Department of Pharmacology and Toxicology
Faculty of Medicine
University of Oulu
Division of Pharmacology and Pharmacotherapy
Faculty of Pharmacy
University of Helsinki

DISCOVERY OF CARDIOPROTECTIVE ISOXAZOLE-AMIDE COMPOUNDS TARGETING THE SYNERGY OF TRANSCRIPTION FACTORS GATA4 AND NKX2-5

MIKA VÄLIMÄKI

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine,
University of Oulu, for public examination in Auditorium 1041,
Viikki Biocenter 2, Helsinki, on December 14th 2018, at 12 noon.

Supervised by
Professor Heikki Ruskoaho
Professor Antti Poso

Reviewed by
Professor Ullamari Pesonen
Docent Pekka Postila

Opponent
Professor Outi Salo-Ahen

ISBN 978-952-94-1251-8 (Paperback)
ISBN 978-952-94-1252-5 (PDF)

Unigrafia – Helsinki University Print
Helsinki 2018

Välimäki, Mika, Discovery of cardioprotective isoxazole-amide compounds targeting the synergy of transcription factors GATA4 and NKX2-5

University of Oulu Graduate School; University of Oulu, Faculty of Medicine, Department of Pharmacology and Toxicology

Drug Research Program; University of Helsinki, Faculty of Pharmacy, Division of Pharmacology and Pharmacotherapy

Abstract

Acute myocardial infarction is a life-threatening condition that occurs as a result of reduced blood flow in the cardiac muscle, eventually leading to tissue damage. In infarcted areas, cardiomyocytes have insufficient ability to proliferate and replace the injured cells, which is associated with a deficient pumping capacity. A strictly regulated combinatorial interplay of transcription factors, e.g., GATA4, NKX2-5, TBX5, and MEF2C, orchestrates cardiac type gene expression during the cardiomyocyte differentiation and maturation processes. The aim of the present study was to (i) characterize the protein-protein interaction of the cardiac transcription factors GATA4-NKX2-5, (ii) evaluate the chemical agents that modify the synergy of GATA4-NKX2-5 *in vitro*, (iii) examine the capacity of the lead compound to promote myocardial repair *in vivo* after myocardial infarction and other cardiac injuries and (iv) study the structural features of the compound important for metabolism and cytotoxicity.

Integration of the experimental mutagenic data with computational modeling suggests that the structural architecture of the GATA4-NKX2-5 interaction resembles the protein structure of the conserved DNA binding domain of nuclear receptors. Fragment-based screening, reporter gene-based optimization and pharmacophore searching were utilized to identify the most potent lead compound targeting the GATA4-NKX2-5 interaction: *N*-[4-(diethylamino)phenyl]-5-methyl-3-phenylisoxazole-4-carboxamide. This compound presented anti-hypertrophic effects *in vitro* and cardioprotective effects *in vivo*. In addition, structural analysis of the lead compound revealed the signature molecular features for metabolism and cytotoxicity. Current drug treatments are able to delay, but not prevent the progress of the heart failure; therefore, modulators of protein-protein interactions of key transcription factors may represent a novel class of pharmaceuticals for cardiac remodeling and repair.

Keywords: cardioprotection, GATA4, hypertrophy, lead compound, protein-protein interaction, reprogramming, transcription factor

Välimäki, Mika, Transkriptiotekijöiden GATA4 ja NKX2-5 yhteisvaikutukseen kohdennettujen isoksatsoli-amidi yhdisteiden suunnittelu ja kehitys

Oulun yliopiston tutkijakoulu; Oulun yliopisto, Lääketieteellinen tiedekunta, Farmakologian ja Toksikologian osasto

Lääketutkimusohjelma; Helsingin yliopisto, Farmasian tiedekunta, Farmakologian ja Lääkehoidon osasto

Tiivistelmä

Sydäninfarkti on henkeä uhkaava verenkierron häiriö, joka syntyy veren virtauksen äkillisen vähentymisen seurauksena sydänlihaksessa aiheuttaen kudosvaurion. Vaurioituneen sydänlihaskudoksen kyky uusiutua tai korvata kuolleet sydänlihassolut uusilla on puutteellinen, ja tämän seurauksena sydämen pumppauskyky heikkenee. Transkriptiotekijöiden GATA4, NKX2-5, TBX5 ja MEF2C muodostamat ja koordinoimat proteiinikompleksit säätelevät sydänsolujen geenien ilmenemistä solujen elinkaaren aikana. Väitöskirjatyön tavoitteena oli (i) karakterisoida geeninsäätelytekijöiden GATA4-NKX2-5 molekyylirakenteet ja niiden keskinäinen vuorovaikutus, (ii) seuloa kemiallisia yhdisteitä, jotka muokkaavat GATA4-NKX2-5 proteiinikompleksin aikaansaamaa geeniaktivaatiota, (iii) tutkia johtoyhdisteen vaikutuksia *in vivo* sydäninfarktia ja painekuormitusta kuvaavissa eläinmalleissa, ja (iv) tutkia johtoyhdisteen molekyylirakenteen yhteyttä yhdisteen metaboliaan ja sytotoksisuuteen.

Väitöskirjatyö osoittaa molekyylimallinuksen ja kokeellisten tulosten perusteella, että geeninsäätelytekijöiden GATA4-NKX2-5 proteiinikompleksin orientaatio matkii tumareseptoriperheen DNA domeenin tertiäärirakennetta. Molekyylifragmenttien, lusiferaasi-reportterikokeen ja farmakoforimallin avulla seulottiin ja optimoitiin sitoutumisvoimakkuudeltaan lupaavin GATA4-NKX2-5 proteiinikompleksin toimintaan vaikuttava johtoyhdiste: *N*-[4-(dietyyliamino)fenyyli]-5-metyyli-3-fenyyli-isoksatsoli-4-karboksamidi. Johtoyhdisteellä havaittiin solu- ja eläinmalleissa hypertrofiaa estäviä vaikutuksia *in vitro* ja sydäntä suojaavia vaikutuksia *in vivo*. Väitöskirjatyö osoitti lisäksi aktiivisten molekyylien rakenneominaisuuksia, jotka keskeisesti vaikuttavat yhdisteiden metaboliaan ja sytotoksisuuteen. Nykyinen lääkehoito hidastaa, mutta ei pysäytä sydänlihasvaurioon liittyvän kroonisen sydämen vajaatoiminnan etenemistä. Lääkevaikutuksen kohdentaminen sydämen keskeisten transkriptiotekijöiden yhteisvaikutukseen avaa uuden mahdollisen tutkimuslinjan sydänlihasvaurion estossa ja korjauksessa.

Asiasanat: GATA4, hypertrofia, johtoyhdiste, proteiinikompleksi, solujen uudelleenohjelmointi, sydäntä suojaava, transkriptiotekijä

To my family

Acknowledgements

This work was carried out at the Department of Pharmacology and Toxicology, Faculty of Medicine, University of Oulu and Division of Pharmacology and Toxicology, Faculty of Pharmacy, University of Helsinki during the years 2002-2018. I wish to thank Professor Emeritus Olavi Pelkonen, the head of the Department of Pharmacology and Toxicology during the years at the University of Oulu, for providing excellent conditions for scientific research and Professor Raimo K. Tuominen, the head of the Division of Pharmacology and Toxicology during the years at the University of Helsinki, for providing an encouraging atmosphere and an excellent environment for research work.

I owe my deepest gratitude to my supervisor Professor Heikki Ruskoaho for providing me with the opportunity to carry out the thesis work in research group of regenerative pharmacology. Your visionary leadership with ambitious scientific goals and high working standards transferred me to a never-to-forget journey into molecular cardiology. I appreciate your patience, thoughtful insights and expertise in cardiology and drug research, which were all essential for the success of these research attempts. It has been a privilege and a great pleasure for me to be a part of your research group.

I am grateful to Professor Antti Poso for co-supervising this thesis work. His expert knowledge and methodological guidance in computational drug design has been of great importance for the completion of this work. I express my warmest thanks for scientific collaboration and valuable advice during this time.

I am truly thankful to Sini Kinnunen and Marja Tölli for the long-term friendship and scientific collaboration over the GATA-research project. It has been a great privilege to work with you and to share research ideas and the unique moments of discoveries. Over the years, you really showed me how commitment and hard work pays off. Without your efforts, this series of GATA-studies would never have been completed.

I express my gratitude to Professor Gösta Brunow, Dr. Jorma Haarala, Professor Tapio Hase, Professor Ossi Renkonen, and Dr. Gerd Wohlfahrt for the introduction into the organic chemistry and molecular sciences. Your insightful teaching and guidance, at the early days, had a major contribution to my understanding of drug research and molecular biology.

My special thanks go to my current and past research colleagues and co-authors at the University of Helsinki: Ingo Aumüller, Samuli Auno, Amanda Coniel, Gustaf Boije af Gennäs, Mikael Jumppanen, Maria Jäntti, Tuuli Karhu, Bogac Kaynak,

Alexandros Kiriazis, Liisa Konttinen, Annika Korvenpää, Robert Leigh, Alexander March, Lotta Pohjolainen, Virpi Talman, Henri Xhaard and Jari Yli-Kauhaluoma, and at the University of Oulu: Jani Aro, Jukka Hakkola, Leena Kaikkonen, Risto Kerkelä, Hanna Leskinen, Johanna Magga, Sampsa Pikkarainen, Jaana Rysä, Raisa Serpi, Heikki Tokola and Miia Turpeinen. I appreciate your cooperation and scientific contribution to this thesis work.

I owe my sincere thanks to my boyhood friends Aki Autio, Sakari Haakana, Jarmo Kuljunlahti and Juha Peltomäki and family friends Riikka Huhtala, Päivi Käyhkö, Arto Käyhkö, Ari Lehtola, Toni Markkanen, Tiina Nurmi, Mika Nurmi, Pete Perälä and Kati Poranen, for sharing long-standing friendship and cheerful moments in life. You fully deserve all the happiness, wealth and prosperity that comes your way.

I am grateful to my parents Maire and Juhani Välimäki for their unconditional encouragement and support over the years and to my sister Satu and brother Niko and their families who have helped me generously with my scientific work as well as during our collective family events.

Finally, I wish to express my gratitude to my wife Anu. I admire your positive attitude, wisdom and beauty. You always bring light and joy to your surroundings, and never-ending love to our beautiful children—Joel, Jonna and Anton. God bless you all.

This research project has been financially supported by the Academy of Finland, Business Finland, the Sigrid Juselius Foundation and the Finnish Foundation for Cardiovascular Research.

Vantaa, 08.08.2018 Mika Välimäki

Abbreviations

ADME	absorption, distribution, metabolism and excretion
ANP	atrial natriuretic peptide
BNP	B-type natriuretic peptide
CAS	chemical abstract service
$CDCl_3$	deuterated chloroform
CF	cardiac fibroblast
ChIP-seq	chromatin immunoprecipitation sequencing
clogP	calculated log octanol/water partition
CREB	cyclic adenosine monophosphate response element-binding protein
D5EB	mESC-derivatives from day 5 embryoid bodies
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulfoxide
DMSO-d_6	deuterated dimethyl sulfoxide
DNA	deoxyribonucleic acid
EHT	extended Hückel theory
ET-1	endothelin 1
ERE	estrogen response element
ERK	extracellular signal-regulated kinases
ESC	embryonic stem cells
ESI	electrospray
FBDD	fragment-based drug discovery
FBS	fetal bovine serum
FDA	Food and Drug Administration
FEP	free-energy perturbation
FOG2	friend of GATA2
ΔG	Gibbs free energy of binding
GATA4	GATA binding protein 4
GBSA	generalized Born and surface area
GB/VI	generalized Born/volume integral
GPCR	G protein-coupled receptor
GSK-3β	glycogen synthase kinase 3β
ΔH	change in enthalpy
HAND2	heart and neural crest derivatives expressed 2
HDAC	histone deacetylase
hiPSC	human induced pluripotent stem cell

hiPSC-CM	hiPSC-derived cardiomyocytes
HPLC	high-performance liquid chromatography
HTS	high-throughput screening
i.p.	intraperitoneal injection
iPSC	induced pluripotent stem cell
Isx	3,5-disubstituted isoxazoles
ITC	isothermal titration calorimetry
K	ligand binding constant
K_{off}	ligand disassociation rate constant
K_{on}	ligand association rate constant
KLF	Kruppel-like factor
LDH	lactate dehydrogenase
MAPK	mitogen-activated protein kinase
MC	Monte Carlo
MD	molecular dynamics
MEF	mouse embryonic fibroblasts
MEF2C	myocyte enhancer factor 2C
mESC	mouse embryonic stem cells
miRNA	microRNA
MM	molecular mechanics
MOE	molecular operating environment
MS	mass spectrometry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium
NCI	National Cancer Institute
NKE	NKX2-5 binding element
NKX2-5	NK2 homeobox 5
NMR	nuclear magnetic resonance
NRVC	neonatal rat ventricular cardiomyocytes
Oct-4	octamer-binding transcription factor-4
p300	histone acetyltransferase p300
PAINS	pan-assay interference compounds
PE	phenylephrine
PBSA	Poisson-Boltzmann and surface area
PDB	protein data bank
PPI	protein-protein interaction
PS	penicillin-streptomycin
PSI-BLAST	position-specific iterative basic local alignment search tool

QM	quantum mechanics
QTOF	quadrupole time-of-flight
R	gas constant
RMSD	root mean square deviation
RNA	ribonucleic acid
ΔS	change in entropy
SAR	structure-activity relationships
SD	standard deviation
Shz	sulfonyl-hydrazone
SME	small to mid-size enterprise
Sox2	sex determining region Y-box 2
SPR	surface plasmon resonance
SRE	serum response element
SRF	serum response factor
SUMO-1	small ubiquitin-like modifier-1
t_R	residence time
T	temperature
TBX5	T-box 5
TF	transcription factor
TGF	transforming growth factor
TOF	time-of-flight
TSA	trichostatin A
UPLC	ultra-performance liquid chromatography
VEGF	vascular endothelial growth factor
WHO	World Health Organization

List of original articles

This thesis is based on the following publications, which are referred throughout the text by their Roman numerals:

I Kinnunen, S.*, Välimäki, M.*, Tölli, M., Wohlfahrt, G., Darwich, R., Komati, H., Nemer, M., Ruskoaho, H. (2015). Nuclear receptor-like structure and interaction of congenital heart disease-associated factors GATA4 and NKX2-5. *PLoS ONE, 10*(12). https://doi.org/10.1371/journal.pone.0144145

II Välimäki, M. J.*, Tölli, M. A.*, Kinnunen, S. M., Aro, J., Serpi, R., Pohjolainen, L., Talman, V., Poso, A., Ruskoaho, H. J. (2017). Discovery of small molecules targeting the synergy of cardiac transcription factors GATA4 and NKX2-5. *Journal of Medicinal Chemistry, 60*(18), 7781–7798. https://doi.org/10.1021/acs.jmedchem.7b00816

III Kinnunen, S. M.*, Tölli, M.*, Välimäki, M. J., Gao, E., Szabo, Z., Rysä, J., Ferreira, M. P. A., Ohukainen, P., Serpi, R., Correia, A., Mäkilä, E., Salonen, J., Hirvonen, J., Santos, H. A., Ruskoaho, H. (2018). Cardiac actions of a small molecule inhibitor targeting GATA4-NKX2-5 interaction. *Scientific Reports, 8*(1). https://doi.org/10.1038/s41598-018-22830-8

IV Karhu, S. T., Välimäki, M. J., Jumppanen, M., Kinnunen, S. M., Pohjolainen, L., Leigh, R. S., Auno, S., Földes, G., Boije af Gennäs, G., Yli-Kauhaluoma, J., Ruskoaho, H., Talman, V. (2018). Stem cells are the most sensitive screening tool to identify toxicity of GATA4-targeted novel small-molecule compounds. *Archives of Toxicology, 92*(9), 2897–2911. https://doi.org/10.1007/s00204-018-2257-1

*co-first authors

Contents

1 Introduction

According to a recent statement by the World Health Organization (WHO, 2017), the cardiovascular diseases are the most prevalent disease-causing deaths globally, taking the lives of 17.7 million people annually, equivalent to 31% of all global deaths. Unhealthy lifestyles are associated with the occurrence of cardiovascular diseases and consist of tobacco smoking, physical inactivity, poor diet and the harmful use of alcohol. These known risk factors may consequently prime the high-risk individuals with raised blood pressure, elevated glucose levels and obesity. The majority of the pathological cardiovascular events arise due to coronary heart disease, heart failure and strokes. General health-promoting objectives are to efficiently reduce the risk factors, and consequently, the incidence and mortality of cardiovascular diseases, thereby reducing the burden of the global cardiovascular disease epidemic.

Insufficient pumping capacity of the heart is the hallmark of congestive heart failure. This physiological deficiency is commonly initiated by pathological events at the cellular level, such as the loss of myocytes after cardiac infarction or the chronic pressure overload leading to cardiac hypertrophy (McMurray & Pfeffer, 2005). Therefore, heart diseases mainly comprise those disorders that chronically increase the cardiac workload, which over time will inevitably lead to the pathological ventricular remodeling, influenced by the hemodynamic load, neurohormonal activation and other factors under investigation (Cohn, Ferrari, & Sharpe, 2000). Due to the current heart disease prevalence and inadequate treatment options, there is an unmet medical need for the development of cost-effective healthcare innovations for the management of heart diseases.

At the cellular level, the network of cardiac transcription factors (TF) controls cardiac gene expression and has a central role in transcriptional regulation during the cardiac differentiation and development and during the adaptive pathophysiological processes in the adult heart. A specific set of evolutionally conserved cardiac transcription factors, GATA binding protein 4 (GATA4), NK2 homeobox 5 (NKX2-5), myocyte enhancer factor 2C (MEF2C), heart and neural crest derivatives expressed 2 (HAND2), serum response factor (SRF) and T-box 5 (TBX5), have been shown to interact with and orchestrate the cardiac gene expression during the differentiation and development and are also involved in cardiac hypertrophy in a context-dependent and dynamically evolving manner (Akazawa & Komuro, 2003; Horb & Thomsen, 1999; Lyons et al., 1995; Molkentin, Lin, Duncan, & Olson, 1997). In particular, the pioneer transcription factor GATA4

has emerged as the nuclear effector of several cardiac signaling pathways that modulate the key cardiac cascades through post-translational modifications and protein-protein interactions (Grepin, Nemer, & Nemer, 1997; Pikkarainen, Tokola, Kerkelä, & Ruskoaho, 2004).

To explore the possibilities to intervene in the function of GATA4 with small molecules, a discovery platform was established and the chemical agents possessing either agonistic or antagonistic effects on synergy arising from the protein-protein interaction of GATA4 were investigated. An extensive chemical screening project, integrating computational and experimental biology, was conducted to uncover the lead candidates acting on the cardiac transcriptional machinery. Moreover, the metabolic stability and chemical properties of the lead candidates responsible for the biological effects and molecular mechanisms at the cellular level were evaluated in detail. Further, the effects of the small molecules were characterized in a number of confirmatory *in vitro* bioassays and their bioactivity assessed *in vitro* in rat cardiac myocytes and *in vivo* in animal models of ischemic injury and angiotensin II-induced pressure overload.

2 Review of the literature

2.1 General overview of the drug discovery process

Over the last decades, the advancement of rational drug design has been a tremendous success story, mainly driven by the growing amount of scientific knowledge together with ever-increasing technical capabilities applied to overcoming the challenges of treating the human diseases. Despite this historic success, the pharmaceutical industry is currently facing the wide-ranging challenges due to declining research and development productivity, increasingly cost-effective healthcare systems and more demanding regulatory environments (Abou-Gharbia & Childers, 2014; Paul et al., 2010; Schuhmacher, Gassmann, & Hinder, 2016). During the last decade, a huge investment in the research and development of novel drugs has not fully achieved the expectations of pharma companies; the cost per launched drug was recently estimated to exceed $2.5 billion (DiMasi, Grabowski, & Hansen, 2016). As the "low-hanging fruit" is no longer available, the achievement of scientific and financial success in today's pharma environment may require an orientation towards the transformative innovations and more challenging target classes, such as TFs and adaptor proteins (Hagenbuchner & Ausserlechner, 2016; Yeh, Toniolo, & Frank, 2013).

In general, drug discovery and development processes are multidisciplinary challenges requiring the application of basic science disciplines, such as biology, chemistry and physics. Over the course of the drug development process, many basic science methodologies are repetitively integrated into development projects to address the scientific issues in question. During drug development, four phases of human clinical trials are conducted prior and after the submission of a new drug application to regulatory authorities for the evaluation of the dossier (van Norman, 2016). Marketing authorization for the new drug includes the documentation for drug safety and efficacy, an estimation of the drug´s benefits versus risks, the drug package information and the drug manufacturing and quality control protocols. In the drug innovation ecosystem, pharma collaborations with universities, public institutions and small to mid-size enterprises (SMEs) have an important role as a source of the novel innovations and specific know-how in niche scientific areas. From 2010-2012, half of the products granted with market approval in the EU were from the in-house efforts of pharma companies, and the remaining half were products based on the ideas originating from SMEs, academic institutions and other

public bodies (Lincker, Ziogas, Carr, Porta, & Eichler, 2014). The supporting role of academic research to provide novel ideas and entities, thereby fostering the drug development pipelines, ultimately enhances the human health and the overall success rate of the pharma industry. However, some intrinsic challenges remain to fully integrate the benefits arising from these fundamentally different research organizations. One of the most decisive issues concerns the reproducibility of scientific results. Based on a survey, the vast majority of scientists in both academic and pharma environments have failed to reproduce the published results (Baker, 2016). Additionally, this study estimated that the reproducibility of scientific results are highly dependent on the research field. At best, over 70% of survey respondents in the field of chemistry and physics expected that the most of published results can be trusted. However, another study in the field of cancer biology reveals the crisis of reproducibility, scientific findings were experimentally confirmed only in 11% of cases (Begley & Ellis, 2012). This creates major obstacles, especially for the pharma industry, to recognize and fully benefit from the scientific results obtained in academic research. As a part of the on-going discussions, the statistical methods and P-values used to evaluate the significance of the scientific results have been heavily criticized as leaving too much room for error (Nuzzo, 2014). The most provocative talks have suggested to lower the P-value from the current 0.05 threshold to 0.005 to improve the quality and reproducibility of academic results (Ioannidis, 2018).

Since the 1990s, the primary approach to drug research has involved target protein-based drug development, where a biologically relevant target protein is first selected and validated and is subsequently evaluated by screening with small molecules to identify the bioactive hit compounds (Fig. 1) (Brown, 2007; Sams-Dodd, 2005). Another general screening strategy has been phenotypic screening, where disease-indicating changes are followed in disease-relevant cells, all without the use of a preconceived target protein. Of the 259 agents that were approved for market between the years 1999-2008, 75 compounds were first-in-class drugs with new molecular mechanisms of action; of these, 50 (67%) were small molecules and 25 (33%) were biologics (Swinney & Anthony, 2011). Surprisingly, an analysis of the first-in-class small molecule subclass reveals that most of the approved drugs were originally found from phenotypic screening (56%), followed by the target-based approaches (34%) and synthetic or enzymatic modifications of natural substances (10%) (Swinney & Anthony, 2011). More recent reviews have reported a more balanced ratio between the target-based and phenotypic research approaches (Eder, Sedrani, & Wiesmann, 2014; Moffat, Rudolph, & Bailey, 2014).

Nevertheless, the outcomes indicate that truly novel entities and innovations affecting to human health are often found without a preconceived target protein, and mechanism of action. Therefore, although the scientific rationale is more complicated without the target protein information (e.g., lead optimization, toxicity testing and trial design), the Food and Drug Administration (FDA) regularly approves drugs without an understanding of their mechanism of action (Mullard, 2015).

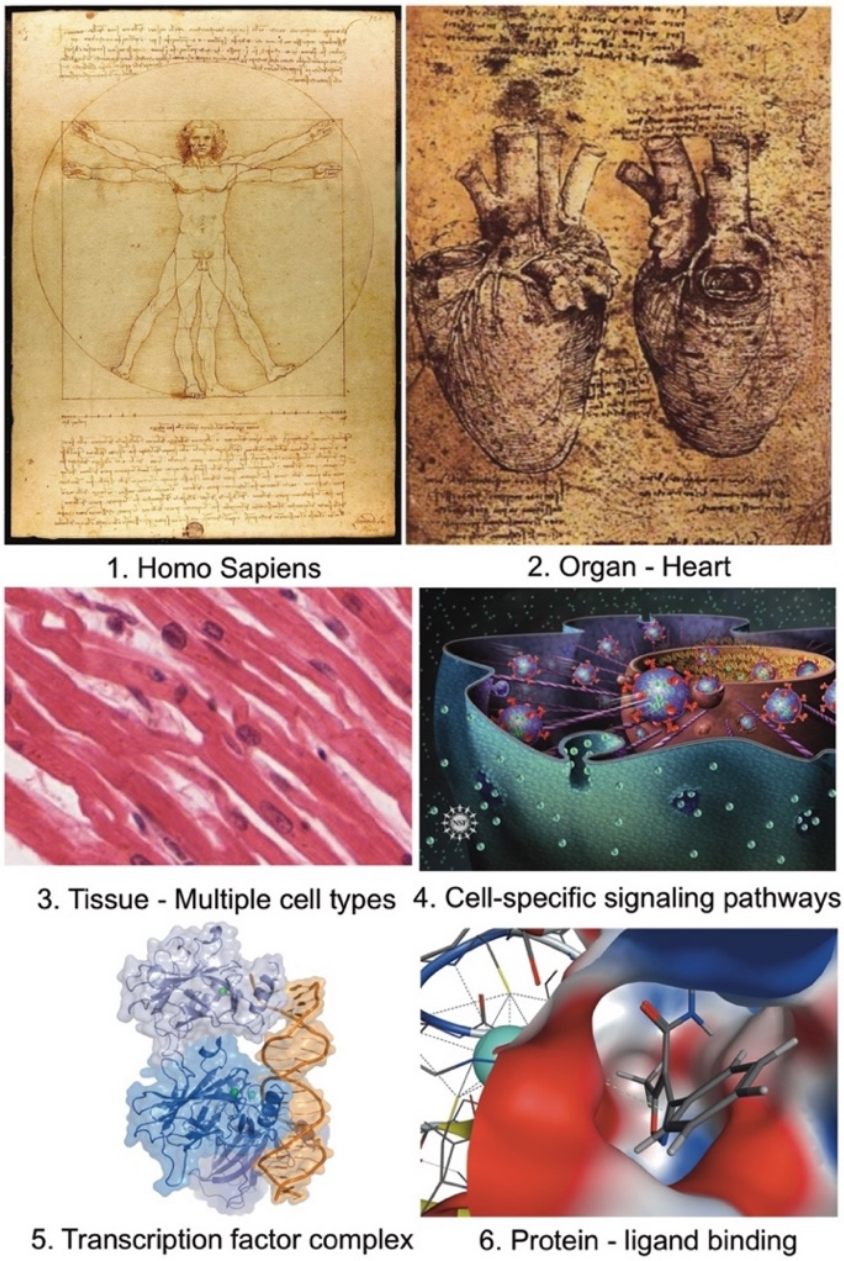

Fig. 1. A gradually evolving landscape of the transcription factor targeted drugs from *Homo sapiens* to atom-level resolution.

2.2 Ligand binding to biological targets

2.2.1 Ligand binding affinity

In all living organisms, biological macromolecules are interacting with each other and various small molecules. As a result of this molecular recognition process, molecular complexes are formed. The expected ligand binding affinity, or more specifically, drug binding affinity, to the target protein is a combination of the energy components linked to molecular structures and the dynamics of interacting partners, along with the solvent molecules (Bronowska, 2011; Du et al., 2016). Typically, during the primary screening for hit compounds, ligand binding affinities are in the low micromolar range, whereas the drug development and optimization process aims to develop the high-affinity ligands, preferably acting in low nanomolar to picomolar range. High-affinity drug candidates are typically favored due to their superior target protein selectivity, potentially leading to fewer adverse effects, smaller doses and lower manufacturing costs (Young & Leeson, 2018).

In thermodynamics, ligand binding to the target protein may occur only if the Gibbs free energy of binding (ΔG) reaches a negative value, as a summary of enthalpy and entropy factors, given that the temperature remains constant.

$$\Delta G = \Delta H - T\Delta S = -RT \ln K \qquad (1)$$

where ΔG is the change in free energy of binding, ΔH is the change in enthalpy, ΔS is the change in entropy, T is the temperature, R is the gas constant and K is the ligand binding constant. A more negative value of ΔG represents a higher binding affinity of the ligand to the target protein. In general, to achieve the maximal energy departure, the ligand conformation should display a perfect shape, size and chemical group complementarity and be in close proximity to the interface of its biological target. Enthalpy change (ΔH) is a net result of weak individual noncovalent molecular contributions between the ligand, protein and solvent, such as electrostatic forces, hydrogen bonding, halogen bonding and van der Waals forces. Upon ligand binding, noncovalent interactions are, at first, distorted between the protein cavity and solvent and between the ligand and solvent and are subsequently formed between the protein and ligand, together with the reorganization of the solvent network. Although noncovalent interactions between

the molecules are reversible by their nature, they orientate according to the specific distance and directional preferences to achieve the minimal global or local energy.

On the other hand, the change in entropy (ΔS) defines the measure of disorder in a system. Changes in molecule conformation (translational, rotational and torsional freedom) and de-solvation (solvent molecules are relocated and reorganized from the interface of the binding partners) are the main sources of changes in entropy (Chang, Chen, & Gilson, 2007). Upon binding, the flexibility of both ligand and protein are reduced, and the resulting loss in conformation entropy is counterbalanced by the attractive enthalpic forces driving the ligand binding. This phenomenon, also called as entropy-enthalpy compensation, has a central role in drug design and development, implicating the difficulty to predict and separate the effect of enthalpic and entropic proportions to the binding affinity of a compound (Chodera & Mobley, 2013; Reynolds & Holloway, 2011). However, the universal acceptance of enthalpy-entropy compensation remains debatable, since several studies have suggested the major contribution of water network reorganization to the thermodynamics of ligand binding (Breiten et al., 2013; Grunwald & Steel, 1995).

2.2.2 Ligand binding kinetics

Over the last decade, drug developers have realized the importance of time-dependent processes linked to ligand binding as selection criteria for the drug candidate optimization (Tummino & Copeland, 2008). An increasing body of evidence demonstrates that the binding kinetics of the drug-target binary complex, rather than ligand binding affinity, dictates the duration of the pharmacologic effect of drugs *in vivo* (Copeland, 2016; Schuetz et al., 2017). In general, *in vitro* experiments are closed equilibrium systems, where the target receptor is constantly exposed to a stable concentration of ligand, whereas during *in vivo* treatments, drug concentrations fluctuate depending on dosing due to drug absorption, distribution, metabolism, and excretion (ADME) (Copeland, Pompliano, & Meek, 2006). The lifetime of a ligand-receptor complex is described by its residence time, which is correlated to its disassociation rate constant (K_{off}).

$$t_R = 1/K_{off} \qquad (2)$$

where t_R is the residence time and K_{off} is the disassociation rate constant. In general, ligand-target association may occur via various molecular mechanisms, including one-step binding by replacing water, the induced fit mechanism, protein conformational selection, and irreversible inhibition. Many times, drug binding occurs so fast that the related association rate constant, K_{on}, is ignored; however, the compound with a slow K_{off} dominate the binding site occupancy and cannot be neglected when judging the quality of the lead candidate. The ligand off-rate is highly dependent on the conformational dynamics of target macromolecules, which affect drug binding and dissociation. Thus, high-affinity drug interactions with targets often involve multistep binding kinetics to achieve a tight binding. Overall, it is rather difficult to fully optimize the pharmacokinetics and adverse effect profiles of drugs without any understanding of the binding kinetics of the ligand interactions (Copeland, 2016).

2.2.3 Determination of affinity and kinetic parameters

Understanding of the biomolecular recognition processes of drugs in detail requires an experimental determination of the binding free energies and a subsequent analysis of structure related thermodynamic data. Isothermal titration calorimetry (ITC) is the most preferred method for measuring the quantitative thermodynamics of ligand binding (Holdgate & Ward, 2005; Perozzo, Folkers, & Scapozza, 2004). The ITC technique measures the heat exchange associated with ligand-protein binding at constant temperature, allowing the simultaneous determination of the ligand binding constant (K), Gibbs free energy of binding (ΔG), change in enthalpy (ΔH) and change in entropy (ΔS) (Freyer & Lewis, 2008; Perozzo et al., 2004). Experimental values for the individual enthalpy and entropy components for ligand binding are possible to determine down to nanomolar ligand affinities from a single ITC experiment. Combining the high-resolution ligand-protein structural data with the ITC-derived thermodynamic data delivers the information about the ligand binding mechanisms and enables a better starting point for rational drug design (Perozzo et al., 2004).

Surface plasmon resonance (SPR) is the most practical experimental method for determination of the ligand binding kinetics of drug candidates (Hahnefeld, Drewianka, & Herberg, 2004; Schuck, 1997). SPR provides an excellent experimental instrumentation (Biacore™) for a real-time and label-free investigation of ligand-protein interactions by using immobilized proteins that are exposed to study ligands with a continuous flow and the subsequent detection of

refractive indexes. SPR methods are able to determine separate association and dissociation rate constants for the ligands affinities in the millimolar to nanomolar range. Instead of solely concentrating on thermodynamics and the affinity of the drug binding, SPR delivers the binding kinetics of the compounds as an additional and complementary information tool for guiding lead optimization towards the optimal structural solutions. Hence, the residence time measurements are on the way to being fully integrated into the drug discovery process of most pharma/biotech companies (Lu & Tonge, 2010; Schuetz et al., 2017).

2.2.4 Computational methods

High-level quantum mechanical (QM) methods are providing the most detailed and accurate descriptions of molecular structures, dynamics, and functions *in silico* by solving the Schrödinger equation (Gohlke & Klebe, 2002; Raha et al., 2007). However, many times chemical systems are too complex for QM methods to resolve, therefore, classical molecular mechanics approaches are more commonly used. Monte Carlo (MC) and molecular dynamics (MD) methods are computational simulations used for studying the large biological systems and their most probable conformations and dynamical trajectories. These methods rely on the atomistic force field parametrization, Newton's law of motion and approximations, where the motion of an atom is equal to the motion of an atom nucleus (Jorgensen & Tirado-Rives, 1996).

Atomistic and coarse-grained MC simulations are the mathematical predictions for the most probable random outcome based on atom parametrization, repeatable random sampling and the subsequent statistical analysis of the results (Donnet, 2012; M. Liu & Wang, 1999). Rather than trying to reproduce the dynamics of the system, the MC approach depends on repeated sampling and the generation of random states in conformational space to achieve the equilibriums that are the most probable, based on the Boltzmann distribution. MC simulations efficiently produce randomized conformations of biosystems, but due to the time-independent nature of the method, separate conformations are not correlated to each other, and therefore, do not create true trajectories of bio-conformations. MD simulations are the most popular methods to study the conformational rearrangements of biomolecules and their interactions with ligands and explicit solvent systems (Borhani & Shaw, 2012; McCammon, Gelin, & Karplus, 1977). The method provides a dynamic description of the time-dependent behavior of atoms and molecular systems by solving the equations of motion for the atoms with an

empirical potential energy function (force field). In comparison to MC methods, MD methods implement the time frame for the simulation, which results in a consistent trajectory of the system over a time window from ten to hundreds of nanoseconds.

Biomolecular systems involving the protein structure, ligand and solvent atoms are in a constant state of motion and may contain up to 1 000 000 atoms in total (Borhani & Shaw, 2012). In this context, MC and MD simulations are computationally intensive and relatively slow for the study of several ligand derivatives. Free-energy perturbation (FEP) methods are used for the accurate prediction of the relative binding free energy differences between two stages and describe the protein-ligand associations from the MD and MC simulations (L. Wang, Wu, et al., 2015; L. Wang, Berne, & Friesner, 2012). In contrast to MD, FEP methods are computationally less intensive due to the restrictions of atom motions in the periphery of the biomolecular system, and therefore, enable the analysis of several compounds per day. They are typically based on restricted MD simulations of the receptor-ligand complex and are therefore intermediate in both the accuracy and computational efforts. Optimally, FEP simulations are applied to a series of structurally similar compounds to preserve the accuracy and maintain the reasonable conformational changes in biomolecular systems. FEP with explicit solvent models has demonstrated the encouraging results in the optimization of pharmaceutical compounds and holds the promise to become a mainstream tool in rational medicinal chemistry (Williams-Noonan, Yuriev, & Chalmers, 2018).

The molecular mechanics energies combined with the Poisson-Boltzmann or generalized Born and surface area continuum solvation (MM/PBSA and MM/GBSA) methods estimate the free energy of the binding of small ligands to biological macromolecules (Shirts, Mobley, & Brown, 2010; C. Wang, Greene, Xiao, Qi, & Luo, 2018). Molecular settings for MM/PBSA and MM/GBSA approaches are relatively simple owing the modular components, implicit solvent model and absence of a training set. Due to the reduced number of atoms in simulation, these methods are efficient but not particularly accurate in comparison to FEP and MD. However, typically, the accuracy of MM/PBSA and MM/GBSA methods is better than docking and scoring methods (Genheden & Ryde, 2015). The total estimation of the binding energy is a combination of six terms; three terms are standard MM energy terms from bonded (bond, angle and dihedral), electrostatic and van der Waals interactions. Two terms are from the polar and non-polar contributions to the solvation free energies. The last term is an estimation of entropy (Genheden & Ryde, 2015).

The most widely applied computational approaches in drug design, especially in the beginning of the research project, are the docking and scoring methods, e.g., for virtual screening (Kitchen, Decornez, Furr, & Bajorath, 2004; Leach, Shoichet, & Peishoff, 2006). These methods are able to effectively screen and rank a large conformation library of small molecules against a selected rigid protein target. Docking programs use placement methods for predicting a compound's binding mode and subsequent scoring calculations for the estimation of the relative binding affinity. The accuracy of the force field has a crucial effect on the accuracy of the virtual screening and biomolecular simulations, since the majority of methods rely on those atomistic parameters. Therefore, the continuous development of force field parameters with more diverse and better descriptors is driving progress for more accurate molecular screening and simulations. The most common force fields used for biomolecular simulations are CHARMM, AMBER, OPLS and GROMOS (Lopes, Guvench, & MacKerell, 2014).

2.3 High-throughput screening

High-throughput screening (HTS) is an enabling technology in drug discovery that utilizes validated automation and large data set processing to quickly assess a compound´s activity against a selected target protein (Broach & Thorner, 1996). After the initial protein target identification and validation, small molecule lead discovery and development programs may involve two main conceptual subprocesses (Hughes, Rees, Kalindjian, & Philpott, 2011; Keseru & Makara, 2006). The first stage is a primary compound screening that contains an *in vitro* assay that, in the case of HTS, may evaluate up to a couple million library compounds to identify the initial hit compounds. Screened compounds are either purchased from commercial vendors, or preferably, selected from in-house compound collections. The assay development, screening campaign and analysis of the results in the drug discovery phase may typically take several years. The second stage of the development process is the lead optimization, carried out through iterative synthesis and testing phases, ultimately leading to the selection of a candidate drug for clinical trials. During the drug development process, lead optimization aims to improve the ligand binding affinity and target selectivity via structural modifications of the lead candidate. In addition to target efficiency, the pharmacokinetic and safety properties of the candidate compounds are taken into consideration during the development process (Davis, Plowright, & Valeur, 2017).

Industry standard HTS settings, with <600 Dalton (Da) compounds, opens up the infinite theoretical chemical space, containing approximately 10^{60} drug-like compounds (Kirkpatrick & Ellis, 2004). This theoretical number of structurally diverse compounds is far beyond the current possibilities and experimental capacity of humans. Presently, the most comprehensive register of experimental organic and inorganic compounds, Chemical Abstract Service (CAS), contains over 142 million unique chemicals, of which approximately half (70 million compounds) are drug-like structures. Therefore, low coverage of the drug-like chemical space and redundancy of experimental library compounds, together with the high costs of cell-based HTS, are major fundamental shortcomings of successful high-throughput screening attempts (MacArron et al., 2011).

Cell-based HTS campaigns are extensively applied in the majority of compound screening programs carried out in the pharma industry. Cellular HTS assays include three main types: second messenger assays, reporter gene assays, and cell proliferation assays (An & Tolliday, 2010). In this context, the luciferase reporter assay represents a target-protein based screen for the identification of particular signaling pathway-intervening compounds. Luciferase reporter assays are widely used, since a specific target protein may markedly accelerate the early phase drug development (W. Xie et al., 2016; Zang, Li, Tang, Wang, & Yang, 2012).

2.4 Fragment-based drug discovery

Fragment-based drug discovery (FBDD) is a research method that utilizes small fragment compounds to identify initial active hit compounds against selected protein targets (Hajduk & Greer, 2007). A validated biophysical assay is classically needed for the detection of the target-bound fragments, involving nuclear magnetic resonance spectroscopy (NMR spectroscopy), surface plasmon resonance (SPR) and X-ray crystallography. Structural determinations for ligand-protein binding conformations are optimally resolved with NMR and X-ray techniques, since NMR chemical shifts and crystal structures are sensitive to changes in ligand binding (Erlanson, Fesik, Hubbard, Jahnke, & Jhoti, 2016; Hajduk & Greer, 2007). Interestingly, introductory studies for the usefulness of weak binding low molecular mass and low-affinity ligands to target proteins for drug discovery were originally called as structure-activity relationships (SAR) by NMR (Hajduk & Greer, 2007).

Small chemical probe-related molecular factors, such as chemical tractability, structure-activity relationships and experimental structural binding data, are of great importance for advancing the chemical biology fragment discovery

approaches. By definition, fragments are organic compounds with a primary limitations in their molecular size, and a secondary limitation in their chemical functionality and stereochemical diversity (Erlanson, 2011). Limiting the small chemical probes to molecular sizes below 300 Dalton (Da) exponentially decreases the number of possible structural derivatives (Erlanson, McDowell, & O'Brien, 2004; Hajduk, 2006). The common molecule description for a fragment category comprises the following criteria: molecular weight <300, number of hydrogen bond donors ≤3, number of hydrogen bond acceptors ≤3 and calculated log octanol/water partition coefficient (clogP) ≤3 (Congreve, Carr, Murray, & Jhoti, 2003).

Since organic compounds are synthesized in a modular manner, the utilization of redundant fragment modules offers a built-in efficiency for screening campaigns in drug discovery. Fragments, in comparison to HTS compounds, produce an efficient research approach to evaluate a much broader chemical space with a minimal number of experiments and compounds (Hajduk & Greer, 2007; Hajduk, Huth, & Tse, 2005). In addition to these previous advantages, the binding of small chemical probes greatly reduces the complexity of the protein-ligand binding due to diminished entropic requirements and a significant emphasis on the enthalpic preferences of the molecular system (Williams, Ferenczy, Ulander, & Keserű, 2017). Use of fragments in the discovery process is particularly beneficial in challenging target classes, such as protein-protein interactions and biological hypotheses not otherwise testable, where *de novo* design is often necessary due to the lack of reference compounds. Studies exploring the global protein druggability have estimated that a surprisingly high number of targets in the protein-protein interaction category have the potential for a drug-binding cavity (29% of proteins) (Hajduk et al., 2005).

In contrast, the initial fragment hits are usually weak and non-selective micromolar binders, and even though they require substantial medicinal chemistry efforts, the benefits of the approach are evident. Structure optimization from small fragments towards the higher binding ligand depends on the rational tradeoff of molecule size versus potency. Retrospective structural analysis of the minimal binding elements of active compounds confirms a nearly linear correlation between molecular weight and binding affinity over the entire range of molecule derivatives (Hajduk, 2006). To date, approximately 40 drug candidates have been successfully developed by fragment-based approaches to enter clinical trials. These breakthrough drug discovery projects further demonstrate how chemical probes can help to bridge the gap between biological research and the development of

medicines and highlights the need for innovative approaches in therapeutic discovery (Erlanson et al., 2016; Schreiber et al., 2015).

However, in some cases, the generation of experimental ligand-protein binding structures remains challenging, and other options such as cell-based phenotypic screens have been used to overcome this deficiency. In phenotypic screens, disease-relevant cells are treated with compounds to induce disease-relevant changes in the absence of target information. Validation of ligand binding in phenotypic screening is complicated, especially in the case of fragments, due to binding to numerous proteins in cells. Various technologies have been applied for the characterization of the direct and indirect targets of bioactive compounds following the phenotypic screening (Baker, 2017; Mullard, 2015; Schirle & Jenkins, 2016). More detailed target protein analyses have revealed that the application of fragment-based approaches in phenotypic screens have the strong ability to target the proteins previously estimated undruggable, especially in the protein classes such as TFs, uncategorized proteins and regulators (Baker, 2017). These findings may shed light on the opportunities for molecule interventions in the field of transcriptional regulation.

2.5 Transcription factors

Transcription factors are proteins which operate in the nucleus in a coordinated manner to recruit the RNA polymerase to the transcription start site and to control the transcription of genetic information from DNA to messenger RNA (Levine & Tjian, 2003; Spitz & Furlong, 2012). By definition, the TFs have at least one DNA binding domain that interacts with the specific *cis*-regulatory sequence at the gene promoter or enhancer region and regulates the expression of their target gene. There are two mechanistically different operating classes of TFs. General TFs are the ubiquitous members of the transcription pre-initiation complexes, e.g., TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, whereas the vast majority of TFs target the recognition sequences present upstream of the transcription start site (Table 1).

Zinc fingers are the most abundant transcription protein class in humans, comprising over 700 of the approximately 2000 putative human TFs (Brivanlou & Darnell, 2002; Weirauch & Hughes, 2011). Typically, structure- and sequence-related diversification of zinc finger TF functions are driven by alterations in DNA sequence specificity, protein–protein interactions and the expression levels of the TF encoding genes (Latchman, 1997; Todeschini, Georges, & Veitia, 2014). The molecular arrangement of zinc finger domain-containing TFs exhibits diverse

structural organizations, namely, Cys_2His_2 (e.g., TFIIIA), Cys_4 (e.g., glucocorticoid receptor) and Cys_6 (e.g., GAL4). In biology, zinc finger proteins are involved in a wide range of different cellular functions, such as replication and repair, transcription and translation, metabolism and signaling, cell proliferation and apoptosis (Cassandri et al., 2017; Krishna, Majumdar, & Grishin, 2003).

Table 1. Summary of the evolutionarily conserved transcription factors expressed in heart.

Transcription factor	Size	Isoforms	Structural classification	Interacting partners
GATA4	442 aa	2	Zinc finger protein	NKX2-5, TBX5, MEF2C
NKX2-5	324 aa	3	Homeobox protein	GATA4, TBX5
MEF2C	473 aa	6	MADS-box superfamily	GATA4, p300
TBX5	518 aa	3	T-box protein	GATA4, NKX2-5
SRF	508 aa	1	MADS-box superfamily	GATA4, ELK4, Myogenin
HAND2	217 aa	2	Basic helix-loop-helix protein	GATA4, NKX2-5
FOG2	1151 aa	3	Zinc finger protein	GATA4

2.6 GATA4

2.6.1 Protein sequence and structure

In humans, the evolutionarily conserved GATA-family of proteins consists of six GATA proteins (GATA1-6), all sharing the similar tertiary protein structure and high amino acid sequence identity over their two DNA binding zinc finger domains. Both N-terminal- and C-terminal zinc fingers in the GATA family are tetrahedrally coordinated and bound to four cysteine residues (Cys_4) to form the protein domains, involving two β-sheets and one α-helix. In comparison to the human and animal genomes, the GATA transcription factor families are comparatively large in plant model organisms, with approximately 30 members in *Arabidopsis thaliana* and 64 members in soya beans. Furthermore, a recent study in plants suggests that two

important processes during the plant development, greening and photosynthesis, as well as stomata formation, and thus, gas exchange, are regulated by plant GATA-factors (Bastakis, Hedtke, Klermund, Grimm, & Schwechheimer, 2018; Behringer & Schwechheimer, 2015).

The protein sequence of human GATA4 contains multiple functional domains, including the C- and N-terminal zinc fingers, in addition to the N-terminal and C-terminal sequences, which have been suggested to constitute transcriptional activation and nuclear localization domains, respectively (Molkentin, 2000; Morrisey, Ip, Tang, & Parmacek, 1997). The reduced protein structure of the C-terminal zinc finger of GATA4 has been experimentally resolved with NMR (Protein Data Bank (PDB) code 2M9W, 2013) by the Northeast Structural Genomics Consortium (Fig. 2). However, the first NMR-structures of the zinc finger domain for GATA1 (PDB code 1GAT) were published in 1993 (Omichinski et al., 1993). Furthermore, X-ray crystallographic binding analyses of other GATA zinc fingers bound to DNA have provided new insights into the DNA recognition mechanisms of GATA-dependent gene regulation (Bates, Chen, Kim, Guo, & Chen, 2008).

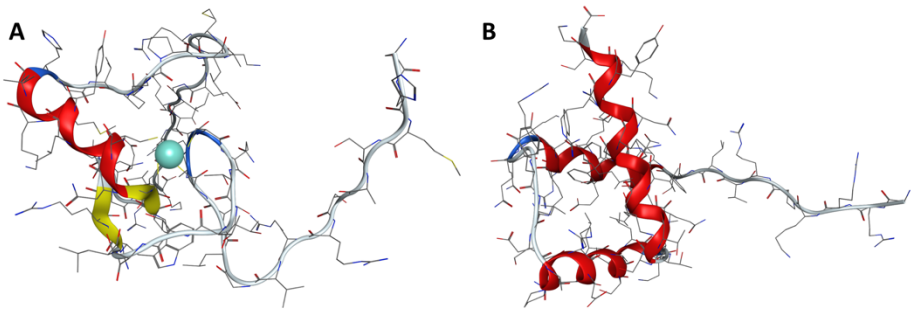

Fig. 2. (A) Molecular structure and folding of C-terminal zinc finger of GATA4 determined by NMR spectroscopy. Study was carried out by Northeast Structural Genomics Consortium (PDB code; 2M9W). The zinc atom is illustrated as a green sphere. (B) Molecular structure for homeodomain of NKX2-5 determined by X-ray crystallography (PDB code; 3RKQ). The secondary structures are illustrated in red (*α*-helix) and yellow (*ß*-sheets).

The human GATA4 protein contains 442 amino acids and includes two structurally stable zinc finger domains, located at amino acid residues 217-241 and 271-295. The protein sequence outside of the zinc finger core domains and C-terminal

extension (residues 210-320) has no globular structure, and it remains completely disordered (Mattapally et al., 2018). In addition, the C-terminal extension of the zinc finger contains the amino acid sequence required for the nuclear localization (Morrisey et al., 1997). Several GATA4 single point mutations identified from humans have been shown to be linked to common developmental anomalies and mortality in newborns. For example, the heterozygous G296S missense mutation of GATA4 results in diminished DNA binding affinity, diminished transcriptional activity and abolishes a physical interaction between GATA4 and TBX5 that is associated with congenital heart diseases (Ang et al., 2016; Garg et al., 2003). Moreover, four heterozygous missense GATA4 mutations, P36S, H190R, S262A, and V399G, have been linked to congenital atrial septal defects in newborns and are responsible for substantial morbidity and mortality in affected individuals (Y.-Q. Yang et al., 2013).

2.6.2 Expression

The GATA4 protein was originally discovered as the one of the earliest molecular markers associated with the initiation of cardiac gene expression (Arceci, King, Simon, Orkin, & Wilson, 1993; Heikinheimo, Scandrett, & Wilson, 1994; Kelley, Blumberg, Zon, & Evans, 1993). In addition to heart, GATA4/5/6 proteins are expressed in various tissues, including liver, lung, gut and gonad (Arceci et al., 1993; Kelley et al., 1993; Laverriere et al., 1994; Morrisey, Ip, Lu, & Parmacek, 1996; Morrisey et al., 1997). The other GATA family proteins, GATA1/2/3, are preferentially expressed in hematopoietic cells (Orkin, 1992). Knock-out studies of GATA4/5/6 proteins during embryonic heart development in Xenopus, zebrafish and mice propose a functional redundancy between these TFs. Moreover, a number of studies have demonstrated that none of the GATA factors are absolutely required for the specification of myocardium, suggesting the compensatory mechanism inside the GATA family (Peterkin, Gibson, & Patient, 2007; Shu et al., 2015; Singh et al., 2010). However, cardiac-specific deletion of GATA4 protein from adult mice resulted in severely compromised basal gene expression, lowered survival of cardiac myocytes, and a hypertrophy response following exercise stimulation (Oka et al., 2006). Additionally, conditional inactivation experiments with GATA4 in mice decreased cardiomyocyte proliferation in the right ventricle, and thus, demonstrates the active role of GATA4 in cell cycle control (Rojas et al., 2008).

One the one hand, it appears that total control of the GATA4 mRNA levels is not a critical predictor of GATA4 activity in hypertrophy model systems. Over the

course of experiments, mRNA levels of GATA4 remained stable in response to arginine-8-vasopressin infusion, nephrectomy *in vivo* and treatment with endothelin-1 (ET-1) *in vitro* (Hautala et al., 2001; Kerkelä, Pikkarainen, Majalahti-Palviainen, Tokola, & Ruskoaho, 2002; Marttila et al., 2001), whereas exposure to phenylephrine (PE), isoproterenol or cardiomyocyte stretch *in vitro* were able to increase the GATA4 mRNA levels (Pikkarainen et al., 2003; Saadane, Alpert, & Chalifour, 1999). On the other hand, treatment with cardiotoxic anthracyclines were associated with myocyte apoptosis and a reduction in both mRNA and protein levels of GATA4 (Suzuki & Evans, 2004). Overall, the mRNA results indicate that activity of myocardial GATA4 protein is preferably controlled by post-transcriptional and post-translational processes (Pikkarainen et al., 2004).

The GATA family of proteins has shown high variance in their cellular protein stability and degradation rate. GATA2 protein has a relatively short half-life (approximately 30 minutes) in comparison to GATA3 and GATA6, expressing half-lives over three hours when studied with cycloheximide, a protein synthesis inhibitor (Izzo et al., 2014; Minegishi, Suzuki, Kawatani, Shimizu, & Yamamoto, 2005; Y. Xie et al., 2015). However, the half-life of GATA1 and GATA4 protein far exceeds the other members of the GATA-family, with a half-life of greater than 6 hours (T. Li, Zhang, Jiang, Liu, & Liu, 2018; Lurie, Boyer, Grass, & Bresnick, 2007; Minegishi et al., 2005). In general, the protein degradation rate plays an important role in protein displacement from the chromatin, especially in the case of the related proteins, and therefore, has a major impact on the establishment of transcription networks that control the gene expressions (Lurie et al., 2007).

Adult human heart has insufficient capacity to repair or regenerate cardiac cells after injury when a significant number of cardiomyocytes are lost. Scar formation and failure to regenerate the injured myocardium are the primary causes for the development of heart failure, arrhythmias and sudden death (Talman & Ruskoaho, 2016). Signaling pathways and regulatory mechanisms that are active during embryogenesis and are involved in heart growth and development may be used to repair the injured adult heart (M. Xin, Olson, & Bassel-Duby, 2013). The overexpression of cardiac GATA4 protein shows a promise to preserve the cardiac function after cardiac injury by promoting increased angiogenesis and reduced fibrosis (J.-G. He et al., 2018; Mathison et al., 2017; Rysä et al., 2010). Moreover, genetic enhancement of GATA4 protein was able to prevent cardiomyocyte apoptosis and drug-induced cardiotoxicity (Aries, Paradis, Lefebvre, Schwartz, & Nemer, 2004).

A study by Malek Mohammadi et al. (2017) demonstrated that high cardiac abundance of the cardiac GATA4 by adenoviral gene transfer at postnatal days 1-7 markedly improved cardiac regeneration after cryoinjury and rescued the loss of regenerative capacity. Accordingly, larger myocardial scars were observed in mice with cardiomyocyte-specific GATA4 knockout after cryoinjury, indicative of reduced cardiac regeneration, accompanied by reduced cardiomyocyte proliferation and reduced myocardial angiogenesis in GATA4 knockout mice (Malek Mohammadi et al., 2017; Oka et al., 2006). In addition, molecular mechanisms of active cell populations responsible for the regenerative capacity of zebrafish have been explored and linked to the triggered expression of the embryonic GATA4 within a week after cardiac injury. These results in zebrafish suggest the primary contribution and association of GATA4-positive cells to heart regeneration and repair (Kikuchi et al., 2010).

2.6.3 DNA binding and chromatin occupancy

Transcription factors regulate gene expression through the co-protein assemblies together with basal transcriptional machinery, by binding to specific *cis*-regulatory sequences in gene promoters and enhancers. The tissue-specific transcription factor GATA4 prefers to bind to the DNA sequence (A/T)GATA(A/G) through its carboxy-terminal zinc finger and is responsible for mediating the site-specific physical interaction with the DNA sequence (Pikkarainen et al., 2004). A number of essential cardiac-expressed genes contains the binding sequence for GATA-factor in their promoter, including atrial natriuretic peptide (ANP) (Liang & Molkentin, 2002), B-type natriuretic factor (BNP) (Thuerauf, Hanford, & Glembotski, 1994), α-myosin heavy chain (Molkentin, Kalvakolanu, & Markham, 1994), β-myosin heavy chain (Hasegawa, Lee, Jobe, Markham, & Kitsis, 1997), cardiac troponin C (Ip et al., 1994), cardiac troponin I (Murphy, Thompson, Peng, & Jones, 1997), and sodium-calcium exchanger (Nicholas & Philipson, 1999).

Cooperative TF binding to the target sequences in chromatin may be sufficient for the activation of gene expression. However, to initiate TF binding to the inaccessible areas where the chromatin is closed and the majority of the potential binding sites are unoccupied, special pioneer TFs are required (Zaret & Carroll, 2011). GATA4 is a pioneer TF and is able to initiate (as opposed to a cooperative process) the sequential process where a single TF is capable of inducing the gene activation procedure. The capacity of pioneer TFs to vary their function in a chromatin acceptability-dependent manner leads to multiple operational benefits in

gene regulation. First, the initial binding of pioneer factors may passively reduce the number of co-factors needed to activate the gene expression. Second, the pioneer factors are shown to physically bind to the genome prior to the period of activation, and therefore, pioneer TF may locally open up the chromatin and prime it to be reachable for the other TFs (Zaret & Carroll, 2011).

Genome-wide analyses have observed the close association between the transcription factor DNA binding and the effects in target gene activity (Johnson, Mortazavi, Myers, & Wold, 2007; MacQuarrie, Fong, Morse, & Tapscott, 2011). It has been shown that in regions of open chromatin, the specific recognition sequence is a reliable binding predictor of TFs, and in those cases, protein interactions play a minimal role (Kaplan et al., 2011). However, genome-wide analyses of TF binding locations in the genome have previously suggested that an extremely small fraction of consensus target sites are actually occupied (A. Yang et al., 2006; Zaret & Carroll, 2011). For example, the chromatin immunoprecipitation sequencing (ChIP-seq) evaluation of estrogen receptor-α binding to corresponding the estrogen response element (ERE) motif sequence identified 10 000–16 000 high confidence estrogen-bound regions, barely equivalent to 2% out of the predicted ERE binding sequences in the genome (Joseph et al., 2010; Welboren et al., 2009).

In the case of GATA4 in adult heart, whole-genome ChIP-seq analysis with GATA4 antibody identified only 1 756 GATA4-bound regions (van den Boogaard et al., 2012), while bioChIP-seq in adult heart ventricles identified more than 15 000 binding sites for the high-affinity FLAG-biotin incorporated into GATA4 (A. He et al., 2014), indicating a major difference in detection sensitivity related to the antibodies used in the experiment. During the cardiac development, a high-affinity bioChIP-seq system identified over 50 000 GATA4-bound regions from the fetal heart ventricles. However, the less sensitive GATA4 antibody-based chromatin immunoprecipitation ChIP-seq identified 11 915 GATA4-bound regions (A. He et al., 2014). Overall, the ChIP-seq experiments indicate the dynamic change of GATA4 chromatin occupancy through normal heart development, in concert with its changing function. In the fetal heart, GATA4-bound regions were predominantly located distal from the transcription start sites, while in adult heart, a significant shift of GATA4 regions to the proximal locations were observed (A. He et al., 2014). In adult heart, pathological stress, such as chronic pressure overload, induced changes in GATA4 chromatin occupancy. Main stress-induced differences of GATA4 recruitment were associated with completely new disease enhancers that were not occupied during the development, as well as the partial revival of the

developmental program through GATA4 binding to a subset of fetal GATA4 enhancers (A. He et al., 2014).

Chromatin remodeling controls gene expression by modifying the access of regulatory transcription machinery proteins to the condensed genomic DNA. This genome-wide remodeling process occurs via two different mechanisms, either by covalent histone modifications or by moving, ejecting or restructuring the nucleosomes. Since gene activation is regulated in a multifaceted manner by the interplay of the TF network and the dynamic modifications of the chromatin landscape, as well as by the interference of microRNA (miRNAs), GATA4 chromatin occupancy alone was not directly associated to the increased cardiac gene expression levels in fetal or adult heart (A. He et al., 2014). However, a number of studies have revealed a high correlation of the genome-wide enrichment of GATA4 binding regions, particularly to acetylated histone H3 at lysine 27 (H3K27ac), a major active transcriptional enhancer marker, together leading to a strong combined effect on gene activation (Ang et al., 2016; A. He et al., 2014; Iyer et al., 2018; Tsankov et al., 2015). Indeed, the binding strength of GATA4, as such, did not correlate with the level of GATA4 target gene transcription assessed by ChIP-seq, whereas the increased expression of GATA4-bound genes were associated with higher H3K27ac enrichment at the GATA4-bound regions (A. He et al., 2014).

In human and mouse, there are approximately 2 000 TFs, more than 100 different modifications of histone residues, and approximately 700 miRNAs that modulate the mRNA profiles corresponding to approximately 20 000 genes. The TF complexes that are associated with GATA4 have a comparable dependency on co-factor binding and modulation by histone modifications, as well as on regulation by miRNAs, and therefore, tissue-specific chromatin co-occurrence with distinct subsets of TFs are preferred to allow a logical and systematic initiation/repression of the transcription. Distinct cardiac TFs, such as NKX2-5, TBX5, SRF and MEF2A, in addition to enhancers such as p300, have been shown to localize together with GATA4 at the chromatin regions and collaborate to direct cardiac gene expression (A. He, Kong, Ma, & Pu, 2011; Schlesinger et al., 2011). Even through these TFs are expressed in multiple tissues, ChIP-seq experiments provide unbiased support for the collaborative TF interactions in driving cardiac-specific gene expression, which is especially linked to the combinatorial localization and interactions between these cardiac TFs (A. He et al., 2011; Schlesinger et al., 2011).

2.6.4 Post-translational modifications

The function of GATA4 protein is modified by enzymes through post-translational processes, where one or more functional groups are covalently attached or detached, to or from, the protein structure. Previous studies show that the post-translational modifications of GATA4 involve the assignment of acetyl-, phosphoryl-, sumo- and ubiquitin moieties (Charron et al., 2001; Pikkarainen et al., 2004; J. Wang, Feng, & Schwartz, 2004; Yanazume et al., 2003). In cells, post-translational modifications have an impact on several different functions of GATA4, involving nuclear localization, DNA binding affinity, co-protein association, and protein degradation.

The sequence of GATA4 protein conveys seven potential phosphorylation sites that are modified by enzymes, such as glycogen synthase kinase 3β (GSK-3β), extracellular signal-regulated kinases (ERK) and p38 mitogen-activated protein kinase (MAPK), extensively reviewed by Suzuki (2011) and Zhou et al. (2012). In response to hypertrophic stimuli (e.g., ET-1, PE, isoproterenol and myocyte stretch), activation of the MAPK kinase signalling cascade significantly augments GATA4 phosphorylation and DNA binding efficiency (Tenhunen et al., 2004). The importance of phosphorylation was further evaluated by *in vivo* experiments with knock-in mice carrying the homozygous GATA4-S105A mutation, which demonstrated the compromised stress response of the myocardium (van Berlo, Elrod, Aronow, Pu, & Molkentin, 2011). In addition, earlier studies have shown that GATA4 phosphorylation via the MEK/ERK pathway at Ser105 gives the tendency to be more resistant to cellular degradation (Suzuki, 2003). In contrast, phosphorylation of the amino-terminal part of GATA4 via activation of GSK-3β resulted in the increased export of GATA4 from the nucleus (Morisco et al., 2001).

Histone acetyltransferases such as p300 and cyclic adenosine monophosphate response element binding protein (CREB) have been shown to induce the acetylation of specific lysine residues through physically interacting with GATA4 (Yanazume et al., 2003; Zhou, He, & Pu, 2012). Analogous to phosphorylation, GATA4 acetylation is similarly recognized as an imperative stimulus-triggered mechanism that regulates cardiac hypertrophy by enhancing its DNA binding efficiency and transcriptional activity (Yanazume et al., 2003). Mutational analysis through alanine scanning by Takaya et al. (2008) identified four lysine residues (K311, K318, K320, K322) as targets of acetylation by p300. Mutation of all four residues blocked GATA4 acetylation and blunted cardiac hypertrophy induced by GATA4 overexpression, thus demonstrating the importance of GATA4 acetylation in the regulation of GATA4 transcriptional activity (Takaya et al., 2008; Zhou et al.,

2012). A recent study identified K311 (corresponding to K313 in the paper) as a primary target of acetyltransferases p300/CREB-binding protein, with an enhanced cellular stability of acetylated GATA4 (You, Song, & Wang, 2018). The study was carefully conducted to simulate the effect of loss-of-function by using lysine to arginine mutations for the optimal structural integrity of the mutated proteins. Other studies have also reported that acetylated GATA4 is more resistant to degradation, perhaps due to competition with lysine ubiquitination (Suzuki, Nagase, Day, & Das, 2004). Furthermore, a pharmacological study with trichostatin A (TSA) demonstrated that acetylation of both GATA4 and histone residues are involved in the differentiation of embryonic stem cells (ESC) into cardiac myocytes (Kawamura et al., 2005).

GATA4 has been identified as a target protein for SUMOylation by small ubiquitin-like modifier-1 (SUMO-1) and ubiquitination by the ubiquitin-proteasome pathway (T. Li et al., 2018; J. Wang et al., 2004). Unlike the activation of the ubiquitin-proteasome pathway leading to protein degradation, SUMOylation enhances GATA4 transcriptional activity through covalent binding of the SUMO motif to exclusively to Lys366. In the cardiac context, the presence of E3 SUMO-protein ligase PIAS1 and SUMO-1 proteins triggered the enhanced SUMOylation of GATA4 and impacted both nuclear localization and cardiac gene activity (J. Wang et al., 2004). Active ubiquitination of GATA4 has been demonstrated in several physiological conditions, e.g., hypoxia, hyperglycemia and oxidative stress (Hae Jun et al., 2013; Kobayashi et al., 2007; T. Li et al., 2018). Based on these observations, it appears that the ubiquitin-proteasome pathway is the major degradation mechanism to regulate the cellular turnover of GATA4 protein.

The study by Aries et al. (2014) demonstrated a specific case of the cellular effects of truncated GATA4 protein. Activation of caspase-1 in cardiomyocytes by doxorubicin led to dominant-negative GATA4 protein with a reduced ability to activate cardiac genes (Aries et al., 2014). Furthermore, it was shown that inhibition of caspase-1 preserved the transcriptional activity, reduced GATA4 protein degradation and reduced myocyte cell death after doxorubicin exposure.

2.6.5 GATA4-targeted small molecule interventions

There is currently only one study that presents research of direct GATA4-targeted small molecule compounds in the literature. A study by El-Hachem and Nemer (2011) utilized the integration of *in silico* and *in vitro* cell-based screening assays to uncover the charged small molecules that selectively and efficiently inhibited the

DNA binding of GATA4 (El-Hachem & Nemer, 2011). Active compounds were identified through a virtual screening campaign with the ZINC 8 database from the National Cancer Institute (NCI). Two corresponding regions of the C-terminal zinc finger domain of the NMR structure of chicken GATA1 and the crystal structure of murine GATA3 were utilized as highly conserved structural templates for the virtual screening campaign. Resolved protein structures are useful templates, since all members of the GATA family proteins share the identical amino acid composition in the interface area responsible for DNA binding.

The study has identified four compounds that were able to inhibit GATA4 binding to DNA, and *in vitro* they blocked the activation of GATA4 downstream target genes and enhanced a mouse model of myoblast differentiation into myotubes. However, testing of the negatively charged study compounds, which all contained the zinc chelating moieties, was restricted entirely to *in vitro* assays due to the compounds unoptimized and insufficient ADME properties. Moreover, the study was not able to confirm direct ligand binding to GATA4 or exclude possible ligand chelation to the zinc ion by destroying its coordinating cysteine bonds. Furthermore, since reduced protein and activity levels of GATA4 are linked to several adverse effects in cardiomyocyte differentiation, cardiomyocyte proliferation, cardiomyocyte apoptosis and drug-induced cardiotoxicity, application of DNA inhibitors of GATA4 protein may include potential risks for unfavorable cardiac effects *in vivo* (Aries et al., 2004; Armiñán et al., 2009; Oka et al., 2006; Rojas et al., 2008; Suzuki et al., 2004; van Berlo et al., 2011; Watt, Battle, Li, & Duncan, 2004).

2.7 NKX2-5

NK-2 transcription factor-related, locus 5 (NKX2-5 or Csx) is a homeodomain-containing TF that is imperative for cardiac gene expression and normal heart development (Akazawa & Komuro, 2005; Serpooshan et al., 2017). The homeodomain of NKX2-5 has a helix-turn-helix motif responsible for binding to the specific consensus DNA sequence T(C/T)AAGTG. During embryogenesis, expression of NKX2-5 is essential for second heart field development, whereas the mesodermal deletion resulted in cardiac defects (L. Zhang et al., 2014). The review by Chung and Rajakumar (2016) highlights the assignment of structurally compromised NKX2-5 to cardiac malformations. To date, approximately 50 mutations of NKX2-5 along the whole protein sequence have been identified that are associated with congenital heart defects responsible for the development of

atrial septal defects, tetralogy of Fallot, and ventricular septal defects (Chung & Rajakumar, 2016). However, transgenic mice with NKX2-5 overexpression showed increased cardiac ANP and BNP expression and normal heart size, indicating that overexpression of NKX2-5 alone in not sufficient to induce cardiac hypertrophy (Akazawa & Komuro, 2003). Similarly, normal heart size was also observed in cardiac failure caused by the cardiac-specific dominant negative form of NKX2-5.

NKX2-5 cooperates with cardiac TFs such as GATA4 and TBX5 and synergistically promotes cardiac gene expression, differentiation and morphogenesis (Durocher, Charron, Warren, Schwartz, & Nemer, 1997; Hiroi et al., 2001). Based on experimental evidence, the physical interactions of NKX2-5 are mainly mediated through the homeodomain and its C-terminal extension. Two extensive structural X-ray studies have revealed the detailed molecular structure for the heterotypic interaction of NKX2-5-TBX5 bound to DNA (Luna-Zurita et al., 2016; Pradhan et al., 2016). Furthermore, mutational studies have identified the importance of Lys193 in the NKX2-5 homeodomain as an essential amino acid for the interaction with GATA4 (Kasahara et al., 2001). In addition, nuclear translocation of NKX2-5 and GATA4 was observed as a common factor for the differentiation of various stem cell types to cardiac fate and was associated with the expression of cardiac-specific markers such as ANP, beta-myosin heavy chain, troponin I and alpha-sarcomeric actin (Armiñán et al., 2009).

2.7.1 NKX2-5-targeted small molecule interventions

An outstanding study by Sadek et al. (2008) identified the small molecules associated with cardiac gene expression by screening a large chemical library for activators of the cardiac signature gene NKX2-5. Cardiac NKX2-5 mRNA and protein expression were triggered by a family of sulfonyl-hydrazone (Shz) small molecules in a variety of embryonic and adult stem/progenitor cells. Improved left ventricular function of cryo-injured rat hearts was observed *in vivo* after injection of Shz pre-treated human mobilized peripheral blood mononuclear cells (M-PBMCs) (Sadek et al., 2008). In the second study, an extensive screening project with NKX2-5 luciferase assays in mouse P19CL6 pluripotent stem cells revealed ten small molecule activators with diverse molecular scaffolds. In that study, 3,5-disubstituted isoxazoles (Isx) were identified in a embryonal carcinoma cell-based screen as small molecules capable of enhancing NKX2-5 luciferase expression *in vitro* (Russell, Goetsch, Aguilar, Frantz, & Schneider, 2012). Isoxazole small

molecules showed promising efficacy *in vivo* after a one-week treatment, enhancing cardiac muscle gene expression and cardiac function without reduction in scar size. However, more detailed studies have suggested that Isx compounds act as an activator of G protein-coupled receptor GPR68 rather than in the modulation of TFs (Russell, Goetsch, Aguilar, Coe, et al., 2012).

2.8 MEF2

The human myocyte enhancer factor-2 (MEF2) protein family consists of four members denoted as MEF2A, MEF2B, MEF2C, and MEF2D. Transcription factor MEF2 proteins are widely expressed in different cell types from yeast to human, especially in inflammatory and muscle cells (skeletal, cardiac, and smooth muscle cells), controlling the cell differentiation, organogenesis and other developmental programs (Potthoff & Olson, 2007). Highly evolutionary conserved adjacent N-terminal MADS and MEF2 domains are required for the establishment of homo- or heterodimeric MEF2 protein complexes, which target the specific DNA binding sequence CAT(A/T)4TAG/A and mediate the interactions with co-proteins (Akazawa & Komuro, 2003). The specific set of genes expressed by MEF2 activation in different cell types depends on extracellular signaling and on cofactor partnerships that directly modulate the function of MEF2 proteins. In adult tissues, MEF2 proteins regulate cardiac hypertrophy and tissue remodeling in cardiac and skeletal muscle (C. Zhang et al., 2002).

In both invertebrate and vertebrate animal models, MEF2C belongs to a small cohort of core TFs that form an evolutionarily conserved gene regulatory network that is necessary for the development of the secondary heart field (Olson, 2006). Transcription factor MEF2C has also been identified as an essential factor, together with GATA4 and TBX5, for the molecular mechanism facilitating the direct reprogramming of skin and cardiac fibroblasts towards the cardiomyocyte fate (Ieda et al., 2010; Inagawa et al., 2012; Song et al., 2012; Wada et al., 2013). Furthermore, it has been shown that TFs stoichiometry (GATA4, MEF2C and TBX5) has a major influence on the transformation efficiency and quality of the induced cardiomyocytes (L. Wang, Liu, et al., 2015).

MEF2C gene knockout studies in mice have further emphasized the essential role of the protein in heart development. MEF2C knockout mice die on embryonic days 9.5-10, with major heart defects, including improper looping, outflow tract abnormalities, and complete lack of the right ventricle (Anderson et al., 2017; Barnes et al., 2016). Thus, MEF2C protein is indispensable for the direct cellular

reprogramming and during the heart development and cardiogenesis, particularly in regulating the development of the second heart field.

2.9 TBX5

T-box transcription factor TBX5 is a member of a phylogenetically conserved family of genes involved in the regulation of developmental processes. TBX5 consists of three domains, an N-terminal domain mediating its interaction with GATA4 and NKX2-5, a T-BOX domain responsible for DNA binding and the C-terminal regulatory domain. In human hearts, TBX5 expression is maintained throughout embryonic development in the epicardium, myocardium, and endocardium (Steimle & Moskowitz, 2017). Proper function of TBX5 protein is critical for human cardiac morphology and is involved in the development of the four chambers in the heart and the electrical conducting system (Horb & Thomsen, 1999). Genetic analyses of mutations in patient populations have revealed the association of TBX5 protein to congenital heart diseases including Holt-Oram syndrome, atrial and ventricular septal defects, and tetralogy of Fallot (Mori & Bruneau, 2004). Particular protein-protein interactions of TBX5 have an important role in cardiac gene regulation including the TBX5–NKX2-5 and TBX5–GATA4 interactions (Ang et al., 2016; Garg et al., 2003; Hiroi et al., 2001; Maitra et al., 2009). Crystal structures of the TBX5–NKX2-5 protein complex bound to an ANP promoter decoy have been resolved (Luna-Zurita et al., 2016; Pradhan et al., 2016).

TBX5 has been identified as an essential core TF for direct reprogramming together with GATA4 and MEF2C (Ieda et al., 2010; Inagawa et al., 2012; Song et al., 2012; Wada et al., 2013). During the *in vitro* or *in vivo* processes, cardiomyocyte-like cells can be directly transformed from mouse cardiac fibroblasts by three defined factors. Although the current reprogramming practices are still inefficient and have room for methodological improvements, the direct reprogramming approach has huge future potential to become a key application in regenerative medicine. Regarding small molecule interventions, a recent report demonstrates that thalidomide, a drug linked to severe malformations including congenital heart defects, reduces TBX5 binding to DNA and suppresses TBX5-mediated activation of the ANP and vascular endothelial growth factor (VEGF) promoters, together with HAND2. The molecular mechanism for this reveals that thalidomide binds directly to TBX5 protein through amino acids R81, R82, and K226, all of which also contribute to DNA binding (Khalil et al., 2017).

2.10 GATA4 association to cardiac transcriptional network

Protein function is completely dependent on the protein three-dimensional structure, which is defined by locations of atoms in space. Even the small changes in protein structure, such as protein mutations, are, at worst, able to lead to improper function and may ultimately cause the disease. Protein function is also equally dependent on protein dynamics. For example, enzymes across the different species are under enormous evolutionary selection pressure to preserve their specific substrates and highly ordered protein structures to confirm strict tolerances and facilitate the catalysis (J. Liu, Faeder, & Camacho, 2009; Shoichet, Baase, Kuroki, & Matthews, 1995). On the other hand, TFs in higher organisms have particularly disordered protein structures demonstrating how evolutionary processes have combined the structural and functional requirements. Transcriptional regulation is orchestrated by TFs relying on a robust collective organization of the gene-specific activators and the general transcription machinery. The high complexity of transcription regulation in humans is accomplished via specific assemblies of TFs to control transcription. The combinatorial nature of TFs is facilitated by protein flexibility, which maximizes the specificity of promiscuous co-protein interactions. Due to combinatorial cooperation, a relatively small subset of TFs are able to control the transcriptional program of an entire cell (Jeronimo et al., 2007; J. Liu et al., 2009; Thomas & Chiang, 2006).

Hereafter, we focus on GATA4 and its association with the interdependent cardiac TFs system involving NKX2-5, MEF2C, HAND2, SRF and TBX5, which strictly controls the context-dependent processes of cardiomyocyte development, maturation and survival (Fig. 3) (Malek Mohammadi et al., 2017; Shu et al., 2015). All these TFs regulate each other's expression and DNA binding preferences in a combinatorial manner, resulting in a buffering capacity of the network (Grunert, Dorn, & Rickert-Sperling, 2016). Perturbation of the core TF network with chemical treatment or genetic alteration may lead to various cardiac phenotypes in mice, and mutations in humans have been associated with congenital heart defects. Although the TFs are the main driving force for the precise control of gene expression, co-regulators, epigenetic marks, and post-transcriptional regulators, such as microRNAs, fine-tune their expression and functional activity (Grunert et al., 2016; Lusis et al., 2010).

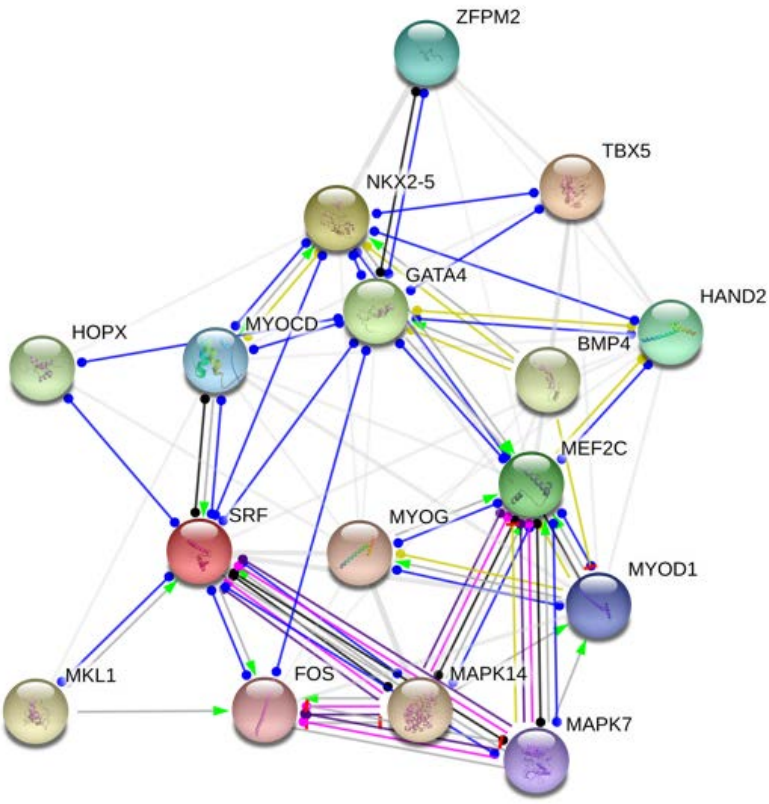

Fig. 3. Cardiac protein association map derived from the STRING database illustrates the network of interactions for selected TFs; GATA4, NKX2-5, MEF2C, HAND2, SRF and TBX5. The associations are meant to be specific and meaningful, and thus, proteins jointly contribute to the shared functions. Interaction map colour codes; blue indicates direct binding, purple indicates post-translational modifications, yellow indicates transcriptional regulation, black indicates reaction, green arrow indicates activation and grey indicates the protein's indirect contribution to shared functions. GATA4, GATA binding protein 4; NKX2-5, NK2 homeobox 5; MEF2C, myocyte enhancer factor 2C; HAND2, heart and neural crest derivatives expressed 2; SRF, serum response factor; TBX5, T-box 5; ZFPM2, zinc finger protein, multitype 2 (also known as FOG2); MYOCD, myocardin; MYOD1, myogenic differentiation 1; MAPK7, mitogen-activated protein kinase 7; FOS, FBJ murine osteosarcoma viral oncogene homolog; MYOG, myogenin; MAPK14, mitogen-activated protein kinase 14; MKL1, megakaryoblastic leukemia 1; BMP4, bone morphogenetic protein 4 and HOPX, HOP homeodomain.

All the recognized protein-protein associations of GATA4 are facilitated by the structurally stable zinc finger domains, since the major areas of the protein sequence and structure remain highly disordered (Mattapally et al., 2018). The stable N-terminal zinc finger of GATA4 is preferable responsible for mediating the physical interaction and known gene repression via binding to friend of GATA2 protein (FOG2) (Svensson, Tufts, Polk, & Leiden, 1999), while the vast majority of the synergistic heterotypic interactions of GATA4 are physically mediated by the C-terminal zinc finger and its C-terminal extension (Pikkarainen et al., 2004).

Increasing evidence shows that a restricted number of regulatory TFs (e.g., GATA4, HAND2, MEF2, NKX2-5, and TBX5) are necessary for the initiation of cardiac-like gene expression and are capable of cooperatively reprograming cardiac fibroblasts into functional cardiac-like myocytes *in vitro* and *in vivo* (Addis et al., 2013; Ieda et al., 2010; Srivastava & Ieda, 2012; M. Xin et al., 2013). Heterotypic pair-wise interactions of GATA4 have revealed that cofactors critical for direct cardiac reprogramming, such as MEF2C, HAND2 and TBX5, have a tendency to synergistically activate the GATA cis-regulatory element (Ang et al., 2016; Dai, Cserjesi, Markham, & Molkentin, 2002; Garg et al., 2003; Morin, Charron, Robitaille, & Nemer, 2000). The dominant expression of either NKX2-5 or SRF consequently leads to activation of the hypertrophic gene program, where synergy is driven through their corresponding DNA binding sites by activation of the GATA4-co-protein complex (Belaguli et al., 2000; Patent WO2012/116064A1, 2012; Sepulveda et al., 1998). Diverse preprogrammed gene activation patterns are therefore consequences of operative selectivity arising from molecular conformations of the core factors at the promoter (Fig. 4). In addition, selected heterotypic GATA4 protein ensembles are able to operate in cardiomyocytes through a single recognition element at the DNA, excluding the GATA4–TBX5 interaction which requires binding elements for the both TFs. Thus, understanding the protein assembly and consequent gene regulation via an inside out approach, starting from pair-wise heterotypic interactions as a core for more complex protein ensembles, may greatly clarify the role of single TFs in gene regulation.

Previous studies have demonstrated that GATA4 and NKX2-5 are critical TFs in gene regulation of cardiac hypertrophy. Together these TFs are required to fully activate the mechanical stretch-responsive genes such as ANP and BNP (Durocher et al., 1997; Lee et al., 1998). Mutational studies using the rat BNP promoter have revealed that the tandem GATA sites of the proximal promoter in combination with NKX2-5 binding element are required for stretch-activated BNP gene transcription (Pikkarainen et al., 2003). In contrast, down-regulation of GATA4 by an antisense

approach in cardiomyocytes blocked a stretch-induced increase in BNP mRNA levels, validating the role of GATA4 as a nuclear mediator of the mechanical stretch-activated hypertrophic program (Pikkarainen et al., 2003). Cardiac hormones ANP and BNP are direct downstream targets of GATA4 and NKX2-5 and are secreted from the heart ventricles in response to increased workload. Plasma BNP measurements are utilized in clinical diagnosis and monitoring of heart failure since the BNP concentrations are significantly upregulated due to myocardial infarction and heart failure (Cowie et al., 2003; Maisel et al., 2002).

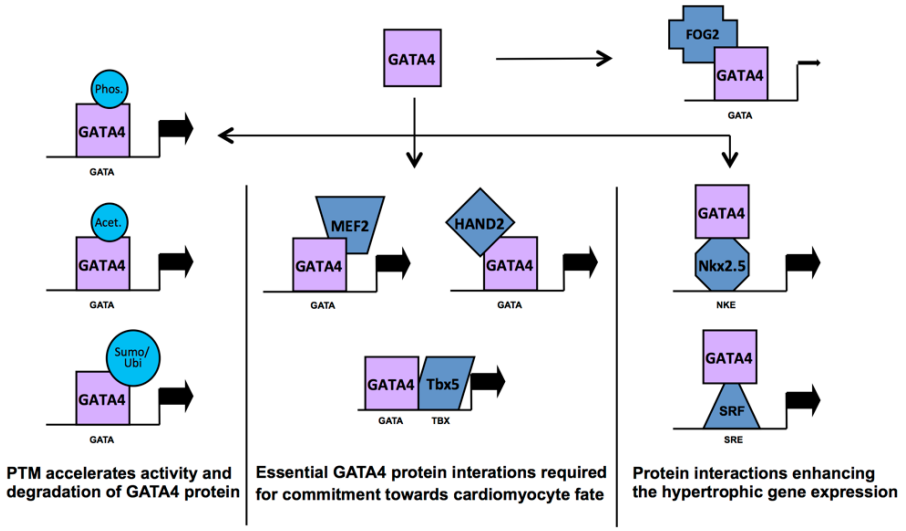

Fig. 4. Cardiac transcriptional activity is regulated by interplay of the GATA4 transcription factor with several other TFs and post-translational modifications. The vast majority of the protein associations of GATA4 are mediated by the C-terminal zinc finger, while the N-terminal zinc finger is responsible for interactions with the friend of GATA2 (FOG2). Cardiac specific heterotypic interactions and DNA occupation preferences for pair-wise GATA4 ensembles are categorized based on experimental measurements of the protein and DNA binding modes. Specific context-dependent GATA4 protein sub-consortiums regulate both the commitment of stem cells toward the cardiac fate and hypertrophic gene expression in mature cardiac cells. GATA4, GATA binding protein 4; NKX2-5, NK2 homeobox 5; MEF2C, myocyte enhancer factor 2C; HAND2, heart and neural crest derivatives expressed 2; SRF, serum response factor; TBX5, T-box 5; NKE, NK2 element; SRE, serum response element; PTM, post-translational modification; Phos., phosphorylation; Acet., acetylation; Sumo/Ubi, sumoylation/ubiquitination.

2.11 Compound-induced cell reprogramming

Breakthrough science by Takahashi and Yamanaka introduced a protocol that altered the programming of somatic cells into induced pluripotent stem cells (iPSCs) by viral transduction of four TFs; octamer-binding transcription factor-4 (Oct-4), Kruppel-like factor 4 (KLF4), sex determining region Y-box 2 (Sox2) and c-Myc (Takahashi & Yamanaka, 2006). Due to the enormous significance of iPSCs to the progress of biochemical sciences and medicine, analogous chemical biology approaches have consequently been explored. Early on, reports indicated that iPSC reprogramming efficiency may increase in the presence of individual small molecules, such as valproic acid (a histone deacetylase inhibitor (HDAC inhibitor)), azacytidine (a DNA methyltransferase inhibitor), butyrate (HDAC inhibitor) and vitamin C (Esteban et al., 2010; Huangfu et al., 2008; Mali et al., 2010; Mikkelsen et al., 2008). The first successful application of a chemical cocktail approach to the conversion of mouse fibroblasts into iPSCs included the combination of four molecules (valproic acid, CHIR99021, repsox and tranylcypromine) in addition to an overexpression of single gene, Oct-4 (Y. Li et al., 2011). These primary attempts to modify the underlying molecular mechanisms of reprogramming by chemical agents to generate iPSCs with almost no genetic modifications were encouraging. Further screening and optimization of chemical cocktail combinations were carried out to eliminate the remaining Oct-4 and achieve the complete chemical reprogramming approach. Finally, Hou et al. (2013) reported a solely chemical-induced pluripotent stem cell protocol with six compounds (valproic acid, CHIR99021, repsox, parnate, forskolin and 3-deazaneplanocin A) (Hou et al., 2013). The study suggests that instead of providing exogenous master genes, the endogenous pluripotency program might be established by the modulation of signaling pathways nonspecific to pluripotency, via small molecules.

Various strategies have been applied after the loss of cardiomyocytes to restore the declining cardiac function and prevent the progress of heart disease. Cell transplantation and virus-mediated TF delivery therapies aim to replace the damaged cells with new functional cardiomyocytes via stem cell and induced cardiomyocyte cell transplantations or direct reprogramming of non-muscle cells with forced overexpression of cardiac TFs or microRNAs (Qian et al., 2012; M. Xin et al., 2013). However, providing therapeutic value with cell transplants and the overexpression of TFs by local virus delivery remains a challenging task in clinics due to difficulties in properly integrating the new cells into the heart and the administration and safety of gene therapy. Conversion of cell types from non-

myocyte cells directly to the cardiac fate by using chemical entities may represent a viable option for therapeutic interventions. Nonimmunogenic and cost-effective small molecule compounds possess the significant advantages of cell permeability and management of standardized industrial production and quality assurance.

To date, various research settings in which non-myocytes (stem cells, progenitor cells, and fibroblasts) are exposed to chemical entities have expressed an enhanced direct cellular transformation toward the cardiac fate (Table 2) (Cahill, Choudhury, & Riley, 2017; Fu et al., 2015). Various signaling pathways have been modified by chemical entities from different compounds classes, including the activation of Wnt/β-catenin signaling (GSK3β inhibitors), inhibition of Wnt-signaling (porcupine and tankyrase inhibitors), or modulation of transforming growth factor (TGF) β-superfamily signaling (bone morphogenetic protein and TGFβ inhibitors) (Davies et al., 2015; Schade & Plowright, 2015). These molecular pathways are highly critical for the cellular lineage commitment and differentiation, especially during mesoderm induction. The development of human pluripotent stem cells towards the cardiomyocyte lineage occurs through a step-wise process that includes the induction of cardiomyocyte mesoderm, the specification of the cardiovascular mesoderm to cardiomyocyte fate, and finally, the maturation of the specified progenitors to functional cardiomyocytes (Cohen, Tian, & Morrisey, 2008; Kimelman, 2006). The optimal use of small molecule modulators of Wnt and TGFβ signaling pathways in cellular assays requires perfect timing for efficient cardiac reprogramming. Nevertheless, powerful directed cardiac differentiation platforms have been developed from human pluripotent stem cells by using small molecule modulators of Wnt signaling (Witty et al., 2014). Recently, Cao et al. (2016) and Mohamed et al. (2017) demonstrated pharmacological lineage-specific approaches for the conversion of human fibroblasts into functional cardiomyocytes *in vitro* by using either a combination of nine small molecules (CHIR99021, A83-01, BIX01294, AS8351, SC1, Y27632, OAC2, SU16F and JNJ10198409) (Cao et al., 2016) or TGFβ and Wnt inhibitors (SB431542 and XAV939) jointly with three cardiac TFs (GATA4, MEF2C, and TBX5) (Mohamed et al., 2017). Therapeutic modulation of these developmentally important Wnt or TGFβ signaling pathways may have benefits for cardiac regeneration, but they also involve the potential risk for pro-oncogenesis (Schade & Plowright, 2015).

Table 2. Summary of compounds linked to the chemical-induced cellular differentiation and reprogramming.

Compound	MW	cLogP	Mechanism of action
Azacytidine	244 Da	-2.4	DNA methyltransferase inhibitor
Butyrate	87 Da	0.8	Warburg effect
Vitamin C	176 Da	-1.6	Co-factor of epigenetic modulators
Valproic acid	144 Da	2.5	HDAC inhibitor
CHIR99021	465 Da	4.9	GSK3 inhibitor
Repsox	287 Da	2.4	ALK5 inhibitor
Parnate	133 Da	1.5	Monoamine oxidase inhibitor
Forskolin	410 Da	1.3	cAMP signaling activator
3-Deazaneplanocin A	262 Da	-1.7	S-Adenosylhomocysteine hydrolase inhibitor and histone methyltransferase EZH2 inhibitor
A83-01	421 Da	5.8	ALK5, ALK4 and ALK7 inhibitor
BIX01294	490 Da	3.9	G9a histone lysine methyltransferase inhibitor
AS8351	291 Da	2.6	KDM5B inhibitor
SC1	550 Da	5.3	ERK1 and Ras-GAP inhibitor
Y27623	320 Da	1.4	ROCK1 inhibitor
OAC2	236 Da	2.7	Oct-4 activator
SU16F	386 Da	4.5	PDGFRβ inhibitor
JNJ10198409	325 Da	3.9	PDGF-β kinase inhibitor
SB431542	384 Da	3.6	ALK5, ALK4 and ALK7 inhibitor
XAV939	312 Da	2.3	Tankyrase inhibitor
Shz-1	355 Da	3.1	NKX2-5 activator
Isx-9	234 Da	1.9	MEF2 modulator
Arotinoid acid	348 Da	6.9	RAR agonist
Rolipram	275 Da	3.2	PDE4 inhibitor

MW = Molecular weight, cLogP = Calculated LogP, Da = Dalton, HDAC = Histone deacetylase, GSK3 = Glycogen synthase kinase 3, ALK5 = Transforming growth factor beta receptor 1, cAMP = Cyclic adenosine monophosphate, EZH2 = Enhancer of zeste homolog 2, ALK4 = Activin A receptor type 1B, ALK7 = Activin A receptor type 1C, G9a = Euchromatic histone lysine methyltransferase 2, KDM5B = Lysine demethylase 5B, ERK1 = mitogen-activated protein kinase 3, Ras-GAP = Ras GTPase-activating protein, ROCK1 = Rho associated coiled-coil containing protein kinase 1, PDGFRβ = Platelet derived growth factor receptor beta, PDGF-β = Platelet-derived growth factor receptor tyrosine kinase beta, RAR = Retinoic acid receptor, PDE4 = Phosphodiesterase type 4.

Compound agents targeting directly to master cardiac TFs are rare. Sadek et al. (2008) identified chemical activators of NKX2-5, one of the earliest lineage-restricted genes expressed in cardiac progenitor cells. Chemical modifiers of the signature gene NKX2-5 were screened by using a luciferase reporter in mouse

P19CL6 pluripotent stem cells. Evaluation of a screening library with 147 000 compounds led to the identification of 10 active molecular scaffolds including the lead compounds, sulfonylhydrazones (Shz-1) and isoxazoles (Isx-9) (Sadek et al., 2008). However, more detailed follow-up studies for molecular mechanisms clarified that Shz-1 (activation of Brachyury T together with early cardiogenic program genes, such as NKX2-5 and myocardin (Sadek et al., 2008)) and Isx-9 (agonist of orphan GPR68, which is an extracellular proton/pH-sensing GPCR (Russell, Goetsch, Aguilar, Coe, et al., 2012)) mediate their actions independently of the recognized procardiogenic pathways, such as the TGFβ and Wnt superfamily.

A recent *in vivo* study demonstrate for the first time that a chemical ensemble of seven compounds (CHIR99021, repsox, forskolin, valproic acid, parnate, arotinoid acid, and rolipram) were able to induce the generation of cardiomyocytes from cardiac fibroblasts in adult mice (Huang, Tu, Fu, Wang, & Xie, 2018). Although the conversion rate was estimated to be only 1%, the treatment significantly improved cardiac function and reduced scar size and cardiac fibrosis after myocardial infarction in mice. Reprogramming in the skeletal muscle cells was not detected; therefore, chemical-induced reprogramming seems to be closely associated with the natural cardiac environment (Huang et al., 2018).

Current therapies for the treatment of myocardial remodeling are shown to attenuate symptoms and prolong the life-span by reducing the workload of the heart (e.g., angiotensin converting enzyme inhibitors, angiotensin receptor blockers, beta-blockers, diuretics and mineralocorticoid receptor antagonists) (Metra & Teerlink, 2017). However, prognosis of the patients remains poor with the present pharmacological treatments of pathological cardiac remodeling because drugs are intervening with the symptoms and not the real cause of the disease (Burchfield, Xie, & Hill, 2013). Rapid progress in the field of cellular reprogramming may eventually lead to efficient *in vitro* and *in vivo* applications for the transformation of cellular fate, representing a potential option for regenerative medicine. The evolution of the specificity and safety of these chemical interventions and gene therapy approaches remains to be seen.

3 Aims of the research

The main aim of the present study was to evaluate the protein structure, the protein-protein interactions and the function of the cardiac transcription factor GATA4, as well as to evaluate small molecules acting upon this system. Specifically, the objectives were:

1. To characterize the protein-protein binding architecture for the cardiac transcription factors GATA and NKX2-5.
2. To discover compound candidates capable of modifying the function of the GATA4-NKX2-5 protein complex *in vitro.*
3. To study the metabolism of the lead compound *in vivo*.
4. To investigate the molecular properties of GATA4-targeted compounds leading to adverse cellular effects *in vitro*.

4 Materials and methods

4.1 Homology modeling (I, II)

Comparative homology modeling predicts protein three-dimensional structures based on the amino acid sequence similarity and existence of the X-ray or NMR structure of homologous proteins (Krieger, Nabuurs, & Vriend, 1963). Domain structures are evolutionary better conserved than protein sequences, and therefore, sequence identity down to approximately 30 % may indicate significant structural similarity between proteins (Chothia & Lesk, 1986). Moreover, protein domains have a limited amount of variability in their protein folds and this evolutionary conservation allows the accurate prediction of unknown protein structures. Presently, there are protein structure homology modeling servers such as SWISS-Model available for fully automated homology modeling workflows (Waterhouse et al., 2018).

A homology model of the GATA4 protein was composed by using the highly conserved N-terminal zinc finger domain of GATA1 as a primary template structure (Protein data bank; 1GNF, sequence identity 84%) for the N-terminal zinc finger of GATA4 (Kowalski, Czolij, King, Crossley, & Mackay, 1999) and by using the highly conserved C-terminal zinc finger domain of GATA3 as a primary template structure (Protein data bank; 3DFX, sequence identity 76%) for the C-terminal zinc finger of GATA4 (Bates et al., 2008). A homology model for the homeodomain of NKX2-5 protein was built by using the highly conserved homeodomain of the thyroid transcription factor 1 as a primary template structure (Protein data bank; 1FTT, sequence identity 61%) (Esposito et al., 1996). Protein models were constructed for the zinc fingers of GATA4 covering the amino acids from 204 to 324 and the homeodomain of NKX2-5 covering the amino acids from 146 to 198. Outside these stable N- and C-terminal zinc finger- and homeodomains, GATA4 and NKX2-5 proteins are highly disordered, and therefore, those flexible areas were excluded from protein modeling (Mattapally et al., 2018). The sequence alignments of proteins were assigned without any amino acid insertions or deletions by using a sequence similarity search method, the position-specific iterative basic local alignment search tool (PSI-BLAST, National Center for Biotechnology Information) (Altschul et al., 1997). The commercial modeling package molecular operating environment (MOE 2014.09; Chemical Computing Group Inc., Montreal, Canada; www.chemcomp. com) was utilized to construct the homology models of

the protein domains. An AMBER99 and AMBER10-extended Hückel theory (EHT) force fields were implemented as the source of the atom parametrization for the protein minimizations and scoring during the homology modeling protocol. The protein side chain orientations were individually adjusted when necessary from the extensive MOE-integrated rotamer library. The following protocol was applied to all comparative protein modeling tasks. First, 10 protein domain intermediate models were generated using the Boltzmann-weighted randomized modeling procedure. Second, individual intermediate models were submitted to an electrostatic-enabled energy minimization calculation until the root mean square deviation (RMSD) gradient fell below 1. Third, the coarse refined protein models were scored and ranked by using the Generalized Born/Volume integral (GB/VI) methodology (Labute, 2008). Fourth, the highest scoring intermediate protein model was further minimized until the RMSD fell below 0.5, and then, selected as the final protein model. Finally, the stereochemical quality of these final protein models was inspected (bond lengths, bond angles, atom clashes and contact energies) and confirmed by using Ramachandran plots (Phi and Psi angles)(Lovell et al., 2003).

4.2 *In silico* fragment screening (II)

An automated identification of potential small molecule binding cavities from the interface of the homology models of GATA4 and NKX2-5 proteins were carried out using the MOE-integrated Alpha Site Finder application with default settings. Alpha Site Finder recognizes and tentatively ranks the propensity of ligand binding scores for the possible ligand binding cavities on the basis of the cavity size, hydrophilic and hydrophobic contact points (Soga, Shirai, Koborv, & Hirayama, 2007). Simultaneous evaluations of multiple copies of diverse chemical probes were carried out against the proposed ligand binding pocket in stable C-terminal zinc finger of GATA4 by using MultiFragment Search application (Miranker & Karplus, 1991). AMBER10-EHT force field was preferred as the appropriate molecular parametrization for the receptor-ligand interactions. The binding cavity next to the DNA binding site of GATA4 was exposed to a default library of 39 randomly placed molecular fragments. The library of fragments with diverse functional groups in the binding cavity were minimized and ranked based on the interaction potential with and without the solvent. The default protocols were applied on the settings concerning the fragment library, a number of fragment copies, energy minimization termination gradient and protein flexibility parameters.

Localizations of the most prominent fragments in the binding cavity were visually inspected and used as principal design components for the fragment screening *in vitro*.

4.3 Virtual screening via molecular docking (II)

Virtual screening is a computational approach to predict and identify chemical structures that are most likely to bind to preselected drug target from the large chemical libraries (Lionta, Spyrou, Vassilatis, & Cournia, 2014; Shoichet, 2004). Mutagenic studies of C-terminal zinc finger recognised the binding interface for GATA4-co-protein interactions. The virtual screening evaluations were carried out with docking software GOLD at CSC, information technology center for science in Finland. The broad binding cavity in the homology protein model was defined as the predicted ligand binding site with radius set to 10.0 Å. Preconceived conformational ligand database with 88 000 drug-like compounds concentrated from the databases of commercial vendors Chembridge and Specs were used as ligand reservoir. The number of dockings performed on each ligand was set to default 10 genetic algorithmic runs and early termination was accepted. Default ligand flexibility, default protein flexibility, automatic atom- and bond-type assignments and automatic genetic parameter settings were allowed in GOLD. GoldScore and ChemScore were selected as the fitness and ranking functions for the virtual screening. After the visual inspection, the highest scoring virtual hits were further tested in immunoprecipitation and luciferase reporter assays *in vitro*.

4.4 Pharmacophore model preparation and screening (II)

Flexible small molecule alignment calculations were implemented to study the molecular similarity in three-dimensional space for three active compounds 3, 4 and 5. The stochastic search procedure aligns the low energy conformations of the selected molecules by maximizing the molecule's chemical similarity with the least amount of internal strain. An AMBER12-EHT force field was assigned for the compound parameterization and energy minimization. Moreover, the default settings were employed to score and rank the database of variable compound superpositions (Chan & Labute, 2010). The highest-ranked alignment of active compounds was exploited to generate the preliminary pharmacophore annotation scheme involving four-point pharmacophore model with two acceptor, one aromatic and one hydrophobic feature. Furthermore, preliminary docking results

suggest the binding of low energy conformations of active compounds into the allosteric GATA4-DNA binding site. Excluded volume constraints derived from the allosteric site were used as shape restrictors in the pharmacophore query. Finally, the small molecule conformation database in MOE (650 000 drug-like compounds from 44 chemical suppliers) were deployed as a ligand reservoir for the pharmacophore search.

4.5 Screening of commercial small molecules (II-IV)

The vast majority of the small molecules for the primary screening and optimization were selected and purchased based on the fragment, virtual and pharmacophore screening from the established suppliers, including ChemBridge (San Diego, CA, USA), Enamine (Kiev, Ukraine), Sigma-Aldrich (St. Louis, MO, USA) and Specs (Zoetermeer, The Netherlands). Chemical quality of the compounds purchased for the study were characterized by the vendors in high-performance liquid chromatography–mass spectrometry (HPLC-MS) and ^{1}H NMR experiments to confirm the compound identity and a minimum purity of 90%. The synthesis of selected study compounds, i.e., 3, 4, 7, M1, and M3 were outsourced to Pharmatory (Oulu, Finland). All reactions were carried out with commercially available solvents and chemicals. All chemicals, solvents, and anhydrous solvents were acquired from Sigma-Aldrich (Schnelldorf, Germany), Fluka (Buchs, Switzerland) and Alfa Aesar (Ward Hill, Massachusetts, USA). The identity and purity of the final compounds was determined by HPLC-MS and ^{1}H NMR experiments (minimum purity >95%). The ^{1}H NMR spectra were observed at the University of Oulu using a Bruker DPX 200 instrument (Rheinstetten, Germany) in deuterated chloroform ($CDCl_3$) or deuterated dimethyl sulfoxide (DMSO-d_6).

4.6 Conformational analysis (II, IV)

Commercial modeling package MOE 2014.09 with LowModeMD module was employed to generate small molecule conformation databases (Labute, 2010). MMFF94x and OPLS-AA force fields suitable for small molecules were assigned for the molecule parameterizations and energy minimizations. Moreover, the default settings were engaged to score and rank the compound databases. The lowest energy conformations were selected to define the molecular conformations, e.g. the torsion angles between the ring planes in the southern part of the compounds (study IV). Mogul v.1.7.2 software (Cambridge Crystallographic Data

Center) relying on knowledge-based library of X-ray crystallographic data with molecular geometries was applied to provide precise experimentally-derived information about preferred ring system geometries of compounds 3 and 3i-1047 (Bruno et al., 2004).

4.7 Plasma concentrations and metabolite profile (III)

The plasma concentrations and metabolite profile of compound 3 were measured following a single administration of 10 mg/kg by intraperitoneal injection (i.p.) into male Sprague-Dawley rats weighing 250-300 g. The blood samples were collected into lithium heparin Microvette tubes from tail vein at 0.5 hours, 2 hours and 6 hours after the dosing. Sample tubes were centrifuged for 10 minutes 1300 *g* at 4°C and the plasma samples were further analysed by Novamass, Finland, by acquiring data with a Waters LCT Premier XE time-of-flight (TOF) mass spectrometer (Waters Corp.).

4.8 Chemical quality assurance

4.8.1 Compound aggregation (II)

Aggregation measurements were carried out to exclude the pan-assay interference compounds (PAINS), often causing false positives in compound screening. The compounds of interest, i.e., compounds 3, 4, and 7 were diluted in dimethyl sulfoxide (DMSO) to obtain a stock of 100 mM, which was further diluted with DMSO to obtain four different concentrations (100, 30, 10 and 3 mM). From these concentrations 1 μL of solution was diluted to 1 mL of Dulbecco's modified Eagle's medium (DMEM, Sigma-Aldrich/Gibco) containing 10% Fetal Bovine Serum (FBS, Gibco) and 1% penicillin-streptomycin (PS, Sigma-Aldrich) (100 U/mL and 0.1 mg/mL, respectively) referring the same conditions as in cell culture. Finally, 100 μL of this these solutions were transferred to a 96-well microplate (BRANDplates® pureGrade). The final concentrations (i.e., 100, 30, 10 and 3 μM) were in 0.1% DMSO. For the measurements, DMEM with FBS and PS and 0.1% DMSO was used as a blank to account for any signals due to the assay conditions. The blank and all four concentrations of the samples were measured as triplicates at three different voltages (300, 400, and 500 V) using a Nepheloskan Ascent® by Labsystems (Helsinki, Finland).

4.8.2 Metabolic stability in cellular context (II)

Extraction method for the extracellular metabolites: Eight samples with no cells were thawed at room temperature, and compound 3 was extracted from the cell culture media (mouse embryonic stem cells, mESC) with 500 μL of chloroform. The samples were vortexed briefly and centrifuged at 16 000 rpm at 4˚C for 10 minutes. The bottom chloroform layer was transferred to a glass vial and analyzed by ultra-performance liquid chromatography-electrospray (+)-quadrupole time-of-flight mass spectrometer (UPLC-ESI(+)-QTOF/MS) in the sensitivity mode.

Extraction method for the intracellular metabolites: Eight samples with cell culture media and mouse embryonic stem cells were thawed on ice and centrifuged at 16 000 rpm at 4˚C for 10 minutes. The supernatants were removed, and cells were washed twice with 200 μL of 0.9% saline (H_2O), followed by centrifugation at 16 000 rpm at 4˚C for 10 minutes. The washed cells were disrupted, and compound **3** was extracted with 200 μL of chloroform in an ultrasonic bath for 10 minutes, followed by centrifugation at 16 000 rpm at 4˚C for 10 minutes. The bottom chloroform phase was transferred to a glass vial and analyzed by UPLC-ESI(+)-QTOF/MS in the sensitivity mode.

Analytical method: A Waters Acquity ultra-performance liquid chromatographic (UPLC) system (Waters Corp., Milford, MA, USA) with an autosampler, vacuum degasser and column oven was used. The analytical column used was a Waters Acquity BEH C18, (2.1×50 mm, 1.7 μm, Waters Corp, Milford, MA, USA), together with an on-line filter. The used eluents were 0.1% formic acid in H_2O (A) and 0.1% formic acid in acetonitrile (B). A linear gradient elution of 7% B -> 80% B was applied for 5 minutes, followed by 1 minute of column equilibration. The flow rate was 0.6 mL/minutes, and the column oven temperature was set to 40°C. The HPLC-MS data were acquired using a Waters Synapt G2 quadrupole-time-of-flight (QTOF) high definition mass spectrometer (Waters Corp., Milford, MA, USA) equipped with a LockSpray electrospray ionization source. A positive ionization mode of electrospray was used with a cone voltage of 40 V and a mass range of m/z 100 – 600. The mass spectrometer and UPLC system were operated using MassLynx 4.1 software. Leucine enkephalin was used as a lock mass compound ($[M+H]^+$: 556.2771).

HPLC-MS CHROMASOLV® grade chloroform, acetonitrile, formic acid and sodium chloride were obtained from Sigma (Darmstadt, Germany). Water was freshly prepared in-house with a Milli-Q (Millipore Oy, Espoo, Finland) purification system and ultrapure grade (18.2 MΩ).

4.9 Ethics

Animal experiments were conducted according the 3R principles of the EU directive 2010/63/EU governing the care and use of experimental animals and following local laws and regulations [Finnish Act on the Protection of Animals Used for Scientific or Educational Purposes (497/2013, Government Decree on the Protection of Animals Used for Scientific or Educational Purposes (564/2013)]. The protocols were authorised by the national Animal Experiment Board of Finland (ESAVI-2028-041007-2014).

4.10 Statistics

Results are expressed as mean and standard deviation (SD). Statistical analyses were performed using SPSS Statistics 21 (IBM). Statistical significance was evaluated by one-way analysis of variance (ANOVA) followed by a least significant difference (LSD) post hoc test for multiple comparisons. To determine the statistical difference between two groups, the independent samples t-test was used. A probability value of $p<0.05$ was considered statistically significant.

5 Results

5.1 Nuclear receptor-like architecture of GATA4-NKX2-5 protein complex (I, II)

Protein structure determination by experimental methods remains a constant challenge, and only a small percentage of initial attempts of structure determinations are successful (Slabinski et al., 2007). The probability of success is restricted by two main factors, challenges either in protein production or protein crystallization. This was also evident with the GATA4 protein, since the solute protein structure was been determined by NMR (Protein Data Bank, 2M9W), but the X-ray structure remains unsolved despite efforts allocated to obtain the protein crystals. Since the current studies were executed prior to the release of the GATA4 and NKX2-5 protein structures, the homologous zinc finger- and homeodomains were used as templates for protein modeling. Retrospective superimposition of homology models of GATA4 and NKX2-5 with the corresponding protein structures (2M9W and 3RKQ) showed excellent overlap of protein backbones with 0.70 and 0.92 Å RMSD for *alpha*-carbons, respectively.

Protein model of GATA4 (zinc fingers) was constructed and utilized for the rational selection of amino acids for mutational studies (Fig. 5). Amino acid mutations are selected based on an even coverage of the protein surface without having an effect on GATA4-DNA binding. The vast majority of the protein-protein interactions of GATA4 are experimentally linked to the C-terminal zinc finger, and that particular domain has been shown to also be responsible for the interaction regarding the homeodomain of NKX2-5 (Sepulveda et al., 1998).

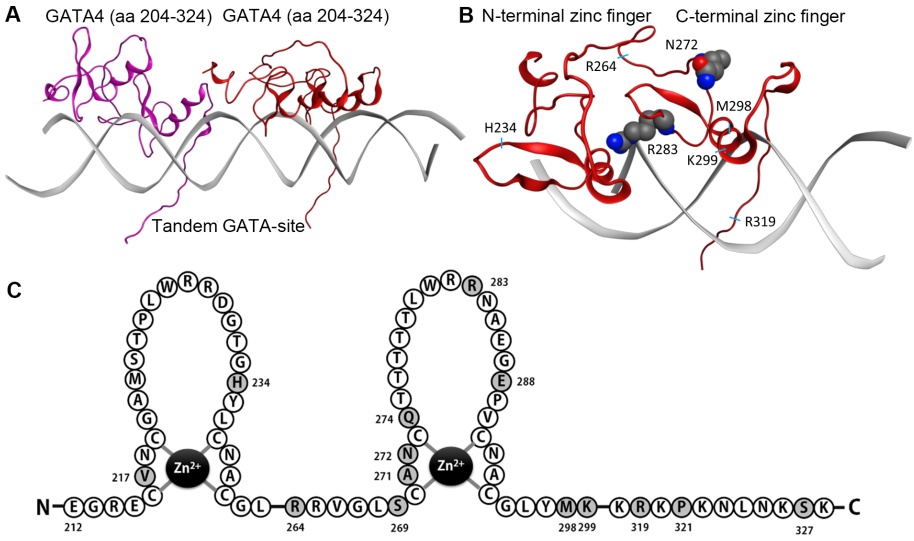

Fig. 5. Structural illustration of GATA4 protein binding to DNA compiled by homology modeling. (A) Truncated GATA4 proteins bound to specific tandem DNA binding sites. (B) Protein model demonstrating that C-terminal zinc finger of GATA4 binds to the major groove of DNA, whereas the N-terminal zinc finger may bind to the minor groove. Amino acids N272 and R283 are highlighted as ball spheres and the location of amino acids H234, R264, M298, K299 and R319 are indicated. (C) Amino acid sequence of mouse GATA4 zinc fingers containing the amino acids 212-328. Darker shades indicate the residues which were mutated during this study.

To obtain an estimate of the protein orientations and interactions of transcriptionally active protein complexes, one may take the inside out approach and determine the pair-wise orientations of the core master TFs. Overall, the GATA4 study consisted of thirteen C-terminal zinc finger point mutations (R264A, S269C, S269C+Q274H, A271V, N272D, N272S, Q274H, R283A, R283Q, E288G, E288K, M298Y, K299A), four C-terminal extension mutations (R319C, R319S, P321C, S327A) and two mutations in the N-terminal zinc finger (V217Y, H234S). The effect of mutated GATA4 proteins, together with NKX2-5, on the transcriptional activity of the BNP promoter, the transcriptional activity of the ANP promoter, physical interactions in co-immunoprecipitation assays and the synergy with another GATA4 cofactor, KLF13, were studied. Furthermore, effects of the GATA4 DNA binding were evaluated to exclude possible uncertainties associated with DNA binding. Together, the results of these studies found five singular amino acids in the second zinc finger (N272, Q274, R283, K299) and C-terminal

extension (R319) that are essential for both physical and functional interactions with the third alpha helix of the NKX2-5 homeodomain (Table 3). Moreover, the mutation of amino acid V217, important for GATA4-FOG2 interaction, increased consistently the physical interaction and synergistic effect of GATA4-NKX2-5 proteins.

Table 3. Summary of GATA4 mutation studies.

GATA4 Mutation	DNA Binding	Promoter Activation	NKX2-5 Binding	Synergy with NKX2-5 on 3xNKE	Synergy with NKX2-5 on ANP	Synergy with KLF13 on ANP
WT	Normal	Normal	Normal	Normal	Normal	Normal
V217Y	Normal	Normal	↑	↑	Normal	↑
H234S	Normal	Normal	Normal	Normal	↓	Normal
R264A	Altered	Normal	↓	Normal	Normal	Normal
S269C	↓	Normal	Normal	Normal	ND	ND
S269C+Q274H	↓	Normal	Normal	Normal	ND	ND
A271V	↓	Normal	Normal	Normal	ND	ND
N272D	↓	↓	↓	↓	ND	ND
N272S	↓	↓	Normal	↓	↓	Normal
Q274H	Normal	Normal	↓	Normal	↓	Normal
R283A	↓	↓	↓	↓	Normal	↓
R283Q	Normal	↓	↓	↓	ND	ND
E288G	Normal	Normal	Normal	↓	↓	Normal
E288K	Normal	Normal	Normal	Normal	ND	ND
M298Y	Normal	Normal	↓	Normal	Normal	Normal
K299A	Altered	↓	↓	↓	↓	Normal
R319C	↓	↓	↓	↓	↓	↓
R319S	Normal	Normal	↓	↓	ND	ND
P321C	Normal	Normal	Normal	↓	ND	ND
S327A	Normal	Normal	Normal	↓	Normal	Normal

↑ = increased vs. wild type, ↓ = decreased vs. wild type, ND = not determined.

Nuclear receptor family proteins have evolutionary conserved DNA binding domains consisting of two zinc fingers packed together. Beyond the structural arrangement at the domain level, nuclear receptors express highly preserved amino acid composition in the helix III of the second zinc finger mediating the interaction between the zinc fingers, such as estrogen receptor residues R63, K66 and C67 (Schwabe, Chapman, Finch, & Rhodes, 1993). Remarkably, a structural assessment of the NKX2-5 homeodomain revealed a nuclear receptor-like binding pattern arranged in a helical motif (R190, K193 and C194). Moreover, it has been previously shown that helix III of the NKX2-5 homeodomain, especially K193, is

necessary for the interaction with the C-terminal zinc finger of GATA4 (Kasahara et al., 2001). Integration of the structural information and the data from single point mutational studies with GATA4 and NKX2-5 suggest that the physical interaction between GATA4 and NKX2-5 domains resembles the architecture of the highly conserved DNA binding domain of nuclear receptors (Fig. 6).

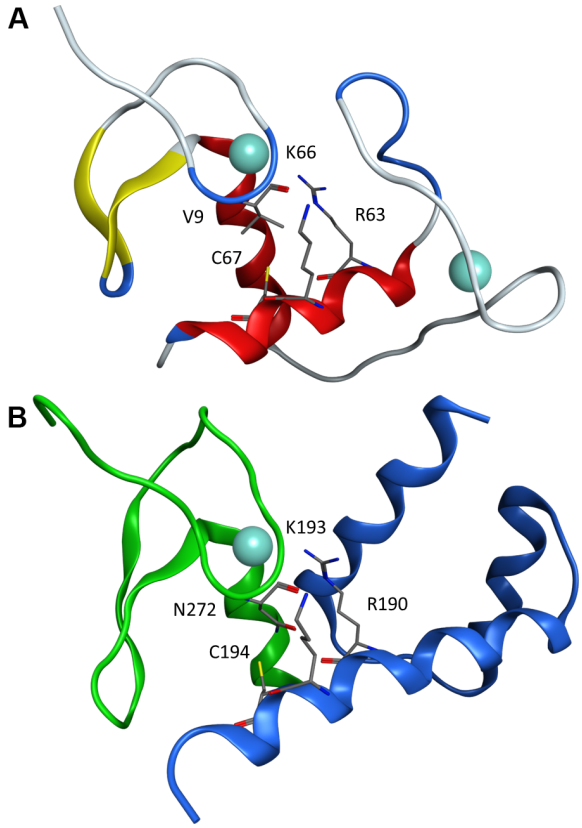

Fig. 6. Nuclear receptor-like structural architecture of GATA4-NKX2-5 interaction. (A) DNA binding domain of the estrogen receptor (PDB, 1HCQ) presenting amino acids V9, R63, K66 and C67 facilitating the interaction between two zinc finger domains. (B) Homology model of the C-terminal zinc finger of GATA4 (green) and the homeodomain of NKX2-5 (blue) indicating amino acids R190, K193, C194 and N272. Structural similarity between GATA4 protein model and estrogen receptor (1HCQ) is 1.81 Å RMSD (*alpha*-carbon). Zinc atoms are represented as light blue spheres.

To assess the wider structural and biological significance of the nuclear receptor-like assembly of the GATA4-co-protein complex, a multiple sequence alignment of 129 non-reductant homeodomains expressed in humans was utilized (Banerjee-Basu & Baxevanis, 2001). Previously, it has been shown that the third helix of NKX2-5, particularly the amino acid K193 is absolutely required for the interaction with GATA4 (Kasahara et al., 2001). On the other hand, our study indicates that the nuclear receptor-like complex prefers the amino acid C193, or to a lesser extent the amino acid S193, in the third helix of the homeodomain. Theoretically, the broadest definition may also include the amino acid G193, which together with localized water resembles the overall outcome of a serine residue (from an electrostatic and volume standpoint). Surprisingly, the amino acid combinations of C193K194, S193K194 and G193K194, which resemble the structure of NKX2-5, are extremely rare among the aligned homeodomain proteins. As a result, the homeodomain sequences of TFs, such as Nanog (S193K194) and Oct-4 (G193K194) are characterised as partial structural equivalents to enable the nuclear receptor-like protein assembly following GATA4 binding. This protein-protein intervention may partly explain both the synchronized protein expressions of GATA4, Nanog and Oct-4 observed during stem cell differentiation (Zwi et al., 2009) and the role of GATA4 as a repressor of Nanog activity contributing adversely to stem cell pluripotency (T. Li et al., 2018; Serrano et al., 2013).

Our proposed GATA4–NKX2-5 interaction model suggests an inability of the protein complex to drive the synergy through the GATA binding site. This directional control arises from the fact that the implemented protein-protein interaction influences to the number of binding options which the protein complex may have. More specifically, upon binding to GATA4, NKX2-5 will partly occupy the same space required for the direct DNA binding of GATA4; therefore, the GATA4-NKX2-5 complex is due to steric hindrances directed towards driving the synergy through the NKX2-5 binding site (Fig. 7).

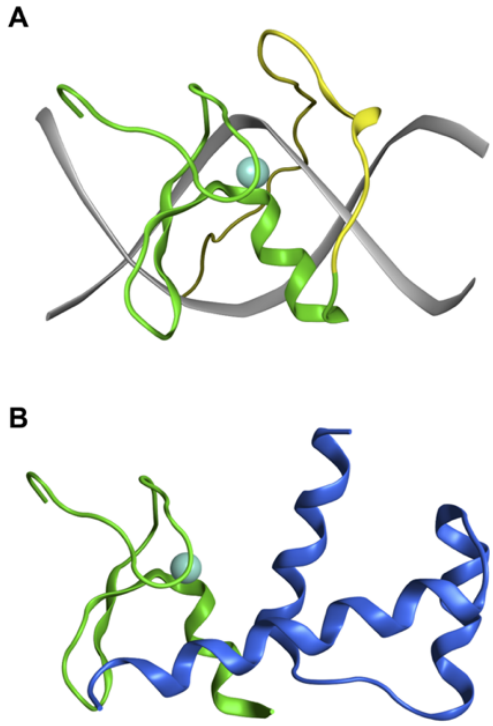

Fig. 7. (A) C-terminal extension in the second zinc finger domain (yellow ribbon) provides outstretched amino acid surface for DNA binding of GATA4. (B) In order to activate the GATA4–NKX2-5 interaction, the C-terminal loop extension needs to relocate and it will lead to structural reorganization of the GATA4 complex in the tandem GATA binding site.

5.2 Discovery of lead compounds targeting the synergy of cardiac transcription factor GATA4 and NKX2-5 (II)

The aim of the study II was to uncover the compound candidates affecting the protein-protein interaction or the synergy of cardiogenic transcription factors GATA4 and NKX2-5. Two separate research approaches were applied to investigate the druggability of the protein targets (Fig. 8). First, an extensive virtual screening campaign with the GATA4-NKX2-5 protein-protein model was carried out and subsequently evaluated by co-immunoprecipitation assay. However,

confirmatory luciferase reporter assays of 15 candidate compounds could not validate the research findings, and this approach was discontinued.

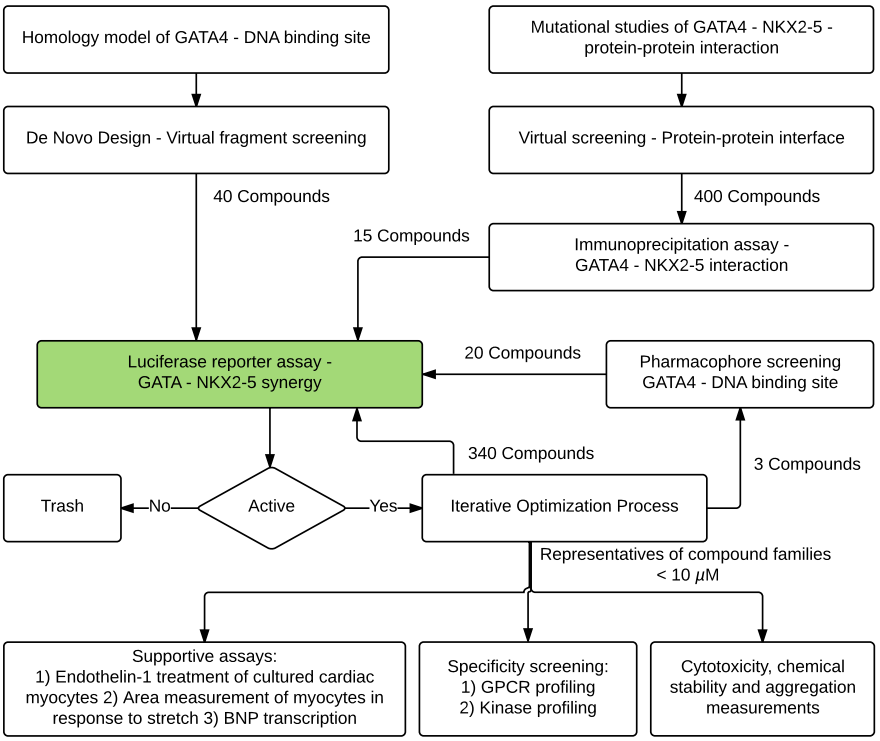

Fig. 8. Overall process workflow describing the step-by-step evolvement of the GATA4-targeted small molecule discovery project.

A parallel *de novo* design strategy was applied for the allosteric GATA4-DNA binding cavity in the C-terminal zinc finger of GATA4. Electrostatic preferences of the proposed binding site were estimated by virtual screening of 39 rigid molecular fragments. Theoretical binding annotations preferred the fragment compounds with negative charge/acceptor and aromatic features with specific distances. In total, 40 fragment compounds were selected and purchased based on defined specification. An in-house established GATA4–NKX2-5 luciferase reporter gene assay was utilized to recognize two fragments that inhibited the transcriptional synergy, 3,4-

diphenyl-1,2,5-oxadiazole (compound 1) and 2-perfluorophenyl-3-phenylacrylic acid (compound 2), at a concentration of 50 μM (Fig. 9).

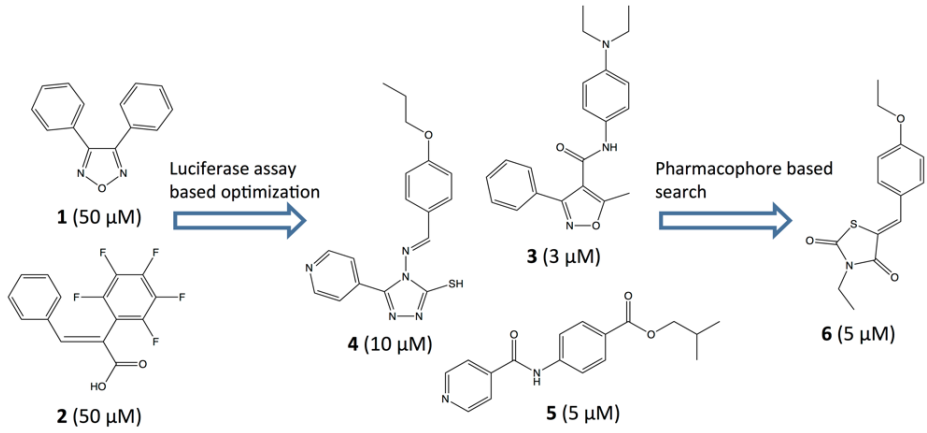

Fig. 9. Small molecule findings in chronological order from the fragment hits to the most potent lead compound 3 (WO2018055235, 2018). Figure indicates the molecular structure, compound code numbers and corresponding IC_{50} values in the GATA4-NKX2-5 luciferase reporter gene assay.

To further improve the affinity of the active fragments, multiple optimization cycles were executed. Initial fragment compounds were modified by inserting the variable substituents and linkers in a sequential manner. Top ranked activities in luciferase assay were achieved by joint compounds with both a para-substitution and two heavy atom linkers. The most potent lead compound, *N*-[4-(diethylamino)phenyl]-5-methyl-3-phenylisoxazole-4-carboxamide (3, IC_{50} 3 μM) (WO2018055235, 2018), dose-dependently inhibited the GATA4–NKX2-5 transcriptional synergy (Fig. 10). All inhibitory compound families, 3-5, identified during this study included at least 15 structurally related compounds. A total of 800 compounds were experimentally evaluated in this project with the primary luciferase reporter and/or immunoprecipitation assays. Additionally, a follow-up study covered an additional 250 in-house synthesized derivatives for a more detailed structure-activity relationship determination. The follow-up study conformed that the inhibition of GATA4-NKX2-5 synergy was not structurally associated with the phenyl ring at the southern part of molecule 3 (Jumppanen et al. manuscript under preparation).

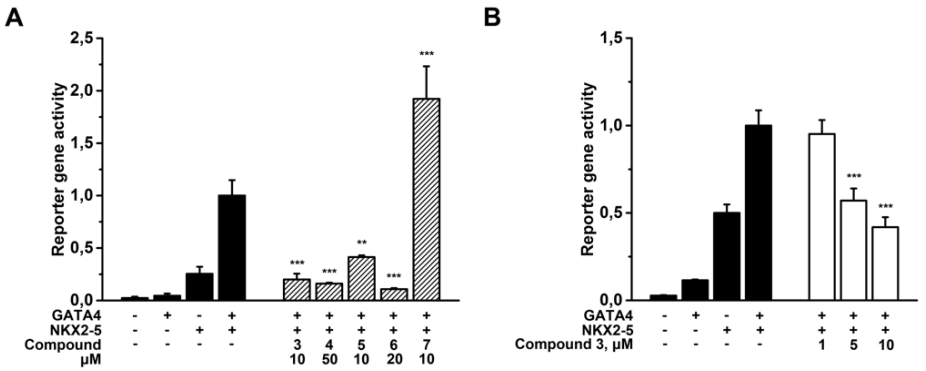

Fig. 10. Cell-based luciferase reporter assay showing either inhibited or enhanced gene expression after treatment with the selected compounds 3, 4, 5, 6 and 7 (A), as well as dose-dependent response for compound 3 (B). All compounds 3-7 were identified during the screening campaign. The data are shown as the mean ± SD, n = 2-9. ** p < 0.01, *** p < 0.001 *vs.* vehicle treatment.

To further evaluate and search the active molecular landscape, a pharmacophore model was built on the basis of three inhibitory compounds: 3, 4-[(4-propoxybenzylidene)amino]-5-(pyridin-4-yl)-4*H*-1,2,4-triazole-3-thiol (4) and isobutyl 4-(isonicotinamido)benzoate (5). The pharmacophore building process, including the conformation generation and the compound superposition was accomplished with a high confidence due to similar structural features and relatively rigid core structures of all three inhibitory compounds (Fig. 11). Moreover, the pharmacophore model was finalized by adding the excluded volumes derived from the homology model of the allosteric GATA4-DNA binding site. Pharmacophore model validation was carried out with 15 inactive and structurally diverse compounds selected from the parallel screening path targeting the implied GATA4-NKX2-5 protein-protein interface. Here, 1 of the 15 compounds was recognized during the validation screening with the pharmacophore model. Finally, the commercial library of 750 000 compounds were screened with the pharmacophore model. After a visual inspection of the positive compound modalities, 20 commercial compounds were purchased and tested *in vitro* by using a GATA4-NKX2-5 luciferase reporter gene assay. As a result, a new active compound, 5-(4-ethoxybenzylidene)-3-ethylthiazolidine-2,4-dione (6) and its derivatives was identified to inhibit GATA4-NKX2-5 synergy at a concentration of 5 μM.

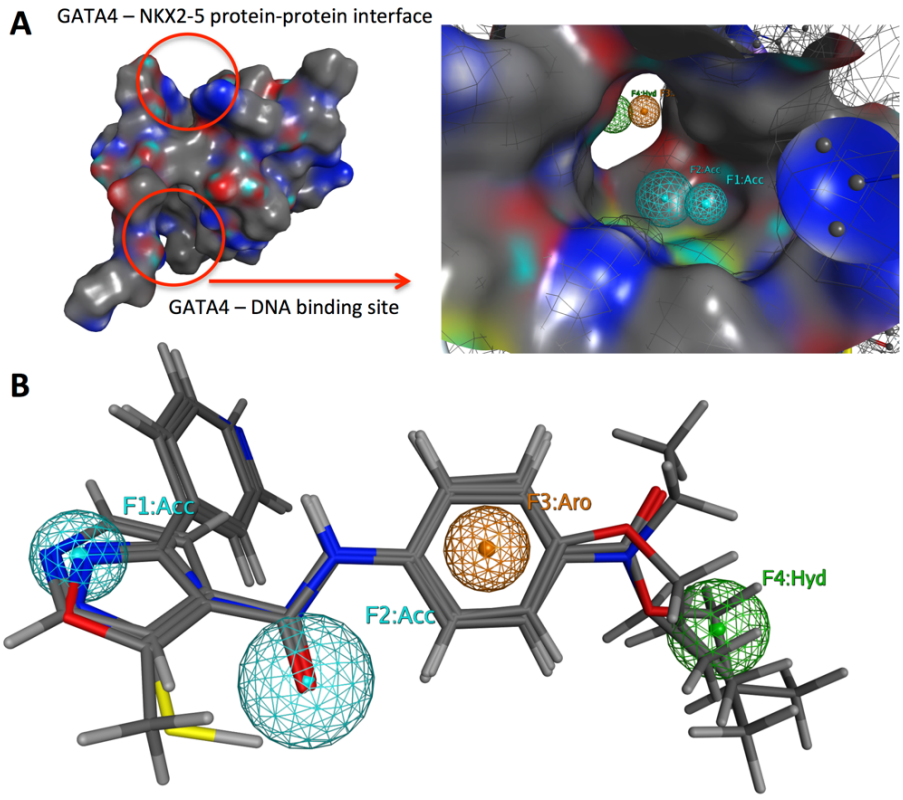

Fig. 11. (A) C-terminal zinc-finger domain of GATA4 highlights both the protein-protein interaction interface and the allosteric DNA binding site. Computational screening methods were applied to discover small molecule binders on both implied binding sites. A pharmacophore model with GATA4 protein derived excluded volumes was constructed for the allosteric DNA binding cavity. (B) Force field-based low energy conformation search for compounds 3, 4 and 5 along with compound superposition were used as a foundation to build up the pharmacophore model. All three compounds are solid inhibitors of the GATA4–NKX2-5 synergy in the luciferase reporter gene assay. Atom color: White = Hydrogen, Gray = Carbon, Red = Oxygen, Blue = Nitrogen and Yellow = Sulfur. Pharmacophore features: Blue sphere = Acceptor, Orange sphere = Aromatic and Green sphere = Hydrophobic.

Four compound families recognized during the *in vitro* screening campaign, either by using the fragment-based or the pharmacophore approach, inhibited the GATA4–NKX2-5-induced synergistic reporter gene activation of the promoter containing the three high-affinity NKX2-5 binding sites. However, weak activation

of the GATA4-NKX2-5 luciferase reporter gene was also noted randomly for the compounds with similar structural elements and the diverse molecular scaffolds in comparison to the inhibitory compounds. Surprisingly, during the small molecule screening and optimization, a compound that strongly augmented the GATA4–NKX2-5 synergy was detected. This activator compound, *N*-(4-chlorophenyl)-5-methyl-*N*-(4-methyl-4,5-dihydrothiazol-2-yl)-3-phenylisoxazole-4-carboxamide (7), structurally resembles and overlaps inhibitory compound 3, with an additional 5-member ring moiety attached. However, due to the complex structure of the activator molecule, the synthesis of its derivatives was not pursued in this study.

A number of confirmatory bioassays, commercial screening assays and chemical quality measurements were carried out to ensure the validity of the primary screening results. First, the effects of the GATA4-targeted compounds on cardiac gene expression (ANP and BNP) in ET-1-treated cardiac myocytes were defined. Second, compound 3 significantly inhibited the increase in the area of the myocytes in response to the stretching, indicating that the compound may inhibit hypertrophic growth. Third, compound 3 also significantly inhibited GATA4-driven transactivation of luciferase reporter constructs containing either BNP minimal promoter or BNP promoter containing minimal promoter and a tandem GATA-site. In additional to the bioassays, effect of compound 3 on function of G protein-coupled receptors and protein kinases were evaluated by commercial screening assays. The screening showed that the protein kinases involved in the regulation of GATA4 phosphorylation were not affected by compound 3. Similarly, the compound's effect on DNA binding was evaluated. Moreover, variable compound structure-related negative effects affecting the interpretation of the results were experimentally determined, including compound aggregation, cytotoxicity and metabolic stability in solution and in the cellular context.

5.3 *In vivo* cardioprotective effects of lead compound - pharmacokinetics and the metabolic stability (III)

To evaluate the pharmacokinetics and the metabolic stability of compound 3 *in vivo*, the concertation of compound 3 was examined in normal rats by injecting a single dose i.p. (10 mg/kg). During the experiment, blood samples were collected at 0.5 hour, 2 and 6 hours after injection. As a result, the plasma level of the parent compound 3 was highest at 0.5 hours, indicating a rapid metabolic degradation of compound 3 in rats (Fig. 12). The major degradation products (compounds M3 and M4) demonstrated both de-ethylation and hydroxylation modifications and showed

a relative stable and abundant plasma levels. Other metabolites identified from the plasma had either short half-lives or levels. Additionally, preliminary biological studies with metabolites M3 and M4 proposed inactive outcomes in comparison to parent molecule 3 (data not shown).

In vivo experimental models of myocardial ischemia and pressure overload showed favourable effects of the compound 3 on cardiac function and associated upregulation of natriuretic peptide gene expression (study III). Moreover, another *in vivo* study with compound 3-loaded multifunctional nanoparticles targeted to the endocardial layer of the injured heart attenuated the hypertrophic signaling cascade (Ferreira et al., 2017). A string of proof-of-principle experiments demonstrated the significant potential of compound 3 to provide cardiac protection after myocardial infarction and other cardiac injuries.

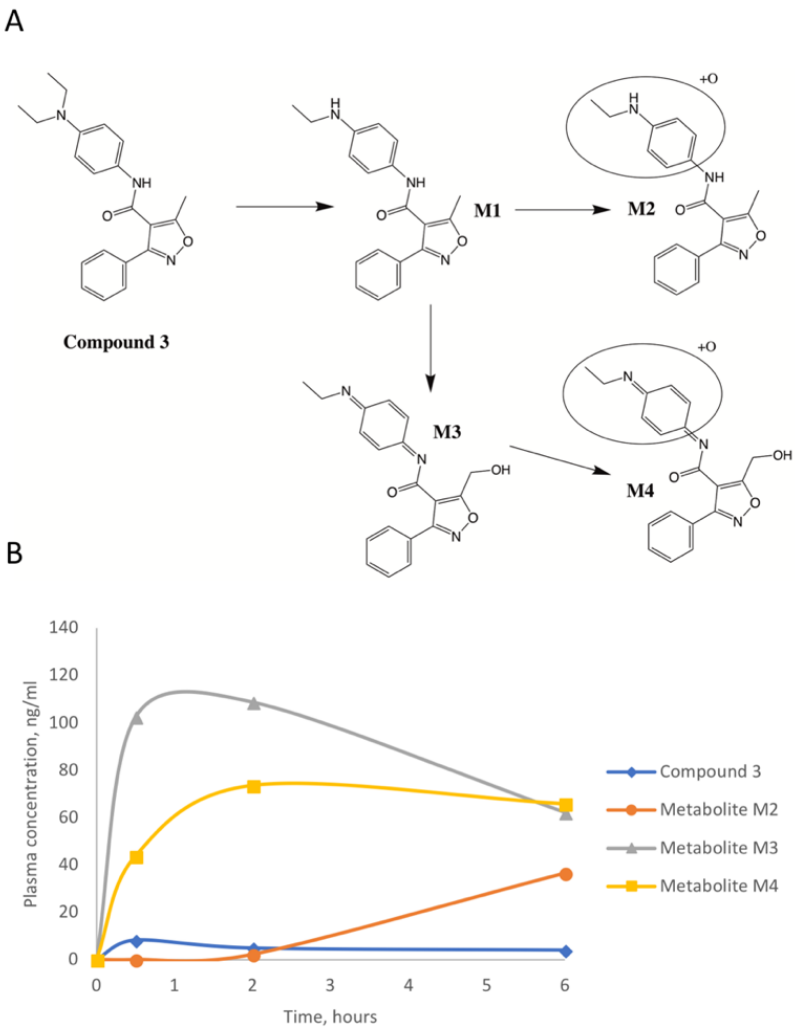

Fig. 12. Pharmacokinetics and metabolite profile in rats after the single injection of compound 3. (A) Molecule structures of compound 3 and the most abundant metabolites identifying the de-ethylation and hydroxylation as preferred metabolic pathways. (B) Plasma concentrations for compound 3 and the metabolites M2-M4 were measured at three-time points during the 6 hours experiment. Concentration of metabolite M1 in plasma remains under 3 ng/ml for 6 hours. Compound 3 was administered i.p. (10 mg/kg) in rats and blood samples were collected from tail vein. The pooled plasma sample from three rats were analyzed by HPLC-TOF mass spectrometer.

5.4 Structural features linked to *in vitro* stem cell cytotoxicity (IV)

Early recognition of compounds possessing cellular toxicity offers a major advantage to lead compound selection during the discovery process. Comprehensive *in vitro* toxicity profiling of GATA4-targeted compounds was executed by utilizing various cardiac and non-cardiac cell lines. Cell viability and toxicity were studied using the lactate dehydrogenase (LDH) assay and the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium (MTT) bromide assay in order to measure both necrosis and mitochondrial redox metabolism. A total of eight structurally conserved compounds were preselected and evaluated using eight different cell lines (Fig. 13): H9c2 myoblasts derived from rat myocardium, primary neonatal rat ventricular cardiomyocytes (NRVCs), primary neonatal rat cardiac fibroblasts (CFs), mouse embryonic fibroblasts (MEFs), mouse embryonic stem cells (mESCs), mESC derivatives from day 5 embryoid bodies (D5EBs), human induced pluripotent stem cells (hiPSCs) and hiPSC-derived cardiomyocytes (hiPSC-CMs).

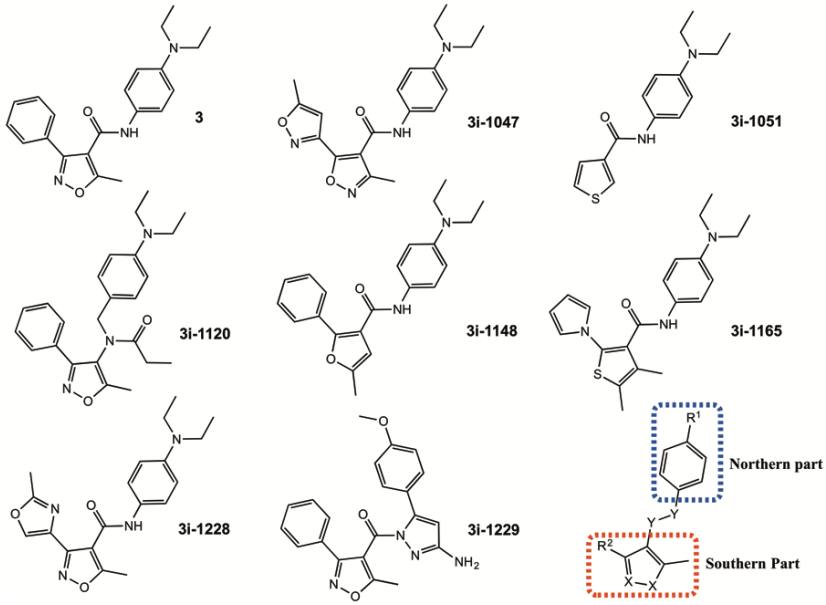

Fig. 13. Molecular structures of compound 3 and the selection of the other GATA4-targeted compounds investigated for cellular toxicity. These test compounds have common molecular scaffolds characterized by northern- and southern parts and the connection via the linker unit.

A comparison of the results composed from the *in vitro* toxicity experiments showed that the test compounds can be divided into two main categories: toxic (3, 3i-1120, 3i-1148 and 3i-1229) and non-toxic (3i-1047, 3i-1051, 3i-1165, and 3i-1228) compounds, depending on their effect on cell viability. The study found that exposure to GATA4-targeted compounds led to cell type-specific toxicity. Cardiomyocytes, fibroblasts, and H9c2 cells were the most resistant cell types, whereas stem cells (mESCs, hiPSCs and D5EBs) were the most sensitive, experiencing the caspase-dependent apoptosis. Overall, 3i-1148 was considered to be the most toxic compound with an IC_{50} of 0.30 μM based on the hiPSC toxicity data. However, none of the compounds induced significant necrosis in any cell type, as measured by LDL release into the culture medium.

Low energy conformations of the compounds are most likely the bioactive/toxic target-protein binding conformations of the molecule. In this context, conformational analysis of the compounds was carried out by force field (MMFF94x and OPLS-AA) and knowledge-based methods (Mogul). Assessment of the compound structures identified two distinct compounds categories (3 and 3i-1047 families), with a characteristic dihedral angle in the ring system (Table 4). The 6-membered ring in the southern part of compound 3 and its derivatives showed calculated dihedral angles from 28 to 51 degrees, whereas the 5-membered ring in the southern part of the 3i-1047 family preferred a flatter orientation, with values ranging from 0 to 19 degrees.

Table 4. Summary of torsional angles measured for the study compounds.

Compound	MMFF94X		OPLS-AA	
	Torsional angle (degrees)	No. of conformations	Torsional angle (degrees)	No. of conformations
3	49.0	46	42.1	149
3i-1047	13.4	56	19.4	308
3i-1051	ND	ND	ND	ND
3i-1120	41.9	309	42.6	488
3i-1148	43.1	39	50.5	91
3i-1165	1.3	22	14.6	66
3i-1228	0.1	13	7.4	263
3i-1229	50.1	32	27.9	26

ND = not determined.

In the case of toxic compounds (compound 3 and derivatives), the critical molecular region in the ring system is more crowded because of the presence of two C-H bonds in the 6-membered phenyl ring (*ortho* position). Steric hindrance increases the internal energy of compound 3 and its derivatives due to the overlapping electron clouds and destabilizes the coplanar orientation. Notably, that toxicity was preserved throughout the extensive structural variations in the middle and northern parts of the compound (3i-1120 and 3i-1229, respectively), while in non-toxic compounds (3i-1047 family), the presence of heteroatoms and a lack of hydrogen allows the southern part to adopt a nearly coplanar orientation. Moreover, additional intramolecular hydrogen bonds in compounds of the 3i-1047 family contribute positively to the low energy conformations.

A parallel conformational analysis was carried out with Mogul (Cambridge Crystallographic Data Center) by comparing the compounds to the data derived from small molecule crystal structures. This knowledge-based approach suggests that the compound 3 family, with a 6-membered ring in the southern part, possess a significantly larger dihedral angle in comparison to the 5-membered ring systems in the 3i-1047 compound family. Overall, these findings confirm that the toxic outcome in stem cells is predominantly linked to the characteristic dihedral angle in the southern part of compound 3 derivatives (Fig. 14).

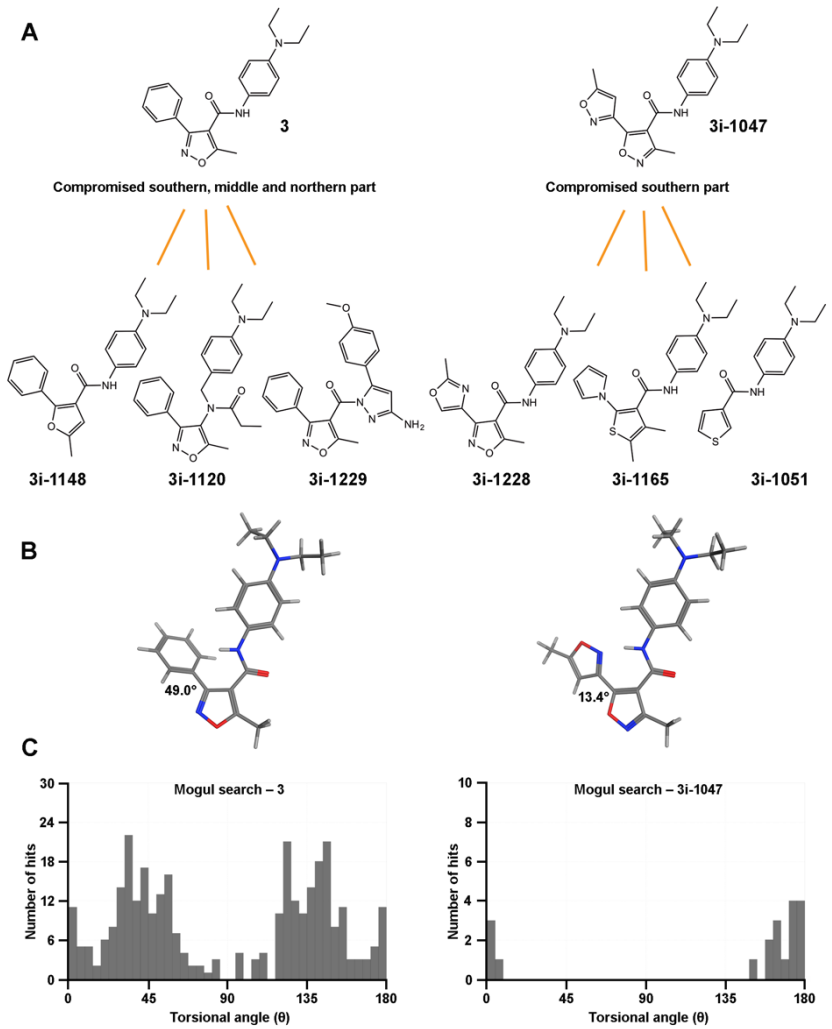

Fig. 14. Structure-based toxicity follows linearly the structural and conformational organization in the southern part of compound. Based on the results, study compounds were arranged into two structural classes containing either 5-member or 6-member ring bound to isoxazole in the southern part (3- or 3i-1047-families) (a). Global low energy conformations estimated with MMFF94x force field parametrization revealed the family-preferred 3D molecular outcome for the representative compounds 3 and 3i-1047 (b). Knowledge-based conformational analysis with Mogul software (Cambridge Structural Database) indicates a unique set of ring system torsion angles for both compound families (c).

6 Discussion

6.1 Protein-protein interaction of the cardiac transcription factors GATA4 and NKX2-5

In biological systems, protein-protein interactions (PPIs) are fundamental molecular recognition processes that are responsible for forming the TF networks important for the cellular homeostasis implicated in health and disease in humans. The complexity of protein-protein associations is enormous in humans, consisting of 650 000 direct protein-protein interactions, referred to as the interactome (Stumpf et al., 2008). Current protein interactome databases, such as the STRING database, covers approximately 10 million proteins from 2 000 organisms with known and predicted protein-protein interactions, including direct physical and indirect functional associations (Szklarczyk et al., 2017). Due to the overwhelming complexity and lack of structural and functional knowledge, the protein-protein interactions are considered to be a challenging target class for drug innovations, and completed research projects have yielded relatively few small molecule drugs directly disturbing the interaction between two proteins (Scott, Bayly, Abell, & Skidmore, 2016). There are basically two main design strategies for the discovery of PPI inhibitors, directly targeting either to binding interface of the protein or the allosteric binding site relevant for the protein mechanism and conformational control. Protein binding interfaces tend to be too large and featureless for efficient small molecule binding. Therefore, one of the major dogmas in the design of direct PPI inhibitors is based on the identification of hot spot residues in the protein interface. Detailed studies have demonstrated that not all amino acids in the binding surface are equally responsible for the association, rather, there are certain residues or regions (hot spots) that are predominantly responsible for protein interactions (Bogan & Thorn, 1998; Guo, Wisniewski, & Ji, 2014).

To search and evaluate small molecules for the GATA4-NKX2-5 interaction, more detailed information was needed to describe the protein binding interface and responsible amino acids. To this end, study I was conducted to clarify the protein binding mode by mutating 19 different residues from the surface of GATA4 and evaluate their ability to bind DNA and to physically and functionally interact with NKX2-5. The mutated amino acids were located predominantly in the C-terminal zinc finger, since this domain has been shown to be relevant for NKX2-5 binding. On the other hand, previous studies have demonstrated that the third helix of

NKX2-5 homeodomain, and especially K193, are responsible for GATA4 interaction (Kasahara et al., 2001). Overall, the study identified five amino acids critical for physical and functional interaction in the second zinc finger of GATA4 (N272, Q274, R283, K299) and its C-terminal extension (R319). Integration of the experimental data with computational modeling suggested that N272 and residues in that region were responsible for direct protein-protein interaction. Based on the protein model, three positively charged amino acids, R264, R283 and K299, were found to be located on the same side of the C-terminal zinc finger, and these single site-mutations seemed to directly regulate the folding of GATA4 at the tandem GATA binding site. In the case of the R264A mutation, N- and C-terminal zinc fingers are tightly packed together, facilitating the simultaneous binding of two GATA4 proteins to the promoter sequence, while the K299A mutation elicited a more open conformation, allowing only one GATA4 protein to harbor the promoter. The experimental transcriptional assays demonstrated the binding preferences and functional roles of the mutations, since R264A activates the promoter containing a tandem GATA site, similar to wtGATA4, whereas K299A was unable to efficiently bind this sequence or subsequently activate the transcription. Alterations in the packing of C- and N-terminal zinc fingers at the promoter may thus have effects on the reduced activity detected with these GATA4 mutants.

However, even though the N272 mutations of GATA4 were not clear-cut in immunoprecipitation assays, the relevance of this residue was more prominent in functional reporter assays where amino acid N272 mutations were unable to activate the tested gene promoters. Occasionally, relatively small structural changes in the protein interface may lead to unpredictable strong effects in protein interactions and function. Importantly, amino acid N272 is precisely at the site equivalent to the binding interface necessary for the internal packing of zinc fingers in nuclear receptor family. Therefore, the structural arrangement of the zinc finger-homeodomain resembles the architecture of the conserved DNA binding domain of the nuclear receptors. To mimic the nuclear receptor analogy, the binding interface should contain at least a successive cysteine and lysine in a helical structure to drive binding. Therefore, mutations R319C and R319S were designed to test the stability of the protein interaction by creating an artificial binding motif in the flexible basic region of the protein, which might resemble and compete with nuclear receptor and/or NKX2-5-like binding. Strikingly, mutations at R319 (to cysteine or serine) showed decreased synergistic activity with NKX2-5, especially on the NKE-dependent promoter. These data suggest that residues in the GATA4 basic domain may interfere with NKX2-5 interaction, consistent with initial reports of GATA4-

NKX2-5 cooperativity and nuclear receptor-like structural architecture. Additionally, a sequence comparison of the homeodomain structures reveals the uniqueness of a cysteine-lysine fingerprint in the third helix of the homeodomain, suggesting the exclusive role for the reprogramming transcription factors NKX2-5, Nanog and Oct-4 as a part of the possible nuclear receptor-like protein arrangement. Recently, Mattapally et al. (2018) proposed a computational model for GATA4 and NKX2-5 association which takes place outside the zinc finger and homeodomain areas, respectively (Mattapally et al., 2018). The relevance of the protein assembly in that study is highly speculative, since the experimental data does not support the proposed protein-protein binding mode (Durocher et al., 1997; Kasahara et al., 2001).

6.2 Compounds interfering with GATA4-NKX2-5 synergy

A proof of the prominence of TFs and their co-interactions as drug targets has been provided by therapeutics such as retinoic acid-, glucocorticoid-, calcitriol-, estrogen- and androgen-compound derivatives that act on the nuclear receptor family (Burris et al., 2013). Overall, more than 50% of FDA-approved small molecule drugs target three protein subclasses, including G-protein coupled receptors (33%), ion channels (18%) and nuclear receptors (16%) (Overington, Al-Lazikani, & Hopkins, 2006; Santos et al., 2016). Nuclear receptors have characteristic ligand and DNA binding domains that mediate the specific ligand and co-factor binding responses.

Study II thoroughly evaluated two complementary molecular docking approaches, targeting the direct protein-protein interface and allosteric binding site, to discover possible chemical agents interfering with the activation of GATA4-NKX2-5-induced synergy. Direct protein-protein interaction-acting compounds were assessed by a virtual screening campaign targeting to the large and shallow binding interface that was identified during GATA4 mutational studies. This approach was exceptionally challenging and ineffective for the search of active compounds in immunoprecipitation assays, possibly due to known issues in achieving an adequate binding affinity against broad protein-protein binding interfaces. On the other hand, the allosteric binding site detected from the GATA4 protein model demonstrated more promising results in preliminary experiments where virtual screening hits were assessed against the DNA binding of GATA4. However, inhibitory screening hits were not suitable for further structure optimization due to molecule size, and other non-drug-like properties (Lipinski,

Lombardo, Dominy, & Feeney, 2001; Young & Leeson, 2018). Instead, a fragment-based approach was applied for screening the allosteric binding pocket for the identification of structurally more relevant compounds. Initial hits were further optimized by fragment growing and linking to obtain the more potent compound structure. As a result, *N*-[4-(diethylamino)phenyl]-5-methyl-3-phenylisoxazole-4-carboxamide (3, IC_{50} 3 μM) was identified by luciferase reporter assay as the most potent inhibitor, along with three other inhibitory compound families (WO2018055235, 2018). Furthermore, the tentative structure-activity relationship reveals a conserved southern part of the molecule in agonist and antagonist compounds, implying a common binding site for active compounds and further validating the significance of the compound findings (Fig. 15).

A comparison of the results from other studies conducted in the search for chemical agents active on transcription factors GATA4 or NKX2-5 shows interesting molecular overlap. Several inhibitory compounds that were identified in our study displayed structural similarity with NKX2-5-inducing compounds previously identified by Sadek et al. (2008), including 3,4,5-trisubstituted isoxazoles vs. 3,5-disubstituted isoxazoles, thiazolidinediones and flavonoids. However, the lead compound of the NKX2-5 screen (Isx-9) did not inhibit the luciferase reporter gene activity in our GATA4–NKX2-5 transcriptional synergy assay (data not shown). Furthermore, our compounds were structurally different and more drug-like in comparison to the DNA binding inhibitors of GATA4 identified by El-Hachem and Nemer (2011). Overall, our lead compound 3 possesses drug-like chemical properties (MW, rotatable bonds, acceptor/donor count, cLogP) and great synthesis accessibility, perfectly suitable for further chemical optimization toward a more efficient drug candidate.

6.3 Pharmacokinetics, metabolism and cytotoxicity

Since absorption, distribution, metabolism, excretion and toxicity (ADMET) of the compound *in vivo* have a major impact on a drug's efficacy, an evaluation of the plasma levels, metabolism and cytotoxicity of compound 3 was carried out in studies III and IV. A previous study indicates that structurally similar compounds are rapidly metabolized in 30 minutes with rat liver microsomal incubation through de-ethylation of the *N*,*N*-diethylaniline (Z. Xin et al., 2005). The relatively fast metabolism of compound 3 was confirmed *in vivo* leading to stable and abundant metabolites M3 and M4 via de-ethylation and hydroxylation reactions. Therefore,

lability and the rapid metabolic rate of the compound may demand further structural modifications to reach its full *in vivo* potential (Fig. 15).

The cytotoxicity of eight preselected compounds was studied in eight different cell lines, including H9c2 myoblast cells derived from rat myocardium, primary neonatal rat ventricular cardiomyocytes, primary neonatal rat cardiac fibroblasts, mouse embryonic fibroblasts, mouse embryonic stem cells, mESC derivatives from day 5 embryoid bodies, hiPSCs and hiPSC-derived cardiomyocytes. Compound structures were classified in two categories, the 5- or 6-ring member configurations of the southern part. Application of the 6-member ring compounds expressed lower cellular viability independent of the structural modifications in the southern, middle or northern part of the compounds. Dose-dependent cytotoxic responses of 6-ring member compounds were especially clear in the stem cell lineages, whereas 5-member ring compounds were far better tolerated in all cell lineages. The structure-toxicity relationship of the compounds suggested that a flatter geometry in the southern part of the 5-ring member compounds correlates favourably to a less toxic outcome in cellular assays. These studies demonstrate the impact of ADME and toxicity properties on the validation and selection of compound candidates (Fig. 15).

6.4 General discussion

Study III examined in detail the cardiac actions of compound 3 targeting GATA4-NKX2-5 interaction *in vitro*, and found *in vivo* cardioprotective effects on cardiac function and gene expression in experimental models of ischemic injury and pressure overload. Overall, the results in mice after a one-week treatment showed a significant improvement in the left ventricular ejection fraction and fractional shortening as well as attenuated myocardial structural changes. In addition, treatment of Wistar rats after myocardial infarction and ischemia either with compound 3 or compound 3-loaded microparticles or nanoparticles with an ANP-coating, resulted in beneficial changes in cardiac gene expression (Ferreira et al., 2017). Moreover, it was shown that *in vitro* compound 3 decreased mechanical stretch, and the hypertrophic agonist PE activated ANP and BNP gene expression at micromolar concentrations without significantly influencing the baseline ANP and BNP mRNA levels. This work is one of the first to demonstrate that chemical intervention of protein–protein interactions of key TFs may present the next compound category of therapeutics for cardiac remodeling and repair. Current therapies for the prevention of adverse cardiac remodeling target neurohumoral

activation and the reduction of the workload of the heart, and novel approaches are urgently needed. The research approach here, which targets the interaction of key TFs regulators of the hypertrophic responses, especially GATA4, represents a completely novel molecular mechanism to provide cardiac protection after myocardial infarction and other cardiac injuries. These studies of TF-influencing, cardioprotective compounds establish a research basis for the possible future academic and industrial research attempts in the cardiac field.

Numerous possible limitations relate to protein target validation and compounds that interfere with the target. Instead of giving absolute resolutions, computational methods provide estimations and enrichments of relevant options to a given research question. Unsuccessful ligand binding site selection for the GATA4 protein was noticed during the identification of protein-protein inhibitors via the virtual and luciferase reporter screening. In the case of GATA4, the underlining reason for the use of a homology model was a lack of protein crystal structures and reference compounds. Due to the nature of the computational methods, the quality of the results is also dependent on the experience and know-how of the personnel. Screening of compounds in cell-based assays, such as the luciferase reporter assay, may result in false positives due to association directly with the luciferase. Furthermore, the compounds may aggregate, have redox behavior, or have other chemical quality-related issues. As always, compounds may mediate their beneficial biological effects through unidentified target proteins. Therefore, the positive screening results obtained here will be further confirmed by various bioassays.

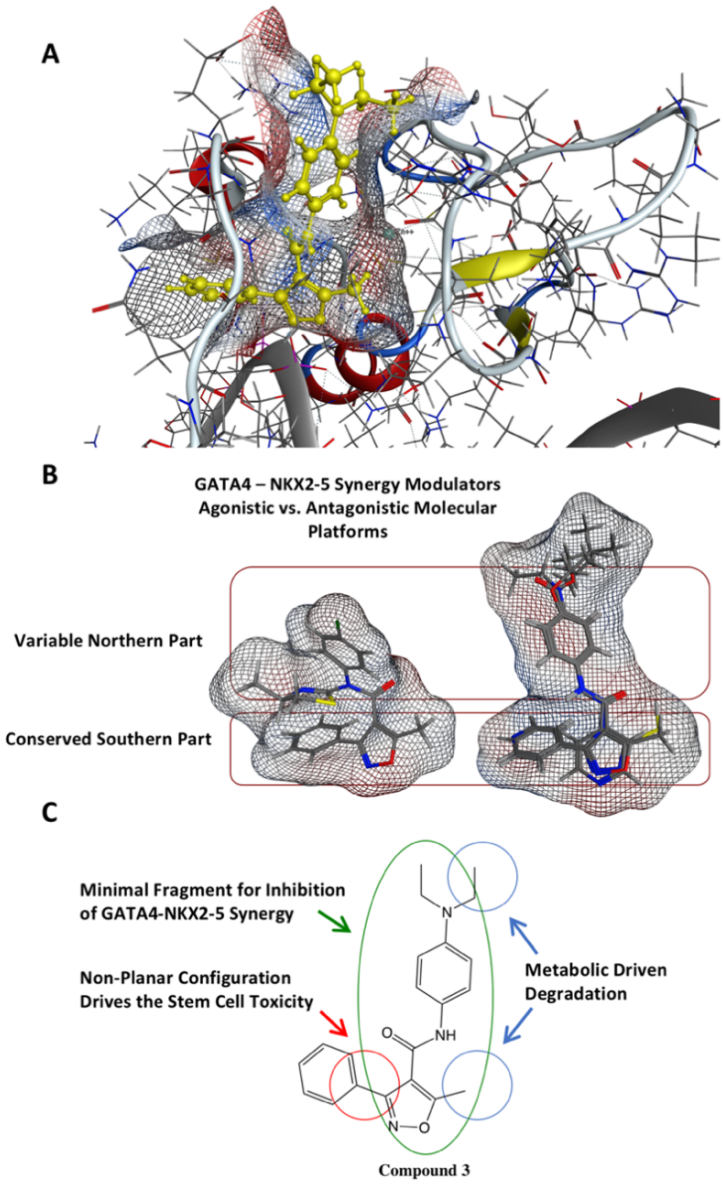

Fig. 15. Summary picture of this work. (A) Proposed binding mode of compound 3 in the C-terminal zinc finger of GATA4. (B) Comparison of molecular structures of inhibitory compounds 3, 4, and 5 and activator compound 7 reveals the structural motifs relevant for compound activity. (C) Structural determinants of compound 3 important for pharmacokinetics, metabolism and stem cell toxicity.

7 Summary and conclusions

The aim of the present study was (i) to characterize the structural architecture of the protein-protein interactions established by two cardiac transcription factors GATA4 and NKX2-5 and (ii) to evaluate small molecules having an effect on the cardiac gene expression via the modulation of the GATA4-NKX2-5 synergy. Moreover, the cardiac actions of the lead compound *in vivo* and adverse cellular effects *in vitro* were investigated. The main findings of the study are summarized as follows:

1. Structure-based design was utilized to explore the effects of protein mutations on GATA4 function. A number of amino acids, including N272, Q274, R283, K299 and R319, were identified in GATA4 to mediate the physical and functional interactions with the homeodomain of NKX2-5. Integration of the structural and experimental information suggests that the assembly of GATA4-NKX2-5 protein complex resembles the architecture of conserved DNA binding domains of the nuclear receptor family.
2. A small molecule discovery project was carried out to uncover the synthetic variants acting *in vitro* on the established GATA4-NKX2-5 assay. Various research approaches, i.e., fragment-based screening, luciferase reporter gene assay, and a pharmacophore search, were applied for small molecule screening, identification, and structure-activity optimization. Four compound families were identified either to inhibit or augment cardiac gene expression and the luciferase reporter assay in a dose-dependent manner. The validity of the most potent lead compound, *N*-[4-(diethylamino)phenyl]-5-methyl-3-phenylisoxazole-4-carboxamide (3, IC_{50} 3 μM) was further assessed by a number of confirmatory cellular assays and by extensive chemical quality control evaluations.
3. A series of proof-of-principle experiments have demonstrated the significant potential of compound 3 to provide cardiac protection after myocardial infarction and other cardiac injuries. The pharmacokinetics and metabolic degradation of compound 3 was measured after a single i.p. injection (30 mg/kg/day) in normal rats. The study revealed a rapid metabolic degradation of the parent compound via the de-ethylation and hydroxylation reactions and the formation of inactive metabolites M3 and M4 in plasma. Further optimization of the pharmacokinetics and metabolic stability of compound 3

derivatives is required to the fully reach the potential of GATA4-targeted compounds.

4. Molecular structure analysis revealed a subset of 6-member ring compounds that were consistently linked to stem cell toxicity, with uniform southern part conformation and stereochemistry defined by force field- and knowledge-based methods. Based on the present study, further compound synthesis work may guide the development of non-toxic derivatives, i.e., compounds with a 5-membered ring in the southern part of the molecule.

References

Abou-Gharbia, M., & Childers, W. E. (2014). Discovery of innovative therapeutics: today's realities and tomorrow's vision. 2. Pharma's challenges and their commitment to innovation. *Journal of Medicinal Chemistry*, *57*(13), 5525–5553. https://doi.org/10.1021/jm401564r

Addis, R. C., Ifkovits, J. L., Pinto, F., Kellam, L. D., Esteso, P., Rentschler, S., … Gearhart, J. D. (2013). Optimization of direct fibroblast reprogramming to cardiomyocytes using calcium activity as a functional measure of success. *Journal of Molecular and Cellular Cardiology*, *60*, 97–106. https://doi.org/10.1016/j.yjmcc.2013.04.004; 10.1016/j.yjmcc.2013.04.004

Akazawa, H., & Komuro, I. (2003). Roles of cardiac transcription factors in cardiac hypertrophy. *Circulation Research*, *92*(10), 1079–1088. https://doi.org/10.1161/01.RES.0000072977.86706.23

Akazawa, H., & Komuro, I. (2005). Cardiac transcription factor Csx/Nkx2-5: Its role in cardiac development and diseases. *Pharmacology & Therapeutics*, *107*(2), 252–268. https://doi.org/10.1016/j.pharmthera.2005.03.005

Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

An, W. F., & Tolliday, N. (2010). Cell-based assays for high-throughput screening. *Molecular Biotechnology*, *45*(2), 180–186. https://doi.org/10.1007/s12033-010-9251-z

Anderson, C. M., Hu, J., Thomas, R., Gainous, T. B., Celona, B., Sinha, T., … Black, B. L. (2017). Cooperative activation of cardiac transcription through myocardin bridging of paired MEF2 sites. *Development*, *144*(7), 1235–1241. https://doi.org/10.1242/dev.138487

Ang, Y.-S., Rivas, R. N., Ribeiro, A. J. S., Srivas, R., Rivera, J., Stone, N. R., … Srivastava, D. (2016). Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell*, *167*(7), 1734–1749. https://doi.org/10.1016/J.CELL.2016.11.033

Arceci, R. J., King, A. A., Simon, M. C., Orkin, S. H., & Wilson, D. B. (1993). Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Molecular and Cellular Biology*, *13*(4), 2235–2246.

Aries, A., Paradis, P., Lefebvre, C., Schwartz, R. J., & Nemer, M. (2004). Essential role of GATA-4 in cell survival and drug-induced cardiotoxicity. *Proceedings of the National Academy of Sciences*, *101*(18), 6975–6980. https://doi.org/10.1073/pnas.0401833101

Aries, A., Whitcomb, J., Shao, W., Komati, H., Saleh, M., & Nemer, M. (2014). Caspase-1 cleavage of transcription factor GATA4 and regulation of cardiac cell fate. *Cell Death & Disease*, *5*(12), e1566. https://doi.org/10.1038/cddis.2014.524

Armiñán, A., Gandía, C., Bartual, Mc., García-Verdugo, J. M., Lledó, E., Mirabet, V., … Sepúlveda, P. (2009). Cardiac Differentiation Is Driven by NKX2.5 and GATA4 Nuclear Translocation in Tissue-Specific Mesenchymal Stem Cells. *Stem Cells and Development*, *18*(6), 907–918. https://doi.org/10.1089/scd.2008.0292

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Baker, M. (2017). Fragment-based phenotypic screening is a hit. *Nature Reviews Drug Discovery*, *16*(4), 225–226. https://doi.org/10.1038/nrd.2017.56

Banerjee-Basu, S., & Baxevanis, A. D. (2001). Molecular evolution of the homeodomain family of transcription factors. *Nucleic Acids Research*, *29*(15), 3258–3269. https://doi.org/10.1093/nar/29.15.3258

Barnes, R. M., Harris, I. S., Jaehnig, E. J., Sauls, K., Sinha, T., Rojas, A., … Black, B. L. (2016). MEF2C regulates outflow tract alignment and transcriptional control of Tdgf1. *Development*, *143*(5), 774–779. https://doi.org/10.1242/dev.126383

Bastakis, E., Hedtke, B., Klermund, C., Grimm, B., & Schwechheimer, C. (2018). LLM-Domain B-GATA Transcription Factors Play Multifaceted Roles in Controlling Greening in Arabidopsis. *The Plant Cell*, *30*(3), 582–599. https://doi.org/10.1105/tpc.17.00947

Bates, D. L., Chen, Y., Kim, G., Guo, L., & Chen, L. (2008). Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. *Journal of Molecular Biology*, *381*(5), 1292–1306. https://doi.org/10.1016/j.jmb.2008.06.072

Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533. https://doi.org/10.1038/483531a

Behringer, C., & Schwechheimer, C. (2015). B-GATA transcription factors - insights into their structure, regulation, and role in plant development. *Frontiers in Plant Science*, *6*, 90. https://doi.org/10.3389/fpls.2015.00090

Belaguli, N. S., Sepulveda, J. L., Nigam, V., Charron, F., Nemer, M., & Schwartz, R. J. (2000). Cardiac tissue enriched factors serum response factor and GATA-4 are mutual coregulators. *Molecular and Cellular Biology*, *20*(20), 7550–7558.

Bogan, A. A., & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, *280*(1), 1–9. https://doi.org/10.1006/jmbi.1998.1843

Borhani, D. W., & Shaw, D. E. (2012). The future of molecular dynamics simulations in drug discovery. *Journal of Computer-Aided Molecular Design*, *26*(1), 15–26. https://doi.org/10.1007/s10822-011-9517-y

Breiten, B., Lockett, M. R., Sherman, W., Fujita, S., Al-Sayah, M., Lange, H., … Whitesides, G. M. (2013). Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *Journal of the American Chemical Society*, *135*(41), 15579–15584. https://doi.org/10.1021/ja4075776

Brivanlou, A. H., & Darnell, J. E. (2002). Signal Transduction and the Control of Gene Expression. *Science*, *295*(5556), 813–818. https://doi.org/10.1126/science.1066355

Broach, J. R., & Thorner, J. (1996). High-throughput screening for drug discovery. *Nature*, *384*, 14–16. https://doi.org/10.1038/384014a0

Bronowska, A. K. (2011). Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design. *Thermodynamics - Interaction Studies - Solids, Liquids and Gases*, 1–49. https://doi.org/10.5772/19447

Brown, D. (2007). Unfinished business: target-based drug discovery. *Drug Discovery Today*, *12*(23–24), 1007–1012. https://doi.org/10.1016/j.drudis.2007.10.017

Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Momerwell, W. D. S., Purkis, L. H., … Orpen, A. G. (2004). Retrieval of crystallographically-derived molecular geometry information. *Journal of Chemical Information and Computer Sciences*, *44*(6), 2133–2144. https://doi.org/10.1021/ci049780b

Burchfield, J. S., Xie, M., & Hill, J. A. (2013). Pathological ventricular remodeling: Mechanisms: Part 1 of 2. *Circulation*, *128*(4), 388–400. https://doi.org/10.1161/CIRCULATIONAHA.113.001878

Burris, T. P., Solt, L. A., Wang, Y., Crumbley, C., Banerjee, S., Griffett, K., … Kojetin, D. J. (2013). Nuclear Receptors and Their Selective Pharmacologic Modulators. *Pharmacological Reviews*, *65*(2), 710–778. https://doi.org/10.1124/pr.112.006833

Cahill, T. J., Choudhury, R. P., & Riley, P. R. (2017). Heart regeneration and repair after myocardial infarction: Translational opportunities for novel therapeutics. *Nature Reviews Drug Discovery*, *16*(10), 699–717. https://doi.org/10.1038/nrd.2017.106

Cao, N., Huang, Y., Zheng, J., Spencer, C. I., Zhang, Y., Fu, J.-D., … Ding, S. (2016). Conversion of human fibroblasts into functional cardiomyocytes by small molecules. *Science (New York, N.Y.)*, *352*(6290), 1216–1220. https://doi.org/10.1126/science.aaf1502

Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., … Raschellà, G. (2017). Zinc-finger proteins in health and disease. *Cell Death Discovery*, *3*, 17071. https://doi.org/10.1038/cddiscovery.2017.71

Chan, S. L., & Labute, P. (2010). Training a scoring function for the alignment of small molecules. *Journal of Chemical Information and Modeling*, *50*(9), 1724–1735. https://doi.org/10.1021/ci100227h

Chang, C. -e. A., Chen, W., & Gilson, M. K. (2007). Ligand configurational entropy and protein binding. *Proceedings of the National Academy of Sciences*, *104*(5), 1534–1539. https://doi.org/10.1073/pnas.0610494104

Charron, F., Tsimiklis, G., Arcand, M., Robitaille, L., Liang, Q., Molkentin, J. D., … Nemer, M. (2001). Tissue-specific GATA factors are transcriptional effectors of the small GTPase RhoA. *Genes & Development*, *15*(20), 2702–2719. https://doi.org/10.1101/gad.915701

Chodera, J. D., & Mobley, D. L. (2013). Entropy-Enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognition and Design. *Annual Review of Biophysics*, *42*(1), 121–142. https://doi.org/10.1146/annurev-biophys-083012-130318

Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO Journal*, *5*(4), 823–826. https://doi.org/060 fehlt

Chung, I.-M., & Rajakumar, G. (2016). Genetics of Congenital Heart Defects: The NKX2-5 Gene, a Key Player. *Genes*, *7*(2). https://doi.org/10.3390/genes7020006

Cohen, E. D., Tian, Y., & Morrisey, E. E. (2008). Wnt signaling: an essential regulator of cardiovascular differentiation, morphogenesis and progenitor self-renewal. *Development*, *135*(5), 789–798. https://doi.org/10.1242/dev.016865

Cohn, J. N., Ferrari, R., & Sharpe, N. (2000). Cardiac remodeling—concepts and clinical implications: a consensus paper from an international forum on cardiac remodeling. *Journal of the American College of Cardiology*, *35*(3), 569–582. https://doi.org/10.1016/S0735-1097(99)00630-0

Congreve, M., Carr, R., Murray, C., & Jhoti, H. (2003). A 'Rule of Three' for fragment-based lead discovery? *Drug Discovery Today*, *8*(19), 876–877. https://doi.org/10.1016/S1359-6446(03)02831-9

Copeland, R. A. (2016). The drug-target residence time model: A 10-year retrospective. *Nature Reviews Drug Discovery*, *15*(2), 87–95. https://doi.org/10.1038/nrd.2015.18

Copeland, R. A., Pompliano, D. L., & Meek, T. D. (2006). Drug-target residence time and its implications for lead optimization. *Nature Reviews Drug Discovery*, *5*(9), 730–739. https://doi.org/10.1038/nrd2082

Cowie, M. R., Jourdain, P., Maisel, A., Dahlstrom, U., Follath, F., Isnard, R., … Francs, G. (2003). Clinical applications of B-type natriuretic peptide (BNP) testing. *European Heart Journal*, *24*(19), 1710–1718. https://doi.org/10.1016/S0195-668X(03)00476-7

Dai, Y. S., Cserjesi, P., Markham, B. E., & Molkentin, J. D. (2002). The transcription factors GATA4 and dHAND physically interact to synergistically activate cardiac gene expression through a p300-dependent mechanism. *The Journal of Biological Chemistry*, *277*(27), 24390–24398. https://doi.org/10.1074/jbc.M202490200

Davies, S. G., Kennewell, P. D., Russell, A. J., Seden, P. T., Westwood, R., & Wynne, G. M. (2015). Stemistry: the control of stem cells in situ using chemistry. *Journal of Medicinal Chemistry*, *58*(7), 2863–2894. https://doi.org/10.1021/jm500838d

Davis, A. M., Plowright, A. T., & Valeur, E. (2017). Directing evolution: the next revolution in drug discovery? *Nature Reviews Drug Discovery*, *16*(10), 681. https://doi.org/10.1038/nrd.2017.146

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, *47*, 20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012

Donnet, S. (2012). Monte Carlo Simulation for the Pharmaceutical Industry: Concepts, Algorithms, and Case Studies by Mark Chang. In *International Statistical Review* (Vol. 80, pp. 186–187). CRC Press. https://doi.org/10.1111/j.1751-5823.2012.00179_10.x

Du, X., Li, Y., Xia, Y.-L., Ai, S.-M., Liang, J., Sang, P., … Liu, S.-Q. (2016). Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *International Journal of Molecular Sciences*, *17*(2), 144. https://doi.org/10.3390/ijms17020144

Durocher, D., Charron, F., Warren, R., Schwartz, R. J., & Nemer, M. (1997). The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. *The EMBO Journal*, *16*(18), 5687–5696. https://doi.org/10.1093/emboj/16.18.5687

Eder, J., Sedrani, R., & Wiesmann, C. (2014). The discovery of first-in-class drugs: Origins and evolution. *Nature Reviews Drug Discovery*, *13*(8), 577–587. https://doi.org/10.1038/nrd4336

El-Hachem, N., & Nemer, G. (2011). Identification of new GATA4-small molecule inhibitors by structure-based virtual screening. *Bioorganic & Medicinal Chemistry*, *19*(5), 1734–1742. https://doi.org/10.1016/j.bmc.2011.01.022

Erlanson, D. A. (2011). Introduction to Fragment-Based Drug Discovery. In *Topics in current chemistry* (Vol. 317, pp. 1–32). https://doi.org/10.1007/128_2011_180

Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., & Jhoti, H. (2016). Twenty years on: The impact of fragments on drug discovery. *Nature Reviews Drug Discovery*, *15*(9), 605–619. https://doi.org/10.1038/nrd.2016.109

Erlanson, D. A., McDowell, R. S., & O'Brien, T. (2004). Fragment-based drug discovery. *Journal of Medicinal Chemistry*, *47*(14), 3463–3482. https://doi.org/10.1021/jm040031v

Esposito, G., Fogolari, F., Damante, G., Formisano, S., Tell, G., Leonardi, A., … Viglino, P. (1996). Analysis of the solution structure of the homeodomain of rat thyroid transcription factor 1 by 1H-NMR spectroscopy and restrained molecular mechanics. *Eur. J. Biochem.*, *2412528*(1), 101–113. https://doi.org/10.1111/j.1432-1033.1996.0101t.x

Esteban, M. A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., … Pei, D. (2010). Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells. *Cell Stem Cell*, *6*(1), 71–79. https://doi.org/10.1016/j.stem.2009.12.001

Ferreira, M. P. A., Ranjan, S., Kinnunen, S., Correia, A., Talman, V., Mäkilä, E., … Santos, H. A. (2017). Drug-Loaded Multifunctional Nanoparticles Targeted to the Endocardial Layer of the Injured Heart Modulate Hypertrophic Signaling. *Small*, *13*(33), 1701276. https://doi.org/10.1002/smll.201701276

Freyer, M. W., & Lewis, E. A. (2008). Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. In *Methods in Cell Biology* (Vol. 84, pp. 79–113). https://doi.org/10.1016/S0091-679X(07)84004-0

Fu, Y., Huang, C., Xu, X., Gu, H., Ye, Y., Jiang, C., … Xie, X. (2015). Direct reprogramming of mouse fibroblasts into cardiomyocytes with chemical cocktails. *Cell Research*, *25*(9), 1013–1024. https://doi.org/10.1038/cr.2015.99

Garg, V., Kathiriya, I. S., Barnes, R., Schluterman, M. K., King, I. N., Butler, C. A., … Srivastava, D. (2003). GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature*, *424*(6947), 443–447. https://doi.org/10.1038/nature01827

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*(5), 449–461. https://doi.org/10.1517/17460441.2015.1032936

Gohlke, H., & Klebe, G. (2002). Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie - International Edition*, *41*(15), 2644–2676. https://doi.org/10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O

Grepin, C., Nemer, G., & Nemer, M. (1997). Enhanced cardiogenesis in embryonic stem cells overexpressing the GATA-4 transcription factor. *Development (Cambridge,*

England), *124*(12), 2387–2395.

Grunert, M., Dorn, C., & Rickert-Sperling, S. (2016). Cardiac Transcription Factors and Regulatory Networks. In *Congenital Heart Diseases: The Broken Heart* (pp. 139–152). Vienna: Springer Vienna. https://doi.org/10.1007/978-3-7091-1883-2_12

Grunwald, E., & Steel, C. (1995). Solvent Reorganization and Thermodynamic Enthalpy—Entropy Compensation. *Journal of the American Chemical Society*, *117*(21), 5687–5692. https://doi.org/10.1021/ja00126a009

Guo, W., Wisniewski, J. A., & Ji, H. (2014). Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorganic and Medicinal Chemistry Letters*, *24*(11), 2546–2554. https://doi.org/10.1016/j.bmcl.2014.03.095

Hae Jun, J., Jung Shin, E., Ho Kim, J., Oh Kim, S., Shim, J.-K., & Kwak, Y.-L. (2013). Erythropoietin Prevents Hypoxia-Induced GATA-4 Ubiquitination via Phosphorylation of Serine 105 of GATA-4. *Biol. Pharm. Bull*, *36*(367), 1126–1133. Retrieved from https://www.jstage.jst.go.jp/article/bpb/36/7/36_b13-00100/_pdf/-char/en

Hagenbuchner, J., & Ausserlechner, M. J. (2016). Targeting transcription factors by small compounds-Current strategies and future implications. *Biochemical Pharmacology*, *107*, 1–13. https://doi.org/10.1016/j.bcp.2015.12.006

Hahnefeld, C., Drewianka, S., & Herberg, F. W. (2004). Determination of kinetic data using surface plasmon resonance biosensors. *Methods in Molecular Medicine*, *94*, 299–320. https://doi.org/10.1385/1-59259-679-7:299

Hajduk, P. J. (2006). Fragment-based drug design: How big is too big? *Journal of Medicinal Chemistry*, *49*(24), 6972–6976. https://doi.org/10.1021/jm060511h

Hajduk, P. J., & Greer, J. (2007). A decade of fragment-based drug design: Strategic advances and lessons learned. *Nature Reviews Drug Discovery*, *6*(3), 211–219. https://doi.org/10.1038/nrd2220

Hajduk, P. J., Huth, J. R., & Tse, C. (2005). Predicting protein druggability. *Drug Discovery Today*, *10*(23–24), 1675–1682. https://doi.org/10.1016/S1359-6446(05)03624-X

Hasegawa, K., Lee, S. J., Jobe, S. M., Markham, B. E., & Kitsis, R. N. (1997). cis-Acting sequences that mediate induction of beta-myosin heavy chain gene expression during left ventricular hypertrophy due to aortic constriction. *Circulation*, *96*(11), 3943–3953. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9403619

Hautala, N., Tokola, H., Luodonpää, M., Puhakka, J., Romppanen, H., Vuolteenaho, O., & Ruskoaho, H. (2001). Pressure overload increases GATA4 binding activity via endothelin-1. *Circulation*, *103*(5), 730–735. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11156886

He, A., Gu, F., Hu, Y., Ma, Q., Ye, L. Y., Akiyama, J. A., … Pu, W. T. (2014). Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nature Communications*, *5*, 4907. https://doi.org/10.1038/ncomms5907

He, A., Kong, S. W., Ma, Q., & Pu, W. T. (2011). Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5632–5637. https://doi.org/10.1073/pnas.1016959108

He, J.-G., Li, H.-R., Han, J.-X., Li, B.-B., Yan, D., Li, H.-Y., … Luo, Y. (2018). GATA-4-expressing mouse bone marrow mesenchymal stem cells improve cardiac function after myocardial infarction via secreted exosomes. *Scientific Reports*, *8*(1), 9047. https://doi.org/10.1038/s41598-018-27435-9

Heikinheimo, M., Scandrett, J. M., & Wilson, D. B. (1994). Localization of transcription factor GATA-4 to regions of the mouse embryo involved in cardiac development. *Developmental Biology*, *164*(2), 361–373. https://doi.org/10.1006/dbio.1994.1206

Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., Nagai, R., & Komuro, I. (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nature Genetics*, *28*(3), 276–280. https://doi.org/10.1038/90123

Holdgate, G. A., & Ward, W. H. J. (2005). Measurements of binding thermodynamics in drug discovery. *Drug Discovery Today*, *10*(22), 1543–1550. https://doi.org/10.1016/S1359-6446(05)03610-X

Horb, M. E., & Thomsen, G. H. (1999). Tbx5 is essential for heart development. *Development (Cambridge, England)*, *126*(8), 1739–1751.

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., … Deng, H. (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science*, *341*(6146), 651–654. https://doi.org/10.1126/science.1239278

Huang, C., Tu, W., Fu, Y., Wang, J., & Xie, X. (2018). Chemical-induced cardiac reprogramming in vivo. *Cell Research*, *28*(6), 686–689. https://doi.org/10.1038/s41422-018-0036-4

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A. E., & Melton, D. A. (2008). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnology*, *26*(7), 795–797. https://doi.org/10.1038/nbt1418

Hughes, J. P., Rees, S. S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, *162*(6), 1239–1249. https://doi.org/10.1111/j.1476-5381.2010.01127.x

Ieda, M., Fu, J. D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B. G., & Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell*, *142*(3), 375–386. https://doi.org/10.1016/j.cell.2010.07.002; 10.1016/j.cell.2010.07.002

Inagawa, K., Miyamoto, K., Yamakawa, H., Muraoka, N., Sadahiro, T., Umei, T., … Ieda, M. (2012). Induction of cardiomyocyte-like cells in infarct hearts by gene transfer of Gata4, Mef2c, and Tbx5. *Circulation Research*, *111*(9), 1147–1156. https://doi.org/10.1161/CIRCRESAHA.112.271148; 10.1161/CIRCRESAHA.112.271148

Ioannidis, J. P. A. (2018). The Proposal to Lower P Value Thresholds to .005. *JAMA*, *319*(14), 1429. https://doi.org/10.1001/jama.2018.1536

Ip, H. S., Wilson, D. B., Heikinheimo, M., Tang, Z., Ting, C. N., Simon, M. C., … Parmacek, M. S. (1994). The GATA-4 transcription factor transactivates the cardiac muscle-specific troponin C promoter-enhancer in nonmuscle cells. *Molecular and Cellular Biology*, *14*(11), 7517–7526. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/7935467

Iyer, L. M., Nagarajan, S., Woelfer, M., Schoger, E., Khadjeh, S., Zafiriou, M. P., … Zelarayán, L. C. (2018). A context-specific cardiac β-catenin and GATA4 interaction influences TCF7L2 occupancy and remodels chromatin driving disease progression in the adult heart. *Nucleic Acids Research*, *46*(6), 2850–2867. https://doi.org/10.1093/nar/gky049

Izzo, F., Mercogliano, F., Venturutti, L., Tkach, M., Inurrigarro, G., Schillaci, R., … Proietti, C. J. (2014). Progesterone receptor activation downregulates GATA3 by transcriptional repression and increased protein turnover promoting breast tumor growth. *Breast Cancer Research : BCR*, *16*(6), 491. https://doi.org/10.1186/s13058-014-0491-x

Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., … Coulombe, B. (2007). Systematic Analysis of the Protein Interaction Network for the Human Transcription Machinery Reveals the Identity of the 7SK Capping Enzyme. *Molecular Cell*, *27*(2), 262–274. https://doi.org/10.1016/j.molcel.2007.06.027

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)*, *316*(5830), 1497–1502. https://doi.org/10.1126/science.1141319

Jorgensen, W. L., & Tirado-Rives, J. (1996). Monte Carlo vs molecular dynamics for conformational sampling. *Journal of Physical Chemistry*, *100*(34), 14508–14513. https://doi.org/10.1021/jp960880x

Joseph, R., Orlov, Y. L., Huss, M., Sun, W., Li Kong, S., Ukil, L., … Liu, E. T. (2010). Integrative model of genomic factors for determining binding site selection by estrogen receptor-α. *Molecular Systems Biology*, *6*, 456. https://doi.org/10.1038/msb.2010.109

Kaplan, T., Li, X.-Y., Sabo, P. J., Thomas, S., Stamatoyannopoulos, J. A., Biggin, M. D., & Eisen, M. B. (2011). Quantitative Models of the Mechanisms That Control Genome-Wide Patterns of Transcription Factor Binding during Early Drosophila Development. *PLoS Genetics*, *7*(2), e1001290. https://doi.org/10.1371/journal.pgen.1001290

Kasahara, H., Usheva, A., Ueyama, T., Aoki, H., Horikoshi, N., & Izumo, S. (2001). Characterization of homo- and heterodimerization of cardiac Csx/Nkx2.5 homeoprotein. *The Journal of Biological Chemistry*, *276*(7), 4570–4580. https://doi.org/10.1074/jbc.M004995200

Kawamura, T., Ono, K., Morimoto, T., Wada, H., Hirai, M., Hidaka, K., … Hasegawa, K. (2005). Acetylation of GATA-4 is involved in the differentiation of embryonic stem cells into cardiac myocytes. *The Journal of Biological Chemistry*, *280*(20), 19682–19688. https://doi.org/10.1074/jbc.M412428200

Kelley, C., Blumberg, H., Zon, L. I., & Evans, T. (1993). GATA-4 is a novel transcription factor expressed in endocardium of the developing heart. *Development (Cambridge, England)*, *118*(3), 817–827. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8076520

Kerkelä, R., Pikkarainen, S., Majalahti-Palviainen, T., Tokola, H., & Ruskoaho, H. (2002). Distinct roles of mitogen-activated protein kinase pathways in GATA-4 transcription

factor-mediated regulation of B-type natriuretic peptide gene. *The Journal of Biological Chemistry*, *277*(16), 13752–13760. https://doi.org/10.1074/jbc.M105736200

Keseru, G. M., & Makara, G. M. (2006). Hit discovery and hit-to-lead approaches. *Drug Discovery Today*, *11*(15–16), 741–748. https://doi.org/10.1016/j.drudis.2006.06.016

Khalil, A., Tanos, R., El-Hachem, N., Kurban, M., Bouvagnet, P., Bitar, F., & Nemer, G. (2017). A HAND to TBX5 Explains the Link between Thalidomide and Cardiac Diseases. *Scientific Reports*, *7*(1), 1416. https://doi.org/10.1038/s41598-017-01641-3

Kikuchi, K., Holdway, J. E., Werdich, A. A., Anderson, R. M., Fang, Y., Egnaczyk, G. F., … Poss, K. D. (2010). Primary contribution to zebrafish heart regeneration by gata4 + cardiomyocytes. *Nature*, *464*(7288), 601–605. https://doi.org/10.1038/nature08804

Kimelman, D. (2006). Mesoderm induction: from caps to chips. *Nature Reviews. Genetics*, *7*(5), 360–372. https://doi.org/10.1038/nrg1837

Kinnunen, S., Tölli, M., Välimäki, M., Jumppanen, M., Boije Af Gennäs, G., Yli-Kauhaluoma, J., & Ruskoaho, H. (2018). *WO2018055235*. Finland.

Kirkpatrick, P., & Ellis, C. (2004). Chemical space. *Nature*, *432*(7019), 823–823. https://doi.org/10.1038/432823a

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery*, *3*(11), 935–949. https://doi.org/10.1038/nrd1549

Kobayashi, S., Mao, K., Zheng, H., Wang, X., Patterson, C., O'Connell, T. D., & Liang, Q. (2007). Diminished GATA4 protein levels contribute to hyperglycemia-induced cardiomyocyte injury. *The Journal of Biological Chemistry*, *282*(30), 21945–21952. https://doi.org/10.1074/jbc.M703048200

Kowalski, K., Czolij, R., King, G. F., Crossley, M., & Mackay, J. P. (1999). The solution structure of the N-terminal zinc finger of GATA-1 reveals a specific binding face for the transcriptional co-factor FOG. *Journal of Biomolecular NMR*, *13*(3), 249–262. https://doi.org/10.1023/A:1008309602929

Krieger, E., Nabuurs, S. B., & Vriend, G. (1963). *Homology modeling*. Retrieved from http://www.cmbi.ru.nl/edu/bioinf4/articles/homologymodeling.pdf

Krishna, S. S., Majumdar, I., & Grishin, N. V. (2003). Structural classification of zinc fingers: SURVEY AND SUMMARY. *Nucleic Acids Research*, *31*(2), 532–550. https://doi.org/10.1093/nar/gkg161

Labute, P. (2008). The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. *Journal of Computational Chemistry*, *29*(10), 1693–1698. https://doi.org/10.1002/jcc.20933

Labute, P. (2010). LowModeMD - Implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. *Journal of Chemical Information and Modeling*, *50*(5), 792–800. https://doi.org/10.1021/ci900508k

Latchman, D. S. (1997). Transcription factors: An overview. *The International Journal of Biochemistry & Cell Biology*, *29*(12), 1305–1312. https://doi.org/10.1016/S1357-2725(97)00085-X

Laverriere, A. C., MacNeill, C., Mueller, C., Poelmann, R. E., Burch, J. B., & Evans, T. (1994). GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. *The Journal of Biological Chemistry*, *269*(37), 23177–23184. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8083222

Leach, A. R., Shoichet, B. K., & Peishoff, C. E. (2006). Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *Journal of Medicinal Chemistry*, *49*(20), 5851–5855. https://doi.org/10.1021/jm060999m

Lee, Y., Shioi, T., Kasahara, H., Jobe, S. M., Wiese, R. J., Markham, B. E., & Izumo, S. (1998). The cardiac tissue-restricted homeobox protein Csx/Nkx2.5 physically associates with the zinc finger protein GATA4 and cooperatively activates atrial natriuretic factor gene expression. *Molecular and Cellular Biology*, *18*(6), 3120–3129.

Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature*, *424*(6945), 147–151. https://doi.org/10.1038/nature01763

Li, T., Zhang, X., Jiang, K., Liu, J., & Liu, Z. (2018). Dural effects of oxidative stress on cardiomyogenesis via Gata4 transcription and protein ubiquitination. *Cell Death & Disease*, *9*(2), y. https://doi.org/10.1038/s41419-018-0281-y

Li, Y., Zhang, Q., Yin, X., Yang, W., Du, Y., Hou, P., … Deng, H. (2011). Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. *Cell Research*, *21*(1), 196–204. https://doi.org/10.1038/cr.2010.142

Liang, Q., & Molkentin, J. D. (2002). Divergent Signaling Pathways Converge on GATA4 to Regulate Cardiac Hypertrophic Gene Expression. *J Mol Cell Cardiol*, *34*, 611–616. https://doi.org/10.1006

Lincker, H., Ziogas, C., Carr, M., Porta, N., & Eichler, H.-G. (2014). Where do new medicines originate from in the EU? *Nature Reviews Drug Discovery*, *13*(2), 92–93. https://doi.org/10.1038/nrd4232

Lionta, E., Spyrou, G., Vassilatis, D., & Cournia, Z. (2014). Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Current Topics in Medicinal Chemistry*, *14*(16), 1923–1938. https://doi.org/10.2174/1568026614666140929124445

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, *46*(1–3), 3–26. https://doi.org/10.1016/S0169-409X(00)00129-0

Liu, J., Faeder, J. R., & Camacho, C. J. (2009). Toward a quantitative theory of intrinsically disordered proteins and their function. *Proceedings of the National Academy of Sciences*, *106*(47), 19819–19823. https://doi.org/10.1073/pnas.0907710106

Liu, M., & Wang, S. M. (1999). MCDOCK: A Monte Carlo simulation approach to the molecular docking problem. *Journal of Computer-Aided Molecular Design*, *13*(5), 435–451. Retrieved from https://link.springer.com/content/pdf/10.1023/A:1008005918983.pdf

Lopes, P. E. M., Guvench, O., & MacKerell, A. D. (2014). Current status of protein force fields for molecular dynamics simulations. In *Molecular Modeling of Proteins: Second Edition* (Vol. 1215, pp. 47–71). NIH Public Access.

https://doi.org/10.1007/978-1-4939-1465-4_3
Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., … Richardson, D. C. (2003). Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins*, *50*(3), 437–450. https://doi.org/10.1002/prot.10286
Lu, H., & Tonge, P. J. (2010). Drug-target residence time: Critical information for lead optimization. *Current Opinion in Chemical Biology*, *14*(4), 467–474. https://doi.org/10.1016/j.cbpa.2010.06.176
Luna-Zurita, L., Stirnimann, C. U., Glatt, S., Kaynak, B. L., Thomas, S., Baudin, F., … Bruneau, B. G. (2016). Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. *Cell*, *164*(5), 999–1014. https://doi.org/10.1016/j.cell.2016.01.004
Lurie, L. J., Boyer, M. E., Grass, J. A., & Bresnick, E. H. (2007). Differential GATA Factor Stabilities: Implications for Chromatin Occupancy by Structurally Similar Transcription Factors†. *Biochemistry*, *47*(3), 859–869. https://doi.org/10.1021/BI701692P
Lusis, A. J., Weiss, J. N., Peralta, C., Lettre, G., Vittinghoff, E., Lubitz, S., … Zamorano, J. (2010). Cardiovascular networks: systems-based approaches to cardiovascular disease. *Circulation*, *121*(1), 157–170. https://doi.org/10.1161/CIRCULATIONAHA.108.847699
Lyons, I., Parsons, L. M., Hartley, L., Li, R., Andrews, J. E., Robb, L., & Harvey, R. P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes & Development*, *9*(13), 1654–1666.
MacArron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., … Sittampalam, G. S. (2011). Impact of high-throughput screening in biomedical research. *Nature Reviews Drug Discovery*, *10*(3), 188–195. https://doi.org/10.1038/nrd3368
MacQuarrie, K. L., Fong, A. P., Morse, R. H., & Tapscott, S. J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. *Trends in Genetics : TIG*, *27*(4), 141–148. https://doi.org/10.1016/j.tig.2011.01.001
Maisel, A. S., Krishnaswamy, P., Nowak, R. M., McCord, J., Hollander, J. E., Duc, P., … McCullough, P. A. (2002). Rapid Measurement of B-Type Natriuretic Peptide in the Emergency Diagnosis of Heart Failure. *New England Journal of Medicine*, *347*(3), 161–167. https://doi.org/10.1056/NEJMoa020233
Maitra, M., Schluterman, M. K., Nichols, H. A., Richardson, J. A., Lo, C. W., Srivastava, D., & Garg, V. (2009). Interaction of Gata4 and Gata6 with Tbx5 is critical for normal cardiac development. *Developmental Biology*, *326*(2), 368–377. https://doi.org/10.1016/j.ydbio.2008.11.004
Malek Mohammadi, M., Kattih, B., Grund, A., Froese, N., Korf-Klingebiel, M., Gigina, A., … Heineke, J. (2017). The transcription factor GATA4 promotes myocardial regeneration in neonatal mice. *EMBO Molecular Medicine*, *9*(2), 265–279. https://doi.org/10.15252/emmm.201606602
Mali, P., Chou, B. K., Yen, J., Ye, Z., Zou, J., Dowey, S., … Cheng, L. (2010). Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting

epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells*, *28*(4), 713–720. https://doi.org/10.1002/stem.402

Marttila, M., Hautala, N., Paradis, P., Toth, M., Vuolteenaho, O., Nemer, M., & Ruskoaho, H. (2001). GATA4 Mediates Activation of the B-Type Natriuretic Peptide Gene Expression in Response to Hemodynamic Stress. *Endocrinology*, *142*(11), 4693–4700. https://doi.org/10.1210/endo.142.11.8468

Mathison, M., Singh, V. P., Sanagasetti, D., Yang, L., Pinnamaneni, J. P., Yang, J., & Rosengart, T. K. (2017). Cardiac reprogramming factor Gata4 reduces postinfarct cardiac fibrosis through direct repression of the profibrotic mediator snail. *The Journal of Thoracic and Cardiovascular Surgery*, *154*(5), 1601–1610.e3. https://doi.org/10.1016/j.jtcvs.2017.06.035

Mattapally, S., Singh, M., Murthy, K. S., Asthana, S., Banerjee, S. K., Mattapally, S., … Banerjee, S. K. (2018). Computational modeling suggests impaired interactions between NKX2.5 and GATA4 in individuals carrying a novel pathogenic D16N NKX2.5 mutation. *Oncotarget*, *9*(17), 13713–13732. https://doi.org/10.18632/oncotarget.24459

McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded proteins. *Nature*, *267*(5612), 585–590. https://doi.org/10.1038/267585a0

McMurray, J. J., & Pfeffer, M. A. (2005). Heart failure. *The Lancet*, *365*(9474), 1877–1889. https://doi.org/10.1016/S0140-6736(05)66621-4

Metra, M., & Teerlink, J. R. (2017). Heart failure. *The Lancet*, *390*(10106), 1981–1995. https://doi.org/10.1016/S0140-6736(17)31071-1

Mikkelsen, T. S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., … Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. *Nature*, *454*(7200), 49–55. https://doi.org/10.1038/nature07056

Minegishi, N., Suzuki, N., Kawatani, Y., Shimizu, R., & Yamamoto, M. (2005). Rapid turnover of GATA-2 via ubiquitin-proteasome protein degradation pathway. *Genes to Cells*, *10*(7), 693–704. https://doi.org/10.1111/j.1365-2443.2005.00864.x

Miranker, A., & Karplus, M. (1991). Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins*, *11*(1), 29–34. https://doi.org/10.1002/prot.340110104

Moffat, J. G., Rudolph, J., & Bailey, D. (2014). Phenotypic screening in cancer drug discovery — past, present and future. *Nature Reviews Drug Discovery*, *13*(8), 588–602. https://doi.org/10.1038/nrd4366

Mohamed, T. M. A., Stone, N. R., Berry, E. C., Radzinsky, E., Huang, Y., Pratt, K., … Srivastava, D. (2017). Chemical Enhancement of In Vitro and In Vivo Direct Cardiac Reprogramming. *Circulation*, *135*, 978–995. https://doi.org/10.1161/CIRCULATIONAHA.116.024692

Molkentin, J. D. (2000). The zinc finger-containing transcription factors GATA-4, -5, and -6: Ubiquitously expressed regulators of tissue-specific gene expression. *The Journal of Biological Chemistry*, *275*(50), 38949–38952. https://doi.org/10.1074/jbc.R000029200

Molkentin, J. D., Kalvakolanu, D. V, & Markham, B. E. (1994). Transcription factor GATA-

4 regulates cardiac muscle-specific expression of the alpha-myosin heavy-chain gene. *Molecular and Cellular Biology*, *14*(7), 4947–4957. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8007990

Molkentin, J. D., Lin, Q., Duncan, S. A., & Olson, E. N. (1997). Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes & Development*, *11*(8), 1061–1072.

Mori, A. D., & Bruneau, B. G. (2004). TBX5 mutations and congenital heart disease: Holt-Oram syndrome revealed. *Current Opinion in Cardiology*, *19*(3), 211–215. https://doi.org/10.1097/00001573-200405000-00004

Morin, S., Charron, F., Robitaille, L., & Nemer, M. (2000). GATA-dependent recruitment of MEF2 proteins to target promoters. *The EMBO Journal*, *19*(9), 2046–2055. https://doi.org/10.1093/emboj/19.9.2046

Morisco, C., Seta, K., Hardt, S. E., Lee, Y., Vatner, S. F., & Sadoshima, J. (2001). Glycogen Synthase Kinase 3β Regulates GATA4 in Cardiac Myocytes. *Journal of Biological Chemistry*, *276*(30), 28586–28597. https://doi.org/10.1074/jbc.M103166200

Morrisey, E., Ip, H., Lu, M., & Parmacek, M. (1996). GATA-6: A Zinc Finger Transcription Factor That Is Expressed in Multiple Cell Lineages Derived from Lateral Mesoderm. *Developmental Biology*, *177*(1), 309–322. https://doi.org/10.1006/dbio.1996.0165

Morrisey, E., Ip, H., Tang, Z., & Parmacek, M. (1997). GATA-4 activates transcription via two novel domains that are conserved within the GATA-4/5/6 subfamily. *The Journal of Biological Chemistry*, *272*(13), 8515–8524. https://doi.org/10.1074/JBC.272.13.8515

Mullard, A. (2015). The phenotypic screening pendulum swings. *Nature Reviews Drug Discovery*, *14*(12), 807–809. https://doi.org/10.1038/nrd4783

Murphy, A. M., Thompson, W. R., Peng, L. F., & Jones, L. (1997). Regulation of the rat cardiac troponin I gene by the transcription factor GATA-4. *The Biochemical Journal*, *322 (Pt 2*, 393–401. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9065755

Nam, Y. J., Song, K., & Olson, E. N. (2012). *Patent WO2012/116064A1*.

Nicholas, S. B., & Philipson, K. D. (1999). Cardiac expression of the Na(+)/Ca(2+) exchanger NCX1 is GATA factor dependent. *The American Journal of Physiology*, *277*(1 Pt 2), 324–330. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10409212

Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature*, *506*(7487), 150–152. https://doi.org/10.1038/506150a

Oka, T., Maillet, M., Watt, A. J., Schwartz, R. J., Aronow, B. J., Duncan, S. A., & Molkentin, J. D. (2006). Cardiac-specific deletion of gata4 reveals its requirement for hypertrophy, compensation, and myocyte viability. *Circulation Research*, *98*(6), 837–845. https://doi.org/10.1161/01.RES.0000215985.18538.c4

Olson, E. N. (2006). Gene regulatory networks in the evolution and development of the heart. *Science*, *313*(5795), 1922–1927. https://doi.org/10.1126/science.1132292

Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., … Gronenborn, A. M. (1993). NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science (New York, N.Y.)*, *261*(5120),

438–446. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8332909

Orkin, S. H. (1992). GATA-binding transcription factors in hematopoietic cells. *Blood*, *80*(3), 575–581. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1638017

Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there? *Nature Reviews Drug Discovery*, *5*(12), 993–996. https://doi.org/10.1038/nrd2199

Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., & Schacht, A. L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, *9*(3), 203. https://doi.org/10.1038/nrd3078

Perozzo, R., Folkers, G., & Scapozza, L. (2004). Thermodynamics of protein-ligand interactions: History, presence, and future aspects. *Journal of Receptors and Signal Transduction*, *24*(1–2), 1–52. https://doi.org/10.1081/RRS-120037896

Peterkin, T., Gibson, A., & Patient, R. (2007). Redundancy and evolution of GATA factor requirements in development of the myocardium. *Developmental Biology*, *311*(2), 623–635. https://doi.org/10.1016/j.ydbio.2007.08.018

Pikkarainen, S., Tokola, H., Kerkelä, R., & Ruskoaho, H. (2004). GATA transcription factors in the developing and adult heart. *Cardiovascular Research*, *63*(2), 196–207. https://doi.org/10.1016/j.cardiores.2004.03.025

Pikkarainen, S., Tokola, H., Majalahti-Palviainen, T., Kerkelä, R., Hautala, N., Bhalla, S. S., … Ruskoaho, H. (2003). GATA-4 is a nuclear mediator of mechanical stretch-activated hypertrophic program. *The Journal of Biological Chemistry*, *278*(26), 23807–23816. https://doi.org/10.1074/jbc.M302719200

Potthoff, M. J., & Olson, E. N. (2007). MEF2: a central regulator of diverse developmental programs. *Development*, *134*(23), 4131–4140. https://doi.org/10.1242/dev.008367

Pradhan, L., Gopal, S., Li, S., Ashur, S., Suryanarayanan, S., Kasahara, H., & Nam, H.-J. (2016). Intermolecular Interactions of Cardiac Transcription Factors NKX2.5 and TBX5. *Biochemistry*, *55*(12), 1702–1710. https://doi.org/10.1021/acs.biochem.6b00171

Qian, L., Huang, Y., Spencer, C. I., Foley, A., Vedantham, V., Liu, L., … Srivastava, D. (2012). In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. *Nature*, *485*(7400), 593–598. https://doi.org/10.1038/nature11044; 10.1038/nature11044

Raha, K., Peters, M. B., Wang, B., Yu, N., Wollacott, A. M., Westerhoff, L. M., & Merz, K. M. (2007). The role of quantum mechanics in structure-based drug design. *Drug Discovery Today*, *12*(17–18), 725–731. https://doi.org/10.1016/j.drudis.2007.07.006

Reynolds, C. H., & Holloway, M. K. (2011). Thermodynamics of ligand binding and efficiency. *ACS Medicinal Chemistry Letters*, *2*(6), 433–437. https://doi.org/10.1021/ml200010k

Rojas, A., Kong, S. W., Agarwal, P., Gilliss, B., Pu, W. T., & Black, B. L. (2008). GATA4 Is a Direct Transcriptional Activator of Cyclin D2 and Cdk4 and Is Required for Cardiomyocyte Proliferation in Anterior Heart Field-Derived Myocardium. *Molecular and Cellular Biology*, *28*(17), 5420–5431.

https://doi.org/10.1128/MCB.00717-08
Russell, J., Goetsch, S., Aguilar, H., Frantz, D., & Schneider, J. (2012). Targeting native adult heart progenitors with cardiogenic small molecules. *ACS Chemical Biology*, *7*(6), 1067–1076. https://doi.org/10.1021/cb200525q
Russell, J., Goetsch, S. C., Aguilar, H. R., Coe, H., Luo, X., Liu, N., … Schneider, J. W. (2012). Regulated expression of pH sensing G Protein-coupled receptor-68 identified through chemical biology defines a new drug target for ischemic heart disease. *ACS Chemical Biology*, *7*(6), 1077–1083. https://doi.org/10.1021/cb300001m
Rysä, J., Tenhunen, O., Serpi, R., Soini, Y., Nemer, M., Leskinen, H., & Ruskoaho, H. (2010). GATA-4 is an angiogenic survival factor of the infarcted heart. *Circulation. Heart Failure*, *3*(3), 440–450. https://doi.org/10.1161/CIRCHEARTFAILURE.109.889642
Saadane, N., Alpert, L., & Chalifour, L. E. (1999). Expression of immediate early genes, GATA-4, and Nkx-2.5 in adrenergic-induced cardiac hypertrophy and during regression in adult mice. *British Journal of Pharmacology*, *127*(5), 1165–1176. https://doi.org/10.1038/sj.bjp.0702676
Sadek, H., Hannack, B., Choe, E., Wang, J., Latif, S., Garry, M. G., … Schneider, J. W. (2008). Cardiogenic small molecules that enhance myocardial repair by stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(16), 6063–6068. https://doi.org/10.1073/pnas.0711507105
Sams-Dodd, F. (2005). Target-based drug discovery: Is something wrong? *Drug Discovery Today*, *10*(2), 139–147. https://doi.org/10.1016/S1359-6446(04)03316-1
Santos, R., Ursu, O., Gaulton, A., Bento, A. P., Donadi, R. S., Bologa, C. G., … Overington, J. P. (2016). A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery*, *16*(1), 19–34. https://doi.org/10.1038/nrd.2016.230
Schade, D., & Plowright, A. T. (2015). Medicinal Chemistry Approaches to Heart Regeneration. *Journal of Medicinal Chemistry*, *58*(24), 9451–9479. https://doi.org/10.1021/acs.jmedchem.5b00446
Schirle, M., & Jenkins, J. L. (2016). Identifying compound efficacy targets in phenotypic drug discovery. *Drug Discovery Today*, *21*(1), 82–89. https://doi.org/10.1016/j.drudis.2015.08.001
Schlesinger, J., Schueler, M., Grunert, M., Fischer, J. J., Zhang, Q., Krueger, T., … Sperling, S. R. (2011). The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, Histone modifications, and microRNAs. *PLoS Genetics*, *7*(2), e1001313. https://doi.org/10.1371/journal.pgen.1001313
Schreiber, S. L., Kotz, J. D., Li, M., Aubé, J., Austin, C. P., Reed, J. C., … Emmitte, K. A. (2015). Advancing biological understanding and therapeutics discovery with small-molecule probes. *Cell*, *161*(6), 1252–1265. https://doi.org/10.1016/j.cell.2015.05.023
Schuck, P. (1997). Reliable determination of binding affinity and kinetics using surface plasmon resonance biosensors. *Current Opinion in Biotechnology*, *8*(4), 498–502. https://doi.org/10.1016/S0958-1669(97)80074-2
Schuetz, D. A., de Witte, W. E. A., Wong, Y. C., Knasmueller, B., Richter, L., Kokh, D. B., … Ecker, G. F. (2017). Kinetics for Drug Discovery: an industry-driven effort to

target drug residence time. *Drug Discovery Today*, *22*(6), 896–911. https://doi.org/10.1016/j.drudis.2017.02.002

Schuhmacher, A., Gassmann, O., & Hinder, M. (2016). Changing R&D models in research-based pharmaceutical companies. *Journal of Translational Medicine*, *14*(1), 105. https://doi.org/10.1186/s12967-016-0838-4

Schwabe, J. W., Chapman, L., Finch, J. T., & Rhodes, D. (1993). The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell*, *75*(3), 567–578. https://doi.org/0092-8674(93)90390-C

Scott, D. E., Bayly, A. R., Abell, C., & Skidmore, J. (2016). Small molecules, big targets: Drug discovery faces the protein-protein interaction challenge. *Nature Reviews Drug Discovery*, *15*(8), 533–550. https://doi.org/10.1038/nrd.2016.29

Sepulveda, J. L., Belaguli, N., Nigam, V., Chen, C. Y., Nemer, M., & Schwartz, R. J. (1998). GATA-4 and Nkx-2.5 coactivate Nkx-2 DNA binding targets: role for regulating early cardiac gene expression. *Molecular and Cellular Biology*, *18*(6), 3405–3415.

Serpooshan, V., Liu, Y.-H., Buikema, J. W., Galdos, F. X., Chirikian, O., Paige, S., … Wu, S. M. (2017). Nkx2.5+ Cardiomyoblasts Contribute to Cardiomyogenesis in the Neonatal Heart. *Scientific Reports*, *7*(1), 12590. https://doi.org/10.1038/s41598-017-12869-4

Serrano, F., Calatayud, C. F., Blazquez, M., Torres, J., Castell, J. V, & Bort, R. (2013). Gata4 blocks somatic cell reprogramming by directly repressing Nanog. *Stem Cells*, *31*(1), 71–82. https://doi.org/10.1002/stem.1272

Shirts, M. R., Mobley, D. L., & Brown, S. P. (2010). Free-energy calculations in structure-based drug design. *Drug Design*, 61–86. https://doi.org/http://dx.doi.org/10.1017/CBO9780511730412.007

Shoichet, B. K. (2004). Virtual screening of chemical libraries. *Nature*, *432*(7019), 862–865. https://doi.org/10.1038/nature03197

Shoichet, B. K., Baase, W. A., Kuroki, R., & Matthews, B. W. (1995). A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences*, *92*(2), 452–456. https://doi.org/10.1073/pnas.92.2.452

Shu, J., Zhang, K., Zhang, M., Yao, A., Shao, S., Du, F., … Deng, H. (2015). GATA family members as inducers for cellular reprogramming to pluripotency. *Cell Research*, *25*(2), 169–180. https://doi.org/10.1038/cr.2015.6

Singh, M. K., Li, Y., Li, S., Cobb, R. M., Zhou, D., Lu, M. M., … Gruber, P. J. (2010). Gata4 and Gata5 cooperatively regulate cardiac myocyte proliferation in mice. *Journal of Biological Chemistry*, *285*(3), 1765–1772. https://doi.org/10.1074/jbc.M109.038539

Slabinski, L., Jaroszewski, L., Rodrigues, A. P. C., Rychlewski, L., Wilson, I. A., Lesley, S. A., & Godzik, A. (2007). The challenge of protein structure determination--lessons from structural genomics. *Protein Science*, *16*(11), 2472–2482. https://doi.org/10.1110/ps.073037907

Soga, S., Shirai, H., Koborv, M., & Hirayama, N. (2007). Use of amino acid composition to predict ligand-binding sites. *Journal of Chemical Information and Modeling*, *47*(2),

400–406. https://doi.org/10.1021/ci6002202

Song, K., Nam, Y. J., Luo, X., Qi, X., Tan, W., Huang, G. N., … Olson, E. N. (2012). Heart repair by reprogramming non-myocytes with cardiac transcription factors. *Nature*, *485*(7400), 599–604. https://doi.org/10.1038/nature11139; 10.1038/nature11139

Spitz, F., & Furlong, E. E. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews.Genetics*, *13*(9), 613–626. https://doi.org/10.1038/nrg3207

Srivastava, D., & Ieda, M. (2012). Critical factors for cardiac reprogramming. *Circulation Research*, *111*(1), 5–8. https://doi.org/10.1161/CIRCRESAHA.112.271452; 10.1161/CIRCRESAHA.112.271452

Steimle, J. D., & Moskowitz, I. P. (2017). TBX5: A Key Regulator of Heart Development. In *Current Topics in Developmental Biology* (Vol. 122, pp. 195–221). NIH Public Access. https://doi.org/10.1016/bs.ctdb.2016.08.008

Stumpf, M. P. H., Thorne, T., de Silva, E., Stewart, R., An, H. J., Lappe, M., & Wiuf, C. (2008). Estimating the size of the human interactome. *Proceedings of the National Academy of Sciences*, *105*(19), 6959–6964. https://doi.org/10.1073/pnas.0708078105

Suzuki, Y. J. (2003). Stress-induced activation of GATA-4 in cardiac muscle cells. *Free Radical Biology & Medicine*, *34*(12), 1589–1598. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12788478

Suzuki, Y. J., & Evans, T. (2004). Regulation of cardiac myocyte apoptosis by the GATA-4 transcription factor. *Life Sciences*, *74*(15), 1829–1838. https://doi.org/10.1016/j.lfs.2003.10.002

Suzuki, Y. J., Nagase, H., Day, R. M., & Das, D. K. (2004). GATA-4 regulation of myocardial survival in the preconditioned heart. *Journal of Molecular and Cellular Cardiology*, *37*(6), 1195–1203. https://doi.org/10.1016/J.YJMCC.2004.09.009

Svensson, E. C., Tufts, R. L., Polk, C. E., & Leiden, J. M. (1999). Molecular cloning of FOG-2: a modulator of transcription factor GATA-4 in cardiomyocytes. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(3), 956–961. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9927675

Swinney, D. C., & Anthony, J. (2011). How were new medicines discovered? *Nature Reviews Drug Discovery*, *10*(7), 507–519. https://doi.org/10.1038/nrd3480

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., … Von Mering, C. (2017). The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research*, *45*(D1), 362–368. https://doi.org/10.1093/nar/gkw937

Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Takaya, T., Kawamura, T., Morimoto, T., Ono, K., Kita, T., Shimatsu, A., & Hasegawa, K. (2008). Identification of p300-targeted acetylated residues in GATA4 during hypertrophic responses in cardiac myocytes. *The Journal of Biological Chemistry*, *283*(15), 9828–9835. https://doi.org/10.1074/jbc.M707391200

Talman, V., & Ruskoaho, H. (2016). Cardiac fibrosis in myocardial infarction—from repair

and remodeling to regeneration. *Cell and Tissue Research*, *365*(3), 563–581. https://doi.org/10.1007/s00441-016-2431-9

Tenhunen, O., Sármán, B., Kerkelä, R., Szokodi, I., Papp, L., Tóth, M., & Ruskoaho, H. (2004). Mitogen-activated Protein Kinases p38 and ERK 1/2 Mediate the Wall Stress-induced Activation of GATA-4 Binding in Adult Heart. *Journal of Biological Chemistry*, *279*(23), 24852–24860. https://doi.org/10.1074/jbc.M314317200

Thomas, M. C., & Chiang, C. M. (2006). The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology*, *41*(3), 105–178. https://doi.org/10.1080/10409230600648736

Thuerauf, D. J., Hanford, D. S., & Glembotski, C. C. (1994). Regulation of rat brain natriuretic peptide transcription. A potential role for GATA-related transcription factors in myocardial cell gene expression. *The Journal of Biological Chemistry*, *269*(27), 17772–17775. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8027030

Todeschini, A.-L., Georges, A., & Veitia, R. A. (2014). Transcription factors: specific DNA binding and specific gene regulation. *Trends in Genetics*, *30*(6), 211–219. https://doi.org/10.1016/j.tig.2014.04.002

Tsankov, A. M., Gu, H., Akopian, V., Ziller, M. J., Donaghey, J., Amit, I., … Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. *Nature*, *518*(7539), 344–349. https://doi.org/10.1038/nature14233

Tummino, P. J., & Copeland, R. A. (2008). Residence time of receptor - Ligand complexes and its effect on biological function. *Biochemistry*, *47*(20), 5481–5492. https://doi.org/10.1021/bi8002023

van Berlo, J. H., Elrod, J. W., Aronow, B. J., Pu, W. T., & Molkentin, J. D. (2011). Serine 105 phosphorylation of transcription factor GATA4 is necessary for stress-induced cardiac hypertrophy in vivo. *Proceedings of the National Academy of Sciences*, *108*(30), 12331–12336. https://doi.org/10.1073/pnas.1104499108

van den Boogaard, M., Wong, L. Y. E., Tessadori, F., Bakker, M. L., Dreizehnter, L. K., Wakker, V., … Christoffels, V. M. (2012). Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *Journal of Clinical Investigation*, *122*(7), 2519–2530. https://doi.org/10.1172/JCI62613

van Norman, G. A. (2016). Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs. *JACC: Basic to Translational Science*, *1*(3), 170–179. https://doi.org/10.1016/j.jacbts.2016.03.002

Wada, R., Muraoka, N., Inagawa, K., Yamakawa, H., Miyamoto, K., Sadahiro, T., … Ieda, M. (2013). Induction of human cardiomyocyte-like cells from fibroblasts by defined factors. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(31), 12667–12672. https://doi.org/10.1073/pnas.1304053110; 10.1073/pnas.1304053110

Wang, C., Greene, D., Xiao, L., Qi, R., & Luo, R. (2018). Recent Developments and Applications of the MMPBSA Method. *Frontiers in Molecular Biosciences*, *4*, 87. https://doi.org/10.3389/fmolb.2017.00087

Wang, J., Feng, X. H., & Schwartz, R. J. (2004). SUMO-1 modification activated GATA4-

dependent cardiogenic gene activity. *The Journal of Biological Chemistry*, *279*(47), 49091–49098. https://doi.org/10.1074/jbc.M407494200

Wang, L., Berne, B. J., & Friesner, R. A. (2012). On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. *Proceedings of the National Academy of Sciences*, *109*(6), 1937–1942. https://doi.org/10.1073/pnas.1114017109

Wang, L., Liu, Z., Yin, C., Asfour, H., Chen, O., Li, Y., … Qian, L. (2015). Stoichiometry of Gata4, Mef2c, and Tbx5 influences the efficiency and quality of induced cardiac myocyte reprogramming. *Circulation Research*, *116*(2), 237–244. https://doi.org/10.1161/CIRCRESAHA.116.305547

Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., … Abel, R. (2015). Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *Journal of the American Chemical Society*, *137*(7), 2695–2703. https://doi.org/10.1021/ja512751q

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., … Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Watt, A. J., Battle, M. A., Li, J., & Duncan, S. A. (2004). GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. *Proceedings of the National Academy of Sciences*, *101*(34), 12573–12578. https://doi.org/10.1073/pnas.0400752101

Weirauch, M. T., & Hughes, T. R. (2011). A Catalogue of Eukaryotic Transcription Factor Types, Their Evolutionary Origin, and Species Distribution. In *A Handbook of Transcription Factors* (pp. 25–73). Springer, Dordrecht. Retrieved from https://link.springer.com/chapter/10.1007/978-90-481-9069-0_3

Welboren, W.-J., van Driel, M. A., Janssen-Megens, E. M., van Heeringen, S. J., Sweep, F. C., Span, P. N., & Stunnenberg, H. G. (2009). ChIP-Seq of ERα and RNA polymerase II defines genes differentially responding to ligands. *The EMBO Journal*, *28*(10), 1418–1428. https://doi.org/10.1038/emboj.2009.88

Williams-Noonan, B. J., Yuriev, E., & Chalmers, D. K. (2018). Free Energy Methods in Drug Design: Prospects of "Alchemical Perturbation" in Medicinal Chemistry. *Journal of Medicinal Chemistry*, *61*(3), 638–649. https://doi.org/10.1021/acs.jmedchem.7b00681

Williams, G., Ferenczy, G. G., Ulander, J., & Keserű, G. M. (2017). Binding thermodynamics discriminates fragments from druglike compounds: a thermodynamic description of fragment-based drug discovery. *Drug Discovery Today*, *22*(4), 681–689. https://doi.org/10.1016/j.drudis.2016.11.019

Witty, A. D., Mihic, A., Tam, R. Y., Fisher, S. A., Mikryukov, A., Shoichet, M. S., … Keller, G. (2014). Generation of the epicardial lineage from human pluripotent stem cells. *Nature Biotechnology*, *32*(10), 1026–1035. https://doi.org/10.1038/nbt.3002

Xie, W., Silvers, R., Ouellette, M., Wu, Z., Lu, Q., Li, H., … Montoute, M. (2016). A

Luciferase Reporter Gene System for High-Throughput Screening of γ-Globin Gene Activators. In *High Throughput Screening* (Vol. 1439, pp. 207–226). https://doi.org/10.1007/978-1-4939-3673-1_14

Xie, Y., Jin, Y., Merenick, B. L., Ding, M., Fetalvero, K. M., Wagner, R. J., … Martin, K. A. (2015). Phosphorylation of GATA-6 is required for vascular smooth muscle cell differentiation after mTORC1 inhibition. *Science Signaling*, *8*(376). https://doi.org/10.1126/scisignal.2005482

Xin, M., Olson, E. N., & Bassel-Duby, R. (2013). Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. *Nature Reviews.Molecular Cell Biology*, *14*(8), 529–541. https://doi.org/10.1038/nrm3619; 10.1038/nrm3619

Xin, Z., Zhao, H., Serby, M. D., Liu, B., Schaefer, V. G., Falls, D. H., … Liu, G. (2005). Synthesis and structure–activity relationships of isoxazole carboxamides as growth hormone secretagogue receptor antagonists. *Bioorganic & Medicinal Chemistry Letters*, *15*(4), 1201–1204. https://doi.org/10.1016/j.bmcl.2004.11.075

Yanazume, T., Hasegawa, K., Morimoto, T., Kawamura, T., Wada, H., Matsumori, A., … Kita, T. (2003). Cardiac p300 is involved in myocyte growth with decompensated heart failure. *Molecular and Cellular Biology*, *23*(10), 3593–3606.

Yang, A., Zhu, Z., Kapranov, P., McKeon, F., Church, G. M., Gingeras, T. R., & Struhl, K. (2006). Relationships between p63 Binding, DNA Sequence, Transcription Activity, and Biological Function in Human Cells. *Molecular Cell*, *24*(4), 593–602. https://doi.org/10.1016/j.molcel.2006.10.018

Yang, Y.-Q., Wang, J., Liu, X.-Y., Chen, X.-Z., Zhang, W., & Wang, X.-Z. (2013). Mutation spectrum of GATA4 associated with congenital atrial septal defects. *Archives of Medical Science : AMS*, *9*(6), 976–983. https://doi.org/10.5114/aoms.2013.39788

Yeh, J. E., Toniolo, P. A., & Frank, D. A. (2013). Targeting transcription factors. *Current Opinion in Oncology*, *25*(6), 652–658. https://doi.org/10.1097/01.cco.0000432528.88101.1a

You, W., Song, L., & Wang, K. (2018). Acetylation of GATA4 on Lysine Residue K313 Promotes Osteoblastic Cells Growth. *Cellular Physiology and Biochemistry*, *46*(1), 269–278. https://doi.org/10.1159/000488428

Young, R. J., & Leeson, P. D. (2018). Mapping the Efficiency and Physicochemical Trajectories of Successful Optimizations. *Journal of Medicinal Chemistry*, acs.jmedchem.8b00180. https://doi.org/10.1021/acs.jmedchem.8b00180

Zang, R., Li, D., Tang, I.-C., Wang, J., & Yang, S.-T. (2012). Cell-Based Assays in High-Throughput Screening for Drug Discovery. *International Journal of Biotechnology for Wellness Industries*, *1*, 31–51. https://doi.org/10.6000/1927-3037.2012.01.01.02

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes & Development*, *25*(21), 2227–2241. https://doi.org/10.1101/gad.176826.111

Zhang, C., McKinsey, T. A., Chang, S., Antos, C. L., Hill, J. A., & Olson, E. N. (2002). Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell*, *110*(4), 479–488. https://doi.org/10.1016/S0092-8674(02)00861-9

Zhang, L., Nomura-Kitabayashi, A., Sultana, N., Cai, W., Cai, X., Moon, A. M., & Cai, C.-L. (2014). Mesodermal Nkx2.5 is necessary and sufficient for early second heart field development. *Developmental Biology*, *390*(1), 68–79. https://doi.org/10.1016/J.YDBIO.2014.02.023

Zhou, P., He, A., & Pu, W. T. (2012). Regulation of GATA4 Transcriptional Activity in Cardiovascular Development and Disease. In *Current topics in developmental biology* (Vol. 100, pp. 143–169). https://doi.org/10.1016/B978-0-12-387786-4.00005-1

Zwi, L., Caspi, O., Arbel, G., Huber, I., Gepstein, A., Park, I. H., & Gepstein, L. (2009). Cardiomyocyte differentiation of human induced pluripotent stem cells. *Circulation*, *120*(15), 1513–1523. https://doi.org/10.1161/CIRCULATIONAHA.109.868885

List of original publications

I Kinnunen, S.*, Välimäki, M.*, Tölli, M., Wohlfahrt, G., Darwich, R., Komati, H., Nemer, M., Ruskoaho, H. (2015). Nuclear receptor-like structure and interaction of congenital heart disease-associated factors GATA4 and NKX2-5. *PLoS ONE, 10*(12). https://doi.org/10.1371/journal.pone.0144145

II Välimäki, M. J.*, Tölli, M. A.*, Kinnunen, S. M., Aro, J., Serpi, R., Pohjolainen, L., Talman, V., Poso, A., Ruskoaho, H. J. (2017). Discovery of small molecules targeting the synergy of cardiac transcription factors GATA4 and NKX2-5. *Journal of Medicinal Chemistry, 60*(18), 7781–7798. https://doi.org/10.1021/acs.jmedchem.7b00816

III Kinnunen, S. M.*, Tölli, M.*, Välimäki, M. J., Gao, E., Szabo, Z., Rysä, J., Ferreira, M. P. A., Ohukainen, P., Serpi, R., Correia, A., Mäkilä, E., Salonen, J., Hirvonen, J., Santos, H. A., Ruskoaho, H. (2018). Cardiac actions of a small molecule inhibitor targeting GATA4-NKX2-5 interaction. *Scientific Reports, 8*(1). https://doi.org/10.1038/s41598-018-22830-8

IV Karhu, S. T., Välimäki, M. J., Jumppanen, M., Kinnunen, S. M., Pohjolainen, L., Leigh, R. S., Auno, S., Földes, G., Boije af Gennäs, G., Yli-Kauhaluoma, J., Ruskoaho, H., Talman, V. (2018). Stem cells are the most sensitive screening tool to identify toxicity of GATA4-targeted novel small-molecule compounds. *Archives of Toxicology, 92*(9), 2897–2911. https://doi.org/10.1007/s00204-018-2257-1

Original publications are not included in the electronic version of the dissertation.