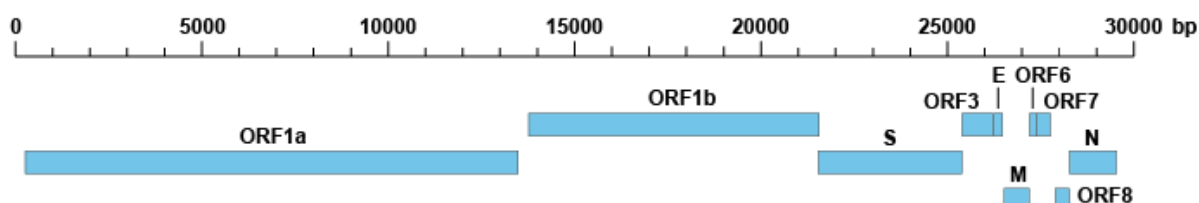


CORONA ONDER DE LOEP (31)

Voor de overzichtelijkheid laat ik hieronder andermaal de schematische afbeelding zien van het genoom van SARS-CoV-2. Aangegeven zijn de verschillende domeinen die men aanduidt met ORFs (*open reading frames*; ze coderen voor de niet-specifieke eiwitten van het virus) én de domeinen S (spike-eiwit), E (envelopeiwit), M (membraaneiwit) en N (nucleocapside-eiwit). S, E, M en N coderen voor specifieke structuureiwitten die respectievelijk uit 1273, 75, 222 en 419 aminozuren (aa) bestaan. We buigen ons allereerst over het genoom.



We herinneren ons dat RNA is opgebouwd uit bouwstenen die we nucleotiden noemen en dat deze nucleotiden weer zijn samengesteld uit een fosfaatgroep, een suiker en een base. In RNA komen we als basen tegen: G (guanine), C (cytosine), A (adenine) en U (uracil).

In de vorige aflevering werd al vermeld dat het hele genoom uit bijna 30.000 van deze basen bestaat. Om een beeld te krijgen op welke wijze de vier basen deel uitmaken van het complete patroon, kijken we naar de onderstaande reeksen van letters (A,U,G en C) die de correcte basenvolgorde aan het begin en aan het einde van het genoom weergeven.



AUGUUUGUUUUUCUUGUUUUUAUUGCCACUAGUCUCUAGUCAGUGUGUUAUCUUACAACCA.....

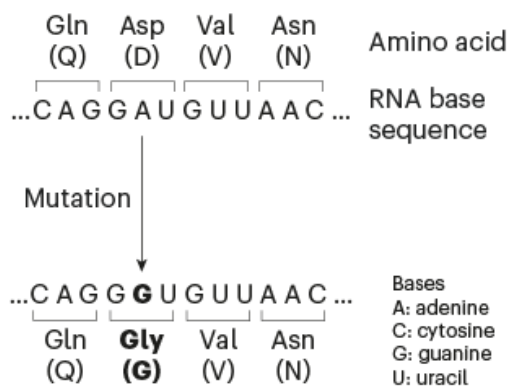
.....CAAUUUGAUGAAGACGACUCUGAGCCAGUGCUCAAAGGAGUCAAAUACA-UUACACAUAA



Ik heb vrij willekeurig gekozen voor een beginstuk met zestig basen (van AUG tot en met CCA) en een even lang eindstuk (van CAA tot en met UAA). De verbindende stippellijn moet de rangschikking van de vier basen op de vele honderden tussenliggende posities verbeelden. Als de punten op de stippellijn de plaatsen van *alle* ontbrekende basen zouden moeten voorstellen, dan zou het aantal punten dus de 30.000 naderen.

Het getoonde patroon is karakteristiek voor het genoom van SARS-CoV-2. Het lijkt weliswaar een chaotische aaneenschakeling, maar de letters (basen) bezetten wel degelijk specifieke plaatsen op het genoom. Als het oorspronkelijke virus (*wildtype*) onze gezonde cellen heeft geïnfecteerd, treffen we dus in het cytoplasma van die cel deze lange ketting van nucleotiden aan. Hierop volgt de synthese van allerlei eiwitten en de vorming van nieuwe virusdeeltjes. Tijdens dit replicatieproces kan door een kopieerfout één of meer van deze 'letters' vervangen worden door andere (insertie), of kunnen één of meer letters 'zomaar' uit dit patroon verdwijnen (deletie).

Ik geef een voorbeeld van een mutatie die zich tijdens deze pandemie daadwerkelijk heeft voorgedaan en die oorzaak is geweest van de 'geboorte' van een mutant waar we niet bepaald op hebben zitten wachten. Ik zal hierop in een later stadium van het verhaal nog ingaan.



We zien in het bovenste deel van deze figuur de basenvolgorde (nucleotidensequentie) van SARS-CoV-2, het *wildtype*. We herkennen in de rij letters de vier basen (CAGU). Ze zijn gerangschikt in codons van elk **drie** nucleotiden, die ieder in het ribosoom coderen voor een bepaald aminozuur. Het codon GAU bijvoorbeeld zorgt ervoor dat asparaginezuur in de peptideketen wordt ingebouwd.¹ In het onderste deel van de afbeelding zien we dat er een mutatie is opgetreden. In plaats van adenine (A) in het oorspronkelijke codon GAU, wordt guanine (G) ingebracht! Dit heeft in dit (bekende) geval consequenties! Het nieuwe codon wordt nu GGU en dat codeert voor het aminozuur glycine. Deze wijziging geeft aan het S-eiwit andere eigenschappen.

Dergelijke mutaties in het RNA worden in de wetenschappelijke literatuur op een bepaalde manier opgeschreven. Dat gaat zo.

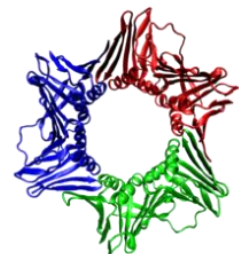
Door middel van een getal wordt de positie in het genoom aangegeven *waar* de afwijking is aangetroffen en *wat* er precies is veranderd. Als we de wijziging beperken tot de plaatsen waar zich een verwisseling van basen heeft voorgedaan, dan kan de mutatie op twee manieren zijn genoteerd, namelijk

1. als bijvoorbeeld C23271A; we lezen dit als een mutatie die zich voordoet op positie 23271 in het genoom, waarbij een C (de base cytosine) is vervangen door A (de base adenine);
2. als bijvoorbeeld 24914G>C; we lezen dit als een mutatie die zich voordoet op positie 24914 in het genoom, waarbij een G (de base guanine) is vervangen door C (de base cytosine).

Je hebt misschien al opgemerkt dat beide mutaties zijn opgetreden in het spike-eiwit. Immers, de posities 23271 en 24914 vinden we terug in het domein S van de figuur bovenaan bladzijde 1. Zouden we in een tekst de mutatie C5388A tegenkomen, dan moeten we deze zoeken in het domein ORF1a.

Van de blauwe balken in de figuur op bladzijde 1, leggen we het domein S onder de microscoop. Immers, als een mutatie in het genoom van het coronavirus een relevante afwijking in de aminozuurvolgorde van de spike (stekel, spijker, S-eiwit) veroorzaakt, kan dit in principe de biologische functie van de spike beïnvloeden. Om hierin enig inzicht te krijgen en om gangbare codenamen als E484K, N501Y en A570D te kunnen ontcijferen en deze een plaats te geven, zullen we de stekels van het virusdeeltje nader moeten analyseren.

Maar eerst even deze aanvulling op de bouw van de spikes op het coronavirus. Om het niet al te ingewikkeld te maken, heb ik er tot nu toe voor gekozen om een spike als één eiwit, of beter, één glycoproteïne te beschouwen. Dat is echter niet hetgeen de werkelijkheid ons laat zien. Een stekel aan het coronavirus bestaat niet uit één glycoproteïne dat naar buiten steekt, maar is een vervlechting van **drie** glycoproteïnen die volkomen identiek zijn. In de biologie noemt men zo'n structuur een *trimeer*. De tekening rechts toont zo'n trimeer. Elke kleur stelt een eiwitmolecuul voor.



De ruimtelijke structuur van de spike op het coronavirus ziet er weer heel anders uit. Deze zien we op bladzijde 4 van deze aflevering.

In verband met het verhaal dat volgt, was het nodig dit nu reeds onder de aandacht te brengen.

¹ De letters tussen haken zijn afkortingen van de overeenkomstige aminozuren; ik kom hier op de volgende bladzijde op terug.

We gaan nu verder met de 'ontleding' van de spike en doen dat uitgaande van één (van de drie identieke) glycoproteïnen. We hebben het dan over het S-eiwit (spijkereiwit of spike-eiwit)

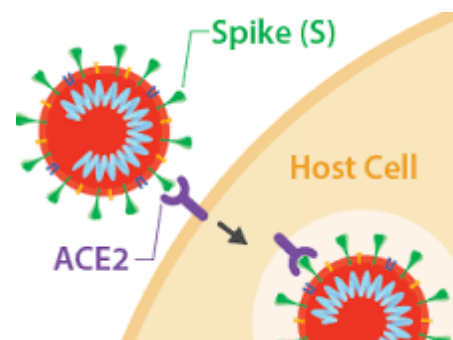
Het is boeiend om te weten *welke* veranderingen zich zoal tijdens het mutatieproces in het S-eiwit kunnen voordoen en vooral welke gevolgen die afwijkingen hebben voor de mate van besmettelijkheid én het ziekmakend vermogen van het nieuwe virus (mutant, variant). Om deze processen in klare taal te kunnen opschrijven, is het didactisch handig om eerst een enigszins gedetailleerd overzicht te geven van het mechanisme waarmee het coronavirus onze cellen binnenkomt.

Hier en daar geef ik een aanvulling op hetgeen je hierover al in aflevering 12 hebt kunnen lezen.

Het S-eiwit als target

Het grote S-eiwit-molecuul, verankerd in het oppervlak van SARS-CoV-2, bevat een gebied (domein) dat het voor het virusdeeltje mogelijk maakt om van buitenaf het binnenste van onze cellen te bereiken. Dit domein, een reeks aan elkaar geregen aminozuren, dient qua structuur aan specifieke eisen te voldoen. Veranderingen in het vaste patroon door bijvoorbeeld wijzigingen in de *volgorde* en/of *identiteit* van één of meer aminozuren kunnen leiden tot andere eigenschappen van de spike. Ook kunnen er zomaar aminozuren ontbreken. Deze afwijkingen kunnen er de oorzaak van zijn dat het vermogen om de cel binnen te dringen wordt afgezwakt, wordt versterkt of helemaal verloren gaat (zie de tabel in afl. 29).

Weten we het nog? De penetratie komt tot stand doordat het hierboven bedoelde domein van het S-eiwit **bindt** aan de buitenmembraan van onze cellen. Dit proces stelt het virusdeeltje vervolgens in staat met de celmembraan te **fuseren**, waarna het als endosoom in het cytoplasma terechtkomt en daar het genetische materiaal loslaat. Is de binding of fusie door een of andere oorzaak niet mogelijk, dan kan de cel niet worden geïnfecteerd en kan het virus zich niet vermenigvuldigen.



Dit is nieuw voor je: aan dit *cruciale* domein, ergens in het S-eiwit, onderscheiden we twee aparte gebieden:

1. een deel dat verantwoordelijk is voor de binding aan een receptor in de celmembraan; deze receptor draagt de naam ACE2;
2. een deel dat de fusie van het virus met de celmembraan mogelijk maakt.

Intermezzo

Het betreffende domein ga ik in kaart brengen. Maar alvorens ik aan de beschrijving van dit deel van het S-eiwit begin, moeten we weten hoe men in de (wetenschappelijke) literatuur de verschillende aminozuren kortheidshalve noteert.

Het komt voor dat men elk aminozuur met de eerste drie letters afkort (bijv. glycine wordt gly), maar in een aantal speciale gevallen kort men in de literatuur een aminozuur met slechts één enkele letter af, omdat de notatie met drie letters te veel ruimte in beslag zou nemen. Die ene letter kan niet altijd de eerste letter van het aminozuur zijn, aangezien er meerdere aminozuren met dezelfde letter beginnen, bijvoorbeeld alanine, arginine en asparagine. Hier is het overzicht:

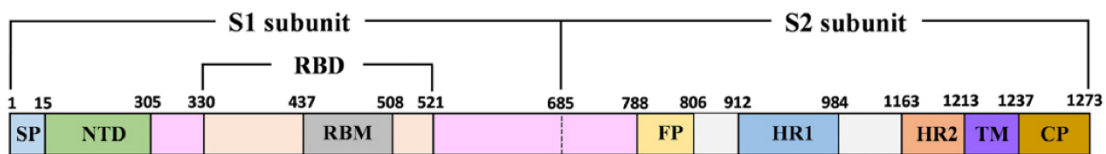
alanine (A), arginine (R), asparagine (N), asparaginezuur (D), cysteine (C), glutamine (Q), glutaminezuur (E), glycine (G) histidine (H), isoleucine (I), leucine (L), lysine (K), methionine (M), fenylalanine (F), proline (P), serine (S), threonine (T), tyrosine (Y), valine (V)

Je vraagt je mogelijk af wat jij als lezer hiermee moet. Dit is het antwoord: het is mijn bedoeling om het voor penetratie cruciale domein van het S-eiwit in letters op te schrijven, althans voor zover het de *aminozuren* in de ketting betreft,² om vervolgens te laten zien hoe het proces van celpenetratie op moleculair niveau verloopt.

Je gaat een wandelingetje maken over het spijkereiwit. Haak aan!

De diverse domeinen

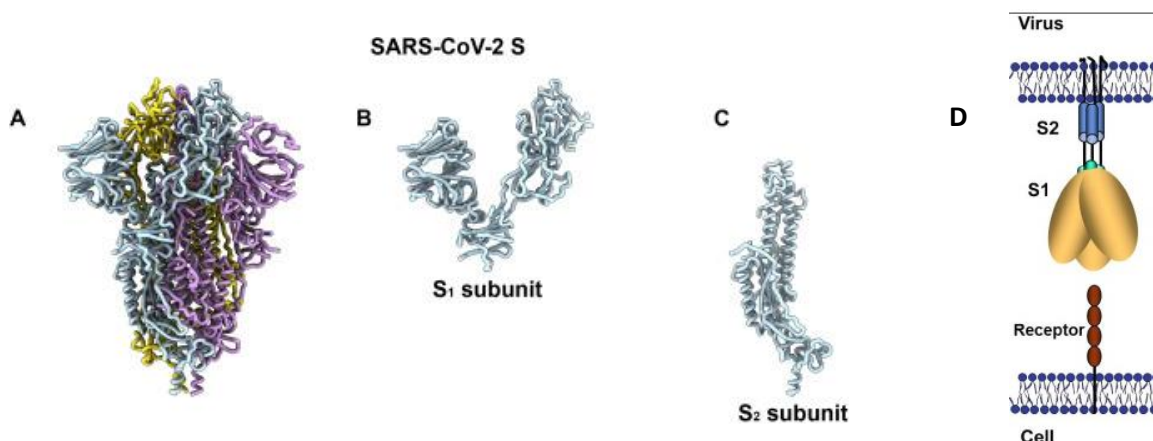
Onderstaande tekening toont - heel schematisch - het uitgerolde glycoproteïne (dus één van de drie in de trimeer). Het gaat hier dus niet langer over nucleotiden met de basen A, C, G en U - die maken immers deel uit van het virale RNA - maar over aminozuren. Aminozuur nr. 1 bevindt zich helemaal aan de linkerkant van de balk; aminozuur nr. 1273 zit uiterst rechts. Tussen het begin en het einde onderscheiden we een aantal gebieden, ofwel, domeinen.



Ik ga ze niet *allemaal* uitvoerig bespreken, maar beperk me in eerste instantie tot de domeinen RBD (= Receptor Bindend Domein), RBM (= Receptor Bindend Motief), en FP (fusiepeptide). Verder onderscheidt men aan het S-eiwit twee *subunits*: S1 en S2. De reden daarvan zul je in de volgende aflevering lezen.

Je kijkt dus tegen het hele S-eiwit aan, schematisch getekend en uitgestrekt in het platte vlak. We stellen vast dat het receptor-bindend domein (RBD; lichtoranje en grijs) het gedeelte van de keten is dat na besmetting een binding aangaat met de ACE2 receptor aan de buitenkant van de celmembraan van de gastheer. Het RBD loopt van positie 330 tot positie 521. Binnen dit domein gaat het in het bijzonder om het receptor-bindend motief (RBM; grijs), dat wil zeggen om het *patroon* van aminozuurketen in dat kleinere gebied binnen het grotere RBD. Dit omvat de posities 437 tot 508.

Ik kom nu terug op de ruimtelijke structuur van de spikes. Zo'n uitsteeksel heeft dus de structuur van een trimeer: de in elkaar verstrengelde ketens van de drie glycoproteïnen, dus drie van de hierboven schematisch weergegeven structuren!

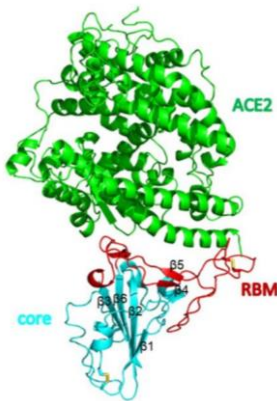


Afbeelding A verbeeldt de ruimtelijke structuur van de spike. We onderscheiden hierin de drie kleuren geel, paars en blauw die ieder een glycoproteïne markeren. We halen het blauw gekleurde glycoproteïne er even uit en bekijken dit in de figuren B en C.

² Er zitten hier en daar ook nog *suikermoleculen* aan vast; we hebben immers te maken met een **glycoproteïne**; deze hebben echter voor ons verhaal geen doorslaggevende betekenis.

Figuur B stelt de S1 *subunit* voor en C toont de daarvan losgeknipte S2 *subunit* (ze komen dus overeen met de S1 en S2 gebieden zoals op het schema op de vorige bladzijde aangegeven).

E



Afbeelding D laat zien op welke wijze de spike in de virusenvelop is ingebouwd. De drie ‘voetjes’ van de drie S2 *subunits* (in plaatje C zie je er een aan de onderkant van S2) zitten stevig verankerd in de buitenmembraan van het coronavirus. De drie *subunits* S1 (van elk glycoproteïne één) met hun RBDs en RBMs vormen de top (knop) van de spike (geel). Deze worden aangetrokken door de receptor (ACE2) op de buitenmembraan van een lichaamscel. Die aantrekkingskracht gaat vooral uit naar het domein RBM. Dit is weergegeven in de tekening bij E. We zien daar het RBD (blauw) met als onderdeel daarvan het RBM (rood) dat de binding is aangegaan met een ACE2-receptor (groen) van bijvoorbeeld een epitheelcel in de luchtweg van een geïnfecteerd persoon.

Een tocht over een pad met aminozuren

We willen wat meer details weten binnen het RBD (RBM) en stellen ons de vraag welke aminozuren tot dit domein behoren en hoe deze zijn geordend. Het is voor wetenschappers niet erg ingewikkeld om deze gedetailleerde structuur te onthullen. Men beschikt namelijk al vele jaren over de techniek om de identiteit én sequentie van aminozuren in (poly)peptiden en eiwitten vast te stellen.

Hieronder heb ik in vier regels de aminozuren + hun specifieke volgorde in het RBD weergegeven.³ Je ziet dat dit domein begint met P, dat wil zeggen met het aminozuur proline (op positie 330), en met P (proline) op positie 521 eindigt. Onderweg zijn we het sub-domein RBM tegengekomen. Hiervan zijn de aminozuren met rode letters afgebeeld.

Het RBM neemt dus op positie 437 met het aminozuur asparagine (N) een aanvang en sluit af met aminozuur tyrosine (Y) op positie 508. Aan weerszijden van deze hele verzameling bevinden zich uiteraard de aminozuren die aan beide kanten van het RBD het S-eiwit completeren (hier dus niet getoond).

←PNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFK
330

CYGVSP TKLNDLCFTNVYADSFVIRGDEVRQIAPGQGTGKIADYNYKLPDDFTGCVIAW

NSNNLDSKVGGNYNLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGF
437 484

QPTNGVGYQPYRVVLSFELLHAP→
501 508 521

Wij richten onze blik op de aminozuren die het RDM vormen, dus op de rode letters. De samenstelling van dit domein is voor het virus van groot belang. Als hierin aminozuren van positie zouden wisselen, of als er aminozuren aan worden toegevoegd (verlenging door insertie), of als er aminozuren tussenuitvallen (verkorting door deletie) dan bestaat de kans dat de binding van het S-

³ Let wel, de afgebeelde aminozuren zijn in werkelijkheid aan elkaar gekoppeld en de vier regels dien je dus te lezen als één keten van aminozuren.

eiwit (de spike) aan een cel van de gastheer niet of veel moeilijker verloopt. Maar het kan ook zijn dat een onverhoopte wijziging in deze structuur leidt tot een veel efficiëntere en snellere binding met de ACE2 receptor! Wij gaan hiervan een voorbeeld zien.

Wanneer is er sprake van zulke veranderingen? Als er sprake is van een afwijking in de opbouw van een S-eiwit. En hoe komt de viroloog hierachter? Door de sequentie van de aminozuren in dat deel van de spike in het laboratorium vast te stellen en na te gaan of de structuur nog zo is als hierboven met rode letters is aangegeven!

De notatie van mutaties in het S-eiwit

Je weet inmiddels hoe we een mutatie in het genoom aanduiden (bijv. C23271A). Op gelijke wijze krijgen de mutaties in het S-eiwit codenamen.

Men volgt hier dezelfde systematiek: het getal dat de codenaam meekrijgt slaat op de positie in het S-eiwit waar zich een verandering ten opzichte van het *wildtype* heeft voorgedaan. De mogelijkheden beperken zich dan tot de getallen 1 t/m 1273, want meer posities zijn er niet. Vóór zo'n getal staat een letter die het aminozuur aangeeft dat op die plaats in het *wildtype* staat. De codenaam wordt afgesloten met een letter die het aminozuur weergeeft dat in de plaats gekomen is van het oorspronkelijke aminozuur. Dus:

De codenaam E484K geeft aan dat we hier te maken hebben met een mutant, gekenmerkt door de insertie van het aminozuur lysine (K) op positie 484, die in het *wildtype* wordt bezet door het aminozuur glutamine (E).

De codenaam N501Y is je nu ook duidelijk: tyrosine (Y) verving het aminozuur asparagine (N) op positie 501. De twee getallen (484 en 501) wijzen erop dat het hier gaat om twee mutaties in receptor-bindend motief (RBM); ze maken immers deel uit van de rode letters.

In de literatuur lezen we als resultaat van onderzoek dat de mutant C23604A (dus in het genoom en, specifieker, in het S-domein; zie het uitgerolde genoom) leidt tot de mutatie met de codenaam P681H. Dat laatste betekent dus dat de vervanging van C door A in het genoom op positie 23604 tot gevolg heeft dat op positie 681 niet langer P, maar H is ingebouwd. We zien in het uitgerolde S-eiwit dat deze positie buiten het RBD ligt. Kan dit dus geen kwaad? Jazeker, die mutatie kan ons parten gaan spelen, want hij is gesitueerd in een domein dat een essentiële rol speelt bij de **fusie** van het virusdeeltje met de gastheercel. Hierover lezen we van alles in aflevering 32.

Waarom willen we dit allemaal weten!?!

Het behoeft nauwelijks gezegd: kennis van de mechanismen die ten grondslag liggen aan de binding van het virusdeeltje aan een lichaamscel én van het fusieproces (beide zijn voorwaarden voor penetratie, replicatie en verspreiding) is van groot belang. Die kennis wordt door wetenschappers gebruikt om niet alleen te kunnen verklaren waarom bepaalde mutaties in het spike-eiwit voordelig voor het virusdeeltje zijn (besmettelijker, schadelijker voor de gastheer), maar vooral óók om trucs te ontwikkelen die deze mechanismen dwarsbomen, blokkeren. Is men eenmaal hiertoe in staat dan ligt met enige kans de ontwikkeling van werkzame antivirale geneesmiddelen tegen Covid-19 in het verschiet.

Want wat zou het prettig zijn om - met een snotneus positief getest - binnen enkele dagen met tweemaal daags een pilletje het virus uit het lijf te kunnen verdrijven!

Je gaat hierover lezen! Maar de opbouw van het verhaal dient geleidelijk te verlopen 😊.

Arijan Porsius
16-03-21