Ajit Elhance
12/14/2015
Final Project
Bio 131

Final Project Write-up

Building on work from my thesis, I will be comparing Pin2/TERF1-interacting protein (PinX1) sequences from mouse, *Xenopus laevis*, and humans. I will do both DNA and amino acid alignments and compare the results to PSIPRED secondary structure predictions. This will hopefully provide some indication of conservation of sequence and domain structure, which may explain previously observed cross-species inhibition effects.

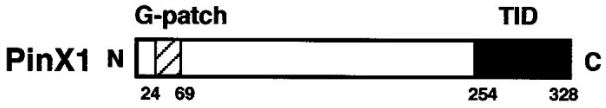

Figure 1: Domain structure of human PinX1.
hPiNX1 contains an N-terminal Glycine rich region and a telomerase inhibitory domain near the C-terminal. Reproduced with permission from Zhou and Lu, 2001.

In order to determine if there were sequential or structural similarities between different PinX1 orthologs, sequence data was obtained from the NCBI database. The DNA and amino acid sequence data was compiled into text files in FASTA format to make them easier to access. File names were trimmed down in order to easily access them in my program, as the original names were too long and difficult to call effectively.

In order to obtain the secondary structure predictions I used the online UCL-CS Bioinformatics PSIPRED protein sequence analysis tool to analyze the protein sequences of different orthologs. I then manual copied out the structure sequence output (C's, H's, or E's depicting the presence of coils, helices or strands respectively) into a new file in a format similar to FASTA. This enabled me to include the PSIPRED structure sequences to my Python file.

After gathering the data, the next step was to flesh out the programs needed to compare the sequences. The Global alignment functions written for DNA and amino acid sequence alignment in Homework 7 provided most of the code that I used. I made small modifications to the programs to get them to work more efficiently, such as including a scoring matrix for DNA alignments and developing a way to visualize matches vs mismatches. I also made a slightly edited version of my amino acid alignment program for use in PSIPRED sequence analysis. This involved creating a new matrix for scoring the C's, H's and E's of the PSIPRED sequence.

DNA sequences were compared first using the global alignment program. The output was very similar to BLAST results, showing a sparsely conserved upstream region and C-terminus. Matches in the C-terminus of mammalian PinX1 and xPinX1 seem to be mostly random, while all species shared a decently conserved N-terminal.

Ajit Elhance
12/14/2015
Final Project
Bio 131

```
Mouse PinX1 vs Xenopus laevis PinX1
Amino Acid Alignment:


MAMLAEPRRKQKWSVDPRNSTWSKDESKFGQKLMEKMGWSKGKGLGAKEQGSTEHIKVQVKNNNLGLGASRNHQDNWLAH
| |||| |||||| |||||  || | ||||||  ||||||||||||| ||| |||||| |||| |||||  |  ||| ||
MSMLAERRRKQKWTVDPRNTAWSNDDSKFGQKMLEKMGWSKGKGLGAQEQGATEHIKVKVKNNHLGLGATNNNEDNWIAH

QDDFNELLAELNNCHGTAESEESPAD-EEKKSFSLEEKSKSSKKRVHYMKFAKGKDLSSRSDTDLACIFGKREKTKKGGP
||||| ||| || |||  |   | |  ||||||||||||| || ||||||| ||||||||| ||| ||||||   ||
QDDFNQLLAALNTCHG-QETADS-SDKKEKKSFSLEEKSKISKNRVHYMKFTKGKDLSSRSETDLDCIFGKR-RNKKLA-

QEEASAESEENEDQGKQSPSGECDPGNTVTSSLSVNEYFAKRMAELKKSQSKHMGKVPQKEEECREEGADESSTKISQKS
|   |  |    |      |       | ||     |||||||| |     |   | ||        | | | |    |
QDGCS-NS--SADE-VNT-S--LTTTTTTTSAFTIQEYFAKRMAQL-----KN--K-PQ--ASA--PGSDLSETPVERK-

KKRKRNKGDTIHDT-NDPDLDNGDLPCVTEVPQSKKKSKTKKYREPESE-RVGRKDEEE-EDGA-SHQNDC-RT-E-EEE
|  | ||     |  | |                              || |    |   |   | |  |       ||  |
KGKKKNKEAADTDVENSPQ-HKAKRHKKKKRVEAERGPVAKK-RD-RAELQPGGPSEDECSDASVEAAEDCVQTPDIQDD

KPKKRKKKNKKRKNETSNQWEEEENSQTQQLKKKKKKNKSE
 ||  | | || |       |          ||||||  |
VPKPKKRKAKK-KLQRPEGVEIDATLDRAPVKKKKKK-VSR


Percentage Match: 49.31 %
```

Figure 2. Example of amino acid alignment from mPinX1 vs *Xenopus laevis* PinX1. Vertical dashes indicate a match between the two sequences. Percentage match for this example is 49.13%. Percentage match is calculated as '# matches/Alignment Length'.

Next, amino acid sequences were analyzed. Despite low DNA sequence alignment scores and percentage of matches, the amino acid sequences of the various species aligned well. As expected the mouse and human orthologs aligned better to one another than to the *Xenopus laevis* PinX1.

```
Mouse PinX1 vs Xenopus laevis PinX1
PSIPRED Alignment:

CCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHCCCCCCCCCCCCCCCCCCCCEEEEEECCCCCCCCCCCCCCCCC
||||     |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
CCCCHHHHHCCCCCCCCCCCCCCCCCCCCHHHHHHHHHCCCCCCCCCCCCCCCCCCCCEEEEEE-CCCCCCCCCCCCCCCCH

HHHHHHHHHHHHHCCCCCCCCCCCCHHHHHCCHHHHHHHHCCCCCCEEEEEECCCCCCCCCCCHHHHHHCCCCCCCCCCC
|||||||||||||| |||||||||| |||| ||||||||| ||||||  |||||||||||||||||||      ||||||
HHHHHHHHHHHHHHCCCCCCCCCCCHHHHHCCCHHHHHHHCCCCCCCCCCEEEECCCCCCCCCCCHHHHHH------CCCCCC

CCCCCCCCCCCCCCCCCCCCCCCCCCCEEECCCHHHHHHHHHHHHHHHHHCCCCCCCCCCHHHHHCCCCCCCCCCCCCCHH
|||||||||||||||||||||||  ||   ||||||||||||||||   ||||||||||  |||    ||||||||||
CCCCCCCCCCCCCCCCCCCCCCCCCEECC---CCCHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCHHH--HHCCCCCCCCCC--

HHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHCCCCCCCCCCCCCCCHHHHCCCCCCCCCCCCCCCCCHHHHHHH
   |||||||||     ||||      |||||||||||||||||||||||  ||||||||||||||||||||    |||
---CCCCCCCCCHHHHHCCCCHHHHHHCCCCCCCCHHHHCCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCCCC---CHHH

HHHHHCCCCCCCCCCCHHHHHHHHHHHHHCCCC
|||||||||||||||       |||||| |||
HHHHHCCCCCCCCCCCCCCCCCCCHHHHHH-CCC


Percentage Match: 82.15 %
```

Figure 3. Example of PSIPRED alignment from mPinX1 vs *Xenopus laevis* PinX1. Vertical dashes indicate a match between the two sequences. Percentage match for this example is 82.15%. Percentage match is calculated as '# matches/Alignment Length'.

Ajit Elhance
12/14/2015
Final Project
Bio 131

PSIPRED analysis predicted similar structures for all of the orthologs (>80% for all comparisons). The N-terminals of all of the proteins is very highly conserved. While again the human and mouse proteins are more similar to one another than to *Xenopus*, each of the PinX1 orthologs share several predicted helices at the C-terminus, which is where the telomerase inhibitory domain is thought to be located.

Experiments with truncated versions of hPinX1 and xPinX1 found that only the C-terminal region was involved in telomerase inhibition. This makes this region the focus of analysis when looking for possible conservation between PinX1 orthologs. Based on the DNA sequence alignments, the mouse and human genes exhibit some sequence similarity, but even in these closely related organisms (relative to *X. laevis*), the C-terminal region does not appear to be very well conserved; C-terminal regions actually appear to be better conserved between humans and *X. laevis* than between humans and mice. In addition to this, the peptide sequences for comparisons between mammals and *X. laevis* displays very little alignment, not much more than would be expected by random chance. This would suggest that the proteins are quite divergent from one another, and that a cross-species inhibition effect would be unlikely.

Secondary structure analysis with PSIPRED does provide some hope though. The PSIPRED alignments of the C-terminal regions of the proteins seem to indicate similar secondary structural characteristics of the proteins. This is one of the most interesting findings from this project. While the amino acid sequences differ significantly in this region, structural similarities may account for at least a part of the observed inhibition effect. Ultimately, more information about the specific residues needed for telomerase inhibition is needed before conclusions can be made about cross-species inhibition, but this analysis may be helpful in dissecting whatever mechanisms are at play.