

Alma Mater Studiorum – Università di Bologna

**DOTTORATO DI RICERCA IN
BIOLOGIA CELLULARE E MOLECOLARE**

Ciclo XXVIII

**Settore Scientifico Disciplinare: 05/E2
Settore Concorsuale di afferenza: BIO/11**

**Structural characterization of
meningococcal vaccine antigen NadA
and of its transcriptional regulator NadR
in ligand-bound and free forms.**

Presentato da: **Alessia Liguori**

Coordinatore Dottorato
Chiar.mo Prof. **Davide Zannoni**

Relatore
Dr. **Matthew James Bottomley**

Co-relatore
Dr. **Enrico Malito**

Co-relatore
Chiar.mo Prof. **Vincenzo Scarlato**

Esame finale anno 2016

Contents

Aim of the thesis	6
Introduction	8
Reverse Vaccinology and Structural Vaccinology.....	9
The Serogroup B Meningococcus Vaccine.....	11
<i>Neisseria meningitidis</i> colonization and invasion.....	12
Meningococcal virulence factors and adhesion.....	13
Transcriptional regulators.....	14
Genome plasticity and phase variation.....	15
Part One	
Crystal structures reveal the molecular basis of ligand-dependent regulation of NadR, the transcriptional repressor of the meningococcal antigen NadA.....	17
Abstract.....	18
The Neisserial adhesin Regulator (NadR).....	19
The MarR family of transcriptional regulators.....	19
Experimental Procedures	21
Bacterial strains, culture conditions and mutant generation.....	21
Protein production and purification.....	22
Size-exclusion high-performance liquid chromatography (SE-HPLC) coupled with Multi-angle laser light scattering (MALLS).....	22
Differential Scanning Calorimetry (DSC).....	23
Surface plasmon resonance (SPR).....	23
Crystallization of NadR in the presence or absence of 4-HPA.....	24
X-ray diffraction data collection and structure determination.....	24
Results	26
NadR is dimeric and is stabilized by specific hydroxyphenylacetate ligands.....	26
NadR displays distinct binding affinities for hydroxyphenylacetate ligands.....	28
Crystal structures of holo-NadR and apo-NadR.....	29
The holo-NadR structure presents only one occupied ligand-binding pocket.....	31

Analysis of the pockets reveals the molecular basis for asymmetry and stoichiometry.....	32
Apo-NadR structures reveal conformational flexibility.....	34
4-HPA stabilizes concerted conformational changes in NadR that prevent DNA-binding...	35
A single conserved leucine residue (L130) is crucial for dimerization.....	37
Discussion	40
 Part Two	
New strategies towards a crystal structure of meningococcal antigen NadAv3.....	47
 Abstract.....	48
The <i>Neisseria meningitidis</i> adhesin A (NadA).....	49
The Trimeric Autotransporter Adhesin (TAA) family – structural organization.....	49
NadA variants.....	51
 Experimental Procedures	52
NadA constructs cloning, expression and mutagenesis.....	52
NadA protein production and purification.....	52
Size-exclusion high-performance liquid chromatography (SE-HPLC) coupled with Multi-angle laser light scattering (MALLS).....	52
Differential Scanning Calorimetry (DSC).....	53
Surface plasmon resonance (SPR).....	53
Crystallization of NadAv3 proteins and X-ray diffraction data collection.....	54
 Results	56
Structure based-design: seeking the minimal N-terminal domain of NadAv3	56
Strategies towards the crystal structure of NadAv3 ₂₄₋₁₇₀ construct.....	58
NadAv3 mutagenesis to promote crystallization.....	62
High-quality NadA reagents for functional studies.....	65
 Discussion.....	67
 Acknowledgments.....	71
 References.....	72

*“A Matthew, guida e stimolo
per la mia crescita professionale e personale.”*

Aim of the thesis

Serogroup B *Neisseria meningitidis* (MenB) is the cause of an acute, potentially severe infection, known as invasive meningococcal disease (IMD) with two peaks in disease incidence occurring among adolescents and young adults 16 to 21 years of age. *Bexsero* is the first genome-derived vaccine against MenB, and it has recently been approved in >35 countries worldwide. Neisserial adhesin A (NadA), a meningococcal trimeric autotransporter adhesin (TAA) that acts in adhesion to, and invasion of, host epithelial cells, is one of the three protein antigens included in *Bexsero*. The main aim of this work was to obtain detailed insights into the structure of NadA variant 3 (NadAv3), the vaccine variant, and into the molecular mechanisms governing its transcriptional regulation by NadR (Neisseria adhesin A Regulator). The amount of NadA exposed on the meningococcal surface influences the antibody-mediated serum bactericidal response measured *in vitro*, which in turn correlates with protection in immunized subjects. A deep understanding of *nadA* expression is therefore important, otherwise the contribution of NadA to vaccine-induced protection against meningococcal disease may be underestimated. The abundance of surface-exposed NadA is regulated by the ligand-responsive transcriptional repressor NadR. The functional, biochemical and high-resolution structural characterization of NadR is presented in the first part of the thesis (Part One). These studies provide detailed insights into how small molecule ligands, such as hydroxyphenylacetate derivatives, found in relevant host niches, modulate the structure and activity of NadR, by 'conformational selection' of inactive forms. These findings shed light on the regulation of a key virulence factor and vaccine antigen of this important human pathogen.

In the second part of the thesis (Part Two), strategies involving both protein engineering and crystal manipulation to increase the likelihood of solving the crystal structure of NadAv3 are described. The first approach was the rational design of new constructs of NadAv3, based on the recently solved crystal structure of a close sequence variant (NadAv5). Then, a comprehensive set of biochemical, biophysical and structural techniques were applied to investigate all the generated NadAv3 constructs, aiming to faithfully represent its natural trimeric status, essential for reliable structural, functional and epitope mapping studies. The well-characterized trimeric NadAv3 constructs represented a set of high quality reagents which were validated as probes for functional studies and as a platform for continued attempts for protein crystallization. Mutagenesis studies and screenings to identify a new crystal form of NadAv3 were performed to improve crystal quality, ultimately allowing the collection of several high quality X-ray diffraction data sets; structure determination is ongoing. The atomic resolution structure of NadAv3 will help to understand its biological role as both an adhesin and a vaccine antigen. For example, the high resolution structure will enable epitope mapping studies using human antibodies and thus permit a

deeper understanding of the molecular determinants of antibody binding and protective epitopes. In addition, it will help to understand the molecular basis of host-pathogen interactions mediated by specific human cell receptors.

Introduction

Reverse Vaccinology and Structural Vaccinology

Genomics tools and the exponentially growing number of bacterial genome sequencing projects have changed the landscape of modern biology providing new opportunities for vaccine development. The complete genome of a bacterium represents a large reservoir of genes encoding for potential antigens that can be selected and tested as vaccine candidates. Therefore, potentially surface-exposed proteins can be identified in a reverse manner, starting from the genome rather than from the microorganism. This approach has been termed Reverse Vaccinology (RV) [1]. Bioinformatics algorithms are used to select open reading frames (ORFs) encoding putative surface-exposed or secreted proteins, which are potentially recognized by antibody and can therefore be considered as vaccine antigens. The identification of such surface proteins is based on specific properties including the presence of signal peptide sequences, membrane spanning regions, lipoprotein signature, and motifs such as LPXTG sortase attachment sites. Sequence homology analyses can additionally help the antigen identification process, comparing homology both to known virulence factors or protective antigens from other pathogens and to human proteins to avoid autoimmune problems [2]. The candidate surface antigens are therefore produced as recombinant proteins and tested for their immunogenicity in a relevant animal model in order to evaluate their potential as vaccine candidates. The reverse vaccinology approach has been strengthened by the development of proteomic techniques to identify vaccine candidates against bacterial infections [3-5]. In the proteomics approach to bacterial vaccine development the surface-located or secreted bacterial proteins are first separated using two-dimensional (2-D) electrophoresis gel, followed by digestion of each protein into its peptide fragments using a specific protease (e.g. trypsin). The molecular mass of each proteolytic digested fragment is then accurately measured using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) [6, 7]. Finally, the generated peptide mass fingerprint is used as an input to allow a database search of predicted masses coming from the digestion of a list of known proteins. If a protein sequence in the reference list does not match the experimental values, the peptide can be identified using tandem mass spectrometry, which provides sequence information on the proteolytic peptides [8]. The effectiveness of the 2-D gel-based platform integrating surface and immune-proteomics analysis was demonstrated by the identification of major meningococcal vaccine antigens [9]. After identification of potential candidates by RV, their testing is facilitated by use of high-throughput screenings for protective immunity and correlation of protection. Combining proteomics with serological analysis is another useful refinement for identifying potential vaccine candidates [8].

In addition, systematic transcriptomic and proteomic gene expression analysis can support RV in the identification of gene-level responses, which are correlated with protection

in vivo and then facilitating the rational design of a hypothetical vaccine candidate [10]. These strategies could help vaccine design for pathogens for which vaccines are not yet available, as well as parasites and viruses.

Once candidates have been identified further potential issues have to be solved before they can be used in a vaccine formulation. For instance, the surface antigens identified are not always abundant, and immune evasion strategies set up by the micro-organism can impact their potential. One of the most frequent examples is the sequence variation of surface antigens across circulating strains. From a more practical viewpoint, the selected antigens may show low stability when expressed as recombinant proteins. Structural vaccinology (SV) can represent the solution for many of these issues [11]. In a process analogous to structure-based drug design of small-molecule pharmaceuticals, where lead candidate inhibitory molecules are rationally-optimized in structure-guided manner, structural information on antigens and their protective epitopes can also be instrumental during the optimization phases.

Notable applications of structural vaccinology in the field of bacterial protein antigens include (i) characterization of the immune response through epitope mapping to provide insights into the molecular features recognized by the host immune response upon infection by the pathogen or following immunization. Epitope mapping experiments produce information about the immunogenic regions of the protein, showing which parts of a surface-bound antigen are exposed and therefore accessible to antibodies. Another important application for structural vaccinology is (ii) the possibility to improve the biochemical stability, and homogeneity of a candidate, stabilizing the folding and reducing degradation and tendency to aggregate. It is also possible (iii) to engineer a protein antigen in order to overcome limits imposed by sequence variability. SV can drive the design of chimeric antigens that display epitopes from multiple proteins to elicit an immune response with wider specificity. Overall, a structurally re-designed molecule can become an antigen with increased immunogenicity, and efficacy. A successful SV approach will also facilitate scale up, and generate an antigen that can be more easily produced, more homogeneous and stable over time. In summary, SV can use knowledge of biochemical, biophysical, structural, immunological & functional properties of biomolecules to benefit vaccine development, encompassing several steps of the process starting from antigen selection up to vaccine approval.

An example of how a structure-based approach has already been used in a preclinical vaccine design program is provided by a combination of NMR spectroscopy and X-ray crystallography to obtain structural insights of the immunodominant domain of GNA1870, a protective antigen of *N. meningitidis* identified by RV. The epitopes of

bactericidal antibodies against several meningococcal strain variants were mapped onto the NMR structure of GNA1870, providing the basis for the rational design of an engineered form of GNA1870 containing several cross-protective, B cell epitopes. A protein domain is defined as an independent unit that can have an independent function in a single-domain protein or can contribute to the function of a multidomain protein in cooperation with other domains [12]. The new GNA1870 antigen had a conserved backbone that carried an engineered surface containing specificities for all three variant groups, demonstrating that the structure-based design of an engineered antigen is an efficient way to generate a broadly protective antigen [13].

The Serogroup B Meningococcus Vaccine

The concept of reverse vaccinology was developed and applied for the first time to *N. meningitidis* serogroup B (MenB). *N. meningitidis* is the major cause of meningitis and sepsis, two devastating diseases that can kill children and young adults within hours, despite the availability of effective antibiotics. *N. meningitidis* is a Gram-negative bacterium that colonizes asymptotically the upper nasopharynx of about 5–15% of the human population, establishing a commensal relationship between the host and the bacterium that fails or becomes dysfunctional in case of disease [14]. This condition represents the only known reservoir for meningococcal infection but may also contribute to establishing host immunity [15]. For unknown factors dependent on both the host and pathogen, the meningococcus can invade the pharyngeal mucosal epithelium and disseminate into the bloodstream causing septicaemia or cross the blood-brain barrier and enter the cerebrospinal fluid, causing meningitis. Although reasons leading to the bacterial invasion are not well known, environmental factors that damage the nasopharyngeal mucosa, together with the lack of a protective immune response could increase the incidence of invasive meningococcal disease. *N. meningitidis* can be classified in 13 serogroups on the basis of the chemical composition of the capsule polysaccharide, five of which (A, B, C, W-135 and Y) are responsible for more than 95 % of total cases of invasive disease. Vaccines against serogroups A, C, W-135 and Y were developed in the 1960s by using the purified capsular polysaccharide as antigen. At the turn of the century, improved second-generation, conjugated vaccines were introduced, where the polysaccharide components were linked to a carrier protein, and which provide effective protection in all age groups [16]. However, the chemical composition of the polysaccharide of serogroup B, which resembles a molecule present in human tissues, makes a polysaccharide-based vaccine poorly immunogenic and a possible cause of autoimmunity. In the last 40 years much effort has been directed to the identification of meningococcus B protein antigens as the basis of new vaccines. However,

the high variability of these proteins among the different MenB strains represents a serious obstacle to the production of a globally effective anti-MenB vaccine [15]. Reverse vaccinology has therefore proven to be a rapid and reliable approach to identifying vaccine candidates. The three most immunogenic antigens on the basis of their ability to induce bactericidal activity or *in vivo* passive protection were selected to be used in a multicomponent vaccine. They were NHBA [17], fHbp [18, 19], and NadA [20, 21]. Two other antigens (named GNA2091 and GNA1030) were also selected. To further enhance their immunogenicity and facilitate large-scale manufacturing of the vaccine components, four of the selected antigens were combined into two fusion proteins so that the resulting protein vaccine contained three recombinant proteins. The antigen NHBA was fused to GNA1030 while GNA2091 was fused to fHbp. NadA was included as a single antigen as it did not perform well when fused to a partner [22-24].

***N. meningitidis* colonization and invasion**

Colonization of the upper respiratory mucosal surfaces by *N. meningitidis* is the first step in the establishment of a human carrier state and invasive meningococcal disease (Figure 1).

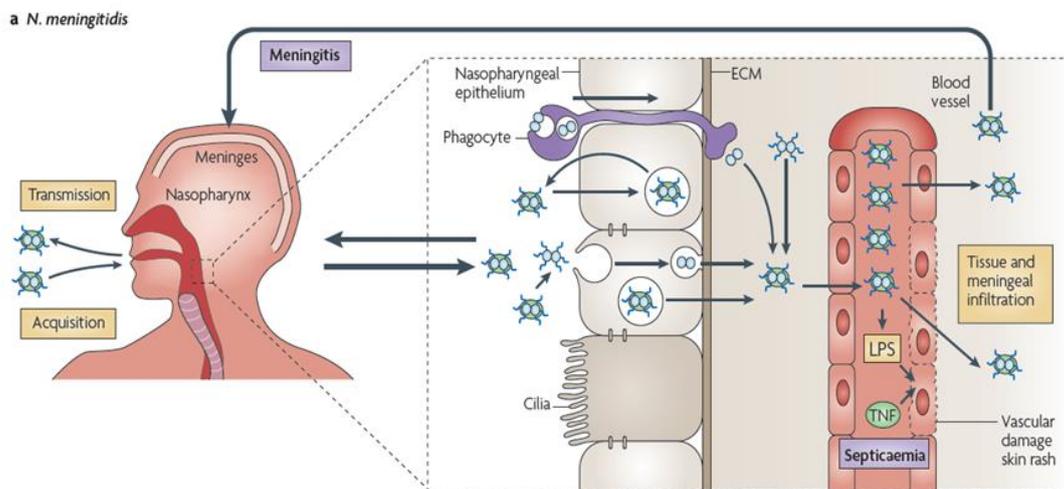


Figure 1. Stages in the pathogenesis of meningococcus. *N. meningitidis* may be acquired through the inhalation of respiratory droplets. The organism establishes intimate contact with non-ciliated mucosal epithelial cells of the upper respiratory tract, where it may enter the cells briefly before migrating back to the apical surfaces of the cells for transmission to a new host. Asymptomatic carriage is common in healthy adults in which bacteria that enter the body by crossing the epithelial barrier are eliminated. In susceptible individuals, once inside the blood, *N. meningitidis* may survive, multiply rapidly and disseminate throughout the body and the brain. Meningococcal passage across the brain vascular endothelium (or the epithelium of the choroid plexus) may then occur, resulting in infection of the meninges and the cerebrospinal fluid [25, 26].

Initial contact of meningococci with nasopharyngeal epithelial cells is mediated by Type IV pili, the host receptor for which may be the I-domain of integrin α chains or possibly CD46 [27]. At this level, the downregulation of the capsule may activate attachment [26], allowing meningococci a closer adherence to the host epithelial cells. It results in the formation of cortical plaques and leads to the recruitment of factors ultimately responsible for the formation and extension of epithelial cell pseudopodia that engulf the meningococcus. Intimate association is mediated by interaction of the bacterial opacity proteins, Opa and Opc, with CD66/CEACAMs and integrins, respectively, on the surface of the epithelial cell and is one trigger of meningococcal internalization [28]. The next steps of meningococcal internalization, intracellular survival and transcytosis through the basolateral tissues and dissemination into the bloodstream are less well studied [15].

Meningococcal virulence factors and adhesins

The virulence of *N. meningitidis* is influenced by multiple factors, including both genetic mechanisms, allowing the bacteria to vary its phenotype and adapt to the host, and iron sequestration mechanisms. Additionally, meningococci express multiple molecules acting as endotoxin, secreted factors or surface proteins, located in different compartments of the meningococcal cell membrane, which interact with host cellular molecules. The key structures at the interface between the meningococcus and the host are the polysaccharide capsule and/or lipopolysaccharide (LPS) that may shield bacterial surfaces from the host innate and adaptive immune effector mechanisms, and the protruding surface proteins that are known as pili [26]. Pili are filamentous structures consisting of protein subunits that extend from the bacterial surface, and these seem to be the main players in the initiation of the interaction between meningococcus and the host cell [29, 30]. Pili facilitate adhesion to host tissues, further aided by the outer membrane adhesins Opa and Opc. The opacity proteins (Opa and Opc) are integral outer membrane proteins that mediate pathogen-host interaction adhering to and invading epithelial and endothelial cells. Both bind the heparan sulphate proteoglycans and sialic acids [31, 32] but they also display a degree of receptor specificity [32].

Numerous additional minor adhesins are generally expressed at low levels during *in vitro* growth but may be important in *in vivo* infections. Neisseria hia homologue A (NhhA) mediates low levels of adhesion to epithelial cells, to heparan sulphate proteoglycans (HspGs) and to laminin [33]. Adhesion penetration protein (App) regulates interactions between the bacteria and the host tissue by mediating adhesion during the early stages of colonization, before it is autocleaved. At later stages, App autocleavage may allow bacterial detachment, therefore facilitating bacterial spread [34]. Meningococcal serine protease A

(MspA) expressed by several but not all virulent *Neisseria* strains mediates binding to both epithelial and endothelial cells and to elicit the production of bactericidal antibodies [35]. Multiple adhesin family (Maf) is a family of glycolipid adhesins characterised first for the gonococcus that may play a role in Opa-independent invasion [36]. Neisserial adhesin A (NadA) is a member of the trimeric autotransporter adhesins (TAAs) belonging to the Oca (oligomeric coiled-coil adhesin) family and is involved in adhesion and invasion of *N. meningitidis* [20, 21].

The regulation of NadA is part of this study and is discussed in Part One of this thesis, while the strategies towards a crystal structure of the NadA variant 3 will be described in Part Two.

Transcriptional regulators

During infection, *N. meningitidis* can invade diverse sites within the human host, which represent different niches with respect to nutrients, environmental factors and competing microorganisms. Therefore it is subjected to constant selective pressures, and its ability to rapidly adapt its metabolism and cellular composition to environmental changes is essential for its survival [37]. Bacteria have two major and complementary mechanisms for adapting to changes in their environment: changing their genotype (genome plasticity) or altering gene expression, both leading to phenotypic variations. The differential expression of potential virulence factors depends largely on the activity of transcriptional regulators, whose activity plays an important role for example in the infection process of *N. meningitidis*. Relatively few transcriptional regulators are found in the pathogenic *Neisseriae* [38]; 36 putative regulators in *N. meningitidis* (strain MC58) and 34 in *N. gonorrhoeae* (strain FA1090), compared to the free-living *E. coli*, which harbours more than 200 transcriptional regulators. The paucity in transcriptional regulators may possibly be related to the restricted ecological niche of the *Neisseria spp.* which are human-adapted pathogens for which there is no other known reservoir. Until now, only few of the predicted 36 transcriptional regulators in *N. meningitidis* MC58 have been characterized. Two of the transcriptional regulators in *N. meningitidis* are members of the MarR family and are encoded by the genes NMB1585 and NMB1843. The structure of the transcription factor NMB1585 has been solved, but its physiological role has not been characterized and therefore the identity of any natural ligand(s) that may modulate its activity is unknown [39]. The product of the NMB1843 gene is the Neisserial adhesin Regulator (NadR) - a MarR-family transcriptional regulator of 16.6 kDa per monomer and it has been demonstrated to repress expression of the meningococcal adhesin NadA [40, 41]. NadR is the subject of an extensive structural and biophysical characterization in this thesis and will be discussed below in detail (Part One).

Several MarR-family transcriptional regulators have previously been identified and described for their activity. In meningococcus: the ferric uptake regulator (Fur) is involved in the response to iron [42-44] and has even been shown to indirectly control gene expression through small regulatory RNA molecules [40, 45]; Zur is the second Fur-like regulator that responds specifically to Zn^{2+} and controls Zn^{2+} uptake by regulating a TonB receptor that functions in high affinity Zn^{2+} acquisition [46]. Adaptation to oxygen-limited conditions as encountered during infection of the human host is mediated by the transcriptional activator FNR (Fumarate and Nitrate Reductase regulator), whose DNA-binding ability is stabilized in the presence of oxygen [47, 48]. Upon conditions of oxygen limitation, this regulator enables the meningococci to survive by switching to enhanced sugar fermentation and expression of a denitrification pathway, utilizing nitrite instead of oxygen as a respiratory substrate [49]. NsrR acts as a repressor of a regulon of genes which responds to nitric oxide [50, 51]. The LysR-type regulator CrgA is upregulated upon contact with human epithelial cells [52]; it acts as a repressor of transcription of its own gene and as an activator of transcription of the *mdaB* gene [53]. NMB0573 (annotated as AsnC) is a global regulator controlling the response to poor nutrient conditions, which are perceived by binding of this regulator to leucine and methionine, two amino acids representing general nutrient abundance [54].

Although extensive transcriptional regulation is expected to accompany both the survival and the infection process of *N. meningitidis*, limited information about transcriptional regulation is available. Only a few of the predicted regulators have been characterized and the regulons of even fewer have been deeply studied, including those involved in the adaptation of meningococcus to iron and oxygen limitation and response to nitric oxide.

Genome plasticity and Phase variation

In order to adapt to changing microenvironments and avoid the host immune defences, the meningococcus possesses mechanisms for rapid genome variation and diversification. The genome plasticity is promoted by spontaneous mutational mechanisms. These events originate either from local genomic changes caused by repeat sequences, phase and antigenic variation, recombination and horizontal gene transfer, or globally from mutated alleles. Repeat sequence elements facilitate the duplication or deletion of regions of the genome, as well as recombination, and thereby establish small and large alterations. The addition and deletion of repeat units lead to the molecular mechanism of phase variation in *Neisseria*, most often owing to slipped-strand mispairing (SSM). The presence of repeat units causes a slippage of the synthesis strand over the template strand during replication that leads to the addition or the deletion of units in the newly synthesised strand [55]. The number of repeats can influence translation or transcription by introducing frameshift mutations or

changing critical promoter spacing, resulting in high frequency on-off switching or modulation of the level of expression of genes usually associated with surface-exposed antigens [56-59]. In meningococcus a considerably high quantity of phase-variable genes have been identified in which phase variation is used to alter surface-exposed molecules such as outer-membrane proteins PorA, Opc, Opa, pili and specific adhesins, as well as LPS and capsule [55, 60, 61]. In particular, the expression of NadA is phase variable and a tetranucleotide tract (TAAA) located upstream of the *nadA* gene promoter has been demonstrated to control this phenomenon, through an altered sigma-factor binding [40]. Whole-genome-sequence analyses have largely confirmed the importance of varying surface-exposed antigens for allowing bacterial commensals and pathogens to evade the immune system of their host and to adapt to changeable environments.

Concluding remarks

Host-pathogen interaction is a dynamic process that can lead to different outcomes such as an equilibrium known as commensalism or the establishment of a disease. The factors that lead *N. meningitidis* to establish the infection, switching from commensal to pathogenic are still poorly understood. For these reasons, a better understanding of the causes and mechanism that mediate the expression of proteins involved in the interaction with host tissues is needed, both for predicting the effectiveness of a vaccine which contains these proteins and for characterizing at the molecular level novel strategies of bacterial populations to changing host environments. *N. meningitidis* has to change gene expression repertoire in order to adapt and survive in the different tissues during an infection of the human host. At the same time the bacteria escape the host immune response, which targets mostly the same structures used by the meningococcus to interact with the host, by surface structure expression variability and redundancy. A multi-disciplinary approach based on molecular genetics, biochemical, biophysical and structural analyses, will provide molecular knowledge of the transcriptional regulation of antigen expression.

Part One

**Crystal structures reveal the molecular basis of
ligand-dependent regulation of NadR,
the transcriptional repressor of the meningococcal
antigen NadA**

Abstract

Neisseria adhesin A (NadA) is present on the meningococcal surface and contributes to adhesion to and invasion of human cells. NadA is also one of three recombinant antigens in the *Bexsero* vaccine, approved in 2012 by the European Medicines Agency (EMA), which protects against serogroup B meningococcus. The amount of NadA on the bacterial surface influences the antibody-mediated serum bactericidal response measured *in vitro*. It is therefore important to understand the mechanisms which regulate *nadA* expression levels, which are predominantly controlled by the transcriptional regulator NadR (*Neisseria* adhesin A Regulator) both *in vitro* and *in vivo*, otherwise the real contribution of NadA to vaccine-induced protection against meningococcal meningitis may be underestimated. NadR binds the *nadA* promoter and represses gene transcription. In the presence of 4-hydroxyphenylacetate (4-HPA), a catabolite present in human saliva both under physiological conditions and during bacterial infection, the binding of NadR to the *nadA* promoter is attenuated and *nadA* expression is induced. NadR also mediates ligand-dependent regulation of many other meningococcal genes, for example the highly-conserved multiple adhesin family (*maf*) genes, which encode proteins emerging with important roles in host-pathogen interactions, immune evasion and niche adaptation. To gain insights into the regulation of NadR mediated by 4-HPA, the work presented here combined X-ray crystallographic, biochemical, and mutagenesis studies. In particular, two new crystal structures of ligand-free and ligand-bound NadR revealed (i) the molecular basis of 'conformational selection' by which one molecule of 4-HPA binds and stabilizes dimeric NadR in a conformation apparently unsuitable for DNA-binding, (ii) molecular explanations for the binding specificities of different hydroxyphenylacetate ligands, including 3Cl,4-HPA which is produced during inflammation, (iii) the presence of a leucine residue essential for dimerization and conserved in many MarR family proteins, and (iv) four residues (His7, Ser9, Asn11 and Phe25), which are involved in binding 4-HPA, and were confirmed *in vitro* to have key roles in the regulatory mechanism in bacteria. Overall, this study deepens our molecular understanding of the sophisticated regulatory mechanisms of the expression of *nadA* and other genes governed by NadR, dependent on interactions with niche-specific signal molecules that may play important roles during meningococcal pathogenesis.

The Neisserial adhesin Regulator (NadR)

Previous studies revealed that *nadA* expression levels are mainly regulated by the *Neisseria* adhesin A Regulator (NadR) [41]. Although additional factors influence *nadA* expression, the attention was focused on the regulation by NadR, the major mediator of *nadA* phase variable expression [62, 63]. Studies of NadR also have broader implications, since a genome-wide analysis of MenB wild-type and *nadR* knock-out strains revealed that NadR influences the regulation of >30 genes, including *maf* genes, from the multiple adhesin family [64]. These genes encode a wide variety of proteins connected to many biological processes contributing to bacterial survival, adaptation in the host niche, colonization and invasion [65, 66]. NadR binds the *nadA* promoter and represses gene transcription [63]. NadR binds *nadA* on three different operators (Opl, OplI and OplII) [64]. The DNA-binding activity of NadR is attenuated *in vitro* upon addition of various hydroxyphenylacetate (HPA) derivatives, including 4-HPA. 4-HPA is a small molecule derived from mammalian aromatic amino acid catabolism and released in human saliva, where it has been detected at micromolar concentration [67]. In the presence of 4-HPA, NadR is unable to bind the *nadA* promoter and *nadA* gene expression is induced [63, 64]. *In vivo*, the presence of 4-HPA in the host niche of *N. meningitidis* serves as an inducer of NadA production, thereby promoting bacterial adhesion to host cells [64]. Further, it was recently reported that 3Cl,4-HPA, produced during inflammation, is another inducer of *nadA* expression [68]. However, the molecular mechanism explaining how this transcriptional regulator interacts with 4-HPA or its derivatives and modulates their DNA-binding affinities accordingly has remained unresolved. The structural analysis of NadR was attempted in order to illustrate precisely how this protein recognizes and binds 4-HPA and to provide the structural basis for the attenuated DNA binding of NadR upon its interaction with 4-HPA.

The MarR family of transcriptional regulators

NadR belongs to the MarR (Multiple Antibiotic Resistance Regulator) family, a group of ligand-responsive transcriptional regulators ubiquitous in bacteria and archaea. MarR family proteins can promote survival in the presence of antibiotics, toxic chemicals, organic solvents or reactive oxygen species [69, 70] and can regulate virulence factor expression [71]. MarR homologues can act either as transcriptional repressors or as activators [72]. To date, >50 MarR family structures are known, revealing a conserved fold of six α -helices (H) and a two-stranded antiparallel β -sheet (B) in the topology: H1-H2-H3-H4-B1-B2-H5-H6. The DNA-binding domains are ascribed to the superfamily of winged helix proteins, containing α -helices 3 and 4, comprise the helix-turn-helix motif, and the β -sheet is called the wing. Helix 4 is termed the recognition helix, as in other HTHs where it binds the DNA major groove. The

α -helices 1, 5 and 6 are involved in dimerisation, as most MarR-like transcription regulators form dimers. Further, a few examples have been obtained in complexes with target DNA ligands. For example, the structure of the *Bacillus subtilis* OhrR-*ohrA* complex revealed the chimeric nature of the wHTH motif and a double-helix DNA binding element, both of which are proposed to be utilized by the entire MarR family to bind cognate DNA [73]. A molecular understanding of their ligand-dependent regulatory mechanisms is still limited, often hampered by lack of identification of their ligands. A potentially interesting exception comes from the ligand-free and salicylate-bound forms of the *Methanobacterium thermoautotrophicum* protein MTH313 which revealed that two salicylate molecules bind to one MTH313 dimer and induce large conformational changes, apparently sufficient to prevent DNA binding [74]. However, the homologous archeal *Sulfolobus tokodaii* protein ST1710 presented essentially the same structure in ligand-free and salicylate-bound forms, apparently contrasting the mechanism proposed for MTH313 [75]. Despite these apparent differences, MTH313 and ST1710 bind salicylate in approximately the same site, between their dimerization and DNA-binding domains. However, it is unknown whether salicylate is a relevant *in vivo* ligand of either of these two proteins, which share ~20% sequence identity with NadR, rendering unclear the interpretation of these findings regarding the regulatory mechanisms of NadR or other MarR family proteins [72]. Other two MarR family homologues TcaR and SAR2349 from *Staphylococcus epidermidis* and *Staphylococcus aureus*, respectively, have been crystallized in the presence of salicylate and antibiotics. In the structure of TcaR complexed with salicylate, multiple binding site were found, one of which (SAL-1) overlaps with the binding site seen in MTH313 [76]. The structures of SAR2349–antibiotic complexes reveals that the binding of antibiotics change the angle between the dimerization domains, inducing conformational changes within the wHTH motifs that interferes with binding to DNA [77].

Experimental procedures

Bacterial strains, culture conditions and mutant generation. In this study *N. meningitidis* MC58 wild type strain and related mutant derivatives were used. The MC58 isolate was kindly provided by Professor E. Richard Moxon, University of Oxford, UK, and was previously submitted to the Meningococcal Reference Laboratory, Manchester, UK [78]. Strains were routinely cultured, stocked, and transformed as previously described [64]. The preparation of the expression construct enabling production of soluble NadR (Uniprot code Q7DD70) with an N-terminal His-tag followed by a thrombin cleavage site and NadR residues M1-S146 was described previously [79]. Site-directed mutagenesis was performed using two 2 couples of mutagenic primers containing the desired mutation to amplify pET15b containing several NMB1843 variants. In short, 1 to 10 ng of plasmid template were amplified using Kapa HiFi DNA polymerase (Kapa Biosystems) and the following cycling conditions: 98°C for 5 min, 15 amplification cycles (of 98°C for 30 s, 60°C for 30 s, 72°C for 6 min) followed by a final extension of 10 min at 72°C. Residual template DNA was digested by 30 min incubation with FastDigest *DpnI* (Thermo Scientific) at 37°C and 1 µl of this reaction was used for transforming competent *E. coli* DH5α.

Strains or plasmid	Relevant characteristics
<i>E. coli</i> strains	
DH5α	<i>supE44 lacU169</i> (w80lacZDM15) <i>hsdR17 recA1 endA1 gyrA96 thi-1 relA1</i>
Plasmids	
pET15b-1843	pET15b derivative for expression of recombinant NMB1843, AmpR
pET15b-PDD0	pET15b derivative for expression of recombinant NMB1843 containing an H7A mutation, AmpR
pET15b-PDD1	pET15b derivative for expression of recombinant NMB1843 containing an S9A mutation, AmpR
pET15b-PDD2	pET15b derivative for expression of recombinant NMB1843 containing an N11A mutation, AmpR
pET15b-PDD3	pET15b derivative for expression of recombinant NMB1843 containing an Y115A mutation, AmpR
pET15b-PDD4	pET15b derivative for expression of recombinant NMB1843 containing an K126A mutation, AmpR
pET15b-PDD5	pET15b derivative for expression of recombinant NMB1843 containing L130K and L133K mutations, AmpR
pET15b-PDD6	pET15b derivative for expression of recombinant NMB1843 containing K126A, L130K and L133K mutations, AmpR
pET15b-PDD7	pET15b derivative for expression of recombinant NMB1843 containing N11A, D112A, R114A and Y115A mutations, AmpR
pET15b-PDD8	pET15b derivative for expression of recombinant NMB1843 containing an L130K mutation, AmpR
pET15b-PDD9	pET15b derivative for expression of recombinant NMB1843 containing an L133K mutation, AmpR

Protein production and purification. The NadR expression constructs (wild-type or mutant clones) were transformed into *E. coli* BL21 (DE3) cells and were grown as 500mL culture volumes in 2L shake flasks at 37°C in Luria-Bertani (LB) medium supplemented with 100µg/mL ampicillin, until an OD₆₀₀ of 0.5 was reached. Target protein production was induced by the addition of 1mM IPTG followed by incubation with shaking overnight at 21°C. (For production of the SeMet derivative form of NadR, essentially the same procedure was followed, but using the *E. coli* B834 strain grown in a modified M9 minimal medium supplemented with 40mg/L L-selenomethionine). Cells were harvested by centrifugation (6400g, 30 min, 4°C), resuspended in 20mM HEPES pH 8.0, 300mM NaCl, 20mM imidazole, and were lysed by sonication (Qsonica Q700). Cell lysates were clarified by centrifugation at 2800g for 30 min, and the supernatant was filtered using a 0.22µm membrane (Corning filter system) prior to protein purification.

NadR was purified by affinity chromatography using an AKTA purifier (GE Healthcare). All steps were performed at room temperature (18-26°C), unless stated otherwise. The filtered supernatant was loaded onto an Ni-NTA resin (5mL column, GE Healthcare), and NadR was eluted using 4 steps of imidazole at 20, 30, 50 and 250mM concentration, at a flow rate of 5mL/min. Eluted fractions were examined by reducing and denaturing SDS-PAGE analysis. Fractions containing NadR were identified by a band migrating at ~17kDa, and were pooled. The N-terminal 6-His tag was removed enzymatically using the Thrombin CleanCleave Kit (Sigma-Aldrich). Subsequently, the sample was reloaded on the Ni-NTA resin to capture the free His tag (or unprocessed tagged protein), thus allowing elution in the column flow-through of tagless NadR protein, which was used in all subsequent studies. The NadR sample was concentrated and loaded onto a HiLoad Superdex 75 (16/60) preparative size-exclusion chromatography (SEC) column equilibrated in buffer containing 20mM HEPES pH 8.0, 150mM NaCl, at a flow-rate of 1mL/min. NadR protein was collected and the final yield of purified protein obtained from 0.5L growth medium was approximately 8mg (~2mg protein per g wet biomass). Samples were used immediately for crystallization or analytical experiments, or were frozen for storage at -20°C.

Size-exclusion high-performance liquid chromatography (SE-HPLC) coupled with Multi-angle laser light scattering (MALLS). SE-HPLC was used to assess the purity and the apparent molecular weight of the recombinant wild-type NadR sample alone or containing a 200-fold molar excess of 4-HPA and of the mutated NadR samples. SE-HPLC experiments were performed by loading 20µl of each sample at a concentration of ~ 50µM on an analytical size exclusion TSK Super SW3000 column of 4 µm particle size and 250Å pore size (Tosoh), with a separation range suitable for globular proteins of 10 to 500 kDa. Samples were eluted isocratically in 0.1M NaH₂PO₄, 0.4M (NH₄)₂SO₄ buffer at pH 6.0, experiments were performed at room temperature (18-26°C).

MALLS analyses were performed in order to determine the absolute molecular mass of NadR alone or in the presence of 4-HPA. MALLS analyses were performed online with SE-HPLC, using a Dawn TREOS MALLS detector (Wyatt Corp., Santa Barbara, CA, USA) and an incident laser wavelength of 658 nm. The intensity of the scattered light was measured at 3 angles simultaneously. Data elaboration was performed using the Astra V software (Wyatt) to determine the weighted-average absolute molecular mass (MW), the polydispersity index (MW/Mn) and homogeneity (Mz/Mn) for each oligomer present in solution. Normalization of the MALLS detectors was performed in each analytical session by use of bovine serum albumin.

Differential Scanning Calorimetry (DSC). The thermal stability of NadR proteins was assessed by DSC using a MicroCal VP-Capillary DSC instrument (GE Healthcare). NadR samples were prepared at a protein concentration of 0.5mg/mL (~30 μ M) in buffer containing 20mM Hepes, 300mM NaCl, pH 7.4, with or without 6mM HPA or salicylate. The DSC temperature scan ranged from 10°C to 110°C, with a thermal ramping rate of 200°C per hour and a 4 second filter period. Data were analyzed by subtraction of the reference data for a sample containing buffer only, using the Origin 7 software. All experiments were performed in duplicate, and mean values of the melting temperature (T_m) were determined.

Surface plasmon resonance (SPR). Determination of equilibrium dissociation constant, K_D : Surface plasmon resonance binding analyses were performed using a Biacore T200 instrument (GE Healthcare) equilibrated at 25 °C. The ligand (NadR) was covalently immobilized by amine-coupling on a CM-5 sensor chip (GE Healthcare), using 20 μ g/mL purified protein in 10 mM sodium acetate buffer pH 5, injected at 10 μ l/min for 120 s until ~9000 response units (RU) were captured. A high level of ligand immobilization was required due to the small size of the analytes. An unmodified surface was used as the reference channel. Titrations with analytes (HPAs or salicylate) were performed with a flow-rate of 30 μ l/min, injecting the compounds in a concentration range of 10 μ M to 20 mM, using filtered running buffer containing Phosphate Buffered Saline (PBS) with 0.05 % Tween-20, pH 7.4. Following each injection, sensor chip surfaces were regenerated with a 30-second injection of 10 mM Glycine pH 2.5. Each titration series contained 20 analyte injections and was performed in triplicate. Titration experiments with long injection phases (> 15 mins) were used to enable steady-state analyses. Data were analyzed using the BIAcore T200 evaluation software and the steady-state affinity model. A buffer injection was subtracted from each curve, and reference sensorgrams were subtracted from experimental sensorgrams to yield curves representing specific binding. The equilibrium dissociation constant, K_D , was determined from the plot of RU_{eq} against analyte concentration, as described previously [80].

Determination of binding stoichiometry: From each plot of RU_{eq} against analyte concentration, obtained from triplicate experiments, the R_{max} value (maximum analyte binding capacity of the surface) was extrapolated from the experimental. Stoichiometry was calculated using the molecular weight of dimeric NadR as ligand molecule (MW_{ligand}) and the molecular weights of the HPA analyte molecules ($MW_{analyte}$), and the following equation:

$$\text{Stoichiometry} = \frac{R_{max} \times MW_{ligand}}{MW_{analyte} \times R_{ligand}}$$

where R_{ligand} is recorded directly from the sensorgram during ligand immobilization prior to the titration series, as described previously [81]). The stoichiometry derived therefore represented the number of HPA molecules bound to one dimeric NadR protein.

Crystallization of NadR in the presence or absence of 4-HPA. Purified NadR was concentrated to 2.7 mg/mL using a centrifugal concentration device (Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-10 membrane with cut-off size 10kDa; Millipore) running at 600 *g* in a bench top centrifuge (Thermo Scientific IEC CL40R) refrigerated at 2-8°C. More concentrated samples were found to induce precipitation of the protein. To prepare holo-NadR samples, HPA ligands were added at a 200-fold molar excess prior to the centrifugal concentration step, and this ratio of protein:ligand concentration was maintained. The concentrated holo- or apo-NadR was subjected to crystallization trials performed in 96-well low-profile Intelli-Plates (Art Robbins) or 96-well low-profile Greiner crystallization plates, using a nanodroplet sitting-drop vapour-diffusion format and mixing equal volumes (200nL) of protein samples and crystallization buffers using a Gryphon robot (Art Robbins). Crystallization trays were incubated at 20° C. Crystals of apo-NadR were obtained at 20°C in 50 % PEG 3350 and 0.13 M di-Ammonium hydrogen citrate, whereas crystals of SeMet–NadR in complex with 4-HPA grew at 20°C in condition H4 of the Morpheus screen (Molecular Dimensions), which contains 37.5% of the pre-mixed precipitant stock MPD_P1K_PEG 3350, buffer system 1 and 0.1 M amino acids, at a pH 6.5. All crystals were mounted in cryo-loops using 10% ethylene glycol or 10% glycerol as cryo-protectant before cooling to 100 K for data collection.

X-ray diffraction data collection and structure determination. X-ray diffraction data from crystals of apo-NadR and SeMet–NadR/4-HPA were collected on beamline PXII-X10SA of the Swiss Light Source (SLS) at the Paul Scherrer Institut (PSI), Villigen, Switzerland. All diffraction data were processed in-house with *XDS* [82] and programs from the CCP4 suite [83]. Crystals of apo-NadR and 4-HPA-bound SeMet-NadR belonged to

space group $P 43 21 2$ (see Table 2). Apo-NadR crystals contained four protein molecules in the asymmetric unit (Matthews coefficient $2.25 \text{ \AA}^3 \cdot \text{Da}^{-1}$, for a solvent content of 45 %), while crystals of SeMet–NadR/4-HPA contained two protein molecules in the asymmetric unit (Matthews coefficient $1.98 \text{ \AA}^3 \cdot \text{Da}^{-1}$, for a solvent content of 38 %). In solving the holo-NadR structure, an initial and marginal molecular replacement (MR) solution was obtained using as template search model the crystal structure of the transcriptional regulator PA4135 (PDB entry 2FBI), the closest structurally-characterized homologue, with which NadR shares ~54% sequence identity. This solution was combined with SAD data to aid identification of two selenium sites in NadR, using *autosol* in *phenix* [84] and this allowed generation of high-quality electron density maps that were used to build and refine the structure of the complex. Electron densities were clearly observed for almost the entire dimeric holo-NadR protein, except for residues 88-90 of chain B, which lie in an exposed region of the winged-helix motif often found to be disordered in MarR family structures. The crystal structure of apo-NadR was subsequently solved by MR in *Phaser* [85] at 2.7 \AA , using the final refined model of SeMet-NadR/4-HPA as the search model. For apo-NadR, electron densities were clearly observed for almost the entire protein, although residues 84-91 of chains A, C, and D, and residues 84-90 of chain B lacked densities suggesting local disorder or flexibility. Both structures were refined and rebuilt using *phenix* [84] and *Coot* [86], and structural validation was performed using Molprobity [87]. Data collection and refinement statistics are reported in Table 2. Atomic coordinates of the two NadR structures have been deposited in the Protein Data Bank, with entry codes 5aip (NadR bound to 4-HPA) and 5aiq (apo-NadR). All crystallographic software was remotely compiled, installed and maintained by SBGrid [88].

Results

NadR is dimeric and is stabilized by specific hydroxyphenylacetate ligands.

Recombinant NadR was produced in *E. coli* using an expression construct prepared from the *nadR* gene of the *N. meningitidis* serogroup B strain MC58. Standard chromatographic techniques were used to purify NadR (see *Materials and Methods*). In analytical size-exclusion high-performance liquid chromatography (SE-HPLC) experiments coupled with multi-angle laser light scattering (MALLS) analyses, NadR presented a single species $\geq 97\%$ pure with an absolute molecular mass of 35 kDa (*Figure 1.1*)

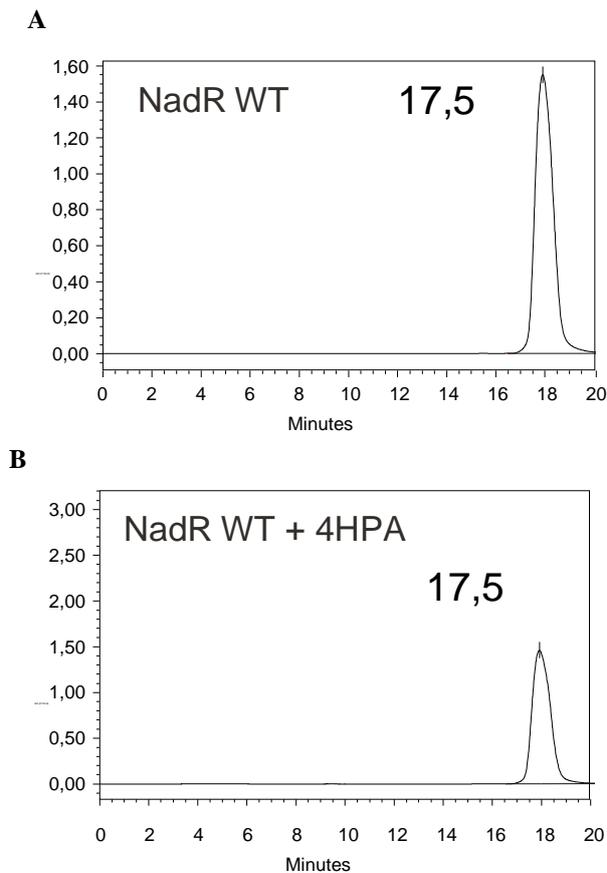


Figure 1.1. Size-exclusion high-performance liquid chromatography profile. SE-HPLC was used to assess the purity and the apparent molecular weight of the recombinant wild-type (WT) NadR sample alone (panel A) or containing a 200-fold molar excess of 4-HPA (panel B). The NadR monomer concentration was approximately $50\mu\text{M}$, and the 4-HPA concentration was approximately 10mM . Data are plotted as Absorbance Units (mAU) at 280nm wavelength, against retention time in minutes. The elution time for the peak at maximum absorbance is indicated in each panel. Analysis by integration of the peaks (peak boundaries were defined as indicated by the red triangles), revealed that the NadR sample was $\geq 97\%$ pure. In both cases, the protein eluted with a retention time (~ 22.5 minutes) indicative of a dimer ($\sim 35\text{kDa}$), determined by calibration of the column using standard molecular weight markers (Bio-rad, cat. no. 151-1901). Notably, the retention time was not significantly changed by the presence of the ligand.

These data showed that NadR was dimeric in solution, since the theoretical molecular mass of the NadR dimer is 33.73 kDa. Subsequently, SE-HPLC/MALLS analyses of NadR

were performed in the presence 4-HPA, revealing that there were no changes in oligomeric state upon addition of the ligand (*Figure 1.2*).

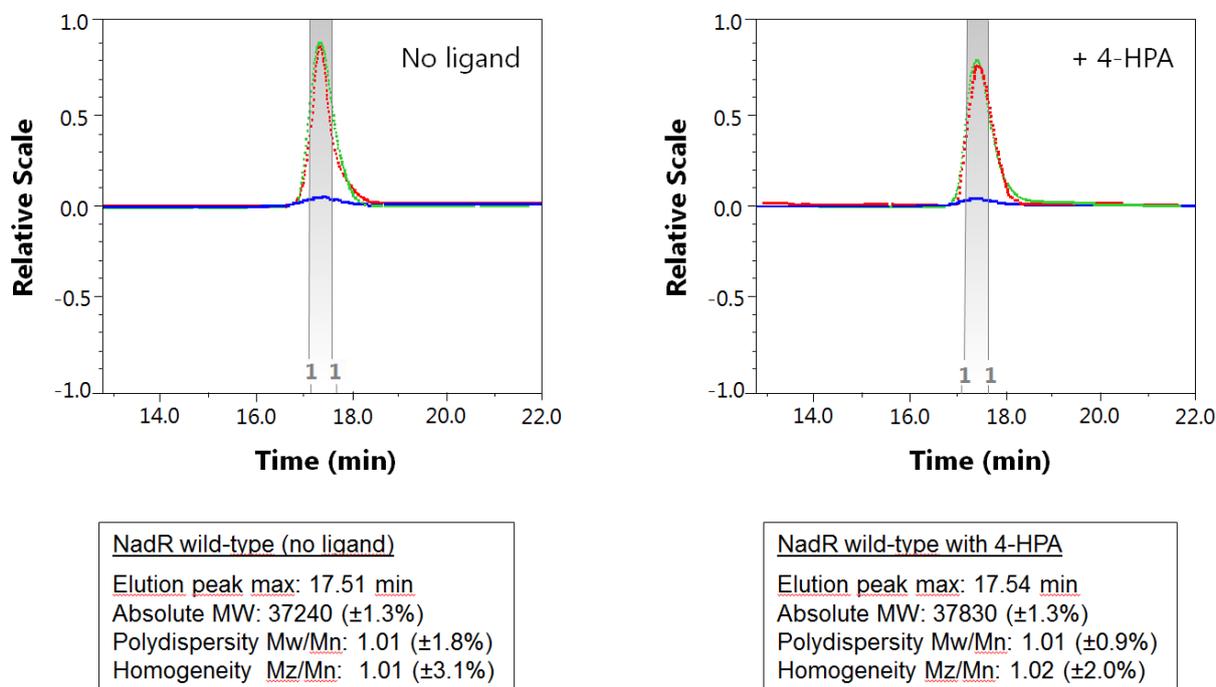


Figure 1.2. Multi-angle laser light scattering. MALLS analyses were performed in order to determine the absolute molecular mass of NadR alone (panel A) or in the presence of 4-HPA (panel B). The curves plotted correspond to Absorbance Units (mAU) at 280nm wavelength (green), light scattering (red), and refractive index (blue). The elution peak maxima were at 17.5 minutes and the numerical data obtained for absolute molecular mass and polydispersity are shown below each image. In both cases, the MALLS data clearly indicated a single monodisperse species of absolute molecular mass ~ 37.5 kDa, corresponding to the dimeric form of NadR. (The numbers '1' at the bottom of the gradient-shaded slice identify the beginning and end of each fraction-1, used for the MALLS analyses)

The thermal stability of NadR was examined using differential scanning calorimetry (DSC). Since ligand-binding can increase protein stability [89], it was also investigated the effect of various HPAs on the melting temperature (T_m) of NadR. As a control of specificity it was tested a salicylate, a ligand of MarR proteins reported to increase the T_m of ST1710 and MTH313 by approximately 3°C and 9°C , respectively [74]. In the absence of ligand, the T_m of apo-NadR was $67.3 \pm 0.1^\circ\text{C}$. An increased thermal stability was induced by 4-HPA ($\Delta T_m \sim 3^\circ\text{C}$) and, to a lesser extent, by 3-HPA ($\Delta T_m \sim 2^\circ\text{C}$) (*Figure 1.3*). Interestingly, NadR displayed the greatest increase in thermal stability upon addition of 3Cl,4-HPA ($\Delta T_m \sim 4^\circ\text{C}$) and was unaffected by salicylate.

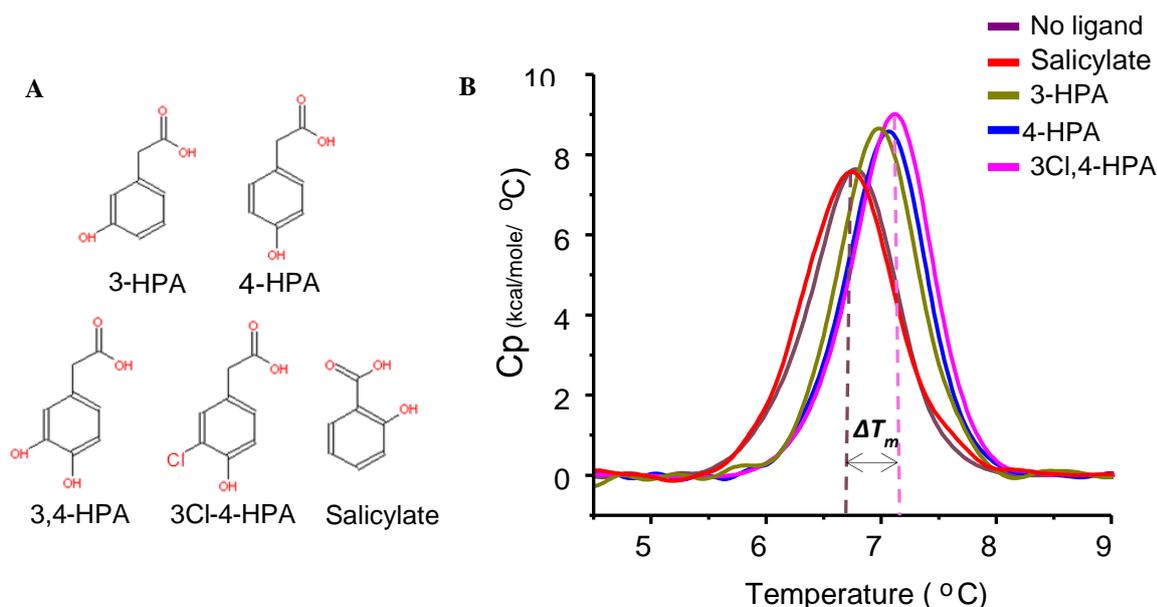


Figure 1.3. Binding and thermostabilization of NadR by small molecule ligands. (A) Molecular structures of salicylate and the hydroxyphenylacetates tested. (B) DSC profiles are colored as follows: apo-NadR (pink), NadR+salicylate (blue), NadR+3-HPA (orange), NadR+4-HPA (green), NadR+3Cl,4-HPA (brown).

NadR displays distinct binding affinities for hydroxyphenylacetate ligands

To further investigate the binding of HPAs to NadR, surface plasmon resonance (SPR) was used. The SPR sensorgrams displayed very fast association and dissociation events typical of small molecule ligands, thus prohibiting a detailed kinetic study. However, steady-state SPR analyses of the NadR-HPA interactions readily allowed determination of the equilibrium dissociation constants (K_D). K_D values of interaction of ligands with NadR are reported in *Table 1.1*, showing that 3Cl,4-HPA was the tightest binder, and thus matched the ranking of ligand-induced T_m increases observed in the DSC experiments. Although these K_D values indicate relatively weak interactions, they are similar to the values determined for the MarR/salicylate interaction ($K_D \sim 1\text{mM}$) [90] and the MTH313/salicylate interaction (K_D 2-3mM) [74], and are approximately 20-fold tighter than the ST1710/salicylate interaction ($K_D \sim 20\text{mM}$) [75].

Ligand	ΔT_m (°C)	K_D (mM)
Salicylate	0	-
3-HPA	2.7	2.7 ± 0.1
4-HPA	3.3	1.5 ± 0.1
3Cl,4-HPA	3.9	1.1 ± 0.1

Table 1.1. Thermal stabilization (ΔT_m) and dissociation constants (K_D) of the NadR/ligand interactions.

Crystal structures of holo-NadR and apo-NadR

To fully characterize the NadR/HPA interactions, the crystal structures of ligand-bound (holo) and ligand-free (apo) NadR was determined. First, NadR was crystallized (a selenomethionine (SeMet)-labelled form) in the presence of a 200-fold molar excess of 4-HPA. The structure of the NadR/4-HPA complex was determined at 2.3 Å resolution using a combination of the single-wavelength anomalous dispersion (SAD) and molecular replacement (MR) methods, and was refined to $R_{\text{work}}/R_{\text{free}}$ values of 20.9/26.0 % (Table 1.2).

	NadR SeMet + 4-HPA (SAD peak) (PDB code 5aip)	NadR apo-form (PDB code 5aiq)
Data collection		
Wavelength (Å)	0.9792	1.0
Beamline	SLS (PXII-X10SA)	SLS (PXII-X10SA)
Resolution range (Å)	39.2 - 2.3	48.2 - 2.7
Space group	P 43 21 2	P 43 21 2
Unit cell	75.3, 75.3, 91.8	69.4, 69.4, 253.8
Total reflections	291132 (41090)	225521 (35809)
Unique reflections	12320 (1773)	17700 (2780)
Multiplicity	23.6 (23.2)	12.7 (12.8)
Completeness (%)	100.0 (100.00)	99.9 (99.7)
Mean I/sigma(I)	25.5 (9.0)	22.6 (3.8)
Wilson B-factor	23.9	49.1
R_{sym}^*	10.9 (39.4)	11.4 (77.6)
R_{meas}^{**}	11.3	11.8
Refinement		
$R_{\text{work}}^{\#}$	20.9	21.7
$R_{\text{free}}^{\#\#}$	26.0	27.2
Number of atoms		
Non-hydrogen atoms	2263	4163
Macromolecules	2207	4144
Ligands	11	0
Water	45	19
Protein residues	275	521
RMS(bonds)	0.008	0.003
RMS(angles)	1.09	0.823
Ramachandran (%)[§]		
Favored	100	98.4
Outliers	0	0
Clashscore	5.0	3.9
Average B-factor		
Macromolecules	34.8	53.3
Ligands	32.9	-
Solvent	37.3 (H ₂ O)	29.0 (H ₂ O)

Statistics for the highest-resolution shell are shown in parentheses.

$$* R_{\text{sym}} = \frac{\sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle|}{\sum_{\text{hkl}} \sum_i I_i(\text{hkl})}$$

** R_{meas} = redundancy-independent (multiplicity-weighted) R_{merge} as reported from AIMLESS [91].

$$\# R_{\text{work}} = \frac{\sum ||F_{(\text{obs})}| - |F_{(\text{calc})}||}{\sum |F_{(\text{obs})}|}$$

R_{free} = as for R_{work} , calculated for 5.0% of the total reflections, chosen at random, and omitted from refinement.

§ Figures from Molprobity [87].

Table 1.2. Data collection and refinement statistics for NadR structures.

X-ray crystallography was selected as the method-of-choice, due to its well-known capacity to provide high-resolution information about protein-small molecule interactions.

NMR spectroscopy was a possible alternative structural technique, but since a number of MarR family proteins had been previously crystallized, it was considered likely that the crystallographic approach would have a reasonable probability-of-success. In contrast, the NadR protein dimer was considered too small to be tractable by the recently-emerging electron cryomicroscopy techniques, which are better suited for larger macromolecules [92]

Despite numerous attempts, it was not possible to obtain high-quality crystals of NadR complexed with 3Cl,4-HPA, 3,4-HPA, 3-HPA or DNA targets. However, it was possible to crystallize apo-NadR, and the structure was determined at 2.7 Å resolution by MR using the NadR/4-HPA complex as the search model. The apo-NadR structure was refined to $R_{\text{work}}/R_{\text{free}}$ values of 19.1/26.8 % (Table 1.2).

The asymmetric unit of the NadR/4-HPA crystals (holo-NadR) contained one NadR homodimer, while the apo-NadR crystals contained two homodimers. In the apo-NadR crystals, the two homodimers are related by a rotation of $\sim 90^\circ$; the observed association of the two dimers was presumably an effect of crystal packing, since the interface between the two homodimers is small ($< 550 \text{ \AA}^2$ of buried surface area), and is not predicted to be physiologically relevant by the PISA software [93]. Moreover, our SE-HPLC/MALLS analyses revealed that in solution NadR is dimeric, and previous studies using native mass spectrometry (MS) also revealed dimers and not tetramers [94].

The holo-NadR homodimer shows a dimerization interface mostly involving the top of its triangular form, while the two DNA-binding domains are located at the base. The overall structure of NadR shows triangular dimensions of $\sim 50 \times 65 \times 50 \text{ \AA}$ and a large homodimer interface burying a total surface area $\sim 4800 \text{ \AA}^2$ (Figure 1.4).

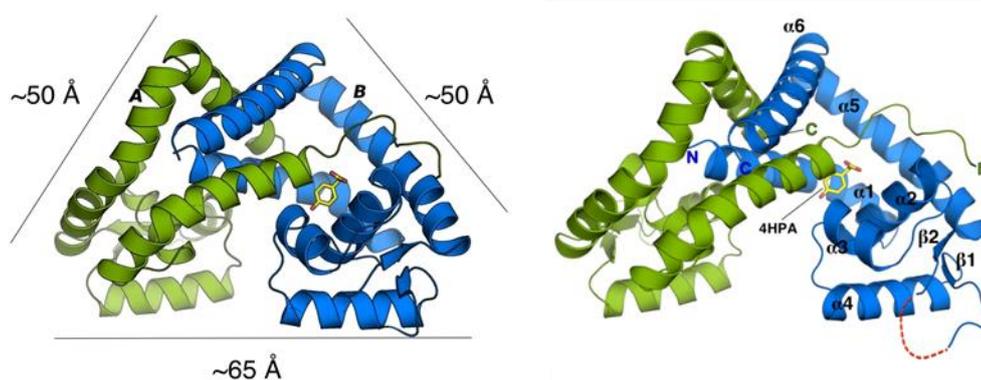


Figure 1.4. The crystal structure of NadR in complex with 4-HPA. (A) The holo NadR homodimer is depicted in green and blue for chains A and B respectively, while yellow sticks depict the 4-HPA ligand. Secondary structures are labelled for chain B only. (B) Orientation as in panel A, showing the secondary structure elements of NadR protein. Red dashes in panels A and B show hypothetical positions of chain B residues 88-90 that were not modeled due to lack of electron density.

Each NadR monomer consists of six α -helices and two short β -strands, with helices α 1, α 5, and α 6 forming the dimerization interface. Helices α 3 and α 4 form a helix-turn-helix motif, followed by the “wing motif” comprised of two short antiparallel β -strands (β 1- β 2). These secondary structure elements constitute the winged helix-turn-helix (wHTH) DNA-binding domain and, together with the dimeric organization, are the hallmarks of MarR family structures [72].

The holo-NadR structure presents only one occupied ligand-binding pocket

As already shown in *Figure 1.4*, the NadR/4-HPA structure revealed the ligand-binding site nestled between the dimerization and DNA-binding domains. High-quality electron density maps allowed clear identification of the bound 4-HPA ligand, which showed a different position and orientation compared to salicylate complexed with MTH313 and ST1710 [74, 75] (see Discussion). The binding pocket was almost entirely filled by 4-HPA and one water molecule, although there also remained a small tunnel 2-4Å in diameter and 5-6Å long leading from the pocket (proximal to the 4-hydroxyl position) to the protein surface. The tunnel was lined with rather hydrophobic amino acids, and did not contain water molecules. Most unexpectedly, only one monomer of the NadR homodimer contained 4-HPA in the binding pocket, whereas the corresponding pocket position of the other monomer was unoccupied by ligand.

Inspection of the protein-ligand interaction network revealed no bonds from NadR backbone NH or CO groups to the ligand, but several key side chain mediated hydrogen (H)-bonds and ionic interactions, most notably between the carboxylate group of 4-HPA and Ser9 (chain A), and Trp39, Arg43 and Tyr115 of chain B (*Figure 1.5A*). At the other end of the ligand, the 4-hydroxyl group was proximal to H-bond donors in the side chains of Asn11 (chain A) and Asp36 (chain B), although these were positioned at slightly greater distances (3.6-4.3 Å) than those atoms contacting the carboxylate group. There was also one water molecule observed in the pocket, bound by the carboxylate group and the side chains of Ser9 and Asn11 from chain A.

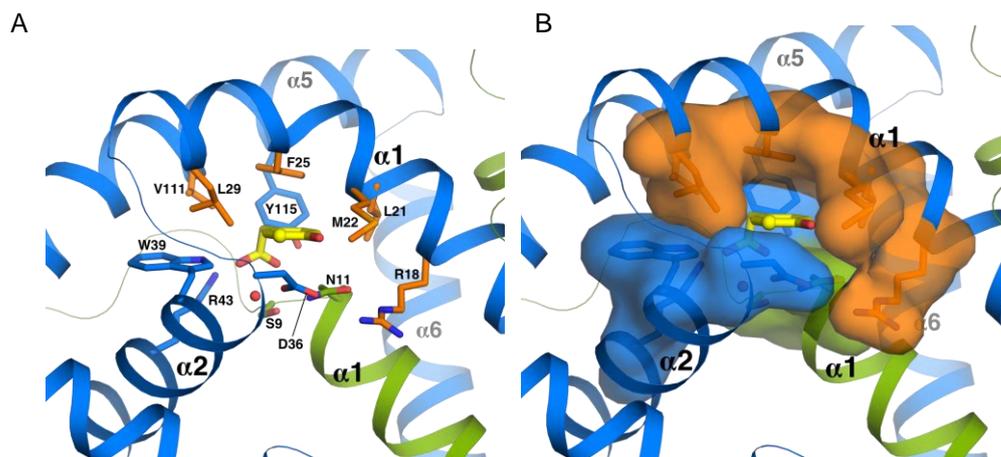


Figure 1.5. Atomic details of the NadR/4-HPA interaction. (A) View of the binding pocket showing side chain sticks for all interactions between NadR and 4-HPA. Green and blue ribbons depict NadR chains A and B, respectively. 4-HPA is shown in yellow sticks, with oxygen atoms in red. A water molecule is shown by the red sphere. A list of the interacting atoms and bond distances is provided in Table 3. Side chains mediating hydrophobic interactions are shown in orange. The yellow sphere on the 4-HPA phenyl ring shows the 3-position at which the chloro group of 3Cl,4-HPA could be readily accommodated. B) 4-HPA is sandwiched by NadR, as shown by the surface representation of residues that line the binding pocket. ‘Ceiling’ residues are colored orange, ‘floor’ residues are colored blue (chain B) or green (chain A).

In addition to the H-bonds involving the carboxylate and hydroxyl groups of 4-HPA, binding of the phenyl moiety appeared to be stabilized by van der Waals’ interactions involving the hydrophobic side chain atoms of Arg18 (via the C β , C γ , C δ methylene groups), Leu21, Met22, Phe25, Leu29 and Val111 of chain B (*Figure 1.5A*). In particular, the phenyl ring of Phe25 was positioned parallel to the phenyl ring of 4-HPA, potentially forming π - π parallel-displaced stacking interactions. Interestingly, NadR residues in the 4-HPA binding pocket effectively created a polar ‘floor’ and a hydrophobic ‘ceiling’, which house the ligand (*Figure 1.5B*). The polar floor is made of residues both from chain A and chain B of the homodimer, while the ceiling is made of residues from chain B only (*Figure 1.5B*). Collectively, this mixed network of polar and hydrophobic interactions endows NadR with a strong recognition pattern for HPAs, with additional medium-range interactions potentially established to the hydroxyl group at the 4-position.

Analysis of the pockets reveals the molecular basis for asymmetry and stoichiometry

The lack of a second 4-HPA molecule in the homodimer suggested negative cooperativity, a phenomenon previously described for the MTH313/salicylate interaction [74] and for other MarR family proteins [72]. To understand the molecular basis of asymmetry in NadR, the ligand-free monomer (chain A) was superposed onto the ligand-occupied monomer (chain B). Overall, the superposition revealed a high degree of structural similarity

(C α root mean square deviation (rmsd) of 1.5Å), though on closer inspection a rotational difference of ~9 degrees of helix α 6 was observed, suggesting that 4-HPA induced a slight conformational change (*Figure 1.6A*)

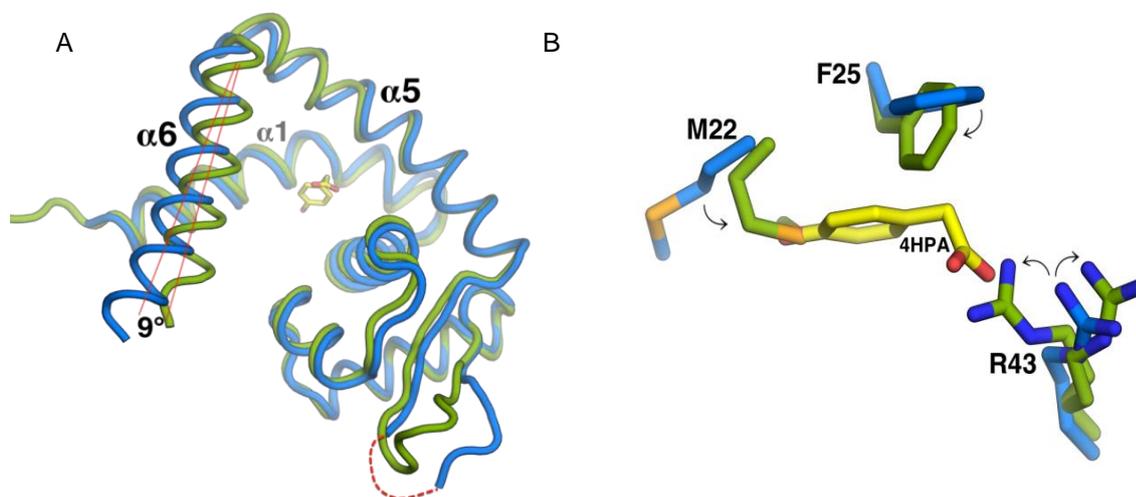


Figure 1.6. Structural differences between the two monomers of holo-NadR. (A) Aligned monomers of the holo-NadR structure, revealing that the major overall difference is the ~9 degree shift in the position of helix α 6 (chain A: green; chain B: blue). (B) A closer comparison of the binding pockets shows that in the ligand-free monomer chain A (green) residues M22, F25 and R43 adopt 'inward' positions (highlighted by arrows) that would prevent binding of 4-HPA due to clashes with the 4-hydroxyl group, the phenyl ring and the carboxylate group, respectively. (Both panels (A) and (B) are rotated compared to *Figure 1.5*).

However, since residues of helix α 6 were not directly involved in ligand binding, an explanation for the lack of 4-HPA in monomer A did not emerge by analyzing only the backbone atom positions suggesting that a more complex series of allosteric events may occur. Indeed, it was noted interesting differences in the side chains of Met22, Phe25 and Arg43, which in monomer B are used to contact the ligand while in monomer A they partially occupied the pocket and collectively reduced its volume significantly. Specifically, upon analysis with the CASTp software [95], the pocket in chain B containing the 4-HPA exhibited a total volume of 368Å³, while the pocket in chain A was occupied by side chains and was divided into three much smaller pockets, each with volumes < 50Å³, evidently rendering chain A unfavorable for ligand binding. Most notably, atomic clashes between the ligand and the side chains of Met22, Phe25 and Arg43 (chain A) would occur if 4-HPA were present in the monomer A pocket (*Figure 1.6B*). Subsequently, analyses of the pockets in apo-NadR revealed that in the absence of ligand the long Arg43 side chain was always in the open 'outward' position compatible with binding to the 4-HPA carboxylate group. In contrast, the apo-form Met22 and Phe25 residues were still encroaching the spaces of the 4-hydroxyl group and the phenyl ring of the ligand, respectively (*Figure 1.6B*). The 'outward' position of

Arg43 generated an open apo-form pocket with volume approximately 380\AA^3 . Taken together, these observations suggest that Arg43 is a major determinant of ligand binding, and that its 'inward' position inhibits the binding of 4-HPA to the empty pocket of holo-NadR.

To support the crystallographic data, the binding stoichiometry was investigated using solution-based techniques. However, studies based on tryptophan fluorescence were confounded by the fluorescence of the HPA ligands, and isothermal titration calorimetry (ITC) was unfeasible due to the need for very high concentrations of NadR in the ITC chamber (due to the relatively low affinity), which exceeded the solubility limits of the protein. However, it was possible to calculate the binding stoichiometry of the NadR-HPA interactions using an SPR-based approach. In SPR, the signal measured is proportional to the total molecular mass proximal to the sensor surface; consequently, if the molecular weights of the interactors are known, then the stoichiometry of the resulting complex can be determined [81]. This approach relies on the assumption that the captured protein ('the ligand', according to SPR conventions) is 100 % active and freely-accessible to potential interactors ('the analytes'). This assumption is likely valid for this pair of interactors, for two main reasons. Firstly, NadR is expected to be covalently immobilized on the sensor chip as a dimer in random orientations, since it is a stable dimer in solution and has sixteen lysines well-distributed around its surface, all able to act as potential sites for amine coupling to the chip, and none of which are close to the ligand-binding pocket. Secondly, the HPA analytes are all very small (MW 150-170) and therefore are expected to be able to diffuse readily into all potential binding sites, irrespective of the random orientations of the immobilized NadR dimers on the chip. The stoichiometry of the NadR-HPA interactions was determined using Equation 1 (see Materials and Methods), and revealed stoichiometries of 1.13 for 4-HPA, 1.02 for 3-HPA, and 1.21 for 3Cl,4-HPA, strongly suggesting that one NadR dimer bound to 1 HPA analyte molecule.

Apo-NadR structures reveal conformational flexibility

After determination of the holo-NadR structure, the structure of apo-NadR was determined. The apo-NadR structure contained two homodimers in the asymmetric unit (chains A+B and chains C+D), which upon superposition revealed a few minor differences and an rmsd of 1.55\AA . Similarly, superpositions of the holo-homodimer onto each of the apo-homodimers resulted in rmsd values of 1.29\AA and 1.31\AA , again showing some slight overall differences between the homodimer pairs. The slightly larger difference between the two apo-homodimers, rather than between apo- and holo-homodimers, indicated that apo-NadR possesses a notable degree of conformational flexibility. The overall structural similarity but

with inherent plasticity of MarR proteins was observed previously upon comparison of the OhrR, MarR, MexR and SarR structures [96].

4-HPA stabilizes concerted conformational changes in NadR that prevent DNA-binding

To further investigate the conformational rearrangements of NadR, local structural alignments were performed using a subset of residues in the DNA-binding helix. By selecting and aligning residues Arg64-Ala77 of one $\alpha 4$ helix from each homodimer, superposition of the holo-homodimer onto the two apo-homodimers revealed differences in the monomer conformations of each structure (*Figure 1.7A*). While one monomer from each structure was closely superimposable (compare green and cyan cartoons, *Figure 1.7A*), the second monomer displayed quite large differences, especially in the DNA-binding helix $\alpha 4$ which shifted by as much as 6Å (*Figure 1.7B*). Accordingly, helix $\alpha 4$ was also found to be one of the most dynamic regions of NadR in previous HDX-MS analyses [94].

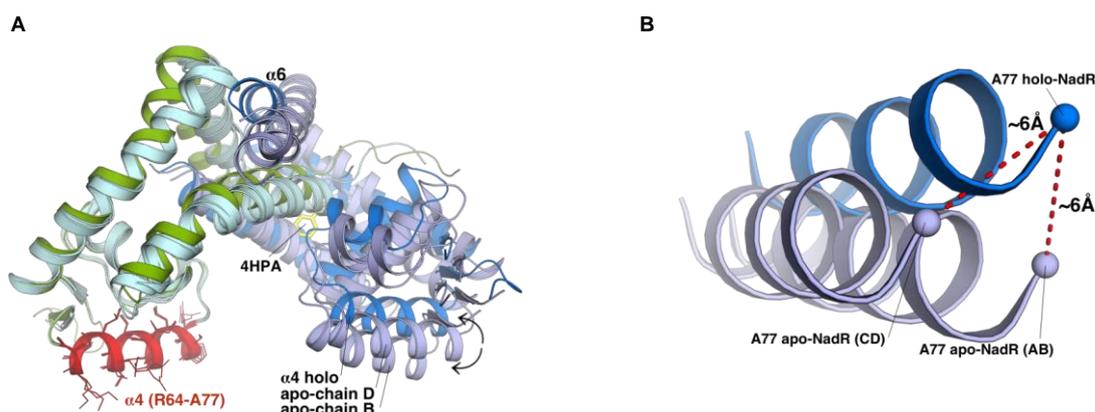


Figure 1.7. Structural comparison of holo- and apo-NadR and modelling of interactions with DNA. (A) The holo-homodimer structure is shown as green and blue cartoons, for chain A and B, respectively, while the two homodimers of apo-NadR are both cyan and pale blue for chains A and B, respectively. The three homodimers (chains AB holo, AB apo, and CD apo) were overlaid by structural alignment of all heavy atoms in residues R64-A77 (shown in red, with side chain sticks) of chains A holo, A apo, and C apo, belonging to helix $\alpha 4$ (left). The $\alpha 4$ helices aligned closely, C α rmsd 0.2Å for 14 residues. (B) The relative positions of the $\alpha 4$ helices of the 4-HPA-bound holo homodimer chain B (blue), and of apo homodimers AB and CD (showing chains B and D) in pale blue. Dashes indicate the Ala77 C α atoms, in the most highly shifted region of the ‘non-fixed’ $\alpha 4$ helix.

However, structural comparisons revealed that the shift of holo-NadR helix $\alpha 4$ induced by the presence of 4-HPA was also accompanied by several changes at the holo dimer interface, while such extensive structural differences were not observed in the apo dimer interfaces, particularly notable when comparing the $\alpha 6$ helices (*Figure 1.7A*). In summary, compared to ligand-stabilized holo-NadR, apo-NadR displayed an intrinsic

flexibility focused in the DNA-binding region. This was also evident in the greater disorder (i.e. less well-defined electron density) in the $\beta 1$ - $\beta 2$ loops of the apo dimers (density for 16 residues per dimer was missing) compared to the holo dimer (density for only 3 residues was missing).

In holo-NadR, the distance separating the two DNA-binding $\alpha 4$ helices was 32 Å, while in apo-NadR it was 29 Å for homodimer AB, and 34 Å for homodimer CD. Thus, the apo-homodimer AB presented the DNA-binding helices in a conformation similar to that observed in the protein:DNA complex OhrR:*ohrA* from *Bacillus subtilis* [96] (Figure 1.8A).

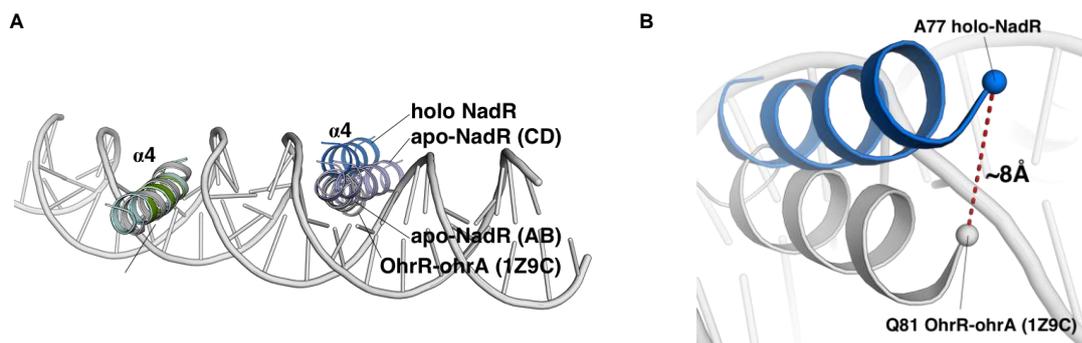


Figure 1.8. Structural comparison of holo- and apo-NadR and modelling of interactions with DNA. (A) The holo- and the apo- NadR homodimer structures are shown and superimposed as already reported in Figure 1.7. (A) The double-stranded DNA molecule (grey cartoon) from the OhrR-*ohrA* complex is shown after superposition with NadR, to highlight the expected positions of the NadR $\alpha 4$ helices in the B-DNA major grooves. The proteins share $\sim 30\%$ amino acid sequence identity. For clarity, only the $\alpha 4$ helices are shown in panels (A) and (B). (B) Upon comparison with the experimentally-determined OhrR:*ohrA* structure (grey), the $\alpha 4$ helix of holo-NadR (blue) is shifted $\sim 8\text{\AA}$ out of the major groove.

Interestingly, OhrR contacted *ohrA* across 22 base pairs (bp), and similarly the main NadR target sites identified in the *nadA* promoter (the operators Op I and Op II) were both shown to span 22 bp [63, 64]. Pairwise superpositions showed that the NadR apo-homodimer AB was the most similar to OhrR (rmsd 2.6Å), while the holo-homodimer was the most divergent (rmsd 3.3Å) (Figure 1.7A). Assuming the same overall DNA-binding mechanism is used by OhrR and NadR, the apo-homodimer AB was ideally pre-configured for DNA binding, while 4-HPA appeared to stabilize holo-NadR in a conformation poorly suited for DNA binding. When aligned with OhrR, the apo-homodimer CD presented another different intermediate conformation (rmsd 2.9Å), apparently not ideally pre-configured for DNA binding, but which in solution can presumably readily adopt the AB conformation due to the intrinsic flexibility described above. In addition to the different inter-helical translational distances, the $\alpha 4$ helices in the holo-NadR homodimer had also rotated, resulting in movement of $\alpha 4$ out of the major groove and preventing efficient DNA binding in the presence of 4-HPA (Figure 1.8B).

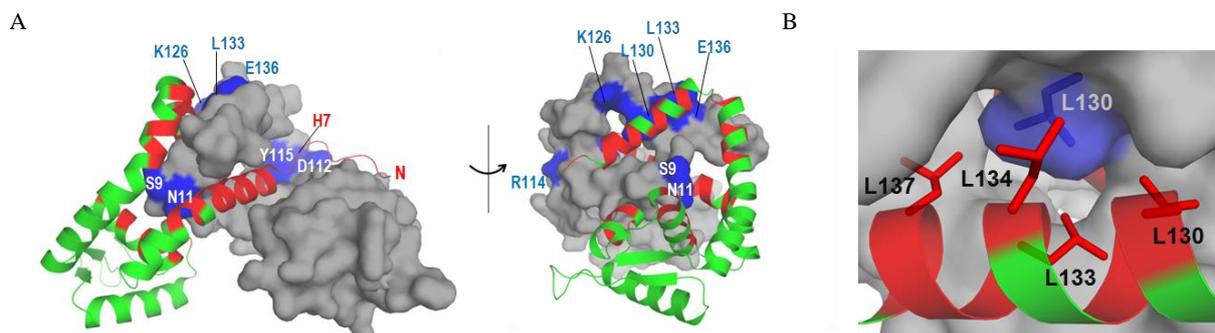
A single conserved leucine residue (L130) is crucial for dimerization

To study the architecture and stability of the NadR homodimer interface, a series of mutations were prepared, with the aim of disrupting the dimer interface. Due to the two-fold symmetry of the interface, each amino acid exchange disrupts twice a given dimer contact.

The NadR dimer interface is formed by at least 32 residues, which establish numerous inter-chain salt bridges or hydrogen bonds, and many hydrophobic packing interactions (*Figure 1.9 A and B*). To determine which residues were most important for dimerization, the interface in silico was studied and several residues were identified as potential mediators of key stabilizing interactions. Using site-directed mutagenesis, a panel of eight mutant NadR proteins was prepared (including mutations H7A, S9A, N11A, D112A, R114A, Y115A, K126A, L130K and L133K), sufficient to explore the entire dimer interface.

The crystal structures presented here allowed a detailed structural analysis for the design of NadR mutants:

- H7, S9, N11 are relevant residues establishing hydrogen bonds interactions both in 4HPA binding pocket;
- Y115 is main-chain interaction with N11 and side-chain with S9, removing it could abolish both potential contacts, between perpendicular helices;
- location of the residue K126 suggests that this might be the main contributor to the interface, being on helix α_6 and symmetrical;
- K130 is located on helix α_6 and mutated in Lys to introduce a long charged residue in place of hydrophobic residues;



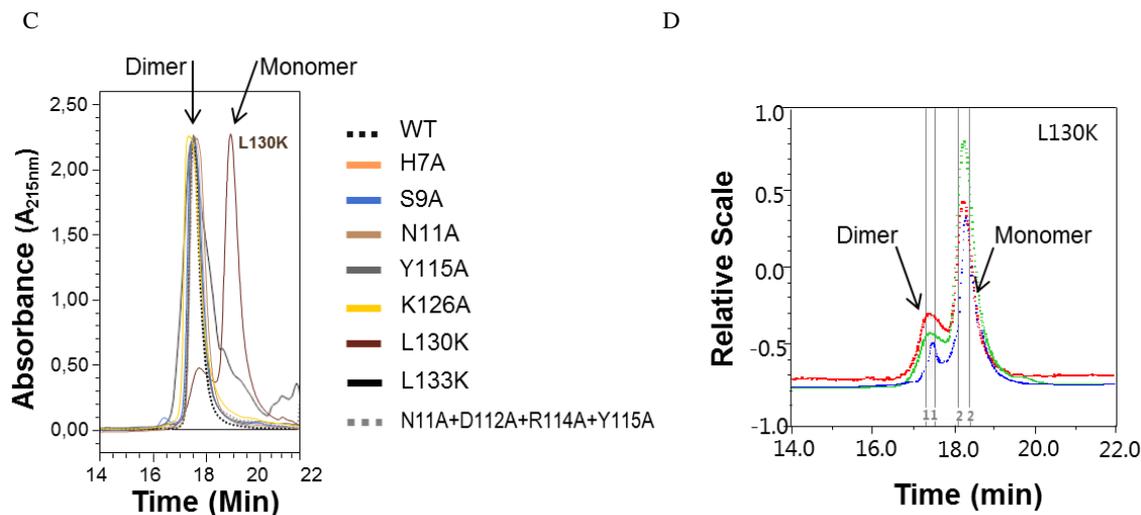


Figure 1.9 Analysis of the NadR dimer interface. (A) Both orientations show chain A, green backbone ribbon, colored red to highlight all locations involved in dimerization; namely, inter-chain salt bridges or hydrogen bonds involving Q4, S5, K6, H7, S9, I10, N11, I15, Q16, R18, D36, R43, A46, Q59, C61, Y104, D112, R114, Y115, D116, E119, K126, E136, E141, N145, and the hydrophobic packing interactions involving I10, I12, L14, I15, R18, Y115, I118, L130, L133, L134 and L137. Chain B, grey surface, is marked blue to highlight residues probed by site-directed mutagenesis (E136 only makes a salt bridge with K126, therefore it was sufficient to make the K126A mutation to assess the importance of this ionic interaction; the H7 position is labelled for monomer A, since electron density was lacking for monomer B). (B) A zoom into the environment of helix $\alpha 6$ to show how residue L130 chain B (blue side chain) is a focus of hydrophobic packing interactions with L130, L133, L134 and L137 of chain A (red side chains). (C) SE-HPLC analyses of all mutant forms of NadR are compared with the wild-type (WT) protein. The WT and most of the mutants show a single elution peak with an absorbance maximum at 17.5 min. Only the mutation L130K has a noteworthy effect on the oligomeric state, inducing a second peak with a longer retention time and a second peak maximum at 18.6 min. To a much lesser extent, the L133K mutation also appears to induce a ‘shoulder’ to the main peak, suggesting very weak ability to disrupt the dimer. (D) SE-HPLC/MALLS analyses of the L130K mutant, shows 20% dimer and 80% monomer. The curves plotted correspond to Absorbance Units (mAU) at 280nm wavelength (green), light scattering (red), and refractive index (blue).

Each mutant NadR protein was overexpressed in *E. coli* and was purified following the same IMAC protocol of NadR wild type protein. The physical effects of these mutations on the protein were analyzed by SEC-MALLS experiments, to determine whether the substitutions had disrupted the interface to form the desired monomeric forms of NadR or led to other structural changes. Almost all the mutants showed the same elution profile as the wild-type (WT) NadR protein. Only the L130K mutation induced a notable change in the oligomeric state of NadR (*Figure1.9C*). Further, in MALLS analyses, the L130K mutant displayed two distinct species in solution, approximately 80% being monomeric (a 19 kDa species), and only 20% retaining the typical native dimeric state (a 35 kDa species) (*Figure1.9D*), demonstrating that Leu130 is crucial for stable dimerization. It is notable that L130 is usually present as Leu, or an alternative bulky hydrophobic amino acid (e.g. Phe, Val), in many MarR family proteins, suggesting a conserved role in stabilizing the dimer

interface. In contrast, most of the other residues identified in the NadR dimer interface were poorly conserved in the MarR family.

Discussion

NadA is a surface-exposed meningococcal protein contributing to pathogenesis, and is one of three main antigens present in the vaccine *Bexsero* [24]. A detailed understanding of the *in vitro* repression of *nadA* expression by the transcriptional regulator NadR is important because it impacts the prediction of vaccine coverage [68], since coverage is estimated using the meningococcal antigen typing system (MATS) through an assay that measures the abundance (and genetic variability) of NadA present in meningococcal strains [97]. The repressive activity of NadR can be relieved by hydroxyphenylacetate (HPA) ligands [68], and HDX-MS studies previously indicated that 4-HPA stabilizes dimeric NadR in a configuration incompatible with DNA binding [94]. Despite these and other studies [72], the molecular mechanisms by which ligands regulate MarR family proteins are relatively poorly understood and likely differ depending on the specific ligand. Given the importance of NadR-mediated regulation of NadA levels in the contexts of meningococcal pathogenesis and vaccine-induced protection, a major aim of this thesis was to characterize NadR, and its interaction with ligands, at atomic resolution.

Firstly, it was confirmed that NadR is dimeric in solution and demonstrated that it retains its dimeric state in the presence of 4-HPA, indicating that induction of a monomeric status is not the manner by which 4-HPA regulates NadR. These observations were in agreement with (i) a previous study of NadR performed using size-exclusion chromatography and mass spectrometry [94], and (ii) crystallographic studies showing that several MarR homologues are dimeric [72]. A structure-guided site-directed mutagenesis was used to identify an important conserved residue, Leu130, which stabilizes the NadR dimer interface, knowledge of which may also inform future studies to explore the regulatory mechanisms of other MarR family proteins. Secondly, the thermal stability and unfolding of NadR in the presence or absence of ligands was assessed. All DSC profiles showed a single peak ($T_m > 65^\circ\text{C}$), suggesting that a single unfolding event simultaneously disrupted the dimer and the monomer (though it cannot be entirely excluded that dimer-to-monomer dissociation occurred sooner with an undetectable signal). HPA ligands specifically increased the stability of NadR. The largest effects were induced by the naturally-occurring compounds 4-HPA and 3Cl,4-HPA, which, in SPR assays, were found to bind NadR with K_D values of 1.5mM and 1.1mM, respectively. Although these NadR/HPA interactions appeared rather weak, their distinct affinities and specificities matched their *in vitro* effects [63, 68] and their biological relevance appears similar to previous proposals that certain small molecules in the millimolar concentration range may be broad inhibitors of MarR family proteins [69, 74]. Indeed, 4-HPA is found in human saliva [67] and 3Cl,4-HPA is produced during inflammatory processes [98], suggesting that these natural ligands might be encountered by *N. meningitidis* in the mucosa of the oropharynx during infections. It is also possible that NadR responds to currently

unidentified HPA analogues. Indeed, in the NadR/4-HPA complex there was a water molecule close to the carboxylate group and also a small unfilled tunnel $\sim 5\text{\AA}$ long, both factors suggesting that alternative larger ligands could occupy the pocket. It is conceivable that such putative ligands may establish different bonding networks and thereby have different allosteric effects, potentially binding in a 2:2 ratio, rather than the 1:2 ratio observed herein. The ability to respond to various ligands might enable NadR *in vivo* to orchestrate multiple response mechanisms and modulate expression of genes other than *nadA*. Ultimately, confirmation of the relevance of each ligand will require a deeper understanding of the available concentration *in vivo* in the host niche during bacterial colonization and inflammation.

Here, the first crystal structures of apo-NadR and holo-NadR were determined, fully refined, and deposited with open-access in the Protein Data Bank (PDB). (Note: previously the only structural data available on NadR was presented in my undergraduate thesis, University of Palermo (2012). In holo-NadR, 4-HPA interacted with at least 11 polar and hydrophobic residues: Ser9, Asn11, (chain A), and Arg18, Leu21, Met22, Phe25, Leu29, Asp36, Trp39, Arg43, Val111 and Tyr115 (chain B). Several, but not all, of these interactions were predicted previously by homology modelling combined with ligand docking *in silico* [94]. More unexpectedly, only one molecule of 4-HPA was bound per NadR dimer. This stoichiometry in solution was confirmed using SPR methods. Our crystallographic observation of this 'occupied vs unoccupied' asymmetry in the NadR/4-HPA interaction is, to our knowledge, the first example of a non-antibiotic small molecule ligand reported for a MarR family protein. Structural analyses suggested that 'inward' side chain positions of Met22, Phe25 and Arg43 precluded binding of a second ligand molecule. Such a mechanism indicates negative cooperativity, which may enhance the ligand-responsiveness of NadR, as proposed previously for HucR [99] and MTH313 [74].

Comparisons of the NadR/4-HPA complex with available MarR family/salicylate complexes revealed that 4-HPA has a previously unobserved binding mode. Briefly, in the *M. thermoautotrophicum* MTH313 dimer, one molecule of salicylate binds in the pocket of each monomer, though with two rather different positions and orientations, only one of which (site-1) is thought to be biologically relevant [74]. In the *S. tokodaii* protein ST1710, salicylate binds to the same position in each monomer of the dimer, in a site equivalent to the putative biologically relevant site of MTH313 [75]. In TcaR, eight molecules of salicylate were found in different sites within and on the surface of the TcaR dimer [100]; and the related protein structure of SAR2349 revealed 6 salicylate binding sites per dimer [77]. In contrast, in NadR, only one molecule of 4-HPA binds per dimer, in a position distinctly different from the more relevant salicylate binding site of MTH313 and ST1710: translated by $>10\text{\AA}$ and with a 180° inverted orientation.

Interestingly, a crystal structure was previously reported for a functionally-uncharacterized meningococcal homologue of NadR, termed NMB1585, which shares only 16% sequence identity with NadR [101]. The two structures can be closely aligned (rmsd 2.3 Å), but NMB1585 appears unsuited for binding HPAs. It can be speculated that MarR family members have evolved separately to engage distinct signaling molecules, thus enabling bacteria to use the overall conserved MarR scaffold to adapt and respond to diverse changing environmental conditions within their natural niches. Alternatively, it is possible that other MarR homologues have no extant functional binding pocket and thus may have lost the ability to respond to a ligand, acting instead as constitutive DNA-binding regulatory proteins.

The apo-NadR structure revealed two dimers with slightly different conformations, most divergent in the DNA-binding domain. It is not unusual for a crystal structure to reveal multiple copies of the same protein in slightly different conformations, which are likely representative of the dynamic ensemble of molecular states naturally sampled by the molecule in solution and with only small energetic differences, as described previously for MexR [102] or more recently for the solute-binding protein FhuD2 [103, 104]. Further, the holo-NadR structure was overall slightly different from the two apo-NadR structures (rmsd values $\sim 1.3\text{\AA}$), suggesting that the ligand selected and stabilized yet another conformation of NadR. These observations suggest that 4-HPA, and potentially other similar ligands, can shift the equilibrium in the molecular landscape, changing the energy barriers that separate active and inactive states, and stabilizing the specific conformation of NadR poorly suited to bind DNA.

Comparisons of the apo- and holo-NadR structures revealed that the largest differences occurred in the DNA-binding helix $\alpha 4$. The shift of helix $\alpha 4$ in holo-NadR was also accompanied by rearrangements at the dimer interface, involving helices $\alpha 1$, $\alpha 5$, and $\alpha 6$, and this holo-form appeared poorly suited for DNA-binding when compared with the known OhrR:*ohrA* protein:DNA complex [96]. While some flexibility of helix $\alpha 4$ was also observed in the two apo-structures, concomitant changes in the dimer interfaces were not observed, possibly due to the absence of ligand. One of the two conformations of apo-NadR appeared ideally suited for DNA-binding. Overall, these analyses suggest that the apo-NadR dimer has a pre-existing equilibrium that samples a variety of conformations, only some of which are compatible with DNA binding. This intrinsically dynamic nature underlies the possibility for different conformations to inter-convert or to be preferentially selected by a regulatory ligand, as generally described in the 'conformational selection' model for protein-ligand interactions (the Monod-Wyman-Changeux model), rather than an 'induced fit' model (Koshland-Nemethy-Filmer) [105]. The noted flexibility may also explain how NadR can adapt to bind various DNA target sequences [64] with slightly different structural features. Subsequently, upon ligand binding, holo-NadR adopts a structure less suited for DNA-binding and this

conformation is selected and stabilized by a network of protein-ligand interactions and concomitant rearrangements at the NadR holo dimer interface. In a similar but less extensive manner, the binding of two salicylate molecules to the *M. thermoautotrophicum* protein MTH313 appeared to induce large changes in the wHTH domain, which was associated with reduced DNA-binding activity [74].

Here two new crystal structures have been presented for the transcription factor, NadR, which regulates expression of the meningococcal surface protein and vaccine antigen NadA. Detailed structural analyses provided a molecular explanation for the ligand-responsive regulation by NadR on the majority of the promoters of meningococcal genes regulated by NadR, including *nadA* [64]. Intriguingly, NadR exhibits a reversed regulatory mechanism on a second class of promoters, including *mafA* of the multiple adhesin family – i.e. NadR represses these genes in the presence but not absence of 4-HPA. The latter may influence the surface abundance or secretion of *maf* proteins, an emerging class of highly conserved meningococcal putative adhesins and toxins with many important roles [65, 66]. Future structure determination of the NadR protein bound to the NadR-regulated promoters (e.g. *nadA* or *mafA* operators) combined with biophysical analyses of the binding could lead to a comprehensive understanding of differential mechanism patterns of this transcriptional regulator. Studies of other known MarR-DNA complexes suggest that regulation of protein-DNA binding could be mediated through conformational change either at protein level [73] or at the DNA binding lobe [106]. However, the molecular basis of NadR-DNA binding appears dependent on communication between two monomers through side chains position. Additional structural data could allow identification of the specific-sequence recognition elements that give rise to differential NadR-DNA regulation. The protein-DNA structure could also confirm the hypothesis of negative cooperativity for protein-ligand regulation systems. However, determining the crystal structure of protein–DNA complex by X-ray requires the preparation of stable and homogenous samples that results to be more challenging when it comes to handle the DNA. Prior crystallization or supporting it, it is also important to correctly identify the specific nucleotide sequence recognized by the protein and to well characterize the interaction between component, in terms of stoichiometry and affinity of the resulting complex, by several biophysical and biochemical experiments. Although X-ray crystallography can provide detailed information about binding site recognition [107, 108], it is also important to consider that crystallography provides only a static view of a protein–DNA complex [109]. A powerful tool that can be used to investigate protein–DNA recognition in the solution state and supports the crystallographic analyses is Nuclear Magnetic Resonance (NMR) spectroscopy. NMR data can be used to elucidate atomic-level conformational dynamics, but can be difficult and time-consuming [110-113]. The availability of a suitable protein-DNA complex for structural studies often hampers efforts in both techniques.

Two others possible routes to such a complex structure are (i) small angle X-ray (SAXS) analyses [114]; and (ii) cryo-electron microscopy (Cryo-EM) or, depending on the available information, fitting atomic structures of the individual protein subunits into a Cryo-EM map of the assembly [92]. In the latter case, atomic structures of the protein-DNA complex are required. SAXS and X-ray diffraction are fundamentally similar [115]. SAXS analysis can be applied to flexible proteins that do not crystallize readily [116]. SAXS differs from X-ray crystallography in that it is applied to proteins in solution rather than crystals; thus, it can be applied to a much wider range of proteins in states more closely resembling their functional forms, but the information is rotationally averaged and so the resulting SAXS profile gives less structural information [117]. Cryo-electron microscopy is also gaining momentum and popularity in structural biology studies [118]. Advances in electron detectors and software for the processing of thousands of images or the correction of beam-induced motion are supporting this development facilitating the gaining of structural information at increasingly high resolution, even approaching the atomic level [119]. SAXS or cryo-EM in combination with X-ray crystallographic data can be very powerful for the analysis of multicomponent systems [120, 121].

The NadR-DNA complex represents a considerable opportunity for synergy between NMR spectroscopy, X-ray crystallography, and complementary structural determination techniques. A hybrid structural approach could take advantage of the most accessible aspects of each structural technique and may be widely applicable for structure complex determination. So, further work is required to investigate how the two different promoter types influence the ligand-responsiveness of NadR during bacterial infection and may provide insights into the regulatory mechanisms occurring during these host-pathogen interactions.

Ultimately, knowledge of the ligand-dependent activity of NadR will continue to deepen our understanding of *nadA* expression levels, which influence meningococcal pathogenesis and vaccine-mediated protection. In that sense, it could be useful perform an extensive experimental or computational small molecule screening with large libraries of ligands to find others inhibitors, providing additional insights into transcriptional regulation mechanism [122]. A variety of screening methodologies exist to identify molecules [123]: a general high throughput screening (HTS) involving the screening of a large compound library against the target protein [124]; a focused or knowledge-based screening based on selection from a smaller library of molecules that are likely to act on the target protein based on previous studies [125]. Both methodologies permit to rapidly identify out of a large library all active compounds against a particular protein target. Finally, the identified small molecules could be tested and validated through *in vitro* and *in vivo* experiments. Such screenings could generate data to develop a hypothesis that the small molecule-modulated NadA

expression inhibition or activation could result in meningococcal mechanisms for adapting to changes in their environment during infection.

Part Two

**New strategies towards a crystal structure of
meningococcal antigen NadAv3**

Abstract

Over the last five years, structural biology has emerged as a major tool supporting the rational design of novel vaccine antigens. In the vaccine research and development pipeline, where Reverse Vaccinology can be used to identify candidates for a protein-based vaccine against a specific pathogen, the structural information can be instrumental during the optimization phases. The structure-based approach, able to generate information on the overall antigen structure, becomes the driving force in the production of engineered antigens with improved immunological properties and biophysical attributes that facilitate manufacturing.

Structural studies of one of the protein components of the *Bexsero* vaccine, the meningococcal NadA variant 3 (NadAv3) antigen, are described in part II of this thesis. The aim was to obtain the 3D crystal structure of NadAv3, in order to enable a deep characterization of its function and its role in eliciting the immune response. Several strategies to crystallize NadAv3 and solve its structure were followed. Firstly, new v3 constructs of distinct C-terminal length were designed (*Figure 2.3*). This initial clone design was inspired by previous successful crystallographic work on NadA variant 5 (NadAv5), a variant that shares approximately 50% sequence identity with NadAv3. A comprehensive set of biochemical, biophysical and structural techniques were applied to investigate all the newly-generated NadAv3 constructs, leading towards samples with improved thermal stability for use in structural studies by X-ray crystallography. Initially, low quality crystals were obtained, and therefore mutagenesis studies were carried out to improve the diffraction quality of these crystals. Ultimately, crystals that diffracted reproducibly up to 2.2 Å resolution have been obtained and the structure determination process is ongoing. The atomic resolution structure of NadA will help to understand its biological role as both an adhesin and a vaccine antigen. Moreover, structural insights will enable a structural vaccinology approach for the design of a more broadly cross-protective antigen, as described previously for the highly variable meningococcal factor H binding protein (fHbp). In addition, the structure of NadA will expand our knowledge of the TAAs family. Finally, the new NadAv3 constructs have been used as 'validated probes' for various functional studies and epitope mapping experiments, contributing to the understanding of the role of NadA in the human immune response and endothelial receptor interactions.

The *Neisseria meningitidis* adhesin A (NadA)

Several studies, carried out using different cell types, including epithelial cells, monocytes, macrophages, and monocyte-derived dendritic cells, have demonstrated the importance of NadA in bacterial adhesion. NadA has been shown to be involved in bacterial uptake in epithelial cells. For example, *E. coli* bacteria expressing NadA on the surface were shown to be internalized in Chang cells in a NadA-dependent fashion. In addition a deletion of the N-terminal globular domain of recombinant NadA abrogated the adhesive phenotype [20]. Additional studies revealed the regions of NadA between amino acids 94 to 110 and 109 to 121 as being involved in NadA-mediated cell-bacterium adhesion [126]. However, the precise mechanism of adhesion is unknown, and a specific human receptor has not yet been clearly identified.

Further studies have shown that recombinant NadA binds to hsp90 *in vitro* and on the surface of monocytes [127]. In the proposed model, NadA binds to an unknown receptor, encounters hsp90 probably by lateral diffusion and then is recruited into a complex also comprising hsp70 and TLR4. This complex is inhibited by polymixin B, which interferes with NadA-hsp90 binding but not with NadA cell binding. The complex is also necessary for full monocyte stimulation and may be important to modulate or enhance the vaccine immune response. Furthermore NadA has been shown to interact with β -integrins on epithelial-like GE-11 and fibroblast-like 2-4-8 cells using flow cytometry. The direct binding data have been corroborated by blocking experiments with anti-human β 1 monoclonal antibodies which have been described to compete with NadA for β -integrin binding [128]. Based on these studies, NadA appears to be a key determinant of meningococcal interactions with the human host at different stages of meningococcal infection. In addition it has recently been demonstrated by large scale protein microarray-based approach that NadA is able to bind to the human LOX-1 (low-density oxidized lipoprotein lectin-like receptor 1) receptor. LOX-1 represented an interesting candidate for further validation since it is known to be involved in bacterial adhesion and invasion [129, 130], endotoxin-induced inflammation [131] and its deletion enhances bacterial clearance in a murine polymicrobial sepsis model [132].

The TAA family –common structural organization

NadA belongs to the trimeric autotransporter adhesion (TAA) family, a class of outer membrane adhesins present in Gram-negative bacteria. These obligate homotrimeric proteins are embedded in the outer membrane (OM) and act primarily as adhesins. Members of the TAA family can bind to diverse molecules, such as cell surface receptors, components of the extracellular matrix (ECM) such as collagen and laminin, and each other (i.e. they act as autoagglutinins). TAAs comprise a group of virulence-related proteins in Gram-negative

bacteria [133], and they typically share a common modular organization that consists of an N-terminal “passenger” domain followed by a C-terminal translocation unit/membrane anchor (Figure 2.1).

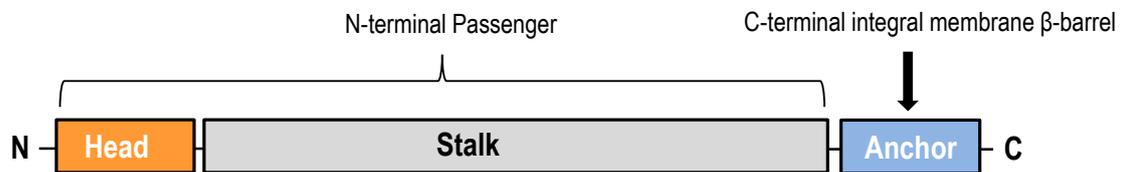


Figure 2.1. TAA Structural Organization. TAAs show a simple head-stalk-anchor organization. The head mediates host-specific binding properties, the stalk projects the head beyond the membrane, and the membrane anchor secretes both previous components while maintaining the protein bound to the outer membrane surface.

The N-terminal part of the TAA, the passenger domain, is responsible for binding to specific host macromolecules. The highly conserved C-terminal domain, the translocation unit, transports the passenger across the outer membrane (OM) into the extracellular space. The β -barrel is the only part of the protein strictly conserved between family members in terms of sequence and structure. Bacteria have evolved a wide array of head domains, and the structural mismatch between the globular head domains and the fiber-like stalk domain requires the presence of an additional class of “neck” elements acting as connectors. These provide a smooth transition to and from the stalk. The coiled coils in TAAs are forced to be trimeric, and typical trimeric coiled coils are left-handed. The sequence motif encoding a coiled coil is composed of hydrophobic (H) residues separated by three and four polar (P) residues $[(HPPHPPP)_n \geq 3]$: where the positions in the heptad repeat are designated *abcdefg*. The crossing angle between the helices in a coiled coil is close to zero, and the packing follows a “knobs-into-holes” arrangement [134], where the knobs formed by hydrophobic residues in positions *a* and *d* pack into cavities formed by residues on a neighboring helix [135]. In TAAs, the three helices are wound in register around each other, so all of the residues are at the same height. Position *a* favours β -branched side-chains (Ile, Val, Thr), while residues in position *d* are closer together and so unbranched residues (Leu, Ala) are favoured [136]. The translocation process seems to be independent of any external source of free energy, such as adenosine triphosphate (ATP), ion gradients, or other proteins; hence the name autotransporter [137]. Overall, the available data suggest that folding, function and structure are very closely linked in TAAs. Nevertheless, the recently solved structure of NadAv5 showed features of a novel trimeric autotransporter adhesin that has no close homologs among other TAAs proteins present in the Protein Data Bank (PDB) [138]. Interestingly, NadA5 is made of a trimeric coiled-coil which includes both the apical N-

terminal region and the main stalk, with a peculiar sequence insertion which gives rise to wing-like structures without altering the coiled-coil geometry (*Figure 2.2*).

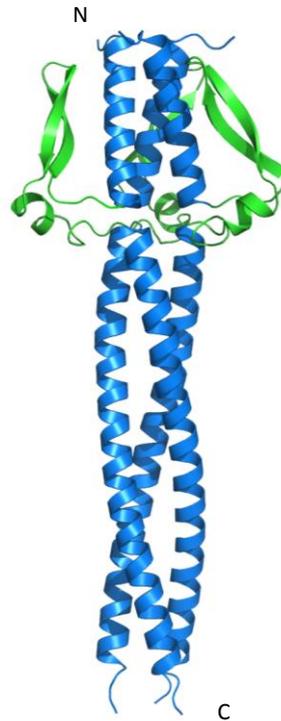


Figure 2.2. Crystal structure of NadAv5. Crystal structure of NadA5 is shown as cartoons; the stalk (coiled coil region) is shown in blue and the wing-like insertions of the head domain are shown in green.

NadA variants

Nucleotide sequence analyses of *nadA* have shown that the gene is present in approximately 30% of *N. meningitidis* major disease-associated strains, and is associated mostly with strains belonging to three out of the four hypervirulent lineages [21, 139]. Updated sequence analyses of NadA indicate the presence of four variants clustering in 2 groups: group I comprising protein variants NadAv1 and NadAv2/3, and group II including protein variants NadAv4/5 and NadAv6. NadAv1 and NadAv2/3 are the most represented and highly cross-protective (i.e. in mouse models, immunization using NadAv3 antigen elicits serum bactericidal activity ‘cross-protective’ against meningococcal strains with surface-expression of NadAv1 and v2 [140]). NadAv4/5 is predominantly associated with carriage isolates and is poorly cross-reactive with variants 1 and 2/3. NadAv6 is closely related to NadAv4/5 and was originally described in one isolate belonging to the ST-11 complex [21, 141, 142]. In the multi-component vaccine *Bexsero*, NadAv3 is present as a well-characterized drug substance termed NadA $\Delta_{351-405}$ [143], which is essentially a soluble trimeric ectodomain form that lacks the transmembrane anchor. However, to date, detailed structural information is available only on NadAv5 [138].

Experimental procedures

NadA constructs cloning, expression and mutagenesis. The *nadA* gene fragments were PCR amplified from the serogroup B *N. meningitidis* strains 2996 (NadAv3) and M01-240320 (NadAv5) and were inserted into the pET-21b(+) vector (Novagen), as described previously [144]. The NadA expression constructs were cloned without the signal peptide and a C-terminal 6-His tag was inserted to facilitate protein purification. The sequence numbering used here refers to the full-length NadAv3 and NadAv5 proteins, UniProt accession numbers Q8KH85 and A0ELI2, respectively. PCR products encoding NadA fragments and point mutants with a 6His tag at the C terminus were cloned using the polymerase incomplete primer extension cloning method (PIPE) method [145]. After sequencing, each plasmid was used to chemically transform *E. coli* BL21 (DE3) cells (Novagen) for protein production.

NadA protein production and purification. Cells were grown following the manufacturer's protocol using BioSilta medium that guarantees a minimum 5-fold increase in protein yield from EnPresso B Animal-free growth systems when compared to yields from typical LB medium. Cultures were aerated in 250ml shake flasks at 30°C for 30 h, and production of the NadA constructs was induced by the addition of 0.1mM IPTG (isopropyl β -D-thiogalactopyranoside). After one day culture, cells were harvested by centrifugation and were suspended in 50 mM NaH₂PO₄, 300 mM NaCl (pH 8.0,) followed by mechanical disruption. Cell lysates were clarified by centrifugation at 30000 g for 30 min at 4°C, and the *E. coli* extract supernatant was removed and filtered using a 0.22 μ m membrane (Corning filter system) prior to protein purification. The soluble cell extract was loaded on a HisTrap 1 ml column (GE Healthcare). Proteins were eluted with 50 mM NaH₂PO₄, 250 mM NaCl, 60 mM Imidazole pH 8.0. Only the NadAv5 constructs required an additional purification step with a Q HP anion exchange resin (GE Healthcare) after dialysis in 20 mM Tris-HCl pH 8.0. Fractions containing NadA were identified by SDS-PAGE (12% gel) analysis and were further purified by preparative size-exclusion chromatography (HiLoad Superdex 75 (16/60), GE Healthcare) in buffer containing 20mM Tris-HCl, 150 mM NaCl, pH 8.0. All proteins were used immediately or frozen for storage at -20°C. Storage of NadA proteins for any length of time can pose stability problems.

Size-exclusion high-performance liquid chromatography (SE-HPLC) coupled with Multi-angle laser light scattering (MALLS). Size-exclusion high-performance liquid chromatography (SE-HPLC), revealed a high level of purity and a lack of any aggregated species. SE-HPLC was performed at RT (18-26°C) on an analytical size exclusion TSK Super SW3000 column by loading 20 μ l of each sample at a concentration of ~ 40 μ M. Samples were eluted isocratically in 0.1M NaH₂PO₄, 0.4M (NH₄)₂SO₄ buffer at pH 6.0.

Coupling SE-HPLC with Multi-angle laser light scattering (SE-HPLC/MALLS) NadA samples were analyzed for absolute molecular size in solution. Data analyses were carried out using Astra V software (Wyatt) to determine the weight-average molecular mass (MW) in Daltons and the polydispersity index (MW/Mn) for each oligomer present in solution. Normalization of the MALLS detectors was performed in each analytical session by use of bovine serum albumin.

Differential Scanning Calorimetry (DSC). The thermal stability of NadA proteins was assessed by DSC using a MicroCal VP-Capillary DSC instrument (GE Healthcare). NadA samples were prepared at a protein concentration of 0.5mg/mL (~10 μ M) in PBS buffer. The DSC temperature scan ranged from 10°C to 110°C, with a thermal ramping rate of 200°C per hour and a 4 second filter period. Data were analyzed by subtraction of the reference data for a sample containing buffer only, using the Origin 7 software. All experiments were performed in duplicate, and mean values of the melting temperature (T_m) were determined.

Surface Plasmon Resonance (SPR). Surface plasmon resonance (SPR) was used to study the binding of NadAv3 constructs to several human and murine monoclonal antibodies (mAbs). All SPR experiments were performed using a Biacore T200 instrument at 25 °C (GE Healthcare). For the single-cycle kinetics (SCK) experiments, which are well-suited for the measurement of high affinity binding events, either a commercially available Mouse Antibody Capture Kit (GE Healthcare) or a Human Fab Capture kit (GE Healthcare) was used to covalently-immobilize respectively anti-mouse or anti-human IgG antibodies by amine coupling on a carboxymethylated dextran sensor chip (CM-5; GE Healthcare). A density level yielding ~10,000 response units (RUs) was prepared for both immobilizations. The anti-mouse IgG was used then to capture ~1200 RU murine mAb 6E3 while the anti-human Fab (anti-huFab) IgG was used to capture also ~1500 RU of the chimeric mAbs. Experimental running buffer contained 10 mM Hepes, 150 mM NaCl, 3mM EDTA, 0.05% (vol/vol) P20 surfactant, pH 7.4. For the determination of K_D and kinetic parameters, a titration series of five consecutive injections of increasing analyte concentration (range 6.25-100 nM; flow rate of 40 μ L/min) followed by a single final surface regeneration step with buffer containing 10 mM glycine pH 1.7 (flow rate of 10 μ L/min) was performed using the standard SCK method implemented by the Biacore T200 Control Software (GE Healthcare). Anti-mouse or anti-human antibody-coated surfaces without captured mAb were used as the reference channel. A blank injection of buffer only was subtracted from each curve and reference sensorgrams were subtracted from experimental sensorgrams to yield curves representing specific binding. The data shown are representative of at least two independent experiments. SPR data were analyzed using the Biacore T200 Evaluation software (GE Healthcare). Each sensorgram was fitted with the 1:1 Langmuir binding model, including a

term to account for potential mass transfer, to obtain the individual k_{on} and k_{off} kinetic constants; the individual values were then combined to derive the single averaged K_D values reported. For the single injection experiments a sample for each protein tested was injected at a concentration of 200 nM over captured either murine or chimeric human IgGs. Surfaces were regenerated between injections with glycine pH 1.7 as described above.

Crystallization of NadAv3 proteins and X-ray diffraction data collection. Purified NadAv3 proteins were concentrated between 10-42mg/mL, depending on the construct, using a centrifugal concentration device (Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-10 membrane with cut-off size 10kDa; Millipore) running at 600 g in a bench top centrifuge (Thermo Scientific IEC CL40R) refrigerated at 2-8°C. The concentrated NadAv3 proteins were subjected to crystallization trials performed in 96-well low-profile Intelli-Plates (Art Robbins) or 96-well low-profile Greiner crystallization plates, using a nanodroplet sitting-drop vapour-diffusion format and mixing equal volumes (200nL) of protein samples and crystallization buffers using a Gryphon robot (Art Robbins).

Each sample was tested using 5 different commercially available screens, namely: JCSG, Morpheus and Structure (Molecular Dimension); PEGlon and Saltrix (Hampton Research). Crystallization trays were incubated at 20° C in the RockImager 182 (Formulatrix) incubator and imager. Crystals of NadAv3₂₄₋₁₇₀ that did not diffract x-rays were obtained at 20°C in condition B6 of JCSG-plus™ HT-96 screen (Molecular Dimensions Ltd), which contains 0.1M sodium phosphate citrate pH 4.2 (Buffer), 40 % (v/v) ethanol (Precipitant) and 5 % (w/v) polyethylene glycol (PEG) 1K (Precipitant). Optimization of this initial crystallization condition yielded reproducible crystals by slightly lowering the concentration of ethanol and PEG 1K to 36.4 % (v/v) and 4.5 % (w/v), respectively. Substitution of ethanol by 2-Methyl-2,4-pentanediol (MPD) allowed the generation of high quality crystals of NadAv3₂₄₋₁₇₀, NadAv3_{24-170_A331-I38L} NadAv3_{24-170_A331-I38L-A39V} that diffracted between 2.2 and 2.8 Å resolution. All crystals grew in 0.1M sodium phosphate citrate, pH 3.9, and 5 % (w/v) PEG 1K, and a range of MPD concentrations between 33.64 and 45.91 %v/v.

X-ray data collection experiments are usually conducted at cryogenic temperatures (100K), in order to reduce radiation damage of the crystals induced by the X-rays. Thus, crystals must be first frozen (without ice formation and damage), and to do so these are soaked into so-called cryo-protectant solutions mixed with the crystallization reagent. However, crystals growing in certain conditions, including high alcohol or MPD as for the NadAv3 crystals described above, can be cryo-cooled without any addition of additives. Thus, prior of data collection crystals of NadA3 were mounted in cryo-loops without additional cryo-protectant and were cooled to 100 K in liquid nitrogen.

X-ray diffraction data from crystals of NadAv3₂₄₋₁₇₀, and NadAv3_{24-170_A331-I38L} and NadAv3_{24-170_A331-I38L-A39V} were collected on beamline ID23 and ID30A-3, respectively, of the European synchrotron radiation facility (ESRF), Grenoble, France.

Results

Structure-based design: seeking the minimal N-terminal domain of NadAv3

Extensive attempts to crystallize the soluble recombinant tagless NadAv3 vaccine construct (NadAv3 $\Delta_{351-405}$) [143], which includes the head and entire stalk domain, but which lacks the transmembrane anchor, were unsuccessful (unpublished results). Similarly, despite being available at much higher degrees of purity, previous efforts were also unsuccessful to crystallize C-terminally 6-His tagged forms of this full-length ectodomain protein (C-His NadAv3 $_{24-342}$), or several C-terminally truncated constructs (spanning residues 24-293, 24-284, 24-274 and 24-268) [138]. The inability to crystallize NadAv3 constructs may well be related to its relatively long stalk region (expected to be approximately 400Å long), which displayed inherent flexibility in negative-stain electron microscopy studies [138]. Typically, long flexible molecules are expected to be less likely to crystallize than more compact stable molecules. Therefore, various strategies, including the design of alternative NadA variants (NadAv4 and v5) to exploit slight variation in amino acidic composition, known to affect crystallizability, were previously employed. Reproducible diffraction-quality crystals were obtained only from a single variant 5 construct, NadAv5 $_{24-220}$, allowing to solve the NadAv5 structure by single anomalous dispersion (SAD) methods by soaking of the crystals with sodium iodide [138]. The work presented here describes new approaches implemented to aid crystallization of v3, using the v5 crystal structure to rationally-design new v3 constructs of distinct C-terminal length (*Figure 2.3*).

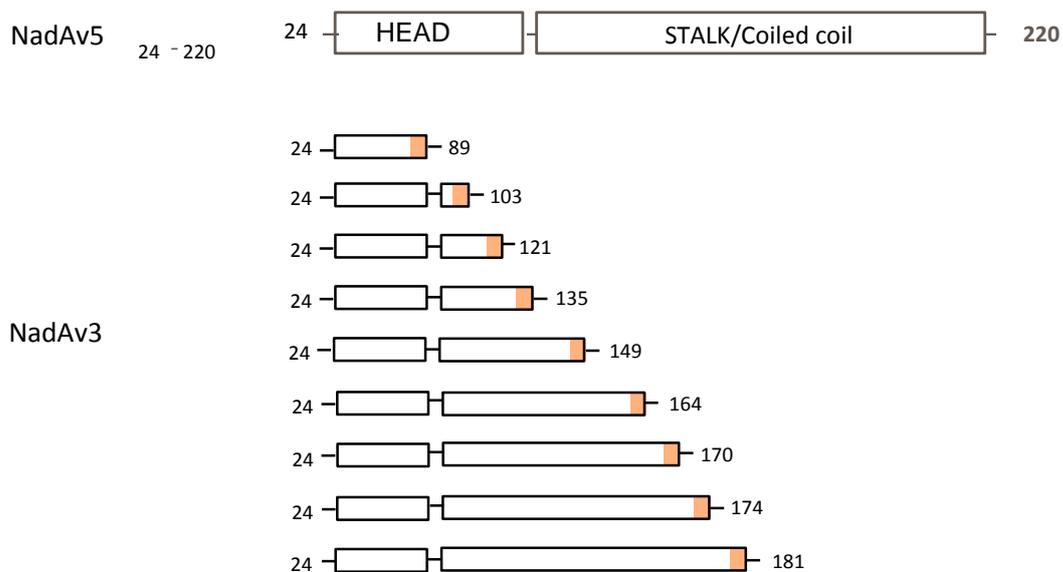


Figure 2.3. Domain organization and constructs of NadAv3. Each expression construct had a C-terminal 6His tag (orange box) to enable affinity purification. The predicted domain organization and domain boundaries are shown schematically by white boxes.

A first expression screen revealed that all the new NadAv3 constructs were expressed in soluble form in large amounts in *E. coli*. All the recombinant NadAv3 protein constructs spanning residues 24-89 to residue 24-181 were purified from the soluble cytoplasmic fraction. The procedure involves a first-step purification of the recombinant protein by immobilized metal ion affinity chromatography (IMAC) and a second step by size-exclusion chromatography (SEC), as described in *Materials and methods*. The corresponding proteins ran at the expected monomeric molecular mass in SDS 4-12% PAGE analyses performed under reducing and non-reducing denaturing conditions. The final sample purity was estimated as >95% in SE-HPLC analysis. The absolute molecular weight values of the proteins were calculated by MALLS analysis and the thermal stability was evaluated by DSC experiments. The extensive biophysical and biochemical analysis revealed that NadAv3₂₄₋₁₇₀ is the shortest v3 construct that is well-folded, stable and fully trimeric (see *Table 2.1*):

Construct	SE-HPLC (Purity %)	MALLS (Absolute MW)	DSC (T _m , °C)
NadAv3_24-89	99%	monomer	N/A
NadAv3_24-103	99%	monomer	N/A
NadAv3_24-121	> 95 %	98% monomer 2% trimer	32.8
NadAv3_24-135		92 % monomer 8 % trimer	31.5
NadAv3_24-149		70 % monomer 30 % trimer	37.5
NadAv3_24-167		60 % monomer 40 % trimer	38.6
NadAv3_24-170	99%	trimer	39.1
NadAv3_24-174	>95%	trimer	39.8
NadAv3_24-181		trimer	41.8

Table 2.1. Biophysical and biochemical properties of NadAv3 constructs. Six constructs were observed to exist in a mixture of oligomeric states on SEC-MALLS. If a measurable trimeric molecular weight (MW) was observed, then this is indicated by the label “trimer”. If no trimeric MW was observed, then the percentage of oligomeric states of the two oligomeric species is provided. N/A, not applicable. The sample purity and the thermal stability of the constructs, evaluated by HPLC and DSC respectively, are listed

For each construct, the final purification yield was approximately 1mg of purified protein per 1gr wet biomass, obtaining a sufficiently pure amount of material to be used in at least five different commercial crystallization screen. Crystallization trials were performed for the three trimeric constructs: NadAv3₂₄₋₁₇₀, NadAv3₂₄₋₁₇₄, and NadAv3₂₄₋₁₈₁.

Strategies towards the crystal structure of the NadAv3₂₄₋₁₇₀ construct

Several approaches were followed in order to grow diffraction-grade protein crystals of NadAv3. Initial crystals of NadAv3₂₄₋₁₇₀ obtained in 0.1M sodium phosphate citrate pH 4.2, 40%v/v ethanol and 5 %w/v PEG 1K (Conditions B6 of JCSG screen) did not diffract. Thus, optimization screenings were performed, slightly lowering the concentration of ethanol and PEG 1K to 36.4 %v/v and 4.5 %w/v, respectively. This approach allowed us to obtain crystals that were diffracting to a maximum resolution of 6Å.

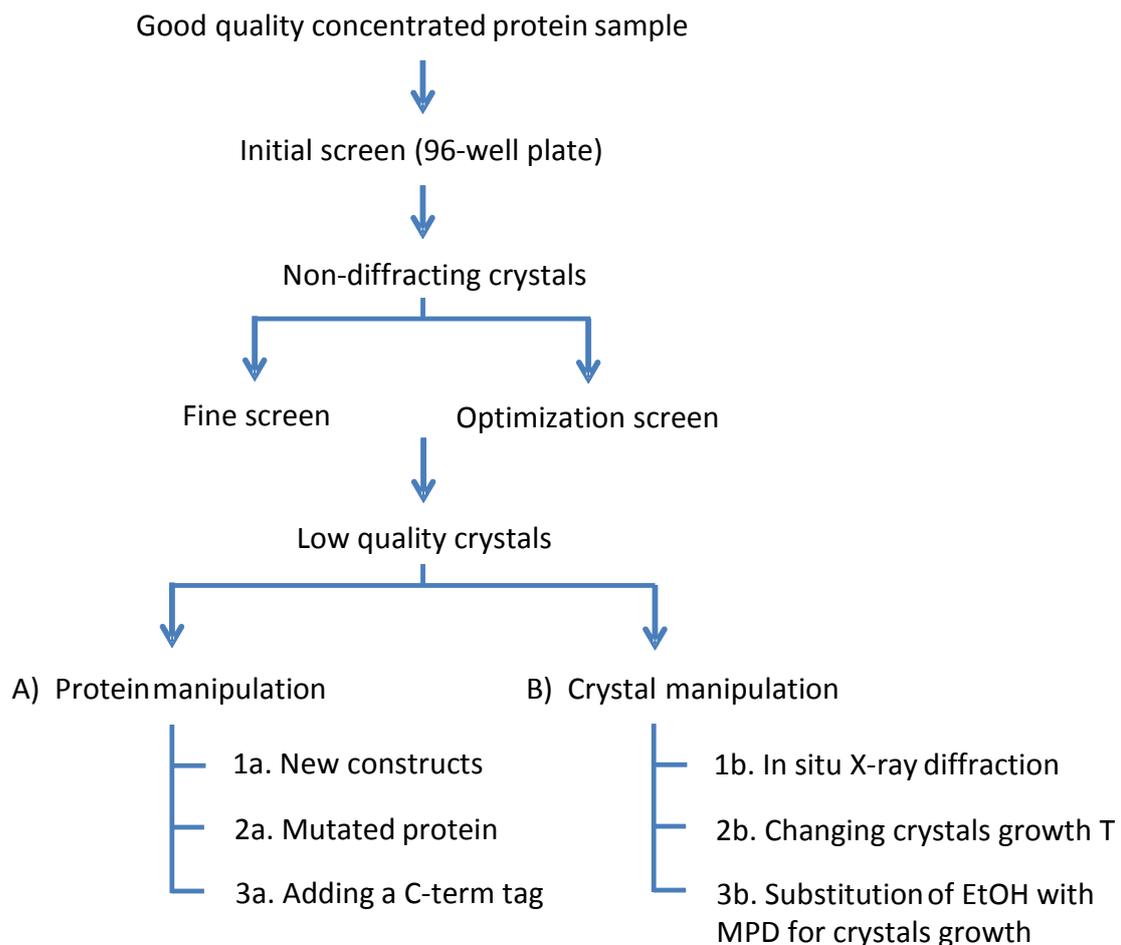


Figure 2.4. Decision tree for optimization experiments. Several approaches can be taken when a hit condition is identified and these can be performed in conjunction with each other or separately.

The flow diagram illustrated in *Figure 2.4* show the various steps towards crystallizing a protein and the options that were followed to overcome problems that were encountered. The most promising construct crystallized readily but generated low quality diffracting crystals. Next, different strategies were applied, ranging from protein to crystal manipulation, to improve diffraction quality of protein crystals.

A) Protein engineering flow-path

New protein constructs of NadAv3 were generated (a) that lacked putatively flexible regions, generating a set of shorter C-terminally truncated constructs (1a) that retain the trimeric conformation. Additional stabilized mutants (2a) were produced by adding a C-terminal GCN4 tag which was derived from a GCN4 leucine zipper to ensure trimerization [146], or a C-terminal foldon trimerization domain of phage T4 fibrin [147] to increase conformational stability (3a). These constructs were tested for increased thermal stability in DSC experiments (*Figure 2.5*).

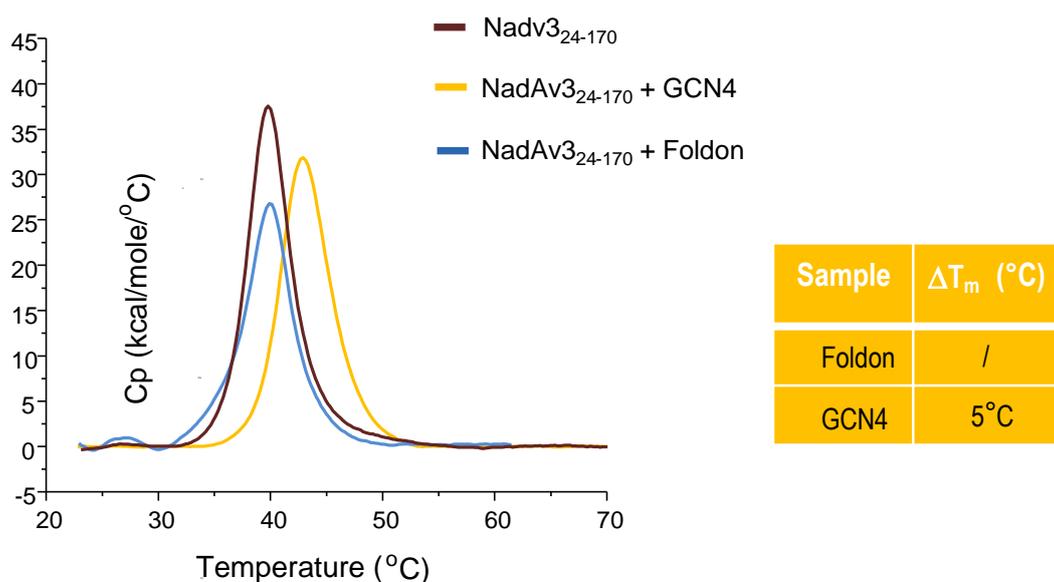


Figure 2.5. DSC profiles of engineering NadA constructs. A symmetric peak with $T_m \sim 40^\circ\text{C}$ (brown) shows that the NadAv3_{24-170} is stably folded. Smaller peak was observed with similar T_m value for the NadAv3_{24-170} + foldon construct (light blue); a shifted peak with an increased T_m was observed for the NadAv3_{24-170} + GCN4 construct (yellow).

B) Crystal manipulation flow-path

In addition, crystal manipulation techniques (b) were followed, since crystals that grow from high concentrations of volatile alcohols are notoriously difficult to handle during mounting and flash-freezing procedures. Cryocooling is in fact successful only when done quickly, hence the term flash-freezing. Briefly, a crystal is manually ‘fished’ from the crystallization drop in which it grew (typical volume < 500 nanolitres) using a tiny nylon loop,

and is then rapidly transferred into liquid nitrogen or a gaseous cryostream at around 100K. The speed of flash freezing prevents formation of crystalline ice, which disrupts the crystal lattice and degrades the data quality. Specifically, these crystals had a tendency to move inside drops due to convection currents and droplet evaporation, and to dissolve when the drop containing the crystal was touched even without directly touching the crystal. Even the addition of a more concentrated solution directly on top of the crystallization drop was found not to be useful in helping with the handling of these crystals. Nevertheless, ultimately about 100 different crystals obtained from that particular condition by numerous repeated efforts over the course of 2 years were screened, but did not yield useful diffraction data (max. resolution only about 6Å in one direction).

Based on these observations, the *in situ* diffraction approach was attempted for X-ray analysis and performed directly in crystallization droplets within their original screening plates, without any manipulation of the crystals (1b) [146]. These experiments were conducted on beamline BM30 at ESRF (Grenoble, France). Unfortunately, this experiment did not yield diffraction data useful to determine the structure. Although this *in situ* approach is potentially useful in some cases, it has several limitations. Few diffraction images were collected, and by observing the low diffraction quality of *in situ* NadAv3 crystals, it was possible to infer that the crystals grown in ethanol were mostly of intrinsically poor quality (i.e. yielded diffraction only to low resolution), suggesting that the *in situ* method would not prove to be a fruitful approach, and that alternative crystal growth conditions were needed.

Temperature can be a significant variable in biological macromolecule and small molecule crystallization [148] and it has been demonstrated that temperature induced crystallization could be a generally useful technique [149]. As an important parameter, the temperature of crystallization experiments was investigated, from 22°C to 4°C (2b) but the crystals that grew at 4°C did not diffract.

Finally, the ethanol was substituted by less volatile alcohols, such as the 2-methylpropane-1,3-diol (MPD). Remarkably, the use of MPD as precipitant readily yielded crystals, which provided a 2.2Å resolution native data set for NadAv3₂₄₋₁₇₀ firstly, and subsequently for two NadAv3 stabilized mutants. The diffraction pattern of NadAv3₂₄₋₁₇₀ crystals, which can be seen as arrays of dots, is showed in *Figure 2.6*. The sharpness of the spots reflects the quality of the crystal organization; the pattern of spots (reflections) and the relative strength of each spot (intensities) can be used to determine the structure. The resulting reflections from a diffraction experiment arises from a complicated relation of the physics of the incoming photon wave interacting with each ordered lattice point in the analyzed protein crystal.

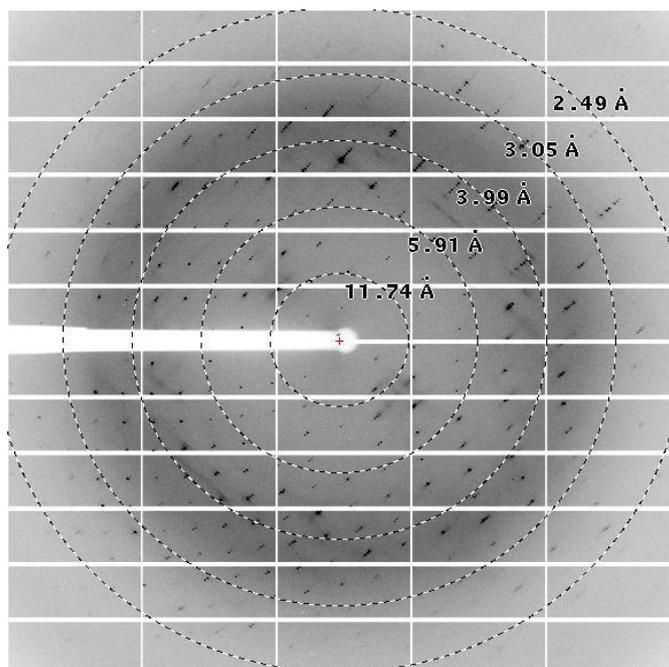


Figure 2.6. Diffraction image of Nadav₃₂₄₋₁₇₀ crystal.

Statistics for data collection are shown in *Table 2.2*

NadAv3 24-170	
Data collection	
Wavelength (Å)	0.977
Beamline	ESRF(ID23-2)
Resolution range (Å)	42.3 - 2.2
Space group	<i>R</i> 3 2 (Space group number 155)
Cell dimensions	
a, b, c (Å)	40.02, 40.02, 762.48
α, β, γ (°)	90, 90, 120
Total reflections	114967
Unique reflections	12055
Multiplicity	9.5 (9)
Completeness (%)	99.3 (96)
Mean I/σ(I)	18.6 (2.1)
Wilson B-factor	64.9
R_{sym} *	5.8 (89.5)
R_{meas} **	6.1 (94.9)

Statistics for the highest-resolution shell are shown in parentheses.

* $R_{\text{sym}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$

** $R_{\text{meas}} =$ redundancy-independent (multiplicity-weighted) R_{merge} as reported from AIMLESS [91].

Table 2.2. Data collection statistics for Nadav₃₂₄₋₁₇₀.

NadAv3 mutagenesis to promote crystallization

As the overall quality of crystals of the C-term truncated NadAv3 constructs remained sub optimal and not suitable for further structural studies, it was decided to explore another method of protein engineering, namely mutagenesis, to mutate residues both in the head and in the stalk domain of NadA. Since an increased thermal stability is correlated with a higher protein tendency to crystallize [150], mutations were designed that might increase the T_m of NadAv3₂₄₋₁₇₀ construct (T_m is $\sim 40^\circ\text{C}$). Note that the T_m of the successfully crystallized NadAv5 is $\sim 60^\circ\text{C}$. A pairwise sequence alignment of NadAv3 vs NadAv5 and multiple sequence alignments of all NadA variants (1 to 5) were inspected in regions where the heptad repeats are not conserved. Analyses of the local environment of the potential mutagenesis sites were performed using a molecular graphics interface (Pymol: www.pymol.org) to evaluate the impact of mutations. Using this approach, nine novel constructs were designed (Figure 2.7).

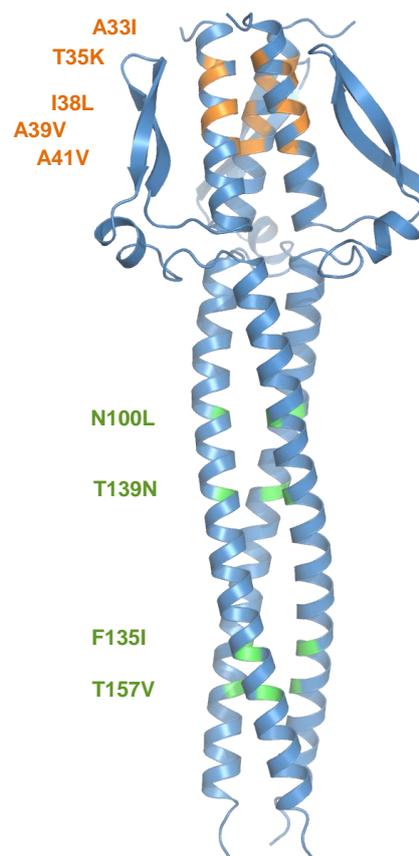


Figure 2.7. Mutagenesis for stabilizing sites in head (orange) and stalk (green) domains of the NadAv5 structure

The aim of mutations was to increase thermal stability by improving the internal hydrophobic packing of the protein, as follows:

a. Coiled coil domain single point mutants:

- i. Asn 100 (v3) vs Leu 99 (v5)
- ii. Thr 139 (v3) vs Asn 106 (v5)
- iii. Phe 153 (v3) vs Ile 120 (v5)
- iv. Thr 157 (v3) vs Val 124 (v5)

The selected residues of NadAv3 were substituted by the equivalent residues of NadAv5, except for the v3 residue Thr 139 substituted by Leu (instead of Asn). Since the Asn is destabilizing at this position of TAA proteins, as shown by Hartmann *et al.* [151], hydrophobic Leu as in mutation Asn 100 (v3) vs Leu 99 (v5) appeared to be more favorable.

b. Head domain single point mutants:

- i. Ala 33 (v3) vs Ile 33 (v5)
- ii. Thr 35 (v3) vs Lys 35 (v5)
- iii. Ile 38 (v3) vs Leu 38 (v5)
- iv. Ala 39 (v3) vs Val 39 (v5)
- v. Ala 41 (v3) vs Val 41 (v5)

The NadAv3₂₄₋₁₇₀ construct mutants were produced by site-direct mutagenesis, were expressed in *E. coli* and purified under identical conditions, as described in *Materials and Methods*. These constructs were screened for increased thermostability using differential scanning calorimetry (DSC). Interestingly, three single point mutations were found to stabilize the NadAv3 protein. The head mutations specifically increased the thermal stability of NadAv3, with the largest effects being induced by the A33I mutation for which it was observed the higher ΔT_m value (*Figure 2.8*).

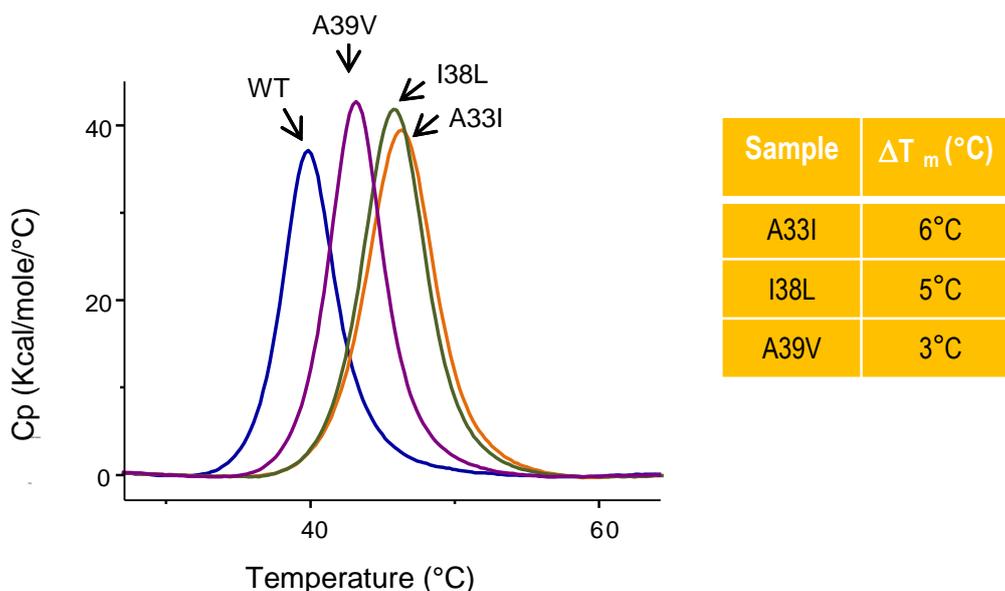


Figure 2.8. Thermostability of NadAv3 stabilized mutants. DSC profiles of wild type NadAv3₂₄₋₁₇₀ protein (blue) and of three single point mutants for which an incremental of melting temperature (T_m) was observed.

After identification of single stabilizing mutations, these were combined in two double mutants and in one triple mutant as shown in Table 2.3:

Construct	Mutation
pET21-CHIS-NadAv3_24-170_A33I+I38L	Head-domain stabilizing mutations
pET21-CHIS-NadAv3_24-170_A33I+A39V	Head-domain stabilizing mutations
pET21-CHIS-NadAv3_24-170_A33I+I38L+A39V	Head-domain stabilizing mutations

Table 2.3. NadAv3 double and triple mutants.

The production of stabilized constructs was scaled-up for production of >10mg quantities for crystallization experiments. As reported in the literature [152, 153], mutagenesis can be an effective tool to aid protein crystallization and may also yield dramatic improvements in crystal quality. Structural integrity of NadAv3 stabilized mutants was tested by the binding of two anti-NadA head domain monoclonal antibodies (mAbs). The two internally available mAbs were used as positive controls in SPR experiments. SPR sensorgrams of each NadAv3 mutant-mAb interactions are shown in Figure 2.9.

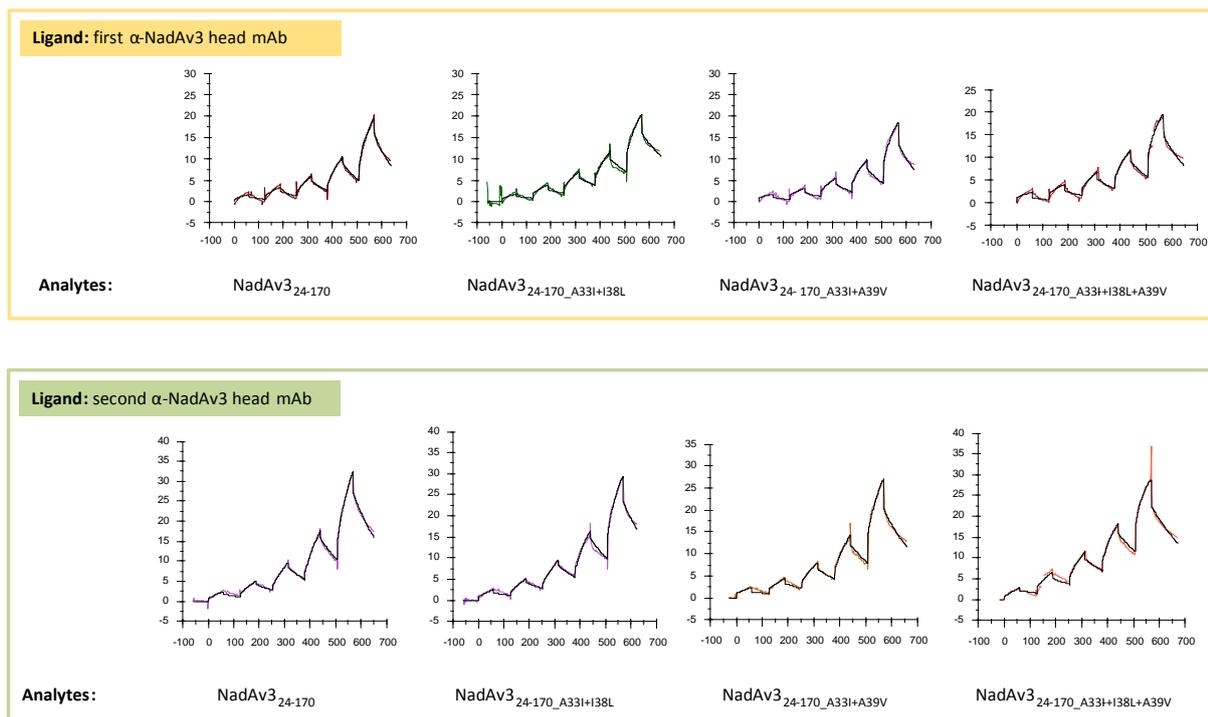


Figure 2.9. SPR sensorgrams for binding of mAbs to NadAv3₂₄₋₁₇₀ and NadAv3 mutant forms.

SPR measurements of ligand binding were performed on the NadAv3₂₄₋₁₇₀ and three NadAv3 mutant constructs (NadAv3_{24-170_A331+I38L}, NadAv3_{24-170_A331+A39V} and NadAv3_{24-170_A331+I38L+A39V}). Each of the two mAbs was captured on a specific sensor chip and tested for binding with each NadAv3 construct that were injected separately in the analyte flow at a range of concentrations.

High-quality NadAv3 reagents for functional studies

Designing several constructs of a single protein enabled the production of large quantities of stable proteins once the length of the coiled coil domain that ensured the trimeric conformation had been determined. The high-quality reagents generated by the structure-based design were validated for functional studies. In particular, the new NadAv3 constructs containing the head and the neck domains (the trimeric NadAv3₂₄₋₁₇₀ and the monomeric NadAv3₂₄₋₈₉) and 'headless' coiled coil constructs were useful for epitope mapping studies. As reported in *Figure 2.10* different collaborations have been carried out in order (i) to define the crucial region of NadAv3 required for its interaction with the human LOX-1 receptor; (ii) to characterize the murine mAb 6E3, able to recognize three cross-protective NadA gene variants and the bactericidal mAb 33E8 specific for group I variants

[138, 154]; (iii) to map the epitopes of three murine mAbs elicited by immunization with MenA and MenW GMMA expressing NadA. All monoclonal antibodies and receptor bound a single epitope.

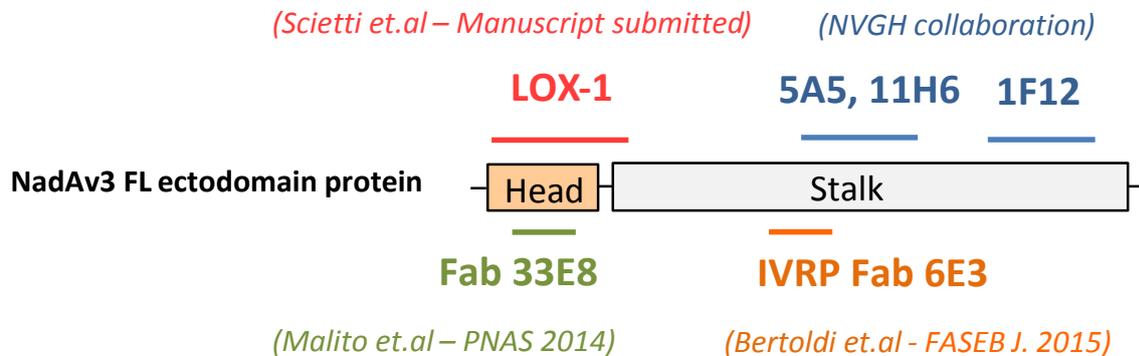


Figure 2.10. Epitope mapping studies on NadAv3 protein. The block cartoon images show the LOX1 site (red) on the NadAv3₂₄₋₁₇₀ region; the Fab 33E8 (green) and the Fab 6E3 sites (orange) on the head and on the central region, respectively; the three murine mAbs sites (blue) on the stalk domain . I would suggest you delete 'IVRP' from the image.

All the previously mentioned epitope mapping experiments were performed on different NadAv3 constructs using the Surface Plasmon Resonance (SPR) technology, except the evaluation of NadAv3/LOX-1 interaction, which was performed by biolayer interferometry (Octet QKe, ForteBio, Pall Instruments) experiments by an internal collaborator. In addition, SPR analysis was used to calculate and compare the kinetics of binding interactions between mAbs and NadAv3 antigen. Epitope mapping experiments using these reagents permitted a deeper understanding of the mechanism of NadA-mediated protection and its role in the interaction with host cells.

Discussion

The long-term goal of this study is to determine the high-resolution structure of NadAv3, in order to provide biological and immunological insights into this vaccine antigen, and potentially to aid the design of an improved vaccine antigen. It was anticipated that this would be a challenging project, since numerous failed efforts had already been reported [138]. Therefore, the immediate goal targeted here was to generate a NadAv3 antigen construct of enhanced thermal stability that could be used as a preferential sample for crystallization trials. The desired protein construct was identified among more than 10 new differently-sized NadAv3 constructs, exploring C-terminal truncations, designed to have three fundamental chemical and structural properties: (1) reduced elongated protein shape, (2) decreased coiled-coil flexibility, (3) increased thermal stability. The recently solved NadAv5 crystal structure [138] was used for the rational design of a new series of NadAv3 constructs. The domains and boundaries in which functional, trimeric and stable recombinant NadAv3 constructs can be produced in large quantities were identified. A comprehensive set of biochemical and biophysical techniques were applied to investigate nine differently-sized v3 constructs, exploring C-terminal truncations. The results revealed that all of the constructs could be expressed and purified with high quality, six of those were not stable and exist in a mixture of monomeric and trimeric states, and only three constructs (NadAv3₂₄₋₁₇₀, NadAv3₂₄₋₁₇₄, and NadAv3₂₄₋₁₈₁) were stable and trimeric. Thus, the shortest stable trimeric construct of the NadAv3 protein was identified: the NadAv3₂₄₋₁₇₀ construct is the shortest construct that is well-folded and trimeric.

It was established that the N-terminal head region alone is insufficient for stable trimerization, which requires additional C-terminal region for stable folding. Hence, it could conclude that to obtain stable NadAv3 trimeric fragments, the presence of an initial coiled-coil region made at least of 80 residues is necessary. This conclusion was further strengthened by the increased stability of constructs containing higher number of residues of coiled coil domain compared with those lacking it, as shown by the results of the thermostability shift assay (*Table 2.1*).

The NadAv3₂₄₋₁₇₀ construct displayed a moderate stability ($T_m \sim 40^\circ\text{C}$, *Figure 2.5*), it is greatly overexpressed in bacteria, and easy to purify, but its immediate use as a crystallization target reproducibly generated low quality crystals. A total number of 100 NadAv3₂₄₋₁₇₀ crystals were screened for diffraction, but this screen did not yield any complete datasets. To improve crystal diffraction quality, stabilized variants of NadAv3₂₄₋₁₇₀ were engineered. Using the sequence alignments of NadA variants and the known v5 structure, molecular graphics tools were used to identify the local environment of potential mutagenesis sites likely to contribute to increased stability. Hydrophobic residues in position *d* of the

heptad repeat in NadAv3 were mutated to Asn, in order to have Asn at position *d* layers in the NadA Coiled-Coil, as in NadAv5 [138]. Regions where the heptad repeats were not conserved were also inspected. For experimental testing, five single head mutations and four single coiled coil mutations were identified. Site-directed mutagenesis of the *nadA* gene was used to generate the designed mutants described above. Like NadAv3 wt protein, the mutated proteins were over-expressed in *E. coli*, and purified to >95% homogeneity using Ni-affinity chromatography. All mutants were expressed at high levels and were soluble. The SEC-MALLS profile and the calculated MW of all the mutants are identical to those of NadA wt, suggesting that the mutations did not affect the tertiary structure.

Protein stability was assessed by following the mutation-induced change of T_m in DSC experiments. Two additional two-point and three-point mutants that are combinations of successfully stabilizing single mutations from the initial experiment were also produced and selected for experimental characterization to assess increase in stabilization and used in crystallization trials. SPR measurements of mAbs binding were performed on the NadAv3₂₄₋₁₇₀ and three Nadav3 mutant constructs. The SPR responses (*Figure 2.9*) were used to calculate the binding affinities to assess the structural integrity of NadAv3 mutants and to compare the effects of each mutation on mAb binding specificity. Binding affinities measured for each mAb with the Nadav3₂₄₋₁₇₀ wt were very similar to those obtained from the binding assay of NadAv3 mutants. The similar behavior of NadAv3 mutants observed in the SPR data provided validation of the trimeric form and folding of the NadA mutants. A second approach involved appending the foldon or the GCN4 domain to the C terminus of the NadAv3₂₄₋₁₇₀. These variants were tested for increased thermal stability in DSC experiments (*Figure 2.5*). Both of the trimerization domains did not have a dramatic effect on the thermal stability of the protein. Nevertheless, the NadA engineering constructs were used for crystallization experiments and, unfortunately, no strongly-diffracting crystals were obtained.

The initial crystals of NadAv3₂₄₋₁₇₀ did not diffract. Subsequent optimization resulted in crystals of NadAv3₂₄₋₁₇₀ that diffracted to maximum resolution of 6 Å. A second crystal form obtained using a different precipitant agent (MPD) provided a 2.3 Å resolution native data set. In addition to the NadAv3₂₄₋₁₇₀ construct, two of the stabilized NadAv3 mutants (NadAv3_{24-170_A33I+I38L} and NadAv3_{24-170_A33I+I38L+A39V}) were successfully crystallized. A key step in the strategy was the substitution of the ethanol with a less volatile alcohol. This probably promoted a more ordered arrangement of the molecule in the crystals that diffracted between 2.3 and 2.8 Å for the three constructs. In order to extract structural information from X-ray diffraction data, the recorded intensities must be supplemented by additional phase information. This introduces the crystallographic phase problem, which can be faced by a number of approaches [155]. One method is the molecular replacement (MR), which is based on the availability of an accurate homologous model of the crystallized

macromolecules that can be correctly placed in the asymmetric unit to obtain the initial phase [156-158]. Although the solved NadAv5 crystal structure shares ~50% sequence identity to NadAv3 and represents a valid model, solving a coiled-coil protein structure can be very difficult with MR techniques [159, 160]. This is due to several factors that work in combination to hamper the phasing of these proteins [135, 161-163]. Specifically, the presence of translational non-crystallographic symmetry (tNCS) in the investigated NadAv3 crystals structure is an additional complicating factor for structure determination. In tNCS, two or more copies of a molecule are arranged in the same orientation in the asymmetric part of the unit cell, or of internal symmetry of the individual molecule [164]. This is common for coiled-coil protein architecture [160]. Solution by MR in the presence of tNCS is not fully characterized and automated, and requires additional efforts.

In addition to structural studies, the C-terminal NadAv3 truncated constructs represented a powerful set of high quality reagents (highly homogenous and pure samples, stable proteins and trimeric conformation as the full length proteins) and were subsequently useful for epitope mapping studies. A functional investigation of the LOX-1 binding site and of different mAb epitopes was performed in SPR experiments to fully characterize the interactions with the NadAv3 antigen. These well-characterized constructs can be used to facilitate validation of the reagents across various assay platforms.

Concluding Remarks

Methods for systematic identification of epitopes are needed for detailed characterization of antigens. The functional characterization of an antigen and its characterization by epitope mapping experiments are important goals of structural vaccinology, in order to selectively present the conserved immunogenic determinants of complex and variable antigens. Using high resolution structures, antigens can be designed to be more efficiently produced and stably stored than native molecules. The biophysical, biochemical, and genetic engineering technologies are essential tools for the structural vaccinology: firstly, it is important to study the native molecular architecture of an antigen and its neutralization determinants; then this knowledge can be used to modify the molecule and engineer new immunogens that are optimally designed for inclusion in protective vaccines. An efficient strategy to generate soluble, well-expressed, trimeric and stable NadA proteins is described. In addition, to enhance crystallization propensity stability of NadAv3, engineered stabilized NadAv3 constructs, that could be characterized antigenically and tested for immunogenicity in mice, were generated. Finally, structural information obtained from the crystal structures of NadAv3₂₄₋₁₇₀ and of the two NadAv3 stabilized constructs could be used to optimize protein and engineered NadAv3 antigens to make it more immunogenic. A new

generation of vaccines, including optimized antigens, could be even more broadly protective, practical, and affordable than those available today. The interplay among biophysical and structural characterization of vaccine antigen, design of an optimized vaccine antigen and its immunogenicity studies provides the basis for further optimization of meningococcal vaccine antigen.

Acknowledgements

Writing this note of thanks is the finishing touch on my thesis. It has been a period of intense learning for me and I would like to reflect on the people who made this thesis possible.

I would first like to express my deepest sense of gratitude to my supervisor Dr. Matthew James Bottomley for your valuable guidance. You definitely provided me with the capabilities that I needed to choose the right direction and successfully complete my PhD.

I would like to thank my day-by-day mentor, Dr. Enrico Malito, for your continuous encouragement throughout the course of PhD. I am also thankful to you for reading my reports, commenting on my views and helping me understand and enrich my ideas.

Acknowledgements are gratefully made here to University of Bologna that provided me the PhD Fellowship and Prof. Vincenzo Scarlato for allowing this opportunity.

I would also like to thank my colleagues from Protein Biochemistry Function for your wonderful collaboration. I would particularly like to single out Dr. ssa Paola Lo Surdo, who has known me since my arrival at GSK Vaccines as a student five years ago, for your excellent cooperation and for all of the opportunities I was given to conduct my research.

In addition, I would like to thank colleagues from the Structure-Based Antigen Design Lab for your numerous discussions and lectures on related topics that helped me improve my knowledge in the area. I am thankful to Dr. ssa Ilaria Ferlenghi for your constructive criticism and helpful comments.

I want to thank my beloved parents, Antonio & Enza, and my brother Andrea for your wise counsels. You are always there for me. I am heartily thankful to my boyfriend Valerio for your love and continuous support.

Finally, there are my friends, Marti & Ila. We were not only able to support each other by deliberating over our problems and findings, but also happily by talking about things other than just our researches.

It has been an honor for me to meet and work with all of you and participate to the research life of this Centre. Thank you very much, everyone!

References

1. Rappuoli, R., *Reverse vaccinology, a genome-based approach to vaccine development*. Vaccine, 2001.
2. Liljeroos, L., et al., *Structural and Computational Biology in the Design of Immunogenic Vaccine Antigens*. J Immunol Res, 2015. **2015**: p. 156241.
3. Vytvytska, O., et al., *Identification of vaccine candidate antigens of Staphylococcus aureus by serological proteome analysis*. Proteomics, 2002. **2**(5): p. 580-90.
4. Couto, N., et al., *Identification of vaccine candidate antigens of Staphylococcus pseudintermedius by whole proteome characterization and serological proteomic analyses*. J Proteomics, 2015.
5. Bittaye, M. and P. Cash, *Streptococcus pneumoniae proteomics: determinants of pathogenesis and vaccine development*. Expert Rev Proteomics, 2015. **12**(6): p. 607-21.
6. Cordwell, S.J., *Technologies for bacterial surface proteomics*. Curr Opin Microbiol, 2006. **9**(3): p. 320-9.
7. Solis, N. and S.J. Cordwell, *Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins*. Proteomics, 2011. **11**(15): p. 3169-89.
8. Grandi, G., *Antibacterial vaccine design using genomics and proteomics*. Trends Biotechnol, 2001. **19**(5): p. 181-8.
9. Dapat, C. and H. Oshitani, *Novel insights into human respiratory syncytial virus-host factor interactions through integrated proteomics and transcriptomics analysis*. Expert Rev Anti Infect Ther, 2016.
10. Delany, I., R. Rappuoli, and E. De Gregorio, *Vaccines for the 21st century*. EMBO molecular medicine, 2014.
11. Dormitzer, P.R., J.B. Ulmer, and R. Rappuoli, *Structure-based antigen design: a strategy for next generation vaccines*. Trends Biotechnol, 2008. **26**(12): p. 659-67.
12. Vogel, C., et al., *Structure, function and evolution of multidomain proteins*. Curr Opin Struct Biol, 2004. **14**(2): p. 208-16.
13. Scarselli, M., et al., *Rational design of a meningococcal antigen inducing broad protective immunity*. Sci Transl Med, 2011. **3**(91): p. 91ra62.
14. Caugant, D.A., G. Tzanakaki, and P. Kriz, *Lessons from meningococcal carriage studies*. FEMS microbiology reviews, 2007. **31**(1): p. 52-63.
15. Stephens, D.S., *Conquering the meningococcus*. FEMS microbiology reviews, 2007. **31**(1): p. 3-14.
16. Gasparini, R. and D. Panatto, *Meningococcal glycoconjugate vaccines*. Human vaccines, 2011. **7**(2).

17. Serruto, D., et al., *Neisseria meningitidis* GNA2132, a heparin-binding protein that induces protective immunity in humans. *Proc Natl Acad Sci U S A*, 2010. **107**(8): p. 3770-5.
18. Massignani, V., et al., *Vaccination against Neisseria meningitidis using three variants of the lipoprotein GNA1870*. *The Journal of experimental medicine*, 2003. **197**(6): p. 789-799.
19. Seib, K.L., et al., *Factor H-binding protein is important for meningococcal survival in human whole blood and serum and in the presence of the antimicrobial peptide LL-37*. *Infection and immunity*, 2009. **77**(1): p. 292-299.
20. Capecchi, B., et al., *Neisseria meningitidis* NadA is a new invasin which promotes bacterial adhesion to and penetration into human epithelial cells. *Molecular Microbiology*, 2005. **55**(3): p. 687-698.
21. Comanducci, M., et al., *NadA, a novel vaccine candidate of Neisseria meningitidis*. *The Journal of experimental medicine*, 2002. **195**(11): p. 1445-1454.
22. Giuliani, M.M., et al., *A universal vaccine for serogroup B meningococcus*. *Proceedings of the National Academy of Sciences of the United States of America*, 2006. **103**(29): p. 10834-10839.
23. Serruto, D., et al., *The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens*. *Vaccine*, 2012. **30 Suppl 2**: p. B87-97.
24. O'Ryan, M., et al., *A multi-component meningococcal serogroup B vaccine (4CMenB): the clinical development program*. *Drugs*, 2014. **74**(1): p. 15-30.
25. Nassif, X., et al., *How do extracellular pathogens cross the blood-brain barrier?* *Trends in microbiology*, 2002. **10**(5): p. 227-232.
26. Virji, M., *Pathogenic neisseriae: surface modulation, pathogenesis and infection control*. *Nature Reviews Microbiology*, 2009. **7**(4): p. 274-286.
27. Coureuil, M., et al., *Meningococcal type IV pili recruit the polarity complex to cross the brain endothelium*. *Science (New York, NY)*, 2009. **325**(5936): p. 83-87.
28. Gray-Owen, S.D. and R.S. Blumberg, *CEACAM1: contact-dependent control of immunity*. *Nat Rev Immunol*, 2006. **6**(6): p. 433-46.
29. Stephens, D.S. and Z.A. McGee, *Attachment of Neisseria meningitidis to human mucosal surfaces: influence of pili and type of receptor cell*. *J Infect Dis*, 1981. **143**(4): p. 525-32.
30. Virji, M., et al., *The role of pili in the interactions of pathogenic Neisseria with cultured human endothelial cells*. *Mol Microbiol*, 1991. **5**(8): p. 1831-41.
31. de Vries, F.P., et al., *Neisseria meningitidis* producing the Opc adhesin binds epithelial cell proteoglycan receptors. *Mol Microbiol*, 1998. **27**(6): p. 1203-12.
32. Moore, J., et al., *Recognition of saccharides by the OpcA, OpaD, and OpaB outer membrane proteins from Neisseria meningitidis*. *J Biol Chem*, 2005. **280**(36): p. 31489-97.

33. Scarselli, M., et al., *Neisseria meningitidis* NhhA is a multifunctional trimeric autotransporter adhesin. *Molecular Microbiology*, 2006. **61**(3): p. 631-644.
34. Serruto, D., et al., *Neisseria meningitidis* App, a new adhesin with autocatalytic serine protease activity. *Mol Microbiol*, 2003. **48**(2): p. 323-34.
35. Turner, D.P., et al., *Characterization of MspA, an immunogenic autotransporter protein that mediates adhesion to epithelial and endothelial cells in Neisseria meningitidis*. *Infect Immun*, 2006. **74**(5): p. 2957-64.
36. van Putten, J.P., T.D. Duensing, and R.L. Cole, *Entry of OpaA+ gonococci into HEp-2 cells requires concerted action of glycosaminoglycans, fibronectin and integrin receptors*. *Mol Microbiol*, 1998. **29**(1): p. 369-79.
37. Hill, D.J., et al., *Cellular and molecular biology of Neisseria meningitidis colonization and invasive disease*. *Clin Sci (Lond)*, 2010. **118**(9): p. 547-64.
38. Dietrich, G., et al., *Transcriptome analysis of Neisseria meningitidis during infection*. *J Bacteriol*, 2003. **185**(1): p. 155-64.
39. Nichols, C.E., et al., *The structure of NMB1585, a MarR-family regulator from Neisseria meningitidis*. *Acta Crystallographica F*, 2009. **65**(Pt 3): p. 204-209.
40. Metruccio, M.M.E., et al., *A novel phase variation mechanism in the meningococcus driven by a ligand-responsive repressor and differential spacing of distal promoter elements*. *PLoS Pathogens*, 2009. **5**(12): p. e1000710.
41. Schielke, S., et al., *Expression of the meningococcal adhesin NadA is controlled by a transcriptional regulator of the MarR family*. *Mol Microbiol*, 2009. **72**(4): p. 1054-67.
42. Delany, I., R. Rappuoli, and V. Scarlato, *Fur functions as an activator and as a repressor of putative virulence genes in Neisseria meningitidis*. *Mol Microbiol*, 2004. **52**(4): p. 1081-90.
43. Delany, I., et al., *Effect of Neisseria meningitidis fur mutations on global control of gene transcription*. *J Bacteriol*, 2006. **188**(7): p. 2483-92.
44. Grifantini, R., et al., *Identification of iron-activated and -repressed Fur-dependent genes by transcriptome analysis of Neisseria meningitidis group B*. *Proc Natl Acad Sci U S A*, 2003. **100**(16): p. 9542-7.
45. Mellin, J.R., et al., *A novel fur- and iron-regulated small RNA, NrrF, is required for indirect fur-mediated regulation of the sdhA and sdhC genes in Neisseria meningitidis*. *J Bacteriol*, 2007. **189**(10): p. 3686-94.
46. Stork, M., et al., *An outer membrane receptor of Neisseria meningitidis involved in zinc acquisition with vaccine potential*. *PLoS Pathog*, 2010. **6**: p. e1000969.
47. Bartolini, E., et al., *Role of FNR and FNR-regulated, sugar fermentation genes in Neisseria meningitidis infection*. *Mol Microbiol*, 2006. **60**(4): p. 963-72.
48. Edwards, J.L., *Neisseria gonorrhoeae survival during primary human cervical epithelial cell infection requires nitric oxide and is augmented by progesterone*. *Infect Immun*, 2010. **78**(3): p. 1202-13.

49. Rock, J.D., et al., *The pathogen Neisseria meningitidis requires oxygen, but supplements growth by denitrification. Nitrite, nitric oxide and oxygen control respiratory flux at genetic and metabolic levels.* Mol Microbiol, 2005. **58**(3): p. 800-9.
50. Rock, J.D., et al., *Regulation of denitrification genes in Neisseria meningitidis by nitric oxide and the repressor NsrR.* J Bacteriol, 2007. **189**(3): p. 1138-44.
51. Heurlier, K., et al., *The nitric oxide (NO)-sensing repressor NsrR of Neisseria meningitidis has a compact regulon of genes involved in NO synthesis and detoxification.* J Bacteriol, 2008. **190**(7): p. 2488-95.
52. Deghmane, A.E., et al., *Intimate adhesion of Neisseria meningitidis to human epithelial cells is under the control of the crgA gene, a novel LysR-type transcriptional regulator.* Embo j, 2000. **19**(5): p. 1068-78.
53. Ieva, R., et al., *CrgA is an inducible LysR-type regulator of Neisseria meningitidis, acting both as a repressor and as an activator of gene transcription.* J Bacteriol, 2005. **187**(10): p. 3421-30.
54. Ren, J., et al., *The structure and transcriptional analysis of a global regulator from Neisseria meningitidis.* J Biol Chem, 2007. **282**(19): p. 14655-64.
55. Davidsen, T. and T. Tonjum, *Meningococcal genome dynamics.* Nat Rev Microbiol, 2006. **4**(1): p. 11-22.
56. Sarkari, J., et al., *Variable expression of the Opc outer membrane protein in Neisseria meningitidis is caused by size variation of a promoter containing poly-cytidine.* Mol Microbiol, 1994. **13**(2): p. 207-17.
57. van der Ende, A., et al., *Variable expression of class 1 outer membrane protein in Neisseria meningitidis is caused by variation in the spacing between the -10 and -35 regions of the promoter.* J Bacteriol, 1995. **177**(9): p. 2475-80.
58. Hammerschmidt, S., et al., *Modulation of cell surface sialic acid expression in Neisseria meningitidis via a transposable genetic element.* Embo j, 1996. **15**(1): p. 192-8.
59. Weiser, J.N., et al., *The phosphorylcholine epitope undergoes phase variation on a 43-kilodalton protein in Pseudomonas aeruginosa and on pili of Neisseria meningitidis and Neisseria gonorrhoeae.* Infect Immun, 1998. **66**(9): p. 4263-7.
60. Moxon, E.R., R.E. Lenski, and P.B. Rainey, *Adaptive evolution of highly mutable loci in pathogenic bacteria.* Perspect Biol Med, 1998. **42**(1): p. 154-5.
61. Feil, E.J. and B.G. Spratt, *Recombination and the population structures of bacterial pathogens.* Annu Rev Microbiol, 2001. **55**: p. 561-90.
62. Cloward, J.M. and W.M. Shafer, *MtrR control of a transcriptional regulatory pathway in Neisseria meningitidis that influences expression of a gene (nadA) encoding a vaccine candidate.* PLoS One, 2013. **8**(2): p. e56097.
63. Metruccio, M.M., et al., *A novel phase variation mechanism in the meningococcus driven by a ligand-responsive repressor and differential spacing of distal promoter elements.* PLoS Pathog, 2009. **5**(12): p. e1000710.

64. Fagnocchi, L., et al., *In the NadR Regulon, Adhesins and Diverse Meningococcal Functions Are Regulated in Response to Signals in Human Saliva*. J Bacteriol, 2012. **194**(2): p. 460-74.
65. Jamet, A., et al., *A new family of secreted toxins in pathogenic Neisseria species*. PLoS Pathog, 2015. **11**(1): p. e1004592.
66. Lamelas, A., et al., *Emergence of a new epidemic Neisseria meningitidis serogroup A Clone in the African meningitis belt: high-resolution picture of genomic changes that mediate immune evasion*. MBio, 2014. **5**(5): p. e01974-14.
67. Takahama, U., T. Oniki, and H. Murata, *The presence of 4-hydroxyphenylacetic acid in human saliva and the possibility of its nitration by salivary nitrite in the stomach*. FEBS Lett, 2002. **518**(1-3): p. 116-8.
68. Fagnocchi, L., et al., *Transcriptional regulation of the nadA gene in Neisseria meningitidis impacts the prediction of coverage of a multicomponent meningococcal serogroup B vaccine*. Infect Immun, 2013. **81**(2): p. 560-9.
69. Alekshun, M.N. and S.B. Levy, *The mar regulon: multiple resistance to antibiotics and other toxic chemicals*. Trends Microbiol, 1999. **7**(10): p. 410-3.
70. Grove, A., *MarR family transcription factors*. Curr Biol, 2013. **23**(4): p. R142-3.
71. Ellison, D.W. and V.L. Miller, *Regulation of virulence by members of the MarR/SlyA family*. Curr Opin Microbiol, 2006. **9**(2): p. 153-9.
72. Perera, I.C. and A. Grove, *Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators*. J Mol Cell Biol, 2010. **2**(5): p. 243-54.
73. Hong, M., et al., *Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family*. Molecular cell, 2005. **20**(1): p. 131-141.
74. Saridakis, V., et al., *Structural insight on the mechanism of regulation of the MarR family of proteins: high-resolution crystal structure of a transcriptional repressor from Methanobacterium thermoautotrophicum*. J Mol Biol, 2008. **377**(3): p. 655-67.
75. Kumarevel, T., et al., *ST1710-DNA complex crystal structure reveals the DNA binding mechanism of the MarR family of regulators*. Nucleic Acids Res, 2009. **37**(14): p. 4723-35.
76. Chang, Y.M., et al., *Structural study of TcaR and its complexes with multiple antibiotics from Staphylococcus epidermidis*. Proc Natl Acad Sci U S A, 2010. **107**(19): p. 8617-22.
77. Chang, Y.M., et al., *Structural analysis of the antibiotic-recognition mechanism of MarR proteins*. Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 6): p. 1138-49.
78. McGuinness, B.T., et al., *Point mutation in meningococcal por A gene associated with increased endemic disease*. Lancet, 1991. **337**(8740): p. 514-7.
79. Fagnocchi, L., et al., *In the NadR Regulon, Adhesins and Diverse Meningococcal Functions Are Regulated in Response to Signals in Human Saliva*. The Journal of Bacteriology, 2012. **194**(2): p. 460-474.

80. Lo Surdo, P., et al., *Mechanistic implications for LDL receptor degradation from the PCSK9/LDLR structure at neutral pH*. EMBO Rep, 2011. **12**(12): p. 1300-5.
81. Mistrik, P., F. Moreau, and J.M. Allen, *BiaCore analysis of leptin-leptin receptor interaction: evidence for 1:1 stoichiometry*. Anal Biochem, 2004. **327**(2): p. 271-7.
82. Kabsch, W., *Xds*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
83. CCP4, *The CCP4 suite: programs for protein crystallography*. Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 5): p. 760-3.
84. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
85. McCoy, A.J., et al., *Phaser crystallographic software*. J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
86. Emsley, P., et al., *Features and development of Coot*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
87. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
88. Morin, A., et al., *Collaboration gets the most out of software*. Elife, 2013. **2**: p. e01456.
89. Celej, M.S., G.G. Montich, and G.D. Fidelio, *Protein stability induced by ligand binding correlates with changes in protein flexibility*. Protein Sci, 2003. **12**(7): p. 1496-506.
90. Martin, R.G. and J.L. Rosner, *Binding of purified multiple antibiotic-resistance repressor protein (MarR) to mar operator sequences*. Proc Natl Acad Sci U S A, 1995. **92**(12): p. 5456-60.
91. Evans, P., *Scaling and assessment of data quality*. Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 72-82.
92. Nogales, E. and S.H. Scheres, *Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity*. Mol Cell, 2015. **58**(4): p. 677-89.
93. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state*. J Mol Biol, 2007. **372**(3): p. 774-97.
94. Brier, S., et al., *Structural Insight into the Mechanism of DNA-Binding Attenuation of the Neisserial Adhesin Repressor NadR by the Small Natural Ligand 4-Hydroxyphenylacetic Acid*. Biochemistry, 2012. **51**(34): p. 6738-52.
95. Liang, J., et al., *Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape*. Proteins, 1998. **33**(1): p. 1-17.
96. Hong, M., et al., *Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family*. Mol Cell, 2005. **20**(1): p. 131-41.

97. Donnelly, J., et al., *Qualitative and quantitative assessment of meningococcal antigens to evaluate the potential strain coverage of protein-based vaccines*. Proc Natl Acad Sci U S A, 2010. **107**(45): p. 19490-5.
98. Mani, A.R., et al., *The metabolism and dechlorination of chlorotyrosine in vivo*. J Biol Chem, 2007. **282**(40): p. 29114-21.
99. Wilkinson, S.P. and A. Grove, *Negative cooperativity of uric acid binding to the transcriptional regulator HucR from Deinococcus radiodurans*. J Mol Biol, 2005. **350**(4): p. 617-30.
100. Chang, Y.-M., et al., *Structural study of TcaR and its complexes with multiple antibiotics from Staphylococcus epidermidis*. Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(19): p. 8617-8622.
101. Nichols, C.E., et al., *The structure of NMB1585, a MarR-family regulator from Neisseria meningitidis*. Acta Crystallogr Sect F Struct Biol Cryst Commun, 2009. **65**(Pt 3): p. 204-9.
102. Lim, D., K. Poole, and N.C. Strynadka, *Crystal structure of the MexR repressor of the mexRAB-oprM multidrug efflux operon of Pseudomonas aeruginosa*. J Biol Chem, 2002. **277**(32): p. 29253-9.
103. Mariotti, P., et al., *Structural and functional characterization of the Staphylococcus aureus virulence factor and vaccine candidate FhuD2*. Biochem J, 2013. **449**(3): p. 683-93.
104. Podkowa, K.J., et al., *Crystal and solution structure analysis of FhuD2 from Staphylococcus aureus in multiple unliganded conformations and bound to ferrioxamine-B*. Biochemistry, 2014. **53**(12): p. 2017-31.
105. Changeux, J.P. and S. Edelman, *Conformational selection or induced fit? 50 years of debate resolved*. F1000 Biol Rep, 2011. **3**: p. 19.
106. Chang, Y.-M., et al., *TcaR-ssDNA complex crystal structure reveals new DNA binding mechanism of the MarR family proteins*. Nucleic Acids Research, 2014.
107. Pabo, C.O. and R.T. Sauer, *Transcription factors: structural families and principles of DNA recognition*. Annu Rev Biochem, 1992. **61**: p. 1053-95.
108. Garvie, C.W. and C. Wolberger, *Recognition of specific DNA sequences*. Mol Cell, 2001. **8**(5): p. 937-46.
109. Boehr, D.D., R. Nussinov, and P.E. Wright, *The role of dynamic conformational ensembles in biomolecular recognition*. Nat Chem Biol, 2009. **5**(11): p. 789-96.
110. Spronk, C.A., et al., *The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator*. Structure, 1999. **7**(12): p. 1483-92.
111. Agback, P., et al., *Architecture of nonspecific protein-DNA interactions in the Sso7d-DNA complex*. Nat Struct Biol, 1998. **5**(7): p. 579-84.
112. Kopke Salinas, R., et al., *Altered specificity in DNA binding by the lac repressor: a mutant lac headpiece that mimics the gal repressor*. ChemBiochem, 2005. **6**(9): p. 1628-37.

113. Wojciak, J.M., K.M. Connolly, and R.T. Clubb, *NMR structure of the Tn916 integrase-DNA complex*. Nat Struct Biol, 1999. **6**(4): p. 366-73.
114. Bizien, T., et al., *A Brief Survey of State-of-the-Art BioSAXS*. Protein Pept Lett, 2016.
115. Boldon, L., F. Laliberte, and L. Liu, *Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application*. Nano Rev, 2015. **6**: p. 25661.
116. Kikhney, A.G. and D.I. Svergun, *A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins*. FEBS Lett, 2015. **589**(19 Pt A): p. 2570-7.
117. Putnam, C.D., et al., *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution*. Q Rev Biophys, 2007. **40**(3): p. 191-285.
118. Carroni, M. and H.R. Saibil, *Cryo electron microscopy to determine the structure of macromolecular complexes*. Methods, 2015.
119. Bai, X.C., G. McMullan, and S.H. Scheres, *How cryo-EM is revolutionizing structural biology*. Trends Biochem Sci, 2015. **40**(1): p. 49-57.
120. Craig, L., et al., *Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions*. Mol Cell, 2006. **23**(5): p. 651-62.
121. Peng, M., D. Cascio, and P.F. Egea, *Crystal structure and solution characterization of the thioredoxin-2 from Plasmodium falciparum, a constituent of an essential parasitic protein export complex*. Biochem Biophys Res Commun, 2015. **456**(1): p. 403-9.
122. Tropsha, A. and J. Bajorath, *Computational Methods for Drug Discovery and Design*. J Med Chem, 2016. **59**(1): p. 1.
123. Hughes, J.P., et al., *Principles of early drug discovery*. Br J Pharmacol, 2011. **162**(6): p. 1239-49.
124. Fox, S., et al., *High-throughput screening: update on practices and success*. J Biomol Screen, 2006. **11**(7): p. 864-9.
125. Boppana, K., et al., *Knowledge based identification of MAO-B selective inhibitors using pharmacophore and structure based virtual screening models*. Eur J Med Chem, 2009. **44**(9): p. 3584-90.
126. Tavano, R., et al., *Mapping of the Neisseria meningitidis NadA cell-binding site: relevance of predicted {alpha}-helices in the NH2-terminal and dimeric coiled-coil regions*. Journal of Bacteriology, 2011. **193**(1): p. 107-115.
127. Cecchini, P., et al., *The Soluble Recombinant Neisseria meningitidis Adhesin NadA(Δ 351-405) Stimulates Human Monocytes by Binding to Extracellular Hsp90*. PLoS ONE, 2011. **6**(9): p. e25089.

128. Nagele, V., et al., *Neisseria meningitidis Adhesin NadA Targets 1 Integrins: FUNCTIONAL SIMILARITY TO YERSINIA INVASIN*. The Journal of biological chemistry, 2011. **286**(23): p. 20536-20546.
129. Campbell, L.A., et al., *Chlamydia pneumoniae binds to the lectin-like oxidized LDL receptor for infection of endothelial cells*. Microbes Infect, 2012. **14**(1): p. 43-9.
130. Shimaoka, T., et al., *LOX-1 supports adhesion of Gram-positive and Gram-negative bacteria*. J Immunol, 2001. **166**(8): p. 5108-14.
131. Honjo, M., et al., *Lectin-like oxidized LDL receptor-1 is a cell-adhesion molecule involved in endotoxin-induced inflammation*. Proc Natl Acad Sci U S A, 2003. **100**(3): p. 1274-9.
132. Wu, Z., et al., *LOX-1 deletion improves neutrophil responses, enhances bacterial clearance, and reduces lung injury in a murine polymicrobial sepsis model*. Infect Immun, 2011. **79**(7): p. 2865-70.
133. Linke, D., et al., *Trimeric autotransporter adhesins: variable structure, common function*. Trends in microbiology, 2006. **14**(6): p. 264-270.
134. Crick, F., *The packing of-helices: simple coiled-coils*. Acta Crystallographica, 1953.
135. Lupas, A.N. and M. Gruber, *The structure of alpha-helical coiled coils*. Adv Protein Chem, 2005. **70**: p. 37-78.
136. Lupas, A., *Coiled coils: new structures and new functions*. Trends in biochemical sciences, 1996. **21**(10): p. 375-382.
137. Henderson, I.R., et al., *Type V protein secretion pathway: the autotransporter story*. Microbiol Mol Biol Rev, 2004. **68**(4): p. 692-744.
138. Malito, E., et al., *Structure of the meningococcal vaccine antigen NadA and epitope mapping of a bactericidal antibody*. Proc Natl Acad Sci U S A, 2014. **111**(48): p. 17128-33.
139. Bambini, S., et al., *NadA diversity and carriage in Neisseria meningitidis*. Infection and immunity, 2004. **72**(7): p. 4217-4223.
140. Bambini, S., et al., *Neisseria adhesin a variation and revised nomenclature scheme*. Clinical and vaccine immunology : CVI, 2014. **21**(7): p. 966-971.
141. Lucidarme, J., et al., *Characterization of fHbp, nhba (gna2132), nadA, porA, sequence type (ST), and genomic presence of IS1301 in group B meningococcal ST269 clonal complex isolates from England and Wales*. Journal of clinical microbiology, 2009. **47**(11): p. 3577-3585.
142. Lucidarme, J., et al., *Characterization of fHbp, nhba (gna2132), nadA, porA, and sequence type in group B meningococcal case isolates collected in England and Wales during January 2008 and potential coverage of an investigational group B meningococcal vaccine*. Clinical and vaccine immunology : CVI, 2010. **17**(6): p. 919-929.
143. Magagnoli, C., et al., *Structural organization of NadADelta(351-405), a recombinant MenB vaccine component, by its physico-chemical characterization at drug substance level*. Vaccine, 2009. **27**(15): p. 2156-70.

144. Pace, D. and A.J. Pollard, *Meningococcal disease: clinical presentation and sequelae*. Vaccine, 2012. **30 Suppl 2**: p. B3-9.
145. Klock, H.E. and S.A. Lesley, *The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis*. 2009, METHODS IN MOLECULAR BIOLOGY-CLIFTON THEN TOTOWA-. p. 91-103.
146. Hernandez Alvarez, B., et al., *A new expression system for protein crystallization using trimeric coiled-coil adaptors*. Protein Eng Des Sel, 2008. **21**(1): p. 11-8.
147. McLellan, J.S., et al., *Structure of RSV Fusion Glycoprotein Trimer Bound to a Prefusion-Specific Neutralizing Antibody*. Science (New York, NY), 2013.
148. McPherson, A., *Current approaches to macromolecular crystallization*. Eur J Biochem, 1990. **189**(1): p. 1-23.
149. Christopher, G.K., A.G. Phipps, and R.J. Gray, *Temperature-dependent solubility of selected proteins*. Journal of Crystal Growth, 1998. **191**(4): p. 820-826.
150. Dupeux, F., et al., *A thermal stability assay can help to estimate the crystallization likelihood of biological samples*. Acta Crystallographica D, 2011.
151. Hartmann, M.D., et al., *A coiled-coil motif that sequesters ions to the hydrophobic core*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(40): p. 16950-16955.
152. Mateja, A., et al., *The impact of GluAla and GluAsp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 Å resolution*. Acta Crystallographica D, 2002. **58**(12): p. 1983-1991.
153. Derewenda, Z.S., *Rational protein crystallization by mutational surface engineering*. Structure (London, England), 2004. **12**(4): p. 529-535.
154. Bertoldi, I., et al., *Exploiting chimeric human antibodies to characterize a protective epitope of Neisseria adhesin A, one of the Bexsero vaccine components*. Faseb j, 2016. **30**(1): p. 93-101.
155. Taylor, G., *The phase problem*. Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 11): p. 1881-90.
156. ROSSMANN, M.G., *The molecular replacement method*. Acta Crystallographica A, 1990.
157. Scapin, G., *Molecular replacement then and now*. Acta Crystallographica D, 2013. **69**(11): p. 2266-2275.
158. Abergel, C., *Molecular replacement: tricks and treats*. Acta Crystallographica D, 2013: p. 1-7.
159. Evans, P. and A. McCoy, *An introduction to molecular replacement*. Acta Crystallographica D, 2008. **64**(Pt 1): p. 1-10.
160. Dauter, Z., *Solving coiled-coil protein structures*. IUCrJ, 2015. **2**(2): p. 164-165.

161. Blocquel, D., et al., *Coiled-coil deformations in crystal structures: the measles virus phosphoprotein multimerization domain as an illustrative example* Acta Crystallographica D, 2014. **70**(6).
162. Kapinos, L.E., et al., *Simultaneous formation of right- and left-handed anti-parallel coiled-coil interfaces by a coil2 fragment of human lamin A*. Journal of molecular biology, 2011. **408**(1): p. 135-146.
163. Communie, G., et al., *Structure of the tetramerization domain of measles virus phosphoprotein*. Journal of Virology, 2013. **87**(12): p. 7166-7169.
164. Read, R.J., P.D. Adams, and A.J. McCoy, *Intensity statistics in the presence of translational noncrystallographic symmetry*. Acta Crystallographica D, 2013: p. 1-8.