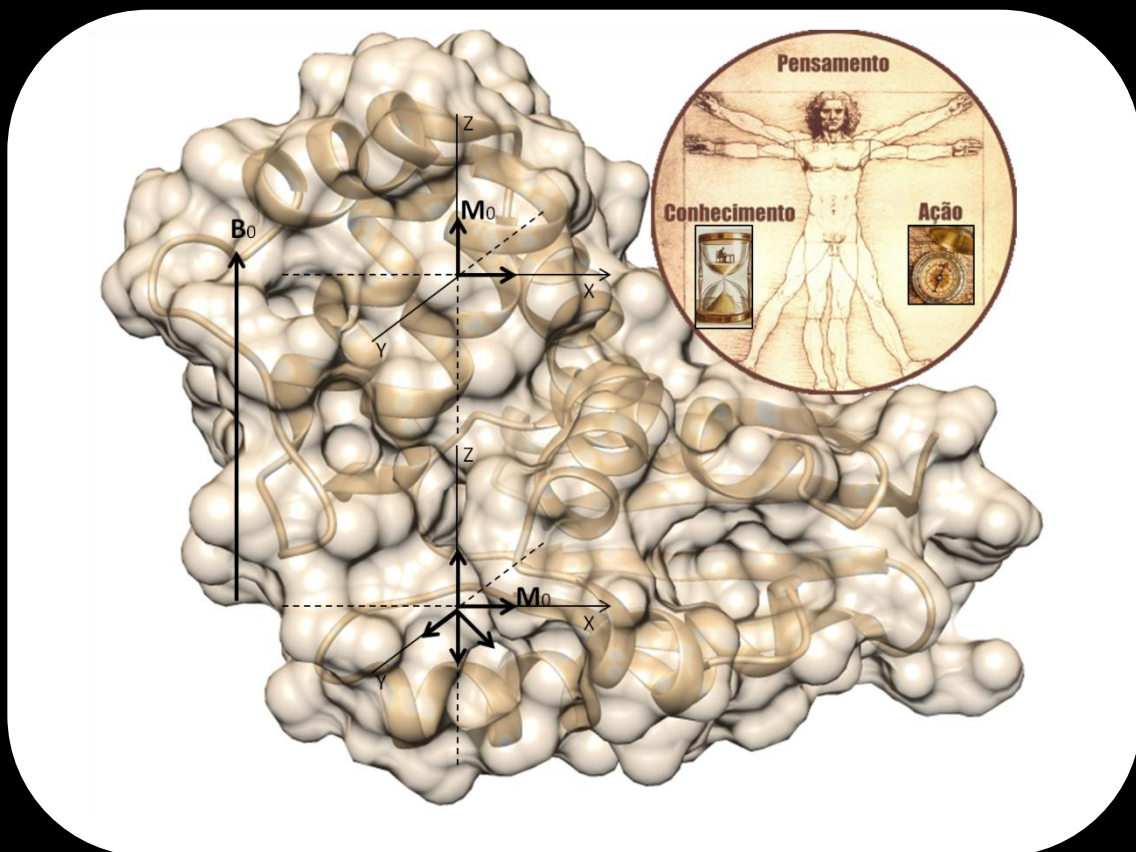# Title: Structure and dynamics of the NO sensing domain of the human soluble Guanylate Cyclase

Author: Meire Coelho de Almeida

Dissertation presented to obtain the Ph.D degree in Structural Biology

Oeiras,
September, 2015

itqb

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

Supervisor:

Doutor Emmanouil Matzapetakis
Investigador Auxiliar do Instituto de Tecnologia Química e Biológica António Xavier da Universidade Nova de Lisboa.

President of the Jury

Professor Doutor Miguel Nuno Sepúlveda Gouveia Teixeira
Professor Catedrático do Instituto de Tecnologia Química e Biológica António Xavier da Universidade Nova de Lisboa, por delegação

Examiners:

Professor Doutor Brian James Goodfellow
Professor Auxiliar do Centro de Investigação em Materiais Cerâmicos e Compósitos (CICECO) da Universidade de Aveiro.

Professor Doutor Carlos Alberto Alves Cordeiro
Investigador Auxiliar do Centro de Química e Bioquímica (CQB) da Faculdade de Ciências da Universidade de Lisboa.

Professor Doutora Sofia Rocha Pauleta
Investigadora Principal da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.

Professor Doutora Smilja Todorovic
Investigadora Principal do Instituto de Tecnologia Química e Biológica António Xavier da Universidade Nova de Lisboa.

To my father (in memory)

Acknowledgments

Acknowledgments

First of all, I want to thank myself for the courage and persistence in following a scientific career against all barriers.

I would like to thank all the people who contributed to the conclusion of my Ph.D thesis. This work has not been possible without the support of:

Acknowledgments

The most important appreciation goes to my father Jacy Bernardo (in memory), my mother Terezinha Rodrigues, my sisters Marcia and Mislene and my nephews Bárbara and Otávio. Despite having a large distance from Portugal to Brazil, they always gave me all emotional support.

Thank you!

Table of Contents

Abstract

Soluble guanylate cyclase (sGC) is a heterodimeric, nitric oxide (NO)-sensing hemoprotein. A heme cofactor is localized in the N-terminal, H-NOX domain. sGC is responsible for the catalytic conversion of GTP to cGMP. The NO-sGC-cGMP-dependent pathway is completely inactivated in the case of heme oxidation and the subsequent heme loss. In those cases reactivation of the enzyme can be achieved by Cinaciguat or BAY 58-2667 which is a heme mimetic compound that is in clinical trials for treatment of Acute Decompensated Heart Failure. In spite of much effort several aspects of sGC structure and function have yet to be elucidated. One of those is the structure of the H-NOX-heme domain and its involvement in the mechanism of activation of the full sGC enzyme.

The main goals of this work were the structure determination of the human H-NOX sGC domain incorporating the cinaciguat activator (H-NOX_c) and study its dynamics behaviour in different H-NOX states that represent the different stages of sGC activation. The later was achieved by using H-NOX complexed with the following ligands as models of different states of activation of the protein, a) heme and Zn^{2+} protoporphyrin IX (ZnPPIX) to represent the native, resting and inactive forms and b) cinaciguat bound to WT and H-NOX^{H105F} to model the active form of the protein.

In the first part of this work, we studied the incorporation of the various ligands into H-NOX, and the stability of their complex in solution. We found that the majority of the apo-protein expressed in *E.coli* (BL21-DE3) is in aggregated form, and we could characterize it to be folded by NMR. We also found that incorporation of cinaciguat (activator) or ZnPPIX (inhibitor) compound resulted in a complex that is in the monomeric state.

In the second part which is described in chapters 3 and 4 we describe the NMR resonance assignment of the human H-NOX_c and the determination of its secondary and tertiary structure. We were able to assign 88% of amide of the

protein backbone, including the 3 prolines and the conserved residues Y2, P118, Y135 S137 and R139, which are involved in heme and/or cinaciguat binding. The secondary structure of human H-NOX sGC domain agrees with that of homologous structures. Using information from NOESY spectra and chemical shift derived dihedral angles, we calculated structures from which, a bundle of the 10 best models of cinaciguat bound protein were selected. (r.m.s.d of 0.77 Å). In those models, cinaciguat is binding in a similar mode to that observed in crystal structures of its complexes with prokaryotic analogues. In addition, the overall structure overlays well with prokaryotic H-NOX homologous structure present in the literature. However, a portion of the protein is not well defined due to the absence of signals corresponding to residues in the region α-helices F and E regions (β-sheet domains).

In the final part of the work, we studied the molecular dynamics of the H-NOX domain using Nuclear Spin Relaxation (NSR) that can probe the dynamics at the ps-ns timescale. Therefore, ^{15}N relaxation data ($R1$, $R2$ and ^{15}N-^{1}H NOE) were collected for the backbone of H-NOX_c, H-NOX^{H105F}_c and H-NOX_Zn. Analysis of the $R1$, R2 and ^{15}N-^{1}H heteronuclear-NOE data for H-NOX_c and H-NOX^{H105F}_c (active state) showed that three clusters of residues exhibited increased dynamics in comparison to the rest of the protein backbone. These residues are located in the N-terminal helices αB-αC and C-terminal β1-β2 linker, respectively. The αB-C region has been previously shown in the literature (using HDX data) to change in backbone H exchange rates upon NO-binding. Therefore, this portion of H-NOX domain might be involved in sGC inter-domain interaction and in full enzyme allosteric regulation. Due to the limited number of resonance assignments of the H-NOX_Zn form of the protein were only able to measure the relaxation of a small number of residues which did not allow us to interpret them in a conclusive manner.

In summary, the work on this dissertation contributed to expand our knowledge of the sGC enzyme. Through our structural work we were able to determine for the first time the structure of the human H-NOX domain. The secondary and tertiary structures were found to be in agreement with homologous H-NOX protein structures. NMR approach was useful for probing the dynamics of the protein in solution, principally for H-NOX sGC domain that could not be crystallized.

Resumo

A forma solúvel da proteína hémica guanilato ciclase (sGC) é heterodimérica. Essa enzima é capaz de captar o óxido nítrico (NO) e converter o substrato GTP em cGMP. cGMP é um importante mensageiro secundário capaz de regular diferentes vias de sinalização intracelular, tais como: vasodilatação, relaxação da musculatura lisa, ativação dos canais de cálcio e a citoproteção de células cardíacas. A via dependente de NO-sGC-cGMP é completamente inactivada, quando a célula está sob stres oxidativo, o qual causa a oxidação do hemo-Fe^{2+} para hemo-Fe^{3+}. O cofactor hemo está localizado na região N-terminal da sGC denominada domínio H-NOX. A oxidação do hemo desestabiliza as interações químicas que mantêm o cofactor ligado à enzima sGC e proporciona a sua saída da proteína. A reactivação da enzima pode ser alcançada por meio da inserção de um fármaco chamado Cinaciguat ou BAY 58-2667. Este composto possui uma estrutura química similar ao cofactor hemo. Essa similaridade é o factor chave para a interação do composto com os resíduos envolvidos na interação proteína-ligando, bem como reactivação da enzima. Além disso, existem também ensaios clínicos que comprovam a eficácia do cinaciguat no tratamento de pacientes que sofrem de falha cardíaca aguda descompensada. Apesar de muito esforço, ainda há vários aspectos da estrutura e função da sGC que não foram elucidados, nomeadamente a estrutura do domínio H-NOX e o seu envolvimento no mecanismo de activação da proteína sGC.

Neste contexto os principais objetivos deste trabalho foram: a) a determinação da estrutura do domínio H-NOX humano incorporando o ativador cinaciguat (H-NOX_c) e b) o estudo do comportamento dinâmico desse domínio em diferentes estados que podem representar as várias fases de activação do sGC. Deste modo, amostras do domínio H-NOX foram complexadas com os diferentes ligandos, modelando os diferentes estados de activação da proteína: a) hemo e protoporfirina IX de Zn^{2+} (ZnPPIX) (H-NOX_Zn) para representar o estado de repouso da forma nativa e inactiva e b) a proteína Wild-type e o mutante

HNOX^{H105F} incorporada ao composto cinaciguat para modelar a forma activa da proteína.

Na primeira parte deste trabalho estudou-se a incorporação dos vários ligandos no domínio H-NOX, bem como a sua estabilidade em solução. Verificou-se que a maior parte da apo-proteína expressa em *E. coli* (BL21-DE3) encontra-se agregada, e o seu enrolamento foi caracterizado por espectroscopia de Ressonância Magnetica Nuclear (RMN). Observou-se também que a incorporação do cinaciguat (activador) ou ZnPPIX (inibidor) resulta na formação de um complexo no estado monomérico.

A segunda parte da tese descreve a atribuição dos sinais de RMN à cadeia principal e lateral dos aminoácidos pertencentes à ao domínio H-NOX, bem como a região de interação com o composto cinaciguat. Além disso, também estão apresentados os resultados a estrutura secundária e terciária do domínio sGC. Fomos capazes de atribuir 88 % da cadeia principal da proteína, incluindo as 3 prolinas e os resíduos conservados Y2, P118, Y135 S137 e R139, os quais já foram identificados como cruciais para a ligação com os ligandos. Os resultados também mostraram que a estrutura secundária do domínio H-NOX é semelhante às estruturas homólogas descritas na literatura. Além disso, a estrutura terciária foi calculada com base nas informações dos espectros de RMN NOESY e dos desvios químicos adquiridos após a atribuição dos resíduos. Um conjunto dos 10 melhores modelos da proteína ligada ao composto cinaciguat foiselecionado (R.m.s.d de 0,77 Å). Após análisar a conformação da estrutura terciária, observamos que a estrutura do H-NOX_c humano apresenta similaridade com várias estruturas análogas de procariotas, as quais possuem os seus dados estruturais depositados no Banco de Dados de Proteinas (PDB). No entanto, uma porção da proteína não se encontra bem definida, devido à ausência de sinais correspondentes aos resíduos na região α-hélices F e E que estão localizadas próximas ao domínio β-sheet.

Na parte final do trabalho, estudamos a dinâmica molecular do domínio H-NOX utilizando experiências de RMN para captar o dados de relaxação de cada resíduo pertencente à proteína. Assim, usando amostras de proteína marcada com ^{15}N para a recolha de dados de relaxação (R1, R2 e ^{15}N-^{1}H NOE) fomos capazes de sondar a dinâmica de cada átomo em ps-ns. Informações de dinâmica dos aminoácidos presentes na cadeia principal foram obtidas para as amostras H-NOX_c, H-NOX^{H105F}_c e H-NOX_Zn. A análise dos dados R1, R2 e NOE heteronuclear ^{15}N-^{1}H para H-NOX_c e H-NOX^{H105F}_c (estado ativo) apresentaram que alguns resíduos possuem uma maior dinâmica em comparação com o demais resíduos pertecentes à cadeia principal da proteína. Alguns desses resíduos estão localizados nas hélices N-terminais αB-C e β1-β2 C-terminal, respectivamente. No entanto, devido ao número limitado de atribuições de ressonâncias da forma H-NOX_Zn da proteína, fomos apenas capazes de medir o relaxamento de um pequeno número de resíduos, não permitindo uma interpretação conclusiva. A região αB-C do domínio H-NOX encontra-se descrita na literatura (usando dados HDX) como sendo capaz de aumentar as taxas de permuta de protão por deutério quando o óxido nítrico se liga à enzima. Por isso, pode-se deduzir que esta porção do domínio H-NOX está envolvida na interação entre os demais domínios e faz parte do evento alostérico de ativação desta enzima.

Em resumo, o trabalho nesta dissertação contribuiu para ampliar o conhecimento da enzima sGC. O nosso trabalho estrutural permitiu determinar, pela primeira vez a estrutura do domínio H-NOX de humano. As estruturas secundárias e terciárias encontradas estão de acordo com estruturas de H-NOX homólogos. Por fim, a técnica de RMN demonstrou ser uma abordagem útil para sondar as propriedades dinâmicas de proteínas em solução, especialmente para o domínio H-NOX que até à data apresenta ser resistente à cristalização.

List of Publication

Saraiva, IH, **Almeida MC**, Matzapetakis M. Structural investigation of the NO sensing domain of the human soluble Guanylate Cyclase <u>Manuscript in preparation.</u>

List of Figures and Tables

List of Abbreviation

Abbreviation	Full forms
cGMP	3´, 5´-cyclic Guanosine MonoPhosphate
CO	Carbon monOxide
CC	Coiled Coil
DTT	DiThioThreitol
EDTA	EthyleneDiamineTetraacetic Acid
eNOS	Endothelial Nitric Oxide Synthase
H-NOX	Heme Nitric Oxide Oxygen domain
HO	Heme Oxygenase
hnNOE	Hetero Nuclear Overhauser Effect
H-NOX_c	Cinaciguat bound to WT H-NOX domain
H-NOX_Zn	Zn^{2+}PPIX bound H-NOX domain
H-NOX^{H105F}_c	Cinaciguat bound to mutant^{H105F} H-NOX domain
HSQC	Heteronuclear Single Quantum Coherence spectrum
iNOS	Inducible Nitric Oxide Synthase
IPTG	IsoPropyl β-D-1-ThioGalactopyranoside
β-Me	2 β-Mercaptoethanol
NO	Nitric Oxide
nNOS,	Neuronal Nitric Oxide Synthase
N-octyl	N-octyl-β-D-glucoside
NOE	Nuclear Overhauser Effect
Ns **H-NOX**	H-NOX structure from *Nostoc sp*
pGC	particulate Guanylate Cyclase
Per/Arnt/Sim (PAS)	Period and Single-Minded Aryl hydrocarbon Receptor Nuclear Transporter
PMSF	Phenyl Methyl Sulfonyl Fluoride
PDEs	PhosphoDiEsterases
PDB	Protein Data Bank
PKG	Protein Kinase G
RDC	Residual dipolar Coupling
NMR	Nuclear Magnetic Resonance
r.m.s.d	root-mean-square deviation
So **H-NOX**	H-NOX structure from *Shewanella oneidensis*
sGC	soluble Guanylate Cyclase
Tt **H-NOX**	H-NOX structure from *Thermoanaerobacter tengcongensis*
TOSCY	Total Correlation Spectroscopy
W.T	Wild Type
ZnPPIX	Zinc $^{2+}$ Protoporphyrin IX
δ-ALA	Aminolevulinic acid

Chapter 1

Introduction

Keywords: sGC, allosteric regulation, NMR, protein dynamics.

Table of Contents

Introduction

1. General Introduction

Soluble guanylate cyclase (sGC) is a heterodimeric heme-protein composed of two subunits called α and β [1-5]. The most common heterodimeric form is the combination of α1 with β1 subunits [1]. The α1 (80KDa) and β1 (70 KDa) subunits are 690 and 619 amino acids in length respectively, and are encoded by the genes, GUCY1A2 and GUCY1A3 respectively [3]. Generally, the highest sequence variability is found at the N-terminus of the α subunits while the greatest sequence identity is at the C-terminus of both α and β proteins [6-11]. Each sGC subunit consists of four distinct domains. The β1 subunit contains an N-terminal heme binding domain called H-NOX, however, the α1 N-terminal domain although similar to β1, does not bind heme and is therefore, called the H-NOB domain. After the H-NOX/H-NOB domain follows a Per/Arnt/Sim (PAS) domain, a putative amphipathic helix coiled coil domain, and the C-terminal catalytic domain [1,12,13]. This enzyme is activated upon binding of nitric oxide (NO) to the H-NOX domain. The allosteric activation of the catalytic site results from conformational changes, initiated after NO binding, that propagate via inter-domain interaction and lead to the generation of the secondary messenger 3′, 5′-cyclic Guanosine MonoPhosphate (cGMP) [5,7]. The increased amount of cGMP activates several secondary pathways, such as NO/sGC/cGMP/Protein Kinase G (PKG) signaling [14,15]. This pathway initiates a cascade of phosphorylation reactions in which the magnitude of each step is enzymatically amplified, a process that is vital for several physiological effects including the cytoprotective effect in cardiac cell, platelet aggregation and smooth muscle relaxation [16-18].

cGMP synthesis is completely inactivated in the case of heme oxidation and its subsequent loss from the H-NOX domain [5,11]. This malfunction leads to sGC instability and eventually leads to degradation by the proteasome. This

failure, compromises the NO/sGC/cGMP/PKG signaling pathway [11]. In such cases, therapeutic reactivation of the enzyme's cGMP production can be achieved by Cinaciguat or BAY 58-2667 (BAY), a potent sGC activator that acts independently of nitric oxide [19]. Cinaciguat bound to sGC is able to convert GTP to cGMP. This drug has been in clinical trials for the treatment of Acute Decompensated Heart Failure [19,20].

Despite much effort, the mechanism of allosteric regulation of the full length sGC has not yet been fully elucidated. However, the structure and function of some individual sGC domains or their bacterial analogues [8-10,21,22], have been characterized in an effort to understand the enzyme architecture [4,12] and inter-domain interactions [12,13]. Molecular details of NO bound sGG shown has NO induce a cleavage of the heme-H105 bond, and elicits a pronounced conformational change in the protein as a result of structural rearrangements of the heme into the protein pocket [23]. Moreover, evidence suggests that NO binding induces widespread effects in several discrete sGC regions [12], which propagate structural perturbation via inter-domain interaction, from the sGC heme domain to the cyclase active site [4,6,12,24]. There is also evidence that changes in protein dynamics are involved in sGC activation.

2 cGMP cell signaling

cGMP is a cyclic nucleotide produced from guanosine triphosphate (GTP) by allosteric enzymes belonging to the guanylate cyclase family, such as particulate guanylate cyclase (pGC) and soluble guanylate cyclase (sGC) [7]. cGMP acts as a secondary messenger much like cyclic adenosine monophosphate (cAMP) [15]. cGMP activates intracellular protein kinases such as, protein kinase G, cyclic AMP-dependent protein kinase A (PKA), cyclic nucleotide-gated (CNG) cation channels and cyclic nucleotide phosphodiesterases (PDEs) [7,15]. Studies have implicated

that endogenous and exogenous molecules, including autacoids, hormones, neurotransmitters, toxins and small gas molecules such as nitric oxide/carbon monoxide produce cellular responses through cGMP [7].

2.1   Protein Kinase G

Protein Kinase G represents the principal intracellular mediator of cGMP signals. Elevation of cGMP concentration leads to cGMP binding and its subsequent activation leading to the catalytic transfer of the γ-phosphate from an ATP molecule to a serine or threonine residue of the target protein. The function of the resulting phosphorylated protein is thus changed, the cascade that started from the extracellular stimulus into a specific biological function [7].

In mammals there are two PKG families PKGI and PKGII that are both derived from separate genes (*prkg1* and *prkg2* respectively). Only the PKGI will be discussed here because its members are more commonly involved when cGMP is mediated by NO.

PKGIs are homodimers containing two identical subunits (approximately 75 kDa and 85 kDa, for PKGI and PKGII respectively) and share common structural features. [14]. PKGI, which is of interest to NO signalinghas been described to have two isoforms, PKGIα and PKGIβ. Both are able to bind two cGMP molecules per monomer [25,26]. The difference between the two is that PKGIα has high and low affinity binding sites that display positive cooperative behavior, while PKGIβ has two cGMP binding sites characterized by lower affinity and cooperativity [27]. PKGIα was detected mainly in particular subcellular membrane fractions, in complex with certain cytosolic proteins, and as free cytosolic protein in the vascular system, kidney, and adrenal gland. The PKGIβ was only detected in the uterus [28].

The isoform PKGIα regulates many secondary pathways increasing smooth muscle relaxation, cell adhesion, vasodilation and cytoprotective effects in heart

cells [29-31]. In cardiac cell the cytoprotective effects are beyond question complex. One such mechanism in cardiac tissues, involves the NO/sGC/cGMP/PKGI signaling pathway to increase the opening of mitochondrial K$^+$/ATP channels and thereby diminish damage incurred on cardiovascular system failure [32-34].

Both isoforms of PKGI are also involved in cyclic nucleotide phosphodiesterases (PDEs) regulation model, for instance in phosphorylation and activation mechanism [35,36]. PDEs constitute an enzyme class that regulates the cellular levels of the second messengers, cAMP and cGMP, by controlling their rate of degradation [35-37]. One of the most common is the PDE5 isoform found in platelets, smooth muscle cells, and cerebellar Purkinje cells [35-37]. Its mechanism of activation is by cGMP binding to allosteric sites of the proteinand by its phosphorylation from the PKGIα leading to negative feedback on the NO/sGC/cGMP/PKGI signaling [38-40].

3 The Guanylate Cyclase Family

Guanylate cyclase family is a class of enzymes that catalyzes the formation of cGMP from GTP [2,41-44]. This family has been studied extensively since the first reports of such enzyme activity in 1969, in which these proteins were indentified existing both in the soluble and particulate fractions of cells. The soluble protein was named soluble guanylate cyclase (sGC), which is a heterodimer hemo-protein that can be regulated by nitric oxide, whereas the particulate form exists as a homodimer that can be regulated by various peptides [45-47].

3.1 Particulate form of Guanylate Cyclase

The particulate guanylate cyclase (pGC) is a homodimer transmembrane protein. In mammals, there are seven isoforms of the particulate guanylate

cyclase (abbreviated GC-A through GC-G) [48,49]. GC-A is expressed in the kidney, smooth muscle vasculature, adrenal gland, and many tissues, whereas GC-B, although found in many tissues, is expressed at particularly high level in the fibroblast [7,50]. GC-C is expressed principally in the intestine but is also found in kidney, testes, liver, and placenta in various animals [51-53]. Guanylate cyclase GC-D was identified in olfactory cell and was denominated neuroepithelial guanylate cyclase (ONE-GC-D) and it is a direct receptor of odorants [54]. GC-E and GC-F were described assensory neuron–sight photoreceptors [55,56]. Guanylate cyclase G is the last member of membrane forms identified thus far. It has been identified only in rat species, where it is expressed in the lung, small intestine, skeletal muscles and kidney [57].

3.2 Soluble form of Guanylate Cyclase

sGC is a heterodimeric heme-protein consisting of α- and β-subunits, and expression of both subunits are required for catalytic activity [5,9,58-60]. Each subunit has an N-terminal regulatory domain and a C-terminal catalytic domain [7]. sGC enzyme belongs to the class III family of cyclases defined by homology within the catalytic domains with the corresponding domains in pGC and Adenylate Cyclases (AC). Adenylate cyclase differentiates from pGC and sGC by catalyzing the conversion of ATP to cAMP. [61].

3.2.1 Nitric oxide sGC activator

In eukaryotic cells nitric oxide is a critical diatomic signaling molecule influencing various physiological processes such as smooth muscle relaxation, neurotransmission, and platelet aggregation. The surge in intracellular Ca^{2+} levels induces the formation of the Ca^{2+} and Calmodulin (CaM) complex. This complex leads to subsequent CaM-dependent activation of nitric oxide synthase (NOS). Afterwards, NOS catalyzes the oxidation of the amino acid L-arginine to produce both NO and L-citrulline (Figure 1) [14,62,63].

There are three isoforms of NOS described, the endothelial, the neuronal and the inducible (eNOS, nNOS, and iNOS). Both eNOS and nNOS are constitutively expressed while iNOS is induced with the proper immunostimulatory signals [14,62,64]. NO is produced by NOS and can rapidly diffuse across a cell membrane to activate sGC as shown in Figure 1 [65-67].
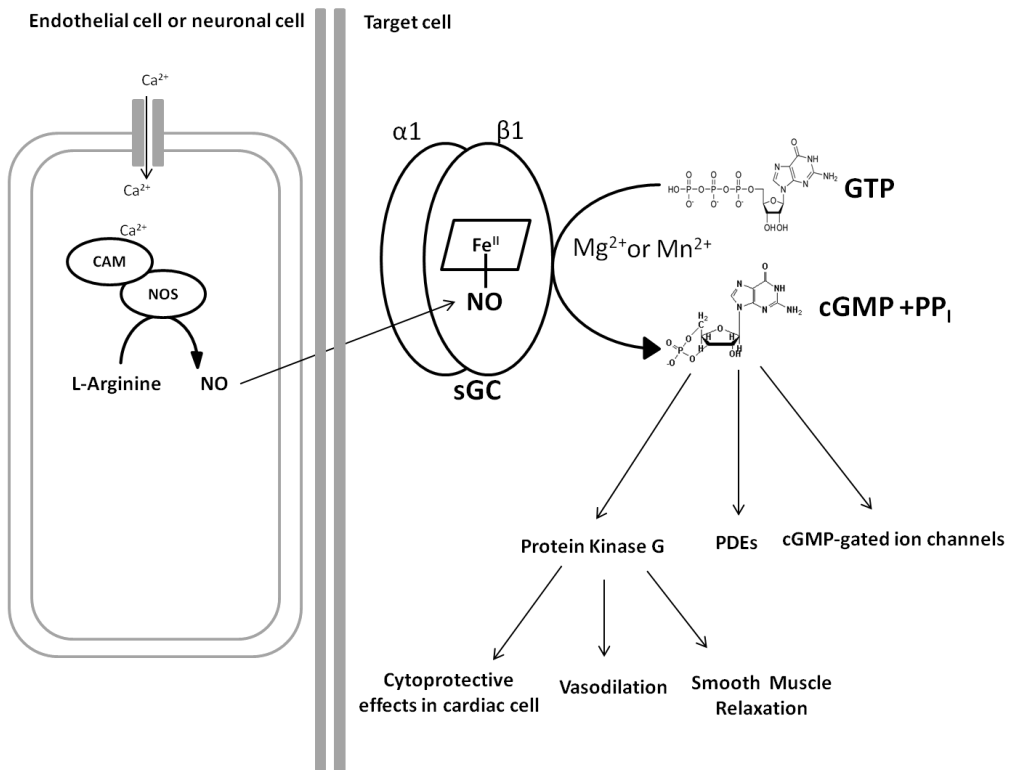
Figure 1: NO-dependent sGC cell signaling.
NO is synthesized by NOS diffuses across cell membranes to a target cell. NO activates sGC. Upon NO bound to sGC it leads to an increase in cGMP synthesis. Adapted from [14,58,59,62,64].

3.2.2 Subunits and Isotypes of sGC

Multiple isotypes containing different subunits in their composition have been found in mammalian sGCs. However, in many tissues, the most abundant subunits are α1 (80-kDa) and β1 (70-kDa). The heterologous expression of individual α1 and β1, produced proteins that do not exhibit catalytic activity.

Therefore, heterodimerization was shown to be essential for function and activation upon NO binding [68,69].

The isoform β2 contains 86 additional amino acids in its C-terminal region compared with β1 [70]. This additional sequence contains a consensus CVVL sequence that was described to be involved in post-translational modifications, such as isoprenylation and carboxymethylation, suggesting that β2 subunit might serve to localize sGC to membranes [70]. According to Gupta and collaborators the β2 subunit is also able to form heterodimer with α1. However, this holoenzyme exhibited lower specific activity compared with α1/β1 heterodimer. COS-7 cells (a fibroblast-like cell line derived from monkey kidney tissue) co-transfected with α1/β1 sGC, upon NO stimulation, resulted in three times more cGMP synthesis in comparison with cells co-transfected with α1/β2. Moreover, the competition between β1 and β2 for binding to α1 indicates that expression of β2 might regulate α1/β1 sGC activity [71].

The α2 (82-kDa) subunit can form heterodimers with β1 or β2, albeit with lower affinity for β1. Moreover, the heterodimerization of the α1/β1 has a higher specific activity in comparison with α2/β1 [72]. Two other subunits of human sGC, α3 (82 kDa) and β3 (70 kDa), have been also described. The α3 and β3 have low homology in their N-terminal regions, however, there is 72% homology between their 310-residue C-terminal regions. [73].

3.2.3 Architecture of sGC

As mentioned earlier, the α1/β1 dimer is the predominant complex which is highly expressed in the brain, lung, heart, kidney spleen and muscle [1]. Each sGC subunit has four domains, the N-terminal domain from α1 has been described as not able to bind heme and is denoted as H-NOB [1]. In contrast, the corresponding domain from β1 binds heme and is termed the H-NOX domain [1,13]. Both subunits have similar structure starting from the N-terminal H-

NOX/H-NOB domains, followed by a Per/ARNT/Sim (PAS), a Helix coiled coil (CC), and a C-terminal catalytic domain as shown in Figure 2.
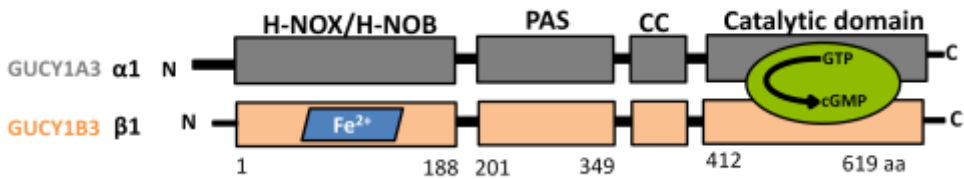
Figure 2: Domain organization of sGC subunits and truncations.
The sGC holoenzyme consists of two homologous subunits, α1 and β1 expressed by separated gene, grey and yellow respectively. Each subunit is composed of four domains. NO gas binds to the ferrous heme cofactor (blue) of the β1 H-NOX domain. PAS and CC domain are involved in inter-domain interaction. The catalytic site of the enzyme is responsible for cyclization of GTP substrate (green) to cGMP. Adapted from [12].

3.2.4 H-NOX domain

The H-NOX domain (22.3 kDa) of sGC has a prosthetic b type heme (heme iron, is bound via a single coordination bond with an amino-acid side-chain) which provides sGC the ability to sense NO/CO, and initiate the signaling process to the catalytic domain resulting in the generation of cGMP from the GTP substrate [7,9,15,63,74,75]. Structurally, this domain is composed of seven alpha helices (αA-G) and four beta sheets (β 1-4) [9,76,77]. The H-NOX domain, has also been described to be composed of two subdomains, a located next to the corresponding distal and the proximal pockets of the prosthetic heme as shown in **Error! Reference source not found.**.The H-NOX domain of sGC shares similar spectroscopic properties with the H-NOX protein found in prokaryotes [78]. The prosthetic heme interacts with the pocket predominantly by hydrophobic interactions, with the exception of the binding to the H105 imidazole ring , which forms a bond with ferrous heme, and interactions of both propionate groups of the heme with the main-chain nitrogen of Y2 and the side chain of YxSxR conserved sequence. Thus, due to the lack of any structures of the eukaryotic

form, most of knowledge present on the literature comes from structural studies of prokaryotic H-NOX proteins that are homologous to the sGC HNOX domain [9].
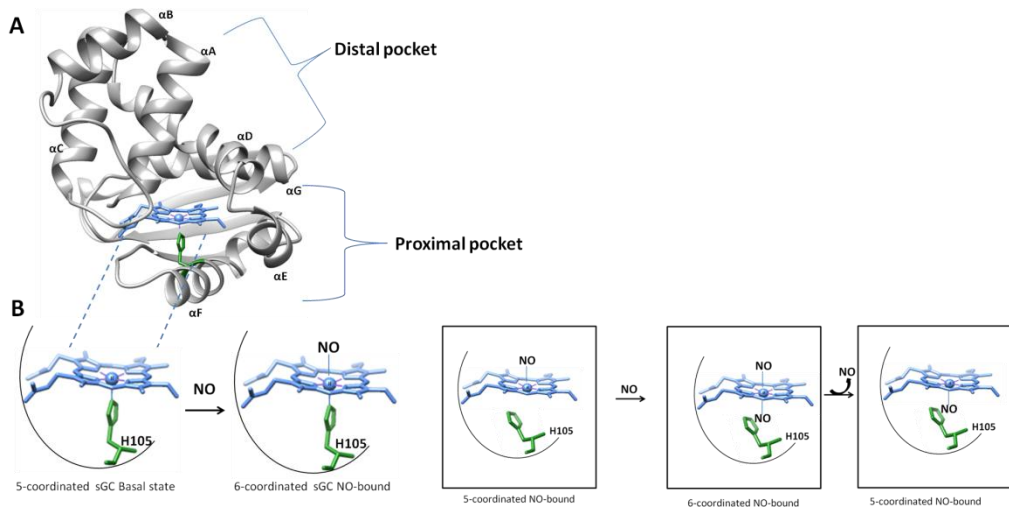
Figure 3: A) Structure of stand-alone *Nostoc sp* homologue of H-NOX sGC .
The heme (light blue), and H105 (green) are highlighted (PBD entry 2O09). **B)** Proposed mechanisms of sGC activation. The first NO molecule from distal pocket causes the ferrous-NO rupture forming a pentacoordinate ferrous nitrosyl complex. A second NO binds to the proximal pocket forming a transient hexa-coordinate species. Release of NO from the distal site results into a proximal site bound NO which corresponds to the high-activity form of H-NOX. Adapted from [9,23,79,80].

H-NOX domain of sGC as ligands sensor

In mammals sGC is a sensor of NO and CO gasses. A single carbon monoxide molecule has been found to form a reversible hexacoordinate low spin complex with the heme [5], in particular in olfactory neurons in which heme oxygenase produces CO [81,82]. Binding of CO to sGC can be monitored by UV-Vis. When CO is binding to sGC the λ_{Soret} band shifts from 431 to 423nm [9]. The CO dissociation constant (K_d) has been determined to be 97 +/- 9 μM [83]. sGC activity increases 6-fold in its activity when CO is bound.

Binding of NO to sGC (K_d 10–100 pM) results in the enzyme activation by several hundred-folds [60,84]. Unlike CO, two molecules of NO have been shown to bind to sGC sequentially. Both NO binding events to sGC can also be monitored

by UV-Vis, from a shift of the λ_{Soret} band from 431 to 399nm [59]. Freeze quench Electron Paramagnetic Resonance (EPR) with anaerobic sequential mixing of [14]NO and [15]NO unambiguously confirmed that the heme Fe^{2+} is the target of the second NO. The first NO molecule binds to the H-NOX domain distal pocket creating a transiently hexacoordinated Fe^{2+} nitrosyl complex. Dissociation of H105 allows for the simultaneous binding of the second NO molecule in the proximal pocket. The distal site NO is released from the heme leaving sGC in a pentacoordinated high spin Fe^{2+} nitrosyl complex corresponding to the active state (Figure 3B) [5,80]. A recent publication of the *So* H-NOX Fe^{2+}-NO crystal structure has also proven the formation of NO in the proximal pocket in H-NOX protein [23]. Moreover, this work suggested that 5 coordinate NO-sGC species with NO ligated at the proximal site, might be the key intermediate that contributes to the cGMP formation during the steady-state catalysis of sGC [23,80].

H-NOX domain in prokaryotes

Prokaryotic H-NOX proteins have been known for some time and are described as belonging to two distinct classes. The first consists of a stand-alone H-NOX protein most often found in a predicted operon with either a histidine kinase or, not as much of with a GGDEF-diguanylate cyclase domain. The second class consists of H-NOX domains fused to methyl-accepting chemotaxis proteins (MCP) [76,77,85,86]. The stand-alone H-NOX domain of *Nostoc sp* (*Ns* H-NOX) and *Shewanella oneidensis* (*So* H-NOX), the fusion protein from *Thermoanaerobacter tengcongensis* (*Tt* H-NOX) shares 18-40% sequence similarity with sGC [9,86-88] (Figure 4 A). Based on phylogenetic analysis it was suggested that the bacterial H-NOX domains were probably transferred to animals, resulting in the mammalian NO/CO-responsive sGC [76]. In addition, *Ns* and *So* H-NOX have similar properties with eukaryotic H-NOX sGC including its ability to bind CO and NO [9,86]. Although *Tt* H-NOX has similar sequence it has been identified to bind oxygen under aerobic conditions [87,88].

In the crystal structure of the *Ns* H-NOX heme domain at 2.1 Å resolution the conserved histidine (H105) could be seen in the proximal site, which examined by Raman Spectroscopy, confirmed the fifth ligand for a pentacoordinated ferrous heme [9]. The residue R138 was found to be key for heme binding in sGC [89,90] and *Tt* H-NOX [87,88]. Additional analysis showed that the *Ns* H-NOX structure has an aromatic residue, in the distal site W74 (F74 in sGC), oriented perpendicularly to the heme Figure 4 B. Such an aromatic residue is absent in *Tt* H-NOX, which contains an asparagines at this position. A possible explanation for the fact that *Ns* H-NOX does not bind oxygen is that it does not have the Y140 that is found in *Tt* H-NOX (M in *Ns* H-NOX), which interacts with the oxygen [87,88]. The binding of oxygen model compounds of ferrous heme ($Fe^{(2+)}$PP (1-MeIm)) was shown to be much weaker than NO and CO binding [91], indicating that ferrous heme proteins require additional polar/electrostatic interactions for oxygen to bind [92]. Therefore, the absence of such polar interactions might explain the absence of oxygen binding properties of *Ns* H-NOX and sGC [9].

A

B

Figure 4: A- Sequence alignment of HNOX.
Residues in the heme cavity are labelled in blue, pink, green and yellow, the latter if the residue is fully conserved. The GenBank identifiers for the selected sequences are *Homo sapiens*, GUCY1A3; *Nostoc sp*, GI:17229770; *T. tengcongensis*, GI:20807169; *Shewanella oneidensis,* GI:470469643. Adapted from Phyre2 [93] output. **B-** 3D structure stand-alone *Nostoc sp* H-NOX sGC homologous. The ferrous heme interact with side chains of the R138, S136, Y134, backbone of the Y2 conserved residues. The W74 residue perpendicular to the ferrous heme and H105 (green) are also highlight (PBD entry 2O09) Adapted from [9].

3.2.5 Per/ARNT/Sim (PAS) domain

The name of the Per-ARNT-Sim (PAS) domain was based on the proteins Period and Single-Minded (Per-Sim), and the vertebrate Aryl hydrocarbon Receptor Nuclear Transporter (ARNT) [94]. *Drosophila* was the first organisms where this domain was found. Thereafter, PAS domain and its function were described in many other organisms. For example it is responsible for regulating processes as diverse as nitrogen fixation in rhizobia, phototropism in plants, circadian behavior in insects, and gating of ion channels in vertebrates. [94,95,96,97,98]. Another function of this domain is to bind cofactors such as metabolites, ions, heme and flavin nucleotides. Moreover, the most common function of this domain is to be a mediator of the interactions between proteins [95] and regulates the activities of a wide range of different effector domains [22,94].
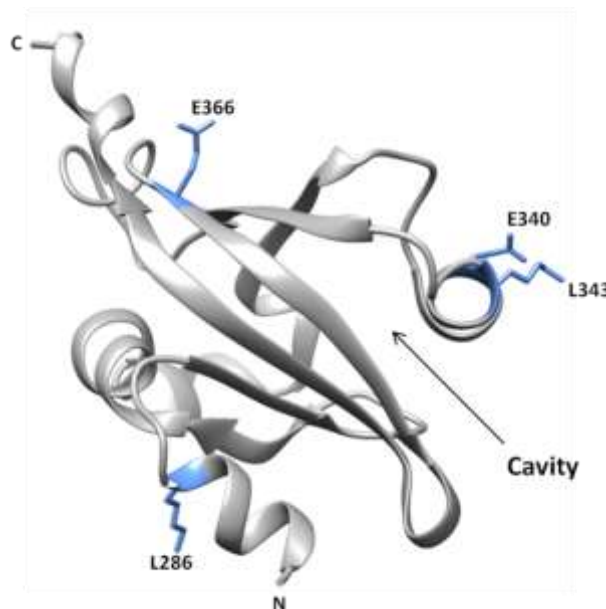
Figure 5: Crystal structure of *Manduca sexta* sGC α1 at 1.8 Å resolution.
The structure emphasize the small internal pocket and inter-domain contact residues highlighted in blue (PBD entry 4GJ4) Adapted from [22].

sGC PAS domain (~17.5 kDa) has been characterized to be involved in heterodimer formation along with the helix coiled coil domain[8,21,22, Ma, 2010 #2473]][9]. Two cyanobacteria, *Nostoc punctiforme* (*Np*) and *Nostoc sp* (*Ns*), have been described as having proteins with PAS-like folds, homologous to the central dimerization region in sGC [8]. *Np* has adjacent genes for Signal Transduction Histidine Kinases (STHK) [96]. *Ns* contains a 2-component hybrid sensor and regulator (2-CHSR). Both protein complexes have a key role for in sensing environmental stresses and are crucial for cell survival. They attain their sensing ability using small molecule binding domains [8,96]. Moreover, recently, sGC *Manduca sexta* (tobacco hornworm) PAS domain was also described to be highly homologous to its mammalian counterparts [6,22]. Its crystal structure at 1.8 Å resolution revealed, four alpha helices and six beta sheets, a small internal pocket and inter-domain contact residues (Figure 5A) [13,22]. Recently it has been suggested that the PAS domain might have other functions than heterodimerization [12,13].

3.2. 6 Helix coiled coil (CC) domain

The α/β subunits homo- and heterodimerization have been thought to be in a physiological equilibrium [97] with heterodimerization being preferred while homodimeric β1β1 [98] and α1α1 are found to be unstable *in vivo* [99].

Dimerization in sGC has been characterized to be mediated by at least three inter-domain interactions involving the Catalytic, Per/ARNT/Sim (PAS) and helix coiled coil (CC) subdomains (~18 kDa). Domain interactions that were found to be the most pronounced include the α1 440-459 (catalytic domain) [100], β1 C-terminal 204-244 (PAS domain) and residues 379- 408 from CC domain. These dimerization regions have been confirmed using deletions studies [101]. The sequence of the CC region of guanylate cyclases is relatively conserved and is postulated to be part of a larger class called the signaling helix (S-helix) [102].

The sGC heterodimerization arrangement has been suggested to have the CC domain in a parallel conformation [102,103]. Nevertheless, the crystal structure of heterologous *Rattus norvegicus* sGC β1CC domain was obtained at 2.15 Å resolution (Figure 6) containing an antiparallel arrangement [21]. However in the same study, experiments using chemical cross linking showed that in solution the coiled coil was in the parallel conformation. These contradictory observations could be either due to crystallization artifacts or due to the fact that in the absence of the α1 subunit the CC β1/β1 exists in equilibrium between parallel and antiparallel conformations.

In these structures, each individual monomer was shown to be comprised of a long α-helix, a turn near residue P399, and a short second α-helix. This structure revealed molecular details related to residues involved in dimerization and also the important role for other faces of the CC dimer (L394 and P399) which might perhaps interact with adjacent sGC domains [21].
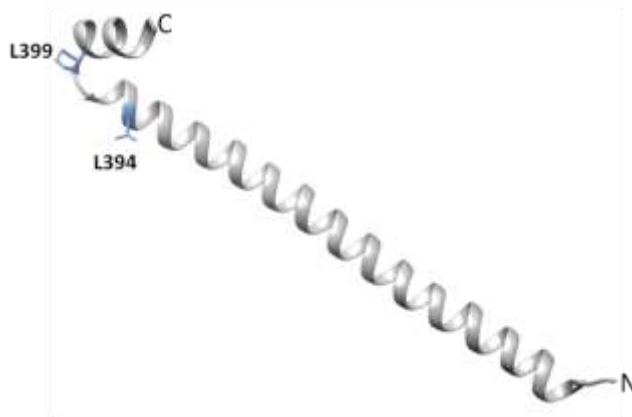
Figure 6: Structure of β1 CC sGC from *Rattus norvegicus*.
The L394 and P399 residues are highlight in light blue (PBD entry 3HLS). Adapted from [21,104].

3.2.7 sGC Catalytic domain

Catalytic domain from both particulate and soluble guanylate cyclases, and also from Adenylate cyclase (AC class III nucleotidyl cyclases) require divalent

cations (Mn^{2+} or Mg^{2+}) as substrate cofactors and allosteric modulators to express maximum catalytic activity. All nucleotide cyclases require that their purine nucleotide substrates form a chelate with a divalent cation to bind to the catalytic domain and undergo enzymatic cyclization of GTP and/or AMP substrate into cGMP and/or cAMP [105]. The crystal structure of human Adenylate cyclase (AC) is the most studied so far. It was determined at 2.3 Å resolution, and showed that the ATP binding site, Mn^{2+}/Mg^{2+} coordination site and transition state stabilization site are formed by two distinct domains, called C1 and C2 [106].

The catalytic sites of the sGC (α1/β1) and AC enzymes share significant homology, with the residues involved in catalysis activation being conserved (Figure 7 A-B). These include, Aspartates in positions 1 and 3 (residues D486 and D530 of the sGC α subunit) which were showed to be key residues for metal coordination. Residues in positions 4, 7 and 8 (R574 of sGCα1, R552 and K593 of sGCβ1, respectively) are known to stabilize the c-phosphate of the substrate. R552 sGCβ1 co-localizes with the equivalent residue in the AC structure (R1029). The side chain of R574 (a) is pointing away from the suggested binding site. A loop (β5–β6) of the beta subunit is shifted inside in sGC β1 compared to the AC-C2 domain, thus, the residue K593 sGCβ1 does not overlap with the equivalent K1065 of AC. This loop may change position upon binding of nucleotide or upon physiological activation of the cyclase. The altered position may allow different interactions with the triphosphate moiety. Residue N548 of sGCβ1 overlaps with the equivalent residue N1025 in AC (position 6), which binds the ribose moiety. Positions 2 and 5 differ between AC and sGC proteins, and underline the preference for adenosine and guanosine nucleotides, respectively. E473 sGCβ1 and C541 sGCβ1 are required for binding to the guanine of GTP (Figure 7B), presumably with the cysteine forming a hydrogen bond to the O6 of guanine with the glutamate accepting hydrogen bonds from the N1 and N2 of guanine [1].

The heterodimeric human α1/β1 catalytic domain of sGC was crystallized at 1.9 Å resolution [10]. The resulting structure showed that the two heterodimers interacted in a head to tail fashion, with the N terminal site of one being in contact with the C terminal site of the other (Figure 7 A and B)

However, in the crystallized structure, there was no metal (Mn^{2+} and Mg^{2+}) or substrate (GTP) bound (Figure 7 B), therefore, it does not represent the catalytic domain in active form [10], when compared to that of active adenylate cyclase structure [107]. Nevertheless, the human α1/β1 catalytic domain crystal structure revealed residues, which have been shown to be the key for sGC activation. These residues fall in two areas of the dimer interface. The first region includes residues located in the region that interacting with the flap on the partner subunit. The second includes residues modulating sGC activity, which are located in the core of the heterodimer at the interface between both subunits (Figure 7 A). This evidence was strongly supported by studies that showed mutations that eliminate the capacity to form hydrogen bonds at this interface severely decreased sGC activity (α1 C595Y, αGC C595D, αGC E526A, and C595D/E526K) [108,109] and adenylate cyclase activity (C1α K436A and D505A) [10].
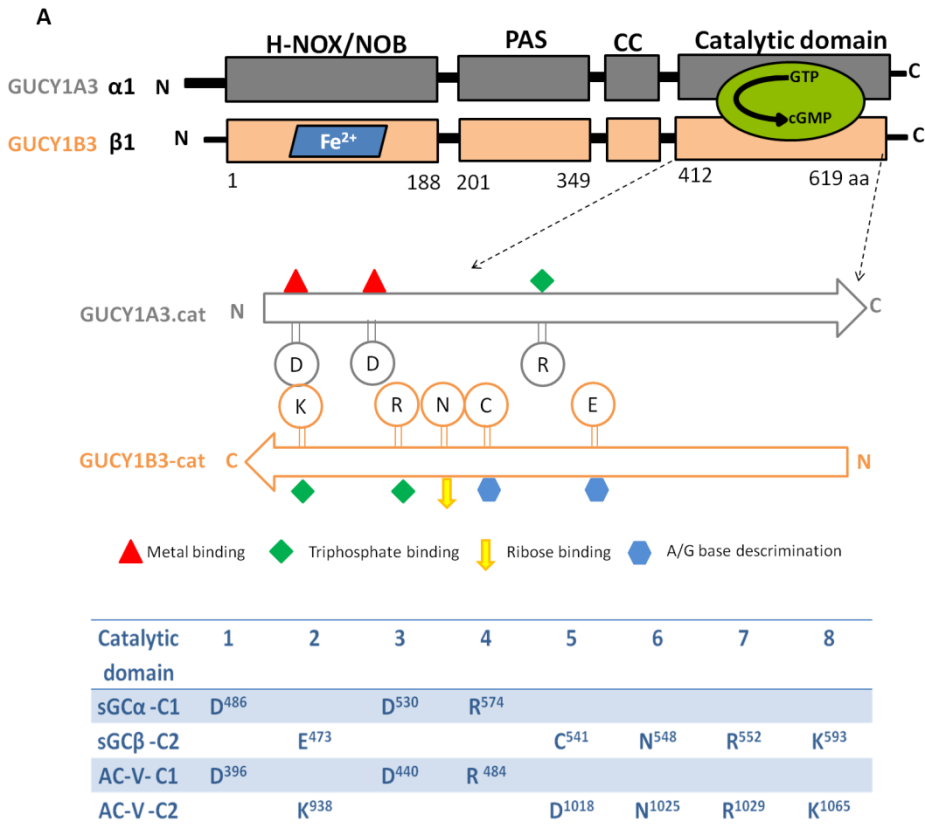
Figure 7: A - **Domain organization of human soluble guanylate cyclase.**
This enzyme is heterodimeric composed of the α/β subunits (encoded by GUCY1A3 and GUCY1B3, respectively). The interaction interfaces spread across the PAS, CC and cyclase catalytic domains. The catalytic domains associate in a head to tail orientation; conserved residues involved in substrate binding and catalysis are distributed between the two subunits. Table list are shown the conserved active-site residues in guanylate cyclase (sGC α1/β1) and the homologous adenylate cyclase (AC-V C1 domain and AC-II C2 domain). Adapted from [1,10].
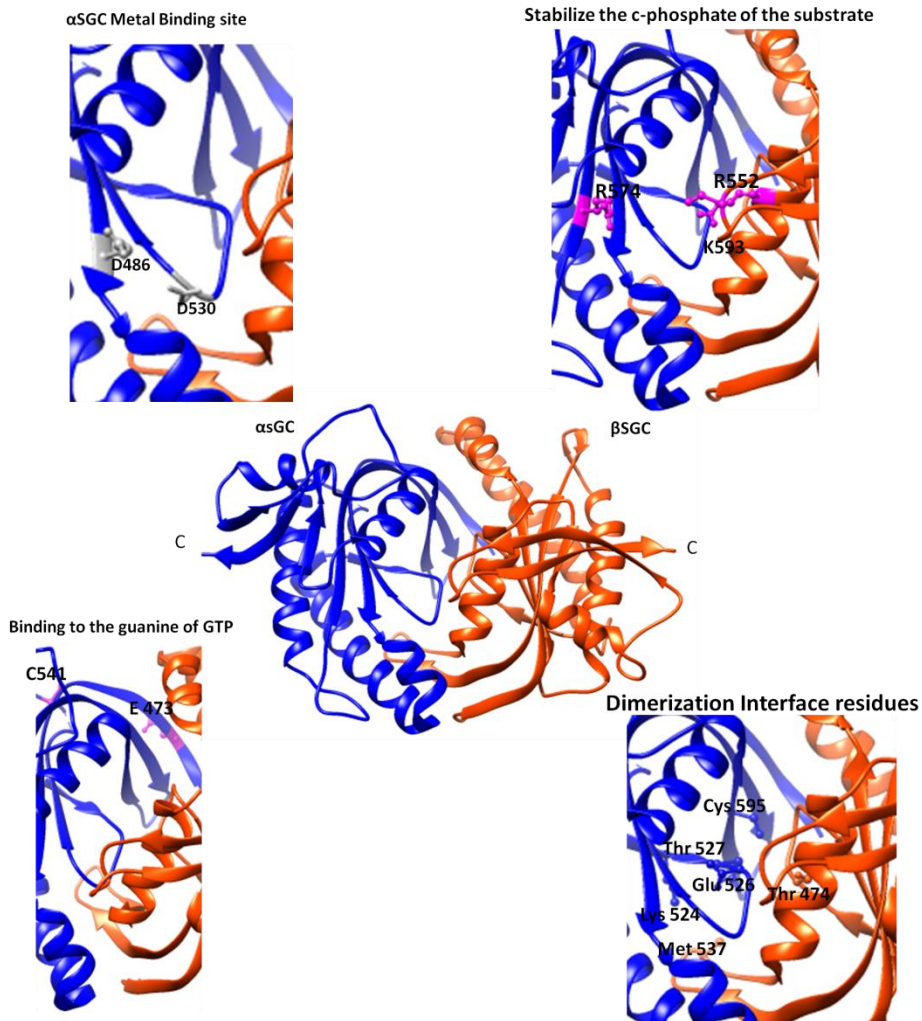
Figure 7: **B** – 3D structure model of human heterodimeric α1/β1 GC catalytic domains. It resembles the Chinese yin-yang symbol with both subunits arranged in a head-to-tail conformation. The α GC subunit is colored blue, and the β GC subunit is colored orange. The metal binding site, the residues responsible to c-phosphate substrate stability, the guanine GTP site and dimerization interface residues are highlight (PBD entry 3UVJ). Adapted from [1,10].

3.2.7 sGC quaternary structure

Focusing on the capability of the β1 H-NOX to inhibit activity of the heterodimer catalytic domains (α1/β1), Winger and collaborators suggested a direct β1 H-NOX/C-terminal inter-domain interaction [110]. In support of such

theory that full sGC activity is regulated by the H-NOX domains, published work indicated that the domains within sGC are organized in a way that allows for direct interaction of the amino-terminal regulatory domains with the carboxy-terminal catalytic site [13,111]. Based on high level of sequence conservation in the α1 C-terminal domain (residues 616–662), these residues were suggested to play an important role in the assembly and/or regulation of sGC. [10]. Mutation of α1 R624A residue, which is in the structural model solvent-exposed, triggered a considerable increase in full-length sGC activity [108]. Thus, this residue was suggested to be a key for domain–domain interaction with the N-terminal regulatory domains [10].

Since no sGC holo-enzyme structure is available so far [10,23], how NO binds to the N-terminal β1 H-NOX sensor domain is communicated to the catalytic site of sGC remains a central question [12]. However, recently the sGC architecture has been reported based on several approaches. For example, using the Förster Resonance Energy Transfer (FRET) technique, it was shown that NO binds to amino-terminal H-NOX domain and propagates the activation signal to the catalytic domain through the coiled-coil helix and by direct interaction between the H-NOX and catalytic domain [24]. This result fits well with Fritz et al study, in which, they analyzed a carboxyl-terminally truncated NO-sGC from *Manduca sexta* (tobacco hornworm) using analytical ultracentrifugation, small angle x-ray scattering (SAX), cross-linking, and tandem mass spectrometry and found 20 intramolecular contact residues. The resulting model displayed a central parallel coiled-coil platform upon which the H-NOX and PAS domains are assembled. The β1 H-NOX and α1 PAS domains were in contact and form the core signaling complex, while the α1 H-NOB domain could be removed without a significant effect on ligand binding or overall shape [6].

The functional role of PAS and CC domain in signal communication was early discovered by mutational screens [100]. Furthermore, the flexibility of the PAS/CC

parallels domains were a result of a recent hydrogen deuterium exchange (HDX) investigation, in which it was found that the NO led to an increase in conformational flexibility at the linker between the PAS and helical domains. [12].

The holo-enzyme quaternary structure model recently published, confirmed an inter-domain interaction [4]. The structural details provided by electronic microscopy (EM) showed numerous pieces of biochemical and structural data and consolidates proposed models to visualize the domain arrangement and interactions. PAS and H-NOX domains directly interact, as previously suggested [6,12]

Furthermore, the quaternary structure model showed that both PAS and H-NOX domains form a tight cluster, sharing large interactions of surfaces, and that the H-NOX/H-NOB domain interacts with both the α1 and β1 PAS domains. This close configuration would allow small-scale changes in the H-NOX domain to be quickly recognized by the adjacent PAS. A growing body of evidence also supports a direct interaction between the two opposite ends of the protein, the N-terminal H-NOX and C-terminal catalytic domain [13,110,111]. In both unligated/ligated NO-sGC, the H-NOX can make direct contact with the catalytic site, which could contribute to allosteric control of the catalytic domain by the H-NOX. The EM data demonstrates that this contact is made possible due to the marked flexibility of the enzyme [4]. Therefore, due to the large enzyme flexibility, the previous holo-sGC inter-domain interaction model suggested by Busker and Fritz et al could represent one of the possible conformations [4,6,24].

4   sGC NO independent pathway

Up to this point, NO has been described as the single activator to sGC/cGMP/PKG signaling. However, endogenously, changes in the redox state of heme-sGC can be induced by reactive oxygen and nitrogen species such as superoxide ($\bullet O_2^-$) and peroxynitrite ($ONOO^-$), which are generated under

conditions of oxidative stress [112]. In these conditions, the ferrous (Fe^{2+}) form of heme of sGC is oxidized to the ferric form (Fe^{3+}). The ferrous heme is required to covalently bind with H105 located in the αF-helix. Because oxidized heme chemically requires a stronger electron donor than the imidazole side chain of histidine, upon oxidation, heme affinity to sGC heme pocket is strongly reduced. This loss of Histidine binding, leads to heme loss from the protein. The apo-sGC is gradually affected by ubiquitin protein leading sGC to be degraded by proteasome enzyme. Thus, as a consequence of oxidative stress, the production of cGMP through NO-sGC dependent pathway is inactivated [5]. This inactivation is associated with heart failure [11].

Most treatments of cardiovascular disorders mainly focus on the use of organic nitrates, which have been used to manage heart disease for over a century [113]. However, organic nitrates have several important limitations like numerous cytotoxic effects, mostly attributed to the reactive oxidant products of NO. Peroxynitrite interacts with proteins and lipids, altering cellular signaling, disrupting mitochondrial function, and damaging DNA, which can eventually culminate in cellular dysfunction and/or death [19]. These limitations have been associated with resistance to NO and organic nitrates [19,113]. For these reasons, nowadays, efforts have been centered on identifying pharmacological agents, which could target sGC in a NO-independent manner [114].

4.1   sGC stimulators

sGC enzyme plays a crucial role in cGMP synthesis and the regulation of cardiovascular process [4]. Pharmaceutical scientists from Bayer HealthCare AG´s, in middle of 1990, were looking for a molecule able to increase NO synthesis to stimulate sGC in porcine endothelial cells. In parallel, the indazole derivative YC-1 was characterized as an NO-independent, heme-dependent stimulator of highly purified sGC [115]. Over the following decade, several compounds were tested in

an attempt to enhance sGC activity in an NO-independent pathway [115-117]. Among these were BAY 41-2272, BAY 41-8543, CMF-1571 and A-350619 compounds. Based on several shared characteristics, such as, dependency on the presence of the reduced prosthetic heme moiety and strong synergistic enzyme activation when combined with NO these drugs were called as heme-dependent sGC-stimulators [115-117].

In order to further pharmacological advances, automated ultra-high-throughput screening was performed. The BAY W 1449 (amino dicarboxylic acid) compound sGC activator was identified upon screening process of several compounds [19,126]. In 2002, Cinaciguat or BAY 58-2667 was selected from a series of analogues as the first NO-independent activator of sGC that showed completely different characteristics to any of the known heme-dependent sGC stimulators [19].

4.1.1 Molecular mechanisms of Cinaciguat sGC activation

Cinaciguat was described to be able to directly activate sGC with EC_{50} and K_d values in the low nanomolar range, making this drug the most potent NO-independent sGC activator [19,20].

The molecular binding of cinaciguat into sGC was described to be via their negatively charged carboxylic groups for the unique sGC heme-binding motif Y-x-S-x-R [88-90]. In fact, activity binding assays and spectroscopic studies supported by structural alignments were able to postulate that just as heme, cinaciguat binds to the sGC free pocket by bonds to the motif Y-x-S-x-R residues [90].

Cinaciguat mimics heme in its structure and binds to the sGC free pocket (Figure 8) [90]. Cinaciguat binding has also been shown to protect the protein from proteasomal degradation. The insertion of cinaciguat into sGC is able to reactivate cGMP production preserving the sGC/cGMP/PKG signaling [75,88-90].
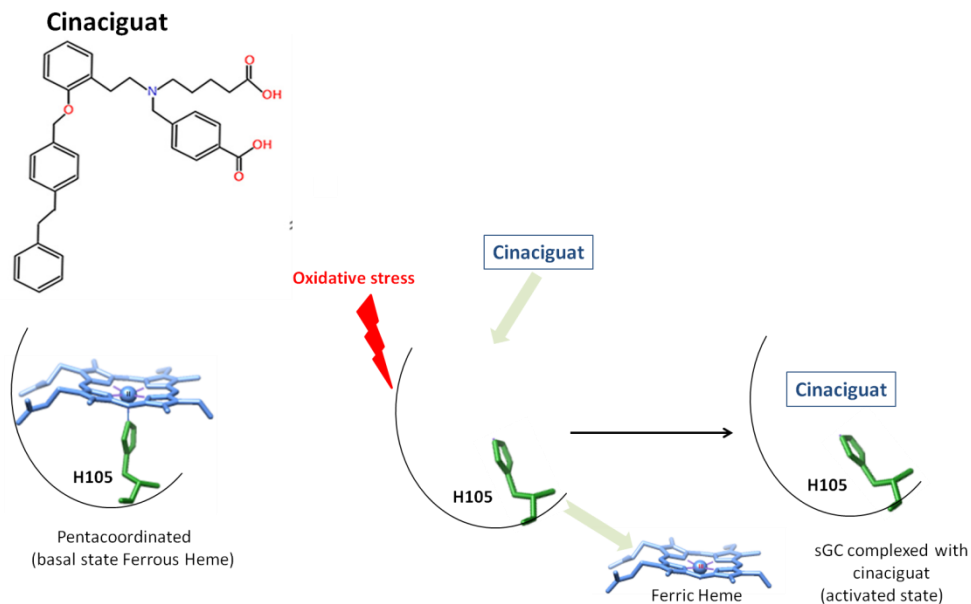
Figure 8: Cinaciguat compound and their mechanisms of sGC activation.
Reactivation of sGC upon oxidative stress requires cinaciguat to occupy the heme free
pocket. Adapted from [75,112].

4.1.2 Cinaciguat therapeutic applications in Heart failure.

Heart failure (HF) is a chronic condition that affects about 26 million
people worldwide. It has been characterized as one of the major causes of
hospitalization, morbidity and mortality. Also HF has recently been estimated to
cause an annual economic burden over $108 billion worldwide [118,119].

The progression of cardiovascular dysfunction to HF is multifaceted and
involves the activation of several intracellular signaling [17,18]. In HF, sGC/sGC–
cGMP signaling is disrupted either as a result of impaired production of NO, their
excessive degradation, or chemical interactions with oxidants such as superoxide.

Anaesthetized dogs that were under autonomic blockade were treated
with intravenous infusion of cinaciguat or glyceryl trinitrate and the results
showed a decrease in arterial blood pressure and reductions in diastolic
pulmonary artery pressure and right atrial pressure. Moreover, the duration of

vasodilator effect was much longer for cinaciguat than glyceryl trinitrate [19]. Based on these experiments, cinaciguat compound was therefore determined as the first non-NO-based compound that has similar heme dynamic effects to glyceryl trinitrate, leading to *in vivo* dilation of both arterial and venous blood vessels [120].

Cinaciguat was included in clinical studies in patients with acute HF [121,122]. The results of the phase I and II, dose-escalation study together with the pre-clinical data indicated this compound as a potent preload and afterload reducing other cardiovascular effects. These findings suggest a novel mechanism for modulating the sGC/cGMP pathway in the treatment of cardiac disease and for eliminating endothelial dysfunction avoiding vascular superoxide product and development of tolerance [112,122] .

4.1.4 Crystal structure of Cinaciguat bound to *Nostoc sp* H-NOX sGC homologous domain

The crystal structure complex of cinaciguat bound to *Nostoc s* H-NOX was determined at 2.3- Å resolution (Figure 9 A) [75]. Based on the structure analysis, *Ns* H-NOX electron difference density in the heme-binding pockets revealed the trifurcated cinaciguat ligand density. The two carboxylate groups of cinaciguat were shown to make extensive interactions. For example, the carboxy-butyl moiety interacts with the side chains of Y134, S136, and R138, and the carboxylate moiety of the benzoic acid interacts with R138 and the main chain nitrogen of Y2. The hydrophobic third branch of cinaciguat makes extensive hydrophobic interactions with residues forming the rest of the heme cavity including L4, W74, T78, K83, F97, L101, L104, V108, L148, and L152 (Figure 9 B). Furthermore, cinaciguat was able to mimic heme and avoid forming a covalent inhibitory bond with H105. The largest conformational shifts showed to be in the C-terminal part of αF *Ns* H-NOX-cinaciguat bound in comparison with the *Ns* H-NOX heme-bound

structure. Moreover, although cinaciguat has a shape dissimilar to heme, it folds up and occupies the heme cavity with its carboxylate groups superimposing on those of the heme propionate groups [9].
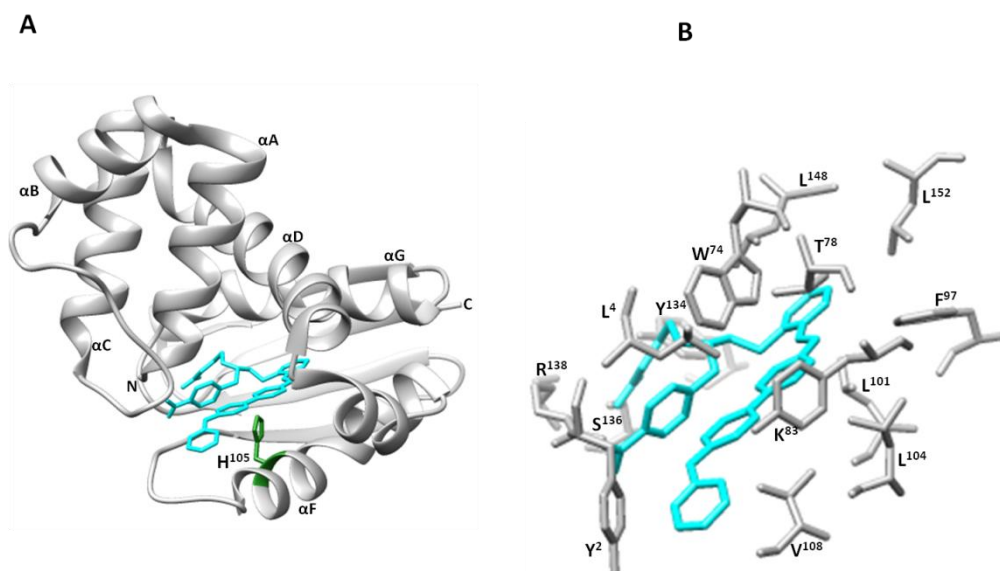
Figure 9: A- 3D structure model of stand-alone *Nostoc sp* H-NOX in complex with cinaciguat activator. Cinaciguat (light blue), H105 (green). **B-** Stereo figure showing interactions of cinaciguat compound with nearby H-NOX residues (PBD entry 4JQH). Adapted from [75].

Together these results corroborate to understand the mode of sGC activation by NO independent compound. These molecular insight information, offers new structure-based strategies to find pharmaceutical important targets for the sGC protein to treat certain cardiovascular diseases.

5 sGC allosteric regulation

Allostery ('allo-steric = other-space') means that action in one part of the molecule causes an effect at another site [123]. sGC allosteric regulation has been

proven to be a complex event, making the sGC activation dependent/independent of the NO, a potential target for drug discovery, able to regulate allosteric mechanism [9,74,75,118,124]. Taking into account the new approach to allosteric regulation, which considers that the native protein state is a conformational ensemble [125-130], three events have been speculated to be the involved in sGC allosteric regulation.

The first event involves the rupture of the Fe^{2+}–H105 bond upon formation of the five-coordinate NO in the proximal pocket that is believed to start a cascade of events that lead to sGC activation [124]. Based on prokaryotic H-NOX sGC homologues, much insight into the structural changes upon NO binding was gathered for *Tt* and *So* H-NOX [78,86]. These structures suggested that removal of the Fe^{2+}–H103 (105 in sGC) penta-coordinate, mimicked by mutant H103G (penta-coordinate imidazole bound to proximal pocket) resulted in conformational changes in the heme cofactor [131]. In *Tt* H-NOX protein the local rearrangements of the heme pocket residues, such as M1 and I5 in H103G made possible by relaxation of the heme translate into large displacements of the N-terminal (greater than 6 Å) by The crystal structure analysis. It is shown that alone H103 is not responsible for such structural changes at the N-terminal. Thus, heme relaxation toward planarity is a second event supposed to be involved in protein activation.

The heme relaxation is a result of van der Waals interactions with residues Isoleucine and Proline that flank the distal and proximal faces of the heme, correspondingly. This hypothesis was supported by elimination (P115A) of steric bulk imparted by Proline (118 in sGC), a residue that is strictly conserved in H-NOX proteins. This single change resulted in a significantly flattened heme [131].

Herzik and collaborators recently reported molecular insight into the mechanism of NO dependent activation in *Shewanella oneidensis* [23]. The analysis of the crystal structure showed several rearrangements in the signaling

αF helix acting to stabilize the heme-histidine in an outside conformation. These rearrangements eliminate the proximal pocket hydrogen-bonding network and release the Proline from its limited position, allowing it to translate away from the heme, and consequently enable the porphyrin to relax back toward planarity. This conformational change communicates NO binding to the effector protein via a binding interface that most probably spans the distal and proximal pocket. Following NO dissociation back to the unligated state, the system is reset, effectively reloading the spring and returning the signaling pathway back to the basal state [23].

Finally, the last event might involve protein dynamics. H-NOX propagates the activation signal to the catalytic site of the protein via inter-domain interactions, and the activation signal might be transferred through changes in protein dynamics, and thus mobility might be involved in sGC allosteric regulation [12,13].

Dynamic properties of proteins have been described to be essential for their biological functions including catalysis, binding, regulation and cellular structure [132,133]. Nuclear Magnetic Resonance (NMR) spectroscopy is a technique that allows the determination of protein structures while at the same time is able to monitor internal protein dynamics phenomena at different timescales. Therefore, in the absence of sGC dynamics information, NMR spectroscopy is a very attractive option for confirming or disproving the adoption of such allosteric mechanisms in individual domains of sGC. Therefore, the goal of this thesis was to solve the 3D structure of human H-NOX and study its dynamic behavior by comparing the domain in the basal, activated and inhibited forms by NMR.

6. References

1. Allerston CK, von Delft F, Gileadi O (2013) Crystal structures of the catalytic domain of human soluble guanylate cyclase. PLoS One 8: e57644.
2. Braughler JM, Mittal CK, Murad F (1979) Purification of soluble guanylate cyclase from rat liver. Proc Natl Acad Sci U S A 76: 219-222.
3. Buechler WA, Nakane M, Murad F (1991) Expression of soluble guanylate cyclase activity requires both enzyme subunits. Biochem Biophys Res Commun 174: 351-357.
4. Campbell MG, Underbakke ES, Potter CS, Carragher B, Marletta MA (2014) Single-particle EM reveals the higher-order domain architecture of soluble guanylate cyclase. Proc Natl Acad Sci U S A 111: 2960-2965.
5. Derbyshire ER, Winter MB, Ibrahim M, Deng S, Spiro TG, et al. (2011) Probing domain interactions in soluble guanylate cyclase. Biochemistry 50: 4281-4290.
6. Fritz BG, Roberts SA, Ahmed A, Breci L, Li W, et al. (2013) Molecular model of a soluble guanylyl cyclase fragment determined by small-angle X-ray scattering and chemical cross-linking. Biochemistry 52: 1568-1582.
7. Lucas KA, Pitari GM, Kazerounian S, Ruiz-Stewart I, Park J, et al. (2000) Guanylyl cyclases and signaling by cyclic GMP. Pharmacol Rev 52: 375-414.
8. Ma X, Sayed N, Baskaran P, Beuve A, van den Akker F (2008) PAS-mediated dimerization of soluble guanylyl cyclase revealed by signal transduction histidine kinase domain crystal structure. J Biol Chem 283: 1167-1178.
9. Ma X, Sayed N, Beuve A, van den Akker F (2007) NO and CO differentially activate soluble guanylyl cyclase via a heme pivot-bend mechanism. Embo j 26: 578-588.
10. Seeger F, Quintyn R, Tanimoto A, Williams GJ, Tainer JA, et al. (2014) Interfacial residues promote an optimal alignment of the catalytic center in human soluble guanylate cyclase: heterodimerization is required but not sufficient for activity. Biochemistry 53: 2153-2165.
11. Zhong F, Pan J, Liu X, Wang H, Ying T, et al. (2011) A novel insight into the heme and NO/CO binding mechanism of the alpha subunit of human soluble guanylate cyclase. J Biol Inorg Chem 16: 1227-1239.
12. Underbakke ES, Iavarone AT, Chalmers MJ, Pascal BD, Novick S, et al. (2014) Nitric oxide-induced conformational changes in soluble guanylate cyclase. Structure 22: 602-611.
13. Underbakke ES, Iavarone AT, Marletta MA (2013) Higher-order interactions bridge the nitric oxide receptor and catalytic domains of soluble guanylate cyclase. Proc Natl Acad Sci U S A 110: 6777-6782.
14. Francis SH, Busch JL, Corbin JD, Sibley D (2010) cGMP-dependent protein kinases and cGMP phosphodiesterases in nitric oxide and cGMP action. Pharmacol Rev 62: 525-563.

15. Lincoln TM, Dey NB, Boerth NJ, Cornwell TL, Soff GA (1998) Nitric oxide--cyclic GMP pathway regulates vascular smooth muscle cell phenotypic modulation: implications in vascular diseases. Acta Physiol Scand 164: 507-515.

16. Boerrigter G, Burnett JC, Jr. (2004) Recent advances in natriuretic peptides in congestive heart failure. Expert Opin Investig Drugs 13: 643-652.

17. Hare JM, Stamler JS (2005) NO/redox disequilibrium in the failing heart and cardiovascular system. J Clin Invest 115: 509-517.

18. Ungvari Z, Gupte SA, Recchia FA, Batkai S, Pacher P (2005) Role of oxidative-nitrosative stress and downstream pathways in various forms of cardiomyopathy and heart failure. Curr Vasc Pharmacol 3: 221-229.

19. Stasch JP, Schmidt P, Alonso-Alija C, Apeler H, Dembowsky K, et al. (2002) NO-and haem-independent activation of soluble guanylyl cyclase: molecular basis and cardiovascular implications of a new pharmacological principle. Br J Pharmacol 136: 773-783.

20. Schmidt P, Schramm M, Schroder H, Stasch JP (2003) Mechanisms of nitric oxide independent activation of soluble guanylyl cyclase. Eur J Pharmacol 468: 167-174.

21. Ma X, Beuve A, van den Akker F (2010) Crystal structure of the signaling helix coiled-coil domain of the beta1 subunit of the soluble guanylyl cyclase. BMC Struct Biol 10: 2.

22. Purohit R, Weichsel A, Montfort WR (2013) Crystal structure of the Alpha subunit PAS domain from soluble guanylyl cyclase. Protein Sci 22: 1439-1444.

23. Herzik MA, Jr., Jonnalagadda R, Kuriyan J, Marletta MA (2014) Structural insights into the role of iron-histidine bond cleavage in nitric oxide-induced activation of H-NOX gas sensor proteins. Proc Natl Acad Sci U S A 111: E4156-4164.

24. Busker M, Neidhardt I, Behrends S (2014) Nitric oxide activation of guanylate cyclase pushes the alpha1 signaling helix and the beta1 heme-binding domain closer to the substrate-binding site. J Biol Chem 289: 476-484.

25. Francis SH, Blount MA, Zoraghi R, Corbin JD (2005) Molecular properties of mammalian proteins that interact with cGMP: protein kinases, cation channels, phosphodiesterases, and multi-drug anion transporters. Front Biosci 10: 2097-2117.

26. Schlossmann J, Desch M (2009) cGK substrates. Handb Exp Pharmacol: 163-193.

27. Pfeifer A, Ruth P, Dostmann W, Sausbier M, Klatt P, et al. (1999) Structure and function of cGMP-dependent protein kinases. Rev Physiol Biochem Pharmacol 135: 105-149.

28. Tamura N, Itoh H, Ogawa Y, Nakagawa O, Harada M, et al. (1996) cDNA cloning and gene expression of human type Ialpha cGMP-dependent protein kinase. Hypertension 27: 552-557.

29. Barger SW, Fiscus RR, Ruth P, Hofmann F, Mattson MP (1995) Role of cyclic GMP in the regulation of neuronal calcium and survival by secreted forms of beta-amyloid precursor. J Neurochem 64: 2087-2096.

30. Nakamizo T, Kawamata J, Yoshida K, Kawai Y, Kanki R, et al. (2003) Phosphodiesterase inhibitors are neuroprotective to cultured spinal motor neurons. J Neurosci Res 71: 485-495.

31. Paquet-Durand F, Hauck SM, van Veen T, Ueffing M, Ekstrom P (2009) PKG activity causes photoreceptor cell death in two retinitis pigmentosa models. J Neurochem 108: 796-810.

32. Costa AD, Garlid KD, West IC, Lincoln TM, Downey JM, et al. (2005) Protein kinase G transmits the cardioprotective signal from cytosol to mitochondria. Circ Res 97: 329-336.

33. Ockaili R, Salloum F, Hawkins J, Kukreja RC (2002) Sildenafil (Viagra) induces powerful cardioprotective effect via opening of mitochondrial K(ATP) channels in rabbits. Am J Physiol Heart Circ Physiol 283: H1263-1269.

34. Salloum FN, Chau VQ, Hoke NN, Abbate A, Varma A, et al. (2009) Phosphodiesterase-5 inhibitor, tadalafil, protects against myocardial ischemia/reperfusion through protein-kinase g-dependent generation of hydrogen sulfide. Circulation 120: S31-36.

35. Shimizu-Albergine M, Rybalkin SD, Rybalkina IG, Feil R, Wolfsgruber W, et al. (2003) Individual cerebellar Purkinje cells express different cGMP phosphodiesterases (PDEs): in vivo phosphorylation of cGMP-specific PDE (PDE5) as an indicator of cGMP-dependent protein kinase (PKG) activation. J Neurosci 23: 6452-6459.

36. Wilson LS, Elbatarny HS, Crawley SW, Bennett BM, Maurice DH (2008) Compartmentation and compartment-specific regulation of PDE5 by protein kinase G allows selective cGMP-mediated regulation of platelet functions. Proc Natl Acad Sci U S A 105: 13650-13655.

37. Wyatt TA, Naftilan AJ, Francis SH, Corbin JD (1998) ANF elicits phosphorylation of the cGMP phosphodiesterase in vascular smooth muscle cells. Am J Physiol 274: H448-455.

38. Corbin JD, Francis SH (1999) Cyclic GMP phosphodiesterase-5: target of sildenafil. J Biol Chem 274: 13729-13732.

39. Francis SH, Corbin JD, Bischoff E (2009) Cyclic GMP-hydrolyzing phosphodiesterases. Handb Exp Pharmacol: 367-408.

40. Rybalkin SD, Rybalkina IG, Feil R, Hofmann F, Beavo JA (2002) Regulation of cGMP-specific phosphodiesterase (PDE5) phosphorylation in smooth muscle cells. J Biol Chem 277: 3310-3317.

41. Baskaran P, Heckler EJ, van den Akker F, Beuve A (2011) Identification of residues in the heme domain of soluble guanylyl cyclase that are important for basal and stimulated catalytic activity. PLoS One 6: e26976.

42. Bian K, Murad F (2014) sGC-cGMP Signaling: Target for Anticancer Therapy. Adv Exp Med Biol 814: 5-13.

43. Boon EM, Huang SH, Marletta MA (2005) A molecular basis for NO selectivity in soluble guanylate cyclase. Nat Chem Biol 1: 53-59.

44. Boon EM, Marletta MA (2005) Ligand discrimination in soluble guanylate cyclase and the H-NOX family of heme sensor proteins. Curr Opin Chem Biol 9: 441-446.

45. Hardman JG, Sutherland EW (1969) Guanyl cyclase, an enzyme catalyzing the formation of guanosine 3',5'-monophosphate from guanosine trihosphate. J Biol Chem 244: 6363-6370.

46. Ishikawa E, Ishikawa S, Davis JW, Sutherland EW (1969) Determination of guanosine 3',5'-monophosphate in tissues and of guanyl cyclase in rat intestine. J Biol Chem 244: 6371-6376.

47. White AA, Aurbach GD (1969) Detection of guanyl cyclase in mammalian tissues. Biochim Biophys Acta 191: 686-697.

48. Wedel BJ, Garbers DL (1997) New insights on the functions of the guanylyl cyclase receptors. FEBS Lett 410: 29-33.

49. Wedel BJ, Garbers DL (1998) Guanylyl cyclases: approaching year thirty. Trends Endocrinol Metab 9: 213-219.

50. Wedel B, Garbers D (2001) The guanylyl cyclase family at Y2K. Annu Rev Physiol 63: 215-233.

51. Laney DW, Jr., Mann EA, Dellon SC, Perkins DR, Giannella RA, et al. (1992) Novel sites for expression of an Escherichia coli heat-stable enterotoxin receptor in the developing rat. Am J Physiol 263: G816-821.

52. London RM, Eber SL, Visweswariah SS, Krause WJ, Forte LR (1999) Structure and activity of OK-GC: a kidney receptor guanylate cyclase activated by guanylin peptides. Am J Physiol 276: F882-891.

53. Vaandrager AB, Schulz S, De Jonge HR, Garbers DL (1993) Guanylyl cyclase C is an N-linked glycoprotein receptor that accounts for multiple heat-stable enterotoxin-binding proteins in the intestine. J Biol Chem 268: 2174-2179.

54. Fulle HJ, Vassar R, Foster DC, Yang RB, Axel R, et al. (1995) A receptor guanylyl cyclase expressed specifically in olfactory sensory neurons. Proc Natl Acad Sci U S A 92: 3571-3575.

55. Pugh EN, Jr., Duda T, Sitaramayya A, Sharma RK (1997) Photoreceptor guanylate cyclases: a review. Biosci Rep 17: 429-473.

56. Sharma RK (2010) Membrane guanylate cyclase is a beautiful signal transduction machine: overview. Mol Cell Biochem 334: 3-36.

57. Schulz S, Wedel BJ, Matthews A, Garbers DL (1998) The cloning and expression of a new guanylyl cyclase orphan receptor. J Biol Chem 273: 1032-1037.

58. Derbyshire ER, Deng S, Marletta MA (2010) Incorporation of tyrosine and glutamine residues into the soluble guanylate cyclase heme distal pocket alters NO and O2 binding. J Biol Chem 285: 17471-17478.

59. Fernhoff NB, Derbyshire ER, Marletta MA (2009) A nitric oxide/cysteine interaction mediates the activation of soluble guanylate cyclase. Proc Natl Acad Sci U S A 106: 21602-21607.

60. Zhang Y, Lu M, Cheng Y, Li Z (2010) H-NOX domains display different tunnel systems for ligand migration. J Mol Graph Model 28: 814-819.

61. Middelhaufe S, Leipelt M, Levin LR, Buck J, Steegborn C (2012) Identification of a haem domain in human soluble adenylate cyclase. Biosci Rep 32: 491-499.

62. Ignarro LJ (1999) Nitric oxide: a unique endogenous signaling molecule in vascular biology. Biosci Rep 19: 51-71.

63. Madhusoodanan KS, Murad F (2007) NO-cGMP signaling and regenerative medicine involving stem cells. Neurochem Res 32: 681-694.

64. Ignarro LJ, Byrns RE, Buga GM, Wood KS (1987) Endothelium-derived relaxing factor from pulmonary artery and vein possesses pharmacologic and chemical properties identical to those of nitric oxide radical. Circ Res 61: 866-879.

65. Celermajer DS, Cullen S, Deanfield JE (1993) Impairment of endothelium-dependent pulmonary artery relaxation in children with congenital heart disease and abnormal pulmonary hemodynamics. Circulation 87: 440-446.

66. Gratzke C, Angulo J, Chitaley K, Dai YT, Kim NN, et al. (2010) Anatomy, physiology, and pathophysiology of erectile dysfunction. J Sex Med 7: 445-475.

67. Musicki B, Burnett AL (2007) Endothelial dysfunction in diabetic erectile dysfunction. Int J Impot Res 19: 129-138.

68. Koesling D, Herz J, Gausepohl H, Niroomand F, Hinsch KD, et al. (1988) The primary structure of the 70 kDa subunit of bovine soluble guanylate cyclase. FEBS Lett 239: 29-34.

69. Nakane M, Saheki S, Kuno T, Ishii K, Murad F (1988) Molecular cloning of a cDNA coding for 70 kilodalton subunit of soluble guanylate cyclase from rat lung. Biochem Biophys Res Commun 157: 1139-1147.

70. Yuen PS, Potter LR, Garbers DL (1990) A new form of guanylyl cyclase is preferentially expressed in rat kidney. Biochemistry 29: 10872-10878.

71. Gupta G, Azam M, Yang L, Danziger RS (1997) The beta2 subunit inhibits stimulation of the alpha1/beta1 form of soluble guanylyl cyclase by nitric oxide. Potential relevance to regulation of blood pressure. J Clin Invest 100: 1488-1492.

72. Harteneck C, Wedel B, Koesling D, Malkewitz J, Bohme E, et al. (1991) Molecular cloning and expression of a new alpha-subunit of soluble

guanylyl cyclase. Interchangeability of the alpha-subunits of the enzyme. FEBS Lett 292: 217-222.

73. Giuili G, Scholl U, Bulle F, Guellaen G (1992) Molecular cloning of the cDNAs coding for the two subunits of soluble guanylyl cyclase from human brain. FEBS Lett 304: 83-88.

74. Kumar V, Martin F, Hahn MG, Schaefer M, Stamler JS, et al. (2013) Insights into BAY 60-2770 activation and S-nitrosylation-dependent desensitization of soluble guanylyl cyclase via crystal structures of homologous nostoc H-NOX domain complexes. Biochemistry 52: 3601-3608.

75. Martin F, Baskaran P, Ma X, Dunten PW, Schaefer M, et al. (2010) Structure of cinaciguat (BAY 58-2667) bound to Nostoc H-NOX domain reveals insights into heme-mimetic activation of the soluble guanylyl cyclase. J Biol Chem 285: 22651-22657.

76. Iyer LM, Anantharaman V, Aravind L (2003) Ancient conserved domains shared by animal soluble guanylyl cyclases and bacterial signaling proteins. BMC Genomics 4: 5.

77. Karow DS, Pan D, Tran R, Pellicena P, Presley A, et al. (2004) Spectroscopic characterization of the soluble guanylate cyclase-like heme domains from Vibrio cholerae and Thermoanaerobacter tengcongensis. Biochemistry 43: 10203-10211.

78. Erbil WK, Price MS, Wemmer DE, Marletta MA (2009) A structural basis for H-NOX signaling in Shewanella oneidensis by trapping a histidine kinase inhibitory conformation. Proc Natl Acad Sci U S A 106: 19753-19760.

79. Derbyshire ER, Marletta MA (2007) Butyl isocyanide as a probe of the activation mechanism of soluble guanylate cyclase. Investigating the role of non-heme nitric oxide. J Biol Chem 282: 35741-35748.

80. Martin E, Berka V, Sharina I, Tsai AL (2012) Mechanism of binding of NO to soluble guanylyl cyclase: implication for the second NO binding to the heme proximal site. Biochemistry 51: 2737-2746.

81. Boehning D, Snyder SH (2003) Novel neural modulators. Annu Rev Neurosci 26: 105-131.

82. Ryter SW, Alam J, Choi AM (2006) Heme oxygenase-1/carbon monoxide: from basic science to therapeutic applications. Physiol Rev 86: 583-650.

83. Stone JR, Marletta MA (1995) The ferrous heme of soluble guanylate cyclase: formation of hexacoordinate complexes with carbon monoxide and nitrosomethane. Biochemistry 34: 16397-16403.

84. Kharitonov VG, Russwurm M, Magde D, Sharma VS, Koesling D (1997) Dissociation of nitric oxide from soluble guanylate cyclase. Biochem Biophys Res Commun 239: 284-286.

85. Price MS, Chao LY, Marletta MA (2007) Shewanella oneidensis MR-1 H-NOX regulation of a histidine kinase by nitric oxide. Biochemistry 46: 13677-13683.

86. Olea C, Jr., Herzik MA, Jr., Kuriyan J, Marletta MA (2010) Structural insights into the molecular mechanism of H-NOX activation. Protein Sci 19: 881-887.

87. Nioche P, Berka V, Vipond J, Minton N, Tsai AL, et al. (2004) Femtomolar sensitivity of a NO sensor from Clostridium botulinum. Science 306: 1550-1553.

88. Pellicena P, Karow DS, Boon EM, Marletta MA, Kuriyan J (2004) Crystal structure of an oxygen-binding heme domain related to soluble guanylate cyclases. Proc Natl Acad Sci U S A 101: 12854-12859.

89. Schmidt PM, Rothkegel C, Wunder F, Schroder H, Stasch JP (2005) Residues stabilizing the heme moiety of the nitric oxide sensor soluble guanylate cyclase. Eur J Pharmacol 513: 67-74.

90. Schmidt PM, Schramm M, Schroder H, Wunder F, Stasch JP (2004) Identification of residues crucially involved in the binding of the heme moiety of soluble guanylate cyclase. J Biol Chem 279: 3025-3032.

91. Martin E, Berka V, Bogatenkova E, Murad F, Tsai AL (2006) Ligand selectivity of soluble guanylyl cyclase: effect of the hydrogen-bonding tyrosine in the distal heme pocket on binding of oxygen, nitric oxide, and carbon monoxide. J Biol Chem 281: 27836-27845.

92. Jain R, Chan MK (2003) Mechanisms of ligand discrimination by heme proteins. J Biol Inorg Chem 8: 1-11.

93. Nema V, Pal SK (2013) Exploration of freely available web-interfaces for comparative homology modelling of microbial proteins. Bioinformation 9: 796-801.

94. Moglich A, Ayers RA, Moffat K (2009) Structure and signaling mechanism of Per-ARNT-Sim domains. Structure 17: 1282-1294.

95. Huang ZJ, Edery I, Rosbash M (1993) PAS is a dimerization domain common to Drosophila period and several transcription factors. Nature 364: 259-262.

96. Ohmori M, Ikeuchi M, Sato N, Wolk P, Kaneko T, et al. (2001) Characterization of genes encoding multi-domain proteins in the genome of the filamentous nitrogen-fixing Cyanobacterium anabaena sp. strain PCC 7120. DNA Res 8: 271-284.

97. Zabel U, Hausler C, Weeger M, Schmidt HH (1999) Homodimerization of soluble guanylyl cyclase subunits. Dimerization analysis using a glutathione s-transferase affinity tag. J Biol Chem 274: 18149-18152.

98. Mergia E, Friebe A, Dangel O, Russwurm M, Koesling D (2006) Spare guanylyl cyclase NO receptors ensure high NO sensitivity in the vascular system. J Clin Invest 116: 1731-1737.

99. Friebe A, Mergia E, Dangel O, Lange A, Koesling D (2007) Fatal gastrointestinal obstruction and hypertension in mice lacking nitric oxide-sensitive guanylyl cyclase. Proc Natl Acad Sci U S A 104: 7699-7704.

100. Rothkegel C, Schmidt PM, Atkins DJ, Hoffmann LS, Schmidt HH, et al. (2007) Dimerization region of soluble guanylate cyclase characterized by bimolecular fluorescence complementation in vivo. Mol Pharmacol 72: 1181-1190.

101. Zhou Z, Gross S, Roussos C, Meurer S, Muller-Esterl W, et al. (2004) Structural and functional characterization of the dimerization region of soluble guanylyl cyclase. J Biol Chem 279: 24935-24943.

102. Anantharaman V, Balaji S, Aravind L (2006) The signaling helix: a common functional theme in diverse signaling proteins. Biol Direct 1: 25.

103. Ramamurthy V, Tucker C, Wilkie SE, Daggett V, Hunt DM, et al. (2001) Interactions within the coiled-coil domain of RetGC-1 guanylyl cyclase are optimized for regulation rather than for high affinity. J Biol Chem 276: 26218-26229.

104. Boyle AL, Woolfson DN (2011) De novo designed peptides for biological applications. Chem Soc Rev 40: 4295-4306.

105. Waldman SA, Murad F (1987) Cyclic GMP synthesis and function. Pharmacol Rev 39: 163-196.

106. Tesmer JJ, Sunahara RK, Gilman AG, Sprang SR (1997) Crystal structure of the catalytic domains of adenylyl cyclase in a complex with Gsalpha.GTPgammaS. Science 278: 1907-1916.

107. Tesmer JJ, Sunahara RK, Johnson RA, Gosselin G, Gilman AG, et al. (1999) Two-metal-Ion catalysis in adenylyl cyclase. Science 285: 756-760.

108. Yuen PS, Doolittle LK, Garbers DL (1994) Dominant negative mutants of nitric oxide-sensitive guanylyl cyclase. J Biol Chem 269: 791-793.

109. Lamothe M, Chang FJ, Balashova N, Shirokov R, Beuve A (2004) Functional characterization of nitric oxide and YC-1 activation of soluble guanylyl cyclase: structural implication for the YC-1 binding site? Biochemistry 43: 3039-3048.

110. Winger JA, Marletta MA (2005) Expression and characterization of the catalytic domains of soluble guanylate cyclase: interaction with the heme domain. Biochemistry 44: 4083-4090.

111. Haase T, Haase N, Kraehling JR, Behrends S (2010) Fluorescent fusion proteins of soluble guanylyl cyclase indicate proximity of the heme nitric oxide domain and catalytic domain. PLoS One 5: e11617.

112. Evgenov OV, Pacher P, Schmidt PM, Hasko G, Schmidt HH, et al. (2006) NO-independent stimulators and activators of soluble guanylate cyclase: discovery and therapeutic potential. Nat Rev Drug Discov 5: 755-768.

113. Chester M, Seedorf G, Tourneux P, Gien J, Tseng N, et al. (2011) Cinaciguat, a soluble guanylate cyclase activator, augments cGMP after oxidative stress and causes pulmonary vasodilation in neonatal pulmonary hypertension. Am J Physiol Lung Cell Mol Physiol 301: L755-764.

114. Evora PR, Evora PM, Celotto AC, Rodrigues AJ, Joviliano EE (2012) Cardiovascular therapeutics targets on the NO-sGC-cGMP signaling pathway: a critical overview. Curr Drug Targets 13: 1207-1214.

115. Ko FN, Wu CC, Kuo SC, Lee FY, Teng CM (1994) YC-1, a novel activator of platelet guanylate cyclase. Blood 84: 4226-4233.

116. Lee FY, Lien JC, Huang LJ, Huang TM, Tsai SC, et al. (2001) Synthesis of 1-benzyl-3-(5'-hydroxymethyl-2'-furyl)indazole analogues as novel antiplatelet agents. J Med Chem 44: 3746-3749.

117. Mulsch A, Bauersachs J, Schafer A, Stasch JP, Kast R, et al. (1997) Effect of YC-1, an NO-independent, superoxide-sensitive stimulator of soluble guanylyl cyclase, on smooth muscle responsiveness to nitrovasodilators. Br J Pharmacol 120: 681-689.

118. Roerecke M, Rehm J (2014) Cause-specific mortality risk in alcohol use disorder treatment patients: a systematic review and meta-analysis. Int J Epidemiol 43: 906-919.

119. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA (2014) Validity of Heart Failure Diagnoses in Administrative Databases: A Systematic Review and Meta-Analysis. PLoS ONE 9: e104519.

120. Boerrigter G, Costello-Boerrigter L, Lapp H, Stasch J-P, Burnett J (2005) Co-activation of soluble and particulate guanylate cyclase by BAY 58-2667 and BNP enhances cardiorenal function in experimental heart failure. BMC Pharmacology 5: P5.

121. Austin S (2006) Drug discovery technology Europe 2006 - IBC's Tenth Annual Conference and Exhibition. IDrugs 9: 256-260.

122. Frey R, Muck W, Unger S, Artmeier-Brandt U, Weimann G, et al. (2008) Pharmacokinetics, pharmacodynamics, tolerability, and safety of the soluble guanylate cyclase activator cinaciguat (BAY 58-2667) in healthy male volunteers. J Clin Pharmacol 48: 1400-1410.

123. Laskowski RA, Gerick F, Thornton JM (2009) The structural basis of allosteric regulation in proteins. FEBS Lett 583: 1692-1698.

124. Derbyshire ER, Marletta MA (2012) Structure and regulation of soluble guanylate cyclase. Annu Rev Biochem 81: 533-559.

125. Formaneck MS, Ma L, Cui Q (2006) Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). Proteins 63: 846-867.

126. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? Proteins 57: 433-443.

127. Ma B, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. Protein Eng 12: 713-720.

128. Popovych N, Sun S, Ebright RH, Kalodimos CG (2006) Dynamically driven protein allostery. Nat Struct Mol Biol 13: 831-838.

129. Tsai CJ, del Sol A, Nussinov R (2008) Allostery: absence of a change in shape does not imply that allostery is not at play. J Mol Biol 378: 1-11.

130. Tsai CJ, Del Sol A, Nussinov R (2009) Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. Mol Biosyst 5: 207-216.

131. Olea C, Boon EM, Pellicena P, Kuriyan J, Marletta MA (2008) Probing the function of heme distortion in the H-NOX family. ACS Chem Biol 3: 703-710.

132. Neudecker P, Lundstrom P, Kay LE (2009) Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. Biophys J 96: 2045-2054.

133. Boehr DD, Dyson HJ, Wright PE (2006) An NMR perspective on enzyme dynamics. Chem Rev 106: 3055-3079.

Chapter 2

Expression, purification and folding characterization of human H-NOX domain ligand binding

The author of this dissertation performed all the experiments described in this chapter with exception of the cloning of the H-NOX and H-NOX^{H105F} constructs which were conducted by Bayer Healthcare.

Table of Contents

Introduction

The detailed information of structure and functional role of the human H-NOX is limited due to the absence of a purification protocol to produce heme or drug bound forms in high quantities. This impedes further investigation related to enzyme activity, and therefore, the improvement of sGC specific drugs for cardiovascular diseases. Especially for NMR structural and dynamics studies of the human H-NOX containing heme or cinaciguat, the use of suitable isotopic labeling, such as ^{15}N and ^{13}C is required [1]. Eukaryotic as well as prokaryotic cells can be used as hosts to drive isotopically labeled protein production.

Expression of uniformly isotopically labeled protein for NMR studies in eukaryotic cells, such as insect-Sf9 cells, is still under development [2]. Insect-Sf9 cells contrary to *E.coli* cells fail to grow in minimal media, where ^{13}C glucose and ^{15}N ammonium chloride are the single sources of carbon and nitrogen respectively. Thus, many efforts have been applied to overcome this difficulty [1,2]. Currently, supplementing the fetal bovine serum media with isotopically labeled amino acids has been shown to be a viable strategy for over-expression of proteins [3,4]. Another issue is heterogeneous glycosylation of proteins expressed in eukaryotic cells, since highly homogenous material is needed to perform NMR studies. To solve it, cell lines should be deficient in complex heterogeneous glycosylation capabilities. In addition, there are other disadvantages associated with the transient transfection approach, where Insect-sf9 cells are infected with baculovirus expression vector and large quantities of DNA are needed to achieve high protein yield [2].

In summary, labelled ^{15}N/^{13}C human H-NOX sGC domain for NMR studies can be produced in eukaryotic cells, but all of these challenges need to be overcome. In

addition, the large scale production of target protein needs a special attention because the cost of goods is prohibitively high, and isotope incorporation efficiency and yields are often sub-optimal [1,4].

Unlabeled individual sGC domains, such as human α1/β1 sGC catalytic domain [5,6], *Rattus norvegicus* β1 CC sGC domain [7], *Manduca sexta* α1 sGC PAS domain [8] and prokaryotic H-NOX domain [9-11] have been obtained in *E. coli* and X-Ray crystallographic studies have been performed on them. Based on these studies, the truncated human β1 H-NOX-PAS domain was also successfully heterologously over-expressed and used to carry out inter-domain interaction studies and NO-sGC conformational changes analysis [12-14]. *E.coli* expressing system has also proved to be highly efficient in producing high protein yield of full length β1 human sGC domain [15].

For NMR structural and dynamics studies, expression of isotopically labeled human H-NOX sGC domain in *E. coli* host system might be a preferred choice for a number of reasons. This system has proved to efficiently over-express the H-NOX domain sGC homologues and to easily incorporate heme or cinaciguat into the heme pocket [9-11]. Also, bacteria grow easily in isotopically labeled minimal media, with a fast growth rates combined with the ability to achieve high protein yields in many cases. It is also less expensive than insect-Sf9 expression systems. Moreover, isotopically labeled protein obtained from *E. coli* expression system can be easily purified based on simple purification protocols [1], contrary to what happens with protein expressed in insect-Spf9 system.

While *E. coli* as an expression system has been proven to be highly efficient in over-expressing recombinant protein, absence of full heme incorporation may occur. However, using *E. coli* to express human H-NOX domain is an advantage since it allows testing different protocols to insert heme into the protein. For instance, full incorporation of heme in recombinant expressed

protein has been achieved in *E.coli* co-expressing the native protein ferrochelatase (FC) [16]. But, co-expression of human H-NOX and FC has not been reported so far.

Other strategies to achieve high protein yield can be planned based not only in expression and purification optimization but also cloning engineering. For instance, expression of the WT human H-NOX domain fused with a maltose-binding protein (MBP) in *E. coli* has been proved to increase its solubility. In these systems, the protein of interest is often expressed as a MBP-fusion protein, preventing aggregation of the protein of interest. The α2 H-NOB/β1 H-NOX and HNOX-PAS truncated domain and full length sGC protein fused with N-terminal MBP tag has been reported to contribute to a more efficient protein expression [17].

The identification of minor fluctuations of a protein structure in solution can be monitored by NMR spectroscopy, allowing the study of protein folding of pure apo/holo human H-NOX domain [18,19]. For instance, a single 1D ^{1}H spectrum can provide information about protein folding and stability [18]. ^{15}N-HSQC spectrum, often referred to as the fingerprint of a protein, allows the access to the above mentioned information but also enables the identification of multiple conformations of protein in solution and the ability to perform stability assays.

The tertiary structure of a protein can be studied with or without ligand incorporation. It can either show to be folded, unfolded, or a mixture of both. If no ligand has been incorporated, resulting in the protein being in the apo-form, they often behave as molten globules. This conformation has been identified for numerous proteins with ligands/cofactors absent [20]. Therefore, it is possible that the apo-H-NOX domain might be over-expressed in this state.

In the present work, we intended to evaluate the structure and dynamics of H-NOX sGC in the active and inactive states. For that, H-NOX was complexed

with multiple ligands that aim to mimic the individual states of activation of the catalytic domain of sGC. Cinaciguat was incorporated into the heme pocket of H-NOX to generate the active form of this domain. We also attempted to produce the native resting/active form of this domain by incorporating heme and generating H-NOX-heme complex. However, this complex was paramagnetic due to the high spin five-coordinate ferrous atom [21], therefore it would severely affect all NMR spectra. So, we intended to convert high spin five-coordinate ferrous into low spin six-coordinate ferrous form through heme-CN binding (resting state), heme-CO binding and high spin five-coordinate ferrous-NO binding (active state) [22].

To generate the inactive state, H-NOX sGC domain was complexed with Zn^{2+} protoporphyrin IX (ZnPPIX) heme analogue, which has high affinity for sGC (16 nM K_d) and it was characterized as a potent sGC inhibitor [23]. Furthermore, it has diamagnetic properties allowing structural and dynamics NMR studies.

In this work, we also over-expressed and optimized the protein purification protocol of the WT and H-NOX[H105F] mutant of H-NOX in *E.coli*. In spite of having the majority of the apo-protein in aggregated form, we were able, using NMR to assess that it had similar structural features with its monomeric counterpart. In addition, we designed an efficient protocol for protein-ligand incorporation capable of promoting full protein dissociation from oligomeric to monomeric form of human H-NOX_c (active state) or H-NOX_Zn (inactive state). We also analyzed the monomeric complex by NMR in order to perform further structural and dynamics studies.

Material and methods

The pETM-44 vector and the HRV-3c protease were obtained from the European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. The pCACy-Duet-1 plasmid with Ferrochelatase gene was provided by

Dr Brian R. Crane (Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, USA). *E. coli* DH5α and BL21 (DE3) strains, DNA polymerase, T4 DNA ligase, dNTP, plasmid purification kits, gel extraction kits were purchased, from NZYtech genes & enzymes. The restriction enzymes (BamHI, NcoI, NotI) were purchased from New England Biolabs. Oligo-nucleotide PCR primer pairs, labeled as NOXFw, HNOXRv, MBP_HNOXFw and HisTag_HNOXRv (Table 1), were synthesized by Metabion, and DNA sequencing reactions were performed by Stab Vida.

HisTrap FF Crude, 5 ml, ultracel 5 KDa ultrafiltration Discs (Millipore Corporation) Diaflo ultra filter (Amicon), Disposable PD-10 Desalting, Superdex 75 10/300 GL, HiLoad™ 16/60 Superdex 200 columns were purchased from GE healthcare Life Science. BAY 58-2667 compound from obtained from Dr J. P. Stasch, Bayer Schering Pharma AG. Zi^{2+} protoporphyrinIX and Hemin chloride from Sigma Aldrich.

Cloning

Constructs C#2, and C#4 were produced using as DNA template the WT C#7 that contained the β1 human H-NOX sGC domain (1-189). Construct C#3 was amplified using C#2 as template since it contained the N-terminal His_MBP_HRV3c_H-NOX (1-605). C#2 PCR products were obtained using the primer pair NOXFw and HNOXRv, C#3 MBP_HNOXFw and HNOXRv, C#4 NOXFw and HisTag_HNOXRv (Table 1). The forward primers were designed to introduce NcoI (C#2 and C#4) and BamHI (C#4) restriction sites. The reverse oligo-nucleotide was used to introduce the NotI restriction sites for all constructs. For the preparation of C#2, a primer introducing the H-NOX gene preceded by an HRV-3c protease recognition site was used. The gene including an N-terminal hexa-His peptide sequence, MBP and an HRV-3c protease recognition site followed by H-NOX was prepared using C#2 as a template resulting in the C#3 construct. The C#4

construct was designed to have a C-terminal hexa-His peptide sequence (Table 2). The resulting PCR products were purified and subsequently digested. The digested PCR products were ligated with T4 DNA ligase to pETM-44 (C#2) and pCACy-Duet-1-FC (C#3 and C#4) plasmid, that had been digested with NCO I and NOT I (C#2 pETM-44), Bam HI and NOT I (C#3 pCACy-Duet-1-FC) and NCO I and NOT I (C#4 pCACy-Duet-1-FC). The resulting ligation mixture was transformed into DH5α competent cells by heat shock. The constructed plasmids were verified by DNA sequencing.

Table 1: Primer sequence used in this study

Construct	Primer	Sequence (5´- 3´)
WT C#2 and WT C#4	H-NOXFw	TATACCATGGGAATGTACGGCTTTGTGAACCA
WT C#2 and WT C#3	HNOXRv	TATAGCGGCCGCTTATTCTTTGCTTTCTTTTTCTTC
WT C#3	MBP_HNOXFw	TATAGGATCCTATGAAAATCGAAGAAGGTAA
WT C#4	HisTag_HNOXRv	TATAGCGGCCGCTTA GTG ATG GTG ATG ATG ATG

Table 2: Constructs used in this study. All Constructs have *T7/Lac* promoter.

Construct	Vector	Gene	Size
WT C#2	pETM-44	His_MBP_H-NOX	1818bp
WT C#3	pCACy-Duet-1	His_MBP_H-NOX	1818bp
		Ferrochelatase	
WT C#4	pCACy-Duet-1	HNOX_ His	615bp
		Ferrochelatase	
WT C#7	PSD-ECO	HNOX_ HistineTag	615bp
C#8	PSD-ECO	HNOX^{H105F}_ HistidineTag	615bp

Expression of the human H-NOX domain

The WT (C#7) and H-NOX^{H105F} (C#8) were codon optimized allowing for the human protein to be expressed in regular *E.coli* systems. The following procedure was followed: Plasmids were transformed into *E.coli* BL21(DE3) competent cells by heat shock. A single colony from Luria–Bertani (LB)-agar plate containing 100mg/mL ampicillin was selected and then applied to inoculate 100ml of LB media containing 100mg/ml ampicillin which was grown overnight. The cell density of the culture was evaluated using the optical dispersion at 600nm (OD$_{600nm}$) as a function of time in hours. The growth of the cells containing C#7

and C#8 in LB medium with 100mg/ml of ampicillin at 37°C was monitored by the OD_{600nm} 0.6-0.8 measurements.

For the optimization of the expression conditions, tests were made using minimal media cultures induced with IPTG in the middle of the log phase (OD_{600nm} 0.6-0.8) (Table 2 SI). These were grown overnight in a range of different temperatures (17, 20 and 37˚C), aeration rates; (120, 140 and 160rpm) and media supplementation, (unlabelled, supplemented with tryptone 0.1-1%) after induction with 1mM IPTG,. The over-expression time and the IPTG concentration were based on recent literature reference of H-NOX domain expression [15,17,24]. All cell cultures were lysed using Bug Buster (Novagen), and OD normalized (OD of 1.4) amounts of the lysates were loaded on a (12.5%) SDS-PAGE for evaluation.

Scale up of the protein expression from 1L to 15L (H-NOX) and 10L (H-NOX[H105F]) were performed in minimal media (S1). Induction was initiated in the middle of the log phase 0.6-0.8 OD_{600nm} using 1mM IPTG at 20˚C, 160 rpm, overnight and cells were harvested by centrifugation. The cell pellets were frozen and stored at -80°C.

Co- expression of WT H-NOX domain and Ferrochelatase

The cloning results of C#2, C#3 and C#4 allowed us to over-express human H-NOX with a cleavable N-terminal hexa-His_MBP_HRV3c protease site (C#2 and C#3) and C-terminal hexa-His (C#4) and Ferrochelatase without any tag (C#3 and C#4). Human H-NOX domain was expressed as reported before in the WT and H-NOX[H105F] section. For the Ferrochelatase expression, the media cultures C#3 and C#4 were supplemented with 25 mg/L of δ-Aminolevulinic acid (δ-ALA) at the time of induction, and growth media was supplemented with 100 μM $FeCl_3$. The antibiotics, kanamycin (50 mg/mL) for C#2, and chloramphenicol (34 mg/mL) were

added to the growth media of C#3 and C#4. The cell pellets from 15L (WT) and 10L (H-NOX^{H105F}) were divide into 7 and 5 pieces respectively and were frozen and stored at -80° C.

Purification of human H-NOX domain and its incorporation with cinaciguat compound

^{15}N-labeled WT and H-NOX^{H105F} domains (pI 5.6) were purified as follows: Each frozen cell pellet was thawed on ice, resuspended in 5 ml/g Histrap buffer A (50 mM Tris, 500 mM NaCl, 10 mM β – mercaptoethanol (β-Me), 10% glycerol, pH 8.0) with 1 tablet of Protease Inhibitor Cocktail without EDTA (Roche life science) 1 mM phenyl methyl sulfonyl fluoride (PMSF). Cell disruption was achieved by passing the solution three times in a large French Press at 1000 psi pressure at room temperature. The cell lysate was subjected to ultracentrifugation using a 70 ti rotor at 302000g for 1 hour at 4°C on a Beckman Optima 80k ultracentrifuge. The supernatant from this centrifugation was applied to nickel affinity chromatography (HisTrap FF Crude, 5 ml column) that had been equilibrated with buffer A. The column was washed with 10 column volume of buffer A and then the protein was eluted with two column volumes of 60 % of Histrap buffer B (50 mM Tris, 500 mM NaCl, 10 mM β – mercaptoethanol (β-Me), 10% glycerol, pH 8.0 containing 500 mM imidazole). The fractions containing the human H-NOX domain were always verified by SDS-PAGE (12.5%). The imidazole in the elute was removed by PD-10 Desalting with 20mM $(PO_4)^{3-}$, 50mM NaCl, 1mM DTT, at room temperature. After desalting, the sample was concentrated using Diaflo ultra filter (3kDa). The protein at different concentrations was used to optimize the protein-cinaciguat incorporation protocol (Table 3).

Table 3: Human H-NOX domain cinaciguat incubation protocol

Incubation protocol	
Protein concentration	44.6, 67, 90 and 134 µM
Protein-compound ratios	$^1/_2$ equivalent of cinaciguat each 2h until 1:1, 1:2, 1:3 and 1:5
Temperature	20 and 37˚C
Reagents added	Protease inhibitor
	0.05% n-Octylgucoside
	5mM DTT
Incubation time	18h

In order to isolate the monomeric form of human H-NOX complexed with cinaciguat, the samples were injected into size exclusion column (Hi Load 16/60 Superdex 200 prep grade or 10/300 Superdex 75). It was equilibrated with SE buffer A (20 mM NaPi pH 6.5, 50 mM NaCl, 10µM cinaciguat, 1mM sodium azide plus 1 mM DTT). The fractions containing the monomeric form of the construct were concentrated using a 3kDa MWCO Vivaspin 15R device to a final volume of 500µL. The concentrations of monomeric ^{15}N-labeled WT (0.4mM) and H-NOX^{H105F} (1mM) of human H-NOX cinaciguat complex were determined by absorbance at 280nm in a nanodrop UV-vis spectrophotometer. Absorbance was measured after background correction with the corresponding buffer solution and the concentration was determined using the Extinction coefficient: $17545M^{-1}$ cm^{-1} [25]. The H-NOX^{H105F}_c sample was diluted to 0.33mM to be used for NMR dynamics data collection.

Purification of co-expression of WT human H-NOX domain with ferrochelatase

The protein samples expressed from constructs C#2, C#3 and C#4 were purified following the procedure reported before in the WT and H-NOX[H105F] section. To perform the purification of the H-NOX co-expressed with FC (C#3 and C#4), the absorption at 280, 410 and 425 nm (human β1 H-NOX, ferric and ferrous Soret bands, respectively [17]) was monitored during nickel affinity and size exclusion chromatography. The construct containing H-NOX domain with N-terminal MBP tag (C#2 and C#3) was cleaved by HRV3C protease following the Thermo scientific protocol and the efficiency of the process was monitored by SDS-PAGE (12.5%).

WT Human H-NOX Heme reconstitution

The apo-H-NOX domain from construct C#7 was first expressed and purified by nickel affinity and then desalted using a PD-10 column equilibrated with buffer (500mM NaCl, 50mM Tris pH 8.0, 10mM β-Me, 10% glycerol). 134 μM of the protein was used for reconstitution. Heme was added in portions of ½ equivalents of reduced hemin at room temperature until the ratio of 1:5 protein-ligand followed by overnight incubation [15,17,24] in an anaerobic chamber. The H-NOX domain-heme mixture was concentrated using Diaflo ultra filter (3kDa) and then applied to the PD-10 desalting column to remove the unbound heme. The sample was loaded into size exclusion column (equilibrated with 20mM NaPi, 50mM NaCl, 5mM DTT) to isolate the monomeric protien. The UV-Vis absorance at 280, 410 and 425nm (human β1H-NOX, ferric and ferrous Soret, respectively [17]) were recorded.

WT Human H-NOX domain Zinc²⁺ protoporphyrin IX incorporation

WT H-NOX domain from C#7 was expressed and purified by nickel affinity and PD-10 based on protocol above described. 134 µM apo-protein in 20mM NaPi pH 6.5, 50mM NaCl, 1mM DTT plus 134 µM ZnPPIX were incubated at 20˚C overnight. In order to isolate the monomeric form of H-NOX_Zn complex, size exclusion chromatography was performed (Hi Load 16/60 Superdex 75 prep grade). The column had previously been equilibrated with buffer (20 mM NaPi pH 6.5, 50 mM NaCl, 10µM ZnPPIX and 1mM DTT). After collection of the aqppropreate fractions, the sample was concentrated using a 3kDa MWCO Vivaspin 15R to produce a suitable NMR sample. The concentration of monomeric ^{15}N-labeled WT H-NOX_Zn complex was determined by absorbance at 280nm (Extinction coefficient: $17545M^{-1}$ cm^{-1}).

CD spectroscopy measurements

CD spectra (Jasco J-815 CD spectrometer) were collected in a 20 mM NaPi buffer, pH 6.5 at 25˚C and 0.1mg/mL of the mixture of the aggregated, oligomeric and monomeric forms of the apo-H-NOX, and monomeric WT H-NOX_ c complex using a 1mm strain free cuvette in the range of 190-240nm.

NMR data collection

All the NMR data were recorded on a 800 MHz Bruker Avance III NMR spectrometer at 298K using a TXI Z axis gradient room temperature probe in 5 mm NMR tubes (NEW ERA UL-5). The NMR samples for H-NOX_c and H-NOX^{H105F}_c and aggregated H-NOX (mixed with 100 µM of cinaciguat) were prepared in a buffer containing 20 mM NaCl pH 6.5, 50 mM NaCl, 1mM DTT 10 µM cinaciguat plus 1mM sodium azide. For H-NOX_Zn, a buffer containing 20 mM NaCl pH 6.5, 50 mM NaCl, 1mM DTT plus 10 µM ZnPPIX was used. For each tube,

500 µL sample and 50uM of D_2O were used. [1]H 1D spectra were collected from these samples. For the Monomeric H-NOX_c, H-NOX[H105F]_c and H-NOX_Zn complexes ^{15}N-HSQC spectra were also collected.

Results

Expression and protein purification

E. coli has been shown to be an effective method for recombinant production of domains of sGC for X-Ray crystallography, and inter-domain interaction studies [5-8,26,9-14]. However, expression of isotopicaly labeled protein in high quantities for NMR structural and dynamic characterization has been a challenge. Therefore, part of our studies was directed at optimizing the expression and purification of human H-NOX domain in minimal media.

Overnight expression of H-NOX domain at 20˚C under 160 rpm showed to be the best condition for expressing WT and H-NOX[H105F] in minimal media (Figure 1A). In previous protein purification protocols, affinity, anionic exchange and size exclusion chromatography were used to obtain the pure protein. To minimize protein loss during the purification process anionic exchange chromatography step was eliminated, since affinity chromatography was shown to be enough to remove the majority of the sample impurities (Figure 1 B).

The monomeric form of apo-HNOX domain was expected to be present. However, size exclusion chromatography using a size calibrated Superdex 200 showed the presence of at least three peask for the apo-protein suggesting different extends of aggregation (**Figure** 1C). The first peak corresponds to a high order aggregated form, followed by an intermediate aggregate and a small amount of monomeric H-NOX domain. Therefore, the majority of human apo-H-NOX domain expressed in bacterial *E.coli* was shown to be a mixture between

these three forms. The concentration of all the forms of the apo protein was 3mg/L of culture.
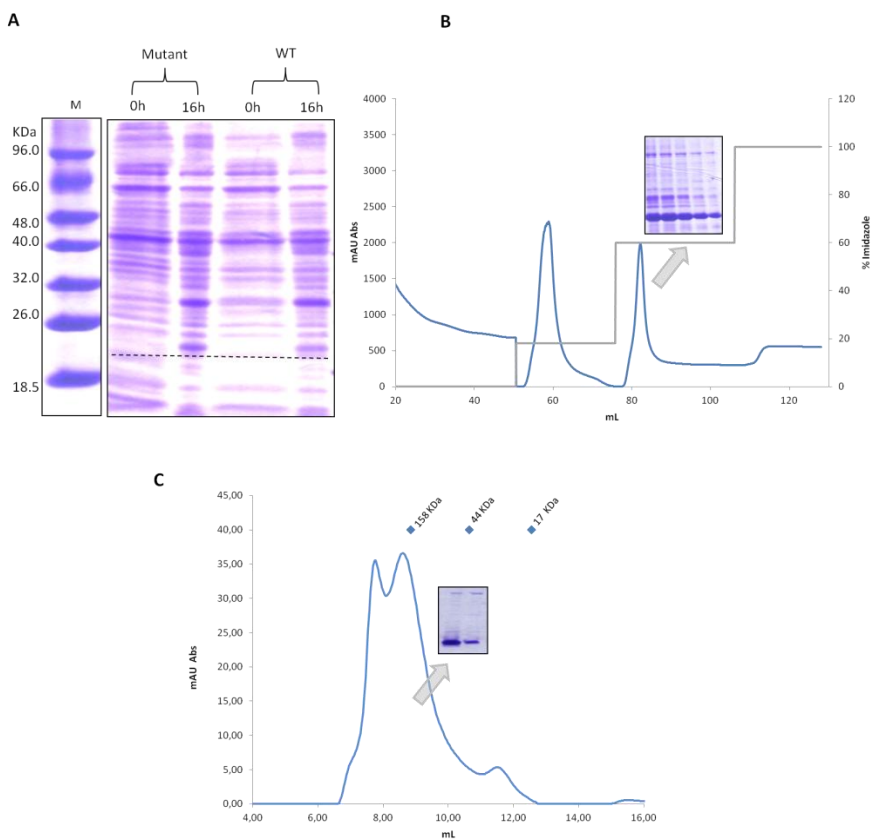
Figure 1: Protein expression and purification profile. A: SDS PAGE (Tris-Glycine 12.5%). OD normalized sample from overnight optimized expression test, 1mM IPTG at 20˚C were used for induction. M- Protein marker, H-NOX[H105F] domain, WT H-NOX domain. Black line: human H-NOX domain (22.3 KDa).   **B** and **C**: Purification profiles of WT and H-NOX[H105F] domain. After purification process the sample were loaded to SDS PAGE (Tris-Glycine 12.5%).   **B:** chromatogram of H-NOX domain purified in Nickel column (HisTrap SP FF). The gradient step of Imidazole buffer from 0, 18, 60 to 100% was used for protein elution. **C:** Size exclusion chromatography (buffer: 20 mM NaCl pH 6.5, 50 mM NaCl, 1mM DTT) of protein sample collected from affinity chromatography (60% imidazole). The molecular weight values are standard calibration curved values of Superdex 75 10/300 GL.

Probing conformational stability of the human apo-HNOX domain

NMR and Circular Dichroism (CD) spectroscopy are powerful techniques for studying protein folding in solution. They are particularly useful when applied to dynamic systems such as partially folded molten-globule states where proteins lack a rigid tertiary packing but maintain a compact shape. These flexible systems are usually not amenable to be studied by X-ray crystallography [18,27].

The mixture of the protein (aggregated and oligomeric form) of the apo-H-NOX domain were analysed by CD, where we observed a signal characteristic of an α-helical secondary structure with absolute peak intensities at 208 and 222 nm elements. When comparing the CD signals of the apo-HNOX mixture with the one of the monomeric form of holo-H-NOX complexed with cinaciguat, they were found to be almost identical (Figure 2A). From this observation we can conclude that the protein aggregation was not due to protein unfolding. A 1D proton NMR spectrum of the apo mixture (aggregated and oligomeric form) was also collected. In NMR, a small protein exhibits well defined signals that cover a large range of chemical shifts, whereas, if the protein is unfolded or partially folded, the peaks are sharper and not as widely dispersed. In the case of the apo protein, NMR results were variable. In some cases(Figure 2B blue) the sample exhibited somewhat broad peaks but well dispersed, ranging from a little over 9 ppm to -0.5 ppm. The features below 0ppm are most likely originating from methyl resonances that correspond to well defined structures. This becomes even more evident when a comparison is made with the case of the folded monomeric H-NOX_c complex. There we could observe well defined proton peaks below 0 ppm (Figure 2C red). In some other protein preparations however, the mixture of the three protein forms exhibited a different behavior. Some protein samples showed different 1D profile, with the signals being less defined, with broader peaks and with absence of proton peaks below 0 ppm. (Figure 2C green) Because of the

absence of sharp signals this protein form is not likely to be unfolded but rather part of a bigger aggregated protein probably in a molten globule state.

In an effort to address the issue of a molten globule, the mixture of the three forms of the protein was subjected to a urea titration. While we were not able to observe a sharp denaturation transition that would be indicative of a well folded protein, due to the very low number of data points the result was inconclusive (data not shown). The combination of the above results suggests that the expression system is not very robust and under, as of yet, unidentified conditions it can produce samples of different quality. Further tests are being developed to elucidate the folding variability in the protein mixture. Meanwhile, we moved on to study the monomeric form of the protein complexed with different ligands.
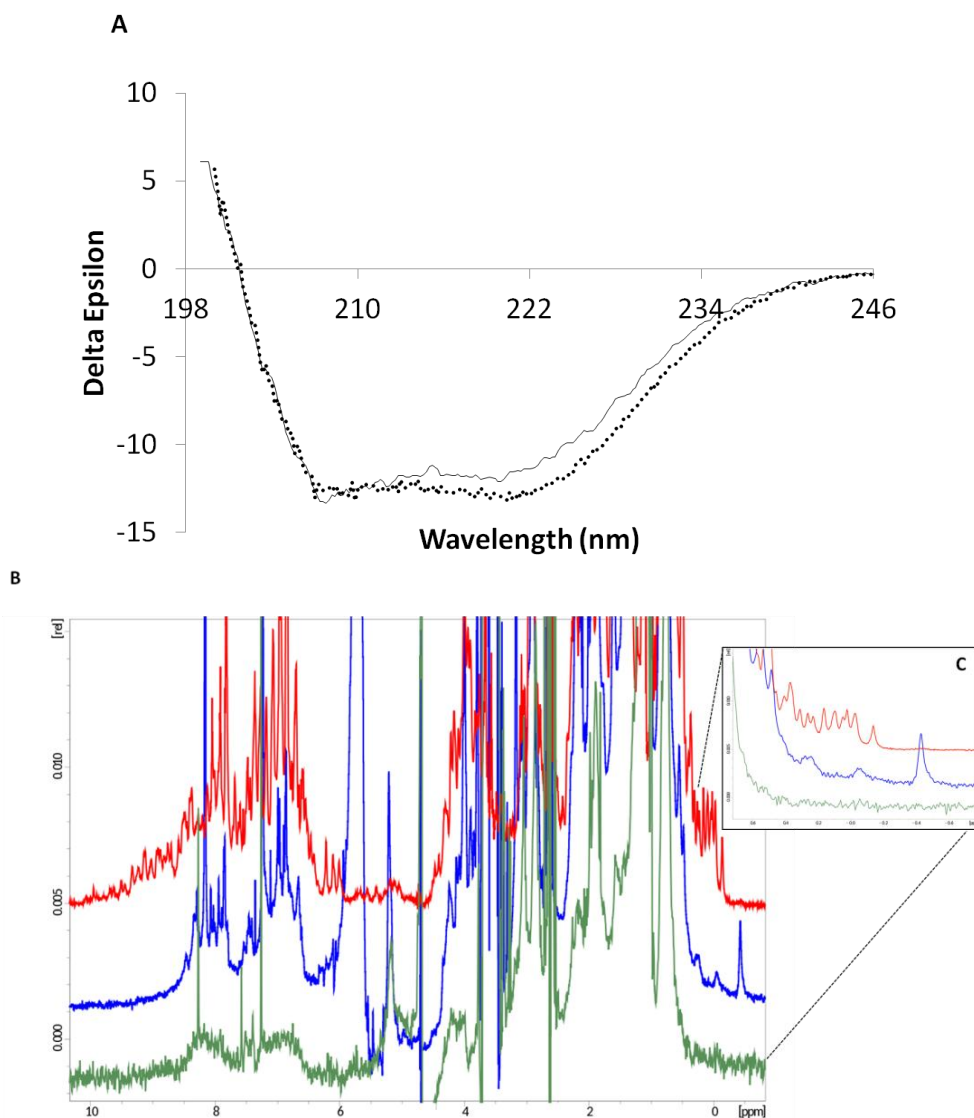
Figure 2: A- H-NOX folding caracterization. Dotted line: aggregated form of the WT apo-protein. Grey line: WT H-NOX_c. **B-** 1H NMR spectrum collected at 25˚C. Green line: aggregated form of the unfolded WT apo- protein. Blue line: aggregated form of the folded W.T apo- protein. Red line: Folded monomeric form of the WT H-NOX_c. **C-** Expanded region of the methyl protons from panel B-.

Human H-NOX domain cinaciguat compound incorporation

In vitro, heme free sGC has been shown to be able to immediately incorporate cinaciguat compound upon its addition, allowing full activation of the cGMP production with an EC_{50} and K_d values in the low nanomolar range. [28,29]. However, the amount of time that cinaciguat takes to incorporate into individual H-NOX sGC domain has not yet reported. Thus, for our H-NOX domain structural studies, a protocol for protein-compound incorporation was tested with different incubation times, to obtain high quantities of monomeric protein-cinaciguat complex.

Using NMR, the solubility limit of cinaciguat in the buffer system used for sample preparation was tested by observing the signal intensity as a function of amount of ligand added. We were able to determine a solubility limit of approximately 55 µM in buffer solution (20 mM NaCl pH 6.5, 50 mM NaCl, 1mM DTT plus 1mM sodium azide) at 25˚C. Therefore, to avoid precipitation in solution, 55 µM was the maximum concentration used to keep the free compound soluble. Various conditions were tested such as, the protein concentration (44.6 µM and 67µM), protein-compound ratio (protein compound 1:1 and 1:2, respectively) and incubation temperature (20˚C and 37˚C). Adding half equivalent of cinaciguat compound in each 2 hours into 67 µM of aggregated form of the protein (protein-cinaciguat ratio 1:2), at 20˚C under 18h. The progress of the incubation processes was monitored by size exclusion chromatography. Each 2h hours, a small fraction of the sample was used in size exclusion chromatography. The peak intensity from aggregated, oligomeric and monomeric fractions were evaluated using the peak volume assuming each form had the same extinction coefficient. The results can be seen in Figure 3A. From this graph it is evident that the protein-cinaciguat complex in monomeric form represents around 30% of the total protein. This suggests that the rest of the protein remained as a mixture between aggregated

forms. (Figure 3 B-C). Therefore, to achieve the maximum monomeric form of protein-compound, the protocol needed to be additionally modified.

In order to further optimize the incubation protocol, higher protein concentrations such as: 90 and 134 µM were used. For each concentration the protein-compound ratio used was 1:2 and 1:3. Based on the same procedure as above, the compound was added in steps into incubation solution. After 18h of incubation, size exclusion chromatography showed that the low molecular weight aggregated form of the protein had dissociated into the monomer protein-cinaciguat complex (Figure 3 D). The complex was shown to be stable between 8 and 18h (Figure 3 E).

To test if the high MW aggregated protein could be further dissociated, the fractions from the size exclusion chromatography were again incubated with cinaciguat. The results showed a small dissociation (23%) from aggregated protein into monomer H-NOX_c (data not shown). A scale up of the WT H-NOX expression from 1 to 15L, and mutant H-NOX^{H105F} (from 1 to 10L) was able to generate approximately 0.5 mg/L and 2.5mg/L of labeled protein-cinaciguat respectively for NMR structural and dynamics data collection.
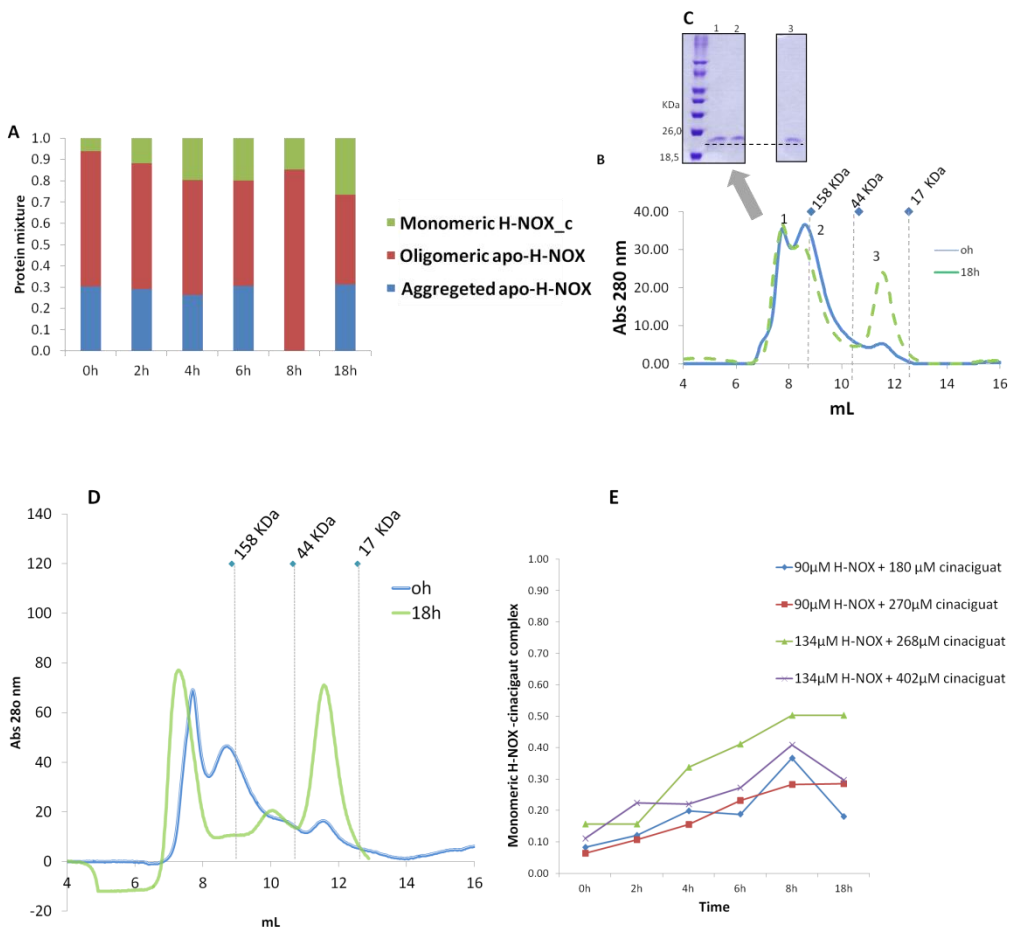
Figure 3: Proportions of aggregated, oligomeric and monomeric forms of the protein mixture during the incubation with cinaciguat. A: Measurements are based on samples collected each 2h hours during the protein-compound incubation. The rations of the different fractions come from size exclusion chromatography (Superdex 75 10/300 GL). **B:** Size Exclusion chromatogram of incubation of 67μM of the protein and between protein-compound ratio of 1:2. Green line: sample loaded after 18h of incubation. Blue line: apo-aggregated protein (sample control). **C:** SDS PAGE (Tris-Glycine 12.5%) of the sample eluted from size exclusion chromatography. Note that the protein from 1, 2 and 3 peak presented the same molecular weight **D:** Size Exclusion chromatogram of incubation of 134μM of the protein and between protein-compound ratio of 1:2 . Green line: sample load after 18h of protein-cinaciguat incubation. Blue line: 0h from apo-aggregated protein (sample control. Note that in this condition no oligomeric form of the protein was observed upon 18h of the incubation with cinaciguat compound. The standard calibration curved values was plotted to indicate the molecular weight of the each eluted protein. **E:** Monomeric H-NOX_c complex collected each 2h during protein compound incubation

indicating the stability of the monomeric protein upon 8h of the protein-compound incubation.

WT human H-NOX domain heme incorporation

Resting state form of H-NOX domain can be represented by the H-NOX-heme complex. However, to incorporate heme into recombinant protein appeared to be a challenge. Here, we are presenting results from different attempts to obtain the HNOX- heme complex.

The first was co-expressing human H-NOX with native protein ferrochelatase (FC) in *E. coli*. It was performed by adding δ-ALA at time of induction into growth media supplemented with $FeCl_3$. Although, human H-NOX domain showed to be well expressed (Figure 4A), the nickel affinity and size exclusion chromatography, monitored by UV at 280, 410 and 425nm (human β1H-NOX, ferric and ferrous Soret, respectively [17]) showed the majority of the H-NOX domain was expressed without heme (Figure 4 B-C)

In the absence of in cell, full heme incorporation into H-NOX, the second strategy was planned based on cloning, expression, and purification of WT human H-NOX with maltose-binding protein (MBP) fusion proteins in *E. coli*.
The gene encoding H-NOX domain was sub cloned successfully into the pETM-44 (C#2) and pCACy-Duet-1 FC (C#3) vector. The C#3 was built to co-express the N-terminal His6x-MBP-H-NOX protein and FC. One of the major advantages of the MBP tag is that it is able to avoid protein aggregation in addition to providing increased yields. Co-expressing H-NOX domain and FC could optimize the heme incorporation in *E. coli*.

Indeed, both constructs C#2 and C#3, over-express the protein with a N-terminal MBP tag (S2A). However, after HRV3C cleavage, only a small fraction of H-NOX domain could be cleaved (S2 B). Moreover, construct C#3 showed no heme incorporation during size exclusion chromatography (data not shown).
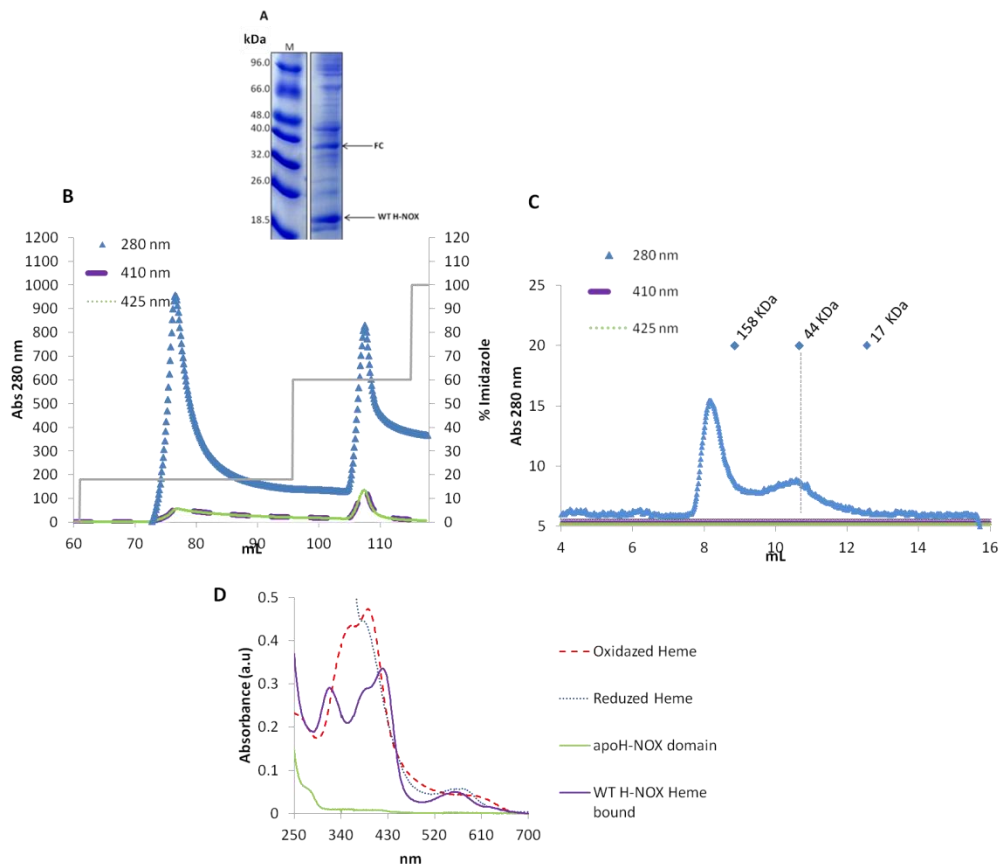
Figure 4: Protein expression and purification profile. **A**: SDS PAGE (Tris-Glycine 12.5%). Normalized sample from overnight optimized expression test, 1mM IPTG at 20°C were used for induction. M- Protein marker. Arrow indicates: human H-NOX domain (22.3 KDa) and ferrochelatase enzyme (35 KDa). **B:** chromatogram of H-NOX domain purified in Nickel column (HisTrap SP FF). The gradient step of Histrap buffer B from 0, 18, 60 to 100% was used for protein elution. **C:** Size exclusion chromatography chart of protein sample collected from affinity chromatography (60% imidazole). The molecular weight values are standard calibration curved values of Superdex 75 10/300 GL. Wavelength at 410 and 425 nm (human β1H-NOX, ferric and ferrous Soret ε, respectively) for both purification process were monitored [17]). **D:** U.V-vis spectrum. The shift change from λ_{Soret} 410 to 425nm indicates heme binding into human WT H-NOX domain [17].

The H-NOX-heme complex can be isolated with a sub-stoichiometric amount of heme. Therefore, the third strategy was to over-express human H-NOX with C-terminal His6x-tag (construct C#7) and incubate the protein with heme following the Derbyshire and collaborators post purification incubation method [24]. Inside an anaerobic chamber, a UV-vis spectra were collected that indicated the heme was binding to the H-NOX domain as we saw a shift in the frequency of the Soret band from λ_{Soret} 410 to 425 nm (human β1 H-NOX, ferric and ferrous Soret ε, respectively) [17] upon protein binding. However, during size exclusion chromatography with UV detection at 410nm, indicated that the protein was being eluted in the dimeric form and no protein-heme complex bands could be detected. (Figure 4 D). This result suggests that either the heme is interacting wtih the surface of the protein and it does not enter into the heme cavity or that it is very labile and it is removed by size exclusion chromatography.

WT human H-NOX domain Zinc^{2+} protoporphyrin IX incorporation

sGC can be converted to an inactive form, when heme is replaced by ZnPPIX. This heme analogue has been described to be a potent sGC inhibitor [23,30]. It is believed to also be bound to H105, since this axial ligand is crucial for heme retention. Therefore, human H-NOX complexed with ZnPPIX can represent a good model for the resting state of the protein. It is also an inactive form because it is not expected to bind NO and thus activate sGC

The protocol for over-expression and purification the protein were previously described on the result section 1 and 3. In order to perform the protein ZnPPXI incorporation, 134 µM of aggregated form apo-HNOX was incubated with the same amount of ZnPPXI (ratio 1:1) at 20˚C. Following the protein purification procedure, HNOX_Zn complex was eluted as a monomeric protein (Figure 5A). A

UV-vis spectrum was collected indicating that ZnPPXI was incorporated into H-NOX domain as we saw a shift change of the free ZnPPXI upon ZnPPXI-protein bond formation (from λ_{Soret} 410 to 425nm) (Figure 5B). Scale the protein expression up from 1 to 15L, approximately 0.5 mg/L of labeled monomeric complex were obtained in the end.

The NMR ^{1}H spectra of H-NOX_c, H-NOX[H105F]_c and H-NOX_Zn were collected. The result showed that all protein complexes were in their monomeric forms and were well folded because we could observe proton signals Hα (5 ppm), the beta sheet conformation (around 10-8 ppm); the amide signals were well resolved spanning more than 2ppm and also the signals of methyl resonances were close or below 0ppm (Figure 5C). A ^{1}H-^{15}N HSQC spectrum was also collected for the H-NOX_c, H-NOX[H105F]_c and H-NOX_Zn. In these spectra one peak is expected for each backbone and side chain NH group of the protein (S3). The dispersion pattern, intensity, and number of observed cross peaks indicates directly the protein folding state and overall sample quality. As the number of peaks in the ^{1}H-^{15}N HSQC spectrum estimate approximately to the number of residues of the protein, conformational heterogeneity can be identified by an excess of peaks. So, the protein assignment, structural and dynamics data analysis were performed for these three protein samples
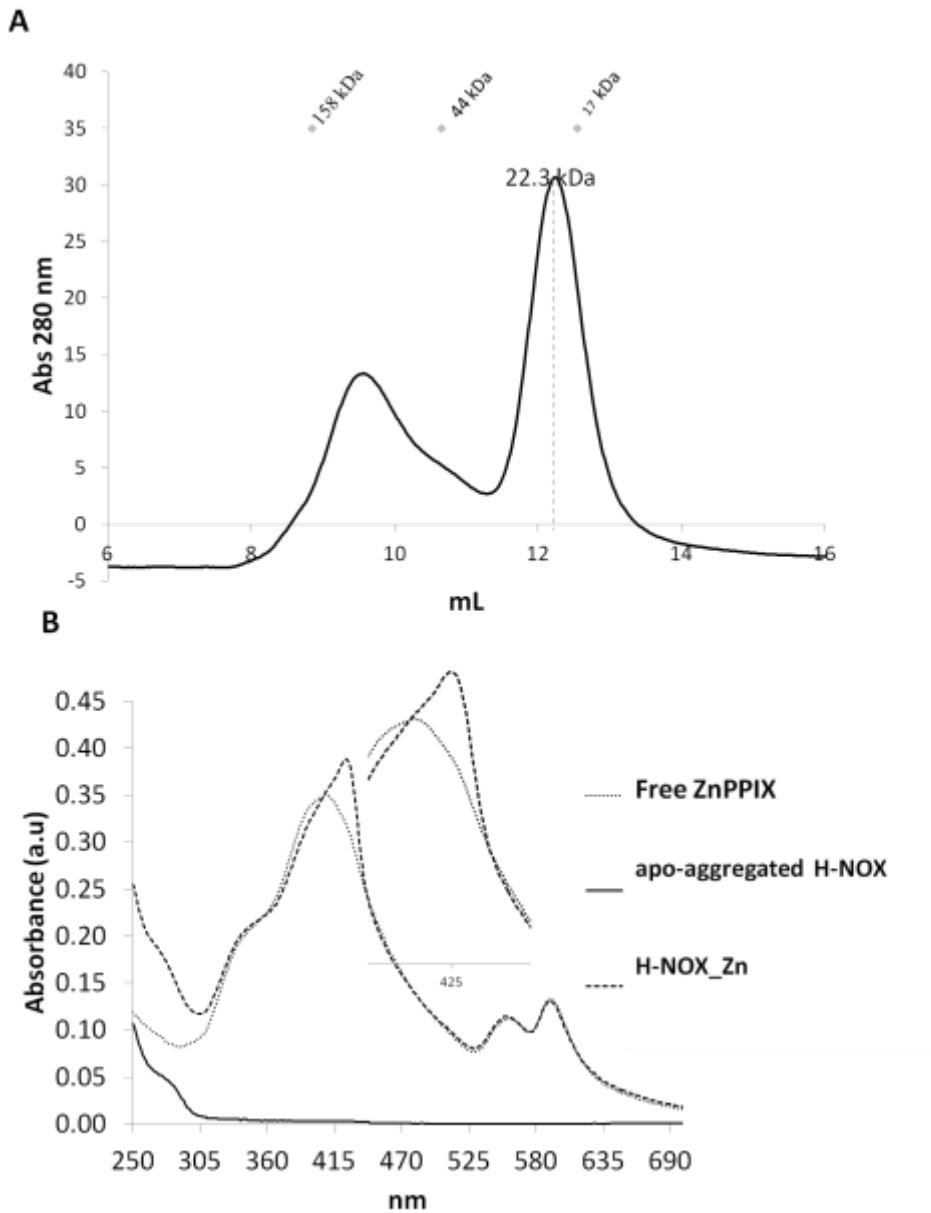
Figure 5: H-NOX ZnPPIX incorporation and folding by NMR profile. A- Size Exclusion chromatography (Superdex 75 10/300 GL) chart of protein sample load after 18h of H-NOX_Zn incubation assay. The molecular weight values are standard calibration curved values from Superdex 75 10/300 GL. **B-** U.V-vis spectrum of monomeric H-NOX_Zn. The shift change from 410 to 425nm indicates that the ZnPPIX was incorporated into H-NOX domain [17].
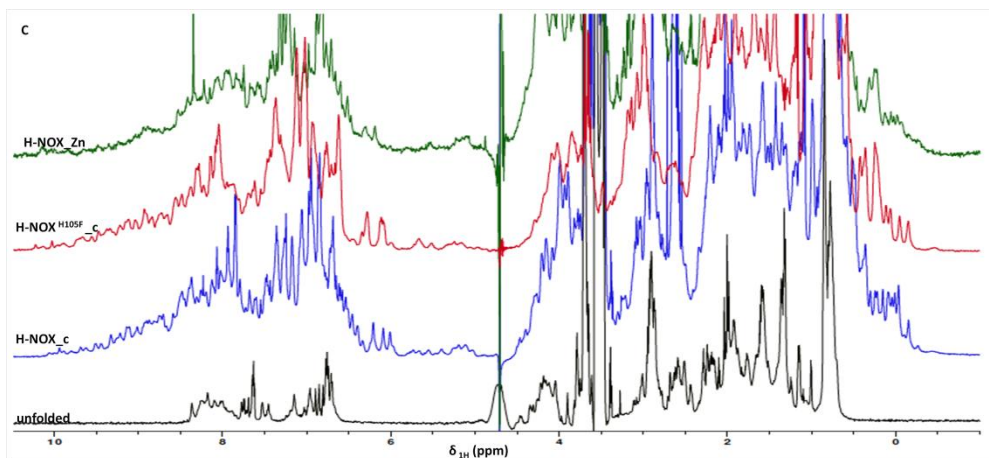
C- 1D proton spectra collected at 25˚C on the 800 MHz NMR. Black line: Unfolded apo- H-NOX. Blue line: Monomeric WT H-NOX in complex with cinaciguat. Red line: Monomeric mutant[H105F] in complex with cinaciguat. Green line: WT H-NOX in complex with Zn^{2+} PPIX.

Discussion

Investigations were initially focused on understanding the human H-NOX domain complexed with different ligands, which allows the domain to be forms representing the resting, active and inactive states. We also wanted to characterize its stability for functional and structural studies by NMR. In order to do that, we established a protocol for expression and purification for human H-NOX domain in *E.coli* that would allow rapid and efficient isolation of the sGC heme domain. Indeed, after testing different conditions, the WT and H-NOX[H105F] forms of the sGC domain were successfully expressed in *E.coli* (Figure 1).

To perform X-ray crystallography [10,11], Resonance Raman spectroscopy [22] and Electron paramagnetic resonance protein studies [15,17], unlabeled H-NOX sGC homologous and human α and β-subunit constructs were obtained in rich media (e.g Lysogeny broth, 2x TY or modified Terrific Broth) [10,11,15,17,22]. However, for NMR measurements, protein samples needed to be highly concentrated (around 0.4–0.8 mM) and isotopically labeled with ^{13}C and/or ^{15}N [31]. So, to obtain labeled protein, *E. coli* strains need to grow in media containing the appropriate isotopes. In cases that large quantities of proteins are required, *E. coli* is often grown in minimal media supplemented with additional carbon and/or nitrogen labeled sources [32]. However, we found no significant increase of H-NOX domain expression in supplemented media, when compared with protein expressed in minimal media where no additional supplement was added.

Purification of the H-NOX domain was also optimized. The initial protocol used of nickel affinity chromatography, followed by anionic exchange (H-NOX domain isoelectric point 5.39), protein ligand incorporation process and size exclusion chromatography. However, after tests we concluded that the ion exchange step did not contribute either to the purity or the stability of the protein, so it was removed (Figure 1B).

Pioneer research on sGC described the protein as an α/β-heterodimeric heme-protein that, upon sensing NO, generates cGMP [33-37]. Although the related family of particulate guanylate cyclases (pGCs) forms active homodimeric complexes, it is not clear whether homodimerization of sGC subunits occurs [38]. Zabel and collaborators, after expressing full length α and β sGC subunits in Sf9 insect cells found α/α and β/β homodimeric complexes. However, upon co-expression of the respective complementary subunits, α/β heterodimeric interactions were preferentially formed. So, as a conclusion, it was proposed the existence of a physiological equilibrium between homo and heterodimeric sGC complexes [39].

In this thesis we describe the expression of the human apo H-NOX domain in *E.coli* cells, and its isolation with protein purification methods. During this process, we observed that in the absence of heme, the majority of this protein domain exists in various, yet distinct states of aggregation (**Figure 1**C). Since the protein was expressed and purified in the apo-form, there is a possibility that the hydrophobic heme cavities could be interacting with each other inducing the protein aggregation. This result has already been reported by Marletta and other researchers, where they described protein aggregation when expressing the H-NOX domain from rat and human [15,22,40]. Moreover, protein aggregation was also observed for the full β sGC subunit upon heme oxidation and consequent release (turning into full β apo-sGC). Since the ββ homodimerization of the full human sCG protein has already been described as a physiological equilibrium between homo- and heterodimeric sGC complexes, we may speculate that the β apo-HNOX subunit aggregation may reflect this physiological effect. So, studying how this aggregation is mediated would aid valuable information to this complex system.

Cinaciguat has been described to be involved in sGC activation in cells that are facing or have faced oxidative stress. This compound has been proved to be

easily and fast incorporated into full length WT and H-NOX[H105F] sGC [28,29,41,42]. However, the absence of monomeric apo-H-NOX domain upon expression in *E.coli* cells, made the protein compound incorporation a challenge.

Initial ligand incorporation experiments were based on a protocol developed by Martin and collaborators, in which *Nostoc sp* H-NOX-cinaciguat bound was obtained by adding a 10-fold molar excess of heme oxidizer and 5-fold molar excess of cinaciguat at 37°C followed by size exclusion chromatography [10]. However, in this protocol the length of the incubation time was not described. So, when that protocol was used in our case, the complex protein-compound was not observed. The aggregation of H-NOX domain could be the cause for this failed attempt of incorporation, since the monomeric form of the heme free H-NOX from *Nostoc sp* has been reported to be easily over-expressed in *E.coli* system [10,11] One additional possibility could be related to the inherent variability in the quality of the produced protein which is not always constant. Thus, the incorporation protocol was slightly modified. An incubation step prior to size exclusion was added to the protocol, where cinaciguat was mixed with H-NOX domain in the aggregated form. So, despite the fact that *in vivo* and *in vitro* cell, cinaciguat is a fast and potent dilator of both arterial and venous blood vessels, the interaction between pure protein and compound may be slower because of the presence of aggregated forms that differ from those present in the natural heme free system [41,43].

Cinaciguat is not very soluble in water, which may cause protein precipitation when mixed at high concentrations. However, according to Hassell and collaborators, it is possible to concentrate the protein and then add the compound to achieve complex formation [44]. Therefore, by increasing the protein concentration from 67 to 134µM, and decreasing the molar excess of cinaciguat over H-NOX in the incubation solution from 1:5 to 1:2 (protein-compound, respectively), indeed, avoided compound precipitation (Figure 2D).

Increasing the temperature to ease protein-ligand incorporation reaction has been described to favor the incorporation process [45]. Nevertheless, in our work was found that temperature change from 37 to 20°C was key in order to preserve protein stability, protect the complex and ultimately avoid degradation by proteases that may have not been completely removed by nickel affinity chromatography. Last but not least, 0.05% of n-octyl glucoside was added during the incubation procedure, since it has been described to increase protein flexibility and promote protein-compound incorporation when added in low concentrations [46].

In an attempt to explain the dissociation of the oligomeric form into monomeric H-NOX_c complex (Figure 3D), we wanted to evaluate if hydrophobic interactions played any role. n-octyl glucoside, but also cinaciguat in solution, can promote hydrophobic interactions. We found that in the absence of n-octyl glucoside and when the protein concentration was increased from 67μM to 134μM, less monomeric protein was observed. Moreover, the protein-cinaciguat ratio 1:1 also showed less oligomeric protein dissociation into monomeric protein-compound complex than protein-cinaciguat ratio 1:2 and 1:3 respectively. Therefore, in our case that the H-NOX domain was initially expressed and purified as a mixture between aggregated/oligomeric proteins, so, the hydrophobic strength in solution might have been important for full oligomeric dissociation into monomeric H-NOX_c complex. Unlike the oligomeric form, the aggregated protein was partially converted into monomeric, since human H-NOX domain in aggregated form was observed even after 18h of protein compound incubation Figure 3D. It could be due to the existence of equilibrium between folded/unfolded, molten globule or a mixture between these state of the protein in solution (Figure 3A-B). However, the partially unfolded protein did not affect the incorporation of cinaciguat into H-NOX domain. Moreover, the monomeric

complex showed to be stable during the incubation/purification process (Figure 2E).

In the literature, several reports have described different strategies for *in cell* heme protein incorporation. [47-50]. However, low incorporation of heme into human H-NOX domain has adversely influenced the characterization of this complex. To overcome this, various strategies have been designed where human H-NOX was co-expressed with FC in *E.coli* cells. For Sudhamsu and collaborators, this method resulted in full heme incorporation into *Geobacillus stearothermophilus* Nitric Oxide Synthase (GsNOS) [16]. In our case, the result showed no significant amount of H-NOX heme incorporation (Figure 4 A-C). Similar result was obtained under co-expression of N-terminal MBP-H-NOX protein together with FC. Despite the fact that maltose-binding protein has been described to have high efficiency to increase the solubility of β1 and α2-β1 human H-NOX and ease heme insertion [17], our result showed that MBP and FC had no role in enhancing the over-expression of heme containing β1 human H-NOX.

According to Derbyshire and Zhong, H-NOX heme binding can be obtained using a heme reconstitution protocol [15,24]. From UV-vis observations, there was an indication of H-NOX heme binding (Figure 4D). However, the purification product was again the aggregated form and oligomeric form as had been previously seen in the case of the H-NOX domain that was in apo-form. An explanation for this can be that the covalent bond between heme Fe^{2+} and H105 residue could not be strong enough to stabilize the protein-heme complex. In spite of having a successful H-NOX heme incorporation, Zhong and collaborators observed a quick oxidation and further aggregation of the complex [15].

In order to both test the capacity of binding of other metalated porphyrins and also to produce a model of the resting yet inactive state of the H-NOX domain, the protein was complexed with the heme analogue ZnPPIX. As an added advantage, ZnPPIX is also diamagnetic which is a great advantage for NMR

studies. This prosthetic group has been described to directly inhibit sGC, being able to be up taken and incorporated into newly synthesized sGC. Thus, the H-NOX_Zn complex would be extremely useful as a model for H-NOX to perform NMR dynamics studies. Our result showed that the incorporation of ZnPPXI into human H-NOX domain was successful. In the end, we were able to purify the protein in the monomeric form. Moreover, based on the shift change observed in an UV-vis spectrum, ZnPPXI was well incorporated into H-NOX domain [17] (figure 5 A-B). In summary, our data showed that the monomeric form of human H-NOX bound with ZnPPIX forms a stable complex, allowing us to execute and follow NMR dynamics studies.

The 2D ^{1}H-^{15}N HSQC spectra showed signals for amide protons and is described as a fingerprint of a protein. This spectrum can be used to observe changes in chemical shifts and map site interactions upon ligand incorporation. These spectra were collected for the H-NOX_c and H-NOX^{H105F}_c monomeric proteins. Monomeric H-NOX_c showed high dispersion pattern quality. By overlaying this spectrum with the one from H-NOX^{H105F}_c and H-NOX_Zn, it was possible to see that there were extended similarities with the majority of the signals appeared at very similar positions. The apo protein could not be observe due to its high molecular weight which severely complicates the acquisition of NMR spectra.

Cinaciguat was developed to be a flexible molecule, allowing it to be easily incorporated into sGC H-NOX domain. Also its interaction with the protein is independent of histine105 binding. Unlike cinaciguat, ZnPPIX heme-analogue has a planar structure and its incorporation into the protein pocket and sub-sequent stability is depend of conserved residues (YxSxR) and might form a metal Zinc^{2+} histine105 covalent bond. Despite the differences between these ligands, the 2D ^{1}H-^{15}N HSQC spectra showed to be highly similar, not losing its signal distribution pattern.

Together, all of these results showed for the first time the incorporation of cinaciguat and a heme-analogue into the human H-NOX domain in the monomeric form, allowing further protein assignment and structural and dynamics studies.

References

1. Assenberg R, Wan PT, Geisse S, Mayr LM (2013) Advances in recombinant protein expression for use in pharmaceutical research. Curr Opin Struct Biol 23: 393-402.
2. Dutta A, Saxena K, Schwalbe H, Klein-Seetharaman J (2012) Isotope labeling in mammalian cells. Methods Mol Biol 831: 55-69.
3. Hansen AP, Petros AM, Mazar AP, Pederson TM, Rueter A, et al. (1992) A practical method for uniform isotopic labeling of recombinant proteins in mammalian cells. Biochemistry 31: 12713-12718.
4. Lustbader JW, Birken S, Pollak S, Pound A, Chait BT, et al. (1996) Expression of human chorionic gonadotropin uniformly labeled with NMR isotopes in Chinese hamster ovary cells: an advance toward rapid determination of glycoprotein structures. J Biomol NMR 7: 295-304.
5. Seeger F, Quintyn R, Tanimoto A, Williams GJ, Tainer JA, et al. (2014) Interfacial residues promote an optimal alignment of the catalytic center in human soluble guanylate cyclase: heterodimerization is required but not sufficient for activity. Biochemistry 53: 2153-2165.
6. Allerston CK, von Delft F, Gileadi O (2013) Crystal structures of the catalytic domain of human soluble guanylate cyclase. PLoS One 8: e57644.
7. Ma X, Beuve A, van den Akker F (2010) Crystal structure of the signaling helix coiled-coil domain of the beta1 subunit of the soluble guanylyl cyclase. BMC Struct Biol 10: 2.
8. Purohit R, Weichsel A, Montfort WR (2013) Crystal structure of the Alpha subunit PAS domain from soluble guanylyl cyclase. Protein Sci 22: 1439-1444.
9. Herzik MA, Jr., Jonnalagadda R, Kuriyan J, Marletta MA (2014) Structural insights into the role of iron-histidine bond cleavage in nitric oxide-induced activation of H-NOX gas sensor proteins. Proc Natl Acad Sci U S A 111: E4156-4164.
10. Martin F, Baskaran P, Ma X, Dunten PW, Schaefer M, et al. (2010) Structure of cinaciguat (BAY 58-2667) bound to Nostoc H-NOX domain reveals insights into heme-mimetic activation of the soluble guanylyl cyclase. J Biol Chem 285: 22651-22657.
11. Ma X, Sayed N, Beuve A, van den Akker F (2007) NO and CO differentially activate soluble guanylyl cyclase via a heme pivot-bend mechanism. Embo j 26: 578-588.
12. Underbakke ES, Iavarone AT, Chalmers MJ, Pascal BD, Novick S, et al. (2014) Nitric oxide-induced conformational changes in soluble guanylate cyclase. Structure 22: 602-611.

13. Campbell MG, Underbakke ES, Potter CS, Carragher B, Marletta MA (2014) Single-particle EM reveals the higher-order domain architecture of soluble guanylate cyclase. Proc Natl Acad Sci U S A 111: 2960-2965.

14. Underbakke ES, Iavarone AT, Marletta MA (2013) Higher-order interactions bridge the nitric oxide receptor and catalytic domains of soluble guanylate cyclase. Proc Natl Acad Sci U S A 110: 6777-6782.

15. Zhong F, Wang H, Ying T, Huang ZX, Tan X (2010) Efficient expression of human soluble guanylate cyclase in Escherichia coli and its signaling-related interaction with nitric oxide. Amino Acids 39: 399-408.

16. Sudhamsu J, Kabir M, Airola MV, Patel BA, Yeh SR, et al. (2010) Co-expression of ferrochelatase allows for complete heme incorporation into recombinant proteins produced in E. coli. Protein Expr Purif 73: 78-82.

17. Wang H, Zhong F, Pan J, Li W, Su J, et al. (2012) Structural and functional insights into the heme-binding domain of the human soluble guanylate cyclase alpha2 subunit and heterodimeric alpha2beta1. J Biol Inorg Chem 17: 719-730.

18. Redfield C (2004) NMR studies of partially folded molten-globule states. Methods Mol Biol 278: 233-254.

19. Shortle DR (1996) Structural analysis of non-native states of proteins by NMR methods. Curr Opin Struct Biol 6: 24-30.

20. Bai P, Song J, Luo L, Peng ZY (2001) A model of dynamic side-chain--side-chain interactions in the alpha-lactalbumin molten globule. Protein Sci 10: 55-62.

21. Schelvis JP, Zhao Y, Marletta MA, Babcock GT (1998) Resonance raman characterization of the heme domain of soluble guanylate cyclase. Biochemistry 37: 16289-16297.

22. Karow DS, Pan D, Davis JH, Behrends S, Mathies RA, et al. (2005) Characterization of functional heme domains from soluble guanylate cyclase. Biochemistry 44: 16266-16274.

23. Serfass L, Burstyn JN (1998) Effect of heme oxygenase inhibitors on soluble guanylyl cyclase activity. Arch Biochem Biophys 359: 8-16.

24. Derbyshire ER, Deng S, Marletta MA (2010) Incorporation of tyrosine and glutamine residues into the soluble guanylate cyclase heme distal pocket alters NO and O2 binding. J Biol Chem 285: 17471-17478.

25. Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins M, et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: Walker J, editor. The Proteomics Protocols Handbook: Humana Press. pp. 571-607.

26. Fritz BG, Roberts SA, Ahmed A, Breci L, Li W, et al. (2013) Molecular model of a soluble guanylyl cyclase fragment determined by small-angle X-ray scattering and chemical cross-linking. Biochemistry 52: 1568-1582.

27. Patra M, Mukhopadhyay C, Chakrabarti A (2015) Probing conformational stability and dynamics of erythroid and nonerythroid spectrin: effects of urea and guanidine hydrochloride. PLoS One 10: e0116991.

28. Schmidt P, Schramm M, Schroder H, Stasch JP (2003) Mechanisms of nitric oxide independent activation of soluble guanylyl cyclase. Eur J Pharmacol 468: 167-174.

29. Stasch JP, Schmidt P, Alonso-Alija C, Apeler H, Dembowsky K, et al. (2002) NO- and haem-independent activation of soluble guanylyl cyclase: molecular basis and cardiovascular implications of a new pharmacological principle. Br J Pharmacol 136: 773-783.

30. Stasch JP, Schmidt PM, Nedvetsky PI, Nedvetskaya TY, H SA, et al. (2006) Targeting the heme-oxidized nitric oxide receptor for selective vasodilatation of diseased blood vessels. J Clin Invest 116: 2552-2561.

31. Erbil WK, Price MS, Wemmer DE, Marletta MA (2009) A structural basis for H-NOX signaling in Shewanella oneidensis by trapping a histidine kinase inhibitory conformation. Proc Natl Acad Sci U S A 106: 19753-19760.

32. Marley J, Lu M, Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. J Biomol NMR 20: 71-75.

33. Koshland DE, Jr., Nemethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5: 365-385.

34. Hardman JG, Sutherland EW (1969) Guanyl cyclase, an enzyme catalyzing the formation of guanosine 3',5'-monophosphate from guanosine trihosphate. J Biol Chem 244: 6363-6370.

35. Ishikawa E, Ishikawa S, Davis JW, Sutherland EW (1969) Determination of guanosine 3',5'-monophosphate in tissues and of guanyl cyclase in rat intestine. J Biol Chem 244: 6371-6376.

36. White AA, Aurbach GD (1969) Detection of guanyl cyclase in mammalian tissues. Biochim Biophys Acta 191: 686-697.

37. Ignarro LJ, Byrns RE, Buga GM, Wood KS (1987) Endothelium-derived relaxing factor from pulmonary artery and vein possesses pharmacologic and chemical properties identical to those of nitric oxide radical. Circ Res 61: 866-879.

38. Lucas KA, Pitari GM, Kazerounian S, Ruiz-Stewart I, Park J, et al. (2000) Guanylyl cyclases and signaling by cyclic GMP. Pharmacol Rev 52: 375-414.

39. Zabel U, Hausler C, Weeger M, Schmidt HH (1999) Homodimerization of soluble guanylyl cyclase subunits. Dimerization analysis using a glutathione s-transferase affinity tag. J Biol Chem 274: 18149-18152.

40. Zhao Y, Marletta MA (1997) Localization of the heme binding region in soluble guanylate cyclase. Biochemistry 36: 15959-15964.

41. Cosyns SM, Huyghe L, Thoonen R, Stasch JP, Brouckaert P, et al. (2014) Influence of cinaciguat on gastrointestinal motility in apo-sGC mice. Neurogastroenterol Motil 26: 1573-1585.
42. Chester M, Seedorf G, Tourneux P, Gien J, Tseng N, et al. (2011) Cinaciguat, a soluble guanylate cyclase activator, augments cGMP after oxidative stress and causes pulmonary vasodilation in neonatal pulmonary hypertension. Am J Physiol Lung Cell Mol Physiol 301: L755-764.
43. Boerrigter G, Costello-Boerrigter L, Lapp H, Stasch J-P, Burnett J (2005) Co-activation of soluble and particulate guanylate cyclase by BAY 58-2667 and BNP enhances cardiorenal function in experimental heart failure. BMC Pharmacology 5: P5.
44. Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL, 3rd, et al. (2007) Crystallization of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr 63: 72-79.
45. Perozzo R, Folkers G, Scapozza L (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects. J Recept Signal Transduct Res 24: 1-52.
46. Ishihara G, Goto M, Saeki M, Ito K, Hori T, et al. (2005) Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors. Protein Expr Purif 41: 27-37.
47. Weickert MJ, Pagratis M, Curry SR, Blackmore R (1997) Stabilization of apoglobin by low temperature increases yield of soluble recombinant hemoglobin in Escherichia coli. Appl Environ Microbiol 63: 4313-4320.
48. Kery V, Elleder D, Kraus JP (1995) Delta-aminolevulinate increases heme saturation and yield of human cystathionine beta-synthase expressed in Escherichia coli. Arch Biochem Biophys 316: 24-29.
49. Varnado CL, Goodwin DC (2004) System for the expression of recombinant hemoproteins in Escherichia coli. Protein Expr Purif 35: 76-83.
50. Woodward JJ, Martin NI, Marletta MA (2007) An Escherichia coli expression-based method for heme substitution. Nat Methods 4: 43-45.

Supplemental Data

S1: Table of the minimal media composition

10X M9 Minimal Media(1L)		**Media Composition (1Liter)**	
$Na_2HPO_4.7H_2O$	64g	Sterile MilliQ H2O	880ml
KH_2PO_4	25g	10x M9 media	100ml
		1M $MgSO_4$	2ml
NaCl	2.5g	1M $CaCl_2$	20 µl
		Metals Stock	20 µl
		Vitamin Mix stock	10ml
		Antibiotic 100 mg/ml	1ml
		C-Glucose	3.2g
		$^{15}NH_4Cl$	1g
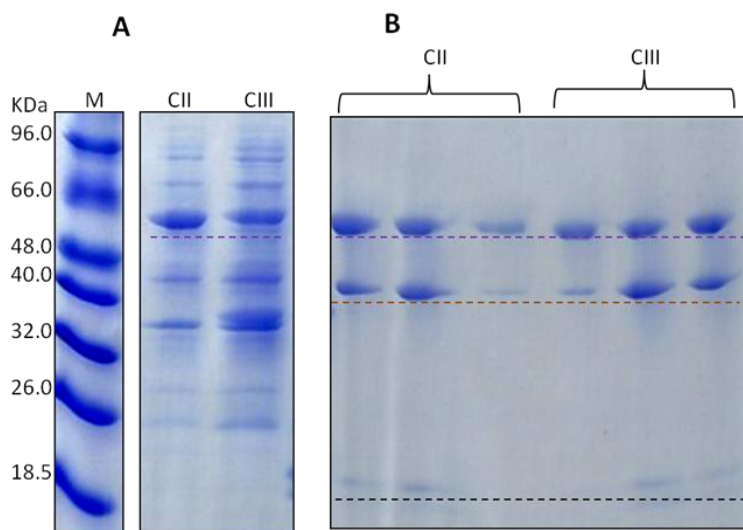
S1: SDS PAGE (Tris-Glycine 12.5%). A: OD 1.4 normalized sample from overnight optimized expression test, 1mM IPTG at 20˚C were used for induction. M- Protein marker. Purple line: WT H-NOX domain with N-terminal MBP tag (66 kDa). **B:** WT H-NOX domain with N-terminal MBP tag (66 kDa) (C#2 and C#3) cleaved with HRV 3C protease for 16h at 4°C. Purple line: WT H-NOX domain with N-terminal MBP tag which were not cleaved by HRV 3C protease. Brown line: MBP tag (~43 kDa). Black line: WT H-NOX domain (22.3 kDa). The protein was cleaved with 5, 10 and 20µg of the HRV3C protease.
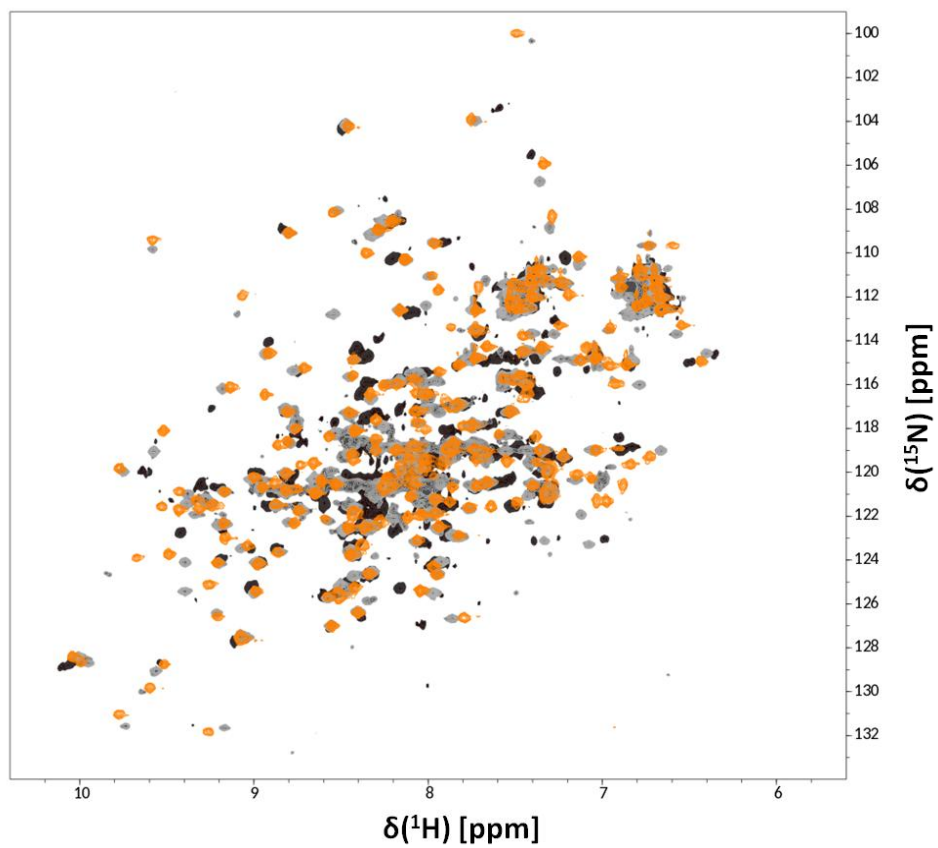
S3: Overlay of the 2D ^{1}H-^{15}N HSQC spectra. Grey line: Monomeric H-NOX_c. Orange line: Monomeric H-NOX^{H105F}_c. Black line: Monomeric H-NOX_Zn.

Chapter 3

Structural characterization of the cinaciguat bound H-NOX domain of human sGC

The preparation of the double labeled ^{13}C/^{15}N cinaciguat bound H-NOX domain sample were performed by Dr. Monica Thomaz from Ibet-Instituto de Biologia Experimental Tecnológica. The NMR protein assignment data collections were done by thesis supervisor Dr. Manolis Matzapetakis. The protein structure calculation was done by the member of Biomolecular NMR group Dr. Ivo Saraiva. Part of protein assignment and structural analysis was made by the author of this dissertation.

Table of Contents

Introduction

1 General Introduction

Under physiological conditions most proteins fold spontaneously into specific three-dimensional structures and these structures are connected to their function. Therefore, unveiling the three-dimensional structure of proteins has always been matter of extreme relevance to understand protein function. The natural algorithms that dictate protein three-dimensional structure from amino acid sequence remain elusive. Therefore, experimental efforts for the determination of biomolecular structures have been subjects of intense interest for a long time. In 1958 Max Perutz and John Kendrew solved the first molecular structures of the two related proteins myoglobin and hemoglobin, by X-ray crystallography [1,2]. Years later, in 1982-1984, the first three-dimensional structure of the bull seminal protease inhibitor was determined by NMR [2]. This approach allowed detailed structural studies of proteins in solution and other non-crystalline states, opposite to the well-established diffraction techniques with crystallized proteins to be performed. This new method allow biomolecules to be placed in a more comparable or identical environment to the physiological one, and promises new insights into the dynamics of protein molecules [3].

Research Collaboratory for Structural Bioinformatics (RCSB) protein data bank (PDB) annual report (RCSB annual report, 2013) listed 91761 atomic coordinate entries available in the PDB repository, with 93% of the entries attributed to proteins and peptide molecules. Both, X-Ray crystallography and NMR represent the two major techniques used for protein structure determination at high resolution quality. According to protein data bank, there are around 40 structures of prokaryote H-NOX proteins/domains homologous to sGC heme domain listed into the data bank. Two of these structures were solved by NMR [4].

2   Protein assignment by NMR

In NMR the observable chemical shift resonances are associated with chemical entities [3,5]. Identification of these resonances is the first step in the many NMR based processes. For proteins larger than 16kDa and smaller than 25kDa the use of triple resonance experiments is needed for resonance assignment. These experiments utilize ^{1}H, ^{13}C and ^{15}N spins in schemes that allow the connection of resonances from sequential aminoacids, while the inherent spectral grouping and resolution of different carbon spin types allows the identification of the different aminoacid types[5,6]. Assignment of protein backbone atoms labeled with ^{1}H, ^{15}N and ^{13}C is possible by using experiments that rely on several well-resolved heteronuclear couplings, which allow multiple step coherence transfers in a time frame that agrees with the intrinsic relaxation times of the three nuclei involved [7,8]. These heteronuclear couplings allow an uninterrupted sequence of the active nuclei throughout protein backbone. These experiments suppress the ambiguity found on older assignment strategies that rely only on nuclear Overhauser effect to supply inter-residue correlations, for instance, ^{1}H-^{1}H NOESY [5] [9,10].

In this work, ^{1}H-^{15}N-HSQC, ^{1}H-^{13}C-HSQC, HNCO, HNcaCO and hcaCONH, HNCA, HNCACB, HNcocaHA, CBCAcoNH and hCCH–TOCSY spectra were used to assign the human H-NOX domain (M.W. 22.3 kDa).   The NMR assignment and structure determination of the *Shewanella oneidensis* HNOX using triple resonance experiments has already shown this to be a functional and efficient methodology [4]. As a first step of the backbone assignment, ^{1}H-^{15}N-HSQC 2D spectrum provides an initial set of HN resonance assignment that will work as a guide through the 3D spectra. This spectrum shows one peak for each covalently bound ^{1}H-^{15}N pair, where the magnetization is transferred from ^{1}H to ^{15}N via J-coupling. The chemical shift is evolved on the nitrogen atom and the magnetization is then transferred

back to the hydrogen atom for detection. Since the backbone of a protein only contains one amide group per residue (except for Proline residues), this experiment can be used to directly identify H-N chemical shifts of each residue [3,8]. After NH atom assignment, HNCO experiment is used to observe CO(i)-N(i + 1),NH(i + 1) correlations, an HNcaCO or in the case of bigger proteins hcaCONH, to observe HN (i), CO(i) correlations. For HN (i), CA (i, i-1) and HN (i), CA (i-1)correlations we use the HNCA the HNcoCA respectively. For the sidechain CB we use the combinations of HNCACB for HN (i), CA (i, i-1), CB (i, i-1) correlations, and CBCAcoNH for HN (i), CACB (i-1) correlations. The backbone intra-residue Hα H(i)-^{15}N(i)-NH(i) correlations are provided by the ^{15}N-edited TOCSY and H(CA)NH experiments, the Hβ H(i)-^{15}N(i)-NH(i) correlations by the ^{15}N-edited TOCSY and HNHB experiments. Aliphatic side-chain spin systems can be assigned with the 3D (^{1}H)-^{13}C-^{13}C-^{1}H correlated hCCH-TOCSY experiment [7].

Validating the precision and completeness of the chemical shifts assignments is critical for the robustness of the subsequent steps of structure determination. Determined chemical shifts values are compared with averages of the BMRB database that has listings of the majority of protein assignments made to date [11,12]. Deviations of the reported values by more than 4 standard deviations are flagged. These values are then manually evaluated to investigate if they are correctly attributed and if they are, why they deviate so significantly from the average. Often, deviations are related to proximity to aromatic rings that can shift resonances due to their ring current effect.

3 NMR Structural characterization

3.1 Nuclear Overhauser Effect (NOE)

The most significant parameters for structure determination by NMR are the NOEs. Long-range NOEs provide essentially tertiary structure information, while secondary structure elements such as α-helices or β-sheets are defined by short-range NOEs. The NOE intensity (I) is related to the distance (r) between the pair of hydrogen atoms: $I = ƒ(τ_c) < r^{-6} >$ in which $ƒ(τ_c)$ is a function of the rotational correlation time [9]. In order to convert intensities to qualitative distance data, usually a constant instead of the function of $τ_c$ is used to avoid the highly variable $τ_c$ values from different molecules at different temperatures and solvent conditions. Many procedures can be applied for that conversion. A calibration constant that will set the median of the intensities to the average distance of the NMR observer distances (3.3-5 Å) is the most common method used. Some errors can interfere with NOE acquisition, including spin diffusion, dynamics and low signal to noise thus NOE peaks can be evaluated as ranges of distances with a lower and higher limit: 1.8-2.5 Å (strong) 1.8-3.5 Å (medium) and 1.8-5-0 Å (weak) where 1.8 Å according to the sum of the van der Waals radii [6].

3.2 Chemical shifts as constrains

Chemical shifts are sensitive probes of molecular structure [13,14]. They are sensitive to minute changes of protein structure and can be used to probe complexation of proteins at an atomic level, hydrogen bonding interactions, ionization and oxidation states, ring current influences from neighboring aromatic residues, and intrinsic exchange dynamics of hydrogen atoms [15]. In structure determination studies, comparison of chemical shifts with database values allow the prediction of the secondary structure of protein [14,16]. It also is able to support the refinement step of complex protein structures and to probe

conformational changes related to partial protein unfolding, as well as, molecular binding events [17,18,19,20]. The amino acid sequence information of the query protein in combination with chemical shift assignments can also be used as constraints for the backbone ϕ and ψ dihedral angles. These can be created using tools such as the DANGLE (Dihedral Angles from Global Likelihood Estimates) algorithm. This tools computes sequence-corrected secondary shifts for all measured $^1H^\alpha$, ^{15}N, $^{13}C'$, $^{13}C^\alpha$ and $^{13}C^\beta$ nuclei and subsequently predict dihedral angles for each protein residue [17,18]. Bayesian inference calculates approximately the conformations throughout the Ramachandran space by focusing the populations distribution specific for the different amino acid types [19]. As an alternative, ^3J-scalar coupling measurements can be used to estimate dihedral angles from empirically-defined relationships using the Karplus equation [20].

The chemical shift information can assist in the determination of the tertiary structure of proteins when used in combination with other NMR information that report on interproton distances (NOEs), dihedral angles and the relative orientations of the different nuclei in a protein structure through residual dipolar couplings (RDC) measurements [14,21-23].

3.3   Preliminary secondary structure determination

Protein secondary structure is formed by a hydrogen bonds pattern between amine hydrogen and carbonyl oxygen atoms from the backbone. Protein backbone dihedral angles in a particular region of the Ramachandran plot define alpha helices or beta sheets regimes. In NMR studies, before to tertiary structure calculation, the dependence of chemical shift on backbone dihedral angles can be used to predict secondary structure [6].

3.4 3D structure determination of a protein-ligand complex

Similar to protein alone, the structural characterization of protein ligand complex can be achieved by NMR spectroscopy. The chemical shifts from protein assignment and NOE accessibility for the ligand-free protein, with the ligand-induced chemical shifts perturbation data is capable to support the assignment of the protein-ligand complex. Equivalent to what occurs with protein homodimers, the major limitation of the assignment is the intermolecular NOEs identification among the protein and the ligand, since the signal observed is similar to intramolecular NOEs in standard isotope-edited NOESY spectra. Thus, 3D F1-^{13}C/^{15}N-filtered, F3-^{13}C-edited NOESY spectrum, ^{15}N/^{13}C labelled protein samples can be used with excess of the unlabelled ligand. These experiments are only able to present NOEs between protons attached to ^{13}C protein and protons attached to unlabeled ligand. It also is able to suppress any other cross peaks by isotope filtering leading to assign the protons from ligand (cinaciguat and/or ZnPPIX) [24].

3.5 Structure calculation software.

The experimental data described above can be used in the protein structure calculation procedure. Several computers software requires experimentally acquired NMR data as restraints. Currently, the most often used are CYANA, CNS (Crystallography & NMR System) [11,12], ARIA (Ambiguous Restraints for Iterative Assignment) [25] and CCPN (Collaborative Computational Project for NMR) [26]. Protein structure calculation software converts the experimental data (input) into *in silico* atomic coordinate information (output). In order to calculate a structure from a random and extended structure, molecular dynamic simulated annealing methods with cartesian or torsion angle space are required. The evaluation of the calculation with the use of a target function is the crucial attribute of all these methods. The target function has many terms that evaluate the agreement of the experimental data with the calculated structure as well as with known physical

properties. The van der Waals and electrostatics interactions can provide information related to molecular geometry of the bond angles and length. These interactions can be monitored by force field, which contain physical energy terms. A subsequent final refinement in water solvent can be done to achieve a 3D protein structure closer to their natural conditions [27,28].

3.6   Homology Modeling

Homology modelling simulations of proteins are a powerful tool to provide an empirical 3D protein structure when no experimental three-dimensional model data is available. Building homology models involves submitting a target sequence to specialized programs with updated sequences and structural databases. SWISS-MODEL [29] and ROSETTA [30] are web-based tools that perform protein structure homology modelling, which accuracy is dependent on the degree of sequence similarity in respect to the amino acid sequences with experimentally determined three-dimensional structures and already deposited in the protein data bank. If the template and target sequences contain high degree of similarity (higher than 50%) prediction quality is considered very high.

Comparing the three-dimensional structure of human H-NOX complexed with cinaciguat with three-dimensional structures of homologous proteins can indicate conserved regions within the amino acid sequence, which can be involved in ligand binding. Moreover, based on current literature, it allows molecular insight into the domain region involved in protein activation.

3.7   Structure validation

The quality of a three-dimensional atomic model has to be evaluated. Three steps can be followed for structure validation. The first is to compare the data to deposited experimental measurements. The second, is to verify the agreement of

the atomic model with the experimental data. The last one is to verify model consistency with known physical and chemical properties.

Material and methods

Homology modeling

The GenBank identifiers for sequential alignment were *Homo sapiens*, GUCY1A3; *Nostoc sp*, GI:17229770; *Thermoanaerobacter tengcongensis*, GI:20807169; *Shewanella oneidensis,* GI:470469643. Each FASTA format of the amino acid sequence was loaded together into Phyre2 program [31]. An estimate of the human H-NOX secondary structure and 3D homology models were obtained by submitting the amino acid sequence (from WT H-NOX) into SWISS-MODEL workspace [29]. The program identified various templates; and the quality of the template was predicted from features of the target-template alignment.

Protein resonance assignment

NMR experiments were carried out at 298.15 K on a Bruker Avance III spectrometer (Bruker, Rheinstetten, Germany) operating at ^{1}H frequency of 800.33 MHz, using a room temperature 5 mm TXI H C/N/-D Z axis gradient probe. Proton chemical shifts were referenced with respect to residual solvent signal (4.7 ppm at 298.15 K, 50 mM NaCl and pH 6.5) and the frequency ratios of ^{15}N/^{1}H = 0.1013 and ^{13}C/^{1}H = 0.251. For sequential backbone assignment the following 2D and 3D experiments were used:2D ^{1}H-^{15}N HSQC, 2D ^{1}H-^{13}C HSQC,3D HNCO,3D HNcaCO, HNCA,3D HNCACB, 3D CBCAcoNH, 3D HNcocaHA and the 3D HBHAcoNH [5,6]. The assignment of the remaining aliphatic and aromatic ^{1}H and ^{13}C side chain resonances was accomplished using combinations of the 3D ^{1}H–^{15}N TOCSY–HSQC the 3D hCCH–TOCSY and the 3D ^{1}H-^{15}N HSQC-NOESY and 3D ^{1}H-^{13}C HSQC-NOESY. Data was processed with Bruker's Topspin 2.1 software (Bruker Biospin) and analyzed with CCPN (v2.4) software.

Secondary structural prediction

After the H-NOX_c assignment by NMR the chemical shift resonances from backbone and Cα, Cβ, CO and Hα were used as input to characterize the protein secondary structure. The data was load into PECAN service program [32].

Protein assignment transfer

After the assignment of the H-NOX_c was complete, the identified resonances in the ^{1}H-^{15}N HSQC spectra were used to support the assignment transfer to the homologous ^{15}N HSQC spectra of both HNOX^{H105F}_c and H-NOX_Zn complexes. The assignment transfer was based on the closeness of chemical shift, only where no ambiguities were observed. Evaluation of the transfers was also made by analyzing the chemical shift differences as a function of sequence. Combined Chemical shift differences are calculated using the following equation:

$$\text{Distance} = \sqrt{(\delta H)^2 + 0.156(\delta N)^2}$$

where 0.156 is a normalization factor for $\pmb{\delta}$N based on the standard deviation of the respective values reported in the diamagnetic of chemical shifts of the Biological Magnetic Resonance Data Bank (BMRDB).

3D structure determination of a protein-ligand complex

To obtain distance restraints for the structure calculation of cinaciguat bound H-NOX complex, the following spectra were collected and used: ^{15}N-HSQC-NOESY, ^{13}C-HSQC-NOESY, 2D ^{1}H-^{1}H NOESY ^{13}C-filtered in one dimension, 2D ^{1}H-^{1}H NOESY ^{13}C-filtered in both dimensions and a ^{13}C-HSQC-NOESY ^{13}C-filtered in one dimension. Given that cinaciguat was not isotopically labelled, the filtered experiments provided both H-NOX intra-residual NOEs and inter-residual NOEs of the cinaciguat-protein complex. Together with a 2D TOCSY ^{13}C-filtered in both

dimensions, these experiments were also used to assign the signals of the ligand. ARIA 2.3 software (Ambiguous Restraints for Iterative Assignment) [25,27] was used to automatically assign parts of the manually picked peak lists of the ^{15}N-HSQC-NOESY and the ^{13}C-HSQC-NOESY, while the filtered experiments were assigned manually. DANGLE [19] was used to obtain dihedral restraints. The distance and dihedral restraints were then used to calculate 100 structural models with the lowest-energy with CNS1.21 software [11,12] The 10 models with the lowest number of NOE violations were then refined in explicit water. To parameterize the bond lengths, dihedral angles and improper and non bonded interactions of cinaciguat the PRODRG Server was used [33].

Results

Homology modeling

An initial profile of H-NOX protein was made using bioinformatics tools. The human H-NOX sGC domain showed high homology with prokaryotic H-NOX proteins already determined and deposited in the protein data bank. So, the protein sequence was submitted into Phyre2 program [31] for sequential alignment. The conserved residues (Y2, P118 and YxSxR) involved in ligand coordination were identified at a to similar position along the primary structure of the H-NOX domain. In order to determine the empirical secondary structure, the human H-NOX sequence was loaded into SWISS MODEL program [29]. The result showed that the secondary structure of human H-NOX domain has higher similarity to *Ns* H-NOX, with around 40% similarity Figure 1A. In agreement with this result, the output of the 3D structure homology modeling also showed that both proteins have a similar tertiary structure arrangement (Figure 1B). The model was built based on the target (human H-NOX)-template (*Ns* H-NOX) alignment using Promod-II web tool [34]. Coordinates conserved between the target and the template were copied from the template to the model. Insertions and deletions were remodelled using a fragment library. Side chains were then rebuilt. Lastly, the geometry of the resulting model was regularized by using a force field. The heme ligand coordinates from the template structure were transferred through homology to the model, meeting the following criteria: (1) The ligand was annotated as biologically relevant in the template library, (2) the ligand was in contact with the model, (3) the ligand was not clashing with the protein, and (4) the residues in contact with the ligand were conserved between the target and the template [35]. In case of the human H-NOX domain model, the ligand information could not be transferred from the template probably due to a

criterion that was not accurately satisfied. The model of the human H-NOX sGC domain (Figure 1B) showed to be composed by seven α-helix (αA–αF) and four β-sheets (β1-β4). And the heme binding pocket was identified, fitting well with the *Ns* H-NOX crystal structure [36].

These preliminary studies were performed to obtain a structural model for subsequent structural human H-NOX domain characterization.
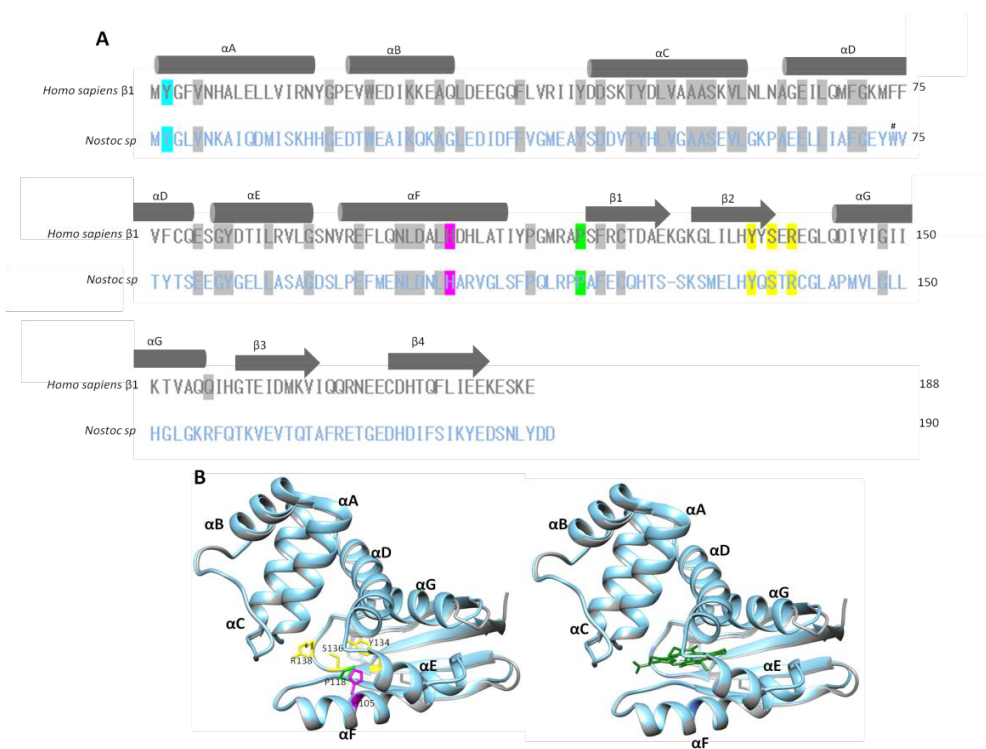
Figure 1: Secondary structure based on homology modelling using *Nostoc* sp H-NOX as template (35% of homology). A- All conserved residues including   hydrophobic core are shaded grey. The *Nostoc* sp H-NOX key residues for the transition from six- to five-coordinated NO bound are labelled with a '#' (W74 and M144). Adapted from SWISS MODEL workspace [29] output. **B-** Theoretical Human H-NOX domain 3D structure model from SWISS MODEL homology modeling program [29] output. The alignment was build using *Nostoc* sp H-NOX as template (PBD entry 2O09). Residues in the structural heme cavity are labelled in blue, pink, green and yellow, the latter if the residue is fully conserved involved in protein activation.

Human H-NOX_c assignment

The ^{1}H-^{15}N HSQC spectra of the WT human H-NOX domain (1-188) complexed with cinaciguat (H-NOX_c) exhibited good peak dispersion, suggesting that the protein was stably folded (Figure 2). From that, approximately 88% of backbone resonances were assigned, the exceptions being residues 85-88 and 105-112 likely arising from signal absence. The assignment was established for all the residues, including the 3 prolines and the first residue. The total extent of the assignment for the ^{1}H, ^{13}C and ^{15}N is 63.9, 72.14 and 76.8 %, respectively (Table 2). Four resonances displayed chemical shifts outside the known distribution of shifts differing to standard deviation, namely $H^{\delta 1\alpha}$ resonance of I42, $H^{\delta\alpha/\beta}$ resonance E10, $H^{\delta/\alpha/\beta}$ resonance K26 (more than 4-fold) and $H^{\delta\alpha}$ resonance of K26. In all these cases the assignment was verified to be correct, therefore, the deviation could be due to ring current effects from neighboring aromatic residues. Based on homology models most likely Y49 in the case of I42, W22 to E10, F37 and W22 to K26 and F112 in case of R40. Although, high percentage of aromatic residues could be assigned (83.3%), their side-chains resonances were not identified.

Figure 2: 2D of the HSQC spectra of WT H-NOX in complex with cinaciguat compound. The assignments are indicated by the *one-letter* amino acid code preceded by the corresponding sequence number along the protein primary sequence. The tryptophan (W) peaks in red color were highlight because in the human H-NOX sequence there is only one, but the second peaks (7.6 ppm) the resonance comes from $N^{\varepsilon 1}$. The peaks connected with horizontal lines are correlations from side-chain NH2 spin pairs belonging to ASN and GLN residues. The peaks indicated with a star, is unassigned/folded residues and peaks inside the contour level are represented below. The spectrum was acquired with at 800 MHz spectrometer equipped with a TXI probe, at 298.15 K.

Table 1: Report of the completeness of the Assignments of human H-NOX sGC domain resonances from residue 1 to 188.

Category	Available	Assigned	% Assigned
Element C	933	645	69
Element H	1134	681	60
Element N	254	172	67
Backbone	564	497	88
Residue Ala	11	10	90
Residue Arg	8	7	78,7
Residue Asn	7	7	100
Residue Asp	12	10	83
Residue Cys	3	3	100
Residue Gln	11	11	100
Residue Glu	19	19	100
Residue Gly	13	13	100
Residue His	6	4	66
Residue Ile	16	14	87,5
Residue Leu	19	16	84,2
Residue Lys	11	11	100
Residue Met	5	5	100
Residue Phe	9	99	100
Residue Pro	3	3	100
Residue Ser	7	7	100
Residue Thr	7	5	71,4
Residue Trp	1	1	100
Residue Tyr	8	6	75
Residue Val	12	11	91,6
All Residues	188	172	91,5
Spin Systems	195	195	100

Secondary structure assessment

The human H-NOX domain secondary structure could be estimated merely by experimental chemical information. It could be achieved from CA, CB, CO and HA chemical shifts values since they are sensitive to differences in secondary structure. Thus, the amino acid arrangement and disorder prediction were probed using

PECAN program, which are commonly accepted procedure to establish the secondary structure of proteins based on chemical shift differences with respect to some predefined standardized values that corresponds to specific secondary structure elements [32]. As shown in Figure 3, the program predicted the N-terminal moiety to be fully α-helix (αA–αF) (approximately seven) and 4 β-sheet (β1-β4) in the C-terminal protein portion [36].
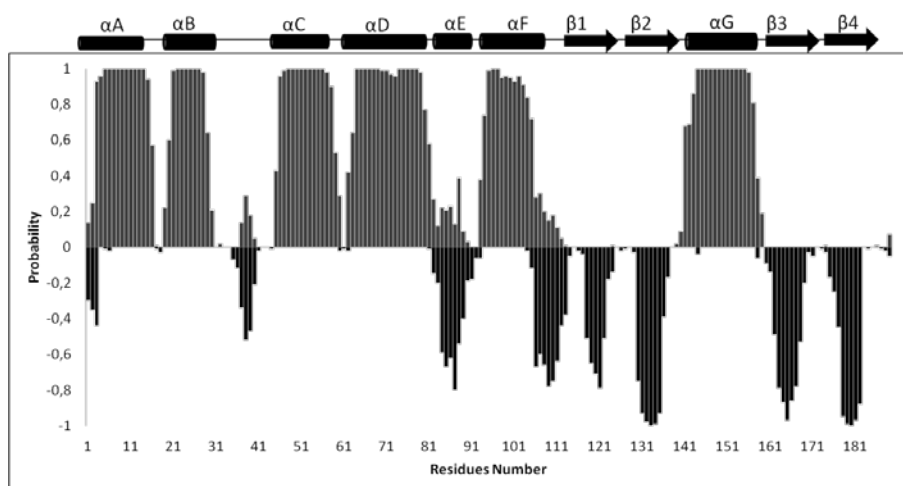
Figure 3: Probability of α-helical (grey bars) and β-sheet (black bars) secondary structure as predicted by PECAN [32]. The corresponding cartoon representation for human H-NOX domain (α-helical (A-G) and β-sheets (1-4) are shown as cube and arrows respectively).

H-NOX_c assignment transfer

A reliable H-NOX_c assignment was possible for 172 of the expected 188 resonances (188 residues minus three prolines residue, the amino terminal and twelve residues, which could not be seen in the ^{1}H-^{15}N HSQC). The H-NOX_c ^{1}H-^{15}N HSQC spectrum supported the assignment transferred (Figure 4A and B), in which 79 and 32% of the residues assigned could be transfer to H-NOX^{H105F}_c H-NOX_Zn respectively. To avoid ambiguity each H-NOX_c chemical shift resonance (^{15}N and ^{1}H) were accurately used as base to calculate the geometric distance from assigned peak in comparison to unassigned H-NOX^{H105F}_c and WT H-NOX_Zn complex. It was

determined based on a combined analysis of the difference value. This examination involves a single value for each amino acid (^{15}N ^{1}H) demonstrating their cumulative deviation from H-NOX_c to H-NOX^{H105F}_c and H-NOX_Zn behaviour. Although, the distance analysis result indicated a few populations deviate from H-NOX_c to H-NOX^{H105F}_c and H-NOX_Zn, these chemical shift differences were too small, proving that there were no ambiguities in the assignment transfer (Figure 5 A and B).
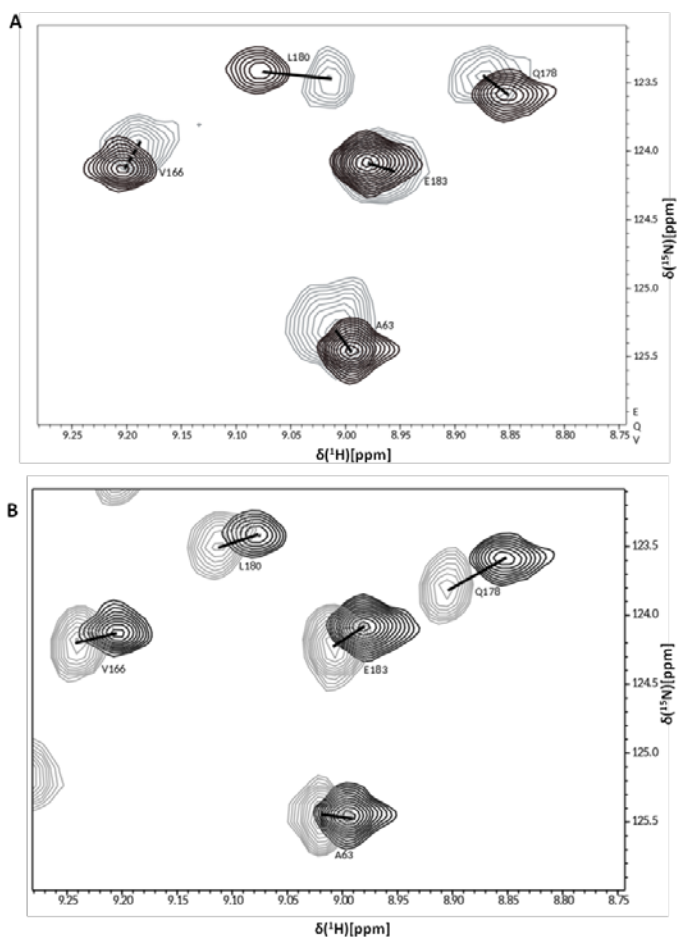
Figure 4: 2D of the HSQC spectra of the assignment transfer from WT H-NOX_c to H-NOX^{H105F}_c and H-NOX_Zn. **A-** ^{15}N (N-HSQC) spectrum overlay of the H-NOX_c (black peak) and mutant H-NOX^{H105F}_c (grey peak). **B-** ^{15}N HSQC spectrum overlay of the H-NOX_c (black peak) bound and H-NOX_Zn (grey peak).
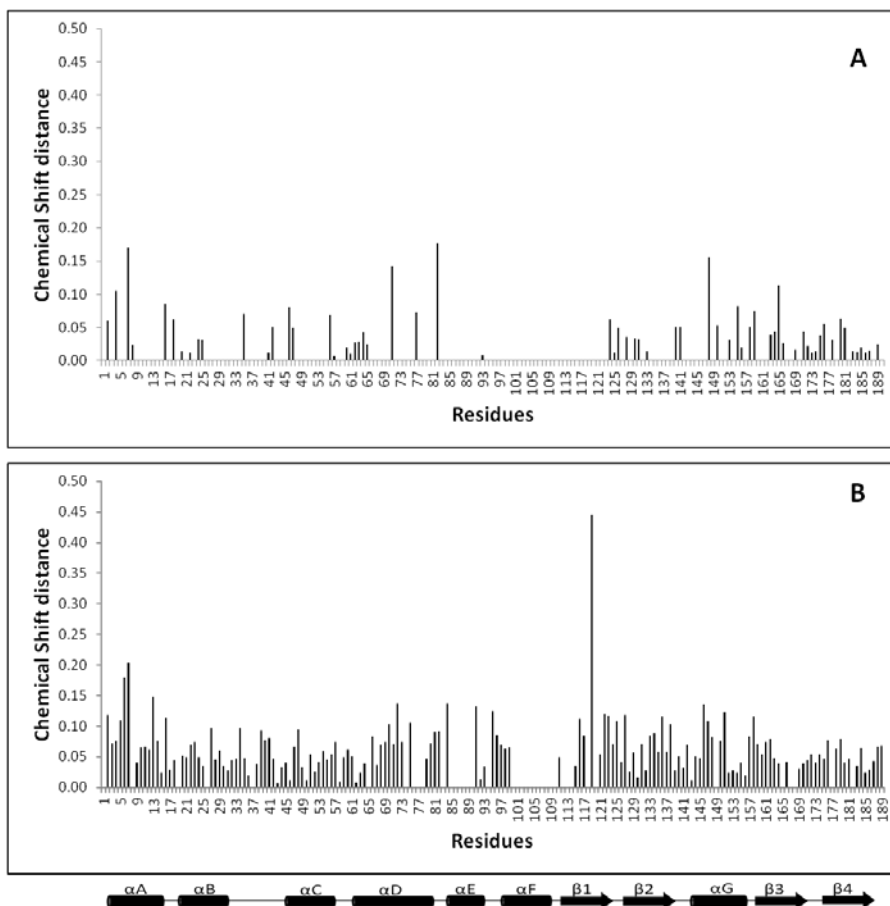
Figure 5: Chemical shift difference. A- Chemical shift difference after assignment transfer from H-NOX_c to H-NOX_Zn. **B-** Assignment transfer from H-NOX_c to H-NOX^{H105F}_c.

Human H-NOX domain structural determination by NMR

The structure determination of WT human H-NOX sGC domain complexed with cinaciguat was performed by NMR using NOE-derived distances and chemical shift-derived dihedral angle restraints. A report of structure statistics is shown in Table 3. From 100 calculated structures, the 10 structures (Figure 6A) which showed lowest energies were selected for further analysis with root-mean-square deviation (r.m.s.d) of distance violations 0.083 Å and r.m.s.d of dihedral violations 3.6 Å. The average r.m.s.d from the mean for this group of 10 lowest-energy

structure was 0.77 Å and the figure 5B showed the r.m.s.d value for each protein residues. Whereas, the α-helixes distal pocket (2-80 residues) showed average r.m.s.d of 1 Å, the β-sheet proximal pocket (81-180 residues) presented average r.m.s.d 1.8 Å value. The reason that proximal pocket showed higher r.m.s.d value (Figure 7A) is because of the low and/or lack of NOEs (Figure 7B). The lack of NOE is due to unassigned residue in this portion of the protein (85-88 105-112) resulting in this region of the protein being less well defined than distal pocket (Figure 5A).

Table 3: Structural statistics for the calculated ensembles of 10 low-energy structures after water refinement of the human H-NOX domain at pH 6.5 and 298K.

Number of restraints	WH-NOX
NOE-distance restraints	3219
Intra-residue	1070
Sequential	932
Medium range (1<|i-j|<5)	629
Long range (|i-j|>4)	588
Ambiguous	172
Protein-cinaciguat	28
Intra cinaciguat	21
H-bonds	5
Dihedral angle restraints	352
Restraint statistics	
rmsd of NOE violations	0.074±0.016
NOE violations per structure (>0.3 Å)	17.2±4.2
rmsd of H-bond violations	0.49±0.97
H-bond violations per structure (>0.3 Å)	0.33±0.75

rmsd of dihedral violations	2.659±0.317
Dihedral violations per structure (>5°)	20.0±3.4
Structure rmsd	0.77±0.18

Structural quality

Ramachandran (PROCHECK)

Most favored regions	79.6%
Allowed regions	16.8%
Disallowed regions	3.6%

WHATIF Z-scores

Backbone conformation	-2.600±0.451
2nd generation packing quality	-4.867±2.027
Ramachandran plot appearance	-1.627±0.516
χ^1/χ^2 rotamer quality	-4.009±0.508

CING ROG analysis

Red	8%
Orange	41%
Green	51%
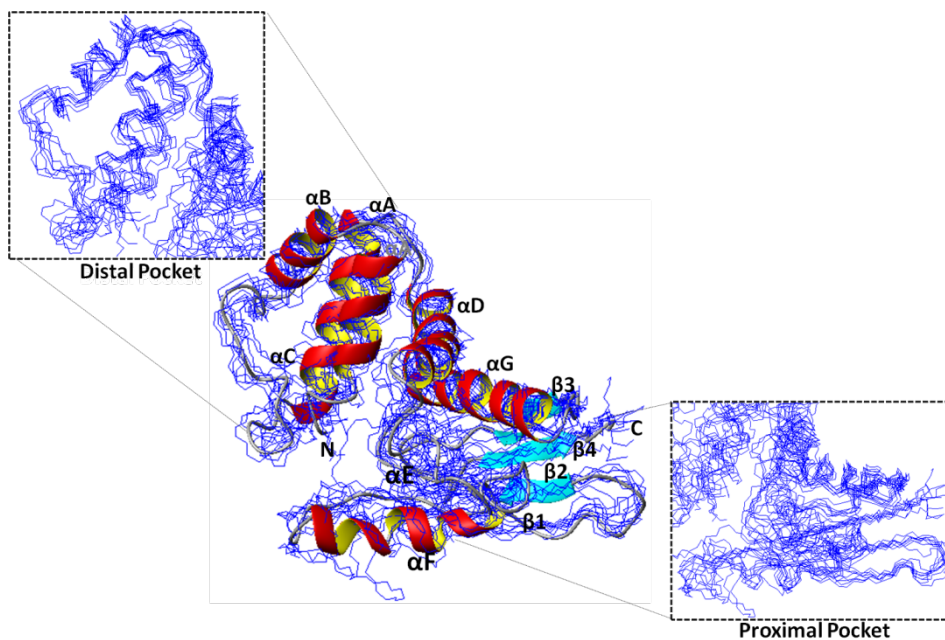
Figure 6: 3D structure model of the human H-NOX sGC domain complexed with cinaciguat compound (not shown). The overlay is the 10 lowest-energy structure model and the backbone atoms of residues 2-80 and 81-180 is the distal and proximal pocket respectively were aligned in MOLMOL program [37].

A

B
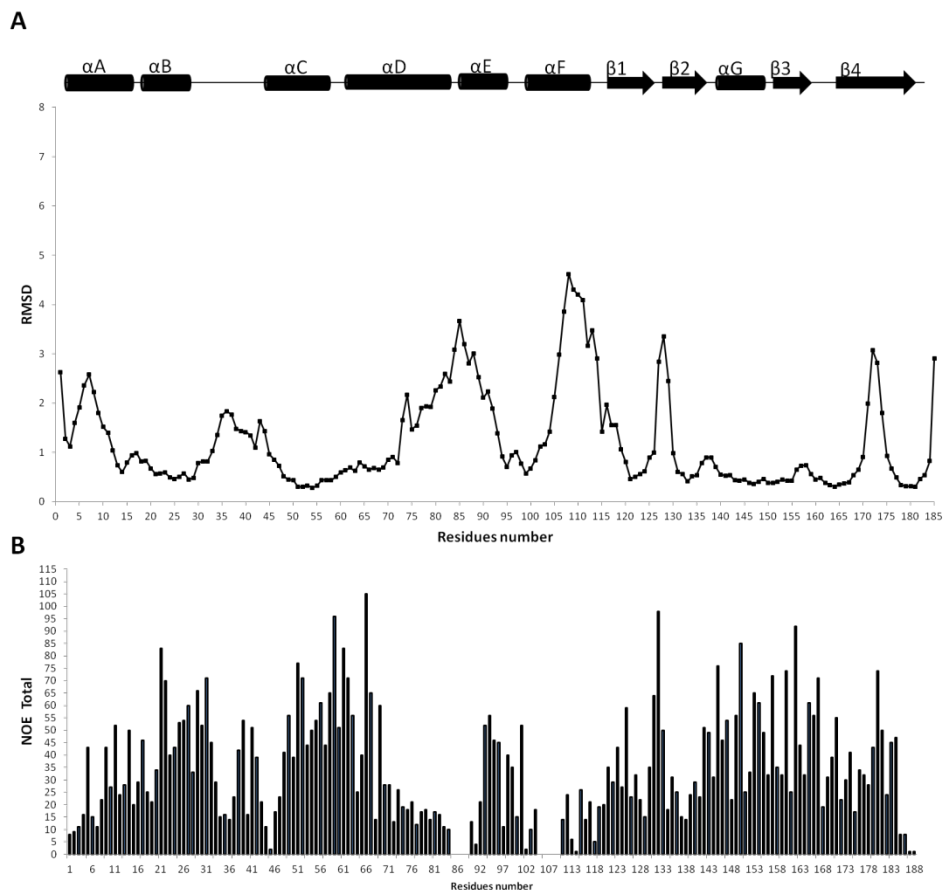
Figure 7: Experimental NOE values evaluation. A-Backbone RMSDs between the lowest energy structures of H-NOX_c. **B-** Experimental NOE values calculated from each amino acid (1-188) including the unassigned residues.

Evaluation of the protein ligand NOEs showed that all the aromatic rings of the ligand had NOEs to the protein. In addition the aliphatic groups linking the phenyl rings of the long chain, also exhibited some NOEs. In contrast the aliphatic chain that connects the carboxylate to the rest of the system did not show any NOEs, a fact that suggests that it is exhibiting increased mobility and/or is relatively delocalized. The residues of the heme cavity that had contacts with the ligand are: M1, F4, V5, V39, G82, L101, L104, M115, A117, V146, I149, I150 and V153 (Figure 8). Cinaciguat has already been reported to be able to occupy the heme cavity with

its carboxylate groups superimposing on these of the heme propionate groups in the apo-H-NOX structure [38]. Despite this compound having a shape dissimilar to heme, our structural analysis showed it was in a similar position as previously found to *Ns* H-NOX-cinaciguat bound [38]. It is a indication that cinaciguat mimics heme by two hydrophilic charged carboxylates that interact with YxSxR motif found to the key also for heme and cinaciguat binding (Figure 9) [39,40].
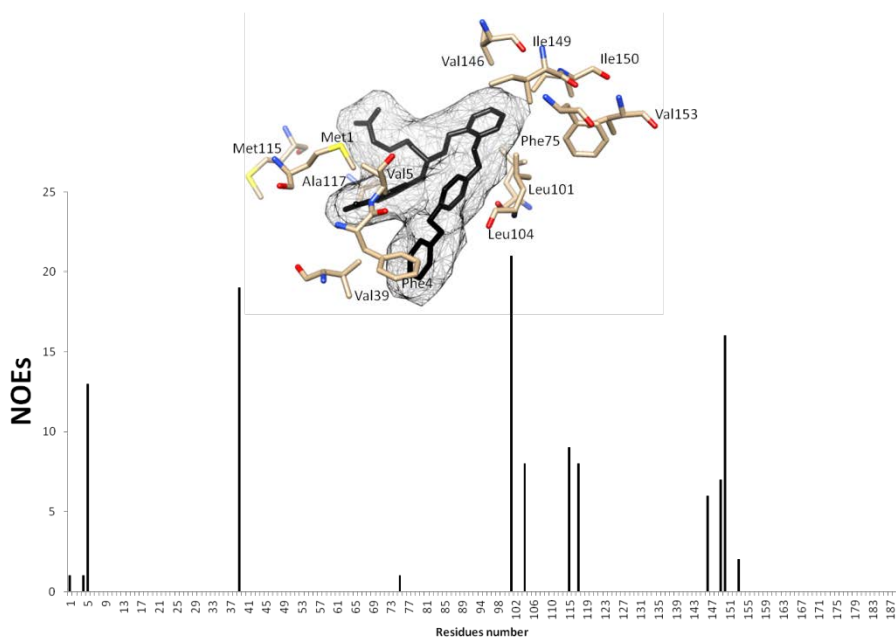
Figure 8: Inter-NOE values from protein-ligand. The interaction showed the cinaciguat compound with nearby H-NOX residues.
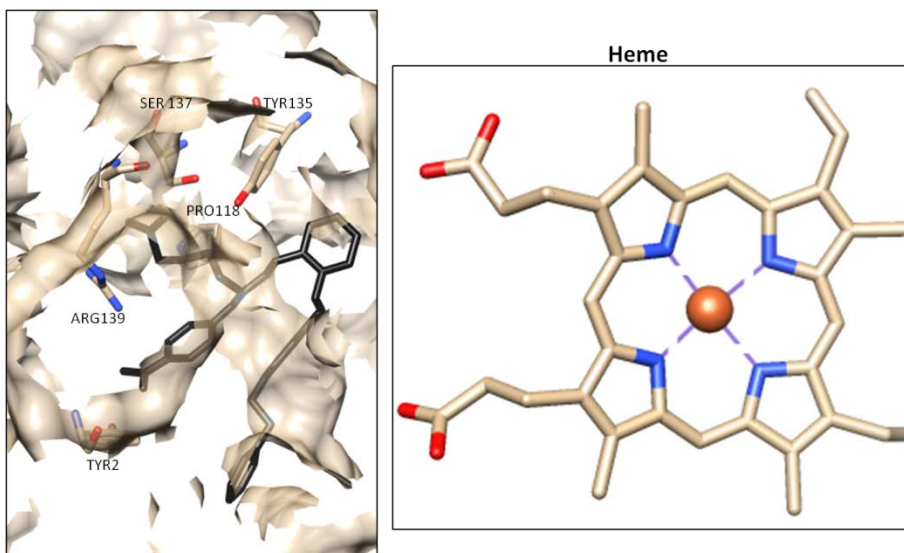
Figure 9: Cinaciguat bound to H-NOX heme cavity and mimics heme/protein interactions. It indicated that hydrogen bound interactions by carboxylates of cinaciguat with the YxSxR motif residues.

The quality of the human H-NOX_c structures was evaluated using the PROCHECK-NMR program [41]. It indicated that 79.6% of residue have phi (φ) and psi (ψ) dihedral angles in the most favoured region of the Ramachandran plot [42] and 16.8% of the residues have phi and psi angles in the allowed regions (Supplementary material). The overall topology of the human H-NOX domain is highly similar to those of the H-NOX domains from prokaryotic *Nostoc sp* heme and/or cinaciguat bound [38], *T. tengcongensis*, and *Shewanella oneidensis* complexed with heme [4,36,43] that have been solved with X-ray and NMR. In addition the topology of the seven α-helices and four β-sheet agrees with those reported (Figure 10).
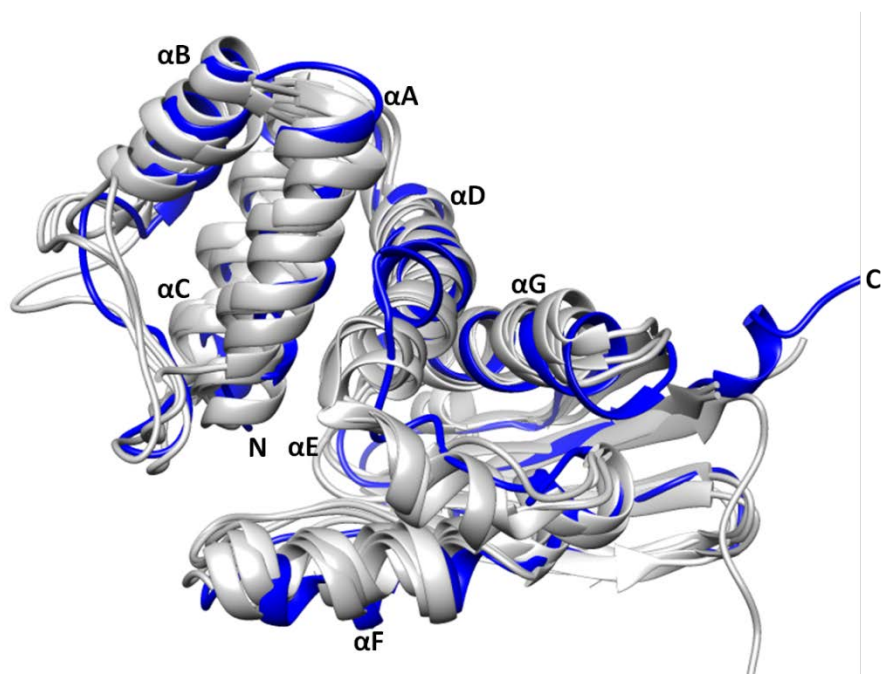
Figure 10: 3D H-NOX structure. Comparison between our structure of human H-NOX_c (in blue) and *Nostoc sp* heme and cinaciguat bound to H-NOX heme (PDB entry 2o09 and 4JQH respectively), *T. Tengcongensis* O_2 unligated (PDB entry 3TF0), *Shewanella oneidensis* complexed with heme-CO (PDB entry 2Kii).

Discussion

Human H-NOX domain contains several conserved residues as suggested by sequence alignment with the several homologous H-NOX proteins. These are Y2, P118, and the Yx,SxR sequence, shown in

Figure 1 [36,44]. The structures of H-NOX sGC homologous (*Ns* H-NOX) complexed with heme and/or cinaciguat have already been solved and the conserved residues were found to be a key for heme/cinaciguat binding to the protein [36,38]. The functional involvement of proline (118) in full sGC activation has been investigated extensively using prokaryote *T. tengcongensis*, and *Shewanella oneidensis* by an X-Ray and NMR approach [4,45]. Recently, this conserved residue was reported to be involved in the porphyrin relaxation toward planarity upon displacement of the axial histidine (H105) from iron caused by heme-NO binding [44]. The first step for the investigation of the significance of these residues in our case is to identify their resonances. All the prolines and serines residues were assigned. High percentage of assignment of arginines (88%) and tyrosines (75%) was also achieved, including all the above mentioned conserved residues (Table 1), the key residue H105 that is involved in heme binding and whole Fe^{2+}–H105 bond rupture that stimulates sGC activation [44] could not be identified. The unassigned regions from residue 85 to 88 (ILRV) and from H105 to Y112 (HLATIYPG) showed to be part of the αF and αE helices in structural model, which are regions very close in space to the heme pocket (
Figure **1**B).

Erbil and collaborators (2009) determined the *So2144* H-NOX structure by NMR and although, in their case, the majority of resonances were assigned, residues 2,20,35,48 (KSVF), 110-112 (YHE) and 114 (S) [4] could not be identified. In our case, the majority of unassigned residues are in portions close to the ligand cavity and may be broadened beyond detection because of conformational

exchange. In fact, conformational exchange may be the reason for the absence of these residues in the ^{1}H-^{15}N HSQC spectra.

Chemical shifts have been extensively correlated with protein secondary structure. For example, the average Cα chemical shifts in α-helices, have a propensity to be higher than those in a random coil, while the average Cα in β-strands are lower. The Cβ atoms behaviour is exactly opposite to Cα atoms, where β-strands have essentially higher signals and in α-helices mostly slower ones [16]. Through chemical shift indexing [49], the patterns observed along the protein sequence give a reliable indication of α-helix or β-strand features [16,50]. If these trends are further refined to take into account the chemical nature of the neighbouring residues it is even possible to correlate chemical shifts with the φ and ψ torsion angles of the Ramachandran plot.[42]. To evaluate the secondary structure of the H-NOX domain, experimental Cα, Cβ, CO and Hα chemical shifts values retrieved from NMR data analysis were used with the program PECAN [32]. The prediction suggests that even though the N-terminal part of the protein has a high propensity for α-helix (αA–αF) secondary features and the C-terminal for β-sheet (β1-β4) features, the data shows to be ambiguous in the 81-91 and 101-111 portion (Figure 3). However, the DANGLE program also predicted the secondary structure based on dihedral angles restrains, and predictions based on the secondary structure of the H-NOX sGC homologous structure (Figure 1A) [36], showed that these two regions belong to αE and αF respectively.

Regarding the three-dimensional protein structure determination, we were able to use all the NOE and dihedral angles to calculate a bundle of the 10 lowest-energy structure of cinaciguat bound to human H-NOX. (Figure 6). The weighed r.m.s.d of the structure calculated with the algorithm of CCPN was 0.77 Å. The structure reaffirmed the ability of cinaciguat to bind into the cavity of heme-depleted sGC, by mimicking heme, without requiring a covalent bond with H105 residue. Not all portions of the protein structure had the same quality however.

The difference between the r.m.s.d values from postulated distal pocket (2-80 residues) and proximal pocket (81-180 residues) was 0.8 Å. This value is significant enough to make the α-helical portion more clearly defined when compared with rich in β-sheet domain figure 6). A low number of NOE values identified for the proximal pocket has negatively affected the r.m.s.d, leading to an uncertainty during the protein structure calculation (Figure 7A and B). Fe^{2+}-CO bound WT and H-NOX[H105G] structures from *Shewanella oneidensis* were also resolved by NMR, resulting in bundles with an r.m.s.d of 0.33 and 0.37 Å for the backbone, respectively [4]. By looking at the 20 lowest-energy structures obtained [4], the proximal pocket showed to be better defined than the structure determined by our group. This could be partially be due to the heme-derived residual dipolar coupling (RDCs) data, which were collected for their work, which were absent from our structure calculation. RDCs give information about individual vector alignment relative to the external magnetic field [51,52]. In other words, this information is about relative orientation of parts of the molecule that are far apart in the protein structure [53]. In addition, heme which is the natural ligand of H-NOX, it is expected to stabilize the protein better than the very flexible cinaciguat that has a large number of degrees of freedom. In relatively large proteins, such as H-NOX domain (22.3 kDa) some portion may have been difficult to record NOEs due to spin diffusion, which the individual nuclear spin, undergo continuous exchange of energy. It allows different polarization levels within the sample to be reduced on a timescale much shorter than the relaxation effect [54]. However, it is not a problem with RDCs [51,55]. Therefore, to improve the structure of the human cinaciguat bound to H-NOX, collection of derived residual dipolar coupling (RDCs) data should be included as input for structure determination. Our efforts to collect RDCs using stretched polyacrylamide gels resulting in spectra with very low signal intensity and they could not be evaluated reliably.

Important insight on structural changes of H-NOX_c was obtained from NMR experimental data. For instance, the inter-NOE evaluation indicated that the third cinaciguat branch possibly makes hydrophobic interactions with residues, occupying the remaining heme cavity including M1, F4, V5, V39, G82, L101, L104, M115, A117, V146, I149, I150 and V153 (Figure 8). Martin *et al*, (2010) also found by X-Ray crystallography that hydrophobic residues from *Nostoc sp* H-NOX interact with cinaciguat [38]. The sequence alignment between human and *Nostoc sp* H-NOX showed that some of these amino acids, such as L101 and L104, were conserved. Nevertheless, due to the limitation in collecting structural data in some part of the protein, especially in the ligand cavity, our structure has restricted information about cinaciguat interactions with the H-NOX domain. Even so, we have an indication that cinaciguat mimics heme by two hydrophilic charged carboxylates that interact with YxSxR motif (Figure 9) as also described by several H-NOX structural studies [4,36,38-40,43-45]. In the absence of higher quality NOE data we have a restriction for an in depth atomic structural analysis. On the other hand, the quality of human H-NOX-cinaciguat structure was evaluated using PROCHECK-NMR program [41], and showed high percentage of the residues in agreement with Ramachandran plot [42]. Our human H-NOX-cinaciguat structure agree well with the most closer in homology (18-40%) prokaryotic H-NOX. In conclusion, despite the absence of some residues from the α-F and E regions, we were able to unambiguously differentiate both helices, as we could observe from the overlay structure using the prokaryotic H-NOX homologous structure and the human sGC heme domain (Figure 10). It should be noted that the quality of the NOE data were better in some parts of the protein than others. This implies that the reasons for the bad quality is not related to data quality but more likely it is linked to increased dynamics and especially conformational exchange due to the flexibility of the ligand. This hypothesis was further tested in the next chapter where the analysis of relaxation data was performed.

For relaxation data collection, samples of the H-NOX[H105F]_c and H-NOX_Zn labelled with ^{15}N were also prepared as described in chapter 2. In the absence of a doubly labelled protein sample with [13]C-^{15}N for assignment, each chemical shift resonance (^{15}N and ^{1}H) from H-NOX_c were accurately used as a support to calculate the geometric distance between H-NOX_c peak and both the unassigned H-NOX[H105F]_c and H-NOX_Z. As a result, the majority of peaks showed few deviations of the signal populations from H-NOX_c to the H-NOX[H105F]_c and H-NOX_Zn. Thus, due to small geometric distances between them, there were no ambiguities in the assignment transfer (Figure 4and 5). The most relevant difference in terms of chemical shift deviation was observed in residue Ser119 transfer from H-NOX_c to the H-NOX[H105F]_c. It could be due to the mutation in a crucial residue involved in heme binding and Fe^{2+}–H105 creating a bond rupture. So, the replacement of histidine by phenylalanine may be responsible for a chemical shift deviation in the neighbouring residue. Still, the shift difference of S119 (β-sheet 1) was below 0.5ppm, being small enough, thereby guaranteeing a reliable assignment transfer (Figure 5 A and B).

The overlay of ^{1}H-^{15}N HSQC H-NOX_c and H-NOX_Zn spectra show certain ambiguity in several resonance signals. Therefore, only 32% of the peaks could be transferred with high degree of reliability (Figure 4B). Whereas the structure of ZnPPIX (inhibitor) is analogous to the heme structure [46], cinaciguat compound (activator) was designed to mimic heme and work independently of NO binding. Therefore, they share a similar chemical features but different structural conformation as described in the chapter 1 [38,47,48]. Thus, the resonances observed in the ^{15}N HSQC spectrum from H-NOX complexed with these two different ligands were slightly different, but enough to create high ambiguities during the assignment transfer procedure. Hence, double labeled ^{13}C/^{15}N of H-NOX_Zn sample is required to perform full protein assignment.

References

1. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, et al. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181: 662-666.
2. Strandberg B (2009) Chapter 1: building the ground for the first two protein structures: myoglobin and haemoglobin. J Mol Biol 392: 2-10.
3. Wuthrich K (1990) Protein structure determination in solution by NMR spectroscopy. J Biol Chem 265: 22059-22062.
4. Erbil WK, Price MS, Wemmer DE, Marletta MA (2009) A structural basis for H-NOX signaling in Shewanella oneidensis by trapping a histidine kinase inhibitory conformation. Proc Natl Acad Sci U S A 106: 19753-19760.
5. Leopold MF, Urbauer JL, Wand AJ (1994) Resonance assignment strategies for the analysis of NMR spectra of proteins. Mol Biotechnol 2: 61-93.
6. Teng Q (2005) Structural Biology Practical NMR Application; Springer, editor. United States of America. 295 p.
7. Powers R, Garrett DS, March CJ, Frieden EA, Gronenborn AM, et al. (1992) 1H, 15N, 13C, and 13CO assignments of human interleukin-4 using three-dimensional double- and triple-resonance heteronuclear magnetic resonance spectroscopy. Biochemistry 31: 4334-4346.
8. Delaglio F, Torchia DA, Bax A (1991) Measurement of 15N-13C J couplings in staphylococcal nuclease. J Biomol NMR 1: 439-446.
9. Wüthrich K (1986) NMR of Proteins and Nucleic Acids; Series BL, editor. 320 pages p.
10. Webb GA (2001) Nuclear Magnetic Resonance; Chemistry RSo, editor.
11. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. Nat Protoc 2: 2728-2733.
12. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54: 905-921.
13. Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13: 289-302.
14. Wishart DS, Case DA (2001) Use of chemical shifts in macromolecular structure determination. Methods Enzymol 338: 3-34.
15. Mielke SP, Krishnan VV (2009) Characterization of protein secondary structure from NMR chemical shifts. Prog Nucl Magn Reson Spectrosc 54: 141-165.
16. Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry 31: 1647-1651.

17. Schwarzinger S, Kroon GJ, Foss TR, Wright PE, Dyson HJ (2000) Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. J Biomol NMR 18: 43-48.

18. Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, et al. (2001) Sequence-dependent correction of random coil NMR chemical shifts. J Am Chem Soc 123: 2970-2978.

19. Cheung M-S, Maguire ML, Stevens TJ, Broadhurst RW (2010) DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. Journal of magnetic resonance 202: 223-233.

20. Karplus M (1959) Contact Electron-Spin Coupling of Nuclear Magnetic Moments. The Journal of Chemical Physics 30: 11-15.

21. Osapay K, Theriault Y, Wright PE, Case DA (1994) Solution structure of carbonmonoxy myoglobin determined from nuclear magnetic resonance distance and chemical shift constraints. J Mol Biol 244: 183-197.

22. Clore GM, Gronenborn AM (1998) New methods of structure refinement for macromolecular structure determination by NMR. Proc Natl Acad Sci U S A 95: 5891-5898.

23. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. Proc Natl Acad Sci U S A 104: 9615-9620.

24. Ziarek JJ, Peterson FC, Lytle BL, Volkman BF (2011) Binding site identification and structure determination of protein-ligand complexes by NMR. Methods Enzymol 493: 241-275.

25. Linge JP, Habeck M, Rieping W, Nilges M (2003) ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics 19: 315-316.

26. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59: 687-696.

27. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M (2003) Refinement of protein structures in explicit solvent. Proteins 50: 496-506.

28. Vranken W, Vuister G, Bonvin AJJ (2015) NMR-Based Modeling and Refinement of Protein 3D Structures. In: Kukol A, editor. Molecular Modeling of Proteins: Springer New York. pp. 351-380.

29. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22: 195-201.

30. Ramelot TA, Raman S, Kuzin AP, Xiao R, Ma LC, et al. (2009) Improving NMR protein structure quality by Rosetta refinement: a molecular replacement study. Proteins 75: 147-167.

31. Nema V, Pal SK (2013) Exploration of freely available web-interfaces for comparative homology modelling of microbial proteins. Bioinformation 9: 796-801.

32. Eghbalnia HR, Wang L, Bahrami A, Assadi A, Markley JL (2005) Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. J Biomol NMR 32: 71-81.

33. Schuttelkopf AW, van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr 60: 1355-1363.

34. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18: 2714-2723.

35. Gallo Cassarino T, Bordoli L, Schwede T (2014) Assessment of ligand binding site predictions in CASP10. Proteins 82 Suppl 2: 154-163.

36. Ma X, Sayed N, Beuve A, van den Akker F (2007) NO and CO differentially activate soluble guanylyl cyclase via a heme pivot-bend mechanism. Embo j 26: 578-588.

37. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14: 51-55, 29-32.

38. Martin F, Baskaran P, Ma X, Dunten PW, Schaefer M, et al. (2010) Structure of cinaciguat (BAY 58-2667) bound to Nostoc H-NOX domain reveals insights into heme-mimetic activation of the soluble guanylyl cyclase. J Biol Chem 285: 22651-22657.

39. Schmidt PM, Rothkegel C, Wunder F, Schroder H, Stasch JP (2005) Residues stabilizing the heme moiety of the nitric oxide sensor soluble guanylate cyclase. Eur J Pharmacol 513: 67-74.

40. Schmidt PM, Schramm M, Schroder H, Wunder F, Stasch JP (2004) Identification of residues crucially involved in the binding of the heme moiety of soluble guanylate cyclase. J Biol Chem 279: 3025-3032.

41. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 26: 283-291.

42. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7: 95-99.

43. Pellicena P, Karow DS, Boon EM, Marletta MA, Kuriyan J (2004) Crystal structure of an oxygen-binding heme domain related to soluble guanylate cyclases. Proc Natl Acad Sci U S A 101: 12854-12859.

44. Herzik MA, Jr., Jonnalagadda R, Kuriyan J, Marletta MA (2014) Structural insights into the role of iron-histidine bond cleavage in nitric oxide-induced activation of H-NOX gas sensor proteins. Proc Natl Acad Sci U S A 111: E4156-4164.

45. Olea C, Jr., Herzik MA, Jr., Kuriyan J, Marletta MA (2010) Structural insights into the molecular mechanism of H-NOX activation. Protein Sci 19: 881-887.

46. Serfass L, Burstyn JN (1998) Effect of heme oxygenase inhibitors on soluble guanylyl cyclase activity. Arch Biochem Biophys 359: 8-16.

47. Evora PR, Evora PM, Celotto AC, Rodrigues AJ, Joviliano EE (2012) Cardiovascular therapeutics targets on the NO-sGC-cGMP signaling pathway: a critical overview. Curr Drug Targets 13: 1207-1214.

48. Boerrigter G, Costello-Boerrigter L, Lapp H, Stasch J-P, Burnett J (2005) Co-activation of soluble and particulate guanylate cyclase by BAY 58-2667 and BNP enhances cardiorenal function in experimental heart failure. BMC Pharmacology 5: P5.

49. Fruchart JS, Lippens G, Warrass R, Seetharaman C, Dhalluin C, et al. (2000) The Chemical Shift Index method applied to resin-bound peptides. J Pept Res 56: 346-351.

50. Wishart DS, Sykes BD (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 4: 171-180.

51. Clore GM, Schwieters CD (2006) Concordance of Residual Dipolar Couplings, Backbone Order Parameters and Crystallographic B-factors for a Small α/β Protein: A Unified Picture of High Probability, Fast Atomic Motions in Proteins. Journal of Molecular Biology 355: 879-886.

52. Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 12: 1-16.

53. Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. Chem Rev 104: 3519-3540.

54. Suter D, Ernst RR (1985) Spin diffusion in resolved solid-state NMR spectra. Phys Rev B Condens Matter 32: 5608-5627.

55. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A (1997) Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Biol 4: 732-738.

Supplemental Data
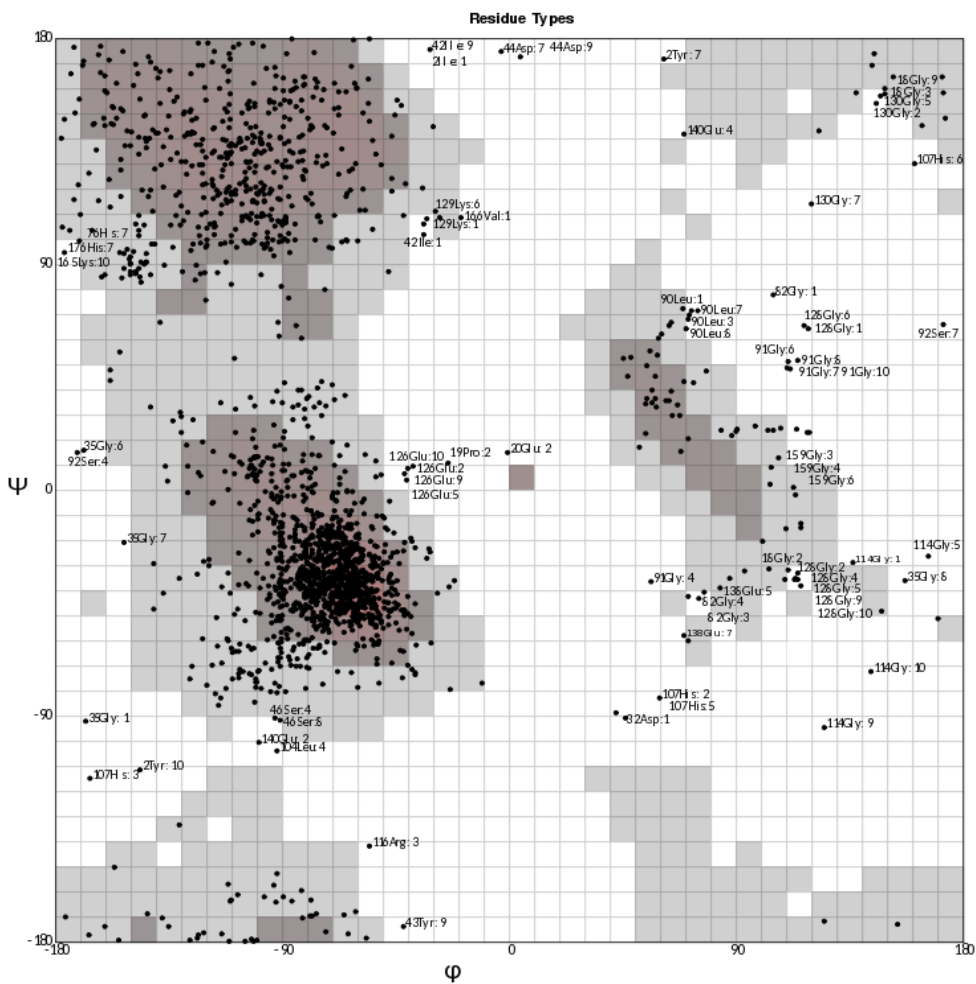
S1: Ramachandran plot. H-NOX_c shows the backbone dihedral angles ψ against φ of amino acid residues in protein structure.

Chapter 4

Dynamics studies of the human sGC H-NOX domain

The NMR protein dynamics data collections were performed by thesis supervisor Dr. Manolis Matzapetakis and Dr. Ivo Saraiva. The analysis of the data was done by the author of this dissertation.

Table of Contents

Introduction

1. General introduction

The spin ½ isotopes ^{1}H, ^{13}C, ^{15}N and ^{31}P and spin 1 isotope ^{2}H can be explored by NMR to provide many site-specific probes of local structure and dynamics [1]. The connection of NMR and dynamics is made via the study of relaxation. Relaxation is described as the process of the return of the magnetization to equilibrium after it has been perturbed. Small perturbations in the local magnetic fields of each nuclei affect the rate of relaxation. In proteins, these perturbations are influenced by both the placement, and the mobility of local structure elements such as the chemical nature of each residue (bond nature and torsion angle geometry) and its environment which is formed by side chains and solutes, (hydrogen bonding, ring current effects) [2].

1.1 Nuclear spin relaxation

Nuclear spin relaxation (NSR) can be affected by events in the ps-ns and ms timescales. These events are collectively described as dynamics and involve motions of different magnitudes. Figure 1 summarizes the several physical processes that might occur in this time window, such as: side chain rotamer interconversion, random coil and loop motions and backbone torsion angle rotations[2].
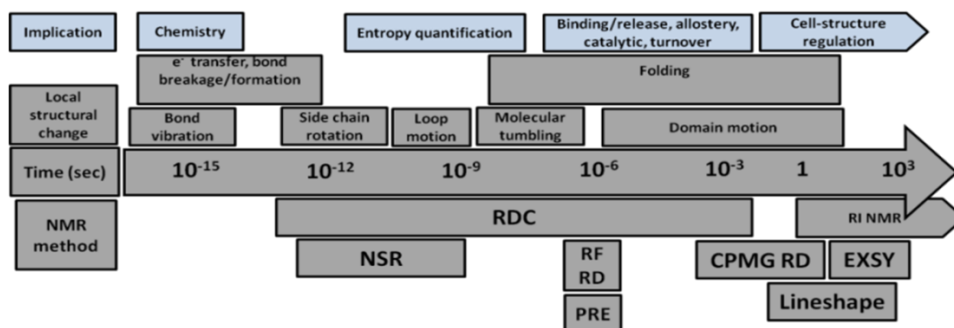
Figure 1: Timescales able to monitor protein conformational changes under their biological function. These dynamic processes can be studied with different NMR methods: (1) Real Time NMR, RT NMR; (2) Exchange Spectroscopy, EXSY (also known as zz-exchange); (3) Lineshape analysis; (4) Carr-Purcell Meiboom-Gill Relaxation Dispersion, CPMG RD; (5) Rotating Frame Relaxation Dispersion, RF RD; (6) Nuclear Spin Relaxation, NSR; (7) Residual Dipolar Coupling, RDC; (8) Paramagnetic Relaxation Enhancement, PRE. Adapted from [2,4-6].

These processes can lead to time dependent loss of signal, or relaxation that can be sensed in various ways. Most commonly it is probed by changes in rates of relaxation and the efficiency of through-space magnetization transfer[4-6]. Two rates of relaxation are used to describe the return to equilibrium of nuclear spins after perturbation, the longitudinal $R1$ (1/T1) and transverse $R2$ (1/T2) relaxation. Longitudinal relaxation refers to the "recovery" of the portion of the net magnetization that is aligned with the applied field B_0 which is commonly described as being along the z axis (Figure 2A). The loss of magnetization along the x,y plane, also called Transverse (x,y) relaxation, is more complicated and is dominated by loss of spin coherence which is not necessarily equivalent to the return to equilibrium (Figure 2 B). The R2 rate is always higher than R1, because after the return of magnetization to the z axis also means that no magnetization is left in the x-y plane.

The through-space magnetization transfer is probed using the *heteronuclear NOE* effect which results from through-space transfer via dipolar coupling between

different types of nuclei. Experimentally, this typically involves transfer from ^{15}N or ^{13}C nucleus to a directly attached ^{1}H [3].
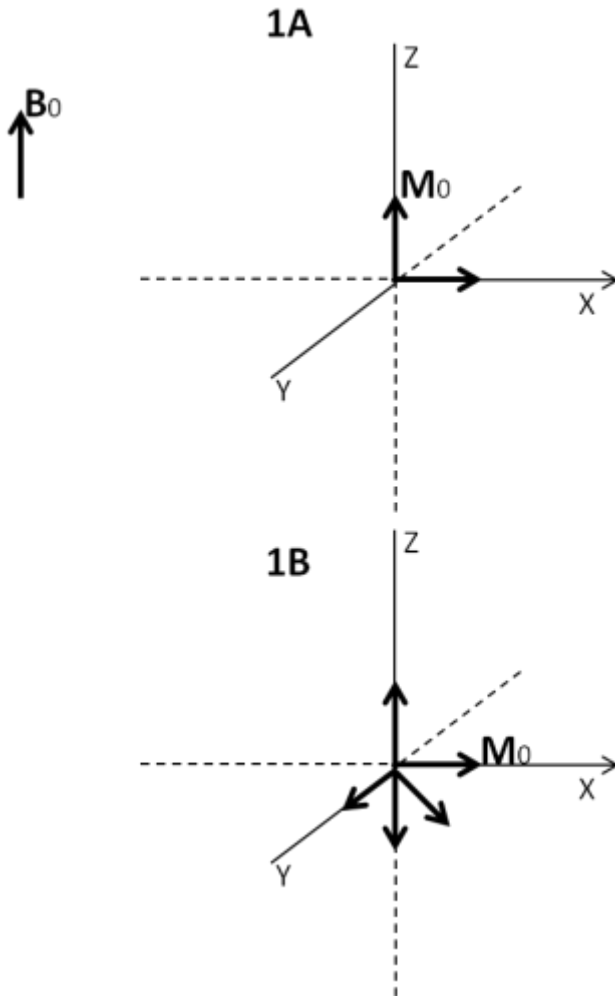
Figure 2: At equilibrium, the sample magnetization M0 is linear with static field B0. A-Upon 90° pulse, M0 lies in the xy plane. In T1 (spin-lattice relaxation) occurs the recovery of the magnetization along the z axis to the equilibrium value of M0. B: In T2 (spin-lattice) occurs the decay of transverse (xy plane) magnetization, based on rotating frame for convenience (relaxation). [3].

1.2   Relaxation measurement methods.

In order to measure the different relaxation rates, different experimental setups are needed for each of R1 and R2. In both cases however, a series of data are collected as functions of specific delays between the application of pulses and the acquisition of signal. The decay in both cases is exponential and the rate is the constant of the exponential functions used to fit the decaying signal.

R1 is measured with the inversion recovery experiment. There, initially the magnetization is stored along the −z axis and after the variable delay periods it is transferred to the xy plane to be measured. The R1 rate is obtained by fitting the time-dependent exponential restoration of intensity using the formula *I(t) = I(0)·(1 − exp(−R_1t))*. The values of in proteins R_1 range from 0.5-5 s^{-1}.

The measurement of R2 is inherently more complicated than R1 because while the magnetization lies along the xy plane it is affected by chemical shift and field inhomogeneities that cause the signal to disperse. These effects can be suppressed using the CPMG sequences (Carr and Purcell Meiboom and Gill) [7] where series of spin echoes continuously refocus magnetization. Different delays are then achieved with increasing number of repletion of these refocusing blocks. R2 is then quantified by a time-dependent exponential reduction of intensity using the formula *I(t) = I(0)·exp(−R_2t)*. In proteins R_2 values are ranging from 5-50 s^{-1}.

For the measurement of heteronuclear {[1]H}-^{15}N NOE enhancement an interleaved experiment is used. Two spectra are collected where protons are saturated for 3-4 seconds prior to the first ^{15}N pulse and in the other not. In these experiments a long recycle delay is required to ensure complete relaxation of water magnetization at the beginning of each scan [3]. The ratio of the two resulting

spectra gives the NOE enhancement. Since the gyromagnetic ratio of ^{15}N is negative, this {^{1}H}-^{15}N enhancement is negative resulting in values smaller than 1.

1.3 Nuclear spin relaxation data interpretation

The relaxation rates R*1*, R*2* and hnNOE are sensitive to random magnetic field oscillations at specific frequencies. These oscillations are caused by random molecular motions, either overall rotational tumbling or local dynamics. The correlation between relaxation and dynamics of a molecule is given through the spectral density function J(w), a function that quantifies these random motions at different frequencies.

Interpretation of nuclear relaxation rates is framed using one of at least four distinct approaches: phenomenological interpretation of site-specific R*1*, R*2* and hnNOE, spectral density mapping, modelfree analysis, or a specific model of internal motions that assumes details about the structure and interactions at the atomic level [9-11].

Phenomenological interpretation of R*1*, R*2* and hnNOE requires no additional processing beyond measuring these site-specific observables. If the protein is internally rigid, then the R*1*, R*2* and hnNOE values will be the same at each site; therefore, values far from the average reflect site-specific internal motions. In general, for proteins, R*1* increases with faster tumbling, lower magnetic field and can increase or decrease with local mobility. R*2* increases with slower tumbling, μs-ms chemical exchange R$_{ex}$, inhomogenity in the static magnetic field and higher magnetic field and decreases with local mobility. For globular, rigid proteins a spherical approximation can be used and the R2/R1 ratio can be used for the calculation of the rotational correlation time (τ_c) using the Stoke's-Einstein law [12]:

$$\tau_c = \frac{4\pi\eta r^3}{3kT} \qquad\qquad \text{Eq. 1}$$

where η is the viscosity of the solvent, r is the effective hydrodynamic radius of the protein molecule, k is the Boltzmann constant and T is the temperature. A closed-form solution for τ_c as a function of the ratio of the longitudinal (R1) and transverse (R2) ^{15}N relaxation rates can then be determined using Equation 2 (a simplified form of Eq. 8 from [13]), where v_N is the ^{15}N Larmor frequency in Hz.

$$\tau_c = \frac{1}{4\pi v_N}\sqrt{6\frac{R2}{R1} - 7} \qquad\qquad \text{Eq. 2}$$

The hnNOE ranges from −4 to 1 for {^{1}H}-^{15}N and from 1 to 3 for {^{1}H}-^{13}C and both are reduced in the presence of internal flexibility.

In spectral density mapping and its more commonly used "reduced" counterpart assumptions based on the usual rotational diffusion coefficients of proteins as well as common magnetic fields and biological nuclei frequencies are used to get site specific J(w)[8].

Modelfree analysis is a strategy used to calculate J(w) without assuming a specific model for the motions causing relaxation. J(w) is calculated using the overall rotational correlation time, local correlation times and local freedom of motion described by the ordered parameter S2[15]. S2 of 1 indicates a completely ordered local system without local motion and a S2 of 0 indicates a completely disordered local system or a system confined to rotation within the magic angle.

Empirically, S^2 values for backbone amide ^{15}N sites are found to be > 0.8 in the secondary structures and between 0.5 and 0.8 for loops, turns and termini [4].

Material and methods

Uniformly ^{15}N-labeled human H-NOX_c, H-NOX[H105F]_c and H-NOX_Zn complex in 9:1 buffer (50mM NaCl, 20mM NaPi pH 6.5 1mM DTT and 10µM cinaciguat or ZnPPIX)/D_2O, were used for all NMR relaxation measurements. Measurements were made using two-dimensional sensitivity-enhanced ^{1}H ^{15}N HSQC based sequences to measure R*1*, R*2* and heteronuclear NOE, through inversion recovery [17], Carr-Purcell-Meiboom-Gill (CPMG) [18], and steady-state NOE [19] respectively. R*1* and R*2* were collected as pseudo 3D experiments while the [15]N-^{1}H NOE was collected in interleaved mode. The experiment sets for HNOX_c and H-NOX_Zn were collected at 25[o]C while those for H-NOX[H105F]_c were collected at 30[o]C. The R*2* experiments used for the H-NOX_c utilized a temperature compensation step which was not used in the rest of the R*2* measurements. [15]N R*1* and R*2* relaxation rates were measured at ^{15}N fields of 81.1 and 50.6 MHz. Sweep widths for R*1* and R*2*: 11160.714 (800MHZ) 7002.801 (500MHz) Hz and 2757.612 (800MHz) 1925.925 (500MHz) Hz were used in the ^{1}H and ^{15}N dimensions respectively. A total of 2K complex data points were used in the ^{1}H dimension in all cases while for ^{15}N, between 140 and 180 increments were collected. In all cases the ^{1}H carrier frequency was positioned on the H_2O resonance, and the ^{15}N carrier was positioned at frequencies corresponding to 117ppm for the 800MHz and 115ppm for the 500MHz. For the R*1* experiment nine relaxation delays were used (0.02, 0.05, 0.1, 0.3, 0.6, 0.9, 1.1, 1.5, 2.0s) for the 800 MHz and eight (0.05, 0.1, 0.25, 0.50, 0.75, 1.00, 1.50, 2.00 s) for the 500 MHz. Recycle delays of 6s (800MHz) and 5s (500MHz) were employed and 32 (800 and 500MHz) scans were acquired per t_1 increment. For the measurement of R*2* in the case of H-NOX_c and H-NOX[H105F]_c, relaxation delays of 0.01584, 0.03168, 0.06336, 0.09504, 0.12672, 0.1584, 0.19008, 0.2376, 0.3168s for 800MHz and 500MHz were achieved by repetition of the spin echo block of total duration of

0.01584s by 1,2,4,6,8,10,12,15,20 times. A recycle delay of 4s (800MHz) and 3s (500MHz) was used, and 32 (800 and 500MHz) scans were acquired per t_1 increment. The delay between 180˚ pulses in the CPMG sequence was set to 450ms (800 and 500MHz). For the H-NOX_Zn sample R2 measurements, a sequence with temperature compensations and shorter spin echo was used using the delays of 0.00848, 0.03392, 0.06784, 0.0848, 0.1272, 0.1696, 0.212, 0.2544 and 0.3392s which were achieved by repeating the spin echo with duration of 0.00848s 1,4,8,10,15,20,25,30 and 40 times respectively.

The heteronuclear NOE data acquisition was performed at both 800 and 500 MHz using the pulse sequence described by [13]. In both cases, a single experiment was performed where two spectra were collected in interleaved mode, one with ^{1}H saturation prior to the {[1]H}-^{15}N NOE step and another without. 2048 complex points and 256 increments were used in the ^{1}H and ^{15}N dimensions, respectively. Recycle delays of 10s (800MHz) and 5s (500MHz) were used, and 16 (800MHz) and 40 (500MHz) scans were acquired. In the saturation spectrum, ^{1}H magnetization was saturated by applying a series of non selective 120˚pulses prior to the first ^{15}N pulse for the duration of the relaxation delay.

All data were processed using Bruker-Topspin software (version 3.2). Data analysis was performed using the CCPN software package. Cross peak intensities were evaluated as peak heights. The relaxation rates R1 and R2 were obtained by non-linear least-squares fitting of single exponential decays to the experimental data. The NOE was calculated as a ratio of the peak intensities measured in the presence and absence of saturation of the proton magnetization.

Results

Relaxation data

^{15}N R1 and R2 relaxation rates as well as the {^{1}H}-^{15}N NOE for human H-NOX_c complex can provide insights on the dynamics of the protein. In the case of R*1* and R*2*, the decay of peak heights as a function of relaxation delay for each HSQC peak was visually inspected to see if there was sufficient signal to noise across the series as well as the presence of severe overlap. Peaks that did not fit the criteria were not considered in the subsequent analysis. The resulting data sets were fit using a monoexponential function and reliable fits were made for 147 of the expected 172 resonances (188 residues minus three prolines, the amino terminal and twelve residues, which could not be observed in ^{15}N HSQC).

Figure 1 A-D shows the plots of R*1*, R*2*, R*2*/R*1* and {^{1}H}-^{15}N NOE values obtained for the H-NOX_c plotted against its amino acid sequence. The average values for R*1*, R*2*, R*2*/R*1* and het NOE excluding the flexible regions are shown in Table 1. These values are consistent with a protein of 25kDa and suggest that our construct is monomeric in solution. The criterion for defining the flexible regions to be excluded from the calculation of the averages was residues whose value was different from the average by more than two standard deviations. Close observation of the residues with excluded values showed that they can be categorized into three types, first those with R*2/R1* values below the average that appear to be clustered in the sequence, and others, still with low R*2/R1* values where their deviations are isolated on a single residue in the sequence. Finally there are those that have R*2/R1* values higher than average.
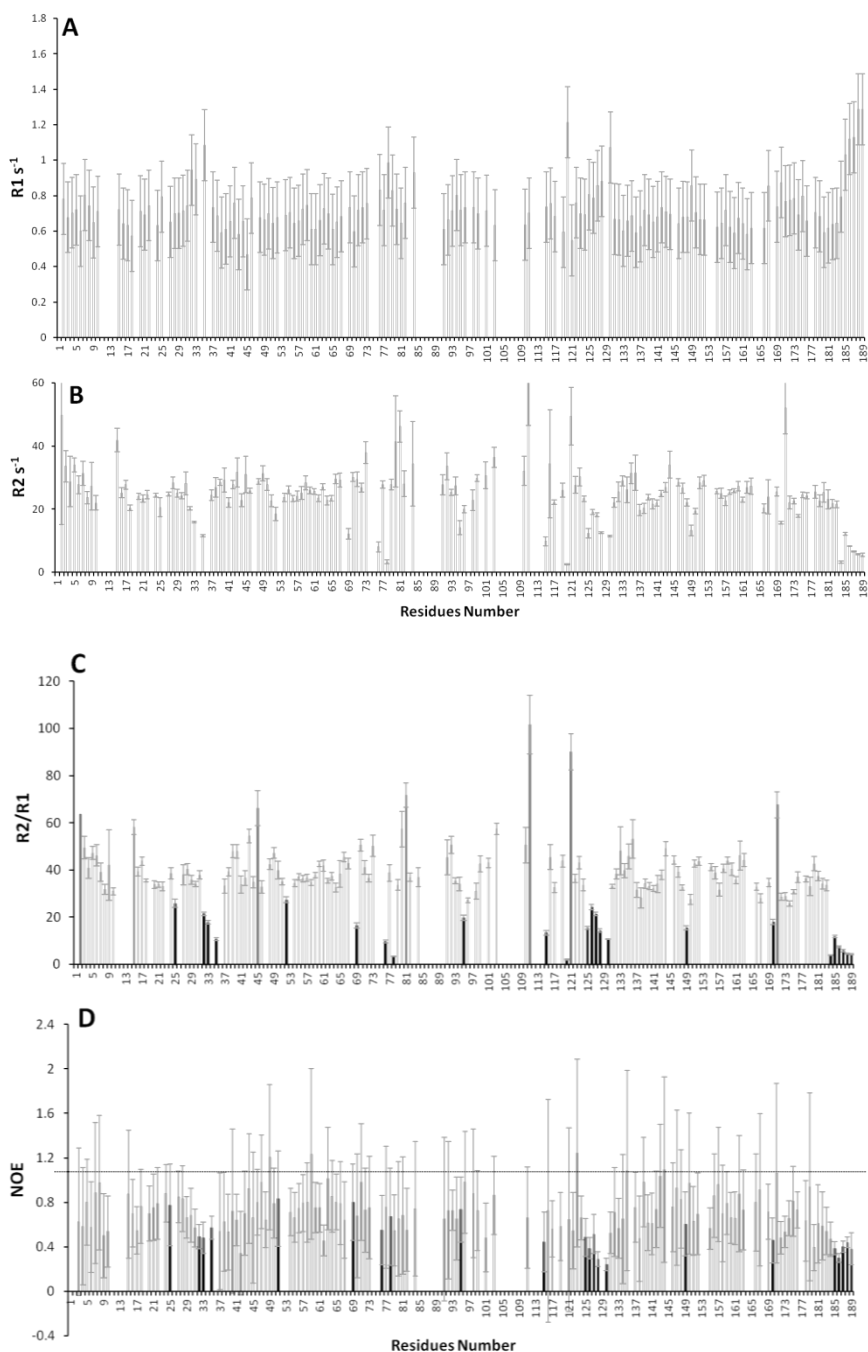
Figure 3: A- R*1*, **B-** R*2*, **C-** R*2*/R*1* and **D-** {^{1}H}-^{15}N hnNOE values from H-NOX_c determined at 800MHZ (data from 500 MHz on Figure S1) as a function of protein sequence. The dark grey line shows the residues which present slow dynamics from the protein backbone

average. The black line shows the residues which present fast dynamics from the protein backbone average.

The residues whose R2/R1 deviate from average values are: Y2, R15, I25, D32, E33, G35, D45, V52, M69, V76, C78, S81, R95, I111, M115, F120, R121, D124, A125, E126, K127, G128, G130, I149, R170, N171,184-189 KESKEH. Those from Y2, R15 D45, S81, I111, R121, and N171 had R2/R1 values higher than average and are shown in Figure 4A. The remaining residues are shown in Figure 4B. It is clear also that three clusters of residues with consistently deviating values are present. These residue ranges are D32-G35, D124-G130 and K184-H189. Highlighting these ranges on the three-dimensional view of the protein model, can be seen in Figure 4B in blue colour. D32-G35 is positioned at the loop linking the N-terminal helices αB-αC, D124-G130 is located on the β1-β2 linker, K184-H189 is on the C-terminal tail.
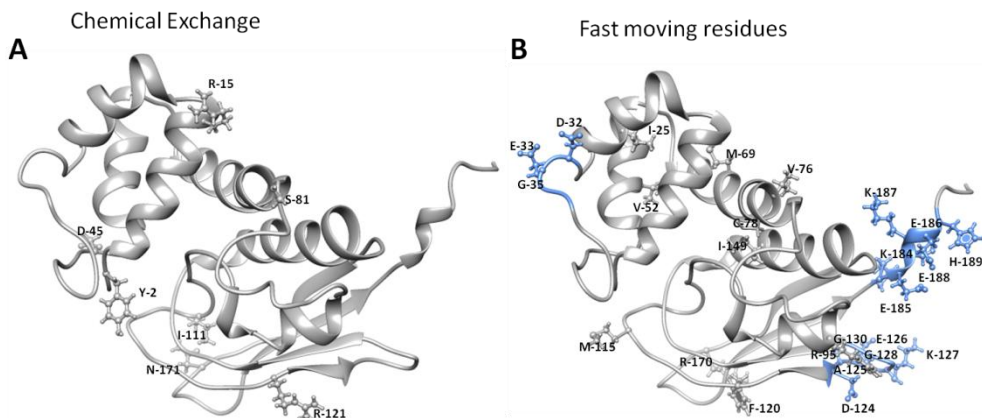
Figure 4: Structure of human H-NOX_c domain calculated by CNS1.21 program [20]. A- Residues which showed slow dynamics than protein average. B- Residues which showed fast dynamics than protein average. The blue region in the protein structure presents amino acid linker cluster to α helixes B-C, amino acids linker to β helixes 1-2 and C-terminal region that also showed fast local motion.

The data of the H-NOX_c (and the H-NOX_Zn) sample are characterized by a low signal/noise ratio making its analysis difficult. One of the main reasons for this was the high R2 at that temperature. It was then decided to collect data at

303K in future samples to improve signal to noise. This was only possible to do in the case of H-NOX^{H105F}_c due to sample availability.

Since a ^{13}C^{15}N labelled sample was not available for the H-NOX^{H105F}_c complex and because the signals were not significantly shifted in respect to those of H-NOX_c, transfer of the ^{1}H and ^{15}N assignments was performed using criteria of signal proximity and pattern similarity between the two species. Trustworthy signal identification was possible for 142 of the expected 147 resonances. Having performed the transfer, dynamics data could be correlated with the H-NOX^{H105F}_c sequence and compared with those of the H-NOX_c sample.

From the results, both protein forms complexed with cinaciguat showed similar dynamics behaviour (Figure 5A-D), presenting the same dynamics regions (Figure 4A-B).
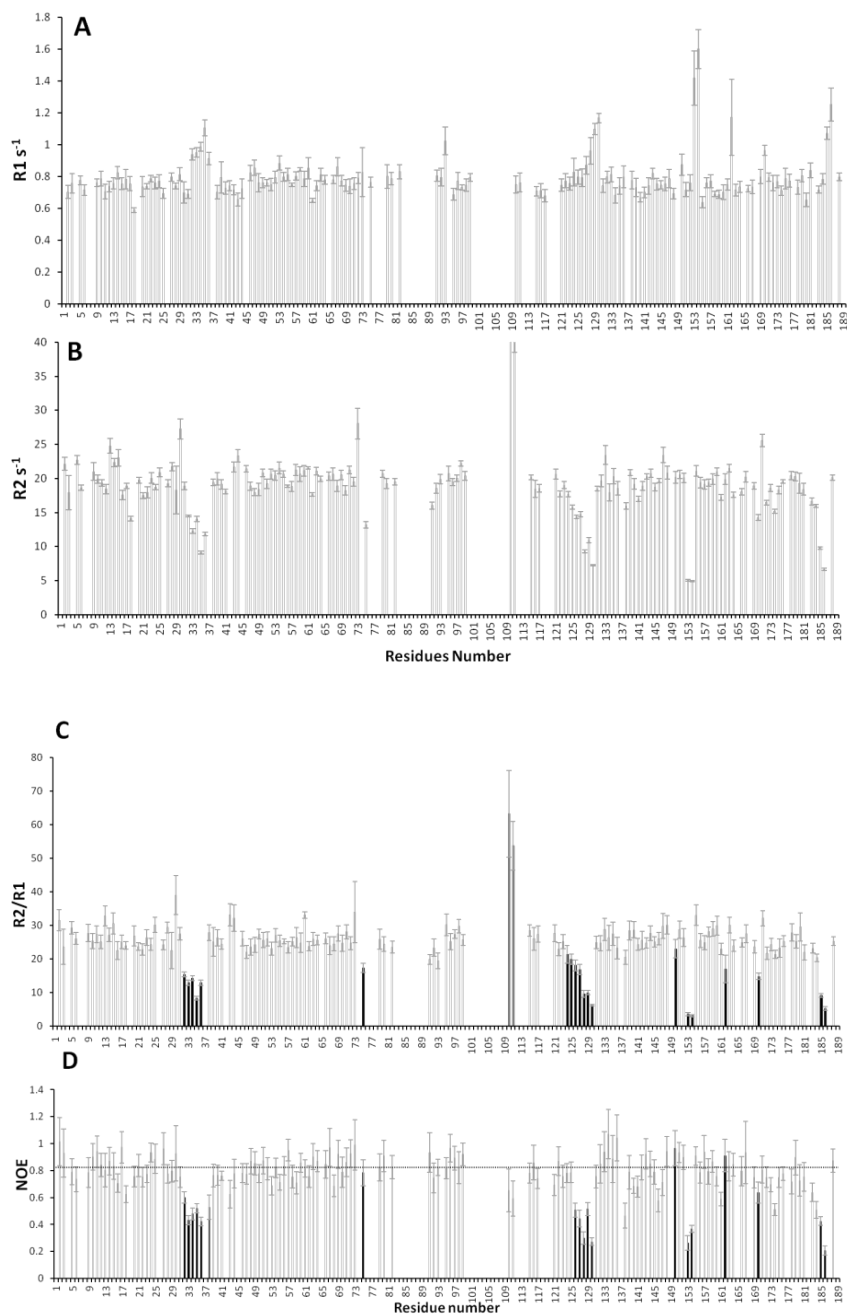
Figure 5: A- R*1*, **B-** R*2*, **C-** R*2*/R*1* and **D- {**[1]H}-[15]N *hnNOE* values from H-NOX[H105F]_c determined at 800 and 500 MHZ (S2) as a function of protein sequence. The dark grey lines correspond to residues exhibiting chemical exchange in the ms timescale. The black lines correspond to residues with faster dynamics than average.

In respect to the H-NOX_Zn spectra, only 60 of the expected 147 resonances could be transferred and relaxation measurements were evaluated only for them. Thus, a complete analysis of the dynamics data between the active (H-NOX_c) and inactive (H-NOX_Zn) form of the protein was compromised due to the absence of full assignment of the latter. Moreover, due to the weak signal to noise of the relaxation data collected at 500 MHz, it was not possible to extract meaningful data and it was not included in the analysis. Some residues, which were accurately transferred, indicated that the protein dynamics regime may be similar to the dynamics behaviour of the protein in active state (Figure 6A-D). Nonetheless, the full protein assignment needs to be performed for more conclusive results.

Table 1: The relaxation data average from H-NOX_c, H-NOXH105F and H-NOX_Zn complex. Data from H-NOX[H105F]_c were collected at 303K while the rest were collected at 298K. Only the R2 measurements of H-NOX_Zn were collected using temperature compensation pulse sequence.

Sample	R1(Hz)	R2(Hz)	R_2/R_1	{¹H}-¹⁵N NOE	MHz
H-NOX_c	0.7 ± 0.04	25.0±1.86	35.0±0.1	0.71± 0.4	800
	1.2 ± 0.10	19.0± 2.50	16.0±4.0	-	500
H-NOX[H105F]_c *	0.77 ± 0.03	19.2 ± 0.45	25.4±2.0	0.77± 0.1	800
	1.36 ± 0.06	13.0 ± 0.03	10.35±0.6	0.70± 0.1	500
H-NOX_Zn	0.71 ± 0.04	20±1	27.76±3.0	0.65± 0.1	800
	-	-	-	-	500

The mean experimental relaxation rates *R1* (and *R2* and the ¹H-¹⁵N NOE values obtained for H-NOX_c, H-NOX[H105F]_c and H-NOX_Zn are shown in Table 1. It should be noted that comparison of these relaxation rates for the three protein samples described on the Table 1 cannot be quantitatively compared because of inconsistencies in the acquisition temperature of each sample. The H-NOX_c and HNOX_Zn samples were both acquired at 298K but due to an upgrade of the NMR software (from Topspin2.1 to Topspin3.1) a sequence using temperature compensation was used for the acquisition of the R2 data of H-NOX_Zn which as it turns out causes increased sample heating. This can be seen from the fact what while the R1 values for H-NOX_c and H-NOX_Zn are practically the same, R2 values of H-NOX_Zn are closer to those of H-NOX[H105F]_c that were collected at 303K. The effect of inefficient temperature control is also evident from the τ_c

values calculated using Equation 2 which is based on the R2/R1 ratios. The mean τ_c values obtained for H-NOX_c, and H-NOX_Zn were 14±1, and 12.1±0.6, respectively. The τ_c value for HNOX^{H105F}_c with the data collected at 303K was 11.6±0.4. The value of τ_c for HNOX_c is consistent with a monomeric protein.

In addition to the problems of temperature control between samples, the relaxation data collected at different magnetic fields (800 and 500 MHz) were also found to be inconsistent, inhibiting a precise quantitative data analysis. This may be due to inefficient temperature control during data acquisition, most likely due to sample heating during the CPMG step of the R2 measurements. In addition, comparison of the ^{1}H-^{15}N HSQC experiments collected before and after data collection of the R2, showed significant chemical shift differences for some residues of the protein (data not shown). However, after repetition of this measurement the sample was stable for the remainder of the data collection.
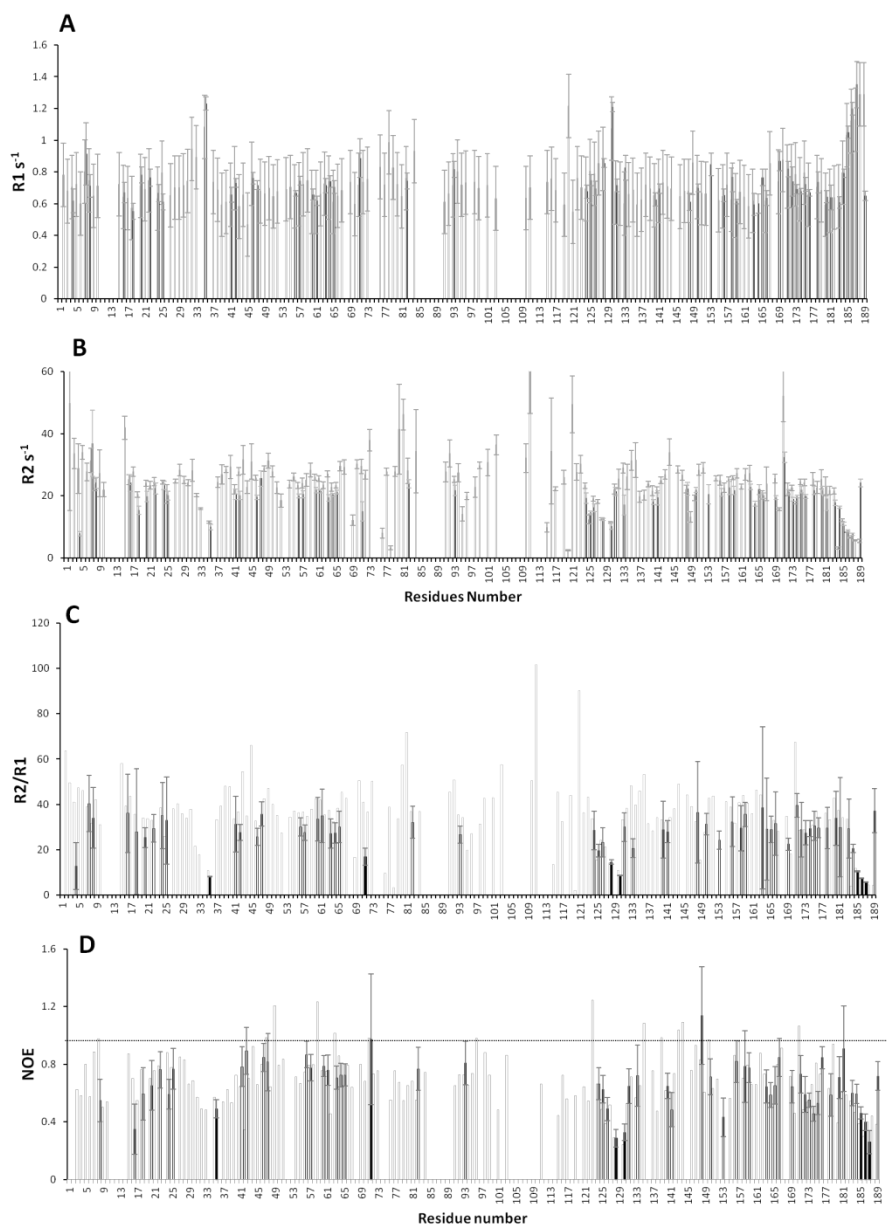
Figure 6: A and B- Comparison between R*1* and R*2* H-NOX_c (grey) and H-NOX_Zn (black border) C R*2*/R*1* and D- {^{1}H}-^{15}N *hnNOE* overlay values from H-NOX_c (grey) and H-NOX_Zn (black border and black fill) as a functions of protein sequence. All values were obtained at 800 MHz.

Discussion

Protein dynamics have been described to be potentially involved in the regulation of allosteric mechanisms [18]. Allostery can regulate cellular assemblies and pathways. It can be done through a perturbation by an effector leading to a functional change at the substrate binding site, thereby altering the shape and/or dynamics of the protein [19]. In the present study, NMR relaxation experiments were collected and analyzed to probe the sub-nanosecond motions of interaction (bond) vectors of H-NOX sGC domain. Observable NMR relaxation parameters, such as the spin-lattice (R1) and spin-spin (R2) relaxation rates and {[1]H}-^{15}N hnNOE of the nuclear spin are directly related to the spectral density function J(ω), describing the motion of the involved interaction vector [4-6].

Several protein backbone dynamics studies based on ^{15}N relaxation data have been reported, and in most cases the main chain largely acts as a rigid scaffold [20,21]. In the case of H-NOX_c the R1, R2 and {[1]H}-^{15}N hnNOE for H-NOX_c (active state) suggest that residues Y2, R15, I25, D32-G35, D45, V52, M69, V76, C78, S81, R95, I111, M115, F120, R121, D124-G130, I149, R170, N171,K184-H189 differ in their dynamics properties in comparison to the rest of the protein backbone scaffold (Figure 1A-B). As previously mentioned, both R1 and R2 relaxation rates for a rigid molecule are not expected to vary from the average values. Deviations are indicative of additional mobility at those sites in excess of molecular tumbling. Similarly, heteronuclear NOE data provide information about the motion of individual N-H bound vectors. Thus, each H-N bond that shows a lower intensity relative to the average of protein backbone residues, possesses a faster motion than the overall domain. Flexible regions of the protein always show a decrease in hnNOE [4,9,22-24]. After a qualitative analysis of the experimental relaxation data, the results were plotted against amino acid

sequence, allowing the identification of local motion within the protein (Figure 1A-D).

Among the dynamic residues, three clusters can be identified, D32-G35, D124-G130 and K184-H188. The first, D32-G35 corresponds to part of the αB-αC loop. The functional importance of αB-αC loop which in our structure extends from A29 to D45, has been explored in previous studies [27,29]. In those studies, mutations in several residues such as the R40E, I41E and D45A resulted in substantial reduction on NO-sGC stimulation of the cyclase activity indicating that the αB-αC loop may take part in the regulatory interactions with the catalytic domain [30]. In our studies of the H-NOX complexed with cinaciguat, which corresponds to the activated form of sGC, we found that residues D32, E33 and G35 of the αB-αC loop exhibit ps-ns dynamics. However, residues 36-44 did not appear to be experiencing ps-ns mobility which is notable based on the absence of specific secondary elements and the fact that it is water exposed. D45 however, exhibited an increased R2/R*1* average R*2* but a lower than average R*1*. The combination of these observations suggests that this residue undergoes, simultaneous ps-ns and ms dynamics. The explanation for this is that R*1* which is not affected by chemical exchange will decrease in the case of ps-ns mobility. Under those circumstances, R*2* has to also decrease. However, in our cases it remained average suggesting that it was increased by an additional term originating from chemical exchange. The presence of increased mobility in the activated form of sGC is interesting but it needs to be compared with the mobility of the non active form in order to be able to assess if it is of interest. Although H-NOX_Zn is described as sGC inhibitor, for our studies, it could be used as model of the non activated protein (ferrous form of heme in its NO free state). As stated previously, not a lot of data are available for this protein; however, Q36 is one of the residues for which relaxation was measured. Comparison of the behavior of this residue with that of the active form shows that they behave in a similar manner, with increased ps-ns dynamics.

This suggests that the mobility in the region of D32-G35 may not be involved directly involved in regulation. D45 cannot be evaluated in a similar way since it was not identified in the H-NOX_Zn sample.

Another cluster of residues, which has been found to display small changes in either basal or NO-stimulated activity is the αF-β1 surface loop region which encompasses residues FLQNLDALHDHLATI (αF) YPGMRAPSF (αF-β1 surface loop) [30]. It was identified by Hydrogen Deuterium Exchange Mass Spectrometry assays (HDX-MS) [30,31]. With this study it was possible to map H-NOX accessibility and dynamics changes induced by PAS domain. The final result indicated that the PAS domain either directly interacts with the signalling helix or strongly influences the local structure around the heme [31]. In the present study the dynamics studies with NMR revealed a higher degree of flexibility in different regions localized between β1-β2 loop (D124, A125, E126, K127, G128 and G130). Further advance studies such as: double labelled $^{13}C/^{15}N$ sample for collecting the dynamics data from side chain residues of H-NOX domain may be required to demonstrate if the fast dynamics showed by β1-β2 loop, indeed is involved in domain-domain interaction as αF-β1 surface loop region.

At this point it should be noted that a number of signals corresponding to the amides of residues T85-R88 and H105-Y112 and could not be identified in the ^{15}N HSQC spectra regardless of testing multiple sample conditions such as temperature, salt concentration, pH and excess ligand (See chapter 2). The most plausible explanations for this is that these amides in the intermediate exchange regime in the ms timescale making them invisible to NMR. Since data collected at pH ranges of 6-7.5 do not result in more signals, it is likely that this exchange regime is caused by conformational motions in the ms time scale, most likely due to the flexibility of the cinaciguat ligand. These missing residues are forming the cinaciguat binding cavity and encompass the αF helix which in previous studies

was shown to be very dynamic. This coupled with the proximity of the β1-β2 loop that was found by us to be very mobile suggests that the whole region of the protein is very flexible. This observation has a lot of significance for all the studies involving the H-NOX domain. Relaxation measurements of the H-NOX_Zn also show that the β1-β2 loop is mobile suggesting that this is a feature of the protein and not related to the nature of the bound ligand.

Finally, protein modeling [25-27] suggests that the c-terminal sequence acts as a linker between the H-NOX and the Per/ARNT/Sim (PAS) domains and in our construct is solvent exposed, thereby increasing its dynamics. To prove this local motion, NMR studies using truncated β1 human H-NOX/PAS [28] domains should be included in the future perspective plans.

In conclusion, our studies have shown that the H-NOX domain of sGC is a very flexible entity. Part of this flexibility might be due to cinaciguat, although the regions of the αB-αC loop and β1-β2 loop seem to be dynamic in both the ZnPPIX and the cinaciguat bound forms. The reasons for this increased flexibility can also be related to the absence of the remaining domains that make up sGC. Therefore, further studies are needed such as; to probe inter-domain interaction by NMR to be able to draw more conclusions.

References

1. Ruschak AM, Kay LE (2010) Methyl groups as probes of supra-molecular structure, dynamics and function. J Biomol NMR 46: 75-87.
2. Kleckner IR, Foster MP (2011) An introduction to NMR-based approaches for measuring protein dynamics. Biochim Biophys Acta 1814: 942-968.
3. Thompson LE, Rovnyak D (2007) Accessible NMR experiments studying the hydrodynamics of (15) N-enriched ubiquitin at low fields. Biochem Mol Biol Educ 35: 49-56.
4. Kempf JG, Loria JP (2003) Protein dynamics from solution NMR: theory and applications. Cell Biochem Biophys 37: 187-211.
5. Jarymowycz VA, Stone MJ (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. Chem Rev 106: 1624-1671.
6. Igumenova TI, Frederick KK, Wand AJ (2006) Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. Chem Rev 106: 1672-1699.
7. Maudsley AA (1986) Modified Carr-Purcell-Meiboom-Gill sequence for NMR fourier imaging applications. Journal of Magnetic Resonance (1969) 69: 488-491.
8. Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE (1995) Spectral density function mapping using 15N relaxation data exclusively. J Biomol NMR 6: 153-162.
9. Halle B (2009) The physical basis of model-free analysis of NMR relaxation data from proteins and complex fluids. J Chem Phys 131: 224507.
10. Baber JL, Szabo A, Tjandra N (2001) Analysis of slow interdomain motion of macromolecules using NMR relaxation data. J Am Chem Soc 123: 3953-3959.
11. Mayo KH, Daragan VA, Idiyatullin D, Nesmelova I (2000) Peptide internal motions on nanosecond time scale derived from direct fitting of (13)C and (15)N NMR spectral density functions. J Magn Reson 146: 188-195.
12. Millikan RA (1923) Stokes' Law of Fall Completely Corrected. Proceedings of the National Academy of Sciences of the United States of America 9: 67-70.
13. Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28: 8972-8979.
14. Levitt MH (2007) Spin Dynamics: Basics of Nuclear Magnetic Resonance.
15. Mandel AM, Akke M, Palmer AG, 3rd (1995) Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. J Mol Biol 246: 144-163.

16. Jin D, Andrec M, Montelione GT, Levy RM (1998) Propagation of experimental uncertainties using the Lipari-Szabo model-free analysis of protein dynamics. J Biomol NMR 12: 471-492.

17. Vold RL, Waugh JS, Klein MP, Phelps DE (1968) Measurement of Spin Relaxation in Complex Systems. The Journal of Chemical Physics 48: 3831-3832.

18. Carr HY, Purcell EM (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. Physical Review 94: 630.

19. Keepers JW, James TL (1984) A theoretical study of distance determinations from NMR. Two-dimensional nuclear Overhauser effect spectra. Journal of Magnetic Resonance (1969) 57: 404-426.

20. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. Nat Protoc 2: 2728-2733.

19. Tsai CJ, del Sol A, Nussinov R (2008) Allostery: absence of a change in shape does not imply that allostery is not at play. J Mol Biol 378: 1-11.

20. Palmer AG (1997) Probing molecular motion by NMR. Current opinion in structural biology 7: 732-737.

21. Stone MJ (2001) NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding. Accounts of chemical research 34: 379-388.

22. Boehr DD, Dyson HJ, Wright PE (2006) An NMR perspective on enzyme dynamics. Chem Rev 106: 3055-3079.

23. Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol 5: 789-796.

24. Konrat R (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. J Magn Reson 241: 74-85.

25. Campbell MG, Underbakke ES, Potter CS, Carragher B, Marletta MA (2014) Single-particle EM reveals the higher-order domain architecture of soluble guanylate cyclase. Proc Natl Acad Sci U S A 111: 2960-2965.

26. Fritz BG, Roberts SA, Ahmed A, Breci L, Li W, et al. (2013) Molecular model of a soluble guanylyl cyclase fragment determined by small-angle X-ray scattering and chemical cross-linking. Biochemistry 52: 1568-1582.

27. Haase T, Haase N, Kraehling JR, Behrends S (2010) Fluorescent fusion proteins of soluble guanylyl cyclase indicate proximity of the heme nitric oxide domain and catalytic domain. PLoS One 5: e11617.

28. Zhao Y, Marletta MA (1997) Localization of the heme binding region in soluble guanylate cyclase. Biochemistry 36: 15959-15964.

29. Martin F, Baskaran P, Ma X, Dunten PW, Schaefer M, et al. (2010) Structure of cinaciguat (BAY 58-2667) bound to Nostoc H-NOX domain reveals insights into heme-mimetic activation of the soluble guanylyl cyclase. J Biol Chem 285: 22651-22657.

30. Underbakke ES, Iavarone AT, Marletta MA (2013) Higher-order interactions bridge the nitric oxide receptor and catalytic domains of soluble guanylate cyclase. Proc Natl Acad Sci U S A 110: 6777-6782.

31. Underbakke ES, Iavarone AT, Chalmers MJ, Pascal BD, Novick S, et al. (2014) Nitric oxide-induced conformational changes in soluble guanylate cyclase. Structure 22: 602-611.

Supplemental Data
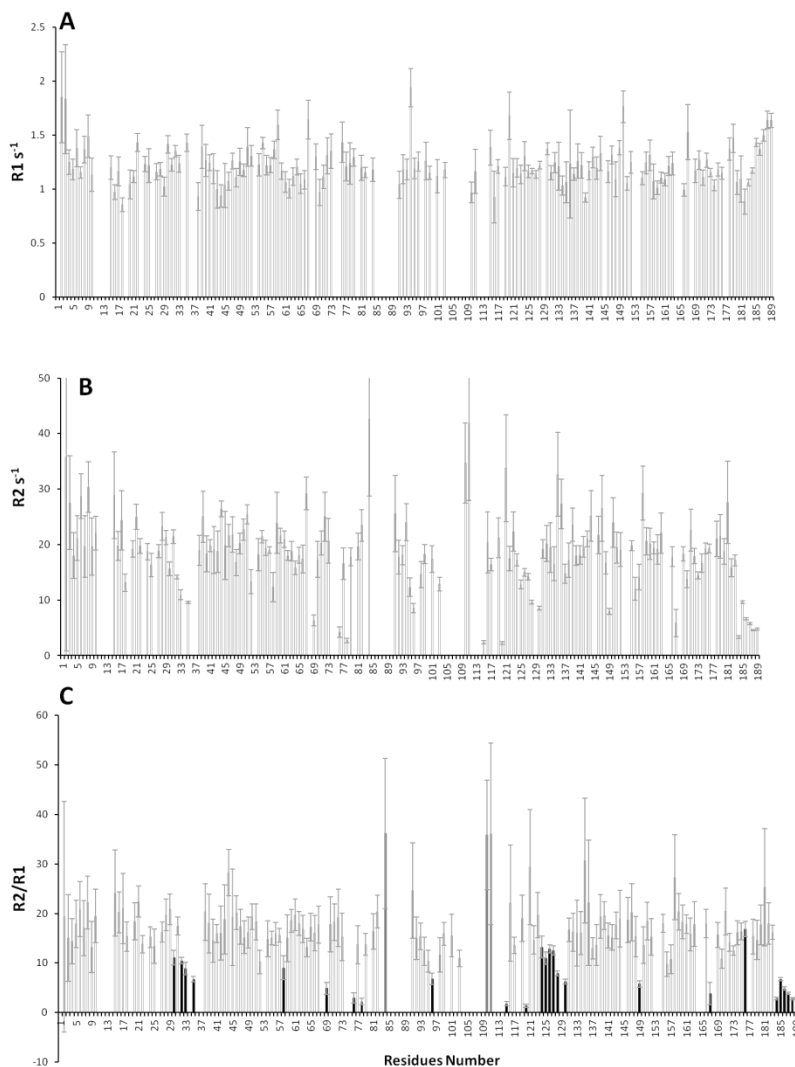
S1: A- R*1*, **B-** R*2*, **C-** R*2*/R*1* values from H-NOX_c determined at 500 MHZ as a function of protein sequence. The dark grey line shows the residues which present slow dynamics from the protein backbone average. The black line shows the residues which present fast dynamics from the protein backbone average.
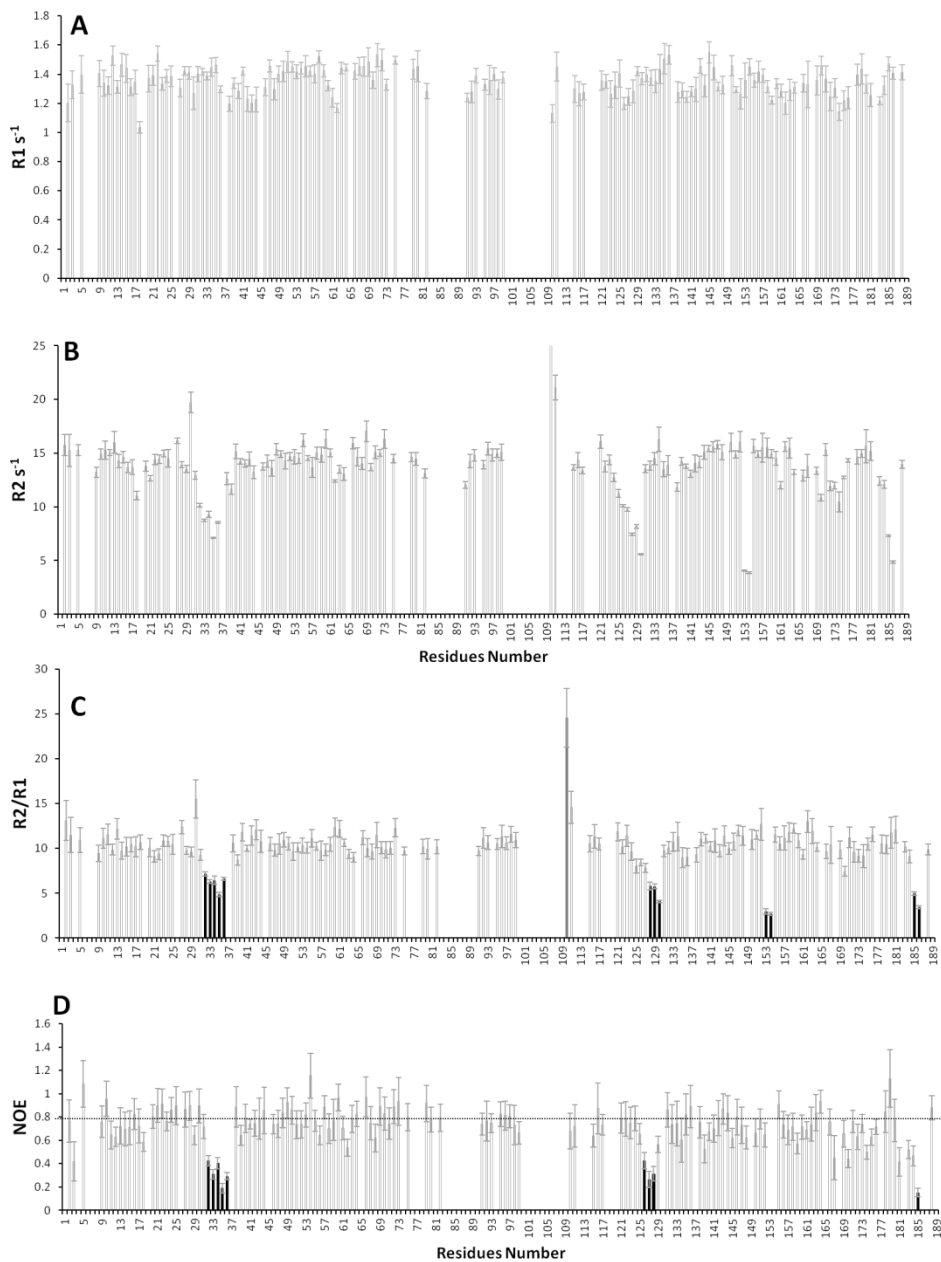
S2: A- R*1*, **B-** R*2*, **C-** R*2*/R*1* and **D-** Heteronuclear NOE values from H-NOX^{H105F}_c determined at 500 MHZ as a function of protein sequence. The dark grey line shows the residues which present slow dynamics from the protein backbone average. The black line shows the residues which present fast dynamics from the protein backbone average.

Chapter 5

General Discussion and Future

Perspectives

Table of Contents

General Discussion

Human H-NOX is one domain among other three (PAS, CC and catalytic domain) that compose the sGC enzyme [1]. The tridimensional arrangement of the sGC domain is still unknown and the way H-NOX domain senses the diatomic NO gas and translates the signal to the catalytic site hasn't been clarified so far. Nevertheless, several pharmaceutical compounds have been found to fully activate sGC in the case of its inactivation due to cellular oxidative stress. One of those activators is the compound cinaciguat. [2-6]. Functional studies of the H-NOX sGC domain are still crucial to fill in the gap about sGC allosteric regulation. Since no structure of the human H-NOX domain was published by the time we started our research, the main goal of this project was to deepen our knowledge of the structure and functional behavior of the H-NOX domain at an atomic level, based on structural and dynamics studies using the potentialities of NMR spectroscopy.

Characterization of human H-NOX domain ligand binding

In this doctoral work I primarily focused in obtaining a well-folded form of the H-NOX domain complexed with different ligands. The initial proposal was to study the structure and dynamics of H-NOX domain in resting (Heme-Fe^{2+} bound to CN), activated (Heme-Fe^{2+} bound to NO/CO or cinaciguat) and inhibited state (Heme-Fe^{2+} bound to ZnPPIX). However, as described in **chapter 2**, the ligand incorporation step revealed to be one of the most challenging steps of our work. Whereas cinaciguat showed to be stable when bound to the H-NOX sGC domain, incorporating and purifying this domain with its natural ligated (Heme-Fe^{2+}) turned into a very demanding work.

Several techniques have been developed to increase heme incorporation into recombinant proteins. One of the explored protocols involves supplying the bacterial growth media with heme cofactor [7,8]. Another approach utilizes the heme-permeability of the *E. coli* strain RP523, which has a porphobilinogen synthase gene, the *hem B gene,* disrupted, to prevent synthesis of native heme. All heme and/or heme analogs are acquired by the cells from the growth media. [9]. Here, in this work we tried to co-express the human H-NOX domain with the native bacterial ferrochelatase enzyme (FC). This method has been described as a straightforward and inexpensive method to increase overproduction and thus its incorporation of heme into proteins over-expressed in *E. coli*. Since the co-expression of FC with GsNOS protein has been proved to be highly efficient [10], high levels of heme incorporated into the human H-NOX domain were expected. However, despite having successful co-expression of FC protein and H-NOX domain, the heme incorporation failed. A Human H-NOX heme reconstitution protocol was also tested [11]. And despite having a positive result from a UV-vis spectrum indicating full heme incorporation into the H-NOX domain, no protein in the monomeric form was obtained, indicating that the heme-H-NOX complex formed might be unstable (**chapter 2**). Derbyshire et al, (2010), worked with the full sGC enzyme, and isolated this protein with a sub-stoichiometric amount of heme. This could only be possible due to the absence of the polymeric form of the protein [11], unlike the apo H-NOX domain studied in this work. Therefore, based on heme-protein incorporation instability, the protocol for protein expression and heme incorporation described by [9] could be, a way to produce the H-NOX domain in the resting and active states. In a recent study, $\beta1$ human H-NOX [12] was not only successfully expressed using RP523 *E. coli* strain, but also the heme cofactor was efficiently incorporated [13].

The resting state of sGC-heme can be converted into the sGC inactive form by replacing heme with ZnPPIX. This heme analogue has been described to

be a potent sGC inhibitor able to make 5-coordination with H-NOX H105, but unable to sense NO or CO gas [14,15]. In fact, ZnPPIX was easily incorporated into the human H-NOX domain, probably due to the fact that this ligand does not undergo oxidation. Since the complex formed is not paramagnetic, it is a good model of the inactive form of the H-NOX domain. Proteins complexed with ZnPPIX have been found to have relevant physiological functions. For instance, Zinc protoporphyrin IX can be found in red blood cells as when heme-Fe^{2+} production is inhibited, a ferrochelatase incorporates a Zn^{2+} ion into PPIX catalysis forming ZnPPIX [16]. As mentioned in **chapter 1**, NO is the most potent physiological activator of sGC, thereby acting as a signalling molecule in the nervous and cardiovascular systems. In spite of its poor sGC-activating ability, CO has also been implicated as a physiological stimulator of sGC in neurotransmission and vasorelaxation [17]. CO is produced by the enzyme heme oxygenase (HO), and ZnPPIX has also been described to be a strong HO inhibitor [18]. In other words, ZnPPIX is able to inhibit HO activity and CO production, lowering endogenous cGMP levels by inhibition of sGC in olfactory receptor neurons, which is dependent on CO production and endogenous cGMP concentration [14]. Since the sGC-ZnPPIX interaction also occurs naturally *in vivo*, structural and dynamics studies of the human H-NOX sGC domain inhibited by ZnPPIX is important form a physiological view point.

Cinaciguat, NO- and heme-independent sGC activation has been proved to be more effective when the heme-group of sGC is oxidized in vascular tissue (**chapter 1**). Our experiments showed that this compound is stable when bound to the WT H-NOX sGC domain after incorporation (**chapter 2**). So, the mutation of the histidine 105 to phenylalanine has been found to allow the full protein activation by cinaciguat compound, therefore it keeps the native protein conformation. Thus, this mutation means the H-NOX domain is unable to

incorporate the heme cofactor but has however, a high affinity for cinaciguat. This variant being the best model to study H-NOX domain after heme-oxidation. H-NOX[H105F] with cinaciguat bound was prepared to compare its structure and dynamic behaviour with the WT-cinaciguat bound form. In recent literature, this mutant was also found to be a good model *in vivo*. In apo-sGC from mice (sGCβ1 (H105F) knockin) both mutated sGC isoforms (sGCα1 β1 and sGCα2 β1) showed to help decrease the blood pressure upon cinaciguat treatment [19]. In the end, the studies with WT and mutant human H-NOX domain complexed with different ligands contributed for our knowledge of the functional aspects of the H-NOX domain. From our preliminary NMR data analysis (**chapter 2**), all protein samples were shown to be well folded, allowing structural characterization to be carried out.

Structural characterization

Over the last fifteen years, the role of NMR spectroscopy in protein structure determination has been increasing gradually [20]. It occupies a unique niche in the biophysical analysis. As we could see in the **chapter 2**, from a single 1D spectrum we can access information related to the folding state of the H-NOX domain. For instance, a folded H-NOX sample complexed with cinaciguat or ZnPPIX would show proton signals in the beta sheet region (around 5 ppm), the amide signals would be well resolved and spanned more than 2 ppm window and also the signals from methyl resonances would be close or below 0 ppm. The described features were indeed observed in the 1D NMR spectra, being compatible with a resulting well folded protein. In spite of the chemical differences between cinaciguat and ZnPPIX ligands, the 2D ^{1}H-^{15}N HSQC spectra of the H-NOX domain showed to be highly similar, not losing its signal distribution pattern. However, we were not able to get a chemical shit deviation patter for the empty WT H-NOX domain, since the protein was always collected complexed with

the ligand, due to the fact that this protein alone, i.e ligand free, does not stay stable in a monomeric form as shown in **chapter 2**. Typically, the chemical shift deviation can be monitored with a ligand titration through ^{1}H-^{15}N HSQC collection. It allows following the chemical shift alteration in the spectral window [20,21]. In respect to the structure determination, two major difficulties were encountered. Firstly, a number of resonances composing the cinaciguat binding cavity were not present in our spectra. The absence of these resonances resulted in a low and/or number of NOEs in those regions, giving rise to high RMSD. The second was related in the difficulties identifying intermolecular NOEs between protein and ligand since those signals are almost indistinguishable from intramolecular protein NOEs in standard isotope-edited NOESY spectra [22]. Because cinaciguat was not isotopically labelled, the filtered experiments described in the **chapter 3** provided information regarding both intra cinaciguat NOEs together with inter cinaciguat-protein NOEs, in combination with a 2D TOCSY ^{13}C-filtered in both dimensions, these experiments were used to assign the ligand signals. After protein structure determination, the inter-NOE data indicated that the cinaciguat is involved in hydrophobic interactions with specific protein residues of the heme cavity, including M1, F4, V5, V39, G82, L101, L104, M115, A117, V146, I149, I150 and V153 (**chapter 3 Figure 6D**). Some of these residues have already been found by another research group to be involved in protein-ligand interaction [3].

Protein-ligand interactions are defined by non-covalent interactions such as Van der Waals interactions, hydrogen bonds and electrostatic forces [23,24]. Similar to Heme, cinaciguat is has two hydrophilic charged carboxylates interacting with YxSxR conserved residues [3,25-29]. Nevertheless, since inter molecular -NOEs between the carboxylate groups and the protein are not possible to observe, there was no direct molecular information to prove that the hydrophilic charged carboxylates from the compound interact with the above

mentioned conserved residues. However, based on the 3D structure determined (**chapter 3**) and from the current literature, this interaction is predicted to happen.

The root mean square deviation (r.m.s.d) from our predicted 10 lowest-energy structures of H-NOX domain complexed with cinaciguat was 0.77 Å. It was higher than the 20 lowest-energy structure determined by *So* WT H-NOX NMR structure (r.m.s.d 0.33 Å) [27]. Dynamics at the interface and chemical exchange most likely inhibited the detection of enough inter-molecular NOE [30,31]. The ligand binding region has a poor global precision and accuracy due to the short-range of NOE restraints [24,32]. The introduction of long-range restraints, such as RDCs increased the probability to decrease the r.m.s.d value. So, to improve our 3D structure, the protocol from Erbil and collaborators for *So* WT H-NOX NMR structure characterization [27] in future could be followed. The RDC experiment not only can help solve ambiguity in the β-sheet domain region, but can also be used as input data for the structure determination of the H-NOX-cinaciguat complex. Moreover, it can improve the assignment statistics based on NOE assignment. Our efforts to collect RDCs were hampered by low signal to noise that produced spectra that were not usable.

In the end, despite having some weaknesses, our work was able to determine for the first time the structure of the H-NOX domain from human. And the resulting secondary and tertiary structure were found to be in agreement with homologous H-NOX protein structures already deposited into the protein data bank (**chapter 3**).

Human H-NOX sGC domain dynamics studies

Allosteric interactions involve a series of discrete changes in bonding interactions that modify the protein conformation. The allosteric effect can be mediated exclusively by transmitted changes in protein motions [33]. Several sGC structural analyses have been published showing a comprehensive assignment of the higher-order domain architecture within the uniquely shaped sGC holoenzyme, with an overall structure landscape [13,34,35]. The models proposed suggest that the holo-enzyme exists in a single conformation [13,34,35]. However, numerous inter-domain contacts have been proposed, making the assumption of a single protein conformation highly unlikely [36]. More recently, sGC was described instead as a highly dynamic protein, sampling a wide range of conformations. NO addition does not lock the protein into a defined conformation, but instead sGC appears to retain its large-scale flexibility (**chapter 1 Figure 8**) [37].

As previously mentioned in **chapter 4**, the α helix A-C on the distal pocket of the protein increased the exchange rates when NO binds to sGC. Upon NO-binding, sGC showed residues more affected by the binding event. For instance, residues from 97 to 121 (FLQNLDALHDHLATIYPGMRAPSFR) of the H-NOX domain (αF-β1) showed higher exchange rates upon NO-binding. Moreover, α1 H-NOX (115-121), α1 PAS/CC (379-410), β1 PAS (267-277) β1 CC (357-406) β1 catalytic domain (459-574) also showed to exchange faster upon NO binding [38]. Unlike β1 H-NOX (33-49 αB-C), β1 PAS/CC (322-358), β1 catalytic domain (544-549), α1 H-NOX (51-80), α1CC (51-80), α1 CC (415-454) and α1 catalytic domain (421-547 and 585-612) [38] showed slow exchange. From NMR relaxation experiments with H-NOX-cinaciguat bound form, we were able to detect local dynamics in a region already proved to have slow exchange rates by an HDX study [38]. This region is

formed between 33 and 36 (EEGQ residues) corresponding to the αB-C loop region **chapter 4.**

Conventional NMR relaxation experiments (R*1*, R*2* and heteronuclear NOE) are used to probe the dynamics of a molecule in the picosecond to nanosecond timeframe, i.e. faster than the tumbling time of a system [39-41]. The dipolar coupling is an interaction that exists between two magnetic nuclei. In isotropic conditions, the dipolar coupling averages to zero as a result of the effects of Brownian motion. A weak force on the protein can generate a molecular alignment and possibly induce an incomplete isotropy, allowing the measurement of the dipolar coupling. Thus, the residual dipolar couplings (RDC) between two spins represent the incomplete averaging of spatially anisotropic dipolar couplings. One intriguing feature of RDCs is that it can explore the blind spot of the conventional NMR relaxation experiments [42-44]. Therefore, RDCs experiments can not only be implemented in H-NOX structure improvement, but also in the investigation of the highly dynamic regions mentioned above.

Relaxation dispersion NMR spectroscopy is a powerful tool to achieve high-resolution structural information related protein folding events on the millisecond timescale [45]. Based on our result, we speculate that the β1 portion of the H-NOX domain (33-49 αB-C), which showed slow hydrogen deuterium exchange rates [38], can have dynamics different to the conventional relaxation timeframe. So, in order to gain more insight into the molecular dynamics of the human H-NOX sGC domain, relaxation dispersion could also be an alternative to optimize the protein dynamics investigation. Another approach can be to use ^{13}C-NMR relaxation. This method allows extending the dynamics information, especially the side-chain orientation motion and their interactions [46]. Therefore, so far, we achieved by NMR dynamics information of an individual domain of sGC.

So, further sGC inter-domain interaction by NMR became our target, since it has been found to be involved in full sGC activation model [37].

The work described in this thesis gave us background to obtain high performance in human H-NOX protein expression and purification. A routine was created to monitor the quality of the protein upon each sample production. Knowledge was acquired regarding cinaciguat incorporation by the protein, and the great effort to dissociate the aggregated protein into monomeric form complexed with cinaciguat was successful. Scientific maturity regarding NMR data collection for protein-ligand assignment, structure determination, and its dynamics probing, was achieved by all members involved in this project. Therefore, our results opened a new landscape for us allowing further sGC inter-domain interaction studies by NMR.

References

1. Allerston CK, von Delft F, Gileadi O (2013) Crystal structures of the catalytic domain of human soluble guanylate cyclase. PLoS One 8: e57644.
2. Evgenov OV, Pacher P, Schmidt PM, Hasko G, Schmidt HH, et al. (2006) NO-independent stimulators and activators of soluble guanylate cyclase: discovery and therapeutic potential. Nat Rev Drug Discov 5: 755-768.
3. Martin F, Baskaran P, Ma X, Dunten PW, Schaefer M, et al. (2010) Structure of cinaciguat (BAY 58-2667) bound to Nostoc H-NOX domain reveals insights into heme-mimetic activation of the soluble guanylyl cyclase. J Biol Chem 285: 22651-22657.
4. Mulsch A, Bauersachs J, Schafer A, Stasch JP, Kast R, et al. (1997) Effect of YC-1, an NO-independent, superoxide-sensitive stimulator of soluble guanylyl cyclase, on smooth muscle responsiveness to nitrovasodilators. Br J Pharmacol 120: 681-689.
5. Stasch JP, Schmidt P, Alonso-Alija C, Apeler H, Dembowsky K, et al. (2002) NO- and haem-independent activation of soluble guanylyl cyclase: molecular basis and cardiovascular implications of a new pharmacological principle. Br J Pharmacol 136: 773-783.
6. Straub A, Stasch JP, Alonso-Alija C, Benet-Buchholz J, Ducke B, et al. (2001) NO-independent stimulators of soluble guanylate cyclase. Bioorg Med Chem Lett 11: 781-784.
7. Graves PE, Henderson DP, Horstman MJ, Solomon BJ, Olson JS (2008) Enhancing stability and expression of recombinant human hemoglobin in E. coli: Progress in the development of a recombinant HBOC source. Biochim Biophys Acta 1784: 1471-1479.
8. Varnado CL, Goodwin DC (2004) System for the expression of recombinant hemoproteins in Escherichia coli. Protein Expr Purif 35: 76-83.
9. Woodward JJ, Martin NI, Marletta MA (2007) An Escherichia coli expression-based method for heme substitution. Nat Methods 4: 43-45.
10. Sudhamsu J, Kabir M, Airola MV, Patel BA, Yeh SR, et al. (2010) Co-expression of ferrochelatase allows for complete heme incorporation into recombinant proteins produced in E. coli. Protein Expr Purif 73: 78-82.
11. Derbyshire ER, Deng S, Marletta MA (2010) Incorporation of tyrosine and glutamine residues into the soluble guanylate cyclase heme distal pocket alters NO and O2 binding. J Biol Chem 285: 17471-17478.
12. Karow DS, Pan D, Davis JH, Behrends S, Mathies RA, et al. (2005) Characterization of functional heme domains from soluble guanylate cyclase. Biochemistry 44: 16266-16274.

13. Underbakke ES, Iavarone AT, Marletta MA (2013) Higher-order interactions bridge the nitric oxide receptor and catalytic domains of soluble guanylate cyclase. Proc Natl Acad Sci U S A 110: 6777-6782.

14. Serfass L, Burstyn JN (1998) Effect of heme oxygenase inhibitors on soluble guanylyl cyclase activity. Arch Biochem Biophys 359: 8-16.

15. Stasch JP, Schmidt PM, Nedvetsky PI, Nedvetskaya TY, H SA, et al. (2006) Targeting the heme-oxidized nitric oxide receptor for selective vasodilatation of diseased blood vessels. J Clin Invest 116: 2552-2561.

16. Chakder S, Rathi S, Ma XL, Rattan S (1996) Heme oxygenase inhibitor zinc protoporphyrin IX causes an activation of nitric oxide synthase in the rabbit internal anal sphincter. J Pharmacol Exp Ther 277: 1376-1382.

17. Verma A, Hirsch DJ, Glatt CE, Ronnett GV, Snyder SH (1993) Carbon monoxide: a putative neural messenger. Science 259: 381-384.

18. Ingi T, Cheng J, Ronnett GV (1996) Carbon Monoxide: An Endogenous Modulator of the Nitric Oxide–Cyclic GMP Signaling System. Neuron 16: 835-842.

19. Cosyns SM, Huyghe L, Thoonen R, Stasch JP, Brouckaert P, et al. (2014) Influence of cinaciguat on gastrointestinal motility in apo-sGC mice. Neurogastroenterol Motil 26: 1573-1585.

20. Ziarek JJ, Peterson FC, Lytle BL, Volkman BF (2011) Binding site identification and structure determination of protein-ligand complexes by NMR. Methods Enzymol 493: 241-275.

21. Wuthrich K (1990) Protein structure determination in solution by NMR spectroscopy. J Biol Chem 265: 22059-22062.

22. Nudelman I, Akabayov SR, Scherf T, Anglister J (2011) Observation of intermolecular interactions in large protein complexes by 2D-double difference NOESY: application to the 44 kDa interferon-receptor complex. Journal of the American Chemical Society 133: 14755-14764.

23. Goldfarb V, Wittekind M, Jeffrey PD, Mueller L, Constantine KL (1993) Production and characterization of an antibody Fv fragment 15N-labeled in the VL domain only. J Mol Biol 232: 15-22.

24. Freund C (2006) Protein-Ligand Interaction Studied by NMR. Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine: Springer Berlin Heidelberg. pp. 1545-1548.

25. Schmidt PM, Rothkegel C, Wunder F, Schroder H, Stasch JP (2005) Residues stabilizing the heme moiety of the nitric oxide sensor soluble guanylate cyclase. Eur J Pharmacol 513: 67-74.

26. Schmidt PM, Schramm M, Schroder H, Wunder F, Stasch JP (2004) Identification of residues crucially involved in the binding of the heme moiety of soluble guanylate cyclase. J Biol Chem 279: 3025-3032.

27. Erbil WK, Price MS, Wemmer DE, Marletta MA (2009) A structural basis for H-NOX signaling in Shewanella oneidensis by trapping a histidine kinase inhibitory conformation. Proc Natl Acad Sci U S A 106: 19753-19760.

28. Ma X, Sayed N, Beuve A, van den Akker F (2007) NO and CO differentially activate soluble guanylyl cyclase via a heme pivot-bend mechanism. Embo j 26: 578-588.

29. Pellicena P, Karow DS, Boon EM, Marletta MA, Kuriyan J (2004) Crystal structure of an oxygen-binding heme domain related to soluble guanylate cyclases. Proc Natl Acad Sci U S A 101: 12854-12859.

30. Wang J, Zuo X, Yu P, Byeon IJ, Jung J, et al. (2009) Determination of multicomponent protein structures in solution using global orientation and shape restraints. J Am Chem Soc 131: 10507-10515.

31. Mueller GA, Choy WY, Yang D, Forman-Kay JD, Venters RA, et al. (2000) Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein. J Mol Biol 300: 197-212.

32. Giesen A, Homans S, Brown J (2003) Determination of protein global folds using backbone residual dipolar coupling and long-range NOE restraints. Journal of Biomolecular NMR 25: 63-71.

33. Popovych N, Sun S, Ebright RH, Kalodimos CG (2006) Dynamically driven protein allostery. Nature structural & molecular biology 13: 831-838.

34. Haase T, Haase N, Kraehling JR, Behrends S (2010) Fluorescent fusion proteins of soluble guanylyl cyclase indicate proximity of the heme nitric oxide domain and catalytic domain. PLoS One 5: e11617.

35. Busker M, Neidhardt I, Behrends S (2014) Nitric oxide activation of guanylate cyclase pushes the alpha1 signaling helix and the beta1 heme-binding domain closer to the substrate-binding site. J Biol Chem 289: 476-484.

36. Fritz BG, Roberts SA, Ahmed A, Breci L, Li W, et al. (2013) Molecular model of a soluble guanylyl cyclase fragment determined by small-angle X-ray scattering and chemical cross-linking. Biochemistry 52: 1568-1582.

37. Campbell MG, Underbakke ES, Potter CS, Carragher B, Marletta MA (2014) Single-particle EM reveals the higher-order domain architecture of soluble guanylate cyclase. Proc Natl Acad Sci U S A 111: 2960-2965.

38. Underbakke ES, Iavarone AT, Chalmers MJ, Pascal BD, Novick S, et al. (2014) Nitric oxide-induced conformational changes in soluble guanylate cyclase. Structure 22: 602-611.

39. Kempf JG, Loria JP (2003) Protein dynamics from solution NMR: theory and applications. Cell Biochem Biophys 37: 187-211.

40. Jarymowycz VA, Stone MJ (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. Chem Rev 106: 1624-1671.

41. Igumenova TI, Frederick KK, Wand AJ (2006) Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. Chem Rev 106: 1672-1699.
42. Stone MJ (2001) NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding. Accounts of chemical research 34: 379-388.
43. Vold RL, Waugh JS, Klein MP, Phelps DE (1968) Measurement of Spin Relaxation in Complex Systems. The Journal of Chemical Physics 48: 3831-3832.
44. Chen K, Tjandra N (2012) The use of residual dipolar coupling in studying proteins by NMR. Top Curr Chem 326: 47-67.
45. Neudecker P, Lundström P, Kay LE (2009) Relaxation Dispersion NMR Spectroscopy as a Tool for Detailed Studies of Protein Folding. Biophysical Journal 96: 2045-2054.
46. Saito H, Ishida M, Yokoi M, Asakura T (1990) Dynamic features of side chains in tyrosine and serine residues of some polypeptides and fibroins in the solid as studied by high-resolution solid-state carbon-13 NMR spectroscopy. Macromolecules 23: 83-88.