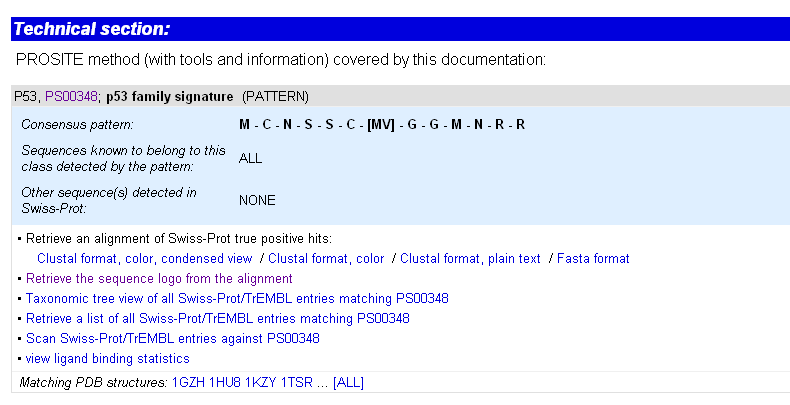
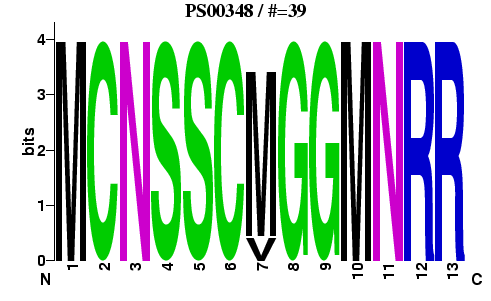
## **https://hanaccount.han.nl/img/han_oh.gif OWE 3a: Python II**

# Week 4: Regular Expressions



Figuur 1: Screenshot van http://kr.expasy.org/cgi-bin/nicedoc.pl?PDOC00301

De p53 familie kent een eigen *signatuur*. Dit betekent dat alle genen van deze familie te herkennen zijn aan een patroon. Dit overeenkomstige patroon noemen we het consensus patroon. Voor p53 is dit een vrij eenvoudig patroon: de aminozuurvolgorde is MCNSSC dan een M of een V en dan GGMNRR.

We kunnen daar het volgende sequence logo voor definiëren:

Figuur 2: afkomstig van screenshot van http://kr.expasy.org/cgi-bin/nicedoc.pl?PDOC00301

1. Breid je code van opgave 3 uit en zorg dat je middels een reguliere expressie in staat bent om in de **aminozuur** sequentie van p53 het consensus patroon te herkennen.
2. Scan een aantal bestanden van <http://www.uniprot.org/uniprot/?query=%22prosite+PS00348%22> op de aanwezigheid van het p53 consensus patroon.
3. Toon met een regular expression aan dat alle letters in de het p53 eiwit voor een aminozuur staan. Toon dus aan dat er bijvoorbeeld geen B in voorkomt. Doe dit in een aparte functie.

Inleveren: opgave4re<naam>.py