UNIVERSIDAD DEL PAÍS VASCO/EUSKAL HERRIKO
UNIBERTSITATEA

NANOMECHANICS OF PATHOGENIC ATTACHMENT: UROPATHOGENIC *ESCHERICHIA COLI* AND HUMAN IMMUNODEFICIENCY VIRUS

Memoria para optar al grado de Doctor en

Physics of Nanostructures and Advanced Materials

Álvaro Alonso Caballero

Director de tesis**: Raúl Pérez Jiménez**

DONOSTIA-SAN SEBASTIÁN, 2017

Agradecimientos

Quiero empezar dando las gracias a los organismos que me han financiado personalmente. En primer lugar al Gobierno Vasco por haberme concedido tanto la beca del programa predoctoral como la beca para realizar una estancia en el extranjero. En segundo lugar al centro de investigación CIC nanoGUNE por haberme financiado durante el comienzo mi tesis y por haberme brindado la oportunidad de realizar mi preparación doctoral. En cuanto a los proyectos de investigación quiero agradecer al Ministerio de Economía y Competitividad del Gobierno de España y a la Unión Europea.

Quiero agradecer a mi supervisor Raúl Pérez Jiménez por haberme dado la oportunidad de realizar mi tesis en un proyecto tan interesante que me ha permitido aprender y crecer como investigador y como persona. Además quiero agradecerle la confianza que ha depositado en mí y que me haya dado la posibilidad de realizar una estancia doctoral en uno de los mejores lugares del mundo para realizar Ciencia.

A mis compañeros de laboratorio Aitor, Nerea, Marie, Borja, Patricia y Bárbara por los buenos momentos que ha habido y por la ayuda y apoyo que me han dado durante estos más de cuatro años. *Eskerrik asko* y *merci beaucoup.* También, agradecer a Simon por la ayuda que me dio durante una parte del proyecto de mi tesis. Un agradecimiento muy especial merecen Leyre *txiki* por ser tan *jatorra* y buena persona, y Jörg, *Ich möchte mich besonders bei Jörg bedanken, dafür dass er mir bei meiner Doktorarbeit geholfen hat und so ein guter Freund ist. Ich hoffe, dass wir für immer Freunde sind.* Quiero agradecer a David por lo mucho que también me ha ayudado y por todas las interesantes discusiones científicas y no científicas que hemos mantenido a lo largo de este tiempo. También quiero agradecer a todas aquellas personas que han pasado por el grupo a lo largo de estos años y de las que también guardo muy buenos recuerdos.

Quiero agradecer a todas las personas de CIC nanoGUNE que me han ayudado, han sido buenas compañeras y también amigas. Entre estas últimas mención especial merecen el *italian team*: Matteo, Lorenzo, Nicolò y Guido (CFM). Jule, por ser una gran *errelaguna* y solo con el *laguna* aún más grande y espero que lo sea por mucho tiempo. Eider *alperrontxi* por ser tan buena gente y por haberme apoyado. El equipo de los viernes que incluye a Okuda y Néstor, y a Bente por pasarse de vez en cuando pero

estar siempre ahí para un café. Olvido a personas, pero les agradezco igualmente todo lo bueno que me han aportado, y que no es poco. *Eskerrik asko denoi*. *Madrilgo lagun bat daukazue*.

No debo olvidar al club Krav Maga Donostia, donde agradezco a los profesores Óscar y Josu por haberme empujado a superar mis límites y por haberme enseñado tanto. A mis compañer@s del nivel cero: Beñat, Javi y Xuria por todas las risas y los buenos momentos, gracias.

Esta tesis no habría sido posible tampoco sin el apoyo de mis amigos de toda la vida de Madrid, a los del barrio: Rafa, Bala, Revi, Álex y a los satélites intermitentes también, que ya son muchos años de amistad y que siga así por mucho tiempo. Mucho tengo que agradecer también a mis amig@s del máster de Biofísica de la UAM: Albert, Adriana, Clara y Dani; que me han aguantado y me han comprendido mejor que nadie a lo largo de estos años. También olvido aquí a personas. Pero no quiero olvidar dar a las gracias a Mónica, cuyo apoyo fue muy importante.

No puedo olvidarme tampoco de mi paso por el Instituto de Catálisis y Petroleoquímica, donde conocí a personas excepcionales y que me enseñaron muchísimo durante mi período allí. Marisela, *tito* Pablo (mi cicerone en AFM), Mario, Ileana, Óscar *ojitos*, Cristinita y muchas otras personas, gracias por todo.

También quiero agradecer el buen trato que me dieron Andrés, Rafa, Shuba, Ted, Dan, Jessica, Paulina y Carmelu en el grupo de Julio Fernández en Columbia University, donde pasé tres meses, aprendí muchísimo y espero seguir haciéndolo.

Por último, y los más importantes de todos, quiero dedicar esta tesis y los agradecimientos más especiales a mi familia. A mi abuelo José, que estaría muy orgulloso y al que le habría encantado discutir conmigo de lo que hago. A mi *tía* Victoria por haber estado siempre ahí desde que tengo uso de razón y, por supuesto, los agradecimientos más grandes para mis queridísimos padres Araceli y Ángel, a los que debo todo y que sin su ayuda y apoyo no habría conseguido nada de esto. A mis padres, que de haber tenido oportunidad en la vida habrían llegado muy lejos, dedico esta tesis por haber creído siempre en mí y por ser cruciales en mi vida.

"El ciego se entera mejor de las cosas del mundo,

los ojos son unos ilusionados embusteros"

R. M. del Valle-Inclán (*Luces de Bohemia*, 1920)

Resumen

En esta tesis doctoral el objetivo ha sido estudiar la nanomecánica de proteínas de superficie celular involucradas en la adhesión de patógenos a células humanas. Concretamente se han estudiado las proteínas que utiliza *Escherichia coli* uropatogénica para unirse al tracto urinario, y el receptor CD4 de los linfocitos T que es empleado por el Virus de la Inmunodeficiencia Humana (VIH) para unirse e infectar a estas células del sistema inmune.

Ambos procesos infecciosos han sido estudiados en profundidad desde diversos campos del conocimiento como la Medicina, la Biología, la Química y la Física, empleando diversas técnicas que han aportado información útil para el desarrollo de tratamientos médicos. Sin embargo, y pese a todos los avances logrados, estas enfermedades aún hoy suponen un desafío para la comunidad científica y médica. Las infecciones del tracto urinario son muy comunes y recurrentes entre la población. Aunque en muchos casos se resuelvan de manera favorable, el problema de la resistencia a antibióticos es cada vez mayor y ya hoy supone un problema grave de salud pública. Las cepas patogénicas están adquiriendo resistencia a cada vez más antibióticos, y tratamientos que antes funcionaban han dejado de hacerlo. Complicaciones derivadas de una infección del tracto urinario pueden desembocar en graves problemas de salud que en última instancia pueden provocar la muerte. En el caso del VIH los avances científicos han permitido transformar una enfermedad mortal en una enfermedad crónica, mejorando sensiblemente la calidad de vida de las personas que viven con la infección. No obstante a día de hoy no existe una cura para esta enfermedad de distribución mundial, de la que cada año

alrededor de un millón de personas se contagia. La enorme capacidad del virus para eludir al sistema inmune y para, de hecho, utilizar y destruir a las mismas células que deberían combatirlo lo convierten en un desafío para la Medicina.

Si bien estos dos patógenos son diferentes existen paralelismos entre ambos que además son exportables a otros casos. Independientemente de la etiología de cada caso, todo proceso infeccioso comienza por el reconocimiento y la adhesión del patógeno a la superficie de la célula que infecta. Para esto las bacterias y los virus utilizan moléculas expuestas tanto en sus propias superficies como en las de las células a las que se unen, siendo mayoritariamente proteínas las moléculas que median el anclaje específico entre los actores involucrados. Una vez superado este primer escollo del proceso infectivo el patógeno puede proceder con su ciclo de vida y expandirse, infectando nuevas células usando mecanismos similares y, en consecuencia, deteriorando la salud del huésped.

Tanto en el caso de *E. coli* uropatogénica como en el del VIH, el patógeno debe unirse a sus células diana de manera específica y superar obstáculos que se oponen a su adhesión física, tales como condiciones químicas adversas o la respuesta del sistema inmune. Entre estas dificultades se encuentran las fuerzas mecánicas, un ámbito muy inexplorado pero a la vez crucial tanto para el éxito como para el fracaso de una infección. El diseño y el uso de medicamentos que alteren mecánicamente a las proteínas empleadas durante la adhesión patogénica podría suponer un avance ante los desafíos que tiene que afrontar la Medicina moderna. A la luz de esta nueva aproximación orientada a alterar mecánicamente proteínas involucradas en

enfermedades humanas nace el campo de la Mecanomedicina o Mecanofarmacología, una nueva forma de estudiar y de tratar patologías donde un componente mecánico juega un papel importante. Debido a que las fuerzas mecánicas intervienen en este proceso, creemos que la adhesión patogénica constituye un ámbito de acción apropiado para el desarrollo de este nuevo campo.

El estudio desde este nuevo punto de vista cobra sentido cuando se atiende a las peculiaridades de la infección del tracto urinario. En su ruta de ascenso desde la uretra, *E. coli* debe avanzar por el tracto urinario y permanecer unida al epitelio en contra del flujo de orina. En ausencia de mecanismos de adhesión sería expulsada durante la micción y por ello, de entre un arsenal de factores, dispone de estructuras filamentosas denominadas pili o fimbrias que le sirven para unirse de manera específica a las células del epitelio urinario. Existen distintos tipos de fimbrias pero las empleadas por este organismo en el tracto urinario inferior son las denominadas de tipo I. Estos filamentos poseen diversas propiedades biomecánicas que le otorgan a la bacteria una adaptación extraordinaria para poder prosperar y resistir en este nicho tan adverso de la anatomía humana. Una cualidad muy interesante de estas estructuras, que pueden llegar a medir varias micras de longitud, es el modo en el que están organizadas y ensambladas. Estos filamentos están compuestos de cientos de proteínas dispuestas linealmente, en el que cada unidad de proteína está conectada a la siguiente mediante interacciones débiles de tipo hidrofóbico y puentes de hidrógeno. De manera que estas largas cadenas de proteínas conectadas, empleadas como anclajes que deben resistir estrés mecánico, mantienen su integridad mediante interacciones en

apariencia débiles pero que se han demostrado como unas de las más resistentes de la naturaleza.

Ante el creciente problema de resistencia a antibióticos, atacar los pili podría suponer una alternativa válida para impedir la adhesión bacteriana en primer lugar. En este sentido comprender las propiedades nanomecánicas de las proteínas que componen el pilus así como el proceso de biogénesis de este puede ser el comienzo para diseñar nuevos tratamientos que alteren la estabilidad mecánica de estos filamentos, o bien el proceso de ensamblaje y secreción del pilus en la superficie de la bacteria.

En el caso del VIH la adhesión se produce preferencialmente a los linfocitos T del sistema inmune, y para esto utiliza un receptor propio de la membrana del linfocito denominado CD4. La partícula vírica expone en su superficie un complejo proteico formado de dos glicoproteínas que median tanto la adhesión como la posterior fusión del virus con la célula. Una de estas proteínas se une de manera específica pero no covalente a CD4, estableciendo el anclaje en la superficie de la célula. Posteriormente esta misma glicoproteína reconoce a un segundo receptor de membrana en la célula y tras su unión se desencadena el proceso que conduce a la fusión del virus con la célula.

Los esfuerzos realizados para encontrar una vacuna efectiva contra el VIH han sido frustrados debido a la naturaleza del virus. El VIH dispone de una serie de ventajas que le permiten eludir al sistema inmune, tales como la glicosilación de las dos únicas proteínas que expone en su superficie así como de su alta tasa de mutación, lo que se traduce en la rápida generación de variantes de estas proteínas. Esto

convierte a la lucha del sistema inmune contra este patógeno en una carrera de fondo en la que el virus es siempre más rápido. Además la habilidad para integrarse en el genoma de las células y permanecer escondido en estas lo hace indetectable hasta que vuelve a reaparecer para infectar nuevas células. La interacción con el receptor CD4 es clave y es la primera que conduce a una infección exitosa. El posterior proceso de unión a un segundo receptor en la membrana del linfocito supone el acercamiento del virus a la membrana mientras está unido a CD4, lo que implicaría que CD4 debería ser una molécula flexible que permitiera esta aproximación. Diversos cambios estructurales y químicos ocurren en las moléculas involucradas en la infección, y las fuerzas mecánicas de nuevo no están exentas en este proceso. Una partícula vírica unida al receptor CD4 está sometida a movimiento Browniano, de manera que el virus podría ejercer fuerzas en CD4 que le permitirían estirar a esta molécula para así acercarse al segundo receptor de membrana. Obviamente el rango de fuerzas en las que esto se podría producir tiene que ser muy bajo, de lo contrario el virus podría separarse de CD4 y alejarse de la superficie celular. En este sentido un delicado equilibrio de fuerzas podría inducir la extensión y la contracción mecánicamente inducida de CD4, pero sin comprometer la unión del virus. Aunque las fuerzas aquí tengan un papel menos conspicuo que en el caso de la infección del tracto urinario, también pueden ser decisivas y por lo tanto soluciones a la infección por el VIH pueden ser encontradas desde un abordaje mecanofarmacológico. Comprender la nanomecánica de CD4 así como el efecto de ciertos factores que se conoce que afectan a esta molécula, podría ser el primer paso para diseñar terapias orientadas a combatir al VIH desde el primer paso de la infección. Bloquear el proceso de infección en este punto

supondría evitar la infección por completo. Aunque no supusiera una cura definitiva para aquellas personas que padecen la infección, limitaría su propagación en estos pacientes y además podría ser utilizado como medida profiláctica para prevenir nuevas infecciones.

Conociendo este contexto nuestro planteamiento ha sido abordar estos dos procesos de adhesión desde un punto de vista mecánico, explorando cómo las fuerzas y ciertas modificaciones químicas afectan mecánicamente a las proteínas involucradas. Para ello en esta tesis se ha empleado la espectroscopia de fuerza atómica para aplicar fuerzas a moléculas únicas y así poder determinar su comportamiento mecánico bajo diferentes condiciones.

En primer lugar nuestro objetivo fue determinar la estabilidad mecánica de las proteínas que componen el pilus tipo I de *E. coli* uropatogénica. Se diseñaron proteínas quiméricas en las que se preservó la interacción nativa que mantiene a cada proteína del pilus unida a su predecesora mediante la creación de permutantes circulares. La espectroscopia de fuerza en este caso nos permite aplicar fuerzas y explorar la resistencia mecánica de estas proteínas y de sus interacciones recreando la misma geometría que tendría lugar en el proceso *in vivo* cuando una bacteria es empujada por el flujo de orina y el pilus anclado al epitelio es estirado longitudinalmente. Nuestros experimentos de fuerza han demostrado la altísima resistencia mecánica de estas proteínas, cuyos valores van desde los 500 pN para la proteína que está más representada en el pilus tipo I, hasta los 100 pN en la última proteína del pilus que media la adhesión al epitelio urinario. De manera muy significativa observamos que la estabilidad mecánica de estas proteínas es jerárquica y está relacionada con la arquitectura del

pilus, siendo las proteínas más débiles a medida que se aproxima el final del pilus. Además detectamos el efecto mecánico que tiene un puente disulfuro muy conservado en estas proteínas, cuya presencia incrementa sustancialmente la estabilidad de estas proteínas. Esta parte de la investigación verificó la enorme estabilidad mecánica de estas proteínas, de lo resistente que es la interacción no covalente entre proteínas del pilus y de lo importante que es la presencia del puente disulfuro en cada una de ellas para garantizar la integridad mecánica de cada una de las proteínas.

En la segunda parte de esta investigación decidimos estudiar el proceso de biogénesis del pilus tipo I. Estudiamos el plegamiento y maduración de una de las proteínas del pilus tipo I a nivel de molécula única y en presencia de los factores que la ayudan en este proceso. La espectroscopia de fuerza en este caso es útil para recrear las condiciones de desplegamiento y estiramiento que tiene la proteína cuando está siendo exportada al periplasma bacteriano, compartimento desde donde estas proteínas son incorporadas y secretadas durante el ensamblaje del pilus. Nuestros primeros hallazgos en presencia de la chaperona que ayuda a estas proteínas a plegarse y a ser incorporadas en el pilus, demostraron que ésta solo actuaba sobre aquellas proteínas que presentaran su puente disulfuro ya formado, discriminando a aquellas que carecen de este enlace. Teniendo en cuenta la diferencia en la estabilidad mecánica de estas proteínas dependiendo de si poseen o no el puente disulfuro formado, este mecanismo de selección por parte de la chaperona garantizaría la incorporación al pilus de únicamente proteínas que presentaran el puente disulfuro y que por lo tanto tuvieran una alta resistencia mecánica. El siguiente descubrimiento que hicimos fue que, en presencia de la enzima oxidorreductasa que induce

precisamente la formación del puente disulfuro, el plegamiento de esta proteína fue incluso mayor que con la chaperona. Este dato arroja una nueva perspectiva al proceso de biogénesis del pilus pues hasta ahora se creía que la chaperona era la única encargada de plegar a las proteínas del pilus y que la oxidorreductasa solo inducía la formación del puente disulfuro. De nuestra investigación concluimos que la oxidorreductasa se encarga de ambas cosas y que entrega a la chaperona proteínas prácticamente plegadas y listas para su incorporación en un pilus de nueva generación. En este escenario la chaperona se encargaría de estabilizar el plegamiento de las proteínas que la oxidorreductasa le proporciona, evitaría procesos de agregamiento o polimerización temprana en el periplasma y escoltaría a las proteínas del pilus hasta la plataforma de ensamblaje del pilus tipo I en la membrana externa bacteriana.

Desde una perspectiva farmacológica, atacar la integridad estructural del pilus o atacar a los factores encargados de la maduración de sus proteínas podrían ser el camino para combatir las infecciones del tracto urinario ocasionadas por *E. coli*.

En la segunda parte de esta tesis decidimos explorar la nanomecánica de los dos primeros dominios del receptor CD4, CD4D1D2. Obtener información sobre las fuerzas a las que despliegan estos dos dominios supone un primer avance para comprender si es factible que el VIH pueda inducir su extensión mecánica. Es por esto que caracterizamos el desplegamiento de estos dos dominios aplicando rangos de fuerzas y velocidades de estiramiento que fueran representativos del contexto fisiológico en el que se produce la infección. Una partícula vírica unida a CD4 puede ejercer fuerza sobre

este pero en ningún caso puede alcanzar los niveles de estrés mecánico a los que se tienen que enfrentar por ejemplo las proteínas del pilus bacteriano. Cuando sometimos a CD4D1D2 a los diferentes protocolos de aplicación de fuerza observamos siempre una jerarquía en el orden de desplegamiento de estos dos dominios, desplegando siempre primero el domino D2 seguido a continuación del dominio D1. Sorprendentemente las fuerzas de desplegamiento de D2 son mayores que las de D1, y sin embargo despliega primero al contrario de lo esperable. Esta característica se reprodujo en cualquiera de las condiciones testadas y la justificamos en base a la estructura de estos dos dominios, que comparten una hebra β. Esto supone que cuando se les somete a fuerza el dominio D2 despliega y deja desprotegido al dominio D1, cuya estabilidad mecánica se ve comprometida una vez la fuerza es transmitida a través de este elemento estructural compartido.

El desarrollo de un modelo matemático que relaciona física de polímeros con grado de infectividad por VIH nos permitió extrapolar, en base a unos experimentos previos que utilizaron variantes de distinta longitud de CD4, que la posibilidad de que algunos dominios de CD4 se desplieguen durante la unión del virus es factible. A mayor longitud de CD4 mayor es la infectividad por VIH, lo que se relaciona con la flexibilidad de CD4 y con la capacidad del virus de explorar áreas más grandes de la superficie celular en busca del coreceptor que necesita para perpetrar la infección de la célula.

Nuestro siguiente paso fue estudiar el efecto que un anticuerpo denominado Ibalizumab, conocido por su potente efecto inhibidor de la infección por VIH, sobre la nanomecánica de CD4D1D2. La mayor parte de los contactos intermoleculares que se establecen entre este

anticuerpo y CD4 se localizan en el dominio D2, y se ha demostrado que las interacciones proteína-proteína pueden tener un efecto positivo en la estabilidad mecánica de las moléculas involucradas. Dado que el origen del efecto inhibitorio de Ibalizumab no ha sido hallado nosotros hemos planteado la posibilidad de que este anticuerpo pueda volver rígido a CD4 tras su unión, reduciendo su flexibilidad y quizás impidiendo que CD4 pueda ser extendido mecánicamente por la partícula vírica. Descubrimos que efectivamente Ibalizumab altera la nanomecánica de CD4 y que las fuerzas requeridas para su desplegamiento son mayores, sugiriendo que este efecto estabilizador puede ser el responsable de la inhibición de la infección y apoyando nuestra hipótesis de que el desplegamiento de CD4 facilita los siguientes eventos previos a la fusión del virus a la célula.

Finalmente quisimos estudiar el efecto que la oxidorreductasa tioredoxina puede tener en la regulación de los puentes disulfuros de CD4. Esta y otras enzimas encargadas de la regulación tiol-disulfuro son secretadas por la célula y afectan al estado de oxidación de los puentes disulfuro de CD4 y otras proteínas. De hecho las reducciones de puentes disulfuro del dominio D2 de CD4 y de las proteínas superficiales del VIH han sido señaladas como prerrequisitos necesarios para la consecución de la infección. Estos enlaces covalentes limitan la extensión mecánica de los dominios de CD4, y sólo la reducción química de estos por enzimas podría permitir mayores extensiones de CD4. Nuestros experimentos de fuerza sobre CD4D1D2 en presencia de tioredoxina demostraron que los puentes disulfuros de estos dos dominios solo son reducidos después de haber sido estirados mecánicamente. En ambos dominios estos enlaces están ocultos en el interior hidrofóbico de sus estructuras, de manera que solo después del

desplegamiento de cada dominio se hacen accesibles para la tioredoxina. Dado que la reducción del puente disulfuro de al menos del dominio D2 parece necesaria para que se produzca la infección, nosotros planteamos que la partícula del VIH puede inducir el desplegamiento mecánico de algunos dominios de CD4 y este hecho puede hacer que la tioredoxina pueda acceder a los puentes disulfuros de estos dominios, reduciéndolos, y permitiendo que el virus pueda extender incluso más a CD4, mejorando sus oportunidades de encontrar al coreceptor de membrana.

Con este conjunto de factores vemos el comportamiento del receptor CD4 como el de un amortiguador, que se expande y se contrae debido a los movimientos del virus que tiene unido en su extremo. El VIH tendría la ventaja de, permaneciendo unido, ejercer fuerzas bajas que desplieguen y plieguen a los dominios de CD4 mientras se produce el encuentro con el coreceptor de membrana que finalmente conduce a la fusión del virus. Junto con los cambios químicos que tienen lugar en CD4, esto puede facilitar su flexibilidad e incrementar las probabilidades de encontrar a su coreceptor.

A la luz de los datos de estas investigaciones proponemos a la investigación mecanofarmacológica como una estrategia de futuro para tratar enfermedades cuyo origen yace en el malfuncionamiento de un componente mecánico. En esta tesis se ha estudiado el proceso de adhesión de patógenos y creemos que esta parte de la infección donde las fuerzas mecánicas juegan un papel relevante puede ser atacada mecánicamente. En el caso de la infección del tracto urinario debilitar sus estructuras de adhesión o deteriorando las propiedades biomecánicas de estas puede ser el camino para combatir esta

enfermedad. En el caso de la adhesión del VIH, hacer más estable al receptor CD4 e impedir su extensión mecanoquímica puede ser el punto de partida para estudiar si esta opción tiene relevancia a nivel fisiológico y puede convertirse en un nuevo tratamiento efectivo para prevenir la infección, tanto de nuevas células como de nuevos individuos.

Abstract

The role of mechanical force in physiology is becoming increasingly recognized as an important factor in the human health. Mechanical force affects the function, the organization and the remodeling of organs and tissues in the human body, which are processes triggered by events occurring at the single-molecule level. In the same way mechanical forces are actively present during disease, and their involvement is calling the attention of the scientific and medical communities. Among diseases, the pathogenic infections are not exempt of mechanical forces, especially during the pathogen attachment to human tissues where surface proteins from both the pathogen and the host interact with each other. Currently, the traditional treatments against infections are facing the challenge of the bacterial antibiotic resistance and the lack of a cure for some pathogenic diseases of worldwide spread. In this sense, the pharmacological targeting of the mechanical attachment of the pathogen at the single-molecule level is seen to be a new approach to fight infections. This new point of view enables the design and discovery of new drugs, and the new fields of Mechanomedicine and Mechanopharmacology provide a suitable framework for the development of new treatments against pathogenic attachment. Although the last years have witnessed an increasing knowledge regarding single-molecule mechanics, little is known about the mechanical features of the proteins used by pathogens for attachment. Hence, basic research is needed to unravel the mechanics of these molecules and to provide a basis from where a pharmacological treatment can be developed.

Abstract

With this goal, my thesis has focused on the study of the nanomechanics of cell-surface proteins involved in viral and bacterial adhesion. Specifically I have studied the proteins used by uropathogenic *Escherichia coli* to attach to the urinary tract, and the human protein receptor CD4 used by Human Immunodeficiency Virus (HIV) for recognizing and infecting immune cells.

We hypothesized that bacterial proteins display a high mechanical resistance based on their adhesion function in the urinary tract. For the CD4 protein it was hypothesized that its mechanical extension is related with the Human Immunodeficiency Virus infection since this process requires the spatial rearrangement of many elements for bringing the virus closer to the cell membrane. Many disciplines and techniques are involved in this thesis but the most important technique applied for the direct manipulation of single protein molecules is single-molecule force spectroscopy.

In the first part of this thesis the mechanical resistance of the proteins that build the attachment organelle of uropahogenic *Escherichia coli* was studied. This extracellular structure is called type I pilus, a filamentous appendage composed of hundreds of proteins used during urinary tract infection. We found that pilus proteins show not only a remarkable resistance to mechanical unfolding but also that they follow an unfolding hierarchy pattern tightly connected to their position in the pilus structure. Our results reveal that conserved disulfide bonds in each of the pilus proteins are a crucial feature for their mechanical stability. Besides unraveling its nanomechanical architecture, the biogenesis of the pilus was studied wherein we recreated *in vitro* the same events that take place during the maturation

and folding of the proteins previous to their incorporation into the pilus. During our investigation we found that the enzyme DsbA involved in disulfide bond creation during the first step of pilus proteins maturation shows an additional pronounced chaperone effect, which helps to fold to pilus proteins in a greater extent than the putative real chaperone of this system. With these findings we provide for the first time a detailed description at the single-molecule level of the type I pilus mechanical architecture and biogenesis.

In the second part of this work the nanomechanics of the lymphocyte T cell-surface protein CD4 were studied. The CD4 protein is used by the Human Immunodeficiency Virus particle to attach to these cells. The subsequent events for viral infection require the approaching of the particle to the cell membrane to interact with other proteins meanwhile attached to CD4. We have discovered that the first two domains of CD4, D1D2, are able to unfold at low forces. We suggest that this is the range of forces that a viral particle could produce due to thermal motion. In this scenario CD4 would act as a shock absorber in which its domains can unfold at low forces by the virus, therefore increasing the length of the tether and allowing the virus to explore the cell membrane looking for the other proteins required for viral fusion. Also the regulation by oxidoreductase enzymes of the disulfide bonds of CD4 during HIV infection is important for the infectivity of the virus. We demonstrated that only the mechanical unfolding of CD4D1D2 makes possible the access of the oxidoreductase enzyme to the buried disulfide bonds, supporting the idea that the viral particle could unfold CD4 *in vivo*. A mathematical model was developed based on the infectivity of cells expressing engineered variants of CD4 showing different tether lengths. The

infectivity increases with the length of the tether, and the model allows interpreting that the *in vivo* infection would require the extension of CD4. Finally the effect of Ibalizumab, an antibody known to be a good neutralizer of HIV infection was tested. This antibody's epitope is placed between CD4D1 and CD4D2 domains and the exact mechanism for its potent neutralizing effect is unknown. We have found that the mechanical stability of CD4D1D2 is increased in the presence of Ibalizumab, suggesting that the binding of the antibody hinders the extension of CD4 by the virus and this avoids the subsequent events of the infection. This supports again that the mechanical extension of CD4 is mandatory for successful virus infection. Altogether our results suggest that CD4 mechanochemical extension may play a critical role during HIV infection.

In the light of these findings, we think that our results could help to develop new therapeutic strategies oriented to alter the nanomechanics of proteins involved in pathogenic attachment, in the context of the new fields of Mechanopharmacology and Mechanomedicine. In the case of bacteria, the use of molecules which disrupt the strength of the pilus proteins would circumvent the current limitations of traditional antibiotic treatments. For HIV, we suggest that an increased mechanical stabilization of CD4 receptor induced by antibody binding could prevent the infection or diminish its chances of success.

Chapter 1: Introduction

In this first chapter the impact of pathogenic infections in the human health are addressed, with a special mention of the two issues covered in this thesis: uropathogenic *Escherichia coli* and HIV-1 attachment. Several examples of how forces affect different physiological processes are presented, and placed in the context of how some diseases have a mechanical malfunction origin that lies at the molecular level. The concepts of Mechanopharmacology and Mechanomedicine are also presented in this chapter as a potential tool for approaching these pathologies.

1.1 Human pathogenic infections

Pathogen infections are one of the main causes of morbidity and mortality in human populations worldwide. In 2015 the third cause of death was caused by lower respiratory infections, accounting for 3.2 million deaths. Diarrheal diseases and tuberculosis produced each 1.4 million deaths the same year. Although the mortality has been reduced during the last years, pathogen infections remain as one of the main public health issues[1].

Among the pathogens affecting the human health, the bacterial infections are increasingly becoming of public concern because of the appearance of antibiotic resistant strains (**Fig. 1.1**). The amount of deaths related with antimicrobial resistance (AMR) in the European Union (EU) is estimated to be about 25,000 per year, meanwhile in the

United States of America (US) it is about 23,000 deaths per year. In low and middle income countries the statistics are not so clear, but taking as an example Thailand the numbers are 3 to 5 times larger than in the EU or the US, although its population is 5 times smaller[2].

Urinary tract infections (UTIs) represent one of the most common bacterial infections, causing from mild cystitis to a severe uroseptic shock[3]. Despite of both women and men can be infected, traditionally UTIs have been considered a female disease since approximately 50 % of the women will suffer this infection at least once along their lives. Moreover, the 25 % of women that experience a cystitis episode will experience recurrent infections (rUTI) in the next months after the first outbreak[4]. Several microorganisms can produce UTIs (*Proteus* spp., *Staphylococcus saprophyticus, Klebsiella* spp., *Candida* spp., *Enterococcus* spp., etc.) but the main one is the uropathogenic *Escherichia coli* (UPEC)[3].

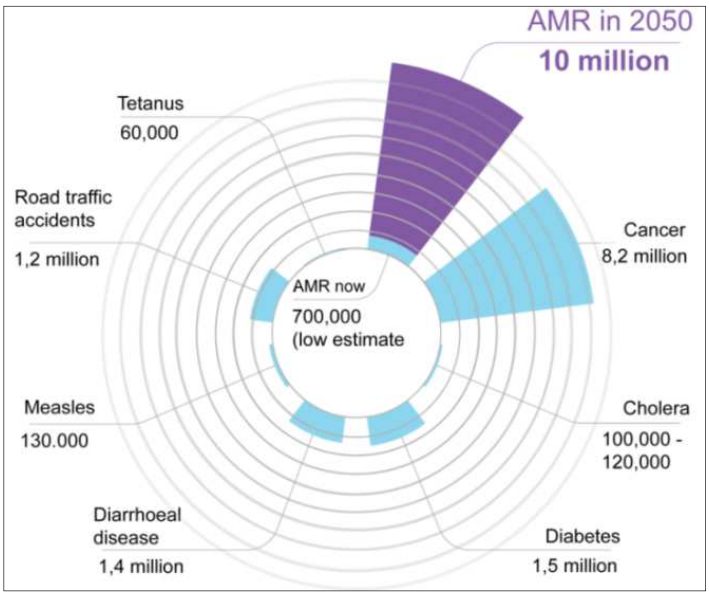


Figure 1.1. World year estimation of deaths caused by several diseases. It is estimated that in 2050 the number of deaths related with AMR will increase from the current 700,000 to 10 million. Graph adapted from reference[5].

Besides the threat posed by resistant bacterial pathogens, the viral infections are also the focus of the international medical community. As an example, the Influenza virus is annually responsible of 3-5 million severe cases and around 250,000 - 500,000 deaths[6]. The recent Ebola virus outbreak in West Africa caused more than 11,000 deaths creating an emergency situation which the international community failed to contain, due to the rapid spread and the high virulence of the virus and the precarious health systems of the affected countries[7].

Despite of these viruses calling the attention of the public periodically, Human Immunodeficiency Virus (HIV) highlights above them because of its worldwide spread and the absence of a cure for it. Currently it is not anymore included in the top 10 cause of death list elaborated by the World Health Organization, although in 2016 caused 1 million deaths (**Fig. 1.2**). In the same year 1.8 million people got infected and it is estimated that around 36.7 million live currently with the infection[8]. Despite of its broad distribution approximately 70% of the people living with the infection are in Sub-Saharan Africa, being the first cause of death in this region of the planet[9].

HIV produces the Acquired Immunodeficiency Disease Syndrome (AIDS), a condition based on the destruction of the immune system of the host which makes the patient more susceptible to infections, cancers and other diseases easily fought by people with a healthy immune system[10]. Huge efforts have been done in order to find a cure for this infection; however because of the nature of the virus it has been elusive to develop a successful treatment. Currently the antiretroviral therapy has proved to be successful converting a deadly

disease into a chronic one, but the increase in the life expectancy of infected people has increased the chances and the number of new infections too[11].

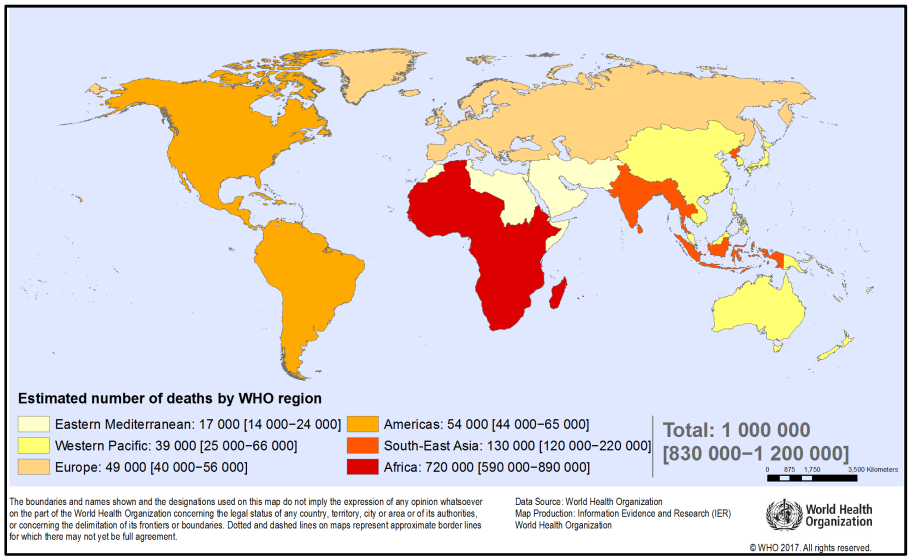


Figure 1.2. Estimated number of people dying from HIV-related causes, 2016 by World Health Organization (WHO) region. Adapted from reference[12].

1.2 The pathogen attachment as a therapeutic target

The medical treatments employed for fighting back pathogenic infections are focused on interfering with the life cycle of the pathogen. For example antibiotics interfere with the ability of bacteria to synthesize proteins, DNA, cell wall components, etc[13]. The main antiretroviral drugs used against HIV infection block the replication of the virus at different stages of its life cycle[11]. In the case of bacteria the antibiotics are becoming useless against resistant strains. There is no

cure for HIV, its ability to integrate into the human DNA and hiding in latent cellular reservoirs makes its eradication impossible[14].

Independently of the differences of each case, in both the infection starts once the bacteria cells or viral particles establish contact with the surface of the target cells. This contact implies the recognition of specific features of the cell surface which serve as an anchor for viral and bacterial proteins used for the attachment. A reasonable approach to treat these diseases could be to interfere and target the proteins used by the pathogens to attach to human cells[15,16]. Disrupting the first step of the infection would avoid the infection itself, reducing the risks to the health and also reducing the costs derived from medical treatments.

1.3 Forces at the nanoscale in biological systems

1.3.1 Mechanobiology and mechanotransduction

Avoiding the successful attachment of the pathogen is a promising target for new therapeutic approaches, however it is important also to broaden the way biological processes are studied and seen. Until recently the importance of mechanical forces on many biological processes which involve from single molecules to organs has received little attention. Tissues like bone or skin are exposed to different mechanical stresses, meanwhile others are able to exert forces like the heart in every beat or the skeletal muscles every time we walk. In many biological events mechanical forces are present, and viral and bacterial attachment is not an exception.

Chapter 1

The study of how physical forces affect cells and tissues and how these processes are related to certain diseases are questions approached by Mechanobiology, a field at the edge of Physics, Biology and Bioengineering[17]. Individual cells in tissues adopt their correct morphology through the modification of their own shape, but they are also affecting the development of their surroundings including other cells. These processes implicate both the generation and the detection of physical disturbances in and from the surrounding media. For example the T cells of the immune system are exposed to several mechanical stresses along their lifetime, from the rigid substrates of the thymus to the fluid flow of the blood circulation[18]. Cells mechanically sense their microenvironment through protein complexes, being able to respond to the mechanical information with changes at the molecular level, a process known as mechanotransduction[17,19,20]. These mechanical stimuli sensed by the cell have a response that can be detected at the tissue level like bone grow and remodeling, wound healing or blood clotting. The mechanosensing of the environment and the mechanotransduction of signals encompasses feedback interactions among external structures like the extracellular matrix (ECM), surface proteins like integrins, and internal complexes like the cytoskeleton[17].

Many proteins are involved in processes of force generation and resistance. For example cadherins are transmembrane proteins involved in the adhesion between cells, contributing to the stability of both epithelial and non-epithelial tissues. These cadherins interact with cytoplasmic and cytoskeletal proteins, connecting the cytoskeleton of adjacent cells through specialized interactions named adherens junctions[20,21]. Another interaction related with force detection and generation are the focal adhesions, which are highly dynamic

complexes involved in cell motility and adhesion which connect the cell's cytoskeleton with the ECM through integrins. These are only a few examples of force-bearing and force-producing complexes of the cell, where a number of proteins intervene not only as mechanical scaffolds but also as chemical and force effectors. An interesting issue is the elucidation of how the mechanical stimuli are sensed by cells. Proteins are the effector molecules of the cells, so the proteins involved in sensing and/or transducing physical signals could experience conformational changes upon mechanical stress application.

1.3.2 Protein unfolding under physiological conditions

The mechanotransduction of signals implies conformational changes of the involved effector proteins, which could include their partial or complete mechanical unfolding. In the above mentioned focal adhesions the linkage between the cytoskeleton and the membrane integrins is made by a protein called talin. This protein recruits another protein called vinculin for the stabilization of the focal adhesion. However the binding sites for vinculin are hidden in the folded talin. In a interesting set of single molecule experiments it was demonstrated that the mechanical unfolding of talin exposed these cryptic binding sites allowing for the recruitment of vinculin (**Fig. 1.3**). The *in vivo* stabilization of focal adhesions through vinculin binding to mechanically unfolded talin could be triggered by the contraction of the actin cytoskeleton**,** exerting forces in the range of the applied ones in these experiments[21–23]. The total or partial mechanical unfolding of ECM proteins such as collagen, fibronectin or laminin is related with cellular responses such as cell differentiation, proliferation and

angiogenesis[24]. The relevance of protein unfolding *in vivo* has been also reported in several cellular processes like the translocation of proteins through membranes or the protein degradation by ATP-dependent proteases[25].

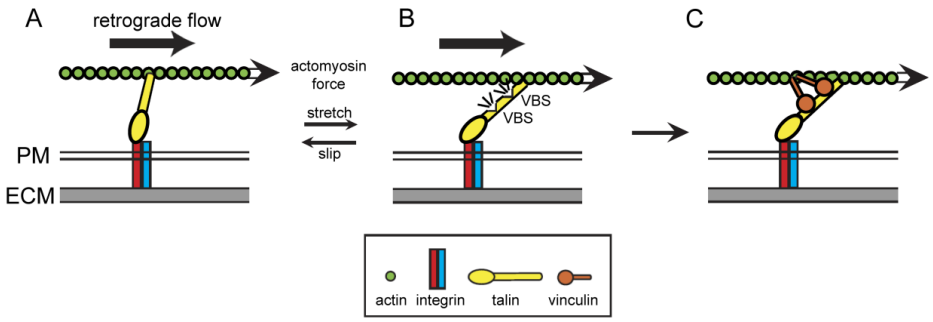


Figure 1.3. Model for force-dependent regulation of vinculin recruitment at focal adhesions and anchoring the actin network. (**A**) In the focal adhesion the ECM is connected to the cytoskeleton through transmembrane proteins called integrins. In the cytoplasmic side of integrins, a protein called talin bridges them with the actin filaments. (**B**) The contraction of the actomyosin cytoskeleton triggers the unfolding of the talin protein bound to the integrins and the cytoskeleton. (**C**) The mechanical unfolding of talin exposes its vinculin's binding sites (VBSs) recruiting vinculing molecules. Adapted from reference[26].

One of the most studied molecules from the mechanical point of view has been titin, the giant muscle protein. Titin is the biggest known protein of the human proteome, being composed of hundreds of domains that connect the M-line with the Z-disc in the sarcomere, the contraction unit of the skeletal muscle (**Fig. 1.4**).

Single molecule techniques have been extensively used for unraveling the mechanical properties of this protein and its implication during muscle stretching[27–34]. The findings done along these last years support the idea that some of the domains of this protein could unfold during muscle stretching, although this is a source of big debate[33,35]. These unfoldings would expose cryptic sites in titin susceptible of

modification[36,37], and titin would also contribute not only passively in the elasticity of the muscle but also generating a restoring force during the refolding of its domains[38,39].

These examples in which protein mechanical unfolding occurs as part of the normal life cycle of tissues and cells, strongly support the need and the importance of studying the effect of forces in biological processes.

1.3.3 Mechanomedicine, a mechanical approach to diseases

All these findings highlight that protein mechanical unfolding is part of the normal physiological performance. But in the same way, the malfunctioning of some of these proteins is directly linked to some pathologies in which physical forces play a crucial role. Many human diseases originate as a consequence of alterations in the mechanotransduction of cells and tissues and the alteration of the mechanical properties of the effector proteins responsible of these processes. For example changes in the composition, stiffness and architecture of the ECM is found in cancer tissues, increasing the proliferation and survival of tumor cells. In another example, abnormal levels of the protein talin are observed in several types of cancers. Another diseases in which these changes are relevant are fibrosis, asthma, atherosclerosis, heart disease among others[17,23,40]. Since forces are an important factor in disease, understanding the mechanics of the proteins involved in these pathologies could be an alternative strategy for the design of new pharmacological treatments.

A multidisciplinary framework where Biology, Chemistry, Physics and Medicine meet is mandatory for investigating these force-related diseases. Mechanomedicine and Mechanopharmacology appear as new fields for approaching pathologies where an altered mechanical feature leads to disease. But not only from the research point of view, Mechanomedicine and Mechanopharmacology could provide tools for the development of new therapies. The design of drugs able to alter the mechanical behavior of proteins involved in these pathologies is part of the issues covered by this field. On the one hand it is known that the binding of some molecules like antibodies to protein epitopes leads to the mechanical stabilization of the last[41,42]. On the other hand posttranslational modifications can weaken the mechanical stability of some proteins like titin[36], cytoskeletal structures like the intermediate filaments[43], and hybrid protein-DNA structures like nucleosomes[44]. Depending on the specific mechanical origin of the problem it is possible to adjust strategies oriented to treat the disease.

As in many diseases, the infective process involves more aspects than biochemical changes. The pathogen attachment to target cells is submitted to forces at the nanoscale, involving cell and pathogen surface single molecules. Protein unfolding under these circumstances may be a crucial factor for a successful infection, and the mechanical disruption or alteration of the proteins involved during the attachment process could be mechanically targeted as a pharmacological treatment.

With this hypothesis, this thesis is aimed to provide a description of the nanomechanics of the proteins used by UPEC to

attach to the urinary tract, and by HIV to bind to lymphocytes T-cells (**Fig. 1.4**).

In the next two chapters the main features of UPEC and HIV-1 biology and attachment will be presented. To unravel the mechanics underlying these processes at the molecular level, and to identify the strengths and weaknesses of the main actors mediating attachment, can constitute the starting point for the designing of new therapeutic strategies in the context of Mechanomedicine.

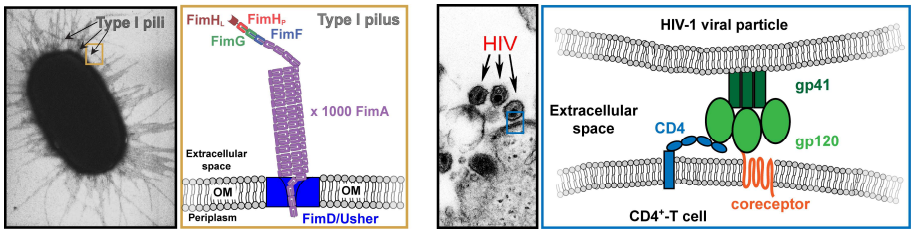


Figure 1.4. UPEC and HIV-1 attachment to target cells. (**Left side**) UPEC uses filamentous structures called pili or fimbriae to attach to epithelial cells from the urinary tract. These pili are made of hundreds of proteins subunits which are secreted from the outer membrane (OM) of the bacteria. *E. coli* image from Manu Forero (DOI:10.1371/journal.pbio.0040314). (**Right side**) HIV-1 particles have macromolecular structures on their surfaces which interact with lymphocyte T- cells surface receptors called CD4, in the first step of the infection. Image of HIV particles on the surface of a cell from Cynthia Goldsmith (https://www.pixnio.com/).

Chapter 2: Type I pilus

In this chapter UPEC type I pilus is presented, with a special focus on its architecture, mechanical properties, and its biogenesis. Before approaching the appendage features, a brief explanation of how UTIs develop is given in order to better understand the mechanical challenges that UPEC faces during uroepithelium attachment.

2.1 UPEC-UTI etiology

Escherichia coli is a Gram-negative bacterial organism which lives in the mucous layer of the mammalian colon. The human gastrointestinal tract is invaded by this bacteria a few hours after being born, establishing a long-term symbiotic relationship with its host[45,46]. The human intestinal microflora is composed of many microorganisms but *E. coli* constitutes the most abundant facultative anaerobe, being one of the most successful competitors in this niche. *E. coli* rarely causes diseases, unless the host is immunosuppressed or the gastrointestinal barriers are compromised.

Despite of the benefits of this relationship some *E. coli* clones have acquired virulence factors that confers them the ability of colonizing and thriving into new niches causing several diseases[46]. The virulence factors enable the bacteria to replicate and to disseminate inside of a host by means of subverting or eluding the immune system of the host[47]. These virulence factors are encoded in genetic elements that can be distributed among different bacteria strains through

horizontal gene transfer. This exchange of genetic material leads to the creation of new combinations of virulence factors that arm the bacteria with an array of resources that allow them to successfully develop an infection in the host. The most successful combinations of virulence factors lead to the appearance of pathotypes. The *E. coli* pathotypes are strains able to cause diseases to healthy hosts and there are three main clinical syndromes derived from the infection with these pathotypes: enteric or diarrhoeal disease, sepsis or meningitis, and UTIs. UPEC is the pathotype responsible for about 70-80 % of the UTIs[48–50], and it constitutes the most common extraintestinal infection[46].

Depending on the anatomical part of the urinary tract colonized by UPEC, UTIs are classified in lower and upper infections. Lower UTIs are confined to the bladder, producing cystitis. Cystitis could be asymptomatic and sometimes it can be resolved without medical treatment. When symptoms are present the main ones are: frequency, urgency, suprapubic pain, haematuria (red blood cells in the urine) and dysuria (pain during micturation)[49]. The upper UTI is known as pyelonephritis and it affects the kidney. Besides symptoms such as loin pain, haematuria and cystitis; the more general severe effects are fever, vomiting and septic shock. **Fig. 2.1** shows a scheme summarizing the main features of both cystitis and pyelonephritis. Other UTIs include asymptomatic bacteriuria (high count of bacterial cells in the urine), urethritis, and prostatitis[4,49].

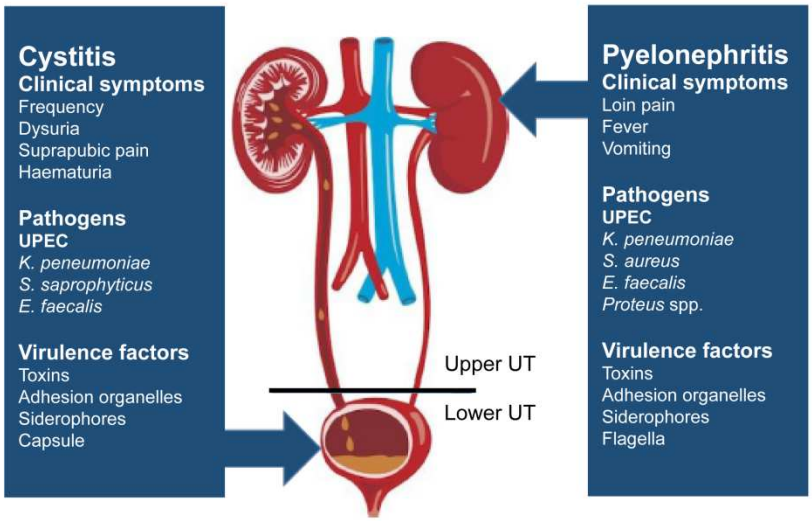


Figure 2.1. Clinical features, causative microorganisms, and virulence mechanisms in cystitis and pyelonephritis. Adapted from reference[4].

2.2 The onset of the infection

The UTI probably starts with the colonization of the periurethral area by an uropathogenic strain coming from the gastrointestinal tract. Then the strain starts its ascension through the urethra to the bladder. Anatomical differences between genders favor the infection in women, since the urethral opening is in the vicinity of the rectum and the shorter urethra facilitates the bladder invasion[4,49,50]. Once in the urinary tract, UPEC faces the difficult physical and chemical conditions imposed to any organism trying to colonize this niche. Any organism occupying the urinary tract is exposed to the harsh chemical environment created by the low pH and the high osmolality of urine. Moreover, the bladder is frequently emptied of urine generating a physical stress on any loose cell. The unidirectional flow of urine generates shear forces that challenge the ability of remaining attached

to the uroepithelium to any organism ascending from the urethra[51]. In order to invade and colonize the urinary tract, UPEC displays its virulence factors. Among these virulence factors it is possible to find several molecules secreted by UPEC like α-hemolysin, cytotoxic necrotizing factor 1 (CNF-1), and the autotransported protease Sat[46]. These virulence factors are toxins that target the uroepithelium, interfering with several cellular processes. But as it was noted before, UPEC faces the forces generated by urine flow standing against its ascending route to the bladder and the kidneys. This mechanical stress is countered by the presence of virulence factors specific for the attachment to the uroepithelium called pili.

2.3 Pili

The pili or fimbriae are extracellular filamentous appendages protruding from the surface of Gram-negative and positive bacteria and archaea. They are used for: adhesion, host recognition, biofilm formation, and secretion of proteins or DNA. In Gram-negative bacteria there are five different classes of pili that differ both in their function and in their assembly mechanism, being categorized according to this second feature[52]. These categories are: chaperone-usher (CU) pili, curli, type IV pili, type V pili, and type IV secretion pili[53]. Among these five assembly systems, UPEC uses the CU-pathway for pilus biogenesis. UPEC uses these CU generated pili for the specific recognition and attachment to certain sugars present in the uroepithelium (**Fig. 2.2**).

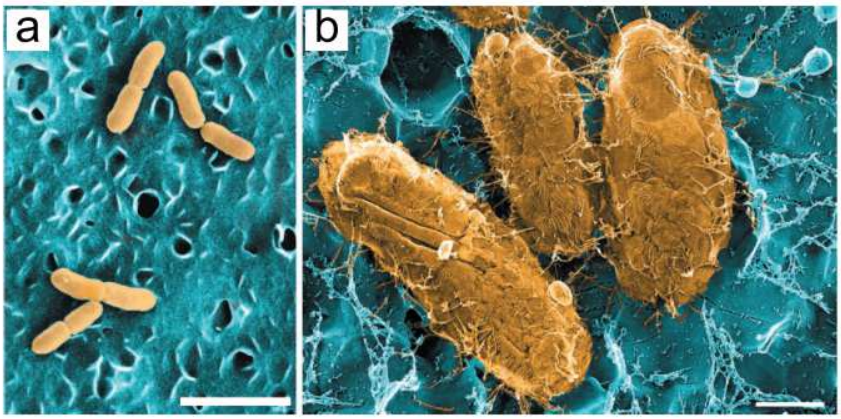


Figure 2.2. Type I pili UPEC attachment to the bladder epithelium of mice. (**a**) Scanning Electron Microscopy image of the bacteria cells (yellow) attached to the mice bladder (blue). Bar 3 µm. (**b**) Freeze-dry/deep-etch Electron Microscopy image where the type I pili are resolved as long filaments surrounding the bacteria surface and interacting with the bladder epithelium. Bar 0.5 µm. Adapted from reference[54].

UPEC is able to express two types of pilus depending on which anatomical part is attached to. In the bladder the type I pilus is expressed, meanwhile in the kidney the type P pilus is chosen. These systems are tightly regulated at the genetic expression level, and when a bacterium ascends from the bladder to the kidney the pilus type expressed on its surface changes from the type I to the type P. In both cases the pilus is made of hundreds of proteins that form a micrometer-size chain in whose tip is placed a lectin protein able to recognize and bind to sugars present in the uroepithelium. In this thesis the focus has been put on the type I pilus, the virulence factor used by UPEC for bladder invasion.

2.4 Type I pilus architecture

2.4.1 Pilus structure

The type I pilus is formed by four different proteins called Fims or pilins: FimA, FimF, FimG and FimH. These four subunits form a filamentous structure of up to ~2 µm of length and ~7 nm of diameter that can be divided in two parts: the pilus rod and the pilus tip.

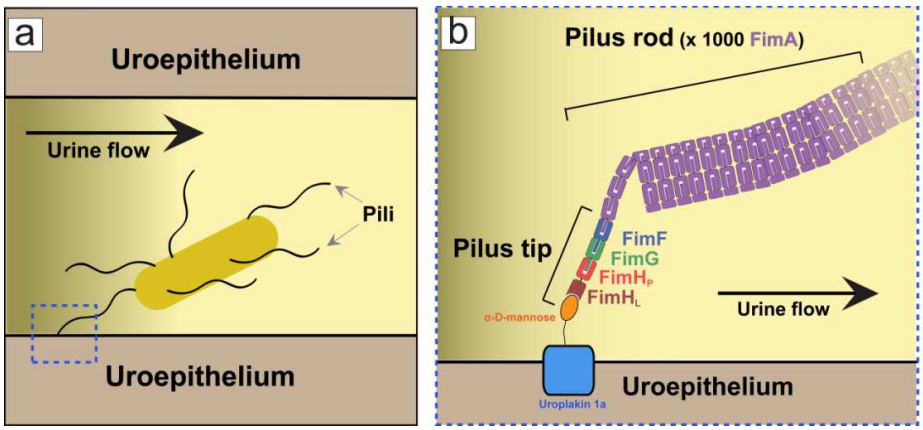


Figure 2.3. UPEC attachment to the uroepithelium with type I pili. (**a**) UPEC cell attached to the uroepithelium through one pilus type I (blue dashed square) meanwhile being pushed by the urine flow. (**b**) Close look to the blue dashed square highlighted in (**a**). Type I pilus is composed of the pilus rod made of hundreds of FimA subunits, and the pilus tip made of the array FimF-FimG-FimH. The lectin domain of FimH ($FimH_L$) recognizes and binds to mannosylated glycoproteins present in the epithelium such as uroplakin 1a.

The vast majority of the pilus is composed by hundreds of copies of the protein FimA, constituting the pilus rod. In the pilus rod the chain of FimA subunits adopt a quaternary structure that generates a right-handed helix configuration, a feature that confers to this rod a spring-like behavior. At the distal end of the pilus rod the pilus tip composed by the array FimF-FimG-FimH is placed (**Fig. 2.3**). The distal end protein FimH has a lectin domain in its N-terminal part that

recognizes and binds to mannosylated glycoproteins present in the uroepithelium[55,56].

2.4.2 Subunit association in the pilus chain

The Fim proteins are very similar between each other, being incomplete immunoglobulin (Ig)-like domains which lack the C-terminal seventh (or G) β-strand. FimA, FimF and FimG have an extra N-terminal β-strand made of 10-20 amino acids which task is to complement the previous domain in the pilus. FimH slightly differs from the rest because it is composed of two domains: the C-terminal domain shares the same incomplete Ig-like structure of the other pilins (termed here as $FimH_P$) meanwhile the N-terminal one is a lectin domain (termed as $FimH_L$) involved in the adhesion to the uroepithelium. The absence of the C-terminal β-strand generates a deep hydrophobic groove in each pilin domain that makes them unstable, and only a β-strand donated by another subunit or by a chaperone can stabilize these protein domains[57].

Hence, in the type I pilus the subunits associate head to tail to each other forming a chain, where each Fim subunit donates its N-terminal β-strand to the incomplete fold of the previous one. This connection between subunits is known as donor-strand interaction. In the pilus rod, each FimA donates to another FimA. In the pilus tip, FimF receives the β-strand from the last FimA subunit of the pilus rod. Then FimF donates its β-strand to FimG (**Fig. 2.4**), and finally FimG donates to the pilin domain of FimH[57].

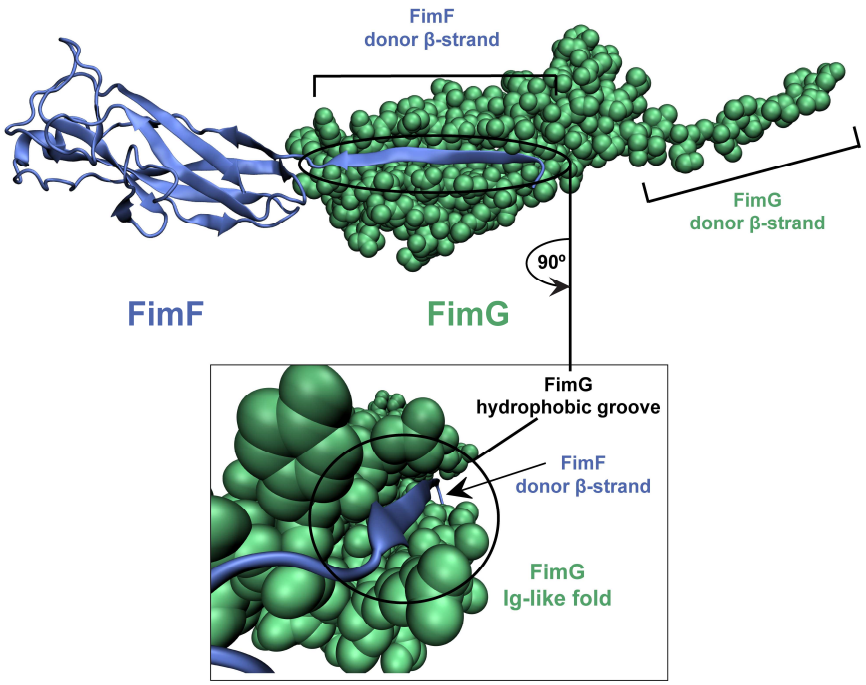


Figure 2.4. Donor-strand interaction between FimF and FimG. FimF N-terminal β-strand complements the fold of FimG. Pilin proteins lack the 7th β-strand causing a deep hydrophobic groove in its fold that makes them unstable. Lower panel shows the insertion of FimF β-strand in the hydrophobic groove of FimG from a longitudinal point of view. FimF displayed as a cartoon representation, FimG atoms shown as van der Waals spheres (PDB 4J3O).

This feature defines one of the most remarkable qualities of the type I pilus, since this whole filament is assembled through non-covalent interactions between hundreds of proteins. It is surprising that the integrity of the structure used by UPEC to attach to the urinary tract relies on apparently weak interactions, precisely when in this environment they are frequently challenged by the shear forces generated by the urine flow.

2.4.3 Pilus mechanical features

Force-induced transition from low to high affinity for mannose binding: the catch-bond

The type I pilus properties have been extensively studied. As a virulence factor it attracts the attention of the medical community since it can be a potential target for drug therapy instead of the traditional treatments with antibiotics. A good alternative approach could be to disrupt these structures used for attachment and colonization of the urinary tract. Since the beginning the focus was put on the receptor-ligand interaction of the lectin domain of FimH with the sugars present in the uroepithelium. FimH binds to mannose-containing glycoproteins present on the surface of the bladder cells, like the protein uroplakin 1a[4,58,59]. The recognition and binding of FimH to the glycoprotein D-mannose has been deeply documented, studied and submitted to research focused on interfering with this bond[58,60–66]. It is common to think that when a pulling force is applied to separate two molecules (here FimH and mannose), the probability of breaking this bond will increase with the force applied[67]. These are known as slip-bonds, but the interaction between FimH and D-mannose is a special type of bond known as a catch-bond. In catch-bonds the strength of the interaction increases with force: FimH changes from a low to a high affinity state for D-mannose when a tensile force is applied to the bond[68].

The attachment of UPEC to the urinary tract is optimized for avoiding being washed out during micturition but also it is optimized for maximizing the spreading of the infection and the colonization of the ascending route. In the absence of flow this loose attachment permits the bacterial cells to roll over the epithelium establishing new

infection focus. When the shear force induced by the urine flow push the cells, the bacteria pili that act as tethers are stretched, and finally this stretching force is transmitted to the bond between FimH and D-mannose. Under this last circumstance, a spatial rearrangement induced by force between the lectin and the pilin domains of FimH increases the affinity for the sugar. The pilin domain acts as a negative allosteric regulator of the lectin domain bond with mannose, and tensile force separates both domains causing the lectin to clamp tightly the mannose ligand yielding a long-lived and resistant interaction (**Fig. 2.5**). In the absence of force, in the pilus tip both domains adopt a hooked shape stabilized by weak inter-domain interactions. On the contrary FimG and FimF do not have these inter-domain interactions neither between them neither with FimH pilin domain and they act as a highly flexible hinge region between the rigid pilus rod and FimH, allowing for this hook conformation of the tip to explore the epithelium[64,69–72].

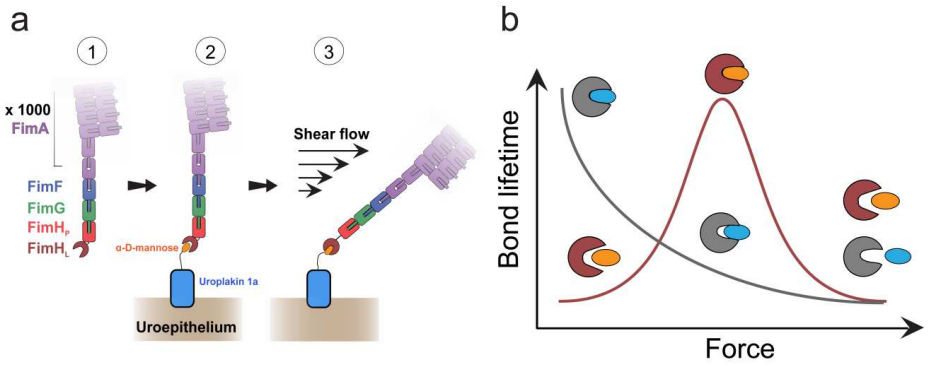


Figure 2.5. Catch-bond mechanism between FimH lectin domain and mannose. (**a**) Transition from the free (1) to the low-affinity bound state (2) of FimH-mannose. In the presence of shear forces (3), a conformational rearrangement between the lectin and the pilin domain of FimH increases the strength of the receptor-ligand bond. (**b**) Force dependence of the bond lifetime of slip-bonds (grey) and catch-bonds (brown). In slip-bonds the probability of receptor-ligand dissociation increases with the tensile force, meanwhile in catch-bonds the

strength increases with the force up to an optimum level and then it decreases again. FimH lectin domain-mannose binding exhibits this force-induced increase in lifetime[72].

Pilus rod uncoiling under shear stress

Besides the optimization of the binding between FimH and the epithelium, the type I pilus also displays an additional mechanism for optimizing its attachment. The pilus rod composed of hundreds of FimA subunits adopts a quaternary structure in which the protein chain forms a rigid right-handed helix[55]. The biomechanical properties of the rod has been studied in single-molecule force experiments which demonstrated that this structure is able to uncoil under force and recoil when the force is reduced[73–75]. Under high shear stress this uncoiling mechanism of the rod increases the length of the pilus several times and it extends the lifetime of the catch-bond, protecting the bacterium from the detachment. When the stress is low enough the pilus experiences an entropic contraction of the chain followed by the recoiling of the helix rod, making possible the movement of the bacteria against the fluid flow (**Fig. 2.6**). This behavior depicts the pilus as a shock-absorber that extends and contracts depending on the flow, reducing the drag force experienced by the bacteria. Pili are present in a high number on the bacteria surface, and each of them acting as a damper would redistribute the stress in all of them, avoiding the overstress on individual pilus[73,76,77].

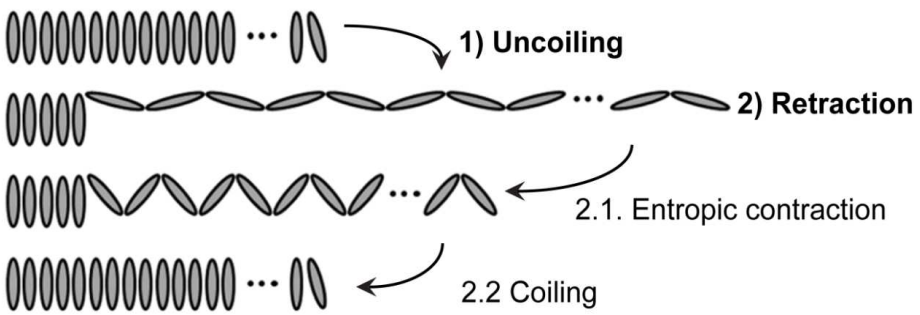


Figure 2.6. Pilus rod uncoiling and retraction. Pilus rod helix configuration uncoils and extends after tensile force application. When the mechanical stress ceases, the retraction of the extended rod occurs in two phases. In the first one the extended rod behaves as an entropic spring, quickly retracting. In the second one the FimA subunits stack and coil again in a helical configuration at lower speed. Adapted from reference[76].

2.5 Type I pilus biogenesis

The pili are placed in the outer plasma membrane (OM) of the bacteria where a transmembrane protein called Usher or FimD serves as a platform for their assembly. The pilus assembly is a complicated multistep process that takes place in several cellular compartments and involves the action of many proteins. The final destination of the pilus proteins is the extracellular space, but before they need to be exported from the cytoplasm to the periplasm through the inner plasma membrane (IM). Once in the periplasm and after the Fims have been folded and maturated, FimD incorporate these proteins to the nascent pilus and it expels them to the extracellular space through the OM.

2.5.1 Exportation to the periplasmic space

This complex process starts with the Fim proteins being synthesized in the bacterial cytoplasm as pre-proteins. These pre-proteins have an N-terminal signal peptide that labels them for being exported to the periplasm. The translocation to the periplasmic space through the IM depends on the Sec secretion system. The Sec translocase is made of three components: the chaperone SecB, the translocation ATPase SecA, and the membrane protein complex SecYEG. The pre-proteins are translocated into the periplasm in an unfolded and extended state through the pore made by the SecYEG complex[78]. Once in the periplasm, the immature Fims are subjected to a maturation process that involves the formation of a disulfide bond through the oxidation of the lateral chains of two cysteine residues, very well conserved not only in the type I pilus subunits but also in other pili[79]. This disulfide bond formation is managed by the Dsb system in the periplasm, whose main actors are the soluble periplasmic protein DsbA and the IM protein DsbB.

2.5.2 Disulfide bond formation by DsbA

DsbA is an oxidoreductase involved in redox reactions taking place in the periplasm. DsbA is part of the thioredoxin-superfamily and like the enzymes of this family it possesses a catalytic motif CXXC (two cysteine residues separated by another two different amino acids) whose cysteine residues alternate between oxidized (disulfide-bonded) and reduced (free thiols) state.

Meanwhile the SecYEG complex translocates the Fim protein, part of the chain appears in the periplasmic side of the IM and then an oxidized DsbA can interact with it. Then one of the free thiols of the Fim proteins performs a nucleophilic attack on DsbA disulfide bond, breaking the bond between Cys30 and Cys33, and forming a mixed disulfide bond with Cys30. Afterwards, the remaining free thiol of the Fim protein attacks the mixed disulfide formed between DsbA and the Fim protein, releasing a reduced DsbA and the disulfide bonded Fim protein. DsbA oxidized state is restored by the reduction of one of the CXXC motifs of the IM protein DsbB. Later on the reoxidation of DsbB is performed by an oxidized quinone, restarting the Dsb system for a new cycle of substrate oxidation (**Fig. 2.7**)[80–83].

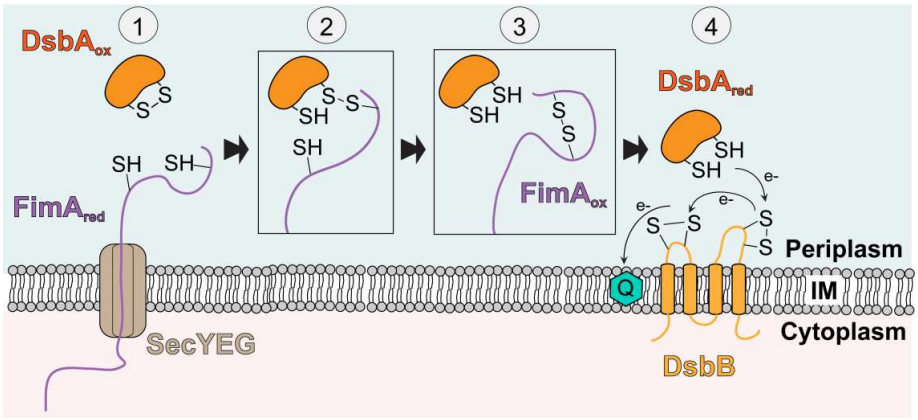


Figure 2.7. Fim protein exportation to the periplasm and disulfide bond formation. (**1**) A FimA preprotein is translocated to the periplasmic space through the SecYEG complex, placed in the inner bacterial membrane (IM). (**2**) A thiol group from one of FimA cysteines attacks an oxidized DsbA enzyme, forming a transient disulfide bond between DsbA and FimA. (**3**) The thiol group from FimA's second cysteine attacks this intermediate, forming a disulfide bond between FimA cysteines and releasing a reduced DsbA. (**4**) DsbA is later oxidated by the transmembrane protein DsbA, restoring its disulfide. An ubiquinone is the final acceptor of the electrons, restoring the cycle for the next DsbA's substrate.

2.5.3 Folding of pilus subunits by the chaperone FimC

As it was mentioned above the four Fim proteins are incomplete Ig-like domains lacking the C-terminal β-strand. This absence creates a deep hydrophobic groove in their fold that makes them thermodynamically unstable unless they are complemented by the donation of the N-terminal β-strand from their cognate subunit, this is the following protein in the pilus assembly (FimH-FimG, FimG-FimF, FimF-FimA, FimA-FimA). Nevertheless this complementation between subunits does not occur until they are incorporated into the growing pilus in the OM platform FimD. After disulfide bond formation induced by DsbA, the chaperone FimC gets into action. This chaperone is made of two connected Ig-fold domains which form a 45º angle between each other[52]. FimC stabilizes the fold of pilus proteins by donating its G1 β-strand, which is inserted between the A' and F strands of the pilin domain, running in a parallel orientation respect to the second (**Fig. 2.8**).

This process is called donor strand complementation (DSC) and it takes place for all the pilin domains in the periplasm[84]. The pilin domains' hydrophobic groove contains five conserved pockets termed P1-P5, which are filled during DSC with FimC's G1 strand P1 to P4 residues (three alternating bulky hydrophobic residues and a fourth hydrophilic amino acid). FimC helps pilin subunits to reach a semifolded metastable state through DSC which makes them competent for assembly into a growing pilus. Besides, FimC protects the pilin subunits from proteolytical degradation, prevents the aggregation with other subunits present in the periplasm, and escort the pilus subunits to

the OM polymerizing catalyst FimD/Usher, the platform in charge of pilus assembly and secretion[84–89].

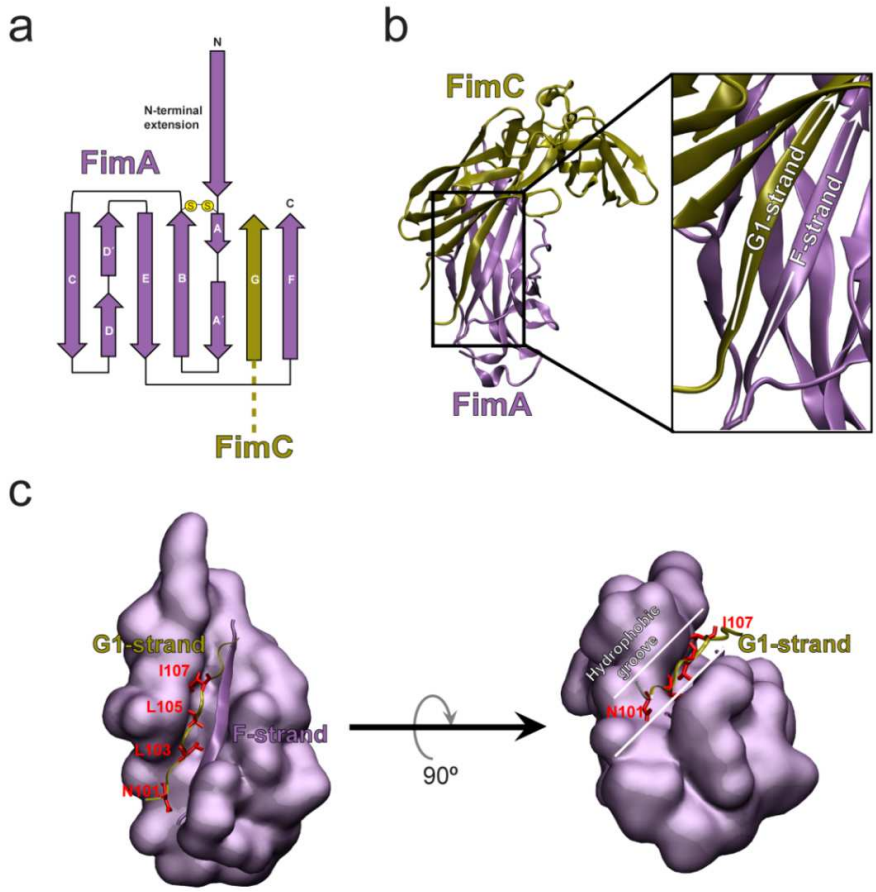


Figure 2.8. DSC of FimA. (**a**) Scheme of the incomplete Ig-fold of FimA with the G1 complementary strand donated by FimC running parallel to FimA F-strand. FimA domain, as FimF and FimG do, has an N-terminal strand for complementing the previously assembled subunit in the pilus. (**b**) Cartoon representation of the FimA-FimC complex, highlighting the parallel orientation of FimC G1-strand and FimA F-strand. Donor N-terminal strand from FimA is not present in this structure. (**c**) QuickSurf representation of FimA fold (lacking the N-terminal strand) with its F-strand depicted as a cartoon and FimC G1-strand depicted as a yellow ribbon, with its P1-P4 residues depicted as red lines. The right figure has been rotated vertically 90º to emphasize how G1-strand fits into FimA hydrophobic groove (PDB: 4DWH).

2.5.4 Pilus assembly and secretion in FimD

The final destination of the pilin-chaperone complexes is the OM protein FimD. FimD is a five-domain OM protein composed of one N-terminal periplasmic domain (NTD), a transmembrane β-barrel domain forming the secretion pore, one periplasmic plug domain that seal the secretion pore, and two periplasmic C-terminal domains (CTD1 and CTD2) (**Fig. 2.9**)[90].

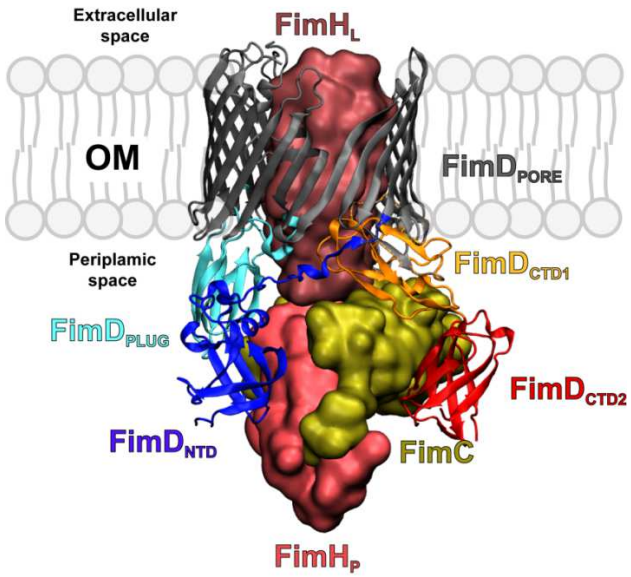


Figure 2.9. FimH-FimC complex recruitment in FimD. FimD is composed of five periplasmic domains and one transmembrane domain. The plug domain occludes the transmembrane pore when FimD is inactive. The NTD receives incoming pilin-chaperone complexes and transfer them to CTD1 and CTD2 for DSE reaction with the next protein in the assembly and latter translocation. Here FimH lectin domain is already inside of the pore and the pilin domain is in DSC state with FimC (PDB: 3RFZ).

The architecture of the pilus is crucial for the attachment of the bacteria, thus a mechanism should control the correct order of pilus subunits incorporaton into the pilus. FimH is the tip end subunit so it should be incorporated the first, followed by FimG, FimF and then

hundreds of FimA subunits. FimD NTD is the primary binding site for pilin-chaperone complexes recruitment, and the affinity for this domain varies among the four different complexes[91]. It has been demonstrated that FimH-FimC complex shows the highest affinity for FimD NTD, explaining why it occupies the first position in the pilus[92]. However the binding affinities of the other complexes do not correlate with their order of assembly into the pilus. The order of assembly is determined by the favorable exchange of FimC G1 complementary strand by the N-terminal β-strand of the incoming subunit to the FimD platform. This process is termed donor-strand exchange (DSE) and it is kinetically favored when the incoming subunit donates the cognate β-strand (this is FimG after FimH, FimF after FimG and so on)[93]. FimD greatly increases the rate of this reaction by bringing closer all the involved elements, and by their correct positioning for optimal DSE[94]. The complementation through antiparallel DSE produces a strong thermodynamic stabilization of the Ig-fold, maintained by hydrophobic effects and an extensive hydrogen bond network between the donated strand and the A and F-strand of the fold[95–97].

After FimH-FimC complex recruitment the plug domain of FimD is moved out to the periplasmic space, allowing the lectin domain of FimH to introduce inside of the pore (**Fig. 2.9**). Then FimH-FimC is transferred to the CTDs of FimD, which provide a high-affinity binding site for the complex. The periplasmic compartment is a very crowded space and probably the NTD of FimD is constantly submitted to incoming complexes which would be randomly incorporated into the pilus. But in order to continue with the pilus translocation the next complex should contain the correct subunit for

DSE with FimH-FimC waiting complex in the CTDs. When a new complex arrives to NTD, the donor β-strand of the subunit is placed in such a way that its P5 residue is close to the FimH P5 pocket. The FimC G1 strand from the FimH-FimC complex only occupies FimH's P1 to P4, leaving free this fifth pocket available for the P5 residue of the incoming donor β-strand. The DSE reaction can only take place if there is a correct steric fit between FimH P5 hydrophobic pocket and the P5 residue of the next recruited subunit, being this correct surface complementarity what finally determines the order of assembly. After P5 residue correctness is addressed, a process of zip-in-zip-out takes place where the incoming β-strand invades the hydrophobic groove and progressively substitutes FimC G1 strand.

After FimH-FimC positioning in the CTDs, a FimG-FimC complex in the NTD proceeds with the DSE with FimH-FimC. The donor β-strand is inserted in an antiparallel fashion respect to the F strand completing and resembling the canonical fold of the pilin domain. After DSE the chaperone is released from FimH, leaving free the CTDs for the transfer from the NTD of the FimG-FimC complex. At the same time FimH translocates through the FimD pore to the extracellular space. This translocation is supported by the high-affinity of the CTDs for the complexes, which compete with NTDs for their binding. An additional support is given because of the presence of a helical pathway inside the pore which imposes a rotation of 110º to each extruded subunit. This generates a rotational uplift through the pore, facilitating this rotation the transfer of the next complex from the NTD to the CTDs (**Fig.2.10**). The free NTD is available for the recruitment of the next complex, this time FimF-FimC, and so on for the next pilins in the assembly[52,56,95,98–104].

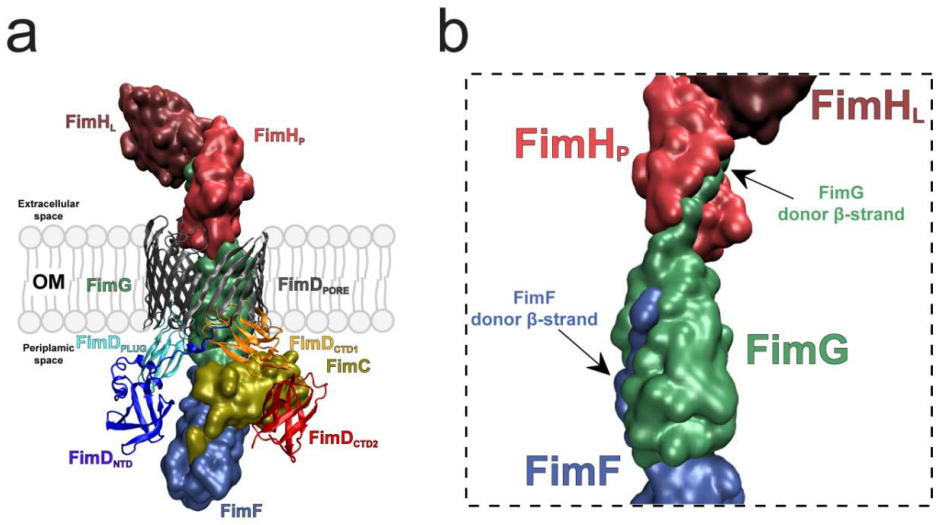


Figure 2.10. Pilus tip assembly in FimD. (**a**) In this moment of the pilus biogenesis, both domains of FimH has been translocated to the extracellular space. FimG is inside the pore and FimF is forming a complex with FimC, waiting for an incoming FimA-FimC complex at the NTD for receiving the donor-strand of FimA. (**b**) Detail of the intersubunit complementation. FimG donor-strand fills the fold of FimH pilin domain, meanwhile FimF donor-strand does the same on FimG (PDB: 4J3O).

It has been demonstrated that subunit translocation only requires the pilus subunits, the FimC chaperone, and FimD for efficient polymerization of pili *in vitro*, being independent of the cellular energy. There is no ATP in the periplasm and no chemical gradient along the OM, thus the competition between the CTDs and the NTD for the pilin-chaperone complex and the helical pathway imposed inside the pore are proposed as the driving-force pushing the secretion of the pilus[94,105]. Besides these two proposed mechanisms for protein translocation, Brownian motions probably also contribute to the secretion of the pilus. However Brownian motions do not impose directionality, so the pilus should be coming in and out of FimD pore with no net extrusion to the extracellular space[52]. It has been proposed that some rearrangements in the pilus subunits take place after being secreted through the pore, such

as an angle change between the lectin and the pilin domain of FimH (**Fig. 2.11**). Both domains are aligned when bound to the CTDs, but after complete extrusion through the pore both domains display a 37.5º angle between each other that would prevent their backward motion. The rotation imposed through the pore plus the quaternary structure adopted by the pilus rod through FimA subunits interactions would give directionality to the pilus growing, but it could not be discarded the role of the plug domain of FimD as a ratchet that avoids the pilus to slide back[90,95].

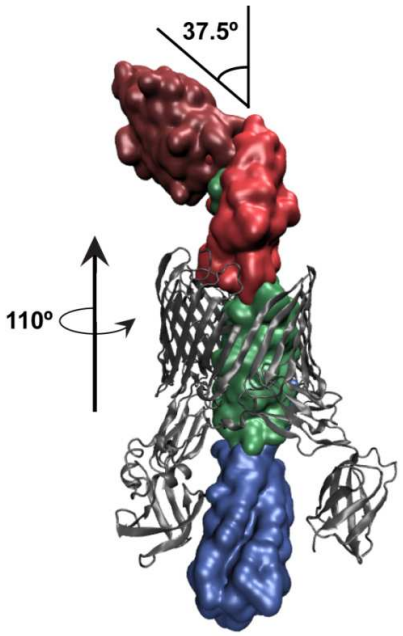


Figure 2.11. Conformational changes guaranteeing pilus extrusion. FimH lectin and pilin domain form a 37.5º angle between each other, a change happening after translocation (both are aligned meanwhile being extruded, see **Fig. 2.8**). Besides, every subunit is 110º counterclockwise rotated meanwhile traversing the pore. This imposed helical pathway make subunits to wind, and in the case of FimA to form a helical rod (PDB: 4J3O).

Chapter 3: CD4 receptor and HIV

In this chapter the main features of HIV-1 infection are explained. The focus is on the cell-surface receptor CD4, the protein used by this virus to establish the first interaction leading to infection. The chemical regulation of CD4 and the HIV-blocking activity of an anti-CD4 antibody are also shown.

3.1 HIV historical perspective

AIDS was identified for the first time in 1981, when a group of young adult people got sick and died from opportunistic infections easily fought by the immune system. The people affected by this disease suffered from slimness and had a weak appearance, besides showing purple lesions on their skin due to a rare type of cancer called Kaposi's sarcoma. Later in 1982 the infectious origin of AIDS was determined, which occurs as a consequence of the exposition to contaminated body fluids or blood. The disease was rapidly spreading worldwide becoming an epidemic, and cases appeared among the previously believed non-risk groups. The origin of the immunodeficiency was found out to be rooted on the diminished levels of CD4 T-cells present in the blood of the affected people[106]. In 1983 it was identified a new human retrovirus in a sample taken from a patient that later on developed AIDS, and it was linked to the human T-cell leukemia virus family (HTLV)[107]. One year later two independent investigations

related this retrovirus, termed as HLTV-III, to lymphadenopathy-associated retroviruses (LAV) regarding morphological and biological criteria. Isolated serum from AIDS patients reacted with antigens present in HLTV-III, suggesting this retrovirus was the etiological agent of AIDS[108–111]. In 1986 this pathogen was named as the human immunodeficiency virus (HIV)[112].

Two HIV types have been identified, termed as HIV-1 and HIV-2, closely related to simian immunodeficiency viruses (SIVs). Both HIV types have an African zoonotic origin, when the simians' pathogens crossed to the human species because of the direct exposition of hunters to chimpanzee blood and mucosal secretions. Both HIV-1 and HIV-2 cause AIDS, but HIV-2 is restricted to West Africa and it is less virulent and transmissible than HIV-1. There are four groups of HIV-1, each one representing a different transmission event (M, N, O and P). The M group is worldwide distributed and it is divided in nine subtypes (A-D, F-H, J, and K) predominating the C and B subtypes[9,11,113,114]. The main transmission mode of HIV-1 is through sexual contact, and since the onset of the pandemic 35 million people have died from AIDS-related diseases and 76 million people have become infected[6].

3.2 HIV-1 tropism

HIV, as SIV does, belongs to the lentivirus genre. These viruses' viral particles are characterized for a lipidic envelope that encloses a protean capsid core where the viral genome and proteins involved during infection are stored (**Fig. 3.1**)[115]. These viruses target the host's

immune system, preventing their eradication by the same immune cells they infect. The HIV-1 (from now referred as HIV) targets the monocyte/macrophage lineage and lymphocyte T-cells establishing persistent infections. The immunodeficiency originates from the infection and depletion of CD4+ helper T cells involved in the adaptive immune response, being these cells the main target of HIV.

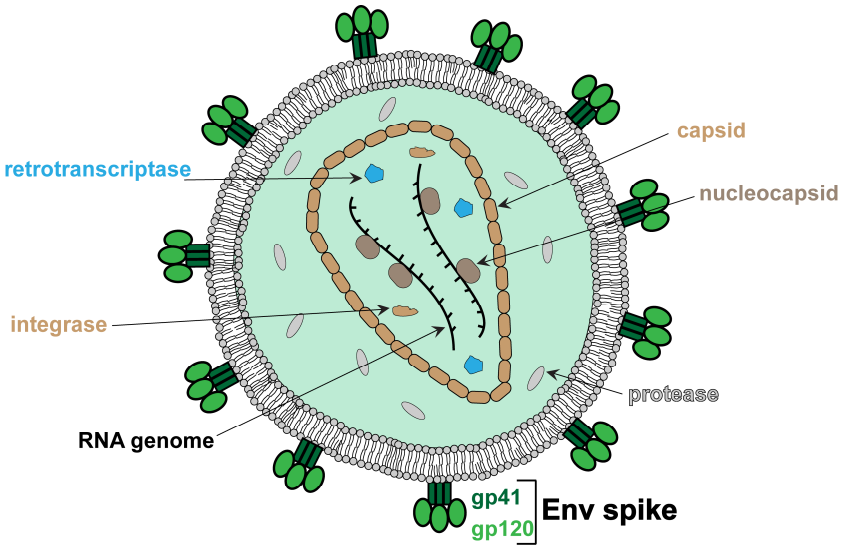


Figure 3.1. HIV-1 structure. HIV virion particle is made of a lipid bilayer envelope which displays on its surface the heterotrimeric protein complexes Env. Each Env complex is made of three gp41 subunits associated with another three gp120 subunits. These two proteins have a crucial role during cell attachment and infection. Inside the virus are stored several proteins and enzymes involved during inside-cell infection process besides a protein capsid which encloses the viral RNA genome. Inspired from Thomas Splettstoesser work (www.scistyle.com). Representation is not at scale.

3.2.1 HIV cell infection at a glance

In order to infect cells HIV free-virion must adhere to the surface of the cell. Cell recognition and adhesion relies on a viral surface glycoprotein complex termed Env. Env binds sequentially to CD4 surface receptor and then to a membrane chemokine coreceptor which can be either

CCR5 or CXCR4. This second interaction with the coreceptor triggers conformational changes on HIV Env glycoprotein inducing membrane fusion between the cell and the virus, liberating the viral capsid into the cell.

Once in the cell cytoplasm the viral capsid is uncoated and the single stranded RNA genome is released for reverse transcription into double stranded DNA by an error-prone reverse transcriptase enzyme. Viral DNA is then transported to the cell nucleus where it integrates into a chromosome. Using host cell machinery the HIV genome is transcribed, translated, and post processed in order to generate new virions that finally bud out from the cell, releasing new infective particles to the extracellular space for viral expansion inside the host (**Fig. 3.2**).

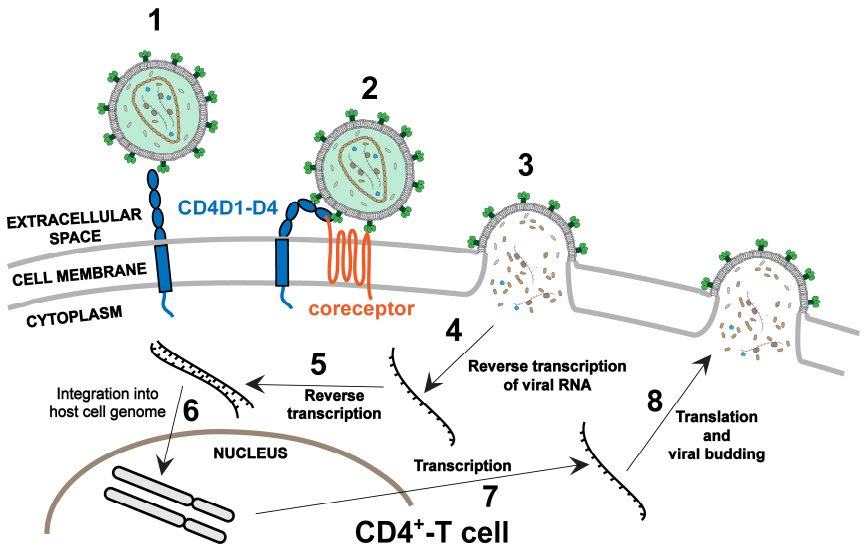


Figure 3.2. HIV-1 life cycle. (**1**) HIV-1 infection starts with the binding of the virion to a CD4 receptor present on the surface of a $CD4^+$-T cell. This attachment is done through the viral surface glycoprotein gp120 from an Env spike. (**2**) This same glycoprotein interacts with a surface coreceptor meanwhile attached to CD4. (**3**) Conformational changes in the viral Env spike leads to the formation of a pore on the host cell surface and the fusion of both

membranes, releasing the content of the virion into the cell cytoplasm. (**4** y **5**) Viral single-stranded RNA is reverse transcribed into double-stranded DNA. (**6**) Viral DNA genome is transported to the cell nucleus and it is inserted into the cell genome. (**7** y **8**) The inserted viral genome is transcribed and translated to assembly new infective virions that finally bud out from the cell surface. Inspired from reference[11]. Representation is not at scale.

3.2.2 Cell-to-cell infection

Although HIV primary targets $CD4^+$ helper T cells, antigen presenting cells (APCs) such as dendritic cells or macrophages can also become infected by HIV since they express the receptors required for Env viral protein binding, but in lower levels. Besides cell-free virion infection HIV has the ability to propagate and infect immune system cells through cell-cell interactions. These interactions are a natural mechanism among immune system cells, but HIV hijacks them to spread. Lymphocyte cells are constantly searching and identifying antigens present in the blood stream and in the lymphatic system. During the adaptive immunity response T-cell receptors (TCR) displayed on the surface of CD4 T-lymphocytes, recognize peptides presented by an APC in the major histocompatibility complex class II (MHC-II). MHC-II molecules are placed on the surface of APCs and the MHC-II/TCR interaction is further stabilized through CD4 receptor binding to MHC-II, forming what is termed as the immunological synapse. The recruitment of CD4 to the MHC-II/TCR intercellular complex besides a stabilization role also triggers the onset of cell signaling for T-cell activation. Cell-to-cell transmission can occur during antigen recognition by T cells of APCs like dendritic cells, which can capture and internalize HIV virion. The virus containing dendritic cell (but maybe not infected) transmits the virus to the T-cell

through a process called trans-infection that requires the formation of what is termed virological synapse between the two cells. This viral synapse is the membrane cell-to-cell contact area characterized by the recruitment of CD4 and the coreceptor, highlighting again the importance of these two membrane receptors for HIV entry.

However as it was noted before the preferential target of HIV are the $CD4^+$ T-cells, and the dissemination of the virus through T-cell/T-cell interactions has proven to be highly efficient for HIV spreading. In the viral synapse between two T-cells the infected cell displays the viral Env glycoprotein, which interacts with the CD4 receptor present in the uninfected T-cell. Although this interaction is crucial for the successful infection, other surface adhesion molecules such as ICAM-1, ICAM-3 and LFA-1 facilitate the adhesion between cells. The Env-CD4 interaction yields a time lasting adhesion that reorganizes the viral production of the infected cell close to the cell-cell contact surface. The viral particles are then internalized by the non-infected cell into an endocytic compartment where the virions are finally maturated into a fusion-competent state[116,117].

3.3 HIV entry key components

HIV entry requires three essential components: viral Env, CD4 receptor, and any of the coreceptors. HIV can attach to cells through several different auxiliary molecules present on the cell surface but viral and cell membrane fusion will not be triggered until the recruitment of CD4 and coreceptor molecules, making their presence indispensable for HIV entry (**Fig. 3.3**).

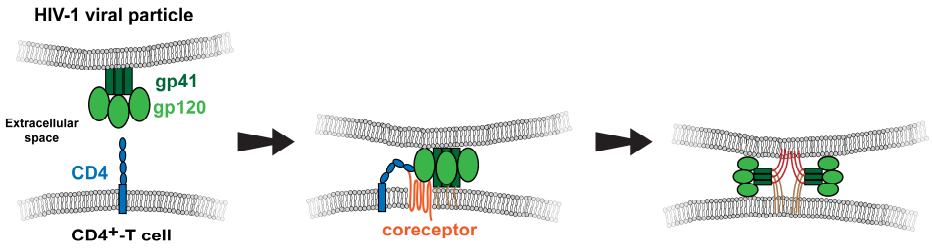


Figure 3.3. HIV-1 adhesion and entry. HIV Env spikes are made of two proteins, called gp41 and gp120. The last one mediates the attachment to CD4D1 domain and later to a surface coreceptor. This two interactions of gp120 trigger structural changes in the Env spike that leads to the release of a fusion peptide from gp41 which generates a fusion pore for membranes fusion. Inspired from reference[118]. Representation is not at scale.

3.3.1 Env spikes

Env is the membrane glycoprotein of HIV and the unique viral antigenic protein exposed. It is first synthesized in the endoplasmic reticulum as a precursor termed gp160. After folding and extensive glycosylation, gp160 forms trimers stabilized through non-covalent interactions. These trimers travel to the Golgi compartment where they are cleaved into two subunits called gp120 and gp41, generating a trimer of heterodimers (three gp120/gp41 dimers)[119]. After maturation these Env trimers are transported to the cell surface where they are incorporated into the budding virions.

Env trimers protrude out of the viral envelope generating what is termed viral spikes. In these spikes gp41 is the transmembrane protein and gp120 the exterior subunit. The gp120 glycoprotein mediates the sequential binding of the virus to CD4 and then to the coreceptor, meanwhile gp41 is involved in the fusion of the viral envelope with the host cell membrane.

The gp41 subunit is composed of three major domains: a C-terminal cytoplasmic domain, a transmembrane domain and an extracellular domain. This last domain contains, among other important sequences, the fusion peptide that penetrates and destabilizes the cell membrane to form the fusion pore. This fusion peptide is buried in the viral spike, and the binding of gp120 to CD4 and the coreceptor triggers conformational changes that expose it. The most apical part of Env is the gp120 glycoprotein. Its structure is made of five variable loop domains (V1-V5) and five constant domains (C1-C5). The sequence variability of the V domains originates from genetic changes like mutations and recombination and together with the heavy glycosylation of gp120, it prevents the recognition from the immune system. The CD4 binding site in gp120 is disseminated within the C1, C3 and C4 domains, which are far from each other in sequence but in close proximity in the folded structure. In the bound state CD4 contacts 26 residues of gp120 triggering a large conformational change in gp120 that exposes the coreceptor binding surface placed in V3 region. V3 region shows dominant epitopes that are recognized by neutralizing antibodies, but its exposition only occurs after CD4 binding. Postposing the exposition until CD4 binding limits the time in which V3 epitopes can be targeted, and also it limits its availability due to the proximity to the cell surface which sterically prevents antibody access[120].

3.3.2 CD4 receptor

CD4 is a cell-surface glycoprotein present in $CD4^+$ T-lymphocytes, macrophages and dendritic cells. It is composed of four extracellular immunoglobulin domains, numbered from the N-terminal most exterior

domain as D1 and the closest one to the cell membrane as D4. All domains with the exception of D3 contain disulfide bonds, and both D3 and D4 contain one glycosylation site. Besides the extracellular part CD4 also possesses a single transmembrane domain and a cytoplasmic C-terminal tail (**Fig. 3.4**)[121].

During the adaptive immunity response the TCR recognizes and binds to antigenic peptides bound to the MHC-II presented by APCs. The T-helper cell is activated due to the recruitment of CD4 receptor which binds to MHC-II in a different site than TCR does it. The cytoplasmic tail of CD4 is associated to the tyrosine kinase Lck, which would be brought into close contact to the cytoplasmic side of the CD3-TCR complex leading to their phosphorylation initiating a cascade signal that leads to T-cell activation[122,123].

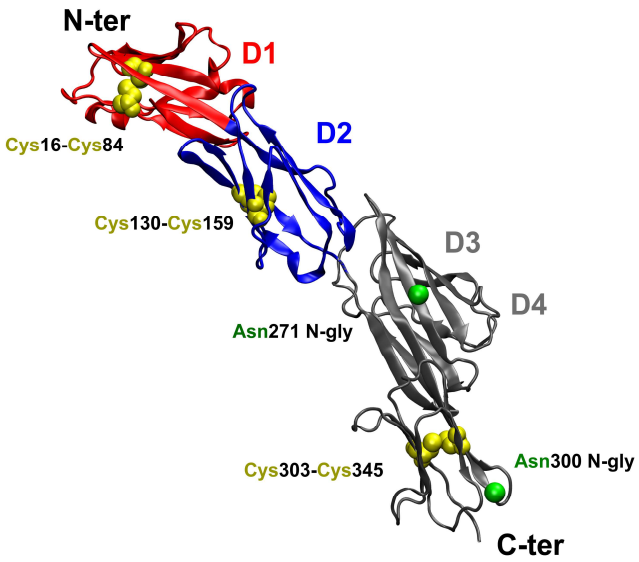


Figure 3.4. CD4 extracellular structure. The extracellular domain of CD4 is composed of four domains. D1 and D2 are colored different from D3-D4 because of the D1 role during HIV binding and because of the role of D2 as a promising target for antibody recognition to prevent HIV infection (D2). Disulfide bonds are highlighted in yellow van der Waals spheres, meanwhile the glycosylation sites are represented with green beads (PDB: 1WIP).

3.3.3 Chemokine receptor CCR5 or CXCR4

CCR5 and CXCR4 chemokine receptors are seven-transmembrane domains coupled to G-proteins. Both present an extracellular N-terminal tail, three intracellular and three extracellular loops, and a C-terminal cytoplasmic extension[119,124]. Chemokines are a group of small peptide molecules which act as chemoattractants and activators of phagocytes. The chemokine interaction with their receptors triggers biochemical and cellular changes important for a rapid first response against infections.

These two receptors are distributed differently among different cell types. Meanwhile CXCR4 is expressed on neutrophils, myeloid cells, and T-lymphocytes; the CCR5 receptor is found in T-lymphocytes and in macrophages[125].

3.4 gp120-CD4/coreceptor binding

HIV entry starts with the virion binding to the cell surface. This binding is enhanced and facilitated by non-covalent interactions with surface molecules such as mannose-binding lectins and heparan sulfate, although the role of this attachment factors differs depending on the cell type. After this initial unspecific adhesion to the cell surface the encounter with CD4 takes place[119,124].

The gp120 structure can be structurally divided in a more conserved inner domain and an external variable domain. CD4 binding site (CD4bs) is placed in a depression at the limit of the inner and the outer domain and limiting also with the bridging-sheet, a four antiparallel strands-structure that lies between de V1/V2 loops. CD4bs

is free of glycosyl chains but it is surrounded by them. The interatomic contacts made in the binding site imply hydrogen bonding and van der Waals interactions formed between 22 CD4 residues and 26 gp120 residues. The CD4 residues implicated range from the 22nd to the 64th amino acid from domain 1 (CD4D1) but in the case of gp120 the residues are distributed over a wide range of its sequence. The critical residues for binding have been identified as Phe 43 and Arg 59 for CD4, and Asp 368, Glu 370, and Trp 427 for gp120. CD4 Phe 43 is placed inside a gp120 cavity made of hydrophobic residues just in the interface between the inner and the outer domain of gp120 (**Fig. 3.1**)[126]. CD4bs constitutes a vulnerability of the viral glycoprotein for immune system evasion, and has been subjected to intense research for vaccine development. Several facts support this approach since soluble versions of CD4 neutralize the virus in a high extent, and broadly neutralizing antibodies isolated from infected patients target this same epitope[127,128].

CD4 binding to gp120 triggers large conformational changes on the quaternary structure of the Env trimer, where the gp41 stalk is exposed and each gp120 subunit is outwardly rotated with respect to the Env trimer central axis. It has been suggested that activation of at least two gp120 subunits of the same Env trimer spike to trigger the changes is required for viral fusion[129,130]. Inside gp120 the V1/V2 stem is moved out from the center of the trimer to a lateral part of this, allowing the rearrangement of the V3 loop towards the apical end of the trimer. After these changes the V3 loop partially exposes the coreceptor binding site facing the cell membrane. The V3 loop tip would act as a molecular hook that sticks out from the Env trimer enabling the interaction with the coreceptor[131]. The tropism for CCR5 or CXCR4 depends on interactions made with the coreceptor by the

V1/V2 and V3 loops and the bridging-sheet, with a difference in the net positive charge of the residues of the V3 loop what determines the usage of one or the other. The selectivity for CXCR4 depends, at least for the subtype-B virus, on the presence of a positively-charged residue either one or both of the 11th and 25th positions of the V3 loop[119,132].

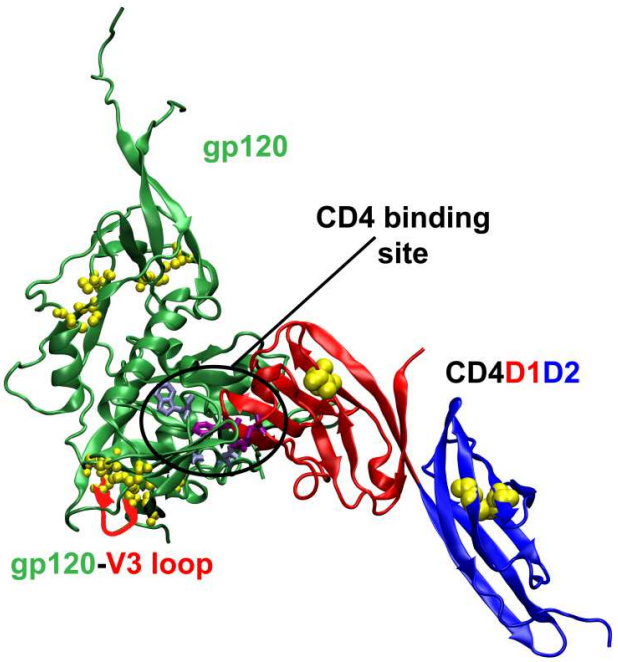


Figure 3.5. gp120-CD4. Crystal structure of the first two domains of CD4 (D1D2) bound to one gp120 subunit. Disulfide bonds are represented as van der Waals spheres in both structures but with a reduced radius in gp120 for the sake of clarity. Besides the gp120 V3-loop is highlighted in red, and the bonds of critical residues for gp120/CD4 binding are represented in grey in gp120 (Asp368, Glu370, Trp427) and purple in CD4 (Phe43, Arg59) (PDB: 3JWD).

Upon coreceptor binding, membranes fusion and fusion pore formation occur with the insertion of the gp41 N-terminal fusion peptide into the cell membrane. This process relies on many other

drastic conformational changes in the Env spike that lead to the fusion of viral and cell membranes.

3.4.1 Thermodynamics of gp120/CD4

The gp120/CD4 binding entails structural changes for both molecules. These changes have been discovered thanks to a big effort put in the determination of different crystal structures of gp120, CD4, and complexes of both, also in the presence of antibodies. A step beyond the static picture provided by Structural Biology techniques was achieved when the thermodynamics of gp120/CD4 interaction were studied. The enormous enthalpic contribution of this binding was determined, supported by the large number of hydrogen bonds and van der Waals interactions formed between both molecules. Despite of the favorable binding enthalpy a large change in entropy is also observed as a consequence of the reduction of the conformational flexibility of the complex, whose decrease comes from the structural stabilization of gp120 after the rearrangements that occur after binding. This reorganization is supported on secondary structure changes after binding where the reduction of random coil parts suggests more structural rigidity. Besides, the slowness in the association rate constant between gp120 and CD4 points to a structural adaptation between both molecules. The high gp120 flexibility in the unbound state could be the reason of the lack of broadly neutralizing antibodies, whose epitopes are placed in discontinuous disordered parts of gp120. Only the binding to CD4 reduces drastically the conformational flexibility of gp120 leading to the exposition of immunogenic sites that could be potentially

targeted, but prevented because of the steric hindrance imposed by the close proximity of the cell membrane[133].

3.5 CD4 mechanochemical modification during viral docking

3.5.1 CD4 bending flexibility

The gp120/CD4 complex formation suggests the stiffening of both molecules after binding and gp120 structural rearrangement. However the downstream events that lead to viral and cell membrane fusion require movements and flexibility in the involved molecules. As it was mentioned above one change after binding is the outward rotation of the gp120 subunits of the Env trimer, a movement that would affect CD4 since both are bound. In this moment the gp120 V3 loop is in a coreceptor binding competent state, nevertheless the cell membrane is still far and the gp120/coreceptor encounter is unreachable[131]. It has been proposed the requirement of CD4 bending to approach gp120 V3 loop to the membrane coreceptor. In the unbound form of CD4 the four extracellular domains exhibit a semi-extended Z shape, and this disposition changes after gp120 binding and rotation. This gp120 movement would induce the bending of CD4 along a hinge region located between its first two domains (D1-D2) and the last two domains (D3-D4), enabling the approaching of the gp120 V3 loop to the membrane coreceptor extracellular loops. Some monoclonal antibodies that bind in the limit between the D2 and D3 domains block the HIV membrane fusion but do not affect gp120/CD4D1 binding. These

evidences suggest that CD4 conformational flexibility, at least at the D2-D3 hinge region, is required for HIV fusion[134,135].

3.5.2 Shock absorber role of CD4 during HIV binding

The role of CD4 structural plasticity during HIV entry gains more importance if the size of the virion and the receptor are considered. The virion is an inert spherical particle of 100 nm of diameter which performs random movements in fluid bodies due to thermal fluctuations. CD4 is, instead, a rod-shape protein of 15 nm length also submitted to Brownian motions as the virion but anchored to the cell membrane. This physical link to the membrane greatly decreases its degrees of freedom although lateral diffusion over the membrane is feasible. The viral particle and the CD4-bearing cell also experience the shear flow of the body fluids, such as the bloodstream[136]. All together configure a complex physical scenario for viral docking on the cell membrane, in which a virus could be rolling and bumping over the cell surface for a while, establishing and breaking bonds between the viral gp120 and the cell CD4. After several attempts, HIV virion could establish a stable bond with one CD4 receptor. Brownian motions would still be acting on the virus, approaching and moving away from the cell membrane, pushing and stretching the intermolecular weak contacts made in the gp120/CD4 complex and challenging the lifetime of this bond[137–139].

At the same time that this viral particle is tethered to the cell surface through a CD4 receptor, it has its coreceptor binding site and its previously masked epitopes exposed and susceptible to immune

surveillance. The coreceptor binding would depend on the proximity to the cell surface and also on the probability of locating a CCR5 or CXCR4 coreceptor nearby the CD4 (**Fig. 3.6**). The structural flexibility of CD4 could facilitate this membrane approaching and the coreceptor finding due to its bending ability, but virus random movements could still break the gp120/CD4 interaction. The range of forces experienced by a CD4 receptor being pulled by the bound viral particle could lead to the additional extension of the protein upon domain unfolding of one/some of its domain/s. The unfolding of some CD4 domains would increase the length of the tether that keeps the virus anchored, allowing for the reconnaissance of a larger area of the cell membrane for coreceptor binding. Of course, a larger tether supposes a drawback since the increased virion-cell separation would facilitate the neutralizing antibodies access to HIV epitopes.

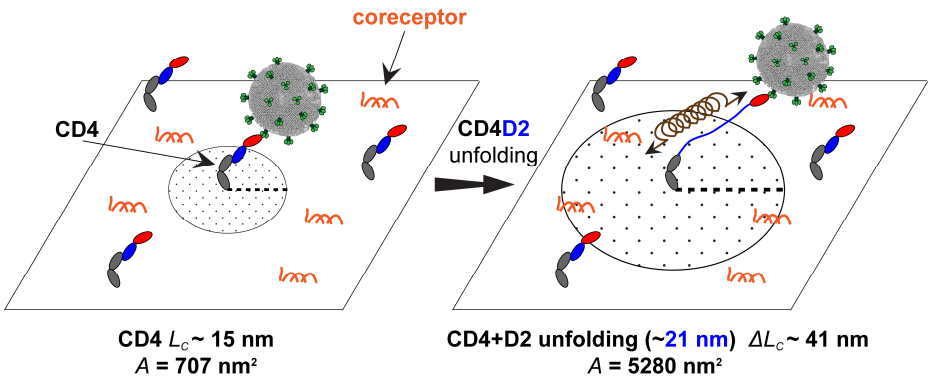


Figure 3.6. CD4-HIV tethering. Virion is attached to CD4. The unfolding of D2 domain (here considered without the disulfide bond) due to the force exerted by the virus on CD4 increases the length of the tether allowing the virus to explore a larger area of the surface cell, incrementing the chances of finding a coreceptor. When the forces are reduced, D2 is able to refold thus acting as a spring. Representation is not at scale.

However this strategy reduces the mechanical stress experienced by the gp120/CD4D1 bond by the unfolding of D2-D4 domains, extending the lifetime of the interaction, increasing the chances of coreceptor binding and subsequent infection. Moreover the random movements of the virus would eventually reduce the strain on CD4 allowing the refolding of its domains, reducing the length of the tether, and bringing closer again the virus to the cell membrane. This hypothesis would depict CD4 extracellular domains as a shock absorber-like structure made of four springs arranged one after the other in a rod, a mechanical feature efficiently exploited for the HIV successful infection of the cell.

3.5.3 gp120/CD4 disulfide bonds redox regulation during the infection

Both gp120 and CD4 possess several disulfide bonds in their structure. The presence of disulfide bonds in extracellular exposed-proteins enhances their stability under changing environmental conditions. Disulfide bonds contribute also in the folding free energy of proteins, decreasing the entropy of the unfolded state and favoring also enthalpically the native state of the protein. However the spatial arrangement of some intramolecular disulfide bonds could affect negatively the thermodynamical stability of the native state of some proteins due to structural-strain induced by this bond. These strained disulfide bonds present higher dihedral energies because of the unfavorable angle disposition of the atoms involved in this bond, making these disulfides more susceptible to redox regulation. These potentially redox-regulated disulfides are termed as allosteric, which

means that rather than being involved in protein stability or catalytic activities they alter protein function depending on their oxidation state. This is the case of CD4D2 domain, whose disulfide bond geometry differs from those ones of D1 and D4 domains. D2 domain disulfide shows properties common with other disulfide redox-regulated proteins, and its reduction while in the cell surface could help for the oligomerization of CD4 receptors during TCR-MHC-II complex stabilization. Free thiols from adjacent CD4D2 domains could react and form intermolecular disulfide bonds, increasing the local colocalization of CD4 cytoplasmic tail-associated enzymes involved in T-cell activation cell signaling[140,141]. CD4D2 disulfide bond could be reduced by oxidoreductase enzymes naturally occurring inside the cell but also known to be secreted and associated with the cell surface, like thioredoxin and/or glutaredoxin.

Regarding gp120, nine disulfide bonds have been identified in its structure. These disulfides connect several loops of its structure and the conformational rearrangement that takes place after binding to CD4 could include the reduction of at least two or three disulfides, located near CD4bs. Their reduction would make possible the reorganization of the V1/V2 loop and the projection of V3 loop for coreceptor binding. Further redox changes in some of the other disulfides could lead to the transformations required for gp41 fusion peptide extrusion. The gp120 disulfide bond reduction would be potentially performed by oxidoreductases like PDI, also secreted to the extracellular space. Interestingly PDI appears to be associated to CD4D3 or D4 domains but it targets only gp120 disulfides, meanwhile thioredoxin would target CD4D2 domain and potentially also the V3 loop disulfide bond[142–144]. It has been shown the direct link between the redox state of

CD4D2 domain and HIV infection, not for the gp120/CD4 complex formation but for critical events occurring before viral fusion[145]. The reduction of CD4D2 disulfide bond could lead to conformational changes critical for HIV fusion. The reduction of CD4D2 domain and some gp120 disulfides is essential for HIV entry[146], as it has been demonstrated by HIV entry prevention by the inhibition of the activity of thioredoxin, PDI and glutaredoxin [140,145,147–149].

However the CD4D2 domain disulfide bond location is buried in the resolved crystal structures available. How thioredoxin access this disulfide remains elusive[147], but the partial unfolding of D2 domain could expose it for enzyme induced-reduction (**Fig. 3.7**).

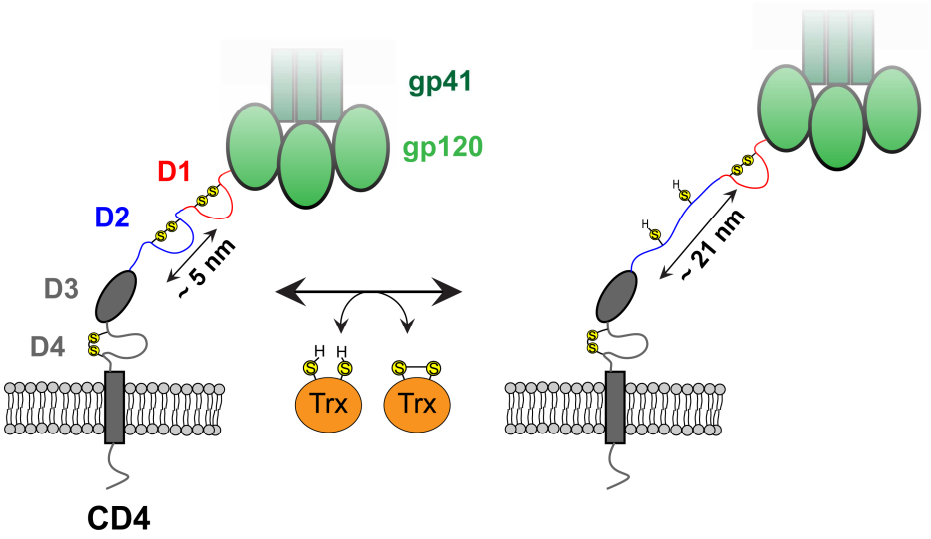


Figure 3.7. CD4 disulfide bonds regulation by oxidoreductases. Thioredoxin (Trx) is secreted to the extracellular space and it has the ability of reducing disulfide bonds, like the present in CD4. CD4D2 domain possesses an allosteric disulfide which when reduced allows the extension of the polypeptide, thus increasing the length of the tether. Representation is not at scale.

Again the viral particle attached to CD4 could induce the mechanical unfolding of CD4 domains, exposing the disulfide bonds of the three extracellular disulfide-containing domains, which allows for their thioredoxin-catalyzed reduction and for a further extension if the domains are fully unfolded. The mechanochemical reduction of CD4 disulfide bonds would be a way for lengthening the tether, covering a wider cell surface for coreceptor binding.

3.5.4 Mechanical role of CD4-binding antibodies

Due to the difficulty to develop broadly neutralizing antibodies directed against gp120 epitopes, antibodies binding CD4 are another therapeutic approach for preventing HIV infection. Targeting CD4 represents the advantage of a low mutation rate since it is a host protein, unlike HIV proteins whose rapid variability makes difficult their long-term effectiveness.

For example the monoclonal antibodies Q425 and Q428 bind to CD4D3 domain, blocking HIV-cell fusion but not modifying the interaction with gp120. It was proposed that these two antibodies block the hinge-like flexibility of CD4, preventing gp120 approach to the cell membrane[134]. Some anti-CD4 antibodies targeting D1 domain inhibit HIV infection directly competing for the binding with gp120, like 15A7 or Leu3A[150,151]. Ibalizumab is an interesting broadly neutralizing anti-CD4 antibody which targets the D2 domain, and its epitope location does not interfere with the binding of gp120 to CD4D1 (**Fig. 3.8**).

Besides not hindering gp120 binding neither does it with MHC-II binding epitope on CD4, avoiding potentially undesirable side effects[153]. Not interfering with gp120 binding, the HIV infection inhibition activity of Ibalizumab develops after gp120/CD4 binding, although the exact mechanism is not known. It has been proposed that the structural rearrangements experienced by CD4 upon gp120 binding could be blocked by Ibalizumab[152]. Some protein-protein interactions like antibody-antigen recognition could lead to the mechanical stabilization of the latter one[41,42]. In this context Ibalizumab could inhibit HIV entry stiffening the CD4 molecule, maybe preventing the mechanical unfolding of D2 domain and abolishing the reduction of its disulfide bond by inhibiting thioredoxin access.

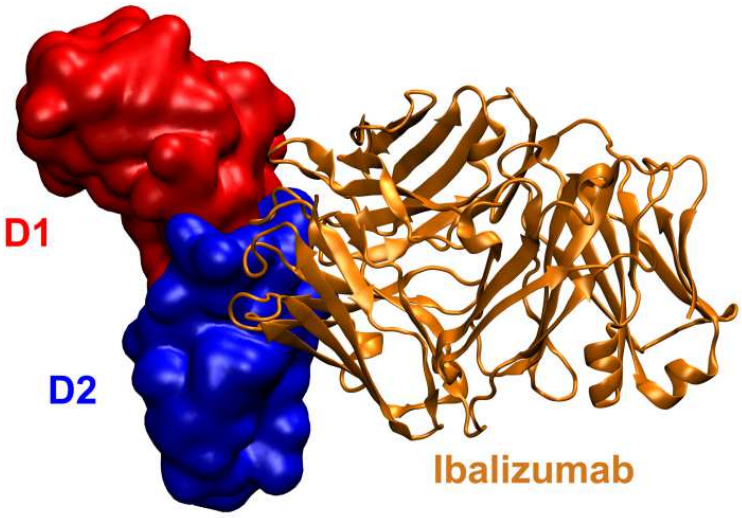


Figure 3.8. CD4D1D2 with Ibalizumab. Neutralizing antibody Ibalizumab bound to CD4. Most of the contacts are made with the B-C loop of domain D2, not overlapping neither with MHC-II binding site neither with gp120 binding site[152]. CD4D1D2 in QuickSurf representation and Ibalizumab in cartoon representation (PDB: 3O2D).

Chapter 4: A mechanical approach to infection

It is clear the existence and the importance of forces at the nanoscale *in vivo* and also during the development of some diseases, including pathogenic infections. In this last introductory chapter the technique used in this thesis for the mechanical manipulation of molecules involved during UPEC and HIV-1 attachment is presented. It is also explained in which way mechanical forces affect protein unfolding and the knowledge that can be obtained from these studies. In the last section is explained how this mechanical manipulations has allowed us to explore force aspects involved during UPEC and HIV-1 attachment.

4.1 Single molecule force spectroscopy (smFS) as a tool for studying force-bearing proteins

4.1.1 smFS features

In the previous introductory chapters several examples have been mentioned for illustrating the importance of physical forces in Physiology, highlighting the UPEC and HIV-1 attachment cases. Until recent times the possibility of studying the mechanical properties of single biomolecules was elusive. The development of single molecule techniques marked a turning point for elucidating the mechanical behavior of proteins, allowing their direct manipulation. The application of mechanical loads to single molecules produces

conformational changes along a well-defined reaction coordinate, different from the ones obtained from thermal or chemical denaturation bulk experiments. smFS techniques also offer a dynamic real-time picture of the force-induced conformational changes in proteins, meanwhile traditional Structural Biology techniques such as X-ray crystallography provide with static images of protein structures[154].

SmFS allow the study of the folding/unfolding of biopolymers like proteins and the characterization of their folding energy landscape under mechanical stress. Unlike chemical or thermal denaturation experiments smFS techniques have a well-defined reaction coordinate called end-to-end extension, which is the distance between the pulled groups of the molecule[155]. This makes possible for smFS to resolve intermediate conformational states on proteins which detection is usually averaged out in bulk experiments, where the most populated states overlap the signal of the rare events[156–158].

The main smFS techniques used for the mechanical manipulation of proteins are Atomic Force Microscopy (AFM), optical (OT) and magnetic tweezers (MT). In all of them a biomolecule is held fixed on one end and on the other end to a probe. Along the experiment the probe position with respect to the fixed end is changed, stretching or relaxing the molecule, and measuring the extension and/or the force applied to the molecule. Each of the smFS techniques has its own advantages and disadvantages, regarding not only the experimental design but also the probe mechanical stiffness, time resolution, and the force-range sensitivity[159]. The first smFS studies conducted on biomolecules focused on DNA and RNA unfolding[160,161], but soon the research was broaden to study muscle proteins like the aforementioned

titin[31,35], and ECM proteins[162]. These first studies on the mechanical behavior of proteins were conducted mainly in AFM, the tool used in this thesis for the mechanical characterization of proteins.

4.1.2 Effect of force on the free energy of proteins

Before studying protein mechanical unfolding with smFS techniques it is necessary to understand which interactions and forces maintain proteins in a folded state. Proteins are polymeric chains made of monomers called amino acids which are connected covalently through peptide bonds. In solution the proteins adopt a collapsed state since this conformation maximizes the conformational freedom of its monomers, thus increasing the entropy. This means that under equilibrium the end points of the protein are on average spatially close between each other, and the mechanical extension of the protein faces the opposing resistance coming from this tendency to maximize the entropy of this system. As a consequence the existence under equilibrium conditions of the extended state of the protein is disadvantaged[31].

The opposing entropic and enthalpic contributions drive the protein folding process from the collapsed to the final folded state. Therein many parallel pathways are possible. The enthalpic contribution, together with the entropic one, determines the structure of the protein, mainly held by weak interactions such as Van der Waals, electrostatic interactions, and hydrogen bonding[163,164]. Besides in proteins two strong interactions also exist: disulfide bonds and isopeptide bonds[165,166]. All of these weak and strong enthalpic interactions are short-ranged thus their strength decays with distance.

When a protein is mechanically stretched the both contributions are responsible of the resistance. Once the force applied to the molecule is high enough the short-ranged weak interactions are broken and the protein goes from a compact state to an extended one. In the unfolded state the remaining short-ranged interactions are usually minimum, and the opposing force of the protein to continue to be extended is entropically driven[158,167].

In the simplest system a protein shows a two-state behavior defined by a folded state and an unfolded state. The free energy landscape of such a protein shows two energy wells separated by an energy barrier (**Fig. 4.1**). One deeper well is placed in an energy minimum which corresponds with the native folded state, and the other shallower well represents the unfolded state which is less energetically favorable. To represent the effect of the force applied in a smFS experiment on the folding energy landscape of a protein, the free energy is plotted *vs* the end-to-end distance of the pulling ends of the protein. The folded state is characterized by a short distance between the ends of the protein, meanwhile in the unfolded state the ends are separated by a larger distance. The energy barrier ($\Delta G^{0}_{x^{\ddagger}}$) represents the free energy of the transition state that separates the folded and the unfolded states of the protein, and its height is due to the enthalpic interactions that keep the protein folded. The folded state and the transition state are separated by a distance $x^{\ddagger}$, and it is the distance between the ends of the protein necessary to reach the transition state. When a force is applied the energy landscape is tilted, lowering both the height of the free energy barrier of the transition state and the free energy of the unfolded state. When the force is high, the opposing entropic forces of the extended unfolded state become energetically

favored[155].

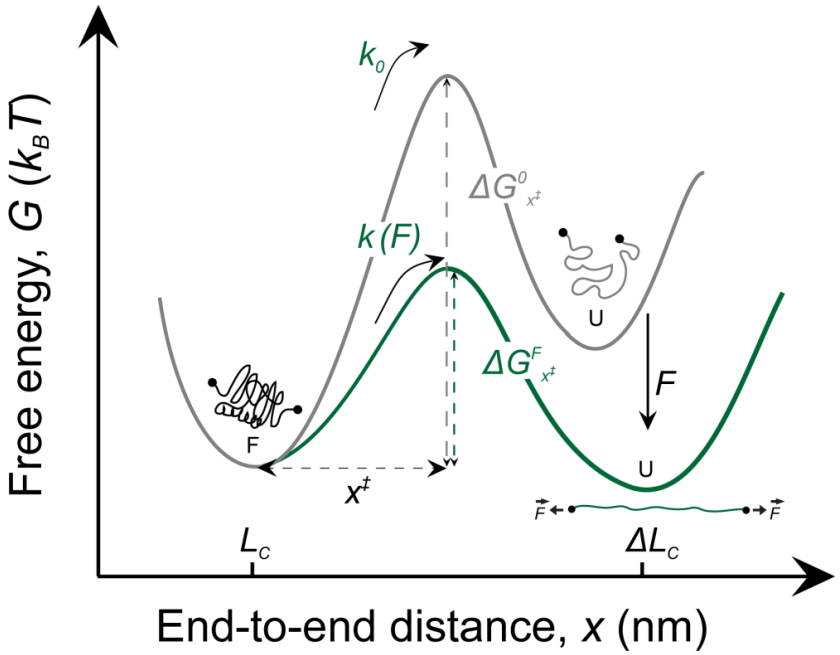


Figure 4.1. Folding free energy landscape of a two-state protein in the absence and in the presence of an external applied force. *k(F)* is the rate constant under a specific force *F*, *A* is an attempt frequency constant with arbitrary value, k_0 is the rate constant in the absence of force, *F* is the mechanical force applied on the protein, $x^\ddagger$ is the distance to the transition state, k_B is the Boltzmann constant and *T* is the absolute temperature. L_c is the contour length of the folded protein and ΔL_c the increment of the contour length after mechanical unfolding.

In order to study the effect of mechanical forces on the kinetics of a protein and to determine its 1D energy landscape, the Bell's model has been widely used. This model predicts an exponential dependency of the rate of bond rupture (the loss of the interactions that keep proteins folded) with the applied force[67]:

$$k_0 = A\exp\frac{-(\Delta G^0_{TS})}{k_BT}; \qquad k(F) = k_0\exp\frac{(Fx^\ddagger)}{k_BT} \qquad (1)$$

Where *k(F)* is the rate constant under a specific force *F*, *A* is an attempt frequency constant with arbitrary value, k_0 is the rate constant

in the absence of force, F is the mechanical force applied on the protein, $x^{\ddagger}$ is the distance to the transition state, k_B is the Boltzmann constant and T is the absolute temperature. As it was noted before this one-dimensional representation of the energy landscape shows a two-state protein. An increasing amount of data available about protein folding point to the presence of conformational intermediates between the folded and the unfolded state, besides the existence of different folding pathways[157,168,169]. Bell's model was further refined by Evans and Ritchie who provided a theoretical framework deriving Kramer's equation for considering the mechanical unfolding as the diffusion of the protein by Brownian motion over the 1D energy landscape, in which the extension of the protein the reaction coordinate. This consideration would place the folded protein in a harmonic well where the protein needs to overcome high energy barriers and the friction of the energy profile[170]. During the last years many theoretical advances have been achieved not only for extract valuable data accurately from smFS experiments but also to correct the effect of the experimental set-up on the unfolding/folding rate of biomolecules[171–175]. Many of the advances and findings done during the last 20 years regarding protein mechanical folding/unfolding have been possible thanks to the data collected with smFS.

4.1.3 AFM for force spectroscopy studies on single-molecules: Atomic Force Spectroscopy (AFS)

In a typical AFS experiment the protein molecules are distributed over a surface placed on the top of a piezoelectric stage, the mechanical moving part of the tool. Above the protein sample the mechanically-

passive element of this set-up, the cantilever, is held. The cantilever is the probe which "fishes" molecules from the surface, proteins may get attached to it due to non-specific interactions like physisorption[176,177], keeping the protein held between the cantilever and the surface. In the easiest experiment the surface is moved apart from the cantilever at a constant speed value while stretching the protein (**Fig. 4.2**). Along the experiment the cantilever's deflection is monitored and this signal is converted into force units, making it possible to know the force which the protein is experiencing in every moment. When the force exerted by the protein on the cantilever is high enough, the last one bends because of the opposing force produced by the protein. However as the piezoelectric actuator is constantly retracting, the force is increasing and at some force value a protein domain unfolds, releasing the amino acid sequence of the protein trapped in the fold and producing a sudden drop in the tension experienced by the cantilever, returning this to its resting state. Usually the proteins tested in AFS are multimodular, so they are single polypeptides composed of a variable number of protein domains arranged one after the other in the chain. In constant speed pulling experiments the unfolding of these protein modules generates a saw-tooth pattern in which both the force (monitored through the cantilever) and the length (monitored through the piezoelectric displacement) are registered. In this pattern, each peak corresponds to the unfolding of one protein domain. These measurements allowed for the determination of the mechanical stability of many proteins suspected to experience forces *in vivo*[167].

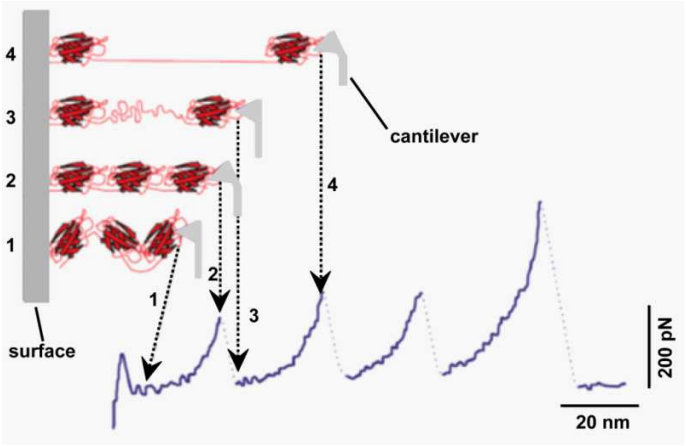


Figure 4.2. Constant speed pulling of a multimodular protein. A three-domain protein is attached on one end to a surface and to the other end to the cantilever tip. Cantilever moves away from the surface at constant speed. The saw-tooth pattern of peaks observed corresponds to the sequential unfolding of the individual domains. From step 1 to 2 the protein is stretched and aligned with the pulling force. In step 2, the cantilever bends as a consequence of the opposing force generated by the protein. In step 3 one of the domains unfolds unraveling the protein sequence trapped in the fold domain an increasing the length of the protein, returning the cantilever to its resting position. In step 4 the situation in step 2 is repeated, and so on. The last peak corresponds to the detachment of the protein from the cantilever. Adapted from reference[167].

Although these constant speed experiments were useful for finding out the mechanical stability of proteins and for correlating this feature with their secondary structure composition[178,179], these measurements are far from equilibrium and thus not physiologically representative. The development of the constant force measurement mode in smFS[180] allowed the application of well-controlled constant forces to proteins near the equilibrium, in the range of the experienced ones *in vivo*, allowing the reconstruction of the energy landscape of biopolymers from the kinetics of force-dependent unfolding. Besides, the use of constant force measurements together with the stability of AFS set-ups allow monitoring the trajectory of single molecules for

long periods of time. The application of different force protocols induce first the unfolding at high force and then favors the refolding at low force. Hence it is possible to monitor the folding process along a single reaction coordinate for several cycles[181].

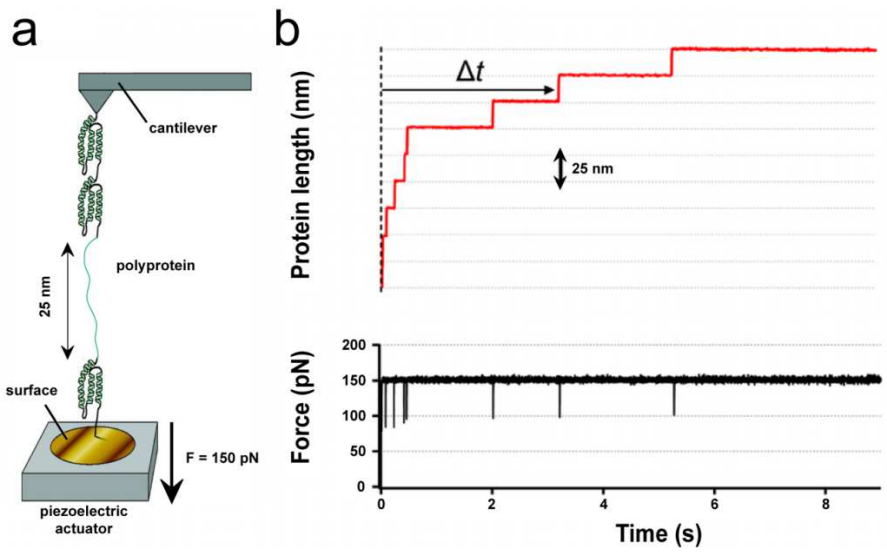


Figure 4.3. Constant force pulling of a multimodular protein. (**a**) A polyprotein is attached on one end to a surface and to the other end to the cantilever tip. Surface moves away from the surface, applying a constant force on the protein of 150 pN. Adapted from reference[181] .(**b**) The length of the protein is registered along time (upper panel), and every time a domain unfolds generates a sudden increase in the total length, giving this staircase pattern. Force (lower panel) is kept constant during the pulling thanks to a feedback electronic controller which restores the force exerted on the protein after each unfolding event. Adapted from reference[182].

4.2 Bacterial and viral attachment, a mechanical target

The current medical treatments target or interfere with the life cycle of the pathogen, and in many cases the drugs used are becoming useless or the treatments do not eradicate the problem completely. The first step

of infection consists in the recognition of the target cell and the attachment of the pathogen to its surface. Therefore the pathogen can use both its own proteins and the cell surface proteins of the host cell, establishing molecular interactions between them. Pathogen attachment happens in many different environments where the physical conditions can be very challenging, like the shear forces generated in the bloodstream or in the urine flow. This implies that pathogens have developed strategies to counteract these obstacles for the successful attachment and the subsequent infection. The proteins used for the attachment to the host are submitted to mechanical stress and as a consequence the pathogens have evolved surface proteins and strategies for counteracting the difficulties of a harsh physical environment. This thesis focuses on the study of the proteins used by UPEC and HIV-1 to attach to host cells with smFS.

Regarding UPEC, the proteins from the pilus type I used for urinary tract infection were studied (**Fig. 4.4**). Bacteria infecting the urinary tract are exposed to the mechanical stress imposed by the urine flow, and the pili are used for resisting these opposing forces. This appendage is interesting not only because of its mechanical role during infection, but also because the interactions maintaining the link between its hundreds of proteins is not covalent and relies on hydrophobic forces. With smFS we can interrogate the strength of the donor-strand interaction with the same geometry as the experienced *in vivo*. Our findings demonstrate that each individual subunit has a high mechanical stability correlated with their position in the pilus structure, following a mechanical hierarchy. These unfolding forces are among the highest ever reported, and the presence of a conserved disulfide

bond in these proteins plays a key role in their stability. The study of the strength of this hydrophobic interaction at the single molecule level could be a starting point for the development of drugs that mechanically weaken or interfere with this stable structure. A mechanically impaired pilus would ease the detachment of the bacteria from the urinary tract, being washed out during micturition.

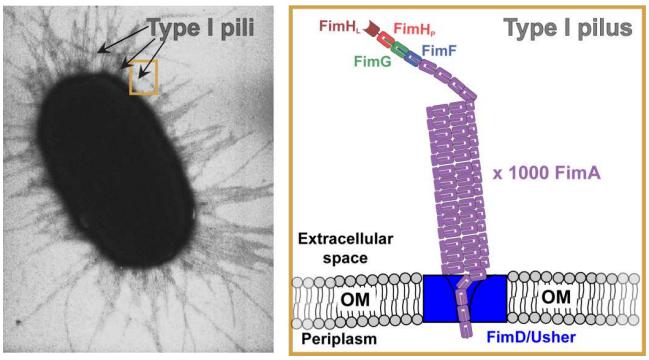


Figure 4.4. *E. coli* type I pilus. *E. coli* bacterium cells display a vast number of filamentous structures called pili protruding from its outer membrane (OM). A closer look into a single pilus shows a multiproteic structure in which hundreds of individual proteins interact head-to-tail through non-covalent interactions along the whole chain. *E. coli* cells use these filamentous structures to attach to surfaces. In the OM, a transmembrane protein called FimD/Usher is involved in the incorporation of proteins during pilus biogenesis and also it acts as a scaffold for the whole pilus. *E. coli* image from Manu Forero (DOI:10.1371/journal.pbio.0040314).

Additionally, the process of pilus biogenesis was also investigated. Pilus proteins are exported to the periplasmic space in an extended and unfolded conformation through the SecYEG pore, and with smFS we can resemble this state and investigate the effect of DsbA and FimC proteins at the single-molecule level. Our results indicate that FimC only recognizes disulfide-bonded domains and it helps them to fold. Surprisingly, DsbA is in fact not only inducing disulfide bond formation but also it contributes the most to pilus subunits folding, a new aspect never reported before in the type I pilus

system. Pilus biogenesis process could be an appropriate therapeutic target since several protein actors and cellular compartments are involved during the assembly of the pilus. Interfering with the chaperones and enzymes required for pilus proteins maturation would abort the pilus assembly, making bacteria unable to attach to the host cells.

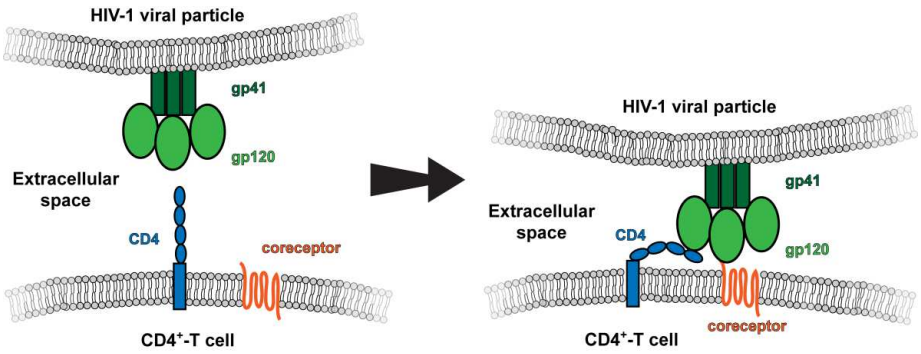


Figure 4.5. HIV-1 viral particle interaction with lymphocyte T-CD4$^+$ cell-surface proteins. HIV-1 viral particle exhibits macromolecular protein assemblies in its lipidic envelope, composed of trimers of the fusion protein gp41 and gp120. The first interaction that leads to viral infection implies the interaction of glycoprotein gp120 with cell-surface receptor CD4. The interaction of these two proteins leads later to an interaction with a membrane coreceptor present on the cell before membrane fusion and viral content release. This requires the approximation of the virus particle to the cell-membrane meanwhile attached to a CD4 protein through its gp120 glycoprotein. In this scenario CD4 should experience conformational changes for physically allow this approaching. Inspired from reference[119].

In the second part of this thesis the focus has been to study the nanomechanics of the first two domains of the cell surface protein CD4, the protein used by HIV to attach to lymphocytes T of the human immune system (**Fig. 4.5**). This is the first interaction of the virus with the cell and several events before virus fusion need to take place. The approaching of the virus to the cell membrane meanwhile bound to CD4 requires a high flexibility and, maybe, the mechanical unfolding

of some of the domains of this protein. It is not known but a viral particle in the fluid flow conditions of the blood or the lymphatic system could exert enough tension to produce the unfolding of the CD4 domains, facilitating the following events leading to viral and cell membrane fusion. We have discovered a hierarchical unfolding in the CD4D1D2 protein tandem which can be mechanically extended at very low forces, like the ones expected to be exerted by a viral particle. Based on previous research, we developed a mathematical model relating infectivity with CD4 mechanical extension, proving that a longer CD4 increases HIV infection. Besides, the mechanochemical reduction of CD4 internal disulfide bonds by thioredoxin was studied with smFS since it has been demonstrated that this chemical regulation increases the infectivity. Our data shows that the internal disulfides of CD4D1D2 are enzymatically reduced only after unfolding, suggesting that *in vivo* the mechanical unfolding of CD4 could be a requirement for this event. Finally, the effect of the HIV-blocking antibody Ibalizumab on the nanomechanics of CD4 was tested. It is known that some antibodies with the ability to bind CD4 reduce the infectivity of HIV. This blocking activity it is not completely understood and maybe the increase of the mechanical stability of CD4 when bound to these antibodies is the explanation for the antiviral activity of these antibodies. We found that Ibalizumab binding mechanically stabilized CD4D1D2, increasing the unfolding force of these two domains.

The mechanical study of CD4 and proteins related with its regulation or recognition could be a promising approach for developing therapies oriented to the mechanical stabilization of this protein. Preventing the attachment avoids the infection of new cells, having a potential use as a prophylactic and as a treatment for people living with

the infection. Not being a complete cure, it would make possible the restriction of the virus in the infected cells but avoiding its dissemination since no new infections could happen.

Chapter 5: Materials and methods

This chapter explains in detail the experimental procedures followed in order to conduct this research. It is divided in four different sections. The first section explains the Molecular Biology techniques and procedures applied for obtaining the different proteins used along this thesis. The second section shows the AFM set-up and its operation. The third section details the data analysis of the experiments done with the AFM. Finally, the fourth section gives an insight into the SMD simulations and its interpretation.

5.1 Molecular Biology

To conduct this research it has been necessary to design, to express and to purify different proteins. During this section all the procedures used starting from the designing of the DNA sequence up to the final proteins measured with smFS will be explained.

5.1.1 Protein design

In order to study the proteins with smFS it has been necessary to create chimeric constructs for most of them. This chimeric protein strategy consists mainly in combining in a single polypeptide chain different protein domains, generating a so called polyprotein. In this case, the

sequence of the proteins of study was flanked in both ends by two I91 protein domains. **Fig. 5.1** shows a cartoon of one of the used chimeric polyproteins.

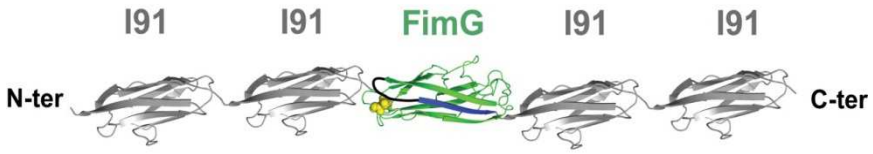


Figure 5.1. Chimeric polyprotein. Cartoon representation of the FimG construct (green, modified from PDB code: 3BFQ), being flanked by two I91 domains (grey, PDB code: 1TIT) on both sides. In yellow is depicted the disulfide bond present in FimG. The donor β-strand of FimF (blue) is connected to FimG through a four amino acid flexible linker (black line).

The I91 domain has been extensively studied with smFS hence its mechanical properties are well described[31,180]. Its unfolding when a single polyprotein is being pulled from its ends will produce a known fingerprint signal that unequivocally belongs to I91. The presence of this fingerprint is critical for smFS data collection and therefore it will be used both as a positive and a quality control of the experiments.

Five proteins were generated with this approach (the **Appendix** contains the protein sequences of all the proteins used). With the exception of the protein CD4 (its first two domains, CD4D1D2), the proteins FimA, FimF, FimG and FimH required a more elaborated design. These proteins are incomplete Ig-like folds lacking the seventh β-strand, and are therefore unstable. They are only stable *in vivo* because of the complementation of their fold with the donation of its cognate subunit β-strand through a process called DSC (see **Chapter 2: Type I pilus**). In order to study the strength of this DSC interaction circular permutant proteins were designed (see **Fig. 5.2**).

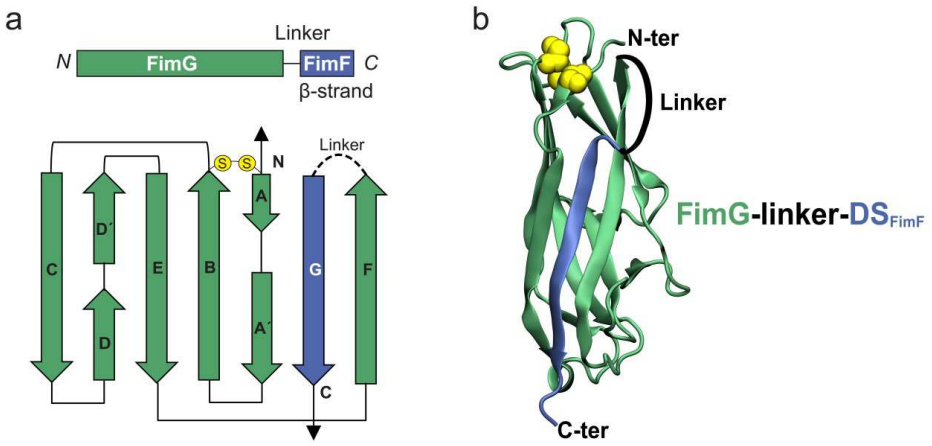


Figure 5.2. Self-complemented FimG. (**a**) Top diagram shows the FimG sequence (green) with the donor β-strand of FimF (blue) placed in its C-terminal position with a linker in between (black line). Below is represented the β-strand disposition of the FimG incomplete Ig-like fold with the FimF β-strand complementation. Disulfide bond is represented as two yellow circles connecting the A and B β-strands. (**b**) Cartoon representation of the FimG protein domain (green), self-complemented with the donor β-strand of FimF (blue). FimG domain and donor strand of FimF are covalently attached through a four amino acid linker (black). In yellow is depicted the disulfide bond present in FimG.

The cognate donor β-strand sequence of each subunit was covalently linked to its fold through a four amino acid flexible linker (DNKQ) that allows for the correct anti-parallel placement of the donor strand with respect to the incomplete Ig-fold which complements. Therefore, the four amino acid linker plus the sequence of the donor strand were added to each C-terminal Fim protein sequence (A, F, G and H). This strategy was already used in the past and it produced stable domains in solution[183,184]. Besides in order to avoid competition with the C-terminal β-strand added, the sequence of their own N-terminal donor β-strand was removed. In the **Fig. 5.3** are depicted the 3D structures of the self-complemented domains studied.

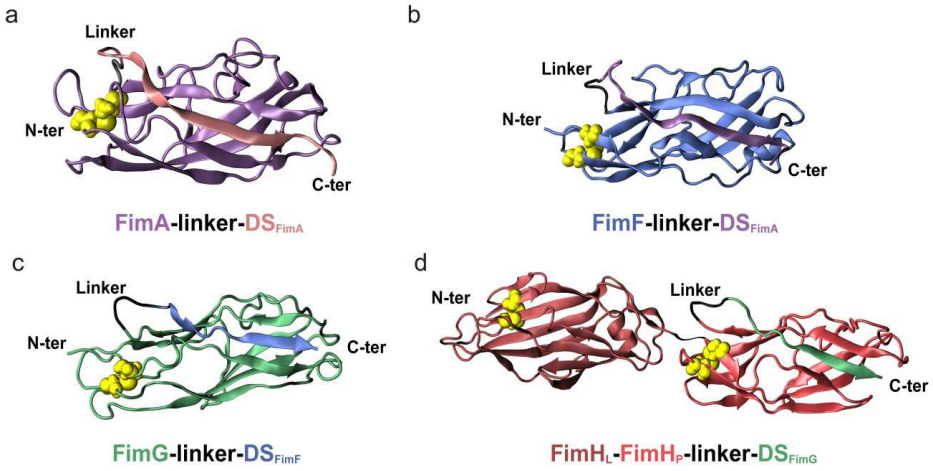


Figure 5.3. Self-complemented Fim proteins. (**a**) Self-complemented FimA (purple) with the donor β-strand of FimA (pink) placed in its C-terminal position with a linker in between (black line). (**b**) Self-complemented FimF (blue) with the donor β-strand of FimA (purple). (**c**) Self-complemented FimG (green) with the donor β-strand of FimF (blue). (**d**) Self-complemented FimH pilin domain (red) with the donor β-strand of FimG (green). FimH lectin domain is covalently linked to the pilin domain in its N-terminal position. All the complementary β-strands are placed in the C-terminal position of the domain and connected to it through a linker (black line). Cysteine residues are represented as yellow spheres, showing the presence of disulfide bonds in all of the domains.

5.1.2 Genetic engineering procedures

As a first step to study these proteins their DNA sequence is required. For that reason, the genes of the proteins were purchased from Life Technologies, which provided them inside of plasmids and with the sequence of the genes codon-optimized for their expression in *Escherichia coli*. Besides the genes with specific restriction sites placed along their sequence were ordered. These restriction sites will allow building the polyproteins, in several rounds of cloning and subcloning.

5.1.2.1 Commercial plasmid amplificationThe first step is to amplify the commercial plasmid containing the gene of interest. *E. coli* XL1-Blue competent cells (Agilent Technologies) were transformed with approximately 50 ng of DNA using the manufacturer protocol[185]. After transformation, the bacteria were grown in SOC media (Invitrogen) for 1 h at 37ºC and 250 rpm. Then, the cells were cultured over plates made of LB +Agar + Carbenicillin (Fisher) at 37ºC. After being grown overnight, single colonies were picked from the plates and grown in 5 mL of LB with 0.1 µg/mL of Carbenicillin for 12-16 h at 37ºC and 250 rpm. Cells were then pelleted centrifuging at 3,200xg for 10 min at 4ºC and plasmidic DNA was extracted from them with the QIAprep® miniprep kit (QIAGEN). Finally, plasmidic DNA was eluted and stored in nuclease-free water (Fisher) and the concentration was determined using the Nanodrop 2000L spectrophotometer (Thermo Scientific). **Fig. 5.4** indicates the main steps of this part.

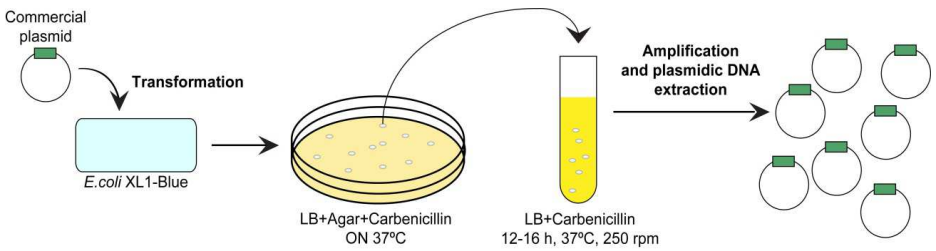


Figure 5.4. Commercial plasmid amplification. Commercial plasmid carrying the gene of interest is used for the transformation of *E. coli* XL1-Blue cells. Cells are grown and selected in the presence of antibiotic and then single colonies are isolated for growing in liquid media also with antibiotic. After several hours, the population of bacteria has increased and then plasmidic DNA extraction is performed for isolating amplified commercial plasmid in high amounts.

5.1.2.2 Commercial plasmid digestion

All of our genes have a *BamH*I restriction site in their 5' end and a *Kpn*I restriction site in their 3' end. At the end of the gene and just before the *Kpn*I site, we have the amino acid sequence RSCC and the stop codon. This four amino acid sequence is flanked in its 5' end by a *Bgl*II site and in its 3' end by the above mentioned *Kpn*I (**Fig. 5.5**). With this gene architecture, we can build our polyproteins splicing the different genes in several rounds of DNA digestion and ligation.

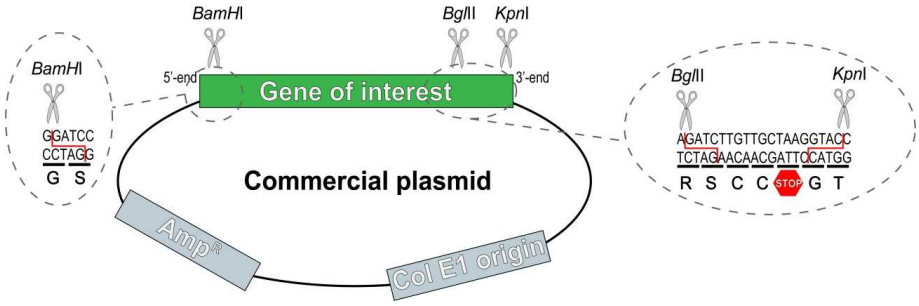


Figure 5.5. Scheme of a commercial plasmid with the gene of interest. A typical commercial plasmid contains three main sequences. One conferring resistance to antibiotic treatment (Amp^R, for ampicillin), another one for plasmid DNA replication (Col E1 origin) and the third one being the gene of interest. In our case, the gene of interest possesses three characteristic restriction sites in its sequence. In the 5'-end of the sequence it has a *BamH*I restriction site and in the 3'-end it has a *Bgl*II and a *Kpn*I restriction sites. The insets show the DNA sequences cut by the different restriction enzymes and their translation into amino acid letter code. *BamH*I and *Bgl*II digestion produce compatible ends between the sequences. Between the *Bgl*II and *Kpn*I restriction sites there are two cysteines and the codon stop. When building a polyprotein with different genes with these architecture, every time we splice two sequences only the second one (in 3'-end) will have the cysteines and the stop codon. In this way the last protein domain of the polyprotein (C-terminal) will have the two cysteines and it will help for protein immobilization over the gold substrate used in AFM experiments.

Commercial plasmid was digested with *BamH*I/*Kpn*I restriction enzymes (Thermo Scientific) following the FastDigest protocol[186]. Digestions were done on approximately 3 µg of total DNA in a volume

of 50 µL for 30' at 37ºC. After digestion, gene insert and commercial plasmid were separated through electrophoresis in agarose 1% gel in Tris-Acetate-EDTA Buffer (Fisher) with ethydium bromide (Sigma). After electrophoresis, gene DNA was extracted from the agarose gel with the QIAquick® Gel Extraction Kit (QIAGEN) and DNA insert was eluted and stored in nuclease-free water and its concentration was determined with a spectrophotometer.

5.1.2.3 Expression plasmid digestion

Once the sequence of the gene of interest has been isolated, the next step is the opening of an expression plasmid, suitable for protein expression in bacteria. The expression plasmid pQE80L (QIAGEN), a kind gift from Julio M. Fernández (Columbia University) was used.

Fig. 5.6 shows a scheme of the pQE80L plasmid. Besides the gene providing resistance to antibiotic treatment and the sequence for the origin of replication, the plasmid possesses the Multiple Cloning Site (MCS) where several restriction sites are placed for different restriction enzymes digestion allowing for the insertion of the gene of interest. Plasmid pQE80L also contains a short sequence encoding for six histidine residues just before the region where the gene of interest will be inserted. This feature will be helpful later during the protein purification process.

Upon gene of interest insertion, this gene will be under the control of the *lac* promoter. The plasmid pQE80L carries itself the gene that encodes the *lac* promoter repressor, *lacI*. The repressor is constitutively expressed by the cell and it binds to the *lac* promoter, inhibiting the expression of the gene the promoter is regulating. When

bacteria are grown in the presence of lactose or an analogous compound, the repression is interrupted, allowing for the expression of the gene under the control of the *lac* promoter.

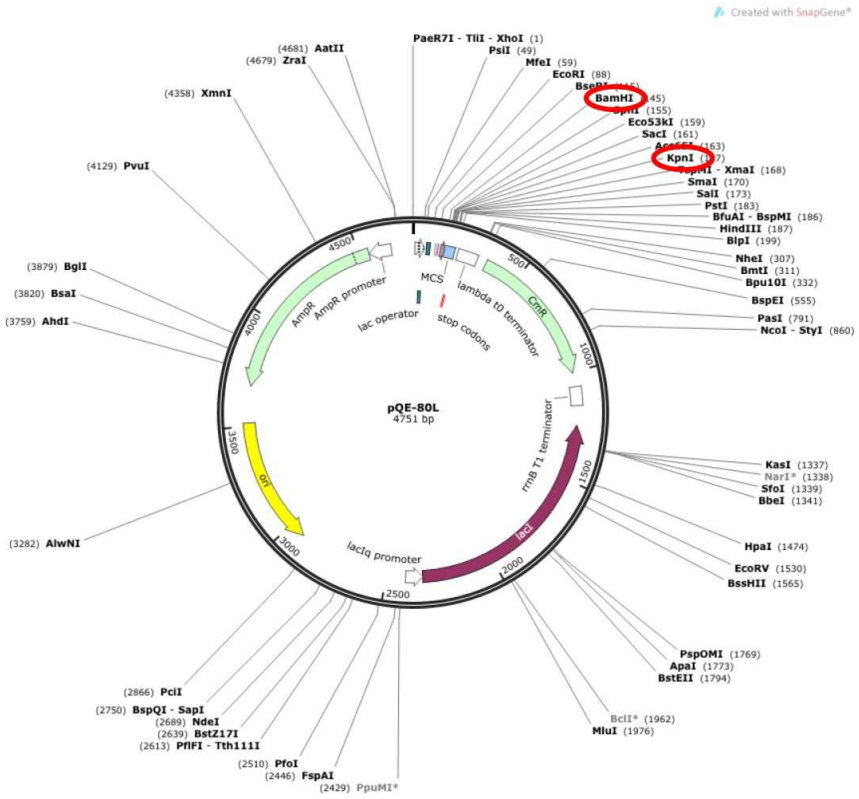


Figure 5.6. pQE80L expression plasmid. Scheme of the expression plasmid pQE80L showing its main features. *BamH*I and *Kpn*I restriction sites are highlighted with a red circle. Gene of interest will be inserted in this plasmid between the sequences of these two restriction sites. Image generated with the software SnapGene® (Marie Fertin).

In order to insert the gene of interest the expression plasmid pQE80L was opened with the same restriction enzymes used for cutting the gene of interest, *BamH*I/*Kpn*I. The same procedure was used as explained for the commercial plasmid, digesting the plasmid with these two enzymes, separating the DNA fragments in an agarose gel electrophoresis and purifying the plasmidic DNA.

5.1.2.4 Gene of interest and expression plasmid ligation

Both the gene of interest and pQE80L are digested with the same enzymes, so the ends of both DNAs are compatible between them. Like the restriction enzymes used generate a different cohesive end, the insertion of the gene will be directional and the expressed protein will have the desired sequence of amino acids.

For inserting the gene of interest inside of pQE80L both DNA sequences were ligated. For that purpose a reaction mixture was prepared in which the main component is the T4-DNA ligase (Invitrogen), an enzyme that catalyzes the formation of the phosphodiester bond between two DNA fragments. Following the protocol specified by the manufacturer[187], a molar ratio 3:1 of insert with respect the plasmid was used, following the next formula:

$$ng\ of\ insert = 3 * \frac{Plasmid\ (ng) * Insert\ (bp)}{Plasmid\ (bp)}$$

In which *ng* specifies the mass in nanograms an *bp* refers to the number of base pairs in the sequence of the DNA fragments involved. Ligation reactions were incubated at room temperature (r.t.) O.N. and the ligation reaction was stopped diluting the reaction mixture five times with nuclease-free water.

5.1.2.5 Expression vector-insert amplification

After the ligation of the gene of interest with the expression plasmid is done, an amplification step is needed. The same procedure as the one described in the subsection 2.1.2.1 was done.

5.1.2.6 Chimeric gene building

In order to create a polyprotein it is needed to splice different gene fragments with the restriction site digestion/ligation strategy. The procedure used for this goal is the same as the one explained above.

For explaining the chimeric DNA building the polyprotein made for FimG will be used as an example. This polyprotein consists in four copies of the I91 domain flanking, two in each side, the FimG protein, and it is referred as $(I91)_2$-FimG-$(I91)_2$. **Fig. 5.7** summarizes the whole process for creating this chimeric gene.

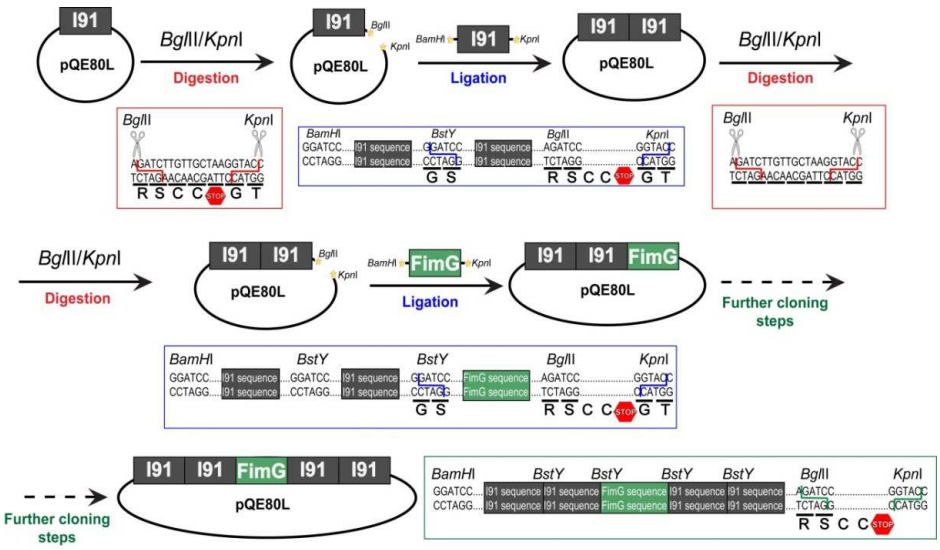


Figure 5.7. Chimeric gene building process. In this diagram it is explained how the polyprotein $(I91)_2$-FimG-$(I91)_2$ is built by splicing the different genes that compose it. Red lines in between the sequences mean digestion pattern meanwhile the blue lines mean ligation between cohesive ends. Below the DNA sequences the codons are indicated and below the amino acids they code. Starting with pQE80L plasmid with one copy of the gene I91, all the steps consist in opening the plasmid digesting with *Bgl*II/*Kpn*I and ligating the next gene previously digested with *BamH*I/*Kpn*I. Every time a gene is fused to the construct the ligation of the cohesive ends *BamH*I/*Bgl*II generates a new restriction site, *BstY*. The last gene added to the sequence will be translated with the last two cysteines and the stop codon.

For this purpose the explanation will start with the pQE80L-Insert construct from subsection 5.1.2.5 (pQE80L-I91), so the gene of interest inserted was the one encoding the I91 protein.

For introducing another copy of the I91 gene after the first one the plasmid needs to be opened again with restriction enzymes. Digesting with *Bgl*II/*Kpn*I enzymes generates *Bgl*II-pQE80L-I91-*Kpn*I. This digestion opens the plasmid an releases a small fragment that encodes for the last two amino acids (two cysteines) and the stop codon. Digesting again the I91 gene with *BamH*I/*Kpn*I will generate the insert *BamH*I-I91-*Kpn*I. *BamH*I and *Bgl*II generate compatible ends, so two different sequences one digested with *BamH*I and the other one with *Bgl*II will have compatible ends and they can be ligated. Upon ligation of *Bgl*II-pQE80L-I91-*Kpn*I with *BamH*I-I91-*Kpn*I the pQE80L-$(I91)_2$ construct will be generated. The ligation of cohesive ends between fragments digested with *BamH*I and *Bgl*II generates a new restriction site called *BstY*. After this, adding a new gene after the previous ones consists in repeating the same steps. Upon digestion of the FimG gene with *BamH*I/*Kpn*I and the digestion of pQE80L-$(I91)_2$ with *Bgl*II/*Kpn*I, these two fragment can be ligated again and so on for the last two copies of I91.

When the last of the gene copies has been added to the plasmid, the whole construct is amplified the same way as it was explained before in subsection 5.1.2.1.

5.1.3 Protein expression and purification

5.1.3.1 Expression test of the chimeric protein

Once the chimeric gene has been built inside of the expression plasmid, the next step consisted in transforming bacterial cells suitable for protein expression. Around 50 ng of the plasmid were used for transforming *E. coli* BL21 (DE3) cells (Agilent) following the manufacturer instructions[188]. Bacteria then were grown with 450 µL of SOC medium for 1 h at 37ºC and 250 rpm. After, the cells were spread out over LB + Agar + Carbenicillin plates and grown O.N. at 37ºC. Several colonies were isolated for protein expression test and grown in 10 mL of LB + Carbenicillin until the optical density (OD) at 600 nm reached a value of 0.6. Then, the volume was splitted in two and one the volumes was treated with 1 mM of Isopropyl-β-D-thiogalactoside (IPTG, Sigma) for protein expression induction. IPTG is a chemical compound analogous to lactose, which inhibits the suppression by the *lac* inhibitor of the *lac* operon under whose control is the gene of interest. Treated and non-treated (control) cultures were incubated O.N. at 37ºC and 250 rpm. After this incubation period, 1 mL of IPTG-treated and control cultures were centrifuged and cells were pelleted. After discarding the supernatant, cell pellets were resuspended in 10 µL of Extraction Buffer (50 mM sodium phosphate pH 7.0, 300 mM NaCl) and then 10 µL of Laemmli sample buffer (Biorad) were added to the mixture. Samples were then boiled for 5 min and centrifuge at 12,000xg for 30'. Upon this treatment bacterial proteins were released and denaturalized. Following with a Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) it is possible to separate the proteins by their size. 10-20 µL of protein sample were

loaded into polyacrylamide gels (stacking 4%, resolving 8-12%) and electrophoresis was done in the Biorad electrophoresis system for 2h at 100 V in running buffer solution (Biorad). After separation, the gels were rinsed three times in milliQ water and proteins were stained with the Bradford solution (Thermo Scientific) for 30 min. Then the gels were rinsed again in water and a picture was taken with the Kodak Image Station 4000R. For every colony tested the control load was placed next to the IPTG-treated. Overexpression of the chimeric polyprotein appears as an intense band only in the IPTG-treated samples at the expected size for the construct. **Fig. 5.8** summarizes the main steps for protein extraction.

The colonies which display the highest protein expression were selected for large-scale protein production. In order to keep the colonies in the long-term the non-induced cultures were aliquoted an kept in 10% v/v of glycerol (Fisher), flash frozen in liquid nitrogen and stored at -80ºC.

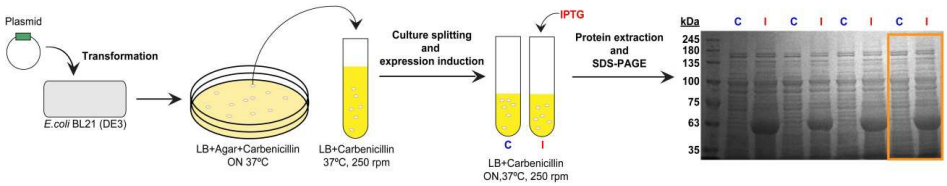


Figure 2.8. Protein expression test. Diagram of protein expression test. Bacteria are transformed with the plasmid and then are plated in selective media. Single colonies are grown in liquid selective media and protein expression is induced in half of the culture, remaining the other half as a control. SDS-PAGE shows the molecular weight ruler (left) and then the controls (C) and the IPTG-treated (I) samples of four individual colonies. Orange rectangle indicates the control and the induced sample of a colony, highlighting the intense band at approximately 60 kDa, corresponding with the size of the polyprotein.

5.1.3.2 Large-scale protein production

In order to conduct AFM experiments a large scale protein purification protocol needs to be applied. For all the proteins studied in this thesis the same protocol was used for protein expression and purification. Only $(I91)_2$-CD4D1D2-$(I91)_2$ was slightly different during the protein expression and DsbA both production and purification were completely different. These differences will be noted along the explanation.

Large-scale culture was induced with one frozen cell aliquot in 800 mL of LB in the presence of Carbenicillin (0.1 µg/mL) and Chloramphenicol (0.05 µg/mL). *E. coli* BL21 (DE3) cells besides its genomic DNA and the expression plasmid introduced for polyprotein expression, carry another plasmid that confers resistance to the antibiotic Chloramphenicol. Culture was incubated at 37ºC and 250 rpm until $DO_{600} > 0.6$ and then IPTG (1 mM) was added for protein expression induction ON in the same conditions.

$(I91)_2$-CD4D1D2-$(I91)_2$ polyprotein was not transformed into BL21 (DE3) cells, instead it was transformed in *E.* coli NEB 5-α (New England Biolabs) following the instructions provided by the company[189]. This strain was chosen since it is the only one able to produce $(I91)_2$-CD4D1D2-$(I91)_2$ in a sufficient quantity suitable for single-molecule experiments. After NEB 5-α cell transformation and plating, a single colony was selected for large-scale culture in the same conditions as explained above but without Chloramphenicol. Upon IPTG adding, protein expression took place ON at 20ºC and 250 rpm. In the case of DsbA, the gene encoding this protein was inside of the plasmid pET11a. For DsbA, the large-scale culture was induced as

explained above and after IPTG induction the culture was incubated ON at 25ºC and 250 rpm.

5.1.3.3 Protein extraction

After o/n protein expression cells were pelleted centrifuging the culture at 4,000xg for 20 min at 4ºC. Supernatant was discarded and cells were resuspended in 16 mL Extraction Buffer (50 mM sodium phosphate pH 7.0, 300 mM NaCl) in the presence of 160 µL of protease inhibitor (Merck Millipore). Then, cells were exposed to a chemical lysis at 4ºC with 30 min incubation in the presence of 100 µg/mL of lysozyme (Thermo Scientific) and gentle shaking (5 rpm). Then, cell resuspension was incubated for 10 min in the presence of Triton X-100 1% (Sigma Aldrich) for cell membrane destabilization, DNAse I (Invitrogen) and RNAse A (Ambion) at 5 µg/mL each for enzymatic nucleic acid degradation, and 10 mM of $MgCl_2$ for optimal performance of these nucleases. After 10 min with gentle shaking, a mechanical disruption of the cells was performed in order to release all the soluble proteins into the media. For this purpose it was used a French press (G. Heinemann HTU DIGI-F Press), a tool that allows for the application of high pressures on the cells. A valve on the bottom of the chamber allows for the regulation of the exit flow. Manually regulating the flow to the lowest speed possible (around 0.5 mL/min) the cells disrupted due to the high shear forces generated were collected. As a consequence, the cell lysis mixture obtained contains the proteins of interest and cell debris. In order to separate the soluble fraction with the proteins from the debris 1h 30 min centrifugation at 33,000xg at 4ºC was performed. Then the pellet was discarded and the

supernatant was filtrated using three consecutive filters of 0.8, 0.45 and 0.22 µm (Merck Millipore).

DsbA protein extraction was performed with an osmotic shock[190,191]. After protein expression, cells were pelleted (10,000xg, 10 min, 4ºC) and resuspended in Extraction Buffer. Then cells were pelleted again and resuspended in Ice Cold Shock Buffer (50 mM Tris pH 8.0, 20% w/v Sucrose, 1 mM EDTA) and gently rocked for 10 min at 4ºC. After a third centrifugation, cells were resuspended and incubated in Ice Cold Leak Buffer (10 mM Tris pH 7.1). Then the cells were centrifuged and the supernatant was the periplasmic extract from the cells. The transference from a highly concentrated buffer to a very diluted one makes possible the periplasmic release of content, where DsbA is exported after translation in the cytoplasm.

5.1.3.4 Protein purification

This filtered supernatant contained the protein of interest and all the soluble proteins that bacteria produce constitutively. Further purification steps were performed taking into advantage the presence of six consecutive histidine residues in the N-terminal part of the protein of interest, added during the translation of the protein from the pQE80L plasmid (see subsection 5.1.2.3). Supernatant was then incubated for 1 h at 4ºC and gentle shaking (5 rpm) in the presence of 3 mL of HisPur Cobalt Resin (Thermo Scientific). This resin contains cobalt divalent ions which bind polyhistidine-tagged proteins. After several rounds of resin cleaning with Extraction Buffer, the protein was released from the resin with Elution Buffer (50 mM Sodium Phosphate pH 7.0, 300 mM NaCl, 150 mM Imidazole). Elution Buffer contains imidazole, a compound that competes with histidine for the binding to cobalt. After

protein solution collection, disulfide bond oxidation or reduction was triggered o/n at RT in the presence of 0.1% of H_2O_2 or 10 mM of DTT, respectively (both from Sigma Aldrich).

A further purification step was done through a size exclusion chromatography in an ÄKTA pure (GE Healthcare) fast protein liquid chromatography (FPLC). Here a Superdex 200 column (GE Healthcare) was used for protein separation by size. Protein elution was done in HEPES buffer pH 7.2, NaCl 150 mM and EDTA 1 mM. A typical chromatogram from a FPLC run depicts the absorbance at 280 nm (A_{280}) for protein detection *vs* the volume eluted through the column (**Fig. 5.9**). Proteins, based on their size, elute at a particular column elution volume and the fractions collected coming from the best absorbance peak were aliquoted, flash frozen in liquid nitrogen and kept at -80ºC until their use in the AFM.

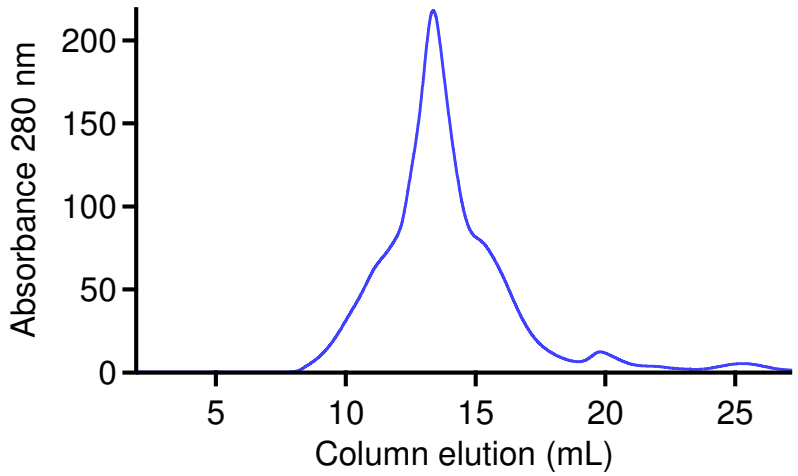


Figure 5.9. Size exclusion chromatogram. This picture depicts a size exclusion chromatogram from the purification of the protein $(I91)_2$-FimA-$(I91)_2$. This protein elutes at ~13.5-14.0 mL, and the fractions collected from the center of the peak are the ones used for AFM experiments.

In the case of DsbA, after periplasmic extraction, the extract was loaded into a 5 mL HiTrap Q Sepharose FF column (GE Healthcare) for anion exchange. Instead of separating proteins by size, this time proteins were separated by electric charge. Negatively charged proteins are retained in the positively charged resin of the column. Upon DsbA binding to the resin, the protein was released from the resin with a linear increase gradient of NaCl, starting with Tris 20 mM pH 8.0 and ending with Tris 20 mM pH 8.0, 1 M NaCl. DsbA elutes around 250-350 mM of NaCl. After ion exchange, the aforementioned size exclusion protocol was applied.

5.2 Single-molecule Force Spectroscopy

Along this section the AFM used in this thesis (Luigs & Neumann), and the different modes of operation applied for the mechanical study of single proteins will be explained. The graphical interface used for AFM operation, the data collection, and the analysis software were a custom-written program for Igor Pro 6 software (Wavemetrics).

5.2.1 AFM set-up

An AFM is composed of three main parts: the piezo-electric actuator (PZA), the cantilever and the photodetector (PD) (**Fig. 5.10**).

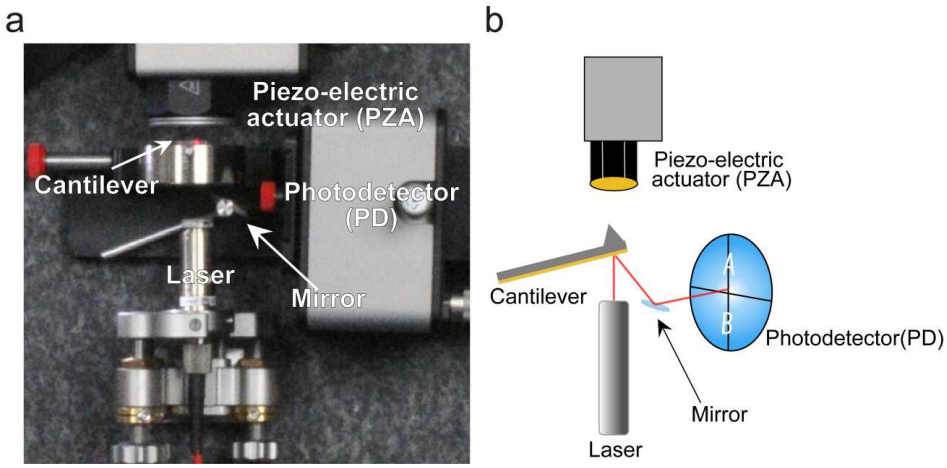


Figure 2.10. AFM set-up. (**a**) Picture of the main AFM parts (modified from http://www.luigs-neumann.com/sites/default/files/_DSC5708.JPG). (**b**) Scheme of the AFM. The laser beam is focused on the back of a cantilever whose back side is covered by a gold layer. The laser is reflected to a mirror which reflects the laser beam to a photodetector (PD). Over the PZA is placed the gold surface with the protein sample adsorbed over it.

The PZA is the active part of the AFM. It is made of a material which shows piezo-electric effect (**Fig. 5.11**). Piezoelectricity is a property that consists in the ability of a material of generating an electric potential once it is mechanically stressed. This effect is related with the presence of electric dipoles in the crystal lattice of the material, and the application of a mechanical stress on the material changes the polarization of these dipoles. Piezoelectric materials exhibit also the inverse property; applying an electric field to these materials leads to mechanical deformation of their structure. This is known as the inverse piezoelectric effect[192] , and it is this property the one exploited in an AFM. The application of a known voltage leads to a well-controlled mechanical disturbance. In the AFM used the PZA can be displaced in its X and Y directions with nm precision. Even more, the precision of the movement in its Z-direction is in the sub-nanometer range. In smFS experiments where the molecules are stretched

vertically this sub-nanometer resolution is crucial for the precise mechanical study of proteins.

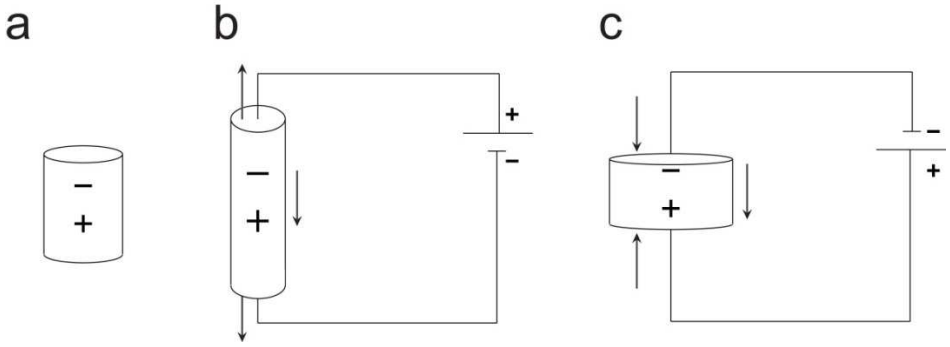


Figure 5.11. Inverse piezo-electric effect. Diagram (**a**) depicts an unperturbed piezo-electric material with its dipole charge. Diagrams (**b**) and (**c**) show how the application of an opposite electrical potential deforms mechanically the piezo-electric material.

The cantilever constitutes the probe with which the single proteins will be captured from the surface. In this thesis two types of cantilevers were used, MLCT and OBL-10 cantilevers (both from Bruker). **Fig. 5.12** shows pictures of both cantilevers. From MLCT, we used the triangular cantilevers C and D and from OBL-10 we used the rectangular A and B cantilevers. Both types are made of silicon nitride and MLCT cantilevers backside had a Ti/Au 60 nm layer, meanwhile OBL-10 cantilevers are covered on both sides including the tip with a 20-30 nm Au layer. Each of the cantilevers has its own dimensions and properties, and they are used differently depending on the operation mode of the AFM used.

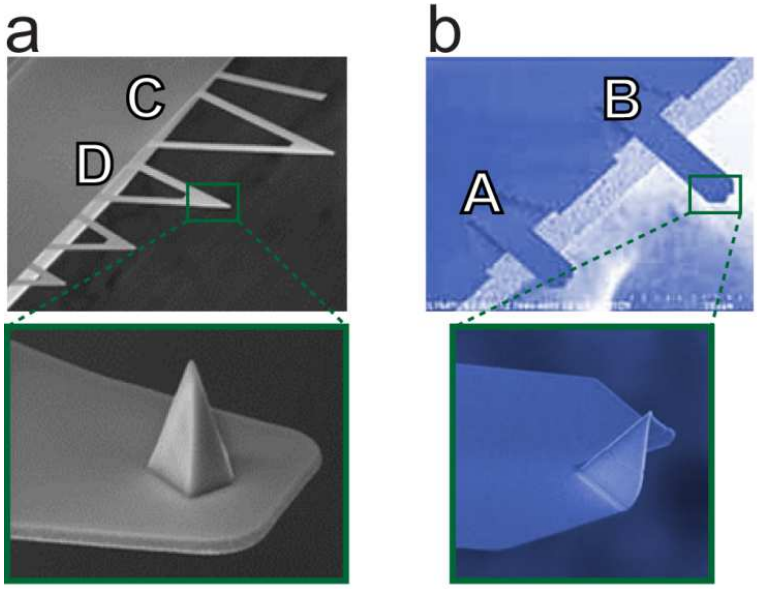


Figure 5.12. AFM cantilevers. (**a**) Top picture shows the front side of the MLCT chip where the cantilevers are placed, highlighting the C and D triangular cantilevers used. Bottom picture shows a close view of the tip of the cantilever. (**b**) OBL-10 chip front view with the used A and B rectangular cantilevers. Bottom picture shows a close view of the tip of one of the cantilevers (Images from http://www.brukerafmprobes.com).

In an AFM, the cantilever deflection is detected as a change in the reflection angle of the laser beam pointing to its backside. The reflection is detected in the PD, an optical device that converts the detected light (laser) into an electric signal (**Fig. 5.10**). The PD in fact is separated in two photodiodes, the upper and the lower one. Each time the cantilever is bent the laser beam will change its reflection angle and one of the halves of the PD will receive more light. The difference of the light registered by each half (upper PD registers a positive signal meanwhile de lower one is negative) generates an output electrical signal that corresponds with the bending of the cantilever.

5.2.2 AFM calibration and sample preparation

In order to make an AFM experiment, it is necessary to place the protein sample over a suitable surface for protein adsorption. We fabricate custom-made substrates consisting of a 40-50 nm gold layer above a 4 nm layer of nickel over a 2 mm borosilicate substrate (brand). Nickel and gold layers are deposited with an Oerlikon 450B Evaporation System (Leybold). The gold substrate is then glued with vacuum grease (Dow Corning®) to the top part of the PZA (**Fig. 5.10**). Once the substrate is ready, we place a drop (10-15 μL) of the protein under study at a final concentration of 5-30 μM. We incubate the protein for 15-20 min, time during which the two last cysteines of the C-terminal end of the polyprotein can make a covalent bond with the gold surface and be immobilized (see **subsection 5.1.2.6**). After the incubation, the sample is rinsed several times with AFM buffer (10 mM HEPES buffer pH 7.2, 150 mM NaCl, 1 mM EDTA) to remove the non-attached molecules.

The cantilever is held inside a chamber and submerged in AFM buffer. The liquid chamber is a circular structure placed on a crystal that allows the passage of the laser beam through it. The chamber is laterally delimited by a silicon O-ring and the whole liquid cell is over a metal holder. Once the cantilever is correctly placed, the laser beam is focused pointing close to the end of the cantilever, near the place where the tip is.

Then the PZA position is placed upside-down so now is facing the cantilever. The PZA is approached to the liquid cell and once it is in contact with the hydrophobic O-ring the liquid cell is sealed and water

evaporation is minimized. Now the experiment calibration can be started (**Fig. 5.13**).

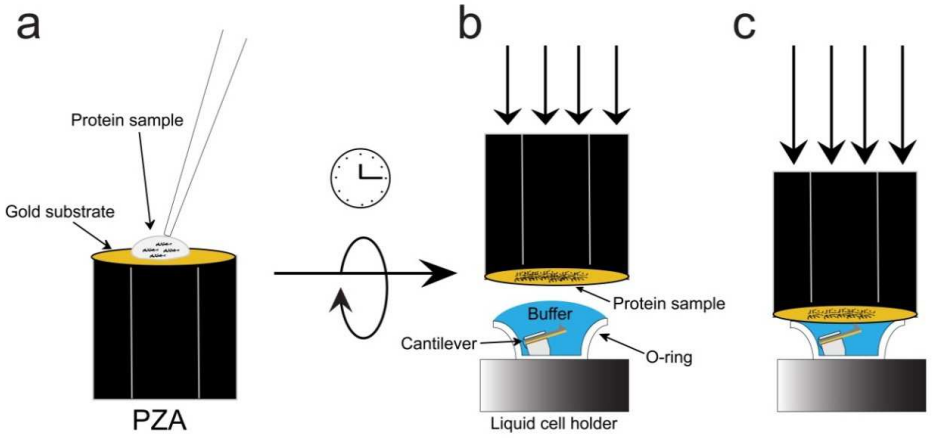


Figure 5.13. Sample preparation and PZA approach to liquid cell. (**a**) Addition of a 10-15 µL drop of protein on the gold substrate placed over the PZA. (**b**) After 15 min of incubation, sample is rinsed and the PZA-gold-sample set is turned upside-down, facing the liquid chamber where the cantilever is immobilized. Then the PZA is motor-approached to the liquid cell. (**c**) When the gold substrate is in contact with the O-ring, the chamber is sealed and water evaporation is minimized.

As it was mentioned in the previous section, the PD registers the bending of the cantilever and converts light into an electrical signal. This signal is given in volts, and it is needed to convert these volts into force units. Assuming the cantilever deflection (z_c) follows Hooke's law (2), the spring-constant (k) can be calculated for small bending distances.

$$F = -kz_c \tag{2}$$

We use the Equipartition theorem (3) to calculate k[193]. This theorem relates the thermal motion of the fundamental oscillation mode (ω_0) of the cantilever with its thermal energy

$$\frac{1}{2}k\langle {z_c}^2\rangle = \frac{1}{2}k_B T \;\; ; \;\; k = \frac{k_B T}{\langle {z_c}^2\rangle} \tag{3}$$

Were k_B is the Boltzmann constant, T is the absolute temperature and $<z_c^2>$ is the variance of the cantilever's Z-position registered by the PD at equilibrium. In order to calculate $<z_c^2>$, the power spectral density (PSD) of the Z-oscillations of the cantilever is calculated along a certain amount of time and the area under the peak of the first oscillation mode. $<z_c^2>$ is in fact calculated as $<\Delta V^2>$ being ΔV=A-B (a difference in the registered voltage between the upper and lower PD).

These calculations take place when the cantilever and the PZA are far away from each other (**Fig. 5.14a**), but in order to calculate the cantilever k it is required to calculate the slope that relates cantilever's deflection (ΔV) with the PZA displacement (z_{PZA}). And for that, the PZA and cantilever have to be brought together. When they are close enough, a force curve is done (**Fig. 5.14b**). When the PZA is moved in Z-direction and the cantilever is pressed, this is deflected and the angle of the laser beam detected by the PD changes (ΔV). Cantilever bending in Z-direction (Δz_c) is the same as the PZA Z-displacement (Δz_{PZA}). Now it is possible to calculate the slope ($s = \Delta V/\Delta z_{PZA}$), which can be used for finding out k from formula 3,

$$k = \frac{k_B T}{\langle {z_c}^2\rangle s^2} \tag{4}$$

Now, it is possible to calculate how much force (F) is being applied to a protein since the cantilever stiffness (k) and its displacement (z_c) are known.

Each cantilever *k* is calculated at the beginning of each new experiment. Typical spring constants are: for MLCT C cantilever, 10-20 pN·nm^{-1}; for MLCT D, 30-60 pN·nm^{-1}; for OBL-10 A cantilever, 9-30 pN·nm^{-1}; and B 2-6 pN·nm^{-1}. MLCT C and OBL-10 B cantilevers are used for force-clamp measurements, which require softer and more sensitive probes. MLCT D and OBL-10 A cantilevers are used for force-extension measurements.

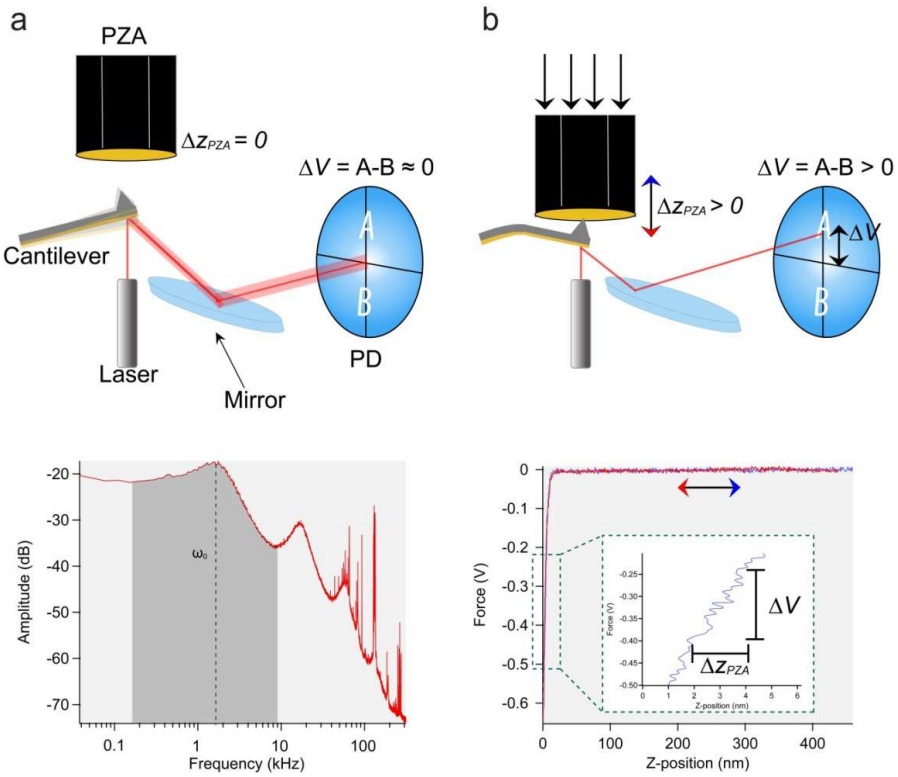


Figure 5.14. Cantilever calibration. (**a**) Cantilever's thermal spectrum is calculated at equilibrium and separated from PZA. For a certain time, cantilever's Brownian movement in Z-axis is registered in the PD as slight changes in the position of the incident light (upper scheme). Then, the power spectrum density (PSD) of this signal is calculated and the area under the fundamental oscillation mode (shaded area) is calculated too (lower graph). Finally $<\Delta V^2>$ is calculated. (**b**) For *k* calculation a force curve is done (lower graph). The PZA movement bends cantilever, changing the incident light on the PD (ΔV). Cantilever's Z-displacement it is the same as the PZA displacement ($\Delta z_{PZA}=\Delta z_c$) (upper scheme). Knowing the cantilever

displacement and the voltage change, now we can calculate the slope ($s = \Delta z_c/\Delta V$) (lower graph) and then the k of the cantilever solving the Equipartition theorem. Red arrow means the approach of the PZA to the cantilever, pressing it. Blue arrow means the retraction of the PZA from the cantilever. In the lower graph, the red approaching trace has been removed from the inset. In both upper schemes the liquid cell has been removed for the sake of simplicity.

5.2.3 AFM operational modes

The proteins studied with smFS in this thesis were measured in three different operational modes, each one of them for a specific purpose. These modes are called force-extension (FX), force-clamp (FC), and force-ramp (FR).

5.2.3.1 Force-extension (FX)

FX measurements consist in pulling a single protein at a constant speed. In this measurement the AFM is constantly doing force curves, with a cycle of PZA approaching to the cantilever, pressing it up to a specific force and then retracting. The typical FX data collected has a sawtooth pattern in which each peak corresponds to the unfolding of one domain. An example FX trace is depicted in **Fig. 5.15** from a real measurement of $(I91)_2$-FimF-$(I91)_2$ polyprotein.

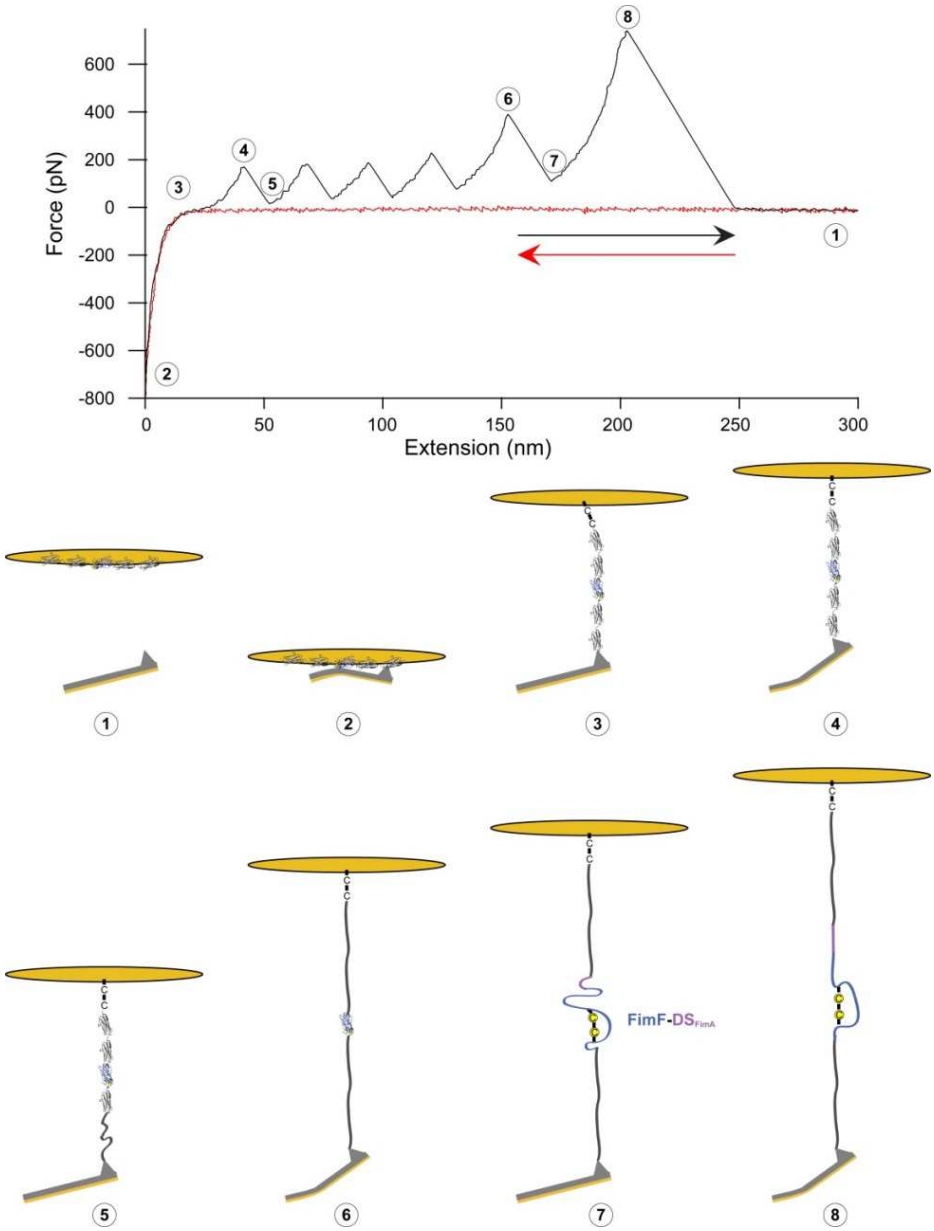


Figure 5.15. FX trace of $(I91)_2$-FimF-$(I91)_2$ polyprotein. The PZA approaches (red trace) to the unperturbed cantilever (1). Afterwards, the PZA pushes the cantilever (2). Then, the PZA retracts (black trace) from cantilever. When one polyprotein is adsorbed to the tip of the cantilever, the molecule becomes raised from the surface up to its C-terminal cysteines (3). Once the molecule is stretched up to its maximum length, the cantilever starts bending because of the tension, increasing the force (4). Force rises up to ~200 pN and one of the I91 domains unfolds (5). The unfolded I91 becomes completely stretched and again force starts to increase

until the next I91 domain unfolds, and so on. When all the I91 domains are unfolded FimF self-complemented domain unfolds, since this domain is mechanically more stable than I91 domains. Force rises up ~400 pN (6) and then FimF unfolding takes place (7). PZA keeps on retracting up to the complete stretching of the unfolded protein polymer up to the disulfide bond of FimF, increasing then the force again (8). Once the tension is above the strength of the interaction between cantilever's tip and protein, the molecule is released from the cantilever and the force drops to zero. This last force peak corresponds with the detachment.

Proteins are immobilized through their C-terminal end to the gold surface, so every time the PZA is pressing and bending the cantilever a protein molecule could get attached to the cantilever's tip. When the retraction of the PZA starts, the cantilever bending gradually disappears, recovering its resting position. If a protein molecule got attached to cantilever's tip, it will be held between the tip and the gold surface. As the PZA is further retracting at a constant speed and it is linked to the cantilever through a protein, the PZA will start pulling the cantilever and bending it in the opposite direction. This bending results in an increment in the force registered by the PD, since the angle of the incident light on it changes. This force has its origin in the resistance made by the polyprotein while being stretched. As the PZA retracts, the protein is more stretched and force continuously increases. At some point, one of the domains of the polyprotein unfolds. Protein domain unfolding corresponds to the loss of the contacts that keep the protein in its native structure. The contacts that can be mechanically disrupted in a smFS experiment are weak interactions such as hydrogen bonding, salt bridges and hydrophobic interactions. When the unfolding of one domain takes place there is a sudden drop in the force and thus the cantilever comes back to its resting position, since the unfolding of the domain releases the length of the amino acid chain previously trapped in the native structure. After the first domain unfolds, as the PZA is

constantly retracting at a constant speed, the polyprotein is stretched again, the force rises again and then another domain unfolds and so on until the last domain is unfolded.

In a homopolyprotein, a polyprotein whose domains are all the same, the mechanical unfolding of the domains is stochastic, happening in a random order one by one but around the same force for all of them. The last peak corresponds to the detachment of the protein from the cantilever, when the polyprotein is completely unfolded and the stretching force exerted is above the strength of the interaction between the cantilever tip and the protein.

In all the FX measurements done in this thesis the polyproteins used were heteropolyproteins, hence not all the domains were the same. Specifically, each protein studied was flanked by two copies of the I91 domain in each side. These I91 domains serve as a molecular fingerprint during the experiment, since their mechanical behavior is known. All the polyproteins of this thesis were pulled in FX at the constant speed of 400 $nm \cdot s^{-1}$, and CD4D1D2 was also pulled at 10 $nm \cdot s^{-1}$. Measurements of CD4D1D2 in the presence of the antibody Ibalizumab were done at 400 $nm \cdot s^{-1}$, and both proteins were previously incubated for 1 h at RT in a 1:5 proportion before incubation over the gold substrate. The amplitude of the PZA approach-retract cycle was set to 400 nm except for FimH which was increased to 500 nm. All the measurements were done in the presence of AFM buffer.

FX operation mode allows for the study of the mechanical stability of protein domains when stretched at constant speed. The unfolding forces of a protein domain, its increment of contour length

and the persistence length of the protein polymer can be estimated with this smFS mode.

5.2.3.2 Force-clamp (FC)

FC measurements consist in pulling a single protein at a specific constant force. As in the case of the FX mode, the AFM is constantly doing force curves, with a cycle of PZA approaching to the cantilever, pressing it and then retracting. However, this time the PZA retracts in a way that the force applied to the protein is held constant using a PID feedback loop. **Fig. 5.16** shows a scheme of a FC experiment. After the PZA pushes the cantilever, it rapidly retracts until the cantilever deflection is the same as the specific force (input). Once a protein domain unfolds, there is a drop in the force registered by the PD (output), since the cantilever is coming back to its resting position.

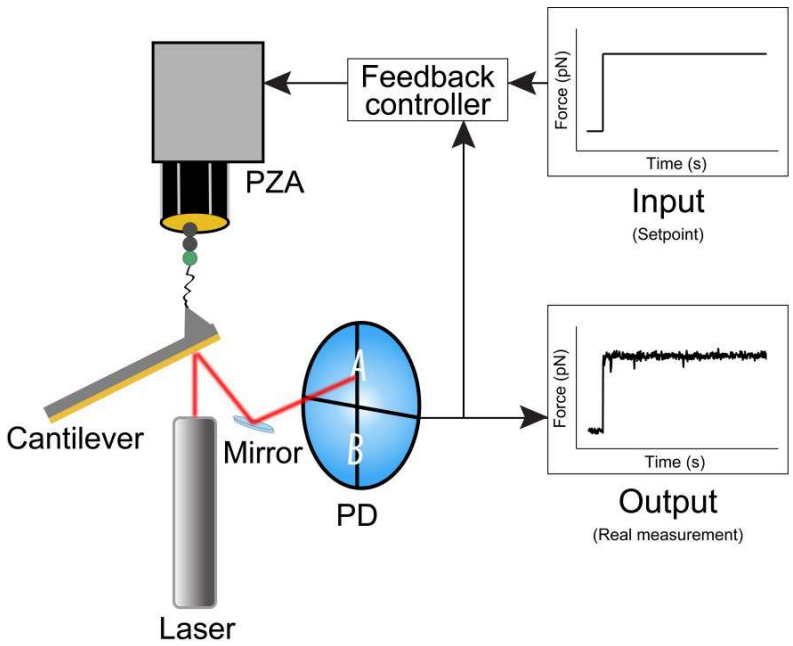


Figure 5.16. FC operation mode of the AFM. An specified force (input) is sent to the feedback loop controller. This controller sets the PZA Z-position in order to apply a specified

force on the polyprotein attached to the cantilever. The real force applied is detected by the PD, which generates the output signal of the experiment. This system is dynamic and every time a protein domain unfolds the force sensed by the protein drops, so along the experiment it is necessary to correct the force. For that reason the output signal is also sent back to the feedback controller. With the information coming from the PD and from the input signal, the feedback loop is able to correct the PZA Z-position along time in order to keep constant the force.

However, PD and the PZA are in constant communication through a feedback loop in order to keep the force constant along time. When a drop in force is detected, the PZA receives the signal to retract as much as needed in order to restore the force on the protein. Every time an unfolding event takes place and the force drops, the PZA will be retracted in order to restore cantilever's deflection (force registered by PD) and maintain the force on the studied protein. In a typical FC trace a staircase signal is registered, in which every step is an increase in length which corresponds with the unfolding of one domain. Besides the extension of the molecule, the force is also registered. In the case of the force the signal oscillates around the specified value with a 10-15 pN error and every time an unfolding event takes place there is a spike in the signal as a consequence of the drop in the force. The feedback loop restores the specified force with a response time below 5 milliseconds. An example FC trace is depicted in **Fig. 5.17** from a real measurement of $(I91)_2$-FimG-$(I91)_2$ polyprotein,

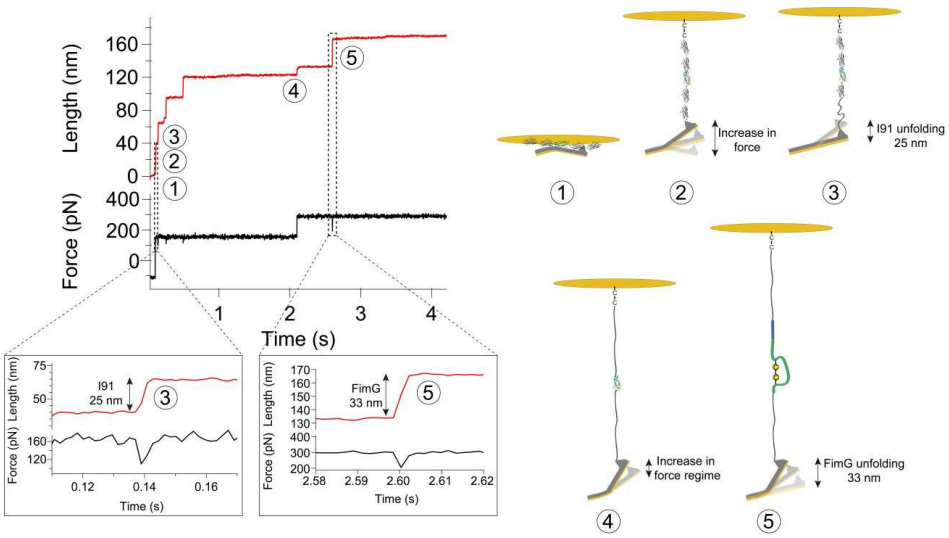


Figure 5.17. FC trace of $(I91)_2$-FimG-$(I91)_2$ polyprotein. In the left part of the figure there is a FC trace of $(I91)_2$-FimG-$(I91)_2$ polyprotein with three force regimes. First, PZA pushes cantilever (right side scheme, 1), and then it retracts until a pulling force of 160 pN are applied on the protein (2). When an I91 domain unfolds, a 25 nm step increase in length is registered, besides a drop in the force registered by the PD (lower left inset). Feedback loop controller then retracts PZA Z-position so the 160 pN are restored (3). After the unfolding of the four I91 domains, the third force regime is applied for FimG unfolding at 300 pN (4). When FimG unfolding occurs, an increase of 33 nm is registered and so it is registered a drop in the force applied. Then, PZA is retracted again for restoring the pulling force in the molecule (5).

FC traces can be set up not only for protein unfolding but also for refolding. In this thesis the unfolding of CD4D1D2 and the unfolding and refolding of self-complemented FimG were studied along single trajectories.

CD4D1D2

The CD4D1D2 nanomechanics were studied pulling at different forces ranging from 20 to 100 pN. A three-pulse force protocol was used, first pushing the cantilever for 0.5 s at 800 pN. CD4D1D2 has lower

mechanical stability than I91, so in order to isolate the unfolding of CD4D1D2, the aforementioned range of forces were applied for different times (2-30 s) in the second pulse and then the force was increased to 150 pN for I91s unfolding for 2-10 s.

In the CD4D1D2 reduction experiments with human thioredoxin, the enzyme was diluted in AFM buffer to a final concentration of 10 μM in the presence of 50 nM of human thioredoxin reductase and 2 mM of NADPH. Thioredoxin reductase and NADPH (Sigma) were added for restoring the oxidation state of the thioredoxin along the experiment. Again a three-pulse protocol was used. The second pulse used was of 50-60 pN for 2-10 s.

FimG

With this protein, the refolding behavior of oxidized (disulfide-bonded) and reduced (non disulfide-bonded) FimG was studied under different conditions. Different force protocols were applied to FimG that consisted in pushing the cantilever at 800 pN for 2 s (1), unfold the four I91s at medium force (2), isolate the unfolding of FimG at high force (3), and then quenching this force to zero for different times for allowing protein folding (4). After these quenching times, the same protocol was repeated to test if FimG refolded. Besides the different refolding quenching times used, the role of proteins involved in Fim proteins maturation was approached. First, the refolding of FimG was studied in the presence of 10 μM of chaperone FimC. Then the same was done but in the presence of 100 μM of the oxidoreductase DsbA. Finally, the effect of both proteins was studied at the same time. A different pulse protocol was used for the measurements in the presence of DsbA, and DsbA with FimC, since the reduction of FimG by

reduced DsbA could be impaired at high force[194]. **Table 5.1** summarizes the force protocols used and the different conditions employed in each experiment.

Protein under study	Unfolding Forces	Quenching forces	Unfolding forces	Other proteins in the media
$(I91)_2$-CD4D1D2-$(I91)_2$	20/160 pN (20-30/2 s)	No	No	-/+ thioredoxin
$(I91)_2$-FimG-$(I91)_2$	160/300 pN (2/20 s)	0 pN (10, 30, 45, 60 s)	160/300 pN (2/20 s)	-/+ FimC (10 μM)
$(I91)_2$-FimG-$(I91)_2$	160/300/100 pN (2/5/2 s) or 160/300 pN (2/7 s)	0 pN (45 s)	160/300/100 pN (2/5/2 s) or 160/300 pN (2/7 s)	-/+ DsbA (100 μM) or +FimC/+DsbA (10/100 μM)

Table 5.1. Force protocols used in constant force smFS experiments.

Fim proteins are exported to the periplasmic space in an unfolded, extended, and non-disulfided state. With constant force smFS experiments unfolded proteins are obtained which resemble the same conditions as the ones taking place during pilus proteins translocation and maturation in the periplasm.

FC operation mode allows for the study of the mechanical stability of a protein at a specific force.

5.2.3.3 Force-ramp (FR)

The last operational mode used was FR. In this mode, the force applied to the molecule is changing linear the time. PZA constantly retracts increasing the deflection of the cantilever and thus the stretching force on the polyprotein. The loading rate is constant and the force triggers the unfolding of the protein domains. Again the feedback loop is in

charge of restoring the tension on the molecule after each unfolding event.

FR experiments were conducted on CD4D1D2 protein in the absence and in the presence of Ibalizumab, incubated at rt for 1h in a 1:5 ratio. The loading rate was 33 $pN \cdot s^{-1}$.

5.3 smFS data analysis

All the experimental data was analyzed with the Igor Pro software (Wavemetrics). The analysis procedure employed for the analysis was kindly provided by Prof. Julio M. Fernández laboratory (AFM_Analysis_V2.40.ipf). Besides, a modification of this analysis procedure for force correction was kindly provided by Mariano Carrión laboratory (LN_AFM_Analysis_V1.1_Filtering).

FX data

The FX data collected had the already mentioned sawtooth pattern in which the unfolding peaks correspond with one domain unfolding. Each of the peaks of the trace is fitted to the Worm-Like Chain (WLC) model for polymer elasticity,

$$F(x) = \frac{k_B T}{p}\left[\frac{1}{4}\left(1 - \frac{x}{L_c}\right)^{-2} - \frac{1}{4} + \frac{x}{L_c}\right] \quad (5)$$

Where k_B is the Boltzmann constant, T is the temperature, p is the persistence length, L_c is the contour length and x is the extension. The WLC model can be applied to describe the elasticity of a polymer submitted to a mechanical stress[195–197]. This model describes a polymer

as a string of monomers set one after the other, of a given total length. Polymers in solution adopt a coil conformation because this allow to all its monomers to maximize their entropy. When the polymer is extended the entropy of the monomers is reduced and this generates an opposite force. The persistence length p defines the flexibility of the chain. In a protein the value of this parameter is the length of one aminoacid (one monomer) being around 0.4 nm. The contour length (L_c) is the end-to-end distance of the polymer. In a FX trace where the WLC model is applied to fit the unfolding of each protein domain, we can modify both L_c and p. The value of p was maintained fixed between 0.25 and 0.5 nm, around the length value of an amino acid. The L_c was calculated as the distance between two consecutive unfolding peaks. It is denoted as ΔL_c since the protein domains when folded already present a certain end-to-end distance, being around 5 nm for Ig-like domains. **Fig. 5.18** shows the fitting of the unfolding peaks from the $(191)_2$-FimF-$(I91)_2$ trace depicted in **Fig 5.15**.

The FX experiments yield a great amount of data; however it is important to select it correctly. Among this body of data the vast majority is constituted by incomplete stretched proteins. Working with homopolyproteins constituted by the repetition of the same protein module is not so problematic since an incomplete trace will still have several peaks which are the same as the missing ones. In this thesis all the proteins stretched were heteropolyproteins, in which the protein domain we wanted to study was flanked on both sides by two copies of the I91 protein domain. As it was mentioned before, I91 mechanical properties are well known thus it serves as a fingerprint for identifying the protein under study. I91 domains have a typical ΔL_c of ~28 nm and an unfolding force of ~200 pN. For that reason, in order to select only

the most suitable data only those traces which present the unfolding peaks of at least three I91 domains are accepted, and they should fulfill with the typical ΔL_c and unfolding forces. To identify the protein domains under study we first calculate theoretically their ΔL_c multiplying the number of aminoacids of the domain by the p of a polypeptide (~0.4 nm·aminoacid^{-1}). In order to calculate the average ΔL_c and the average unfolding force of a protein domain it is necessary to collect as many traces as possible to get a solid statistical support. Usually this number is above 50 traces but for FimH lectin domain this amount was below because of the difficulty to measure this protein. The dispersion of the data was calculated as the standard deviation.

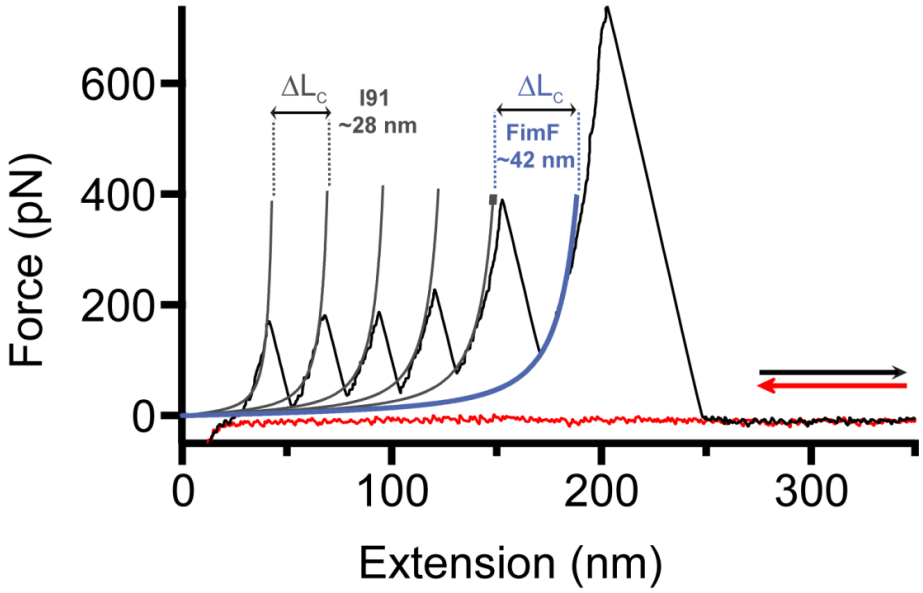


Figure 5.18. FX trace of (I91)$_2$-FimF-(I91)$_2$ polyprotein with WLC model fitting. Same trace as shown in **Fig. 2.15** with the WLC fitting of the unfolding peaks. The The fitting of the I91 domains is showed in grey meanwhile FimF unfolding is shown in blue. ΔL_c is calculated as the distance between the fittings of two consecutive peaks. The red arrows indicates the approaching of the PZA to the cantilever (red trace) and the black arrow shows the retraction of the PZA (black trace).

FC and FR data

The FC and FR data had a staircase pattern in which every step corresponds with the unfolding of one protein domain. The increase in length is measured as the difference in height before and after the unfolding of the domain. In this case the ΔL_c of the I91 is ~25 nm because at constant force the protein domains are not stretched to its maximum as it happens at constant speed. Again, in order to accept a trace at least three I91 domains should be present during the unfolding. In the refolding experiments it is expected to see at least the unfolding of two I91s after the quenching time. For the unfolding rates of CD4D1D2 and FimG calculated at a specified force, the unfolding trajectories were summed and averaged, and then fitted to single or double exponential functions. The refolding probability of the experiments done at constant force with FimG was calculated as the ratio of trajectories showing the refolding of FimG divided by the total amount of trajectories collected under a specific condition and quenching time. The standard deviation of this refolding probability was calculated as the one followed by a binomial distribution.

5.4 Steered Molecular Dynamics simulations

Besides the smFS experiments of the type I pilus proteins, these ones were also stretched *in silico*. The goal was to get a detailed view of the unfolding process at the atomic scale, since these details are undetectable in an AFM. For that purpose all-atom Steered Molecular Dynamics (SMD) simulations were done using the GROMACS 4.6.5 package[198]. Self-complemented structures of the four Fim proteins

equivalent to the ones used in smFS experiments were used for the pulling experiments, based on previous 3D resolved structures and available in the Protein Data Bank (2JTY for the self-complemented FimA and 4J3O for the complex FimC-FimF-FimG-FimH). The subunits were isolated individually with their correspondent donor β-strand and the DNKQ linker was manually added to each of them. In the case of FimF both the linker and the donor β-strand of FimA were manually inserted in the structure since there are no available structures in the PDB database. In the case of FimH we only used the pilin domain.

Also, it was tested how could affect the self-complementation strategy followed in the design of the Fim proteins for smFS. For that reason another structure of FimH pilin domain complemented with the donor β-strand of FimG but not covalently linked together, was stretched in SMD simulations. This approach allows to test if our experimental constructs are mechanically different from the real interaction that takes place in the pilus. This experiment is only possible *in silico* since non complemented Fim proteins are very prone to aggregation and unable to fold if they lack the complementation of their cognate donor β-strand or from the donor β-strand of FimC.

All the structures were placed in boxes of 7x7x28 nm dimensions with their pulling axis aligned in the z-dimension. Explicit solvent of TIP3P water molecules[199] and 0.15 M of NaCl were added to the boxes. The force-field used was CHARMM27-CMAP[200] and the systems were prepared with a steepest descent minimization and an equilibration simulation of 100 ps with velocity rescaling[201]. Upon relaxation and minimization steps of the structures the proteins were

stretched along the z-axis taking as reference pulling groups the nitrogen atom of the N-terminal aminoacid from the pilin domain and the carbonyl atom from the last C-terminal aminoacid of the donor β-strand. The pulling was done at 1 $nm \cdot ns^{-1}$ for 10 ns with Nosé-Hoover temperature coupling and the timestep was 0.001 ps. At this pulling rate and with this simulation time it is possible to capture the complete ripping out of the donor β-strand from the pilin fold of each structure. The simulations were conducted with periodic boundary conditions in all directions and with the Smooth Particle Mesh Ewald (SPME) method for long-range electrostatics[202] and a real-space cut-off for van der Waals interactions of 1.5 nm. No pressure coupling was used and the temperature of the system was set to 300 K. The hydrogen-bonding pattern of each donor β-strand with the pilin fold was calculated based on the instantaneous distances among oxygen and hydrogen atoms analyzed along the equilibration run and the pulling simulation.

Chapter 6: Results

In this chapter all the results obtained along this thesis are detailed. These results are organized in two sections that approach some of the nanomechanical aspects involved in bacterial and viral attachment to target cells.

The first part of the results is dedicated to the bacterial attachment system of the type I pilus. This set of results is at the same time divided in three sections. One is dedicated to the mechanical study with smFS of the Fim proteins that compose the type I pilus; the second one gives an atomic-level insight to the smFS experiments with Molecular Dynamics simulations; and the third one is focused on the study of the folding and the maturation of the FimG subunit for pilus genesis with smFS.

The second part of the results is dedicated to the nanomechanical study of the first two domains of the T-cell surface receptor CD4. Besides, the effect of other proteins in the mechanical behavior of CD4D1D2 was studied and a theoretical model was elaborated for relating the extension of the CD4 protein with the infectivity of the HIV-1 virus.

6.1 Type I-pilus nanomechanical architecture and genesis

6.1.1 Nanomechanical architecture of the type I pilus

In order to test the mechanical resistance and stability of the type I pilus we designed self-complemented pilus subunits in which each protein domain bears the sequence of the donor β-strand of the next subunit of the pilus. The whole pilus structure and stability relies on this donor strand complementation interaction between subunits, thus pulling along the N-terminal to the C-terminal axis of the subunits we resemble the pulling direction experienced by the pilus when attached to a surface.

We pulled the different proteins at 400 $nm \cdot s^{-1}$ in FX mode. Most of the type I pilus is composed of FimA subunits which form the pilus rod, being the FimA-FimA interaction the most abundant in the structure. Therefore we pulled from the chimeric heteropolyprotein $(I91)_2$-FimA-$(I91)_2$ in which the FimA subunit is complemented with its donor β-strand but placed in the C-terminal end through the linker. Besides, FimA domain has a conserved disulfide bond between its A and B β-strands as the other Fim proteins. So the mechanical stability of the oxidized (disulfide-bonded) and reduced (non disulfide-bonded) domains was tested. In the **Fig. 6.1** are shown two FX traces of FimA, each one in one oxidation state.

In **Fig. 6.1a** after the four I91 unfolding peaks and before the detachment it can be seen a peak whose ΔL_c and unfolding force is 42 nm and 500 pN respectively, corresponding with the oxidized domain ($FimA_{S\text{-}S}$). The theoretical contour length (L_c) of FimA is ~62 nm,

assuming each aminoacid contributes with 0.38 nm to the L_c[158] and knowing that FimA sequence has 163 aminoacids. However the force regime we can reach allows us to stretch the protein up to its disulfide bond, and there are 41 aminoacids enclosed between the two cysteines involved in the disulfide bond. This leaves the maximum L_c to ~46 nm (122 aminoacids). However, it is necessary to take into account that a folded protein domain already has an initial end-to-end distance before the stretching. In the self-complemented FimA with no N-terminal donor β-strand this length is ~4 nm between the α-carbons of the first and the last amino acids. Therefore the expected ΔL_c is 42 nm for $FimA_{S\text{-}S}$, matching the value shown in the trace of the **Fig. 6.1a**.

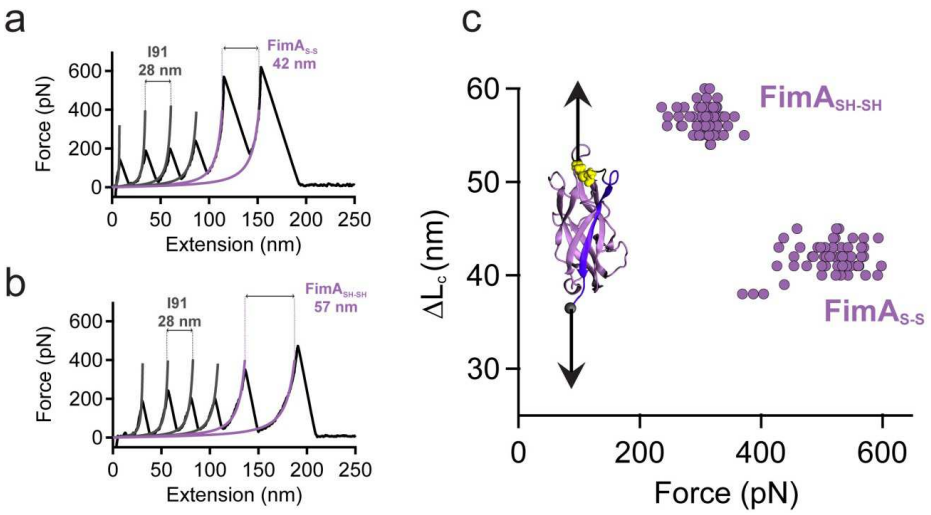


Figure 6.1. FX of $(I91)_2$-FimA-$(I91)_2$. (**a**) FX trace of oxidized FimA ($FimA_{S\text{-}S}$) (**b**) FX trace of reduced FimA ($FimA_{SH\text{-}SH}$). Grey lines are the WLC fitting of the I91 unfolding peaks. The purple one represents the fitting to FimA unfolding. (**c**) Increment of contour length (ΔL_c) *vs* unfolding force. Oxidized (n=63) and reduced FimA (n=68) populations differ clearly both in ΔL_c and force. The cartoon figure has two arrows showing the pulling axis of the self-complemented FimA domain, where the yellow spheres represent the cysteine residues and the blue strand represents the donor β-strand of FimF. Linker appears in black.

On the other hand, reduced FimA does not have the intramolecular bond so the whole sequence can be stretched, leaving us a ΔL_c of ~58 nm. **Fig. 6.1b** shows a FX trace of reduced FimG ($FimA_{SH-SH}$). The ΔL_c is 57 nm as expected but the unfolding force is decreased (300 pN) when compared with the oxidized domain. The absence of the disulfide bond makes the protein weaker when this is mechanically unfolded. **Fig. 6.1c** shows ΔL_c *vs* unfolding force plot of all the unfolding peaks collected for FimA both in the absence and in the presence of the disulfide bond. The calculated average for ΔL_c was 42 ± 2 nm (mean±SD) and for the unfolding force was 516 ± 49 pN (mean±SD). Meanwhile, for $FimA_{SH-SH}$ the ΔL_c was 57 ± 1 nm and for the unfolding force was 307 ± 25 pN. This represents a 33 % decrease in the mechanical stability of the reduced domain.

Following the type I pilus architecture, after the pilus rod composed of hundreds of FimA subunits we find the pilus tip fibrillum. This structure is composed of the array of FimF-FimG-FimH although two subunits of FimF could be present instead of only one. FimF and FimG constitute what is known as the hinge region of the pilus, bridging the pilus rod with the lectin binding protein. Therefore we pulled the oxidized and reduced domains of both proteins. In the case of $FimF_{S-S}$, the ΔL_c was 42 ± 2 nm and the unfolding force was 409 ± 39 pN. The theoretical ΔL_c is 42.5 nm since in the oxidized domain there are 125 aminoacids available for stretching and the L_c is ~5 nm. In the reduced domain, the ΔL_c was 55 ± 1 nm and the unfolding force was 272 ± 25 pN (**Fig. 6.2**).

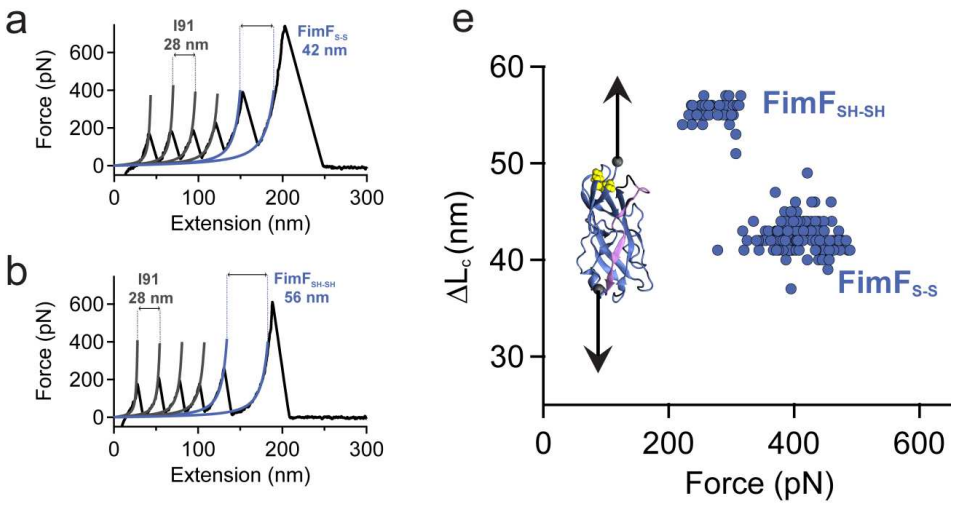


Figure 6.2. FX of $(I91)_2$-FimF-$(I91)_2$. (**a**) FX trace of oxidized FimF ($FimF_{S-S}$) (**b**) FX trace of reduced FimF ($FimF_{SH-SH}$). Grey lines are the WLC fitting of the I91 unfolding peaks. The blue line represents the fitting to FimF unfolding. (**e**) Increment of contour length (ΔL_c) *vs* unfolding force. Oxidized (n=150) and reduced FimF (n=40). (**c**) The cartoon figures show the pulling axis of the self-complemented FimF domain (in **c**) where the yellow spheres represent the cysteine residues and the purple strand represents the donor β-strand of FimA. Linker is shown as a black line.

In the case of FimG, the oxidized domain displayed an ΔL_c of 40 ± 2 nm and an unfolding force of 440 ± 36 pN, meanwhile the reduced one had an ΔL_c of 52 ± 1 nm and an unfolding force of 339 ± 27 pN (**Fig. 6.3**). Again, we observe a pronounced decrease in the mechanical stability of the non-disulfided proteins, ~50 % in the case of FimF and ~30 % in the case of FimG.

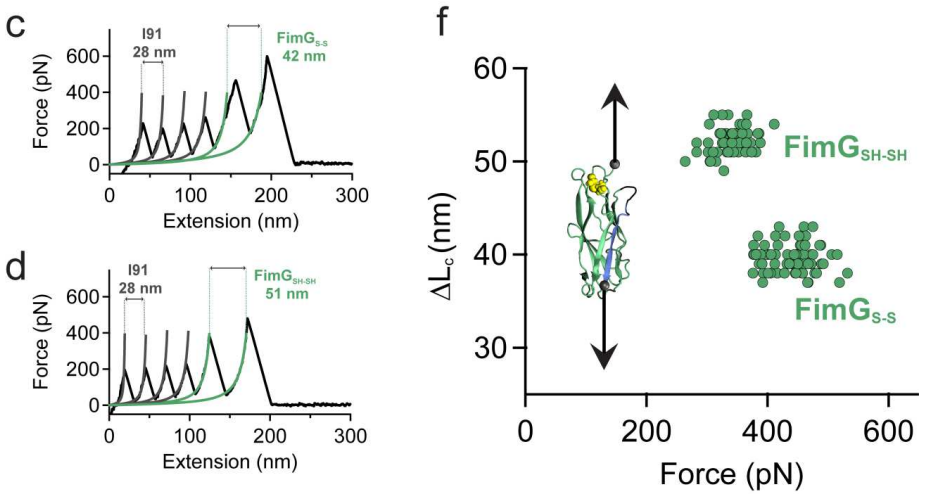


Figure 6.3. FX of $(I91)_2$-FimG-$(I91)_2$. (**a**) FX trace of oxidized FimG ($FimG_{S-S}$) (**b**) FX trace of reduced FimG ($FimG_{SH-SH}$). Grey lines are the WLC fitting of the I91 unfolding peaks. The green line represents the fitting to FimG. (**c**) Increment of contour length (ΔL_c) *vs* unfolding force. Oxidized (n=83) and reduced FimG (n=69). The cartoon figures show the pulling axis of the self-complemented FimG (in **c**) where the yellow spheres represent the cysteine residues and the blue strand represents the donor β-strand of FimF. Linker is shown as a black line.

FimH is the last of the pilus subunits. It is divided in two connected domains, the N-terminal lectin domain involved in the recognition and binding to the sugars present in the cells of the bladder, and the C-terminal pilin domain. The pilin domain is complemented by the donor β-strand of FimG, thus it shares the same characteristics as the other Fim proteins. Both pilin and lectin domains have conserved disulfide bonds near their N-terminal sequence. In **Fig. 6.4** are shown FX traces of the polyprotein $(I91)_2$-FimH-$(I91)_2$. In the oxidized state the unfolding of FimH is detected as two independent peaks, one corresponding to the lectin domain ($FimH_L$) and the other corresponding to the pilin domain ($FimH_P$). $FimH_L$ is the weakest domain in the polyprotein thus it unfolds at the beginning of the stretching. After it, the I91 domains unfold and then $FimH_P$ unfolds just

before the detachment peak. The average ΔL_c is 40 ± 2 nm and the unfolding force is 122 ± 24pN, for $FimH_L$. In one third of the traces the unfolding of $FimH_L$ domain takes place through an intermediate with two peaks, one with an ΔL_c of 6 ± 1 nm and 95 ± 30 pN and the other with ΔL_c of 35 ± 2 nm and 116 ± 19 pN (**Fig. 6.4b**). When reduced, the pilin domain presents a ΔL_c of 47 ± 2 nm and a mechanical stability of 235 ± 37 pN, following the trend observed in the other Fim proteins. Capturing the unfolding of the reduced lectin domain was elusive only being possible to register it in two traces, probably due to the misfolding or instability of the domain in the reduced state.

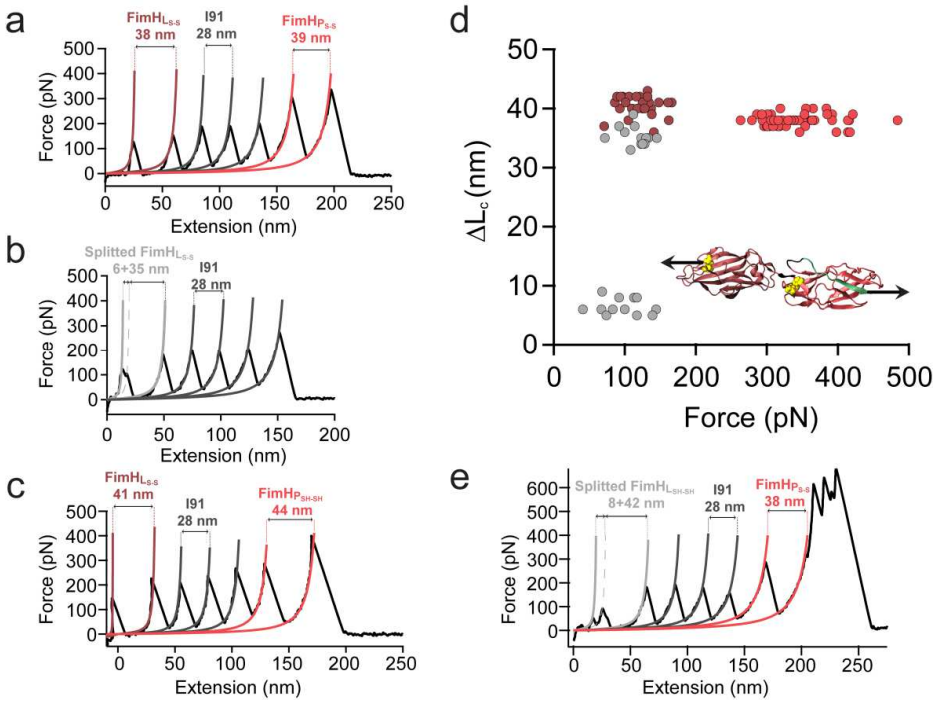
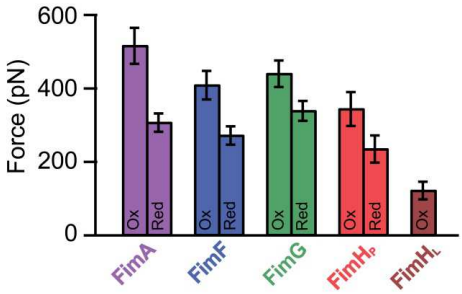


Figure 6.4. FX of $(I91)_2$-FimH-$(I91)_2$. (**a**) FX trace of oxidized FimH, with both lectin ($FimH_{LS-S}$) and pilin domains oxidized ($FimH_{PS-S}$) (**b**) FX trace of oxidized lectin domain showing an intermediate ($FimH_{LS-S}$). (**c**) FX trace of reduced pilin domain ($FimH_{LSH-SH}$) (**d**) Increment of contour length (ΔL_c) *vs* unfolding force. Oxidized $FimH_{PS-S}$ (n=48), oxidized $FimH_{LS-S}$ (n=29) and $FimH_{LS-S}$ with an intermediate (n=13). The cartoon figure shows the pulling axis of the self-complemented FimH. $FimH_L$ (brown) and $FimH_P$ (red) present both a

disulfide bond, showed as yellow spheres. The green strand represents the donor β-strand of FimG. Linker is shown as a black line. (**e**) FX trace of reduced $FimH_L$ ($FimH_{LSH-SH}$). Grey lines are the WLC fitting of the I91 unfolding peaks. The red one represents the fitting to $FimH_P$ unfolding, the brown one represents the fitting to $FimH_L$.

In **Fig. 6.5** is shown the unfolding forces of the Fim subunits both in the oxidized and reduced state. It is clear there is an unequivocal decreasing trend in the mechanical stability of the domains as we approach to the tip of the type I pilus. The FimA subunit is the most stable domain followed by FimG and FimF and then by $FimH_P$ and $FimH_L$. This trend is reproduced at lower forces by the reduced domains of each Fim protein. It seems clear the type I pilus architecture and its mechanical properties are intimately related, yielding a very resistant structure which relies on non-covalent interactions and submitted to challenging shear forces.

a



b

Protein domain	Force (pN)	ΔL_c (nm)
$FimA_{S-S}$	516±49	42±2
$FimA_{SH-SH}$	307±25	57±1
$FimF_{S-S}$	409±39	42±2
$FimF_{SH-SH}$	272±25	55±1
$FimG_{S-S}$	440±36	40±2
$FimG_{SH-SH}$	339±27	52±1
$FimHP_{S-S}$	342±46	38±1
$FimHP_{SH-SH}$	235±37	47±2
$FimH_{LS-S}$	122±24	40±2

Figure 6.5. Hierarchical unfolding forces of type I pilus subunits. (**a**) Histogram showing the values of the unfolding forces of both oxidized and reduced pilus domains. Bars show the mean value and the error bars show the SD (**b**) Summary table of the unfolding forces and the increment of contour length values of the subunits in their oxidized and reduced state. Data showed as mean±SD.

6.1.2 Molecular Dynamics pulling simulations of type I pilus subunits

smFS experiments show a general view of the unfolding mechanism of the pilus subunits. However a detailed description of the events taking place during the unfolding is unreachable with the AFM. For that reason, Molecular Dynamics simulations are a useful tool for exploring atomic-level events that fall outside the reachable experimental range. This computational study was done in collaboration with Dr Fabiano Corsetti and Dr Emilio Artacho.

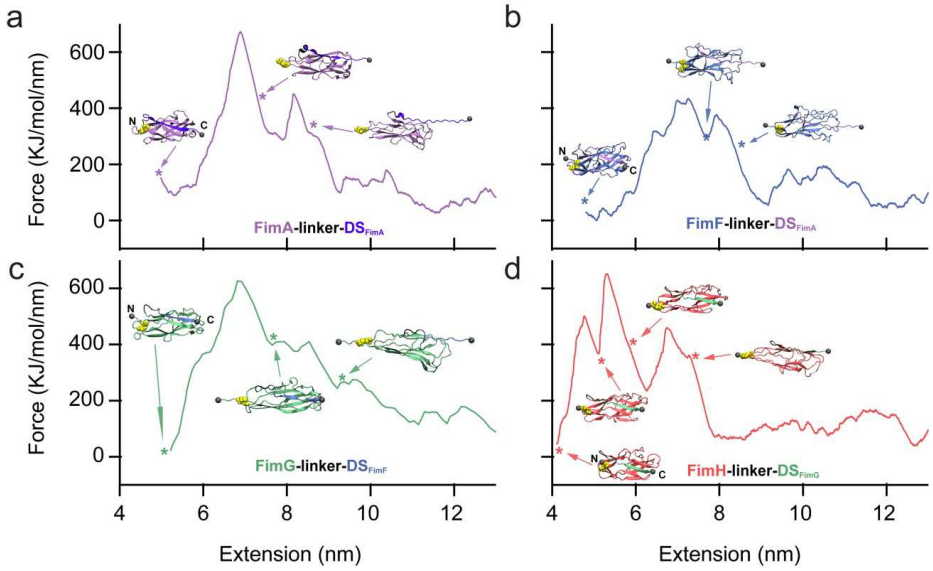


Figure 6.6. Force vs Extension plots of the SMD simulations of the type I pilus subunits. From top to bottom, FimA-DS$_{FimA}$ (**a**), FimF-DS$_{FimA}$ (**b**), FimG-DS$_{FimF}$ (**c**) and FimH$_P$-DS$_{FimG}$ (**d**). The cartoon insets show the proteins at different extensions during the pulling. N and C terminal aminoacids are highlighted as grey beads, except for FimA whose N-terminal aminoacid is a cysteine, shown with a van der Waals representation. The four aminoacid linkers (DNKQ) are represented in black.

Steered Molecular Dynamics (SMD) simulations were performed in the four self-complemented Fim proteins. Proteins were

pulled from their N-terminal end at a constant speed (1 $nm \cdot ns^{-1}$), mimicking the FX experiments done in the AFM. The proteins were stretched during 10 ns, a time window long enough for β-strand removal from the protein fold. **Fig. 6.6** shows the force *vs* extension plots of the SMD simulations of the four pilin proteins. At first sight the force traces look very similar and the force peaks observed coincide at similar protein extensions. Most of the force builds up from the strength of the interaction between the donor β-strand and the rest of the fold.

Several hydrogen-bonds connect the donor β-strand with two β-strands, one running parallel and the other one in an antiparallel disposition. **Fig. 6.7** shows the hydrogen bonds between the donor β-strands and the fold of the protein they are complementing. The hierarchical forces observed in the AFM are somehow conserved in the simulations although FimH climbed two positions (FimA: 673 $KJ \cdot nm \cdot mol^{-1}$ > FimH: 652 $KJ \cdot nm \cdot mol^{-1}$ > FimG: 626 $KJ \cdot nm \cdot mol^{-1}$ > FimF: 434 $KJ \cdot nm \cdot mol^{-1}$). The pulling rate used in the SMD simulations is six orders of magnitude faster than the attained experimentally in the AFM, a possible explanation for this rearrangement in the forces.

Nevertheless the differences in the force maximums are negligible at the force ranges registered. What is common to the four proteins is that the force peaks disappear once the correspondent donor β-strand is ripped out from the protein fold and the force drops to near zero once this event takes place, pointing to the fact that all the mechanical resistance of the pilus domains lies on the strength of the interactions between the donor β-strand and the protein fold.

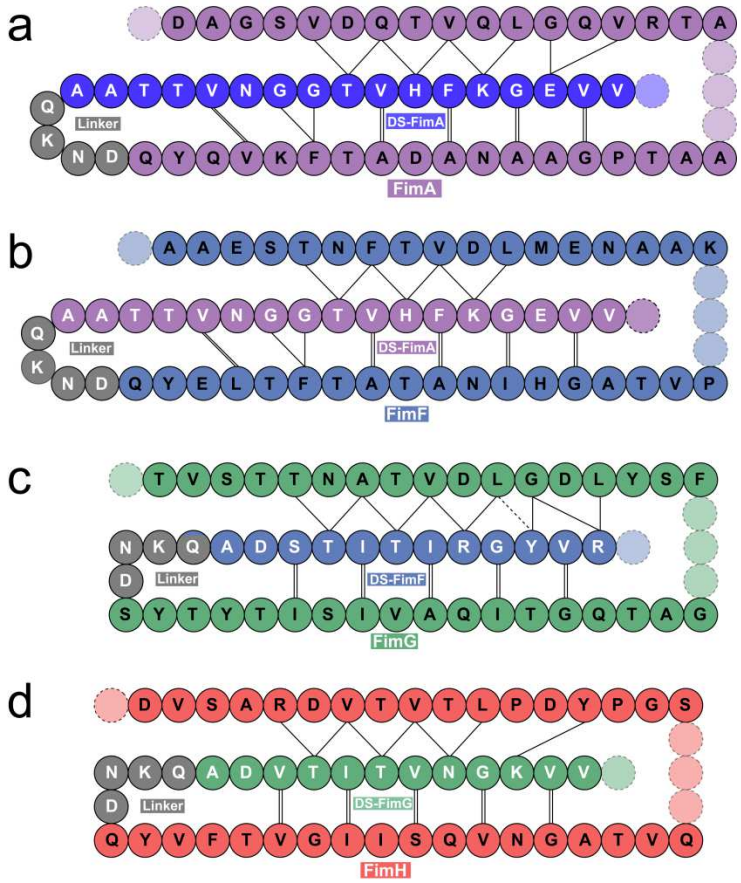


Figure 6.7. Hydrogen bond pattern of the self-complemented Fim proteins. The bonding pattern is recovered by examining the fluctuation of O-H bond lengths in the equilibrated structure before the start of the pulling. Each circle represents a residue (specified in letter code), and a single line connecting two circles shows a hydrogen bond between them. Double lines indicate two hydrogen bonds within the same pair of residues. The dotted line indicates a weak bond. Within each protein, the top strand and donor strand exhibit parallel β-sheet bonding (given by the zig-zag bonding pattern), while the bottom strand and donor strand exhibit antiparallel β-sheet bonding (given by the row of double bonds).

In the case of FimA, 20 hydrogen bonds are stitching the fold; FimG and FimF show 18 and 19 bonds respectively, and FimH shows 17. Based on this finding, we can infer that FimA shows the highest unfolding force in smFS experiments since it presents the highest number of hydrogen bonds between the donor β-strand and its

surroundings. FimG and FimF show a similar number and FimH shows the lowest number, something we can relate with the different unfolding forces registered in the AFM in which FimH pilin domain shows the lowest unfolding force among the all pilins.

In order to validate the experimental design of the self-complemented domains, a SMD simulation with the same conditions was done on the FimH pilin domain complemented with the donor β-strand of FimG but not covalently linked to it. **Fig. 6.8** shows the comparison between the pulling simulations of the complemented (688 KJ·nm·mol^{-1}) and the self-complemented (652 KJ·nm·mol^{-1}) FimH pilin domains. The force-extension plot shows almost no difference between pullings, validating our experimental constructs and the data obtained in the smFS experiments. In the complemented protein the force drops to zero after the removal of the donor strand of FimG since no physical connection links strand with the rest of the fold, as it happens in the self-complemented structure.

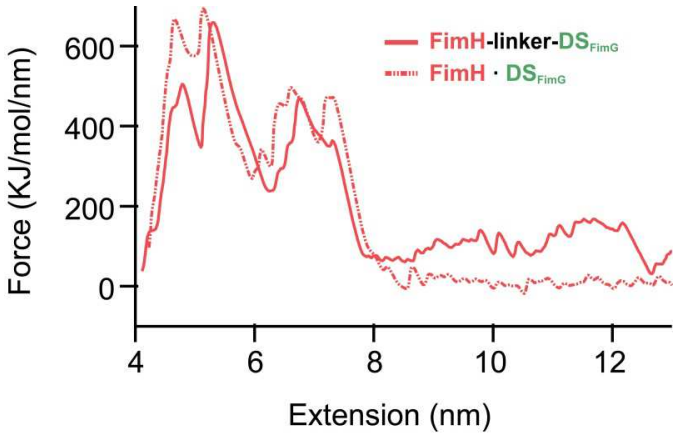


Figure 6.8. Force *vs* time of the Steered Molecular Dynamics simulations of the self-complemented FimH pilin domain and the FimH pilin domain with the donor β-strand of FimG. The self-complemented FimH shows the same unfolding pattern and the same mechanical stability as the FimH domain complemented with the donor β-strand of FimG.

6.1.3 FimG subunit maturation and folding for pilus incorporation

smFS experiments allowed us to study the mechanical stability of the Fim proteins. Besides, it allows us to study at the single-molecule level the process of maturation of the pilus subunits. Fim proteins are exported to the periplasmic space in an extended and non-disulfided conformation. Under these conditions they cross the inner membrane through the SecYEC pore and once in the periplasm an orchestrate process takes place for the formation of their disulfide bond, their correct folding and their transport to the outer membrane FimD/Usher protein platform for incorporation to the pilus.

With the AFM we can unfold and extend the protein mimicking the same conditions as the ones the pilus proteins face when they are exported to the periplasmic space. In the *in vivo* process, it is known that the first actor involved after the subunit exportation to the periplasm is DsbA, the oxidoreductase enzyme involved in the formation of the conserved disulfide bond present in the Fim proteins. Upon oxidation, the chaperone FimC donates a β-strand to the incomplete folds of the Fim proteins. The established model stipules that once the FimC-Fim complex is formed, the Fim protein folds with the help of FimC. The complex (i.e. FimC-FimF) is then recruited in the outer membrane by the FimD platform where there is an exchange between the N-terminal donor strand of the incoming subunit (i.e. FimF) and the one of FimC of a previously incorporated Fim protein (i.e FimG). When the donor strand complementation (DSC) between subunits is done, the FimC (i.e. the FimC from the complex FimG-FimC) chaperone is released and ready for the interaction with a new incoming Fim protein. When a new Fim complexed with the chaperone

arrives to the FimD (i.e. FimA-FimC), DSC takes place and the FimC from the previous subunit (FimF-FimC) is released, and so on.

We tested this process with the self-complemented FimG domain using the force-clamp (FC) mode of the AFM. In the case of FimG, the DSC between FimG and its cognate donor β-strand from FimF is already formed in the protein structure since it is covalently linked to it. However, the unfolding of the protein yields an extended polypeptide, separating the FimF donor strand from the rest of the protein fold. When the protein is held in an extended state for a period of time it enables the intervention of third parties such as FimC or DsbA.

FimG assisted refolding in the presence of FimC

Under these considerations, a constant force protocol was designed in order to test the ability of FimG-DS_{FimF} to refold after its mechanical unfolding. First, a pulse of 160 pN for 2 s was applied for the I91 unfoldings, and then the force was increased to 300 pN for 20 s in order to unfold FimG. Then, the force was quenched to 0 pN for different times ranging from 10 to 60 s for allowing protein refolding. After, the same two pulses of the beginning were applied to test if protein refolding took place. We tested this ability to refold both in the absence and in the presence of the chaperone FimC. Besides, we wanted to test the ability of FimC to discriminate between oxidized and reduced FimG domains, since it has been suggested the need of the disulfide bond for FimC binding and Fim subunits maturation previous to pilus incorporation[203].

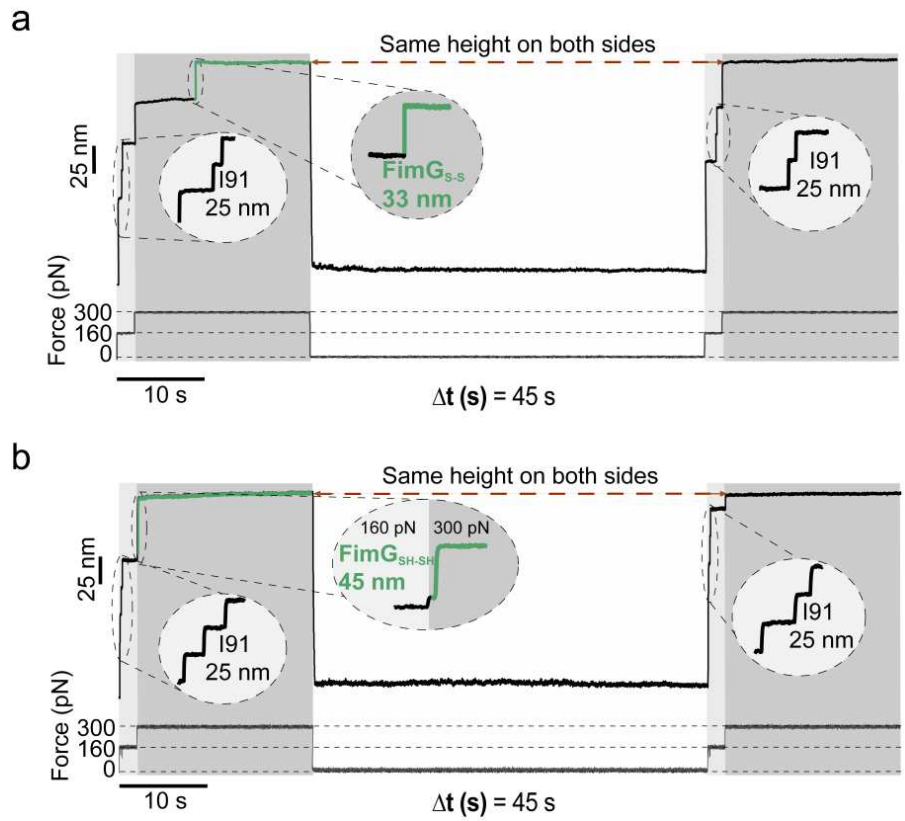


Figure 6.9. Force-clamp refolding experiments of FimG. The figure shows two example traces from unfolding and refolding experiments of FimG at 45 s of quenching time. In each trace I91 unfoldings are detected as 25 nm increase in length, meanwhile FimG is detected as an increase of 33 nm (when disulfide-bonded, (**a**)) or 45 nm (reduced, (**b**)). Dotted brown line indicates the stretched molecule has the same height on both sides. The different gray scale of the shaded areas indicates the constant force applied along the experiment (grey, 160 pN for 2 s; dark grey, 300 pN for 20 s; and light grey, 0 pN for different times).

In a FC experiment both the force applied and the PZA displacement are registered. The unfolding of one protein domain is captured as a step increase in length and a sudden drop in the force applied. The feedback loop then corrects and restores the constant force value retracting the PZA as much as needed for restoring the force value. **Fig. 6.9** shows two traces of FimG-DS$_{FimF}$, one oxidized (step size ~33 nm) and the other reduced (step size ~45 nm), in which the

refolding of FimG did not take place after the quenching time at 0 pN. We know FimG did not refold since there is no any step length increase in the second part of the experiment, but the total extension of the protein is the same before and after the quenching time. We do not know if the FimC recognition of the FimG domain occurs when the protein is extended or once it is collapsed during the quenching time. For that reason we tested a wide range of quenching times (10, 30, 45 and 60 s). **Fig. 6.10** and **Fig. 6.11** shows oxidized and reduced refolded traces at 10, 30, 45 and 60 s of quenching time.

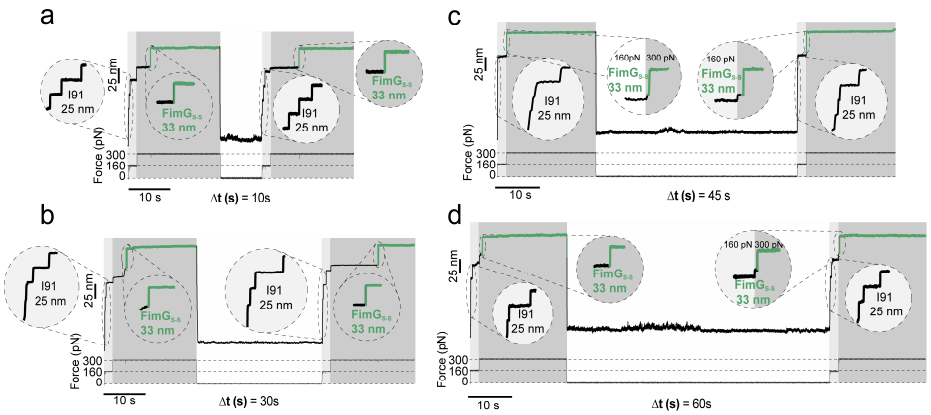


Figure 6.10. Force-clamp refolding experiments of oxidized FimG in the presence of FimC. The figure shows four example traces from unfolding and refolding experiments of FimG in the presence of FimC at different quenching times. In each trace I91 unfoldings are detected as 25 nm increase in length, meanwhile is FimG is detected as an increase of 33 nm. Oxidized FimG ($FimG_{S\text{-}S}$) at 10 s (**a**), 30 s (**b**), 45 s (**c**) and 60s (**d**) of quenching time. The different gray scale of the shaded areas indicates the constant force applied along the experiment (grey, 160 pN for 2 s; dark grey, 300 pN for 20 s; and light grey, 0 pN for different times).

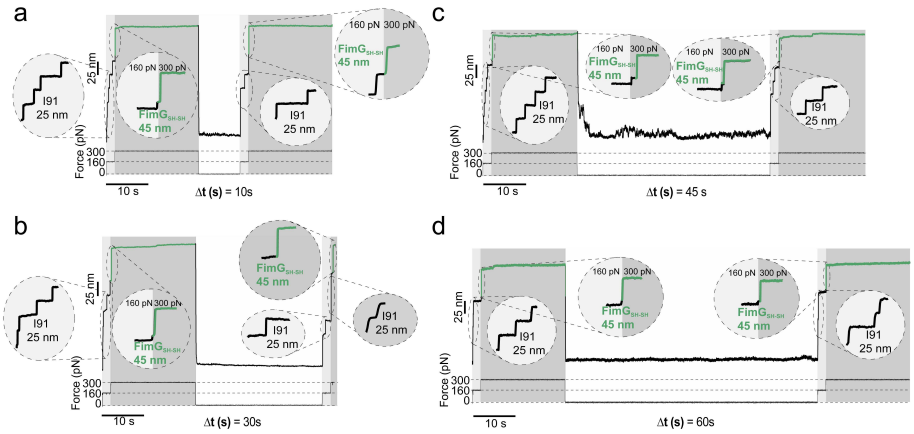


Figure 6.11. Force-clamp refolding experiments of reduced FimG in the presence of FimC. The figure shows four example traces from unfolding and refolding experiments of FimG in the presence of FimC at different quenching times. In each trace I91 unfoldings are detected as 25 nm increase in length, meanwhile is FimG is detected as an increase of 45 nm. Reduced FimG ($FimG_{SH-SH}$) at 10 s (**a**), 30 s (**b**) 45 s (**c**) and 60 s (**d**) of quenching time. The different gray scale of the shaded areas indicates the constant force applied along the experiment (grey, 160 pN for 2 s; dark grey, 300 pN for 20 s; and light grey, 0 pN for different times).

After collecting a significant number of traces under each different condition, we calculated the refolding probability of oxidized and reduced FimG in the absence and in the presence of FimC for four different quenching times. **Fig. 6.12** shows the refolding probability of FimG under different conditions and the single-exponential fitting of the data.

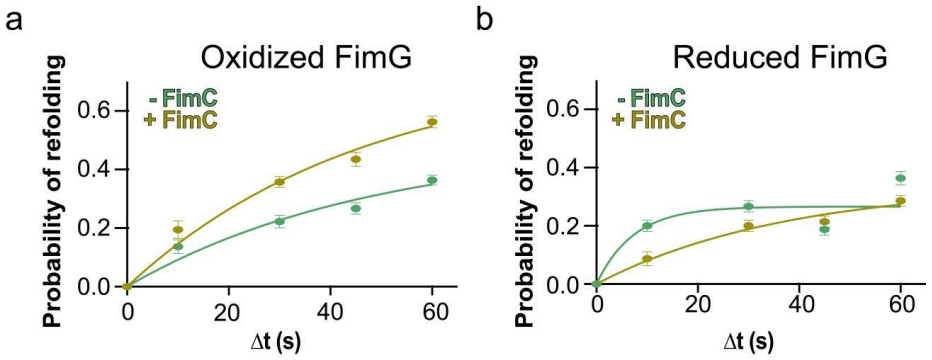


Figure 6.12. Single-exponential fitting of the Probability of refolding of FimG in the presence and in the absence of FimC. (**a**) Fitting to a single-exponential function of the Probability of refolding of oxidized FimG in the absence (green, for Δt=10s, n=22; Δt=30s, n=18; Δ=45s, n=15 and Δt=60s, n=11) and in the presence of FimC (yellow, for Δt=10s, n=36; Δt=30s, n=14; Δt=45s, n=23 and Δt=60s, n=16). Refolding rates were determined to be 0.021 ± 0.008 s^{-1} (-FimC) and 0.022 ± 0.006 s^{-1} (+FimC). (**b**) Fitting to a single-exponential function of the Probability of refolding of reduced FimG in the absence (green, for Δt=10s, n=15; Δt=30s, n=15; Δt=45s, n=16 and Δt=60s, n=22) and in the presence of FimC (yellow, for Δt=10s, n=23; Δt=30s, n=15; Δt=45s, n=14 and Δt=60s, n=14). Refolding rates were determined to be 0.137 ± 0.034 s^{-1} (-FimC) and 0.027 ± 0.011 s^{-1} (+FimC).

On one hand, the refolding probability of the oxidized FimG in the presence of FimC increases moderately if we compare it with the non-treated protein. However, this trend is not observed in the reduced FimG. This form of the protein, both in the absence and in the presence of FimC refolded in a lower extent than the oxidized form. In fact, the presence of FimC did not change the refolding probability at all. This indicates that FimC only acts on disulfide bonded subunits, being this feature crucial for the refolding process. On the other hand, the refolding kinetics of FimG does not change in the presence of FimC (~0.02 s^{-1}), neither in the oxidized (~0.02 s^{-1}) neither in the reduced forms of FimG (~0.03 s^{-1}). In fact, the fastest refolding rate is observed in reduced FimG (~0.14 s^{-1}), but the maximum is reached very early. This last finding could be counterintuitive since it was expected a

higher rate of refolding in the presence of FimC. However it looks like FimC helps in the refolding of oxidized FimG but does not accelerate it.

Besides the folding/refolding of FimG, other interesting features can be deduced from constant force experiments. As we discovered in the FX measurements the reduced domain unfolds at lower force (339±27 pN) than the oxidized one (440±36 pN), so it is plausible that at constant force the reduced domain should unfold faster than the oxidized one. **Fig. 6.13** shows the cutted, summed, and averaged traces of FimG oxidized and reduced when submitted to a constant force of 300 pN. As expected and in good agreement with the FX data, the reduced FimG unfolds at a faster rate ($1/\tau_1$~10 s^{-1} and $1/\tau_2$~182 s^{-1}) than the oxidized one ($1/\tau_1$~0.5 s^{-1} and $1/\tau_2$~10 s^{-1}). The data shown in this figure is well fitted by a double-exponential function. A reasonable explanation for this behavior could be that the molecule is first submitted to a 160 pN pulse for 2 s and then the force is clamped at 300 pN for 20 s. This first pulse at 160 pN could be biasing the unfolding rate of FimG at 300 pN, triggering the early unfolding of some domains. However even if these measurements cannot be assumed as "pure" dwell-times, it is clear reduced FimG is less mechanically stable not only at constant pulling speed but also at constant force. This trend is expected to be the same for the other Fim proteins from the type I pilus.

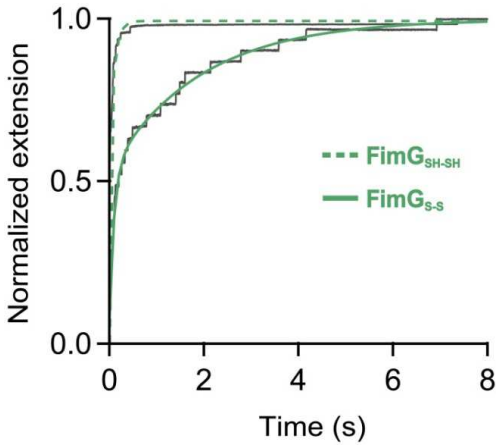


Figure 6.13. Unfolding rate at 300 pN. Double-exponential fitting of the summed and averaged traces of the unfolding of oxidized ($FimG_{S-S}$, n=29) and reduced FimG ($FimG_{SH-SH}$, n=58) at 300 pN ($FimG_{S-S}$: $1/\tau_1$~0.5 s^{-1} and $1/\tau_2$~10 s^{-1}; $FimG_{SH-SH}$: $1/\tau_1$~10 s^{-1} and $1/\tau_2$~182 s^{-1}).

FimG oxidation and refolding in the presence of DsbA

Now that we know in which extent FimC contributes to the folding of FimG, we decided to study the whole process of subunit maturation before being incorporated to the pilus. The main actors involved in this process are FimC and DsbA. Combining them at the same time we can have a clearer and more complete picture of the maturation process. But before testing if they cooperate and have a synergistic effect on the folding of FimG, we need to test first the effect of DsbA on FimG.

A similar protocol to the one followed for FimC was used for DsbA. The oxidation state of both FimG and DsbA were difficult to fix along the experiments.

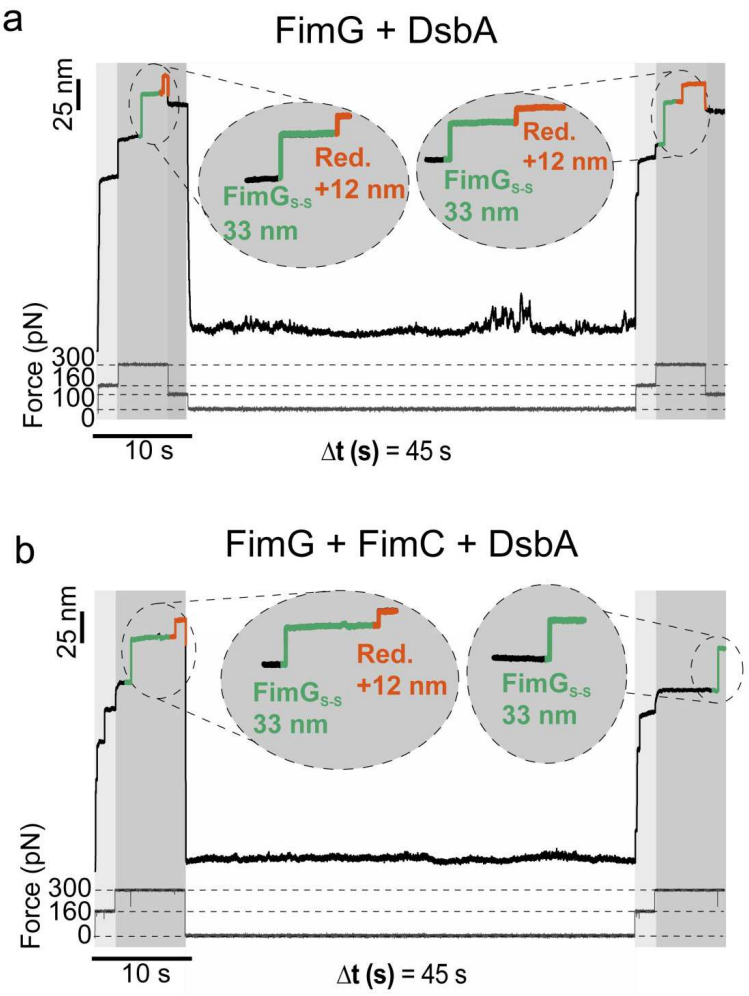


Figure 6.14. FC traces of FimG in the presence of DsbA and in the presence of DsbA and FimC. (**a**) Force-clamp trace of FimG in the presence of DsbA. Protein was submitted to a 4-pulse protocol (or 3-pulse protocol). First it was stretched at 160 pN for 2 seconds (I91 fingerprint detection) and then the force was increased to 300 pN for 5 seconds (or 7 seconds) for FimG unfolding. After that, for some traces (the ones submitted to 5 seconds at 300 pN) the force was quenched for 2 seconds to 100 pN in order to ease the reduction of the protein by DsbA. Then, force was quenched to 0 pN for 45 seconds. In order to test if FimG refolded and if its disulfide bond was reformed, we applied the same pulse protocol. I91 unfolding steps appear as a 25 nm increase in length meanwhile oxidized FimG yields a 33 nm step size increase, and its reduction produces a 12 nm increment in length. (**b**) Force-clamp trace of FimG in the presence of both FimC and DsbA. The same force protocols as for FimG with DsbA experiments were applied.

At least a 20% of the FimG molecules stretched are in the opposite oxidation state of the expected one. This means that oxidized molecules could be detected even if the protein came from a purification in which DTT was added for disulfide bond reduction, and vice versa with the oxidized samples. Besides, along the experiment the oxidation state of the molecules oscillates since every time an oxidized DsbA induces disulfide bond formation on a FimG, a reduced DsbA is generated so the opposite reaction (disulfide bond reduction) can be also detected later. For that reason, after the fingerprint detection with a force pulse at 160 pN for 2 s, we applied either a pulse of 300 pN for 7 s or two sequential pulses of 300 pN for 5 s followed by a pulse of 100 pN for 2 s. This second pulse of 100 pN was used in order to ease the reduction of oxidized FimG by reduced DsbA, since it has been demonstrated that the enzymatic mechanochemical reduction of disulfide bonds is impaired at high forces[194]. After these force pulses, the force is quenched to 0 pN for 45 s and then the same pattern is repeated for checking both the refolding and the oxidation state of the FimG.

Ideally the experiments are done with reduced FimG and oxidized DsbA. **Fig. 6.14a** shows one trace of the unfolding and refolding of FimG in the presence of DsbA. Upon the unfolding of the I91 domains, the unfolding of oxidized FimG (~33 nm) is detected. Then, the reduction of the disulfide bond of FimG takes place even before the force is quenched to 100 pN, releasing the previously sequestered sequence (~12 nm). During 45 s it is allowed the refolding of FimG in the presence of DsbA. When the force pulse protocol is repeated, we observe that FimG appears again oxidized (~33 nm) and it is reduced by DsbA once more meanwhile held at 300 pN (~12 nm).

From this trace we can see that FimG's disulfide experienced three oxidation states changes. First it was registered oxidized, and then it was reduced by DsbA. After the 45 s allowed for refolding, we detected that DsbA oxidized the disulfide and upon unfolding it reduced it again. DsbA is able to both oxidize and reduce disulfide bonds, depending on its own oxidation state. From this kind of experiments it is difficult to say if it is the same DsbA molecule involved in all of these oxidoreduction events. Most probably, two populations of different oxidation states of DsbA molecules live together since the beginning. Also during the experiment these two populations shift from one state to the other as long as they are reacting with the FimG molecules. Also, we detected the disulfide bond reduction during the 100 pN quench, but apparently the enzyme catalyzed reduction of FimG was not impaired at 300 pN and for that reason we extended the 300 pN pulse for 2 seconds more and remove the 2 s at 100 pN.

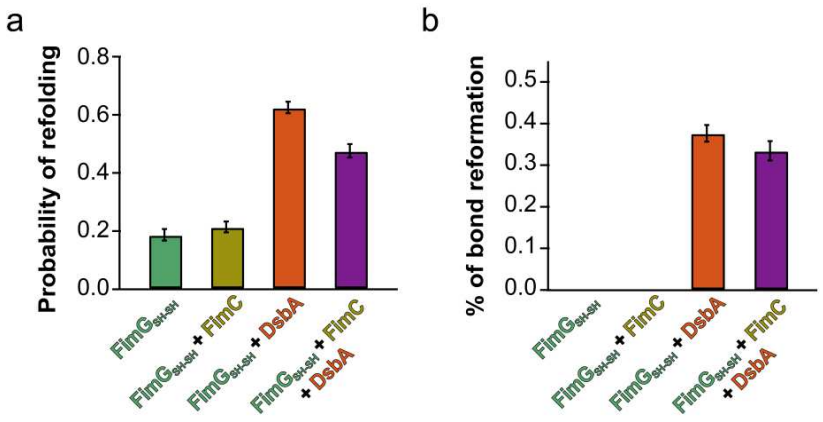


Figure 6.15. Refolding probability and percentage of disulfide reformation of FimG in the presence of DsbA and in the presence of DsbA and FimC. (**a**) Refolding probability of FimG after 45 seconds of quenching time under different conditions ($FimG_{SH\text{-}SH}$, n=16 $FimG_{SH\text{-}SH}$ + FimC, n=14, $FimG_{SH\text{-}SH}$ + DsbA, n=16, $FimG_{SH\text{-}SH}$ + FimC + DsbA, n=21). (**b**) Percentage of disulfide bond reformation of FimG after 45 seconds of quenching time under different conditions (same dataset as specified in (**a**)).

Fig. 6.15a shows the refolding probability of FimG under several circumstances. Above all the maximum refolding probability is achieved when in the presence of DsbA (~60 %), indicating that DsbA is increasing the refolding probability of FimG in a greater extent than FimC does.

In **Fig. 6.15b** is represented the percentage of bond reformation under different conditions. We assume bond reformation when FimG is detected in its reduced state before the quenching time (~45 nm or ~33 + ~12 nm) and it is detected in the oxidized state after the quenching (~33 nm). This implies DsbA formed the disulfide bond during the 0 pN quench. Only in the presence of DsbA it is possible to detect the bond formation, since FimC it is unable to induce its formation.

FimG maturation both in the presence of DsbA and FimC

As it was explained the maturation of pilus subunits implies the sequential action of DsbA followed by FimC. As the Fim preproteins are exported to the periplasmic space, DsbA should recognize their free thiols as they appear in the periplasmic side of the inner membrane. Once the disulfide is formed the protein is then delivered for FimC, so this chaperone can stabilize the preproteins before their incorporation into the pilus. We tested both the refolding probability and the disulfide bond formation as we did with the experiments in the presence of only DsbA. **Fig. 6.14b** shows an example trace of FimG, which is reduced by DsbA and oxidized again during the quenching time, being detected oxidized in the second part of the trace. In this case, the refolding probability is increased to almost ~50 % (**Fig. 6.15a**), and the bond reformation is on the same range as the previously detected with DsbA only (**Fig. 6.15b**). These experiments resemble the natural process of

pilus subunit maturation for pilus biogenesis. More importantly, these experiments uncover DsbA as the true chaperone of this system, being able by itself to increase the ability of refolding of FimG. Also proves that the presence of both proteins do not have a synergistic effect on the refolding probability of FimG, being greatly determined by DsbA more than from FimC.

6.2 Nanomechanics of cell-surface protein CD4

This part of the results is part of a publication from 2014 in the journal *ACS nano*[204]. The results presented here are part of the research conducted by Raul Perez-Jimenez, Alvaro Alonso-Caballero, Ronen Berkovich, David Franco, Ming-Wei Chen, Patricia Richard, Carmen L. Badilla and Julio M. Fernandez. Specifically, my contribution to this work was the study of CD4D1D2 at low pulling speed and low constant force. From the latest it was possible to obtain unfolding rates at the lowest tested constant force, making possible to extrapolate an unfolding rate value at zero force.

6.2.1 Nanomechanics of the HIV binding domains of CD4: D1D2

In order to study the mechanical properties of the first two domains of CD4, a polyprotein was designed. The sequence of CD4D1D2 was flanked on both sides with two copies of the I91 protein, as we did with the Fim proteins. Same conditions as before were used for studying the mechanical stability of CD4D1D2 when stretched at constant speed. The kind of pulling velocities exerted *in vivo* by a HIV-1 particle bound to a CD4 molecule are unknown, so besides pulling CD4D1D2 at 400 $nm \cdot s^{-1}$, we decided to pull it also at a lower speed of 10 $nm \cdot s^{-1}$, in order to test more physiological conditions. Pulling at lower speed yields lower unfolding forces, following a lineal relation, something that has been already demonstrated in other proteins and in I91 too[155].

Fig. 6.16 shows the data collected for both conditions and some example FX traces.

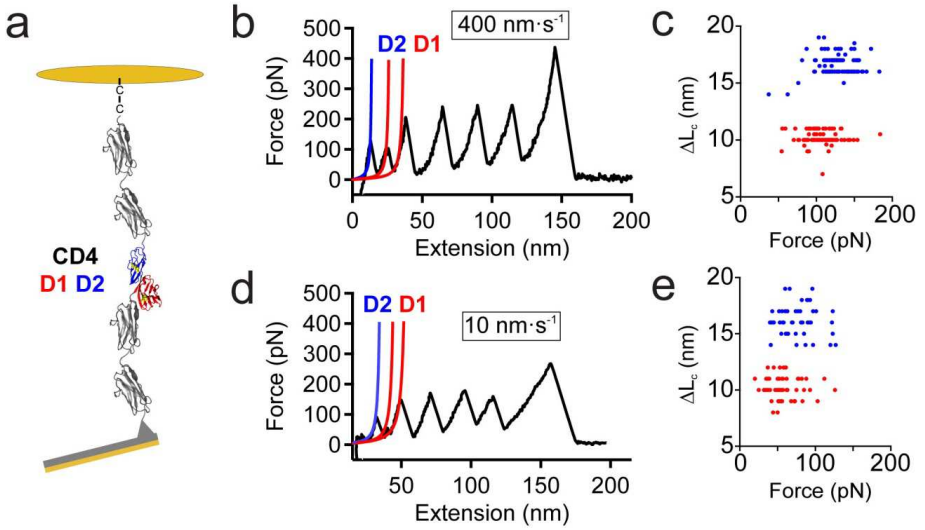


Figure 6.16. Mechanical stability of CD4D1D2 domains. (**a**) $(I91)_2$-CD4D1D2-$(I91)_2$ polyprotein in the AFM set-up. (**b**) FX trace of $(I91)_2$-CD4D1D2-$(I91)_2$ pulled at 400 nm·s^{-1}. The unfolding peaks of the two domains are fitted with the WLC model (D2 in blue, D1 in red). (**c**) ΔL_c (nm) *vs* unfolding force (pN) scatter plot of $(I91)_2$-CD4D1D2-$(I91)_2$ pulled at 400 nm·s^{-1} (D1, n = 132; D2, n = 125). (**d**) FX trace of $(I91)_2$-CD4D1D2-$(I91)_2$ pulled at 10 nm·s^{-1}. (**e**) ΔL_c (nm) *vs* unfolding force (pN) scatter plot of $(I91)_2$-CD4D1D2-$(I91)_2$ pulled at 10 nm·s^{-1} (D1, n = 56; D2, n = 51).

Both CD4 domains are weaker than the I91 domains, in contrast to the results obtained for type I pilus proteins. Interestingly and independent of the pulling speed, a mechanical hierarchic behavior is observed since D2 domain most of the times unfolds before D1 domain, but even more interestingly is the fact that D2 unfolding force (119 ± 32 pN, 16 ± 4 nm, the theoretical value up to the disulfide bond is 20 nm) is greater than the one of D1 (101 ± 30 pN, 10 ± 1 nm, the theoretical value up to the disulfide bond is 12 nm). This counterintuitive mechanical behavior could be explained by the existence of a shared β-strand between D1 and D2. D1 mechanical stability probably is compromised upon the unfolding of D2 and the loss of this shared structural element. With this design D2 would be

acting as a mechanical protector of the D1 domain. As expected, when pulling at 10 nm·s^{-1} the unfolding forces of both domains decrease (D1, 57 ± 21 nm; D2 75 ± 23 nm). It would be expected that a viral particle could exert these kind of pulling forces more probably than the forces obtained at higher pulling rates, resembling more accurately and realistically the initial stage of the infection.

The mechanical extension of CD4 could be important for HIV-1 infection, but we do not know how this event could happen. Besides the constant speed, we decided to apply a constant force to trigger the extension of CD4D1D2, in order to know at which time scale this could be happening. We used a double pulse force protocol: first we applied a pulse of forces ranging from 20 to 100 pN, and then we increased the force to 150 pN for the I91 fingerprint detection. We decided to cover again this wide range of forces since we do not know how much force a virus particle could exert meanwhile attached to CD4. **Fig. 6.17a** shows an example trace obtained when applying 20 pN for 30 s. After collecting enough traces under each force tested, the CD4D1D2 unfolding traces were summed and averaged and the data is fitted with a single-exponential function, as can be seen in **Fig. 6.17b**. From these fittings the unfolding rate of both domains is calculated at each different force. **Fig.6.17c** plots the unfolding rate of CD4D1D2 *vs* force. Using Bell-Evans-Ritchie model it can be extrapolated the unfolding probability at zero force (0.08 s^{-1} at 0 pN). This could mean that CD4 could be mechanically extended at very low forces, in the range of what a viral particle could exert. The CD4 extension at low forces could even happen through intermediates, as we observed for the unfolding of the D2 domain in approximately 5% of the traces.

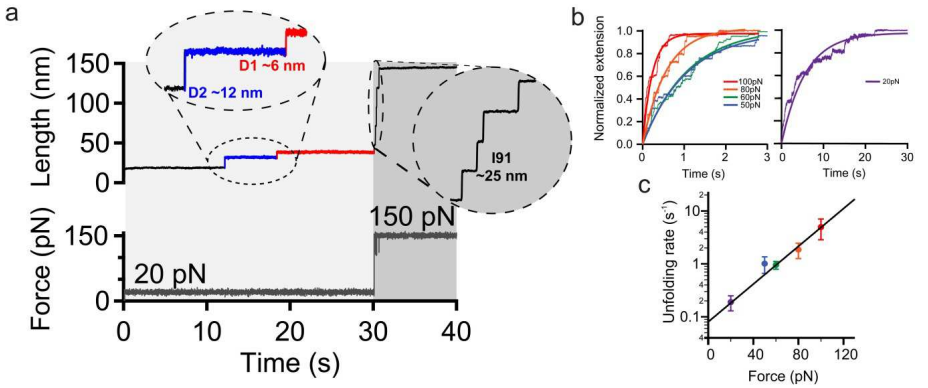


Figure 6.17. Mechanical stability of CD4D1D2 domains. (**a**) FC trace of $(I91)_2$-CD4D1D2-$(I91)_2$ pulled at 20 pN for 30 s and then pulled at 150 pN for the fingerprint detection. (**b**) Summed and averaged unfolding traces of CD4D1D2 at different forces fitted with a single-exponential function (At 100, n = 11; at 80 pN, n = 8; at 60 pN, n = 26; at 50 pN, n = 11; at 20 pN, n = 23). (**c**) Unfolding rate *vs* constant force of CD4D1D2. An unfolding rate of 0.08 s^{-1} is observed if an extrapolation to zero force is done. Data shown is the unfolding rate determined from the single-exponential fittings of panel (**b**) and the error was calculated with the bootstrap method.

6.2.2 CD4 extension correlation with HIV-1 Infectivity

Experiments done on 293T cells demonstrated a strong dependency between the extension of CD4 and the infectivity by HIV-1[152]. Cells expressing variants of the CD4D1D2 domains connected to the cell membrane with flexible linkers of 2, 6, 12 and 18 nm allowed the infectivity with an increasing efficiency related with the length (**Fig. 6.18a**). The total length of this constructs was the length of the flexible linkers plus the length of the folded domains D1D2, which was determined to be around 6 nm based on the X-ray structure of CD4 (PDB code: 1WIP). **Fig. 6.18b** shows the relative infectivity *vs* the contour length of the construct with the different linkers. Longer linkers are correlated with higher infectivity, however the folded wild-type

CD4 (CD4WT, black dot) does not follow this trend. If we assume that two of the domains of CD4WT are unfolded the extension is increased (CD4WT, red dot) and the correlation between the infectivity and the total extension is well fitted by a single-exponential function.

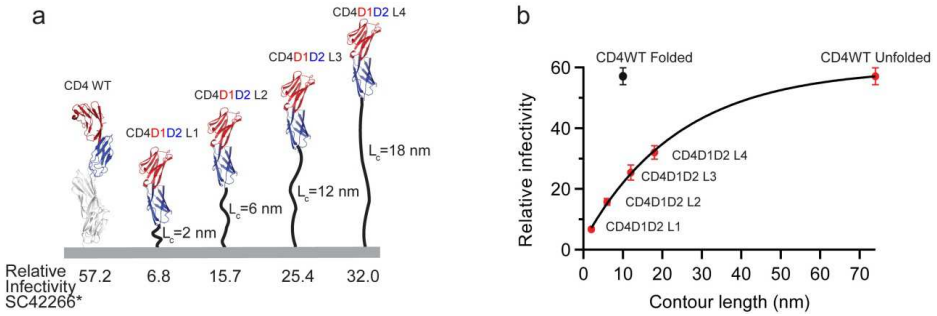


Figure 6.18. Correlation between HIV-1 infectivity and the total extension of CD4. (**a**) Schematic representation of the constructs used in[152], with the wild-type CD4 (CD4WT) and four constructs made of flexible linkers of different length that connect the cell membrane with the domains CD4D1D2. The linkers are made of repetitions of the peptide GGGGS. The relative infectivity (relative light unit, RLU) in SC42266 isolates is indicated below each construct, being maximum for CD4WT. Data extracted from[152]. (**b**) Relative infectivity *vs* contour length extension for the different constructs. In all of them is considered that the total contour length comes from sum of the folded CD4D1D2 domains and the complete extension of the linkers assuming a contribution of 0.4 nm per aminoacid. The black dot corresponds with the infectivity of the CD4WT assuming its folded. The red dot for CD4WT assumes that two domains are unfolded, following the same trend as the Linker-CD4D1D2 constructs follow.

The mechanical extension of a protein is well described in the low force regime by the freely jointed chain (FJC) model for polymer elasticity. During the hypothesized mechanical unfolding of CD4 triggered by a HIV-1 viral particle one or several domains may be unfolded, contributing to the final extension of CD4. We applied the FJC model to predict the final length of the four Linker-CD4D1D2 constructs when different forces are applied. First the relation between the extension and the force, $x(F)$, is described by the FJC model:

$$x(F) = x_c\left[tanh^{-1}\left(\frac{Fa}{k_BT}\right) - \frac{k_BT}{Fa}\right]$$

in which x_c is the total contour length of the molecule taking into account both folded and unfolded domains, F is de applied force, a is the Kuhn length, k_B is the Boltzmann constant and T is the temperature. Now the infectivity relationship with the force-induced extension of CD4

$$I(x) = I_0 - Aexp\left\{\frac{x(F)}{x_0}\left(tanh^{-1}\left(\frac{Fa}{k_BT}\right) - \frac{k_BT}{Fa}\right)\right\}$$

The maximum value of infectivity for CD4WT is 57 RLU (I_0) and in this model is considered an asymptote (see **Fig. 6.19**). A is the length of the folded protein at zero infectivity as $x_0 ln(I_0/A)$ ~ 5.3 nm, and x_0 is a length-scale coefficient of the infectivity at a given force. **Fig. 6.19** shows the behavior of the different linkers with CD4D1D2 and the CD4WT at different forces. Also is shown how the unfolding of the domains affects the final extension. It is clear that at least two domains of CD4WT should be unfolded in order to follow the same correlation as the Linker-CD4D1D2 variants do, as it can be seen in **Fig. 6.19a and b**.

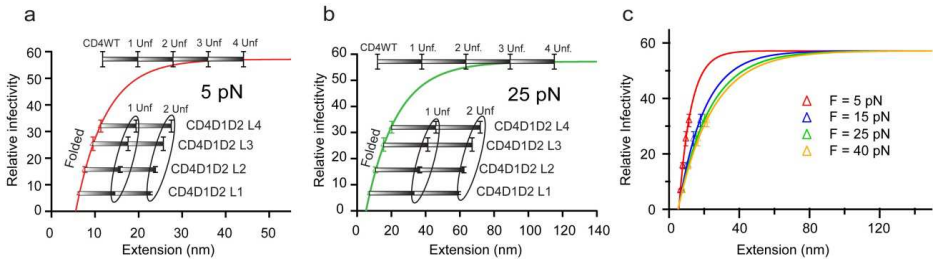


Figure 6.19. Infectivity *vs* extension of CD4 constructs calculated from the FJC model. (**a**) Infectivity *vs* calculated extension at 5 pN for the different constructs, both folded and unfolded. (**b**) Same as (**a**) but at 25 pN. In both panels the grey bars indicate the extension of

unfolded domains. (**c**) Infectivity *vs* extension at 5, 15, 25 and 40 pN of aplied force. In the three panels the lines are exponential fittings of the function that relates infectivity with the force-induced extension of the protein. In panels (**a**) and (**b**) it can be seen that at least the unfolding of two domains of CD4WT are necessary in order to follow the same correlation as the other Linker-CD4D1D2 constructs follow.

6.2.3 Mechanical effect of the HIV-1 neutralizing antibody Ibalizumab on CD4D1D2 domains

The previous subsection suggests that a longer extension of CD4 improves the infectivity by HIV-1 virus. This hypothesis suggests that a mechanically more stable CD4 in which the unfolding of its domains requires higher forces would reduce the chances of infection. Ibalizumab is a humanized monoclonal anti-CD4 antibody that blocks HIV-1 infectivity binding to the interface between domains D1 and D2 of CD4 with a very high affinity, and it is not clear why it is such a good inhibitor of HIV-1 infection. We decided to test if the mechanical stability of the polyprotein $(I91)_2$-CD4D1D2-$(I91)_2$ changes in the presence of the antibody Ibalizumab.

FX measurements at 400 $nm \cdot s^{-1}$ revealed that CD4D2 domain was stabilized in ~20 pN by the presence of Ibalizumab, meanwhile the CD4D1 domain mechanical stability remained unchanged (CD4D1, 101 ± 30 pN; CD4D1+Iba, 100 ± 24 pN; CD4D2, 119 ± 32 pN; CD4D2+Iba, 138 ± 23 pN). A striking feature of the FX traces is the position of the unfolding peaks of CD4D1D2 domains (**Fig. 6.20b, c and d**), where we are able to detect them after the unfolding of one, two or even three I91 domain unfoldings. The mechanical stabilization exerted by the presence of Ibalizumab is not very different when it is

not present and one would expect that the weaker domains are the first ones unfolding when the protein is stretched, but the fact that the unfolding peaks of CD4D1D2 appear after the unfolding of I91 domains whose unfolding forces are ~200 pN indicates that clearly Ibalizumab is altering the mechanical properties of CD4D1D2.

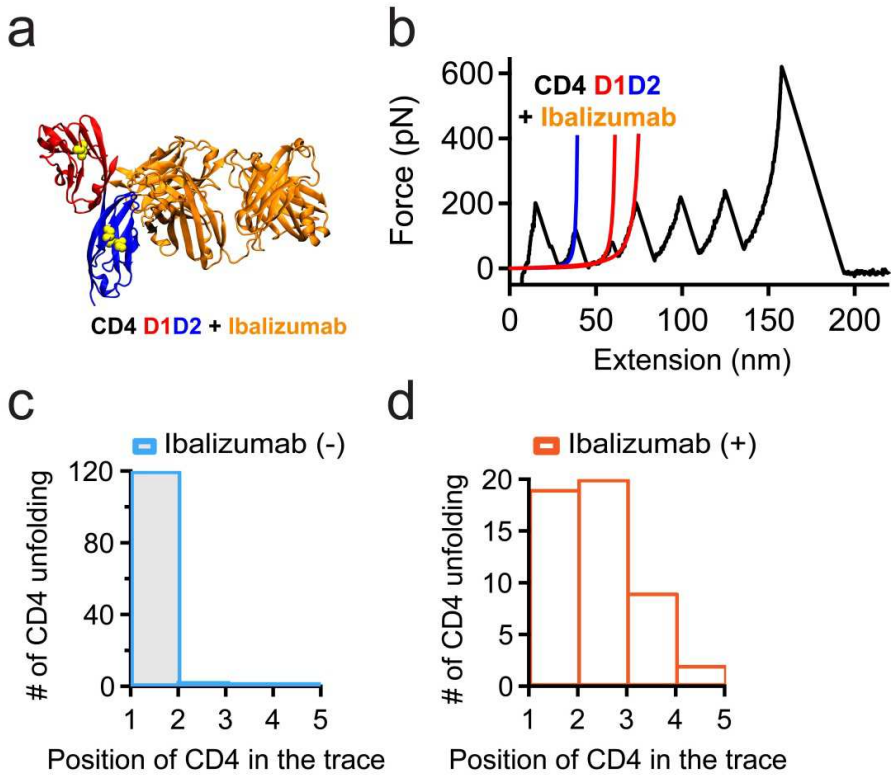


Figure 6.20. Mechanical stability of CD4D1D2 in the presence of the antibody Ibalizumab. (**a**) Cartoon representation of CD4D1D2 with the Ibalizumab antibody bound to it (PDB: 3O2D). (**b**) FX trace of $(I91)_2$-CD4D1D2-$(I91)_2$ pulled at 400 nm·s^{-1} in the presence of Ibalizumab. (**c** and **d**) Position of CD4D1D2 unfolding peaks in the FX traces in the absence (**c**, n = 123) and in the presence (**d**, n = 50) of Ibalizumab.

The force-ramp (FR) experiments in which we pull the protein at a constant force rate of 33 pN·s^{-1} (**Fig. 6.21a** and **b**) revealed the late simultaneous unfolding of the CD4D1D2 tandem in the presence of

Ibalizumab. In **Fig. 6.21c** and **d** is shown the step length increases of CD4D1D2 in the absence and in the presence of the antibody. Both domains present a tendency to unfold in a single ~22 nm step at elevated forces (up to ~250 pN), opposite to the unfolding behavior with no antibody where both domains unfold separately at low forces. The initial unfolding forces (**Fig. 6.21e**) under each condition confirm that the Ibalizumab presence shifts the forces from ~80 pN to ~150 pN.

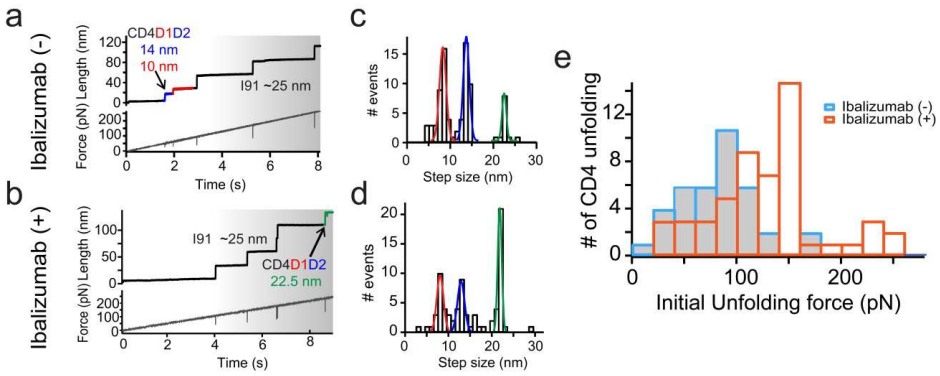


Figure 6.21. Mechanical stability of CD4D1D2 in the presence of the antibody Ibalizumab. (**a** and **b**) FR trace of $(I91)_2$-CD4D1D2-$(I91)_2$ pulled at 33 pN·s^{-1} in the absence (**a**) and in the presence (**b**) of Ibalizumab. (**c** and **d**) Step size length increase histograms in the absence (**c**, n = 72) and in the presence (**d**, n = 76) of Ibalizumab. In the presence of the antibody there are more events of CD4D1D2 unfolding in one single step of ~22 nm. (**e**) Initial unfolding forces in the absence (n = 37) and in the presence (n = 54) of Ibalizumab.

6.2.4 Mechanochemical extension of CD4D1D2 through the reduction of its disulfide bonds by thioredoxin

Disulfide bonds tend to increase the thermodynamic stability of proteins[166]. From the mechanical point of view, the presence of disulfide bonds increases the robustness of proteins however this depends on the position and the structural features of the protein that

the disulfide in bridging. Three out of four of the extracellular domains of CD4 present a disulfide bond in its structure, being CD4D3 domain the only one lacking this bond. It looks important during HIV-1 infection the fine tuning of the oxidation state of the disulfides both from CD4 and gp120 viral protein[149]. The inhibition of oxidoreductase enzymes apparently prevents HIV-1 infection, and the regulation of the disulfide of CD4D2 domain seems to be managed by thioredoxin[141,148]. Besides, HIV-1 is prone to infect cells where CD4D2 domain disulfide is reduced[146]. However, the disulfide of CD4D2 is buried in the inner core of the protein so thioredoxin is not able to reach it. We hypothesized that mechanical unfolding of CD4 triggered upon the attachment of the HIV-1 viral particle could expose this cryptic disulfide and make it available for disulfide/thiol exchange by thioredoxin.

In **Fig. 6.22b** it is shown a FC trace of the construct in the presence of thioredoxin. Upon the unfolding of CD4D1D2, the buried disulfides of both domains are exposed and susceptible of being reduced by thioredoxins. CD4D2 reduction is detected as a ~8 nm step that overlaps with the length observed for the oxidized CD4D1 (in **Fig. 6.22c**, the first distribution of the histogram highlighted with a grey Gaussian curve), and then the reduction of CD4D1 is detected as an increase of ~20 nm. In some of the traces it was possible to register the unfolding of CD4D1D2 but only the reduction of CD4D2. This could be related with a different dependency of force for the enzymatic cleavage of the disulfide of these two domains. The unfolding of CD4D1D2 domains already reduced would yield length increments of ~29 nm and ~22 nm, respectively. However these events are rare, supporting the idea that their disulfide bonds remain oxidized in the

core of these two domains and are only available for reduction upon their mechanical unfolding.

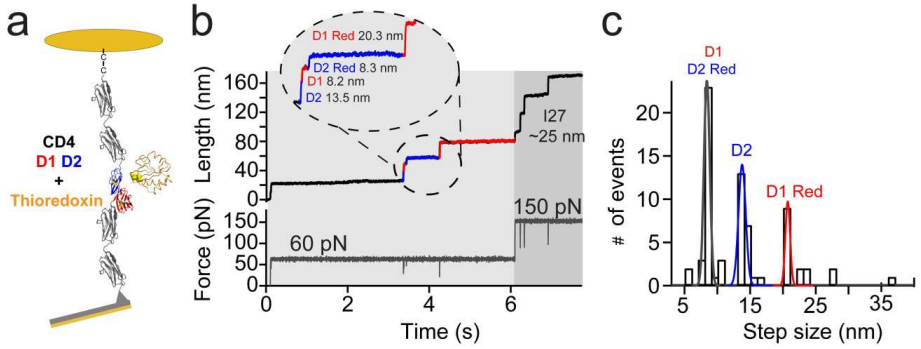


Figure 6.22. Force-induced exposition of CD4D1D2 cryptic disulfides and reduction by thioredoxin. (**a**) $(I91)_2$-CD4D1D2-$(I91)_2$ polyprotein in the AFM set-up in the presence of human thioredoxin (PDB code: 1UVZ). (**b**) FC trace of $(I91)_2$-CD4D1D2-$(I91)_2$ pulled at 60 pN in the presence of thioredoxin. After the extension of the oxidized domains of CD4D1D2, the reduction of CD4D2 takes place followed by the reduction of CD4D1. Then the force is increased to 150 pN to detect the fingerprint (**c**) Histogram of the step size increase in length of CD4D1D2 in the presence of thioredoxin (n = 25). Three population of steps are detected. The first one of 8 ± 1 nm corresponds both with the unfolding of CD4D1 and the step size increase produced after the reduction of the disulfide of CD4D2. The second one of 13.4 ± 0.7 nm corresponds with the unfolding of CD4D2 and the third one shows the reduction of CD4D1 domain with 20.2 ± 1.8 nm.

Chapter 7: Discussion

In this final chapter the results obtained in this thesis will be discussed and compared with the most recent available literature for the two issues covered. A final section is used to conclude and connect the main ideas and findings achieved. Finally a focus on the prospective research which could be conducted in the Mechanomedicine framework for the development of new therapeutic strategies is given.

7.1 Nanomechanical architecture and biogenesis of the type I pilus

UPEC type I pilus is a filamentous multiproteic structure used by bacteria as a virulent factor for uroepithelium attachment. UTIs develop under chemical and physical harsh conditions which pathogenic bacteria have to fight against to in order to colonize the urinary tract. Bacteria ascend from the urethra to the bladder, and then to the ureters and kidneys. Along this path, but especially from the urethra to the bladder, bacteria have to advance and avoid to be dragged away during micturition. Type I pili provide mechanical anchors that help bacteria to remain attached when the shear forces generated by the urine flow push their cell bodies out of the urinary tract.

Type I pilus has been widely studied as it is one of the reasons of the persistence and recurrence of UTIs, and it represents a potential target for therapeutic treatment. These appendages are highly adapted

for their task and for the environment where they are required, and several features are responsible for this specialization.

First, the presence of FimH protein at the pilus tip end. UPEC/Uroepithelium attachment occurs as a protein-ligand interaction between mannose-containing proteins from the epithelium and FimH lectin domain. This bond is allosterically regulated by FimH pilin domain, whose interfacial contacts with the lectin domain down-regulate the affinity for the mannose[205,206]. This behavior permits the colonization and spreading of bacteria along the urinary tract since the lectin-mannose interaction is weak and the bond rupture rate is high, allowing bacteria rolling and spreading. But in the event of an urine flow burst the lectin affinity for mannose increases as a consequence of the tensile force experienced by FimH[65]. As urine flow pushes bacteria cells their pili, establishing weak contacts with the epithelium through their tip-end FimH, start to be stretched. This stretching force finally is transmitted to the mannose-lectin bond and then to the interfacial contacts between the lectin and pilin domain of FimH. A conformational rearrangement induced by the tensile stress aligns both domains along their longitudinal axis, disrupting the inhibitory effect of the pilin domain and leading to a lectin high affinity state for mannose. This finger trap-like toy mechanism is known as a catch-bond and it is optimized for bacterial spreading in the urinary tract meanwhile prevents bacteria detachment[64,68,71].

Second, the pilus rod quaternary structure provide with an additional mechanical feature which favors both pilus and catch-bond integrity. Although the whole pilus is a passive structure its architecture is dynamic and experiences conformational changes upon the

application and removal of tensile force[74]. Almost the entire pilus is made of FimA subunits which form the rigid pilus rod. These domains arrange in a right-handed helical configuration forming a 7 nm diameter structure, presenting 3.6 stacked subunits per helix turn. The rod helical conformation is stabilized through the turn-to-turn weak interactions made between the residues of non-sequential subunits, occurring between the *i* domain and the *i* + 3 domain[77,207]. Upon tensile stress application the pilus is completely stretched and the tension is finally transmitted to the lectin-mannose catch-bond. Catch-bonds experience an increased affinity for the ligand and an increased bond life time up to an optimal force, but above that force the bond rupture probability increases. Before the catch-bond interaction is broken, pilus rod uncoiling takes place and the different layers of stacked FimA subunits start to break the weak contacts that held their configuration. The rod starts unraveling in the closest stacked layer to the pilus tip, and progresses along it extending the length up to 5 to 6 times[77,208,209]. This unraveling behavior protects the catch-bond from being broken during the shear flow of the urine, maintaining an optimal force at the protein-ligand interface[210,211]. It is important to bring the attention to the fact that each bacteria cell is surrounded by hundreds of pili. Many of these appendages can be attached to the uroepithelium through catch-bonds and the tensile force can be distributed among several of them, protecting the structural integrity of the individual pilus[76]. Once the tensile force is removed, i.e. when the urine flow finishes, the unraveled pili recoil reforming the turn-to-turn contacts of the subunits that lead to the quaternary structure. This re-stacking process shortens the length of the extended pilus and this contraction enables the bacteria to advance and recover their pre-stress position[76,210].

All these features confer elastic tether properties to the type I pilus whose overextension and recoiling behavior is optimized for extending the bond lifetime at its tip-end, ensuring bacteria attachment to the uroepithelium but also allowing the spreading and rolling at low or zero shear force. One of the most surprising qualities of these structures is the fact that the proteins which form the pilus are connected head-to-tail through non-covalent interactions, donating each subunit an N-terminal β-strand to the incomplete Ig-like fold of the previous subunit in the assembly. This interaction is also specific for each subunit protein, being what finally determines the order of assembly during pilus biogenesis. It has been demonstrated that this hydrophobic interaction is one of the strongest found in nature, conferring high kinetic stability against dissociation or unfolding once the antiparallel β-strand complementation is done[96,97]. High energy barriers against unfolding guarantee the integrity of the pilus structure preventing its disassembly. Hundreds of subunits form the pilus and the spontaneous rupture of only one single donor-strand interaction would lead to tether breakage and detachment.

In this thesis we have explored the mechanical stability of the four Fim proteins that form the type I pilus. We designed self-complemented proteins where the cognate donor β-strand of each subunit was placed in its C-terminal sequence. In the tertiary folded structure of the domain this design ensures the correct antiparallel insertion of the β-strand, forming the native contacts made in the mature pilus. During pilus stretching *in vivo*, after the pilus quaternary structure has been completely unraveled, the force is transmitted along the head-to-tail axis of the subunits. This longitudinal pulling is what finally challenges the resistance of the hydrophobic interaction between

proteins. With smFS we could resemble the geometry of the force experienced by the pilus and interrogate the mechanical strength of the donor-strand interaction. Although the thermal stability of these proteins has been explored, their evident exposition to mechanical forces has not been addressed, and a mechanical description of them is required for the development of new therapeutical strategies oriented to target bacterial attachment.

We first applied constant speed experiments pulling at 400 $nm \cdot s^{-1}$ to these constructs finding strikingly high unfolding forces for all of them, standing out as some of the most mechanically resistant proteins discovered until today, in the range of the values obtained for cohesin I modules of scaffoldin[212]. In the case of FimA (**Fig. 6.1**), the most abundant protein in the pilus, the mechanical stability was above 500 pN being the most stable type I pilus Fim protein. This is in agreement with the values obtained for other shaft pilus proteins (most abundant proteins forming a pilus) for Gram-positive bacteria, where SpaA from *Corynebacterium diphteriae* and FimA from *Actinomyces oris* showed the highest unfolding forces ever reported for globular proteins[213]. Unlike type I pilus subunits, these two examples of Gram-positive pili are assembled through interdomain isopeptide bonds, forming a continuous covalent backbone along the whole pilus. In our case the mechanical strength builds from the self-complemented FimA fold (FimA-DS_{FimA}), in which the donor-strand of another FimA subunit is included and mimics the native fold. Taking into account that almost the whole pilus is made of an array of hundreds of FimA subunits, we suggest the strength of this complementation evolved in order to reduce the probability of pilus disassembly due to FimA-FimA

interaction rupture. A mechanically strong pilus rod is mandatory for the UPEC attachment to the urine tract.

We observed an even more interesting property of the type I pilus when we realized that the mechanical stability of Fim domains follows a hierarchical trend. After FimA, FimG and FimF showed the highest unfolding forces (**Fig. 6.2** and **Fig. 6.3**) followed by FimH pilin domain, and finally by the weakest FimH lectin domain (**Fig. 6.4**). This mechanical hierarchy matches the pilus architecture, becoming less stable as we approach to the pilus tip. FimG and FimF are part of what is known to be the hinge region of the pilus. It adopts a hook-like shape in the non-attached pilus which serves the FimH lectin domain to explore the surface for mannose recognition. FimF is complemented with the FimA donor-strand and the first pilus tip protein after the rod. The FimF donor-strand complements the FimG fold. Both Fim proteins have similar unfolding force values, which gives them a similar role when bridging the two very different parts formed by the lectin and the pilus rod. At the end of the pilus tip we find FimH protein, formed by two domains connected by a peptide linker. The pilin domain is complemented by the FimG donor-strand, in the same way as FimA, FimF and FimG. Of all the pilin domains this is the less mechanically stable one, although its values are above the not negligible quantity of 300 pN. Following both the mechanical trend and the pilus architecture, the lowest force value corresponds to the FimH lectin domain, the ligand-recognizing protein of the pilus which sometimes unfolded through an intermediate (**Fig. 6.4b**). This lowest value is not surprising since the lectin domain possesses a jelly-roll topology, an elongated β-barrel made of 11 strands and which conformation changes depending on the allosteric regulation exerted by the pilin domain. It tightly

clamps the mannose ligand when the down-regulation exerted by the pilin domain is removed after tensile force application.

Our findings reveal a high mechanical stability for the four pilin domains and the lectin domain, hierarchically decreasing from the pilus rod to the pilus tip. Although protein-ligand interaction breaks at values below the least mechanically stable domain (FimH pilin domain), we think these donor-strand interdomain connections are designed for guaranteeing pilus integrity in the harsh conditions of the urinary tract. We think this strength is based on the surface complementarity between the residues of the donor-strand and the residues of the fold it complements. The mechanical hierarchy observed in the experiments was partially affirmed by the SMD simulations, which showed a slightly different order of mechanical stability (**Fig. 6.6**), but supported the force hierarchy observed in smFS experiments due to the different number of hydrogen bonds made in the self-complemented proteins (**Fig. 6.7**). The simulations also confirmed us that the self-complemented strategy was suitable for testing the mechanical strength of the complementation, since covalently and non-covalently self-complemented FimH pilin domains showed the same unfolding pattern (**Fig. 6.8**). Although this same design has been used before in bulk experiments in the past[183,184,214], this simulation validates our self-complemented constructs as pilus native-like interactions.

All these previous findings where obtained with disulfide-bonded domains, and the mechanical contribution of this bond to these proteins has not been addressed before. All pilus type I proteins contain conserved disulfides in their structure, including the FimH lectin domain. In the lectin case, the disulfide is placed near the mannose-

binding pocket and its presence has been demonstrated to be crucial for the protein-ligand interaction[215]. In the reduced form, capturing the unfolding of the lectin domain was elusive (**Fig. 6.4e**), and a possible explanation could be the lack of stability or misfolding. But more importantly is the fact that all four pilin domains exhibited decreased unfolding forces in their reduced states, with a lost in the mechanical stability up to 50 % (**Fig. 6.5**). The position of the disulfides is not only conserved among type I pili subunits but also in the proteins from the Dr fimbriae and P pilus, highlighting the stabilizing effect of this bonds in adhesive organelles[79]. In all these fimbrial proteins the disulfide bond bridges A' and B antiparallel strands and confers high thermodynamical and kinetic stability, which together with the strength of the donor-strand interaction confer high resistance to the domains. These configuration places two mechanical clamps on each pilus subunit. On the one hand the disulfide bond, and on the other hand the hydrophobic interaction and the hydrogen bond network made between the donor-strand and the A and F-strands of the fold. When a tensile force is applied, domain unfolding starts from the C-terminal end of the protein which consists in removing the donor β-strand from the fold, as it can be seen in the SMD simulations performed where the force maximums were reached just before the donor-strands are ripped out (**Fig. 6.6**). In the absence of the disulfide bond clamp, the mechanical unfolding pathway shifts and it starts from the N-terminal end, the weakest part of the protein. This different unfolding pathway was revealed in coarse-grained simulations performed and published by our group on oxidized and reduced self-complemented FimG, which showed the mechanical locker effect of the disulfide bond[216]. It is clear that disulfide bonds are relevant for the structural integrity of adhesive

structures and they confer not only thermodynamical stability but also mechanical resistance.

These findings highlight the importance of both the disulfide bond and the strand complementation between domains as the main contributors to pilus mechanical integrity. Hence, these two elements are potential targets for bacterial attachment ablation. The use of reducing agents would break disulfide bonds, and the usage for example of small peptides with the ability of competing with the strand complementation interaction would lead to pilus disassembly. However targeting the mature pilus is not the only option available for disrupting bacterial attachment. The process of pilus assembly *in vivo* is another interesting target for drug therapy, since several events occur before the pilus secretion. We wanted to study even deeper the attachment system of UPEC approaching the process of pilus biogenesis in the bacterial periplasm. For this purpose we used the FimG construct and we monitored its unfolding/folding behavior with constant-force experiments within single trajectories. With this approach we resemble the extended state of the protein when is exported to the periplasmic space through the SecYEG pore system. We observed the refolding probability of disulfide-bonded FimG was greater than the one of reduced domain (**Fig. 6.9**), showing that the disulfide bond as expected contributes to the folded state both enthalpically and reducing the entropy of the chain[217]. It is necessary to consider the residual structure existing in the oxidized FimG, which also contribute to ease the refolding of the stretched protein.

Our next step was to identify how much FimC contributes to the folding of FimG, as this chaperone has been identified as the main actor

for pilus subunits folding in the periplasm. FimC contributed moderately in the refolding of oxidized FimG but its effect was null for the reduced version (**Fig. 6.10**, **Fig. 6.11** and **Fig. 6.12**), supporting the already proposed additional function of FimC of acting as a quality control system which only supplies disulfide-bonded domains to the assembly platform FimD[203]. Unfolding kinetics of both states of the protein revealed a higher rate of unfolding for the reduced protein (**Fig. 6.13**), which directly connects with the lower mechanical stability observed in the FX experiments. Considering the lower mechanical stability of the reduced domains, we think FimC would ensure the incorporation to the pilus of only properly folded and oxidized domains.

It is important to remember that the design of our FimG contains the cognate antiparallel donor-strand provided by FimF (*cis*-DSE), an interaction more favorable than the parallel G1-strand donated by FimC (*trans*-DSC). After quenching to zero force the refolding of FimG can happen without an external help, and this design could be masking the real performance of FimC. However the lack of activity on the reduced domains suggests that FimC recognizes and helps to fold oxidized FimG.

The modest folding effect of FimC was puzzling since previous extensive research has always supported this chaperone as the only factor contributing to pilus proteins folding in the periplasm. We wanted to explore in detail pilus protein maturation involving DsbA oxidoreductase, the enzyme responsible for disulfide bond formation in Fim proteins. Surprisingly, DsbA not only contributes in thiol-pair oxidation but also contributes, and in a greater extent than FimC, to

FimG folding (**Fig. 6.14** and **Fig. 6.15**). We believe this newly uncovered behavior corresponds to a real chaperone effect of DsbA not only enhanced by disulfide-bond formation. Although this is the first time DsbA has been directly identified as a chaperone in type I pilus biogenesis, previous studies already suggested that oxidoreductases from the thioredoxin family could be involved in more tasks than disulfide bond formation.

When the crystal structure of DsbA was reported some structural elements were identified as being capable of binding partially folded proteins[218]. Later a chaperone-like activity of DsbA during the oxidative folding and maturation of the periplasmic chaperone PapD (P pilus type equivalent of FimC) was proposed. Native PapD chaperone possesses a disulfide bond, being its correct formation and the PapD folding critical events for P type pilus assembly[105]. This study could not verify if PapD folding was being facilitated only by thiol-pair oxidation or if there was a chaperone activity in DsbA. Zheng *et al.*[219] identified the chaperone-like activity of DsbA on proteins lacking disulfide bonds, separating it from the thiol-oxidoreductase activity. Therein, DsbA was able to help in the refolding of chemically-denatured protein substrates, binding to unfolded intermediates and preventing aggregation. In the presence of small peptides the DsbA-assisted refolding of these denatured proteins was diminished as a consequence of the competition between the small peptide and the protein substrate, demonstrating the peptide-binding ability of DsbA. DsbA, as its eurkaryotic counterpart PDI, possesses a thioredoxin-like domain which presents a hydrophobic groove just below the active site, which would allow the binding of partially folded proteins before the disulfide bond formation.

Frech *et al.*[220] demonstrated the establishment of non-covalent interactions between DsbA and a protein substrate during the intermolecular disulfide bond complex generated between both species before complete oxidative folding. The fact that DsbA-catalyzed disulfide exchange with unfolded proteins is much faster than the provided one by small molecules like glutathione, suggests that the increased rates are favored by non-covalent interactions between DsbA and the protein substrate. Non-covalent interactions are usually made with chaperones which provide suitable molecular surfaces for the correct folding of the protein substrates, as it happens with the periplasmic chaperone Spy[221,222]. Proteins secreted to the periplasmic space usually contain few disulfide bonds and frequently they can fold before oxidation. It is important that DsbA rapidly recognizes and binds to these proteins meanwhile they are being secreted to the periplasm, thus catalyzing oxidative folding as it is required for each pilus subunit. The mixed disulfide intermediates stabilized through non-covalent interactions would be solved once the protein substrates fold in the surface of DsbA. This would decrease the affinity of the folded protein for DsbA, which would be released in an almost complete folded state or competent for FimC-assisted folding[220].

The fact that when together DsbA and FimC do not increase even more the refolding probability of FimG suggests that DsbA delivers almost completely mature subunits to FimC. In our design folded self-complemented FimG is stable by itself but *in vivo* FimC would stabilize more than helping to fold, the different subunits through parallel DSC and facilitate DSE during pilus incorporation at the FimD platform. Moreover FimC would discriminate between correctly folded and oxidized subunits from the reduced ones, and

forming complex with the individual subunits would prevent subunit oligomerization and aggregation in the periplasm through donor-strand donation.

In summary our results suggest the type I pilus is a mechanically stable structure. This stability is built on the interdomain non-covalent interactions that hold the structure, and this strength decreases hierarchically from the pilus rod to the pilus tip. The presence of disulfide bonds in each of the proteins mechanically clamps the domains, increasing even more their stability. The importance of this intramolecular covalent bond is shown during pilus biogenesis, where chaperone FimC discriminates between oxidized strong domains and reduced weak domains. In our experiments, DsbA was not only inducing disulfide bond formation but also demonstrated to be a true chaperone. This new scenario strongly supports DsbA as the real chaperone during pilus proteins maturation, and FimC as the stabilization chaperone which prevents aggregation and addresses the correct oxidation state of the subunits.

7.2 Nanomechanics of CD4 receptor

CD4 protein is a cell-surface receptor with a crucial role during the adaptive immune response, involved in the stabilization of the interactions between the TCR and MHC-II, and in the triggering of the T-cell activation. HIV uses this receptor to attach to $CD4^+$-lymphocytes, to infect them and to replicate. Therefore the consequence is the destruction of the immune system cells leading to the host immunodepression, leaving it vulnerable to opportunistic

infections and cancer. Targeting the same cells that should be fighting the virus is one of the several reasons of the success of the infection. Another important reason is the virus ability to integrate into the host genome and hide from immunologic surveillance, remaining latent for future replication. This Trojan horse strategy is complemented with morphological features of the virion that make it undetectable to the immune system. The only viral-origin protein exposed on the surface of the virus is the glycoprotein complex Env. This protein is extensively glycosylated, making it invisible to the immune system, and its potential epitopes are subjected to high sequence variability because of the errors made during the retrotranscription of the viral genome. All together, these strategies shape the arsenal of a pathogen which produces a persistent and incurable infection which when non-treated leads to a poor prognosis of the patient.

Since the onset of the world pandemic huge efforts have been done to find a cure and crucial achievements have been accomplished, turning a deadly disease into a chronic one with a significant improvement of the life expectancy of those who live with the infection. Thanks to the knowledge acquired about the virus biology it has been possible to develop drugs that interfere with different steps of its life cycle, however it remains unreachable the complete eradication of the infection once it is started, with the only possibility of maintaining a low baseline viremia load and to avoid further spreading of the infection.

Especially relevant are those drugs used to prevent infection in the first instance. These drugs interfere with the fusion of the virus disrupting conformational changes required for membranes fusion, like

enfuvirtide, or with events taking place before fusion such as the binding of gp120 to the coreceptor, like maraviroc[124]. There are several antibodies targeting the most conserved sites of gp120, the CD4bs and the coreceptor binding site in the V3 loop[127]. The success of these antibodies greatly varies, some of them are effective *in vitro* with laboratory adapted strains but not useful against strains isolated from patients, and others show different effectiveness *in vivo* depending on the isolate. The coreceptor binding site and other gp120 susceptible regions become exposed after CD4 binding, and they are termed as CD4 induced epitopes (CD4i). Several antibodies target this site like 17b, E51, or m36, but again they only exhibit mild neutralization potency[223,224]. The treatment with these antibodies in clinical trials usually is effective for a short period of time, after which resistant strains appear and are positively selected[225]. The successful viremia suppression comes when the antiretroviral therapy combines several neutralizing antibodies[225].

The problem about targeting the viral spikes is not only the glycan masking of the epitopes but also the high mutation rate. This mutability generates virus quasispecies varying up to 5% in their genetic content inside the same host[226]. A reasonable approach is to interfere with the receptors used in viral attachment, like CD4 and the chemokine receptor[151]. Being host proteins they are not exposed to a high mutation rate thus they can be targeted with the certainty of not becoming obsolete due to sequence variability[227]. Some problems may arise from this strategy since the binding of antibodies to these two receptors could alter their natural function. This requires inhibitory molecules able to prevent HIV infection but to not interfere with the performance of these receptors[228].

Chapter 7

The effect of forces on proteins has not been approached until recently and, as it was indicated in the Introduction chapter, their relevance in human diseases cannot be underestimated. In the same way that bacteria attachment is strongly influenced by forces, the viral attachment process also experiences mechanical stress. During HIV entry, CD4 conformational flexibility is required for viral particle approaching to the membrane coreceptor[135,229]. This malleability implies structural changes on CD4 which could be triggered due to the forces exerted by the bound virion. A stiffer CD4 molecule would complicate downstream events required for HIV entry. Understanding the mechanics of CD4 under small stretching forces which resemble the viral attachment *in vivo*, constitutes the first step for designing a strategy to stiffen CD4 and block HIV infection.

In this thesis the mechanical behavior of the first two domains of CD4 under different conditions has been investigated. For this purpose we studied CD4D1D2 protein with smFS. In the first set of experiments the mechanical stability of CD4D1D2 was studied with constant velocity experiments. Two pulling speeds were tested, 400 $nm \cdot s^{-1}$ and 10 $nm \cdot s^{-1}$ (**Fig. 6.16**). The first speed has been widely used for determining the mechanical stability of proteins like I91 and constitutes a reasonable starting condition for identifying the protein mechanical stability[230], but the second lower speed reflects the attempt of resembling the hypothetical moving speed of a viral particle attached to the cell surface *in vivo*. Under both pulling speeds both CD4 domains exhibited a lower mechanical stability than the I91 domains of the polyprotein, and D2 showed a higher unfolding force than D1. Counterintuitively the strongest domain, D2, unfolds before the weaker, D1, when the expected behavior is just the opposite[31]. This trend is

observed at low and high pulling speed indicating a mechanical hierarchy in the behavior of CD4D1D2 protein. D1 and D2 are connected through a β-strand shared by both domains (termed as G-strand in D1, and as A-strand in D2). This shared structural element would confer to the D1-D2 tandem a common mechanical response, acting as one unit. After D2 unfolding D1is left unprotected, easing its mechanical unfolding. Once D2 is unfolded the force would be propagated through the shared strand between domains, breaking the hydrogen bonds made between D1 A-strand and G-strand (the strand shared with D2). This intimate interdomain connection has been suggested as a proof of modularity in CD4, since the same kind of relationship exists also between D3 and D4 domains[231].

In the set of FC experiments different constant forces where tested to unravel the kinetics of CD4D1D2 unfolding, from 20 to 100 pN (**Fig. 6.17**). Following Bell-Evans-Ritchie model[67,232], the unfolding rate of both domains increased exponentially with the force applied (**Fig. 6.17c**). The force range used allows for the extrapolation of the unfolding rate at force-values close to zero, in the range of forces expected to be exerted by a viral particle attached to CD4. These measurements demonstrated that CD4D1D2 extension is possible at very low forces, like the ones HIV could exert *in vivo*. This extra length would increase the chances for coreceptor binding on the cell surface, since a longer tether permits the exploration of a larger area surrounding the anchoring point provided by CD4. Again the hierarchical unfolding of the tandem was reproduced as it occurred in the FX measurements. Moreover D2 was observed to be able to unfold through an intermediate, a feature that could be related with the redox regulation of its allosteric disulfide bond. This intermediate could

expose the buried disulfide making it available for reduction by thioredoxin enzyme, enabling a further extension of the tether.

Longer tethers support greater HIV infectivity, as it can be deduced from the experiments of Freeman *et al.*[152]. In this work four constructs where expressed on the surface of 293T cells, presenting all of them the tandem CD4D1D2 followed by polypeptide linkers of different length. We developed a mathematical expression that relates the infectivity data from this work with the length of the different CD4D1D2 constructs. This expression accounts for the force-induced extension of CD4 under different loads. The infectivity levels of the four constructs were correlated with the length of the tether and this trend follows a single-exponential function (**Fig. 6.18** and **Fig. 6.19**). Wild-type CD4 provides the maximum infectivity level but it falls out from the exponential fit if we assume this infectivity is achieved without the extension of any of its domains. Nevertheless if the unfolding of one or two domains is considered, the extra extension would place wild-type CD4 in the exponential trend. This suggests the unfolding of CD4 D3 and D4 would help HIV to infect the cells, and the unfolding of D2 also cannot be ruled out from this behavior.

From these findings it is feasible that HIV virion could trigger the partial unfolding at low forces of CD4 supporting the infection. Several antibodies interfere with the infection, and those ones recognizing CD4 are of special importance since they provide an opportunity for targeting a host element instead of the highly mutable epitopes of the virus[134,150,151]. It has been hypothesized that some of these antibodies that do not compete with the binding site of gp120, could interfere with the conformational rearrangements required for

viral entry. Among these antibodies is Ibalizumab, a strong inhibitor of HIV infection but whose neutralizing mechanism is unknown[153]. It has been demonstrated that protein-protein interactions can lead to mechanical stabilization[41,42], and Ibalizumab binding to CD4 could mechanically stabilize CD4 hindering its force-induced extension. For this reason we performed smFS experiments on CD4D1D2 in the presence of Ibalizumab. First, FX experiments did not show remarkable differences in the unfolding force of the tandem CD4D1D2, only with a slight increase in the case of D2 domain. Ibalizumab contacts are placed mainly in D2 domain and some in D1 domain, and a greater stabilization effect was expected. However although forces did not increase significantly, the FX traces showed a change in the position of the CD4D1D2 unfolding peaks suggesting the stabilization of the tandem. In the absence of Ibalizumab both CD4 domains unfold first followed by the unfolding peaks of the fingerprint I91 domains. When Ibalizumab is present the tandem unfolding was detected even after the unfolding of the four I91 domains (**Fig. 6.20d**). A possible explanation for this behavior could be that Ibalizumab stabilizes mainly D2 domain, which protects D1 domain from mechanical unfolding, but under increasing tensile stress the antibody unbinding occurs and leaves D2 domain susceptible of mechanical unfolding. FR experiments provided a better understanding of this behavior since unlike FX they make possible the application of constant loading rates. FR traces in the presence of Ibalizumab showed again a change in the position of the unfolding steps of CD4D1D2, occurring after the I91 unfolding (**Fig. 6.21b**). If the initial unfolding forces (the first unfolding step coming from any of the domains of CD4D1D2) are compared it is possible to see a shift to higher forces when the antibody is present. In this case the

average unfolding force increased around 80 pN (**Fig. 6.21e**), four times the values detected in FX experiments. We also detected this stabilization effect in the length increase of CD4D1D2, whose simultaneous unfolding was more probable in the presence of Ibalizumab (**Fig. 6.21d**). This could be explained in the context of the force range where the unfolding of the tandem takes place. Ibalizumab delays the unfolding of D2 domain meanwhile the tensile force increases, but when it unbinds it leaves the tandem submitted to forces way higher than the ones the CD4D1D2 tandem is able to resist by itself. At this high force the unfolding of both domains occurs in one single step that reveals that Ibalizumab binding mechanically stabilizes CD4. During HIV entry Ibalizumab could prevent CD4 unfolding decreasing the success of coreceptor binding because of the rigidity of the linker.

These results suggested us that CD4 extension and flexibility play a significant role for the HIV successful infection. Another important feature of HIV infection is the redox regulation of the proteins involved[147]. CD4 domains possess intramolecular disulfide bonds whose presence imposes mechanical inextensibility beyond a certain point. In the force range we can apply with smFS these bonds are unbreakable due to their covalent nature. Besides, the virus could never pull at such high forces without breaking the weak contacts made with D1 domain. It is known that several oxidoreductases are secreted to the extracellular space for signaling and also for the regulation of membrane and extracellular proteins, and it has been demonstrated that the reduction of disulfide bonds present in gp120 and in CD4 are required for viral entry[148,233]. Specifically the disulfide bond reduction of CD4D2 domain is apparently required for the successful infection[146].

This disulfide presents some particularities that make it different from the disulfide bonds of domains 1 and 4. For example in D1 the disulfide bond connects the B and F-strands, bridging the two β-sheets that form the β-sandwich structure of the domain. This is the typical architecture of Ig-like domains which present disulfide bonds. However in D2 the disulfide bond bridges the C and F-strands and both are part of the same β-sheet. This kind of disulfides are unusual and are related with the allosteric regulation of the protein[231]. In the case of CD4 the reduction of the disulfide bond of D2 domain would allow the dimerization of CD4 molecules through the formation of intermolecular disulfides between D2 domains of neighboring receptors[234]. Dimers of CD4 would help to stabilize the interaction with the MHC-II during antigen presentation and also would amplify the downstream signals that lead to T-cell activation.

During viral attachment thioredoxin oxidoreductases secreted to the extracellular space could reduce D2 disulfide, making possible the total extension of this domain. Despite of being different from D1 and D4 disulfides, as in them D2 disulfide is buried in the hydrophobic core of the domain. As it is hidden, oxidoreductases cannot access the disulfide thus cannot reduce it. Nevertheless the viral particle bound to CD4 could unfold partially D2 domain making its disulfide solvent-exposed and available for thioredoxin catalytic activity. We hypothesized that CD4 viral-induced extension could be mechanochemically enhanced by cell enzymes. In order to test this hypothesis we conducted FC experiments in the presence of thioredoxin. FC measurements are best suited for mechanochemical experiments since oxidoreductase activity is impaired when the substrate is held at high forces[194], thus maintaining a constant

intermediate force guarantees the optimal catalytic performance of the enzyme. The data collected shows the unfolding of both domains up to its disulfide bonds and its posterior reduction as a consequence of the action of thioredoxin (**Fig. 6.22**). In very rare events was possible to unfold completely the domains showing no disulfide. Along the experiment thioredoxin could reduce the disulfide bonds of CD4D1D2 molecules before stretching, but the low number of events registered suggests that their disulfides are inaccessible for the enzyme and only after mechanical extension of the protein they become available. In some occasions after the initial unfolding we detected the reduction of D2 domain disulfide but not the one from D1, suggesting that this disulfide is more prone to be regulated by thioredoxin.

D2 domain disulfide bond shares the same configuration as the one of gp120 V3-loop, with two antiparallel β-strands connected[235]. This configuration, termed as –RHStaple, is typical of allosteric disulfides and provides the adequate positioning of the three sulfur atoms involved in the reaction, forming a 180º angle between thioredoxin Cys^{32} and CD4D2 Cys^{130} and Cys^{159}, which favors reduction[143]. Moreover the –RHStaple configuration imposes an unfavorable conformation of the χ angles of the disulfide, generating a mechanical strain in the protein structure which destabilizes the bond. These topological constraints modulate the disulfide reactivity and stretching forces facilitate their reduction[236]. Therefore external applied forces (like the exerted by the virus) induce a conformational distortion which affects the redox potential of the disulfide, lowering its activation energy[237,238]. Once the disulfide is reduced, the prestress topology of the protein could be released triggering conformational changes in both gp120 and CD4, which are required in the subsequent

events of infection[143,235]. This scenario strongly supports the mechanochemical regulation of CD4, where the virus could exert force enough for unfolding CD4 domains, destabilizing the oxidized state of the disulfide and easing the access of thioredoxin. Again the extra lengthening provided by a reduced and an unfolded CD4 domain would increase the chances of the virus for binding to its coreceptor, covering a larger area of the cell surface.

We propose that HIV takes advantage from CD4 conformational malleability in order to commit a successful infection. Once bound, the virion remains attached to CD4 through weak interactions but the random movements of the particle can exert forces on CD4 high enough for inducing its mechanical extension. The unfolding of some of CD4 domains exposes buried disulfide bonds which are susceptible of redox regulation by thioredoxin, an oxidoreductase secreted to the extracellular space as part of the normal cell physiological processes. CD4 D2 domain is highlighted as a potential target for this redox regulation since its disulfide bond presents features proper of allosterism, facilitating the dimerization of CD4 molecules on the cell surface. This would allow HIV to extend CD4 even more, covering larger areas of the cell surface and increasing the probability of coreceptor encounter. CD4 dimerization could also help HIV binding, securing the attachment of one Env spike with two gp120/CD4 interactions. However we think the mechanochemical extension of CD4 during HIV binding contributes more to the infection than CD4 dimerization. Mechanically extended proteins are able to refold in the presence of low forces, and we think that viral Brownian motions can both extend and collapse CD4, hopping around the folding equilibrium of the domains of the receptor.

Taking into account all of these features we propose CD4 behaves as a shock-absorber element during HIV binding, a structure able to extend and contract at low forces due to the motion of the virus. The ability of releasing the tension experienced by the contacts of the gp120/CD4 interface would prevent the viral detachment from the receptor, as it happens with the unraveling mechanism of the type I pilus rod. Together with the hinge-like flexibility of CD4 and its mechanochemical extension, the opportunities for coreceptor binding of HIV increase. Our results using Ibalizumab show that the mechanical stabilization of CD4 could be the reason of the inhibitory effect of this antibody. This finding supports our hypothesis of the flexible spring-like structure hijacked for infection, and also highlights the mechanomedical research as a field full of possibilities for the future of Medicine.

7.3 Conclusions and outlook

In the presented work we have approached the mechanochemical aspects involved during UPEC and HIV-1 attachment to their target cells on the single molecule level. Our work strongly highlights the importance of forces on processes occurring at the nanoscale like bacterial and viral attachment, where single molecules and protein-protein/ligand interactions govern the fate of these two pathogenic infections.

We think these findings are exportable to other models of pathogenic infection and that the nanomechanical properties of the

molecules involved in the attachment should be explored and considered for medical treatment development.

In the case of UPEC smFS has been proven to be useful for probing the mechanical stability of pilus proteins, and to resemble the extended state of pilus subunits for interrogating the maturation process in the periplasm. As it can be seen from our results, the redox state of pilus subunits disulfide bonds can be targeted for mechanical destabilization. Also, designed small peptides or molecules able to compete/disrupt for the donor-strand complementation in the mature pilus could be used for triggering pilus disassembly. This last strategy should be oriented to target specially the donor-strand interactions from the pilus tip, since the quaternary structure of the pilus rod would shield the access to FimA-FimA linkage unless it is mechanically uncoiled. Besides targeting the mature pilus, the main actors involved during pilus proteins periplasmic maturation and incorporation offer alternative approaches. DsbA defective mutants display problems for type I and P pilus maturation[105]. DsbA/DsbB system is widely spread across species, and targeting it affects several functions of the bacteria, many of them related with the assembly and secretion of virulence factors such as Tcp pili, cholera toxin and enterotoxin[239].

Other potential targets for drug development could be the blockage of FimD secretion pore, which would inhibit pilus secretion. This target has some advantages since it is located in the OM of bacteria and it is easier to access. The blockage of SecYEG channel for avoiding periplasmic importation of proteins, or disrupting the favorable molecular surfaces provided by the periplasmic domains of

FimD, responsible of chaperone/subunit complex management during pilus assembly, could be more alternatives.

SmFS can be helpful in the developing and testing the effect of new drugs, as it was done with the CS20 fimbriae. In this research the mechanical effect of antibody binding to these adhesive structures of enterotoxigenic *E. coli* was tested with smFS experiments, showing that antibody binding decreased pilus elasticity hindering uncoiling. Higher forces were required to uncoil these pili because of the stabilization effect of the antibody, and also extended pilus recoiling was hindered due to this protein-protein interaction[240]. Peptides or molecules with the ability of disrupting pilus proteins mechanical stability could be also approached with smFS, and experiments on whole purified pilus should be conducted in order to test their effect on higher order structures.

Regarding CD4 and HIV infection we observed strong parallelisms with the UPEC attachment. The redox regulation of CD4D2 disulfide bond by thioredoxin, as DsbA does on pilus proteins, could be a potential target for viral infection prevention. However other oxidoreductases like PDI or glutaredoxin could substitute thioredoxin and also interfering with them probably would alter many cellular processes. In the case of CD4, the mechanical stabilization of its domains through antibody binding could be the better approach. Ibalizumab has shown us that its neutralizing activity could be explained in the context of CD4 stiffening due to antibody binding.

In the future the four extracellular domains of CD4 should be tested in order to obtain a complete description of the nanomechanics of this receptor. The glycosylation of D3 and D4 domains prevented us to produce this protein in our bacterial expression system, but as we

proved the first two domains of CD4 already gave us an interesting picture of how forces could be affecting them during infection. Future work involving smFS would require measurements of CD4 mechanical extensibility at forces close to the range of thermal motions, like for example using OT or MT setups. An interesting experiment would be to functionalize MT probes with gp120 and pull from full extracellular CD4 molecules attached to the surface. After gp120/CD4 binding, pulling at very low forces would reveal us the kinetics of CD4 domains unfolding when pulled from the specific binding site of HIV, resembling the *in vivo* interaction. These experiments should test also the effect of thioredoxin and Ibalizumab on this system, and see how these two proteins separated and in combination affect CD4 mechanical unfolding.

Our findings provide a framework for mechanomedical research. Research focused towards to find molecules able to bind to CD4, increasing its stiffness, and preventing its mechanical extension, could be the therapeutic strategy of the future to avoid HIV entry in cells. The same situation is applicable for UPEC pilus attachment where instead of protein stiffening, the mechanical weakening of its individual domains could be the best approach. The Mechanopharmacology field may develop as the future strategy for the designing of fine-tuned drugs able to mechanically affect proteins involved in human diseases.

Bibliography

1. http://www.who.int/mediacentre/factsheets/fs310/en/.

2. Lim, C. *et al.* Epidemiology and burden of multidrug-resistant bacterial infection in a developing country. *Elife* **5,** 1–18 (2016).

3. Tandogdu, Z. & Wagenlehner, F. M. E. Global epidemiology of urinary tract infections. *Curr Opin Infect Dis* **29,** 73–79 (2016).

4. McLellan, L. K. & Hunstad, D. A. Urinary Tract Infection: Pathogenesis and Outlook. *Trends Mol. Med.* **22,** 946–957 (2016).

5. O 'Neill, J. Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. The Review on Antimicrobial Resistance. (2016). doi:10.1016/j.jpha.2015.11.005

6. World Health Organization – Fact sheet – Influenza, November 2016 http://www.who.int/mediacentre/factsheets/fs211/en/.

7. Keusch, G., Mcadam, K. & Cuff, P. A. *Integrating Clinical Research into Epidemic Response*. (National Academies Press, 2017). doi:10.17226/24739

8. UNAIDS - Fact sheet – Latest statistics on the status of the AIDS epidemic, July 2017, http://www.unaids.org/en/resources/fact-sheet.

9. del Rio, C. The global HIV epidemic: What the pathologist needs to know. *Semin. Diagn. Pathol.* **34,** 314–317 (2017).

10. World Health Organization – Fact sheet HIV/AIDS, July 2017, http://www.who.int/mediacentre/factsheets/fs360/en/.

11. Maartens, G., Celum, C. & Lewin, S. R. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet* **384,** 258–271 (2014).

12. http://gamapserver.who.int/mapLibrary/Files/Maps/HIV_deaths_2016.png.

13. Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nat. Rev. Microbiol.* **8,** 423–435 (2010).

14. Darcis, G., Van Driessche, B. & Van Lint, C. HIV Latency: Should We Shock or Lock? *Trends Immunol.* **38,** 217–228 (2017).

15. Bhella, D. The role of cellular adhesion molecules in virus attachment and entry. *Phil. Trans. R. Soc. B* **370,** 20140035 (2015).

16. Ribet, D. & Cossart, P. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect.* **17,** 173–183 (2015).

17. Jansen, K. A. *et al.* A guide to mechanobiology: Where biology and physics meet. *Biochim. Biophys. Acta - Mol. Cell Res.* **1853,** 3043–3052 (2015).

18. Chen, W. & Zhu, C. Mechanical regulation of T-cell functions.

Immunol. Rev. **256,** 160–176 (2013).

19. Hu, X., Margadant, F. M., Yao, M. & Sheetz, M. P. Molecular stretching modulates mechanosensing pathways. *Protein Sci.* **26,** 1337–1351 (2017).

20. Ladoux, B., Nelson, W. J., Yan, J. & Mège, R. M. The mechanotransduction machinery at work at adherens junctions. *Integr. Biol.* 1109–1119 (2015). doi:10.1039/C5IB00070J

21. Eyckmans, J., Boudou, T., Yu, X. & Chen, C. S. A Hitchhiker's Guide to Mechanobiology. *Dev. Cell* **21,** 35–47 (2011).

22. del Rio, A. *et al.* Stretching single talin rod molecules activates vinculin binding. *Science (80-.).* **323,** 638–641 (2009).

23. Haining, A. W. M., Lieberthal, T. J. & Del R??o Hern??ndez, A. Talin: A mechanosensitive molecule in health and disease. *FASEB J.* **30,** 2073–2085 (2016).

24. Mortimer, G. M. & Minchin, R. F. Cryptic epitopes and functional diversity in extracellular proteins. *Int. J. Biochem. Cell Biol.* **81,** 112–120 (2016).

25. Matouschek, A. Protein unfolding - An important process in vivo? *Curr. Opin. Struct. Biol.* **13,** 98–109 (2003).

26. Hirata, H., Tatsumi, H., Lim, C. T. & Sokabe, M. Force-dependent vinculin binding to talin in live cells: a crucial step in anchoring the actin cytoskeleton to focal adhesions. *AJP Cell Physiol.* **306,** C607–C620 (2014).

27. Garcia, T. I., Oberhauser, A. F. & Braun, W. Mechanical

stability and differentially conserved physical-chemical properties of titin Ig-domains. *Proteins Struct. Funct. Bioinforma.* **75,** 706–718 (2009).

28. Kawakami, M., Byrne, K., Brockwell, D. J., Radford, S. E. & Smith, D. A. Viscoelastic study of the mechanical unfolding of a protein by AFM. *Biophys. J.* **91,** L16–L18 (2006).

29. Li, H. & Fernandez, J. M. Mechanical design of the first proximal Ig domain of human cardiac titin revealed by single molecule force spectroscopy. *J. Mol. Biol.* **334,** 75–86 (2003).

30. Best, R. B. *et al.* Mechanical unfolding of a titin Ig domain: Structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. *J. Mol. Biol.* **330,** 867–877 (2003).

31. Rief, M. Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science (80-.).* **276,** 1109–1112 (1997).

32. Chen, H. *et al.* Dynamics of Equilibrium Folding and Unfolding Transitions of Titin Immunoglobulin Domain under Constant Forces. *J. Am. Chem. Soc.* **137,** (2015).

33. Tskhovrebova, L., Trinick, J., Sleep, J. A. & Simmons, R. M. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387,** 308–312 (1997).

34. Kellermayer, M. S., Smith, S. B., Granzier, H. L. & Bustamante, C. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science (80-.).* **276,** 1112–

1116 (1997).

35. Rief, M., Gautel, M., Schemmel, A. & Gaub, H. E. The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by atomic force microscopy. *Biophys. J.* **75,** 3008–14 (1998).

36. Alegre-Cebollada, J. *et al.* S-glutathionylation of cryptic cysteines enhances titin elasticity by blocking protein folding. *Cell* **156,** 1235–1246 (2014).

37. Grützner, A. *et al.* Modulation of titin-based stiffness by disulfide bonding in the cardiac titin N2-B unique sequence. *Biophys. J.* **97,** 825–834 (2009).

38. Rivas-Pardo, J. A. A. *et al.* Work Done by Titin Protein Folding Assists Muscle Contraction. *Cell Rep.* **14,** 1339–1347 (2016).

39. Mártonfalvi, Z., Bianco, P., Naftz, K., Ferenczy, G. G. & Kellermayer, M. Force generation by titin folding. *Protein Sci.* **26,** 1380–1390 (2017).

40. Ingber, D. Mechanobiology and diseases of mechanotransduction. *Ann. Med.* **35,** 564–577 (2003).

41. Cao, Y., Yoo, T., Zhuang, S. & Li, H. Protein-Protein Interaction Regulates Proteins' Mechanical Stability. *J. Mol. Biol.* **378,** 1132–1141 (2008).

42. Cao, Y., Li, Y. D. & Li, H. Enhancing the mechanical stability of proteins through a cocktail approach. *Biophys. J.* **100,** 1794–1799 (2011).

43. Snider, N. T. & Omary, M. B. Post-translational modifications of intermediate filament proteins: mechanisms and functions. **15,** 163–177 (2014).

44. Brower-Toland, B. *et al.* Specific contributions of histone tails and their acetylation to the mechanical stability of nucleosomes. *J. Mol. Biol.* **346,** 135–146 (2005).

45. Guglietta, A. *et al.* Recurrent urinary tract infections in women: risk factors, etiology, pathogenesis and prophylaxis. *Future Microbiol.* **12,** 239–246 (2017).

46. Kaper, J. B., Nataro, J. P. & Mobley, H. L. T. Pathogenic Escherichia coli. *Nat Rev Microbiol* **2,** 123–140 (2004).

47. Cross, A. S. What is a virulence factor? *Crit. Care* **12,** 197 (2008).

48. Wu, J., Miao, Y. & Abraham, S. N. The multiple antibacterial activities of the bladder epithelium. *Ann. Transl. Med.* **5,** 35 (2017).

49. Sheerin, N. S. Urinary tract infection. *Medicine (Baltimore).* **39,** 384–389 (2011).

50. Foxman, B. The epidemiology of urinary tract infection. *Nat Rev Urol* **7,** 653–660 (2010).

51. Persat, A. *et al.* The mechanical world of bacteria. *Cell* **161,** 988–997 (2015).

52. Waksman, G. Structural and Molecular Biology of a protein-polymerizing nanomachine for pilus biogenesis. *J. Mol. Biol.*

429, 2654–2666 (2017).

53. Costa, T. R. D. *et al.* Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **13,** 343–359 (2015).

54. Mulvey, M. A., Schilling, J. D., Martinez, J. J. & Hultgren, S. J. Bad bugs and beleaguered bladders: Interplay between uropathogenic Escherichia coli and innate host defenses. *Proc. Natl. Acad. Sci.* **97,** 8829–8835 (2000).

55. Hahn, E. *et al.* Exploring the 3D molecular architecture of Escherichia coli type 1 pili. *J. Mol. Biol.* **323,** 845–857 (2002).

56. Lillington, J., Geibel, S. & Waksman, G. Reprint of 'Biogenesis and adhesion of type 1 and P pili'. *Biochim. Biophys. Acta - Gen. Subj.* **1850,** 554–564 (2015).

57. Hospenthal, M. K., Costa, T. R. D. & Waksman, G. A comprehensive guide to pilus biogenesis in Gram-negative bacteria. *Nat. Rev. Microbiol.* **15,** 365–379 (2017).

58. Lewis, A. J., Richards, A. C., Mulvey, M. A. & City, S. L. Invasion of Host Cells and Tissues by Uropathogenic Bacteria. *Microbiol. Spectr.* **4,** 1–18 (2016).

59. Zhou, G. *et al.* Uroplakin Ia is the urothelial receptor for uropathogenic Escherichia coli: evidence from in vitro FimH binding. *J. Cell Sci.* **114,** 4095–103 (2001).

60. Ozer, A. *et al.* Advanced glycation end products facilitate bacterial adherence in urinary tract infection in diabetic mice.

Pathog. Dis. **73,** ftu004-ftu004 (2015).

61. Kalas, V. *et al.* Evolutionary fine-tuning of conformational ensembles in FimH during host-pathogen interactions. *Sci. Adv.* **3,** e1601944 (2017).

62. Rodriguez, V. B. *et al.* Allosteric coupling in the bacterial adhesive protein FimH. *J. Biol. Chem.* **288,** 24128–24139 (2013).

63. Anderson, B. N. *et al.* Weak Rolling Adhesion Enhances Bacterial Surface Colonization. *J. Bacteriol.* **189,** 1794–1802 (2007).

64. Le Trong, I. *et al.* Structural Basis for Mechanical Force Regulation of the Adhesin FimH via Finger Trap-like β Sheet Twisting. *Cell* **141,** 645–655 (2010).

65. Thomas, W. E., Nilsson, L. M., Forero, M., Sokurenko, E. V. & Vogel, V. Shear-dependent 'stick-and-roll' adhesion of type 1 fimbriated Escherichia coli. *Mol. Microbiol.* **53,** 1545–1557 (2004).

66. Lao, B. B. *et al.* Rational design of topographical helix mimics as potent inhibitors of protein-protein interactions. *J. Am. Chem. Soc.* **136,** 7877–7888 (2014).

67. Bell, G. Models for the specific adhesion of cells to cells. *Science (80-.).* **200,** 618–627 (1978).

68. Thomas, W. E., Vogel, V. & Sokurenko, E. Biophysics of Catch Bonds. *Annu. Rev. Biophys.* **37,** 399–416 (2008).

69. Hertig, S. & Vogel, V. Catch bonds. *Curr. Biol.* **22,** R823–R825 (2012).

70. Sauer, M. M. *et al.* Catch-bond mechanism of the bacterial adhesin FimH. *Nat. Commun.* **7,** 10738 (2016).

71. Aprikian, P. *et al.* The bacterial fimbrial tip acts as a mechanical force sensor. *PLoS Biol.* **9,** e1000617 (2011).

72. Sokurenko, E. V., Vogel, V. & Thomas, W. E. Catch-Bond Mechanism of Force-Enhanced Adhesion: Counterintuitive, Elusive, but ... Widespread? *Cell Host Microbe* **4,** 314–323 (2008).

73. Miller, E., Garcia, T., Hultgren, S. & Oberhauser, A. F. The Mechanical Properties of E. coli Type 1 Pili Measured by Atomic Force Microscopy Techniques. *Biophys. J.* **91,** 3848–3856 (2006).

74. Andersson, M., Axner, O., Almqvist, F., Uhlin, B. E. & Fällman, E. Physical Properties of Biopolymers Assessed by Optical Tweezers: Analysis of Folding and Refolding of Bacterial Pili. *ChemPhysChem* **9,** 221–235 (2008).

75. Andersson, M., Uhlin, B. E. & Fällman, E. The Biomechanical Properties of E. coli Pili for Urinary Tract Attachment Reflect the Host Environment. *Biophys. J.* **93,** 3008–3014 (2007).

76. Rangel, D. E., Marín-Medina, N., Castro, J. E., González-Mancera, A. & Forero-Shelton, M. Observation of Bacterial Type I Pili Extension and Contraction under Fluid Flow. *PLoS One* **8,** e65563 (2013).

77. Zakrisson, J., Wiklund, K., Axner, O. & Andersson, M. Helix-like biopolymers can act as dampers of force for bacteria in flows. *Eur. Biophys. J.* **41,** 551–560 (2012).

78. Tsirigotaki, A., De Geyter, J., Šoštaric´, N., Economou, A. & Karamanou, S. Protein export through the bacterial Sec pathway. *Nat. Rev. Microbiol.* **15,** 21–36 (2016).

79. Piątek, R., Bruździak, P., Wojciechowski, M., Zalewska-Piątek, B. & Kur, J. The noncanonical disulfide bond as the important stabilizing element of the immunoglobulin fold of the Dr fimbrial DraE subunit. *Biochemistry* **49,** 1460–1468 (2010).

80. Hatahet, F., Boyd, D. & Beckwith, J. Disulfide bond formation in prokaryotes: history, diversity and design. *Biochim. Biophys. Acta* **1844,** 1402–14 (2014).

81. Inaba, K. *et al.* Crystal Structure of the DsbB-DsbA Complex Reveals a Mechanism of Disulfide Bond Generation. *Cell* **127,** 789–801 (2006).

82. Inaba, K. & Ito, K. Structure and mechanisms of the DsbB-DsbA disulfide bond generation machine. *Biochim. Biophys. Acta - Mol. Cell Res.* **1783,** 520–529 (2008).

83. Grauschopf, U. *et al.* Why is DsbA such an oxidizing disulfide catalyst? *Cell* **83,** 947–955 (1995).

84. Choudhury, D. *et al.* X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic Escherichia coli. *Science* **285,** 1061–6 (1999).

85. Sarowar, S., Hu, O. J., Werneburg, G. T., Thanassi, D. G. & Li, H. The Escherichia coli P and type 1 pilus assembly chaperones PapD and FimC are monomeric in solution. *J. Bacteriol.* **198,** 2360–2369 (2016).

86. Sauer, F. G. Structural Basis of Chaperone Function and Pilus Biogenesis. *Science (80-.).* **285,** 1058–1061 (1999).

87. Hermanns, U., Sebbel, P., Eggli, V. & Glockshuber, R. Characterization of FimC, a periplasmic assembly factor for biogenesis of type 1 pili in Escherichia coli. *Biochemistry* **39,** 11564–70 (2000).

88. Pellecchia, M., Guntert, P., Glockshuber, R. & Wuthrich, K. NMR solution structure of the periplasmic chaperone FimC. *Nat Struct Biol* **5,** 885–890 (1998).

89. Eidam, O., Dworkowski, F. S. N., Glockshuber, R., Grütter, M. G. & Capitani, G. Crystal structure of the ternary FimC-FimF(t)-FimD(N) complex indicates conserved pilus chaperone-subunit complex recognition by the usher FimD. *FEBS Lett.* **582,** 651–5 (2008).

90. Geibel, S., Procko, E., Hultgren, S. J., Baker, D. & Waksman, G. Structural and energetic basis of folded-protein transport by the FimD usher. *Nature* **496,** 243–246 (2013).

91. Nishiyama, M., Vetsch, M., Puorger, C., Jelesarov, I. & Glockshuber, R. Identification and Characterization of the Chaperone-Subunit Complex-binding Domain from the Type 1 Pilus Assembly Platform FimD. *J. Mol. Biol.* **330,** 513–525

(2003).

92. Saulino, E. T. Ramifications of kinetic partitioning on usher-mediated pilus biogenesis. *EMBO J.* **17,** 2177–2185 (1998).

93. Remaut, H. *et al.* Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol. Cell* **22,** 831–842 (2006).

94. Nishiyama, M., Ishikawa, T., Rechsteiner, H. & Glockshuber, R. Reconstitution of pilus assembly reveals a bacterial outer membrane catalyst. *Science* **320,** 376–9 (2008).

95. Hospenthal, M. K. *et al.* Structure of a Chaperone-Usher Pilus Reveals the Molecular Basis of Rod Uncoiling. *Cell* **164,** 269–278 (2016).

96. Puorger, C. *et al.* Infinite Kinetic Stability against Dissociation of Supramolecular Protein Complexes through Donor Strand Complementation. *Structure* **16,** 631–642 (2008).

97. Puorger, C., Vetsch, M., Wider, G. & Glockshuber, R. Structure, folding and stability of FimA, the main structural subunit of type 1 Pili from uropathogenic escherichia coli strains. *J. Mol. Biol.* **412,** 520–535 (2011).

98. Werneburg, G. T. *et al.* The pilus usher controls protein interactions via domain masking and is functional as an oligomer. *Nat. Struct. & Mol. Biol.* **22,** 1–9 (2015).

99. Allen, W. J., Phan, G. & Waksman, G. Pilus biogenesis at the outer membrane of Gram-negative bacterial pathogens. *Curr.*

Opin. Struct. Biol. **22,** 500–506 (2012).

100. Nishiyama, M. & Glockshuber, R. The Outer Membrane Usher Guarantees the Formation of Functional Pili by Selectively Catalyzing Donor-Strand Exchange between Subunits That Are Adjacent in the Mature Pilus. *J. Mol. Biol.* **396,** 1–8 (2010).

101. Allen, W. J., Phan, G., Hultgren, S. J. & Waksman, G. Dissection of Pilus Tip Assembly by the FimD Usher Monomer. *J. Mol. Biol.* **425,** 958–967 (2013).

102. Phan, G. *et al.* Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. *Nature* **474,** 49–53 (2011).

103. Vetsch, M. *et al.* Mechanism of fibre assembly through the chaperone-usher pathway. *EMBO Rep.* **7,** 734–738 (2006).

104. Geibel, S. & Waksman, G. The molecular dissection of the chaperone-usher pathway. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843,** 1559–1567 (2014).

105. Jacob-Dubuisson, F. *et al.* PapD chaperone function in pilus biogenesis depends on oxidant and chaperone-like activities of DsbA. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 11552–6 (1994).

106. Greene, W. C. A history of AIDS: Looking back to see ahead. *Eur. J. Immunol.* **37,** S94–S102 (2007).

107. Barre-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (80-.).* **220,** 868–871 (1983).

108. Gallo, R. C. *et al.* Frequent detection and isolation of cytopathic

retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224,** 500–3 (1984).

109. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **224,** 497–500 (1984).

110. Schüpbach, J. *et al.* Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science* **224,** 503–5 (1984).

111. Levy, J. *et al.* Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science (80-.).* **225,** 840–842 (1984).

112. Coffin, J. *et al.* What to call the AIDS virus? *Nature* **321,** 10 (1986).

113. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS Pandemic. *Cold Spring Harb. Perspect. Med.* **1,** a006841–a006841 (2011).

114. Sharp, P. M. *et al.* The origins of acquired immune deficiency syndrome viruses: where and when? *Philos. Trans. R. Soc. B Biol. Sci.* **356,** 867–876 (2001).

115. Perilla, J. R. & Gronenborn, A. M. Molecular Architecture of the Retroviral Capsid. *Trends Biochem. Sci.* **41,** 410–420 (2016).

116. Law, K. M., Satija, N., Esposito, A. M. & Chen, B. K. Cell-to-Cell Spread of HIV and Viral Pathogenesis. in *Advances in Virus*

Research **95,** 43–85 (Elsevier Inc., 2016).

117. Hübner, W. *et al.* Quantitative 3D video microscopy of HIV transfer across T cell virological synapses. *Science* **323,** 1743–7 (2009).

118. Wilen, C. B., Tilton, J. C. & Doms, R. W. HIV: Cell Binding and Entry. *Cold Spring Harb. Perspect. Med.* **2,** a006866–a006866 (2012).

119. Wilen, C. B., Tilton, J. C. & Doms, R. W. *Viral Molecular Machines*. **726,** (Springer US, 2012).

120. Checkley, M. A., Luttge, B. G. & Freed, E. O. HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. *J. Mol. Biol.* **410,** 582–608 (2011).

121. HARRIS, R. J., CHAMOW, S. M., GREGORY, T. J. & SPELLMAN, M. W. Characterization of a soluble form of human CD4. Peptide analyses confirm the expected amino acid sequence, identify glycosylation sites and demonstrate the presence of three disulfide bonds. *Eur. J. Biochem.* **188,** 291–300 (1990).

122. Zamoyska, R. CD4 and CD8: modulators of T-cell receptor recognition of antigen and of immune responses? *Curr. Opin. Immunol.* **10,** 82–87 (1998).

123. Yin, Y., Wang, X. X. & Mariuzza, R. A. Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 5405–10 (2012).

124. Haqqani, A. A. & Tilton, J. C. Entry inhibitors and their use in the treatment of HIV-1 infection. *Antiviral Res.* **98,** 158–170 (2013).

125. Murdoch, C. & Finn, A. Chemokine receptors and their role in inflammation and infectious diseases. *Blood* **95,** 3032–43 (2000).

126. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393,** 648–659 (1998).

127. Georgiev, I. S., Gordon Joyce, M., Zhou, T. & Kwong, P. D. Elicitation of HIV-1-neutralizing antibodies against the CD4-binding site. *Curr. Opin. HIV AIDS* **8,** 382–392 (2013).

128. Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393,** 705–11 (1998).

129. Yang, X., Kurteva, S., Ren, X., Lee, S. & Sodroski, J. Stoichiometry of Envelope Glycoprotein Trimers in the Entry of Human Immunodeficiency Virus Type 1. *J. Virol.* **79,** 12132–12147 (2005).

130. Yang, X., Kurteva, S., Lee, S. & Sodroski, J. Stoichiometry of antibody neutralization of human immunodeficiency virus type 1. *J. Virol.* **79,** 3500–3508 (2005).

131. Liu, J., Bartesaghi, A., Borgnia, M. J., Sapiro, G. & Subramaniam, S. Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455,** 109–113 (2008).

132. Schuitemaker, H., van ’t Wout, A. B. & Lusso, P. Clinical

significance of HIV-1 coreceptor usage. *J. Transl. Med.* **9,** S5 (2010).

133. Myszka, D. G. *et al.* Energetics of the HIV gp120-CD4 binding reaction. *Proc. Natl. Acad. Sci.* **97,** 9026–9031 (2000).

134. Healey, D. *et al.* Novel anti-CD4 monoclonal antibodies separate human immunodeficiency virus infection and fusion of CD4+ cells from virus binding. *J. Exp. Med.* **172,** 1233–42 (1990).

135. Ashish *et al.* Conformational rearrangement within the soluble domains of the CD4 receptor is ligand-specific. *J. Biol. Chem.* **283,** 2761–2772 (2008).

136. Sandoval, M. & Jimenez, A. Two-dimensional motion of Brownian swimmers in linear flows. *J. Biol. Phys.* **42,** 199–212 (2016).

137. English, T. J. & Hammer, D. A. Brownian adhesive dynamics (BRAD) for simulating the receptor-mediated binding of viruses. *Biophys. J.* **86,** 3359–72 (2004).

138. English, T. J. & Hammer, D. a. The effect of cellular receptor diffusion on receptor-mediated viral binding using Brownian adhesive dynamics (BRAD) simulations. *Biophys. J.* **88,** 1666–1675 (2005).

139. Burckhardt, C. J. & Greber, U. F. Virus movements on the plasma membrane support infection and transmission between cells. *PLoS Pathogens* **5,** (2009).

140. Moolla, N., Killick, M., Papathanasopoulos, M. & Capovilla, A.

Thioredoxin (Trx1) regulates CD4 membrane domain localization and is required for efficient CD4-dependent HIV-1 entry. *Biochim. Biophys. Acta - Gen. Subj.* **1860,** 1854–1863 (2016).

141. Matthias, L. J., Yam, P. T., Jiang, X. M. & Hogg, P. J. Disulfide exchange in CD4. *BioFactors* **17,** 241–248 (2003).

142. Ryser, H. J. P. & Flückiger, R. Keynote review: Progress in targeting HIV-1 entry. *Drug Discovery Today* **10,** 1085–1094 (2005).

143. Azimi, I., Matthias, L. J., Center, R. J., Wong, J. W. H. & Hogg, P. J. Disulfide Bond That Constrains the HIV-1 gp120 V3 Domain Is Cleaved by Thioredoxin. *J. Biol. Chem.* **285,** 40072–40080 (2010).

144. Reiser, K. *et al.* Thioredoxin-1 and protein disulfide isomerase catalyze the reduction of similar disulfides in HIV gp120. *Int. J. Biochem. Cell Biol.* **44,** 556–562 (2012).

145. Matthias, L. J. *et al.* Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1. *Nat. Immunol.* (2002). doi:10.1038/ni815

146. Cerutti, N., Killick, M., Jugnarain, V., Papathanasopoulos, M. & Capovilla, A. Disulfide Reduction in CD4 Domain 1 or 2 Is Essential for Interaction with HIV Glycoprotein 120 (gp120), which Impairs Thioredoxin-driven CD4 Dimerization. *J. Biol. Chem.* **289,** 10455–10465 (2014).

147. Matthias, L. J. & Hogg, P. J. Redox Control on the Cell Surface:

Implications for HIV-1 Entry. *Antioxid. Redox Signal.* **5,** 133–138 (2003).

148. Gallina, A. *et al.* Inhibitors of protein-disulfide isomerase prevent cleavage of disulfide bonds in receptor-bound glycoprotein 120 and prevent HIV-1 entry. *J. Biol. Chem.* **277,** 50579–50588 (2002).

149. Auwerx, J. *et al.* Human glutaredoxin-1 catalyzes the reduction of HIV-1 gp120 and CD4 disulfides and its inhibition reduces HIV-1 replication. *Int. J. Biochem. Cell Biol.* **41,** 1269–1275 (2009).

150. Bui, V.-C. & Nguyen, T.-H. Insights into the interaction of CD4 with anti-CD4 antibodies. *Immunobiology* **222,** 148–154 (2017).

151. Yuan, R., Qi, J., Zhang, Z., Li, S. & Gu, Y. Anti-CD4 : An Alternative Way to Inhibit HIV Infection. *J. HIV Retro Virus* **2,** 1–6 (2016).

152. Freeman, M. M. *et al.* Crystal structure of HIV-1 primary receptor CD4 in complex with a potent antiviral antibody. *Structure* **18,** 1632–1641 (2010).

153. Song, R. *et al.* Epitope mapping of ibalizumab, a humanized anti-CD4 monoclonal antibody with anti-HIV-1 activity in infected patients. *J. Virol.* **84,** 6935–6942 (2010).

154. Fisher, T. E., Marszalek, P. E. & Fernandez, J. M. Stretching single molecules into novel conformations using the atomic force microscope. *Nat. Struct. Biol.* **7,** 719–24 (2000).

155. Hughes, M. L. & Dougan, L. The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding. *Reports Prog. Phys.* **79,** 76601 (2016).

156. Ferreon, A. C. M. & Deniz, A. A. Protein folding at single-molecule resolution. *Biochim. Biophys. Acta - Proteins Proteomics* **1814,** 1021–1029 (2011).

157. Schönfelder, J., Perez-Jimenez, R. & Muñoz, V. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution. *Nat. Commun.* **7,** (2016).

158. Carrion-Vazquez, M. *et al.* Mechanical and chemical unfolding of a single protein: A comparison. *Biophysics (Oxf).* **96,** 3694–3699 (1999).

159. Schönfelder, J., De Sancho, D. & Perez-Jimenez, R. The Power of Force: Insights into the Protein Folding Process Using Single-Molecule Force Spectroscopy. *J. Mol. Biol.* **428,** 4245–4257 (2016).

160. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science (80-.).* **271,** 795–799 (1996).

161. Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I. & Bustamante, C. Reversible unfolding of single RNA molecules by mechanical force. *Science* **292,** 733–7 (2001).

162. Co, D., Present, D. M. & Experiments, a a. The molecular elasticity of the extracellular matrix protein tenascin. *Nature*

424, 181–185 (2003).

163. Pace, C. N. *et al.* Contribution of hydrogen bonds to protein stability. *Protein Sci.* **23,** 652–661 (2014).

164. Rico, F., Rigato, A., Picas, L. & Scheuring, S. Mechanics of proteins with a focus on atomic force microscopy. *J. Nanobiotechnology* **11,** S3 (2013).

165. Kang, H. J. & Baker, E. N. Intramolecular isopeptide bonds: Protein crosslinks built for stress? *Trends Biochem. Sci.* **36,** 229–237 (2011).

166. Fass, D. Disulfide Bonding in Protein Biophysics. *Annu. Rev. Biophys.* **41,** 63–79 (2012).

167. Fisher, T. E., Oberhauser, A. F., Carrion-Vazquez, M., Marszalek, P. E. & Fernandez, J. M. The study of protein mechanics with the atomic force microscope. *Trends Biochem. Sci.* **24,** 379–384 (1999).

168. Yu, H., Siewny, M. G. W., Edwards, D. T., Sanders, A. W. & Perkins, T. T. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science (80-.).* **355,** 945–950 (2017).

169. Ziegler, F. *et al.* Knotting and unknotting of a protein in single molecule experiments. *Proc. Natl. Acad. Sci.* **113,** 7533–7538 (2016).

170. Evans, E. & Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **72,** 1541–1555 (1997).

171. Dudko, O. K., Hummer, G. & Szabo, A. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci.* **105,** 15755–15760 (2008).

172. Cossio, P., Hummer, G. & Szabo, A. On artifacts in single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 14248–53 (2015).

173. Woodside, M. T. & Block, S. M. Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annu. Rev. Biophys.* **43,** (2014).

174. Hummer, G. & Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci.* **98,** 3658–3661 (2001).

175. Hummer, G. & Szabo, A. Free energy profiles from single-molecule pulling experiments. *Proc. Natl. Acad. Sci.* **107,** 21441–21446 (2010).

176. Popa, I., Kosuri, P., Alegre-Cebollada, J., Garcia-Manyes, S. & Fernandez, J. M. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. *Nat. Protoc.* **8,** 1261–1276 (2013).

177. Beedle, A. E. M., Lynham, S. & Garcia-Manyes, S. Protein S-sulfenylation is a fleeting molecular switch that regulates non-enzymatic oxidative folding. *Nat. Commun.* **7,** 12490 (2016).

178. Carrion-Vazquez, M. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.* **74,** 63–91 (2000).

179. Hoffmann, T., Tych, K. M., Hughes, M. L., Brockwell, D. J. & Dougan, L. Towards design principles for determining the mechanical stability of proteins. *Phys. Chem. Chem. Phys.* **15,** 15767 (2013).

180. Oberhauser, a F., Hansma, P. K., Carrion-Vazquez, M. & Fernandez, J. M. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 468–72 (2001).

181. Javadi, Y., Fernandez, J. M. & Perez-Jimenez, R. Protein Folding Under Mechanical Forces: A Physiological View. *Physiology* **28,** 9–17 (2013).

182. Garcia-Manyes, S., Kuo, T. L. & Fernández, J. M. Contrasting the individual reactive pathways in protein unfolding and disulfide bond reduction observed within a single protein. *J. Am. Chem. Soc.* **133,** 3104–3113 (2011).

183. Vetsch, M. *et al.* Pilus chaperones represent a new type of protein-folding catalyst. *Nature* **431,** 329–333 (2004).

184. Barnhart, M. M., Sauer, F. G., Pinkner, J. S. & Hultgren, S. J. Chaperone-subunit-usher interactions required for donor strand exchange during bacterial pilus assembly. *J. Bacteriol.* **185,** 2723–2730 (2003).

185. https://www.genomics.agilent.com/files/Manual/200249.pdf.

186. https://www.thermofisher.com/es/es/home/brands/thermo-scientific/molecular-biology/thermo-scientific-restriction-modifying-enzymes/restriction-enzymes-thermo-

scientific/fastdigest-thermo-scientific/reaction-conditions-for-fastdigest-enzymes.html.

187. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/t4dnaligase_1U_man.pdf.

188. http://www.agilent.com/cs/library/usermanuals/public/200133.pdf.

189. https://www.neb.com/protocols/0001/01/01/5-minute-transformation-protocol-c2987.

190. Koshland, D. & Botstein, D. Secretion of beta-lactamase requires the carboxy end of the protein. *Cell* **20,** 749–760 (1980).

191. Kahn, T. B., Fernández, J. M. & Perez-Jimenez, R. Monitoring oxidative folding of a single protein catalyzed by the disulfide oxidoreductase DsbA. *J. Biol. Chem.* **290,** 14518–14527 (2015).

192. Manbachi, A. & Cobbold, R. S. C. Development and application of piezoelectric materials for ultrasound generation and detection. *Ultrasound* **19,** 187–196 (2011).

193. Hutter, J. L. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64,** 1868–1873 (1993).

194. Wiita, A. P. *et al.* Probing the chemistry of thioredoxin catalysis with force. *Nature* **450,** 124–127 (2007).

195. Bustamante, C., Marko, J., Siggia, E. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science (80-.).* **265,** 1599–1600 (1994).

196. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28,** 8759–8770 (1995).

197. Kratky, O. & Porod, G. Röntgenuntersuchung gelöster Fadenmoleküle. *Recl. des Trav. Chim. des Pays-Bas* **68,** 1106–1122 (2010).

198. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91,** 43–56 (1995).

199. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926 (1983).

200. MacKerell, A. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* **102,** 3586–3616 (1998).

201. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 14101 (2007).

202. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J Chem Phys* **103,** 8577–8593 (1995).

203. Crespo, M. D. *et al.* Quality control of disulfide bond formation in pilus subunits by the chaperone FimC. *Nat. Chem. Biol.* **8,** 707–13 (2012).

204. Perez-Jimenez, R. *et al.* Probing the Effect of Force on HIV-1 Receptor CD4. *ACS Nano* **8,** 10313–10320 (2014).

205. Aprikian, P. *et al.* Interdomain interaction in the FimH adhesin

of Escherichia coli regulates the affinity to mannose. *J. Biol. Chem.* **282,** 23437–23446 (2007).

206. Yakovenko, O. *et al.* FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation. *J. Biol. Chem.* **283,** 11596–11605 (2008).

207. Jass, J. *et al.* Physical Properties of Escherichia coli P Pili Measured by Optical Tweezers. *Biophys. J.* **87,** 4271–4283 (2004).

208. Axner, O. *et al. Bacterial Adhesion*. **715,** (Springer Netherlands, 2011).

209. Zakrisson, J. *et al.* Rigid multibody simulation of a helix-like structure: the dynamics of bacterial adhesion pili. *Eur. Biophys. J.* **44,** 291–300 (2015).

210. Forero, M., Yakovenko, O., Sokurenko, E. V., Thomas, W. E. & Vogel, V. Uncoiling mechanics of Escherichia coli type I fimbriae are optimized for catch bonds. *PLoS Biol.* **4,** 1509–1516 (2006).

211. Whitfield, M. J., Luo, J. P. & Thomas, W. E. Yielding Elastic Tethers Stabilize Robust Cell Adhesion. *PLoS Comput. Biol.* **10,** (2014).

212. Valbuena, A. *et al.* On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 13791–6 (2009).

213. Echelman, D. J. *et al.* CnaA domains in bacterial pili are

efficient dissipaters of large mechanical shocks. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 2490–5 (2016).

214. Barnhart, M. M. *et al.* PapD-like chaperones provide the missing information for folding of pilin proteins. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7709–14 (2000).

215. Nilsson, L. M. *et al.* The cysteine bond in the Escherichia coli FimH adhesin is critical for adhesion under flow conditions. *Mol. Microbiol.* **65,** 1158–1169 (2007).

216. Manteca, A. *et al.* The Influence of Disulfide Bonds on the Mechanical Stability of Proteins is Context Dependent. *J. Biol. Chem.* **292,** 13374–13380 (2017).

217. Betz, S. F. Disulfide bonds and the stability of globular proteins. *Protein Sci.* **2,** 1551–1558 (1993).

218. Martin, J. L., Bardwell, J. C. A. & Kuriyan, J. Crystal structure of the DsbA protein required for disulphide bond formation in vivo. *Nature* **365,** 464–468 (1993).

219. Zheng, W. D., Quan, H., Song, J. L., Yang, S. L. & Wang, C. C. Does DsbA have chaperone-like activity? *Arch. Biochem. Biophys.* **337,** 326–31 (1997).

220. Frech, C., Wunderlich, M., Glockshuber, R. & Schmid, F. X. Preferential binding of an unfolded protein to DsbA. *EMBO J.* **15,** 392–98 (1996).

221. Koldewey, P., Stull, F., Horowitz, S., Martin, R. & Bardwell, J. C. A. Forces Driving Chaperone Action. *Cell* **166,** 369–379

(2016).

222. Stull, F., Koldewey, P., Humes, J. R., Radford, S. E. & Bardwell, J. C. A. Substrate protein folds while it is bound to the ATP-independent chaperone Spy. *Nat. Struct. Mol. Biol.* **23,** 53–58 (2015).

223. DeVico, A. L. CD4-induced epitopes in the HIV envelope glycoprotein, Gp120. *Curr HIV Res* **5,** 561–571 (2007).

224. Gardner, M. R. & Farzan, M. Engineering antibody-like inhibitors to prevent and treat HIV-1 infection. *Curr. Opin. HIV AIDS* **12,** 294–301 (2017).

225. Jacobson, J. M. & Flexner, C. W. Universal antiretroviral regimens. *Curr. Opin. HIV AIDS* **12,** 343–350 (2017).

226. Ahmed, Y., Tian, M. & Gao, Y. Development of an anti-HIV vaccine eliciting broadly neutralizing antibodies. *AIDS Res. Ther.* **14,** 50 (2017).

227. Hou, W. *et al.* Molecular insights into the inhibition of HIV-1 infection using a CD4 domain-1-specific monoclonal antibody. *Antiviral Res.* **122,** 101–111 (2015).

228. Luo, Z. *et al.* Pathological Role of Anti-CD4 Antibodies in HIV-Infected Immunologic Nonresponders Receiving Virus-Suppressive Antiretroviral Therapy. *J. Infect. Dis.* **216,** 82–91 (2017).

229. Yachou, A. & Sékaly, R.-P. Binding of Soluble Recombinant HIV Envelope Glycoprotein, rgp120, Induces Conformational

Changes in the Cellular Membrane-Anchored CD4 Molecule. *Biochem. Biophys. Res. Commun.* **265,** 428–433 (1999).

230. Szoszkiewicz, R. Spectroscopies in Investigating Mechanochemical Reactions and Mechanical Properties of Single Biomolecules. *Scanning Probe Microsc. Nanosci. Nanotechnol. SE - 12* 395–423 (2010). doi:10.1007/978-3-642-03535-7_12

231. Ryu, S.-E. *et al.* Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* **348,** 419–426 (1990).

232. Evans, E. & Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **72,** 1541–1555 (1997).

233. Stantchev, T. S. *et al.* Cell-type specific requirements for thiol/disulfide exchange during HIV-1 entry and infection. *Retrovirology* **9,** 97 (2012).

234. Sanejouand, Y.-H. Domain swapping of CD4 upon dimerization. *Proteins Struct. Funct. Bioinforma.* **57,** 205–212 (2004).

235. Zhou, B., Baldus, I. B., Li, W., Edwards, S. A. & Gräter, F. Identification of Allosteric Disulfides from Prestress Analysis. *Biophys. J.* **107,** 672–681 (2014).

236. Anjukandi, P. *et al.* The Effect of Tensile Stress on the Conformational Free Energy Landscape of Disulfide Bonds. *PLoS One* **9,** 2–8 (2014).

237. Baldus, I. B. & Gräter, F. Mechanical force can fine-tune redox

potentials of disulfide bonds. *Biophys. J.* **102,** 622–629 (2012).

238. Liang, J. & Fernández, J. M. Mechanochemistry: One bond at a time. *ACS Nano* **3,** 1628–1645 (2009).

239. Smith, R. P., Paxman, J. J., Scanlon, M. J. & Heras, B. Targeting bacterial Dsb proteins for the development of anti-virulence agents. *Molecules* **21,** 811 (2016).

240. Singh, B. *et al.* Antibody-mediated disruption of the mechanics of CS20 fimbriae of enterotoxigenic Escherichia coli. *Nat. Sci Rep* **5,** 13678 (2015).

Appendix

Protein sequences

(I91)$_2$-CD4D1D2-(I91)$_2$

LIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEI
IEDGKKHILILHNCQLGMTGEVSFQAANAKSAANLKVKELIEVEKPLY
GVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKKHIL
ILHNCQLGMTGEVSFQAANAKSAANLKVKERSRSGSSKKVVLGKKG
DTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADS
RRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDT
HLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG
TWTCTVLQNQKKVEFKIDIVVLAFQKASSRSLIEVEKPLYGVEVFVGE
TAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKKHILILHNCQLG
MTGEVSFQAANAKSAANLKVKELIEVEKPLYGVEVFVGETAHFEIELS
EPDVHGQWKLKGQPLTASPDCEIIEDGKKHILILHNCQLGMTGEVSFQ
AANAKSAANLKVKERS

(I91)$_2$-FimA-(I91)$_2$

LIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEI
IEDGKKHILILHNCQLGMTGEVSFQAANAKSAANLKVKELIEVEKPLY
GVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKKHIL
ILHNCQLGMTGEVSFQAANAKSAANLKVKERSRSACAVDAGSVDQT
VQLGQVRTASLAQEGATSSAVGFNIQLNDCDTNVASKAAVAFLGTAI
DAGHTNVLALQSSAAGSATNVGVQILDRTGAALTLDGATFSSETTLN
NGTNTIPFQARYFATGAATPGAANADATFKVQYQDNKQAATTVNGG
TVHFKGEVVNARSLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQW
KLKGQPLTASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANAKSAA
NLKVKELIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLT
ASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANAKSAANLKVKERS

(I91)$_2$-FimF-(I91)$_2$

LIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEI
IEDGKKHILILHNCQLGMTGEVSFQAANAKSAANLKVKELIEVEKPLY

GVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKKHIL
ILHNCQLGMTGEVSFQAANAKSAANLKVKERSRSDGNGCSVAAEST
NFTVDLMENAAKQFNNIGATTPVVPFRILLSPCGNAVSAVKVGFTGV
ADSHNANLLALENTVSAASGLGIQLLNEQQNQIPLNAPSSALSWTTLT
PGKPNTLNFYARLMATQVPVTAGHINATATFTLEYQDNKQAATTVN
GGTVHFKGEVVNARSLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQ
WKLKGQPLTASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANAKSA
ANLKVKELIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQP
LTASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANAKSAANLKVKE
RS

(I91)$_2$-FimG-(I91)$_2$

LIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEI
IEDGKKHILILHNCQLGMTGEVSFQAANAKSAANLKVKELIEVEKPLY
GVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKKHIL
ILHNCQLGMTGEVSFQAANAKSAANLKVKERSRSVVAKPCTVSTTNA
TVDLGDLYSFSLMSAGAASAWHDVALELTNCPVGTSRVTASFSGADS
TGYYKNQGTAQNIQLELQDDSGNTLNTGATKTVQVDDSSQSAHFPVR
ALTVNGGATQGTIQAVISITYTYSDNKQADSTITIRGYVRRSLIEVEKP
LYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKK
HILILHNCQLGMTGEVSFQAANAKSAANLKVKELIEVEKPLYGVEVF
VGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKKHILILHNC
QLGMTGEVSFQAANAKSAANLKVKERS

(I91)$_2$-FimH-(I91)$_2$

LIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEI
IEDGKKHILILHNCQLGMTGEVSFQAANAKSAANLKVKELIEVEKPLY
GVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKKHIL
ILHNCQLGMTGEVSFQAANAKSAANLKVKERSRSFACKTANGTAIPIG
GGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRG
SAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYL
TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVP
TGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADA
GNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLT
ANYARTGGQVTAGNVQSIIGVTFVYQDNKQADVTITVNGKVRSLIEV
EKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIED

GKKHILILHNCQLGMTGEVSFQAANAKSAANLKVKELIEVEKPLYGV
EVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKKHILIL
HNCQLGMTGEVSFQAANAKSAANLKVKERS