Diffraction methods for the crystallography of membrane proteins

Igor Melnikov

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

pour l'obtention du Grade de

DOCTEUR DE L'UNIVERSITE DE GRENOBLE ALPES

Spécialité : **Biologie Structurale et Nanobiologie**
Arrêté ministériel : 7 août 2006

Présentée par :

Igor MELNIKOV

Thèse dirigée par **Gordon LEONARD**
Co-dirigée par **Alexander POPOV**

préparée au sein du **Structural Biology Group, European Synchrotron Radiation Facility** dans **l'École Doctorale de Chimie et Sciences du Vivant**

Méthodes de diffraction pour la cristallographie des protéines membranaires

Diffraction methods for the crystallography of membrane proteins

Thèse soutenue le **27 Octobre 2017**
devant le jury composé de :

Prof. Dr. Eva PEBAY-PEYROULA	Présidente
Prof. Dr. Arwen PEARSON	Rapporteure
Dr. Meitian WANG	Examinateur
Prof. Dr. Valentin GORDELIY	Examinateur
Dr. Gleb BOURENKOV	Examinateur
Dr. Alexander POPOV	Directeur de Thèse

Table of contents

Acknowledgements

As with almost every scientific project, the work that is presented in this thesis manuscript is not the product of a single person but that of quite a large community of collaborators who have made contributions by discussing, performing experiments, advising or providing me with the words of encouragement.

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Alexander Popov for his continuous support of my study and related research, for his patience, for daily discussions and for his great scientific expertise. I appreciate all the time, ideas and efforts spent by him to make my Ph.D. work productive and informative. I also would like to thank my co-supervisor Dr. Gordon Leonard for his scientific comments and advice, for his willingness to listen and help at any time in his busy working schedule and for his tremendous help with scientific proof reading. His valuable contribution brought many of the ideas in this manuscript into their present shape.

I would like to thank the two members of my Thesis Committee: Prof. Valentin Gordeliy and Dr. Gleb Bourenkov for their remarks that allowed me to maintain the right direction of my research. I am thankful to them for their precious advice and very efficient collaboration.

I also address my acknowledgements to the members of the Joint Structural Biology Group at ESRF/EMBL Grenoble Outstation. Particularly, I thank Ulrich Zander for his help, especially in my early days at the ESRF, for providing the crystals and experimental data for algorithm tests and for his inimitable sense of humour. I also thank Olof Svensson who possesses a deep knowledge in computer science and who made a significant contribution

in allowing the methods described here to be easily used. Olof also helped me a lot in mastering my programming skills. I thank Gianluca Santoni for valuable discussions and his words of encouragement. I thank all the rest of the Structural Biology Group, it was such a pleasure working with you in such a friendly atmosphere created by you. The group seminars were always informative and interesting as well as entertaining were the other group events.

I particularly acknowledge all beamline scientists and beamline support staff on all the ESRF macromolecular crystallography-dedicated beamlines where, thanks to their work, diffraction data collection procedures went smoothly and routinely.

Special thanks go to Vitaly Polovinkin who, thanks to his strong scientific experience in membrane protein crystallisation and in structural biology in general made most of the research described here possible. Vitaly also made a very big impact by providing high-quality membrane protein crystals diffracting to highest possible resolution. I also especially thank Ivan Gushchin who provided his scientific excellence and expertise and became a central figure in the NarQ protein structural studies. His advice and remarks helped a lot in my research.

I would like also to thank the ESRF staff: administration, HR department, infrastructure, etc. Their timely support made it a real joy from doing science in Grenoble. For me and my wife the support and care of the ESRF made our life as foreigners in Grenoble much easier and happier. I thank the PSB partner institutes for creating such a strong scientific environment here on EPN campus.

Finally, and most importantly, I address my warm and sincere thanks to my family. To my parents Alexander and Elena who raised me in love and care, who worked hard to

give me and my brother a happy childhood and who greatly supported me in my aspiration to become a scientist. To my brother Denis who helped me many times in many ways. Most of all though, I would like to thank my loving wife Julia for being so supportive, patient and sensitive during our stay in Grenoble. Her unconditional love carries me through and is the source of my life energy. She gives me motivation and confidence in every single moment of my life.

Abstract

Membrane proteins are essential components of all cellular life forms. Apart from helping the lipid membrane maintain its main barrier function, membrane proteins are responsible for a vast variety of cellular functions ranging from the perception of external stimuli to membrane fusion. Huge effort is thus directed towards the study of membrane proteins particularly as their defective function is often a cause of serious medical disorders. However, a major factor that makes *in vitro* investigations of membrane proteins complicated is the problem of reconstruction of their natural environment, i.e. the lipid bilayer. This is especially relevant to the structural investigations of membrane proteins. Currently the main source of high-resolution information for membrane proteins is X-ray crystallography and here the production chain is full of difficulties, starting from protein expression, purification and crystallisation and ending with diffraction data collection and structure solution.

In the work presented here two methodological developments aimed at helping to produce high-resolution crystal structures of membrane proteins are reported and applied to the biologically-relevant, 'real world' case of the elucidation of the mechanism of signal transduction of a transmembrane histidine kinase.

In X-ray crystallography samples vary in size, shape and diffraction quality. Thus, in order to collect diffraction data in the most efficient way the crystal sample should be properly characterised prior to the actual data collection step. Raster X-ray scanning has proved to be the most viable technique to do this and in this context a method of sample pre-characterisation, applicable throughout protein crystallography is presented. The method developed estimates the positions, sizes and quality of diffraction of each crystal in

the scanned area, information which can then be used for the rational design of the data collection process. The performance of the method is demonstrated on several different types of protein crystal samples.

A bottleneck in production of crystal structures of proteins is often in the solution of the so-called 'phase problem'. When molecular replacement cannot be used for this the most widespread techniques are SAD or MAD-based experimental phasing, both of which imply incorporation of anomalous scatterers into the crystal(s) being analysed. In the list of potential useful anomalous scatterers halides particularly stand out, mostly because their low toxicity and ease of handling, but also because of the potentially large anomalous signals introduction of such ions into a crystal can produce. Indeed, the protocol of the cryo-soaking of protein crystals in halide-containing solutions has been shown to be successful in the resolution of many crystal structures of soluble proteins. Here, similar protocols were tested in the solution of the crystal structures of four different classes of membrane protein, work which suggests that the introduction of iodide into crystals of membrane proteins could be a routinely successful method for addressing the phase problem in the case of membrane protein crystal structures.

Sensor histidine kinases are one of the most common transmembrane receptors, present in all kingdoms of Life. However, understanding the mechanism of signal transduction employed by sensor histidine kinases is a fundamental question that currently remains unanswered. In order to shed light on the structural changes involved in the transmission of the signal across the lipid membrane also presented here is the crystal structure of a construct of the histidine kinase NarQ – a sensor of nitrates/nitrites – from *Escherichia coli*. The construct studied contains periplasmic sensor domain, the transmembrane helices and the cytoplasmic HAMP domain and crystal structures were

determined in both ligand-bound and mutated *apo* states. The crystal structures presented provide insight into the conformational changes occurring in the transmembrane domain and in the downstream HAMP domain during the ligand-induced signal transduction. The progress of the structural investigation was greatly enhanced by applying methodological developments presented in this work.

Résumé

Les protéines membranaires sont des composants essentiels de toutes les formes de vie cellulaire. En plus que d'aider la membrane lipidique à maintenir sa principale fonction de barrière, les protéines membranaires sont responsables d'une grande variété de fonctions cellulaires allant de la perception des stimuli externes à la fusion membranaire. Un énorme effort est donc dirigé vers l'étude des protéines membranaires, d'autant plus que leur fonction défectueuse est souvent cause de graves maladies. Cependant, un facteur majeur qui rend compliqué les recherches *in vitro* des protéines membranaires est le problème de la reconstruction de leur environnement naturel, c'est-à-dire la bicouche lipidique. Ceci est particulièrement pertinent pour les études structurelles des protéines membranaires. Actuellement, la principale source d'informations à haute résolution pour les protéines membranaires est la cristallographie aux rayons X et, ici, la chaîne de production des structures est pleine de difficultés, à partir de l'expression des protéines, de la purification et de la cristallisation et se terminant par la collecte de données de diffraction et la solution structurale.

Dans le travail présenté ici, deux développements méthodologiques visant à aider à produire des structures cristallines à haute résolution des protéines membranaires sont rapportés et appliqués au cas réel de l'élucidation du mécanisme de transduction du signal d'une histidine kinase transmembranaire.

Dans la cristallographie aux rayons X, les échantillons varient en taille, en forme et en qualité de diffraction. Ainsi, afin de collecter les données de diffraction de manière plus efficace, l'échantillon devrait être correctement caractérisé avant l'étape de collecte de données. L'analyse raster par rayons X est la technique la plus viable pour ce faire et, dans

ce contexte, une méthode de pré-caractérisation de l'échantillon, applicable dans toute la cristallographie protéique, est présentée. La méthode développée estime les positions, les tailles et la qualité de la diffraction de chaque cristal dans la zone scannée, des informations qui peuvent ensuite être utilisées pour la conception rationnelle du processus de collecte de données. La performance de la méthode est démontrée sur différents types d'échantillons de cristaux de protéines.

Un goulot d'étranglement dans la production des structures cristallines des protéines est souvent dans la solution du soi-disant «problème de phase». Lorsque le remplacement moléculaire ne peut pas être utilisé pour cela, les techniques les plus répandues sont les techniques de mise en phase expérimental basé sur SAD ou MAD, ce qui implique l'incorporation de diffuseurs anormaux dans le(s) cristal(s) à analyser. Dans la liste des diffuseurs potentiellement utiles, les halogénures se distinguent particulièrement, principalement parce que leur faible toxicité et leur facilité de manipulation, mais aussi parce que la grande capacité à produire un signal anormal de ces ions dans un cristal. En effet, le protocole du cryo-trempage des cristaux de protéines dans les solutions contenant des halogénures a été démontré pour réussir dans la résolution de nombreuses structures cristallines de protéines solubles. Ici, des protocoles similaires ont été testés dans la solution des structures cristallines de quatre classes différentes de protéines membranaires, ce qui suggère que l'introduction de l'iodure dans des cristaux de protéines membranaires pourrait être une méthode facilement et efficacement utilisé pour traiter le problème de phase dans le cas des cristaux de protéines membranaires.

Le capteur histidine kinases est l'un des récepteurs transmembranaires les plus courants, présents dans tous les royaumes de la Vie. Cependant, la compréhension du mécanisme de la transduction du signal utilisée par les histidine kinases sensorielles est une

question fondamentale qui reste actuellement sans réponse. Afin d'éclairer les changements structurels impliqués dans la transmission du signal à travers la membrane, on retrouve ici la structure cristalline d'une construction de l'histidine kinase NarQ – un capteur de nitrates/nitrites – de *Escherichia coli*. La construction étudiée contient le domaine du capteur périplasmique, les hélices transmembranaires et le domaine HAMP cytoplasmique et les structures cristallines ont été déterminés dans les états lie à un ligand et *apo* muté. Les structures cristallines présentées fournissent un aperçu des changements de conformation qui se produisent dans le domaine transmembranaire et dans le domaine HAMP pendant la transduction en aval du signal induite par le ligand. Les avancements de la recherche structurelle ont été grandement améliorés en appliquant les développements méthodologiques présentés dans ce travail.

Popularised summary of the thesis

Membrane proteins are important components of a living cell located at or partially inside cellular membranes. Investigations of the three-dimensional structure of membrane proteins is of fundamental importance as they shed light on the mechanism of action of membrane proteins and provide a structural framework for drug development and other important applied disciplines. To date the most efficient way to elucidate the structure of a membrane protein is X-ray crystallography. The work presented here showcases the development of two methods aimed at optimising diffraction data collection and structure solution processes in membrane protein crystallography and shows how these can be applied in the solution of crystal structures of a transmembrane receptor histidine kinase. Analysis of the crystal structures obtained sheds new light on the mechanism of transmembrane signal transduction used by histidine kinases.

Résumé vulgarisé

Les protéines membranaires sont des composants importants d'une cellule vivante située sur ou partiellement dans la membrane cellulaire. Les recherches de la structure tridimensionnelle des protéines membranaires sont d'une importance fondamentale car elles mettent en lumière le mécanisme d'action des protéines membranaires et fournissent un cadre structurel pour le développement de médicaments et d'autres disciplines appliquées importantes. À ce jour, le moyen le plus efficace d'élucider la structure d'une protéine est la cristallographie aux rayons X. Le travail présenté ici combine le développement de deux méthodes visant à optimiser les processus de collecte de données de diffraction et de solution de la structure dans la cristallographie des protéines membranaires et montre comment elles peuvent être appliquées dans le cas réel de la solution des structures cristallines des récepteurs transmembranaires. L'analyse des structures cristallines obtenues éclaire une nouvelle lumière sur le mécanisme de transduction du signal utilisé par les histidine kinases transmembranaires.

Abbreviations

MP	membrane protein
GPCR	G protein-coupled receptor
MX	macromolecular crystallography
NMR	nuclear magnetic resonance
PDB	protein data bank
LCP	lipidic cubic phase
TM	transmembrane
UV	ultraviolet
TPE-UVF	two-photon-excitation ultraviolet fluorescence
Cryo-EM	cryo-electron microscopy
SHG	second harmonic generation
CCD	charge-coupled device
ESRF	European Synchrotron Radiation Facility
XFEL	X-ray free-electron laser
FWHM	full width at half maximum
r.m.s.d.	root-mean square deviation
HCA	hierarchical cluster analysis
SIR/MIR	single/multiple isomorphous replacement
SAD/MAD	single/multi-wavelength anomalous diffraction
SIRAS	single-wavelength isomorphous replacement with anomalous diffraction

FFT	fast Fourier transform
TCS	two-component system
σECF	σ-factors of the extracytoplasmic function
HK	histidine kinase
RR	response regulator
HAMP	a domain present in **h**istidine kinases, **a**denylate cyclases, **m**ethyl-accepting chemotaxis proteins and **p**hosphatases
PAS	a domain first discovered in **p**eriodic circadian proteins, **a**ryl hydrocarbon nuclear translocator proteins, and **s**ingle-minded proteins
GAF	a domain present in c**G**MP-specific phosphodiesterases, **a**denylate cyclases, and **f**ormate hydrogenases
WT	wild-type
CC	colied coil

List of figures and tables

Tables

Figures

Preamble

Membrane proteins (MPs, (Langel *et al.*, 2010)) are the key functional machinery of a living organism. Being located in, or close to, the lipid membrane of a cell, MPs help maintain the essential barrier function of the membrane. At the same time, however, they are also responsible for the interaction of a cell with its environment and a wide variety of MPs are involved in crucial cellular functions including ion and molecule transport, signal transduction and energy accumulation, to name but a few. Because of their important role in the life cycle of an organism malfunctioning MPs have a severe impact on its health and although MPs are encoded only by one third of the human genome (Wallin & Heijne, 2008) they currently constitute up to 60% of the targets for drugs (Overington *et al.*, 2006). Thus, the importance of investigating and elucidating the mechanisms of action of MPs, as well as being of fundamental interest, is strongly reinforced by prospective medical applications, particularly drug design.

The study of MPs is made difficult by a range of factors which are mainly associated with their amphiphilic nature (Figure 1a): certain surfaces of MPs are hydrophobic, allowing them to be embedded into the lipid bilayer and to interact with it. Lipid bilayers are complex, heterogeneous systems which vary in lipid composition in different organisms or even different cell types of the same organism. Indeed, MPs require a natural lipid environment for proper folding and function and are normally insoluble in aqueous solution. *In vitro* studies thus imply reconstitution of a MP in synthetic objects (Figure 1) such as lipid vesicles or detergent micelles (Seddon *et al.*, 2004). Lipid vesicles (Figure 1b) are artificial objects of near-spherical shape formed by lipid bilayers. They are able to carry few MP molecules inserted into the bilayer and in such manner allow them to freely diffuse across aqueous solution. Due to their flexibility in many aspects such as size or lipid composition lipid

vesicles offer a broad potential for single-molecule studies of MPs. Detergent micelles (Figure 1c) create a shell around the hydrophobic surface of an MP and within this shell the MP molecule can then be dissolved in aqueous solutions. However, the detergent shell may seriously interfere with the internal interactions within a MP molecule, affecting its usual mode of action. Employing lipid vesicles or detergent micelles in the study of MPs requires some effort but has now become routine as have, in the last decade or so, more complex techniques such as reconstitution of MPs in nanodiscs (Figure 2) (Nikolaev *et al.*, 2017).

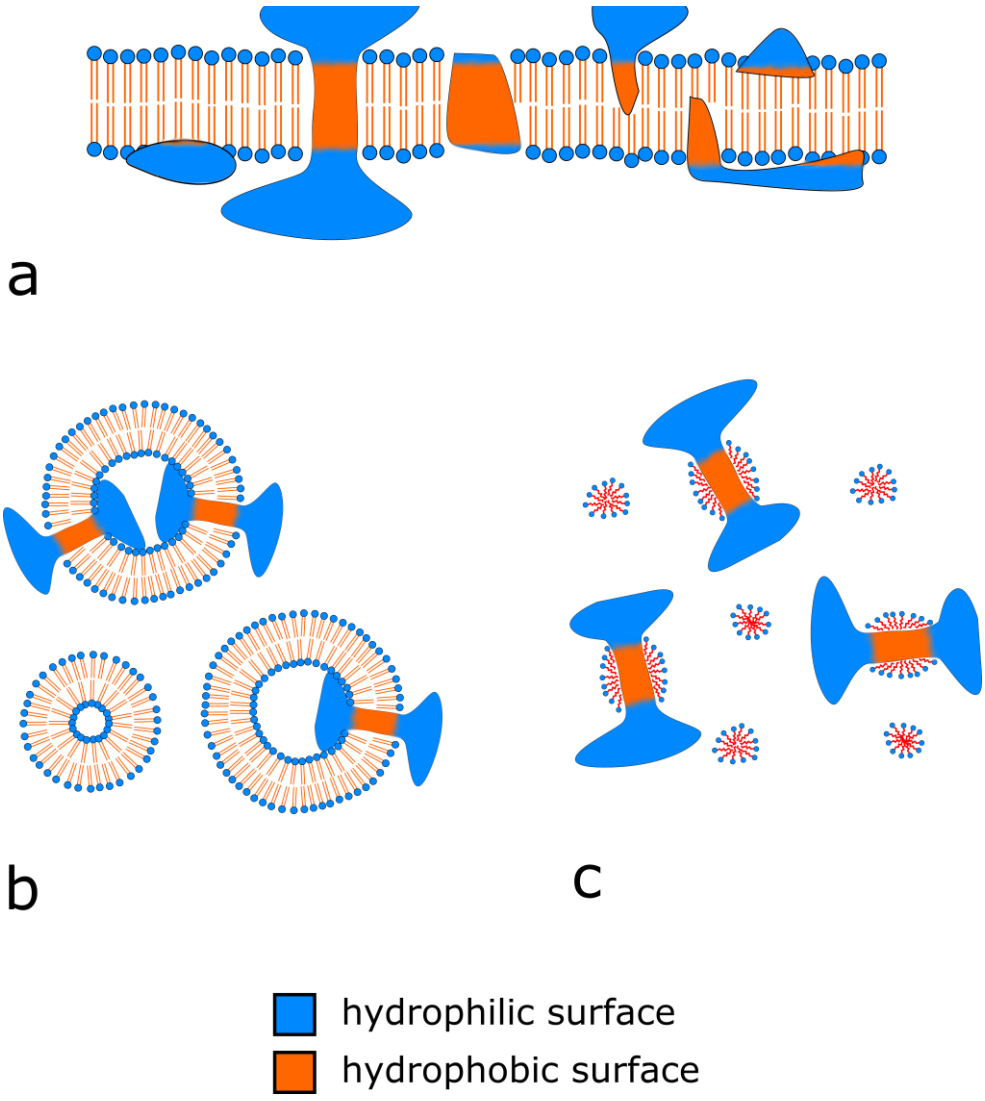


***Figure 1**. Membrane proteins (MPs) embedded in various environments. The hydrophobic surfaces of MPs are shown in orange, their hydrophilic surfaces in blue. Lipid molecules are shown (a and b) with orange tails, detergent molecules (c) with red tails. **a**. Schematic illustration of different types of membrane proteins in their natural environment – a lipid bilayer. The proteins interact with the bilayer and are anchored in it differently owing to the amphiphilic nature of a MP. **b**. Schematic illustration of a MP inserted into lipid vesicles. **c**. Schematic illustration of a MP assembled into detergent micelles.*

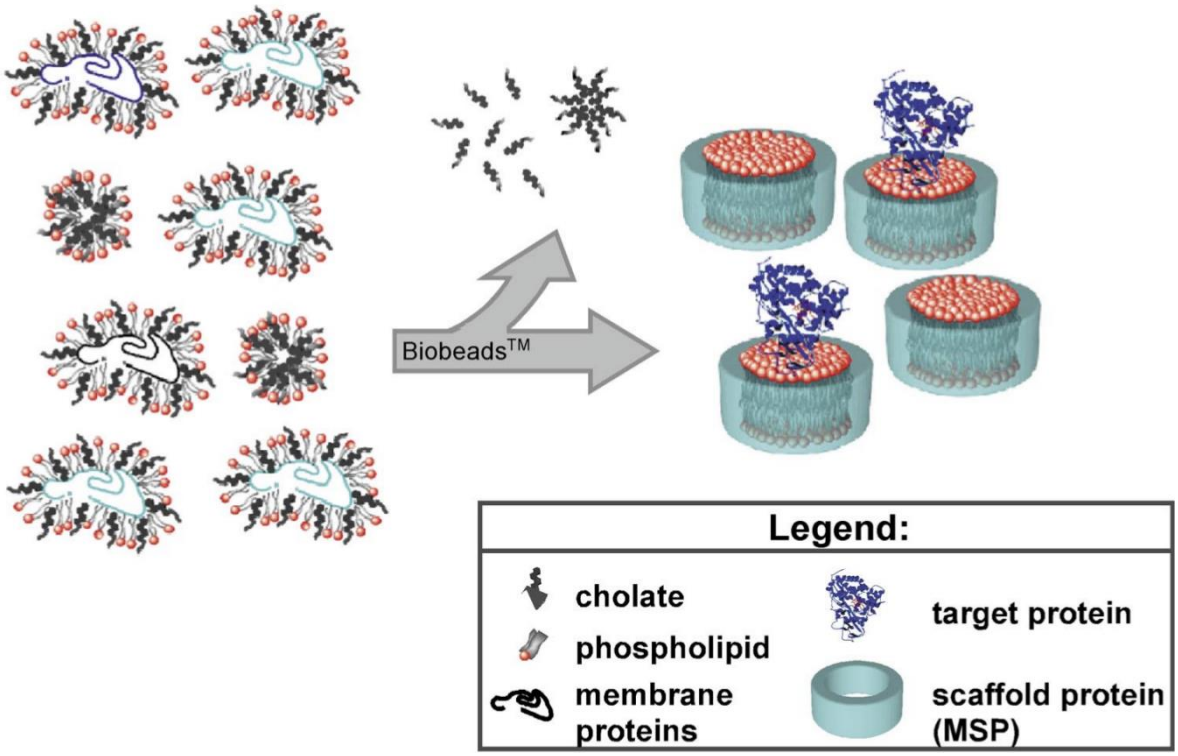


Figure 2. *Formation of MP-carrying nanodiscs formed through the self-assembly of lipid-detergent micelles. A scaffold protein (MSP) wraps around the lipid bilayer with a MP inserted. Adapted from (Civjan et al., 2003).*

A further bottleneck in the study of MPs is in their recombinant expression and purification. Lipid composition may vary considerably between the host cell and the expression system and this often affects the final protein yield. Moreover, a given MP is usually present in low amounts at the surface of the cell and when overexpressed it gathers into inclusion bodies. Some MPs also require post-translational modification, making production much more difficult (Walsh *et al.*, 2005). For example, human GPCRs cannot be expressed in *E. coli* which is the most popular expression system.

The function of any protein, including MPs, is closely connected with its three-dimensional (3-D) structure. If we are to reveal the mechanism of a MP we must therefore obtain its 3-D structure, usually in several states. The dawn of structural biology of MPs began with the elucidation of a low-resolution structure (Figure 3) (Henderson & Unwin, 1975) of the purple membrane of *Halobacterium Halobium* in which molecules of the MP bacteriorhodopsin (Oesterhelt & Stoeckenius, 1971) were self-organised into two-

dimensional crystals. The structure was obtained by combining electron microscopy and electron diffraction. Since this first membrane 3-D structure, those of roughly 700 of unique MPs have been determined and, of these, about 500 were produced in the last ten years (http://blanco.biomol.uci.edu/mpstruc/). However, despite this exponential growth (see Figure 4), the 3-D structures of MPs deposited in the Protein Data Bank (PDB; (Berman *et al.*, 2003)) still constitute only ~2% of the total number of known protein structures (Hendrickson, 2016). To date the most prominent method in obtaining the high-resolution structures of MPs has been X-ray crystallography, accounting for almost 88% of the deposited structures. Other methods used include Nuclear Magnetic Resonance (NMR, 3.3% of deposited MP structures) and cryo-electron microscopy (Cryo-EM, 8.2% of deposited MP structures).

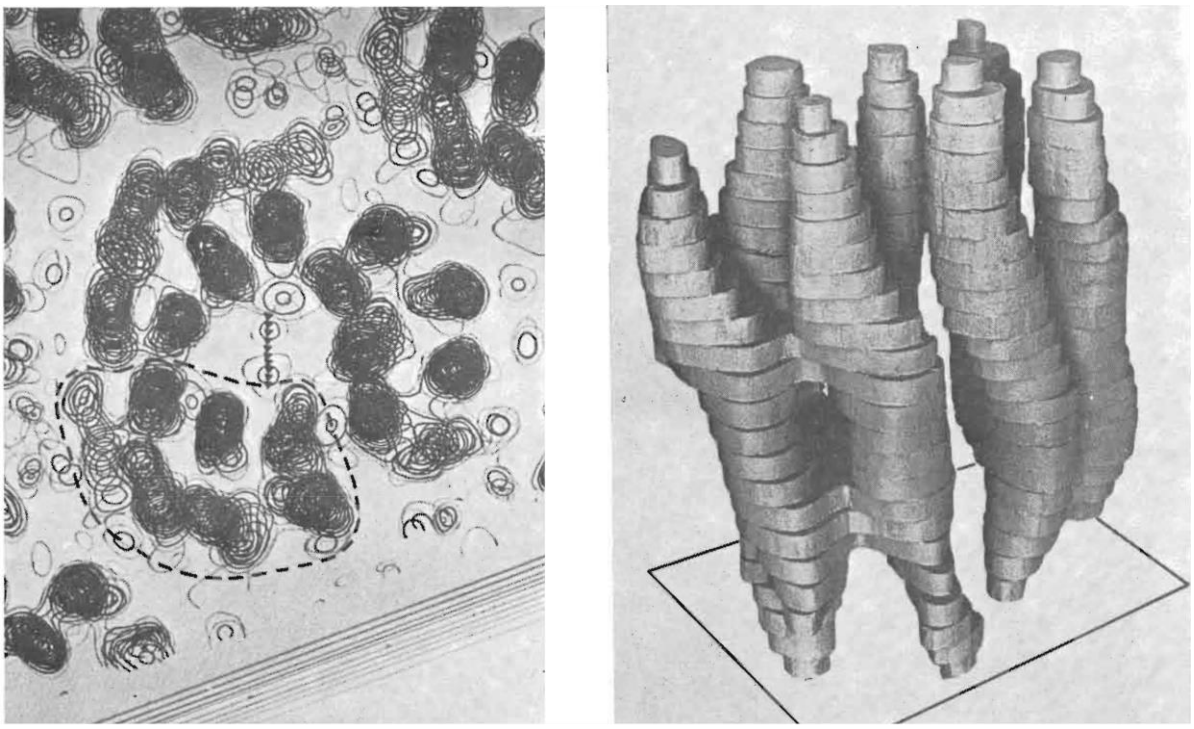


Figure 3. *The 7 Å-resolution structure of bacteriorhodopsin (bR) obtained by Henderson and Unwin (Henderson & Unwin, 1975).* ***a.*** *A section of three-dimensional electron density map showing 50 Å-thick region of the purple membrane of Halobacterium Halobium. A trimer of bR molecules can be easily identified.* ***b.*** *A model of the structure of a bR constructed based on the studies of Henderson and Unwin. Adapted from (Henderson & Unwin, 1975).*

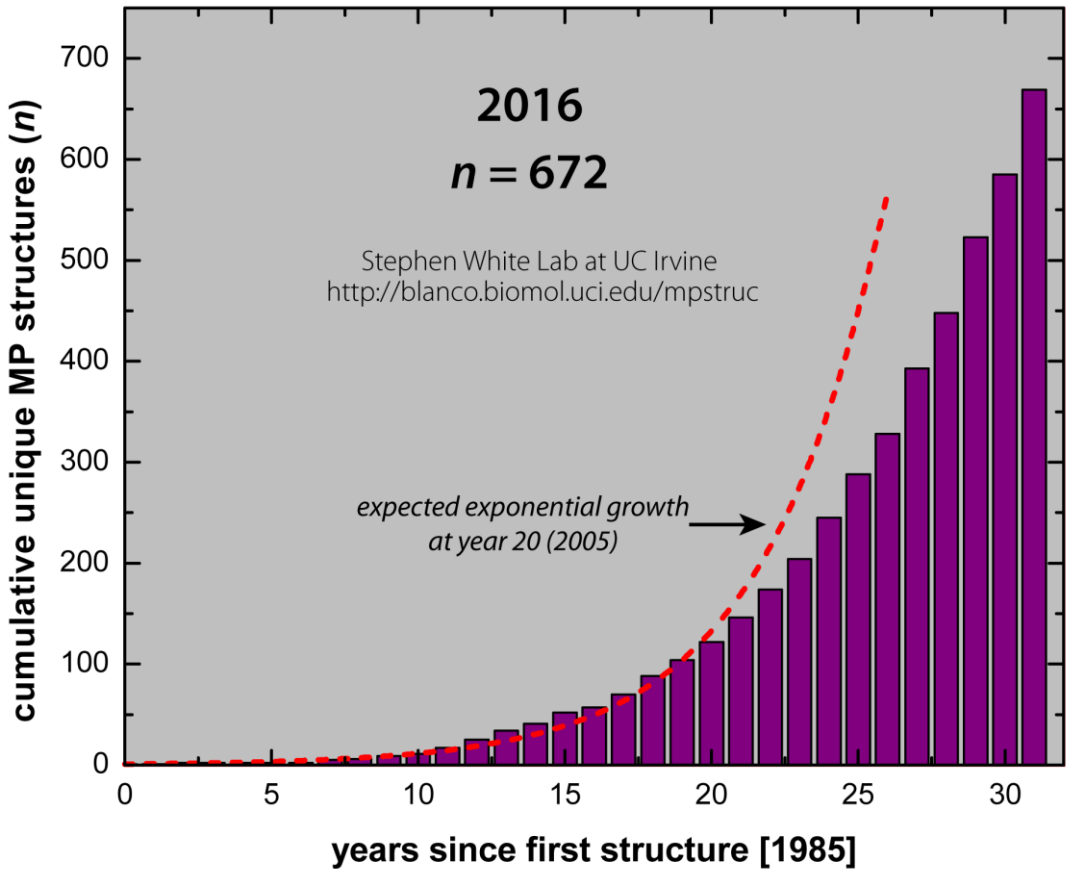


***Figure 4**. The number of unique membrane protein structures deposited in the PDB since 1985. The red line represents expected exponential growth of the number of structures calculated based on 1985-2005 data. Adapted from http://blanco.biomol.uci.edu/mpstruc/.*

The role of NMR in the determination of the 3-D structures of MPs is, at least in the near future, likely to remain limited the maximal macromolecule size that can be analysed employing conventional Overhauser effects is ~25kDa (Duss *et al.*, 2015) due to the fact that for large macromolecular ensembles peaks in the spectra overlap. The role of EM in the determination of the 3-D structures of MPs is likely to increase sharply thanks to recent breakthroughs in instrumentation and experimental methodology (Kuhlbrandt, 2014; Subramaniam *et al.*, 2016; Callaway, 2015) that have meant it is now possible to obtain the 3-D structures, using cryo-EM, of relatively small proteins at medium resolution (~90 kDa at 3.8 Å in (Merk *et al.*, 2016); (Bartesaghi *et al.*, 2015; Banerjee *et al.*, 2016)). Nevertheless, and as is currently the case, Macromolecular crystallography (MX) – especially on

synchrotron beamlines – will likely remain the technique of choice for the determination of high resolution 3-D structures of MPs.

MX implies producing relatively large three-dimensional crystals of a protein which, particularly in the case of MPs, is the main bottleneck of the method. The first three-dimensional crystals of MPs suitable for high-resolution MX were obtained using so the called *in surfo* method. Here a MP is solubilised in detergent micelles and, thus, with its hydrophobic area covered can be handled and subject to crystallisation trials using standard protocols in the same way as a soluble protein (Michel, 1983). Using this technique, the crystal structure of the photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis* was determined at the resolution of 3 Å (Deisenhofer *et al.*, 1984). The *in surfo* method remains widely employed in MP MX. However the technique has certain drawbacks: firstly, detergents used in *in surfo* protocols often decreases the number of possible protein crystal contacts (Caffrey, 2003; Seddon *et al.*, 2004); secondly, being wrapped into the detergent micelle a MP could adopt a state far from its natural conformation. These drawbacks can thus limit the value of the structural information obtained from *in surfo* grown crystals.

In the 1990s the field of MP crystallisation was revolutionised by the appearance of the *in meso* method (Landau & Rosenbusch, 1996). The term "*in meso*" implies the use of a mesophase, mostly cubic or swollen cubic phases (Landau & Rosenbusch, 1996; Cherezov *et al.*, 2006) formed by lipids, as a container for crystallisation. Lipidic cubic phases (LCPs) are complex three-dimensional ordered lipid bilayer arrays permeated by water channels (see Figure 5). MP molecules introduced into the lipid bilayer in such structures can often crystallise more readily and the MP itself has better chances to assume a conformation reflecting that adopted *in vivo.*

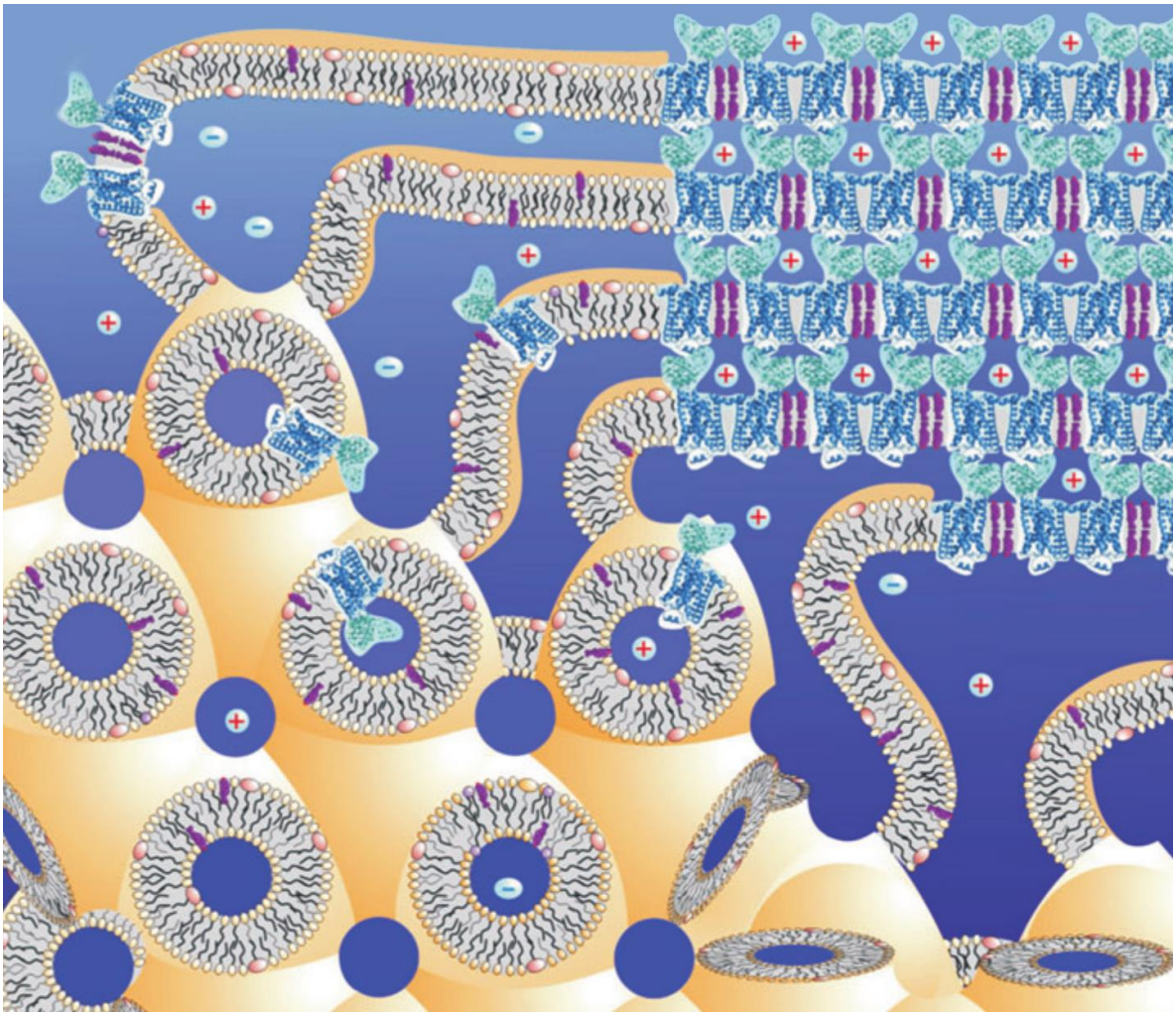


Figure 5. *Principle of MP crystallisation in the lipidic cubic phase. The protein is integrated into the curved bilayer of the lipidic cubic phase (bottom left corner) penetrated by bicontinuous water channels and is able to diffuse throughout the whole its volume. The addition of the precipitant shifts the equilibrium and the protein assembles into a layer-structured crystal (upper right corner). Adapted from (Caffrey, 2009).*

The *in meso* method of MP crystallisation was used for the first time in experiments to produce large crystals of bacteriorhodopsin (bR) from *Halobacterium Halobium* which were then used to determine the crystal structure of bR at high resolution (Figure 6) (Pebay-Peyroula *et al.*, 1997). Since this first use *in meso* MP crystallisation has been used to produce crystals from which diffraction data collection has resulted about 20% of all MP crystal structures deposited in the PDB. More than half of these depositions have occurred during the past two years, indicating that *in meso* method of MP crystallisation is becoming widely used.

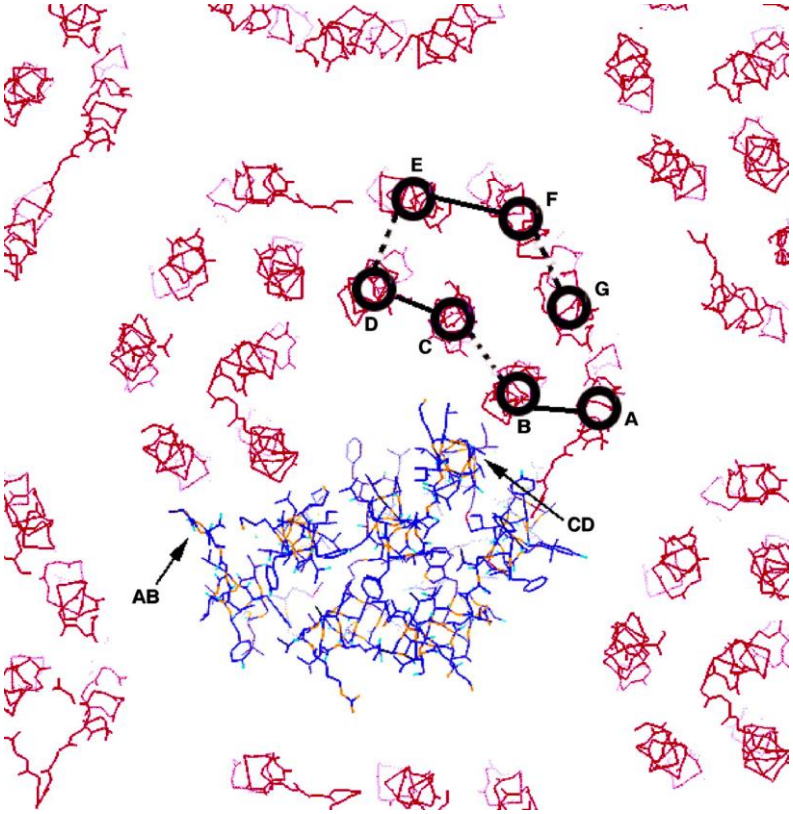


Figure 6. *The bR trimers, viewed from above the membrane in the 2.5 Å-resolution crystal structure of bR determined using the in meso method by Pebay-Peyroula et al. Each bR monomer comprises 7 transmembrane helices (A-G, highlighted for one bR monomer). Two monomers are shown as C_α traces, the third in stick representation. The colours of the original figure were inverted to adapt to white background. Adapted from (Pebay-Peyroula et al., 1997).*

However, although structural studies of MPs are of great importance in both medical applications and in fundamental biology and despite the advances in MP protein handling and crystallisation briefly outlined above, the number of structurally characterised MPs is still negligible compared to that for soluble proteins. A major reason for this is that, even once they are obtained, crystals of MPs are often not straightforwardly amenable to structure solution by MX. The crystals obtained (see, for example, GPCRs (Cherezov *et al.*, 2007) are often small and/or fragile and/or mechanically twinned and/or sensitive to radiation damage and the collection of high-quality diffraction data from such samples is a very difficult undertaking. The work presented in this thesis contributes to the development of protocols facilitating MP MX: Chapter 1 presents X-ray based methods for the rational design of MX data collection via pre-interrogation of a single crystal or of many crystals contained in the same sample holder; Chapter 2 demonstrates a potentially universal

method, based on the incorporation of iodide ions, for the *de novo* solution of MP crystal structures; Chapter 3 then shows the result of applying of the methods developed in Chapters 1 and 2 on a real case study aimed at revealing the mechanism of signal transduction of sensor histidine kinase proteins.

1. The rational design of data collection by the determination of crystal sample parameters

1.1. Summary

In Chapter 1 I present *MeshBest* which comprises novel methods and a software program for the X-ray based on-axis pre-analysis of protein crystals prior to macromolecular crystallography data collections. The method is based on the X-ray diffraction mesh (raster) scan technique and produces a two-dimensional crystal map of the sample area showing estimates of each crystal's dimensions, positions and diffraction qualities. Sample regions where several crystals are superimposed at the beam producing multi-crystal diffraction pattern are distinguished from the regions with single-crystal diffraction. The applicability of the method is demonstrated on several cases.

1.2. Introduction

In MX experiments the measured quantity is the intensity of the X-rays diffracted by a protein crystal. Many factors affect the quality of the intensity data obtained. These include the flux of the incident beam, measurement time and the general 'quality' of the crystal being studied. A wide range of factors affect crystal quality including crystal size, the conformational and dynamic heterogeneity of the molecules making up the crystal, potential lack of long range order, mechanical or merohedral twinning and sensitivity to radiation damage (MacPherson, 1982). The size of the X-ray beam used to illuminate crystals is also an important factor in determining diffraction data quality. In MX experiments crystals are nearly always mounted in crystallisation mother liquor (or mesophase) supplemented with a cryo-protectant and, here, X-ray beam size should ideally be optimised to maximise diffraction signal from the crystal and reduce the noise (diffuse scattering) resulting from the medium on which the crystal is mounted.

In recent decades rapid advances in technology, software and experimental protocols have dramatically improved data quality in MX experiments. At synchrotron sources, notable amongst these advances are the development of highly automated, highly advanced, high brilliance MX beamlines at 3rd generation sources including the ESRF (Mueller-Dieckmann *et al.*, 2015) and the advent of single photon counting pixel-array detectors (Henrich *et al.*, 2009) capable of shutterless (Hülsen *et al.*, 2006) data acquisition that enables rapid data collection using fine-slicing techniques (Pflugrath, 1999). Taken together, these have facilitated experiments on more challenging samples, particularly micron sized crystals, allowing diffraction data collection employing complex sample motions from a series of crystals or from different regions of the same crystal (Flot *et al.*, 2010; Svensson *et al.*, 2015). However, such experiments can only be optimised via a pre-

interrogation of the sample(s) contained in a sample holder (Bourenkov & Popov, 2006; Zander *et al.*, 2015). Such pre-interrogation provides essential information concerning crystal shape, size, position and diffraction strength. The conventional way of estimating crystal sizes and positions is to use optical microscope and most, if not all, modern synchrotron-based MX beamlines are equipped with on-(X-ray beam)-axis optical microscopes which allow experimenters to manually (Snell *et al.*, 2004) or automatically (Andrey *et al.*, 2004; Karain *et al.*, 2002; Lavault *et al.*, 2006; Pothineni *et al.*, 2006) detect and align crystals in the X-ray beam (Figure 7Figure 8).

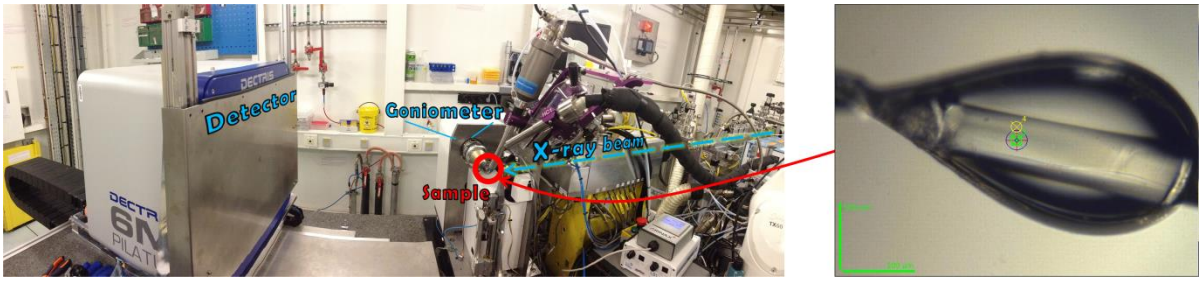


Figure 7*. The experimental setup for MX on synchrotron beamline (ESRF ID23-1). The microphotograph of the sample area with a trypsin crystal inside a micro loop is shown in the right.*

Automatic detection and alignment of crystals using on-axis visible light microscopes implies using intensity-based pixel bright-field image processing algorithms to retrieve the position(s) of the crystal sample(s). Such techniques gets very complicated when the sample holder is opaque, for example due to improper cryo-protection or due to the handling of crystals grown in lipidic mesophases. In such cases, another option for crystal detection is to use fluorescent microscopy. Mostly targeted on the excitation of tryptophan fluorescence in protein crystals, this method implies using an UV laser for single-photon excitation (Figure 9a) (Vernede *et al.*, 2006) or a 532 nm laser for two-photon excitation (TPE-UVF, (Madden *et al.*, 2011; Newman *et al.*, 2016)). In addition to fluorescence microscopy, the non-linear technique based on the phenomenon of second-harmonic generation (SHG, Figure 9b) has

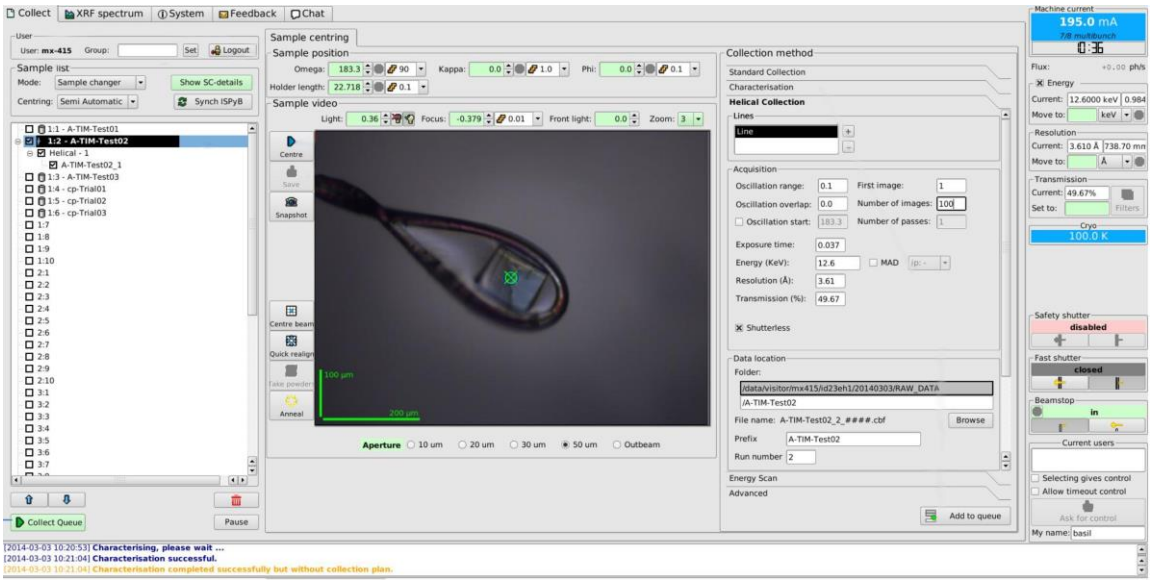


Figure 8. *The graphical user interface (GUI) for the MX beamline control software MxCUBE (Gabadinho et al., 2010; de Sanctis et al., 2016). The on-line image of the on-axis light microscope is shown in the centre of the interface. Crystal centring could be performed manually via interactive communication with goniometer through the interface. Adapted from www.esrf.eu/.*

also been employed for crystal detection prior to MX experiments (Haupert & Simpson, 2011; Kissick *et al.*, 2013; Madden *et al.*, 2011, 2013; Newman *et al.*, 2016). SHG is observed in crystals of which the point group lacks a centre of symmetry (i.e. crystals of chiral protein and/or nucleic acid molecules). Upon excitation of such a crystal with light, typically in the infrared region (λ ~ 1000 nm), the energy of the re-emitted photons is doubled and detectable at λ ~ 500 nm. Both fluorescence-based and SHG techniques give high signal-to-noise ratios due to the application of spectral-selective filters which remove the (Rayleigh) light scattering contribution of the crystal environment to the image background. However, both have their drawbacks: UV-based methods will not work in some cases where the protein is lacking aromatic amino acid residues or will give false positive results for non-crystalline protein precipitates and some sample holder materials; SHG is better suited to low-symmetry crystals as the signal is reduced with higher order symmetry crystals. Indeed, SHG is not allowed in the cubic 432 symmetry class (Haupert & Simpson, 2011). Moreover, some salts or other organic compounds in precipitants may form non-centrosymmetric crystals resulting in false-positive protein crystal detection. However, by using these two

techniques in conjunction it could be possible to eliminate all false-positive detections. Another method that has been tried for the detection of protein crystals is Raman spectroscopy (Nitahara *et al.*, 2012). This, however, does not seem a routinely applicable technique as measurement times are very long even on beamlines equipped with an on-line Raman spectroscopy set up such as on ESRF beamline ID29 (von Stetten *et al.*, 2015).

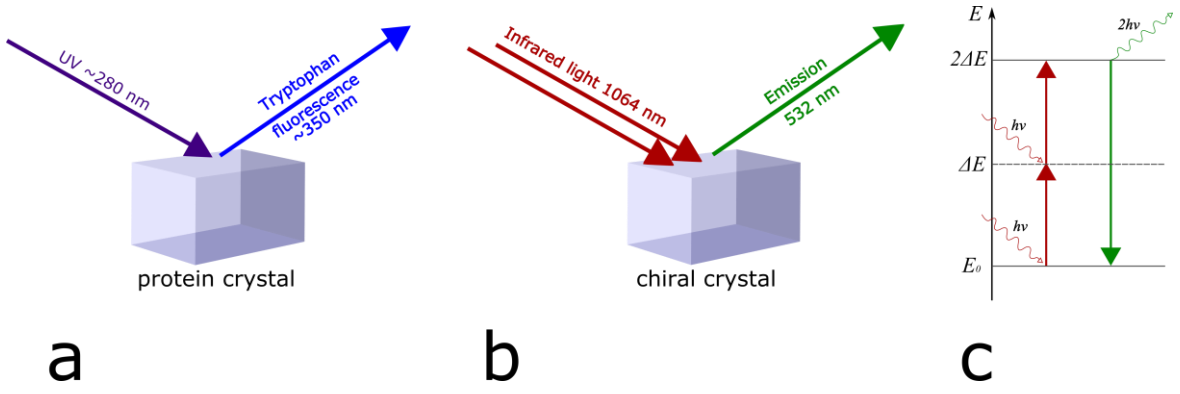


***Figure 9**. Principles of UV- and SHG-based protein crystal detection. **a.** A protein crystal under exposure of UV light produces fluorescence from aromatic side chains, primarily tryptophan. **b.** A crystal of chiral molecule exhibits non-linear SHG with absorption of two infrared photons and reemission of one visible photon with doubled energy (c). **c.** The mechanism of SHG described on an energy diagram. Two photons with energy ΔE are absorbed and the system goes back to the ground state with single photon emission.*

A major drawback of all light-based techniques for crystal detection is that they can be affected by refraction, particularly from the curved surface of the mother liquor in a micro loop in which protein crystals may be mounted. This may cause significant inaccuracy in determining crystal positions and sizes. Thus, a better option is to use X-ray based techniques. One possibility here is X-ray tomography (Brockhauser *et al.*, 2008; Warren *et al.*, 2013). However, the application of this is limited as a major requirement is a relatively large X-ray beam size (several hundreds of microns) needed to capture the entire sample area. It is therefore quite inconvenient to modify modern MX beamline installations by unmounting X-ray-focusing optics and adding further tomography-coupled equipment such as scintillator and a suitable area detector. X-ray scanning transmission microscopy, which does not need modifications of the 'standard' beamline setup (Figure 7), has thus been proposed as an alternative approach (Wojdyla *et al.*, 2016).

Despite the developments outlined above, by far the most popular approach for crystal detection is the two-dimensional raster X-ray diffraction scan, implemented with slight variations on several MX synchrotron beamlines worldwide (Cherezov *et al.*, 2009; Bowler *et al.*, 2010; Aishima *et al.*, 2010; Svensson *et al.*, 2015; Zander *et al.*, 2015; Wojdyla *et al.*, 2016). In Figure 10 the principle of the two-dimensional raster X-ray scan, which does not require modification of the standard set up, is presented. During such a scan the sample area is translated under the exposure of the X-ray beam and diffraction images are accumulated at each point of a pre-defined two-dimensional grid, the number of points in which is defined by the size of the incident X-ray beam. The resulting diffraction images are then analysed (i.e. using *DISTL&Spotfinder* (Zhang *et al.*, 2006), *CrystFEL* (White *et al.*, 2012), *MOSFLM* (Battye *et al.*, 2011), *XDS* (Kabsch, 2010), *EDNA/BEST* (Bourenkov & Popov, 2010), *Dozor* (Zander *et al.*, 2015) etc.) for the presence of diffraction spots and the resolution range to which they extend. The diffraction signal on each particular image is then ranked and the results displayed in a two-dimensional heat map.

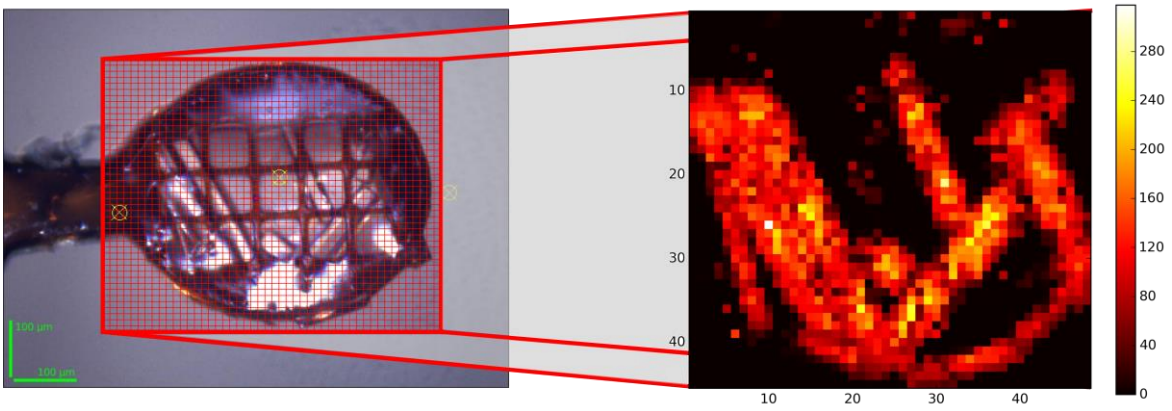


Figure 10. *The sample area with thermolysin crystals mounted in the sample holder (left) and the diffraction heat map (right) based on the mesh scan performed in the sample area depicted with red rectangular mesh. Mesh scan diffraction images were recorded at the positions of unit elements of the mesh using beam aperture with a size of a unit mesh element. The color-coded value on the heat map is based on diffraction signal estimates produced by Dozor for each image of the mesh scan, namely Dozor-scores.*

At the ESRF this two-dimensional diffraction-scanning technique is called an 'X-ray mesh scan' (Svensson *et al.*, 2015; Zander *et al.*, 2015) and is implemented on all its MX

beamlines with the software program *Dozor* (Zander *et al.*, 2015) being used for the analysis of diffraction images. *Dozor* determines the coordinates and strength of diffraction spots and produces an estimate of diffraction signal – the *Dozor* score – for each image. The main feature of *Dozor* is in its complex diffraction ranking algorithm that allows protein crystal diffraction to be distinguished from any other signal such as ice rings. Figure 10 shows a typical diffraction heat map based on the *Dozor*-score. On such map it is possible to locate the areas of stronger protein crystal diffraction signal and determine accurately positions in the mesh scan for potential subsequent MX data collection.

As implied above, for a large single crystal a *Dozor*-score heat map straightforwardly shows the regions of best quality diffraction which should be used in subsequent data collection experiments. Likewise, for a sample holder containing many crystals a *Dozor*-score heat map will show the regions in the sample holder showing the highest diffraction signal. However, when many crystals are present on the sample holder the *Dozor* heat-map does not provide information as to which (well-diffracting) adjacent regions in the map – if any – a given individual crystal belongs (i.e. *Dozor* does not provide specific information concerning the number, size and disposition of crystals contained in a sample holder). Such information can be extremely important in MX experiment planning and in this Chapter a method is presented for the recognition, based on the X-ray mesh scan technique and image analysis carried out by *Dozor*, of individual protein crystals contained in the sample area. The method developed both distinguishes regions belonging to different crystals and detects regions where several crystals are exposed to the beam simultaneously, producing multi-crystal diffraction patterns. The algorithms developed are implemented in the program *MeshBest* which outputs descriptions of all individual crystals in the sample area including crystal sizes, shapes, centre positions and diffraction strength.

1.3. Methods

The *MeshBest* workflow is presented in Figure 11 and comprises three major steps: firstly, mesh scan images containing multi-crystal diffraction patterns, if any, are detected. This is done by analysing inter-spot distance statistics on each diffraction image (see §1.3.1). Secondly, diffraction images containing diffraction only from one crystal are analysed and grouped into those belonging to the same crystal (see §1.3.2). Finally, the size and disposition of all single crystals contained in the sample loop are described using an elliptical shape approximation (§1.3.3).

1.3.1. Detecting superposition of multiple diffraction patterns

Mesh scan regions in which several crystals are simultaneously illuminated by the incident X-ray beam produce diffraction images on which the diffraction patterns from the crystals illuminated are superposed (see, for example, Figure 12c).

Such 'multi-pattern' diffraction images can cause problems for diffraction image processing packages including the misindexing of the diffraction patterns and/or degraded estimation of the intensities, and their standard deviations, of reflections. In severe cases, such areas are thus best removed from subsequent diffraction data collection protocols based on mesh scans. To recognise multi-pattern regions in a mesh scan *MeshBest* uses the distribution of inter-spot distances (Figure 12) observed on the diffraction images collected during the scan (full algorithm described in §1.3.1.3).

In the case of diffraction from a single crystal, the inter-spot distances observed in a diffraction pattern can be only of certain values because all the spots are at the positions of

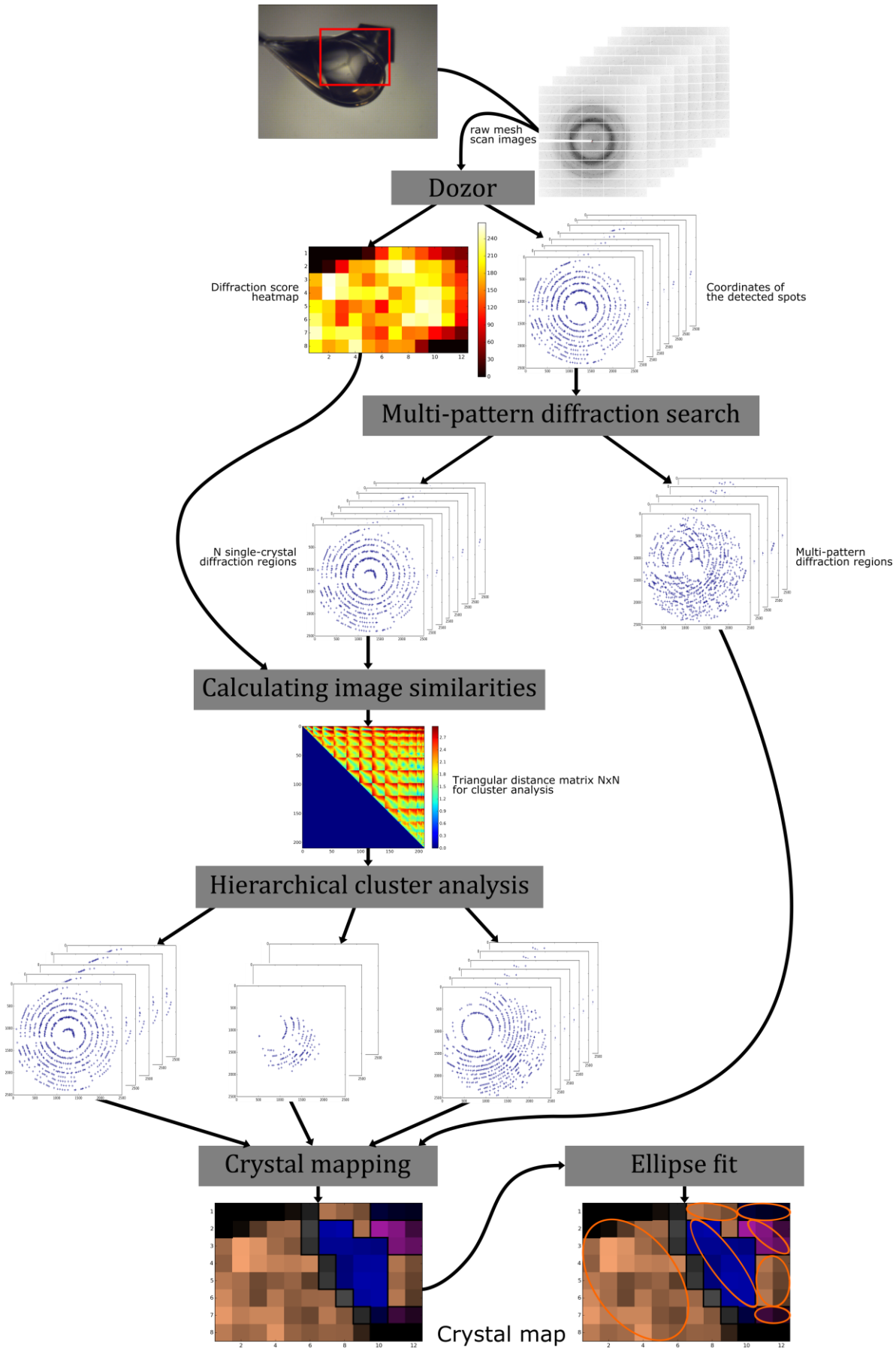


Figure 11. *Overview of the workflow of the MeshBest method. Each X-ray mesh scan produces M×N diffraction images (M and N are the numbers of grid points in orthogonal directions). These are individually analysed by Dozor which produces the estimate of diffraction signal and determines diffraction spot coordinates and partial intensities in each image. MeshBest then carries out analyses described in the main text.*

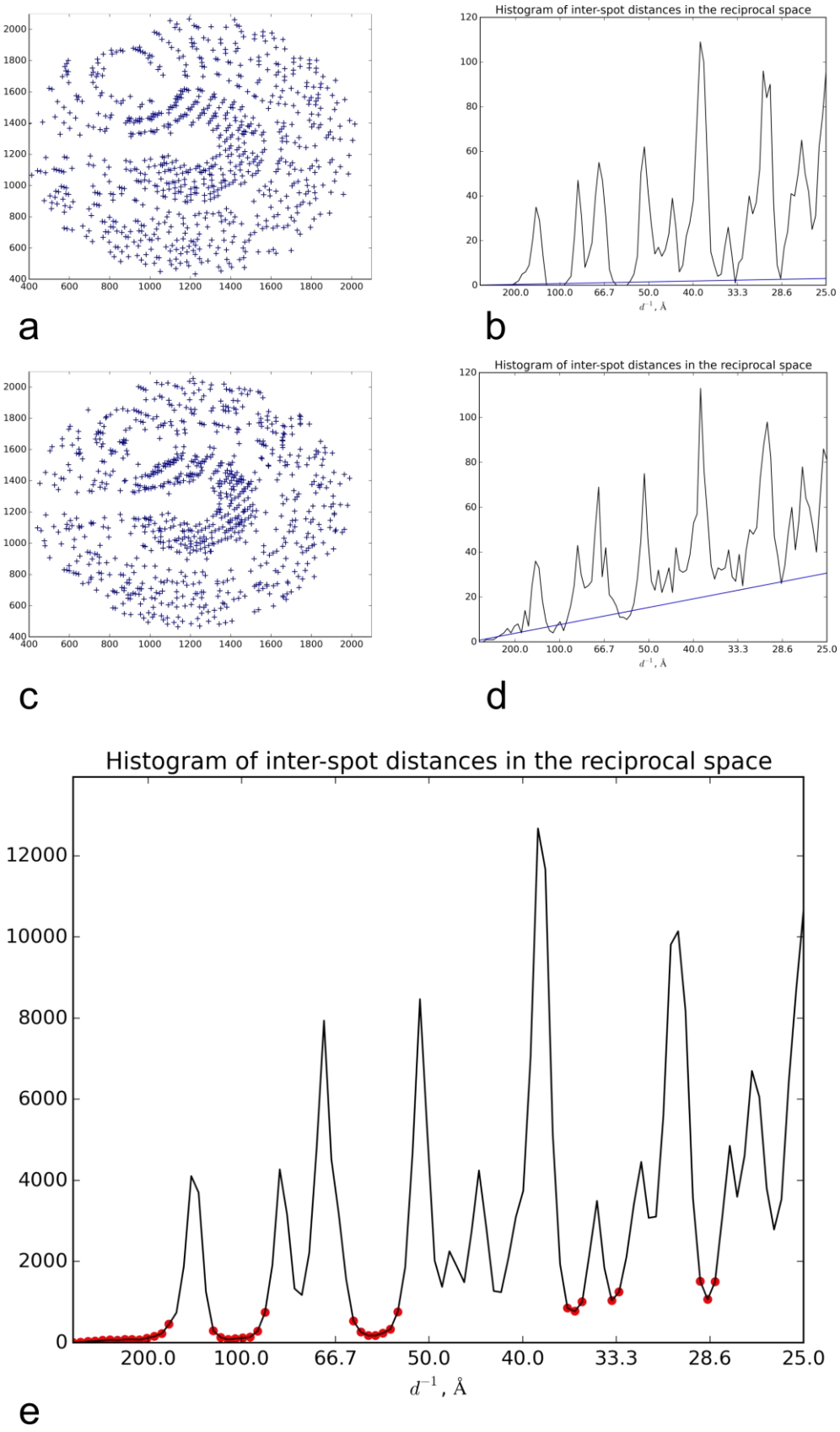


***Figure 12**. Multi-pattern diffraction analysis. **a-b.** A spot diagram in the detector plane of an example of single-pattern diffraction (a) and the corresponding inter-spot distance histogram (b) for a crystal of thermolysin (§1.4.2). **c-d.** A spot diagram in the detector plane of an example of multi-pattern diffraction (c) and the corresponding inter-spot distance histogram (d) for thermolysin crystals (§1.4.2). Blue lines in the histograms show the fitted baselines. **e.** A cumulative inter-spot distance histogram of all images of the mesh scan for crystals of thermolysin (§1.4.2) with the determined baseline regions depicted by red circles. The numbers in the histograms are presented for 100 bins in the interval of the analysis.*

the nodes of the crystal's reciprocal lattice. The distribution of inter-spot distances thus consists of several peaks with each peak representing a particular subset of distances between reciprocal lattice nodes (Figure 12a, b). The frequency at which these peaks occur depends on crystal orientation in the X-ray beam. The superposition of two or more diffraction patterns on a single diffraction image leads to an inter-spot distance distribution composed of single lattice inter-spot distances following the usual multi-peak distribution and the distances between the spots from different lattices. As will be shown below, the resulting distribution of all inter-spot distances will comprise the peaks of intra-lattice distances imposed on a monotonous, roughly linear baseline (Figure 12c, d). As the slope of the baseline is proportional to the total number of the inter-spot vectors between the different lattices in a diffraction image, fitting the baseline (i.e. determining its slope) thus allows a quantitative judgement as to whether a diffraction image contains multi-pattern diffraction. The theoretical model describing this is discussed in the next paragraph.

1.3.1.1. Approximating multi-pattern diffraction image inter-spot distributions

Here it is considered how the distribution of inter-spot distances is changed when two independent crystals contribute to the diffraction pattern observed on a single diffraction image. Inter-spot distances between the spots of the same crystal lattice obey a multi-peak distribution as described in above. However, for two random spots of different lattices inter-spot distance has the same probability distribution as the distance between a pair of random points in certain area due to the two crystals being oriented randomly. This area is, in fact, a resolution circle in the detector plane. The distribution of the distance (d) between two random points chosen in a circular area of radius r is calculated as shown in equation (1) (García-Pelayo, 2005):

$$\rho(d) = \frac{4d}{\pi r^2}\arccos\frac{d}{2r} - \frac{2d^2}{\pi r^3}\sqrt{1-\frac{d^2}{4r^2}} \qquad (1)$$

In our method, the range of inter-spot distance analysis is limited to 0.001 – 0.04 Å^{-1}. This covers the typical unit cell parameter range observed for protein crystals. The radius of the resolution circle r (discussed below) is always much larger than 0.04 Å^{-1}, for example, if the crystal diffracts to the resolution of 3 Å then it equals ~0.3 Å^{-1}. Therefore, in the range of the analysis (1) can be linearly approximated to (2) (see Figure 13).

$$\rho(d) = \frac{2d}{r^2} \qquad (2)$$

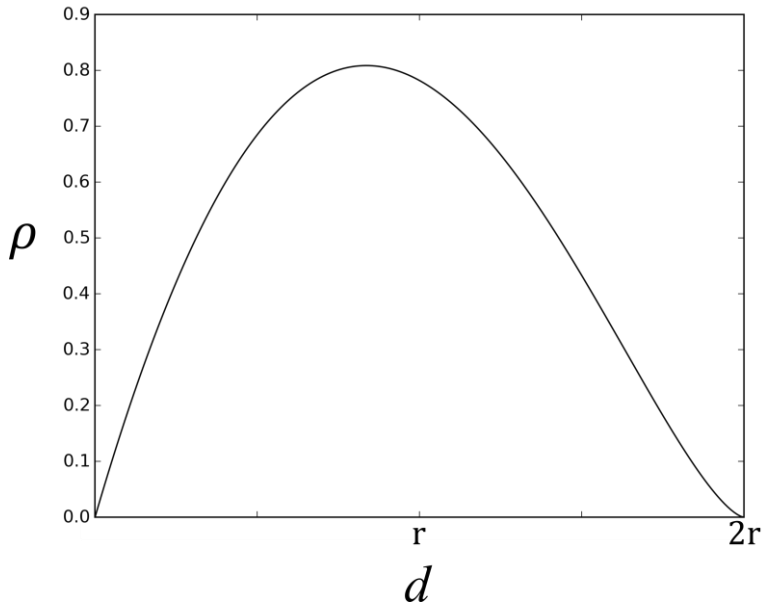


***Figure 13**. The probability distribution $\rho(d)$, calculated using equation (1), of the distances between random points in a circle with radius r.*

Considering that the first (main) crystal contributes to the diffraction image with N_m spots and the second (satellite) crystal contributes with N_s spots we expect to count $\frac{2d\Delta d}{r^2} \cdot N_m N_s$ distances of value $(d \pm \frac{\Delta d}{2})$ between the spots from different crystals. As already mentioned above, this will add a roughly straight baseline to the peaks in the distribution. The slope of this baseline in the probability distribution (equation (2)) is:

$$k = \frac{2}{r^2} \cdot N_m N_s \qquad (3)$$

and in the case of a histogram with the bin size Δd is:

$$k_0 = \frac{2\Delta d}{r^2} \cdot N_m N_s \qquad (4)$$

where r represents the radius of the highest-resolution circle where diffraction is detected.

Protein crystal diffraction image may contain rings originating from diffraction due to salt crystals or crystalline ice. Such regions are usually detected and removed by spot-searching software programs and if not – seriously violate the initial assumption (1) that the area in which two random spots from different lattices are found is circular. However, in the linear range of the distribution ($d \ll r$, equation (1), Figure 13) the distribution of inter-spot distances is not affected by the shape of the area because they are relatively small. This is further supported by the linear approximation (2) where only the area term r^2 stays. Therefore, based on simple additivity of area regardless of the shape of the spot-containing region the term r^2 could be replaced by the effective area taking into account all removed regions. Besides, the diffraction spot density is different in the centre and at the edge of detector due to the projections from the Ewald sphere. After all corrections r^2 in (2) is replaced by the area *S* of the Ewald's sphere cap that contain the detected spots in the diffraction pattern from which the corresponding salt/ice rings have been removed. Therefore, based on (4):

$$k_0 = \frac{2\Delta d}{S} \cdot N_m N_s \qquad (4a)$$

where k_0 is the slope value of the baseline in the inter-spot distance histogram with the bin size Δd, based on the diffraction pattern with N_m spots from the main crystal and with N_s spots from the satellite crystal, S – the mentioned above area of the Ewald's sphere cap.

1.3.1.2. Measuring multi-pattern diffraction

Based on the analysis shown above, the slope value of the linear fit of the baseline (hereinafter, k_0, or the slope) is used as a marker for determining whether an image from a mesh scan (or any other single diffraction image) contains multi-pattern diffraction. Indeed, according to (4a) the slope is proportional to the number of spots N_s in the satellite crystal pattern:

$$k_0 \propto N_s \qquad (5)$$

To examine the applicability of the model derived in §1.3.1.1 simulations of diffraction images containing diffraction patterns from two separate crystals were constructed using pairs of diffraction images collected from a series of single crystals (thermolysin, thaumatin and NarQ) during standard MX experiments (Figure 14).

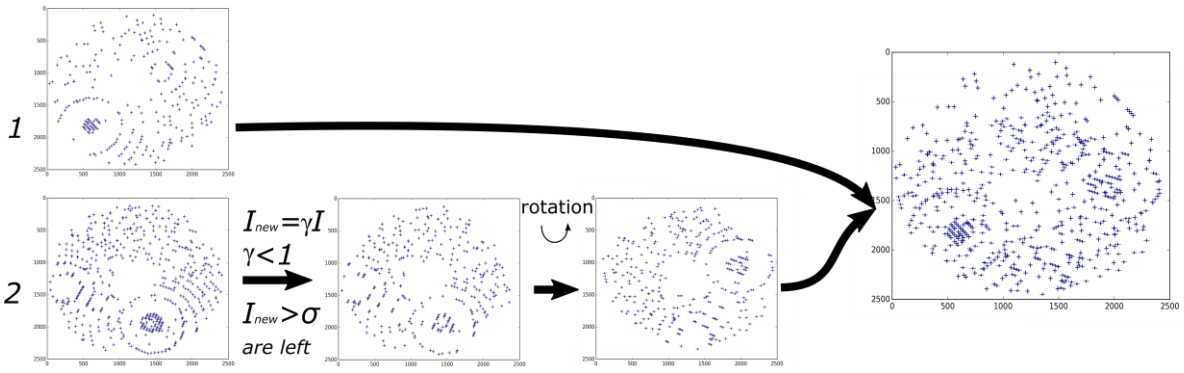


Figure 14. *Simulating multi-crystal diffraction patterns. Two single-crystal patterns (left, 1 and 2) are chosen for combination. An intensity filter (γ < 1) is applied to pattern 2 and only spots of which the resulting intensity is still above background (σ, as determined by Dozor) is retained. Then, the new pattern 2 is rotated around the beam axis by the random angle. Finally, the lists of spot coordinates of patterns 1 and modified pattern 2 are merged to produce a simulated multi-crystal diffraction pattern (right).*

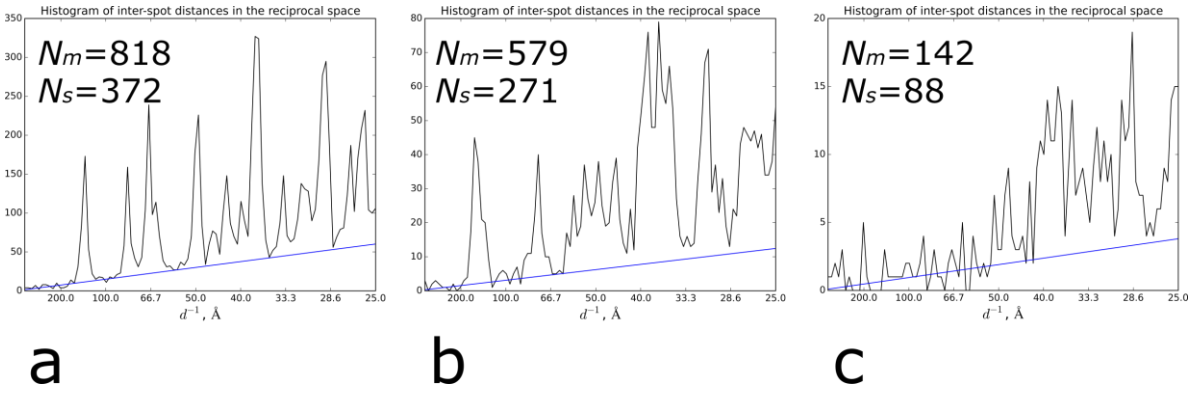


***Figure 15**. The sample inter-spot distance histograms calculated in the multi-pattern simulation experiment based on thermolysin (a), thaumatin (b) and NarQ (c) crystal diffraction patterns. The numbers N_m and N_s in the simulated pattern are presented. The blue lines show the fitted linear approximations of the baselines.*

Dozor was used to produce a list of detector spot coordinates and corresponding reflection intensities for each image. The spot lists of the second image in a pair were then computationally modified by applying an intensity filter (γ in Figure 14) and/or rotating the spots by the random angle around the X-ray beam axis. Such a manipulation allowed simulations of decreased diffraction strength and of randomness in satellite crystal orientation, respectively. The spot lists of both images in the pair were then merged to obtain simulated multi-pattern diffraction images. Inter-spot distance histograms were then generated and the slope of the resulting baselines estimated (Figure 15). As can clearly be seen from Figure 15, the inter-spot histograms produced from the simulated multi-pattern diffraction images all have a baseline with a slope much greater than 0.

The slope values were then normalised by the main crystal surface spot density $\rho = N_m/S$ where N_m is the number of spots in the main crystal pattern (pattern 1) and S the surface of the Ewald sphere cap containing all the detected spots, and plotted against the number of spots N_s in the satellite pattern (pattern 2) remaining after the intensity filter was applied (see Figure 16). As can be seen, the slope of the histogram baselines is clearly proportional to N_s, further validating the correctness of the model outlined in §1.3.1.1.

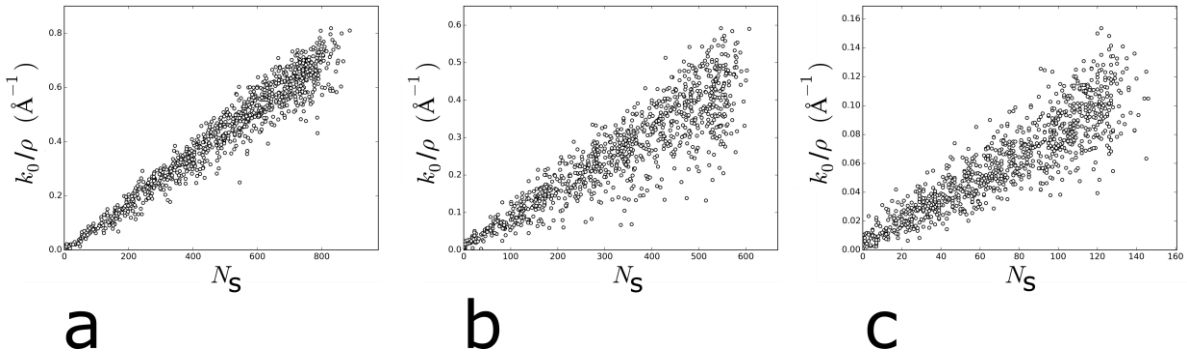


***Figure 16**. The dependence of the slope normalised by the spot surface density k_0/ρ (where $\rho = N_m/S$) from the number of satellite spots N_s in the simulations of multi-pattern diffraction. There are 1000 data points in total; each point represents one simulated pattern. The simulations were based on the data sets from standard rotational data collection from individual crystals of thermolysin (a), thaumatin (b) and NarQ (c).*

Unfortunately, however, using k_0 presented as a determinant of multi-pattern diffraction is not ideal because the exact numbers of spots from each lattice contributing to the diffraction image is usually unknown and only the total number N of spots in the image is available. To account for all this a better measure is

$$\mathrm{K} = \frac{k_0}{\rho N} = k_0 \cdot \frac{S}{N^2} \qquad (6)$$

which is k_0 normalised by spot density N/S and the total number of spots, N, on a diffraction image.

Taking into account, that from (4a):

$$k_0 = 2\Delta d \frac{N_m N_s}{S}$$

where $2\Delta d$ is the proportionality constant, and that in two-crystal case:

$$N = N_m + N_s$$

one can state from (6):

$$\mathrm{K} = 2\Delta d \frac{N_m N_s}{S} \cdot \frac{S}{(N_m + N_s)^2} = 2\Delta d \frac{N_m N_s}{(N_m + N_s)^2} = 2\Delta d \frac{{}^{N_s}/_{N_m}}{\left(1 + {}^{N_s}/_{N_m}\right)^2} \qquad (7)$$

In (7) the dependence of K on the fraction of satellite crystal spots N_s/N_m is shown. This hypothetical behaviour of K was analysed by multi-pattern diffraction simulations. The patterns were again simulated in the manner described above, and the value of K was calculated based on the histogram baseline fit. The results of this analysis are shown in Figure 17 in which the dependence of K on N_s is in accordance with (7) taking into account that the number of spots in the main pattern N_m was nearly constant for each protein.

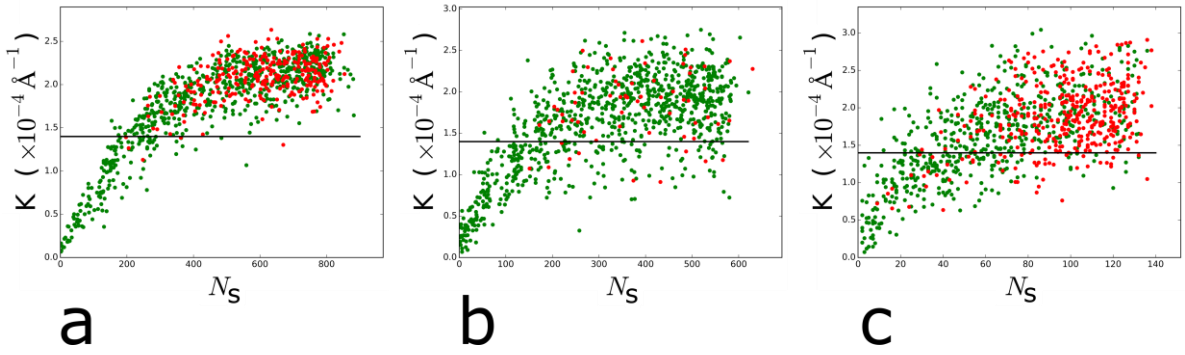


Figure 17*. The dependence of the K value and of the possibility of pattern indexing from the number of satellite spots N_s in the simulations of multi-pattern diffraction. Each point represents one simulated pattern, 1000 points in total. The successful indexing of the simulated pattern is shown by green point, the failed indexing is shown by red point. The simulations were based on the data sets from standard rotational data collection from individual crystals of thermolysin (a), thaumatin (b) and NarQ (c) which had nearly 900, 600 and 150 as the number N_m of spots in the original diffraction pattern, respectively, as determined by Dozor. The black horizontal line indicates the threshold level.*

The relation between K value, the number of spots N_s in the satellite pattern and the possibility to index multi-pattern diffraction images was also examined using *XDS* (Kabsch, 2010). Here, simulated multi-pattern diffraction images were introduced by modification of the spot-list file SPOT.XDS and a total of 5 multi-pattern images (which corresponded to maximum 0.5° of rotation) was used for indexing starting with the image on which the histogram (see Figure 15) was calculated. The results are also presented in Figure 17 which shows, at least for the three crystalline systems on which the simulations were based, that *XDS* successfully indexes multi-pattern diffraction images producing inter-spot histograms with $K < 1.4 \cdot 10^{-4}$ Å^{-1} and this threshold was set for determining at which point multi-pattern diffraction regions of mesh scans should be removed from subsequent data collection

protocols. The threshold chosen corresponds (equation (7), Figure 17) to a two-pattern diffraction image in which the fraction of satellite crystal spots $N_s/N_m = 0.3$.

1.3.1.3. Technical details of the baseline estimation

To construct an inter-spot distance histogram for a given diffraction image the coordinates of the spots on the image are transformed from the detector plane system to three-dimensional reciprocal space, all inter-spot distances are then calculated and the histogram is computed in 100 bins in the boundaries from 1000^{-1} Å^{-1} to 25^{-1} Å^{-1} which cover typical protein unit cell parameters.

The estimate of the baseline is made in every image histogram by linear fit to the data in the chosen baseline regions (Figure 12b, d) which are different for different types of crystals, but the same within the same type of protein crystal lattice. A cumulative histogram from all mesh scan images exhibiting diffraction gives improved accuracy and is used in determination of the baseline regions (see red dots in Figure 12e). Baseline regions are determined in three steps. In the first step the local minima of the cumulative inter-spot distance histogram are found using the algorithm for automatic multiscale-based peak detection (AMPD) (Scholkmann *et al.*, 2012). This algorithm employs a computation of scale-dependent occurrences of local maxima and is used for peak detection in noisy periodic or quasi-periodic signals. Then, an initial baseline of the cumulative histogram is found by linear fitting on the data in local minima. In the second step the baseline of the cumulative histogram is estimated more precisely by applying weights to the least squares minimisation to downgrade the data points which lie higher than the initial baseline. Finally, baseline regions are determined by those data points lying in the cumulative histogram below its baseline or not more than one standard deviation (of these data points from the baseline) higher its baseline.

1.3.2. Discriminating between distinct crystals using hierarchical clustering

Following a mesh scan, once *MeshBest* has identified and discarded all images with unacceptable multi-pattern diffraction every pair of the remaining images is then analysed to find how similar their diffraction patterns are. This is done by comparing the positions of the common diffraction spots on the two images. To be independent of instrument parameters, the angle between two reflections (or between their scattering vectors) is taken as the criterion for position similarity. The image with the smaller number of spots is taken as a reference image. We define the distance score *D* between two images as the following (8):

$$D = \sqrt{\frac{\sum_{i=1}^{N} \Delta_i^2}{N}} \qquad (8)$$

where N is the number of spots in the reference image, Δ_i – the angle expressed in degrees between the centres of the ith spot in reference image and the corresponding spot in the compared image.

For every diffraction spot in the reference image, the corresponding spot in the second (compared) image is searched for in an angular vicinity of 0.1° of the same coordinates as in the reference image. The angular deviation between centre positions of the two spots is then calculated. If no spot is found within 0.1° of the reference spot Δ_i is set to 0.1°. Finally, the distance score, which is independent of the numbers of spots in the two images being compared, is calculated as the root mean square value of the spot centre deviations between the two images (D, equation (8)).

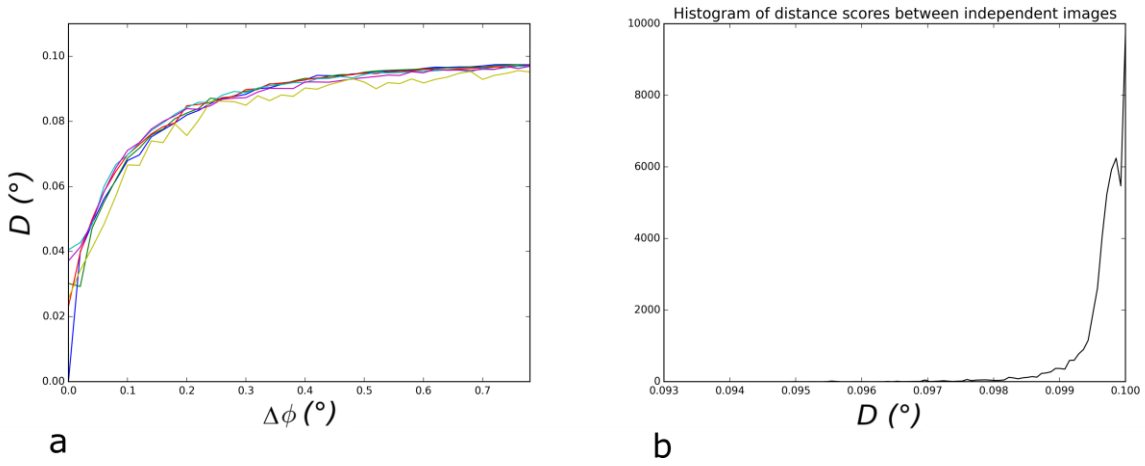


Figure 18. a. *The behaviour of the distance score D as a function of the angular difference in crystal orientations ($\Delta\phi$). In this experiment several standard fifty-frame data collections were carried out from the same crystal of thaumatin using the same data collection parameters except beam transmission which was changed to imitate different diffraction strength. D was then calculated between pairs of images and plotted against ($\Delta\phi$). The average crystal mosaicity was 0.04° as determined by XDS (Kabsch, 2010). Different colours show the results with different X-ray beam transmissions (from 100%, blue, to 0.1%, yellow).* ***b.*** *The histogram of the distance scores obtained between randomly selected images of the same rotational data set. Several data sets were used with different crystal mosaicities.*

Figure 18a shows how D behaves as a function of the angular difference ($\Delta\phi$) between crystal orientations when pairs of images from a standard rotation data collection are compared. As can be seen, D tends to a value of 0.1° as the orientation of the crystal in the X-ray beam becomes more and more different. Figure 18b shows that the random error in the determination of the distance score can be estimated from the histogram of the distance scores between totally independent (i.e. from non-adjacent images in the rotational data set) diffraction patterns as shown in Figure 18b. The histogram has been calculated based on comparing independent images of the same rotational data set and averaged on different protein crystals with different mosaicities and shows the deviation from the expected (for independent images) value of $D = 0.1°$ due to the random spot position coincidence. For mesh scans, a threshold of 0.093° in D is thus introduced to distinguish diffraction images arising from different crystals. The value of the threshold is chosen as ten standard deviations from the centre of the peak in the histogram of the distance scores between independent images.

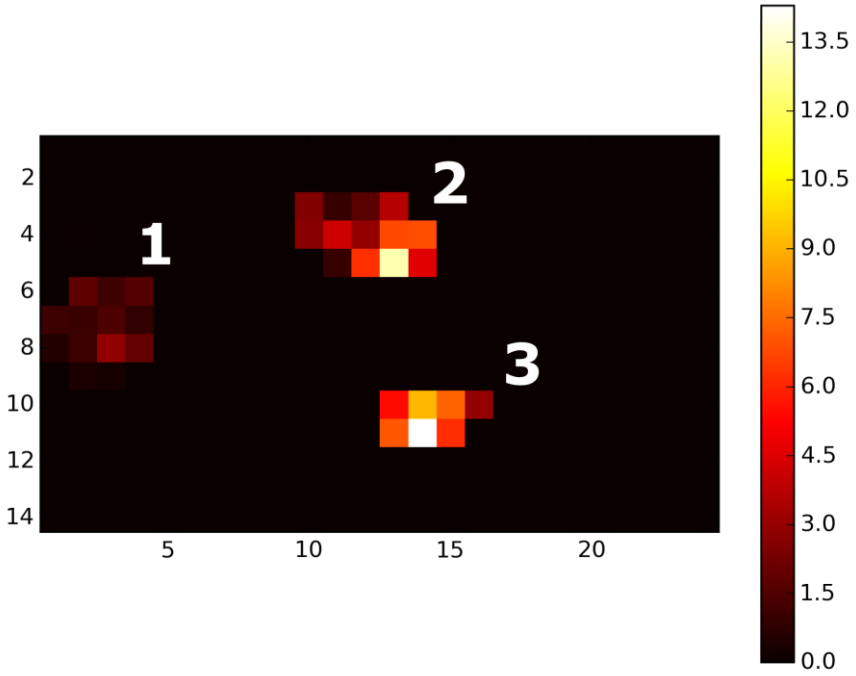


Figure 19. *Sample diffraction heat map of the mesh scan of protein crystals based on the Dozor-score (indicated on colour bar). The mesh scan regions 1, 2 and 3 are not connected in terms of diffraction signal and therefore belong to different crystals.*

MeshBest searches for the regions of the mesh scan that are interconnected in terms of diffraction signal. Those regions that are not connected obviously are of different crystals (see Figure 19). The comparison of images is thus done only inside each interconnected region. The values of D (equation (8)) for every pair of images are then accumulated in the upper triangular matrix $L \times L$, where L is the total number of diffraction images in a particular interconnected region of the mesh scan. When the matrix is complete hierarchical cluster analysis (HCA, (Sneath & Sokal, 1962)) is performed. Clustering is performed using the "average" linkage method which takes the averaged distance between the elements of two different clusters as the real distance between them. If the mesh scan images are taken with rotation then the linkages are calculated excluding the scores between the images with crystal orientations farther than 0.5 degrees from each other. This is done because when the sample is rotated during mesh scan diffraction images from the same crystals can be collected at different orientations. This obviously limits the ability of comparing mesh scan regions that are far from each other by the similarity of their diffraction pattern. The final

clusters are formed by applying the distance cut-off of 0.093° determined above to the dendrogram. Subsequently, these images that have been clustered together are treated as from individual crystals.

1.3.3. Creating a crystal map and fitting elliptic crystal shapes

Once the HCA step has been carried out, *MeshBest* estimates the size and shape of each individual crystal region determined (Figure 20 left). This is an essential step for the determination of optimal data collection strategies. For this 2D projections of the positions of individual crystals identified by HCA are approximated by an elliptical shape. However, as different regions of a crystal may have different diffraction signal, additional information concerning this should be taken into account when describing the 2D area of a crystal. For this reason on crystal maps produced by *MeshBest* the *Dozor* scores for each grid point represented on the map are plotted on a third axis perpendicular to the mesh scan plane and the resulting three-dimensional diagram of each crystal approximated using a semi-ellipsoid shape (see Figure 20 right) parameterised by five quantities. The optimization algorithm of differential evolution (Storn & Price, 1997) is used to fit the shape into the diagram region by the least squares minimisation of the residual function (equation (9)).

$$Residue(x_i, y_i) = \alpha \cdot \left(DozorScore(x_i, y_i) - H \cdot \sqrt{1 - X_i^2 - Y_i^2} \right),$$

$$\begin{pmatrix} X_i \\ Y_i \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{a} & \frac{1}{b} & 1 \end{pmatrix} \cdot \boldsymbol{R}(\varphi) \cdot \begin{pmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}$$

$$\text{where } \alpha = \begin{cases} 2, & DozorScore(x_i, y_i) = 0 \\ 1, & otherwise \end{cases} \qquad (9)$$

where x_i, y_i – mesh scan coordinates of i[th] image; $\boldsymbol{R}(\varphi)$ – basis rotational matrix; H, x_0, y_0, a, b, φ – parameters of the ellipse being optimised; when the expression under the square root turns negative it is set to zero.

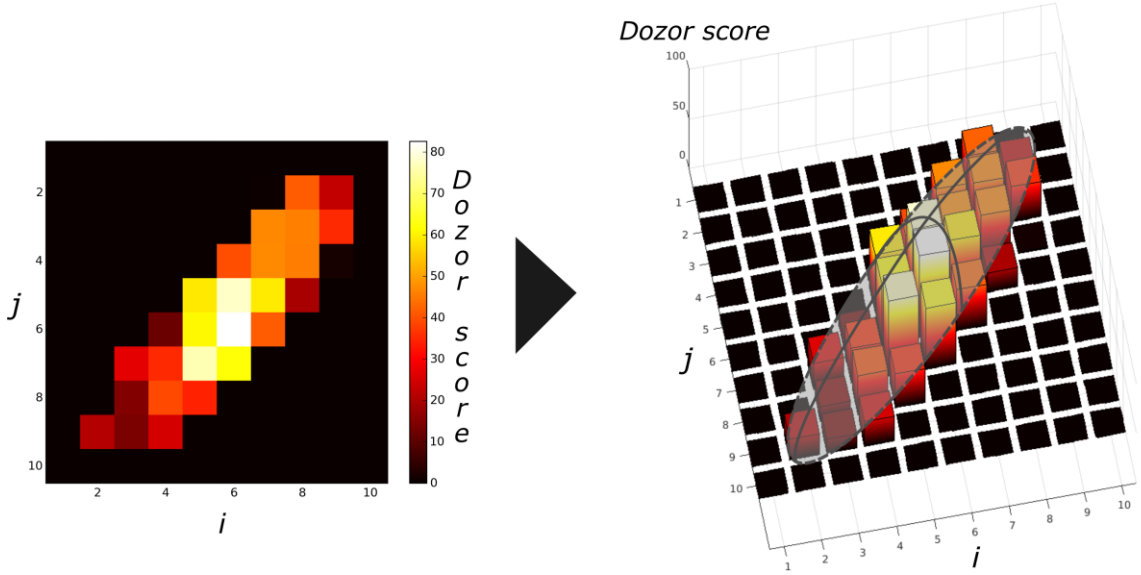


***Figure 20**. The principle of fitting an elliptical shape into the mesh scan map. The mesh scan of a single crystal of NarQ is taken here as an example. Left: a diffraction score (estimated by Dozor) heat map of the mesh scan. Right: A schematic view of a semi-ellipsoid shape fitted to the three-dimensional representation of the map with the Dozor-score indicated on the third axis.*

In this way the characteristics of each crystal region are determined as those of the corresponding semi-ellipsoid including size, centre position and average diffraction strength. The individual crystals identified are sorted in a list according to the integral diffraction signal calculated as the volume of semi-ellipsoid approximating each crystal ($I = \frac{2}{3}\pi abH$). This list can then be used to select the best diffracting crystals for subsequent data collection. To visualise the result *MeshBest* creates a crystal map in which every crystal region is illustrated in one of nine colours tinted according to the diffraction signal (for example see Figure 25g). Use of a standard palette of nine colours allows neighbouring regions of different crystals to be clearly separated by eye. The appropriate ellipses are drawn over each individual crystal region. Those regions of the mesh identified as producing unacceptable multi-pattern diffraction images are coloured in grey.

1.4. Results and discussion

To test the applicability of the algorithms and methodology implemented in *MeshBest* five experiments were carried out (Table 1). These include the *MeshBest* analyses of a large, homogeneous, nearly perfect crystal (§1.4.1); a long crystal with high mosaicity (§1.4.2); a mesh scan performed prior to subsequent multi-crystal data collection (§1.4.3); a mesh scan containing crystals of lysozyme where diffraction was measured only to moderate resolution (~3.7 Å, (§1.4.4)); a mesh scan from a sample holder containing a crystal mess with "dirty" diffraction patterns (§1.4.5). In all cases *MeshBest* demonstrated itself to be a robust tool in sample characterisation for protein crystallography.

Table 1. Mesh scan experiment parameters for experiments aimed at testing the applicability of MeshBest

Case №	1	2	3	4	5
Protein	Trypsin	Thermolysin	NarQ	Lysozyme	Thaumatin
Crystal size, μm	700×70×70	600×120×100	20−100 (range)	~10 (average)	40−100 (range)
Space group	$P2_12_12_1$	$P6_122$	$I2_12_12_1$	$P4_32_12$	$P4_12_12$
Unit cell parameters, Å	a = 62, b = 64, c = 69	a = b = 93, c = 130	a = 40, b = 59, c = 240	a = b = 79, c = 38	a = b = 58, c = 151
Beamline (ESRF)	ID23-1	ID23-1	ID29	ID23-1	MASSIF3
Wavelength, Å	0.972	0.972	1.00	1.85	0.97
X-ray beam size, μm	10×10	30×30	20×20	10×10	15×15
Mesh scan grid dimensions, points	13×22	13×21	26×19	18×37	33×43
Flux, photons/s	$1.6\cdot10^{10}$	$5.6\cdot10^{10}$	$8.6\cdot10^{11}$	$8.0\cdot10^{9}$	$4\cdot10^{11}$
Sample rotation per image, °	0.05	0.14	0.05	0.1	0.02
Detector edge resolution, Å	1.6	1.5	2.5	3.7	1.7
Exposure time per image, ms	37	37	62.4	37	50.6

Table 1 shows the data collection parameters used in these experiments. Data were collected at ESRF beamlines ID23-1, ID29 and MASSIF-3 (Nurizzo *et al.*, 2006; de Sanctis *et al.*, 2012; Theveneau *et al.*, 2013) using X-ray beams with near gaussian profiles and FWHM (denoted as the beam size) ranging from 10 to 50 µm (see Table 1 for details). Prior to data collection, grid locations and sizes were defined manually in the MxCuBE2 beamline control interface (Gabadinho *et al.*, 2010; de Sanctis *et al.*, 2016) with the number of grid points in each direction automatically calculated based on the incident X-ray beam size and allowing for no overlap between adjacent grid points. Mesh scans were then carried out in a shutterless fashion and diffraction images were collected using noise-free Dectris PILATUS 6M or PILATUS3 2M pixel array detectors. In all experiments there was a constraint – defined by the need to trigger detector readout – to perform a slight rotation around Ω axis when collecting each row of the mesh. The program *MeshBest* was implemented as a Python module using standard libraries.

1.4.1.Case 1: A large, homogenous crystal

A large crystal of bovine trypsin (Sigma-Aldrich) was cryocooled and mounted on the beamline. The mesh scan was performed in the area shown in Figure 21a. No signs of multi-crystal diffraction were recognised in the images of this mesh scan (Figure 21b). Based on the 210 interconnected diffracting regions of the mesh scan a distance matrix, 210×210 in size was calculated. As might be expected based on the appearance of the crystal in region of the mesh scan, subsequent HCA (Figure 21c) determined this to be a single crystal and an ellipse was fitted to the crystal shape (Figure 21g).

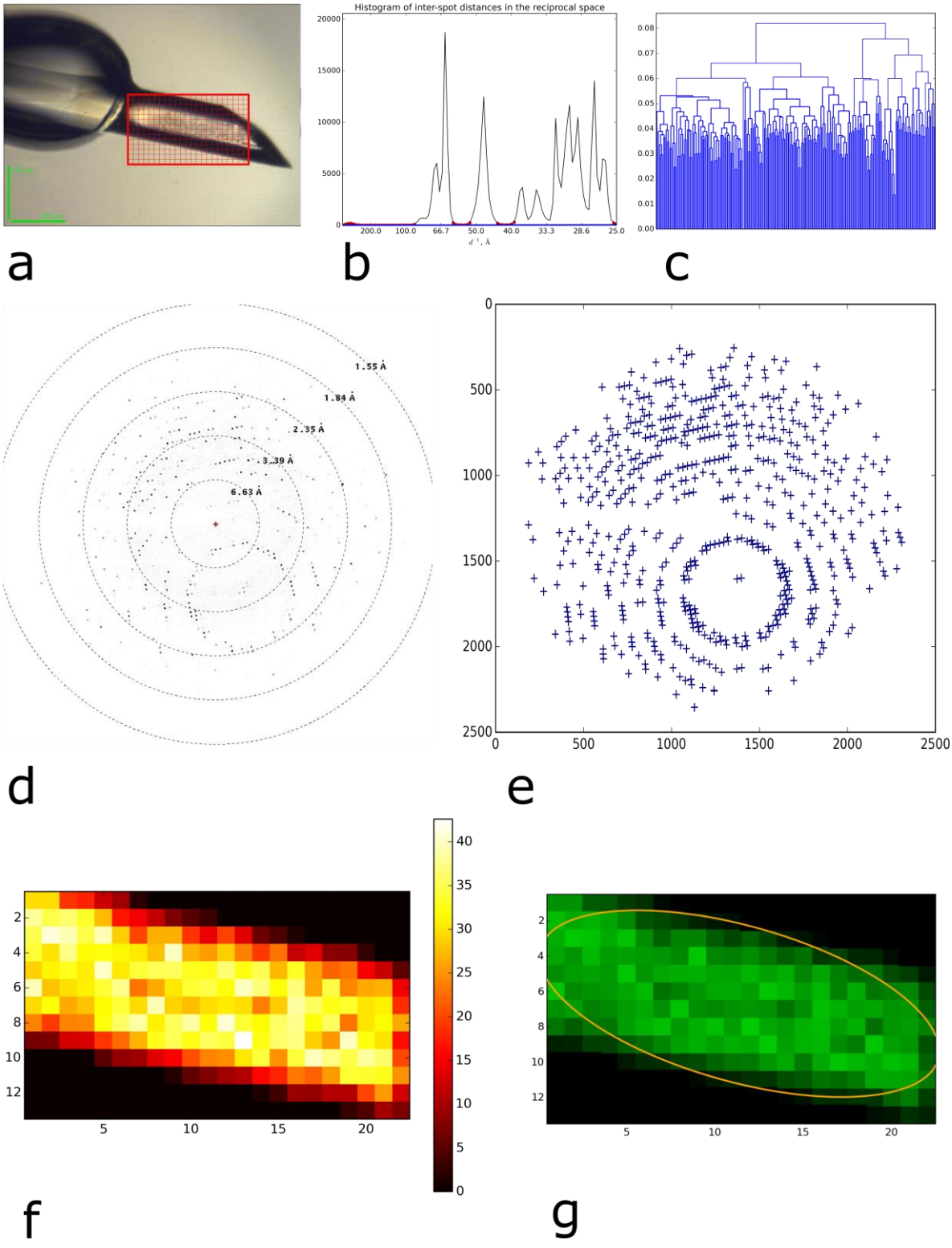


***Figure 21**. MeshBest analysis of the mesh scan of a large, homogeneous, nearly perfect crystal of trypsin. **a.** A snaphot of the crystal as mounted on the beamline goniometer, red rectangular mesh indicates the area and the dimensions of the mesh scan. **b.** Cumulative inter-spot distance histogram used in multi-pattern diffraction analysis with the baseline regions determined marked with red circles. The blue line shows the fitted baseline of the cumulative histogram the slope of which indicates no presence of multi-pattern diffraction in the images of the mesh scan. **c**. Dendrogram based on HCA of mesh scan images (see §1.3.2), the colours correspond to the crystal map in g. **d**. A sample of raw diffraction image from the mesh scan. **e.** A diagram of spots detected by Dozor corresponding to the image in d. **f.** Dozor-score heat map of the mesh scan. **g.** A crystal map of the mesh scan generated by MeshBest.*

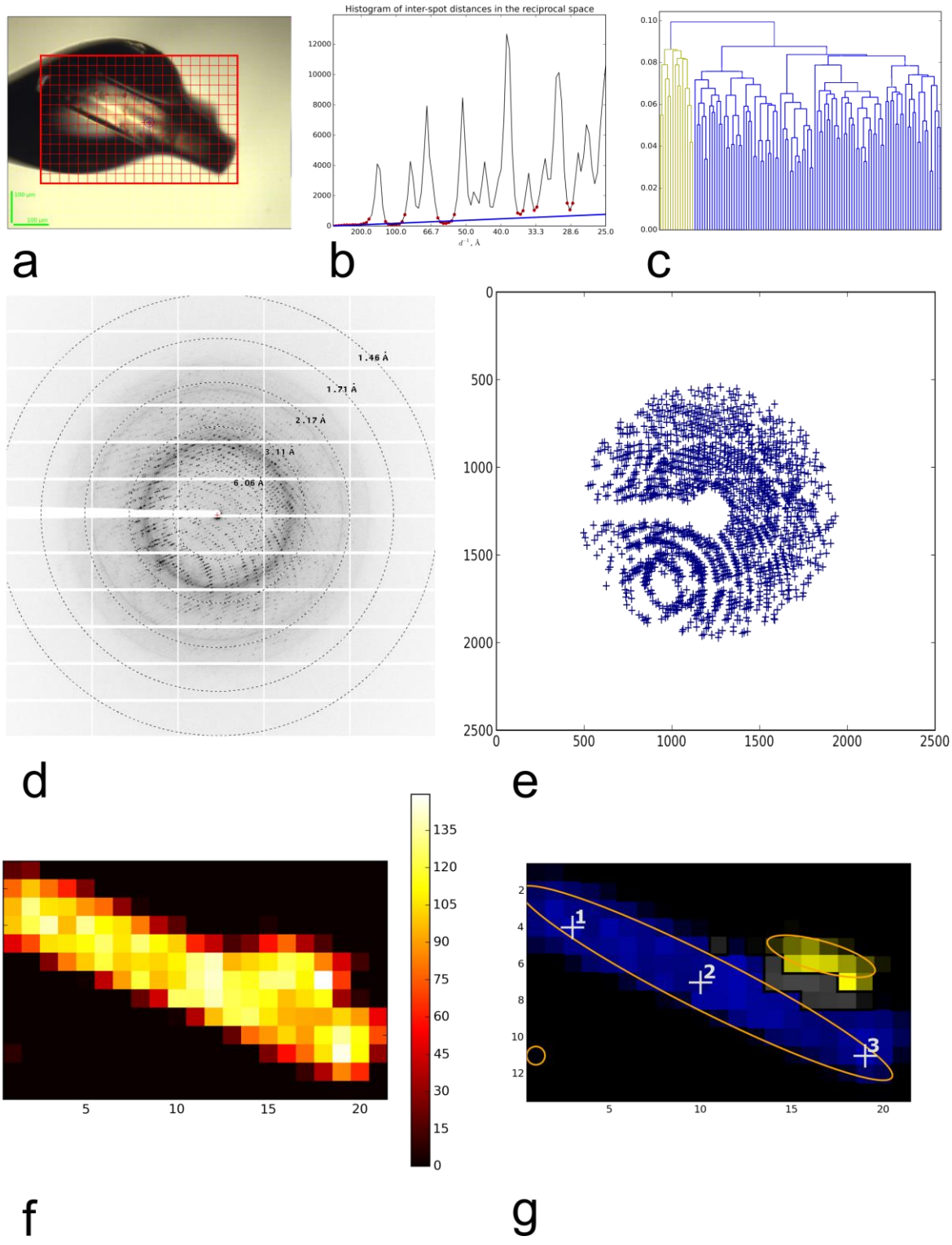


Figure 22. *MeshBest analysis of the mesh scan of large thermolysin crystal.* ***a****. A snapshot of the sample as mounted on the beamline goniometer, red rectangular mesh indicates the area and the dimensions of the mesh scan.* ***b****. Cumulative inter-spot distance histogram used in multi-pattern diffraction analysis with determined baseline regions marked with red circles. The blue line shows the fitted baseline of the cumulative histogram the slope of which indicates the presence of multi-pattern diffraction in the images of the mesh scan.* ***c****. Dendrogram based on HCA of mesh scan images (see §1.3.2), the colours correspond to the crystal map in g.* ***d****. A sample of raw diffraction image from the region of the mesh scan with multi-pattern diffraction (shown in grey in (g)).* ***e.*** *A diagram of spots detected by Dozor corresponding to the image in d.* ***f.*** *Dozor-score heat map of the mesh scan.* ***g.*** *A crystal map of the mesh scan generated by MeshBest. Numbered white crosses show the positions at which the partial data sets were collected to determine lattice orientations.*

1.4.2.Case 2: a large crystal with lattice disorientation

A large crystal of thermolysin (from *Bacillus thermoproteolyticus*, Sigma-Aldrich) was analysed with a mesh scan as shown in Figure 22. The inter-spot distance distribution analysis found multi-pattern diffraction in the region (grey in Figure 22g) between the principal crystal and, most likely, a small satellite subsequently identified (in yellow, Figure 22g). One main interconnected region of diffraction was found in the mesh scan area (neglecting very small crystal below the main one). The distance matrix 115×115 in size was calculated and used to perform HCA which indicated the presence of two different crystals in the sample holder (Figure 22cg).

Figure 22a shows that the mesh scan area used to analyse the diffraction properties of the crystal of thermolysin includes a section of crystal within the sample loop and a section outside the loop. Anecdotal evidence suggest that in such cases there is often a misorientation between the two ends of the crystal due to the bending action of the loop. To examine whether *MeshBest* had not taken into account such a misorientation, different parts of the main crystal a series of partial data collections were carried out at different positions, shown in Figure 22g, on the main crystal identified. Fifty images of 0.1° rotation data were collected at each position and the crystal orientation in the X-ray beam was determined using *XDS* (Kabsch, 2010). Figure 23 shows the orientations of the **c*** vector, expressed in Euler angles, at the different crystal positions tested. Clearly, the crystal has a discrepancy in lattice orientation between the two ends of about 1°. However, Figure 23 also shows that this misorientation is gradual over the length of the crystal which is thus better considered one large single crystal rather than two or three smaller ones.

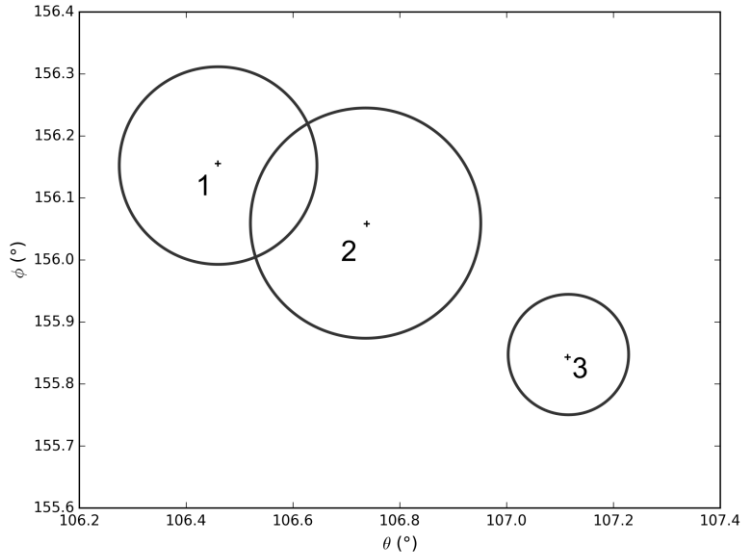


Figure 23. *A diagram of orientations of c* vector of the reciprocal lattice in different regions of the thermolysin crystal shown in Figure 22. The orientations are presented here by two Euler angles with the values calculated from the data collected at the positions given by corresponding numbers in Figure 22g. The precision is denoted by circles with the radius being equal to the standard deviation of the spindle position.*

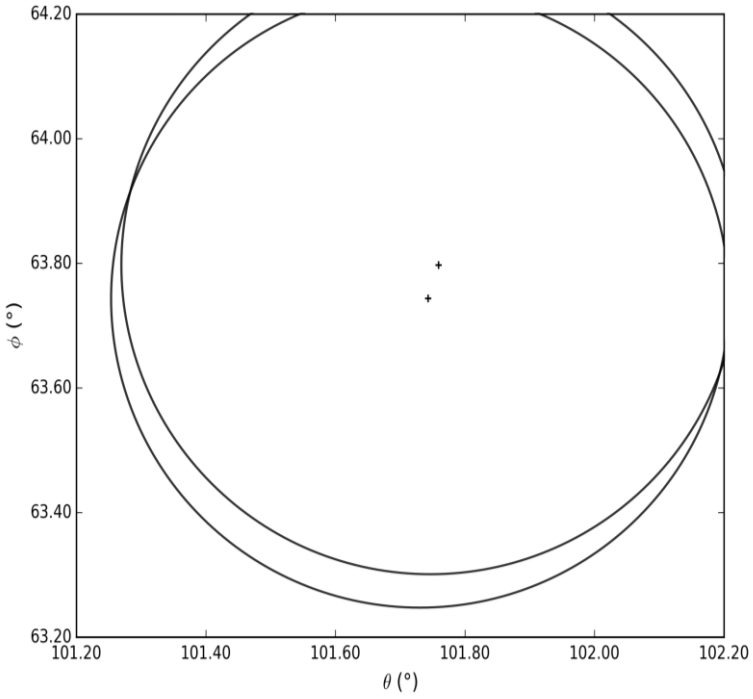


Figure 24. *A diagram of orientations of c* vector of the reciprocal lattice in two different regions of the trypsin crystal shown in Figure 21. The orientations are presented by two Euler angles with the values calculated from the diffraction images of the mesh scan: first 10 images in row 3 (see Figure 21g) were taken for calculation of the orientation for the point 1 on the graph and last 10 images in row 10 (see Figure 21g) were taken for calculation of the orientation for point 2 on the graph. The precision here is denoted by circles with the radius being equal to the standard deviation of the spindle position which was ~0.5°.*

The discrepancy in the crystal orientations between the parts of trypsin crystal from §1.4.1 was also examined. The orientations were determined with *XDS* based on the images of the mesh scan itself. The difference in the lattice orientation in that case did not exceed 0.06° with the precision of 0.5° confirming that trypsin crystal was homogeneous.

1.4.3. Case 3: Multi-crystal data collection in the MeshAndCollect pipeline

Crystals of recombinant transmembrane construct of nitrate/nitrite-dependent histidine kinase NarQ of *E.coli* (R50K mutant, (Gushchin *et al.*, 2017)) were obtained *in meso* as detailed in (Gushchin *et al.*, 2017). Crystals were harvested in a micromesh sample holder (Figure 25a) and a mesh scan was performed. Around 4% of images in this mesh scan were found to be affected by multi-crystal diffraction, however in all cases this was below the threshold determined in §1.3.1. The mesh scan area consisted of two large and two small diffraction regions that were not connected to each other (see Figure 25f). *MeshBest* divided these areas into 41 individual crystal regions (Figure 25g in different colours). The mesh scan was then characterised for subsequent multi-crystal data collection by retrieving the dimensions and the centre coordinates of each crystal by elliptical shape approximations.

Partial crystallographic data sets were collected from NarQ crystals similarly to (Zander *et al.*, 2015) from the best positions ranked by integral diffraction score given by *MeshBest*. Each data set was collected using a beam aperture in accordance with the size of diffracting area targeted (white crosses in Figure 25g denote the beam size used at each position). Where an identified crystal was significantly larger than a particular beam size, more than one partial data set was collected at different positions on the crystal (i.e. crystals

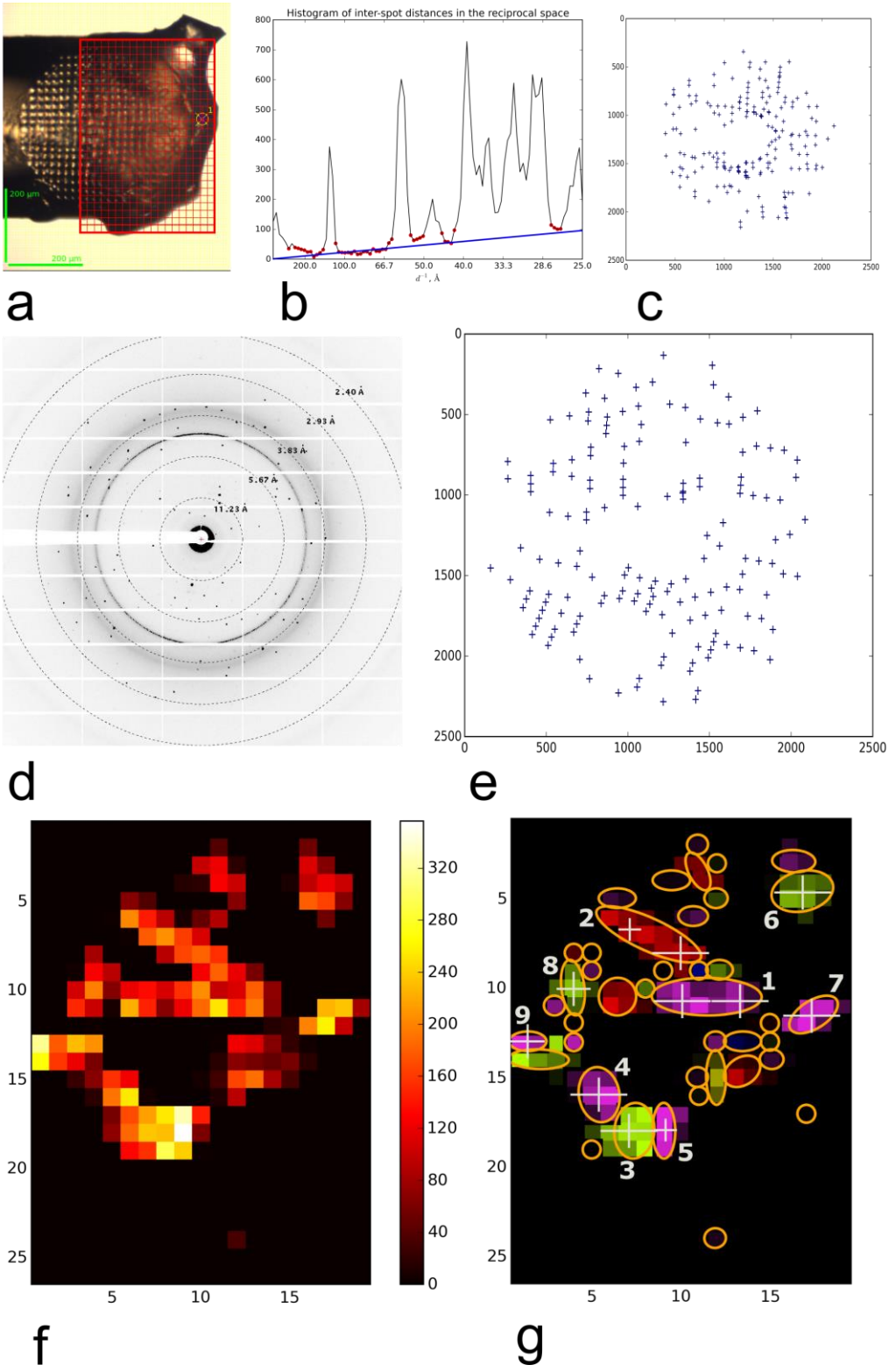


Figure 25. *MeshBest analysis of the mesh scan of NarQ crystals for subsequent multi-crystal data collection* ***a.*** *A snapshot of the sample as mounted on the beamline goniometer, red rectangular mesh indicates the area and the dimensions of the mesh scan.* ***b.*** *Cumulative inter-spot distance histogram used in multi-pattern diffraction analysis with baseline regions marked with red circles. The blue line shows the fitted baseline of the cumulative histogram the slope of which indicates the presence of multi-pattern diffraction in the images of the mesh scan.* ***c.*** *A diagram of spots detected by Dozor in the sample image with small portion of multi-pattern diffraction (below the threshold for detection).* ***d.*** *An example of a raw diffraction image from the mesh scan.* ***e.*** *A diagram of spots detected by Dozor corresponding to the image in d.* ***f.*** *Dozor-score heat map of the mesh scan.* ***g.*** *A crystal map of the mesh scan generated by MeshBest. The crystals used for data collection are numbered according to their diffraction score rank. White crosses with a size of beam aperture mark the positions on which the partial data sets were collected.*

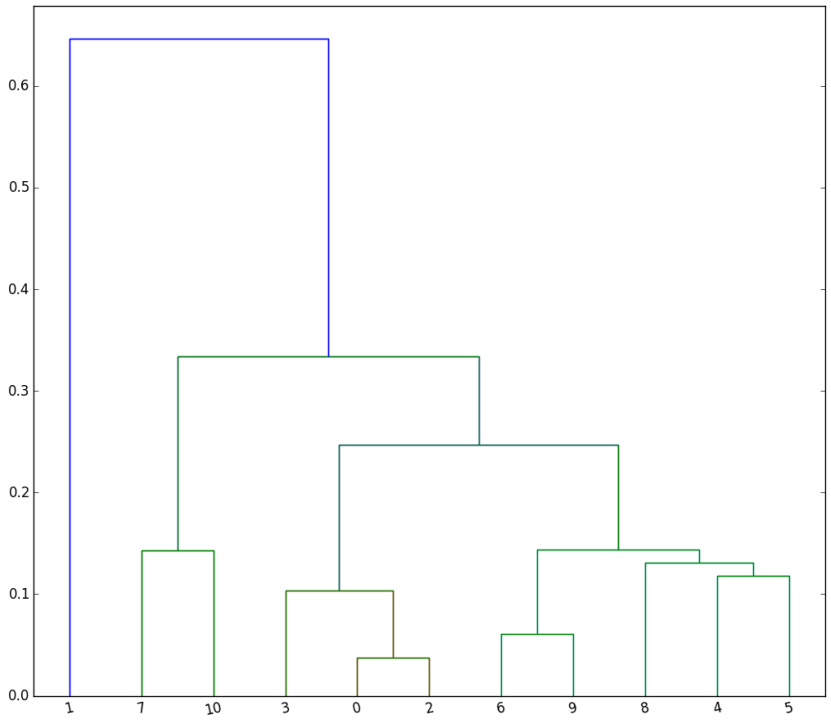


Figure 26. *Dendrogram of the HCA carried out on the partial data sets collected from NarQ crystals at the positions shown in Figure 25. The final merged data set was composed of the cluster shown in green.*

Table 2. Data collection and processing for the production of a MeshBest-guided Mesh&Collect dataset from crystals of the R50K mutant of nitrate/nitrite-dependent histidine kinase NarQ of E. coli. Values for the outer resolution shell are given in parentheses.

Diffraction source	ESRF ID29 MX beamline
Wavelength, Å	1.00
Detector	Dectris Pilatus 6M
Crystal-detector distance, mm	378.8
Rotation range per image, °	0.1
Total rotation range (per partial data set), °	20
Exposure time per image, s	0.05
Number of merged partial data sets	10
Space group	$I2_12_12_1$
a, *b*, *c*, Å	40.0, 59.7, 239.2
α, β, γ, °	90, 90, 90
Resolution range, Å	60.0-2.0 (2.1-2.0)
Total No. of reflections	138753 (19314)
No. of unique reflections	34226 (4734)
Completeness, %	92.1 (93.4)
Redundancy	4.05 (4.08)
$\langle I/\sigma(I)\rangle$	7.37 (1.13)
$CC_{1/2}$, %	99.1 (73.2)
$R_{r.i.m.}$, %	8.4 (203.3)
Overall *B* factor from Wilson plot, Å^2	61.4

1 and 2 in Figure 25g). In total, 11 partial sets of 20°-rotation data (φ in the range from -10° to 10°) were acquired. The diffraction images from each partial data set were then processed with *XDS* (Kabsch, 2010) and HCA (Figure 26) (Giordano *et al.*, 2012) applied to decide which partial data sets to merge (XSCALE; (Kabsch, 2010)) to produce a final data set to d_{min} = 2.0 Å, the statistics for which are presented in Table 2.

1.4.4. Case 4: *MeshBest* at moderate diffraction resolution

Crystals of hen egg-white lysozyme (Roche) of a size of 10 µm were harvested in a micromesh sample holder and a mesh scan was performed (Table 1; Figure 27) by Ulrich Zander.

The recorded diffraction of the crystals was limited by the detector position to the maximum resolution (at the edge of detector) of 3.7 Å (Figure 27c); thus this experiment could be considered as a test of the performance of *MeshBest* at low resolution as well as demonstrating how *MeshBest* performs when there are relatively few diffraction spots on the images being analysed. It was determined that the mesh scan region consisted of 78 different groups/crystals (Figure 27f) with few crystals stacked together and producing multi-pattern diffraction. Every particular crystal region was approximated with an elliptical shape.

1.4.5. Case 5: The case of a crystal mishmash

Thaumatin from *Thaumatococcus daniellii* (Sigma-Aldrich) crystals were harvested in a micromesh container in a messy way and then subjected to a mesh scan (Figure 28) by Ulrich Zander.

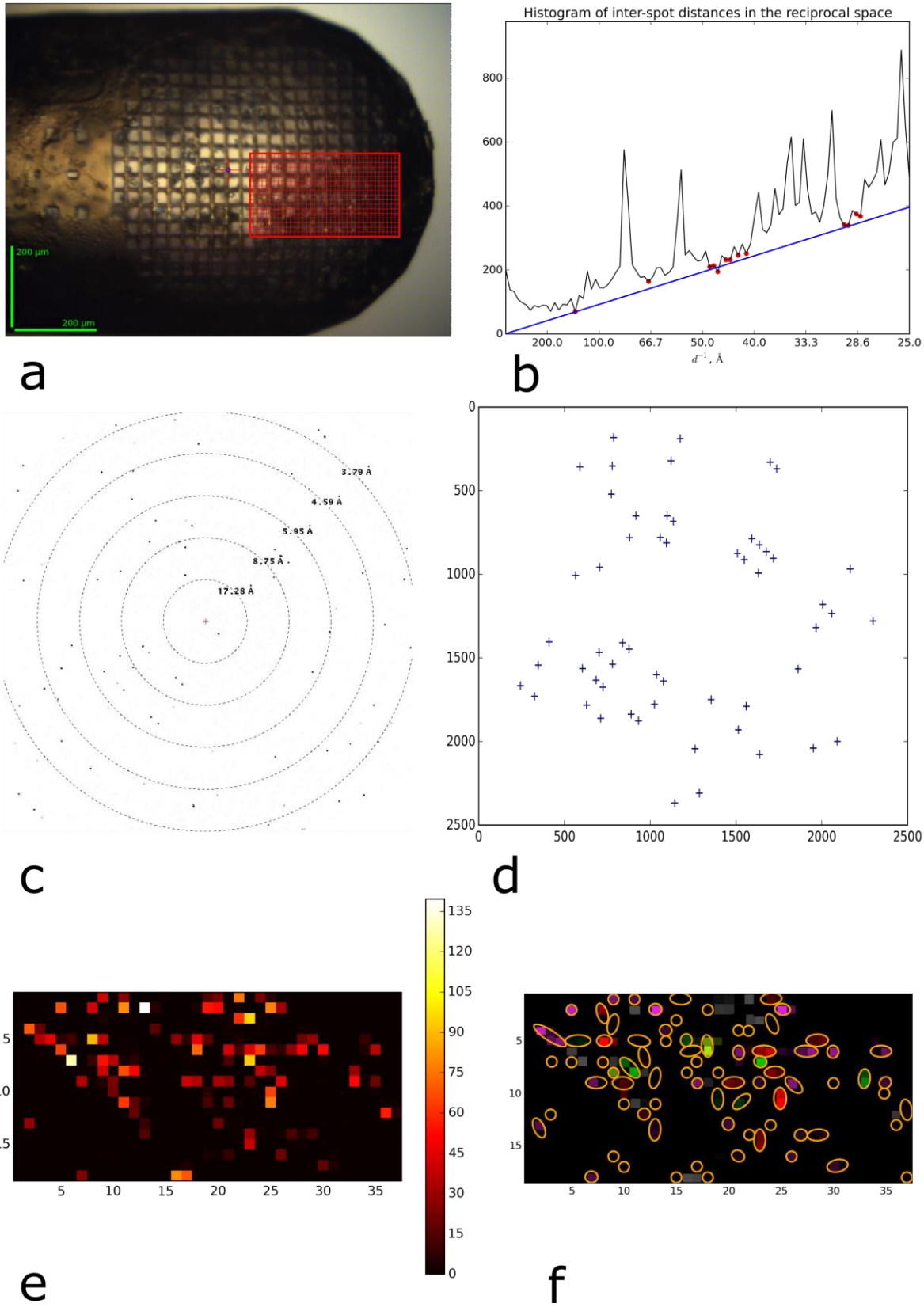


Figure 27. *MeshBest analysis of the mesh scan of lysozyme crystals.* ***a.*** *A snapshot of the sample holder as mounted on the beamline goniometer, red rectangular mesh indicates the area and the dimensions of the mesh scan.* ***b.*** *Cumulative inter-spot distance histogram used in multi-pattern diffraction analysis with baseline regions marked with red circles. The blue line shows the fitted baseline of the cumulative histogram the slope of which indicates the presence of multi-pattern diffraction in the images of the mesh scan.* ***c.*** *An example of a raw diffraction image from the mesh scan.* ***d.*** *A diagram of spots detected by Dozor corresponding to the image in c.* ***e.*** *Dozor-score heat map of the mesh scan.* ***f.*** *A crystal map of the mesh scan generated by MeshBest.*

Many of the diffraction images collected during the scan were polluted by intensive salt ring sectors, reflection splitting, the superposition of diffraction patterns from many crystals and other artefacts of unclear origin (Figure 28e, g). From the diffraction heat map itself (Figure 28c) it was unclear which regions would be suitable for diffraction data collection. As might be expected, based on image spot lists produced by *Dozor* and in which regions originating from salt or crystalline ice had been removed, *MeshBest* analysis revealed (and a steep slope of the baseline of the cumulative inter-spot distance histogram in Figure 28b indicate that) a large area in the central region of the sample holder (grey in Figure 28d) containing almost exclusively images affected by superposed multi-crystal diffraction. The images from individual crystals with similar diffraction patterns were then clustered together into 37 different crystal groups (see Figure 28d in different colours). The dimensions and centre coordinates of each crystal were then estimated by elliptical shape approximation.

1.4.6. Discussion

The results presented above clearly show that the *MeshBest* analysis of X-ray mesh scans provides reliable information about the positions, sizes and relative diffraction strengths of crystals of macromolecules mounted in the sample holder. Such information is critical for the proper organization and design of subsequent diffraction data collection protocols. Where *MeshBest* indicates that the sample holder contains one (or relatively few) crystal(s) the assessment of crystal size would allow a more precise description of radiation damage effects arising during measurements when the size of the sample is smaller than X-ray beam (Zeldin *et al.*, 2013) and thus define more realistic and correct strategy of data collection (Bourenkov & Popov, 2010). In cases where *MeshBest* indicates that large crystals

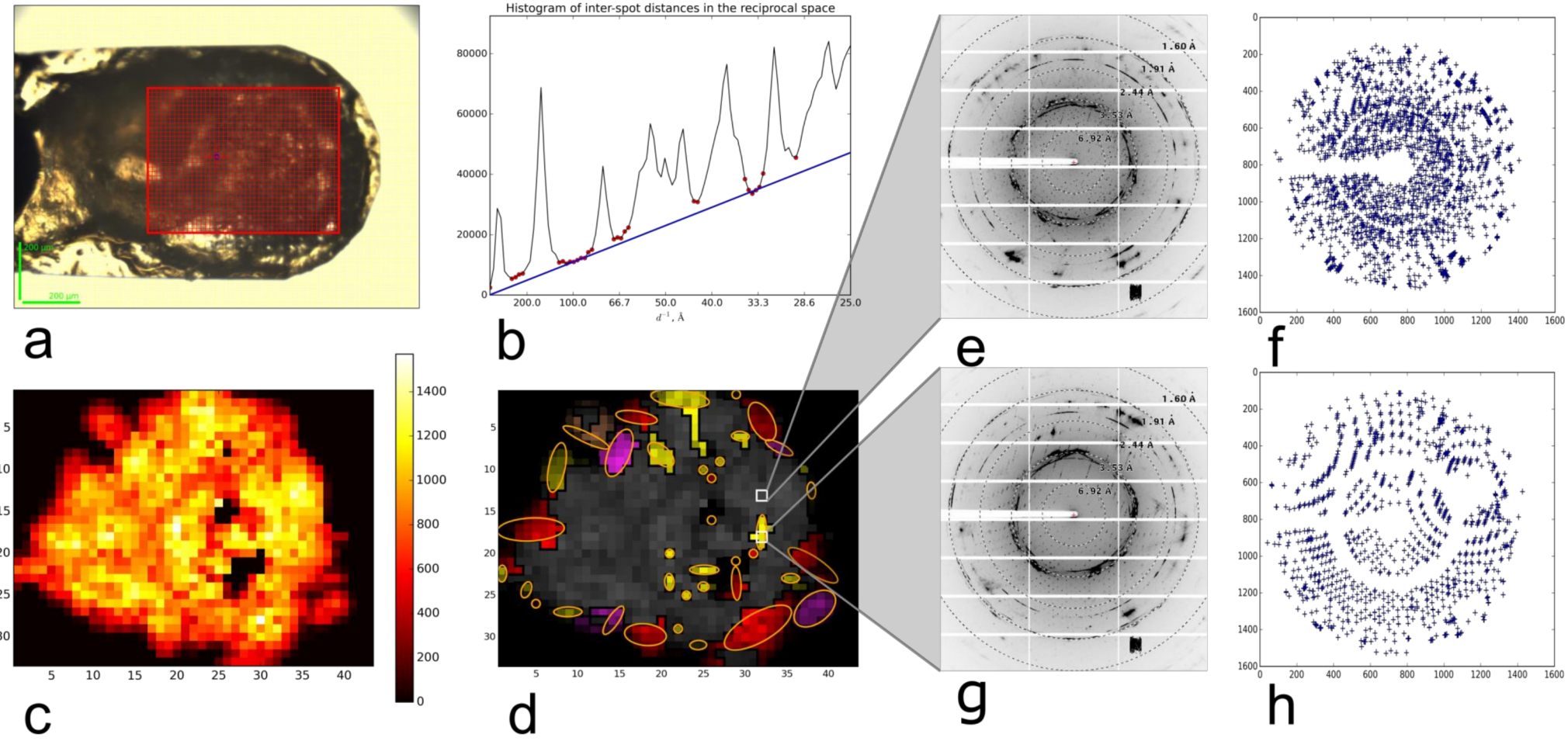


Figure 28. *MeshBest analysis of the mesh scan of thaumatin crystals harvested in a way which produces a crystal 'mismash'.* ***a.*** *A snapshot of the sample holder as mounted on the beamline goniometer, the red rectangle indicates the area of the mesh scan.* ***b.*** *Cumulative inter-spot distance histogram used in multi-pattern diffraction analysis with determined baseline regions marked with red circles.* ***c.*** *Dozor-score heat map of the mesh scan. The blue line shows the fitted baseline of the cumulative histogram the slope of which indicates the presence of multi-pattern diffraction in the images of the mesh scan.* ***d.*** *A crystal map of the mesh scan generated by MeshBest. Two images from the positions indicated by white rectangles on the map are shown in* ***e*** *and* ***g****. Their spot diagrams are shown in* ***f*** *and* ***h*** *accordingly. The image in e is sampled from multi-crystal diffraction zone (grey on the map) whereas the image in g is sampled from the region with single-crystal diffraction.*

are essentially homogenous (i.e. (§1.4.1, §1.4.2) then such an approach combined with 'helical' style data collections (Flot *et al.*, 2010) might be the preferred mechanism of optimising data quality. Here though, the analysis presented in §1.4.2 shows that care should be exercised to define a protocol in which the direction of the helical scan and or the beam size used avoids illuminating satellite crystals and the production of multi-pattern diffraction images.

In cases where *MeshBest* indicates the presence of multiple small crystals packed tightly together on the sample support, data collection might be best carried out by accumulating many partial data sets from different individual crystals with subsequent HCA-guided merging producing complete data sets (i.e. (Zander *et al.*, 2015)). Here, in order to avoid noise from the diffraction patterns of adjacent crystals or from the medium in which the crystals are mounted care should be taken to adjust incident beam size to crystals size. Indeed, it can be imagined that future versions of *MeshBest* will pass information concerning crystal size directly to beamline control interfaces in order to further help in the automation of the optimisation of diffraction data collection in MX. In such experiments the ability of *MeshBest* to separate areas of single-crystal and multi-crystal diffraction based on the threshold (§1.3.1.2) is also important: processing diffraction images collected in the latter regions can be problematic for data processing packages, however, those images when indexable can provide useful intensity information, and thus should not always be excluded from potential data collection protocol.

However, multi-pattern diffraction images should not be associated with single crystal region on the map though. The reason for this is that including them would break the logic of the pattern similarity-based HCA algorithm aimed at finding distinct crystal

areas (see §1.3.2, as multi-pattern images in terms of defined distance metric could be proximal to several clusters/crystals and in that case violate the hard-clustering linkage). Nevertheless, the protocol for data collection strategy calculation must ensure that diffraction data from those areas where multi-crystal diffraction is above the threshold can be included in data collection protocols whenever this might be necessary.

2. Experimental phasing for crystal structure determination of membrane proteins

2.1. Summary

Halide-SAD is a promising method to obtain the phases for crystal structures of soluble proteins. Based on the common for MPs 'positive-inside' rule it was discovered that MP crystal derivatisation with iodide can be a potentially universal tool for phasing MP crystal structures via iodide-SAD (I-SAD). The hypothesis of the universality was tested on four different target MP – representatives of different species. The crystals of these proteins were derivatised employing the protocol of cryo-soaking in iodide-containing solutions. In all cases the anomalous data collected were sufficient for successful MP structure solution. At the same time, an alternative approach of Br-SAD was tested and none of the attempts at structure solution were successful. In this chapter the work on halide soaking attributed to the solution of crystal structures of MP via I-SAD and Br-SAD is presented and discussed.

2.2. Introduction

Although, once X-rays were first described by W. Röntgen and then proved, by Laue and the Braggs, to be diffracted by the periodically arranged molecules in a crystal, it took almost fifty years to determine the first crystal structure of a macromolecule, X-ray crystallography has evolved as the prominent method of elucidating the 3-D structures of biological molecules. The first protein structures of myoglobin and haemoglobin were solved by X-ray diffraction by Kendrew and Perutz in the late 1950s (Kendrew *et al.*, 1958; Perutz *et al.*, 1960). Since then macromolecular crystallography (MX) has been used to determine thousands of crystal structures of proteins, nucleic acids and their complexes, revolutionising our understanding of Nature.

Since the 1990s, the original methodology employed in MX experiments has evolved extensively. Modern, high-brilliance synchrotron X-ray beamlines produce fluxes greater than 10^{12} photons/s in a highly-collimated, micron-sized beam (i.e. ID23-1, ESRF; (Nurizzo *et al.*, 2006)). Such beamlines are equipped with high-precision goniometers (Brockhauser *et al.*, 2013) and noise-free pixel detectors (Henrich *et al.*, 2009) which allow the recording diffraction patterns even from micron-sized crystals of poor quality. Moreover, advances in automation (Svensson *et al.*, 2015) and diffraction image processing (Kabsch, 2010) during past two decades has transformed how MX experiments are carried out.

$$\rho(x,y,z) = FT^{-1}[\boldsymbol{F}_{hkl}] = \frac{1}{V}\sum_{h,k,l}^{\infty} \boldsymbol{F}_{hkl} \cdot e^{-2\pi i(hx+ky+lz)}$$

$$= \frac{1}{V}\sum_{h,k,l}^{\infty} |\boldsymbol{F}_{hkl}| \cdot e^{i\varphi_{hkl}} \cdot e^{-2\pi i(hx+ky+lz)} \qquad (10)$$

In order to obtain the electron density distribution $\rho(x, y, z)$ – and thus the structure – of molecules making up a crystal one must use a Fourier transform (equation (10)). However, in order to calculate this one must know both the amplitude and the phase of the structure factor $\boldsymbol{F}_{hkl}$ ($|\boldsymbol{F}_{hkl}|$ and φ_{hkl}, respectively, in equation (10)). However, while an X-ray diffraction experiment provides information as to the amplitude of $\boldsymbol{F}_{hkl}$, it provides no information as to the phase and this must be retrieved indirectly. Retrieving the correct value of the phase for each $\boldsymbol{F}_{hkl}$ is commonly known as the phase problem and solving this can, in some cases be very challenging.

Three main approaches that have been developed to solve the phase problem: direct methods; molecular replacement; methods implying the introduction of a substructure of heavy atoms into a crystal.

Direct methods use implicit relationships (Karle and Hauptman, 1956; Sayre, 1952) within the phase values of the structure factors, such as triplet relations, to make an initial estimate of the phases. The resolution limit for their implementation is named Sheldrick's rule: direct methods are generally only successful if more than a half of the full number of possible reflections in the resolution shell of 1.1-1.2 Å are measured well enough ($F > 4\sigma(F)$) (Sheldrick, 1990). This is usually the case for small molecule crystals exhibiting high order. However, crystals of biological macromolecules rarely diffract to 'atomic resolution' and thus, direct methods are hardly ever applied in protein crystal structure solution.

The Molecular Replacement (MR, (Rossmann, 1990), Figure 29) method for solution of the phase problem requires a pre-existing model – usually the crystal structure of a protein with a high amino acid sequence similarity to the protein under study. This model (the 'search model') is then placed at the right position in the crystal lattice of the target

protein by rotation and translation while, at the same time, searching for the best correlation of Patterson functions of the model structure and that calculated from structure factors measured from the target crystals. MR is the most popular method for solution of the phase problem in MX due to the increasing number of solved crystal structures which can be used as search models. However, MR has a major drawback in that the phases derived are initially biased by the search model which can lead the solution to converge to a local minimum, i.e. to an incorrect electron density map.

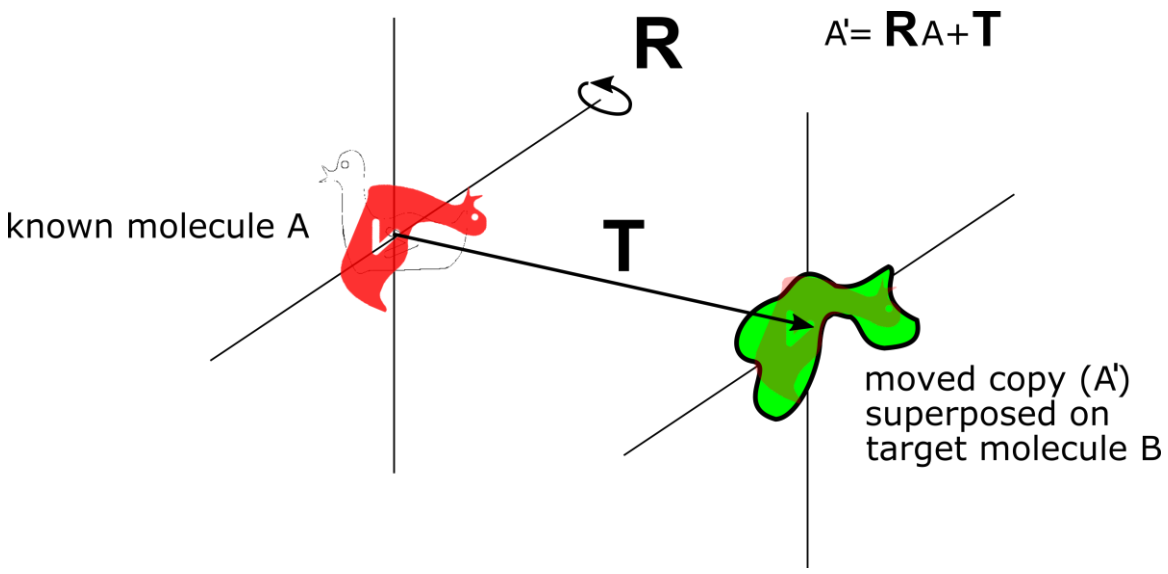


Figure 29. *The principle of molecular replacement (MR). Molecule A with known structure (red duck) shows high sequence similarity to target molecule B (in green). Upon rotation R and translation T molecule A is moved to the position which is held by B. The calculation of Patterson map correlation would indicate how well A at this position reflects the unknown structure (B).*

When MR protocols are not successful or suitable search models are not available, experimenters must resort to experimental (*de novo*) approaches for solving the phase problem. These usually rely on the introduction of a substructure of heavy atoms into a crystal. This is known as heavy atom derivatisation and there are a number of ways this can be done: soaking a crystal in heavy atom-containing solution (Morth *et al.*, 2006); co-crystallisation in the presence of heavy atoms (Morth *et al.*, 2006); modifying residues of protein polypeptide chain (i.e. substitution of methionine for selenomethionine (Walden,

2010)). The basis of experimental methods is the determination of the resulting heavy atom substructure in the crystals. This is done by the use of diffraction intensity differences caused by heavy atom incorporation.

The resulting structure factors $\boldsymbol{F}_{PA}$ for a crystal structure with heavy atoms added are calculated by simple summation of native structure factors $\boldsymbol{F}_{P}$ and the structure factors of heavy atom substructure $\boldsymbol{F}_{A}$ (see equation (11)).

$$\boldsymbol{F}_{PA} = FT[\rho_{PA}] = FT[\rho_P + \rho_A] = FT[\rho_P] + FT[\rho_A] = \boldsymbol{F}_P + \boldsymbol{F}_A \qquad (11)$$

If the exact positions of the heavy atoms in the unit cell of the crystal are known we can calculate $\boldsymbol{F}_{A}$ precisely. As amplitudes of both $\boldsymbol{F}_{P}$ and $\boldsymbol{F}_{PA}$ are known experimentally the original phase of $\boldsymbol{F}_{P}$ can be determined from equation (11) simply by the use of the law of cosines:

$$\varphi = \pm \arccos\left(\frac{{F_{PA}}^2 - {F_P}^2 - {F_A}^2}{2F_P F_A}\right) \qquad (12)$$

A better understanding of the solution is given by a geometrical interpretation called Harker diagram (Figure 30).

Here, we draw circles of the sizes of structure factor amplitudes F_P and F_{PA} centred at the origin point and moved by the vector $-\boldsymbol{F}_{A}$ from the origin correspondingly. The intersections of these circles occur where the equation (11) is correct. As it is clearly seen, the solution provides two values of phase (as we got in algebraic form in the equation (12)) and at this point it is impossible to say which one is correct. To calculate initial electron density maps (equation (10)) one thus uses a weighted mean of these solutions, also called

centroid phase value. This method is called Single Isomorphous Replacement (SIR). Following calculation of the initial electron density map, this is then subject then to density modification procedures (see the end of this chapter) which allow retrieval of the correct phase angle.

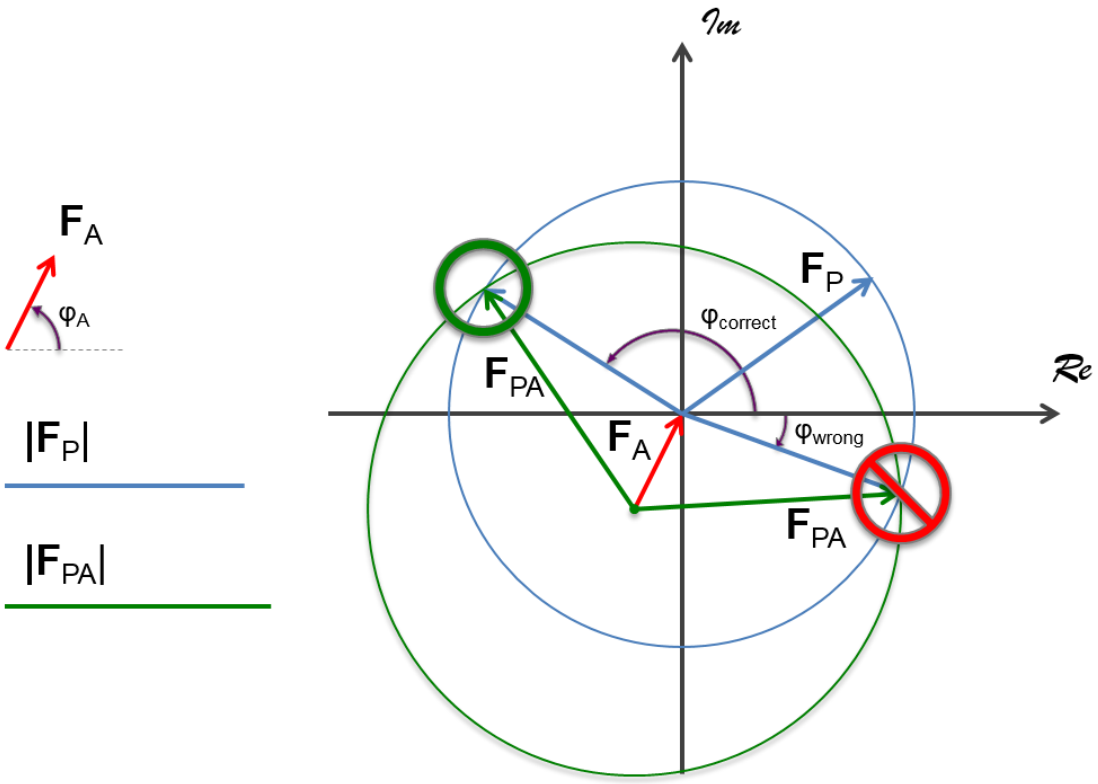


Figure 30. *A Harker diagram showing solution to the phase problem using the single isomorphous replacement (SIR) method. The two possible solutions, one correct and one wrong can be found at the intersection of the structure factor circles.*

The phase ambiguity resulting from a SIR experiment can be resolved by using several derivatives of protein crystals. In the MIR (Multiple Isomorphous Replacement) method multiple protein crystals are derivatised by different heavy atom molecules and diffraction data collected from these. Figure 31 illustrates how MIR resolves the SIR phase ambiguity. As can be seen, by using at least two derivatives we obtain a set of equations with the only solution. However, MIR experiments are often not successful: firstly, it is often not straightforward to derivatise a protein crystal even with one type of heavy atom;

secondly, adding heavy atoms to a crystal lattice often induces non-isomorphism between crystals of the heavy atom derivative and crystals of the native protein.

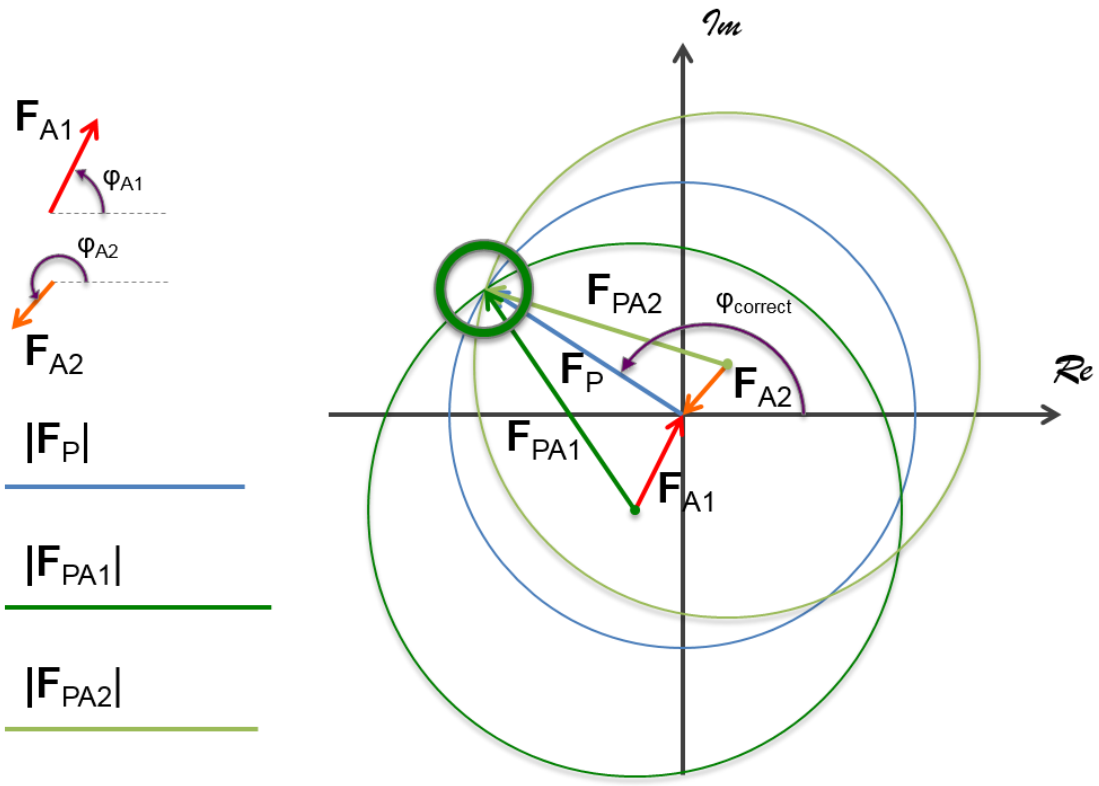


Figure 31. *A Harker diagram showing phase calculation using the MIR method. Here, two heavy atom derivatives with structure factors* $\boldsymbol{F}_{PA1}$ *and* $\boldsymbol{F}_{PA2}$ *are shown. A unique solution can be found at the intersection of all three structure factor circles.*

Increasingly, experimental (*de novo*) approaches relying on the introduction of heavy atoms into a crystal lattice take advantage of the phenomenon of anomalous scattering. This phenomenon arises from the fact that heavy atoms can absorb X-ray photons changing their energy state. When this phenomenon, known as anomalous scattering, is significant at the energy of an incident X-ray beam, it alters the atom's contribution to the diffraction pattern as shown in equation (13). The terms $f'(\lambda)$ and $f''(\lambda)$ are known as anomalous scattering coefficients related to inelastic scattering. They do not depend on the scattering direction, but only on the energy of the X-ray photons.

Conversely, basic term $f_0(\boldsymbol{s})$ is a Thomson scattering component depending only on scattering vector $\boldsymbol{s}$.

$$f = f_0(\boldsymbol{s}) + f'(\lambda) + i \cdot f''(\lambda) \qquad (13)$$

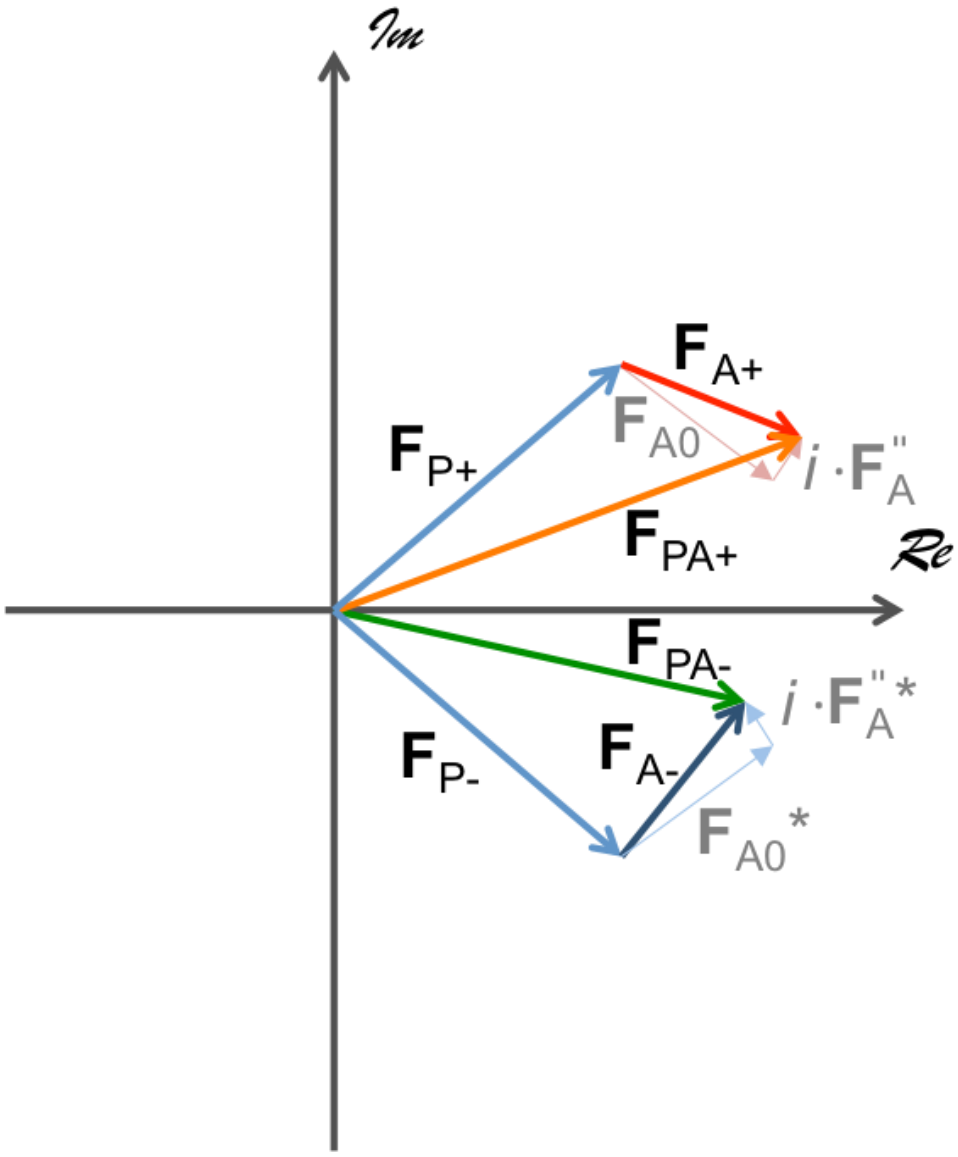


Figure 32. *A schematic diagram showing the breakdown of Friedel's law. The protein structure factors* $\boldsymbol{F}_{P+}$ *and* $\boldsymbol{F}_{P-}$ *are Friedel mates. The addition of* $\boldsymbol{F}_A$ *from the anomalous substructure makes the resulting structure factors* $\boldsymbol{F}_{PA+}$ *and* $\boldsymbol{F}_{PA-}$ *different.*

This modification to the structure factor breaks Friedel's law because of the imaginary term (see Figure 32) and for a protein heavy atom derivative the two Friedel mates $\mathbf{F}_{PA+}$ and $\mathbf{F}_{PA-}$ are no longer complex conjugates. As shown in Figure 33, this

phenomenon can be used in Single-wavelength Anomalous Dispersion (SAD) experiments to retrieve the phase value. Knowing the amplitudes of $\mathbf{F}_{PA+}$ and $\mathbf{F}_{PA-}$ two circles of these radii can be drawn, shifted from the origin by $-\mathbf{F}_{A+}$ and $-\mathbf{F}_{A-}$, correspondingly. The solution can be found at the intersection of two circles. Again, and as for SIR, the problem of phase ambiguity occurs (only one of the two intersections corresponds to the correct value of the phase). At this point, and again as for SIR, this problem can often be solved using density modification.

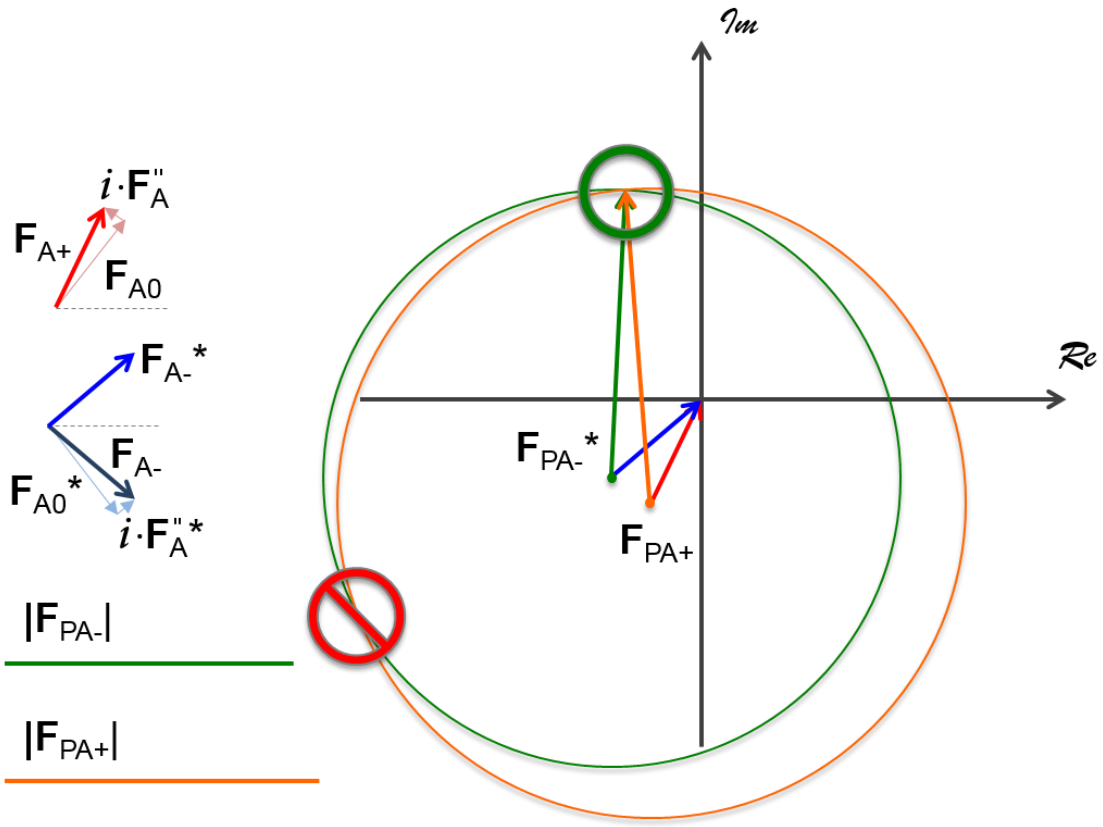


Figure 33. *A Harker diagram showing the solution for single wavelength anomalous dispersion method. The two solutions, correct and wrong ones can be found at the intersection of the structure factor circles* $\boldsymbol{F}_{PA+}$ *and* $\boldsymbol{F}_{PA-}$.

As should be clear from Figure 33, SAD phasing does not require the measurement of diffraction data from a native crystal of a protein; only the measurement of diffraction data from a single heavy atom derivative is required and, importantly the phase angles derived are for the structure factors of the derivative crystal. However, if a diffraction data

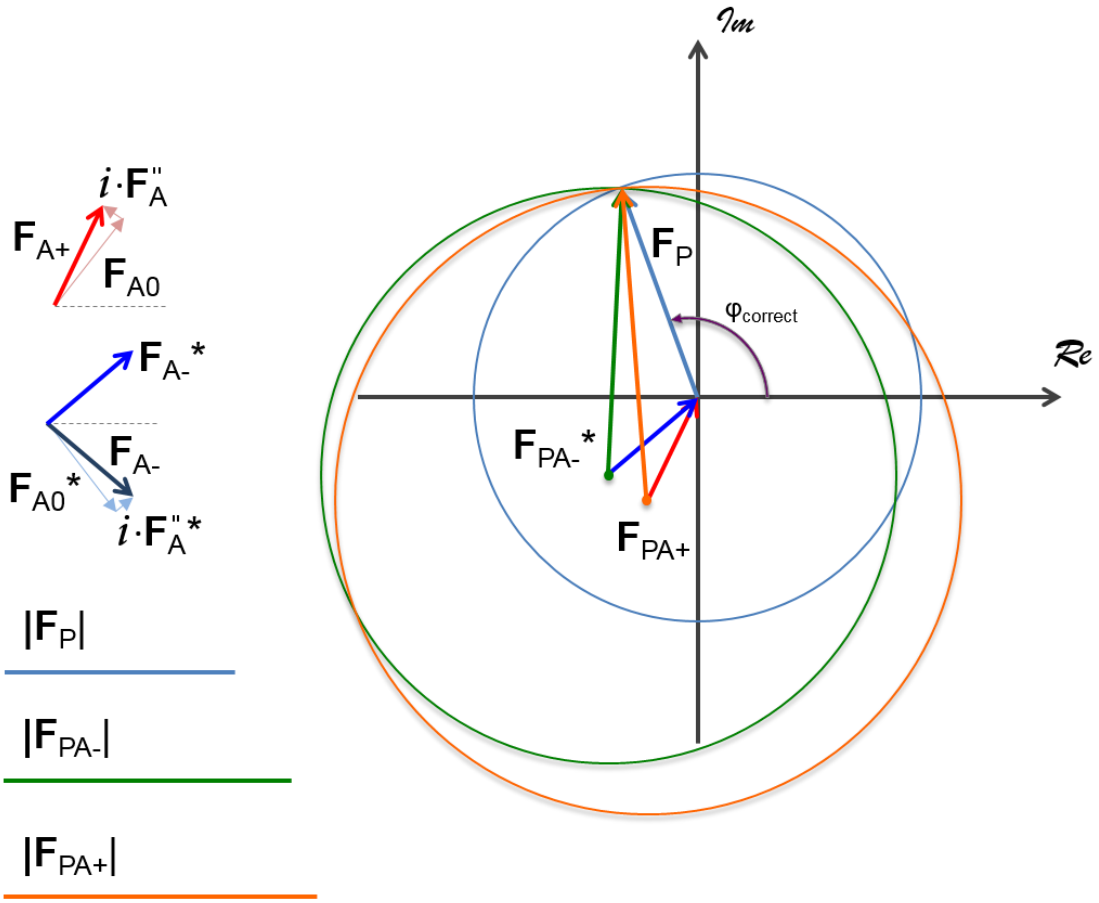


Figure 34. *A Harker diagram showing the solution for the single isomorphous replacement with anomalous scattering (SIRAS) method. The only solution can be found at the intersection of the structure factor circles* $\boldsymbol{F}_P$, $\boldsymbol{F}_{PA+}$ *and* $\boldsymbol{F}_{PA-}$.

from a native crystal are available phases can be derived using Single Isomorphous Replacement with Anomalous Scattering (SIRAS). As can be seen from Figure 34, addition of a F_P circle to a Harker diagram allows for a unique determination of the phase angle.

Proper use of the anomalous scattering properties of a heavy atom derivative can also produce a unique value for structure factor phase angles without the need to collect diffraction data from a native crystal. In Figure 35 and Figure 36 experimentally obtained normalised X-ray absorption spectra of iodide and bromide are shown in the energy ranges 4500 – 4750 eV and 13460 – 13550 eV, respectively. These show the iodide L_{III}-edge and bromide K-edge, respectively. The corresponding anomalous scattering factors $f'(\lambda)$ and $f''(\lambda)$ (equation (13)) can be calculated from equations (14) and (15). Figure 37 shows the

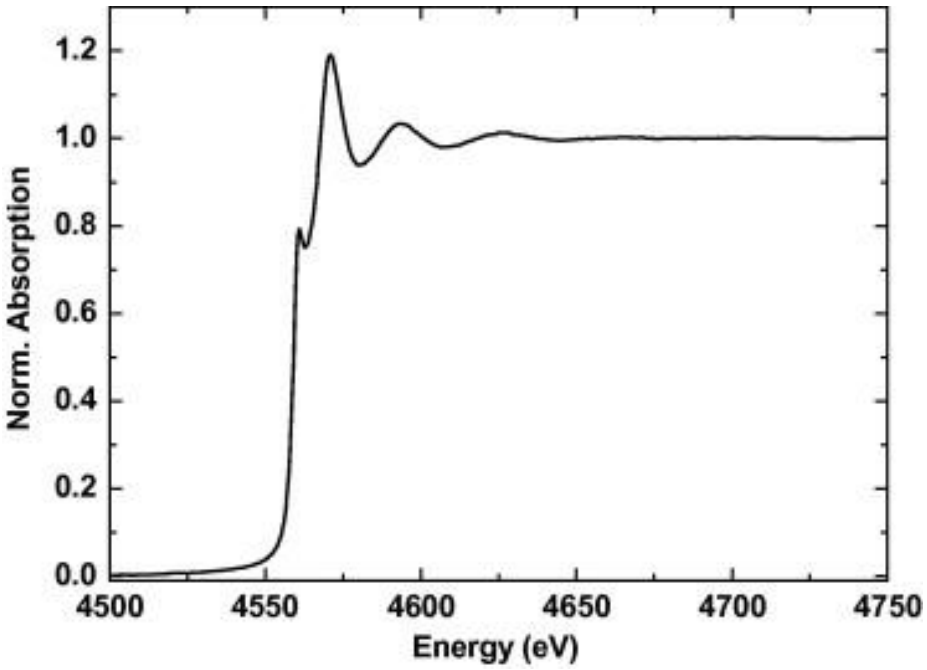


Figure 35. *L_{III}-edge of iodide absorption spectra. The normalised X-ray absorption coefficient is denoted on vertical axis. Adapted from (Pham et al., 2010).*

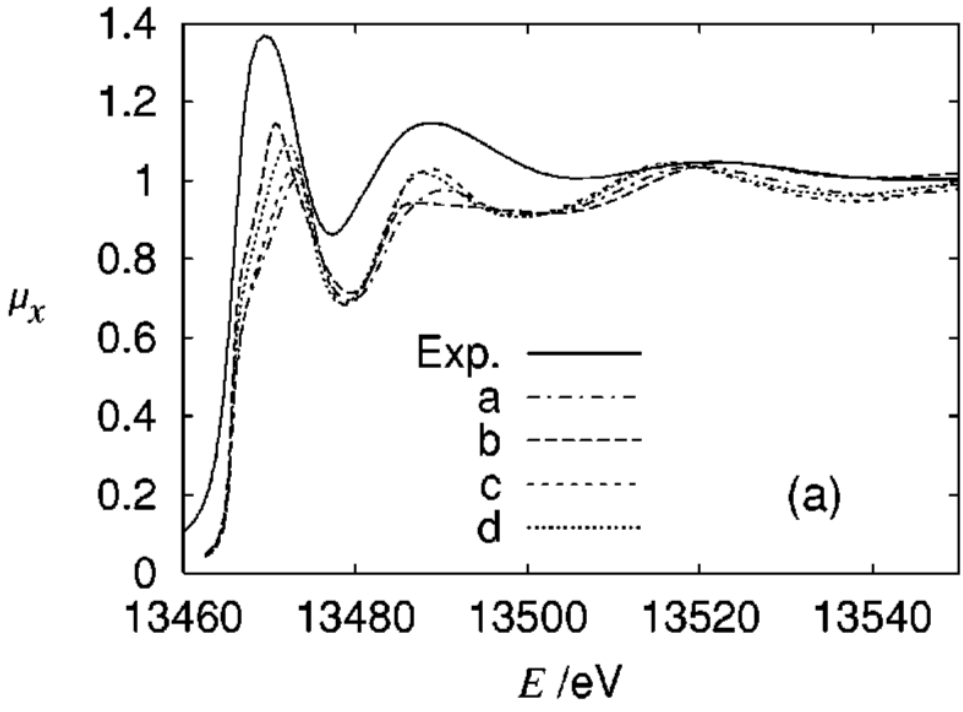


Figure 36. *K-edge of bromide absorption spectra. The normalised X-ray absorption coefficient is denoted on vertical axis. Adapted from (Merkling et al., 2003).*

theoretical variation of the anomalous scattering factors $f'(\lambda)$ and $f''(\lambda)$ near the L_{III} absorption edge of iodine and the K absorption edge of bromine (Cromer & Liberman, 1970).

$$f''_{\lambda} = \frac{mc}{4e^2\hbar} E\mu_{\lambda} \qquad (14)$$

$$f'_{\lambda} = \frac{2}{\pi}\int_0^{\infty} E_t \cdot \frac{f''_{E_t}}{E^2 - {E_t}^2}\, dE_t \qquad (15)$$

where E is the X-ray energy, μ_{λ} is the atomic X-ray absorption coefficient.

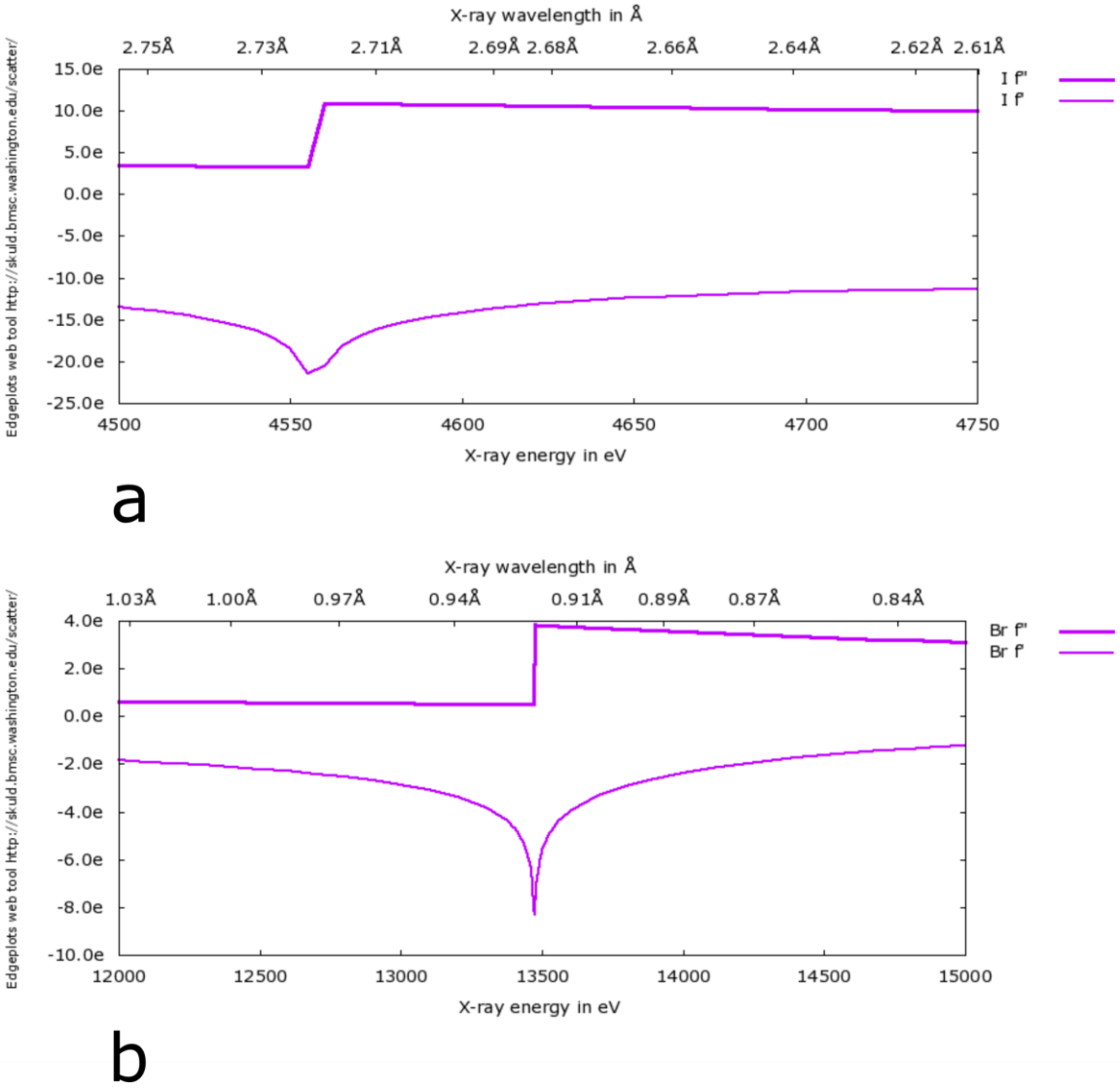


Figure 37. *The theoretical variation of the anomalous scattering factors ($f'(\lambda)$ and $f''(\lambda)$) near the (**a**) the L_{III} absorption edge of iodine and (**b**) the K absorption edge of bromine. Calculated at http://skuld.bmsc.washington.edu/.*

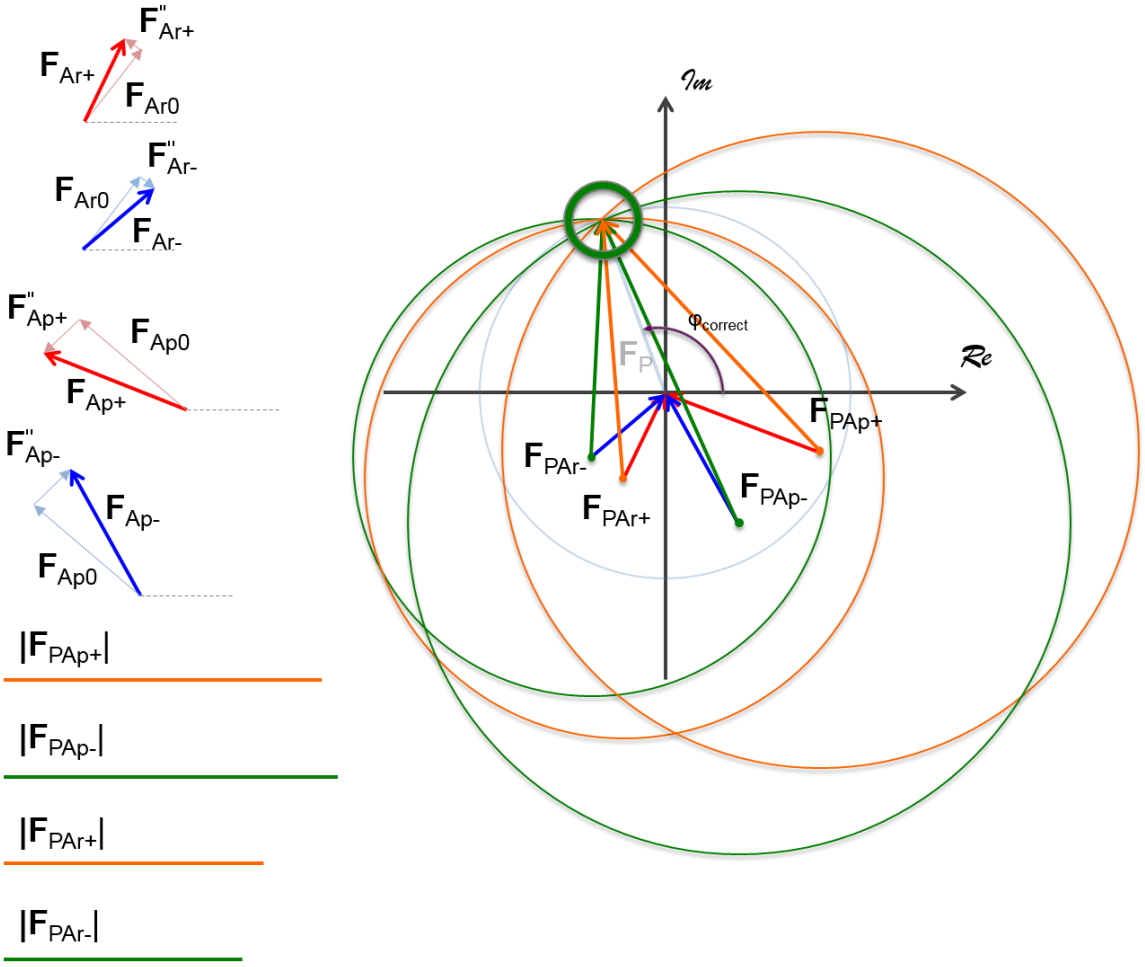


Figure 38. *A Harker diagram for the MAD method. The only solution of the system can be found at the intersection of at least four structure factor circles* $\boldsymbol{F}_{PAp-}$, $\boldsymbol{F}_{PAp+}$, $\boldsymbol{F}_{PAr-}$, $\boldsymbol{F}_{PAr+}$ *obtained at different X-ray energies ("peak" and "remote"). The native structure factor circle is drawn for the convenience in transparent blue.*

As can be seen, there is a rapid variation of $f'(\lambda)$ and $f''(\lambda)$ around an absorption edge which will result in different structure factors at different wavelengths. This particular phenomenon is employed in the Multi-wavelength Anomalous Dispersion (MAD) technique. Typically, when employing MAD diffraction data are collected at least at three points around an absorption edge: "peak" (corresponding to the peak of absorption spectra), "inflection point" (as corresponds to the point of inflection of the spectra) and "high or low remote" (corresponding to the low or high energy sides of absorption edge). Then, anomalous structure factors are retrieved for each data set. Figure 38 shows geometrically how the correct phase value can be found using the structure factors from

"peak" and "low remote" data sets. Indeed, there is no phase ambiguity problem here as it is resolved by using more equations. So, MAD is a very powerful technique. However, it requires an X-ray facility with tuneable beam energy.

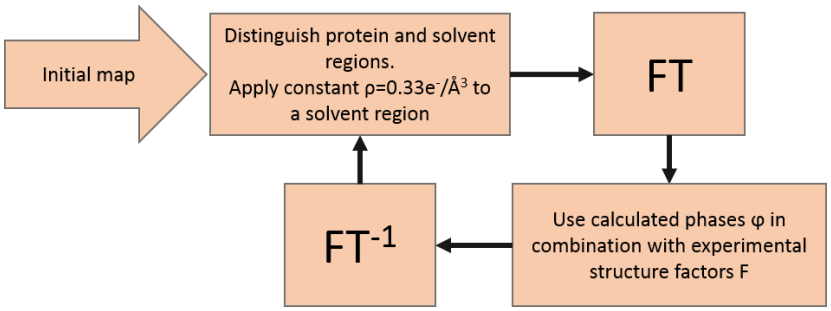


Figure 39. *Scheme of a solvent flattening algorithm. Cycling is done until no possible further map improvement.*

Once experimental phases have been determined using one of the methods described above, they usually need to be improved before an interpretable electron density map can be produced. Phase improvement is usually carried out using density modification techniques which include solvent flattening (Wang, 1985), solvent flipping (Abrahams & Leslie, 1996), density averaging and histogram matching (Lunin, 1988). In solvent flattening first the protein mask area is defined in the experimentally-phased electron density map and distinguished from the solvent environment. Typically, solvent region of a crystal has an average electron density of $0.32e^{-}/Å^{3}$ ((Terwilliger, 1999), excluding extreme salty solutions). The density in solvent region is set to this average value and new structure factors are calculated via Fourier transform of the modified electron density map. Improved phases are then combined with original experimental amplitudes and used to compute a new map where next cycle of solvent flattening is applied (Figure 39). More complicated algorithms are introduced in software programs for density modification such as sphere of influence method in SHELX (Sheldrick, 2002). In most cases,

density modification is used to improve the initial phases obtained by the other phasing techniques rather than being itself the main method of structure determination.

2.2.1.Halide SAD

It has been reported that so-called 'halide-SAD' is a promising technique for solving the phase problem in the crystal structures of soluble proteins (Dauter *et al.*, 2000; Abendroth *et al.*, 2011). Here, the simple protocols of soaking and fast cryo-soaking of protein crystals in a halide-containing solution were used for successful structure solution of the structures via SAD experiments using iodide (I-SAD) and bromide (Br-SAD) derivatives of native crystals. In the cryo-soaking protocol the crystals are soaked in solution for several minutes and then are frozen immediately in liquid nitrogen. More recently, the structure of a MP (Gu *et al.*, 2016) has been solved via I-SAD based on the introduction of iodide ions by co-crystallisation.

Several studies (Lund *et al.*, 2008; Heyda *et al.*, 2009) suggest a high affinity of halide ions to positively charged protein residues. MPs typically contain positively charged residues located at the membrane hydrophobic/hydrophilic interface. This can favour the interaction between MPs and halide ions. Indeed, the concept, widely known as the "positive-inside rule" (vonHeijne, 1989; von Heijne, 1992; Heijne, 1994) suggests that positively charged residues of a MP are necessarily present at the inner interface of the membrane. This was defined as the major determinant of topology of integral MPs and is essential for proper folding and insertion of the entire MP polypeptide into the membrane. However, positive charges close to the lipidic membrane surface seeks to be compensated due to the polarizing properties of the membrane and halides ions are thus expected to strongly interacting with MPs.

Due to the ion size and increased polarizability, bulkier halides also have an increased tendency for non-polar interactions with hydrophobic residues and cavities close to a protein surface. The prevalence of some aromatic residues at or close to the hydrophobic/hydrophilic interface in the structures of MPs (Yau *et al.*, 1998) again supports that this region would also be a good target for the binding of ions such as iodide. This, coupled with the observations above, suggests that halide-SAD might prove a routine technique for solving the phase problem in the determination of the crystals structures of MPs. Part of this thesis work was dedicated to testing this hypothesis. A manuscript describing the results of this work was recently published and is, along with the relevant Supplementary Materials, reproduced here.

2.3. Fast iodide-SAD phasing for high-throughput membrane protein structure determination

Igor Melnikov[1], Vitaly Polovinkin[234], Kirill Kovalev[34], Ivan Gushchin[34], Mikhail Shevtsov[4], Vitaly Shevchenko[345], Alexey Mishin[4], Alexey Alekseev[34], Francisco Rodriguez-Valera[6], Valentin Borshchevskiy[4], Vadim Cherezov[47], Gordon A. Leonard[1], Valentin Gordeliy[234] and Alexander Popov[1]

[1]European Synchrotron Radiation Facility, 38043 Grenoble, France.

[2]Institut de Biologie Structurale Jean-Pierre Ebel, Université Grenoble Alpes–Commissariat à l'Energie Atomique et aux Energies Alternatives–CNRS, F-38000 Grenoble, France.

[3]Institute of Complex Systems (ICS), ICS-6, Structural Biochemistry, Research Centre Jülich, 52425 Jülich, Germany.

[4]Moscow Institute of Physics and Technology, 141700 Dolgoprudny, Moscow Region, Russia.

[5]Institute of Crystallography, University of Aachen (RWTH), Aachen, Germany.

[6]Department of Vegetal Production and Microbiology, Universidad Miguel Hernández de Elche, San Juan de Alicante, Valencia, Spain.

[7]Departments of Chemistry, Biological Sciences, and Physics & Astronomy, Bridge Institute, University of Southern California, Los Angeles, CA 90089, USA.

Science Advances ***3*** *(2017): e1602952*

BIOCHEMISTRY

Fast iodide-SAD phasing for high-throughput membrane protein structure determination

Igor Melnikov,[1]* Vitaly Polovinkin,[2,3,4]* Kirill Kovalev,[3,4] Ivan Gushchin,[3,4] Mikhail Shevtsov,[4] Vitaly Shevchenko,[3,4,5] Alexey Mishin,[4] Alexey Alekseev,[3,4] Francisco Rodriguez-Valera,[6] Valentin Borshchevskiy,[4] Vadim Cherezov,[4,7] Gordon A. Leonard,[1] Valentin Gordeliy,[2,3,4†] Alexander Popov[1†]

We describe a fast, easy, and potentially universal method for the de novo solution of the crystal structures of membrane proteins via iodide–single-wavelength anomalous diffraction (I-SAD). The potential universality of the method is based on a common feature of membrane proteins—the availability at the hydrophobic-hydrophilic interface of positively charged amino acid residues with which iodide strongly interacts. We demonstrate the solution using I-SAD of four crystal structures representing different classes of membrane proteins, including a human G protein–coupled receptor (GPCR), and we show that I-SAD can be applied using data collection strategies based on either standard or serial x-ray crystallography techniques.

INTRODUCTION

Membrane proteins (MPs) are the key functional components of cell membranes. They carry out the main functions of cells including ion and solute transport and energy and signal transduction. MPs represent roughly one-third of the proteins encoded in any genome, including the human one (*1*). However, although they are also extremely important drug targets [up to 60% of all existing drug targets are MPs (*2*)], they still remain poorly structurally characterized compared to soluble proteins, mainly because their production and crystallization are two major bottlenecks of structural biology. Structural investigations by x-ray crystallography thus remain a significant challenge. This challenge becomes even more demanding if the crystal structure cannot be solved by molecular replacement (MR) methods. In these cases, the most popular approaches are the use of heavy atom derivatization (*3*), the substitution of methionine with selenomethionine (*4*), and native sulfur single-wavelength anomalous diffraction (SAD) phasing (*5*, *6*). However, these techniques can be expensive and hazardous and are often not very efficient (*6*–*8*). A reliable, fast, and easy method of preparing these derivatives would thus be a huge step forward in the production of MP crystal structures.

Several reports have shown halide-SAD to be a promising method for phasing the crystal structures of water-soluble proteins (*9*, *10*), and more recently, iodide-SAD (I-SAD) has been used to solve the structure of an MP (*7*). The distribution of positively charged amino acids at, or close to, the membrane hydrophobic-hydrophilic interface is considered to be a major determinant of the transmembrane topology of integral MPs (*11*) and is one of their most universal features. In this context, the "positive-inside" rule suggests that regions of polytopic (multispanning) MPs facing the cytoplasm are enriched with arginine, lysine, and, to a lesser extent, histidine amino acid residues (*11*–*13*). In addition, there are also observations of an enrichment of tryptophan and tyrosine residues at the membrane surface interfaces (*14*). From an electrostatic point of view, the existence of uncompensated charge in close proximity to a medium with lower dielectric constant (for example, the membrane hydrophobic-hydrophilic interface) is unfavorable, and to compensate for this, the positively charged residues placed at the interface are likely to interact particularly strongly with bulk halide ions. Thus, we hypothesized that halide-SAD might be a universal phasing method in x-ray MP crystallography.

To test this hypothesis and to develop protocols for halide-SAD phasing, we selected four target representatives for different large classes of MPs varying in the sizes of their membrane and soluble components: the recently discovered and characterized light-driven sodium pump from the marine bacterium *Krokinobacter eikastus* (*KR2*) (*15*), the light-driven proton pump from marine actinobacterial clade rhodopsins (MACRs) (*16*), a fragment of histidine protein kinase NarQ from *Escherichia coli* (*17*), and a human adenosine A_{2A} G protein–coupled receptor (GPCR) (*18*). We attempted to solve their crystal structures by preparing iodide or bromide derivatives via the cryo-soaking of native crystals in solutions of either NaI (I-SAD) or NaBr (Br-SAD). The results we report here suggest that I-SAD is an efficient, fast, nontoxic, and potentially universal technique for the de novo (that is, not MR-based) solution of MP crystal structures that will help to improve throughput in MP structure determination.

RESULTS

Iodide–single-wavelength anomalous diffraction

Diffraction data sets were collected from several different crystals of each of the target proteins after first soaking them in precipitant solution supplemented with 0.5 M NaI (Table 1). In all cases, diffraction data were collected using x-rays of λ = 1.85 Å, where f'' for iodine/iodide = 9.2 e^- (6.7 keV; all values of f'' quoted are taken from http://skuld.bmsc.washington.edu/scatter/). For each target, several individual data sets were merged to obtain a single high-redundancy data set for use in structure solution (Table 1). In all four cases, I-SAD

[1]European Synchrotron Radiation Facility, 38043 Grenoble, France. [2]Institut de Biologie Structurale Jean-Pierre Ebel, Université Grenoble Alpes–Commissariat à l'Energie Atomique et aux Energies Alternatives–CNRS, F-38000 Grenoble, France. [3]Institute of Complex Systems (ICS), ICS-6, Structural Biochemistry, Research Centre Jülich, 52425 Jülich, Germany. [4]Moscow Institute of Physics and Technology, 141700 Dolgoprudny, Moscow Region, Russia. [5]Institute of Crystallography, University of Aachen (RWTH), Aachen, Germany. [6]Department of Vegetal Production and Microbiology, Universidad Miguel Hernández de Elche, San Juan de Alicante, Valencia, Spain. [7]Departments of Chemistry, Biological Sciences, and Physics & Astronomy, Bridge Institute, University of Southern California, Los Angeles, CA 90089, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: valentin.gordeliy@ibs.fr (V.G.); apopov@esrf.fr (A.P.)

Table 1. I-SAD/I-SIRAS data collection statistics. Entries in parentheses represent values for the highest-resolution bin. All statistics and data quality indicators are calculated, treating Friedel's pairs as separate reflections. Linear merging R value $R_{merge} = \frac{\Sigma_h \Sigma_{i=1}^{N} |I_{(h)i} - \langle I_{(h)} \rangle|}{\Sigma_h \Sigma_{i=1}^{N} I_{(h)i}}$; redundancy-independent merging R value $R_{meas} = \frac{\Sigma_h \sqrt{\frac{N}{N-1}} \Sigma_{i=1}^{N} |I_{(h)i} - \langle I_{(h)} \rangle|}{\Sigma_h \Sigma_{i=1}^{N} I_{(h)i}}$; $CC_{1/2}$—correlation coefficient between intensities from random half–data sets; SigAno $= \left\langle \frac{|F_+ - F_-|}{\sigma(F_+ - F_-)} \right\rangle$.

Protein/ space group	Crystal/ data set	Unit cell dimensions (Å, °)						Resolution range (Å)	Multiplicity	Completeness (%)	R_{merge} (%)	R_{meas} (%)	$\langle I/\sigma(I) \rangle$	$CC_{1/2}$ (%)	SigAno	CC_{anom} (%)	$SigAno_{inner}$	CC_{anom_inner} (%)
		a	*b*	*c*	α	β	γ											
NarQ *F*222	Native	57.58	73.87	235.56	90	90	90	60–2.0 (2.1–2.0)	3.61 (3.50)	99.8 (99.7)	8.8 (79.7)	10.4 (94.2)	10.58 (1.58)	99.9 (62.3)	0.76 (0.73)	−3 (−4)	0.68	−4
	1	57.27	73.64	236.15	90	90	90	60–2.9 (3.1–2.9)	6.69 (5.99)	99.7 (99.1)	15.8 (77.1)	17.2 (84.6)	10.23 (2.45)	99.8 (87.5)	1.24 (0.88)	47 (26)	4.12	92
	2	57.40	73.79	237.02	90	90	90	60–3.3 (3.5–3.3)	6.73 (6.78)	99.7 (99.8)	20.8 (75.9)	22.6 (82.2)	7.33 (2.36)	99.6 (86.1)	1.15 (0.76)	41 (8)	3.17	90
	3	57.22	73.59	236.42	90	90	90	60–3.1 (3.3–3.1)	6.74 (6.53)	99.7 (99.7)	14.2 (62.0)	15.4 (67.5)	9.87 (3.20)	99.8 (88.8)	1.43 (1.08)	57 (32)	4.03	93
	4	56.92	73.20	236.25	90	90	90	60–2.9 (3.1–2.9)	6.69 (6.69)	99.7 (99.8)	11.3 (50.3)	12.3 (54.6)	12.37 (3.30)	99.8 (93.3)	1.57 (0.78)	64 (5)	5.82	96
	5	57.21	73.52	236.54	90	90	90	60–2.7 (2.9–2.7)	6.69 (6.59)	99.8 (99.8)	11.4 (73.3)	12.4 (79.6)	11.93 (2.47)	99.9 (85.3)	1.51 (0.82)	61 (16)	6.03	97
	6	57.12	73.34	236.74	90	90	90	60–2.9 (3.1–2.9)	6.72 (6.52)	99.4 (99.5)	11.3 (67.4)	12.3 (73.2)	12.20 (2.67)	99.9 (87.1)	1.53 (0.80)	60 (1)	5.85	96
	7	56.99	73.12	236.07	90	90	90	60–2.9 (3.1–2.9)	6.73 (6.68)	98.9 (98.4)	11.1 (71.4)	12.0 (77.4)	12.56 (2.51)	99.7 (82.5)	1.57 (0.80)	62 (8)	6.00	97
	8	57.24	73.66	236.43	90	90	90	60–2.7 (2.9–2.7)	6.73 (6.68)	99.8 (99.8)	12.6 (71.2)	13.6 (77.4)	11.03 (2.67)	99.9 (86.8)	1.41 (0.94)	54 (16)	5.15	96
	Merged	57.17	73.48	236.45	90	90	90	60–2.7 (2.9–2.7)	42.95 (13.26)	99.9 (99.8)	17.5 (76.0)	17.7 (79.1)	20.68 (3.12)	99.8 (91.7)	2.26 (0.85)	78 (17)	10.14	98
KR2 *I*222	Native	40.62	83.66	234.46	90	90	90	60–2.5 (2.65–2.5)	6.58 (6.45)	99.7 (99.6)	9.6 (85.9)	10.5 (93.7)	13.73 (2.10)	99.8 (81.2)	0.84 (0.81)	15 (19)	1.42	44
	1	40.50	83.66	234.13	90	90	90	60–2.8 (3.0–2.8)	6.85 (6.72)	99.7 (99.8)	13.5 (83.0)	14.6 (89.9)	11.87 (2.44)	99.8 (80.7)	1.22 (0.80)	41 (10)	6.76	97
	2	40.52	83.63	233.80	90	90	90	60–2.8 (3.0–2.8)	6.82 (6.71)	99.8 (99.7)	13.7 (77.0)	14.8 (83.5)	11.92 (2.60)	99.6 (87.2)	1.24 (0.79)	44 (7)	7.26	98
	3	40.51	83.75	234.10	90	90	90	60–2.5 (2.65–2.5)	6.73 (6.60)	99.9 (100)	11.3 (75.4)	12.3 (82.0)	11.95 (2.35)	99.9 (79.7)	1.22 (0.76)	42 (0)	7.64	98
	4	40.72	83.99	234.21	90	90	90	60–2.65 (2.8–2.65)	6.71 (6.40)	99.8 (99.7)	10.6 (64.0)	11.5 (69.7)	12.70 (2.66)	99.6 (89.0)	1.25 (0.77)	46 (5)	6.96	98
	Merged	40.56	83.76	234.06	90	90	90	60–2.5 (2.65–2.5)	22.06 (6.60)	100 (100)	13.7 (75.1)	14.0 (81.6)	19.43 (2.30)	99.9 (80.0)	1.64 (0.75)	60 (1)	13.23	99
MACR *P*1	Native	40.76	56.85	57.65	63.67	90.08	0.1	60–2.0 (2.2–2.0)	1.05 (1.07)	76.2 (73.1)	7.3 (26.7)	10.3 (37.9)	6.43 (1.81)	99.5 (86.2)	n/a	n/a	n/a	n/a
	1	40.77	56.89	57.28	63.87	86.80	2.2	60–2.2 (2.4–2.2)	1.86 (1.81)	70.1 (23.0)	5.1 (27.2)	7.2 (38.5)	10.40 (2.61)	99.5 (87.0)	1.16 (0.88)	43 (8)	3.61	91
	2	40.75	56.87	57.32	63.67	84.80	2.2	60–2.4 (2.6–2.4)	1.85 (1.79)	72.5 (20.0)	6.4 (44.9)	9.0 (63.4)	7.86 (1.55)	99.5 (72.4)	1.07 (0.87)	36 (22)	3.14	89
	3	40.79	56.77	57.29	63.67	85.80	2.2	60–2.0 (2.2–2.0)	1.88 (1.84)	81.9 (72.3)	5.1 (38.3)	7.2 (54.2)	8.64 (1.86)	99.7 (70.4)	1.07 (0.76)	35 (−5)	3.49	90
	Merged	40.77	56.84	57.30	63.77	85.80	2.2	60–2.0 (2.2–2.0)	3.68 (1.84)	89.4 (71.8)	7.9 (38.7)	9.0 (54.8)	11.03 (1.81)	99.7 (69.5)	1.22 (0.75)	46 (−5)	4.85	95

continued on next page

Protein/ space group	Crystal/ data set	Unit cell dimensions (Å, °)						Resolution range (Å)	Multiplicity	Completeness (%)	R_{merge} (%)	R_{meas} (%)	$\langle I/\sigma(I)\rangle$	$CC_{1/2}$ (%)	SigAno	CC_{anom} (%)	$SigAno_{inner}$	CC_{anom_inner} (%)
		a	*b*	*c*	α	β	γ											
A_{2A}AR-BRIL-ΔC $C222_1$	Native	39.62	179.54	139.71	90	90	90	60–2.4 (2.6–2.4)	2.94 (2.98)	99.6 (99.8)	9.4 (41.1)	11.5 (50.2)	9.60 (2.74)	99.5 (80.4)	0.77 (0.77)	−4 (−1)	0.74	−14
	1	39.63	179.66	139.57	90	90	90	60–3.4 (3.6–3.4)	6.93 (6.96)	100 (100)	24.8 (73.6)	26.8 (79.5)	7.33 (2.56)	99.1 (81.2)	0.96 (0.79)	17 (0)	1.94	71
	2	39.65	179.79	139.81	90	90	90	60–2.8 (3.0–2.8)	6.87 (6.87)	99.7 (99.0)	18.5 (78.1)	20.0 (84.5)	8.83 (2.29)	99.4 (76.0)	0.98 (0.76)	18 (−5)	2.57	82
	3	39.40	179.87	139.52	90	90	90	60–3.6 (3.8–3.6)	6.67 (6.54)	99.9 (99.8)	23.5 (72.1)	25.5 (78.5)	8.45 (3.21)	99.2 (82.8)	1.13 (1.08)	46 (41)	1.96	72
	4	39.57	179.76	139.99	90	90	90	60–3.8 (4.0–3.8)	6.83 (6.79)	99.9 (99.9)	30.7 (84.7)	33.3 (91.9)	6.67 (2.97)	98.6 (72.0)	0.96 (0.91)	19 (8)	1.59	63
	5	39.60	179.10	139.42	90	90	90	60–3.0 (3.2–3.0)	3.02 (2.98)	98.9 (99.0)	14.5 (65.7)	17.6 (80.1)	7.06 (1.67)	98.7 (56.6)	0.93 (0.75)	16 (−3)	2.06	74
	6	39.50	179.46	139.40	90	90	90	60–3.4 (3.6–3.4)	6.85 (6.52)	99.9 (100)	19.9 (69.1)	21.5 (75.2)	9.61 (3.20)	99.4 (80.6)	1.07 (0.89)	27 (5)	2.46	81
	7	39.63	179.65	139.96	90	90	90	60–3.2 (3.4–3.2)	6.27 (5.83)	99.9 (99.7)	21.1 (68.3)	23.1 (75.1)	8.86 (2.61)	98.9 (84.4)	1.02 (0.92)	24 (12)	2.06	76
	Merged	39.57	179.61	139.67	90	90	90	60–2.8 (3.0–2.8)	27.10 (6.88)	99.8 (99.1)	24.5 (77.6)	24.9 (83.9)	14.55 (2.25)	99.8 (76.1)	1.24 (0.73)	35 (−3)	4.83	94

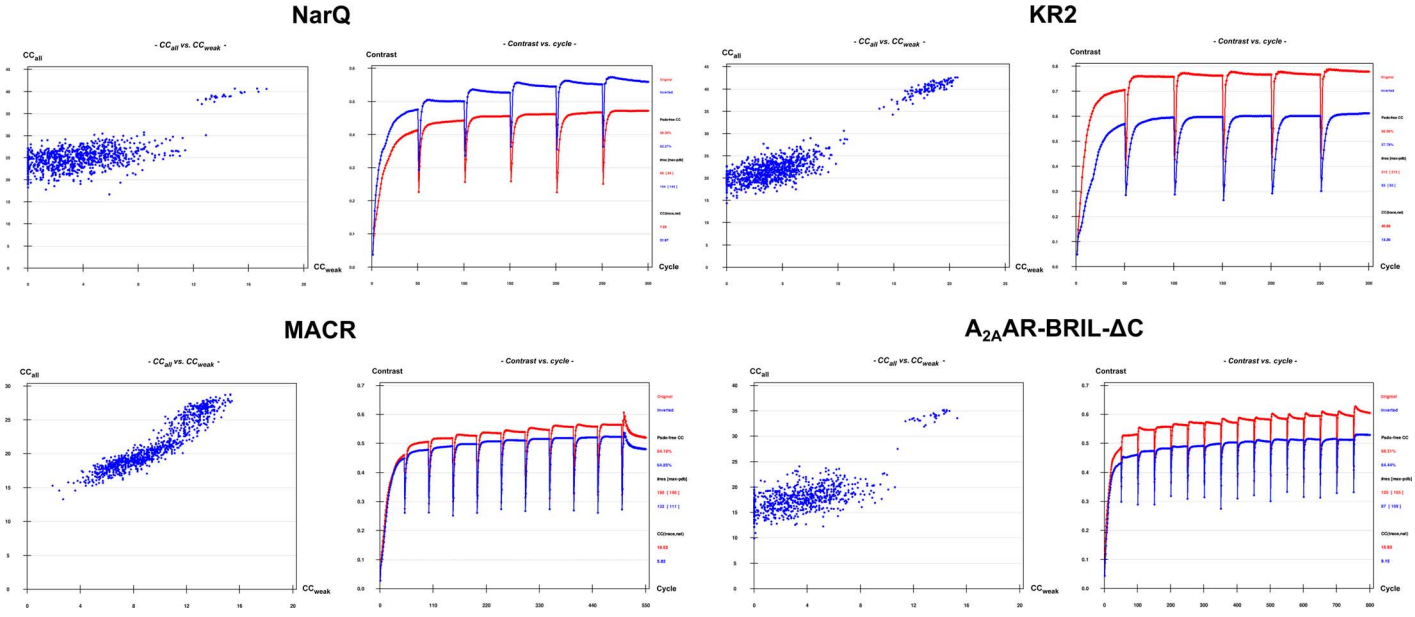


Fig. 1. I-SAD phasing of MP crystal structures. Heavy atom substructure determination (SHELXD, left) and phase calculation and extension (SHELXE, right) for each of the four targets investigated. In all four cases, plots of CC_{all} versus CC_{weak} (1000 trials) from SHELXD show a bimodal distribution indicative of a correct iodide ion substructure solution, whereas the difference in contrast (SHELXE) between the two possible hands (red, original; blue, inverted) indicates successful structure solution.

phasing of the crystal structure was straightforward (Fig. 1 and Table 2), providing experimental phases that allowed the automated building of initial models, which could be further refined. By including native data sets, we could also test the efficiency of I-SIRAS (single isomorphous replacement with anomalous scattering) phasing of all four crystal structures. This approach (Fig. 2) provided even higher contrast solutions than I-SAD protocols, suggesting that introduction of iodide ions into the crystals had not resulted in significant non-isomorphism. In addition to the above methods, in experiments based on the serial x-ray crystallography MeshAndCollect pipeline (*19*), we also collected a succession of partial data sets from a series of iodide-soaked crystals of KR2, each with a maximum dimension of ~10 to 15 μm (Table 3 and Fig. 3). Here, structure solution was also successful using either I-SAD or I-SIRAS (Fig. 3), suggesting that large crystals are not necessary for the collection

Table 2. I-SAD structure solution and refinement statistics.

	Protein							
	NarQ		KR2		MACR		A_{2A}AR-BRIL-ΔC	
Structure solution and initial model building								
Data	Eight anomalous + native	Data set #8 + native	Four anomalous + native		Three anomalous + native		Seven anomalous + native	
Method	I-SAD	I-SIRAS	I-SAD	I-SIRAS	I-SAD	I-SIRAS	I-SAD	I-SIRAS
Initial resolution range (Å)	60–2.7	60–2.7	60–2.5	60–2.5	60–2.0	60–2.0	60–2.8	60–2.8
Number of residues in the protein	237		288		220		447	
Substructure	14·I	10·I	16·I	17·I	26·I[†] 14·I	16·I	8·I	10·I
Resolution cutoff for substructure search (Å)	3.0	2.8	2.8	2.8	2.5	2.9	3.7	3.9
Resolution extension by native data (Å)	1.95	1.95	2.5	2.5	2.0	2.0	2.4	2.4
CC_{all}/CC_{weak}	40/15.4	31.8/18.1	42.6/20.7	37.7/25.1	28.8/15.3	29.8/19.6	35.0/14.8	36.7/21.4
Solvent content used in SHELXE (%)	47	47	58	59	50	50	49	49
CC_{map}, highest-resolution shell (%)	74	77	67	84	87	81	73	74
Number of Ala residues traced by SHELXE	178	187	212	208	190	272	155	250
R_{work}/R_{free} after initial model building (%)	23.0/28.5*	24.8/32.9*	22.4/30.1*	22.6/29.3*	24.2/29.2[†‡] 29.4/32.7[‡]	29.7/46.2* 29.9/32.7[‡]	30.7/32.9[‡]	29.5/45.8*
Number of residues built in initial model building	215*	216*	235*	265*	343[†‡] 300[‡]	69* 319[‡]	277[‡]	168*
Structure refinement								
Data set (see Table 1)	8		3		3		2	
R_{work}/R_{free} final (%)	19.0/26.2		19.0/22.6		17.7/21.9		23.4/29.4	
Number in the ASU (asymmetric unit)								
Protein residues	223		268		431		388	
Water molecules	60		49		78		64	
Iodide ions	12		20		22		6	
Averaged *B*-factors (Å^2)								
Protein residues	43		40		38		38	
Water molecules	34		44		47		27	
Iodide ions	70		66		71		50	
Ramachandran plot								
Preferred	216 (99.1%)		259 (97.4%)		419 (98.4%)		373 (97.9%)	
Allowed	2 (0.9%)		6 (2.2%)		7 (1.6%)		7 (1.8%)	
Outliers	0		1 (0.4%)		0		1 (0.3%)	

*Initial model built by ARP-wARP. †Phasing carried out using phenix.autosol (all others used SHELXC/D/E). ‡Initial model built by phenix.autobuild.

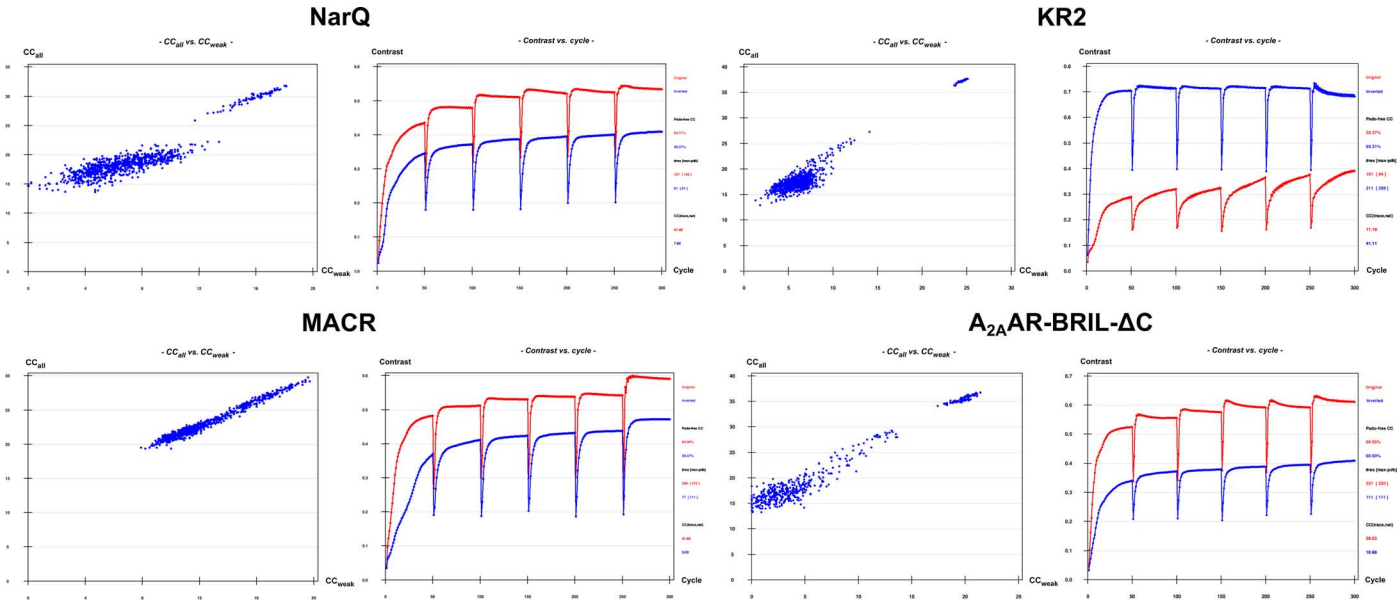


Fig. 2. I-SIRAS phasing of MP crystal structures. Heavy atom substructure determination (SHELXD, left) and phase calculation and extension (SHELXE, right) for I-SIRAS phasing for each of the four targets investigated. As might be expected, I-SIRAS produces both clearer substructure solutions and better contrast in electron density maps phased using the two possible hands (see Fig. 1 for a comparison).

of high-quality I-SIRAS/I-SAD diffraction data for the solution of MP crystal structures.

The positions of the iodide binding sites for all four target proteins are shown in Fig. 4. Iodide binding sites (12, 20, 22, and 6), reproduced in most of the individual data sets for each target (fig. S1), are seen for NarQ, KR2, MACR, and A_{2A}AR-BRIL-ΔC, respectively. As can be seen, the highest-occupancy (as estimated from anomalous difference Fourier maps) sites are clustered close to the predicted membrane interfaces. In the same way as was previously observed in the I-SAD crystal structure solution of soluble proteins (*9*, *10*), the iodide ions interact with a wide variety of positively charged amino acid residues, ordered water molecules, and hydrophobic side chains (Fig. 4 and fig. S3).

To examine the potential universality of the I-SAD method in the de novo solution of MP crystal structures, we examined the availability at the membrane interface of five amino acid residues, namely, arginine, lysine, histidine, tryptophan, and tyrosine in 445 unique transmembrane protein structures deposited in the Protein Data Bank [PDB; www.rcsb.org (*20*); the database of unique MP structures: http://blanco.biomol.uci.edu/mpstruc/]. According to our analysis (Table 4), 96% of the structures contain aromatic amino acids close to the membrane interface, and 100% of the structures contain positively charged amino acids in this region. This, coupled with the recent report of the crystal structure of the outer membrane β-barrel assembly machinery complex, solved by I-SAD phasing (*7*), suggests that I-SAD could be a universal technique for the solution of MP crystal structures.

Bromide–single-wavelength anomalous diffraction

To test whether derivatization with bromide might be an alternative method to the I-SAD/I-SIRAS approach described above, we soaked crystals of NarQ and KR2 in precipitant solutions supplemented with 0.5 M NaBr and collected diffraction data sets as presented in Table 5. For both proteins, despite constructing merged data sets with multiplicities greater than 20, all attempts at Br-SAD/Br-SIRAS structure solution were unsuccessful regardless of whether diffraction data were collected far from (NarQ, λ = 0.873 Å; E = 14.2 keV; f'' = 3.4 e^-) or close to the peak of the Br K-absorption edge (KR2, λ = 0.920 Å; E = 13.476 keV; f'' = 3.8 e^-).

DISCUSSION

The results outlined above suggest that I-SAD is a potentially universal method for the de novo solution of MP crystal structures. Further advantages of I-SAD are speed, nontoxicity, and simplicity because soaking can be easily done at the crystal-harvesting stage and does not require any additional equipment or safety precautions. In addition, iodide soaking does not, at least in the experiments described here, appreciably disturb crystal quality. Another—not insignificant—advantage lies in the anomalous scattering properties of iodide itself. Although anomalous signals in I-SAD experiments would be optimized by targeting the peak of the L_1 absorption edge of iodine (f'' ~ 13.4 e^-; λ = 2.4 Å; E = 5.19 keV), routinely accessing the photon energy required is not always possible even on tuneable synchrotron macromolecular crystallography beamlines. However, at λ = 1.85 Å, a wavelength usually accessible on these beamlines, the anomalous scattering properties of iodide are still significant (f'' = 9.2 e^-), and in many cases, particularly because our studies suggest the presence of several iodide binding sites in each of the different classes of MPs (Fig. 4), experiments at, or close to, this wavelength will be more than sufficient for structure solution. Therefore, we carried out our experiments at λ = 1.85 Å, using a multicrystal approach to produce the highly redundant data (Table 1), which is often required for SAD structure solution (*21*). Our experiments also show that I-SAD phasing of MP crystal structures using data collected in a serial fashion from many microcrystals mounted on a so-called fixed target is also possible. Considering that often it is hard to grow MP crystals of sufficient size for standard data collection, this latter conclusion is very important, suggesting that serial I-SAD data collection methods might also be successful for MP crystal

Table 3. Data collection, structure solution, and initial model building statistics resulting from I-SAD/I-SIRAS phasing of the crystal structure of KR2 using diffraction data collected using serial methods.

		I-SAD	I-SIRAS
Data collection	No. of crystals	136	39
	Method	SAD	SIRAS
	Space group	*I*222	*I*222
Unit cell	*a*	40.6	41.0
	b	83.6	84.0
	c	233.5	234.3
	No. of frames	13,600	3900
	Oscillation range (°)	0.1	0.1
	Resolution range (Å)	60–2.8 (3.0–2.8)	60–2.9 (3.1–2.9)
	Measured reflections	469,213 (85,461)	124,613 (22,610)
	Multiplicity	24.79 (24.13)	7.19 (7.13)
	Completeness (%)	100 (100)	99.8 (100)
	R_{merge} (%)	61.7 (220.7)	23.8 (154.6)
	R_{meas} (%)	63.0 (225.4)	25.6 (166.6)
	$\langle I/\sigma(I)\rangle$	10.64 (2.27)	7.71 (1.69)
	$CC_{1/2}$ (%)	99.8 (70.1)	99.3 (56.2)
	SigAno	1.21 (0.78)	1.03 (0.73)
	CC_{anom} (%)	39 (0)	29 (−1)

		I-SAD	I-SIRAS
Structure solution and Refinement	Initial resolution range (Å)	60–2.8	60–2.9
	Number of residues in the protein	288	288
	Substructure	16·I	12·I
	Resolution cutoff for substructure search (Å)	3.9	3.4
	Resolution extension by native data (Å)	2.5	2.5
	CC_{all}/CC_{weak}	38.7/11.0	26.7/9.6
	Solvent content (%)	63	60
	CC_{map}, highest-resolution shell (%)	47	67
	Number of Ala residues traced by SHELXE	122	165
	R_{work}/R_{free} after initial model building (%)	29.6/35.7†	24.2/31.6* 29.9/35.4†
	Number of residues built in initial model building	207†	248* 196†

*Model built by ARP-wARP. †Model built by phenix.autobuild.

structure solution using diffraction data collected at x-ray free-electron lasers (*22*) or at next-generation diffraction-limited storage ring synchrotron sources.

A notable result of the investigations described here is that Br-SAD does not appear to be a routinely viable method for the de novo solution of the crystal structures of MPs. This is intriguing, particularly because Br-SAD is a relatively well-used technique for the solution of the crystal structures of soluble proteins (*9*). Subsequent analysis of NarQ and KR2 Br-SAD anomalous difference Fourier maps (Fig. 5) revealed peaks indicating both that Br^- binds to these proteins and, moreover, that many of the Br^- binding sites observed coincide with the sites observed for binding of I^- in the corresponding I-SAD experiments. Thus, Br-SAD/Br-SIRAS phasing of MP crystals structures should be possible. However, a

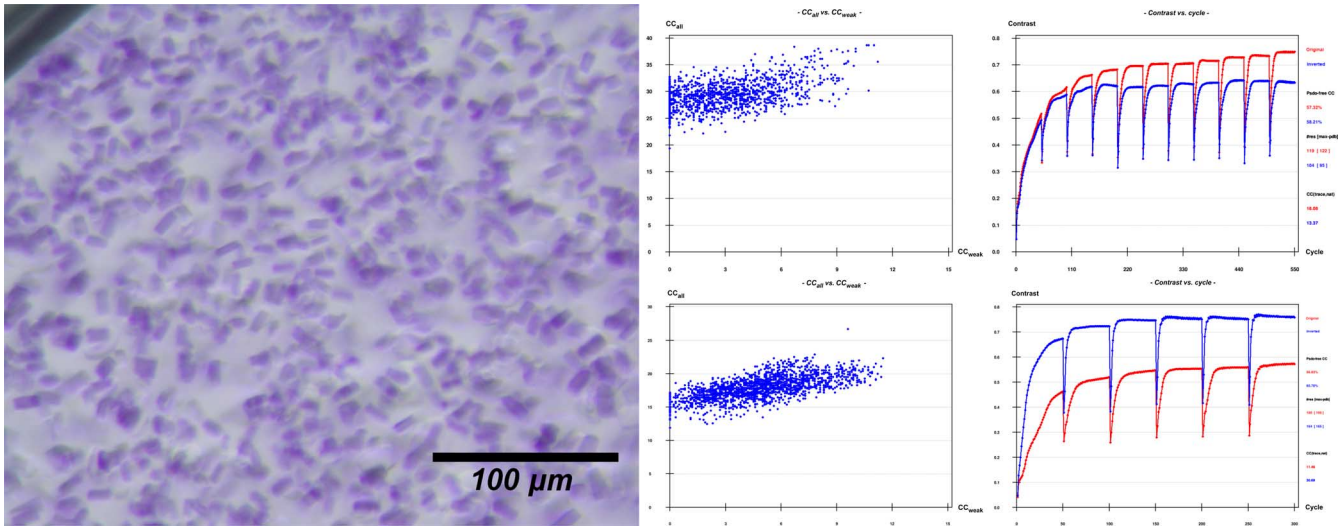


Fig. 3. I-SAD/I-SIRAS phasing of the crystal structure of KR2 using diffraction data collected using serial methods. Left: Photograph of crystals, 10 to 15 µm in maximum dimensions, of KR2 in its monomeric blue form. Top right: Heavy atom substructure determination (SHELXD, left) and phase calculation and extension (SHELXE, right) for I-SAD phasing using data collected via serial methods. Bottom right: Heavy atom substructure determination (SHELXD, left) and phase calculation and extension (SHELXE, right) for I-SIRAS phasing using data collected via serial methods.

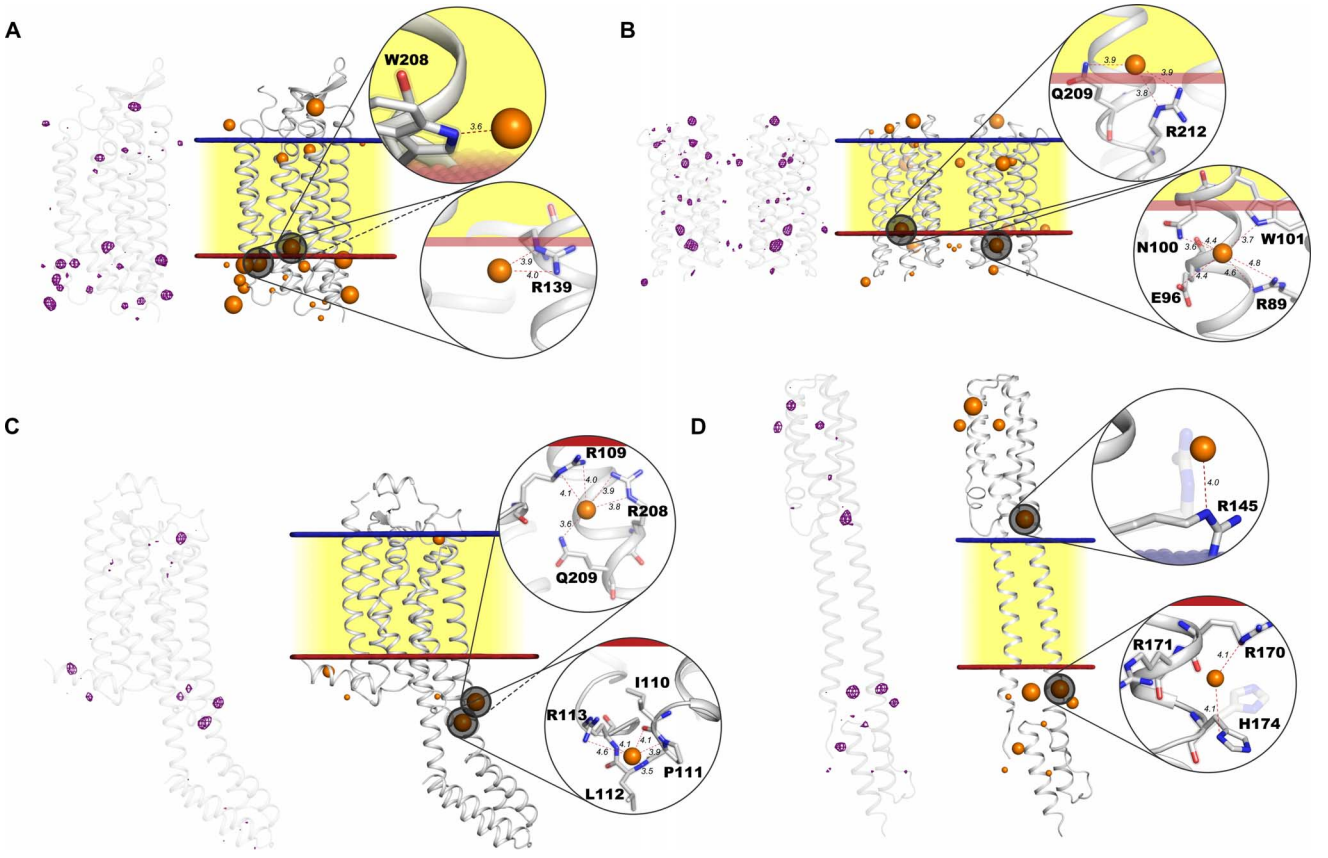


Fig. 4. Cartoon representations of the crystal structures of the four classes of MP solved using the I-SAD/I-SIRAS technique. (**A**) KR2. (**B**) MACR (noncrystallographic dimer). (**C**) A_{2A}AR-BRIL-ΔC. (**D**) NarQ. On the left of each representation, peaks in I-SAD anomalous difference Fourier maps [purple chicken wire, highest-resolution I-SAD data sets for each target (Table 1), contoured at the 3.5 × r.m.s. (root mean square) level] are shown superposed on the C_α backbone of the protein. On the right of each panel, the blue and red lines represent outer and inner lipidic membrane surfaces, respectively, with the hydrophobic region of the lipidic membrane represented in yellow. Iodide ions are shown as orange spheres, divided into three sizes based on the height of anomalous difference map peaks. Two iodide binding sites per crystal structure are highlighted to illustrate their environment. In (D), it is clear that protein residues may change their side-chain conformation upon binding of iodide (conformation in native structures shown as shadows).

Table 4. The analysis of presence at the membrane hydrophobic-hydrophilic interface of arginine (R), lysine (K), histidine (H), tryptophan (W), and tyrosine (Y). As outlined in the main text, 445 unique transmembrane protein crystal structures deposited in the PDB were analyzed with, in each case, the membrane surfaces represented by two parallel planes calculated by Lomize *et al.* (*42*). Amino acid residues were considered to be present at the membrane interfaces if its C_α atom was not more than 7 Å from the membrane surfaces defined above.

	R/K/H	R/K	W/Y	R	K	H	W	Y
Number of entries where the amino acid residues are present at the membrane interface	445 (100%)	442 (99.3%)	428 (96.2%)	429 (96.4%)	414 (93.0%)	353 (79.3%)	393 (88.3%)	422 (94.8%)

Table 5. Br-SAD data collection statistics. Entries in parentheses represent values for the highest-resolution bin. All statistics and data quality indicators are calculated treating Friedel's pairs as separate reflections. Linear merging *R* value $R_{merge} = \frac{\sum_h \sum_{i=1}^{N} |I_{(h)i} - \langle I_{(h)} \rangle|}{\sum_h \sum_{i=1}^{N} I_{(h)i}}$; redundancy-independent merging *R* value $R_{meas} = \frac{\sum_h \sqrt{\frac{N}{N-1}} \sum_{i=1}^{N} |I_{(h)i} - \langle I_{(h)} \rangle|}{\sum_h \sum_{i=1}^{N} I_{(h)i}}$; $CC_{1/2}$—correlation coefficient between intensities from random half–data sets; SigAno $= \left\langle \frac{|F_+ - F_-|}{\sigma(F_+ - F_-)} \right\rangle$.

Protein/ space group	Crystal/ data set	λ (Å)	Unit cell dimensions (Å, °)						Resolution (Å)	Multiplicity	Completeness (%)	R_{merge} (%)	R_{meas} (%)	<I/σ(I)>	$CC_{1/2}$ (%)	SigAno	CC_{anom} (%)	$SigAno_{inner}$	CC_{anom_inner} (%)
			a	*b*	*c*	α	β	γ											
NarQ *F*222	1	0.8729	57.87	74.79	237.85	90	90	90	60–2.8 (3.0–2.8)	2.09 (2.11)	83.2 (45.2)	14.5 (105.4)	18.2 (132.8)	5.53 (1.21)	99.4 (47.7)	0.82 (0.71)	6 (10)	1.12	37
	2	0.8729	57.70	74.34	237.70	90	90	90	60–2.8 (3.0–2.8)	4.03 (4.12)	98.2 (92.6)	18.1 (106.9)	20.8 (122.7)	6.99 (1.36)	99.5 (58.3)	0.83 (0.74)	8 (3)	1.56	55
	3	0.8729	58.21	73.57	237.42	90	90	90	60–2.8 (3.0–2.8)	4.81 (4.87)	99.2 (99.2)	21.3 (105.1)	24.0 (117.9)	5.85 (1.43)	99.5 (65.2)	0.82 (0.67)	13 (1)	1.27	45
	14	0.8729	57.53	73.91	237.23	90	90	90	60–2.6 (2.8–2.6)	4.31 (4.06)	93.2 (68.4)	15.7 (101.1)	17.9 (116.5)	7.38 (1.26)	99.5 (60.1)	0.86 (0.71)	11 (3)	1.50	61
	15	0.8729	57.52	74.44	237.76	90	90	90	60–2.8 (3.0–2.8)	2.58 (2.40)	93.5 (87.6)	14.0 (59.6)	17.5 (74.6)	5.83 (1.47)	98.9 (66.1)	0.85 (0.74)	18 (–2)	1.14	56
	16	0.8729	57.76	74.22	238.68	90	90	90	60–2.6 (2.8–2.6)	2.89 (2.98)	97.4 (98.4)	10.8 (73.5)	13.0 (88.8)	8.13 (1.55)	99.6 (59.7)	0.87 (0.75)	9 (2)	1.80	74
	24	0.8729	57.51	74.20	237.26	90	90	90	60–2.2 (2.4–2.2)	2.31 (2.22)	91.6 (76.8)	11.4 (125.6)	14.5 (160.1)	5.85 (0.62)	99.6 (69.2)	0.83 (0.65)	12 (4)	1.69	79
	Merged		57.73	74.21	237.70	90	90	90	60–2.8 (3.0–2.8)	22.3 (21.3)	94.2 (100)	19.7 (85.9)	20.1 (88.0)	17.3 (5.0)	99.9 (96.6)	0.98 (0.75)	20 (3)	2.77	90
KR2 *I*222	32	0.91968	40.67	83.73	234.46	90	90	90	60–2.8 (3.0–2.8)	6.68 (6.45)	89.4 (44.7)	24.0 (176.7)	26.1 (191.6)	7.44 (1.11)	99.6 (49.4)	0.97 (0.65)	43 (3)	1.75	67
	33	0.91968	40.78	83.62	234.38	90	90	90	60–3.0 (3.2–3.0)	6.82 (7.05)	100 (100)	26.3 (159.7)	28.6 (172.5)	7.43 (1.33)	99.6 (47.5)	0.92 (0.70)	21 (–8)	1.87	71
	37	0.91968	40.67	83.32	234.10	90	90	90	60–2.8 (3.0–2.8)	6.77 (6.64)	99.5 (99.2)	16.1 (118.5)	17.4 (128.7)	9.85 (1.61)	99.8 (78.1)	0.88 (0.72)	17 (7)	1.69	67
	38	0.91968	40.58	83.31	234.04	90	90	90	60–2.8 (3.0–2.8)	6.85 (7.05)	99.6 (99.4)	14.4 (103.8)	15.6 (112.3)	10.06 (1.73)	99.8 (84.9)	0.90 (0.76)	19 (7)	1.75	69
	39	0.91968	40.56	83.46	233.45	90	90	90	60–3.0 (3.2–3.0)	6.74 (6.54)	96.1 (79.3)	17.0 (121.3)	18.4 (131.7)	9.52 (1.35)	99.9 (74.9)	0.90 (0.73)	21 (14)	1.65	61
	Merged		40.65	83.49	234.09	90	90	90	60–2.8 (3.0–2.8)	31.56 (23.48)	100 (100)	23.9 (166.0)	24.3 (169.7)	16.72 (2.49)	99.9 (85.7)	1.03 (0.74)	23 (0)	2.63	80

major disadvantage of Br-SAD experiments is that the maximum f'' that can be induced around the Br K-absorption edge is ~4 e^-. This compares to f'' ~9.2 e^- for I-SAD experiments carried out at λ = 1.85 Å. Anomalous signals in Br-SAD experiments will thus be significantly lower than for long-wavelength I-SAD (see Fig. 6 for an illustration of this). Successful MP Br-SAD experiments will thus demand higher-accuracy data than I-SAD. This will generally require higher multiplicity measurements and/or the merging of more individual data sets,

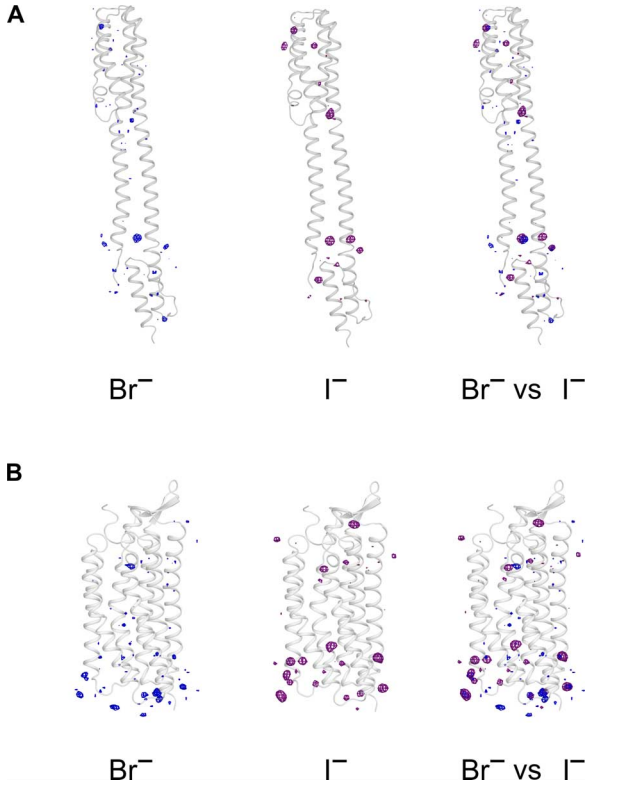


Fig. 5. Comparison of peaks in anomalous difference Fourier maps in I-SAD and Br-SAD experiments. (**A**) Peaks in anomalous difference Fourier maps in Br-SAD (left) and I-SAD (middle) experiments for NarQ, with a superposition of the two maps shown on the right. (**B**) Peaks in anomalous difference Fourier maps in Br-SAD (left) and I-SAD (middle) experiments for KR2, with a superposition of the two maps shown on the right. Only the regions close to the proteins are shown, and the maps are contoured at the 3.5 × r.m.s. level. For both NarQ and KR2, the anomalous difference Fourier maps shown are those calculated using anomalous differences in the individual I-SAD data sets of the highest resolution, as detailed in Table 1, and in the merged Br-SAD data sets, as detailed in Table 5.

and Br-SAD will thus be less conducive to high-throughput MP structure determination.

MATERIALS AND METHODS

Protein expression and purification

Expression and purification of the sodium-pumping rhodopsin KR2 (UniProt ID N0DKS8) and the chimeric protein A_{2A}AR-BRIL-ΔC of human A_{2A} adenosine receptor (A_{2A}AR) with a thermostabilized apocytochrome b_{562}RIL (UniProt IDs P0ABE7 and P29274) in complex with ZM241385 were performed as described by Gushchin *et al.* (*23*) and Liu *et al.* (*24*), respectively.

Expression and purification of the nitrate/nitrite sensor kinase NarQ

The nucleotide sequence encoding residues 1 to 230 of the nitrate/nitrite sensor kinase NarQ (*17*) (UniProt ID P27896) was cloned from *E. coli* strain BL21 (DE3) and introduced into the pSCodon1.2 expression vector (StabyCodon T7, Eurogentec) via Nde I and Xho I restriction sites. Consequently, the construct harbored a C-terminal His_6 tag. NarQ was then expressed in *E. coli* strain SE1 (StabyCodon T7, Eurogentec). Cells were cultured in shaking baffled flasks in ZYP-5052 auto-inducing medium (*25*) containing ampicillin (100 mg/liter). After incubating for 2 hours at 37°C, the temperature was decreased to 30°C, and incubation continued overnight. Harvested cells were disrupted in M-110P Lab Homogenizer (Microfluidics) at 25,000 psi in a phosphate-buffered saline buffer with addition of deoxyribonuclease (DNase) I (50 mg/liter; Sigma-Aldrich) and EDTA-free protease inhibitor cocktail (cOmplete, Roche). The membrane fraction of cell lysate was isolated by ultracentrifugation at 90,000*g* for 1 hour at 4°C. The pellet was resuspended in a buffer containing 50 mM NaH_2PO_4/Na_2HPO_4 (pH 8.0), 0.3 M NaCl, and 2% *n*-dodecyl-β-D-maltoside (DDM) (Anatrace, Affymetrix) and stirred overnight for solubilization. The insoluble fraction was then removed by ultracentrifugation at 90,000*g* for 1 hour at 4°C. The supernatant was loaded on a gravity flow column containing Ni-NTA resin (Qiagen), and the protein was eluted in a buffer containing 20 mM tris-HCl (pH 8.0), 0.3 M NaCl, 0.2 M imidazole, and 0.1% DDM. Imidazole was then removed by dialysis against 20 mM tris-HCl (pH 8.0), 0.3 M NaCl, and 0.1% DDM for 3 hours. The eluate was subjected to size-exclusion chromatography on a 125-ml Superdex 200 PG column (GE Healthcare Life Sciences) in a buffer containing 20 mM tris-HCl (pH 8.0), 0.3 M NaCl, and 0.1% DDM. Protein-containing fractions were pooled and concentrated to 30 mg/ml for crystallization.

Expression and purification of proton-pumping rhodopsin MACR

The nucleotide sequence encoding *Candidatus Actinomarina minuta* opsin gene (MACR, UniProt ID S5DM51) (*16*) was cloned from metagenomic fosmid MedDCM-OCT-S44-C50 without any optimization and was introduced into the pSCodon1.2 vector via Xba I and Bam HI restriction sites. Consequently, the expressed construct harbored an additional C-terminal tag with a sequence PGGGSHHHHHH. *E. coli* strain SE1 cells were transformed with the pSC-MACR-His6 plasmid. The cells were grown at 37°C in shaking baffled flasks in ZYP-5052 auto-inducing medium containing ampicillin (100 mg/liter). After the glucose level in the growing bacterial culture dropped below 10 mg/liter, 10 μM all-trans-retinal (Sigma-Aldrich) was added, the temperature was reduced to 20°C, and incubation continued overnight. Collected cells were disrupted using the M-110P Lab Homogenizer at 25,000 psi in a buffer containing 20 mM tris-HCl (pH 8.0), 5% glycerol, 0.5% Triton X-100 (Sigma-Aldrich), and DNase I (50 mg/liter). The membrane fraction of cell lysate was obtained by ultracentrifugation at 90,000*g* for 1 hour at 4°C. The pellets were resuspended in a buffer containing 50 mM NaH_2PO_4/Na_2HPO_4 (pH 8.0), 0.1 M NaCl, and 1% DDM. The mixture was left overnight for solubilization. The insoluble fraction was removed by ultracentrifugation at 90,000*g* for 1 hour at 4°C. The supernatant was loaded on a Ni-NTA column, and the His-tagged protein was eluted in a buffer containing 50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5), 0.1 M NaCl, 0.5 M imidazole, and 0.2% DDM. The eluate was subjected to size-exclusion chromatography on a 125-ml Superdex 200 PG column in a buffer containing 50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5), 0.1 M NaCl, and 0.01% DDM. Protein-containing colored fractions were collected and concentrated to 40 mg/ml for crystallization.

Protein crystallization and iodide soaking

All crystals were grown using the in meso approach used in some of our previous work (*23*, *24*, *26*, *27*). NarQ, MACR, and KR2 in their respective crystallization buffers (see above) were added to the monoolein-formed lipidic phase (Nu-Chek Prep), whereas the solubilized A_{2A}AR-BRIL-ΔC/ZM241385 complex was mixed with monoolein containing 10% (w/w) cholesterol. The protein-LCP (lipid cubic phase) mixtures (40- to 50-nl aliquots in the case of A_{2A}AR-BRIL-ΔC and 100-nl aliquots in all other cases) were spotted on 96-well LCP glass sandwich plates

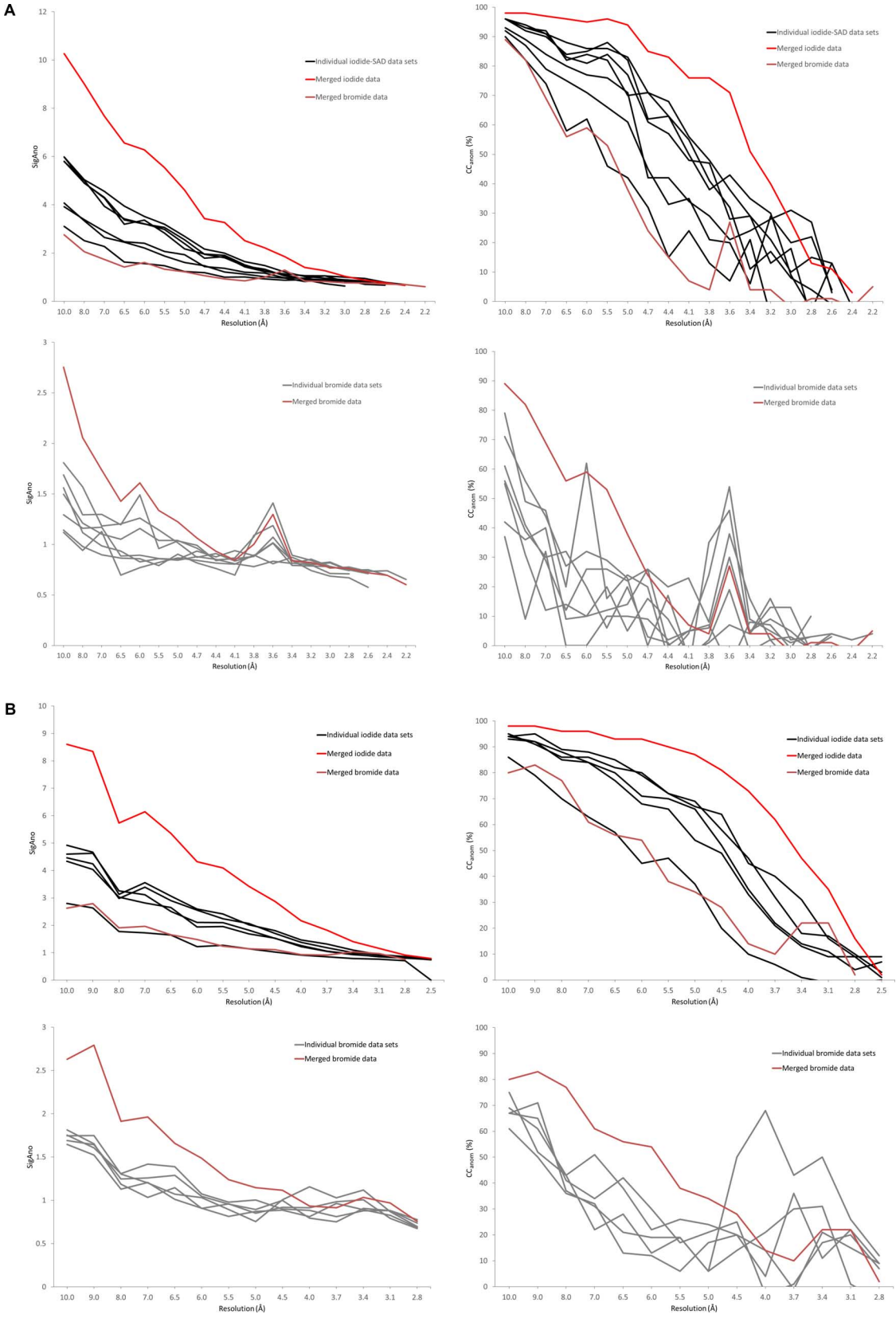


Fig. 6. Comparison of anomalous signal in I-SAD and Br-SAD diffraction data sets for NarQ (A) and KR2 (B). (**A**) Comparisons of SigAno (left) and CC_{anom} (right) in I-SAD (top panel) and Br-SAD (bottom) panel data sets measured from derivatized NarQ crystals. (**B**) Comparisons of SigAno (left) and CC_{anom} (right) in I-SAD (top panel) and Br-SAD (bottom) panel data sets measured from KR2 crystals. Individual and merged data sets for I-SAD experiments are as detailed in Table 1, whereas those for Br-SAD are as detailed in Table 5.

(Marienfeld) and covered with 800 nl of precipitant solution (see below) using the NT8-LCP crystallization robot (Formulatrix). Crystals were grown at 20°C and reached their final sizes within 2 to 8 weeks.

The best crystals of NarQ were obtained using 0.6 M KH_2PO_4/Na_2HPO_4 (pH 4.6) (Qiagen) and 5 mM $NaNO_3$ as the precipitant solution and grew to 50 to 100 µm in size. For KR2, the best crystals were obtained using 2.0 M sodium malonate (pH 4.3) (Hampton Research) as a precipitant solution and grew to 70 to 100 µm in size. The best crystals of MACR were obtained using 2.6 M $(NH_4)_2SO_4$ and 0.1 M sodium acetate pH 5.2 (Qiagen) as the precipitant and were 100 to 150 µm in size. The best crystals of the A_{2A}AR-BRIL-ΔC/ZM241385 complex were obtained using a precipitant solution comprising 32% polyethylene glycol 400, 25 mM NaSCN, and 0.1 M sodium acetate (pH 5.0) and grew to 40 to 60 µm in size. For serial crystallography SAD experiments, we obtained crystals of KR2 (10 to 15 µm in size) (Fig. 3) using 2.5 M sodium malonate (pH 4.3) (Hampton Research) as the precipitant.

Once crystals reached their final size, crystallization wells were opened as described elsewhere (*28*), and drops containing the protein-mesophase mixture were covered with 50 µl of the respective precipitant solution. For native data collection (Table 1), harvested crystals were incubated for ~5 min in the respective precipitant solutions. In the case of NarQ, this was supplemented with 20% glycerol as a cryoprotectant. For I-SAD experiments, the incubation solution was supplemented with 0.5 M NaI. For Br-SAD experiments, the incubation solution was supplemented with 0.5 M NaBr. After incubation, crystals were loaded onto MicroMounts and MicroMeshes (MiTeGen), flash-cooled, and stored in liquid nitrogen.

X-ray crystallography

X-ray diffraction data were collected on the European Synchrotron Radiation Facility (ESRF) beamlines ID23-1 (*29*) or ID23-2 (*30*) equipped with Dectris pixel array detectors. For I-SAD experiments, diffraction data from crystals soaked in NaI were collected at λ = 1.85 Å. The x-ray beam size at the sample position was varied from 10 to 50 µm in maximum dimension depending on the crystal size. Experimental parameters for optimal data collection were designed using the program BEST (*31*). Diffraction images were processed with XDS (*32*) and XSCALE (*32*), which was also used to merge and scale different data sets (see Table 1 for details). Structure factor amplitudes, anomalous differences (DANO), and FreeR labels were then generated using the CCP4 programs POINTLESS and AIMLESS (*33*).

The KR2 I-SAD serial data collection was carried out as described by Zander *et al.* (*19*) using x-rays of λ = 1.85 Å. Here, 171 partial data sets were successfully processed, and hierarchical cluster analysis (*34*) was then used to select 136 of these for merging/scaling to produce the final data set. Structure solution and refinement were then carried out as described below for data sets obtained in a "standard" fashion. For the serial crystallography I-SIRAS experiments, the native data set was collected in a standard way, and only 39 of the I-SAD partial data sets were merged to provide the derivative data set.

For Br-SAD experiments, diffraction data from crystals of NarQ soaked in 0.5 M NaBr were collected at λ = 0.872 Å, whereas for KR2 crystals soaked in 0.5 M NaBr, diffraction data were collected at λ = 0.9198 Å. Data processing, merging, and reduction were then carried out as described above for I-SAD experiments (see Table 5 for details). The native data sets were required for both I-SIRAS and Br-SIRAS structure solution protocols and were recorded at either λ = 1.85 Å (KR2 only) or λ = 0.972 Å, with data processing and reduction carried out as described above for I-SAD experiments (see Table 1 for details).

Structure solution and refinement

I-SAD and I-SIRAS phasing protocols were carried out using the SHELXC/D/E pipeline as implemented in HKL2MAP (1000 trials for substructure determination) (*35*, *36*). ARP-wARP (*37*) or phenix. autobuild (*38*) was then used for automated model building. Final structural models (Table 2) were obtained by alternating cycles of manual building in Coot (*39*) with refinement in REFMAC5 (*40*). Figures illustrating the final models obtained were prepared using PyMOL (*41*). Anomalous difference Fourier maps were calculated with fast Fourier transform (FFT) (CCP4) using DANO and α_{calc} + 90° as coefficients (α_{calc} = calculated phases from the final refined models) and were then used to verify the number, position, and occupancy of iodide ions in the individual and the merged I-SAD data sets (Tables 1 and 2 and fig. S1), with, for each target, only those common to at least half of the data sets included in the final models.

Br-SAD and Br-SIRAS de novo phasing protocols for NarQ and KR2 were also carried out using the SHELXC/D/E pipeline as implemented in HKL2MAP (10,000 trials for substructure determination). However, none of our attempts resulted in successful structure solution. This being the case, structure refinement based on the model obtained from native crystals was carried out using REFMAC5 and Coot. Anomalous difference Fourier maps were then calculated with FFT (CCP4) using DANO and α_{calc} + 90° as coefficients (α_{calc} = calculated phases from the final refined models) and were used to verify the number, position, and occupancy of bromide ions in each of the individual and the merged Br-SAD data sets.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/3/5/e1602952/DC1

fig. S1. Comparison of peaks in anomalous difference Fourier maps calculated from individual data sets in I-SAD experiment.

fig. S2. The distribution of positively charged and aromatic residues in the crystal structures obtained in I-SAD experiment.

fig. S3. The bound iodide ions and their environment.

REFERENCES AND NOTES

1. E. Wallin, G. Von Heijne, Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7**, 1029–1038 (1998).
2. J. P. Overington, B. Al-Lazikani, A. L. Hopkins, How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).
3. J. P. Morth, T. L. Sørensen, P. Nissen, Membrane's eleven: Heavy-atom derivatives of membrane-protein crystals. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 877–882 (2006).
4. S. Doublié, Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
5. T. Weinert, V. Olieric, S. Waltersperger, E. Panepucci, L. Chen, H. Zhang, D. Zhou, J. Rose, A. Ebihara, S. Kuramitsu, D. Li, N. Howe, G. Schnapp, A. Pautsch, K. Bargsten, A. E. Prota, P. Surana, J. Kottur, D. T. Nair, F. Basilico, V. Cecatiello, S. Pasqualato, A. Boland, O. Weichenrieder, B.-C. Wang, M. O. Steinmetz, M. Caffrey, M. Wang, Fast native-SAD phasing for routine macromolecular structure determination. *Nat. Methods* **12**, 131–133 (2015).
6. D. Li, V. E. Pye, M. Caffrey, Experimental phasing for structure determination using membrane-protein crystals grown by the lipid cubic phase method. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 104–122 (2015).
7. Y. Gu, H. Li, H. Dong, Y. Zeng, Z. Zhang, N. G. Paterson, P. J. Stansfeld, Z. Wang, Y. Zhang, W. Wang, C. Dong, Structural basis of outer membrane protein insertion by the BAM complex. *Nature* **531**, 64–69 (2016).

8. Y. Guo, R. C. Kalathur, Q. Liu, B. Kloss, R. Bruni, C. Ginter, E. Kloppmann, B. Rost, W. A. Hendrickson, Structure and activity of tryptophan-rich TSPO proteins. *Science* **347**, 551–555 (2015).
9. Z. Dauter, M. Dauter, K. R. Rajashankar, Novel approach to phasing proteins: Derivatization by short cryo-soaking with halides. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 232–237 (2000).
10. J. Abendroth, A. S. Gardberg, J. I. Robinson, J. S. Christensen, B. L. Staker, P. J. Myler, L. J. Stewart, T. E. Edwards, SAD phasing using iodide ions in a high-throughput structural genomics environment. *J. Struct. Funct. Genomics* **12**, 83–95 (2011).
11. G. von Heijne, Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**, 456–458 (1989).
12. G. V. Heijne, Membrane proteins: From sequence to structure. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 167–192 (1994).
13. G. von Heijne, Membrane protein structure prediction: Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**, 487–494 (1992).
14. W.-M. Yau, W. C. Wimley, K. Gawrisch, S. H. White, The preference of tryptophan for membrane interfaces. *Biochemistry* **37**, 14713–14718 (1998).
15. K. Inoue, H. Ono, R. Abe-Yoshizumi, S. Yoshizawa, H. Ito, K. Kogure, H. Kandori, A light-driven sodium ion pump in marine bacteria. *Nat. Commun.* **4**, 1678 (2013).
16. R. Ghai, C. M. Mizuno, A. Picazo, A. Camacho, F. Rodriguez-Valera, Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci. Rep.* **3**, 2471 (2013).
17. V. Stewart, Nitrate- and nitrite-responsive sensors NarX and NarQ of proteobacteria. *Biochem. Soc. Trans.* **31**, 1–10 (2003).
18. V.-P. Jaakola, M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. T. Chien, J. R. Lane, A. P. IJzerman, R. C. Stevens, The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. *Science* **322**, 1211–1217 (2008).
19. U. Zander, G. Bourenkov, A. N. Popov, D. de Sanctis, O. Svensson, A. A. McCarthy, E. Round, V. Gordeliy, C. Mueller-Dieckmann, G. A. Leonard, MeshAndCollect: An automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 2328–2343 (2015).
20. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
21. Z. Dauter, D. A. Adamiak, Anomalous signal of phosphorus used for phasing DNA oligomer: Importance of data redundancy. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 990–995 (2001).
22. A. Batyuk, L. Galli, A. Ishchenko, G. W. Han, C. Gati, P. A. Popov, M.-Y. Lee, B. Stauch, T. A. White, A. Barty, A. Aquila, M. S. Hunter, M. Liang, S. Boutet, M. Pu, Z.-j. Liu, G. Nelson, D. James, C. Li, Y. Zhao, J. C. H. Spence, W. Liu, P. Fromme, V. Katritch, U. Weierstall, R. C. Stevens, V. Cherezov, Native phasing of x-ray free-electron laser data for a G protein-coupled receptor. *Sci. Adv.* **2**, e1600292 (2016).
23. I. Gushchin, V. Shevchenko, V. Polovinkin, K. Kovalev, A. Alekseev, E. Round, V. Borshchevskiy, T. Balandin, A. Popov, T. Gensch, C. Fahlke, C. Bamann, D. Willbold, G. Büldt, E. Bamberg, V. Gordeliy, Crystal structure of a light-driven sodium pump. *Nat. Struct. Mol. Biol.* **22**, 390–395 (2015).
24. W. Liu, E. Chun, A. A. Thompson, P. Chubukov, F. Xu, V. Katritch, G. W. Han, C. B. Roth, L. H. Heitman, A. P. IJzerman, V. Cherezov, R. C. Stevens, Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **337**, 232–236 (2012).
25. F. W. Studier, Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
26. V. I. Gordeliy, J. Labahn, R. Moukhametzianov, R. Efremov, J. Granzin, R. Schlesinger, G. Büldt, T. Savopol, A. J. Scheidig, J. P. Klare, M. Engelhard, Molecular basis of transmembrane signalling by sensory rhodopsin II–transducer complex. *Nature* **419**, 484–487 (2002).
27. V. I. Gordeliy, R. Schlesinger, R. Efremov, G. Büldt, J. Heberle, Crystallization in lipidic cubic phases: A case study with bacteriorhodopsin. *Methods Mol. Biol.* **228**, 305–316 (2003).
28. D. Li, C. Boland, D. Aragao, K. Walsh, M. Caffrey, Harvesting and cryo-cooling crystals of membrane proteins grown in lipidic mesophases for structure determination by macromolecular crystallography. *J. Vis. Exp.* **2012**, e4001 (2012).
29. D. Nurizzo, T. Mairs, M. Guijarro, V. Rey, J. Meyer, P. Fajardo, J. Chavanne, J.-C. Biasci, S. McSweeney, E. Mitchell, The ID23-1 structural biology beamline at the ESRF. *J. Synchrotron Radiat.* **13**, 227–238 (2006).
30. D. Flot, T. Mairs, T. Giraud, M. Guijarro, M. Lesourd, V. Rey, D. van Brussel, C. Morawe, C. Borel, O. Hignette, J. Chavanne, D. Nurizzo, S. McSweeney and E. Mitchell, The ID23-2 structural biology microfocus beamline at the ESRF. *J. Synchrotron Radiat.* **17**, 107–118 (2010).
31. G. P. Bourenkov, A. N. Popov, Optimization of data collection taking radiation damage into account. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 409–419 (2010).
32. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
33. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
34. R. Giordano, R. M. F. Leal, G. P. Bourenkov, S. McSweeney, A. N. Popov, The application of hierarchical cluster analysis to the selection of isomorphous crystals. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 649–658 (2012).
35. T. Pape, T. R. Schneider, HKL2MAP: A graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843–844 (2004).
36. G. M. Sheldrick, Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
37. G. Langer, S. X. Cohen, V. S. Lamzin, A. Perrakis, Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
38. T. C. Terwilliger, R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L.-W. Hung, R. J. Read, P. D. Adams, Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69 (2008).
39. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
40. G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
41. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.
42. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).

Acknowledgments

Funding: The work on protein expression, purification and crystallization was supported by the program "Chaires d'excellence" edition 2008 of ANR France, CEA(IBS) – HGF(FZJ) STC 5.1 specific agreement, and used the platforms of the Grenoble Instruct Centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The work on photoactive proteins was supported by Ministry of Education and Science of the Russian Federation (RFMEFI58716X0026). The work on GPCR expression, purification, and crystallization was supported by the Russian Science Foundation (project no. 16-14-10273). I.M. is the recipient of an ESRF PhD Studentship. **Author contributions:** A.P. and V.G. conceived the research. I.M., V.G., and A.P. prepared the manuscript with the contribution of V.P., G.A.L., and all other authors. K.K., I.G., A.A., and M.S. expressed and purified the proteins. V.S., A.M., and V.C. oversaw the expression and purification. V.P., I.M., and M.S. crystallized the proteins. V.C. supervised GPCR crystallization and data analysis. V.B. and A.M. oversaw the crystallization. I.M., V.P., V.B., M.S., and A.P. collected and analyzed the diffraction data. I.M. and A.P. solved the structures. I.G., A.A., K.K., M.S., and V.C. helped with structure analysis. F.R.-V. contributed to the work on MACR protein production. A.P., G.A.L., and V.G. oversaw the research. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. All diffraction data and refined models have been deposited in the PDB: 5JGP (NarQ), 5JRF (KR2), 5JSI (MACR), and 5JTB (A_{2A}AR-BRIL-ΔC).

Submitted 24 November 2016
Accepted 15 March 2017
Published 12 May 2017
10.1126/sciadv.1602952

Citation: I. Melnikov, V. Polovinkin, K. Kovalev, I. Gushchin, M. Shevtsov, V. Shevchenko, A. Mishin, A. Alekseev, F. Rodriguez-Valera, V. Borshchevskiy, V. Cherezov, G. A. Leonard, V. Gordeliy, A. Popov, Fast iodide-SAD phasing for high-throughput membrane protein structure determination. *Sci. Adv.* **3**, e1602952 (2017).

ScienceAdvances

Fast iodide-SAD phasing for high-throughput membrane protein structure determination

Igor Melnikov, Vitaly Polovinkin, Kirill Kovalev, Ivan Gushchin, Mikhail Shevtsov, Vitaly Shevchenko, Alexey Mishin, Alexey Alekseev, Francisco Rodriguez-Valera, Valentin Borshchevskiy, Vadim Cherezov, Gordon A. Leonard, Valentin Gordeliy and Alexander Popov

Sci Adv **3** (5), e1602952.
DOI: 10.1126/sciadv.1602952

Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/3/5/e1602952/DC1

Supplementary Materials for

Fast iodide-SAD phasing for high-throughput membrane protein structure determination

Igor Melnikov, Vitaly Polovinkin, Kirill Kovalev, Ivan Gushchin, Mikhail Shevtsov, Vitaly Shevchenko, Alexey Mishin, Alexey Alekseev, Francisco Rodriguez-Valera, Valentin Borshchevskiy, Vadim Cherezov, Gordon A. Leonard, Valentin Gordeliy, Alexander Popov

Published 12 May 2017, *Sci. Adv.* **3**, e1602952 (2017)
DOI: 10.1126/sciadv.1602952

This PDF file includes:

Supplementary Materials

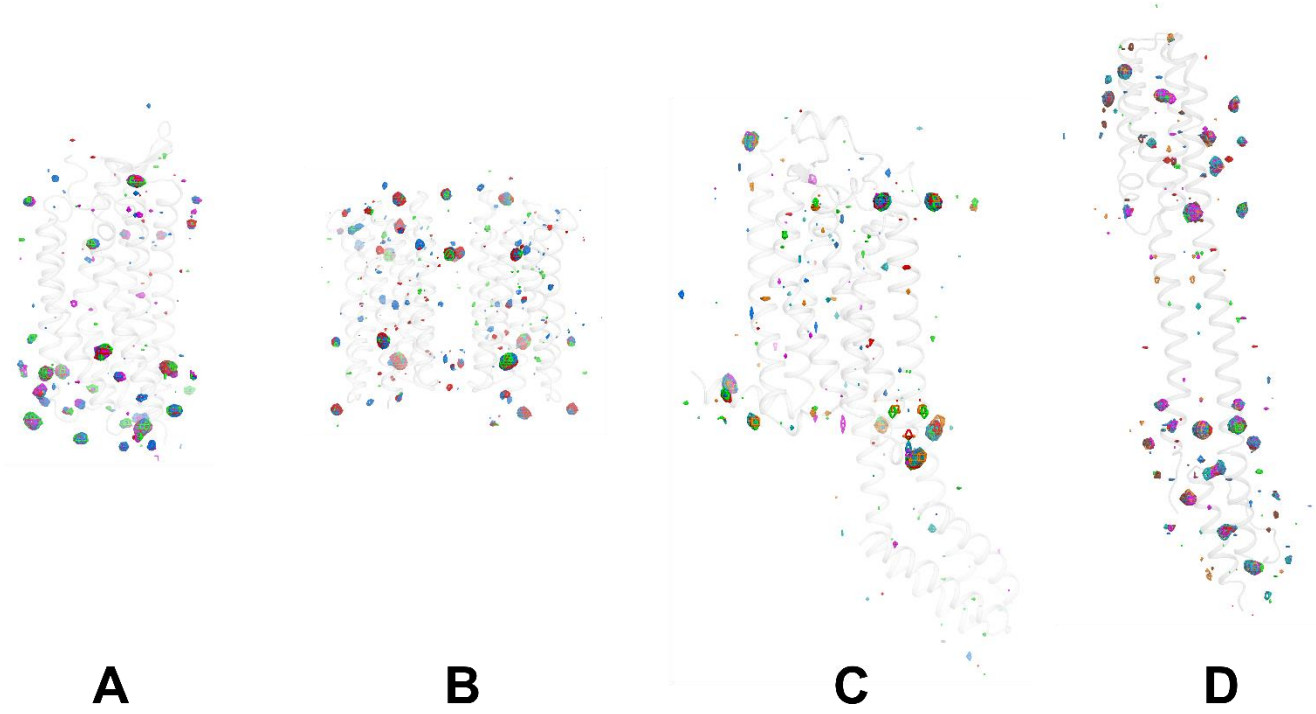


fig. S1. Comparison of peaks in anomalous difference Fourier maps calculated from individual data sets in I-SAD experiment. Only the regions close to the proteins are shown and the maps are contoured at the 3.5 x r.m.s. (root mean square) level. Each individual data set (see Table 1) is represented by its individual color (red, green, marine, teal, orange, magenta, firebrick, skyblue) for each of the structures displayed: (**A**) KR2, (**B**) MACR, (**C**) A_{2A}AR-BRIL-ΔC, (**D**) NarQ.

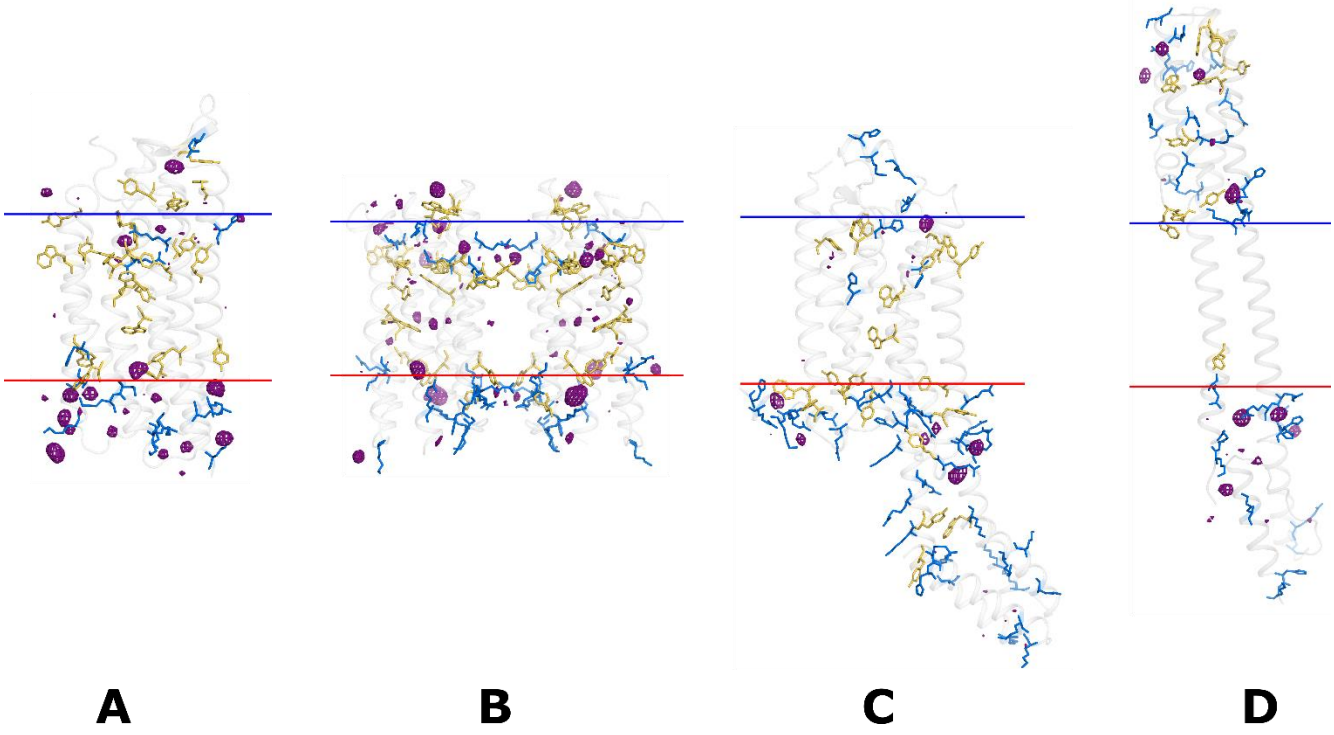


fig. S2. The distribution of positively charged and aromatic residues in the crystal structures obtained in I-SAD experiment. Positively charged protein residues (Arg, Lys, His) are represented as blue sticks, aromatic residues (Trp, Tyr) are represented as yellow sticks. The anomalous difference Fourier maps from I-SAD experiment (purple chicken wire, calculated from highest-resolution I-SAD data sets for each target (Table 1), contoured at 3.5 x r.m.s. level) are shown superposed on the structures of KR2 (**A**), MACR (**B**), A_{2A}AR-BRIL-ΔC (**C**), NarQ (**D**). The blue and red lines represent outer and inner lipidic membrane surfaces, respectively.

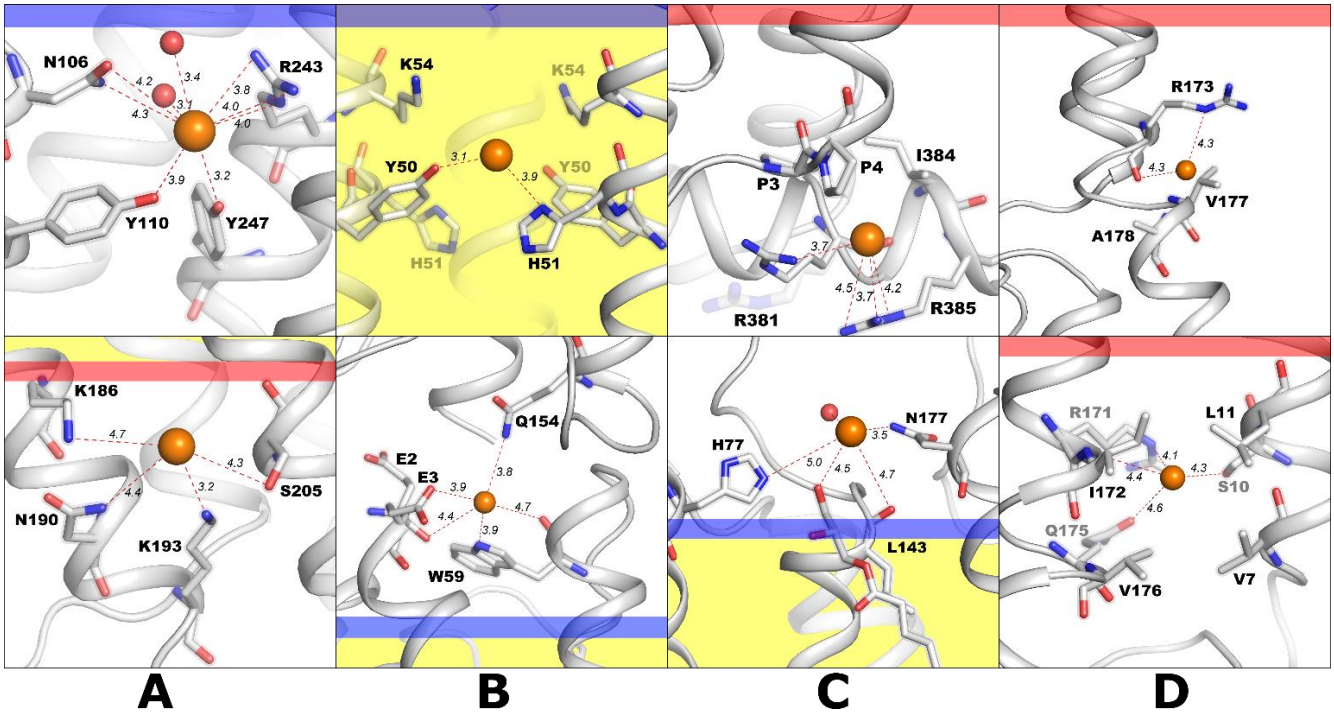


fig. S3. The bound iodide ions and their environment. Two figures (top and bottom) illustrating iodide environment at different binding sites are shown for each of the proteins investigated: (**A**) KR2, (**B**) MACR, (**C**) A_{2A}AR-BRIL-ΔC, (**D**) NarQ. The blue and red stripes represent outer and inner lipidic membrane surfaces respectively with the hydrophobic region of the lipidic membrane represented in yellow. Iodide ions are shown as orange spheres. Water molecules are shown as red spheres.

2.4. Discussion

The work described in (Melnikov *et al.*, 2017) aimed at testing the applicability of I-SAD and Br-SAD as methods of solving the phase problem in the determination of the crystals structures of MPs. The results obtained suggest that I-SAD is a potentially universal method for the *de novo* solution of MP crystal structures, combining speed, nontoxicity, and simplicity in derivative preparation with the advantageous anomalous scattering properties of iodide even when experiments are carried out relatively far from the L-absorption edges of iodide. This latter property is strengthened by the fact that for each of the systems investigated, several I^- sites were identified, meaning that anomalous signals remain high even if the anomalous scattering properties of I^- are not optimised during an experiment. One of the experiments described also suggests that the I-SAD phasing of MPs should be relatively straightforward using diffraction data collected in a serial fashion (i.e. (Zander *et al.*, 2015), see also §1.4.3) from many microcrystals mounted on a fixed target that can be moved through the X-ray beam. This is an important observation as it is often difficult to obtain MP crystals of sufficient size for standard data collection. It also suggests that similar serial data collection protocols might also be successful for MP crystal structure solution using diffraction data collected at X-ray free-electron lasers (Batyuk *et al.*, 2016) or at next-generation diffraction-limited storage ring synchrotron sources.

An interesting result of the investigations described in (Melnikov *et al.*, 2017) is that Br-SAD does not appear to be as routinely viable as I-SAD for the *de novo* solution of the crystal structures of MPs. Indeed, Br-SAD did not result successful structure for any of the four targets studied. For two of the targets investigated anomalous difference Fourier

maps contained peaks indicating that Br^- binds to these MPs. Moreover, many of the Br^- binding sites identified were the same as those observed for binding of I^- in corresponding I-SAD experiments. This, suggests that Br-SAD and/or Br-SIRAS phasing of the crystal structures of MPs should be feasible. A clear disadvantage, however, in Br-SAD experiments is that the maximum f'' that can be induced around the Br K-absorption edge is $\sim 4\ e^-$. This compares to $f'' \sim 9.2\ e^-$ for I-SAD experiments carried out, as in (Melnikov *et al.*, 2017), at λ = 1.85 Å. Because of this, anomalous signals in Br-SAD experiments will be appreciably lower than in their I-SAD equivalents. Successfully applying Br-SAD to the crystals structures of MPs will thus clearly require higher-accuracy data (i.e. higher multiplicity measurements probably based on the merging of more individual (partial) data sets) than when applying I-SAD. Br-SAD will therefore be less favourable to high-throughput MP structure determination.

3. Transmembrane signal transduction by bacterial sensor histidine kinases

3.1. Summary

In meso and single wavelength anomalous dispersion (SAD) approaches were used to crystallise and determine the crystal structures of a fragment of the nitrate sensor NarQ that contains the sensor, TM and HAMP domains. The symmetric *apo* state structure of a conserved mutant and the symmetric (*holo*-S) and highly asymmetric (*holo*-A) *holo* state structures of the wild type protein were obtained. Comparison of the *apo* and *holo* states reveals extensive structural rearrangements and provides a direct demonstration of signal transduction through the TM and HAMP domains, including a piston-like motion of TM helices and a lever-like rotation of HAMP domain protomers. Comparison of *holo*-S and *holo*-A structures shows that the signal transduction in NarQ can be achieved via different sets of conformational changes. Overall, the presented findings clarify the mechanism of TM signalling in TCS receptors and particularly the role of HAMP domain in it.

3.2. Introduction

3.2.1. Two-component systems are major bacterial means of adaptation to the environment

Sensing and reacting to the environment is essential for the survival of every living organism and in order to adapt to rapid environmental changes the sensory signals must be rapidly transformed into a regulatory response of a cell. There are three main signal transduction systems in bacteria (Mascher, 2014): one-component systems, two-component system (TCS) and the alternative σ-factors of extracytoplasmic function (σECF). One component systems are responsible mainly for intracellular perception and comprise sensor and regulatory as two separate domains of one molecule. TCS and σECF respond mainly to extracellular cues and, in contrast to minimalistic one-component systems, have a multi-modular organisation. The principle of σECF regulation (see Figure 40a) is usually based on signal-dependent binding of the σ-factor to the anti-σ domain of the transmembrane sensor. Because σ-factors initiate DNA transcription directly, signal-dependent capturing or releasing of them by the transmembrane sensor modulates their activity and thus directly regulates the response (Raivio & Silhavy, 2001; Heimann, 2002).

TCS represent the most prevalent bacterial signalling pathway for responding to environmental changes (Egger *et al.*, 1997). TCS comprise two proteins (see Figure 40b): a transmembrane sensor histidine kinase (HK) which receives the signal at its extracellular sensor module and an intracellular response regulator (RR) which acts as a DNA transcription factor. Through TCS bacteria react to various environmental factors such as

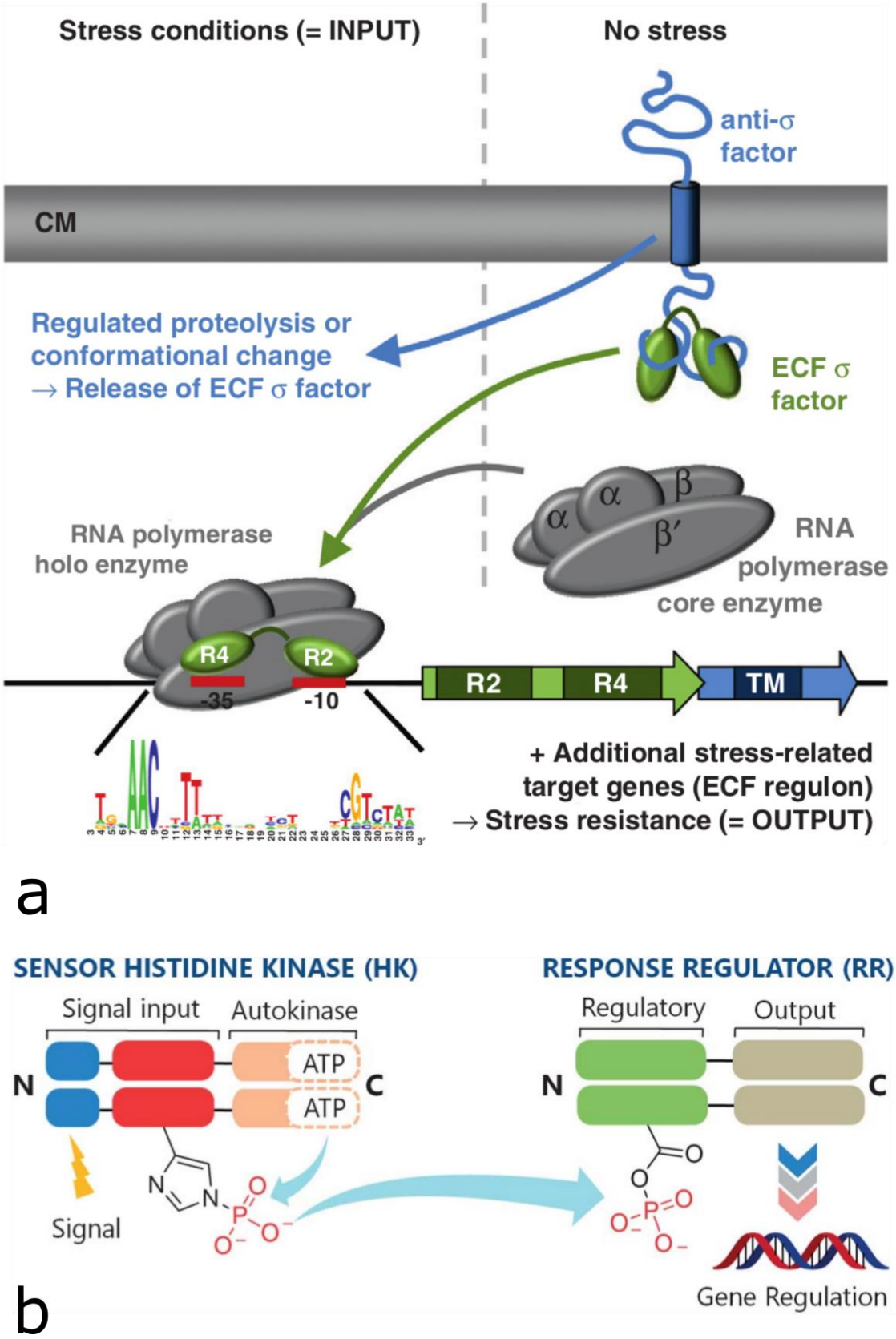


Figure 40. *Schemes showing different mechanisms of bacterial signalling by TCS and σECF.* ***a.*** *Overview of σECF-dependent signalling. The σECF factor is shown in green. The transmembrane sensor – the cognate anti-σ factor is shown in blue. Upon receiving the signal σECF factor is released and binds to RNA polymerase initiating transcription. Adapted from (Mascher, 2013).* ***b.*** *Mechanism of signalling by TCS. A signal is received by the histidine kinase sensor domain which causes ATP-dependent phosphorylation of the kinase domain histidine. This phosphate is then transferred to the cognate response regulator protein which in turn modulates its activity in transcription activation or repression. Adapted from (Francis et al., 2013).*

pH level, osmotic pressure or concentration of various nutrients, etc. (Forst *et al.*, 1989; Stewart, 2003; Watts *et al.*, 2004). Bacterial mechanisms for quorum sensing, virulence and resistance to the innate immune system attacks are also driven by TCS (Walker *et al.*, 2013; Mike *et al.*, 2014; Herrera *et al.*, 2014). Most bacterial species possess tens to hundreds of different TCS (Podgornaia & Laub, 2013) (in *E. coli,* the number of unique TCS identified is around 30 (Mizuno, 1997)) which share similar modes of action. Importantly, TCS are almost absent in eukaryotes (Parkinson & Kofoid, 1992) making them potential targets for anti-microbial treatments.

The principle of TCS response is in transmitting the signal from the sensor protein (HK) to the response regulator (RR) (Figure 40b). Having received a stimulus at the sensor, HKs undergo structural rearrangements which pass downstream to its auto-kinase domains (Figure 40b, Figure 41). Here, the conformational rearrangements lead to the favouring of ATP-binding and the transfer of its γ-phosphate to a conserved histidine residue in the kinase module. This γ-phosphate is then transferred to a conserved aspartate residue of the receiver domain of the RR. The free energy change induced by phosphorylation of this aspartate then trigger conformational changes in the whole RR molecule (Stock *et al.*, 2000). Thus, by altering RR conformations aspartate phosphorylation can activate or repress DNA transcription. Indeed, conformational changes in RRs can induce their binding or release to or from DNA or affecting directly (i.e. induce degradation) of the σ-subunit of RNA-polymerase (Zhou, 2001; Yamamoto *et al.*, 2005).

The domain structure of a typical HK is shown in Figure 41. Comparisons of different HKs show areas of structural similarity and difference. Here, the greatest diversity is observed in the structures of the sensor domains of different HKs (blue in Figure 41) while,

for example, the amino acid sequences and structures of HK catalytic domains located at the C-terminus (navy blue and yellow in Figure 41) are quite conserved (Song *et al.*, 2004; Xie *et al.*, 2010; Ferris *et al.*, 2014). RR receiver domains are also relatively structurally conserved (Volz & Matsumura, 1991; Gouet *et al.*, 1999; Solà *et al.*, 1999; Buckler *et al.*, 2002). This raises the question as to how the specificity of an individual HK to its cognate RR is maintained is such a structurally conserved environment and, major determinants of TCS specificity are likely to include molecular recognition (Skerker *et al.*, 2005), phosphatase activity and competition of substrates (Podgornaia & Laub, 2013).

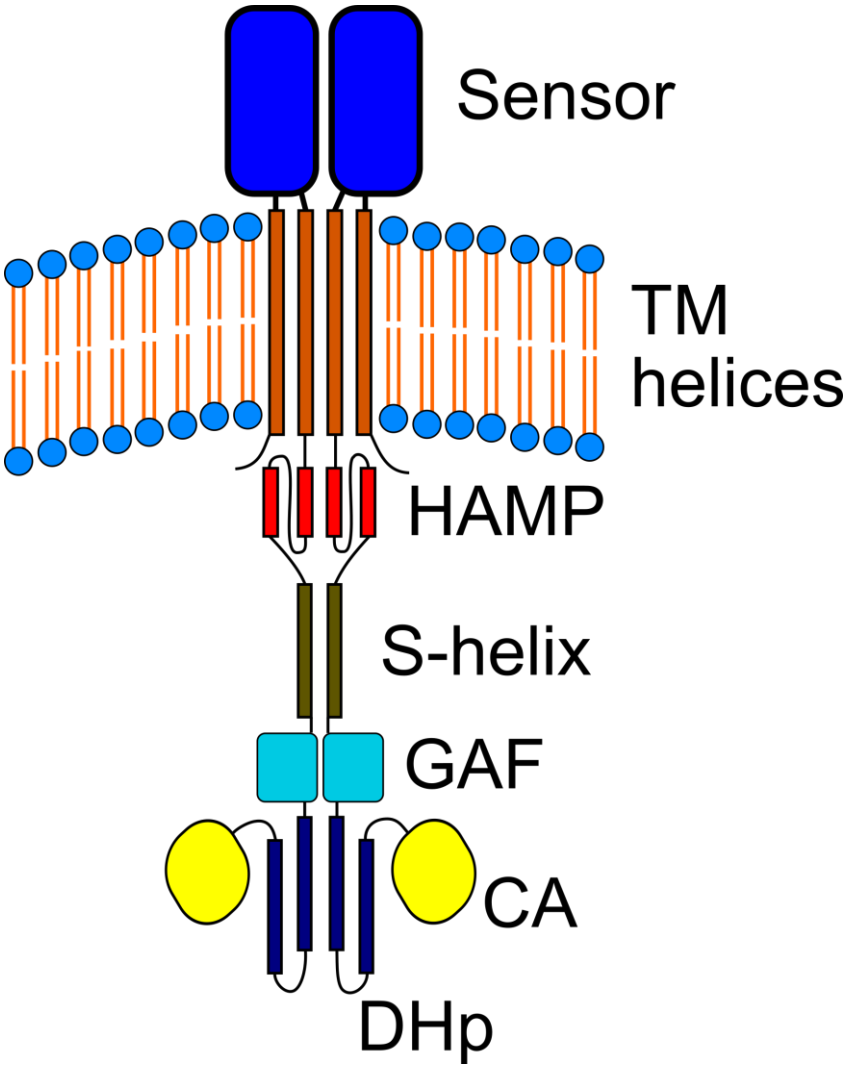


Figure 41. *The domain architecture of a typical bacterial TM sensor histidine kinase (HK). A HK dimer is shown inserted in the cytoplasmic membrane: cytosol at the bottom, periplasm is at the top. The sensor domains are shown in blue, the transmembrane helices in orange, the HAMP domains in red, signalling S-helices in pickle, GAF domains in teal and the dimerisation and histidyl phosphotransfer (DHp) and catalytic and ATP-binding domains in navy blue and yellow, respectively.*

3.2.2. Signal transduction by transmembrane sensor histidine kinases

HKs generally function as dimers (see Figure 41). However some pathways include inactivation of HK by homo- or hetero-oligomerisation (Krell *et al.*, 2010; Willett & Crosson, 2017). As can be seen from Figure 41, the extracellular signal captured is usually transmitted across the membrane by an anti-parallel bundle of four transmembrane helices, two from each HK monomer. Downstream of the TM helices HKs comprise various transmitter domain homodimers. Such domains include HAMP (present in **h**istidine kinases, **a**denylate cyclases, **m**ethyl-accepting chemotaxis proteins and **p**hosphatases (Aravind & Ponting, 1999)), PAS (first discovered in **p**eriodic circadian proteins, **a**ryl hydrocarbon nuclear translocator proteins, and **s**ingle-minded proteins, (Taylor & Zhulin, 1999)) and GAF (also found in c**G**MP-specific phosphodiesterases, **a**denylate cyclases, and **f**ormate hydrogenases, (Aravind & Ponting, 1997)) which are supplemented with coiled-coil elements such as signalling helices (S-helix, (Winkler *et al.*, 2012)) which further transmit the signal to, and regulate the activity of, the catalytic unit composed of **d**imerisation and **h**istidine-**p**hosphotransfer (DHp) and catalytic and ATP-binding (CA) domains (Bhate *et al.*, 2015).

Several studies have reported the successful construction of HK chimeras comprising sensor, transducer and catalytic domains from different HKs (Ward *et al.*, 2002; Zhu & Inouye, 2003; Ward *et al.*, 2006). These engineered chimeras possess the functionality to receive and transmit extracellular signals, clearly suggesting that the mechanisms of TM signal transduction are shared amongst HKs. However, exactly how a signal captured in the sensor domain is transmitted across the membrane to the

downstream domains of transmembrane HKs remains a central issue of investigation and debate.

The structure of the extracellular sensor domain varies between different HK proteins. These can be PAS domain-like structure with α/β-fold (Cheung & Hendrickson, 2008; Cheung *et al.*, 2008; Sevvana *et al.*, 2008; Affandi *et al.*, 2016) or fully α-helical (Cheung & Hendrickson, 2009; Moore & Hendrickson, 2012) (see Figure 42). The conformational signal transmitted by sensor domain to the TM region is proposed to be a helical displacement towards the membrane plane ((Cheung & Hendrickson, 2009), Figure 43a-left, Figure 45). However, this hypothesis is not proven because it is based on the isolated structures of HK sensor domains (Cheung & Hendrickson, 2009).

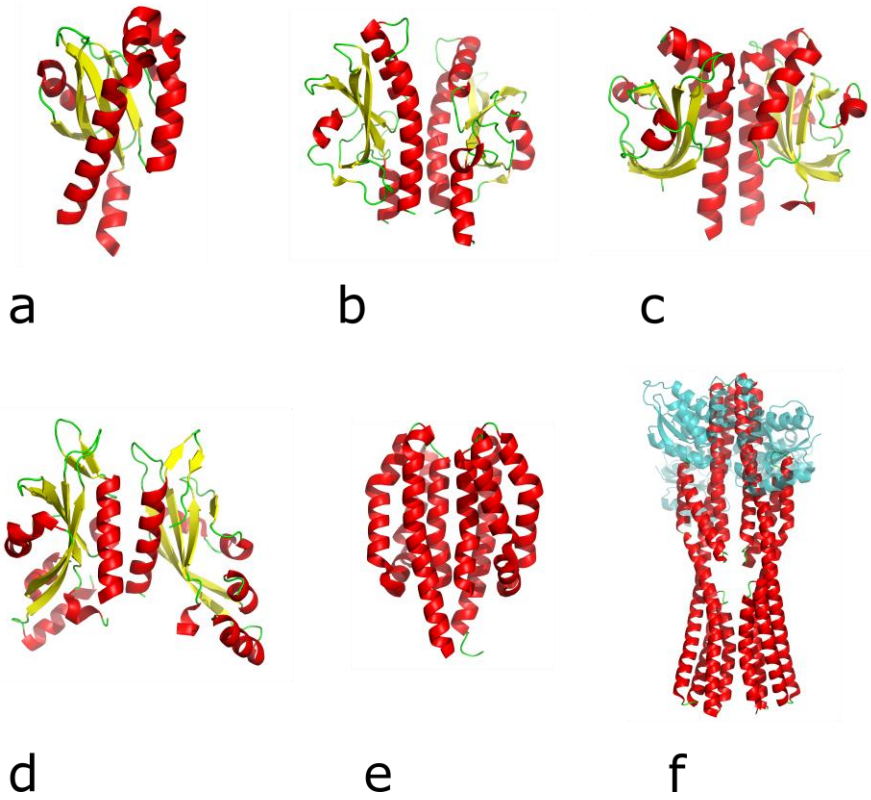


***Figure 42**. Cartoon representation of the structures of extracellular sensor domains of HKs. The crystal structures of **a.** E. coli DcuS sensor (of malate) domain monomer (PDB code: 3BY8); **b.** CusS sensor (of copper and silver) domain dimer in the ligand-bound state (PDB code: 5KU5); **c.** CitA sensor (of citrate) domain dimer in the ligand-bound state (PDB code: 2J80); **d.** E.coli PhoQ sensor (of bivalent cations) domain dimer (PDB code: 3BQ8); **e.** E. coli NarX sensor (of nitrate) domain dimer in the ligand-bound state (PDB code: 3EZH); **f.** TorS (in transparent teal) in complex with histidine kinase TorT sensor domain (in red) in the absence of ligand (TMAO; PDB code: 3O1I). Secondary structure is colour-coded with α-helices shown in red, β-sheets shown in yellow whereas the loops are shown in green.*

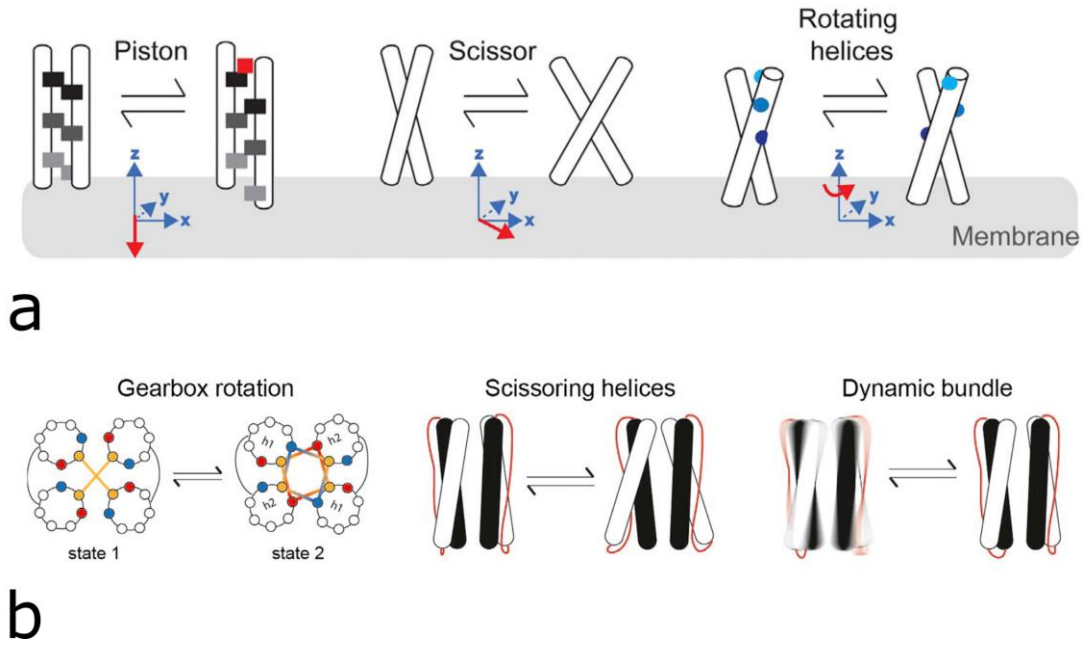


Figure 43. *Schematic representation of hypothetical mechanisms of signal transduction employed by sensor/TM domain helices (**a**) and by HAMP domain helices (**b**). The sensor domain helical endings are only shown in (a) which prolong to TM helices. Adapted from (Bhate et al., 2015).*

How the signal is transmitted across the membrane by HK TM helices is also far from clear as, until now, insight into this phenomenon has been provided either by low resolution structural studies or structural studies of isolated fragments (Pakula & Simon, 1992; Moukhametzianov *et al.*, 2006; Maslennikov *et al.*, 2010; Molnar *et al.*, 2014). Nevertheless, three hypotheses have been made (Figure 43a, (Bhate *et al.*, 2015)) proposing piston-like helix displacement (Cheung & Hendrickson, 2009; Moore & Hendrickson, 2012), scissoring helical movement (Molnar *et al.*, 2014) and helical rotation (Neiditch *et al.*, 2006; Wang *et al.*, 2014) as potential mechanism of signal transduction by TM helices.

The critically important part in signal propagation is taken by the cytoplasmic transmitter domains. Of the transmitter domains found in HKs, HAMP is present in one fifth of all histidine kinases (Hulko *et al.*, 2006). The structure of a HAMP domain dimer represents parallel four-helix coiled-coil (see Figure 44) connected by a flexible linker; this was predicted (Butler & Falke, 1998) and later proved (Hulko *et al.*, 2006) by multiple

reported structures of HAMPs from different proteins. Studies on mutated HAMP domains showed that these can seriously impede functioning of different sensor proteins. It is considered thus that HAMP is more an active element of signal transduction chain rather than a simple connector (Zhu & Inouye, 2004).

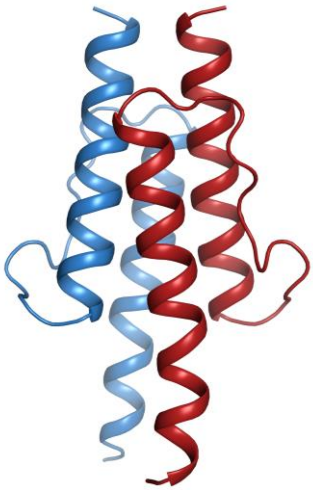


Figure 44. *Cartoon representation of the dimeric structure of the HAMP domain of hypothetic kinase Af1503 of the hyperthermal archae Archaeoglobus fulgidus, obtained using NMR. N-terminal end is up while C-terminal end of polypeptide chain is down in the figure. The four helices of the homodimer (the two monomers are shown in red and blue, respectively) constitute a parallel coiled-coil fold. The motile linker is also shown which connects the two helices of one monomer. Reconstructed based on PDB entry 2L7H (Hulko et al., 2006).*

No evidence has been provided for the mechanism of signal transduction by HAMP domain dimers. However, three main hypotheses have been made to try to match the available experimental facts (Figure 43b): the dynamic-bundle hypothesis propose that the downstream to HAMP domain activity is regulated via changing the stability of four-helix HAMP bundle from tight to loose packing or even to helix dissociation (Parkinson, 2010; Stewart, 2014); the gearbox hypothesis suggests that the helices of the HAMP are rotated coherently around the bundle axis like gears between the two states (Hulko *et al.*, 2006; Ferris *et al.*, 2011); the scissoring mechanism proposes that the HAMP helices undergo scissoring motion from the tightly-packed bundle thus creating the large output amplitude (Airola *et al.*, 2010; Gushchin *et al.*, 2013). However all proposed mechanisms (Figure 43)

of HAMP signalling are different, the real case might be also in their combination as they are not mutually exclusive.

3.2.3. NarX and NarQ – *E. coli* sensors of nitrates and nitrites

The *E. coli* TCSs which have been the most extensively studied are the paralogous systems NarX-NarL and NarQ-NarP which regulate anaerobic oxidative metabolism in bacteria. In the absence of oxygen, bacteria are able to use many other electron acceptors to perform anaerobic respiration. Among these acceptors are nitrates, nitrites, fumarates, sulfates, etc. The necessity of sensing the environmental concentrations of these acceptors is caused by the fact that energy gain is different for different anaerobic pathways and, bacteria use the most profitable oxidation that is available. In the presence of oxygen the production of all anaerobic reductase ferments is repressed. However, in the absence of oxygen, the production of reductases for the next most valuable acceptor is activated. Most often, the next best substrate for respiration is nitrate or nitrite.

Initially NarX-NarL was discovered and assumed to be the only pathway for the sensing of nitrates in *E. coli* (Stewart & Parales, 1988). However, after it had been shown that nitrate sensing was not disturbed in NarX-deletion strains a NarX homologue, NarQ and its cognate RR, NarP were discovered (Chiang *et al.*, 1992; Rabin & Stewart, 1992, 1993). It was also shown that while NarX-NarL is responsible only for nitrate sensing, NarQ-NarP is also sensitive to nitrites (Rabin & Stewart, 1993). As typical bifunctional HKs, NarX and NarQ can stimuli-dependently phosphorylate and dephosphorylate their cognate RRs. However studies on phosphorylation kinetics (Noriega *et al.*, 2010) have shown that NarQ can cross-phosphorylate NarL and, at a much reduced level, NarX can cross-phosphorylate

NarP. Such cross-phosphorylation might be essential for the fine tuning of bacterial machinery.

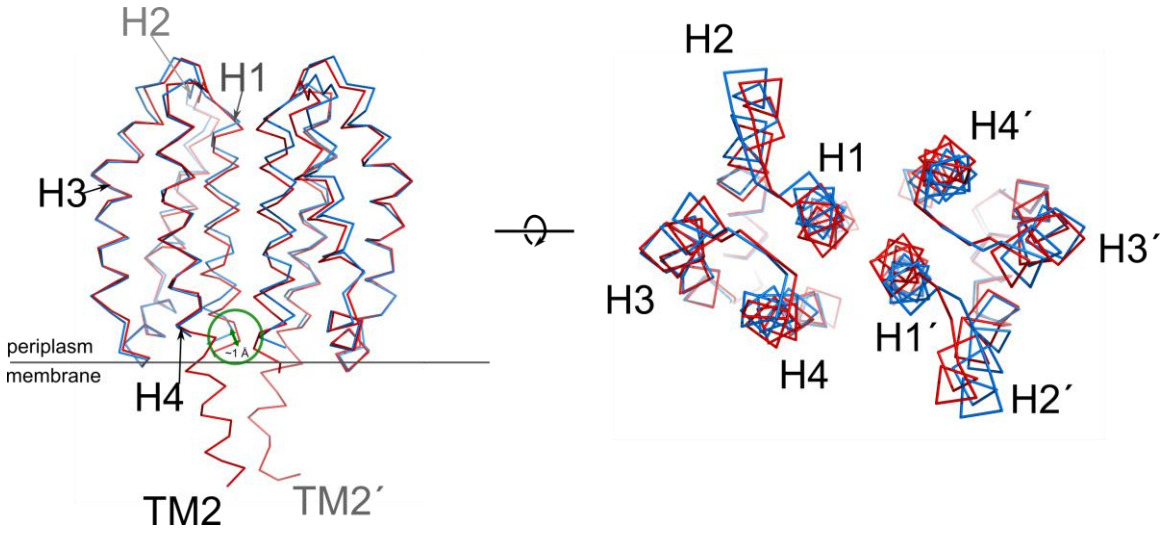


Figure 45. *The comparison of the apo (in blue, PDB code: 3EZI) and ligand-bound (holo, in red, PDB code: 3EZH) structures of NarX truncated sensor domain (Cheung & Hendrickson, 2009). The structures of each monomer of the apo-state molecules (in blue) are aligned by the residues 79-105 (membrane-anchoring part of helices H2 and H3) of the amino acid sequence to the corresponding monomers of the dimeric structure of the holo-state molecule (in red). The alignment shows the displacement of ~1 Å between the ends of the H1 helices of the sensor in two states (shown in green). The sensor domain helices H1, H2, H3, H4 and a part of transmembrane helix TM2 are shown. Left: side view; right: top view.*

NarX and NarQ architecture is similar as they are homologues and monomers of both comprise a periplasmic sensor domain, two TM helices, HAMP domain, then S-helix downstream, GAF domain, DHp and CA domains at the C-terminus (Figure 41). No particular structural information is available for NarQ. However, there are crystal structures available for the isolated sensor domain of NarX in the *apo* and nitrate-bound states (Cheung & Hendrickson, 2009). In that work it was determined that binding of the ligand induces piston-like displacement perpendicular to the membrane of the end of the sensor helix that in the full-length protein is prolonged by the second TM helix (Figure 45). However, since the crystal structures are of an isolated fragment, it is not clear whether this conformational change occurs in fuller length NarX constructs. There are also several crystal structures available of isolated HAMP, GAF, DHp and CA domains of other proteins (Bhate *et al.*, 2015; Zschiedrich *et al.*, 2016) As for the crystal structures of the isolated

sensor domain of NarX, these also provide little insight on how these domains cause the transmission of a signal when assembled in the whole protein structure. Structural analysis of constructs of HKs containing several domains would clearly help in this regard. Part of this thesis work therefore was dedicated to the elucidation of the crystal structure in both *apo*- and ligand-bound states, of a construct of NarQ, comprising periplasmic sensor domain, transmembrane helices and HAMP domain (Figure 41). A manuscript describing the results of this work was recently published and is, along with the relevant Supplementary Materials, reproduced here.

3.3. Mechanism of transmembrane signalling by sensor histidine kinases

Ivan Gushchin[12], Igor Melnikov[3], Vitaly Polovinkin[124], Andrii Ishchenko[15†], Anastasia Yuzhakova[12], Pavel Buslaev[2], Gleb Bourenkov[6], Sergei Grudinin[789], Ekaterina Round[14], Taras Balandin[1], Valentin Borshchevskiy[12], Dieter Willbold[1 10], Gordon Leonard[3], Georg Büldt[2], Alexander Popov[3], Valentin Gordeliy[124]

[1]Institute of Complex Systems (ICS), ICS-6: Structural Biochemistry, Research Centre Jülich, 52425 Jülich, Germany.

[2]Moscow Institute of Physics and Technology, 141700 Dolgoprudniy, Russia.

[3]European Synchrotron Radiation Facility, F-38043 Grenoble, France.

[4]Institut de Biologie Structurale J.-P. Ebel, Université Grenoble Alpes-CEA-CNRS, F-38000 Grenoble, France.

[5]Institute of Crystallography, University of Aachen (RWTH), 52056 Aachen, Germany.

[6]European Molecular Biology Laboratory, Hamburg Outstation, 22607 Hamburg, Germany.

[7]Université Grenoble Alpes, LJK, F-38000 Grenoble, France.

[8]CNRS, LJK, F-38000 Grenoble, France.

[9]Inria, F-38000 Grenoble, France.

[10]Institute of Physical Biology, Heinrich Heine University, 40225 Düsseldorf, Germany.

[†] Present address: Department of Chemistry, Bridge Institute, University of Southern California, Los Angeles, CA 90089, USA.

*Science **356** (2017): aah6345*

RESEARCH ARTICLE SUMMARY

STRUCTURAL BIOLOGY

Mechanism of transmembrane signaling by sensor histidine kinases

Ivan Gushchin,* Igor Melnikov, Vitaly Polovinkin, Andrii Ishchenko, Anastasia Yuzhakova, Pavel Buslaev, Gleb Bourenkov, Sergei Grudinin, Ekaterina Round, Taras Balandin, Valentin Borshchevskiy, Dieter Willbold, Gordon Leonard, Georg Büldt, Alexander Popov, Valentin Gordeliy*

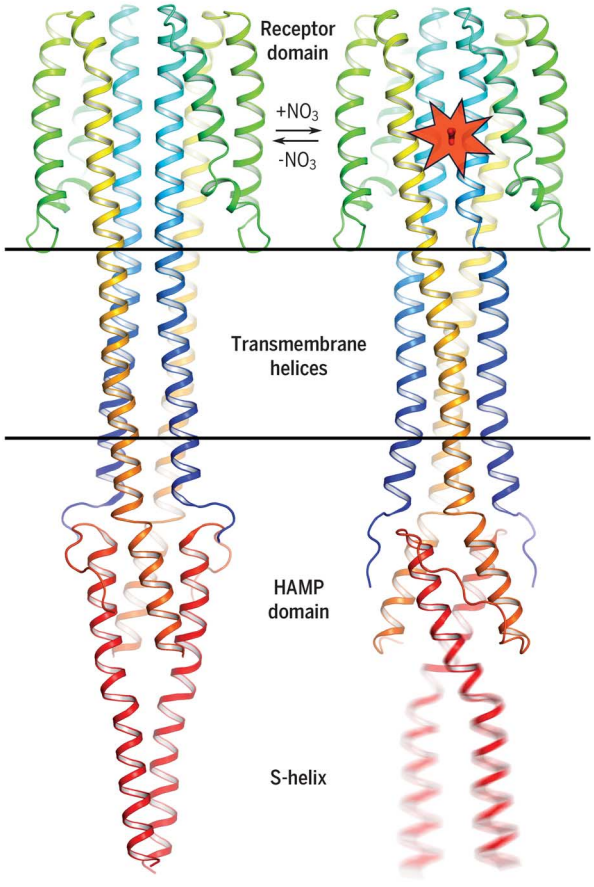
INTRODUCTION: Microorganisms obtain most of the information about their environments through membrane-associated signaling systems. One of the most abundant classes of membrane receptors, present in all domains of life, is sensor histidine kinases, members of two-component signaling systems (TCSs). Tens of thousands of TCSs are known. Many of these systems are essential for cell growth, survival, or pathogenicity and consequently can be targeted to reduce virulence. Several large families of transmembrane (TM) TCS receptors are known: (i) sensor kinases, which generally possess a periplasmic, membrane, or intracellular sensor module; a transmembrane domain; often one or more intracellular signal transduction domains such as HAMP, PAS, or GAF; and an intracellular autokinase module (DHp and CA domains), which phosphorylates the response regulator protein; (ii) chemoreceptors, which also possess the sensor module and the TM domain but lack the kinase domain and control a separate kinase protein (CheA) via a kinase control module; and (iii) phototaxis systems, which are similar to chemotaxis systems except that the sensor module—a light receptor sensory rhodopsin—is a separate protein.

RATIONALE: Despite the wealth of biochemical data, the structural mechanisms of transmembrane signaling by TCS sensors are poorly understood at the atomic level. In particular, high-resolution structures of the TM segments connected to the adjacent domains are lacking. Deciphering of the signaling-associated conformational changes would shed light on the details of long-range transmembrane signal transduction and might help in the development of novel classes of antimicrobials targeting TCSs.

RESULTS: We used the in meso crystallization approach and single-wavelength anomalous dispersion to determine the crystal structures, at resolutions of up to 1.9 Å, of a fragment of *Escherichia coli* nitrate/nitrite sensor histidine kinase NarQ that contains the sensor, TM, and HAMP domains in a symmetric ligand-free apo state and in symmetric and asymmetric ligand-bound holo-S and holo-A states. In all of the structures, the TM domain is an antiparallel four-stranded coiled coil (CC) consisting of nine CC layers. The sensor domain is connected to the TM domain through continuous α-helical linkers that are partially disrupted in the holo state. The intracellular HAMP domain is connected to the TM helices via flexible proline junctions and robust hydrogen bonds conserved in all signaling states. The structures reveal the mechanism of transmembrane signal transduction in NarQ and show that binding of ligand induces displacement of the sensor domain helices by ~0.5 to 1 Å. This displacement translates into rearrangements and ~2.5 Å pistonlike shifts of transmembrane helices and is later converted, via leverlike motions of the HAMP domain protomers, into 7 Å shifts of the output helices and changes of the CC helical phase. The structures also demonstrate that the signaling-associated conformational changes in the TM domain do not need to be symmetric.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aah6345



The structures of histidine kinase NarQ in ligand-free and ligand-bound forms. The structures reveal rearrangement of transmembrane α helices during signal transduction and show that pistonlike shifts of the transmembrane helices result in leverlike motions of the HAMP domain protomers.

CONCLUSION: The determined structures of the transmembrane and membrane-proximal domains of the nitrate/nitrite receptor NarQ in ligand-free and ligand-bound forms present a template for studies of other TCS receptors, establish the importance of the pistonlike displacements of the TM helices for TM signal transduction, and highlight the role of the HAMP domain as an amplifier and converter of a pistonlike displacement into helical rotation. Overall, the results show how a mechanistic signal is generated and amplified while being transduced through the protein over distances of 100 Å or more. Because membrane-associated TCSs are ubiquitous in microorganisms and are central for bacterial sensing, we believe that our results will help to elucidate a broad range of cellular processes such as basic metabolism, sporulation, quorum sensing, and virulence. They may also provide insights useful for the development of novel antimicrobial treatments targeting TCSs. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ivan.gushchin@phystech.edu (I.G.); valentin.gordeliy@ibs.fr (V.G.)
Cite this article as I. Gushchin *et al.*, *Science* **356**, eaah6345 (2017). DOI: 10.1126/science.aah6345

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

Mechanism of transmembrane signaling by sensor histidine kinases

Ivan Gushchin,[1,2]* Igor Melnikov,[3] Vitaly Polovinkin,[1,2,4] Andrii Ishchenko,[1,5]† Anastasia Yuzhakova,[1,2] Pavel Buslaev,[2] Gleb Bourenkov,[6] Sergei Grudinin,[7,8,9] Ekaterina Round,[1,4] Taras Balandin,[1] Valentin Borshchevskiy,[1,2] Dieter Willbold,[1,10] Gordon Leonard,[3] Georg Büldt,[2] Alexander Popov,[3] Valentin Gordeliy[1,2,4]*

One of the major and essential classes of transmembrane (TM) receptors, present in all domains of life, is sensor histidine kinases, parts of two-component signaling systems (TCSs). The structural mechanisms of TM signaling by these sensors are poorly understood. We present crystal structures of the periplasmic sensor domain, the TM domain, and the cytoplasmic HAMP domain of the *Escherichia coli* nitrate/nitrite sensor histidine kinase NarQ in the ligand-bound and mutated ligand-free states. The structures reveal that the ligand binding induces rearrangements and pistonlike shifts of TM helices. The HAMP domain protomers undergo leverlike motions and convert these pistonlike motions into helical rotations. Our findings provide the structural framework for complete understanding of TM TCS signaling and for development of antimicrobial treatments targeting TCSs.

Microorganisms obtain most of the information about their environments through membrane-associated signaling systems. One of the most abundant classes of membrane receptors, present in all domains of life, is sensor histidine kinases, members of two-component signaling systems (TCSs) (*1*–*3*). Tens of thousands of TCSs are known (*4*). Many of these systems are essential for cell growth, survival, or pathogenicity and consequently can be targeted to reduce virulence (*5*–*7*).

Membrane-associated sensor kinases generally function as homodimers and possess a periplasmic, membrane, or intracellular sensor module; a transmembrane (TM) domain; often one or more intracellular signal transduction domains (such as HAMP, PAS, or GAF); and an intracellular autokinase module (DHp and CA domains), which phosphorylates the response regulator protein (*1*, *8*, *9*) (fig. S1). Other TCS TM sensors, chemo- and photoreceptors, have domain organization similar to that of sensor histidine kinases; they possess the sensor module, TM, and HAMP domains but lack the autokinase module and control a separate kinase protein (CheA) via the kinase control module (fig. S1). Also, the core functional unit of chemoreceptors and phototaxis systems is not a dimer but a trimer of dimers (*10*–*12*), which may then pack into higher-order oligomeric assemblies (*13*).

The molecular mechanisms of TM signal transduction in different TCS classes are expected to be somewhat similar, because a vast majority of the sensors feature a dimeric four-helix TM core in the TM domain, often followed by the HAMP domain downstream. At present, only two reports on atomic-resolution structures of the TM domains are available, both describing isolated TM fragments and thus providing little insight into the signal transduction in and out of the TM domain of a typical TCS sensor. One report (*14*) provided a crystallographic structure of the complex between the photoreceptor sensory rhodopsin and the TM domain of its transducer protein. The other report (*15*) described nuclear magnetic resonance (NMR) structures of nonphysiological monomeric TM domains of histidine kinases from three different classes. Additionally, several low-resolution models of the TM helices in the context of full-length proteins such as chemoreceptor Tar or histidine kinase PhoQ have been constructed on the basis of disulfide cross-linking and modeling (*16*, *17*). Because of the limitations of all these studies, atomic details of TM signal transduction are still not resolved (*8*, *9*). As for the HAMP domain, multiple experimental structures of the domains from different proteins are available, and there are several competing hypotheses of HAMP domain signaling (*8*, *9*, *18*).

To complement the above reports, we present our studies of the *E. coli* nitrate/nitrite sensor kinase NarQ. Similarly to NarX, another *E. coli* nitrate/nitrite sensor, NarQ consists of seven major structural elements: the periplasmic sensor domain, the TM domain, the HAMP domain, the signaling helix, the GAF-like domain, the DHp domain, and the kinase CA domain (*18*–*20*) (fig. S1). Whereas no structural information is available for NarQ, crystal structures of NarX sensor domain in apo and holo states have been determined (*21*), as well as multiple structures of the HAMP, GAF, DHp, and kinase domains of other proteins (*8*, *9*).

Here, we used the in meso (*22*–*24*) and single-wavelength anomalous dispersion (SAD) (*25*) approaches to crystallize and determine the crystal structures, at resolutions of up to 1.9 Å, of a fragment of the nitrate sensor NarQ that contains the sensor, TM, and HAMP domains. We determined the symmetric apo (ligand-free) state structure of a conservative mutant and symmetric and highly asymmetric holo-S and holo-A (ligand-bound) state structures of the wild-type protein (Fig. 1 and fig. S2). Comparison of the apo and holo structures reveals extensive structural rearrangements and provides a direct demonstration of signal transduction through the TM and HAMP domains, including a pistonlike motion of TM helices and a leverlike rotation of HAMP domain protomers. Comparison of holo-S and holo-A structures shows that the signal transduction in NarQ can be achieved via different sets of conformational changes. Overall, our findings clarify the mechanism of TM signaling in TCS receptors.

Structure of the sensor domain

In all of the presented structures, including the asymmetric ligand-bound form, the NarQ sensor domain is a symmetric dimer of two monomers comprising four α helices, H1 to H4, with helices H2 and H4 broken into subhelices (fig. S3). The loop between helices H2 and H3 is anchored into the membrane by residues Trp^{89} and Tyr^{90}. Nitrate binds at the symmetry axis between helices H1 and H1′ of the two protomers, where it is coordinated by the side chains of the residues Arg^{50} and $Arg^{50'}$ (fig. S3E). The ligand was observed in all crystals of the wild-type protein, including those obtained without supplementation of the protein purification and crystallization solutions with nitrate. This might be a consequence of several factors: (i) the high ligand affinity of the crystallized fragment for nitrate, (ii) the presence of nitrate impurities in the aforementioned solutions, and (iii) the very small amount of nitrate needed to saturate the binding sites in the ~0.5 mM receptor solution used for crystallization. Consequently, we generated a conservative Arg^{50} → Lys (R50K) mutant to obtain the crystal structure of NarQ in its apo state. This mutation is known to lock the receptor in the off state (*26*) but was not expected to alter the local structure of the sensor domain, because Arg and Lys side chains are similar in size and charge. Indeed, the structures of the NarQ sensor domain in the apo R50K and holo-S and holo-A wild-type forms reported here closely resemble those of the apo and holo wild-type NarX sensor

[1]Institute of Complex Systems (ICS), ICS-6: Structural Biochemistry, Research Centre Jülich, 52425 Jülich, Germany. [2]Moscow Institute of Physics and Technology, 141700 Dolgoprudniy, Russia. [3]European Synchrotron Radiation Facility, F-38043 Grenoble, France. [4]Univ. Grenoble Alpes, CEA, CNRS, IBS, F-38000 Grenoble, France. [5]Institute of Crystallography, University of Aachen (RWTH), 52056 Aachen, Germany. [6]European Molecular Biology Laboratory, Hamburg Outstation, 22607 Hamburg, Germany. [7]Université Grenoble Alpes, LJK, F-38000 Grenoble, France. [8]CNRS, LJK, F-38000 Grenoble, France. [9]Inria, F-38000 Grenoble, France. [10]Institute of Physical Biology, Heinrich Heine University, 40225 Düsseldorf, Germany.
*Corresponding author. Email: ivan.gushchin@phystech.edu (I.G.); valentin.gordeliy@ibs.fr (V.G.) †Present address: Department of Chemistry, Bridge Institute, University of Southern California, Los Angeles, CA 90089, USA.

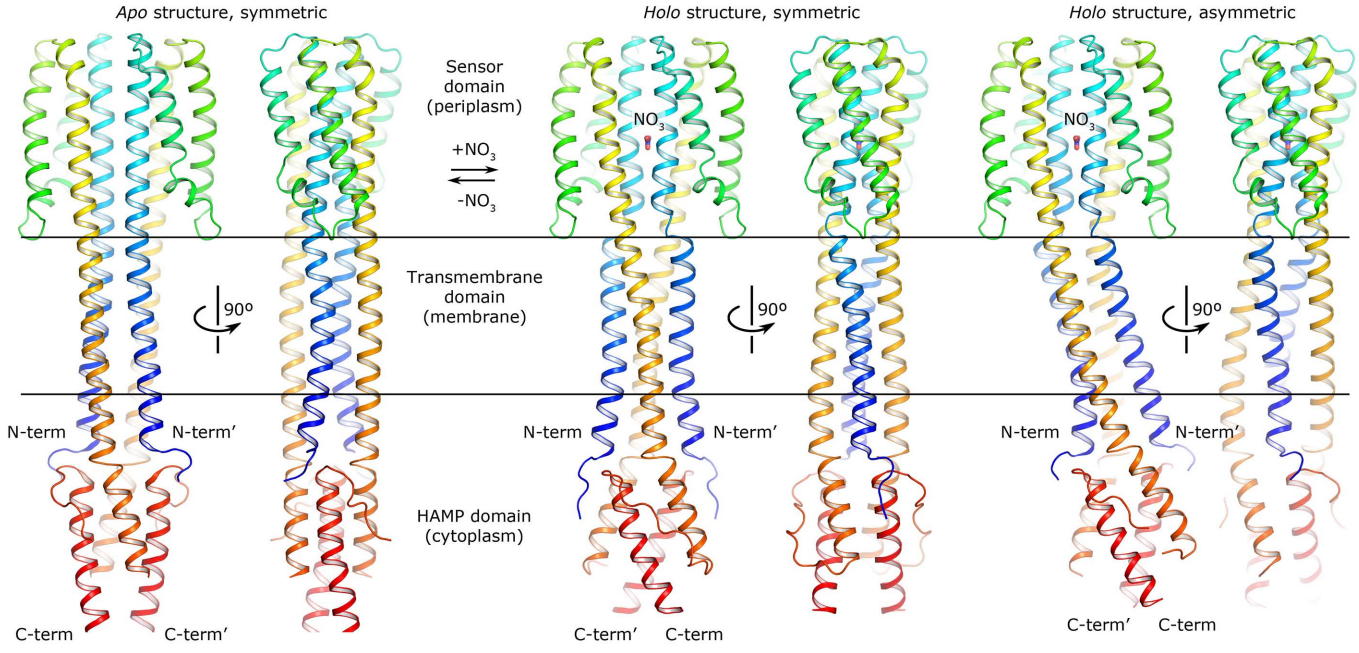


Fig. 1. Overall structure of the crystallized NarQ fragment in apo and holo states. The protein is observed as a symmetric dimer in both the apo and holo-S states in the space groups $I2_12_12_1$ and $F222$, respectively, and as an asymmetric dimer in holo-A state in the space group $P2$. Termini of the second protomer are denoted with primes. The crystallized fragment comprises the transmembrane domain (helices TM1 and TM2), the periplasmic sensor domain (helices H1 to H4) with the nitrate ion binding at the dimerization interface, and the intracellular HAMP domain (helices AS1 and AS2). The structures are aligned by sensor domains. Hydrophobic membrane boundaries (black) were calculated using the OPM server (*69*) with the symmetric ligand-bound structure as a template.

domain (*21*), respectively. In particular, the positions of the membrane-proximal ends of helices H1 and H4, responsible for downstream signaling, are identical in the corresponding structures (fig. S3, C and F), and we thus conclude that introduction of the mutation results in correct reproduction of the structural features of the wild-type ligand-free protein.

Structure of the TM domain

In all of the structures presented here, the NarQ TM domain is an antiparallel four-stranded coiled coil (CC) consisting of nine CC layers (Figs. 1 and 2 and fig. S4). Several ordered lipid fragments are observed in the holo-S structure, one fragment is observed in the holo-A structure, and no fragments are observed in the apo structure. However, reliable identification of the corresponding lipidic moieties is not possible because of the small size of the fragments.

There are notable variations in the packing of the TM helices. The apo state structure is best described as a dimeric CC comprising helices TM1 and TM1′, with TM2 and TM2′ flanking this. The symmetric holo-S state structure is a traditional four-helix CC, switching from a 7/2-period left-handed CC at the periplasmic side to an 11/3-period structure with close to zero supercoil twist on the cytoplasmic side. Finally, in the asymmetric holo-A state, the TM domain is a highly distorted CC, where the relative arrangement of the helices in layers 1 to 4 and 9 resembles that observed in the symmetric holo-S state and the arrangement in layers 7 and 8 resembles that seen in the apo state (fig. S4).

In general, the cytoplasm-proximal layers are more hydrophobic and are packed tightly, whereas those proximal to the periplasm contain hydrated cavities and multiple polar (Ser^{22}, Ser^{25}, Thr^{26}, Thr^{32}, Ser^{154}, and Thr^{163}) and somewhat disordered glycine residues (Gly^{27}, Gly^{157}, Gly^{158}, and Gly^{160}; fig. S2E). Although the role of water molecules and glycine residues in the NarQ TM domain is not clear, similar α helix–destabilizing features are observed in the TM domains of other histidine kinases (fig. S5), and they probably impart to the TM domain the flexibility needed to accommodate signaling-associated transitions, similarly to the glycine hinge in the cytoplasmic part of chemoreceptors (*27*).

The α-helical composition of the TM region of TCS receptors has long been predicted (*8*, *9*, *14*–*17*). However, the structure of the NarQ TM domain is notably different from that of the sensory rhodopsin transducer TM domain (fig. S5). In NarQ, the TM region is either a simple four-helix CC or has a dimeric CC core consisting of the TM1 helices, whereas *Np*HtrII has a dimeric CC core consisting of the TM2 helices (*14*, *28*). Also, in NarQ, there is either zero or left-handed supercoil twist, whereas in *Np*HtrII this changes from close to zero at the periplasmic side to right-handed at the cytoplasmic side (fig. S5). As for the comparison with the TM fragments of ArcB, QseC, and KdpD, the NMR structures unfortunately do not provide information about the dimerization interfaces and CC packing in the physiological dimers of any of these kinases (*15*).

Junctions between TM and adjacent domains

Our crystal structures of NarQ reveal details of the interactions of the TM domain with the membrane-proximal domains (fig. S6). On the periplasmic side of the protein, in the apo state, there is a continuous transition of the TM1 and TM2 helices into the H1 and H4 helices, respectively, whereas in both holo-A and holo-S states there is a break in the α-helical structure between TM1 and H1 (see fig. S6, A and D, for comparison of apo and holo-S states).

At the cytoplasmic side, in all states the cytoplasmic side of TM1 and the protein's N terminus make extensive contacts with the HAMP domain. The highly conserved residue Glu^{207} of the membrane-proximal region of HAMP domain helix AS2 (*29*) forms two hydrogen bonds with the backbone nitrogens of helix TM1 residues (fig. S6, B and E), similarly to what has been observed in the structure of the soluble mutant of protein Af1503 (*30*) (fig. S7A). As for the TM2-AS1 junction, there is a proline-induced kink (fig. S6, C and F), again as also seen for Af1503 (*30*) (fig. S7B).

The HAMP domain

The NarQ HAMP domain displays a typical parallel four-stranded CC fold (*18*), with each protomer consisting of two α helices, AS1 and AS2, separated by an unstructured linker (fig. S8).

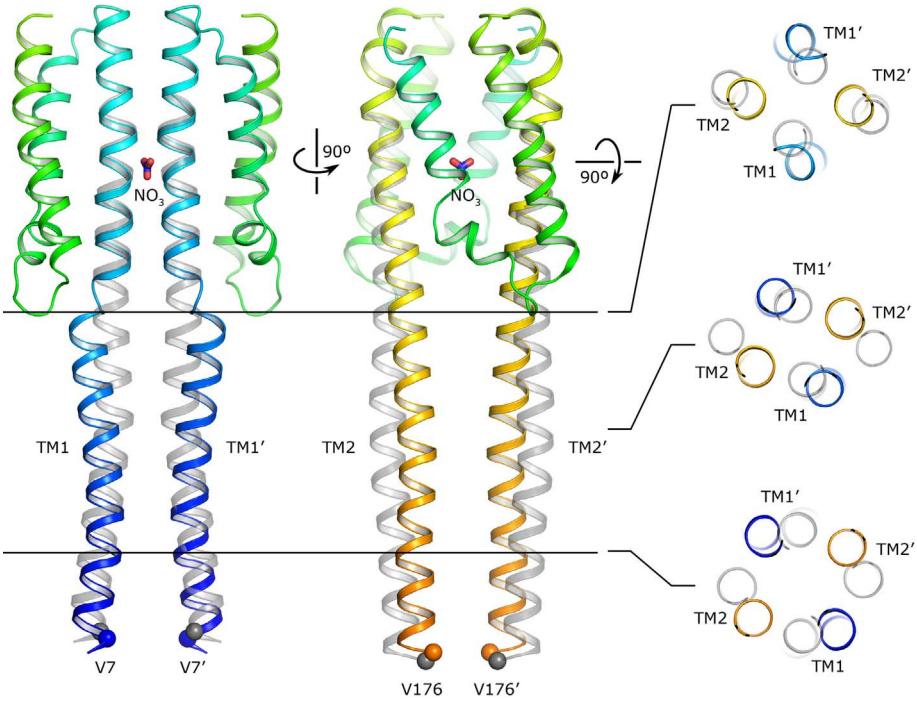


Fig. 2. Signaling-associated conformational changes in the NarQ TM domain. Superposition of the apo state structure (gray) and symmetric holo state structure (colored) is shown. Left: Changes in conformation of helix TM1. Center: Changes in conformation of helix TM2. Right: Changes in the arrangement of the TM helices. C_α atom positions of the residues from CC layer 9 are marked with the spheres to highlight the displacements in the direction perpendicular to the membrane. The structures are aligned by the sensor domains.

In all of the structures, the HAMP domain is less ordered than the sensor and TM domains, and in the apo state structure, the positions of the side chains of Leu^{225} and Tyr^{226} are not resolved. Overall, the apo state structure is similar to the structure of the Af1503 HAMP domain (fig. S9). The structure observed for the symmetric holo-S state HAMP domain is influenced by crystal contacts: The AS2 helices are interspersed with the same helices of an adjacent dimer (figs. S2 and S8), and consequently there is no contact between the Leu^{225} side chains facing toward the CC core. Although such conformation would be consistent with an antiparallel four-helix CC downstream domain, such as in bacterial chemoreceptors (*31*), it is unusual in the structural context of histidine kinases. The crystal contacts mentioned above are absent in the asymmetric holo-A state structure (space group *P*2). There, the hydrophobic Leu^{225} side chains are less exposed to the solvent and are in contact with each other. Thus, we conclude that the holo-A structure represents a better model of the activated NarQ HAMP domain.

The signaling helix region

To obtain insights into signal transduction through and downstream of the HAMP domain, we constructed atomic models of the adjacent signaling helix region (residues 224 to 245; see NarQ domain architecture in fig. S1). The signaling helix is a common sequence motif in sensor kinases (*32*) and sensory rhodopsin transducers (*33*), and the positions of hydrophobic and hydrophilic residues indicate that two CC heptad repeat pattern assignments are possible for NarQ (fig. S10). Neither of the models is compatible with the holo state HAMP domain structures, but both—especially model 2—are compatible with the structure of the apo state HAMP domain (see layer 6 in fig. S8 and layer 1 in fig. S10). Model 1 comprises a phase stutter that is also observed in other histidine kinases and chemoreceptors (*18*, *19*, *34*) and may be required for destabilization of the signaling helix for it to be able to adopt different signaling states. However, model 2 appears to be more probable, judging from the protein sequence (*19*).

Discussion

Signal transduction in NarQ

Comparison of the apo and holo state crystal structures of NarQ reveals that binding of nitrate causes extensive rearrangements in the TM region that can be represented as a combination of changes in the lateral arrangement of the TM helices and pistonlike shifts of the helices in the direction perpendicular to the membrane plane. At the same time, comparison of the holo-S and holo-A states reveals profound differences in the middle of the TM segment but identical conformations at its periplasmic and intracellular sides (see movie S1 for comparison of apo and holo-S state structures, and movie S2 for comparison of holo-S and holo-A structures). We make several observations from these comparisons.

First, binding of nitrate in the vicinity of Gly^{47} causes disruption in the α helix H1 and appearance of a 3_{10}-helix element: Gly^{47} carbonyl switches from Met^{51} amide hydrogen to Arg^{50} amide hydrogen (from i → i + 4 helical hydrogen bonding to i → i + 3; figs. S3 and S11). 3_{10} helices have a higher helical rise per residue than α helices (*35*), and as a result the membrane-proximal part of H1 moves slightly toward the membrane and rotates. This rotation of the membrane-proximal part of H1 opens a "hole" near residues Ile^{43} and Ala^{46}, which is then occupied by a Val^{136} "knob" (*36*). Consequently, H4 rotates as well and moves away from the membrane, which results in a pistonlike displacement of H4 relative to H1 for ~0.5 to 1.0 Å. Also, the H4 and H4′ helices are brought much closer to the H1 and H1′ helices. This interpretation is supported by mutagenesis data: Mutation of Ala^{46} to Ser is inconsequential, mutation to Thr results in impaired signaling, and mutations to the much bulkier Asn and Ile lock NarQ in the "off" state (*37*). The observed conformational changes in NarQ are identical to the changes observed in isolated NarX sensor domain (fig. S3) (*21*).

The rotation of the membrane-proximal parts of H1 and H1′ and the approach of H4 and H4′ upon binding of nitrate result in disruption of the TM1/H1-TM1′/H1′ CC interface, appearance of discontinuity between α helices TM1 and H1, and distancing of the TM1 and TM1′ periplasmic sides (Fig. 3, fig. S11, and movie S1). The introduced discontinuity between TM1 and H1 further increases the relative pistonlike displacement from ~1 Å for H1 relative to H4 [as in NarX (*21*)] to ~2.5 Å for TM1 relative to TM2 (fig. S11). The importance of the TM1-H1 helical break for kinase activation is also highlighted by a mutagenesis study in which the introduction of a proline into the TM1-H1 junction (mutation E41P) locks the receptor in the "on" state, whereas other nonconservative mutations of Glu^{41} to Arg, His, and Leu (E41R, E41H, and E41L) do not disrupt the function of the receptor (*37*).

Conformational changes in the sensor domain and at the sensor-TM junction are moderate and identical in holo-A and holo-S structures; however, the rearrangements in the TM domain are much more pronounced and are different between holo-A and holo-S. The TM1 helices, which form a tight CC interface in the apo state, are not fully separated in the holo-A structure (fig. S4). In the holo-S state, the four TM helices form a traditional four-helix CC (Fig. 2 and fig. S4). But there is a similarity between the holo-A and holo-S structures: In both, the ~2.5 Å relative pistonlike displacement of the TM helices is fully transduced toward the HAMP domain (Figs. 2 and 3 and movies S1 and S2). Conformational changes at the TM-HAMP interface are also identical in the holo-S and holo-A structures

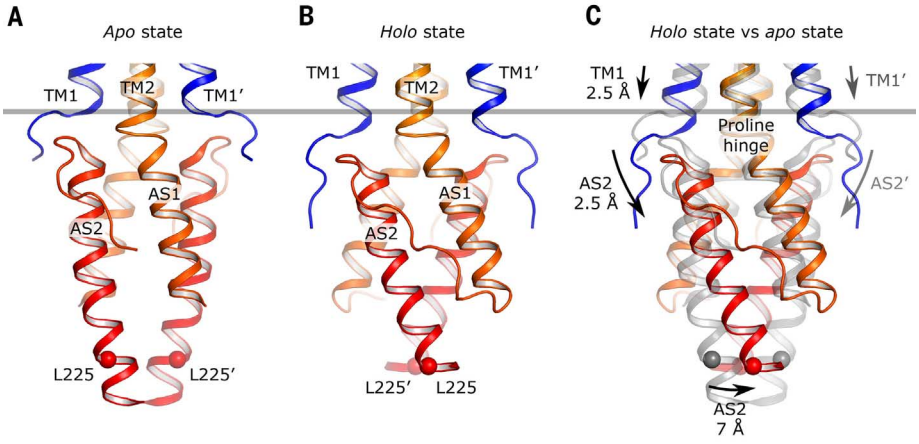


Fig. 3. Details of the signal transduction from the TM domain to and through the HAMP domain. (**A**) Inactive apo state. (**B**) Active holo-S state. (**C**) Superposition of the apo (gray) and holo-S (colored) states. A pistonlike displacement of the cytoplasmic end of TM1 relative to TM2 and the TM2-AS1 hinge is transmitted to the membrane-proximal end of AS2 and results in leverlike rotations of the HAMP domain protomers around the hinges. Because the HAMP domain protomers move in opposite directions, the positions of membrane-distal ends of the AS2 helices also change relative to each other. Positions of the Leu^{225} C_α atom are marked by spheres. The gray bar shows the position of TM1 ends in the apo state structure. The domains are aligned by residues 175 to 177.

as compared to the apo state structure. Thus, comparison of the asymmetric holo-A and symmetric holo-S state structures offers insight into the receptor flexibility and robustness of signal transduction by the TM domain despite the ~9 Å difference in the position of the HAMP domain (Figs. 1 and 3 and figs. S4 and S8).

Finally, in the cytoplasmic part of NarQ, the pistonlike displacements of the TM helices, caused by binding of the ligand, are amplified through the leverlike rotations of the HAMP domain protomers around the Pro^{179} hinge into roughly 7 Å displacements of the output ends of the AS2 helices (Fig. 3). During these conformational rearrangements, each HAMP domain protomer behaves essentially as a rigid body. The RMSD values of the backbone atom positions of helices AS1 (residues 179 to 192) and AS2 (residues 206 to 225) are ~0.6 Å between the apo and both holo-S and holo-A structures, and ~0.4 Å between the holo-S and holo-A structures; these values are comparable to crystal-to-crystal variation and typical crystallographic error of atomic positions. However, as a result of the relative motions of the HAMP domain protomers, the whole dimeric HAMP domain itself changes profoundly: The RMSDs of the backbone atom positions of helices AS1 and AS2 are ~2.9 Å between the apo and holo-S structures and ~3.0 Å between the apo and holo-A structures, as opposed to ~1.2 Å between the holo-S and holo-A structures.

Details of the signal transduction in NarQ downstream of the HAMP domain are less clear because of truncation of the crystallized construct. In the apo state, the CC register of the AS2 output ends is consistent with the CC register of the signaling helix. However, in the holo-A and holo-S states the CC helical phase is much less consistent with that of the signaling helix. Therefore, it appears that the signaling helix should be destabilized and/or dissociated upon binding of the ligand (Fig. 4). Signaling-related dissociation or destabilization of signaling helix was also observed in Af1503-EnvZ chimeras (*34*) and in the histidine kinases BvgS and DesK (*38*, *39*) and may thus be the general mechanism for signal propagation toward the DHp domain in TCS sensor kinases.

Mechanism of transmembrane signaling by TCS receptors

Several hypotheses concerning TM signaling mechanisms in TCS receptors have been proposed (*8*, *9*). For sensor histidine kinases, the crystal structures of α-helical NarX and TorS sensor domains hinted at pistonlike conformational changes (*21*, *40*). The crystal structures of the mixed α/β tandem PAS sensor domain of HAMP-less kinase LuxQ revealed rotation of signal output helices (*41*); this was also demonstrated in another HAMP-less kinase, AgrC, where the sensor domain is integral to the membrane (*42*). A recent cross-linking study revealed scissorlike diagonal displacement of the TM helices in the kinase PhoQ (*17*). For chemoreceptors with α-helical sensor domains, such as aspartate receptor Tar, the major model is pistonlike displacement of the TM helices in the direction normal to the membrane plane (*13*, *31*). For chemoreceptor McpB with mixed α/β tandem PAS sensor domain, the signal was found to be transduced via rotation of TM helices (*43*). Finally, in the sensory rhodopsin–cognate transducer complex, where the sensor is integral to the membrane, a combination of a pistonlike motion and a rotation of the signal output helices was observed (*44*). Although the pistonlike displacements, helical rotations, CC phase changes, and scissoring motions are not mutually exclusive and are possibly coupled to each other, consensus on the TM signaling mechanism in TCS sensors is currently lacking (*8*, *9*).

Our results show that binding of ligand to NarQ causes a pistonlike displacement of the TM helices, which is accompanied by extensive symmetric or asymmetric rearrangements and scissoring of the TM helices. The rearrangements are different in the two presented holo state structures, but the pistonlike displacement is perfectly conserved. Thus, the latter appears to be a more robust mechanism of TM signal transduction.

Also, our data show that symmetrical changes in the sensor domain may result in both symmetrical changes in the TM domain (as in the holo-S structure) and asymmetrical changes (as in the holo-A structure). Asymmetry only changes the position of the HAMP domain relative to the TM domain and not its signaling state (Fig. 1). Sensor domains of several histidine kinases are known to reside in asymmetric conformations in the inactive state and in symmetric conformations in the active state (*40*). The aspartate chemoreceptor can be activated by asymmetric changes in the sensor domain upon binding of a ligand (*31*, *45*) and presumably also by symmetric conformational changes upon mutations of amino acids at the membrane-water interfaces (*46*, *47*). Thus, it is still unclear whether asymmetric changes in the sensor domain can result in symmetric changes in the TM domain, and the exact relations between asymmetry and activation also remain obscure.

Signal transduction through the HAMP domain

Currently, there are several competing hypotheses concerning the mechanism of signal transduction through the HAMP domain (*8*, *9*, *18*): The gearbox model postulates rotation of the HAMP domain helices in opposite directions (*48*, *49*); the dynamic bundle model proposes that signal transduction is associated with changes in HAMP domain stability (*50*, *51*); and several reports based on cross-linking and crystallographic studies propose a variety of scissoring models (*52*–*54*).

Our results indicate that in NarQ, the HAMP domain serves as an amplifier and converter of the pistonlike conformational changes in the TM domain. Pistonlike displacements of TM1 and TM1′ cytoplasmic ends for 2.5 Å relative to the proline hinge result in leverlike rotations of the protomers around their respective proline hinges and consequent 7 Å displacements in opposite directions of AS2 and AS2′ membrane-distal ends. This results in a ~90° change in the helical phase and the output CC register (Fig. 3 and fig. S8). Thus, the membrane-proximal HAMP domain acts as a converter of the pistonlike motions of TM domain helices into helical rotation at its output helices. These conformational changes are similar to the scissoring models (*52*–*54*) and are possibly accompanied by changes in the

stability of the HAMP or adjacent domains. However, our results do not support the gearbox hypothesis. We observe that the HAMP domain protomers move relative to each other in a rigid-body fashion, without rotation of the AS1 and AS2 helices within each protomer. Also, observations of slightly different HAMP domain conformations in holo-A and holo-S structures and increased *B* factors in the HAMP domain region support its dynamic nature. It is possible that the domain exists in an ensemble of similar conformations in vivo, whereas crystallographic structures provide only static snapshots. Finally, as a result of the absence of the adjacent signaling helix region in the crystallized construct, the observed conformational changes may under- or overestimate the actual transformations associated with signal transduction.

The signal converter role of the HAMP domain might reconcile the data on TM signal transduction in histidine kinases that contain and those that lack the HAMP domains. In sensors containing a membrane-proximal HAMP domain, binding of a ligand causes a pistonlike displacement of the TM helices (which may be accompanied by scissoring); this displacement is then converted into helical rotation by the HAMP domain. In sensors lacking the HAMP domain, binding of ligand directly causes rotation of the TM signal output helices (*41*, *42*). In both cases, the signal that is passed downstream is the helical rotation. Whether pistonlike displacements are present during PhoQ activation remains to be tested.

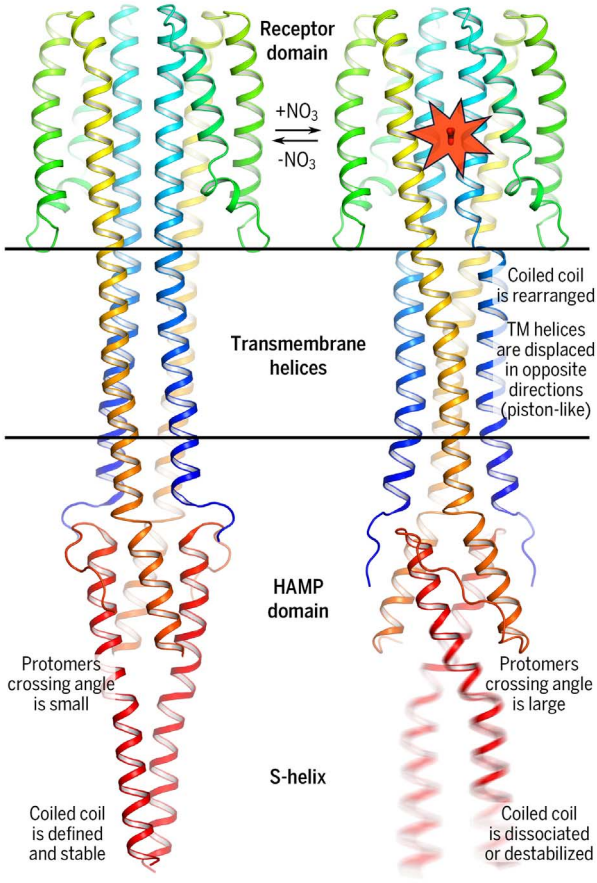


Fig. 4. Mechanism of NarQ transmembrane signaling. Binding of the ligand results in a pistonlike displacement toward the periplasm of the TM1 helices relative to TM2 and consequent leverlike conformational changes in the HAMP domain, which cause dissociation or destabilization of the signaling helix.

The multiple conformational states observed for the NarQ TM domains present a template for studies of other TCS receptors, establish the importance of the pistonlike displacements of the TM helices, and highlight the role of the HAMP domain as an amplifier and converter of a pistonlike displacement into helical rotation. Overall, our results show how a mechanistic signal is generated and amplified while being transduced through the protein over distances of 100 Å or more. Because membrane-associated TCSs are ubiquitous in microorganisms and are central for bacterial sensing (*1*–*4*), we believe that the results reported here will help in understanding of a broad array of cellular processes, ranging from basic metabolism to sporulation, quorum sensing, and virulence. They may also provide insights for development of novel antimicrobial treatments targeting TCSs (*5*–*7*).

Materials and methods

Cloning, protein expression, and purification

The nucleotide sequence encoding residues 1 to 230 of NarQ was cloned from *E. coli* strain BL21 (DE3) and introduced into the pSCodon1.2 expression vector (StabyCodon T7, Eurogentec, Belgium) via Nde I and Xho I restriction sites. Consequently, the construct harbored a C-terminal 6×His tag. Mutation R50K was introduced by site-directed mutagenesis. Both the wild-type protein and the mutant version were expressed in *E. coli* strain SE1 (StabyCodon T7, Eurogentec). Cells were cultured in shaking baffled flasks in ZYP-5052 autoinducing medium (*55*) containing ampicillin (100 mg/liter). After 2 hours of incubation at 37°C, the temperature was decreased to 30°C and incubation continued overnight. Harvested cells were disrupted in M-110P Lab Homogenizer (Microfluidics) at 25,000 psi in a PBS buffer with addition of DNase I (50 mg/liter; Sigma-Aldrich) and EDTA-free protease inhibitor cocktail Complete (Roche). The membrane fraction of cell lysate was isolated by ultracentrifugation at 90,000*g* for 1 hour at 4°C. The pellet was resuspended in a buffer containing 50 mM NaH_2PO_4/Na_2HPO_4, pH 8.0, 0.3 M NaCl, and 2% DDM (Anatrace, Affymetrix) and stirred overnight for solubilization. The insoluble fraction was then removed by ultracentrifugation at 90,000*g* for 1 hour at 4°C. The supernatant was loaded on a gravity flow column containing Ni-NTA resin (Qiagen) and the protein was eluted in a buffer containing 20 mM Tris, pH 8.0, 0.3 M NaCl, 0.2 M imidazole, and 0.1% DDM. Imidazole was removed by dialysis against 20 mM Tris, pH 8.0, 0.3 M NaCl, and 0.1% DDM for 3 hours. The eluate was subjected to size-exclusion chromatography on 125 ml Superdex 200 PG column (GE Healthcare Life Sciences) in a buffer containing 20 mM Tris, pH 8.0, 0.3 M NaCl, and 0.1% DDM. Protein-containing fractions were pooled and concentrated to 30 mg/ml for crystallization.

Crystallization

The crystals were grown using the in meso approach (*22*–*24*), similarly to our previous work (*14*, *56*, *57*). The solubilized protein in the crystallization buffer was added to the monooleoyl-formed lipidic phase (Nu-Chek Prep). Crystallization trials were set up using the NT8 robotic system (Formulatrix). The crystals were grown at 22°C and reached the final size of 50 to 150 μm within 2 weeks. The best holo state crystals in space groups $F222$ and $P2$ were obtained using the precipitants 0.6 M KH_2PO_4/Na_2HPO_4, pH 4.6, 5 mM $NaNO_3$, and 0.3 M KH_2PO_4/Na_2HPO_4, pH 7.6, 250 mM NaBr, respectively. The best apo state crystals in the space group $I2_12_12_1$ were obtained using the precipitant 1.6 M $(NH_4)_3PO_4$, pH 5.8. Before harvesting, the crystals were incubated for 5 min in the respective precipitant solutions supplemented with 20% glycerol. For iodide-SAD experiments, the soaking solution additionally contained 0.5 M NaI. All crystals were harvested using micromounts (MiTeGen) and were flash-cooled and stored in liquid nitrogen.

Acquisition and treatment of diffraction data

The diffraction data were collected at 100 K at ESRF beamline ID23-1 (*58*) equipped with a PILATUS 6M-F detector. The data collection statistics are reported in table S1. In all cases the diffraction was anisotropic as determined by decay of the $CC_{1/2}$ values in 20° cones along the reciprocal cell directions a^*, b^*, and c^* (*59*). In the space group $I2_12_12_1$, the resolution limits along the directions a^* and b^* were 2.55 and 2.17 Å, and along the direction c^* the $CC_{1/2}$ value was ~0.9 at the resolution cutoff of 1.9 Å. In the native data set in the space group $F222$, the resolution limits along the directions a^* and b^* were 2.26 and 1.94 Å, and along the direction c^* the $CC_{1/2}$ value was ~0.8 at the resolution cutoff of 1.94 Å. In the space group $P2$, the resolution limit along the direction a^* was 2.82 Å, and along the directions b^* and c^* the $CC_{1/2}$ values were ~0.6 and 0.9 at the resolution cutoff of 2.42 Å. Diffraction images were processed using XDS (*60*). XSCALE (*60*) was used to merge different data sets and to scale the data for the phasing steps.

POINTLESS (*59*) and AIMLESS (*59*) were used to merge, scale, assess the quality, convert intensities to structure factor amplitudes, and generate Free-R labels.

Structure determination and refinement

The holo-S state structure in the space group $F222$ was solved using experimental phasing (*61*), and the apo and holo-A state structures in the space groups $I2_12_12_1$ and $P2$, respectively, were solved using molecular replacement with MOLREP (*62*) and the sensor domain from the structure in the space group $F222$ as a search model. For the $F222$ structure solution, eight isomorphous data sets from NaI-soaked crystals were collected and merged for subsequent SAD phasing. The iodide sites were determined using ShelxD (*63*) with the HKL2MAP (*64*) interface, with CC_{all} and CC_{weak} values of 40.0% and 15.4%. The resolution was extended to 1.94 Å using the native $F222$ data collected from a single crystal. 178 Ala residues were placed in 5×50 cycles of autotracing and density modification using ShelxE (*63*). The initial model was improved using automatic model building software ARP-wARP (*65*). The final $F222$ model was built using only the native data. For the $I2_12_12_1$ and $P2$ structure solution, native data sets collected from single crystals were used in each case. All models were refined manually using Coot (*66*) and REFMAC5 (*67*). Intermediate conformations shown in movies S1 and S2 were calculated using the NOLB algorithm (*68*).

REFERENCES AND NOTES

1. T. Krell *et al*., Bacterial sensor kinases: Diversity in the recognition of environmental signals. *Annu. Rev. Microbiol.* **64**, 539–559 (2010). doi: 10.1146/annurev.micro.112408.134054; pmid: 20825354
2. A. M. Stock, V. L. Robinson, P. N. Goudreau, Two-component signal transduction. *Annu. Rev. Biochem.* **69**, 183–215 (2000). doi: 10.1146/annurev.biochem.69.1.183; pmid: 10966457
3. P. M. Wolanin, P. A. Thomason, J. B. Stock, Histidine protein kinases: Key signal transducers outside the animal kingdom. *Genome Biol.* **3**, S3013 (2002). doi: 10.1186/gb-2002-3-10-reviews3013; pmid: 12372152
4. L. E. Ulrich, I. B. Zhulin, The MiST2 database: A comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res.* **38** (suppl. 1), D401–D407 (2010). doi: 10.1093/nar/gkp940; pmid: 19900966
5. Y. Gotoh *et al*., Two-component signal transduction as potential drug targets in pathogenic bacteria. *Curr. Opin. Microbiol.* **13**, 232–239 (2010). doi: 10.1016/j.mib.2010.01.008; pmid: 20138000
6. B. LaSarre, M. J. Federle, Exploiting quorum sensing to confuse bacterial pathogens. *Microbiol. Mol. Biol. Rev.* **77**, 73–111 (2013). doi: 10.1128/MMBR.00046-12; pmid: 23471618
7. D. A. Rasko, V. Sperandio, Anti-virulence strategies to combat bacteria-mediated disease. *Nat. Rev. Drug Discov.* **9**, 117–128 (2010). doi: 10.1038/nrd3013; pmid: 20081869
8. C. P. Zschiedrich, V. Keidel, H. Szurmant, Molecular mechanisms of two-component signal transduction. *J. Mol. Biol.* **428**, 3752–3775 (2016). doi: 10.1016/j.jmb.2016.08.003; pmid: 27519796
9. M. P. Bhate, K. S. Molnar, M. Goulian, W. F. DeGrado, Signal transduction in histidine kinases: Insights from new structures. *Structure* **23**, 981–994 (2015). doi: 10.1016/j.str.2015.04.002; pmid: 25982528
10. M. Li, G. L. Hazelbauer, Core unit of chemotaxis signaling complexes. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9390–9395 (2011). doi: 10.1073/pnas.1104824108; pmid: 21606342
11. C. M. Khursigara, X. Wu, S. Subramaniam, Chemoreceptors in *Caulobacter crescentus*: Trimers of receptor dimers in a partially ordered hexagonally packed array. *J. Bacteriol.* **190**, 6805–6810 (2008). doi: 10.1128/JB.00640-08; pmid: 18689468
12. P. S. Orekhov *et al*., Signaling and adaptation modulate the dynamics of the photosensoric complex of *Natronomonas pharaonis*. *PLOS Comput. Biol.* **11**, e1004561 (2015). doi: 10.1371/journal.pcbi.1004561; pmid: 26496122
13. G. L. Hazelbauer, J. J. Falke, J. S. Parkinson, Bacterial chemoreceptors: High-performance signaling in networked arrays. *Trends Biochem. Sci.* **33**, 9–19 (2008). doi: 10.1016/j.tibs.2007.09.014; pmid: 18165013
14. V. I. Gordeliy *et al*., Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* **419**, 484–487 (2002). doi: 10.1038/nature01109; pmid: 12368857
15. I. Maslennikov *et al*., Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10902–10907 (2010). doi: 10.1073/pnas.1001656107; pmid: 20498088
16. A. A. Pakula, M. I. Simon, Determination of transmembrane protein structure by disulfide cross-linking: The *Escherichia coli* Tar receptor. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4144–4148 (1992). doi: 10.1073/pnas.89.9.4144; pmid: 1315053
17. K. S. Molnar *et al*., Cys-scanning disulfide crosslinking and bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. *Structure* **22**, 1239–1251 (2014). doi: 10.1016/j.str.2014.04.019; pmid: 25087511
18. J. S. Parkinson, Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. *Annu. Rev. Microbiol.* **64**, 101–122 (2010). doi: 10.1146/annurev.micro.112408.134215; pmid: 20690824
19. V. Stewart, L.-L. Chen, The S helix mediates signal transmission as a HAMP domain coiled-coil extension in the NarX nitrate sensor from *Escherichia coli* K-12. *J. Bacteriol.* **192**, 734–745 (2010). doi: 10.1128/JB.00172-09; pmid: 19966007
20. V. Stewart, Nitrate- and nitrite-responsive sensors NarX and NarQ of proteobacteria. *Biochem. Soc. Trans.* **31**, 1–10 (2003). doi: 10.1042/bst0310001; pmid: 12546643
21. J. Cheung, W. A. Hendrickson, Structural analysis of ligand stimulation of the histidine kinase NarX. *Structure* **17**, 190–201 (2009). doi: 10.1016/j.str.2008.12.013; pmid: 19217390
22. V. I. Gordeliy, R. Schlesinger, R. Efremov, G. Büldt, J. Heberle, in *Membrane Protein Protocols: Expression, Purification, and Characterization*, B. S. Selinsky, Ed. (Humana, 2003), pp. 305–316.
23. M. Caffrey, Crystallizing membrane proteins for structure determination: Use of lipidic mesophases. *Annu. Rev. Biophys.* **38**, 29–51 (2009). doi: 10.1146/annurev.biophys.050708.133655; pmid: 19086821
24. V. Cherezov, Lipidic cubic phase technologies for membrane protein structural studies. *Curr. Opin. Struct. Biol.* **21**, 559–566 (2011). doi: 10.1016/j.sbi.2011.06.007; pmid: 21775127
25. Z. Dauter, M. Dauter, K. R. Rajashankar, Novel approach to phasing proteins: Derivatization by short cryo-soaking with halides. *Acta Crystallogr. D* **56**, 232–237 (2000). doi: 10.1107/S0907444999016352; pmid: 10666615
26. S. B. Williams, V. Stewart, Discrimination between structurally related ligands nitrate and nitrite controls autokinase activity of the NarX transmembrane signal transducer of *Escherichia coli* K-12. *Mol. Microbiol.* **26**, 911–925 (1997). doi: 10.1046/j.1365-2958.1997.6262002.x; pmid: 9426129
27. M. D. Coleman, R. B. Bass, R. S. Mehan, J. J. Falke, Conserved glycine residues in the cytoplasmic domain of the aspartate receptor play essential roles in kinase coupling and on-off switching. *Biochemistry* **44**, 7687–7695 (2005). doi: 10.1021/bi0501479; pmid: 15909983
28. A. Ishchenko *et al*., Ground state structure of D75N mutant of sensory rhodopsin II in complex with its cognate transducer. *J. Photochem. Photobiol. B* **123**, 55–58 (2013). doi: 10.1016/j.jphotobiol.2013.03.008; pmid: 23619282
29. S. Dunin-Horkawicz, A. N. Lupas, Comprehensive analysis of HAMP domains: Implications for transmembrane signal transduction. *J. Mol. Biol.* **397**, 1156–1174 (2010). doi: 10.1016/j.jmb.2010.02.031; pmid: 20184894
30. M. D. Hartmann *et al*., A soluble mutant of the transmembrane receptor Af1503 features strong changes in coiled-coil periodicity. *J. Struct. Biol.* **186**, 357–366 (2014). doi: 10.1016/j.jsb.2014.02.008; pmid: 24568954
31. J. S. Parkinson, G. L. Hazelbauer, J. J. Falke, Signaling and sensory adaptation in *Escherichia coli* chemoreceptors: 2015 update. *Trends Microbiol.* **23**, 257–266 (2015). doi: 10.1016/j.tim.2015.03.003; pmid: 25834953
32. V. Anantharaman, S. Balaji, L. Aravind, The signaling helix: A common functional theme in diverse signaling proteins. *Biol. Direct* **1**, 25 (2006). doi: 10.1186/1745-6150-1-25; pmid: 16953892
33. I. Y. Gushchin, V. I. Gordeliy, S. Grudinin, Role of the HAMP domain region of sensory rhodopsin transducers in signal transduction. *Biochemistry* **50**, 574–580 (2011). doi: 10.1021/bi101032a; pmid: 21162553
34. H. U. Ferris *et al*., Mechanism of regulation of receptor histidine kinases. *Structure* **20**, 56–66 (2012). doi: 10.1016/j.str.2011.11.014; pmid: 22244755
35. R. S. Vieira-Pires, J. H. Morais-Cabral, 3_{10} helices in channels and other membrane proteins. *J. Gen. Physiol.* **136**, 585–592 (2010). doi: 10.1085/jgp.201010508; pmid: 21115694
36. F. H. Crick, The packing of α-helices: Simple coiled-coils. *Acta Crystallogr.* **6**, 689–697 (1953). doi: 10.1107/S0365110X53001964
37. R. C. Chiang, R. Cavicchioli, R. P. Gunsalus, 'Locked-on' and 'locked-off' signal transduction mutations in the periplasmic domain of the *Escherichia coli* NarQ and NarX sensors affect nitrate- and nitrite-dependent regulation by NarL and NarP. *Mol. Microbiol.* **24**, 1049–1060 (1997). doi: 10.1046/j.1365-2958.1997.4131779.x; pmid: 9220011
38. E. Lesne *et al*., Balance between coiled-coil stability and dynamics regulates activity of BvgS sensor kinase in *Bordetella*. *MBio* **7**, e02089-15 (2016). doi: 10.1128/mBio.02089-15; pmid: 26933056
39. E. Saita *et al*., A coiled coil switch mediates cold sensing by the thermosensory protein DesK. *Mol. Microbiol.* **98**, 258–271 (2015). doi: 10.1111/mmi.13118; pmid: 26172072
40. J. O. Moore, W. A. Hendrickson, An asymmetry-to-symmetry switch in signal transmission by the histidine kinase receptor for TMAO. *Structure* **20**, 729–741 (2012). doi: 10.1016/j.str.2012.02.021; pmid: 22483119
41. M. B. Neiditch *et al*., Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. *Cell* **126**, 1095–1108 (2006). doi: 10.1016/j.cell.2006.07.032; pmid: 16990134
42. B. Wang, A. Zhao, R. P. Novick, T. W. Muir, Activation and inhibition of the receptor histidine kinase AgrC occurs through opposite helical transduction motions. *Mol. Cell* **53**, 929–940 (2014). doi: 10.1016/j.molcel.2014.02.029; pmid: 24656130
43. H. Szurmant, M. W. Bunn, S. H. Cho, G. W. Ordal, Ligand-induced conformational changes in the *Bacillus subtilis* chemoreceptor McpB determined by disulfide crosslinking in vivo. *J. Mol. Biol.* **344**, 919–928 (2004). doi: 10.1016/j.jmb.2004.09.093; pmid: 15544802
44. R. Moukhametzianov *et al*., Development of the signal in sensory rhodopsin and its transfer to the cognate transducer. *Nature* **440**, 115–119 (2006). doi: 10.1038/nature04520; pmid: 16452929
45. S. A. Chervitz, J. J. Falke, Molecular mechanism of transmembrane signaling by the aspartate receptor: A model. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2545–2550 (1996). doi: 10.1073/pnas.93.6.2545; pmid: 8637911
46. A. S. Miller, J. J. Falke, Side chains at the membrane-water interface modulate the signaling state of a transmembrane receptor. *Biochemistry* **43**, 1763–1770 (2004). doi: 10.1021/bi0360206; pmid: 14967017
47. R. R. Draheim, A. F. Bormans, R. Z. Lai, M. D. Manson, Tryptophan residues flanking the second transmembrane helix (TM2) set the signaling state of the Tar chemoreceptor. *Biochemistry* **44**, 1268–1277 (2005). doi: 10.1021/bi048969d; pmid: 15667220
48. M. Hulko *et al*., The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell* **126**, 929–940 (2006). doi: 10.1016/j.cell.2006.06.058; pmid: 16959572
49. H. U. Ferris *et al*., The mechanisms of HAMP-mediated signaling in transmembrane receptors. *Structure* **19**, 378–385 (2011). doi: 10.1016/j.str.2011.01.006; pmid: 21397188
50. Q. Zhou, P. Ames, J. S. Parkinson, Mutational analyses of HAMP helices suggest a dynamic bundle model of input-output signalling in chemoreceptors. *Mol. Microbiol.* **73**, 801–814 (2009). doi: 10.1111/j.1365-2958.2009.06819.x; pmid: 19656294
51. V. Stewart, The HAMP signal-conversion domain: Static two-state or dynamic three-state? *Mol. Microbiol.* **91**, 853–857 (2014). doi: 10.1111/mmi.12516; pmid: 24417364
52. K. E. Swain, J. J. Falke, Structure of the conserved HAMP domain in an intact, membrane-bound chemoreceptor: A disulfide mapping study. *Biochemistry* **46**, 13684–13695 (2007). doi: 10.1021/bi701832b; pmid: 17994770

53. M. V. Airola, K. J. Watts, A. M. Bilwes, B. R. Crane, Structure of concatenated HAMP domains provides a mechanism for signal transduction. *Structure* **18**, 436–448 (2010). doi: 10.1016/j.str.2010.01.013; pmid: 20399181
54. I. Gushchin, V. Gordeliy, S. Grudinin, Two distinct states of the HAMP domain from sensory rhodopsin transducer observed in unbiased molecular dynamics simulations. *PLOS ONE* **8**, e66917 (2013). doi: 10.1371/journal.pone.0066917; pmid: 23843970
55. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005). doi: 10.1016/j.pep.2005.01.016; pmid: 15915565
56. I. Gushchin *et al*., Crystal structure of a light-driven sodium pump. *Nat. Struct. Mol. Biol.* **22**, 390–395 (2015). doi: 10.1038/nsmb.3002; pmid: 25849142
57. I. Gushchin *et al*., Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12631–12636 (2013). doi: 10.1073/pnas.1221629110; pmid: 23872846
58. D. Nurizzo *et al*., The ID23-1 structural biology beamline at the ESRF. *J. Synchrotron Radiat.* **13**, 227–238 (2006). doi: 10.1107/S0909049506004341; pmid: 16645249
59. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006). doi: 10.1107/S0907444905036693; pmid: 16369096
60. W. Kabsch, XDS. *Acta Crystallogr. D* **66**, 125–132 (2010). doi: 10.1107/S0907444909047337; pmid: 20124692
61. I. Melnikov *et al*., Fast iodide-SAD phasing for high-throughput membrane protein structure determination. *Sci. Adv.* **3**, e1602952 (2017). doi: 10.1126/sciadv.1602952
62. A. Vagin, A. Teplyakov, Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010). doi: 10.1107/S0907444909042589; pmid: 20057045
63. G. M. Sheldrick, Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 (2010). doi: 10.1107/S0907444909038360; pmid: 20383001
64. T. Pape, T. R. Schneider, HKL2MAP: A graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844 (2004). doi: 10.1107/S0021889804018047
65. G. Langer, S. X. Cohen, V. S. Lamzin, A. Perrakis, Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008). doi: 10.1038/nprot.2008.91; pmid: 18600222
66. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
67. G. N. Murshudov *et al*., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011). doi: 10.1107/S0907444911001314; pmid: 21460454
68. A. Hoffmann, S. Grudinin, NOLB: Nonlinear rigid block normal-mode analysis method. *J. Chem. Theory Comput.* **13**, 2123–2134 (2017). doi: 10.1021/acs.jctc.7b00197
69. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012). doi: 10.1093/nar/gkr703; pmid: 21890895

ACKNOWLEDGMENTS

Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank under accession codes 5IJI, 5JEF, and 5JEQ. The work was done in the framework of CEA(IBS)–HGF(FZJ) STC 5.1 specific agreement and supported by ERA.Net RUS PLUS and Ministry of Education and Science of the Russian Federation (project ID 323, RFMEFI58715X0011). The work used the platforms of the Grenoble Instruct Centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology. I.M. is the recipient of a studentship funded by the Ph.D. program of the European Synchrotron Radiation Facility. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities. Author contributions: I.G., A.I., and V.G. designed the study; A.I. expressed and purified the wild-type protein; I.G., P.B., and T.B. helped with expression and purification of the wild-type protein; A.Yu. expressed and purified the R50K mutant; I.G., I.M., V.P., and A.I. crystallized the protein; E.R. helped with crystallization; A.P. solved the structures; G. Bourenkov and G.L. helped with structure solution; I.G. and I.M. refined the structures; I.G. analyzed the structures; S.G., V.B., D.W., and G. Büldt helped with structure analysis; I.G. and V.G. oversaw the study, analyzed the results, and prepared the manuscript with contributions from all other coauthors.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/356/6342/eaah6345/suppl/DC1
Figs. S1 to S11
Table S1
Movies S1 and S2

24 July 2016; accepted 8 May 2017
Published online 18 May 2017
10.1126/science.aah6345

Mechanism of transmembrane signaling by sensor histidine kinases

Ivan Gushchin, Igor Melnikov, Vitaly Polovinkin, Andrii Ishchenko, Anastasia Yuzhakova, Pavel Buslaev, Gleb Bourenkov, Sergei Grudinin, Ekaterina Round, Taras Balandin, Valentin Borshchevskiy, Dieter Willbold, Gordon Leonard, Georg Büldt, Alexander Popov and Valentin Gordeliy

Science **356** (6342), eaah6345.
DOI: 10.1126/science.aah6345originally published online May 18, 2017

Bacterial sensing mechanism revealed

Escherichia coli use a transmembrane sensor protein to sense nitrate in their external environment and initiate a biochemical response. Gushchin *et al.* compared crystal structures of portions of the NarQ receptor that included the transmembrane helices in ligand-bound or unbound states. The structures suggest a signaling mechanism by which piston- and lever-like movements are transmitted to response regulator proteins within the cell. Such two-component systems are very common in bacteria and, if better understood, might provide targets for antimicrobial therapies.

Science, this issue p. eaah6345

www.sciencemag.org/cgi/content/full/science.aah6345/DC1

Supplementary Materials for

Mechanism of transmembrane signaling by sensor histidine kinases

Ivan Gushchin,* Igor Melnikov, Vitaly Polovinkin, Andrii Ishchenko, Anastasia Yuzhakova, Pavel Buslaev, Gleb Bourenkov, Sergei Grudinin, Ekaterina Round, Taras Balandin, Valentin Borshchevskiy, Dieter Willbold, Gordon Leonard, Georg Büldt, Alexander Popov, Valentin Gordeliy*

*Corresponding author. Email: ivan.gushchin@phystech.edu (I.G.); valentin.gordeliy@ibs.fr (V.G.)

Published 18 May 2017 on *Science* First Release
DOI: 10.1126/science.aah6345

This PDF file includes:

Other supplementary material for this manuscript includes the following:

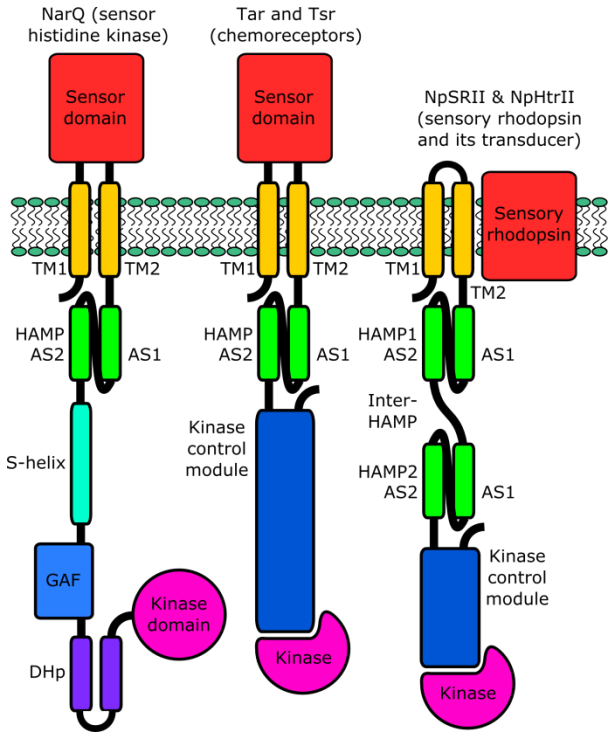


Figure S1. Domain architectures of representative microbial membrane sensor proteins. TM1 and TM2, transmembrane helices. AS1 and AS2, HAMP domain helices. NarQ is *E. coli* nitrate and nitrite sensor, Tsr is *E. coli* serine chemoreceptor, *Np*SRII and *Np*HtrII are *Natronomonas pharaonis* sensory rhodopsin II and sensory rhodopsin transducer II. In chemo- and photoreceptor systems, the kinase is a separate protein. While only monomers are shown, all of the proteins are normally dimeric with chemo- and photoreceptor dimers forming higher-order oligomeric assemblies.

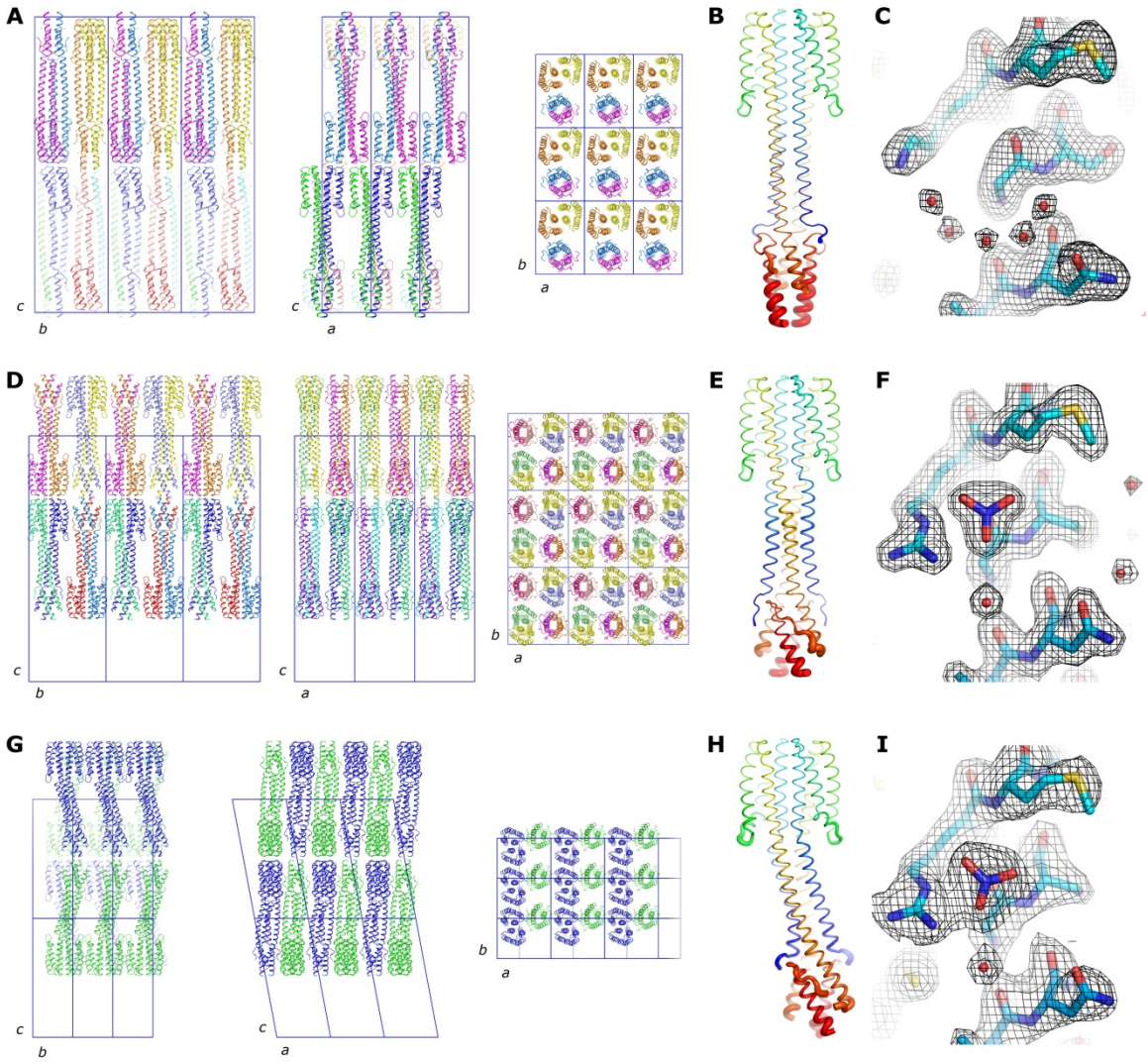


Figure S2. Crystallographic details. (A-C) Data for the space group $I2_12_12_1$ (symmetric *apo* state structure). **(D-F)** Data for the space group F222 (symmetric *holo* state structure). **(G-I)** Data for the space group P2 (asymmetric *holo* state structure). **(A, D, G)** Crystal packing. The NarQ dimers pack in layers, as is typical for Type I membrane protein crystals, obtained through *in meso* crystallization. **(B, E, H)** Relative distributions of the B-factor values. The scale in each drawing is different. In each space group, the membrane-distal parts of the HAMP domain are notably disordered. **(C, F, I)** Example of the electron density maps for the chain A residues around the ligand binding site. Weighted $2F_o–F_c$ electron density maps are contoured at the level of 1.5 σ.

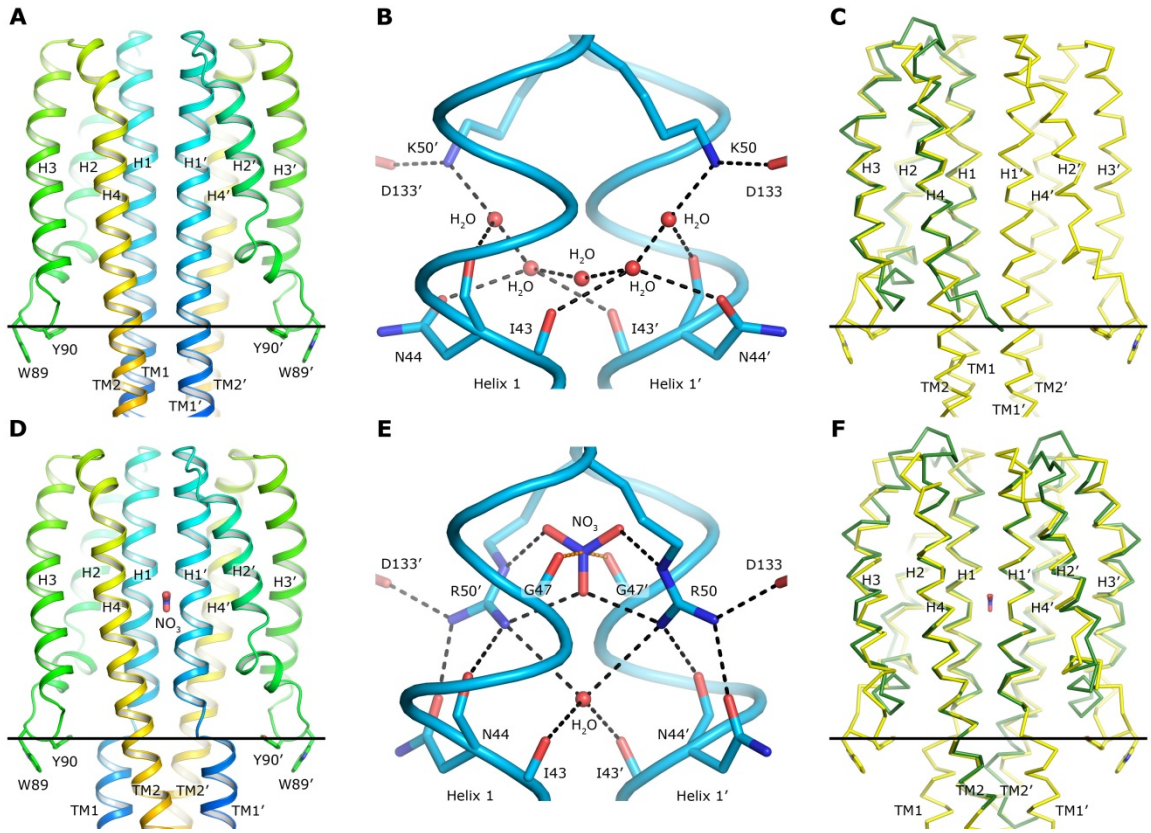


Figure S3. Structure of the NarQ periplasmic sensor domain. (A-C) Structure of the sensor domain in the *apo* state. The R50K mutant was used for structure determination. **(A)** Overall structure of the domain. The H2-H3 loop is anchored into the membrane with the residues Trp89 and Tyr90. **(B)** Ligand-binding site of *apo* NarQ. Three ordered water molecules are observed. **(C)** Comparison of the NarQ (yellow) and NarX (green, chain A from PDB: 3EZI (*21*)) sensor domains in the absence of ligands. Positions of the backbone atoms of the helices H1 and H4 are virtually indistinguishable. **(D-F)** Structure of the sensor domain in the *holo* state. **(D)** Overall structure of the domain. The ligand is bound at the dimerization interface between the helices H1 and H1'. The H2-H3 loop is anchored into the membrane with the residues Trp89 and Tyr90. **(E)** Nitrate-binding site of NarQ. The ion is bound by Arg50 side chains and Gly47 backbone oxygen atoms. Below the nitrate ion, a tetrahedrally coordinated water molecule is observed. Potential hydrogen bonds are shown as black dashed lines. The interaction of NO_3^- partially positively charged nitrogen atom with the Gly47 partially negatively charged backbone oxygen atom is shown with orange dashing. **(F)** Comparison of the ligand-bound NarQ (yellow) and NarX (green, PDB: 3EZH (*21*)) sensor domains. Positions of the backbone atoms of the helices H1 and H4 are virtually indistinguishable. Positions of the TM2 backbone atoms present in the NarX construct used for crystallization(*21*) are also identical to those of NarQ. Positions of the NarX TM2' backbone atoms are different, possibly due to the absence of corresponding TM1 and TM1' residues that would stabilize the correct conformation. Note that NarX position relative to the membrane boundaries probably differs from that of NarQ (not shown).

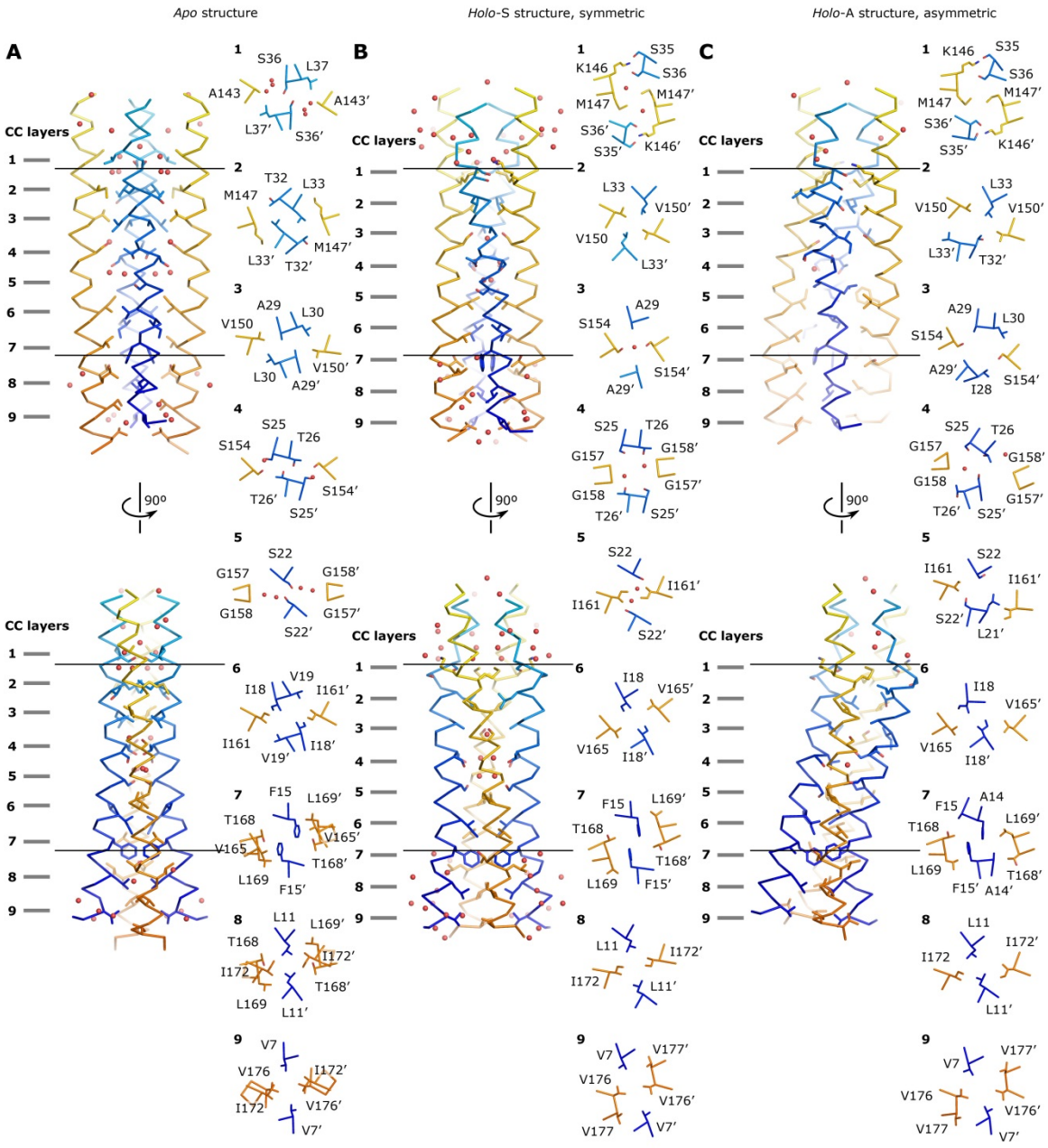


Figure S4. Structure of the NarQ transmembrane domain. TM1 is in blue, and TM2 is in yellow. **(A)** *Apo* state. **(B)** Symmetric *holo* state. **(C)** Asymmetric *holo* state. In each panel, overall structure (left) and side chain packings in the 9 CC layers (right, view from the periplasm) are shown. In each structure, ordered water molecules are observed in or around CC layer 4.

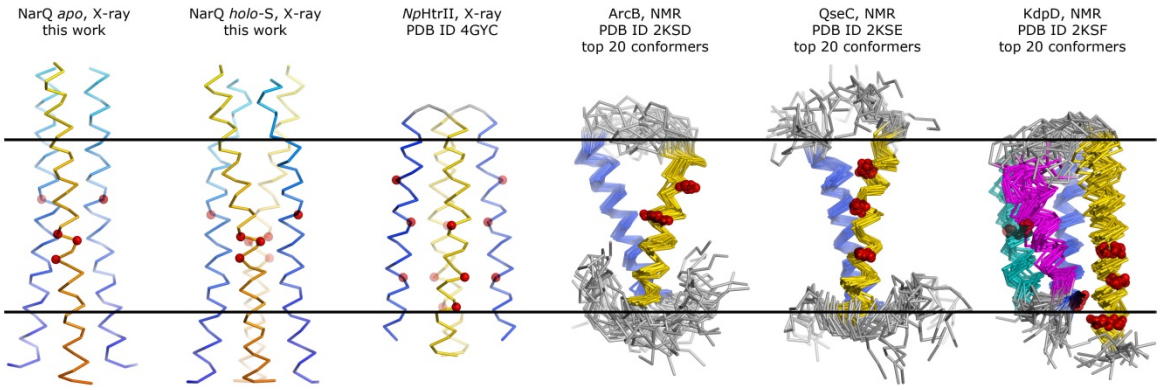


Figure S5. Comparison of the experimentally determined structures of TCS TM domains. In each case, the N-terminal α-helix TM1 is shown in blue and the C-terminal one is shown in yellow. Glycine and proline residues of the TM helices are shown with red spheres. There are generally more glycine and proline residues in the last TM helix of the histidine kinases. The scale is the same for all of the proteins. X-ray structure of *Np*HtrII (*28*) is aligned to ligand-bound NarQ structure using the positions of the HAMP-domain proximal residues that are similar, opposite to the positions of the residues in the middle of the membrane and at the periplasmic side. Since there is no sequence similarity between the TM domains, NMR structures of ArcB, QseC and KdpD (*15*) are aligned to the NarQ structure so that their helices are roughly at the same positions relative to the membrane. Structural comparison with ArcB, QseC and KdpD is hampered since the structures are monomeric and information about the dimerization interfaces is lacking (*15*). It is not clear whether the ArcB and QseC fragments are in the correct physiological state. In the KdpD structure, presence of the additional TM helices (TM2 and TM3, cyan and magenta) probably ensures correct folding of the TM domain and proper orientations of the N-terminal and C-terminal helices.

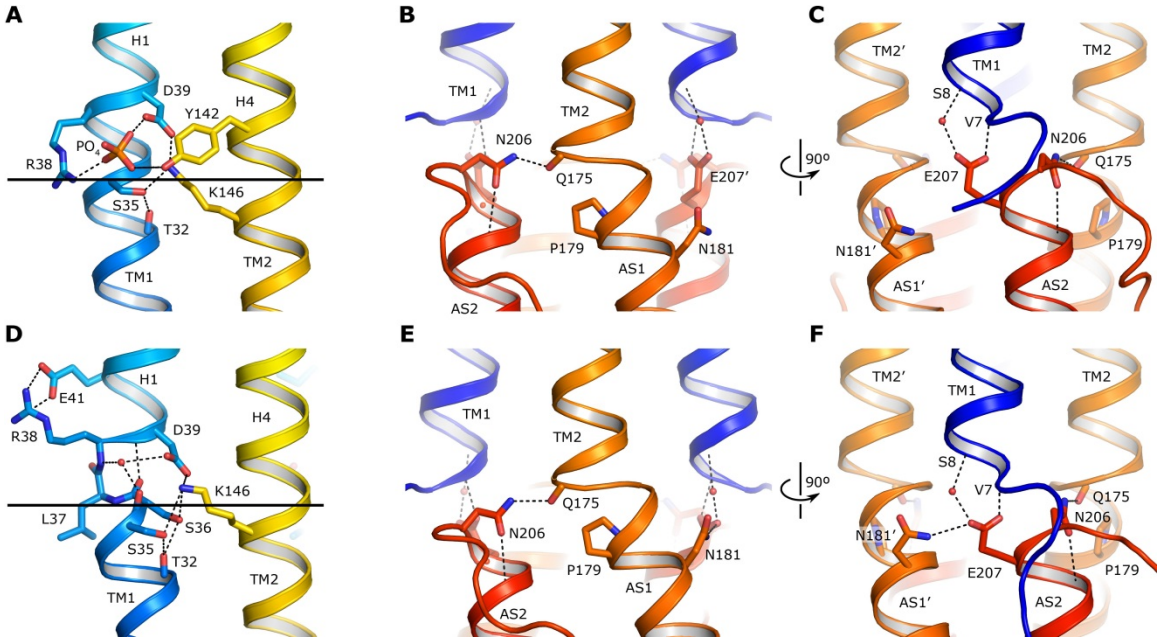


Figure S6. Structure of the junctions between the NarQ TM helices and adjacent domains. (A-C) *Apo* state structure. **(A)** Connections between the TM helices and the sensor domain. Both TM1-H1 and TM2-H4 transitions maintain the helical structure. A phosphate anion is bound close to the hydrophobic membrane boundary. Putative hydrogen bonds are shown with the black dashes. **(B)** and **(C)** Connections between the TM helices and the HAMP domain. **(B)** Structure of the proline kink in the TM2-AS1 junction. **(C)** Structure of the TM1-AS2 link. Cytoplasmic terminus of TM1 is connected to the membrane-proximal end of AS2 via two hydrogen bonds. One bond is direct (Val7-Glu207) and the other is mediated by a conserved water molecule (Ser8-Glu207). **(D-F)** *Holo*-S symmetric state structure. **(D)** Connections between the TM helices and the sensor domain. There is a break in the helical structure at the TM1-H1 junction, while the H4-TM2 transition is continuous. The break is stabilized by a water molecule and several hydrogen bonds (black dashes). **(E)** and **(F)** Connections between the TM helices and the HAMP domain. **(E)** Structure of the proline kink in the TM2-AS1 junction. There is a hydrogen bond between Asn206 and Gln175 backbone carbonyl oxygen atom that connects the cytoplasmic end of TM2 and membrane-proximal end of AS2. **(F)** Structure of the TM1-AS2 link. The cytoplasmic terminus of TM1 is connected to the membrane-proximal end of AS2 via two hydrogen bonds. One bond is direct (Val7-Glu207) and the other is mediated by a conserved water molecule (Ser8-Glu207). Glu207 side chain conformation is also stabilized by a hydrogen bond with Asn181’ of the adjacent helix AS1’.

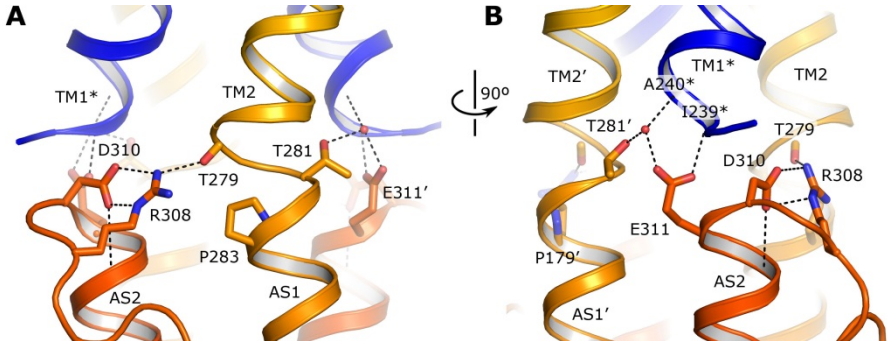


Figure S7. Structure of the junctions between the TM helices and HAMP domain in Af1503 (PDB: 4CQ4 (*30*)**). (A)** Structure of the proline kink in the TM2-AS1 junction. **(B)** Structure of the TM1*-AS2 link. Cytoplasmic terminus of the helix TM1* is connected to the membrane-proximal end of AS2 via two hydrogen bonds. One bond is direct (Ile239*-Glu311) and the other is mediated by a water molecule (Ala240*-Glu311). The water molecule is also stabilized by a hydrogen bond with the side chain of Thr181' of the adjacent helix AS1'. Notation TM1* is chosen to emphasize that the structure corresponds to unnatural antiparallel tetramer of the construct comprising the mutated TM2 helix and the HAMP domain of the Af1503 protein (*30*). Consequently, positions of TM1 in the correctly folded full-length Af1503 can differ from those presented.

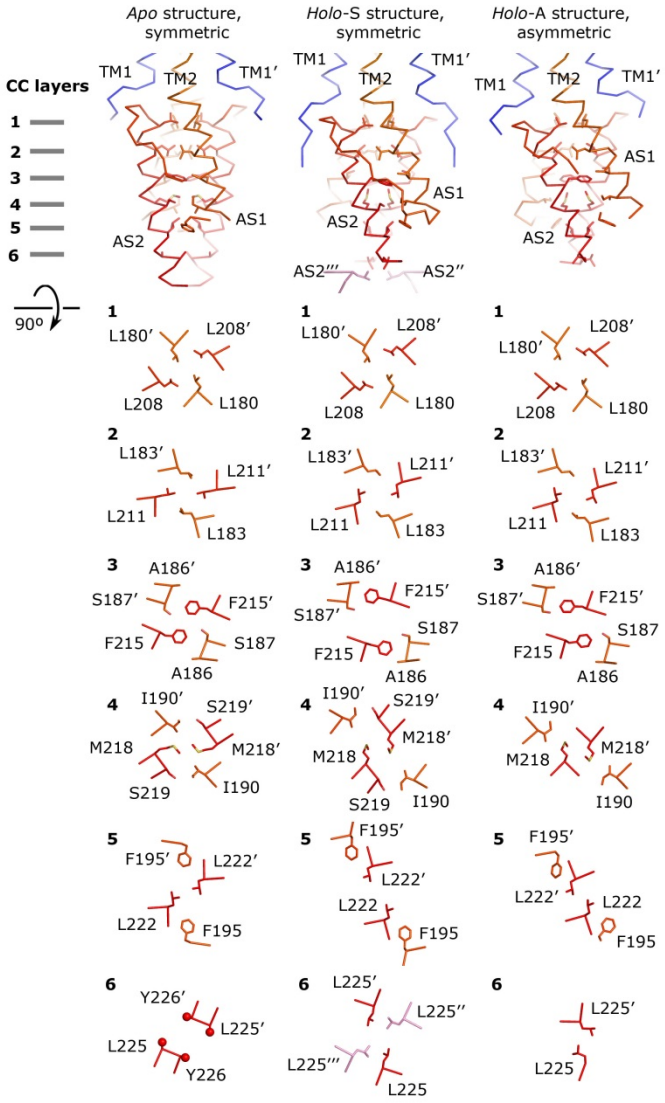


Figure S8. Structure of the NarQ HAMP domain in *apo*, symmetric *holo* and asymmetric *holo* states. Structures are aligned by the residues of the TM2-AS1 junction. In the *apo* state structure, side chains of the residues Leu225 and Tyr226 are not resolved and consequently are marked with spheres. In the symmetric *holo* state structure, AS2 helices of symmetry-related molecules are marked AS2'' and AS2''', and only one of two alternative side chain conformations of Leu225 is shown in each case. Note that Phe195 of layer 5 resides on the linker and not on AS1 or AS2, and that not all of the linker residues are resolved in the *apo* and asymmetric *holo* state structures.

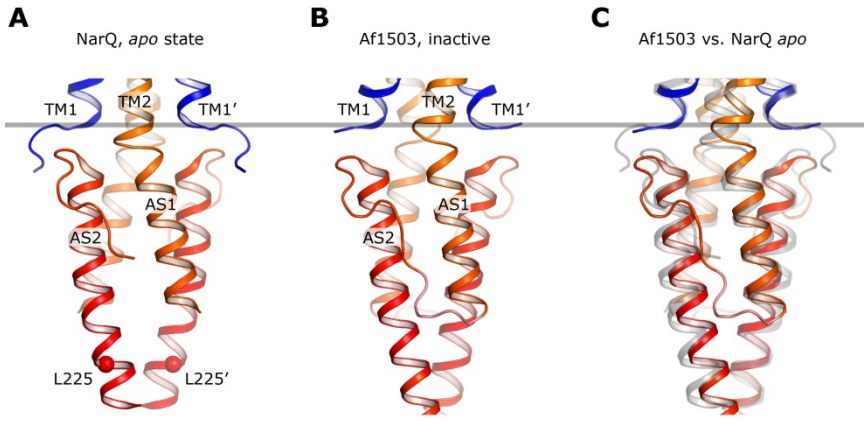


Figure S9. Comparison of the NarQ *apo* state HAMP domain and Af1503 soluble mutant HAMP domain. (A) Structure of the HAMP domain in the *apo* state of NarQ. **(B)** Structure of the HAMP domain in the soluble mutant of Af1503 receptor (PDB: 4CQ4) (*30*). **(C)** Superposition of the *apo*-state NarQ HAMP domain (grey) and Af1503 HAMP domain (colored). The HAMP domain structures are similar. The grey line shows the position of TM1 ends in the *apo* state NarQ structure.

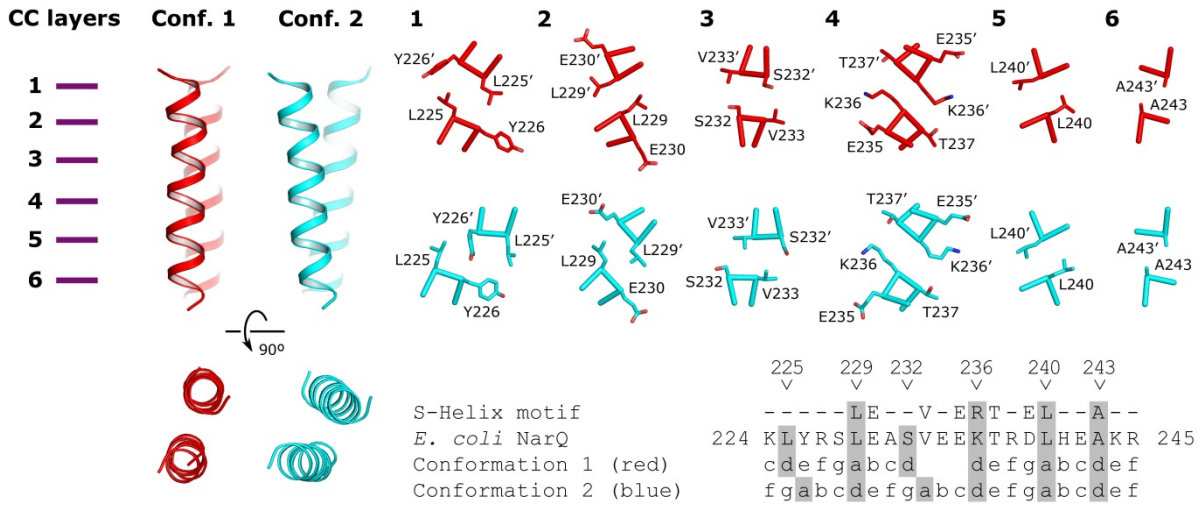


Figure S10. Coiled coil models of NarQ signaling helix domain. Two conformations are possible. Conformation 1 is shown in red and conformation 2 is shown in blue. The packing in the layers 1-3 is different, while the packing in the layers 4-6 is similar. The phase stutter is possible in the region of the residues 232-236. Residues Glu235, Lys236, Thr237 belong to the characteristic conserved signaling helix motif (*32*).

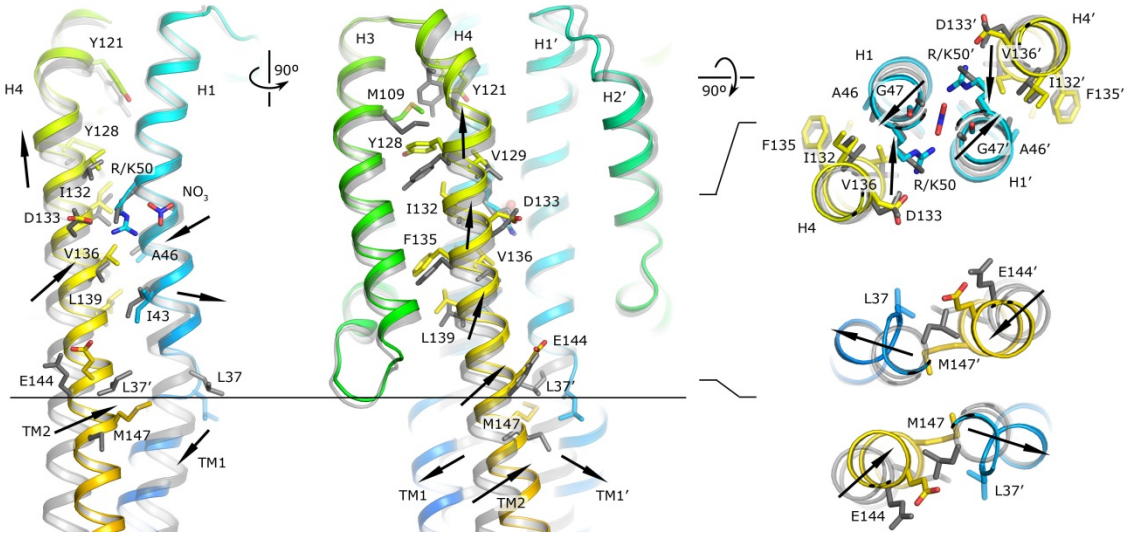


Figure S11. Signaling-associated conformational changes in the NarQ sensor domain. Superposition of the *apo* state structure (grey) and symmetric *holo* state structure (colored) is shown. The black arrows indicate the directions of the displacements of NarQ elements upon binding of the ligand (not drawn to scale). The structures are aligned using the residues 48–53 (membrane-distal part of the helix H1).

Table S1. Data collection, phasing and refinement statistics

	Apo	*Holo* F222 NaI	*Holo* F222 Native	*Holo* P2
Data collection				
Space group	$I2_12_12_1$	F222	F222	P2
Cell dimensions				
a, *b*, *c* (Å)	39.59, 59.37, 239.58	57.17, 73.48, 236.45	57.58, 73.87, 235.56	53.46, 39.87, 120.36
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 100.62, 90
Wavelength (Å)	0.972	1.85	0.972	0.972
Resolution (Å)	47.64-1.90 (1.94-1.90)*	60-2.7 (2.9-2.7)	39.4-1.94 (1.99-1.94)	52.55-2.42 (2.51-2.42)
R_{merge} (%)	7.5 (156.4)	17.1 (74.4)	8.3 (98.9)	9.4 (66.6)
$I/\sigma I$	11.8 (1.2)	21.7 (3.4)	14.5 (2.0)	7.6 (1.8)
$CC_{1/2}$ (%)	99.8 (80.2)	100 (93.9)	99.9 (77.9)	99.5 (67.2)
CC_{anom} (%)		79 (14)		
Completeness (%)	99.9 (99.9)	99.9 (99.8)	98.9 (98.1)	96.6 (96.6)
Unique reflections	22891 (1426)	13242 (2584)	18674 (1221)	18809 (1732)
Redundancy	6.4 (6.6)	43.0 (13.3)	5.5 (5.8)	2.8 (2.9)
Refinement				
Resolution (Å)	47.64-1.90		39.4-1.94	52.55-2.42
No. reflections	21742		17728	17932
R_{work}/ R_{free} (%)	23.2/28.5		19.0/25.4	21.8/28.6
No. atoms				
Protein	1763		1849	3388
Ligand	0		4	4
Water and ions	119		172	62
Lipids	0		34	6
B-factors				
Protein	51.6		34.3	52.1
Ligand	-		27.7	28.2
Water and ions	61.5		42.8	41.0
Lipids	-		57.0	62.3
R.m.s deviations				
Bond lengths (Å)	0.007		0.008	0.006
Bond angles (°)	1.0		1.1	0.9
Ramachandran analysis				
Favored (%)	97		98	97
Allowed (%)	3		2	3
Outliers (%)	0		0	0

*Highest resolution shell is shown in parenthesis.

Movie S1. Mechanism of transmembrane signaling by sensor histidine kinase NarQ.

Movie S2. Comparison of symmetric and asymmetric ligand-bound NarQ structures.

4. Conclusions

In the first chapter of this thesis manuscript a methodology for the analysis of protein crystal X-ray mesh scans and the program *MeshBest* based on the methodology are presented. *MeshBest* determines the regions of the sample belonging to diffraction from any distinct individual crystal and as well regions where superposition of crystals in the X-ray beam leads to multi-pattern diffraction. The method developed facilitates postmesh scan diffraction data collection procedures by providing a crystal map in the mesh scan plane showing estimates of sizes and shapes of crystals which are ranged according to their diffraction quality. This information allows the potentially automatic tuning of experiment parameters in order to obtain the best data quality.

In Chapter 2, based on the 'positive-inside' rule, the usefulness of halide-SAD in the solutions of the crystal structures of membrane proteins is assessed. I-SAD and Br-SAD protocols were applied to crystals of four different MP targets. Although, Br-SAD did not provide satisfactory structure solution with the given data, the structure solution via I-SAD was successful and straightforward in all four cases. The diversity in the target structure along with reports of other MP crystal structures solved by I-SAD supports the idea of the potential universality of this method which, moreover, is easy in implement, fast and non-toxic compared to the use of other potential heavy atom labels, in *de novo* MP crystal structure solution.

In the Chapter 3 structural studies of the HK of *E.coli* are presented. The crystal structures of a truncated construct of HK NarQ in *holo*-state in two crystal forms and in the *apo*-state acquired by the mutation were elucidated. The structures both show the domain conformations of a HK and provide much needed insight into the mechanisms of TM

signalling by sensor HK. The study defines the role of the HAMP domain as a signal amplifier and by its transformation of TM helix displacements into rotation and scissoring movements. The results show how for a bacterial sensor extracellular cues can be transformed into a cascade of conformational rearrangements on the cytoplasmic side of the cell. This knowledge could provide the basis for the new TCS target-based antibacterial drug design.

List of references

Abendroth, J., Gardberg, A. S., Robinson, J. I., Christensen, J. S., Staker, B. L., Myler, P. J., Stewart, L. J. & Edwards, T. E. (2011). *Journal of Structural and Functional Genomics*. **12**, 83–95.

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Crystallographica Section D Biological Crystallography*. **52**, 30–42.

Affandi, T., Issaian, A. V. & McEvoy, M. M. (2016). *Biochemistry*. **55**, 5296–5306.

Airola, M. V., Watts, K. J., Bilwes, A. M. & Crane, B. R. (2010). *Structure*. **18**, 436–448.

Aishima, J., Owen, R. L., Axford, D., Shepherd, E., Winter, G., Levik, K., Gibbons, P., Ashton, A. & Evans, G. (2010). *Acta Crystallographica Section D Biological Crystallography*. **66**, 1032–1035.

Andrey, P., Lavault, B., Cipriani, F. & Maurin, Y. (2004). *Journal of Applied Crystallography*. **37**, 265–269.

Aravind, L. & Ponting, C. P. (1997). *Trends in Biochemical Sciences*. **22**, 458–459.

Aravind, L. & Ponting, C. P. (1999). *FEMS Microbiology Letters*. **176**, 111–116.

Banerjee, S., Bartesaghi, A., Merk, A., Rao, P., Bulfer, S. L., Yan, Y., Green, N., Mroczkowski, B., Neitz, R. J., Wipf, P., Falconieri, V., Deshaies, R. J., Milne, J. L. S., Huryn, D., Arkin, M. & Subramaniam, S. (2016). *Science*. **351**, 871–875.

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J. L. S. & Subramaniam, S. (2015). *Science*. **348**, 1147–1151.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Crystallographica Section D Biological Crystallography*. **67**, 271–281.

Batyuk, A., Galli, L., Ishchenko, A., Han, G. W., Gati, C., Popov, P. A., Lee, M.-Y., Stauch, B., White, T. A., Barty, A., Aquila, A., Hunter, M. S., Liang, M., Boutet, S., Pu, M., Liu, Z. -j., Nelson, G., James, D., Li, C., Zhao, Y., Spence, J. C. H., Liu, W., Fromme, P., Katritch, V., Weierstall, U., Stevens, R. C. & Cherezov, V. (2016). *Science Advances*. **2**, e1600292–e1600292.

Berman, H., Henrick, K. & Nakamura, H. (2003). *Nature Structural Biology*. **10**, 980–980.

Bhate, M. P., Molnar, K. S., Goulian, M. & DeGrado, W. F. (2015). *Structure*. **23**, 981–994.

Bourenkov, G. P. & Popov, A. N. (2006). *Acta Crystallographica Section D Biological Crystallography*. **62**, 58–64.

Bourenkov, G. P. & Popov, A. N. (2010). *Acta Crystallographica Section D Biological Crystallography*. **66**, 409–419.

Bowler, M. W., Guijarro, M., Petitdemange, S., Baker, I., Svensson, O., Burghammer, M., Mueller-Dieckmann, C., Gordon, E. J., Flot, D., McSweeney, S. M. & Leonard, G. A. (2010). *Acta Crystallographica Section D Biological Crystallography*. **66**, 855–864.

Brockhauser, S., Di Michiel, M., McGeehan, J. E., McCarthy, A. A. & Ravelli, R. B. G. (2008). *Journal of Applied Crystallography*. **41**, 1057–1066.

Brockhauser, S., Ravelli, R. B. G. & McCarthy, A. A. (2013). *Acta Crystallographica Section D Biological Crystallography*. **69**, 1241–1251.

Buckler, D. R., Zhou, Y. & Stock, A. M. (2002). *Structure*. **10**, 153–164.

Butler, S. L. & Falke, J. J. (1998). *Biochemistry*. **37**, 10746–10756.

Caffrey, M. (2003). *Journal of Structural Biology*. **142**, 108–132.

Caffrey, M. (2009). *Annual Review of Biophysics*. **38**, 29–51.

Callaway, E. (2015). *Nature*. **525**, 172–174.

Cherezov, V., Clogston, J., Papiz, M. Z. & Caffrey, M. (2006). *Journal of Molecular Biology*. **357**, 1605–1618.

Cherezov, V., Hanson, M. A., Griffith, M. T., Hilgart, M. C., Sanishvili, R., Nagarajan, V., Stepanov, S., Fischetti, R. F., Kuhn, P. & Stevens, R. C. (2009). *Journal of The Royal Society Interface*. **6**, S587–S597.

Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). *Science*. **318**, 1258–1265.

Cheung, J., Bingman, C. A., Reyngold, M., Hendrickson, W. A. & Waldburger, C. D. (2008). *Journal of Biological Chemistry*. **283**, 13762–13770.

Cheung, J. & Hendrickson, W. A. (2008). *Journal of Biological Chemistry*. **283**, 30256–30265.

Cheung, J. & Hendrickson, W. A. (2009). *Structure*. **17**, 190–201.

Chiang, R. C., Cavicchioli, R. & Gunsalus, R. P. (1992). *Molecular Microbiology*. **6**, 1913–1923.

Civjan, N. R., Bayburt, T. H., Schuler, M. A. & Sligar, S. G. (2003). *BioTechniques*. **35**, 556–560, 562–563.

Cromer, D. T. & Liberman, D. (1970). *The Journal of Chemical Physics*. **53**, 1891–1898.

Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Crystallographica Section D Biological Crystallography*. **56**, 232–237.

Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1984). *Journal of Molecular Biology*. **180**, 385–398.

Duss, O., Yulikov, M., Allain, F. H. T. & Jeschke, G. (2015). *Methods in Enzymology*, Vol. *558*, pp. 279–331. Elsevier.

Egger, L. A., Park, H. & Inouye, M. (1997). *Genes to Cells*. **2**, 167–184.

Ferris, H. U., Coles, M., Lupas, A. N. & Hartmann, M. D. (2014). *Journal of Structural Biology*. **186**, 376–379.

Ferris, H. U., Dunin-Horkawicz, S., Mondéjar, L. G., Hulko, M., Hantke, K., Martin, J., Schultz, J. E., Zeth, K., Lupas, A. N. & Coles, M. (2011). *Structure*. **19**, 378–385.

Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S. & Mitchell, E. (2010). *Journal of Synchrotron Radiation*. **17**, 107–118.

Forst, S., Delgado, J. & Inouye, M. (1989). *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6052–6056.

Francis, S., Wilke, K. E., Brown, D. E. & Carlson, E. E. (2013). *Med. Chem. Commun.* **4**, 269–277.

Gabadinho, J., Beteva, A., Guijarro, M., Rey-Bakaikoa, V., Spruce, D., Bowler, M. W., Brockhauser, S., Flot, D., Gordon, E. J., Hall, D. R., Lavault, B., McCarthy, A. A., McCarthy, J., Mitchell, E., Monaco, S., Mueller-Dieckmann, C., Nurizzo, D., Ravelli, R. B. G., Thibault, X., Walsh, M. A., Leonard, G. A. & McSweeney, S. M. (2010). *Journal of Synchrotron Radiation*. **17**, 700–707.

García-Pelayo, R. (2005). *Journal of Physics A: Mathematical and General*. **38**, 3475–3482.

Giordano, R., Leal, R. M. F., Bourenkov, G. P., McSweeney, S. & Popov, A. N. (2012). *Acta Crystallographica Section D Biological Crystallography*. **68**, 649–658.

Gouet, P., Fabry, B., Guillet, V., Birck, C., Mourey, L., Kahn, D. & Samama, J.-P. (1999). *Structure*. **7**, 1517–1526.

Gu, Y., Li, H., Dong, H., Zeng, Y., Zhang, Z., Paterson, N. G., Stansfeld, P. J., Wang, Z., Zhang, Y., Wang, W. & Dong, C. (2016). *Nature*. **531**, 64–69.

Gushchin, I., Gordeliy, V. & Grudinin, S. (2013). *PLoS ONE*. **8**, e66917.

Gushchin, I., Melnikov, I., Polovinkin, V., Ishchenko, A., Yuzhakova, A., Buslaev, P., Bourenkov, G., Grudinin, S., Round, E., Balandin, T., Borshchevskiy, V., Willbold, D., Leonard, G. A., Büldt, G., Popov, A. & Gordeliy, V. (2017). *Science*.

Haupert, L. M. & Simpson, G. J. (2011). *Methods*. **55**, 379–386.

Heijne, G. V. (1994). *Annual Review of Biophysics and Biomolecular Structure*. **23**, 167–192.

von Heijne, G. (1992). *Journal of Molecular Biology*. **225**, 487–494.

Heimann, J. D. (2002). *Advances in Microbial Physiology*, Vol. *46*, pp. 47–110. Elsevier.

Henderson, R. & Unwin, P. N. T. (1975). *Nature*. **257**, 28–32.

Hendrickson, W. A. (2016). *Nature Structural & Molecular Biology*. **23**, 464–467.

Henrich, B., Bergamaschi, A., Broennimann, C., Dinapoli, R., Eikenberry, E. F., Johnson, I., Kobas, M., Kraft, P., Mozzanica, A. & Schmitt, B. (2009). *Nuclear Instruments and*

Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. **607**, 247–249.

Herrera, C. M., Crofts, A. A., Henderson, J. C., Pingali, S. C., Davies, B. W. & Trent, M. S. (2014). *mBio*. **5**, e02283-14.

Heyda, J., Hrobárik, T. & Jungwirth, P. (2009). *The Journal of Physical Chemistry A*. **113**, 1969–1975.

Hulko, M., Berndt, F., Gruber, M., Linder, J. U., Truffault, V., Schultz, A., Martin, J., Schultz, J. E., Lupas, A. N. & Coles, M. (2006). *Cell*. **126**, 929–940.

Hülsen, G., Broennimann, C., Eikenberry, E. F. & Wagner, A. (2006). *Journal of Applied Crystallography*. **39**, 550–557.

Kabsch, W. (2010). *Acta Crystallographica Section D Biological Crystallography*. **66**, 125–132.

Karain, W. I., Bourenkov, G. P., Blume, H. & Bartunik, H. D. (2002). *Acta Crystallographica Section D Biological Crystallography*. **58**, 1519–1522.

Karle, J. & Hauptman, H. (1956). *Acta Crystallographica*. **9**, 635–651.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). *Nature*. **181**, 662–666.

Kissick, D. J., Dettmar, C. M., Becker, M., Mulichak, A. M., Cherezov, V., Ginell, S. L., Battaile, K. P., Keefe, L. J., Fischetti, R. F. & Simpson, G. J. (2013). *Acta Crystallographica Section D Biological Crystallography*. **69**, 843–851.

Krell, T., Lacal, J., Busch, A., Silva-Jiménez, H., Guazzaroni, M.-E. & Ramos, J. L. (2010). *Annual Review of Microbiology*. **64**, 539–559.

Kuhlbrandt, W. (2014). *Science*. **343**, 1443–1444.

Landau, E. M. & Rosenbusch, J. P. (1996). *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14532–14535.

Langel, Ü., Cravatt, B., Graslund, A., von Heijne, G., Zorko, M., Land, T. & Niessen, S. (2010). Introduction to peptides and proteins Boca Raton, Fla.: CRC Press.

Lavault, B., Ravelli, R. B. G. & Cipriani, F. (2006). *Acta Crystallographica Section D Biological Crystallography*. **62**, 1348–1357.

Lund, M., Vácha, R. & Jungwirth, P. (2008). *Langmuir*. **24**, 3387–3391.

Lunin, V. Y. (1988). *Acta Crystallographica Section A Foundations of Crystallography*. **44**, 144–150.

MacPherson, A. (1982). Preparation and analysis of protein crystals New York: Wiley.

Madden, J. T., DeWalt, E. L. & Simpson, G. J. (2011). *Acta Crystallographica Section D Biological Crystallography*. **67**, 839–846.

Madden, J. T., Toth, S. J., Dettmar, C. M., Newman, J. A., Oglesbee, R. A., Hedderich, H. G., Everly, R. M., Becker, M., Ronau, J. A., Buchanan, S. K., Cherezov, V., Morrow, M. E., Xu, S., Ferguson, D., Makarov, O., Das, C., Fischetti, R. & Simpson, G. J. (2013). *Journal of Synchrotron Radiation*. **20**, 531–540.

Mascher, T. (2013). *Current Opinion in Microbiology*. **16**, 148–155.

Mascher, T. (2014). *Trends in Microbiology*. **22**, 559–565.

Maslennikov, I., Klammt, C., Hwang, E., Kefala, G., Okamura, M., Esquivies, L., Mors, K., Glaubitz, C., Kwiatkowski, W., Jeon, Y. H. & Choe, S. (2010). *Proceedings of the National Academy of Sciences*. **107**, 10902–10907.

Melnikov, I., Polovinkin, V., Kovalev, K., Gushchin, I., Shevtsov, M., Shevchenko, V., Mishin, A., Alekseev, A., Rodriguez-Valera, F., Borshchevskiy, V., Cherezov, V., Leonard, G. A., Gordeliy, V. & Popov, A. (2017). *Science Advances*. **3**, e1602952.

Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M. I., Pragani, R., Boxer, M. B., Earl, L. A., Milne, J. L. S. & Subramaniam, S. (2016). *Cell*. **165**, 1698–1707.

Merkling, P. J., Ayala, R., Martínez, J. M., Pappalardo, R. R. & Sánchez Marcos, E. (2003). *The Journal of Chemical Physics*. **119**, 6647–6654.

Michel, H. (1983). *Trends in Biochemical Sciences*. **8**, 56–59.

Mike, L. A., Choby, J. E., Brinkman, P. R., Olive, L. Q., Dutter, B. F., Ivan, S. J., Gibbs, C. M., Sulikowski, G. A., Stauff, D. L. & Skaar, E. P. (2014). *PLoS Pathogens*. **10**, e1004044.

Mizuno, T. (1997). *DNA Research*. **4**, 161–168.

Molnar, K. S., Bonomi, M., Pellarin, R., Clinthorne, G. D., Gonzalez, G., Goldberg, S. D., Goulian, M., Sali, A. & DeGrado, W. F. (2014). *Structure*. **22**, 1239–1251.

Moore, J. O. & Hendrickson, W. A. (2012). *Structure*. **20**, 729–741.

Morth, J. P., Sørensen, T. L.-M. & Nissen, P. (2006). *Acta Crystallographica Section D Biological Crystallography*. **62**, 877–882.

Moukhametzianov, R., Klare, J. P., Efremov, R., Baeken, C., Göppner, A., Labahn, J., Engelhard, M., Büldt, G. & Gordeliy, V. I. (2006). *Nature*. **440**, 115–119.

Mueller-Dieckmann, C., Bowler, M. W., Carpentier, P., Flot, D., McCarthy, A. A., Nanao, M. H., Nurizzo, D., Pernot, P., Popov, A., Round, A., Royant, A., de Sanctis, D., von Stetten, D. & Leonard, G. A. (2015). *The European Physical Journal Plus*. **130**,.

Neiditch, M. B., Federle, M. J., Pompeani, A. J., Kelly, R. C., Swem, D. L., Jeffrey, P. D., Bassler, B. L. & Hughson, F. M. (2006). *Cell*. **126**, 1095–1108.

Newman, J. A., Zhang, S., Sullivan, S. Z., Dow, X. Y., Becker, M., Sheedlo, M. J., Stepanov, S., Carlsen, M. S., Everly, R. M., Das, C., Fischetti, R. F. & Simpson, G. J. (2016). *Journal of Synchrotron Radiation*. **23**, 959–965.

Nikolaev, M., Round, E., Gushchin, I., Polovinkin, V., Balandin, T., Kuzmichev, P., Shevchenko, V., Borshchevskiy, V., Kuklin, A., Round, A., Bernhard, F., Willbold, D., Büldt, G. & Gordeliy, V. (2017). *Crystal Growth & Design*. **17**, 945–948.

Nitahara, S., Maeki, M., Yamaguchi, H., Yamashita, K., Miyazaki, M. & Maeda, H. (2012). *The Analyst*. **137**, 5730.

Noriega, C. E., Lin, H.-Y., Chen, L.-L., Williams, S. B. & Stewart, V. (2010). *Molecular Microbiology*. **75**, 394–412.

Nurizzo, D., Mairs, T., Guijarro, M., Rey, V., Meyer, J., Fajardo, P., Chavanne, J., Biasci, J.-C., McSweeney, S. & Mitchell, E. (2006). *Journal of Synchrotron Radiation*. **13**, 227–238.

Oesterhelt, D. & Stoeckenius, W. (1971). *Nature New Biology*. **233**, 149–152.

Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. (2006). *Nature Reviews Drug Discovery*. **5**, 993–996.

Pakula, A. A. & Simon, M. I. (1992). *Proceedings of the National Academy of Sciences*. **89**, 4144–4148.

Parkinson, J. S. (2010). *Annual Review of Microbiology*. **64**, 101–122.

Parkinson, J. S. & Kofoid, E. C. (1992). *Annual Review of Genetics*. **26**, 71–112.

Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. & Landau, E. (1997). *Science*. **277**, 1676–1681.

Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. & North, A. C. T. (1960). *Nature*. **185**, 416–422.

Pflugrath, J. W. (1999). *Acta Crystallographica Section D Biological Crystallography*. **55**, 1718–1725.

Pham, V. T., Tavernelli, I., Milne, C. J., van der Veen, R. M., D'Angelo, P., Bressler, C. & Chergui, M. (2010). *Chemical Physics*. **371**, 24–29.

Podgornaia, A. I. & Laub, M. T. (2013). *Current Opinion in Microbiology*. **16**, 156–162.

Pothineni, S. B., Strutz, T. & Lamzin, V. S. (2006). *Acta Crystallographica Section D Biological Crystallography*. **62**, 1358–1368.

Rabin, R. S. & Stewart, V. (1992). *Proceedings of the National Academy of Sciences*. **89**, 8419–8423.

Rabin, R. S. & Stewart, V. (1993). *Journal of Bacteriology*. **175**, 3259–3268.

Raivio, T. L. & Silhavy, T. J. (2001). *Annual Review of Microbiology*. **55**, 591–624.

Rossmann, M. G. (1990). *Acta Crystallographica Section A Foundations of Crystallography*. **46**, 73–82.

de Sanctis, D., Beteva, A., Caserotto, H., Dobias, F., Gabadinho, J., Giraud, T., Gobbo, A., Guijarro, M., Lentini, M., Lavault, B., Mairs, T., McSweeney, S., Petitdemange, S., Rey-

Bakaikoa, V., Surr, J., Theveneau, P., Leonard, G. A. & Mueller-Dieckmann, C. (2012). *Journal of Synchrotron Radiation*. **19**, 455–461.

de Sanctis, D., Oscarsson, M., Popov, A., Svensson, O. & Leonard, G. (2016). *Acta Crystallographica Section D Structural Biology*. **72**, 413–420.

Sayre, D. (1952). *Acta Crystallographica*. **5**, 60–65.

Scholkmann, F., Boss, J. & Wolf, M. (2012). *Algorithms*. **5**, 588–603.

Seddon, A. M., Curnow, P. & Booth, P. J. (2004). *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1666**, 105–117.

Sevvana, M., Vijayan, V., Zweckstetter, M., Reinelt, S., Madden, D. R., Herbst-Irmer, R., Sheldrick, G. M., Bott, M., Griesinger, C. & Becker, S. (2008). *Journal of Molecular Biology*. **377**, 512–523.

Sheldrick, G. M. (1990). *Acta Crystallographica Section A Foundations of Crystallography*. **46**, 467–473.

Sheldrick, G. M. (2002). *Zeitschrift Für Kristallographie - Crystalline Materials*. **217**,.

Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G. & Laub, M. T. (2005). *PLoS Biology*. **3**, e334.

Sneath, P. H. A. & Sokal, R. R. (1962). *Nature*. **193**, 855–860.

Snell, G., Cork, C., Nordmeyer, R., Cornell, E., Meigs, G., Yegian, D., Jaklevic, J., Jin, J., Stevens, R. C. & Earnest, T. (2004). *Structure*. **12**, 537–545.

Solà, M., Gomis-Rüth, F. X., Serrano, L., González, A. & Coll, M. (1999). *Journal of Molecular Biology*. **285**, 675–687.

Song, Y., Peisach, D., Pioszak, A. A., Xu, Z. & Ninfa, A. J. (2004). *Biochemistry*. **43**, 6670–6678.

von Stetten, D., Giraud, T., Carpentier, P., Sever, F., Terrien, M., Dobias, F., Juers, D. H., Flot, D., Mueller-Dieckmann, C., Leonard, G. A., de Sanctis, D. & Royant, A. (2015). *Acta Crystallographica Section D Biological Crystallography*. **71**, 15–26.

Stewart, V. (2003). *Biochemical Society Transactions*. **31**, 1–10.

Stewart, V. (2014). *Molecular Microbiology*. **91**, 853–857.

Stewart, V. & Parales, J. (1988). *Journal of Bacteriology*. **170**, 1589–1597.

Stock, A. M., Robinson, V. L. & Goudreau, P. N. (2000). *Annual Review of Biochemistry*. **69**, 183–215.

Storn, R. & Price, K. (1997). *Journal of Global Optimization*. **11**, 341–359.

Subramaniam, S., Kühlbrandt, W. & Henderson, R. (2016). *IUCrJ*. **3**, 3–7.

Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D. & Bowler, M. W. (2015). *Acta Crystallographica Section D Biological Crystallography*. **71**, 1757–1767.

Taylor, B. L. & Zhulin, I. B. (1999). *Microbiol. Mol. Biol. Rev.* **63**, 479–506.

Terwilliger, T. C. (1999). *Acta Crystallographica Section D Biological Crystallography*. **55**, 1863–1871.

Theveneau, P., Baker, R., Barrett, R., Beteva, A., Bowler, M. W., Carpentier, P., Caserotto, H., Sanctis, D. de, Dobias, F., Flot, D., Guijarro, M., Giraud, T., Lentini, M., Leonard, G. A., Mattenet, M., McCarthy, A. A., McSweeney, S. M., Morawe, C., Nanao, M., Nurizzo, D., Ohlsson, S., Pernot, P., Popov, A. N., Round, A., Royant, A., Schmid, W., Snigirev, A., Surr, J. & Mueller-Dieckmann, C. (2013). *Journal of Physics: Conference Series*. **425**, 012001.

Vernede, X., Lavault, B., Ohana, J., Nurizzo, D., Joly, J., Jacquamet, L., Felisaz, F., Cipriani, F. & Bourgeois, D. (2006). *Acta Crystallographica Section D Biological Crystallography*. **62**, 253–261.

Volz, K. & Matsumura, P. (1991). *J. Biol. Chem.* **266**, 15511–15519.

vonHeijne, G. (1989). *Nature*. **341**, 456–458.

Walden, H. (2010). *Acta Crystallographica Section D Biological Crystallography*. **66**, 352–357.

Walker, J. N., Crosby, H. A., Spaulding, A. R., Salgado-Pabón, W., Malone, C. L., Rosenthal, C. B., Schlievert, P. M., Boyd, J. M. & Horswill, A. R. (2013). *PLoS Pathogens*. **9**, e1003819.

Wallin, E. & Heijne, G. V. (2008). *Protein Science*. **7**, 1029–1038.

Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. (2005). *Angewandte Chemie International Edition*. **44**, 7342–7372.

Wang, B., Zhao, A., Novick, R. P. & Muir, T. W. (2014). *Molecular Cell*. **53**, 929–940.

Wang, B.-C. (1985). *Methods in Enzymology*, Vol. *115*, pp. 90–112. Elsevier.

Ward, S. M., Bormans, A. F. & Manson, M. D. (2006). *Journal of Bacteriology*. **188**, 3944–3951.

Ward, S. M., Delgado, A., Gunsalus, R. P. & Manson, M. D. (2002). *Molecular Microbiology*. **44**, 709–719.

Warren, A. J., Armour, W., Axford, D., Basham, M., Connolley, T., Hall, D. R., Horrell, S., McAuley, K. E., Mykhaylyk, V., Wagner, A. & Evans, G. (2013). *Acta Crystallographica Section D Biological Crystallography*. **69**, 1252–1259.

Watts, K. J., Ma, Q., Johnson, M. S. & Taylor, B. L. (2004). *Journal of Bacteriology*. **186**, 7440–7449.

White, T. A., Kirian, R. A., Martin, A. V., Aquila, A., Nass, K., Barty, A. & Chapman, H. N. (2012). *Journal of Applied Crystallography*. **45**, 335–341.

Willett, J. W. & Crosson, S. (2017). *Molecular Microbiology*. **103**, 197–202.

Winkler, K., Schultz, A. & Schultz, J. E. (2012). *Journal of Biological Chemistry*. **287**, 15479–15488.

Wojdyla, J. A., Panepucci, E., Martiel, I., Ebner, S., Huang, C.-Y., Caffrey, M., Bunk, O. & Wang, M. (2016). *Journal of Applied Crystallography*. **49**, 944–952.

Xie, W., Dickson, C., Kwiatkowski, W. & Choe, S. (2010). *Protein & Peptide Letters*. **17**, 1383–1391.

Yamamoto, K., Hirao, K., Oshima, T., Aiba, H., Utsumi, R. & Ishihama, A. (2005). *Journal of Biological Chemistry*. **280**, 1448–1456.

Yau, W.-M., Wimley, W. C., Gawrisch, K. & White, S. H. (1998). *Biochemistry*. **37**, 14713–14718.

Zander, U., Bourenkov, G., Popov, A. N., de Sanctis, D., Svensson, O., McCarthy, A. A., Round, E., Gordeliy, V., Mueller-Dieckmann, C. & Leonard, G. A. (2015). *Acta Crystallographica Section D Biological Crystallography*. **71**, 2328–2343.

Zeldin, O. B., Gerstel, M. & Garman, E. F. (2013). *Journal of Applied Crystallography*. **46**, 1225–1230.

Zhang, Z., Sauter, N. K., van den Bedem, H., Snell, G. & Deacon, A. M. (2006). *Journal of Applied Crystallography*. **39**, 112–119.

Zhou, Y. (2001). *Genes & Development*. **15**, 627–637.

Zhu, Y. & Inouye, M. (2003). *Journal of Biological Chemistry*. **278**, 22812–22819.

Zhu, Y. & Inouye, M. (2004). *Journal of Biological Chemistry*. **279**, 48152–48158.

Zschiedrich, C. P., Keidel, V. & Szurmant, H. (2016). *Journal of Molecular Biology*. **428**, 3752–3775.