Caracterización del complejo mayor de histocompatibilidad clase II en primates del género *Aotus*

Carlos Fernando Suárez Martínez

"Tesis de Doctorado presentada como requisito para optar por el título de Doctor en Ciencias Biomédicas y Biológicas de la Universidad del Rosario"

Bogotá D.C., 2017

Caracterización del complejo mayor de histocompatibilidad clase II en primates del género *Aotus*

Estudiante

Carlos Fernando Suárez Martínez

Directores

Manuel Alfonso Patarroyo Gutiérrez M.D., Dr.Sc.

Fundación Instituto de Inmunología de Colomba (FIDIC)

Universidad del Rosario

Luis Fernando Cadavid Gutiérrez M.D., Ph.D.

Universidad Nacional de Colombia

DOCTORADO EN CIENCIAS BIOMÉDICAS Y BIOLÓGICAS

UNIVERSIDAD DEL ROSARIO

Bogotá. D.C., 2017

Agradecimientos

Quiero expresar mi gratitud a mi familia, especialmente a mis padres, por su apoyo constante y por ser mi brújula moral.

A mis directores, el Doctor Manuel Alfonso Patarroyo y el Doctor Luis Fernando Cadavid, por sus aportes, por la libertad y por la confianza con la que me permitieron desarrollar el proyecto.

Al Profesor Manuel Elkin Patarroyo, por su generosidad, tenacidad e inspiración.

A mis colegas de la FIDIC, especialmente a Carolina López, Hugo Bohórquez y Ronald González, pues sin su apoyo y aportes, este proyecto no habría sido posible.

A la Universidad del Rosario, por la extraordinaria oportunidad de desarrollar mis estudios, especialmente a la Doctora Luisa Matheus, por su diligencia y colaboración para hacer todos los procesos lo más sencillos posibles.

Contenido

LISTA DE FIGURAS Y TABLAS

Resumen

El estudio del complejo mayor de histocompatibilidad (CMH) de los monos del género *Aotus*, y la comprensión del proceso de unión CMH-péptido, son importantes para entender las semejanzas y diferencias en la respuesta inmune entre humanos y los monos del género *Aotus*. Esto tiene implicaciones para el uso apropiado y la validez de las conclusiones alcanzadas, cuando se utilizan estos animales como modelos experimentales en el desarrollo de vacunas y fármacos.

El presente trabajo tiene como propósito contribuir al conocimiento del complejo mayor de histocompatibilidad clase II de los monos *Aotus*. Con la determinación de la secuencia de los genes del CMH-DPA y CMH-DRA, se ha completado la caracterización del CMH de los monos *Aotus*, contribuyendo a la validación de este primate como modelo experimental, y aumentando el conocimiento en la evolución de los genes del CMH en primates. Además, se profundizó en el análisis de convergencia y polimorfismo de los genes del CMH-DR en primates.

Adicionalmente, se implementaron metodologías de modelación computacional de la unión CMH-péptido (basadas en química quántica y redes neurales), como herramientas necesarias para entender los mecanismos de presentación de péptidos por parte del CMH clase II a los linfocitos T. El estudio del polimorfismo de la región de unión al péptido, permitió el desarrollo de estrategias (perfiles de bolsillos) para reducir eficientemente el número de sistemas a considerar en el diseño de péptidos a ser usados como candidatos a vacuna contra la malaria.

Usando minería de datos sobre distribuciones de Ramachandran, se desarrolló una escala de similitud estructural de aminoácidos, con el fin de implementar su uso en el desarrollo de péptidos candidatos a vacunas. Adicionalmente, se encontró que la estructura secundaria de las proteínas tiene una relación clara con los patrones evolutivos de sustitución y la mutabilidad de los aminoácidos.

Así, se ha generado un marco de conceptual que contribuye al desarrollo de vacunas basadas en péptidos, que tiene como base el estudio del polimorfismo del complejo mayor de histocompatibilidad, las restricciones fisicoquímicas/estructurales que moldean el proceso de reconocimiento molecular involucrado en la interacción CMH-péptido y la aplicación de metodologías computacionales para cuantificar el proceso de unión CMH-péptido.

Summary

Studying the *Aotus* major histocompatibility complex (MHC) and understanding MHC-peptide binding are important issues for recognizing similarities and differences regarding immune response between humans and *Aotus*. This has implications for the appropriate use and validity of the conclusions reached when these animals are used as experimental models when developing vaccines and drugs.

This work was aimed to contribute to increase our knowledge on the MHC class II in monkeys from the genus *Aotus*. Determining the sequences of MHC-DPA and MHC-DRA genes has allowed to complete the characterisation of the *Aotus* MHC, contributing towards validating the role of this primate as experimental model and increasing our knowledge regarding MHC gene evolution in primates. It also dealt with in–depth analysis of MHC-DR genes' convergence and polymorphism in primates.

The study involves computational modelling of MHC-peptide binding methodologies (based on quantum chemistry and neural networks) as necessary tools for understanding the mechanisms of MHC class II peptide presentation to T-lymphocytes. Studying peptide binding region polymorphism has enabled developing strategies (pocket profiles) for efficiently reducing the amount of systems to be considered when designing peptides to be used as candidates for an antimalarial vaccine.

Data-mining regarding Ramachandran distribution led to developing an amino acid structural similarity scale for use in developing/designing peptides as vaccine candidates. It was found that protein secondary structure has a clear relationship with amino acid substitution and mutability evolutionary patterns.

A conceptual framework thus emerged aimed at developing peptide-based vaccines as a basis for studying the mayor histocompatibility complex polymorphism, the physicochemical/structural restrictions shaping the molecular recognition involved in MHC-peptide interaction and using computational methodologies for quantifying MHC-peptide binding.

Introducción

Aotus, generalidades y distribución

Todas las especies de este género se caracterizan por tener una talla pequeña (50 – 80 cm), con un peso entre 500 -1000 gramos. Su pelaje varía entre gris y marrón brillante, con una coloración rojiza alrededor de su cuello, en la cara interna de sus extremidades y en la base de la cola, que no es prensil. La clasificación taxonómica de las especies de *Aotus* es compleja debido a su enorme similitud morfológica, lo que ha dificultado establecer un consenso sobre su número. Varios estudios taxonómicos con base en las características fenotípicas y citogenéticas, y la distribución geográfica de los monos *Aotus*, han permitido proponer la existencia de 9 a 12 especies de *Aotus* desde Panamá hasta el norte de Argentina (3-7).

Existe registro fósil del género en la fauna del mioceno medio en la Venta, Colombia, datado en 12 – 15 Millones de años (*Aotus dindensis*) (8, 9). El origen del género se data hace aproximadamente 20 Millones de años (~19,3, usando 54 genes nucleares (10) o ~20,0 millones de años usando genomas mitocondriales (11)). Se ha estimado la divergencia de las especies actuales, con base en la caracterización de varias regiones mitocondriales entre 3,1 – 6,4 millones de años (12) o usando genes nucleares entre 3,2 – 7,9 millones de años (10).

Las especies de este género se encuentran en altitudes que van desde el nivel del mar hasta 3.200 metros en bosques húmedos tropicales y subtropicales. Es el único grupo de primates neotropicales nocturnos, lo que representa una ventaja adaptativa para su reproducción y supervivencia (5-7, 13-18). Siete especies han sido reportadas en Colombia hasta la fecha: *Aotus zonalis* (región del pacífico norte), *A. griseimembra* (costa atlántica y región andina), *A. lemurinus* (costa atlántica y región andina), *A. brumbacki* (departamento del Meta), *A. vociferans* (región amazónica), *A. nancymaae* (región amazónica) y *A. jorgehernandezi* (región andina) (3-7, 19).

Aotus como modelo experimental

La disponibilidad de modelos experimentales animales bien caracterizados es fundamental para el desarrollo de métodos terapéuticos, contribuyendo además a la investigación en inmunología comparada y en la evolución del sistema inmune. La necesidad de primates como modelos animales se resalta por la inhabilidad de otros modelos animales ampliamente usados (como el murino) de presentar susceptibilidad a enfermedades o procesos infecciosos específicos de los seres humanos (por ejemplo, la hipertensión y la osteoporosis ocurren naturalmente en todos los primates). La información experimental obtenida en primates es más fácilmente extrapolable a seres humanos y a otros primates, lo que permite determinar la eficacia de tratamientos en casos donde otros modelos animales fallan (20-22). Durante los últimos 35 años, los monos del género *Aotus* (Familia Aotidae, Parvorden Platyrrihini) han sido usados en el desarrollo de una vacuna contra la malaria por el Instituto de Inmunología del Hospital San Juan de Dios, y posteriormente por la Fundación Instituto de Inmunología de Colombia (FIDIC) (23, 24).

Algunas especies de *Aotus* han sido usadas desde hace más de 50 años como modelo para el estudio de la malaria (25, 26). A diferencia de otros modelos primates, los *Aotus* son susceptibles a la infección con esporozoítos, lo que permite el desarrollo de vacunas y fármacos para el tratamiento en todas las fases de la enfermedad (21). Estos monos también son susceptibles a otras enfermedades humanas, como leishmaniosis, esquistosomiasis, hepatitis, tuberculosis, y varios tipos de infecciones entéricas como campylobacteriosis, siendo también usados para el desarrollo de fármacos y estudio de estas enfermedades (27-33). *Aotus* también es uno de los modelos primates mejor conocidos de fisiología de la visión y electrofisiología del sistema nervioso central (34). Todo lo anterior, sumado a su facilidad de manejo en laboratorio (talla, adaptación, longevidad) son ventajas que hacen de los primates de este género un valioso modelo, y justifican la profundización en el conocimiento biológico de las especies que a él pertenecen.

Caracterización de las moléculas del sistema inmune de *Aotus* para corroborar su idoneidad como modelo experimental

Distintos componentes clave del sistema inmune de los monos *Aotus* han sido caracterizados: KIRs (35), CD1 (36), CD3 (37), CD45 (38, 39), IGKV (40), IGHV (41), TCR (42-44), algunas de las citoquinas (45), receptores similares a Toll (en inglés, *Toll-like receptors*) (46), células dendríticas (47), células T (48), perfil linfo-proliferativo (49), y los esplenocitos (50). Además de las anteriores, son de especial interés los genes del complejo mayor de histocompatiblidad (CMH). Las proteínas codificadas por los genes del CMH juegan un papel central en el reconocimiento de lo propio y lo ajeno, al efectuar la presentación de los péptidos para su reconocimiento por las células T, siendo fundamentales en la defensa contra los agentes extraños. La variación genética del CMH es clave para entender la respuesta a las vacunas por parte de los hospederos (51, 52). En *Aotus*, se han caracterizado tanto los genes de clase I (53-55), como los de clase II (56-63). *Aotus* muestra una alta identidad (>~80%) al compararlo con humanos, en todas las moléculas del sistema inmune caracterizadas hasta el momento, demostrando la viabilidad de su uso para obtener resultados extrapolables a humanos (64).

Complejo mayor de histocompatibilidad. Generalidades

Los genes del CMH conforman una familia multigénica que codifica para glicoproteínas receptoras expresadas en la membrana celular. Estas proteínas juegan un papel central en el reconocimiento de lo propio y lo ajeno, siendo piezas clave en la defensa contra agentes extraños, al efectuar la presentación de los péptidos para su reconocimiento por las células T. En humanos y otros primates, éstos se organizan en un clúster con otros genes mayoritariamente relacionados con el sistema inmune, y se dividen en tres regiones cromosómicas (I, II y III), reflejando también especializaciones funcionales.

Este arreglo está relativamente conservado en todos los mamíferos: (I) La región de los genes de CMH clase I, cuya región de unión al péptido está constituida por dos dominios

(α1 y α2) que son codificados por un solo gen, y son expresados en todos los tipos celulares nucleados. El CMH clase I presenta péptidos de origen intracelular a los linfocitos T CD8+. En esta región, también se encuentran otros genes críticos para el procesamiento de antígenos como la tapasina. (II) En la siguiente región, se encuentran los genes de clase II, su región de unión al péptido es codificada por dos genes (cadenas α1 y β1), y son expresados en células presentadoras de antígeno como los monocitos, macrófagos, linfocitos B, etc., presentando péptidos a los linfocitos T CD4+, que han sido adquiridos primordialmente por endocitosis/fagocitosis de proteínas exógenas o por carga directa en la superficie; y (III) la región de los genes de clase III, que codifican para otros componentes del sistema inmune, como el sistema de complemento (vg. C2, C4, factor B) y citoquinas (vg. TNF-α) (Figura 1).

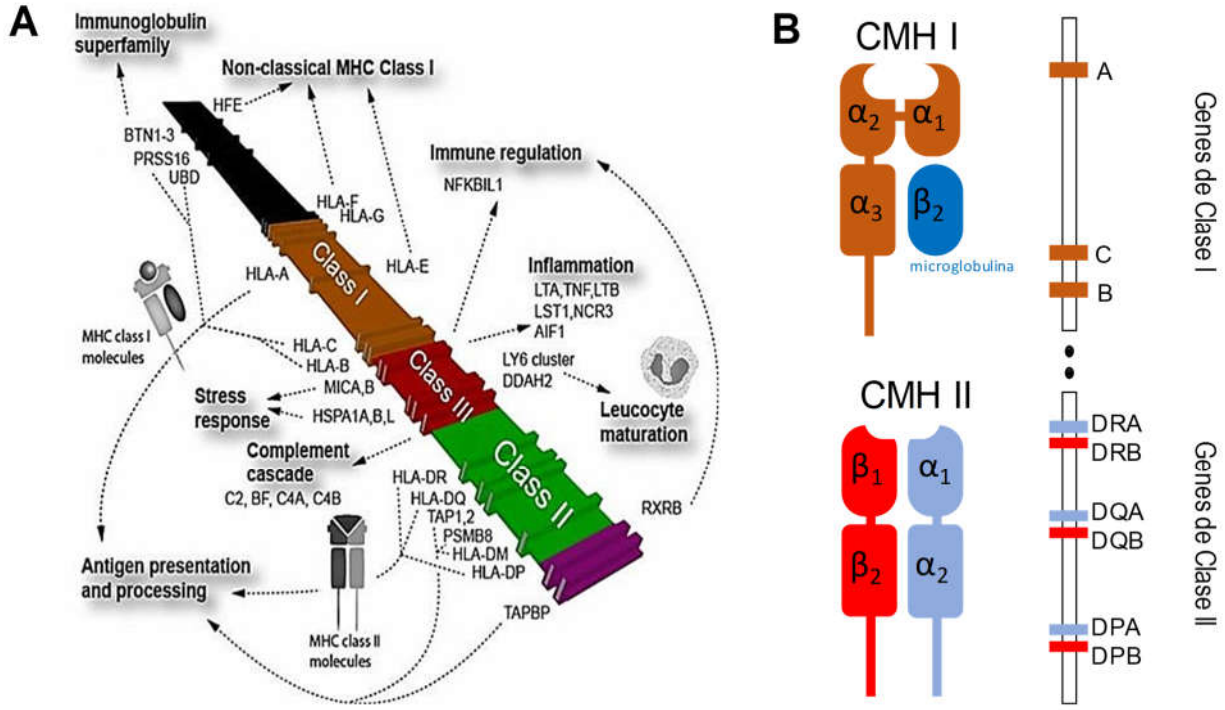


Figura 1. Organización genómica del HLA y disposición de los dominios de CMH I y II.

A. *Representación del complejo mayor de histocompatibilidad humano en el cromosoma 6p21 (Tomado de (1)).* ***B.*** *Arquitectura de dominios de CMH I y II (gráfica propia).*

La región del MHC muestra sintenía entre todos los mamíferos, y en humanos se encuentra en el cromosoma 6, comprendiendo 140 genes y un tamaño de 3,6 Mpb (65), siendo posible usarla como patrón para la caracterización del sector en otros primates, como es el caso de *Macaca fascicularis*, en donde la región tiene un tamaño de 4,3 Mpb (66).

CMH. Polimorfismo y convergencia

El CMH incluye los genes más polimórficos en los vertebrados y se constituyen en un modelo paradigmático para el estudio de los mecanismos de adaptación a nivel molecular (67-69). A manera de ejemplo, para los loci más polimórficos en humanos (HLA-B en clase I y HLA-DRB en clase II), se han reportado a la fecha en la base de datos IMGT-HLA (70) más de 4.800 alelos de HLA-B, así como más de 2.100 alelos de HLA-DRB.

La dinámica poblacional (cambios en el tamaño de las poblaciones y deriva génica), la recombinación, conversión génica, y la selección natural, son fuerzas que causan y modelan el polimorfismo del CMH (71-73). Se han propuesto múltiples procesos que mantienen el polimorfismo del CMH (74): por selección balanceada, bien sea por sobredominancia (75, 76), por selección dependiente de frecuencia (77, 78), o por variación espacial / temporal de la presión ejercida por patógenos (52, 79, 80).

Otros mecanismos, no directamente asociados a la relación hospedero-patógeno, tienen que ver con patrones de apareamiento dependientes de características relacionadas a olores o simetría, que buscan obtener la mayor heterocigosidad posible para la progenie, evitando la endogamia y favoreciendo el emparejamiento con individuos con distinto repertorio de CMH (81-89). También existen mecanismos reproductivos, relacionados con la fertilización selectiva (90, 91). El CMH también muestra polimorfismo trans-específico (TSP) adaptativo, presentándose alelos de larga duración, que son compartidos por varias especies (92-94).

Adicionalmente, el estudio de la evolución del CMH-DRB ha mostrado que existe una convergencia a nivel molecular en la región de unión al péptido entre primates del nuevo mundo (Platyrrhini) y primates del viejo mundo (Catarrhini) (60, 62, 95, 96). La existencia de alelos en común entre primates del nuevo mundo y primates del viejo mundo, puede estar relacionada con la conservación de motivos de unión a péptidos (97).

La iniciativa para el estudio de los monos *Aotus* en la FIDIC, nos ha permitido caracterizar diversos genes del CMH en las especies *A. nancymaae*, *A. vociferans* y *A. nigriceps*, centrándose en la variabilidad, comparándolos con humanos y otros primates; todos estos estudios se han enfocado esencialmente en la región de unión al péptido (en el caso de CMH clase II, ésta es codificada por el exón 2, y para clase I, en los exones 2 y 3). Se ha realizado la caracterización del CMH clase I (53, 54), así como del CMH clase II: DQA y DQB (58), DPA (61), DPB (59) y DRB (56, 60, 62, 63).

Hasta el momento, la evidencia indica que el locus más polimórfico del complejo mayor de histocompatibilidad clase II en *Aotus* es el CMH-DRB (como en humanos y en otros primates), en contraposición de un CMH-DRA con muy bajo polimorfismo, seguido por CMH-DQ (A y B) y por ultimo CMH-DP (A y B) (a diferencia de humanos, y similar al mono Rhesus) (98-100). A pesar que la divergencia entre monos del nuevo mundo y humanos se puede datar en aproximadamente ~43 millones de años (10, 11, 101), estos estudios señalan que los genes del CMH de *Aotus* y humanos presentan algunas semejanzas, bien sea por homología (CMH clase I) (53, 54) o por convergencia (CMH clase II DRB) (60, 62).

CMH. Polimorfismo y repertorio de presentación

El proceso de presentación de antígenos, tiene un paso critico en la unión de los péptidos al CMH para su presentación al receptor de los linfocitos T (Figura 2A). El receptor del CMH está constituido por una región de unión al péptido que está formado por un conjunto de subreceptores denominados bolsillos de unión (*pockets*, en inglés) (Figura 2B). Típicamente para clase II, el péptido es anclado en una región de unión de 9

aminoácidos, existiendo múltiples marcos de unión, dado que el surco es abierto (Figura 2C), a diferencia de CMH clase I, en donde el surco de unión es cerrado, lo que impone un marco de unión único (102, 103). Estas características, hacen que las moléculas del CMH clase II posean un repertorio de ligandos mayor que las moléculas de clase I (104).

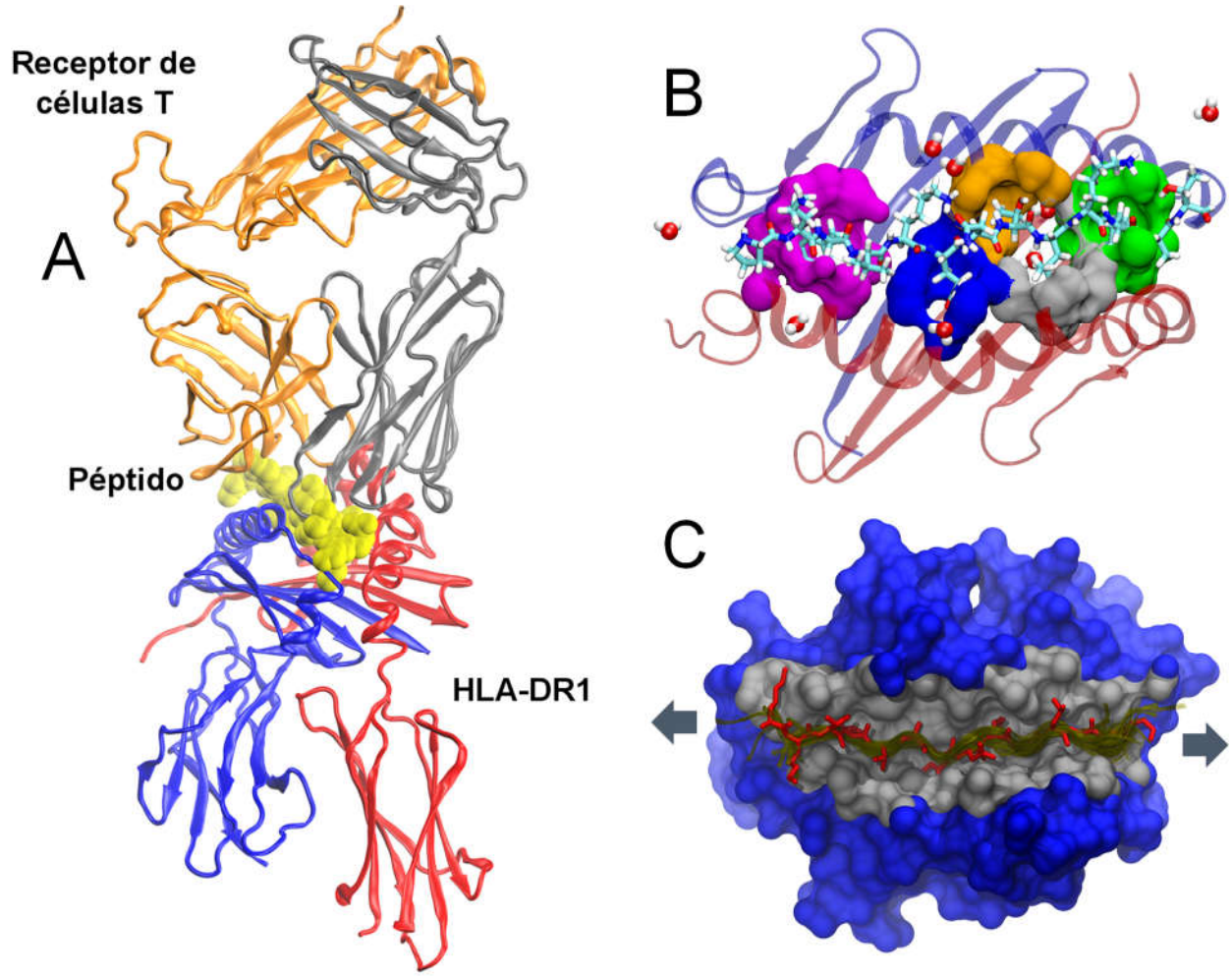


Figura 2. Arquitectura del CMH-DR. A. *Estructura del HLA-DR1 presentando el péptido de la Triosa-fosfato isomerasa al TCR (PDB=2IAN).* ***B.*** *Vista desde arriba del HLA-DR1 (PDB=1DLH) presentando el péptido de hemaglutinina; en purpura el bolsillo 1, en azul oscuro el bolsillo 4, en naranja el bolsillo 6, en gris el bolsillo 7 y en verde el bolsillo 9.* ***C.*** *Vista desde arriba de la región de unión al péptido (en gris), mostrando los posibles marcos de unión del péptido al CMH. La estructura abierta del receptor permite la existencia de múltiples marcos de unión (Gráfica propia generada a partir de las coordenadas descargadas del Protein Data Bank (PDB)).*

Así, el repertorio de péptidos capaces de ser reconocidos por una molécula específica de CMH, a pesar de ser vasto, cuenta con restricciones que obedecen a la arquitectura del receptor. Esta restricción en el repertorio de presentación está relacionada con diversidad de moléculas observada en el CMH y evidencia de ello es que tal diversidad alélica está concentrada en los residuos que intervienen en el proceso de unión al péptido (67-69). Existe una relación entre la diversificación de linajes en busca de la mayor capacidad de presentación posible y la existencia de variantes alélicas y polimorfismo trans-específico (105).

El espectro de péptidos que pueden ser unidos por las moléculas de CMH puede superponerse. Así, se pueden definir *supertipos* de CMH con base en este espectro. Esta similitud en la capacidad de unión generalmente está relacionada con una significativa similitud en las secuencias que constituyen los bolsillos de unión y tiene implicaciones en la resistencia de las poblaciones naturales (106-111). La estimación de la capacidad de unión de péptidos al CMH es de primordial interés para optimizar el diseño de vacunas (103), y desde el punto de vista de la validación del modelo experimental, es de interés estudiar sí la similitud entre los bolsillos de unión de humanos y monos *Aotus* implica la existencia de repertorios similares de unión de péptidos.

CMH. Predicción de péptidos de unión

Estimar experimentalmente la unión de péptidos al CMH es un procedimiento complejo. La obtención de un receptor viable para los ensayos de unión, bien sea por purificación a partir de líneas celulares inmortalizadas, o por la expresión de estas moléculas usando la tecnología de ADN recombinante, requieren de procedimientos dispendiosos y costosos. Adicionalmente, el número de sistemas a estudiar es enorme, dado el polimorfismo de las moléculas del CMH y la diversidad de péptidos con potencial de unión (112, 113). Teniendo en cuenta lo anterior, la implementación y desarrollo de metodologías computacionales para estudiar y predecir la interacción CMH-péptido, es una alternativa racional y necesaria.

Los métodos computacionales para la estimación de la interacción CMH-péptido pueden ser divididos en *métodos basados en secuencia* y *métodos basados en estructura*. Los primeros, usan datos de unión experimentales como punto de partida, para la generación de motivos de unión por posición (114-121), métodos de inteligencia artificial (redes neuronales como NetMHCIIpan) (122-125), modelos ocultos de Markov (126-128) y máquinas vectoriales/kernels (129-132). Estas aproximaciones, pretenden resolver el problema de la predicción *únicamente*, pero que no aportan conocimiento en términos de la *naturaleza* del proceso de unión entre el péptido y el CMH.

Los *métodos basados en estructura*, estiman la energía de unión péptido-CMH, basándose en las propiedades estructurales y no requieren de entrenamiento con datos de unión experimentales. A primera vista, resulta más atractivo este enfoque, pues además de ser predictivo, permite disecar los procesos involucrados en el proceso de unión. El enfoque mayoritariamente usado para calcular la energía de unión, usa aproximaciones de la mecánica molecular clásica, como dinámica molecular, en donde usando campos de fuerza que describen los tipos y magnitud de las interacciones involucradas, se estima el cambio en la energía libre de Gibbs durante la formación del complejo CMH-péptido, el cual se define como la diferencia en la energía libre entre el péptido libre y ligado (133-141).

Tanto los métodos de predicción basados en estructura, como los basados en secuencia, ofrecen la promesa de predecir la unión CMH-péptido, reduciendo el costo de la verificación experimental de tal proceso “en húmedo”. Sin embargo, el enfoque basado en inteligencia artificial, se ha desarrollado más rápidamente que el enfoque estructural, produciendo resultados prometedores (especialmente para moléculas de clase I), superiores hasta ahora a los métodos estructurales, pero con resultados dependientes del set de datos (cantidad y calidad) usado para su entrenamiento. Desarrollar un enfoque metodológico basado únicamente en las propiedades estructurales inferidas de la secuencia, resulta lo más adecuado en el caso de *Aotus*, en donde no se han desarrollado los medios necesarios para hacer ensayos de unión CMH-péptido.

Estudio de la interacción CMH-péptido usando métodos cuánticos

El estudio de la interacción entre CMH-péptido usando métodos de química teórica computacional, ha sido una de las líneas de la investigación en la FIDIC, en donde hemos apostado por el análisis de estos sistemas desde la química cuántica, usando métodos *ab initio* (142-148). Esta aproximación, se ha centrado en la comprensión de los mecanismos de interacción entre receptor-ligando (centrándose principalmente en el análisis de los residuos de los bolsillos de unión), usando propiedades electrostáticas como los momentos multipolares, potencial electrostático y análisis de la función de onda, para identificar los orbitales que contribuyen a la unión CMH-péptido. Como resultado, se han identificado los residuos clave en la interacción CMH-péptido, la descripción del paisaje electrostático, así como la importancia relativa de cada bolsillo en el proceso de unión y la estimación de los perfiles de unión de aminoácidos por bolsillo. Estos hallazgo reproducen las tendencias experimentales observadas (132, 149), demostrando la plausibilidad y el poder descriptivo de esta aproximación.

A pesar de la solidez de este enfoque, el costo computacional (hardware *vs.* tiempo) que impone el estudio de macromoléculas desde la mecánica cuántica usando enfoques *ab initio*, limita el tamaño de los sistemas a estudiar. El desarrollo en la última década de métodos semi-empíricos, estrategias de procesamiento paralelo y técnicas de fragmentación, han permitido solucionar este problema, haciendo posible analizar proteínas completas en tiempos razonables (150, 151).

Así, hemos implementado los métodos semi-empíricos PM7 (152) y DFTB (153) para tratar proteínas; estos métodos semi-empíricos de química cuántica, se fundamentan en los mismos formalismos que los métodos *ab initio* (teoría de Hartree-Fock para el primero y teoría del funcional de la densidad en el segundo), pero hacen diversas simplificaciones y obtienen algunos parámetros de datos empíricos para compensar las imprecisiones derivadas de tales simplificaciones (151, 154). Adicionalmente, nuestro grupo ha implementado el método de fragmentación orbital molecular (FMO, *Fragment Molecular Orbital*) (155, 156) junto con PIEDA (*pair interaction decomposition analysis*) (157), que

dividen la molécula en fragmentos (en este caso, en la escala de aminoácidos) y hace cálculos de energía para cada uno de ellos, permitiendo obtener las propiedades del sistema global o de partes del mismo, por la combinación de las de los fragmentos. Como resultado, hemos sido capaces de simular el proceso de unión entre CMH-péptido, considerando la totalidad del sistema, con resultados que superan en precisión, los obtenidos por otros enfoques basados en estructura (158). Este enfoque permite la evaluación detallada de los efectos causados por substituciones de aminoácidos en proteínas, enfoque que puede ser aplicado tanto al análisis del CMH, como de otros sistemas (159).

Arquitectura del CMH y diseño de vacunas

En el desarrollo de vacunas basadas en péptidos, la FIDIC ha implementado una metodología fundamentada en la modificación de péptidos derivados de regiones conservadas de las proteínas de los parásitos, que resultan ser críticas en múltiples funciones biológicas, incluyendo el proceso de invasión a las células hospederas (HABPs, *high activity binding peptides*) (24, 160). La modificación de tales péptidos, obedece a principios de substitución, que involucran propiedades fisicoquímicas (como masa, volumen y polaridad) y estructurales, como la distancia entre los residuos de anclaje al CMH (161), orientación de las cadenas laterales (162), y su estructura secundaria (163); que en ultimas, producen cambios que modifican la afinidad del péptido al CMH, desencadenando una respuesta inmune contra estos sectores, que de otra forma, son inmunológicamente silentes (24).

En particular, el ajuste al CMH-DR tiene especial relevancia en el desarrollo de una vacuna contra la malaria, dado que la inmunidad al parásito es principalmente controlada por esta molécula (164-166), no solo en humanos, sino en otras especies (110, 111). Nuestros estudios han demostrado la similitud a nivel de polimorfismo, presiones selectivas y correlación con actividad inmune, entre el CMH-DR de humanos y *Aotus* (57, 60, 64, 161, 166, 167).

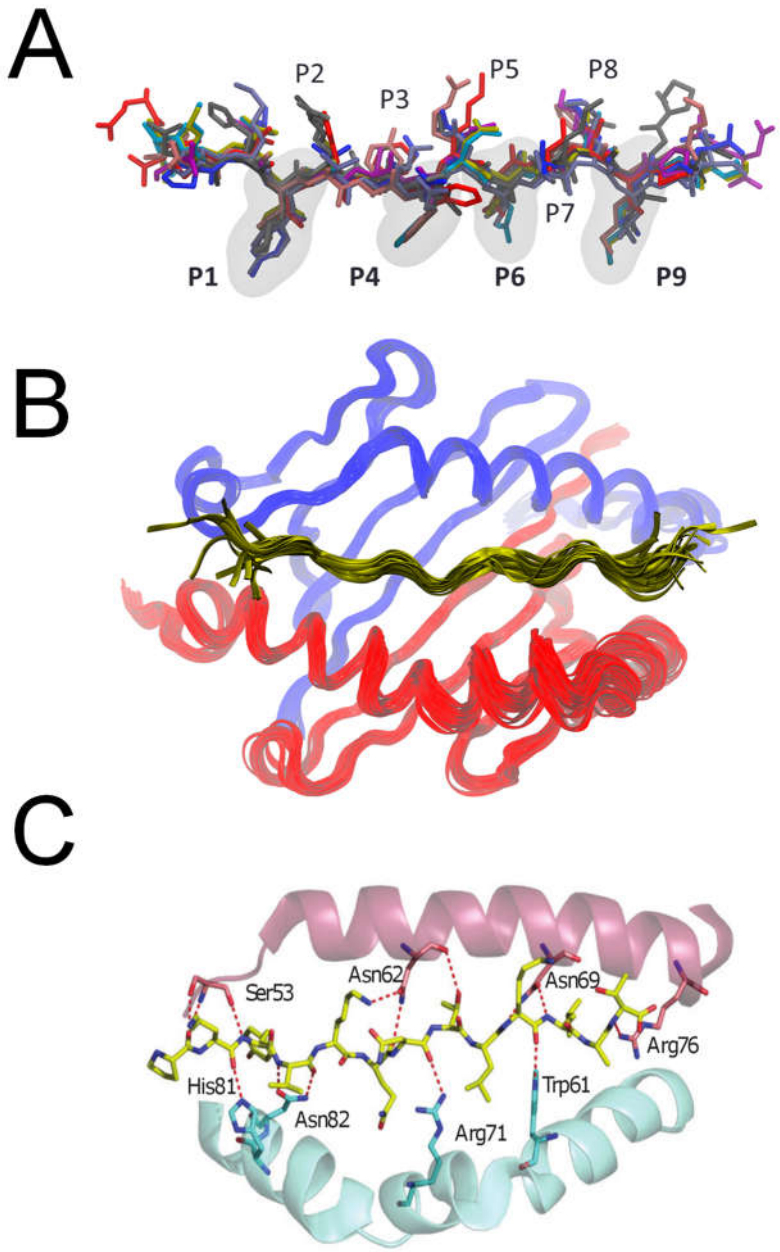


Figura 3. Persistencia en la estructura secundaria y red de enlaces de hidrógeno en el CMH-DR. A. *Diez péptidos unidos al CMH-DR de humanos y de murinos. Nótese la notable conservación en la estructura secundaria y en la orientación de las cadenas laterales. En negrilla, las posiciones que se anclan a los bolsillos 1, 4, 6 y 9.* ***B.*** *Conformación de estructura secundaria de 50 complejos CMH-péptido, incluyendo moléculas de CMH-DR tanto de humano como de ratón. (A. y B. son gráficas propias).* ***C.*** *Vista desde arriba de la red de puentes de hidrogeno del péptido de hemaglutina con el HLA-DR1 (tomado de (2)).*

La mayor contribución a la unión entre péptido y el CMH, está dada por la contribución de un conjunto de enlaces de hidrógeno conservados que interactúan con el esqueleto del péptido. Esto implica que los péptidos unidos, poseen estructuras secundarias que son *variaciones alrededor de un mismo tema* (Figura 3). Las cadenas laterales de los aminoácidos que interaccionan con los bolsillos de unión, aportan una interacción especifica que modula la afinidad de la unión (103, 168-170). La estructura secundaria consenso de los péptidos que se unen al CMH, es denominada hélice de poliprolina (PPII) y junto a las hélices alfa y hojas beta, son las tres estructuras estables observadas en proteínas naturales. Esta estructura favorece los procesos de interacción proteína-proteína y es frecuente encontrarla en sitios de unión (163, 171).

Así, teniendo en cuenta la influencia de los variables estructurales, especialmente las tendencias de estructura secundaria en las interacciones CMH-péptido, para el diseño de péptidos de unión, es necesario el establecimiento de un instrumento que permita realizar substituciones, siguiendo un criterio de similitud estructural. Previamente, habíamos caracterizado los aminoácidos de acuerdo a propiedades no estructurales, lo que ha permitido el establecimiento de principios de substitución (144). Usando la información de estructuras cristalizadas disponibles y analizando sus distribuciones de Ramachandran, hemos establecido una medida cuantitativa de la similitud estructural de los aminoácidos, que puede mejorar el diseño péptidos con la capacidad de adoptar la configuración favorable para su unión al CMH, y que además muestra un hallazgo fundamental: las tendencias estructurales, junto con la masa, explican los patrones de substitución evolutivos de los aminoácidos (172).

Objetivos

Objetivo General

- Completar la caracterización del complejo mayor de histocompatibilidad de clase II (CMH-DPA, CMH-DRA y CMH-DRB) en las especies *Aotus nancymaae* y *Aotus vociferans*.

Objetivos Específicos

- Caracterizar las secuencias de los genes CMH-DPA, CMH-DRA, y CMH-DRB en *Aotus nancymaae* y *Aotus vociferans*.

- Realizar un análisis comparativo de la evolución de los genes CMH-DPA, CMH-DRA y CMH-DRB en el contexto de los primates.

- Estudiar los patrones y la naturaleza de las variaciones a nivel de proteína de las moléculas clásicas del complejo mayor de histocompatibilidad clase II CMH-DRA y CMH-DRB, modelando su estructura y perfiles de unión de péptidos con métodos computacionales.

Preámbulo a los capítulos

El hilo conductor de este trabajo, se centra en la resolución de problemas relacionados con el desarrollo de vacunas para uso en humanos, usando como modelo animal los monos del género *Aotus*. Así, este trabajo es la continuación de los esfuerzos realizados para caracterizar el sistema inmune de estos primates y establecer la magnitud de la similitud entre humanos y *Aotus*, representado la oportunidad de comprender los modos de evolución de estas moléculas. También es la continuación del desarrollo y aplicación de métodos para dilucidar los mecanismos involucrados en la unión CMH-péptido, usando un enfoque computacional.

En el caso del desarrollo de una vacuna contra la malaria por parte de la FIDIC, la metodología desarrollada se centra en el diseño de péptidos que deben unirse exitosamente al CMH como condición *sine qua non* para que ocurra una respuesta de protección exitosa.

Existen varios problemas a tener en cuenta en las estrategias de diseño de “péptidos a la medida” para el CMH:

- Polimorfismo (tanto en humanos como en *Aotus*).
- Tipos de sustitución de aminoácidos que deben hacerse, para garantizar el ajuste de los péptidos al CMH.
- Evaluación y análisis de la unión CMH-péptido.

Polimorfismo

El hecho de que las moléculas del CMH sean tan polimórficas, dificulta enormemente el diseño de péptidos, dado que, a diferencia de otros problemas de diseño molecular, los receptores no son únicos y, por lo tanto, el número de soluciones posibles se incrementa enormemente. Así, además de estimar la magnitud de este polimorfismo, es necesario diseñar estrategias para manejarlo.

En este trabajo, se completó la caracterización de los genes que codifican para el dominio alfa del CMH-DP (61) (capítulo 1) y CMH-DR, ambos mostrando un limitado polimorfismo. Por otra parte, el polimorfismo del CMH-DRB de *Aotus* es muy grande, encontrándose en la caracterización efectuada, no solamente nuevos alelos, sino nuevos linajes alélicos en estos primates (62) (capítulo 2). La caracterización experimental de este polimorfismo es un reto en si misma, por lo que se describió un microsatélite asociado al intrón 2 del CMH-DRB en *Aotus*, para evaluar su capacidad de discriminación de los distintos alelos del CMH-DRB, con resultados prometedores (62) (capítulo 2). Así, en cuanto al polimorfismo del CMH-DR en *Aotus*, encontramos que éste es similar al observado en humanos, con un CMH-DRA con un limitado polimorfismo, y un CMH-DRB muy polimórfico.

La modelación computacional de sistemas con miles de receptores y millones de péptidos es impensable. Con el fin reducir el número de moléculas para describir la arquitectura de los bolsillos de unión y poder hacer inferencias basadas en modelos computacionales, se siguió la estrategia de enfocarse únicamente en los residuos críticos en el proceso de unión definidos cristalográficamente. De esta forma, se generó un "diccionario de bolsillos" del CMH-DRB (Anexo 1), por medio del cual se puede reducir el número de sistemas a considerar de manera efectiva.

Por ejemplo, solamente 27 bolsillos 1 se encuentran en humanos y *Aotus*, representando dos de ellos del 91% en humanos y el 72% en *Aotus* (el dimorfismo V↔G en la posición 86, Figura A, anexo 1). Cada alelo del CMH-DRB puede ser descrito como la concatenación de distintos los distintos bolsillos para generar un "perfil de bolsillos" que permite reducir directamente el número de alelos a considerar para el diseño de péptidos. A manera de ilustración, para el linaje alélico HLA-DRB1*01, existen 130 alelos reportados, pero, dos perfiles caracterizan el 68% de los alelos descritos. Los perfiles se nombran de acuerdo a un "alelo prototipo" que es representante del conjunto de alelos que comparten el mismo perfil en un linaje alélico determinado. En el anexo 1, tablas 1 (HLA-DRB) y 2 (*Aotus* CMH-DRB), se muestran los perfiles que cubren al menos el 60% de los alelos estudiados para cada linaje.

Sobre cada perfil de bolsillo, se puede realizar el diseño de péptidos usando la información de unión experimental disponible y los principios metodológicos previamente descritos de sustitución de aminoácidos. Para evaluar la capacidad de unión de forma rápida y relativamente precisa, se optimizó el uso del algoritmo NetMHCIIpan 3 (122), usando un conjunto reducido de alelos prototipo, que cubren la mayoría de perfiles de bolsillo de todos los linajes alélicos humanos. Esta estrategia se ha implementado con éxito en el diseño de péptidos que inducen protección de largo término (161) (Anexo 2).

Adicionalmente, los perfiles de bolsillos para cada linaje alélico pueden ser usados para extrapolar el cubrimiento potencial de los péptidos diseñados sobre éstos en las poblaciones. Para ello, se realizó una minería de sobre base de datos AFND (*Allele Frequency Net Database*) (173). Una estimación de la frecuencia de los linajes alélicos del CMH-DRB en poblaciones humanas (Anexo 3), permite evaluar el cubrimiento potencial como el producto de la probabilidad de encontrar un determinado linaje alélico en una población, por la probabilidad del perfil de bolsillo en tal linaje. Este enfoque ha sido seguido para calcular el cubrimiento potencial de péptidos diseñados para unirse tanto a alelos humanos como de *Aotus* (64) (Capítulo 3). Adicionalmente, en este artículo, se explora el alcance de la similitud de los perfiles de bolsillos entre humanos y *Aotus* desde un punto de vista estructural y fisicoquímico.

Tipos de sustitución de aminoácidos

Siendo la tendencia a adquirir una conformación extendida (PP_{II}) necesaria para el ajuste de los péptidos al CMH, se propuso determinar una clasificación basada en las propiedades estructurales de los aminoácidos, analizando los patrones de estructura secundaria en proteínas biológicas, haciendo minería sobre la base de datos PGD (*Protein Geometry Database*) (174). Como resultado, se obtuvo una clasificación de aminoácidos acompañada de una medida cuantitativa de su similitud, que puede ser usada en el modelamiento y diseño de péptidos También se logró hacer un aporte inédito

en el entendimiento de los patrones de sustitución evolutivos en proteínas biológicas y su relación con la estructura secundaria (172) (Capítulo 4).

Evaluación y análisis de la unión CMH-péptido.

El uso de una estrategia optimizada para la estimación de la unión CMH-péptido usando el método basado en redes neurales NetMHCIIpan 3 (122), permite una evaluación rápida de esta interacción. Sin embargo, queda aún mucho espacio para innovar en este campo, usando aproximaciones más precisas y con la capacidad de brindar información de las fuerzas interactuantes en el proceso de unión CMH-péptido. Así, se ha implementado el uso de métodos cuánticos con resultados que sobrepasan la capacidad predictiva de los métodos disponibles (158) (Capítulo 5). Cabe anotar, que el uso de estas alternativas constituye una segunda línea metodológica, que permite profundizar en el análisis de las fuerzas interactuantes que moldean el proceso de unión y no son, por el momento, métodos de tamización. Sin embargo, la implementación de estas estrategias para el estudio de los efectos de sustituciones en sistemas proteicos resulta muy prometedora, dada su precisión y capacidad explicativa (159) (Anexo 4).

* * *

Capítulo 1. Characterisation and comparative analysis of MHC-DPA1 exon 2 in the owl monkey (*Aotus nancymaae*)

Suarez CF, Patarroyo MA, Patarroyo ME. Characterisation and comparative analysis of MHC-DPA1 exon 2 in the owl monkey (*Aotus nancymaae*). Gene. 2011;470(1-2):37-45.

La versión publicada del artículo puede ser consultada en:

http://www.sciencedirect.com/science/article/pii/S0378111910003823

Title: Characterisation and comparative analysis of MHC-DPA1 Exon 2 in the owl monkey (*Aotus nancymaae*)

Authors: Carlos F. Suárez M. [1, 2], Manuel A. Patarroyo [1, 2], Manuel E. Patarroyo [1, 3] ✉

Addresses and Affiliations: [1] Fundación Instituto de Inmunología de Colombia (FIDIC), Carrera 50 No. 26-20, Bogotá, Colombia. [2] Universidad del Rosario, Calle 63D No. 24-31, Bogotá, Colombia. [3] Universidad Nacional de Colombia, Carrera 45 No. 26-85 Bogotá, Colombia.

✉ Corresponding Author: Prof. Manuel Elkin Patarroyo.

E-mail: mepatarr@gmail.com

Fax: (57-1) 4815269

Telephone: (57-1) 4815219

Abstract: The *Aotus nacymaae* (owl monkey) is an important animal model in biomedical research, particularly for the pre-clinical evaluation of vaccine candidates against *Plasmodium falciparum* and *Plasmodium vivax,* which require a precisely typed major histocompatibility complex. The exon 2 from *Aotus nancymaae* MHC-DPA1 gene was characterised in order to infer its allelic diversity and evolutionary history. Aona-DPA1 shows no polymorphism, and is related to other primate DPA alleles (including Catarrhini and Platyrrhini); constituting an ancient trans-specific and strongly-supported lineage with different variability and selective patterns when compared to other primate-MHC-DPA1 lineages. *A. nancymaae* monkeys have thus a smaller MHC-DP polymorphism than MHC-DQ or MHC-DR.

Key words: Animal model; MHC class-II molecule; molecular evolution; new world monkeys; Platyrrhini.

Abbreviations: Major histocompatibility complex (MHC), antigen-presenting cells (APC), peptide binding region (PBR), new world monkeys (NWM), old world monkeys (OWM), neighbour joining (NJ), minimum-evolution (ME), maximum likelihood (ML), local rearrangements of tree topology around an edge (LRSH), parsimony (Pars), global rate minimum deformation method (GRMD), million years (MY), single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), random effects likelihood (REL), substitution per site per million years (Sub/S/MY), trans-specific polymorphism (TSP).

1. Introduction

Major histocompatibility complex (MHC) Class II molecules display peptides on the surface of antigen-presenting cells (APC) for subsequent recognition by T-cells, thereby performing a key defence role against pathogens. MHC Class II molecules are heterodimers assembled from an α and a β glycopeptide chains encoded by the MHC Class II A and B genes, respectively. Three main MHC Class II loci, named HLA-DR, -DQ, and -DP, encode functional antigen-presenting molecules in primates. Genetic polymorphism and diversifying selection tied to functional and structural restrictions are common characteristics of these main loci. Such polymorphism is mainly restricted to the second exon of MHC class II A and B genes, constituting the molecule's peptide binding region (PBR) (Klein et al., 1993b).

MHC-DP is an ancient locus, shared by divergent mammalian orders (Takahashi et al., 2000; Yuhki et al., 2003). However, its polymorphism and functionality vary. For example, MHC-DP acquires a pseudo-genic nature in felines, as also occurs in murinae (mouse-like rodents), even though MHC-DP is the most polymorphic MHC Class II locus in other rodents, such as the mole rat (*Spalax* genus) (Klein et al., 1993a; Yuhki et al., 2003; Kelley et al., 2005).

MHC-DP is the most centromeric locus within the primate MHC gene cluster region, being constituted by four genes: DPA1 and DPB1 genes and DPA2 and DPB2 pseudo-genes. This arrangement (position and number) is apparently the same in all primates and was established before the split between Platyrrhini and Catarrhini ~ 43 million years ago (MY) (Klein et al., 1993a; Steiper and Young, 2006).

MHC-DPA1 variability in primates varies amongst non-existent and low polymorphism whilst for MHC-DPB1, it fluctuates from moderate to high polymorphism (Otting and Bontrop, 1995; Slierendregt et al., 1995; Bontrop et al., 1999; Doxiadis et al., 2001). HLA-DPA1 exhibits low polymorphism in humans, where 28 alleles have been reported to date, compared to the 138 alleles described for HLA-DPB1 (Robinson, et al., 2003). In contrast, *Callithrix jacchus* (the common marmoset, a neo-tropical primate), has the MHC-DP region inactive, not expressing any MHC-DP molecule (Antunes et al., 1998). In spite of such low polymorphism, MHC-DPA1 can be important in modulating an immune response, since HLA-DPA1*0301 appears to be involved in the genetic susceptibility to *Schistosoma haematobium* and several chronic inflammatory diseases (May et al., 1998; Dai et al., 2010).

Previous studies have characterised *Aotus* MHC Class II genes and molecules: MHC DQA-DQB (Diaz et al., 2000), MHC-DRB1 (Niño-Vasquez et al., 2000; Suarez et al., 2006) and MHC-DPB1 (Diaz et al., 2002). These neo-tropical primates have been shown to be susceptible to various human infectious diseases (Lujan et al., 1986; Polotsky et al., 1994; Noya et al., 1998). They can develop human malaria, particularly *Plasmodium falciparum* (Gysin, 1988; Rodriguez et al., 1990; Collins, 1994) and *Plasmodium vivax* asexual/blood stage infections (Pico de Coana et al., 2003). This makes the owl monkey a highly valuable animal model for biomedical research. To complete this landmark, the study of MHC-DPA1 might play a key role in understanding the immune response against *Plasmodium* (Diaz et al., 2002) and contributes towards gaining a deeper knowledge about the immune system of owl monkeys. The exon 2 from *A. nancymaae* MHC-DPA1 gene was characterised to infer its allelic diversity, variability patterns, the amount

and kind of its variation, the type of changes involved, as well as the extent of natural selection and evolutionary relationships within the primate context.

2. Materials and Methods

2.1 Animals

Six *Aotus nancymaae* monkeys (4 males, 2 females) were randomly caught from different familiar groups in Lagos de Leticia and Atacuari River, two widely separated zones (80 Km) in the Colombian Amazon. The monkeys were captured with the authorisation of the official environmental authority of Colombia in this region, CORPOAMAZONIA, which granted the Fundación Instituto de Inmunología de Colombia (FIDIC) permission for the capture, study and scientific research with these primates in the Colombian Amazon (Resolutions #1966/2006 and 0028/2010 and previous authorisations beginning in 1982). This research has been performed following the guidelines approved by FIDIC's ethics committee. The studied animals have been always under the supervision of expert veterinarians and biologists, and after experimental procedures they are released back into the Amazon jungle in optimal health conditions in the presence of a representative from CORPOAMAZONIA.

2.2 RNA extraction, cDNA synthesis, PCR, cloning and sequencing

Leukocytes were obtained from six healthy *A. nancymaae* monkeys by density gradient separation of peripheral blood obtained by venous puncture. Total cellular RNA was isolated from peripheral blood mononuclear cells using the TRIzol one-step procedure (Invitrogen Life Technologies, CA, USA). Moloney murine leukaemia virus reverse transcriptase (Promega, Madison, WI, USA) was used for cDNA synthesis, according to the Manufacturer's instructions.

Two PCR of MHC DPA1 exon 2 were independently performed for each monkey; PCR primers used were GH98 (5'-CGCGGATCCTGTGTCAACTTATGCCGCG-3') and GH99 (5'-CTGGCTGCAGTGTGGTTGGAACGCTG-3') (Otting and Bontrop, 1995) at a final 0.8 μM concentration. The PCR mixture contained 1.5 μM MgCl2, 50 mM Tris (pH 8.3) and 2.5 U Taq DNA polymerase (Promega). Five microlitres of cDNA were added to each reaction for a 25 μl final volume. These reactions were heated to 95°C for 5 min and then amplified for 40 cycles as follows: denaturing for 30 s at 94°C, annealing for 1 min at 65°C and extension for 2 min at 68°C. A final extension cycle was run at 65°C for 1 min and 68°C for 5 min.

A WIZARD PCR Preps Purification kit (Promega) was used for purifying PCR products which were then ligated into pGEM T vector (Promega). MiniPreps Purification Kit (Mo Bio, Carlsbad, CA, USA) was used for isolating double-strand plasmid DNA. Three clones from each PCR were randomly chosen and sequenced using fluorescent dye-labelled dideoxy terminators (Applied Biosystems, Foster City, CA, USA) in an ABI Prism 310 genetic analyser (Applied Biosystems).

2.3 MHC-DPA1 sequences

64 exon 2 MHC-DPA1 gene sequences from 11 primates (suborder Anthropoidea) were used. Platyrrhini (New World monkeys – NWM): *Aotus nancymaae* –Owl Monkey- (Aona, one sequence, reported here) and *Saimiri sciureus* -squirrel monkey- (Sasc, three sequences); Catarrhini: Cercopithecoidea (Old World monkeys – OWM): *Macaca arctoides* -stump-tailed macaque (Maar, one sequence), *Macaca fascicularis* -crab-eating macaque- (Mafa, six sequences), *Macaca mulatta* -rhesus monkey- (Mamu, 17 sequences) and *Papio hamadryas* -hamadryas baboon- (Paha, one sequence); Hominoidea (humans and apes): *Homo sapiens* –human- (HLA, 25 sequences), *Pan troglodytes* –chimpanzee- (Patr, three sequences), *Gorilla gorilla* –gorilla- (Gogo, three sequences), *Pongo pygmaeus* –Bornean orangutan- (Popy, three sequences), and Pongo abelii -Sumatran orangutan- (Poab, one sequence).

The following are the GenBank accession numbers of the studied sequences: Aona-DPA1*01-AF529200, Gogo-DPA1*0401-AF026701, Gogo-DPA1*0402-AF026702, Gogo-DPA1-CU104655, HLA-DPA1*010302-AF074848, HLA-DPA1*010304-DQ274060, HLA-DPA1*0104-X78198, HLA-DPA1*0105-X96984, HLA-DPA1*010601-U87556, HLA-DPA1*010602-EU729350, HLA-DPA1*0107-AF076284, HLA-DPA1*0108-AF346471, HLA-DPA1*0109-AY650051, HLA-DPA1*0110-DQ274061, HLA-DPA1*020101-X78199, HLA-DPA1*020102-L31624, HLA-DPA1*020103-AF015295, HLA-DPA1*020104-AF074847, HLA-DPA1*020105-AF098794, HLA-DPA1*020106-AF165160, HLA-DPA1*020203-AF092049, HLA-DPA1*02021-X79475, HLA-DPA1*02022-X79476, HLA-DPA1*0203-Z48473, HLA-DPA1*0204-EU304462, HLA-DPA1*0301-M83908, HLA-DPA1*0302-AF013767, HLA-DPA1*0303-AY618553, HLA-DPA1*0401-L11643, Maar-DPA1*0201-

AF026703, Mafa-DPA1*0201-AF026704, Mafa-DPA1*0202-EF208806, Mafa-DPA1*0204-AM943632, Mafa-DPA1*0401-EF208808, Mafa-DPA1*0701-EF208809, Mafa-DPA1*0702-EF208810, Mamu-DPA1*0101-Z32411, Mamu-DPA1*0201-EF204945, Mamu-DPA1*0203-EF204950, Mamu-DPA1*0208-FJ544416, Mamu-DPA1*0401-FJ544417, Mamu-DPA1*0402-FJ544415, Mamu-DPA1*0403-GQ471885, Mamu-DPA1*0601-EF204949, Mamu-DPA1*0701-EF204946, Mamu-DPA1*0801-EU305663, Mamu-DPA1-AB219099, Mamu-DPA1-AB219100, Mamu-DPA1-AB219101, Mamu-DPA1-AB250754, Mamu-DPA1-AB250756, Mamu-DPA1-AB219102, Mamu-DPA1-AB250757, Paha-DPA1*0201-AF026706, Patr-DPA1*0201-AF026707, Patr-DPA1*0202-AF026693, Patr-DPA1*0301-AF026694, Poab-DPA1-AC207096, Popy-DPA1*0201-AF026695, Popy-DPA1*0202-AF026696, Popy-DPA1*0401-AF026697, Sasc-DPA1*0501-AF026698, Sasc-DPA1*0502-AF026699, Sasc-DPA1*0601-AF026700

2.4 Sequence analysis

Clustal X (Thompson et al., 1997) was used for aligning the MHC-DPA1 exon 2 sequences. The *A. nancymaae* sequence was included and an amino acid alignment was also performed. HLA-DRA1*010101 and HLA-DQA1*010101 were used as outgroups. The resulting alignment had a total of 189/63 nucleotide/amino acid positions (supplementary material 1 and 2).

GENEDOC (Nicholas, et al., 1997) was used for calculating the percent of identity (*ie*, equal positions between sequences) and similarity (*ie*, positions with conservative substitutions between sequences, in this case, assessed by the PAM 250 substitution matrix) in the considered alignments. Means and standard deviations of pairwise nucleotide and amino acid identity and similarity (this

last one for amino acid sequences only) inside each group of sequences were analytically calculated.

Each position's variation for MHC-DPA1 exon 2 amino acid aligned sequences was represented by using WebLogo (Crooks et al., 2004). All amino acids occupying each position were indicated, in which the height of every amino acid letter represented its relative frequency in that position. The logo also allowed conservative and non-conservative substitutions for each position to be determined, where the variation in an amino acid symbol's colour indicated non-conservative changes and its preservation represented conservative changes based on PAM 250 substitution matrix groups (DENQH/ SAT/ KR/ FYW/ LIVM/ C/ G and P) (Dayhoff M et al., 1978).

2.5 Phylogenetic analysis

Neighbour Joining (NJ) and Minimum-evolution (ME) (Rzhetsky and Nei, 1993) trees were constructed using MEGA 4.0 (Tamura et al., 2007). Genetic distances were estimated by using Kimura 2-parameter (Kimura 1980), Log-Det (Tamura and Kumar, 2002) and Maximum Composite Likelihood (Tamura et al., 2004) substitution models for nucleotide sequences and JTT (Jones et al., 1992) and Dayhoff (Schwarz and Dayhoff, 1979) substitution models for amino acid-deduced sequences. Bootstrap analysis (Hillis and Bull, 1993) and interior branch test (IBT) (Sitnikova, 1996), both with 10000 replicates, were used for assigning confidence levels to branch nodes. Nodes having bootstrap values greater than 70% were statistically significant, as well as internal branch test values greater than 95%.

Maximum likelihood (ML) (Felsenstein, 1981) trees were constructed using TREEFINDER (Jobb et al., 2004) and DNAML / PROTML included in the PHYLIP package (Felsenstein, 1989); Bootstrap analysis (Hillis and Bull, 1993), with 10000 replicates, was used for assigning confidence levels to branch nodes. Genetic distances for TREEFINDER were calculated by using the estimated model from data following AICc criteria, in this case, HKY (Hasegawa et al., 1985) substitution model for nucleotide sequences and JTT (Jones et al., 1992) substitution model for amino acid sequences. Bootstrap analysis (Hillis and Bull, 1993) and local rearrangements of tree topology around an edge (LRSH) (Shimodaira and Hasegawa, 1999), both having 10000 replicates, were used for assigning confidence levels to branch nodes. Nodes having LRSH values greater than 95% were considered statistically significant.

Parsimony (Pars) (Felsenstein, 1983) trees were constructed using MEGA 4.0 (Tamura et al., 2007) and DNAPARS, both included in the PHYLIP package (Felsenstein, 1989). Bootstrap analysis (Hillis and Bull, 1993), with 10000 replicates, was used for assigning confidence levels to branch nodes.

A Bayesian approach was also used for inferring phylogenetic relationships using Mr. Bayes (Ronquist and Huelsenbeck, 2003). Default settings for the GTR model with gamma-distributed rate variation across sites and a proportion of invariable sites for nucleotide sequences and a mixed model for amino acid sequences, were used. Two simultaneous Markov chain Monte Carlo analyses were performed using one cold and three heated chains (temperature set to default 0.2) for each analysis. Simulations were run for 15.000.000 generations with a tree being saved each 100^{th} generation. At approximately ten million generations for the nucleotide alignment and 11

million generations for the amino acid alignment, the standard deviation of split frequencies reached a <0.01 value, indicating that both analyses converged on similar trees. The last 25% generations were preserved as burn-in and generated a consensus tree. Nodes having posterior probability values of 85 to 89 were considered to have low statistical support, 90 to 94 to have moderate support and nodes greater than 95 to be highly supported (Huelsenbeck and Ronquist, 2001).

2.6 Tree Calibration

Global Rate Minimum Deformation method (GRMD), implemented in TREEFINDER software (Jobb et al., 2004), was used to estimate the evolutionary rates of DPA groups deduced from the Bayesian tree (calculated in MrBayes for nucleotide sequences). As calibration points the divergence time amongst: Catarrhini – Platyrrhini 42.9 million years (MY) (36.1–51.1 MY) Platyrrhini - Platyrrhini, 21.0 (19.15-22.05 MY), Catarrhini - Catarrhini, 30.5 MY (26.9–36.4 MY), Hominoidea - Hominoidea, 18.3 MY (16.3–20.8 MY), *Homo* - *Pan* 6.6 MY (6 - 7 MY), *M. mulatta* - *M. fascicularis* 0.9 MY, were used (Goodman et al., 1998; Opazo et al., 2006; Osada et al., 2008).

2.7. Natural selection analysis

Natural selection was detected using single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL) and random effects likelihood (REL) methods using HYPHY (Kosakovsky-Pond et al., 2005). These maximum likelihood-based methods estimated the rates of non-synonymous

and synonymous changes at each site in the sequence alignment and identified sites under positive or negative selection (Kosakovsky-Pond and Muse, 2005; Kosakovsky-Pond and Frost, 2005b). For SLAC and FEL methods, a p-value ≤ 0.1, whilst for REL, the Bayes factor of 50 were considered as significant. The algorithms are available on the Datamonkey Web (Kosakovsky-Pond and Frost, 2005a; Poon et al., 2009). Also MEGA 4.0 software was used for calculating synonymous and non-synonymous substitutions and associated variance rates (assessed by the bootstrap method with 1,000 replicates) by Nei–Gojobori's method (Nei and Gojobori, 1986).

2.8. 3D representations.

Positions under variation/selection were represented in a 3D model of each Pocket (including adjacent residues within a range of 5Å) for DPA, from crystallized DPA1 complex (PDB 3LQZ from DPA1*0103 - DPB1*0201) (Dai, et al., 2010) using VMD 1.87 (Humphrey et al., 1996).

3. Results

3.1. MHC Aona-DPA1 sequence

The MHC-DPA1 exon 2 from six *A. nancymaae* monkeys was amplified by RT-PCR. Amplification products had a 189 bp size, corresponding to exon 2 positions 34 -222 (12 - 74 in α domain). 36 clones were sequenced yielding an identical sequence. Analysed sequences, including the Aona-DPA1 sequence, are shown in supplementary material 1 (Exon 2) and supplementary material 2 (α domain).

3.2 Evolutionary analysis of Aona-DPA1 exon 2

Independently of the tree construction method (Bayesian, Parsimony, NJ, ME or ML) or the substitution model assumed, MHC-DPA1 exon 2 sequences analysed clustered into similar groups. For sake of simplicity, five MHC-DPA1 groups were defined (Fig. 1): Group one, supported by a high posterior probability value and LRSH value, formed by alleles DPA1*05 and DPA1*07 from all Antropoidea groups, including the *A. nancymaae* DPA sequence. MHC Aona-DPA1 was clearly included in MHC-DPA1*05 lineage, having high statistically supported values in all phylogenetic methods used. Group two, supported by LRSH, formed mainly by DPA1*01 and DPA1*03 alleles from Catarrhini groups, but mainly conformed by human sequences. Group three, formed mainly by HLA-DPA1*02 sequences and supported by LRSH. Group four contains sequences from all Antropoidea groups distributed in four well supported subgroups: DPA1*04 from Hominoidea; DPA1*04 from Cercopithecidae; DPA1*06 from *S. sciureus* (Platirrhini) and *M. mulatta* (Cercopithecidae) and a subgroup conformed of two unnamed alleles from *M. mulattta*. Group five comprises Catarrhini sequences, primarily DPA1*02 sequences from Cercopithecidae and also DPA1*02 from *P. pygmaeus*. All group associations were relatively well conserved at protein-deduced sequence level, but some not so well supported (data not shown). Group 1, Group 2, Group 4 and Group 5 displayed a trans-species or convergent nature (Fig. 1). Moreover, some sequences were identical amongst species.

3.3 Evolutionary rate estimation in primate MHC-DPA1 exon 2

Aona-DPA1*01 exon 2 appears as one of the most divergent sequence amongst primate MHC-DPA1 sequences. A tree calibration was carried out in order to establish whether divergence corresponds to a high evolutionary rate or corresponds to a long time of existence (Fig. 1). For sake of simplicity, it has been assumed that the divergence times used as calibration points for MHC-DPA1 exon 2, correspond to the divergence time amongst species. As can be seen, primate MHC-DPA1 groups are divided in two tendencies: groups 1, 4 and 5 have similar rates, between 3.8 to 4.5 $\times 10^{-3}$ Sub/S/MY, evolving about 4 – 4.5 times slower than groups 2 and 3, which have a rate between 1.7 to 1.8 $\times 10^{-2}$ Sub/S/MY.

Within groups, the rates are often very variable. For example, in group 1, the subgroup MHC-DPA1*05 is formed by Sasc-DPA1*0501, *0502 and Aona-DPA1*01, with an evolutionary rate about 10 times slower than the rate of the subgroup formed by sequences from *Macaca* MHC-DPA1*07, being the rate of this group the highest observed in the analysis (7.3 $\times 10^{-3}$ vs. 7.9 $\times 10^{-2}$ Sub/S/MY). In contrast, Mafa-DPA1*0702 shows the lowest evolutionary rate observed (9.1 $\times 10^{-4}$ Sub/S/MY). This pattern of variability occurs within all groups considered. The different evolutionary constrains amongst alleles and species may be reflected by the rate variation within and amongst the studied groups.

3.4 Primate MHC-DPA1 exon 2 variability

Overall identity at nucleotide level was high, having a 94% mean (88% - 100% range) (Fig. 1). The logo of the deduced amino acid sequence of MHC-DPA1 α domain for the set of all analysed species which was remarkably conserved, having 95.1% mean similarity (88% - 100% range) and 90.7% identity (75% - 100% range) (Fig. 2). In general, most amino acid substitutions were non-conservative (24 from 33 variable positions) considering all sequences analysed (Antropoidea DPA, Fig. 2). Group 1 and Group 4 displayed a greater amount of sequence variability, followed by Group 5, whereas the remaining lineages showed a most conservative nature (at nucleotide and amino acid identity, and at amino acid similarity, Fig. 1 and Fig. 2).

Aona-DPA1 possessed distinctive nucleotide and amino-acid substitutions (16Q→H, 31I→M, 54V→F, 56V→A, 65A→I), being the most non-conservative (Fig. 2, supplementary material 1 and supplementary material 2). This characteristic highlights its divergent nature, shared with other NWM-DPA sequences.

Most variable positions at nucleotide and amino acid levels were grouped within positions 50 and 74 in amino acid sequence (150 and 222 in nucleotide sequence, red line in Fig. 2). This sector includes most of the residues involved in the interaction with peptide (Pocket residues) at PBR, as assigned by homology with HLA-DPA1*0103 (Dai et al., 2010). The region between amino acids 12 and 49 was more conserved (34 to 150 in nucleotide sequence, black line in Fig. 2).

The variability of MHC-DPA1 exon 2 is concentrated especially in Pocket residues and their neighbours. The most variable is the Pocket 9, followed by Pockets 6 and 1. Each group varies in a distinct way at Pocket level, *i.e.*, for both nucleotide and amino acids, group 4 is the most variable at Pocket 1, group 5 is the most variable at Pocket 6, group 1 is the most variable at Pocket 9 and group 3 only varies at Pocket 9 (Fig. 2). The substitution pattern at codon level in the PBR is concentrated in first and second positions in all groups, with exception of group 4, in which all codon positions exhibit equivalent variability. In the remaining sequences, substitutions in the third codon position prevail (supplementary material 3).

3.5 Natural selection in primate MHC-DPA1 exon 2

No complete correspondence between SLAC, FEL and REL selection tests for all the analysed positions was observed, only some common positions being detected (Fig. 2, supplementary material 4). MHC-DPA1 exon 2 for the set of all analysed groups displayed an accumulation of negatively selected positions (when present) in the less variable region (codons 12 to 49) and an accumulation of positively selected positions (when present) in the most variable region (codons 50 to 74). This pattern also occurred in most MHC-DPA1 groups (with the exception of Group 5). With a few exceptions, no common positions under selection occurred between groups. Pocket positions assigned by homology (Dai et al., 2010) and near residues showed greater variability, accumulation of non-synonymous and non-conservative substitutions and, in some cases, are under positive selection. Positions submitted to negative selection tend to occur with greater frequency in non-PBR sectors, as has been reported previously (Hughes and Yeager, 1998) (Fig. 2).

All Pockets suffer selective pressures, but not in the same way depending of the group. Group 1 and Group 2 show more positions under positive selection than the other lineages analysed (seven of seven in Group 2, seven of eight in Group 1), but in Group 1, those positions are more variable than the observed in Group 2, and comprise Pockets 6 and 9. On the other hand, Group 2 shows selective forces in Pockets 1, 6 and 9. Groups 3 and 4 show only positions under negative selection and in both groups these positions occur outside the Pockets. Group 5 Pockets 1 and 6 are under positive selection, having this group more positively selected positions in the less variable sector, and two of these positions occur at the Pockets 1 and 6. The result of the analysis of all sequences together (Anthropoidea DPA) shows the occurrence of positively selected positions (four of seven) interspersed amongst the negatively selected positions in the less variable sector of the molecule. Two of them occur in residues with potential contact with the peptide (13 and 42), and two in Pocket residues (23 and 31). No positions under negative selection pressure were observed in Pocket residues. The remaining positions under positive selection occur in the most variable region, at Pocket 9 (72 and 73) and at a neighbour residue (68).

A detailed analysis of positions under variation and/or selection for MHC-DPA1 has been performed (Fig. 3). Considering all positions under selection for groups and sequences together, all Pockets are under positive pressure (Fig. 3A, 3C, 3E), and that condition might be extended to some neighbour residues (considered as residues in direct contact with Pocket residues, at distances < 5Å, Fig. 3B, 3D, 3F). In Pocket 1, the definition of neighbour comprises the major number of residues amongst DPA1 Pockets, showing all possible tendencies. Six neighbour positions are under positive selection (in red) and two are variable (green) in positions that might involve peptide contact (Fig. 3B). Some of these positions are considered Pocket residues in DRA

locus (Stern, et al., 1994, Cardenas, et al., 2004). Pocket 1 is highly conserved amongst DPA1 sequences analysed (Fig. 2), showing the non-variant residue α32F (blue) and negatively selected neighbour positions (ice blue) (Fig. 3A and 3B, respectively). Despite its conservation, the subtle variation observed is a consequence of diversifying forces. Pocket 6 shows six variable neighbour positions above the Pocket residues, and a negatively selected position in the Pocket base alongside a positively selected position (Fig. 3D). Pocket 6 is also highly conserved, but less than Pocket 1, showing variable positions such as α31 (Fig. 2). Pocket 9 is the most variable amongst Pockets considered, showing variable positions as 69, 72 and 73, all under positive selection (Fig. 2, and Fig. 3E). Only one neighbour position shows high variability and positive pressure, (α68), and as in the previous cases, it is considered Pocket position for DRA loci.

Nei–Gojobori's method confirms these results, showing a significant accumulation of non-synonymous substitutions in PBR region for all sequences together; Groups 1, 2 and 5 show significant positive selection in PBR, group 3 shows a near neutral substitution pattern, and group 4 a non-significant accumulation of synonymous *vs*. non-synonymous substitutions. The non-PBR region displays the opposite behaviour, showing a significant accumulation of synonymous *vs.* non-synonymous substitutions for all sequences together; Groups 3 and 4 show significant negative selection in this region, whilst the remaining groups show the same tendency, but statistically unsupported. When analysing the entire sequence, all DPA1 groups and all DPA1 sequences together show accumulation of synonymous *vs.* non-synonymous substitutions, being significant only for group 3. In the less variable region (codons 12 to 49), all groups display the same pattern, and groups 3, 4 and 5 show a significant negative selection. In the most variable

region (codons 50 to 74) all groups show more non-synonymous than synonymous substitutions, but without statistical support (Supplementary material 3 and 5).

4. Discussion

The study of MHC-DPA1 represents an essential task in order to improve our understanding of both the MHC Class II and the immune system in the owl monkey. The central role of MHC Class II in defence against pathogens and its continuous struggle with changing pathogen strategies has caused a complex evolutionary scenario, in which multiple factors such as adaptive evolution by over-dominance, gene conversion, intra-allelic recombination and other recombination processes have shaped MHC polymorphism. The degree of polymorphism varies between MHC loci, as a result of different functional constrains (adaptive diversification or conservation), and stochastic processes (such as a bottleneck in population structure); these differences became relevant when comparing different immune systems (Hughes and Yeager, 1998; Bontrop et al., 1999; Yuhki et al., 2003).

A. nancymaae MHC Class II polymorphism and evolutionary relationships have been previously explored using similar strategies to those used in this article. In the case of MHC-DQ, Diaz *et al*. (Diaz et al., 2000) found 5 MHC-DQA1 (Aona-DQA1*27) alleles isolated from 3 owl monkeys, 14 MHC-DQB1 (Aona-DQB1*22 and Aona-DQB1*23) alleles and two Aona-DQB2 alleles isolated from 19 monkeys. Suarez *et al*. (Suarez et al., 2006) have found 98 alleles for MHC-DRB (split into 12 lineages), isolated from 86 owl monkeys and Diaz *et al*. (Diaz et al., 2002), reported 3 alleles for MHC-DPB1, isolated from 7 owl monkeys.

This work reports one Aona-DPA1 sequence isolated from 6 owl monkeys, suggesting that Aona-DPA1 may display a limited or non-existent polymorphism. Aona-DPA1 constitutes a divergent sequence located in one of the most variable groups within the context of primate MHC-DPA1. Despite the MHC-DPA1*05 lineage support, internal similarity and identity were often lower than similarity and identity observed between Aona-DPA1 and other MHC-DPA1 sequences from different loci (not shown). However, they share exclusive substitutions (supplementary material 1 and Fig. 2). Such high variability is caused by the age of the group and it has also been associated with greater variable positions and non-conservative changes; when high positively selected positions are added up to these findings (Fig. 2), particular functional constrains might be inferred for the evolution of this group.

The most polymorphic locus in the owl monkey, as in humans and other primates, is thus MHC-DR. Such polymorphism is concentrated in MHC-DRB, whilst MHC-DRA is conserved in the studied primates. MHC-DRB is the only polymorphic gene in the common marmoset (*C. jacchus*) (Antunes et al., 1998; Bontrop et al., 1999).

The second most polymorphic MHC Class II locus varies between different species; it is MHC-DQ for the owl monkey and rhesus monkey (*M. mulatta*) whereas it is MHC-DP for humans. The least variable locus is MHC-DP for the owl monkey and rhesus monkey (Slierendregt et al., 1995) and in humans it is MHC-DQ (Bontrop et al., 1999; Robinson et al., 2003).

Although more sampling may be necessary, these results support *A. nancymaae* as having a smaller polymorphism in MHC-DP than in MHC-DQ or MHC-DR. These data establish differences between *Aotus* and *Callithrix*, denoting the different MHC class II restrictions and specialisation in new world monkeys, and in a global view, the different strategies used by each primate species regarding the specialisation and diversification of their MHC class II repertories.

The existence of trans-species polymorphism (TSP) has been well-established for several MHC loci in primates (Klein 1987; Klein et al., 1998) but it can be mimicked by molecular convergence phenomena, as established for exon 2 from DRB1, DQA, DQB and DPB MHC Class II genes (O'HUigin, 1995; Trtkova et al., 1995; Kriener et al., 2000a; Kriener et al., 2000b; Kriener et al., 2001). All the above shows that only trans-species polymorphism has been found within Anthropoidea infraorders such as Catarrhini and Platyrrhini. If the TSP occurs, its duration is greater in MHC-DPA1 than MCH-DPB1, as can be observed in *A. nancymaae* (Diaz et al., 2002) and in other primates (Otting and Bontrop, 1995).

Association between Aona-DPA1 and Sasc-DPA1*05 in a highly supported NWM clade (Fig. 1) becomes a trans-specific lineage. Other Catarrihini-exclusive trans-specific MHC-DPA1 lineages have been detected (Fig. 2). Group 1, formed by MHC-DPA1*05 and MHC-DPA1*07, includes Catarrihini sequences from *Pongo* and *Macaca*. This indicates the noticeable antiquity of this group, being the best supported clade in primate order. This group and MHC-DPA1*06 lineage (Group 4), show long evolutionary times, predating the divergence between Catarrihini and Platyrrini. The absence of Platyrrhini sequences in other groups might obey to the small sampling of MHC DPA1 in these primates. Interestingly, human, the best sampled primate, restricts most of

its allelic repertoire to two groups (2 and 3) with a high conservation, but also high evolutionary rate. The evolutionary significance of this apparent specialisation may be explained by the birth and death model (Takahashi et al., 2000; Piontkivska and Nei, 2003), or by the origin of sequences, frequently derived from expressed genes only. In Group 2, an almost human lineage (with the exception of Patr-DPA1*0301), the virtual identity amongst HLA-DPA1*05 and Mamu-DPA1*0101 may indicate the existence of an ancient TSP, or a molecular convergence. Interestingly, this lineage shows a high number of positively selected positions, indicating a strong process of diversifying selection within the human lineage. These results show the existence of MHC-DPA specific lineages in some primate clades, but also, long term lineages as Group 1 or MHC-DPA1*06 in Group 4. This emphasises the need for a greater sampling amongst primate species to better understand MHC-DPA evolution.

In spite of its low variation, MHC-DPA1 exon 2 displayed differential variability constrains along the sequence, exhibiting a conserved region (residues 12 - 49) in which synonymous substitutions and negatively selected positions prevailed, and a mostly variable region (residues 50 - 74) in which non-synonymous substitutions and positively selected positions predominated (Fig. 2). This observation may have functional relevance indicating compartmentalisation. As other MHC genes, the positive selection is focused on PBR positions, and negative selection occurs on non-PBR positions, however, all groups analysed show specific variation and selection patterns, *e.g*. Group 1 shows a relatively high sequence diversity, a slow evolutionary rate and predominance of diversifying selection; Group 4 also shows a relatively high diversity and slow evolutionary rate, but evidence of purifying selection, explained by the accumulation of substitutions in the third position of the codon that lead to accumulating synonymous substitutions. On the other hand,

Group 2 shows a relatively low diversity and a high evolutionary rate as Group 3, but Group 2 shows evidence of a diversifying selection whilst Group 3 displays evidence of a purifying selection. These differences amongst Groups also involve the Pockets themselves, being stressed to different selection patterns depending on the group. All the above suggests the existence of differences between the evolutionary restrictions modelling the peptide binding boundaries for each group analysed.

The detection of variation and selective constrains beyond the Pocket residues may have a functional importance. In some cases, those positions might be involved in peptide contact or in the modification of electrostatic properties of the Pocket by surrounding residues (Fig. 3). The visualisation of that "extended Pockets" suggests that the binding interactions described by crystallographic studies might be fuzzier, and the evolutionary analysis provides evidence of different binding capacities for non-crystallised alleles. These subtle residue variations might be functionally relevant, as has been described in other MHC contexts (Posch et al., 1995; Posch et al., 1996).

The above results led us to conclude that Aona-DPA1 shows a limited or non-existent polymorphism and is associated with Sasc-DPA1*05, forming a strongly-supported lineage with distinctive variability and selective patterns from the other primate-MHC-DPA1 lineages. Our results show differences in the evolutionary pattern of HLA-DPA, suggesting a recent but strong diversifying process in the human lineage. The groups delimited from our analyses possess a set of distinctive features at diversity and selection patterns, indicating several modes of evolution in primate MHC-DPA.

Acknowledgements

This work was funded by COLCIENCIAS; contract RC-140-2009. We would like to thank Monica Estupiñan for laboratory technical support in the obtaining of sequences and Gisselle Rivera for helping in the translation of this manuscript.

References

- Antunes, S.G., De Groot, N.G., Brok, H., Doxiadis, G., Menezes, A.A., Otting, N. and Bontrop, R.E., 1998. The common marmoset: a new world primate species with limited MHC class II variability. Proc Natl Acad Sci U S A. 95, 11745-11750.
- Bontrop, R.E., Otting, N., De Groot, N.G. and Doxiadis, G.G., 1999. Major histocompatibility complex class II polymorphisms in primates. Immunol Rev. 167, 339-350.
- Cardenas, C., Villaveces, J.L., Bohorquez, H., Llanos, E., Suarez, C., Obregon, M. and Patarroyo, M.E., 2004. Quantum chemical analysis explains hemagglutinin peptide-MHC Class II molecule HLA-DRbeta1*0101 interactions. Biochem Biophys Res Commun. 323, 1265-1277.
- Collins, W.E., 1994. The owl monkey as a model for malaria, in: W. K. Baer (Eds.), Aotus: the owl monkey. Academic Press pp. 245-258.
- Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188-1190.
- Dai, S., Murphy, G.A., Crawford, F., Mack, D.G., Falta, M.T., Marrack, P., Kappler, J.W. and Fontenot, A.P., 2010. Crystal structure of HLA-DP2 and implications for chronic beryllium disease. Proc Natl Acad Sci U S A. 107, 7425-7430.
- Dayhoff, M.O., Schwartz R.M., Orcutt, B., 1978. A model of evolutionary change in proteins, in: Dayhoff M. (Eds.), Atlas of protein sequence and structure. National Biomedical Research Foundation, pp. 345-352.
- Diaz, D., Daubenberger, C.A., Zalac, T., Rodriguez, R. and Patarroyo, M.E., 2002. Sequence and expression of MHC-DPB1 molecules of the New World monkey *Aotus nancymaae*, a primate model for *Plasmodium falciparum*. Immunogenetics. 54, 251-259.
- Diaz, D., Naegeli, M., Rodriguez, R., Nino-Vasquez, J.J., Moreno, A., Patarroyo, M.E., Pluschke, G. and Daubenberger, C.A., 2000. Sequence and diversity of MHC DQA and DQB genes of the owl monkey *Aotus nancymaae.* Immunogenetics. 51, 528-537.
- Doxiadis, G.G., Otting, N., De Groot, N.G. and Bontrop, R.E., 2001. Differential evolutionary MHC class II strategies in humans and rhesus macaques: relevance for biomedical studies. Immunol Rev. 183, 76-85.
- Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 17, 368-376.
- Felsenstein, J., 1983. Parsimony in Systematics: Biological and Statistical Ann. Rev. Ecol. Syst. 313-333.
- Felsenstein, J., 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 164-166.
- Goodman, M., Porter, C.A., Czelusniak, J., Page, S.L., Schneider, H., Shoshani, J., Gunnell, G. and Groves, C.P., 1998. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol. 9, 585-598.

- Gysin, J., 1988. Animal models: primates, in: Sherman I.W. (Eds.), Malaria: parasite biology, pathogenesis, and protection. ASM Press pp. 419-439.
- Hasegawa, M., Kishino, H. and Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 22, 160-174.
- Hillis, D.B. and Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol 182-192.
- Huelsenbeck, J.P. and Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17, 754-755.
- Hughes, A.L. and Yeager, M., 1998. Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet. 32, 415-435.
- Humphrey, W., Dalke, A. and Schulten, K., 1996. VMD: visual molecular dynamics. J Mol Graph. 14, 33-38, 27-38.
- Jobb, G., Von Haeseler, A. and Strimmer, K., 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol. 4, 18.
- Jones, D.T., Taylor, W.R. and Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8, 275-282.
- Kelley, J., Walter, L. and Trowsdale, J., 2005. Comparative genomics of major histocompatibility complexes. Immunogenetics. 56, 683-695.
- Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16, 111-120.
- Klein, J., 1987. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum Immunol. 19, 155-162.
- Klein, J., O'huigin, C., Figueroa, F., Mayer, W.E. and Klein, D., 1993a. Different modes of MHC evolution in primates. Mol Biol Evol. 10, 48-59.
- Klein, J., Satta, Y., O'huigin, C. and Takahata, N., 1993b. The molecular descent of the major histocompatibility complex. Annu Rev Immunol. 11, 269-295.
- Klein, J., Sato, A., Nagl, S. and O'huigín, C., 1998. Molecular Trans-Species Polymorphism Annual Review of Ecology and Systematics. 29, 1-21.
- Kosakovsky-Pond, S.K. and Muse, S.V., 2005. Site-to-site variation of synonymous substitution rates. Mol Biol Evol. 22, 2375-2385.
- Kosakovsky-Pond, S.L. and Frost, S.D., 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics. 21, 2531-2533.
- Kosakovsky-Pond, S.L. and Frost, S.D., 2005b. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 22, 1208-1222.
- Kosakovsky-Pond, S.L., Frost, S.D. and Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 21, 676-679.
- Kriener, K., O'huigin, C. and Klein, J., 2000a. Alu elements support independent origin of prosimian, platyrrhine, and catarrhine Mhc-DRB genes. Genome Res. 10, 634-643.
- Kriener, K., O'huigin, C., Tichy, H. and Klein, J., 2000b. Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. Immunogenetics. 51, 169-178.
- Kriener, K., O'huigin, C. and Klein, J., 2001. Independent origin of functional MHC class II genes in humans and New World monkeys. Hum Immunol. 62, 1-14.
- Lujan, R., Chapman, W.L., Jr., Hanson, W.L. and Dennis, V.A., 1986. *Leishmania braziliensis*: development of primary and satellite lesions in the experimentally infected owl monkey, *Aotus trivirgatus*. Exp Parasitol. 61, 348-358.
- May, J., Kremsner, P.G., Milovanovic, D., Schnittger, L., Loliger, C.C., Bienzle, U. and Meyer, C.G., 1998. HLA-DP control of human *Schistosoma haematobium* infection. Am J Trop Med Hyg. 59, 302-306.
- Nei, M. and Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 3, 418-426.

- Nicholas, K.B., Nicholas, H.B. and Deerfield, D.W., 1997. Genedoc. Analysis and visualization of genetic variation. EMBNEW NEWS 4, 14.
- Niño-Vasquez, J.J., Vogel, D., Rodriguez, R., Moreno, A., Patarroyo, M.E., Pluschke, G. and Daubenberger, C.A., 2000. Sequence and diversity of DRB genes of *Aotus nancymaae*, a primate model for human malaria parasites. Immunogenetics. 51, 219-230.
- Noya, O., Gonzalez-Rico, S., Rodriguez, R., Arrechedera, H., Patarroyo, M.E. and Alarcon De Noya, B., 1998. *Schistosoma mansoni* infection in owl monkeys (*Aotus nancymai*): evidence for the early elimination of adult worms. Acta Trop. 70, 257-267.
- O'huigin, C., 1995. Quantifying the degree of convergence in primate Mhc-DRB genes. Immunol Rev. 143, 123-140.
- Opazo, J.C., Wildman, D.E., Prychitko, T., Johnson, R.M. and Goodman, M., 2006. Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). Mol Phylogenet Evol. 40, 274-280.
- Osada, N., Hashimoto, K., Kameoka, Y., Hirata, M., Tanuma, R., Uno, Y., Inoue, I., Hida, M., Suzuki, Y., Sugano, S., Terao, K., Kusuda, J. and Takahashi, I., 2008. Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between M. fascicularis and M. mulatta. BMC Genomics. 9, 90.
- Otting, N. and Bontrop, R.E., 1995. Evolution of the major histocompatibility complex DPA1 locus in primates. Hum Immunol. 42, 184-187.
- Pico De Coana, Y., Rodriguez, J., Guerrero, E., Barrero, C., Rodriguez, R., Mendoza, M. and Patarroyo, M.A., 2003. A highly infective *Plasmodium vivax* strain adapted to *Aotus* monkeys: quantitative haematological and molecular determinations useful for P. vivax malaria vaccine development. Vaccine. 21, 3930-3937.
- Piontkivska, H. and Nei, M., 2003. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. Mol Biol Evol. 20, 601-609.
- Polotsky, Y.E., Vassell, R.A., Binn, L.N. and Asher, L.V., 1994. Immunohistochemical detection of cytokines in tissues of Aotus monkeys infected with hepatitis A virus. Ann N Y Acad Sci. 730, 318-321.
- Poon, A.F., Frost, S.D. and Pond, S.L., 2009. Detecting signatures of selection from DNA sequences using Datamonkey. Methods Mol Biol. 537, 163-183.
- Posch, P.E., Araujo, H.A., Creswell, K., Praud, C., Johnson, A.H. and Hurley, C.K., 1995. Microvariation creates significant functional differences in the DR3 molecules. Hum Immunol. 42, 61-71.
- Posch, P.E., Hurley, C.K., Geluk, A. and Ottenhoff, T.H., 1996. The impact of DR3 microvariation on peptide binding: the combinations of specific DR beta residues critical to binding differ for different peptides. Hum Immunol. 49, 96-105.
- Robinson, J., Waller, M.J., Parham, P., De Groot, N., Bontrop, R., Kennedy, L.J., Stoehr, P. and Marsh, S.G., 2003. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res. 31, 311-314.
- Rodriguez, R., Moreno, A., Guzman, F., Calvo, M. and Patarroyo, M.E., 1990. Studies in owl monkeys leading to the development of a synthetic vaccine against the asexual blood stages of *Plasmodium falciparum*. Am J Trop Med Hyg. 43, 339-354.
- Ronquist, F. and Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19, 1572-1574.
- Rzhetsky, A. and Nei, M., 1993. Theoretical foundation of the minimum-evolution method of phylogenetic inference. Mol Biol Evol. 10, 1073-1095.
- Schwarz, R. and Dayhoff, M., 1979. Matrices for detecting distant relationships, in: Dayhoff, M. (Eds.), Atlas of protein sequences. National Biomedical Research Foundation, pp. 353 - 358.
- Shimodaira, H. and Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16, 1114-1116.

- Sitnikova, T., 1996. Bootstrap method of interior-branch test for phylogenetic trees. Mol Biol Evol. 13, 605-611.
- Slierendregt, B.L., Otting, N., Kenter, M. and Bontrop, R.E., 1995. Allelic diversity at the Mhc-DP locus in rhesus macaques (*Macaca mulatta*). Immunogenetics. 41, 29-37.
- Steiper, M.E. and Young, N.M., 2006. Primate molecular divergence dates. Mol Phylogenet Evol. 41, 384-394.
- Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L. and Wiley, D.C., 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature. 368, 215-221.
- Suarez, C.F., Patarroyo, M.E., Trujillo, E., Estupiñan, M., Baquero, J.E., Parra, C. and Rodriguez, R., 2006. Owl monkey MHC-DRB exon 2 reveals high similarity with several HLA-DRB lineages. Immunogenetics. 58, 542-558.
- Takahashi, K., Rooney, A.P. and Nei, M., 2000. Origins and divergence times of mammalian class II MHC gene clusters. J Hered. 91, 198-204.
- Tamura, K., Dudley, J., Nei, M. and Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24, 1596-1599.
- Tamura, K. and Kumar, S., 2002. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol Biol Evol. 19, 1727-1736.
- Tamura, K., Nei, M. and Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A. 101, 11030-11035.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876-4882.
- Trtkova, K., Mayer, W.E., O'huigin, C. and Klein, J., 1995. Mhc-DRB genes and the origin of New World monkeys. Mol Phylogenet Evol. 4, 408-419.
- Yuhki, N., Beck, T., Stephens, R.M., Nishigaki, Y., Newmann, K. and O'brien, S.J., 2003. Comparative genome organization of human, murine, and feline MHC class II region. Genome Res. 13, 1169-1179.

Figure legends

Fig. 1. Phylogenetic tree calculated using a Bayesian approach for primate MHC-DPA1 exon 2 sequences. Topologies obtained for Parsimony (Pars), maximum likelihood (ML) and minimum evolution (ME) are similar, and significant node support from these analyses are also shown (Bootstrap >70%; IBT >90%, LRSH >95%. See the code at the bottom of the figure). Allelic lineages are shown in different colours. Primate species divergence time in million years (MY) and mean substitution per site per million years (Sub/S/My) are shown for each group and subgroup, the average nucleotide identity obtained from all possible pairwise comparisons of exon 2 is also shown. The scale indicates 0.2 substitutions per site. See Materials and Methods section for species abbreviations and calculation details.

Fig. 2. MHC-DPA1 exon 2-deduced amino acid sequence logo. PAM 250 substitution matrix groups (DENQH (green), SAT (blue), KR (red), FYW (black), LIVM (purple), C (Gray), G (Brown) and P (Yellow)) are used to show conservative or non-conservative substitutions; colour changes imply non-conservative substitutions. Above each logo, sites under positive selection (combined results for SLAC, FEL and REL tests) are marked with +, whilst those under negative selection are shown with –; the remaining sites are unmarked and are considered neutral. Coloured numbers below each logo denote Pocket positions: fuchsia P1, orange P6, green P9, coloured arrows indicate other residues in contact with Pocket residues. At the right-hand side, the amino acid identity and amino acid similarity in primate MHC-DPA1 are shown. The average was obtained from all possible pairwise comparisons of deduced MHC-DPA1 protein sequences within each group. Similarity was calculated based on PAM 250 substitution matrix.

Fig. 3. Pockets of MHC-DPA1. Based in the PDB 3LQZ (DPA1*0103, DPB1*0201), the pockets and their neighbouring residues are shown. *A.* Pocket 1, *B.* Pocket 1 neighbour residues, *C.* Pocket 6, *D.* Pocket 6 neighbour residues, *E.* Pocket 9 and *F.* Pocket 9 neighbour residues. In red, positively selected residues, in ice blue, negatively selected residues, in blue, invariant residues, in green, variable residues, and in white, non-considered residues.

Figure 1.

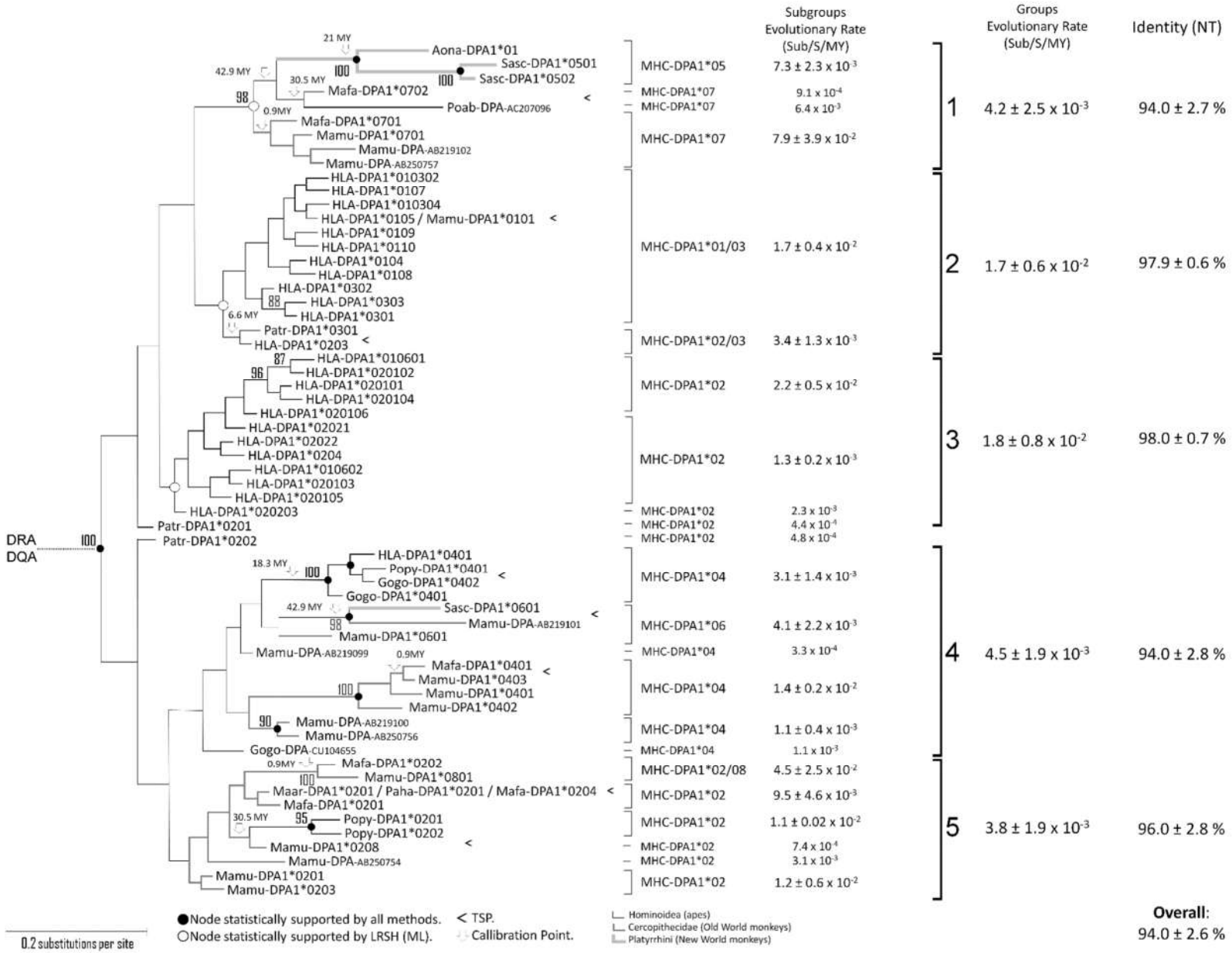

Subgroups Evolutionary Rate (Sub/S/MY)
Groups Evolutionary Rate (Sub/S/MY)
Identity (NT)
21 MY
Aona-DPA1*01
Sasc-DPA1*0501
42.9 MY
100
30.5 MY
100
Sasc-DPA1*0502
Mafa-DPA1*0702
98
Poab-DPA-AC207096
0.9MY
Mafa-DPA1*0701
Mamu-DPA1*0701
Mamu-DPA-AB219102
Mamu-DPA-AB250757
HLA-DPA1*010302
HLA-DPA1*0107
HLA-DPA1*010304
HLA-DPA1*0105 / Mamu-DPA1*0101
HLA-DPA1*0109
HLA-DPA1*0110
HLA-DPA1*0104
HLA-DPA1*0108
HLA-DPA1*0302
88
HLA-DPA1*0303
HLA-DPA1*0301
6.6 MY
Patr-DPA1*0301
HLA-DPA1*0203
87
HLA-DPA1*010601
96
HLA-DPA1*020102
HLA-DPA1*020101
HLA-DPA1*020104
HLA-DPA1*020106
HLA-DPA1*02021
HLA-DPA1*02022
HLA-DPA1*0204
HLA-DPA1*010602
HLA-DPA1*020103
HLA-DPA1*020105
HLA-DPA1*020203
Patr-DPA1*0201
DRA
DQA
100
Patr-DPA1*0202
18.3 MY
HLA-DPA1*0401
100
Popy-DPA1*0401
Gogo-DPA1*0402
Gogo-DPA1*0401
42.9 MY
Sasc-DPA1*0601
98
Mamu-DPA-AB219101
Mamu-DPA1*0601
Mamu-DPA-AB219099
0.9MY
Mafa-DPA1*0401
Mamu-DPA1*0403
100
Mamu-DPA1*0401
Mamu-DPA1*0402
90
Mamu-DPA-AB219100
Mamu-DPA-AB250756
Gogo-DPA-CU104655
0.9MY
Mafa-DPA1*0202
100
Mamu-DPA1*0801
Maar-DPA1*0201 / Paha-DPA1*0201 / Mafa-DPA1*0204
Mafa-DPA1*0201
30.5 MY
95
Popy-DPA1*0201
Popy-DPA1*0202
Mamu-DPA1*0208
Mamu-DPA-AB250754
Mamu-DPA1*0201
Mamu-DPA1*0203
MHC-DPA1*05 7.3 ± 2.3 x 10-3
MHC-DPA1*07 9.1 x 10-4
MHC-DPA1*07 6.4 x 10-3
MHC-DPA1*07 7.9 ± 3.9 x 10-2
MHC-DPA1*01/03 1.7 ± 0.4 x 10-2
MHC-DPA1*02/03 3.4 ± 1.3 x 10-3
MHC-DPA1*02 2.2 ± 0.5 x 10-2
MHC-DPA1*02 1.3 ± 0.2 x 10-3
MHC-DPA1*02 2.3 x 10-3
MHC-DPA1*02 4.4 x 10-4
MHC-DPA1*02 4.8 x 10-4
MHC-DPA1*04 3.1 ± 1.4 x 10-3
MHC-DPA1*06 4.1 ± 2.2 x 10-3
MHC-DPA1*04 3.3 x 10-4
MHC-DPA1*04 1.4 ± 0.2 x 10-2
MHC-DPA1*04 1.1 ± 0.4 x 10-3
MHC-DPA1*04 1.1 x 10-3
MHC-DPA1*02/08 4.5 ± 2.5 x 10-2
MHC-DPA1*02 9.5 ± 4.6 x 10-3
MHC-DPA1*02 1.1 ± 0.02 x 10-2
MHC-DPA1*02 7.4 x 10-4
MHC-DPA1*02 3.1 x 10-3
MHC-DPA1*02 1.2 ± 0.6 x 10-2
1 4.2 ± 2.5 x 10-3 94.0 ± 2.7 %
2 1.7 ± 0.6 x 10-2 97.9 ± 0.6 %
3 1.8 ± 0.8 x 10-2 98.0 ± 0.7 %
4 4.5 ± 1.9 x 10-3 94.0 ± 2.8 %
5 3.8 ± 1.9 x 10-3 96.0 ± 2.8 %
Overall: 94.0 ± 2.6 %
Node statistically supported by all methods.
Node statistically supported by LRSH (ML).
< TSP.
Callibration Point.
0.2 substitutions per site
Hominoidea (apes)
Cercopithecidae (Old World monkeys)
Platyrrhini (New World monkeys)

Figure 2.

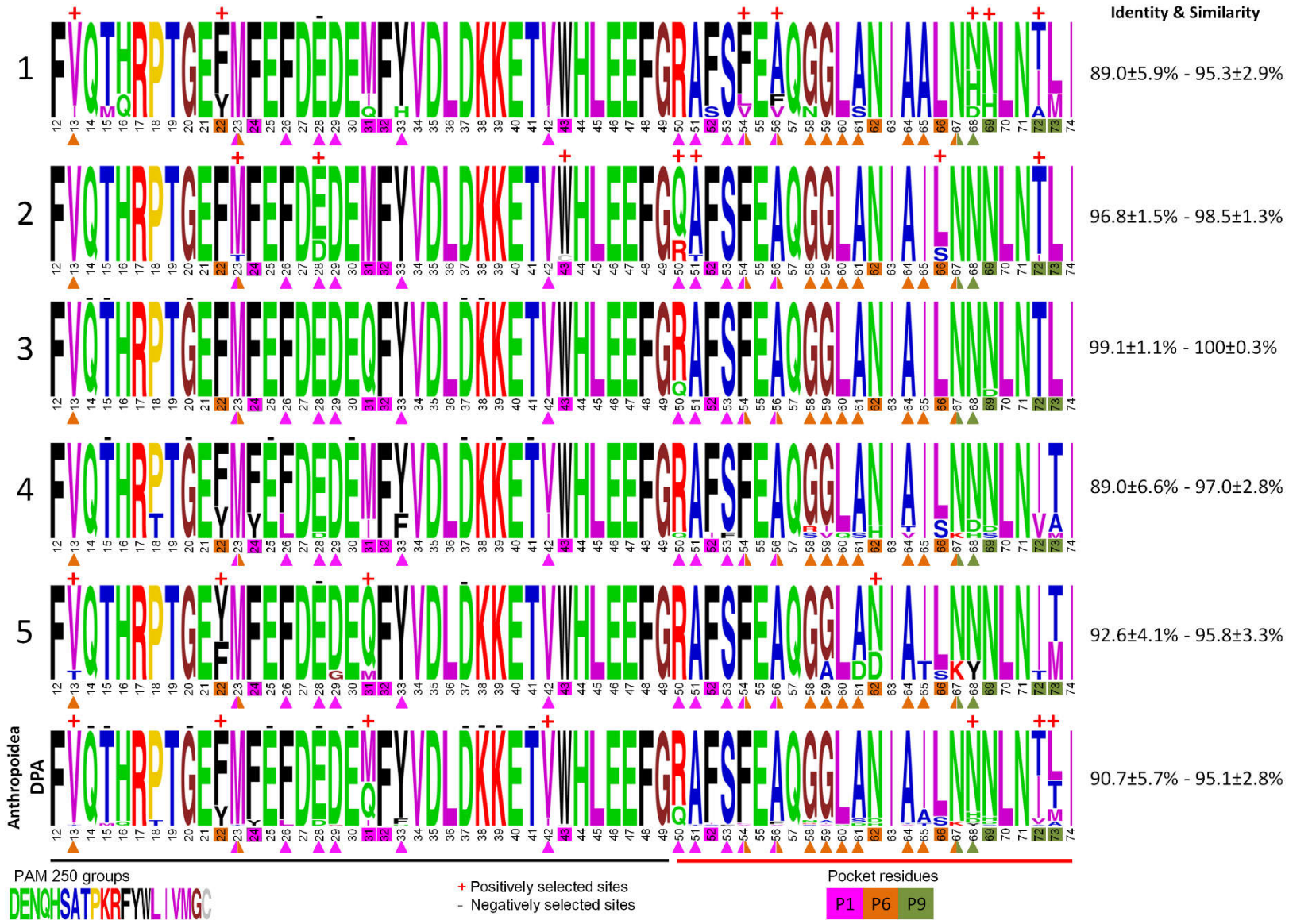
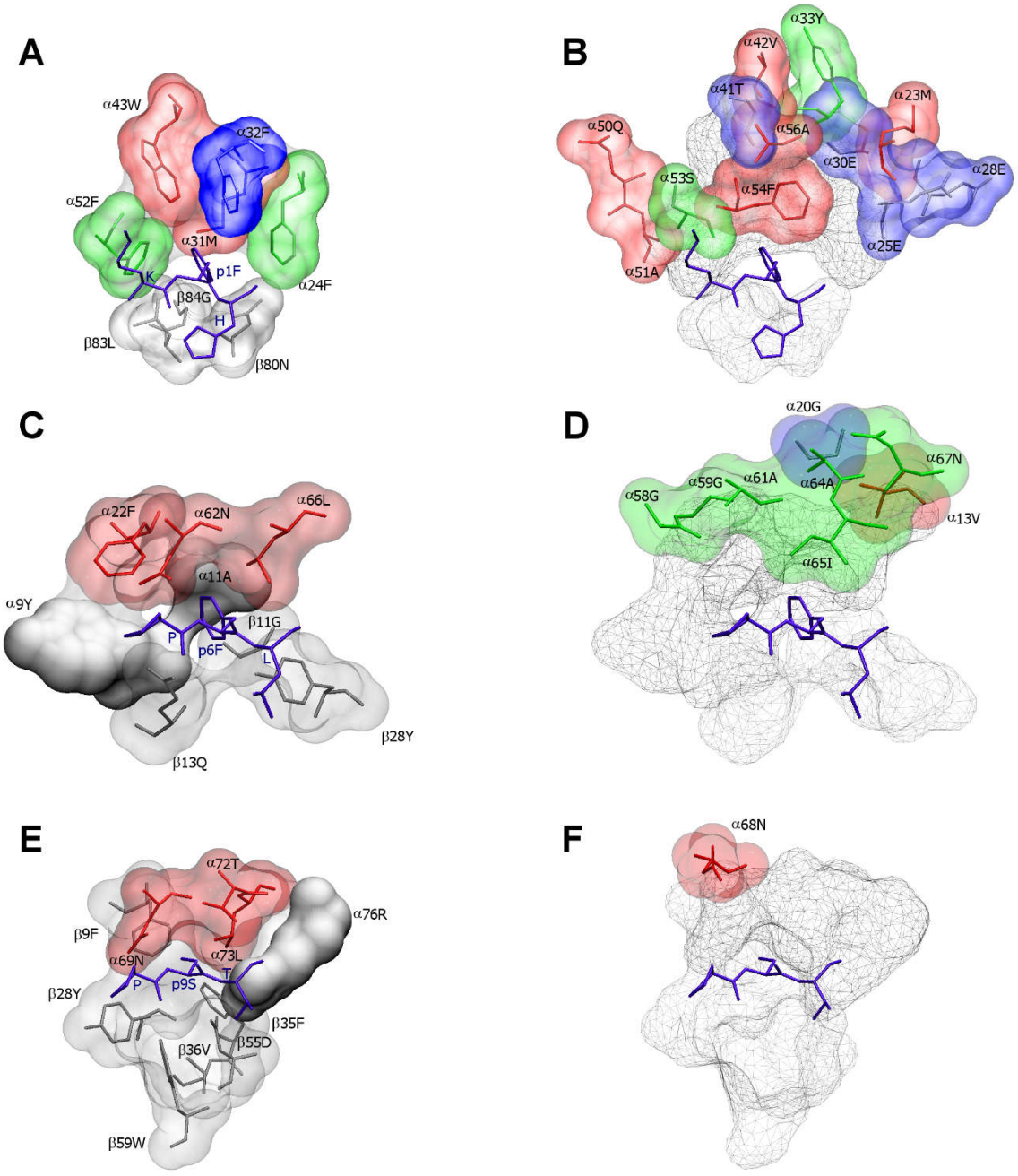

Identity & Similarity
1
89.0±5.9% - 95.3±2.9%
2
96.8±1.5% - 98.5±1.3%
3
99.1±1.1% - 100±0.3%
4
89.0±6.6% - 97.0±2.8%
5
92.6±4.1% - 95.8±3.3%
Anthropoidea DPA
90.7±5.7% - 95.1±2.8%
PAM 250 groups
DENQHSATPKRFYWLIVMGC
+ Positively selected sites
- Negatively selected sites
Pocket residues
P1
P6
P9

Figure 3.


A
α43W
α32F
α52F
α31M
p1F
K
α24F
β84G
H
β83L
β80N
B
α33Y
α42V
α41T
α23M
α50Q
α56A
α30E
α28E
α53S
α54F
α25E
α51A
C
α22F
α62N
α66L
α11A
α9Y
β11G
P
p6F
L
β28Y
β13Q
D
α20G
α67N
α59G
α61A
α64A
α58G
α13V
α65I
E
α72T
α76R
β9F
α69N
α73L
T
P
p9S
β28Y
β35F
β55D
β36V
β59W
F
α68N

Supplementary Material

1. ***Exon 2 nucleotide sequence alignment of MHC-DPA1 alleles from 11 primates.*** Position is indicated by the top numbers and asterisks (*) symbolise 10 base intervals. A dot (.) denotes identity with regards to Aona-DPA1 sequence. GenBank Accession numbers appear after each sequence name.

```
                                34        *         *         *         *         *         *         *         *         *         *       147
Aona-DPA1*01 - AF529200    : TTT GTA CAG ACG CAG AGA CCA ACA GGG GAG TTT ATG TTT GAG TTT GAT GAG GAT GAG ATA TTC TAC GTG GAT CTG GAC AAG AAG GAG ACC GTC TGG CAT CTG GAG GAG TTT GGC
Sasc-DPA1*0501 - AF026698  : ... ... ... .T. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ...
Sasc-DPA1*0502 - AF026699  : ... ... ... .T. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mafa-DPA1*0702 - EF208810  : ... ... ... ..A ..T ... ... ... ... ... .A. ... ... ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Poab-DPA - AC207096        : ... A.. ... ... ..T ... ..G ... ... ... .A. ... ... ... ... ... ... ... ... ..G ... C.T ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ...
Mafa-DPA1*0701 - EF208809  : ... ... ... ..A ..T ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA1*0701 - EF204946  : ... ... ... ..A ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA - AB219102        : ... ... ... ..A ..T ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA - AB250757        : ... ... ... ..A ..T ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*010302 - AF074848 : ... ... ... ... ..T ... ... ... ..A ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0107 - AF076284   : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*010304 - DQ274060 : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0105 - X96984     : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA1*0101 - Z32411    : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0109 - AY650051   : ... ... ... ... ..T ... ... ... ... ... ... ... .C. ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0110 - DQ274061   : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ...
HLA-DPA1*0104 - X78198     : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..C ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0108 - AF346471   : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..C ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0302 - AF013767   : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0303 - AY618553   : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..C ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0301 - M83908     : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Patr-DPA1*0301 - AF026694  : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0203 - Z48473     : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*010601 - U87556   : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ..T ..A ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*020102 - L31624   : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ..T ..A ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*020101 - X78199   : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ..T ..A ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*020104 - AF074847 : ... ... ..A ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ..T ..A ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*020106 - AF165160 : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*02021 - X79475    : ... ... ... ..C ..T ... ... ... ..A ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*02022 - X79476    : ... ... ... ..C ..T ... ... ... ..A ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0204 - EU304462   : ... ... ... ..C ..T ... ... ... ..A ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*010602 - EU729350 : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*020103 - AF015295 : ... ... ... ... ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*020105 - AF098794 : ... ... ... ... ..T ... ... ... ..A ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*020203 - AF092049 : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Patr-DPA1*0201 - AF026707  : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Patr-DPA1*0202 - AF026693  : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Gogo-DPA1 - CU104655       : ... ... ... ..C ..T ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ..G ... ..T ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ...
HLA-DPA1*0401 - L11643     : ... ... ... ... ..T ... A.. ... ..A ... ... ... ... ... ... ... ..T ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Popy-DPA1*0401 - AF026697  : ... ... ... ... ..T ... A.. ... ..A ... ... ... ... ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Gogo-DPA1*0402 - AF026702  : ... ... ... ... ..T ... A.. ... ..A ... ... ... ... ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Gogo-DPA1*0401 - AF026701  : ... ... ... ... ..T ... A.. ... ..A ... ... ... ... ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Sasc-DPA1*0601 - AF026700  : ... ... ... ..A ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA1 - AB219101       : ... ... ... ..A ..T ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA1*0601 - EF204949  : ... ... ... ..A ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA1 - AB219099       : ... ... ... ..A ..T ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mafa-DPA1*0401 - EF208808  : ... ... ... ... ..T ... ... ... ... ... ... ... ... .A. ... ..G ... ..A ... ..A ... ... .TT ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ...
Mamu-DPA1*0403 - GQ471885  : ... ... ... ... ..T ... ... ... ... ... ... ... .A. ... .A. ... ..G ... ..A ... ..A ... ... .TT ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ...
Mamu-DPA1*0401 - FJ544417  : ... ... ... ... ..T ... ... ... ... ... ... ... .A. ... .A. ... ..G ... ..A ... ..A ..G ... .TT ... ... ... ... ... ... ..A ... ... A.. ... ... ... ... ... ...
Mamu-DPA1*0402 - FJ544415  : ... ... ... ... ..T ... ... ... ... ... ... ... .A. ... .AC ... ..G ... ..A ... ..A ... ... .TT ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ... ...
Mamu-DPA1 - AB219100       : ... ... ... ..A ..T ... ... ... ... ... ... ... .A. ... ... ... ... ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA1 - AB250756       : ... ... ... ..A ..T ... ... ... ... ... ... ... .A. ... ... ... ... ... ... ..A ... ... ..G ... ..T ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ...
Maar-DPA1*0201 - AF026703  : ... ... ... ..A ..T ... ... ... ... ... ... ... .A. ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Paha-DPA1*0201 - AF026706  : ... ... ... ..A ..T ... ... ... ... ... ... ... .A. ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mafa-DPA1*0204 - AM943632  : ... ... ... ..A ..T ... ... ... ... ... ... ... .A. ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mafa-DPA1*0201 - AF026704  : ... ... ... ..A ..T ... ... ... ... ... ... ... .A. ... ... ... ... ... ... ..A .G. ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mafa-DPA1*0202 - EF208806  : ... ... ... ..A ..T ... ... ... ... ... ... ... .A. ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA1*0801 - EU305663  : ... ... ... ..A ..T ... ... ... ... ... ... ... .A. ... ... ... ... ... ... ..G ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA1*0208 - FJ544416  : ... ... ... ..A ..T ... ... ... ... ... ... ... .A. ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Popy-DPA1*0201 - AF026695  : ... ... ... ... ..T ... ..G ... ..A ... ... ... ... ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ...
Popy-DPA1*0202 - AF026696  : ... ... ... ... ..T ... ..G ... ... ... ... .A. ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA1 - AB250754       : ... AC. ... ..A ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ...
Mamu-DPA1*0201 - EF204945  : ... ... ... ..A ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Mamu-DPA1*0203 - EF204950  : ... ... ... ..A ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... CAG ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T
```

```
148 *           *           *           *           *           *           *           * 222
Aona-DPA1*01 - AF529200    : CGG GCC TTT TCC GTT GAG GTT CAG GGT GGG CTG GCT AAC ATT GCT GCA TTG AAC AAC AAC TTG AAT ATC CTG ATC
Sasc-DPA1*0501 - AF026698  : ..A ... .C. ... T.. ... T.. ... AA. ... ... T.. ... ... ... ... ... ... ... C.. ... ... ... A.. ...
Sasc-DPA1*0502 - AF026699  : ..A ... ... ... T.. ... T.. ... AA. ... ... T.. ... ... ... ... ... ... ... C.. ... ... ... A.. ...
Mafa-DPA1*0702 - EF208810  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... .C. T.. ...
Poab-DPA - AC207096        : ..A ... ... ... T.G ... .C. ... ..C ... ... ... ... ... ... ... ... ... G.. C.. ... ... GC. A.. ...
Mafa-DPA1*0701 - EF208809  : ..A ... ... ... T.. ... .C. ... ..A ... ... ... ... ... ... ... ... ... ... ... ... ... .C. T.. ...
Mamu-DPA1*0701 - EF204946  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... ... ... ... C.. ... ... ... .C. T.. ...
Mamu-DPA - AB219102        : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... ... ... ... C.. ... ... ... ... T.. ...
Mamu-DPA - AB250757        : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... ... ... ... C.. ... ... ... .C. T.. ...
HLA-DPA1*010302 - AF074848 : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*0107 - AF076284   : .AA A.. ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*010304 - DQ274060 : .AA ... ... ... T.. ... .C. ... ... ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*0105 - X96984     : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
Mamu-DPA1*0101 - Z32411    : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*0109 - AY650051   : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*0110 - DQ274061   : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*0104 - X78198     : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*0108 - AF346471   : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*0302 - AF013767   : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*0303 - AY618553   : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. .C. ... ... ... ... ... .C. T.. ...
HLA-DPA1*0301 - M83908     : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. .C. ... ... ... ... ... .C. T.. ...
Patr-DPA1*0301 - AF026694  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... T.. ...
HLA-DPA1*0203 - Z48473     : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*010601 - U87556   : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*020102 - L31624   : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*020101 - X78199   : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*020104 - AF074847 : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*020106 - AF165160 : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*02021 - X79475    : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*02022 - X79476    : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*0204 - EU304462   : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... G.. ... ... .C. T.. ...
HLA-DPA1*010602 - EU729350 : .AA ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*020103 - AF015295 : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*020105 - AF098794 : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
HLA-DPA1*020203 - AF092049 : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... .C. T.. ...
Patr-DPA1*0201 - AF026707  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... T.. ...
Patr-DPA1*0202 - AF026693  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... AC. ...
Gogo-DPA1 - CU104655       : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... AC. ...
HLA-DPA1*0401 - L11643     : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... GCT ...
Popy-DPA1*0401 - AF026697  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ..T ... ... ... GCT ...
Gogo-DPA1*0402 - AF026702  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... GCT ...
Gogo-DPA1*0401 - AF026701  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... ACT ...
Sasc-DPA1*0601 - AF026700  : ..A ... ... ... T.. ... .C. ... A.C ATC .A. ... C.. ... A.. AT. ... ..G G.. G.. ... ... ... AC. ...
Mamu-DPA1 - AB219101       : ..A ... ... .T. T.. ... .C. ... A.G .T. ... T.. C.. ... .T. AT. ... ..T G.. .G. ... ... ... A.. ...
Mamu-DPA1*0601 - EF204949  : .AA ... A.. ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... C.. ... ... ... ... AC. ...
Mamu-DPA1 - AB219099       : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... AC. ...
Mafa-DPA1*0401 - EF208808  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. .C. ... ... ... ... ... G.. AC. ...
Mamu-DPA1*0403 - GQ471885  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. .C. ... ... ... ... ... G.. AC. ...
Mamu-DPA1*0401 - FJ544417  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. .C. ... ... ... ... ... G.. AC. ...
Mamu-DPA1*0402 - FJ544415  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... G.. AC. ...
Mamu-DPA1 - AB219100       : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... ACC ...
Mamu-DPA1 - AB250756       : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... ACC ...
Maar-DPA1*0201 - AF026703  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... AC. ...
Paha-DPA1*0201 - AF026706  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... AC. ...
Mafa-DPA1*0204 - AM943632  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... AC. ...
Mafa-DPA1*0201 - AF026704  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... AC. ...
Mafa-DPA1*0202 - EF208806  : ..A ... ... ... T.. ... .C. ... ..C ... ... .A. G.. ... ... A.. ... ..G T.. ... ... ... ... AC. ...
Mamu-DPA1*0801 - EU305663  : ..A ... ... ... T.. ... .C. ... ..C ... ... .A. G.. ... ... A.. ... ..G T.. ... ... ... ... AC. ...
Mamu-DPA1*0208 - FJ544416  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... A.. ...
Popy-DPA1*0201 - AF026695  : ..A ... ... ... T.. ... .C. ... ..C .C. ... ... ... ... ... AT. ... ... ... ... ... ... ... A.. ...
Popy-DPA1*0202 - AF026696  : ..A ... ... ... T.. ... .C. ... ..C .C. ... ... G.. ... ... AT. ... ... ... ... ... ... ... A.. ...
Mamu-DPA1 - AB250754       : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. .C. ... ... ... ... ... .C. A.. ...
Mamu-DPA1*0201 - EF204945  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... AC. ...
Mamu-DPA1*0203 - EF204950  : ..A ... ... ... T.. ... .C. ... ..C ... ... ... ... ... ... AT. ... ... ... ... ... ... ... AC. ...
```

2. ***Alpha domain sequence alignment of MHC-DPA1 alleles from 11 primates.*** Position is indicated by the top numbers and asterisks (*) symbolise 10 amino acid intervals. A dot (.) denotes identity with regards to Aona-DPA1 sequence. GenBank Accession numbers appear after each sequence name.

```
                          12        *         *         *         *         *         *   74
Aona-DPA1*01 - AF529200    : FVQTQRPTGEFMFEFDEDEIFYVDLDKKETVWHLEEFGRAFSVEVQGGLANIAALNNNLNILI
Sasc-DPA1*0501 - AF026698  : ...M.............M.........I.........S.F.F.N..S.......H...M.
Sasc-DPA1*0502 - AF026699  : ...M.............M.....................F.F.N..S.......H...M.
Mafa-DPA1*0702 - EF208810  : ....H.....Y........M.....................F.A.............T..
Poab-DPA - AC207096        : .I..H.....Y........M.H...................L.A...........DH..AM.
Mafa-DPA1*0701 - EF208809  : ....H.............M.....................F.A.............T..
Mamu-DPA1*0701 - EF204946  : ....H.............M.....................F.A...........H...T..
Mamu-DPA - AB219102        : ....H.............Q.....................F.A...........H......
Mamu-DPA - AB250757        : ....H.............M.....................F.A...........H...T..
HLA-DPA1*010302 - AF074848 : ....H.............M...................Q...F.A........I......T..
HLA-DPA1*0107 - AF076284   : ....H.............M...................QT..F.A........I......T..
HLA-DPA1*010304 - DQ274060 : ....H.............M...................Q...F.A........I......T..
HLA-DPA1*0105 - X96984     : ....H.............M...................Q...F.A........I......T..
Mamu-DPA1*0101 - Z32411    : ....H.............M...................Q...F.A........I......T..
HLA-DPA1*0109 - AY650051   : ....H......T.......M...................Q...F.A........I......T..
HLA-DPA1*0110 - DQ274061   : ....H.............M...........C.......Q...F.A........I......T..
HLA-DPA1*0104 - X78198     : ....H...........D..M...................Q...F.A........I......T..
HLA-DPA1*0108 - AF346471   : ....H...........D..M.....................F.A........I......T..
HLA-DPA1*0302 - AF013767   : ....H.............M...................Q...F.A........I......T..
HLA-DPA1*0303 - AY618553   : ....H...........D..M...................Q...F.A........IS.....T..
HLA-DPA1*0301 - M83908     : ....H.............M...................Q...F.A........IS.....T..
Patr-DPA1*0301 - AF026694  : ....H.............M.....................F.A........I........
HLA-DPA1*0203 - Z48473     : ....H.............M.....................F.A........I......T..
HLA-DPA1*010601 - U87556   : ....H.............Q...................Q...F.A........I......T..
HLA-DPA1*020102 - L31624   : ....H.............Q.....................F.A........I......T..
HLA-DPA1*020101 - X78199   : ....H.............Q.....................F.A........I......T..
HLA-DPA1*020104 - AF074847 : ....H.............Q.....................F.A........I......T..
HLA-DPA1*020106 - AF165160 : ....H.............Q.....................F.A........I......T..
HLA-DPA1*02021 - X79475    : ....H.............Q.....................F.A........I......T..
HLA-DPA1*02022 - X79476    : ....H.............Q.....................F.A........I......T..
HLA-DPA1*0204 - EU304462   : ....H.............Q.....................F.A........I...D..T..
HLA-DPA1*010602 - EU729350 : ....H.............Q...................Q...F.A........I......T..
HLA-DPA1*020103 - AF015295 : ....H.............Q.....................F.A........I......T..
HLA-DPA1*020105 - AF098794 : ....H.............Q.....................F.A........I......T..
HLA-DPA1*020203 - AF092049 : ....H.............Q.....................F.A........I......T..
Patr-DPA1*0201 - AF026707  : ....H.............Q.....................F.A........I........
Patr-DPA1*0202 - AF026693  : ....H.............Q.....................F.A........I.......T.
Gogo-DPA1 - CU104655       : ....H.............M.....................F.A........I.......T.
HLA-DPA1*0401 - L11643     : ....H.T.........D..M.....................F.A........I.......A.
Popy-DPA1*0401 - AF026697  : ....H.T...........M.....................F.A........I.......A.
Gogo-DPA1*0402 - AF026702  : ....H.T...........M.....................F.A........I.......A.
Gogo-DPA1*0401 - AF026701  : ....H.T...........M.....................F.A........I.......T.
Sasc-DPA1*0601 - AF026700  : ....H.............M.....................F.A.SIQ.H.TI.KDD...T.
Mamu-DPA1 - AB219101       : ....H.............M.....................FF.A.RV.SH.VI..DS...M.
Mamu-DPA1*0601 - EF204949  : ....H.............M...................Q.I.F.A........I..H....T.
Mamu-DPA1 - AB219099       : ....H.............M.....................F.A........I.......T.
Mafa-DPA1*0401 - EF208808  : ....H.......Y.L......F........I.........F.A........IS.....VT.
Mamu-DPA1*0403 - GQ471885  : ....H.....Y.Y.L......F........I.........F.A........IS.....VT.
Mamu-DPA1*0401 - FJ544417  : ....H.....Y.Y.L....M.F........I.........F.A........IS.....VT.
Mamu-DPA1*0402 - FJ544415  : ....H.....Y.Y.L......F.....................F.A........I......VT.
Mamu-DPA1 - AB219100       : ....H.....Y........M.....................F.A........I.......T.
Mamu-DPA1 - AB250756       : ....H.....Y........M.........I...........F.A........I.......T.
Maar-DPA1*0201 - AF026703  : ....H.....Y........Q.....................F.A........I.......T.
Paha-DPA1*0201 - AF026706  : ....H.....Y........Q.....................F.A........I.......T.
Mafa-DPA1*0204 - AM943632  : ....H.....Y........Q.....................F.A........I.......T.
Mafa-DPA1*0201 - AF026704  : ....H.....Y......G.Q.....................F.A........I.......T.
Mafa-DPA1*0202 - EF208806  : ....H.....Y........Q.....................F.A....DD..T.KY....T.
Mamu-DPA1*0801 - EU305663  : ....H.....Y........M.....................F.A....DD..T.KY....T.
Mamu-DPA1*0208 - FJ544416  : ....H.....Y........Q.....................F.A........I.......M.
Popy-DPA1*0201 - AF026695  : ....H.............Q.....................F.A..A.....I.......M.
Popy-DPA1*0202 - AF026696  : ....H.....Y........Q.....................F.A..A..D..I.......M.
Mamu-DPA1 - AB250754       : .T..H.............Q.....................F.A........IS.....TM.
Mamu-DPA1*0201 - EF204945  : ....H.............Q.....................F.A........I.......T.
Mamu-DPA1*0203 - EF204950  : ....H.............Q.....................F.A........I.......T.
```

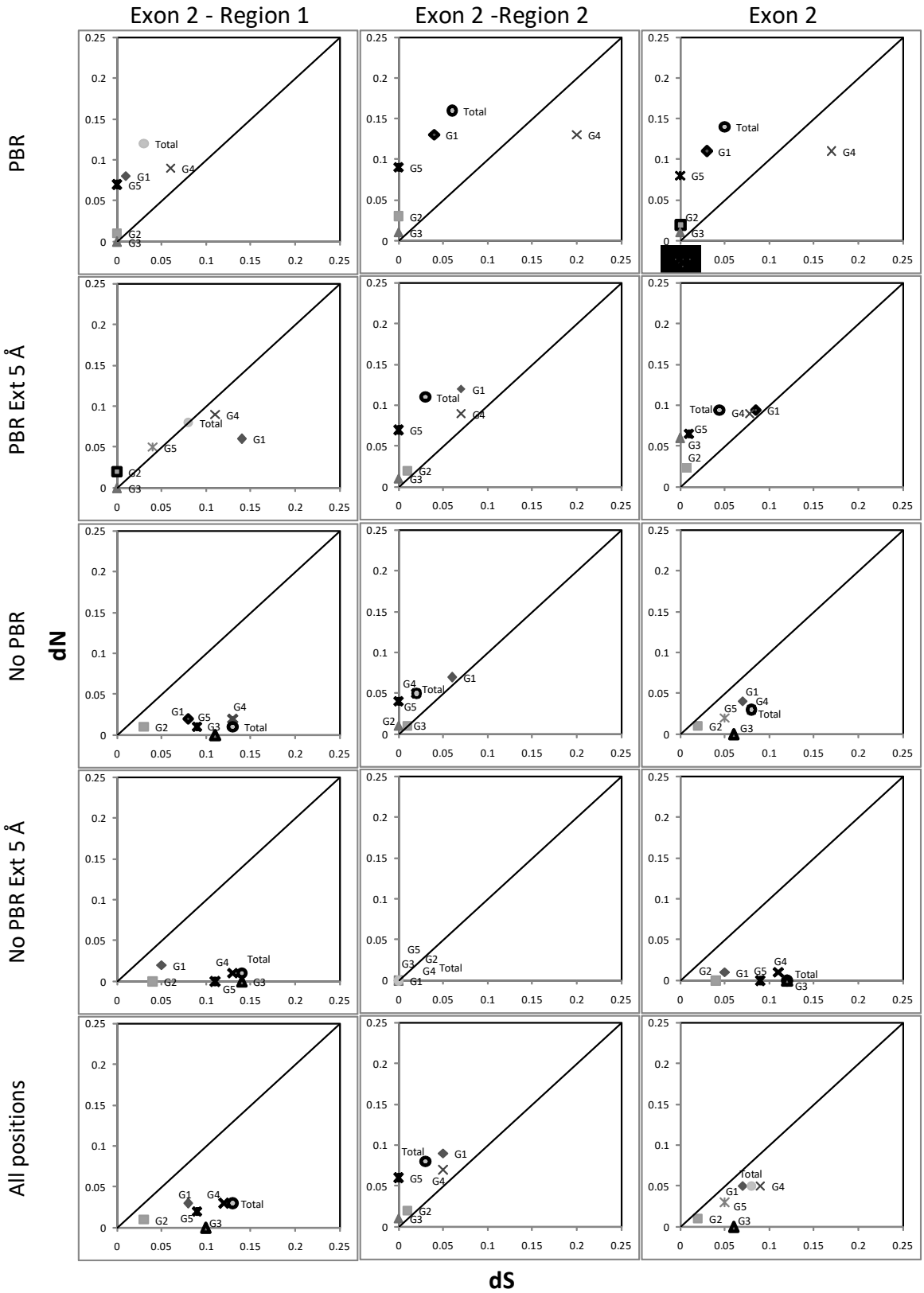
3. *Codon positions variability and Nei-Gojobori test.* P distance (pd) and standard error (S.E.) were calculated for the codon and each of its positions, for each group of analysed sequences in different sectors of the DPA exon 2. Synonymous (dS) and non synonymous substitutions (dN) and associated variance rates (assessed by the bootstrap method with 1,000 replicates) for every group of analysed sequences in each sector of exon 2 were calculated by Nei–Gojobori's method. Z values for statistically significant tests (significance levels of 5% ≥ 1.64) were marked in red for positive selection (dN > dS) and in gray for negative selection (dS > dN). Exon 2 region 1 comprises the less variable sector (codons 12 to 49) and exon 2 region 2 contains the most variable region (codons 50 to 74). Definition of PBR and extended PBR positions are according to figure 2.

			Exon2 - Region 1						Exon 2 - Region 2						Exon 2					
			G1	G2	G3	G4	G5	Total	G1	G2	G3	G4	G5	Total	G1	G2	G3	G4	G5	Total
PBR	Codon first position	pd	0.04	0.00	0.00	0.00	0.04	0.10	0.23	0.00	0.03	0.21	0.08	0.18	0.02	0.01	0.00	0.11	0.00	0.04
		S.E.	0.04	0.00	0.00	0.00	0.04	0.08	0.11	0.00	0.03	0.06	0.07	0.07	0.02	0.01	0.00	0.06	0.00	0.02
	Codon second position	pd	0.12	0.00	0.00	0.18	0.15	0.19	0.13	0.07	0.00	0.10	0.16	0.20	0.13	0.04	0.00	0.14	0.15	0.20
		S.E.	0.07	0.00	0.00	0.10	0.09	0.09	0.08	0.04	0.00	0.05	0.08	0.08	0.06	0.03	0.00	0.05	0.06	0.06
	Codon third position	pd	0.04	0.03	0.00	0.10	0.00	0.04	0.00	0.00	0.00	0.12	0.00	0.04	0.02	0.01	0.00	0.11	0.00	0.04
		S.E.	0.04	0.03	0.00	0.06	0.00	0.02	0.00	0.00	0.00	0.12	0.00	0.03	0.02	0.01	0.00	0.06	0.00	0.02
	Codon	pd	0.07	0.01	0.00	0.09	0.06	0.11	0.12	0.02	0.01	0.14	0.04	0.14	0.10	0.03	0.01	0.12	0.07	0.13
		S.E.	0.03	0.01	0.00	0.04	0.04	0.04	0.05	0.02	0.01	0.04	0.04	0.04	0.03	0.01	0.01	0.03	0.03	0.03
	Nei-Gojobori Test	dS	0.01	0.00	0.00	0.06	0.00	0.03	0.04	0.00	0.00	0.20	0.00	0.06	0.03	0.00	0.00	0.17	0.00	0.05
		dN	0.08	0.01	0.00	0.09	0.07	0.12	0.13	0.03	0.01	0.13	0.09	0.16	0.11	0.02	0.01	0.11	0.08	0.14
		Z-value	1.00	0.17	nc	0.43	**1.75**	1.00	**1.80**	1.50	1.00	0.64	**3.00**	**2.00**	**1.75**	**2.00**	1.00	0.67	**2.67**	**2.25**
PBR Ext 5 Å	Codon first position	pd	0.07	0.00	0.00	0.03	0.03	0.06	0.18	0.01	0.01	0.12	0.04	0.11	0.07	0.02	0.00	0.09	0.02	0.05
		pd S.E.	0.03	0.00	0.00	0.03	0.02	0.04	0.05	0.01	0.01	0.03	0.03	0.03	0.03	0.01	0.00	0.03	0.01	0.02
	Codon second position	pd	0.05	0.01	0.00	0.11	0.09	0.10	0.09	0.04	0.02	0.07	0.11	0.12	0.07	0.03	0.01	0.09	0.10	0.11
		pd S.E.	0.03	0.01	0.00	0.05	0.04	0.04	0.04	0.02	0.02	0.02	0.04	0.04	0.03	0.02	0.01	0.03	0.03	0.03
	Codon third position	pd	0.08	0.04	0.00	0.12	0.02	0.07	0.06	0.01	0.00	0.07	0.02	0.03	0.07	0.02	0.00	0.09	0.02	0.05
		pd S.E.	0.05	0.03	0.00	0.06	0.02	0.03	0.04	0.01	0.00	0.03	0.02	0.01	0.03	0.01	0.00	0.03	0.01	0.02
	Codon	pd	0.07	0.02	0.00	0.09	0.05	0.08	0.11	0.02	0.01	0.09	0.06	0.09	0.09	0.02	0.01	0.09	0.05	0.08
		pd S.E.	0.02	0.01	0.00	0.03	0.02	0.02	0.03	0.01	0.01	0.02	0.02	0.02	0.02	0.01	0.00	0.01	0.01	0.01
	Nei-Gojobori Test	dS	0.14	0.00	0.00	0.11	0.04	0.08	0.07	0.01	0.00	0.07	0.00	0.03	0.09	0.01	0.00	0.08	0.01	0.04
		dN	0.06	0.02	0.00	0.09	0.05	0.08	0.12	0.02	0.01	0.09	0.07	0.11	0.09	0.02	0.06	0.09	0.07	0.09
		Z-value	0.82	**2.00**	nc	0.20	0.40	0.00	1.50	0.50	1.00	0.75	**3.50**	**4.00**	**1.86**	1.17	1.20	0.34	**2.89**	**1.97**
No PBR	Codon first position	pd	0.02	0.00	0.00	0.03	0.01	0.01	0.11	0.01	0.00	0.05	0.02	0.05	0.05	0.00	0.00	0.04	0.01	0.03
		pd S.E.	0.01	0.00	0.00	0.02	0.01	0.01	0.04	0.01	0.00	0.02	0.02	0.02	0.02	0.00	0.00	0.01	0.01	0.01
	Codon second position	pd	0.01	0.00	0.00	0.01	0.01	0.01	0.05	0.02	0.02	0.04	0.06	0.06	0.02	0.01	0.01	0.02	0.03	0.03
		pd S.E.	0.01	0.00	0.00	0.01	0.01	0.00	0.03	0.02	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01
	Codon third position	pd	0.07	0.03	0.07	0.09	0.06	0.09	0.06	0.01	0.00	0.03	0.02	0.02	0.06	0.02	0.04	0.07	0.04	0.07
		pd S.E.	0.03	0.02	0.03	0.03	0.02	0.03	0.04	0.01	0.00	0.03	0.02	0.01	0.02	0.01	0.02	0.02	0.01	0.02
	Codon	pd	0.03	0.01	0.02	0.04	0.02	0.04	0.07	0.01	0.01	0.04	0.03	0.04	0.05	0.01	0.02	0.04	0.03	0.04
		pd S.E.	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	Nei-Gojobori Test	dS	0.08	0.03	0.11	0.13	0.09	0.13	0.06	0.01	0.00	0.02	0.00	0.02	0.07	0.02	0.06	0.08	0.05	0.08
		dN	0.02	0.01	0.00	0.02	0.01	0.01	0.07	0.01	0.01	0.05	0.04	0.05	0.04	0.01	0.00	0.03	0.02	0.03
		Z-value	**1.78**	1.02	**2.23**	**2.06**	**2.55**	**2.58**	0.46	0.07	1.03	**1.97**	**2.64**	**1.91**	1.23	0.95	**2.12**	**1.65**	1.59	**1.96**
No PBR Ext 5 Å	Codon first position	pd	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
		pd S.E.	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
	Codon second position	pd	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
		pd S.E.	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	Codon third position	pd	0.06	0.03	0.09	0.08	0.06	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.02	0.07	0.06	0.05	0.07
		pd S.E.	0.03	0.02	0.04	0.03	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.03	0.03	0.02	0.03
	Codon	pd	0.02	0.01	0.03	0.03	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.02	0.03	0.02	0.03
		pd S.E.	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01
	Nei-Gojobori Test	dS	0.05	0.04	0.14	0.13	0.11	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.04	0.12	0.11	0.09	0.12
		dN	0.02	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00
		Z-value	1.55	1.32	**2.25**	**2.25**	**2.85**	**2.60**	nc	nc	nc	nc	nc	nc	1.57	1.34	**2.23**	**2.28**	**2.77**	**2.60**
All positions	Codon first position	pd	0.02	0.00	0.00	0.02	0.01	0.02	0.14	0.01	0.01	0.09	0.03	0.08	0.07	0.00	0.00	0.05	0.02	0.05
		pd S.E.	0.01	0.00	0.00	0.02	0.01	0.01	0.04	0.01	0.01	0.03	0.02	0.03	0.02	0.00	0.00	0.01	0.01	0.01
	Codon second position	pd	0.03	0.00	0.00	0.03	0.03	0.03	0.07	0.03	0.01	0.06	0.08	0.09	0.04	0.02	0.00	0.04	0.05	0.06
		pd S.E.	0.01	0.00	0.00	0.02	0.02	0.01	0.03	0.02	0.01	0.02	0.03	0.03	0.02	0.01	0.00	0.01	0.02	0.02
	Codon third position	pd	0.06	0.03	0.06	0.09	0.05	0.08	0.04	0.01	0.00	0.05	0.01	0.03	0.06	0.02	0.04	0.08	0.04	0.06
		pd S.E.	0.02	0.02	0.03	0.03	0.02	0.02	0.03	0.01	0.00	0.03	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02
	Codon	pd	0.04	0.01	0.02	0.05	0.03	0.05	0.08	0.02	0.01	0.07	0.04	0.07	0.06	0.01	0.01	0.06	0.03	0.05
		pd S.E.	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01
	Nei-Gojobori Test	dS	0.08	0.03	0.10	0.12	0.09	0.13	0.05	0.01	0.00	0.05	0.00	0.03	0.07	0.02	0.06	0.09	0.05	0.08
		dN	0.03	0.01	0.00	0.03	0.02	0.03	0.09	0.02	0.01	0.07	0.06	0.08	0.05	0.01	0.00	0.05	0.03	0.05
		Z-value	1.46	0.98	**2.26**	**1.90**	**2.18**	**2.24**	1.04	0.61	1.41	0.49	**3.51**	**2.82**	0.60	0.71	**2.05**	1.51	0.80	1.19

4. Selected sites using SLAC, FEL and REL methods. For SLAC and FEL methods, a p-value ≤ 0.1 was considered as significant, and for REL, the Bayes factor of ≥ 50 was considered as significant. Significant positively selected codons has been marked in red, and significant negatively selected codons has been marked in gray.

	Codon	SLAC dN-dS	SLAC p-value	FEL dN-dS	FEL p-value	REL dN-dS	REL Bayes Factor
Group 1	13	4.381	0.701	2.008	0.249	2.182	**54.212**
	22	8.203	0.626	4.096	0.303	2.157	**72.289**
	28	-13.876	0.245	-15.503	**0.046**	-1.571	3.840
	54	8.467	0.539	3.719	0.318	2.158	**73.199**
	56	9.027	0.462	3.731	0.211	2.209	**109.947**
	68	11.177	0.559	4.610	0.213	2.183	**87.433**
	69	7.451	0.679	2.957	0.313	2.182	**86.489**
	72	13.811	0.296	4.443	0.156	2.216	**117.933**
Group 2	23	3.519	0.991	10.738	0.870	1.815	**320.525**
	28	8.987	0.746	27.692	0.404	1.840	**1.8E+05**
	43	4.416	0.996	-9.1E+04	0.953	1.816	**339.136**
	50	9.612	0.673	26.301	0.303	1.840	**1.7E+05**
	51	5.230	0.667	14.578	0.357	1.819	**381.336**
	66	5.258	0.742	15.439	0.521	1.819	**391.442**
	72	5.206	0.670	13.690	0.373	1.818	**369.870**
Group 3	14	-25.190	0.201	-193.283	**0.056**	-7.924	**5.6E+12**
	15	-25.171	0.111	-91.231	**0.052**	-7.931	**8.4E+09**
	20	-37.757	**0.037**	-246.682	**0.004**	-7.940	**2.4E+38**
	37	-16.672	0.252	-95.249	0.103	-7.924	**3.6E+12**
	38	-19.729	0.213	-122.547	**0.090**	-7.920	**4.4E+12**
Group 4	15	-5.586	**0.037**	-6.536	**0.007**	-2.575	**3.2E+14**
	20	-3.724	0.114	-6.080	**0.016**	-2.565	**3.5E+13**
	25	-3.080	0.216	-9.523	**0.042**	-2.525	**3.1E+03**
	30	-3.024	0.220	-5.553	**0.063**	-2.521	**1.1E+03**
	37	-2.990	0.208	-3.098	**0.098**	-2.536	**2.0E+03**
	39	-3.082	0.216	-10.204	**0.037**	-2.544	**4.9E+03**
	41	-1.862	0.333	-2.099	0.116	-2.575	**1.4E+08**
Group 5	13	8.832	0.453	8.159	0.140	0.979	**111.828**
	22	12.293	0.651	10.881	0.350	1.018	**1.2E+03**
	28	-23.357	0.143	-13.958	**0.069**	-0.586	2.143
	31	7.567	0.768	5.927	0.590	0.968	**88.879**
	37	-26.673	0.111	-20.159	**0.044**	-0.608	2.349
	62	6.687	0.790	6.074	0.412	0.969	**90.300**
Anthropoidea DPA	13	0.745	0.300	0.381	**0.070**	0.962	25.815
	14	-0.842	0.249	-0.765	**0.055**	-1.707	20.904
	15	-4.800	**0.000**	-2.530	**0.000**	-6.771	**2.5E+07**
	20	-2.991	**0.001**	-2.082	**0.000**	-6.988	**1.5E+06**
	22	1.506	0.191	0.873	**0.055**	2.693	**102.582**
	25	-1.617	**0.049**	-1.122	**0.009**	-3.151	**270.393**
	28	-3.201	**0.019**	-1.665	**0.022**	-5.871	**205.564**
	30	-0.837	0.213	-0.669	**0.054**	-1.517	24.624
	31	2.620	0.252	1.390	0.217	2.618	**66.889**
	37	-2.418	**0.009**	-1.005	**0.004**	-3.104	**1.9E+03**
	38	-0.838	0.213	-0.733	**0.043**	-1.680	34.413
	39	-0.841	0.212	-0.760	**0.042**	-1.735	34.964
	41	-0.997	0.111	-0.419	**0.029**	-1.066	**151.425**
	42	0.745	0.299	0.394	**0.084**	0.966	20.834
	68	1.259	0.250	0.659	**0.058**	2.027	31.837
	72	1.676	**0.078**	0.719	**0.025**	2.467	**74.320**
	73	1.363	0.209	0.047	0.954	0.197	**50.089**

5. ***Relationships between synonymous (dS) and non synonymous substitutions (dN) in the different sectors of the DPA exon 2***. Values above neutrality line (dS = dN) denotes accumulation of non-synonymous substitutions (positive selection pressure), values below neutrality line denotes accumulation of synonymous substitutions (negative selection pressure). Bold markers indicate a statistically significant Nei-Gogobori's test. Test significances, definition of exon 2 sectors, PBR and extended PBR positions are according to supplementary material 3 and figure 2.



Capítulo 2. Characterising a Microsatellite for DRB Typing in *Aotus vociferans* and *Aotus nancymaae*

López C, Suárez CF, Cadavid LF, Patarroyo ME, Patarroyo MA. Characterising a microsatellite for DRB typing in *Aotus vociferans* and *Aotus nancymaae* (Platyrrhini). PLoS One. 2014;9(5):e96973.

La versión publicada del artículo puede ser consultada en:

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0096973

OPEN ACCESS Freely available online

PLOS ONE

Characterising a Microsatellite for DRB Typing in *Aotus vociferans* and *Aotus nancymaae* (Platyrrhini)

Carolina López[1,2,3☯], Carlos F. Suárez[1,2☯], Luis F. Cadavid[4], Manuel E. Patarroyo[5], Manuel A. Patarroyo[1,2]*

1 Molecular Biology and Immunology Department, Fundación Instituto de Inmunología de Colombia (FIDIC), Bogotá, Cundinamarca, Colombia, **2** School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Cundinamarca, Colombia, **3** MSc Microbiology Programme, Instituto de Biotecnología (IBUN), Universidad Nacional de Colombia, Bogotá, Cundinamarca, Colombia, **4** Genetics Institute, Universidad Nacional de Colombia, Bogotá, Cundinamarca, Colombia, **5** School of Medicine, Universidad Nacional de Colombia, Bogotá, Cundinamarca, Colombia

Abstract

Non-human primates belonging to the *Aotus* genus have been shown to be excellent experimental models for evaluating drugs and vaccine candidates against malaria and other human diseases. The immune system of this animal model must be characterised to assess whether the results obtained here can be extrapolated to humans. Class I and II major histocompatibility complex (MHC) proteins are amongst the most important molecules involved in response to pathogens; in spite of this, the techniques available for genotyping these molecules are usually expensive and/or time-consuming. Previous studies have reported MHC-DRB class II gene typing by microsatellite in Old World primates and humans, showing that such technique provides a fast, reliable and effective alternative to the commonly used ones. Based on this information, a microsatellite present in MHC-DRB intron 2 and its evolutionary patterns were identified in two *Aotus* species (*A. vociferans* and *A. nancymaae*), as well as its potential for genotyping class II MHC-DRB in these primates.

Citation: López C, Suárez CF, Cadavid LF, Patarroyo ME, Patarroyo MA (2014) Characterising a Microsatellite for DRB Typing in *Aotus vociferans* and *Aotus nancymaae* (Platyrrhini). PLoS ONE 9(5): e96973. doi:10.1371/journal.pone.0096973

Editor: Roscoe Stanyon, University of Florence, Italy

Received October 17, 2013; **Accepted** April 14, 2014; **Published** May 12, 2014

Funding: This research was supported by the "Departamento Administrativo de Ciencia, Tecnología e Innovación (COLCIENCIAS)", contract RC#0309-2013. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mapatarr.fidic@gmail.com

☯ These authors contributed equally to this work.

Introduction

Using non-human primates in the field of biomedical research is useful for validating methodologies for diagnosing and treating diseases affecting human beings [1,2]. Monkeys from the *Aotus* genus are used for studying the main types of human malaria (*Plasmodium falciparum* and *Plasmodium vivax*), being suitable models due to their susceptibility to the infection, thereby facilitating the evaluation of vaccines and drugs for treating and controlling this disease. These primates have also been used for studying leishmaniasis, schistosomiasis, hepatitis, tuberculosis and various types of enteric infection [3–9].

Previous studies have shown that this animal model is similar to humans regarding immune system molecules, particularly concerning MHC class II and especially those corresponding to human HLA-DR. Such similarity enables evaluating the immune response to different pathogens and evaluating the potential of molecules which are candidates for a vaccine aimed at controlling diseases of importance for human health [10–12].

The high degree of polymorphism and allele diversity shown by MHC-DRB molecules in humans and other primates, as well as their importance in interaction with peptides so that they can be presented to the T-lymphocyte receptor, makes their typing relevant for evaluating an immune response to malaria and vaccines designed for controlling it [13]. MHC-DR variability is mainly concentrated in MHC-DRB exon 2 and to a lesser extent in MHC-DRA exon 2 [14], both regions encoding the peptide binding region (PBR). This sector mainly defines the alleles observed in vertebrates and is subject to diversifying selection and recombination, thereby modelling its variability [15–17]. Twelve allele lineages have been characterised for *Aotus* MHC class II DRB, having considerable similarity with human HLA-DRB lineages [12,18,19].

Precise typing of MHC genes implies using laborious and costly techniques due to their complex genomic organisation (usually into different haplotypes) and their individual (expressing different genes) and population variability (polymorphism) [13]. Current techniques would include restriction fragment length polymorphism (RFLP), single strand conformation polymorphism (SSCP), denaturing gradient gel electrophoresis (DGGE), reference strand-mediated conformational analysis (RSCA) and amplifying, cloning and sequencing fragments of interest, especially exon 2. The latter represents the most precise approach but does involve some disadvantages such as its high cost and the longer time involved in obtaining results. The other approaches offer results having variable agreement with the data obtained by sequencing [20–22].

In addition to the above, a microsatellite located at the start of intron 2 in humans, macaques and chimpanzees has been used for typing MHC-DRB [23,24]. Short tandem repeat (STR) polymorphism has been shown to be well-correlated with the diversity shown by exon 2. The microsatellite is basically presented as the repeat of $(GT)_x$ $(GA)_y$ dinucleotides, showing different degrees of complexity, according to the species being analysed [23].

Regarding HLA-DRB, the STR has been called D6S2878, being present in all HLA-DRB genes/pseudogenes, except HLA-DRB2, HLA-DRB8 and HLA-DRB9 where the first part of intron 2 is lost. It is highly polymorphic in composition and length and can specifically differentiate between HLA-DRB gene alleles [25]. This sector also exhibits polymorphism in *Macaca mulatta*, having high variability regarding length and sequence, thus allowing the characterisation of different MHC-DRB alleles in this primate [24]. DRB-STR microsatellite ancestral structure in Old World monkeys (OWM) contains a simple nucleotide repeat, whilst HLA and Mamu-DRB-associated microsatellite structure is more complex [25]. Taking into account that this microsatellite's variability pattern in humans and macaques is correlated with exon 2 polymorphism, making it an attractive option for typing these genes [25,26], it was thus of interest to verify whether the same occurs in New World monkeys (NWM). The MHC-DRB intron 2 in Platyrrhini is very variable in length, ranging from 50 bp to 1 Kbp [27], including a simple repeat sequence of around 50 bp downstream the limit between exon 2 and intron 2 [28,29].

The microsatellite present at the start of MHC-DRB genes' intron 2 in individuals from the *A. vociferans* and *A. nancymaae* species has thus been verified and characterised here, this being the first systematic characterisation of this marker in NWM, indicating the feasibility of its use in these primates for typing MHC-DRB.

Materials and Methods

Sample origin

Monkeys from the *Aotus nancymaae* (25 adults) and *Aotus vociferans* species (23 adults) were studied; they came from FIDIC's primate station in Leticia, Amazonas, Colombia. Blood samples from *A. vociferans* were collected fresh, whilst those from *A. nancymaae* had been collected in 2001. All primates were kept in conditions laid down by Colombian Ministry of Health (law 84/1989) and Colombian Institute of Health regulations for animal care, monitored weekly by CORPOAMAZONIA (resolutions 0202/1999 and 0028/2010). All procedures were approved and supervised by the Health Research Ethics Committee and FIDIC's Primate Station Ethics Committee.

The US Committee on the Care and Use of Laboratory Animals' guidelines were followed for all animal handling procedures, in turn complying with Colombian regulations for biomedical research (resolution 8430/1993 and law 84/1989).Monkeys at the station were numbered, sexed, weighed, given a physical-clinical exam and kept temporally in individual cages, prior to all experimental procedures. They were kept in controlled conditions regarding temperature (25°–30° centigrade) and relative humidity (83%), similar to those present in their natural environment. The monkeys' diet was based on a supply of fruit typical of the Amazon region (i.e. such primates' natural diet), vegetables and a nutritional supplement including vitamins, minerals and proteins. Environmental enrichment included visual barriers to avoid social conflict, feeding devices, some branches and vegetation, perches and habitat. Any procedure requiring animal handling was undertaken by trained veterinary personnel and animals were submitted to sedation and analgesia procedures to reduce stress when necessary [30].

Molecular characterisation of species of owl monkeys studied

Mitochondrial gene cytochrome c oxidase subunit II (mtCOII) sequences were used for determining the species to which the owl monkeys being studied belonged to, following the methodology described by Ashley & Vaughn [31]. PCR was used for amplifying the gene, using high fidelity Taq DNA polymerase. Two independent PCR reactions were performed and the amplified products were purified using a Wizard SV gel and PCR clean-up system kit (Promega, Madison, WI, USA); these were sent for sequencing with mtCOII-specific primers using the BigDye Terminator method (MACROGEN, Seoul, South Korea). The sequences so obtained were analysed for constructing phylogenetic trees and these were then compared to previously described sequences from databases for mtCOII from primates.

DNA, RNA extraction and cDNA synthesis

Genomic DNA (gDNA) from each specimen was isolated for *A. vociferans* from 300 µL peripheral blood samples using an UltraClean Blood DNA Isolation kit (Carlsbad, CA, USA), following the manufacturer's instructions. Total RNA was isolated from 2 mL peripheral blood in EDTA diluted 1:1 with PBS. A Ficoll-Hypaque density gradient (Lymphocyte Separation Medium, ICN Biomedicals, CA, USA) was used for isolating mononuclear cells, according to the manufacturer's recommendations. The lymphocytes so recovered were immediately homogenised with TRIzol reagent (Life Technologies, NY, USA). cDNA was synthesised with a SuperScript III First-Strand Synthesis System for RT-PCR kit (Life Technologies, NY, USA), using Oligo$(dT)_{20}$ (Invitrogen, NY, USA) as primer, according to the manufacturer's instructions.

Genomic DNA was isolated from leucocytes for *A. nancymaae*, using a NucleoSpin C+T kit (Macherey-Nagel AG, Oensingen, Switzerland), according to the manufacturer's protocol. Total RNA was isolated from PBMC using a NucleoSpin RNA kit (Macherey-Nagel AG, Oensingen, Switzerland), according to the manufacturer's recommendations. Reverse transcription was performed using SuperScript and Oligo$(dT)_{12–18}$ primer (Gibco BRL Life Technologies, Basel, Switzerland). Both gDNA and cDNA were preserved in 95% ethanol at -80°C until use. DNA integrity was verified by electrophoresis on 1% agarose gel, stained with SYBR Safe (Invitrogen) for visualisation under UV light. NanoDrop 2000 (Thermo Scientific) was used for calculating the concentration.

Amplifying, cloning and sequencing

The primers used here were designed by aligning available genome sequences for the *Callithrix jaccus*, *Homo sapiens* and *Macaca mulatta* MHC-DRB region (Table S1 in File S1), using Netprimer software [32] for optimising parameters. Two sets of primers were used for amplifying exon 2+ intron 2 sequences. The first primer set included direct primer GEX2DRBf (5′-GGTCAAGGTTCCCAGAGC-3) to the end of intron 1 and reverse GEX2DRBr (5′-CTCCAAGGATAAGAAGAAGCC-3′) located about 100 bp downstream of the end of the microsatellite. The second set included direct primer F-DRBINT1-2 (5′-TTCGTGTCCCCACAGCAC-3′) to the end of intron 1 and reverse R-DRBINT2-2 (5′-TAAACCCTCACCCCAGCC-3′) situated about 160 bp downstream of the end of the microsatellite (Figure 1). Direct primer DRBExon1PF (5′-CACTGGCTTTGGCTGGGGAC-3′) in exon 1 was used for amplification from cDNA with either DRBExon6PR1 (5′-CCACAAGGGAGGACATTTCTGC-3′) or DRBExon6PR2 (5′-CCAAGGGCAGAAGCTGAGGAA-3′) reverse primers in exon 6.

Two independent PCR reactions were carried out for each primate; the reactions followed recommendations made by Lenz *et al.*, [33] for avoiding chimera formation. The KAPA HiFi HotStart Readymix enzyme (Kapa Biosystems, Woburn, MA,

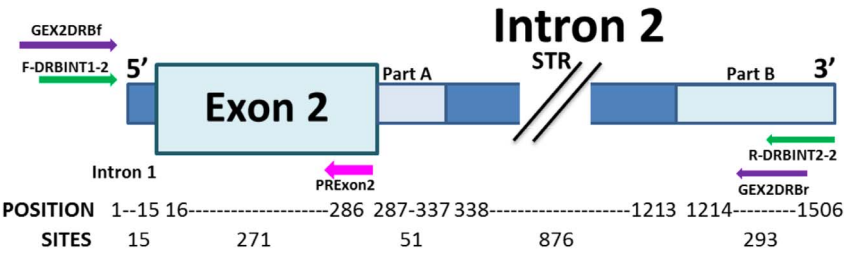


Figure 1. Diagram of the MHC-DRB region studied. The primers used for amplifying the exon 2+ intron 2 (partial) from gDNA are shown as arrows (purple and green); the PRExon2 primer was designed for confirmatory colony PCR (pink arrow). The MHC-DRB amplified sector (exon 2, intron alignable sectors 2 (A and B) and STR) was partitioned for sequence analysis (position and sites).
doi:10.1371/journal.pone.0096973.g001

USA) was used with 0.3 μM each primer and 10–40 ng DNA (in the case of gDNA) or 2 μL recently synthesised cDNA for 25 μL final volume. The PCR reaction at saturation was carried out in a PerkinElmer GeneAmp 9600 thermocycler. The following thermal profile was used for cDNA: 95°C for 5 min, 35 cycles at 98°C for 20 s, 66°C/67°C (when using the first or the second reverse primer, respectively) for 15 s, 72°C for 30 s and a final 5 min extension step at 72°C. The following thermal profile was used for gDNA: 95°C for 5 min, 35 cycles at 98°C for 20 s, 57°C/66°C (for the set of primers 1 or 2, respectively) for 15 s, 72°C for 30 s and the final extension step at 72°C for 5 min.

Amplified products were purified using a Wizard SV Gel and PCR Clean-Up System kit (Promega, USA) and a protocol was used for extending A with GoTaq Flexi DNA polymerase (Promega) to enable ligating them with the pGem-T Easy Vector Systems (Promega, Madison, WI, USA)vector, following the manufacturer's recommendations. The transformation was carried out in *Escherichia coli* JM109 strain competent cells. The clones were selected using positive selection with ampicillin and lacZ gene α-complementation. Plasmid DNA was extracted using an UltraClean 6 Minute Mini Plasmid Prep kit (MO BIO, USA).

Given that other targets were observed for the pairs of primers used for amplifying the exon 2+ intron 2 STR sector, a primer was designed at the end of exon 2 (PRExon2) (5′-TCGCCGCTGCACTGTGAAG-3′), enabling confirmatory colony PCR, using those used in amplifying gDNA as direct primers (Figure 1). The reaction contained 1 μL enzyme buffer, 0.6 μL $MgCl_2$ [25 mM], 1.6 μL dNTPs [1.25 mM], 0.8 μL of each primer [5 μM], 0.12 μL GoTaq Flexi DNA polymerase (Promega) and 10–40 ng colony DNA at 10 μL final volume. PCR conditions consisted of one cycle at 95°C for 5 min, 35 cycles at 95°C for 1 min, 60°C for 1 min, 72°C for 1 min and a final extension step at 72°C for 5 min.

At least 8 clones (confirmed from each amplification) were selected for sequencing; their DNA was sequenced in both directions using T7 and SP6 primers, following the BigDye Terminator method (MACROGEN, Seoul, South Korea).

Sequence analysis

The MHC-DRB sequence electropherograms were assembled using CLC Main Workbench software v.5 (CLC bio, Cambridge, MA, USA). The sequences so obtained had to comply with the following requirements to be considered as being valid: having been found in at least two independent PCR from the same individual, or coming from two different individuals (including previously reported sequences in this category). The alleles found were validated and named by a curator from the Immuno Polymorphism Database (IPD) [34,35].

Clustal X software (v2.1) was used for aligning all the MHC-DRB exon 2 and exon 2+ intron 2 sequences [36], using BioEdit Sequence Alignment Editor software for manual editing [37].

MEGA software (v5.2) was used for selecting the best nucleotide substitution model using Bayesian Information Criteria (BIC); phylogenetic trees were constructed using minimum evolution, neighbour joining, parsimony and maximum likelihood methods. The bootstrap test was used for supporting the trees so obtained, in addition, the interior branch test was used for supporting trees constructed using the minimum evolution and neighbour joining methods. 1,000 replicates were carried out; those groups having greater than or equal to 70% by bootstrap and greater than or equal to 95% by interior branch test were considered as supported groups [38,39].

Microsatellite analysis

Microsatellite search and building database (MSDB) software [40] was used for identifying the microsatellite, using the imperfect search mode; valid repeats were considered as those having 12 or more mononucleotide segments and repeats having 4 or more di-tri-tetra-penta-hexa nucleotides. Their descriptors were constructed using previous results and manual edition as guidelines. A compressibility method was used, given the difficulty of obtaining an unambiguous alignment of repeat sectors when they were analysed exclusively. The sequences were organised as 100 tandem repeats and compressed into separate files using an adaptive Lempel-Ziv algorithm (using the Linux command *compress*). From the resulting vector obtained from the bytes for each compressed sequence, a distance matrix was then calculated using either the Euclidean, Maximum or Manhattan metrics through the DIST package from R [41].Hierarchical clusters were constructed with the R hclust package [41], using single and complete methods.

Results

Amplicons ranging from ~700 bp to ~1,000 bp were obtained for *A. vociferans* and *A. nancymaae* samples (Figure 2); 289 sequences were obtained from exon 2+STR intron 2. One to five different MHC-DRB sequences per animal were observed from two independent PCR reactions; this implied the duplication of this loci, as has been reported previously [12]. A total of 34 distinct nucleotide sequences were validated, 28 of which were also isolated from cDNA: two new sequences belonging to two new *A. nancymaae* lineages (Aona-DRB*W9101 and Aona-DRB*W8901), 7 new sequences belonging to five new *A. vociferans* lineages (Aovo-DRB*W9101, Aovo-DRB*W9102, Aovo-DRB*W9201, Aovo-DRB*W9202, Aovo-DRB*W9301, Aovo-DRB*W8801, Aovo-DRB*W9001), 11 new sequences from previously reported *A. vociferans* lineages (Aovo-DRB1*0304, Aovo-DRB1*0305, Aovo-

DRB1*0306, Aovo-DRB1*0307, Aovo-DRB3*0601, Aovo-DRB*W1801, Aovo-DRB*W1802, Aovo-DRB*W1803, Aovo-DRB*W2901, Aovo-DRB*W3001, Aovo-DRB*W4501), 6 new from previously reported *A. nancymaae* lineages (Aona-DRB1*031701, Aona-DRB1*0329, Aona-DRB3*062502, Aona-DRB3*0628, Aona-DRB*W1808, Aona-DRB*W3002) and 8 already reported sequences for *A. nancymaae* lineages (Aona-DRB1*0328, Aona-DRB3*0615, Aona-DRB3*062501, Aona-DRB3*0626, Aona-DRB3*0627, Aona-DRB*W1806, Aona-DRB*W2908, Aona-DRB*W2910) (see Table S1 in File S1).

The MHC-DRB amplified sector was divided into the following partitions for sequence analysis: intron 1 (positions 1–15: 15 sites), exon 2 (positions 16–285: 270 sites), intron 2A (alignable; positions 286-325: 40 sites), intron 2R (STR sector; positions 326–1,110: 785 sites), intron 2B (alignable; positions 1,111–1,378: 268 sites) (Figure 1). These size ranges were related to aligning the sequences given in Figure S1 (within File S1).

Greater conservation of alignable areas was observed in intron 2 (A+B, 95±1% identity) compared to exon 2 (91±1% identity). An unambiguous alignment could not be made for intron 2 STR. This had substantial variation regarding its size, representing an 83 bp (Aovo-DRB*W9301) to 761 bp (Aona-DRB1*0329GA) interval.

Exon 2 in the sequences reported here were analysed together with 57 representative sequences of *Aotus* MHC-DRB allele lineages reported in previous studies by Suárez *et al.*, and Niño *et al.*,[12,18] and others available in Genbank. The evolutionary analysis methods described in the methodology were used on an alignment of 268 positions. Figure 3 shows the tree with the maximum likelihood method using a GTR+G+I model.

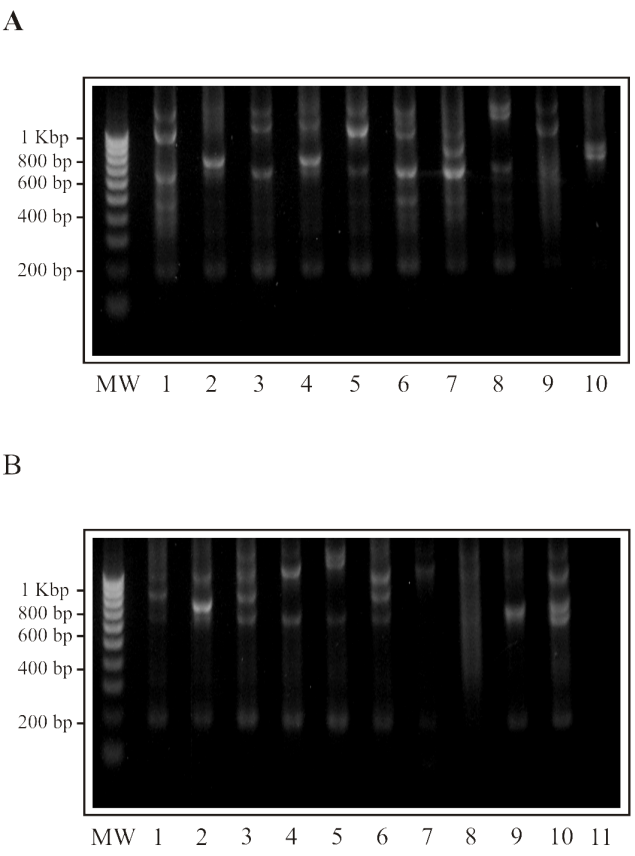


Figure 2. *A. nancymaae* and *A. vociferans* exon 2+ intron 2 partial amplicons. Amplicons ranging from ~700 bp to ~1,000 bp were obtained from *A. vociferans* and *A. nancymaae* samples. A. Lanes 1–10 show *A. nancymaae* amplicons. B. Lanes 1–10 show *A. vociferans* amplicons, lane 11 negative control. MW. molecular weight.
doi:10.1371/journal.pone.0096973.g002

The alleles observed came from some lineages previously reported by Suárez *et al.*, [12] thereby highlighting the existence of seven new lineages. Most lineages were supported by all the phylogenetic reconstruction and support methods (those only supported by some of them are indicated by circles in the node); however, the relationships between such lineages had low support (Figure 3). Based on the sequences studied here, most observed lineages were trans-specific, DRB1*03 GB and DRB*W89 lineages being species-specific for *A. nancymaae* and DRB*W88, DRB*W92, DRB*W90, DRB*W45 and DRB*W93 for *A. vociferans*.

Molecular phylogenetic analysis was made regarding the 34 sequences reported here, examining separately either exon 2 or the concatenated intron 2 alignable sectors (2A+2B) using previously described evolutionary analysis methods. Figure 4A shows the tree obtained by aligning exon 2 sequences (271 positions) with the maximum likelihood method, using a HKY+G+I model. Figure 4B shows the tree obtained by aligning intron 2 alignable sectors (344 positions) using the maximum likelihood method and an HKY+G+I model.

Most groups' identity was maintained regarding intron 2 alignable sectors compared to those observed in exon 2, although some became fused (i.e. DRB3*06 - DRB1*03 GA, DRB*W45 - DRB*W92 and DRB*W89 - DRB*W29), changing their relationships for each partition. However, lineage differentiation was well supported and even the association between some lineages (e.g. DRB3*06 - DRB1*03 GA, DRB*W30 – DRB*W92) was very clear, being maintained for the sets of data and methods analysed.

Compressibility was used for estimating similarity between sequences, given that the intron 2 repeat sector was not unequivocally alignable due to its repeat nature. The Lempel–Ziv algorithm was used with the Linux standard command *compress* for compressing files. Each sequence was repeated 100 times in tandem to ensure better resolution, so that files were 734–7,249 bytes after having been compressed (Figure 4C). Equivalent results were obtained using different metrics and grouping/clustering methods. Figure 4C shows the results using Manhattan metrics and the complete linkage agglomeration method. The STR grouping pattern is an intermediate between that of exon 2 and that generated from intron 2 A+B sectors.

It was observed that DRB3*06 and DRB1*03 GA lineages were associated in all the sectors analysed, being included in this grouping the DRB1*03GB lineage sequence in intron 2 A+B sectors and in STR. Each lineage's definition became lost in the STR, Aona-DRB1-0329GA, Aona-DRB1-031701GA and Aona-DRB1-0328GB sequences being differentiated by differences in STR length but being maintained in a common cluster with the remaining DRB3*06 and DRB1*03 sequences.

DRB*W88, DRB*W29, DRB*W30, DRB*W92, DRB*W91 and DRB*W90 lineages were associated in both exon 2 and the STR, the difference being that DRB*W89 and DRB*W45 lineages were inserted in the latter analysis, grouping with DRB*W29 and DRB*W30/*W91 lineages, respectively, in the STR and intron 2 A+B sectors. DRB*W89 and DRB*W45 were grouped in exon 2 with the DRB1*03GA - DRB3*06 - DRB*W18 group. The DRB*W30 and DRB*W92 lineages formed a cluster with the DRB1*03GA and DRB3*06 group in the intron 2 A+B sectors. The DRB*W18 lineage was always well characterised, having a cluster in STR and exon 2 which included DRB1*03GA - DRB3*06 – DRB1*03 GB lineages. The DRB*W92/*W91/

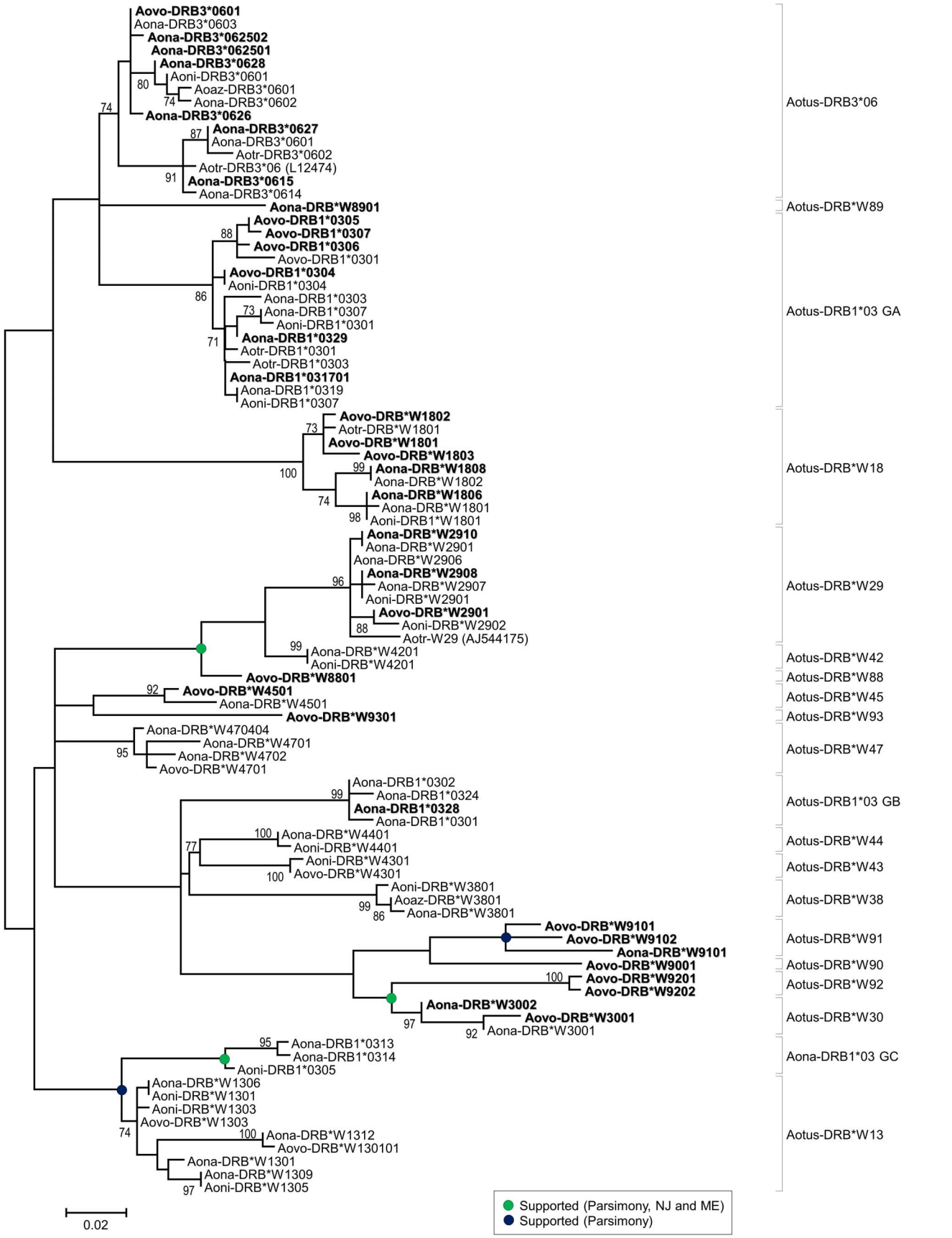

Aovo-DRB3*0601
Aona-DRB3*0603
Aona-DRB3*062502
Aona-DRB3*062501
Aona-DRB3*0628
Aoni-DRB3*0601
Aoaz-DRB3*0601
Aona-DRB3*0602
Aona-DRB3*0626
Aona-DRB3*0627
Aona-DRB3*0601
Aotr-DRB3*0602
Aotr-DRB3*06 (L12474)
Aona-DRB3*0615
Aona-DRB3*0614
Aona-DRB*W8901
Aovo-DRB1*0305
Aovo-DRB1*0307
Aovo-DRB1*0306
Aovo-DRB1*0301
Aovo-DRB1*0304
Aoni-DRB1*0304
Aona-DRB1*0303
Aona-DRB1*0307
Aoni-DRB1*0301
Aona-DRB1*0329
Aotr-DRB1*0301
Aotr-DRB1*0303
Aona-DRB1*031701
Aona-DRB1*0319
Aoni-DRB1*0307
Aovo-DRB*W1802
Aotr-DRB*W1801
Aovo-DRB*W1801
Aovo-DRB*W1803
Aona-DRB*W1808
Aona-DRB*W1802
Aona-DRB*W1806
Aona-DRB*W1801
Aoni-DRB1*W1801
Aona-DRB*W2910
Aona-DRB*W2901
Aona-DRB*W2906
Aona-DRB*W2908
Aona-DRB*W2907
Aoni-DRB*W2901
Aovo-DRB*W2901
Aoni-DRB*W2902
Aotr-W29 (AJ544175)
Aona-DRB*W4201
Aoni-DRB*W4201
Aovo-DRB*W8801
Aovo-DRB*W4501
Aona-DRB*W4501
Aovo-DRB*W9301
Aona-DRB*W470404
Aona-DRB*W4701
Aona-DRB*W4702
Aovo-DRB*W4701
Aona-DRB1*0302
Aona-DRB1*0324
Aona-DRB1*0328
Aona-DRB1*0301
Aona-DRB*W4401
Aoni-DRB*W4401
Aoni-DRB*W4301
Aovo-DRB*W4301
Aoni-DRB*W3801
Aoaz-DRB*W3801
Aona-DRB*W3801
Aovo-DRB*W9101
Aovo-DRB*W9102
Aona-DRB*W9101
Aovo-DRB*W9001
Aovo-DRB*W9201
Aovo-DRB*W9202
Aona-DRB*W3002
Aovo-DRB*W3001
Aona-DRB*W3001
Aona-DRB1*0313
Aona-DRB1*0314
Aoni-DRB1*0305
Aona-DRB*W1306
Aoni-DRB*W1301
Aoni-DRB*W1303
Aovo-DRB*W1303
Aona-DRB*W1312
Aovo-DRB*W130101
Aona-DRB*W1301
Aona-DRB*W1309
Aoni-DRB*W1305
0.02
Supported (Parsimony, NJ and ME)
Supported (Parsimony)
Aotus-DRB3*06
Aotus-DRB*W89
Aotus-DRB1*03 GA
Aotus-DRB*W18
Aotus-DRB*W29
Aotus-DRB*W42
Aotus-DRB*W88
Aotus-DRB*W45
Aotus-DRB*W93
Aotus-DRB*W47
Aotus-DRB1*03 GB
Aotus-DRB*W44
Aotus-DRB*W43
Aotus-DRB*W38
Aotus-DRB*W91
Aotus-DRB*W90
Aotus-DRB*W92
Aotus-DRB*W30
Aona-DRB1*03 GC
Aotus-DRB*W13

Figure 3. Maximum likelihood tree constructed from *Aotus* MHC-DRB exon 2 sequences (91 OTUs, 268 aligned positions). The analysis involved using the general time reversible model with invariable positions and Gamma distribution (5 categories, +G, parameter = 0.3371), > 70% bootstrap values are displayed. Green dots represent nodes supported by parsimony (>70% bootstrap), Neighbour joining and minimum evolution tests (>70% bootstrap and >95% interior branch test), but not in maximum likelihood analysis. Nodes represented by blue dots were supported only by parsimony (>70% bootstrap), but not in the maximum likelihood analysis. Bootstrap and interior branch tests involved using 1,000 replicates. The scale bar represents substitutions per site. New sequences reported in this study are shown in bold. Abbreviations and GenBank accession numbers for the sequences compared here are shown in Table S1 (within File S1).
doi:10.1371/journal.pone.0096973.g003

*W45 lineages were also included in intron 2 A+B sectors in this group.

The DRB*W93 lineage appeared in all analysis as a divergent member of the cluster formed by DRB3*06 - DRB1*03 GA - DRB1*03 GB - DRB*W18 and was related to the DRB*W45 lineage in exon 2, losing such relationship in intron 2. This lineage had a similar pattern to that of DRB*W89, whose grouping was very different between exon 2 and intron 2.

MSDB software [40] was used for characterising the amplified sequences (exon 2+ intron 2 (partial)) for analysing motifs (Table S2 in File S1). The different types of microsatellite agreed with the results found by the compression method.

It was observed that the microsatellite was characteristic for some lineages, being clearly differentiated by length and structure, forming 3 groups which included the 34 sequences described for the *Aotus* species included in this study. The STR could be divided into 3 sectors (Table S2 in File S1), the initial and final sectors being similar in all sequences; greater variability (intra and inter lineages) was observed in the microsatellite's central region. $(GA)_y$ was the main repeat motif found in all cases.

The STR had a similar structure throughout the DRB1*03 and DRB3*06 lineage sequence repeat sector, but there were differences regarding the number of repeats. The microsatellite had lengths ranging from 294 to 354 bp in *A. nancymaae* and *A.*

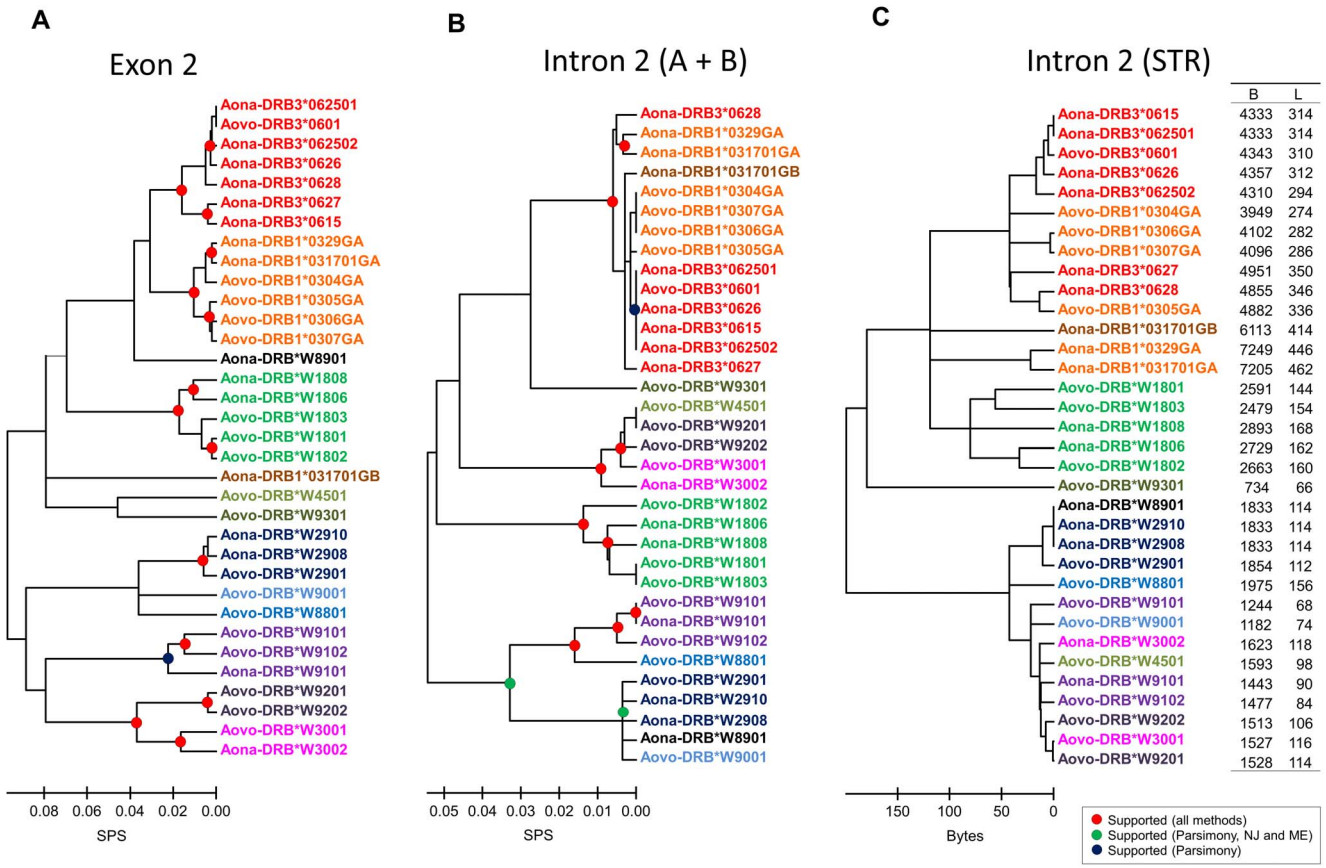


Figure 4. Comparison amongst exon 2, alignable sectors of intron 2 and intron 2 STR. A. Maximum likelihood tree constructed from *Aotus* MHC-DRB exon 2 sequences (34 OTUs, 344 aligned positions). The analysis used the Hasegawa-Kishino-Yano model with invariable positions and Gamma distribution (5 categories, +G, parameter = 0.2659, +I, 51.7393% sites). **B. Maximum likelihood tree constructed from *Aotus* MHC-DRB intron 2 (A+B) sequences (34 OTUs, 271 aligned positions).** The analysis involved using the Hasegawa-Kishino-Yano model with invariable positions and Gamma distribution (5 categories, +G, parameter = 0.2378, +I, 0.0% sites). **C. Complete linkage tree constructed from *Aotus* MHC-DRB intron 2 STR sequences.** The analysis was done using a Manhattan distance over Lempel-Ziv compression. Compression in bytes (B) and length in nucleotides (L) are also shown. Nodes indicated by red dots were supported by all methods. Nodes shown by green dots were supported by parsimony (>70% bootstrap), Neighbour joining and minimum evolution tests (>70% bootstrap and >95% interior branch test), but not in maximum likelihood analysis. Nodes represented by blue dots were supported only by parsimony (bootstrap >70%), but not in maximum likelihood analysis. Bootstrap and interior branch tests were performed using 1,000 replicates. The scale bar represents substitutions per site (A and B), and bytes (C). Abbreviations and GenBank accession numbers of the analysed sequences are shown in Table S1 (within File S1).
doi:10.1371/journal.pone.0096973.g004

vociferans DRB3*06 lineage sequences, a very similar structure being maintained in the initial and final part. There were slight differences in the repeats towards its central part and identical sequences were even observed in the STR, such as Aona-DRB3*062501/*0615. The DRB1*03 lineage sequences did not have a specific STR pattern, length varying from 274 to 462 bp. However, two defined groups were identified, one for the Aovo-DRB1*0304, 1*0307 and 1*0306GA sequences and another for the Aona-DRB1*0328GB, Aona-DRB1*031701 and 1*0329GA sequences, having similar structure and length. The Aona-DRB1*0329GA and *031701GA sequences had very similar distribution, having minimal differences regarding length at the start of the STR. Aovo-DRB1*0305GA had an STR having a particular structure, but maintaining similarity concerning lineage. The Aona-DRB3*0627 and Aona-DRB3*0628 sequences' repeat sector had similar distribution with DRB1*03 lineage sequences regarding repeats and length.

Regarding DRB*W18 lineage sequences, the STR had a size ranging from 144 to 160 bp, having similar distribution concerning composition and number of repeats at the beginning and end of the STR. Each sequence varied specifically at the central part in both nucleotide sequence and number of repeats. The Aovo-DRB*W9301 sequence had a 66 bp STR, being the smallest of all the sequences. It maintained a similar structure in the initial and final part to that described in other lineages, having a relatively short central region (26 bp).

The microsatellite had similar structure at the start and end in the DRB*W89/*W29/*W88/*W90/*W91/*W45/*W30/and *W92 lineages, having a length ranging from 68 to 156 bp. Various sequences had practically identical STR in this group, such as Aovo-DRB*W9201 and Aovo-DRB*W3001 (only one repeat being different), or identical STR, such as Aona-DRB*W8901, Aona-DRB*W2910 and Aona-DRB*W2908. Regarding this group, the Aovo-DRB*W2901 sequence had very similar organisation in the STR, having slight differences regarding structure and the number of repeats, given that even though belonging to the same lineage (W29), it came from a different species. The Aovo-DRB*W8801 sequence was similar to the DRB*W29 lineage, but had differences concerning the number of repeats in the central region. The Aovo-DRB*W9102 and Aona-DRB*W9101 sequences in the DRB*W91 lineage had similar microsatellite structure, having few differences concerning the number of repeats in the central region.

Regarding primates, 34 sequences from the MHC-DRB gene's exon 2+ intron 2 (partial) were analysed in *A. nancymaae* and *A. vociferans*; sequences related to the sector being studied were selected from previous typing reports [14,27] and a search of available complete or ongoing primate genomes using the BLAST algorithm [42]. This led to 86 primate sequences being included, including representatives for distinct human lineages (Table S1 in File S1). Clustal X v2.1 software was used for aligning the sequences [36]; these were then edited manually (especially in the repeat sector). The MHC-DRB sector was divided into the partitions shown in Figure 1 for their analysis.

A satisfactory alignment could not be made for the intron 2 repeat area (which is why it has not been considered in the phylogenetic analysis); however, the alignable sectors from intron 2 (A and B) had a notable degree of identity (90±0.8% for all primates), this being 94.1±0.7% for NWM and 90.2±0.7% for OWM. Such degree of conservation was even greater than that observed for exon 2, whose average identity for the primates studied here was 87.3±0.1% (similar values being obtained for both OWM and NWM). The intron 2 repeat region had notable variation regarding length between the primates analysed here; however, the presence of a central motif (GA)y was constant, being very idiosyncratic for each allelic lineage analysed. The sequences obtained from the *C. jacchus* genome were illustrative in this respect; whilst Caja-DRB*04 only had 3 base pairs in the repeat sector, Caja-DRB*05 was 849 bp.

The selected sequences were subjected to two molecular phylogenetic analysis; one used just exon 2 and another used intron 2 alignable sectors (A+B). Figure 5A shows the maximum likelihood analysis for exon 2. Several Catarrhini and Platyrrhini sequences were associated, presenting a mixture of alleles from both types of primate in several groups. For Catarrrihini, some groups were formed by a mixture of species belonging to different genera and families. This did not happen for NWM; the Callitrichidae maintained their identity in well-supported nodes, whilst the *Cebus* sequence was associated with one of the groups of sequences formed by *Aotus* sequences.

Figure 5B shows the maximum likelihood analysis for intron 2 alignable sectors (2A+2B). having clear division of Platyrrhini and Catarrhini sequences. Regarding Catarrhini, most groups were seen to be well-differentiated, being mainly groups exclusively containing Anthropoidea (*Homo, Pan, Gorilla*) or Cercopithecoidea (*Macaca, Chlorocebus*), few cases involving both groups occurring simultaneously. A genus-specific disposition predominated in Platyrrhini. The *Aotus* sequences were configured into three groups, whilst Callitrichidae formed multiple genus- specific clusters. The result for this sector was similar to that observed for exon 2+ intron 2 (A+B) (not shown).

Discussion

Analysis of *Aotus* MHC-DRB gene exon 2 sequences showed how the number of trans-specific lineages for the genus were increased and defined by improving *A. vociferans* sampling. Except for DRB*W41, DRB*W43, DRB*W44, DRB*W38, DRB*W42, DRB*W47, DRB*W13 and DRB1*03GC lineages, the remaining *Aotus* lineages were sampled in the present study (Figure 3).

The definition of two sub-lineages could be observed in lineages like DRB*W18 (having no report of alleles for *A. vociferans*), one belonging to species typically from the north of the Amazon region (*A. vociferans* and *A. trivirgatus*) and another related to species typically from the south of the Amazon region (*A. nancymaae* and *A. nigriceps*). Such tendency (although less marked) was observed for the DRB1*03GA lineage where a well-supported sub-lineage was exclusively grey-neck (there were also exclusively red-neck sub-lineages). An *A. vociferans* sequence (Aovo-DRB3*0601) was reported for the DRB3*06 lineage (apparently exclusive to red-neck monkeys) which was identical to an *A. nancymaae* sequence (Aona-DRB3*062501). This was also true for the DRB*W45 and DRB*W30 lineages where *A. vociferans* sequences were described (Figure 3). Apparently exclusive lineages exist, such as the DRB1*03GB lineage, which has just *A. nancymaae* sequences; however, differing degrees of trans-specificity were observed in the rest of the lineages, even though there could be specific sub-lineages.

There were differences regarding frequencies but not regarding the repertoires of the two *Aotus* species studied here, indicating that each had undergone diversification; however, they maintained notable identity between their MHC-DRB repertoires over a relatively long period of time (from 13–8 mya) [43]. Such trans-specific polymorphism in repertoires suggests that using both species as animal models could be equivalent for MHC-DRB-mediated processes [44].

Comparative analysis of *Aotus* DRB genes' exon 2 phylogenies (Figure 4A) and intron 2 alignable sectors (Figure 4B) showed that

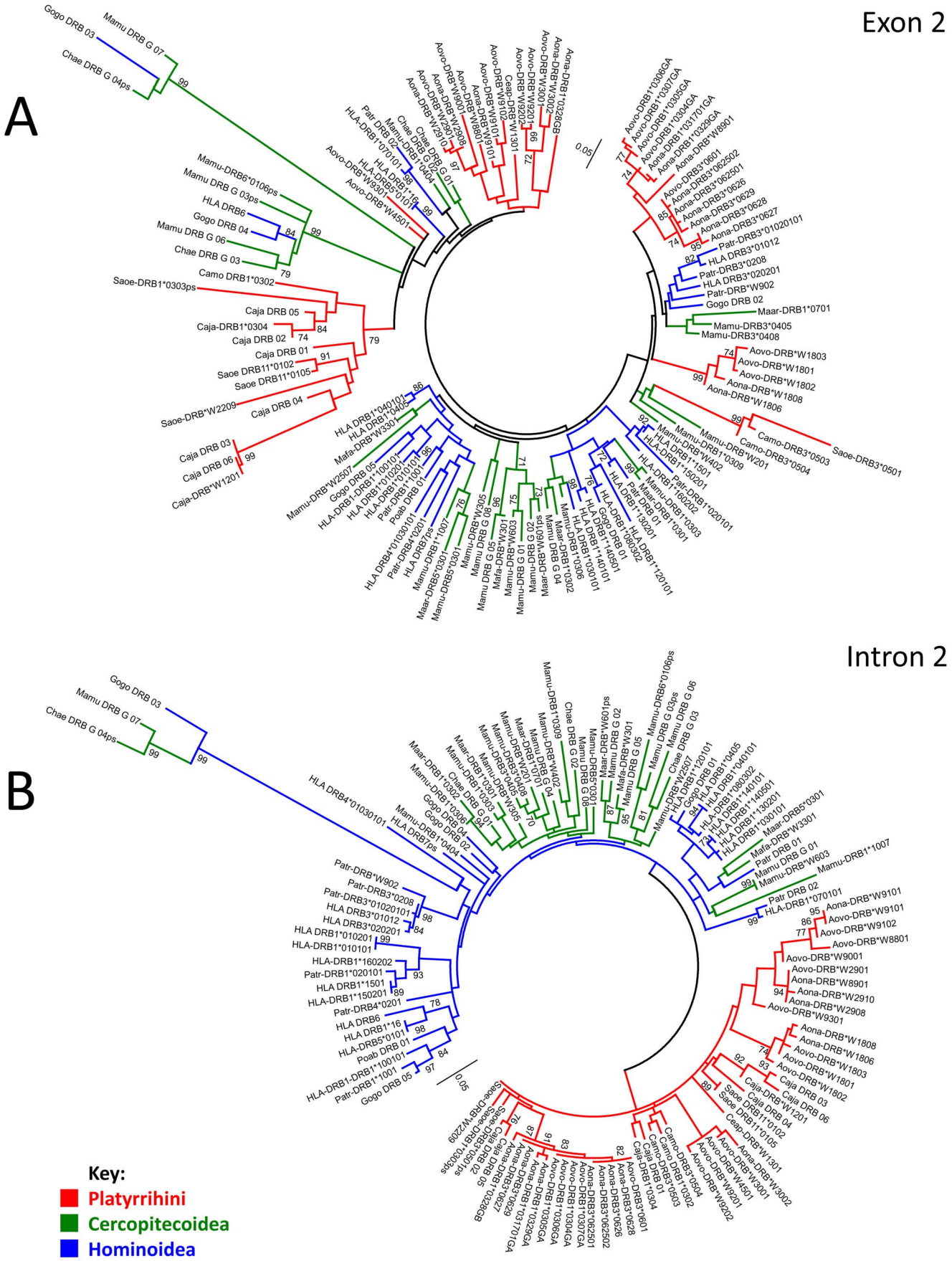

A
Exon 2
B
Intron 2
Key:
Platyrrihini
Cercopitecoidea
Hominoidea
0.05

Figure 5. Maximum likelihood trees. A. Maximum likelihood tree constructed from *Aotus* MHC-DRB exon 2 sequences (120 OTUs, 271 aligned positions). The analysis involved using Kimura's 2 parameter model with invariable positions and Gamma distribution (5 categories, + G, parameter = 0.5550). **B Maximum likelihood tree constructed from *Aotus* MHC-DRB alignable sectors of intron 2 (132 OTUs, 359 aligned positions).** The analysis involved using the general time reversible model with invariable positions and Gamma distribution (5 categories, + G, parameter = 1.2072). >70% bootstrap values are displayed. The bootstrap test involved using 1,000 replicates. The scale bar represents substitutions per site. Abbreviations and GenBank accession numbers for the sequences compared here are shown in Table S1 (within File S1).
doi:10.1371/journal.pone.0096973.g005

some of the lineages clearly maintained their identity, whilst others became merged. The relationship between lineages also changed from one sector to another, groups of well-supported lineages becoming formed in analysis of intron 2 (this did not happen in exon 2). The degree of intron 2 A+B sector conservation was notable compared to exon 2, thereby highlighting the magnitude of the latter's selection process.

Differential grouping showed that distinct forces have modulated each DRB gene sector's evolution, thereby posing the question, "Which one reflects more accurately the origin of DRB genes en *Aotus*?" If the intron 2 alignable sectors were to be chosen (given that they apparently have not undergone the previously described phenomena generating diversity in exon 2), then one would have a scenario where the number of lineages would be less than that proposed based on exon 2 polymorphism, and the relationships between them would have been different. Positive selection and recombination would thus have generated variability which would have grouped (by convergence and/or recombination) the sequences in previously described lineages. If exon 2 were to be chosen, the scenario would be marked by intron 2 recombination which would lead to the different groups' homogenisation in fewer lineages.

Recombination substantially affects support for trees [45,46], thereby making the first scenario more probable, given that the tree for the intron 2 alignable sectors was better supported than that for exon 2. However, complete DRB gene sequences (including coding and non-coding sectors) are needed to clarify this point.

STR in *Aotus* mainly had $(GA)_y$ repeats interrupted by CT motifs and a similar structure between sequences at the 5′ and 3′ extremes belonging to the same group according to phylogeny for the intron 2 alignable sectors (Figure S1 and Table S2 in File S1, Figure 4B). The $(GA)_y$ repeats form part of the ancestral structure described for *Catarrhini* [29,47,48].

The *Aotus* MHC-DRB microsatellite is variable in length, as has been described for humans, macaques and chimpanzees. Exon 2 analysis led to observing that the microsatellite for the DRB3*06 lineage (the Aovo-DRB3*0601, Aona-DRB3*062502, Aona-DRB3*0626, Aona-DRB3*0628 and Aona-DRB3*0627 sequence group) could differentiate them due to their variable length, except for the Aona-DRB3*062501 and Aona-DRB3*0615 sequences which had identical length and sequence, meaning that sequencing methods were needed for identifying these alleles.

The microsatellite had highly variable length in the DRB1*03GA, DRB*W18, DRB*W91, DRB*W93, DRB*W88, DRB*W90, DRB*W91, DRB*W45 and DRB*W30 lineage and could differentiate the sequences to which it belonged in *A. nancymaae* and *A. vociferans*, except for the Aona-DRB*W8901, Aona-DRB*W2910/*W2908 and Aovo-DRB*W9201 sequences where the microsatellite had the same length thereby differentiating it as a group, but not individually, and thus working as a screening but not as a typing method for these alleles.

According to the results reported here, the composition of the microsatellite described for MHC-DRB sequences in *A. nancymaae* and *A. vociferans* was more variable and complex than in humans and other *Catarrhini* (Figure 6). Comparison of the groups deduced from exon 2 and those observed for the STR was not always consistent, just as in previous reports concerning OWM published by Bontrop *et al.*, [23,24,49].

The ancestral structure of the microsatellite in *Catarrhini* has evolved from dinucleotide repeats $(GT)_x$ $(GA)_y$. Current structure of the HLA- and Mamu-DRB-associated microsatellite was seen to be more complex (Figure 6). The repeat in the 5′ extreme was the longest, uninterrupted part; the second part (GA)z was short and interrupted by other dinucleotides, being able to correlate well with different DRB gene lineages. The length of the third segment (GA)y could also be correlated with some DRB gene lineages in *M. mulatta*. The 3′ extreme consisted of a short (GC)n repeat part. It is known that mutation tendency depends on repeat length, since there is less microsatellite stability in the longer dinucleotide repeats than in the shorter ones [23,28,47].

The $(GA)_y$ dinucleotide in *Aotus* was maintained in STR structure and the $(GT)_x$ repeat was not present. Initial and final extreme repeat length in the microsatellites was similar between lineages, whilst repeat composition and number in the middle part could have been associated with specific lineages, sequences or groups; this could have been explained by the inherent differences in mutation rate between the different parts of the microsatellite.

The *A. nancymaae* and *A. vociferans* MHC-DRB microsatellite was present in all the DRB genes studied here, having considerable differences regarding length and variability, enabling it to differentiate some lineages, and even DRB sequences, thereby agreeing with exon 2 diversity. STR variability in other primate species was not always consistent with a given lineage; however, others could be characterised by a unique pattern [23,26].

Analysis of the repeat region of 5 sequences from another Platyrrhini genus, *Callithrix jacchus* (Caja-DRB*01/*02/*03/*05/*06), revealed the same organisational pattern described for *Aotus*, having a $(GA)_y$ repeat in the central sector which was complex, interrupted by CT motifs, highly variable in length and number of repeats; it came within the same ranges observed for *Aotus*, having 130-554 bp repeats. The initial and final parts of the Caja-DRB STR had similar length and sequence, the initial part being similar to that for *Aotus*, but having a more complex final part (Table S2 in File S1).

Using techniques which did not require sequence alignment for comparing them was useful in cases where this was impractical (i.e. analysis of complete genomes). As compression gives a basic measurement of a sequence of characters' algorithmic complexity, it could be especially useful when dealing with biological sequences. Using Lempel-Ziv complexity as a tool for data-mining and classifying nucleic acid and protein sequences has already been proposed [50,51].

Compression in the present work measured two relevant parameters in microsatellite analysis, given that compressed size (in bytes) would have depended on a sequence's length and degree of simplicity (monotony), being very correlated with length in this case ($R^2 = 0.9793$) given that the repeats between sequences were the same type and had the same complexity, mainly varying regarding number (Figure S1 and Table S2 in File S1).

Results for the repeat sector and exon 2 and intron 2 alignable sectors (Figures 4A and B) highlighted sector agreement. There

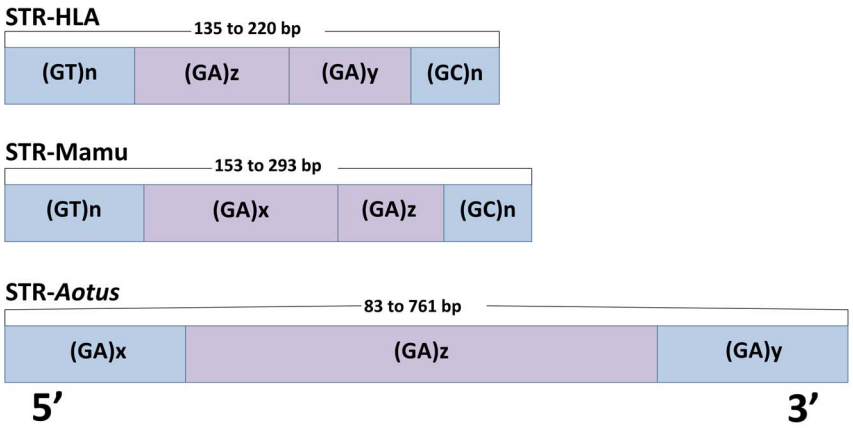


Figure 6. MHC-DRB STR model for Platyrrhini *cf* Catarrhini. The Figure shows the STR structure described by Bontrop *et al*., for Human HLA-DRB (STR-HLA) and *Macaca mulatta* MHC-DRB (STR-Mamu); and our proposed *Aotus* MHC-DRB model (STR-*Aotus*). The lengths ranges for each STR are shown. The ancestral structure of the microsatellite in Catarrhini has evolved from dinucleotide repeats (GT)x (GA)y; the (GA)y dinucleotide in *Aotus* was maintained in STR structure and the (GT)x repeat was not present. STR in *Aotus* mainly had (GA)y repeats interrupted by CT motifs, this being more complex and bigger than Catarrhini STR.
doi:10.1371/journal.pone.0096973.g006

were two large groups, one formed by DRB3*06, DRB1*03, DRB*W18 and DRB*W93 and another formed by DRB*W29, DRB*W91 and DRB*W88, DRB*W89, DRB*W45, DRB*W30 - DRB*W92 lineages being associated with one of the two, according to the sector being analysed. The DRB*W89 and DRB*W45 lineages had the greatest differences regarding grouping pattern between exon 2 and the STR, whilst this occurred between the STR and intron 2 alignable sectors in DRB*W30 - DRB*W92 groups.

There was no differentiation between lineages for the DRB3*06-DRB1*03, DRB*W89, DRB*W29, DRB*W30, DRB*W92 and DRB*W91 groups, suggesting that exon 2 origin and diversity represented a characteristic which could have been derived from a less diverse original set. This agreed with the origin of NWM arising from these primates' African transfer during the Eocene age (35 mya) [52], implying that current class I and class II MHC lineages were generated from a founding event [53].

Phylogenetic analysis of MHC-DRB gene exon 2 in primates (Figure 5A) highlighted the difficulty of inferring this gene's evolutionary relationships based just on this sector. Previous studies [12,27,54] have shown that even though the alleles being studied have been associated in assigned lineages, there has been poor support for such relationships, given the occurrence of phenomena guaranteeing PBR functional and structural stability. However, as a response to the diversity exhibited by pathogen proteins as a mechanism for avoiding the immune response, variation in the PBR has been produced by several mechanisms, thereby establishing a co-evolutionary arms race [55]. The most relevant features would include balanced selection (for conserving both functional integrity and diversifying the receptor) and recombination (intra-locus and inter-loci) [15–17,56].

Analysis of just exon 2 has revealed the occurrence of groups of multiple primate species, thus showing the existence of groups containing Platyrrhini and Catarrhini sequences, even though most groups of sequences were biased regarding the types of primate forming them (i.e. showing some group as being predominant) (Figure 5A). The inferences drawn regarding exon 2 did not lead to concluding whether such grouping reflected a common origin for these lineages or convergence.

Concerning the particular case of MHC-DRB, molecular convergence at exon 2 level has been described in both primates [27,53,57] and other orders of mammals [58–60]. Evidence sustaining such observation has been based on independent analysis of other MHC-DRB sectors not implicated in PBR formation, where sequences belonging to Catarrhini and Platyrrhini have been shown to cluster apart, whilst for exon 2, they cluster within common allelic lineages [27,53,57], thus favouring the appearance of common motifs between different lineages, thereby contributing towards reducing bootstrap support [45].

Phylogenetic comparison of exon 2 (Figure 5A) and intron 2 alignable sectors (Figure 5B) from the *Aotus* sequences so obtained and a representative sample from other primates, showed that whilst the last displays a clear division between Platyrrhini MHC-DRB sequences (shown in red) and Catarrhini (Hominoidea in blue and Cercopithecoidea in green), the analysis of exon 2 presented a mixture of alleles from both types of primate, and thus molecular convergence between several groups is observed. This agreed with previous reports [27,53].

Differently to the convergence regarding phenotypical features, convergence at molecular level is a rare phenomenon producing the same effect as another phenomenon which has shaped MHC evolution, trans-specific polymorphism, implying the maintenance of allele diversity going beyond speciation events due to balanced selection [61].

The extent of the convergence between related groups' lineages has not been previously described for DRB genes in primates; our analysis showed that the phylogenies obtained from exon 2 and those obtained for intron 2 differed regarding the relationship inside Platyrrhini and Catarrhini. The occurrence of groups containing Hominoidea and Cercopithecoidea sequences was greater in analysis inferred from exon 2 (Figure 5A) than in clusters obtained from intron 2 (Figure 5B). The same was true for Platyrrhini, where the *C. apella* sequence appeared to be included within a group of *Aotus* sequences in analysis of exon 2 (Figure 5A), whilst this did not occur regarding inference from intron 2 (Figure 5B). The foregoing could imply more recent convergence than that described to date. It also shows that MHC-DRB in primates has had a complex evolutionary mode in which trans-specific evolution has occurred at the same time as convergence between the different species analysed, underlining a predominantly intra-generic TSP pattern.

The molecular study in primates of the DRB gene in intron 2 (without considering the repeat sector) showed a high degree of identity for all the primates, indicating a clear division between NWM and OWM and between DRB gene lineages, demonstrating an independent origin for each DRB repertoire in Platyrrhini and Catarrhini. The study also verified that the microsatellite present in *A. nancymaae* and *A. vociferans* MHC-DRB gene intron 2 could be a useful marker for high and medium resolution genotyping of the MHC-DRB gene in these species, and probably in NWM. The microsatellite sequences could have been associated with the polymorphism observed for the corresponding *Aotus* MHC-DRB exon 2, making this a valuable tool for studying these genes' variability.

Supporting Information

File S1 Supporting tables and figure. Table S1. Sequences used for designing primers and analysis of exon 2+ intron 2. Available genome sequences for the *Callithrix jaccus*, *Homo sapiens* and *Macaca mulatta* MHC-DRB region were used for designing the primers. Sequences used for comparative analysis of *Aotus* MHC-DRB exon 2+ intron 2 (partial), as well as those used for analysing MHC-DRB exon 2+ intron 2 (partial) in primates. **Table S2. Microsatellite sequence and length in Platyrrhini MHC-DRB.** STR structure corresponding to each DRB gene sequence for *A. nancymaae* and *A. vociferans*. The colours signify microsatellite identity or similarity and microsatellite sequences corresponding to MHC-DRB *Callithrix jacchus* (in bold) are shown at the end. **Figure S1. Aligning *A. vociferans* and *A. nancymaae* MHC-DRB gene exon 2+ intron 2 (partial) sequences.**
(PDF)

Acknowledgments

We would like to thank Wendy Ortiz, Luis Alfredo Baquero and Yoelis Yepes for their technical assistance, and Jason Garry for translating the manuscript.

Author Contributions

Conceived and designed the experiments: CL CFS. Performed the experiments: CL. Analyzed the data: CFS CL. Wrote the paper: CFS CL LFC MEP MAP.

References

1. Ward JM, Vallender EJ (2012) The resurgence and genetic implications of New World primates in biomedical research. Trends Genet 28: 586–591.
2. Bontrop RE (2001) Non-human primates: essential partners in biomedical research. Immunol Rev 183: 5–9.
3. Bone JF, Soave OA (1970) Experimental tuberculosis in owl monkeys (Aotus trivirgatus). Lab Anim Care 20: 946–948.
4. Gysin J (1988) Animal models: primates. In: Sherman IW, editor.Malaria: parasite biology, pathogenesis and protection. Washington DC: ASM. pp. 419–439.
5. Jones FR, Baqar S, Gozalo A, Nunez G, Espinoza N, et al. (2006) New World monkey Aotus nancymae as a model for Campylobacter jejuni infection and immunity. Infect Immun 74: 790–793.
6. Lujan R, Dennis VA, Chapman WL Jr, Hanson WL (1986) Blastogenic responses of peripheral blood leukocytes from owl monkeys experimentally infected with Leishmania braziliensis panamensis. Am J Trop Med Hyg 35: 1103–1109.
7. Noya O, Gonzalez-Rico S, Rodriguez R, Arrechedera H, Patarroyo ME, et al. (1998) Schistosoma mansoni infection in owl monkeys (Aontus nancymai): evidence for the early elimination of adult worms. Acta Trop 70: 257–267.
8. Pico de Coana Y, Rodriguez J, Guerrero E, Barrero C, Rodriguez R, et al. (2003) A highly infective Plasmodium vivax strain adapted to Aotus monkeys: quantitative haematological and molecular determinations useful for P. vivax malaria vaccine development. Vaccine 21: 3930–3937.
9. Polotsky YE, Vassell RA, Binn LN, Asher LV (1994) Immunohistochemical detection of cytokines in tissues of Aotus monkeys infected with hepatitis A virus. Ann N Y Acad Sci 730: 318–321.
10. Diaz D, Naegeli M, Rodriguez R, Nino-Vasquez JJ, Moreno A, et al. (2000) Sequence and diversity of MHC DQA and DQB genes of the owl monkey Aotus nancymaae. Immunogenetics 51: 528–537.
11. Guerrero JE, Pacheco DP, Suarez CF, Martinez P, Aristizabal F, et al. (2003) Characterizing T-cell receptor gamma-variable gene in Aotus nancymaae owl monkey peripheral blood. Tissue Antigens 62: 472–482.
12. Suarez CF, Patarroyo ME, Trujillo E, Estupinan M, Baquero JE, et al. (2006) Owl monkey MHC-DRB exon 2 reveals high similarity with several HLA-DRB lineages. Immunogenetics 58: 542–558.
13. Bontrop RE, Otting N, de Groot NG, Doxiadis GG (1999) Major histocompatibility complex class II polymorphisms in primates. Immunol Rev 167: 339–350.
14. Doxiadis GG, de Groot N, de Groot NG, Doxiadis, II, Bontrop RE (2008) Reshuffling of ancient peptide binding motifs between HLA-DRB multigene family members: old wine served in new skins. Mol Immunol 45: 2743–2751.
15. Yeager M, Hughes AL (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunol Rev 167: 45–58.
16. Takahata N, Satta Y (1998) Selection, convergence, and intragenic recombination in HLA diversity. Genetica 102–103: 157–169.
17. Takahata N, Satta Y (1998) Footprints of intragenic recombination at HLA loci. Immunogenetics 47: 430–441.
18. Nino-Vasquez JJ, Vogel D, Rodriguez R, Moreno A, Patarroyo ME, et al. (2000) Sequence and diversity of DRB genes of Aotus nancymaae, a primate model for human malaria parasites. Immunogenetics 51: 219–230.
19. Middleton SA, Anzenberger G, Knapp LA (2004) Identification of New World monkey MHC-DRB alleles using PCR, DGGE and direct sequencing. Immunogenetics 55: 785–790.
20. Ujvari B, Belov K (2011) Major histocompatibility complex (MHC) markers in conservation biology. Int J Mol Sci 12: 5168–5186.
21. Baquero JE, Miranda S, Murillo O, Mateus H, Trujillo E, et al. (2006) Reference strand conformational analysis (RSCA) is a valuable tool in identifying MHC-DRB sequences in three species of Aotus monkeys. Immunogenetics 58: 590–597.
22. Knapp LA, Cadavid LF, Eberle ME, Knechtle SJ, Bontrop RE, et al. (1997) Identification of new mamu-DRB alleles using DGGE and direct sequencing. Immunogenetics 45: 171–179.
23. Doxiadis GG, de Groot N, Claas FH, Doxiadis II, van Rood JJ, et al. (2007) A highly divergent microsatellite facilitating fast and accurate DRB haplotyping in humans and rhesus macaques. Proc Natl Acad Sci U S A 104: 8907–8912.
24. de Groot NG, Heijmans CM, de Groot N, Doxiadis GG, Otting N, et al. (2009) The chimpanzee Mhc-DRB region revisited: gene content, polymorphism, pseudogenes, and transcripts. Mol Immunol 47: 381–389.
25. Doxiadis GG, de Groot N, Dauber EM, van Eede PH, Fae I, et al. (2009) High resolution definition of HLA-DRB haplotypes by a simplified microsatellite typing technique. Tissue Antigens 74: 486–493.
26. de Groot N, Doxiadis GG, de Vos-Rouweler AJ, de Groot NG, Verschoor EJ, et al. (2008) Comparative genetics of a highly divergent DRB microsatellite in different macaque species. Immunogenetics 60: 737–748.
27. Kriener K, O'Huigin C, Tichy H, Klein J (2000) Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. Immunogenetics 51: 169–178.
28. Riess O, Kammerbauer C, Roewer L, Steimle V, Andreas A, et al. (1990) Hypervariability of intronic simple (gt)n(ga)m repeats in HLA-DRB genes. Immunogenetics 32: 110–116.
29. Andersson G, Larhammar D, Widmark E, Servenius B, Peterson PA, et al. (1987) Class II genes of the human major histocompatibility complex. Organization and evolutionary relationship of the DR beta genes. J Biol Chem 262: 8748–8758.
30. National Research Council (U.S.). Committee for the Update of the Guide for the Care and Use of Laboratory Animals, Institute for Laboratory Animal Research (U.S.), National Academies Press (U.S.) (2011) Guide for the care and use of laboratory animals. Washington, D.C.: National Academies Press. xxv, 220 p. p.
31. Ashley A (1995) Owl monkeys (Aotus) are highly divergent in mitochondrial cytochrome C oxidase (COII) sequences. Journal of Primatology 16: 793–806.
32. PREMIER Biosoft International PA, CA, USA (2013) Netprimer.
33. Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci–implications for evolutionary analysis. Gene 427: 117–123.
34. de Groot NG, Otting N, Robinson J, Blancher A, Lafont BA, et al. (2012) Nomenclature report on the major histocompatibility complex genes and alleles of Great Ape, Old and New World monkey species. Immunogenetics 64: 615–631.
35. Robinson J, Halliwell JA, McWilliam H, Lopez R, Parham P, et al. (2013) The IMGT/HLA database. Nucleic Acids Res 41: D1222–1227.

36. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.
37. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41: 95–98.
38. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
39. Sitnikova T (1996) Bootstrap method of interior-branch test for phylogenetic trees. Mol Biol Evol 13: 605–611.
40. Du L, Li Y, Zhang X, Yue B (2013) MSDB: a user-friendly program for reporting distribution and building databases of microsatellites from genome sequences. J Hered 104: 154–157.
41. R Core Team (2013) R: A language and environment for statistical computing. Available: http://www.R-project.org/: R Foundation for Statistical Computing.
42. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.
43. Menezes AN, Bonvicino CR, Seuanez HN (2010) Identification, classification and evolution of owl monkeys (Aotus, Illiger 1811). BMC Evol Biol 10: 248.
44. Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum Immunol 19: 155–162.
45. Efron B, Halloran E, Holmes S (1996) Bootstrap confidence levels for phylogenetic trees. Proc Natl Acad Sci U S A 93: 7085–7090.
46. Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16: 562–563.
47. Bergstrom TF, Engkvist H, Erlandsson R, Josefsson A, Mack SJ, et al. (1999) Tracing the origin of HLA-DRB1 alleles by microsatellite polymorphism. Am J Hum Genet 64: 1709–1718.
48. Epplen C, Santos EJ, Guerreiro JF, van Helden P, Epplen JT (1997) Coding versus intron variability: extremely polymorphic HLA-DRB1 exons are flanked by specific composite microsatellites, even in distant populations. Hum Genet 99: 399–406.
49. Doxiadis GG, de Groot N, de Groot NG, Rotmans G, de Vos-Rouweler AJ, et al. (2010) Extensive DRB region diversity in cynomolgus macaques: recombination as a driving force. Immunogenetics 62: 137–147.
50. Otu HH, Sayood K (2003) A new sequence distance measure for phylogenetic tree construction. Bioinformatics 19: 2122–2130.
51. Gusev VD, Nemytikova LA, Chuzhanova NA (1999) On the complexity measures of genetic sequences. Bioinformatics 15: 994–999.
52. Schrago CG, Russo CA (2003) Timing the origin of New World monkeys. Mol Biol Evol 20: 1620–1625.
53. Trtkova K, Mayer WE, O'Huigin C, Klein J (1995) Mhc-DRB genes and the origin of New World monkeys. Mol Phylogenet Evol 4: 408–419.
54. Suarez CF, Cardenas PP, Llanos-Ballestas EJ, Martinez P, Obregon M, et al. (2003) Alpha1 and alpha2 domains of Aotus MHC class I and Catarrhini MHC class Ia share similar characteristics. Tissue Antigens 61: 362–373.
55. Acevedo-Whitehouse K, Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? Trends Ecol Evol 21: 433–438.
56. Reusch TB, Langefors A (2005) Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback Gasterosteus aculeatus. J Mol Evol 61: 531–541.
57. O'Huigin C (1995) Quantifying the degree of convergence in primate Mhc-DRB genes. Immunol Rev 143: 123–140.
58. Srithayakumar V, Castillo S, Mainguy J, Kyle CJ (2012) Evidence for evolutionary convergence at MHC in two broadly distributed mesocarnivores. Immunogenetics 64: 289–301.
59. Gustafsson K, Andersson L (1994) Structure and polymorphism of horse MHC class II DRB genes: convergent evolution in the antigen binding site. Immunogenetics 39: 355–358.
60. Gustafsson K, Brunsberg U, Sigurdardottir S, Andersson L (1991) A Phylogenetic Investigation of MHC Class II DRB Genes Reveals Convergent Evolution in the Antigen Binding Site. In: Klein J, Klein D, editors.Molecular Evolution of the Major Histocompatibility Complex: Springer Berlin Heidelberg. pp. 119–130.
61. Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. Annu Rev Genet 41: 281–304.

Table S1. Sequences used in this article

Sequence	Remarks	Accesion Number Genbank
Sequences used for primer designing		
Caja_DRB_G_01	43380-44258 Callithrix jacchus BAC clone CH259-77F15 from chromosome unknown, complete sequence	AC242730
Caja_DRB_G_02	105908-107178 Callithrix jacchus BAC clone CH259-77F15 from chromosome unknown, complete sequence	AC242730
Caja_DRB_G_03	192613-193403 Callithrix jacchus BAC clone CH259-77F15 from chromosome unknown, complete sequence	AC242730
Caja_DRB_G_04	138963-138357 Callithrix jacchus BAC clone CH259-15O7 from chromosome unknown, complete sequence	AC243457
Caja_DRB_G_05	76772-75338 Callithrix jacchus BAC clone CH259-15O7 from chromosome unknown, complete sequence	AC243457
Caja_DRB_G_06	161835-162584 Callithrix jacchus BAC clone CH259-49P2 from chromosome unknown, complete sequence	AC242576
Mamu_DRB_G_01	140068-140822 Macaca mulatta Major Histocompatibility Complex BAC MMU012K22	AC148663
Mamu_DRB_G_02	28917-29606 Macaca mulatta Major Histocompatibility Complex BAC MMU012K22	AC148663
Mamu_DRB_G_03ps	111593-112214 Macaca mulatta Major Histocompatibility Complex BAC MMU012K22	AC148663
Mamu_DRB_G_05	c75283-74620 Macaca mulatta Major Histocompatibility Complex BAC MMU248N17	AC148697
Mamu_DRB_G_06	c151148-150454 Macaca mulatta Major Histocompatibility Complex BAC MMU248N17	AC148697
Mamu_DRB_G_07	c29018-28359 Macaca mulatta Major Histocompatibility Complex BAC MMU248N17	AC148697
Mamu_DRB_G_08	c26068-25409 Macaca mulatta Major Histocompatibility Complex BAC MMU281E18	AC148700
Mamu_DRB_G_04	173697-174407 Macaca mulatta Major Histocompatibility Complex BAC MMU370O02	AC148706
HLA_DRB1*040101	Homo sapiens major histocompatibility complex, class II, DR53 haplotype (DR53)	NG_002433
HLA_DRB1*070101	Human DNA sequence from clone DADB-102D14 on chromosome 6	CR753309
HLA_DRB1*1501	Homo sapiens major histocompatibility complex, class II, DR51 haplotype (DR51) on chromosome 6	NG_002432
HLA_DRB1*1602	Homo sapiens MHC class II antigen (HLA-DRB1) gene DR51	AB774985
HLA_DRB3*01012	Homo sapiens major histocompatibility complex, class II, DR52 haplotype (DR52) on chromosome 6	NG_002392
HLA_DRB3*020201	Human DNA sequence from clone DAQB-97F8 on chromosome 6	AL929581
HLA_DRB4*01030101	Homo sapiens major histocompatibility complex, class II, DR53 haplotype (DR53)	NG_002433
HLA_DRB5*0101	Homo sapiens major histocompatibility complex, class II, DR51 haplotype (DR51) on chromosome 6	NG_002432
HLA_DRB6ps	Homo sapiens major histocompatibility complex, class II, DR51 haplotype (DR51) on chromosome 6	NG_002432
HLA_DRB7ps	Homo sapiens major histocompatibility complex, class II, DR53 haplotype (DR53)	NG_002433
HLA_DRB1*03	c91958-91273 Human DNA sequence from clone DAQB-97F8 on chromosome 6	AL929581
SEQUENCES HERE REPORTED		
Aovo-DRB1*03:05	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447756
Aovo-DRB1*03:06	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447757
Aovo-DRB1*03:07	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447758
Aovo-DRB*W91:01	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447733
Aovo-DRB*W92:01	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447735
Aovo-DRB*W92:02	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447736
Aovo-DRB*W91:02	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447737
Aovo-DRB*W93:01	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447740
Aovo-DRB*W88:01	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447741
Aovo-DRB*W29:01	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447742
Aovo-DRB1*03:04	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447759
Aovo-DRB*W18:01	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447762
Aovo-DRB*W18:02	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447763
Aovo-DRB*W18:03	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447764
Aovo-DRB*W90:01	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447765
Aovo-DRB3*06:01	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447766
Aovo-DRB*W30:01	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447738
Aovo-DRB*W45:01	Aotus vociferans MHC class II antigen beta chain (Aovo-DRB) gene	KF447739
Aona-DRB*W91:01	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447734
Aona-DRB*W29:10	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447743
Aona-DRB*W29:08	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447744
Aona-DRB*W30:02	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447745
Aona-DRB1*03:28	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447746
Aona-DRB*W89:01	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447747
Aona-DRB3*06:25:01	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447748
Aona-DRB3*06:26	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447749
Aona-DRB3*06:27	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447750
Aona-DRB3*06:25:02	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447751
Aona-DRB3*06:15	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447752
Aona-DRB3*06:28	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447753
Aona-DRB1*03:29	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447754
Aona-DRB1*03:17:01	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447755
Aona-DRB*W18:08	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447760
Aona-DRB*W18:06	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene	KF447761

Sequence	Remarks	Accesion Number Genbank
Sequences of Aotus MHC-DRB (molecular phylogenetic analysis exon 2)		
Aoaz-DRB*W3801	Aotus azarai MHC class II antigen (Aoaz-DRB) gene Aoaz-DRB*W3801 allele	AY429143
Aoaz-DRB3*0601	Aotus azarai MHC class II antigen (Aoaz-DRB3) gene Aoaz-DRB3*0601 allele	AY429142
Aona-DRB*W1301	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene Aona-DRB*W1301 allele	AF132767
Aona-DRB*W1305	Aotus nancymaae isolate 20896_2 MHC class II antigen beta chain (DRB) mRNA DRB*W1305 allele	AY563223
Aona-DRB*W1306	Aotus nancymaae isolate 21100_1 MHC class II antigen beta chain (DRB) mRNA DRB*W1306 allele	AY563218
Aona-DRB*W1309	Aotus nancymaae isolate 22417-7 MHC class II antigen beta chain (DRB) mRNA DRB*W1309 allele	AY563255
Aona-DRB*W1312	Aotus nancymaae MHC class II antigen (DRB) mRNA DRB*W1312 allele	DQ162705
Aona-DRB*W1801	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene Aona-DRB*W1801 allele	AF132768
Aona-DRB*W1802	Aotus nancymaae MHC-DRB (DRB*W) mRNA DRB*W1802 allele	AF169487
Aona-DRB*W2901	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene Aona-DRB*W2901 allele	AF129806
Aona-DRB*W2906	Aotus nancymaae MHC class II antigen (DRB) mRNA DRB*W2906 allele	DQ162688
Aona-DRB*W2907	Aotus nancymaae isolate 20894_3 MHC class II antigen beta chain (DRB) mRNA DRB*W2907 allele	AY563201
Aona-DRB*W3001	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB) gene Aona-DRB*W3001 allele	AF132766
Aona-DRB*W3801	Aotus nancymaae isolate 16606_7 MHC class II antigen beta chain (DRB) mRNA DRB*W3801 allele	AY563194
Aona-DRB*W4201	Aotus nancymaae isolate 20337_3 MHC class II antigen beta chain (DRB) mRNA DRB*W4201 allele	AY563209
Aona-DRB*W4401	Aotus nancymaae isolate 20559_2 MHC class II antigen beta chain (DRB) mRNA DRB*W4401 allele	AY563206
Aona-DRB*W4501	Aotus nancymaae isolate 20249_12 MHC class II antigen beta chain (DRB) mRNA DRB*W4501 allele	AY563180
Aona-DRB*W4701	Aotus nancymaae isolate 20465_3 MHC class II antigen beta chain (DRB) mRNA DRB*W4701 allele	AY563181
Aona-DRB*W4702	Aotus nancymaae isolate 22822_13 MHC class II antigen beta chain (DRB) mRNA DRB*W4702 allele	AY563183
Aona-DRB*W470404	Aotus nancymaae MHC class II antigen (DRB) mRNA DRB*W470404 allele	DQ162645
Aona-DRB1*0301	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB1) gene Aona-DRB1*0301 allele	AF129793
Aona-DRB1*0302	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB1) gene Aona-DRB1*0302 allele	AF129792
Aona-DRB1*0303	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB1) gene Aona-DRB1*0303 allele	AF129794
Aona-DRB1*0305	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB1) gene Aona-DRB1*0305 allele	AF129796
Aona-DRB1*0307	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB1) gene Aona-DRB1*0307 allele	AF129798
Aona-DRB1*0313	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB1) gene Aona-DRB1*0313 allele	AF132760
Aona-DRB1*0314	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB1) gene Aona-DRB1*0314 allele	AF132761
Aona-DRB1*0319	Aotus nancymaae isolate 20719_2 MHC class II antigen beta chain (DRB1) mRNA DRB1*0319 allele	AY563188
Aona-DRB1*0324	Aotus nancymaae isolate 21955_10 MHC class II antigen beta chain (DRB1) mRNA DRB1*0324 allele	AY563193
Aona-DRB3*0601	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB3) gene Aona-DRB3*0601 allele	AF129799
Aona-DRB3*0602	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB3) gene Aona-DRB3*0602 allele	AF129800
Aona-DRB3*0603	Aotus nancymaae MHC class II antigen beta chain (Aona-DRB3) gene Aona-DRB3*0603 allele	AF129801
Aona-DRB3*0614	Aotus nancymaae isolate 20444_7 MHC class II antigen beta chain (DRB3) mRNA DRB3*0614 allele	AY563212
Aoni-DRB*W1301	Aotus nigriceps isolate 21921_5 MHC class II antigen beta chain (DRB) mRNA DRB*W1301 allele	AY563261
Aoni-DRB*W1303	Aotus nigriceps MHC class II antigen (DRB) mRNA DRB*W1303 allele	DQ162732
Aoni-DRB*W2901	Aotus nigriceps isolate 20596_8 MHC class II antigen beta chain (DRB) mRNA DRB*W2901 allele	AY563259
Aoni-DRB*W2902	Aotus nigriceps isolate 21919_18 MHC class II antigen beta chain (DRB) mRNA DRB*W2902 allele	AY563246
Aoni-DRB*W3801	Aotus nigriceps isolate 16584_5 MHC class II antigen beta chain (DRB) mRNA DRB*W3801 allele	AY563245
Aoni-DRB*W4201	Aotus nigriceps isolate 20483_1 MHC class II antigen beta chain (DRB) mRNA DRB*W4201 allele	AY563253
Aoni-DRB*W4301	Aotus nigriceps isolate 20848_16 MHC class II antigen beta chain (DRB) mRNA DRB*W4301 allele	AY563249
Aoni-DRB*W4401	Aotus nigriceps isolate 20596_4 MHC class II antigen beta chain (DRB) mRNA DRB*W4401 allele	AY563247
Aoni-DRB1*0301	Aotus nigriceps MHC class II antigen beta chain (Aoni-DRB1) gene Aoni-DRB1*0301 allele	AF129797
Aoni-DRB1*0304	Aotus nigriceps isolate 21791_15 MHC class II antigen beta chain (DRB1) mRNA DRB1*0304 allele	AY563242
Aoni-DRB1*0307	Aotus nigriceps MHC class II antigen (DRB1) mRNA DRB1*0307 allele	DQ162711
Aoni-DRB1*W1801	Aotus nigriceps isolate 20456_8 MHC class II antigen beta chain (DRB1) mRNA DRB1*W1801 allele	AY563257
Aoni-DRB3*0601	Aotus nigriceps isolate 20506_2 MHC class II antigen beta chain (DRB3) mRNA DRB3*0601 allele	AY563229
Aotr-DRB*W1801	Aotus trivirgatus MHC class II DRB*W1801 gene exon	L12477
Aotr-DRB1*0301	Aotus trivirgatus MHC class II DRB1*0301 gene exon	L12472
Aotr-DRB1*0303	Aotus trivirgatus partial DRB1 gene for MHC class II antigen DRB1*0303 allele exon 2	AJ544176
Aotr-DRB3*0602	Aotus trivirgatus partial DRB3 gene for MHC class II antigen DRB3*0602 allele exon 2	AJ544174
Aotr-DRB3*0603	Aotus trivirgatus partial DRB3 gene for MHC class II antigen DRB3*0603 allele exon 2	AJ544175
Aotr-DRB3*06	Aotus trivirgatus MHC class II DRB gene exon	L12474
Aovo-DRB*W130101	Aotus vociferans MHC class II antigen (DRB) mRNA DRB*W130101 allele	DQ162634
Aovo-DRB*W1303	Aotus vociferans isolate 20789_1 MHC class II antigen beta chain (DRB) mRNA DRB*W1303 allele	AY563258
Aovo-DRB*W4301	Aotus vociferans MHC class II antigen (DRB) mRNA DRB*W4301 allele	DQ162630
Aovo-DRB*W4701	Aotus vociferans isolate 16704_9 MHC class II antigen beta chain (DRB) mRNA DRB*W4701 allele	AY563227
Aovo-DRB1*0301	Aotus vociferans MHC class II antigen (DRB1) mRNA DRB1*0301 allele	DQ162628

Sequence	Remarks	Accesion Number Genbank
Sequences of primates MHC-DRB (molecular phylogenetic analysis exon 2 + intron 2)		
Caja_DRB_G_01	43380-44258 Callithrix jacchus BAC clone CH259-77F15 from chromosome unknown, complete sequence	AC242730
Caja_DRB_G_02	105908-107178 Callithrix jacchus BAC clone CH259-77F15 from chromosome unknown, complete sequence	AC242730
Caja_DRB_G_03	192613-193403 Callithrix jacchus BAC clone CH259-77F15 from chromosome unknown, complete sequence	AC242730
Caja_DRB_G_04	138963-138357 Callithrix jacchus BAC clone CH259-15O7 from chromosome unknown, complete sequence	AC243457
Caja_DRB_G_05	76772-75338 Callithrix jacchus BAC clone CH259-15O7 from chromosome unknown, complete sequence	AC243457
Caja_DRB_G_06	161835-162584 Callithrix jacchus BAC clone CH259-49P2 from chromosome unknown, complete sequence	AC242576
SAOE-DRB1*0303	Saguinus oedipus MHC class II antigen (SAOE-DRB1) pseudogene	AF173332
SAOE-DRB3*0501	Saguinus oedipus MHC class II antigen (SAOE-DRB3) pseudogene	AF173333
SAOE-DRB11*0102	Saguinus oedipus MHC class II antigen (SAOE-DRB11) gene	AF173334
SAOE-DRB11*0105	Saguinus oedipus MHC class II antigen (SAOE-DRB11) gene	AF173335
SAOE-DRB*W2209	Saguinus oedipus MHC class II antigen (SAOE-DRB) gene	AF173336
CAJA-DRB1*0304	Callithrix jacchus MHC class II antigen (CAJA-DRB1) gene	AF173337
CAMO-DRB1*0302	Callicebus moloch MHC class II antigen (CAMO-DRB1) gene	AF173338
CAMO-DRB3*0503	Callicebus moloch MHC class II antigen (CAMO-DRB3) gene	AF173339
CAMO-DRB3*0504	Callicebus moloch MHC class II antigen (CAMO-DRB3) pseudogene	AF173340
CEAP-DRB*W1301	Cebus apella MHC class II antigen (CEAP-DRB) gene	AF173341
CAJA-DRB*W1201	Callithrix jacchus MHC class II antigen (CAJA-DRB) gene	AF173348
Mamu_DRB_G_01	140068-140822 Macaca mulatta Major Histocompatibility Complex BAC MMU012K22	AC148663
Mamu_DRB_G_02	28917-29606 Macaca mulatta Major Histocompatibility Complex BAC MMU012K22	AC148663
Mamu_DRB_G_03ps	111593-112214 Macaca mulatta Major Histocompatibility Complex BAC MMU012K22	AC148663
Mamu_DRB_G_05	c75283-74620 Macaca mulatta Major Histocompatibility Complex BAC MMU248N17	AC148697
Mamu_DRB_G_06	c151148-150454 Macaca mulatta Major Histocompatibility Complex BAC MMU248N17	AC148697
Mamu_DRB_G_07	c29018-28359 Macaca mulatta Major Histocompatibility Complex BAC MMU248N17	AC148697
Mamu_DRB_G_08	c26068-25409 Macaca mulatta Major Histocompatibility Complex BAC MMU281E18	AC148700
Mamu_DRB_G_04	173697-174407 Macaca mulatta Major Histocompatibility Complex BAC MMU370O02	AC148706
Chae_DRB_G_01	63309-64000 Chlorocebus aethiops BAC clone CH252-249I23 from chromosome 6	AC241608
Chae_DRB_G_02	143590-144262 Chlorocebus aethiops BAC clone CH252-249I23 from chromosome 6	AC241608
Chae_DRB_G_03	111557-112224 Chlorocebus aethiops BAC clone CH252-249I23 from chromosome 6	AC241608
Chae_DRB_G_04	21256-21901 Chlorocebus aethiops BAC clone CH252-249I23 from chromosome 6	AC241608
MAFA-DRB*W3301	Macaca fascicularis MHC class II antigen (MAFA-DRB) gene	AF173349
MAAR-DRB1*0301	Macaca arctoides MHC class II antigen (MAAR-DRB1) gene	AF173350
MAAR-DRB1*0302	Macaca arctoides MHC class II antigen (MAAR-DRB1) gene	AF173351
MAAR-DRB1*0701	Macaca arctoides MHC class II antigen (MAAR-DRB1) gene	AF173352
MAFA-DRB*W301	Macaca fascicularis MHC class II antigen (MAFA-DRB) gene	AF173353
MAMU-DRB*W402	Macaca mulatta MHC class II antigen (MAMU-DRB) gene	AF173354
MAAR-DRB*W601	Macaca arctoides MHC class II antigen (MAAR-DRB) pseudogene	AF173355
MAAR-DRB5*0301	Macaca arctoides MHC class II antigen (MAAR-DRB5) gene	AF173356
MAMU-DRB6*0106	Macaca mulatta MHC class II antigen (MAMU-DRB6) pseudogene	AF173357
Mamu-DRB*W201	7681-8413 Macaca mulatta partial Mamu-DRB gene for MHC class II antigen	AM910410
Mamu-DRB*W305	7704-8400 Macaca mulatta partial Mamu-DRB gene for MHC class II antigen	AM910411
Mamu-DRB*W603	6501-7247 Macaca mulatta partial Mamu-DRB gene for MHC class II antigen	AM910412
Mamu-DRB*W2507	6540-7190 Macaca mulatta partial Mamu-DRB gene for MHC class II antigen	AM910413
Mamu-DRB1*0303	3087-3748 Macaca mulatta partial Mamu-DRB gene for MHC class II	AM910414
Mamu-DRB1*0306	Macaca mulatta partial Mamu-DRB1 gene for MHC class II antigen	AM910415
Mamu-DRB1*0309	7622-8327 Macaca mulatta partial Mamu-DRB1 gene for MHC class II antigen	AM910417
Mamu-DRB1*0404	2994-3688 Macaca mulatta partial Mamu-DRB1 gene for MHC class II antigen	AM910419
Mamu-DRB1*1007	8014-8684 Macaca mulatta partial Mamu-DRB1 gene for MHC class II antigen	AM910420
Mamu-DRB3*0405	Macaca mulatta partial Mamu-DRB3 gene for MHC class II antigen	AM910421
Mamu-DRB3*0408	2995-3667 Macaca mulatta partial Mamu-DRB3 gene for MHC class II antigen	AM910422
Mamu-DRB5*0301	976-8612 Macaca mulatta partial Mamu-DRB5 gene for MHC class II antigen	AM910423

Sequence	Remarks	Accesion Number Genbank
Sequences of primates MHC-DRB (molecular phylogenetic analysis exon 2 + intron 2)		
HLA_DRB1*010101	Homo sapiens HLA-DRB1 gene for MHC class II antigen	AM493435
HLA_DRB1*010201	Homo sapiens voucher Coriell Cell Repository DNA sample NA01018	AY663400
HLA_DRB1*040101	Homo sapiens major histocompatibility complex, class II, DR53 haplotype (DR53)	NG_002433
HLA_DRB1*0405	Homo sapiens MHC class II antigen (HLA-DRB1) gene, HLA-DRB1*0404	AB715390
HLA_DRB1*070101	Human DNA sequence from clone DADB-102D14 on chromosome 6	CR753309
HLA_DRB1*08:03:02	Homo sapiens HLA-DRB1 gene for MHC class II antigen	FN823238
HLA_DRB1*110101	Homo sapiens voucher Coriell Cell Repository DNA sample NA00576	AY663412
HLA_DRB1*110401	Homo sapiens voucher Coriell Cell Repository DNA sample NA14661	AY663394
HLA_DRB1*12:01:01	Homo sapiens HLA-DRB1 gene for major histocompatibility complex	AB715399
HLA_DRB1*130201	Homo sapiens voucher Coriell Cell Repository DNA sample NA14663	AY663413
HLA_DRB1*140101	Homo sapiens voucher Coriell Cell Repository DNA sample NA10540	AY663405
HLA_DRB1*140501	Homo sapiens voucher Coriell Cell Repository DNA sample NA04535	AY663408
HLA_DRB1*1501	Homo sapiens major histocompatibility complex, class II, DR51 haplotype (DR51) on chromosome 6	NG_002432
HLA_DRB1*16	Homo sapiens major histocompatibility complex, class II	NG_002432
HLA_DRB1*1602	Homo sapiens MHC class II antigen (HLA-DRB1) gene DR51	AB774985
HLA_DRB3*01012	Homo sapiens major histocompatibility complex, class II, DR52 haplotype (DR52) on chromosome 6	NG_002392
HLA_DRB3*020201	Human DNA sequence from clone DAQB-97F8 on chromosome 6	AL929581
HLA_DRB4*01030101	Homo sapiens major histocompatibility complex, class II, DR53 haplotype (DR53)	NG_002433
HLA_DRB5*0101	Homo sapiens major histocompatibility complex, class II, DR51 haplotype (DR51) on chromosome 6	NG_002432
HLA_DRB6ps	Homo sapiens major histocompatibility complex, class II, DR51 haplotype (DR51) on chromosome 6	NG_002432
HLA_DRB7ps	Homo sapiens major histocompatibility complex, class II, DR53 haplotype (DR53)	NG_002433
HLA_DRB1*10:01:01	Homo sapiens MHC class II antigen (HLA-DRB1) gene	JN157606
HLA_DRB1*15:02:01	Homo sapiens HLA-DRB1 gene for MHC class II antigen	AB774991
HLA_DRB1*03-AL929581	c91958-91273 Human DNA sequence from clone DAQB-97F8 on chromosome 6	AL929581
Patr-DRB1*020101	Pan troglodytes partial Patr-DRB1 gene for MHC class II antigen, Patr-DRB1*020101	AM910425
Patr-DRB3*0208	Pan troglodytes partial Patr-DRB3 gene for MHC class II antigen	AM910428
Patr-DRB1*10:01	Pan troglodytes verus partial patr-DRB1 gene for MHC class II antigen	HE800526
Gogo_DRB_01	44822-45507 Gorilla gorilla voucher Coriell Cell Repository DNA sample NG05251	AY663402
Patr-DRB*W902	Pan troglodytes partial Patr-DRB gene for MHC class II antigen	AM910424
Patr-DRB3*01020101	Pan troglodytes partial Patr-DRB3 gene for MHC class II antigen	AM910426
Patr_DRB_01	Pan troglodytes voucher Coriell Cell Repository DNA sample NS03646	AY663401
Gogo_DRB_02	c218276-217597 Gorilla DNA sequence from clone CH255-114D6, complete sequence	CU104652
Gogo_DRB_03	c218276-217597 Gorilla DNA sequence from clone CH255-114D6, complete sequence	CU104652
Gogo_DRB_04	73853067:151462-152106 Gorilla DNA sequence from clone CH255-351B13	CT025711
Gogo_DRB_05	73853067:151462-152106 Gorilla DNA sequence from clone CH255-351B13	CT025711
Patr-DRB4*02:01	5938-6579 Pan troglodytes troglodytes partial patr-DRB4	HE800525
Poab_DRB_01	200269-200926 Pongo abelii BAC clone CH276-191M9 from chromosome 6	AC206450
Patr_DRB_01	27628-28281 Pan troglodytes genomic DNA, chromosome 5	AP006503

Table S2. Microsatellite sequence and length in Platyrrhini MHC-DRB.

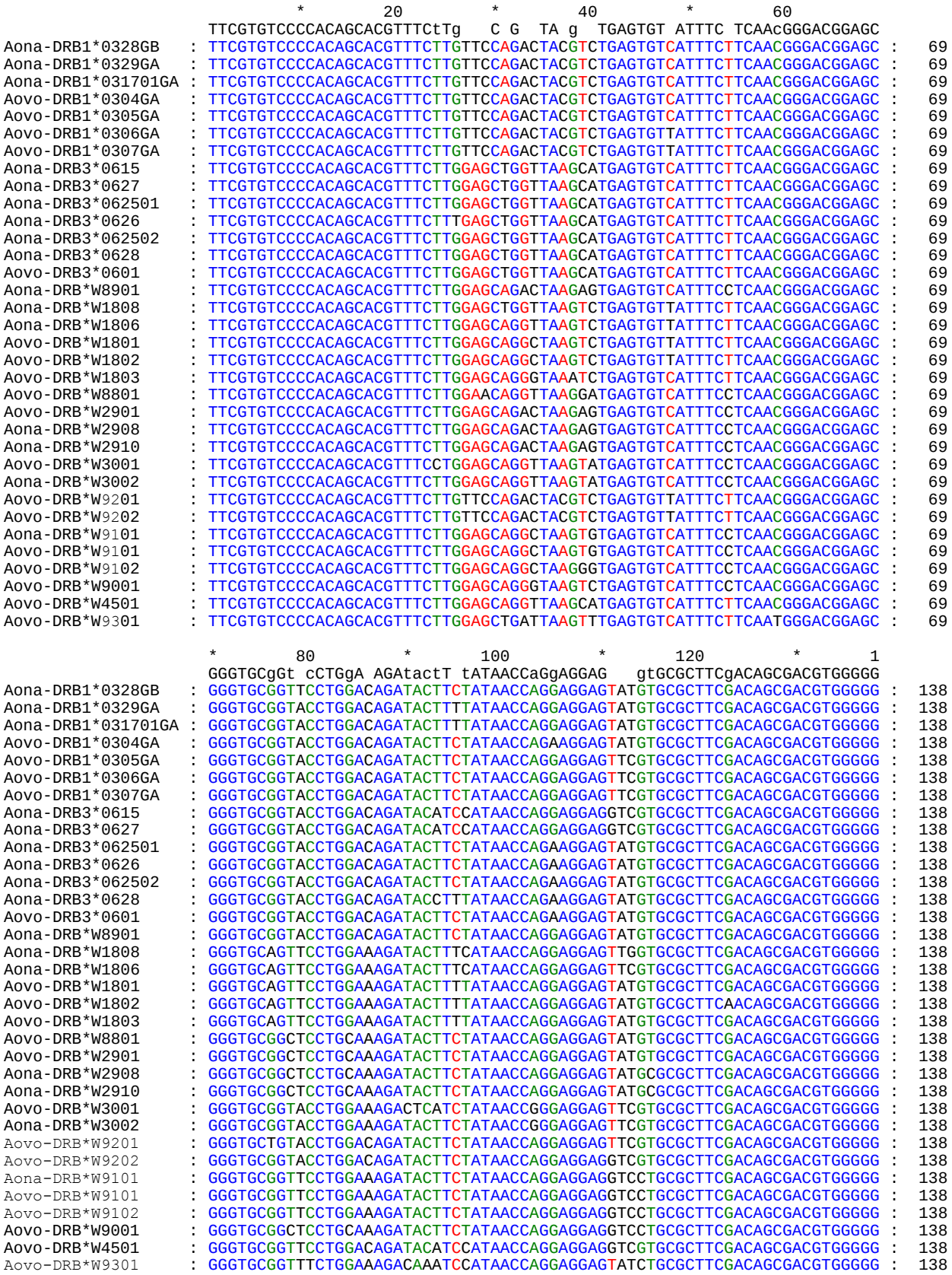
SEQUENCE	MICROSATELLITE			LENGTH	LZSize (Bytes)
	FIRST PART	CENTRAL PART	FINAL PART		
Aona DRB3*062501	CA(GA)4CT(GA)4CT(GA)4CC	(GA)5CACA(GA)2CACT(GA)4CA(GA)6CT(GA)4CAGT((GA)3,3,4CT)(GA)4CACT(GA)4CACAGT(GA)2CA(GA)2AACA(GA)2AAGACT(GA)3TACACT(GA)4CA(GA)2CT(GA)3CA(GA)3CACA(GA)3CA(GA)3CTGAAAGACT(GA)4CA((GA)3,4CT)(GA)3AAGACT((GA)4,3CT)(GA)3AACA((GA)2,3CT)(GA)3CACT(GA)3CT(GA)3CACAGATA(GA)2TT((GA)3,5CT)(GA)3CACTGATA((GA)2,5,3CT)	(GA)15CT(GA)3GCGCCTTG	314	4333
Aovo DRB3*0601	CA(GA)4CT(GA)4CT(GA)4CC	(GA)5(CA)2(GA)2CACT(GA)4CAGAG(GA)4CT(GA)4CAGT((GA)3,3,4CT)(GA)4CACT(GA)4CACAGT(GA)2CA(GA)2AACA(GA)2AAGACT (GA)3TACACT(GA)4CA(GA)2CT(GA)3CA(GA)3CACA(GA)3CA(GA)3CTGAAAGACT(GA)4CA((GA)3,4CT)(GA)3AAGACT((GA)4,3)(GA)3AACA((GA)2,3CT)(GA)3CACT(GA)3CT(GA)3 CACAGATA(GA)2TT((GA)3,5CT)(GA)3CACTGATA((GA)2,5,3CT)	(GA)15CT(GA)3GCGCCTTG	310	4343
Aona DRB3*062502	CA(GA)4CT(GA)4CC(GA)4CC	(GA)5CACA(GA)2CACT(GA)4CA(GA)6CT(GA)4CAGT((GA)3,3,4CT)(GA)4CACT(GA)4CACAGT(GA)2CA(GA)2AACA(GA)2AAGACT(GA)3TACACT(GA)4CA(GA)2CT((GA)3,3CA)CA(GA)3CA(GA)3CTGAAAGACT(GA)4CA((GA)3,4CT)GAAAGACT((GA)4,3CT)(GA)3AACA((GA)2,3CT)(GA)3CACT(GA)3CT(GA)3CACAGATA(GA)2TT((GA)3,5CT)(GA)3CACTGATA(GA)2CT	(GA)16CT(GA)3GCGCCTTG	294	4310
Aona DRB3*0626	CA(GA)4CT(GA)4CC	(GA)5CACA(GA)2CACT(GA)4CA(GA)5CT(GA)4CAGT((GA)3,3,4CT)(GA)4CACT(GA)4CACAGT(GA)2CA(GA)2AACA(GA)2AAGACT(GA)3TACACT(GA)4CA(GA)2CT((GA)3,3CA)CA(GA)3CA(GA)3CTGAAAGACT(GA)4CA((GA)3,4CT)(GA)3AAGACT((GA)4,3CT)(GA)3AACA((GA)2,3CT)(GA)3CACT(GA)3CT(GA)3CACAGATA(GA)2TT((GA)3,5CT)(GA)3CACTGATA((GA)2,5,3CT)	(GA)15CT(GA)3GCGCCTTG	312	4357
Aona DRB3*0628	CA(GA)4CT(GA)4CC	(GA)5CACA(GA)2CACT(GA)4CA(GA)6CT(GA)4CAGT((GA)3,3,4CT)(GA)4CACT(GA)4CAGAG((GA)2,3CA)(GA)2AAGACT(GA)3CA(GA)2CT(GA)3TACACT(GA)4CA(GA)2CT((GA)3,3CA)CT(GA)4CA(GA)3CTGAAAGAT(GA)4CA((GA)4,5,3CT)(GA)3TACTTAGG(GA)2CT(GA)4CA(GA)3CT(GA)3AAGACT((GA)4,3,3CT)(GA)3AACA((GA)2,3CT)(GA)3CACT(GA)3CT(GA)3CACAGATA(GA)2TT(GA)3CT(GA)3CACTGATA((GA)5,3)	(GA)11AACT(GA)3GCACTTG	346	4855
Aona DRB3*0627	CA(GA)4CT(GA)4CC	(GA)5CACA(GA)2CACT(GA)4CA(GA)6CT(GA)4CAGT((GA)3,3,4CT)(GA)4CACT(GA)4CAGAG((GA)2,3CA)(GA)2AAGACT(GA)3CA(GA)4CT(GA)3TACACT(GA)4CA(GA)2CT((GA)3,3CA)CT(GA)4CA(GA)3CTGAAAGACTCA(GA)4CA(GA)4CT(GA)4GGCT(GA)3CT(GA)3TACTTAGG(GA)2CT(GA)4CA(GA)3CT(GA)3AAGACT((GA)4,3,3CT)(GA)3AACA((GA)2,3CT)(GA)3CACT(GA)3CT(GA)3CACAGATA(GA)2TT(GA)3CT(GA)3CACTGATA((GA)2,5,3CT)	(GA)10AACT(GA)3GCACTTG	350	4951
Aona DRB3*0615	CA(GA)4CT(GA)4CC	(GA)5CACA(GA)2CACT(GA)4CA(GA)6CT(GA)4CAGT((GA)3,3,4CT)(GA)4CACT(GA)4CACAGT(GA)2CA(GA)2AACA(GA)2AAGACT(GA)3TACACT(GA)4CA(GA)2CT((GA)3,3CA)CA(GA)3CA(GA)3CTGAAAGACT(GA)4CA((GA)3,4CT)(GA)3AAGACT((GA)4,3CT)(GA)3AACA((GA)2,3CT)(GA)3CACT(GA)3CT(GA)3CACAGATA (GA)2TT((GA)3,5CT)(GA)3CACTGATA((GA)2,5,3CT)	(GA)15CT(GA)3GCGCCTTG	314	4333
Aona DRB1*0329GA	CA(GA)4CTGAAA(GA)2CT(GA)4CACTGAGT	(GA)2GT((GA)2,3CA)(GA)2AAGACT(GACA)2(GA)4CT(GA)2GGTACACT(GA)4CAGAGT((GA)2,3CA)GAGTAAGACTGACA(GA)4CT(GA)3TACACT(GA)4CA(GA)2CT((GA)3,3CA)CT(GA)4CA(GA)3CTGAAAGACT(GA)4CA((GA)4,5,3CT)(GA)3TACTTAGG(GA)2CT(GA)4CA(GA)3CT(GA)2CAGACT(GA)3AAGACT(GA)6CT(GA)3GGCT(GA)5CTGAGG(GA)2CA(GA)2AT(GA)5CT(GA)3CAAGACA(GA)2CT(GA)4T(GA)2CA((GA)4,4CT)AA (GA)3CTGAGG((GA)2,4CT)GAAAGACT(GA)5CT(GA)4CAAGACA(GA)2CT(GA)4T(GA)2CA(GA)4CT(GA)3 AAGACT(GA)3CA((GA)2,3,2CT)(GA)3AATAGACT(GA)3CT(GA)3CACT(GA)3CT(GA)3CACAGATA(GA)2TT ((GA)3,5CT)(GA)3CACTGATA((GA)3,5,3CT)	(GA)23CT(GA)3GCGCCGTG	446	7249
Aona DRB1*031701	CA(GA)4CT(GA)4CACT	(GA)4GT((GA)2,3CA)(GA)2AAGACTGACAGACA(GA)4CT(GA)2GGTACACT(GA)4CAGAGT((GA)2,3CA)GAGTAAGACTGACA(GA)4CT(GA)3TACACT(GA)4CA(GA)2CT(GA)3CAGC(GA)2CACT(GA)4CA(GA)3CTGAAAGACT(GA)4CA((GA)4,5,3CT)(GA)3TACTTAGG(GA)2CT(GA)4CA(GA)3CT(GA)2CAGACT(GA)3AAGACT(GA)6CT(GA)3GGCT(GA)5CTGAGG(GA)2CC(GA)2AT(GA)5CT(GA)3CAAGACA(GA)2CT(GA)4T(GA)2CA((GA)4,4CT)AA(GA)3CTGAGG((GA)2,4CT)GAAAGACT(GA)5CT(GA)4CAAGACA(GA)2CT(GA)4T(GA)2CA(GA)4CT(GA)3AAGACT(GA)3CA((GA)2,3CT)(GA)3AATAGACT(GA)3CT(GA)3CACT(GA)3CT(GA)3CACAGATA(GA)2TT((GA)3,5CT)(GA)3CACTGATA ((GA)2,5,3CT)	(GA)25CT(GA)3GCGCCGTG	462	7205

Aovo DRB1*0304	CA(GA)4CT(GA)4CC(GA)5CACA	(GA)2CACT(GA)4CA(GA)6CT(GA)4CAGT((GA)3,3CT)(GA)4CACT(GA)4CAGAGT(GA)2CA(GA)2AACA(GA)2AAGACT (GA)3TACACT(GA)4CA(GA)2CT((GA)3,3CA)CT(GA)4CA(GA)3CTGAAAGACT(GA)4CA(GA)3CT(GA)3AAGACT((GA)4 ,3CT)(GA)3AACA((GA)2,3CT)(GA)3CACT(GA)3CT(GA)3CACAGATA(GA)2TT(GA)3CT(GA)4CACTGATA((GA)2,4,3CT)	(GA)11AACT(GA)3GCACTTG	274	3949
Aovo DRB1*0305	CA(GA)4CT(GA)4CC(GA)5CACA	(GA)2CACT(GA)4CA(GA)6CT(GA)4CAGT((GA)3,3,2CT)(GA)4CACT(GA)4CAGAGG((GA)2,3CA)(GA)2AAGACT(GA)3CA(GA)3CT(GA)3 TACACT(GA)4CA(GA)2CT(GA)3CAGG(GA)2CACT(GA)4CA(GA)3CTGAAAGACT(GA)4CA(GA)4CT((GA)5,3CT)(GA)3TACTTAGG(GA)2CT (GA)4CA(GA)3CT(GA)3AAGACTCT(GA)4CT(GA)3AACAGGGACT(GA)3CT(GA)3CACT(GA)3CT(GA)3CACAGATAGAAATT(GA)3CT(GA)3CACTGATA ((GA)2,5,3CT)(GA)3CA(GA)2CT	(GA)9AACT(GA)3GCACTTG	336	4282
Aovo DRB1*0306	CA(GA)4CT(GA)4CC(GA)5CACA	(GA)2CACT(GA)4CA(GA)6CT(GA)4CAGT((GA)3,3CT)(GA)4CT(GA)4CACT(GA)4CAGAGT(GA)2CA(GA)2 AACA(GA)2AAGACT(GA)2AATACACT(GA)4CA(GA)2CT((GA)3,3CA)CT(GA)4CA(GA)3CTGAAAGACT (GA)4CA(GA)3CT(GA)3AAGACT((GA)4,3CT)(GA)3AACA((GA)2,3CT)(GA)3CACT(GA)3CT(GA)3CACAGATA (GA)2TT(GA)3CT(GA)3CACTGATA((GA)2,4,3CT)	(GA)13AACT(GA)3GCACTTG	282	4102
Aovo DRB1*0307	CA(GA)4CT(GA)4CC(GA)7CACA	(GA)2CACT(GA)4CA(GA)6CT(GA)4CAGT((GA)3,3,4CT)(GA)4CACT(GA)4CAGAGT(GA)2CA(GA)2AACA (GA)2AAGACT(GA)2AATACACT(GA)4CA(GA)2CT((GA)3,3CA)CT(GA)4CA(GA)3CTGAAAGACT(GA)4CA (GA)3CT(GA)3AAGACT((GA)4,3CT)(GA)3AACA((GA)2,3CT)(GA)3CACT(GA)3CT(GA)3CACAGATA(GA)2TT (GA)3CT(GA)3CACTGATA((GA)2,4,3CT)	(GA)13AACT(GA)3GCACTTG	286	4096
Aona DRB*W8901	CA(GA)2CA(GA)2CT(GA)4CT	(GA)7AT((GA)4,5,4CT)(GA)3TA(GA)2CTGAAAGAC(GA)2CA(GA)3AACT(GA)3AACATA(GA)2CT(GA)5CACA	(GA)5CAGAGC(GA)4GCG	114	1833
Aona DRB*W1808	CA(GA)5CT(GA)4CT(GA)10CACA	(GA)4TAGACA(GA)2CA((GA)3,3CT)GT((GA)4,4CT)GTGCAAGACC((GA)5,2CT)(GA)3TT(GA)4CA(GA)2CT (GA)3GGCT(GA)3CTCG(GA)3CTGAGC(GA)3CT(GA)2TAGACT(GA)3AACT(GA)3TACT	(GA)6(GC)3GACAGCG	168	2893
Aona DRB*W1806	CA(GA)5CT(GA)4CT(GA)10CACA	(GA)4TAGACA(GA)2CA((GA)3,3CT)GT(GA)4CTGTGCAAGACC((GA)5,2CT)(GA)3TT(GA)4CA((GA)2,4CT) (GA)3CTCG(GA)3CTGAGC(GA)3CT(GA)2TAGACT(GA)3AACT(GA)3TACT	(GA)6(GC)3GACAGCG	162	2729
Aovo DRB*W1803	CA(GA)5CT(GA)15CACA	(GA)4TAGACA(GA)2CA((GA)3,3CT)GT(GA)4CT(GA)4CTGTGCAAGACC((GA)5,2,4CT)(GA)3CTCG(GA)3 CTGAGC(GA)3CT(GA)2TAGACT(GA)3AACT(GA)3TACT	(GA)6(GC)3GACAGCG	154	2479
Aovo DRB*W1801	CA(GA)7CACA	(GA)4TAGACA(GA)2CA((GA)3,3CT)GT((GA)4,4CT)GTGCAAGACC(GA)5TT(GA)2CT(GA)2TT(GA)4CA ((GA)2,4CT)(GA)3CTCG(GA)3CTGAGC(GA)3CT(GA)2TAGACT(GA)3AACT(GA)3TACT	(GA)6(GC)3GACAGCG	144	2591
Aovo DRB*W1802	CA(GA)5CT(GA)4CT(GA)13CACA	(GA)4TAGACA((GA)2,3CT)GT(GA)4CTGTGCAAGACC((GA)5,2CT)(GA)3TT(GA)4CA((GA)2,4CT)(GA)3 CTCG(GA)3CTGAGC(GA)3CT(GA)2TAGACT(GA)3AACA(GA)2TACT	(GA)6(GC)3GACAGCG	160	2663
Aona DRB1*0328	CA(GA)4CT(GA)4CC	(GA)5CACA(GA)2CACT(GA)4CA(GA)6CT(GA)4CAGT((GA)3,3,2CT)(GA)4CACT(GA)4CAGAGT((GA)2,3CA)(GA)2AAGGCTGACAGACA (GA)4CT(GA)3TACACT(GA)4CA(GA)2CT(GA)3CAGACACTCAGACACT((GA)4,3CT)GT(GA)5CT(GA)4CA(GA)4CT(GA)4CACT(GA)4CA (GA)3CTGAAAGACT(GA)3TACTTAGG(GA)2CT(GA)3AACA((GA)3,4CT)(GA)3AAGACT(GA)5CT(GA)3GGCT(GA)3CTGAGG(GA)2CA(GA)2AT(GA)5CT(GA)3TAAGACA(GA)2CT(GA)4T(GA)2CA(GA)4CT(GA)3AAGACT(GA)3CA((GA)3,3CT)(GA)4CA((GA)2,3,3CT)(GA)3CACAGATA(GA)2TT((GA)3,5CT)GGGGGGG	(GA)11CT(GA)3GCACGTG	414	6113

Aovo DRB*W4501	CA(GA)2CA(GA)2CT(GA)4CT	(GA)3CAGACT(GA)5GT(GA)12G((GA)2,4CA)(GA)2GGCT(GA)3GCCA(GA)3CT(GA)2AAGACA	(GA)2GCCAGAGCCA(GA)3GCA	98	1593
Aovo DRB*W8701	CA(GA)4CT	(GA)6CT(GA)5CT(GA)15CT	(GA)3GCGCGTG	66	734
Aona DRB*W2910	CA(GA)2CA(GA)2CT(GA)4CT	(GA)7AT((GA)4,5,4CT)(GA)3TA(GA)2CTGAAAGAC(GA)2CA(GA)3AACT(GA)3AACATA(GA)2CT(GA)5CACA	(GA)5CAGAGC(GA)4GCG	114	1833
Aona DRB*W2908	CA(GA)2CA(GA)2CT(GA)4CT	(GA)7AT((GA)4,5,4CT)(GA)3TA(GA)2CTGAAAGAC(GA)2CA(GA)3AACT(GA)3AACATA(GA)2CT(GA)5CACA	(GA)5CAGAGC(GA)4GCG	114	1833
Aovo DRB*W2901	CA(GA)2CA(GA)2CT(GA)4CT	(GA)7ATGAAA((GA)2,5,4CT)(GA)3TA(GA)2CTGAAAGAC(GA)2CA(GA)4CT(GA)3AACATA(GA)2CT(GA)5CACA	(GA)5CAGAGC(GA)4GCG	112	1854
Aovo DRB*W9001	CA(GA)2CA(GA)2CT(GA)4CT	(GA)4AT((GA)4,4CT)(GA)3TA(GA)2CT(GA)3CAGACACA	(GA)5CAGAGC(GA)4GCG	74	1182
Aovo DRB*W8801	CA(GA)4CA(GA)7CT(GA)7CT	(GA)7CT(GA)4AT((GA)4,5,4,6CT)(GA)6CTGAAAGAC(GA)2CA(GA)4CT(GA)3AACA(GA)3CT(GA)3GGGACACA	(GA)5CAGAGC(GA)4GCG	156	1975
Aovo DRB*W8501	CA(GA)2CA(GA)3CC(GA)5CT	(GA)6CTGAAAGAC(GA)2CA(GA)4CT(GA)3GGGACACA	(GA)5CAGAGC(GA)4GCG	68	1244
Aovo DRB*W8502	CA(GA)2CA(GA)3CC	(GA)3GCGACT((GA)4,6CT)(GA)6CTGAAAGAC(GA)2CA(GA)4CT(GA)3GGGACACA	(GA)5CAGAGC(GA)4GCG	84	1477
Aona DRB*W8501	CA(GA)2CA(GA)3CC	((GA)5,4CT)GC((GA)5,6CT)GAAAGAC(GA)2CA(GA)4CT(GA)3GGGACACACACA	(GA)5CAGAGC(GA)4GCG	90	1443
Aovo DRB*W8601	CA(GA)2CA(GA)2CA	(GA)10AA(GA)7AGACA(GA)14G((GA)2,4CA)((GA)3,3CT)(GA)2AAGACA(GA)3GCCA	(GA)2GCCAGAGCCA(GA)3GCA	114	1528
Aovo DRB*W8602	CA(GA)2CA(GA)2CA	(GA)9AA(GA)7AGACA(GA)15G((GA)2,4CA)(GA)3CT(GA)2AAGACA(GA)3GCCA	(GA)2GCCAGAGCCA(GA)2GCA	106	1513

Aovo DRB*W3001	CA(GA)2CA(GA)2CA	(GA)10AA(GA)7AGACA(GA)15G((GA)2,4CA)((GA)3,3CT)(GA)2AAGACA(GA)3GCCA	(GA)2GCCAGAGCCA(GA)3GCA	116	1527
Aona DRB*W3002	CA(GA)2CA(GA)2CA(GA)9AA	(GA)11AA(GA)7AGACA(GA)9GTG((GA)2,4CA)(GA)3GCCA(GA)3CT(GA)2AAGACA	(GA)2GCCAGAGCCA(GA)3GCA	118	1623
Caja-DRB03	**CA(GA)3CA(GA)2TACA**	**(GA)4CT(GA)3GG(GA)2CA(GA)2CT(GA)4CA(GA)9CA(GA)2TACA(GA)4CT(GA)4CA(GA)2GC(GA)2GCCA(GA)3AA**	**(GA)3GGAAA(GA)30GGGCG**	**158**	
Caja-DRB06	**CA(GA)3CA(GA)2TACA**	**(GA)4CT(GA)4CA(GA)2GC(GA)2GC(GA)2GC(GA)2GCCA**	**(GA)7GGAAA(GA)37GGGCG**	**130**	
Caja-DRB02	**CA(GA)11CA(GA)9CA**	**(GA)10TTGACA(GA)2CA(GA)3CT(GA)3CGAGGCAAAGACT(GA)2CAGACGAGAAG(GA)4CA(GA)2TT(GA)3AAGACT(GA)3CA(GA)4CT(GA)2AACT(GA)4CA(GA)2CT(GA)2CACT(GA)3CT(GA)3CC(GA)4C(GA)2CA(GA)4TA(GA)2AAGACA(GA)4CT(GA)3CA(GA)2CAGACT(GA)4CT(GA)2CT(GA)4CTTA(GA)2CT(GA)3CTGAAAGACT(GA)5CT(GA)3CACT(GA)4CT(GA)2GC(GA)2CA(GA)3CT(GA)5CCGAGG(GA)2CT(GA)2CAGAAACT(GA)3CT(GA)2AAGACT(GA)3CTAA(GA)2TACT(GA)5CT(GA)2CACTGATA(GA)2CT(GA)4CA(GA)2CT(GA)4TA(GA)4CTGAGCGACTAA(GA)2CACT(GA)3CT(GA)2CACTCATA(GA)2CT(GA)3CT(GA)2GGGACT(GA)3CTGACAGACACTGATA(GA)2CTCA(GA)4CT(GA)3CACT(GA)4CT**	**(GA)2GCCAGAGTGACACT(GA)16GCG**	**434**	
Caja-DRB05	**CA(GA)34CA(GA)13TTGACA**	**(GA)2CT(GA)3CT(GA)3CGAGGCAAAGACT(GA)2CAGAC(GA)2AG(GA)4CA(GA)2TT(GA)3AAGACT(GA)3CA(GA)4CT(GA)2AACT(GA)4CA(GA)2CT(GA)2CACT(GA)3CT(GA)3CTGAAAGTAAGACT(GA)5CT(GA)2GT(GA)2CA(GA)3CA(GA)3GCCT(GA)3CAGACA(GA)2CA(GA)4CT(GA)2CT(GA)5TT(GA)4CT(GA)3CACT(GA)7CT(GA)3CC(GA)4C(GA)2CA(GA)3GATAAAGACA(GA)4CT(GA)3CA(GA)2CAGACT(GA)4CT(GA)2CT(GA)4CTTA(GA)2CT(GA)3CTGAAAGACT(GA)3CACT(GA)4CT(GA)2GC(GA)2CA(GA)3CT(GA)5CCGAGG(GA)2CT(GA)2CAGAAACT(GA)3CT(GA)4CT(GA)3CTAA(GA)2TACT(GA)5CT(GA)2CACTGATA(GA)2CT(GA)4CA(GA)2CT(GA)4TA(GA)4CT(GA)3CTAA(GA)2CACT(GA)3CT(GA)2CACTCATA(GA)2CT(GA)2GGGACT(GA)3CTGACAGACACTGATA(GA)2CTCA(GA)4CT(GA)3CACT(GA)4CT**	**(GA)2GCCAGAGTGACACT(GA)15(GCGA)2GCGCAAGCG**	**554**	
Caja-DRB01	**(GA)16CA(GA)9CT(GACA)2CT**	**(GA)5CT(GA)3GTCTCA(GA)2CT(GA)4CA(GA)2CT(GA)3CTGT(GA)5CT(GA)4C(GA)2GG(GA)2(CAGA)2(GA)2AAGACT(GA)2CT(GA)5CT(GA)4CTGAAAGACT(GA)3CT(GA)3CAAGACA(GA)2CT(GA)4TAAGACAGACT(GA)4CTGATG(GA)4CT(GA)9GG**	**(GA)2AAGG(GA)3GGGCG**	**212**	

Figure S1. *Aotus* MHC-DRB Exon 2 + Intron 2 (partial)



```
                    40         *       160         *       180         *       200
                    AGTaCCGGGCGGTGACGGAGCTGGGGcGGCCctg  GC gAGtactggAACaGcCaGaAGGAc  c TGG
Aona-DRB1*0328GB    : AGTACCGGGCGGTGACGGAGCTGGGGCGGCGGAGCGCAGAGTACTGGAACAGCCAGAAGGACTTCCTGG :  207
Aona-DRB1*0329GA    : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aona-DRB1*031701GA  : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aovo-DRB1*0304GA    : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aovo-DRB1*0305GA    : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aovo-DRB1*0306GA    : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aovo-DRB1*0307GA    : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aona-DRB3*0615      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aona-DRB3*0627      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aona-DRB3*062501    : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aona-DRB3*0626      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aona-DRB3*062502    : AGTACCGGGCGGTGACGGAGCTGGGCCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aona-DRB3*0628      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aovo-DRB3*0601      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aona-DRB*W8901      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCGGACCGCAGAGTACTGGAACAGCCAGAAGGACTACGTGG :  207
Aona-DRB*W1808      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGATGCAAAGTACTGGAACGGTCAGCAGGACATCCTGG :  207
Aona-DRB*W1806      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGATGCAAAGTACTGGAACGGTCAGCAGGACATCCTGG :  207
Aovo-DRB*W1801      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGATGCAAAGTACTGGAACGGTCAGCAGGACATCCTGG :  207
Aovo-DRB*W1802      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGATGCAAAGTACTGGAACGGTCAGCAGGACATCCTGG :  207
Aovo-DRB*W1803      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGATGCAAAGTACTGGAACGGTCAGCAGGACATCCTGG :  207
Aovo-DRB*W8801      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCAGAAGGACATCCTGG :  207
Aovo-DRB*W2901      : AGTTCCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCTGAAGGAACGCCTGG :  207
Aona-DRB*W2908      : AGTTCCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCTGAAGGAACGCCTGG :  207
Aona-DRB*W2910      : AGTTCCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCCGAGTACTGGAACAGCCTGAAGGAACGCCTGG :  207
Aovo-DRB*W3001      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGTTGCAGAGAAGCTCAACAGCCAGAAGGACATCCTGG :  207
Aona-DRB*W3002      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTAGCGCAGAGAAGCTCAACAGCCAGAAGGACATCCTGG :  207
Aovo-DRB*W9201      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGTTGCAGAGAAGCTCAACAGCCAGAAGGACATCCTGG :  207
Aovo-DRB*W9202      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGTTGCAGAGAAGCTCAACAGCCAGAAGGACATCCTGG :  207
Aona-DRB*W9101      : AGTTCCGGGCGGTGACGGAGCTGGGGCGGCCTGAGGCCGAGTCCTGGAACAGCCAGAAGGACATCCTGG :  207
Aovo-DRB*W9101      : AGTTCCGGGCGGTGACGGAGCTGGGGCGGCCTGAAGCAGAGAAGTACAACAGCCAGAAGGACTTCCTGG :  207
Aovo-DRB*W9102      : AGTTCCGGGCGGTGACGGAGCTGGGGCGGCCTAGCGCAGAGAAGTACAACAGCCAGAAGGACATCCTGG :  207
Aovo-DRB*W9001      : AGTTCCGGGCGGTGACGGAGCTGGGGCGGCCTAGCGCAGAGAAGTTAAACAGCCAGAAGGAAGCCTGG :  207
Aovo-DRB*W4501      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCGGAGTACTGGAACAGCCAGAAGGACATCCTGG :  207
Aovo-DRB*W9301      : AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGCGGAGTCCTGGAACAGCCAGAAGGACTTAATGG :  207

                      *       220         *       240         *       260         *
                    AG   a GCGGG C  GGTgGACA CTaCTGcAgAcACAAcTACGGGGTTg TGAGAGCTTCACAGTGC
Aona-DRB1*0328GB    : AGGAGAGGCGGGCCTTGGTGGACACCTACTGTAGATACAACTACGGGGGTTGCTGAGAGCTTCACAGTGC :  276
Aona-DRB1*0329GA    : AGCGGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aona-DRB1*031701GA  : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aovo-DRB1*0304GA    : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aovo-DRB1*0305GA    : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
Aovo-DRB1*0306GA    : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aovo-DRB1*0307GA    : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
Aona-DRB3*0615      : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
Aona-DRB3*0627      : AGCAGAAGCGGGGCCAGGTGGACAACTACTGCAGACACAACTACGGGGTTTTTGAGAGCTTCACAGTGC :  276
Aona-DRB3*062501    : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
Aona-DRB3*0626      : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
Aona-DRB3*062502    : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
Aona-DRB3*0628      : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
Aovo-DRB3*0601      : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
Aona-DRB*W8901      : AGCAGAAGCGGGGCCGGGTGGACAACTACTGCAGACACAATTACGGGGTTGCTGAGAGCTTCACAGTGC :  276
Aona-DRB*W1808      : AGCTCAAGCGGGGCCAGGTAGACAACTACTGCAGACACAACTACGGGGTTTTTGAGAGCTTCACAGTGC :  276
Aona-DRB*W1806      : AGCTGAAGCGGGGCCAGGTAGACAACTACTGCAGACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W1801      : AGCTGAAGCGGGGCCAGGTAGACAACTACTGCAGACACAACTACGGGGTTTTTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W1802      : AGCTGAAGCGGGGCCAGGTAGACAACTACTGCAGACACAACTACGGGGTTTTTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W1803      : AGCTGAAGCGGGGCCAGGTAGACAACTACTGCAGACACAACTACGGGGTTTTTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W8801      : AGTATCTGCGGGCCGCGGTGGACAACTACTGCAGACACAACTACGGGGTTGCTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W2901      : AGTATCTGCGGGCCGCGGTGGACACCTGCTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
Aona-DRB*W2908      : AGTATCTGCGGGCCGCGGTGGACACCTACTGCAGACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aona-DRB*W2910      : AGTATCTGCGGGCCGCGGTGGACACCTACTGCAGACACAACTACGGGGTTTTTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W3001      : AGGACAGGCGGGCCTCGGTGGACACCTACTGTAAACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aona-DRB*W3002      : AGGACAGGCGGGCCGCGGTGGACACCTACTGTAAACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W9201      : AGGACAGGCGGGCCTCGGTGGACACCTACTGTAAACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W9202      : AGGACAGGCGGGCCTCGGTGGACACCTACTGTAAACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aona-DRB*W9101      : AGACCAGGCGGGCCGCGGTGGACACCTTCTGCAGACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W9101      : AGACCAGGCGGGCCGCGGTGGACACCTTCTGCAGACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W9102      : AGACCAGGCGGGCCGCGGTGGACACCTTCTGCAGACACAACTACGGGGTTTTTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W9001      : AGTATCTGCGGGCCGCGGTGGACACCTACTGCAGACACAACTACGGGGTTGGTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W4501      : AGGACAAGCGGGCCTCGGTGGACACCTACTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
Aovo-DRB*W9301      : AGGACAGGCGGGCCGCGGTGGACACCTACTGCAGACACAACTACGGGGTTGTTGAGAGCTTCACAGTGC :  276
```

```
                    280         *       300         *       320         *       340
                    AGCGGAgAGGTGAGcGCGGCGGGgCGGGGCCTCCCTGTGAgCTGCcgaTCAGAGA      gaga  ga
Aona-DRB1*0328GB    : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aona-DRB1*0329GA    : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAAAGAGA :  345
Aona-DRB1*031701GA  : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aovo-DRB1*0304GA    : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aovo-DRB1*0305GA    : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aovo-DRB1*0306GA    : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aovo-DRB1*0307GA    : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aona-DRB3*0615      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aona-DRB3*0627      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aona-DRB3*062501    : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aona-DRB3*0626      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aona-DRB3*062502    : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aona-DRB3*0628      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aovo-DRB3*0601      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACTGAGAGAGA :  345
Aona-DRB*W8901      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA----CAGAGACTGA :  341
Aona-DRB*W1808      : AGCGGAGAGGTGAGTGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA----GAGAGACTGA :  341
Aona-DRB*W1806      : AGCGGAGAGGTGAGTGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA----GAGAGACTGA :  341
Aovo-DRB*W1801      : AGCGGAAAGGTGAGTGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA----GAGAGA---- :  337
Aovo-DRB*W1802      : AGCGGAAAGGTGAGTGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA----GAGAGACTGA :  341
Aovo-DRB*W1803      : AGCGGAAAGGTGAGTGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA----GAGAGACTGA :  341
Aovo-DRB*W8801      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGAGAGACAGAGAGAGA :  345
Aovo-DRB*W2901      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA----CAGAGACTGA :  341
Aona-DRB*W2908      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA----CAGAGACTGA :  341
Aona-DRB*W2910      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA----CAGAGACTGA :  341
Aovo-DRB*W3001      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCGAATCAGAGA----CAGAGACAGA :  341
Aona-DRB*W3002      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCGGATCAGAGA----CAGAGACAGA :  341
Aovo-DRB*W9201      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCGAATCAGAGA----CAGAGACAGA :  341
Aovo-DRB*W9202      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCGAGTCAGAGA----CAGAGACAGA :  341
Aona-DRB*W9101      : AGCGGAGAGGTGAGCGCGGCGGGACGGGGCCTCCCTGTGAACTGCCAATCAGAGA-------------- :  331
Aovo-DRB*W9101      : AGCGGAGAGGTGAGCGCGGCGGGACGGGGCCTCCCTGTGAACTGCCAATCAGAGA-------------- :  331
Aovo-DRB*W9102      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA-------------- :  331
Aovo-DRB*W9001      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA-------------- :  331
Aovo-DRB*W4501      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCGAATCAGAGA----CAGAGACTGA :  341
Aovo-DRB*W9301      : AGCGGAGAGGTGAGCGCGGCGGGGCGGGGCCTCCCTGTGAGCTGCCGATCAGAGA-------------- :  331

                        *       360         *       380         *       400         *
                                                                    gAga agagaGA  gA
Aona-DRB1*0328GB    : CTGAGAGAGAC------CGAGAGAGAGACACAGAGACACTGAGAGA--GACAGAGAGAGAGAGACTGAG :  406
Aona-DRB1*0329GA    : CTGAGAGAGACACTGAGTGAGAGTGAGACAGAGAGACAGAGAAAGACTGACAGACAGAGAGAGACTGAG :  414
Aona-DRB1*031701GA  : CTGAGAGAGACACTGAGAGAGAGTGAGACAGAGAGACAGAGAAAGACTGACAGACAGAGAGAGACTGAG :  414
Aovo-DRB1*0304GA    : CCGAGAGAGA---------------GACACAGAGACACTGAGAGA--GACAGAGAGAGAGAGAGACTGAG :  396
Aovo-DRB1*0305GA    : CCGAGAGAGA---------------GACACAGAGACACTGAGAGA--GACAGAGAGAGAGAGAGACTGAG :  396
Aovo-DRB1*0306GA    : CCGAGAGAGA---------------GACACAGAGACACTGAGAGA--GACAGAGAGAGAGAGAGACTGAG :  396
Aovo-DRB1*0307GA    : CCGAGAGAGA-----------GAGAGACACAGAGACACTGAGAGA--GACAGAGAGAGAGAGACTGAG :  400
Aona-DRB3*0615      : CTGAGAGAGAC------CGAGAGAGAGACACAGAGACACTGAGAGA--GACAGAGAGAGAGAGACTGAG :  406
Aona-DRB3*0627      : CTGAGAGAGAC------CGAGAGAGAGACACAGAGACACTGAGAGA--GACAGAGAGAGAGAGACTGAG :  406
Aona-DRB3*062501    : CTGAGAGAGAC------CGAGAGAGAGACACAGAGACACTGAGAGA--GACAGAGAGAGAGAGACTGAG :  406
Aona-DRB3*0626      : CTGAGAGAGAC------CGAGAGAGAGACACAGAGACACTGAGAGA--GACAGAG--AGAGAGACTGAG :  404
Aona-DRB3*062502    : CCGAGAGAGAC------CGAGAGAGAGACACAGAGACACTGAGAGA--GACAGAGAGAGAGAGACTGAG :  406
Aona-DRB3*0628      : CTGAGAGAGAC------CGAGAGAGAGACACAGAGACACTGAGAGA--GACAGAGAGAGAGAGACTGAG :  406
Aovo-DRB3*0601      : CTGAGAGAGAC------CGAGAGAGAGACACAGAGACACTGAGAGA--GACAGAG-GAGAGAGACTGAG :  405
Aona-DRB*W8901      : GA-----------------------------------------------------GAGACTGAGAGAGAGAG :  360
Aona-DRB*W1808      : GAGAGACTGAG----------------------AGAGAGAGAGAGA------GAGACACAGAGAGAGAT :  382
Aona-DRB*W1806      : GAGAGACTGAG----------------------AGAGAGAGAGAGA------GAGACACAGAGAGAGAT :  382
Aovo-DRB*W1801      : -----------------------------------------------------GAGACACAGAGAGAGAT :  354
Aovo-DRB*W1802      : GAGAGACTGAG----------------------AGAGAG----AGA------GAGAGAGAGAGAGACAC :  378
Aovo-DRB*W1803      : GAGAGAGAGAG----------------------AGAGAGAGAGAGA------GAGACACAGAGAGAGAT :  382
Aovo-DRB*W8801      : GA-----------------------------------------------------GAGACTGAGAGAGAGAG :  364
Aovo-DRB*W2901      : GA-----------------------------------------------------GAGACTGAGAGAGAGAG :  360
Aona-DRB*W2908      : GA-----------------------------------------------------GAGACTGAGAGAGAGAG :  360
Aona-DRB*W2910      : GA-----------------------------------------------------GAGACTGAGAGAGAGAG :  360
Aovo-DRB*W3001      : GA-----------------------------------------------------GAGAGAGAGAGAGAGAA :  360
Aona-DRB*W3002      : GA-----------------------------------------------------GAGAGAGAGAGAGAAAG :  360
Aovo-DRB*W9201      : GA-----------------------------------------------------GAGAGAGAGAGAGAGAA :  360
Aovo-DRB*W9202      : GA-----------------------------------------------------GAGAGAGAGAGAGAAAG :  360
Aona-DRB*W9101      : -----------------------------------------------------CAGAGAGACCGAGAGAG :  348
Aovo-DRB*W9101      : -----------------------------------------------------CAGAGAGACCGAGAGAG :  348
Aovo-DRB*W9102      : -----------------------------------------------------CAGAGAGACCGAGAGAG :  348
Aovo-DRB*W9001      : -----------------------------------------------------CAGAGACTGAGAGAGAC :  348
Aovo-DRB*W4501      : GA-----------------------------------------------------GAGACTGAGAGACAGAC :  360
Aovo-DRB*W9301      : -----------------------------------------------------GAGACTGAGAGAGAGAG :  348
```

```
                      420         *       440         *       460         *       480
                    aga
Aona-DRB1*0328GB   : AGA-GACAGTGA--GAGA----------------------------CTGAGAGACTGAGACTGAGAG :  442
Aona-DRB1*0329GA   : AGG-TACACTGA--GAGAGACAGAGTGAGACAGAGAGACAGAGTAAGACTGACAGAGAGAGACTGAGAG :  480
Aona-DRB1*031701GA : AGG-TACACTGA--GAGAGACAGAGTGAGACAGAGAGACAGAGTAAGACTGACAGAGAGAGACTGAGAG :  480
Aovo-DRB1*0304GA   : AGA-GACAGTGA--GAGACT------------------------GAGAGACTGAGAGAGAC------ :  430
Aovo-DRB1*0305GA   : AGA-GACAGTGA--GAGA----------------------------CTGAGAGACTGAGACTGAGAG :  432
Aovo-DRB1*0306GA   : AGA-GACAGTGA--GAGACT------------------------GAGAGACTGAGAGAGACTGAGAG :  436
Aovo-DRB1*0307GA   : AGA-GACAGTGA--GAGACT------------------------GAGAGACTGAGAGAGACTGAGAG :  440
Aona-DRB3*0615     : AGA-GACAGTGA--GAGACT------------------------GAGAGACTGAGAGAGACTGAGAG :  446
Aona-DRB3*0627     : AGA-GACAGTGA--GAGACT------------------------GAGAGACTGAGAGAGACTGAGAG :  446
Aona-DRB3*062501   : AGA-GACAGTGA--GAGACT------------------------GAGAGACTGAGAGAGACTGAGAG :  446
Aona-DRB3*0626     : AGA-GACAGTGA--GAGACT------------------------GAGAGACTGAGAGAGACTGAGAG :  444
Aona-DRB3*062502   : AGA-GACAGTGA--GAGACT------------------------GAGAGACTGAGAGAGACTGAGAG :  446
Aona-DRB3*0628     : AGA-GACAGTGA--GAGACT------------------------GAGAGACTGAGAGAGACTGAGAG :  446
Aovo-DRB3*0601     : AGA-GACAGTGA--GAGACT------------------------GAGAGACTGAGAGAGACTGAGAG :  445
Aona-DRB*W8901     : AGA------------------------------------------------------------- :  363
Aona-DRB*W1808     : AGAC--AGAGACAGA--GAG-------------------------------------------- :  398
Aona-DRB*W1806     : AGAC--AGAGACAGA--GAG-------------------------------------------- :  398
Aovo-DRB*W1801     : AGAC--AGAGACAGA--GAG-------------------------------------------- :  370
Aovo-DRB*W1802     : AGAG--AGAGATAGACAGAG-------------------------------------------- :  396
Aovo-DRB*W1803     : AGAC--AGAGACAGA--GAG-------------------------------------------- :  398
Aovo-DRB*W8801     : AGAC--TGAGAGAGAGAGAG-------------------------------------------- :  382
Aovo-DRB*W2901     : AGA------------------------------------------------------------- :  363
Aona-DRB*W2908     : AGA------------------------------------------------------------- :  363
Aona-DRB*W2910     : AGA------------------------------------------------------------- :  363
Aovo-DRB*W3001     : AGA------------------------------------------------------------- :  363
Aona-DRB*W3002     : AGA------------------------------------------------------------- :  363
Aovo-DRB*W9201     : AGA------------------------------------------------------------- :  363
Aovo-DRB*W9202     : AGA------------------------------------------------------------- :  363
Aona-DRB*W9101     : AGA------------------------------------------------------------- :  351
Aovo-DRB*W9101     : A--------------------------------------------------------------- :  349
Aovo-DRB*W9102     : CGA------------------------------------------------------------- :  351
Aovo-DRB*W9001     : TGA------------------------------------------------------------- :  351
Aovo-DRB*W4501     : TGA------------------------------------------------------------- :  363
Aovo-DRB*W9301     : ACT------------------------------------------------------------- :  351

                            *       500         *       520         *       540         *
                                  gagaga a
Aona-DRB1*0328GB   : AGAC------ACTGAGAGAGACAGAGT----GAGACAGAGAGACA--GAGAAAGGCTGACAGACAGAGA :  499
Aona-DRB1*0329GA   : ATAC------ACTGAGAGAGACAGAGACTGAGAGACAGAGAGACACTGAGAGAGACAGAGAGACTGAAA :  543
Aona-DRB1*031701GA : ATAC------ACTGAGAGAGACAGAGACTGAGAGACAGCGAGACACTGAGAGAGACAGAGAGACTGAAA :  543
Aovo-DRB1*0304GA   : ----------ACTGAGAGAGACAGAGT----GAGACAGAGAAACA--GAGAAAGAC------------ :  470
Aovo-DRB1*0305GA   : AGAC------ACTGAGAGAGACAGAGG----GAGACAGAGAGACA--GAGAAAGACTGAGAGACAGAGA :  489
Aovo-DRB1*0306GA   : AGAC------ACTGAGAGAGACAGAGT----GAGACAGAGAAACA--GAGAAAGAC------------ :  480
Aovo-DRB1*0307GA   : AGAC------ACTGAGAGAGACAGAGT----GAGACAGAGAAACA--GAGAAAGAC------------ :  484
Aona-DRB3*0615     : AGAC------ACTGAGAGAGACAGAGT----GAGACAGAGAAACA--GAGAAAGAC------------ :  490
Aona-DRB3*0627     : AGAC------ACTGAGAGAGACAGAGG----GAGACAGAGAGACA--GAGAAAGACTGAGAGACAGAGA :  503
Aona-DRB3*062501   : AGAC------ACTGAGAGAGACAGAGT----GAGACAGAGAAACA--GAGAAAGAC------------ :  490
Aona-DRB3*0626     : AGAC------ACTGAGAGAGACAGAGT----GAGACAGAGAAACA--GAGAAAGAC------------ :  488
Aona-DRB3*062502   : AGAC------ACTGAGAGAGACAGAGT----GAGACAGAGAAACA--GAGAAAGAC------------ :  490
Aona-DRB3*0628     : AGAC------ACTGAGAGAGACAGAGG----GAGACAGAGAGACA--GAGAAAGACTGAGAGACAG--- :  500
Aovo-DRB3*0601     : AGAC------ACTGAGAGAGACAGAGT----GAGACAGAGAAACA--GAGAAAGAC------------ :  489
Aona-DRB*W8901     : -------------ATGAGAGAGACTGA------------------------------------------ :  377
Aona-DRB*W1808     : --ACTGAGAGACTGTGAGAGAGACTGA------------------------------------------ :  423
Aona-DRB*W1806     : --ACTGAGAGACTGTGAGAGAGACTG------------------------------------------- :  422
Aovo-DRB*W1801     : --ACTGAGAGACTGTGAGAGAGACTGA------------------------------------------ :  395
Aovo-DRB*W1802     : --ACTGAGAGACTGTGAGAGAGACTG------------------------------------------- :  420
Aovo-DRB*W1803     : --ACTGAGAGACTGTGAGAGAGACTG------------------------------------------- :  422
Aovo-DRB*W8801     : --ACTGAGAGAGAATGAGAGAGACTGA------------------------------------------ :  407
Aovo-DRB*W2901     : -------------ATGAAAGAGACTGA------------------------------------------ :  377
Aona-DRB*W2908     : -------------ATGAGAGAGACTGA------------------------------------------ :  377
Aona-DRB*W2910     : -------------ATGAGAGAGACTGA------------------------------------------ :  377
Aovo-DRB*W3001     : -------------GAGAGAGAGAG--------------------------------------------- :  374
Aona-DRB*W3002     : -------------GAGAGAGAGAG--------------------------------------------- :  374
Aovo-DRB*W9201     : -------------GAGAGAGAGAG--------------------------------------------- :  374
Aovo-DRB*W9202     : -------------GAGAGAGAGAA--------------------------------------------- :  374
Aona-DRB*W9101     : -------------CTGAGAGAGACTGC------------------------------------------ :  365
Aovo-DRB*W9101     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9102     : -------------CTGAGAGAGACTGA------------------------------------------ :  365
Aovo-DRB*W9001     : -------------GAGA----------------------------------------------------- :  355
Aovo-DRB*W4501     : -------------GAGAGAGAGTG--------------------------------------------- :  374
Aovo-DRB*W9301     : -------------GAGAG---------------------------------------------------- :  356
```

```
                          560         *       580         *       600         *       620

Aona-DRB1*0328GB   : GAGACTGAGAGATACACTGAGAG----AGACAGAGACTGAGAGAC----AGACACT----------GAG :  550
Aona-DRB1*0329GA   : GA--CTGAGAGAGACAGAGAGAGACTGAGAGAGAGACTGAGAGACTGAGAGATACTTAGGGAGACTGAG :  610
Aona-DRB1*031701GA : GA--CTGAGAGAGACAGAGAGAGACTGAGAGAGAGACTGAGAGACTGAGAGATACTTAGGGAGACTGAG :  610
Aovo-DRB1*0304GA   : -----TGAGAGATACACTGAGAG----AGACAGAGACTGAGAGACAGAGAGACACT----------GAG :  520
Aovo-DRB1*0305GA   : GA--CTGAGAGATACACTGAGAG----AGACAGAGACTGAGAGACAGGGAGACACT----------GAG :  542
Aovo-DRB1*0306GA   : -----TGAGAAATACACTGAGAG----AGACAGAGACTGAGAGACAGAGAGACACT----------GAG :  530
Aovo-DRB1*0307GA   : -----TGAGAAATACACTGAGAG----AGACAGAGACTGAGAGACAGAGAGACACT----------GAG :  534
Aona-DRB3*0615     : -----TGAGAGATACACTGAGAG----AGACAGAGACTGAGAGACAGAGAGACAC------------AG :  538
Aona-DRB3*0627     : GAGACTGAGAGATACACTGAGAG----AGACAGAGACTGAGAGACAGAGAGACACT----------GAG :  558
Aona-DRB3*062501   : -----TGAGAGATACACTGAGAG----AGACAGAGACTGAGAGACAGAGAGACAC------------AG :  538
Aona-DRB3*0626     : -----TGAGAGATACACTGAGAG----AGACAGAGACTGAGAGACAGAGAGACAC------------AG :  536
Aona-DRB3*062502   : -----TGAGAGATACACTGAGAG----AGACAGAGACTGAGAGACAGAGAGACAC------------AG :  538
Aona-DRB3*0628     : -AGACTGAGAGATACACTGAGAG----AGACAGAGACTGAGAGACAGAGAGACACT----------GAG :  554
Aovo-DRB3*0601     : -----TGAGAGATACACTGAGAG----AGACAGAGACTGAGAGACAGAGAGACAC------------AG :  537
Aona-DRB*W8901     : ---------------GA----------------GAGAGACTGAGAGAGACTGAGA-------------- :  401
Aona-DRB*W1808     : ---------------GAGAGACT------GTGCAAGACCGAGAGAGAGACTGAGA-------------- :  457
Aona-DRB*W1806     : ----------------------------TGCAAGACCGAGAGAGAGACTGAGA-------------- :  447
Aovo-DRB*W1801     : ---------------GAGAGACT------GTGCAAGACCGAGAGAGAGATTGAGA-------------- :  429
Aovo-DRB*W1802     : ----------------------------TGCAAGACCGAGAGAGAGACTGAGA-------------- :  445
Aovo-DRB*W1803     : ------------------------------------------AGAGAGACTGTG-------------- :  434
Aovo-DRB*W8801     : ---------------GA----------------GAGAGACTGAGAGAGACTGAGA-------------- :  431
Aovo-DRB*W2901     : ---------------GA----------------GAGAGACTGAGAGAGACTGAGA-------------- :  401
Aona-DRB*W2908     : ---------------GA----------------GAGAGACTGAGAGAGACTGAGA-------------- :  401
Aona-DRB*W2910     : ---------------GA----------------GAGAGACTGAGAGAGACTGAGA-------------- :  401
Aovo-DRB*W3001     : ---------------------------------AAGACAGAGAGAGAGAGAGAGA-------------- :  396
Aona-DRB*W3002     : ----------------A----------------GAGAGAAAGAGAGAGAGAGAGAA-------------- :  398
Aovo-DRB*W9201     : ---------------------------------AAGACAGAGAGAGAGAGAGAGA-------------- :  396
Aovo-DRB*W9202     : ----------------------------------GACAGAGAGAGAGAGAGAGAGA-------------- :  396
Aona-DRB*W9101     : ----------------------------------------GAGAGAGA---------------------- :  373
Aovo-DRB*W9101     : ------------------------------------------------------------------------ :    -
Aovo-DRB*W9102     : ----------------------------------------GAGAGAGA---------------------- :  373
Aovo-DRB*W9001     : ------------------------------------------------------------------------ :    -
Aovo-DRB*W4501     : ---------------------------------AGAGAGAGAGAGAGAGA-------------- :  391
Aovo-DRB*W9301     : ------------------------------------------------------------------------ :    -

                            *       640         *       660         *       680         *
                                                          agaga   a
Aona-DRB1*0328GB   : AGAGACTGAGAGACTG--------TGAGAGAGAGACTGAGAGAGACAGA----GAGAGACTGAGAGAGA :  607
Aona-DRB1*0329GA   : AGAGACAGAGAGACTGAGACAGACTGAGAGAAAGACTGAGAGAGAGAGACTGAGAGAGGCTGAGAGAGA :  679
Aona-DRB1*031701GA : AGAGACAGAGAGACTGAGACAGACTGAGAGAAAGACTGAGAGAGAGAGACTGAGAGAGGCTGAGAGAGA :  679
Aovo-DRB1*0304GA   : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA------GAGACTGAGAGAAA :  569
Aovo-DRB1*0305GA   : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA----GAGAGACTGAGAGAGA :  593
Aovo-DRB1*0306GA   : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA------GAGACTGAGAGAAA :  579
Aovo-DRB1*0307GA   : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA------GAGACTGAGAGAAA :  583
Aona-DRB3*0615     : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA------GAGACTGAGAGAGA :  587
Aona-DRB3*0627     : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA----GAGAGACTGAGAGAGA :  609
Aona-DRB3*062501   : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA------GAGACTGAGAGAGA :  587
Aona-DRB3*0626     : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA------GAGACTGAGAGAGA :  585
Aona-DRB3*062502   : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA------GAGACTGAGAGAGA :  587
Aona-DRB3*0628     : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA----GAGAGACTGAGAGAGA :  605
Aovo-DRB3*0601     : AGAGACAGAGAGACTGA--------------AAGACTGAGAGAGACAGA------GAGACTGAGAGAGA :  586
Aona-DRB*W8901     : ----------------------------------GATAGAGACTGA---------------------- :  413
Aona-DRB*W1808     : ----------------------------------CTGAGAGATTGAGA------------GAGACAGA :  479
Aona-DRB*W1806     : ----------------------------------CTGAGAGATTGAGA------------GAGACAGA :  469
Aovo-DRB*W1801     : ----------------------------------CTGAGAGATTGAGA------------GAGACAGA :  451
Aovo-DRB*W1802     : ----------------------------------CTGAGAGATTGAGA------------GAGACAGA :  467
Aovo-DRB*W1803     : ----------------------------------CAAGACCGAGAGA------------GAGACTGA :  455
Aovo-DRB*W8801     : ----------------------------------GAGAGAGACTGAGA------------GAGAGAGA :  453
Aovo-DRB*W2901     : ----------------------------------GATAGAGACTGA---------------------- :  413
Aona-DRB*W2908     : ----------------------------------GATAGAGACTGA---------------------- :  413
Aona-DRB*W2910     : ----------------------------------GATAGAGACTGA---------------------- :  413
Aovo-DRB*W3001     : ----------------------------------GAGAGAGA--GA---------------------- :  406
Aona-DRB*W3002     : ----------------------------------GACAGAGAGAGA---------------------- :  410
Aovo-DRB*W9201     : ----------------------------------GAGAGAGA--GA---------------------- :  406
Aovo-DRB*W9202     : ----------------------------------GAGAGAGA--GA---------------------- :  406
Aona-DRB*W9101     : ------------------------------------------------------------------------ :    -
Aovo-DRB*W9101     : ------------------------------------------------------------------------ :    -
Aovo-DRB*W9102     : ------------------------------------------------------------------------ :    -
Aovo-DRB*W9001     : ------------------------------------------------------------------------ :    -
Aovo-DRB*W4501     : ----------------------------------GAGAGAGG---------------------------- :  399
Aovo-DRB*W9301     : ------------------------------------------------------------------------ :    -
```

```
                          700         *       720         *       740         *       76

Aona-DRB1*0328GB   : CACTGAGAGAGACAGAGA--GACTGAAAGACTGAGAGATACTTAGGGAGACTGAGAGAAAC-------A :  667
Aona-DRB1*0329GA   : GACTGAGGGAGACAGAGAATGAGAGAGAGACTGAGAGACAAG-ACAGAGACTGAGAGAGATGAGACAGA :  747
Aona-DRB1*031701GA : GACTGAGGGAGACCGAGAATGAGAGAGAGACTGAGAGACAAG-ACAGAGACTGAGAGAGATGAGACAGA :  747
Aovo-DRB1*0304GA   : ------------------------------------------------------------------- :    -
Aovo-DRB1*0305GA   : GACT--------------------------------------------------------------- :  597
Aovo-DRB1*0306GA   : ------------------------------------------------------------------- :    -
Aovo-DRB1*0307GA   : ------------------------------------------------------------------- :    -
Aona-DRB3*0615     : ------------------------------------------------------------------- :    -
Aona-DRB3*0627     : GGCT----------------------------------------------------------------- :  613
Aona-DRB3*062501   : ------------------------------------------------------------------- :    -
Aona-DRB3*0626     : ------------------------------------------------------------------- :    -
Aona-DRB3*062502   : ------------------------------------------------------------------- :    -
Aona-DRB3*0628     : GACT----------------------------------------------------------------- :  609
Aovo-DRB3*0601     : ------------------------------------------------------------------- :    -
Aona-DRB*W8901     : ------------------------------------------------------------------- :    -
Aona-DRB*W1808     : ------------------------------------------------------------------- :    -
Aona-DRB*W1806     : ------------------------------------------------------------------- :    -
Aovo-DRB*W1801     : ------------------------------------------------------------------- :    -
Aovo-DRB*W1802     : ------------------------------------------------------------------- :    -
Aovo-DRB*W1803     : ------------------------------------------------------------------- :    -
Aovo-DRB*W8801     : ------------------------------------------------------------------- :    -
Aovo-DRB*W2901     : ------------------------------------------------------------------- :    -
Aona-DRB*W2908     : ------------------------------------------------------------------- :    -
Aona-DRB*W2910     : ------------------------------------------------------------------- :    -
Aovo-DRB*W3001     : ------------------------------------------------------------------- :    -
Aona-DRB*W3002     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9201     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9202     : ------------------------------------------------------------------- :    -
Aona-DRB*W9101     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9101     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9102     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9001     : ------------------------------------------------------------------- :    -
Aovo-DRB*W4501     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9301     : ------------------------------------------------------------------- :    -

                    0         *       780         *       800         *       820

Aona-DRB1*0328GB   : GAGAGACTGAGAGAGACTGAGAGAAAGACTGAGAGAGAGACTGAGAGAGGCTGAGAGACTGAGGGAGA- :  735
Aona-DRB1*0329GA   : GAGAGACTGAGAGAGACTAAGAGA--GACTGAGGGAGA--CTGAGAGAGACTGAAAGACTGAGAGAGAG :  812
Aona-DRB1*031701GA : GAGAGACTGAGAGAGACTAAGAGA--GACTGAGGGAGA--CTGAGAGAGACTGAAAGACTGAGAGAGAG :  812
Aovo-DRB1*0304GA   : ----------------------GA---------------CTGAGAGAGACTGAGAGACTGAGAGAAA- :  599
Aovo-DRB1*0305GA   : ----------GAGAGACTGAGAGA--TACTTAGGGAGA--CTGAGAGAGACAGAGAGACTGAGAGAGA- :  651
Aovo-DRB1*0306GA   : ----------------------GA---------------CTGAGAGAGACTGAGAGACTGAGAGAAA- :  609
Aovo-DRB1*0307GA   : ----------------------GA---------------CTGAGAGAGACTGAGAGACTGAGAGAAA- :  613
Aona-DRB3*0615     : ----------------CTGAGAGA----------AAGA--CTGAGAGAGACTGAGAGACTGAGAGAAA- :  627
Aona-DRB3*0627     : ----------GAGAGACTGAGAGA--TACTTAGGGAGA--CTGAGAGAGACAGAGAGACTGAGAGAAAG :  668
Aona-DRB3*062501   : ----------------CTGAGAGA----------AAGA--CTGAGAGAGACTGAGAGACTGAGAGAAA- :  627
Aona-DRB3*0626     : ----------------CTGAGAGA----------AAGA--CTGAGAGAGACTGAGAGACTGAGAGAAA- :  625
Aona-DRB3*062502   : ----------------CTGAAAGA----------------CTGAGAGAGACTGAGAGACTGAGAGAAA- :  623
Aona-DRB3*0628     : ----------GAGAGACTGAGAGA--TACTTAGGGAGA--CTGAGAGAGACAGAGAGACTGAGAGAAAG :  664
Aovo-DRB3*0601     : ----------------CTGAGAGA----------AAGA--CTGAGAGAGACTGAGAGACTGAGAGAAA- :  626
Aona-DRB*W8901     : ------------------------------------------------------------AAGAC- :  418
Aona-DRB*W1808     : ---------------------------------------------------------GACTGAGA-- :  487
Aona-DRB*W1806     : ---------------------------------------------------------GACTGAGA-- :  477
Aovo-DRB*W1801     : ---------------------------------------------------------GACTGAGA-- :  459
Aovo-DRB*W1802     : ---------------------------------------------------------GACTGAGA-- :  475
Aovo-DRB*W1803     : ---------------------------------------------------------GACTGAGA-- :  463
Aovo-DRB*W8801     : ------------------------------------------------------CTGAAAGAC- :  462
Aovo-DRB*W2901     : ------------------------------------------------------------AAGAC- :  418
Aona-DRB*W2908     : ------------------------------------------------------------AAGAC- :  418
Aona-DRB*W2910     : ------------------------------------------------------------AAGAC- :  418
Aovo-DRB*W3001     : ------------------------------------------------------------GAGAG- :  411
Aona-DRB*W3002     : ------------------------------------------------------------GAG--- :  413
Aovo-DRB*W9201     : ------------------------------------------------------------GAG--- :  409
Aovo-DRB*W9202     : ------------------------------------------------------------GAG--- :  409
Aona-DRB*W9101     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9101     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9102     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9001     : ------------------------------------------------------------------- :    -
Aovo-DRB*W4501     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9301     : ------------------------------------------------------------------- :    -
```

```
                      *       840         *       860         *       880         *

Aona-DRB1*0328GB   : ----------------CAGAGAATGAGAGAGA---GACTGAGAGA---TAAGACAGAGACTGAGAGAGA :  782
Aona-DRB1*0329GA   : ACTGAGAGAGACAAGACAGAGACTGAGAGAGATGAGACAGAGAGAGACTGAGAGAAAGACTGAGAGACA :  881
Aona-DRB1*031701GA : ACTGAGAGAGACAAGACAGAGACTGAGAGAGATGAGACAGAGAGAGACTGAGAGAAAGACTGAGAGACA :  881
Aovo-DRB1*0304GA   : ------------------------------------------------------------------- :    -
Aovo-DRB1*0305GA   : ------------------------------------------------CTGAGAGAAAGACTGAGAGA-- :  671
Aovo-DRB1*0306GA   : ------------------------------------------------------------------- :    -
Aovo-DRB1*0307GA   : ------------------------------------------------------------------- :    -
Aona-DRB3*0615     : ------------------------------------------------------------------- :    -
Aona-DRB3*0627     : ACT---------------------------------------------GAGAGAGACTGAGAGACT :  689
Aona-DRB3*062501   : ------------------------------------------------------------------- :    -
Aona-DRB3*0626     : ------------------------------------------------------------------- :    -
Aona-DRB3*062502   : ------------------------------------------------------------------- :    -
Aona-DRB3*0628     : ACT---------------------------------------------GAGAGAGACTGAGAGACT :  685
Aovo-DRB3*0601     : ------------------------------------------------------------------- :    -
Aona-DRB*W8901     : ------------------------------------------------------------------- :    -
Aona-DRB*W1808     : ------------------------------------------------------------------- :    -
Aona-DRB*W1806     : ------------------------------------------------------------------- :    -
Aovo-DRB*W1801     : ------------------------------------------------------------------- :    -
Aovo-DRB*W1802     : ------------------------------------------------------------------- :    -
Aovo-DRB*W1803     : ------------------------------------------------------------------- :    -
Aovo-DRB*W8801     : ------------------------------------------------------------------- :    -
Aovo-DRB*W2901     : ------------------------------------------------------------------- :    -
Aona-DRB*W2908     : ------------------------------------------------------------------- :    -
Aona-DRB*W2910     : ------------------------------------------------------------------- :    -
Aovo-DRB*W3001     : ------------------------------------------------------------------- :    -
Aona-DRB*W3002     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9201     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9202     : ------------------------------------------------------------------- :    -
Aona-DRB*W9101     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9101     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9102     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9001     : ------------------------------------------------------------------- :    -
Aovo-DRB*W4501     : ------------------------------------------------------------------- :    -
Aovo-DRB*W9301     : ------------------------------------------------------------------- :    -

                      900         *       920         *       940         *       960
                                                        ga ag               g        agagac
Aona-DRB1*0328GB   : T-GAGACAGAGAGA----GACTGA--GAGAAAGACTGAGAGACAGAGA----CTGAGAGACTGAGAGA- :  839
Aona-DRB1*0329GA   : --GAGACTGAGAGACTGAGACTGAGAGAAATAGACTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  948
Aona-DRB1*031701GA : --GAGACTGAGAGACTGAGACTGAGAGAAATAGACTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  948
Aovo-DRB1*0304GA   : ------CAGAGA---------------------CTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  640
Aovo-DRB1*0305GA   : ----GACTGAGAGA-----------AACAGGGACTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  724
Aovo-DRB1*0306GA   : ------CAGAGA---------------------CTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  650
Aovo-DRB1*0307GA   : ------CAGAGA---------------------CTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  654
Aona-DRB3*0615     : ------CAGAGA---------------------CTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  668
Aona-DRB3*0627     : GAGAGACTGAGAGA-----------AACAGAGACTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  746
Aona-DRB3*062501   : ------CAGAGA---------------------CTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  668
Aona-DRB3*0626     : ------CAGAGA---------------------CTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  666
Aona-DRB3*062502   : ------CAGAGA---------------------CTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  664
Aona-DRB3*0628     : GAGAGACTGAGAGA-----------AACAGAGACTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  742
Aovo-DRB3*0601     : ------CAGAGA---------------------CTGAGAGACTGAGAGACACTGAGAGACTGAGAGAC :  667
Aona-DRB*W8901     : ----------------------------------GAGACAG-----------AG--------AGAAAC :  433
Aona-DRB*W1808     : ----------------------------------GAGGCTG-----------AGAGACTCGGAGAGAC :  510
Aona-DRB*W1806     : ----------------------------------GAGACTG-----------AGAGACTCGGAGAGAC :  500
Aovo-DRB*W1801     : ----------------------------------GAGACTG-----------AGAGACTCGGAGAGAC :  482
Aovo-DRB*W1802     : ----------------------------------GAGACTG-----------AGAGACTCGGAGAGAC :  498
Aovo-DRB*W1803     : ----------------------------------GAGACTG-----------AGAGACTCGGAGAGAC :  486
Aovo-DRB*W8801     : ----------------------------------GAGACAG-----------AG--------AGAGAC :  477
Aovo-DRB*W2901     : ----------------------------------GAGACAG-----------AG--------AGAGAC :  433
Aona-DRB*W2908     : ----------------------------------GAGACAG-----------AG--------AGAAAC :  433
Aona-DRB*W2910     : ----------------------------------GAGACAG-----------AG--------AGAAAC :  433
Aovo-DRB*W3001     : ----------------------------------GAGACAG-----------AG--------AGAGAC :  426
Aona-DRB*W3002     : ----------------------------------AGAGAG-----------AGT-------GGAGAC :  428
Aovo-DRB*W9201     : ----------------------------------GAGACAG-----------AG--------AGAGAC :  424
Aovo-DRB*W9202     : ----------------------------------GAGACAG-----------AG--------AGAGAC :  424
Aona-DRB*W9101     : ----------------------------------GACTGAG-----------AGAGA----GAGACTG :  392
Aovo-DRB*W9101     : ----------------------------------GACTGAG-----------AGAGA----GAGACTG :  368
Aovo-DRB*W9102     : ----------------------------------GACTGAG-----------AGAGA----GAGACTG :  392
Aovo-DRB*W9001     : ----------------------------------GAATGAG-----------AGAGA----CTGAGAG :  374
Aovo-DRB*W4501     : ----------------------------------AGACAG-----------AG--------AGAGAC :  413
Aovo-DRB*W9301     : ------------------------------------------------------------------- :    -
```

```
                          *       980         *      1000         *      1020         *
                            ag  aga a                                ActGA agAga   agA A
Aona-DRB1*0328GB    : ---GACAGAGACTGAGAGACTGAGAGA----CTGAGAGACACAGATAGAGATTGAGAGACTGAGAGAGA :  901
Aona-DRB1*0329GA    : ACAGATAGAGATTGAGAGACTGAGAGAGAGACTGAGAGACACTGATAGAGACTGAGAGAGAGACTGAGA : 1017
Aona-DRB1*031701GA  : ACAGATAGAGATTGAGAGACTGAGAGAGAGACTGAGAGACACTGATAGAGACTGAGAGAGAGACTGAGA : 1017
Aovo-DRB1*0304GA    : ACAGATAGAGATTGAGAGACTGAGAGA------------CACTGATAGAGACTGAGAGA--GACTGAGA :  695
Aovo-DRB1*0305GA    : ACAGATAGAAATTGAGAGACTGAGAGA------------CACTGATAGAGACTGAGAGAGAGACTGAGA :  781
Aovo-DRB1*0306GA    : ACAGATAGAGATTGAGAGACTGAGAGA------------CACTGATAGAGACTGAGAGA--GACTGAGA :  705
Aovo-DRB1*0307GA    : ACAGATAGAGATTGAGAGACTGAGAGA------------CACTGATAGAGACTGAGAGA--GACTGAGA :  709
Aona-DRB3*0615      : ACAGATAGAGATTGAGAGACTGAGAGAGAGACTGAGAGACACTGATAGAGACTGAGAGAGAGACTGAGA :  737
Aona-DRB3*0627      : ACAGATAGAGATTGAGAGACTGAGAGA------------CACTGATAGAGACTGAGAGAGAGACTGAGA :  803
Aona-DRB3*062501    : ACAGATAGAGATTGAGAGACTGAGAGAGAGACTGAGAGACACTGATAGAGACTGAGAGAGAGACTGAGA :  737
Aona-DRB3*0626      : ACAGATAGAGATTGAGAGACTGAGAGAGAGACTGAGAGACACTGATAGAGACTGAGAGAGAGACTGAGA :  735
Aona-DRB3*062502    : ACAGATAGAGATTGAGAGACTGAGAGAGAGACTGAGAGACACTGATAGAGACTGAGAGAGAGA------ :  727
Aona-DRB3*0628      : ACAGATAGAGATTGAGAGACTGAGAGA------------CACTGATAGAGACTGAGAGAGAGACTGAGA :  799
Aovo-DRB3*0601      : ACAGATAGAGATTGAGAGACTGAGAGAGAGACTGAGAGACACTGATAGAGACTGAGAGAGAGACTGAGA :  736
Aona-DRB*W8901      : TGAG--AGAAACAT--------------------AGA--GACTGAGAGAGAG--ACACA---------- :  466
Aona-DRB*W1808      : TGAGCGAGAGACTG--------------------AGATAGACTGAGAGAAACTGAGAGA---------- :  549
Aona-DRB*W1806      : TGAGCGAGAGACTG--------------------AGATAGACTGAGAGAAACTGAGAGA---------- :  539
Aovo-DRB*W1801      : TGAGCGAGAGACTG--------------------AGATAGACTGAGAGAAACTGAGAGA---------- :  521
Aovo-DRB*W1802      : TGAGCGAGAGACTG--------------------AGATAGACTGAGAGAAAC--AGAGA---------- :  535
Aovo-DRB*W1803      : TGAGCGAGAGACTG--------------------AGATAGACTGAGAGAAACTGAGAGA---------- :  525
Aovo-DRB*W8801      : TGAG--AGAAACAG--------------------AGA--GACTGAGAGAGGG--ACACA---------- :  510
Aovo-DRB*W2901      : TGAG--AGAAACAT--------------------AGA--GACTGAGAGAGAG--ACACA---------- :  466
Aona-DRB*W2908      : TGAG--AGAAACAT--------------------AGA--GACTGAGAGAGAG--ACACA---------- :  466
Aona-DRB*W2910      : TGAG--AGAAACAT--------------------AGA--GACTGAGAGAGAG--ACACA---------- :  466
Aovo-DRB*W3001      : --AG--AGAGACTG--------------------AGA--GACTGAGAAAGAC--AGAGA---------- :  457
Aona-DRB*W3002      : AGAG--AGAGACAG--------------------AGA--GACTGAGAAAGAC--AGAGA---------- :  461
Aovo-DRB*W9201      : --AG--AGAGACTG--------------------AGA--GACTGAGAAAGAC--AGAGA---------- :  455
Aovo-DRB*W9202      : --AG--AGAGACTG--------------------AGA--GACTGAGAAAGAC--AGAGA---------- :  455
Aona-DRB*W9101      : AAAGACGAGACA------------------------GAGAGAGACTGAGAGAGGGACA---------- :  426
Aovo-DRB*W9101      : AAAGACGAGACA------------------------GAGAGAGACTGAGAGAGGGACA---------- :  402
Aovo-DRB*W9102      : AAAGACGAGACA------------------------GAGAGAGACTGAGAGAGGGACA---------- :  426
Aovo-DRB*W9001      : AGACTGAGAGAT------------------------AGAGACTGAGAGACAGACA---------- :  405
Aovo-DRB*W4501      : --AG--AGAGGCTG--------------------AGA--GACTGAGAAAGAC--AGAGA---------- :  444
Aovo-DRB*W9301      : ----------------------------------AGAGACTGAGAGAGAGAGAGA---------- :  379

                          1040         *      1060         *      1080         *      1100
                              ag  agaGAGa AgaG ga                                GaGagaGc
Aona-DRB1*0328GB    : GACTGGGGGGG-GAGAGAGAGAGAGAGAGAGAGA--------------------CTGAGAGAGCACGTG :  949
Aona-DRB1*0329GA    : GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA----CTGAGAGAGCGCGTG : 1082
Aona-DRB1*031701GA  : GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACTGAGAGAGCGCGTG : 1086
Aovo-DRB1*0304GA    : GACTGAGAGAGAGAGAGAGAGAGAAA-------------------------CTGAGAGAGCACTTG :  738
Aovo-DRB1*0305GA    : GACTGAGAGACAGAGAGAGAGAGAGAGAAA---------------------CTGAGAGAGCACTTG :  828
Aovo-DRB1*0306GA    : GACTGAGAGAGAGAGAGAGAGAGAGAGAAA---------------------CTGAGAGAGCACTTG :  752
Aovo-DRB1*0307GA    : GACTGAGAGAGAGAGAGAGAGAGAGAGAAA---------------------CTGAGAGAGCACTTG :  756
Aona-DRB3*0615      : GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA--------------------CTGAGAGAGCGCTTG :  786
Aona-DRB3*0627      : GACTGAGAGAGAGAGAGAGAGAGAA--------------------------CTGAGAGAGCACTTG :  844
Aona-DRB3*062501    : GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA--------------------CTGAGAGAGCGCTTG :  786
Aona-DRB3*0626      : GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA--------------------CTGAGAGAGCGCTTG :  784
Aona-DRB3*062502    : ------GAGAGAGAGAGAGAGAGAGAGA-----------------------CTGAGAGAGCGCTTG :  764
Aona-DRB3*0628      : GACTGAGAGAGAGAGAGAGAGAGAAA-------------------------CTGAGAGAGCACTTG :  842
Aovo-DRB3*0601      : GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA--------------------CTGAGAGAGCGCTTG :  785
Aona-DRB*W8901      : ----GAG--AGAGAGACAGAGCGA-------------------------------GAGAGAGCG---- :  493
Aona-DRB*W1808      : ----TACTGAGAGAGAGAGAGCGC-------------------------------GCGACAGCG---- :  578
Aona-DRB*W1806      : ----TACTGAGAGAGAGAGAGCGC-------------------------------GCGACAGCG---- :  568
Aovo-DRB*W1801      : ----TACTGAGAGAGAGAGAGCGC-------------------------------GCGACAGCG---- :  550
Aovo-DRB*W1802      : ----TACTGAGAGAGAGAGAGCGC-------------------------------GCGACAGCG---- :  564
Aovo-DRB*W1803      : ----TACTGAGAGAGAGAGAGCGC-------------------------------GCGACAGCG---- :  554
Aovo-DRB*W8801      : ----GAG--AGAGAGACAGAGCGA-------------------------------GAGAGAGCG---- :  537
Aovo-DRB*W2901      : ----GAG--AGAGAGACAGAGCGA-------------------------------GAGAGAGCG---- :  493
Aona-DRB*W2908      : ----GAG--AGAGAGACAGAGCGA-------------------------------GAGAGAGCG---- :  493
Aona-DRB*W2910      : ----GAG--AGAGAGACAGAGCGA-------------------------------GAGAGAGCG---- :  493
Aovo-DRB*W3001      : ----GAGCCAGAGAGCCAGAGCCA-------------------------------GAGAGAGCA---- :  486
Aona-DRB*W3002      : ----GAGCCAGAGAGCCAGAGCCA-------------------------------GAGAGAGCA---- :  490
Aovo-DRB*W9201      : ----GAGCCAGAGAGCCAGAGCCA-------------------------------GAGAGAGCA---- :  484
Aovo-DRB*W9202      : ----GAGCCAGAGAGCCAGAGCCA-------------------------------GAGAGAGCA---- :  484
Aona-DRB*W9101      : ----CAGAGAGAGAGACAGAGCGA-------------------------------GAGAGAGCG---- :  455
Aovo-DRB*W9101      : ----CAGAGAGAGAGACAGAGCGA-------------------------------GAGAGAGCG---- :  431
Aovo-DRB*W9102      : ----CAGAGAGAGAGACAGAGCGA-------------------------------GAGAGAGCG---- :  455
Aovo-DRB*W9001      : ----CAGAGAGAGAGACAGAGCGA-------------------------------GAGAGAGCG---- :  434
Aovo-DRB*W4501      : ----GAGCCAGAGAGCCAGAGCCA-------------------------------GAGAGAGCA---- :  473
Aovo-DRB*W9301      : ----GAGAGAGAGAGACTGAGA---------------------------------GAGCGCGTG---- :  408
```

```
                          *      1120         *      1140         *      1160         *
                    C aTCTGTGAG  TT AGaATCCTcTc ATCCTGAGCagGGAGcTCtGaGGGCACAggTGTgTGTgt
Aona-DRB1*0328GB    : CCATCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- : 1016
Aona-DRB1*0329GA    : CCATCTGTGAGCATTCAGAATCCTGTCCATCCTGAGCAGGGAGCTCTGGGGGCACAGGTGTGTGTAT-- : 1149
Aona-DRB1*031701GA  : CCATCTGTGAGCATTCAGAATCCTGTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTTTGTAT-- : 1153
Aovo-DRB1*0304GA    : CAATCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  805
Aovo-DRB1*0305GA    : CAATCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  895
Aovo-DRB1*0306GA    : CAATCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  819
Aovo-DRB1*0307GA    : CAATCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  823
Aona-DRB3*0615      : CAGTCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGT---- :  851
Aona-DRB3*0627      : CCATCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  911
Aona-DRB3*062501    : CAGTCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGT---- :  851
Aona-DRB3*0626      : CAGTCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGT---- :  849
Aona-DRB3*062502    : CAGTCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGT---- :  829
Aona-DRB3*0628      : CCATCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  909
Aovo-DRB3*0601      : CAGTCTGTGAGCATTCAGAATCCTCTCCATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGT---- :  850
Aona-DRB*W8901      : CCATCTGTGAGAGTTCAGTATCCTCTCAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  560
Aona-DRB*W1808      : CCATCTGTGAGCGTTTAGAATCCTCTCAATCCTGAGCGGGGAGCTCTGAGGGCACAAGTGTGTGTGT-- :  645
Aona-DRB*W1806      : CCATCTGTGAGCGTTTAGAATCCTCTCAATCCTGAGCAGGGAGCTCCGAGGGCACAAGTGTGTGTGT-- :  635
Aovo-DRB*W1801      : CCATCTGTGAGCGTTTAGAATCCTCTCAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGTGT :  619
Aovo-DRB*W1802      : CCATCTGTGAGAGTTTAGAATCCTCTCAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  631
Aovo-DRB*W1803      : CCATCTGTGAGCGTTTAGAATCCTCTCAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  621
Aovo-DRB*W8801      : CCATCTGTGAGAGTTTAGAATCCTCTCAATCCTGAGCAAGGAGTTCTGAGGGCACAGATGTGTGTGT-- :  604
Aovo-DRB*W2901      : CCATCTGTGAGAGTTCAGTATCCTCTCAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  560
Aona-DRB*W2908      : CCATCTGTGAGAGTTCAGTATCCTCTCAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  560
Aona-DRB*W2910      : CCATCTGTGAGAGTTCAGTATCCTCTCAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  560
Aovo-DRB*W3001      : CCATCTGTGAGAGTTTAGAATCCTCTAAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  553
Aona-DRB*W3002      : CCATCTGTGAGAGTTTAGAATCCTCTAAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  557
Aovo-DRB*W9201      : CCATCTGTGAGAGTTTAGAATCCTCTAAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  551
Aovo-DRB*W9202      : CCATCTGTGAGAGTTTAGAATCCTCTAAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  551
Aona-DRB*W9101      : CCATCTGTGAGAGTTTAGAATCCTCTCAATCCTGAGCAAGGAGTTCTGAGGGCACAGATGTGTGTGT-- :  522
Aovo-DRB*W9101      : CCATCTGTGAGAGTTTAGAATCCTCTCAATCCTGAGCAAGGAGTTCTGAGGGCACAGATGTGTGTGT-- :  498
Aovo-DRB*W9102      : CCATCTGTGAGAGTTTAGAATCCTCTCAATCCTGAGCAAGGAGTTCTGAGGGCACAGATGTGTGTGT-- :  522
Aovo-DRB*W9001      : CCATCTGTGAGAGTTCAGTATCCTCTCAATCCTGAGCAAGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  501
Aovo-DRB*W4501      : CCATCTGTGAGAGTTTAGAATCCTCTAAATCCTGAGCAGGGAGCTCTGAGGGCACAGGTGTGTGTGT-- :  540
Aovo-DRB*W9301      : CCATCTGTGAGCATTCAGTATCCTCTCAATCCTGAGCAGGGAGCTCTGAGGGCACAGATGTGTGTGT-- :  475

                             1180         *      1200         *      1220         *      1240
                    AGAGTGTGGATTTGTGTG G GGCTGTTGTGgGagGgGAgGCAGGAGGGGGGCTTCTTC TA CCTTGGA
Aona-DRB1*0328GB    : AGAGTGTGGATTTGTGTGTGTGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTACCCTTGGA : 1085
Aona-DRB1*0329GA    : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA : 1218
Aona-DRB1*031701GA  : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA : 1222
Aovo-DRB1*0304GA    : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  874
Aovo-DRB1*0305GA    : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  964
Aovo-DRB1*0306GA    : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  888
Aovo-DRB1*0307GA    : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  892
Aona-DRB3*0615      : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  920
Aona-DRB3*0627      : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  980
Aona-DRB3*062501    : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  920
Aona-DRB3*0626      : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  918
Aona-DRB3*062502    : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  898
Aona-DRB3*0628      : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  978
Aovo-DRB3*0601      : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  919
Aona-DRB*W8901      : AGAGTGTGGATTTGTGTGTGTGGCTGTTGTGAG-AGGGAGGCAGGAGGGGGCTTCTTCTTACCCTTGGA :  628
Aona-DRB*W1808      : AGAGTGTGGATTTGTGTGTGTGGCTGTTGTGGGAGGGGAAGCAGGAGGGGGCTTCTTCTTATCCTTGGA :  714
Aona-DRB*W1806      : AGAGTGTGGATTTGTGTGTGTGGCTGTTGTGGGAGGGGAAGCAGGAGGGGGCTTCTTCTTATCCTTGGA :  704
Aovo-DRB*W1801      : AGAGTGTGGATTTGTGTGTGTGGCTGTTGTGGGAGGGGAAGCAGGAGGGGGCTTCTTCTTATCCTTGGA :  688
Aovo-DRB*W1802      : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  700
Aovo-DRB*W1803      : AGAGTGTGGATTTGTGTGTGTGGCTGTTGTGGGAGGGGAAGCAGGAGGGGGCTTCTTCTTATCCTTGGA :  690
Aovo-DRB*W8801      : AGAGTGTGGATTTGTGTGCGTGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCTTACCCTTGGA :  673
Aovo-DRB*W2901      : AGAGTGTGGATTTGTGTGTGTGGCTGTTGTGAGAGGGGAGGCAGGAGGGGGCTTCTTCTTACCCTTGGA :  629
Aona-DRB*W2908      : AGAGTGTGGATTTGTGTGTGTGGCTGTTGTGAGAGGGGAGGCAGGAGGGGGCTTCTTCTTACCCTTGGA :  629
Aona-DRB*W2910      : AGAGTGTGGATTTGTGTGTGTGGCTGTTGTGAGAGGGGAGGCAGGAGGGGGCTTCTTCTTACCCTTGGA :  629
Aovo-DRB*W3001      : AGAGTGTGGATTTGTGTGAGAGGCTGTTGTGGGAGGAGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  622
Aona-DRB*W3002      : AGAGTGTGGATTTGTGTGAGAGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCCTATCCTTGGA :  626
Aovo-DRB*W9201      : AGAGTGTGGATTTGTGTGAGAGGCTGTTGTGGGAGGAGAGGCAGGAGGGGGCTTCTTCTTATCCTTGGA :  620
Aovo-DRB*W9202      : AGAGTGTGGATTTGTGTGAGAGGCTGTTGTGGGAGGAGAGGCAGGAGGGGGCTTCTTCTTATCCTTGGA :  620
Aona-DRB*W9101      : AGAGTGTGGATTTGTGTGCGTGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCTTACCCTTGGA :  591
Aovo-DRB*W9101      : AGAGTGTGGATTTGTGTGCGTGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCTTACCCTTGGA :  567
Aovo-DRB*W9102      : AGAGTGTGGATTTGTGTGCGTGGCTGTTGTGGGAGGGGAGGCAGGAGGGGGCTTCTTCTTACCCTTGGA :  591
Aovo-DRB*W9001      : AGAGTGTGGATTTGTGTGTGTGGCTGTTGTGAGAGGGGAGGCAGGAGGGGGCTTCTTCTTACCCTTGGA :  570
Aovo-DRB*W4501      : AGAGTGTGGATTTGTGTGAGAGGCTGTTGTGGGAGGAGAGGCAGGAGGGGGCTTCTTCTTATCCTTGGA :  609
Aovo-DRB*W9301      : AGAGTGTGGATTTGTGTGTGAGGCTGTTGTGAGAGGCGAGGCAGGAGGGGGCTTCTTCTTATCCTTGGA :  544
```

```
                                *      1260          *      1280          *      1300          *
                          Ggcctct gtg  gagg gaca   gagg gg t cagggg tggaga ggaggagacct gattgtcc
Aona-DRB1*0328GB    : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTGGATTGTCC : 1152
Aona-DRB1*0329GA    : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC : 1285
Aona-DRB1*031701GA  : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC : 1289
Aovo-DRB1*0304GA    : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC :  941
Aovo-DRB1*0305GA    : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC : 1031
Aovo-DRB1*0306GA    : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC :  955
Aovo-DRB1*0307GA    : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC :  959
Aona-DRB3*0615      : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC :  987
Aona-DRB3*0627      : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC : 1047
Aona-DRB3*062501    : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC :  987
Aona-DRB3*0626      : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC :  985
Aona-DRB3*062502    : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC :  965
Aona-DRB3*0628      : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC : 1045
Aovo-DRB3*0601      : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGGGG-TGGAGAGGGAGGAGACCTCGATTGTCC :  986
Aona-DRB*W8901      : GGCCTCT-GTGAGGAGGTGACATGGGAGGCGGGTGCAGGGG-TGGAGAGGGAGGAGACCTGGATTGTCC :  695
Aona-DRB*W1808      : GGCCTCTTGTGTTGAGGGGACATGTGAGGTGGCTGCAGGGGCTGGAGAGGGAGGAGACCTCGATTGTCC :  783
Aona-DRB*W1806      : GGCCTCTTGTGTTGAGGGGACATGTGAGGTGGCTGCAGGGGCTGGAGAGGGAGGAGACCTCGATTGTCC :  773
Aovo-DRB*W1801      : G------------------------------------------------------------------- :  689
Aovo-DRB*W1802      : GGCCTCTTGTGTTGAGGGGACATGTGAGGTGGCTGCAGGGGCTGGAGAGGGAGGAGACCTCGATTGTCC :  769
Aovo-DRB*W1803      : GGCCTCTTGTGTTGAGGGGACATGTGAGGTGACTGCAGGGGCTGGAGAGGGAGGAGACCTCGATTGTCC :  759
Aovo-DRB*W8801      : GGCCTCT-GTGAGGAGGTGACATGGGAGGCGGGTGCAGGGG-TGGAGACGGAGGAGACCTGGATTGTCC :  740
Aovo-DRB*W2901      : GGCCTCT-GTGAGGAGGTGACATGGGAGGCGGGTGCAGGGG-TGGAGAGGGAGGAGACCTGGATTGTCC :  696
Aona-DRB*W2908      : GGCCTCT-GTGAGGAGGTGACATGGGAGGCGGGTGCAGGGG-TGGAGAGGGAGGAGACCTGGATTGTCC :  696
Aona-DRB*W2910      : GGCCTCT-GTGAGGAGGTGACATGGGAGGCGGGTGCAGGGG-TGGAGAGGGAGGAGACCTGGATTGTCC :  696
Aovo-DRB*W3001      : G------------------------------------------------------------------- :  623
Aona-DRB*W3002      : GGCCTCT-GTGGGGAGGTGACACAGGAGGTGGGTGCAGAGG-TGGGGAGGGAGGAGACCTCGTTTGTCA :  693
Aovo-DRB*W9201      : G------------------------------------------------------------------- :  621
Aovo-DRB*W9202      : G------------------------------------------------------------------- :  621
Aona-DRB*W9101      : GGCCTCT-GTGAGGAGGTGACATGGGAGGCGGGTACAGGGG-TGGAGACGGAGGAGACCTGGATTGTCC :  658
Aovo-DRB*W9101      : GGCCTCT-GTGAGGAGGTGACATGGGAGGCGGGTACAGGGG-TGGAGACGGAGGAGACCTGGATTGTCC :  634
Aovo-DRB*W9102      : GGCCTCT-GTGAGGAGGTGACATGGGAGGCGGGTACAGGGG-TGGAGACGGAGGAGACCTGGATTGTCC :  658
Aovo-DRB*W9001      : GGCCTCT-GTGAGGAGGTGACATGGGAGGCGGGTGCAGGGG-TGGAGAGGGAGGAGACCTGGATTGTCC :  637
Aovo-DRB*W4501      : G------------------------------------------------------------------- :  610
Aovo-DRB*W9301      : G------------------------------------------------------------------- :  545

                          1320          *      1340          *      1360          *
                          t ggtccttagagat caggaa  g aa   tga gtgtgtgtggctggggtgagggttta
Aona-DRB1*0328GB    : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1212
Aona-DRB1*0329GA    : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1345
Aona-DRB1*031701GA  : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1349
Aovo-DRB1*0304GA    : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1001
Aovo-DRB1*0305GA    : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1091
Aovo-DRB1*0306GA    : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1015
Aovo-DRB1*0307GA    : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1019
Aona-DRB3*0615      : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1047
Aona-DRB3*0627      : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1107
Aona-DRB3*062501    : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1047
Aona-DRB3*0626      : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1045
Aona-DRB3*062502    : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1025
Aona-DRB3*0628      : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1105
Aovo-DRB3*0601      : TTGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAAGTGTGTGTGGCTGGGGTGAGGGTTTA : 1046
Aona-DRB*W8901      : TGGGTCCTTAGAGATGCAGGAAGGGAAATG-TGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  755
Aona-DRB*W1808      : TGGGTCCTTAGAGATGCAGGAAGGGAAATG-TGAAGTGTGTGTGGCTGGGGTGAGGGTTTA :  843
Aona-DRB*W1806      : TGGGTCCTTAGAGATGCAGGAAGGGAAATG-TGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  833
Aovo-DRB*W1801      : ---------------------------------------------------------- :    -
Aovo-DRB*W1802      : TGGGTCCTTAGAGATGCAGGAAGGGAAATG-TGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  829
Aovo-DRB*W1803      : TTGGTCCTTAGAGATGCAGGAAGGGAAATG-TGAAGTGTGTGTGGCTGGGGTGAGGGTTTA :  819
Aovo-DRB*W8801      : TGGGTCCTTAGAGATTCAGGAATGGAAATG-TGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  800
Aovo-DRB*W2901      : TGGGTCCTTAGAGATGCAGGAAGGGAAATG-TGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  756
Aona-DRB*W2908      : TGGGTCCTTAGAGATGCAGGAAGGGAAATG-TGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  756
Aona-DRB*W2910      : TGGGTCCTTAGAGATGCAGGAAGGGACCTG-TGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  756
Aovo-DRB*W3001      : ---------------------------------------------------------- :    -
Aona-DRB*W3002      : TTGGTCCTTAGAGATGCAGGAATGGAAATG-TGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  753
Aovo-DRB*W9201      : ---------------------------------------------------------- :    -
Aovo-DRB*W9202      : ---------------------------------------------------------- :    -
Aona-DRB*W9101      : TGGGTCCTTAGAGATTCAGGAA-TGGAAATGTGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  718
Aovo-DRB*W9101      : TGGGTCCTTAGAGATTCAGGAA-TGGAAATGTGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  694
Aovo-DRB*W9102      : TGGGTCCTTAGAGATTCAGGAA-TGGAAATGTGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  718
Aovo-DRB*W9001      : TGGGTCCTTAGAGATGCAGGAA-GGGAAATGTGAGGTGTGTGTGGCTGGGGTGAGGGTTTA :  697
Aovo-DRB*W4501      : ---------------------------------------------------------- :    -
Aovo-DRB*W9301      : ---------------------------------------------------------- :    -
```

Capítulo 3. Structural analysis of owl monkey MHC-DR shows that fully protective malaria vaccine components can be readily used in humans

Suárez CF, Pabón L, Barrera A, Aza-Conde J, Patarroyo MA, Patarroyo ME. Structural analysis of owl monkey MHC-DR shows that fully-protective malaria vaccine components can be readily used in humans. Biochem Biophys Res Commun. 2017;491(4):1062-1069.

La versión publicada del artículo puede ser consultada en:

http://www.sciencedirect.com/science/article/pii/S0006291X17315486

Structural analysis of Owl monkey MHC-DR shows that fully-protective malaria vaccine components can be readily used in humans

Carlos F. Suárez[a,b,c], Laura Pabón[a], Ana Barrera[a], Jorge Aza-Conde[a],

Manuel Alfonso Patarroyo[a,b], Manuel Elkin Patarroyo[a,d,*]

[a] Fundación Instituto de Inmunología de Colombia (FIDIC), Bogotá D.C., Colombia

[b] Universidad del Rosario, Bogotá D.C., Colombia

[c] Universidad de Ciencias Aplicadas y Ambientales (UDCA), Bogotá, Colombia

[d] Universidad Nacional de Colombia, Bogotá DC, Colombia.

* Corresponding author. e-mail: mepatarr@gmail.com

Abstract

More than 50 years ago the owl monkey (genus *Aotus*) was found to be highly susceptible to developing human malaria, making it an excellent experimental model for this disease. Microbes and parasites' (especially malaria) tremendous genetic variability became resolved during our malaria vaccine development, involving conserved peptides having high host cell binding activity (cHABPs); however, cHAPBs are immunologically silent and must be specially modified (mHABPs) to induce a perfect fit into major histocompatibility complex (MHC) molecules (HLA in humans). Since malarial immunity is mainly antibody-mediated and controlled by the HLA-DRB genetic region, ~1,000 *Aotus* have been molecularly characterised for MHC-DRB, revealing striking similarity between human and *Aotus* MHC-DRB repertories. Such convergence suggested that a large group of immune protection-inducing protein structures (IMPIPS), highly immunogenic and protection inducers against malarial intravenous challenge in *Aotus,* could easily be used in humans for inducing full protection against malaria. We highlight the value of a logical and rational methodology for developing a vaccine in an appropriate animal model: *Aotus* monkeys.

Keywords:

MHC-DR, animal model, IMPIPS, malarial-vaccine, HLA-peptide binding

Introduction

Searching for an appropriate experimental model for human malaria research, Carl Johnson's group [1] demonstrated that *Plasmodium vivax* malaria could be transmitted to *Aotus* monkeys through infected human blood and to humans via an infected mosquito's bite, thereby replicating the malaria parasite's biological cycle. Contacos & Collins [2] repeated the trial, infecting *Aotus* with *P. falciparum*-infected blood and humans by mosquito bite, concluding that *Aotus* is an excellent experimental model for human malaria research. Many human *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale* parasite strains have now been adapted to grow in *Aotus*. Such primates are native to Panama and tropical South America; we have been using them during the last 35 years in the search for a logical and rational vaccine development methodology.

Aotus reduce dangerous, cumbersome and expensive human trials to a minimum as they involve thousands of people who have to be followed-up for years. Our experimental guidelines regarding the animal model follow stringent methodology, followed by meticulous immunological analysis [3, 4]. A very robust, sensitive and specific methodology has emerged from working with modified high-activity binding peptides (mHABPs) derived from conserved high-activity binding peptides (cHABPs) from the most relevant proteins involved in host (red blood, hepatocyte and endothelial cell) binding and invasion. This has led to defining some specific principles and rules for vaccine development [3, 4].

MHC-mediated antigen presentation represents the first step in inducing immune protection; HLA-DR molecules have two very deep pockets (P1, P9) in their peptide binding region (PBR). Along with two shallow ones (P4, P6), they enable a perfect antigen fit for establishing H-bonds and

becoming fixed for proper presentation to TCR molecules, thereby activating an appropriate immune response.

We thus characterised the main *Aotus* immune system components by molecular biology such as MHC-I/II [5-11] and other molecules as TCR [12, 13], finding 80%-100% similarity with human counterparts, thus enabling information regarding *Aotus* to be extrapolated to vaccines for human use. In-depth understanding of antigen presentation involved studying ~1,000 *Aotus* monkeys; 215 MHC-DRB sequences were obtained [8-11], analysed and grouped into lineages according to their sequences. Similarity with human HLA-DRB allele lineages was investigated by generating pocket profiles. Molecular modelling methods were used to generate *Aotus*-MHC-DR pockets from HLA-DR molecules whose structure had already been determined by X-ray crystallography; *Aotus*/human variant residues were replaced to determine their impact on volume and electrostatic characteristics regarding experimentally-obtained results in *Aotus* for using such peptides as fully-effective vaccine components.

Materials and methods

Pocket profiling

The main problem when dealing with great MHC-DRB allele diversity was resolved by abstracting sequences to a "pocket dictionary" which estimated unique pocket variety, defined by key contact-residues involved in peptide binding. Pocket profiles were determined by the occurrence of a specific amino-acid (aa) combination in MHC pockets defined from previous crystallographic studies [14]. The most frequently occurring profiles, named by allele prototype were determined for each allele lineage (PPF in Figure 1); translated HLA-DRB sequences reported in the IMGT

[15] were used for humans along with *Aotus*-MHC-DRB sequences previously reported by our group [8-11], Allele Frequency Net Database (http://www.allelefrequencies.net) was used for calculating allele lineage frequency for humans and our previous surveys for *Aotus* (% in red, Figure 1) [8-11]. IMPIPS' potential population coverage was calculated as the product of MHC lineage probability and the probability of the profile on which it was designed (% in blue, Figure 1). PAM250 matrix was used for calculating average percentage identity and similarity between HLA-DRB and *Aotus*-DRB lineages.

HLA-DR peptide-binding prediction

NETMHCIIPAN-3.1 [16], the best available tool for peptide-binding prediction, was used for predicting peptide-HLA-DRB allele binding affinity with peptide vaccine candidates and evaluating residue affinity for each pocket. We categorised epitopes as being strong binders (≤100 nM), binders (>100 to ≤500 nM) and non-binders (>500 nM). The pocket profile approach selected 65 HLA-DRB allele prototypes for predictions, covering at least 60% of the pocket profiles displayed in each HLA-DRB1 lineage (% in green, Figure 1): DRB1*0101/02/04/09/06, DRB1*0301/02/05/25/13, DRB1*0401/02/03/04/05/06/07/08/22, DRB1*0701/04/03/06/24, DRB1*0801/02/04/05/06/12/24/34, DRB1*0901/02, DRB1*1001/02, DRB1*1101/02/04/06/11/10, DRB1*1201/16, DRB1*1301/02/03/12/07/05, DRB1*1401/05/03/04/14/06/08/25/32, DRB1*1501/02/03, DRB1*1601/04/15.

MHC-DR modelling and analysis

HLA-DRβ1*0101 (PDB-1DLH), HLA-DRβ1*1501 (PDB-1BX2), HLA-DRβ1-03 (PDB-1A6A) and HLA-DRβ1*04 (PDB-1J8H) crystallographic structures were used as templates for sterically localising residue/aa differences between humans and *Aotus*. Since *Aotus* MHC-DR structure has

not been described, molecular modelling (Insight II energy minimisation analysis) involved replacing β-chain residues for obtaining energetically-favoured structures [17].

Residues forming P1, P4, P6, P9 (Figure 1 and Figure 2 for β-chain residues, since α-chain residues are conserved) were used for each complex; human and *Aotus* electrostatic potential surface and volume were determined via UCSF Chimera package. APBS was used to evaluate each pocket's electrostatic surface potential; solvent-accessible potential surface values were set from 7 *kT/e* (negative charge, red) to 7 *kT/e* (positive, blue) [18].

Peptides used for immunisation, protection and 3D structure determination

Chemically-synthesised peptides used for 600 MHz ^{1}H-NMR spectrometry 3D structure determination, assessing *Aotus* immunisation, challenge, protection and infection, determining immunofluorescence antibody test (IFA) and western blot (WB) reactivity with *P. falciparum* lysates have been thoroughly described [4].

Results and Discussion

Analysing *Aotus* class II gene MHC-DRB sequences revealed 17 allele lineages' striking convergence with human HLA-DRβ lineages [8-11] (~82% mean similarity in MHC-DRB β1 domain) (Figure 1). Remarkably, no allele differences were observed between humans/*Aotus* in large hydrophobic P1, since both had dimorphic variation β86G (accepting aromatic residues W, Y, F) or β86V (accepting large aliphatic residues L, I, M, V) in all allele lineages [19]. Human HLA-DRβ1* and *Aotus* DR-like allele lineages' variant ratio (β86G↔β86V) was the main difference between humans and *Aotus* regarding P1. A detailed analysis follows regarding alleles having differences or similarities between humans and *Aotus*.

HLA-DRβ1*03 lineage

The human HLA-DRβ1*03 lineage covers 19.7% of the global population, 5 allele-prototypes accounting for 65.9% of HLADRβ1*03 pocket profiles whilst Aotus-MHC-DRβ1*03 lineage accounts for 57.1% of its population. Figure 1 shows *Aona*- DRβ1*0305/07/04/09 alleles as being almost identical to human HLA-DRβ1*0302/05 regarding aa sequence and *Aona*-DRβ1*0311 being identical to HLA-DRβ1*0301, 0325 and 0313 alleles.

β86V as predominant dimorphic allele (~80%) in humans compared to ~20% in *Aotus* represented the difference between humans and *Aotus* in P1; P4 was almost identical electrostatically and volumetrically in both species, accommodating D and S.

HLA-DRβ1*03 structure revealed differences in P6 (adjoining the PBR groove), as Fβ9E and Qβ10Y had similar volume (131.8 $Å^3$ *cf* 139.7 $Å^3$) and charge, were far apart in P6 side wall in humans and did not interact directly with a peptide, having no impact on antigen presentation; so, binding prediction preferences for R, K, P, S could be equivalent for both species.

P9 Fβ9E and Yβ37N differences (Figure 1) made it slightly larger (202.7 $Å^3$ *cf* 196.4 $Å^3$) and more pi(π)-charged in *Aotus* (Figure 2, row 1, columns 5-6). The aforementioned residues plus Yβ30, Yβ60 and Wβ61 conserved residues formed P9 in both species. Such electrostatic and volumetric difference induced *Aotu*s to prefer aromatic residue Y rather than R as in humans; peptide-binding prediction gave R, Y, S as classical binding motifs for P9. However, alleles HLA-DRβ1*0338, 0319, 0313, 0310 also bound apolar residues V, L, I. The five HLA-DRβ1*03-binding IMPIPS protecting *Aotus* monkeys against intravenous challenge with fresh, living *P. falciparum* parasites could thus be readily used to protect ~13.0% of the human population, since both allele lineages are almost identical in both species.

Figure 3A gives an excellent example of AMA-1 cHABP 4313-derived **10022** IMPIPS fully protecting *Aotus*, displaying all HLA-DRβ1*0302 and 1312 allele binding molecular characteristics. Theoretically, interaction with HLA-DRβ1*0302 could protect ~1.0% and ~1.2% of the world's population, respectively, according to this pocket profile's frequency in humans (Figure 1); 4 more IMPIPS binding to the other HLA-DRβ1*03 lineage would thus be required to protect the remaining human population.

HLA-DRβ1*04 lineage

Aona-DRβW4704/01/02/03/05/08/09 alleles (in 21.5% of the *Aotus* population) are quite similar to HLA-DRβ1*0401/05/03/04/02/06 alleles (in 26.1% of humans). Dimorphic variants in P1 were quite similar between both species (26.3% in humans, 23.1% in *Aotus*). P4 differences (*Aotus* P4, Dβ70Q and Qβ74A/E, Figure 1) made it slightly smaller (134.0 $Å^3$ *cf* 149.5 $Å^3$) and more apolar in *Aotus* (Figure 2, row 2, columns 5-6), mainly accepting residues L, M, I; all predominated in peptide-binding prediction HLA-DRβ1*04 binding motifs. Ile was always avoided when designing peptides due to its unfavourable $PPII_L$-forming propensity. However, 1/3 of HLA-DRβ1*04 alleles (i.e. HLA-DRβ1*0422/01,07,09,16,17,21,33,34,35) also received D in P4, common in *Aotus.* P6 was identical in both species and allele lineage ratios were very similar.

P9 was almost identical, receiving S, T, D in both species; however, alleles HLA-DRβ1*0422, 25, 44, 55 also received R, K. Thus, nine HLA-DRβ1*04-binding IMPIPS (having highly immunogenic and protection-inducing characteristics in *Aotus* which could be readily used in humans) could protect ~15.6% of the human population (Figure 1). Figure 3B gives another clear example regarding *Aotus* EBA-175 cHABP 1758-derived **13790** IMPIPS (equivalent to HLA-DRβ1*0422-binding IMPIPS for human use) could protect ~0.26% of the world's population.

HLA-DRβ1*15 lineage

This lineage covers 24.6% of the human population; the four allele-prototypes shown here represent 76.9% of lineage pocket profiles. As this lineage is present in only 2% of the *Aotus* population, it hampered identifying IMPIPS for these alleles. Differences with *Aona*-DRβ*03GC in HLA-DRβ1*15 lineages were minimal in P1, where large aliphatic residues L, M, V, I were preferred due to β86V dimorphic variant predominance (>85% in humans and almost 100% in *Aotus*) in both groups (Figure 1). Eβ70Q and Tβ71A replacements slightly reduced P4 size in *Aotus,* preferring large aromatic F, Y or large aliphatic L, M, I residues, as described by peptide-binding prediction (F, L, M, I, Y, respectively). P6 replacements F9βW (far apart in pocket wall), Tβ71A (not forming this pocket), Tβ11P, Tβ12K and Sβ13R had little relevance, since P6 had little involvement in this lineages' peptide binding. Critical, specific P7 was identical for humans and *Aotus.*

Regarding P9 replacements, HLA-DRβ1*1501 3D-structure (Figure 2, row 3, column 4) showed Sβ13R far away from P9 floor whilst Yβ37S and Eβ57D (on pocket floor) made it negatively-charged in *Aotus*, showing a slightly greater volume than in humans (244.0 $Å^3$ *cf* 223.6 $Å^3$) (Figure 2, row 3, column 5-6). Positively-charged residues R, K were preferred in *Aotus* rather than L, V, also accepted in humans. Fβ61W replacement was a relevant difference, since nitrogen from the indole ring in aromatic β9W (absent in aromatic *Aona* β9F) established one H-bond with the peptide's backbone, stabilising this MHCII-peptide complex.

Therefore, four IMPIPs having *Aotus*-protecting characteristics could be readily used in humans could protect ~18.9% of the human population against *P. falciparum* malaria. Unfortunately, we lack 3D-structures for the few HLA-DRβ1*15-binding IMPIPS identified to date.

HLA-DRβ1*01 lineage

This HLA-DRβ1* lineage accounts for 17% of the global population; its 5 allele-prototypes comprises 72.3% of HLA-DRβ*01 pocket profiles. However, the counterpart (Aona-DRβ*W43 lineage) in only 2.4% of the *Aotus* population hampered finding monkeys having genetic traits equivalent to HLA-DRβ1*01 and HLA-DRβ1*15 for identifying IMPIPs fitting into these alleles' PBR.

Gβ86V dimorphism frequency was similar in P1 (75% in *Aotus*, 60% in humans). P4 was almost identical; Dβ28E and Nβ70Q differences did not make any substantial volumetric (150.5 $Å^3$ *cf* 163.4 $Å^3$) or electrostatic differences since replacements were electrostatically similar (D↔E, N↔Q). Binding motifs were the same, preferentially accepting apolar residues L, M, V, I in both species.

Regarding highly specific and relevant HLA-DRβ1*01, *Aotus* and humans had differences regarding Kβ9W in P6, far apart in pocket wall according to 3D-model (Figure 2, row 4, column 3), and Vβ11L, Cβ13F, Hβ30C (on top of P6) having equivalent volume in *Aotus* and humans (147.9 $Å^3$ *cf* 149.6 $Å^3$ respectively) being somewhat negatively-charged in *Aotus* (Figure 2, row 4, columns 5-6). Therefore, positively-charged (N), alcohol-derived S, T or apolar P residues were preferred for binding in *Aotus* P6. Apolar residues A, P, G were preferred in humans, according to peptide-binding prediction.

Cβ13F and Yβ37S replacements in deep hydrophobic P9 were structurally far apart on P9 floor (Figure 2, row 4, column 4); Kβ9W and Hβ30C replacements were directly located on P9 floor thereby modifying its volume (247.0 $Å^3$ *cf* 198.4 $Å^3$) and electrostatic landscape, making it larger and negatively-charged in *Aotus* (Figure 2, row 4, column 5) (as in HLA-DRβ1*0104). This

enabled *Aotus* to accept large, positively-charged residues R, Y, while preferred residues in humans were apolar or large aliphatic L, I, V, M. Tβ57D replacement seemed a critical modification (Figure 1) due to canonical α76R=β57D salt bridge rupture, making P9 wider than deeper, small apolar residues A, S, T being preferred according to peptide-binding prediction. IMPIPs modified according to these characteristics could thus be used to protect ~12.3% of the human population against *P. falciparum* malaria. Figure 3C shows that MSP1 cHABP 1585-derived 10014 IMPIPS characterised as HLA-DRβ1*0101 allele-prototype could protect ~8.8% of the human population, based on this allele's frequency (Figure 1), or ~3.2% if bound to HLA-DRβ1*0901.

HLA-DRβ1*08 lineage

Comparing *Aona*-DRβ1*03-GB lineage (*Aona*-DRβ1*0302/01/26 covering 18.8% *Aotus* population) to human HLA-DRβ1*08 lineage (covering 8.1% human population) showed that pocket profiles shown here represent 69.5% of the HLA-DRβ1*08 lineage. The Aβ86G/V difference in P1 made it intermediate in size in *Aotus* between β86G and β86V dimorphic sequences; this large hydrophobic pocket could bind F, Y, L, I, V, M, but not W.

P4 was almost identical in both species; the Eβ70D difference had no impact on binding preference, fitting apolar residues L, M, V, S, A according to peptide-binding prediction. Fβ9E Qβ10Y differences (as in HLA-DRβ*03) were distant in P6 side wall and did not interact with peptide. Sβ13G slightly reduced P6 space in *Aotus*, maintaining polarity, and could bind residues R, K. Peptide-binding prediction indicated that P, S, A were equally accepted.

Replacements in Fβ9E (on the floor) and Sβ13G were especially distant in P9 and did not interact with peptide, therefore having no influence on human HLA-DRβ1*08 residue preference

regarding *Aona*-DRβ1*03GB. However, compared to HLA-DRβ1*03, the Yβ37N difference rendered P9 smaller, preferentially accepting apolar residues S, G, A and to a lesser extent L, V, I, M, as in HLA-DRβ1*0803, HLA-DRβ1*0810, HLA-DRβ1*0815, HLA-DRβ1*0830.

Dβ57S variation was another difference due to the aforementioned canonical α76R=β57D salt bridge rupture in both species (~40% in humans, 100% in *Aotus*), inducing preference for S, G, A and in aforesaid alleles like HLA-DRβ1*0803, preferring L, I, V, M. Similar to *Aotus*, mice I-Ag^7 MHC-II, HLA-DRβ*08 and *Aona*-DRβ0306B have β57S in P9, preferentially accepting residues G, S, A, D, E in optimum fitting conditions. P9 is wider than deeper in I-Ag^7, having greater lateral freedom than in other class II molecules; it accepts L, V, I, M in non-optimum conditions [20], as could happen in humans and *Aotus*.

Eight HLA-DRβ1*08-binding IMPIPS, inducing protection in *Aotus*, could thus be readily used to protect ~5.6% of the human population. Figure 3D provides another excellent example; MSP2 cHABP 4044-derived IMPIPS 24112, classified as HLA-DRβ1*0802, could cover ~1.1% of the world population (Figure 1). Characterised as HLA-DRβ1*1312, it would protect ~1.2% of the world's population (though having greater affinity for HLA-DRβ1*0802).

HLA-DRβ1*07 lineage

Human HLA-DRβ1*07 lineage covers 22.4% of the global population whilst convergent *Aona*-DRβ*W30 allele lineage is found in 20% of the *Aotus* population. Five HLA-DRβ1*07 pocket profiles represent 69.1% of the HLA-DRβ1-07 lineage (Figure 1).

β86G is the predominant dimorphic allele (>80% worldwide) in P1 in humans while this dimorphic allele is almost exclusive to *Aotus*-DRβ*W30, preferentially receiving aromatic residues F, Y, W.

HLA-DRβ1*04 modelling gave Eβ14K and Aβ73G replacements in HLA-DRβ1*07 lineage in P4 wall and Qβ74S and Yβ78V on the floor. Such electrostatic and volumetric differences made *Aotus* accept small apolar residues S or T, whilst humans could accept also larger apolar residues V, I, A, according to peptide-binding prediction.

The Eβ9W difference in P6 was far apart within the pocket and did not intervene in interaction with peptide whilst Vβ11G made P6 smaller and thus preferentially accept small aa S, A, G, as in humans according to peptide-binding prediction.

Eβ9W, Sβ57V, Kβ60S, Lβ61W in P9 showed that Kβ60S was above and far apart. Canonical α76R=β57D rupture meant that the salt bridge replaced here by Sβ57 would allow large aliphatic residues L, I to fit. The same could be happening in HLA-DRβ1*15 and HLA-DRβ1*08 as in HLA-DRβ1*07, as Lβ61W did not establish one H-bond (out of 13) with P9 due to the lack of pyrrole nitrogen, since the change involved β9W in HLA to β9L in *Aotus*-MHC, making IMPIPS having such weaker optimal characteristics fit this pocket.

Five HLA-DRβ1*07-binding IMPIPS, inducing protection in *Aotus*, could thus be readily used to protect 15.5% of the human population (Figure 1). Figure 3E shows that HRPII cHABP 6800-derived 24230 IMPIPS, characterised as HLA-DRβ1*0701, could protect as much as 11.2% of the human population, unfortunately being a short-memory protection inducing IMPIPS [23].

Implications for a vaccine development methodology

The foregoing suggests that the five IMPIPS shown here (Figure 3), inducing immunogenicity and full-protection regarding the most stringent challenge against *P. falciparum* malaria in *Aotus,* could

be used for human immunisations and in so doing protect 22.4% (considering the strongest binder only) of such population.

New IMPIPS derived from functionally-relevant cHABPs from proteins involved in RBC invasion designed to fit the lineages' pocket profiles presented above would suggest that the 36 IMPIPS mentioned below could protect ~80.9% of the world's population against *P. falciparum* malaria. This would involve 5 having HLA-DRβ1*03-binding characteristics (totally protecting *Aotus*) which could cover ~13% of the world's population, plus another 9 HLA-DRβ*04-binding IMPIPS (~15.6%), another 4 HLA-DRβ1*15-binding IMPIPS (~18.9%), plus 5 HLA-DRβ1*01-binding IMPIPS (~12.3%), 5 HLA-DRβ1*07-binding IMPIPS (~15.5%) and 8 HLA-DRβ1*08-binding IMPIPS (~5.6%).

According to our calculations, 14 additional IMPIPS covering all allele lineages representing the most frequently-occurring pocket profiles (giving 50 IMPIPS in total) could protect ~96.6% of the world's population [21] with a minimum of 1.19% IMPIPS recognised by 90% of the world's population. This approach could achieve the objective of developing a complete, totally-effective vaccine against pathogens, even complex parasites like *P. falciparum* which uses multiple proteins and complex strategies during invasion to escape the immune response [4].

The aforementioned volumetric and electrostatic findings regarding IMPIPS side-chains enabling a perfect fit into MHC-DR pockets according to allele lineage suggests their immediacy for use in humans as they have completely protected *Aotus*.

The great immunological similarity between humans and *Aotus* has allowed the development of a logical and rational methodology for developing complete, fully-protective, minimal subunit-based, multi-epitope, multi-stage chemically-synthesised universal vaccines for human use. This

has had to be complemented with already-described steric, electronic [3, 4, 22] and topological rules (i.e. 26.5 Å ± 1.5 Å) distance between P1 and P9 residues [23], φ and ψ torsion angles to induce $PPII_L$ conformation [24], correct side-chain orientation [25] and peripheral flanking residue preference [26]. These emerging rules, combined with a quantum chemistry approach to studying MHC-peptide binding [27], provides a strong framework for peptide-based vaccine design.

The forgoing, based on the aforementioned principles together with the use of *Aotus* as appropriate experimental model, has paved the way forward for effective vaccine development regarding malaria and other infectious diseases, as well as cancer induced by viruses, bacteria or parasites [28].

Conflict of interest

The authors declare that they have no financial/commercial conflicts of interest.

Acknowledgments

We would like to thank Mr Jason Garry for translating and revising the manuscript. This research was supported by Colciencias, contract 860-2015.

References

[1] M.D. Young, J.A. Porter, Jr., C.M. Johnson, Plasmodium vivax transmitted from man to monkey to man, Science, 153 (1966) 1006-1007.

[2] P.G. Contacos, W.E. Collins, Falciparum malaria transmissible from monkey to man by mosquito bite, Science, 161 (1968) 56-56.

[3] L.E. Rodriguez, H. Curtidor, M. Urquiza, G. Cifuentes, C. Reyes, M.E. Patarroyo, Intimate molecular interactions of P. falciparum merozoite proteins involved in invasion of red blood cells and their implications for vaccine design, Chem Rev, 108 (2008) 3656-3705.

[4] M.E. Patarroyo, A. Bermudez, M.A. Patarroyo, Structural and immunological principles leading to chemically synthesized, multiantigenic, multistage, minimal subunit-based vaccine development, Chem Rev, 111 (2011) 3459-3507.

[5] D. Diaz, M. Naegeli, R. Rodriguez, J.J. Nino-Vasquez, A. Moreno, M.E. Patarroyo, G. Pluschke, C.A. Daubenberger, Sequence and diversity of MHC DQA and DQB genes of the owl monkey Aotus nancymaae, Immunogenetics, 51 (2000) 528-537.

[6] C.F. Suarez, M.A. Patarroyo, M.E. Patarroyo, Characterisation and comparative analysis of MHC-DPA1 exon 2 in the owl monkey (Aotus nancymaae), Gene, 470 (2011) 37-45.

[7] P.P. Cardenas, C.F. Suarez, P. Martinez, M.E. Patarroyo, M.A. Patarroyo, MHC class I genes in the owl monkey: mosaic organisation, convergence and loci diversity, Immunogenetics, 56 (2005) 818-832.

[8] J.J. Nino-Vasquez, D. Vogel, R. Rodriguez, A. Moreno, M.E. Patarroyo, G. Pluschke, C.A. Daubenberger, Sequence and diversity of DRB genes of Aotus nancymaae, a primate model for human malaria parasites, Immunogenetics, 51 (2000) 219-230.

[9] J.E. Baquero, S. Miranda, O. Murillo, H. Mateus, E. Trujillo, C. Suarez, M.E. Patarroyo, C. Parra-Lopez, Reference strand conformational analysis (RSCA) is a valuable tool in identifying MHC-DRB sequences in three species of Aotus monkeys, Immunogenetics, 58 (2006) 590-597.

[10] C.F. Suarez, M.E. Patarroyo, E. Trujillo, M. Estupinan, J.E. Baquero, C. Parra, R. Rodriguez, Owl monkey MHC-DRB exon 2 reveals high similarity with several HLA-DRB lineages, Immunogenetics, 58 (2006) 542-558.

[11] C. Lopez, C.F. Suarez, L.F. Cadavid, M.E. Patarroyo, M.A. Patarroyo, Characterising a microsatellite for DRB typing in Aotus vociferans and Aotus nancymaae (Platyrrhini), PLoS One, 9 (2014) e96973.

[12] C.A. Moncada, E. Guerrero, P. Cardenas, C.F. Suarez, M.E. Patarroyo, M.A. Patarroyo, The T-cell receptor in primates: identifying and sequencing new owl monkey TRBV gene sub-groups, Immunogenetics, 57 (2005) 42-52.

[13] J.E. Guerrero, D.P. Pacheco, C.F. Suarez, P. Martinez, F. Aristizabal, C.A. Moncada, M.E. Patarroyo, M.A. Patarroyo, Characterizing T-cell receptor gamma-variable gene in Aotus nancymaae owl monkey peripheral blood, Tissue Antigens, 62 (2003) 472-482.

[14] L.J. Stern, J.H. Brown, T.S. Jardetzky, J.C. Gorga, R.G. Urban, J.L. Strominger, D.C. Wiley, Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide, Nature, 368 (1994) 215-221.

[15] J. Robinson, J.A. Halliwell, H. McWilliam, R. Lopez, P. Parham, S.G. Marsh, The IMGT/HLA database, Nucleic Acids Res, 41 (2013) D1222-1227.

[16] M. Andreatta, E. Karosiene, M. Rasmussen, A. Stryhn, S. Buus, M. Nielsen, Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification, Immunogenetics, 67 (2015) 641-650.

[17] M.S. Inc, Insight II User Guide, in: M.S. Inc (Ed.), Molecular Simulations Inc, San Diego, 1998.

[18] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis, J Comput Chem, 25 (2004) 1605-1612.

[19] C. Cardenas, J.L. Villaveces, H. Bohorquez, E. Llanos, C. Suarez, M. Obregon, M.E. Patarroyo, Quantum chemical analysis explains hemagglutinin peptide-MHC Class II molecule HLA-DRbeta1*0101 interactions, Biochem Biophys Res Commun, 323 (2004) 1265-1277.

[20] T. Stratmann, V. Apostolopoulos, V. Mallet-Designe, A.L. Corper, C.A. Scott, I.A. Wilson, A.S. Kang, L. Teyton, The I-Ag7 MHC class II molecule linked to murine diabetes is a promiscuous peptide binder, J Immunol, 165 (2000) 3214-3225.

[21] H.H. Bui, J. Sidney, K. Dinh, S. Southwood, M.J. Newman, A. Sette, Predicting population coverage of T-cell epitope-based diagnostics and vaccines, BMC Bioinformatics, 7 (2006) 153.

[22] A. Moreno-Vranich, M.E. Patarroyo, Steric-electronic effects in malarial peptides inducing sterile immunity, Biochem Biophys Res Commun, 423 (2012) 857-862.

[23] M.P. Alba, C.F. Suarez, Y. Varela, M.A. Patarroyo, A. Bermudez, M.E. Patarroyo, TCR-contacting residues orientation and HLA-DRbeta* binding preference determine long-lasting protective immunity against malaria, Biochem Biophys Res Commun, 477 (2016) 654-660.

[24] M.E. Patarroyo, A. Moreno-Vranich, A. Bermudez, Phi (Phi) and psi (Psi) angles involved in malarial peptide bonds determine sterile protective immunity, Biochem Biophys Res Commun, 429 (2012) 75-80.

[25] A. Bermudez, D. Calderon, A. Moreno-Vranich, H. Almonacid, M.A. Patarroyo, A. Poloche, M.E. Patarroyo, Gauche(+) side-chain orientation as a key factor in the search for an immunogenic peptide mixture leading to a complete fully protective vaccine, Vaccine, 32 (2014) 2117-2126.

[26] C. Reyes, R. Rojas-Luna, J. Aza-Conde, L. Tabares, M.A. Patarroyo, M.E. Patarroyo, Critical role of HLA-DRbeta* binding peptides' peripheral flanking residues in fully-protective malaria vaccine development, Biochem Biophys Res Commun, 489 (2017) 339-345.

[27] R. González, C.F. Suárez, H.J. Bohórquez, M.A. Patarroyo, M.E. Patarroyo, Semi-empirical quantum evaluation of peptide–MHC class II binding, Chemical Physics Letters, 668 (2017) 29-34.

[28] H. zur Hausen, The search for infectious causes of human cancers: where and why (Nobel lecture), Angew Chem Int Ed Engl, 48 (2009) 5798-5808.

Figure Legends

Figure 1.

Human HLA-DRβ1* and *Aona* DRβ* convergent allele lineages, showing their identical aa sequences in P1 (fuchsia), P4 (blue), P6 (orange) and P9 (green). Similar amino acids using volumetric or electrostatic criteria, are shown by lighter colours and dissimilar aa are not shaded. Allelic lineage percentage in the global population (% in red), number of HLA-DRB pocket profiles considered (n), IMPIPS´ potential global population coverage (% in blue), pocket profile frequency (PPF) and the final percentage covered by such profiles (% in green) are displayed. 36 IMPIPS fitting into these allele prototypes would thus protect ~80.9% of the human population. *Aotus nancymaae* (Aona), *A. vociferans* (Aovo), *A. nigriceps* (Aoni).

Figure 2.

The first column shows the HLA-DR molecule a-chain in magenta, β-chain in pale blue (both shown in ribbon); the aa forming Pocket 1 are shown by fuchsia balls, Pocket 4 in dark blue, Pocket 6 orange and Pocket 9 green. Residues differing amongst HLA-DR β1* and Aotus-MHC-DR are highlighted by red balls. Columns 2, 3 and 4 show Pockets 4, 6 and 9 surface conforming residues (differences highlighted in red). Columns 5 (Aotus) and 6 (Human) give a top/side view of selected pockets, showing determined volume (Å3)

Figure 3.

Side, front and top view of cHABP-derived protein 3D-structure (bold letters) (mHABPs, bold numbers). Corresponding aa sequences highlighted in colour; residues fitting into HLA-DRβ1*

are indicated and regions having PPIIL conformation are underlined. The yellow box contains HLA-DRβ1* allele binding (≤ 100 nM, in bold highest affinity) with IC (<200), IFA antibody titre reciprocals (II_{20}/III_{20}: 20 days post-second and 20 days post-third dose, respectively) and the amount of monkeys protected after intravenous challenge (Prot, highlighted in red). Below the side-view the distance (Å) between residues fitting into HLA-DRb1* molecules PBR P1 to P9 is shown.

Figure 1.

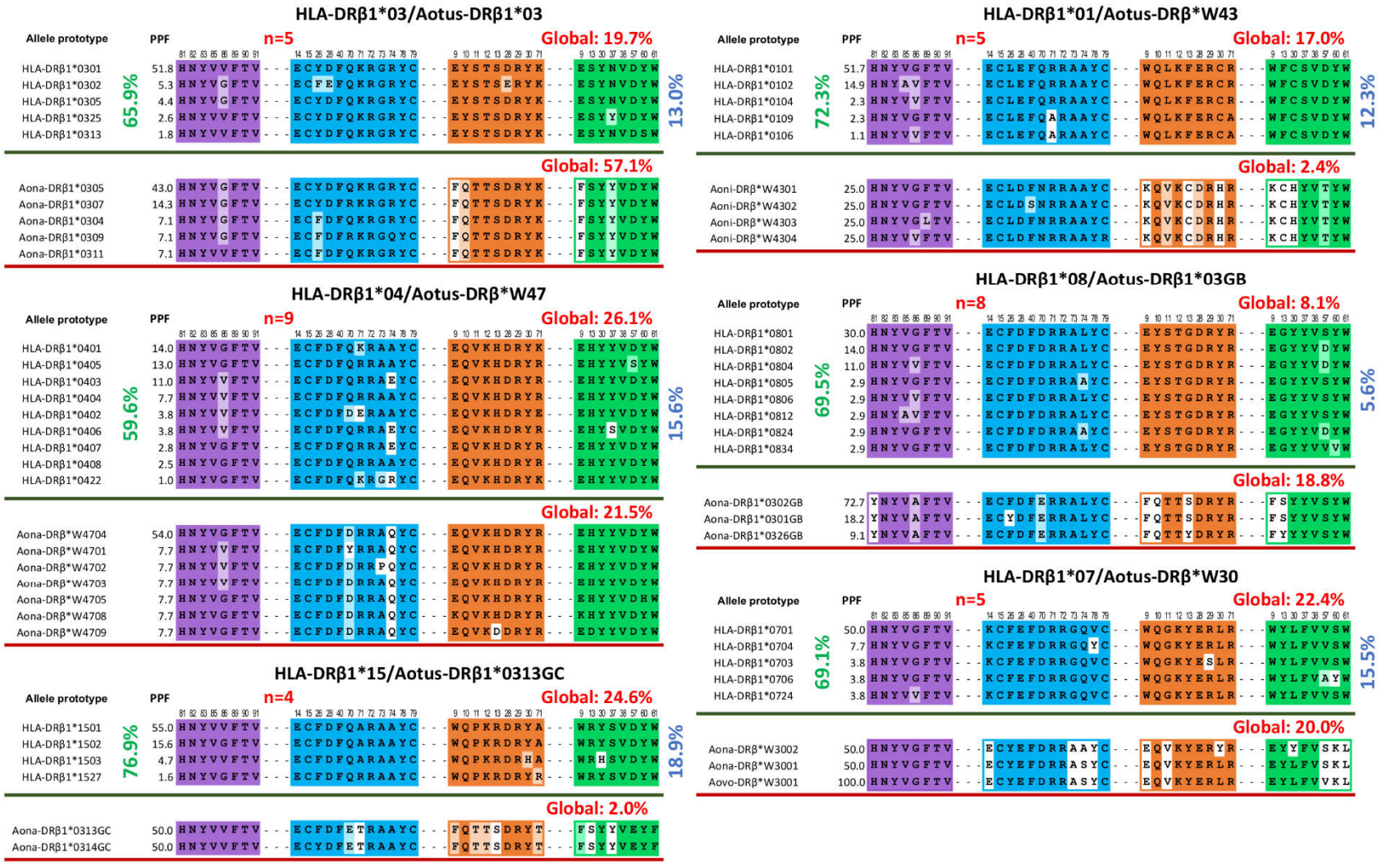

HLA-DRβ1*03/Aotus-DRβ1*03
Allele prototype
PPF
n=5
Global: 19.7%
HLA-DRβ1*0301 51.8
HLA-DRβ1*0302 5.3
HLA-DRβ1*0305 4.4
HLA-DRβ1*0325 2.6
HLA-DRβ1*0313 1.8
65.9%
13.0%
Global: 57.1%
Aona-DRβ1*0305 43.0
Aona-DRβ1*0307 14.3
Aona-DRβ1*0304 7.1
Aona-DRβ1*0309 7.1
Aona-DRβ1*0311 7.1
HLA-DRβ1*04/Aotus-DRβ*W47
Allele prototype
PPF
n=9
Global: 26.1%
HLA-DRβ1*0401 14.0
HLA-DRβ1*0405 13.0
HLA-DRβ1*0403 11.0
HLA-DRβ1*0404 7.7
HLA-DRβ1*0402 3.8
HLA-DRβ1*0406 3.8
HLA-DRβ1*0407 2.8
HLA-DRβ1*0408 2.5
HLA-DRβ1*0422 1.0
59.6%
15.6%
Global: 21.5%
Aona-DRβ*W4704 54.0
Aona-DRβ*W4701 7.7
Aona-DRβ*W4702 7.7
Aona-DRβ*W4703 7.7
Aona-DRβ*W4705 7.7
Aona-DRβ*W4708 7.7
Aona-DRβ*W4709 7.7
HLA-DRβ1*15/Aotus-DRβ1*0313GC
Allele prototype
PPF
n=4
Global: 24.6%
HLA-DRβ1*1501 55.0
HLA-DRβ1*1502 15.6
HLA-DRβ1*1503 4.7
HLA-DRβ1*1527 1.6
76.9%
18.9%
Global: 2.0%
Aona-DRβ1*0313GC 50.0
Aona-DRβ1*0314GC 50.0
HLA-DRβ1*01/Aotus-DRβ*W43
Allele prototype
PPF
n=5
Global: 17.0%
HLA-DRβ1*0101 51.7
HLA-DRβ1*0102 14.9
HLA-DRβ1*0104 2.3
HLA-DRβ1*0109 2.3
HLA-DRβ1*0106 1.1
72.3%
12.3%
Global: 2.4%
Aoni-DRβ*W4301 25.0
Aoni-DRβ*W4302 25.0
Aoni-DRβ*W4303 25.0
Aoni-DRβ*W4304 25.0
HLA-DRβ1*08/Aotus-DRβ1*03GB
Allele prototype
PPF
n=8
Global: 8.1%
HLA-DRβ1*0801 30.0
HLA-DRβ1*0802 14.0
HLA-DRβ1*0804 11.0
HLA-DRβ1*0805 2.9
HLA-DRβ1*0806 2.9
HLA-DRβ1*0812 2.9
HLA-DRβ1*0824 2.9
HLA-DRβ1*0834 2.9
69.5%
5.6%
Global: 18.8%
Aona-DRβ1*0302GB 72.7
Aona-DRβ1*0301GB 18.2
Aona-DRβ1*0326GB 9.1
HLA-DRβ1*07/Aotus-DRβ*W30
Allele prototype
PPF
n=5
Global: 22.4%
HLA-DRβ1*0701 50.0
HLA-DRβ1*0704 7.7
HLA-DRβ1*0703 3.8
HLA-DRβ1*0706 3.8
HLA-DRβ1*0724 3.8
69.1%
15.5%
Global: 20.0%
Aona-DRβ*W3002 50.0
Aona-DRβ*W3001 50.0
Aovo-DRβ*W3001 100.0

Figure 2.

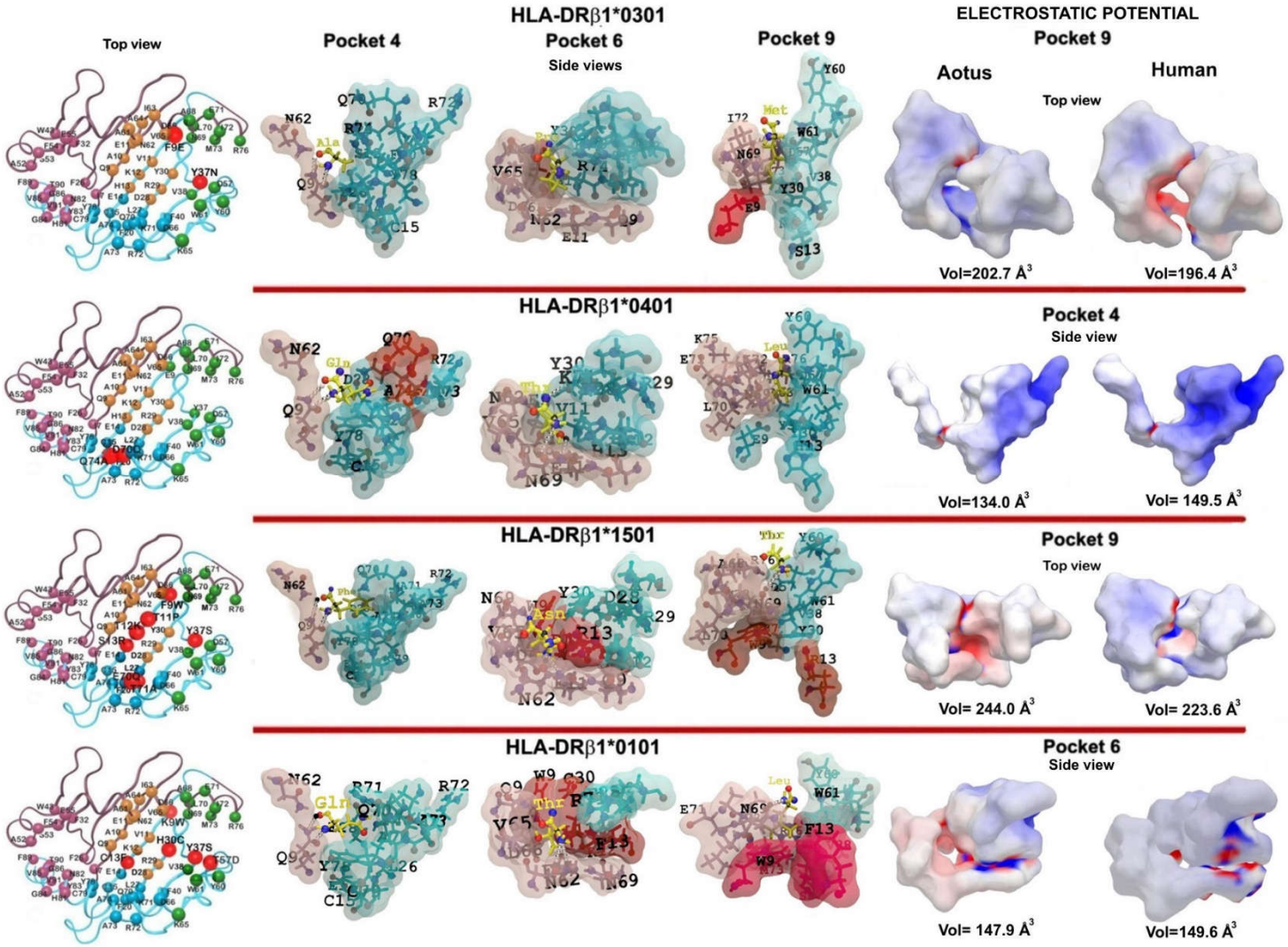
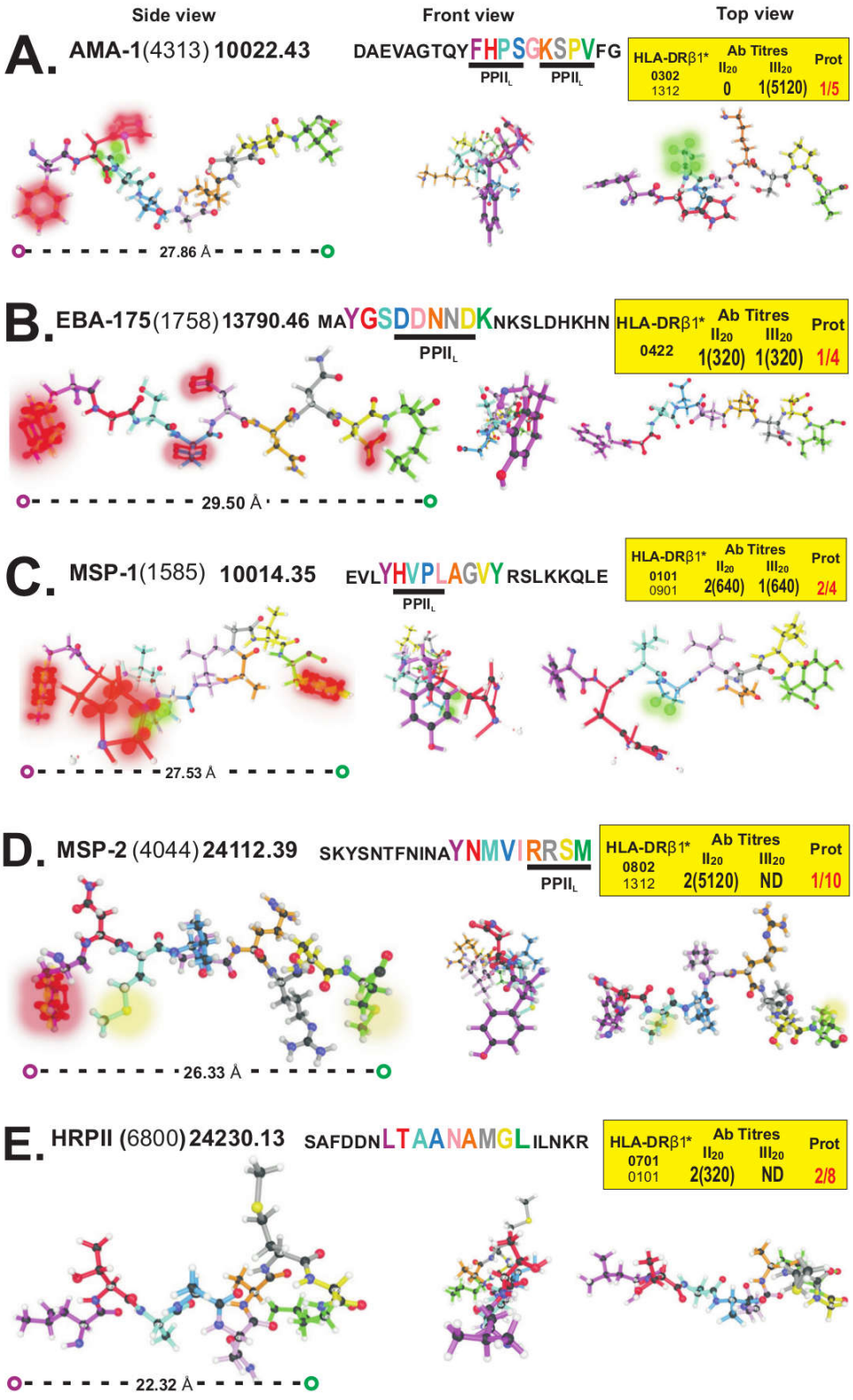

HLA-DRβ1*0301
Top view
Pocket 4
Pocket 6
Side views
Pocket 9
ELECTROSTATIC POTENTIAL
Pocket 9
Aotus
Human
Top view
Vol=202.7 Å3
Vol=196.4 Å3
HLA-DRβ1*0401
Pocket 4
Side view
Vol=134.0 Å3
Vol= 149.5 Å3
HLA-DRβ1*1501
Pocket 9
Top view
Vol= 244.0 Å3
Vol= 223.6 Å3
HLA-DRβ1*0101
Pocket 6
Side view
Vol= 147.9 Å3
Vol=149.6 Å3

Figure 3.

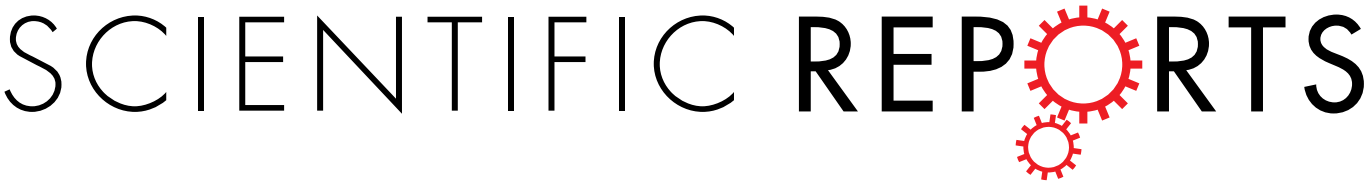

Side view
Front view
Top view
A. AMA-1(4313) 10022.43
DAEVAGTQYFHPSGKSPVFG
PPII$_L$
PPII$_L$
HLA-DRβ1* | Ab Titres II$_{20}$ | Ab Titres III$_{20}$ | Prot
0302 | 0 | 1(5120) | 1/5
1312
27.86 Å
B. EBA-175(1758)13790.46 MAYGSDDNNDKNKSLDHKHN
PPII$_L$
HLA-DRβ1* | Ab Titres II$_{20}$ | Ab Titres III$_{20}$ | Prot
0422 | 1(320) | 1(320) | 1/4
29.50 Å
C. MSP-1(1585) 10014.35 EVLYHVPLAGVYRSLKKQLE
PPII$_L$
HLA-DRβ1* | Ab Titres II$_{20}$ | Ab Titres III$_{20}$ | Prot
0101 | 2(640) | 1(640) | 2/4
0901
27.53 Å
D. MSP-2 (4044) 24112.39 SKYSNTFNINAYNMVIRRSM
PPII$_L$
HLA-DRβ1* | Ab Titres II$_{20}$ | Ab Titres III$_{20}$ | Prot
0802 | 2(5120) | ND | 1/10
1312
26.33 Å
E. HRPII (6800) 24230.13 SAFDDNLTAANAMGLILNKR
HLA-DRβ1* | Ab Titres II$_{20}$ | Ab Titres III$_{20}$ | Prot
0701 | 2(320) | ND | 2/8
0101
22.32 Å

Capítulo 4. Mass & secondary structure propensity of amino acids explain their mutability and evolutionary replacements

Bohórquez HJ, Suárez CF, Patarroyo ME. Mass & secondary structure propensity of amino acids explain their mutability and evolutionary replacements. Scientific Reports. 2017;7(1):7717.

La versión publicada del artículo puede ser consultada en:

https://www.nature.com/articles/s41598-017-08041-7



OPEN

Mass & secondary structure propensity of amino acids explain their mutability and evolutionary replacements

Received: 30 January 2017
Accepted: 28 June 2017
Published online: 10 August 2017

Hugo J. Bohórquez[1], Carlos F. Suárez[1,2,3] & Manuel E. Patarroyo[1,4]

Why is an amino acid replacement in a protein accepted during evolution? The answer given by bioinformatics relies on the frequency of change of each amino acid by another one and the propensity of each to remain unchanged. We propose that these replacement rules are recoverable from the secondary structural trends of amino acids. A distance measure between high-resolution Ramachandran distributions reveals that structurally similar residues coincide with those found in substitution matrices such as BLOSUM: Asn ↔ Asp, Phe ↔ Tyr, Lys ↔ Arg, Gln ↔ Glu, Ile ↔ Val, Met → Leu; with Ala, Cys, His, Gly, Ser, Pro, and Thr, as structurally idiosyncratic residues. We also found a high average correlation ($\overline{R} = 0.85$) between thirty amino acid mutability scales and the *mutational inertia* (I_X), which measures the energetic cost weighted by the number of observations at the most probable amino acid conformation. These results indicate that amino acid substitutions follow two optimally-efficient principles: (a) amino acids interchangeability privileges their secondary structural similarity, and (b) the amino acid mutability depends directly on its biosynthetic energy cost, and inversely with its frequency. These two principles are the underlying rules governing the observed amino acid substitutions.

In molecular evolution, protein stability is a solid indicator of function preservation thanks to a positive correlation between protein functionality and native stability[1,2]. Natural protein sequences evolved to avoid aggregation and increase functional diversity[3], and once a protein fold is established, the selection pressure at most positions in the protein will preserve fold stability. Homologous families of proteins have related functions, and structures are similar although sequences have diverged[4], even in regions with less than 30% sequence identity[5,6]. Accordingly, mutation events over time may replace a residue by another while keeping the backbone dihedral angles at that position unchanged[7]. These facts indicate that the amino acid sequence alone is an incomplete measure of evolutionary relationships between proteins. Indeed, structural similarities better reflect homology than sequence similarities[8]. Therefore, sequence variation around a conserved molecular architecture could be traced through amino acid substitution patterns fixed during protein evolution.

The intrinsic secondary structure propensities of amino acids are given by the statistics of Ramachandran distributions[9–11]. In this way, we could know the conformational bias of each amino acid towards specific secondary structures[12,13]. For instance, long polypeptide chains with the same backbone conformation are found exclusively in $\alpha - helix$, *PPII*, and β strands structures[14]. In general, examining the frequency of occurrence of particular amino acid residues in stable secondary structures have been useful for determining protein structure, folding, and energetics[15]. We propose that, in addition, the statistics of the secondary structure of proteins may reveal their evolutionary information.

To confirm this assumption, we explore a combination of extensive physical quantities with the statistics of Ramachandran distributions $P_X(\phi, \psi)$. In particular, we investigate the molecular mass as a measure of the amino acids biosynthetic cost. In addition, we use the protein geometry database (PGD 1.1)[16] for obtaining

[1]Bio-mathematics, Fundación Instituto de Inmunología de Colombia, FIDIC, Cra. 50 No. 26-00, Of. 102, Bogotá DC, 111321160 Cundinamarca, Colombia. [2]Universidad de Ciencias Aplicadas y Ambientales, UDCA, Bogotá DC, Colombia. [3]Universidad del Rosario, Bogotá DC, Colombia. [4]Universidad Nacional de Colombia, Bogotá DC, Colombia. **Hugo J. Bohórquez and Carlos F. Suárez contributed equally to this work**. Correspondence and requests for materials should be addressed to H.J.B. (email: hugo.j.bohorquez@fidic.org.co)

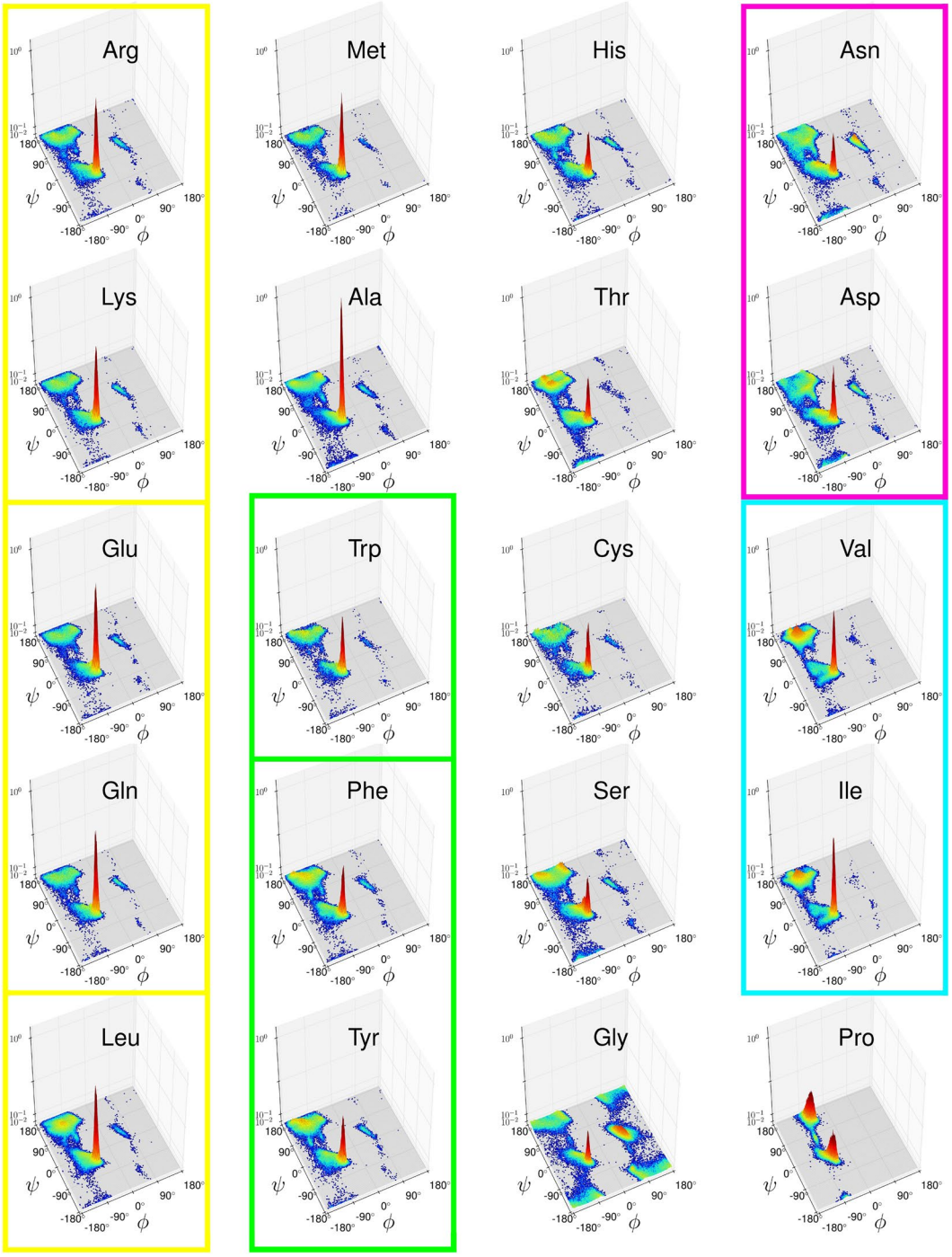


Figure 1. High-resolution Ramachandran probability distributions $P_X(\phi, \psi)$ (logarithmic scale) as derived from the PGD 1.1 database at 1.895° × 1.895° bin size. Structurally similar open sets: yellow, $S_I = \{\{\text{Arg, Lys}\}, \{\text{Glu, Gln}\}, \text{Leu}\}$; green, $S_{II} = \{\text{Trp, }\{\text{Phe, Tyr}\}\}$; magenta, $S_{III} = \{\text{Ans, Asp}\}$; cyan, $S_{IV} = \{\text{Val, Ile}\}$. Ala, Met, and Ser have their first neighbor in S_I; His, Thr, and Cys are adjacent to S_{II}. Larger images of each Ramachandran distribution are given by Supplementary Figs. S1–S20.

high-resolution Ramachandran distributions as 2D-binned probability histograms (Fig. 1). This choice has some practical advantages, including the possibility of directly applying distance measures between the distributions. The secondary structure distance between the amino acids (Fig. 2) is the main task in our research because the emerging close-distance pairs can be straightforwardly compared to pairwise mutations. The optimal bin area ($\Delta\phi\Delta\psi$) dividing the Ramachandran map is given by the method of Shimazaki & Shinomoto[17]. This is a key element in histogram binning because a very small bin size will result in noise amplification whereas a very large value will overpass important details of the distribution.

	Arg	Lys	Glu	Gln	Leu	Met	Ala	Trp	Phe	Tyr	His	Thr	Cys	Ser	Asn	Asp	Val	Ile	Gly	Pro
Arg	0.000	0.371	0.404	0.399	0.415	0.450	0.496	0.547	0.561	0.574	0.567	0.644	0.644	0.636	0.755	0.680	0.746	0.754	1.260	1.350
Lys	0.371	0.000	0.407	0.400	0.413	0.473	0.509	0.551	0.598	0.601	0.590	0.658	0.658	0.654	0.767	0.685	0.763	0.773	1.265	1.336
Glu	0.404	0.407	0.000	0.376	0.420	0.467	0.423	0.617	0.696	0.703	0.665	0.737	0.732	0.677	0.811	0.709	0.828	0.813	1.285	1.302
Gln	0.399	0.400	0.376	0.000	0.411	0.459	0.499	0.605	0.644	0.652	0.621	0.688	0.688	0.668	0.756	0.678	0.813	0.813	1.276	1.364
Leu	0.415	0.413	0.420	0.411	0.000	0.439	0.577	0.556	0.555	0.579	0.640	0.649	0.656	0.759	0.796	0.713	0.643	0.620	1.316	1.394
Met	0.450	0.473	0.467	0.459	0.439	0.000	0.570	0.590	0.621	0.632	0.642	0.701	0.700	0.714	0.833	0.761	0.765	0.769	1.312	1.397
Ala	0.496	0.509	0.423	0.499	0.577	0.570	0.000	0.686	0.775	0.792	0.731	0.843	0.799	0.649	0.877	0.768	0.961	0.940	1.260	1.231
Trp	0.547	0.551	0.617	0.605	0.556	0.590	0.686	0.000	0.502	0.527	0.597	0.674	0.651	0.679	0.825	0.785	0.744	0.769	1.295	1.360
Phe	0.561	0.598	0.696	0.644	0.555	0.621	0.775	0.502	0.000	0.355	0.523	0.578	0.601	0.684	0.802	0.815	0.666	0.697	1.288	1.493
Tyr	0.574	0.601	0.703	0.652	0.579	0.632	0.792	0.527	0.355	0.000	0.532	0.557	0.612	0.687	0.806	0.823	0.665	0.713	1.281	1.487
His	0.567	0.590	0.665	0.621	0.640	0.642	0.731	0.597	0.523	0.532	0.000	0.632	0.627	0.635	0.691	0.690	0.798	0.827	1.239	1.402
Thr	0.644	0.658	0.737	0.688	0.649	0.701	0.843	0.674	0.578	0.557	0.632	0.000	0.656	0.665	0.844	0.848	0.653	0.736	1.259	1.435
Cys	0.644	0.658	0.732	0.688	0.656	0.700	0.799	0.651	0.601	0.612	0.627	0.656	0.000	0.693	0.793	0.781	0.788	0.811	1.274	1.429
Ser	0.636	0.654	0.677	0.668	0.759	0.714	0.649	0.679	0.684	0.687	0.635	0.665	0.693	0.000	0.748	0.724	0.988	1.029	1.177	1.151
Asn	0.755	0.767	0.811	0.756	0.796	0.833	0.877	0.825	0.802	0.806	0.691	0.844	0.793	0.748	0.000	0.520	1.046	1.041	1.162	1.382
Asp	0.680	0.685	0.709	0.678	0.713	0.761	0.768	0.785	0.815	0.823	0.690	0.848	0.781	0.724	0.520	0.000	1.036	1.016	1.186	1.284
Val	0.746	0.763	0.828	0.813	0.643	0.765	0.961	0.744	0.666	0.665	0.798	0.653	0.788	0.988	1.046	1.036	0.000	0.292	1.420	1.621
Ile	0.754	0.773	0.813	0.813	0.620	0.769	0.940	0.769	0.697	0.713	0.827	0.736	0.811	1.029	1.041	1.016	0.292	0.000	1.427	1.625
Gly	1.260	1.265	1.285	1.276	1.316	1.312	1.260	1.295	1.288	1.281	1.239	1.259	1.274	1.177	1.162	1.186	1.420	1.427	0.000	1.557
Pro	1.350	1.336	1.302	1.364	1.394	1.397	1.231	1.360	1.493	1.487	1.402	1.435	1.429	1.151	1.382	1.284	1.621	1.625	1.557	0.000

Figure 2. Distance matrix ordered according to structurally similar amino acids. The smallest distance is represented in yellow, and the largest distance in blue, with intermediate values in green. Open subsets appear, consistently, in yellow. Additionally, Gly, and Pro appear as the most distant elements, followed by Asn, Val-Ile, Ala, and Thr.

We explore the twenty amino acid distributions through some of their distinctive features such as the most probable conformation, which is given by the highest peak of each distribution. Additionally, we propose a plausible mutability parameter that combines structural information with the molecular mass of the amino acids. Our results indicate that amino acid evolutionary substitutions occur by following two optimal-efficiency principles: (a) interchangeability between amino acids occurs by preserving secondary structural propensity, and (b) the mutability of an amino acid depends directly on its mass, and inversely with its frequency. The methodology introduced here gives the basis for developing a new kind of scoring matrices involving physical quantities and secondary structure statistics. Hopefully, these future efforts will further help to improve the peptide design strategies, which can contribute to close the gap between the primary sequence and the 3D structure of proteins.

Results and Discussion

High-resolution Ramachandran Probability Distributions. We distinguish two concepts regarding the backbone dihedral angles of proteins, as suggested by Dunbrack Jr. *et al.*[11]. The first is a *Ramachandran plot* or *Ramachandran map*, which is simply a scatter plot of the ϕ, ψ values for the amino acids in a single protein structure or a set of protein structures. It provides a simple view of the conformation of a protein. The second is a *Ramachandran probability distribution* $P(\phi, \psi)$ which is a statistical representation of Ramachandran data, usually in the form of a probability density function. $P_X(\phi, \psi)$ gives the probability of finding an amino acid conformation in a specific range of (ϕ, ψ) values.

We obtained non-parametric density estimates of $P_X(\phi, \psi)$ for each amino acid X from 1,153,791 residues retrieved from the high-resolution protein geometry database (PGD 1.1)[16]. In our approach—frequentist—events have a specific probability whose determination depends on the number of observations. Therefore each

Amino acid	M_X (Da)	B_X	Δ_X^{min}	$P_X^{\max}$ (%)	N_X	W_X	I_X
Ala	71.079	4	1.176°	0.437	113609	496.654	0.143
Arg	156.188	10	1.593°	0.265	45373	120.333	1.298
Asn	114.104	2	2.535°	0.156	46573	72.701	1.569
Asp	115.089	1	2.169°	0.192	56963	109.191	1.054
Cys	103.139	5	2.951°	0.173	15823	27.298	3.778
Gln	128.131	2	2.118°	0.307	35633	109.470	1.170
Glu	129.116	1	1.748°	0.321	48458	155.431	0.831
Gly	57.052	5	2.118°	0.124	98983	122.840	0.464
His	137.141	13	2.609°	0.173	27675	47.910	2.862
Ile	113.159	7	1.488°	0.285	74768	213.090	0.531
Leu	113.159	7	1.463°	0.276	116941	322.560	0.351
Lys	128.174	10	1.856°	0.276	40135	110.584	1.159
Met	131.193	7	1.782°	0.284	20968	59.610	2.201
Phe	147.177	11	2.169°	0.190	56511	107.242	1.372
Pro	97.117	4	2.222°	0.110	54555	60.167	1.614
Ser	87.078	4	1.978°	0.141	66612	93.593	0.930
Thr	101.105	6	2.069°	0.178	68557	121.726	0.831
Trp	186.213	14	2.687°	0.200	21118	42.340	4.398
Tyr	163.176	11	2.400°	0.184	48972	90.250	1.808
Val	99.133	4	1.622°	0.241	95564	230.082	0.431

Table 1. Properties of the Amino acids used in the present study. M_X is the residue average mass (without water). B_X gives Davis' biosynthetic steps[37]. $\Delta_X^{min}(deg)$ is the optimal bin angle determined by MISE method[17]. $P_X^{\max}$ corresponds to the peak of the Ramachandran distribution $P_X(\phi, \psi)$. N_X is the number of points used for determining $P_X(\phi, \psi)$. $W_X = P_X^{\max} \times N_X$ is an estimator of the maximum possible observations at the most frequent conformation. $I_X = M_X/W_X$ is the *mutational inertia*.

distribution $P_X(\phi, \psi)$ is given by a joint histogram. Such an approach depends on finding an optimal grid size, which can be determined with Shimazaki & Shinomoto method[17]. Said strategy requires a heuristic exhaustive sampling of a cost function whose minimum corresponds to an optimal binning of the distribution—see methods for details. Table 1 reports the optimal bin width for each Ramachandran probability distribution, Δ_X^{min}. The weighted average of these optimal bin widths gave us the bin size used (1.895°) in the present study. Thus, we obtained a grid with a total of 190 × 190 bins (36,100), each one covering an area of 1.895° × 1.895° of the dihedral space (Fig. 1), which is a significant improvement on the resolution of Ramachandran distributions previously reported.

For comparison, the 3D representation of the Ramachandran distributions for the first version of PGD uses a grid of 20.0° × 20.0° (i.e. a total of 324 bins), from a dataset containing 72,376 residues[10]. In another approach, the predicted protein backbone torsion angles from NMR chemical shifts made by the TALOS+ program uses an identical bin size (20.0° × 20.0°)[18,19], other studies on folding trends uses a resolution of 10.0° × 10.0° (i.e 1,296 bins)[11]. An early report on detailed Ramachandran distributions used bin widths of 4.0° × 4.0° (i.e. 90 × 90 bins), involving 237,384 amino acids from 1,042 proteins[20]. Our distributions have a resolution 4.5 times higher, which translates into a higher accuracy in the distance computations between the set of distributions $P_X(\phi, \psi)$. This high resolution was possible thanks to the fact that at least 84% of the structures reported at the protein data bank (PDB) were obtained during the last decade alone, most of which have atomic resolution.

Figure 1 reports the 3D plots of the twenty Ramachandran distributions determined for the present study; the dihedral angles are given in degrees, while the percentage probability per bin is given on a logarithmic scale. All the plots have the same height to facilitate their comparison. Larger plots are included in Supplementary Figs. S1–S20. While most distributions look similar one to another, there are some key differences. The probability distribution of glycine is very symmetrical and occupies all the allowed regions of the Ramachandran map. It is the only residue having a maximum at the left-handed α-helix conformation with a peak almost as high as the one at the α-helix region; these features are a consequence of its lack of a side chain[21]. On the other hand, proline—an imino acid—has two highly-populated states, with a slightly higher probability at the PPII conformation than at the α-helix conformation. It belongs to the set of structurally restricted amino acids composed by {Ile, Pro, Thr, Val}, which have an extremely low probability of occupying the right-hand side of the Ramachandran map. Indeed, the corresponding plots (Fig. 1) show few points within the quadrants I and IV ($\phi > 0$). The conformational restrictions of proline arise from its pyrrolidine ring, whose flexibility is coupled to the backbone[22]. Isoleucine, threonine, and valine are the only amino acids with C-β branching, which means that they have more bulkiness near to the protein backbone than the rest of amino acids[23]. They also have a local maximum within the β-sheet region—shown as red shaded peaks in Fig. 1—a feature only shared with the three aromatic residues, Phe, Tyr, Trp, and Leu. The remaining amino acids occupy the allowed regions in a generic fashion[20,24], whose distributions agree with the original Ramachandran and co-workers explanation in terms of steric clashes[25].

All these observations point to the qualitative aspects of the distributions. However, a systematic comparison of the twenty Ramachandran distributions requires the use of a quantitative evaluation of their similarities. In the

following subsection, we show a distance matrix accounting for dissimilarities between the secondary-structural trends of amino acids.

Secondary-structural vs BLOSUM replacements. A quantitative assessment of the similarities between the twenty distributions $P_X(\phi, \psi)$ requires a distance measure. We used the *city-block* distance, which can be used to assess the differences in discrete frequency distributions. It gives more weight to the most probable dihedral conformations of the Ramachandran distributions.

Each amino acids X has a set of twenty distances, D_X, including with itself, (in which case $||P_X - P_X|| = 0$):

$$D_X = \{||P_X - P_{Ala}||, ||P_X - P_{Arg}||, \ldots, ||P_X - P_{Tyr}||, ||P_X - P_{Val}||\} \quad (1)$$

The most plausible secondary-structural replacement to X is that amino acid Y having the smallest positive distance to X, or the minimum positive value from the set of distances: $min_+\{D_X\}$. That $\min_+\{D_X\} = ||P_X - P_Y||$ does not imply necessarily that $\min_+\{D_Y\} = ||P_Y - P_X||$. In other words, the structural replacement is not always a reciprocal operation; hence if Y is the replacement of X, we denote this by $X \rightarrow Y$. In the case of a reciprocal replacement, we denote it by $X \leftrightarrow Y$.

The secondary-structural distance matrix between the amino acids is shown in Fig. 2. The proximity between amino acids is given by a color scheme: the smallest distance is represented in yellow, and the largest distance in blue, with intermediate values in green. We found *open subsets* by a nearest-neighbor criterion: any element within an open subset has exactly the remaining elements of said subset as its nearest neighbors—the procedure is explained in the methods section. For instance, the simplest open subset is composed by two elements for which the other one is the closest element—i.e. those elements for which $D_{min}(P_X, P_Y) = D_{min}(P_Y, P_X)$ or, equivalently, $X \leftrightarrow Y$.

We found the following open sets (Fig. 3): a five-member set including a couple of two-member subsets: $S_I = \{\{\text{Arg, Lys}\}, \{\text{Glu, Gln}\}, \text{Leu}\}$—in yellow; a three-member set containing a two-member set, $S_{II} = \{\text{Trp}, \{\text{Phe, Tyr}\}\}$—in green; and a pair of two-member sets: $S_{III} = \{\text{Val, Ile}\}$, and $S_{IV} = \{\text{Asn, Asp}\}$—in cyan and magenta, respectively. Within this topology, Met appears as a boundary element of the first set S_I; Fig. 3 shows that Met first five neighbors are exactly the elements of S_I. In turn, every residue in S_I has Met as the fifth neighbor but Glu, which has Ala closer; this proximity may result from Ala and Glu being the strongest α-helix formers, as their respective $P_X^{\max}$ values indicate (Table 1). The S_I group includes aliphatic saturated side chains, while S_{II} contains the aromatic residues. Adjacent to these two major sets we found residues sharing their physiochemical characteristics—as shown by their close distances to the main groups in the distance matrix (Fig. 2). Specifically, four residues have their nearest neighbor within a major open set: Ala have its first neighbor in S_I, whereas His, Thr, and Cys have their first neighbor in S_{II}. Those amino acids outside an open set or its boundaries were considered structurally idiosyncratic: Ala, Cys, His, Gly, Ser, Pro, and Thr. Gly and Pro are the farthest ones from any other residue, as the last column of Fig. 3 shows. Certainly, these amino acids populate the Ramachandran map in a unique way. The Ramachandran distribution of glycine is widespread over the allowed regions; while Pro is the most structurally restricted. Alanine has twice the probability of forming an α-helix ($P_{Ala}^{\max} = 0.437\%$ from Table 1) than any other residue ($P_{aver \neq Ala}^{\max} = 0.214\%$). The Ramachandran distribution of Thr has four peaks around the β and π regions unlike any other residue, including the C-β branched amino acids (Fig. 1). While Thr is chemically similar to Ser[26], they have different structural propensities. According to our distance matrix (Fig. 2), Thr is closer to Tyr & Phe, while Ser is closer to His & Arg. A recent study shows that the phosphorylation of Ser increases its propensity of forming PPII, whereas that of Thr has the opposite effect[27]. This result indicates that Ser and Thr are far from being ideal secondary structural replacements. In summary, our classification reflects the intrinsic structural trends of amino acids; in particular, the S_I set and its adjacent elements Met and Ala are the same alpha formers found by Fujiwara *et. al.*[28]. Within the same scale, the aromatic set, S_{II}, and its adjacent elements (Cis, Thr) and S_{III} are beta formers. The remaining amino acids are turn/bend formers, including S_{IV} and Gly, Ser, and Pro, most of which have the lowest P_X^{max} values in Table 1.

More importantly, nevertheless, is the fact that an unexpected pattern emerged: our structurally similar pairs of amino acids matches with most BLOSUM matrices pair replacements[29], which are shown as shadowed boxes in Fig. 3. More details about the substitution matrices are in the methods section. Our list of structural replacements is: Asn $\leftrightarrow$ Asp, Phe $\leftrightarrow$ Tyr, Lys $\leftrightarrow$ Arg, Gln $\leftrightarrow$ Glu, Ile $\leftrightarrow$ Val, Met $\rightarrow$ Leu. In BLOSUM matrices, Thr and Ser are replacements. For all BLOSUM matrices, Gly, Pro, Cys, His, and Ala are idiosyncratic residues. In general, our set of structurally-similar amino acids coincide with most canonical residue substitutions given by scoring matrices such as BLOSUM62 and BLOSUM100[29], and consensus replacements[30]. This is a remarkable finding considering the extremely low probability of randomly finding six out of seven replacement pairs: less than one in a 681 million, as detailed in the methods section. In consequence, our result reveals an underlying correlation between mutation matrices and structural propensities. Hence, the replacement rules implied by the secondary structure distance (Fig. 2) may be directly used for for exploring structural amino acid replacements in peptide design strategies.

We conclude that during evolution, mutational replacements occurred between structurally similar amino acids. Hence, mutations followed a process that privileges structure and hence preserves function. But BLOSUM and PAM substitution matrices give additional information about the mutational trends of amino acids. The diagonal of these matrices determine how easy is for an amino acid to be replaced. A large value means more resistance to change. However, our distance matrix (Fig. 2) has a diagonal of zeros. For studying the mutability, we explored a parameter that combines the statistical information at the $P_X^{\max}$ with a basic extensive property.

Molecular mass and optimum evolutionary cost. Molecular mass is a fundamental extensive property that might have played a central role in defining the actual protein landscape. Previously, our group revealed a

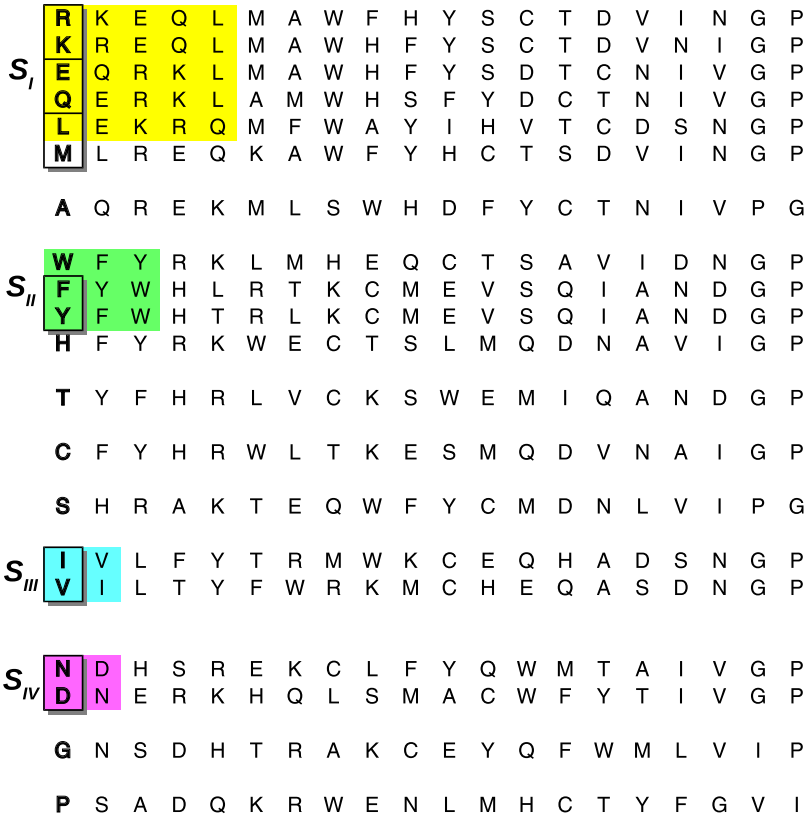


Figure 3. Rows ordered according to the cityblock distance. Open sets are indicated by the same color code used in Fig. 1. The shadowed boxes contain the BLOSUM100 pair replacements. The procedure for determining an open set consists on finding rows with the same set of first neighbors. For instance, the first neighbor of Arg (top row) is Lys; after placing the Lys row under the top row, we see that they share the seven first neighbors (up to Trp). The third row corresponds to Arg second neighbor, i.e. Glu, which also shares the same first neighbors with the previous ones up to Trp. The fourth row corresponds to Arg third neighbor, i.e. Gln, whose fifth neighbour is Ala, unlike the previous rows. The fifth row corresponds to Arg fourth neighbor, i.e. Leu, which has all the previous rows as its first neighbors. In this way, the yellow box includes those elements whose first four neighbors are completely contained within the set. Methionine is a frontier element of this set: its first five neighbors are exactly the elements of the whole closed set; however, Glu does not include Met within its first five neighbours and for that reason Met is not contained in the set. The remaining open sets S_{II} to S_{IV} were obtained in the same way. Notice that Pro and Gly are the farthest residues from any other one, as a consequence of their structural propensity uniqueness.

very high correlation ($R = 0.98$) between mass and the electronic energy of amino acids—excluding the two sulfur-containing side chains[31]. In the present study, we found a complex relationship between the amino acids mass M_X and the structural trends via the probability at the most frequent conformational state, $P_X^{\max}$; this quantity is given by the highest peak of each Ramachandran distribution—$\max(P_X(\phi, \psi))$. $P_X^{\max}$ corresponds to the most frequent conformation and, therefore, it is an indicator of structural persistence[32].

The α-helix conformation is the highest peak for all amino acids (but proline) with alanine at the top as the strongest helix former. While mass has an overall poor correlation with $P_X^{\max}$ ($R = 0.05$), we identified two main and opposite trends delimited by separate ranges of $P_X^{\max}$: (a) $P_X^{\max} > 0.200\%$ defines the set of strong helix formers {Ala, Glu, Gln, Ile, Met, Leu, Lys, Arg, Val} (in descending order), with a negative correlation $R = -0.61$; and, (b) $P_X^{\max} \leq 0.200\%$ defines the weak helix formers: {Trp, Asp, Phe, Tyr, Thr, His, Cys, Asn, Ser, Gly, Pro}, with a positive correlation of $R = 0.76$. The small set of C-β branched amino acids ({Ile, Thr, Val}) plus proline shows a correlation of $R = 0.78$ between mass and $P_X^{\max}$. After excluding these four elements from the two main sets, their respective correlations rise to $R = -0.87$ for the strong helix formers, and to $R = 0.87$ for the set of weak helix formers. In strong helix formers, the negative correlation between $P_X^{\max}$ and the molecular mass indicates that light side chains have a better chance of forming an alpha helix than heavy ones. These three correlations reveal a direct involvement of the molecular mass on the α-helical propensities of the amino acids.

A recent observation by Lehmann *et. al.* reports a negative correlation between the background frequency and codon degeneracy of amino acids with mass[33]. Seligmann already observed that the evolutionary rate of amino acid replacements correlates negatively with mass[34]. Accordingly, heavier amino acids are less frequent, which suggests that the genomes preserve a fundamental distribution ruled by simple energetics. Inverse correlations between the average amino acid biosynthetic cost and the levels of gene expression are consistent with natural

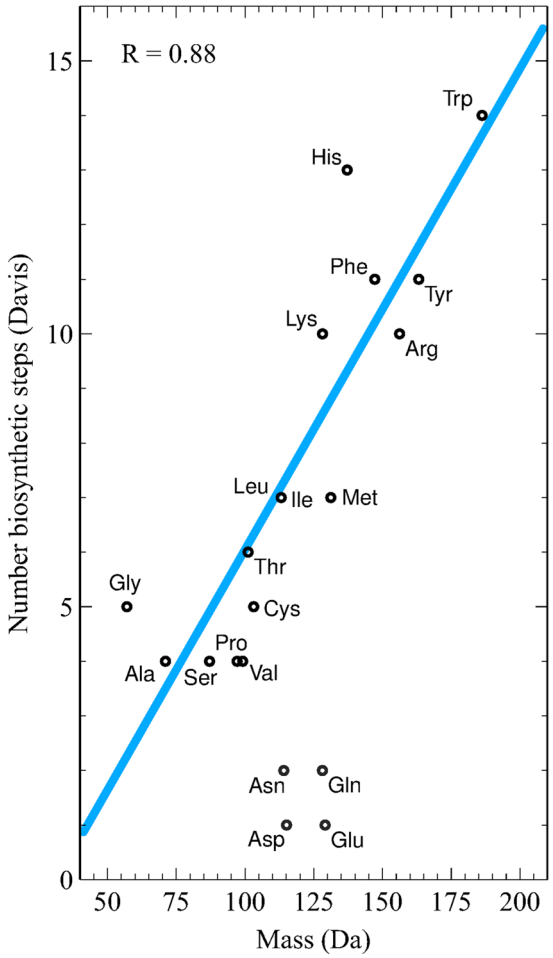


Figure 4. Correlation between the molecular mass of the amino acids M_X and their energetic cost as accounted by the number of biosynthetic steps B_X proposed by Davis[37]. The outliers {Asn, Asp, Gln, Glu} are excluded from the Pearson's correlation and from the linear interpolation.

selection to minimize costs[35]. Seligmann also shows a positive correlation ($R = 0.80$) between the molecular mass M_X and the total energetic cost per amino acid (in ATPs)[34], as reported by Akashi & Gojobori[36]. According to Lehmann *et al.*, highly expressed proteins tend to use amino acids with relatively low synthetic costs[33]. Therefore, heavy amino acids are less frequent because they are biosynthetically more expensive. We found a further confirmation of this statement: the molecular mass grows with the number of biosynthetic steps, as shown in Fig. 4. The values proposed by Davis[37], are included in Table 1 as B_X. The number of biosynthetic steps has been proposed as a natural way of determining the evolutionary history of amino acids[38], and so does the amino acids molecular mass. We found a correlation of $R = 0.64$ between mass and biosynthetic steps, which rises up to $R = 0.88$ after excluding the set of outliers {Asn, Asp, Gln, Glu} (Fig. 4).

In summary, we found a high correlation—by parts—between the molecular mass and the probability at the most frequent conformational state ($P_X^{\max}$). We also found a high correlation between mass and the number of biosynthetic steps (B_X). These correlations are consistent with the fact that evolution privileges energetically optimal costs[34,39]. Thus, in the quest for a physical quantity that can explain amino acid's mutability, mass is irreplaceable as a fundamental measure of energetic cost.

Mass over the frequency at the most probable conformation correlates with mutability. The background frequency or natural abundance of amino acids, N_X, may be indicative of their evolutionary age: more abundance reflects an early adoption in molecular evolution[40]. The values of N_X were obtained from the PGD 1.1 database (Table 1). The quantity $W_X = P_X^{\max} \times N_X$ is an estimator of the maximum observations at the most frequent conformation. In this way, W_X combines the probability at the most probable conformation with the background frequency. In the previous section we showed that an amino acid has less probability to be changed if it is more energetically expensive, and therefore mass directly measures the resistance to be changed. Additionally, less frequent amino acids are also less replaceable, indicating an inverse correlation with the mutability. Under these considerations, we define a "replacement inertia" as the mass M_X weighted by W_X: $I_X = M_X/W_X$. It summarizes the energetic cost per number of observations at the most probable conformation. We hypothesize that I_X might reflect the mutability of amino acids—i.e. the diagonal of substitution matrices (see more details in the Methods).

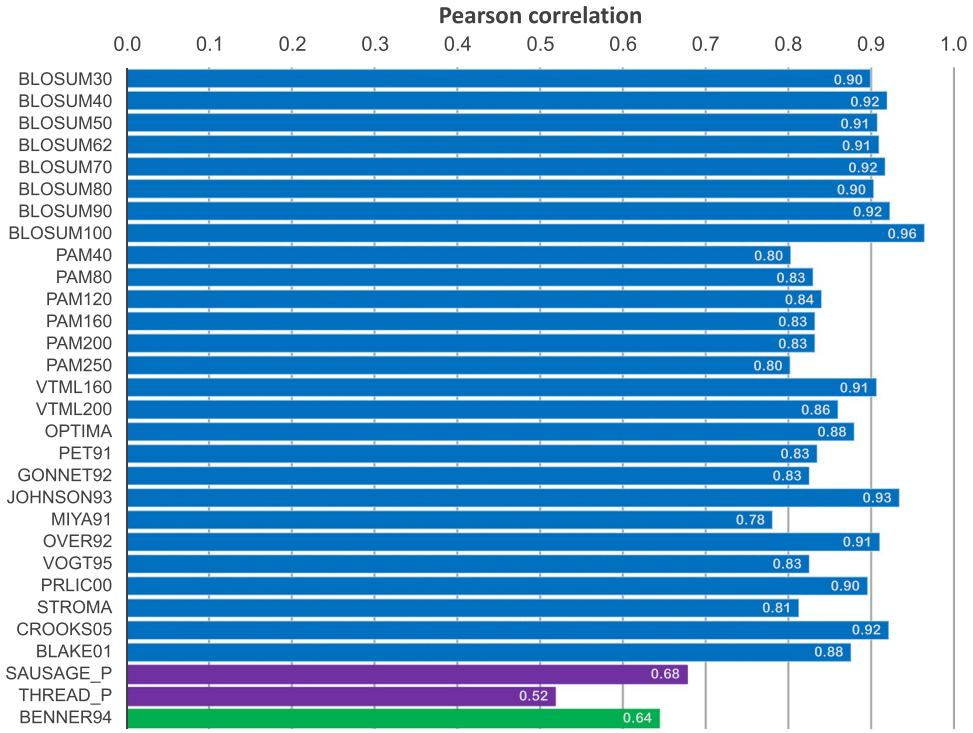


Figure 5. Pearson correlation coefficients between the replacement inertia I_X (Table 1) and the mutability of thirty replacement matrices. Alignment derived matrices are shown in blue, force field derived matrices in purple, and the genetic code derived matrix in green. See Supplementary Table S1 for the abbreviations.

In order to test if I_X reflects the mutability of amino acids, we selected thirty replacement matrices reported by the AAindex[41]: twenty-seven that were built from sequence alignments—including a selection of six PAM and eight BLOSUM matrices; two more that were crafted from force fields (THREADER and SAUSAGE)[42]; and a last one that was obtained from replacements at the genetic code level[43]. Supplementary Table S1 contains the list of matrices used in our survey. We computed the Pearson correlation coefficient between I_X and each mutability, which is shown in Fig. 5; in this figure, the correlation with alignment-derived matrices is colored in blue; the correlation with force-field derived appears in purple; and the correlation with the genetic code based matrix is plotted in green.

We found a very strong average correlation between I_X and the whole mutability set of $\overline{R}_{30} = 0.85$. This average value can be explained by the strong correlation found between I_X and the mutability of matrices derived from sequence alignments, which have values $R > 0.78$, as Fig. 5 shows. For the family of BLOSUM matrices, R values were obtained between 0.90 and 0.96, with an average correlation of $\overline{R}_B = 0.92$. For PAM matrices, the correlation was lower with an average value of $\overline{R}_P = 0.82$ for the six PAM matrices included in our survey.

On the other hand, the correlation between I_X and the mutability of the THREADER substitution matrix was the lowest we found, $R_{THREADER} = 0.52$. The second lowest correlation for was with the matrix based on the genetic code ($R_{BENNER} = 0.64$). The other force field derived matrix gave a correlation of $R_{SAUSAGE} = 0.68$. These low correlations may have an interesting explanation: while force field based substitution matrices do not include evolutionary information, BENNER matrix, on the other hand, assumes that the genetic code is the only determinant of amino acid substitutions. As a consequence, the underlying factors controlling these matrices are poorly reflected on I_X. Therefore, we must conclude that the very high correlation between I_X and the mutability of matrices derived from sequence alignments implies that molecular mass, abundance, and the most probable secondary structure conformation may have played a decisive role on shaping the molecular evolution of proteins.

However, how significant an average correlation of $\overline{R} = 0.85$ between I_X and the mutability set is? We evaluated the correlation coefficients between the mutability of all the substitution matrices, which yields a total of 430 correlations for the thirty matrices considered. The average value for these correlations is $\overline{R}_{430} = 0.84$. This value differs little from $\overline{R}$, which means that I_X describes amino acids mutability as well as any the mutability of the accepted mutation matrices. The correlation matrix with significance levels for I_X and the mutability of the whole set of matrices is shown in Supplementary Fig. S1. An excerpt of this plot is shown in Fig. 6, which includes the following matrices: BLOSUM30, BLOSUM62, BLOSUM100, PAM40, PAM160, and PAM250. This plot reveals that the correlations between PAM and BLOSUM fall within 0.70 and 0.83. Expectedly, correlations between matrices of the same family are higher, up to 0.96 for BLOSUM and up to 0.97 for PAM. It is surprising that I_X had better simultaneous correlations with both matrix families than they have with each other. This observation holds for the eight BLOSUM and six PAM matrices included in our study, as shown in Supplementary Fig. S21.

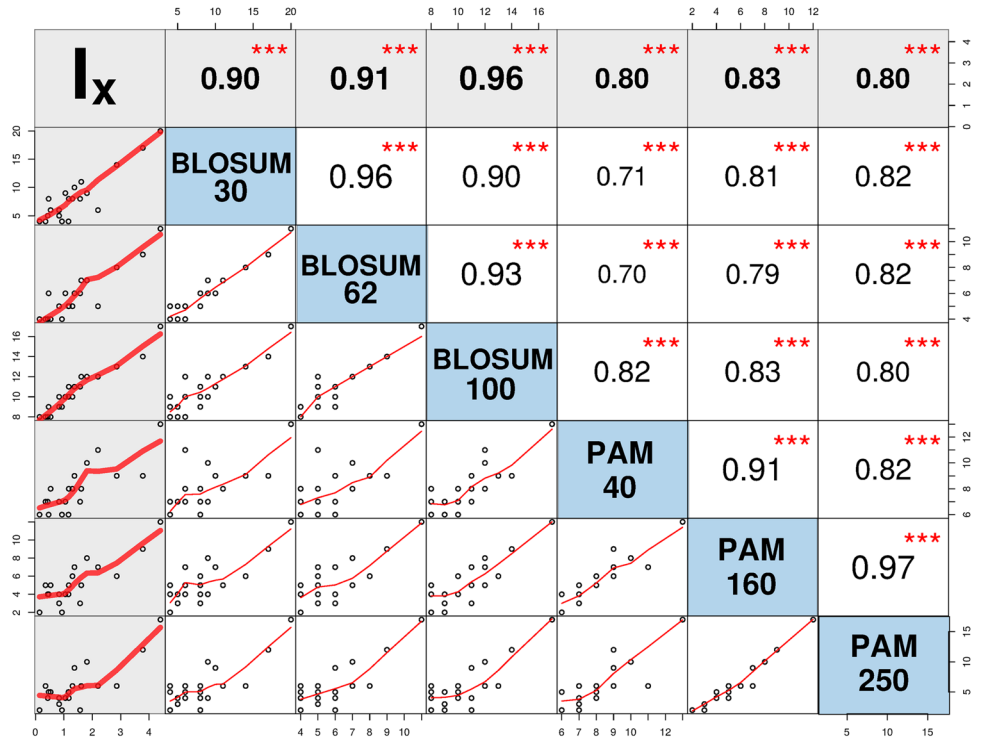


Figure 6. Correlation matrix plot with significance levels between the replacement inertia (I_X) and the mutability of a representative set of BLOSUM and PAM matrices. The lower triangular matrix is composed by the bivariate scatter plots with a fitted smooth line. The upper triangular matrix shows the Pearson correlation plus significance level (as stars). Each significance level is associated to a symbol: p-values 0.001 (***), 0.01 (**), 0.05 (*). This plot was generated with the Performance Analytics package in R program[57]. The correlation matrix for the complete mutability set is plotted in Supplementary Fig. S1.

Our results indicate that amino acids mutability may be an evolutionary invariant that depends on the biosynthetic cost per amino acid and on the background frequency. These observations might have relevant consequences for future developments and improvements of the actual scoring matrices, as well on structure prediction and design.

Conclusions

Our study provides compelling evidence about the physiochemical nature of the substitution matrices. Taylor's early work[44] on *evolutionary biochemistry*[45] proposes an integrative amino acid classification schema based on Dayhoff's PAM matrix and properties such as volume and polarity. In a complementary way, our approach puts the evolutionary concepts closer to physiochemical properties, which might be helpful for treating proteins as integrated physical and historical wholes.

The main findings of the present work agree with accepted ideas about the molecular evolution of proteins. In the first place, we claim that secondary structural similarities resemble to a great extent the canonical replacements given by substitution matrices (Figs 2 and 3). We interpret this result as a manifestation of an underlying structural preservation principle according to which amino acids interchangeability is highly determined by their secondary structural similarity. It might be a consequence of the fact that less structurally important parts of a protein evolve faster than more important ones. In this way, conservative substitutions occur more frequently in evolution than more disruptive ones. Our result agrees with Koonin & Wolf view according to which the primary causes of protein evolution could have more to do with fundamental principles of protein folding than with unique biological functions[46]. In the second place, we showed that amino acids mutability is correlated with the *replacement inertia* I_X (Fig. 5). Therefore, amino acids mutability depends on the biosynthetic cost, the most probable conformation, and the background frequency. Davis proposes that the timeline of genetically encoded amino acids correlates with the number of chemical reactions required to synthesize each amino acid[37,38,47]. As a consequence, the correlation between mass and biosynthetic steps (Fig. 4) indicates that the mutability of amino acids might be a timeline of protein evolution as well.

Undeniably, the biosynthetic cost, structural preservation, and frequency distribution of amino acids, all played a significant role in the molecular evolution of proteins. Indeed, two main selective factors determining the evolution of proteins are structural robustness against misfolding, and energy-cost efficiency[46,48,49]. Protein synthesis is very error-prone in comparison to DNA replication, and hence many folding-recognition mechanisms seem to have evolved to minimize costs of erroneous protein synthesis[49]. This energy-cost efficiency may explain why highly expressed proteins evolve slowly and at rates largely unrelated to their functions[48].

We can summarize our two main findings in similar terms with the following optimal-efficiency principles: (a) amino acids interchangeability occurs by preserving the secondary structural propensity, and (b) the amino acid mutability depends directly on its biosynthetic energy cost, and inversely with its frequency at the most probable conformation. We believe that these two principles are the underlying rules governing the observed amino acid substitutions. They provide a unified interpretation to mutation matrices, outside the statistical realm alone. Our results also indicate that amino acids mutability might be an invariant scale that differs little from one substitution matrix to another (Supplementary Fig. S21). These results may offer a new understanding of the evolutionary processes determining the structure of proteins.

Finally, the statistical similarities between secondary structural propensities used here offer a viable methodology for systematically exploring amino acid structural replacements. For instance, one can determine a structural distance matrix limited to the β-strand region, which may differ from the one of the whole Ramachandran map. With this type of sectoral statistics one can envision new rules for the design of polypeptide chains.

Methods

Data source. We calculated the Ramachandran distributions from the protein geometry database PGD 1.1, retrieved in June 2016[16]. We selected crystallized protein geometries with resolution equal or less than 2Å, a R-factor equals to 0.2, and a R-free maximum of 0.3. In order to avoid over-representation bias of some protein families, we used 7,398 proteins with a maximum identity of 25%. A total of 1,153,791 residues were considered.

Data analysis. The statistical analysis of the present work was implemented in Python 2.7 programming language[50,51]. A Python routine extracts the observed (ϕ, ψ) values from the PGD database for each amino acid (PGDread.py). The 2D optimization process was done with a routine that computes the cost function by changing the bin width equally for both dihedral variables $\Delta = \Delta\phi = \Delta\psi$, (MISE.py). The Ramachandran distribution histograms were computed and plotted with Matplotlib libraries (3DRamadistr.py)[52]. The cityblock distance was taken from the SCIPY package. A total of 600 code lines were written for the complete analysis shown here. The Python codes are available upon request.

Histogram optimization. Histograms are a type of non-parametric density estimates for which the number of parameters equals the number of data points[53]. A different approach uses analytic functions for obtaining smooth distributions that minimize low resolution and outliers effects[54]. The discrete (histogram) representation of the joint probability distribution $P_X(\phi_i, \psi_j)$ depends on the bin width of the dihedral variables, i.e. $\Delta\phi$ and $\Delta\psi$. A coarse binning size decreases the data noise but it might overpass relevant details of the structural information. On the other hand, a very fine grain bin size might highlight underlying statistical noise. The mean integrated squared error (MISE) can be estimated from the data through a cost function $C(\Delta)$. A histogram with the bin size that minimizes the MISE is optimal[17]. This method guarantees that a substantial increasing in the observations will further increase the accuracy of the histogram representation of probability distributions even more. The main assumption underlying this method is that the distribution can be represented by a smooth continuum function. Previous works have proven that Ramachandran distributions obey such assumption[11]. We assumed a regular partitioning of the Ramachandran maps i.e having the same bin size Δ for both dihedral variables: $\Delta = \Delta\phi = \Delta\psi$. The cost function for two variables is therefore given by

$$C(\Delta) = \frac{2n - v}{\Delta^4} \tag{2}$$

where the mean n and the variance v of the number of occurrences are given, respectively, by $n = \frac{1}{N}\sum_i^N n_i$, and $v = \frac{1}{N}\sum_i^N (n_i - n)^2$. The obtained optimal bin value for each amino acid is Δ_X (Table 1). We used the weighted average as the bin with for all the Ramachandran distributions: $\overline{\Delta} = \sum_X^{20} N_X \Delta_X / \sum_X^{20} N_X$. From the obtained Δ_X values, $\overline{\Delta} = 1.887°$, which was approximated by the integer fraction $360°/190 \simeq 1.895°$, i.e. we used 190 bins in each angular coordinate, for a total of $190 \times 190 = 36{,}100$.

Amino acid classification. We classified the amino acids according to the city-block (Manhattan) distance. Our grouping method takes advantage of the fact that a metric induces a topology on a set. Accordingly, we determined the topology induced by the city-block distance over the set of amino acids. The increasing distance between a given element X and the remaining ones determines an ordered list. Therefore, for the present case, we have twenty ordered lists, one for each amino acid. The intersection between the first neighbors of these lists gave us *open subsets*. An open subset consists on those elements such that, for every element within the subset, its neighbors belong to the same subset. Figure 3 reports the twenty ordered lists with an example about how to obtain open sets.

Substitution matrices and mutability. The most common method of evaluating the amino acid substitution patterns is through substitution matrices such as PAM[55] or BLOSUM[29]. A typical substitution matrix has 20×20 elements, in which non-diagonal pairwise scores (log odds) represent the probability of one amino acid could be substituted by other in protein evolution. The diagonal scores of the matrix are estimators of amino acid mutability. For each amino acid, a greater score implies lesser possibilities to be substituted, on the other hand, lesser scores implies a greater chance to be substituted[55,56]. We used a set of thirty substitution matrices reported in the AAindex[41] and NCBI (ftp://ftp.ncbi.nih.gov/blast/matrices/).

Probability of randomly finding six out of seven sets. Substitution matrices, such as BLOSUM62 & BLOSUM100, define seven replacement pairs of amino acids. Our structural similar pairs do coincide with six of them. We need an assessment of the probability for correctly obtaining six out of seven pairs. The probability of

obtaining the first element of a pair is the number of elements of such pair (2) divided by the total of elements (14). Then, the probability of finding the match is the number of pair elements still in the set (1) divided by the total left (13). Hence, the combined probability of randomly finding the first pair out of seven is $P_1 = 2/14 \times 1/13$. By a similar reasoning, the probability of obtaining a second pair is $P_2 = 2/12 \times 1/11$, and so on. Therefore, the probability of simultaneously finding six out of seven pairs is $\prod_{i=1}^{6} P_i$, or equivalently, $\prod_{k=2}^{7} \frac{2}{2k(2k-1)} = 1/681{,}080{,}400 = 1.468 \times 10^{-9}$. In other words, there is a chance of one in 681 million of simultaneously obtaining six correct pairs from a set of seven pairs.

References

1. Sikosek, T. & Chan, H. S. Biophysics of protein evolution and evolutionary protein biophysics. *Journal of The Royal Society Interface* **11**, 20140419 (2014).
2. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences* **103**, 5869–5874 (2006).
3. Yu, J.-F. *et al*. Natural protein sequences are more intrinsically disordered than random sequences. *Cellular and Molecular Life Sciences* **15**, 2949–2957 (2016).
4. Worth, C. L., Gong, S. & Blundell, T. L. Structural and functional constraints in the evolution of protein families. *Nature Reviews Molecular Cell Biology* **10**, 709–720 (2009).
5. Levy, E. D., Erba, E. B., Robinson, C. V. & Teichmann, S. A. Assembly reflects evolution of protein complexes. *Nature* **453**, 1262–1265 (2008).
6. Yang, Y. *et al*. Sixty-five years of the long march in protein secondary structure prediction: the final stretch? *Briefings in Bioinformatics* bbw 129 (2016).
7. Orengo, C. A. & Thornton, J. M. Protein families and their evolutionâ€"a structural perspective. *Annu. Rev. Biochem.* **74**, 867–900 (2005).
8. Dokholyan, N. V. & Shakhnovich, E. I. Scale-free evolution. In *Power Laws, Scale-Free Networks and Genome Biology*, 86–105 (Springer, 2006).
9. Ramachandran, G. t. & Sasisekharan, V. Conformation of polypeptides and proteins. *Advances in protein chemistry* **23**, 283–437 (1968).
10. Hollingsworth, S. A. & Karplus, P. A. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular concepts* **1**, 271–283 (2010).
11. Ting, D. *et al*. Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS computational biology* **6**, e1000763 (2010).
12. Levitt, M. Conformational preferences of amino acids in globular proteins. *Biochemistry* **17**, 4277–4285 (1978).
13. Koehl, P. & Levitt, M. Structure-based conformational preferences of amino acids. *Proceedings of the National Academy of Sciences* **96**, 12524–12529 (1999).
14. Hollingsworth, S. A., Berkholz, D. S. & Karplus, P. A. On the occurrence of linear groups in proteins. *Protein Science* **18**, 1321–1325 (2009).
15. DeBartolo, J., Jha, A., Freed, K. F. & Sosnick, T. R. Local Backbone Preferences and Nearest-Neighbor Effects in the Unfolded and Native States. *Protein and Peptide Folding, Misfolding, and Non-Folding* 79–98 (2012).
16. Berkholz, D. S., Krenesky, P. B., Davidson, J. R. & Karplus, P. A. Protein Geometry Database: a flexible engine to explore backbone conformations and their relationships to covalent geometry. *Nucleic Acids Res.* **38**, D320–D325 (2010).
17. Shimazaki, H. & Shinomoto, S. A method for selecting the bin size of a time histogram. *Neural Computation* **19**, 1503–1527 (2007).
18. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. Talos+: a hybrid method for predicting protein backbone torsion angles from nmr chemical shifts. *Journal of biomolecular NMR* **44**, 213–223 (2009).
19. Shen, Y. & Bax, A. Protein structural information derived from nmr chemical shift with the neural network program talos-n. In *Artificial Neural Networks*, 17–32 (Springer, 2015).
20. Hovmöller, S., Zhou, T. & Ohlson, T. Conformations of amino acids in proteins. *Acta Crystallographica Section D: Biological Crystallography* **58**, 768–776 (2002).
21. Ho, B. K. & Brasseur, R. The ramachandran plots of glycine and pre-proline. *BMC structural biology* **5**, 14 (2005).
22. Ho, B. K., Coutsias, E. A., Seok, C. & Dill, K. A. The flexibility in the proline ring couples to the protein backbone. *Protein Science* **14**, 1011–1018 (2005).
23. Betts, M. J. & Russell, R. B. Amino acid properties and consequences of substitutions. *Bioinformatics for geneticists* **317**, 289 (2003).
24. Ho, B. K., Thomas, A. & Brasseur, R. Revisiting the ramachandran plot: Hard-sphere repulsion, electrostatics, and h-bonding in the α-helix. *Protein Science* **12**, 2508–2522 (2003).
25. Ramachandran, G. & Ramakrishnan, C. t. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of molecular biology* **7**, 95 (1963).
26. Bohórquez, H. J. *et al*. Electronic energy and multipolar moments characterize amino acid side chains into chemically related groups. *The Journal of Physical Chemistry A* **107**, 10090–10097 (2003).
27. Kim, S.-Y., Jung, Y., Hwang, G.-S., Han, H. & Cho, M. Phosphorylation alters backbone conformational preferences of serine and threonine peptides. *Proteins: Structure, Function, and Bioinformatics* **79**, 3155–3165 (2011).
28. Fujiwara, K., Toda, H. & Ikeguchi, M. Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type. *BMC structural biology* **12**, 18 (2012).
29. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences* **89**, 10915–10919 (1992).
30. Bordo, D. & Argos, P. Suggestions for "safe" residue substitutions in site-directed mutagenesis. *Journal of molecular biology* **217**, 721–729 (1991).
31. Bohórquez, H. J., Cárdenas, C., Matta, C. F., Boyd, R. J. & Patarroyo, M. E. Methods in biocomputational chemistry: a lesson from the amino acids. *Quantum Biochemistry* 403–421.
32. Chatterjee, P. & Sengupta, N. Effect of the a30p mutation on the structural dynamics of micelle-bound α synuclein released in water: a molecular dynamics study. *European Biophysics Journal* **41**, 483–489 (2012).
33. Lehmann, J., Libchaber, A. & Greenbaum, B. D. Fundamental amino acid mass distributions and entropy costs in proteomes. *Journal of Theoretical Biology* **410**, 119–124 (2016).
34. Seligmann, H. Cost-minimization of amino acid usage. *Journal of molecular evolution* **56**, 151–161 (2003).
35. Raiford, D. W. *et al*. Do amino acid biosynthetic costs constrain protein evolution in saccharomyces cerevisiae? *Journal of molecular evolution* **67**, 621–630 (2008).
36. Akashi, H. & Gojobori, T. Metabolic efficiency and amino acid composition in the proteomes of escherichia coli and bacillus subtilis. *Proceedings of the National Academy of Sciences* **99**, 3695–3700 (2002).
37. Davis, B. K. Evolution of the genetic code. *Progress in biophysics and molecular biology* **72**, 157–243 (1999).
38. Griffiths, G. Cell evolution and the problem of membrane topology. *Nature Reviews Molecular Cell Biology* **8**, 1018–1024 (2007).

39. Guilloux, A. & Jestin, J.-L. The genetic code and its optimization for kinetic energy conservation in polypeptide chains. *Biosystems* **109**, 141–144 (2012).
40. Brooks, D. J., Fresco, J. R., Lesk, A. M. & Singh, M. Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. *Molecular Biology and Evolution* **19**, 1645–1655 (2002).
41. Kawashima, S. & Kanehisa, M. Aaindex: amino acid index database. *Nucleic acids research* **28**, 374–374 (2000).
42. Dosztanyi, Z. & Torda, A. E. Amino acid similarity matrices based on force fields. *Bioinformatics* **17**, 686–699 (2001).
43. Benner, S., Cohen, M. A. & Gonnet, G. H. Amino acid substitution during functionally constrained divergent evolution of protein sequences. *Protein Engineering* **7**, 1323–1332 (1994).
44. Taylor, W. R. The classification of amino acid conservation. *Journal of theoretical Biology* **119**, 205–218 (1986).
45. Harms, M. J. & Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nature Reviews Genetics* **14**, 559–571 (2013).
46. Koonin, E. V. & Wolf, Y. I. Constraints, plasticity, and universal patterns in genome and phenome evolution. In *Evolutionary Biology–Concepts, Molecular and Morphological Evolution*, 19–47 (Springer, 2010).
47. Davis, B. K. Molecular evolution before the origin of species. *Progress in biophysics and molecular biology* **79**, 77–133 (2002).
48. Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O. & Arnold, F. H. Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14338–14343 (2005).
49. Drummond, D. A. & Wilke, C. O. The evolutionary consequences of erroneous protein synthesis. *Nature Reviews Genetics* **10**, 715–724 (2009).
50. van Rossum, G. & de Boer, J. Linking a stub generator (ail) to a prototyping language (python). In *Proceedings of the Spring 1991 EurOpen Conference, Troms, Norway*, 229–247 (1991).
51. Python Software Foundation. Python language reference. URL http://www.python.org.
52. Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering* **9**, 90–95 (2007).
53. Shapovalov, M. V. & L., D. J. R. Non-Parametric Statistical Analysis Of The Ramachandran Map. *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map* 76 (2013).
54. Lovell, S. C. *et al*. Structure validation by C α geometry: ϕ, ψ and C β deviation. *Proteins: Structure, Function, and Bioinformatics* **50**, 437–450 (2003).
55. Dayhoff, M. O. & Schwartz, R. M. A model of evolutionary change in proteins. In *In Atlas of protein sequence and structure* (Citeseer, 1978).
56. Valdar, W. S. Scoring residue conservation. *Proteins: Structure, Function, and Bioinformatics* **48**, 227–241 (2002).
57. Peterson, B. G. *et al*. Performanceanalytics: Econometric tools for performance and risk analysis. r package version 1.4. 3541 (2014).

Acknowledgements

We would like to thank Professor Mario Amzel for his insightful comments on the paper.

Author Contributions

C.F.S. and H.J.B. proposed the project and developed the methodology of the study. H.J.B. wrote the Python codes. C.F.S. and H.J.B. carried out computations. C.F.S. and H.J.B. analyzed the data. M.E.P. supervised the project. H.J.B. wrote the manuscript whose final version include contributions by all authors.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-08041-7

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary material
Mass & secondary structure propensity of amino acids explain their mutability and evolutionary replacements

Hugo J. Bohórquez[1,+,*], **Carlos F. Suárez**[1,2,3,+], **and Manuel E. Patarroyo**[1,4]

[1]Fundación Instituto de Inmunología de Colombia, FIDIC, Biomathematics, Cra. 50 No. 26-00, Bogotá D. C., Colombia
[2]Universidad de Ciencias Aplicadas y Ambientales, UDCA, Bogotá D. C., Colombia
[3]Universidad del Rosario, Bogotá D. C., Colombia
[4]Universidad Nacional de Colombia, Bogotá D. C., Colombia
[+]Hugo J. Bohórquez and Carlos F. Suárez contributed equally to this work.

ABSTRACT

We use the protein geometry database (PGD 1.1)[1] for obtaining the high-resolution Ramachandran distributions as 2D-binned probability histograms (Figures S1 to S20). The optimal bin area ($1.895^\circ \times 1.895^\circ$) dividing the Ramachandran map was obtained with the method of Shimazaki & Shinomoto.[2] Figure S21 shows the correlation matrix plot with significance levels between the replacement inertia I_X and the mutability of the full set of replacement matrices used in the present study (Table S1).

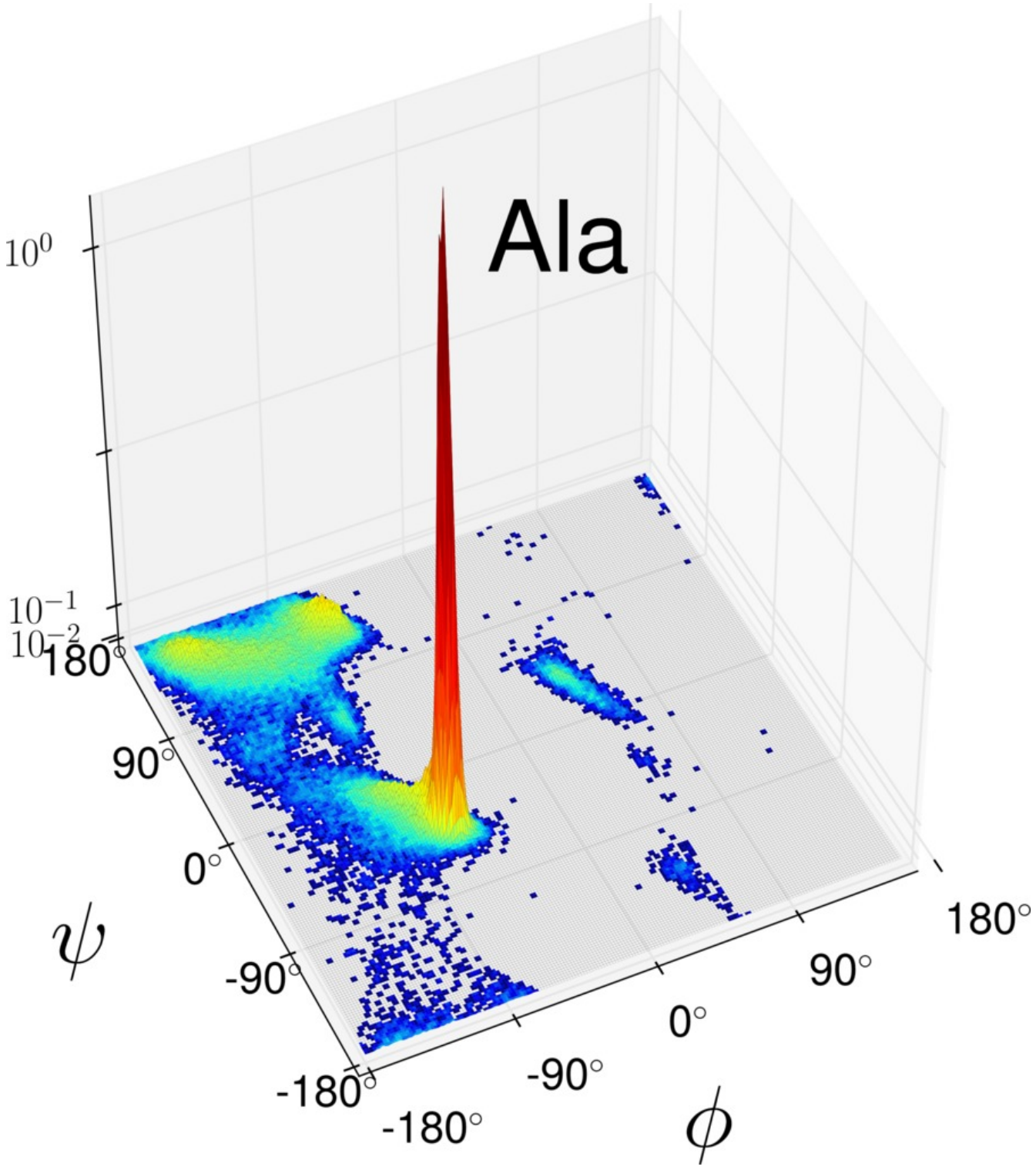


Figure S1. High-resolution Ramachandran distribution $P_{Ala}(\phi, \psi)$ of alanine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

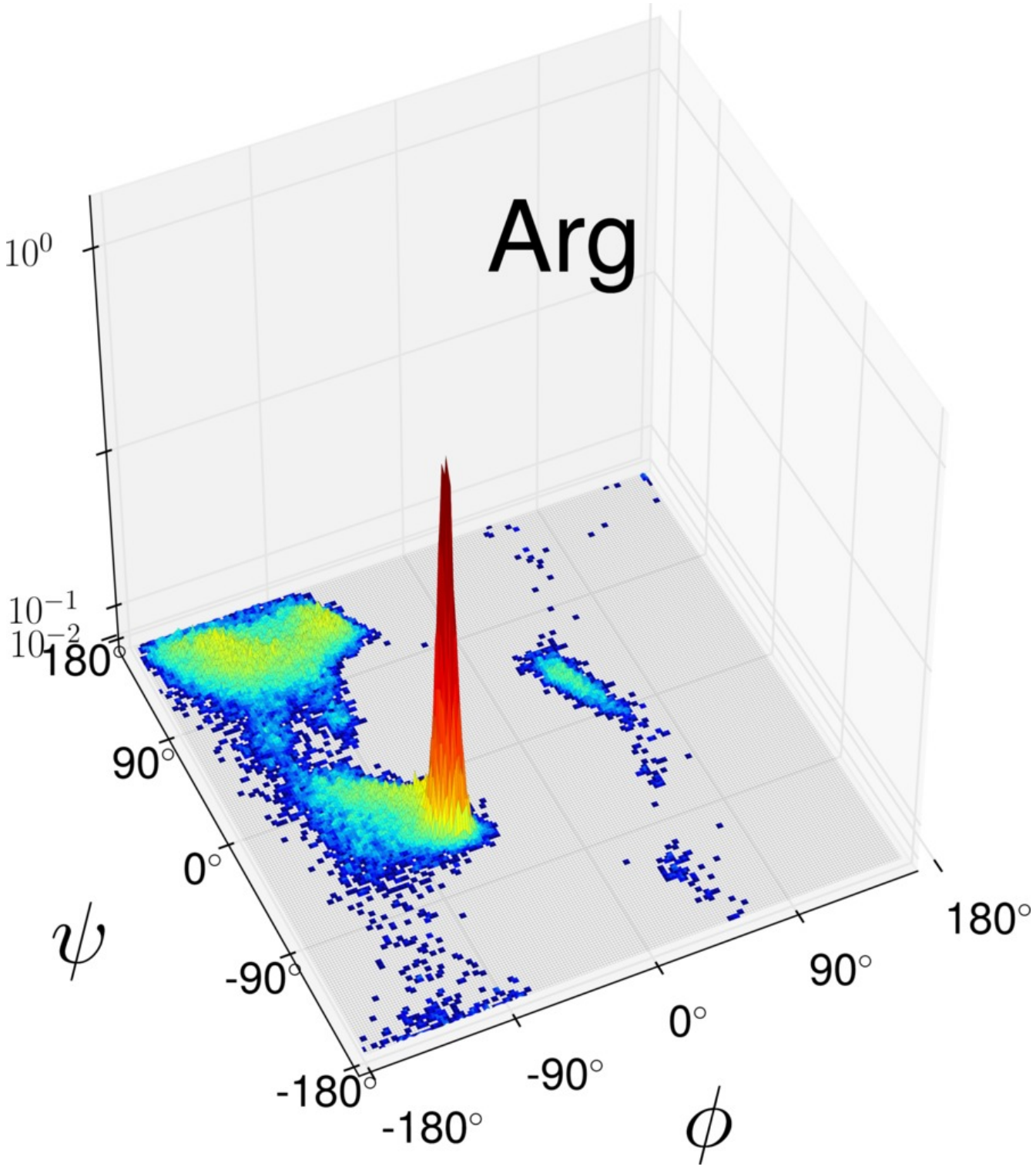


Figure S2. High-resolution Ramachandran distribution $P_{Arg}(\phi, \psi)$ of arginine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

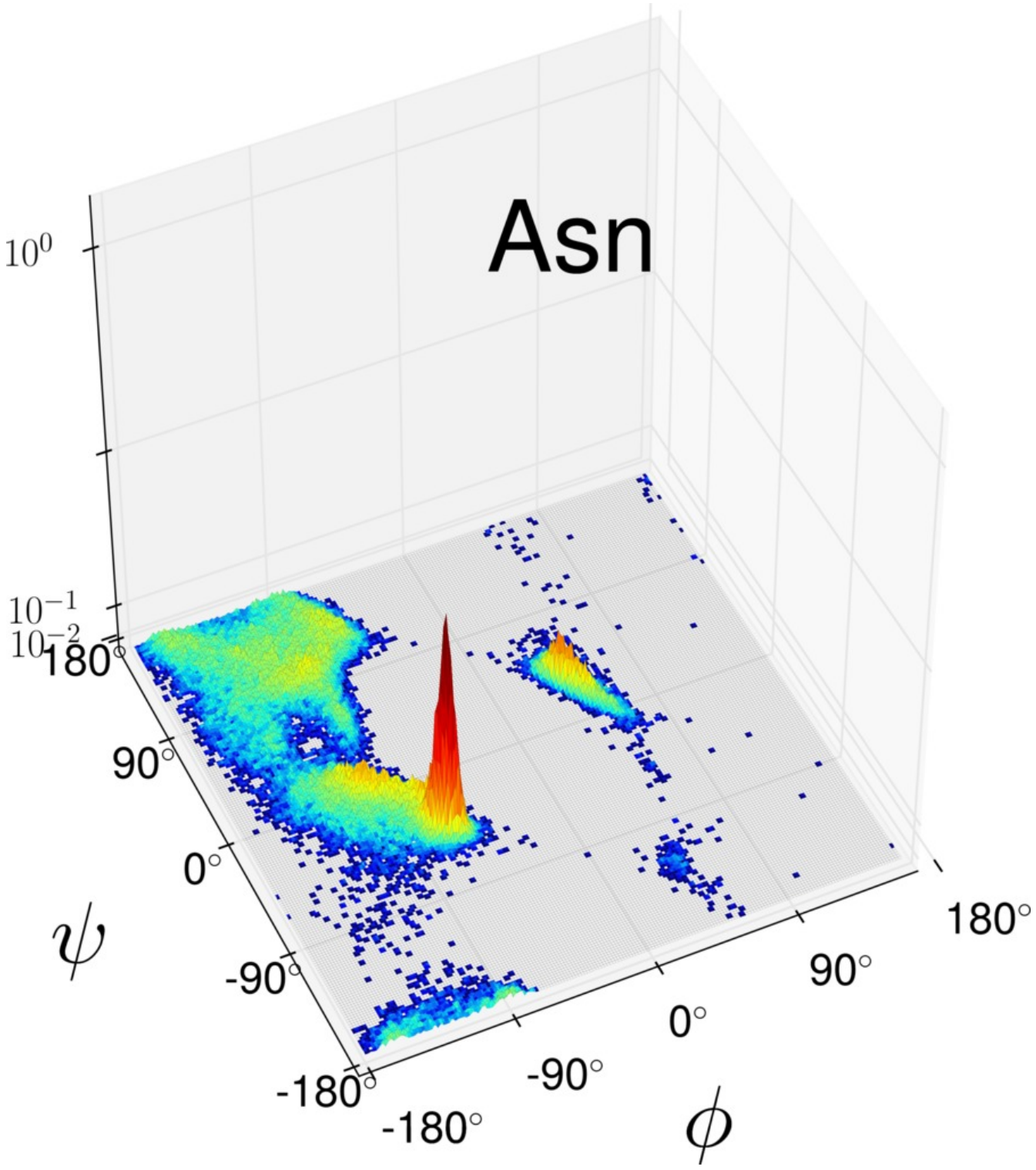


Figure S3. High-resolution Ramachandran distribution $P_{Asn}(\phi, \psi)$ of asparagine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

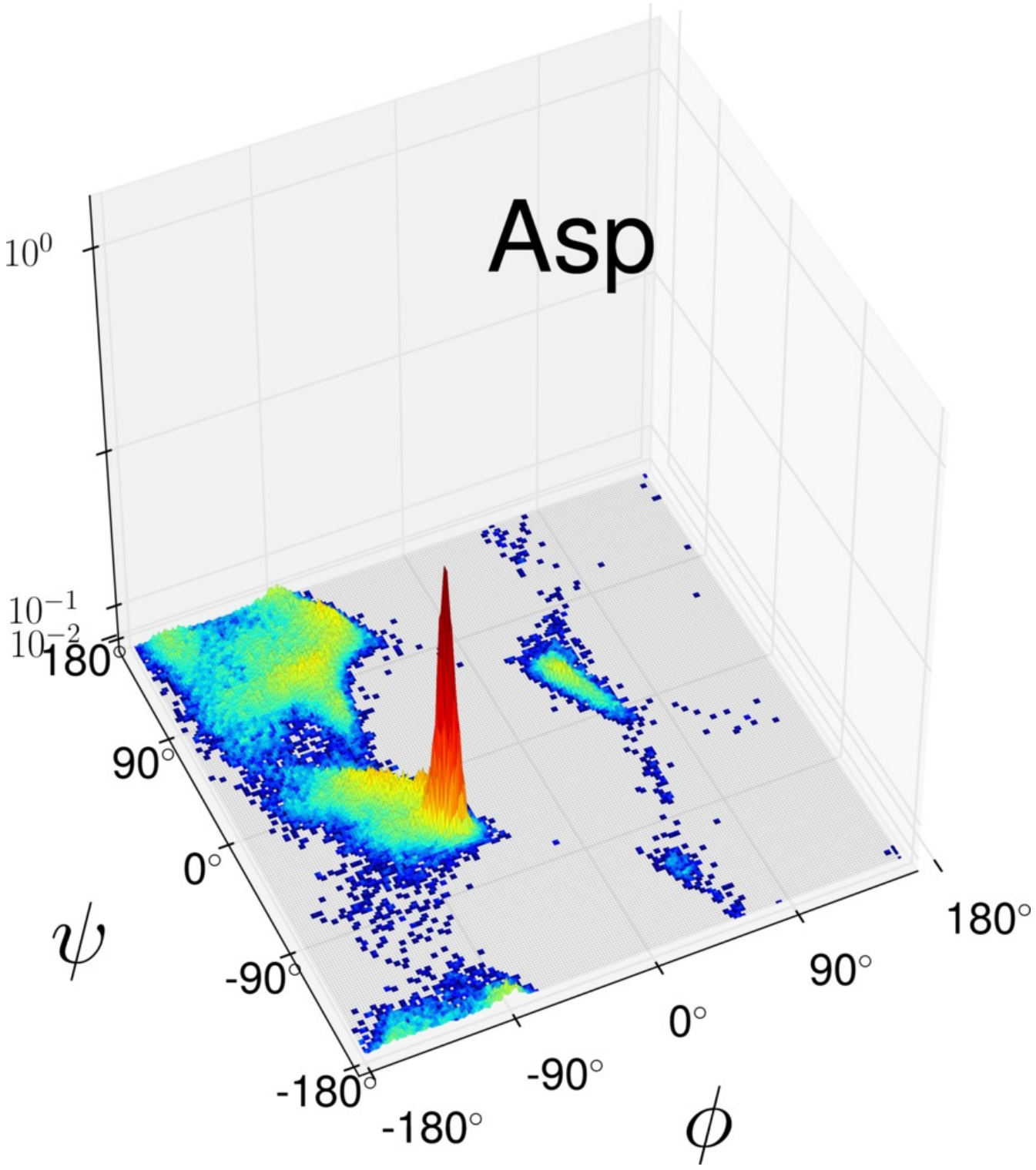


Figure S4. High-resolution Ramachandran distribution $P_{Asp}(\phi, \psi)$ of aspartic acid as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

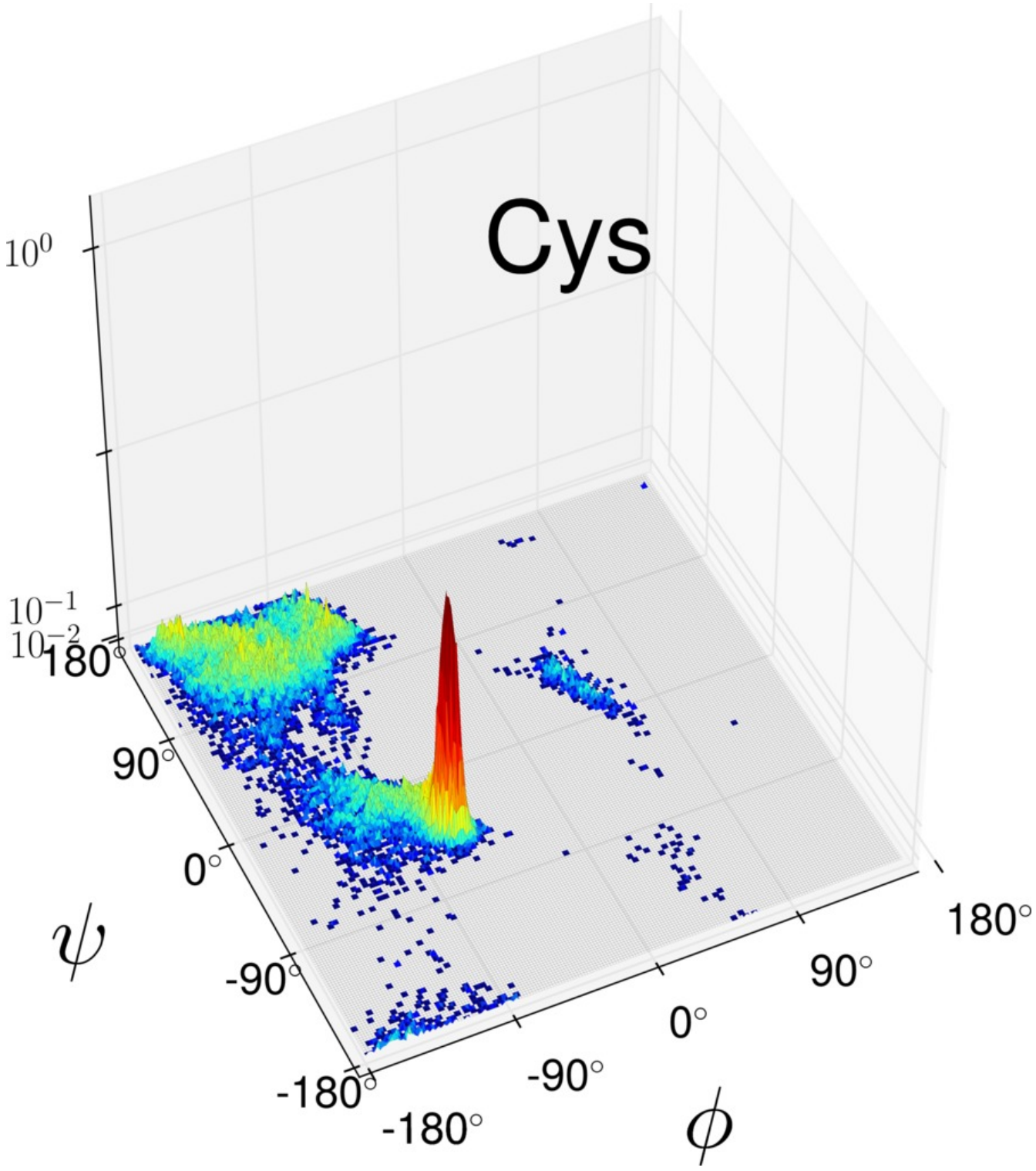


Figure S5. High-resolution Ramachandran distribution $P_{Cys}(\phi, \psi)$ of cysteine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

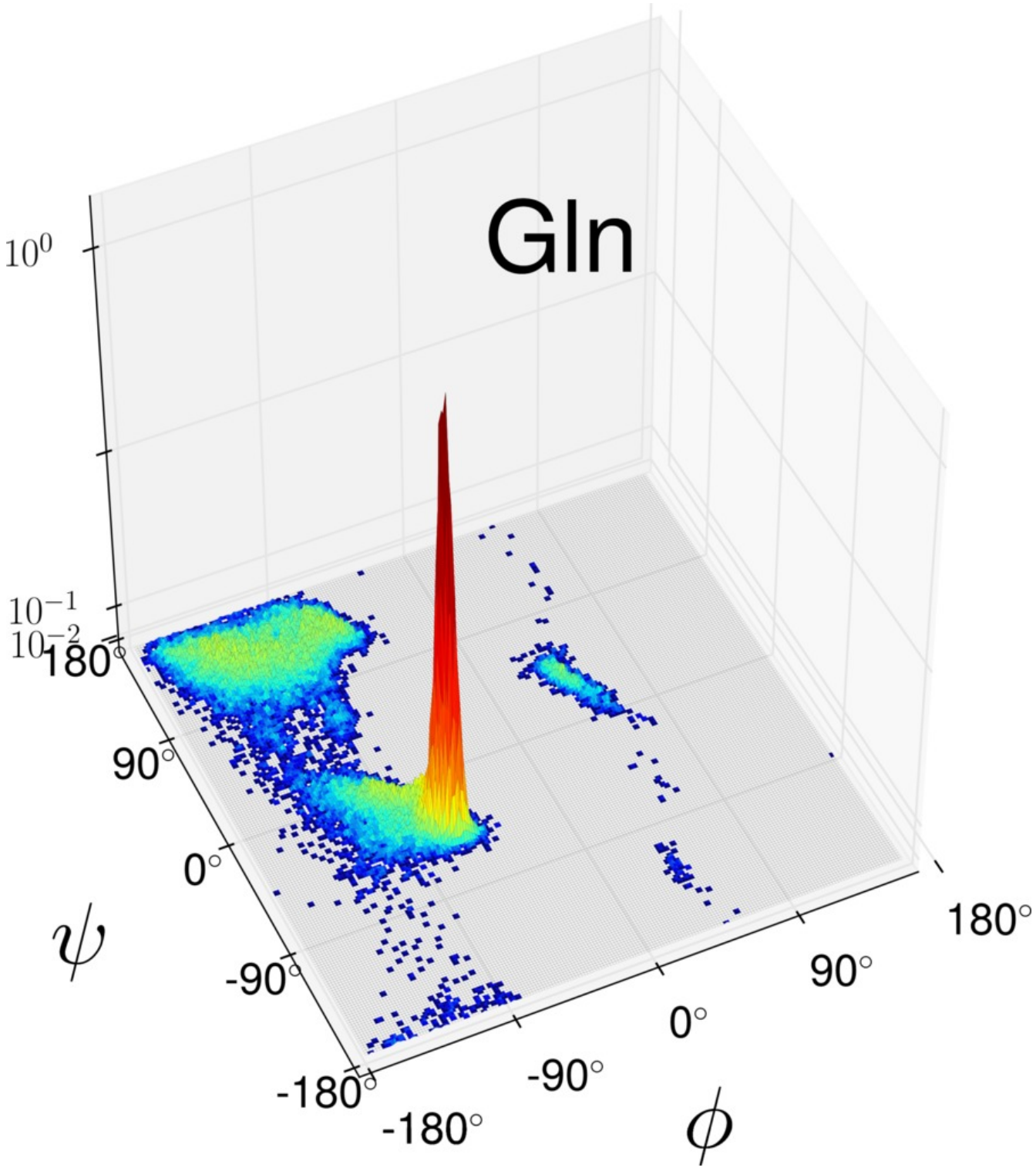


Figure S6. High-resolution Ramachandran distribution $P_{Gln}(\phi, \psi)$ of glutamine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

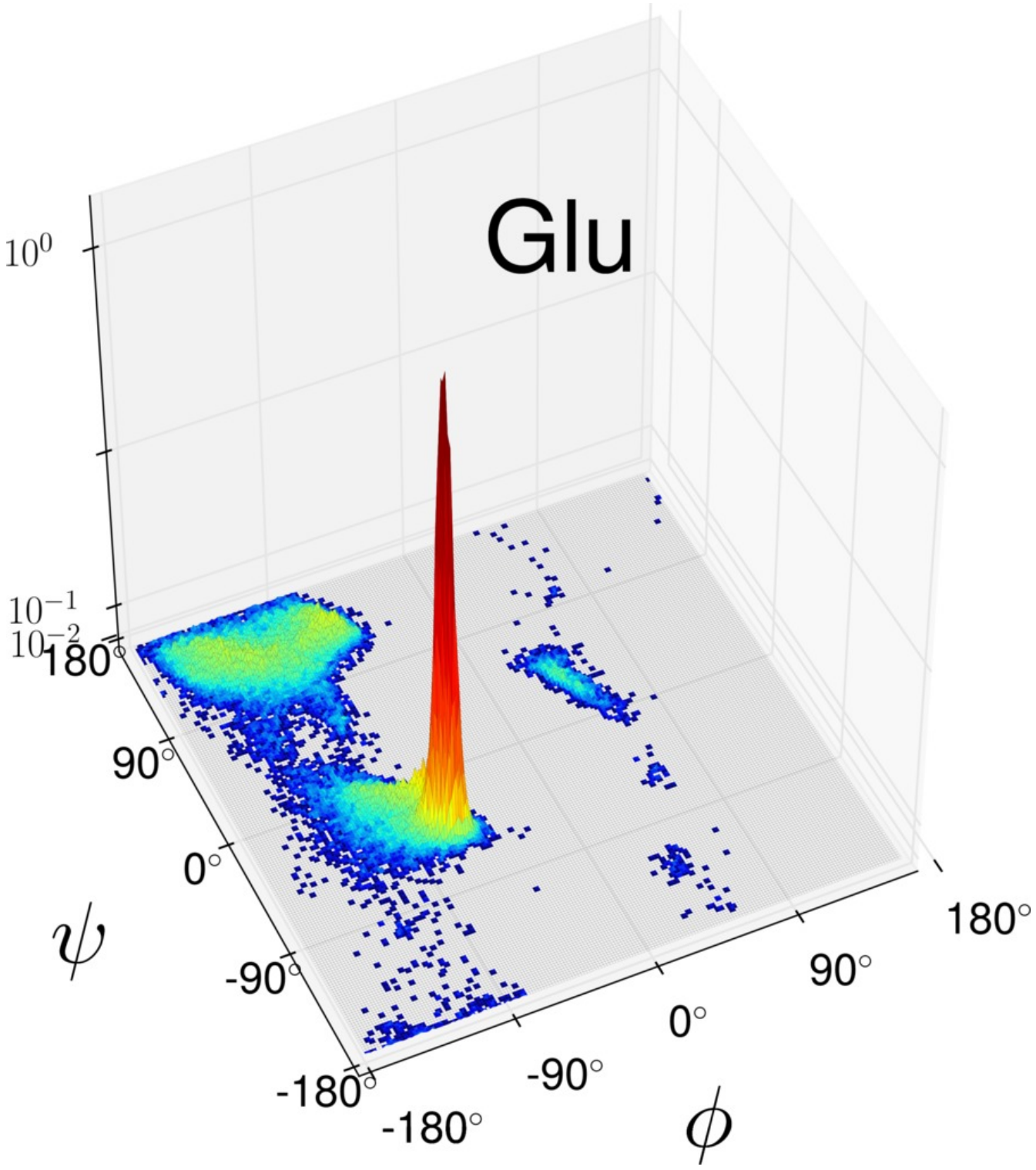


Figure S7. High-resolution Ramachandran distribution $P_{Glu}(\phi, \psi)$ of glutamic acid as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

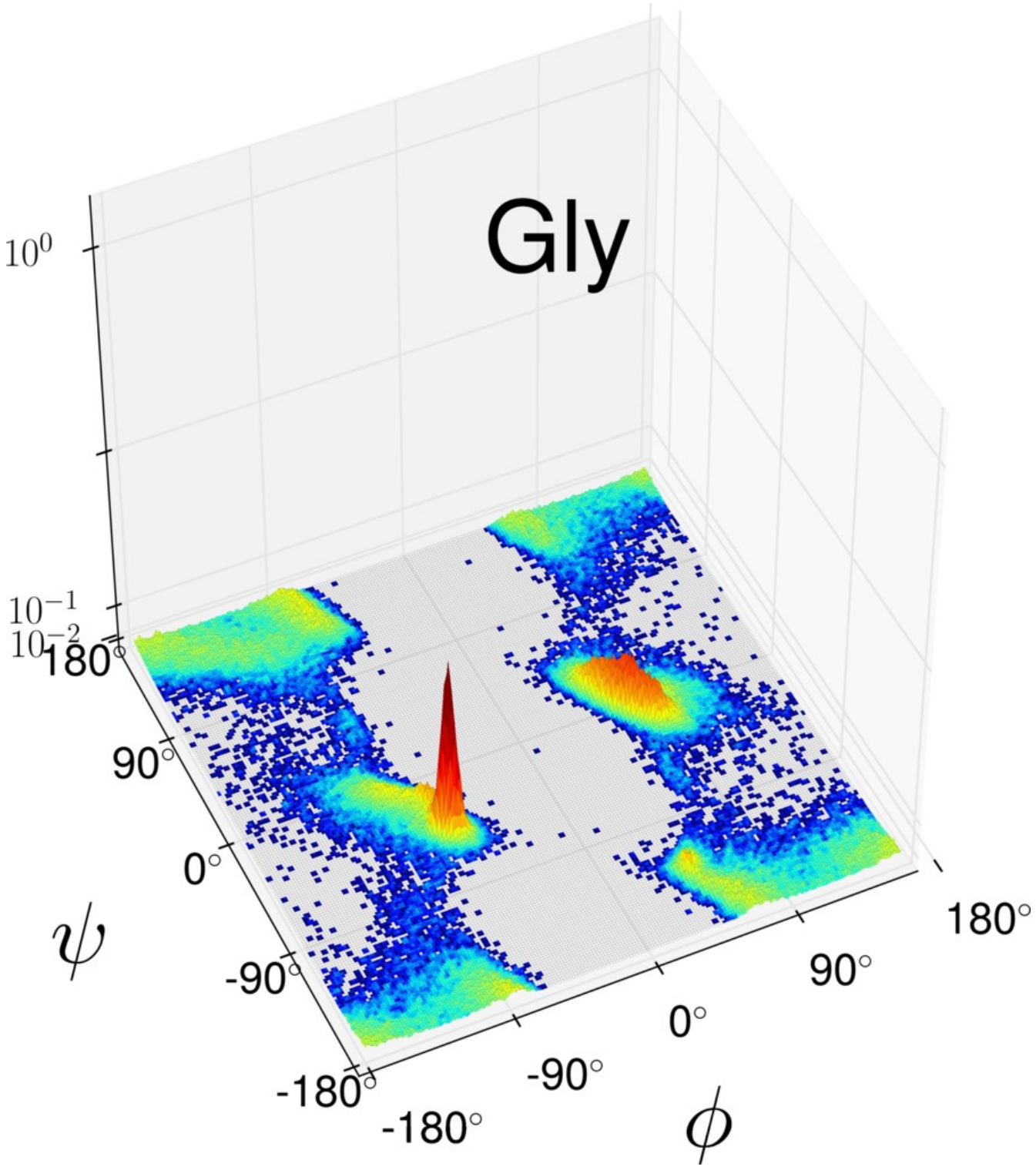


Figure S8. High-resolution Ramachandran distribution $P_{Gly}(\phi, \psi)$ of glycine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

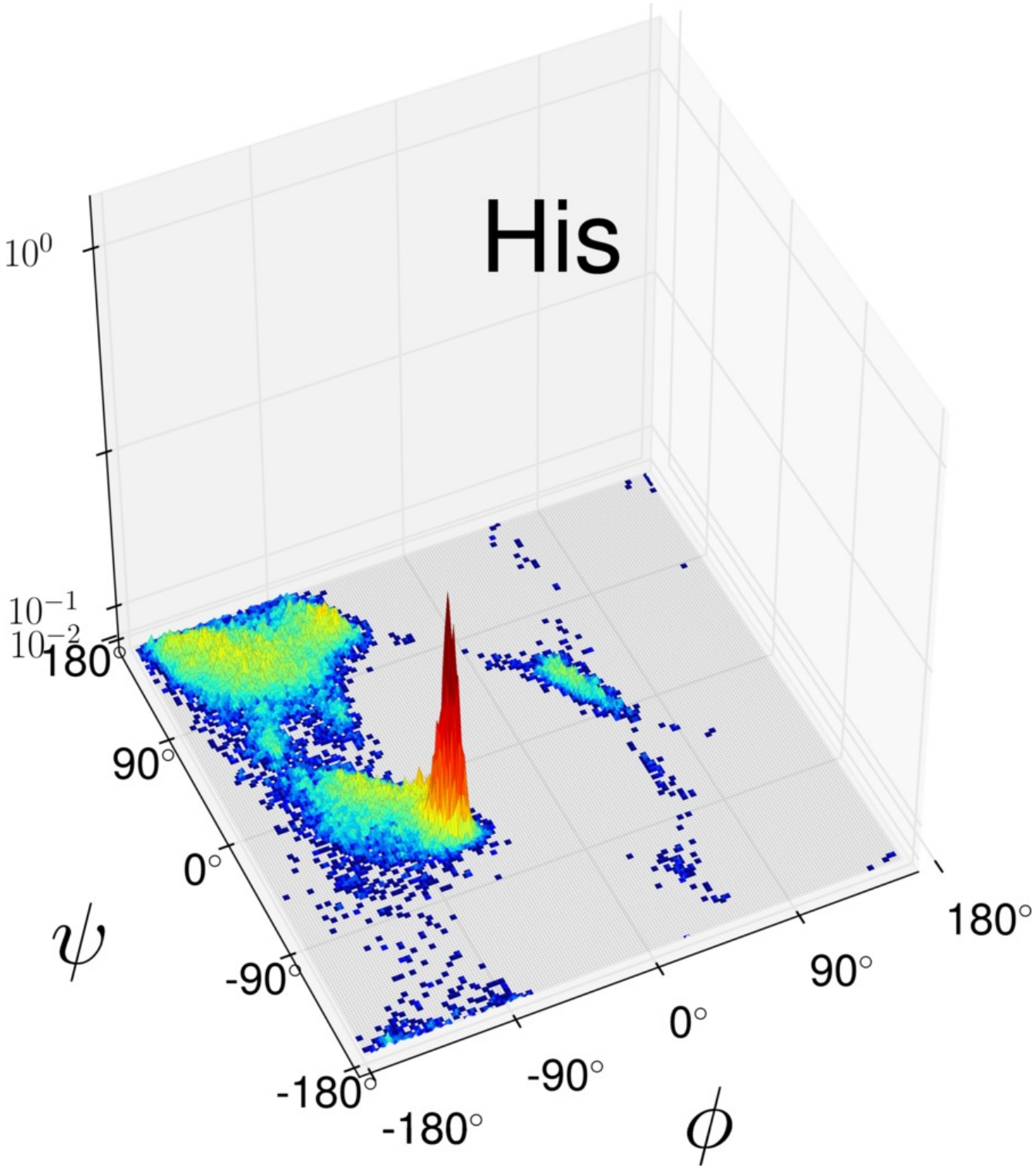


Figure S9. High-resolution Ramachandran distribution $P_{His}(\phi, \psi)$ of histidine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

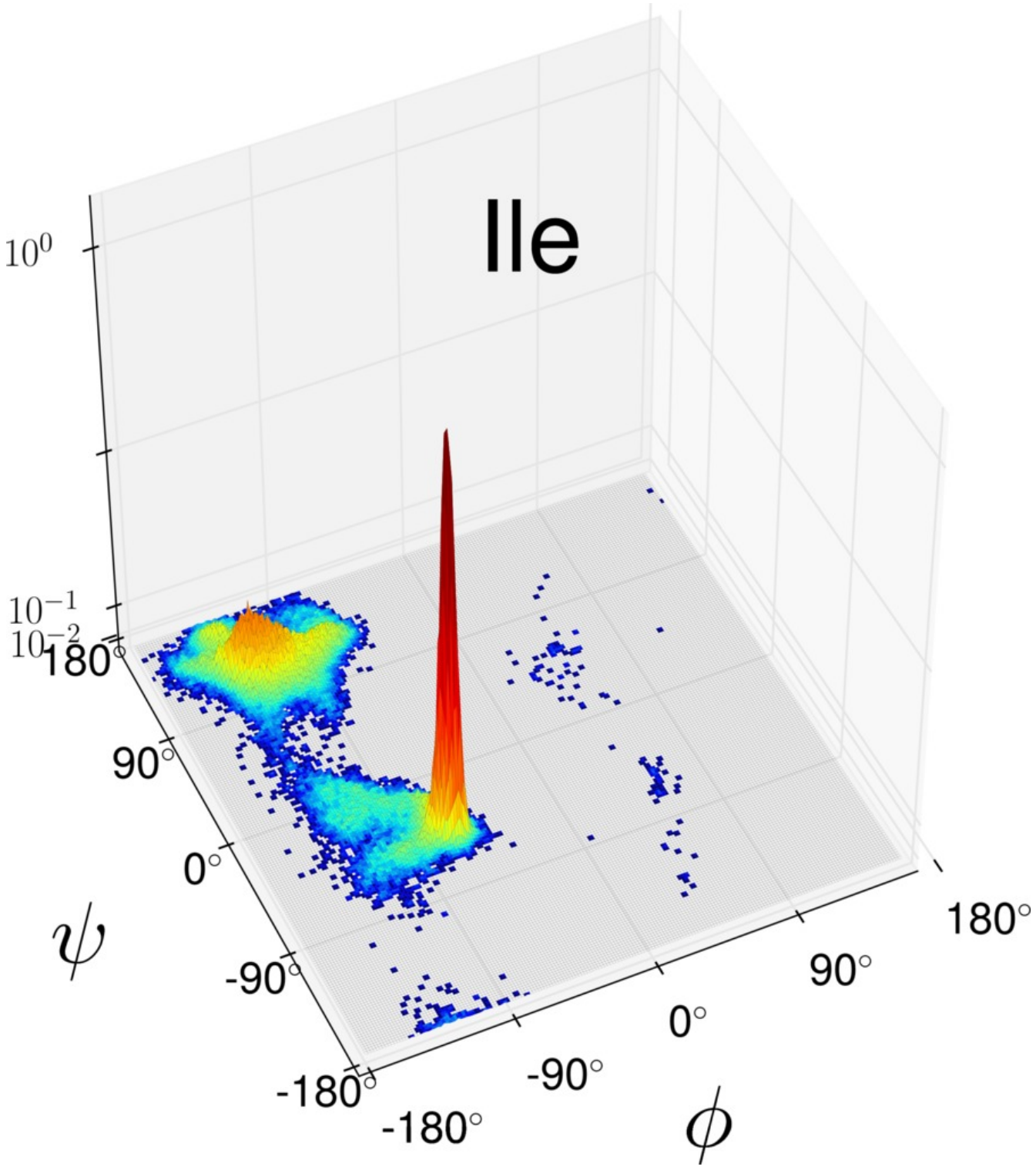


Figure S10. High-resolution Ramachandran distribution $P_{Ile}(\phi, \psi)$ of isoleucine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

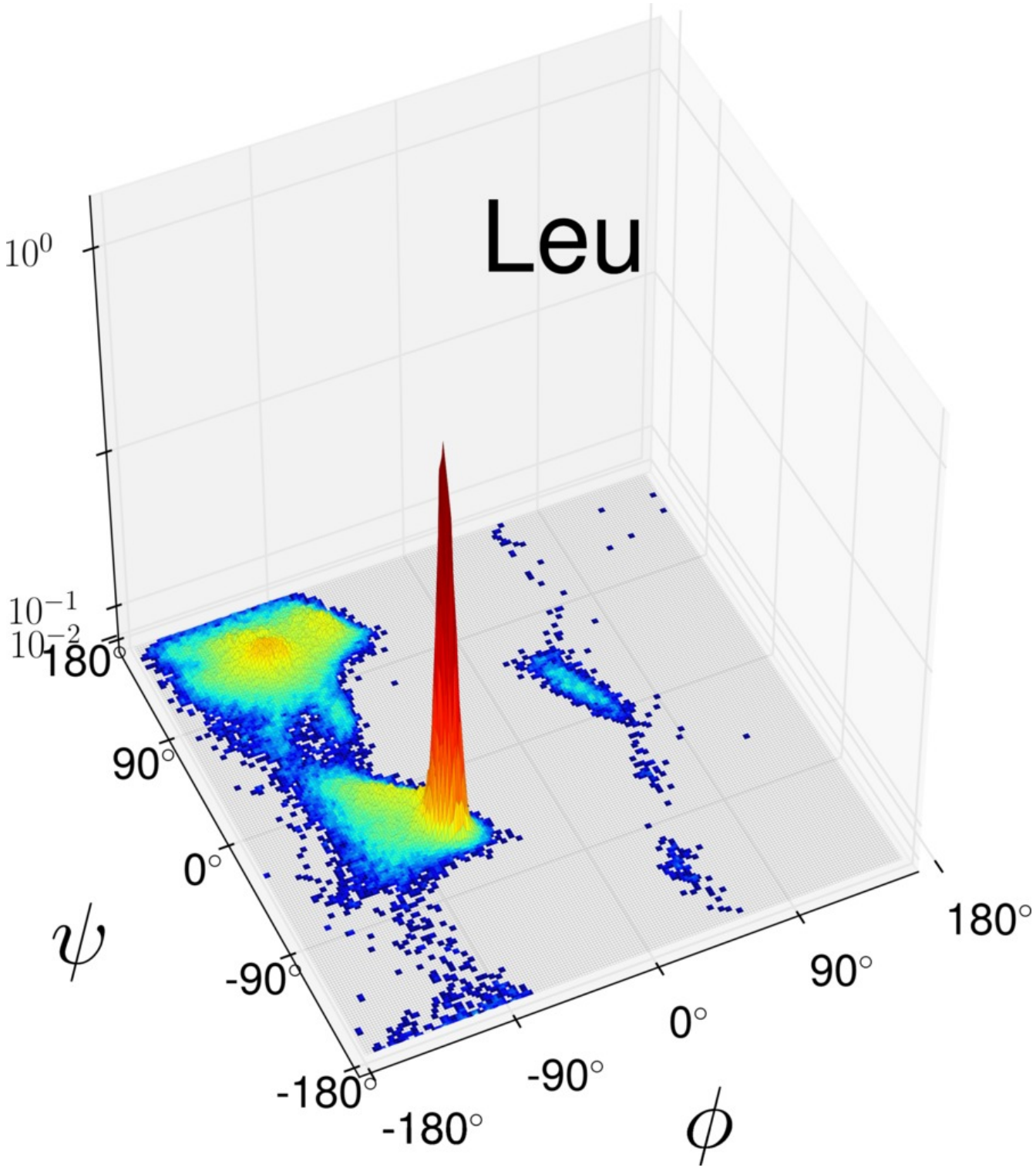


Figure S11. High-resolution Ramachandran distribution $P_{LeuX}(\phi, \psi)$ of leucine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

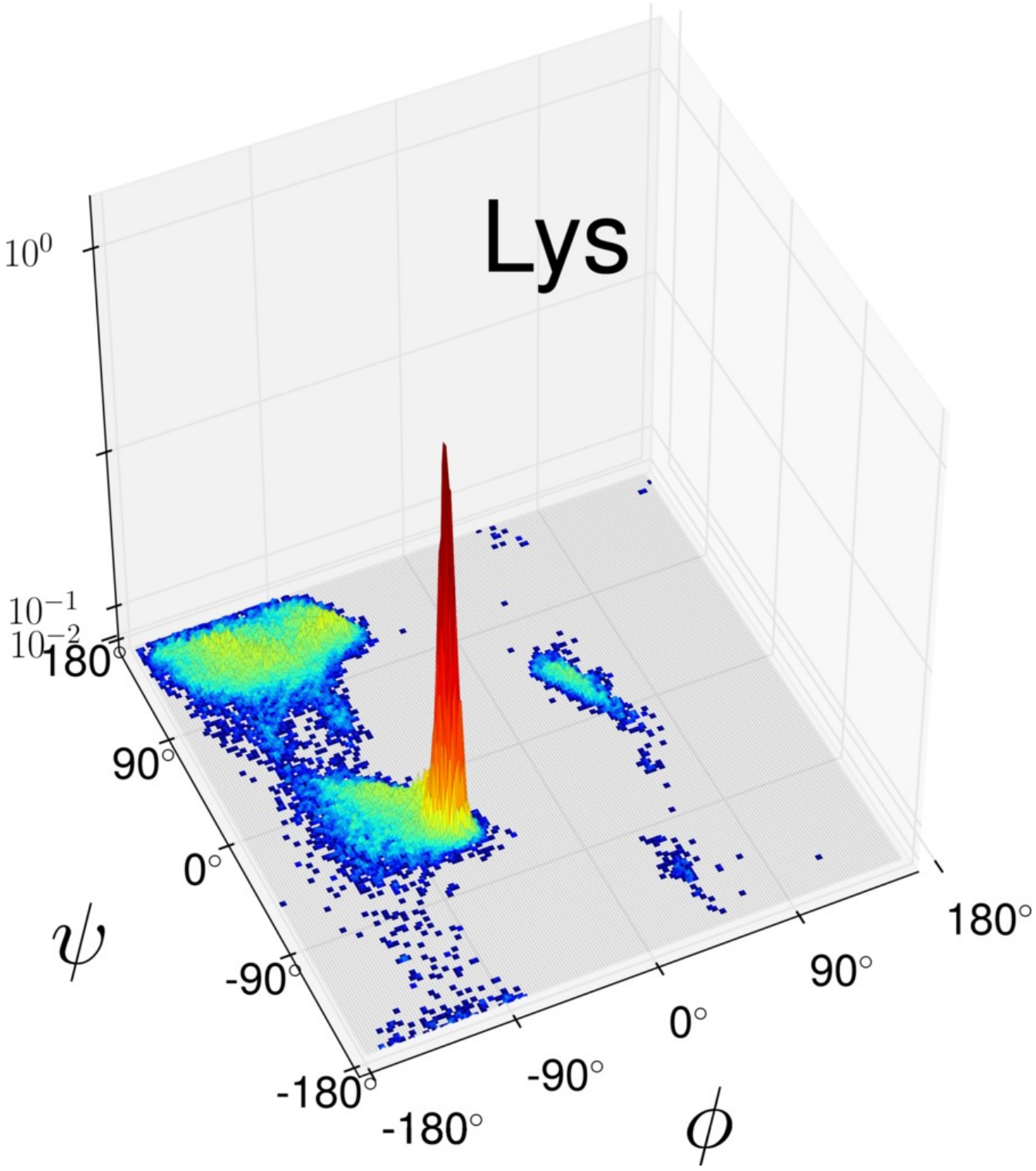


Figure S12. High-resolution Ramachandran distribution $P_{Lys}(\phi, \psi)$ of lysine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

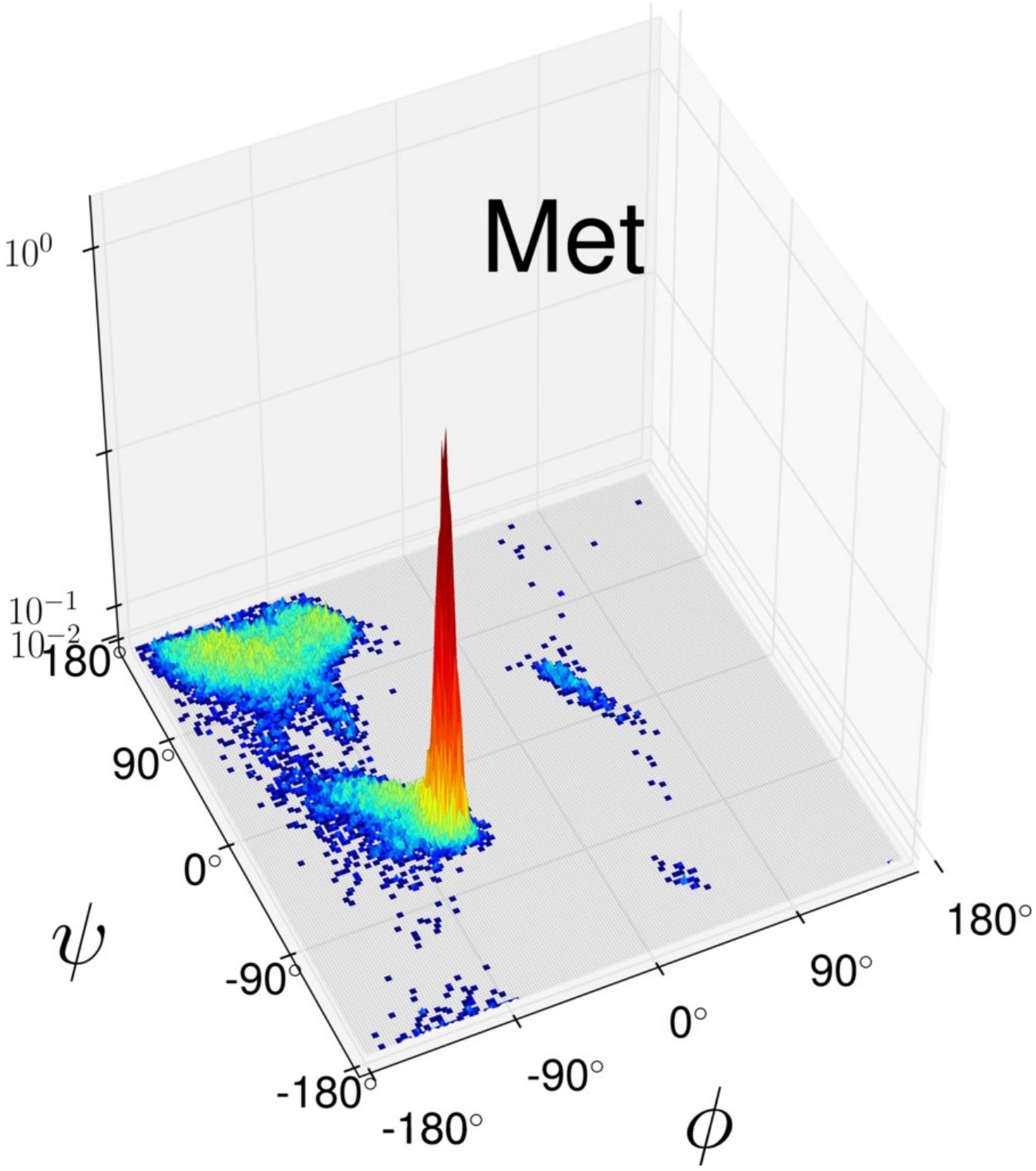


Figure S13. High-resolution Ramachandran distribution $P_{Met}(\phi, \psi)$ of methionine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

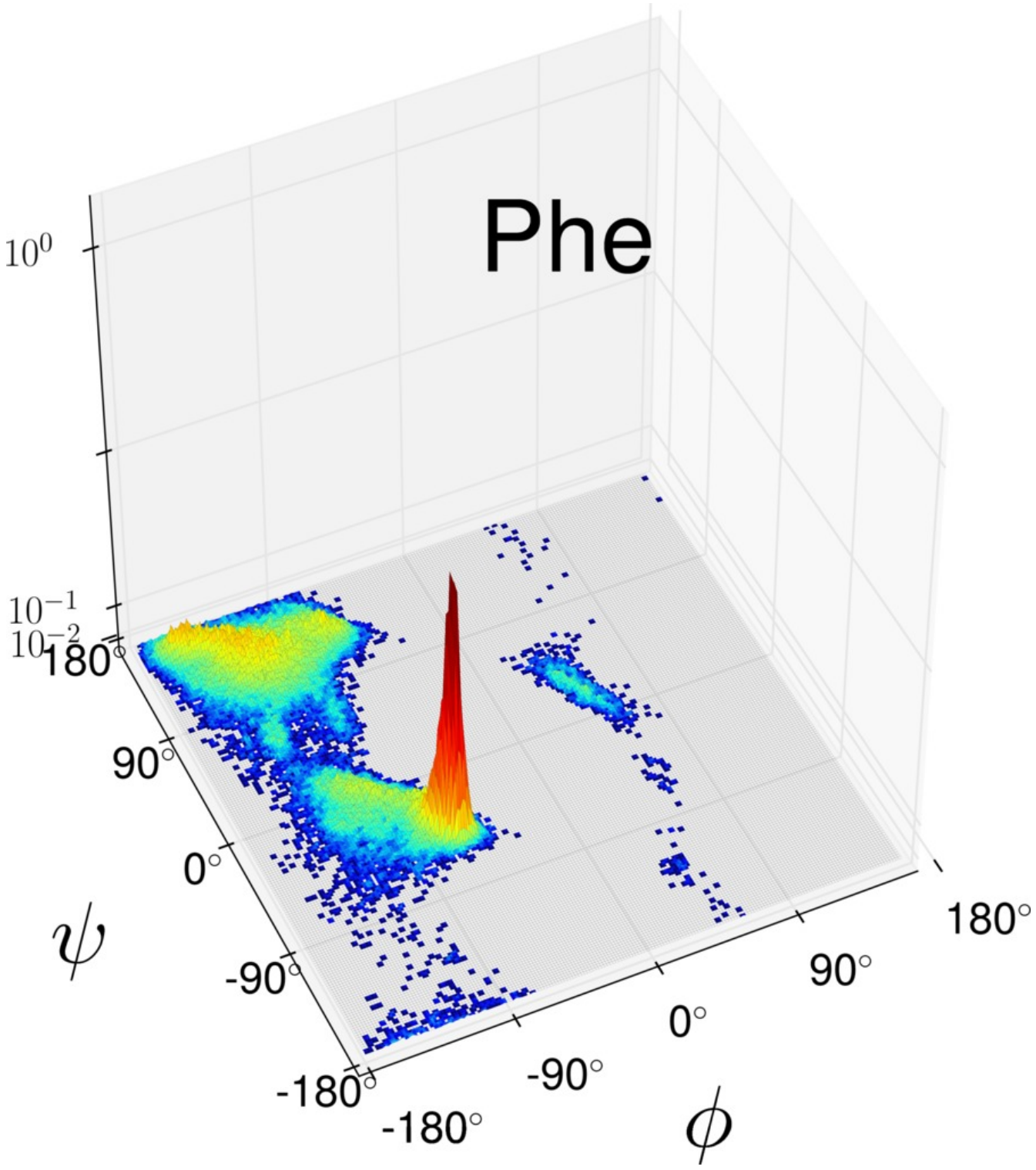


Figure S14. High-resolution Ramachandran distribution $P_{Phe}(\phi, \psi)$ of phenilalanine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

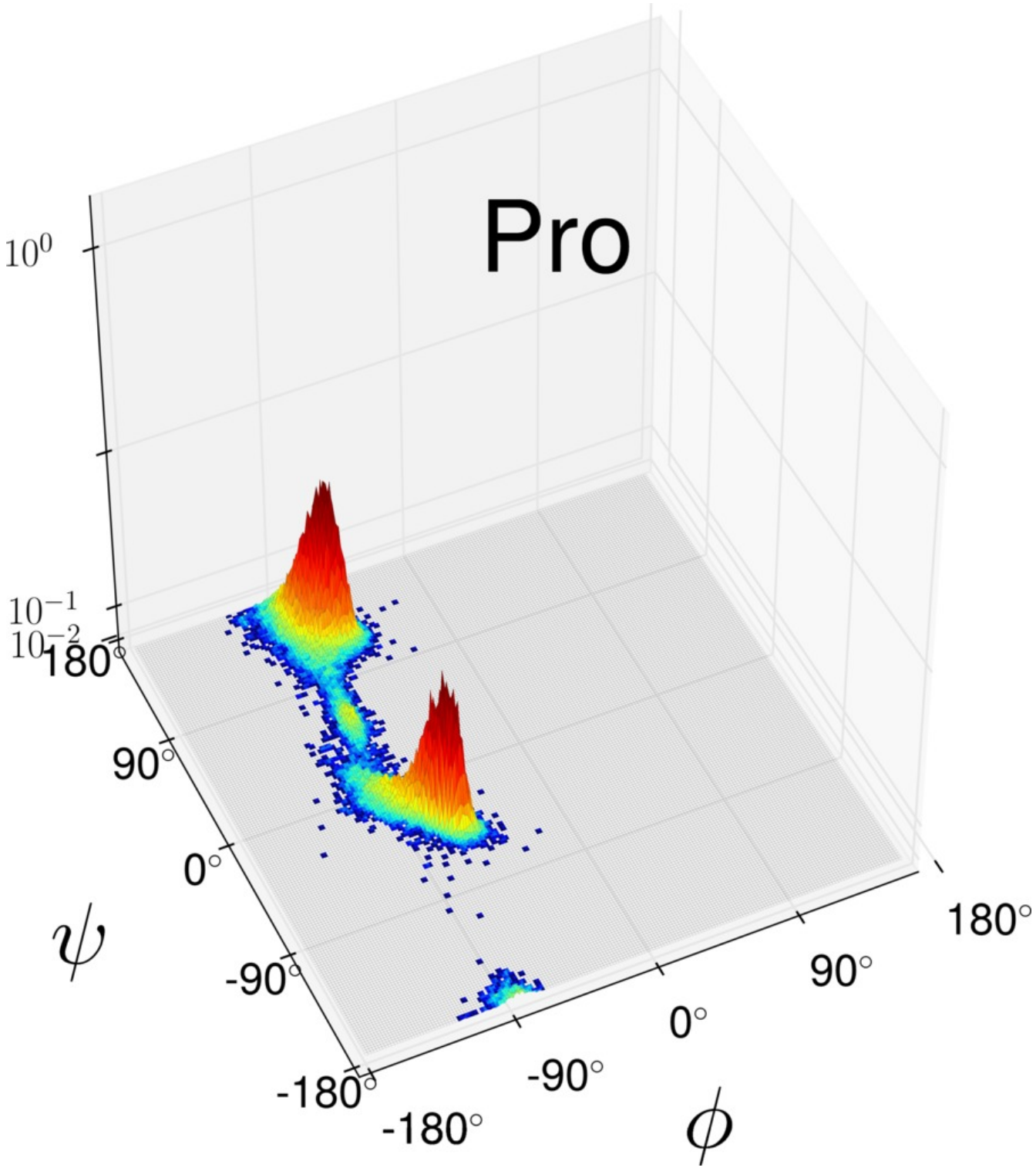


Figure S15. High-resolution Ramachandran distribution $P_{Pro}(\phi, \psi)$ of proline as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

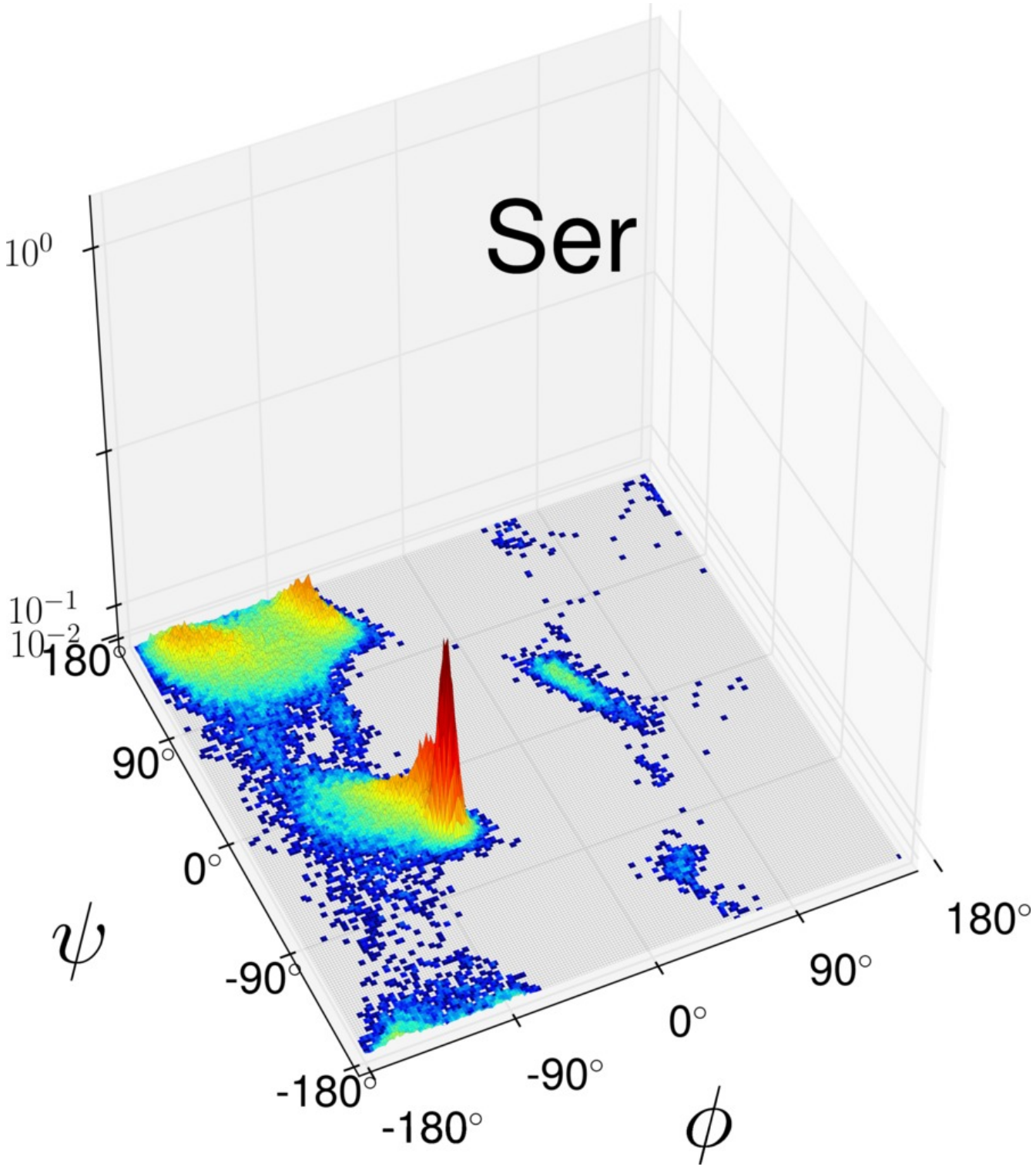


Figure S16. High-resolution Ramachandran distribution $P_{Ser}(\phi, \psi)$ of serine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

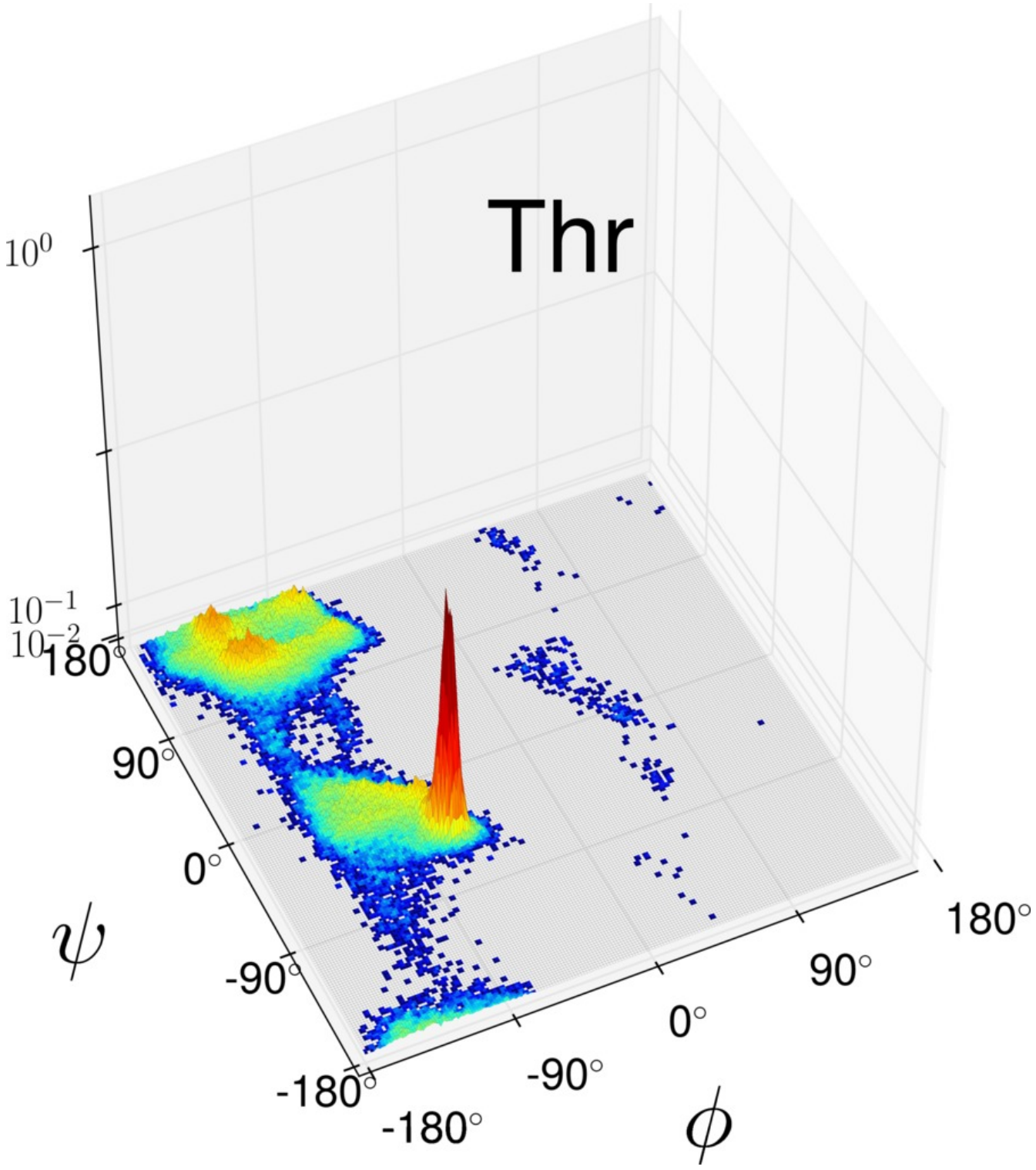


Figure S17. High-resolution Ramachandran distribution $P_{Thr}(\phi, \psi)$ of threonine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

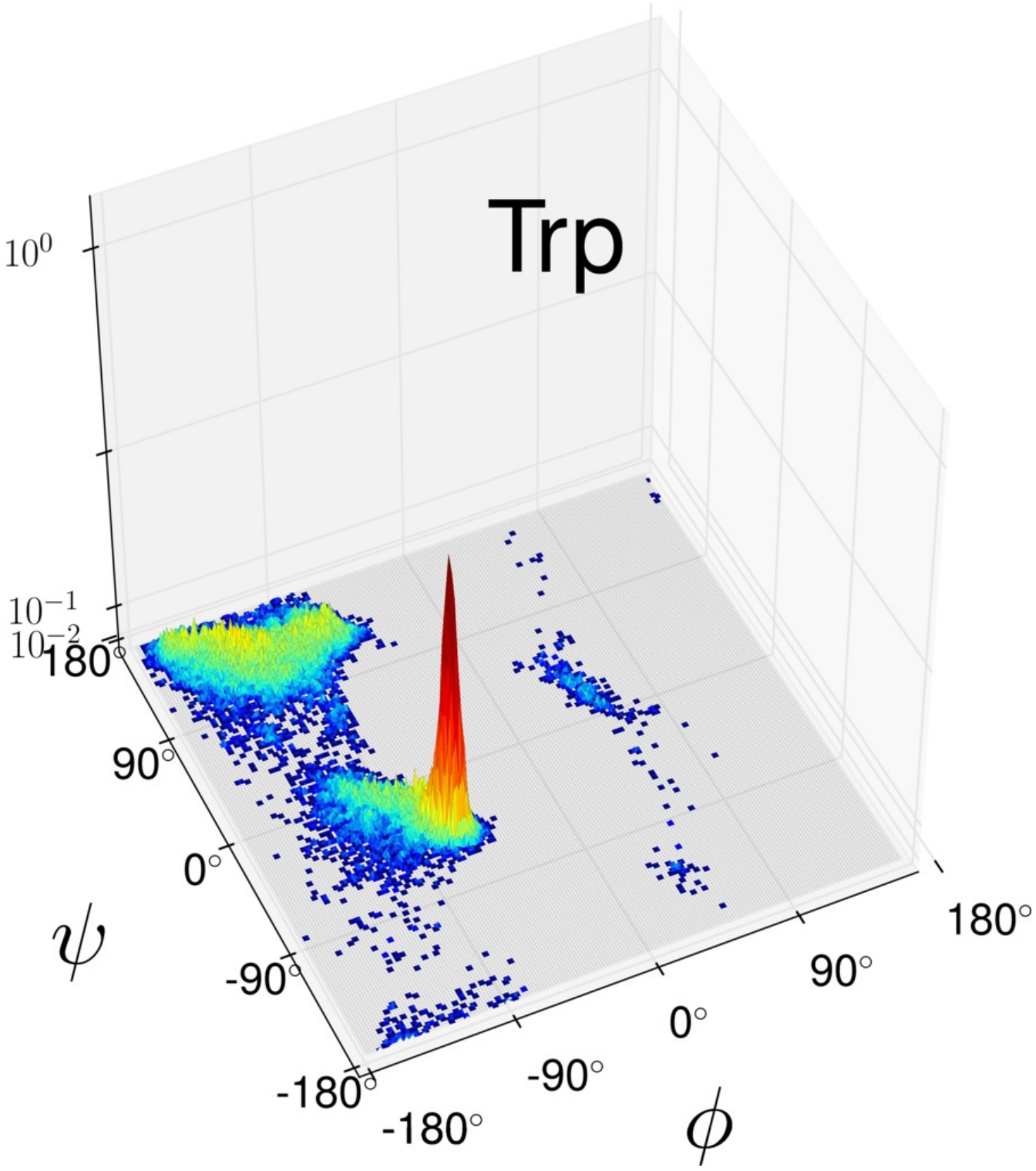


Figure S18. High-resolution Ramachandran distribution $P_{Trp}(\phi, \psi)$ of tryptophan as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

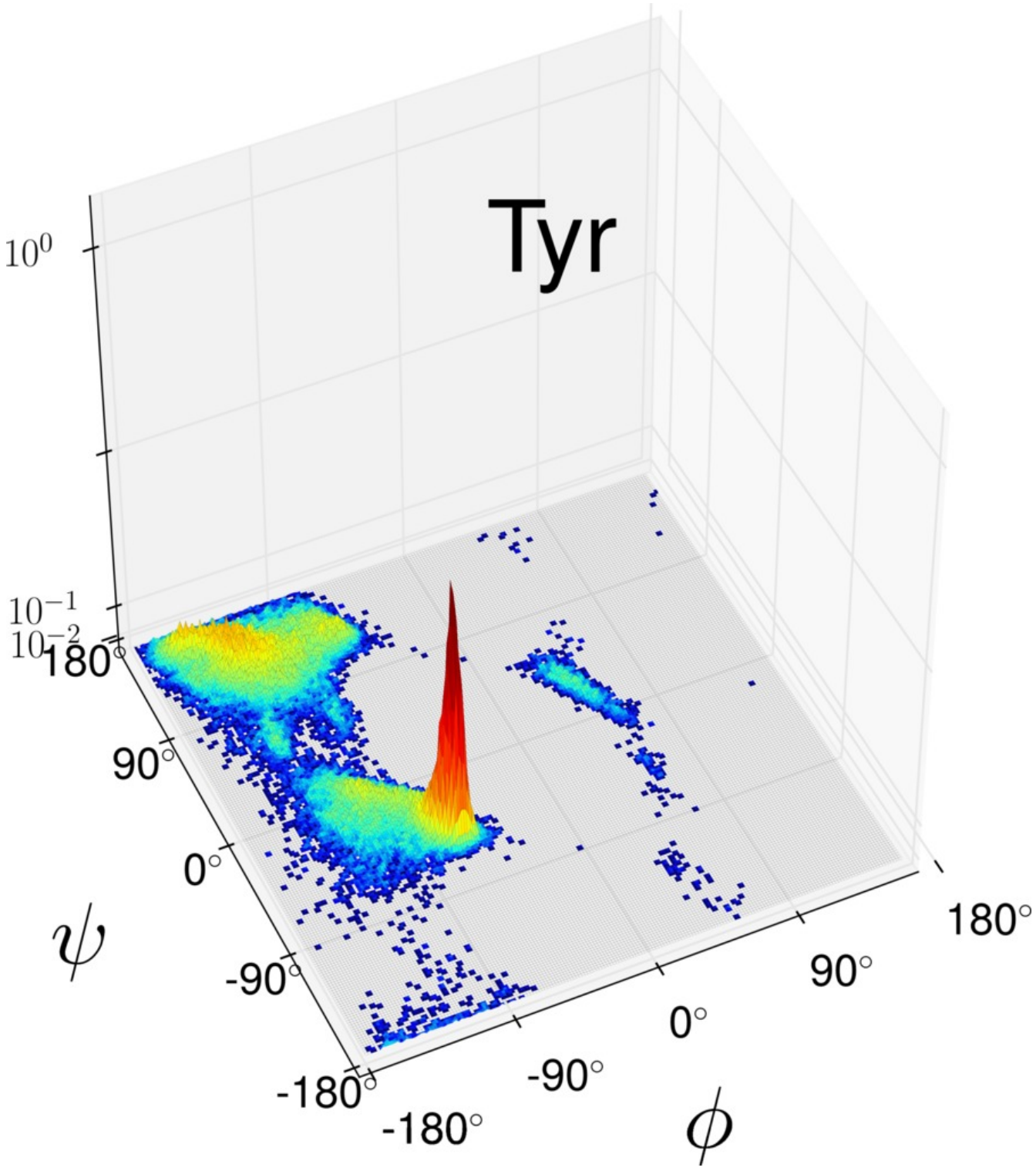


Figure S19. High-resolution Ramachandran distribution $P_{Tyr}(\phi, \psi)$ of tyrosine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

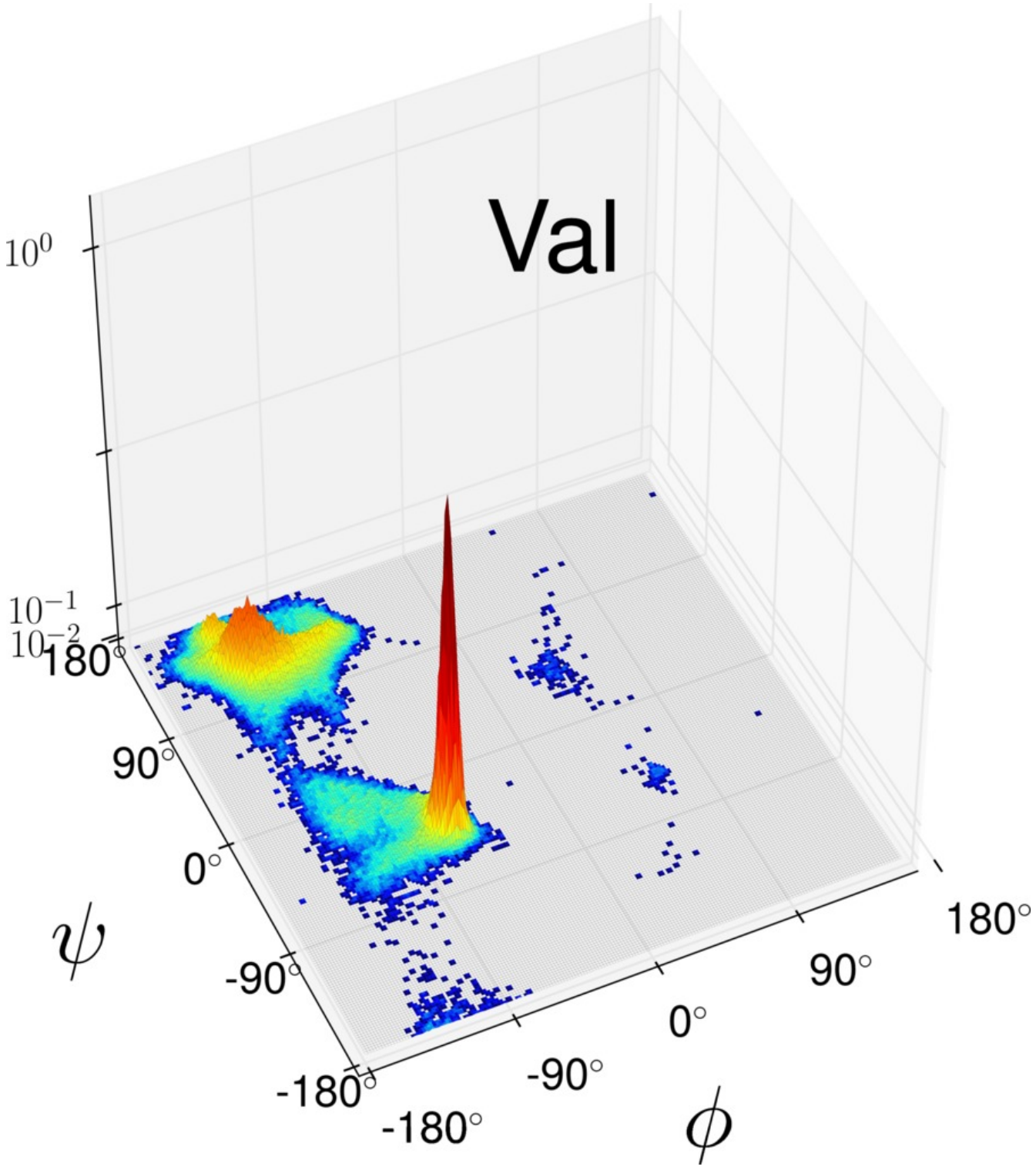


Figure S20. High-resolution Ramachandran distribution $P_{Val}(\phi, \psi)$ of valine as derived from the PGD 1.1 database at $1.895^\circ \times 1.895^\circ$ bin size (logartihmic scale).

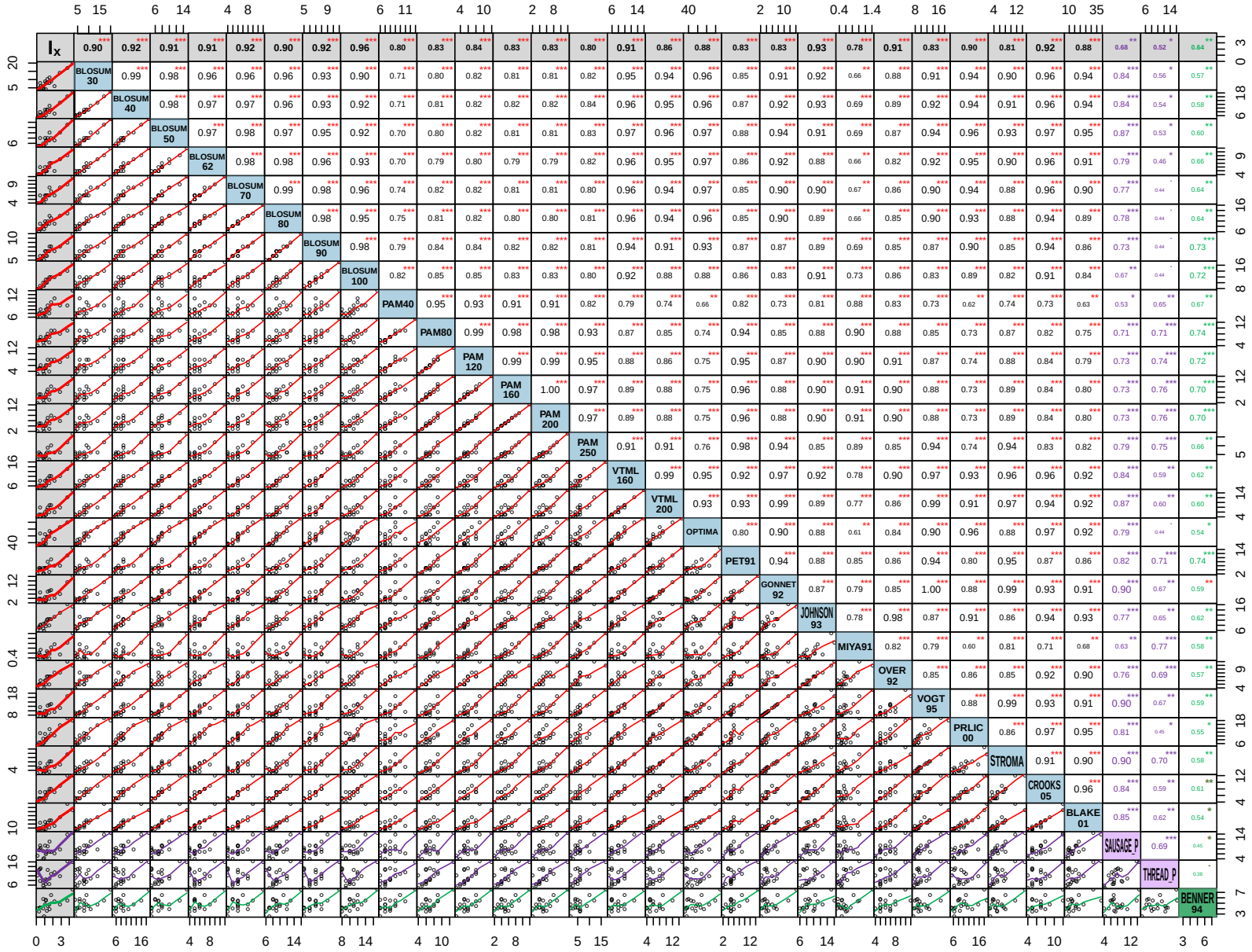


Figure S21. Correlation matrix plot with significance levels between the replacement inertia I_X and the mutability of the full set of replacement matrices used in the present study (Table S1). The lower triangular matrix is composed by the bivariate scatter plots with a fitted smooth line. The upper triangular matrix show the Pearson correlation plus significance level (as stars). Each significance level is associated to a symbol: p-values 0.001 (***), 0.01 (**), 0.05 (*). This plot was generated with the Performance Analytics package in the R program.[3] The abbreviations used in this plot are detailed in Table S1.

Table S1. Abbreviations used in the present study (left) and the corresponding description (center) of the set of substitution matrices with their respective source or AAindex code (right).

Name	Description	AAindex Entry/Source
BLOSUM30	The BLOSUM30 matrix	ftp://ftp.ncbi.nih.gov/blast/matrices/
BLOSUM40	The BLOSUM40 matrix	ftp://ftp.ncbi.nih.gov/blast/matrices/
BLOSUM50	The BLOSUM50 matrix	HENS920104
BLOSUM62	The BLOSUM62 matrix	HENS920102
BLOSUM70	The BLOSUM70 matrix	HENS920103
BLOSUM80	The BLOSUM80 matrix	ftp://ftp.ncbi.nih.gov/blast/matrices/
BLOSUM90	The BLOSUM90 matrix	ftp://ftp.ncbi.nih.gov/blast/matrices/
BLOSUM100	The BLOSUM100 matrix	ftp://ftp.ncbi.nih.gov/blast/matrices/
PAM40	The PAM40 matrix	DAYM780302
PAM80	The PAM80 matrix	ftp://ftp.ncbi.nih.gov/blast/matrices/
PAM120	The PAM120 matrix	ALTS910101
PAM160	The PAM160 matrix	ftp://ftp.ncbi.nih.gov/blast/matrices/
PAM200	The PAM200 matrix	ftp://ftp.ncbi.nih.gov/blast/matrices/
PAM250	The PAM250 matrix	DAYM780301
VTML160	The VTML160 matrix	MUET020101
VTML200	The VTML250 matrix	MUET020102
OPTIMA	The OPTIMA matrix	KANM000101
PET91	The 250 PAM PET91 matrix	JOND920103
GONNET92	The mutation matrix for initially aligning	GONG920101
JOHNSON93	Structure-based amino acid scoring table	JOHM930101
MIYA91	Base-substitution-protein-stability matrix	MIYS930101
OVER92	STR matrix from structure-based alignments	OVEJ920101
VOGT95	Amino acid exchange matrix	VOGG950101
PRLIC00	Homologous structure derived matrix	PRLA000102
STROMA	STROMA score matrix for the alignment of known distant homologs	QUIB020101
CROOKS05	Substitution matrix computed from the Dirichlet Mixture Model	CROG050101
BLAKE01	Matrix built from structural superposition data for identifying potential	BLAJ010101
SAUSAGE_P	Amino acid similarity matrix based on the SAUSAGE force field	DOSZ010101
THREAD_P	Amino acid similarity matrix based on the THREADER force field	DOSZ010103
BENNER94	Genetic code matrix	BENS940104

References

1. Berkholz, D. S., Krenesky, P. B., Davidson, J. R. & Karplus, P. A. Protein Geometry Database: a flexible engine to explore backbone conformations and their relationships to covalent geometry. *Nucleic Acids Res.* **38**, D320–D325 (2010).

2. Shimazaki, H. & Shinomoto, S. A method for selecting the bin size of a time histogram. *Neural Computation* **19**, 1503–1527 (2007).

3. Peterson, B. G. *et al.* Performanceanalytics: Econometric tools for performance and risk analysis. r package version 1.4. 3541 (2014).

Capítulo 5. Semi-empirical quantum evaluation of peptide – MHC class II binding

González R, Suárez CF, Bohórquez HJ, Patarroyo MA, Patarroyo ME. Semi-empirical quantum evaluation of peptide–MHC class II binding. Chemical Physics Letters. 2017;668:29-34

La versión publicada del artículo puede ser consultada en:

http://www.sciencedirect.com/science/article/pii/S0009261416309642

Semi-empirical quantum evaluation of peptide – MHC class II binding

Ronald González[a,b,e], Carlos F. Suárez[a,b,c,e], Hugo J. Bohórquez[a,b,c], Manuel A. Patarroyo[a,b], Manuel E. Patarroyo[a,d,*]

[a] *Fundación Instituto de Inmunología de Colombia (FIDIC), Bogotá D. C., Colombia*
[b] *Universidad del Rosario, Bogotá D. C., Colombia*
[c] *Universidad de Ciencias Aplicadas y Ambientales (UDCA), Bogotá D. C., Colombia*
[d] *Universidad Nacional de Colombia, Bogotá D. C., Colombia*
[e] *Both authors equally contributed as first author*

Abstract

Peptide presentation by the major histocompatibility complex (MHC) is a key process for triggering a specific immune response. Studying peptide-MHC (pMHC) binding from a structural-based approach has potential for reducing the costs of investigation into vaccine development. This study involved using two semi-empirical quantum chemistry methods (PM7 and FMO-DFTB) for computing the binding energies of peptides bonded to HLA-DR1 and HLA-DR2. We found that key stabilising water molecules involved in the peptide binding mechanism were required for finding high correlation with IC_{50} experimental values. Our proposal is computationally non-intensive, and is a reliable alternative for studying pMHC binding interactions.

Keywords: FMO-DFTB, PM7, HLA-DR, Receptor-ligand interactions

*Corresponding author

1. Introduction

The major histocompatibility complex (MHC) —or human leukocyte antigen (HLA) in humans— plays a key role in an adaptive immune response against pathogens and cancer, presenting self and non-self peptides to T-cells. Researching peptide-MHC (pMHC) binding mechanisms should improve our understanding of pathogenic diseases, autoimmunity and cancer; consequently, this is of paramount importance in designing drugs and vaccines [1].

MHC molecules involved in antigen presentation can be divided into two classes: I and II. MHC class I molecules bind especially to endogenous peptides and are present in all nucleated cells. MHC class II molecules are expressed in professional antigen-presenting cells (such as dendritic and B-cells) and bind to exogenous antigens. Although MHC class I and II peptide binding region (PBR) have similar architecture —a groove that attaches antigenic peptides within a binding frame of nine amino acids (P1 to P9)—, MHC class I having a unique binding frame while MHC class II PBR has an open groove, consequently, calculations of pMHC binding for MHC class II is difficult, because peptide length variation and multiple binding frames increasing the amount of required calculations [2].

Studying peptide binding to MHC is extremely challenging: First, the receptor isolation and the binding assays themselves require extensive and expensive testing[3, 4]; second, the high MHC polymorphism increases the number of molecular systems to be studied [5]; and third, up to 1.2 x 10^{19}

potential peptides might bind to each receptor. A promising line of attack is the use of computational methods to evaluate whether a given pMHC binding occurs, thereby reducing the number of experimental measurements required.

The computational methods for pMHC binding estimation can be divided into *sequence-based methods* —which use experimental binding data as training input for several kind of algorithms (*e.g.* neural networks) [6]; and *structure-based methods* —which use mainly the pMHC binding energy from structural information alone[7]; this approach is specially advantageous for studying pMHC interactions, due to its independence from experimental data and the possibility of obtaining structures of non-crystallised complexes using homology modelling [8].

The present work describes a structure-based approach, using quantum mechanical semi-empirical methods for calculating pMHC-DR binding energies. Semi-empirical methods can be defined as the simplest version of electronic structure theory; by performing a large number of approximations and parameterisations it is possible to obtain an efficient computational approach [9]. The PM7 method and the density-functional tight-binding method (DFTB) are two of the most used and efficient semi-empirical methods for studying large bio-molecular systems [10, 11]. Furthermore, the recent implementation of the fragment molecular orbital method (FMO) on DFTB [12] has reduced computation times by dividing large bio-molecules into smaller pieces [13, 14].

We calculated the binding energy of 22 peptides bound to MHC class II: HLA-DR1 (8 peptides) and HLA-DR2 (14 peptides). Ligand-receptor binding has high sensitivity to water molecules in the interface[15]; the role of water molecules has already been described regarding pMHC binding [16]. Thus, we including crystallographic waters located near the pMHC interface and correlated these values with the corresponding experimental binding affinities (IC_{50}), estimating the capacity of discriminate binders from non-binders using receiver operating characteristic (ROC) analyses.

2. Methodology

2.1. Studied sets

Two sets of MHC class II molecules were studied: 1) a crystallised HLA-DR1 structure, (HLA-DRA*01:01/HLA-DRB1*01:01, pdb code 1DLH) complexed with haemagglutinin peptide ($HA_{306-318}$) [17], using IC_{50} experimental values for native $HA_{306-318}$ and 7 mono-substituted (Asp) analogues from Geluḱs *et al's.* study [18] (FIG 2A) and 2) a crystallised HLA-DR2 structure (HLA-DRA*01:01/HLA-DRB1*15:01, pdb code 1BX2) complexed with myelin peptide (MY_{86-96}) [19], using IC_{50} experimental values for native MY_{86-96} and 13 mono-substituted (Ala) analogues from Krogsgaard́s *et al's.* study [20] (FIG 3A). The sequence variation in HLA-DR molecules focused on the HLA-DRB gene (being the most polymorphic MHC class II in humans), HLA-DRA being almost monomorphic [21]. In this case, DRB1*01:01 vs. DRB1*15:01 had 5% sequence divergence in the β_1 domain, showing very

different peptide-binding profiles [22]. The HLA-DR1 set had well differentiated IC_{50} values (separated into four orders of magnitude, IC_{50} values ranging from 5 to >12,500 nM) Figure 2A, while the HLA-DR2 set had a more challenging IC_{50} range, having narrow IC_{50} values (4 to 199 nM) Figure 3A, some repeated several times. The chosen HLA-DR sets enabled evaluating peptide mono-substitutions using two different kinds of amino acids, Asp for DR1 set and Ala for the case of DR2 set.

2.2. Structure preparation and modelling

Amino acid substitutions were made in peptides using the UCSF Chimera *swapaa* function, using the Dunbrack backbone-dependent rotamer library [23, 24]. The first preparation step involved adding hydrogen atoms to the protein structures using MOPAC2016 software[25]. It should be noted that crystal structures must be optimised before any kind of calculation can be made (for example, binding energies) since minor errors in protein atom coordinates could become in non-realistic energies. We explored several optimization strategies, and found that the best result was obtained optimising hydrogen atoms using the PM7 method with conductor-like screening model (COSMO) as an implicit solvent model with fixed heavy atoms in their crystallographic positions. All residues were neutralised. This strategy has been used previously in ligand-receptor studies [26]. Calculations included all crystallographic water molecules within a radius of ≤ 8.0 Å around the peptide. The computing time for the minimisation of the near 3050 hydrogen atoms

($\sim$ 50% of atoms for each system) took 6 hours using 4 CPU cores.

2.3. Binding calculations using the PM7 method

Using the previously optimised models, binding enthalpies for the pMHC complexes were calculated according to the following equation:

$$\Delta H_{bind}^{PM7} = \Delta H_{complex} - \Delta H_{receptor} - \Delta H_{peptide}, \quad (1)$$

where $\Delta H_{complex}$ is the calculated enthalpy of formation for the pMHC complex, $\Delta H_{receptor}$ is the calculated enthalpy of formation for the MHC protein without peptide and $\Delta H_{peptide}$ is the calculated enthalpy of formation for the peptide. Binding energies calculated by the PM7/COSMO method took some minutes (15 minutes) on 4 CPU cores.

2.4. Binding calculations using the FMO-DFTB method

We used the FMO-DFTB method (version 5.1) [27] as implemented in General Atomic and Molecular Electronic Structure System (GAMESS)[28]. The first step in this method consisted of assigning every atom to a fragment. The second step involved calculating a self-consistent field (SCF) for every fragment due to the presence of electrostatic field generated by the remaining fragments. The third step consisted of fragment pair SCF calculations (i.e., the inter-fragment interaction energy) and total properties evaluation, for instance: total energy, gradient, minimisation, etc. These steps summarise the two-body FMO approach.

Total energy E in the two-body FMO expansion is:

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} \left(E_{IJ} - E_I - E_j \right), \tag{2}$$

where E_I is the energy of monomer I immersed in the external electrostatic potential generated by the remaining monomers; E_{IJ} is the interaction energy of dimer IJ, which is also immersed in the external electrostatic potential of the other fragments.

Using the optimised models with the PM7/COSMO method, total energies for the pMHC complexes and its components were calculated using equation 2 and binding energies were calculated using equation 1 at the FMO2-DFTB level of theory. Binding energies calculated by FMO-DFTB method took only some minutes (5 minutes) on 1 CPU core.

2.5. Statistical test of pMHC binding energies vs. experimental IC_{50}

We used a linear model for $\ln(\mathrm{IC}_{50})$ vs. ΔH_{bind} to calculate determination coefficients R^2. Receiver operating characteristic (ROC) analyses were performed —using the R program pROC package [29]— to estimate the values of the area under the curve (AUC). Affinity IC_{50} cutoffs for binary codification were: very strong binders ($\leq$ 5 nM), strong binders ($\leq$ 50 nM) and weak binders ($\leq$ 500 nM).

3. Results and discussion

We only found strong correlations between ΔH_{bind} and IC_{50} by keeping the crystallographic water molecules. These results agreed with Petrone et al. [16] who studied class I pMHC complexes, finding that bound water molecules in the interface have two main tasks: filling empty spaces and bridging hydrogen bonds between the MHC and a peptide. Interestingly, Li et al., [30] found that breaking the water-mediated hydrogen bond network produced a binding energy loss of at least 8 kcal/mol, as for class I pMHC complexes. We only focused on crystallographic waters located within a radius of ≤ 8.0 Å from the peptide. The correlations observed with this approach were the same as those including all water molecules in the calculations. Hence, only water molecules in close proximity to the pMHC contact region were required for an accurate estimation of binding energy. The correlation plots for experimentally measured IC_{50} values and calculated binding energies are shown in Figure 2 for the HLA-DR1 set and in Figure 3 for the HLA-DR2 set.

The same high correlation value ($R^2 = 0.81$) for the HLA-DR1 set was found with the semi-empirical methods used; however, FMO-DFTB gave a higher AUC value for discriminating strong binders ($AUC_{DFTB} = 0.86$) than PM7 ($AUC_{PM7} = 0.71$). On the other hand, FMO-DFTB outperformed PM7 with the HLA-DR2 set, having a correlation of $R^2_{DFTB} = 0.74$ *vs.* $R^2_{PM7} = 0.61$ for FMO-DFTB. This was also true for strong binders discriminated by AUC values: $AUC_{DFTB} = 0.94$ *vs.* $AUC_{PM7} = 0.74$. Overall, FMO-DFTB showed better predictability than PM7. Moreover, compared to the best

sequence-based method, NetMHCIIpan 3.1 [6] (HLA-DR1 set $R^2 = 0.75$ and HLA-DR2 set $R^2 = 0.66$), our results had better or equivalent correlation with experimental IC_{50} values.

Entropic contributions are important during the binding process, since peptide's flexibility entails large conformational changes [31]. In addition, some solvent molecules must be displaced from the corresponding binding region during a specific ligand's docking; ergo, a desolvation energy could play an important role in determining binding energies [32]. Therefore, the strong correlation between the computed values of enthalpy ΔH_{bind} and IC_{50} experimental values indicate that these contributions were small regarding the present cases.

The receptor cavities interacting with the peptide side-chains of positions P1, P4, P6, P7 and P9 are called pockets, and named according their interacting peptide amino acid. Our binding energy calculations indicated the following pocket order for the HLA-DR1 set (see Figure 2): Pocket-1 $\gg$ Pocket-7 $\gg$ Pocket-6 $>$ Pocket-4, which is in perfect agreement with the experimentally measured IC_{50} values. Remarkably, Tyr 308 substitution in peptide position 1 (P1) yielded a four orders of magnitude variation in IC_{50} values, making this one of the most important anchoring residues for HLA-DR1 set studied here. Moreover, it is well known that Pocket-1 has a strong preference for large hydrophobic side-chains, presumably being the most determinant binding site[17]. Consequently, substituting Leu for Asp in peptide position 314 (P7) (Figure 2) changed peptide binding to HLA-

DR1 by up to two orders of magnitude. On the other hand, replacing Thr by Asp in position 313 (P6) produced a one order of magnitude change in binding energy —big enough for altering binding affinity from a high binder to a non-binder. Substituting Val, Lys, Gln and Asn for Asp in positions 309 (P2), 310 (P3), 311 (P4), and 312 (P5), respectively, all gave high binding energies.

PM7 and FMO-DFTB binding energies agreed with the respective IC_{50} values for the HLA-DR2 set, yielding the following pocket order: Pocket-4 $\gg$ Pocket-1 $>$ Pocket-6 = Pocket-7 = Pocket-9. In this case, hydrophobic pocket 4 is the primary binding site in the PBR[19]. Substituting Val for Ala in position 89 (P1) produced a substantial change in binding energy; In this case, pocket 1 had a secondary role according to the HLA-DR2 set's peptide binding energies —unlike the HLA-DR1 set. Furthermore, replacing Asn, Ile and Thr by Ala in peptide positions 94 (P6), 95 (P7), and 97 (P9), respectively, left HLA-DR2 binding energies unaltered. Our results for both sets revealed definite variability regarding HLA pocket binding hierarchy, relative to anchoring residues. This may well be a result of PBR differences due to each receptor's specific pocket architecture.

We explored the stabilising role of water molecules regarding the mechanism of peptide binding to a class II MHC —HLA-DR2 set— by replacing Asn-94 (P6) for Ala in the $Myelin_{86-98}$ peptide. According to the protein crystal structure, Myelin's Asn-P6 side-chain is buried within HLA polar pocket 6 (Figure 4A.). This amino acid makes a stabilising network consist-

ing of five hydrogen bonds involving Glu α11, Arg β13, and Asn α62, amino acids. The guanidinium group of Arg β13 participates in two hydrogen bonds with the carboxyl oxygen from the Asn-P6 side-chain. Simultaneously, Asn-P6 side-chain amide group establishes two hydrogen bonds: one with Asn α62 backbone oxygen and another with the unprotonated oxygen from the Glu α11 side-chain carboxylic acid. The backbone hydrogen of the Asn-P6 amide group makes a hydrogen bond with Asn α62 side-chain carboxyl oxygen. This latter hydrogen bond remained unchanged after replacing Asn-P6 for Ala-P6, as indicated by the arrow in Fig. 4B. However, the missing hydrogen bonds destabilised anchoring by 16.5 and 7.6 kcal/mol with PM7 and FMO2-DFTB, respectively. Such computations contradicted the binding reported by IC_{50} values for the myelin$_{86-98}$ (IC_{50} = 5 nM) and MY A94 (IC_{50} = 4 nM) peptides, thereby indicating similar stabilising interactions. Accordingly, these results lowered the correlations between enthalpies and IC_{50} values for the whole set, at both levels of theory, to $R^2_{PM7} = 0.15$ and $R^2_{DFTB} = 0.47$.

Interestingly, adding a water molecule at the location of the former amide Asn-P6 group created three hydrogen bonds locally stabilising the Ala-P6 side-chain. The hydrogen atoms of this water molecule coordinate the Asn α62 backbone carboxylic carbon and the unprotonated oxygen of the Glu α11 side-chain carboxylic acid, i.e. similar to the Asn-P6 side-chain. The water molecule's oxygen makes a hydrogen bond with a hydrogen from the Ala-P6 side-chain. As can be seen in Figure 4B, the water molecule re-

constructed a great part of the former hydrogen bond network, which was consistently reflected in stronger binding energy. This correction alone raised correlation values between the binding energies and the IC_{50} values for both semi-empirical methods: PM7 (R^2 = 0.61) and FMO-DFTB (R^2 = 0.74) (Fig. 3). These results demonstrate the stabilising role of water molecules at the pMHC interface.

4. Conclusions

Studying two different pMHC systems gave strong correlation between calculated binding energies and experimental IC_{50} values. Our binding energy calculations discriminated weak from strong and even very strong binders having a high level of accuracy, thereby showing the advantages of the approach proposed here. It provides valuable proof that semi-empirical quantum mechanical methods are reliable and cost-effective for studying high complex systems —such as the pMHC HLA-DR1 and HLA-DR2 systems. The two levels of theory used here (DFTB and PM7) are fast enough — assuming conventional computational resources— to understand the pMHC binding. We anticipate increasing use of these quantum methods in the near future for drug and synthetic vaccine design.

5. Acknowledgments

We would like to thank Jason Garry for revising the text. We also want to thank Dmitri Fedorov for his support in implementing the FMO-DFTB

method.

6. References

[1] Manuel E Patarroyo and Manuel A Patarroyo. Emerging rules for subunit-based, multiantigenic, multistage chemically synthesized vaccines. *Accounts of chemical research*, 41(3):377–386, 2008.

[2] Linus Backert and Oliver Kohlbacher. Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome medicine*, 7(1):1, 2015.

[3] Peng Wang, John Sidney, Courtney Dow, Bianca Mothe, Alessandro Sette, and Bjoern Peters. A systematic assessment of mhc class ii peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol*, 4(4):e1000048, 2008.

[4] John Sidney, Scott Southwood, Carrie Moore, Carla Oseroff, Clemencia Pinilla, Howard M Grey, and Alessandro Sette. Measurement of mhc/peptide interactions by gel filtration or monoclonal antibody capture. *Current protocols in immunology*, pages 18–3, 2013.

[5] John Trowsdale and Julian C Knight. Major histocompatibility complex genomics and human disease. *Annual review of genomics and human genetics*, 14:301, 2013.

[6] Massimo Andreatta, Edita Karosiene, Michael Rasmussen, Anette Stryhn, Søren Buus, and Morten Nielsen. Accurate pan-specific prediction of peptide-mhc class ii binding affinity with improved binding core identification. *Immunogenetics*, 67(11-12):641–650, 2015.

[7] Atanas Patronov and Irini Doytchinova. T-cell epitope vaccine design by immunoinformatics. *Open biology*, 3(1):120139, 2013.

[8] Bernhard Knapp, Samuel Demharter, Reyhaneh Esmaielbeiki, and Charlotte M Deane. Current status and future challenges in t-cell receptor/peptide/mhc molecular dynamics simulations. *Briefings in bioinformatics*, page bbv005, 2015.

[9] Anders S Christensen, Tomás Kubar, Qiang Cui, and Marcus Elstner. Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. *Chemical reviews*, 116(9):5301–5337, 2016.

[10] James J. P. Stewart. Optimization of parameters for semiempirical methods vi: more modifications to the nddo approximations and re-optimization of parameters. *J Mol Model*, 19:1–32, 2013.

[11] M Elstner. The scc-dftb method and its application to biological systems. *Theoretical Chemistry Accounts*, 116(1-3):316–325, 2006.

[12] Yoshio Nishimoto, Dmitri G. Fedorov, and Stephan Irle. Density-

functional tight-binding combined with the fragment molecular orbital method. *J. Chem. Theory Comput.*, 10:4801–4812, 2014.

[13] Kazuo Kitaura, Eiji Ikeo, Toshio Asada, Tatsuya Nakano, and Masami Uebayasi. Fragment molecular orbital method: an approximate computational method for large molecules. *Chemical Physics Letters*, 313, 1999.

[14] D. G. Fedorov, T. Nagata, and K. Kitaura. Exploring chemistry with the fragment molecular orbital method. *Phys. Chem. Chem. Phys*, 14, 2012.

[15] Caterina Barillari, Justine Taylor, Russell Viner, and Jonathan W Essex. Classification of water molecules in protein binding sites. *Journal of the American Chemical Society*, 129(9):2577–2587, 2007.

[16] Paula M. Petrone and Angel E. Garcia. Mhc-peptide binding is assisted by bound water molecules. *Journal of Molecular Biology*, 338:0–435, 2004.

[17] Lawrence J Stern, Jerry H Brown, Theodore S Jardetzky, Joan C Gorga, Robert G Urban, Jack L Strominger, and Don C Wiley. Crystal structure of the human class ii mhc protein hla-dr1 complexed with an influenza virus peptide. 1994.

[18] A Geluk, KE Van Meijgaarden, and TH Ottenhoff. Flexibility in t-cell

receptor ligand repertoires depends on mhc and t-cell receptor clonotype. *Immunology*, 90(3):370, 1997.

[19] Kathrine J Smith, Jason Pyrdol, Laurent Gauthier, Don C Wiley, and Kai W Wucherpfennig. Crystal structure of hla-dr2 (dra* 0101, drb1* 1501) complexed with a peptide from human myelin basic protein. *The Journal of experimental medicine*, 188(8):1511–1520, 1998.

[20] Michelle Krogsgaard, Kai W Wucherpfennig, Barbara Canella, Bjarke E Hansen, Arne Svejgaard, Jason Pyrdol, Henrik Ditzel, Cedric Raine, Jan Engberg, and Lars Fugger. Visualization of myelin basic protein (mbp) t cell epitopes in multiple sclerosis lesions using a monoclonal antibody specific for the human histocompatibility leukocyte antigen (hla)-dr2–mbp 85–99 complex. *The Journal of experimental medicine*, 191(8):1395–1412, 2000.

[21] James Robinson, Jason A Halliwell, James D Hayhurst, Paul Flicek, Peter Parham, and Steven GE Marsh. The ipd and imgt/hla database: allele variant databases. *Nucleic acids research*, page gku1161, 2014.

[22] Nicolas Rapin, Ilka Hoof, Ole Lund, and Morten Nielsen. Mhc motif viewer. *Immunogenetics*, 60(12):759–765, 2008.

[23] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. Ucsf

chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612, 2004.

[24] Roland L Dunbrack. Rotamer libraries in the 21 st century. *Current opinion in structural biology*, 12(4):431–440, 2002.

[25] James J. P. Stewart. Mopac2016. *Stewart Computational Chemistry*, Version 7.263W, 2016.

[26] Alexander Heifetz, Giancarlo Trani, Matteo Aldeghi, Colin H MacKinnon, Paul A McEwan, Frederick A Brookfield, Ewa I Chudyk, Mike Bodkin, Zhonghua Pei, Jason D Burch, et al. Fragment molecular orbital method applied to lead optimization of novel interleukin-2 inducible t-cell kinase (itk) inhibitors. *Journal of medicinal chemistry*, 59(9):4352–4363, 2016.

[27] Alexeev Yuri, P. Mazanetz Michael, Ichihara Osamu, and G. Fedorov Dmitri. Gamess as a free quantum-mechanical platform for drug research. *Current Topics in Medicinal Chemistry*, 12, 2012.

[28] Michael W. Schmidt, Kim K. Baldridge, Jerry A. Boatz, Steven T. Elbert, Mark S. Gordon, Jan H. Jensen, Shiro Koseki, Nikita Matsunaga, Kiet A. Nguyen, Shujun Su, Theresa L. Windus, Michel Dupuis, and John A. Montgomery Jr. General atomic and molecular electronic structure system. *Journal of Computational Chemistry*, 14, 1993.

[29] Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, and Markus Müller. proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC bioinformatics*, 12(1):1, 2011.

[30] Yuanchao Li, Yadong Yang, Ping He, and Qingwu Yang. Qm/mm study of epitope peptides binding to hla-a*0201: The roles of anchor residues and water. *Chemical Biology & Drug Design*, 74:611–618, 2009.

[31] Andrea Ferrante; Jack Gorski. Enthalpy–entropy compensation and co-operativity as thermodynamic epiphenomena of structural flexibility in ligand–receptor interactions. *Journal of Molecular Biology*, 417, 2012.

[32] Dmitri G. Fedorov and Kazuo Kitaura. Subsystem analysis for the fragment molecular orbital method and its application to protein-ligand binding in solution. *Journal of Physical Chemistry A*, 120, 2016.

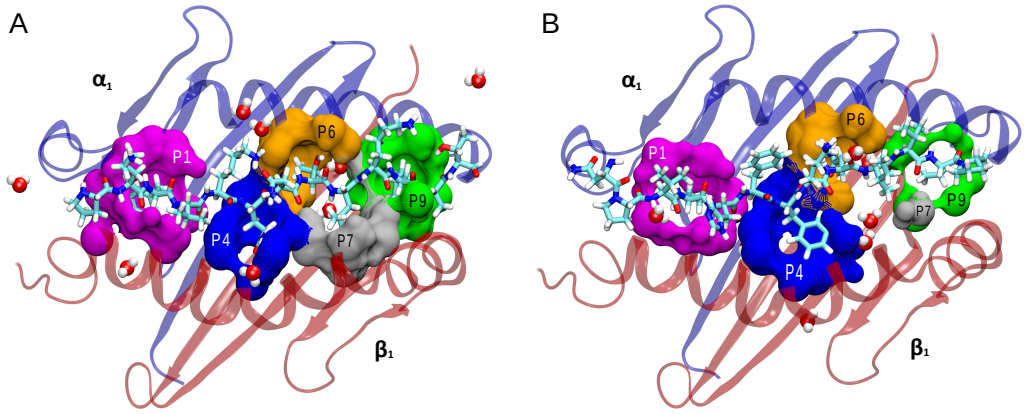
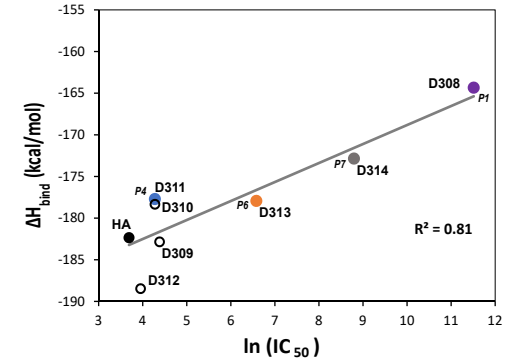


Figure 1: Top view of the A. HLA DR1 (1DLH) and B. HLA DR2 (1BX2) PBR, including water molecules in a range of 8 Å from each peptide. Peptides and water molecules are depicted in a ball & stick model and coloured by atoms (C: cyan, H: white, O: red, N: blue). α_1 (blue) and β_1 (red) domains are shown as cartoons. Pockets, showed here as receptor contact atoms in a range of 3.5 Å from peptide anchor residues, are represented as surfaces. P1 (magenta), P4 (blue), P6 (orange), P7 (grey), P9 (green).

A

Peptide	Sequence (1 4 6 7 9)	PM7/COSMO	FMO-DFTB	IC_{50}
HA $_{306-318}$	PK<u>YVKQNTLKLAT</u>	-182.35	-181.54	40.0
HA D308	PK<u>YVKQNTLKLAT</u>	-164.34	-159.02	100000.0
HA D309	PK<u>YDKQNTLKLAT</u>	-182.85	-181.23	80.0
HA D310	PK<u>YVDQNTLKLAT</u>	-178.33	-180.99	72.0
HA D311	PK<u>YVKDNTLKLAT</u>	-177.71	-174.42	72.0
HA D312	PK<u>YVKQDTLKLAT</u>	-188.47	-184.30	52.0
HA D313	PK<u>YVKQNDLKLAT</u>	-177.93	-174.62	720.0
HA D314	PK<u>YVKQNTDKLAT</u>	-172.85	-174.32	6600.0
R^2		0.81	0.81	
AUC (50/500 nM)		0.71/0.93	0.86/0.93	

B



C

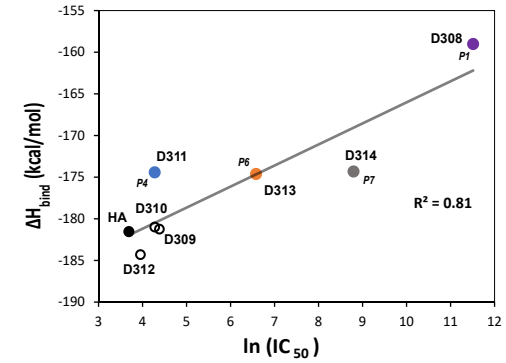
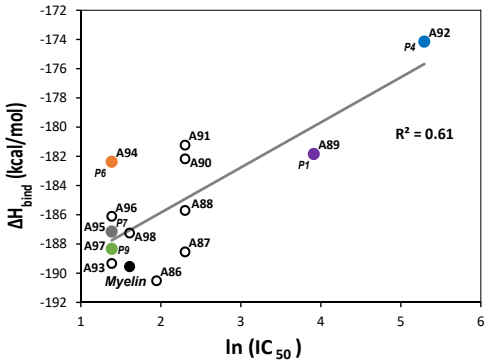


Figure 2: HLA-DR1/HA and mono-substituted analogue (Asp) set. A. Values of experimentally measured affinity (IC_{50}, nM) along with binding energies ΔH_{bind}, kcal/mol), coefficient of determination (R^2) and ROC AUC (50 and 500 nM cutoff) for each method evaluated. Binding cores are underlined. B. Correlation plot between ln of IC_{50} and binding energies calculated using the PM7 method. C. Correlation plot between ln of IC_{50} and binding energies calculated using the FMO-DFTB method. Substitutions in anchor residues are represented by colours: P1 (magenta), P4 (blue), P6 (orange), P7 (grey), and P9 (green).

A

Peptide	Sequence (1 4 6 7 9)	PM7/COSMO	FMO-DFTB	IC_{50}
MY $_{86\text{-}98}$	NPV<u>VHFFKNIVTP</u>	-189.55	-215.71	5.0
MY A86	APV<u>VHFFKNIVTP</u>	-190.52	-210.39	7.0
MY A87	NAV<u>VHFFKNIVTP</u>	-188.54	-211.70	10.0
MY A88	NPA<u>VHFFKNIVTP</u>	-185.70	-211.94	10.0
MY A89	NPV<u>AHFFKNIVTP</u>	-181.84	-207.52	50.0
MY A90	NPV<u>VAFFKNIVTP</u>	-182.17	-206.45	10.0
MY A91	NPV<u>VHAFKNIVTP</u>	-181.23	-207.44	10.0
MY A92	NPV<u>VHFAKNIVTP</u>	-174.15	-199.13	199.0
MY A93	NPV<u>VHFFANIVTP</u>	-189.33	-214.87	4.0
MY A94	NPV<u>VHFFKAIVTP</u>	-182.37	-209.99	4.0
MY A95	NPV<u>VHFFKNAVTP</u>	-187.15	-212.30	4.0
MY A96	NPV<u>VHFFKNIATP</u>	-186.10	-212.96	4.0
MY A97	NPV<u>VHFFKNIVAP</u>	-188.33	-213.87	4.0
MY A98	NPV<u>VHFFKNIVTA</u>	-187.25	-213.50	5.0
R^2		0.61	0.75	
AUC (5/50 nM)		0.74/1.0	0.94/1.0	

B



C

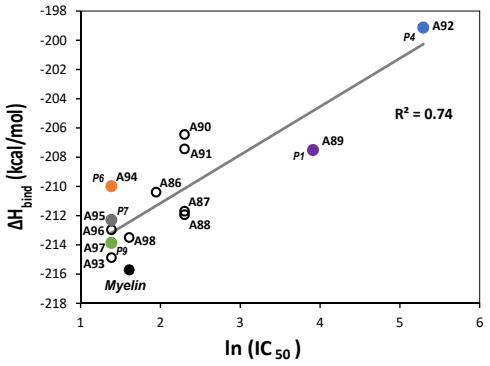


Figure 3: HLA-DR2/myelin & mono-substituted analogues (Ala) set. A. Values of experimentally measured affinity (IC_{50}, nM) along with binding energies (ΔH_{bind}, kcal/mol), coefficient of determination (R^2) and ROC AUC (5 and 50 nM cutoff) for each method evaluated. Binding cores are underlined. B. Correlation plot between ln of IC_{50} and binding energies calculated using the PM7 method. C. Correlation plot between Ln of IC_{50} and binding energies calculated using the FMO-DFTB method. Substitutions in anchor residues are represented by colours: P1 (magenta), P4 (blue), P6 (orange), P7 (grey), and P9 (green).

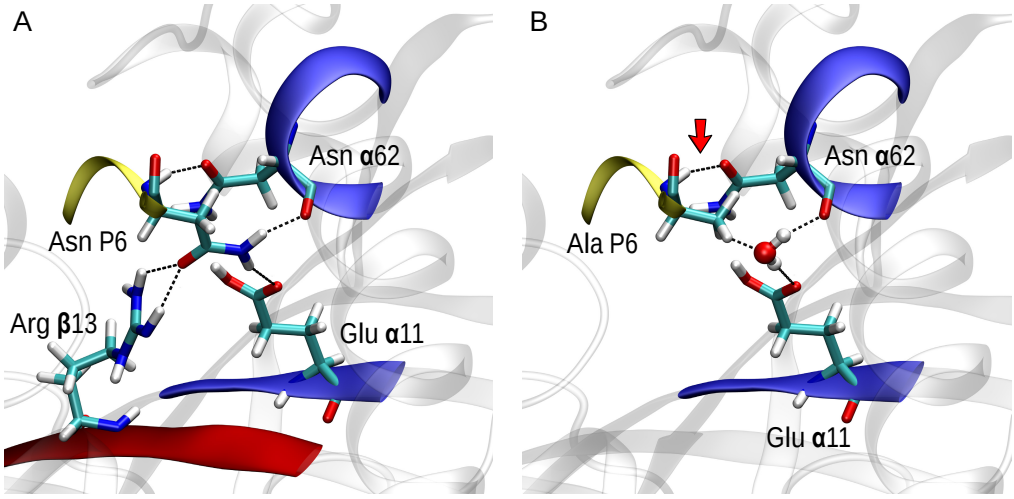
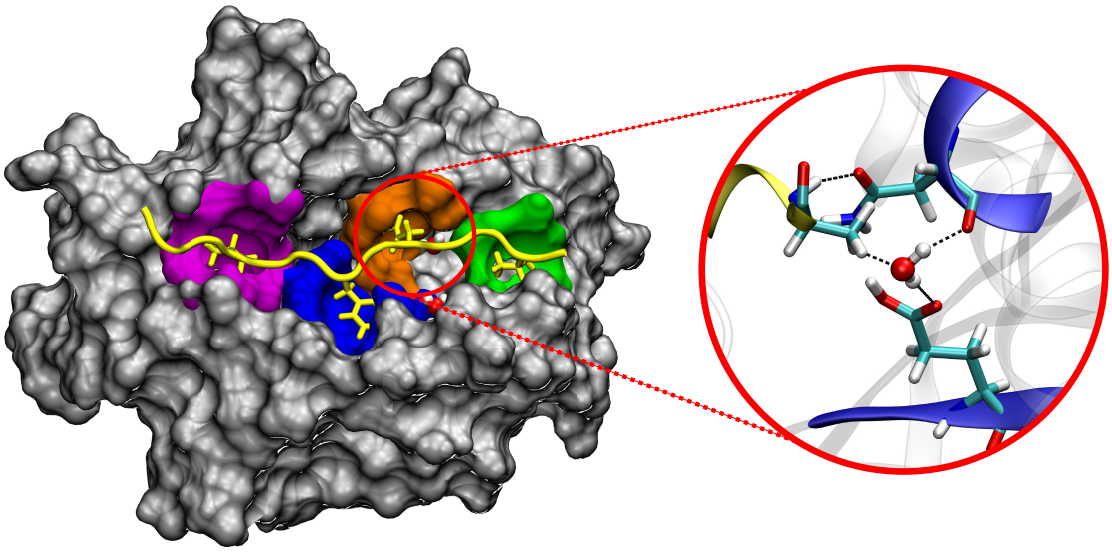


Figure 4: Hydrogen-bonding network in the P6-binding site for: A. native myelin$_{86-98}$ peptide (Asn 94) and B. mono-substituted analogue (Ala 94). Interacting HLA-DR2 and P6 residues, including a water molecule that stabilises binding in the analogue peptide, are shown by a ball & stick model and coloured by atoms (C: cyan, H: white, O: red, N: blue). Peptide (yellow), α_1 (blue) and β_1 (red) domains are shown as cartoons. Hydrogen-bond distances are between 1.8 to 2.2 Å . A red arrow indicates the only H-bond peptide formed without the addition of a water molecule in pocket 6

.



Graphical Abstract

Conclusiones generales

Aumentar el conocimiento sobre la biología de los primates no humanos, tiene un impacto directo en la mejora de la salud humana por medio de la investigación científica. Dada la estrecha relación evolutiva e identidad biológica (genética, anatómica y fisiológica) entre todos los primates -incluyendo a los seres humanos-, éstos son referentes obligados en el campo de la biología comparada y en la investigación biomédica. Siguiendo este planteamiento, este trabajo ha contribuido a la caracterización de las moléculas del complejo mayor de histocompatibilidad los monos *Aotus*, buscando estimar y analizar su polimorfismo. Los aportes realizados, si bien tienen como objeto contribuir al desarrollo de vacunas, también implican una contribución a aspectos más básicos de la biología del CMH en primates y de la evolución de estas proteínas. Como resultado, se han estudiado por primera vez los loci CMH-DPA y CMH-DRA de *Aotus* y se profundizó en el estudio del CMH-DRB, analizando los modos de evolución de estos genes y proponiendo estrategias para manejar su polimorfismo.

Desde el punto de vista experimental, se realizó análisis de un microsatélite del CMH-DRB que puede constituirse en un sensible método de tipificación. Desde el punto de vista computacional, se diseñaron y aplicaron estrategias para manejar el polimorfismo del CMH-DRB tanto en humanos como en *Aotus*, con el fin de optimizar el proceso de diseño de péptidos modificados como candidatos a vacuna, su evaluación en el modelo animal y se brinda una estrategia para estimar su cubrimiento potencial en poblaciones humanas.

Adicionalmente, se implementaron protocolos computacionales para modelar la unión CMH-péptido, usando estrategias basadas en redes neurales y se desarrollaron protocolos basados en métodos cuánticos semi-empíricos, que permiten un modelamiento más preciso y detallado de este proceso.

En la búsqueda de una escala de similitud estructural para los aminoácidos, se encontró una relación entre las tendencias de estructura secundaria, masa y los patrones de

sustitución y mutabilidad de los aminoácidos, mostrando alta correlación con matrices de sustitución como las BLOSUM. Esta relación es inédita y muestra cómo los procesos históricos que gobiernan evolución de las proteínas tienen un contrapunto con las propiedades estructurales de los aminoácidos.

Esta investigación parte de un enfoque multidisciplinario que trata con el problema central la unión de péptidos al CMH. La evolución de estas secuencias puede considerarse como un experimento, en donde la selección natural ha probado múltiples soluciones, y se han mantenido aquellas que resultan adecuadas (aunque sin garantía que sean las mejores). El análisis de estos patrones en busca de identificar cuales propiedades fisicoquímicas describen este proceso, nos muestra una perspectiva valiosa, señalando que la búsqueda de explicaciones que incorporen, tanto información evolutiva como fisicoquímica, es clave para la comprensión de este complejo proceso.

Perspectivas y recomendaciones

El desarrollo de métodos para modelar los procesos de interacción proteína - proteína (como la interacción CMH-péptido) es uno de los campos de enorme interés para comprender las funciones de las proteínas, y son clave para estudiar procesos como metabolismo celular, transducción de señales, y reconocimiento molecular, entre otros. Los enfoques propuestos no solamente tienen aplicación al campo concreto del estudio del CMH en *Aotus* y Humanos, sino que tienen el potencial de aplicarse a problemas similares en otros sistemas.

Las metodologías desarrolladas permitirán caracterizar con gran detalle la interacción CMH-péptido, siendo especialmente promisorio el uso de FMO-PIEDA en el estudio de residuos claves en la región de unión al péptido (bien sea por su conservación y variabilidad), lo que permitirá una visión de los factores fisicoquímicos que determinan los procesos selectivos y los patrones de variabilidad en el CMH.

Las metodologías de modelamiento de la unión CMH-péptido propuestas, permitirán evaluar computacionalmente los perfiles de unión de moléculas de interés, para lo cual se pueden usar modelos estructurales generados por homología. Esto es de especial interés, dado el grado de dificultad que implica el establecimiento de datos de unión en húmedo.

Usando estrategias similares, se puede generalizar la metodología propuesta para otros loci de CMH clase I y CMH clase II, con interés biomédico para otras patologías.

A partir de la minería de datos sobre información cristalográfica, se adelantará el análisis de los patrones de secuencia relacionados con estructuras secundarias estables (hélice alfa, beta extendidas y hélice de PP_{II}), con el fin de completar un marco para el diseño de péptidos basados en parámetros estructurales.

Referencias

1. Julian K. Professor Julian C Knight - Nuffield Department of Medicine https://www.ndm.ox.ac.uk/principal-investigators/researcher/julian-knight: Nuffield Department of Medicine, University of Oxford; 2017 (08/11/2017)
2. Neefjes J, Ovaa H. A peptide's perspective on antigen presentation to the immune system. Nature chemical biology. 2013;9(12):769-75.
3. Hershkovitz P. Two new species of night monkeys, genus Aotus (Cebidae: Platyrrhini): A preliminary report on Aotus taxonomy. Am J Primatol. 1983;4:209–43.
4. Torres O, Enciso S, Ruiz F, Silva E, Yunis I. Chromosome diversity of the genus Aotus from Colombia. Am J Primatol. 1998;44:255–75.
5. Fernandez-Duque E. Primates in Perspective. New York: Oxford University Press; 2007. p. 139–54.
6. Defler T, Bueno M. Aotus diversity and the species problem. Primate Conservation. 2007; 22: 55-70.
7. Defler T. Historia Natural de los Primates Colombianos. Bogotá D.C.: Universidad National de Colombia; 2010.
8. Setoguchi T, Rosenberger AL. A fossil owl monkey from La Venta, Colombia. Nature. 1987;326(6114):692-4.
9. Takai M, Nishimura T, Shigehara N, Setoguchi T. Meaning of the canine sexual dimorphism in fossil owl monkey, Aotus dindensis from the middle Miocene of La Venta, Colombia. Front Oral Biol. 2009;13:55-9.
10. Perelman P, Johnson WE, Roos C, Seuanez HN, Horvath JE, Moreira MA, et al. A molecular phylogeny of living primates. PLoS Genet. 2011;7(3):e1001342.
11. Finstermeier K, Zinner D, Brameier M, Meyer M, Kreuz E, Hofreiter M, et al. A mitogenomic phylogeny of living primates. PLoS One. 2013;8(7):e69504.
12. Menezes AN, Bonvicino CR, Seuanez HN. Identification, classification and evolution of owl monkeys (Aotus, Illiger 1811). BMC Evol Biol. 2010;10:248.
13. Aquino R, Encarnación F. Characteristics and use of sleeping site in Aotus (Cebidae: Primates) in the Amazonian lowland of Perú. Am J Primatol. 1986;11:319-31.
14. Aquino R, Encarnación F. Population densities and geographic distribution of night monkeys (Aotus nancymai and Aotus vociferans) (Cebidae: Primates) in Northeastern Perú. American Journal of Primatology. 1988;14:375–81.
15. Aquino R, Encarnación F. Aotus: The Owl Monkey. San Diego: Academic Press; 1994. p. 59–95.
16. Fernandez-Duque E, Rotundo M, Sloan C. Density and population structure of owl monkeys (Aotus azarai) in the Argentinean Chaco. Am J Primatol. 2001;53:99–108.
17. Chapman A, Chapman J. Implications of Small Scale Variation in Ecological Conditions for the Diet and Density of Red Colobus Monkeys. Primates. 1999; 40: 215-31.
18. Ankel-Simons F, Rasmussen DT. Diurnality, nocturnality, and the evolution of primate visual systems. Am J Phys Anthropol. 2008;Suppl 47:100-17.
19. Hernández A, Díaz A. Estado preliminar poblacional del mono nocturno (Aotus sp. Humboldt, 1812) en las comunidades Indígenas Siete de Agosto y San Juan de Atacuari- Puerto Nariño, Departamento de Amazonas, Colombia. Ibagué, Colombia.: Universidad del Tolima; 2011.
20. Bontrop R. Non-human primates: essential partners in biomedical research. Immunol Rev. 2001;183:5-9.
21. Langhorne J, Buffet P, Galinski M, Good M, Harty J, Leroy D, et al. The relevance of non-human primate and rodent malaria models for humans. Malar J. 2011;10(1):23.
22. Ward JM, Vallender EJ. The resurgence and genetic implications of New World primates in biomedical research. Trends Genet. 2012;28(12):586-91.
23. Rodriguez LE, Curtidor H, Urquiza M, Cifuentes G, Reyes C, Patarroyo ME. Intimate molecular interactions of P. falciparum merozoite proteins involved in invasion of red blood cells and their implications for vaccine design. Chem Rev. 2008;108(9):3656-705.

24. Patarroyo ME, Bermudez A, Patarroyo MA. Structural and immunological principles leading to chemically synthesized, multiantigenic, multistage, minimal subunit-based vaccine development. Chem Rev. 2011;111(5):3459-507.
25. Young MD, Porter JA, Jr., Johnson CM. Plasmodium vivax transmitted from man to monkey to man. Science. 1966;153(3739):1006-7.
26. Contacos PG, Collins WE. Falciparum malaria transmissible from monkey to man by mosquito bite. Science. 1968;161(3836):56-.
27. Gysin J. Malaria: parasite biology, pathogenesis and protection. Washington DC.: ASM.; 1988. p. 419–39.
28. Lujan R, Dennis V, Chapman WJ, Hanson W. Blastogenic responses of peripheral blood leukocytes from owl monkeys experimentally infected with Leishmania braziliensis panamensis. Am J Trop Med Hyg. 1986;35(6):1103-9.
29. Pico de Coaña Y, Rodriguez J, Guerrero E, Barrero C, Rodriguez R, Mendoza M, et al. A highly infective Plasmodium vivax strain adapted to Aotus monkeys: quantitative haematological and molecular determinations useful for P. vivaxmalaria vaccine development. Vaccine. 2003;21:3930–7.
30. Polotsky Y, Vassell R, Binn L, Asher L. Immunohistochemical detection of cytokines in tissues of Aotus monkeys infected with hepatitis A virus. Ann N Y AcadSci. 1994;730:318–21.
31. Noya O, Gonzalez-Rico S, Rodriguez R, Arrechedera H, Patarroyo M, Alarcon D. Schistosomamansoniinfection in owl monkeys (Aotus nancymai): evidence for the early elimination of adult worms. Acta Trop. 1998;70:257–67.
32. Bone J, Soave O. Experimental tuberculosis in owl monkeys (Aotus trivirgatus). Lab Anim Care. 1970;5(946-8).
33. Jones F, Baqar S, Gozalo A, Nunez G, Espinoza N, Reyes S, et al. New World monkey Aotus nancymae as a model for Campylobacter jejuni infection and immunity. Infect Immun. 2006;74(1):790-3.
34. Ding Y, Casagrande V. The distribution and morphology of LGN K pathway axons within the layers and CO blobs of owl monkey V1. Vis Neurosci. 1997;14:691-704.
35. Cadavid LF, Lun CM. Lineage-specific diversification of killer cell Ig-like receptors in the owl monkey, a New World primate. Immunogenetics. 2009;61(1):27-41.
36. Castillo F, Guerrero C, Trujillo E, Delgado G, Martinez P, Salazar LM, et al. Identifying and structurally characterizing CD1b in Aotus nancymaae owl monkeys. Immunogenetics. 2004;56(7):480-9.
37. del Castillo H, Vernot JP. Characterizing the CD3 epsilon chain from the New World primate Aotus nancymaae. Biomedica. 2008;28(2):262-70.
38. Montoya GE, Vernot JP, Patarroyo ME. Partial characterization of the CD45 phosphatase cDNA in the owl monkey (Aotus vociferans). Am J Primatol. 2002;57(1):1-11.
39. Montoya GE, Vernot JP, Patarroyo ME. Comparative analysis of CD45 proteins in primate context: owl monkeys vs humans. Tissue Antigens. 2004;64(2):165-72.
40. Diaz OL, Daubenberger CA, Rodriguez R, Naegeli M, Moreno A, Patarroyo ME, et al. Immunoglobulin kappa light-chain V, J, and C gene sequences of the owl monkey Aotus nancymaae. Immunogenetics. 2000;51(3):212-8.
41. Hernandez EC, Suarez CF, Parra CA, Patarroyo MA, Patarroyo ME. Identification of five different IGHV gene families in owl monkeys (Aotus nancymaae). Tissue Antigens. 2005;66(6):640-9.
42. Favre N, Daubenberger C, Marfurt J, Moreno A, Patarroyo M, Pluschke G. Sequence and diversity of T-cell receptor alpha V, J, and C genes of the owl monkey Aotus nancymaae. Immunogenetics. 1998;48(4):253-9.
43. Guerrero JE, Pacheco DP, Suarez CF, Martinez P, Aristizabal F, Moncada CA, et al. Characterizing T-cell receptor gamma-variable gene in Aotus nancymaae owl monkey peripheral blood. Tissue Antigens. 2003;62(6):472-82.
44. Moncada CA, Guerrero E, Cardenas P, Suarez CF, Patarroyo ME, Patarroyo MA. The T-cell receptor in primates: identifying and sequencing new owl monkey TRBV gene sub-groups. Immunogenetics. 2005;57(1-2):42-52.
45. Hernandez EC, Suarez CF, Mendez JA, Echeverry SJ, Murillo LA, Patarroyo ME. Identification, cloning, and sequencing of different cytokine genes in four species of owl monkey. Immunogenetics. 2002;54(9):645-53.

46. Spirig R, Peduzzi E, Patarroyo ME, Pluschke G, Daubenberger CA. Structural and functional characterisation of the Toll like receptor 9 of Aotus nancymaae, a non-human primate model for malaria vaccine development. Immunogenetics. 2005;57(3-4):283-8.
47. Delgado G, Parra C, Patarroyo M. Phenotypical and functional characterization of non-human primate Aotus spp. dendritic cells and their use as a tool for characterizing immune response to protein antigens. Vaccine. 2005;23(26):3386-95.
48. Daubenberger CA, Salomon M, Vecino W, Hubner B, Troll H, Rodriques R, et al. Functional and structural similarity of V gamma 9V delta 2 T cells in humans and Aotus monkeys, a primate infection model for Plasmodium falciparum malaria. J Immunol. 2001;167(11):6421-30.
49. Pinzon-Charry A, Vernot JP, Rodriguez R, Patarroyo ME. Proliferative response of peripheral blood lymphocytes to mitogens in the owl monkey Aotus nancymae. J Med Primatol. 2003;32(1):31-8.
50. Daubenberger CA, Spirig R, Patarroyo ME, Pluschke G. Flow cytometric analysis on cross-reactivity of human-specific CD monoclonal antibodies with splenocytes of Aotus nancymaae, a non-human primate model for biomedical research. Vet Immunol Immunopathol. 2007;119(1-2):14-20.
51. Glass EJ. Genetic variation and responses to vaccines. Anim Health Res Rev. 2004;5(2):197-208.
52. Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc Biol Sci. 2010;277(1684):979-88.
53. Suarez CF, Cardenas PP, Llanos-Ballestas EJ, Martinez P, Obregon M, Patarroyo ME, et al. alpha(1) and alpha(2) domains of Aotus MHC Class I and Catarrhini MHC class Ia share similar characteristics. Tissue Antigens. 2003;61(5):362-73.
54. Cardenas PP, Suarez CF, Martinez P, Patarroyo ME, Patarroyo MA. MHC class I genes in the owl monkey: mosaic organisation, convergence and loci diversity. Immunogenetics. 2005;56(11):818-32.
55. Cadavid LF, Shufflebotham C, Ruiz FJ, Yeager M, Hughes AL, Watkins DI. Evolutionary instability of the major histocompatibility complex class I loci in New World primates. Proceedings of the National Academy of Sciences of the United States of America. 1997;94(26):14536-41.
56. Nino-Vasquez JJ, Vogel D, Rodriguez R, Moreno A, Patarroyo ME, Pluschke G, et al. Sequence and diversity of DRB genes of Aotus nancymaae, a primate model for human malaria parasites. Immunogenetics. 2000;51(3):219-30.
57. Patarroyo ME, Cifuentes G, Baquero J. Comparative molecular and three-dimensional analysis of the peptide-MHC II binding region in both human and Aotus MHC-DRB molecules confirms their usefulness in antimalarial vaccine development. Immunogenetics. 2006;58(7):598-606.
58. Diaz D, Naegeli M, Rodriguez R, Nino-Vasquez JJ, Moreno A, Patarroyo ME, et al. Sequence and diversity of MHC DQA and DQB genes of the owl monkey Aotus nancymaae. Immunogenetics. 2000;51(7):528-37.
59. Diaz D, Daubenberger CA, Zalac T, Rodriguez R, Patarroyo ME. Sequence and expression of MHC-DPB1 molecules of the New World monkey Aotus nancymaae, a primate model for Plasmodium falciparum. Immunogenetics. 2002;54(4):251-9.
60. Suarez CF, Patarroyo ME, Trujillo E, Estupinan M, Baquero JE, Parra C, et al. Owl monkey MHC-DRB exon 2 reveals high similarity with several HLA-DRB lineages. Immunogenetics. 2006;58(7):542-58.
61. Suarez CF, Patarroyo MA, Patarroyo ME. Characterisation and comparative analysis of MHC-DPA1 exon 2 in the owl monkey (Aotus nancymaae). Gene. 2011;470(1-2):37-45.
62. Lopez C, Suarez CF, Cadavid LF, Patarroyo ME, Patarroyo MA. Characterising a microsatellite for DRB typing in Aotus vociferans and Aotus nancymaae (Platyrrhini). PLoS One. 2014;9(5):e96973.
63. Baquero JE, Miranda S, Murillo O, Mateus H, Trujillo E, Suarez C, et al. Reference strand conformational analysis (RSCA) is a valuable tool in identifying MHC-DRB sequences in three species of Aotus monkeys. Immunogenetics. 2006;58(7):590-7.
64. Suárez CF, Pabón L, Barrera A, Aza-Conde J, Patarroyo MA, Patarroyo ME. Structural analysis of owl monkey MHC-DR shows that fully-protective malaria vaccine components can be readily used in humans. Biochemical and Biophysical Research Communications. 2017.
65. Stephens R, Horton R, Humphray S, Rowen L. Gene organisation, sequence variation and isochore structure at the centromeric boundary of the human MHC. J Mol Biol. 1999;291:789-99.
66. Watanabe A, Shiina T, Shimizu S, Hosomichi K, Yanagiya K, Kita Y, et al. A BAC-based contig map of the cynomolgus macaque (Macaca fascicularis) major histocompatibility complex genomic region. Genomics. 2007;89(3):402-12.

67. Tregenza T, Wedell N. Genetic compatibility mate choice and patterns of parentage. Invited Review Mol Ecol. 2000;9:1013-27.
68. Hughes A, Hughes M. Natural selection on the peptide-binding regions of major histocompatibility complex molecules. Immunogenetics. 1995;42:233-43.
69. Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool. 2005;2(16:1–16:18).
70. Robinson J, Halliwell JA, McWilliam H, Lopez R, Parham P, Marsh SG. The IMGT/HLA database. Nucleic Acids Res. 2013;41(Database issue):D1222-7.
71. Sutton JT, Nakagawa S, Robertson BC, Jamieson IG. Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. Mol Ecol. 2011;20(21):4408-20.
72. Yeager M, Hughes AL. Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunol Rev. 1999;167:45-58.
73. Hughes AL, Yeager M. Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet. 1998;32:415-35.
74. Hedrick PW. Pathogen resistance and genetic variation at MHC loci. Evolution. 2002;56(10):1902-8.
75. Potts WK, Wakeland EK. Evolution of MHC genetic diversity: a tale of incest, pestilence and sexual preference. Trends Genet. 1993;9(12):408-12.
76. Worley K, Collet J, Spurgin LG, Cornwallis C, Pizzari T, Richardson DS. MHC heterozygosity and survival in red junglefowl. Mol Ecol. 2010;19(15):3064-75.
77. Ejsmond MJ, Babik W, Radwan J. MHC allele frequency distributions under parasite-driven selection: A simulation model. BMC Evol Biol. 2010;10:332.
78. Apanius V, Penn D, Slev PR, Ruff LR, Potts WK. The nature of selection on the major histocompatibility complex. Crit Rev Immunol. 1997;17(2):179-224.
79. Potts WK, Slev PR. Pathogen-based models favoring MHC genetic diversity. Immunol Rev. 1995;143:181-97.
80. Borghans JA, Beltman JB, De Boer RJ. MHC polymorphism under host-pathogen coevolution. Immunogenetics. 2004;55(11):732-9.
81. Potts WK, Manning CJ, Wakeland EK. The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. Philos Trans R Soc Lond B Biol Sci. 1994;346(1317):369-78.
82. Jordan WC, Bruford MW. New perspectives on mate choice and the MHC. Heredity. 1998;81 (Pt 2):127-33.
83. Huchard E, Raymond M, Benavides J, Marshall H, Knapp LA, Cowlishaw G. A female signal reflects MHC genotype in a social primate. BMC Evol Biol. 2010;10:96.
84. Huchard E, Knapp LA, Wang J, Raymond M, Cowlishaw G. MHC, mate choice and heterozygote advantage in a wild social primate. Mol Ecol. 2010;19(12):2545-61.
85. Setchell JM, Huchard E. The hidden benefits of sex: evidence for MHC-associated mate choice in primate societies. Bioessays. 2010;32(11):940-8.
86. Roberts SC, Little AC, Gosling LM, Jones BC, Perrett DI, Carter V, et al. MHC-assortative facial preferences in humans. Biol Lett. 2005;1(4):400-3.
87. Havlicek J, Roberts SC. MHC-correlated mate choice in humans: a review. Psychoneuroendocrinology. 2009;34(4):497-512.
88. Manning CJ, Wakeland EK, Potts WK. Communal nesting patterns in mice implicate MHC genes in kin recognition. Nature. 1992;360(6404):581-3.
89. Yamazaki K, Beauchamp GK. Genetic basis for MHC-dependent mate choice. Adv Genet. 2007;59:129-45.
90. Wedekind C, Chapuisat M, Macas E, Rulicke T. Non-random fertilization in mice correlates with the MHC and something else. Heredity. 1996;77 (Pt 4):400-9.
91. Dorak MT, Lawson T, Machulla HK, Mills KI, Burnett AK. Increased heterozygosity for MHC class II lineages in newborn males. Genes Immun. 2002;3(5):263-9.
92. Klein J, Sato A, Nagl S, O'hUigín C. Molecular trans-species polymorphism. Annu Rev Ecol Syst. 1998;29:1-21.
93. Klein J, Sato A, Nikolaidis N. MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. Annu Rev Genet. 2007;41:281-304.

94. Klein J, Satta Y, Takahata N, O'HUigin C. Trans-specific Mhc polymorphism and the origin of species in primates. J Med Primatol. 1993;22(1):57-64.
95. Trtkova K, Mayer WE, O'Huigin C, Klein J. Mhc-DRB genes and the origin of New World monkeys. Molecular phylogenetics and evolution. 1995;4(4):408-19.
96. O'HUigin C. Quantifying the degree of convergence in primate Mhc-DRB genes. Immunol Rev. 1995;143:123-40.
97. Doxiadis GG, de Groot N, de Groot NG, Doxiadis, II, Bontrop RE. Reshuffling of ancient peptide binding motifs between HLA-DRB multigene family members: old wine served in new skins. Mol Immunol. 2008;45(10):2743-51.
98. Slierendregt BL, Otting N, Kenter M, Bontrop RE. Allelic diversity at the Mhc-DP locus in rhesus macaques (Macaca mulatta). Immunogenetics. 1995;41(1):29-37.
99. Bontrop RE, Otting N, de Groot NG, Doxiadis GG. Major histocompatibility complex class II polymorphisms in primates. Immunol Rev. 1999;167:339-50.
100. Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ, et al. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res. 2003;31(1):311-4.
101. Steiper M, Young N. Primate molecular divergence dates. Molecular phylogenetics and evolution. 2006;41:384–94.
102. Wang JH, Reinherz EL. Structural basis of T cell recognition of peptides bound to MHC molecules. Mol Immunol. 2002;38(14):1039-49.
103. Backert L, Kohlbacher O. Immunoinformatics and epitope prediction in the age of genomic medicine. Genome Med. 2015;7:119.
104. Lafuente EM, Reche PA. Prediction of MHC-peptide binding: a systematic and comprehensive overview. Curr Pharm Des. 2009;15(28):3209-20.
105. Lenz TL. Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. Evolution. 2011;65(8):2380-90.
106. Doytchinova IA, Flower DR. In silico identification of supertypes for class II MHCs. Journal of Immunology. 2005;174(11):7085-95.
107. Doytchinova IA, Guan P, Flower DR. Identifiying human MHC supertypes using bioinformatic methods. Journal of Immunology. 2004;172(7):4314-23.
108. Lund O, Nielsen M, Kesmir C, Petersen AG, Lundegaard C, Worning P, et al. Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics. 2004;55(12):797-810.
109. Schwensow N, Fietz J, Dausmann K, Sommer S. Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. Heredity. 2007;99(3):265-77.
110. Sepil I, Lachish S, Hinks AE, Sheldon BC. Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. Proceedings of the Royal Society of London B: Biological Sciences. 2013;280(1759):20130134.
111. Hill AV. Common West African HLA antigens are associated with protection from severe malaria. Nature. 1991;352(6336):595-600.
112. Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol. 2008;4(4):e1000048.
113. Sidney J, Southwood S, Moore C, Oseroff C, Pinilla C, Grey HM, et al. Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture. Curr Protoc Immunol. 2013;Chapter 18:Unit 18 3.
114. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol. 1999;17(6):555-61.
115. Zhang L, Chen Y, Wong HS, Zhou S, Mamitsuka H, Zhu S. TEPITOPEpan: extending TEPITOPE for peptide binding prediction covering over 700 HLA-DR molecules. PLoS One. 2012;7(2):e30483.
116. Rothbard JB, Taylor WR. A sequence pattern common to T cell epitopes. Embo J. 1988;7(1):93-100.

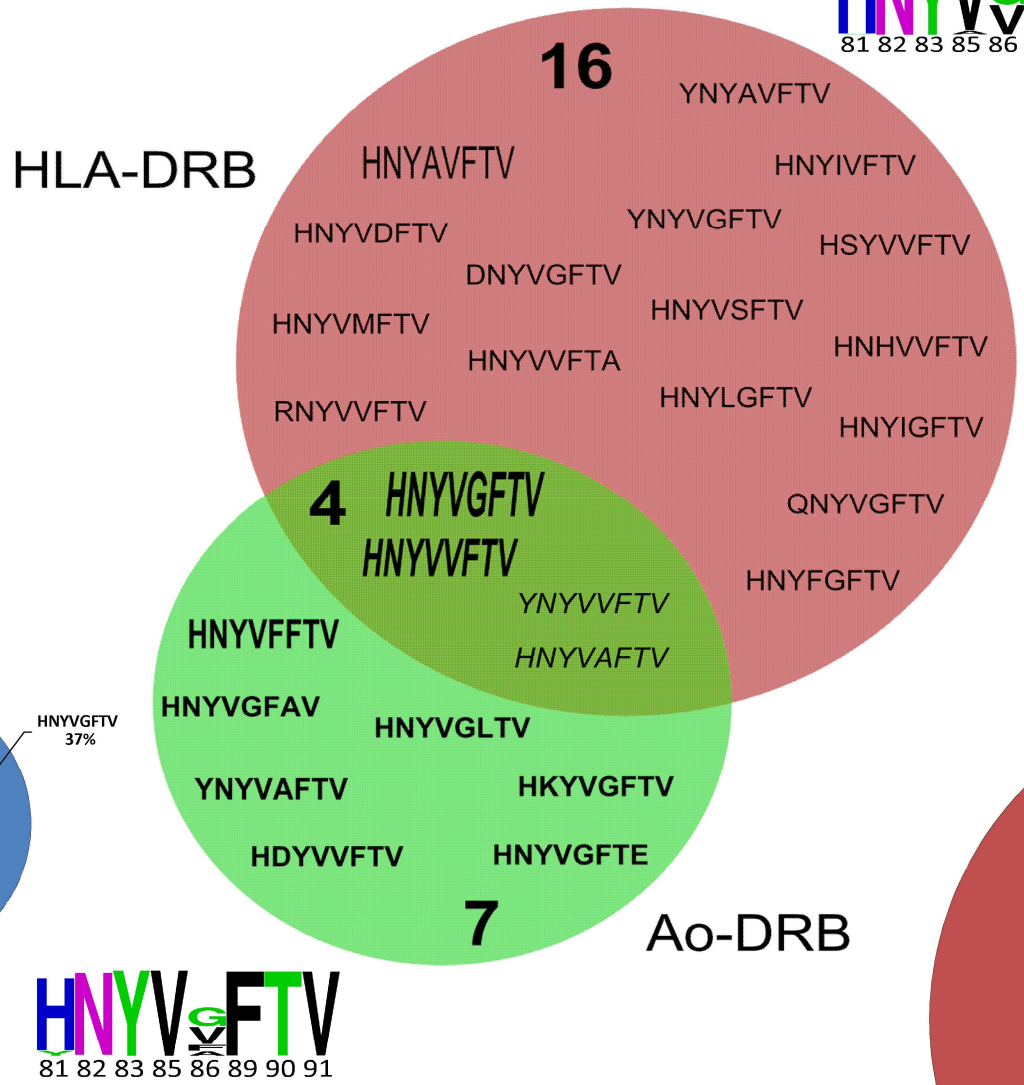
117. Udaka K, Wiesmuller KH, Kienle S, Jung G, Tamamura H, Yamagishi H, et al. An automated prediction of MHC class I-binding peptides based on positional scanning with peptide libraries. Immunogenetics. 2000;51(10):816-28.
118. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics. 2005;6:132.
119. Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, et al. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. Immunome Res. 2008;4:2.
120. Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinformatics. 2007;8:238.
121. Zhang W, Liu J, Niu Y. Quantitative prediction of MHC-II binding affinity using particle swarm optimization. Artif Intell Med. 2010;50(2):127-32.
122. Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. Immunogenetics. 2015;67(11-12):641-50.
123. Lundegaard C, Lund O, Nielsen M. Prediction of epitopes using neural network based methods. J Immunol Methods. 2011;374(1-2):26-34.
124. Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 2003;12(5):1007-17.
125. Roomp K, Antes I, Lengauer T. Predicting MHC class I epitopes in large datasets. BMC Bioinformatics. 2010;11:90.
126. Nielsen M, Justesen S, Lund O, Lundegaard C, Buus S. NetMHCIIpan-2.0 - Improved pan-specific HLA-DR predictions using a novel concurrent alignment and weight optimization training procedure. Immunome Res. 2010;6:9.
127. Noguchi H, Kato R, Hanai T, Matsubara Y, Honda H, Brusic V, et al. Hidden Markov model-based prediction of antigenic peptides that interact with MHC class II molecules. J Biosci Bioeng. 2002;94(3):264-70.
128. Nielsen M, Lund O, Buus S, Lundegaard C. MHC class II epitope predictive algorithms. Immunology. 2010;130(3):319-28.
129. Vider-Shalit T, Louzoun Y. MHC-I prediction using a combination of T cell epitopes and MHC-I binding peptides. J Immunol Methods. 2011;374(1-2):43-6.
130. Liu W, Meng X, Xu Q, Flower DR, Li T. Quantitative prediction of mouse class I MHC peptide binding affinity using support vector machine regression (SVR) models. BMC Bioinformatics. 2006;7:182.
131. Donnes P. Support vector machine-based prediction of MHC-binding peptides. Methods Mol Biol. 2007;409:273-82.
132. Agudelo W, Patarroyo M. Quantum chemical analysis of MHC-peptide interactions for vaccine design. Mini reviews in medicinal chemistry. 2010;10(8):746-58.
133. Wan S, Knapp B, Wright DW, Deane CM, Coveney PV. Rapid, Precise, and Reproducible Prediction of Peptide–MHC Binding Affinities from Molecular Dynamics That Correlate Well with Experiment. J Chem Theory Comput. 2015;11(7):3346-56.
134. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. Open Biol. 2013;3(1):120139.
135. Bordner AJ, Abagyan R. Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. Proteins. 2006;63(3):512-26.
136. Zhang H, Wang P, Papangelopoulos N, Xu Y, Sette A, Bourne PE, et al. Limitations of Ab initio predictions of peptide binding to MHC class II molecules. PLoS One. 2010;5(2):e9272.
137. Bordner AJ. Towards universal structure-based prediction of class II MHC epitopes for diverse allotypes. PLoS One. 2010;5(12):e14383.
138. Yanover C, Bradley P. Large-scale characterization of peptide-MHC binding landscapes with structural simulations. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(17):6981-6.
139. Knapp B, Omasits U, Schreiner W. Side chain substitution benchmark for peptide/MHC interaction. Protein Sci. 2008;17(6):977-82.

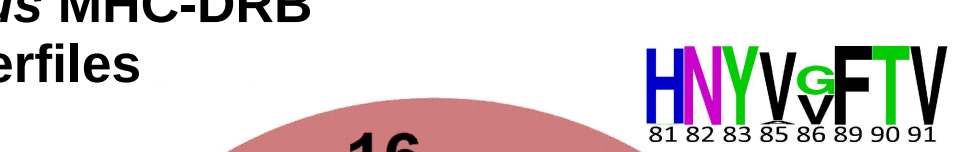
140. Tong JC, Tan TW, Ranganathan S. Modeling the structure of bound peptide ligands to major histocompatibility complex. Protein Sci. 2004;13(9):2523-32.
141. Bui HH, Schiewe AJ, von Grafenstein H, Haworth IS. Structural prediction of peptides binding to MHC class I molecules. Proteins-Structure Function and Genetics. 2006;63(1):43-52.
142. Cárdenas C, Ortiz M, Balbín A, Villaveces JL, Patarroyo ME. Allele effects in MHC–peptide interactions: A theoretical analysis of HLA-DRβ1* 0101-HA and HLA-DRβ1* 0401-HA complexes. Biochemical and biophysical research communications. 2005;330(4):1162-7.
143. Balbín A, Cárdenas C, Villaveces JL, Patarroyo ME. A theoretical analysis of HLA-DRβ1* 0301–CLIP complex using the first three multipolar moments of the electrostatic field. Biochimie. 2006;88(9):1307-11.
144. Bohorquez HJ, Obregon M, Cárdenas C, Llanos E, Suárez C, Villaveces JL, et al. Electronic energy and multipolar moments characterize amino acid side chains into chemically related groups. The Journal of Physical Chemistry A. 2003;107(47):10090-7.
145. Cárdenas C, Villaveces JL, Bohórquez H, Llanos E, Suárez C, Obregón M, et al. Quantum chemical analysis explains hemagglutinin peptide–MHC Class II molecule HLA-DRβ1* 0101 interactions. Biochemical and biophysical research communications. 2004;323(4):1265-77.
146. Cárdenas C, Villaveces JL, Suárez C, Obregón M, Ortiz M, Patarroyo ME. A comparative study of MHC Class-II HLA-DRβ1* 0401-Col II and HLA-DRβ1* 0101-HA complexes: a theoretical point of view. Journal of structural biology. 2005;149(1):38-52.
147. Cárdenas C, Obregón M, Balbín A, Villaveces JL, Patarroyo ME. Wave function analysis of MHC–peptide interactions. Journal of Molecular Graphics and Modelling. 2007;25(5):605-15.
148. Agudelo WA, Galindo JF, Ortiz M, Villaveces JL, Daza EE, Patarroyo ME. Variations in the electrostatic landscape of class II human leukocyte antigen molecule induced by modifications in the myelin basic protein peptide: a theoretical approach. PLoS One. 2009;4(1):e4164.
149. Bohórquez HJ, Cárdenas C, Matta CF, Boyd RJ, Patarroyo ME. Methods in biocomputational chemistry: a lesson from the amino acids. Quantum Biochemistry. 2010:403-21.
150. Stone JE, Hardy DJ, Ufimtsev IS, Schulten K. GPU-accelerated molecular modeling coming of age. Journal of Molecular Graphics and Modelling. 2010;29(2):116-25.
151. Akimov AV, Prezhdo OV. Large-scale computations in chemistry: a bird's eye view of a vibrant field. Chemical reviews. 2015;115(12):5797-890.
152. Stewart JJ. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. Journal of molecular modeling. 2013;19(1):1-32.
153. Elstner M. The SCC-DFTB method and its application to biological systems. Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta). 2006;116(1):316-25.
154. Christensen AS, Kubar Ts, Cui Q, Elstner M. Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. Chemical reviews. 2016;116(9):5301-37.
155. Kitaura K, Ikeo E, Asada T, Nakano T, Uebayasi M. Fragment molecular orbital method: an approximate computational method for large molecules. Chemical Physics Letters. 1999;313(3):701-6.
156. Fedorov DG, Nagata T, Kitaura K. Exploring chemistry with the fragment molecular orbital method. Physical Chemistry Chemical Physics. 2012;14(21):7562-77.
157. Fedorov DG, Kitaura K. Pair interaction energy decomposition analysis. Journal of computational chemistry. 2007;28(1):222-37.
158. González R, Suárez CF, Bohórquez HJ, Patarroyo MA, Patarroyo ME. Semi-empirical quantum evaluation of peptide–MHC class II binding. Chemical Physics Letters. 2017;668:29-34.
159. Patiño LC, Beau I, Carlosama C, Buitrago JC, González R, Suárez CF, et al. New mutations in non-syndromic primary ovarian insufficiency patients identified via whole-exome sequencing. Human Reproduction. 2017:1-9.
160. Patarroyo ME, Arévalo-Pinzón G, Reyes C, Moreno-Vranich A, Patarroyo MA. Malaria parasite survival depends on conserved binding peptides' critical biological functions. Current issues in molecular biology. 2016;18:57-78.

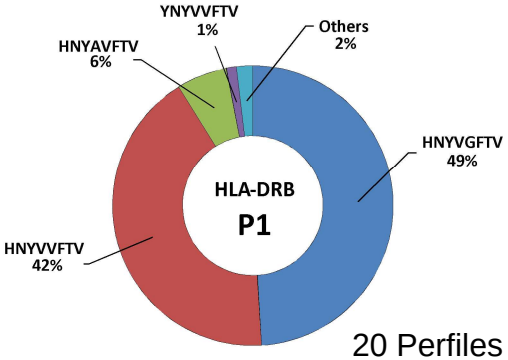
161. Alba MP, Suarez CF, Varela Y, Patarroyo MA, Bermudez A, Patarroyo ME. TCR-contacting residues orientation and HLA-DRbeta* binding preference determine long-lasting protective immunity against malaria. Biochem Biophys Res Commun. 2016;477(4):654-60.
162. Bermudez A, Calderon D, Moreno-Vranich A, Almonacid H, Patarroyo MA, Poloche A, et al. Gauche(+) side-chain orientation as a key factor in the search for an immunogenic peptide mixture leading to a complete fully protective vaccine. Vaccine. 2014;32(18):2117-26.
163. Patarroyo ME, Moreno-Vranich A, Bermudez A. Phi (Phi) and psi (Psi) angles involved in malarial peptide bonds determine sterile protective immunity. Biochem Biophys Res Commun. 2012;429(1-2):75-80.
164. Beck HP, Felger I, Barker M, Bugawan T, Genton B, Alexander N, et al. Evidence of HLA class II association with antibody response against the malaria vaccine SPF66 in a naturally exposed population. Am J Trop Med Hyg. 1995;53(3):284-8.
165. Patarroyo ME, Vinasco J, Amador R, Espejo F, Silva Y, Moreno A, et al. Genetic control of the immune response to a synthetic vaccine against Plasmodium falciparum. Parasite Immunol. 1991;13(5):509-16.
166. Patarroyo MA, Bermudez A, Lopez C, Yepes G, Patarroyo ME. 3D analysis of the TCR/pMHCII complex formation in monkeys vaccinated with the first peptide inducing sterilizing immunity against human malaria. PLoS One. 2010;5(3):e9771.
167. Cifuentes G, Patarroyo ME, Urquiza M, Ramirez LE, Reyes C, Rodriguez R. Distorting malaria peptide backbone structure to enable fitting into MHC class II molecules renders modified peptides immunogenic and protective. J Med Chem. 2003;46(11):2250-3.
168. Stern LJ, Wiley DC. Antigenic peptide binding by class I and class II histocompatibility proteins. Structure. 1994;2(4):245-51.
169. Madden DR. The three-dimensional structure of peptide-MHC complexes. Annu Rev Immunol. 1995;13:587-622.
170. Barber LD, Parham P. Peptide binding to major histocompatibility complex molecules. Annu Rev Cell Biol. 1993;9:163-206.
171. Adzhubei AA, Sternberg MJ, Makarov AA. Polyproline-II helix in proteins: structure and function. Journal of molecular biology. 2013;425(12):2100-32.
172. Bohórquez HJ, Suárez CF, Patarroyo ME. Mass & secondary structure propensity of amino acids explain their mutability and evolutionary replacements. Scientific Reports. 2017;7(1):7717.
173. González-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MHT, Silva ALSd, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. Nucleic acids research. 2014;43(D1):D784-D8.
174. Berkholz DS, Krenesky PB, Davidson JR, Karplus PA. Protein Geometry Database: a flexible engine to explore backbone conformations and their relationships to covalent geometry. Nucleic acids research. 2009;38(suppl_1):D320-D5.

Anexo 1. Diccionario de bolsillos del CMH-DRB

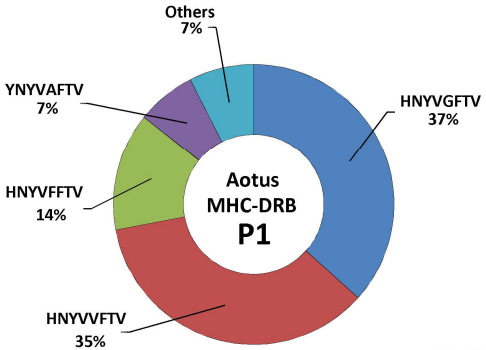
Humano/*Aotus* MHC-DRB
Bolsillo 1 - Perfiles







20 Perfiles
2100 alelos



11 perfiles
215 alelos

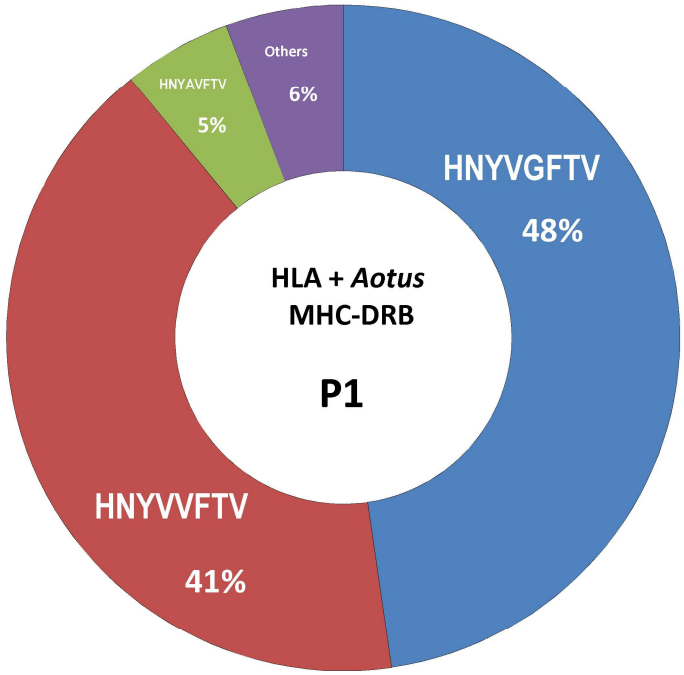


Figura A

Tabla 1

Perfiles de bolsillo más frecuentes en el HLA-DRB (>60%)

			Pocket 1								Pocket 4								Pocket 6								Pocket 9							
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*01	HLA-DRB1*010101	**51.7**	H	N	Y	V	G	F	T	V	E	C	L	E	F	Q	R	R	A	A	Y	C	W	Q	L	K	F	E	R	C	R	W	F	C	S	V	D	Y	W
	HLA-DRB1*010201	**14.9**	H	N	Y	A	V	F	T	V	E	C	L	E	F	Q	R	R	A	A	Y	C	W	Q	L	K	F	E	R	C	R	W	F	C	S	V	D	Y	W
	HLA-DRB1*0104	**2.3**	H	N	Y	V	V	F	T	V	E	C	L	E	F	Q	R	R	A	A	Y	C	W	Q	L	K	F	E	R	C	R	W	F	C	S	V	D	Y	W
	HLA-DRB1*0109	**2.3**	H	N	Y	V	G	F	T	V	E	C	L	E	F	Q	A	R	A	A	Y	C	W	Q	L	K	F	E	R	C	A	W	F	C	S	V	D	Y	W

			Pocket 1								Pocket 4								Pocket 6								Pocket 9							
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*12	HLA-DRB1*120101	**59.6**	H	N	Y	A	V	F	T	V	E	C	L	E	F	D	R	R	A	A	Y	C	E	Y	S	T	G	E	R	H	R	**E**	**G**	**H**	**L**	**L**	**V**	**S**	**W**
	HLA-DRB1*121601	**8.5**	H	N	Y	V	G	F	T	V	E	C	L	E	F	D	R	R	A	A	Y	C	E	Y	S	T	G	E	R	H	R	**E**	**G**	**H**	**L**	**L**	**V**	**S**	**W**
	HLA-DRB1*120302	**6.4**	H	N	Y	V	V	F	T	V	E	C	L	E	F	D	R	R	A	A	Y	C	E	Y	S	T	G	E	R	H	R	**E**	**G**	**H**	**L**	**L**	**V**	**S**	**W**
	HLA-DRB1*1204	**4.3**	H	N	Y	A	V	F	T	V	E	C	L	E	F	D	R	R	A	A	Y	C	E	Y	S	T	G	E	R	H	R	**E**	**G**	**H**	**L**	**L**	**D**	**Y**	**W**
	HLA-DRB1*1205	**4.3**	H	N	Y	A	V	F	T	V	E	C	L	E	F	D	R	R	A	A	Y	C	E	Y	S	T	G	E	R	H	R	**E**	**G**	**H**	**F**	**L**	**V**	**S**	**W**

			Pocket 1								Pocket 4								Pocket 6								Pocket 9							
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*03	HLA-DRB1*03010101	**51.8**	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	N	V	D	Y	W
	HLA-DRB1*030201	**5.3**	H	N	Y	V	G	F	T	V	E	C	F	E	F	Q	K	R	G	R	Y	C	E	Y	S	T	S	E	R	Y	K	E	S	Y	N	V	D	Y	W
	HLA-DRB1*030501	**4.4**	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	N	V	D	Y	W
	HLA-DRB1*0325	**2.6**	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*0357	**1.8**	H	N	Y	V	A	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	N	V	D	Y	W
	HLA-DRB1*0340	**1.8**	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	K	R	G	R	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*0326	**1.8**	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	N	A	D	Y	W
	HLA-DRB1*031301	**1.8**	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	N	V	D	S	W
	HLA-DRB1*030401	**1.8**	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	S	V	D	Y	W

			Pocket 1								Pocket 4								Pocket 6								Pocket 9							
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*04	HLA-DRB1*040101	**13.9**	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	K	R	A	A	Y	C	E	Q	V	K	H	D	R	Y	K	E	H	Y	Y	V	D	Y	W
	HLA-DRB1*040501	**13.0**	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	R	R	A	A	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	S	Y	W
	HLA-DRB1*040301	**10.6**	H	N	Y	V	V	F	T	V	E	C	F	D	F	Q	R	R	A	E	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	Y	W
	HLA-DRB1*040401	**7.7**	H	N	Y	V	V	F	T	V	E	C	F	D	F	Q	R	R	A	A	Y	C	E	Q	V	K	H	D	R	Y	E	E	H	Y	Y	V	D	Y	W
	HLA-DRB1*040201	**3.8**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	E	R	A	A	Y	C	E	Q	V	K	H	D	R	Y	E	E	H	Y	Y	V	D	Y	W
	HLA-DRB1*040601	**3.8**	H	N	Y	V	V	F	T	V	E	C	F	D	F	Q	R	R	A	E	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	S	V	D	Y	W
	HLA-DRB1*040701	**2.9**	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	R	R	A	E	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	Y	W
	HLA-DRB1*040801	**2.4**	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	R	R	A	A	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	Y	W
	HLA-DRB1*0415	**1.4**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	Y	W
	HLA-DRB1*0418	**1.4**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	L	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	Y	W
	HLA-DRB1*041001	**1.4**	H	N	Y	V	V	F	T	V	E	C	F	D	F	Q	R	R	A	A	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	S	Y	W
	HLA-DRB1*041101	**1.4**	H	N	Y	V	V	F	T	V	E	C	F	D	F	Q	R	R	A	E	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	S	Y	W

			Pocket 1								Pocket 4								Pocket 6								Pocket 9							
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA - DRB1 * 07	HLA-DRB1*07010101	**50.0**	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	R	G	Q	V	C	W	Q	G	K	Y	E	R	L	R	W	Y	L	F	V	S	W
	HLA-DRB1*0704	**7.7**	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	R	G	Q	Y	C	W	Q	G	K	Y	E	R	L	R	W	Y	L	F	V	V	S	W
	HLA-DRB1*0703	3.8	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	R	G	Q	V	C	W	Q	G	K	Y	E	S	L	R	W	Y	L	F	V	V	S	W
	HLA-DRB1*0706	3.8	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	R	G	Q	V	C	W	Q	G	K	Y	E	R	L	R	W	Y	L	F	V	A	Y	W
	HLA-DRB1*0708	3.8	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	R	G	Q	V	C	W	Q	G	K	Y	E	R	L	R	W	Y	L	V	V	V	S	W
	HLA-DRB1*0709	3.8	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	R	G	Q	V	C	W	Q	G	K	Y	E	R	F	R	W	Y	F	F	V	V	S	W
	HLA-DRB1*0712	3.8	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	R	G	Q	V	C	W	Q	G	K	Y	E	R	L	R	W	Y	L	F	V	I	S	W
	HLA-DRB1*0717	3.8	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	W	G	Q	V	C	W	Q	G	K	Y	E	R	L	R	W	Y	L	F	V	V	S	W
	HLA-DRB1*0718	3.8	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	R	S	Q	V	C	W	Q	G	K	Y	E	R	L	R	W	Y	L	F	V	V	S	W
	HLA-DRB1*0720	3.8	H	N	Y	V	D	F	T	V	K	C	F	E	F	D	R	R	G	Q	V	C	W	Q	G	K	Y	E	R	L	R	W	Y	L	F	V	V	S	W
	HLA-DRB1*0722	3.8	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	R	G	Q	V	C	W	Q	G	K	C	E	R	L	R	W	C	L	F	V	V	S	W
	HLA-DRB1*0723	3.8	H	N	Y	V	G	F	T	V	K	C	F	E	F	D	R	R	G	Q	V	C	W	R	G	K	Y	E	R	L	R	W	Y	L	F	V	V	S	W
	HLA-DRB1*0724	3.8	H	N	Y	V	V	F	T	V	K	C	F	E	F	D	R	R	G	Q	V	C	W	Q	G	K	Y	E	R	L	R	W	Y	L	F	V	V	S	W

			Pocket 1								Pocket 4								Pocket 6								Pocket 9							
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*08	HLA-DRB1*080101	**30.0**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	L	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	Y	V	S	Y	W
	HLA-DRB1*080201	**14.3**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	L	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	Y	V	D	Y	W
	HLA-DRB1*080401	**11.4**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	L	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	Y	V	D	Y	W
	HLA-DRB1*0805	**2.9**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	Y	V	S	Y	W
	HLA-DRB1*0806	**2.9**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	L	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	Y	V	S	Y	W
	HLA-DRB1*0812	**2.9**	H	N	Y	A	V	F	T	V	E	C	F	D	F	D	R	R	A	L	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	Y	V	S	Y	W
	HLA-DRB1*0824	**2.9**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	Y	V	D	Y	W
	HLA-DRB1*0834	**2.9**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	L	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	Y	V	V	S	W

			Pocket 1								Pocket 4								Pocket 6								Pocket 9							
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*09	HLA-DRB1*090102	**57.1**	H	N	Y	V	G	F	T	V	E	C	Y	H	F	R	R	R	A	E	V	C	K	Q	D	K	F	H	R	G	R	K	F	G	N	V	V	S	W
	HLA-DRB1*090201	**7.1**	H	N	Y	V	G	F	T	V	E	C	Y	H	F	R	R	R	A	E	V	C	K	Q	D	K	F	H	R	G	R	K	F	G	N	V	D	Y	W
	HLA-DRB1*0903	3.6	H	N	Y	V	G	F	T	V	E	C	Y	H	F	D	R	R	A	E	V	C	K	Q	D	K	F	H	R	G	R	K	F	G	N	V	V	S	W
	HLA-DRB1*0905	3.6	H	N	Y	V	G	F	T	V	E	C	Y	H	F	R	R	R	A	E	Y	C	K	Q	D	K	F	H	R	G	R	K	F	G	N	V	V	S	W

			Pocket 1								Pocket 4								Pocket 6								Pocket 9							
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*10	HLA-DRB1*100101	**42.9**	H	N	Y	V	G	F	T	V	E	C	L	E	Y	R	R	R	A	A	Y	C	E	E	V	K	F	E	R	R	R	E	F	R	Y	A	D	Y	W
	HLA-DRB1*1002	14.3	H	N	Y	V	V	F	T	V	E	C	L	E	Y	R	R	R	A	A	Y	C	E	E	V	K	F	E	R	R	R	E	F	R	Y	A	D	Y	W
	HLA-DRB1*1003	14.3	H	N	Y	V	G	F	T	V	E	C	L	E	F	R	R	R	A	A	Y	C	E	E	V	K	F	E	R	R	R	E	F	R	Y	A	D	Y	W

			Pocket 1								Pocket 4								Pocket 6								Pocket 9							
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA - DRB1 * 11	HLA-DRB1*110101	**27.3**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*110401	**14.9**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*110201	**6.2**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	E	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	E	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*110601	**2.6**	H	N	Y	A	V	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*111101	**2.6**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	E	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	E	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*111001	**2.1**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	F	V	D	Y	W
	HLA-DRB1*11103	**2.1**	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	K	R	G	R	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*11113	**2.1**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	N	V	D	Y	W
	HLA-DRB1*1109	**1.5**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	N	V	D	Y	W
	HLA-DRB1*1123	**1.5**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	L	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	Y	V	D	Y	W

Tabla 1 (cont)

			Pocket 1								Pocket 4										Pocket 6										Pocket 9									
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*13	HLA-DRB1*130101	**18.5**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	E	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	E	E	S	Y	N	V	D	Y	W
	HLA-DRB1*130201	**11.0**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	E	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	E	E	S	Y	N	V	D	Y	W
	HLA-DRB1*130301	**7.5**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	K	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	Y	V	S	Y	W
	HLA-DRB1*1312	**5.0**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	Y	V	S	Y	W
	HLA-DRB1*130701	**4.0**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*130501	**3.0**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	N	V	D	Y	W
	HLA-DRB1*13149	**2.5**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	E	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	E	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*132301	**2.0**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	E	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	E	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*1304	**1.5**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	E	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	E	E	S	Y	Y	V	S	Y	W
	HLA-DRB1*1308	**1.5**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	E	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	E	E	S	Y	F	V	D	Y	W
	HLA-DRB1*131101	**1.5**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	Y	V	D	Y	W
	HLA-DRB1*1313	**1.5**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	L	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	Y	V	S	Y	W
	HLA-DRB1*1389	**1.5**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	K	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	K	E	S	Y	Y	V	S	Y	W

			Pocket 1								Pocket 4										Pocket 6										Pocket 9									
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*14	HLA-DRB1*140101	**15.4**	H	N	Y	V	V	F	T	V	E	C	F	D	F	R	R	R	A	E	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	F	V	A	H	W
	HLA-DRB1*140501	**8.3**	H	N	Y	V	V	F	T	V	E	C	F	D	F	R	R	R	A	E	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	F	V	D	Y	W
	HLA-DRB1*140301	**4.8**	H	N	Y	V	G	F	T	V	E	C	F	E	F	D	R	R	A	L	Y	C	E	Y	S	T	S	E	R	Y	R	E	S	Y	N	V	D	Y	W
	HLA-DRB1*1404	**4.2**	H	N	Y	V	V	F	T	V	E	C	F	D	F	R	R	R	A	E	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	F	V	A	H	W
	HLA-DRB1*1414	**4.2**	H	N	Y	V	G	F	T	V	E	C	F	D	F	R	R	R	A	E	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	F	V	D	Y	W
	HLA-DRB1*140601	**2.6**	H	N	Y	V	V	F	T	V	E	C	F	E	F	Q	R	R	A	A	Y	C	E	Y	S	T	S	E	R	Y	R	E	S	Y	N	V	D	Y	W
	HLA-DRB1*1408	**2.0**	H	N	Y	V	V	F	T	V	E	C	F	D	F	R	R	R	A	E	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	F	V	D	H	W
	HLA-DRB1*1425	**2.0**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	Y	V	A	H	W
	HLA-DRB1*143201	**2.0**	H	N	Y	V	V	F	T	V	E	C	F	D	F	R	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	F	V	A	H	W
	HLA-DRB1*1402	**2.0**	H	N	Y	V	G	F	T	V	E	C	F	E	F	Q	R	R	A	A	Y	C	E	Y	S	T	S	E	R	Y	R	E	S	Y	N	V	D	Y	W
	HLA-DRB1*140701	**1.3**	H	N	Y	V	G	F	T	V	E	C	F	D	F	R	R	R	A	E	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	F	V	A	H	W
	HLA-DRB1*1409	**1.3**	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	N	V	D	Y	W
	HLA-DRB1*14100	**1.3**	H	N	Y	V	V	F	T	V	E	C	F	D	F	R	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	F	V	D	Y	W
	HLA-DRB1*14105	**1.3**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	F	V	A	H	W
	HLA-DRB1*14107	**1.3**	H	N	Y	V	V	F	T	V	E	C	F	D	F	Q	K	R	G	R	Y	C	E	Y	S	T	G	D	R	Y	K	E	G	Y	F	V	A	H	W
	HLA-DRB1*1411	**1.3**	H	N	Y	V	V	F	T	V	E	C	F	D	F	R	R	R	A	E	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	F	V	D	Y	W
	HLA-DRB1*141201	**1.3**	H	N	Y	V	V	F	T	V	E	C	F	E	F	D	R	R	A	L	Y	C	E	Y	S	T	S	E	R	Y	R	E	S	Y	N	V	D	Y	W
	HLA-DRB1*1417	**1.3**	H	N	Y	V	V	F	T	V	E	C	F	D	F	Q	R	R	A	A	Y	C	E	Y	S	T	S	D	R	Y	R	E	S	Y	N	V	D	Y	W
	HLA-DRB1*1463	**1.3**	H	N	Y	V	G	F	T	V	E	C	F	E	F	D	R	R	A	L	Y	C	E	Y	S	T	S	E	R	Y	R	E	S	Y	N	V	S	Y	W
	HLA-DRB1*1468	**1.3**	H	N	Y	V	G	F	T	V	E	C	F	D	F	R	R	R	A	E	Y	C	E	Y	S	T	G	D	R	Y	R	E	G	Y	F	V	A	H	W

			Pocket 1								Pocket 4										Pocket 6										Pocket 9									
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*15	HLA-DRB1*15010101	**55.5**	H	N	Y	V	V	F	T	V	E	C	F	D	F	Q	A	R	A	A	Y	C	W	Q	P	K	R	D	R	Y	A	W	R	Y	S	V	D	Y	W
	HLA-DRB1*150201	**15.6**	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	A	R	A	A	Y	C	W	Q	P	K	R	D	R	Y	A	W	R	Y	S	V	D	Y	W
	HLA-DRB1*15030101	**4.7**	H	N	Y	V	V	F	T	V	E	C	F	D	F	Q	A	R	A	A	Y	C	W	Q	P	K	R	D	R	H	A	W	R	H	S	V	D	Y	W
	HLA-DRB1*1538	**1.6**	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	A	R	A	A	Y	C	W	Q	P	K	R	D	R	Y	A	W	R	Y	S	V	D	S	W
	HLA-DRB1*1527	**1.6**	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	R	R	A	A	Y	C	W	Q	P	K	R	D	R	Y	R	W	R	Y	S	V	D	Y	W

			Pocket 1								Pocket 4										Pocket 6										Pocket 9									
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB1*16	HLA-DRB1*160101	**64.0**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	W	Q	P	K	R	D	R	Y	R	W	R	Y	S	V	D	Y	W
	HLA-DRB1*1604	**8.0**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	L	Y	C	W	Q	P	K	R	D	R	Y	R	W	R	Y	S	V	D	Y	W
	HLA-DRB1*1615	**8.0**	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	A	Y	C	W	Q	P	K	R	D	R	Y	R	W	R	Y	S	V	D	Y	W

			Pocket 1								Pocket 4										Pocket 6										Pocket 9									
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB3*01	HLA-DRB3*01010201	**42.1**	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	L	R	K	S	D	R	Y	K	E	S	Y	F	L	V	S	W
	HLA-DRB3*0102	5.3	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	L	C	K	S	D	R	Y	K	E	S	Y	F	L	V	S	W
	HLA-DRB3*0103	5.3	H	N	Y	V	G	F	T	V	E	C	Y	E	F	Q	K	R	G	R	Y	C	E	L	R	K	S	E	R	Y	K	E	S	Y	F	L	V	S	W
	HLA-DRB3*0105	5.3	H	N	Y	V	G	F	T	V	E	C	Y	N	F	Q	K	R	G	R	Y	C	E	L	R	K	S	N	R	Y	K	E	S	Y	F	L	V	S	W
	HLA-DRB3*0106	5.3	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	L	R	K	S	D	R	Y	K	E	S	Y	F	V	V	S	W

			Pocket 1								Pocket 4										Pocket 6										Pocket 9									
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB3*02	HLA-DRB3*02020101	**39.4**	H	N	Y	V	G	F	T	V	E	C	F	E	F	Q	K	R	G	Q	Y	C	E	L	L	K	S	E	R	H	K	E	S	H	Y	A	D	Y	W
	HLA-DRB3*0201	**6.1**	H	N	Y	V	V	F	T	V	E	C	F	E	F	Q	K	R	G	Q	Y	C	E	L	L	K	S	E	R	H	K	E	S	H	Y	A	D	Y	W
	HLA-DRB3*0209	**6.1**	H	N	Y	V	G	F	T	V	E	C	F	E	F	Q	K	R	G	Q	Y	C	E	L	L	K	S	E	R	H	K	E	S	H	Y	A	V	S	W
	HLA-DRB3*0203	3.0	H	N	Y	V	G	F	T	V	E	C	F	E	F	Q	K	R	G	Q	Y	C	E	L	L	K	S	E	R	H	K	E	S	H	S	V	D	Y	W
	HLA-DRB3*0204	3.0	H	N	Y	V	V	F	T	V	E	C	F	E	F	Q	K	R	G	R	Y	C	E	L	L	K	S	E	R	H	K	E	S	H	Y	A	D	Y	W
	HLA-DRB3*0205	3.0	H	N	Y	V	G	F	T	V	E	C	F	E	F	Q	K	R	G	Q	Y	C	E	L	L	K	S	E	R	Y	K	E	S	Y	Y	A	D	Y	W

			Pocket 1								Pocket 4										Pocket 6										Pocket 9									
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB3*03	HLA-DRB3*030101	**60.0**	H	N	Y	V	V	F	T	V	E	C	F	E	F	Q	K	R	G	Q	Y	C	E	L	L	K	S	E	R	Y	K	E	S	Y	F	V	V	S	W
	HLA-DRB3*0303	20.0	H	N	Y	V	G	F	T	V	E	C	F	E	F	Q	K	R	G	R	Y	C	E	L	L	K	S	E	R	Y	K	E	S	Y	F	V	V	S	W
	HLA-DRB3*0302	20.0	H	N	Y	V	V	F	T	V	E	C	F	E	F	Q	K	R	G	Q	Y	C	E	L	L	K	S	E	R	H	K	E	S	H	F	V	V	S	W

			Pocket 1								Pocket 4										Pocket 6										Pocket 9									
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB4*01	HLA-DRB4*01010101	**91.7**	Y	N	Y	V	V	F	T	V	E	C	N	I	Y	R	R	R	A	E	Y	C	E	Q	A	K	C	I	R	Y	R	E	C	Y	Y	A	D	Y	W
	HLA-DRB4*0105	8.3	H	N	Y	V	V	F	T	V	E	C	N	I	Y	R	R	R	A	E	Y	C	E	Q	A	K	C	I	R	Y	R	E	C	Y	Y	A	D	Y	W

			Pocket 1								Pocket 4										Pocket 6										Pocket 9									
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB5*01	HLA-DRB5*010101	**30.8**	H	N	Y	V	G	F	T	V	E	C	F	H	F	D	R	R	A	A	Y	C	Q	Q	D	K	Y	H	R	D	R	Q	Y	D	D	L	D	Y	W
	HLA-DRB5*0102	7.7	H	N	Y	V	G	F	T	V	E	C	F	H	F	D	R	R	A	A	Y	C	Q	Q	D	K	Y	H	R	G	R	Q	Y	G	N	V	D	Y	W
	HLA-DRB5*0103	7.7	H	N	Y	V	G	F	T	V	E	C	F	H	F	D	T	R	A	A	Y	C	Q	Q	D	K	Y	H	R	G	T	Q	Y	G	N	V	D	Y	W
	HLA-DRB5*0104	7.7	H	N	Y	V	G	F	T	V	E	C	F	H	F	D	R	R	A	L	Y	C	Q	Q	D	K	Y	H	R	D	R	Q	Y	D	D	L	D	Y	W
	HLA-DRB5*0105	7.7	H	N	Y	V	G	F	T	V	E	C	F	H	F	D	R	R	A	A	Y	C	Q	Q	D	K	Y	H	R	D	R	Q	Y	D	D	V	D	Y	W

			Pocket 1								Pocket 4										Pocket 6										Pocket 9									
	Allele prototype	**PPF**	**81**	**82**	**83**	**85**	**86**	**89**	**90**	**91**	**14**	**15**	**26**	**28**	**40**	**70**	**71**	**72**	**73**	**74**	**78**	**79**	**9**	**10**	**11**	**12**	**13**	**28**	**29**	**30**	**71**	**9**	**13**	**30**	**37**	**38**	**57**	**60**	**61**
HLA-DRB5*02	HLA-DRB5*0202	**60.0**	H	N	Y	A	V	F	T	V	E	C	F	H	F	Q	A	R	A	A	Y	C	Q	Q	D	K	Y	H	R	G	A	Q	Y	G	N	V	D	Y	W
	HLA-DRB5*0205	20.0	H	N	Y	A	V	F	T	V	E	C	F	H	F	Q	R	R	A	A	Y	C	Q	Q	D	K	Y	H	R	G	R	Q	Y	G	N	V	D	Y	W
	HLA-DRB5*0203	20.0	H	N	Y	V	G	F	T	V	E	C	F	H	F	Q	A	R	A	A	Y	C	Q	Q	D	K	Y	H	R	G	A	Q	Y	G	N	V	D	Y	W

Tabla 2

Perfiles de bolsillo más frecuentes en el *Aotus*-MHC-DRB (>60%)

			Pocket 1								Pocket 4										Pocket 6									Pocket 9								
	Allele prototype	PPF	81	82	83	85	86	89	90	91	14	15	26	28	40	70	71	72	73	74	78	79	9	10	11	12	13	28	29	30	71	9	13	30	37	38	57	60	61
Ao-DRB*W38	Aoaz-DRB*W3801	100.0	H	N	Y	V	G	F	T	V	E	C	F	E	F	D	R	R	A	Q	V	C	E	Q	A	K	Y	E	R	H	R	E	Y	H	Y	A	T	Y	W
	Aona-DRB*W3802	50.0	H	N	Y	V	G	F	T	V	E	C	F	E	F	D	R	R	A	Q	V	C	E	Q	A	K	Y	E	R	H	R	E	Y	H	Y	A	T	Y	W
	Aona-DRB*W3801	50.0	H	N	Y	V	G	F	T	V	E	C	F	E	F	V	R	R	A	Q	V	C	E	Q	A	K	Y	E	R	H	R	E	Y	H	Y	A	T	Y	W
	Aoni-DRB*W3801	100.0	H	N	Y	V	V	F	T	V	E	C	F	E	F	D	R	R	A	Q	V	C	E	Q	A	K	Y	E	R	H	R	E	Y	H	Y	A	T	Y	W
Ao-DRB*W13	Aona-DRB*W1302	**25.0**	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	D	Y	F
	Aona-DRB*W1308	**25.0**	H	N	Y	V	A	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	E	Y	F
	Aona-DRB*W1301	**16.7**	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	E	Y	F
	Aona-DRB*W1303	8.3	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	L	D	R	Y	T	E	L	Y	Y	V	E	Y	F
	Aona-DRB*W1307	8.3	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	D	Y	L
	Aona-DRB*W1310	8.3	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	D	Y	W
	Aona-DRB*W1312	8.3	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	F	V	T	Y	F
	Aoni-DRB*W1301	**33.3**	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	D	Y	F
	Aoni-DRB*W1306	**22.2**	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	D	Y	W
	Aoni-DRB*W1302	11.1	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	D	H	F
	Aoni-DRB*W1305	11.1	H	N	Y	V	A	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	E	Y	F
	Aoni-DRB*W1307	11.1	H	N	Y	V	G	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	D	Y	F
	Aoni-DRB*W1308	11.1	H	D	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	D	Y	F
	Aovo-DRB*W130101	**50.0**	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	F	V	T	Y	F
	Aovo-DRB*W1302	25.0	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	F	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	F	V	T	Y	F
	Aovo-DRB*W1304	25.0	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	E	Q	F	K	P	D	R	Y	T	E	P	Y	Y	V	D	Y	F
Ao-DRB*W18	Aona-DRB*W1802	**42.9**	H	N	Y	V	F	F	T	V	E	C	F	E	F	L	K	R	G	Q	Y	C	E	L	V	K	S	E	R	Y	K	E	S	Y	L	V	D	Y	W
	Aona-DRB*W1801	**28.6**	H	N	Y	V	G	F	T	V	E	C	F	E	F	L	K	R	G	Q	Y	C	E	Q	V	K	S	E	R	Y	K	E	S	Y	F	V	D	Y	W
	Aona-DRB*W1803	14.3	H	N	Y	V	V	F	T	V	E	C	F	E	F	L	K	R	G	Q	Y	C	E	Q	V	K	S	E	R	Y	K	E	S	Y	F	V	D	Y	W
	Aona-DRB*W1804	14.3	H	N	Y	V	F	F	T	V	E	C	F	E	F	L	K	R	G	Q	Y	C	E	L	V	K	S	E	R	Y	K	E	S	Y	L	A	D	Y	W
	Aoni-DRB1*W1801	100.0	H	N	Y	V	G	F	T	V	E	C	F	E	F	L	K	R	G	Q	Y	C	E	Q	V	K	S	E	R	Y	K	E	S	Y	F	V	D	Y	W
	Aotr-DRB*W1801	100.0	H	N	Y	V	F	F	T	V	E	C	F	E	F	L	K	R	G	Q	Y	C	E	Q	A	K	S	E	R	Y	K	E	S	Y	Y	V	D	Y	W
	Aovo-DRB*W1801	**66.7**	H	N	Y	V	F	F	T	V	E	C	F	E	F	L	K	R	G	Q	Y	C	E	Q	A	K	S	E	R	Y	K	E	S	Y	Y	V	D	Y	W
	Aovo-DRB*W1803	33.3	H	N	Y	V	F	F	T	V	E	C	F	E	F	L	K	R	G	Q	Y	C	E	Q	G	K	S	E	R	Y	K	E	S	Y	Y	V	D	Y	W
Ao-DRB*W29	Aona-DRB*W2901	**50.0**	H	N	Y	V	V	F	T	V	E	C	L	Q	F	Y	L	R	A	A	Y	C	E	Q	T	K	S	Q	R	Y	L	E	S	Y	Y	A	D	Y	W
	Aona-DRB*W2906	**25.0**	H	N	Y	V	V	F	T	V	E	C	L	Q	F	Y	L	R	A	A	Y	C	E	Q	T	K	S	Q	R	Y	L	E	S	Y	Y	A	D	Y	W
	Aona-DRB*W2907	12.5	H	N	Y	V	G	F	A	V	E	C	L	Q	F	Y	L	R	A	A	Y	C	E	Q	T	K	S	Q	R	Y	L	E	S	Y	Y	A	D	Y	W
	Aona-DRB*W2908	12.5	H	N	Y	V	G	F	T	V	E	C	L	Q	F	Y	L	R	A	A	Y	C	E	Q	T	K	S	Q	R	Y	L	E	S	Y	Y	A	D	Y	W
	Aoni-DRB*W2902	**80.0**	H	N	Y	V	V	F	T	V	E	C	L	Q	F	Y	L	R	A	A	C	C	E	Q	T	K	S	Q	R	Y	L	E	S	Y	Y	V	D	Y	W
	Aoni-DRB*W2901	20.0	H	N	Y	V	G	F	T	V	E	C	L	Q	F	Y	L	R	A	A	Y	C	E	Q	T	K	S	Q	R	Y	L	E	S	Y	Y	A	D	Y	W
	Aovo-DRB*W2901	100.0	H	N	Y	V	V	F	T	V	E	C	L	Q	F	Y	L	R	A	A	C	C	E	Q	T	K	S	Q	R	Y	L	E	S	Y	Y	V	D	Y	W
Ao-DRB*W30	Aona-DRB*W3002	50.0	H	N	Y	V	G	F	T	V	E	C	Y	E	F	D	R	R	A	A	Y	C	E	Q	V	K	Y	E	R	Y	R	E	Y	Y	F	V	S	K	L
	Aona-DRB*W3001	50.0	H	N	Y	V	G	F	T	V	E	C	Y	E	F	D	R	R	A	S	Y	C	E	Q	V	K	Y	E	R	L	R	E	Y	L	F	V	S	K	L
	Aovo-DRB*W3001	100.0	H	N	Y	V	G	F	T	V	E	C	Y	E	F	D	R	R	A	S	Y	C	E	Q	V	K	Y	E	R	L	R	E	Y	L	F	V	V	K	L
W42	Aona-DRB*W4201	100.0	H	N	Y	V	V	F	T	V	E	C	F	E	F	Y	L	R	A	A	Y	C	E	Q	V	K	D	E	R	Y	L	E	D	Y	Y	V	D	Y	W
	Aoni-DRB*W4201	100.0	H	N	Y	V	V	F	T	V	E	C	F	E	F	Y	L	R	A	A	Y	C	E	Q	V	K	D	E	R	Y	L	E	D	Y	Y	V	D	Y	W
Ao-DRB*W43	Aoni-DRB*W4301	25.0	H	N	Y	V	G	F	T	V	E	C	L	D	F	N	R	R	A	A	Y	C	K	Q	V	K	C	D	R	H	R	K	C	H	Y	V	T	Y	W
	Aoni-DRB*W4302	25.0	H	N	Y	V	G	F	T	V	E	C	L	D	S	N	R	R	A	A	Y	C	K	Q	V	K	C	D	R	H	R	K	C	H	Y	V	T	Y	W
	Aoni-DRB*W4303	25.0	H	N	Y	V	G	L	T	V	E	C	L	D	F	N	R	R	A	A	Y	C	K	Q	V	K	C	D	R	H	R	K	C	H	Y	V	T	Y	W
	Aoni-DRB*W4304	25.0	H	N	Y	V	V	F	T	V	E	C	L	D	F	N	R	R	A	A	Y	R	K	Q	V	K	C	D	R	H	R	K	C	H	Y	V	T	Y	W
	Aovo-DRB*W4301	100.0	H	N	Y	V	G	F	T	V	E	C	L	D	F	N	R	R	A	A	Y	C	K	Q	V	K	C	D	R	H	R	K	C	H	Y	V	T	Y	W
W44	Aona-DRB*W4401	100.0	H	N	Y	V	G	F	T	V	E	C	Y	D	F	D	R	R	A	A	Y	C	E	Q	A	K	S	D	R	Y	R	E	S	Y	Y	V	T	Y	W
	Aoni-DRB*W4401	100.0	H	K	Y	V	G	F	T	V	E	C	Y	D	F	D	R	R	A	A	Y	C	E	Q	A	K	S	D	R	Y	R	E	S	Y	Y	V	T	Y	W
W45	Aona-DRB*W4501	100.0	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	Q	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	V	V	D	Y	W
	Aovo-DRB*W4501	100.0	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	K	R	A	S	Y	C	E	Q	V	K	H	D	R	Y	K	E	H	Y	V	V	D	Y	W
Ao-DRB*W47	Aona-DRB*W470401GA	**53.8**	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	Q	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	Y	W
	Aona-DRB*W4701GA	7.7	H	N	Y	V	V	F	T	V	E	C	F	D	F	Y	R	R	A	Q	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	Y	W
	Aona-DRB*W4702GA	7.7	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	P	Q	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	Y	W
	Aona-DRB*W4703GA	7.7	H	N	Y	V	V	F	T	V	E	C	F	D	F	D	R	R	A	Q	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	Y	W
	Aona-DRB*W4705GA	7.7	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	Q	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	H	W
	Aona-DRB*W4708GA	7.7	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	Q	Y	C	K	Q	V	K	H	D	R	Y	R	K	H	Y	Y	V	D	Y	W
	Aona-DRB*W4709GA	7.7	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	Q	Y	C	E	Q	V	K	D	D	R	Y	R	E	D	Y	Y	V	D	Y	W
	Aovo-DRB*W4701GA	100.0	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	Q	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	V	D	Y	W
W47	Aoni-DRB*W4701GB	100.0	H	N	Y	V	G	F	T	V	E	C	F	D	F	D	R	R	A	Q	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	A	D	Y	W
	Aovo-DRB*W4702GB	100.0	H	N	Y	V	G	F	T	E	E	C	F	D	F	D	R	R	A	Q	Y	C	E	Q	V	K	H	D	R	Y	R	E	H	Y	Y	A	D	Y	W

Tabla 2 (cont)

	Allele prototype	PPF	Pocket 1								Pocket 4								Pocket 6								Pocket 9							
			81	82	83	85	86	89	90	91	14	15	26	28	40	70	71	72	73	74	78	79	9	10	11	12	13	28	29	30	71	9	13	30	37	38	57	60	61
W88	Aovo-DRB*W8801	100.0	H	N	Y	V	A	F	T	V	E	C	L	Q	F	Y	L	R	A	A	Y	C	E	Q	V	K	D	Q	R	Y	L	E	D	Y	Y	V	D	Y	W
W89	Aona-DRB*W8901	100.0	H	N	Y	V	A	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	Q	T	K	S	D	R	Y	K	E	S	Y	Y	V	T	Y	W
W90	Aovo-DRB*W9001	100.0	H	N	Y	V	G	F	T	V	E	C	L	Q	F	Y	L	R	A	A	Y	C	E	Q	G	K	S	Q	R	Y	L	E	S	Y	V	L	S	K	L
W91	Aona-DRB*W9101	100.0	H	N	Y	V	G	F	T	V	E	C	F	E	F	T	R	R	A	A	F	C	E	Q	A	K	C	E	R	Y	R	E	C	Y	V	L	E	S	W
	Aovo-DRB*W9102	50.0	H	N	Y	V	F	F	T	V	E	C	F	E	F	T	R	R	A	A	F	C	E	Q	A	K	G	E	R	Y	R	E	G	Y	V	L	S	K	Y
	Aovo-DRB*W9101	50.0	H	N	Y	V	G	F	T	V	E	C	F	E	F	T	R	R	A	A	F	C	E	Q	A	K	C	E	R	Y	R	E	C	Y	V	L	E	K	Y
W92	Aovo-DRB*W9202	50.0	H	N	Y	V	G	F	T	V	E	C	Y	D	F	D	R	R	A	S	Y	C	F	Q	T	T	S	D	R	Y	R	F	S	Y	F	V	V	K	L
	Aovo-DRB*W9201	50.0	H	N	Y	V	G	F	T	V	E	C	Y	D	F	D	R	R	A	S	Y	C	F	Q	T	T	S	D	R	Y	R	F	S	Y	V	V	V	K	L
W93	Aovo-DRB*W9301	100.0	H	N	Y	V	V	F	T	V	E	C	F	E	F	D	R	R	A	A	Y	C	E	L	I	K	F	E	R	Q	R	E	F	Q	Y	L	D	S	W
Ao-DRB1*03GA	Aona-DRB1*0305GA	**42.9**	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aona-DRB1*0307GA	**14.3**	H	N	Y	V	G	F	T	V	E	C	Y	D	F	R	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aona-DRB1*0303GA	7.1	H	N	Y	V	F	F	T	V	E	C	Y	D	F	Q	K	R	G	Q	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aona-DRB1*0304GA	7.1	H	N	Y	V	G	F	T	V	E	C	F	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aona-DRB1*0309GA	7.1	H	N	Y	V	G	F	T	V	E	C	F	D	F	R	K	R	G	Q	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aona-DRB1*0311GA	7.1	H	N	Y	V	V	F	T	V	E	C	F	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aona-DRB1*0312GA	7.1	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aona-DRB1*0319GA	7.1	H	N	Y	V	G	F	T	V	E	C	H	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aoni-DRB1*0303GA	**33.3**	H	N	Y	V	G	F	T	V	E	C	Y	D	F	R	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aoni-DRB1*0304GA	**33.3**	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aoni-DRB1*0301GA	16.7	H	N	Y	V	G	F	T	V	E	C	Y	D	F	R	K	R	G	Q	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aoni-DRB1*0307GA	16.7	H	N	Y	V	G	F	T	V	E	C	H	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aotr-DRB1*0303GA	33.3	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	A	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aotr-DRB1*0301GA	33.3	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aotr-DRB1*0302GA	33.3	H	N	Y	V	G	F	T	V	E	C	Y	D	F	R	K	R	G	Q	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aovo-DRB1*0302GA	**28.6**	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aovo-DRB1*0305GA	**28.6**	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	F	V	D	Y	W
	Aovo-DRB1*0301GA	14.3	H	N	Y	V	G	F	T	V	E	C	Y	D	F	R	K	R	G	Q	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	F	V	D	Y	W
	Aovo-DRB1*0303GA	14.3	H	N	Y	V	G	F	T	V	E	C	Y	H	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	H	R	Y	K	F	S	Y	Y	V	D	Y	W
	Aovo-DRB1*0306GA	14.3	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	F	Q	T	T	S	D	R	Y	K	F	S	Y	F	V	D	Y	W
Ao-DRB1*03GB	Aona-DRB1*0302GB	**72.7**	Y	N	Y	V	A	F	T	V	E	C	F	D	F	E	R	R	A	L	Y	C	F	Q	T	T	S	D	R	Y	R	F	S	Y	Y	V	S	Y	W
	Aona-DRB1*0301GB	**18.2**	Y	N	Y	V	A	F	T	V	E	C	Y	D	F	E	R	R	A	L	Y	C	F	Q	T	T	S	D	R	Y	R	F	S	Y	Y	V	S	Y	W
	Aona-DRB1*0326GB	9.1	Y	N	Y	V	A	F	T	V	E	C	F	D	F	E	R	R	A	L	Y	C	F	Q	T	T	Y	D	R	Y	R	F	Y	Y	Y	V	S	Y	W
Ao-DRB1*03GC	Aona-DRB1*0313GC	50.0	H	N	Y	V	V	F	T	V	E	C	F	D	F	E	T	R	A	A	Y	C	F	Q	T	T	S	D	R	Y	T	F	S	Y	Y	V	E	Y	F
	Aona-DRB1*0314GC	50.0	H	N	Y	V	V	F	T	V	E	C	Y	D	F	E	T	R	A	A	Y	C	F	Q	T	T	S	D	R	Y	T	F	S	Y	Y	V	E	Y	F
	Aoni-DRB1*0305GC	50.0	H	N	Y	V	V	F	T	V	E	C	Y	D	F	E	T	R	A	A	Y	C	F	Q	T	T	S	D	R	Y	T	F	S	Y	Y	V	D	Y	F
Ao-DRB3*06	Aoaz-DRB3*0601	100.0	H	N	Y	V	F	F	T	V	E	C	Y	D	F	Q	K	R	G	Q	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	Y	V	D	Y	W
	Aona-DRB3*0603	**36.8**	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	Y	V	D	Y	W
	Aona-DRB3*0601	**15.8**	H	N	Y	V	F	F	T	V	E	C	Y	D	F	Q	K	R	G	Q	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	V	V	D	Y	W
	Aona-DRB3*0613	**15.8**	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	V	V	D	Y	W
	Aona-DRB3*0602	**10.5**	H	N	Y	V	F	F	T	V	E	C	Y	D	F	Q	K	R	G	Q	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	Y	V	D	Y	W
	Aona-DRB3*0604	5.3	H	N	Y	V	G	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	Y	V	D	Y	W
	Aona-DRB3*0607	5.3	H	N	Y	V	F	F	T	V	E	C	Y	D	F	Q	K	R	G	Q	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	S	V	D	Y	W
	Aona-DRB3*0618	5.3	H	N	Y	V	F	F	T	V	E	C	Y	Y	F	Q	K	R	G	Q	Y	C	E	L	V	K	H	Y	R	Y	K	E	H	Y	Y	V	D	Y	W
	Aona-DRB3*0624	5.3	Y	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	Y	V	D	Y	W
	Aoni-DRB3*0601	100.0	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	Q	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	Y	V	D	Y	W
	Aotr-DRB3*06L	100.0	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	V	V	D	Y	W
	Aovo-DRB3*0601	100.0	H	N	Y	V	V	F	T	V	E	C	Y	D	F	Q	K	R	G	R	Y	C	E	L	V	K	H	D	R	Y	K	E	H	Y	Y	V	D	Y	W

Anexo 2. TCR-contacting residues orientation and HLA-DRβ* binding preference determine long-lasting protective immunity against malaria

Alba MP, Suarez CF, Varela Y, Patarroyo MA, Bermudez A, Patarroyo ME. TCR-contacting residues orientation and HLA-DRβ* binding preference determine long-lasting protective immunity against malaria. Biochem Biophys Res Commun. 2016;477(4):654-60.

La versión publicada del artículo puede ser consultada en:

http://www.sciencedirect.com/science/article/pii/S0006291X16310336

TCR-contacting residues orientation and HLA-DRβ* binding preference determine long-lasting protective immunity against malaria

Martha P. Alba [a, b, c], Carlos F. Suarez [a, b, c], Yahson Varela [a], Manuel A. Patarroyo [a, b], Adriana Bermudez [a, b], Manuel E. Patarroyo [a, d, *]

[a] Fundación Instituto de Inmunología de Colombia (FIDIC), Bogotá D.C., Colombia

[b] Universidad del Rosario, Bogotá D.C., Colombia

[c] Universidad de Ciencias Aplicadas y Ambientales (UDCA), Bogotá, Colombia

[d] Universidad Nacional de Colombia, Bogotá DC, Colombia.

* Corresponding author. e-mail: mepatarr@gmail.com

Abstract

Fully-protective, long-lasting, immunological (FPLLI) memory against *Plasmodium falciparum* malaria regarding immune protection-inducing protein structures (IMPIPS) vaccinated into monkeys previously challenged and re-challenged 60 days later with a lethal *Aotus* monkey-adapted *P. falciparum* strain was found to be associated with preferential high binding capacity to HLA-DRβ1* allelic molecules of the major histocompatibility class II (MHC-II), rather than HLA-DRβ3*, β4*, β5* alleles. Complete $PPII_L$ 3D structure, a longer distance (26.5 Å ± 1.5 Å) between residues perfectly tting into HLA-DRβ1*PBR pockets 1 and 9, a *gauche*$^-$ rotamer orientation in p8 TCR-contacting polar residue and a larger volume of polar p2 residues was also found. This data, in association with previously-described p3 and p7 apolar residues having *gauche*$^+$ orientation to form a perfect MHC-II-peptide-TCR complex, determines the stereo-electronic and topochemical characteristics associated with FPLLI immunological memory.

Keywords:

Antimalarial-vaccine, T-cell-receptor, MHC-II, Immunological memory, Rotamer-orientation.

Introduction

One of the main problems in vaccine development is the in- duction of FPLLI memory. Microbes (viruses, bacteria, parasites, etc.) have developed an incredible number of escape mechanisms against immune pressure, such as antigenic diversity where a single amino acid (aa) mutation or replacement can completely avert previously developed immunity, as occurs with *Plasmodium falciparum* malaria proteins apical membrane antigen-1 (AMA-1) [1,2], merozoite surface protein-1 (MSP-1) [3], etc., to quote a few. Microbes can also induce suppression, blocking [4], impeding [5,6] and many other escape mechanisms [7] rendering new or previously acquired immunity useless.

In continents like Africa, the development of FPLLI poses a tantalizing and insurmountable problem as one person can receive as many as eighteen *P. falciparum* infectious mosquito bites per day during the high transmission season. The putative vaccine candidate RTS,S/ASO1 provides a clear example [8], since the suggested protective immunity (considering protection to be less than 5000 parasites per microliter of blood) was short-lived (less than 6 months) [8] and observed in only 27% of the vaccinated population after the fourth booster immunisation 6 months later [9]. The WHO thus did not recommend its use for infants [9].

For more than three decades, we have pursued the idea that fully-protective immunity: zero parasites in the blood or spontaneous rapid and permanent recovery after very low parasitaemia (less than 0.1%) can be induced with chemically-synthesized vaccines, based on the concept that functionality relevant conserved high activity binding peptides (cHABP) have to be recognized in the corresponding [10] protein to properly modify them (mHABP) and render them highly immunogenic and protection-inducing [11]. Such minimal subunit-based mHABPs must ful l a

set of physicochemical and topochemical rules (previously described) to properly display a perfect fitting into MHCII-pep-TCR complex [12].

That goal was achieved when a large number of highly immunogenic protection-inducing peptide structures (IMPIPS) [13] fulfilled those requirements when used as individual epitopes in primary challenges.

Therefore, these merozoite-derived IMPIPS which had demonstrated clear FPLLI against experimental challenge with the highly- infectious *Aotus* monkey-adapted *P. falciparum* FVO strain were used to solve the immunological memory problem. Protected monkeys and some non-protected ones kept in captivity after challenge, after all of them had received anti-malarial treatment (to clear any residual parasites), were then re-challenged 60 days later (after all traces of anti-malarial drugs had disappeared) to determine the development of FPLLI. By the same token, sera from *Aotus* monkeys immunized with Spz-derived IMPIPS and kept in captivity for up to 900 days (~2 ½ years) after the first immunization were analysed for the presence of very high long-lasting antibody (VHLLA) titres against *P. falciparum* Spz, as determined by immunofluorescence assay (IFA), and their corresponding recombinant proteins by western blot (WB), to determine antibody titre duration [14].

Materials and methods

IMPIPS

mHABPs were synthesized according to Merrifield's peptide synthesis methodology, as modified by Houghten and thoroughly described [10]; a 600 MHz spectrometer was used for determining the 1H NMR 3D structure of a large panel of mHABPs [11].

Monkeys

Wild-caught *Aotus* monkeys from the Amazon jungle were used for trials authorized by Colombian environmental authorities (CORPOAMAZONIA, permission number 0632 and 0042/2010); they were kept in our eld-station in Leticia (Amazon department capital), looked after by expert veterinarians and workers supervised weekly by expert biologists and veterinarians from the local environmental authorities and ethics committee. After the study was completed, they were treated with paediatric doses of quinine, kept in quarantine for 20 more days and released back into the jungle close to their capture site, accompanied by environmental authority of cials. Those participating in this trial were kept according to methods above described.

Immunization

After arriving at our eld station, monkeys were deparasitized, kept in quarantine for twenty days and fed on a hypercaloric, hyperproteic diet before experiments commenced. Each monkey received 150 mg polymerized IMPIPS subcutaneously, in complete Freund's adjuvant, on day zero; a second dose of the same IMPIPS with incomplete Freund's adjuvant was administered 20 days later. They were challenged 20 days later.

First challenge

This involved intravenous inoculating 100,000 erythrocytes infected with the highly-virulent *Aotus*-adapted *P. falciparum* FVO strain freshly obtained from another infected *Aotus* monkey [11]; intravenous challenge with a 100% infectious, virulent *P. falciparum* malaria strain being the most stringent vaccine testing methodology.

Assessing infection

Parasitaemia was determined by uorescence microscopy using Acridine Orange staining; the percentage of parasitized RBC in their blood was counted, starting on day ve. Protected monkeys in their rst or primary challenge had no parasites in their blood while non-protected ones started showing parasites by day ve, reaching ≥6% parasitaemia on days eight to ten; they were immediately treated with paediatric doses of chloroquine. All protected and non-protected monkey; were treated after the experiment ended (day 20 after challenge) and kept in quarantine.

Determining antibodies

IFA titres were determined as previously described [11], blood samples being taken one day prior to the first immunization (i.e. preimmune - PI) or ten days after the second dose (II_{10}) and 20 days after the second immunization (II_{20}), the day before challenge.

Total schizont lysate or recombinant proteins containing the aa sequence from which the IMPIPS were derived were used for WB.

Re-challenge

No further immunizations were performed after the second dose was given at the beginning of the experiment. All protected and some non-protected monkeys were kept in quarantine for a further 60 days and re-challenge with 100,000 iRBC freshly-obtained from another previously-infected monkeys parasitaemia was assessed as before. Two trials (A and B) were performed with two different groups of IMPIPS used for immunization in the rst challenge.

HLA-DRβ* binding IMPIPS

The NetMHCIIpan-3.0 algorithm (predictor of peptide binding to MHC-II molecules), having >95% speci city and 90% sensitivity accurately, predicting (≥90%) correct HLA-DR peptide binding cores (previously determined by X-ray crystallography) was used. This in silico method identi es peptides having very high theoretical binding to speci c HLA-DRβ1* alleles and alternative β-chain isotypes like HLA-DRβ3*, β4* and β5* alleles measured as peptides half inhibitory capacity (IC ≤ 100 nM), based on the Immune Epitope Database [15].

Determining 3D structure

600 MHz spectrometer 1H NMR 3D structures were determined with RP-HPLC-puri ed IMIPS; their sequential connectivities and dihedral angles have already been described [13,14,16 - 19]. Only χ^1 angle degrees of residues considered TCR contacting (positions p2, p5, p8) regarding their binding in the HLA-DR peptide binding region (PBR) are described, based on their predicted binding to HLA- DR molecules. For other very relevant TCR contacting residues (p3, p7) their rotameric orientation and relevant immunological functions have been already described [14].

Results and discussion

Reminder: All participating monkeys were immunized only twice with a single IMPIPS; immune protection was therefore elicited by just two doses of individual IMPIPS.

Antibodies

Remarkably, Group I (protected) and Group II (non-protected) antibody (Ab) patterns, titres and reactivities (assessed by IFA and WB) were extremely similar prior to the first-challenge (Table 1), as can be appreciated when comparing cHABP 4044-derived MSP-2 24112 (protected) and 22774 (non-protected) analogue mHABPs (Table 1) as assessed by IFA and WB (Figure 1B, *Aotus* 16087 and 12877 respectively). Similarly the Ab reactivity by IFA of SERA-5 6725-derived 22830 (protected) and 24216 (non-protected) derived from 6746 were very similar by WB analysis (not shown).

It is thus extremely difficult to distinguish between permanent long-lasting protective epitopes and permanent short-protective ones based on actual Ab reactivity; such thoroughly-described phenomenon shows the exquisite reactivity of the immune response regarding FPLLI induction.

Furthermore, IFA, ELISA or WB serological analysis involving recombinants fragments prior to high malarial transmission seasons have shown that the bulk of immune response is directed against highly polymorphic, hypervariable regions of the molecule, the same occurs when immunizing humans or experimental animals with X-ray attenuated whole Spz, recombinant proteins, DNA vector based fragments, etc., showing that polymorphism is a very common mechanism used by microbes to escape immune pressure. Such approach (immunological) to

epitope selection has been exhaustively shown to be inappropriate in countless human vaccine trials [20] due to skewing the immune response towards highly polymorphic hypervariable regions.

Immunogenetic analysis

Genetic restriction ascribed to a particular HLA-DRβ1* allele represents an alternative to such long-lasting protective response but it is extremely dif cult to ascertain due to the tremendous polymorphism this region displays. The NetMHCIIpan-3.0 algorithm revealed no preference for any HLA-DRβ1* allele, since the same alleles were present in both groups (I and II) but showing a skewing towards binding to alternative β-chain HLA-DRβ3*, β4* and β5* alleles in the non-protected group II (Table 2). Such preference deserves further analysis.

Protection against re-challenge

Two trials (A and B) performed with different Mrz-derived IMPIPS to cover the MHC-II genetic restriction, trying to address memory or FPLLI phenomena, produced similar results (Figure 2) when these previously protected monkeys were re-challenge.

Three IMPIPS-induced FPLLI: 1585 MSP-1-derived 22770 (*Aotus* 12824), 6737 SERA-5-derived 22834 (*Aotus* 12984) and 4044 MSP-2-derived 24112 (*Aotus* 16006 and 16087) in some immunized monkeys having complete absence of parasites in their blood during the whole trial. All these IMPIPS showed high binding capacity to HLA-DRβ1* alleles but none bound to HLA-DRβ3*, β4* or β5* alleles.

Short-lived (~5 days), very low parasitaemia (<0.1%) that spontaneously recovered, not showing any more parasites during the rest of the experiment was seen in some previously-protected monkeys participating in re-challenge trial involving other IMPIPS (cHABP 4313 AMA-1-derived

22780, cHABP 6725 SERA-5-derived 22830, and cHABP 1783 EBA-175-derived 22814). Therefore, they were considered protective IMPIPS since parasitaemia was very low and rapidly cleared being this behaviour totally different to the well-known semi-immune chronicity phenomena. The latter two IMPIPS: 22830 and 22814 bound with high capacity to HLA-DRβ1* molecules and simultaneously to HLA-DRβ5*0101/ 0102 and HLA-DRβ5*0202 alleles (Table 2).

Another striking nding that correlates with the previous observation is that all non-protection inducing in re-challenge IMPIPS (group II) display shorter (22.5 Å ± 1.5 Å) structures (Table 1) as determined by 1H NMR (Figure 1C. IMPIPS 22774.47 and 24216.48, for example) when compared with all group I IMPIPS having 26.5 Å ± 1.5 Å (Table 1) distances between residues tting into HLA-DRβ1* PBR pockets 1 to 9 (Figure 1C. IMPIPS 24112.39 and 25608.37, for example). Group I IMPIPS totally displayed complete polyproline type II left-handed ($PPII_L$) structures while group B displayed a mixture of α-helical and $PPII_L$ structures, making them ±3.0 Å shorter. Such clear and neat difference had not been observed previously, due to the fact that re-challenge experiments had not been performed beforehand; therefore, our previously reported distances for IMPIPS were 26.5 Å ± 3.5 Å which included both groups (I and II).

Most monkeys which were not protected in re-challenge trials displayed greater binding capacity to HLA-DRβ3*, β4* or β5* alleles (Table 2), suggesting these IMPIPS clear skewing regarding their binding to these MHC-II alleles. It might be speculated that such preferential HLA-DRβ3*, β4* or β5* binding could bias the immune response towards short-lived memory protective immunity. Supporting such information, we have previously shown that peptides inducing short-lived antibody responses against *P. falciparum* malaria have shorter structure registers between aa tting into the HLA-DRβ1* peptide binding region (PBR) as determined by ^{1}H NMR spectrometry and are read in a different MHC-II functional register [21].

X-ray crystallography has shown that HLA-DRβ3* molecules are 2.0 Å wider in Kβ71 than DRβ1*, that Wβ61 is rotated 90 and more distant from pocket 9, that α76R is notably displaced upwards leaving pocket 9 highly hydrophobic, that H-bonds between peptide backbone atoms and DRβ3* interacting residues are >4 Å distant, making these interaction between DRβ3*-IMPIPS longer, unstable and weaker for stimulating an appropriate immune response. All of these stereo chemical characteristics could probably be associated with short memory induction [22,23].

Since IMPIPS cannot be involved in Spz challenge, due to irreproducible results regarding the only Anopheles mosquito-derived *P. falciparum* strain (Santa Lucia) adapted to *Aotus* monkeys, such antibodies' permanence in Spz-derived IMPIPS immunized monkey sera was determined by IFA and WB with recombinant fragments corresponding to the protein from which the aa sequence was derived. Monkeys, kept in our eld station in the Amazon jungle for 900 days after the 1st vaccination, followed-up for 840 days after the 3rd dose (~2½ years) with IMPIPS CSP-1 4383-derived 25608; 4389-derived 32958 and STARP 24230 20546-derived produced very high and long-lasting Ab titres (Figure 1B). Some others like 3289-derived TRAP 24246 and SPECT-2 34938 derived 38890 had high Ab titres that slowly declined over a 6-month period. These short-lived antibody inducer mHABP also had high binding to HLA-DRβ5*0202 allele molecules.

p2 volume in long-lasting protective immunity

Besides the distance between P1 and P9 residues, and ϕ and φ angles having $PPII_L$ conformation, we have found volume and charge to be critical physicochemical characteristics for a proper t into the HLA-DRβ1* PBR. Something similar occurs with upwardly-orientated TCR-contacting residues, as previously shown for p3 and p7 [14]. Table 1 in the present manuscript clearly shows that most FPLLI- and VHLLAI-inducing IMPIPS in group I had a larger volume in p2 than those

in groups II, whereas positively-charged residues having p electrons (H, R, K) predominated in group I. Smaller polar residues predominating in group II had alcohol groups (S, T) in their side chains acting as nucleophiles or acidic negatively-charged aa (E, D).

p8 residue orientation determines long-lasting protective immunity

Protein and peptide studies have thoroughly demonstrated that aa side-chain orientation has trimodal distribution based on $\chi 1$ angle rotation related to a protein or peptide's frontal plane; *gauche*$^+$ (trans to the carbonyl group), *gauche*$^-$ (trans to the Hα atom) or trans (trans to the amino group), except for Gly, Ala and Pro, the later (warning) an iminoacid having different $\chi 1$ angle rotation, depending on the preceding's residue f angle. According to $\chi 1$ angle rotation degree, aa have been divided into *gauche*$^+$ (-120° to 0°), *gauche*$^-$ (0° to +120°) and trans (trans +120° to +240°). Therefore when the 600 MHz 3D structures of our IMPIPS used for immunization were determined it was found that, strikingly, all protected *Aotus* monkeys during re-challenge had been vaccinated with IMPIPS having $\chi 1$ angles ranging from +89.9° to +8.1° in residues located in p8 (Table 2), therefore having *gauche*$^-$ aa side-chain orientation in p8. By contrast, all non-protected monkeys in re-challenge had been immunized with IMPIPS having -167.1° to -12.3° rotation angles, therefore *gauche*$^+$ orientation in p8 (Table 2).

VHLLAI Spz-derived IMPIPS (25608, 32958 and 24320) and Mrz- derived FPLLI 24112 included in mixtures [14] not blocking, interfering or suppressing each other's activity had also *gauche*$^-$ sidechain orientation in p8.

When analysing the aa sequence of IMPIPS used to immunise re-challenge protected monkeys, p8 was occupied by polar residues (S, T, E, D), the same as those Spz derived VHLLAI IMPIPS (N,

N, P), except for AMA-1-derived 22780 and STARP-derived 24320 both having the iminoacid Pro (which could be puckered up or down) and 22770 having Val in this position (Table 1. Group I). Strikingly, all non-protected monkeys during re-challenge were immunised with IMPIPS having apolar residues in p8 (M, I, M, L, N, G, A, F), except for 38890 (E) (Table 1) including the Spz derived IMPIPS 24246 and 38890 inducing short lived Abs titres.

Our previous data regarding IMPIPS previously reported 3D structures has shown the critical role of $\chi 1$ angle in residues p3 and p7, having *gauche*$^+$ orientation associated with being able to be mixed to induce FPLLI and VHLLAI without interfering, blocking, suppressing, abolishing or poisoning each other's immunological activity in the process of developing a complete multi-epitope, multi-stage minimal subunit-based, chemically-synthesized anti-malarial vaccine. Conversely, the completely abolished immune response induced in mixtures with other IMPIPS when mixing them i.e. 24148 and 24246 corresponds to the same IMPIPS which could not induce either re-challenge protection or VHLLAI memory; these IMPIPS also displayed *gauche*$^+$ orientation in p8 (Table 2. Group II) suggesting some stereo-chemical interference in memory induction and combination in mixture composition.

In intracellular pathogenic diseases, the development of polyfunctional, rapidly proliferating T-cells, with low apoptosis seems to be the key issue [24] to clear infection and develop a robust T-cell memory [25] and many hypotheses have arisen for explaining the absence of memory induction, i.e. T-cell exhaustion after infection [26] leading to the loss of parasite-speci c memory T-cells inducing protection from re-infection [27], (as in this manuscript), or alternative up-regulation of FOXP3 expressing $CD4^+$ $CD25^+$ T-regulatory cells associated with more rapid parasite growth during infection [28] or elevated number of highly suppressive T-regulatory cells in severe malaria [29].

Alternative explanations are the induction of programed cell death-1 (PD-1) molecules on activated $CD4^+$ or $CD8^+$ T-cells that in conjunction with LAG-3+ T-cells modulate immunity against malaria [30]. It has been recently demonstrated in mice having the PD-1 gene deleted (PD-1KO) that such deletion generates sterile protective immunity, unlike wild type mice infected with *Plasmodium chabaudi* which maintained ~1% parasitaemia [31] equivalent to human chronic subclinical malaria.

There are many more alternative hypothesis associated with the lack or absence of protective immunity memory but this 3D structural analysis of 20 IMPIPS clearly suggested that p8 residue χ1 angle rotation and orientation is associated with or determines long-lasting protective memory.

We therefore suggest that in a complete, fully-protective minimal subunit-based, chemically-synthesized vaccine able to induce very long-lasting protective immunological memory, besides the previously-described physicochemical principles regarding a perfect t, into the HLA-DRβ1*PBR, TCR-contacting residues p3 (apolar) and p7 (also apolar) should have *gauche*$^+$ rotamer orientation [14] while p8 (polar) should have *gauche*$^-$ orientation and p2 should have the polar characteristics shown here. These ndings allow us to propose that such stereo chemical and topological rules mediate FPLLI memory.

This is the rst time protective memory induction has been shown, at 3D structural level to be associated with speci c electronic and rotamer orientation of a particular TCR-contacting residue (p8) while negatively associated also with a binding capacity to HLA-DRβ3*, β4* or β5* allelic molecules, paving the way for a logical rational methodology for long-lasting protective immunity.

Con ict of interest

The authors declare that they have no nancial or commercial con icts of interest.

Acknowledgments

This research was supported by “The Colombian Science, Technology and Innovation Department (Colciencias)”, Contract RC#0309-2013.

We would like to thank Mr. Jason Garry for his collaboration in the translation of this manuscript.

References

[1] S. Dutta, S.Y. Lee, A.H. Batchelor, D.E. Lanar, Structural basis of antigenic escape of a malaria vaccine candidate, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 12488-12493.

[2] D.P. Eisen, A. Saul, D.J. Fryauff, J.C. Reeder, R.L. Coppel, Alterations in *Plasmodium falciparum* genotypes during sequential infections suggest the presence of strain speci c immunity, Am. J. Trop. Med. Hyg. 67 (2002) 8-16.

[3] W.D. Morgan, M.J. Lock, T.A. Frenkiel, M. Grainger, A.A. Holder, Malaria parasite-inhibitory antibody epitopes on *Plasmodium falciparum* merozoite surface protein-1(19) mapped by TROSY NMR, Mol. Biochem. Parasitol. 138 (2004) 29-36.

[4] W.D. Morgan, T.A. Frenkiel, M.J. Lock, M. Grainger, A.A. Holder, Precise epitope mapping of malaria parasite inhibitory antibodies by TROSY NMR cross- saturation, Biochemistry 44 (2005) 518-523.

[5] C.Q. Schmidt, A.T. Kennedy, W.H. Tham, More than just immune evasion: hijacking complement by *Plasmodium falciparum*, Mol. Immunol. 67 (2015) 71-84.

[6] J.Y.A. Doritchamou, VAR2CSA domain-speci c analysis of naturally acquired functional antibodies to *P. falciparum* placental malaria, J. Infect. Dis. (2016).

[7] F. Farooq, E.S. Bergmann-Leitner, Immune escape mechanisms are *Plasmodium*'s secret weapons foiling the success of potent and persistently ef cacious malaria vaccines, Clin. Immunol. 161 (2015) 136-143.

[8] Ef cacy and safety of RTS, S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: nal results of a phase 3, individually randomised, controlled trial, Lancet 386 (2015) 31-45.

[9] W.H.O. (WHO), Malaria vaccine, Wkly. Epidemic 9 (2016) 33-52.

[10] L.E. Rodriguez, H. Curtidor, M. Urquiza, G. Cifuentes, C. Reyes, M.E. Patarroyo, Intimate molecular interactions of *P. falciparum* merozoite proteins involved in invasion of red blood cells and their implications for vaccine design, Chem. Rev. 108 (2008) 3656-3705.

[11] M.E. Patarroyo, A. Bermudez, M.A. Patarroyo, Structural and immunological principles leading to chemically synthesized, multiantigenic, multistage, minimal subunit-based vaccine development, Chem. Rev. 111 (2011) 3459-3507.

[12] M.A. Patarroyo, A. Bermudez, C. Lopez, G. Yepes, M.E. Patarroyo, 3D analysis of the TCR/pMHCII complex formation in monkeys vaccinated with the rst peptide inducing sterilizing immunity against human malaria, PLoS One 5 (2010) e9771.

[13] M.E. Patarroyo, A. Bermudez, M.P. Alba, M. Vanegas, A. Moreno-Vranich, L.A. Poloche, M.A. Patarroyo, IMPIPS: the immune protection-inducing protein structure concept in the search for steric-electron and topochemical principles for complete fully-protective chemically synthesised vaccine development, PLoS One 10 (2015) e0123249.

[14] A. Bermudez, D. Calderon, A. Moreno-Vranich, H. Almonacid, M.A. Patarroyo, A. Poloche, M.E. Patarroyo, *Gauche*(+) side-chain orientation as a key factor in the search for an immunogenic peptide mixture leading to a complete fully protective vaccine, Vaccine 32 (2014) 2117-2126.

[15] M. Andreatta, E. Karosiene, M. Rasmussen, A. Stryhn, S. Buus, M. Nielsen, Accurate pan-speci c prediction of peptide-MHC class II binding af nity with improved binding core identi cation, Immunogenetics 67 (2015) 641-650.

[16] M.E. Patarroyo, A. Moreno-Vranich, A. Bermudez, Phi (Phi) and psi (Psi) angles involved in malarial peptide bonds determine sterile protective immunity, Biochem. Biophys. Res. Commun. 429 (2012) 75-80.

[17] M.E. Patarroyo, A. Bermudez, M.P. Alba, The high immunogenicity induced by modi ed sporozoites' malarial peptides depends on their phi (varphi) and psi (psi) angles, Biochem. Biophys. Res. Commun. 429 (2012) 81-86.

[18] M.E. Patarroyo, M.A. Patarroyo, L. Pabon, H. Curtidor, L.A. Poloche, Immune protection-inducing protein structures (IMPIPS) against malaria: the weapons needed for beating Odysseus, Vaccine 33 (2015) 7525-7537.

[19] M.E. Patarroyo, G. Arevalo-Pinzon, C. Reyes, A. Moreno-Vranich, M.A. Patarroyo, Malaria parasite survival depends on conserved binding peptides' critical biological functions, Curr. Issues Mol. Biol. 18 (2015) 57-78.

[20] S. Li, M. Plebanski, P. Smooker, E.J. Gowans, Editorial: why vaccines to HIV, HCV, and malaria have So far failed-challenges to developing vaccines against immunoregulating pathogens, Front. Microbiol. 6 (2015) 1318.

[21] M.E. Patarroyo, M.P. Alba, L.E. Vargas, Y. Silva, J. Rosas, R. Rodriguez, Peptides inducing short-lived antibody responses against *Plasmodium falciparum* malaria have shorter structures and are read in a different MHC II functional register, Biochemistry 44 (2005) 6745-6754.

[22] C.S. Parry, J. Gorski, L.J. Stern, Crystallographic structure of the human leukocyte antigen DRA, DRB3*0101: models of a directional alloimmune response and autoimmunity, J. Mol. Biol. 371 (2007) 435-446.

[23] S. Dai, F. Crawford, P. Marrack, J.W. Kappler, The structure of HLA-DR52c: comparison to other HLA-DRB3 alleles, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 11893-11897.

[24] J.R. Lukens, M.W. Cruise, M.G. Lassen, Y.S. Hahn, Blockade of PD-1/B7-H1 interaction restores effector CD8+ T cell responses in a hepatitis C virus core murine model, J. Immunol. 180 (2008) 4875-4884.

[25] E.J. Wherry, T cell exhaustion, Nat. Immunol. 12 (2011) 492-499.

[26] M.N. Wykes, J.M. Horne-Debets, C.Y. Leow, D.S. Karunarathne, Malaria drives T cells to exhaustion, Front. Microbiol. 5 (2014) 249.

[27] R. Stephens, J. Langhorne, Effector memory Th1 CD4 T cells are maintained in a mouse model of chronic malaria, PLoS Pathog. 6 (2010) e1001208.

[28] M. Walther, J.E. Tongren, L. Andrews, D. Korbel, E. King, H. Fletcher, R.F. Andersen, P. Bejon, F. Thompson, S.J. Dunachie, F. Edele, J.B. de Souza, R.E. Sinden, S.C. Gilbert, E.M. Riley, A.V. Hill, Upregulation of TGF-beta, FOXP3, and CD4þCD25þ regulatory T cells correlates with more rapid parasite growth in human malaria infection, Immunity 23 (2005) 287-296.

[29] G. Minigo, T. Woodberry, K.A. Piera, E. Salwati, E. Tjitra, E. Kenangalem, R.N. Price, C.R. Engwerda, N.M. Anstey, M. Plebanski, Parasite-dependent expansion of TNF receptor II-positive regulatory T cells with enhanced sup- pressive activity in adults with severe malaria, PLoS Pathog. 5 (2009) e1000402.

[30] N.S. Butler, J. Moebius, L.L. Pewe, B. Traore, O.K. Doumbo, L.T. Tygrett, T.J. Waldschmidt, P.D. Crompton, J.T. Harty, Therapeutic blockade of PD-L1 and LAG-3 rapidly clears established blood-stage *Plasmodium* infection, Nat. Immunol. 13 (2012) 188-195.

[31] J.M. Horne-Debets, D.S. Karunarathne, R.J. Faleiro, C.M. Poh, L. Renia, M.N. Wykes, Mice lacking Programmed cell death-1 show a role for CD8(þ) T cells in long-term immunity against blood-stage malaria, Sci. Rep. 6 (2016) 26210.

Table legends

Table 1.

IMPIPS molecule of origin and our laboratory's serial number in bold; below the native cHABP number, aa sequence; distance between the farthest atoms in pockets 1 and 9, measured in Å; (NA = not-applicable), antibody titres as assessed by IFA, the pre x the number of monkeys displaying such titre, PI = pre-immune, 20 days after the second dose (II20) and performance after rst challenge including the number of fully protected monkeys and those protected after rechallenge (+ o -). Colours indicate residues tting into HLA-DRβ1* PBR pockets: fuchsia pocket 1, blue pocket 4, orange pocket 6 and green pocket 9. TCR-contacting residues in this study (p2, p5, p8) are indicated.

Table 2.

IMPIPS inducing merozoite-FPLLI or sporozoite-VHLLAI. HLA-DRβ1* or β3*, β4*, β5* alleles binding activity and their IC below in parenthesis based on the NetMHCIIpan-3.0 method. According to their PBR register, TCR-contacting residues p2, p5, p8 side-chain c1 angles are described.

Figure Legends

Figure 1.

A. Immuno uorescence patterns recognised by sera from Aotus monkeys immunised with speci c IMPIPS and determined by immuno uorescence. MSP-2 and MSP-1 detected on the membrane

surface proteins; SERA-5, serine repeat antigen-5 intracytoplasmic; AMA-1, apical merozoite antigen: present on the apical and Mrz membrane; HRP-II histidine-rich protein II: identi ed as small intra-erythrocyte dots; EBA-175, erythrocyte binding antigen-175 present in micronemes. CSP-1 membranal circumsporozoite protein-1; SPECT-1 sporozoite microneme protein essential for cell traversal-1 identi ed in membrane and micronemal small dots; STARP sporozoite threonine and asparagine-rich protein, and TRAP thrombospondin-related anonymous protein, identi ed in rhoptries and micronemes. **B.** WB analysis of MSP-2 (4044) 24112 immunised and re-challenge protected monkeys compared to (4044) 22774 MSP-2 immunized re-challenge and non-protected monkey. **C.** IMPIPS lowest energy conformer 3D structure determined by 600 MHz 1H NMR identi ed by our serial number followed by dot corresponding to conformer number. Amino-acid colour based on HLA-DRβ1* binding activities, binding motifs, and binding registers as follows: pocket 1, fuchsia; p2, red; p3, turquoise; pocket 4, dark blue; p5, rose; pocket 6, light brown; p7, gray; p8, yellow and pocket 9, green. The distances between the farthest atoms of residues tting into pockets 1 and 9 are measured in angstroms (Å).

Figure 2.

Parasitaemia levels, percentage of infected RBC (%) displayed in a semi-logarithmic scale as assessed by AO staining in, monkeys participating in re-challenge trials A and B group I (protected); group II (non-protected) on days after re-challenge.

Table 1.

PROTEIN PEPTIDE		SEQUENCE	P1-P9 Distance (Å)	IFA TITERS PI	IFA TITERS II20	Protection First Challenge	Protection RE Challenge
		GROUP I					
		P1 p2 p5 p8 P9					
MSP-2	24112 4044	S K Y S N T F N I N A Y N M V I R R S M	26.5	0	1(1280)	II/8	+
AMA-1	22780 4313	G E D A E V A G T Q Y F H P S G K V P V F G	26.6	0	1(1280)	II/10	+
MSP-1	22770 1585	Y H L P L G G V Y R A L K K Q	26.6	0	2(640)	II/9	+
EBA-175	22814 1783	N D K L Y R M E Y W K T I K K D V W	25.9	0	2(5120)	I/10	+
SERA-5	22830 6725	L K M T N N A I S F M S P S S S L E K K	26.5	0	1(160)	I/10	+
SERA-5	22834 6737	D H I H V K M F K V I E N N D K S E L I	25.6	0	1(2560)	II/9	+
CSP-1	25608 4383	K N S F S L G E N P N A N P	27.5	0	2(2560)	ND	ND
CSP-1	32958 4388	G N G Q G L N M N N P P N F N V D E N A	27.1	0	3(1280)	ND	ND
STARP	24320 20546	V I K H M R F H A D Y Q A P F L G G G Y	NA	0	1(320)	ND	ND
		GROUP II					
		P1 p2 p5 p8 P9					
MSP-2	22774 4044	K N E S K Y S N T F E V N A Y N M V N R	22.3	0	1(2560)	II/10	-
MSP-1	24148 5501	M L N I S M L Q T V M M M T P Q K	19.0	0	1(2560)	II/8	-
EBA-175	22812 1779	N N D R I Y D M N H L M I K M H I L A I	21.5	0	1(2560)	I/9	-
EBA-175	24166 1818	F N N I P S R Y N L Y D K M L P L D D	21.7	0	1(160)	II/5	-
EBA-175	24150 1758	W K S Y S V D D N I P M N M S L I H K H	21.2	0	1(1280)	II/7	-
HRP-II	24230 6800	S A F D D N L T A A N A M G L I L N K R	21.6	0	2(320)	II/7	-
SERA-5	24216 6746	D Q G N T I T A W N R A A K F H L E T I	22.3	0	2(640)	II/8	-
TRAP	24246 3289	S P T S V T V G K G A F S F K A E	21.0	0	1(1280)	ND	ND
SPECT-2	38890 34938	S D Y T K A L A A E A K V S G S Y W G I	18.6	0	1(640)	ND	ND

Table 2.

		HLA-DRβ1*	HLA-DRβ3*,β5*	p2	p5	p8
				χ1	χ1	χ1
GROUP I						
MSP-2	**24112.39**	0101/0102 25.6-30.0	--	-83.7	+68.1	+66.5
AMA-1	**22780.43**	0101/0102/1303/1305 94.0-123.0-128.0-111.0	--	-170	Gly	Pro
MSP-1	**22770.15**	1330/1303/1425/1201 20.7-67.5-61.7-24.6	--	-173	+65	+89.9
EBA-175	**22814.42**	1303/1330 67.6-20.7	β5*0101/0102 29.0-29.6	-71	-146	+56.4
SERA-5	**22830.20**	0901/0902 76.5-53.5	β5*0202 23.6	-169	-174	+66.0
SERA-5	**22834.42**	0102 50.9	--	+64	-60.1	+56.9
CSP-1	**25608.37**	0101 278.6	--	-174	-58	+80.0
CSP-1	**32958.2**	1303 132.9	--	-77.7	-171.9	+8.1
STARP	**24320.18**	1303/1330 168.4-107.6	β3*0101/0201 56.2-170.1	-113	-166	+19.3
GROUP II						
MSP-2	**22774.36**	0101/0102 122.0-124.0	β5*0202 111,6	-173	Ala	-167.1
EBA-175	**22812.20**	1102/1301 94.2-94.2	β3*0201/0301 48.1-32.3	-168	-167	-68,9
MSP-1	**24148.7**	0101/0102/1201 17.7-57.0-8.2	β3*0301/β4*0401/β5*0202 67.4-75.8-93.0	+63.0	-94.3	-72.1
EBA-175	**24166.13**	1201/1330/1406 119.0-102.0-110.0	β5*0101/0102 29.0-74.0	-168	-59.3	-58.6
EBA-175	**24150.47**	0301 230.4	β3*0101 100.7	-170	-61	-62
HRPII	**24230.13**	0701/0101/0102 92.3-24.9-14.5	β5*0202 30.5	+18.5	+103	Gly
SERA-5	**24216.48**	1104/1106/0804 13.0-14.2-34.0	β4*0101 331.0	+55	-164	Ala
TRAP	**24246.41**	0102/0101 70.5-142.9	β5*0202 78.5	+178	+161	Ala
SPECT-2	**38890.32**	0901/0902/0101/0102 33.7-32.8-8.9-10.9	β5*0101/0102 88.1-68.0	-152	-68	-69.9

Figure 1

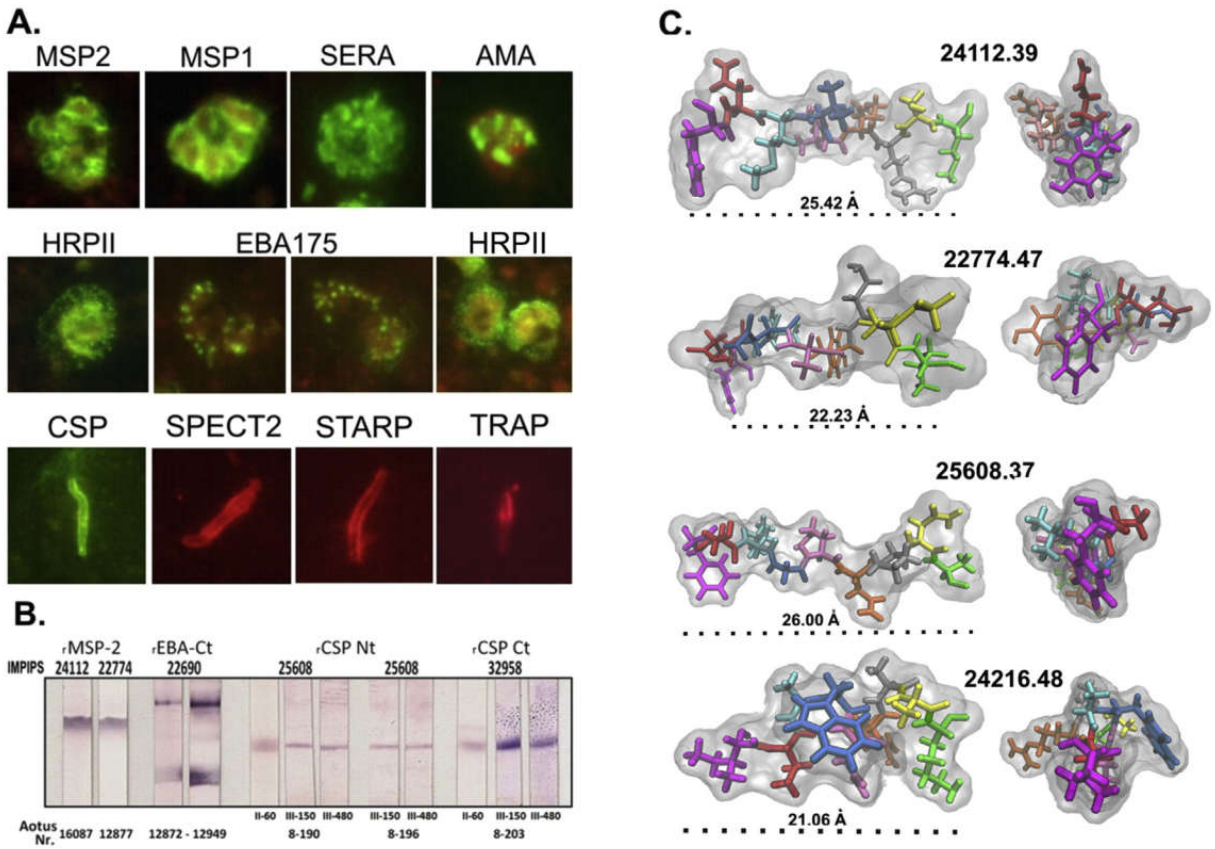
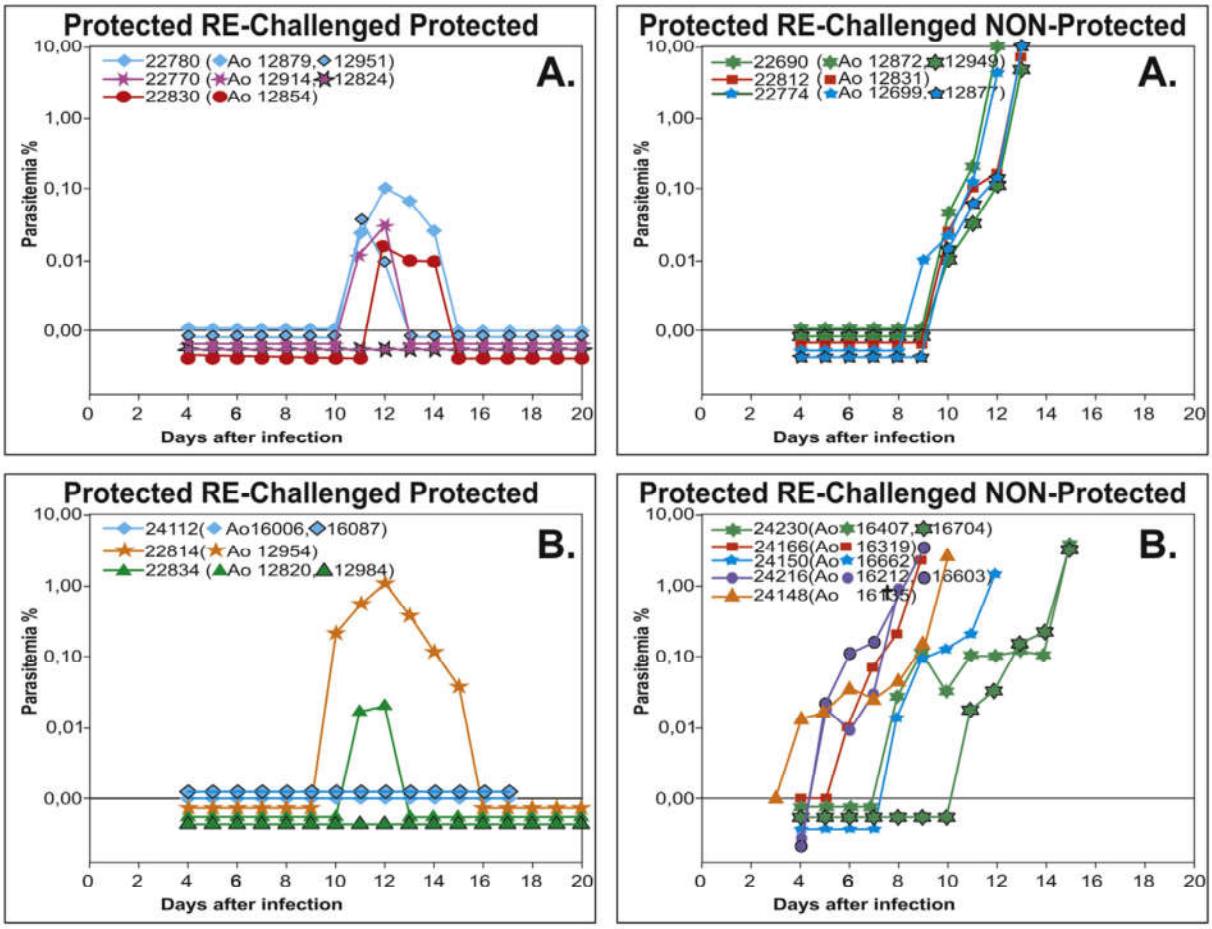

A.
MSP2
MSP1
SERA
AMA
HRPII
EBA175
HRPII
CSP
SPECT2
STARP
TRAP
B.
rMSP-2
rEBA-Ct
rCSP Nt
rCSP Ct
IMPIPS 24112 22774
22690
25608
25608
32958
II-60 III-150 III-480 III-150 III-480 II-60 III-150 III-480
Aotus Nr. 16087 12877
12872 - 12949
8-190
8-196
8-203
C.
24112.39
25.42 Å
22774.47
22.23 Å
25608.37
26.00 Å
24216.48
21.06 Å

Figure 2


Protected RE-Challenged Protected
A.
22780 (Ao 12879, 12951)
22770 (Ao 12914, 12824)
22830 (Ao 12854)
Parasitemia %
Days after infection
Protected RE-Challenged NON-Protected
A.
22690 (Ao 12872, 12949)
22812 (Ao 12831)
22774 (Ao 12699, 12877)
Parasitemia %
Days after infection
Protected RE-Challenged Protected
B.
24112(Ao16006, 16087)
22814(Ao 12954)
22834 (Ao 12820, 12984)
Parasitemia %
Days after infection
Protected RE-Challenged NON-Protected
B.
24230(Ao 16407, 16704)
24166(Ao 16319)
24150(Ao 16662)
24216(Ao 16212, 16603)
24148(Ao 16135)
Parasitemia %
Days after infection

Anexo 3. Estimación de la frecuencia en poblaciones humanas de los linajes alélicos del CMH-DRB

	Frecuencias linajes alelicos. HLA-DRB																					
	Aleut	Amerindian	Arab	Asian	Aust. Abor	Austronesian	Bashkir	Berber	Black	Caucasoid	Gypsy	Jew	Kurd	Melanesian	Mestizo	Micronesian	Mulatto	Oriental	Persian	Polynesian	Siberian	Tatar	Global
DRB1*01	21.2	2.4	9.8	8.3	0.9	1.4	15.1	25.7	14.0	20.3	13.5	16.5	8.0	0.3	19.3	0.0	18.9	4.7	4.9	2.7	4.0	22.2	17.0
DRB1*03	4.7	1.8	24.4	17.7	2.7	8.0	10.3	16.9	25.7	22.5	37.1	13.1	18.4	0.1	17.0	2.3	15.8	7.5	17.4	4.7	4.8	12.6	19.7
DRB1*04	35.3	65.2	31.8	18.7	23.0	12.6	23.3	19.1	9.9	25.6	15.1	33.9	28.2	32.4	31.0	35.7	23.7	26.7	20.8	50.5	33.0	17.8	26.1
DRB1*07	32.9	2.8	23.4	23.9	2.4	17.7	43.2	31.6	18.3	23.7	19.7	28.4	20.7	0.2	23.0	0.0	20.6	16.9	17.7	5.3	11.4	39.3	22.4
DRB1*08	18.8	35.8	4.5	6.6	58.2	3.3	5.5	14.0	11.8	6.7	1.2	2.4	1.7	29.4	15.7	17.8	14.5	14.9	2.6	23.3	12.1	5.9	8.1
DRB1*09	2.4	7.5	0.7	8.8	0.0	6.7	9.6	5.1	5.1	1.6	0.0	0.6	0.0	3.4	3.5	0.0	4.4	28.3	1.4	22.7	20.7	5.9	5.6
DRB1*10	2.4	0.6	7.6	10.5	0.3	3.8	1.4	5.1	5.0	2.4	5.8	5.0	4.6	0.0	4.1	0.0	3.5	2.8	4.6	1.6	2.7	3.0	2.8
DRB1*11	9.4	4.1	37.7	16.1	0.0	6.4	12.3	28.7	28.3	29.1	16.6	43.4	48.3	23.3	18.8	8.5	18.4	11.8	35.6	27.4	10.9	20.7	26.8
DRB1*12	12.9	0.6	1.7	11.0	14.9	58.7	16.4	0.0	8.0	2.1	1.2	4.3	2.3	5.1	2.6	67.4	6.6	25.6	1.6	37.4	18.2	4.4	5.9
DRB1*13	8.2	2.8	28.0	17.6	1.2	5.9	30.8	29.4	37.7	26.2	10.4	26.3	23.6	0.1	28.7	1.6	27.2	11.5	19.4	3.8	11.4	28.9	23.8
DRB1*14	21.2	50.9	7.2	13.0	80.6	8.7	8.9	0.0	3.5	4.8	38.2	7.3	7.5	14.8	10.4	19.4	10.1	13.2	9.1	23.6	34.6	6.7	6.7
DRB1*15	30.6	2.5	18.4	34.9	13.7	58.4	18.5	17.6	30.3	24.6	23.6	14.0	27.6	69.4	17.9	47.3	19.7	27.5	21.6	15.8	14.1	22.2	24.6
DRB1*16	0.0	21.6	2.8	2.5	0.9	8.0	4.8	2.2	2.5	4.6	14.3	3.9	9.2	21.7	7.6	0.0	5.7	5.0	7.5	0.3	2.3	10.4	4.8
N	85	4800	19145	6860	335	2760	146	136	8745	628401	259	23926	174	1560	15423	129	228	120744	15996	639	1409	135	853309

Global: Basado en la tipificación de 853309 individuos

Minería de datos a partir de: Allele Frequency Net Database (AFND). Nucleic Acid Research 2011 39:D913-D919. http://www.allelefrequencies.net/

Anexo 4. Uso de la metodología FMO-PIEDA en el análisis del efecto de mutaciones en proteínas

“New mutations in non-syndromic primary ovarian insufficiency patients identified via whole-exome sequencing”

Patiño LC, Beau I, Carlosama C, Buitrago JC, González R, Suárez CF, et al. New mutations in non-syndromic primary ovarian insufficiency patients identified via whole-exome sequencing. Human Reproduction. 2017:1-9.

La versión publicada del artículo puede ser consultada en:

https://academic.oup.com/humrep/article-abstract/32/7/1512/3823627/New-mutations-in-non-syndromic-primary-ovarian?redirectedFrom=fulltext

New mutations in non-syndromic primary ovarian insufficiency patients identified via whole-exome sequencing

Liliana Catherine Patiño[1], Isabelle Beau[2], Carolina Carlosama[1], July Constanza Buitrago[1], Ronald González[3], Carlos Fernando Suárez[3,4], Manuel Alfonso Patarroyo[3,5] Brigitte Delemer[6], Jacques Young[2,7], Nadine Binart[2], Paul Laissue[1,*]

[1]Center For Research in Genetics and Genomics (CIGGUR). GENIUROS Research Group. School of Medicine and Health Sciences. Universidad del Rosario. Bogotá, Colombia.

[2]Inserm 1185, Le Kremlin-Bicêtre, Université Paris-Saclay, Faculté de Médecine Paris Sud, Le Kremlin-Bicêtre, France; [3]Fundación Instituto de Inmunología de Colombia (FIDIC), Bogotá D.C., Colombia.;[4]Universidad de Ciencias Aplicadas y Ambientales (UDCA), Bogotá D.C., Colombia; [5]Basic Sciences Department, School of Medicine and Health Sciences, Universidad del Rosario, Bogotá D.C., Colombia; [6]Service d'Endocrinologie-Diabète-Nutrition, CHU de Reims-Hôpital Robert-Debré, Reims, France ; [7]APHP, Hôpital de Bicêtre, Service d'Endocrinologie et des Maladies de la Reproduction, Le Kremlin-Bicêtre, France.

***Correspondence address:** Paul Laissue MD, PhD, HDR, Center For Research in Genetics and Genomics (CIGGUR). GENIUROS Research Group. School of Medicine and Health Sciences. Universidad del Rosario. Bogotá, Colombia.

Address: Carrera 24 N° 63C-69, CP 112111, Bogotá DC, Colombia.

Tel : +5712970200; Fax : +5712970200; E-mail: paul.laissue@urosario.edu.co

Running Title: Mutations in primary ovarian insufficiency

Abstract

STUDY QUESTION: It is able to identify new mutations potentially associated to non-syndromic primary ovarian insufficiency (POI) via whole-exome sequencing (WES)?

SUMMARY ANSWER: WES is an efficient tool to study genetic causes of POI as we have identified new mutations, some of which lead to protein destabilisation potentially contributing to the disease aetiology.

WHAT IS KNOWN ALREADY: POI is a frequently occurring complex pathology leading to infertility. Mutations in only few candidate genes, mainly identified by Sanger sequencing, have been definitively related to the pathogenesis of the disease.

STUDY DESIGN, SIZE, DURATION: This is a retrospective cohort study performed on 69 women affected by POI.

PARTICIPANTS/MATERIALS, SETTING, METHODS: WES and an innovative bioinformatics analysis were used on non-synonymous sequence variants in a subset of 420 selected POI candidate genes. Mutations in BMPR1B and GREM1 were modelled by using fragment molecular orbital analysis.

MAIN RESULTS AND THE ROLE OF CHANCE:

Fifty-five coding variants in 49 genes potentially related to POI were identified in 33 out of 69 patients (48%). These genes participate in key biological processes in the ovary, such as meiosis, follicular development, granulosa cell differentiation/proliferation and ovulation. The presence of at least two mutations in distinct genes in 45% of the patients argued in favor of a polygenic nature of POI.

LARGE SCALE DATA: Exome data was uploaded at the Open Science Framework.

LIMITATIONS, REASONS FOR CAUTION: It would be possible that regulatory regions, not analysed in the present study, carry further variants related to POI.

WIDER IMPLICATIONS OF THE FINDINGS: WES and the *in silico* analyses presented here represent an efficient approach for mapping variants associated with POI etiology. Computational modelling of variants suggested a significant change in protein stability secondary to BMPR1B-p.Arg254His, BMPR1B-p.Phe272Leu and p.GREM1-p.Arg169Thr mutations. Taken together, our findings add valuable information regarding POI molecular origin. Sequence variants presented here represents potential future genetic biomarkers.

STUDY FUNDING/COMPETING INTERESTS: This study was supported by the Universidad del Rosario and Colciencias (Grants CS/CIGGUR-ABN062-2016 and 672-2014). Colciencias supported Liliana Catherine Patiño´s work (Fellowship: 617, 2013).

Key words: whole-exome sequencing; primary ovarian insufficiency; female infertility; molecular etiology

Introduction

Primary ovarian insufficiency (POI), is a frequently occurring complex pathology affecting 1% of women under 40 years old (Conway, 2000). Clinically, it is characterized by amenorrhea, hypoestrogenism, and high gonadotropin levels reflecting precocious ovarian depletion of the follicular reserve (Nelson, 2009; De Vos *et al.*, 2010). POI has been proposed as a progressive condition describing ovarian dysfunction (e.g. ovarian function impairment and irregular ovulation) leading to infertility (premature ovarian failure, POF) (Welt, 2008). Although most POI cases are considered idiopathic, genetic anomalies have been described in syndromic and non-syndromic forms of the disease, such as chromosomal abnormalities and point mutations in POI genes' coding regions (autosomes and X-linked genes) (Laissue, 2015; Qin *et al.*, 2015). Mutations in only a few candidate genes have been definitively related to pathogenesis of the disease, despite numerous attempts at identifying sequence variants via Sanger sequencing (Laissue, 2015; Qin *et al.*, 2015) (and references therein). This might have been due to the fact that female reproduction requires numerous steps, from sex determination/gametogenesis to ovulation, to guarantee oocyte health for normal fecundation.

It has been shown that several transcription factors (e.g. NR5A1, NOBOX, FIGLA, FOXL2) play key roles during female gonadal development and their mutations lead to POI (Laissue, 2015). TGF-β molecules and their downstream molecular pathways have also demonstrated to be essential for ovary physiology in distinct mammalian species. BPM15 and GDF9 are especially interesting as they participate as major regulators of mammalian ovulation rate. Furthermore, their mutations have been related to POI origin (Laissue, 2015). Meiotic genes as *MCM8*, *MCM9*, *STAG3*, *SYCE1*, *MSH3*, *MSH4* and *MLH3* have been considered as important molecules for determining the oocyte pool. To date, more than 60 mouse models presenting a well-defined phenotype of ovarian failure have been described (Barnett, 2006; Roy and Matzuk, 2006; Edson *et al.*, 2009; Jagarlamudi *et al.*, 2010; Sullivan and Castrillon, 2011; Monget *et al.*, 2012 and www.jax.org). Such a scenario, in which hundreds of genes are involved in complex dynamic regulatory networks, has hampered selecting relevant candidates to be screened by Sanger sequencing. This constraint, as well as the rarity of families affected by the disease (theoretically facilitating classical genetic mapping), has made research concerning POI genetic causes particularly challenging. Very recently, some studies based on next generation sequencing (NGS) have been successfully undertaken as they have led to new genes being proposed, as well as mutations associated with POI etiology (Caburet *et al.*, 2014; de Vries *et al.*, 2014; Wood-Trageser *et al.*, 2014; Fonseca *et al.*, 2015; Bouilly *et al.*, 2016; Bramble *et al.*, 2016; Fauchereau *et al.*, 2016). However, experiments have not been performed on large genomic regions in unrelated POI individuals.

The present study involved whole-exome sequencing of 69 unrelated Caucasian women affected by POI. Innovative bioinformatics analysis was used on non-synonymous sequence variants in a subset of 420 selected POI candidate genes. Fifty-five coding variants in 49

genes potentially related to the phenotype were identified in 33 out of 69 patients (48%). These genes participate in key biological processes in the ovary, such as meiosis, follicular development, granulosa cell differentiation/proliferation and ovulation. The presence of at least two mutations in distinct genes in 45% of the patients argued in favour of a polygenic nature of POI. Computational 3D modelling, via fragment molecular orbital method, of three mutations (two in BMPR1B and one in GREM1) argued strongly in favour of pathogenic effects. The novel genes and mutations described here represent potential future genetic biomarkers for POI.

Materials and Methods

Women affected by POI

Sixty-nine women (Pt-1 through Pt-69) affected by idiopathic POI were included in the study. These patients were Caucasians living in France who were referred for evaluation to the Reproductive Endocrinology Department at Bicêtre Teaching Hospital and the Endocrinology Department at Robert Debré Hospital, both in France. All patients exhibited at least 6 months of amenorrhea before age 40 with FSH values >20 IU/L measured in two samples at least 1 month apart and had a normal 46,XX karyotype. Turner syndrome, X-chromosome karyotypic abnormalities and *FMR1* premutations were excluded and none of the patients had circulating ovarian antibodies. Women having antecedents of pelvic surgery, ovarian infections, chemotherapy and/or autoimmune disease were also excluded from the study. Twelve and 57 displayed primary or secondary amenorrhea, respectively.

NGS, Sanger sequencing and bioinformatics analysis

Total DNA from patients was extracted from blood leucocytes by conventional salting-out procedure. Experimental details of NGS experiments, Sanger sequencing and bioinformatics analysis have been included as **Supplemental Methods**.

Structure preparation, modelling and fragment molecular orbital (FMO) calculations

Details on the *in silico* approaches for modelling BMPR1B-p.Arg254His, BMPR1B-p.Phe272Leu and p.GREM1-p.Arg169Thr mutations have been included as **Supplemental Methods.**

Ethical approval

All clinical and experimental steps of this study were approved by Institutional Review Board (reference PHRC No. A0R03 052) and by Bicêtre Ethical committee (CPP # PP 16-024 Ile-de-France VII). The clinical investigation was performed according to Helsinki Declaration guidelines (1975, as revised in 1996). All the women had given their informed consent to participate.

Results

The percentage of reads on target (coverage) ranged from 80%–95%. Coverage was defined as the percentage of target bases that are sequenced a given number of times. More than 85% of the target was covered at 40X depth. Exome data was uploaded at the Open Science

Framework (Patiño, L. 2016, December 16, http://doi.org/10.17605/OSF.IO/EY9ME). 43337 sequence variants were identified in the POI-420 subset (**Figure 1**). 2544 variants having MAF <0.05 were present in the POI-420 group while 137996 were found throughout the exome (all exome data, *All-ex*). Among POI-420, 488 induced a protein change: 7 nonsense, 4 splice site, 53 frameshift and 424 missense variants. Among these 460 missense variants, 120 had scores compatible with deleterious effects by using PolyPhen-2 and SIFT bioinformatics tools. 55 sequence variants were definitely confirmed by Sanger sequencing (**Table 1, Figure 1**). All variants were found at heterozygous state. In this series of 69 POI patients, 33 presented one or more confirmed variant (**Table 1**). The frequency of each variation in the ExAC database was indicated. Four genes displayed at least two mutations: *NOTCH2* (n=3), *ADAMTS16* (n=2), *BMPR1A* (n=2), *BMPR1B* (n=2) and *C3ORF77* (n=2).

Clinical characteristics of patients having candidate mutations are shown in **Table 1**. Four patients presented with primary amenorrhea with varying pubertal development. The other patients presented with normal puberty and secondary amenorrhea. Symptoms appeared between 15-39 years of age (median 32±8 yrs). Hormonal characteristics included markedly elevated FSH (73,6 ± 6.2 IU/L), LH (36,5 ± 3,8 IU/L) and low levels of estradiol (14,7 ± 2,5 ng/L). In sum, among the 33 patients, 19, 9, 2 and 3 patients were found to carry 1, 2, 3 and 4 mutations, respectively. Interestingly 43% of these patients had at least two mutations in different genes arguing in favor of a polygenic origin for POI.

The BMPR1B modelled mutations by FMO analysis involved changes in stabilising interactions (**Supplemental Figure S2**). The mutations highlighted a major change in total interaction energy from -54.75 (WT) to -29.54 (MT) kcal/mol (in position Arg254) and -

44.69 (WT) to -33.38 (MT) kcal/mol (in position Phe272). Replacing a charged amino acid by a neutral amino acid and the loss of a non-classical H-bond (CH-π interactions) contributed to BMPR1B-MT protein destabilisation. Similarly, changes of one order of magnitude were found (-239.86 kcal/mol WT vs. -27.86 kcal/mol MT) concerning stabilising interactions between GREM1-WT (wild type) and GREM1-MT (mutant) (**Supplemental Figure S3**). Detailed information on results from FMO analysis has been included as **Supplemental Results**.

Discussion

The present work describes whole-exome sequencing in 69 patients who were affected by classical clinical signs of POI. Primary analysis of data was focused on 420 POI candidate genes which had been systematically selected from public databases. Stringent filters (e.g. low MAF, non-synonymous mutations, SIFT and PolyPhen2 software screening) were used to facilitate the selection of rare mutations having (theoretically) moderate/strong pathogenic functional effects. These mutations affected genes involved in several key biological processes, such as meiosis, follicular development, granulosa cell differentiation/proliferation, ovulation, cell metabolism and extracellular matrix regulation (**Table 1**). Although all the 55 filtered variants (and genes) may have contributed to the POI phenotype (some of them probably in an additive/epistatic fashion), several of them belonging to distinct molecular cascades are especially interesting because of their previously described roles in ovary physiology.

GDF9, BMPR1B, GREM1, which participate in the TGF-β (transforming growth factor) signalling pathway, have been clearly linked to specific ovary biological functions, such as granulosa cell proliferation, ovulation and/or follicular development regulation (**Figure 2**). *GDF9* (as well as its close homologue *BMP15*) is a soluble oocyte-secreted factor which binds to specific serine/threonine kinase types I and II receptors located on granulosa cell surface (Weiss and Attisano, 2013; Laissue, 2015). Several mutations in humans, most located in the protein's pro-region, have been identified in POI patients and women displaying twinning (Montgomery *et al.*, 2004; Palmer *et al.*, 2006; Laissue *et al.*, 2008; Persani *et al.*, 2014). Functional tests of mutant GDF9 have been seen to have deleterious effects, such as the synthesis of defective mature products, the reduction of mature protein expression/secretion and the inhibition of granulosa cell proliferation (Inagaki and Shimasaki, 2010; Wang *et al.*, 2013; Persani *et al.*, 2014; Simpson *et al.*, 2014). Some mutations, especially those located at the end (C-ter) of the pro-domain, have been related to an increase in granulosa cell proliferation (Simpson *et al.*, 2014). The GDF9 p.Ser83Cys mutation identified in Pt-34 was located in the protein's pro-region which is important for proper protein folding, dimerization, secretion and stability. Similar to other GDF9 mutations located in the pro-region, GDF9-p.Ser83Cys might lead to mature peptide dysfunction and granulosa cell proliferation inhibition.

BMP15:GDF9 heterodimers (which have greater biological activity than either BMP15 or GDF9 homodimers alone) act in human and mouse species via a receptor complex constituted by the BMPR2 receptor, the ALK4/5/7 type I receptor and the BMPR1B (ALK6) co-receptor (Peng *et al.*, 2013). ALK6 has been shown to be essential for downstream intracellular signalling by triggering SMAD1/5/8 phosphorylation. *Alk6* knockout females have been

shown to suffer infertility secondary to cumulus expansion impairment while the p.Gln249Arg mutation in sheep (located in the protein's highly conserved intracellular kinase signalling domain) has been linked to hyperfertility, due to an increase in ovulation rate (Souza *et al.*, 2001; Yi *et al.*, 2001; Davis, 2004). Overexpression of BMPR1B has been described in women having a reduced ovarian reserve (Regan *et al.*, 2016). Both mutations identified in BMPR1B (p.Arg254His and p.Phe272Leu) in the present study were located in the functional intracellular kinase domain, suggesting that they might be associated with POI pathogenesis. In addition, results from FMO analysis suggested a significant change in protein stability secondary to these mutations, which might related to and impairment of the TGF-β signalling between oocytes and granulosa cells (**Supplemental Figure S2**).

Regarding TGF-β signalling regulation, *GREM1* (Gremlin1), a member of the DAN family of BMP inhibitors, binds to BMP proteins, preventing them from activating specific receptors (Kattamuri *et al.*, 2012). Although the mechanism used by DAN proteins during BMP ligand inhibition is not well understood, it has been shown that GREM1 regulates important factors having roles during folliculogenesis, such as BMP2, BMP4 and BMP15 (Hsu *et al.*, 1998; Pangas *et al.*, 2004; Nilsson *et al.*, 2014; Church *et al.*, 2015; Bayne *et al.*, 2016) (**Figure 2**). *Grem1* knockout mice have displayed delayed meiotic progression, defects regarding primordial follicle assembly dysfunction and a reduced amount of oocytes (Myers *et al.*, 2011). GREM1 is expressed in humans during early and until late stages of follicular development, and has been linked to granulosa cell development (Kristensen *et al.*, 2014; Bayne *et al.*, 2016). Furthermore, a significant decrease in its expression has been reported in women having reduced ovarian reserve (Jindal *et al.*, 2012).

The GREM1-p.Arg169Thr mutation found in Pt-24 strongly suggests a functional role since it is located in a critical region (Pro^{145} to Gln^{174} residues) of the DAN domain which directly interacts with BMP4 (Sun *et al.*, 2006). Furthermore, the GREM1-Arg^{169} residue is conserved in other DAN-family members and among numerous vertebrate species (Sun *et al.*, 2006; Veverka *et al.*, 2009). Indeed, abnormal folding of the β2/β3 (finger 2) sheet could modify the protein's local chemical properties which might then lead to interaction disturbances with BMP4 (or other BMP factors). As for BMPR1B mutations, the FMO analysis showed that the GREM1-p.Arg169Thr mutation led to changes in protein stability which might contribute to the phenotype (**Supplemental Figure S3**). These findings strongly suggest a relevant role for TGF-β proteins, especially those involved in oocyte-to-granulosa cell signalling, during POI pathogenesis.

Concerning molecules involved in meiosis, the present study was able to identify 16 mutations potentially contributing to the phenotype. Functional protein association networks of some meiotic proteins have been included as supplemental material (**Supplemental Figure S1**). STAG3 and MCM9 are especially interesting due to their well-established role during female fertility and POI. To date, all mutations in meiotic genes linked to POI etiology have been found in biallelic state (homozygous or compound heterozygous) thereby underlining meiosis' key role in reproduction and species maintenance (Caburet *et al.*, 2014; de Vries *et al.*, 2014; Wang *et al.*, 2014; Wood-Trageser *et al.*, 2014; AlAsiri *et al.*, 2015; Fauchereau *et al.*, 2016). Mutations in meiotic genes were present at heterozygous state in our present study, which might be associated with a background of POI predisposition. Further variants would be necessary to originate the phenotype in such hypothetically scenario. Interestingly, we found that 64% (7 out 11) of patients having a heterozygous

mutation in a meiotic gene were carriers of at least one further variant in the same or a distinct gene.

Interestingly, we have found three different mutations in *NOTCH2,* a gene encoding one of the four NOTCH family single-pass Type I (SPTI) transmembrane receptors (Andersson *et al.*, 2011). The NOTCH2-p.Ser1804Leu, p.Gln1811His and p.Leu2408His mutations identified in the present study were located in the intracellular domain of the protein which translocates to the nucleus where it mediates transactivation/repression (Kopan and Ilagan, 2009). Thus, it would be possible that these mutant forms lead to expression disturbances of key target genes involved during oocyte development.

We consider that additional mutations in genes participating in follicular development, granulosa cell differentiation and proliferation, ovulation and extracellular matrix regulation could also contribute to the phenotype due to their molecular behaviour during ovary development and physiology. For example, this is the case of ATG7-p.Phe403Leu, THBS1-p.Gln96Arg, PTCH1-p.Val1131Ala, PCSK6-p.Thr964Met, UMODL1-p.Ile1330Asn, ADAMTS16-p.Arg100Trp, p.Arg789Cys and PTX3-p.Pro303Arg.

To note, in clinical practice it has been observed that patients affected by POI report similar phenotypes in some women from their families which suggests a genetic origin of the disease.

In our case, although candidate mutations have not shown to be clustered in particular familial cases, incomplete penetrance cannot be excluded. Thus, it would be interesting to study potential segregation analysis of interesting variants but, unfortunately, although we

did propose to most of our POI patients the idea of contacting their parents regarding their participation in our study they decided not to involve their families.

The genetic approach presented here revealed that 33 out of 69 (48%) patients were carriers of mutations potentially related to the phenotype. Interestingly, 42% of these patients had at least two mutations in different genes and 49 out 55 variants were identified in distinct genes, thereby arguing in favor of a polygenic origin for POI. Furthermore, our findings evoke the importance of rare variants in complex disease pathogenesis and contribute information for resolving genomic concerns such as "missing hereditability" (Manolio *et al.*, 2009; Gibson, 2012; Lee *et al.*, 2014; Laissue, 2015).

Concerning our methodological approach it is clear that correct gene subset configuration depends on multiple variables, such as the availability of previous accurate data relating specific genes to ovarian biology and the rigor (and method) used when investigating potential candidates. This approach may lose further candidates contributing to the phenotype. However, we consider that it represents interesting middle ground between a large amount of genomic data (e.g. All-ex variants) and the results obtained from other sequencing designs (custom array sequencing or single Sanger approaches). An advantage of the present design is that the availability of sequences from all encoding regions enables future reanalysing of data by including additional genes and/or by setting up alternative methods (e.g. interactome approaches).

We estimate that whole-exome sequencing and the *in silico* analysis presented here represent an efficient approach for mapping variants (having potentially moderate/strong functional

effects) associated with POI etiology. Further NGS studies, performed in larger panels of women affected by POI, would be a valuable exercise to identify novel causative mutations. Taken together, our findings add valuable information regarding POI molecular etiology and ought to form the starting point for further functional *in vitro* and *in vivo* studies.

Authors' Roles

Clinical work was performed by BD, JY, NB and IB. The experiments were performed by LCP, CC, JCB. MAP, CFS and RG performed the FMO analysis. All authors contributed to interpretation of findings. The study was designed and directed by PL. The manuscript was draft by PL with contributions to revision and final version by all authors.

Funding

This study was supported by the Universidad del Rosario, Grant CS/CIGGUR-ABN062-2016.

Conflict of Interest

The authors declare no conflict of interest.

References

AlAsiri S, Basit S, Wood-Trageser MA, Yatsenko SA, Jeffries EP, Surti U, Ketterer DM, Afzal S, Ramzan K, Faiyaz-Ul Haque M, *et al.* Exome sequencing reveals MCM8 mutation underlies ovarian failure and chromosomal instability. *J Clin Invest*

2015;**125**:258–262.

Andersson ER, Sandberg R, Lendahl U. Notch signaling: simplicity in design, versatility in function. *Development* 2011;**138**:3593–3612.

Barnett KR. Ovarian follicle development and transgenic mouse models. *Hum Reprod Update* 2006;**12**:537–555.

Bayne RA, Donnachie DJ, Kinnell HL, Childs AJ, Anderson RA. BMP signalling in human fetal ovary somatic cells is modulated in a gene-specific fashion by GREM1 and GREM2. *Mol Hum Reprod* 2016;**22**:622–633.

Bouilly J, Beau I, Barraud S, Bernard V, Azibi K, Fagart J, Fèvre A, Todeschini AL, Veitia RA, Beldjord C, *et al.* Identification of multiple gene mutations accounts for a new genetic architecture of primary ovarian insufficiency. *J Clin Endocrinol Metab* 2016;jc.2016-2152.

Bramble MS, Goldstein EH, Lipson A, Ngun T, Eskin A, Gosschalk JE, Roach L, Vashist N, Barseghyan H, Lee E, *et al.* A novel follicle-stimulating hormone receptor mutation causing primary ovarian failure: a fertility application of whole exome sequencing. *Hum Reprod* 2016;**31**:905–914.

Caburet S, Arboleda VA, Llano E, Overbeek PA, Barbero JL, Oka K, Harrison W, Vaiman D, Ben-Neriah Z, García-Tuñón I, *et al.* Mutant cohesin in premature ovarian failure. *N Engl J Med* 2014;**370**:943–949.

Church RH, Krishnakumar A, Urbanek A, Geschwindner S, Meneely J, Bianchi A, Basta B, Monaghan S, Elliot C, Strömstedt M, *et al.* Gremlin1 preferentially binds to bone

morphogenetic protein-2 (BMP-2) and BMP-4 over BMP-7. *Biochem J* 2015;**466**:55–68.

Conway GS. Premature ovarian failure. *Br Med Bull* 2000;**56**:643–649.

Davis GH. Fecundity genes in sheep. *Anim Reprod Sci* 2004;**82**–**83**:247–253.

Edson MA, Nagaraja AK, Matzuk MM. The mammalian ovary from genesis to revelation. *Endocr Rev* 2009;**30**:624–712.

Fauchereau F, Shalev S, Chervinsky E, Beck-Fruchter R, Legois B, Fellous M, Caburet S, Veitia RA. A non-sense MCM9 mutation in a familial case of primary ovarian insufficiency. *Clin Genet* 2016;**89**:603–607.

Fonseca DJ, Patiño LC, Suárez YC, Jesús Rodríguez A de, Mateus HE, Jiménez KM, Ortega-Recalde O, Díaz-Yamal I, Laissue P. Next generation sequencing in women affected by nonsyndromic premature ovarian failure displays new potential causative genes and mutations. *Fertil Steril* 2015;**104**:154–162.e2.

Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet* 2012;**13**:135–145.

Hsu DR, Economides AN, Wang X, Eimon PM, Harland RM. The Xenopus dorsalizing factor Gremlin identifies a novel family of secreted proteins that antagonize BMP activities. *Mol Cell* 1998;**1**:673–683.

Inagaki K, Shimasaki S. Impaired production of BMP-15 and GDF-9 mature proteins derived from proproteins WITH mutations in the proregion. *Mol Cell Endocrinol* 2010;**328**:1–7.

Jagarlamudi K, Reddy P, Adhikari D, Liu K. Genetically modified mouse models for premature ovarian failure (POF). *Mol Cell Endocrinol* 2010;**315**:1–10.

Jindal S, Greenseid K, Berger D, Santoro N, Pal L. Impaired Gremlin 1 (GREM1) expression in cumulus cells in young women with diminished ovarian reserve (DOR). *J Assist Reprod Genet* 2012;**29**:159–162.

Kattamuri C, Luedeke DM, Nolan K, Rankin SA, Greis KD, Zorn AM, Thompson TB. Members of the DAN Family Are BMP Antagonists That Form Highly Stable Noncovalent Dimers. *J Mol Biol* 2012;**424**:313–327.

Kopan R, Ilagan MXG. The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell* 2009;**137**:216–233.

Kristensen SG, Andersen K, Clement CA, Franks S, Hardy K, Andersen CY. Expression of TGF-beta superfamily growth factors, their receptors, the associated SMADs and antagonists in five isolated size-matched populations of pre-antral follicles from normal human ovaries. *Mol Hum Reprod* 2014;**20**:293–308.

Laissue P. Aetiological coding sequence variants in non-syndromic premature ovarian failure: From genetic linkage analysis to next generation sequencing. *Mol Cell Endocrinol* 2015;**411**:243–257.

Laissue P, Vinci G, Veitia RA, Fellous M. Recent advances in the study of genes involved in non-syndromic premature ovarian failure. *Mol Cell Endocrinol* 2008;**282**:101–111.

Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014;**95**:5–23.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747–753.

Monget P, Bobe J, Gougeon A, Fabre S, Monniaux D, Dalbies-Tran R. The ovarian reserve in mammals: A functional and evolutionary perspective. *Mol Cell Endocrinol* 2012;**356**:2–12.

Montgomery GW, Zhao ZZ, Marsh AJ, Mayne R, Treloar SA, James M, Martin NG, Boomsma DI, Duffy DL. A deletion mutation in GDF9 in sisters with spontaneous DZ twins. *Twin Res* 2004;**7**:548–555.

Myers M, Tripurani SK, Middlebrook B, Economides AN, Canalis E, Pangas SA. Loss of Gremlin Delays Primordial Follicle Assembly but Does Not Affect Female Fertility in Mice. *Biol Reprod* 2011;**85**:1175–1182.

Nelson LM. Primary Ovarian Insufficiency. *N Engl J Med* 2009;**360**:606–614.

Nilsson EE, Larsen G, Skinner MK. Roles of Gremlin 1 and Gremlin 2 in regulating ovarian primordial to primary follicle transition. *Reproduction* 2014;**147**:865–874.

Palmer JS, Zhao ZZ, Hoekstra C, Hayward NK, Webb PM, Whiteman DC, Martin NG, Boomsma DI, Duffy DL, Montgomery GW. Novel variants in growth differentiation factor 9 in mothers of dizygotic twins. *J Clin Endocrinol Metab* 2006;**91**:4713–4716.

Pangas SA, Jorgez CJ, Matzuk MM. Growth differentiation factor 9 regulates expression of the bone morphogenetic protein antagonist gremlin. *J Biol Chem* 2004;**279**:32281–32286.

Peng J, Li Q, Wigglesworth K, Rangarajan A, Kattamuri C, Peterson RT, Eppig JJ, Thompson TB, Matzuk MM. Growth differentiation factor 9:bone morphogenetic protein 15 heterodimers are potent regulators of ovarian functions. *Proc Natl Acad Sci* 2013;**110**:E776–E785.

Persani L, Rossetti R, Pasquale E Di, Cacciatore C, Fabre S. The fundamental role of bone morphogenetic protein 15 in ovarian function and its involvement in female fertility disorders. *Hum Reprod Update* 2014;**20**:869–883.

Qin Y, Jiao X, Simpson JL, Chen Z-J. Genetics of primary ovarian insufficiency: new developments and opportunities. *Hum Reprod Update* 2015;**21**:787–808.

Regan SLP, Knight PG, Yovich JL, Stanger JD, Leung Y, Arfuso F, Dharmarajan A, Almahbobi G. Dysregulation of granulosal bone morphogenetic protein receptor 1B density is associated with reduced ovarian reserve and the age-related decline in human fertility. *Mol Cell Endocrinol* 2016;**425**:84–93.

Roy A, Matzuk MM. Deconstructing mammalian reproduction: using knockouts to define fertility pathways. *Reproduction* 2006;**131**:207–219.

Simpson CM, Robertson DM, Al-Musawi SL, Heath DA, McNatty KP, Ritter LJ, Mottershead DG, Gilchrist RB, Harrison CA, Stanton PG. Aberrant GDF9 expression and activation are associated with common human ovarian disorders. *J Clin Endocrinol Metab* 2014;**99**:E615-24.

Souza CJ, MacDougall C, MacDougall C, Campbell BK, McNeilly AS, Baird DT. The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic

receptor type 1 B (BMPR1B) gene. *J Endocrinol* 2001;**169**:R1-6.

Sullivan S, Castrillon D. Insights into Primary Ovarian Insufficiency through Genetically Engineered Mouse Models. *Semin Reprod Med* 2011;**29**:283–298.

Sun J, Zhuang F-F, Mullersman JE, Chen H, Robertson EJ, Warburton D, Liu Y-H, Shi W. BMP4 activation and secretion are negatively regulated by an intracellular gremlin-BMP4 interaction. *J Biol Chem* 2006;**281**:29349–29356.

Veverka V, Henry AJ, Slocombe PM, Ventom A, Mulloy B, Muskett FW, Muzylak M, Greenslade K, Moore A, Zhang L, *et al.* Characterization of the structural features and interactions of sclerostin: molecular insight into a key regulator of Wnt-mediated bone formation. *J Biol Chem* 2009;**284**:10890–10900.

Vos M De, Devroey P, Fauser BCJM. Primary ovarian insufficiency. *Lancet (London, England)* 2010;**376**:911–921.

Vries L de, Behar DM, Smirin-Yosef P, Lagovsky I, Tzur S, Basel-Vanagaite L. Exome Sequencing Reveals SYCE1 Mutation Associated With Autosomal Recessive Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* 2014;**99**:E2129–E2132.

Wang J, Zhang W, Jiang H, Wu B-L, Primary Ovarian Insufficiency Collaboration. Mutations in HFM1 in recessive primary ovarian insufficiency. *N Engl J Med* 2014;**370**:972–974.

Wang T-T, Ke Z-H, Song Y, Chen L-T, Chen X-J, Feng C, Zhang D, Zhang R-J, Wu Y-T, Zhang Y, *et al.* Identification of a mutation in GDF9 as a novel cause of diminished ovarian reserve in young women. *Hum Reprod* 2013;**28**:2473–2481.

Weiss A, Attisano L. The TGFbeta Superfamily Signaling Pathway. *Wiley Interdiscip Rev Dev Biol* 2013;**2**:47–63.

Welt CK. Primary ovarian insufficiency: a more accurate term for premature ovarian failure. *Clin Endocrinol (Oxf)* 2008;**68**:499–509.

Wood-Trageser MA, Gurbuz F, Yatsenko SA, Jeffries EP, Kotan LD, Surti U, Ketterer DM, Matic J, Chipkin J, Jiang H, *et al.* MCM9 mutations are associated with ovarian failure, short stature, and chromosomal instability. *Am J Hum Genet* 2014;**95**:754–762.

Yi SE, LaPolt PS, Yoon BS, Chen JY, Lu JK, Lyons KM. The type I BMP receptor BmprIB is essential for female reproductive function. *Proc Natl Acad Sci U S A* 2001;**98**:7994–7999.

Figure Legends

Figure 1

POI gene subset included 420 candidate genes. Among 2 244 677 total variants, 55 were selected and confirmed by Sanger sequencing. **All ex**: variants found throughout the exome. **MAF**: minor allele frequency. **Missense S&P+:** missense mutations displaying potential deleterious effects by both SIFT and PolyPhen2 bioinformatic tools.

Figure 2

Signaling pathways and proteins involved in follicular development. **a)** Autophagy; **b)** P13K/AKT pathway; **c)** SOHLH1 pathway; **d)** TGF-β´s pathway; **e)** KIT-L and c-Kit **f)** leptin pathway; **g)** NOTCH pathway; **h)** connexins.

Supplemental Figure S1

Protein-protein interaction network made by STRING software for different meiosis proteins. Main proteins are shown into red circles: **A)** STAG3; **B)** MLH3; **C)** MEI1; **D)** PRDM1. Colored lines display known and predicted interactions. **Light blue**: curated databases; **pink:** experimentally determined; **green:** gene neighborhood; **red:** gene fusions; **blue**: gene co-occurrence; yellow: textmining; **black**: co-expression; **purple**: protein homology.

Supplemental Figure S2

FMO results for BMPR1B: **A.** PIEDA contributions of amino acids interacting with positions 254 -- Arg (WT) and His (MT) and 272 -- Phe (WT) and Leu (MT). Energies are expressed in kcal/mol **B.** Overall view of the analysed system. BMPR1B (chain A) and FKBP12 (chain B) are shown in blue and red, respectively. The mutation zones for positions 254 and 272 are shown in green and purple boxes, respectively. **C.** Arg254 WT**, D.** His254 MT**, E.** Phe272 WT**, F.** Leu272 MT**:** Bar plots describe the PIEDA of energy interaction terms: electrostatics (green), exchange-repulsion (red), charge-transfer (blue), dispersion (yellow), and solvation (cyan). Positive values are considered destabilising and negative stabilising. **G.** Detail of the amino acids interacting with Arg254 in BMPR1B WT. **H.** Detail of the amino acids interacting with His254 in BMPR1B MT. **I.** Detail of the amino acids interacting with Phe272 in BMPR1B WT. **J.** Detail of the amino acids interacting with Leu272 in BMPR1B MT. Hydrogen bonds are shown as dotted lines. The backbone-backbone hydrogen bond with Glu256A could not be calculated due the limitations of fragmentation model.

Supplemental Figure 3

FMO results for GREM1: **A.** PIEDA contributions of amino acids interacting with positions 169 chain A -- Arg (WT) and Thr (MT) and the same position, but in chain B. Energies are expressed in kcal/mol **B.** Overall view of the analysed system. Chain A (blue), chain B (red), chain C (gray) and chain D (orange). The mutation zone for position 169 in the chain A (site 1) and the mutation zone for position 169 in the chain B (site 2) are shown in green and purple boxes, respectively. **C.** Site 1 WT**, D.** Site 1 MT**, E.** Site 2 WT**:** Bar plots describe the PIEDA of energy interaction terms: electrostatics (green), exchange-repulsion (red), charge-transfer (blue), dispersion (yellow), and solvation (cyan). Positive values are

considered destabilising and negative stabilising. **F.** Detail of the amino acids interacting with Arg169A (site 1) in GREM1 WT (side chain view). **G.** Detail of the amino acids interacting with Arg169A (site 1) in GREM1 WT (backbone view) **H.** Detail of the amino acids interacting with Thr169A in in GREM1 MT. **I.** Detail of the amino acids interacting with Arg169B in in GREM1 WT. Hydrogen bonds are shown as dotted lines.

Supplemental Results

The FMO method was used for studying BMPR1B and GREM1 WT and MT structures regarding the effect of amino acid substitutions (BMPR1B-p.Arg254His, BMPR1B p.Phe272Leu and p.GREM1-p.Arg169Thr). Supplementary Figures 2 and 3 show the calculated values for BMPR1B and GREM1 models, respectively.

Regarding BMPR1B, FMO analysis showed that Arg254A when replaced by His254A evoked a deleterious effect on stabilising interactions. Two major interactions were found concerning the WT protein (**Supplementary Figure 2 A, C and G**). A hydrogen bond (H-bond) was formed between Glu204A backbone and Arg254A side chain. Another significant interaction between charged Arg254A and charged Glu55B (corresponding to FKBP12 protein) was detected. These two interactions were dominated by the electrostatic term. Regarding His254A-MT, four interactions were identified by FMO (**Supplementary Figure 2 A, D and H**). The side chain of His254A formed a H-bond with Gln233A side chain. An important electrostatic interaction was detected between His254A and charged Glu55B (corresponding to FKBP12 protein). Two interactions dominated by the solvation component of PIEDA were found between His254A and Glu204A and Glu256A.

Concerning the Phe272A in WT, four interactions were identified by FMO (**Supplementary Figure 2 A, E and I**). Two H-bonds were formed between Phe272A backbone and the backbone of Glu276A and Glu268A. A non-classical H-bond CH-π interaction was detected between Phe272A side chain and Pro89B side chain of FKBP12. This interaction was dominated by the dispersion term. An additional H-bond was found between Phe272A side

chain and Glu268A side chain. FMO only identified two interactions for Leu272A-MT (**Supplementary Figure 2 A, F and G**). It is worth noting that the CH-π interaction is missing due to substituting Phe272A by Leu272A. As in the previous case, two H-bonds were formed between Phe272A backbone and the backbone of Glu276A and Glu268A.

Analysis of the GREM1 model led to identifying eight major interactions by means of the FMO method for residue Arg169A in WT (Site 1) (**Supplementary Figure 3 A, C, F and G**). A salt bridge was formed between deprotonated Glu135B and protonated Arg169A; this interaction consisted of a combination of two non-covalent interactions: hydrogen bonding and electrostatic interactions. Four additional H-bonds were detected by FMO. Three H-bonds between Arg169A side chain and the backbone of Asp184C, Asp182C and Leu183C, respectively. An additional H-bond was formed between Arg169A backbone and Met153A side chain. Two other interactions dominated by the electrostatic term were found between the guanidinium group of Arg169A with Glu134B and Gln139C side chains. The weakest interaction was driven by the solvation component of PIEDA between Arg169A side chain and Thr151A side chain. FMO only identified two interactions for Thr169A MT (**Supplementary Figure 3 A, D, and H**). It is worth noting that the salt bridge between deprotonated Glu135B and protonated Arg169A was missing due to the charged Arg169A being replaced by the non-charged Thr169A. Two H-bonds were formed between the Thr169A side chain and the backbone of Asp182C and Asp184C. These two interactions were dominated by the electrostatic term. The remaining interactions stabilising the interaction in the WT became lost in the Thr169A MT structure.

Concerning the Arg169B in WT (**Supplementary Figure 3** A, E and I), three interactions were identified by FMO. The guanidinium group of Arg169B formed a H-bond with

Met153B side chain, this interaction was dominated by the dispersion term. Another important interaction driven by the electrostatic term was detected between Arg169B and Glu105B. As in the previous case, the weakest interaction was driven by the solvation component of PIEDA between the Arg169B side chain and the Thr151B side chain. The FMO method did not detect significant interactions for Thr169B MT.

Supplemental methods

NGS, Sanger sequencing and bioinformatics analysis

Library preparation and Ion Proton sequencing were performed following certified protocols from Life Technology. Briefly, 100 ng of genomic DNA was used to amplify exonic target regions, and were enriched and amplified for the 69 DNA samples using Ion AmpliSeq™ Exome RDY Library Preparation kit (Thermo Scientific, A27192). Each sample was processed separately. The amplicons were partially digested with FuPa reagent (proprietary to Thermo Scientific) and phosphorylated prior to ligation of Ion Xpress™ Barcode Adapters followed by cleanup using HighPrep PCR clean up system (Magbio, AC 60050). The final libraries were quantified on Qubit® Flurometer using Qubit® dsDNA HS Assay Kit (Thermo Scientific, Q32854) and Agilent® Bioanalyzer using Agilent High Sensitivity DNA Kit (Agilent, 5067-4626). 2 samples were pooled according to the concentrations on the Bioanalyzer and loaded on Ion PI™ Chip to be sequenced on Ion Proton™ system. The samples were sequenced with Ion Proton Sequencer and analyzed with Torrent suite v 4.4.3. The raw reads undergo the process of trimming and filtering to get only the high quality reads. Only those which pass these filters will be considered for the downstream analysis. The raw reads obtained are aligned to the reference HG19 with the TMAP algorithm. The

variants detected with the variant caller plugin were further annotated using the Ion Reporter 4.2 to give location (intronic/exonic/utr), gene name, protein change, function and dbSNP Id (from the dbSNP database 137) and Variant effect predictor for SIFT and Polyphen prediction. Library preparation and sequencing were carried out at Genotypic Technology's Genomics facility (Bangalore, Karnataka, India).

The POI gene subset (POI-420) consisted of 420 genes (**Supplementary Table 1**) which were considered candidates as they had been reported as having expression/function during distinct reproductive processes (e.g. sex determination, meiosis, folliculogenesis and ovulation). Several websites were used for creating this list of genes, such as Highwire, PubMed, MGI-Jackson Laboratory, Geoprofiles, Genecards and Illumina NextBio. These databases were exhaustively mined for pertinent information by using numerous combinations of keywords: premature ovarian failure, primary ovarian insufficiency, POI/POF genetics, hypergonadotropic hypogonadism, gametogenesis, molecular regulation of meiosis, folliculogenesis, ovulation genetics, sex determination, granulosa cell physiology and hypothalamic/pituitary/gonadal axis.

R software programming and Excel (Microsoft) functions were used for exome data filtering. Sequence variants (synonymous and non-synonymous) in the POI-420 subset reported as having minor allele frequencies (MAF) <0.05 were selected for subsequent analysis. Variants having a potential effect at sequence protein level (e.g. missense, nonsense, splice site, frameshift) were then filtered for downstream analysis. Concerning missense mutations, all those displaying potential deleterious effects by both PolyPhen2 and SIFT bioinformatics tools (n=119) were filtered for subsequent analysis. The PolyPhen2 prediction software

includes an algorithm that uses distinct variables such as interspecific protein alignments, mapping residues to 3-dimensional protein structures and physicochemical characteristics of the interchanged amino acids. The SIFT algorithm is based on calculations of evolutionary conservation of amino acids. All filtered candidate sequence variants were checked by PCR/Sanger sequencing. Technical conditions for PCR/sequencing assays, including oligonucleotide sequences, are available upon request. Clustal W software was used for aligning human protein sequences with those from orthologous species. STRING software (string-db.org) was used for constructing functional protein association networks for STAG3 MLH3 MEI1 and PRDM1.

Structure preparation and modelling

BMPR1B (pdb: 3MDY) and GREM1 (pdb: 5AEJ) crystal structures (WT versions) and their respective mutants (MT) (BMPR1B-p.Arg254His, BMPR1B-p.Phe272Leu and p.GREM1-p.Arg169Thr) were analysed (Chaikuad *et al.*, 2012; Kišonaitė *et al.*, 2016). The UCSF Chimera *swapaa* function was used to make amino acid substitutions in crystal structures, using the Dunbrack backbone-dependent rotamer library (Dunbrack, 2002; Pettersen *et al.*, 2004). The Poisson-Boltzmann method was used for calculating residue protonation states, using the H++ web server and a pH of 7.4 for both proteins (Gordon *et al.*, 2005).

The structures were subjected to a restrained minimization procedure with the ff14SB classical force field implemented in the AMBER14 program. Each structure was solvated in an octahedral box of TIP3P water molecules containing chloride as the counter-ion. The minimum distance between the protein surface and the edge of the box was set at 10 Å for

the solvated box. Only the protein without water molecules was included in the fragment molecular orbital (FMO) calculations after the minimization procedure.

Fragment molecular orbital (FMO) calculations

The FMO method was used for studying the effect induced by amino acid substitutions (Fedorov *et al.*, 2012). This approach allows a comprehensive evaluation of variation types and energy changes caused by mutations. This *ab initio* quantum method enables an accurate evaluation of large molecular systems by means of a partition scheme (fragments). Total interaction energy can be decomposed into electrostatic, repulsion, charge transfer, dispersion and solvation terms by using a pair interaction decomposition analysis (PIEDA) for each fragment pair (Fedorov and Kitaura, 2007). The FMO method (version 5.2) implemented in the GAMESS 2016 software and the Hartree Fock (HF) theory with the 6-31G* basis set was used (Schmidt *et al.*, 1993). Solvent effects were included with the polarizable continuum model (PCM). Grimme's dispersion model D3 was used for correcting all HF energies (Grimme *et al.*, 2011). All the models were fragmented using Facio v. 19.2.1 (Suenaga, 2005). Interactions between fragments having a $\geq$ 3 kcal/mol absolute value were considered significant (Heifetz *et al.*, 2016). Only interactions within 6.5 Å from the studied amino acid were included for each structure.

References

Chaikuad A, Alfano I, Kerr G, Sanvitale CE, Boergermann JH, Triffitt JT, Delft F von, Knapp S, Knaus P, Bullock AN. Structure of the bone morphogenetic protein receptor ALK2 and implications for fibrodysplasia ossificans progressiva. *J Biol Chem* 2012;**287**:36990–36998.

Dunbrack RL. Rotamer libraries in the 21st century. *Curr Opin Struct Biol* 2002;**12**:431–440.

Fedorov DG, Kitaura K. Pair interaction energy decomposition analysis. *J Comput Chem* 2007;**28**:222–237.

Fedorov DG, Nagata T, Kitaura K. Exploring chemistry with the fragment molecular orbital method. *Phys Chem Chem Phys* 2012;**14**:7562.

Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* 2005;**33**:W368–W371.

Grimme S, Ehrlich S, Goerigk L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* 2011;**32**:1456–1465.

Heifetz A, Chudyk EI, Gleave L, Aldeghi M, Cherezov V, Fedorov DG, Biggin PC, Bodkin MJ. The Fragment Molecular Orbital Method Reveals New Insight into the Chemical Nature of GPCR–Ligand Interactions. *J Chem Inf Model* 2016;**56**:159–172.

Kišonaitė M, Wang X, Hyvönen M. Structure of Gremlin-1 and analysis of its interaction with BMP-2. *Biochem J* 2016;**473**:1593–1604.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera-A visualization system for exploratory research and analysis. *J Comput Chem* 2004;**25**:1605–1612.

Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, *et al.* General atomic and molecular electronic structure system. *J Comput Chem* 1993;**14**:1347–1363.

Suenaga M. Facio 3D-Graphic program for molecular modeling and visualization of quantum chemical calculations. *J Comput Chem Japan* 2005;**4**:25–32.

Table 1. Clinical and molecular findings of POI patients studied via whole-exome sequencing

Patient ID	Phenotype	Age at diagnosis	Hormone values			Gene	Locus (position)	Accession number	Mutation		ExAC Allele Frecuency	Biological process
			FSH (IU/L)	LH (IU/L)	E2 (ng/L)				Sequence variation	Protein Position		
Pt-2	Primary	29	50	16	17	*HK3*	5;176318162	NM_002115.2	c.290G>A	p.Gly97Glu	0.001034	Cell metabolism
						NOTCH2	1;120458122	NM_024408.3	c.7223T>A	p.Leu2408His	0.001788	Granulosa cell differentiation and proliferation
Pt-3	Secondary	17	91	34	9	*GATA4*	8;11615928	NM_002052.3	c.1273G>A	p.Asp425Asn	0.002117	Granulosa cell proliferation and differentiation
						INHBC	12;57843255	NM_005538.3	c.509T>A	p.Leu170Gln	0.002969	Meiosis
						MLH3	14;75515926	NM_001040108.1	c.433A>G	p.Thr145Ala	ND	Meiosis
						PCSK5	9;78796345	NM_001190482.1	c.2035T>C	p.Tyr679His	0.000008274	Ovulation
Pt-6	Secondary	21	64	28	55	*TSC1*	9;135781014	NM_000368.4	c.1951A>G	p.Arg651Gly	ND	Follicular development
Pt-7	Secondary	37	83	17	1	*ATG7*	3;11389434	NM_006395.2	c.1209T>A	p.Phe403Leu	0.000008238	Ovarian reserve
Pt-11	Secondary	35	44	23	6	*UMODL1*	21;43547856	NM_173568.3	c.3989T>A	p.Ile1330Asn	0.0002650	Granulosa cell differentiation and proliferation
Pt-14	Secondary	39	64	27	11	*HTRA3*	4;8295883	NM_053044.3	c.1006C>T	p.Arg336Cys	0.00004640	Granulosa cell differentiation and proliferation
						NBL1	1;19981530	NM_182744.3	c.112C>T	p.Leu38Phe	0.0005640	Follicular development
Pt-16	Secondary	39	141	58	7	*UBR2*	6;42571438	NM_015255.2	c.644C>T	p.Pro215Leu	0.0001484	Meiosis
Pt-17	Secondary	35	22	14	31	*PCSK1*	5;95730629	NM_000439.4	c.1823C>T	p.Thr608Met	0.00002471	Other
						BMP6	6;7862681	NM_001718.4	c.1154G>A	p.Arg385His	0.00004124	Follicular development
Pt-22	Secondary	20	101	28	2	*CXCR4*	2;136873083	NM_003467.2	c.415G>A	p.Val139Ile	ND	Ovulation
Pt-23	Secondary	32	37	5	8	*FGFR2*	10;123353268	NM_022970.3	c.64C>T	p.Arg22Trp	0.00009078	Follicular development

Pt-24	Secondary	37	58	7	12	*GREM1*	15;33023397	NM_013372.6	c.506G>C	p.Arg169Thr	ND	Follicular development
Pt-25	Primary	29	54	25	9	*MEI1* *GJA4* *IPO4* *ADAMTS16*	22;42095664 1;35260779 14;24649689 5;5239880	NM_152513.3 NM_002060.2 NM_024658.3 NM_139056.2	c.122C>A c.965G>A c.3205G>C c.2365C>T	p.Pro41His p.Arg322His p.Asp1069His p.Arg789Cys	0.00006178 0.00009366 0.000008760 0.001292	Meiosis Meiosis Meiosis Regulation of the extracellular matrix
Pt-34	Secondary	16	18	19	21	*GDF9* *PDE3A*	5;132199978 12;20769270	NM_005260.4 NM_000921.4	c.248C>G c.1376G>A	p.Ser83Cys p.Arg459Gln	ND 0.001566	Granulosa cell differentiation and proliferation Meiosis
Pt-35	Secondary	39	72	38	60	*PTCH1*	9;98215817	NM_000264.3	c.3392T>C	p.Val1131Ala	ND	Granulosa cell differentiation and proliferation
Pt-36	Secondary	27	102	64	5	*BMPR1B* *TSC2*	4;96051153 16;2138096	NM_001256793.1 NM_000548.3	c.816C>G c.5116C>T	p.Phe272Leu p.Arg1706Cys	0.000008247 0.0002665	Ovulation Follicular development
Pt-37	Primary	17	56	26	8	*BMPR1A*	10;88681384	NM_004329.2	c.1274A>G	p.Tyr425Cys	ND	Follicular development
Pt-38	Secondary	34	105	69	20	*LAMC1*	1;183079729	NM_002293.3	c.961C>T	p.Pro321Ser	0.0005848	Regulation of the extracellular matrix
Pt-39	Secondary	28	86	84	7	*ADAMTS16*	5;5146365	NM_139056.2	c.298C>T	p.Arg100Trp	0.0008860	Regulation of the extracellular matrix
Pt-41	Primary	17	42	17	45	*PTX3*	3;157160530	NM_002852.3	c.908C>G	p.Pro303Arg	0.0005518	Ovulation
Pt-42	Secondary	32	96	82	20	*FANCG*	9;35078733	NM_004629.1	c.176G>A	p.Gly59Glu	0.00003301	Meiosis
Pt-43	Secondary	35	77	52	9	*NOTCH2*	1;120462920	NM_024408.3	c.5411C>T	p.Ser1804Leu	0.00002472	Granulosa cell differentiation and
Pt-45	Secondary	34	136	46	2	*MCM9* *BMPR1B*	6;119234579 4;96051098	NM_017696.2 NM_001256793.1	c.911A>G c.761G>A	p.Asn304Ser p.Arg254His	0.003325 0.001081	Meiosis Ovulation
Pt-47	Secondary	35	12	24	7	*SEBOX*	17;26691490	NM_001080837.2	c.362_371delGCACCTCAGT	p.Ser116Ala*fs7	ND	Meiosis

Pt-49	Secondary	24	136	37	1	*FANCL*	2;58386928	NM_004629.1	c.1114_1115insATTA	p.Thr372Asnfs*11	ND	Meiosis
						ZP1	11;60637010	NM_207341.3	c.319G>A	p.Asp107Asn	0.002254	Follicular development
						BMPER	7;34086005	NM_133468.4	c.664C>T	p.Pro222Ser	0.0002637	Follicular development
Pt-51	Secondary	38	137	78	1	*NOTCH2*	1;120462898	NM_024408.3	c.5433G>C	p.Gln1811His	ND	Granulosa cell differentiation and proliferation
						CYP26B1	2;72362437	NM_019885.3	c.541G>A	p.Val181Met	0.00009900	Granulosa cell differentiation and proliferation
						PRDM1	6;106554919	NM_001198.3	c.2036G>A	p.Arg679His	0.00004120	Meiosis
						STAG3	7;99797247	NM_012447.3	c.1657G>A	p.Gly553Ser	ND	Meiosis
Pt-54	Secondary	16	65	22	12	*PADI6*	1;17698849	NM_207421.3	c.109C>T	p.Leu37Phe	ND	Follicular development
						KIT	4;55524204	NM_000222.2	c.23G>C	p.Trp8Ser	ND	Regulation of follicular development
						THBS1	15;39874613	NM_003246.2	c.287A>G	p.Gln96Arg	ND	Regulation of follicular development
Pt-55	Secondary	23	96	43	10	*MTHFR*	1;11850895	NM_005957.4	c.1813T>C	p.Ser605Pro	0.000008245	Cell metabolism
Pt-56	Secondary	31	75	68	10	*BRD2*	6;32942354	NM_001199456.1	c.4G>T; c.5C>G	p.Ala2Cys	ND	Meiosis
Pt-58	Secondary	15	60	29	15	*SOX15*	17;7492861	NM_006942.1	c.134C>T	p.Pro45Leu	ND	Other
						BMPR1A	10;88681435	NM_004329.2	c.1325G>A	p.Arg442His	ND	Follicular development
Pt-59	Secondary	23	23	20	19	*LEPR*	1;66064368	NM_002303.5	c.875C>A	p.Ser292Tyr	0.0001735	Ovulation

Pt-64	Secondary	37	114	33	10	*PCSK6* *SAPCD1*	15;101845484 6;31731303	NM_002570.3 NM_001039651.1	c.2891C>T c.226C>T	p.Thr964Met p.Gln76Ter	0.003727 0.0009166	Granulosa cell differentiation and proliferation Other
Pt-67	Secondary	35	38	15	8	*BMP5*	6;55739432	NM_021073.2	c.232C>T	p.Pro78Ser	ND	Granulosa cell differentiation and proliferation
Pt-68	Secondary	39	76	59	30	*C3orf77* *C3orf77*	3;44284349 3;44284351	NM_001145030.1 NM_001145030.1	c.351G>T c.353A>T	p.Lys117Asn p.Glu118Val	ND ND	Meiosis Meiosis

Figure 1

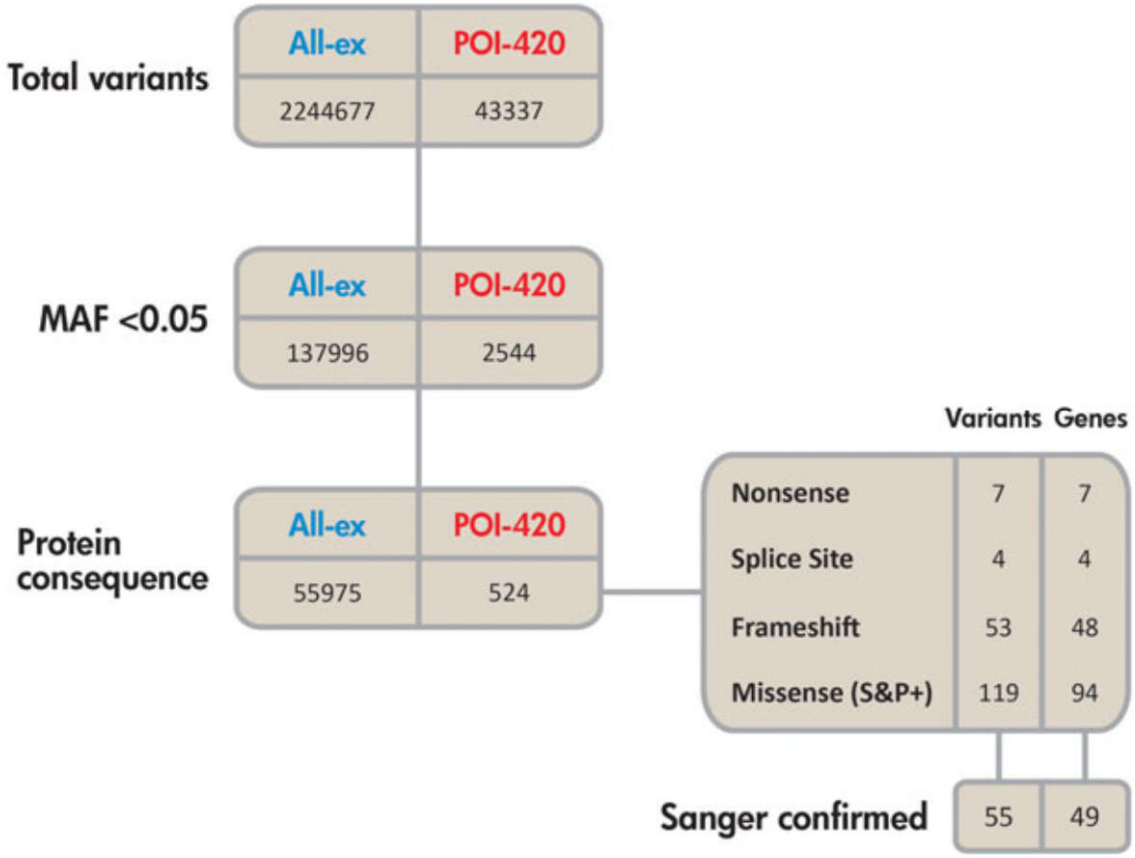
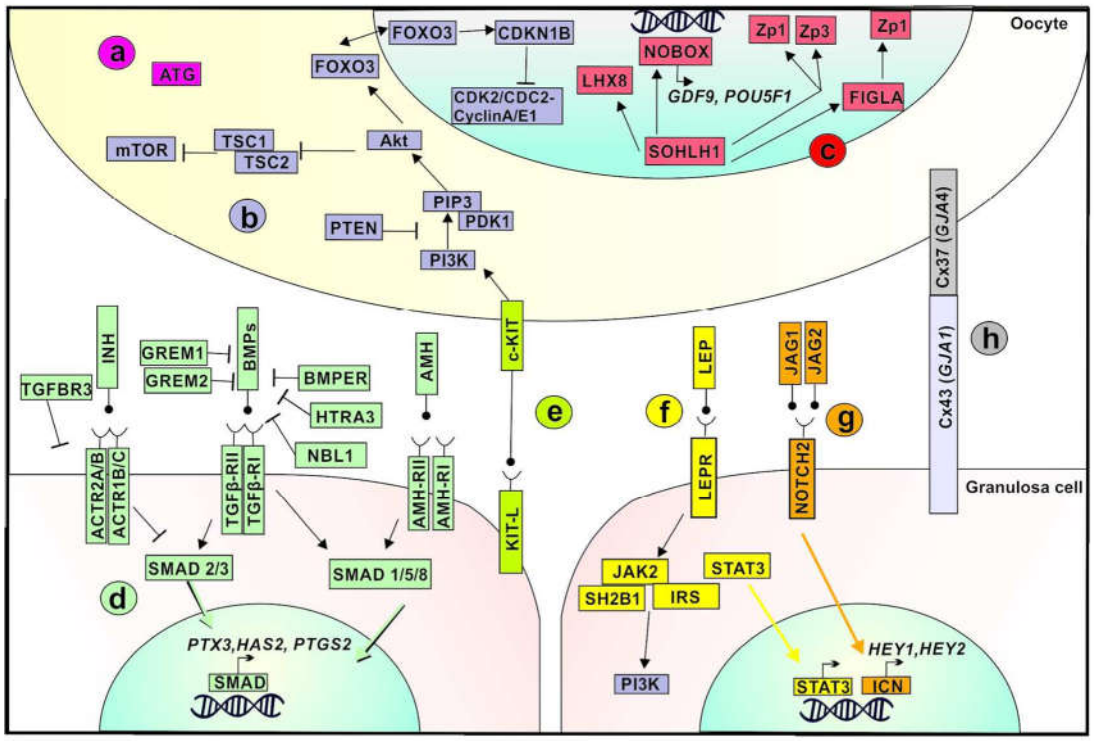

Total variants
All-ex
POI-420
2244677
43337
MAF <0.05
All-ex
POI-420
137996
2544
Protein consequence
All-ex
POI-420
55975
524
Variants
Genes
Nonsense
7
7
Splice Site
4
4
Frameshift
53
48
Missense (S&P+)
119
94
Sanger confirmed
55
49

Figure 2


Oocyte
a
ATG
FOXO3
FOXO3
CDKN1B
CDK2/CDC2-CyclinA/E1
NOBOX
LHX8
GDF9, POU5F1
Zp1
Zp3
Zp1
FIGLA
SOHLH1
c
mTOR
TSC1
TSC2
Akt
b
PIP3
PTEN
PDK1
PI3K
c-KIT
Cx37 (GJA4)
Cx43 (GJA1)
h
Granulosa cell
INH
TGFBR3
GREM1
GREM2
BMPs
BMPER
HTRA3
NBL1
AMH
e
f
LEP
JAG1
JAG2
g
ACTR2A/B
ACTR1B/C
TGFβ-RII
TGFβ-RI
AMH-RII
AMH-RI
KIT-L
LEPR
NOTCH2
SMAD 2/3
SMAD 1/5/8
d
JAK2
STAT3
SH2B1
IRS
PTX3,HAS2, PTGS2
SMAD
PI3K
HEY1,HEY2
STAT3
ICN

Supplemental table S1. POI gene subset (POI-420) analyzed via NGS

Gene	Gene name
ACVR2A	Activin a receptor, type iia
ADAMTS1	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 1
ADAMTS15	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 15
ADAMTS16	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 16
ADAMTS19	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 19
ADAMTS4	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 4
ADAMTS5	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 5
ADAMTS6	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 6
ADIPOR1	Adiponectin receptor 1
ADIPOR2	Adiponectin receptor 2
AFP	Alpha-fetoprotein
AHR	Aryl hydrocarbon receptor
AKT	Akt serine/threonine kinase 1
ALK4	Activin a receptor type 1b
ALK6	Bone morphogenetic protein receptor type 1b
ALK7	Activin a receptor type 1c
AMBP	Alpha-1 microglobulin/bikunin precursor
AMH	Anti-mullerian hormone
AMHR2	Anti-mullerian hormone receptor type 2
AR	Androgen receptor
AREG	Amphiregulin
ARHGEF7	Rho guanine nucleotide exchange factor 7
ARNTL	Aryl hydrocarbon receptor nuclear translocator like
ATG10	Autophagy-Related Protein 10
ATG16L1	Autophagy related 16 like 1
ATG2A	Autophagy-related protein 2 homolog A
ATG4A	Autophagy related 4A Cysteine Peptidase
ATG4B	Autophagy Related 4B Cysteine Peptidase
ATG4C	Autophagy related 4C Cysteine Peptidase
ATG5	Autophagy related 5
ATG7	Autophagy related 7
ATG9A	Autophagy-related protein 9A
ATG9B	Autophagy-related protein 9B
ATM	Ataxia-telangiectasia mutated gene
AURKA	Aurora kinase a
AURKB	Aurora kinase b
AURKC	Aurora kinase c
BAX	BCL2 associated X, apoptosis regulator

BCL2	BCL2, apoptosis regulator
BCL2L1	BCL2 like 1
BCL2L2	BCL2 like 2
BCL6	B-cell CLL/lymphoma 6
BDNF	Brain derived neurotrophic factor
BMAL1	Aryl hydrocarbon receptor nuclear translocator-like
BMP15	Bone morphogenetic protein 15
BMP2	Bone morphogenetic protein 2
BMP4	Bone morphogenetic protein 4
BMP5	Bone morphogenetic protein 5
BMP6	Bone morphogenetic protein 6
BMP7	Bone morphogenetic protein 7
BMP8B	Bone morphogenetic protein 8b
BMPER	BMP binding endothelial regulator
BMPR1A	Bone morphogenetic protein receptor type 1A
BMPR1B	Bone morphogenetic protein receptor type 1B
BMPR2	Bone morphogenetic protein receptor type 2
BOLL	Boule-Like RNA Binding Protein
BRCA1	Breast cancer 1 gene
BRD2	Bromodomain containing 2
BRD3	Bromodomain containing 3
BRD4	Bromodomain containing 4
BRDT	Bromodomain testis associated
BRSK1	BR serine/threonine kinase 1
BRWD1	Bromodomain and WD repeat domain containing 1
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B
BVES	Blood vessel epicardial substance
C1GALT1	Core 1 synthase, glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1
CASP2	Caspase 2
CBX2	Chromobox 2
CCNA1	Cyclin A1
CCNB1IP1	Cyclin B1 interacting protein 1
CCND2	Cyclin D2
CDC25B	Cell division cycle 25B
CDK2	Cyclin dependent kinase 2
CDK4	Cyclin dependent kinase 4
CDKN1B	Cyclin dependent kinase inhibitor 1B
CDKN1C	Cyclin dependent kinase inhibitor 1C
CEBPA	CCAAT/enhancer binding protein alpha
CEBPB	CCAAT/enhancer binding protein beta
CGA	Glycoprotein hormones, alpha polypeptide

CITED2	Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2
CKS2	CDC28 protein kinase regulatory subunit 2
CMYC	V-myc avian myelocytomatosis viral oncogene homolog
CPE	Carboxypeptidase E
CPEB1	Cytoplasmic polyadenylation element binding protein 1
CRTC1	CREB regulated transcription coactivator 1
CTGF	Connective tissue growth factor
CTNNB1	Catenin beta 1
CUGBP1	CUGBP, Elav-like family member 1
CXCL19	Chemokine (C-X-C motif) ligand 19
CXCR4	C-X-C motif chemokine receptor 4
CYP11B1	Cytochrome P450 family 11 subfamily B member 1
CYP17A1	Cytochrome P450 family 17 subfamily A member 1
CYP19A1	Cytochrome P450 family 19 subfamily A member 1
CYP21A2	Cytochrome P450 family 21 subfamily A member 2
CYP26B1	Cytochrome P450 family 26 subfamily B member 1
CYP27B1	Cytochrome P450 family 27 subfamily B member 1
DAND5	DAN domain BMP antagonist family member 5
DAZL	Deleted in azoospermia like
DDR2	Discoidin domain receptor tyrosine kinase 2
DHCR24	24-dehydrocholesterol reductase
DICER1	Dicer 1, ribonuclease III
DLX5	Distal-less homeobox 5
DLX6	Distal-less homeobox 6
DMC1	DNA meiotic recombinase 1
DMRT1	Doublesex and mab-3 related transcription factor 1
DMRT3	Doublesex and mab-3 related transcription factor 3
DND1	DND microrna-mediated repression inhibitor 1
DPPA2	Developmental pluripotency associated 2
EDNRB	Endothelin receptor type B
EGR1	Early growth response 1
EIF4ENIF1	Eukaryotic translation initiation factor 4E nuclear import factor 1
EPAB	Poly(A) binding protein cytoplasmic
ERCC1	ERCC excision repair 1, endonuclease non-catalytic subunit
ERCC2	ERCC excision repair 2, endonuclease non-catalytic subunit
EREG	Epiregulin
ERK1	Mitogen-activated protein kinase 3
ERK2	Mitogen-activated protein kinase 1
ESCO2	Establishment of sister chromatid cohesion N-acetyltransferase 2
ESR1	Estrogen receptor 1
ESR2	Estrogen receptor 2

EVI1	MDS1 and EVI1 complex locus
EXO1	Exonuclease 1
FABP6	Fatty acid binding protein 6
FANCA	Fanconi anemia complementation group A
FANCC	Fanconi anemia complementation group C
FANCG	Fanconi anemia complementation group G
FANCL	Fanconi anemia complementation group L
FGF2	Fibroblast growth factor 2
FGF9	Fibroblast growth factor 9
FGFR1	Fibroblast growth factor receptor 1
FGFR2	Fibroblast growth factor receptor 2
FHL2	Four and a half LIM domains 2
FIGLA	Olliculogenesis specific bhlh transcription factor
FKTN	Fukutin
FMN2	Formin 2
FMR1	Fragile X mental retardation 1
FOG2	Zinc finger protein, FOG family member 2
FOXC1	Forkhead box C1
FOXE1	Forkhead box E1
FOXG1B	Forkhead box G1
FOXL2	Forkhead box L2
FOXL3	Forkhead box L3
FOXO3	Forkhead box O3
FOXO4	Forkhead box O4
FSD1L	Fibronectin type III and SPRY domain containing 1 like
FSHB	Follicle stimulating hormone beta subunit
FSHR	Follicle stimulating hormone receptor
FST	Follistatin
FSTL3	Follistatin like 3
FZD1	Frizzled class receptor 1
FZD4	Frizzled class receptor 4
FZR1	Fizzy/cell division cycle 20 related 1
GADD45G	Growth arrest and DNA damage inducible gamma
GATA4	GATA binding protein 4
GATA6	GATA binding protein 6
GCM2	Glial cells missing homolog 2
GCX1	TOX high mobility group box family member 2
GDF9	Growth differentiation factor 9
GGN1	Gametogenetin 1
GGT1	Gamma-glutamyltransferase 1
GGT5	Gamma-glutamyltransferase 5

GJA1	Gap junction protein alpha 1
GJA4	Gap junction protein alpha 4
GLI1	GLI family zinc finger 1
GLP1	Glucagon-like peptide 1, included
GNRH1	Gonadotropin releasing hormone 1
GNRHR	Gonadotropin releasing hormone receptor
GOLT1A	Golgi transport 1A
GPR3	G protein-coupled receptor 3
GREM1	Gremlin 1, DAN family BMP antagonist
GREM2	Gremlin 2, DAN family BMP antagonist
GULP	GULP, engulfment adaptor PTB domain containing 1
H2AFX	H2A histone family member X
HACE1	HECT domain and ankyrin repeat containing E3 ubiquitin protein ligase 1
HAS2	Hyaluronan synthase 2
HDAC1	Histone deacetylase 1
HDAC2	Histone deacetylase 2
HDX	Highly divergent homeobox
HES1	Hes family bhlh transcription factor 1
HEY2	Hes related family bhlh transcription factor with YRPW motif 2
HHIP	Hedgehog interacting protein
HK3	Hexokinase 3
HNRNPK	Heterogeneous nuclear ribonucleoprotein K
HORMAD1	HORMA domain containing 1
HOXA5	Homeobox A5
HPGD	Hydroxyprostaglandin dehydrogenase 15-(NAD)
HPRT1	Hypoxanthine phosphoribosyltransferase 1
HSD17B4	Hydroxysteroid 17-beta dehydrogenase 4
HSF2	Heat shock transcription factor 2
HSP10	Heat-shock 10-kd protein
HSP27	Heat shock protein family B (small) member 1
HTRA1	Htra serine peptidase 1
HTRA3	Htra serine peptidase 3
IGF1	Insulin like growth factor 1
IGF2R	Insulin like growth factor 2 receptor
IL6ST	Interleukin 6 signal transducer
IMMP2L	Inner mitochondrial membrane peptidase subunit 2
INHA	Inhibin alpha subunit
INHBA	Inhibin beta A subunit
INHBB	Inhibin beta B subunit
INHBC	Inhibin beta C subunit
INSL3	Insulin like 3

IRS2	Insulin receptor substrate 2
JAGGED1	Jagged 1
JAK2	Janus kinase 2
JMJD1A	Lysine demethylase 3A
KDR	Kinase insert domain receptor
KISS1	Kiss-1 metastasis-suppressor
KISS1R	KISS1 receptor
KIT	KIT proto-oncogene receptor tyrosine kinase
KITLG	KIT ligand
LAMC1	Laminin subunit gamma 1
LARS2	Leucyl-trna Sybthetase 2
LATS1	Large Tumor Suppressor Kinase 1
LBX2	Ladybird homeobox 2
LEP	Leptin
LEPR	Leptin receptor
LFNG	Lunatic fringe
LGR4	Leucine rich repeat containing G protein-coupled receptor 4
LHB	Luteinizing hormone beta polypeptide
LHCGR	Luteinizing hormone/choriogonadotropin receptor
LHX8	Lim homeobox gene 8
LHX9	LIM homeobox 9
LIN28A	Protein lin-28 homolog A
LIN28B	Protein lin-28 homolog B
LOX1	Low density lipoprotein, oxidized, receptor 1
MAP3K4	Mitogen-activated protein kinase kinase kinase 4
MAPK14	Mitogen-activated protein kinase 14
MCL1	Myeloid cell leukemia sequence 1
MCM8	Minichromosome maintenance complex component 8
MCM9	Minichromosome maintenance complex component 9
MEI1	Meiosis inhibitor protein 1
MGARP	Mitochondria localized glutamic acid rich protein
MLH1	DNA mismatch repair protein Mlh1
MLH3	DNA mismatch repair protein Mlh3
MMP2	Matrix metallopeptidase 2
MOGAT1	Monoacylglycerol o-acyltransferase 1
MSH4	Muts homolog 4
MSH5	Muts homolog 5
MSX1	Msh homeobox 1
MSX2	Homeobox protein MSX-2
MTHFR	5,10-methylenetetrahydrofolate reductase
MTOR	Mammalian target of rapamycin

MTRR	Methionine synthase reductase
NALP5	NLR family pyrin domain containing 5
NANOS2	Nanos homolog 2
NANOS3	Nanos C2HC-type zinc finger 3
NAT9	N-acetyltransferase 9
NBL1	Neuroblastoma candidate region, suppression of tumorigenicity 1
NBN	Nibrin
NHLH2	Nescient helix-loop-helix 2
NOBOX	Homeobox protein NOBOX
NOHLH	Spermatogenesis and oogenesis specific basic helix-loop-helix 1
NOS1	Nitric oxide synthase 1
NOS3	Nitric oxide synthase 3
NOTCH2	Neurogenic locus notch homolog protein 2
NR2C2	Nuclear receptor subfamily 2, group c, member 2
NR5A1	Nuclear receptor subfamily 5 group A member 1
NR5A2	Nuclear receptor subfamily 5 group A member 2
NRG1	Neuregulin 1
NRIP1	Nuclear receptor interacting protein 1
NTF4	Neurotrophin 4
NTRK2	Neurotrophic tyrosine kinase, receptor, type 2
NUR77	Nuclear receptor subfamily 4 group A member 1
OOSP1	Oocyte secreted protein 1, pseudogene
P2Y2	Purinergic receptor P2Y, g protein-coupled, 2
P2Y2R	Purinergic receptor P2Y2
P2Y6	Pyrimidinergic receptor P2Y6
P2Y6R	Pyrimidinergic receptor P2Y6
PADI6	Peptidylarginine deiminase, type vi
PCNA	Proliferating cell nuclear antigen
PCSK1	Proprotein convertase, subtilisin/kexin-type, 1
PCSK5	Proprotein convertase, subtilisin/kexin-type, 5
PCSK6	Proprotein convertase, subtilisin/kexin-type, 6
PCYT1B	Phosphate cytidylyltransferase 1, choline, beta
PDE3A	Phosphodiesterase 3A, cGMP-Inhibited
PDE4D	Phosphodiesterase 4D, cAMP-Specific
PDPK1	3-phosphoinositide dependent protein kinase 1
PER1	period circadian clock 1
PGD2	Prostaglandin D2 synthase, brain
PGR	Progesterone receptor
PGRMC1	Progesterone receptor membrane component 1
PHB	Prohibitin
PIK3CA	Phosphatidylinositol 3-kinase, catalytic, alpha

PIK3CG	Phosphatidylinositol 3-kinase, catalytic, gamma
PMS2	PMS1 homolog 2, mismatch repair system component
POPDC3	Popeye domain containing 3
POR	Cytochrome P450 oxidoreductase
POU1F1	Pou domain, class 1, transcription factor 1
POU5F1	POU class 5 homeobox 1
PPM1A	Protein phosphatase, Mg^{2+}/Mn^{2+} dependent 1A
PPP2R1A	Protein phosphatase 2, structural/regulatory subunit a, alpha
PRDM1	PR domain-containing protein 1
PRDX2	Peroxiredoxin 2
PRL	Prolactin
PRLR	Prolactin receptor
PROP1	Prop paired-like homeobox 1
PSMC3IP	PSMC3-interacting protein
PTCH1	Protein patched homolog 1
PTEN	Phosphatase and tensin homolog
PTGER2	Prostaglandin e receptor 2, EP2 subtype
PTGS2	Prostaglandin-endoperoxide synthase 2
PTX3	Pentraxin 3, long
RAD51C	RAD51 paralog C
RBMS1	RNA-binding motif protein, single strand-interacting, 1
REC8	REC8 meiotic recombination protein
RHOX13	Reproductive homeobox 13
RHOX5	Rhox homeobox family, member 1
RHOX8	Rhox homeobox family, member 8
RHOXF2	Rhox homeobox family, member 2
RHOXF2B	Rhox homeobox family member 1, pseudogene 1
RICTOR	Rapamycin-insensitive companion of MTOR
RNF35	Tripartite motif-containing protein 40
RPS6KB1	Ribosomal protein S6 kinase, 70-KD, 1
RSPO1	R-spondin family, member 1
RUNX2	Ribosomal protein s6 kinase, 70-KD, 1
SAM68	KH domain-containing, RNA-binding, signal transduction-associated protein 1
SCARB1	Scavenger receptor class b, member 1
SDF1	Chemokine, CXC motif, ligand 12
SEBOX	Skin-, embryo-, brain-, and oocyte-specific homeobox
SETDB2	Set domain protein, bifurcated, 2
SGOL2	Shugoshin-like 2
SH2B1	Sh2b adaptor protein 1
SIGLEC11	Sialic acid-binding immunoglobulin-like lectin 11
SIRT1	Sirtuin 1

SIX1	Sine Oculis Homeobox Homolog 1
SIX4	Sine Oculis Homeobox Homolog 4
SKP2	S-phase kinase-associated protein 2
SLC44A1	Solute carrier family 44, member 1
SMAD1	SMAD family member 1
SMAD2	SMAD family member 2
SMAD3	SMAD family member 3
SMAD4	SMAD family member 4
SMAD5	SMAD family member 5
SMAD8	SMAD family member 8
SMAD9	SMAD family member 9
SMC1B	Structural maintenance of chromosomes 1b
SMOM2	Smoothened
SOD1	Superoxide dismutase 1
SOHLH1	Spermatogenesis and oogenesis-specific basic helix-loop-helix protein 1
SOHLH2	Spermatogenesis and oogenesis-specific basic helix-loop-helix protein 2
SOX15	Sry-box 15
SOX3	Sry-box 3
SOX8	Sry-box 8
SOX9	Sry-box 9
SPO11	SPO11, initiator of meiotic double stranded breaks
SRC	V-src avian sarcoma (schmidt-ruppin a-2) viral oncogene
SSTR2	Somatostatin receptor 2
STAG3	Stromalin 3
STAR	Steroidogenic acute regulatory protein
STAT3	Signal transducer and activator of transcription 3
STRA8	Stimulated by retinoic acid 8
SULT1E1	Sulfotransferase family 1e, estrogen-preferring, member 1
SUV420H2	Suppressor of variegation 4-20
SYCE1	Synaptonemal complex central element protein 1
SYCE2	Synaptonemal complex central element protein 2
SYCE3	Synaptonemal complex central element protein 3
SYCP1	Synaptonemal complex protein 1
SYCP2	Synaptonemal complex protein 2
SYCP2L	Synaptonemal complex protein 2-like
SYCP3	Synaptonemal complex protein 3
TAF4B	TAF4B RNA polymerase ii, tata box-binding protein-associated factor
TAL2	T-cell acute lymphocytic leukemia 2
TBB8	Tubulin beta 8 class VIII
TCF21	Transcription factor 21
TERT	Telomerase reverse transcriptase

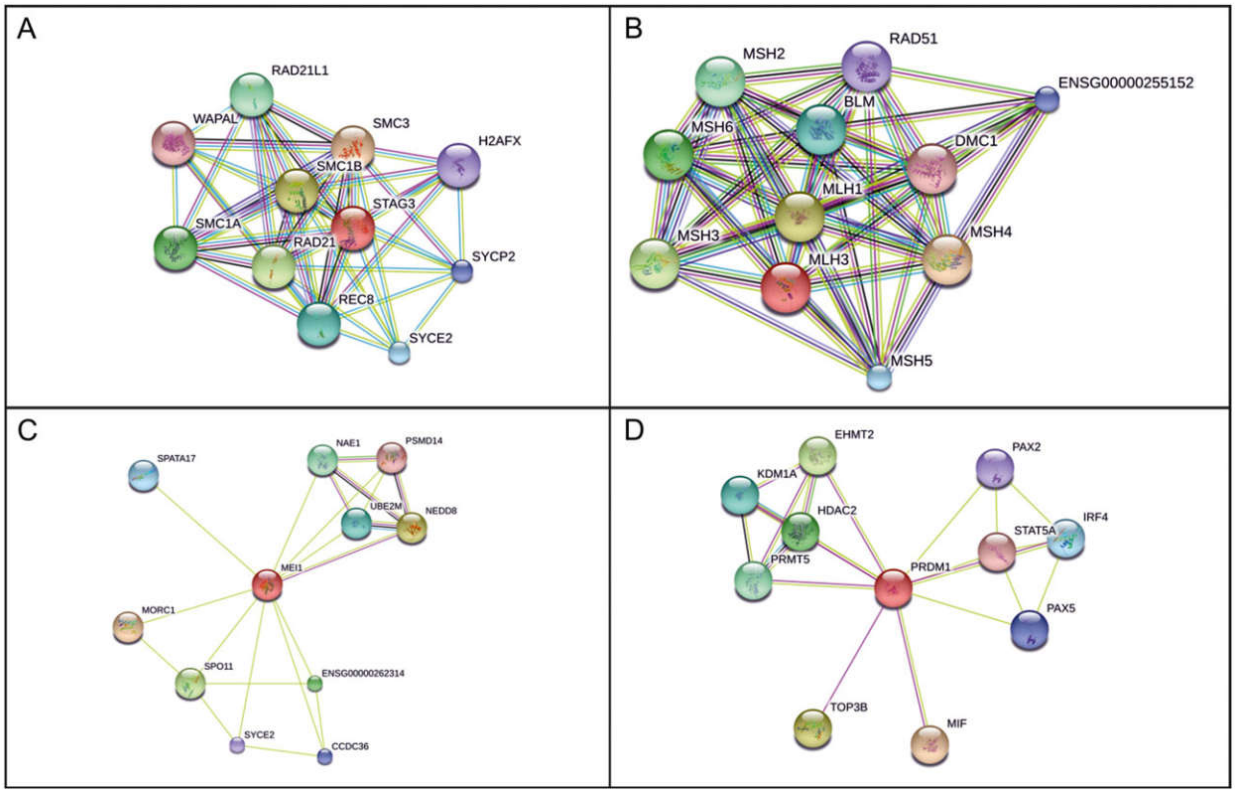
TGFB1	Transforming growth factor, beta-1
TGFBR3	Transforming growth factor-beta receptor, type III
THBS1	Thrombospondin I
TIAL1	Tia1 cytotoxic granule-associated rna-binding protein-like 1
TIMP3	Tissue inhibitor of metalloproteinase 3
TMEM38B	Transmembrane protein 38b
TNFAIP6	Tumor necrosis factor-alpha-induced protein 6
TOP3B	Topoisomerase, DNA, III, beta
TOPAZ1	Chromosome 3 open reading frame 77
TORC1	Creb-regulated transcription coactivator 1
TP53	Tumor protein p53
TP73	Tumor protein p73
TRIP13	Thyroid hormone receptor interactor 13
TRKB	Neurotrophic tyrosine kinase, receptor, type 2
TRMT6	tRNA methyltransferase 6
TSC1	Tuberous sclerosis 1
TSC2	Tuberous sclerosis 2
TWSG1	Twisted gastrulation BMP signaling modulator 1
UBB	Ubiquitin b
UBE3A	Ubiquitin-protein ligase E3A
UBR2	Ubiquitin-protein ligase E3 component n-recognin 2
UIMC1	Ubiquitin interaction motif-containing protein 1
UMODL1	Uromodulin-like 1
UNC5A	UNC-5 netrin receptor A
USP9X	Ubiquitin-specific protease 9, x-linked
USP9Y	Ubiquitin-specific protease 9, y chromosome
VDR	Vitamin D receptor
VRK1	Vaccinia-related kinase 1
VWC2	Von willebrand factor c domain-containing protein 2
WNT2	Wingless-type MMTV integration site family, member 2
WNT4	Wingless-type MMTV integration site family, member 4
WNT5A	Wingless-type MMTV integration site family, member 5a
WNT7A	Wingless-type MMTV integration site family, member 7a
WT1	Wilms tumor 1
YBX2	Y box-binding protein 2
YY1	Transcription factor yy1
ZFAND3	zinc finger, AN1-type domain 3
ZFP36L2	Zinc finger protein 36-like 2
ZFX	Zinc finger protein, x-linked
ZNF346	Zinc finger protein 346
ZNF462	Zinc finger protein 462

ZP1	Zona pellucida glycoprotein 1
ZP2	Zona pellucida glycoprotein 2
ZP3	Zona pellucida glycoprotein 3

Supplemental table S2. Available protein structures for modelling mutations identified via next generation sequencing. Structures used for fragment molecular orbital analysis are indicated in bold

Gene	Mutation		Patient ID	PDB ID	Fragment (aa)
	DNA	Protein			
BMPR1B	**c.761G>A**	**p.Arg254His**	**Pt-45**	**3MDY**	**168-502**
BMPR1B	**c.816C>G**	**p.Phe272Leu**	**Pt-36**	**3MDY**	**168-502**
CXCR4	c.415G>A	p.Val139Ile	Pt-22	3ODU	2-319
				3OE0	2-319
				3OE6	2-325
				3OE8	2-319
				3OE9	2-319
				4RWS	2-228 and 231-319
FANCL	c.1114_1115insATTA	p.Thr372Asnfs*11	Pt-49	4CCG	288-375
GREM1	**c.506G>C**	**p.Arg169Thr**	**Pt-24**	**5AEJ**	**72-184**
HTRA3	c.1006C>T	p.Arg336Cys	Pt-14	4RI0	130-453
THBS1	c.287A>G	p.Gln96Arg	Pt-54	1Z78	19-233
				1ZA4	19-233
				2ERF	25-233
				2ES3	25-233
				2OUH	19-257

Supplementary Figure S1



Supplementary Figure S2

A

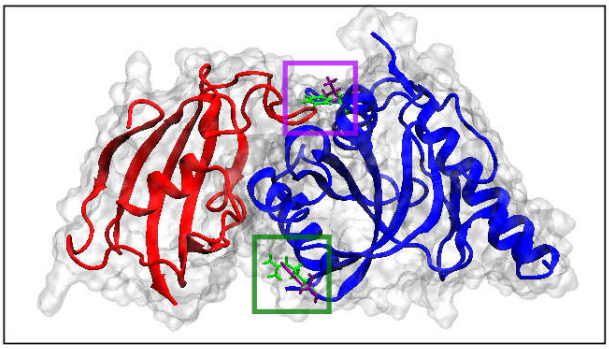
BMPR1B_WT (Arg254A)						
Residue	ΔE_{es}	ΔE_{ex}	ΔE_{ct}	ΔE_{disp}	ΔG_{sol}	Total
Glu204A	-60.61	2.45	-1.71	-5.12	34.35	**-30.64**
Gln233A						**-12.64**
Glu256A						**-4.43**
Glu55B	-42.83	0.00	0.00	0.00	18.71	**-24.12**
					Total interaction =	**-54.75**

BMPR1B_MT (His254A)						
	ΔE_{es}	ΔE_{ex}	ΔE_{ct}	ΔE_{disp}	ΔG_{sol}	Total
	9.49	1.10	-1.73	-3.22	-10.51	**-4.88**
	-13.31	3.81	-1.55	-1.80	0.20	**-12.64**
	0.83	6.75	-3.89	-2.57	-5.56	**-4.43**
	-7.06	0.00	0.00	0.00	-0.53	**-7.59**
					Total interaction =	**-29.54**

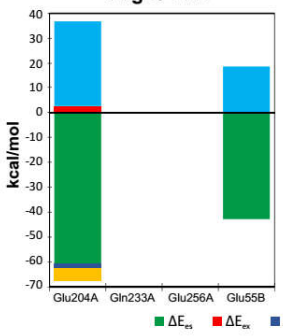
BMPR1B_WT (Phe272A)						
Residue	ΔE_{es}	ΔE_{ex}	ΔE_{ct}	ΔE_{disp}	ΔG_{sol}	Total
Glu268A	-5.53	1.58	-1.44	-3.13	2.09	**-6.43**
Glu268A	-15.53	10.15	-3.87	-3.99	-0.15	**-13.39**
Glu276A	-31.37	9.27	-3.58	-3.30	11.41	**-17.57**
Pro89B	-3.08	5.28	-2.13	-7.83	0.44	**-7.31**
					Total interaction =	**-44.69**

BMPR1B_MT (Leu272A)						
	ΔE_{es}	ΔE_{ex}	ΔE_{ct}	ΔE_{disp}	ΔG_{sol}	Total
	-17.20	10.64	-4.31	-4.63	2.73	**-12.77**
	-39.66	9.34	-3.85	-3.68	17.22	**-20.62**
					Total interaction =	**-33.38**

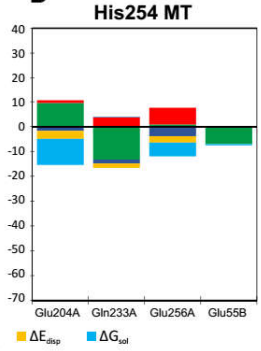
B



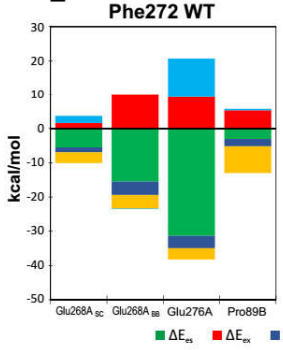
C Arg254 WT



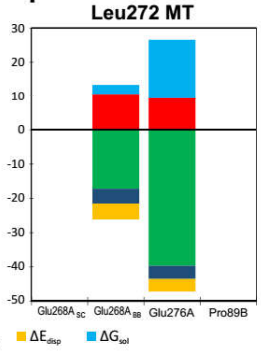
D His254 MT



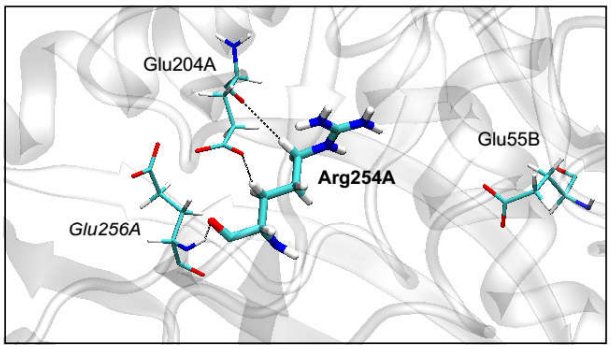
E Phe272 WT



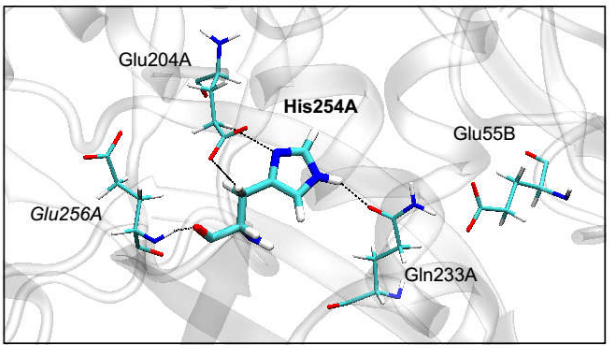
F Leu272 MT



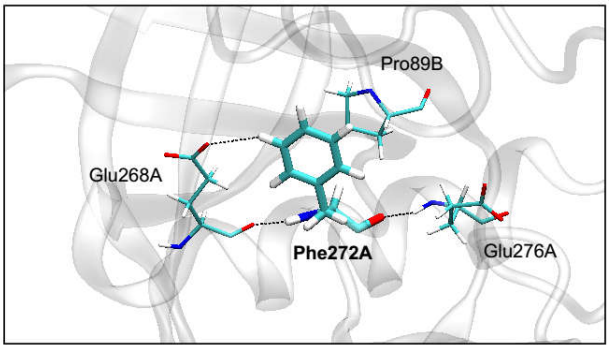
G



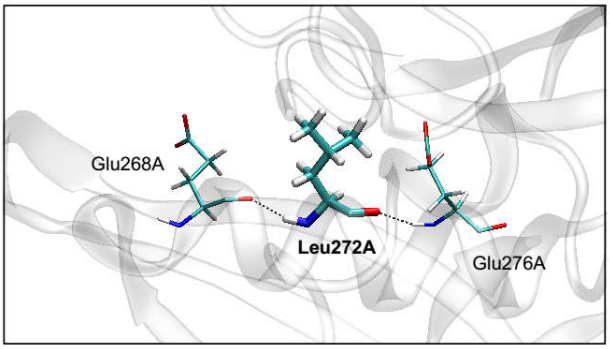
H



I



J



Supplementary Figure S3

A

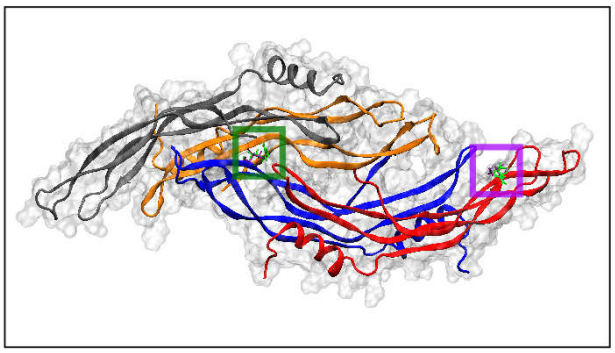
GREM1_WT S1 (Arg169A)						
Residue	ΔE_{es}	ΔE_{ex}	ΔE_{ct}	ΔE_{disp}	ΔG_{sol}	Total
Thr151A	-0.21	1.04	-0.18	-2.22	-3.61	**-5.18**
Met153A	-10.71	8.62	-3.95	-7.21	1.35	**-11.91**
Glu134B	-35.63	0.00	0.00	0.00	16.25	**-19.37**
Glu135B	-118.36	16.22	-7.57	-2.94	9.31	**-103.35**
Gln139C	-5.27	0.03	-0.31	-0.88	-0.52	**-6.94**
Asp182C	-42.00	2.07	-1.62	-5.37	18.77	**-28.15**
Leu183C	-35.49	9.83	-2.90	-2.98	5.56	**-25.98**
Asp184C	-53.24	14.74	-4.34	-4.47	8.34	**-38.97**
					Total interaction =	**-239.86**

GREM1_MT S1 (Thr169A)						
ΔE_{es}	ΔE_{ex}	ΔE_{ct}	ΔE_{disp}	ΔG_{sol}	Total
-19.2	10.9	-5.5	-4.1	2.2	**-15.7**
-12.7	1.9	-0.9	-3.1	2.5	**-12.2**
				Total interaction =	**-27.86**

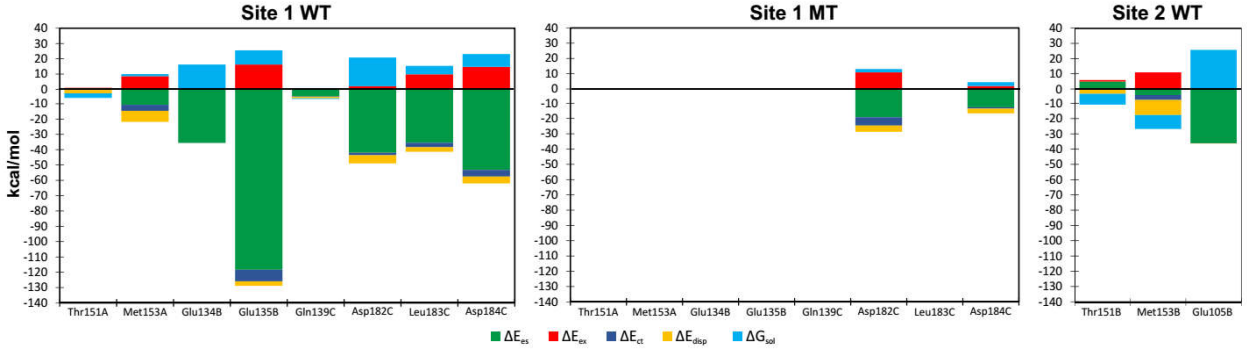
GREM1_WT S2 (Arg169B)						
Residue	ΔE_{es}	ΔE_{ex}	ΔE_{ct}	ΔE_{disp}	ΔG_{sol}	Total
Thr151B	5.16	0.94	-0.40	-2.52	-7.79	**-4.61**
Met153B	-3.93	11.09	-3.73	-9.87	-9.09	**-15.53**
Glu105B	-36.05	0.00	0.05	-0.07	25.80	**-10.27**
					Total interaction =	**-30.41**

GREM1_MT S2 (Thr169B)						
ΔE_{es}	ΔE_{ex}	ΔE_{ct}	ΔE_{disp}	ΔG_{sol}	Total
				Total interaction =	**0.00**

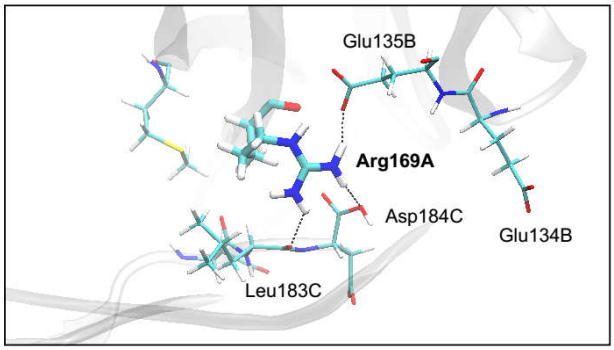
B



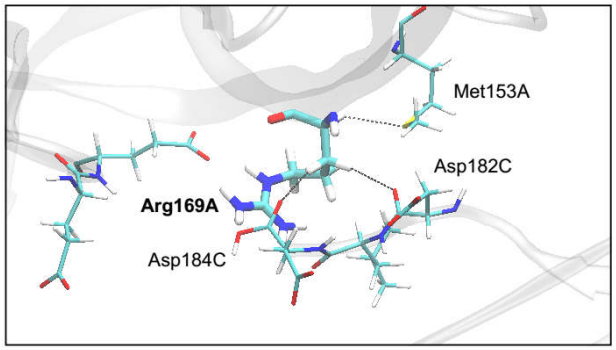
C



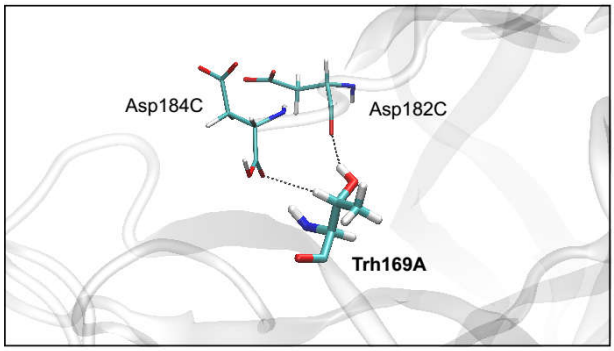
F



G



H



I