



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Industrial

Diseño de un Modelo Conceptual para caracterizar
proteoma y pathways humanos: aplicaciones al
metabolismo de la CoVid-19

Trabajo Fin de Máster

Máster Universitario en Ingeniería Biomédica

AUTOR/A: Romero Jaque, Laura

Tutor/a: Pastor López, Oscar

CURSO ACADÉMICO: 2021/2022



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIERÍA
INDUSTRIAL VALENCIA

Curso Académico:

AGRADECIMIENTOS

A Óscar y Alberto,

por darme la oportunidad de realizar este Trabajo Fin de Máster y su inestimable paciencia y apoyo, ha sido un placer trabajar contigo, gracias por tu energía y apoyo durante toda la elaboración del trabajo has sido mi guía e inspiración.

A mi familia,

por su apoyo constante e incondicional en cada proyecto de mi vida, enseñándome a perseguir mis sueños y no rendirme nunca, gracias por haberme dado la fuerza y el estímulo durante la redacción de este proyecto.

A la universidad,

y todo el profesorado que me han formado durante estos cuatro años

RESUMEN

En el siguiente trabajo Fin de Máster se proponen aspectos relevante para la base del diseño de una nueva vista para el modelo conceptual del genoma humano CSHG v3 buscando ampliar el marco de aplicación del modelo al estudio enfermedades infecciosas de origen vírico por medio de la caracterización de su patogénesis gracias a la instanciación de un caso real, en este caso en de las vistas proteína y pathways que actualmente existen en el modelo. Por medio de la instanciación de un caso real de patogénesis de la enfermedad del coronavirus (COVID19), a partir de la aproximación con los mecanismos de infección del SARS-CoV1 en vez del SARS-CoV1, debido a la gran similitud entre ambos virus y el mayor índice de bibliografía y estudios concluyentes en cuanto al comportamiento del primero.

Para ello se hizo una descripción textual detallada de las vistas proteoma y pathways definiendo cada clase, relación y cardinalidad presente en cada vista. Con ellos se forma la herramienta vista estructural que permite generar el primer acercamiento al análisis y caracterización de patogénesis virales. A continuación se diseño la vista funcional en base a la bibliografía científica de la patogénesis del SARS-CoV1. El siguiente paso fue la ejecución de un análisis de consistencia de la instanciación de la vista funcional por medio del mapeo, con el que se detectaron varias limitaciones de expresión y modelización claves para la caracterización correcta de las infecciones virales.

Por último se redacto una serie aspectos, enfoques y conceptos que sirven de base para poder realizar un rediseño de las vistas preexistentes del modelo así como una nueva centrada en los procesos de enfermedades infecciosas como el COVID19 que tenga como objetivo de entender las interacciones relevantes a nivel proteico que puedan ayudar a comprender el funcionamiento biológico-molecular de la enfermedad.

Palabras clave: Modelado conceptual, patogénesis, infección vírica, diseño, caracterización, COVID19, coronavirus, SARS-COV, proteína.

RESUM

Al següent treball Final de Màster es proposen aspectes rellevants per a la base del disseny d'una nova vista per al model conceptual del genoma humà CSHG v3 buscant ampliar el marc d'aplicació del model a l'estudi malalties infeccioses d'origen víric per mitjà de la caracterització de la seua patogènesi gràcies a la instanciació d'un cas real, en aquest cas en les vistes proteïna i pathways que actualment existeixen al model. Per mitjà de la instanciació d'un cas real de patogènesi de la malaltia del coronavirus (COVID19), a partir de l'aproximació amb els mecanismes d'infecció del SARS-CoV1 en comptes del SARS-CoV1, a causa de la gran similitud entre ambdós virus i el major índex de bibliografia i estudis concloents pel que fa al comportament del primer.

Per això es va fer una descripció textual detallada de les vistes proteoma i pathways definint cada classe, relació i cardinalitat present a cada vista. Amb ells es forma l'eina vista estructural que permet generar el primer apropament a l'anàlisi i la caracterització de patogènesis virals. A continuació es va dissenyar la vista funcional sobre la base de la bibliografia científica de la patogènesi del SARS-CoV1. El pas següent va ser l'execució d'una anàlisi de consistència de la instanciació de la vista funcional per mitjà del mapatge, amb la qual es van detectar diverses limitacions d'expressió i modelització claus per a la caracterització correcta de les infeccions virals.

Finalment es redacta una sèrie d'aspectes, enfocaments i conceptes que serveixen de base per poder realitzar un redisseny de les vistes preexistents del model així com una nova centrada en els processos de malalties infeccioses com el COVID19 que tingui com a objectiu entendre les interaccions rellevants a nivell proteic que puguin ajudar a comprendre el funcionament biològic-molecular de la malaltia.

Paraules clau: Modelat conceptual, patogènesi, infecció vírica, disseny, caracterització, COVID19, coronavirus, SARS-COV, proteïna.

ABSTRACT

In the following Master's Thesis, relevant aspects are proposed for the basis of the design of a new view for the conceptual model of the human genome CSHG v3, seeking to broaden the framework of the application of the model to the study of infectious diseases of viral origin through the characterization of its pathogenesis thanks to the instantiation of a real case, in this case in the protein and pathways views that currently exist in the model. Through the instantiation of a real case of the pathogenesis of the coronavirus disease (COVID19), from the approximation with the infection mechanisms of SARS-CoV1 instead of SARS-CoV1, due to the significant similarity between both viruses and the highest index of bibliography and conclusive studies regarding the behaviour of the first.

For this, a detailed textual description of the proteome and pathways views was made, defining each class, relationship and cardinality present in each view. With them, the structural view tool is formed that allows generating the first approach to the analysis and characterization of viral pathogenesis. Next, the functional view was designed based on the scientific literature on the pathogenesis of SARS-CoV1. The next step was to carry out a consistency analysis of the instantiation of the functional view through mapping, with which several key modelling and expression limitations for proper characterization of viral infections.

Finally, a series of aspects, approaches and concepts were drafted that serve as a basis for redesigning the pre-existing views of the model as well as a new one focused on the processes of infectious diseases such as COVID19 that aims to understand the interactions relevant to protein level that can help to understand the biological-molecular functioning of the disease.

Keywords: Conceptual modeling, pathogenesis, viral infection, design, characterization, COVID19, coronavirus, SARS-COV, protein.

DOCUMENTOS CONTENIDOS EN LA MEMORIA

- DOCUMENTO I: MEMORIA
- DOCUMENTO II: PRESUPUESTO
- DOCUMENTO III: ANEXOS

ÍNDICE

DOCUEMENTO I: MEMORIA

CAPITULO 1: MOTIVACIÓN Y OBJETIVOS

1.1 Motivación.....	10
1.2 Objetivos	11

CAPITULO 2: INTRODUCCIÓN

2.1 Modelado conceptual	13
2.1.1 Fundamentos del modelado conceptual.....	13
2.1.2 Caracterización de elementos conceptuales.....	14
2.1.3 Técnicas de modelado conceptuales	14
2.2 Coronavirus	17
2.2.1 Características generales y taxonomía.....	17
2.2.2 Estructura vírica general.....	18
2.3 Patogénesis viral.....	19
2.3.1 Definición	19
2.3.2 Relación virus – célula	19
2.3.3 Tipos de infecciones virales.....	20
2.4 Proteomica	20
2.4.1 Proteínas y proteoma.....	20
2.4.2 Biosíntesis y funciones biológicas	21
2.4.3 Estructura proteica.....	21
2.4.4 Clasificación de proteínas.....	22

CAPITULO 3: MATERIALES Y MÉTODOS

3.1 Conceptual Schema for the Human Genome (CSHG)	26
3.2 Vista estructural	27
3.2.1 Vista proteoma.....	27
3.2.2 Vista pathways	34
3.3 Vista funcional <i>de los mecanismos de patogénesis Vírica</i>	40
3.3.1 Características generales de la patogénia del SARS-CoV-1	40
3.3. 2 Unión y entrada a la célula	40
3.3.2 Respuesta inmune	43

CAPÍTULO 4: ANÁLISIS DE CONSISTENCIA

4.1 Mapeo	48
4.1.1 Instanciación proteínas virales no estructurales.....	52
4.1.2 Instanciación de las proteínas virales accesorias	53
4.1.3 Instanciación de las proteínas virales estructurales.....	54
4.2 Limitaciones del modelo CSGH v3	55

CAPÍTULO 5. CONCLUSIONES Y APORTACIONES

CAPÍTULO 6. BIBLIOGRAFÍA.....

DOCUMENTO II: PRESUPUESTO

1. Coste personal.....	67
2. Coste de ejecución	67
2.1 Coste del software.....	67
2.2 Coste del hardware	68
3. Coste total	68

DOCUMENTO III: ANEXOS

ANEXO I	71
ANEXO II	73

LISTA DE FIGURAS

Figura 1. Simbología de la composición (rombo negro) y la agregación (rombo blanco). Fuente: [7]	16
Figura 2. Ejemplo de asociación entre la entidad vehículo con las entidades automóvil, motocicleta y camión por medio de la relación <i>es-un</i> . Fuente [2]	16
Figura 3. Objetos A y B relacionados por medio de una asociación en la que se indica un rol y multiplicada cualquiera. Fuente:[2]	16
Figura 4. Taxonomía de los coronavirus, los marcados en azul corresponden a HCoV. Fuente: 11	18
Figura 5. Estructura vírica Fuente: [8]	18
Figura 6. Estructura secundaria (Lámina β y Hélice α). Fuente: [14]	21
Figura 7. Diagrama del mecanismo de actividad de una enzima. Fuente: [14]	24
Figura 8. Parte de la vista proteoma Fuente: [1]	27
Figura 9. Parte de la vista proteoma Fuente: [1]	29
Figura 10. Parte de la vista proteoma Fuente: [1]	29
Figura 11. Parte de la vista proteoma Fuente: [1]	31
Figura 12. Parte de la vista pathways Fuente: [1]	34
Figura 13. Parte de la vista pathways Fuente: [1]	36
Figura 14. Parte de la vista pathways Fuente: [1]	37
Figura 15. Parte de la vista pathways Fuente: [1]	37
Figura 16. Parte de la vista pathways Fuente: [1]	38
Figura 17. Parte de la vista pathways Fuente: [1]	39
Figura 18. Adhesión del coronavirus con la membrana celular. Fuente [24]	41
Figura 19. Endocitosis del SARS-CoV [2]	42
Figura 20. Se muestra la relación S/ACE2 con ADAM17 y TMPRSS2 para lograr la fusión con la membrana celular, la endocitosis del coronavirus y los efectos sobre otros sistemas celulares responsables de una respuesta inflamatoria AGTR1, TGFbR. El color rojo indica que la enzima está activada Fuente: 20	43
Figura 21. Diagrama de la vista estructural. Fuente: Elaboración propia	51
Figura 22 Basado en [37]	60

LISTA DE TABLAS

Tabla 1. Tipos de citoquinas. Fuente: [15,16]	22
Tabla 2. Tabla resumen de las vistas modeladas de CSHG v3. [4,5,6]	26
Tabla 3. Resumen de los procesos víricos asociados a cada proteína viral del SARS-CoV, explicando las alteración que produce en la célula. Fuente: [20]	48
Tabla 3 Lista de limitaciones detectadas en el modelo CSGH v3 para expresar la patogénesis del cors vírica. Fuente: Elaboración propia.....	55
Tabla 5. Listado de modificaciones postraduccionales más conocidas divididas según su actividad sobre la proteína diana. Fuente [38][39]	61
Tabla 6. Coste de personal desglosado por individuos implicados en el trabajo. Fuente: Elaboración propia	67
Tabla 7. Coste del software desglosado. Fuente: Elaboración propia	68
Tabla 8. Coste del hardware desglosado. Fuente: Elaboración propia.....	68
Tabla 9. Coste total del proyecto. Fuente: Elaboración propia.....	68

LISTA DE ABREVIATURAS, SIGLAS Y SÍMBOLOS

CSHG	Conceptual Schema for the Human Genome
SARS-CoV-1	Síndrome respiratorio agudo grave por coronavirus 1
SARS-CoV-2	Síndrome respiratorio agudo grave por coronavirus 2
COVID-19	Enfermedad del coronavirus
SI	Sistemas de información
CM	Conceptual Modelling
ACE2	Enzima convertidora de angiotensina 2
TMPRSS2	Proteasa transmembrana tipo 2

CTSL	Cathepsin L
ADAM17	Desintegrina y metaloproteinasa dominio 17
RAS	Sistema renina-angiotensina
SDRA	Síndrome de distrés respiratorio agudo
AGT1R	Receptores de angiotensina tipo 1
ANG	Angiotensina
-CO-	Grupo carboxilo
-NH-	Grupo amino
ARN	Ácido ribonucleico
ARNm	Ácido ribonucleico mensajero
ARNt	Ácido ribonucleico de transferencia
ADN	Ácido desoxirribonucleico
MTP	Modificación postraducciona
IL	Interleucinas
TNF	Factores de necrosis tumoral
IFN	Interferones
TGF- β	Factor de Crecimiento Transformante β
AT1-7	Angiotensina 1-7
GTP	Guanosín trifosfato
ATP	Adenosín trifosfato
ncRNA	ARN no codificante
CDS	Secuencias de codificación
UTR	Región no traducida
EC	Número de Comisión de Enzimas
KM	Constante de Michaelis

DOCUMENTO 1: MEMORIA

CAPÍTULO 1. MOTIVACIÓN Y OBJETIVOS

Esta tesis de máster se enmarca en el ámbito de investigación de la ingeniería biomédica, concretamente en la calidad y análisis de datos genéticos y su uso en el diagnóstico clínico. En este capítulo se describe la motivación para realizar este proyecto, los objetivos y la estructura seguida.

1.1 MOTIVACIÓN

La Medicina de Precisión es un nuevo paradigma de prevención, diagnóstico y tratamiento de enfermedades. En especial han alcanzado un gran protagonismo durante la pandemia de COVID-19 con el objetivo de poder describir y entender su comportamiento virulento, así como sus graves consecuencias en el organismo a nivel celular y tisular. La correcta descripción y análisis es pues clave para poder desarrollar estrategias de intervención multidimensionales que van desde del desarrollo de vacunas, terapias, mecanismo de detección, origen y surgimiento de nuevas variaciones del virus.

No obstante, para poder aplicar exitosamente la medicina de precisión a partir de la descripción y análisis de la patología es necesario integrar información proveniente de diversas áreas del conocimiento. Sin embargo, se hallan dos obstáculos en el camino, por una parte la información disponible es heterogénea y dispersa con cientos de diferentes fuentes de datos genómicos. Y por otro lado la complejidad de los procesos biológicos relacionados con la patogénesis de un patógeno viral, que dan como resultado numerosos datos experimentales fruto de la diversidad de tecnologías de utilizadas por lo que existe una gran variabilidad en la calidad de la información disponible.

El análisis genómico se perfila, así como herramienta que ayude a dar con las respuestas a estas problemáticas. El análisis genómico se basa en la concurrencia de la biología molecular y la ciencia de la computación, para la creación de herramientas bioinformáticas que se fundamentan en el modelado abstracto de la información genómica con distintos niveles de organización, también conocido como modelado conceptual.

Mediante el modelado conceptual es posible tener una visión holística de todos los niveles genéticos y biológicos que participan, en este caso en el desarrollo de una infección por medio de la caracterización de cada uno de los elementos que componen el esquema de patogénesis en este caso del COVID19. De cada elemento se genera su perfil genético, proteico y de pathways, de forma que se facilita la comprensión y estudio de los mecanismos de patogénesis ya que también se referencia a bibliografía científica y registros de bases de datos que contengan al elemento de estudio.

Esto cobra especial importancia para identificar marcadores o indicadores biológicos relativos a una infección, que pueden ser tan específicos con identificar el marcador biológico que causa que el virus puede entrar en ellas mientras tiene vetada la entrada en otras. Esto es pues un paso hacia la medicina personalizada donde a partir del estudio de nuestro genoma y proteoma sea posible, predecir como de grave será el avance y pronóstico de una enfermedad infecciosa vírica o incluso en el diseño de fármacos y terapias

Con esto en mente, resulta claro descubrir porque la caracterización de enfermedades infecciosas mediante modelado conceptual, es tan importante para la investigación desde una visión holística de la patogénesis de virus, en este caso el COVID19. La enfermedad del coronavirus ha sido elegida dada la relevación actual que tiene en el panorama médico como resultado de la pandemia, sirviendo como un ejemplo claro de como se podría beneficiar la comunidad científica de la aplicación del modelado conceptual a una enfermedad, aunque dicho modelo pueda servir para cualquier otra enfermedad infecciosa de origen viral.

1.2 OBJETIVOS

El principal objetivo del presente trabajo consistió en la instanciación del esquema conceptual del genoma humano proporcionado por el Centro de Investigación en Métodos de Producción de Software (PROS) aplicado a la patogénesis de la enfermedad del coronavirus proporcionado por [1] en incluido en el anexo I con el fin de servir para la extracción de información de valor clínico diagnóstico o para el tratamiento a partir del modelo instanciado.

Para ello se establecieron los siguientes objetivos secundarios:

- Caracterización de las vistas proteína y pathway para generar una herramienta de modelización denominada vista estructural
- Compresión y descripción de los mecanismos de infección del SARS COV
- Caracterización de los elementos presentes en la patogénesis del virus
- Identificación y caracterización elementos para ampliar el esquema conceptual del genoma humano CSHG en trabajo futuro, si procediese.

CAPÍTULO 2. INTRODUCCIÓN

RESUMEN: Para realizar el análisis y moderación de datos biológicos, es necesario familiarizarse con los fundamentos del campo desde la perspectiva de la biología molecular y la genética. Describir el genoma humano como un dominio en un sistema de información requiere la integración de todos los conceptos importantes en un esquema conceptual y este es un paso esencial en el análisis genético variación. El dominio debe ser descrito bajo los principios de la estructura biológica. Por lo que en el siguiente apartado se ha realizado la descripción de los fundamentos del modelado conceptual, la patogénesis, el coronavirus y las proteínas

2.1 MODELADO CONCEPTUAL

2.1.1 Fundamentos del modelado conceptual

El modelado conceptual (CM) es un artefacto (también llamado esquema, guion, documento, diagrama) que tiene como objetivo obtener y definir conocimiento sobre un sistema, entendiéndose por sistema un conjunto de elementos relacionados entre sí y con su entorno [2]. Para modelar conceptualmente dicho sistema se requiere de la especificación de un dominio, pues permite definir la identificación de las propiedades relevantes del sistema, en un instante determinado, y el comportamiento esperado desde del punto de vista del observador. Por lo tanto se puede describir en los siguientes puntos como ha de ser un CM y sus requisitos [3]:

- Es una representación de la realidad o una representación de lo que alguien tiene en mente para un dominio; la conceptualización de alguien de un dominio
- Debe tener un nivel de abstracción suficientemente alto;
- Debe tener semántica explícita
- Es creado y utilizado por personas
- Requiere un propósito

El modelado conceptual también tiene fundamentos en matemáticas, en especial la teoría y lógica de conjuntos y conceptos de filosofía tales como las ontologías que se conciben como un marco común o una estructura conceptual sistematizada y de consenso no sólo para almacenar la información, sino también para poder buscarla y recuperarla. [3,4]

Por tanto cada sistema tiene una serie de objetivos, que deben ser tomados en cuenta cuando se realiza la descripción de sus características relevantes. En el campo de aplicación de este proyecto, este sistema toma la forma de los sistemas de información (SI) que se describen en un entorno de ingeniería de software que posteriormente se convierten en implementaciones que ayudan a sistemas

de cualquier tipo a cumplir sus objetivos a través de artefactos software. Por lo que SI es el sistema que recoge, almacena, procesa y distribuye información sobre las propiedades relevantes de un sistema y su entorno (dominio), con la intención de que el sistema en cuestión cumpla sus objetivos. [5]

Actualmente en el análisis genómico en las herramientas bioinformáticas trabajan fundamentalmente con modelos abstractos de la información genómica obtenidos a partir de estudios de biología molecular. Por lo que bajo esta perspectiva, el genoma humano puede ser visto como un SI y toda la información asociada a su dominio puede ser descrita en un modelo conceptual y lograr que biólogos e informáticos hablen un lenguaje común.[4]

2.1.2 Caracterización de elementos conceptuales

Un modelo conceptual se basa en la relación concepto-instancia, donde se relacionan mediante asociaciones los conceptos que son relevantes en el dominio de un sistema, en este caso, el genoma humano especialmente relacionado con el proteoma y los pathways celulares.

Un concepto es una idea formada por la generalización de las propiedades más relevantes que tienen en común un conjunto de instancias; por otro lado una instancia o entidad es la ocurrencia de un concepto determinado con valores específicos para las propiedades relevantes. La relación concepto-instancia, también puede ser vista como clase conceptual-objeto conceptual, no confundir esta relación con su significado en programación orientada a objetos. [2]

La conceptualización que se haga del dominio en un modelo conceptual debe ser flexible para insertar o borrar clases de conceptos por lo que sumado a todos estos elementos, se debe conseguir que el CM posea dos propiedades muy importantes: la completitud y la corrección. Esto con la finalidad de garantizar que la descripción que se hace en el modelo conceptual del dominio, es una representación válida en cualquier momento. [2,3]

La completitud se refiere a que la base de información del SI debe contener todos los conceptos relevantes del dominio. Mientras que la corrección se refiere a que todos los conceptos almacenados en la base de información del SI deben ser correctos.

Por ello los modelos conceptuales incluyen restricciones de integridad en sus descripciones del dominio para garantizar su completitud y su corrección. Una restricción de integridad es una condición que la base de información debe satisfacer para garantizar que es una representación válida del dominio. La base de información es consistente si satisface todas sus restricciones de integridad.

2.1.3 Técnicas de modelado conceptuales

Existen muchas técnicas de modelado, con distintos grados de sofisticación, para describir el modelo conceptual.[1]

- **Diccionario/Glosario:** Es una lista de clases de con sus atributos y relaciones, es decir, es un diccionario de sinónimos o diccionario de datos asociado; los nombres pueden considerarse estándar/normativos
- **Diagrama de Entidad-Relación:** Es un lenguaje gráfico que introduce estructura para las clases de conceptos. Es utilizado para el diseño de bases de datos.

- **Diagrama de Clases:** Es una extensión del diagrama Entidad Relación que añade características expresivas como herencia, modificadores, etc.
- **Sistemas ontológicos.** Además de estos enfoques de MC puros, las llamadas ontologías genómicas componen una familia de soluciones complementarias, como la Fundación de Ontología Biológica y Biomédica Abierta (OBO) [4]. OBO es una entidad cuya misión es proporcionar un conjunto de principios de ontología de diseño. Las denominadas ontologías OBO son gráficos acíclicos dirigidos libremente jerárquicos, por ejemplo, un concepto puede tener más de un término padre. Intentan organizar el conocimiento del dominio en dos dimensiones diferentes: granularidad y relación con el tiempo.

2.1.4 Diagrama de clases

Un diagrama de clases es un tipo de diagrama que describe la estructura de un sistema mostrando sus clases, atributos y las relaciones entre ellos. Los diagramas de clases son utilizados durante el proceso de análisis y diseño de los sistemas, donde se crea el diseño conceptual de la información que se manejará en el sistema, y los componentes que se encargaran del funcionamiento y la relación entre uno y otro.

En diagrama de clase se identifican los siguientes elementos: [1,7]

- **Clase :** Representa un propiedad relevante del dominio del problema, sobre el que el sistema debe almacenar información. Una clase es la agrupación de un conjunto de propiedades y características comunes que presentan elementos del dominio, a dichas características se las denomina atributos tienen un nombre y un rango posible de valores que pueden ser mutables o inmutables.
- **Asociación:** Representa un tipo de relación bidireccional entre dos o más conceptos relevantes del dominio del problema que el sistema debe conocer. Una relación es una característica que vincula conceptualmente a varias clases conceptuales distintas. Cada clase conceptual u objeto conceptual juega un rol conceptual en ese vínculo.
- **Composición:** Se trata de una asociación especial que representa el concepto de ser-parte-de, en concreto: Una parte sólo puede pertenecer a un todo, una parte no puede existir sin pertenecer a un todo, la eliminación del todo implica la eliminación de todas sus partes. Se caracteriza por ser una relación transitiva y antisimétrica, puede ser recursiva.
- **Agregación:** En el otro extremo de la composición tenemos esta asociación puesto aunque es una asociación que describe una relación entre un todo y sus partes de modo que las partes pueden existir por sí mismas. La representación gráfica de ambas asociaciones se puede observar en la Figura

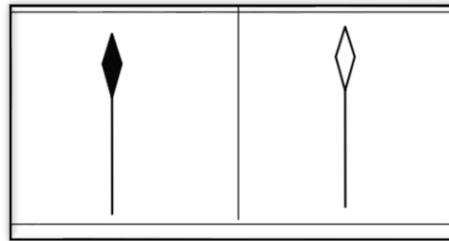


Figura 1. Simbología de la composición (rombo negro) y la agregación (rombo blanco).
Fuente: [7]

- **Los roles:** indican el papel que juega cada una de las clases que participan en una asociación, en asociaciones de una entidad consigo misma o cuando existe más de una asociación entre dos clases.
- **Generalización/especialización:** A veces, algunos de los conceptos del dominio del problema presentan entre ellos relaciones del tipo es-un, (Figura 1). Estos conceptos suelen tener propiedades comunes, que al modelarlos conceptualmente aparecen como atributos o asociaciones comunes. La clase más general (la superclase), contiene todas las propiedades (atributos y asociaciones) comunes, que son heredados por las clases más específicas (las subclasses). Todas las instancias de las subclasses se consideran también instancias de la superclase. La generalización es una relación transitiva y antisimétrica

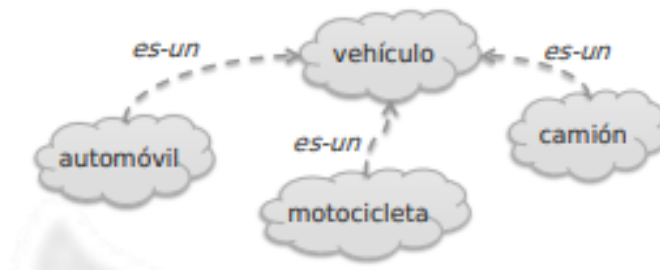


Figura 2. Ejemplo de asociación entre la entidad vehículo con las entidades automóvil, motocicleta y camión por medio de la relación *es-un*. Fuente [2]

- **Objeto:** Cada ocurrencia o instancia de una clase. (Figura 3)
- **Enlaces:** Cada ocurrencia o instancia de una asociación.



Figura 3. Objetos A y B relacionados por medio de una asociación en la que se indica un rol y multiplicada cualquiera. Fuente:[2]

- **Cardinalidad:** La cardinalidad indica los números mínimo y máximo de objetos de la otra clase con los que puede estar asociado mediante enlaces de la asociación (Figura 2). Indica los números mínimo y máximo de instancias de la clase que se interrelacionan con una instancia concreta de la otra clase.

2.2 CORONAVIRUS

2.2.1 Características generales y taxonomía

Los coronavirus (CoV) son unos patógenos zoonóticos, es decir, su origen es en el mundo animal y puede ser transmitidos a través del traspaso de la barrera de especie por medio de un agente etiológico existente en otra especie diferente a la huésped, que actúe como puente de adaptación. Esta familia de virus puede infectar los sistemas respiratorio, gastrointestinal, hepático y nervioso central de humanos aunque son principalmente virus respiratorios, que causan inflamación severa y patología pulmonar que pueden ser fatales, las enfermedades más comúnmente asociadas son la neumonía y el síndrome de dificultad respiratoria aguda (SDRA) [8]

- **Neumonía:** La neumonía es una infección que inflama uno o ambos pulmones, provocando que los alvéolos pulmonares se llenen de material purulento, lo que provoca tos con flema, fiebre, escalofríos y dificultad para respirar.
- **SRDA:** El síndrome de dificultad respiratoria aguda ocurre cuando se acumula líquido en los alvéolos pulmonares. Este impide que los pulmones se llenen con suficiente aire, por lo que llega menos oxígeno al torrente sanguíneo.
- Los coronavirus son un grupo de virus envueltos con genomas de ARN no segmentados, monocatenarios y de sentido positivo. Según el Comité Internacional de Taxonomía de Virus, los coronavirus están clasificados en el orden Nidovirales, familia Coronaviridae, subfamilia Coronavirinae . Esta subfamilia se divide en cuatro géneros: Alfacoronavirus, Betacoronavirus, Gammacoronavirus y Deltacoronavirus. El género betacoronavirus a su vez se divide en cuatro linajes distintos: A, B, C y D se han asignado dentro del género Betacoronavirus (Figura 4). [8,9]

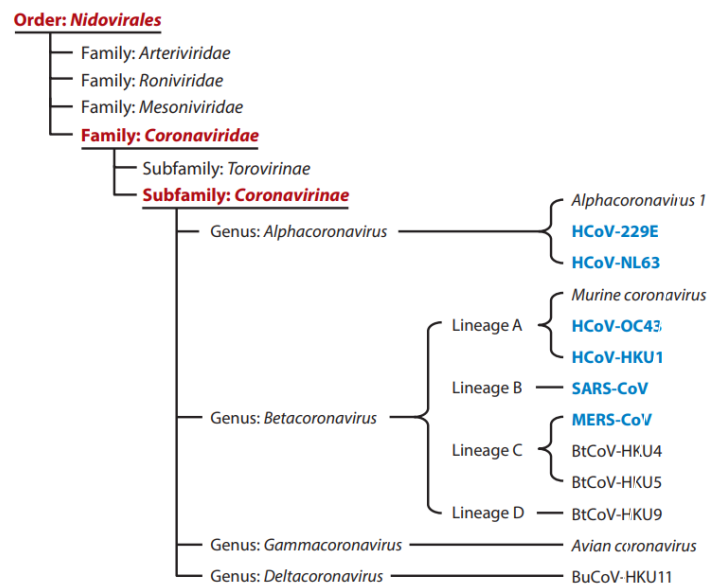


Figura 4. Taxonomía de los coronavirus, los marcados en azul corresponden a HCoV. Fuente: 11

2.2.2 Estructura vírica general

Estructuralmente, los coronavirus son virus esféricos, con envuelta y que contienen ARN monocatenario (ssRNA) de polaridad positiva de entre 26 y 32 kilobases de longitud [8].

Estos virus de ARN monocatenario de sentido positivo ingresan células epiteliales a través de proteínas de acoplamiento específicas, como ACE2 y TMPRSS2, y se replican, en el caso del SARS, en el citoplasma. [9]

El genoma del virus SARS-CoV-codifica cuatro proteínas estructurales: la proteína S (spike protein), la proteína E (envelope), la proteína M (membrane) y la proteína N (nucleocapsid).(Figura 5)

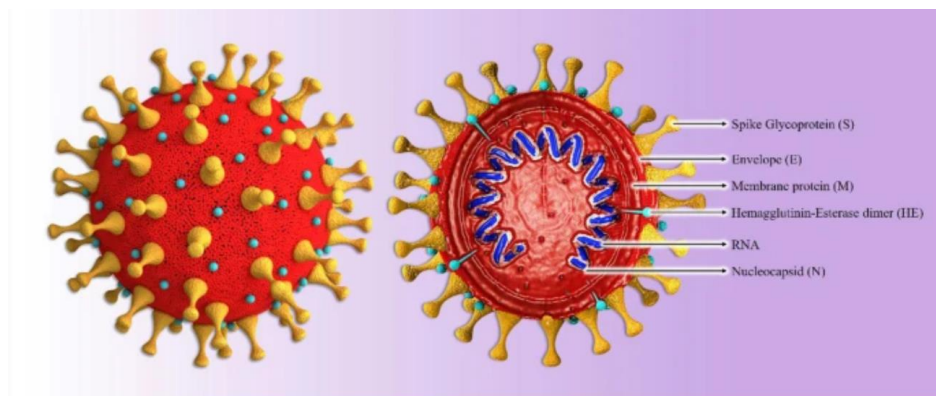


Figura 5. Estructura vírica Fuente: [8]

La proteína N está en el interior del virión asociada al RNA viral, y las otras cuatro proteínas están asociadas a la envuelta viral.

La proteína S contienen el dominio de unión al receptor celular y por lo tanto es la proteína determinante del tropismo del virus y además es la proteína que tiene la actividad de fusión de la membrana viral con la celular y de esta manera permite liberar el genoma viral en el interior de la célula que va a infectar [8].

Además, una pequeña proteína transmembrana conocida como proteína de la envoltura (E) también está presente en una pequeña cantidad en la envoltura.

2.3 PATOGÉNESIS VIRAL

2.3.1 Definición

La patogénesis viral es el proceso por el cual los virus producen enfermedad en el huésped, esta se inicia cuando entran en contacto una partícula viral y una célula susceptible que tras lograr superar las barreras protectoras naturales del organismo y evadir las defensas locales, hícticas e inmunitarias, se propagan por el cuerpo, y destruyen células ya sea en forma directa o desencadenando una respuesta inflamatoria e inmunitaria que puede causar daño al propio organismo. [10]

2.3.2 Relación virus – célula

Los factores que determinan la transmisión, diseminación y multiplicación viral y el desarrollo de la enfermedad en el huésped implican interacciones complejas y dinámicas entre el virus y el hospedador susceptible. Esta relación se lleva a cabo en diferentes fases [10]:

1. **Transmisión e ingreso del virus en el hospedador:** sucede la entrada en el organismo y se da inicio a la infección en el lugar primario.
2. **Periodo de incubación:** replicación en el lugar primario y diseminación al secundario. A través del aparato circulatorio y el sistema linfático el virus puede llegar a todo el organismo.
3. **Tropismo:** el virus colonizara a distintos tipos celulares siempre que estas contengan receptores específicos a las proteínas de superficie viral, por lo que se define como la capacidad de los virus para infectar una población de células separadas dentro de un órgano.
4. **Replicación viral en el tejido diana**
5. **Respuesta inmunitaria y posible acción inmunopatológica**
6. **Liberación del virus y posibilidad de transmisión y contagio**

Según como evolucione la infección es posible distinguir entre tres tipos: líticas, persistentes y abortivas. En todos estos tipos de infección, el evento inicial consiste en la interacción entre el virus y el receptor correspondiente presente en la superficie de la célula. No obstante, si una célula carece del receptor apropiado para un virus en particular; entonces es automáticamente resistente a la infección por ese tipo de virus. [10]

2.3.3 Tipos de infecciones virales

- **Infecciones líticas:** Los virus pueden matar a la célula hospedera al inhibir la síntesis de los ácidos nucleicos y proteínas celulares se provoca la destrucción celular debido al daño provocado por la replicación del virus y la alteración en la regulación del transporte y concentraciones de iones en la célula infectada [12]
- **Infecciones persistentes:** se caracterizan por la producción continua de virus infecciosos sin que las células hospederas mueran o sean destruidas, el virus se replica sin provocar la muerte de las células. Las infecciones persistentes son consecuencia de un delicado equilibrio que a veces se establece entre el virus y su célula hospedera. Según su persistencia se clasifican en crónicas, latentes donde sucede una replicación parcial vírica pero sin la capacidad de formar partículas víricas completas, recurrentes donde habrá periodos de latencia y replicación alternadamente y transformantes fruto de la transformación tumoral. [9,10]
- **Infecciones abortivas:** se caracterizan por una reducción en el rendimiento total de partículas virales debido a que la aparición de mutantes víricas que no pueden multiplicarse dentro de la célula. Estas surgen por defectos en la producción o procesamiento de cualquiera de los componentes necesarios para la multiplicación viral: ARN, ADN o proteína [8]

2.4 PROTEOMICA

2.4.1 Proteínas y proteoma

La estructura del organismo está constituida por proteínas. Las proteínas son el componente principal de las células y tienen como función básica la regeneración y reparación de tejidos corporales como el músculo, el cabello, las uñas, etc. No obstante, la labor que mayor protagonismo ha tomado en este trabajo ha sido su papel en la regulación de funciones metabólicas y el mantenimiento de la homeostasis celular, realizando funciones enzimáticas, hormonales, reguladoras, entre otras. [12]

Las proteínas son macromoléculas formadas por cadenas de unidades de aminoácidos que se encuentran unidos por medio de enlaces peptídicos entre los grupos carboxilo y el grupo amino [14]

El proteoma por su parte se trata del grupo completo de proteínas que se expresan o pueden expresarse a partir del genoma de un organismo. Las proteínas se elaboran en cantidades y tiempos diferentes, según cómo funcionen, cuando se necesiten y cómo interactúen con otras proteínas dentro de las células. La información sobre el proteoma puede ayudar a encontrar las proteínas involucradas en enfermedades víricas como las causadas por el coronavirus humano entre otras y a producir fármacos que impidan su acción. [13]

Dado que el marco de actuación de este trabajo es delimitado por la caracterización del proteoma celular durante la colonización viral del coronavirus, los proteomas estudiados han sido el proteoma celular de las células diana del SARS-CoV y CoVid19 tratándose estas del células respiratorias, cardiomiocitos y el sistema inmune así como el proteoma viral del coronavirus. Pues este, se determina a raíz del análisis de los cambios de las proteínas del huésped tras la infección del virus.

2.4.2 Biosíntesis y funciones biológicas

La síntesis de proteínas tiene dos etapas fundamentales: etapa de transcripción y etapa de traducción [14]

▪ Etapa de transcripción

Esta etapa ocurre en el núcleo de las células eucariotas, la secuencia se transcribe en una molécula de ácido ribonucleico (ARN), el cual se denomina como ARN mensajero (ARNm). La transcripción tiene tres etapas importantes.

- *Iniciación:* La enzima ARN Polimerasa II y los factores de transcripción, se unen a la caja TATAAA y sirve de señal para identificar el sitio donde inicia la transcripción.
- *Elongación:* En esta fase, el ADN se encuentra desdoblado y las bases nitrogenadas complementarias separadas.
- *Terminación:* El ADN de las células eucariotas contiene regiones de proteínas llamadas exones, y otras que no codifican proteínas llamadas intrones. El ARNm transcrito se le llama el nombre de ARNm inmaduro o pre-ARNm, y tiene que pasar por el siguiente por la eliminación de los intrones para convertirse en un ARN maduro.

▪ Etapa de traducción

Este proceso sucede en los ribosomas, que se encuentran en el citoplasma de la célula y se obtiene la cadena polipeptídica asociada al gen.

2.4.3 Estructura proteica

- **Estructura Primaria:** Está constituida por la secuencia de aminoácidos de la cadena polipeptídica.
- **Estructura secundaria:** Es el plegamiento que forma la cadena polipeptídica debido a la formación de puentes de hidrógeno entre los átomos que forman el enlace peptídico. Los puentes de hidrógeno se establecen entre los grupos -CO- y -NH- del enlace peptídico. Se pueden observar las estructuras: Hélice α o Lamina β (Figura 6) [14]
 - *Hélice α :* La estructura secundaria en la Hélice- α se forma cuando la cadena polipeptídica se enrolla de manera helicoidal, como una estructura en espiral, sobre un eje imaginario.
 - *Lámina β :* Esta estructura es conocida también como lamina plegada. La cadena queda estirada y en forma de zigzag formando láminas.

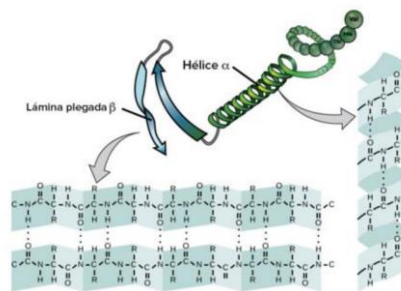


Figura 6. Estructura secundaria (Lámina β y Hélice α). Fuente: [14]

- **Estructura Terciaria:** La estructura terciaria ocurre cuando existen atracciones entre Láminas β y Hélices- α . Esta estructura es específica para cada proteína y determinará su función. Esta surge por la agrupación de conjuntos de estructuras denominadas dominios, que luego se articularán para formar la estructura terciaria definitiva. Se le llama dominio a las regiones de la proteína que tienen una estructura secundaria definida. [14]

La estructura terciaria da lugar a dos tipos de proteínas [14]:

- *Proteínas con estructura terciaria de tipo fibroso:* las hélices- α o láminas β que lo conforman, mantienen su orden y no tienen grandes modificaciones, solo ligeros giros longitudinales.
- *Proteínas con estructura terciaria de tipo globular:* su forma es aproximadamente esférica. En este tipo de estructuras se forman regiones con estructuras al azar, hélices- α y láminas β y acodamientos.

Cuando esta estructura se pierde, ocurre la desnaturalización de la proteína y por lo tanto también esta pierde su función. La mayoría de las proteínas se pueden desnaturalizar por calor, pH extremos, entre otros.[14]

2.4.5 Clasificación de proteínas

- **Citoquinas**

Su función en el organismo es coordinar la respuesta del sistema inmunológico, para ello reclutan o inhiben células específicas como células dendríticas o neutrófilos, a la vez que regulan la proliferación y diferenciación celular [15]. Existen distintas clases de citoquinas, con múltiples funciones biológicas, por lo que es posible encontrarse que algunas tienen funciones similares y otras antagónicas o incluso la misma citoquina puede tener además una función concreta sobre un tipo de células y una función completamente distinta sobre otro. [16]

Las citoquinas se pueden dividir en varios grupos, según su contexto de activación, la clase de células que las producen, etc. Por lo general, se encuentran los siguientes (Tabla 1)

Tabla 1. Tipos de citoquinas. Fuente: [15,16]

Grupo de citoquinas	Funciones
Interleucinas (IL)	<ul style="list-style-type: none"> ○ Regulación de la activación de las células del sistema inmune ○ Control de la diferenciación y proliferación de algunas subpoblaciones celulares. ○ Activación del endotelio y aumento la permeabilidad vascular. ○ Promueven la secreción de anticuerpos y controlan la respuesta de los linfocitos T.

Factores de necrosis tumoral (TNF):	<ul style="list-style-type: none">○ Presentes en las primeras etapas de la respuesta inflamatoria.○ Producidas por una gran variedad de células○ Papel central en infecciones virales, así como en la proliferación y muerte celular (apoptosis).
Interferones (IFN):	<ul style="list-style-type: none">○ Respuesta innata ante virus u otros microorganismos patógenos (los interferones α, β son secretados como señales de peligro p○ Promueven la actividad antiviral y la activación de las células Natural Killer (células NK).
Quimiocinas:	Estimulación de la motilidad de las células del sistema inmune (neutrófilos), y hacia el lugar de inflamación, mediante un fenómeno denominado quimiotaxis.

▪ Factores de crecimiento

Los factores de crecimiento, o factores tróficos, son un conjunto de proteínas que en conjunto con las hormonas y los neurotransmisores desempeñan una importante función en la comunicación intercelular. De todos ellos cabe resaltar el Factor de Crecimiento Transformante β (TGF- β) que presenta multitud de funciones. Entre ellas, destaca: la regulación de la proliferación y diferenciación de varios tipos celulares, pero también de la respuesta inmunitaria. Por lo que varios tipos de células inmunes, tales como linfocitos, macrófagos y células dendríticas sintetizan el TGF- β , donde tiene una acción inmunosupresora y antiinflamatoria. De hecho, es importante en la resolución de la inflamación y regreso al estado homeostático.¹ Si no se produce adecuadamente, puede persistir la activación inmunitaria, favoreciendo el daño al tejido y la aparición de enfermedades autoinmunes. [17]

▪ Hormonas proteicas

Las hormonas proteicas son de alto peso molecular y están formadas por aminoácidos. Algunas de las más importantes son las hormonas hipofisarias (TSH, prolactina, hormona de crecimiento, gonadotropinas), las del sistema inmune y las del páncreas (glucagón e insulina). Dada la naturaleza del coronavirus en esta sección tan solo analizaremos la angiotensina II (ANG2) que forma parte del sistema renina-angiotensina. Este sistema regula el control de la función cardiovascular y del equilibrio hidroelectrolítico e integrado por diferentes péptidos y enzimas que conducen a la síntesis de ANG2. La ANG2 se une y estimula receptores específicos, los Receptores de Angiotensina 1-7 (AT1-7), responsables de las acciones clásicas del péptido como vasoconstricción, retención de agua y sodio o proliferación celular. [18]

La Angiotensina 1-7 (A1-7) es un heptapéptido con efecto vasodilatador y antiproliferativo, que se produce a partir de la Angiotensina I por acción de una endopeptidasa o desde la Angiotensina II por acción de la ECA 2, enzima que puede también producir Angiotensina 1-9 (A1-9) a partir de la Angiotensina II. [18]

¹ Proceso de autorregulación mediante el cual los sistemas biológicos mantienen la estabilidad mientras se ajustan a las condiciones externas cambiantes. [X] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7076167/>

- **Enzimas**

Una enzima es un catalizador biológico, se trata de una proteína que acelera la velocidad de una reacción química específica en la célula pero no se consume durante la misma y se utiliza una y otra vez. Se caracterizan por ser proteínas globulares formadas por una o más cadenas polipeptídicas plegadas, creando un hueco donde encaja el sustrato y tiene lugar la reacción (Figura 7). Esta zona de la enzima se denomina centro activo y sólo unos pocos aminoácidos están implicados en él. La configuración tridimensional del centro activo es complementaria a la del sustrato y posee una distribución complementaria de sus cargas sobre la superficie de unión. Es decir, si una región del sustrato tiene una carga negativa, la zona correspondiente del centro activo tendrá una carga positiva y viceversa. [19]

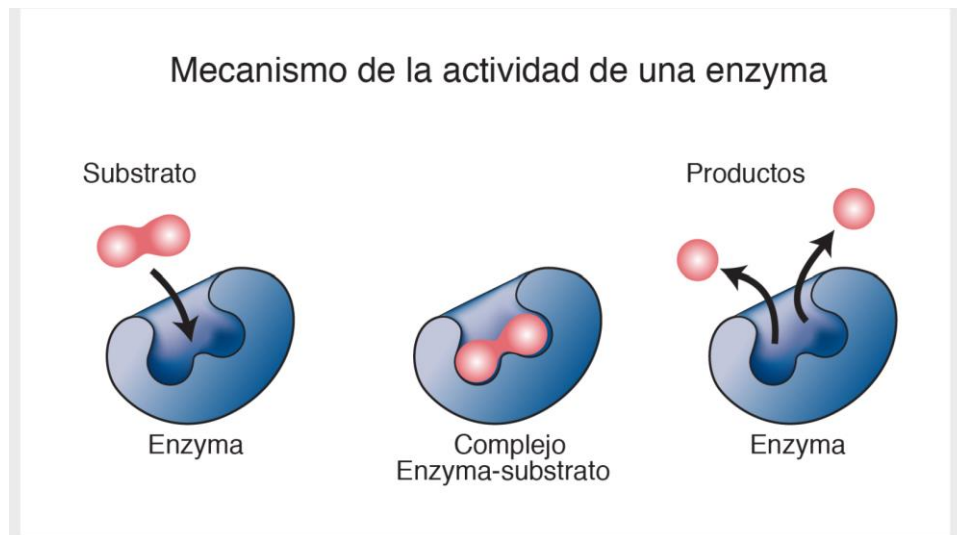


Figura 7. Diagrama del mecanismo de actividad de una enzima. Fuente: [14]

- **Canales de iones**

Los canales iónicos son proteínas integrales transmembrana oligoméricas² que, en respuesta a un estímulo adecuado forman un poro hidrofílico, que comunica los espacios intra y extracelular y permite el rápido paso a su través de determinados iones a favor de gradiente de concentración y de potencial eléctrico (gradiente electroquímico) generando una corriente iónica. [19]

- **Quinasas**

Es un tipo de enzima que transfiere grupos fosfatos desde ATP³ a un sustrato específico o diana mediante la fosforilación. Todas las quinasas necesitan un ion metálico divalente como el Mg²⁺ o

² Proteínas formadas por el ensamblaje de varias subunidades

³ ATP (adenosín trifosfato), biomolécula energética básica del metabolismo celular.

el Mn^{2+} para transferir el grupo fosfato [24]. La fosforilación forma parte de los MTP llevando a la activación o inhibición funcional en la transducción de señales y regulación a nivel celular.[19]

- **Peptidasa**

Las peptidasas son enzimas cuya activa se basa en romper ya sea enlaces peptídicos específicos (Proteólisis limitada), dependiendo en la secuencia de aminoácidos de la proteína, o reducir un péptido completo a aminoácidos. (proteólisis ilimitada). [19]

- **Fosfatasas**

Se engloban dentro del grupo de las fosfatasas todas aquellas enzimas que presentan actividad hidrolasa capaz de eliminar un grupo fosfato inorgánico a un sustrato. Dicho sustrato puede ser bien otra proteína u otro compuesto orgánico aminoacídico o bien un nucleótido. Al proceso bioquímico de la escisión de un fosfato se le denomina desfosforilación. [19]

- **Regulador/factor de la transcripción**

Un factor de transcripción es, por definición, cualquier proteína necesaria para iniciar el proceso de transcripción. Muchos de los factores de transcripción actúan mediante el reconocimiento de posiciones en los promotores o intensificadores de los genes [19]

- **Regulador/factor de la traducción**

La regulación de la traducción permite realizar cambios rápidos en la expresión del genoma transcrito para coordinar el crecimiento y diferenciación celulares con las condiciones fisiológicas y ambientales. La iniciación es la fase en la que actúan la mayoría de los mecanismos reguladores de la traducción eucariótica y la más relevante desde el punto de vista biológico y médico, ya que alteraciones de estos afectan al crecimiento y diferenciación celular en distintos organismos y son causa de enfermedades humanas. [14]

- **Transportador**

Las proteínas transportadoras son proteínas que transportan sustancias de un lado de una membrana biológica al otro. Muchas proteínas transportadoras se encuentran en la membrana de una célula, aunque también pueden encontrarse en las membranas de orgánulos internos como las mitocondrias, cloroplastos, nucléolos y otros.[19]

- **Grupo/Complejo**

La unión de varias proteínas en un grupo o complejo proteico se puede dar para llevar a cabo una acción o proceso conjunto. A continuación, se listan los siguientes proteínas que pueden formar grupos y/o complejos durante la patogénesis del SARS-CoV.[20] (Tabla 3, se encuentra su lugar de actuación (E: extracelular, C: citoplasma, N: núcleo, TRANS E: transmembrana) acompañado de un segundo paréntesis con el tipo de proteína. Otro: hace referencia a que no es ninguno de los tipos descritos en este documento pudiendo incluir ribosomas, proteínas *pocket*, proteínas nucleares)

CAPÍTULO 3. MATERIALES Y MÉTODOS

En este capítulo se ha realizado la descripción completa del esquema conceptual del genoma humano CSHG v3[1] (ver Anexo 1), que ha sido diseñado aplicando técnicas de Modelado Conceptual, específicamente diagramas de clases, a un dominio determinado: el genoma humano. La descripción es realizada de forma narrativa con la intención de asociar los elementos del dominio con las clases, atributos y asociaciones diseñadas en el CSHG v3. Con una breve descripción general de todas las vistas que lo componen y profundizando en la vista *proteoma* y *pathways*, dado que son las que forman el marco de interés este trabajo. A continuación se incluye la vista estructural y funcional desarrolladas a partir del dominio del modelado CSHG v3 y los mecanismos de patogénesis del modelo real de QIAGEN [20] (ver Anexo 2) por medio de la caracterización de todos los elementos involucrados y sus procesos celulares asociados. Para realización de la vista funcional se parte sobre la se diseño la base a la bibliografía científica de la patogénesis del SARS-CoV1 ya que debido a la gran similitud entre ambos virus y el mayor índice de bibliografía y estudios concluyentes en cuanto al comportamiento del SARS-CoV1, es más sencillo hacer la aproximación a los mecanismos de infección del SARS-CoV2.

3.1 CONCEPTUAL SCHEMA FOR THE HUMAN GENOME (CSHG)

CSHG versión 3 se trata del modelado conceptual del genoma humano diseñado y proporcionado por PROS para realización de esta tesis de máster. Este modelo parte de CSHG perspectiva holística del genoma humano pasando por distintas actualizaciones en la relación entre entidades, así como la adición de nuevos elementos con tal mejorar la caracterización de la complejidad del genoma humano [6]. CSHG actúa al estilo de una de base de datos conceptual proporcionando una perspectiva unificada de todos los datos relevantes que, en la práctica, se distribuyen a través de un conjunto de diferentes fuentes de datos como las ontologías [4].

Para lograr su objetivo CSHG está dividida en seis vistas diferentes interconectadas entre si [5] explicadas en la tabla 2:

Tabla 2. Tabla resumen de las vistas modeladas de CSHG v3. [4,5,6]

Nombre de la vista	Descripción
Vista gen	Responsable de modelar los elementos básicos de la secuencia de ADN
Vista de transcripción	Se hallan los componentes relacionados con la transcripción de ADN a ARN
Vista de variaciones	Responsable de caracterizar cambios en las secuencias de referencias con futuras implicaciones funcionales en la expresión del genoma

Vista proteoma	Caracterizar el conjunto de componentes que contienen los genes encargados de codificar las diferentes proteínas. Así como la conformación y estructura de estas que dará origen a distintas funciones y actividades a cada familia de proteínas
Vista pathways	Contiene información sobre rutas metabólicas celulares caracterizando el papel de componentes del genoma en dichas vías con las expresiones fenotípicas correspondientes.
Vista de bibliografía y banco de datos	Evaluación e identificación de la fuente de datos utilizada para la extracción de información de elementos del genoma humano secuenciados.

3.2 VISTA ESTRUCTURAL

Para formar la vista estructural que se ha utilizado de herramienta para el análisis y caracterización de la patogénesis del coronavirus, se ha utilizado la vista proteoma y la vista pathways.

La representación gráfica de la vista estructural completa explicada en este apartado se puede observar en el Anexo 1.

3.2.1 Vista proteoma

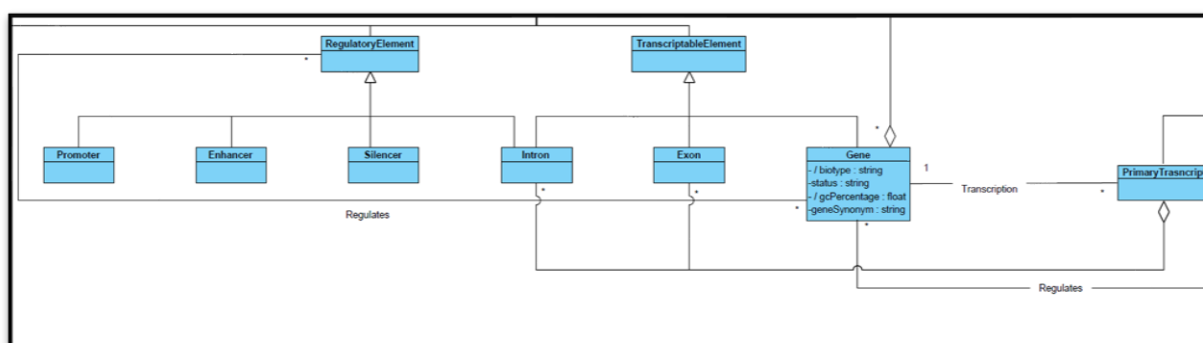


Figura 8. Parte de la vista proteoma Fuente: [1]

Analizando la Figura 8 podemos caracterizar los siguiente elementos y asociaciones:

- **Clase Gene:** Modela el concepto de gen genérico independiente y su descripción se realiza desde un enfoque estructural. Cada proteína puede estar codificada por uno o más genes y producida por diferentes componentes celulares en la célula.

En esta clase están presentes atributos como *biotype* que representa la forma típica o características generales del gen, el atributo *gcPercentage* hace referencia al el porcentaje de bases nitrogenadas en una molécula de ADN o ARN que son guanina (G) o citosina (C). Por último el atributo *geneSynonym* que hace referencia a otros denominaciones o nombres asociados a un gen determinado.

Esta clase presenta dos relaciones *regulates*, la primera con cardinalidad muchos a muchos con la clase *RegulatoryElement* para describir para controlar la regulación de un gen se pueden usar distintos elementos y a su vez que un elementos de regulación esta asociado a múltiples genes independientes. La segunda con cardinalidad de muchos a muchos con la clase *ncRNA*, esta clase representa a una molécula de ARN funcional, que a diferencia del ARN mensajero (ARNm) no se traduce en una proteína, esta asociación pues representa la función regulatoria del mismo durante la expresión génica.

Otras relaciones que contiene son la asociación *Transcription* con la clase *PrimaryTranscript* con cardinalidad de uno a muchos, por lo que un gen puede tener varios transcrito primario del ARN mensajero pero varios transcritos primarios solo pueden ser identificados a un único gen.

También existe una relación de agregación que va desde la clase *ChromosomeElement* a la clase *Gene* para describir que un gen genérico solo puede ser instancia de un cromosoma en el que se encuentra localizado y posicionado.

- **Clase *RegulatoryElement*:** Modela un elemento de regulación de la expresión génica genérico que presenta una especialización entre 4 subclases denominadas:
 - ***Promoter***, que hace referencia a una región de un gen, en la que proteínas relevantes (como la ARN polimerasa y factores de transcripción) se unen para iniciar la transcripción de ese gen
 - ***Enhacer***, que representa a cortas secuencias de ADN que ayudan a establecer la transcripción aumentado la actividad de la polimerasa.
 - ***Silencer***, un elemento específico de secuencia que induce un efecto negativo en la transcripción de su gen particular.
 - ***Intron***, es una región que reside dentro de un gen pero no permanece en la molécula final de ARNm maduro que aumenta los niveles de transcripción al afectar la tasa de transcripción, la exportación nuclear y la estabilidad de la transcripción.
- **Clase *TranscriptibleElement*:** Modela un elemento genérico que sea formado a partir de la transcripción de un gen o se pueda aplicar sobre él la transcripción. Contiene una especialización en tres subclases:
 - ***Intron***: tal y como se ha mencionado es una región que reside dentro de un gen pero no será traducida.
 - ***Exon***: representa la región transcrita del gen que forma parte del ARN transcrito y no será eliminado durante la maduración de este, pudiendo ser traducido a proteínas
 - ***Gene***: elemento explicado con anterioridad, representa secuencias del ADN humano que contienen el código para una proteína específica.

Las subclases *Exon* e *Intron* presentan una asociación tipo agregación con la clase *PrimaryTranscript* para describir que ambos elementos forman parte del transcrito primario del ARNm sobre el que se llevará a cabo un proceso de maduración.

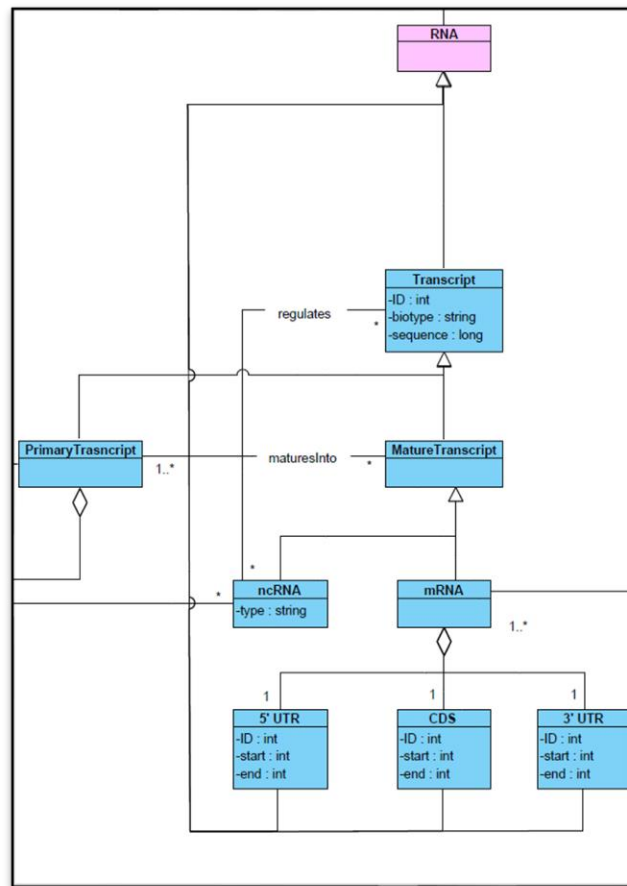


Figura 9. Parte de la vista proteoma Fuente: [1]

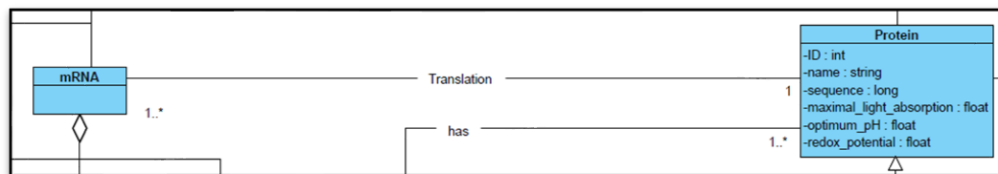


Figura 10. Parte de la vista proteoma Fuente: [1]

A continuación en la Figura 9 y 10 podemos ver como continua el esquema conceptual con los siguientes elementos y relaciones:

- **Clase *Transcript*:** Modela los ARNm de los que es subclase por medio de la especificación de la secuencia de transcripción de dicho ARN genérico. El atributo *ID* referencia nomenclatura identificativa de la secuencia, el atributo *biotype* representa la forma típica o características generales del transcrito y el atributo *sequence* contiene la secuencia de ARN que constituye el transcrito.

Esta clase presenta dos relaciones, la primera *regulates* con una cardinalidad de muchos a muchos con la clase *ncRNA*, esta asociación pues representa la función regulatoria del mismo

durante la expresión génica como el splicing⁴ de intrones. La segunda relación es una especialización en dos subclases *PrimaryTranscript* y *MatureTranscript*.

- **PrimaryTranscript:** Representa la secuencia de ARN transcrito del gen que está formado por la unión de los intrones o fragmentos no codificantes y los exones o regiones codificantes. Y requiere de modificación post transcripcional para la obtención de moléculas de ARN plenamente funcionales.
- **MatureTranscript:** Representa el ARN que ha sufrido transformaciones post transcripcional como el splicing de intrones, formación de caperuza cap y la adición de la cola poli A.

Existe una relación *maturesInto* entre *PrimaryTranscript* y *MatureTranscript* con cardinalidad uno a muchos y muchos a muchos respectivamente ya que el proceso de maduración del ARN en el transcrito primario depende de diversos procesos en especial el splicing que puede dar lugar empalmes y cortos alternativos que den lugar a ARN maduros diferentes aunque siempre asociados inequívocamente a un ARN primario.

Así, también podemos observar una relación de especialización de la clase *MatureTranscript* en las subclases:

- **mRNA:** Representa la formación del ARNm después de los procesos de maduración y que podrá ser traducida en proteínas. Esta subclase a su vez forma una asociación de agregación con *5'UTR*, *CDS* y *3' UTR* puesto que dichas subclases son dependientes y solo pueden estar asociadas a un ARNm.
 - **5' UTR:** representa a la porción de un ARNm desde el extremo 5' hasta la posición del primer codón utilizado en la traducción.
 - **3' UTR:** representa es la porción de un ARNm desde el extremo 3' del ARNm hasta la posición del último codón utilizado en la traducción.
 - **CDS:** Las secuencias de codificación (CDS) son regiones de ARN o ADN cuya secuencia determina la secuencia de aminoácidos en una proteína determinada.

Cada una de estas subclases están formadas por los atributos *ID*, *start* y *end* que hacen referencia al inicio, final e identificación de las porciones del ARN correspondientes a las subclases.

Otra relación a destacar de la clase *mRNA* es la denominada *Translation* con una cardinalidad de uno a muchos con la clase *Protein*, esta relación modela como la traducción del ARNm da lugar a diversas proteínas, pero como solo una proteína específica puede surgir de la traducción de un ARNm concreto. (Figura 10)

- **ncRNA:** Tal y como se mencionado antes, esta clase representa a las regiones del transcrito que durante el proceso de maduración no se traducirán pero servirán para

⁴ Proceso post-transcripcional de maduración del ARN, que se basa en el corte y empalme de regiones del ARN transcrito

regular los mecanismo que dan lugar a dicha maduración. Contiene el atributo *type* que referencia al tipo de ncRNA de un ncRNA cualquiera.

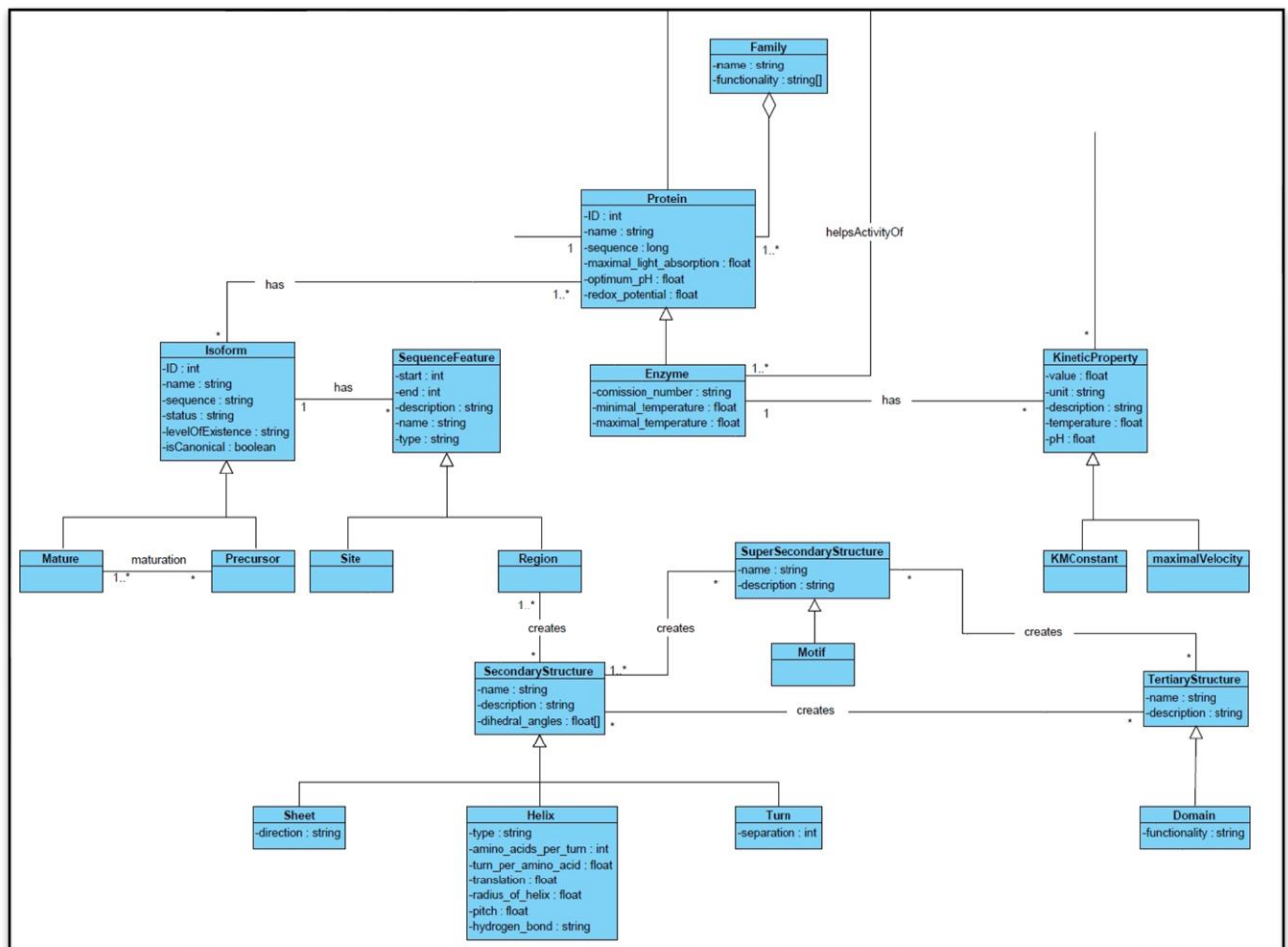


Figura 11. Parte de la vista proteoma Fuente: [1]

Para finalizar a partir de la Figura 11, se pueden ver los siguientes elementos y relaciones:

- Clase *Protein*:** Modela a una proteína genérica y se constituye de los siguientes atributos: El atributo *ID* y *name*, funcionan como identificador de la clase y al mismo tiempo describe el nombre de la proteína, por su parte el atributo *sequence* almacena la cadena de aminoácidos que describen una proteína, que recordemos es la estructura primaria de la proteína. Los atributos *maximal_light_absorption*, *optimum_pH*, *redox_potential* hacen referencia a propiedades físicas de la proteína que le permiten realizar su función, cambios externos que modifiquen o varíen algunas de estas características puede llevar a la desnaturalización de la proteína y su pérdida de función.

Forma cuatro asociaciones con otras clases entre las cuales ya se ha mencionado con anterioridad la relación *Transalation* que forma con la clase *mRNA*, por lo que a continuación se detallarán las otras tres.

Una de ellas es la relación *has* con una cardinalidad de uno a muchos con la clase *Isoform* hace referencia a la existencia de proteínas que son similares entre sí y realizan funciones similares

dentro de las células denominadas isoforma, estas pueden ser generadas por genes relacionados, o podrían generarse por el mismo gen a través del proceso de splicing alternativo, o maduración diferencial.

La segunda relación se trata de una agregación con la clase *Family* con una cardinalidad de uno a muchos, esto representa como cada proteína genérica puede estar asociada a una o más familias, algunos ejemplos de familia son las citoquinas, las fosfatas, las hormonas proteicas, los canales iónicos, etc.

- **Family:** esta formada por el atributo *name* que permiten identificar con un nombre a la familia de interés y el atributo *functionality* que describe la función de la familia

La tercera relación se trata de una especialización en la subclase *Enzyme*, que se realizó debido al interés biológico que rodea a estas proteínas pues controlan y regulan las reacciones químicas del organismo.

- **Enzyme:** modela a una enzima genérica especificando los atributos *comission_number*, *minimal_temperature* y *maximal_temperature*. El primero hace referencia al Número de Comisión de Enzimas (Número EC), se trata de una clasificación numérica para las enzimas, basado en las reacciones químicas que catalizan. La reacción química catalizada es la propiedad específica que distingue a una enzima de otra. El segundo y tercero por su parte hacen referencia a las propiedades físicas de la enzima relacionadas con la temperatura a la que pueden trabajar.

Esta subclase presenta una asociación denominada *has* con la clase *KineticProperty*, de forma que en conjunto con los atributos de *Enzyme* queda definidas todas sus propiedades. La cardinalidad es de uno a muchos, ya que cada enzima tiene asociada una cinética enzimática a sus reacciones, es decir, que dependiendo de las reacciones químicas que sean catalizadas por una enzima su velocidad variara.

- **KineticProperty:** Esta clase modela las propiedades cinéticas de las reacciones enzimáticas por medio de los atributos *value* (valor cinético), *unit* (unidades de medida), *description* (descripción de la reacción), *temperature* (temperatura a la que sucede la reacción) y *pH*.

Esta clase presenta una relación de especialización dando lugar a las subclases *KMConstant* y *maximalVelocity*, las cuales modelan la actividad enzimática por medio de la ecuación de Michaelis-Menten

- **Clase Isoform:** Como se ha mencionado con anterioridad, la isoforma hace referencia a cada una de las distintas formas de la misma proteína ya que por splicing alternativo la transcripción de un gen puede implicar a grupos diferentes de exones, resultando en distintas proteínas. Si bien muchos desempeñan funciones biológicas iguales o similares, determinadas isoformas tienen diferente actividad funcional. Sus atributos son su *ID*, que refiere a un identificador único, su *name*, nombre por el cual se conoce, su *sequence* que es la secuencia de aminoácidos que la forman, su *levelOfExistence* que es su prevalencia en la naturaleza y su *isCanonical* que hace referencia a si la isoforma es canónica o no, una isoforma de proteína se etiqueta como la

secuencia canónica basándose en criterios como su prevalencia y similitud con secuencias ortólogas ⁵(o funcionalmente análogas) de otras especies.

Esta clase presenta una especialización en las dos subclases *Precursor* y *Mature* que a su vez están asociadas por la relación *maturation*.

También cuenta con la relación *has* con la clase *SequenceFeature* con una cardinalidad de uno a muchos, ya que las isoformas asociadas a una proteína tienen varias posibilidades distintas de secuencias.

- **Clase *SequenceFeature*:** Esta clase modela todas las secuencias asociadas a la isoforma de una proteína por medio de los atributos *start* y *end*, que indican el inicio y fin de la secuencia así como su identificación por medio de *description*, *name* y *type*.

Contiene una especialización en las dos subclases *Site* y *Region* que hacen referencia a la estructura que presenta la isoforma. *Sites* hace referencia a los sitios de unión en la cadena polipeptídica donde los ligandos se unen mediante fuerzas débiles. Mientras que *Region* son las cadenas polipeptídicas de la proteína que se autoestabiliza y se pliega independientemente del resto. Cabe señalar que la subclase *Region* presenta a su vez una asociación por medio de la relación *creates* con la clase *SecondaryStructure* con cardinalidad uno a muchos.

- **Clase *SecondaryStructure*:** Tal y como se explica en el capítulo 2, en el apartado de estructuras protéicas, la proteína por medio de puentes de hidrógeno se pliega sobre sí misma da lugar a dos tipos de estructuras secundarias (alfa-hélice y lámina beta). Los atributos que modelan la clase son *name* para la identificación por su nombre, *description*, para su descripción y *dihedral_angles*, para conocer el ángulo entre dos planos o semiplanos que se cortan en la cadena polipeptídica.

Así, como habíamos adelantado se modulan los distintos tipos de estructuras secundarias mediante una relación de especialización así como la separación entre regiones de la cadena polipeptídica al tomar una forma u otra pues nunca toda cadena polipeptídica será puramente alfa-hélice o lámina-beta.

- ***Helix*:** hace referencia a la forma alfa-hélice y se modela por medio de los atributos *type*, *amino_acids_per_turn*, *turn_per_amino_acid*, *translation*, *radius_of_helix*, *pitch* y *hydrogen_bond*
- ***Sheet*:** hace referencia a la forma lámina-beta y se modela por medio del atributo *direction*
- ***Turn*:** hace referencia al cambio de forma de la cadena de una forma a otra y la separación de existe entre ambas ***separation***

La clase *SecondaryStructure* contiene un tipo más de asociación y esta es una relación *creates* con cardinalidad uno a muchos con la clase *SuperSecondaryStructure*, puesto que es frecuente encontrar combinaciones de estructuras al azar, a y b, con una disposición característica que se repite en distintos tipos de proteínas. Es posible que aparezcan motivos

⁵ Aquellas proteínas que provienen de un mismo gen y cuya divergencia se debe a un proceso de especiación.

estructurales también llamadas estructuras supersecundarias a las que da origen previamente una o varias estructuras secundarias.

- **SuperSecondaryStructure:** está modelado por los atributos *name* y *description* para su identificación. Y cómo hemos adelantado la especialización *Motif*, hará referencia a los motivos estructurales que se modelan como una subclase debido a la especificación que refiere en las proteínas.
- **Clase TertiaryStructure:** Al igual que la estructura secundaria, este tipo de estructura se explica en profundidad en el Capítulo 2, en el apartado sobre estructuras proteicas, donde se detalla como se trata de la disposición tridimensional de la proteína que es estabilizada gracias fuerzas de enlaces covalentes y no covalentes. Sus atributos son relativos a su identificación mediante un nombre y una descripción (*name*, *description*).

Esta clase contiene una especialización a la subclase **Domain** cuyo único atributo modelado es la funcionalidad (*functionality*). El *domain* se trata de una región de la cadena polipeptídica de una proteína que se autoestabiliza y que se pliega independientemente del resto.

3.2.2 Vista pathways

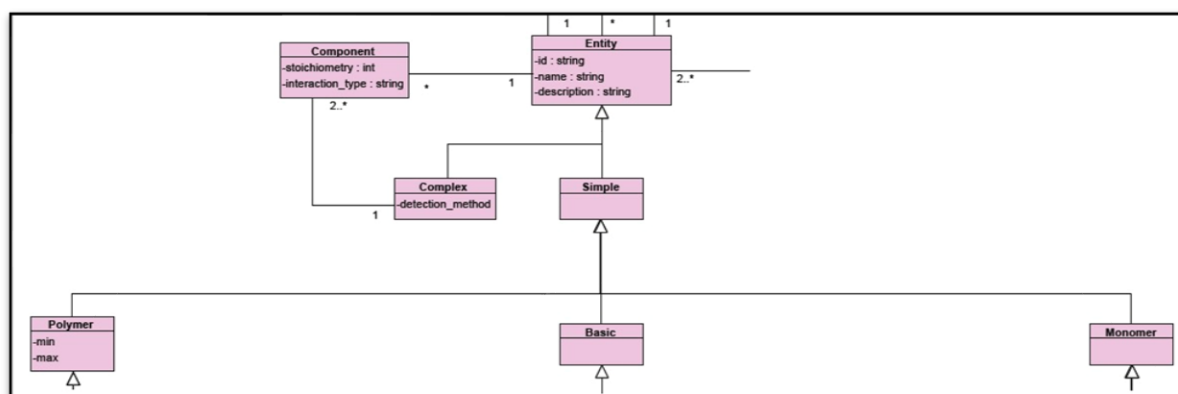


Figura 12. Parte de la vista pathways Fuente: [1]

- **Clase Entity:** En este modelo, se entiende como entidad cualquier elemento que forme parte de procesos y pathways celulares, pudiendo este elemento ser por ejemplo una proteína, una base nitrogenada, una hormona, etc. Para llevar a cabo esta modelización se utilizan los atributos *id*, *name* y *description* para identificar y delimitar el elemento junto a una asociación de especialización a clases que a su vez se especializan en más subclases para representar distintos tipos de biomoléculas. Estas subclases asociadas a **Entity** son **Simple** y **Complex**. (Figura 12)
 - **Complex** : Modela elementos complejos con el atributo *detection_method* que permite la identificación del elemento según su método de detección. Esta clase presenta una asociación con la clase **Component** con una cardinalidad de uno a dos o

muchos, esto quiere decir que harán falta por lo menos dos instancias *Complex* para formar una instancia *Component*. Los atributos de *Component* son *stoichiometry*, *interaction_type*.

○ **Simple:** Esta subclase modelo elementos simples que por medio de la asociación de especialización a otras subclases, puede ser definida como *Polymer*, *Basic* y *Monomer*

▪ **Polymer:** Se modela un polímero genérico que se trata de una sustancia compuesta por grandes moléculas, formadas por la unión mediante enlaces covalentes de una o más unidades simples llamadas monómeros. Presentan una especialización en tres subclases diferentes que hacen referencia polímeros biológicos que son *Saccharide*, *NucleicAcid* y *Peptide* (Figura 13)

• **Saccharide:** modela a los sacáridos que son la estructura unitaria de los carbohidratos. Esta clase se especializa en las subclases: *Disaccharide* (cadena de dos monosacáridos), *OligoSaccharide* (cadena de 3 a 10 monosacaridos) y *Polisaccharide* (cadenas de más de 10 monosacáridos)

• **NucleicAcid:** modela los ácidos nucleicos que son largas cadenas de nucleótidos concatenados. Existe una asociación de especialización en las dos subclases *RNA* y *DNA*.

• **Peptide:** modela las cadenas cortas de aminoácidos que cuando se organizan en cadenas largas dan lugar a oligopéptidos o proteínas según el número de péptidos. Tiene lugar una especificación con la subclase *Oligopeptide* y *Protein*

▪ **Basic:** se modela en base a su asociación de especialización en la subclase *Cofactor*, un cofactor es un compuesto químico no proteico que se une a una enzima y es necesario para su actividad cinética. Así, tal y como se puede prever las asociaciones de *Cofactor* será con la clase *Enzyme* mediante la relación *helpsActivityOf* con una cardinalidad de uno a muchos por lo que una enzima puede tener múltiples cofactores pero un cofactor solo puede estar asociado a una enzima. Y otra con la clase *KIneticProperty* mediante la relación *forA* con una cardinalidad de uno a muchos. (Figura 14)

Los atributos de *Cofactor* referencian sus propiedades físicas y químicas mediante *is_organic*, *is_water_soluble* y *is_lipid_soluble*, para caracterizar su solubilidad en agua o lípidos y su naturaleza.

▪ **Monomer:** Modela la unidad estructural de los polímeros con la que mediante la repetición y enlace de monómeros se forman dichos polímeros. Se especializa en tres subclases *Aminoacid*, *Monosaccharide* y *Nucleotide* (Figura 15)

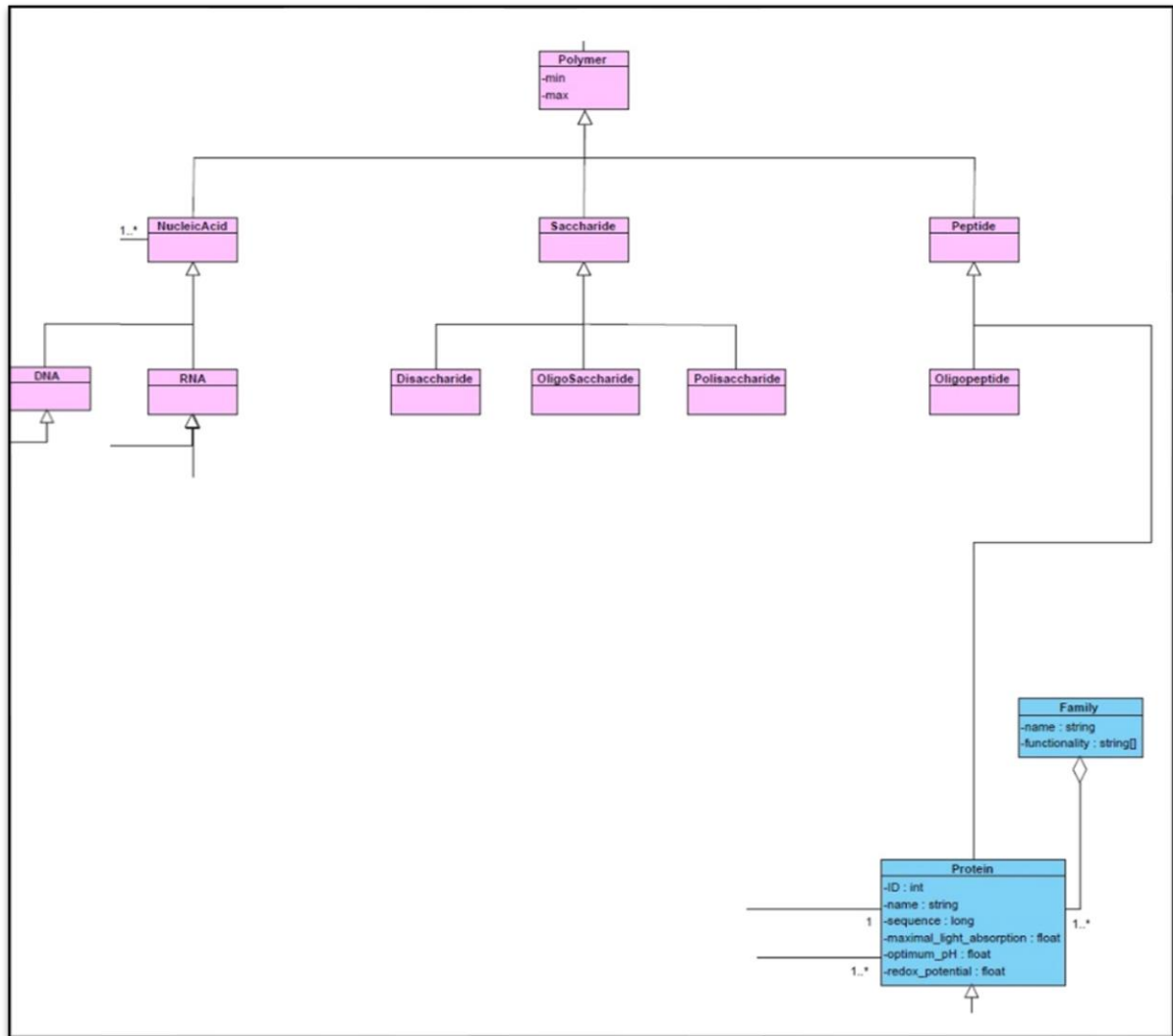


Figura 13. Parte de la vista pathways Fuente: [1]

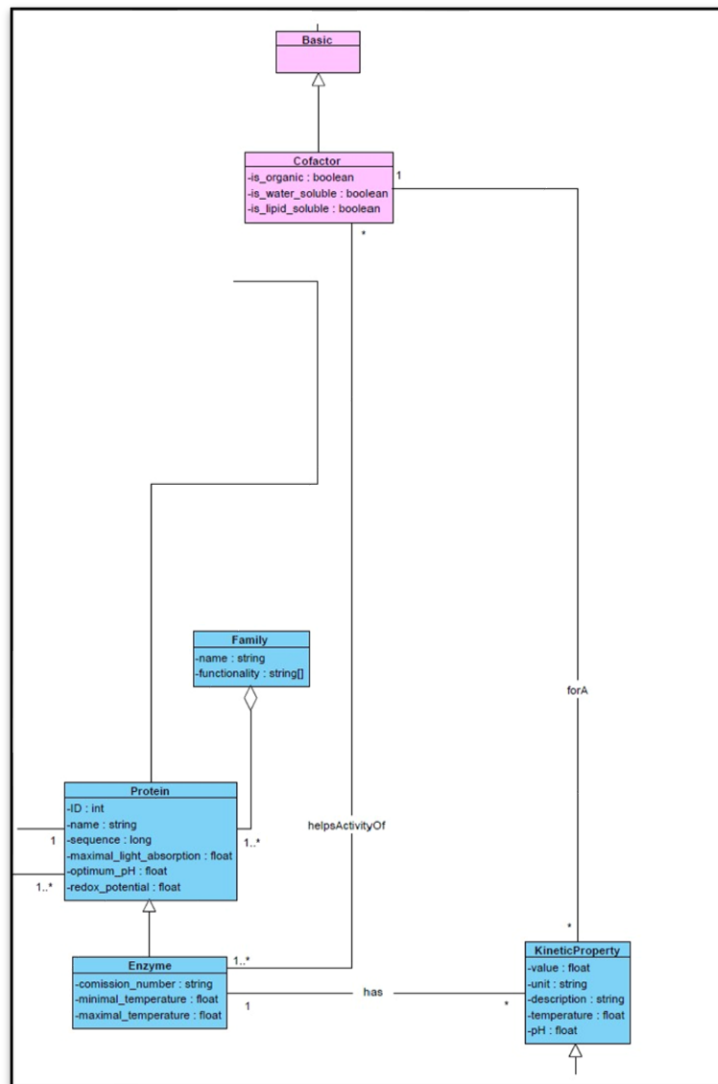


Figura 14. Parte de la vista pathways Fuente: [1]

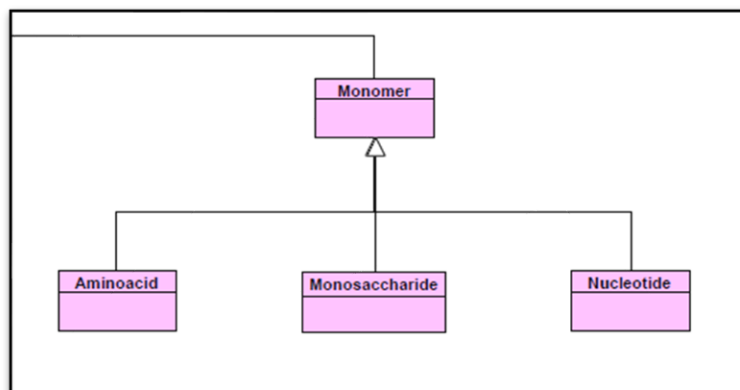


Figura 15. Parte de la vista pathways Fuente: [1]

A parte de la asociación por especialización **Entity** también mantiene asociaciones tres otras clases. La primera asociación que analizaremos será la que se forma entre **Entity** y **EntityRole**, por medio de la relación *has* con una cardinalidad de uno a muchos pues una entidad puede tener más de un rol pero hay un rol que se asocia de manera inequívoca a una entidad, por ejemplo el ADN tiene el rol de almacenar a largo plazo de información para construir otros componentes de las células que solo él puede realizar pero también tiene otros roles como la autoduplicación. (Figura 16)

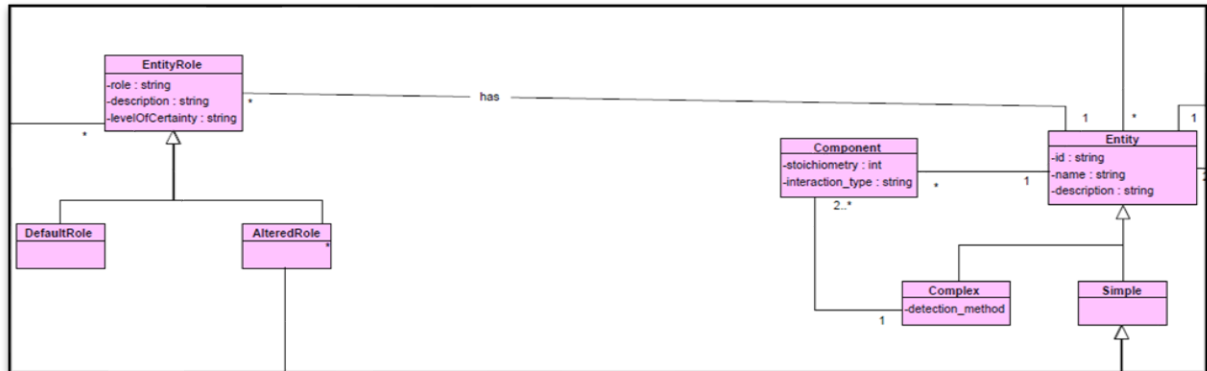


Figura 16. Parte de la vista pathways Fuente: [1]

- **EntityRole:** Sus atributos son *role*, como el rol que toma en un proceso o acción, *description*, y *levelOfCertainty*. Esta clase se especializa en otras dos subclases por un lado **DefaultRole** que hace referencia al rol estándar en un contexto fisiológico de la célula de un elemento, por lo que no funciona con normalidad, y por otro lado esta **AlteredRole** que ocurre cuando por motivos externos o internos el elemento deja de cumplir su rol asignado y este muta a otro diferente al original. (Figura 16)

La segunda asociación de **Entity** tiene lugar con la clase **entitySet** con una cardinalidad de dos a muchos, esto significa que harán falta por lo menos dos entidades para formar un set de entidades. Es destacable en esta relación la conexión que se produce con la clase **Event** desde la clase **entitySet** mediante la relación *interchangeableFor*, esto referencia que ambas clases son intercambiables unas por otras y por tanto, se puede tener una serie de instancia de **entitySet** puede también expresar un evento genérico que haya tenido lugar en la célula. (Figura 17)

La tercera asociación de **Entity** se con la clase **Takes_part** con una cardinalidad de uno a muchos, por lo que esta relación modela correspondencia entre una entidad los procesos en los que toma parte, pudiendo ser un solo proceso o diversos. Por ello, **Takes_part** lleva a cabo una especialización en tres subclases con tal de determinar específicamente cómo toma parte la entidad en el proceso. Distinguiendo entre *input*, *output* y *regulator*:

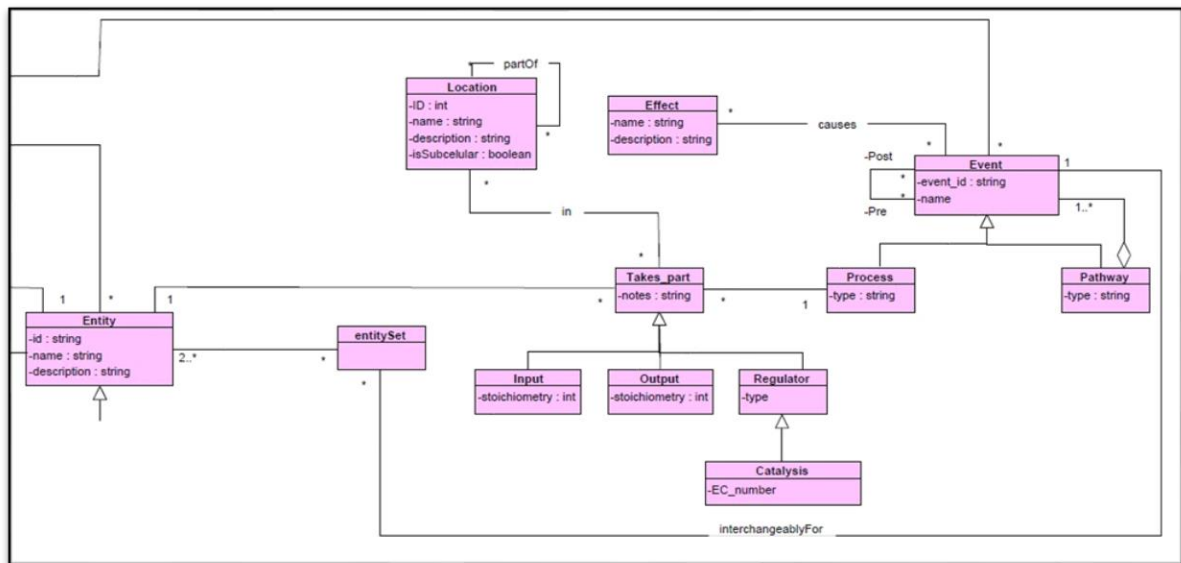


Figura 17. Parte de la vista pathways Fuente: [1]

- **Input:** se modela de forma que la acción del elemento es una condición necesaria para la realización del proceso y quien lo inicie.
- **Output:** se modela de forma que el elemento surge o se genera como consecuencia de un de la aplicación de un proceso
- **Regulator:** se modela de forma que la acción del elemento medie en el proceso favoreciendo que este se lleve a término como acelerándolo o actuando de forma antagónica inhibiéndolo. Por ello el atributo de esta clase es *type* que permite definir que tipo de regulador actuará. A su vez, presenta una especialización para el caso de la catálisis por medio de la subclase *Catalysis* con el atributo *EC_number*.

La clase **Takes_part** cuenta con otros dos tipos de asociaciones, una se trata de una relación *in* con la clase **Location** con cardinalidad de muchos a muchos, que modela en que lugar del espacio celular está teniendo lugar el proceso del que una entidad está tomando parte, en otras palabras, es el espacio de actuación de la entidad. Esta clase también cuenta con un tipo especial de asociación referida así misma *partOf* con una cardinalidad de muchos a muchos por lo que en una misma localización puede haber otras localizaciones relacionadas. Los atributos de esta clase son *ID*, *name*, *description*, para la identificación del lugar y *isSubcelular* para determinar si son localizaciones subcelulares. (Figura 17)

La otra asociación es referida a la clase **Process** con una cardinalidad de muchos a uno, por lo que varias entidades participantes del **Takes_part** son asociadas a un proceso correspondientes (Figura 17)

- **Clase Event:** esta clase modela los eventos celulares por medio de sus atributos *event_id* y *name* para conseguir la correcta identificación, así como por medio de una asociación referida a así misma para ordenar los distintos eventos que sucedan en pre y post, por tanto se le asocia un orden cronológico. Para continuar la modelización, sufre una especialización en dos subclases **Process** y **Pathway**. Por un lado la primera subclase alude a la serie de acciones

Llevadas a cabo por distintas entidades que con un principio y un fin, por otro lado la segunda subclase alude a todos los procesos que se han llevado a cabo, es decir, es la agrupación de procesos relacionados bajo un mismo nombre.

Conviene resaltar que la clase **Event** forma una asociación extra con la subclase **Pathway** por medio de una agregación con una cardinalidad de uno a muchos, por lo tanto, se modela cada evento teniendo asociado uno o varios pathways.

- **Clase Effect:** Esta clase modela por medio de sus atributos *name* y *description*, efectos debidamente identificados y varios eventos de forma que la dirección de la relación sea bidireccional, y por tanto efectos genéricos causen varios eventos y varios eventos puedan causar efectos en otros eventos, por medio de la asociación *causes* con una cardinalidad de muchos a muchos.

3.3 VISTA FUNCIONAL DE LOS MECANISMOS DE PATOGÉNESIS VÍRICA

Se ha llevado a cabo en este apartado una descripción narrativa de los mecanismos principales de patogénesis junto a la caracterización de los elementos implicados que se presentan ordenados en tablas con el objetivo de que puedan ser identificados con mayor facilidad en el rol empleado en el mecanismo y la naturaleza del elemento dentro de los tipos explicados en el capítulo 2 en el apartado de biología molecular. Para la realización de la caracterización de los mecanismos de patogénesis del coronavirus, se ha usado como base el artículo Coronavirus Pathogenesis Pathway [5] proporcionado por QIAGEN donde se describe mediante esquemas los mecanismos de respuesta antiviral del huésped.

3.3.1 Características generales de la patogénia del SARS-CoV-1

La patogénesis de la infección del SARS-CoV se relaciona con la propiedad del virus para modificar las vías normales de señalización celular (hipótesis de redirección de la señalización) a través de las proteínas virales capaces de desarrollar actividad de proteínas kinasas y citoquinas, que manipulan la maquinaria celular para favorecer la transcripción, traducción y multiplicación virales. Cuando el virión se liga a sus receptores y correceptores en la membrana celular, se desencadena una serie de señales a través de cascadas de fosforilación y defosforilación, ya sea para inhibir la expresión de genes que participan en el mecanismo de defensa celular, como es la vía de interferón, o bien para reprimir el mecanismo de muerte celular por apoptosis que permita realizar la multiplicación viral, hasta que finalmente la célula huésped muere, con la consecuente liberación de las partículas virales infecciosas.

Por lo que es importante tener en cuenta la evolución de la enfermedad de acuerdo a sobre que tejidos diana actúa para poder tener una imagen global de todos los mecanismos que encuentra el SARS-CoV1 para infectar el organismo. Por lo que teniendo el esquema general de infección vírica [21, 22]

3.3.2 Unión y entrada a la célula

Una vez el virus a logrado penetrar en el organismo trata de adherirse a la membrana plasmática de las células huésped por medio para ser fagocitadas y entrar en el citoplasma celular para dar comienzo a la multiplicación vírica. Así pues se ha dividido esta fase en dos procesos con tal de facilitar el análisis

y descripción de las interacciones entre el virus y la células, siendo estas **adhesión viral** y **endocitosis viral**.

- **Adhesión viral**

El virus SARS-CoV-1 se adhiere a la célula huésped por medio de la unión específica entre la glicoproteína espiga (S) de SARS-CoV y con el dominio catalítico de la enzima convertidora de angiotensina 2 (ECA2), una proteína transmembrana con actividad de monoxipeptidasa anclada a la membrana plasmática huésped. [23]. Este un paso crítico para la entrada del virus, en el cual la glicoproteína espiga (S) utiliza sus dos subunidades funcionales: la subunidad S1, responsable de la unión con el receptor de la célula huésped y la subunidad S2, responsable de la fusión del virus con las membranas celulares.[22]

La unión del virus a la ECA2 desencadena un cambio conformacional en la proteína S del coronavirus, dando lugar a que el complejo resultante active las proteasas transmembrana tipo 2 (TMPRSS2) y la proteína peptidasa desintegrina y metaloproteína dominio 17 (ADAM17) provocando la escisión proteolítica del ectodominio ACE2 y a la activación de la glicoproteína espiga (S). El resultado se esta serie de activaciones da lugar a la fusión del virus a la fusión del virus a la membrana celular y facilitando su ingreso del virus en el interior de la célula huésped. (Figura 18A y 18B) [24]

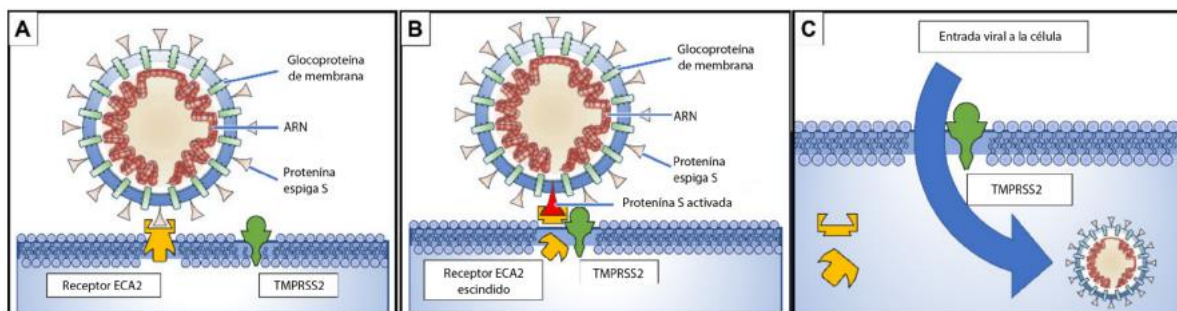


Figura 18. Adhesión del coronavirus con la membrana celular. Fuente [24]

- **Endocitosis viral del SARS-CoV**

Para estudiar la endocitosis del SARS-CoV debemos tener en cuenta que existen dos vías de acceso al citoplasma celular una endosómica y otra no-endosómica. En la vía no-endosómica una vez realizada la adhesión viral el ARN genómico viral se libera en el citoplasma y se desnuda para permitir la formación de las poliproteínas, la transcripción de los ARNs subgenómicos y replicación del genoma viral (Figura 18C). Por otro lado la vía endosómica tiene la particularidad de ser independiente de la activación de las TMPRSS2 (vía independiente de TMPRSS2). En este caso, tras la unión del virus a la

ECA2 se forma un endosoma dentro del cual la proteína S es clivada y activada por la cisteína-proteasa catépsina L (CTSL) dependiente de pH. (Figura 19) [25]

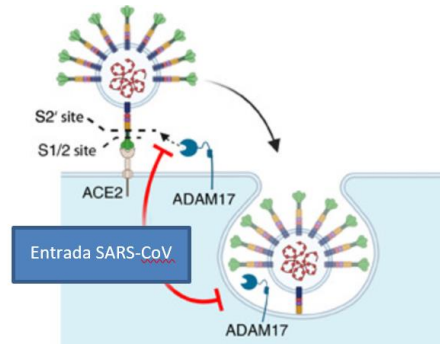


Figura 19. Endocitosis del SARS-CoV [2]

Como consecuencia de esta primera actividad viral, aparecen las primeras alteraciones en el comportamiento celular debido a la activación de las proteínas participantes puesto que se hayan interconectadas con otros procesos biológicos que a su vez afectaran a otros ocasionando finalmente una retroalimentación dañina en la regulación de las rutas o vías de señalización y/o expresión celular. Las principales proteínas activadas en estos procesos fueron: ACE2, TMPRSS2 y ADAM17 y en la **Figura 19** se incluye un esquema de todas las interacciones que se han llevado a cabo.[24]

Después de la interacción con la proteína S viral, ACE2 se internaliza junto con las partículas virales, lo que reduce la expresión de ACE2 en los tejidos superficiales. La internalización de ACE2 puede desencadenar una retroalimentación positiva dañina para los tejidos. La modulación negativa de ACE2 conduciría a un aumento en el nivel de Ang II y la activación del receptor A1T, que a su vez activaría más ADAM17 (**Figura 19**). Por lo tanto, la reducción exacerbada de ACE2 es, indirectamente, responsable de su propia escisión por ADAM17 y, en consecuencia, del agotamiento de ACE2 en el tejido y la liberación de la forma sACE2 (68).[25]

Dado que la molécula ACE2, además de ser un receptor de SARS-CoV y SARS-CoV-2 (29–32), reduce la actividad del sistema renina-angiotensina (RAS) al convertir la angiotensina tipo 1 (ANG I) y tipo 2 (ANG II) en Ang 1 -9 y Ang 1-7 respectivamente. Los descensos en su actividad juegan un papel importante en la protección frente a algunos trastornos como las complicaciones cardiovasculares, la enfermedad pulmonar obstructiva crónica (EPOC), miocarditis aguda, síndrome de distrés respiratorio agudo (SDRA) entre otras comorbilidades del coronavirus. Debido a que se debilita la acción antagonista del eje ACE2/Ang 1-7 cuyo papel es disminuir los niveles de Ang II mediante la activación de los receptores

[26] de angiotensina tipo 1 (AGT1R) para lograr una disminución de los efectos fisiopatológicos en los tejidos, como la inflamación y la fibrosis que dan lugar a las condiciones médicas nombradas. [4]

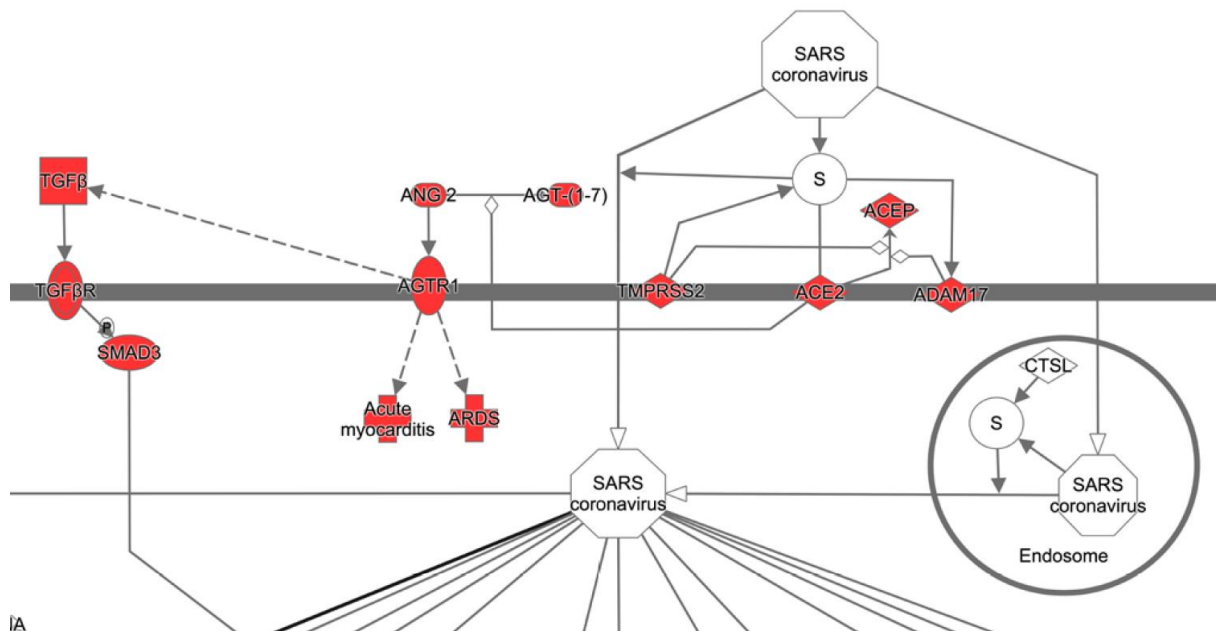


Figura 20. Se muestra la relación S/ACE2 con ADAM17 y TMPRSS2 para lograr la fusión con la membrana celular, la endocitosis del coronavirus y los efectos sobre otros sistemas celulares responsables de una respuesta inflamatoria AGTR1, TGFβR. El color rojo indica que la enzima está activada **Fuente: 20**

Esta situación conlleva a que los receptores AGTR1 sean el detonante de una respuesta inflamatoria sistémica puesto que se inicia con una tormenta de citocinas, la cual consiste en una liberación de grandes cantidades de citocinas proinflamatorias (IFN- α , IFN- γ , IL-1 β , IL-6, IL-12, IL-18, IL-33, TNF- α , TGF β , etc.) y quimiocinas (CCL2, CCL3, CCL5, CXCL8, CXCL9, CXCL10)^{23,24}. Se trata de un violento ataque del sistema inmune, causando daño alveolar difuso, insuficiencia orgánica múltiple y en el peor de los casos muerte. [25]

3.3.2 Respuesta inmune

El sistema inmune representa el conjunto de moléculas y células encargadas de diferenciar lo propio de lo extraño, lo ofensivo de lo inofensivo, y actúa de forma coordinada frente a los microorganismos infecciosos y los tumores. Los coronavirus en concreto están particularmente adaptados para evadir la detección inmune. Durante las distintas epidemias de SARS-CoV y MERS-CoV este hecho fue muy estudiado. La interferencia en el reconocimiento de interferón, el cross-talking intercelular y la señalización intracelular suponen los 3 modos fundamentales para lograr evadir la actividad inmunitaria. [27]

Los dos tipos principales de defensas del hospedador son respuestas inmunitarias inespecíficas (innatas) y específicas (adaptativas). La respuesta inmunitaria innata incluye a los interferones (α , β), células asesinas naturales, macrófagos (fagocitosis), α -defensinas, depuración mucociliar, enzima de edición del RNA de la apolipoproteína B (APOBEC3G, una enzima que combate el HIV) y fiebre, entre

muchos otros factores, en tanto que la respuesta inmunitaria adaptativa implica inmunidad humoral y mediada por células.[28]

- **Respuesta interferón antiviral normal [28]**

- **IFN tipo I:** el interferón tipo I (IFN-I) es un mecanismo de defensa temprano sofisticado posicionado para combatir infecciones microbianas y ayudar en el inicio de la respuesta inmunitaria adaptativa. Por selección evolutiva, múltiples especies de IFN-I desarrollaron esa señal a través del mismo receptor de IFN-I (IFNAR)
- **IFNAR:** el receptor IFNAR es heterodimérico y consta de dos subunidades, IFNAR-1 e IFNAR-2 (Figura 1). IFNAR-2 es la proteína de unión primaria y se une a moléculas de IFN-I con gran afinidad; Posteriormente, IFNAR-1 se une a la molécula con baja afinidad. El evento de unión desencadena la fosforilación de la porción citoplásmica de IFNAR-1 y la consiguiente señalización a través de la vía JAK/STAT a través de STAT1 y STAT2. La capacidad de transmitir de manera sensible los parámetros de unión extracelular a distintos resultados funcionales se logra a través de una combinación de efectos sobre la internalización del receptor, reguladores intracelulares negativos y positivos y circuitos de retroalimentación

- **Vía señalización JAK/STAT**

Una vez transcritos y traducidos como consecuencia de la infección por el virus, el interferón- β y el interferón- α inducen la expresión génica en las células vecinas mediante la unión a los receptores de interferón de la superficie celular y la posterior activación de la Janus quinasa (Jak)/transductor de señales y activador de la transcripción (STAT) vías de señalización. La unión del ligando de interferón da como resultado la activación de las tirosinas quinasas Jak asociadas al receptor y el posterior reclutamiento de la fosforilación de las proteínas STAT [29]. Los heterodímeros de STAT-1/2 interactúan con IRF-9 para formar el factor del gen estimulado por interferón (ISG) (ISGF)-3, que se une a los elementos de respuesta estimulados por interferón que se encuentran en numerosos genes antivirales inducidos por interferón [30]

- **STAT1/STAT2:** transductor de señales y activador de la transcripción que media las respuestas celulares a los interferones (IFN), la citoquina KITLG/SCF y otras citoquinas y otros factores de crecimiento. Después de la unión del IFN tipo I (IFN-alfa e IFN-beta) a los receptores de la superficie celular, la señalización a través de las proteínas quinasas conduce a la activación de las quinasas Jak (TYK2 y JAK1) y a la fosforilación de tirosina de STAT1 y STAT2. Los STAT fosforilados se dimerizan, se asocian con IRF9/ISGF3G para formar un complejo denominado factor de transcripción ISGF3, que ingresa al núcleo. ISGF3 se une al elemento de respuesta estimulada por IFN (ISRE) para activar la transcripción de genes estimulados por interferón, que impulsan a la célula a un estado antiviral [29,30]
- **IRF-9:** Factor de transcripción que juega un papel esencial en la inmunidad antiviral. Interviene en la señalización por IFN tipo I (IFN-alfa e IFN-beta). Después de la unión del IFN tipo I a los receptores de la superficie celular, se activan las quinasas Jak (TYK2

y JAK1), lo que lleva a la fosforilación de tirosina de STAT1 y STAT2. IRF9/ISGF3G se asocia con el dímero STAT1:STAT2 fosforilado para formar un complejo denominado factor de transcripción ISGF3, que ingresa al núcleo. ISGF3 se une al elemento de respuesta estimulado por IFN (ISRE) para activar la transcripción de genes estimulados por interferón, que conducen a la célula a un estado antiviral. [56] Defensa antiviral, Transcripción, Regulación de la transcripción

- **Supresión traslacional inducida por SARS-CoV Nsp1 y SARS-CoV-2 Nsp1:**

Normalmente, cuando un virus ataca a las células huésped, las helicasas de ARN huésped detectan el ARN viral en las células infectadas, lo que da como resultado la activación de los factores de transcripción. Estos factores inducen posteriormente la producción de citoquinas, incluidos los miembros de la familia IFN, que se unen a sus respectivos receptores y desencadenan una segunda ola de señalización de citoquinas. Pero nsp1 se une de forma estable a la subunidad ribosómica 40S e inactiva la función de traducción de la subunidad ribosómica 40S, lo que conduce a la inhibición de la síntesis de proteínas a la vez que evita la inducción de citoquinas.

3.3.3 Apoptosis

Inicio de la apoptosis

Las funciones mediante las cuales se puede activar el programa apoptótico están estrechamente reguladas a través de los mecanismos de activación e inhibición dentro de la célula, controlados a través de productos génicos, modificaciones postraduccionales. Morfológicamente, la apoptosis se caracteriza por contracción celular, condensación de cromatina, generación de ROS y formación de ampollas en la membrana.

La apoptosis tiene el propósito esencial de asegurar la homeostasis de tejidos y organismos a través de dos vías distintas: intrínseca y extrínseca. [31]

- La vía intrínseca se inicia a partir de señales internas, como el daño del ADN y el estrés oxidativo, y la vía extrínseca es inducida por una variedad de factores, como la infección viral. La apoptosis es un mecanismo de defensa poderoso y el más eficiente del huésped en las células infectadas por virus que detiene la progresión viral y evita que la infección se salga de control.
- La vía extrínseca de la apoptosis puede estar mediada por receptores de muerte, como TNFR y TRAIL en la superficie de la célula huésped y es aprovechada por los virus como mecanismo de diseminación. El programa apoptótico con respecto a los mecanismos de defensa inmunológica se desarrolla para eliminar las células que se han infectado por virus de participar en el programa de reproducción viral, mediante el cual orquesta la destrucción de la célula y su maquinaria constituyente.

La apoptosis, puede permitir que se propaguen algunos parciales virales, pero una apoptosis estrictamente controlada reduce el título viral general e inhibe la cantidad potencial de virus que se puede liberar para infectar otras células vecinas dentro de un organismo. La apoptosis representa un sistema ordenado de eventos que orquestan el destino final de una célula, de modo que una sola célula puede eliminarse sistemáticamente. [31]

Formación del inflamosoma

La inflamación suele ser la primera respuesta cuando las células huésped se encuentran con una infección viral que, en parte, induce la apoptosis mediada en gran medida por las citocinas proinflamatorias, como la interleucina 1 β (IL-1 β). La IL-1 β se sintetiza como la forma inactiva pro-IL-1 β y se activa mediante escisión proteolítica. La escisión de pro-IL-1 β en la forma activa está mediada por la caspasa 1. La caspasa 1, que también se sintetiza en la forma inactiva pro-caspasa 1, se activa a través del inflamosoma NLRP3 [34, 35]. El inflamosoma NLRP3 se forma en presencia de un patrón molecular asociado a patógenos (PAMP, por ejemplo, LPS). El reconocimiento, a través de la activación de receptores Toll, trae como consecuencia la activación del factor de transcripción NF- κ B, el cual estimula la producción de citocinas proinflamatorias. Adicionalmente, mediante un mecanismo complejo de señalización y transducción, la activación de los receptores RIG-I activa una familia de factores de transcripción que se denominan IRF los cuales pueden translocarse al núcleo. Una vez en el núcleo, estos IRF interactúan con genes que contienen en sus promotores secuencias nucleotídicas denominadas ISRE y que conducen a la expresión de los interferones de tipo I (INF- α e IFN- β) y otros genes antivirales importantes. [35]

Adicionalmente y para hacer la respuesta inflamatoria contra SARS-CoV-2 aún más compleja, la identificación de ACE-2 como el receptor de SARS-CoV-2 coloca al Sistema Renina Angiotensina Aldosterona (SRAA), como un elemento clave en la patogénesis de la COVID-19, debido a que esta proteasa está involucrada en el metabolismo de la angiotensina II, la cual se sabe tiene un papel preponderante en la respuesta inflamatoria. De hecho, la evidencia científica indica que el inflamosoma NLRp3 se activa en los pulmones, las células renales y los cardiomiocitos, en respuesta a la estimulación con angiotensina II vía unión a receptores AT1 [35]

NF- κ B activación

La expresión de citocinas proinflamatorias está impulsada por la vía de señalización del factor nuclear kappa B (NF- κ B). NF- κ B es una familia de factores de transcripción, NF- κ B regula muchos comportamientos celulares importantes como las respuestas inflamatorias, el crecimiento celular y la apoptosis. NF- κ B también contribuye a los cánceres y las funciones del sistema nervioso y mitocondrial. La vía de NF- κ B responde a diversos estímulos que incluyen ligandos de varios receptores de citocinas, receptores de reconocimiento de patrones (PRR), superfamilia de receptores de TNF (TNFR), así como receptores de células T y células B. A su vez, los virus pueden utilizar NF- κ B para sus propios beneficios. El mecanismo principal para la activación de NF- κ B es la degradación inducible de I κ B α desencadenada por un complejo de I κ B quinasa de múltiples subunidades (IKK). La IKK puede ser activada por citocinas, factores de crecimiento, mitógenos, componentes microbianos y agentes infecciosos. [35]

Para el SARS-CoV-1, tanto las proteínas estructurales como las proteínas accesorias activan la señalización de NF- κ B. La proteína nucleocápsida (N) del SARS activa NF- κ B e induce la expresión de IL-6 en las células A54921,22. La respuesta de NF- κ B por la proteína de membrana (M) es controvertida. En un estudio, la proteína M suprimió la actividad de NF- κ B en las células HeLa y Vero E6 al afectar la translocación nuclear de NF- κ B23. Por el contrario, M activó las cascadas de señalización de NF- κ B y promovió aún más la producción de interferón-beta (IFN- β) en las células HEK293T24. Para las proteínas accesorias, las proteínas ORF3a y ORF7a fueron capaces de activar la quinasa NF- κ B y c-Jun N-terminal (JNK) y mejoraron significativamente la expresión de IL-825. La proteína ORF3a también

indujo la transcripción de pro-IL-1 β a través de la activación de NF- κ B y promovió el inflamasoma 3 (NLRP3) de la familia del receptor similar a NOD (NLR) que contiene el dominio pirina 3 (NLRP3). Dada la similitud genética del SARS-CoV-1 con el SARS-CoV-2, sus proteínas virales pueden poseer estrategias conservadas para manipular la respuesta de las citocinas. [36]

CAPÍTULO 4. ANÁLISIS DE CONSISTENCIA

En este capítulo se realizó la instanciación de la vista funcional de la patogénesis del coronavirus por medio del mapeo utilizando como herramienta caracterizadora la vista estructural perteneciente al modelo CSGH, ambas vistas han sido descritas en el capítulo 3 de Materiales y métodos. Dicha instanciación se compuso partiendo de la interacción virus/huésped descrita en la vista funcional por lo que para facilitar la modelización y comprensión de las dinámicas se ha decidido centrarse solo en las respuestas virales. Una vez mapeada la vista funcional, se detectaron y analizaron las limitaciones de expresión del modelo, para posteriormente proponer posibles mejoras para su subsanación y nuevas características añadidas.

4.1 MAPEO

Para realizar el mapeo de la vista funcional generada en el capítulo 3, se ha realizado un diagrama de la vista estructural (Figura 21) que se ha usado como base para instanciar cada uno de los elementos que componen los mecanismos de patogenia relativos al SARS-COV, es decir las proteínas víricas, puesto que son las que originan la respuesta de procesos celulares en el huésped, por tanto resulta útil de cara al uso del modelo en otras infecciones víricas como quedará la especificación de los procesos de patogénesis relativos a un virus específico de la fuente [20]

Previamente al mapeo se ha realizado un listado de las proteínas víricas incluyendo su estructura proteica así como los procesos asociados a cada una durante la patogénesis (Tabla 3).

Tabla 3. Resumen de los procesos víricos asociados a cada proteína viral del SARS-CoV, explicando las alteraciones que produce en la célula. Fuente: [20]

Nombre entidad	Alteraciones inducidas
Proteínas no estructurales	
NSP1	<ul style="list-style-type: none">– Degradación RNA mensajero– Inhibición del transporte nucleocitoplasmático de proteínas

	<ul style="list-style-type: none"> - Inhibición de las tres vías de señalización dependientes del virus que conducen a la activación de los factores de transcripción ATF2/c-Jun, IRF3/IRF7 y NF-kB
NSP3	<ul style="list-style-type: none"> - Inhibición del complejo formado por STING1, TBK1 y TRAF3 - Activación de la señalización NFkB
Proteína accesorias	
SARS 3A	<ul style="list-style-type: none"> - Unión al virus - Activación de la señalización de inflamosoma NALP3 - Activación de la ruta EIF2AK3 - Activación ruta para la señalización de apoptosis p38 MAPK - Activación de la señalización NFkB - Activación de la ruta JNK ruta de la apoptosis
SARS 3B	<ul style="list-style-type: none"> - Activación de la ruta JNK ruta de señalización de la apoptosis - Activación ERK 1/2
SARS 6	<ul style="list-style-type: none"> - Inhibición KPNB
SARS 7A	<ul style="list-style-type: none"> - Inhibición BCL2 ruta señalización de apoptosis - Activación de la señalización NFkB que causará más adelante hipercitokinemia que desencadenará una inflamación pulmonar - Activación ruta para la señalización de apoptosis p38 MAPK - Activación de la ruta JNK ruta de la apoptosis
SARS 8A	<ul style="list-style-type: none"> - Disfunción mitocondrial <p>Debido a la inducción de la inhibición de la respiración celular y fosforilación oxidativa que causa la activación del CASP3 que genera la apoptosis</p>
SARS 9B	<ul style="list-style-type: none"> - La activación del CASP3 que genera la apoptosis.
Proteínas estructurales	
E	<ul style="list-style-type: none"> - Activación de la señalización de inflamosoma NALP3. - Inhibición BCL2 ruta señalización de apoptosis

M	<ul style="list-style-type: none"> - Activación CASP8 y CASP9 que señalizan la apoptosis (finalización del proceso de respuesta inmune y transmisión viral)
N	<ul style="list-style-type: none"> - Inhibición del TRIM25 - Inhibición del EF1a induce la inhibición del transporte nucleoplasmático
S	<ul style="list-style-type: none"> - Entrada del virus a la célula - Unión a la proteína TMPRSS2 - Unión a la proteína ACE2 - Unión a la proteína ADAM17 - Unión a CSTL en el endosoma - Activación de SMAD4 y SMAD3 - Inhibición del complejo CCND1 y CDK4 - Inhibición de NPM1 (duplicación del centrosoma, causando la inhibición de la progresión del ciclo celular) - Inhibición de la CDK2- CyclinE - Activación de la COX2 que causa inflamación pulmonar y lleva a daños del tejido pulmonar

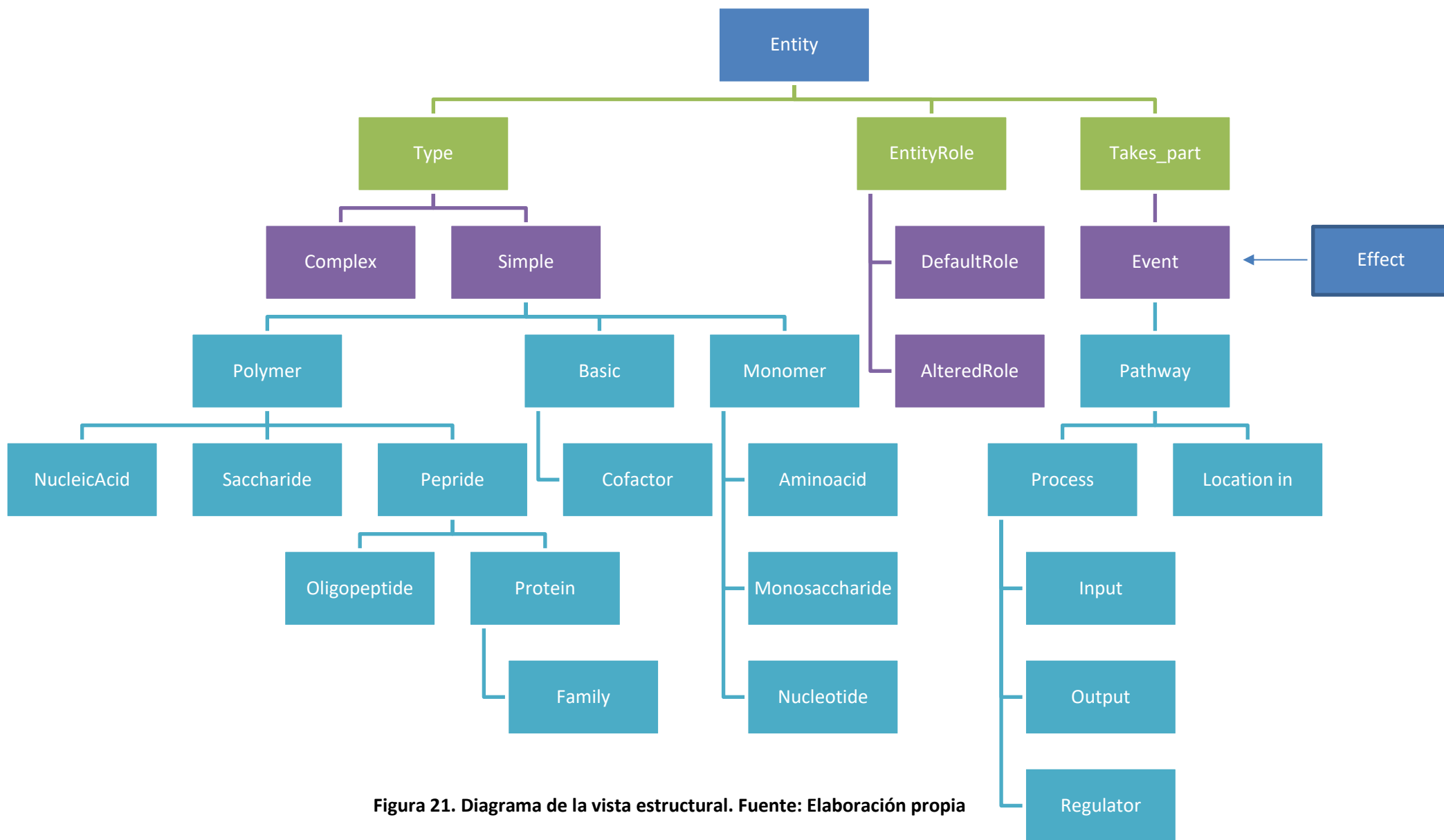
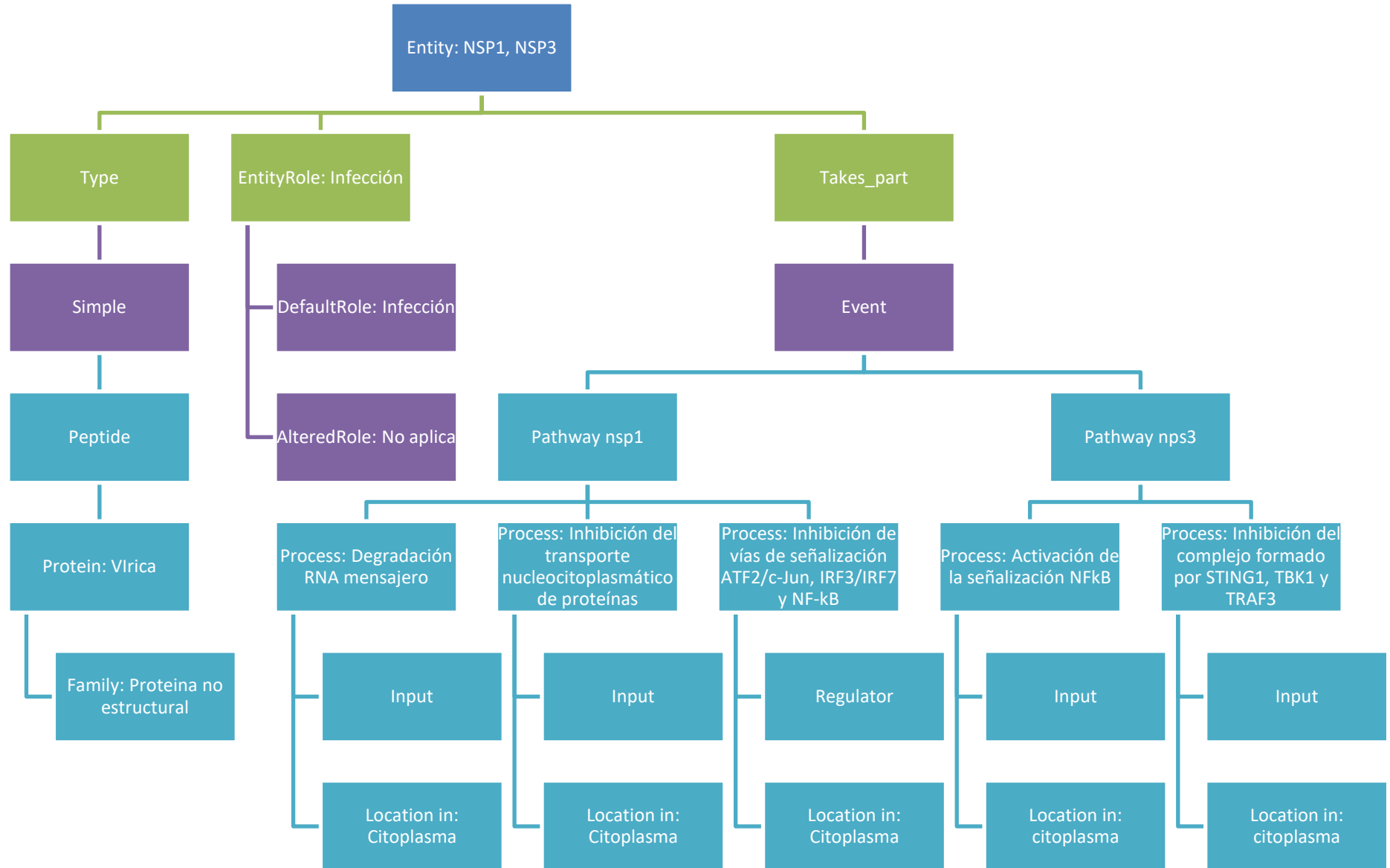
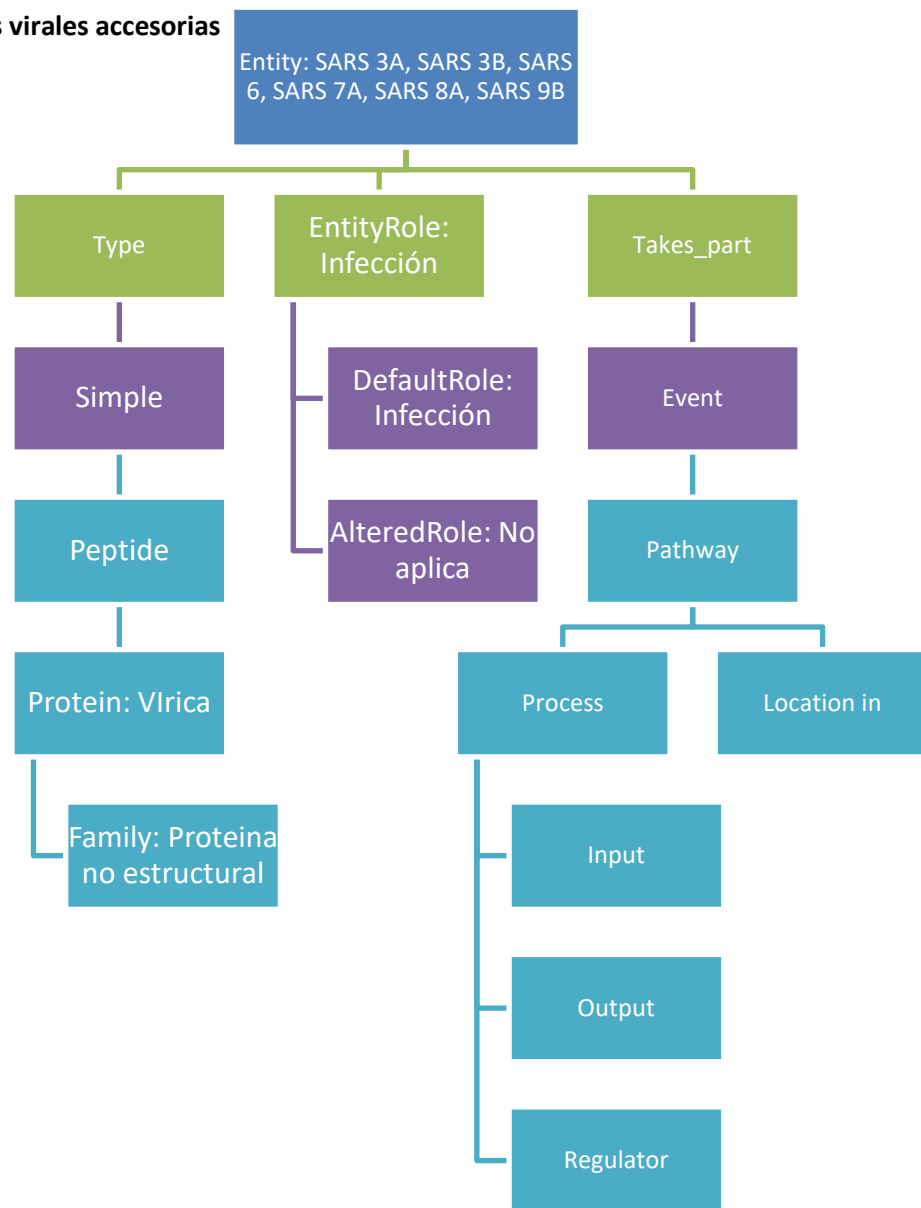


Figura 21. Diagrama de la vista estructural. Fuente: Elaboración propia

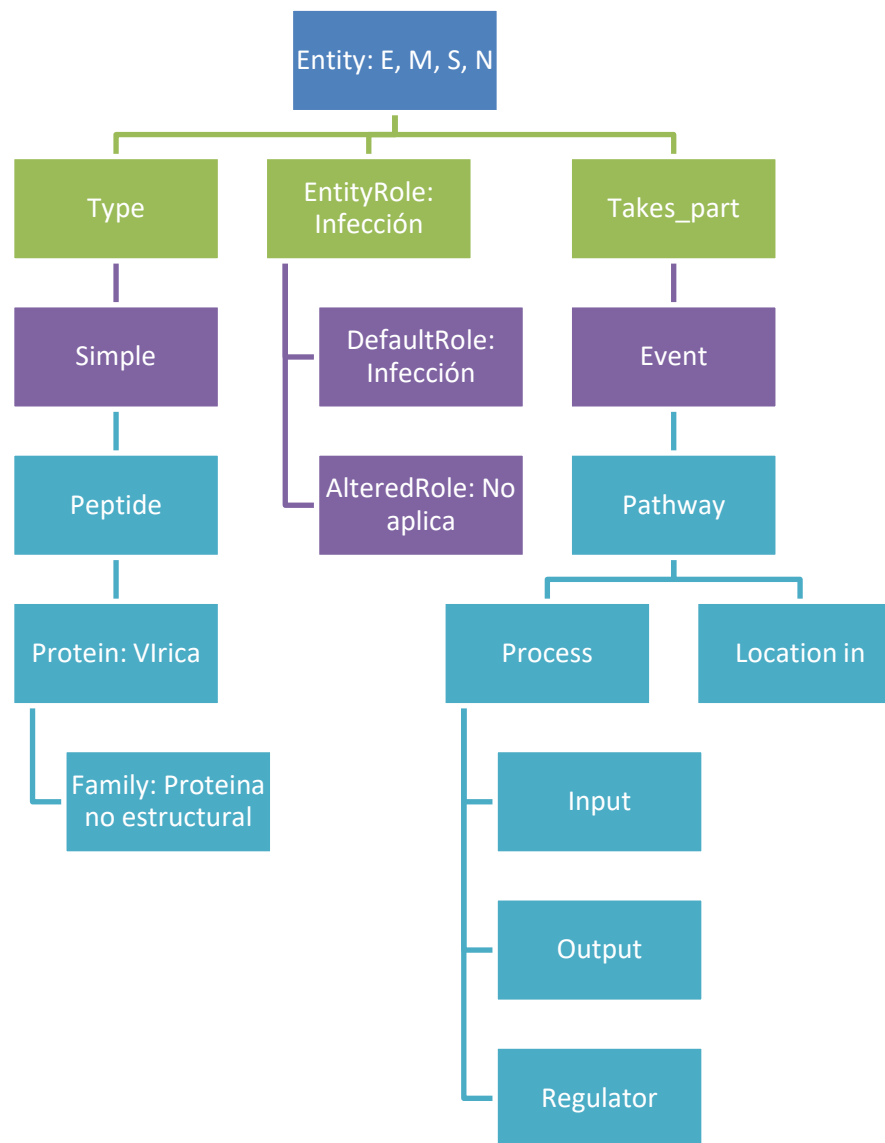
4.1.1 Instanciación proteínas virales no estructurales



4.1.2 Instanciación de las proteínas virales accesorias



4.1.3 Instanciación de las proteínas virales estructurales



4.2 LIMITACIONES DEL MODELO CSGH V3

Una vez realizado el mapeo del modelo aplicado a la vista funcional, fue posible apreciar las siguiente lista de limitaciones o imposibilidad para expresar ciertos mecanismos de patogénesis y en la por medio de la caracterización. Han sido recopilados en la Tabla 6.

Tabla 4. Lista de limitaciones detectadas en el modelo CSGH v3 para expresar la patogénesis del cors vírica. Fuente: Elaboración propia

Limitaciones detectadas	
1	No es capaz de explicar o modelizar el comportamiento de entidades que no formen parte de la propia célula.
2	No es posible distinguir las causas y consecuencias que llevan a los proceso a la respuesta infecciosa en el huésped.
3	No es posible distinguir entre eventos alterados por la influencia del patógeno y no alterados.
4	No es posible establecer relaciones entre los eventos más allá de la posición <i>Post</i> y <i>Pre</i> , como la interdependencia o paralelismo entre eventos (o procesos)
5	La clase <i>EntityRole</i> no es capaz de explicar el origen de la mutación con el proceso y elemento causante.
6	La clase <i>AlteredRole</i> no puede ser aplicada a entidades virales (o cualquier patógeno)
7	La clase <i>Location</i> solo hace referencia localizaciones en la misma célula pero no hace referencia a localizaciones que se den en el entorno intracelular de otros tipo celulares
8	No es posible modelar cómo ciertos eventos como los pathways indican un fase del la evolución de la infección.
9	No es posible conectar el fenotipo de la enfermedad infecciosa con la actuación de ciertos eventos como la neumonía
10	No es posible determinar como propiedades físicas extracelulares ambientales como el pH y la temperatura (entidades que no se corresponde con ninguna de las clase especializadas de <i>Entity</i>) afectan a los eventos.
11	Falta caracterización de los procesos relacionados con transformación postraduccionales de proteínas, así como de la forma nativa de las proteínas.
12	Al modelar un elemento de la clase <i>Effect</i> no es posible expresar sobre que entidad del evento se ha ejercido influencia para que este se desencadene.

1. No es capaz de explicar o modelizar el comportamiento de entidades que no formen parte de la propia célula.

La primera limitación con la que se encuentra el modelo a la hora de representar mecanismos de patogénesis del coronavirus, es que el modelo original de CSHG fue pensado para explicar y modelizar entidades del propio organismo y como mutaciones genéticas podían dar lugar a enfermedades. Por tanto, su objetivo como herramienta es rastrear y comparar secuencias de ADN, ARN o de aminoácidos para encontrar tanto el elemento original como el que surge fruto de una mutación, caracterizar ambos así como el tipo de mutación que ha tenido lugar para relacionarlo con un fenotipo y gravedad así como con bibliografía clínica existente y bases de datos que contengan los elementos de estudio de interés.

No obstante, si únicamente se pone atención sobre la caracterización de la vista proteoma, si que es posible poder expresar las proteínas pese a que se ha necesario mejorar en la profundización de las clases y relaciones que constituyen esta vista. De hecho tal y como se ha visto tanto las proteínas huésped como las víricas han podido modelizarse, aunque presentan la dificultad de cómo puede indicarse en el modelo que proteínas son de origen vírico correspondiente a una patología concreta, y cuales son del huésped.

Pero al observar la vista pathways, debemos tener en cuenta que la patogénesis de un virus, en este caso el coronavirus se basa en la interacción entre el huésped y este mismo por lo que en la clase *takes_part* a un más importante que saber si la entidad es *input*, *output* o *regulator* es saber, si el elemento del invasor o del huésped. Por ejemplo, en el pathway de unión y entrada viral en la célula huésped, ocurría una unión específica entre ACE2 y la glicoproteína S, esta especificidad que se da entre receptores virales y huésped explica la propagación y transmisión del virus.

2. No es posible distinguir las causas y consecuencias que llevan a los procesos de la respuesta en el huésped ante la infección

Tras cada interacción del virus con el huésped se produce un proceso que lleva como consecuencia daños y alteraciones tanto en las biomoléculas como en la célula, para ser capaz de expresar y modelar la virulencia del virus. La virulencia se designa como el carácter patogénico, nocivo y violento de un virus par poder expresar esta propiedad hace pues falta modelar estas consecuencias a las acción de un virus que se puede ver en la respuesta que provoca en la célula huésped reaccionando a los daños fruto de la infección retroalimentándose y dañando más su equilibrio celular.

3. No es posible distinguir entre eventos alterados por la influencia del patógeno y no alterados.

El efecto de un proceso viral o una proteína viral afecta a la ejecución de otro proceso, a veces inhibiéndolo y otras alterándolo. Un ejemplo es como la proteína SARS7 altera la actividad celular y provoca la activación de la señalización NFkB que causará más adelante una tormenta que desencadenará una inflamación pulmonar. Este evento es una alteración del camino de señalización NFkB puesto que en vez ser capaz de regular el proceso inflamatorio se descontrola y permite que aumente. En el modelo actual estos procesos aparecerían como procesos independientes y en vez de explicitar la relación entre alteración y estándar.

4. No es posible establecer relaciones entre los eventos más allá de la posición Post y Pre, como la interdependencia o paralelismo entre eventos (o procesos).

Eventos como la adhesión y endocitosis viral causan de forma paralela el inicio de la respuesta sistémica inflamatoria mediante una tormenta de citocinas, por lo que ambos sucesos son interdependientes pero también semi-simultáneos, ya que no se siguen sucesivamente uno a otro, ya que cuando aún no finalizado la endocitosis viral ya se ha activado mediante elementos como ADAM17 y ACE2 que desencadenen la respuesta inflamatoria. Tal y como se ha diseñado el modelo actual contar únicamente con la relación pre y post es ineficiente para expresar el comportamiento patogénico.

5. La clase EntityRole no es capaz de explicar el origen de la mutación con el proceso y elemento causante.

No obstante, no ocurre igual cuando se analiza la vista pathways, ya que resulta ineficiente para poder expresar las interacciones virus/huésped especialmente cuando se trata del virus. Esto se debe, a que cómo se ha dicho con anterioridad el modelo CSHG está pensado para explicar el funcionamiento interno de las células, aspecto que se ve principalmente al analizar la clase *EntityRole*.

La clase *EntityRole* modeliza el rol o función que cumple una célula en un proceso particular, si una proteína como la STA1 cuya labor, durante el proceso de entrada viral a la célula, es proporcionar instrucciones para producir proteínas que participan en el sistema inmunológico, dejase de hacerlo debido a alteraciones inducidas o provocada por la interacción de la células con las proteínas virales del coronavirus o actuará de forma que provoque el deterioro celular. Es posible expresar claramente cuál es el rol normal (*defaultRole*) y el rol alterado (*alteredRole*) pero no podríamos explicar como se ha dado esta alteración, puesto que esta alteración es el output de un proceso alterado por la presencia e interacción con un virus.

6. La clase AlteredRole no puede ser aplicada a entidades virales (o cualquier patógeno)

Si en vez de ser una proteína del huésped es viral la modelización en cuanto cuál es su rol normal y cuál el alterado carece de sentido, ya que no ocurre ninguna transformación o alteración en la estructura vírica siempre actuará para lograr la replicación viral y daño celular.

7. La clase Location solo hace referencia localizaciones en la misma célula pero no hace referencia a localizaciones que se den en el entorno intracelular de otros tipo celulares.

En la patogénesis se invaden distintos tipos celulares, en el caso del coronavirus estos son células respiratorias, tejido miocárdico, tejido renal y células del sistema inmunitario, cada uno con funciones y mecanismos propios que no se haya en todas las células. Por lo tanto habrá procesos que no solo dependan de en que punto del espacio intracelular (núcleo o citoplasma) sino también en qué célula se lleva a cabo.

El ejemplo más claro es como los procesos relativos a la respuesta inmune como la activación del interferón tipo 1 solo sucede en células del sistema inmunológicos como las dendríticas. Pero no solo sirve para esta diferenciación entre los tipos celulares diana de un virus como el coronavirus, si no que se pueden expresar procesos comunes en todas las células diana para lograr identificar marcadores biológicos de la enfermedad. Un ejemplo sería la unión y entrada a la célula proceso común a todos

los tipos celulares diana donde el elemento clave que actúa en todas y por tanto corresponde a un marcador biológico de infección del coronavirus es el receptor ACE2.

Pero el modelo no cuenta con dichas referencias por lo que no se puede explicar

8. No es posible modelar cómo ciertos eventos como los pathways indican un fase de la evolución de la infección y gravedad.

Una enfermedad está formada por varias fases de la enfermedad como la infección del tejido diana, la respuesta autónoma celular, el periodo de incubación, respuesta local inmune, impacto en el organismo como el RAS que lleva a la inflamación y restauración de la homeostasis.

Cada una de estas fases tiene una gravedad y fisiopatología asociadas por lo que al expresar un pathway este no solo indica los procesos y elementos que lo forman sino que también en momento de evolución de la enfermedad está y por tanto la gravedad de los daños producidos por la patogénesis.

Un ejemplo es como el pathway Inflammosoma da origen a la inflamación pulmonar que causa daños en el tejido de los pulmones durante la respuesta local inmune.

No obstante, en el modelo no cuenta con una o varias clases que modelen esta relación.

9. No es posible conectar el fenotipo de la enfermedad infecciosa como el SRAS con la actuación de ciertos eventos.

Como resultado de los eventos virales pueden surgir daños celulares o en los tejidos, llegando a causar sintomatologías como la neumonía o el SRAS en el caso del coronavirus. Conocer en que punto del pathway suceden estos daños que conforman el fenotipo de la enfermedad permite una identificación de marcadores biológicos de tales sucesos. No obstante, en el modelo no se puede expresar esta relación pues la única clase que hace referencia al resultado de un proceso es *Output* que está referido a la *Entity* que no cuenta en sus especializaciones de fenotipos centrándose en biomoléculas

10. No es posible determinar como propiedades físicas extracelulares ambientales como el pH y la temperatura (entidades que no se corresponde con ninguna de las clases especializadas de Entity) afectan a los eventos.

Aspectos como el pH y la temperatura son claves en la función de las proteínas, especialmente de las enzimas. Durante una infección estos son dos parámetros que presentan oscilaciones durante toda la infección, un ejemplo sería la fiebre. Por tanto, las reacciones que suceden en los procesos debe tener en cuenta a que temperatura se están realizando puesto que se puede ver alterada su velocidad de reacción. En el modelo de CSHG v3 no es posible expresar esta relación.

11. Falta caracterización de los procesos relacionados con transformación postraduccionales de proteínas, así como de la forma nativa de las proteínas.

Si nos fijamos en el desarrollo del modelo de patogénesis podemos ver como a cada uno de los procesos víricos les acompañan subprocesos enzimáticos que causan la activación o inhibición de distintas vías celulares por medio de la alteración de las proteínas del huésped, para llevar a cabo todas estas actividades se utilizaban procesos de fosforilación, hidrólisis, glicosilación, etc. Todos estos procesos tienen en común que son transformación postraduccionales (MTP) que se dan sobre la entidad proteína en su estructura primaria que se refleja en su conformación espacial alterando su

función. Por lo que de cara a poder caracterizar como se traduce esos cambios en la disposición espacial de la proteína, es decir, en la estructura terciaria, es interesante poder modelarlo para poder ver la relación entre la forma nativa de una proteína, el MTP correspondiente y la estructura espacial resultante.

12. Al modelar un elemento de la clase Effect no es posible expresar sobre que entidad del evento se ha ejercido influencia para que este se desencadene.

Un efecto puede ser un proceso o elemento producto de evento anterior que ejerce una influencia sobre esto desencadenándolo, o también pueden ser una serie de condiciones que favorecen y dan lugar a un evento, desde la visión de patogénesis, ese efecto puede modelar como un proceso o elemento viral. El problema radica en dicho efecto actúa sobre un elemento o elemento concreto y serán estos los que generen el evento, por lo tanto manteniendo la relación actual *cause* solo entre la clase *Effect* y *Event* no es posible determinar esa correspondencia. Esto es de gran interés clínico puesto que podrían servir como marcadores biológicos de un de los eventos infecciones.

CAPÍTULO 5. CONCLUSIONES Y APORTACIONES

Después de la realización de este proyecto, en especial tras el análisis de constancia que conon el modelo actual no es posible distinguir entre mecanismos fisiológicos y los patológicos causados por el virus, dividiendo entre aquellos que son propios de la célula ante la presencia de un virus y aquellos que altera el virus para lograr la replicación y virulencia. Así pues se trata de poder caracterizar los efectos del virus en los procesos de una manera clara y concisa de forma que ante distintos organismos ajenos al organismos podamos parametrizar sus efectos concretos de las misma forma que podemos verlo con las mutaciones genéticas.

Por lo tanto tiene que ser optimizado para enfermedades infecciosas de forma que así como con las enfermedades de origen genético, el modelo es capaz de relacionarlos a un fenotipo, población y prevalencia, se ha de ser capaz de identificar las enfermedades infecciosas en este caso de origen vírico que deben incluir los siguientes parámetros recogidas en el siguiente esquema (Figura 22) [37]

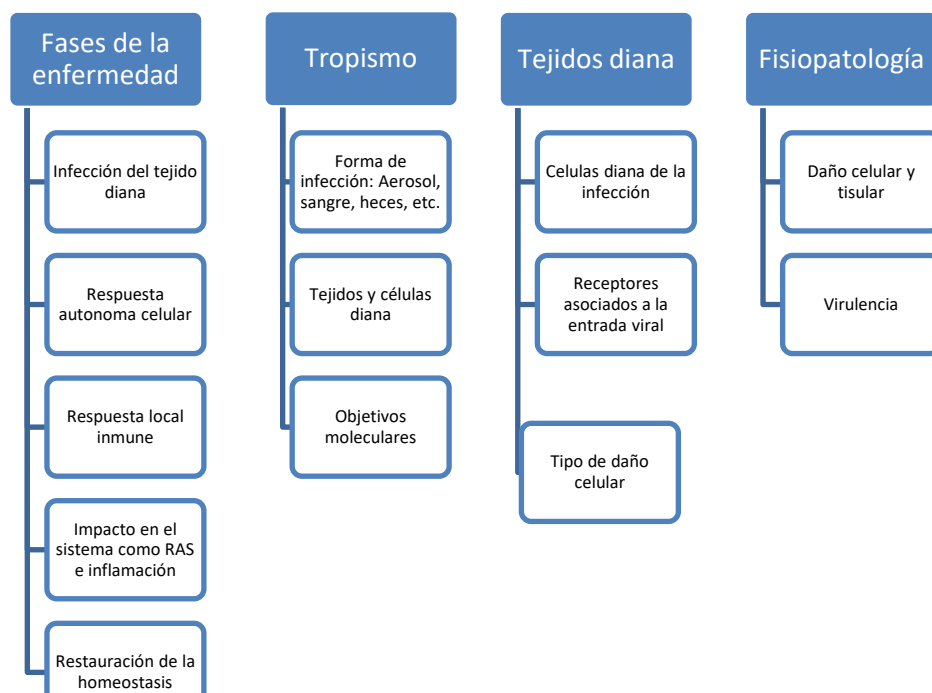


Figura 22 Basado en [37]

Añadir otras categorías como la fisiopatología, las manifestaciones clínicas relacionados con la severidad de la enfermedad, tal y como se ha explicado en el apartado limitaciones del capítulo 4

Así como mejorar la caracterización de los procesos fisiopatológicos por medio de la modificación postraduccional de proteínas. Algunas de las MTP mencionadas en la vista funcional se han incluido en la Tabla 5

Tabla 5. Listado de modificaciones postraducionales más conocidas divididas según su actividad sobre la proteína diana. Fuente [38][39]

Adición de grupos funcionales	
Acilación	Adición de un grupo acilo.
Fosforilación	Adición de un grupo fosfato.
Metilación.	Adición de un grupo metilo.
Hidroxilación.	Adición de un grupo hidroxilo.
Glicosilación.	Adición de un glúcido.
Glicación.	Modificación no enzimática de los grupos amino de las proteínas por la acción de azúcares reductores.
Sulfonilación	Adición covalente normalmente irreversible de tres átomos de oxígeno al grupo tiol de un residuo de cisteína , que da como resultado la formación de un residuo de ácido cisteico
Prenilación.	Adición de moléculas hidrofóbicas.
Nitrosilación.	Adición de un grupo nitroxilo.
Nitración.	Adición de un grupo nitro.
Acilación.	Adición de un grupo acilo.
Enlace con otras proteínas	
Sumoilación.	Adición de otra pequeña proteína llamada SUMO. SUMO es también conocida por otros nombres, como por ejemplo sentrina, Smt-3 y GPM-1
Ubiquitinación.	Adición de la proteína ubiquitina a proteínas diana.
Rotura de enlaces covalentes	
Proteolisis	Escisión de un fragmento proteico y degradación de la proteína

En base a estas aportaciones más al análisis de los objetivos que nos habíamos planteado es posible afirmar que se han cumplido todos los objetivos del trabajo:

- Caracterización de las vistas proteína y pathway para generar una herramienta de modelización denominada vista estructural
- Compresión y descripción de los mecanismos de infección del SARS COV
- Caracterización de los elementos presentes en la patogénesis del virus
- Identificación y caracterización elementos para ampliar el esquema conceptual del genoma humano CSHG en trabajo futuro, si procediese.

CAPÍTULO 6. BIBLIOGRAFÍA

1. Modelo CSHG v3
2. Lsi.us.es, 2022. [Online]. Available: <http://www.lsi.us.es/docencia/get.php?id=6889>. [Accessed: 09- Mar- 2022].
3. Uschold, M., Gruninger, M.: Ontologies: principles, methods and applications. The knowledge engineering review 11, 2 (1996)
4. A. Garcia S., A. Palacio, J. Reyes Roman, J. Casamayor and O. Pastor, "Towards the Understanding of the Human Genome: A Holistic Conceptual Modeling Approach", IEEE Access, vol. 8, pp. 197111-197123, 2020. Available: 10.1109/access.2020.3034793 [Accessed 9 March 2022].
5. Ó. Pastor, A. León, J. Reyes, A. García and J. Casamayor, "Using conceptual modeling to improve genome data management", Briefings in Bioinformatics, vol. 22, no. 1, pp. 45-54, 2020. Available: 10.1093/bib/bbaa100 [Accessed 9 March 2022].
6. Ó. López, A. Palacio, J. Román and J. Casamayor, "Modeling Life: A Conceptual Schema-centric Approach to Understand the Genome", Conceptual Modeling Perspectives, pp. 25-40, 2017. Available: 10.1007/978-3-319-67271-7_3 [Accessed 9 March 2022].
7. Ocw.unizar.es, 2022. [Online]. Available: https://ocw.unizar.es/ciencias-experimentales/modelos-matematicos-en-bases-de-datos/uml/03UML_DiagramaClases.pdf. [Accessed: 15- Sep- 2022].
8. Fung, S. and Liu, D.X. (2019) Human coronavirus: host-pathogen interaction. Annu. Rev. Microbiol. 73, 529.
9. Lim, Y.X., et al. (2016) Human coronaviruses: A review of virus-host interactions. Diseases 4, 26.
10. P. Murray, M. Pfaller, K. Rosenthal and P. Di Francesco, Microbiología medica. Milano: Edra, 2021, p. Chapter 7.
11. Nieto-Torres, J.L., et al. (2015) Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome. Virology 485, 330.
12. G. Pastian-Soto, "Bases Genéticas y Moleculares del COVID-19 (SARS-CoV-2). Mecanismos de Patogénesis y de Respuesta Inmune", 2022. "Neumonía - Síntomas y causas - Mayo Clinic", MayoClinic.org, 2022. [Online]. Available: <https://www.mayoclinic.org/es-es/diseases-conditions/pneumonia/symptoms-causes/syc-20354204>. [Accessed: 15- Sep- 2022].
13. "Diccionario de cáncer del NCI", Instituto Nacional del Cáncer, 2022. [Online]. Available: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/proteoma>. [Accessed: 15- Sep- 2022]
14. Uv.es, 2022. [Online]. Available: https://www.uv.es/tunon/pdf_doc/proteinas_09.pdf. [Accessed: 15- Sep- 2022].

15. M. Inmune, "Las citoquinas y su función en la respuesta inmunológica", MiSistemaInmune, 2022. [Online]. Available: <https://www.misistemainmune.es/inmunologia/componentes/las-citoquinas-y-su-funcion-en-la-respuesta-inmunologica>. [Accessed: 15- Sep- 2022].
16. Zhang and J. An, "Cytokines, Inflammation, and Pain", *International Anesthesiology Clinics*, vol. 45, no. 2, pp. 27-37, 2007. Available: 10.1097/aia.0b013e318034194e [Accessed 15 September 2022].
17. F. Gálvez-Gastélum, A. Sandoval-Rodríguez and J. Armendáriz-Borunda, "El factor de crecimiento transformante beta como blanco terapéutico", *Salud Pública de México*, vol. 46, no. 4, pp. 341-350, 2004. Available: 10.1590/s0036-36342004000400009.
18. F. Morales Olivas and L. Estañ Yago, "Conceptos nuevos sobre el sistema renina angiotensina", *Hipertensión y Riesgo Vascular*, vol. 27, no. 5, pp. 211-217, 2010. Available: 10.1016/j.hipert.2009.09.002 [Accessed 15 September 2022].
19. "Capítulo I. Las enzimas", *Addi.ehu.es*, 2022. [Online]. Available: <https://addi.ehu.es/bitstream/handle/10810/14292/4-%20Cap%C3%ADtulo%20I.%20Las%20enzimas.pdf?sequence=4&isAllowed=y>. [Accessed: 15- Sep- 2022].
20. I. Pathway and A. Ipa, "Coronavirus Pathogenesis Pathway Sample to Insight."
21. L. Maxwell, Karen; F., Lori (2007). *Viral proteomics*. *Microbiol Mol Biol Rev*. pp. 398-411.
22. Viswanathan, Kasinath; Früh, Klaus (de diciembre de 2007). «Viral proteomics: global evaluation of viruses and their interaction with the host». *Expert Review of Proteomics* 4 (6): 815-829. doi:10.1586/14789450.4.6.815.
23. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J et al. (2020) Characterization of spike glycoprotein of SARS-CoV-1 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* 11, 1620.
24. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220 (2020). Together with Shang, J. et al., this study provides one of two early crystal structures of the SARS-CoV-2 RBD–ACE2 complex, revealing how the S protein recognizes its receptor.
25. Mercer J, Schelhaas M & Helenius A (2010) Virus entry by endocytosis. *Annu Rev Biochem* 79, 803–833.
26. Li, W. et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450–454 (2003)
27. C. Ng, J. Mendoza, K. Garcia and M. Oldstone, "Alpha and Beta Type 1 Interferon Signaling: Passage for Diverse Biologic Outcomes", *Cell*, vol. 164, no. 3, pp. 349-352, 2016. Available: 10.1016/j.cell.2015.12.027 [Accessed 9 March 2022].
28. N. Grandvaux, B. tenOever, M. Servant and J. Hiscott, "The interferon antiviral response: from viral invasion to evasion", *Current Opinion in Infectious Diseases*, vol. 15, no. 3, pp. 259-267, 2002. Available: 10.1097/00001432-200206000-00008 [Accessed 9 March 2022].
29. "STAT1 protein expression summary - The Human Protein Atlas", *Proteinatlas.org*, 2022. [Online]. Available: <https://www.proteinatlas.org/ENSG00000115415-STAT1>. [Accessed: 09- Mar- 2022].
30. "STAT2 protein expression summary - The Human Protein Atlas", *Proteinatlas.org*, 2022. [Online]. Available: <https://www.proteinatlas.org/ENSG00000170581-STAT2>. [Accessed: 09- Mar- 2022].

31. Azkur, A. K., Akdis, M., Azkur, D., Sokolowska, M., van de Veen, W., Brügger, M. C., ... & Akdis, C. A. (2020). Immune response to SARS-CoV-2 and mechanisms of immunopathological changes in COVID-19. *Allergy*, 75(7), 1564-1581.
32. Chaparro-M, N. A., & Franco-L, A. O. (2020). Clinical and immunological aspects of SARS-CoV-2 infection. *Revista de la Universidad Industrial de Santander. Salud*, 52(3), 295-309.
33. Johnson, D. T.; Harris, R. A.; French, S.; Blair, P. V.; You, J.; Bemis, K. G.; Wang, M.; Balaban, R. S. (2006). «Tissue heterogeneity of the mammalian mitochondrial proteome». *American Journal of Physiology. Cell Physiology* 292: c689-c697. PMID 16928776. doi:10.1152/ajpcell.00108.2006.
34. Uetz, Peter; Rajagopala, Seesandra V.; Dong, Yu-An; Haas, Jürgen (15 de octubre de 2004). «From ORFeomes to Protein Interaction Maps in Viruses». *Genome Research* 14 (10b): 2029-2033. doi:10.1101/gr.2583304.
35. Fang, X., Gao, J., Zheng, H., Li, B., Kong, L., Zhang, Y., ... & Ye, L. (2007). The membrane protein of SARS-CoV suppresses NF-κB activation. *Journal of medical virology*, 79(10), 1431-1439.
36. LIAO, Q. J., YE, L. B., Timani, K. A., ZENG, Y. C., SHE, Y. L., Ye, L., & WU, Z. H. (2005). Activation of NF-κB by the full-length nucleocapsid protein of the SARS coronavirus. *Acta biochimica et biophysica Sinica*, 37(9), 607-612.
37. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7885626/>. [Accessed: 15-Sep- 2022].
38. M. Mann and O. Jensen, "Proteomic analysis of post-translational modifications", *Nature Biotechnology*, vol. 21, no. 3, pp. 255-261, 2003. Available: 10.1038/nbt0303-255 [Accessed 15 September 2022].
39. Y. Deribe, T. Pawson and I. Dikic, "Post-translational modifications in signal integration", *Nature Structural & Molecular Biology*, vol. 17, no. 6, pp. 666-672, 2010. Available: 10.1038/nsmb.1842 [Accessed 15 September 2022].

DOCUMENTO 2: PRESUPUESTO

PRESUPUESTO

El presupuesto tiene como objetivo principal valorar el coste monetario de la tesis de máster desarrollado en esta memoria. El presupuesto tiene en cuenta: el personal, el software y hardware. La amortización de los costes del software y hardware se calculan utilizando la siguiente formula.

$$\text{Coste imputable(sin IVA)} = t \cdot \frac{C}{T}$$

Donde

- t es el tiempo de uso del equipo (en meses)
- C es el coste del equipo o licencia (en euros)
- T es el tiempo de amortización (en meses)

1. COSTE PERSONAL

En el coste del personal se tienen en cuenta el coste del ingeniero que realiza el trabajo y las horas que han invertido los tutores (miembros del PROS) a la asesoría del trabajo teniendo en cuenta las horas invertidas.

Tabl a6. Coste de personal desglosado por individuos implicados en el trabajo. Fuente: Elaboración propia

	Cantidad(horas)	Coste unitario (€/h)	Coste total (€)
Tutor (Ingeniero informático)	30	25,00	750,00
Tutor (Ingeniero informático)	30	25,00	750,00
Ingeniera biomédica	500	15,00	7.500,0
			Total: 9.000,00

2. COSTE DE EJECUCIÓN

2.1 Coste del software

En esta parte del proyecto se detallan los costes respecto de las licencias y las bases de datos. Se incluyen el coste del sistema operativo, del paquete de Microsoft Office y de las bases de datos utilizadas.

Tabla 7. Coste del software desglosado. Fuente: Elaboración propia

	Coste sin IVA (€)	Duración licencia (meses)	Uso (meses)	Coste imputable sin IVA (€)
Windows 10	145,00	Ilimitada	10	0,00
Microsoft Office 365	69,00	12	10	57,50
Acceso a recursos bibliográficos	0,00	Ilimitada proporcionada por la UPV	10	0,00
				Total: 57,50

2.2 Coste del hardware

Tabla 8. Coste del hardware desglosado. Fuente: Elaboración propia

	Coste sin IVA (€)	Vida útil (meses)	Uso (meses)	Coste imputable sin IVA (€)
Portatil ASUS GL552VW	709,92	84	10	84,51
Disco duro 1TB 7200rpm SATA	88	28	10	31,43
				Total: 115,94

El coste de ejecución total es la suma de los costes del hardware y del software, lo cual resulta en 173,44€.

3. COSTE TOTAL

El coste total del proyecto es la suma del coste de personal y del coste de ejecución, ambos calculados con anterioridad. Antes de realizar la suma total aplicaremos según a la ley vigente, tanto al coste de personal como al de ejecución un Impuesto sobre el Valor Añadido (IVA) del 21%.

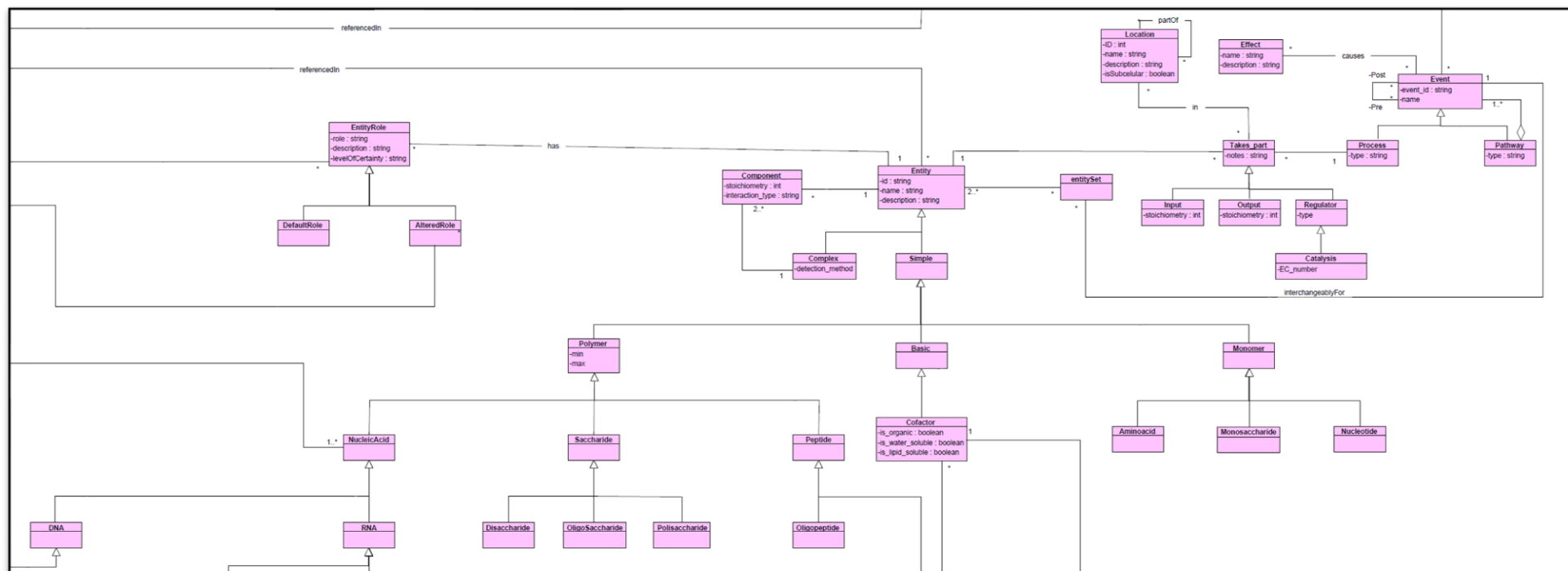
Tabla 9. Coste total del proyecto. Fuente: Elaboración propia

	Coste imputable sin IVA (€)	Coste con IVA (€)
Coste de personal	9.000,00	10.890,00
Coste de ejecución	173,44	209,86
		Total: 11.099,86

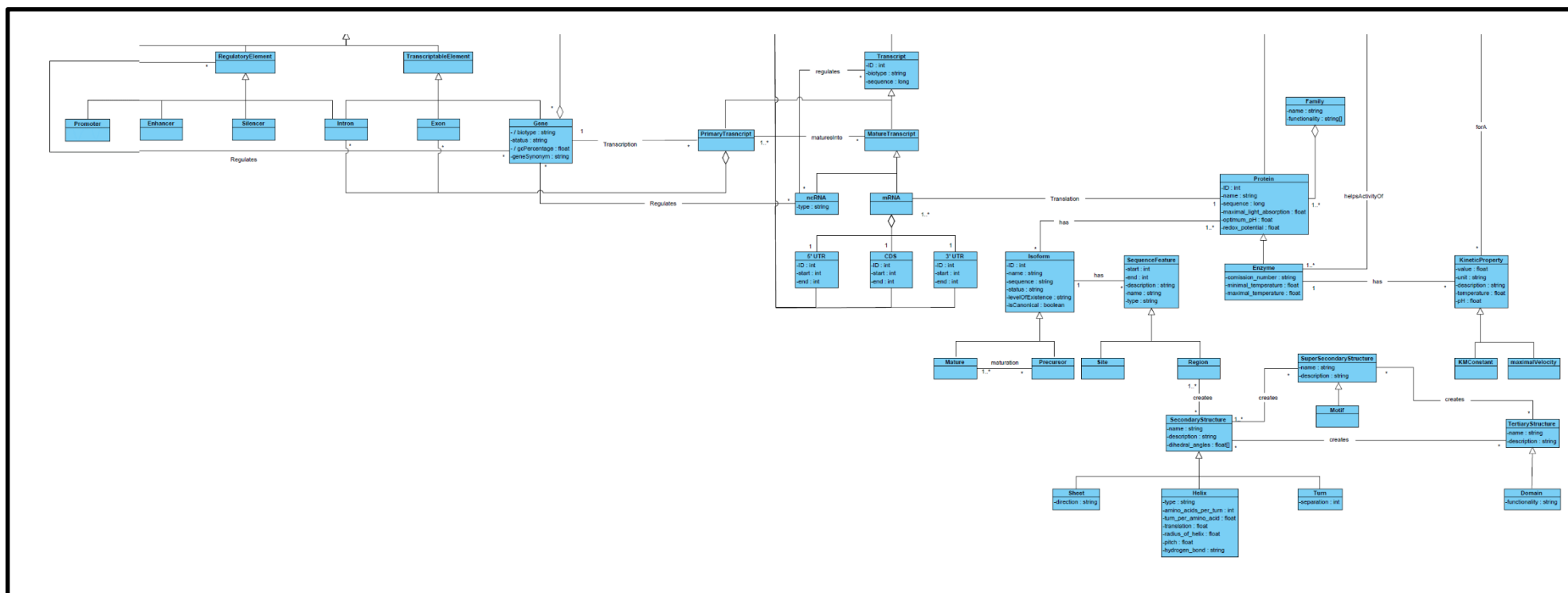
DOCUMENTO 3: ANEXOS

ANEXO I

VISTA PATHWAY COMPLETA



VISTA PROTEINA COMPLETA



ANEXO II

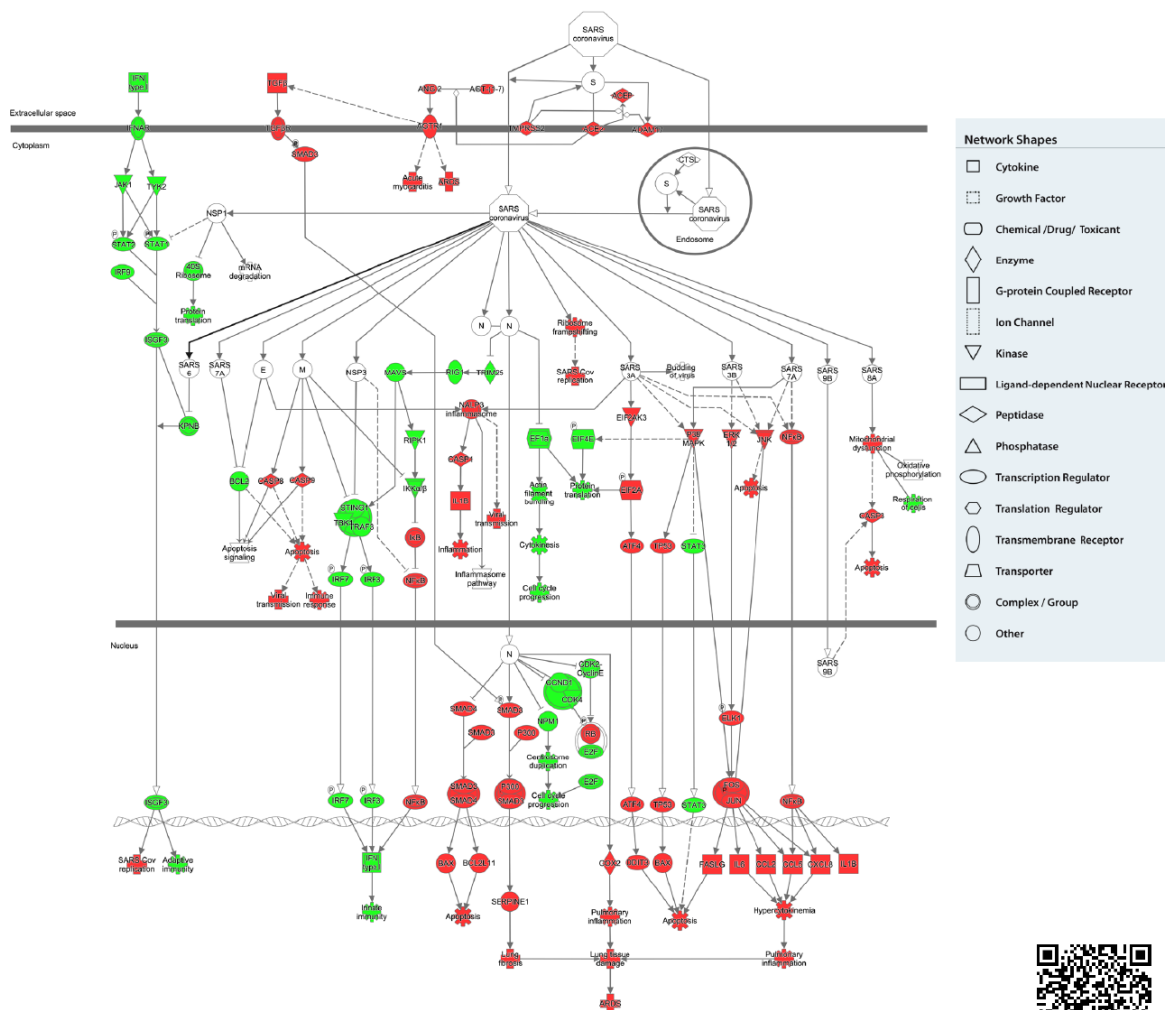


Ingenuity
Pathway
Analysis



Coronavirus Pathogenesis Pathway

from QIAGEN® Ingenuity® Pathway Analysis (IPA®)



Coronaviruses alter key cellular pathways to enhance their replication and virulence (see back page for detailed description). Nodes are color coded according to inference that they are inhibited (green) or activated (red).

Learn more at digitalinsights.qiagen.com/IPA



Sample to Insight

