STRUCTURAL STUDIES OF SOLUBLE GUANYLYL CYCLASE
AND ITS BACTERIAL HOMOLOGS

By

Xiao-lei MA

Submitted in partial fulfillment of the requirements
For the degree of Doctor of Philosophy

Thesis Advisor: Dr. Focco van den Akker

Department of Biochemistry
CASE WESTERN RESERVE UNIVERSITY

May, 2008
We hereby approve the thesis/dissertation of

Xiaolei MA

candidate for the Doctor of Philosophy degree *.

(signed) Menachem Shoham

(chair of the committee)

Focco van den Akker

Paul R. Carey

Irene Nga Wing Lee

Saurav Misra

(dated) March 25, 2008

*We also certify that written approval has been obtained for any proprietary material contained therein.
Dedicated to my parents
Guang-de MA and Yun HU

献给我的父母

马广德和胡芸
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMMITTEE SIGNATURE SHEET</td>
<td>2</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>3</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>9</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>10</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>13</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>15</td>
</tr>
<tr>
<td>LIST OF PDB ID</td>
<td>16</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>17</td>
</tr>
<tr>
<td>CHAPTER 1: Structural biology approaches to unravel soluble guanylyl cyclase signaling and function</td>
<td>19</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>19</td>
</tr>
<tr>
<td>1.2 A short history of NO-sGC-cGMP signaling cascade</td>
<td>20</td>
</tr>
<tr>
<td>1.3 Isoforms and tissue distribution</td>
<td>21</td>
</tr>
<tr>
<td>1.4 Structure of soluble guanylyl cyclase</td>
<td>22</td>
</tr>
<tr>
<td>1.4.1 H-NOX domain</td>
<td>22</td>
</tr>
<tr>
<td>1.4.1.1 The discovery of H-NOX domain in sGC and in prokaryotes</td>
<td>22</td>
</tr>
<tr>
<td>1.4.1.2 H-NOX gene family display distinct ligand preference and UV/Vis spectroscopic characteristics</td>
<td>23</td>
</tr>
<tr>
<td>1.4.1.3 Crystal structures of H-NOX domain from strict anaerobe and facultative aerobe</td>
<td>24</td>
</tr>
</tbody>
</table>
1.4.2 H-NOXA domain……………………………………………………………..25
  1.4.2.1 The discovery of H-NOXA domain in sGC and in prokaryotes……25
  1.4.2.2 The H-NOXA domain adopts PAS-like fold…………………….26
1.4.3 CC region……………………………………………………………………27
1.4.4 Catalytic domain…………………………………………………………….28
1.5 Current knowledge on sGC activation mechanism…………………………28

CHAPTER 2: The basic procedure of protein X-ray crystallography………………….33
  2.1 Introduction……………………………………………………………………33
  2.2 Protein preparation……………………………………………………………33
  2.3 Crystallization……………………………………………………………………34
    2.3.1 Screens……………………………………………………………………..34
    2.3.2 Experimental setup…………………………………………………………35
  2.4 Diffraction data collection……………………………………………………36
    2.4.1 Crystal handling…………………………………………………………….36
    2.4.2 X-ray sources………………………………………………………………37
    2.4.3 Recording the reflections and data processing…………………………38
  2.5 Phasing……………………………………………………………………………39
    2.5.1 Direct Methods……………………………………………………………..39
    2.5.2 Molecular replacement……………………………………………………39
    2.5.3 Anomalous scattering (MAD or SAD) ………………………………..40
    2.5.4 Heavy atom methods (MIR or SIR) ………………………………..41
  2.6 Model building and refinement………………………………………………41
  2.7 Validation………………………………………………………………………..42
CHAPTER 3: Crystal structure of a Heme–Nitric oxide and OXygen binding (H-NOX) domain from *Nostoc sp.*: the common structural module used for Nitric Oxide (NO) sensing in bacterial NO sensor and mammalian soluble guanylate cyclase……………..47

3.1 Introduction…………………………………………………………………………...47

3.2 Materials and methods………………………………………………………………47

3.2.1 Cloning and mutagenesis of *Ns* H-NOX…………………………………48
3.2.2 Expression and purification of *Ns* H-NOX…………………………………48
3.2.3 UV/Vis absorption spectroscopic analysis of *Ns* H-NOX………………..48
3.2.4 Crystallization and data collection of *Ns* H-NOX………………………..49
3.2.5 *Ns* H-NOX structure determination and refinement…………………..49
3.2.6 Mutagenesis of rat sGC and transfection in COS-7 cells………………51
3.2.7 Cytosol preparation and Western blot analysis…………………………51
3.2.8 sGC activity assay……………………………………………………………52

3.3 Results and discussion…………………………………………………………...52

3.3.1 Structure of the *Ns* H-NOX domain………………………………………52
3.3.2 The *Ns* H-NOX domain in the NO and CO bound form……………..53
3.3.3 Conformation changes upon ligands binding……………………………..54
3.3.4 Heme pivoting and heme-bending activation mechanism………………55
3.3.5 Mutagenesis studies to test H-NOX activation mechanism…………….56

3.4 Conclusions…………………………………………………………………………56

CHAPTER 4: Crystal structure of the dimerized PAS-like domain from a *Nostoc punctiforme* signal transduction histidine kinase (STHK): Implications for soluble
guanylate cyclase dimerization and organization………………………………………..69

4.1 Introduction……………………………………………………………………..69

4.2 Materials and methods………………………………………………………70

4.2.1 Cloning and mutagenesis of the Np STHK H-NOXA domain………....70

4.2.2 Expression and purification of Np STHK H-NOXA domain…………..70

4.2.3 Crystallization and data collection of Np STHK H-NOXA domain……71

4.2.4 Np STHK H-NOXA domain structure determination and refinement…..71

4.2.5 Subcloning, mutagenesis, expression, and purification of the H-NOXA
domain of sGCβ1………………………………………………………..72

4.2.6 Mutagenesis of Rat sGC and Transfection in COS-7 Cells……………...73

4.2.7 Cytosol Preparation and Western Blot Analysis…………………………...73

4.2.8 sGC Activity Assay…………………………………………………………...73

4.3 Results and discussion……………………………………………………….....74

4.3.1 The PAS fold………………………………………………………………..74

4.3.2 Np STHK H-NOXA dimer organization…………………………………...75

4.3.3 sGC H-NOXA forms a similar PAS-like dimer…………………………...75

4.3.4 Possible implications for preferential heterodimerization of sGC……..77

4.3.5 Possible role for PAS-dimer in signaling or allosteric regulation in sGC
and Np STHK…………………………………………………………………………78

4.3.6 Np STHK-type H-NOXA dimer in sGC and the implications for sGC
domain organization…………………………………………………………...79

4.4 Conclusions………………………………………………………………………79

CHAPTER 5: Cloning, purification and crystallization of the Coiled Coil (CC) region
from *Rattus norvegicus* soluble guanylyl cyclase (sGC) β1 subunit

5.1 Introduction

5.2 Materials and methods

5.2.1 Cloning and mutagenesis of rsGCβ1 CC domain

5.2.2 Expression and purification of rsGCβ1 CC domain

5.2.3 rsGCβ1 CC domain crystallization

5.2.4 rsGCβ1 CC domain structure determination and refinement

BIBLIOGRAPHY
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Data collection, phasing, and refinement statistics for ( \text{Ns} ) H-NOX</td>
<td>67</td>
</tr>
<tr>
<td>2.2</td>
<td>Guanylyl cyclase activity measurements of wild-type and mutants of rat sGC under basal and NO-stimulated conditions</td>
<td>68</td>
</tr>
<tr>
<td>3.1</td>
<td>Data collection, phasing, and refinement statistics for ( \text{Np} ) STHK</td>
<td>88</td>
</tr>
<tr>
<td>5.1</td>
<td>Data collection, phasing, and refinement statistics for rat ( \text{sGC}\beta1 ) CC domain</td>
<td>94</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Domain organization of sGC</td>
</tr>
<tr>
<td>1.2</td>
<td>Model for NO-binding and activation of sGC</td>
</tr>
<tr>
<td>1.3</td>
<td>Proposed models for activation of sGC (A) by PPIX (B), NO (C) and BAY 58-2667 (D)</td>
</tr>
<tr>
<td>2.1</td>
<td>Protein crystallization flowchart involves the following steps</td>
</tr>
<tr>
<td>2.2</td>
<td>Crystallization experiment setup methods</td>
</tr>
<tr>
<td>2.3</td>
<td>Se absorption edge plot from 12600eV to 12700eV with the edge, peak and short/long remote labeled as ( \lambda_1 ), ( \lambda_2 ) and ( \lambda_3 )</td>
</tr>
<tr>
<td>3.1</td>
<td>Structure and domain organization of H-NOX and H-NOXA containing proteins. (A) Schematic diagram of the H-NOX and H-NOXA domains present in proteins from animals and <em>Nostoc</em> cyanobacteria. (B) Structure-based sequence alignment of HNOX domains</td>
</tr>
<tr>
<td>3.2</td>
<td>Crystal structure of unliganded <em>Ns</em> H-NOX. (A) Schematic diagram of the structure of <em>Ns</em> H-NOX. (B) Electronic absorption spectra of wild-type and mutants of <em>Ns</em> H-NOX. (C) Stereo figure showing electron density for the heme group in the free <em>Ns</em> H-NOX structure. (D) Superposition of the <em>Ns</em> and <em>Tt</em> H-NOX structures revealing an N-terminal domain shift for helices ( \alpha A-\alpha C )</td>
</tr>
<tr>
<td>3.3</td>
<td>Structural consequences of NO and CO binding to <em>Ns</em> H-NOX. (A) NO binding to the distal face of the heme in the NO bound <em>Ns</em> H-NOX structure. (B) As above, but for CO binding to H-NOX, omit density contoured at 4( \sigma ). (C)</td>
</tr>
</tbody>
</table>
Molecular shifts within the heme region of Ns H-NOX upon NO and CO binding.
(D) CO binding to Ns H-NOX leads to increased flexibility in the loop following helix αF and C-terminal part of this helix……………………………………….62

3.4 Evidence for heme pivoting and heme bending in activation mechanism of H-NOX. (A) Top (distal) view of the heme groups of the superimposed H-NOX structures. (B) Side-view stereo figure of the heme groups of the superimposed HNOX structures. (C) Western blot analysis of wild-type and mutant sGC proteins……………………………………………………………………64

3.5 Schematic diagram of NO and CO dependent activation of sGC………………66

4.1 Structure of H-NOXA domain of Np STHK and its evolutionary, structural, and sequence relationships. (A) Schematic diagram of the H-NOX and H-NOXA domains present in proteins from animals and Nostoc cyanobacteria. (B) Structure-based sequence alignment of the H-NOXA domains. (C) Ribbon diagram of the Np STHK domain monomer structure. (D) Stereo-figure depicting superposition of the monomer of the H-NOXA domain of Np STHK and the Ec DOS PAS domain (PDBid 1V9Z)…………………………………..80

4.2 Homo-dimeric structure of H-NOXA domains of Np STHK. (A) Schematic diagram of the Np STHK dimer in two side views related to each other by a 90° rotation around the axis shown. (B) Stereo diagram of the dimer interface. (C) Same as in (B) except rotated 90° along the horizontal axis such that the 2 N-terminal helices are now oriented vertically. (D) Electron density of the N-terminal dimer interface region of Np STHK. (E) Different dimer configurations of full length and truncated Np STHK…………………………………..82

4.3 Structural similarity of the Np STHK dimer and other PAS domain dimer
4.4 Probing the putative H-NOXA dimer interface in sGC by mutagenesis. (A) Conservation of surface residues of the *Np* STHK H-NOXA monomer with the α1 and β1 sGC subunits. (B) Probing dimerization state of sGC β1 H-NOXA domain mutants using size exclusion chromatogram. (C) Guanylyl cyclase activity and Western blot analysis of co-immunoprecipitated wt and 6 sGC β1 mutants of sGC. (D) Guanylyl cyclase activity measurements and Western blot co-immunoprecipitation analysis of additional mutants of sGC generated in both α1 and β1 subunits.

4.5 Possible subunit arrangement of sGC
ACKNOWLEDGMENTS

Above all, it is with my deepest and sincere gratitude that I thank my advisor, Dr. Focco van den Akker, for allowing me the opportunity to study such a challenging project in the field of structural biology. I will be forever indebted to his careful guidance, his open mindedness, his extended patience, and his invariable availability. Most importantly, I am grateful to his enthusiasm and encouragement in times of difficulty. I am honored to be his first graduate student to get a Ph.D. degree and I am sure there will be many more to follow.

I would also like to thank Dr. Vivien Yee for her persistent encouragement for me to pursue my aspirations of obtaining my Ph.D.

A special thank you is reserved for Dr. Priyaranjan Pattanaik, who joined Dr. van den Akker’s lab as a research associate two weeks before I did. In the initial stages of my research, I benefited greatly from his help. He was second only to Dr. van den Akker as far as offering me advice to overcome the problems encountered in this project. I valued a lot our friendship, our numerous scientific discussions and your precious hints on life.

I am extremely thankful to Dr. Pius S. Padayatti. He has introduced me to many aspects of protein crystallography and showed me the beauty of macromolecular structures. It was a rewarding experience for me to work with him since I am now able to solve the structures independently.

Many thanks are also extended to current and former lab mates in Dr. van den Akker and Dr. Yee’s labs including Dennis Wilk, Jared Sampson, Kaustubha Qanungo, Laura Fromondi, Lucas Donovan, Pamela Hall, Rajaganapathi, Rune Hartmann, Seungjoo Lee, Wei Ke, Yi Peng: for many valuable suggestions that have aided me in the completion of
this project, as well as a wonderful work environment and technical support.

I am also grateful to my dissertation committee Dr. Irene Lee, Dr. Menachem Shoham, Dr. Misra Saurav, and Dr. Paul Carey for their valuable insights and much needed advice throughout my graduate career.

I wish to extend my appreciation to our collaborators, in particular Dr. Annie Beuve and her group who carried out extensive biochemical and enzymatic characterization on sGC that complement our structural studies.

I would like to thank Dr. Jay Nix at ALS 4.2.2 for his instrumental support and help with MAD/SAD data collection. Had it not been for his assistance, I would not have been able to complete my graduate studies as timely as I did.

In addition I gratefully acknowledge those who have funded this work including the American Heart Association and the National Institute of Health.

It goes without saying that my most heart-felt thanks goes to my parents for their continual and unlimited support. It is amazing how much I have continued to learn from them, in spite of the fact they have no idea what x-ray crystallography is. I hope this dissertation and the degree make them feel proud of their son.
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-CHSR</td>
<td>two-component hybrid sensor and regulator</td>
</tr>
<tr>
<td>ALS</td>
<td>advanced light source</td>
</tr>
<tr>
<td>APS</td>
<td>advanced photon source</td>
</tr>
<tr>
<td>cAMP</td>
<td>cyclic adenosine 3’,5’-monophosphate</td>
</tr>
<tr>
<td>CC</td>
<td>Coiled coil</td>
</tr>
<tr>
<td>cGMP</td>
<td>cyclic guanosine 3’,5’-monophosphate</td>
</tr>
<tr>
<td>CNBD</td>
<td>cyclic nucleotide-binding domains</td>
</tr>
<tr>
<td>CO</td>
<td>carbon monoxide</td>
</tr>
<tr>
<td>EDRF</td>
<td>endothelium derived relaxing factor</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine 5’-triphosphate</td>
</tr>
<tr>
<td>H-NOX</td>
<td>Heme nitric oxide and oxygen binding</td>
</tr>
<tr>
<td>H-NOXA</td>
<td>Heme nitric oxide and oxygen binding associated</td>
</tr>
<tr>
<td>NO</td>
<td>nitric oxide</td>
</tr>
<tr>
<td>NOS</td>
<td>nitric oxide synthase</td>
</tr>
<tr>
<td>NSLS</td>
<td>national synchrotron light source</td>
</tr>
<tr>
<td>ODQ</td>
<td>1H-[1,2,4]oxadiazolo[4,3-a]quinoxalin-1-one</td>
</tr>
<tr>
<td>PAS</td>
<td>an acronym formed from three proteins: Per, Arnt, Sim</td>
</tr>
<tr>
<td>PDB</td>
<td>protein data bank</td>
</tr>
<tr>
<td>sGC</td>
<td>soluble guanylyl cyclase</td>
</tr>
<tr>
<td>STHK</td>
<td>signal transduction histidine kinase</td>
</tr>
<tr>
<td>YC-1</td>
<td>3-(5’-hydroxymethyl-2’-furyl)-1-benzyl indazole</td>
</tr>
</tbody>
</table>
LIST OF PDB ID

2O09
Crystal structure of the H-NOX domain from Nostoc sp. PCC 7120

2O0C
Crystal structure of the H-NOX domain from Nostoc sp. PCC 7120 complexed to NO

2O0G
Crystal structure of the H-NOX domain from Nostoc sp. PCC 7120 complexed to CO

2P04
2.1 Ang structure of the dimerized PAS domain of signal transduction histidine kinase from Nostoc punctiforme PCC 73102 with homology to the H-NOXA/H-NOBA domain of the soluble guanylyl cyclase

2P08
Structure of the N-terminally truncated PAS domain of signal transduction histidine kinase from Nostoc punctiforme PCC 73102 with homology to the H-NOXA/H-NOBA domain of the soluble guanylyl cyclase
Abstract

By

Xiao-lei MA

The most well known receptor for NO is soluble guanylate cyclase (sGC), which converts guanosine-5’-triphosphate (GTP) to cyclic guanosine-3’, 5’-monophosphate (cGMP) upon NO activation. cGMP serves as an important second messenger to regulate a wide range of physiological process in cardiovascular and neurological systems. This pathway is targeted for therapeutic purposes as Glyceryl trinitrate (GTN), which generates the signaling molecule NO inside the cell, has been used to treat angina pectoris and heart failure for more than a century. sGC is a heterodimeric hemoprotein composed of two different subunits: α (73-82 kDa) and β (70-76 kDa). Both subunits share a similar domain organization: an N-terminal domain (which harbors a heme only in β1 called H-NOX), a central H-NOXA domain, a Coiled-coil (CC) domain and a C-terminal guanylyl cyclase domain. An appreciation of the atomic details of the mechanism by which NO activates sGC is critical to our understanding of how such a small gaseous molecule NO is able to recognized by the target cell and in turn elicit a wide variety of responses. This thesis is focused on the structural characterization of H-NOX domain, H-NOXA domain and CC domain, little structural information was available by the time this project was initiated.
The first target was the stand-alone H-NOX domain from *Nostoc sp* PCC 7120, which shares 33% sequence identity with human sGCβ1. We have determined three *Ns* H-NOX structures including unliganded, NO bound and CO bound form at 2.1Å, 2.5Å and 2.6Å respectively. We have identified a critical aromatic residue above the heme plane conferring binding advantage to NO over CO by steric hindrance. Comparison of these structures and previously published *Tt* H-NOX structure has revealed a heme pivot-bend mechanism that correlates with the H-NOX structural changes with respect to their presumed activation state.

The second target was the H-NOXA domain from a STHK gene in the *Nostoc punctiforme* PCC 73102 genome, which is 35% identical to human sGCβ1 H-NOXA domain. We have determined the *Np* STHK N-terminal H-NOXA domain in a dimerized form at 2.1Å. The monomer adopts a PAS-like fold and dimerizes in a similar fashion as observed in other PAS dimers by juxtapositioning of the N-terminal helices with their preceding residues and the face of the β-sheet. We proposed sGC H-NOXA domain could hetero-dimerize in the same fashion and we confirmed our hypothesis by Ala scanning mutagenesis studies. The structural information provides broad implications including PAS-mediated dimerization for preferential heterodimerization of sGC, allosteric regulation of sGC and *Np* STHK, and sGC domain organization.

The third target was the CC domain from rat sGC. We have obtained single crystals for this domain and successfully prepared selenomethionine substituted crystals for phasing. However due to the presence of pseudo merohedral twinning in both of the crystal forms, structural determination and refinement is still in progress.
CHAPTER 1

Structural biological approaches to unravel soluble guanylyl cyclase signaling and function

1.1 Introduction

sGC is the primary receptor of the signaling molecule nitric oxide (NO). As a heterodimer of α and β subunits, sGC has the ability to convert guanosine-5’-triphosphate (GTP) to cyclic guanosine-3’,5’-monophosphate (cGMP) upon activation. The second messenger cGMP in turn modulates the activity of cGMP-regulated ion channels, cGMP-dependent protein kinases, and cGMP-regulated phosphodiesterases. These effectors are involved in the regulation of several physiological functions in controlling smooth muscle relaxation, inhibiting platelet aggregation, and modulating synaptic transmission.1,2

In spite of sGC being the central player in the NO-cGMP pathway, most of studies have been focusing on aspects such as the spectral, enzymatic and NO-binding properties of sGC. A detailed understanding of sGC signaling and function at molecular level has been hampered by the lack of precise structural information mainly due to the difficulty in obtaining large amounts of homogenous enzyme from recombinant expression system or from the native source. The recent explosion of microbial genomic data and the improvement in bioinformatics algorithms provide novel opportunities by identifying sGC prokaryotic domain homologues such as the H-NOX and H-NOXA domain containing genes in cyanobacteria sources.3,4 The recent determination of the \(N_s\) H-NOX and \(N_p\) H-NOXA domain crystal structures (Figure 1.1) has provided unique insights...
into the diatomic gaseous molecules ligand discrimination, heme and H-NOX protein movement upon ligand binding, the fold of H-NOXA and the PAS-mediated dimerization.

The following review will attempt to recapture the highlights of current knowledge regarding sGC structure-function relationships from a structural perspective.

1.2 A short history of NO-sGC-cGMP transduction system

The studies of NO-sGC-cGMP pathway have a long and distinguished history dating back almost 50 years. The discoveries of cAMP and cGMP as mammalian cells metabolites were made around the same time\textsuperscript{3,5,6}. The GPCR-adenylyl cyclase (AC)-cAMP cascade controlling a diverse range of cellular processes was extensively studied while the progress in understanding the physiological significance of sGC-cGMP signaling has been comparatively slow\textsuperscript{7}. Although the enzyme responsible for cGMP production in mammalian tissues was identified ten years later\textsuperscript{8,9}, the endogenous ligand of this enzyme remained elusive for more than another decade. In 1980 a hormone produced by endothelium namely EDRF was shown to be indispensable for ACh-induced vasodilation in vivo\textsuperscript{10}. Several years later, the chemical nature of EDRF was identified to be the free radical NO\textsuperscript{11-13}. Shortly afterwards the biosynthesis of NO was traced to a flavoprotein nitric oxide synthase (NOS) utilizing L-arginine as the substrate\textsuperscript{14,15}. Interestingly, physicians began to prescribe nitroglycerin to alleviate chest pain more than 100 years before it was clarified that these vasodilators acts by releasing NO gas to stimulate cGMP formation. Nitric oxide (NO) was proclaimed the molecule of the year by "Science" in 1992. Robert F. Furchgott, Louis J. Ignarro and Ferid Murad were awarded 1998 Nobel Prize in physiology or medicine for their discoveries concerning
nitric oxide as a signaling molecule in the cardiovascular system.

1.3 Isoforms and their heterologous expression efforts

Four independent genes termed α1, α2 and β1, β2 that belong to the large and small subunit of a functional sGC have been cloned and located in the human genome. Although up to 10 homodimeric and heterodimeric combinations of sGC subunits are theoretically possible, only three of them have been shown to be catalytically active: α1/β1, α2/β1 heterodimer and β2/β2 homodimer. The two heterodimeric forms share similar kinetic properties, sensitivities to NO and pharmacological response to ODQ and YC-16. The α1/β1 isoform is the most abundant form in most of mammalian tissues except brain, where the highest expression of α2 subunit was detected, suggesting a putative role in synaptic transmission17,18. β2 subunit is unlikely to exist at the protein level in human due to the presence of a frameshift mutation in the coding sequence, which was absent in non-human primates genes19. A novel variant of the rat β2 subunit has been shown to display NO-sensitive GC activity when expressed in the absence of a second subunit in Sf9 cells20. In addition to rat β2, two atypical soluble guanylyl cyclase genes from insects Manduca sexta β3 and Drosophila Gyc-88E also forms functional homodimer when heterologous expressed21,22.

Several labs (including ours) have attempted to develop expression systems to bypass laborious and problematic sGC purification from bovine lung23,24. Mammalian expression cell lines such as COS cells and HEK293 cells have been used to transiently or stably express sGC subunit (s) for GC activity assay and mutational analysis25-27. Baculovirus/Sf9 expression system for sGC has been developed by different groups in order to generate ample amounts of sGC suitable for more extensive
biochemical/structural characterization\textsuperscript{28-30}. \textit{E. coli} expression system has been extensively tested to generate soluble sGC fragments representing single function domain or domains in combination. The successfully overexpressed sGC constructs include the heme binding region $\beta_1$ 1-385, $\beta_1$ 1-194, $\beta_2$ 1-217\textsuperscript{31-34}, the GC domain of $\alpha_1/\beta_1$ heterodimer $\alpha_1$ 467-690, $\beta_1$ 414-619\textsuperscript{35} in Marletta’s lab and the dimerization domain $\beta_1$ 202-344; the coiled coil region $\beta_1$ 348-409; the $\alpha_1$ N-terminal domain $\alpha_1$ 66-260 in our lab. Nevertheless, no crystal structure of sGC or any of its functional domains has been reported.

1.4 Structure of soluble guanylyl cyclase

1.4.1 H-NOX domain

1.4.1.1 The discovery of H-NOX domain in sGC and in prokaryotes

sGC has been known as a heme containing protein since the early 1980s\textsuperscript{36,37}, and shown to be activated by NO binding to the heme Fe center\textsuperscript{38}. The stoichiometry of heme per enzyme unit has been definitely determined to be one heme per heterodimer\textsuperscript{39}. The deletion of the first 259 amino acids of human sGC $\alpha_1$ has little effect on heme binding or sensitivity to NO\textsuperscript{40} while the \textit{E. coli} expressed $\beta_1$ 1-385 contained one equivalent of heme per monomer\textsuperscript{31}. In addition, this $\beta_1$ fragment display nearly identical spectral properties to those observed with heterodimeric sGC purified from native source\textsuperscript{41}. These results have demonstrated that heme binds to the N-terminal region of the $\beta_1$ subunit. Extensive mutational analysis has been performed to replace His105 of the $\beta_1$ subunits to alanine and glycine\textsuperscript{32} in $\beta_1$ 1-385 or to phenyl-alanine\textsuperscript{28} and cysteine\textsuperscript{42} in $\alpha_1/\beta_1$ heterodimer. All histidine mutations render sGC $\beta_1$ 1-385 or $\alpha_1/\beta_1$ heterodimer heme-deficient enzyme indicating His105 is the proximal ligand of the heme group.
Moreover, both α1/H105Fβ1 and α1/H105Cβ1 lack NO-responsiveness although α1/H105Cβ1 exhibits higher basal activity while α1/H105Fβ1 retained near wild type activity. The minimal domain structures still capable of heme incorporation have been mapped to β1 1-194 and β2 1-217\(^\textsuperscript{33}\) only recently.

As the ~200 residues N-terminal extension of sGC subunits does not map to any previously known domains, a sequence profile search were performed using sGCβ1 1-185 as seed\(^\textsuperscript{4}\). This PSI-BLAST\(^\textsuperscript{43}\) search was run to convergence and detected a family of bacterial proteins with significant homology (15-40\%) to the heme domain of sGC. This cluster of proteins is present in, other than sGC, several distinct contexts with other domains in the same polypeptide such as methyl-accepting chemotaxis receptors from obligate anaerobes, or in a stand-alone form in a histidine kinase operon from facultative aerobes. Due to their divergent ligand preference (reviewed in the next section), Marletta has termed this family the H-NOX (heme nitric oxide and/or oxygen binding)\(^\textsuperscript{44}\), although other names such as HNOB (heme NO binding)\(^\textsuperscript{4}\) or SONO (sensor of NO)\(^\textsuperscript{45}\) have been used solely indicative of NO binding.

1.4.1.2 H-NOX gene family display distinct ligand preference and UV/Vis spectroscopic characteristics

Extensive efforts have been devoted to investigate the ligand binding preference of H-NOX from different origins. Members of this family from strict anaerobe form extremely high affinity O\(_2\) complex and might function as an oxygen sensor. \(Tt\) H-NOX, part of a predicted methyl-accepting chemotaxis protein (MCP) from the strict anaerobe \textit{Thermoanaerobacter tengcongensis} forms a low-spin, 6-coordinate ferrous-oxy complex, the first of this H-NOX family that binds O\(_2\). In contrast members that belong to the
facultative aerobes prefer NO or CO as ligands while rigorously excluding O₂ similar to sGC. In the absence of NO or CO, the H-NOX proteins from facultative aerobes remain 5-coordinated even when high concentration of O₂ is present. CO binding resulted in a low-spin, 6-coordinated heme while NO binding triggers the Fe-His bond breakage forming high-spin, 5-coordinated heme in *Vibrio cholerae*<sup>44</sup>, *Legionella pneumophila* L1<sup>46</sup>, *Shewanella oneidensis*<sup>47</sup> MR-1 H-NOX, or in a temperature dependent manner in *Legionella pneumophila* L2, *Nostoc punctiforme*<sup>46</sup> H-NOX.

1.4.1.3 Crystal structures of H-NOX domain from strict anaerobe and facultative aerobe

The N-terminal regulatory domain in the β1 subunit is the best characterized part in sGC. The crystal structure of oxygen-binding H-NOX (heme-NO and oxygen binding) domain from a thermophilic organism *Thermoanaerobacter tengcongensis*, which shares 17% sequence identity with the first 180 residues of mammalian sGC, was determined in an oxygen-bound form<sup>45,48</sup>. In the full length protein, it is fused through a membrane-spanning region to a hypothetical methyl-accepting chemotaxis protein domain. The H-NOX consists of a proximal domain with an αββαββ motif and a distal domain with 4 α helixes. The heme is tightly sandwiched within a central pocket between the two domains. The proximal ligand to its iron is His-102 (His-105 in sGC) in helix α5, while the distal ligand is oxygen. Intensive interactions are observed between the propionate group of the heme and a “YxSxR” motif<sup>49</sup> in the protein. The phenolic –OH group of Tyr-140 is within hydrogen-bonding range of this oxygen ligand and the orientation of this particular Tyr-140 is constrained by a hydrogen bonding network including Trp-9 and Asn-74. The distal heme cavity is apolar and there are no polar atoms within 7 Å of the heme iron except Tyr-140. The ability of Tyr-140 in the distal cavity to
modulate the bond strength of the proximal Fe-His linkage is unprecedented in heme proteins\textsuperscript{50}.

Although the general fold is most likely preserved in the H-NOX family, specific amino acid changes in the heme proximity are presumably required to promote binding of one diatomic gas or another. To gain molecular insights into the NO and CO activation mechanism of H-NOX, we have determined \textit{Nostoc sp.} H-NOX crystal structures in the free, NO, and CO liganded \textit{Ns} H-NOX domain to 2.1 Å, 2.6 Å, and 2.5 Å respectively (see Chapter 3). Our structures together with structure-guided mutational and biophysical characterization provide evidence that the NO activation mechanism in sGC involves a stepwise heme pivot to facilitate a ~20° rotational shift of the N-terminal helical region in H-NOX causing the heme to bend in the activated 5-coordinated state. CO signals differently from NO by having a larger heme pivot shift without proceeding to the heme-bend 5-coordinated state.

1.4.2 H-NOXA domain

1.4.2.1 The discovery of H-NOXA domain in sGC and in prokaryotes

The central region of the sGC subunits is involved in the formation of homo/hetero-dimers, which is a pre-requisite for forming a functional enzyme. Bioinformatics analyses have predicted that sGC dimerization sites are localized at sequence segments spanning positions 312-377\textsuperscript{21} or 340-385\textsuperscript{31}. Indeed, using multiple N- and C-terminal \(\beta\)1 deletion constructs, two distinct segments contributing to \(\alpha\)1 binding have been mapped to residues 204-244 (N-terminal binding site, NBS) and residues 379-408 (C-terminal binding site, CBS)\textsuperscript{51}. Both sites are critical for sGC function and the CBS contains the major binding site, while the NBS seems to be an accessory or
regulatory binding site. Conclusive identification of the H-NOXA domain of sGC used a
construct encoding rsGCβ1 202-34452, which excludes the H-NOX domain and the
proceeding CC region. This small construct is expressed as a soluble homodimer in E.
coli, indicating this region of sGC is capable of folding properly on its own and is
responsible, at least in part, for the homo/hetero-dimerization of sGC subunits.

On the other hand, bacterial homologous searches have been beneficial revealing a
family of H-NOXA domains that always associated with the H-NOX domain either
resides in the same protein or proteins encoded by the same operon. Sequence-structure
threading based structural prediction program 3D-PSSM suggested a PAS-like fold for
the H-NOXA domain prior to any structural analysis4,53. Phylogenetic analyses have
pointed out the possibility that animal versions of H-NOX/H-NOXA were probably
acquired by horizontal transfer from a cyanobacterial source3,4.

1.4.2.2 The H-NOXA domain adopts PAS-like fold

The α1 and β1 H-NOXA domains are 38% sequence identical and also share ~35%
sequence identity with the H-NOXA domains in the STHK from Nostoc punctiforme
PCC 73102 and the 2-CHSR from Nostoc sp. PCC 7120. Unlike the sGC subunits, the
H-NOXA domain of the STHK was amendable to crystallographic analysis. We have
determined the 2.1 Å crystal structure of the dimerized H-NOXA domain of Np STHK
(see Chapter 4). STHK PAS-like domain adopts a few α-helices with a 7-stranded
anti-parallel β-barrel, while a five-stranded anti-parallel β sheet embedding consecutive 4
α-helices is usually contained in a canonical PAS domain fold. The difference of this
structure to other known PAS domains is demonstrated by the high-backbone R.M.S.D
values (2.8-3.3 Å) of pairwise comparisons between our current structure and the most
similar ones in the protein data bank (PDB) according a DALI search.

1.4.3 CC region

The coiled-coil is a ubiquitous structural motif in proteins, which plays essential roles in protein assembly and molecular recognition. Notable examples of CCs include the basic-leucine zippers (bZIP), viral fusion peptides (HIV), SNARE complexes, certain tRNA synthetases and tropomyosin. The CC usually adopts a long helical conformation consisting of multiple copies of repeated seven amino acid residues pattern known as heptad repeats. Between 2 and 7 CC monomers intertwine each other to form a higher order super helical coil by segregating hydrophobic as well as some polar residues between helices, such as leucine arranged in a so called leucine zipper. The resulting CCs may be parallel or antiparallel, and usually arranged in a left hand fashion. Although energetically disfavored, right-handed coiled coils have also been reported in naturally occurring proteins and in designed proteins.

sGC α/β subunits and pGC have been known to harbor such CC elements according to primary structural analysis using program COILS. A recent in silico study has attributed the GC CC elements to a novel class of signaling modules after careful sequence analysis. This well-conserved version of CC with distinct sequence features is termed the Signaling helix (S-helix), given its exclusive co-occurrence with other major signaling domains. According to domain architecture analysis, most frequently S-helix bridges N-terminal sensory domains or extra-cellular sensor domains connected via transmembrane helices and various C-terminal catalytic domains. Alternatively S-helix is found between two small-molecule binding domains of same or different types. In comparisons to the great diversity of CC in bacterial lineages, S-helix has been only
detected in different classes of GCs in animals. The S-helices in GCs were demonstrated to be related to specific bacterial forms despite limited sequence similarity, suggesting the CC of GC arose in bacteria and was acquired later along with genes encoding other signaling modules.

1.4.4 Catalytic domain

The C-terminal catalytic domain exhibits a high degree of homology among sGC subunits, the C-terminal regions of pGC and Adenylyl cyclase (AC). Separately these domains form homodimers that are inactive when expressed in Sf9 or E. coli, but when co-expressed or combined following individual expression, they yield a catalytically active enzyme, which is unresponsive to NO. Similar phenomenon has been observed for the catalytic domains of the adenylyl cyclases. An enzymatically active enzyme can be obtained only when the C1 and C2 domains are co-expressed, while a homodimer formed by C2 domain is not active. The crystal structure of the catalytic domains of adenylyl cyclase shows the ATP binding site, Mg$^{2+}$ coordination site, transition state stabilization site are formed by both the C1 and the C2 domains. Similarly, the α1 subunit is deprived of residues that stabilize the transition state (like C1), while β1 subunit lacks residues to coordinate the Mg$^{2+}$ ions (like C2). The sGCβ1 H-NOX domain inhibited the activity of the α$_{\text{cat}}$/β$_{\text{cat}}$ complex by direct interaction that is independent of NO. By analogy, Shewanella oneidensis H-NOX protein bound NO in a 5-coordinate complex and repressed a downstream sensor histidine kinase activity in trans.

1.5 Current knowledge on sGC activation mechanism
Recently an updated NO-binding and activation of sGC based on the reciprocal communication of the catalytic and ligand-binding domains have been described\(^\text{64}\) (Figure 1.2). A six-coordinated heme is generated after NO binding to the distal site of heme. In the presence of the products (Mg\(^{2+}\)/cGMP/PPi) the cleavage of histidine-heme happens immediately, which is thought to activate the enzyme by a series of yet unknown conformational changes so far. In the absence of the products, the six-coordinated heme will be trapped in a non-activated state by adding a second NO molecule at the proximal site. This model not only consolidates results from NO-binding and dissociation/deactivation of sGC but also agrees with an early resonance Raman spectroscopy analysis, which indicated GTP binds proximally to the bound NO thus potentially regulating NO binding\(^\text{65}\). Important sGC modulators include 1H-[1,2,4]oxadiazolo[4,3-a]-quinoxalin-1-one (ODQ), 3-(5’-hydroxymethyl-2’-furyl)-1-benzylindazole (YC-1) and BAY 58-2667\(^\text{18}\). ODQ oxidizes the heme iron and therefore abolishing sGC activity. YC-1 and BAY 58-2667 (Figure 1.3) on the other hand can activate sGC independent of NO. Nevertheless, there are no definitive conclusions about the allosteric sites involved in NO sensitization of sGC.
Figure 1.1 Domain organization of sGC. Each subunit of sGC contains 4 functional domains as determined by BLAST searches, sequence alignments and structure prediction: an N-terminal H-NOX domain, a predicted PAS-like domain, a putative amphipathic helix, and the C-terminal cyclase catalytic domain.
Figure 1.2 Model for NO-binding and activation of sGC (Figure taken from Ref64). NO binding to the distal site of heme generates a six-coordinated state (with soret maximum at 420nm), which can either be trapped in a non-activated state (399nm) in the absence of the enzyme’s products or result in the active state (399nm) if the reaction products are present.
Figure 1.3 Proposed models for activation of sGC (A) by PPIX (B), NO (C) and BAY 58-2667 (D) (Figure taken from Ref\textsuperscript{49}).
CHAPTER 2

The basic procedure of protein X-ray crystallography

2.1 Introduction

3D structural information of proteins, protein complexes bound to substrates or inhibitors, protein-protein complexes, protein-RNA/DNA complexes and RNA/DNA have become an extremely valuable asset in the modern biology by providing atomic level understanding of macromolecular function. Specifically, we can study how proteins interact with other molecules, how they undergo conformational changes, and how they perform catalysis in the case of enzymes. Armed with this information we can design novel drugs that target a particular protein, or rationally engineer an enzyme for a specific industrial process. There are a number of techniques for studying 3D structure such as nuclear magnetic resonance (NMR), electron microscopy (EM), neutron diffraction, powder diffraction, atomic force microscopy (AFM) and X-ray crystallography, among which X-ray crystallography remains the oldest yet arguably most powerful of these methods at present. My graduate studies have focused on using protein crystallography and other complement techniques to study structures of soluble guanylyl cyclase and related proteins. The basic procedure involved in a crystal structure determination is the following: protein preparation, crystallization, diffraction data collection, phasing, model building, refinement and validation (Figure 2.1).

2.2 Protein preparation

Ample amount of pure and homogenous protein sample is a prerequisite before any
crystallization trials. It can be obtained either by isolating from native source or recombinant expression. Ideally the final purified protein sample need to meet the following criteria before crystallization experiment can be carried out: water soluble, properly folded, maintains its biological function, stable when stored for an extended time frame, relatively pure, and monodisperse.

2.3 Crystallization

2.3.1 Screens

In most cases, obtaining diffraction quality crystals from biological macromolecules is the major obstacle in solving its atomic-resolution structure. Although a great deal of efforts have been devoted to rational prediction of protein crystallization conditions\textsuperscript{66-69}, unfortunately macromolecule crystallizations still depends heavily on empirical screening approaches. Biological macromolecules generally have many degrees of freedom that require a very narrow range of chemical components, pH, temperature, sample purity to be crystallized. Currently the most efficient and standard approach to identify the initial crystallization condition is to use a sparse matrix sampling method\textsuperscript{70} to select reasonable numbers of trials from known or published crystallization conditions of various proteins in the past. However, provided a large batch of protein sample is present, a comprehensive search for the crystallization condition with a much finer grid is available as a joint effort of structure genomics projects and high-throughput crystallography\textsuperscript{71}.

During my graduate studies the following sets of commercially available screens were routinely used: Hampton research PCT\textsuperscript{72}, Crystal Screen 1\textsuperscript{70} and 2\textsuperscript{73}, INDEX\textsuperscript{74}, SaltRx\textsuperscript{75} and PEG/Ion\textsuperscript{76}.

If the sample fails to crystallize in any of the screening conditions, it may be
worthwhile tracing back to the expression and purification steps and try again with a slightly modified version of the protein molecule. Changing buffer conditions, addition of ligands, adding extra purification steps, making mutated or truncated construct, going to different expression systems are the common tricks to introduce some changes in macromolecular properties that might lead to huge differences in crystallization behavior.

2.3.2 Experimental setup

In this section I will offer an overview of the most popular techniques used to obtain crystals including sitting drop vapor diffusion, hanging drop vapor diffusion and microbatch under oil while more sophisticated setups are also occasionally used such as sandwich drop vapor diffusion, free interface diffusion, and microdialysis.

In sitting drop or hanging drop vapor diffusion (Figure 2.2A and B), a small (0.5-20 μl) droplet of protein sample is mixed with 0.5-20 μl crystallization reagent, and placed on a platform in vapor equilibration with the reagent or on a siliconized glass cover slide inverted on top of the reservoir in vapor equilibration with the reagent. At the beginning, the concentration of the crystallization reagent in the droplet is less than that in the reservoir. As water molecules are pulled over to the reservoir via vapor diffusion process, the protein concentration as well as the crystallization reagent concentration gradually increases in the droplet, both of which promote the degree of saturation of protein sample. Under the ideal condition, crystals form. Both of the methods demand a small amount of sample, and allow easy access to the crystal once formed. Sitting drop is superior to hanging drop in terms of speed and simplicity, while hanging drop offers the ability to visualize the drop without the optical interference from plastic and reduced chance of crystals sticking to the surface of the plastic.
In microbatch under oil crystallization (Figure 2.2C), a small volume of proteins combined with the crystallization solutions is placed under a layer of Paraffin Oil and Silicon Oil mixture. Such oils allow water vapor to permeate from the drop thus sample and reagent is concentrated. The rate of evaporation can be easily controlled by the recipe of the oil mixture. Paraffin Oil allows little if any water permeability while pure Silicon Oil allows the maximum evaporation rate. This crystallization under oil method is well suitable for high-throughput method and require very low amount of protein samples, although the crystals are difficult to retrieve in some cases.

2.4 Diffraction data collection

Diffraction data collection as the final experimental step in crystallography includes mounting the crystals in the x-ray beam, selecting X-rays with appropriate wavelength and intensity, and measuring the intensity and pattern of the diffraction spots using an automated detector. Due to the difficulty of getting crystals, much effort is devoted toward maximizing the quantity and quality of the data that can be recorded from a single crystal. During the past 10-15 years, many techniques have been developed that have enhanced the rate of structure determination by crystallography. These include the general availability of synchrotron radiation sources, widely used cryogenic techniques, increased speed of computing, and advanced software.

2.4.1 Crystal handling

Once the crystals reach their full size, they can be mounted so that they can be held still while embraced in the X-ray beam and rotated. Traditionally, a crystal is mounted into glass capillaries and the mother liquor is dried afterwards allowing the crystal to
attach to the interior of the capillary. More recently, the loop mounting method developed by Teng\textsuperscript{77} become the most simple and standard approach as a major advancement in cryo crystallography (around 100K). A single crystal is placed in a tiny loop, made of plastic or nylon and attached to a solid rod base, and dunked immediately into a cryogen such as liquid nitrogen or propane. The main reason for using this cryo technique is to minimize X-ray induced radiation damage to the protein crystals, as well as reduced B factors due to thermal motion (the Debye-Waller effect). However an appropriate cryo-protectant containing both mother liquor and an anti-freeze agent need to be identified on a trial and error basis.

### 2.4.2 X-ray sources

Currently, most structures are solved using synchrotron X-rays owing to their higher beam intensity and tunability. In synchrotron facilities, electrons that travel at a speed that is close to the speed of light are held in a circular orbit through magnetic forces. The acceleration of the negatively charged electrons generates intense and broad electromagnetic radiation, including X-rays. X-ray at synchrotron sources are several orders of magnitude brighter than the beam generated by conventional rotating anodes, which allow data to be collected rapidly from very small crystals at better resolution. In addition, as the electromagnetic radiation in synchrotron is continuous, and with the help of a tunable monochromator it is convenient to change the wavelength to the optimal energy of choice, which is critical for multi-wavelength anomalous dispersion (MAD) phasing, described below.

Synchrotron data for my graduate studies were collected at the Beamline 4.2.2, Advanced Light Source (ALS) at Berkeley, CA, Beamline X25, X29, National
Synchrotron Light Source (NSLS) at Brookhaven, NY, and Beamline 19ID, 19BM, Advanced Photon Source (APS) at Argonne, IL.

2.4.3 Recording the reflections and data processing

When a mounted crystal is exposed to X-ray beam, it scatters the X-rays into a pattern of spots that can be recorded with a charge-coupled device (CCD) image sensor behind the crystal. The relative intensities of these reflections provide direct information regarding the arrangement of macromolecules in 3-D space within the crystal in atomic detail. The pattern of these reflections within the first one or few frames can be used to ascertain the crystal symmetry, the unit cell parameters, the crystal orientation and the resolution limit known as INDEX step, within which a Miller index can be assigned to each single reflection. Having assigned symmetry, a data collection strategy is used which will maximize both the resolution and completeness of the dataset can be derived. Usually the crystal is rotated through a small angle 0.5-2 degree along the omega angle and the diffraction pattern is recorded. In the following INTEGRATION step, the intensity for each reflection on every frame is calculated, sorted according to their Miller index and stored in a single file. In the last step SCALE, two procedures are usually applied. First the program tries to identify which reflection appears in two or more images and to scale the relative images so that they have a consistent intensity scale. Second, many symmetry-equivalent reflections are recorded multiple times and merged together giving rise to the symmetry related R-factor, $R_{\text{sym}}$ or $R_{\text{merge}}$. This value is usually used to assess the quality of the data based upon how similar the measured intensities of symmetry equivalent reflections are.

Data collection and processing software suite I have used in my graduate studies
include HKL2000\textsuperscript{78} and d*TREK\textsuperscript{79}.

2.5 Phasing

In order to obtain an interpretable electron density map of crystallized macromolecule, both amplitude and phase information of each individual reflection must be known. During data collection, however, only the intensity of the diffraction spots is recorded while the other half of information i.e. phases are lost; this is known as the phase problem. Multiple techniques have been developed to circumvent this challenging phase problem in macromolecular crystallography.

2.5.1 Direct Methods

Direct methods are limited to structures with, at most, 1000 non-hydrogen atoms in the unit cell and have been used successfully to solve the phase problems for small proteins by exploring statistical relationships between sets of structure factors. Due to its requirement for near atomic resolution, direct methods are not generally applicable to the vast majority of macromolecular crystal structures.

2.5.2 Molecular replacement

MR remains the most widely used technique for macromolecular crystal structure solution, accounting to 67% of all X-ray structures released in 2006\textsuperscript{80}. MR relies upon the information of a previously determined protein structure that is homologous to our target protein from which the diffraction data is obtained. The rule of thumb is "Conventional wisdom holds that an RMSD of about 1.5Å between the search model and the actual structure constitutes the limit at which MR can still be used"\textsuperscript{39} Gerard Kleywegt. This
limit roughly corresponds to 29% identity in protein primary sequence. In this method, self rotation, cross rotation and translation function searches are carried out to find an orientation and position of the molecules within the unit cell. The phases obtained this way can be used to generate electron density maps that contain certain features of the protein of interest.

MR programs I have used in my graduate studies include Phaser, MOLREP and EPMR.

2.5.3 Anomalous scattering (MAD or SAD)

Over the past few years, MAD and SAD method have become the methods of choice in solving a novel structure, each representing about 10% of all X-ray structures released in 2006 respectively. Bragg reflections related by inversion through the origin known as Friedel pairs usually have equal amplitude and opposite phase. Friedel's Law is broken whenever anomalous scatters are present in the crystal such as selenium incorporated during protein expression, heavy atom derivatives, metals inherent to protein, halides, and noble gases. When the wavelength of the incident x-ray approaches the absorption maximum of an anomalous scatters in the crystal, the scattering enlarges dramatically (Figure 2.3). By collecting data at selected wavelengths, the real (f') or imaginary (f'') anomalous contributions to the atomic scattering factor differences are maximized, which is used for calculating the substructure of the anomalously diffracting atoms and allowing for the phase angles of the whole protein to be determined. Traditionally, MAD data are collected at three different wavelengths, corresponding to the inflection point of the absorption edge (λ1), the absorption maximum above the edge (λ2), and a remote point at higher energy (λ3). More recently anomalous data collected at a single wavelength SAD
are becoming popular as the phase ambiguity intrinsic in SAD data can be partially resolved by the phase relationship determined by direct methods\textsuperscript{85}.

MAD or SAD programs I have used in my graduate studies include Solve/Resolve\textsuperscript{86}, SHELX-97\textsuperscript{87}, and PHENIX\textsuperscript{88}.

2.5.4 Heavy atom methods (MIR or SIR)

This used to be the most common method by which novel protein crystal structures were solved, although nowadays it has largely been replaced by MAD/SAD phasing. Typically heavy atoms are introduced into the crystal by soaking in a heavy atom-containing solution or by co-crystallization. Historically the most successful compounds are named “magic seven”: K$_2$PtCl$_4$ (platinum potassium chloride), KAu(CN)$_2$ (aurous potassium cyanide), K$_2$HgI$_4$ (mercuric potassium iodide), UO$_2$(C$_2$H$_3$O$_2$)$_2$ [uranium (VI) oxyacetate], HgCl$_2$ (mercuric chloride), K$_3$UO$_2$F$_5$ (potassium uranyl fluoride) and para-chloromercurybenzoic sulfate (PCMBS)\textsuperscript{89,90}. These electron-dense metal atoms can be located by comparing their diffraction intensity difference via direct methods or Patterson-space methods allowing the calculation of the initial phases for the whole molecule.

2.6 Model building and refinement

Once the initial phase angles are obtained, an electron density map can be generated by combining the diffraction intensity and the estimated phases. The map can be significantly improved by applying density modification step such as solvent flattening, histogram matching, non-crystallographic subunit (NCS) averaging and pattern recognition of structural motifs. Map interpretation and protein model building are
performed by recognizing secondary structure elements, locating heavy atom sites, main
chain atom tracing and assigning amino acid sequence to the electron density map. Once
a relatively complete model is built, it is subject to refinement against the diffraction data,
while the new phase information is derived exclusively from this new model. The
parameters of this model such as atom positions and temperature factors are optimized to
obtain a best fit to the experimental data. The R factors, which describe the difference
between the modeled structure and the experimental data, are constantly calculated
through every step of refinement. A fraction of the data is kept aside from the refinement
and used as a cross-validation set to minimize the unwanted effects of overfitting. This
process allows us to calculate an $R_{\text{free}}$ value, which is an unbiased indication of the
quality of the structure. As a rule of thumb, $R/R_{\text{free}}$ should be approximately the
resolution in Ångstoms divided by 10. The model building, structure refinement and
electron density map calculation are performed in an iterative fashion until the
convergence of the final model and the diffraction data is reached.

Model building programs I have used in my graduate studies include coot$^{91}$ and O$^{92}$.

Refinement programs I have used in my graduate studies include Refmac$^{93}$, CNS$^{94}$,
SHELX-97$^{87}$, and PHENIX$^{88}$.

2.7 Validation

Crystal structures need to be validated before they can be deposited into the public
Protein Data Bank (PDB). First of all, the bond lengths, bond angels, close contacts and
other parameters need to be kept in a reasonable range. Second the distribution of
main-chain torsion angles as shown in a Ramachandran plot is a good indicator of the
quality of the structure.
Validation programs I have used in my graduate studies include PROCHECK$^{95}$, SFCHECK$^{96}$, and MOLPROBITY$^{97}$.  


Figure 2.1 Protein crystallization flowchart involves the following steps: subcloning, expression, purification, crystallization, diffraction data collection, recording the reflections and data processing, phasing, model building and refinement, and validation.
Figure 2.2 Crystallization experiment setup methods. (A) Sitting drop vapor diffusion. (B) Hanging drop vapor diffusion. (C) Microbatch under oil. The pictures are adapted from Hampton research website (http://www.hamptonresearch.com/).
Figure 2.3 Se absorption edge plot from 12600eV to 12700eV with the edge, peak and short/long remote labeled as $\lambda_1$, $\lambda_2$ and $\lambda_3$. The picture is a theoretical estimation of $f'$ and $f''$ for Se generated by X-ray Anomalous Scattering web tool (http://skuld.bmsc.washington.edu/scatter/).
CHAPTER 3

Crystal structure of a Heme–Nitric oxide and OXygen binding (H-NOX) domain from *Nostoc sp.*: the common structural module used for Nitric Oxide (NO) sensing in bacterial NO sensor and mammalian soluble guanylate cyclase

3.1 Introduction

Diatomic ligand discrimination by soluble guanylyl cyclase (sGC) is paramount to cardiovascular homeostasis and neuronal signaling. Nitric oxide (NO) stimulates sGC activity 200-fold compared with only four-fold by carbon monoxide (CO). The molecular details of ligand discrimination and differential response to NO and CO are not well understood. These ligands are sensed by the heme domain of sGC, which belongs to the heme nitric oxide oxygen (H-NOX) domain family, also evolutionarily conserved in prokaryotes. Here we report crystal structures of the free, NO-bound, and CO-bound H-NOX domains of a cyanobacterial homolog. These structures and complementary mutational analysis in sGC reveal a molecular ruler mechanism that allows sGC to favor NO over CO while excluding oxygen, concomitant to signaling that exploits differential heme pivoting and heme bending. The heme thereby serves as a flexing wedge, allowing the N-terminal subdomain of H-NOX to shift concurrent with the transition of the six- to five-coordinated NO-bound state upon sGC activation. This transition can be modulated by mutations at sGC residues 74 and 145 and corresponding residues in the cyanobacterial H-NOX homolog.

3.2 Materials and methods
3.2.1 Cloning and mutagenesis of Ns H-NOX

The genomic DNA of *Nostoc sp* PCC 7120 (*Anabaena sp* PCC 7120) was used for PCR of the 189-amino-acid H-NOX gene using the forward primer 5’-ggaattccatatgtatggtttagtgaacaaagcc-3’ and the backward primer 5’-ccggaattctcagtcgtcatataattcgagtc-3’. The amplified fragment was inserted into the pET22b (Novagen) expression vector. The W74F and M144I mutants were introduced using the QuickChange kit (Stratagene) and were confirmed by DNA sequencing.

3.2.2 Expression and purification of Ns H-NOX

The Ns H-NOX protein was expressed in *Escherichia coli* BL21 (DE3) Star cells (Invitrogen). After overnight 0.1mM IPTG induction at 18°C, cells were pelleted and lysed in buffer A containing 20mM Tris–HCl (pH 7.5), 100mM NaCl, and 2mM β-mercaptoethanol. The cleared cell lysate was loaded onto a Q-Sepharose FF 5ml column and the H-NOX protein was eluted using an NaCl gradient. The 420 nm peak fractions containing the H-NOX protein were concentrated and loaded onto a Sephadex 200 column with buffer A as the running buffer. The fractions with OD420>OD280 were loaded onto a high-resolution MonoQ anion exchange column and an NaCl gradient was used for elution. Finally, the H-NOX protein was cleaned up using a high-resolution prepacked Sephadex75 column.

3.2.3 UV/Vis absorption spectroscopic analysis of Ns H-NOX

A SHIMADZU UV-1700 UV–visible spectrophotometer was used for all spectral analysis performed at room temperature. The concentration of the wt and the mutant H-NOX proteins was 1 mM in a buffer containing 10mM Tris–HCl (pH 7.5), 100mM
NaCl, and 1mM β-mercaptoethanol. The NO complex was prepared by adding the NO
donor S-nitroso-N-acetylpenicillamine (SNAP) to a final concentration of 200 mM. The
CO complex was obtained by using a saturated CO solution that was prepared by
bubbling CO gas (Praxair) through the above Tris buffer.

3.2.4 Crystallization and data collection of Ns H-NOX

Ns H-NOX crystals were grown at 20ºC by the sitting-drop vapor diffusion method.
The protein was concentrated to 15 mg/ml in 5mM Tris (pH 7.5), 100mM NaCl, and
1mM β-mercaptoethanol, and was mixed with an equal volume of the reservoir solution
containing 1.4M tri-sodium citrate dihydrate (pH 5.6–6.6). Native Ns H-NOX crystals
were prepared for data collection by fast transfer into cryo-protectant solution containing
1.36M tri-sodium citrate dihydrate with 15% ethylene glycol. Selenomethionine-substituted protein crystals (produced similar to Ref98) were used for phasing. The F74W and M145I Ns H-NOX mutants unfortunately resisted crystallization.

3.2.5 Ns H-NOX structure determination and refinement

The native 2.1Å resolution data set was collected at the Brookhaven NSLS
synchrotron (beamline X29) and processed by HKL200078. For phasing, a SAD data set
was collected from a selenomethionine-substituted crystal at the Advanced Light Source
ALS (beamline 4.2.2.) and processed by D*trek79. SOLVE/RESOLVE86 was used to
calculate initial phases and density modification, resulting in locating 10 selenium atoms
yielding a figure of merit of 0.7 (Table 2.1). A partial model of two molecules in the
asymmetric unit generated by RESOLVE was used as a starting point for ARP/wARP99
for additional automatic model building, using the 2.1Å native data set. Iterative cycles of
model building in Coot\textsuperscript{91} and refinement using REFMAC\textsuperscript{93} yielded a model comprising H-NOX residues 1–182 for both molecules A and B, two heme moieties, and 230 water molecules (if above 3\(\sigma\) observed in omit |Fo|-|Fc| map and kept if peak remained above 1\(\sigma\) in 2|Fo|-|Fc| map) with a final R-factor of 18.1\% and \(R_{\text{free}}\) of 22.2\% (see Table 2.1 for additional refinement statistics). One high-temperature factor water molecule was observed in the general vicinity of Fe in only molecule B, but was located at >4.5\AA\ from this Fe, weakly bound, and therefore was not included in the analysis.

The 2.6\AA\ NO:H-NOX complex data set was collected at the Advanced Photon Source APS (beamline 19ID) and processed by HKL2000. The NO complex crystals were prepared by soaking wt H-NOX crystals in the similar cryoprotectant solution with 20mM of the NO donor SNAP (Cayman) while varying the soaking time from 1, 2, 4, 6, 8, 10 min to several hours before crystal freezing by dunking it in liquid nitrogen for X-ray data collection. Tens of data sets were collected and analyzed. We found the 2 min soaking time point to be the most structurally informative regarding resolution of diffraction and NO electron density. The 2.1\AA\ native H-NOX structure was used as a starting point for crystallographic refinement with REFMAC. Strong NO density in the vicinity of Fe at the distal side of the heme was observed (7.7\(\sigma\) peak for molecule A and 4.5\(\sigma\) for molecule B) (Figure 3.1A). One NO molecule for each heme was positioned in the electron density and further refined using REFMAC. The final refined model contained residues 1–182 for both molecules A and B, two heme molecules, two NO molecules, and 64 waters. CO-bound crystals were obtained in a similar manner. The cryoprotection solution was bubbled with CO (Praxair) for an extended period of time to achieve a saturated CO soaking solution. The soaking time was varied and the most optimal time for H-NOX crystals was 45 min, and a 2.5\AA\ resolution data set was
collected at APS (beamline 19ID). Initial refinement revealed strong $|\text{Fo}|-|\text{Fc}|$ difference density, representing CO at the distal side of the heme (6.3$\sigma$ and 7.0$\sigma$ for molecules A and B, respectively). Both CO molecules were refined at 100% occupancy. The final refined model contained residues 1–182 of both molecules, two heme molecules, two CO molecules, and 122 water molecules. Heme distortion calculations were calculated using normal-coordinate structural decomposition\textsuperscript{100}. Owing to the 2.6/2.5 Å resolution of the NO/CO H-NOX complex structures, the angles are likely only accurate to within about 10° of the refined values, yet these refined angles are within the range of the Fe–N/C–O angles observed for other NO and CO heme-protein complexes\textsuperscript{101,102}. The program ESCET\textsuperscript{103} was used in the analysis of conformational changes among the different H-NOX structures.

### 3.2.6 Mutagenesis of rat sGC and transfection in COS-7 cells

Templates were cDNAs encoding the $\alpha1$ and $\beta1$ subunits of rat sGC cloned into the mammalian expression vector pCMV5\textsuperscript{104}. The F74W and I145M mutations were introduced by PCR (Quickchange, Stratagene) and checked by DNA sequencing. COS-7 cells were grown in DMEM supplemented with 10% heat-inactivated fetal bovine serum (FBS), penicillin, and streptomycin (100 U/ml). Cells were transfected for 48 h with Superfect reagent, according to the supplier’s protocol (Qiagen).

### 3.2.7 Cytosol preparation and Western blot analysis

COS-7 cells were washed twice with ice-cold PBS and then scraped off the plate in cold lysis buffer: PBS buffer contained protease inhibitors, 50 mM HEPES (pH 8.0), 1 mM EDTA, and 150 mM NaCl. Cells were broken by sonication (three pulses of 3 s).
The resulting lysate was centrifuged at 16 000 g for 10 min at 4°C to collect the cytosol. To determine the expression levels for wt sGC or sGC mutants in COS-7 cells, 15 μg of cytosol was resolved on 10% SDS–PAGE and analyzed by immunoblotting with anti-sGC (anti-α1 subunit and anti-β1 subunit; Cayman Chemicals).

3.2.8 sGC activity assay

GC activity was determined by the formation of [α-32P]cGMP from [α-32P]GTP, as previously described105. Reactions were performed for 5 min at 33°C in a final volume of 100 μl, in 50 mM HEPES (pH 8.0) reaction buffer containing 500 M GTP, 1 mM DTT, and 5 mM MgCl2. Typically, 40 μg of COS-7 cytosol transfected with either wt or mutants was used in each assay reaction. All assays were performed in duplicate and each experiment was repeated twice. Enzymatic activity was stimulated with the NO donor SNAP (Calbiochem) at 100 μM. sGC activity is expressed in pmol/min mg and as mean±s.e.

3.3 Results and discussion

3.3.1 Structure of the Ns H-NOX domain

The Ns H-NOX domain represents the second member in the H-NOX family that has been structurally characterized. The protein fold mirrors those found in the Tt H-NOX domain45,48: seven α helices, labeled αA-G, and one four-stranded antiparallel β sheet, labeled β1-4 (Figure 3.2A). The heme prosthetic group is covalently attached to the protein via the His105-Fe bond at the proximal side. Intensive hydrophobic interactions are observed between the propionate group of the heme and the conserved “YxSxR” motif49,106 in the protein (Figure 3.2C).
Despite the overall fold similarity, the \( Ns \) and \( Tt \) H-NOX display distinct ligand preferences\(^4\). \( Tt \) H-NOX has extremely high affinity for its oxygen ligand while the \( Ns \) H-NOX only recognize NO and CO with no apparent affinity for O\(_2\), which is reflected by the flat heme electron density at the distal side. The ligand preference observed in the crystal structures is consistent with the solution studies (Figure 3.2B). This difference can be readily explained by comparing the residues in the heme pocket: 17 of 27 residues are identical between \( Ns \) H-NOX and sGC, compared with only 8 of 27 between \( Tt \) H-NOX and sGC. This similarity applies to the whole H-NOX gene (Figure 3.1B): \( Ns \) H-NOX shares a much higher sequence identity with sGC (33%) than \( Tt \) H-NOX with sGC (18%).

Superposition of the C-terminal proximal subdomain of \( Tt \) H-NOX and \( Ns \) H-NOX reveals an 18° difference in the orientation of the N-terminal distal subdomain (Figure 3.2D). This N-terminal shift in the \( Tt \) H-NOX causes a strong heme distortion characterized by the -1Å heme saddling and ruffling\(^4\). In contrast, the heme in \( Ns \) H-NOX is less bent, with only moderate ruffling (~0.35 Å) and dome (~0.5 Å) distortions. The other independently refined \( Tt \) H-NOX molecule (monoclinic molecule B) adopts an intermediate N-terminal domain shift of 11° regarding \( Ns \) H-NOX. We hypothesize that the H-NOX domain is capable of doing a series of subdomain orientation shifts accompanied by heme distortion.

3.3.2 The \( Ns \) H-NOX domain in the NO and CO bound form

We obtained \( Ns \) H-NOX NO/CO complex structures at 2.6 and 2.5 Å resolution by soaking \( Ns \) H-NOX crystals in solutions containing NO or CO. Strong omit electron density peaks at the distal site of the heme strongly suggest ligand binding (Figure 3.3A,
B). The diatomic ligands are located in a hydrophobic pocket formed by V5, F70, W74, M144, and L148. In NO-bound \( Ns \) H-NOX, the nitrogen atom of NO forms a 1.8Å bond with Fe and the Fe–N–O angle is 143° and 157° for molecules A and B, respectively. These angles are more perpendicular (163° and 168°) for the CO-bound complex. The observed bent/tilt angle differences are in general agreement with the published NO and CO heme geometries\(^{101,102,107}\).

### 3.3.3 Conformation changes upon ligands binding

Since the first H-NOX structure from \( Tt \) was solved only in the oxygen bound form, the conformational changes in the protein and heme group upon binding of the ligands could not be addressed. In our NO and CO bound \( Ns \) H-NOX structures, the diatomic gaseous molecules bind to the distal face of the heme resulting in close van der Waals interactions between the ligands and W74. In order to provide the steric space for NO and CO, the heme Fe-center pivots away from W74 while keep the propionate group attached to pyrrole ring D (P\( _D \)) in place serve as the pivot point (Figure 3.3C, 3.4A). A larger Fe shift (0.8–0.9Å\(^\circ\)) was observed upon CO binding compared with NO binding (0.2–0.3Å\(^\circ\)) due to the larger atomic radius of its carbon atom and its preference to bind perpendicular to the plane of the heme\(^{107}\). We propose this aromatic residue serves as a molecular ruler to discriminate NO over CO\(^{108}\). Besides several protein regions surrounding the heme moiety in \( Ns \) H-NOX protein also displays several small but significant changes as a result of heme movement: H105, \( \alpha A \) region near Y2 and \( \alpha B-\alpha C \) loop near residue E41. A loop region near F112 and C-terminal part of helix \( \alpha F \), which interact with heme becomes substantially flexible due to the heme moving away motion (Figure 3.3D). These structural changes are observed in both independently refined \( Ns \) H-NOX molecules in...
the asymmetric unit.

3.3.4 Heme pivoting and heme-bending activation mechanism

By comparing the inactive apo \( Ns \) H-NOX structure, the NO/CO-bound low activity \( Ns \) H-NOX structures and the final active form of \( Tt \) H-NOX structures, a full scheme of heme pivoting motion was revealed. The iron center together with the heme group moves progressively in the counter clockwise direction when viewed from the distal side (Figure 3.4A). The extent to which the heme pivots correlates well with their presumably cyclase activity.

However at room temperature, the \( Ns \) H-NOX remains predominately in the six coordinated state although a noticeable proportion of the protein is in the active five coordinated state. The limited capability for \( Ns \) H-NOX to fully reach five coordinated state can be readily explained when the following heme bending motions are taken into consideration (Figure 3.4B). The heme bending at the side of the pyrrole A ring with attached propionate \( P_A \) provides the space to accommodate the large shift of the N-terminal \( \alpha A - \alpha C \) helices. The helical shift of \( \alpha A \) would require, in addition to the heme bending, space in the direction of residue W74 as the \( \alpha A \) residue A8 is postulated to shift about 3–4Å towards W74. Owing to the size of W74 (which is the smaller F74 in rat sGC\( \beta 1 \)) and the close 3.7Å van der Waals interaction between the side chains of A8 and W74, the \( \alpha A \) helix has limited space for this large shift in \( Ns \) H-NOX. Besides the pyrrole D ring attached to the propionate \( P_D \) group tilts in the opposite (distal) direction towards M144. The tilting of the pyrrole D ring with attached propionate \( P_D \) of the heme involves a 0.7Å shift of this propionate group when comparing the unliganded \( Ns \) H-NOX to the activated oxygen bound \( Tt \) H-NOX structure. However, the CE atom of M144 forms a
snug 3.4Å van der Waals interaction with the C<sub>AD</sub> atom of this propionate P<sub>D</sub> group in the unliganded <i>Ns</i> H-NOX structure, thereby causing steric hindrance for such a heme bending at this side as well.

### 3.3.5 Mutagenesis studies to test H-NOX activation mechanism

Based on our structural analysis, we have pinpointed the switch controlling the six or five heme coordination states to residues W74 and M144. To test this hypothesis, we mutated them to F and I, respectively, the smaller amino acids counterpart found in rat sGCβ1 respectively. The W74F and M144I <i>Ns</i> H-NOX displayed considerable increases in the soret peak near 400 nm (Figure 3.2B), representing increased population of the five-coordinated NO-bound state. The removal of steric clashes indeed facilitates the conversion from six to five coordination state. These results are in agreement with our activation model, indicating that reaching the activated five-coordinated NO-bound state is directly correlated with heme bending and the N-terminal subdomain shift. On the other hand, when we introduced the larger <i>Ns</i> H-NOX residues into rat sGCβ1, NO-simulated guanylyl cyclase activity are significantly decreased from 21 fold in wild-type sGC to only 2 or 5 fold for F74W and I145M sGC mutants, respectively (Table 3.2 and Figure 3.4C). These reverse mutants have thus limited their capability to reach the optimal 5-coordinated state due to the larger size of W and M side chains.

### 3.4 Conclusions

In conclusion, we provide evidence for a molecular ruler mechanism in sGC favoring NO over CO binding, using steric hindrance involving an aromatic residue at position 74. Both ligands induce differential heme pivoting and heme bending, which appear to
correlate with differences in sGC activation levels. CO induces a larger initial heme pivot shift compared with NO, whereas NO only proceeds to the five-coordinated fully activated state concomitant with a strongly bent heme and large N-terminal subdomain shift. These H-NOX heme pivoting/bending signaling insights for sGC activation offer new opportunities for sGC to be therapeutically exploited.
Figure 3.1 Structure and domain organization of H-NOX and H-NOXA containing proteins. (A) Schematic diagram of the H-NOX and H-NOXA domains present in proteins from animals and Nostoc cyanobacteria. The sGC subunits contain additional coiled-coil (CC) and guanylyl cyclase (GC) domains. The purple interrupted line indicates the domain that was targeted for crystallographic studies. The PAS-like domain is termed H-NOXA, since it is often associated with H-NOX and was previously named H-NOBA4 prior to the discovery that these heme domains could also bind oxygen48. The proteins transcribed adjacent to the H-NOX gene in cyanobacteria are the signal transduction histidine kinase (STHK) and the two-component hybrid sensor and regulator (2-CHSR). (B) Structure-based sequence alignment of H-NOX domains. Residues in the heme cavity are labeled in orange or red, the latter if the residue is fully conserved. Additional conserved residues in the hydrophobic core are shaded yellow. Residues key for the transition from 6- to 5- coordinated NO bound are labeled with a ‘#’ (W74 and M144). The three N-terminal helical rods are red. Residues identical to rat sGCβ1 are in bold. Conserved glycine residues are shaded green. The GeneBank identifiers for the selected sequences are: N.sp., GI:17229770; N.punctiforme, GI:23129606; L.pneumophila1, GI:52841290; L.pneumophila2 GI:52629778; V.cholerae GI:15601476; T.tengcongensis GI:20807169; C.acetobutylicum GI:15896488; Rattus_sGCβ1 GI:28564567; Rattus_sGCβ2 GI:21956635; C.elegans_GCY35 GI:52782806; D.melanogaster_sGCβ1, GI:7302016.
**Animals (HNOX fused to HNOXA/CC/cyclase)**

- sGC α1
- sGC β1

**Cyanobacteria (stand-alone HNOX and HNOXA/CC)**

- Nostoc punctiforme PCC 71110
  - ORF: NpunG20003819
  - ORF: NpunG02000820
- Nostoc sp. PCC 7120
  - ORF: air2278
  - ORF: air2279

**Bacterial ORF**

- Nsp
- Npun
- Lpsma1
- Vchol
- Tten
- Cst
- rs0071
- Cst
- Dsel
- Cons.

- Nsp
- Npun
- Lpsma1
- Vchol
- Tten
- Cst
- rs0071
- Cst
- Dsel
- Cons.
Figure 3.2 Crystal structure of unliganded *Ns* H-NOX. (A) Schematic diagram of the structure of *Ns* H-NOX. The three N-terminal helices αA-αC (red), the heme (blue), and H105 (green) are highlighted. The heme propionate groups attached to pyrrole rings A and D are labeled P_A and P_D, respectively. Figure generated using MOLSCRIPT\(^\text{109}\) and Raster3D\(^\text{110}\). (B) Electronic absorption spectra of wild-type and mutants of *Ns* H-NOX. The spectra before (solid line) and after adding 200μM SNAP (dotted line) as the NO donor are shown. The 5-coordinate free heme signature peak is near 430nm, the 6-coordinated NO-bound heme signature peak near 416nm, and the 5-coordinated NO-bound heme is near 400nm. For the wt *Ns* H-NOX, also a spectra in the presence of CO was measured (interrupted line) yielding a peak near 423nm indicating the presence of a 6-coordinated CO-bound state. (C) Stereo figure showing electron density for the heme group in the free *Ns* H-NOX structure. The main chain is depicted in coil representation with side chains in close proximity to the heme in stick representation; residues from the N-terminal residues 1-60 are shown in red. The omit |Fo|-|Fc| density is drawn in blue (2.1 Å resolution, contoured at 3σ). Figure generated using BOBSCRIPT\(^\text{111}\). (D) Superposition of the *Ns* and *Tt* H-NOX structures revealing an N-terminal domain shift for helices αA-αC. Depicted are the *Ns* H-NOX structure (blue), partially shifted *Tt* H-NOX (green; monoclinic molecule B in PDBid 1U55), and maximally shifted *Tt* H-NOX (red, monoclinic molecule A in PDBid 1U55)\(^\text{48}\). The arrow shown indicates the direction of the shift.
Figure 3.3 Structural consequences of NO and CO binding to Ns H-NOX. (A) NO binding to the distal face of the heme in the NO bound Ns H-NOX structure. Omit |Fo|-|Fc| electron density (contoured blue at 5σ) and |2|Fo|-|Fc| electron density (contoured grey at 1.2σ) for the heme and its two ligands, NO and H105, are shown. (B) As above, but for CO binding to H-NOX, omit density contoured at 4σ, |2|Fo|-|Fc| contoured at 1σ level. (C) Molecular shifts within the heme region of Ns H-NOX upon NO and CO binding. Superpositioning of free (blue), NO-bound (yellow), CO-bound Ns H-NOX (magenta) structures reveals a shift in the heme, αA including residues 1-3, and loop αB-αC containing residues E41. Hydrogen bonds between the heme propionate and main chain of Y2 and between the main chains of G3 and E41 are shown (dashed lines). (D) CO binding to Ns H-NOX leads to increased flexibility in the loop following helix αF and C-terminal part of this helix. Depicted are a close-up view of the heme region of the free Ns-HNOX structure (left) and CO-bound Ns H-NOX structure (right). The temperature factor of the mainchain is color ramped from blue (~35Å²), green (~45Å²), to orange (~80Å²). The heme and the H105 and F112 side chains are color coded by atom type. Figure generated using Pymol (http://pymol.sourceforge.net/).
Figure 3.4 Evidence for heme pivoting and heme bending in activation mechanism of H-NOX. (A) Top (distal) view of the heme groups of the superimposed H-NOX structures. Ns H-NOX residues M144 and W74 are shown as well as the residues interacting with the propionate groups of the heme (M1, Y2, Y134, S136, and R138) and residues belonging to the αA-αC (red). The five heme moieties are color coded as in Figure 2D and 3C. The pivot directions of Fe and heme are indicated by arrows. (B) Side-view stereo figure of the heme groups of the superimposed H-NOX structures. The five heme moieties are color coded as in Figure 4A. The van der Waals interactions between the A8 side chain and CZ3 atom of W74 as well as between CE atom of M144 and the propionate PD group in the unliganded Ns H-NOX are depicted as transparent dashed lines. The position of the αA helix in Tt H-NOX molecule A and its M1 side chain are shown in transparent blue to indicate their potential clash with the W74 side chain and the propionate PD group, respectively. (C) Western blot analysis of wild-type and mutant sGC proteins. Cytosols of the wild-type, H-NOX β1 heme domain mutants (F74W and I145M) for both the α1 and β1 subunits were probed with anti-sGC α1 and anti-sGC β1 antibodies.
Figure 3.5 Schematic diagram of NO and CO dependent activation of sGC. NO and CO bind to the distal face of the heme and cause the heme to pivot, with varying degrees, away from F74 (W74 in Ns H-NOX). The larger pivot for CO is likely a result of the need to accommodate its carbon atom, having a larger radius compared to nitrogen in NO, adjacent to F74 and CO’s preference to bind perpendicular to the heme. The second step in the NO activation mechanism involves heme bending and N-subdomain movement (only helix αA is shown for illustrative purposes). Whether a similar heme bending event occurs in CO/YC-1 activated sGC is unknown and speculative. Also unknown is whether a second NO molecule binds to the proximal side of the heme thereby displacing the histidine ligand as well as the distal NO as was postulated previously.\textsuperscript{64,112} NO could therefore be bound to the distal or proximal face of the heme in the NO-bound 5-coordinated activated state (both possibilities are depicted).
Table 3.1 Data collection, phasing, and refinement statistics for Ns H-NOX

<table>
<thead>
<tr>
<th>Datacollection</th>
<th>Native</th>
<th>Se-peak</th>
<th>NO complex</th>
<th>CO complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.1000</td>
<td>0.9789</td>
<td>0.9793</td>
<td>1.0719</td>
</tr>
<tr>
<td>Space group</td>
<td>P2_13</td>
<td>P2_13</td>
<td>P2_13</td>
<td>P2_13</td>
</tr>
<tr>
<td>Cell dimension (Å)</td>
<td>a=b=c=123.42</td>
<td>a=b=c=123.76</td>
<td>a=b=c=123.56</td>
<td>a=b=c=123.63</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.1</td>
<td>2.9</td>
<td>2.6</td>
<td>2.5</td>
</tr>
<tr>
<td>Total observations</td>
<td>153963</td>
<td>148837</td>
<td>52349</td>
<td>99121</td>
</tr>
<tr>
<td>Unique observations</td>
<td>36257</td>
<td>14307</td>
<td>18960</td>
<td>21865</td>
</tr>
<tr>
<td>I/SigI</td>
<td>18.4 (1.7)</td>
<td>14.8 (4.5)</td>
<td>12.0 (2.0)</td>
<td>29.3 (3.7)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.2 (3.9)</td>
<td>10.40 (10.55)</td>
<td>2.8 (2.8)</td>
<td>4.5 (4.5)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.9 (98.1)</td>
<td>100.0 (100.0)</td>
<td>96.6 (99.0)</td>
<td>99.1 (99.6)</td>
</tr>
<tr>
<td>Rsym (%)</td>
<td>6.7 (43.9)</td>
<td>9.1 (42.3)</td>
<td>9.3 (60.6)</td>
<td>5.8 (47.5)</td>
</tr>
<tr>
<td>SAD phasing</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure of merit</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>25-2.1</td>
<td>40-2.6</td>
<td>39-2.5</td>
<td>39-2.5</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>2866</td>
<td>2866</td>
<td>2866</td>
<td>2866</td>
</tr>
<tr>
<td>No. of hetero atoms</td>
<td>86</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>No. of waters</td>
<td>229</td>
<td>64</td>
<td>122</td>
<td>122</td>
</tr>
<tr>
<td>R_factor</td>
<td>18.5</td>
<td>19.2</td>
<td>18.7</td>
<td>18.7</td>
</tr>
<tr>
<td>R_free</td>
<td>22.7</td>
<td>23.6</td>
<td>22.8</td>
<td>22.8</td>
</tr>
<tr>
<td>R.m.s.d. bond lengths (Å)</td>
<td>0.009</td>
<td>0.011</td>
<td>0.009</td>
<td>0.009</td>
</tr>
<tr>
<td>R.m.s.d. bond angles (°)</td>
<td>1.22</td>
<td>1.33</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>Average B-factors (Å²):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-Protein</td>
<td>45.1</td>
<td>52.8</td>
<td>51.0</td>
<td>51.0</td>
</tr>
<tr>
<td>-Heme moieties</td>
<td>32.0</td>
<td>38.0</td>
<td>38.6</td>
<td>38.6</td>
</tr>
<tr>
<td>-NO or CO ligands</td>
<td>-</td>
<td>49.1</td>
<td>45.8</td>
<td>45.8</td>
</tr>
<tr>
<td>-Solvent</td>
<td>53.9</td>
<td>48.4</td>
<td>53.4</td>
<td>53.4</td>
</tr>
<tr>
<td>Ramachandran plot statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residues in:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-most favored regions</td>
<td>93.7</td>
<td>93.0</td>
<td>94.0</td>
<td>94.0</td>
</tr>
<tr>
<td>-additional allowed reg.</td>
<td>6.3</td>
<td>7.0</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>-in generously allowed reg.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-disallowed regions</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3.2 Guanylyl cyclase activity measurements of wild-type and mutants of rat sGC under basal and NO-stimulated conditions

<table>
<thead>
<tr>
<th>H-NOX mutants</th>
<th>Basal (pmol min⁻¹ mg⁻¹)</th>
<th>+ NO donor SNAP (pmol min⁻¹ mg⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt α₁/β₁</td>
<td>245 ± 15</td>
<td>5240 ± 35</td>
</tr>
<tr>
<td>α₁/β₁ F74W</td>
<td>115 ± 7</td>
<td>263 ± 22</td>
</tr>
<tr>
<td>α₁/β₁ I145M</td>
<td>155 ± 16</td>
<td>770 ± 40</td>
</tr>
</tbody>
</table>
CHAPTER 4

Crystal structure of the dimerized PAS-like domain from a *Nostoc punctiforme* signal transduction histidine kinase (STHK): Implications for soluble guanylate cyclase dimerization and organization

4.1 Introduction

Signal transduction histidine kinases (STHK) are key for sensing environmental stresses, crucial for cell survival, and attain their sensing ability using small molecule binding domains. An N-terminal domain in an STHK from *Nostoc punctiforme* is of unknown function yet is homologous to the central region in soluble guanylyl cyclase (sGC), the main receptor for nitric oxide (NO). This domain is termed H-NOXA since it is often associated with the heme-nitric-oxide/oxygen binding (H-NOX) domain. A structure-function approach was taken to investigate the role of H-NOXA in STHK and sGC. We report the 2.1 Å resolution crystal structure of the dimerized H-NOXA domain of STHK, which reveals a Per-Arnt-Sim (PAS)-like fold often used as a sensory module. The H-NOXA monomers dimerize in a parallel arrangement juxtaposing their N-terminal helices and preceding residues. Deletion of 7 N-terminal residues affects dimer organization. Sequence analysis suggest that the H-NOXA domains of heterodimeric sGC could adopt a similar dimer organization and putative H-NOXA interface mutations in either sGCα1 or sGCβ1 yielded varying effects ranging from loss of overall heterodimer formation and concomitant cyclase activity, to heterodimeric sGC with decreased NO-stimulated activity suggesting a role for H-NOXA in the formation of a functional sGC heterodimer. Furthermore, the PAS-like fold and dimer organization were
unexpectedly found to be similar to those in the heme containing sensors Ec DOS and Rm FixL raising the possibility that the H-NOXA dimer might have allostERIC capabilities in sGC and STHK. Our structural and mutational results provide new insights into sGC and STHK dimerization.

4.2 Materials and methods

4.2.1 Cloning and mutagenesis of the \textit{Np} STHK H-NOXA domain

The genomic DNA of \textit{Nostoc punctiforme} PCC 73102 was used as the template to PCR the 1-121 H-NOXA fragment of STHK COG0642 using the forward primer 5'-ggaattccatatggctcctcctcaccttacg-3' and backward primer 5'-ccggaattctcatttgagtttgatacccaaaggagc-3'. To obtain the 8-121 (Δ7) H-NOXA construct, the 5'-ggaattccatgccttcacctaccgc-3' forward primer was used. The PCR amplified fragments were inserted into a pET28a (Novagen) expression vector.

4.2.2 Expression and purification of \textit{Np} STHK H-NOXA domain

Both the 1-121 and truncated Δ7 H-NOXA domains were expressed as N-terminal his-tagged protein in \textit{E. coli} BL21 (DE3) Star cells (Invitrogen) using IPTG for induction. The cells were pelleted and sonicated for 5 minutes on ice in the following buffer: 20mM Tris pH7.5, 100mM NaCl, 5mM β-mercaptoethanol. The cell lysate was centrifuged at 16,000g for 10 min at 4°C and the supernatant was incubated with Ni-NTA (Qiagen) beads. The beads were washed with the washing buffer: 20mM Tris pH7.5, 100mM NaCl, 5mM β-mercaptoethanol, 20mM imidazole. The protein was released from the beads by thrombin (Enzyme research labs) digestion. Further purification was performed by gel filtration in a Sephadex75 (Pharmacia) column equilibrated with 5mM Tris pH7.5,
100mM NaCl, and 1mM β-mercaptoethanol.

4.2.3 Crystallization and data collection of Np STHK H-NOXA domain

Crystals of the Np STHK 8-121 protein construct were grown at 4°C by sitting-drop vapor diffusion. Protein was concentrated to 20mg/ml in 5mM Tris pH7.5, 100mM NaCl, 1mM β-mercaptoethanol and was mixed with an equal volume of the reservoir solution: 1.7-1.9M ammonium sulfate, 100mM Tris-HCl (pH 7.7-9.0). Crystals were prepared for data collection by fast transfer into cryoprotectant solution containing 2.0M ammonium sulfate 100mM Tris-HCl with 15% glycerol. The 1-121 Np STHK protein was concentrated to 15mg/ml after the gel-filtration step and mixed with equal volume of 0.1M HEPES pH7.5 and 1.5M lithium sulfate monohydrate. Crystals appeared after three days at 20°C. For data collection, the crystals were soaked in cryo-protectant containing 25% glycerol in addition to the crystallization solution and dunked into liquid nitrogen prior to data collection. Crystals of selenomethionine substituted 8-121 protein were used for SAD phasing.

4.2.4 Np STHK H-NOXA domain structure determination and refinement

Due to great difficulty in obtaining diffraction quality crystals for the 1-121 Np STHK construct, the initial structure determination was carried out using crystals of the shorter 8-121 construct. A native 2.0 Å resolution dataset for the 8-121 Np STHK protein was collected at ALS (Beamline 4.2.2.) and processed with D*trek79. The crystals belong to space group P6₁2₂, with cell dimensions a=b=72.3 Å, c=169.1 Å and two molecules in the asymmetric unit. To obtain crystallographic phase information, a SeMet crystal of the 8-121 Np STHK construct was used for SAD phasing. A 2.6 Å SAD data set was
collected at the selenium peak wavelength at ALS. SOLVE/RESOLVE[86] was used for phasing and automatic model building. Manual model building was carried out in Coot[91] and REFMAC[93] was used for refinement. The final model includes \( \text{Np} \) STHK residues 8-113 of molecule A and residues 10-119 of molecule B and 95 waters and 2 sulfate ions. The final model yielded an \( R/R_{\text{free}} \) of 23.8/29.2%.

For 1-121 H-NOXA, a native 2.1 Å resolution data set was collected at the Advanced Photon Source (Beamline 19ID) belonging to space group C2 with cell dimensions \( a=95.2 \, \text{Å}, b=44.4 \, \text{Å}, c=59.7 \, \text{Å} \) and 2 molecules in the asymmetric unit. The structure of this larger construct was determined via Molecular Replacement using PHASER[82] using one of the monomers from the P6\(_1\)22 space group. Coot and REFMAC were used for model building and crystallographic refinement yielding a final model includes residues 1-107 of molecule A, 1-105 of molecule B, and 55 waters (\( R/R_{\text{free}} \) are 20.2/26.6%). The stereochemistry was checked using PROCHECK[95].


### 4.2.5 Subcloning, mutagenesis, expression, and purification of the H-NOXA domain of sGC\( \beta \)1

Residues 202-344 of the H-NOXA domain of rat sGC\( \beta \)1 was subcloned into pET22b using the forward primer: gga aat tcc ata tgg gta ccc agg act ccc gta tc, and backward primer: gcg aat tct cag tgg tgg tgg tgg tgt cca ggg atg tca etc agg tac ag. The C-termially his-tagged sGC\( \beta \)1 H-NOXA protein was expressed and purified similar to the homologous \( \text{Np} \) STHK counterpart except that the protein was eluted from the Ni-NTA beads by an imidazole gradient instead of thrombin digestion. The Ala-scanning
mutants of sGCβ1 H-NOXA domain were introduced into pET22b plasmid using the QuikChange site-directed mutagenesis kit (Quickchange, Stratagene) sequence confirmed.

4.2.6 Mutagenesis of Rat sGC and Transfection in COS-7 Cells

Templates were cDNAs encoding the α1 and β1 subunits of rat sGC cloned into the mammalian expression vector pCMV5104. Mutations described in the text were introduced by PCR (QuikChange, Stratagene) and sequenced. COS-7 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal bovine serum, penicillin, and streptomycin (100 units/ml). Cells were transfected for 48 h with Superfect reagent using the protocol of the supplier (Qiagen).

4.2.7 Cytosol Preparation and Western Blot Analysis

COS-7 cells were washed twice with ice-cold phosphate-buffered saline and then scraped off the plate in cold lysis buffer: phosphate-buffered saline containing protease inhibitors, 50 mM HEPES, pH 8.0, 1 mM EDTA, and 150 mM NaCl. Cells were broken by sonication (3 pulses of 3 s) and centrifuged at 16,000 x g for 10 min at 4 °C to collect the soluble fraction (referred to as cytosol in the text). To determine the efficiency of transfection for wt sGC or mutants in COS-7 cells, 15 µg of cytosol was resolved on 10% SDS-PAGE and analyzed by immunoblotting with anti-sGC (anti-α1 subunit and anti-β1 subunit; Cayman Chemicals).

4.2.8 sGC Activity Assay

GC activity was determined by formation of [α-32P]cGMP from [α-32P]GTP, as
previously described\textsuperscript{105}. Reactions were performed for 5 min at 33°C in a final volume of 100 µl, in a 50 mM HEPES pH 8.0 reaction buffer containing 500 µM GTP, 1 mM dithiothreitol, and 5 mM MgCl\textsubscript{2}. Typically, 40 µg of COS-7 cytosol transfected with either wt or mutants were used in each assay reaction. All assays were done in duplicate and each experiment repeated twice. Enzymatic activity was stimulated with the NO-donor SNAP (Calbiochem) at 100 µM. sGC activity is expressed in pmol min\textsuperscript{-1} mg\textsuperscript{-1} and mean ± S.E.

4.3 Results and discussion

4.3.1 \textit{Np} H-NOXA structure reveals a PAS-like fold

The structure of the Δ7 N-terminally truncated H-NOXA was solved using SAD method, the monomer of which was used as a MR model to solve the full length \textit{Np} H-NOXA domain structure (Table 4.1 and Figure 4.1C).

The \textit{Np} H-NOXA monomer structure is comprised of a 7-stranded anti-parallel β-barrel flanked by a few α-helices with the N- and C-termini located on opposite ends of the monomer (Figure 4.1C). DALI\textsuperscript{113} search have revealed that all \textit{Np} H-NOXA structural neighbours adopt PAS fold including: Bj FixL (PDBid 1DRM-A\textsuperscript{114}), \textit{Ec} DOS (PDBid 1V9Y\textsuperscript{115}), the dPER fragment (PDBid 1WA9-A\textsuperscript{116}), PAS domain of HIF-2 (PDBid 1P97\textsuperscript{117}), and the photoactive yellow protein (PDBid 3PYP\textsuperscript{118}) with DALI Z-scores of 8.3 to 7.6 and r.m.s.d. values range from 2.5-3.0 Å for ~90 superimposed Cα atoms. Superposition of the monomers of \textit{Np} H-NOXA and the \textit{Ec} DOS PAS domain revealed a helical region embracing the heme group in the \textit{Ec} DOS structure is substituted by a well compacted βD strand in \textit{Np} H-NOXA resulting a slightly different fold (Figure 4.1D).
The primary structure of \( Np \) H-NOXA and the homologous sGC subunits were compared with each other to identify sequence conservation and divergence. Other than smaller 1 residue insertions/deletions around \( \beta C \) strand, a relatively long insertion was located between the \( \beta E \) and \( \beta F \) strands (Figure 4.1B). To confirm if both sGC subunits are capable of adopting PAS-like fold, homology models were constructed by MODELLER\(^{119} \) and validated by VERIFY3D\(^{120} \).

### 4.3.2 \( Np \) H-NOXA dimer organization

\( Np \) H-NOXA protein was purified as a dimer and two H-NOXA molecules are found in the asymmetric unit related by a 2-fold non-crystallographic axis. They are very similar conformationally (superposition of the 105 Ca atoms yields an r.m.s.d of 0.45Å) and interact with each other extensively by the juxtapositioning of the N-terminal helices and the extended stretch of residues preceding these helices resulting in 2,395Å\(^2 \) buried surface area\(^{121} \) (Figure 4.2A). Dimerization involves juxtapositioning of the N-terminal helices with its preceding residues and the face of the \( \beta \)-sheet (Figure 4.2A-4.2D).

Strikingly, H-NOXA monomers dimerize in a similar fashion as the \( Ec \) DOS dimers by utilizing the similar dimerization interface in spite of low sequence identity (10%) (Figure 4.3). The same fold and dimer organization is also present in a different heme sensor, \( Rm \) FixL\(^{122} \), also an STHK protein (Figure 4.3). More recently, such PAS dimerization was also observed in \( Av \) NifL\(^{123} \) (Figure 4.3), the N-terminal amphipathic helix constitutes part of the dimerization interface suggesting it functions as a conserved dimerization motif despite the limited sequence similarity in this helix.

### 4.3.3 sGC H-NOXA forms a similar PAS-like dimer
We compared the sequences between *Np* H-NOXA and sGC α1 and β1 subunits H-NOXA domains, a much higher degree of conservation was mapped to the dimerization region than the rest of the protein surface residues (Figure 4.4A). The hydrophobic nature of all 5 residues of the central cluster at the interface (L8, L13, F17, F100, and L102) is conserved as well. Previous studies also showed two sequence separated regions in the sGC β1 subunit, region 202-244 and 379-408 contribute to the heterodimerization of sGC α1/β1 and the homodimerization of sGC β1/β1. While region 379-408 has been a well known predicted coiled coil structure, it’s corresponding region was not present in the crystallized *Np* H-NOXA since it did not include the coiled-coil region. The region in *Np* H-NOXA corresponding to sGC β1 202-244 was including in the construct and formed the major part of the H-NOXA dimerization interface. Besides, the sequence identity in the N-terminal dimerization region between sGC α1 and β1 (43.9%) are much higher than the sequence identity between two full length subunits (29.8%), suggesting functional importance in this region. This indicates that sGC α1 and β1 H-NOXA domains could use the same interface to heterodimerize or homodimerize as observed in solution for sGC β1 H-NOXA domain comprising residues 202-344.

To experimentally validate the presence of a similar PAS-like dimer interface in sGC we performed alanine scanning mutagenesis studies on eight putative interface residues in the H-NOXA domain of sGC based on their importance in the dimerization of *Np* H-NOXA (Figure 4.4B, C and D). These residues are systematically mutated to Ala in sGCβ1 H-NOXA domain or full-length sGC to evaluate their importance in dimerization and enzyme function. Similar effects on both sGCβ1 H-NOXA homodimerization and full-length sGC heterodimerization are observed in a number of mutations. The most
significant decrease in dimerization was caused by Q309A, Q231A, and L322A, whereas S206 and F217 result in a small negative effect on H-NOXA homodimerization and to a lesser degree on full-length sGC heterodimerization (Figure 4.4B, C and D). The β1 F217A and α1 F285A mutations caused a 3.6- and 6.1-fold drop in NO stimulated activity, respectively, whereas the β1 Q309A and α1 Q368A mutations both lead to a complete loss of basal and NO stimulated guanylyl cyclase activity despite the presence of both subunits detected in the COS-7 cell lysate. All the other mutants had only a modest negative effect on guanylyl cyclase (Figure 4.4C and D) yielding a maximal decrease of NO-stimulated activity of just under 3-fold.

Taking into account the structural and sequence conservation and these structure-based mutagenesis data, it is very likely that the H-NOXA PAS-like domains of both α1 and β1 form a heterodimer similar to the Np H-NOXA dimer and that this PAS-like heterodimeric interface is crucial for proper functioning of sGC.

4.3.4 Possible implications for preferential heterodimerization of sGC

As an entirely reversed dimer orientation was formed up on deleting the first 7 residues of Np STHK 1-121, this region must play an important role in the correct positioning of the monomers within the dimer. After close inspection, a small network of molecular interactions has been identified involving residues Leu6, Ala22, Val30, and Gln31. The main chain N and O atoms of Leu6 make hydrogen bonds with the conserved Gln31 while its side chains interacts with a relatively conserved hydrophobic region around Ala22 and Val30 (Figure 4.2B and 4.2C). These interactions not only dictate the orientation of Np STHK but also could contribute to the well known preferential sGC heterodimerization. The corresponding residues are Leu274, Met290, and Leu298 in
sGCα1, or Ser206, Ile222, and Thr230 in sGCβ1. The larger sized residues at these positions in α1 homodimer, if present at all, would cause steric clash thus disfavoring such dimerization perhaps analogous to the disruptive effects of the Δ7 deletion in Np STHK. The smaller sized residues at these positions in β1 allow homodimer formation yet perhaps a heterodimeric α1/β1 is favored due to the complementary large/small residue interface differences between α1 and β1 subunits.

4.3.5 Possible role for PAS mediated allosteric regulation in sGC and STHK

Over the past decade, PAS domains have been identified in many signal transduction molecules and different forms have been found in animals, plants, and prokaryotes. A whole variety of ligands such as oxygen and light, redox potential and even protein-protein interactions can be sensed using PAS domains. The identification of these versatile sensory domains in STHK and sGC provides the possibility that they are used for sensing unknown small molecules. For the PAS-like domain in Np STHK, only small ligands are expected due to the limited size of the binding pocket. However for the sGC PAS-like domain, an insertion between βE and βF could result in a deeper binding pocket thus accommodate bigger ligands. An intriguing candidate could be the sGC allosteric activator YC-1 as no consensus result has been resolved regarding the binding site(s) on sGC. Various PAS domains can adopt a series conformational changes upon ligand binding including FG loop in Rm FixL, scissor-type movement in Ec DOS, N-terminal PAS helix unfolding in Hh PYP, a speculated helix-swap event, or PAS pocket mediated inter-subunit interactions in drosophila Per. Whether or not the PAS-like domain has a physiological sensory role, its presence in sGC does provide a unique opportunity for this domain to be therapeutically exploited as a drug.
target for sGC regulation. Such an approach was successful for the PAS kinase domain, whose pocket is mainly closed as well, using the SAR-by-NMR approach\(^{133}\).

### 4.3.6 PAS-like H-NOXA dimer implications for domain organization sGC

As H-NOXA domains are present in the center of both the \(\alpha 1\) and \(\beta 1\) subunits of sGC, the dimerized PAS-like domain of the homologous STHK H-NOXA therefore provides a framework on which additional sGC domains can be built on. We have generated a composite structural model for sGC \(\alpha 1/\beta 1\) heterodimer based on the following structural and functional properties (Figure 5). The guanylyl cyclase domains are most likely to be catalytically active as a heterodimer\(^{35}\) and modeled according to homologous adenylyl cyclase crystal structure\(^ {62,134}\). Although no experimental structural information is available for the CC region, a parallel coiled-coil segment is expected according to sensitive sequence analysis\(^ {58}\). The N-terminal domain of \(\beta 1\) subunit is modeled using the homologous \(Tt\) H-NOX structure\(^ {45,48}\) and \(Ns\) H-NOX structure\(^ {135}\). The H-NOX domain and the guanylyl cyclase domains are positioned close to each other due to the demonstrated trans interaction between them\(^ {66}\). The interactions likely involve D44-D45 in the H-NOX domain\(^ {136}\) and perhaps the surface where Gs binds adenylyl cyclase\(^ {62}\) in the guanylyl cyclase domain.

### 4.4 Conclusions

Our studies have elucidated critical hetero-dimerization interactions within sGC via a PAS-like domain dimer which, with its key central position and its \(Ec\) DOS-like allosteric ‘tools’, could function as a second allosteric/sensory module to regulate sGC either directly or indirectly.
Figure 4.1 Structure of H-NOXA domain of \textit{Np} STHK and its evolutionary, structural, and sequence relationships. (A) Schematic diagram of the H-NOX and H-NOXA domains present in proteins from animals and \textit{Nostoc} cyanobacteria. The purple interrupted line indicates the crystallized domain. (B) Structure-based sequence alignment of the H-NOXA domains of STHK, 2-CHSR, sGCa1, -\textit{\beta}1, and -\textit{\beta}2 subunits, and EcDOS (PDB id 1V9Z) and RmFixL (PDB id 1D06). Residues identical to sGCB1 are in bold. Semi-conserved hydrophobic residues, excluding heme PAS domains since they have a heme as their hydrophobic core, are highlighted in yellow; fully conserved residues in green. Residues at the \textit{Np} STHK dimer interface are labeled with a ‘#’; the putative sGC interface residues probed by mutagenesis are circled. Residues of Ec DOS and \textit{Rm} FixL that are in structurally equivalent positions with the H-NOXA structure are in uppercase. The Ec DOS and \textit{Rm} FixL sequences contains a few insertions at positions labeled ‘\textsuperscript{\^}’ and their insertion sizes are 3, 6, and 2, respectively for both Ec DOS and \textit{Rm} FixL. sGCB1 residues 204-244 have previously been shown to be key for dimerization\textsuperscript{51} and are boxed. The expressed construct end (K121) and start of the CC of \textit{Np}STHK are shown (hooked arrow). Domain abbreviations used: CC (coiled-coil), GAF (domain present in cGMP-regulated PDEs, adenylyl cyclases, and the \textit{E. coli} protein Fh1A), REC (receiver domain), and HPT (histidine phosphotransfer domain). The red underlines stretches of residues highlight deletions that resulted in lack of sGC heterodimerization and concomitant loss of activity; deletion of the orange underlines stretches of residues resulted in sGC heterodimerization but severely compromised cyclase activity; deletion of the blue underlines stretches of residues resulted in increased cyclase activity and decreased EC\textsubscript{50} values for BAY41-2272 in combination with nitric oxide; deletion of the interrupted blue underlined stretch of residues had a mixed effect in that this deletion caused a decrease in cyclase activity yet a decrease in EC\textsubscript{50} for BAY41-2272 in combination with nitric oxide\textsuperscript{137}. (C) Ribbon diagram of the \textit{Np} STHK domain monomer structure. Secondary structure elements are labeled. (D) Stereo-figure depicting superposition of the monomer of the H-NOXA domain of \textit{Np} STHK and the Ec DOS PAS domain (PDBid 1V9Z). \textit{Np} STHK and Ec DOS share common \textalpha{}A, \textalpha{}B, and \textalpha{}C helices and even helix \textalpha{}D although this helix is longer in Ec DOS and in a somewhat different orientation. The major difference between the \textit{Np} STHK and Ec DOS structure is that the Ec DOS pocket that harbors the heme (ball-and-stick) flanked by the long helix is collapsed in \textit{Np} STHK and filled with the new \textbeta{}D strand.
A. Animals (HNOX fused to HNOXA/CC/GC)

- sGC α1
- sGC β1

Cyanobacteria (stand-alone HNOX and adjacent HNOXA/CC containing gene)

Nostoc punctiforme
PCC 73102
ORF: Npun02000819
Npuno20000820

STHK

Nostoc sp.
PCC 7120
ORF: alr2278

2-CHSR

B. Protein sequences

STHK Npun
MAPPHLCLSPILILAPPPHRHAIRASPRPLVEWIPFEP-LVGL-MAQPVFFVTMTAEHELIDPLISGRFLGFLIPPI 815
sGCa1 rat
KPQEGLFLPSSTSTEMSPEEMAMNPEEMPSFRRSEGRTSARKQKKCRKSNPRLVTPPILEKTSEHMKICMKPI 987
sGCb1 rat
NOSQSIKLPLPELPRVNPEDFPLNFVPLVPRHDPFLHLHCVPHNLPP183
sGCb2 rat
VPPESLVEEPVCPAPPHRHAIRASPRPLVEWIPFEP-LVGL-MAQPVFFVTMTAEHELIDPLISGRFLGFLIPPI 847
2-CHSR Np
SSSKKSSKLSQ Oculus FLPSSTSTEMSPEEMAMNPEEMPSFRRSEGRTSARKQKKCRKSNPRLVTPPILEKTSEHMKICMKPI 892
Ec. Dos
qifPFLEGtrnGvliNEDVEFHFEPAPAVKLSGYKLQV-QNN-IDMLIPAtlhphkejiihneegkgqerBL 933
RmflxL
vrSrdahllRSLDTVpaKTVSWATDGTIVSIFNOXAVRPAFAYVI-GQN-LGILHpeyrhcdgylgrynajlSqvW 210

Coiled-coil domain

C. Diagrams

D. Diagrams with text

Np STHK
Ec. DOS

Np STHK
Ec. DOS
Figure 4.2 Homo-dimeric structure of H-NOXA domains of *Np* STHK. (A) Schematic diagram of the *Np* STHK dimer in two side views related to each other by a 90° rotation around the axis shown. (B) Stereo diagram of the dimer interface. View is that of from the top of H-NOXA dimer with the two N-terminal helices perpendicular to the plane of the paper. Hydrogen bonds are depicted as dashed lines. Residues that are equivalent to the ones targeted by Ala scanning mutagenesis in sGC are underlined. (C) Same as in (B) except rotated 90° along the horizontal axis such that the 2 N-terminal helices are now oriented vertically. (D) Electron density of the N-terminal dimer interface region of *Np* STHK. Depicted are residues 7-17 of both monomers and the omit |Fo|-|Fc| density (contoured at 2.5σ) after these residues were omitted (omitted portion is about 10% of the total structure). Residues 7-17 contains helix αA and a few preceding residues that are involved in dimer interface interactions. (E) Different dimer configurations of full length and truncated *Np* STHK. The full length *Np* STHK monomers entailing the crystallized dimer are shown in blue and yellow. A different dimer arrangement can be found in the asymmetric unit of the crystallized truncated Δ1-7 *Np* STHK: in keeping one of the H-NOXA monomers in a fixed superimposed position (yellow), the second H-NOXA subunit (red) interacts in an almost flipped orientation compared to the blue H-NOXA monomer resulting in a possible novel dimer organization (yellow and red). This results in drastic differences in relative positioning of the N- and C-termini of these different dimer organizations.
Figure 4.3 Structural similarities of the $Np$ STHK dimer and other PAS domain dimer containing proteins. The $Np$ STHK dimer is depicted along with $Ec$ DOS, $Rm$ FixL, and $Av$ NifL. The latter has an FAD molecule bound in each monomer and $Ec$ DOS and $Rm$ FixL both have heme moieties bound (ligands are shown in ball-and-stick).
Figure 4.4 Probing the putative H-NOXA dimer interface in sGC by mutagenesis. (A) Conservation of surface residues of the \( Np \) STHK H-NOXA monomer with the \( \alpha_1 \) and \( \beta_1 \) sGC subunits. The surface facing the dimer interface (left) and the opposite side (right) are shown. Non-conserved surface H-NOXA residues are shown as white. Conserved residues at the dimer interface or in the vicinity are labeled. (B) Probing dimerization state of sGC\( \beta_1 \) H-NOXA domain mutants using size exclusion chromatogram. A superdex 75 column (HR10/30, GE Biosciences) was used to test the oligomeric state of mutant H-NOXA proteins in which the following sGC\( \beta_1 \) residues were mutated to Ala (corresponding residue in \( Np \) STHK are in parentheses): S206 (L6), I208 (L8), F217 (F17), Q231 (Q31), Q309 (Q89), I311 (M91), L320 (F100), and L322 (L102). A mixture of molecular weight protein standards was also applied to the column for reference (grey). The theoretic molecular weight for the monomer of \( wt \) sGC\( \beta_1 \) H-NOXA domain is 17kD, which includes the His-tag. For comparison, the crystallized \( Np \) STHK elutes at 10.6ml (not shown) which is similar to that of the \( wt \) sGC\( \beta_1 \) H-NOXA domain confirming the dimeric state for both proteins. Note that theoretical molecular weight of the monomeric \( Np \) STHK protein is 14kD after His-tag cleavage. The injecting volume for each run is 100 \( \mu l \) and protein concentration is 1mg/ml. (C) Guanylyl cyclase activity and Western blot analysis of co-immunoprecipitated \( wt \) and 6 sGC\( \beta_1 \) mutants of sGC. Basal and NO-stimulated guanylyl cyclase activity of cell lysates with transfected sGC subunits are shown (upper panel). The Western blot (lower panel) includes co-immunoprecipitated sGC subunits pulled down with anti-sGC\( \beta_1 \) antibody (top gel) as well as the contents of the cell lysates (bottom). The Western blot was probed simultaneously with anti-sGC\( \alpha_1 \) and anti-sGC \( \beta_1 \) antibodies. (D) Guanylyl cyclase activity measurements and Western blot co-immunoprecipitation analysis of additional mutants of sGC generated in both \( \alpha_1 \) and \( \beta_1 \) subunits. These mutations include the sGC\( \beta_1 \) F217A and Q309A mutations and are also generated at the equivalent positions in sGC\( \alpha_1 \) (\( \alpha_1 \) mutations F285A and Q368A, respectively). Data presentation similar as in (C) but includes additional co-IP using an anti-sGC\( \alpha_1 \) antibody (top gel). This Western blot was also probed similarly as in (C). The difference in specific activity observed between experiments 4C and 4D (as reflected in the WT activity) is probably due the fact that two different batches of COS-7 cells were used in 4C and 4D. For each set of experiments, COS-7 cells were transfected three times and activity assays repeated three times with each measurement done in duplicate.
Figure 4.5 Possible subunit arrangement of sGC. Composite figure depicting possible subunit arrangement of heterodimeric sGC. The structure of the helical putative CC region is not known (labeled ‘?’) as well as the structure of the α1 N-terminal domain. The H-NOX domain is oriented such that helix αF and loop L1 in H-NOX are in proximity to the site that corresponds to where Gsα binds and regulates the homologous adenylyl cyclase (marked ‘X’). In H-NOX, both αF and the N-terminal helical subdomain (αA-αC), which includes the loop L1 containing the potential switch residues D44-D45, are postulated to shift upon activation. To illustrate the N-terminal subdomain shift, we have depicted the superimposed Ns H-NOX (red) and Tt H-NOX (blue) structures which are postulated to represent the basal and activated state, respectively, of an H-NOX domain. Note that we cannot rule out direct interactions between the H-NOXA domain and the GC catalytic domain, since we do not know the conformation of the intervening sequences and the structure and position of the CC region.
Table 4.1 Data collection, phasing, and refinement statistics for Np STHK.

<table>
<thead>
<tr>
<th></th>
<th>Native (8-121)</th>
<th>Se-peak (8-121)</th>
<th>Native (1-121)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.2398</td>
<td>0.9790</td>
<td>0.97934</td>
</tr>
<tr>
<td>Space group</td>
<td>P6122</td>
<td>P6122</td>
<td>C2</td>
</tr>
<tr>
<td>Cell dimension (Å)</td>
<td>a=b=72.33</td>
<td>a=b=72.26</td>
<td>a=95.21</td>
</tr>
<tr>
<td></td>
<td>c=169.08</td>
<td>c=169.11</td>
<td>b=44.36</td>
</tr>
<tr>
<td></td>
<td>α=β=γ=90º</td>
<td>α=β=γ=90º</td>
<td>c=59.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β=104.51º</td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.0</td>
<td>2.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Total observations</td>
<td>104640</td>
<td>150704</td>
<td>213337</td>
</tr>
<tr>
<td>Unique observations</td>
<td>18435</td>
<td>8621</td>
<td>12364</td>
</tr>
<tr>
<td>I/SigI</td>
<td>12.2 (2.5)</td>
<td>16.2 (6.9)</td>
<td>10.0 (1.6)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.68 (4.05)</td>
<td>17.48 (19.27)</td>
<td>3.4 (1.9)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (99.2)</td>
<td>100.0 (100.0)</td>
<td>87.4 (59.8)</td>
</tr>
<tr>
<td>R_sym (%)</td>
<td>8.0 (36.9)</td>
<td>9.3 (38.3)</td>
<td>11.9 (45.6)</td>
</tr>
<tr>
<td><strong>SAD phasing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of sites</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Figure of merit</td>
<td></td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.0-42.3</td>
<td>2.1-27.8</td>
<td></td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>1773</td>
<td>1732</td>
<td></td>
</tr>
<tr>
<td>No of waters</td>
<td>111</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Sulfate ions</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>R_factor (%)</td>
<td>23.9</td>
<td>19.9</td>
<td></td>
</tr>
<tr>
<td>R_free (%)</td>
<td>27.8</td>
<td>26.2</td>
<td></td>
</tr>
<tr>
<td>R.m.s.d. for bond lengths (Å)</td>
<td>0.012</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>R.m.s.d. for bond angles (º)</td>
<td>1.37</td>
<td>1.35</td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residues in:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-most favored regions</td>
<td>88.7</td>
<td>88.8</td>
<td></td>
</tr>
<tr>
<td>-additional allowed regions</td>
<td>11.3</td>
<td>11.1</td>
<td></td>
</tr>
<tr>
<td>-in generously allowed regions</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>-disallowed regions</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Values for the highest resolution shell are listed in parenthesis.
Cloning, purification and crystallization of the Coiled Coil (CC) region from *Rattus norvegicus* soluble guanylyl cyclase (sGC) β1 subunit

5.1 Introduction

As considerable amount of structural knowledge on sGC has accumulated from H-NOX domains, PAS-like domain and adenylyl cyclase, the only domain for which structural information is lacking is the CC domain. A recent *in silico* study has unveiled a novel class of signaling module that bridges diverse N-terminal sensory domains and various C-terminal catalytic domains ranging from histidine kinases, PP2C phosphatases, NtrC-like AAA+ ATPases and diguanylate cyclases to guanylyl cyclases. Although members in this signaling helix (S-helix) family shared limited but detectable sequence identity among each other, their secondary structures are predicted to be entirely helical and likely form parallel coiled-coil structure. To our knowledge, no S-helix coiled-coil structures have yet been determined so the elucidation of sGC CC domain would not only contribute to the understanding of guanylyl cyclase dimerization, activation, domains cross-talk but also provide a structural prototype for the entire S-helix family. Moreover, coiled-coil structure have been an interesting drug target in the cases of HIV type 1 gp41, human T-cell leukemia virus envelope glycoprotein and human Bcr-Abl oncoprotein. The structural information gained from the sGC CC could present an unexplored opportunity for developing novel sGC inhibitors by targeting the amphipathetic interactions between coiled-coils or even cavities formed upon dimerization.
5.2 Materials and methods

5.2.1 Cloning and mutagenesis of rsGCβ1 CC domain

The coding sequence (residues 348-409) for the CC region from rsGCβ1 was subcloned into pET15b vector between Nde I and BamH I restriction sites using the following primer set:

*Forward primer: 5’- gga att cca tat ggc tac acg aga cct ggt cct ttt- 3’*
*Backward primer: 5’- cgc gga tcc tca ctt gtg tct cag etc att ggc aac- 3’*

The resulting pET15b_rsGCβ1_348-409 plasmid encodes the following polypeptides in the T7 expression region:

```
mgsshhhhhsglvpr/gshmATRDLVLLGEQFREEYKLTQERLTLQRLALEDEKKKTDTLLYSVLPPSVANELRK (The residues that belong to rsGCβ1 are capitalized. Other residues are introduced by cloning artifact.)
```

In addition I371M mutation generated to facilitate Se-Met SAD-phasing structure determination was introduced into pET15b_rsGCβ1_348-409 by site directed mutagenesis method using the following primer set:

*Forward primer: 5’- ca caa gag ctc gag cca atG etc aca gac agg ctg c- 3’*
*Backward primer: 5’- g cag cct gtc tgt gag Cat ttc cag ctc ttg tg- 3’*

5.2.2 Expression and purification of rsGCβ1 CC domain

The pET15b_rsGCβ1_348-409 vectors were transformed into E. coli BL21 (DE3) pLysS cells (Invitrogen). The bacteria were grown in LB media containing 50 μg/mL ampicillin and 37μg/mL chloramphenicol at 37°C until a cell density of 1.2–1.4 OD600 was reached. The cultures were induced with 300μM isopropyl-β-D-thiogalactopyranoside (IPTG) for 8 hours at 37°C. The cells were
harvested by centrifugation at 6000 rpm, followed by freezing at -80 °C. The pellet was thawed on ice, resuspended in buffer A containing 20mM Tris-HCl, 100mM NaCl, and lysed by sonication. The crude lysate was clarified by centrifugation at 16,000 rpm for 15 minutes, incubated with 3mL Ni-NTA (Qiagen) for 4 hours and washed with buffer A plus 15mM Imidazole extensively. The protein was eluted using 50mM buffer A plus 350mM Imidazole and dialyzed against 4L buffer A overnight. Thrombin (Sigma) digestion was carried out afterwards at 20°C for 8 hours and monitored by SDS-PAGE. The untagged protein product was loaded onto a Hitrap Q-sepharose (Amersham-Pharmacia Biotech) column and eluted with NaCl gradient. Fractions containing protein of interest were pooled together, concentrated and frozen in aliquots at -80 °C. Prior to crystallization trials, a gel filtration polishing step using superdex75 (Amersham-Pharmacia Biotech) was applied and the final protein buffer is 10mM Tris-HCl, 100mM NaCl.

Selenomethionine substituted rsGCβ1_348-409-I371M mutant protein was expressed and purified as described above except that M9 minimal medium is used supplemented with essential amino acids and selenomethionine.

5.2.3 rsGCβ1 CC domain crystallization

Initial screening (Index, Screen I, Screen II, and PEG/Ion, Hampton Research) indicated PEG 3350 solutions in different buffers at pH values close to 8.0 as the most promising precipitants. The presence of divalent metal cations such as MgCl₂ was also essential to obtain crystals. The crystallization conditions were optimized by varying the PEG type and concentration, salt type and concentration, the pH and the buffer. Preliminary crystallization condition (0.2 M Magnesium chloride hexahydrate, 0.1 M Tris
pH 8.0, 18% PEG 3350) routinely gave needle cluster crystals that failed to diffract in ALS 4.2.2, NSLS X29 synchrotron X-ray sources. Further optimization was performed using Additive Screen (Hampton Research). Two chemicals (phenol and n-butanol) helped the CC needle crystals to obtain the third dimension. Eventually, X-ray diffraction-quality crystals were obtained at 4 ºC with the sitting-drop method by mixing 2µl of the protein solution with 1µl of the precipitant solution and were equilibrated with a well containing 200 µl of the precipitant solution (0.2 M Magnesium chloride hexahydrate, 0.1 M Tris pH 8.0, 8% PEG 3350, 8mM phenol). Crystals of average dimensions of 300*20*20µm³ were commonly grown in the above conditions and used for data collection with additional 20% glycerol as cryoprotectant.

The selenomethionine substituted rsGCβ1_348-409-I371M mutant protein was crystallized in a similar yet different condition that contains 0.2 M Lithium Sulfate monohydrate, 0.1 M Tris pH 8.0, 6% PEG 3350.

5.2.4 rsGCβ1 CC domain structure determination and refinement

A native 2.6 Å resolution dataset for the rsGCβ1_348-409 protein was collected at NSLS X29 and processed with HKL2000. The crystals belong to space group C2, with cell dimensions a=152.739Å, b=65.964Å, c=98.385Å, β=129.840 and eight molecules in the asymmetric unit. To obtain crystallographic phase information, a SeMet crystal of the rsGCβ1_348-409-I371M mutant protein was used for SAD phasing. To our surprise, a higher resolution (2.5Å) dataset was obtained from a selenomethionine substituted crystal collected at the selenium peak wavelength at the same beamline and was used for phasing as well as refinement. The rsGCβ1_348-409-I371M monomer structure was initially solved by the SOLVE/RESOLVE in space group I2₁2₁2₁, with cell dimensions a=68.938Å,
b=99.376Å, c=115.690Å. An electron density map of a workable quality was obtained and allowed us to build a complete model of rsGCβ1_348-409-I371M monomer. However, non-documented high solvent content (86%) and high R/R_free values (42%/45%) indicated errors in space group assignment. The indexing step was revisited revealing the presence of higher metric symmetry, which was indicative of pseudo-merohedral twinning. Thus the data was reduced to C2 with cell dimension (a=152.366Å, b=68.907Å, c=99.365Å, β=130.606) that is almost isomorphous with the native crystal. The serendipitous geometry relationship of a*cosβ almost equal to –c allowed a real space rotation about an axis perpendicular to the crystallographic 2-fold as the twin operation (h+2l, -k, -l). PHASER was used to locate the eight monomers in the C2 asymmetrical unit using the monomer structure as the search model. The refinement using the twin operator improved the R factor and eliminated much of the disorder in the electron-density maps. The amino-terminal residues 1-5 and the carboxy-terminal residues 59-64 were omitted from all subunits of the refined structure owing to the lack of electron density. The final refinement used diffraction data from 10.0 to 2.5 Å resolution and gave an R factor of 0.30 with an R_free of 0.35.
Table 5.1 X-ray data collection, phasing and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Native</th>
<th>Se-Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.008</td>
<td>0.9795</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Cell dimension (Å)</td>
<td>a=152.739, b=65.964, c=98.385, β=129.840</td>
<td>a=152.366, b=68.907, c=99.365, β=130.606</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0-2.6, (2.69-2.6)</td>
<td>50.0-2.5, (2.59-2.5)</td>
</tr>
<tr>
<td>Total observations</td>
<td>65155</td>
<td>393394</td>
</tr>
<tr>
<td>Unique observations</td>
<td>21986</td>
<td>53540</td>
</tr>
<tr>
<td>I/SigI</td>
<td>20.2 (2.3)</td>
<td>25.5 (6.1)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.0 (2.8)</td>
<td>7.3 (6.8)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>94.6 (96.2)</td>
<td>99.5 (99.4)</td>
</tr>
<tr>
<td>Rmerge a (%)</td>
<td>5.8 (58.8)</td>
<td>7.2 (43.0)</td>
</tr>
</tbody>
</table>

MAD phasing

<table>
<thead>
<tr>
<th>Resolution</th>
<th>No. of Se sites b</th>
<th>FOMRESOLVE c</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20-2.5</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Refinement

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>No. of protein atoms</th>
<th>No. of hetero atoms</th>
<th>No. of waters</th>
<th>R-factor d</th>
<th>R-free</th>
<th>RMSD bond lengths</th>
<th>RMSD bond angles</th>
</tr>
</thead>
</table>

Ramachandran plot

<table>
<thead>
<tr>
<th>Most favored (%)</th>
<th>Additional allowed (%)</th>
<th>Generously allowed (%)</th>
<th>Disallowed (%)</th>
</tr>
</thead>
</table>

Values in parentheses represent the highest resolution shell.
BIBLIOGRAPHY

Reference List


105. Lamothe, M., Chang, F.J., Balashova, N., Shirokov, R. & Beuve, A. Functional characterization of


