
EEOB563 – Assignment #3

Dennis Psaroudakis

February 20th 2019

I

1. a)

* A L I G N M E

* 0 -4 -8 -16 -24 -32 -40 -48

A -4

I -8

L -16

M -24

E -32

N -40

T -48

that's a
bit tiny, I'll
make it bigger

* A L I G N M E

A 0 -4 -8 -12 -16 -20 -24

I -4 2 6 10 14 18 22

L -8 6 8 4 1 3 0

M -12 8 9 6 4 9 7

E -16 5 6 4 6 8 14

N -20

T -24

Wait, that's not
right

* A L I G N M E

* 0 -4 -8 -12 -16 -20 -24 -28

A -4 4 0 -4 -8 -12 -16 -20

I -8 0 6 4 0 -4 -8 -12

L -12 -4 4 8 4 0 -2 -6

M -16 -8 0 5 5 2 5 1

E -20 -12 -4 1 3 5 1 10

N -24 -16 -8 -3 1 9 5 6

T -28 -20 -12 -7 -3 5 8 4

A A L I G N M E
A I L M E N T -

b) Yes, this is the best global alignment*

because it is built on all the smaller
local optima.

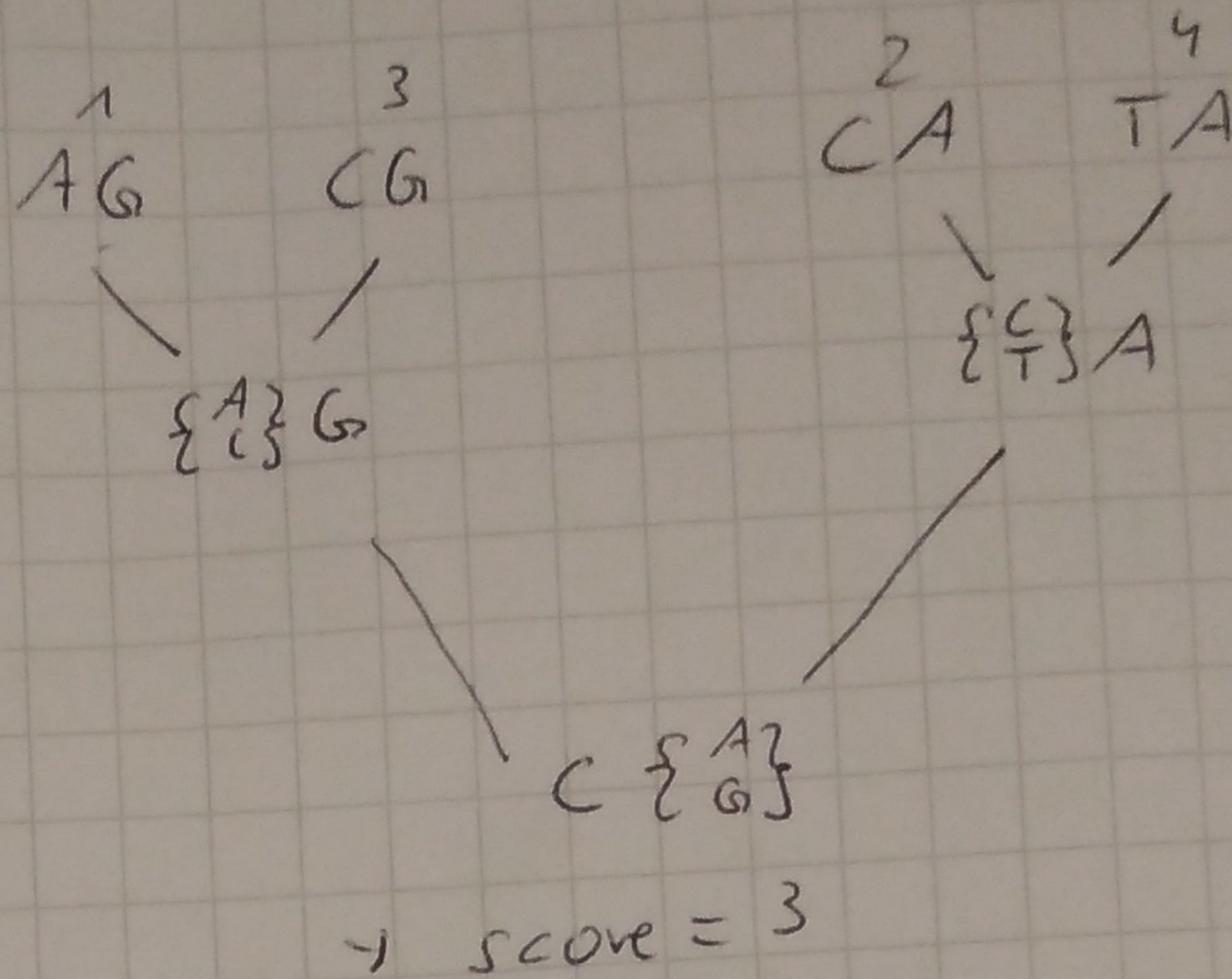
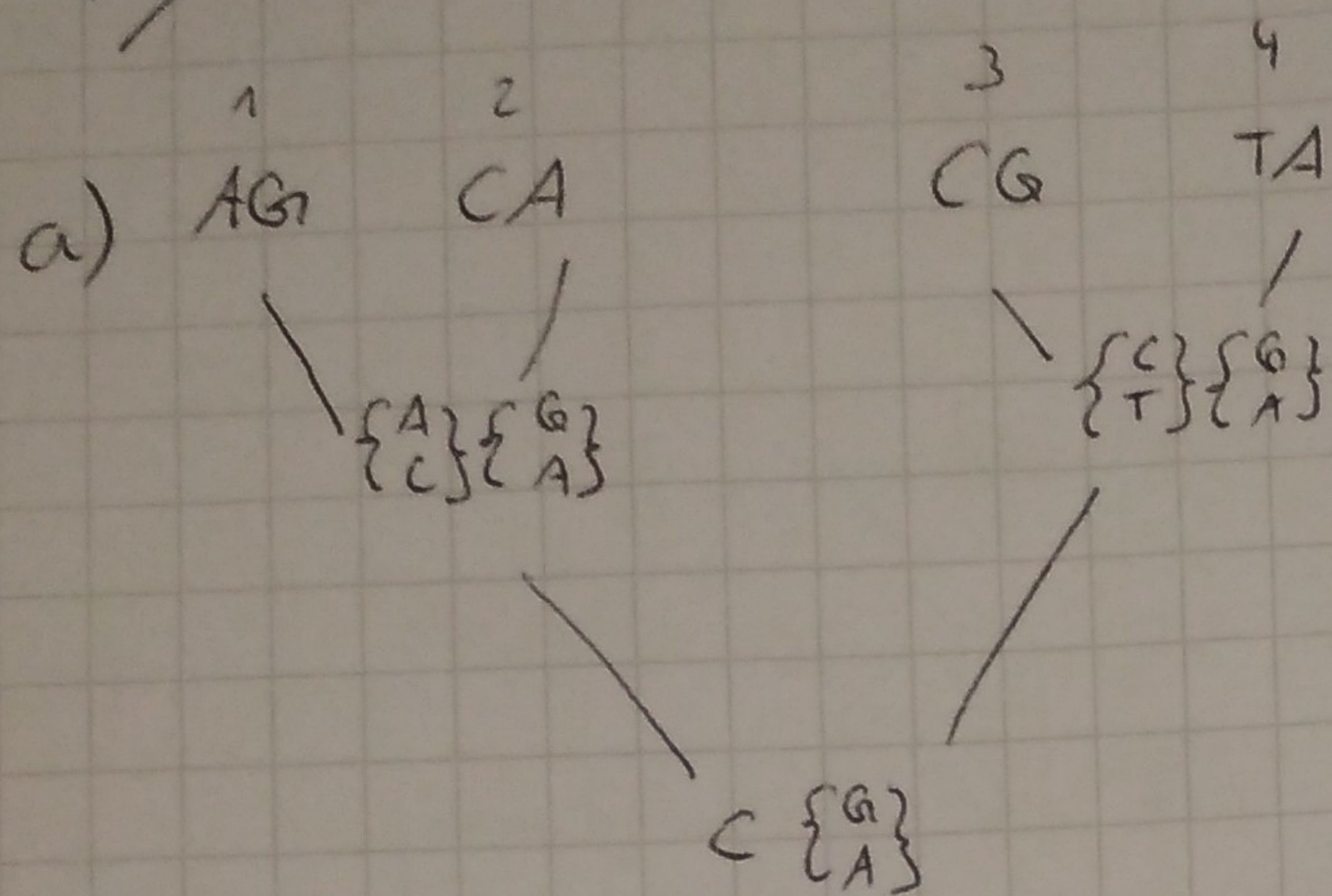
I still refuse to make an assumption about

homology just from the alignment. Etymo-

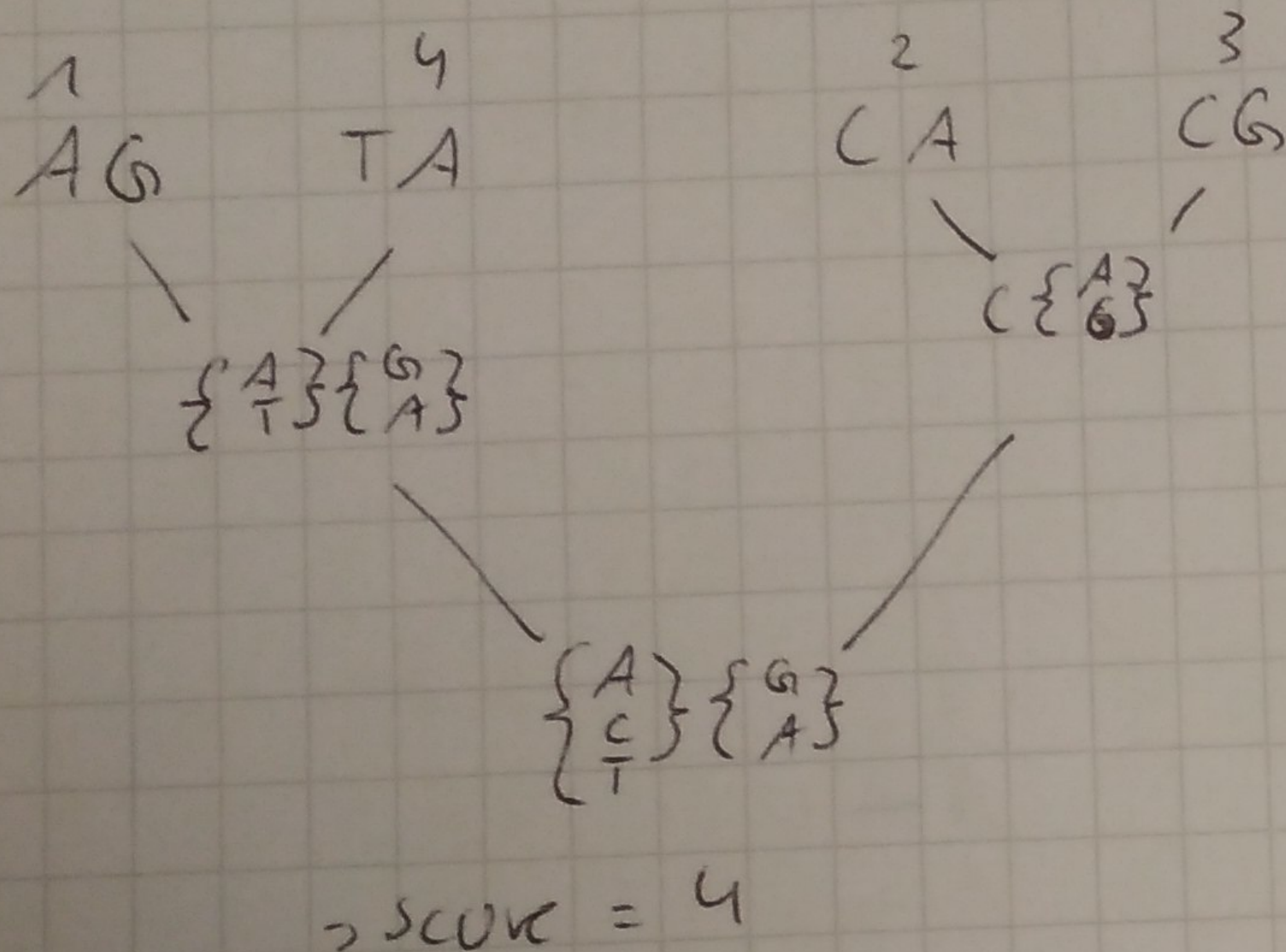
logically, the words are probably not related
except for the -ment suffix.

* given this score matrix (blosum 62)

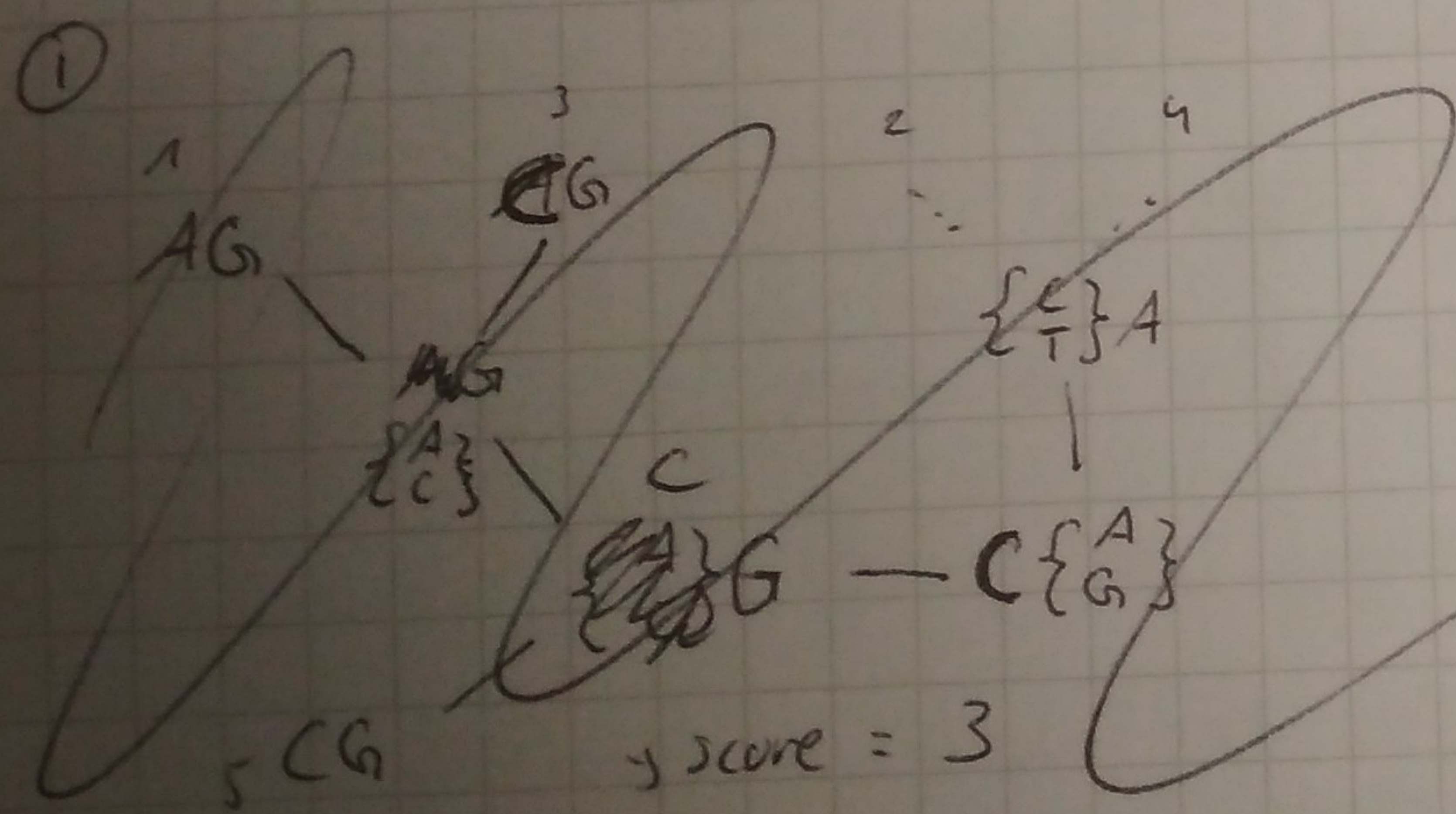
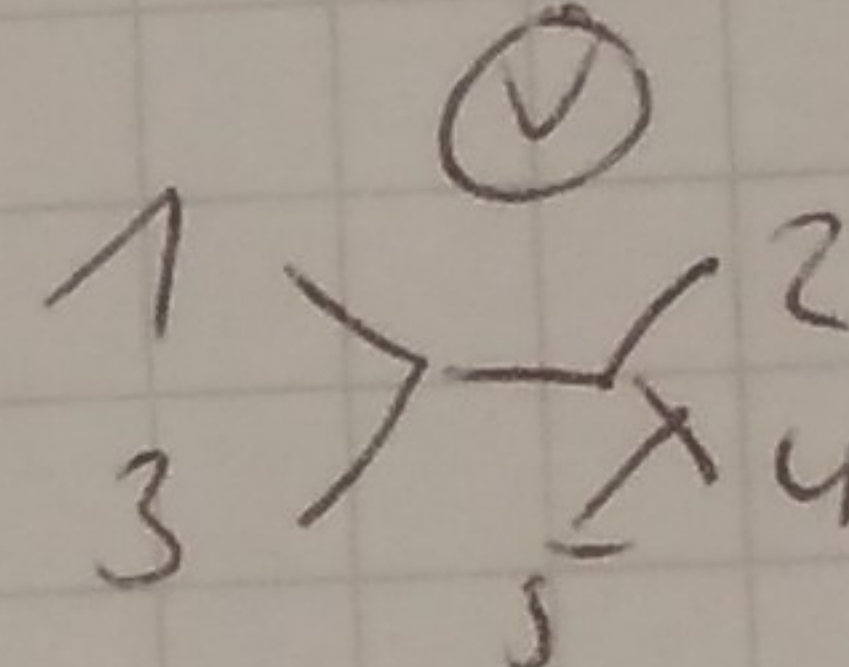
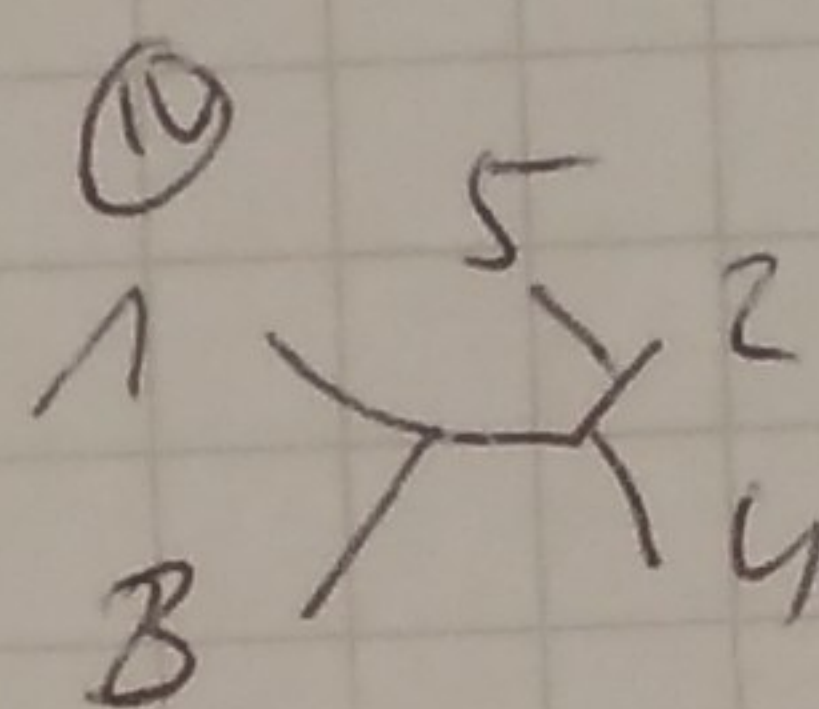
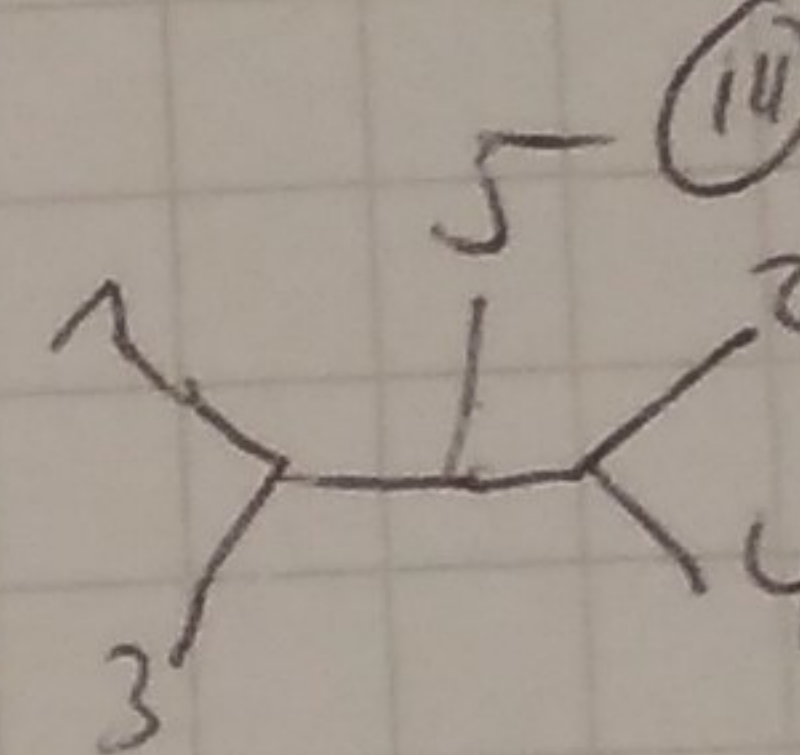
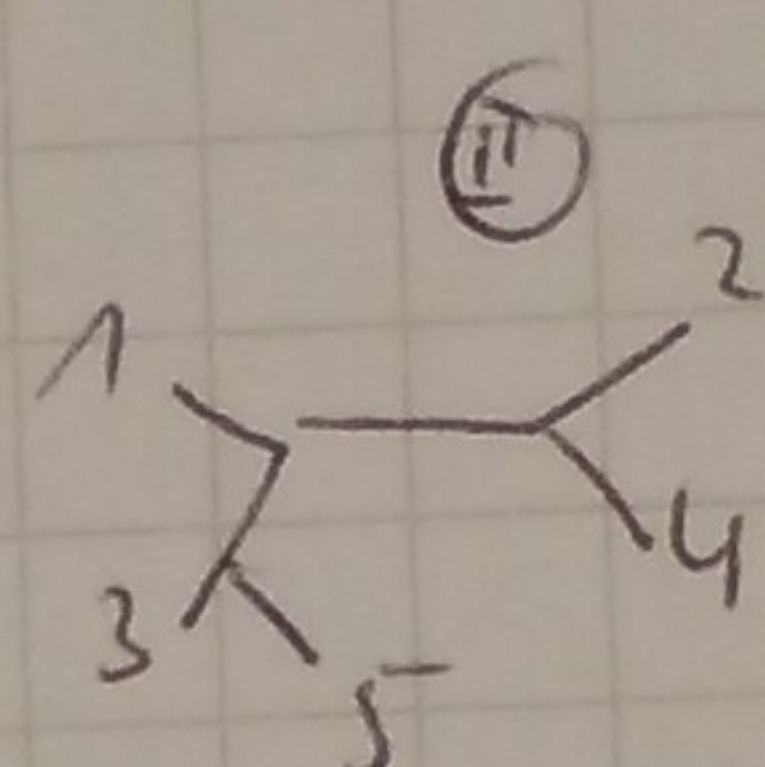
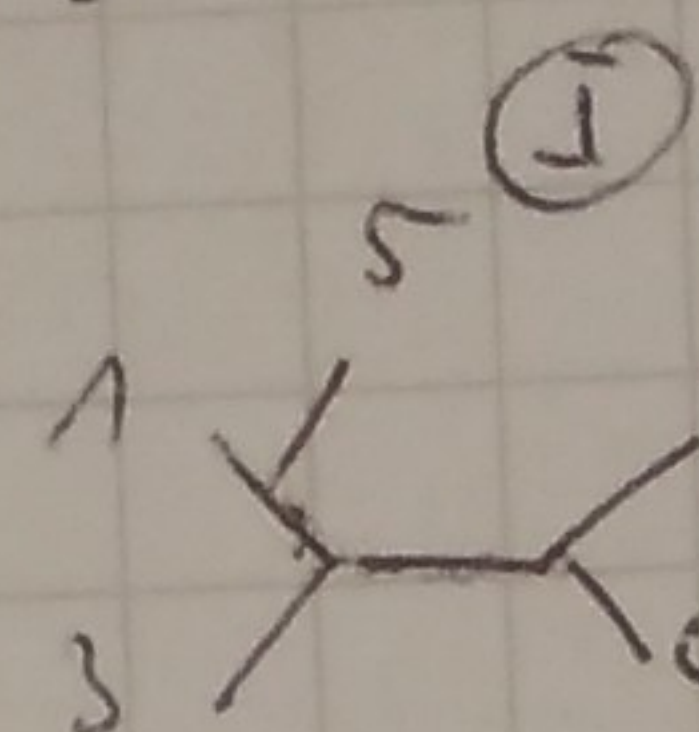
2/



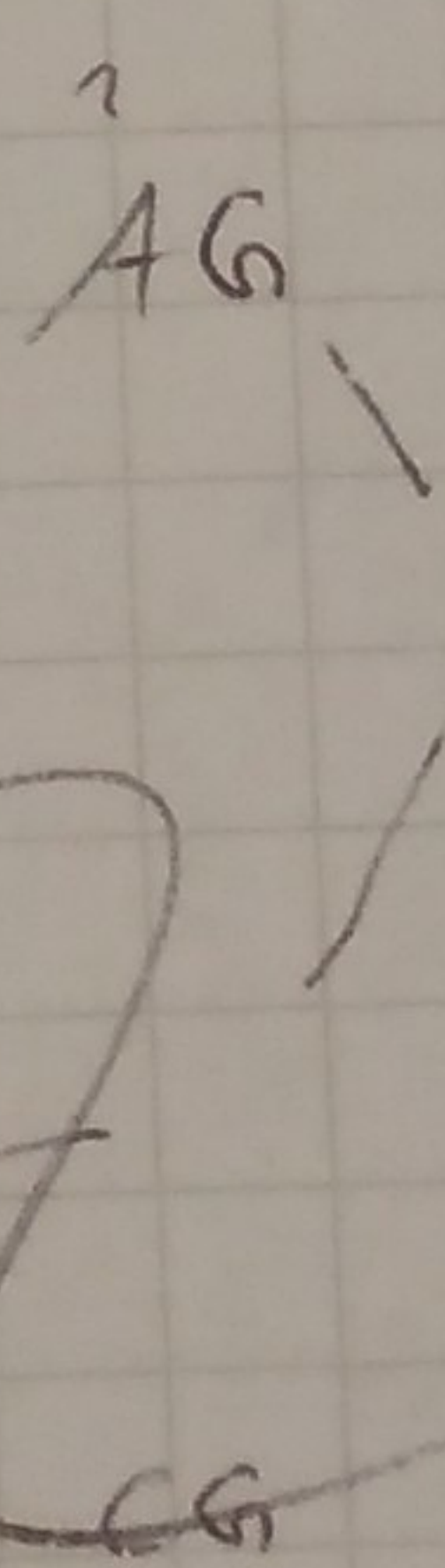
best tree.



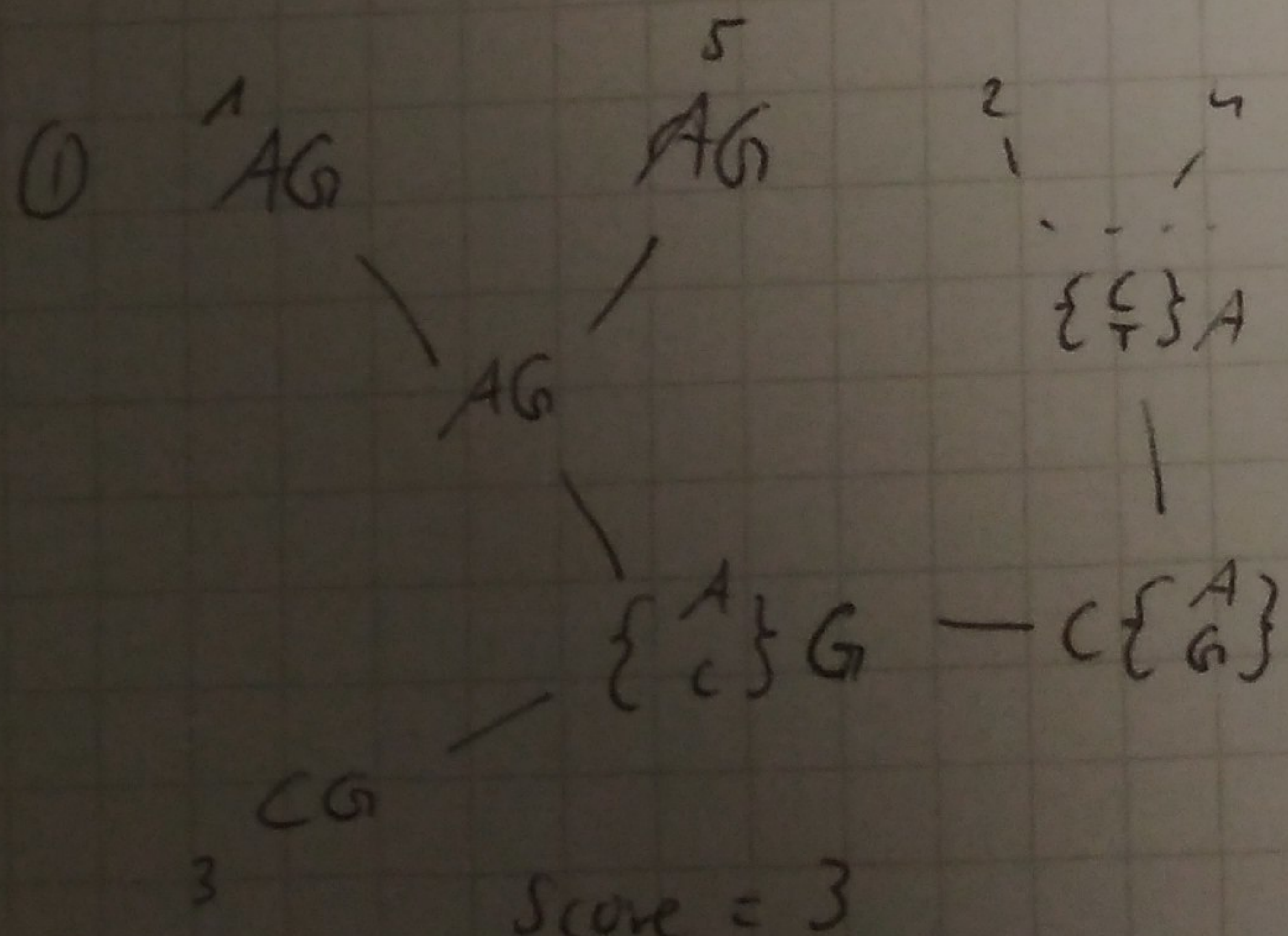
Now there are 5 new trees:



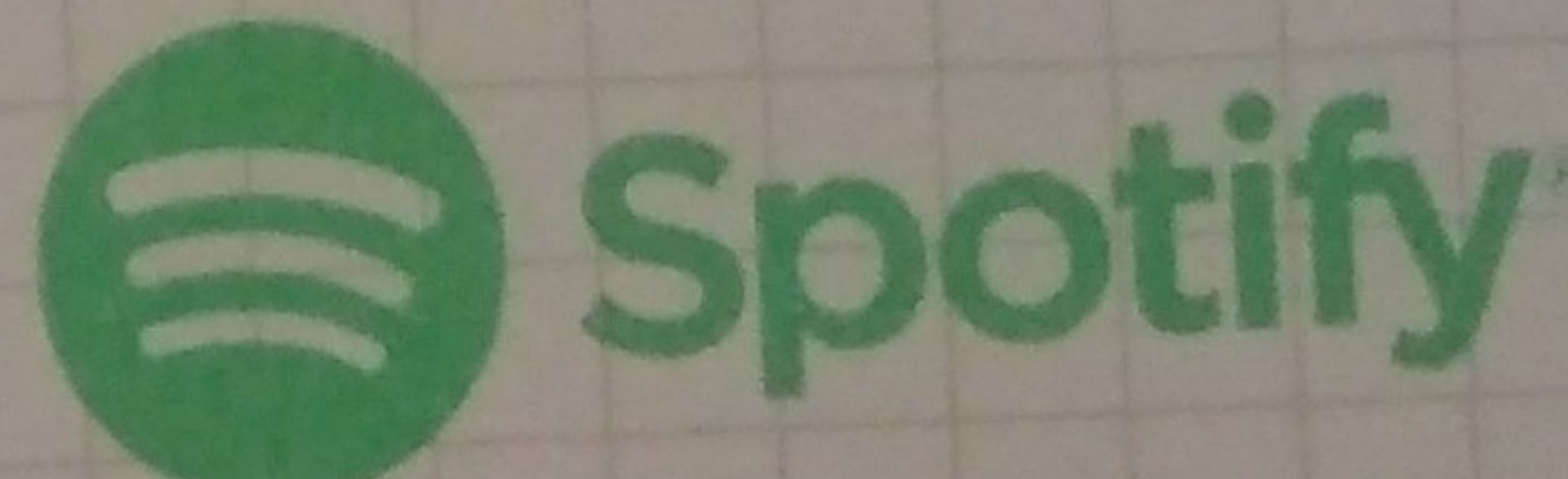
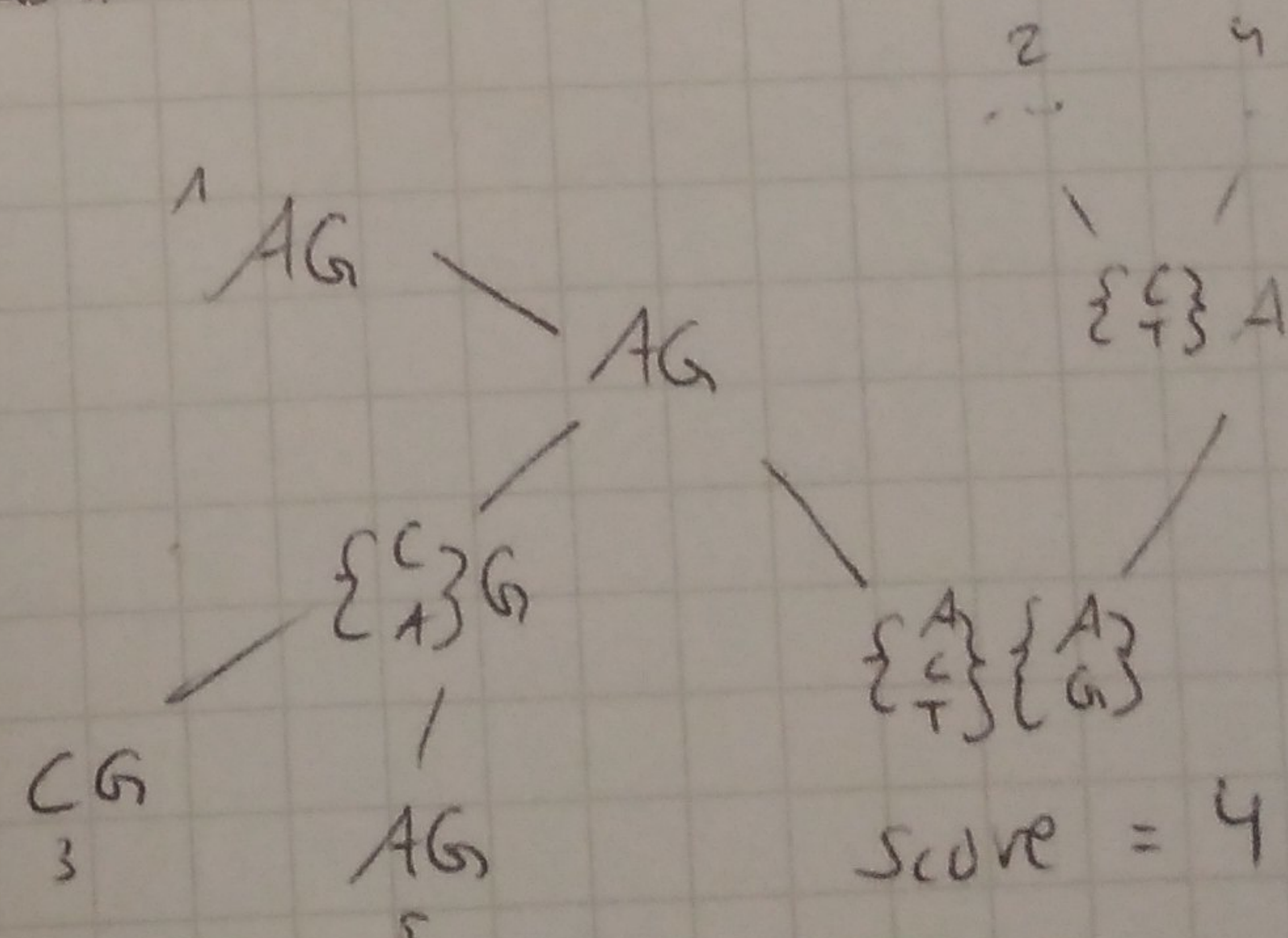
II



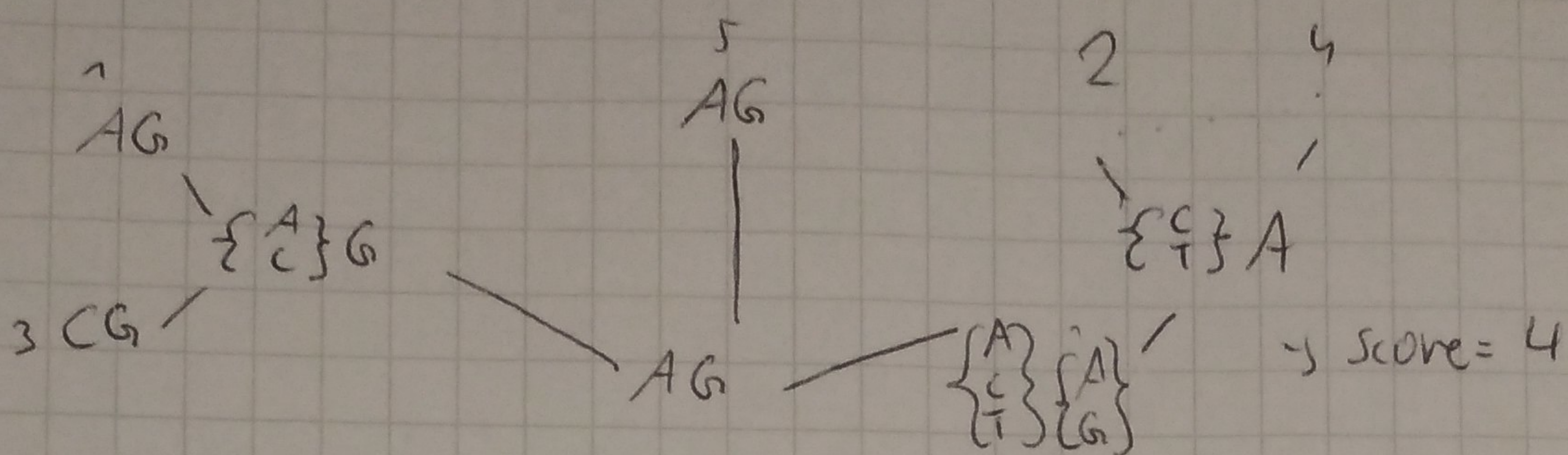
This is why I hate doing things by hand, I make so many mistakes!



II

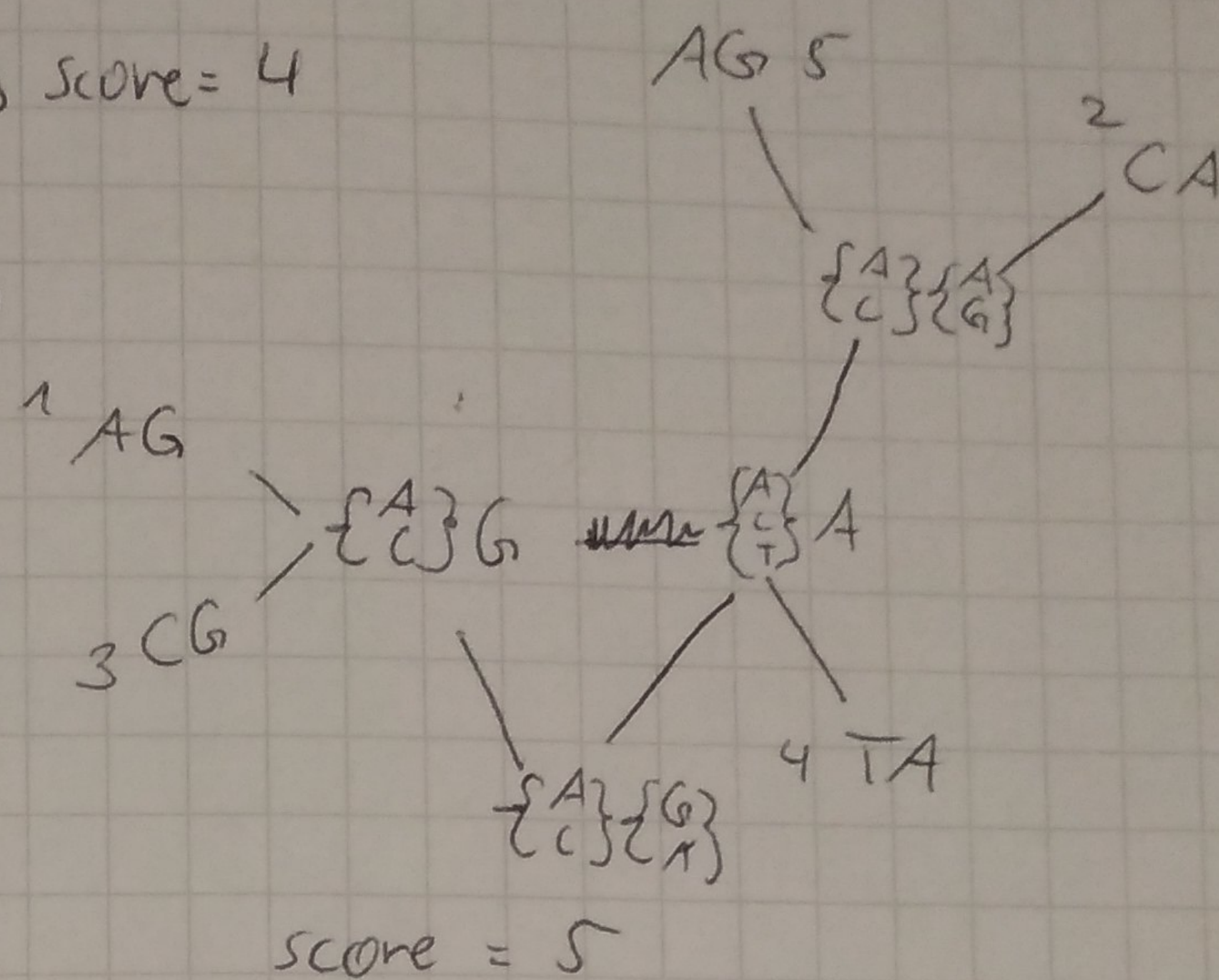


III

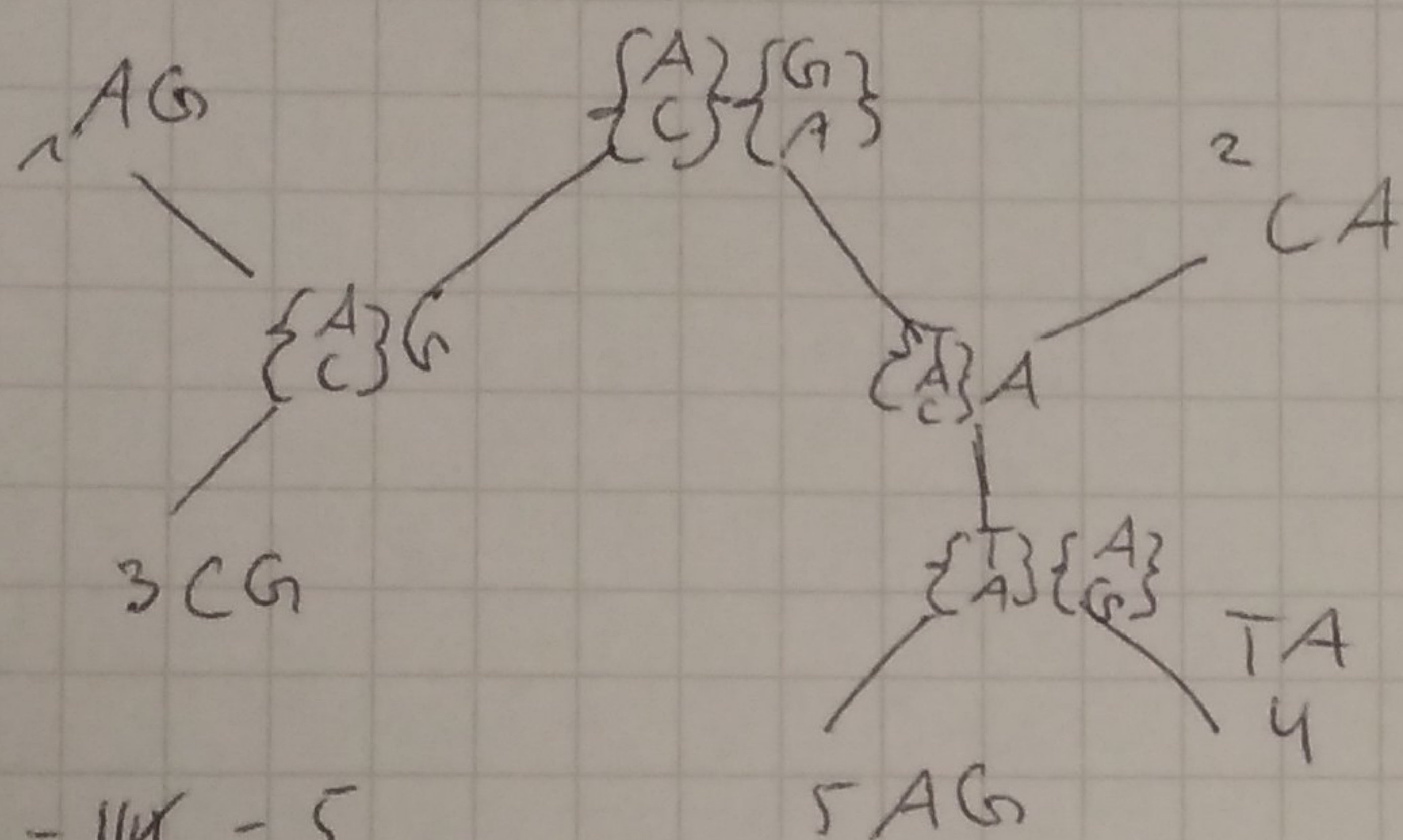


I have no time for this right now, maybe later, maybe not. It would take hours to score all the different rooted versions of these 5 trees, and where you root them makes a difference.

IV

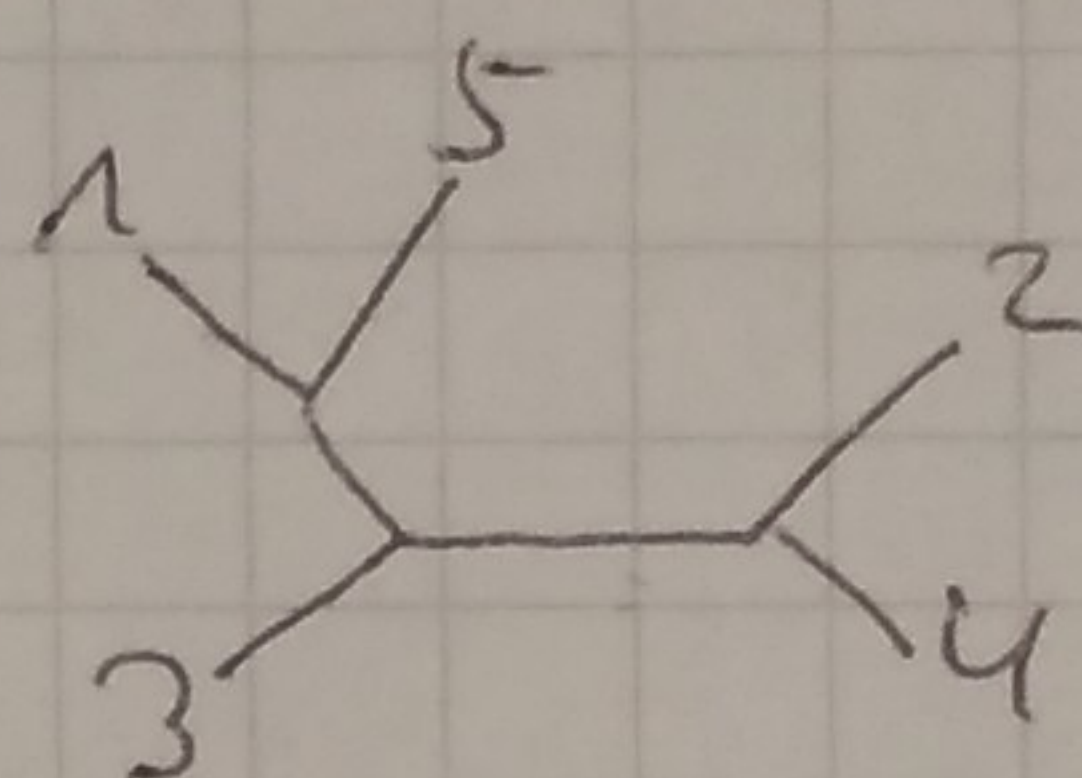


V



Score = III = 5

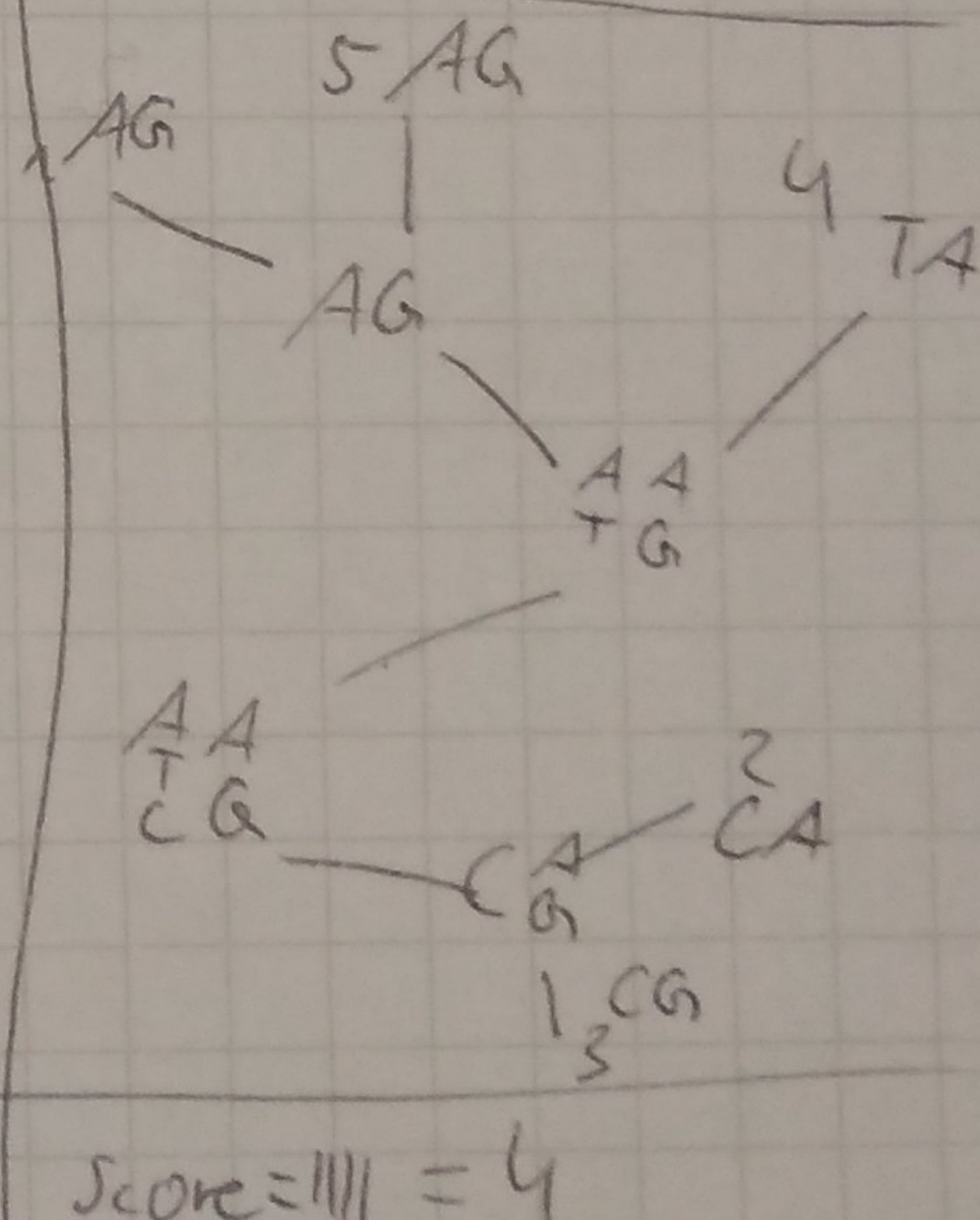
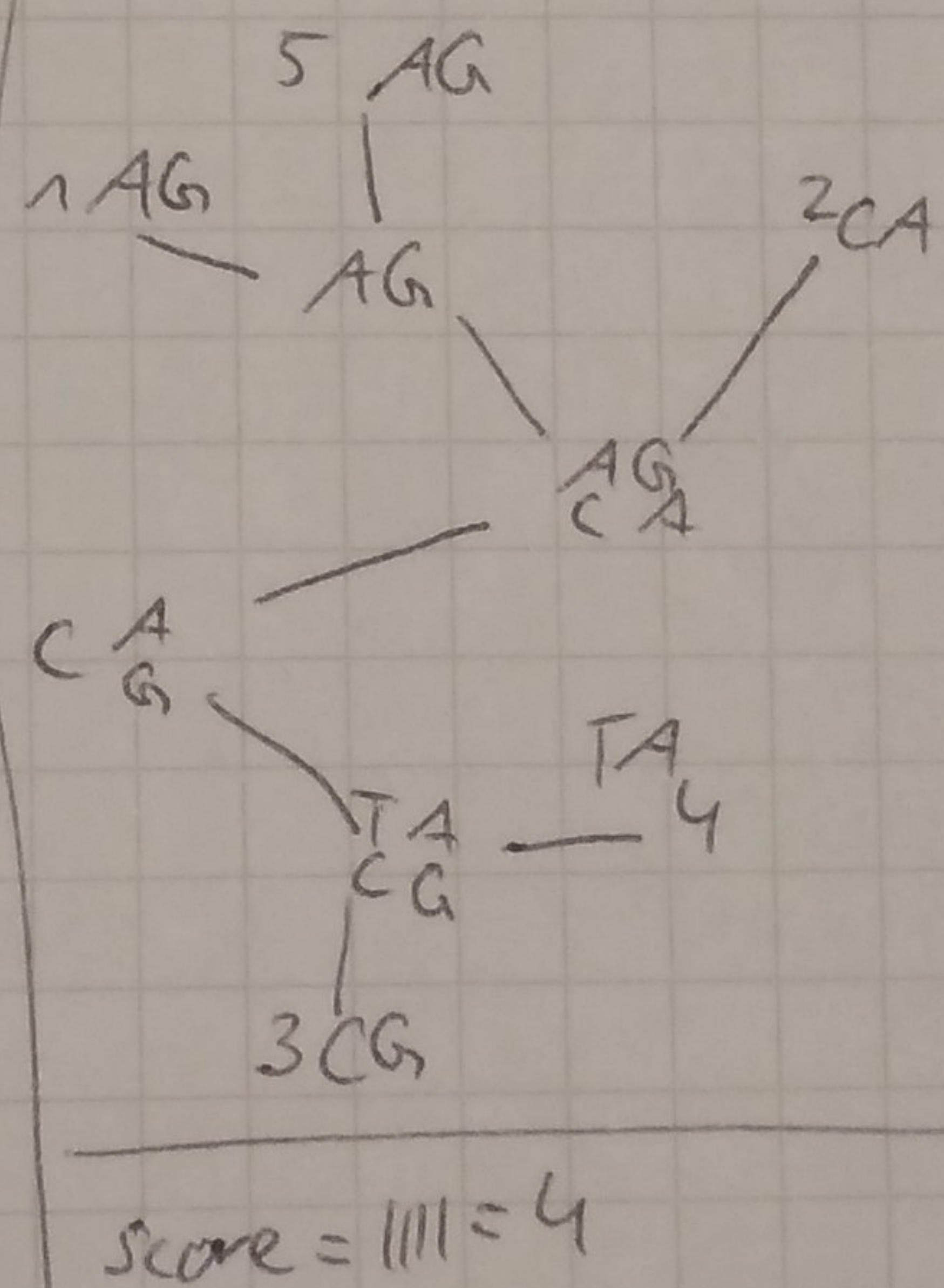
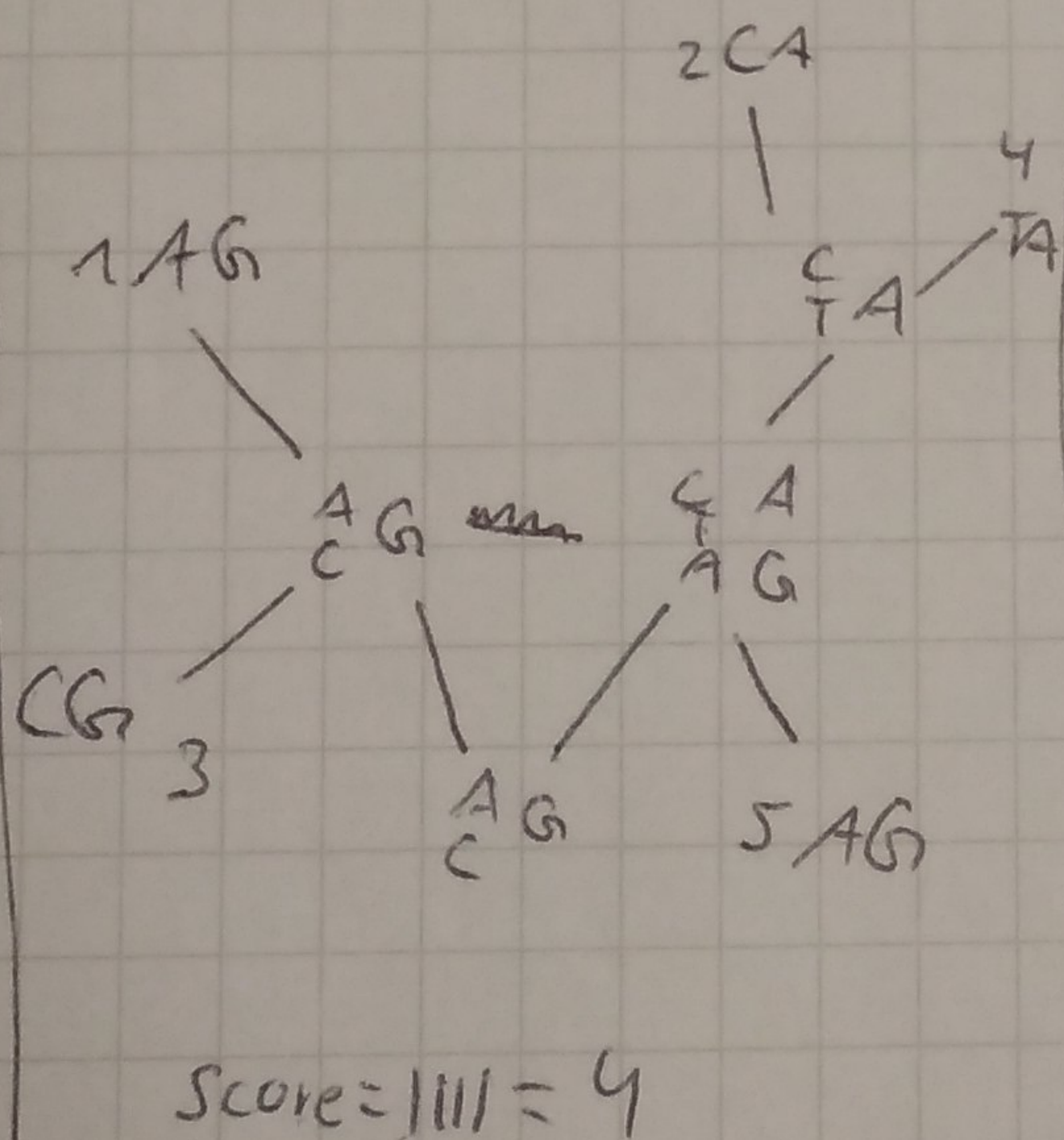
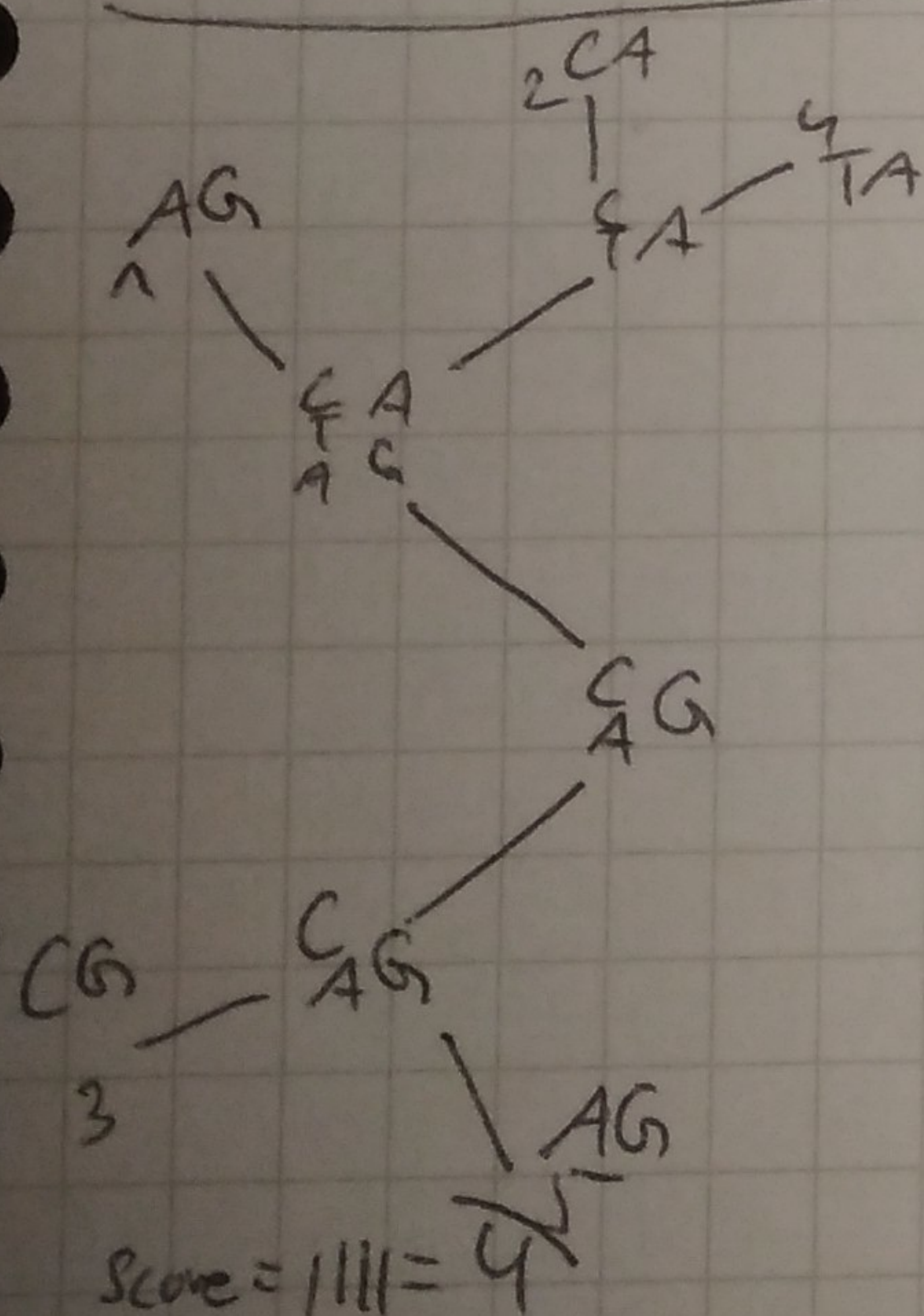
=> best tree is I = score = 3



b)

1 branch (1,5), (3,...)

2 branch (3,...), (2,4)



C) Possibly, but possibly not. It is a heuristic search so we didn't check all the trees. So we can't know for sure.

=> Tree I remains the best with a score of 3.

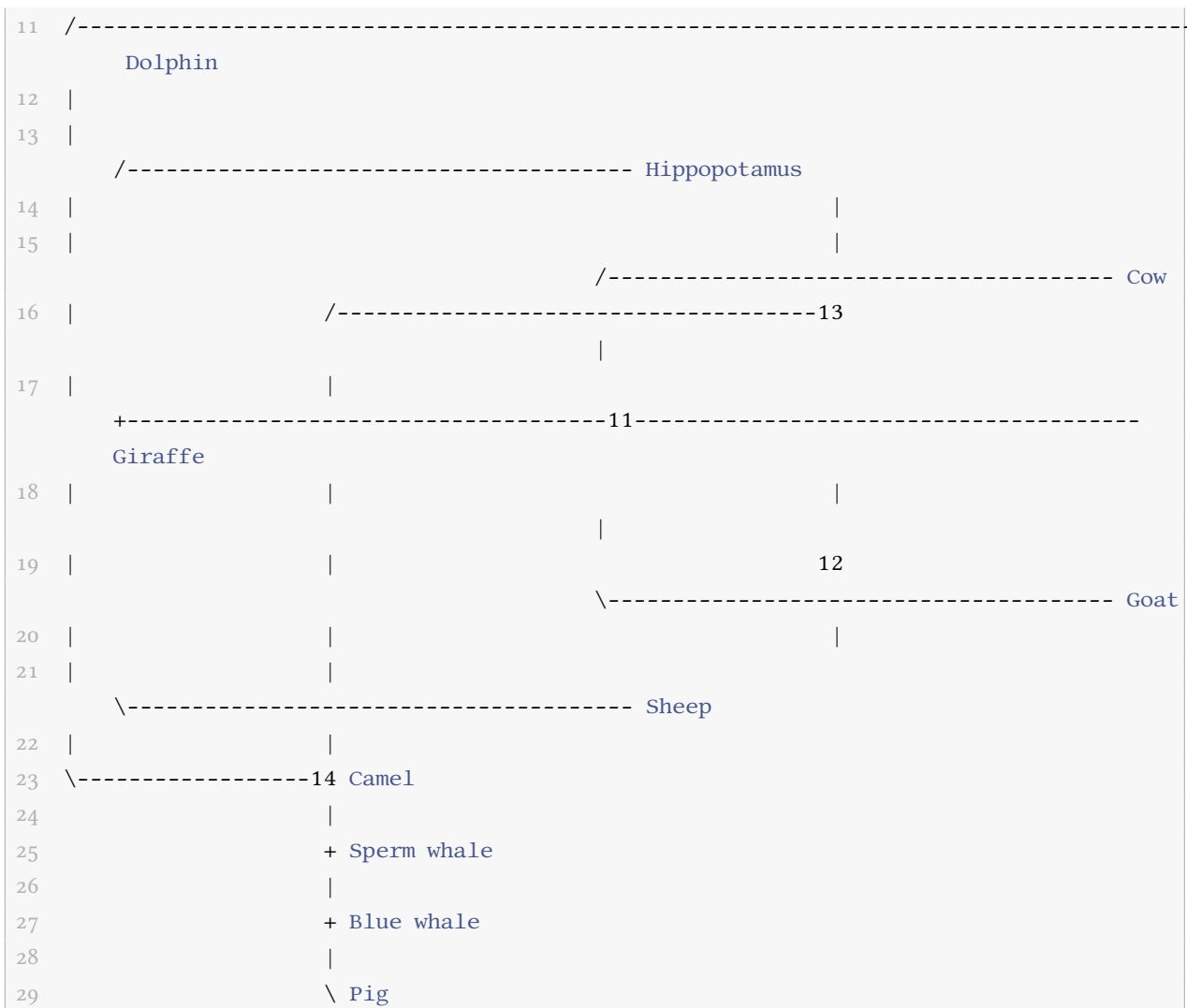
3

c) I did an alltree search on the sequences and this is the result

```
1 Exhaustive search settings:
2   Optimality criterion = parsimony
3   Character-status summary:
4     Of 30 total characters:
5       All characters are of type 'unord'
6       All characters have equal weight
7       22 characters are constant (proportion = 0.733333)
8       6 variable characters are parsimony-uninformative
9       Number of parsimony-informative characters = 2
10    Gaps are treated as "missing"
11    Initial 'Maxtrees' setting = 100
12    Branches collapsed (creating polytomies) if maximum branch length is zero
13    'MulTrees' option in effect
14    No topological constraints in effect
15    Trees are unrooted
16
17 Exhaustive search completed:
18   Number of trees evaluated = 2027025
19   Score of best tree found = 11
20   Score of worst tree found = 16
21   Number of trees retained = 3
22   Time used = 0.54 sec (CPU time = 0.54 sec)
```

So the best score is 11, according to PAUP. I don't know why it is so different from my score, but I guess it just uses a different algorithm and/or different values in the algorithm. The best tree it found looks like this:

```
1 Tree 3 (rooted using default outgroup)
2
3 Tree length = 11
4 Consistency index (CI) = 1.0000
5 Homoplasy index (HI) = 0.0000
6 CI excluding uninformative characters = 1.0000
7 HI excluding uninformative characters = 0.0000
8 Retention index (RI) = 1.0000
9 Rescaled consistency index (RC) = 1.0000
10
```



Please look at my [paup log](#) file for additional info, here is also the [FASTA file](#) I wrote and the [Nexus file](#) I generated from it.

Part II

I aligned the sequences with Pal2Nal (vertebrate table, output format FASTA, I told it to remove gaps, otherwise elephant and human were shorter than the others. Otherwise I left the standard options), I got a lot of warnings that I does not correspond to ATA but it seems like it worked ([result file](#)). I converted it

into a [Nexus file](#).

4

Please look at the complete log file of the analysis [here](#). All the trees that were found had a very similar score and only marginally different topologies, the run with 100 reps was the most different from the rest. It reminds me of the trees I did by hand where I also found the best tree very early. So either this is just the global optimum and very easy to find or it is a local optimum and none of the methods were thorough enough to find the global optimum.

5

I used the mafft alignment for this one because I somehow didn't trust the Pal2Nal result. No, the trees are not the same:

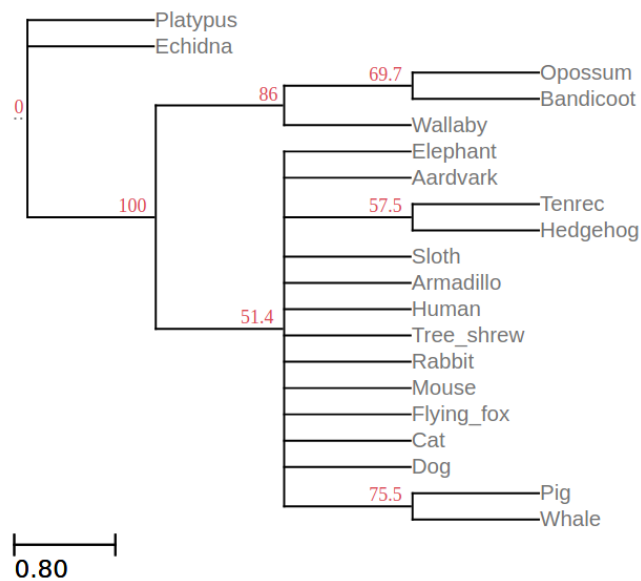


Figure 1: Majority consensus tree

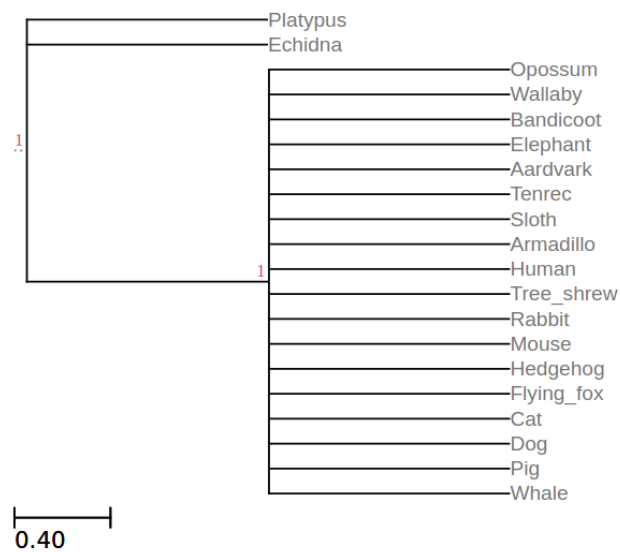


Figure 2: Strict consensus tree

Their topology is different and even more importantly the majority consensus tree contains distances while the strict consensus tree does not.