UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS

DIVERSIDAD Y DISTRIBUCIÓN EN EL ESPACIO QUÍMICO DE PÉPTIDOS PARA EL DESARROLLO DE FÁRMACOS.

PROYECTO DE INVESTIGACIÓN

PARA OPTAR POR EL GRADO DE

MAESTRA EN CIENCIAS

PRESENTA

QFB. Bárbara Itzel Díaz Eufracio

Dr. José Luis Medina Franco
Facultad de Química, UNAM

Ciudad Universitaria. CDMX. Junio 2018

Universidad Nacional Autónoma de México

Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS

DIVERSIDAD Y DISTRIBUCIÓN EN EL ESPACIO QUÍMICO DE PÉPTIDOS PARA EL DESARROLLO DE FÁRMACOS.

PROYECTO DE INVESTIGACIÓN

PARA OPTAR POR EL GRADO DE

MAESTRA EN CIENCIAS

P R E S E N T A

QFB. Bárbara Itzel Díaz Eufracio

CDMX, Junio de 2018

El presente trabajo fue realizado en el grupo DIFACQUIM ubicado en el cubículo 108 del Edificio F de la Facultad de Química de la UNAM. Durante el periodo entre enero de 2017 y junio de 2018 bajo de supervisión del Dr. José Luis Medina Franco.

Vo.Bo.
Dr. José Luis Medina Franco

QFB. Bárbara Itzel Díaz Eufracio

Durante la realización de este proyecto, se obtuvieron como producto las siguientes publicaciones. Se anexan al final de este trabajo escrito:

- Díaz-Eufracio, B. I., Naveja, J. J., & Medina-Franco, J. L. (2018). Chapter Three - Protein–Protein Interaction Modulators for Epigenetic Therapies. In R. Donev (Ed.), Protein-Protein Interactions in Human Disease, Part A (Vol. 110, pp. 65–84). Academic Press.

- Díaz-Eufracio, B. I., Palomino-Hernández, O., Houghten, R. A., & Medina-Franco, J. L. (2018). Exploring the chemical space of peptides for drug discovery: a focus on linear and cyclic penta-peptides. Molecular Diversity. En prensa. DOI:10.1007/s11030-018-9812-9

Adicionalmente, los avances y resultados parciales fueron presentados en los siguientes congresos:

- "Diversidad y Distribución en el Espacio Químico de Péptidos Para el Desarrollo de Fŕmacos"Bárbara Itzel Díaz Eufracio, Oscar Palomino Hernández y José Luis Medina-Franco. 52ºCongreso Mexicano de Química y 36ºCongreso Nacional de Educación Química. División de Química Teótica y Computacional. 26 al 29 de Septiembre de 2017 en Puerto Vallarta, Jalisco, México.

- "D-Peptide Builder: A Web-based application to enumerate the chemical space of peptides"Bárbara Itzel Díaz Eufracio, Oscar Palomino Hernández y José Luis Medina-Franco. 255th American Chemica Society (ACS) 2018 National Meeting and Exposition. Chemical Information Division. 16 al 22 de Marzo de 2018 en Nueva Orleans, Estados Unidos.

Agradecimientos Institucionales

- Especialmente a los miembros del jurado, por sus valiosas observaciones para enriquecer el presente trabajo.
- Agradezco al CONACyT por la beca 620289 otorgada para la realización de este trabajo.
- Al programa PAEP otorgado por el posgrado en Ciencias Químicas de la UNAM, el cual contribuyó la presentación de este trabajo en la reunión 255th American Chemica Society (ACS).
- A la SQM por la beca otorgada en el 52ºCongreso Mexicano de Química y 36ºCongreso Nacional de Educación Química.
- Al programa Nuevas Alternativas para el Tratamiento de Enfermedades Infecciosas(NUATEI), Instituto de Investigaciones Biomédicas, UNAM por facilitar la adquisición de la licencia de Molecular Operating Enviroment (MOE).

Resumen

Recientemente los péptidos y sus derivados han aumentado su interés en el diseño de fármacos, puesto que son considerados como un puente entre las moléculas pequeñas y las proteínas. Estudios previos sugieren que los péptidos que exhiben actividad biológica poseen estructuras químicas muy diversas cuyas propiedades fisicoquímicas pueden ser atractivas para desarrollar nuevos candidatos. Sin embargo, son limitados los estudios quimionformáticos que caracterizan dicha diversidad. Es precisamente la falta de información acerca de la diversidad de los péptidos la razón que motiva la realización de este estudio.
La relavancia de este trabajo se asocia con el estudio del espacio químico de péptidos, el cual a diferencia del de fármacos de moléculas pequeñas ha sido poco estudiado. También se analiza el impacto de las modificaciones estructurales en el espacio químico y la diversidad estructural de los péptidos enumerados.
En el presente trabajo se realizó una caracterización de las propiedades fisicoquímicas y un análisis del espacio químico de cuatro bibliotecas peptidicas cíclicas y lineales. Específicamente esta investigación se enfoca en pentapéptidos.
La enumeración de los péptidos se automatizó con un *script* en el lenguaje de programación en R, actividad que facilitó la enumeración de cuatro bibliotecas con 32^5 compuestos cada una. El espacio químico de dichas bibliotecas fué comparado con colecciones de compuestos de interés farmaceútico.
Los resultados indican que existe un mayor traslape entre la bibliotecas de cíclicos N-metilados con las bibliotecas de inhibidores de interacciones proteína - proteína y macrocíclicos naturales. Otra aportación relevante

es la observación de un traslapamiento entre el espacio químico de los péptidos sintéticos enumerados con los péptidos aprobados como fármacos (o en estudios clínicos) asi como con otros fármacos aprobados que se encuentran fuera del espacio químico tradicional de fármacos de moléculas pequeñas. Los resultados también respaldan que los pentapéptidos sintéticos son compuestos adecuados para usar en proyectos de descubrimiento de fármacos.

Como perspectivas se propone la implementación de un servidor en línea que sea capaz de enumerar bibliotecas peptídicas a partir de los aminoácidos naturales y muestre su representación en el espacio químico respecto a otras bibliotecas.

Índice general

1. Antecedentes **1**

2. Objetivos **4**
2.1. General 4
2.2. Particulares 4

3. Metodología **5**
3.1. Construcción de las bibliotecas moleculares 5
3.1.1. Péptidos 5
3.1.2. Bibliotecas de referencia 7
3.2. Propiedades fisicoquímicas 8
3.3. Espacio químico 8
3.4. Diversidad basada en huellas digitales (Fingerprint) 9

4. Resultados y discusión **10**
4.1. Propiedades fisicoquímicas 10
4.2. Espacio químico 12
4.2.1. Análisis de componentes principales (PCA) 12
4.2.2. Mapas de auto-organización (SOM) 19
4.3. Análisis de diversidad 21

5. Conclusiones **23**

6. Perspectivas **24**
6.1. D-Peptide Builder: A Web-based application to enumerate the chemical space of peptides 24

6.1.1. Funciones principales del servidor 25
6.1.2. Perspectivas del servidor 27
6.2. Explorar bibliotecas de diferente número de aminoácidos . . 27
6.3. Búsqueda de estructuras de interés farmaceútico 28

Bibliografía **29**

Capítulo 1

Antecedentes

En los últimos años se ha observado que la investigación alrededor de los péptidos ha aumentado en el área de diseño de fármacos [1] lo que se traduce en gran número de moléculas peptídicas posicionadas como candidatos en ensayos clínicos[2].
La creciente investigación en torno a péptidos puede atribuirse a que los péptidos representan una clase de compuestos de interés farmaceútico molecularmente posicionada entre moléculas pequeñas y proteínas. Además fungen un papel importante en vías de señalización intrínsecas para muchas funciones fisiológicas, por lo que representan una opción terapéutica que imita las vías naturales de señalización [3].
El éxito de los péptidos aprobados para su empleo terapeútico o en ensayos clínicos es resultado de su baja toxicidad, asi como su alta afinidad de unión, selectividad a dianas. Las características mencionadas hacen de los péptidos moléculas atractivas para el desarrollo de fármacos [4].
Sin embargo también poseen desventajas como baja biodisponibilidad, inestabilidad metabólica, y tiempos de vida media muy cortos [5],[6].
Actualmente observamos un amplio número de péptidos aprobados cómo fármacos, las indicaciones terapeúticas incluyen antibióticos, antidiabéticos y antituberulosos[7], además de antidepresivos, inmunomoduladores [8] y fárnacos para el tratamiento de enfermedades gastrointestinales[9].
Las estructuras químicas de los péptidos con actividad terapeútica es muy diversa. De tal forma que los péptidos de origen natural o sintético poseen estructuras con alta diversidad y número de aminoácidos. Sin embargo la

diversidad estructural ha sido poco estudiada.
También poco se sabe sobre el efecto que generan en el espacio químico de las modificaciones estructurales, a pesar de que ha sido reportado que tales modificaciones tienen efectos directos en las propiedades químicas y biológicas de los péptidos[9]. Una modificación importante es la N-metilación la cual genera un impacto significativo en la disponibilidad oral. Entre las modificaciones estructurales más comunes de péptidos destacan también los giros β y la cilación; esta última juega un rol importante en la modificación de las propiedades fisicoquímicas [10]. En particular el diseño de péptidos cíclicos es un estrategia empleada comúnmente en el estudio de péptidos. Algunos ejemplos son hormonas o análogos de hormonas así como otros compuestos que han demostrado actividad antimicrobiana [4]. La ciclación mejora la biodisponibilidad haciendo a los péptidos menos susceptibles a las enzimas proteolíticas como las aminopeptidasas y carboxipeptidasas que escinden aminoácidos desde el nitrógeno o el carbono terminal, respectivamente[11, 12].
A pesar de la importancia de los péptidos y sus modificaciones estructurales en el diseño de fármacos, existen pocos estudios sistemáticos que exploren su diversidad y espacio químico. Recientemente, Santos et al. realizaron un análisis en el que clasifican a los péptidos por ruta de administración y efecto terapeútico. En ese trabajo los autores sugieren que los péptidos pueden ocupar el espacio entre las moléculas pequeñas y los medicamentos biotecnológicos en términos de propiedades fisicoquímicas[2]. Entre tanto, para nuestro conocimiento, una caracterización quimioinformática que incluya un análisis de espacio químico y evaluación de diversidad estructural de péptidos no se ha reportado previamente.
En este estudio, se exploró el efecto de la N-metilación y ciclación de péptidos en su distribución en el espacio químico. Como caso de estudio, el presente análisis se centró en pentapéptidos los cuales son sintéticamente accesibles mediante química combinatoria[13]. Se eligieron pentapéptidos porque tal longitud de aminoácidos se encuentra dentro del rango de los péptidos empleados como fármacos. Se sabe que el 86 % de péptidoa aprobados por la Administración de Fármacos y Alimentos de EUA (FDA) cuentan con una longitud máxima de 10 aminoácidos[2]. También se sabe que el peso molécular de los péptidos se encuentra en un rango entre 330 y 850 Daltons[14]. Como parte de este proyecto, estructuras pentapentídicas

lineales, cíclicas y sus respectivas estructuras N-metiladas fueron comparadas con bibliotecas de compuestos de interés farmaceútico incluyendo péptidos aprobados por FDA e inhibidores de interacciones proteína − proteína (PPIs).
Esta última biblioteca se eligió puesto que en una publicación previa a este trabajo, se sugirió que los péptidos comparten espacio químico por propiedades con moduladores de PPIs [15]. Dicho trabajo es un producto de la investigación desarrollada en conjunto con el presente trabajo en el cual se hace una revisión bibliográfica del estado actual las PPIs como dianas epigenéticas importantes en el diseño de fármacos novedosos y especialmente se enfatiza en la importancia de los péptidos y sus derivados como potenciales candidatos de moduladores de PPIs[15].

Capítulo 2

Objetivos

2.1. General

Analizar la diversidad estructural y cobertura del espacio químico de bibliotecas de péptidos enumerados.

2.2. Particulares

- Automatizar la construcción *in-silico* de bibliotecas de péptidos lineales, lineales N-metilados, cíclicos y cíclicos N-metilados.
- Identificar si las modificaciones estructurales como la ciclación y la metilación modifican las propiedades fisicoquímicas de los péptidos.
- Analizar la distribución de las bibliotecas enumeradas en el espacio químico respecto a fármacos de moléculas pequeñas, fármacos de péptidos aprobados y otras bibliotecas de interes farmaceútico.

Capítulo 3

Metodología

3.1. Construcción de las bibliotecas moleculares

3.1.1. Péptidos

En total se emplearon 32 aminoácidos para enumerar las biblioteca de péptidos. 20 aminoácidos naturales en conjunto con otros doce no naturales, todos ellos disponibles comercialmente. En este caso de estudio, los aminoácidos fueron seleccionados considerando las moléculas comúnmente usadas en Torrey Pines Institute for Molecular Studies, una organización con amplia experiencia en la síntesis de bibliotecas empleando química combinatoria como una herramienta útil en el diseño de fármacos[13][16]. Dos modificaciones estructurales diferentes fueron consideradas: la ciclación y la N-metilación. Ambas modificaciones estructurales son comunmente empleadas en el diseño de péptidos. [17]. Los péptidos cíclicos fueron seleccionados debido a su habilidad para mejorar la permeabilidad pasiva mediante la inducción de un cambio conformacional gracias a los puentes de hidrógeno intramoleculares [18]. Los péptidos cíclicos N-metilados han mostrado transporte pasivo en ensayos con membranas permeables artificiales. Estas estructuras también han exhibido un aumento en la estabilidad química [19]. Finalmente, las bibliotecas basadas en péptidos fueron enumeradas para su análisis: péptidos lineales(LIN), péptidos cíclicos (CYC), péptidos lineales N-metilados(LIN NM) y lineáles cíclicos N-metilados (CYC NM).

Script de automatización

Para lograr la automatización de la generación de las bibliotecas se desarrolló un *script* empleando el lenguaje de programación R [20], cuya implementación permitió la enumeración de 32^5 compuestos para cada una de las bibliotecas. El diseño del *script* se basó en el principio de la síntesis de péptidos en fase sólida, donde la primera posicion permanece fija y posteriormente se adicionan aminoacidos en un particular orden de adición para formar enlaces peptídicos. Por ejemplo, si se desea combinar dos aminoácidos: Alanina(A) y Cisteína(C), el resultado será un total de 2^4 es decir 16 péptidos. En la Figura 3.1 se muestra un esquema que representa la secuencia lógica en la que se combinan los aminoácidos, de forma tal que en la última columna encontramos péptidos de 5 aminoácidos de longitud.

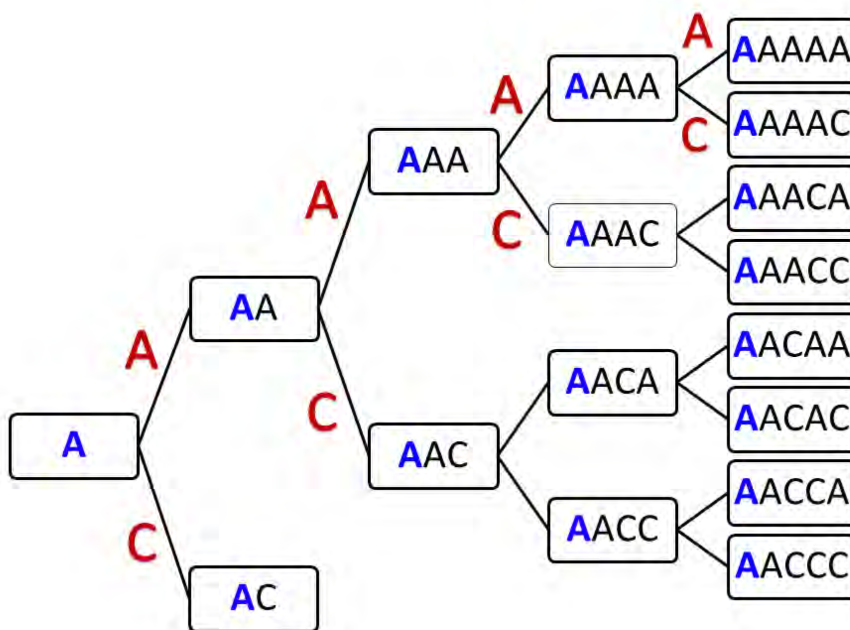


Figura 3.1: Diagrama representativo de la combinación de aminoácidos.

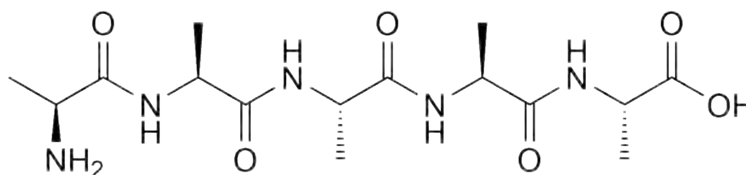


Figura 3.2: Esctructura péptido A-A-A-A-A

Las combinaciones se generan a partir de la representación de SMILES [21] para cada aminoácido. Por ejemplo para la secuencia de aminoácidos A-A-A-A-A se obtiene la secuencia de SMILES siguiente:
N[C@@H](C)C(N[C@@H](C)C(N[C@@H](C)C(N[C@@H](C)C(N[C@@H](C)C(O)=O)=O)=O)=O)=O. Correspondiente a la estructura peptídica en la figura 3.2

3.1.2. Bibliotecas de referencia

Se emplearon como referencia seis colecciones de compuestos (Veáse Cuadro 3.1). Las bibliotecas péptidos aprobados como fármacos por FDA (FDA PEP) y moléculas pequeñas de FDA (FDA) fueron construidas de información obtenida de la base de datos DrugBank [22]. La bilioteca de moduladores de interacciones proteína - proteína (PPI) contiene estructuras no peptídicas de 18 familias de PPIs [23]. También se construyeron tres colecciones de compuestos de productos naturales mismas que se listan a continuación: Base de Datos de Medicina Tradicional China (TCM)[24], colección de Productos Naturales comerciales (NP) [25], y una colección de productos Naturales macrocíclicos (MACRO). Esta última corresponde a una biblioteca comercial disponible para virtual screening que comprende productos naturales derivados de plantas y microorganismos [26].

Cuadro 3.1: Bibliotecas de referencia

Bibliotecas	**Número de elementos**
Péptidos Aprobados por FDA (PEP FDA	108
Moléculas Pequeñas de FDA (FDA)	1, 622
Medicina Tradicional China (TCM)	18, 280
Macrociclos Naturales (MACRO)	1,826
Productos Naturales (NP)	247
Inhibidores de interacciones Proteína - Proteína (PPI)	2,583

3.2. Propiedades fisicoquímicas

Seis propiedades de interés farmaceútico fueron calculadas con Molecular Operating Enviroment (MOE)[27]: donadores de puente de hidrógeno (HBD), aceptores de puente de hidrógeno (HBA), coeficiente de partición octanol/agua (logP), peso molecular, área topológica superficial (TPSA), y número de enlaces rotables (RTB) [28]. Estas propiedades son comúnmente empleadas en análisis quimioinformáticos para comparar bibliotecas de compuestos[29, 30].

3.3. Espacio químico

La visualización del espacio químico se realizó empleando las técnicas de reducción de componentes prinpales (PCA) y mapas de autoorganización (SOM). La representación se realizó a partir de las seis propiedades fisicoquímicas calculadas en el punto 3.2. El análisis de los SOMs fué realizado usando el paquete kohonen del lenguaje de programación R[20].

3.4. Diversidad basada en huellas digitales (Fingerprint)

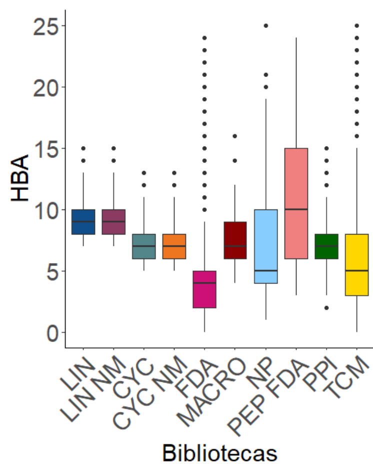
Se calcularon cinco huellas digitales diferentes: Extended Connectivity de radio 2 (ECFP4) y radio 6 (ECFP6), Molecular Access System (MACCS) Keys(166-bits), PubChem y AtomPair. Con el propósito de cuantificar la diversidad estructural de cada biblioteca, la similitud estructural fué calculada para cada par de compuestos empleando el coeficiente de Tanimoto [31, 32] y analizada mediante la función de distribución acumulada, la cual se define como la probabilidad de que la variable tome valores menores o iguales a x. Este análisis de diversidad estructural fué realizado con KNIME[33], y los valores estadísticos se calcularon con un *script* en R [20].

Capítulo 4

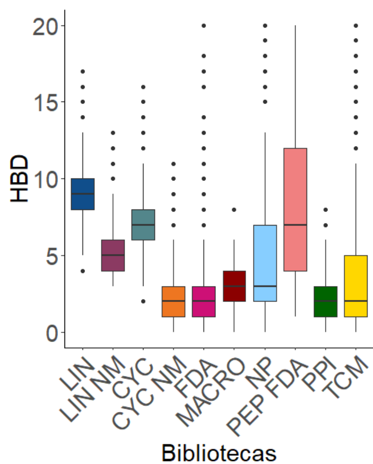
Resultados y discusión

4.1. Propiedades fisicoquímicas

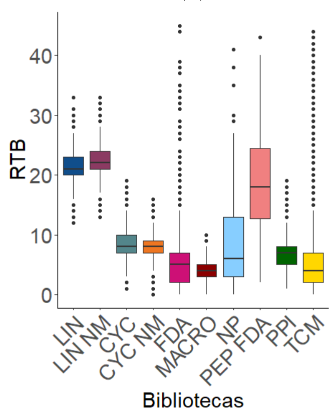
La Figura 4.1 muestra gráficos de cuartiles de la distribución de las seis propiedades fisicoquímicas. En el caso de las bibliotecas de péptidos enumerados, las cíclicas respecto a las lineales difieren significativamente en algunas propiedades y específicamente en número de enlaces rotables. Las bibliotecas metiladas y no metiladas también muestran importantes diferencias en donadores de hidrógeno y peso molecular. Además,las bibliotecas metiladas en constraste con los no metiladas difieren en sus valores de coeficiente de partición octanol/agua y área topológica superficial. La metilación incrementa significativamente los valores de coeficiente de partición octanol/agua y disminuye los valores de área topológica superficial. Ambas propiedades son consideredas importantes en términos de permeabilidad, de forma que las modificaciones de estas propiedades potencialmente significa la mejora en la biodisponilidad de las moléculas.
Respecto a las bibliotecas de referencias, PEP FDA y NP cubren un rango amplio de valores para las seis propiedades calculadas. TCM tiene en general un rango menor de valores para las mismas propiedades, comparada con otras biblioteca de referencia. No obstante, TCM tiene un número amplio de puntos satélite (outliers). Dicho lo anterior podemos decir que las bibliotecas son muy diversas en términos de propiedades. Las 4 bibliotecas peptídicas enumeradas en este trabajo poseen valores estadísticos dentro del intervalo de valores de las colecciones de referencia.



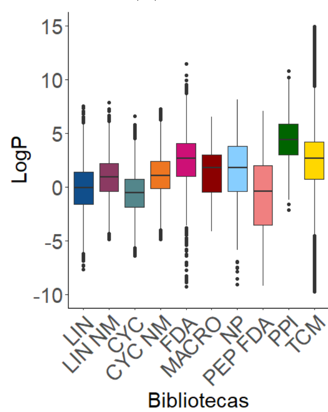
(a)



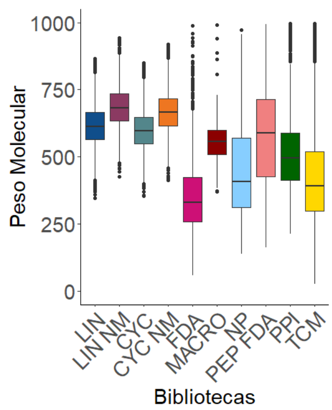
(b)



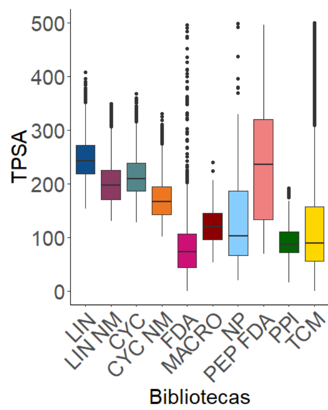
(c)



(d)



(e)



(f)

Figura 4.1: Propiedades Fisicoquímicas

Cuadro 4.1: Eienvectores de los componentes principales

Eigenvectores **Descriptor**	**PC 1**	**PC 2**
HBA	-0.46	-0.09
HBD	-0.51	+0.23
RTB	-0.35	-0.34
SlogP	+0.03	-0.67
TPSA	-0.52	+0.09
MW	-0.33	-0.60

4.2. Espacio químico

4.2.1. Análisis de componentes principales (PCA)

Para una mejor comparación de las propiedades químicas de las cuatro bibliotecas enumeradas respecto a las bibliotecas de referencia, se efectuó la visualización del espacio químico. La Figura 4.3 muestra una visualización en dos dimensiones del espacio químico obtenido mediante la técnica de reducción de dimensiones conocida como componentes principales (PCA). Para ello los valores de los descriptores fueron normalizados. La visualización de los datos se realizó en dos dimensiones, correspondientes al PC 1 y PC 2, el porcentaje de varianza que se recupera de PC1 corresponde a un 51.52 % y PC 2 un 28.28 % lo que significa que ambos componentes permiten representar un 79.8 % de la variación total de los datos. Los valores de los eigenvectores se muestran en el Cuadro 4.1, a partir de estos datos podemos apreciar que los descriptores que contribuyen mayormente al PC 1 son TPSA, HBD, HBA, en contraparte para PC 2 son SlogP, MW y RTB

En la Figura 4.2 se observa la distribución en el espacio químico de las bibliotecas enumeradas. Se aprecia que la distribución de las librerias enumeradas es similar entre las 4 librerias. La Figura 4.3 muestra el espacio químico de las diez bibliotecas: cuatro bibliotecas de péptidos enumerados y seis de interés farmaceútico. Para facilitar el anális, en las Figuras

4.3b - 4.4b se muestran cinco bibliotecas en cada plano y se mantienen las mismas coordenadas que en la Figura 4.3. En cada uno de los cuatro paneles, las bibliotecas de péptidos enumerados se comparan con las siguientes bibliotecas: PEP FDA, FDA, MACRO Y PPIs.
La Figura 4.3a muestra que TCM, NP y PEP FDA exhiben una distribución extensa en el espacio químico mostrandose ampliamente diversos en términos de propiedades fisicoquímicas. El espacio de la biblioteca PEP FDA es muy extenso a pesar de ser una biblioteca relativamente pequeña (108 compuestos, la biblioteca de referencia con menor número de elementos de este trabajo Cuadro 3.1), lo anterior puede atribuirse al número de aminoácidos de los péptidos. Se ha observado que el 86 % de los péptidos aprobados por FDA corresponden a secuencias con una longitud máxima de 10 aminoácidos, mientras que el 14 % restante incluye secuencias entre 11 y 45 aminoácidos [2].
La biblioteca de PEP FDA muestra un traslape con el espacio químico de las cuatro biblioteca de péptidos enumerados 4.3b. Este resultado indica que existe un gran número de compuestos en las biblioteca pentapeptídicas mismas que poseen propiedades similares a los péptidos aprobados para uso clínico o que se encuentran en fases avanzadas de ensayos clínicos. Algunos ejemplos se muestran en la Figura 4.5 como el Almurtide, un péptido con potencial para estimular el sístema inmune y que ha mostrado también actividad antineoplásica [34]. Cilegentide es un péptido cíclico con potencial actividad antineoplásica que se encuenta en ensayos clínicos [35] y Daclastavir, un derivado peptídico con antividad agonísta contra el virus de la hepatitis C[36], su ubicación en el espacio químico se representa en la Figura 4.3b.
En cuanto a las bibliotecas enumeradas, fué posible observar que las modificaciones estructurales como la ciclación y la N-metilación no modifican drásticamente la distribución completa de las bibliotecas en el espacio químico, sin embargo existen algunas diferencias en cuanto a la ubicación de la distribución (Figura 4.2). En partícular la N-metilación de las bibliotecas de péptidos cambian la posición de la distribución respecto a sus análogos no metilados a lo largo de los valores de PC 2, cambio asociado con un incremento en el MW y SlogP. El traslapamiento de los péptidos N-metilados (lineales y cíclicos) con péptidos de FDA (Fig 4.3b) sugiere que algunos de los compuestos que integran las bibliotecas poseen valo-

res de propiedades similares a los péptidos aprobados como fármacos o en ensayos clínicos. La Figura 4.3c indica que la mayoría de las moléculas pequeñas de FDA están localizadas en áreas diferentes del espacio químico respecto a los péptidos. Igualmente, la Figura 4.3c muestra claramente que hay muchos fármacos aprobados que se encuentran fuera del espacio químico tradicional. Lo anterior en relación con comparaciones previas entre el espacio químico de fármacos aprobados respecto al de productos naturales y otras biblioteca [37]. La Figura 4.3c también muestra un traslape entre elementos del espacio químico de las bibliotecas de péptidos enumerados y fármacos aprobados.
Las comparaciones en la Figura 4.4a muestra que los pentapétidos cíclicos enumerados (en particular CYC NM), exhiben un mayor traslape con el espacio químico de la biblioteca MACRO. Conclusiones similares fueron obtenidas comparando el espacio químico de la biblioteca CYC NM con PPI en la Figura 4.4b. Estos resultados sugieren firmemente que las bibliotecas cíclicas N-metiladas pueden emplearse en pruebas experimentales o en ensayos de cribado virtual como una fuente importante de moléculas bioactivas (similar a los actuales usos de cribado virtual de bibliotecas de inhibidores PPIs o MACRO).

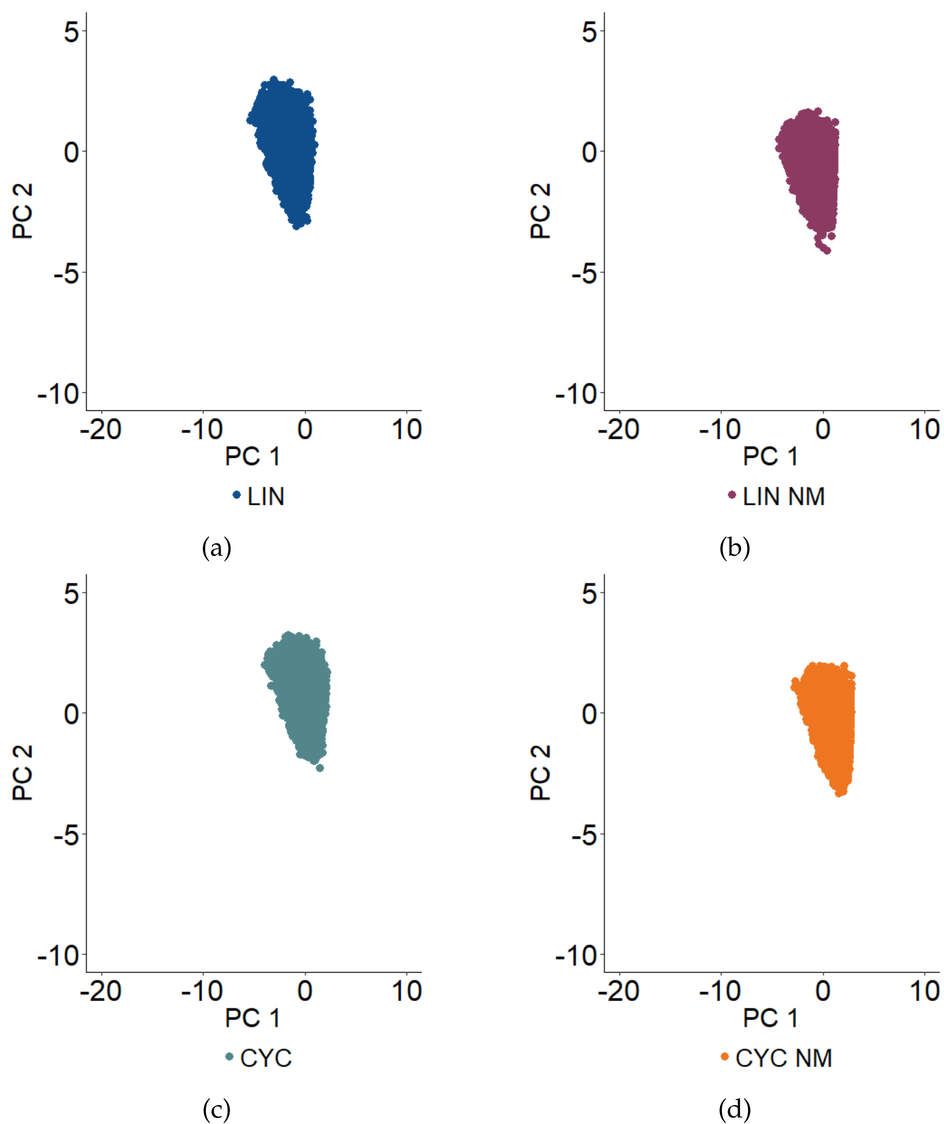
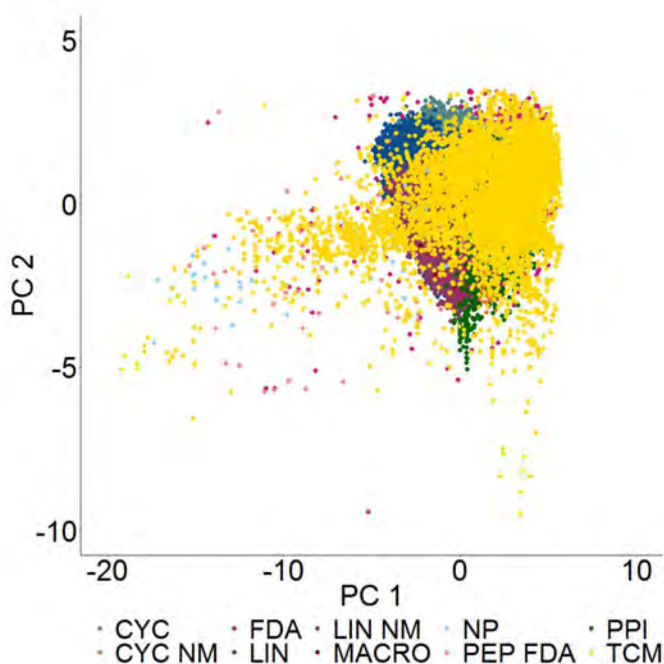
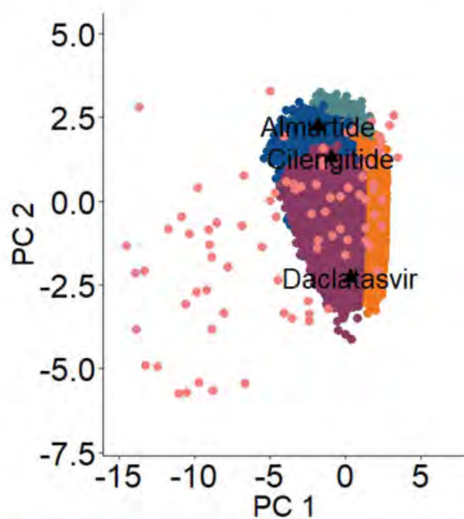


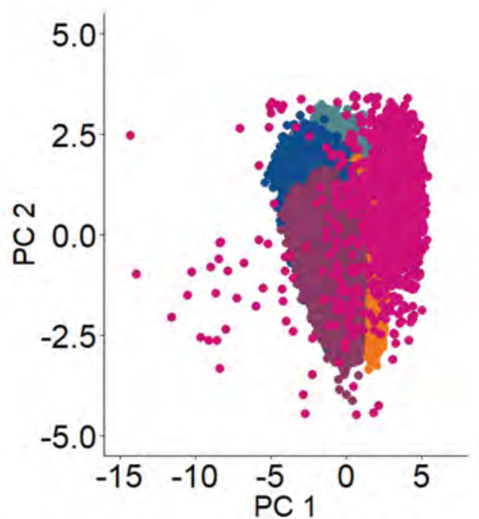
Figura 4.2: Espacio químico de las librerias enumeradas por separado. a) LIN ; b) LIN NM ; c) CYC; d) CYC NM



(a)

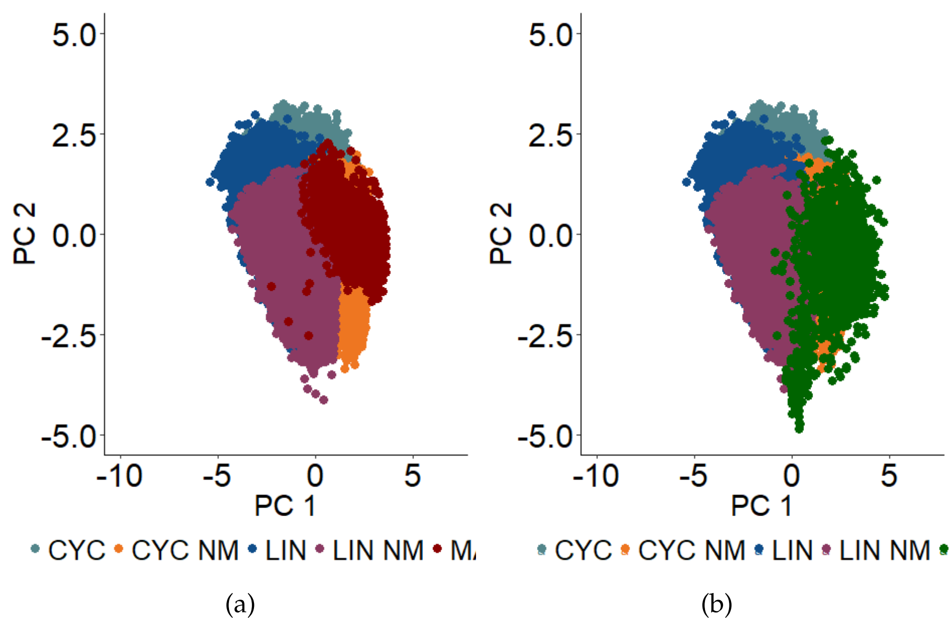


(b)



(c)

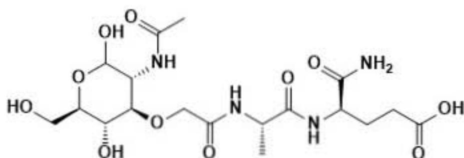
Figura 4.3: Visualización del espacio químico basado en propiedades fisicoquímicas. a) 10 bibliotecas. Bibliotecas peptídicas comparadas con: b) péptidos de FDA; c) fármacos de FDA



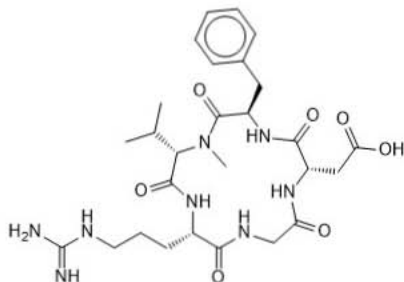
(a) (b)

Figura 4.4: Visualización del espacio químico basado en propiedades fisicoquímicas. a) macrociclos; b) inhibidores de PPIs

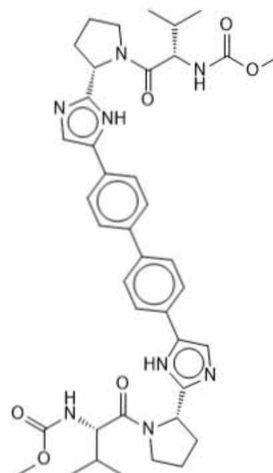
PÉPTIDOS APROBADOS POR FDA



ALMURTIDE

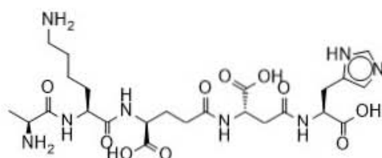


CILENGITIDE

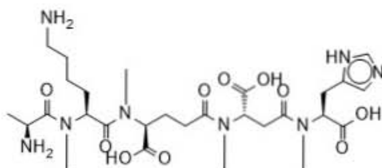


DACLATASVIR

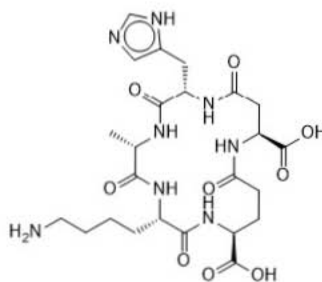
PÉPTIDOS ENUMERADOS



L-ALA L-LYS L-GLU L-ASP L-HIS
LINEAL



L-ALA L-LYS L-GLU L-ASP L-HIS
LINEAL N-METILADO

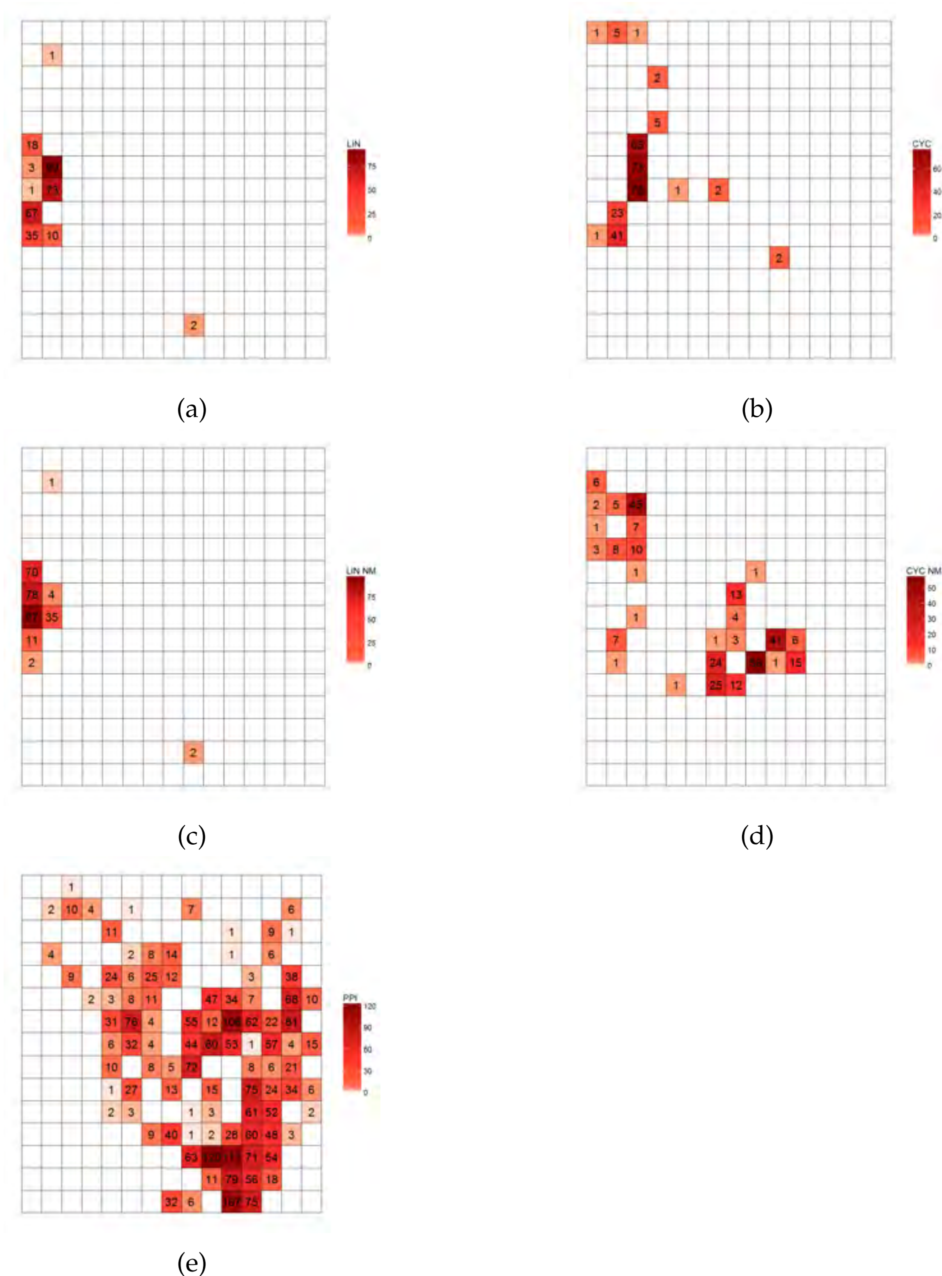


L-ALA L-LYS L-GLU L-ASP L-HIS
CÍCLICO

Figura 4.5: Estructuras químicas representativas de los péptidos discutidos en este trabajo.

4.2.2. Mapas de auto-organización (SOM)

El espacio químico además se visualizó empleando una proyección conocida como mapa de auto-organización (SOM) . Algunas de las bibliotecas se encuentran representadas en la figura 4.6. El número de cada celda en el mapa SOM indica el número de compuestos. La intensidad del color representa el número de compuestos relativo en cada celda: un color más intenso se asocia a un mayor número de compuestos. En las Figuras 4.6b y 4.6d la mayoría de los péptidos enumerados CYC y CYC NM ocupan celdas comparables en el mapa SOM. Dicho resultado indica que las propiedades fisicoquímicas alrededor de ambas bibliotecas son muy similares. Este resultado concuerda con lo observado en PCA, sin embargo gracias a los mapas SOM la comparación es más informativa.
Comparando las bibliotecas lineales y cíclicas fig 4.6a - 4.6d indica que la mayoría de los péptidos lineales ocupan diferentes regiones respecto a sus contrapartes metiladas. Comparando la biblioteca PPI con las de péptidos enumerados (Figura 4.6e) se concluyó que la biblioteca pentapeptídica cíclica N-metilada enumerada en este trabajo tiene un traslape importante con la de PPI. Dicho resultado concuerda también con los resultados de PCA.



(a) (b)

(c) (d)

(e)

Figura 4.6: Representación 2D del espacio químico de cinco bibliotecas empleando mapas de autoorganización(SOM). a)LIN, b)CYC, c)LIN NM, d)CYC NM, e)PPIs

4.3. Análisis de diversidad

La Figura 4.7 muestra la curva de distribución de frecuencia acumulada para los valores de similitud de pares de compuestos empleando el coeficiente de Tanimoto y cinco huellas digitales para cada una de las diez bibliotecas. Varias huellas digitales se emplearon para cuantificar aspectos diferentes de las estructuras moleculares [38]. Basado en las huellas moleculares de diccionarios como MACCS keys y PubChem, las estructuras lineales son estructuralmente menos diversas que las PEP FDA y PPI, pero menos diversas que MACRO 4.7a y 4.7b. Considerando ECFP4, ECFP6 y Atompair Fingerprint las bibliotecas de péptidos enumeradas son las menos diversas. Para las bibliotecas de referencia, los valores de similitud intrabibliotecas indica que FDA es la más diversa considerando los cinco fingerprints, seguida de TCM, NP, y PEP FDA.

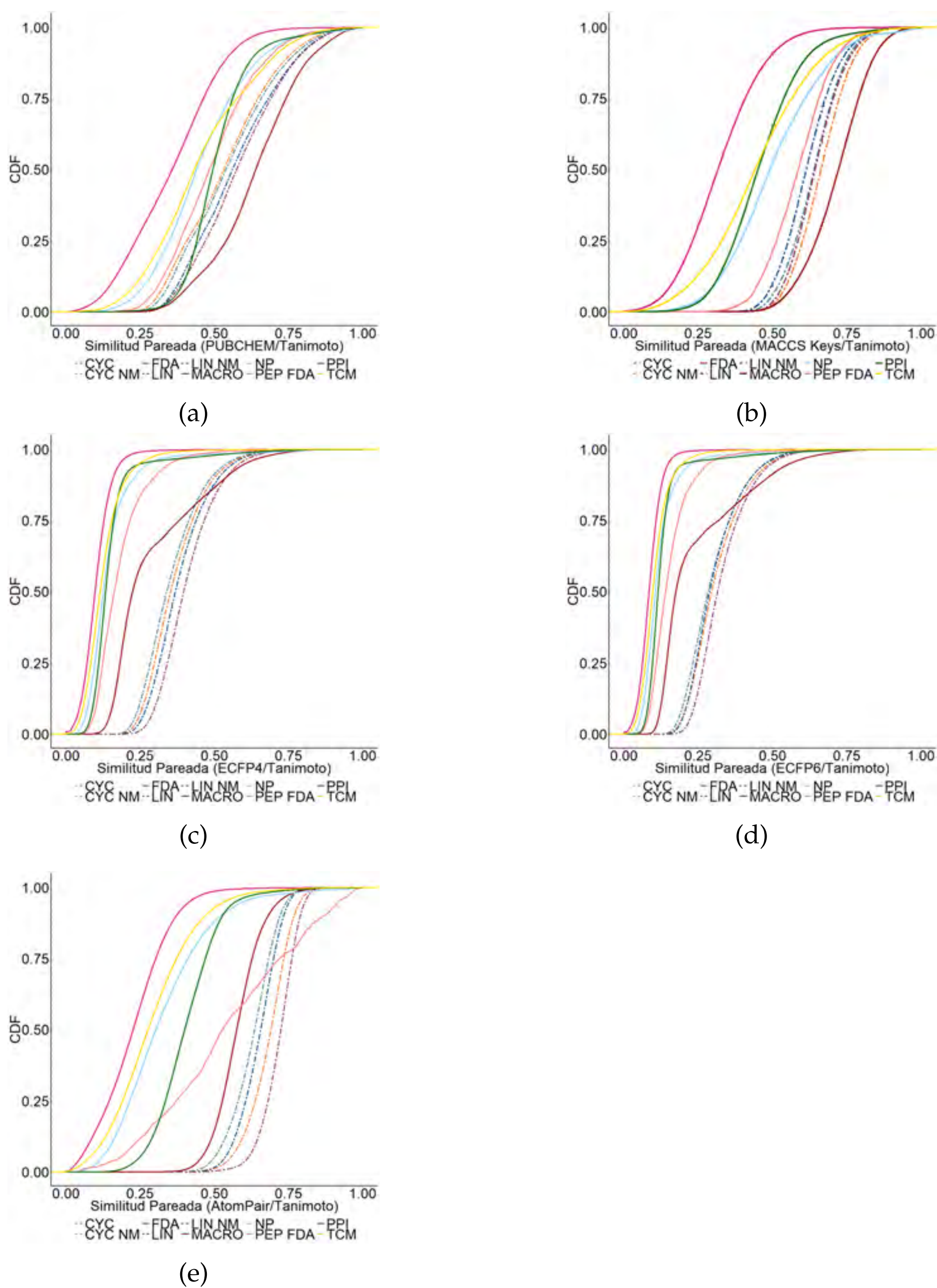


Figura 4.7: Función de distribución acumulada de similitud pareada evaluada con coeficiente de Tanimoto. a) PUBCHEM, b) MACCS keys, c) ECFP4, d) ECFP6, e) AtomPair

Capítulo 5

Conclusiones

En este trabajo se efectuó un análisis exploratorio quimioinformático de 4 bibliotecas de pentapéptidos enumeradas (LIN, LIN NM, CYC y CYC NM). Mediante la enumeración automática a través de un *script* que generó todas las combinaciones de los aminoácidos esenciales y algunos comerciales. El análisis del espacio químico sugirió que la N-metilación y la ciclación modifican la distribución de los péptidos acercandola a la distribución de las moléculas pequeñas de FDA. También estas modificaciones podrían resultar en compuestos con actividad biológica y permeabilidad adecuada ya que los resultados sugieren que las bibliotecas con modificaciones estructurales son más similares por por propiedades a los fármacos de moléculas pequeñas.
Las bibliotecas peptídicas, en partícular las cíclicas N-metiladas, se sobrepone en el espacio químico con los inhibidores de interacciones proteína - proteína, productos naturales macrocíclicos y moléculas FDA aprobadas o en ensayos clínicos de origen peptídico. Por lo tanto, la N-metilación de pentapéptidos representa una prometedora fuente para explorar novedosas y biológicamente relevantes regiones del espacio químico escasamente cubiertas por el espacio químico de bibliotecas medicamente relevantes.[39].

Capítulo 6

Perspectivas

6.1. D-Peptide Builder: A Web-based application to enumerate the chemical space of peptides

Después de automatizar la enumeración de las bibliotecas de péptidos a través de la implementación de un *script* de programación en el lenguaje R, se decidió trabajar en el desarrollo de un servidor en línea que facilite la rápida enumeración de péptidos combinatorios. Este servidor enumera bibliotecas lineales, lineales N-metiladas, cíclicas y cíclinas N-Metiladas, mismas que se analizaron en este proyecto. Nuestro principal objetivo es desarrollar una herramienta útil para grupos de síntesis orgánica que estudien péptidos naturales o sintéticos. La lógica del funcionamiento del servidor se describe en la Figura 6.1. En este caso el desarrollo de los algoritmos involucrados se realiza empleando el lenguaje de programación Python y algunas librerias como rdkit package.

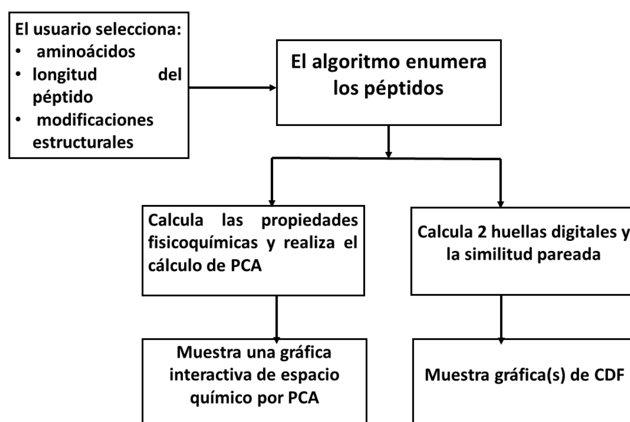


Figura 6.1: Diágrama de funcionamiento del servidor.

6.1.1. Funciones principales del servidor

Enumeración de péptidos y cálculo de propiedades fisicoquímicas

La interfaz de usuario permite la seleccion de diferentes longitudes de aminoácidos para generar desde dipeptidos hasta hexapeptidos, se pretende mejorar esta opción para poder generar decapeptidos. Es posible elegir las librerias a enumerar, las opciones incluyen librerias lineales o con modificaciones estructurales como la ciclación o la N-metilación. Una perpectiva importante es aumentar el número de modificaciones estructurales elegibles. Una muestra de la ventana de selección de especificaciones se muestra en la figura 6.2. Al final de esta sección el usuario puede descargar un archivo con extensión .csv que contenga los SMILES de las bibliotecas enumeradas y los valores de HBA, HVD, RB, SLogP, TPSA y MW.

Espacio químico

El servidor hace una reducción de componentes principales (PCA) para generar un gráfico en dos dimensiones que muestre PC1 vs PC2 y ofrecer al usuario un gráfico interactivo como el que se muestra en la Figura 6.3 para comparar las bibliotecas enumeradas con otras colecciones de

compuestos.

Similitud

Tal como se muestra en la Figura 6.1 el servidor calcula dos huellas digitales (MACCS Keys y ECFP4) para posteriormente calcular la similitud pareada empleando el coeficiente de Tanimoto. Los resultados se muestran al usuario en una gráfica de función de distribución acumulada.

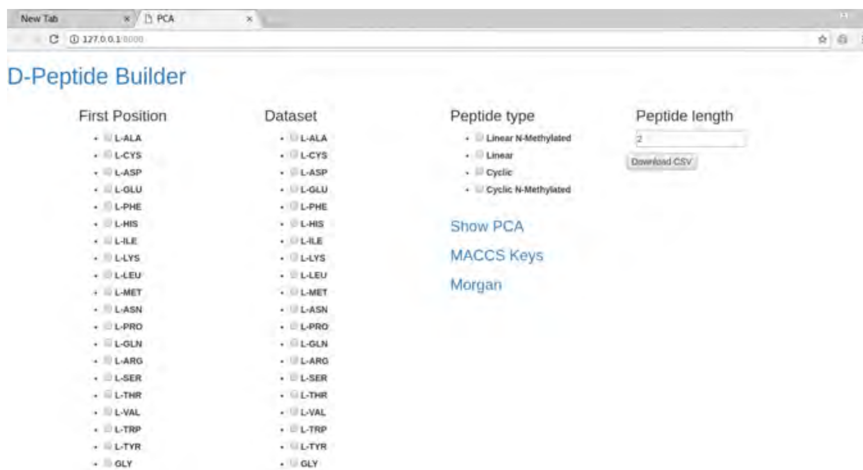


Figura 6.2: Visualización selección de especificaciones

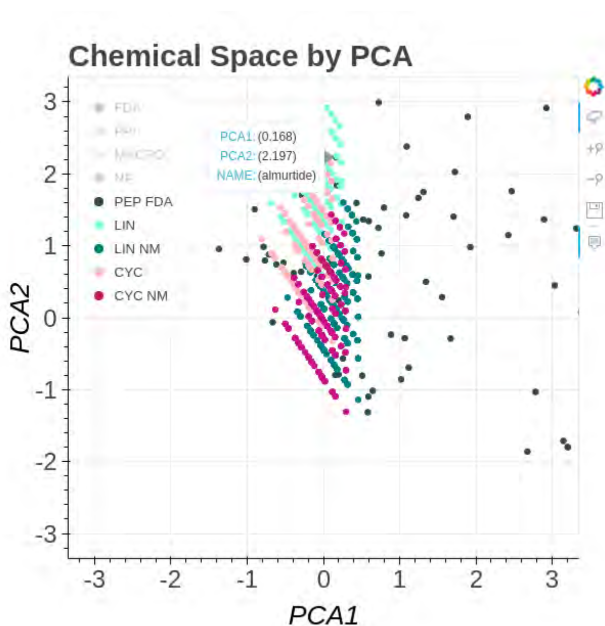


Figura 6.3: Espacio químico generado con D-Peptide Builder

6.1.2. Perspectivas del servidor

Las características descritas anteriormente están desarrolladas en su totalidad. Sin embargo el desarrollo del proyecto aún continúa puesto que es necesario hacer algunas implementaciones para asegurar la funcionalidad del servidor antes de su liberación e implementación en la sección D-TOOLS (DIFACQUIM Tools for Chemoinformatics).

6.2. Explorar bibliotecas de diferente número de aminoácidos

Otra perspectiva de este trabajo es explorar el espacio químico, asi como la similitud estructural y por propiedades de bibliotecas péptidos cíclicos y lineales (incluyendo metilados) con varias composiciones de ami-

noácidos y péptidos de diferente longitud como hexapéptidos o heptapéptido; Esta actividad se facilitará empleando a D-Peptide Builder.

6.3. Búsqueda de estructuras de interés farmaceútico

La intensión de este servidor es proporcionar bibliotecas peptídicas de diferentes longitudes de aminoácidos y modificaciones estructurales que puedan porteriormente ser empleadas en otros estudios computacionales como el cribado virtual o en la implimentación de modelos de aprendizaje automático para la predicción de estructuras con actividad biólogica. Recientemente Wong et al [40] reportaron la comprensión de las propiedades y estructuras de péptidos con actividad antimibrobiana, mediante la implementación de una maquina de soporte vectorial, un algoritmo de aprendizaje automático; El modelo reportado fué entrenado a partir de descriptores calculados por los autores para péptidos que obtuvieron de bases de datos como CAMP R3, una base de datos que proporciona colecciones de estructuras peptídicas antimicrobianas [40]. Este servidor pretende proporcionar bibliotecas péptidicas enumeradas que contribuyan al estudio computacional de péptidos, como lo hacen otras bases de datos.

Bibliografía

[1] James C Nuss John M Henninot, Antoine Collins. The Current State of Peptide Drug Discovery: Back to the Future? *J Med Chem (in press).*, 61(4):1382–1414, 2018.

[2] Gabriela B Santos, A Ganesan, and Flavio S Emery. Oral Administration of Peptide-Based Drugs : Beyond Lipinski ′ s Rule. *Chem Med Chem*, 11(20):2245—2251, 2016.

[3] Dunn MK Lau JL. Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg Med Chem*, 26(10):2700–2707.

[4] Kaycie Heinis Christian Zorzi, Alessandro Deyle. Cyclic peptide therapeutics: past, present and future. *Curr Opin Chem Biol*, 38:24–29, 2017.

[5] Carol S Lim Benjamin J Bruno, Geoffrey D Miller. Basics and recent advances in peptide and protein drug delivery. *Ther Deliv*, 4(11):1443–1467, 2013.

[6] Torsten Fosgerau, Keld Hoffmann. Peptide therapeutics: current status and future directions. *Drug Discov Today*, 20(1):122–128, 2015.

[7] Essam A Yakıcı Gülfer Var Işıl Kayar Begüm Köksal Fatih AlMatar, Manaf Makky. Antimicrobial peptides as an alternative to antituberculosis drugs. *Pharmacol Res*, 128:288–305, 2018.

[8] Sylvia J. Bogaards Medine Cansoy André M. Deelder Hans C. van Leeuwen Jan W. Drijfhout Jaap T. van Dissel Anne M. van der Does,

Paul J. Hensbergen and Peter H. Nibbering. The Human Lactoferrin-Derived Peptide hLF1-11 Exerts Immunomodulatory Effects by Specific Inhibition of Myeloperoxidase Activity. *J Immunol*, 188(10):5012–5019, 2012.

[9] Sarit Laufer Burkhardt Gilon Chaim Hoffman Amnon Kessler Horst Ovadia, Oded Greenberg. Improvement of drug-like properties of peptides: the somatostatin paradigm. *Expert Opin Drug Discov*, 5(7):655–671, 2010.

[10] Florian Weinmüller Michael Kessler Horst Räder, Andreas F B Reichart. Improving oral bioavailability of cyclic peptides by N-methylation. *Bioorg Med Chem*, 0(1):01 – 02, 2017.

[11] Daniel S. Nielsen Huy N. Hoang Rink-Jan Lohman Timothy A. Hill Andrew J. Lucke David J. Craik David J. Edmonds David A. Griffith Charles J. Rotter Roger B. Ruggeri David A. Price Spiros Liras David P. Fairlie. Improving on Nature: Making a Cyclic Heptapeptide Orally Bioavailable. *Angew Chemie Int Ed*, 53(45):12059–63, 2014.

[12] Tessa Nall Praveen K. Madala, Joel D. A. Tyndall† and David P. Fairlie. Update 1 of: Proteases Universally Recognize Beta Strands In Their Active Sites. *Chem Rev*, 110(6):PR1–PR31, 2010.

[13] Giulianotti MA Appel JR Dooley CT Nefzi A Ostresh JM Yu Y Maggiora GM Medina-Franco JL Brunner D Schneider J Houghten RA, Pinilla C. Strategies for the use of mixture-based synthetic combinatorial libraries: scaffold ranking, direct testing, in vivo, and enhanced deconvolution by computational methods. *J Comb Chem*, 10:3–19, 2008.

[14] Kann MG Gonzalez MW. Chapter 4: protein interactions and disease. *PLoS Comp Biol*, 8, 2012.

[15] Medina-Franco JL Díaz-Eufracio BI, Naveja JJ. Proteinprotein interaction modulators for epigenetic therapies. *Adv Protein Chem Str*, 110:65–84, 2018.

[16] Santos RG-Morales A Nefzi A Welmaker GS Medina-Franco JL Giulianotti MA Houghten RA-ShawLN Fleeman R, LaVoi TM. Combinatorial libraries as a tool for the discovery of novel, broad-spectrum antibacterial agents targeting the eskape pathogens. . *J Med Chem*, 58:3340—3355, 2015.

[17] Hoffman A-Kessler H Chatterjee J, Gilon C. N-methylation of peptides: a new perspective in medicinal chemistry. *Acc Chem Res*, 41:1331—1342, 2008.

[18] Zhou MV-Kalyanaraman C Lokey RS Jacobson MP Rezai T, Bock JE. Conformational flexibility, internal hydrogen bonding, and passive membrane permeability? Successful in silico prediction of the relative permeabilities of cyclic peptides. *J Am Chem Soc*, 128:14073–14080, 2006.

[19] Albericio F Marcucci E, Tulla-Puche J. Solid-phase synthesis of NMe-IB-01212, a highly n-methylated cyclic peptide. *Org Lett*, 14:612–615, 2012.

[20] Team RC R core team. R: A language and environment for statistical computing. *R foundation for statistical computing, Vienna, Austria.*

[21] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comp Sci*, 28:31–36.

[22] Djoumbou Y Jewison T Guo AC Liu Y Maciejewski A Arndt D Wilson M Neveu V Tang A Gabriel G Ly C Adamjee S Dame ZT Han B Zhou Y Wishart DS Law V, Knox C. Drugbank 4.0: shedding new light on drug metabolism. *Nucl Acids Res*, 12:D1091–D1097, 2014.

[23] Kuenemann MA Villoutreix BO Sperandio O Labbé CM, Laconde G. iPPI-DB: a manually curated and interactive database of small non-peptide inhibitors of protein-protein interactions. *Drug Discov Today*, 18:958–968, 2013.

[24] Chen CY-C. TCM Database@taiwan: the world's largest traditional chinese medicine database for drug screening in silico. *PLoS ONE*, 6:e15939, 2011.

[25] The Natura Product Company A discovery Collections of purified Natural Products– continuously growing. `https://ac-discovery.com/substancelibrary/purified-natural-products/`. Accessed: 2017-11-22.

[26] TThe Natural Product Company A discovery Natural Macrocyclic Compounds. `https://ac-discovery.com/substancelibrary/natural-macrocyclic-compounds/`. Accessed: 2017-11-22.

[27] Molecular Operating Environment (MOE). Version 2015.08, Chemical Computing Group INC., Montreal, Quebec.

[28] Lipinski CA. Lead- and drug-like compounds: the ruleof-five revolution. *Drug Discov Today: Technologies*, 1:337–341.

[29] Mendez-Lucio O Medina-Franco JL Prieto-Martinez FD, Gortari EF. A chemical space odyssey of inhibitors of histone deacetylases and bromodomains. *RSC Adv*, 6:56225–56239, 2016.

[30] Gupta MP Medina-Franco JL Olmedo DA, González-Medina M. Cheminformatic characterization of natural products from panama. *Molecular Divers*, 21:779–789, 2017.

[31] Maggiora GM Medina-Franco JL. Molecular similarity analysis. In: Bajorath J (ed) Chemoinformatics for drug discovery. *Wiley*, pages 343—399.

[32] Héberger K Bajusz D, Rácz A. Why is tanimoto index an appropriate choice for fingerprint based similarity calculations? *J Cheminf*, 7:1—-13, 2015.

[33] Dill F Gabriel TR-Kötter T Meinl T Ohl P Sieb C Thiel K Wiswedel B Berthold MR, Cebron N. KNIME:the Konstanz information miner. In: Preisach C, Burkhardt H, Schmidt-Thieme L(eds) Data analysis, machine learning and applications, studies in classification, data analysis, and knowledge organization. *Springer*, pages 319 – 326, 2008.

[34] Koichi SK O, Yuen-Joyce L. Muramyl dipeptide and its derivatives: peptide adjuvant in immunological disorders and cancer therapy. *Curr Bioactive Compds*, 7:180–197, 2011.

[35] Kim B DeWitt JP Lee JE Han JH Kim SK Oh CW Kim CY Kim Y-H, Lee JK. Combination therapy of cilengitidewithbelotecanagainstexperimentalglioblastoma. *Int J Cancer*, 133:749—756, 2013.

[36] Bachand C Deon DH Goodrich JT James CA Lavoie R Lopez OD Martel A Romine JL Ruediger EH Snyder LB St Laurent DR Yang F Zhu J Wong HS Langley DR Adams SP Cantor GH Chimalakonda A Fura A Johnson BM Knipe JO Parker DD Santone KS Fridell RA Lemm JA O'Boyle DR 2nd Colonno RJ Gao M Meanwell NA Hamann LG Belema M, Nguyen VN. Hepatitis C virus nS5A replication complex inhibitors: the discovery of daclatasvir. *J Med Chem*, 57:2013–2032, 2014.

[37] Houghten RA Medina-Franco JL López-Vallejo F, Giulianotti MA. Expanding the medicinally relevant chemical space with compound libraries. *Drug Discov Today*, 17:718–726, 2012.

[38] Bender A-Marín RM Giulianotti MA Pinilla C Houghten RA Medina-Franco JL, Martínez-Mayorga K. Characterization of activity landscapes using 2D and 3D similarity methods:consensus activity cliffs. *J Chem Inf Model*, 49:477–491, 2009.

[39] Meurice N Medina-Franco JL, Martinez-Mayorga K. Balancing novelty with confined chemical space in modern drug discovery. *Expert Opin Drug Discovery*, 9:151–165, 2014.

[40] Barai RS Jayaraman VK-Idicula-Thomas S. Thomas S, Karnik S. CAMP: a useful resource for research on antimicrobial peptides. *Nucleic Acids Res*, 38(10):D774–D780, 2010.

Molecular Diversity
https://doi.org/10.1007/s11030-018-9812-9

ORIGINAL ARTICLE

Exploring the chemical space of peptides for drug discovery: a focus on linear and cyclic penta-peptides

Bárbara I. Díaz-Eufracio[1] · Oscar Palomino-Hernández[1] · Richard A. Houghten[2] · José L. Medina-Franco[1]

Received: 13 December 2017 / Accepted: 23 January 2018
© Springer International Publishing AG, part of Springer Nature 2018

Abstract
Peptide and peptide-like structures are regaining attention in drug discovery. Previous studies suggest that bioactive peptides have diverse structures and may have physicochemical properties attractive to become hit and lead compounds. However, chemoinformatic studies that characterize such diversity are limited. Herein, we report the physicochemical property profile and chemical space of four synthetic linear and cyclic combinatorial peptide libraries. As a case study, the analysis was focused on penta-peptides. The chemical space of the peptide and N-methylated peptides libraries was compared to compound data sets of pharmaceutical relevance. Results indicated that there is a major overlap in the chemical space of N-methylated cyclic peptides with inhibitors of protein–protein interactions and macrocyclic natural products available for screening. Also, there is an overlap between the chemical space of the synthetic peptides with peptides approved for clinical use (or in clinical trials), and to other approved drugs that are outside the traditional chemical space. Results further support that synthetic penta-peptides are suitable compounds to be used in drug discovery projects.

Keywords Cheminformatics · Chemical space · Combinatorial chemistry · Protein–protein inhibitors · Small molecules · Synthetic peptides

Introduction

The last years have seen a significant renaissance in peptide drug discovery [1]. There are several drug candidates in clinical trials and numerous drugs approved for clinical use that are or come from peptides [2]. Despite the fact that peptide drugs have, in general, several drawbacks such as low bioavailability, metabolic liability, and short half-lives [3,4], there are several peptides approved as drugs for several indications such as antibiotic, antidiabetic, and anti-tuberculosis [5]. Additional uses of peptides are as antidepressants, immunomodulators [6], and for the treatment of diseases of the gastrointestinal tract [8]. The success of peptides approved for clinical use can be attributed to favorable physicochemical properties such as good binding affinity, target selectivity, and low toxicity. These features make peptides attractive for the development of therapeutics [7].

The chemical structure of pharmaceutical active peptides is diverse. Indeed, peptides from natural or synthetic origin have chemical structures with large diversity in size and amino acid content. Structural modifications can further change the chemical and biological properties of peptides [8]. A major modification is N-methylation which has a strong impact on oral bioavailability. Other modifications such as size, Β-turns and cyclization also play important roles to modify physicochemical properties [9]. For instance, cyclic peptides have shown great success as therapeutics. Examples of broadly applied cyclic peptide drugs are hormones or hormone analogues and other compounds that have shown antimicrobial activity [7]. Cyclization improves bioavailability by making peptides more stable to proteolytic enzymes such as aminopeptidases and carboxypeptidases that cleave off amino acids from their N termini and C termini, respectively [10,11].

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11030-018-9812-9) contains supplementary material, which is available to authorized users.

✉ José L. Medina-Franco
jose.medina.franco@gmail.com; medinajl@unam.mx

1 School of Chemistry, Department of Pharmacy, Universidad Nacional Autónoma de México, Avenida Universidad 3000, 04510 Mexico City, Mexico

2 Torrey Pines Institute for Molecular Studies, Port St. Lucie, FL 34987, USA

Published online: 14 February 2018

Despite the high importance of peptides and modified peptides in drug discovery, there are few systematic studies that explore their diversity and chemical space. Just recently, Santos et al. analyzed their route of administration, therapeutic class, and physical properties to analyze the chemical space of peptides. The authors suggested that peptides can fill a gap between small molecules and biologics in terms of physicochemical properties [2]. However, to our knowledge, a chemoinformatic characterization, which includes a chemical space analysis and assessment of the structural diversity of peptides, has not been recently reported.

In this work, we explored the effect of N-methylation and cyclization of peptides in their distribution in chemical space. As a case study, in the current analysis we focused on penta-peptides which are synthetically accessible through combinatorial chemistry [12]. For this analysis, penta-peptides were selected because the size is within the range of peptide-based drugs. Indeed, 86% of peptides approved by the US Food and Drug Administration (FDA) have a maximum length of ten amino acids [2]. In addition, the molecular weight of the peptides considered in this work is reasonable (molecular weight between 330 and 850 Dalton) [13]. As part of this study, linear and cyclic penta-peptides and N-methylated penta-peptides were compared to compound data sets of pharmaceutical relevance, including peptides approved for clinical use and inhibitors of protein–protein interactions (PPIs). Indeed, it has been suggested that peptides share the chemical space of modulators of PPIs [14].

Methods

Compound data sets

Peptides

A total of 32 amino acids were used to enumerate the peptide libraries: 20 naturally occurring amino acids plus 12 unnatural amino acids, all commercially available (Table S1 in the Supporting information). For this case study, the amino acids were selected considering common amino acids used at the Torrey Pines Institute for Molecular Studies, a non-for-profit organization with vast experience on combinatorial peptide libraries for drug discovery [12,15]. Two different chemical modifications were considered: cyclization and N-methylation. These modifications are the most common in peptide synthesis [16]. Cyclic peptides were chosen due to their ability to improve passive permeability by inducing a conformational change via intramolecular hydrogen bonds [17]. Cyclic N-methylated peptides have been shown to be passively permeable in artificial membrane permeability assays and also demonstrated an improvement in stability [18]. Thus, four peptide-based data sets were enumerated for the analysis: linear peptides (LIN), cyclic peptides (CYC), linear *N*-methylated peptides (LIN NM), and cyclic *N*-methylated peptides (CYC NM). A script in the R programming language [19] based on SMILES [20] substrings was implemented to enumerate the 32^5 compounds in each data set.

Table 1 Reference compound data sets used in this analysis

Data set	Size
Peptide FDA approved drugs (PEP FDA)[a]	108
FDA approved drug (FDA)	1,622
Traditional Chinese medicine database (TCM)	18,280
Natural macrocyclic compounds (MACRO)	1,826
Natural products (NP)	247
Protein–protein interactions inhibitors (PPI)	2,583

[a]Contains some compounds in clinical trials

Reference data sets

Six compound collections of pharmaceutical interest were used as reference (Table 1). The Peptide FDA approved drugs (PEP FDA) and FDA approved drug library (FDA) obtained from DrugBank [21]; a protein–protein interactions inhibitors database (iPPI-DB) which contains nonpeptide inhibitors across 18 families of PPIs [22]. Also three compound collections of natural products were used as reference: the Traditional Chinese Medicine Database (TCM) [23], commercial collections of purified Natural Products (NP) [24], and Natural Macrocyclic Compounds (MACRO). The latter is a commercial macrocyclic library available for screening that comprises plant and microbial-derived natural products with ring sizes ranging from 8 to 35 [25].

Physicochemical properties

Six properties of pharmaceutical relevance were calculated with Molecular Operating Environment (MOE) [26]: hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), octanol/water partition coefficient (SLogP), molecular weight (MW), topological polar surface area (TPSA) and number of rotatable bonds (RTB) [27]. These properties are commonly used in chemoinformatic analysis to compare compound databases [28,29].

Chemical space

Visualization of the chemical space was done using principal component analysis (PCA) and self-organizing maps (SOM) based on the six properties computed in the previous point (2.2). The analysis was performed using the *kohonen* package in R [19].

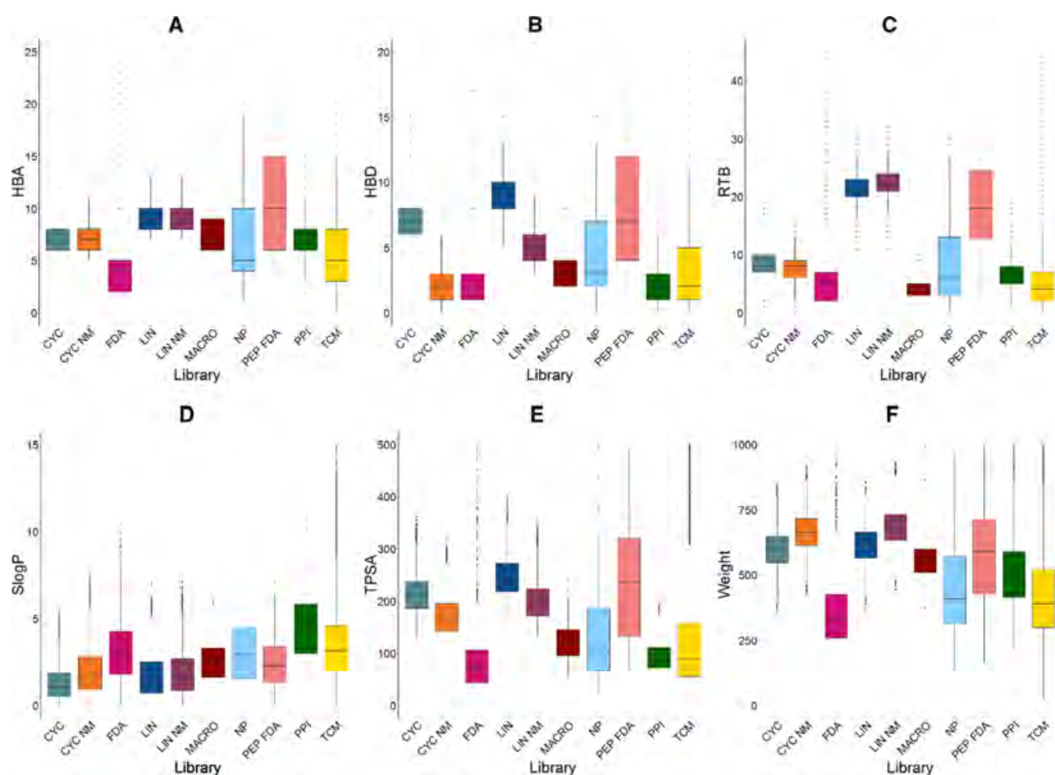


Fig. 1 Box plot and of **a** hydrogen bond acceptors (HBA); **b** hydrogen bond donors (HBD); **c** number of rotatable bonds (RTB). **d** octanol/water partition coefficient (SLogP); **e** topological polar surface area (TPSA); and **f** molecular weight (MW). Summary statistics are presented in the supporting information

Fingerprint-based diversity

Five commonly used fingerprints with different design were used, namely Extended Connectivity radius 2 (ECFP4), radius 3 (ECFP6), Molecular Access System (MACCS) keys (166-bits), PubChem, and AtomPair. To measure the structural diversity of each compound data set, the pairwise structural similarity was calculated using the Tanimoto coefficient [30,31] and analyzed with a cumulative distribution function. The diversity analysis was performed with KNIME [32], and statistical values were obtained with R [19].

Results and discussion

Physicochemical properties

Figure 1 shows boxplots with the distribution of the six properties. Table S2 in the Supporting information presents summary statistics of the distributions. For the peptide data sets, cyclic and linear libraries showed large differences in some properties such as RTB. Methylated and non-methylated sets also had marked differences in HBD and MW. Additionally, methylated versus non-methylated libraries had small but marked differences of SlogP and TPSA: methylation slightly increases SlogP and decreases TPSA values. These properties are important to permeability, so changes in these properties could mean an improvement in bioavailability.

Regarding the reference libraries, PEP FDA and NP cover a broad range of values for most of the six calculated properties. TCM had, overall, a relatively smaller range of values for most properties (as compared to the other reference collections). However, TCM had a large number of outliers. Therefore, the reference libraries can be regarded as very diverse. Of note, the four peptide-based data sets were within the range of values of the reference collections.

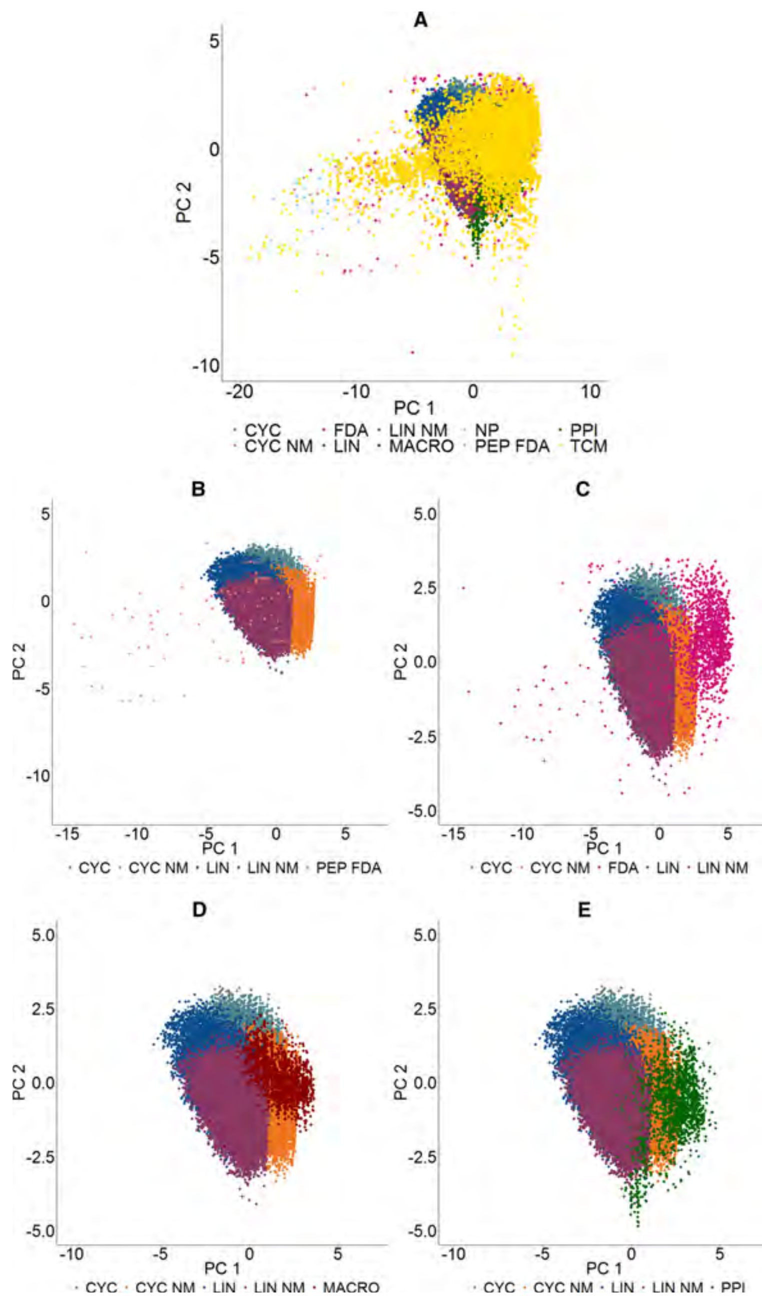
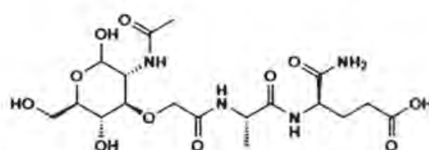
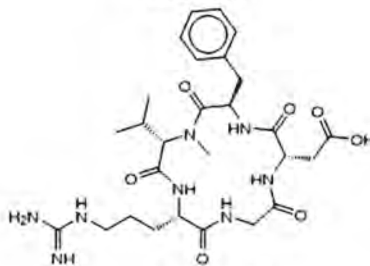
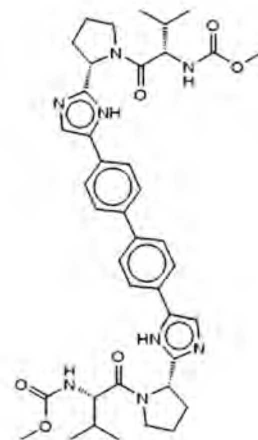


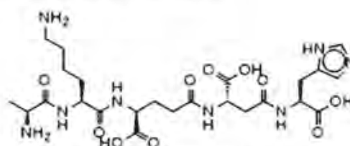
Fig. 2 Visualization of the chemical space based on six properties of pharmaceutical relevance. The first two principal components (PC) recover 80% of the total variance. **a** Comparison of all 10 data sets. Peptide data sets compared to: **b** peptides approved by the FDA and in clinical trials; **c** FDA approved drugs; **d** macrocyclic natural products; **e** inhibitors of protein–protein interactions. The color code is indicated below each panel

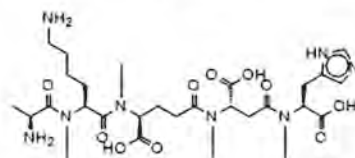
Fig. 3 Chemical structures of representative peptides discussed in this work

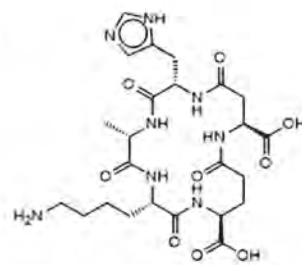












Chemical space

PCA

For further comparison of the six properties of the four peptide-based data sets with the reference compounds, a chemical space representation was analyzed with two approaches. Figure 2 shows a 2D visualization of the chemical space obtained with PCA. The chemical descriptors that contributed most to the first principal component (PC 1; 51.52% of the total variance) were TPSA, HBD and HBA with values of 0.52, 0.51 and 0.46, respectively. The descriptors that contributed most to the second principal component (PC 2; 28.27%) were MW, SlogP and RTB with values of − 0.67, − 0.52 and − 0.47, respectively. Figure 2 shows the chemical space of all 10 data sets: four peptide-based sets plus six reference collections. To aid in the visualization, Fig. 2b–e shows only five data sets in each plot keeping the same coordinates as in Fig. 2a. In each of the four panels, the peptide-based data sets are compared to a different reference set: PEP FDA, FDA, MACRO and PPIs.

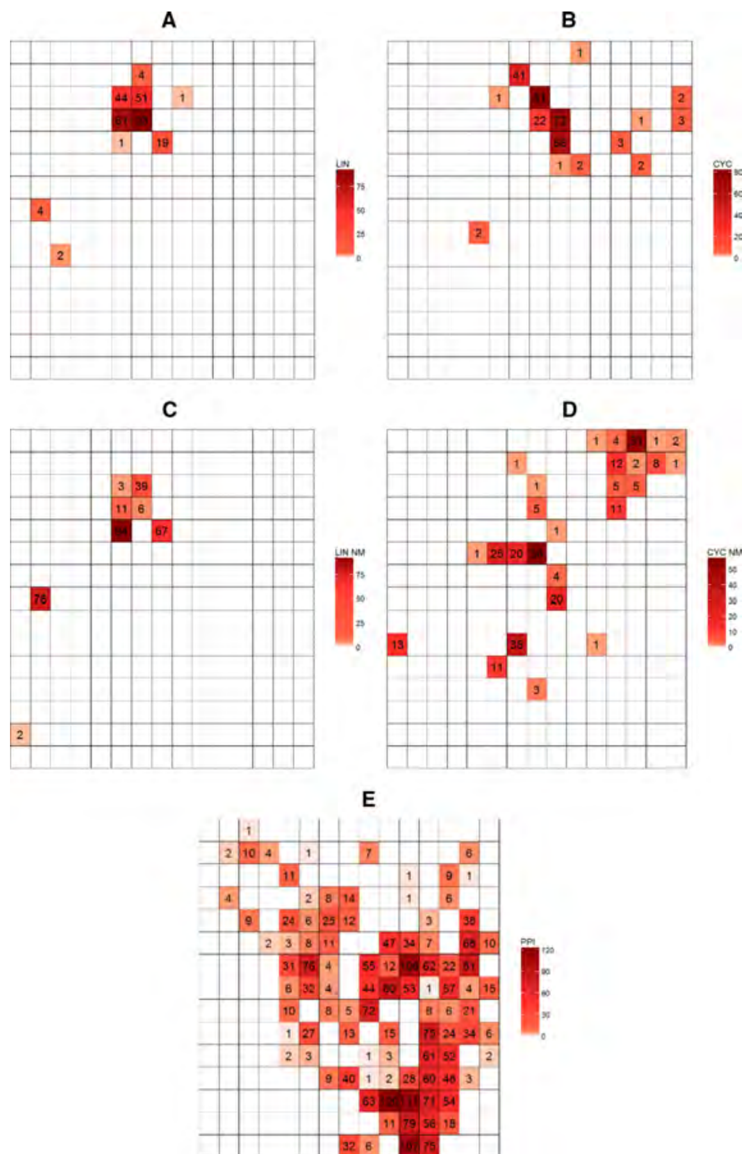


Fig. 4 2D representation of the chemical space of five compound data sets using self-organizing maps (SOMs). The color represents the relative population of a cluster and the number indicates its exact population. **a** LIN; **b** CYC; **c** LIN NM; **d** CYC NM, and **e** PPI. The SOMs of other reference sets are in the Supporting information

Figure 2 shows that TCM, NP and PEP FDA have a large distribution in chemical space thus being rendered as very diverse in terms of their physicochemical properties. The large coverage in chemical space of PEP FDA, despite the relatively small size (108 compounds, the smallest reference set considered in this work, Table 1), can be attributed to the broad range of amino acid sequences and lengths. Indeed, it has been noted that 86% of peptides approved by FDA correspond to sequences with a maximum length of ten amino acids, while the other 14% includes sequences between 10 and 45 amino acids [2].

Interestingly, the PEP FDA set shows a large overlap with the chemical space of the four enumerated peptide-based sets (Fig. 2b). These results indicate that there are several compounds in the penta-peptides libraries which possess similar properties to peptides approved for clinical use or that are

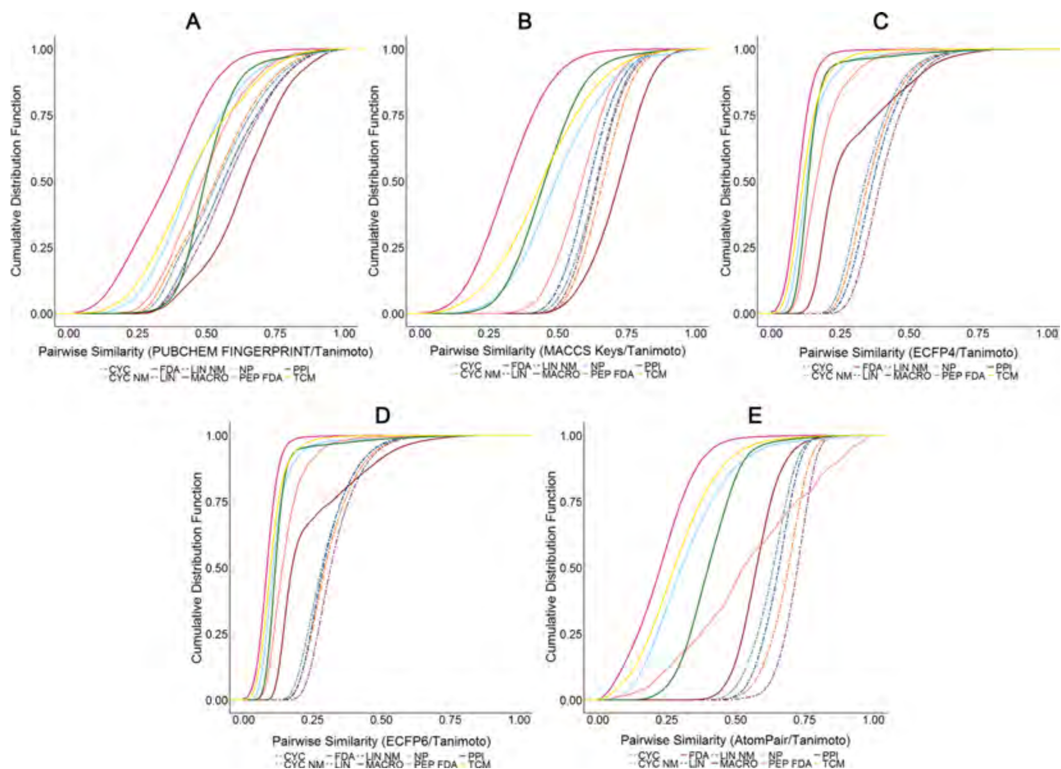


Fig. 5 Cumulative distribution function of the pairwise similarity of the different data sets evaluated with Tanimoto and **a** PUBCHEM fingerprint; **b** MACCS keys; **c** ECFP4; **d** ECFP6; **e** AtomPair fingerprint

in advanced clinical trials. Examples are shown in Fig. 3. Almurtide, a muramyl dipeptide derivative, is a molecule with potential immune stimulating and antineoplastic activity [33]; Cilengitide is a cyclic peptide with potential antineoplastic activity and has been under clinical trials [34], and Daclatasvir, a peptide derivative with activity against the hepatitis C virus [35].

Regarding the enumerated peptide libraries, it was noted that the chemical modifications such as cyclization and methylation did not modify dramatically the overall distribution of the library in chemical space, but there is a noticeable shift. In particular, the methylation of the peptide libraries changed their position with respect to their non-methylated analogues toward lower values of PC 2, which is associated with an increase in MW and SlogP values. The overlap of the N-methylated peptides (linear and cyclic) with PEP FDA (Fig. 2b) suggests that some compounds from these libraries have similar properties to other peptides already used as drug or that are in clinical trials.

Figure 2c indicates that most of the FDA approved drugs are located in a different area of chemical space that peptides. However, Fig. 2c clearly shows that there are several approved drugs that are outside the traditional or most populated chemical space. This is in agreement with previous comparisons between the chemical space of approved drugs with natural products and other compound databases [36]. Notably, Fig. 2c also shows important overlaps between the chemical space of enumerated peptide-based data sets and approved drugs.

The comparison in Fig. 2d shows that the cyclic pentapeptides enumerated in this work (in particular CYC NM) have a major overlap with the chemical space of MACRO. Similar conclusions were obtained comparing the chemical space of CYC NM with PPI in Fig. 2e. These results strongly suggest that synthetic N-methylated cyclic libraries can be used in experimental or virtual screenings as sources of compounds to identify bioactive molecules (similar to the current uses of screening libraries based on inhibitors of PPIs or MACRO).

SOM

The chemical space was also visualized using a SOM projection. Selected data sets are shown in Fig. 4 and all other reference sets are in Figure S1 in the Supporting Information. The number in each cell of the SOM indicates the number of compounds in that cell. The color represents the relative number of compounds in each cell: darker color indicates more compounds. In Fig. 4a and b most of the synthetic cyclic peptides CYC and CYC NM occupy comparable cells in SOM. This result indicates that the physicochemical properties among those two libraries are similar. This result is in agreement with PCA, but in SOM the comparison is visually clearer.

Comparing the linear and cyclic libraries Fig. 4a–d indicates that most of the linear peptides occupy different regions compared to their counterparts. Comparing the PPI with the peptide-based data sets (Fig. 4e), it was concluded that the N-methylated cyclic penta-peptides enumerated in this work had the largest overlap. These results were also in agreement with the results of PCA.

Diversity analysis

Figure 5 shows the CDF of the pairwise similarity values calculated with the Tanimoto coefficient and five fingerprints for the ten data sets. Several fingerprints were used to capture different aspects of the molecular structures [37]. Summary statistics of the distributions of the pairwise similarity values are in Table S3 in the Supporting information. Based on dictionary fingerprints such as MACCS keys and PubChem fingerprints, enumerated peptide libraries (depicted in Fig. 5 with dashed lines) are structurally less diverse than PEP FDA and PPI, but more diverse than MACRO. However, considering ECFP4, ECFP6 and AtomPair fingerprints, enumerated peptide libraries are the least diverse. For the reference libraries, the pairwise intra-library similarity values indicate that FDA is the most diverse set considering all five fingerprints, followed by TCM, NP, and PEP FDA.

Conclusion

In this work, a chemoinformatic exploratory analysis of four synthetic penta-peptide combinatorial data sets (LIN, LIN NM, CYC and CYC NM) was performed by enumerating all combinations of essential and a number of commercially available amino acids to build linear and cyclic penta-peptides. The chemical space analysis suggested that N-methylation and cyclization shifts their position toward the FDA-defined chemical space. Therefore, this modification could result in compounds with biological activity and suitable permeability. Peptide-based data sets, in particular N-methylated cyclic peptides, overlap the chemical space of inhibitors of protein–protein interactions, macrocyclic natural products, and FDA approved drugs or in clinical trials that have a peptide origin. Therefore, N-methylated penta-peptides represent promising sources to explore novel and biologically relevant regions of chemical space scarcely covered by the medicinally relevant chemical space [38].

A perspective of this work is the chemoinformatic analysis of linear and cyclic peptides (including N-methylated) with various compositions of amino acids and peptides of different lengths, e.g., hexa-peptides.

Acknowledgements B.I.D.-E. and O.P. acknowledges *Consejo Nacional de Ciencia y Tecnología* (CONACyT) for Scholarships Number 620289 and 606003, respectively. This work was supported by the *Programa de Apoyo a la Investigación y el Posgrado* (PAIP) Grant 5000-9163, Facultad de Química, UNAM. This work is dedicated to the loving memory of Nicolás Medina Sandoval.

References

1. Henninot A, Collins JC, Nuss JM (2018) The current state of peptide drug discovery: Back to the future? J Med Chem (in press). https://doi.org/10.1021/acs.jmedchem.7b00318
2. Santos GB, Ganesan A, Emery FS (2016) Oral administration of peptide-based drugs: beyond Lipinski's rule. ChemMedChem 11:2245–2251. https://doi.org/10.1002/cmdc.201600288
3. Bruno BJ, Miller GD, Lim CS (2013) Basics and recent advances in peptide and protein drug delivery. Ther Deliv 4:1443–1467. https://doi.org/10.4155/tde.13.104
4. Fosgerau K, Hoffmann T (2015) Peptide therapeutics: current status and future directions. Drug Discov Today 20:122–128. https://doi.org/10.1016/j.drudis.2014.10.003
5. AlMatar M, Makky EA, Yakıcı G, Var I, Kayar B, Köksal F (2017) Antimicrobial peptides as an alternative to anti-tuberculosis drugs. Pharmacol Res. https://doi.org/10.1016/j.phrs.2017.10.011
6. van der Does AM, Hensbergen PJ, Bogaards SJ, Cansoy M, Deelder AM, van Leeuwen HC, Drijfhout JW, van Dissel JT, Nibbering PH (2012) The human lactoferrin-derived peptide hlF1-11 exerts immunomodulatory effects by specific inhibition of myeloperoxidase activity. J Immunol 188:5012. https://doi.org/10.4049/jimmunol.1102777
7. Zorzi A, Deyle K, Heinis C (2017) Cyclic peptide therapeutics: past, present and future. Curr Opin Chem Biol 38:24–29. https://doi.org/10.1016/j.cbpa.2017.02.006
8. Ovadia O, Greenberg S, Laufer B, Gilon C, Hoffman A, Kessler H (2010) Improvement of drug-like properties of peptides: the somatostatin paradigm. Expert Opin Drug Discov 5:655–671. https://doi.org/10.1517/17460441.2010.493935
9. Räder AFB, Reichart F, Weinmüller M, Kessler H (2017) Improving oral bioavailability of cyclic peptides by n-methylation. Bioorg Med Chem. https://doi.org/10.1016/j.bmc.2017.08.031
10. Nielsen DS, Hoang HN, Lohman R-J, Hill TA, Lucke AJ, Craik DJ, Edmonds DJ, Griffith DA, Rotter CJ, Ruggeri RB, Price DA, Liras S, Fairlie DP (2014) Improving on nature: making a cyclic heptapeptide orally bioavailable. Angew Chemie Int Ed 53:12059–12063. https://doi.org/10.1002/anie.201405364
11. Madala PK, Tyndall JDA, Nall T, Fairlie DP (2010) Update 1 of: proteases universally recognize beta strands in their active sites. Chem Rev 110:PR1–PR31. https://doi.org/10.1021/cr900368a

12. Houghten RA, Pinilla C, Giulianotti MA, Appel JR, Dooley CT, Nefzi A, Ostresh JM, Yu Y, Maggiora GM, Medina-Franco JL, Brunner D, Schneider J (2008) Strategies for the use of mixture-based synthetic combinatorial libraries: scaffold ranking, direct testing, in vivo, and enhanced deconvolution by computational methods. J Comb Chem 10:3–19. https://doi.org/10.1021/cc7001205
13. Gonzalez MW, Kann MG (2012) Chapter 4: protein interactions and disease. PLoS Comp Biol 8:e1002819. https://doi.org/10.1371/journal.pcbi.1002819
14. Díaz-Eufracio BI, Naveja JJ, Medina-Franco JL (2018) Protein-protein interaction modulators for epigenetic therapies. Adv Protein Chem Str 110:65–84. https://doi.org/10.1016/bs.apcsb.2017.06.002
15. Fleeman R, LaVoi TM, Santos RG, Morales A, Nefzi A, Welmaker GS, Medina-Franco JL, Giulianotti MA, Houghten RA, Shaw LN (2015) Combinatorial libraries as a tool for the discovery of novel, broad-spectrum antibacterial agents targeting the eskape pathogens. J Med Chem 58:3340–3355. https://doi.org/10.1021/jm501628s
16. Chatterjee J, Gilon C, Hoffman A, Kessler H (2008) N-methylation of peptides: a new perspective in medicinal chemistry. Acc Chem Res 41:1331–1342. https://doi.org/10.1021/ar8000603
17. Rezai T, Bock JE, Zhou MV, Kalyanaraman C, Lokey RS, Jacobson MP (2006) Conformational flexibility, internal hydrogen bonding, and passive membrane permeability? Successful in silico prediction of the relative permeabilities of cyclic peptides. J Am Chem Soc 128:14073–14080. https://doi.org/10.1021/ja063076p
18. Marcucci E, Tulla-Puche J, Albericio F (2012) Solid-phase synthesis of NMe-IB-01212, a highly n-methylated cyclic peptide. Org Lett 14:612–615. https://doi.org/10.1021/ol203231q
19. Team RC R core team (2013). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. http://www.R-project.org/. Accessed December 13, 2017
20. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comp Sci 28:31–36. https://doi.org/10.1021/ci00057a005
21. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, Adamjee S, Dame ZT, Han B, Zhou Y, Wishart DS (2014) Drugbank 4.0: shedding new light on drug metabolism. Nucl Acids Res 42:D1091–D1097. https://doi.org/10.1093/nar/gkt1068
22. Labbé CM, Laconde G, Kuenemann MA, Villoutreix BO, Sperandio O (2013) iPPI-DB: a manually curated and interactive database of small non-peptide inhibitors of protein-protein interactions. Drug Discov Today 18:958–968. https://doi.org/10.1016/j.drudis.2013.05.003
23. Chen CY-C (2011) TCM Database@taiwan: the world's largest traditional chinese medicine database for drug screening in silico. PLoS ONE 6:e15939. https://doi.org/10.1371/journal.pone.0015939
24. The Natura Product Company A discovery Collections of purified Natural Products– continuously growing. https://ac-discovery.com/substancelibrary/purified-natural-products/. Accessed 22 Nov 2017
25. The Natural Product Company A discovery Natural Macrocyclic Compounds. https://ac-discovery.com/substancelibrary/natural-macrocyclic-compounds/. Accessed 22 Nov 2017
26. Molecular Operating Environment (MOE), Version 2015.08, Chemical Computing Group INC., Montreal, Quebec. http://www.chemcomp.com
27. Lipinski CA (2004) Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today: Technologies 1:337–341. https://doi.org/10.1016/j.ddtec.2004.11.007
28. Prieto-Martinez FD, Gortari EF, Mendez-Lucio O, Medina-Franco JL (2016) A chemical space odyssey of inhibitors of histone deacetylases and bromodomains. RSC Adv 6:56225–56239. https://doi.org/10.1039/C6RA07224K
29. Olmedo DA, González-Medina M, Gupta MP, Medina-Franco JL (2017) Cheminformatic characterization of natural products from panama. Molecular Divers 21:779–789. https://doi.org/10.1007/s11030-017-9781-4
30. Medina-Franco JL, Maggiora GM (2014) Molecular similarity analysis. In: Bajorath J (ed) Chemoinformatics for drug discovery. Wiley, New York, pp 343–399. https://doi.org/10.1002/9781118742785.ch15
31. Bajusz D, Rácz A, Héberger K (2015) Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminf 7:1–13. https://doi.org/10.1186/s13321-015-0069-3
32. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Sieb C, Thiel K, Wiswedel B (2008) KNIME: the Konstanz information miner. In: Preisach C, Burkhardt H, Schmidt-Thieme L (eds) Data analysis, machine learning and applications, studies in classification, data analysis, and knowledge organization. Springer, Berlin, pp 319–326. https://doi.org/10.1007/978-3-540-78246-9_38
33. Chikako O, Yuen-Joyce L, Koichi SK (2011) Muramyl dipeptide and its derivatives: peptide adjuvant in immunological disorders and cancer therapy. Curr Bioactive Compds 7:180–197. https://doi.org/10.2174/157340711796817913
34. Kim Y-H, Lee JK, Kim B, DeWitt JP, Lee JE, Han JH, Kim SK, Oh CW, Kim CY (2013) Combination therapy of cilengitide with belotecan against experimental glioblastoma. Int J Cancer 133:749–756. https://doi.org/10.1002/ijc.28058
35. Belema M, Nguyen VN, Bachand C, Deon DH, Goodrich JT, James CA, Lavoie R, Lopez OD, Martel A, Romine JL, Ruediger EH, Snyder LB, St Laurent DR, Yang F, Zhu J, Wong HS, Langley DR, Adams SP, Cantor GH, Chimalakonda A, Fura A, Johnson BM, Knipe JO, Parker DD, Santone KS, Fridell RA, Lemm JA, O'Boyle DR 2nd, Colonno RJ, Gao M, Meanwell NA, Hamann LG (2014) Hepatitis C virus nS5A replication complex inhibitors: the discovery of daclatasvir. J Med Chem 57:2013–2032. https://doi.org/10.1021/jm401836p
36. López-Vallejo F, Giulianotti MA, Houghten RA, Medina-Franco JL (2012) Expanding the medicinally relevant chemical space with compound libraries. Drug Discov Today 17:718–726. https://doi.org/10.1016/j.drudis.2012.04.001
37. Medina-Franco JL, Martínez-Mayorga K, Bender A, Marín RM, Giulianotti MA, Pinilla C, Houghten RA (2009) Characterization of activity landscapes using 2D and 3D similarity methods: consensus activity cliffs. J Chem Inf Model 49:477–491. https://doi.org/10.1021/ci800379q
38. Medina-Franco JL, Martinez-Mayorga K, Meurice N (2014) Balancing novelty with confined chemical space in modern drug discovery. Expert Opin Drug Discovery 9:151–165. https://doi.org/10.1517/17460441.2014.872624

Supporting Information

Exploring the chemical space of peptides for drug discovery: A focus on linear and cyclic penta-peptides

Bárbara I. Díaz-Eufracio, Oscar Palomino-Hernández, Richard A. Houghten,

José L. Medina-Franco*

Contents

		Page
Table S1	Commercial amino acids employed to enumerate the peptide libraries.	S2
Table S2	Summary statistics of the distribution of six physicochemical properties for the 10 data sets.	S3
Figure S1	2D representation of the chemical space of five compound data sets using self-organizing maps. See main text for the SOMs of the remaining data sets.	S4
Table S3	Summary statistics of the distributions of the pairwise similarity values computed with Tanimoto and different fingerprints.	S5

Table S1. Commercial amino acids employed to enumerate the peptide libraries.

L-2-Aminobutyric acid

L-Norvaline

L-Norleucine

L-Ornithine

L-2-Naphthyl alanine

L-CyclohexylAlanine

L-4-Nitro-Phenylalanine

L-4-Fluorophenylalanine

L-(3-Pyridyl)alanine

L-2-Thienylalanine

L-Ethyl-L-Tyrosine

L-Citrulline

L-2,6-Dimethyltyrosine

Table S2. Summary statistics of the distribution of six physicochemical properties for the 10 data sets.

HBA	CYC	CYC NM	FDA	PEP FDA	LIN	LIN NM	MACRO	NP	PPI	TCM
MIN	5.00	5.00	0.00	3.00	7.00	7.00	4.00	1.00	2.00	0.00
1Q	6.00	6.00	2.00	7.50	8.00	8.00	6.00	4.00	6.00	3.00
MEDIAN	7.00	7.00	4.00	12.00	9.00	9.00	7.00	5.00	7.00	5.00
MEAN	7.20	7.20	4.50	15.27	9.19	9.19	7.34	9.85	7.04	6.68
3Q	8.00	8.00	5.00	21.00	10.00	10.00	9.00	12.00	8.00	9.00
MAC	13.00	13.00	37.00	71.00	15.00	15.00	16.00	48.00	15.00	61.00
STD	1.31	1.31	3.62	11.83	1.31	1.31	1.76	9.77	2.17	6.17

HBD	CYC	CYC NM	FDA	PEP FDA	LIN	LIN NM	MACRO	NP	PPI	TCM
MIN	2.00	0.00	0.00	1.00	4.00	3.00	0.00	0.00	0.00	0.00
1Q	6.00	1.00	5.00	5.00	8.00	5.00	2.00	2.00	1.00	1.00
MEDIAN	7.00	2.00	2.00	10.00	9.00	4.00	3.00	4.00	2.00	2.00
MEAN	7.30	2.43	2.26	14.00	9.32	5.45	2.69	5.68	2.21	3.75
3Q	8.00	3.00	3.00	20.00	10.00	6.00	4.00	8.00	3.00	5.00
MAC	15.00	10.00	27.00	76.00	17.00	13.00	8.00	27.00	8.00	43.00
STD	1.70	1.61	3.09	12.80	1.71	1.65	1.33	5.59	1.58	4.05

RTB	CYC	CYC NM	FDA	PEP FDA	LIN	LIN NM	MACRO	NP	PPI	TCM
MIN	0.00	0.00	2.00	12.00	12.00	12.00	0.00	0.00	1.00	0.00
1Q	7.00	7.00	13.00	20.00	21.00	21.00	3.00	3.00	5.00	2.00
MEDIAN	8.00	8.00	19.00	21.00	22.00	22.00	4.00	4.00	7.00	4.00
MEAN	8.23	8.02	28.45	21.50	22.52	22.53	3.85	3.85	7.02	5.52
3Q	10.00	10.00	32.00	23.00	14.00	24.00	5.00	5.00	8.00	7.00
MAX	18.00	16.00	209.00	32.00	34.00	34.00	10.00	10.00	19.00	89.00
STD	2.45	2.21	30.00	2.55	2.57	2.57	1.62	1.62	3.41	5.58

Slogp	CYC	CYC NM	FDA	PEP FDA	LIN	LIN NM	MACRO	NP	PPI	TCM
MIN	-6.48	-4.45	-17.87	-17.87	-7.86	-5.39	-4.13	-9.10	-2.14	-4.10
1Q	-1.85	-0.15	-4.05	-4.05	-1.64	-1.64	-0.48	-0.44	2.97	0.71
MEDIAN	-0.59	1.11	-0.50	-0.50	-0.11	-0.11	1.78	1.76	4.40	2.63
MEAN	0.58	1.11	-1.60	-1.60	-0.11	-0.11	1.31	1.44	4.44	2.51
3Q	0.69	2.37	1.76	1.76	1.40	1.40	2.69	3.81	5.82	4.18
MAX	6.60	7.13	7.05	7.05	7.05	7.05	6.49	8.10	10.78	39.89
STD	1.83	1.83	4.86	4.86	4.86	2.21	2.12	3.39	2.07	3.17

TPSA	CYC	CYC NM	FDA	PEP FDA	LIN	LIN NM	MACRO	NP	PPI	TCM
MIN	119.13	101.55	0.00	69.72	132.14	130.57	53.09	20.03	16.13	0.00
1Q	184.30	142.01	43.45	127.60	218.63	171.03	95.28	95.28	72.20	55.76
MEDIAN	208.82	167.53	7345.00	308.26	244.54	169.55	119.33	119.33	86.71	89.90
MEAN	212.39	159.71	90.09	396.48	247.12	199.15	120.62	120.62	91.71	122.92
3Q	237.85	194.05	108.22	530.68	272.14	224.49	145.32	145.32	110.69	158.67
MAC	363.46	330.34	687.66	2064.83	404.46	357.83	210.04	240.04	191.77	316.54
STD	38.12	37.30	79.14	330.10	38.41	37.88	33.28	33.28	30.78	104.51

WEIGHT	CYC	CYC NM	FDA	PEP FDA	LIN	LIN NM	MACRO	NP	PPI	TCM
MIN	341.37	411.50	59.07	132.20	331.33	429.52	371.44	213.10	213.10	27.03
1Q	546.63	656.78	260.32	560.15	564.68	633.84	509.60	414.24	414.24	301.35
MEDIAN	594.71	664.81	334.29	848.82	614.76	638.83	556.66	297.19	497.19	400.47
MEAN	596.52	666.39	376.71	1070.65	615.73	685.05	554.63	518.81	518.81	462.68
3Q	645.76	715.90	431.29	1312.99	665.75	734.86	598.53	590.30	590.30	552.49
MAC	848.96	914.03	1816.71	5036.65	861.01	938.05	1015.15	#####	1061.28	######
STD	72.12	72.99	206.32	777.39	73.40	73.27	66.43	148.12	148.12	256.82

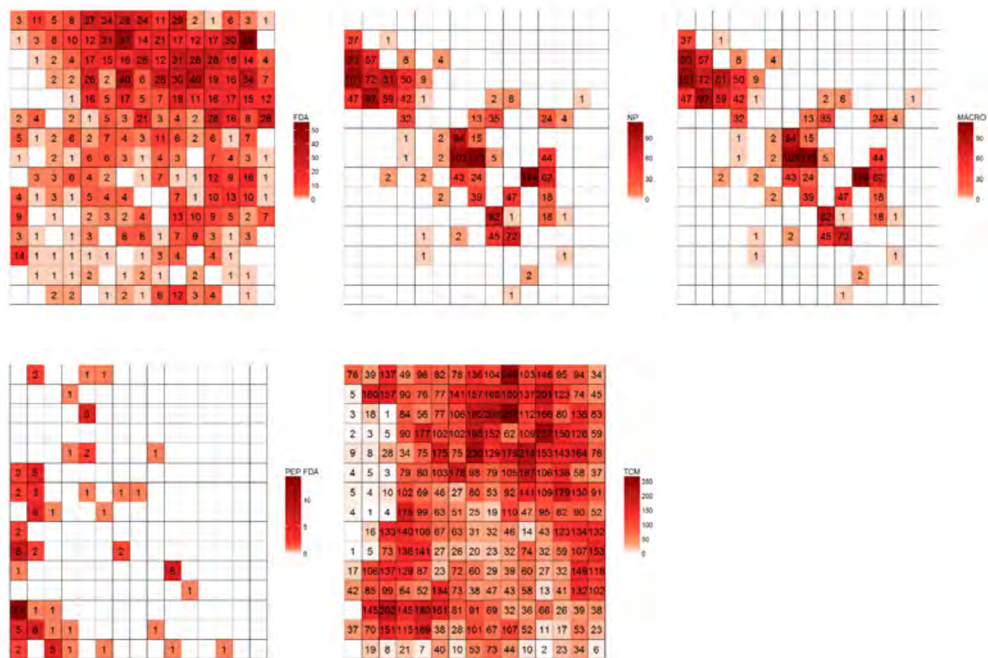


Figure S1. SOM libraries pharmacological interest. 2D representation of the chemical space of five compound data sets using self-organizing maps. See main text for the SOMs of the remaining data sets.

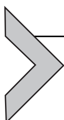
Table S3. Summary statistics of the distributions of the pairwise similarity values computed with Tanimoto and different fingerprints.

	MACCS Keys								ECFP4						
	Min	1st Q	Median	Mean	3rd	Max	SD		Min	1st Q	Median	Mean	3rd	Max	SD
CYC	0.037	0.57	0.64	0.64	0.70	1.00	0.09		0.11	0.25	0.30	0.32	0.38	1.00	0.10
CYC NM	0.39	0.60	0.66	0.66	0.75	1.00	0.08		0.11	0.25	0.31	0.33	0.39	1.00	0.10
LIN	0.35	0.55	0.61	0.61	0.67	1.00	0.09		0.15	0.30	0.36	0.37	0.42	1.00	0.09
LIN NM	0.38	0.58	0.64	0.64	0.70	1.00	0.09		0.10	0.33	0.40	0.39	0.45	0.89	0.09
FDA	0.00	0.23	0.31	0.32	0.40	1.00	0.13		0.00	0.04	0.60	0.70	0.69	0.90	0.10
PEP FDA	0.24	0.51	0.58	0.58	0.70	1.00	0.10		0.02	0.08	0.11	0.14	0.17	1.00	0.09
MACRO	0.36	0.64	0.72	0.71	0.78	1.00	0.10		0.05	0.14	0.18	0.25	0.35	1.00	0.16
NP	0.07	0.39	0.49	0.50	0.60	1.00	0.10		0.00	0.06	0.08	0.10	0.11	0.99	0.09
TCM	0.00	0.32	0.44	0.44	0.56	1.00	0.17		0.00	0.08	0.11	0.11	0.15	1.00	0.06
PPI	0.04	0.37	0.45	0.46	0.53	1.00	0.12		0.00	0.11	0.11	0.15	0.16	1.00	0.09

	Pubchem								ECFP6						
	Min	1st Q	Median	Mean	3rd	Max	SD		Min	1st Q	Median	Mean	3rd	Max	SD
CYC	0.21	0.41	0.53	0.54	0.65	1.00	0.15		0.06	0.19	0.24	0.26	0.31	1.00	0.09
CYC NM	0.20	0.40	0.53	0.53	0.64	1.00	0.15		0.07	0.20	0.25	0.27	0.32	1.00	0.09
LIN	0.23	0.44	0.56	0.57	0.68	1.00	0.15		0.10	0.23	0.27	0.29	0.34	1	0.08
LIN NM	0.23	0.46	0.57	0.57	0.69	1.00	0.14		0.11	0.26	0.30	0.31	0.36	1.00	0.08
FDA	0.00	0.23	0.36	0.35	0.46	1.00	0.15		0	0.03	0.04	0.05	0.06	1.00	0.03
PEP FDA	0.16	0.38	0.48	0.49	0.58	1.00	0.14		0.01	0.05	0.07	0.09	0.11	1.00	0.07
MACRO	0.23	0.53	0.63	0.63	0.74	1.00	0.14		0.03	0.09	0.12	0.19	0.26	1.00	0.15
NP	0.08	0.34	0.43	0.45	0.55	1.00	0.15		0.00	0.04	0.06	0.07	0.08	0.96	0.08
TCM	0.02	0.32	0.43	0.45	0.57	1.00	0.17		0.00	0.05	0.06	0.08	0.08	1.00	0.08
PPI	0.07	0.44	0.50	0.50	0.56	1.00	0.11		0.00	0.05	0.06	0.064	0.08	1.00	0.08

Table S3. (Continued)

	ATOM PAIR						
	Min	1st Q	Median	Mean	3rd	Max	SD
CYC	0.34	0.57	0.63	0.63	0.68	0.95	0.07
CYC NM	0.35	0.63	0.69	0.68	0.73	0.98	0.07
LIN	0.38	0.60	0.65	0.65	0.70	0.93	0.06
LIN NM	0.39	0.68	0.72	0.71	0.76	0.96	0.06
FDA	0.00	0.14	0.22	0.22	0.30	1.00	0.11
PEP FDA	0.043	0.37	0.52	0.54	0.71	1.00	0.22
MACRO	0.25	0.52	0.57	0.57	0.62	1.00	0.07
NP	0.00	0.21	0.30	0.32	0.41	0.99	0.15
TCM	0.00	0.19	0.27	0.28	0.37	0.98	0.13
PPI	0.09	0.33	0.44	0.44	0.46	1.00	0.10



CHAPTER THREE

Protein–Protein Interaction Modulators for Epigenetic Therapies

Bárbara I. Díaz-Eufracio*, J. Jesús Naveja*,†, José L. Medina-Franco*,1

*Facultad de Química, Universidad Nacional Autónoma de México, Mexico City, Mexico
†PECEM, Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City, Mexico
[1]Corresponding author: e-mail addresses: medinajl@unam.mx; jose.medina.franco@gmail.com

Contents

1. Introduction 66
2. PPI Modulators in Epigenetic Drug Discovery 67
   2.1 Bromodomains 67
   2.2 BET 68
   2.3 Chromatin Remodelers 70
   2.4 Histone Acetyltransferases (HATs) 70
   2.5 TRIM-Containing Proteins 71
   2.6 Histone Methyltransferases (HMTs) 71
   2.7 Histone Methylation Readers (HMRs) 72
   2.8 DNA Methyltransferases (DNMTs) 72
3. Peptides in Drug Discovery 73
   3.1 Peptides Approved for Clinical Use 73
   3.2 Peptides as Modulators of PPIs 76
4. Peptides as Modulators of Epigenetic-Related PPIs 77
   4.1 Chemical Space 78
5. Conclusions and Perspectives 80
Acknowledgments 80
References 80

Abstract

Targeting protein–protein interactions (PPIs) is becoming an attractive approach for drug discovery. This is particularly true for difficult or emerging targets, such as epi-targets that may be elusive to drugs that fall into the traditional chemical space. The chemical nature of the PPIs makes attractive the use of peptides or peptidomimetics to selectively modulate such interactions. Despite the fact peptide-based drug discovery has been challenging, the use of peptides as leads compounds for drug discovery is still a valid strategy. This chapter discusses the current status of PPIs in epigenetic drug discovery. A special emphasis is made on peptides and peptide-like compounds as potential drug candidates.

Advances in Protein Chemistry and Structural Biology, Volume 110
ISSN 1876-1623
https://doi.org/10.1016/bs.apcsb.2017.06.002
© 2018 Elsevier Inc.
All rights reserved.

1. INTRODUCTION

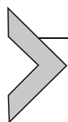
For many years, proteins have been recognized as molecular targets for discovery new drugs. More recently, it has been demonstrated that proteins are capable of regulating themselves through the direct interaction with other proteins. Thus protein–protein interactions (PPIs) have also emerged as promising targets for drug discovery.

A distinctive feature of PPI is the size of the interaction surface. Typical PPI interfaces span the contact area that is considerably larger than the interfaces for protein–small molecules (Smith & Gestwicki, 2012). A major challenge of PPIs interfaces as macromolecular targets for drug discovery is that most PPI interfaces are relatively featureless, lacking preformed, and well-defined hydrophobic cavities which can fully accommodate a small molecule ligand. Such single, deep binding pockets that entirely surround the bound ligands are usually found on traditional "druggable" targets, and on average, they occupy a volume of $\sim$270 A^3 (Buchwald, 2010). Because of the different features of PPIs as drug targets the scientific community for many years considered that PPIs could not be modulated (e.g., inhibitors or stabilizers) by drug-like compounds. Since around 2005, the situation has shifted and remarkable efforts now are being made to rationally design PPI modulators. In fact, it has been recently discussed the many roles that PPIs are playing in many disease conditions (Gonzalez & Kann, 2012).

An example is the potential therapeutic application of modulators of PPIs for epigenetic drug discovery that is discussed thoroughly this chapter. For instance, there are some epigenetic targets well identified as PPI targets as antiapoptotic proteins Bcl-2 and Bcl-XL bind and suppress NALP1, reducing caspase-1 activation, and interleukin-1β production (Bruey et al., 2007). Other example is the inhibition of VEGFR-1–VEGF. Many publications have described the role of VEGFR-1 in cancers such as breast cancer, acute myeloid leukemias, and acute lymphoblastic leukemia know we know that some molecules can prevent VEGFR-1 signaling pathway by the specific inhibition of the formation of the VEGFR-1–VEGF complex (Gautier et al., 2011).

Because new drug targets are needed to improve drug discovery efforts, new projects aiming at identifying compound modulators of PPIs started in several academic and private laboratories (Villoutreix et al., 2014). These efforts have been added to the traditional strategies of modulating PPIs with monoclonal antibodies and other types of proteins and peptides.

The goal of this chapter is to discuss the role of PPIs for epigenetic drug discovery. We emphasized on the potential use of peptides as promising starting points to cover novel regions of epigenetic relevant chemical space (Prieto-Martínez, Gortari, Méndez-Lucio, & Medina-Franco, 2016). The chapter is organized in five major sections. After this introduction, Section 2 discusses the current status of PPI modulators in epigenetic drug discovery. Section 3 is focused on peptide-based drug discovery as strategy to modulate PPIs. Section 4 addresses the potential use of peptides as modulators of PPIs in epigenetic drug discovery. Section 5 presents major conclusions and perspectives.



2. PPI MODULATORS IN EPIGENETIC DRUG DISCOVERY

Through their dynamic posttranslational modifications (PTMs), histones harbor a crucial signaling interplay that regulates many chromatin-related processes, such as gene expression, DNA damage response, and repair (Jenuwein & Allis, 2001). A myriad of PPIs are involved around the process of writing, reading, and erasing these histone marks. Although the inhibition of PPIs might seem more linked to histone PTMs readers, they are not exclusive. Indeed, many other epigenetic processes not directly related to histone PTMs involve PPIs in their functioning, as we will discuss later.

The acknowledgment of the potential of PPI modulators in epigenetics is quite recent, and therefore few but progressive advances have been achieved in this area. There is significant interest in developing epigenetic PPIs, and the results so far are promising. In the following sections, we revise insights in the literature that underscore the advances and potential of developing PPI modulators in epigenetics.

2.1 Bromodomains

Quite recently, the very first cases of PPIs inhibitors in epigenetics started to emerge, with bromodomains (Brds) inhibitors as a proof-of-principle (Mullard, 2012). Brds are protein motifs that bind to acetyl-lysines that can be found, remarkably (although not exclusively), in histones (Brand et al., 2015). There is significant variability in the structure of Brds, thereby allowing the design of compounds that target specifically some of them (Brand et al., 2015).

2.2 BET

The bromodomain and extra C-terminal domain (BET) is a family of proteins containing tandem Brd motifs (most importantly, BRD2, BRD3, BRD4, and BRDT). BET proteins act as epigenetic readers, recruiting other proteins to the chromatin in regions with lysine-acetylated histones (Dawson, Kouzarides, & Huntly, 2012; Zeng & Zhou, 2002). BET proteins interact with many regulators of transcription and cell growth (Garnier, Sharp, & Burns, 2014). The diazepine scaffold in general has been found to be a potent inhibitor of Brds (Smith, Sanchez, & Zhou, 2014). Also, fused diazepines able of inhibiting BET proteins have been synthesized (e.g., **JQ1**, **I-BET762**, and **CPI203**). Chemical structures of selected compounds are shown in Fig. 1 (Garnier et al., 2014).

JQ1 has been shown to inhibit both transcription and recruitment to the chromatin of c-Myc. Thereby it modifies the transcriptional profile of multiple myeloma cells and hurdles their proliferation in vitro and in vivo (Delmore et al., 2011). Similar findings have arisen in hepatocellular and squamous cell carcinoma (where BRD4 or BRD3 is sometimes found fused with NUT) (Filippakopoulos et al., 2010; Li et al., 2016). In parallel, **I-BET762** has been proposed as a potential therapy in acute myeloid leukemia, where some related mutations enhance BRD4 activity (Dawson et al., 2014). This inhibitor mimics histone and therefore interfere with chromatin reading by Brds (Nicodeme et al., 2010). Other Brd inhibitors have been tested in melanoma xenografts models, exhibiting an inhibition pattern that is reversible upon drug discontinuation (Segura et al., 2013). Accordingly, inhibition of BRD4 through iRNA, showed cell growth inhibition in acute myeloid leukemia and melanoma (Segura et al., 2013; Zuber et al., 2011). These pioneering studies highlight the applicability of bromodomains inhibitors in modifying relevant transcription profiles in the maintenance of cancer phenotypes. Interestingly, many mutations in these and other epigenetic readers have been linked to carcinogenesis, which provides a potential source of selective inhibition (Dawson et al., 2012).

BET inhibitors have demonstrated a wide range of applicability in diseases beyond cancer. For instance, it has been found that in heart failure in vitro models, **JQ1** regresses the pathological gene expression profile, and in vivo it suppresses the pathologic cardiac remodeling associated with this disease (Anand et al., 2013). Also, **JQ1** is able to reactivate latent HIV, without cytotoxic effects, therefore positioning this drug as a potential compound in aiming to achieve HIV eradication from the host's

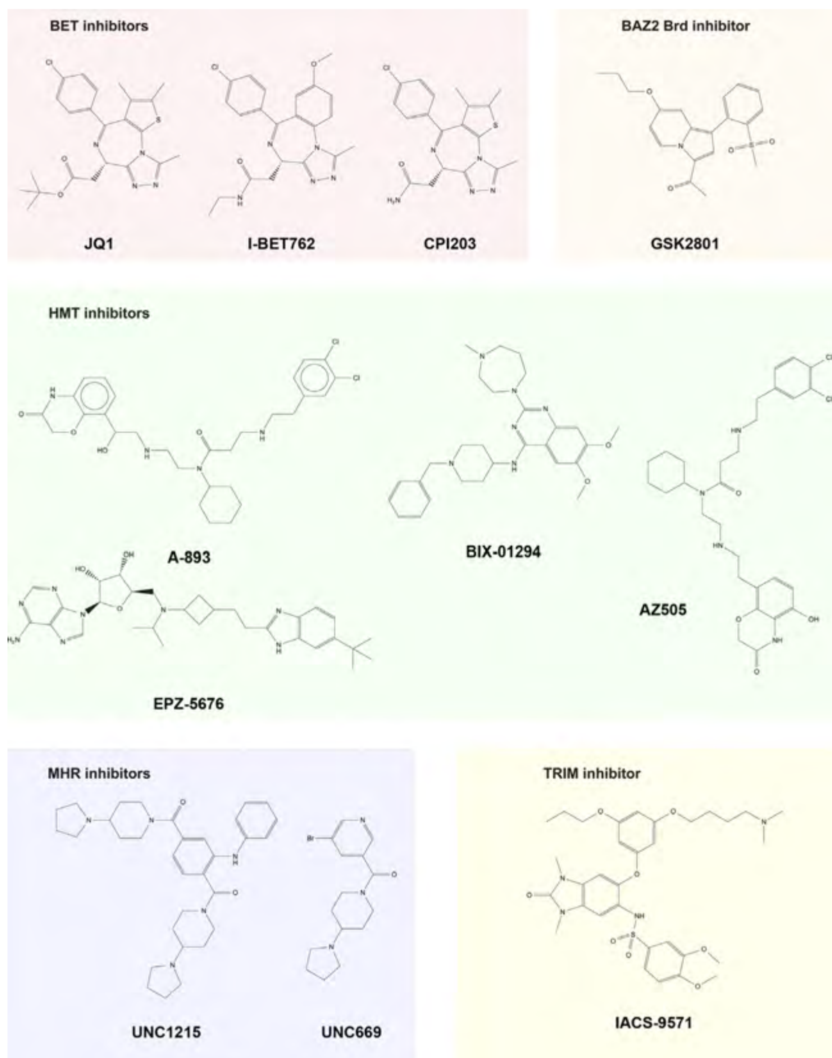


Fig. 1 Chemical structures of small molecules discussed in the chapter.

reservoirs (Banerjee et al., 2012). Moreover, in a periodontitis mouse model, **JQ1** inhibits inflammatory response and osteoclast differentiation (Meng et al., 2014).

Among non-BET bromodomain proteins are chromatin remodelers, as well as histone acetyltransferases (HATs) and tripartite motif (TRIM) proteins, which have been shown to have an important role in gene expression

in carcinogenesis, for which incipient biological assays have been taking place recently (Brand et al., 2015; Gallenkamp, Gelato, Haendler, & Weinmann, 2014). For a detailed review on bromodomain inhibitors, see Brand et al. (2015).

2.3 Chromatin Remodelers

Another quite exciting area under a therapeutic scope is that regarding chromatin remodelers, particularly those that are ATP dependent. This kind of proteins has a significant impact in chromatin structure and function by influencing the disposition of histones and of some histone variants (Mayes, Qiu, Alhazmi, & Landry, 2014). Significant amount of evidence has linked numerous chromatin remodelers to a variety of cancer types (Mayes et al., 2014). As they tend to form protein complexes, inhibiting the formation of these could be an adequate approach. Chromatin remodelers have been categorized in different families (SWI/SNF, INO80, CHD, and ISWI), but we will not discuss each separately (for an in-depth review, see Mayes et al., 2014). Interestingly, the proteins SWI/SNF family have a bromodomain (Mayes et al., 2014). As discussed earlier, the druggability of bromodomains has been proved.

Functional cell assays have been developed to screen compounds for their activity against BRG1 and RUVBL1 (belonging to the SWI/SNF and INO80 families, respectively). Thereby hits have been obtained, which can also inhibit cancer cells growth in vitro (Mayes et al., 2014). Also, peptide-like compounds have been developed de novo to inhibit the PPI between BAF57 (that participates in the SWI/SNF complex) and the androgen receptor in prostate cancer (Link et al., 2008).

BAZ2 proteins are members of the ISWI family of chromatin remodelers. These contain Brds as well.

2.4 Histone Acetyltransferases (HATs)

CREBBP and EP300 were mentioned earlier, as they contain bromodomain motifs. However, their enzymatic activity involves histone acetylation. These proteins interact with many others involved in transcription and cell cycle regulation, as well as with p53 (Brand et al., 2015), and products of oncogenes and fused genes (Giotopoulos et al., 2016). Either functional or pharmacological inhibition of CREBBP and EP300 has an inhibitory effect on the growth of acute myeloid leukemia cells (Giotopoulos et al., 2016). However, targeting the bromodomains present in these proteins can alter

their recruitment to chromatin and disrupt their function. Rooney et al. reported the design of dihydroquinoxalinones that potently displaced acetyl-lysine from CREBBP bromodomains, hindering its normal binding to chromatin (Rooney et al., 2014). Another research group designed 5-isoxazolyl-benzimidazoles derivatives that were selective and potent bromodomain inhibitors for CREBBP and EP300. These results were demonstrated by crystallography, enzymatic, and cellular assays (Hay et al., 2014). Borah et al. focused on the CREBBP interaction with p53 for preventing apoptosis in doxorubicin-treated cardiomyocytes (Borah et al., 2011). This group performed a screening of 115 K compounds against the CREBBP bromodomain, yielding hits that shared the azobenzene scaffold (Borah et al., 2011).

2.5 TRIM-Containing Proteins

TRIM-containing proteins are yet another family of epigenetic readers that have been found to be dysregulated in cancer (Barbieri, Cannizzaro, & Dawson, 2013). These proteins contain different motifs that make them capable of recognizing multiple marks in the same nucleosome; they have been proven relevant in carcinogenesis (Barbieri et al., 2013). **IACS-9571** (Fig. 1) was developed de novo through a structure-based approach guiding the modification of acetyl-lysine mimetics that contained a sulfonamide linked to an aryl group (Palmer et al., 2016). A significant inhibition of other bromodomains (although lesser in potency) was found for this compound, namely against BRPF proteins. This study published the first known TRIM inhibitors (Palmer et al., 2016).

2.6 Histone Methyltransferases (HMTs)

Histone 3 lysine 4- and histone 3 lysine 36 (H3K4 and H3K36, respectively) methylation regulate diverse processes implicated in cancer, including transcriptional regulation, alternative splicing, and DNA repair. SMYD2 is a protooncogene methyltransferase that represses the functional activity of the tumor suppressor protein p53 (Huang et al., 2006).

Cowen et al. designed substrate competitive SMYD2 inhibitors. Authors concluded that perhaps the most interesting feature of the binding mode of the synthetic compound **AZ505** (Fig. 1), is that the dichlorophenethyl moiety of the inhibitor extends across the peptide-binding groove of SMYD2 and is inserted into a secondary hydrophobic pocket adjacent to the cofactor binding site (Cowen et al., 2016).

Among other HMTs inhibitors, **A-893** was found to function as a peptide-competitive inhibitor (Sweis et al., 2015). **BIX-01294** is a very interesting compound acting through occupying the binding pocket of histones in HMTs (Di Costanzo, Del Gaudio, Migliaccio, & Altucci, 2014). In turn, **EPZ-5676** (Fig. 1) is a DOT1L inhibitor undergoing clinical trials for its use in leukemia (Abdel-Wahab & Levine, 2013). Remarkably, DOT1L interacts with fusion proteins including MLL, in a variety of leukemias (Campbell & Tummino, 2014).

EZH2, a HMT that forms part of Polycomb repressive complex 2, regulates cell differentiation and homeostasis. Its hyperactivity has been linked to tumor suppression genes repression and carcinogenesis. Besides direct enzymatic inhibition, its critical PPI with EED (a methyl-lysine reader) have been targeted through peptides, with antineoplastic results (Kim et al., 2013). Similarly to EED, BHC80, an unmethylated lysines reader, has been found to be important in LSD1 function as a HMT (Lan et al., 2007).

2.7 Histone Methylation Readers (HMRs)

These proteins contain modules rich in aromatic amino acid residues that interact with methyl lysine mainly through cation–π and CH–π interactions (Itoh, Suzuki, & Miyata, 2013). Through a structure-based design approach, **UNC669** (Fig. 1) was designed to inhibit L3MBTL1. **UNC669** fills the aromatic cavity and forms the interactions required in recognition of methyl lysine. **UNC1215**, a L3MBTL3 inhibitor has been designed by the same research group (James & Frye, 2013).

2.8 DNA Methyltransferases (DNMTs)

DNMTs are a family of proteins that catalyze the methylation of CpG regions in DNA. DNA methylation plays an outstanding role in the regulation of gene expression. Typically, hypermethylation of gene promoters is associated to gene silencing, while methylation in the gene body has an enhancing effect on the transcript elongation (Baylin & Jones, 2016). The dysregulation of this cellular process has been found to be a hallmark of cancer, where it affects the expression of oncogenes and tumor suppressor genes, and rises genome instability (Baylin & Jones, 2016).

DNMTs have been targeted pharmacologically in leukemia and myelodysplastic syndromes, mainly through the elimination of the catalytic activity of DNMTs by covalent binding of the ligand to the active site (Heerboth et al., 2014). This mechanism of action leads to reactivation of

tumor suppressor genes (Stresemann & Lyko, 2008), and it also may reduce resistance to other pharmacological agents, e.g., to estrogen antagonists in triple negative breast cancer cell lines that have hypermethylation in the promoter of the estrogen receptor gene (Sharma, Saxena, Davidson, & Vertino, 2006).

Nonetheless, as the knowledge on the interaction of the proteins in this family grows, it is becoming apparent that targeting them could be biologically relevant and more selective against cancer cells (Castillo-Aguilera, Depreux, Halby, Arimondo, & Goossens, 2017). Some of the interesting DNMT1 protein interactors are PCNA (involved in DNA replication), DMAP1 (involved in the interaction DNMT1-PCNA), and UHRF1 (which binds to hemimethylated DNA and is involved in DNA damage response), which are crucial for maintenance of DNA methylation (Jeltsch & Jurkowska, 2016; Tien et al., 2011). Although DNMT3L lacks the enzymatic ability of DNA methylation that characterizes its family, it is an epigenetic reader of unmethylated H3K4 that facilitates recruitment of de novo DNMTs (DNMT3A and DNMT3B). The interaction DNMT3L–H3K4 could be inhibited in order to inhibit de novo DNA methylation (Ooi et al., 2007). To date, there are no reported inhibitors against these PPIs.

3. PEPTIDES IN DRUG DISCOVERY

3.1 Peptides Approved for Clinical Use

Proteins and peptides are the building blocks of life and are now evolving as promising candidates of therapeutic drugs (Muheem et al., 2016). Over the last decades, peptides have gained a growing research interest as promising therapeutic for conditions such as autoimmune diseases, diabetes, cancer, mental disorder, hypertension, and certain cardiovascular and metabolic diseases (Ismail & Csóka, 2017). In cancer, for example, several intracellular signaling pathways and extracellular communication networks that are hijacked by the tumors became promising targets for peptide-based therapy (Wu et al., 2014).

Peptides, which are distinguished from proteins based on their smaller size (50 amino acids or less), mediate various essential biological functions, such as signal transduction, heart rate regulation, food intake, and growth. Natural peptides such as insulin, oxytocin, and cyclosporine are successful drugs (Fig. 2). Similar to biologics, peptides can bind large protein targets with high potency and great selectivity, which translates into fewer

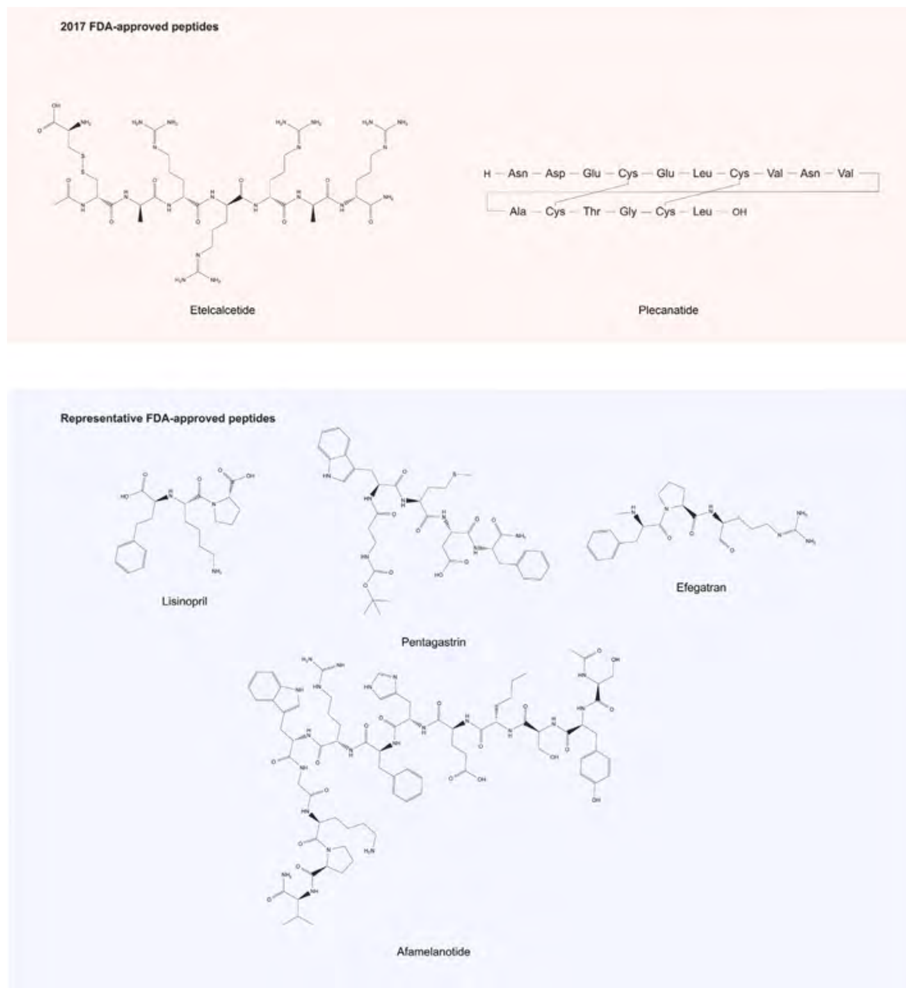


Fig. 2 Chemical structures of representative peptides approved for clinical use.

off-target side effects and less potential for toxicity than small molecule drugs (Craik, Fairlie, Liras, & Price, 2013). Overall, the high therapeutic value of peptides is associated to their ability to provide effective and potent action (Pawar et al., 2014). Additionally, peptides can be highly selective, efficacious, relatively safe, and well tolerated (Ellert-Miklaszewska, Poleszak, & Kaminska, 2017; Fosgerau & Hoffmann, 2015). Short peptides show high binding affinity to the target protein and have several advantages in comparison with other protein-based drugs, such as reduced immunogenicity,

improved safety, ease of synthesis, and modifications (Ellert-Miklaszewska et al., 2017).

Nevertheless, peptide drugs have drawbacks, including low bioavailability, metabolic liability, and short half-lives. For instance, naturally occurring peptides are often not directly suitable for use as convenient therapeutics because they have intrinsic weaknesses, including poor chemical, physical stability, and a short circulating plasma half-life (Bruno, Miller, & Lim, 2013; Fosgerau & Hoffmann, 2015). To address this issue, different reactions have been developed to stabilize the secondary structures of peptides, such as stapling, hydrogen bond surrogates, b-hairpin mimetics, grafting on stable scaffolds, and macrocyclization. Many of these strategies have successfully demonstrated their ability to generate potent, selective, and proteolytically stable inhibitors that modulate previously undruggable or challenging targets. For instance, phage display-based bicyclic peptides can generate structurally constrained peptide libraries with complex loop architectures and enhanced proteolytic stability. In addition, their cell-penetrating capacity has been recently explored (Santos, Ganesan, & Emery, 2016). However, tuning pharmacokinetic properties and achieving cell permeability remains a hurdle.

Besides traditional peptide design a range of peptide technologies have emerged that represent the opportunities and future directions within the peptide field. These include multifunctional and cell penetrating peptides, peptide drug conjugates, and technologies focusing on alternative routes of administration. Emergent and innovative avenues of research, advanced drug discovery tools, and new platform technologies are promising to engineer next-generation peptide-based therapeutics with desired properties (Tsomaia, 2015).

Santos et al. analyzed all FDA-approved drugs from the last 5 years (2012–16) and observed that, from a total of 269 approved drugs, 131 (49%) are small molecules, followed by 29 (11%) monoclonal antibodies (CenterWatch, n.d.). Peptides represent a total of 20 drugs (7%) and 24 are proteins (9%). These numbers are similar to the reported in 2012 (Albericio & Kruger, 2012; Santos et al., 2016). Taking into account that there are fewer academic and industrial research groups devoted to peptide drug synthesis than those focused on small molecule synthetic chemistry, it can be concluded that peptides are currently more successful as therapeutic agents than initially thought (Santos et al., 2016).

In all, there is an increased interest in peptides in pharmaceutical research and development. Currently, there are about 140 peptide therapeutics

currently being evaluated in clinical trials (Fosgerau & Hoffmann, 2015). The number of FDA-approved peptides is constantly increasing. Just in 2017, two peptides have been marketed. One is etelcalcetide (Fig. 2), a linear peptide for the treatment of secondary hyperparathyroidism in patients undergoing hemodialysis (Hamano, Komaba, & Fukagawa, 2017). Plecanatide, a drug for the treatment of chronic idiopathic constipation (Valentin, Acosta, & Camilleri, 2015) has also been approved recently.

3.2 Peptides as Modulators of PPIs

It is well known that specific PPIs are involved in the pathogenesis of numerous diseases. Therefore designing a peptide that mimics a particular protein-binding site and thereby inhibits a specific interaction is a promising therapeutic strategy (Ellert-Miklaszewska et al., 2017). As discussed earlier, PPIs are involved at all levels of cellular organization, thus making the development of PPI inhibitors extremely valuable. The identification of selective inhibitors is challenging because of the shallow and extended nature of PPI interfaces. Inhibitors can be obtained by mimicking peptide-binding epitopes in their bioactive conformation. For this purpose, several strategies have been evolved to enable a projection of side chain functionalities in analogy to peptide secondary structures, thereby yielding molecules that are generally referred to as peptidomimetics (Pelay-Gimeno, Glas, Koch, & Grossmann, 2015).

Pelay-Gimeno et al. introduced a classification of peptidomimetics that modulate PPI based on the degree of their similarity to the natural peptide precursor. Four classes are proposed:

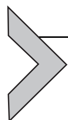
- *Class A*: The backbone and side chains align closely with the bioactive conformation of the precursor peptide.
- *Class B*: Involve peptides with various nonnatural amino acids, isolated small molecule building blocks, and/or major backbone alterations. This class also includes foldamers, which align their side chains topologically similar to the precursor peptide.
- *Class C*: Includes highly modified structures with small molecule character that replace the peptide backbone completely. The central scaffold projects substituents in analogy to the orientation of key residues in the bioactive conformation of the parent peptide.
- *Class D*: Molecules that mimic the mode of action of a bioactive peptide without a direct relationship to its structure.

Peptide-inspired PPI inhibitors have been developed for a broad range of targets. Some proteins, such as G protein-coupled receptors, apoptosis regulators MDM2/MDMX, and BCL-2 family proteins, evolved into model systems that were widely used to test the applicability of peptidomimetics side chain functionalities.

There are various investigations that have demonstrated the PPI modulation by peptides. An example is the design of small peptide that mimics the VEGF helix region. It adopts in pure water a helical conformation and binds to VEGF receptors, these peptide acts as an agonist and induces angiogenesis in vitro; all these conclude the potential applications for this peptide are in the diagnostic field and in therapy of cardiovascular diseases by modification a PPI (D'Andrea et al., 2005).

Inhibition of PPIs is rapidly becoming an attractive approach for the generation of new therapeutics. Progress in the field is aided by the establishment of systematic approaches for target assessment, screening, and structure-based design. While it remains nontrivial to discover new inhibitors for protein complexes, computational tools now provide reliable assessments of hot-spots and pockets found in PPIs, which offer reasonable starting points for inhibitor design. The optimism in the field stands in stark contrast to the rejection of PPIs as largely undruggable less than a decade ago.

It is now recognized that molecules larger than traditional drug compounds will be required for specific disruption of PPIs. While small molecules remain attractive as therapeutics, especially for oral bioavailability, peptides, and proteins are increasingly considered viable candidates. The success in modulating proteins with peptides and peptidomimetics has been reflected in the number of FDA-approved peptide therapeutics. An unprecedented number of peptides have been recently approved or are in clinical trials (Modell, Blosser, & Arora, 2016).



4. PEPTIDES AS MODULATORS OF EPIGENETIC-RELATED PPIs

As commented earlier, peptides are reemerging as drugs due to their high specificity and low toxicity. Peptides are particularly attractive because of their extremely tight binding to their molecular targets (Khazanov & Carlson, 2013). Peptides have also demonstrated the possibility to modulate PPI and this is probably the most important reason to continue with the investigation and designs of peptides as drugs.

Studies about modulation of epigenetic PPIs with peptides are limited. However, there is published evidence that show the activity of peptide-like compounds as chromatin remodelers due to the inhibition of the PPI between BAF57 and the androgen receptor in prostate cancer (see Section 2.3). Other research suggest that the development of DNMTi based on the inhibition/disruption of interaction existing between a DNMT and a DNMT-binding protein appears as a promising approach to design successful therapeutic protocols based on rational molecular targeting.

Cheray et al. develop peptides that inhibit the DNMT3A/ISGF3γ interaction such as a biomarker whose the presence level is associated with a poor survival prognosis and with a poor prognosis of response to the conventional chemotherapeutic treatment of glioblastoma multiforme. Authors demonstrated that the disruption DNMT3A/ISGF3γ interactions increases the efficiency of chemotherapeutic treatment on established tumors in mice and also suggest an innovative alternative to the development of specific DNMT inhibitors (Cheray et al., 2016).

4.1 Chemical Space

As commented earlier, one of the major challenges to target PPIs is the need for new and diverse scaffold libraries to allow for better sampling of the PPI chemical space. In the absence of small natural substrates or ligands, high-throughput screening is often used to identify hit compounds. However, the success rate is rather low. One of the reasons is lack of diversity in the commercial small molecule compound libraries used for screening. The chemical space of existing libraries is strongly influenced by binding pockets of traditional targets, e.g., G protein-coupled receptors, enzymes. Consequently, "drug-like" or medicinally relevant compound libraries (López-Vallejo, Giulianotti, Houghten, & Medina-Franco, 2012) are not diverse enough to contain PPI surface-compatible molecular scaffolds. A number of studies have analyzed known PPI inhibitor molecules and delineated some of the features that differentiate PPI inhibitors from "drug-like" molecules: more hydrophobic, rigid aromatic scaffolds combined with charged or polar groups, larger in size, macrocyclic, higher number of chiral centers, and three-dimensional in bound conformation (Santos et al., 2016; Tsomaia, 2015).

A better understanding of the chemical space of PPI inhibitors will help to create novel but focused chemical libraries that could ensure more successful high-throughput screening in the future (Medina-Franco,

Martinez-Mayorga, & Meurice, 2014; Sperandio, Reynès, Camproux, & Villoutreix, 2010). It may be obvious that PPI inhibitors and peptides have a different distribution in the representation of the chemical space. To illustrate this point, Fig. 3 shows a visual representation of the chemical space of small molecules and peptides approved by the USA Food and Drug Administration (FDA), synthetic peptides (linear, linear *N*-methylated, cyclic, and cyclic *N*-methylated), and PPI inhibitors. The visual representation was generated with principal component analysis of six properties of pharmaceutical relevance, namely: molecular weight, hydrogen bond donors and acceptors, number of rotatable bonds, topological surface area, and log *P* (González-Medina et al., 2017). Of note, the structures of PPIs were retrieved from the inhibitors of Protein–Protein Interaction Database (iPPI-DB) that contains, in total, 1756 nonpeptide inhibitors across 18 families of PPIs (http://www.ippidb.cdithem.fr).

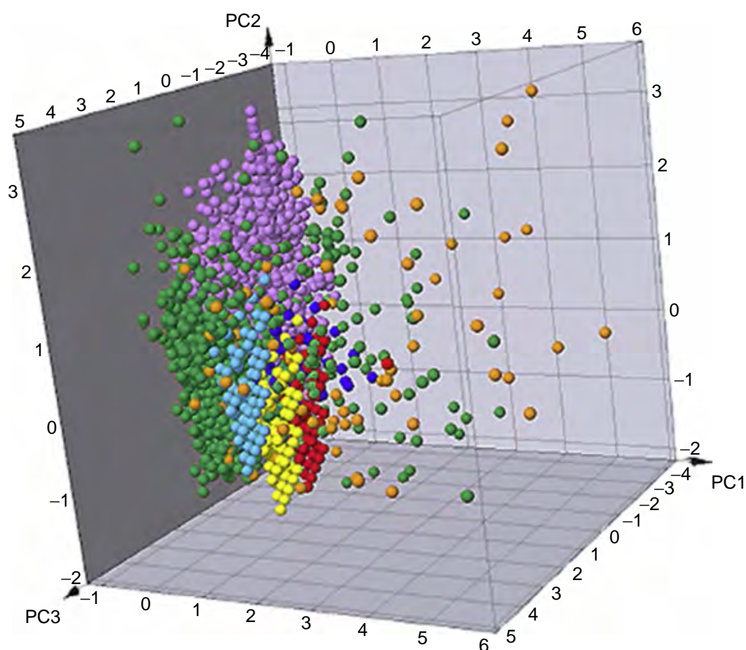


Fig. 3 Visual representation of the chemical space of FDA-approved small molecules (*green*), FDA-approved peptides (*orange*), inhibitors of protein–protein interactions (*pink*), and synthetic penta-peptides as reference: linear (*yellow*), linear methylated (*light blue*), cyclic peptides (*red*), and cyclic *N*-methylated (*blue*).

This representation of the chemical space in Fig. 3 shows that peptides and PPIs occupy different regions of the chemical space as compared to FDA small molecules drugs. This observation suggests the potential to expand the medicinally relevant chemical space with peptides and PPIs.

5. CONCLUSIONS AND PERSPECTIVES

Inhibition of PPIs has rapidly becoming an attractive approach for the generation of new therapeutics including epigenetic drugs. Systematic approaches for target assessment, screening, and structure-based design have facilitated the advancement in the field. Computational tools now provide reliable assessments of hot spots and pockets found in PPIs, which offer reasonable starting points for inhibitor design. Preliminary data support the potential of peptides to develop effective modulators of PPIs. It is anticipated a continuous increase in the expansion of the epigenetic relevant chemical space with peptide or peptide-like compounds.

One of the major perspectives in the field of epigenetic PPI drug discovery is to increase the scaffold diversity of screening libraries. This can be achieved through the chemical synthesis of compound datasets with novel chemical scaffolds. A related perspective is to include in the screening libraries chemical compounds that sit outside the currently medicinally relevant chemical space.

ACKNOWLEDGMENTS

This work was supported by the Universidad Nacional Autónoma de México (UNAM), grants PAPIIT IA204016 and PAIP 5000-9163. B.I.D.-E. and J.J.N. acknowledges Consejo Nacional de Ciencia y Tecnología (CONACyT) for scholarships number 620289 and 622969, respectively.

REFERENCES

Abdel-Wahab, O., & Levine, R. L. (2013). Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. *Blood*, *121*, 3563–3572.

Albericio, F., & Kruger, H. G. (2012). Therapeutic peptides. *Future Medicinal Chemistry*, *4*, 1527–1531.

Anand, P., Brown, J. D., Lin, C. Y., Qi, J., Zhang, R., Artero, P. C., et al. (2013). BET bromodomains mediate transcriptional pause release in heart failure. *Cell*, *154*, 569–582.

Banerjee, C., Archin, N., Michaels, D., Belkina, A. C., Denis, G. V., Bradner, J., et al. (2012). BET bromodomain inhibition as a novel strategy for reactivation of HIV-1. *Journal of Leukocyte Biology*, *92*, 1147–1154.

Barbieri, I., Cannizzaro, E., & Dawson, M. A. (2013). Bromodomains as therapeutic targets in cancer. *Briefings in Functional Genomics*, *12*, 219–230.

Baylin, S. B., & Jones, P. A. (2016). Epigenetic determinants of cancer. *Cold Spring Harbor Perspectives in Biology*, *8*, a019505.
Borah, J. C., Mujtaba, S., Karakikes, I., Zeng, L., Muller, M., Patel, J., et al. (2011). A small molecule binding to the coactivator CREB-binding protein blocks apoptosis in cardiomyocytes. *Chemistry & Biology*, *18*, 531–541.
Brand, M., Measures, A. R., Wilson, B. G., Cortopassi, W. A., Alexander, R., Höss, M., et al. (2015). Small molecule inhibitors of bromodomain-acetyl-lysine interactions. *ACS Chemical Biology*, *10*, 22–39.
Bruey, J.-M., Bruey-Sedano, N., Luciano, F., Zhai, D., Balpai, R., Xu, C., et al. (2007). Bcl-2 and Bcl-XL regulate proinflammatory caspase-1 activation by interaction with NALP1. *Cell*, *129*, 45–56.
Bruno, B. J., Miller, G. D., & Lim, C. S. (2013). Basics and recent advances in peptide and protein drug delivery. *Therapeutic Delivery*, *4*, 1443–1467.
Buchwald, P. (2010). Small-molecule protein–protein interaction inhibitors: Therapeutic potential in light of molecular size, chemical space, and ligand binding efficiency considerations. *IUBMB Life*, *62*, 724–731.
Campbell, R. M., & Tummino, P. J. (2014). Cancer epigenetics drug discovery and development: The challenge of hitting the mark. *The Journal of Clinical Investigation*, *124*, 64–69.
Castillo-Aguilera, O., Depreux, P., Halby, L., Arimondo, P. B., & Goossens, L. (2017). DNA methylation targeting: The DNMT/HMT crosstalk challenge. *Biomolecules*, *7*, 3.
CenterWatch. (n.d.). FDA approved drugs. Retrieved April 24, 2017, from http://www.centerwatch.com/drug-information/fda-approved-drugs/year.
Cheray, M., Pacaud, R., Nadaradjane, A., Oliver, L., Vallette, F. M., & Cartron, P.-F. (2016). Specific inhibition of DNMT3A/ISGF3γ interaction increases the temozolomide efficiency to reduce tumor growth. *Theranostics*, *6*, 1988–1999.
Cowen, S. D., Russell, D., Dakin, L. A., Chen, H., Larsen, N. A., Godin, R., et al. (2016). Design, synthesis, and biological activity of substrate competitive SMYD2 inhibitors. *Journal of Medicinal Chemistry*, *59*, 11079–11097.
Craik, D. J., Fairlie, D. P., Liras, S., & Price, D. (2013). The future of peptide-based drugs. *Chemical Biology & Drug Design*, *81*, 136–147.
D'Andrea, L. D., Iaccarino, G., Fattorusso, R., Sorriento, D., Carannante, C., Capasso, D., et al. (2005). Targeting angiogenesis: Structural characterization and biological properties of a de novo engineered VEGF mimicking peptide. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 14215–14220.
Dawson, M. A., Gudgin, E. J., Horton, S. J., Giotopoulos, G., Meduri, E., Robson, S., et al. (2014). Recurrent mutations, including NPM1c, activate a BRD4-dependent core transcriptional program in acute myeloid leukemia. *Leukemia*, *28*, 311–320.
Dawson, M. A., Kouzarides, T., & Huntly, B. J. P. (2012). Targeting epigenetic readers in cancer. *The New England Journal of Medicine*, *367*, 647–657.
Delmore, J. E., Issa, G. C., Lemieux, M. E., Rahl, P. B., Shi, J., Jacobs, H. M., et al. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell*, *146*, 904–917.
Di Costanzo, A., Del Gaudio, N., Migliaccio, A., & Altucci, L. (2014). Epigenetic drugs against cancer: An evolving landscape. *Archives of Toxicology*, *88*, 1651–1668.
Ellert-Miklaszewska, A., Poleszak, K., & Kaminska, B. (2017). Short peptides interfering with signaling pathways as new therapeutic tools for cancer treatment. *Future Medicinal Chemistry*, *9*, 199–221.
Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W. B., Fedorov, O., et al. (2010). Selective inhibition of BET bromodomains. *Nature*, *468*, 1067–1073.
Fosgerau, K., & Hoffmann, T. (2015). Peptide therapeutics: Current status and future directions. *Drug Discovery Today*, *20*, 122–128.

Gallenkamp, D., Gelato, K. A., Haendler, B., & Weinmann, H. (2014). Bromodomains and their pharmacological inhibitors. *ChemMedChem*, *9*, 438–464.
Garnier, J.-M., Sharp, P. P., & Burns, C. J. (2014). BET bromodomain inhibitors: A patent review. *Expert Opinion on Therapeutic Patents*, *24*, 185–199.
Gautier, B., Miteva, M. A., Goncalves, V., Huguenot, F., Coric, P., Bouaziz, S., et al. (2011). Targeting the proangiogenic VEGF-VEGFR protein–protein interface with drug-like compounds by in silico and in vitro screening. *Chemistry & Biology*, *18*, 1631–1639.
Giotopoulos, G., Chan, W.-I., Horton, S. J., Ruau, D., Gallipoli, P., Fowler, A., et al. (2016). The epigenetic regulators CBP and p300 facilitate leukemogenesis and represent therapeutic targets in acute myeloid leukemia. *Oncogene*, *35*, 279–289.
Gonzalez, M. W., & Kann, M. G. (2012). Chapter 4: Protein interactions and disease. *PLoS Computational Biology*, *8* , e1002819.
González-Medina, M., Owen, J. R., El-Elimat, T., Pearce, C. J., Oberlies, N. H., Figueroa, M., et al. (2017). Scaffold diversity of fungal metabolites. *Frontiers in Pharmacology*, *8*, 180.
Hamano, N., Komaba, H., & Fukagawa, M. (2017). Etelcalcetide for the treatment of secondary hyperparathyroidism. *Expert Opinion on Pharmacotherapy*, *18*, 529–534.
Hay, D. A., Fedorov, O., Martin, S., Singleton, D. C., Tallant, C., Wells, C., et al. (2014). Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomains. *Journal of the American Chemical Society*, *136*, 9308–9319.
Heerboth, S., Lapinska, K., Snyder, N., Leary, M., Rollinson, S., & Sarkar, S. (2014). Use of epigenetic drugs in disease: An overview. *Genetics & Epigenetics*, *6*, 9–19.
Huang, J., Perez-Burgos, L., Placek, B. J., Sengupta, R., Richter, M., Dorsey, J. A., et al. (2006). Repression of p53 activity by Smyd2-mediated methylation. *Nature*, *444*, 629–632.
Ismail, R., & Csóka, I. (2017). Novel strategies in the oral delivery of antidiabetic peptide drugs-insulin, GLP 1 and its analogs. *European Journal of Pharmaceutics and Biopharmaceutics*, *115*, 257–267.
Itoh, Y., Suzuki, T., & Miyata, N. (2013). Small-molecular modulators of cancer-associated epigenetic mechanisms. *Molecular BioSystems*, *9*, 873–896.
James, L. I., & Frye, S. V. (2013). Targeting chromatin readers. *Clinical Pharmacology and Therapeutics*, *93*, 312–314.
Jeltsch, A., & Jurkowska, R. Z. (2016). Allosteric control of mammalian DNA methyltransferases—A new regulatory paradigm. *Nucleic Acids Research*, *44*, 8556–8575.
Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science*, *293*, 1074–1080.
Khazanov, N. A., & Carlson, H. A. (2013). Exploring the composition of protein–ligand binding sites on a large scale. *PLoS Computational Biology*, *9* , e1003321.
Kim, W., Bird, G. H., Neff, T., Guo, G., Kerenyi, M. A., Walensky, L. D., et al. (2013). Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. *Nature Chemical Biology*, *9*, 643–650.
Lan, F., Collins, R. E., De Cegli, R., Alpatov, R., Horton, J. R., Shi, X., et al. (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature*, *448*, 718–722.
Li, G.-Q., Guo, W.-Z., Zhang, Y., Seng, J.-J., Zhang, H.-P., Ma, X.-X., et al. (2016). Suppression of BRD4 inhibits human hepatocellular carcinoma by repressing MYC and enhancing BIM expression. *Oncotarget*, *7*, 2462–2474.
Link, K. A., Balasubramaniam, S., Sharma, A., Comstock, C. E. S., Godoy-Tundidor, S., Powers, N., et al. (2008). Targeting the BAF57 SWI/SNF subunit in prostate cancer: A novel platform to control androgen receptor activity. *Cancer Research*, *68*, 4551–4558.
López-Vallejo, F., Giulianotti, M. A., Houghten, R. A., & Medina-Franco, J. L. (2012). Expanding the medicinally relevant chemical space with compound libraries. *Drug Discovery Today*, *17*(13–14), 718–726.

Mayes, K., Qiu, Z., Alhazmi, A., & Landry, J. W. (2014). ATP-dependent chromatin remodeling complexes as novel targets for cancer therapy. *Advances in Cancer Research*, *121*, 183–233.
Medina-Franco, J. L., Martinez-Mayorga, K., & Meurice, N. (2014). Balancing novelty with confined chemical space in modern drug discovery. *Expert Opinion on Drug Discovery*, *9*(2), 151–165.
Meng, S., Zhang, L., Tang, Y., Tu, Q., Zheng, L., Yu, L., et al. (2014). BET inhibitor JQ1 blocks inflammation and bone destruction. *Journal of Dental Research*, *93*, 657–662.
Modell, A. E., Blosser, S. L., & Arora, P. S. (2016). Systematic targeting of protein–protein interactions. *Trends in Pharmacological Sciences*, *37*, 702–713.
Muheem, A., Shakeel, F., Jahangir, M. A., Anwar, M., Mallick, N., Jain, G. K., et al. (2016). A review on the strategies for oral delivery of proteins and peptides and their clinical perspectives. *Saudi Pharmaceutical Journal*, *24*, 413–428.
Mullard, A. (2012). Protein–protein interaction inhibitors get into the groove. *Nature Reviews. Drug Discovery*, *11*, 173–175.
Nicodeme, E., Jeffrey, K. L., Schaefer, U., Beinke, S., Dewell, S., Chung, C.-W., et al. (2010). Suppression of inflammation by a synthetic histone mimic. *Nature*, *468*, 1119–1123.
Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, *448*, 714–717.
Palmer, W. S., Poncet-Montange, G., Liu, G., Petrocchi, A., Reyna, N., Subramanian, G., et al. (2016). Structure-guided design of IACS-9571, a selective high-affinity dual TRIM24-BRPF1 bromodomain inhibitor. *Journal of Medicinal Chemistry*, *59*, 1440–1454.
Pawar, V. K., Meher, J. G., Singh, Y., Chaurasia, M., Surendar Reddy, B., & Chourasia, M. K. (2014). Targeting of gastrointestinal tract for amended delivery of protein/peptide therapeutics: Strategies and industrial perspectives. *Journal of Controlled Release*, *196*, 168–183.
Pelay-Gimeno, M., Glas, A., Koch, O., & Grossmann, T. N. (2015). Structure-based design of inhibitors of protein–protein interactions: Mimicking peptide binding epitopes. *Angewandte Chemie*, *54*, 8896–8927.
Prieto-Martínez, F. D., Gortari, E. F., Méndez-Lucio, O., & Medina-Franco, J. L. (2016). A chemical space odyssey of inhibitors of histone deacetylases and bromodomains. *RSC Advances*, *6*, 56225–56239.
Rooney, T. P. C., Filippakopoulos, P., Fedorov, O., Picaud, S., Cortopassi, W. A., Hay, D. A., et al. (2014). A series of potent CREBBP bromodomain ligands reveals an induced-fit pocket stabilized by a cation-π interaction. *Angewandte Chemie*, *53*, 6126–6130.
Santos, G. B., Ganesan, A., & Emery, F. S. (2016). Oral administration of peptide-based drugs: Beyond Lipinski's rule. *ChemMedChem*, *11*, 2245–2251.
Segura, M. F., Fontanals-Cirera, B., Gaziel-Sovran, A., Guijarro, M. V., Hanniford, D., Zhang, G., et al. (2013). BRD4 sustains melanoma proliferation and represents a new target for epigenetic therapy. *Cancer Research*, *73*, 6264–6276.
Sharma, D., Saxena, N. K., Davidson, N. E., & Vertino, P. M. (2006). Restoration of tamoxifen sensitivity in estrogen receptor-negative breast cancer cells: Tamoxifen-bound reactivated ER recruits distinctive corepressor complexes. *Cancer Research*, *66*, 6370–6378.
Smith, M. C., & Gestwicki, J. E. (2012). Features of protein–protein interactions that translate into potent inhibitors: Topology, surface area and affinity. *Expert Reviews in Molecular Medicine*, *14*, e16.
Smith, S. G., Sanchez, R., & Zhou, M.-M. (2014). Privileged diazepine compounds and their emergence as bromodomain inhibitors. *Chemistry & Biology*, *21*, 573–583.

Sperandio, O., Reynès, C. H., Camproux, A.-C., & Villoutreix, B. O. (2010). Rationalizing the chemical space of protein–protein interaction inhibitors. *Drug Discovery Today*, *15*, 220–229.

Stresemann, C., & Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. *International Journal of Cancer*, *123*, 8–13.

Sweis, R. F., Wang, Z., Algire, M., Arrowsmith, C. H., Brown, P. J., Chiang, G. G., et al. (2015). Discovery of A-893, a new cell-active benzoxazinone inhibitor of lysine methyltransferase SMYD2. *ACS Medicinal Chemistry Letters*, *6*, 695–700.

Tien, A. L., Senbanerjee, S., Kulkarni, A., Mudbhary, R., Goudreau, B., Ganesan, S., et al. (2011). UHRF1 depletion causes a G2/M arrest, activation of DNA damage response and apoptosis. *The Biochemical Journal*, *435*, 175–185.

Tsomaia, N. (2015). Peptide therapeutics: Targeting the undruggable space. *European Journal of Medicinal Chemistry*, *94*, 459–470.

Valentin, N., Acosta, A., & Camilleri, M. (2015). Early investigational therapeutics for gastrointestinal motility disorders: From animal studies to Phase II trials. *Expert Opinion on Investigational Drugs*, *24*, 769–779.

Villoutreix, B. O., Kuenemann, M. A., Poyet, J.-L., Bruzzoni-Giovanelli, H., Labbé, C., Lagorce, D., et al. (2014). Drug-like protein–protein interaction modulators: Challenges and opportunities for drug discovery and chemical biology. *Molecular Informatics*, *33*, 414–437.

Wu, D., Gao, Y., Qi, Y., Chen, L., Ma, Y., & Li, Y. (2014). Peptide-based cancer therapy: Opportunity and challenge. *Cancer Letters*, *351*, 13–22.

Zeng, L., & Zhou, M. M. (2002). Bromodomain: An acetyl-lysine binding domain. *FEBS Letters*, *513*, 124–128.

Zuber, J., Shi, J., Wang, E., Rappaport, A. R., Herrmann, H., Sison, E. A., et al. (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature*, *478*, 524–528.