | 10/2001     | Marti Chalfie     |
| Yuling An          | Columbia/Chemistry           | 07/2002     | Rich Friesner     |
| Chuck Duarte       | Columbia/Biochemistry        | 09/2002     | Ann-Marie Pyle    |
| Cathy S Gunther    | Rockefeller University       | 11/2002     | Terry Gaasterland |
| Samuel K Handelman | Columbia/Biology             | 10/2003     | John Hunt         |
| Trevor Siggers     | Columbia/Biochemistry        | 2005        | Barry Honig       |
| Oleg Trott         | Columbia/Biochemistry        | 2005        | Arthur Palmer     |
| Cinque Soto        | Columbia/Biochemistry        | 2005        | Barry Honig       |
| Christopher Tang   | Columbia/Biochemistry        | 2006        | Barry Honig       |
| Erroll Rueckert    | Columbia/Integrated          | 2006        | Richard Axel      |
| Marina Gimpelev    | Columbia/Biochemistry        | 2006        | Barry Honig       |
| Chen Peter Chien   | Columbia/MD,PhD              | 2006        | Barry Honig       |
| Murat Cokol        | Columbia/Biology             | 2006        | Andrej Rzhetsky   |

*Rotation students*

| <i>Name</i>      | <i>University/Department</i> | <i>Period</i> |
|------------------|------------------------------|---------------|
| Murat Cokol      | Columbia/Biology             | Spring 2000   |
| Trevor Siggers   | Columbia/Biochemistry        | Summer 2000   |
| Hong Yu          | Columbia/Medical Inform.     | Summer 2000   |
| Hepan Tan        | Columbia/Biochemistry        | Summer 2000   |
| Enrique Tadique  | Bronx Highschool             | Summer 2000   |
| Henry Bigelow    | Columbia/Biochemistry        | Spring 2001   |
| Chen Peter Chien | Columbia/MD-PhD              | Summer 2001   |
| Seth Gale        | Columbia/Undergraduate       | Fall 2001     |

|                      |                          |                  |
|----------------------|--------------------------|------------------|
| Andrew Kernytsky     | Columbia/Biochemistry    | Summer 2002      |
| Avner Schlessinger   | Columbia/Biochemistry    | Summer 2002      |
| Anthony J DeCostanzo | Columbia/Pharmacology    | Fall 2002        |
| Roman Trakhtenberg   | Columbia/Medical Inform. | Winter 2002/2003 |
| Shoshana L Posy      | Columbia/Integrated      | Spring 2003      |
| Shameek Biswas       | Columbia/Computer Sci.   | Summer 2003      |
| Gabor Halasz         | Columbia/Integrated      | Summer 2003      |

### *Visiting graduate and under-graduate students*

| <i>Name</i>       | <i>From</i>              | <i>Period</i>         |
|-------------------|--------------------------|-----------------------|
| Bastian Bruning   | Nijmegen, Netherlands    | Spring 2001           |
| Phil Carter       | Imperial College, London | Fall 2001             |
| Seth Gale         | Columbia                 | Fall 2001             |
| Claus AF Andersen | Copenhagen, Denmark      | Fall 2001/Spring 2002 |
| Jackson Ng        | High school graduate     | Spring/Summer 2000    |
| Enrique Tadique   | High school graduate     | Summer 2000           |

## **Research Support**

### *Grants past*

PI: Burkhard Rost  
 Period : August 2002  
 Agency: NSF  
 Total Dir Costs: \$40 K Overhead 0  
 Title: ISMB'2002  
 Summary: Travel support for young participants of the ISMB 2002 meeting

PI: Burkhard Rost  
 Period : August 2002  
 Agency: DOE  
 Total Dir Costs: \$35 K Overhead 0  
 Title: ISMB'2002  
 Summary: Travel support for young participants of the ISMB 2002 meeting

### *Grants active*

PI: Gaetano T. Montelione (Rutgers University)  
 Co-PI: Burkhard Rost  
 Period: 10/01/00 to 31/10/05  
 Agency: NIH (P50 GM62413-01)  
 Ann Dir Costs: \$180,000 Total (5 years) \$750,000 Overhead (4 years) \$580,000  
 Title: Structural genomics of eukaryotic model organisms  
 Summary: Goal of the pilot project in structural genomics (one of nine in the USA; of about 14 world-wide) to develop high-throughput techniques for large-scale structure determination. This project combines over 20 groups from 12 research institutions in the USA, Canada and Israel. Our particular task is the development of techniques that rationalize the automatic target selection, i.e. prioritize which proteins to pursue experimentally.

PI: Burkhard Rost  
 Period: 05/01/01 to 04/30/05

Agency: NIH (1 R01 GM63029-01)  
 Ann Dir Costs: \$190,000 Total (4 years) \$750,000 Overhead (4 years) \$660,000  
 Title: Intruding into the midnight zone of protein sequence comparisons  
 Summary: The 'twilight zone' of protein sequence comparison is the region in which sequence similarity does not suffice to conclude, e.g., structural similarity. The vast majority of all protein pairs of similar structure populate a 'midnight zone', i.e., their sequences differ too much for sequence-based inferences. We proposed to refine, extend, and specialize methods combining sequence alignment, structure prediction and functional information. Goal is to unravel hidden similarities in entirely sequenced organisms by a reliable, automatic tool. Towards the end of our project, the sequences for most protein families realized by life will supposedly be available.

PI: Burkhard Rost  
 Period: 02/01/02 to 01/31/05  
 Agency: NSF (DBI-0131168)  
 Ann Dir Costs: \$157,000 Total (3 years) \$470,000 Overhead (4 years) \$240,000  
 Title: Ab initio prediction of sub-cellular localization  
 Summary: The major goals of this project are to develop a system predicting the sub-cellular localization of a protein based on sequence alignments and signal peptide motifs. The ultimate objective is to combine a series of novel methods into a comprehensive system that we can use to automatically annotate genomes, in particular those of the fly and the subset of known human proteins. We expect to have our first comprehensive system for the major compartments available by fall 2003.

PI: Burkhard Rost  
 Period: 05/01/03 – 04/30/07  
 Agency: NIH/NLM (1 R01 LM07329-01 A1)  
 Ann Dir Costs: \$175,000 Total (4 years) \$900,000 Overhead (4 years) \$770,000  
 Title: Improve predictions of structure and function by PredictProtein  
 Summary: The major objective is to improve the internet prediction server PredictProtein. The particular goals are to (1) implement technical improvements, (2) and to systematically combine the results from various methods.

PI: Burkhard Rost  
 Period: 05/01/03 – 04/30/07  
 Agency: NIH (1 R01 GM64633-01 A2)  
 Ann Dir Costs: \$200,000 Total (4 years) \$800,000 Overhead (4 years) \$680,000  
 Title: Predict putative protein-protein interface segments at low resolution  
 Summary: Here, we propose to develop methods predicting interface segments, i.e. regions of residues consecutive in sequence that are in contact with other interface segments. Separate methods address internal and external interfaces. The basic means explored will be combinations of statistics and neural networks using evolutionary information as contained in multiple sequence alignments. The goal is a low-resolution prediction succeeding often enough to distinguish between internal and external interfaces to assist the design of experiments in molecular and medical biology. We focus on predicting interfaces from sequence.

PI: Barry Honig (Biochemistry, Columbia)  
 Co-PI: Burkhard Rost

Submitted: 03/01/01  
 Period: 09/01/01 to 08/31/03  
 Agency: NIH (5-P20-LM7276)  
 Ann Dir Costs: \$20,000 Overhead 0  
 Summary: Grant supporting to prepare submitting a grant proposal to begin a center of excellence.

### *Grants pending*

PI: Burkhard Rost  
 Submitted: 12/15/03  
 Period: 07/01/04 – 06/30/09  
 Agency: NIH (1-R13-GM072354-01)  
 Ann Dir Costs: \$45,000 Total (5 years) \$225,000  
 Title: Meeting for Critical Assessment of protein Structure Prediction (CASP)  
 Summary: This grant will cover the funding needed for the next 3 CASP meetings that have evolved to the major event in the structure prediction field.

PI: Barry Honig (Biochemistry, Columbia)  
 Project leader: Burkhard Rost  
 Submitted: 01/22/04  
 Period: 09/01/04 to 08/31/09  
 Agency: NIH ()  
 Ann Dir Costs: \$200,000 Overhead \$120,000  
 Summary: Grant proposal to begin a national center of excellence.

### *Total budget of group*

|   |             |
|---|-------------|
| Estimated Total Budget from 1999-future, i.e. covered by grants | \$5,844,000 |
| Estimated Total Revenue for Columbia University 1999-future     | \$3,644,534 |
| 2002: Total Budget  | \$982,224   |
| 2002: Salaries  | \$625,216   |
| 2002: Equipment, etc. (estimate)                                | \$357,008   |
| 2002: Estimated total overhead for Columbia University          | \$331,612   |

## List of publications $\diamond$

### Theses

- A. **B Rost** (1988) Lernalgorithmen für verdünnte Spin-glas-artige Neuronale Netzwerke (Learning algorithms for spin-glass-like neural networks). **Master Thesis**, Inst. for Theoretical Physics, Heidelberg University, Germany.
- B. **B Rost** (1993) Neural networks and evolution - advanced prediction of protein secondary structure. **PhD Thesis**, Dep. of Physics and Astronomy, Heidelberg Univ.

### Publications on Arms Control

- \* 1. **B Rost** (1990) Report on conventional weapons. In 'Verification and arms control implications for European security', HG Brauch (ed.) Mosbach, Germany: AFES Press, 120-122.
- \* 2. **B Rost** (1991) Sensor Networks for Verifying Disarmament of Aircraft. Bochum: UVB Universitätsverlag Brockmeyer, Book.
- \* 3. **B Rost** (1992) Aircraft Monitoring by Automatic Sensor Networks. In 'Verification at Vienna. Monitoring Reductions of Conventional Armed Forces', J Altmann, H Graaf, PM Lewis, and P Markl (eds.). Philadelphia et al.: Gordon & Breach, 179-187.
4. JLA Altmann, W Baus, A DeVolpi, B Gonsior, J Grin, J Hanousek, V Journé, W Kaiser, J Klinger, P Lewis, J Málek, J Matousek, M Pospisil, **B Rost**, V Rudajev, I Sabo & P Stein (1993) Ground Vibration, Acoustic Waves and Magnetic Disturbances Produced by Land Vehicles of the North-Atlantic Treaty Organization - Results of the 1989 Measurements of Baumholder, FRG. Bochum: UVB Universitätsverlag Brockmeyer.

### Publications on Computational Biology and Bioinformatics

- \* 5. **B Rost** & Sander (1992) Jury returns on structure prediction. **Nature** 360, 540.
- \* 6. **B Rost** & C Sander (1992) Exercising multi-layered networks on protein secondary structure. In 'Neural Networks: From Biology to High Energy Physics', O Benhar, S Brunak, P DelGiudice, and M Grandolfo (eds.) Elba, Italy: International Journal of Neural Systems, 209-220.
7. T Meitinger, A Meindl, P Bork, **B Rost**, C Sander, M Haasemann & J Murken (1993) Molecular modelling of the Norrie disease protein predicts a cysteine knot growth factor tertiary structure. **Nature Genetics** 5, 376-380.
- \* 8. **B Rost** & C Sander (1993) Prediction of protein secondary structure at better than 70% accuracy. **J Mol Biol** 232, 584-599.
- \* 9. **B Rost** & C Sander (1993) Improved prediction of protein secondary structure by use of sequence profiles and neural networks. **PNAS** 90, 7558-7562.
- \* 10. **B Rost** & C Sander (1993) Secondary structure prediction of all-helical proteins in two states. **Protein Engineering** 6, 831-836.
- \* 11. **B Rost**, C Sander and R Schneider (1993) Progress in protein structure prediction? **TIBS** 18, 120-123.
- \* 12. B Rost & G Vriend (1993) Neural networks in chemistry. **CDA News** 8, 24-27.

$\diamond$  Group members in bold letters; journals underlined; asterisk marks first/senior author publications.

13. L Holm, B Rost, C Sander, R Schneider & G Vriend (1994) Data based modeling of proteins. In 'Statistical Mechanics, Protein Structure, and Protein Substrate Interactions', Doniach, S (eds.) New York: Plenum Press, 277-296.
- \* 14. **B Rost & C Sander (1994) Structure prediction of proteins - where are we now? *Cur Opin Biotech* 5, 372-380.**
- \* 15. **B Rost & C Sander (1994) 1D secondary structure prediction through evolutionary profiles. In 'Protein Structure by Distance Analysis', H Bohr and S Brunak (eds.) Amsterdam, Oxford, Washington: IOS Press, 257-276.**
- \* 16. **B Rost & C Sander (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 19, 55-72.**
- \* 17. **B Rost & C Sander (1994) Conservation and prediction of solvent accessibility in protein families. *Proteins* 20, 216-226.**
- \* 18. **B Rost, C Sander & R Schneider (1994) Evolution and neural networks - protein secondary structure prediction above 71% accuracy. In '27th Hawaii International Conference on System Sciences', L Hunter (ed.) Wailea, Hawaii; Los Alamitos, CA: IEEE Society Press, 385-394.**
- \* 19. **B Rost, C Sander & R Schneider (1994) PHD - an automatic server for protein secondary structure prediction. *CABIOS* 10, 53-60.**
- \* 20. **B Rost, C Sander & R Schneider (1994) Redefining the goals of protein secondary structure prediction. *J Mol Biol* 235, 13-26.**
- \* 21. **B Rost (1995) Fitting 1-D predictions into 3-D structures. In 'Protein folds: a distance based approach', H Bohr and S Brunak (eds.) Boca Raton, FL: CRC Press, 132-151.**
- \* 22. **B Rost (1995) TOPITS: Threading One-dimensional Predictions Into Three-dimensional Structures. In 'Third International Conference on Intelligent Systems for Molecular Biology', C Rawlings, D Clark, R Altman, L Hunter, T Lengauer, and S Wodak (eds.) Cambridge, England: Menlo Park, CA: AAAI Press, 314-321.**
- \* 23. **B Rost, R Casadio, P Fariselli & C Sander (1995) Prediction of helical transmembrane segments at 95% accuracy. *Protein Science* 4, 521-533.**
- \* 24. **B Rost & C Sander (1995) Progress of 1D protein structure prediction at last. *Proteins* 23, 295-300.**
- \* 25. **B Rost & C Sander (1995) Protein structure prediction by neural networks. In 'The handbook of brain theory and neural networks', M Arbib (ed.) Cambridge, MA: Bradford Books/The MIT Press, 772-775.**
26. T Hubbard, A Tramontano, G Barton, D Jones, M Sippl, A Valencia, A Lesk, J Moulton, **B Rost**, C Sander, R Schneider, A Lahm, R Leplae, C Buta, M Eisenstein, O Fjellström, H Floeckner, JG Grossmann, J Hansen, M Helmer-Citterich, FS Joergensen, A Marchler-Bauer, J Osuna, J Park, A Reinhardt, L Ribas de Pouplana, A Rojo-Dominguez, V Saudek, J Sinclair, S Sturrock, C Venclovas & C Vinals (1996) Update on protein structure prediction: results of the 1995 IRBM workshop. ***Folding & Design* 1, R55-R63.**
- \* 27. **B Rost (1996) PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods in Enzymology* 266, 525-539.**
- \* 28. **B Rost, R Casadio & P Fariselli (1996) Refining neural network predictions for helical transmembrane proteins by dynamic programming. In 'Fourth International Conference on Intelligent Systems for Molecular Biology', D States, P Agarwal, T Gaasterland, L Hunter, and RF Smith (eds.) St. Louis, Menlo Park: AAAI Press, 192-200.**
- \* 29. **B Rost, R Casadio & P Fariselli (1996) Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Science* 5, 1704-1718.**

- \* 30. **B Rost** & C Sander (1996) Bridging the protein sequence-structure gap by structure predictions. **Ann Rev Biophys and Biomol Structure** 25, 113-136.
- \* 31. **B Rost** & A Valencia (1996) Pitfalls of protein sequence analysis. **Cur Op Biotech** 7, 457-461.
- \* 32. **B Rost** (1997) Learning from evolution to predict protein structure. In 'BCEC97: Bio-Computing and Emergent Computation', B Olsson, D Lundh, and A Narayanan (eds.) Skövde, Sweden: World Scientific, 87-101.
- \* 33. **B Rost** (1997) NN which predicts protein secondary structure. In 'Handbook of Neural Computation', E Fiesler and R Beale (eds.) New York: Oxford Univ. Press, G4.1.
- \* 34. **B Rost** (1997) Protein structures sustain evolutionary drift. **Folding & Design** 2, S19-S24.
- \* 35. **B Rost** & SI O'Donoghue (1997) Sisyphus and prediction of protein structure. **Bioinformatics** 13, 345-356.
- \* 36. **B Rost**, R Schneider and C Sander (1997) Protein fold recognition by prediction-based threading. **J Mol Biol** 270, 471-480.
- \* 37. MA Andrade, SI O'Donoghue & **B Rost** (1998) Adaptation of protein surfaces to subcellular location. **J Mol Biol** 276, 517-525.
- \* 38. **B Rost** (1998) Better 1D predictions by experts with machines. **Proteins Suppl.** 1, 192-197.
- \* 39. **B Rost** (1998) Marrying structure and genomics. **Structure** 6, 259-263.
- \* 40. **B Rost** (1998) Protein structure prediction in 1D, 2D, 3D. In 'The Encyclopaedia of computational chemistry' P von Ragué Schleyer, NL Allinger, T Clark, J Gasteiger, PA Kollman, HF Schaefer III, and PR Schreiner (eds.) Chichester: Wiley, 2242-2255.
- \* 41. R Zidovetzki, **B Rost** & I Pecht, (1998) The role of transmembrane domains in the functions of B- and T-cell receptors. **Immunology Lett** 64, 97-107.
- \* 42. **B Rost** (1999) Twilight zone of protein sequence alignments. **Protein Engineering** 12, 85-94.
- \* 43. A Zemla, C Venclovas, K Fidelis & **B Rost** (1999) A modified definition of SOV, a segment-based measure for protein secondary structure prediction assessment. **Proteins** 34, 220-223.
- \* 44. **B Rost** (1999) Evolution teaches neural networks. In 'Scientific applications of neural nets', JW Clark, T Lindenau, and ML Ristig (eds.) Heidelberg: Springer, 207-223.
- 45. O Olmea, **B Rost**. & A Valencia (1999) Effective use of sequence correlation and conservation in fold recognition. **J Mol Biol** 293, 1221-1231.
- 46. D Fischer, C Barret, K Bryson, A Elofsson, A Godzik, D Jones, KJ Karplus, LA Kelley, RM MacCallum, K Pawowski, **B Rost**, L Rychlewski & M Sternberg (1999) CAFASP-1: critical assessment of fully automated structure prediction methods. **Proteins Suppl** 3, 209-21.
- 47. F Pazos, **B Rost** & A Valencia (1999) A platform for integrating threading results with protein family analyses. **Bioinformatics** 15, 1062-1063.
- \* 48. **B Rost** & C Sander (2000) Third generation prediction of secondary structure. **Meth Mol Biol** 143, 71-95.
- \* 49. **M Cokol**, **R Nair** & **B Rost** (2000) Finding nuclear localization signals. **EMBO Reports** 1, 411-415.
- \* 50. **B Rost** (2001) Protein secondary structure prediction continues to rise. **J Struct Biol** 134, 204-218.

- \* 51. **J Liu & B Rost** (2001) Comparing function and structure between entire proteomes. **Protein Science** 10, 1970-1979.
- \* 52. **V Eyrych, MA Marti-Renom, D Przybylski, A Fiser, F Pazos, A Valencia, A Sali & B Rost** (2001) EVA: continuous automatic evaluation of protein structure prediction servers. **Bioinformatics** 17, 1242-1243.
- 53. **D Fischer, A Elofsson, L Rychlewski, F Pazos, A Valencia, B Rost, AR Ortiz & RL Dunbrack** (2001) CAFASP2: the second critical assessment of fully automated structure prediction methods. **Proteins** 45 (S5), 171-183.
- \* 54. **B Rost & V Eyrych** (2001) EVA: large-scale analysis of secondary structure prediction. **Proteins** 45 (S5), 192-199.
  
- \* 55. **D Przybylski & B Rost** (2002) Alignments grow, secondary structure prediction improves. **Proteins** 46, 195-205.
- \* 56. **CAF Andersen, AG Palmer, S Brunak & B Rost** (2002) Continuum secondary structure captures protein flexibility. **Structure** 10, 175-184.
- 57. **G Pollastri, D Przybylski, B Rost & P Baldi** (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. **Proteins** 47, 228-235.
- 58. **MA Marti-Renom, MS Madhusudhan, A Fiser, B Rost, A Sali** (2002) Reliability of assessment of protein structure prediction methods at CASP. **Structure** 10, 435-440.
- \* 59. **B Rost** (2002) Enzyme function less conserved than anticipated, **J Mol Biol** 318, 595-608.
- \* 60. **B Rost** (2002) Did evolution leap to create the protein universe? **Curr Opin Struct Biol** 12, 409-416.
- \* 61. **CP Chen & B Rost** (2002) State-of-the-art in membrane protein prediction. **Appl Bioinformatics** 1, 21-35.
- \* 62. **J Liu & B Rost** (2002) Target space for structural genomics revisited. **Bioinformatics** 18, 922-933.
- \* 63. **B Rost, B Honig & A Valencia** (2002) Bioinformatics in structural genomics. **Bioinformatics** (Editorial) 18, 897-898.
- \* 64. **J Glasgow & B Rost** (2002) ISMB 2002. **Bioinformatics** (Editorial) 18 Suppl 1, S1
- \* 65. **R Nair & B Rost** (2002) Inferring sub-cellular localization through automated lexical analysis. **Bioinformatics** 18 Suppl 1 (ISMB Proceedings), S78-S86.
- \* 66. **J Liu, H Tan & B Rost** (2002) Loopy proteins appear conserved in evolution. **J Mol Biol** 322, 53-64.
- \* 67. **CP Chen, A Kernytsky & B Rost** (2002) Transmembrane helix predictions revisited. **Protein Science** 11, 2774-2791.
- \* 68. **R Nair & B Rost** (2002) Sequence conserved for sub-cellular localization. **Protein Science** 11, 2836-2847.
- \* 69. **CP Chen & B Rost** (2002) Long membrane helices and short loops predicted less accurately. **Protein Science** 11, 2766-2773.
  
- \* 70. **Y Ofran & B Rost** (2003) Analysing six types of protein-protein interfaces. **J Mol Biol** 325, 377-387.
- \* 71. **CAF Andersen & B Rost** (2003) Automatic secondary structure assignment. **Methods Biochem Anal** 44, 341-363.
- \* 72. **B Rost** (2003) Prediction in 1D: secondary structure, transmembrane helices and accessibility. **Methods Biochem Anal** 44, 559-587.
- \* 73. **P Carter, J Liu & B Rost** (2003) PEP: Predictions for entire proteomes. **Nucl Acids Res** 31, 410-413.
- \* 74. **R Nair, P Carter & B Rost** (2003) NLSdb: database of nuclear localization signals. **Nucl Acids Res** 31, 397-399.
- \* 75. **J Liu & B Rost** (2003) Domains, motifs, and clusters in the protein universe. **Curr Opin Chem Biol** 7, 5-11.

- \* 76. **B Rost** (2003) Rising accuracy of protein secondary structure prediction. In 'Protein structure determination, analysis, and modeling for drug discovery', D Chasman (ed.) New York: Dekker, 207-249.
- 77. R Zidovetzki, **B Rost**, Don L Armstrong & I Pecht (2003) Transmembrane domains in the functions of Fc receptors. **J Biophysical Chemistry** 15, 555-575.
- \* 78. **Y Ofran & B Rost** (2003) Predict protein-protein interaction sites from local sequence information. **FEBS Letters** 544, 236-239.
- \* 79. **B Rost** (2003) Neural networks predict protein structure: hype or hit? In 'Artificial intelligence and heuristic methods in bioinformatics' P Frasconi and R Shamir (eds.) Amsterdam: IOS Press, 34-50.
- \* 80. **KO Wrzeszczynski & B Rost** (2003) Cataloguing proteins in cell cycle control. In 'Cell cycle checkpoint control protocols' H Lieberman (ed.) Totowa: Humana Press, 219-233.
- \* 81. **B Rost, J Liu, D Przybylski, R Nair, H Bigelow, K Wrzeszczynski & Y Ofran** (2003) Predicting protein structure through evolution. In 'Chemoinformatics - From Data to Knowledge' J Gasteiger & T Engel (eds.) New York: Wiley, 1789-1811.
- \* 82. **B Rost & J Liu** (2003) The PredictProtein server. **Nucl Acids Res** 31, 3300-3304.
- \* 83. **J Liu & B Rost** (2003) NORSp: predictions of long regions without regular secondary structure. **Nucl Acids Res** 31, 3833-3835.
- \* 84. **R Nair & B Rost** (2003) LOC3D: annotate sub-cellular localization for protein structures. **Nucl Acids Res** 31, 3337-3340.
- \* 85. **S Mika & B Rost** (2003) UniqueProt: creating sequence-unique protein data sets. **Nucl Acids Res** 31, 3789-3791.
- \* 86. **A Kernytsky & B Rost** (2003) Static benchmarking of membrane helix predictions. **Nucl Acids Res** 31, 3642-3644.
- \* 87. **IYY Koh, VA Eyrich, MA Martí-Renom, D Przybylski, MS Madhusudhan, E Narayanan, O Graña, A Valencia, A Sali & B Rost** (2003) EVA: evaluation of protein structure prediction servers. **Nucl Acids Res** 31, 3311-3315.
- \* 88. **VA Eyrich & B Rost** (2003) META-PP: single interface to crucial prediction servers. **Nucl Acids Res** 31, 3308-3310.
- \* 89. **P Carter, CAF Andersen, B Rost** (2003) DSSPcont: a database of continuous secondary structure assignments for proteins, **Nucl Acids Res** 31, 3308-3310.
- \* 90. **R Nair & B Rost** (2003) Better prediction of sub-cellular localization by combining evolutionary and structural information. **Proteins** 53, 917-930.
- \* 91. **VA Eyrich, IYY Koh, D Przybylski, O Graña, F Pazos, A Valencia & B Rost** (2003) CAFASP3 in the spotlight of EVA. **Proteins** 53 S6, 548-560.
- 92. JM Aramini, YJ Huang, JR Cort, S Goldsmith-Fischman, R Xiao, L Shih, CK Ho, **J Liu, B Rost, B Honig, MA Kennedy, TB Acton & GT Montelione** (2003) Solution NMR structure of the 30S ribosomal protein S28E from *Pyrococcus horikoshii*. **Protein Science** 12, 2823-2830.
- \* 93. **B Rost, J Liu, R Nair, K Wrzeszczynski & Y Ofran** (2003) Automatic prediction of protein function. **Cellular and Mol Life Sciences** 60, 2637-2650.

### Publications in press

- \* 94. **R Nair & B Rost** (2004) Annotating protein function through lexical analysis. AAI Magazine special issue, in press.
- \* 95. **J Liu, H Hegyi, TB Acton, GT Montelione & B Rost** (2004) Automatic target selection for structural genomics on eukaryotes. **Proteins** , in press.
- 96. Z Wunderlich, TB Acton, **J Liu, G Kornhaber, J Everett, P Carter, N Lan, N Ecols, Mark Gerstein, B Rost, & GT Montelione** (2004) ZebaView: a web-based tool for organizing the protein target list of the Northeast Structural Genomics consortium. **Proteins** , in press.

- \* 97. **J Liu & B Rost** (2004) CHOP proteins into structural domains. *Proteins* in press.
- \* 98. **B Rost** (2004) How to use protein 1D structure predicted by PROFphd. **Meth Mol Biol**, in press.
- \* 99. **B Rost, G Yachdav & J Liu** (2004) The PredictProtein server. **Nucl Acids Res**, in press.
- \*100. **Y Ofran & B Rost** (2004) Predictive methods using protein sequence. In 'Bioinformatics' AD Baxevanis & F Ouelette (eds.) New York: Wiley, in press.

### Publications submitted

- \*101. **KO Wrzeszczynski & B Rost** (2004) Annotating proteins from Endoplasmic Reticulum and Golgi apparatus in eukaryotic proteomes. **Cellular and Molecular Life Sciences** 2004-01-06.
- \*102. **S Mika & B Rost** (2004) Protein names peeled precisely off free text. **Bioinformatics**, submitted 2004-01-09.
- \*103. **M Punta & B Rost** (2004) Toward good 2D predictions in proteins. **Bioinformatics**, submitted 2004-01-14.
- \*104. **J Liu & B Rost** (2004) Sequence-based prediction of protein domains. **Nucl Acids Res**, submitted 2004-01-20.
- \*105. **H Bigelow, D Petrey, J Liu, D Przybylski & B Rost** (2004) Predicting transmembrane beta-barrels for entire proteomes. **Nucl Acids Res**, submitted 2004-01-21.
- \*106. **D Przybylski & B Rost** (2004) Improving fold recognition without folds. **J Molecular Biology**, submitted 2004-02-03.
- \*107. **S Mika & B Rost** (2004) NLProt: extracting protein names and sequences from papers. **Nucl Acids Res**, submitted 2004-02-15.
- \*108. **J Liu & B Rost** (2004) CHOP: parsing proteins into structural domains. **Nucl Acids Res**, submitted 2004-02-14.
- \*109. **R Nair & B Rost** (2004) LOCnet and LOCtarget: Sub-cellular localization for structural genomics targets. **Nucl Acids Res**, submitted 2004-02-15.
- 110. R Powers, T Acton, Y Chiang, R Paranjy, JR Cort, LC Ma, **J Liu, B Rost**, GT Montelione (2003) 1H, 13C and 15N Assignments for the *Archaeoglobus fulgidis* Protein AF2095 Northeast Structural Genomics Consortium target GR4. Proteins in preparation.

### Publications in preparation

- \*111. **Y Ofran & B Rost** (2004) Predicting protein-protein interfaces through evolution. *PNAS*, in preparation.
- \*112. **S Mika & B Rost** (2004) Cataloguing nuclear matrix proteins. **Bioinformatics**, in preparation.
- \*113. D Murray, **J Liu, H Hegyi**, TB Acton, S Goldsmith, B Honig, GT Montelione & **B Rost** (2004) Target selection strategy for Northeast Structural Genomics project. **Proteins**, in preparation.
- \*114. **Y Ofran & B Rost** (2004) Rescue for statistical tests in high-throughput biology. **Science**, in preparation.
- \*115. **KO Wrzeszczynski, G Yachdav, P Carter & B Rost** (2004) PredictCellCycle: identifying cell cycle control proteins. **Bioinformatics**, in preparation.
- \*116. **A Schlessinger & B Rost** (2004) Protein flexibility prediction from sequence, in preparation.
- \*117. **P Carter & B Rost** (2004) PEP: database for predictions of entire proteomes. **Bioinformatics**, in preparation

## Preprints/Electronic material

- E1. **B Rost**: Protein fold recognition by merging 1D structure prediction and sequence alignments. Preprint; 1996; [http://cubic.bioc.columbia.edu/papers/pre1996\\_topits/](http://cubic.bioc.columbia.edu/papers/pre1996_topits/)
- E2. **B Rost & SI O'Donoghue**: Sisyphus and protein structure prediction. Fifth International Conference on Intelligent Systems for Molecular Biology; Tutorial; 1997; [http://cubic.bioc.columbia.edu/papers/pre1997\\_ismb/](http://cubic.bioc.columbia.edu/papers/pre1997_ismb/)
- E3. **B Rost, SI O'Donoghue and C Sander**: Midnight zone of protein structure evolution. Preprint; 1998; [http://cubic.bioc.columbia.edu/papers/pre1998\\_midnight/](http://cubic.bioc.columbia.edu/papers/pre1998_midnight/)
- E4. **B Rost**: Short yeast ORFs: expressed protein or not? Preprint; 1999; [http://cubic.bioc.columbia.edu/papers/1999\\_globe/](http://cubic.bioc.columbia.edu/papers/1999_globe/)
- E5. **B Rost**: Neural networks for protein structure prediction: hype or hit? Preprint; 1999; [http://cubic.bioc.columbia.edu/papers/pre1999\\_tics/](http://cubic.bioc.columbia.edu/papers/pre1999_tics/)
- E6. **B Rost and R Schneider**: Pedestrian guide to analysing sequence databases. Preprint; 1999; [http://cubic.bioc.columbia.edu/papers/1999\\_pedestrian/](http://cubic.bioc.columbia.edu/papers/1999_pedestrian/)
- E7. **R Nair and B Rost**: Surface profiles predict sub-cellular localisation. Preprint; 2001; [http://cubic.bioc.columbia.edu/papers/2001\\_loci\\_surface/](http://cubic.bioc.columbia.edu/papers/2001_loci_surface/)
- E8. **B Rost, P Baldi, G Barton, J Cuff, VA Eyrich, D Jones, K Karplus, R King, G Pollastri and D Przybylski**: Simple jury predicts protein secondary structure best. Preprint; 2001; [http://cubic.bioc.columbia.edu/papers/2001\\_sec\\_jury/](http://cubic.bioc.columbia.edu/papers/2001_sec_jury/)
- E9. **B Rost, J Liu, D Przybylski, R Nair, KO Wrzeszczynski, H Bigelow and Y Ofran**: Predict protein structure and function through evolutionary information. Preprint; 2002; [http://cubic.bioc.columbia.edu/papers/2002\\_rev\\_chem/](http://cubic.bioc.columbia.edu/papers/2002_rev_chem/)

## Web services (\$root = <http://cubic.bioc.columbia.edu>)

- W1. **B Rost & J Liu**: PredictProtein: Service for sequence analysis and protein structure prediction. Service (first internet server for protein structure prediction; one of the first in molecular biology); \$root/predictprotein/
- W2. **VA Eyrich & B Rost**: META-PP: Single-page interface to validated sequence analysis services. Service; \$root/predictprotein/submit\_meta.html
- W3. **IYY Koh, VA Eyrich, MA Martí-Renom, D Przybylski, MS Madhusudhan, E Narayanan, O Graña, A Valencia, A Sali & B Rost**: EVA: Evaluation of automatic protein structure prediction servers. Service; \$root/eva/
- W4. **R Nair & B Rost**: PredictNLS: prediction and analysis of nuclear localization signals. Service; \$root/services/predictNLS/
- W5. **P Carter, CAF Andersen & B Rost**: DSSPcont: Continuous assignment of protein secondary structure. Service and database; URL: \$root/services/DSSPcont/
- W6. **S Mika & B Rost**: UniqueProt: creating sequence-unique protein databases. Service; \$root/services/uniqueprot/
- W7. **J Liu & B Rost**: NORSp: prediction of Non-Regular Secondary Structure (NORS). Service; \$root/services/NORSp/
- W8. **A Kernytsky & B Rost**: TMH-Benchmark: evaluate new methods for membrane helix structure prediction. Service; \$root/services/tmh\_benchmark/
- W9. **P Carter, J Liu & B Rost**: PEP: predictions for entire proteomes. Database; \$root/db/PEP/
- W10. **R Nair & B Rost**: NLSdb: database of nuclear localization signals. Database; \$root/db/NLSdb/
- W11. **R Nair & B Rost**: LOC3d: database of predicted sub-cellular localization for eukaryotic PDB chains. Database; \$root/db/LOC3d/
- W12. **R Nair & B Rost**: LOChom: database of sub-cellular localization predictions based on sequence homology. Database; \$root/db/LOChom/
- W13. **S Mika & B Rost**: NLProt: extracting protein names and sequences from scientific papers. Service; \$root/services/NLProt/

- W14. **R Nair & B Rost:** LOCnet and LOctarget: De novo prediction of sub-cellular localization and predictions for targets in structural genomics. Database and Service; \$root/services/LOCnet/
- W15. **H Bigelow & B Rost:** PROFtmb: Predictions of transmembrane beta barrels. Service; \$root/services/PROFtmb/
- W16. **M Punta & B Rost:** PROFcon08: Predictions of inter-residue contacts. Service; \$root/services/PROFcon08/
- W17. **J Liu & B Rost:** CHOP: Chopping proteins into structural domains. Service; \$root/services/CHOP/

## Overall numbers of publications

### *Cumulative 1999-2003:*

- 71% of the peer-reviewed and 65% of all publications since 1999
- **57% of the peer-reviewed and 53% of all publications 2002-now**

### *Types of publications:*

|                         |    |
|-------------------------|----|
| Peer-reviewed:          | 70 |
| Original Peer-reviewed: | 58 |
| Reviews:                | 22 |
| Books:                  | 1  |
| Book chapters:          | 21 |
| Conference Proceedings: | 8  |
| Editorials:             | 2  |
| First/Last author:      | 87 |
| First author:           | 31 |
| Single author:          | 20 |
| Last author:            | 36 |

## The most often quoted publications (numbering as above)

*Publications (subset with >10 citations; sorted by impact over last year)*

| Excerpt from ISI Index, Feb 2003 |                                       |                 |             |
|----------------------------------|---------------------------------------|-----------------|-------------|
| No                               | Quote                                 | Last 12 months* | Total       |
| 8                                | 1993 JMB 232, 584-599                 | 152             | <b>1525</b> |
| 27                               | 1996 Meth Enzymol 266, 525-539        | 90              | <b>596</b>  |
| 16                               | 1994 Proteins 19, 55-72               | 84              | <b>979</b>  |
| 29                               | 1996 Protein Science 5, 1704-18       | 41              | 206         |
| 23                               | 1995 Protein Science 4, 521-533       | 36              | <b>417</b>  |
| 49                               | 2000 EMBO Reports 1, 411-5            | 35              | 61          |
| 42                               | 1999 Prot Engng 12, 85-94             | 33              | 103         |
| 19                               | 1994 CABIOS 10, 53-60                 | 31              | <b>453</b>  |
| 36                               | 1997 JMB 270, 471-80                  | 30              | 112         |
| 52                               | 2001 Bioinformatics 17, 1242-3        | 27              | 35          |
| 50                               | 2001 J Structural Biology 134, 204-18 | 23              | 42          |
| 20                               | 1994 JMB 235, 13-26                   | 21              | 120         |
| 9                                | 1993 PNAS 90, 7558-7562               | 20              | <b>258</b>  |
| 17                               | 1994 Proteins 20, 216-226             | 19              | 153         |
| 59                               | 2002 JMB 318, 595-608                 | 17              | 17          |
| 43                               | 1999 Proteins 34, 220-3               | 15              | 39          |
| 66                               | 2002 JMB 322, 53-64                   | 12              | 12          |
| 62                               | 2002 Bioinformatics 18, 922-933       | 12              | 12          |
| 51                               | 2001 Protein Science 10, 1970-9       | 12              | 28          |
| 70                               | 2003 JMB 325, 377-87                  | 11              | 11          |
| 67                               | 2002 Prot Sci 11, 2774-91             | 11              | 11          |
| 30                               | 1996 Ann Rev 25, 113-36               | 8               | 61          |
| 24                               | 1995 Proteins 23, 295-300             | 7               | 53          |
| 34                               | 1996 Fold & Des 2, S19-S24            | 7               | 33          |
| 39                               | 1998 Structure 6, 259-63              | 4               | 41          |
| 35                               | 1997 CABIOS/Bioinformatics 13, 345-56 | 2               | 31          |
| 5                                | 1992 Nature 360, 540                  | 2               | 81          |
| 11                               | 1993 TIBS 18, 120-123                 | 1               | 82          |
| 37                               | 1998 JMB 276, 517-25                  | 1               | 25          |
| 10                               | 1993 Prot Engng 6, 831-6              | 1               | 24          |
| 7                                | 1993 Nature Genetics 5, 376-380       | 0               | 84          |
| 22                               | 1995 ISMB                             | -               | 73 **       |
| 28                               | 1996 ISMB                             | -               | 22 **       |

### Statistics

|                                       | Sum  | Average Citation index |
|---------------------------------------|------|------------------------|
| All 76*** publications:               | 5320 | 70                     |
| 45 ** First/Last author papers listed | 5036 | <b>112</b>             |

\* more precisely: citations from Feb 2003 - Feb 2004

\*\* numbers from Feb 2001, no longer listed in ISI

\*\*\* averages are compile over ALL papers listed in the Publication list; since many publications do not appear in ISI (too new, book chapters, proceedings), the averages constitute lower bounds.