

a

Una compuerta de glicano controla la apertura de la proteína espiga del SARS-CoV-2

- Terra Sztain^{1,8}, Surl- Hee Ahn^{1,8}, Anthony T. Bogetti², Lorenzo Casalino¹, Jory A. Goldsmith³, Evan Seitz⁴, Ryan S. McCool³, Fiona L. Kearns¹, Francisco Acosta-Reyes⁵, Suvrajit Maji⁵,
- Ghoncheh Mashayekhi⁶, J. Andrew McCammon^{1,7}, Abbas Ourmazd⁶, Joachim Frank^{4,5},
- Jason S. McLellan³, Lillian T. Chong²  y Rommie E. Amaro¹ 

La infección por SARS-CoV-2 se controla mediante la apertura del dominio de unión al receptor de proteína espiga (RBD), que pasa de un estado "hacia abajo" protegido por glicanos a un estado expuesto "hacia arriba" para unirse al receptor humano de la enzima convertidor de angiotensina 2 e infectar las células. Si bien se han obtenido instantáneas de los estados "arriba" y "abajo" mediante microscopía crio-electrónica y crio-tomografía electrónica, los detalles de la transición de apertura RBD evaden la caracterización experimental. Aquí, más de 130 μ s de simulaciones de conjunto ponderado del ectodominio de pico totalmente glicosilado nos permiten caracterizar más de 300 vías de apertura RBD continuas y cinéticamente imparciales. Junto con el análisis de ManifoldEM de los datos de microscopía crio-electrónica y los experimentos de interferometría de biocapa, revelamos un papel de GATILLO para el *N-glicano* en la posición N343, lo que facilita la apertura de RBD. También participan los residuos D405, R408 y D427. La caracterización a nivel atómico del mecanismo de activación de picos glicosilados proporcionado en este documento representa un estudio histórico para simulaciones de vías de conjunto y ofrece una base para comprender los mecanismos fundamentales de Entrada e infección viral por SARS-CoV-2.

¹ Departamento de Química y Bioquímica, Universidad de California–San Diego, La Jolla, CA, Estados Unidos. ² Departamento de Química, Universidad de Pittsburgh, Pittsburgh, PA, ESTADOS UNIDOS. ³ Departamento de Molecular Biosciences, Universidad de Texas en Austin, Austin, TX, ESTADOS UNIDOS. ⁴ Departamento de Biología, Columbia Universidad, Nueva York, NY, Estados Unidos. ⁵ Departamento de Bioquímica y Molecular Biofísica, Columbia Universidad Medical Center, Nueva York, NY, Estados Unidos. ⁶ Departamento de Física, Universidad de Wisconsin–Milwaukee, Milwaukee, WI, ESTADOS UNIDOS. ⁷ Departamento de Farmacología, Universidad de California–San Diego, La Jolla, CA, Estados Unidos. ⁸ Estos Autores Contribuido igualmente: Terra Sztain, Surl-Hee Ahn.  correo electrónico: ltchong@pitt.edu; ramaro@ucsd.edu



El coronavirus 2 del síndrome respiratorio agudo severo (SARS-CoV-2) es un virus de ARN envuelto y el agente causal de la enfermedad por coronavirus 2019 (COVID-19), una enfermedad que ha causado morbilidad y mortalidad esencial en todo el mundo^{1,2}. La principal maquinaria de infección del virus, la proteína espiga que se encuentra en el exterior del virus, es el primer punto de contacto que el virión hace con la célula huésped, y es un antígeno viral importante³. Recientemente se ha informado de un número sustancial de estructuras de microscopía crio-electrónica (crio-EM) de la proteína espiga, informando colectivamente sobre los estados estructurales de la proteína espiga. La gran mayoría de las estructuras resueltas caen en estados "abajo" o "arriba", según lo definido por la posición del dominio de unión al receptor (RBD), que modula la interacción con el receptor de la enzima convertidoras de angiotensina 2 (ACE2) para la entrada celular⁴⁻⁶.

El RBD mosto transición De a 'abajo' Para un 'arriba' estado para el receptor encuadernación motivo (RBM) Para ser accesible para ACE2 encuadernación (Figura. 1), y por lo tanto la activación mecanismo es esencial para celda entrada. Lu et al.⁷ usado soltero-molécula En este estudio, caracterizamos la vía de apertura de RBD de pico para la proteína espiga de SARS-CoV-2 completamente glicosilada para obtener una comprensión detallada del mecanismo de activación. Se utilizó la estrategia de muestreo de trayectoria de conjunto ponderado (WE)^{12,13} (Fig. 1 suplementaria) para permitir la simulación de vías atómicas para el proceso de apertura de picos. Como estrategia de muestreo de trayectoria, WE centra la potencia de cálculo en las transiciones funcionales entre estados estables en lugar de los estados estables en sí mismos¹⁴. Esto se logra ejecutando múltiples trayectorias en paralelo y replicando periódicamente trayectorias que han pasado de regiones del espacio configuracional 15 visitadas con anterioridad a regiones recién visitadas, minimizando así el tiempo de espera en el estado estable inicial para transiciones "afortunadas" sobre la barrera de energía libre. Dado que estas transiciones son mucho más rápidas que los tiempos de espera^{16,17}, la estrategia WE puede ser órdenes de magnitud más eficiente que las simulaciones convencionales de MD en la generación de vías para eventos raros como el plegamiento de proteínas y la unión a proteínas^{18,19}. Esta eficiencia es aún mayor para procesos más lentos, aumentando exponencialmente con la barrera efectiva de energía libre²⁰. No solo la

a
objetivo, la definición de este estado objetivo no necesita fijarse de antemano cuando se aplica en condiciones de equilibrio²¹, lo que nos permite refinar la definición del estado objetivo "abierto" de la proteína espiga sobre la base de la distribución de probabilidad de las conformaciones proteicas muestreados por la simulación.

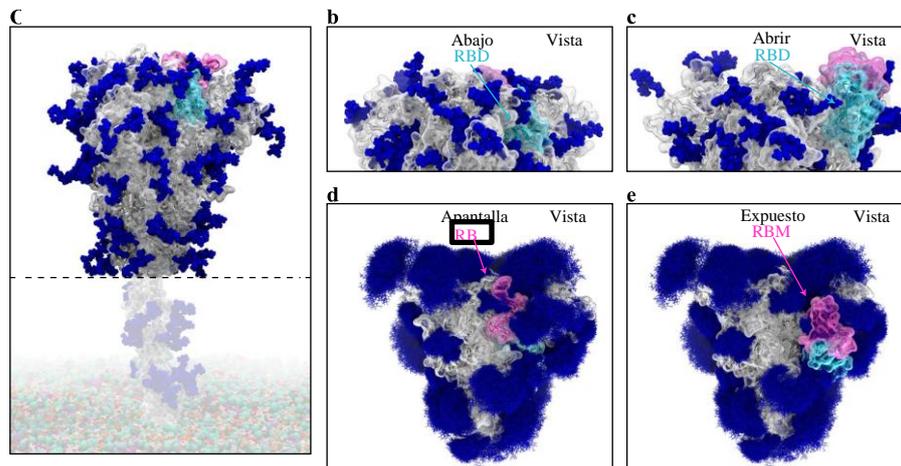


Figura 1 | Espiga glicosilada RBD 'hacia abajo' y conformaciones 'abiertas'. a, La cabeza de la espiga del SARS-CoV-2 (gris) con glicanos (azul oscuro) simulados, con el dominio del tallo y la membrana (no simulados aquí, pero se muestran como transparentes para su integridad). RBD se muestra en cian, RBM en rosa. **b,c**, Vista lateral del RBD hacia abajo (blindado, **b**) y RBD abierto (expuesto, **c**). **d,e**, Vista superior de la RBM cerrada (blindada, **d**) y RBM abierta (expuesta, **e**). La imagen compuesta de glicanos (líneas azul oscuro) muestra muchas instantáneas superpuestas de los glicanos durante las simulaciones de microsegundos.

Nuestro trabajo caracteriza una serie de vías de transición de la apertura de la espiga, de acuerdo con las conformaciones detectadas en el conjunto de datos crio-EM por ManifoldEM²², e identifica residuos clave, incluido un glicano en la posición N343, que participan en el mecanismo de apertura. Nuestros hallazgos de simulación están corroborados por experimentos de interferometría de biocapa (BLI), que muestran una reducción en la capacidad del pico para interactuar con ACE2 después de la mutación de estos residuos clave.

Resultados y discusión

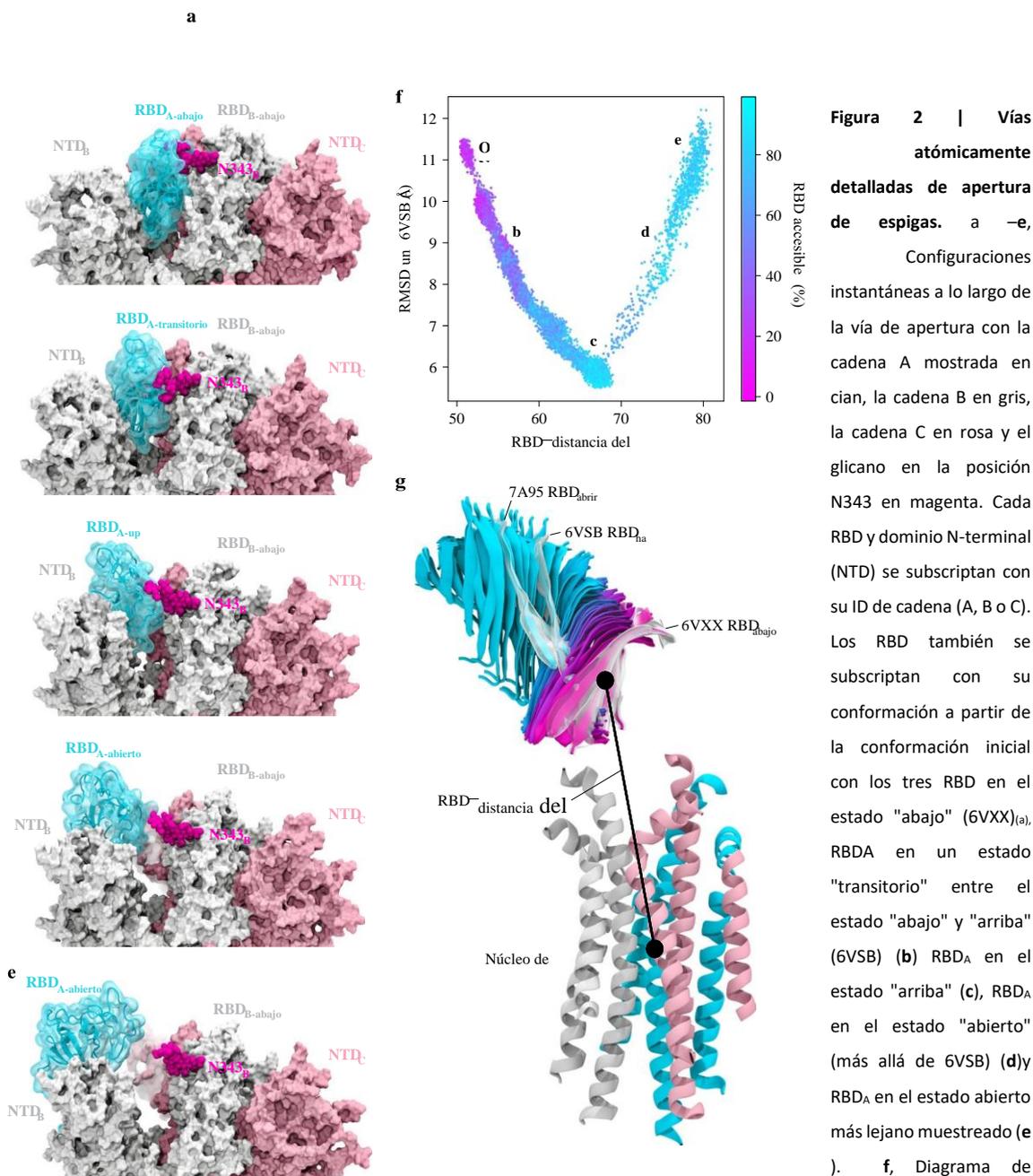
Simulaciones WE de apertura de espiga. Como se mencionó anteriormente, las simulaciones del proceso de apertura de picos requieren una estrategia de muestreo mejorada, ya que el proceso ocurre más allá de la escala de tiempo de microsegundos

^a
(es decir, la escala de tiempo de segundos⁷). Por lo tanto, utilizamos la estrategia en el panorama de la energía libre, sino que la asignación adecuada de pesos estadísticos a las trayectorias proporciona una caracterización imparcial de las propiedades del conjunto dependiente del tiempo del sistema¹³. Por lo tanto, la estrategia WE genera vías continuas con dinámicas imparciales, produciendo las vistas más directas y atomistas para analizar el mecanismo de las transiciones funcionales, incluida la elucidación de estados transitorios que son demasiado fugaces para ser capturados por experimentos de laboratorio. Además, si bien la estrategia requiere una coordenada de progreso hacia el estado de muestreo de ruta WE, que permitió la generación de vías continuas y atomísticas para el proceso de apertura de picos con dinámicas imparciales (Fig. 2a–e y Video Complementario 1); estas vías tenían cientos de nanosegundos de largo, excluyendo los tiempos de espera en el estado inicial "abajo". El modelo proteico se basó en la región de la cabeza (residuos 16 a 1.140) del pico glicosilado del SARS-CoV-2 de Casalino et al.⁸ (Fig. 1), que a su vez se construyó sobre la estructura crio-EM del pico de tres RBD -down (Protein Data Bank (PDB) ID, 6VXX (ref. 5)). Todo el sistema de simulación, incluidos los iones explícitos de agua y sal, alcanza casi medio millón de átomos. Enfocamos el muestreo a lo largo de una coordenada de progreso bidimensional para rastrear la apertura de RBD: la diferencia en el centro de masa del núcleo de espiga con respecto a la RBD y la desviación raíz-media-cuadrada de la RBD desde el estado_{RBD hacia arriba} (Fig. 2f,g). En las supercomputadoras Comet del San Diego Supercomputer Center (SDSC) y Longhorn del Texas Advanced Computing Center (TACC), 100 unidades de procesamiento de gráficos (GPU) ejecutaron las simulaciones WE en paralelo durante más de un mes, generando más de 130 μ s de trayectorias de pico glicosiladas y más de 200 TB de datos de trayectoria. Simulamos un total de 310 vías independientes, incluyendo 204 vías desde la conformación RBD_{hacia abajo} (PDB ID, 6VXX (ref. 5)) a la conformación RBD_{up} (PDB ID, 6VSB (ref. 4)) y 106 vías desde el RBD_{hasta el estado abierto} RBD, en el que el RBD se abre más allá del 6VSB (ref. 4) estructura crio-EM. Sorprendentemente, el estado_{abierto} de RBD que muestreamos incluye conformaciones que se alinean estrechamente con la estructura crio-EM de pico unida a ACE2 (PDB ID, 7A95 (ref. 6)) a pesar de que esta estructura no era un estado objetivo de nuestra coordenada de progreso (Fig. 2f,g, Video Suplementario 1 y Figs. 2 y 3 Suplementarios). Este resultado subraya el valor de

a

usar (1) simulaciones WE de equilibrio que no requieren una definición fija del estado objetivo y (2) una coordenada de progreso bidimensional que permite a las simulaciones muestrear un espacio conformacional inesperado a lo largo de múltiples grados de libertad. La conformación de picos ligada a ACE2 también ha sido muestreada por el proyecto de computación distribuida por **Comparación con conformaciones de pico detectadas por ManifoldEM.**

Para validar nuestra ruta simulada de RBD_{hasta} RBD_{hacia arriba}, se aplicó el marco ManifoldEM²² utilizando el conjunto de datos crio-EM de PDB 6VSB de McLellan y sus colegas⁴. El método ManifoldEM permite la caracterización de variaciones conformacionales obtenidas a partir de un conjunto crio-EM de una sola partícula de una molécula en equilibrio térmico. Dos coordenadas conformacionales (es decir, coordenadas de movimiento colectivo) CC1 y CC2 fueron descubiertas a partir de este conjunto de datos, y observadas a partir de varias direcciones de proyección ejemplares. (PD) que muestran una (1) vía de RBD_{hacia arriba} y (2) vía de apertura hacia afuera de RBD (Fig. 4 suplementaria y Videos suplementarios 2 y 3). ding@home¹¹, y la rotación rbd se ha detectado en experimentos crio-EM⁶.



Estas proyecciones se alinearon a continuación con las proyecciones bidimensionales correspondientes de mapas de potencial de coulomb generados con fotogramas de la

^a
simulación WE (Fig. 5 suplementaria y Videos suplementarios 2 y 3). En general, hubo muy buena concordancia entre las coordenadas conformacionales de ManifoldEM y la trayectoria WE, aparte de dos discrepancias. Primero, el CC2 observado en el ManifoldEM incluyó la apertura concertada de los tres RBD, mientras que el WE centró el muestreo en la apertura de un solo RBD (Video Complementario 2). En segundo lugar, la trayectoria WE finalmente se abre a una distancia RBD-núcleo 11 Å mayor que la conformación más abierta en el ManifoldEM. Esto es probable porque las simulaciones muestrean la subunidad S1 en ruta hacia la conformación posterior a la fusión, mientras que el conjunto de datos experimentales no lo hace.

Las puertas de glicano N343 RBD se abren. En el estado 'abajo', el RBD del pico del SARS-CoV-2 está protegido por glicanos en posiciones investigadas para los estados RBD_{hacia abajo} y RBD_{hacia arriba}⁸. Nuestras simulaciones WE permitieron la caracterización del blindaje durante el proceso de apertura, revelando una disminución abrupta en el blindaje de glicanos cuando el RBD pasa del estado "hacia abajo" al estado "arriba". Los glicanos en la posición N165 y N234 protegen consistentemente el RBM, mientras que el blindaje por el glicano N343 disminuye con la apertura de RBD (Fig. 6 suplementaria). Más allá del blindaje, recientemente se ha reportado un papel estructural para los glicanos en las posiciones N165 y N234, estabilizando el RBD en la conformación 'up' a través de un mecanismo de 'carga y bloqueo'⁸.

Nuestras simulaciones WE revelan un papel aún más específico y crítico de un glicano en el mecanismo de apertura de la espiga: el glicano N343 actúa como una "puerta de glicano" que empuja el RBD desde la conformación "hacia abajo" hasta la conformación "hacia arriba" intercalando entre los residuos F490, Y489, F456 y R457 del motivo de unión ACE2 en un movimiento de "mano-gire" (Fig. 2a-e, 3 y Video Complementario 4). Por lo tanto, el glicano N343 desempeña un papel activo en el inicio de la transición, distinto de los roles estabilizadores de los glicanos N165 y N234. Este mecanismo de apertura se visualizó inicialmente en varias vías exitosas de apertura de espigas y luego se confirmó a través del análisis de las 310 vías exitosas en las que se encontró que el glicano N343 formaba contactos (dentro de 3.5 Å) con cada uno de los residuos mencionados anteriormente en cada vía exitosa (Fig. 7 suplementaria). El mismo comportamiento mecanicista del glicano N343 se observó en dos simulaciones WE

a

totalmente independientes, lo que sugiere que el resultado es robusto a pesar del muestreo potencialmente incompleto que puede desafiar a WE y otros métodos mejorados de simulación de muestreo¹⁵.

Para probar el papel del glicano N343 como residuo de llave, realizamos experimentos BLI. Los experimentos BLI evalúan el nivel de unión de la RBM de espiga (residuos 438 a 508) a ACE2, actuando como un proxy de la proporción relativa de RBD en la posición "arriba" para cada variante de espiga. No se mutaron residuos directamente involucrados en la unión (es decir, en la interfaz RBM-ACE2) para garantizar la detección controlada del impacto de la apertura de RBD en respuesta a las mutaciones. Aunque los resultados N165A y N234A en la proteína^{8S-2P} del SARS-CoV-2, la variante N343A mostró una disminución aún mayor en la unión a ACE2, reduciendo el nivel de unión al pico en $\sim 56\%$ (Fig. 4 y cuadro complementario 1). Como control negativo, la variante 24 de S383C/D985C, que se espera que sea bloqueada por disulfuros en la conformación de tres RBD- hacia abajo, no mostró asociación con el receptor ACE2. Estos resultados apoyan la hipótesis de que la conformación RBD_{up} se ve sustancialmente afectada por la glicosilación en la posición N343.

Detalles atómicos del mecanismo de apertura. El estado_d de RBD presenta un enlace de hidrógeno entre T415 del RBD_A y K986 de la cadena C, un puente de sal entre R457 de RBD_A y D364 de RBD_B, y un puente de sal entre K462 de RBD_A y D198 de dominio N-terminal C (NTD_C) (Fig. 5a–c,e y Fig. 8 suplementaria). El enlace de hidrógeno T415_A–K986_C gasta un promedio del 12% de las vías exitosas hacia el estado 'up' antes de que K986_C haga un puente de sal de corta duración (2% de duración promedio al estado 'up') con RBD_A–D427. (Figura. 5b,e y Fig. 8 suplementaria). A continuación, K986_C forma puentes de sal con E990_C y E748_C a medida que el RBD_A continúa abriendo. Estos contactos se forman en las 310 vías exitosas (Fig. 8 suplementaria). La mutación de K986 a prolina se ha utilizado para estabilizar el pico de prefusión^{25,26}, incluso en el desarrollo de lavacuna²⁷, y estas simulaciones proporcionan contexto molecular a un papel adicional de este residuo en la apertura de RBD. Posteriormente, a un promedio del 16% del camino a través de las vías exitosas hacia el estado 'up', el puente de sal R457_A–D364_B se rompe, lo que lleva al RBD_A a girar hacia arriba, lejos de RBD_B hacia RBD_C y formando un puente de sal entre R408 de RBD_A y D405 de RBD_C (Fig. 5c,e y Fig. 8

a
suplementaria). Este puente de sal persiste durante el 20% de las trayectorias exitosas hacia el estado 'up' y está presente en las 310 vías exitosas.

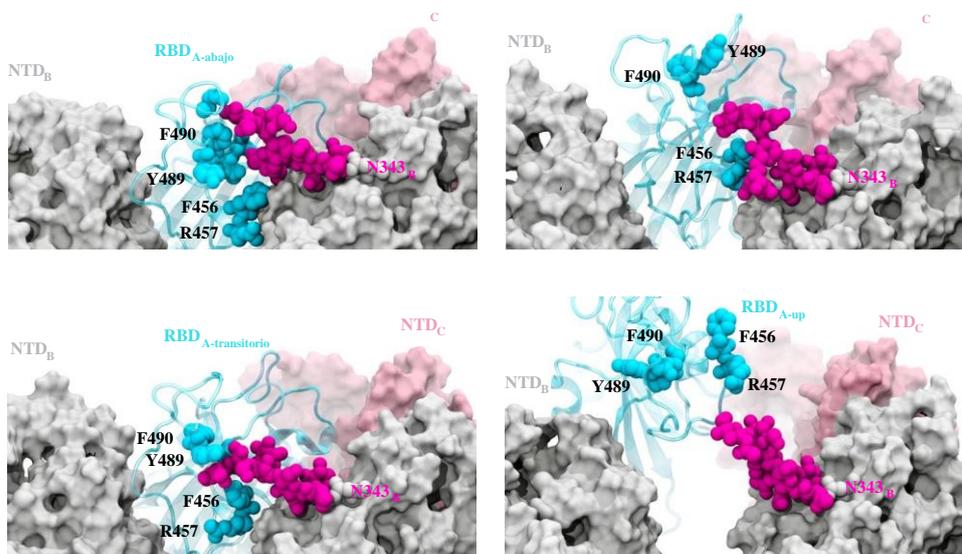


Figura 3 | Glycan gating por N343. a–d, Configuraciones instantáneas a lo largo de la vía de apertura con la cadena A mostrada en cian, la cadena B en gris, la cadena C en rosa y el glicano en la posición N343 en magenta. RBD_A en la conformación 'abajo' está protegida por el glicano en la posición N343 del RBD_B adyacente (a). El glicano N343 intercala entre (b) y debajo (c) los residuos F490, Y489, F456 y F457 para empujar el RBD hacia arriba y abrir (d).

Un puente de sal entre R466 de RBD_A y E132 de NTD_B está presente en 189 de las 204 vías exitosas hacia el estado "arriba", y las 106 vías hacia el estado "abierto". Este contacto es más frecuente durante la transición entre el estado "arriba" y "abierto". Finalmente, el puente de sal entre D428 de RBD_A y R454 de RBD_C está presente solo en las 106 vías desde el estado "hacia arriba" hasta el estado "abierto" y es el último puente de sal entre el RBD y el pico en el estado "abierto" antes de que la subunidad S1 comience a desprenderse (Fig. 5d,e y Fig. 8 suplementaria), momento en el que el último contacto restante con el RBD_A es el glicano en la posición N165 de NTD_B.

Experimentos adicionales de BLI de los residuos de picos clave identificados R408A, D405A y D427A corroboran las vías observadas en nuestras simulaciones. Cada uno de estos reduce las interacciones de unión del pico con ACE2 en ~ 13%, ~ 27% y ~ 52%, respectivamente (Fig. 4 y Tabla suplementaria 1). También observamos que los residuos identificados D198, N343, D364, D405, R408, T415, D427, D428, R454, R457, R466, E748, K986 y E990 se conservan entre los picos de SARS-CoV y SARS-CoV-2, lo que respalda su importancia en la coordinación de la función primaria de pico de la apertura de RBD. Las cepas mutantes emergentes de SARS-CoV-2, B.1 (D614G), B.1.1.7 (deleción H69–V70 y deleciones Y144–Y145, N501Y, A570D, D614G, P681H, T716I, S982A y D1118H), B.1.351

^a
(L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G y A701V), P.1 (L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y y T1027I) y CAL.20 C (L452R y D614G)²⁸, no contienen mutantes en los residuos que identificamos aquí para facilitar la apertura de RBD. El análisis de los residuos y glicanos vecinos a los mutados en las cepas emergentes a lo largo de de apertura de RBD. Las vías de apertura simuladas se alinean muy bien con las conformaciones detectadas a partir de crio-EM con el método ManifoldEM. El análisis de estas vías a partir de simulaciones WE independientes indica un claro papel de cierre para el glicano en N343, que levanta y estabiliza el RBD a lo largo de la transición de apertura. También caracterizamos un estado "abierto" de la espiga RBD, en la que el glicano N165 de la cadena B es el último contacto restante con la RBD en el camino hacia una mayor apertura de S1. Los experimentos BLI de residuos identificados como clave en las transiciones de apertura, incluidos N343, D405, R408 y D427, respaldaron ampliamente nuestros hallazgos computacionales. En particular, una disminución del 56% en la unión a ACE2 del mutante N343A, en comparación con una disminución del 40% en el mutante N234A y una disminución del 10% en el mutante N165A reportado anteriormente⁸ evidenció el papel clave de N343 en la apertura de la RBD, destacando la importancia de muestrear las transiciones funcionales para comprender completamente los detalles mecanicistas. Ninguna de las mutaciones individuales abolió completamente la unión a ACE2, lo que indica que el virus ha desarrollado un mecanismo que involucra múltiples residuos para coordinar la apertura de la espiga. Nuestro trabajo indica un papel crítico del glicano N343 en la apertura de espigas y proporciona nuevos conocimientos sobre los mecanismos de infección viral para este importante patógeno.

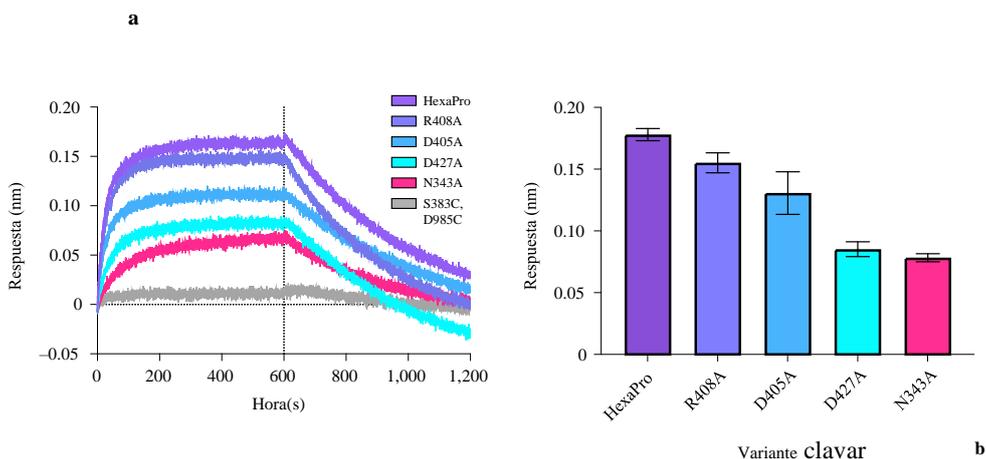


Figura 4 | La unión a ACE2 se reduce por mutación del sitio de glicosilación N343 y residuos clave del puente salino. **a**, sensorgramas BLI de variantes de pico HexaPro que se unen a ACE2. Para mayor claridad, solo se muestran los rastros de la primera réplica. **b**, Gráfico de la respuesta vinculante para los datos BLI recogidos por triplicado con barras de error que representan la desviación estándar de la media.

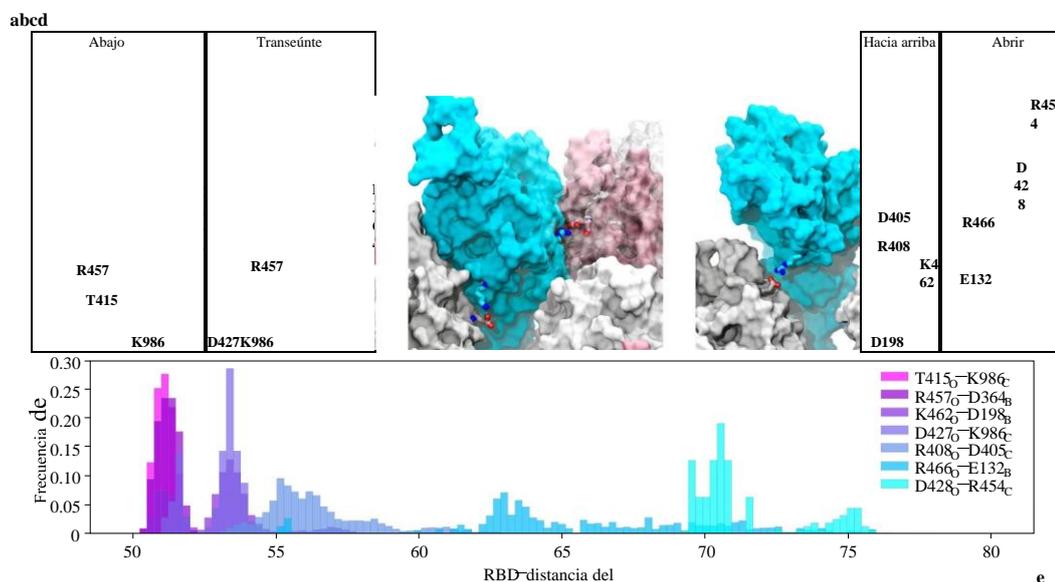


Figura 5 | Puentes de sal y enlaces de hidrógeno a lo largo de la vía de apertura. **a-d**, Contactos de puente de sal o de enlace de hidrógeno realizados entre RBD_A, mostrado en azul, y RBD_B, mostrado en gris, o RBD_C, mostrado en rosa, dentro de las conformaciones 'down' (**a**), 'transitoria' (**b**), 'up' (**c**) y 'open' (**d**). Los átomos de nitrógeno son de color azul y los átomos de oxígeno son de color rojo en la representación de los aminoácidos para mostrar contactos de enlace de hidrógeno y puente de sal. **e**, Histograma que muestra la frecuencia a la que los residuos de **a-d** están dentro de 3.5 Å entre sí en relación con la distancia RBD-núcleo. Las frecuencias se normalizan a 1.

Contenido en línea

Cualquier método, referencias adicionales, resúmenes de informes de Nature Research, datos de origen, datos ampliados, información complementaria, agradecimientos, información de revisión por pares; detalles de las contribuciones de los autores y los intereses contrapuestos; y las declaraciones de disponibilidad de datos y códigos están disponibles en <https://doi.org/10.1038/s41557-021-00758-3>.

Recibido: 15 de febrero de 2021; Aceptado: 21 de febrero de 2021
Published online: 19 August 2021

Referencias

1. Chan, J. F.-W. et al. Un grupo familiar de neumonía asociada con el nuevo coronavirus de 2019 que indica transmisión de persona a persona: un estudio de un grupo familiar. *Lancet* **395**, 514–523 (2020).

a

2. Lu, R. et al. Caracterización genómica y epidemiología del nuevo coronavirus 2019: implicaciones para los orígenes del virus y la unión al receptor. *Lancet* **395**,565–574 (2020).
3. Li, F. Estructura, función y evolución de las proteínas espiga del coronavirus. *Annu. Reverendo Virol.* **3**,237–261 (2016).
4. Wrapp, D. et al. Estructura crio-EM del pico 2019-NCov en la conformación de prefusión. *Ciencia* **367**,1260–1263 (2020).
5. Muros, A.C. et al. Estructura, función y antigenicidad de la glicoproteína espiga del SARS-CoV-2. *Celda* **181**,281–292.e6 (2020).
6. Benton, D. J. et al. Unión al receptor y cebado de la proteína espiga del SARS-CoV-2 para la fusión de membranas. *Naturaleza* **588**,327–330 (2020).
7. Lu, M. et al. Dinámica conformacional en tiempo real de los picos de SARS-CoV-2 en partículas de virus. *Cell Host Microbe* **28**, 880–891.e8 (2020).
8. Casalino, L. et al. Más allá del blindaje: el papel de los glicanos en la proteína espiga del SARS-CoV-2. *ACS Cent. Ciencia* **6**,1722–1734 (2020).
9. Gur, M. et al. Transición conformacional de la glicoproteína espiga del SARS-CoV-2 entre sus estados cerrado y abierto. *J. Química. Phys.* **153**, 075101 (2020).
10. Fallon, L. et al. Paisajes de energía libre para la apertura de RBD en SARS-CoV-2
11. Las simulaciones de glicoproteínas de pico sugieren interacciones clave y una bolsa alotérica potencialmente farmacogable. Preimpresión en *ChemRxiv* <https://doi.org/10.26434/chemrxiv.13502646.v1> (2020). Zimmerman, M. I. et al. Las simulaciones del SARS-CoV-2 van a exaescala para predecir la apertura dramática de picos y bolsas crípticas a través del proteoma. *Nat. Química.* **13**,651–659 (2021).
12. Huber, G. A. & Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* **70**,97–110 (1996).
13. Zhang, B. W., Jasnow, D. & Zuckerman, D.M. El método de muestreo de la ruta del "conjunto ponderado" es estadísticamente exacto para una amplia clase de procesos estocásticos y procedimientos de binning. *J. Química. Phys.* **132**, 054107 (2010).
14. Chong, L. T., Saglam, A. S. & Zuckerman, D.M. Estrategias de muestreo de trayectorias para simular eventos raros en sistemas biomoleculares. *Curr. Opin. Estructura. Biol.* **43**,88–94 (2017).
15. Zuckerman, D.M. & Chong, L. T. Simulación de conjunto ponderado: revisión de metodología, aplicaciones y software. *Annu. Biophys.* **46**,43–57 (2017).
16. Pratt, L. R. Un método estadístico para identificar estados de transición en problemas de alta dimensión. *J. Química. Phys.* **85**,5045–5048 (1986).
17. Zuckerman, D.M. & Woolf, T.B. Eventos de transición en simulaciones de butano: similitudes entre modelos. *J. Química. Phys.* **116**,2586–2591 (2002).
18. Adhikari, U. et al. Estimación computacional de tiempos de plegamiento de microsegundos a segundo atómico. *J. Am. Química. Soc.* **141**, 6519–6526 (2019).
19. Saglam, A. S. & Chong, L. T. Proteína: Vías de unión a proteínas y cálculos de constantes de velocidad utilizando simulaciones totalmente continuas y explícitas de disolventes. *Química. Ciencia.* **10**, 2360–2372 (2019).
20. DeGrave, A. J., Ha, J.-H., Loh, S. N. & Chong, L. T. Gran mejora de los tiempos de respuesta de un interruptor conformacional de proteínas por diseño computacional. *Nat. Comunidad.* **9**,1013 (2018).
21. Suárez, E. et al. Cálculo simultáneo de información dinámica y de equilibrio utilizando un conjunto ponderado de trayectorias. *J. Química. Teoría de la Computación.* **10**,2658–2667 (2014).
22. Dashti, A. et al. Trayectorias del ribosoma como nanomáquina browniana. *Proc. Natl Acad. Ciencia. Estados Unidos* **111**,17492–17497 (2014).
23. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Análisis de glicanos específicos del sitio del pico del SARS-CoV-2. *Ciencia* **369**,330–333 (2020).
24. Henderson, R. et al. Controlling the SARS-CoV-2 spike glycoprotein conformation (Control de la conformación de glicoproteínas de espiga del SARS-CoV-2). *Nat. Estructura. Mol. Biol.* **27**,925–933 (2020).
25. Hsieh, C.-L. et al. Diseño basado en la estructura de picos de SARS-CoV-2 estabilizados previamente a la fusión. *Ciencia* **369**,1501–1505 (2020).
26. Pallesen, J. et al. Inmunogenicidad y estructuras de un antígeno de pico del MERS-CoV de prefusión diseñado racionalmente. *Proc. Natl Acad. Ciencia. USA* **114**, E7348–E7357 (2017).
27. Cruz, R. El pequeño retoque detrás de las vacunas contra el COVID-19. *Química.* 18–20 (2020) (18–2020).
28. Corum, J. & Zimmer, C. Variantes y mutaciones del coronavirus. *New York Times* (10 de febrero de 2021).
29. Amaro, R. E. & Mulholland, A. J. Una carta de la comunidad sobre el intercambio de datos de simulación biomolecular para COVID-19. *J. Química. Inf. Modelo.* **60**,2653–2656 (2020).
30. Bogetti, A. T. et al. Un conjunto de tutoriales para el software de muestreo de eventos raros WESTPA. *Vivir J. Informática. Mol. Ciencia.* **1**, 10607 (2019).

Disponibilidad de datos

Los datos que respaldan los hallazgos de este estudio se incluyen en el artículo y sus archivos de información complementaria. Respalamos los principios de la comunidad en torno al intercambio abierto de datos de simulación de COVID-19²⁹. Todos los archivos y datos de entrada de simulación están disponibles en el NSF MolSSI COVID-19 Molecular Structure and Therapeutics Hub en <https://covid.molssi.org> y en el sitio web de Amaro Lab <http://amarolab.ucsd.edu>. Los datos de origen se proporcionan con este documento.

a

Disponibilidad de código

Este estudio utilizó las compilaciones estándar del software de simulación WESTPA 2020.02 (<https://github.com/westpa/westpa>) y AMBER 18 (<https://ambermd.org>) de acuerdo con las mejores prácticas para ejecutar simulaciones WE³⁰ con sin modificaciones especiales.

Agradecimientos

Estamos agradecidos por los esfuerzos del Centro de Computación Avanzada de Texas (TACC) Equipo de Longhorn y por el tiempo de cómputo disponible a través de una Asignación Discrecional del Director (posible gracias al premio OAC-1818253 de la National Science Foundation (NSF)). Agradecemos a Z. Gaieb para discusiones útiles sobre la construcción del sistema.

Agradecemos a M. Tatineni por su ayuda con la computación en SDSC Comet, así como un Premio del Consorcio HPC COVID-19 por el tiempo de computación. También agradecemos a C. Simmerling y su grupo de investigación (SUNY Stony Brook), y A. Mulholland y su grupo de investigación (Universidad de Bristol), para discusiones útiles relacionadas con la proteína espiga, así como D. Zuckerman, J. Copperman, M. Zwiery y A. Saglam para discusiones metodológicas útiles. T.S. está financiado por una subvención NSF GRFP (DGE-1650112). Este trabajo fue apoyado por una subvención de los Institutos Nacionales de Salud (NIH) (GM132826); una subvención NSF RAPID (MCB-2032054); un premio de RCSA Research Corp. y un UC San Diego Moores Cancer Center 2020 Subvención semilla sars-CoV-2 a R.E.A.; una subvención de los NIH (R01-GM31749) a J.A.M.; una subvención de los NIH (R01-AI127521) a J.S.M.; una subvención de los NIH (R01 GM115805) y una subvención de la NSF (CHE-1807301) a L.T.C.; y subvenciones de NIGMS (R01 GM29169 y R35 GM139453) a J.F. A.O. y G.M. reconocer el apoyo de el Departamento de Energía de los Estados Unidos, la Oficina de Ciencia, Ciencias Básicas de la Energía bajo el premio DE-SC0002164 (algoritmos dinámicos subyacentes), y por la Fundación Nacional de Ciencias de los Estados Unidos bajo los premios STC 1231306 (subyacente data- técnicas analíticas) y DBI-2029533 (datos subyacentes- modelos analíticos).

Contribuciones de los autores

T.S. y S.-H.A. contribuyeron igualmente a este trabajo. R.E.A. y L.T.C. supervisó el proyecto. T.S. y L.C. preparó el modelo de simulación. T.S. y S.-H.A. realizó simulaciones WE y A.T.B. proporcionó scripts WESTPA. S.-H.A., A.T.B., T.S. y L.T.C. realizó un análisis WE. T.S. y F.L.K. realizó análisis de simulación. L.C., T.S. y F.L.K. creó figuras y videos. J.S.M. diseñó y supervisó experimentos BLI.

R.S.M. y J.A.G. realizó experimentos BLI y escribió las partes correspondientes en Resultados y Métodos. J.F. y A.O. dirigió el estudio ManifoldEM, y E.S., F.A.-R., S.M. y G.M. realizó el estudio ManifoldEM. E.S. y F.A.-R. analizó los resultados y los comparó con simulaciones WE. E.S. y J.F. describió los métodos y resultados de ManifoldEM. T.S., S.H.-A., L.T.C. y R.E.A. escribió el manuscrito con contribuciones de todos los autores.