

Application of in-silico approaches to cardiovascular disease

Citation for published version (APA):

Du, J. (2014). *Application of in-silico approaches to cardiovascular disease*. Maastricht University. <https://doi.org/10.26481/dis.20141008jd>

Document status and date:

Published: 01/01/2014

DOI:

[10.26481/dis.20141008jd](https://doi.org/10.26481/dis.20141008jd)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary

Over the last decades, the application of *in-silico* approaches in cardiovascular research has been increasing. In this thesis, we describe the development of an optimal protocol to perform a virtual ligand screening campaign. We also applied various *in-silico* approaches to guide and support the structure-function study of IRAK-M. Further, we have conducted coarse-grained molecular dynamic simulation to study and simulate the membrane-binding process of FVIIIa, the haemophilia A protein.

In **chapter 2** of this thesis, we have developed a new method to perform structure based VLS and validated our approach by a stochastic compound database. We combined two structure based VLS programs FRED and Surflex in a parallel computational environment, aiming to accelerate the docking time while insuring the enrichment of hit molecules. Multiple conformations for each of a total of 10 protein targets have been generated so as to sample the conformational space that is available to the target as a result of intrinsic protein flexibility. We found that the use of multiple conformers for a given target is preferable over that of the use of a single target conformation. It is advisable to apply a consensus of the FRED docking results for consecutive Surflex screening. We have presented a novel approach (stepwise improvement) that is useful to detect the optimal cut-off from a rigid docking, as is FRED, to a flexible docking, like Surflex. To generate an optimal consensus result from eight conformations, we tested 7 different methods and we present a new consensus method, which is named *S_{opt}*, which is designed to

produce consensus lists for a protein target with an optimal enrichment for both FRED and Surflex. As described in chapter 2, we have tested the VLS performance on several cardiovascular related protein targets, including *FXa*, *thrombin*, *tyrosine kinase* and propose the optimal approach for a VLS campaign for those targets. Moreover, several other homologous targets, that are of relevance to the cardiovascular system, can now be more optimally studied, guided by the findings of our studies. In **chapter 3** of this thesis, an *in-silico* study has been performed to understand the membrane-binding mechanism of human FVIII, and more in particular that of the binding of the FVIII C domains to membranes containing phosphatidylserine. We constructed coarse-grained models for the C1, C2 and C1+C2 domains and also for seven variant domains for which have been studied experimentally before. By performing MD simulations we were able to study membrane binding times (the time it takes for a domain to stably bind to the membrane), to identify the individual residues that are involved in membrane binding and which of these residues become buried in the membrane including mode of membrane-burial, the depth of buried residues, the tilting of the domains, the C1+C2 angle movement and the membrane-binding energies. We also studied the relationship between 1,2-Dioleoyl-sn-glycero-3-phosphoserine (DOPS) lipid content in the membrane and the membrane-binding. We found that the C1 and C2 domains show different membrane binding properties. Some C domain variants had abnormal membrane-binding modes compared to that of the wild type. The C1 domain membrane binding was two times slower than that of the C2 domain, while the C1+C2 domain bound similarly to the C2

domain. The C domain variants which were experimentally confirmed to possess a higher binding affinity we found to bind faster than wild type C2 domain for binding. We concluded that the membrane binding speed is dependent on electrostatic interactions and identified the important residues for membrane-binding time. The tilting of the C domains, after they have been bound to the membrane, is likewise important for membrane-binding. The simulation results showed that 5% of DOPS lipids in the membrane is optimal for membrane-binding. Through analysis of the compiled contributions of electrostatic and van der Waals interactions and the identified membrane-buried residues, we present a list of residues which we identify as key residues for FVIII membrane binding. The results help to explain the molecular causes for haemophilia A, in those instances where membrane-binding is compromised, and predict likely candidates for novel causative missense mutations in haemophilia A.

In **chapter 4** of this thesis, we describe the structure-function relationship of the death domain of IRAK-M by analysis of mutant variants of this protein which have been rationally designed after creation and analysis of a 3D model for the IRAK-M death domain. The toll-like receptor (TLR) signaling inhibitor IRAK-M is a major player involved in the proper functioning of monocytes and macrophages. We generated a high quality 3D structure model for the IRAK-M death domain and identified two most likely interaction areas that are involved in IRAK-M ligand/modulators binding. These areas were then targeted by means of structure-guided mutagenesis. We provided firstly a list of residues and/or combinations of several residues that are most likely to be involved in the interaction of the IRAK-M death

domain with other protein ligands in the TLR pathway. We then prepared a series of recombinant IRAK-M mutants according to the *in-silico* prediction in different cell lines allowing the study of the function of these IRAK-M variant molecules. Furthermore, we proposed a structure for a tetramer of the IRAK-M death domain based on the X-ray structure of IRAK-4/2 tetramers (3MOP.pdb). The models and mutagenesis results showed that the NF- κ B activating activity of IRAK-M is dependent on two different sites on the DD, with respectively W74 and R97 as critical IRAK-4 binding residues. Residues W74 and R97 may interact with IRAK4 and prevent the formation of the IRAK-4-IRAK-2-MyD88 complex and further inhibit the TLR2 and TLR4 mediated TNF and IL-6 production in human monocytes. Our results suggest that the tetramers may sandwich an IRAK-4-DD tetramer between the R97-exposing surface and the W74-exposing surface that are situated at respectively the top and bottom of the IRAK-4 tetramer. Residues W74 and R97 are also important for NF- κ B and ERK activation as well as for the inhibitory action of IRAK-M on TLR induced release of cytokines. Residue R70 and the stretch D19-A23 are specifically involved in ERK activation and IRAK-M expression levels. Our study provides novel insights in the molecular mechanisms of IRAK-M by differential use of its death domain. This may channel a way for the rational design of IRAK-M inhibitors, which may be applied in future treatment strategies and which could potentially improve innate immunity in vulnerable patients or CVD patients.

In the **chapter 5** of this thesis, we expanded the *in-silico* approaches from chapter 4 to the kinase domain of IRAK-M. In this chapter, we analyzed the IRAK-M protein solely via bioinformatics methods. By

homology modeling and data analysis, we confirmed that the likely reason that causes the kinase domain to be inactive. For example, the kinase domain lacks a primary phosphorylation site, followed by a histidine instead of an arginine at the P+1 site. Serine residues are missing in the activation region and the ATP binding pocket is too narrow to accommodate ATP substrate although tyrosine (Y241) is present as the gatekeeper residue in the pocket. The chapter 5 provides a more extensive description of the 3D structure of the death domain as compared to chapter 4 and moreover describes the 3D structure of the kinase domain. The quality of the structures is discussed after evaluation by means of different evaluation factors such as 3D packing quality, bond quality, hydrogen bonding and structural stability. The IRAK-M kinase model structure is compared to that of other active kinase domains. Finally, residues which are most likely to interact with the kinase substrates are predicted.

Samenvatting

De toepassing van onderzoek dat expliciet gebruik maakt van computers, ook wel *in silico* methoden genoemd, heeft gedurende de laatste tien tot twintig jaar ook binnen het cardiovasculaire onderzoeksveld een vlucht genomen. In dit proefschrift wordt de ontwikkeling van een optimaal protocol beschreven dat gebruikt kan worden voor het vinden van kleine moleculen die binden aan een bepaald doeleiwit door middel van zogenaamde virtuele ligand screening. Verder hebben wij *in silico* methoden toegepast om de structuurfunctie studie van IRAK-M te sturen en te ondersteunen. Tenslotte hebben wij grofschalige moleculaire dynamica simulaties uitgevoerd, om daarmee de binding van FVIIIa, het hemofilie A eiwit, aan een lipide membraan te kunnen simuleren en te kunnen bestuderen.

Hoofstuk 2 beschrijft de ontwikkeling van een nieuwe methode om structuurgedreven virtuele ligand screening te bedrijven. Hierbij zijn twee bestaande programma's, FRED en Surflex, in een parallelle rekenomgeving toegepast, met als doel om de totale tijd van de screening te versnellen en de doelmatigheid van de gehele procedure te verhogen. Hierbij zijn er tien model eiwitten gebruikt om onze nieuwe methode mee op te zetten. Door gebruik te maken van verschillende conformaties van eenzelfde doeleiwit verkregen wij betere resultaten dan wanneer wij slechts een enkele conformatie van dit eiwit gebruikten. Om te bepalen hoe het beste de beide programma's, in combinatie met het gebruik van meerdere doeleiwit structuren konden worden gecombineerd, hebben wij zeven

verschillende methodes getest. Eén van deze methoden, Sopt genaamd, presteerde beter dan de andere consensus methoden.

Hoofdstuk 3 van dit proefschrift beschrijft een *in silico* studie die is uitgevoerd om de membraan bindende eigenschappen van de humane stollingsfactor VIII te bestuderen, en meer in het bijzonder de binding van de carboxy-terminale C-domeinen van factor VIII aan modelmembranen die het fosfolipide molecuul fosfatidylserine bevatten. Door middel van grofschalige moleculaire dynamica experimenten waren we in staat de bindingstijden te berekenen die behoren bij de binding van een C1 domein, een C2 domein of een C1+C2 domein aan een lipidelaag. Hierbij waren we in staat individuele aminozuren te identificeren die betrokken zijn bij het bindingsmechanisme en meerdere individuele parameters te kwantiteren die de binding van C-domeinen aan de membraan beschrijven.

Naast de gesimuleerde wildtype domeinen, werden er ook enkele gesimuleerde varianten van de C-domeinen getest ter controle van de juistheid van de door ons gebruikte *in silico* methoden. In alle gevallen waren de gevonden resultaten in overeenstemming met data uit de literatuur. Dit hoofdstuk wordt afgesloten met een voorstel dat het bindingsmechanisme beschrijft waarmee C-domeinen op calcium onafhankelijke wijze kunnen binden aan negatief geladen membranen.

Met dit hoofdstuk presenteren wij niet alleen een algemeen mechanisme voor binding waarbij individuele aminozuren kunnen

worden aangewezen die van groot belang zijn voor de membraan binding van factor VIII. Deze informatie en de door ons toegepaste *in silico* simulaties kunnen gebruikt worden om een moleculaire verklaring te geven voor de bloedingsneiging die bij hemofilie A patiënten bestaat, zowel in geval van bekende mutaties, maar ook voor nieuw ontdekte mutaties.

Hoofdstuk 4 van dit proefschrift geeft een beschrijving van de structuurfunctie relaties van het zogenaamde death domein van humaan IRAK-M, een eiwit dat deel is van het menselijk immuunsysteem. De manier waarbij we hier onderzoek hebben verricht is door het maken van een 3D homologie model voor dit eiwit domein en dit model vervolgens te gebruiken om voorspellingen te doen over potentiële interactie gebieden tussen IRAK-M en zijn bindingspartners. Deze informatie is op zijn beurt gebruikt om op rationele wijze mutaties aan te brengen in het IRAK-M. Door functionele analyse te doen van, tot expressie gebrachte recombinante varianten van het IRAK-M, is er waardevolle informatie verkregen over de functie van dit eiwit. Belangrijke aminozuren zijn geïdentificeerd en voorts is er een voorstel gedaan over een mogelijke quaternaire structuur waarin het IRAK-M een remmende rol zou kunnen vervullen door complexering met IRAK-4.

Hoofdstuk 5 beschrijft een uitbreiding van de *in silico* studies die gedaan zijn aan IRAK-M door ook expliciet het kinase domein van humaan IRAK-M te bestuderen. Het betreft hier uitsluitend de toepassing van bioinformatica methodes waarbij een 3D model is gemaakt voor het kinase domein van IRAK-M. Bestudering van dit

model, en een vergelijking met andere kinase structuren geeft ons een beter inzicht in de functionaliteit van dit domein, en geeft met name een verklaring waarom dit kinase inactief is. Tenslotte beschrijft dit hoofdstuk een uitgebreide analyse van de 3D structuur van het IRAK-M death domein, die voortbouwt op de reeds in hoofdstuk 4 genoemde resultaten.

总结

在过去的几个十年中，生物信息学方法广泛的应用于心血管基础医学领域。在该部论文中，我们优化了计算机辅助药物筛选(virtual ligand screening, VLS)的算法；运用结构生物信息学和点突变来研究免疫蛋白 IRAK-M 的结构与功能；运用分子动力学计算模拟研究血友病 A 型相关凝血因子 FVIIIa 并阐释其与细胞膜结合的机理。

在本论文的第二章中，我们优化了基于结构 VLS 的算法，并通过随机选取小分子数据库来验证此算法。该算法加快了筛选时间，同时也提高了筛选准确率。本实验选取十个靶蛋白并且使每个蛋白分子产生八个构象以便于模拟蛋白质高级结构的可变性。我们发现，使用多种构象做虚拟筛选的效果优于仅仅使用单一构象。我们推荐使用整合后的 FRED 结果然后进而运行 Surfex 筛选。我们提出了一种新的方法（逐步提高检测法）对刚性筛选结果的选择是有效的。为了得到最佳筛选效果，我们测试了七种不同的方法。其中最佳的组合方法被命名为 *Sopt*。正如第二章所述，我们已经在多个心血管疾病相关蛋白分子，例如凝血因子 Xa，凝血酶，酪氨酸激酶上面做了实验验证，并为这些蛋白靶点提供了最佳筛选参数。第 3 章描述了结构生物信息学在研究人类凝血因子 FVIII 的膜结合的生物学机制。我们构建了 FVIII 蛋白分子 C1, C2, C1 + C2 结构域和七个突变体的三维模型。通过进行分子动力学模拟，我们能够计算这些模型结合到细胞膜上所需的时间（结合时间），并且能够鉴别出直接作用于细胞膜结合的具体氨基酸残基，插入到细胞膜的深度，结构域的最佳结合角度以及结合能。我们还研究了在细胞膜中 1,2 - 二油酰基-sn-甘油基-3 - 磷酸丝氨酸（

DOPS) 脂质含量和膜结合之间的关系。我们发现, C1 和 C2 结构域具有不同的膜结合特性。相比于野生型中的 C 结构域, 有些 C 结构域的突变体有异常膜结合模式。C1 结构域的膜结合时间比的 C2 结构域慢 2 倍, 而 C1 + C2 结构域的结合时间基本类似于 C2 结构域。实验验证具有比野生型结合能高的突变体也具有比野生型更短的结合时间。我们推论 C 结构域的膜结合特性依赖于静电相互作用, 并鉴定出与结合时间相关的重要氨基酸残基。当结合到细胞膜后, C 结构域的倾斜被认为对其结合能力产生重要影响。研究结果表明, 当细胞膜里面含有 5% 的 DOPS, C 结构域的膜结合效果最佳。结合静电和范德华相互作用的因素, 我们预测出一系列参与到细胞膜结合的氨基酸残基。该研究结果有助于从分子水平解释血友病 A 的致病原因。

在本论文的第四章中, 我们通过分析所构建的 IRAK-M death domain 三维结构的模型, 并结合点突变实验研究了其结构与功能的关系。Toll 样受体 (TLR) 信号传导抑制剂 IRAK-M 是参与单核细胞和巨噬细胞的正常运作的主要蛋白。我们构建了高品质的 3D 结构模型, 并预测了两个配体结合区域。我们在这些区域研究其氨基酸定点突变。我们得出了一系列最有可能参与到分子互作的氨基酸。此外, 基于模板晶体结构, 我们构建了 IRAK-M death domain 的四聚体的结构。该模型和点突变研究结果表明, IRAK-M 在 NF- κ B 激活途径依赖两个不同的位点, W74 和 R97 分别为关键氨基酸。氨基酸残基 W74 和 R97 可与 IRAK4 相互作用并阻止 IRAK-4-IRAK-2-MyD88 的复合物的形成, 并进一步抑制 TLR2 和 TLR4 介导的 TNF 和 IL-6 的产生。研究结果表明该四聚体可能夹裹 IRAK-4 四聚体。在 NF- κ B 和 ERK 的激活途径中, 氨基酸 W74 和 R97 同样行使重要功能。R70 和 D19-A23 区域可能特异的参与到

ERK 活化过程和影响到 IRAK-M 表达水平。本研究揭示了 IRAK-M 参与到多种生物途经的分子机制。本实验为针对 IRAK-M 为靶点的理性药物研发提供了重要思路。这可能被应用到未来治疗心血管方面的疾病或者改善患者的生活质量。

在本论文的第五章中，我们将生物信息学的方法从 death domain 拓展至激酶结构域。我们从生物信息学方面分析了 IRAK-M 激酶结构域没有活性的原因。例如，该激酶结构域缺乏主磷酸化位点，在随后的位点上组氨酸代替精氨酸。激活区域缺失丝氨酸残基，并且 ATP 结合区域太窄进而无法容纳 ATP 底物。本章详细的描述了 IRAK-M 的两个结构域的三维结构信息并讨论了其三维模型的质量，例如分析其原子排列质量，分子键，氢键和结构稳定性。研究对 IRAK-M 激酶 ATP 结合区和潜在的催化位点做了充分的描述，并对比了其他具有活性的激酶。最终，一些最有可能参与到其激酶底物结合的氨基酸被预测。