

Assignment 2 – Using Online Molecular Biology Resources

Start: Week 4

Due: Week 6

Credit Weight: 15%

Complete the following activities using the online resources indicated below. Your answers should be prepared in full sentences. Proofread a printed version of your submission for both spelling and grammatical errors since there will be deductions for these.

By the end of this assignment you should be able to:

- Appreciate the information that can be recovered easily via OMIM and GenBank resources
- Compare gene and cDNA sequences
- Carry out a BLAST analysis
- Generate a strategy to synthesize a riboprobe.

You should not start this assignment until you have completed the GenBank task.

Imagine you have just started your Co-op work term in a lab. Your supervisor, a traditional marine biologist who has spent the last 25 years studying dolphins, is interested in a connection between learning abilities and mood swings in sea mammals. He explained to you that contrary to general public understanding there are individual dolphins that are not always very friendly (he calls them “moody dolphins” and according to him “... they go from playful to almost vicious in a second...”). Over the years he has noticed that the “moody dolphins” are able to learn new tricks really fast; they seem to be “smarter” than their mates. He believes that this trait is partially due to the different levels of expression of dopamine receptor D2 (DRD2) in the brains of “moody dolphins”. Your project is to initiate a study of the transcription of the gene for the dopamine receptor by northern hybridization using a DIG-labeled riboprobe. Unfortunately, nobody in the lab has ever performed northern blotting before; they are not really good with molecular biology techniques and they consider you an expert. All they have is the complete cDNA for human DRD2. When you asked for information about it, your supervisor told you that he received the cDNA as a present from Dr Kenneth K. Kidd from the Yale University School of Medicine, who submitted the sequence to the GenBank in 1998. According to the letter that came with the cDNA, the GeneBank accession number for the sequence is [AF050737](#).

Your supervisor wants you to find out the current research interests of Dr. Kenneth Kidd. But your priority is to investigate the full sequence of DRD2 gene, to determine the number of exons and introns it contains and to find out if there is a polyadenylation signal in the cDNA. Finally, he wants to know if the cDNA has a Kozak sequence. He has a few questions regarding the DRD2 protein, including the exact size in kDa. To your surprise, your supervisor suggests that you look for regions of nucleotide identity between human DRD2 cDNA and a puffer fish (*Fugu rubripes*) DRD2 cDNA. He gave you an accession number for the puffer fish sequence: [X80175](#).

Everything is now up to you. You even have to find and purchase the suitable vector for making a probe. The final goal is to make the DIG-labeled antisense riboprobe from human DRD2 gene.

Part I: PubMed at NCBI

You first have to locate a publication of a particular scientist. To do this, go to the PubMed <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi> and enter the Dr. Kenneth Kidd's name in the query box, using the correct formatting for the query.

1. How many total publications has Dr. Kenneth Kidd written based on the information contained in PubMed?
2. How many articles did Dr. Kenneth Kidd publish in 2004?
3. What is his current research interest based on his publications in 2005?
4. Briefly explain in 100 words or less the role of human DRD2. (You might find OMIM useful here.) You query OMIM easily from the NCBI homepage:

The screenshot shows the NCBI homepage. At the top left is the NCBI logo. To its right is the text 'National Center for Biotechnology Information' with 'National Library of Medicine' and 'National Institutes of Health' below it. A navigation bar contains links for 'PubMed', 'All Databases', 'BLAST', 'OMIM', 'Books', 'TaxBrowser', and 'Structure'. Below this is a search bar with 'OMIM' in a dropdown menu, 'for' text, and 'DRD2' in a text input field, followed by a 'Go' button. On the left side, there is a 'SITE MAP' section with links for 'Alphabetical List' and 'Resource Guide', and an 'About NCBI' link. To the right of the search bar, there are two main sections: 'What does NCBI do?' with a sub-link 'Established in 1988 as a national resource for molecular biology information, NCBI creates' and 'Hot Spots' with sub-links 'Assembly Archive' and 'Clusters of'.

5. Find the article published by Rocheville M. *et al.* in Science, vol. 288(5463): 54-7. **Print the first page of the pdf version of this article.**

5a) How many other articles have cited this article (see below)?

The screenshot shows a sidebar of links for a PubMed article. The links are: 'Abstract of this Article', 'Full Text (HTML) of this Article', 'E-Letters: Submit a response to this article', 'Download to Citation Manager', 'Alert me when: new articles cite this article', 'Search for similar articles in: Science Online, ISI Web of Science, PubMed', 'Search Medline for articles by: Rocheville, M. || Patel, Y. C.', and 'Search for citing articles in: ISI Web of Science (260)'.

5b) Choose three of the articles published in 2005 and submit the titles, authors and the names of the Journals in which they were published.

Find the GenBank record for [AF050737](#)

6. Scroll down the page and find the mRNA record (click on mRNA). Print it. What is the size of mRNA in nucleotides?
7. What is the GI number?
8. Find the open reading frame (CDS). Click on it. CDS is from nucleotide_____ to nucleotide_____ in the mRNA.
9. a) Find a polyA signal. PolyA signal is from nucleotide_____ to nucleotide_____ in the gene and
b) from nucleotide_____ to nucleotide_____ in the mRNA.
10. What is the role of a polyA signal?
11. a) Have you managed to find the Kozak sequence? If you did, label it on the mRNA print-out.
b) What is the role of a Kozak sequence?
12. What is the PID for the corresponding protein sequence of this record?
13. What is the size of the protein in KDa? (Try using the following web site: http://pir.georgetown.edu/pirwww/search/comp_mw.shtml)
14. What is the major difference between gene variations 7423 and 22640?
15. What is "gene variation"?
16. What is the GenBank common name for the taxonomy ID "9606"?
17. Compare the record for the CDS with the record for the mRNA.
 - a) Explain the differences in the sequences between these two records.
 - b) Which sequence does your cDNA resemble most, the CDS or the mRNA?
 - c) Explain your reasoning.
18. a) How many exons does the human DRD2 contain?
b) What kind of information (coding? junk? regulatory?) does exon 8 contain?
19. Download (or make your own) schematic, labeled graph of the features of biological significance (introns, exons, 5' and 3' UTRs etc.) for AF050737.

Now perform a **pairwise BLAST analysis** of human DRD2 **cDNA** sequence and the puffer fish DRD2 **cDNA** sequence ([X80175](#)). For the comparison you have to use Pairwise BLAST -BLAST 2 sequences, at <http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>. Use the default settings for the alignment.

20. Explain why did you have to compare cDNA sequences and not genomic sequences of human and puffer fish DRD2.
21. Attach the printed result from your BLAST analysis to your submission.
22. Describe your results in two sentences or less, including the regions of sequence identity between the two sequences (i.e. indicate which nucleotides of each sequence align).

23. a) Why did you think your supervisor asked you to perform this comparison of human and fish cDNAs if he is interested in dolphins?
b) Considering the degree of similarity shown in your pairwise Blast analysis of the puffer fish and the human DRD2 sequences, do you think it is feasible to use your human cDNA as a probe for corresponding dolphin transcripts? Explain your reasoning.
c) Would you have to think about stringency conditions for your northern hybridization and washes? Explain your reasoning.

Part II: Using a program for restriction enzyme digestion

Go to <http://tools.neb.com/NEBcutter2/index.php>

This program accepts sequence queries written in FASTA format (check the display options of the NCBI record).

Paste the FASTA sequence for **human DRD2 cDNA** into the sequence box.

Choose "All commercially available specificities".

24. Print the record.
25. Print the list of enzymes that cut the sequence only once.

Part III: Buying a vector from ATCC

Go to the ATCC web site <http://www.atcc.org/>

26. Explain the role of the ATCC in few sentences.

Choose **All collections** from the pull-down search menu. Click on **Vectors**.

27. What is the difference between pUC18 and pUC19 vectors (pay attention at the enzymes from the MCS)?
28. What is the difference between pBAD18 and pCMVbeta? (**Hints**: check the promoters that they contain; what does CMV stand for? Draw your own conclusions about systems in which they could be expressed.)

Choose pGEM-7Zf(+)/LIC-F vector.

29. Print the map.
30. a) What is the ATCC accession number for pGEM-7Zf(+)/LIC-F?
b) What is the price of the vector?
31. How is the vector shipped?
32. What is the suggested host for this vector?
33. Given how the vector is supplied from ATCC you will not have enough vector DNA to use immediately for your experiments. How do you propose to get more of it? (**Hint**: think about the answer for the question 32.)
34. How many RNAP binding sites (promoters) does this vector have?
35. Why is the information about the number of RNAP promoters important to know when you plan to make an antisense riboprobe?
36. Make a list of the restriction enzymes found in the **vector's MCS**. Include the sequence of each restriction enzyme recognition site and their locations in the plasmid. (You could use the NEBcutter website to make and print the table.)

Part IV: Developing a strategy for cloning and making a riboprobe

37. Keep in mind: you do not have to make a probe from the full-length cDNA. Theoretically, 20 nucleotides should be unique enough, but you want to try to make as long probe as possible. Explain why.

Develop a cloning strategy: you have to check the NEBcutter record for human DRD2 cDNA and the vector printout with unique enzymes from MCS in order to find two restriction enzymes in the MCS that would also cut DRD2 cDNA, generating as long a fragment as possible.

Explain your cloning strategy by answering the following questions:

38. a) Which restriction enzymes would you use to clone the fragment of human DRD2 cDNA into the pGEM vector? (**Hint:** there are recognition sites for two commonly used restriction enzymes present in both cDNA and vector's MCS; you could use them for directional cloning and you would have only one possible orientation for making an antisense riboprobe).

b) Circle these enzymes on the NEBcutter record for DRD2. Label them as 5' (closer to the beginning of the cDNA CDS) and the 3' (closer to the end of cDNA CDS) ends.

39. a) Which RNAP would you use to make the antisense riboprobe?

b) Which enzyme would you use to linearize recombinant plasmid? (**Hint:** if you followed the hint given to you in question 38, you could make use of one of the restriction enzymes utilized in cloning.)

40. What will the expected (approximate) size of the riboprobe be? (Express your answer as a number of nucleotides.)