IUPattern

The DPstructure software package, which includes IUPattern, is intended for the analysis of intrinsically unstructured and disordered protein regions. IUPattern is used for the prediction of binding sites on nucleic acid biding IUPs.

IUPattern is an improvement upon SeqCom, our previous nucleic acid binding site predictor, as IUPattern analyzes specific patterns in the predicted sequence before producing the final predicted binding sites.

Instructions for the use of IUPattern:

The code for IUPattern is written in C++. It can simply be extracted into a .cpp file and compiled. We compile the source code in a terminal of a Unix-based machine using the command

```
g++ -o IUPattern IUPattern.cpp.
```

Once the application is compiled, it can be run in the terminal of a Unix-based machine by typing the command

```
./IUPattern [sequence file] [desired output file]
```

in the directory containing the compiled IUPattern file. If users have difficulty getting the application to compile and run, be sure to send an email to goodman.russell.c@gmail.com.

Interpreting the output of IUPattern:

There will be five columns in the structure files produced by IUPattern. One column will contain letters corresponding to the sequence, and the other columns will contain **0**, **1**, or **2**. A **0** indicates that the amino acid in that row does not participate in the particular physical property corresponding to that column. A **1** indicates that the amino acid does participate in the physical property corresponding to that column. A **2** indicates that the algorithm wasn't able to produce the result for that row. From left to right, the columns represent the sequence, the location of the intrinsically unstructured region, the binding site of the IDP, alpha helix formation, and beta-pleated sheet formation.

Here's an example of a structure file:

```
L 2 0 1 2
E 2 1 1 2
E 2 1 1 2
A 2 1 1 2
L 2 0 1 2
A 2 1 1 2
A 2 1 1 2.
```

Here, the location of the intrinsically unstructured regions of the protein cannot be determined, which is represented by the **2**s in the second column. Every amino acid except the two leucine amino acids were expected to participate in the binding site, which is indicated by the **0**s and **1**s in column three. An alpha helix was formed along the entire sequence of the IDP (N.B., IUPattern does not predict secondary structure). Finally, IUPattern could not calculate the beta-pleated sheet formation for this IDP, which can be seen by the **2**s in the last column, column five.