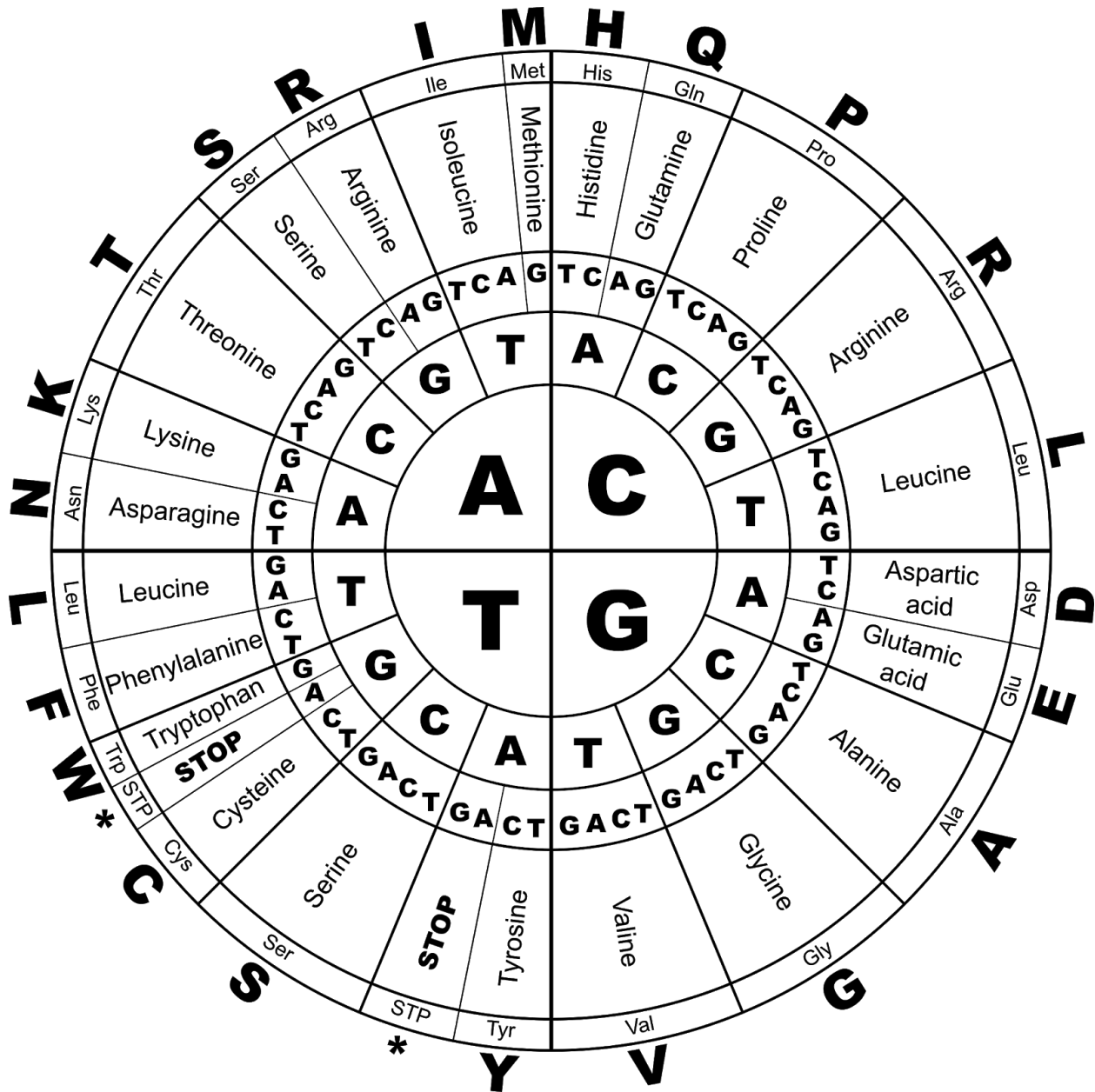


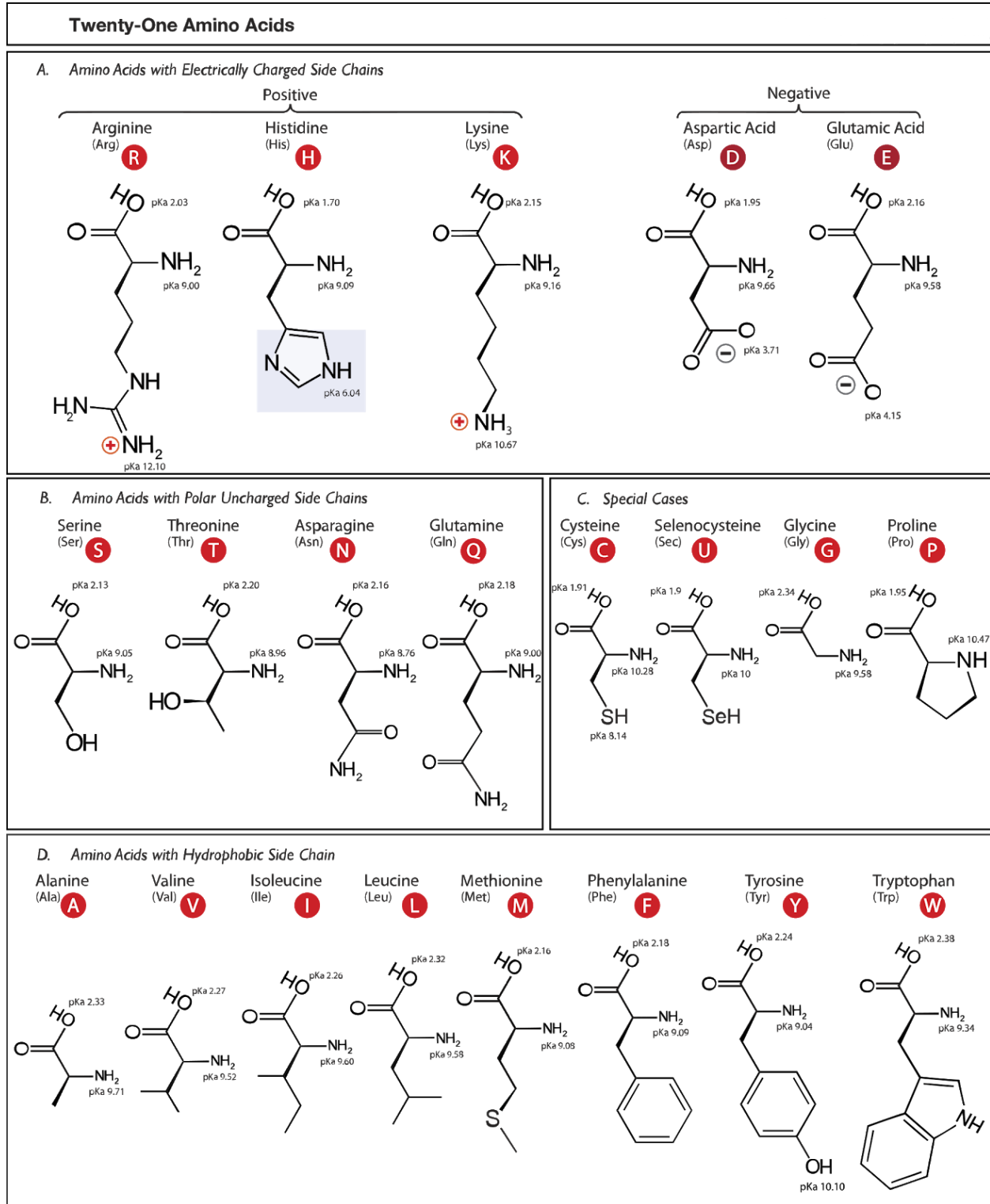
## AMIDD Lecture 2: Biological Sequence Analysis

### Genetic code



Downloaded from <https://pixabay.com/vectors/dna-amino-acids-biology-code-152136/>. Original work by OpenClipart-Vectors. Reused under the Pixabay License.

Amino acids



Modified from the original work by Dancojocari, available at [https://commons.wikimedia.org/wiki/File:Amino\\_Acids.svg](https://commons.wikimedia.org/wiki/File:Amino_Acids.svg), under CC-SA 3.0

## Exercises of Levenshtein distance

Try to identify the Levenshtein's distances and optimal alignments between the following pairs and words, and check the results with any programming language or web-tools you like (for instance <http://www.let.rug.nl/~kleiweg/lev/>):

1. *physicist* and *physician*
2. *flout* and *flaunt*

## A subset of BLOSUM 50 values per aligned residue pair

	H	E	A	G	A	W	G	H	E	E
P	-2	-1	-1	-2	-1	-4	-2	-2	-1	-1
A	-2	-1	5	0	5	-3	0	-2	-1	-1
W	-3	-3	-3	-3	-3	15	-3	-3	-3	-3
H	10	0	-2	-2	-2	-3	-3	10	0	0
E	0	6	-1	-3	-1	-3	0	0	6	6
A	2	-1	5	0	5	-3	0	2	-1	-1
E	0	6	1	3	1	-3	-3	0	6	6

Adapted from *Biological Sequence Analysis* (R. Durbin, S. Eddy, A. Krogh, G. Mitchison), Figure 2.3. We assume that a gap cost per unaligned residue of  $d=-8$ . Try to use the information to perform global alignment between the two amino-acid sequences:

1. HEAGAWGHEE
2. PAWHEAE

## What does Fomivirsen target?

It is possible to search for local sequence matches in large databases of nucleotides, for instance using the BLAST algorithm. An implementation is freely available at National Institute of Health (NIH, US): <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Try to search for the RNA/protein targeted by fomivirsen, given its sequence 5'-GCG TTT GCT CTT CTT CTT GCG-3'.