

Computationally Searching the Genome for Novel JNK Substrates

Thomas Whisenant^{1,2*}, David Ho¹, Ryan Benz², Frank Antilla¹, Pierre Baldi¹,
Lee Bardwell^{1,2}

1. Developmental and Cell Biology, University of California, Irvine, California, USA
2. Institute for Genomics and Bioinformatics, University of California, Irvine, California, USA

*email: twhisena@uci.edu

Mitogen-Activated Protein Kinase (MAPK) pathways enable cells to respond appropriately to stress and a vast array of other extracellular signals. The pathways include a conserved three-tiered cascade of kinases, which activate each other in succession to propagate the signal. MAPK substrates and/or associated proteins have been shown to interact with the kinase through a docking site motif (D-site) separate from phosphoacceptor residues involved in phosphorylation. The D-site consensus sequence has been identified experimentally as (basic₁₋₃-X₂₋₆-hydrophobic-X-hydrophobic) for substrates of the MAPK c-Jun N-terminal Kinase (JNK). The availability of this motif data enables a computational approach which allows us to search the genome for novel JNK substrates where there is currently little available data.

Hidden Markov Models (HMMs) are a class of probabilistic graphical models used in sequence alignment and other bioinformatic problems. Using D-site sequences from 20 known JNK interacting proteins, a HMM was trained to predict additional proteins in the predicted human translated transcriptome that contain this motif. Prior to running the algorithm against a large database of full length sequences, multiple separate validations were performed. These included the verification of the HMM's ability to identify the known D-site sequences in the full length protein sequences of the training set members. Following validation, the HMM algorithm was run against a database of over 36,000 human protein sequences and yielded a total of 422 predicted D-site sequences above a statistically determined threshold.

To directly test the results of the HMM, we synthesized the predicted D-site peptides onto a peptide macroarray and probed for binding with ³⁵S-JNK1. The binding of the array was quantified with a phosphoimager and normalized to peptides for the positive control, MKK7. Peptides that bound as well as or better than the positive control were classified as accurate predictions by the HMM. The approximate accuracy of the HMM based on the peptide macroarray validation was 64% (36/56).

Molecular cloning was used to create proteins of the top candidates from both the HMM and peptide macroarray. Fragments of the full length proteins with and without the D-site were used in GST pull down assays to verify binding. Specifically, the top novel prediction from the HMM, GLI-Kruppel family member 3 (GLI3), was cloned and tested as N-terminal fragments with and without the D-site. The results showed improved binding of GLI3 to JNK2 in the presence of the D-site. Using these techniques, we have identified multiple new JNK substrates that appear to bind through the D-site. Further progress will focus on understanding the biological implications of these interactions.

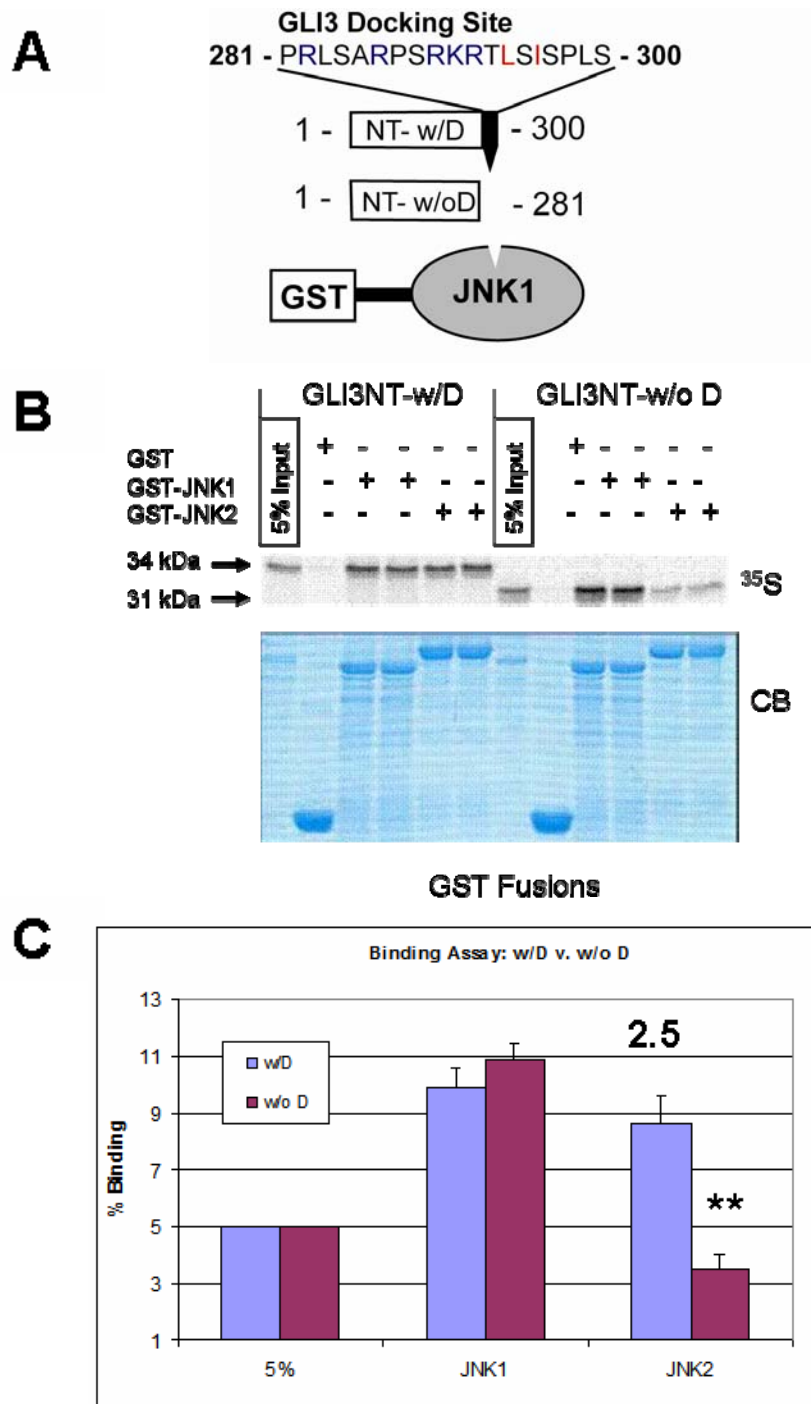


Figure 2: A) Schema for GST pull down assay between JNK1 and the two N-terminal (NT) fragments of GLI3 (w/D – with D-site, w/o D – without D-site). The sequence of the predicted D-site is shown. B) Image of GST pull down assay (above-phosphoimage, below-Coomassie Blue stain), Lanes: 5% – positive control; GST–negative control, 40ug; GST-JNK1–30ug; GST-JNK2–30ug. C) Quantification of GST pull down assay (n=4). ** = $p < 0.005$.