



**CALIFORNIA STATE SCIENCE FAIR
2005 PROJECT SUMMARY**

Name(s) Brad J. Jensen	Project Number J1210												
Project Title RNA Alignment Scoring Based on Covariance													
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The objective of my project was to test a new method of comparing RNA sequences. The new method uses covariance to see how similar the 2 sequences are. I will look at the results of the tests, and other aspects such as time needed to make the test (It will be compared using a program I wrote) to see if this new method is valid.</p> <p>Methods/Materials The purpose of my project is to test a new method of RNA comparison to see if it#s accurate. My hypothesis was that it would prove to be a very effective method of comparing RNA sequences. The materials I used for this project were: Fedora Core 3 for 64 bit computers, VI (Which comes with Fedora Core 3), PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi), ClustalW (http://www.ebi.ac.uk/clustalw/), A program written by me for the comparison tests.</p> <p>Results When I ran the sequences through my program (for example, Sequence a1 and Sequence a2 I got these results for both comparison methods (The base comparison method results are show first, then the Covariance comparison results are shown):</p> <table border="0"><tr><td>Sequence A</td><td>23.349056%</td><td>53.949463%</td></tr><tr><td>Sequence B</td><td>33.141212%</td><td>50.033535%</td></tr><tr><td>Sequence C</td><td>33.163483%</td><td>52.011089%</td></tr><tr><td>Sequence D</td><td>20.452732%</td><td>Failed</td></tr></table> <p>Conclusions/Discussion By looking at the results I have concluded that the new comparison method does not produce the same results and that it is not a valid form of comparison. I have reached this conclusion by observing that the covariance comparison results are all in the same general area when compared to the standard comparison results. Another factor that makes this new comparison method invalid is that it finds every spot the RNA sequence could co vary with itself. So if a sequence like gcgc were used, it would score 100% covariance similarity to the sequence tata, when in reality they have nothing in common (Other than they both only use 2 bases#) This method is also very inefficient in comparing RNA strands because the buffer that holds where the RNA sequence co varies with its self grows cubically. That makes it very hard for a standard computer to compare long sequences (I tried a 30,000 length sequence and it failed.) This method also takes a very long time to compare the RNA sequences. My conclusion is that this new comparison method is invalid in all respects. This goes against my hypothesis which believed that it would have some validity</p>		Sequence A	23.349056%	53.949463%	Sequence B	33.141212%	50.033535%	Sequence C	33.163483%	52.011089%	Sequence D	20.452732%	Failed
Sequence A	23.349056%	53.949463%											
Sequence B	33.141212%	50.033535%											
Sequence C	33.163483%	52.011089%											
Sequence D	20.452732%	Failed											
Summary Statement I was testing a new way to compare RNA with a program I wrote.													
Help Received Inspiration for the project came from my brother Jeff.													