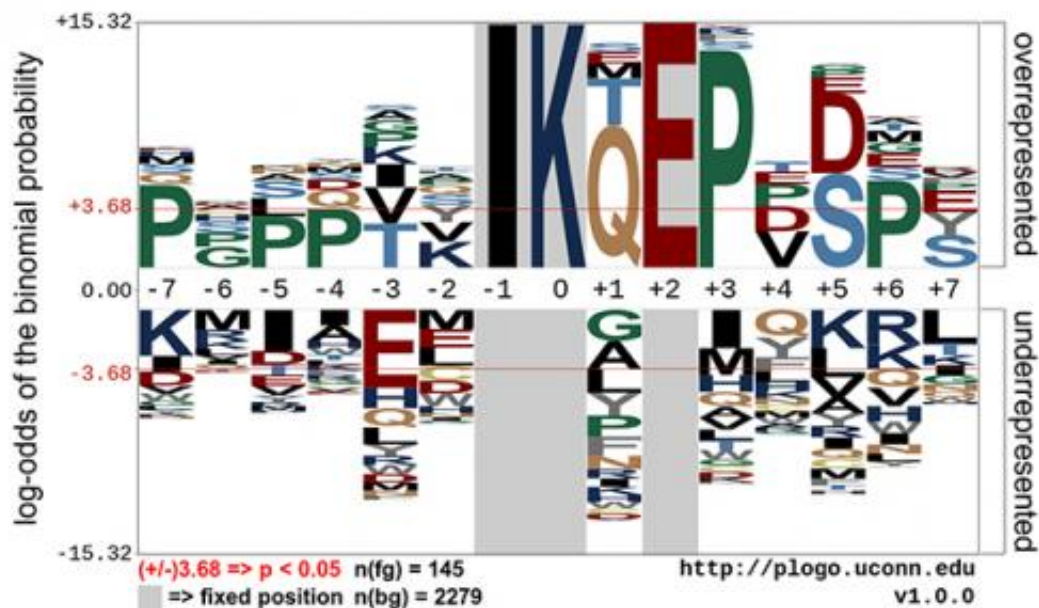# A new tool for visualizing DNA, protein sequences

October 10 2013, by Bri Diaz

A pLogo representing the protein sequences modified by the SUMO-family of enzymes. Credit: Daniel Schwartz

(Medical Xpress)—A group of researchers and students from UConn and Harvard Medical School have developed a new Web program that will help scientists visually analyze DNA and protein sequence patterns faster and more efficiently than ever before.

Called the probability logo (pLogo) generator, the program produces graphical representations of amino acids and nucleotides – the building blocks of biological molecules, like protein and DNA.

The methodology for the pLogo generator was published online in the journal *Nature Methods* on Oct. 6.

"This project represents a major goal of our lab, which is to create resources for the molecular biology research community that are user-friendly and as interactive as possible," says Daniel Schwartz, assistant professor of physiology and neurobiology and leader of the pLogo development team.

Schwartz conceived the algorithm that became the pLogo visualization strategy when he was a postdoctoral researcher at Harvard Medical School along with Michael Chou, who is currently a lecturer at Harvard Medical School. Schwartz and Chou both worked in the lab of George Church, professor of genetics at Harvard Medical School; all three are co-authors on the paper.

The program is remarkable not only for the new features it provides to the scientific community, but also for the work that went into its public debut: Schwartz drew on the computational and Web development expertise of several UConn computer science and engineering students to convert his visualization strategy into the Web-based tool.

Computer science Ph.D. student Saad Quader and undergraduate students Joey O'Shea '14 (ENG) and Kevin Ryan '14 (ENG) are also co-authors on the paper, with O'Shea billed as a co-first author.

"It is quite an achievement for an undergraduate student to be lead programmer on a project and co-first author on a paper published in Nature Methods," says Schwartz.

pLogo at work

The new program allows users to visualize short linear patterns, or "motifs," in a biological molecule by producing a series of scaled, color-coded letters that represent the biological residues that make up the molecule. The size of each letter indicates the relative significance of a residue occurring at a particular position in a motif.

While pLogo is not the first open access logo generator, it does introduce several groundbreaking interactive features. Users supply a foreground data set, which they collect from a sample organism, and pLogo automatically generates a background data set that represents the entire set of proteins in that organism.

"With a foreground and background data set, we can compare and scale letters relative to their overall statistical significance instead of just their frequency of occurrence," says Schwartz. "This means we can determine if a data set generated in a lab is special and to what extent it is special."

In addition to the visualization, pLogo allows users to interact with their motif data in real time, which Schwartz says has never been offered before. Researchers can generate specific statistical information and new visualizations based on conditional probabilities by simply dragging their cursor over a letter.

The program has other "smart" attributes; it automatically detects and corrects formatting inconsistencies and proposes parameters for analyzing user data, which according to Schwartz may easily contain 5,000 or 10,000 sequences.

"We wanted the interface to be virtually effortless, with researchers being able to spend time on analysis rather than troubleshooting minor formatting errors in their data sets," he says.

A challenging experience for students

Since the project's inception, Schwartz intended to build pLogo into a highly interactive Web tool. It wasn't until he arrived at UConn that he was able to achieve this goal by recruiting talented graduate and undergraduate students from the Department of Computer Science and Engineering to work in his lab.

In the process, students like O'Shea, Quader, and Ryan were able to take advantage of what they describe as an invaluable learning opportunity: building a professional-grade product from the ground up.

"We took ownership of this project and were encouraged to come up with our own solutions to problems, like how to make relatively complex calculations run as fast as possible," says Quader, who has worked in Schwartz's lab since summer of 2011.

Additionally, Quader emphasizes that Schwartz's high standards for the project in terms of performance, interactivity, and visual aesthetics set an example to the young scholars working on the team.

"It not only elevated the pLogo generator to its current state, but also motivates me to set similar standards for my own research projects," he says.

Adds Ryan, "This is real work that is challenging and enjoyable … I don't know if I could have gotten experience like this anywhere else on campus."

Schwartz and his team are continuing to develop interactive tools for the scientific community, a process he describes as "rewarding all around."

"I try to give the students a high degree of freedom to use new

technologies that benefit them beyond their time at UConn, but I also benefit tremendously from the expertise, motivation, and passion they bring to lab every day," he says.

O'Shea cites the opportunities presented in Schwartz's lab as a major part of why he transferred to UConn from Marist College.

"I have learned so much working on this project, and I feel like we created something that we are really proud of," he says. "It is exciting to make things that benefit people, like this tool, which will be used by scientists around the world."

Provided by University of Connecticut

Citation: A new tool for visualizing DNA, protein sequences (2013, October 10) retrieved 3 May 2024 from https://medicalxpress.com/news/2013-10-tool-visualizing-dna-protein-sequences.html