

**BIOGRAPHICAL SKETCH**

Provide the following information for the key personnel in the order listed for Form Page 2.  
Follow the sample format on for each person. (See attached sample). **DO NOT EXCEED FOUR PAGES.**

NAME		POSITION TITLE	
Burkhard Rost		Associate Professor, Biochemistry, Columbia Univ	
EDUCATION/TRAINING ( <i>Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.</i> )			
INSTITUTION AND LOCATION	DEGREE (if applicable)	YEAR(s)	FIELD OF STUDY
Gießen University (Germany)	BS	1984	Physics
Heidelberg University (Germany)	BA	1988	Philosophy
Heidelberg University (Germany)	MS	1988	Physics
Heidelberg University (Germany)	Dr. rer. nat.	1994	Physics

**A. Positions and Honors.****Positions and Employment**

1986 - 1988	<b>Assistant</b> at Institute for Theoretical Physics, Heidelberg University (Germany)
1988 - 1990	Research fellow at Institute for Theoretical Physics, Heidelberg University (Germany)
1990 - 1992	Visitor at the European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
1993 - 1994	Research fellow at the EMBL Heidelberg, Germany
1995	Scientist at the European Bioinformatics Institute (EBI), Hinxton, Cambridge, England
1996 - 1998	Scientist at EMBL Heidelberg, Germany
1998	Researcher at the company LION-Biosciences, Heidelberg, Germany
1999 - 2000	<b>Assistant Professor</b> at Columbia Univ., Dept. Biochemistry & Molecular Biophysics
2000 - present	<b>Associate Professor</b> at Columbia Univ., Dept. Biochemistry & Molecular Biophysics

**Other Experience and Professional Memberships**

1996 - present	Member of ISCB (International Society for Computational Biology)
2002 - present	Member of New York Academy of Sciences
1995 - present	Program Committee of ISMB meetings (Intelligent Systems for Molecular Biology)
2002 - present	Board of Directors ISCB
1992 - present	Involved in Organization of 14 International meetings, including most important meeting in bioinformatics ISMB 2002 in Edmonton, Canada (1500 participants), and most important meeting in the field of structure prediction (CASP6, Italy)
2001 - 2003	Ad hoc panels at NIH, including one as chair of the panel
2002 - 2003	Ad hoc panels for NSF
2003	Ad hoc panel for European Community
1999 - present	Over 300 referee reports (137 in 2003) for peer-reviewed journals
2001 - present	Associate Editor Journal of Medical Informatics
1999 - present	Ad hoc panels in England, Austria, Switzerland, Denmark, Israel, Sweden, Spain, Italy
1992 - present	More than 92 invited talks at international meetings in 16 countries since 1988

**B. Selected peer-reviewed publications (in chronological order).**

(First- and last-author publications selected from 53 peer-reviewed publications)

- 8 **B Rost** & C Sander (1993) Prediction of protein secondary structure at better than 70% accuracy. **J Mol Biol** 232, 584-599.
- 17 **B Rost** & C Sander (1994) Conservation and prediction of solvent accessibility in protein families. **Proteins** 20, 216-26.
- 30 **B Rost**, R Casadio & P Fariselli (1996) Topology prediction for helical transmembrane proteins at 86% accuracy. **Protein Science** 5, 1704-1718.

- 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.
- 41 **B Rost** (1999) Twilight zone of protein sequence alignments. **Prot Engng** 12, 85-94.
- 48 **M Cokol, R Nair & B Rost** (2000) Finding nuclear localisation signals. **EMBO Rep** 1, 411-415.
- 50 **J Liu & B Rost** (2001) Comparing function and structure between entire proteomes. **Protein Sci** 10, 1970-1979.
- 51 **V Eyrich**, 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.
- 52 **D Przybylski & B Rost** (2002) Alignments grow, secondary structure prediction improves. **Proteins** 46, 195-205.
- 53 **CAF Andersen**, AG Palmer, S Brunak & **B Rost** (2002) Continuous secondary structure assignment correlates with protein flexibility. **Structure** 10, 175-184.
- 58 **B Rost** (2002) Enzyme function less conserved than anticipated. **J Mol Biol** 318, 595-608.
- 61 **J Liu & B Rost** (2002) Target space for structural genomics revisited. **Bioinformatics** 18, 922-933.
- 64 **R Nair & B Rost** (2002) Inferring sub-cellular localization through automated lexical analysis. **Bioinformatics** 18 (ISMB Proceedings), S78-S86.
- 65 **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. **Prot Science** 11, 2774-91.
- 68 **R Nair & B Rost** (2002) Sequence conserved for sub-cellular localization. **Prot Science** 11, 2836-47.
- 69 **CP Chen & B Rost** (2002) Long membrane helices and short loops predicted less accurately. **Prot Science** 11, 2766-73.
- 70 **Y Ofran & B Rost** (2003) Analysing six types of protein-protein interfaces. **J Mol Biol** 325, 377-387.
- 73 **P Carter, J Liu & B Rost** (2003) PEP: Predictions for entire proteomes. **Nucl Acids Res** 31, 410-413.
- 78 **Y Ofran & B Rost** (2003) Predict protein-protein interaction sites from local sequence information. **FEBS Letters** 544, 236-239.
- 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.
- 94 **J Liu, H Hegyi**, TB Acton, GT Montelione & **B Rost** (2004) Automatic target selection for structural genomics on eukaryotes. **Proteins**, in press.
- 97 **J Liu & B Rost** (2004) CHOP proteins into structural domain-like fragments. **Proteins** in press.

Note 1: numbers according to full publication list sorted by date (total incl. submitted=99); members of my group in bold face.

Note 2: all 45 first-/last-author papers had been quoted more than 5000 times by Jun 2003.

## C. Research Support

### Active

<b>P50</b> GM62413	Montelione (PI)	Date:	09/30/00 to 08/31/05	
<b>NIH</b>		<b>Annual Direct:</b>	<b>\$100,000</b>	11%
<b>Role:</b>	<b>Co-PI</b>			

Title: Structural genomics of eukaryotic model organisms

The major goal of this project is to develop high-throughput techniques for experimental determination of protein structures. The objective of the bioinformatics component is to propose targets and to develop ways of using experimental structures to infer aspects of function.

<b>R01</b> GM63029	Rost (PI)	Date:	04/01/01 to 03/31/04	
<b>NIH</b>		<b>Annual Direct:</b>	<b>\$150,000</b>	12%

Title: Intruding into the midnight zone of protein comparisons

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 comparisons. 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.

GM63029	Rost (PI)	Date:	02/15/02 to 01/31/05	
<b>NSF</b>		<b>Annual Direct:</b>	<b>\$100,000</b>	12%

**Title:** Ab initio prediction of sub-cellular localization

The major goals of this project are to develop methods automatically predicting sub-cellular localization of proteins from protein alignments. The ultimate objective is to develop and combine a series of new methods into a comprehensive system supporting automatic genome annotation. The novel aspects are the rigorous testing of conservation of localization, the use of alignment information, and observed/predicted surface composition, the proposed testing of known signaling motifs, and the particular way of combining the components into a comprehensive system.

R01 GM64633-02A2      Rost (PI)      Date:      04/01/03 to 03/31/07  
**NIH**      **Annual Direct:      \$150,000      15%**

**Title:** Predict putative protein-protein interface segments at low resolution

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.

R01 LM07329-01A1      Rost (PI)      Date:      04/01/03 to 03/31/07  
**NIH/NLM**      **Annual Direct:      \$175,000      15%**

**Title:** Improve predictions of structure and function by PredictProtein

The major goals of this project are to continue offering the protein structure prediction server PredictProtein to the community and to implement a variety of technical and scientific solutions improving the functionality of PredictProtein. (1) The technical solutions address job handling, data handling, database update, user interface, layout of web pages, presentation of results, and directly linking the output to original resources. (2) The systematic combination of methods focus on improving predictions for membrane helical proteins and on using structure predictions to more accurately infer functional information.

**Pending**

   Rost (PI)      Date:      08/01/04 to 07/31/08  
**NIH**      **Total Direct:      \$224,572**

**Title:** Meeting on Critical Assessment of Protein Structure Prediction (CASP6-CASP8)

Over the last decade the bi-annual CASP meetings have evolved into undoubtedly the most important event in the field of protein structure prediction. This grant will support the next three meetings CASP6-9.

M15 NCBC/Roadmap      Califano (PI)      Date:      06/01/04 to 05/31/09  
**NIH**      **Annual Direct:      \$3,000,000 (entire center grant)**

**Role: Investigator/Project leader**

**Title:** National Center of Excellence in Bio-Computing

**Completed**

**Startup**      Hirsh (PI)      Date:      01/01/96 to 12/31/01  
**HHMI**      **Total Direct:      \$450,000**

This grant constituted my startup capital since I arrived at Columbia in 12/98. I have used it to support my own salary and to purchase equipment.

ISMB      Rost (PI)      Date:      2002  
**NSF**      **Annual Direct:      \$40,000**

**Title:** Support for the sixth international meeting on 'Intelligent Systems for Molecular Biology'

The ISMB meeting has evolved to the largest meeting on computational biology. The 2002 meeting that I co-organized was held in Edmonton, Canada; it had over 1800 participants.

ISMB      Rost (PI)      Date:      2002  
**DOE**      **Annual Direct:      \$35,000**

**Title:** Support for the sixth international meeting on 'Intelligent Systems for Molecular Biology'

The ISMB meeting has evolved to the largest meeting on computational biology. The 2002 meeting that I co-organized was held in Edmonton, Canada; it had over 1800 participants.

## **Overlap**

### **Completed Research General**

Achieved in 1999: (1) Moving the internet prediction service PredictProtein from the EMBL Heidelberg to Columbia University ([cubic.bioc.columbia.edu/predictprotein](http://cubic.bioc.columbia.edu/predictprotein)), improving the interface, and including additional programs. (2) In summer 1999, launching the first version of META-PredictProtein. This new server enables users to access many different additional tools (currently 15) through one single interface ([cubic.bioc.columbia.edu/predictprotein/submit\\_meta.html](http://cubic.bioc.columbia.edu/predictprotein/submit_meta.html)).

Achieved in 2000: (1) Program and expert database to predict presence of nuclear localization signals in proteins (#50). (2) In collaboration with the groups of Andrej Sali (Rockefeller) and Alfonso Valencia (Madrid), we started the first automatic, continuous server analyzing the performance of methods for comparative modeling, and prediction of secondary structure and inter-residue distances (EVA: [cubic.bioc.columbia.edu/eva](http://cubic.bioc.columbia.edu/eva), #54). (3) Analysis of 28 entire proteomes with various structure prediction tools (#52). (4) Improvement of secondary structure prediction through iterated PSI-BLAST (#53).

Achieved in 2001: (1) First target list for structural genomics. (2) Inclusion of fold recognition/threading servers into EVA. (3) First server for the detection of proteins with long regions depleted of regular secondary structure (natively unstructured regions/loopy proteins)..

Achieved in 2002: (1) Completion of continuous assignment of secondary structure (DSSPcont) and launch of database and program ([cubic.bioc.columbia.edu/services/DSSPcont](http://cubic.bioc.columbia.edu/services/DSSPcont)). (2) Preliminary release of database with predictions for entire proteomes (PEP: [cubic.bioc.columbia.edu/pep](http://cubic.bioc.columbia.edu/pep)). (3) First attempt at predicting enzymatic activity through homology (server under construction). (4) First version of method predicting sub-cellular localization through homology (server under construction). (5) Tool for automatic text analysis to infer sub-cellular localization through homology (server under construction).

### **Ongoing Research No Support Requested, yet**

Our group is currently also working on the following projects. (1) Improve and extend the EVA server continuously evaluating structure prediction, (2) improve methods to predict secondary structure and solvent accessibility, (3) build a database of predictions for entire genomes (first prototype available since summer 2002), (4) predict structural domains from sequence.

More generally, we attempt to shift focus from predicting structure to predicting function.