

## **The Grammatical Structure of Protein Sequences**

**Chris Bystroff**

### **Abstract**

We describe a hidden Markov model, HMMSTR, for general protein sequence based on the I-sites library of sequence-structure motifs. Unlike the linear HMMs used to model individual protein families, HMMSTR has a highly branched topology and captures recurrent local features of protein sequences and structures that transcend protein family boundaries. The model extends the I-sites library by describing the adjacencies of different sequence-structure motifs as observed in the database, and achieves a great reduction in parameters by representing overlapping motifs in a much more compact form.

The HMM attributes a considerably higher probability to coding sequence than does an equivalent dipeptide model, predicts secondary structure with an accuracy of 74.6% and backbone torsion angles better than any previously reported method, and predicts the structural context of beta strands and turns with an accuracy that should be useful for tertiary structure prediction. HMMSTR has been incorporated into a public, fully-automated protein structure prediction server.

## **Curriculum Vitae**

### **Christopher Bystroff**

Date of birth: 2 January 1960

Home address: 8 Lansing Ave

Troy, NY 12180

Home phone: 518-272-0071

Work address: Dept of Biology, MRC 303

Rensselaer Polytechnic Institute

Troy, NY 12180

Work phone: 518-276-3185

fax: 518-276-2162

Nationality: USA

Wife: Maria Victoria

Children(2): Lindsay (6), Veronica (15 mo)

email: bystrc@rpi.edu

website: [isites.bio.rpi.edu/bystrc/index.html](http://isites.bio.rpi.edu/bystrc/index.html)

### **Higher Education**

<i>years</i>	<i>institution</i>	<i>degree</i>
1978-1983	Carleton College, Minnesota	B.A. cum laude, (chemistry)
1983-1985	UCSD, La Jolla, CA	M.S. (chemistry)
1985-1988	UCSD, La Jolla, CA	Ph.D. (chemistry)

### **Postdoctoral Work**

<i>year</i>	<i>position/ P.I.</i>	<i>activities</i>	<i>place</i>
1988	postdoc, Joseph Kraut	xray crystallography	UCSD
1989-1992	postdoc, R.J.Fletterick	xray crystallography	UCSF
1992-1994	Fulbright fellow	Teaching, research	UNI, Managua
1995-1999	senior fellow, D.Baker	Research	UW, Seattle
1999-	Asst. professor	bioinformatics	RPI, Troy NY

### **Awards**

1993-1994 Fulbright research/teaching fellowship to Nicaragua

1989-1992 NIH PHS post-doctoral training grant, UCSF

1978-1983 Carleton College scholarship

### **Publications (most recent first)**

Zaki, M., Jin, C., Bystroff, C. Mining residue contacts in proteins using predicted local structure. Sci. Prog. (submitted, 2000)

Bystroff, C., Thorsson, V., Baker, D. HMMSTR: A hidden Markov model for sequence-structure correlations in proteins. J. Mol. Biol. 301(1):173-190, 2000.

Thayer, E., Bystroff, C., Baker, D. Detection of protein sequences using a mixture model for local protein amino acid sequence. J. Comput. Biol. 7:317-27, 2000.

Simons, K.T., Ruczinski, I., Kooperberg, C., Fox, B.A., Bystroff, C., Baker, D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins, Structure, Function and Genetics 34(1):82-95, 1999.

Yi, Q., Bystroff, C., Rajagopal, P., Klevit, R.E., Baker, D. Prediction and structural characterization of an independently folding substructure in the src SH3 domain. J. Mol. Biol. 283(1):293-300, 1998.

Bystroff, C., Baker, D. Prediction of local structure in proteins using a library of sequence-structure motifs. J. Mol. Biol. 281:565-577, 1998.

Bystroff, C., Baker, D. Blind ab initio local structure predictions using a library of sequence-structure motifs. Proteins: Structure, Function, and Genetics, Suppl. 1:167-171, 1997.

Han, K.F., Bystroff, C., Baker, D. Three dimensional structures and contexts associated with recurrent amino acid sequence patterns. Protein Science 6:1587-1590, 1997.

Bystroff, C., Simons, K.T., Han, K.F., Baker, D. Local sequence-structure correlations in proteins. Curr Op Biotech 7:417-421, 1996.

McGrath, M.E., Erpel, T., Bystroff, C., Fletterick, R.J. Macromolecular chelation as an improved mechanism of protease inhibition: structure of the ecotin-trypsin complex. Embo J 13:1502-1507, 1994.

Bystroff, C., Baker, D., Fletterick, R.J., Agard, D.A. PRISM: application to the solution of two protein structures. *Acta Crystallographica*, Section D 5:440-448, 1993.

Baker, D., Bystroff, C., Fletterick, R.J., Agard, D.A. PRISM: topologically constrained phase refinement for macromolecular crystallography. *Acta Crystallographica*, Section D 5:429-439, 1993.

Goddette, D.W., Paech, C., Yang, S.S., Mielenz, J.R., Bystroff, C., Wilke, M.E., Fletterick, R.J. The crystal structure of the *Bacillus lentus* alkaline protease, subtilisin BL, at 1.4 Å resolution. *J Mol Biol* 228:580-595, 1992.

Bystroff, C., Kraut, J. Crystal structure of unliganded *Escherichia coli* dihydrofolate reductase. Ligand-induced conformational changes and cooperativity in binding. *Biochemistry* 30:2227-2239, 1991.

Bystroff, C., Oatley, S.J., Kraut, J. Crystal structures of *Escherichia coli* dihydrofolate reductase: the NADP<sup>+</sup> holoenzyme and the folate.NADP<sup>+</sup> ternary complex. Substrate binding and a model for the transition state. *Biochemistry* 29:3263-3277, 1990.

### **Recent Collaborators**

Eric Alm, David Baker, Richard Bonneau Karen Butner, Brian Fox, Viara Grancharova, David Kim, Charles Kooperberg, Rachel E. Klevit, Ponni Rajagopal, David Riddle, Ingo Ruczinski, Kim T Simons, Charles E. Strauss, Edward Thayer, Vestienn Thorsson, Jerry Tsai, Qian Yi :: University of Washington, Seattle WA  
Karen F. Han :: University of California, San Francisco.

Wilfredo Colon, Mark Goldberg, John Salerno, Mohammed Zaki :: Rensselaer Polytechnic Institute, Troy NY

Susan Baxter, Charles Lawrence :: Wadsworth Laboratories, Albany NY.

#### Advisees

Yu Shao, Kirsten Piotrowski, Roumen Dimitrov, Xin Yuan

#### Advisors

Joseph Kraut, University of California, San Diego.

Robert Fletterick, University of California, San Francisco.

David Baker, University of Washington, Seattle.