

Содержание

1	Введение	2
1.1	Цели и актуальность работы	2
1.2	Список используемых сокращений	3
2	Обзор литературы	4
2.1	Рецепторы, сопряженные с G-белком	4
2.2	Сайты связывания и специфичность	5
2.3	Пространство лигандов для GPCR	6
2.4	Полифармакология	6
2.5	Off-target взаимодействия	8
3	Методы	10
3.1	Построение модели белка	10
3.2	Поиск off-target взаимодействий	14

1. Введение

(пока что оформление не по ГОСТу, не ставится на TexStudio, но надежда есть)

Цели и актуальность работы

Целью дипломной работы является разработка гибкого протокола поиска и оценки **off-target** эффектов разного типа у выбранного белка и его лигандов. Исходный белок не имеет экспериментально полученной пространственной структуры. Основная цель протокола – работа с рецепторами, сопряженными с G-белком.

Насколько нам (мне) известно, в мире не существует публично доступных протоколов, позволяющих по последовательности мишени произвести основанный на структурах лиганда, мишени и связывающего кармана поиск всевозможных **off-target** эффектов как со стороны лигандов, так и со стороны мишени.

Возникновение такого протокола позволит оценивать перспективность разработки или **respurposing** лекарств и необходимость расшифровки пространственной структуры мишени экспериментальными методами.

Список используемых сокращений

GPCR – рецепторы, сопряженные с G-белком (англ. G-protein-coupled receptors)

ТМ – трансмембранный участок

МГ – моделирование по гомологии

МД – молекулярная динамика

2. Обзор литературы

Рецепторы, сопряженные с G-белком

Семейство GPCR состоит из около 800 многофункциональных белков-рецепторов [1], регулирующих разнообразные внутриклеточные сигнальные каскады в ответ на гормоны, нейротрансмиттеры, ионы, фотоны, одоранты и другие стимулы [2]. Поэтому они играют важнейшую роль в физиологии и разработке лекарств, представляя собой привлекательную мишень для лекарственных средств. Около трети всех лекарств, одобренных Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов США, действуют именно на мишени из этого класса, хотя это всего чуть более ста различных рецепторов, то есть порядка десятой части всего семейства [3].

Белки, принадлежащие семейству GPCR, состоят из семи трансмембранных спиралей, связанных тремя внутриклеточными и внеклеточными петлями. Внеклеточная часть, с которой связывается лиганд, также включает в себя N-конец. Внутриклеточная часть содержит кроме петель также восьмую спираль и C-конец и взаимодействует с G белками, аррестинами и другими **downstream effectors** [4].

Трансмембранная часть является наиболее консервативной в структуре белков семейства, что не мешает рецепторам из различных подсемейств обеспечивать крайнее разнообразие в форме, размере и электростатических свойствах связывающихся с ними лигандов за счет вариаций

в структуре связывающих карманов. Их гидрофобность и закрытость с внеклеточной стороны связаны с функциями рецептора [4].

Несмотря на колоссальный прогресс в кристаллографии GPCR, пространственная структура известна только у небольшой доли рецепторов. Важная роль компьютерного моделирования состоит в раскрытии структур остальных рецепторов и их комплексов [5] и последующем рациональном создании лекарств.

Сайты связывания и специфичность

Хотя природные лиганды внутри семейства GPCR очень разнообразны, белки одного подтипа имеют практически одинаковые конформации активных сайтов, что позволяет моделировать их компьютерными методами с высокой точностью [6]. Некоторые подсемейства рецепторов взаимодействуют с одним и тем же **endogenous** лигандом, и в этом случае отсутствие больших различий в ортостерических связывающих карманах, где и происходит взаимодействие **endogenous** лиганда и рецептора, представляет вызов для поиска селективных лигандов и является одной из главных проблем в разработке безопасных и эффективных лекарств, действующих на GPCR [7].

Исторически дизайн лекарств был направлен на создание лигандов, подобных **endogenous**, которые искусственно активировали сигнальные пути. Другой классический подход состоит в создании антагонистов — веществ, способных конкурировать с природным лигандом, не активируя при этом работу рецептора.

В последнее время произошел взрывной рост количества новых методов использования GPCR в качестве мишени. Например, многие недавно разработанные лиганды действуют в активном сайте, топологически отделенном от ортостерического. Такие сайты и лиганды называются

«аллостерическими», причем лиганды могут как усиливать работу рецептора, так и ослаблять ее [8]. Возможна даже комбинация фармакофоров ортостерических и аллостерических лигандов (**bitopic ligands**) в одном лиганде, который будут теоретически иметь лучшую аффинность и селективность за счет большего количества связей с рецептором [9].

Простейший механизм активации GPCR включает в себя два состояния, между которыми балансирует рецептор: активное, в котором происходит передача сигнала внутрь клетки, и неактивное. Агонисты смещают это равновесие в сторону активной конформации, обратные агонисты - в сторону неактивного. Эффективность агониста определяется преобладанием *аффинности к активному состоянию над неактивным* (звучит коряво). К настоящему моменту стало очевидно, что различные лиганды, воздействуя на один и тот же рецептор, могут стабилизировать его в различных конформациях, так что активными остается только часть всех возможных сигнальных путей, в который вовлечен этот рецептор. Такой процесс называется **biased** агонизм [10].

Пространство лигандов для GPCR

(не очень понятно, что нужно писать здесь. Перечисление типов найденных для GPCR лигандов, их особенности? (но у всех разные вроде))

Полифармакология

При разработке лекарств важно добиться селективности, избавившись от побочных действий. Именно эта парадигма «одно лекарство – одна мишень», так называемая таргетированная терапия, до недавнего времени широко использовалась в фармакологии. С другой стороны, в

последнее время стала осознаваться важность полифармакологии, которая означает множественное, но специфичное воздействие лекарства на многие мишени, позволяющее добиться синергетического эффекта и более эффективного лечения комплексных заболеваний, таких как рак [11].

При этом полифармакология может выгодно отличаться от комбинирования нескольких лекарств, так как: (а) единственная молекула обычно имеет более предсказуемую и безопасную фармакокинетику; (б) часто действующие на несколько мишеней лекарства имеют большую эффективность на поздних стадиях заболевания; (в) не нужно учитывать эффекты перекрестного взаимодействия лекарств, которые, являясь негативными, переносятся хуже в случае комбинационной терапии; (г) при прочих равных меньше вероятность выработки лекарственной устойчивости к одному лекарству, чем к хотя бы одному из набора лекарств [11].

Стоит заметить, что каждый белковый домен в среднем содержит 3-5 связывающих карманов достаточного размера для связывания с типичными малыми лигандами [12]. Таким образом, существует возможность выбрать новый карман, отличный от ранее использовавшихся, для разработки лекарства. К тому же, количество видов связывающих карманов со статистически значимыми различиями оценивается, как меньшее 400 [12], что позволяет считать полифармакологическую картину взаимодействий лиганд-мишень неизбежной, и потому более перспективной, чем таргетированная.

Новая парадигма подчеркивает важность поиска всевозможных пар взаимодействий лиганд-мишень. Такой анализ может аккумулировать результаты уже известных связей, приводя к построению сложных сетей [11], но важнее уметь предсказывать такие взаимодействия. Так как перебор и оценка силы всех взаимодействий лиганд-мишень *in vivo* и *in vitro* является непрактичной, в этом направлении развиваются компью-

терные методы [13].

Off-target взаимодействия

Off-target взаимодействия – дополнительные взаимодействия выбранного лиганда/мишени с другими, кроме основных, мишенями/лигандами. Одной из основных проблем в поиске таких взаимодействий исходя из структуры является то, что часто эти взаимодействия в большой степени определяются подвижными частями рецептора, которые сложно или невозможно исследовать с достаточной атомарной точностью [14].

Принципиально, поиск субъектов **off-target** взаимодействия может осуществляться по структуре: (а) мишени; (б) лиганда; (в) связывающего сайта [15].

(а) при поиске возможных лигандов по известной структуре мишени, воспроизводится обычный процесс современного дизайна лекарств, при котором производится высокопроизводительный скрининг по базе возможных лигандов. Таким образом, в сущности, оценивается **druggability** рецептора. Процесс можно ускорить, используя поиск по так называемым «горячим точкам», то есть набору мест на поверхности мишени, где максимальна энергия связывания с потенциальным лигандом [16], что напоминает концепцию фармакофорного поиска.

(б) поиск по структуре лиганда по сути своей близок к понятию **repurposing’a**, которое заключается в поиске новых мишеней и применений для лекарств, которые уже выпущены на рынок. Это позволяет сократить расходы на преклиническую стадию и оптимизацию [17]. [16]

(в) связывающие сайты могут сравниваться по различным характеристикам, таким как геометрические и физикохимические свойства поверхности мишени, профили взаимодействия или с структура остова. Также нахождение связывающих карманов само по себе сложная задача,

к которой существует несколько подходов (добавить, как ее решать) [18].

Связывающие сайты могут описываться разными способами. Например, как трехмерный граф из вершин-атомов, содиненный ребрами-длинами. Или же как облако точек, то есть чисто геометрически. В этих моделях могут выделяться основанные на фармакофорном принципе черты, которые в дальнейшем позволяют значительно ускорить поиск. Один из наиболее затратных в вычислениях, но и чувствительных методов – построение карт электронной плотности [18]. [19] [20] [21]

3. Методы

посмотреть MOE by Chemical Computing Group?

Построение модели белка

Итогом развития геномного секвенирования в последнее время стал резкий рост количества известных белковых последовательностей, в то время, как только около одной сотой доли последовательностей охарактеризована с атомистической точностью и с использованием экспериментальных методов определения структуры [22].

В таких условиях полученные компьютерными методами модели структур белков часто являются ценными для выдвижения проверяемых гипотез. Такие модели, в целом, создаются с использованием методов сравнительного моделирования или свободного моделирования (**free modelling**), также называемых «ab initio» или «de novo» [22].

Сравнительное моделирование, или моделирование на основе гомологии (МГ), базируется на построении модели по известным структурам близких (**related** белков, как по шаблонам. Принцип **Free modelling** подхода в использовании не структуры близких белков, а в применении разнообразных методов, комбинирующих физику и известные особенности структур белков, например, сопоставление большого количества небольших фрагментов, выделенных из известных структур белков.

Конструирование белков этим методом обычно чрезвычайно затратно в плане вычислений [22]. Современные пакеты моделирования зачастую комбинируют эти два подхода, используя, если доступны шаблоны, МГ для построения основы-скелета белка, а затем уточняя положения петель, боковых цепей и частей без шаблона.

Моделирование требует наличия схемы сэмпирования конформаций с целью получения набора альтернативных структур. Также необходима оценивающая функция для ранжирования этих конформаций по качеству. Для этих целей было предложено большое количество физически обоснованных (**physics-based**) функций энергии и статистических потенциалов, полученных из анализа известных структур [23].

Так, например, MODELLER сначала накладывает входную последовательность на шаблонный остов, а затем, внося случайные смещения атомов, ищет локальный минимум оценивающей функции, повторяя эту процедуру несколько раз [22].

Значительное увеличение количества решенных структур GPCR позволяет строить с атомистической точностью пространственные структуры большого количества рецепторов при отсутствии экспериментально полученной структуры целевого белка [24].

Построение модели состоит из нескольких этапов, результатом которых является физически и биологически адекватная модель белка. Ими являются: (1) определение сходства набора последовательностей с целевой и выделение шаблонов; (2) выравнивание целевой последовательности и шаблона(ов); (3) построение модели, основанной на выравнивании с выбранными шаблонами; (4) предсказание точности модели [22].

Выделение шаблонов и множественное выравнивание последовательностей

Сначала, беря аминокислотную последовательность заданного белка с неизвестной пространственной структурой, необходимо получить набор белков с известной 3D-структурой для последующего «сшивания» соответствующих участков.

Для этого используются различные подходы к поиску шаблонов и множественному выравниванию соответствующих им последовательностей:

1) PSI-BLAST [25,26]: с помощью матрицы вероятностей замен аминокислот BLOSUM62 и штрафа за пропуски производится поиск наиболее близких последовательностей в базе до некоторого порога близости. Далее, с использованием программы Clustal [27] эти последовательности выравниваются с исходной в совокупности. А именно, ... **алгоритм**

2) HHBlits [28] – быстрее и чувствительнее BLAST, но ... (**недостатки**). Использует предварительно кластеризованные с помощью kClustal [29] базы Uniprot и nr.

Общее описание алгоритма: (а) приведение запроса из одной последовательности (выравнивание) или нескольких (множественное выравнивание) к скрытой марковской модели (СММ, сопоставление каждой позиции в последовательности вектора вероятностей размерности 1x20 обнаружить ту или иную аминокислоту на этом месте) производится добавлением к исходной последовательностей, отличающихся от нее заменами аминокислот на похожие по физико-химическим свойствам. При этом учитывается локальный контекст в 13 оснований вокруг замены, а суммарная энергия этих замен должна быть меньше некоторого задаваемого порога, что важно для скорости и чувствительности алгоритма. Далее, каждая строка в исходной СММ причисляется к одно-

му из 219 типичных профилей-кластеров последовательностей, создавая строку-профиль этой СММ. После этого сначала проводится выравнивание этого профиля СММ с предварительно созданными профилями из базы данных, потом – обычное выравнивание уже входной последовательности внутри соответствующего профилю СММ кластера базы данных. (**наверное, так много тут не надо, просто разобрался в алгоритме**)

3) GPCRdb.org – при исследовании GPCR проще всего использовать готовые выравнивания с сайта базы данных GPCRdb (*зачем тогда вообще другие нам?*).

Моделирование белка и оценка качества модели

Моделирование из выровненного набора шаблонов может производиться при помощи различных программных пакетов для предсказания четвертичной структуры белков. Среди них наиболее известным является MODELLER, но и, например, Rosetta [30] [31], Itasser [32], RaptorX [33] могут быть использованы. [34]

При использовании пакета MODELLER в случае идентичности последовательностей более 30%, в среднем более $\sim 60\%$ скелетных атомов моделируются корректно со средним квадратическим отклонением позиций $C\alpha$ атомов менее 3,5 Å. При меньшей идентичности, как правило, результат хуже [22].

какие у разных программ особенности и различия?

Оценка качества модели [35],

Уточнение модели

Хотя современные программы МГ производят проверки физической и химической адекватности полученных моделей, за счет неточностей в

выбранном силовом поле и просто недостаточности исходного покрытия последовательностями шаблонами, итоговые модели могут иметь значительные отличия от реальной структуры.

Улучшение точности модели может производиться средствами молекулярной динамики (МД) [36]. Производя множественные траектории МД с полученной МГ моделью, помещенной в нативную среду, могут быть получены несколько устойчивых подсостояний белка с локально минимальными свободными энергиями, которые и будут считаться наиболее вероятными конформациями белка в мембране, что важно для GPCR.

Поиск off-target взаимодействий

Определение, где находится активный центр и взаимодействует ли лиганд с рецептором, может быть проведено различными методами. Возможно использование докинга [37] и основанного на машинном обучении поиска....(*что еще?*)

Заметки:

слишком много слов «взаимодействие», надо придумать синонимов

Литература

- [1] Robert Fredriksson, Malin C. Lagerström, Lars-Gustav Lundin, and Helgi B. Schiöth. The g-protein-coupled receptors in the human genome form five main families. phylogenetic analysis, paralogon groups, and fingerprints. *Molecular Pharmacology*, 63(6):1256–1272, 2003.
- [2] Daniel Hilger, Matthieu Masureel, and Brian K. Kobilka. Structure and dynamics of GPCR signaling complexes. *Nature Structural and Molecular Biology*, 25(1):4–12, 2018.
- [3] Alexander S. Hauser, Misty M. Attwood, Mathias Rask-Andersen, Helgi B. Schiöth, and David E. Gloriam. Trends in gpcr drug discovery: new agents, targets and indications. *Nature Reviews Drug Discovery*, 16:829 EP –, Oct 2017.
- [4] Vsevolod Katritch, Vadim Cherezov, and Raymond C. Stevens. Diversity and modularity of G protein-coupled receptor structures. *Trends in Pharmacological Sciences*, 33(1):17–27, 2012.
- [5] Irina Kufareva, Vsevolod Katritch, Raymond C. Stevens, and Ruben Abagyan. Advances in gpcr modeling evaluated by the gpcr dock 2013 assessment: Meeting new challenges. *Structure*, 22(8):1120 – 1139, 2014.
- [6] Vsevolod Katritch, Vadim Cherezov, and Raymond C. Stevens. Structure-function of the g protein-coupled receptor superfamily.

Annual Review of Pharmacology and Toxicology, 53(1):531–556, 2013.
PMID: 23140243.

- [7] Vsevolod Katritch, Irina Kufareva, and Ruben Abagyan. Structure based prediction of subtype-selectivity for adenosine receptor antagonists. *Neuropharmacology*, 60(1):108 – 115, 2011. High Resolution.
- [8] Jeremy Shonberg, Ralf C. Kling, Peter Gmeiner, and Stefan Löber. Gpcr crystal structures: Medicinal chemistry in the pocket. *Bioorganic & Medicinal Chemistry*, 23(14):3880 – 3906, 2015. Selective GCPR Ligands.
- [9] Denise Wootten, Arthur Christopoulos, and Patrick M. Sexton. Emerging paradigms in gpcr allostery: implications for drug discovery. *Nature Reviews Drug Discovery*, 12:630 EP –, Aug 2013. Review Article.
- [10] J. Robert Lane, Lauren T. May, Robert G. Parton, Patrick M. Sexton, and Arthur Christopoulos. A kinetic view of gpcr allostery and biased agonism. *Nature Chemical Biology*, 13:929 EP –, Aug 2017. Perspective.
- [11] Andrew Anighoro, Jürgen Bajorath, and Giulio Rastelli. Polypharmacology: Challenges and opportunities in drug discovery. *Journal of Medicinal Chemistry*, 57(19):7874–7887, 2014.
- [12] Jeffrey Skolnick, Mu Gao, Ambrish Roy, Bharath Srinivasan, and Hongyi Zhou. Implications of the small number of distinct ligand binding pockets in proteins for drug discovery, evolution and biochemical function. *Bioorganic Medicinal Chemistry Letters*, 25(6):1163 – 1170, 2015.
- [13] Rajan Chaudhari et al. Computational polypharmacology: a new

- paradigm for drug discovery. *Expert Opinion on Drug Discovery*, 12(3):279–291, 2017. PMID: 28067061.
- [14] Kathryn A. Loving, Andy Lin, and Alan C. Cheng. Structure-based druggability assessment of the mammalian structural proteome with inclusion of light protein flexibility. *PLOS Computational Biology*, 10(7):1–13, 07 2014.
 - [15] Didier Rognan. Structure-based approaches to target fishing and ligand profiling. *Molecular Informatics*, 29(3):176–187, mar 2010.
 - [16] David R. Hall, Dima Kozakov, Adrian Whitty, and Sandor Vajda. Lessons from hot spot analysis for fragment-based drug discovery. *Trends in Pharmacological Sciences*, 36(11):724–736, Nov 2015.
 - [17] Eric March-Vila, Luca Pinzi, Noé Sturm, Annachiara Tinivella, Ola Engkvist, Hongming Chen, and Giulio Rastelli. On the integration of in silico drug design methods for drug repurposing. *Frontiers in Pharmacology*, 8(MAY):1–7, 2017.
 - [18] Christiane Ehrt, Tobias Brinkjost, and Oliver Koch. Impact of binding site comparisons on medicinal chemistry and rational molecular design. *Journal of Medicinal Chemistry*, 59(9):4121–4151, May 2016.
 - [19] Alvaro Cortés-Cabrera, Garrett M Morris, Paul W Finn, Antonio Morreale, and Federico Gago. Comparison of ultra-fast 2d and 3d ligand and target descriptors for side effect prediction and network analysis in polypharmacology. *British Journal of Pharmacology*, 170(3):557–567.
 - [20] Michal Brylinski. Local alignment of ligand binding sites in proteins for polypharmacology and drug repositioning. In *Protein function prediction*, volume 1611 of *Methods in Molecular Biology*, pages 109 – 122. Humana Press, New York, NY, 2017.

- [21] Rajiv Gandhi Govindaraj and Michal Brylinski. Comparative assessment of strategies to identify similar ligand-binding pockets in proteins. *BMC Bioinformatics*, 19(1):91, Mar 2018.
- [22] Benjamin Webb and Andrej Sali. Protein structure modeling with MODELLER. In *Methods in Molecular Biology*, pages 39–54. Springer New York, 2017.
- [23] Guang Qiang Dong, Hao Fan, Dina Schneidman-Duhovny, Ben Webb, and Andrej Sali. Optimized atomic statistical potentials: assessment of protein interfaces and loops. *Bioinformatics*, 29(24):3158–3166, 2013.
- [24] Christofer S. Tautermann. Gpcr homology model generation for lead optimization. In *Methods in Molecular Biology*, pages 115–131. Springer New York, nov 2017.
- [25] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403 – 410, 1990.
- [26] Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402, 1997.
- [27] Julie D. Thompson, Desmond G. Higgins, and Toby J. Gibson. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22):4673–4680, 1994.
- [28] Michael Remmert, Andreas Biegert, Andreas Hauser, and Johannes Söding. Hhblits: lightning-fast iterative protein sequence searching by hmm-hmm alignment. *Nature Methods*, 9:173 EP –, Dec 2011.

- [29] Maria Hauser, Christian E Mayer, and Johannes Söding. kClust: fast and sensitive clustering of large protein sequence databases. *BMC Bioinformatics*, 14(1):248, 2013.
- [30] Carol A. Rohl, Charlie E.M. Strauss, Kira M.S. Misura, and David Baker. Protein structure prediction using rosetta. In *Numerical Computer Methods, Part D*, volume 383 of *Methods in Enzymology*, pages 66 – 93. Academic Press, 2004.
- [31] Yifan Song, Frank DiMaio, Ray Yu-Ruei Wang, David Kim, Chris Miles, TJ Brunette, James Thompson, and David Baker. High-resolution comparative modeling with RosettaCM. *Structure*, 21(10):1735–1742, oct 2013.
- [32] Jianyi Yang, Renxiang Yan, Ambrish Roy, Dong Xu, Jonathan Poisson, and Yang Zhang. The i-TASSER suite: protein structure and function prediction. *Nature Methods*, 12(1):7–8, jan 2015.
- [33] Sheng Wang, Wei Li, Shiwang Liu, and Jinbo Xu. Raptorx-property: a web server for protein structure property prediction. *Nucleic Acids Research*, 44(W1):W430–W435, 2016.
- [34] Haiyou Deng, Ya Jia, and Yang Zhang. Protein structure prediction. *International Journal of Modern Physics B*, 32(18):1840009, 2018.
- [35] Min-Yi Shen and Andrej Sali. Statistical potential for assessment and prediction of protein structures. *Protein Sci*, 15(11):2507–2524, Nov 2006. 17075131[pmid].
- [36] Amin Nowroozi and Mohsen Shahlaei. A coupling of homology modeling with multiple molecular dynamics simulation for identifying representative conformation of gpcr structures: a case study on human

bombesin receptor subtype-3. *Journal of Biomolecular Structure and Dynamics*, 35(2):250–272, 2017. PMID: 26922838.

- [37] H. Pradeep and G. K. Rajanikant. A rational approach to selective pharmacophore designing: an innovative strategy for specific recognition of gsk3 β . *Molecular Diversity*, 16(3):553–562, Aug 2012.