1. Describe all the features you used for each model. Which features were the most useful?

I used 77 features total for the contextual model and the Viterbi model. I used 50 features for the non-contextual models. All the features used in the non-contextual models were used in the other two. For contextual features, I used n=1, which means that my contextual model was a Markov chain

For non-contextual features, I used the following features, in order of usefulness.

1. If the word contains Greek letters (alpha, beta, kappa, zeta, …)
2. If the word contains the string 'jun'
3. If the word contains the string 'endorphin'
4. If the word contains the string 'globin'
5. If the word is all lowercase and then a capital at end (e.g. 'syzX')
6. If the word is a combination of alphabet and numbers (e.g. 'MNP3')
7. If the word contains the string 'protein'
8. If the word contains string 'synthetase'
9. If the word contains substring '-rin'
10. If the word contains string 'kinase'
11. If the word contains string 'mutant'
12. If a vowel alphabet is not present
13. If a word has more than one capital letters
14. If a word contains substring '-ase'
15. If a word contains the word 'factor'

The rest of non-contextual features were less impactful.

For the contextual features, I used the following features, again in order of usefulness.

1. A single capital letter that does not appear in beginning of sentence
2. String 'gene' that follows a gene
3. A single capital letter that follows a gene
4. String 'enhancer' that follows a gene
5. String 'receptor' that follows a gene
6. A single capital letter that follows the word 'protein'
7. String 'transcript' that follows a gene
8. String 'element' that follows a gene
9. String 'chain' that follows a gene
10. Character '-' that follows a gene
11. Any word that follows '-' which is a gene
12. String 'factor' that follows a gene

13. String 'site' that follows a gene

The rest of the contextual features were less impactful.

2. Which model performs the best on the development data in terms of F1 score? Does it also have the highest Precision and Recall?

The contextual model performs best on the dev data. However, the Viterbi model has a slightly higher precision than the contextual model. The results are found in the table below

	Non-contextual	Contextual	Viterbi
Precision	0.482	0.538	0.553
Recall	0.286	0.349	0.331
F1	0.359	0.424	0.415

3. Provide scatter plots for precision, recall and F1 of the different models.