Histidine and Tyrosine-based Heme-binding Motifs for the Prediction of Heme-Regulated Proteins

Dissertation

zur Erlangung des Doktorgrades (*Dr. rer. nat.*)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Benjamin Franz Schmalohr, geb. Syllwasschy

aus
Bad Honnef am Rhein

Bonn, Februar 2021

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

"Das Edelste, was am Menschen ist, das ist Blut, wenn es gut will. Aber das Ärgste, das am Menschen ist, das ist Blut, wenn es übel will."

Eckhart von Hochheim

"The noblest thing that is in man is blood when it wants well. But the most wicked thing that is in man is blood when it wants evil."

Eckhart von Hochheim

Abstract

The versatile molecule heme (iron protoporphyrin IX) fulfils numerous vital functions as a part of hemoproteins, such as hemoglobin and cytochromes, in which it is essential for oxygen transport, electron transport, and detoxification. Under certain conditions, it can be released from hemoproteins and then regulate cellular processes, but also exert toxic effects. In recent years, significant progress has been made towards the understanding of this regulatory heme. Surface-exposed sequence stretches were found to play a crucial role and with the cysteine-proline dipeptide, the first heme-regulatory motif (HRM) was identified. Histidine and tyrosine were also frequently identified in heme-regulated proteins and heme-binding peptides, but a distinct histidine/tyrosine (H/Y)-based motif has not been discovered yet.

In this thesis, H/Y-based motifs were analyzed systematically. For this purpose, four subclasses of heme-binding peptides (A-D) were established from all possible combinations of histidine and tyrosine, and divided according to spacer length (0-3 amino acids). Over 50 model peptides were synthesized and analyzed in depth by ultraviolet-visible (UV/Vis), resonance Raman, and NMR spectroscopy. It was found that motifs with spacer lengths of 1 (subclass B) and 3 (subclass D) exhibited the strongest heme-binding affinities and most binders were found in these classes. Structural studies revealed that these classes occupy mixed conformational states of penta- and hexacoordination and two NMR structures were solved. Overall, the motifs HXH, HXXXY, and HXXXH were found to be the most promising H/Y-based heme-binding motifs. These findings were combined with those of earlier studies and implemented into a web application called HeMoQuest. This tool allows users to predict HRMs from protein sequence and features a machine learning algorithm, which was trained with experimental peptide data. As an example of H/Y-based motifs, two proteins were studied herein. The first protein is Janus kinase 2 (JAK2), which is critical in nascent erythrocytes to propagate growth signals and increase hemoglobin production. Heme was confirmed to activate JAK2 and its corresponding downstream signaling in the K562 cell line. Furthermore, a YXH and a cysteine-proline (CP) motif were suggested as heme-binding sites in the catalytically active Janus homology 1 (JH1) domain. The second protein, Toll-like receptor 4 (TLR4), was found to be connected to three major heme-related pathologies, i.e. inflammation, thrombosis, and hemolysis. A systematic *in silico* analysis of heme binding to TLR4 was therefore performed with the aid of HeMoQuest and docking experiments. Therein, a suitable HXXXY motif on TLR4 itself and an interesting interaction with the lipopolysaccharide binding pocket was predicted.

The results presented in this thesis show distinct H/Y-based HRMs on the peptide level, which are used to successfully predict protein candidates. The combined knowledge is made available to the scientific community through a web-based algorithm. Better understanding of regulatory heme binding and heme biology may allow for targeted treatment and prevention of heme-related diseases.

Zusammenfassung

Das vielseitige Molekül Häm (Eisenprotoporphyrin IX) erfüllt lebenswichtige Funktionen als Bestandteil zahlreicher Hämoproteine, wie z. B. Hämoglobin und Cytochrome, in denen es unerlässlich für Sauerstofftransport, Elektronentransport und Detoxifikation ist. Unter bestimmten Bedingungen kann es aus Hämoproteinen freigesetzt werden und dann zelluläre Prozesse regulieren, aber auch selbst toxische Wirkungen ausüben. In den letzten Jahren wurden bedeutende Fortschritte hinsichtlich des Verständnisses von diesem regulatorischen Häm gemacht. Man fand heraus, dass oberflächenexponierte Sequenzabschnitte eine entscheidende Rolle spielen und identifizierte mit dem Cystein-Prolin-(CP-)Dipeptid das erste Häm-regulatorische Motiv (HRM). Auch Kombinationen von Histidin und Tyrosin wurden häufig in Häm-regulierten Proteinen und Häm-bindenden Peptiden identifiziert, aber ein eindeutiges Histidin/Tyrosin (H/Y)-basiertes Motiv wurde bisher noch nicht vorgestellt.

In dieser Arbeit wurden H/Y-basierte Motive systematisch analysiert. Dazu wurden aus allen theoretisch möglichen Kombinationen von Histidin und Tyrosin vier Unterklassen von Häm-bindenden Peptiden (A-D) gebildet und entsprechend der Anzahl an „Spacer"-Aminosäuren (0-3) zwischen den koordinierenden Resten unterteilt. Über 50 Modellpeptide wurden synthetisiert und mittels UV/Vis-, Resonanz-Raman- und NMR-Spektroskopie eingehend analysiert. Es zeigte sich, dass Motive mit Spacerlängen von 1 (Unterklasse B) und 3 (Unterklasse D) die stärksten Häm-Bindungsaffinitäten aufwiesen und auch insgesamt die meisten bindenden Peptide enthielten. Strukturelle Untersuchungen zeigten, dass diese Klassen gemischte Konformationszustände aus Penta- und Hexakoordination einnehmen können und zwei NMR-Strukturen wurden gelöst. Insgesamt erwiesen sich die Motive HXH, HXXXY und HXXXH als die vielversprechendsten H/Y-basierten Motive. Diese Ergebnisse wurden mit denen früherer Studien kombiniert und zu einer Webanwendung namens HeMoQuest zusammengefügt. Der dieser Anwendung zugrundeliegende Algorithmus ermöglicht es Benutzern, HRMs mittels maschinellem Lernen aus einer gegebenen Proteinsequenz vorherzusagen. Als Beispiel für H/Y-basierte Motive wurden in dieser Arbeit zwei Proteine intensiver untersucht. Das erste Protein ist die Janus-Kinase 2 (JAK2), die in Präerythrozyten entscheidend für die Weiterleitung von Wachstumssignalen und die Steigerung der Hämoglobinproduktion ist. Es wurde bestätigt, dass Häm in der K562-Zelllinie JAK2 und nachgeschaltete Signale aktiviert. Ein YXH- und ein CP-Motiv wurden als Häm-Bindungsstellen in der katalytisch aktiven Janus homology 1 (JH1)-Domäne vorgeschlagen. Das zweite Protein, der Toll-like Rezeptor 4 (TLR4), wurde bereits mit drei zentralen Häm-bezogenen Pathologien in Verbindung gebracht: Entzündung, Thrombose und Hämolyse. Daher wurde eine systematische *in silico* Analyse der Häm-Bindung an TLR4 mit Hilfe von HeMoQuest und Docking-Experimenten durchgeführt. Dabei wurde ein HXXXY-Motiv auf TLR4 selbst und eine Interaktion mit der Lipopolysaccharid-Bindungstasche vorhergesagt.

Die in dieser Arbeit vorgestellten Ergebnisse zeigen eindeutige H/Y-basierte HRMs auf Peptidebene, die zur erfolgreichen Vorhersage von Proteinkandidaten genutzt werden können. Die kombinierten Erkenntnisse werden der Öffentlichkeit durch einen webbasierten Algorithmus zur Verfügung gestellt. Ein besseres Verständnis der regulatorischen Häm-Bindung und der Häm-Biologie könnte eine gezielte Behandlung und Vorbeugung von Häm-bezogenen Krankheiten ermöglichen.

Contents

1 Introduction

Heme, one of life's most central molecules, continuously keeps each of us alive. In fact, almost all eukaryotes are capable of synthesizing heme, and those who are not depend on a continuous dietary supply[1]. In humans and many other higher organisms, heme represents the crucial part of hemoglobin, in which it directly binds and releases oxygen and it is the substance which gives blood its distinct red color[2]. However, the functions of heme go far beyond the transport of oxygen. As a prosthetic group, it is an integral part of numerous human proteins[3]. In proteins of the cytochrome P450 family, it oxidizes xenobiotics, whereas in sensor proteins, such as soluble guanylyl cyclase (sGC), it binds gaseous ligands and confers a biological response[4,5]. Furthermore, it is involved in the transport of electrons within the respiratory chain and its catalytic properties can be exploited for the removal or the generation of toxic H_2O_2 in peroxidases and the immune defense enzyme catalase[6,7]. As vital as heme is to the body, it can be detrimental if it is not bound to proteins. Defective erythrocytes can release hemoglobin and in turn labile heme, which can wreak havoc by generating reactive oxygen species (ROS), intercalating into membranes, or modifying proteins, lipids, and deoxyribonucleic acid (DNA)[8–10]. In fact, the human body has developed its own heme scavenging system, consisting chiefly of hemopexin, human serum albumin (HSA), and haptoglobin (Hp) which remove labile heme and hemoglobin, respectively, in order to prevent these deleterious effects[11–13].

Intriguingly, there is a further quality to labile heme: it is a regulator of biological processes. In this context, the term regulatory heme has been coined, which is entirely different from permanent heme binding seen in hemoglobin and other hemoproteins[14]. The physiological and pathological functions conferred by regulatory heme are numerous and have been subject to vivid scientific discussion[14–17]. The first human heme-regulated protein was discovered in δ-aminolevulinic acid synthase 1 (ALAS1), which performs the rate-limiting step of heme biosynthesis[18]. Later studies expanded the regulatory properties to ion channels, mRNA processing, and regulation of the circadian rhythm[19–23]. By now, over 40 proteins have been reported to transiently bind heme and in many cases, to be regulated by it[24]. Notably, it has also been linked to inflammatory processes, either via binding to proteins of the immune system, such as complement component 1q (C1q) or interleukin 36α (IL-36α), or by activating the immune response as a damage-associated molecular pattern (DAMP)[9,25,26]. Heme interaction with proteins utilizes short, surface exposed sequence stretches, HRMs[27]. The first motif to be identified was the CP dipeptide[28]. The last decade of research has built up a body of characteristics, which mark suitable HRMs[27,29–32]. Based on these characteristics, it was possible to successfully predict previously unknown heme-regulated proteins[26,33,34]. Interestingly, many of the identified proteins were not solely dependent on the CP motif, but instead contained H/Y-based HRMs[32]. Two especially intriguing proteins in this regard were JAK2 and Toll-like receptor 4 (TLR4). In this thesis, efforts are undertaken to pinpoint a distinct H/Y-based motif, to investigate heme-binding motifs in proteins and to make the combined knowledge from previous studies publicly available by creating an HRM prediction algorithm.

2 Theoretical Background

2.1 *Heme – Structure and functions of an omnipresent molecule*

Heme has constantly been a subject of scientific investigation throughout the last two centuries. It was first discovered in the early 19th century, when Friedrich Hünefeld first described hemoglobin in 1840, and subsequently Anton Teichmann-Stawiarski isolated heme from it in 1853[35,36]. William Küster was the first to propose a structure for heme in 1912, while Hans Fischer succeeded in its total synthesis, for which he was honored with the Nobel Prize in 1930[37,38]. Subsequently, Max Perutz and John Kendrew were able to solve the three-dimensional structure of hemoglobin through crystallographic analysis and were awarded the second heme-related Nobel Prize in 1962[39]. With these findings the interest in heme declined to some extent, since it was thought that relevant amounts of heme only occur in hemoproteins[14]. In the 1990s, it was discovered that heme also acts as regulator of protein function, which has sparked a new subfield of heme research devoted to the investigation of regulatory heme[14,15,18,28]. This subfield has recently made tremendous advances through the elucidation of the first heme transporter in humans, heme-responsive gene 1 protein homolog (HRG1), and the structural characteristics of regulatory heme binding[1,17,27].

A glimpse of the relevance of heme in the human body can be obtained by considering some of the involved numbers. The majority of heme can be found in the approximately 25 trillion (2.5×10^{13}) red blood cells per human, each of which contains 250-300 million ($2.5\text{-}3 \times 10^8$) hemoglobin molecules[17,40]. This accounts for a staggering 6.25×10^{21} hemoglobin molecules weighing 670 g[40]. In conclusion, approximately 25 g of heme can be found circulating in the blood, which accounts for 80% of heme in the human body[14]. A further 15% are found in the liver, where the majority of heme is synthesized and where it is part of numerous enzymes, while the remaining 5% are found throughout other tissues[14].

2.1.1 *Structure of heme and its derivatives*

In order to comprehend the diversity of functions performed by heme, it is crucial to appreciate the chemical characteristics of heme. Although "heme" is often used as an umbrella term, there is more than one type of heme. "Hemes" are members of a family of molecules, which share an iron (+2/+3) ion complexed by a porphyrin ring (see Figure **1**). The porphyrin moiety of heme consists of four pyrrole rings (A-D), which are linked via methene bridges to form a large aromatic system[41]. The nitrogen atoms of the four pyrroles act as a tetradentate ligand by forming coordinative bonds with the iron ion. Between different hemes, the substitution pattern of the porphyrin ring is varied. Most common in humans is heme *b*, which is a non-covalently bound prosthetic group, for example in hemoglobin[39]. Heme *b* is based on a protoporphyrin IX scaffold, which has four methyl substituents at positions 2, 7, 12, and 18, two vinyl substituents at positions 3 and 8, as well as two propionate side chains at positions 13 and 17 (see Figure **1**B).

In eukaryotes and prokaryotes, the protoporphyrin IX substitution pattern of heme *b* can be modified to yield other hemes (see Figure **1**)[42]. For example, heme *o* is synthesized by farnesylation of the 3' vinyl group and consequently, heme *a* is biosynthesized by a two-step oxidation of the 18' methyl substituent (see Figure **1**, A and D)[43]. Heme *a* was first isolated from bovine heart muscle and is part of the respiratory chain complex IV in most aerobic organisms (see Figure **1**A)[44,45]. Heme *c* is generated, when the vinyl side chains of heme *b* are covalently attached to the thiol functions of

cysteine residues (see Figure **1**C)[46]. This is typically the case in cytochromes c, where heme is a covalently bound prosthetic group[47]. Numerous further heme variants, such as heme *d*, *s*, *i*, and *m*, have been reported, but play a minor role in humans[42,48].
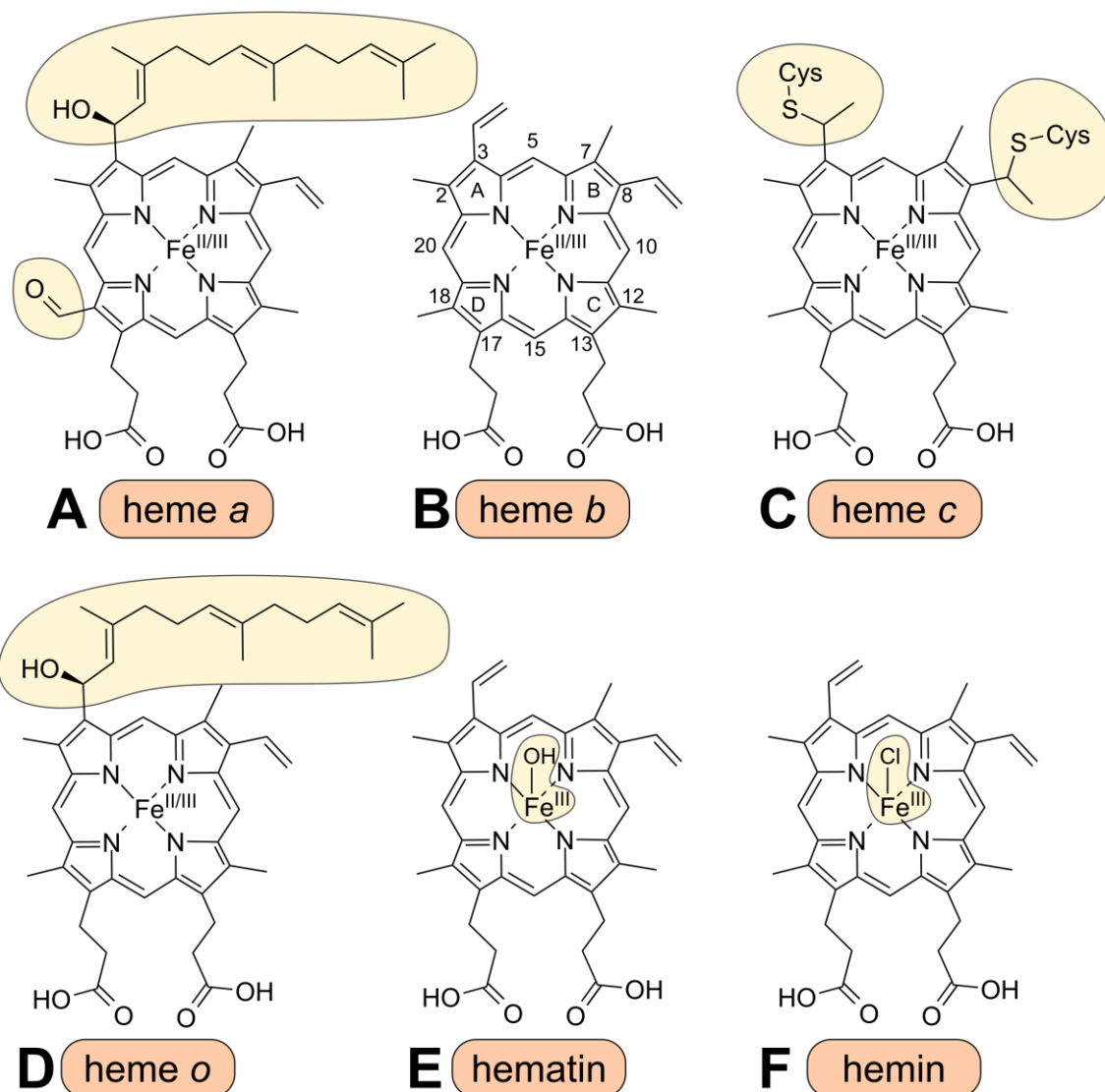
Figure 1: Structure of the most common hemes.
Nomenclature of select carbon atoms (1-20) and the pyrrole rings (A-D) is indicated in heme *b*[49]. Shown are heme *a*, heme *b*, heme *c*, heme *o*, hematin, and hemin (**A-F**, respectively). The difference compared to heme *b* is marked for each structure in yellow. Modified from[50].

  In some cases, the nomenclature of heme *b* is further distinguished by the oxidation state of iron and its fifth ligand. Under physiological conditions, the central iron ion is most frequently in the ferrous (Fe^{2+}) or ferric (Fe^{3+}) oxidation state, but oxidation states up to Fe^{5+} have been reported[51]. When heme is in the ferric state, it can be isolated with a fifth ligand (see Figure **1**, E and F). It is then termed hemin, when it possesses a chloride ion as fifth ligand, or hematin, which binds a hydroxyl group instead[52]. Other physiological ligands of heme, which occur especially in hemoproteins, include electron-rich side chains, such as cysteine, histidine, and tyrosine, or diatomic gases[53].

  In the early twentieth century, it was demonstrated that the total synthesis of heme is principally possible[38]. Nonetheless, the synthesis of heme is cumbersome and of low yield, so that all current applications use heme isolated from porcine or human blood[54,55].

2.1.2   Properties of heme and its behavior in solution

Its unique structure also gives rise to heme's extraordinary properties. Due to the large aromatic ring structure, and the lipophilic methyl and vinyl substituents, heme is a relatively lipophilic molecule and sometimes even described as a metallolipid[56]. The lipophilicity of heme enables it to, specifically or unspecifically, bind to proteins, cellular membranes, and even laboratory glass[57–59]. The central iron ion is crucial for the majority of the various biological functions fulfilled by heme[1]. By virtue of its electronic configuration, it can bind diatomic gas molecules, such as O_2 in hemoglobin or NO in sGC[4,39]. These binding events lead to changes in the porphyrin ring system and the surrounding protein, which in turn provides these proteins with cooperativity or gas-sensing properties[60–62]. In addition, heme is highly redox active and can readily change between its different oxidation states[63]. It is presumably due to this characteristic that heme has continued its evolutionary advance from the time when early organisms employed it in their respiratory chains[64,65]. Today, our mitochondria use different heme oxidation states to transport electrons and to create the proton gradient needed to produce adenosine triphosphate (ATP)[66]. However, this is not the only process in which heme changes its oxidation state. Proteins of the cytochrome P450 family use heme in the metabolism of lipids and steroids, or to detoxify harmful substances or drugs by oxidation, while at the same time reducing cellular reductants, such as Nicotinamide adenine dinucleotide (NAD^+) or flavines[67]. Its redox properties also allow heme to catalyze Fenton-type reactions, in which H_2O_2 is reduced to water, while a second reactant is oxidized[68]. In a pathological context, the second reactant can be DNA, lipids or proteins, which can lead to severe mutations and defects[69–71]. As a further example, catalase also employs heme to reduce reactive oxygen species[72]. Interestingly, the redox potential of heme is also influenced by the pH[73].

It has been a matter of debate, which configurations of heme exist physiologically, both in humans and in other eukaryotes[16,57,74,75]. In humans, it can be assumed that most heme is permanently or transiently bound to proteins, and that high concentrations of extracellular heme are a hallmark of disease[1,17,57,76]. In consequence, the frequently used expression "free heme", which refers to non-hemoprotein bound heme, is slightly misleading because physiologically, heme will preferably be bound to other chemical entities[77]. A more precise term, which should be used instead is "labile heme", referring to heme only loosely attached to other molecules[78]. Organisms, which are confronted with large amounts of heme, such as the malaria parasites of the genus *Plasmodium*, have developed a strategy to avoid heme toxicity: They produce hemozoin (see Figure **2**A)[75]. Heme in their food vacuoles is brought to aggregate and crystallize by acidification yielding innocuous hemozoin crystals[79]. One deprotonated propionate serves as axial ligand of the iron ion, while the second is involved in a hydrogen bond to the next pair of heme molecules (see Figure **2**A)[80]. Very recently, it has been observed that in HRG-1 null mice, which are faced with severe defects in heme recycling, hemozoin can also be produced[81].
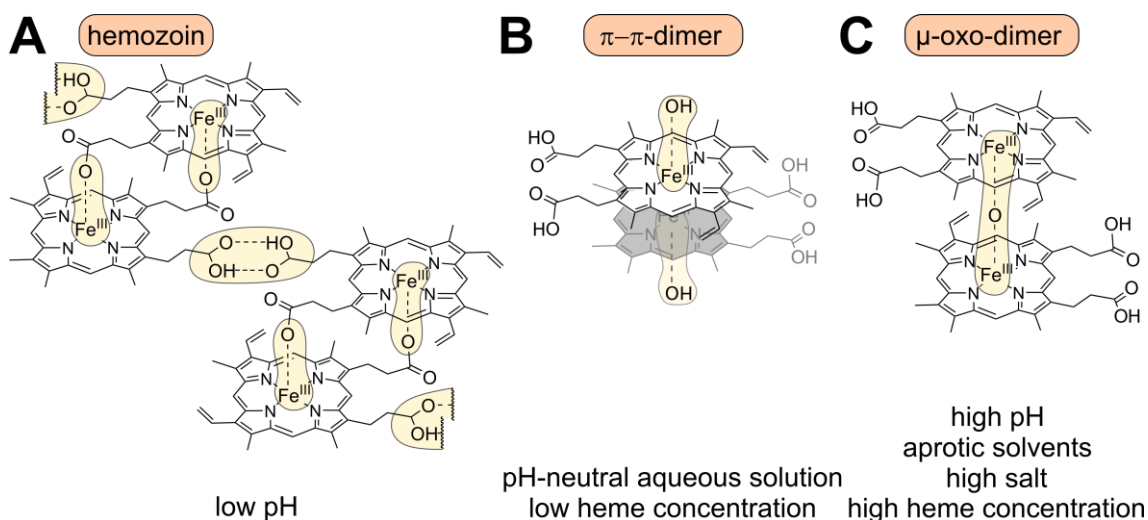
Figure 2: Aggregation states of heme depend on the pH and iron ligand.
A: At acidic pH two hemes are connected by ligation of a propionate side chain to the iron ion and larger crystals are connected by hydrogen bonds, forming hemozoin[79]. **B**: The π-π-dimer can be formed at neutral pH with outwardly facing hydroxo ligands and cannot aggregate further. **C**: The μ-oxo-dimer is found under basic conditions with an O^{2-}-ligand shared by two heme molecules and can form larger aggregates by π-stacking. Modified from[50,59,80,82].

In vitro, it is possible to investigate monomeric heme in aqueous systems, if the correct conditions are employed[83]. In most *in vitro* studies, ferric heme in the form of hemin is employed, because the air oxygen would quickly oxidize ferrous heme[50]. Crystalline hemin is not soluble in water and therefore typically dissolved in either dilute NaOH or neat dimethyl sulfoxide (DMSO)[84]. In dilute solutions, heme will remain in the monomeric hematin form[59]. In concentrated solutions, heme aggregates along three possible routes: At neutral pH, π-π-dimers form, which cannot aggregate further due to their outwards facing hydroxo ligands (see Figure **2**B)[59,82]. At basic pH, in the presence of aprotic solvents (such as DMSO), high heme, or high salt concentrations, μ-oxo-dimers can form, with a single oxygen atom coordinated by two iron ions, which are prone to further aggregation via π-π interactions of the unligated distal faces (see Figure **2**C)[59,85,86]. By acidification of the solvent, hemozoin can be produced *in vitro*[87].

2.1.3 Experimental methods for the investigation of heme interactions

One of the reasons why heme has been identified relatively early is its distinct color[41]. Using UV/Vis spectroscopy, several spectral features in a range from 230-750 nm can be analyzed[88]. The most conspicuous band in the absorbance spectrum of heme is the B-band, or Soret band at λ≈400 nm, named after its discoverer Jaques-Louis Soret[89]. Additionally, there is a set of bands between 450 nm and 700 nm, termed the Q-bands according to the model established by Gouterman and colleagues[88]. These bands arise primarily from π →π* transitions to the first (Q-bands) or second (Soret band) excited state[82,90]. Deconvolution and density-functional theory (DFT) calculations have been performed recently with the heme absorbance spectrum and may prospectively be able to distinguish the heme monomer from the μ-oxo-dimer based on UV/Vis spectra alone[82]. However, all of these bands are responsive to changes in the environment of the chromophore, such as solvent, pH, temperature, oxidation state, and coordination of the iron ion[84,90]. This characteristic can be exploited by titrating heme against its interaction partners, such as heme-binding peptides[29]. If a binding event occurs, the Soret band undergoes a batho- or hypsochromic shift (see Figure **3**)[84]. The change in absorbance can

then be fitted to thermodynamic equations yielding dissociation constants (K_D)[91]. UV/Vis spectroscopy therefore represents a method for rapid and facile assessment on the interaction strength of heme and a ligand, albeit other methods may be more accurate and sensitive (see Figure **3**A)[84].

When a heme-peptide/protein complex is investigated, the complex geometry is often of interest. This can be probed using resonance Raman spectroscopy[93]. In this technique, the heme-containing sample is irradiated with a laser beam of $\lambda \approx 400$ nm[94]. Most photons are scattered with a change in direction but retained energy, but a small fraction of photons transfers energy by inducing vibrations in the heme ring[94]. This inelastic scattering produces photons with a wavelength different from the source[94]. These photons can be recorded and form vibrational bands, which are distinct for different molecules[94]. From changes in the heme-characteristic vibrational bands, structural information on the complex can be deduced[93,94]. The most meaningful band for heme-ligand complex geometries is the v_3-band, which is visible at wavenumbers of $\sim$1491 cm^{-1} in pentacoordinated heme and shifted to higher wavenumbers of $\sim$1505 cm^{-1} for hexacoordinated heme[95]. This band is caused by stretching vibrations of the C_α-C_β-atoms of the porphyrin ring (see Figure **3**B)[93]. Three further bands are of interest for the study of heme complexes: 1: The v_7-band at $\sim$677 cm^{-1}, which presumably correlates with planarity of the porphyrin ring system and disappears in ruffled complexes with strong pentacoordinate ligands[96]. 2: The v_2-band, which is formed by C_β-C_β stretching, and 3: the v_4-band, which responds to changes in the oxidation state of the iron ion[97].

A second, more complex method for structure elucidation is nuclear magnetic resonance (NMR) spectroscopy. It relies on the excitation of the atomic nuclei of a sample using radio waves in a strong magnetic field[98]. After excitation, the relaxation of nuclear spins can be measured as the so called free induction decay (FID)[98]. Consequently, Fourier transformation can be used to convert the FID to characteristic chemical shifts[98]. These chemical shifts can then be assigned to atoms of the investigated compound[99]. Sophisticated two-dimensional experiments can be used to correlate the chemical shifts of two different atoms with each other (see Figure **3**C)[98]. For heme-peptide/protein complexes, the experiments ^{1}H,^{1}H-<u>N</u>uclear <u>O</u>verhauser <u>E</u>nhancement <u>S</u>pectroscop<u>Y</u> (NOESY), ^{1}H,^{1}H-<u>CO</u>rrelated <u>S</u>pectroscop<u>Y</u> (COSY), ^{1}H,^{1}H-<u>TO</u>tal <u>C</u>orrelated <u>S</u>pectroscop<u>Y</u> (TOCSY), and ^{1}H,^{13}C-<u>H</u>eteronuclear <u>S</u>ingle <u>Q</u>uantum <u>C</u>oherence (HSQC) are most suited[30,32,92]. They allow for the so-called sequential walk, in which the resonances of every backbone hydrogen can be assigned incrementally (see Figure **3**C)[30,32,92]. The major limitation of NMR spectroscopy with heme is that it requires nuclei with a non-zero spin value[99]. In the porphyrin ring, all hydrogens and naturally occurring ^{13}C atoms are therefore amenable to NMR. However, the central iron ion is paramagnetic, which leads to broadening of the signals and drastic loss of resolution, often preventing the identification of heme signals within complexes or hemoproteins[100]. Nonetheless, it is possible to overcome these limitations by using the diamagnetic heme analogue Ga(III) protoporphyrin IX or, for larger proteins, by comparing the assignments with and without heme[30,101–103]. In this way it is possible to create structural ensembles from NMR datasets, which show the entire structure of a heme-peptide/protein complex (see Figure **3**D)[30,101–103].
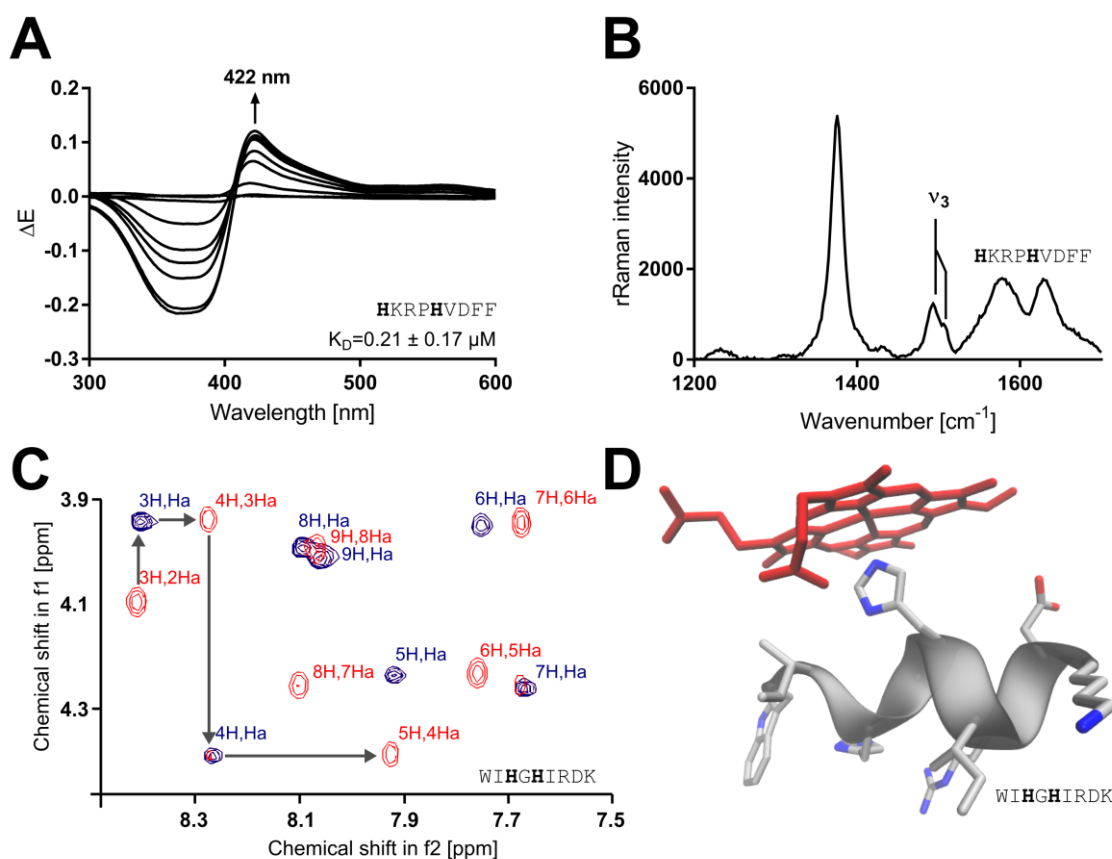
Figure 3: Different methods to study interactions of heme.
A: Representative UV/Vis difference spectrum of the peptide HRKPHVDFF with heme. A shift of the Soret band to 422 nm can be observed, which is indicative of a binding event[92]. **B**: High-wavenumber sector of the resonance Raman spectrum of the same peptide and heme[92]. The split ν_3-band is indicative of a 5c/6c mixture. **C**: Exemplary cross correlation peaks of the NH-Hα region of the peptide WIHGHIRDK in 2D-NMR ^{1}H-^{1}H-NOESY (red) and ^{1}H-^{1}H-TOCSY (blue) experiments. Assigning the resonances is possible via the "sequential walk" shown as grey arrows for the first residues. **D**: Docking pose of the NMR structure of WIHGHIRDK and heme[92].

2.2   Permanent heme binding versus transient heme binding

In eukaryotes, the majority of heme can be found deep within the binding pockets of hemoproteins[3,56]. The most prominent hemoproteins are hemoglobin and myoglobin, but in eukaryotes at least 153 different families of hemoproteins exist[53,104]. Within these proteins, heme is permanently bound and essential for their correct function, which depends on the unique chemical characteristics of heme (2.1)[3]. The binding pockets of hemoproteins are typically characterized by a hydrophobic environment, which complements the hydrophobic porphyrin ring[3,105]. Additionally, the central iron ion of heme is usually coordinated by an electron-rich amino acid residue[53]. Common amino acids include cysteine, histidine, or tyrosine, but also methionine and lysine have been reported[53]. In hemoproteins, the heme group is typically not surface accessible and remains within the protein core for the entire lifetime of the protein (see Figure **4**A)[106,107]. Binding affinities of heme to hemoproteins normally range from 10^{-12}-10^{-15} M (fM – pM range), which reflects that they are *de facto* inseparable from heme under physiological conditions when intact[108]. Hemoproteins are often disordered without heme and only obtain their correct folding after incorporation of heme[107].

There is, however, a functionally and structurally different mode of heme binding to proteins, i.e. transient heme binding[27]. By temporary association to proteins, heme can

act as a regulatory molecule instead of a cofactor[15]. Transient heme binding is structurally distinct from permanent heme binding. The most striking difference is that it occurs on the protein surface via so-called heme-binding motifs (see Figure **4**B)[28]. Additionally, transiently heme-binding proteins are fully structured without heme and undergo changes in the orientation of folded domains, which affects their function[14]. By now, over 40 human proteins have been identified as transiently heme-binding or heme-regulated proteins[24]. In the following paragraphs, an exemplary overview over human heme-regulated proteins involved in the biosynthesis of heme and hemoglobin will be given.
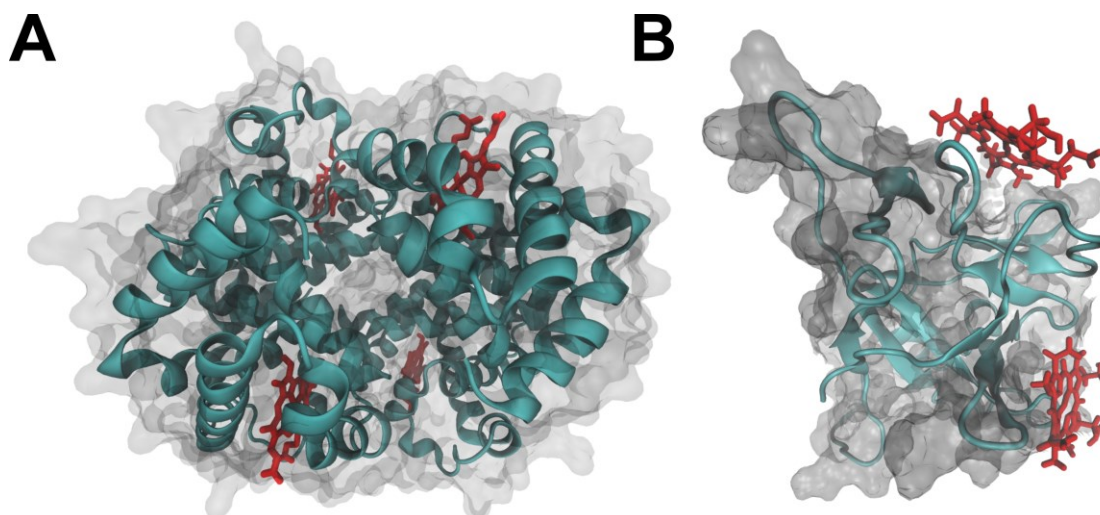

Figure 4: Comparison of permanent and transient heme binding.
A: Hemoglobin (cyan ribbons) as an example of permanent heme binding to proteins (PDB: 5WOG). Heme (red sticks) is bound tightly and buried within a binding pocket. **B**: IL-36α (cyan ribbons) as an example of transient heme binding. Heme is bound to two short sequence stretches on the surface of the protein[103]. Surface representations are drawn in gray. Visualization was performed with YASARA 20.4.24[109].

The first process, in which regulatory heme was identified was its own synthesis[110]. The rate-limiting first step of heme biosynthesis is the condensation of glycine and succinyl-CoA to δ-aminolevulinic acid (ALA)[110]. This step is catalyzed by the enzyme ALAS[28]. Two functionally distinct isoforms of this enzyme exist in man: ALAS1 and δ-aminolevulinic acid synthase 2 (ALAS2)[20]. ALAS1 performs basal heme synthesis in hepatocytes and other non-erythroid cells, whereas ALAS2 is expressed only in erythrocytes and their precursors in order to synthesize large amounts of heme required for hemoglobin synthesis[111]. ALAS1 is negatively heme-regulated on several levels[18]. Direct binding of heme to two sites leads to a reduction of ALAS1 on both, the mRNA and the protein level[20]. This is realized by a reduction in ALAS1 transcription and translation and destabilization of the ALAS1 mRNA[112]. Additionally, the import of the ALAS1 precursor protein into the mitochondria is attenuated by heme[20]. Interestingly, ALAS2 is not directly inhibited by heme, but instead responds to iron[113]. The iron regulatory protein 2 (IRP2) reduces ALAS2 translation when cellular iron pools are low[1]. Upon an increase in iron concentration, IRP2 dissociates from the ALAS2 mRNA and it can be translated[1]. However, IRP2 itself is also regulated by heme, which leads to its ubiquitination and subsequent degradation[114].

It was also found that the consumption and degradation of heme are subject to regulation through heme-regulated transcription factor BTB domain and CNC homolog 1 (Bach1)[115]. It is a transcription repressor, which, in complex with a musculoaponeurotic fibrosarcoma (Maf) protein, inhibits the expression of several target genes, such as heme

oxygenase 1 (HO-1), and the α- and β-chains of hemoglobin[116,117]. At higher heme concentrations, when globin chains are needed for the production of hemoglobin, and HO-1 is required to degrade excess heme, Bach1 is inactivated[115]. The inactivation has been suggested to take place at three levels: 1. Heme can directly bind to the Bach1-Maf dimer, causing it to dissociate from the DNA[115]. 2. Heme binding to Bach1 can lead to its export from the nucleus into the cytosol[118]. 3. Through RanBP-type and C3HC4-type zinc finger-containing protein 1 (RBCK1), Bach1 is ubiquitinated and proteosomally degraded in a heme-dependent manner[119]. The regulation of Bach1 via heme thereby allows cells to respond to changing cellular concentrations of heme by upregulation of heme-removing proteins.

A further protein involved in the heme-regulated expression of globin chains is the heme-regulated inhibitor kinase (HRI)[120]. HRI is able to reduce the translation of globin chains by phosphorylating the eukaryotic initiation factor 2 (eIF2)[120]. Under high heme conditions, HRI binds heme and is inactivated, so that globin biosynthesis can proceed freely[28].

The presented proteins represent a physiological mechanism of adjusting the biosynthesis of hemoglobin to the availability of heme in the cell and to degrade excess heme by the action of HO-1. However, this is only one of several fields in which transient heme binding confers a regulatory role. It has been implicated in regulation of the circadian rhythm through interaction with period circadian clock 2 (Per2), Brain and Muscle ARNT-Like 1 (BMAL1), and nuclear receptor subfamily 1 group D member 1/2 (Rev-ERB α/β)[23,121–125]. It was shown to affect miRNA processing via DiGeorge syndrome critical region 8 (DGCR8) as well as ion channel currents in large conductance calcium-activated potassium channels (Slo1), ATP-sensitive potassium channel (K_{ATP}), and Potassium voltage-gated channel subfamily A member 4 ($K_V1.4$)[21,126–128]. Furthermore, several proteins of the immune system have been shown to be regulated by heme. Notable examples include IL-36α, immunoglobulin G (IgG), C1q and TLR4[25,26,129,130]. Interestingly, it was also reported to interact with several proteins of the coagulation cascade, i.e. Factor VIII, and activated protein C (APC)[131,132]. Despite the large number of heme-regulated proteins that have already been identified, there are likely many more which remain to be discovered[14,15]. In addition, the relevance of heme regulation both in physiological and pathophysiological situations is still a matter of dispute, and remains to be conclusively evaluated.

2.3 Heme-binding motifs are a prerequisite for regulatory heme-protein interactions

It was clearly demonstrated that HRMs are involved in virtually all cases of regulatory heme binding[14]. HRMs represent short, conserved, surface-exposed amino acid sequences, to which heme can bind[14,18]. HRMs were first discovered in 1993 by Lathrop and Timko, who identified the CP-motif in ALAS1[18]. Over the years, numerous other proteins containing CP-motifs were identified and classified as heme-binding[15]. It was firmly established that heme-coordination is the most crucial step of the interaction and performed by the cysteine side chain[20,111,133]. Intriguingly, in some heme-regulated proteins, heme-coordination was performed by histidine or tyrosine instead of cysteine[134–136]. In order to perform a thorough and comprehensive analysis of heme-binding motifs, a peptide-based library screening analysis was performed in 2011[29]. Peptides had previously been identified as suitable models for heme binding and it was shown that up to four amino acids from the coordinating amino acid could have an effect on heme

binding[28]. Consequently, the library was composed around a coordinating amino acid in the P^0-position according to the formula: X$_4$(C/H/Y)X$_4$, with X representing any natural amino acid, except for cysteine and methionine, or norleucine as an isosteric substitute for methionine[29]. Unexpectedly, histidine and tyrosine were more prevalent (each ~40% occurrence) than cysteine (~20% occurrence) in the screening hits[29].

These hits were divided into eight classes of heme-binding peptides (I-VIII, see Figure **5**), according to their coordinating amino acid and additional amino acids[27,31]. The first four classes were assigned to cysteine-based peptides: Cysteine without proline and further coordinating amino acids (I), cysteine without proline but with an additional histidine or tyrosine (II), the CP-dipeptide without further coordinating amino acids (III), and CP plus at least one further histidine or tyrosine (IV)[31]. A further two classes were established for histidine-based peptides containing one (V) or multiple coordinating amino acids (VI), and likewise for tyrosine without (VII) and with additional coordination sites (VIII)[32]. It was furthermore shown that any coordinating amino acid alone or CP alone are not sufficient to confer heme binding, by testing control peptides, such as AAAACAAAA[29–32]. Consequently, the attention was focused on the surrounding amino acids. Following extensive studies demonstrated that the occurrence of multiple coordinating residues, basic amino acids, and aromatic amino acids is beneficial for heme binding, especially in H/Y-based motifs[32]. CP-motifs did not necessarily require further basic amino acids[31]. Determination of the K$_D$ values revealed strong binders throughout all classes, with K$_D$S ranging from $0.24 \pm 0.17\ \mu M$ to $6.36 \pm 2.61\ \mu M$[29–32]. Cysteine motifs often showed low dissociation constants with heme, indicating strong interactions[27]. This was also reflected in peptide models for transient heme binding, in which the rank order of affinities according to the coordinating amino acid was $C > H > Y$[29–32]. Interestingly also peptides from classes VI and VIII, containing multiple coordinating amino acids, showed high affinities[29–32].
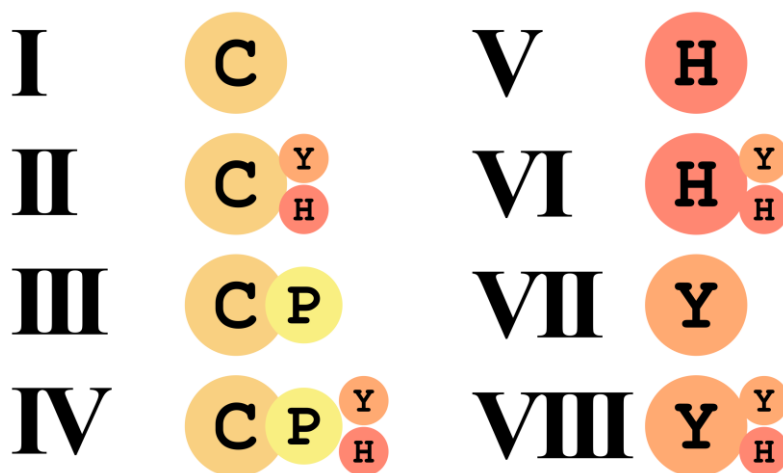
Figure 5: Classes of heme-binding peptides.
As established from a peptide library screening, heme-binding motifs can be assigned to one of eight classes, which are distinct by their coordinating amino acid (large circle) and occurrence of supporting amino acids, such as proline, histidine, and tyrosine (smaller circles)[27,29–32]. Modified from[92].

In addition to the classification according to the occurrence of the different coordinating amino acids (see Figure **5**), the peptides were also grouped into five groups according to their behavior in heme titration difference spectra (see Section 2.3.1) obtained by UV/Vis spectroscopy[29–32]. UV-group 1 exhibited a minimum at ~370 nm and

a maximum at ~420 nm and contained various peptides from classes which possess histidine or tyrosine, i.e. II, and IV-VIII[31,32]. UV-group 2 was found to be typical of the CP-motif containing classes III and IV, but also found in few peptides from classes V-VIII, and was characterized by only one maximum at ~370 nm[31]. Interestingly, UV-group 3 with two maxima at ~370 nm and ~420 nm was mostly populated by peptides from classes I and II, containing cysteine but not CP[31]. UV-groups 4 (single maximum at ~400 nm) and 5 (minimum at ~370 nm) were only found for few peptides and can therefore be considered less common[29–32].

Several of the most interesting hits were structurally analyzed with an array of methods including resonance Raman, NMR, and electron paramagnetic resonance (EPR) spectroscopy[30–32]. It was found that both penta- and hexacoordination can occur in these heme-binding peptides[30–32]. CP-containing peptides were found to be structurally rigid and only induced pentacoordinate heme complexes[30,31]. Several other peptides also adopted pentacoordinate conformations with heme, although less structurally defined[30,31]. Hexacoordinated complexes occurred in two principal arrangements. From classes I and VIII, a sandwich-like complex was identified, in which both faces of one heme molecule were occupied by two peptide molecules[27]. From classes II and VI, on the other hand, clamp-like complexes were found, with a single peptide coordinating one heme molecule on both sides[27]. *In silico* studies established that a distance of at least two amino acids between the two coordinating amino acids is required for this arrangement[27].

In the following section, an overview on the occurrence and characteristics of the known motifs in protein examples is given.

2.3.1 Cysteine-based heme-binding motifs

Of the currently established heme-regulated proteins, many possess CP-motifs as mediators of heme binding[20,28,115,121,137,138]. Cysteine provides the heme-coordination site, while the C-terminal proline increases the binding affinity via structural stabilization[14]. Bach1 possesses six CP motifs, four of which (at positions 428, 464, 495, and 649) were suggested to mediate heme regulation[115]. Interestingly, different CP motifs appear to be involved in different aspects of the regulation. $C^{438}P$ and $C^{464}P$ are required for the heme-dependent nuclear export of Bach1, while $C^{495}P$ and $C^{649}P$ are not[118,139]. ALAS1 also contains three CP motifs, HRI binds heme through two CP motifs, and IRP2 has one CP motif[112,140,141]. As detailed previously (see Section 2.3), especially CP-motifs induce a characteristic shift of the Soret band towards lower wavelengths of around 370 nm. This shift can also been seen in several heme-regulated proteins, such as ALAS2 (371 nm), Bach1 (371 nm), or hPer2 (369 nm)[111,122,139]. A recent report even suggested the use of this characteristic for the *in vitro* detection of labile heme levels[142].

In addition to the CP-motif, there is a second cysteine-based motif, the CXXCH-motif. Canonically, this motif is required for the covalent attachment of heme *c* to c-type cytochromes and other proteins containing heme c^{53}. Slight variations of this motif were also reported for covalent heme-linkage, such as extended motifs $(C(X)_{3-4}CH)^{143,144}$, and incomplete motifs (FXXCH or AXXCH), which attach only one cysteine to the heme moiety[145]. Structurally, the CXXCH motif is predominantly embedded in alpha helices, where the two cysteines point in the same direction and have the same distance as the porphyrin propionates[146]. In most cytochromes, the C-terminal histidine can then act as an axial ligand to the heme iron ion[146]. Strikingly, this motif was also identified in voltage-gated potassium ion BK channel Slo1[19]. It does not provide the basis for covalent attachment of heme to Slo1, but is thought to function as a redox-sensitive switch

regulating ion current[147]. More recently, the CXXCH-motif was identified in iron-regulatory protein 1 (IRP1) and a similar CXXHXH-motif in K_{ATP} channels[127,148]. In conclusion, the CXXCH-motif can in some instances also serve as a heme-regulatory motif.

2.3.2   *Histidine- and tyrosine-based heme-binding motifs*

Despite the prevalence of the CP-motif, histidine and tyrosine can also function as the basis for heme binding. This has long been known for hemoproteins, where histidine as an axial ligand is more frequent than cysteine, for example in hemoglobin[39,53]. Indeed, histidine has been identified in some transiently heme-binding human proteins, often as part of a regulatory interaction, but not always. Bacterial H/Y-based proteins are by far more numerous than human ones, but are of minor relevance to this thesis and have been discussed extensively elsewhere[149]. The following paragraphs give a brief overview of known H/Y-based transient heme-interactions with human proteins.

Non-regulatory transient heme binding is well-established for several human plasma proteins, some of which have been crystallized together with heme[11]. The most well-known example is hemopexin, in which heme is coordinated by the two histidines H^{236} and H^{293} in a hexacoordinate fashion[150]. Hemopexin represents a natural scavenger for toxic heme in the human body[12]. Its heme-binding affinity is the highest among all studied human non-hemoproteins and K_D values were determined as ~5 fM in 1974[151] and, more recently by Detzel et al., as 0.32 ± 0.04 nM[152]. The most abundant protein in human plasma, HSA, has also been found to interact with heme and crystallized with it[153,154]. In the HSA crystal structure, heme coordination is performed by a single tyrosine, Y^{161}, in a pentacoordinated complex, while further amino acids stabilize the complex via salt bridges and hydrophobic interactions[153]. In both proteins one high-affinity interaction site exists, which presumably corresponds to the crystallized structure[150,154]. However, the occurrence of additional, lower-affinity heme-binding sites has been suggested in the literature and is subject of an ongoing discussion in the field[152,153,155].

Regulatory heme binding mediated by histidine or tyrosine has also been shown for a number of human proteins. In 2004, the amyloid beta (Aβ) peptide, which is implicated in the pathogenesis of Alzheimer's disease, was found to display increased peroxidase activity when in complex with heme[135]. The heme-binding site was identified to be located in a sequence stretch containing three histidines and one tyrosine ($RH^6DSGYEVH^{13}H^{14}$), with all three histidines potentially involved in the interaction[156]. A second example can potentially be found in human tryptophanyl-tRNA-synthetase (TrpRS), which was suggested to bind heme through one of its 11 histidine residues[157]. Extensive mutation studies failed to show reduced heme-binding capacity upon H→A mutation of 8 of the 11 histidines[157]. Therefore, binding to the remaining three histidines would be conceivable, but the $C^{62}P$ or the PC^{309} motifs in TrpRS might also have to be considered[157].

An example of regulatory heme binding mediated exclusively through tyrosine is the progesterone receptor membrane component 1 (PGRMC1)[158,159]. This protein was recently crystallized with heme and features a unique regulatory mechanism[158]. Heme binds to the single tyrosine Y^{113} to form a pentacoordinated complex[158]. Consequently, two units of PGRMC1 can form a dimer by π-π-stacking of the two hemes, similar to what is observed in heme solutions[59,158]. Only the dimerized form of PGRMC1 can interact with the epidermal growth factor receptor (EGFR) to provide chemoresistance of

cancer cells[158]. Binding of carbon monoxide as a sixth heme-ligand can inhibit this dimerization and thereby restore sensitivity to cytotoxic agents[158].

Some proteins contain mixtures of different coordinating amino acids. In IL-36α, two heme molecules can bind to $Y^{108}H^{109}$ and $C^{136}P$, respectively[103]. In the circadian locomotor output cycles kaput (CLOCK) protein, which is a key factor in the regulation of the circadian rhythm, a combination of histidine and cysteine has been suggested to mediate hexacoordinate heme binding[160]. Depending on the oxidative state of the heme iron, either a H/H or H/C ligand pair could perform heme coordination[160]. A further protein involved in the circadian rhythm, Rev-erbβ, also binds heme through H/C hexacoordination with a recently re-determined K_D of < 0.1 nM in the oxidized state and a 20-fold reduction of affinity in the reduced state[125,161].

In conclusion, there are several examples of human proteins in which histidine or tyrosine provide at least one of the axial ligands for heme. Interestingly, no distinct heme-regulatory motif based on either histidines or tyrosines has been identified so far. In an extensive screening approach, H/Y-based peptides from classes VI and VIII, in which more than one coordinating amino acid was present, exhibited rather low K_D values[32]. Several high-affinity peptides, however, contained more than two H/Y, such as AA**HYH**TYER (K_D 0.83 ± 0.33 µM), FKQ**YHH**ELI (K_D 0.24 ± 0.17 µM), and KPFK**YD**H**HY** (K_D 1.94 ± 0.72 µM)[32]. Furthermore, in bacterial heme-regulated proteins, combinations of H/Y occur, such as the motif HXH in Staphyloferrin B cluster 9-gene product (SbnI) from *S. aureus* and iron response regulator (Irr) from *R. leguminosarum*[162,163]. The frequent occurrence of H/Y in functionally heme-regulated proteins is an indication that there might be a distinct heme-binding motif, which has not been identified yet[15,27,149].

2.4 Prediction of heme-binding sites in proteins

Since the physiological role of regulatory heme is only now being unraveled, several attempts at the identification of previously unknown heme-regulated proteins have been undertaken. For this purpose, two fundamentally different approaches have been utilized. The first approach was to generate data using an appropriate model system and then, based on this tailor-made data, to deduce characteristics which are shared among heme-binding peptides/proteins. This approach was realized with a custom-made peptide library[29]. Over several years, the peptides and characteristics, which emerged in this screening have been successfully used to identify heme-regulated proteins[30,32,33]. By comparing high-affinity peptides, consensus sequences were derived for each coordinating amino acid, which in turn aided in the identification of previously unknown heme-regulated proteins[29–32].

This approach has the advantage that the data used for the establishment of the consensus sequences is specifically generated from transiently heme-binding peptides. The importance lies in the fundamental differences between transient and permanent heme binding, which have been discussed previously (see Section 0). On the other hand, by analyzing sequence only, key factors, which determine heme binding in the end, might be overlooked and have to be analyzed in tedious manual work[164]. Such factors are glycosylation, disulfide bridges, and other structural features, such as surface exposure[24].

As an entirely separate approach, several computer-assisted tools, based on existing structural data of hemoproteins, have been created in the last decade (see Figure 6)[165–169]. The first of these services were HemeBIND and HemeNET in 2011, which both used the same datasets[165,166]. For the creation of the main dataset, available protein

structures with either heme *b* or heme *c* as prosthetic group were collected[165]. Due to the high redundancy of the obtained structures, the authors compared their sequence identity and applied a cutoff at < 30%[165]. Additionally, a previously generated dataset was used[53]. From the datasets, both structural and sequential features, such as a position-specific scoring matrix, relative accessible surface area, depth index and protrusion index were calculated[165]. Based on these features, a support vector machine was trained, five-fold cross validated and achieved an overall accuracy of 80.7% for hemoproteins[165]. With the subsequent iteration of this algorithm, HemeNET, network-derived features, as well as a dataset consisting of 10 apo-holo hemoprotein pairs, were included[166]. However, HemeNET performed only slightly better than the previous version with an accuracy of 84.6%[166]. In both cases, the prediction did not distinguish between heme *b* and *c*, nor was transient heme binding considered. In addition, both algorithms rely on structural information, which is often not available.
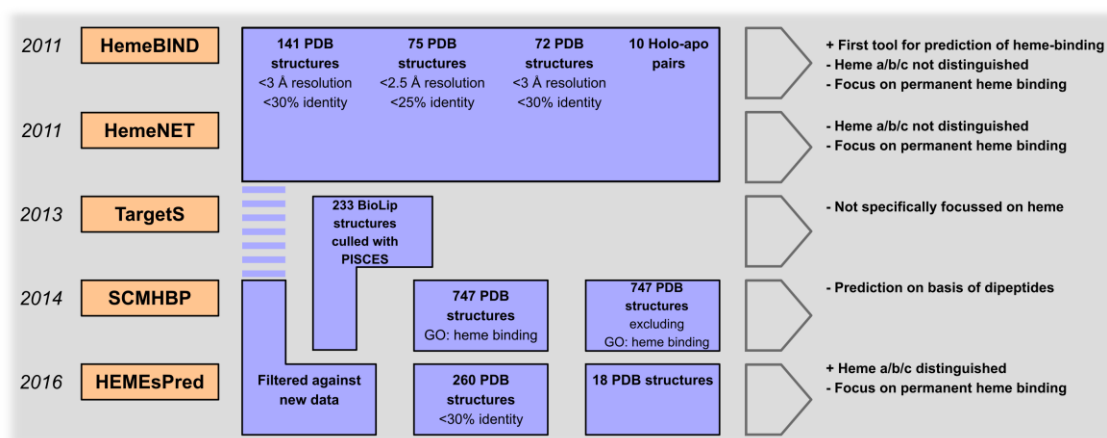

Figure 6: Existing algorithms for the prediction of heme binding.
Several different datasets for the training and testing of the previously established algorithms have been established and reused in subsequent implementations[165–169]. Most datasets were derived from publicly available databases, such as the protein data bank (PDB) and gene ontology (GO). The majority of algorithms did not distinguish between different types of heme and none distinguished permanent and transient heme binding. Modified from[24].

A distinctly different method has been described with SCMHBP, in which dipeptides formed the basis of heme-binding prediction[168]. Therein, all previous datasets were combined and expanded. Subsequently, a scoring card method on the basis of dipeptides was applied to predict heme-binding propensity[168]. As a control group for training, 90,914 putative non-heme-binding proteins were used resulting in a highly unbalanced dataset[168]. Nonetheless, a mean accuracy of 74.2% was reported[168]. Interestingly, aromatic and hydrophobic amino acids emerged as beneficial for heme binding in this study, which is in agreement with the previously mentioned screening studies. However, not all possible coordinating amino acids were pronounced, which might hint at an underrepresentation of the coordinative protein-iron bond. This might have occurred because transient heme binding was not differentiated from permanent heme binding.

The latest algorithm, HEMEsPred, was the first to distinguish between heme *a*, *b*, and *c*[169]. HEMEsPred utilized the largest collection of datasets to date, consisting of over 500 hemoproteins and over 100,000 non-hemoproteins[169]. Using structure and sequence features, the authors were able to identify well-known heme-binding characteristics from the dataset, i.e. the frequent involvement of cysteine and histidine in heme binding as well

as the CXXCH motif[169]. The prediction accuracies ranged from 86.0% to 94.4%, depending on the dataset[169]. This study demonstrates that the differentiation between different heme subtypes is of crucial importance for the prediction of heme-binding events. As an extension of this thought, it can be deduced that the differences between transient and permanent heme binding must be accounted for as well, in order to correctly predict transient heme binding. However, this has not been done by any of the algorithms published so far.

Other, complimentary *in silico* methods, which can be used for hypothesis generation are molecular docking and molecular modeling[170]. These methods can predict binding events based on structural models. In both of these broadly applicable approaches, the starting point is an atomic representation of both the investigated protein and heme. Molecular docking, which is the computationally more economic method, samples several configurations of heme and, optionally, rotamers of the protein[171]. A number of different possible interaction 'poses' are consequently predicted and assessed according to a scoring function[171]. With an increasing number of sample conformers of both the receptor and the ligand, more computing power is needed, but more accurate results are achieved[172]. Since computational resources are always finite, not all possible conformations can be sampled. Therefore, docking is inherently limited and has been criticized for the fact that protein flexibility cannot be fully accounted for[173]. This is especially problematic since many proteins change their conformation after ligand binding, which is described as 'induced fit'[174].

Molecular dynamics simulations try to overcome these limitations by regarding the entirety of the protein, ligand, and surroundings as a composite of individual atoms. The protein is therefore added to a virtual 'soup', together with the ligand, water molecules, and counter ions, such as Na^+, K^+, or Cl^-[175,176]. On all individual atoms, a so-called force field is applied, which is a mathematical model for all possible atomic interactions and forces, such as bond stretching, angle bending, short range Van der Waal's forces, and long range interactions[170]. As a consequence, the entire protein-ligand system remains flexible and can be observed over a time course. Typically, both molecular docking and molecular dynamics can be used in tandem, by using high-scoring docking poses as a starting point for molecular dynamics simulations[177]. In this way, the stability of the docked conformations as well as relative binding energies can be derived[177].

The first method to circumvent this constraint on structural data availability was TargetS, published in 2013[167]. This generalized algorithm was designed to predict 12 different ligands using sequential, structural and ligand-specific features[167]. As training dataset, 233 hemoprotein sequences were used, which exhibited relatively high sequence identity of < 40%[167]. The authors reported an astounding accuracy of 94.4%, however, only on the relatively homogenous dataset of hemoproteins used in the study. A differentiation between permanent and transient heme binding, and between heme *a, b,* or *c* was not made[167]. Despite significant advances in the prediction of heme binding, no algorithm has been established to specifically predict transient heme-binding motifs from sequence data. Such an algorithm would greatly reduce the experimental effort needed to pinpoint transient interaction sites with heme and facilitate the discovery of yet unidentified heme-regulated proteins.

2.5 Regulatory and inflammatory effects of heme on erythrocytes and endothelial cells

As detailed previously (see Section 2.1), the largest share of heme in the human body is found in erythrocytes. Heme and iron are scarce resources for the human body, considering that an estimated 25% of the world population is anemic. Thus, there is a need for efficient heme and iron recycling[178]. Red blood cells only have a limited life-span of around 120 days[10,179]. As a consequence, heme degradation is performed continuously by macrophages to render both heme and iron available again[10,179]. In humans, two heme-degrading enzymes, HO-1 and heme oxygenase 2 (HO-2), oxidize heme to biliverdin, iron, and CO[180]. HO-2 is constitutively expressed, while HO-1 is induced by excess heme as detailed previously (see Section 0)[14].

Erythrocyte precursors take up heme or iron and incorporate it into nascent (hemo)globin molecules, completing this iron-cycle of life[181]. To assure the smooth and reliable handling of heme, a myriad of proteins are engaged in its subjugation[17]. Physiologically, this leads to undetectably low heme concentrations in the plasma[76]. In pathological situations, however, erythrocytes can rupture and heme is released in an uncontrolled manner[11]. For this case, several emergency mechanisms exist to prevent damage, but they are not always sufficient[71]. Consequently, numerous proteins come into contact with heme, bind it, and may be regulated by it[142]. Many of the newly identified heme-regulated proteins are plasma proteins or surface proteins, and are often associated with the immune system[182]. Contradictory roles have been attributed to heme in immune reactions. In some inflammatory conditions of the gastrointestinal tract, heme has been proposed as antiinflammatory agent[183,184]. However, the majority of research points to a proinflammatory effect exerted via the generation of ROS and activation of proinflammatory cytokines[185–187]. A cell type that is especially sensitive to the effects of heme are macrophages, where one of the key molecules is TLR4[130].

In the following paragraphs, the mechanisms regulating heme transport and incorporation into erythrocytes and hemoglobin, respectively, as well as hemolytic diseases and subsequent inflammatory effects of heme on select proteins will be discussed.

2.5.1 Regulatory effect of heme in erythroid precursor cells

Erythrocytes are produced from common myeloid progenitors via several subtypes of erythroblasts and reticuloblasts[188,189]. During this developmental process, vast amounts of hemoglobin are produced so that, in mature erythrocytes, 95% of the total cellular protein is hemoglobin[188]. The generation of hemoglobin requires the concerted ribosomal synthesis of globin and the synthesis or import of heme[189,190]. In early erythroid precursors, this is initiated by the presence of erythropoietin (EPO), stem cell factor (SCF), interleukin 3 (IL-3), and granulocyte macrophage-colony stimulating factor (GM-CSF)[191]. Ongoing signaling is then sustained by EPO throughout the development of erythrocytes[191]. EPO is the natural ligand of the erythropoietin receptor (EPO-R) and causes it to homodimerize[189,192]. Upon dimerization of the receptor, the associated signal transducing enzyme JAK2 can autophosphorylate its critical tyrosine residues Y1007/Y1008 and is activated[193]. Active JAK2 phosphorylates tyrosine residues of EPO-R and thereby several downstream pathways are induced (see Figure 7)[188,194]. Signal transducers and activators of transcription 3 and 5 (STAT3/5) bind to the phosphorylated residues via their Src homology 2 (SH2) domains and are consequently phosphorylated

16

by JAK2 as well[195]. Once phosphorylated, STATs dimerize and translocate to the nucleus, where they initiate the transcription of a plethora of downstream genes, which lead to the cells proliferation, survival, and differentiation[196]. Furthermore, the mitogen-activated protein kinase (MAPK) pathway and the phosphatidylinositol 3-kinase (PI3K) pathway are activated and culminate in the phosphorylation of globin transcription factor 1 (GATA1), which has been termed the "master" transcription factor of erythropoiesis[197]. GATA1 is induced when myeloid stem cells differentiate towards erythroid precursors, reaches peak levels during the pre-erythroblast stage, and declines towards mature erythrocytes[188,197]. GATA1 increases the expression of all proteins needed in mature erythrocytes, most importantly α and β globin chains[198]. By incorporation of heme into globin, holo-hemoglobin can be formed[199].
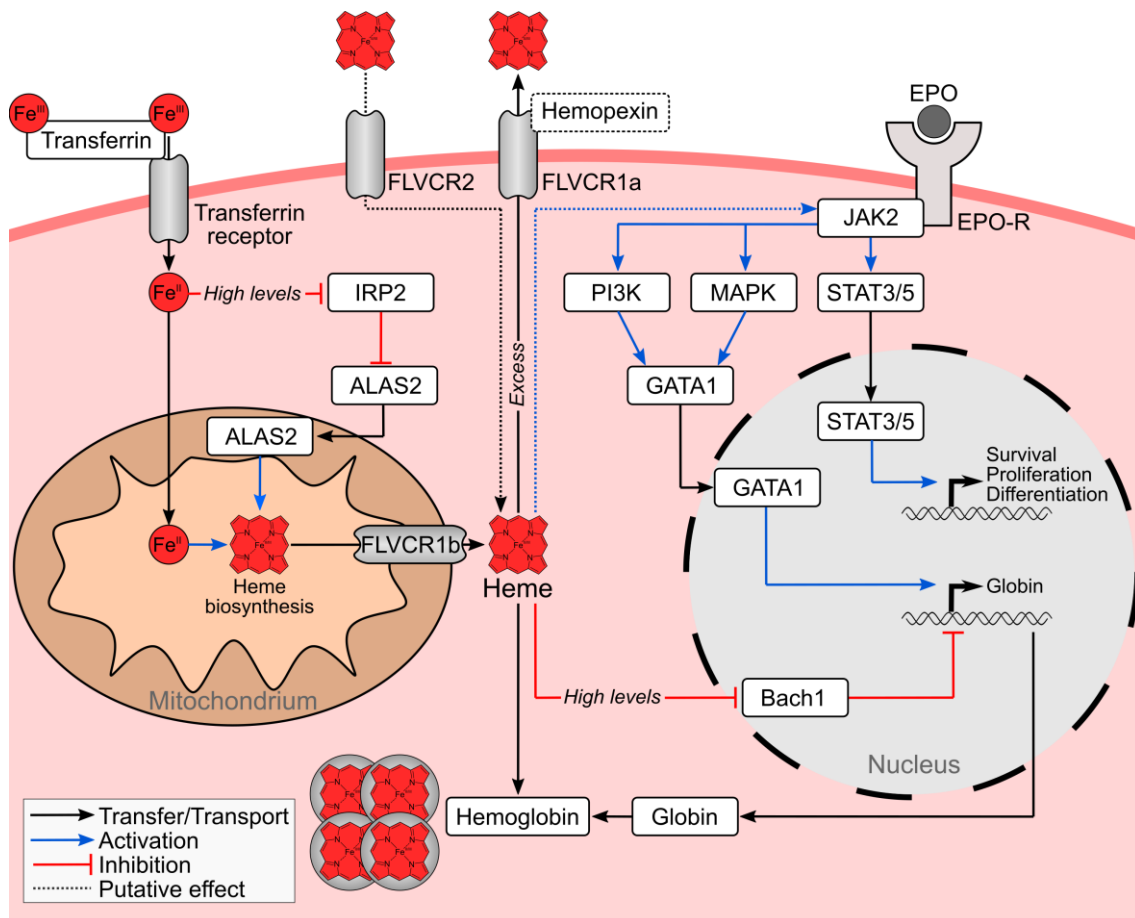

Figure 7: Overview of heme transport and signaling pathways in erythroid precursors and impact of heme thereon.
Shown are the involved proteins, heme, and iron ions with arrows indicating the relationship between them. Black arrows indicate transfer or transport, blue arrows represent activation, red blunted arrows indicate inhibition, dotted lines represent effects that have been suggested in individual reports[71,200,201]. EPO induces EPO-R dimerization and subsequent JAK2 signaling. Through JAK2, downstream signaling of PI3K, MAPK, STAT3, and STAT5 is initiated, which leads to differentiation, proliferation, and increased globin production via GATA1. At the same time, transferrin transports iron ions to the cell, which are utilized for the production of large amounts of heme in the mitochondria. Presence of heme further increases globin synthesis via Bach1 and its own synthesis via IRP2 and ALAS2. Putatively, JAK2 is also activated by heme. Excess heme is exported from mitochondria via feline leukemia virus subgroup C receptor-related protein 1 isoform b (FLVCR1b) and from the cell via feline leukemia virus subgroup C receptor-related protein 1 isoform a (FLVCR1a), which might act cooperatively with hemopexin. Intact heme might be imported via feline leukemia virus subgroup C receptor-related protein 2 (FLVCR2).

17

Interestingly, the synthesis of globin chains is also positively regulated by heme (see Figure **7**) through the transcription repressor Bach1 (see Section 0). This regulation ensures that globin production is balanced with the availability of heme by increasing it, when ample heme is available[116]. Canonically, the availability of heme depends critically on the availability of iron ions *per se*, which can be transported to the erythrocytes by transferrin[202]. As detailed in Section 0, under high iron conditions, IRP2 is inactivated and ALAS2 protein is produced. In addition, IRP2 is also inactivated by high heme concentrations, creating a feed-forward loop (see Figure **7**). The extreme increase of heme synthesis is thought to increase heme levels to such an extent that erythrocytes have to export excess heme via FLVCR1a lest they undergo apoptosis[203]. It was also suggested that hemopexin is required as an external heme carrier for FLVCR1a to function correctly, hinting at a role of hemopexin beyond that of a mere heme scavenger[204]. In addition, some evidence points towards an import mechanism for intact heme utilizing the putative heme importer FLVCR2[200].

In summary, hemoglobin synthesis is initialized primarily by the presence of external EPO signaling and iron availability. It is then sustained and amplified by the IRP2-ALAS2 feed-forward loop for heme synthesis and Bach1 for globin synthesis. Intriguingly, Yao et al. have reported that JAK2 could be activated by heme in HeLa cells[201]. Although no structural analysis of the interaction was performed, this finding implies a highly interesting hypothesis, where potentially presence of heme could sustain globin synthesis. Consequently, it would be of great interest to validate this observation in a directly relevant cell line, such as the erythroleukemia cell line K562[205].

It is well-established that JAK2 gain-of-function mutations, especially V617F, are causal for myeloproliferative disorders of the erythrocyte precursors[206]. Especially in polycythemia vera, more than 90% of the patients carry this single mutation[207]. JAK2 is structured in four domains from N-terminal to C-terminal: The band 4.1, ezrin, radixin, moesin (FERM) domain, the SH2-like (SH2L) domain, the Janus homology-2 (JH2) domain and the JH1 domain (see Figure **8**)[206]. FERM and SH2L are responsible for JAK2 binding to the EPO-R, while JH1 is the kinase domain performing the phosphorylation of Y1007/Y1008 in JAK2 and of activating sites in target proteins[206]. The JH2 domain has also been termed the pseudokinase domain, due to its homology to JH1[208]. However, JH2 autophosphorylates two negative regulatory sites in JAK2 and thereby terminates signaling[209]. The mechanism how JAK2 is hyperactivated by mutations such as V617F is not yet fully understood[206].
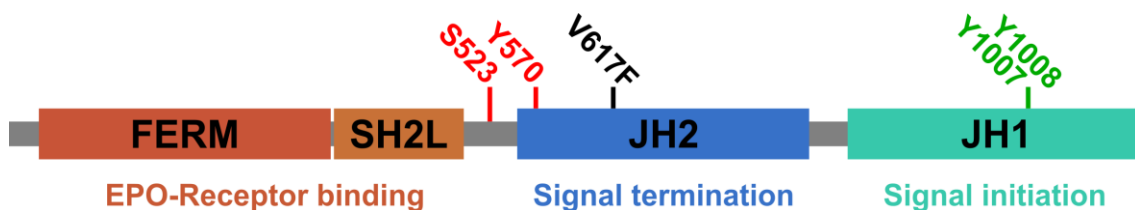
Figure 8: Domain structure of JAK2 and primary domain functions in signaling.
JAK2 consists of four domains: The FERM domain (red) and the SH2L domain (orange) are responsible for receptor binding and presumably remain in contact with JH1 (cyan) in the inactive state[206,210]. JH1, the kinase domain, autophosphorylates Y1007, Y1008, and subsequently downstream targets such as STAT3[206]. JH2 (blue), termed the pseudokinase domain, performs inhibitory autophosphorylation at S523 and Y570 at a slow rate to terminate signal transduction[208]. Modified from[206].

A key element to advance the understanding would be a full-length three-dimensional structure, which has not been published yet. Nonetheless, several domain

structures have been reported and recently, extensive full-length molecular dynamics simulations have hypothesized an interface between JH2 and FERM as well as JH1 and SH2, which may be destabilized by these mutations[210].

Considering hyperproliferative disorders, it would be especially interesting to elucidate the binding site of heme on JAK2, in order to assess the impact of heme in this process, since it is highly concentrated in the affected cells.

2.5.2 Release of heme in hemolytic disorders and the human heme scavenging system

In healthy humans, heme is almost exclusively confined to erythrocytes. In diseased states, however, it can be released in a process called hemolysis[11]. In hemolytic diseases, red blood cells are destroyed and their contents spill into the blood[211]. The reasons for erythrocyte rupture are manifold. In principle, they can be classified into the following categories: 1. Inherent erythrocyte defects, such as in sickle cell disease (SCD) or beta thalassemia[212–214]. These defects destabilize erythrocytes *per se* via mutations in hemoglobin, leading to increased turnover rates and drastically shortened erythrocyte lifespan[212–214]. 2. Pathogen-associated erythrocyte damage, caused for example by *Plasmodium spp.* in Malaria, or bacterial toxins, such as hemolysins from *Staphylococcus aureus*[215–217]. 3. Transfusion-related hemolysis, which can either be caused by storage-induced lysis of packed red blood cells or by blood-type incompatibility[211]. 4. Erythrocyte rupture as a secondary cause of other diseases, which can occur via several different routes[218–221]. Upon erythrocyte destruction, the most abundant substance to be released is hemoglobin. Subsequently, hemoglobin can break down into dimers, then heme is oxidized and finally it dissociates from globin[12,222]. The consequences of such uncontrolled heme release are dire. Patients exposed to heme over extended time periods develop acute chest syndrome and a variety of vascular dysfunctions, such as pulmonary hypertension, leg ulcers, ischemic stroke and systemic vasculopathy[223,224]. Furthermore, one of the primary symptoms is acute kidney injury, which can eventually lead to complete renal failure[225]. Furthermore, heme was implicated in a number of neurological complications, such as intracerebral hemorrhage, subarachnoid hemorrhage, and ischemia reperfusion injury[226–228]. Increased heme levels were also observed in connection with neurodegenerative diseases, such as Alzheimer's disease or Parkinson's disease, for instance by increasing the peroxidase activity of Aβ[135,229]. It is apparent that these pathologies are connected and part of overall dysregulation of heme[8]. Not only heme, but also hemoglobin itself has been associated with oxidative damage and vascular complications, such as nephropathy[222,230,231].

To prevent damage caused by released heme and hemoglobin, the human body has developed a protective system composed of heme scavengers[10,232]. Haptoglobin, a 388 amino acid plasma protein, binds hemoglobin dimers specifically and transports them to macrophages by binding to the cluster of differentiation 163 (CD163) receptor[11,233]. Interestingly, the haptoglobin gene displays polymorphisms, such as duplication and triplication of the gene and three major isoforms, and is highly upregulated in acute hemolysis and inflammation[13]. Upon binding to CD163, the Hp-hemoglobin complex is internalized and degraded[234]. The concentration of Hp in the plasma has been estimated to be 5-34 µM in healthy adults, resulting in a hemoglobin scavenging capacity of 2.5-17 µM, which would result in 10-68 µM heme if it was completely released from the hemoglobin molecules[13]. When this capacity is exceeded, two further proteins intervene: hemopexin and HSA[235]. HSA is the most abundant plasma protein with a concentration of approximately 600 µM and is known as a versatile transport molecule for metal ions,

lipids, bilirubin, and other substances[236,237]. HSA has also been shown to bind heme in a stoichiometry of at least 1:1 and to rapidly transfer it to the specific heme scavenger hemopexin, which in turn exhibits a heme affinity several orders of magnitudes higher than HSA[153,238]. As discussed earlier, each hemopexin molecule binds at least one heme molecule with extremely high affinity and its plasma concentration has been determined to be 12-20 μM[13,76]. Furthermore, a number of other putative plasma heme scavengers have been identified, whose relevance is a matter of ongoing debate, such as $α_1$-microglobulin, low-density lipoprotein, $α_1$-antitrypsin, $α_1$-proteinase inhibitor, and likely others[239–242]. Interestingly, the overall heme scavenging capacity of human plasma has been a matter of discussion and recently been estimated to be as high as 1.8 mM, predominantly attributed to HSA[142]. However, the determination of labile heme concentrations in hemolysis patients has been performed and revealed heme concentrations of more than 300 μM in hemolytic patients[76,227,231]. In these patients, the heme scavenging system has apparently been overwhelmed and was not able to remove all labile heme[76,227,231].

2.5.3 TLR4 as a receptor of heme as a proinflammatory signal in endothelial cells

When excessive heme concentrations do accumulate despite the aforementioned heme clearance mechanisms, they can act on several heme-regulated proteins as outlined in Section 0. Heme has been termed a DAMP (see Section 1) and induces a marked inflammatory response[9,243]. One of the molecules involved in sensing and signal transduction of heme is the receptor TLR4, which will be discussed in detail in the following.

TLR4 belongs to the class of pattern recognition receptors and is the human homolog to the name-giving Toll gene from *Drosophila*[244]. Like all pattern recognition receptors, it is part of the innate immune system and allows cells to recognize pathogenic structures and to respond accordingly[245]. The canonical ligand for TLR4 is the bacterial lipopolysaccharide (LPS), which induces a strong immune response in humans that can even lead to sepsis[246]. TLR4 is a transmembrane protein consisting of 838 amino acids divided into three main domains (see Figure **9**): the extracellular domain (1-624), the transmembrane domain (625-658), and the intracellular Toll/interleukin receptor (TIR) domain (659-839)[247]. In the cellular context, TLR4 can form a 1:1 dimer with myeloid differentiation protein 2 (MD2), the obligatory co-receptor for LPS[247]. The activation of TLR4 results in a highly complex signaling cascade and a subsequent proinflammatory response[248]. Classical activation through LPS initializes with the formation of an MD2-LPS complex, which then binds to TLR4[248]. Two of these heterotrimers dimerize again, so that the vicinity of two TIR domains enable the binding of adaptor proteins and subsequent signal transduction (see Figure **9**)[249]. Prior to MD2 binding, LPS is recognized by the LPS-binding protein and consequently by cluster of differentiation 14 (CD14), which transfers it to MD2[248]. In regard to downstream signaling, two distinct pathways exist: the Myeloid differentiation primary response 88 (MyD88)-dependent pathway, which primarily results in the expression of proinflammatory cytokines, and the MyD88-independent pathway, which results in the expression of type I interferon and its target genes[247,249].
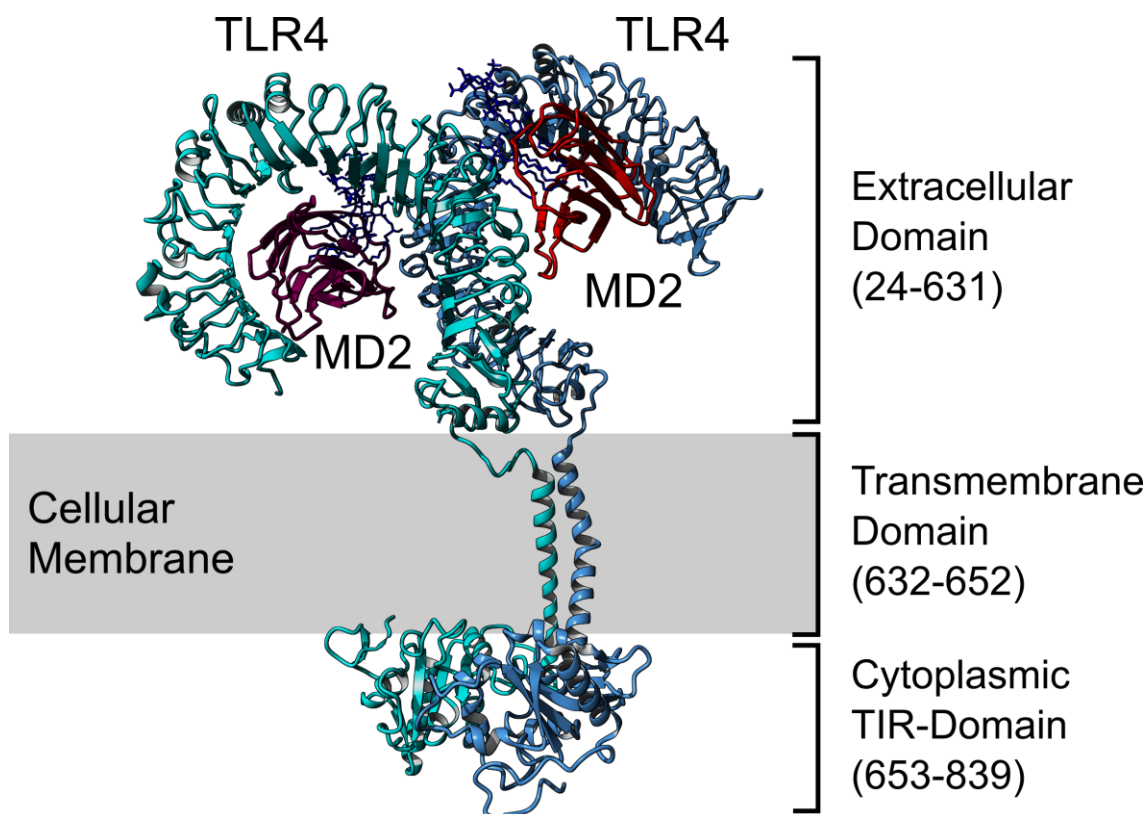
Figure 9: Structure of the (TLR4/MD2/LPS)₂-hexamer.
Depiction of an *in silico* homology model of full-length structure of the TLR4, which is based on crystal structure data (PDB: 3FXI and 2J67)[250]. LPS (dark blue sticks) binds to MD2 (red) and the complex associates with TLR4 (blue and cyan ribbons), which can subsequently dimerize and initiate signal transduction. Domain names are shown on the right with the corresponding amino acid numbers in brackets. Visualization was performed with YASARA 20.4.24[109].

Apart from the well-known LPS-induced pathway, over 30 molecules have been reported to display either agonistic or antagonistic properties on TLR4[247]. Interestingly, heme is among these substances and has been characterized as an activator of TLR4[130]. Heme leads to the secretion of tumor necrosis factor α (TNFα) in mouse and human macrophages[130]. This effect was found to be specific for heme with an iron ion, i.e. neither protoporphyrin IX (PPIX), even if substituted with palladium, nor mesoporphyrin elicited a response[130]. However, the effect seems to be relatively weak, requiring concentrations of 10 μM and higher[130,182]. Recently, an over-additive effect of heme and LPS on TLR4 activation was reported, albeit contradictory effects were found between human and mouse primary macrophages[251]. Furthermore, in mouse endothelial cells, a reduction of the TLR4-mediated LPS response by induction of HO-1 was observed[252]. Even the reverse effect was suggested, namely an effect of extracellular LPS addition on intracellular labile heme levels[251]. Heme-mediated activation of TLR4 was proposed as the major mechanism of heme toxicity to the endothelium, resulting in vaso-occlusion in a mouse model of SCD[253,254]. Hemopexin therapy was shown to reduce this vaso-occlusion in mouse models of SCD and β-thalassemia[255]. Heme toxicity has been termed a "second hit" for the endothelium in hemolytic diseases and the endothelial damage was characterized as cause for renal dysfunction[9]. By other groups, this relevance of TLR4 activation by heme was criticized in regard to the occurrence of sufficiently high heme concentrations in (patho-)physiological conditions[182]. In order to resolve the true nature and pathological importance of heme-induced TLR4 signaling, further studies are

required. Recently, MD2 was proposed to harbor a binding site for heme and binding has been tested by UV/Vis and agarose pull-down experiments[256]. However, these reports were not corroborated yet. Overall, TLR4 remains one of the most interesting human protein candidates for further investigation of the pathological consequences of heme binding.

3   Thesis outline

This thesis aims at a comprehensive and systematic analysis of transient heme-binding motifs (HBMs) and HRMs based on histidine and tyrosine using peptide models in order to better understand potential heme regulation under physiological conditions. Three major challenges are addressed.

The first objective is the identification and characterization of motifs with combinations of histidine and/or tyrosine. The resulting motifs should represent the counterpart of the well-known cysteine motifs CP and CXXCH. The results are to be integrated into the well-established class scheme of heme-binding peptides and critically discussed in regard to their affinity and natural occurrence.

The second objective is the prediction of previously unknown heme-binding motifs in proteins within a publicly available service. Insight gained from previous studies with heme-binding peptides and proteins is to be incorporated into an algorithm, and used to train machine learning methods for motif and K_D prediction.

A third objective is the validation of the first two objectives through the analysis of heme-binding motifs in suitable protein candidates. As prime examples, JAK2 and TLR4 should be investigated. JAK2 as a key factor responsible for orchestration of the synthesis of hemoglobin in erythrocytes and TLR4 as receptor of heme in endothelial cells and macrophages. Both proteins were connected to heme in previous studies, but the location of heme interaction sites is not precisely known for neither of them. By studying these proteins, valuable insight into specific heme-regulated processes should be gained.

These objectives are addressed in four chapters, which are described below.

Chapter I addresses the first objective and encompasses the extension of the classification system for HBMs by four subclasses containing H/Y-based motifs. In order to achieve this goal, new subclasses of HBMs should be established to reflect all possible arrangements of exactly two histidine and/or tyrosine coordinating residues separated by different spacer lengths. Nonapeptide representatives of these subclasses are to be recruited from previously identified library screening hits, a database search, and known protein examples. At least ten representatives per class are to be synthesized via solid-phase peptide synthesis, purified and analytically characterized. Subsequently, each peptide should be assayed for its heme-binding affinity and binding mode. A subset of peptides should be characterized structurally using resonance Raman and 2D-NMR spectroscopy. Furthermore, several peptides are to be analyzed in regard to their peroxidase activity with heme. The results should give insight into the affinity, prevalence, and relevance of H/Y-based motifs and pinpoint distinct motifs, which are suitable for further investigation in proteins.

Chapter II covers the second objective by utilizing the previously generated peptides to train a tailor-made algorithm for the prediction of potential transient heme-binding sites from protein sequences. For the validation of the algorithm an extensive manual collection of literature-known heme-binding proteins and their interaction sites and affinities is to be assembled. On this dataset, and the in-house generated peptide data, the accuracy of the developed algorithm should be validated. Furthermore, the algorithm should be compared to previously published algorithms. The generated algorithm is to be

implemented into a web interface, in order to make it available to the scientific community.

Chapter III explores JAK2 as a prime example of a heme-regulated protein. The known activating effect of heme on JAK2 is to be examined and validated in the erythroleukemia cell line K562. The tools and knowledge generated in the first two chapters should be used to predict potential motifs for heme binding. These motifs are to be analyzed on the peptide level and critically discussed in regard to their localization and possible mechanistic explanations for the observed effect.

Chapter IV contains an analysis of TLR4 and its interaction with heme. Prediction of possible heme-binding sites on the TLR4/MD2 complex should be performed using the results from the first chapter and the algorithm established in the second chapter. These results are to be compared with literature reported interaction sites and corroborated by molecular docking experiments. Putative HBMs should be critically discussed in regard to their suitability and possible routes of experimental validation of the *in silico* data should be explored.

In the concluding section of this thesis, the obtained H/Y-based motifs are put into perspective with previous work on the CP motif and other known heme-regulatory motifs. The limitations of peptide models and predictive algorithms and their applicability on proteins are reviewed. Furthermore, physiological relevance of JAK2 and TLR4 and the impact of putative heme binding to these proteins is analyzed. Further experiments needed to understand the impact of heme on JAK2 and TLR4, and to develop possible treatments for heme-related diseases are discussed.

4   Manuscripts

4.1   Chapter I - High-affinity binding and catalytic activity of His/Tyr-based sequences: Extending heme-regulatory motifs beyond CP.

Authors*

Benjamin Franz Syllwasschy, Maximilian Steve Beck, Ivona Družeta, Marie-Thérèse Hopp, Anuradha Ramoji, Ute Neugebauer, Senada Nozinovic, Dirk Menche, Dieter Willbold, Oliver Ohlenschläger, Toni Kühl, and Diana Imhof

This article was published in:

4.1.1   Introduction

Despite their frequent occurrence in a peptide library screening and in known heme-regulated proteins, no distinct heme-regulatory motif based on histidine and/or tyrosine is known so far. In order to systematically investigate and potentially identify such a motif, the previously established class system for heme-binding proteins (see Section 2.3) is extended by four subclasses containing exactly two histidine and/or tyrosine residues. Peptides from a library screening and example sequences from (putative) heme-binding proteins are chosen to represent each possible combination of $(H/Y)X_{0-3}(H/Y)$. These peptides are probed for heme binding using UV/Vis spectroscopy and dissociation constants are derived. In order to elucidate the complex geometry and structure of chosen complexes, resonance Raman and NMR spectroscopy are performed. In addition, the peroxidase activity of select peptides is analyzed, since it has been suggested as mechanism of pathogenesis for instance in Alzheimer's disease. In conclusion, putative motifs are derived from the data and their potential structure within heme-regulated proteins is discussed.

*Contributions:

BFS, TK, and DI designed the research. BFS, MSB, ID, and MH performed the research with AR, UN, SN, DM, OO, and DW. BFS, MSB, MH, and TK synthesized and analyzed the peptides and performed UV/Vis binding studies with heme. MSB, ID, and TK performed the studies on peroxidase activity. BFS, MH, AR, and UN performed and analyzed data from resonance Raman spectroscopy. BFS, SN, DM, DW, and OO measured free peptides and heme-peptide complexes by NMR spectroscopy and calculated the structures. Docking studies were performed and analyzed by BFS and OO. BFS, MSB, UN, OO, TK, and DI wrote the manuscript. All authors reviewed the results and approved the final version of the manuscript.
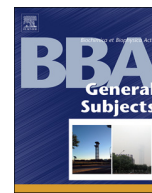
4.1.2   Article

On the following pages, the article is printed in its published form with permission of Elsevier B.V., Amsterdam, Netherlands.

High-affinity binding and catalytic activity of His/Tyr-based sequences: Extending heme-regulatory motifs beyond CP

Benjamin Franz Syllwasschy[a], Maximilian Steve Beck[a], Ivona Družeta[a], Marie-Thérèse Hopp[a], Anuradha Ramoji[b,c], Ute Neugebauer[b,c], Senada Nozinovic[d], Dirk Menche[d], Dieter Willbold[e,f], Oliver Ohlenschläger[g], Toni Kühl[a,*], Diana Imhof[a,*]

[a] Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute, University of Bonn, 53121 Bonn, Germany
[b] Center for Sepsis Control and Care (CSCC), Jena University Hospital, 07747 Jena, Germany
[c] Leibniz Institute of Photonic Technology, Albert-Einstein-Str. 9, 07745 Jena, Germany
[d] Institute for Organic Chemistry and Biochemistry, University of Bonn, 53121 Bonn, Germany
[e] Jülich Research Centre, Institute of Complex Systems – Structural Biochemistry (ICS-6), 52425 Jülich, Germany
[f] Institute of Physical Biology, University of Düsseldorf, 40225 Düsseldorf, Germany
[g] Leibniz Institute on Aging – Fritz Lipmann Institute, 07745 Jena, Germany

ARTICLE INFO

ABSTRACT

Background & motivation: Peptides and proteins can interact with heme through His, Tyr, or Cys in heme-regulatory motifs (HRMs). The Cys-Pro dipeptide is a well investigated HRM, but for His and Tyr such a distinct motif is currently unknown. In addition, many heme-peptide complexes, such as heme-amyloid β, can display a peroxidase-like activity, albeit there is little understanding of how the local primary and secondary coordination environment influences catalytic activity. We thus systematically evaluated a series of His- and Tyr-based peptides to identify sequence features for high-affinity heme binding and their impact on the catalytic activity of heme.
Methods: We employed solid-phase peptide synthesis to produce 58 nonapeptides, which were investigated by UV/vis, resonance Raman, and 2D NMR spectroscopy. A chromogenic assay was used to determine the catalytic activity of the heme-peptide complexes.
Results: Heme-binding affinity and binding mode were found to be dependent on the coordinating amino acid and spacer length between multiple potential coordination sites in a motif. In particular, HXH and HXXXH motifs showed strong heme binding. Analysis of the peroxidase-like activity revealed that some of these peptides and also HXXXY motifs enhance the catalytic activity of heme significantly.
Conclusions: We identify HXH, HXXXH, and HXXXY as potential new HRMs with functional properties. Several peptides displayed a strikingly high peroxidase-like activity.
General significance: The identification of HRMs allows to discover yet unknown heme-regulated proteins, and consequently, enhances our current understanding of pathologies involving labile heme.

1. Introduction

Heme ($Fe^{II/III}$ protoporphyrin IX) is well established as a regulatory molecule which can influence crucial cellular functions. It can regulate and modulate its own biosynthesis (Bach1, ALAS1/2), circadian rhythm (mPer2), and different immune responses on the cellular (IL-36α, JAK2, TLR4) and humoral level (IgG, C1q, C3) [1–9]. In pathological conditions, such as hemolysis or sickle cell disease, labile heme is released in
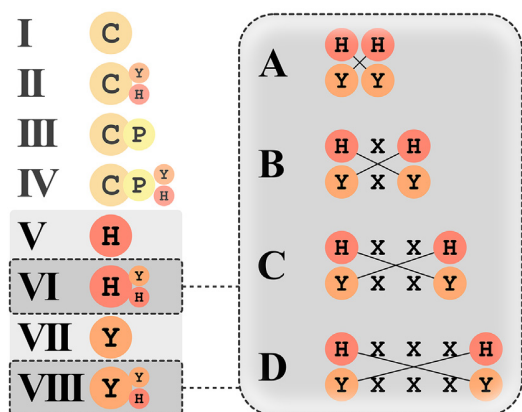
Fig. 1. Conception of subclasses A–D of HRMs.
In previous studies, HRM classes I–VIII were established based on the co-ordination site (C/H/Y at P0). From classes VI and VIII subclasses A–D were derived to cover all possible arrangements of HRMs with two coordinating His and/or Tyr residues differing in spacer length X (X: any amino acid except cysteine and methionine, including norleucine).

large quantities [10]. Recent evidence has also pointed out that there is an intracellular pool of labile heme, which differs in each cellular compartment [11,12].

When associated with proteins, heme is frequently coordinated via histidine, tyrosine, cysteine, or methionine as the initial binding event [13–16]. In case of regulatory heme binding, the coordinating amino acids are surface-exposed and the surrounding residues have an additional strong impact on heme association. The term heme-regulatory or heme-binding motif (HRM/HBM) has been established for such transient heme-binding events [2,13].

A minimal length of 9 amino acids in a peptide sequence has been proven sufficient and suitable for high-affinity heme binding and in particular cases, the development of a peroxidase-like activity in complex with heme [16–21]. In our earlier studies we established a classification system for HRMs based on the coordinating amino acid (Fig. 1), a selection scheme, which allowed for successful prediction and subsequent validation of potential heme-regulated proteins (HRPs) [4,17–20,22], as well as the establishment of a search algorithm for potential HRMs in such proteins [16,23]. The individual HRM classes were divided according to the central coordination site (C/H/Y at P0) as well as specific basic and hydrophobic amino acids in the immediate environment [16–20,24,25]. For cysteine-based HRMs, the Cys-Pro dipeptide motif (CP-motif) is the best known and established determinant for heme binding. It occurs in numerous proteins, such as ALAS1, ALAS2, Irr, HRI, Bach1, mPer2, and IRP2 [2,3,26–30]. However, for proteins containing His or Tyr as the coordination site no such explicit motif has been described so far. Nonetheless, the existence of multiple coordination sites has been identified as a major factor within classes II, IV, VI, and VIII of the aforementioned HRM classification system (Fig. 1). Up to now, only few HRPs are described to bind heme exclusively via histidine and/or tyrosine (e.g., CLOCK, Rev-erbβ, and PGRMC1 [31–37]), and not in conjunction with cysteine (e.g., K_{ATP} channels, CXXH{X_{16}H} [38–41]).

This prompted us to further investigate H/Y-based HRMs to describe them more precisely and possibly disclose motifs as strong heme binders as CP motifs. We earlier found that the strongest heme binders from classes V–VIII contained more than one potential coordination site and thus belonged to classes VI and VIII (Fig. 1) [18]. The respective peptides are structurally diverse and contain up to four heme-coordinating residues with spacer lengths ranging from 0 to 3 [18]. However, no distinct HRM could be identified so far due to the limited number of peptides investigated [17,18].

Binding of heme to H/Y-containing peptides/proteins may induce a

peroxidase-like activity [21,42–48]. The generation of free radicals by such complexes can have severe pathological consequences as has been suggested for amyloid β (Aβ) in Alzheimer's disease [21,42–48]. Aβ can bind heme via a sequence stretch that contains multiple His and Tyr residues, and in complex with heme enhances its intrinsic catalytic activity [21]. The latter is discussed to significantly influence the physiological changes that are observed during the development of Alzheimer's disease (e.g., oxidation of neurotransmitters, functional heme deficiency) [42,43,48–50]. Furthermore, a His-based 23mer peptide derived from fatty acyl-CoA reductase 1 was earlier discovered to display a remarkably high peroxidase-like activity in complex with heme, which was even higher than the Aβ-heme complex [21]. In addition, other H/Y-based non-Aβ-derived nonapeptides in complex were also found to increase the peroxidase-like activity of heme [21].

From the points raised here, the question arises whether or not there exist conserved H/Y-based HRMs beyond CP, which would allow for reliable prediction of both heme-binding sites and peroxidase activity in proteins. Here, the impact of various H/Y combinations as well as effects of other amino acids on heme binding and peroxidase activity was investigated in detail. A total of 58 heme-peptide complexes were analyzed using UV/vis spectroscopy and in part resonance Raman (rRaman) and nuclear magnetic resonance (NMR) spectroscopy. A peroxidase assay was used to explore their functional properties. In summary, the results presented herein shed light on functional and structural prerequisites for heme binding to H- and Y-based motifs as HRMs and suggest distinct motifs similar or superior to CP, and their potential to develop a peroxidase-like activity.

2. Results

2.1. Rationale for peptide selection

It was found in earlier studies that most of the strong heme-binding His- and Tyr-based sequences contained more than one coordinating amino acid [16–18], and were representatives of HRM-classes VI and VIII. However, a distinct H/Y-based motif could not be derived. We thus hypothesized that yet undiscovered HRMs, comprising a combination of two histidine and/or tyrosine residues could exist. Consequently, we subdivided classes VI and VIII taking into account two criteria: i) different combinations of histidine and tyrosine, and ii) different spacer lengths between the two coordinating residues (Fig. 1). Subclass A contained the motifs HH, HY, YH, and YY, subclass B HXH, HXY, YXH, and YXY (X: any amino acid except cysteine and methionine, including norleucine). In a similar fashion subclasses C and D were constructed with spacers X_2 and X_3, respectively (Fig. 1). In total, 58 peptides were investigated in this study: 51 were assessed for heme binding by UV/vis spectroscopy, of which 22 and further 7 from classes V and VII were examined with the peroxidase assay (Tables 1 and 2). 13 peptides recruited for this study originated from a high-throughput screening (HTS) reported earlier [18,19], 7 peptides from known and 10 were predicted in previous studies and by the SeqD-HBM algorithm [16,18,19]. In order to investigate the influence of the position of the potential second coordination site relative to P0, 5 peptides from previous studies were produced as inverted sequences. Finally, a set of control peptides was synthesized, resulting in another 16 peptides.

In addition to the heme-binding study, representatives from each subclass (22 peptides in total) were selected for a TMB-based peroxidase assay [21]. The chosen peptides contained different motifs (e.g., HH, HY, YH, HXH) and possessed varying heme-binding affinities. Furthermore, 7 H/Y-based peptides with only one potential heme-coordination site were also included for comparison (Table 2) [18,51]. With the selection we strived to cover a large variety of peptides to examine possible interdependencies between an enhanced peroxidase activity and other factors, such as coordinating state or UV class (Soret band shift) [17–20].

Table 1
Analytical and functional data of peptides used in this study.

Nr.	(Sub-) class	Motif	Sequence	Origin	K_D [μM]	Soret [nm]	rR	Ref.
	A	**HH/HY/YH/YY**						
control	A	HH	$A_3H/YH/YA_4$	control	n.b.	n.b.	–	
1	A	HH	RQRHHKEFK	BMP6 from *H. sapiens*	n.b.	421	5c/(6c)	
2	A	YH	GGFYHLAAD	HTS	n.sat.	370	–	[18]
3	A	YH	QKRYHEDIF	iNOS from *H. sapiens*	16.12 ± 0.70	417	5c	
4	A	YH	FLFYHSQSG	IL36α from *H. sapiens*	4.48 ± 2.20	415	–	[4]
5	A	YH	TRQYHENIK	NCOR2 from *H. sapiens*	1.88 ± 0.27	418	5c	
	B	**HXH/HXY/YXH/YXY**						
control	B	HXH	$A_2H/YAH/YA_4$	control	n.b.	n.b.	–	
6	B	HXH	BVQLHKHSG	HTS	2.65 ± 0.78	418	5c/(6c)	[18]
6i	B	HXH	GSHKHLQVB	peptide 6 inverted	n.b.	423	5c > 6c	
7	B	HXH	QGHGHQNFP	HTS	n.sat.	419	–	[18]
8	B	HXH	FKAAHKHVR	IsdG from *S. aureus*	0.56 ± 0.23	421	5c < 6c	[18]
8i	B	HXH	RVHKHAAKF	peptide 8 inverted	14.06 ± 0.34	421	5c < 6c	
9	B	HXH	WIHGHIRDK	APC from *H. sapiens*	1.65 ± 0.51	418	5c < 6c	[64]
10	B	HXY	PRQAHVYRA	HTS	5.69 ± 1.35	423	5c	[18]
11	B	YXH	TVBLYNHRI	HTS	n.sat.	362(min)	–	[18]
12	B	YXH	KRYIHRDLA	JAK2 from *H. sapiens*	2.49 ± 0.13	421	5c/(6c)	
13	B	YXH	MNYIHRDLR	Fyn from *H. sapiens*	7.32 ± 0.79	421	5c/6c	
14	B	YXH	RNYIHRDLR	Hck from *H. sapiens*	8.86 ± 0.67	421	5c/(6c)	
15	B	YXH	MNYVHRDLR	Src from *H. sapiens*	4.58 ± 0.51	422	5c/(6c)	
16	B	YXY	FLENYLYER	HTS	n.sat.	424	5c/6c	[19]
	C	**HXXH/HXXY/YXXH/YXXY**						
control	C	HXXH	$AH/YA_2H/YA_4$	Control	n.b.	n.b.	–	
17	C	HXXH	QHLVHFDRR	HTS	n.b.	423	5c/(6c)	
18	C	HXXY	RQRDHQBYA	HTS	3.40 ± 0.70	417	5c	[18]
19	C	YXXH	GFGTYSWHE	HTS	6.39 ± 1.06	415	5c	[18]
20	C	YXXY	WYAAYTKPS	HTS	n.sat.	427	5c/(6c)	[19]
	D	**HXXXH/HXXXY/YXXXH/YXXXY**						
control	D	HXXXH	$H/YA_3H/YA_4$	Control	n.b.	n.b.	–	
21	D	HXXXH	HKRPHVDFF	CTDNEP1 from *H. sapiens*	0.21 ± 0.17	422	5c/6c	
22	D	HXXXH	HRELHGDFV	POMK from *H. sapiens*	8.39 ± 0.50	416	5c	
23	D	HXXXH	HRKEHFEAF	Cat from *E. coli*	n.sat.	417	–	[18]
23i	D	HXXXH	FAEFHEKRH	peptide 23 inverted	1.26 ± 0.51	418	5c < 6c	
24	D	YXXXH	YNLLHBAFD	HTS	n.sat.	423	5c	[19]
25	D	HXXXY	KSVIHNLVY	HrtR from *L. lactis*	n.sat.	421	–	[18]
26	D	HXXXY	VRREHFEFY	CatIII from *E. coli*	n.sat.	419	–	[18]
27	D	HXXXY	AEFRHDSGY	Aβ from *H. sapiens*	n.b.	n.b.	–	
28	D	HXXXY	HPFPYIWKA	HTS	0.33 ± 0.25	423	5c/6c	[19]
28m	VII	AXXXY	APFPYIWKA	peptide 28 mutant	1.94 ± 0.54	422	–	
28i	D	YXXXH	AKWIYPFPH	peptide 28 inverted	16.30 ± 0.89	421	5c < 6c	
29	D	HXXXY	HADTYFGWR	HTS	19.01 ± 1.76	415	5c/(6c)	[18]
29i	D	YXXXH	RWGFYTDAH	peptide 29 inverted	11.47 ± 0.92	418	5c	

51 peptides were investigated using different spectroscopic methods. K_D values > 10 μM were defined as very low and > 20 μM as non-binding. The peptides are organized in subclasses A–D with spacer lengths of 0–3 amino acids and the previously established classes V and VII [18]. K_D values and maximum wavelength in difference spectra were determined using UV/vis spectroscopy. rR, resonance Raman spectroscopy. 5c, pentacoordination; 6c, hexacoordination; Values in brackets represent weak rRaman signals. B, norleucin; HTS, high-throughput screening hits from [19]; n.b., non-binding; n.sat., not saturated; n.sol., not soluble.

2.2. UV/vis studies

51 peptides from subclasses A–D were analyzed for their heme-binding affinity by UV/vis spectroscopy using an earlier established experimental set-up [20].

26 peptides were found to bind heme with a clear concentration-dependent shift of the Soret band and a dissociation constant (K_D) below 10 μM (Table 1). The K_D values ranged from 0.21 ± 0.17 μM to 8.86 ± 0.67 μM. Peptides for which no K_D could be determined (not saturated, n.sat.) were classified as weak binders. 9 peptides showed no binding, which was seen by either a lack of the Soret shift or a $K_D \geq 10$ μM. Subclass B contained both, absolutely and relatively, the largest amount of binders (11 peptides, 78%) and least non-binders (3 peptides, 21%, Fig. 2). Subclasses A and C contained very few peptides in general and no strong binders, while subclasses B and D contained several strong and medium binders (Fig. 2).

In agreement with previous studies [18,20], the controls did not bind heme (Table 1). Surprisingly, several peptides (e.g., **1**, **6i**, **17**) did not bind heme even though they fulfill criteria established earlier, such as basic residues, positive net charge, and multiple coordinating

residues [18,19,21]. From the presented data it can be suggested that in addition to the existence of more than one coordinating amino acid, the distance between the coordinating residues seems to be crucial.

Analysis of the subclasses A–D, according to the combinations of coordinating amino acids (H/Y), revealed striking differences. Generally, the presence of histidine induced lower K_D values and higher binding affinity than tyrosine, which has been found for H/Y-based classes V–VIII as well [18]. All strong binders ($K_D < 1$ μM) contained at least one histidine residue. In addition, for motifs HXH and HXXXH containing two histidine residues one strong binder was found, which was not the case for HXXH and HH. Noteworthy is also the marginal occurrence of peptides with only tyrosine as coordinating amino acid, which mainly originated from HTS. Only peptides **16** (FLENYLYER) and **20** (WYAAYTKPS) containing a YXY and YXXY motif, respectively, were found to be weak heme binders.

Surprisingly, the preference for histidine was diminished in subclasses A and C. HH and HXXH appeared to be unsuited as HBM since none of the peptide representatives was able to bind heme. In addition, no peptide with HY was identified as heme binding. However, several tyrosine-containing peptides with YH and HXXY motifs showed

3

Table 2
Peroxidase activity of investigated peptides.

| Sequence	K_D [μM]	Peroxidase activity [%]
1 coordinating residue		
TVEIHDLFF [18]	weak	218 ± 33
QVRLHWLSP [18]	weak	71 ± 14
VFKEHPAFR [18]	weak	27 ± 8
WELDYFQWK [21]	weak	142 ± 35 [21]
APFPYIWKA(28 m)	mod.	138 ± 46
DNFRYBIPN [18]	weak	88 ± 13
IBFRYSSLK [18]	strong	62 ± 20
2 coordinating residues		
RQRHHKEFK(1)	n.b.	106 ± 36
GGFYHLAAD(2)	weak	104 ± 11
TRQYHENIK(5)	mod.	111 ± 27
FKAAHKHVR(8)	strong	120 ± 39 [21]
RVHKHAAKF(8i)	n.b.	176 ± 21
WIHGHIRDK(9)	mod.	364 ± 20
PRQAHVYRA(10)	weak	89 ± 19
TVBLYNHRI(11)	weak	66 ± 24
KRYIHRDLA(12)	mod.	64 ± 28
FLENYLYER(16)	weak	112 ± 23
RQRDHQBYA(18)	mod.	89 ± 19
GFGTYSWHE(19)	weak	107 ± 22
WYAAYTKPS(20)	weak	113 ± 10
HKRPHVDFF(21)	strong	56 ± 12
HRKEHFEAF(23)	weak	98 ± 6
FAEFHEKRH(23i)	mod.	406 ± 42
YNLLHBAFD(24)	weak	162 ± 29
AEFRHDSGY(27)	n.b.	129 ± 40 [21]
HPFPYIWKA(28)	strong	299 ± 25
AKWIYPFPH(28i)	n.b.	185 ± 25
HADTYFGWR(29)	n.b.	199 ± 27
RWGFYTDAH(29i)	n.b.	181 ± 40
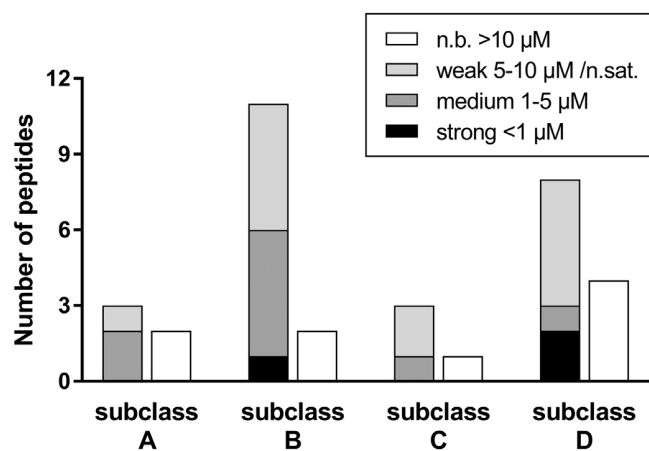
Fig. 2. Heme-binding affinity distribution per subclass.
The K_D values of the investigated peptides (Table 1) were classified as strong (< 1 μM); moderate (1–5 μM); weak (5–10 μM or not saturated); n.b., not binding (> 10 μM). Subclass B and D contain the largest number of medium and strong binders.

moderate binding affinity.

The motif HXH from subclass B emerged as the most promising for a new type of HRM, because it is present in one strong and two moderate binders. Indeed, no other motif investigated herein showed a better record. In addition, only very few non-binders with HXH were identified. Peptides from subclass B with mixed coordinating residues (HXY, YXH) exhibited moderate to weak heme-binding affinities, but surprisingly no non-binding peptides were identified. Peptides **12–15** originate from a set of highly homologous sequences of tyrosine kinases. Comparing peptides **13**, **14**, and **15** it can be concluded that valine instead of isoleucine between the two coordinating amino acids

improved the K_D value, which might be caused by reduced steric hindrance. Exchanging the C-terminal arginine for alanine also resulted in a minor increase in affinity. Peptide **12** (KRYIHRDLA) from JAK2 showed the best heme-binding affinity (K_D = 2.49 ± 0.13 μM) among this set. JAK2 has been reported to bind heme via a CP motif [5], which was confirmed as a heme-binding peptide in a previous study [20]. However, the exact heme-binding site of the protein still remains unknown. In addition, only one candidate with the YXY motif, i.e. peptide **16** (FLENYLYER), bound heme, but saturation was not achieved.

The majority of peptides in subclass C displayed weak or no heme binding. Only one peptide (**18**) displayed moderate affinity.

Subclass D is even more promising with respect to heme binding. Two strong binders, among them peptide **21** (HKRPHVDFF) with the lowest K_D value (0.21 ± 0.17 μM) measured for a heme-binding nonapeptide so far were found in this class. On the other hand, 4 peptides without heme-binding capability belong to this class. This obvious deviation can be explained when taking the involved amino acids into account. While all investigated HXXXH peptides showed heme binding, only approximately half of the H/Y mixed peptides had this feature, and none with the YXXXY motif. In conclusion, HXXXH could be a potential new heme-binding motif.

With regard to the influence on the UV/vis spectrum of heme, almost all peptides induced a similar bathochromic shift of the Soret band of heme. Consequently, the vast majority of the peptides belonged to UV class 1 with absorbance maxima at ~420 nm and minima at ~360 nm in the difference spectra with respect to free heme, as observed earlier (Fig. 3, Fig. S1) [18]. The only exceptions were peptides **2** and **11** which are members of UV class 2 (maximum at ~370 nm) and 5 (minimum at ~ 370 nm), respectively [18].

Analysis of the five inverted peptide sequences revealed that the heme-binding mode is strongly influenced by the sequential arrangement of the amino acids. The inversion of high-affinity peptides such as **6**, **8**, and **28** (to **6i**, **8i**, **28i**) decreased the binding affinity 10–20 fold. On the other hand, the inversion of low-affinity peptides **23** and **29** (to **23i** and **29i**) lead to mixed results. For peptide **23** the resulting inverted sequence **23i** (FAEFHEKRH) showed a low K_D of 1.26 ± 0.51 μM. For peptide **29i** a twofold improvement was seen. This led us to conclude that there is no general and uniform rule for motif and sequence inversion, respectively, but a sole dependency on the individual sequence itself.

In conclusion, the heme binding studies revealed subclasses B and D as the most effective in promoting heme binding. Especially the motifs HXH and HXXXH were found most frequently in strong and medium heme binders.

2.3. Resonance Raman studies

In order to investigate how the peptides of each subclass A–D interact with heme on a structural basis, we performed resonance Raman (rRaman) spectroscopy with at least one heme-peptide complex from every possible combination of H and Y in each subclass. The only exceptions were YXXXY, HY, and YY, as no peptides with these motifs were found in the literature or from HTS. In total, 20 nonapeptides were analyzed in complex with heme (Fig. S2, Table 1). The analysis of the rRaman spectra focused on the ν_3 band, which is visible at 1490 cm^{-1} in unbound heme and pentacoordinated (5c) heme-peptide complexes, whereas it occurs at ~1500 cm^{-1} in hexacoordinated (6c) states [20].

Peptides from subclass A preferentially bound heme in a pentacoordinate fashion, which was especially pronounced when the motif YH was present. In HH peptides, a small fraction of 6c complexes was also detected, which is likely due to the formation of 1:2 heme-peptide complexes [18,20].

Subclass B displayed mixed coordination states with different preferences depending on the respective coordinating amino acids. HXH peptides showed a clear tendency to 6c. However, a spacer of only one amino acid is insufficient for a clamp-like arrangement [18]. Therefore,
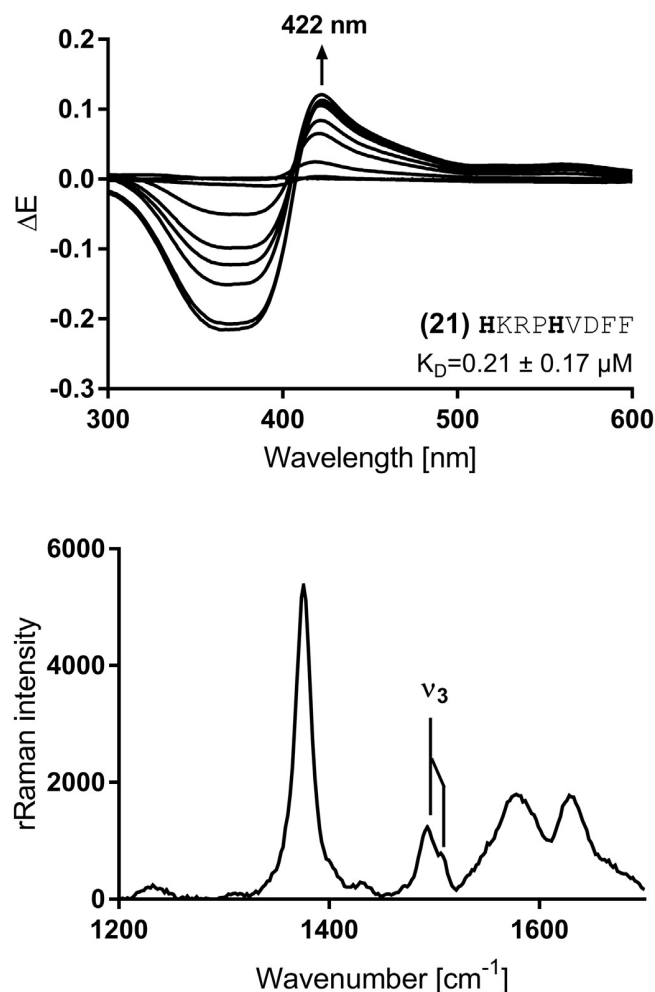
4

Fig. 3. Representative UV/vis and resonance Raman spectra.
Top: Peptide **21** (HKRPHVDFF) at 20 μM was incubated with heme concentrations from 0.4 μM to 35 μM and the change in UV/vis absorbance relative to free heme was recorded. Bottom: A 1:1 complex of peptide **21** and heme was measured with resonance Raman spectroscopy (λ_{ex} = 405 nm). A split ν_3 band indicates a mixture of penta- and hexacoordination [18].

Fig. 4. NMR structures of representatives of subclass B.
Peptide **16** (FLENYLYER), top, and **9** (WIHGHIRDK), bottom, as representatives of subclass B were investigated using 2D NMR spectroscopy. The 20 best structures out of 100 are shown in free (left) and Ga(III)PPIX complexed (right) state. For peptide **9** the complex structure was generated by in silico docking and MD simulation. Structure representations were created as trajectories in VMD [65] using a smoothing range of 1.

it probably occurs in a sandwich conformation. With any other amino acid combination, i.e. YXH, HXY, and YXY, the 5c/6c balance is shifted towards 5c.

A spacer length of two amino acids (e.g., HXXH), as represented by subclass C, almost exclusively led to pentacoordinate heme-peptide complexes. Only peptides **17** (QHLVHFDRR) and **20** (WYAAYTKPS) showed portions of a 6c state.

Subclass D exhibited mixtures of penta- and hexacoordination without a clear preference for either state (Fig. 3, Table 1). With a spacer length of three amino acids, hexacoordination in a clamp-like state in addition to a sandwich conformation are equally possible [17,18], but might not necessarily occur as can be exemplified with peptides **22**, **24**, and **29i**, which exclusively formed 5c complexes.

2.4. NMR studies

In earlier studies of H/Y-based HRMs we already reported on NMR structural analyses of individual complexes, most notably **HH**QYHA-RVA and KPFKYD**HH**Y, representing classes VI and VIII, as well as subclasses A, C, and D, respectively (Fig. 1) [18]. It was shown that only subclass D was able to bind heme in a clamp-like hexacoordination. Additional in silico studies on the spacer length confirmed that more
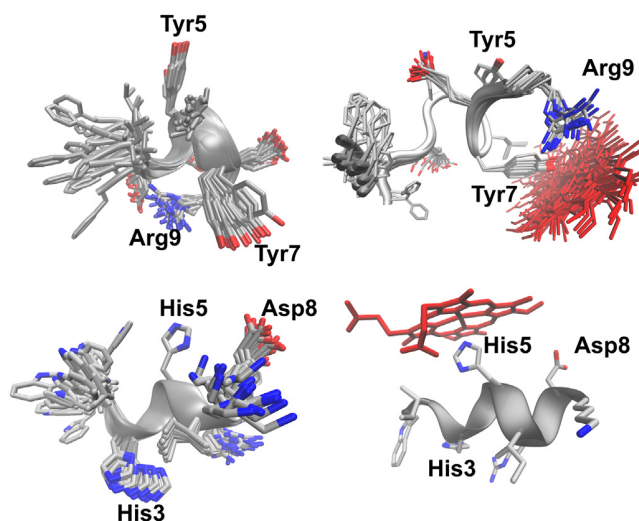
than two amino acids as a spacer are required for this arrangement. Subclass B (H/YXH/Y), however, remained unexplored by NMR herein because of precipitation of several respective peptides in complex with heme at high concentrations (e.g., **6i**, **8**). This prompted us to investigate peptides **16** (FLENYLYER) and **9** (WIHGHIRDK) for structural analysis instead (Fig. 4).

In case of peptide **16**, 142 NOEs for the uncomplexed peptide and 98 NOEs for the complex were detected and used for structure calculation. As observed earlier for other peptides [17,18], the free peptide was already structurally defined with some alpha-helical tendencies and an average backbone RMSD of 1.21 ± 0.41 Å when considering residues 2–8 (Fig. 4, Tables S2 and S3). Addition of gallium(III) protoporphyrin IX (Ga(III)PPIX) [17,18,20] induced only minor changes in chemical shifts of the amino acids surrounding the binding site. However, the H^{γ}-resonance of Leu6 and the H^{ϵ}-resonance of Arg9 disappeared in the complex spectra probably due to a broadening effect upon Ga(III)PPIX binding. On the other hand, the resonances of several aromatic protons emerged (Phe1, Tyr5, Tyr7), which may be due to a stabilization by π-π interactions with the porphyrin ring.

The structure of the Ga(III)PPIX-peptide **16** complex showed a marked increase in rigidity with a backbone RMSD of 0.46 ± 0.15 Å but dispersion of the helical tendencies in the central part. Surprisingly, structure calculations with heme coordinated by Tyr5 led to excessive violation of restraints. In agreement with previous studies [17], a linker length of only one amino acid appeared to be too short to enable clamp-like binding with both Tyr5 and Tyr7. Indeed, Tyr7 was found to be the coordinating amino acid of peptide **16** (Fig. 4) and Arg9 was reoriented towards the porphyrin system. Tyr5 is not involved in heme binding, but rather facing away from the heme surface. Both, Leu6 and Arg9 face towards the porphyrin moiety and may stabilize the binding via interactions with the methyl/vinyl groups and the propionates, respectively. Glu8 is facing away from the porphyrin as it would otherwise repel the propionates. The emergence of 6c contributions in rRaman is most likely due to 1:2 heme:peptide interactions.

HXH emerged as an especially promising motif in UV/vis. Due to the aforementioned problem of precipitation of other complexes, we solved the structure of peptide **9** (WIHGHIRDK) using 79 NOEs, with a backbone RMSD of 0.83 ± 0.38 Å between residues 2–8. Upon addition of Ga(III)PPIX, a major part of the complex precipitated and the remaining
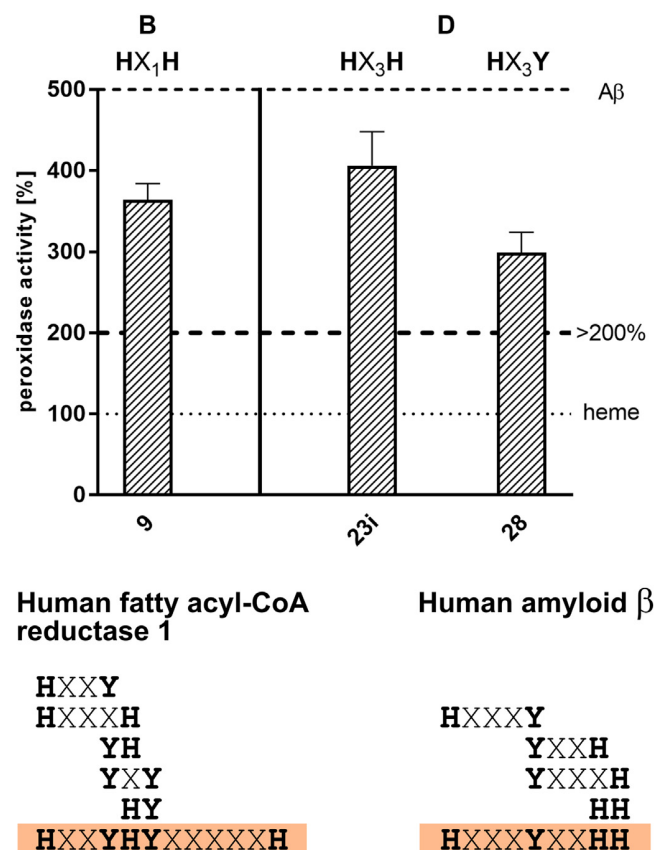
5

Fig. 5. Peroxidase activity of the most potent peptides.
Top: Only 3 peptides from subclasses B and D showed a significant increase in peroxidase activity (> 200%) Bottom: Schematic representation of potential HRMs and included motifs according to subclasses A–D found in sequence stretches from human fatty acyl-CoA reductase 1 (H^{346}–H^{357}) and human amyloid β (H^6–H^{14}).

dissolved sample showed very low signal intensity making structure determination impossible. Thus, we performed in silico docking and molecular dynamics (MD) simulation of heme to the structure of the free peptide. Peptide **9** occurred in α-helical fold before and after heme binding, and, in the latter case, primarily formed interactions to the porphyrin via the coordinating His5 and Ile2. In good agreement with previous studies, clamp-like heme binding was not possible with only one linker amino acid [17]. The 6c contributions in rRaman are therefore likely due to 1:2 heme:peptide interactions. Interestingly, Asp8 was faced towards the porphyrin, while both Arg7 and Lys9 were directed away from the heme plane. Heme binding to His3 was ruled out, because heme was exclusively coordinated by His5 in the simulation experiment with both histidines present.

2.5. Peroxidase activity of peptides with H/Y motifs

The heme-peroxidase activity of a total of 29 heme-peptide complexes was determined using 3,3′,5,5′-tetramethylbenzidine (TMB) as the substrate as described earlier [21], and normalized against the activity of free heme (100%, Table 2, Fig. 5). In order to elucidate which arrangement of H and Y within the motifs yields the strongest peroxidase-like activity, representatives of each subclass were investigated with 3 peptides for A, 7 for B, 3 for C, and 9 for subclass D. In addition, 7 peptides from classes V and VII containing only one potential coordinating amino acid (either His or Tyr) were tested (Table 2).

A total number of 9 peptides, which mainly belong to subclass D, showed at least a slight enhancement (> 150%) of the catalytic activity

(Fig. 5, Table 2). The highest peroxidase activity after incubation with heme, however, was found for peptides **23i** (FAEFHEKRH, 406 ± 42%), **9** (WIHGHRDK, 364 ± 20%), and **28** (HPFPYIWKA, 299 ± 25%) (Fig. 5, Table 2). With respect to the K_D values of active and inactive peptides, a high to moderate binding affinity was favorable for an increase of peroxidase activity compared to heme alone, but does not appear to be an absolute criterion.

In general, subclasses B and D were preferred over A and C. Motifs with an additional histidine enhanced the activity of both H- and Y-based peptides stronger than sequences containing an additional tyrosine. Y-based peptides with an additional tyrosine from subclass B (YXY) neither bound heme strongly nor enhanced the peroxidase activity (Tables 1 and 2). No member of subclasses A and C had an effect on the activity.

None of the peptides with only one coordinating amino acid (either H or Y) had a considerable effect on the catalytic activity, with the exception of peptide TVEIHDLFF (218 ± 33%), which showed a slight increase. These results match the findings from earlier studies, in which e.g. peptide WELDYFQWK did not show any activity [21].

In case of subclass B, only peptides with the HXH motif, i.e. peptides **9** (WIHGHIRDK, 364 ± 20%) and **8i** (RVHKHAAKF, 176 ± 21%) had an enhancing effect on the peroxidase activity. For subclass D, the H-based peptide **24** (YNLLHBAFD, 162 ± 29%) as well as the three Y-based peptides **29** (HADTYFGWR, 199 ± 27%), **28i** (AKWIYPFPH, 185 ± 25%), and **29i** (RWGFYTDAH, 181 ± 40%) showed a slight increase, whereas peptides **23i** and **28** showed a high peroxidase activity (> 300%) as mentioned before. As a control experiment, the activity of peptide **28** (299 ± 25%) could be reproduced (324 ± 15%) from earlier studies [21].

It can be concluded that an additional histidine at positions P + 4, P − 4, and P − 2 may be beneficial for the catalytic activity. Tyrosine appeared to be less preferred in these positions as most of the respective peptides (**10**, **12**, **16**, **27**) exhibited no or only weak peroxidase-like activity. The presence of the motifs alone does not necessarily induce a peroxidase-like activity, as can be exemplified with peptides **8**, **21**, and **23**. As observed in UV/vis studies, additional amino acids appear to play a vital role [21].

We observed several similarities and differences between the peptides with good heme-binding affinity and peroxidase-like activity. The N-terminal residues (P − 4 to P − 1) as well as the amino acids at positions P + 1 and P + 4 seem to have a stronger impact on both, binding and activity, than amino acids at positions P + 2 and P + 3. As already stated, an additional histidine at P − 4, P + 4 or P − 2 can be favorable for peroxidase activity. Comparing the two Y-based peptides **28** (HPFPYIWKA, 299 ± 68%) and **28m** (APFPYIWKA, 138 ± 46%) revealed an obvious beneficial effect of the second histidine at P − 4.

As a shared sequence characteristic, basic and hydrophobic residues emerged in all active peptides with the exception of TVEIHDLFF, which did not contain a basic amino acid K or R. Surprisingly, the most active peptides displayed a reduced number of basic amino acids, in particular for the N-terminal residues and the position P + 1. In general, active peptides contained less basic amino acids at P − 2 and P + 1 than inactive peptides. Conversely, we observed an increased occurrence of acidic and polar amino acids (e.g., at position P − 2). Furthermore, all investigated peptides with histidine at the C-terminal position P + 4 showed an increased peroxidase-like activity (**23i**, **28i**, **29i**), which highlights histidine as an important amino acid for this position. In addition, it should be noted that for all active peptides a mixed coordination state is observed in resonance Raman spectroscopy (Table 1).

From the presented data it is evident that heme binding not necessarily correlates with peroxidase activity. Instead, specific properties such as a reduced number of basic amino acids, availability of polar and/or acidic amino acids as well as a certain number of hydrophobic residues in a suitable arrangement can yield peptides that are able to increase the peroxidase-like activity of heme. The introduction of an

6

additional histidine in distinct positions also promotes an increase in catalytic activity, although it may not be beneficial in every position. It is also necessary for a peroxidase-like activity that the heme-peptide complex contains a pentacoordinate Fe(III) moiety. This is realized for complexes that show mixtures of penta- and hexacoordinated species.

Promising motifs from the presented peptide series are HXH, HXXXH, and HXXXY, representing a further refinement of the results from the heme-binding studies on subclasses B and D. Interestingly, those motifs also occur in the earlier investigated sequences from Aβ and a 23mer peptide from human fatty acyl-CoA reductase 1 (Fig. 5) that show a significant increase of the peroxidase-like activity in complex with heme (Aβ: ~500%, 23mer: ~1000%) [21]. For Aβ, this finding is further supported by the nonapeptide GYEVHHQKL that was earlier found to increase the peroxidase activity in complex with heme [21]. However, it should be noted that for Aβ the heme coordination site is described to be H13 resulting in a YXXH^{13}H motif [52]. The underlying reason may be possible synergistic effects of more than two coordinating amino acids.

3. Discussion

The discovery of heme-regulated proteins depends substantially on the available knowledge of sequential and structural data on heme-regulatory motifs. Although there is increasing evidence of proteins containing HRMs, still little information on motifs other than CP exists [31–37]. For sequences containing histidine or tyrosine as the iron-coordinating amino acid, the importance of further histidines and/or tyrosines in the sequence was revealed earlier, but information about the localization and motif composition was still missing [18,19]. With the results presented here, a detailed analysis of His- and Tyr-based motifs containing one additional histidine or tyrosine is provided. In addition, their influence on functional properties of the respective complexes with ferric heme is demonstrated.

Based on a more elaborate subdivision of HRM classes VI and VIII, we investigated a set of peptides of different combinations of histidines and/or tyrosines with varying spacer length according to subclasses A–D (Fig. 1). In agreement with previous studies, a positive net charge and more than one coordinating amino acid were found to be beneficial for heme binding [16]. However, these characteristics were not sufficient to produce strong heme binding, as can be exemplified with **1** (RQR**HH**KEFK, K$_D$: 31.51 ± 0.80 μM) and **6i** (GS**H**K**H**LQVB, K$_D$: 39.24 ± 4.84 μM). In our data set, including only peptides with two potential coordination sites, His-containing peptides frequently bound stronger to heme than Tyr-containing ones. In this context it might be interesting that in naturally occurring Tyr-based hemoproteins, Tyr can be involved in hydrogen bonds, which amplify its tyrosinate character [53,54]. For example in peptides **28** (**H**PF**PY**IWKA) and **28m** (APFP-**Y**IWKA) this effect might occur, whereas in peptide **16** (FLEN**Y**L**Y**ER), this effect was less pronounced.

A distance of one or three spacer amino acids between two potential coordination sites (subclass B and D) was preferred over zero or two amino acids (subclass A and C) (Fig. 6 top). While subclasses A and C did not show peptides with promising heme-binding capabilities, in subclasses B and D certain motifs were more prominent.

The HXH motif (subclass B) occurred in peptides with moderate and strong heme binding. Structural investigations on heme-peptide complexes of peptides **9** (HXH) and **16** (YXY) indicated two slightly different scenarios. Both peptides displayed α-helical secondary structure elements in the free state, which dispersed upon heme binding in case of **16**, but was maintained for **9**. Subclass D revealed several strong and moderate heme binders, which especially contain HXXXH and HXXXY motifs.

In summary, our data on H/YX$_{0-3}$H/Y peptides allow for a hypothesis on possible heme-binding modes. i) Two directly adjacent histidines/tyrosines seem to be too close and sterically hindered to synergistically bind heme. ii) One spacer amino acid is found in several of

the strong and moderate heme binders (**6**, **8**, **9**, **12**, **15**). iii) The presence of two spacer amino acids preferably produces non-binders. iv) Finally, for peptides from subclass D it is possible to form a clamp-like hexacoordinative state, which in turn may be reflected in the peptides with the highest heme-binding affinities reported in this study. It has to be acknowledged that the results presented herein are based on ferric heme and that ferrous heme can have different properties [55].

In accordance with the results of the binding experiments, peptides of subclasses B (HXH) and D (HXXXH, HXXXY) also appeared to be favorable for developing a peroxidase-like activity in complex with heme. Histidine exhibited a much stronger propensity of increasing the catalytic activity of heme as tyrosine. Tyr-based peptides only achieved significant activity when a further histidine was present, as can be seen from the aforementioned heme-binding motifs. Interestingly, N-terminal amino acids, i.e. acidic and polar residues at P − 2, seem to contribute to the development of a peroxidase-like activity. In addition, in several positions (P − 4 to P − 1, and P + 1) a reduction of the number of basic amino acids was observed compared to peptides with less or without an effect. Histidine as C-terminal amino acid at P + 4 represents a key residue for catalytic activity. For the nonapeptide with the highest activity in our study (**23i**, FAEF**H**EKR**H**, 406 ± 42%), all beneficial factors for heme binding and peroxidase activity upon complex formation are found in the sequence: i) increased electron density due to a polar or acidic residue in P − 2, ii) no basic amino acid in positions P − 2 and P + 1, and an additional His in P + 4.

These findings are in good agreement with the conserved residues at the proximal site of natural heme peroxidases, e.g. horseradish peroxidase (His, Asp, Phe) and catalases (Tyr, Arg, Asp). In these enzymes an increased electron density due to aromatic π-stacking interactions or hydrogen bonds can stabilize the positive net charge of the emerging radical complex in the catalytic reaction [56–58]. Consequently, cleavage of the peroxide is facilitated by the proximal heme ligand due to electron donation ('push'-effect) (Fig. 6, bottom) [59]. Peptides **28** and **29** (HXXXY) resemble catalases to some extent, i.e. a basic amino acid (catalases: Arg, peptides: His) with a spacer of three amino acids to the heme-coordinating Tyr [60]. A hydrogen bond formed between Arg and Tyr increases the electron density at the proximal site to facilitate the reaction at the distal heme site and might be mimicked by a His-Tyr hydrogen bond in peptides **28** and **29** (Fig. 6, bottom) [58]. Three notions support this hypothesis: i) A substitution of His in **28** for Ala in **28i** showed a complete loss in catalytic activity, ii) All peptides with catalytic activity showed mixed coordination states, and iii) In the NMR structure of the peptide **9**-Ga(III)PPIX complex, the 'push effect' is seen in His5, which is stabilized by hydrogen bonds with Asp8.

Taken together, we confirmed the motifs HXH, HXXXH, and HXXXY as heme-regulatory motifs beyond CP. These motifs display preferential heme binding in a nonapeptide model and can also be found in larger peptides, such as Aβ40/42 and a 23mer from fatty acyl-CoA reductase 1 [19,21], where they might be involved in the induction of a peroxidase activity (Fig. 5). We expect these results to shed light on heme binding to proteins via histidine and tyrosine, and allow for the targeted discovery of heme-regulated proteins with a potential for exerting a catalytic activity. In the context of heme-induced pathological states, such as for Aβ, this work may provide valuable impetus to increase our understanding of the role of heme as a catalyst and signaling molecule.

4. Materials and methods

4.1. Materials

Hemin BioXtra, from porcine, was purchased from Sigma Aldrich, St. Louis, Missouri, USA. Ga(III) protoporphyrin IX (Ga(III)PPIX) was purchased from Frontier Scientific, Logan, Utah, USA. TMB was purchased from AppliChem GmbH, Darmstadt, Germany.
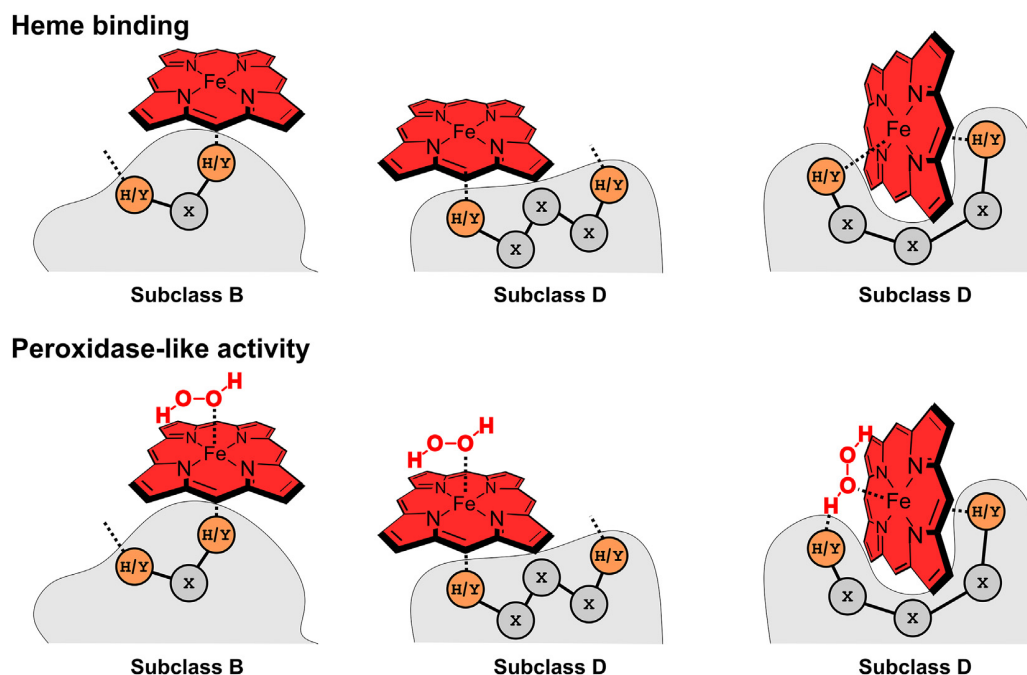
Heme binding

Subclass B   Subclass D   Subclass D

Peroxidase-like activity

Subclass B   Subclass D   Subclass D

Fig. 6. Spacer lengths of 1 or 3 are preferred due to spatial arrangement.
Top: Schematic depiction of the spatial arrangement of coordinating amino acids in peptides from subclasses B and D. Bottom: Different possible penta- and hexacoordinative states and the transition between them allows for H_2O_2 binding and conversion. Dotted lines on H/Y represent amino acid side chains or interactions with the iron ion.

4.2. Methods

4.2.1. Peptide synthesis

Standard Fmoc solid-phase peptide synthesis, purification by high-performance liquid chromatography and concentration determination using amino acid analysis were performed as described earlier (Table S1) [20].

4.2.2. Mass spectrometry

Molecular mass measurements for the synthesized peptides were performed using either an ultrafleXtreme TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) with α-cyano-4-hydroxycinnamic acid as matrix or a LC-ESI-MS system (Dionex UltiMate 3000 LC (ThermoScientific, Dreieich, Germany) in combination with a micrOTOF-Q III system (Bruker Daltonics GmbH, Bremen, Germany)) equipped with a C18 column (EC100 mm × 2 mm Nucleoshell RP18 Gravity, 2.7 μm particle size, 90 Å pore size (Macherey-Nagel, Düren, Germany)). The column was kept at 25 °C during measurement and the absorbance at 220 nm was monitored during measurement. MALDI data was evaluated using FlexAnalysis (Bruker Daltonics Bremen, Germany). Analysis of the LC-ESI-MS data was processed using Compass Data Analysis 4.2 software (Bruker Daltonics, Bremen, Germany).

4.2.3. UV/vis spectroscopy

Heme-binding analysis with UV/vis spectroscopy was performed as described earlier using peptide concentrations of 5–20 μM [18]. K_D values were determined using the following equation published by Pîrnau and colleagues [61]:

$$\Delta A = 0.5 \times \Delta \varepsilon \times ((C_H + n \times C_P + K_D) \pm ((C_H + n \times C_P + K_D)^2 - 4 \times C_H \times n \times C_P)^{0.5})$$

where ΔA corresponds to the difference in absorbance at the wavelength of maximum in the difference spectrum, Δε corresponds to the difference in the molar extinction coefficients of heme and the heme-peptide complex, C_H corresponds to the heme concentration, n corresponds to the number of heme binding sites on the peptide, C_P corresponds to the peptide concentration, and K_D corresponds to the dissociation constant. PRISM 7 (GraphPad Software, San Diego, California, USA) was used as the software program. ΔA was determined from the UV/Vis difference spectra at the wavelength of the maximum. C_H and C_P were known from the experiment. C_H was normalized according to an extinction coefficient of 32.482 mM^{-1} cm^1 at 398 nm, which was determined by a linear fit of 19 independently dissolved heme solutions in the range of 0.4–40 μM [18]. In the initial fit n was set to 0.5, 1, and 2, as these values are the only possible stoichiometries of a nonapeptide-heme interaction, while both Δε and K_D were variable. K_D values (± SEM) were given for the best possible fit.

Statistical analysis was performed using PRISM 7 (GraphPad Software, San Diego, California, USA). K_D values for peptides from previous studies were reevaluated using the aforementioned extinction coefficient.

4.2.4. Resonance Raman and NMR spectroscopy

NMR and rRaman measurements were performed as described earlier with minor modifications [17,20]. In brief, the porphyrin was dissolved in 1 N NaOH corresponding to 5.13% of the final volume by occasional agitation in the dark for 30 min. The solution was diluted with water and phosphate buffer (25 mM phosphate buffer, pH 7.0) and neutralized with 1 N HCl to a final porphyrin concentration of 400 μM. This solution was then added to the peptide powder to yield the complex in 1:1 M ratio. For NMR spectroscopy, D_2O was added to a final concentration of 8% v/v.

Raman spectra were collected with a commercial upright micro-Raman set-up (CRM 300, WITec GmbH, Germany) equipped with TopMode-405 nm diode laser as excitation source. The spectrometer is fitted with a grating of 600 g/mm and a back-illuminated CCD camera (DV401A BV-532, ANDOR, 1024 × 127 pixels) cooled down to −60 °C. The samples were measured in liquid condition by focusing the 405 nm laser light through a Zeiss 10× objective (NA 0.2) giving 22 mW at the object plane. The sample was taken in a quartz cuvette (~80 μl of

sample volume) and placed on a custom-made sample holder which was spirally rotated with a speed of 1 rotation/60s. Raman spectra were collected in triplicate with 30 s integration time per spectrum.

NMR spectra were recorded with an Avance III HD 700 MHz or 750 MHz cryo spectrometer (Bruker BioSpin, Rheinstetten, Germany) with a He-CP 700 or 750 QCI H-P/C/N cryo probe head at 10 °C. ^{1}H resonances were assigned in CcpNmr [62] using 2D [^{1}H$-^1$H]-DQF-COSY, [^{1}H$-^1$H]-TOCSY (80 or 60 ms spinlock time) and [^{1}H,^{1}H]-NOESY (100 ms or 450 ms mixing time) spectra. Upper distance and dihedral constraints were derived as described previously [17]. In addition, lower distance constraints were derived from the [^{1}H$-^1$H]-NOESY peak intensity. Structure calculation was performed in YASARA version 19.7.5 [63] using the nmr_solve macro with default parameters. Heme was treated as a ligand and was added to the roughly folded peptide structure derived from the nmr_fold macro and used in all further structure refinement steps. For peptide **9** the resolution in complex with Ga(III)PPIX was insufficient for proton assignment due to low solubility. In this case, heme was docked to the first structure of the NMR ensemble of the free peptide using the dock_runensemble macro in YASARA. The best-docked heme-peptide complex was further subjected to refinement via MD simulation using the md_refine macro in YASARA and the final snapshot of this simulation was presented.

4.2.5. TMB-based peroxidase assay

Determination of peroxidase activity of selected heme-peptide-complexes was performed as described before [21]. Briefly, 42 μM heme was incubated with an equimolar amount of peptide in PBS buffer for 30 min. As substrate a TMB/H_2O_2 mix composed of equal amounts of 163.2 mM H_2O_2 in citrate buffer (pH 5) and a 1.66 mM TMB solution dissolved in 0.12 M HCl was used. 10 μl of a pre-incubated heme-peptide or buffer-heme solution were added to 200 μl substrate mix and measured for 8 min at 652 nm. Samples without heme and/or the substrate mix were included into the measurements as controls as well as a blank containing only buffer and substrate mix. In each measurement of the peroxidase activity the approaches containing the heme-peptide-complexes or free heme were recorded in triplicates and the average value was determined. All data have been normalized against the activity of heme only (100%). All measurements were performed at room temperature.

Author contributions

BFS, TK, and DI designed the research. BFS, MSB, ID, and MH performed the research with AR, UN, SN, DM, OO, and DW. BFS, MSB, MH, and TK synthesized and analyzed the peptides and performed UV/Vis binding studies with heme. MSB, ID, and TK performed the studies on peroxidase activity. BFS, MH, AR, and UN performed and analyzed data from resonance Raman spectroscopy. BFS, SN, DM, DW, and OO measured free peptides and heme-peptide complexes by NMR spectroscopy and calculated the structures. Docking studies were performed and analyzed by BFS and OO. BFS, MSB, UN, OO, TK, and DI wrote the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagen.2020.129603.

References

[1] K. Ogawa, J. Sun, S. Taketani, O. Nakajima, C. Nishitani, S. Sassa, N. Hayashi, M. Yamamoto, S. Shibahara, H. Fujita, K. Igarashi, Heme mediates derepression of Maf recognition element through direct binding to transcription repressor Bach1, EMBO J. 20 (2001) 2835–2843, https://doi.org/10.1093/emboj/20.11.2835.

[2] J.T. Lathrop, M.P. Timko, Regulation by heme of mitochondrial protein transport through a conserved amino acid motif, Adv. Sci. 259 (1993) 522–526, https://doi.org/10.1126/science.8424176.

[3] K. Kaasik, C.C. Lee, Reciprocal regulation of haem biosynthesis and the circadian clock in mammals, Nature. 430 (2004) 467–471, https://doi.org/10.1038/nature02724.

[4] A. Wißbrock, N.B. Goradia, A. Kumar, A.A. Paul George, T. Kühl, P. Bellstedt, R. Ramachandran, P. Hoffman, K. Galler, U. Neugebauer, K. Hampel, B. Zimmermann, S. Adams, M. Wiendl, G. Krönke, I. Hamza, S.H. Heinemann, S. Frey, A. Hueber, O. Ohlenschläger, D. Imhof, Structural insights into heme binding to IL-36α proinflammatory cytokine, Sci. Rep. 9 (2019) 16893, https://doi.org/10.1038/s41598-019-53231-0.

[5] X. Yao, P. Balamurugan, A. Arvey, C. Leslie, L. Zhang, Heme controls the regulation of protein tyrosine kinases Jak2 and Src, Biochem. Biophys. Res. Commun. 403 (2010) 30–35, https://doi.org/10.1016/j.bbrc.2010.10.101.

[6] R.T. Figueiredo, P.L. Fernandez, D.S. Mourao-Sa, B.N. Porto, F.F. Dutra, L.S. Alves, M.F. Oliveira, P.L. Oliveira, A.V. Graça-Souza, M.T. Bozza, Characterization of heme as activator of toll-like receptor 4, J. Biol. Chem. 282 (2007) 20221–20229, https://doi.org/10.1074/jbc.M610737200.

[7] J.A. McIntyre, D.R. Wagenknecht, W.P. Faulk, Autoantibodies unmasked by redox reactions, J. Autoimmun. 24 (2005) 311–317, https://doi.org/10.1016/j.jaut.2005.03.005.

[8] L.T. Roumenina, M. Radanova, B.P. Atanasov, K.T. Popov, S.V. Kaveri, S. Lacroix-Desmazes, V. Frémeaux-Bacchi, J.D. Dimitrov, Heme interacts with C1q and inhibits the classical complement pathway, J. Biol. Chem. 286 (2011) 16459–16469, https://doi.org/10.1074/jbc.M110.206136.

[9] M. Frimat, F. Tabarin, J.D. Dimitrov, C. Poitou, L. Halbwachs-Mecarelli, V. Fremeaux-Bacchi, L.T. Roumenina, Complement activation by heme as a secondary hit for atypical hemolytic uremic syndrome, Blood. 122 (2013) 282–292, https://doi.org/10.1182/blood-2013-03-489245.

[10] U. Muller-Eberhard, J. Javid, H.H. Liem, A. Hanstein, M. Hanna, Plasma concentrations of hemopexin, haptoglobin and heme in patients with various hemolytic diseases, Blood. 32 (1968) 811–815, https://doi.org/10.1182/blood.V32.5.811.811.

[11] D.A. Hanna, R.M. Harvey, O. Martinez-Guzman, X. Yuan, B. Chandrasekharan, G. Raju, F.W. Outten, I. Hamza, A.R. Reddi, Heme dynamics and trafficking factors revealed by genetically encoded fluorescent heme sensors, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 7539–7544, https://doi.org/10.1073/pnas.1523802113.

[12] X. Yuan, N. Rietzschel, H. Kwon, A.B. Walter Nuno, D.A. Hanna, J.D. Phillips, E.L. Raven, A.R. Reddi, I. Hamza, Regulation of intracellular heme trafficking revealed by subcellular reporters, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) E5144–E5152, https://doi.org/10.1073/pnas.1609865113.

[13] L. Zhang, Heme Biology: The Secret Life of Heme in Regulating Diverse Biological Processes, World Scientific, 2011, https://doi.org/10.1142/9789814287937.

[14] T. Li, H.L. Bonkovsky, J.T. Guo, Structural analysis of heme proteins: implication for design and prediction, BMC Struct. Biol. 11 (2011) 13, https://doi.org/10.1109/BIBMW.2010.5703932.

[15] C. Fufezan, J. Zhang, M.R. Gunner, Ligand preference and orientation in b-and c-type heme-binding proteins, Proteins Struct. Funct. Genet. 73 (2008) 690–704, https://doi.org/10.1002/prot.22097.

[16] A. Wißbrock, A.A. Paul George, H.H. Brewitz, T. Kühl, D. Imhof, The molecular basis of transient heme-protein interactions: analysis, concept and implementation, Biosci. Rep. 39 (2019) BSR20181940, , https://doi.org/10.1042/BSR20181940.

[17] H.H. Brewitz, T. Kühl, N. Goradia, K. Galler, J. Popp, U. Neugebauer, O. Ohlenschläger, D. Imhof, Role of the chemical environment beyond the coordination site: structural insight into FeIII protoporphyrin binding to cysteine-based heme-regulatory protein motifs, ChemBioChem. 16 (2015) 2216–2224, https://doi.org/10.1002/cbic.201500331.

[18] H.H. Brewitz, N. Goradia, E. Schubert, K. Galler, T. Kühl, B.F. Syllwasschy, J. Popp, U. Neugebauer, G. Hagelueken, O. Schiemann, O. Ohlenschläger, D. Imhof, Heme interacts with histidine-and tyrosine-based protein motifs and inhibits enzymatic activity of chloramphenicol acetyltransferase from *Escherichia coli*, Biochim.

Biophys. Acta Gen. Subj. 1860 (2016) 1343–1353, https://doi.org/10.1016/j.bbagen.2016.03.027.

[19] T. Kühl, N. Sahoo, M. Nikolajski, B. Schlott, S.H. Heinemann, D. Imhof, Determination of hemin-binding characteristics of proteins by a combinatorial peptide library approach, ChemBioChem. 12 (2011) 2846–2855, https://doi.org/10.1002/cbic.201100556.

[20] T. Kühl, A. Wißbrock, N. Goradia, N. Sahoo, K. Galler, U. Neugebauer, J. Popp, S.H. Heinemann, O. Ohlenschläger, D. Imhof, Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins, ACS Chem. Biol. 8 (2013) 1785–1793, https://doi.org/10.1021/cb400317x.

[21] A. Wißbrock, T. Kühl, K. Silbermann, A.J. Becker, O. Ohlenschläger, D. Imhof, Synthesis and evaluation of amyloid β derived and amyloid β independent enhancers of the peroxidase-like activity of heme, J. Med. Chem. 60 (2017) 373–385, https://doi.org/10.1021/acs.jmedchem.6b01432.

[22] S. Peherstorfer, H.H. Brewitz, A.A. Paul George, A. Wißbrock, J.M. Adam, L. Schmitt, D. Imhof, Insights into mechanism and functional consequences of heme binding to hemolysin-activating lysine acyltransferase HlyC from *Escherichia coli*, Biochim. Biophys. Acta Gen. Subj. 1862 (2018) 1964–1972, https://doi.org/10.1016/j.bbagen.2018.06.012.

[23] A.A. Paul George, M. Lacerda, B.F. Syllwasschy, M.-T. Hopp, A. Wißbrock, M. Hofmann-Apitius, D. Imhof, HeMoQuesT: A webserver for qualitative prediction of transient heme binding to protein motifs, BMC Bioinforma. 21 (2020) 124, https://doi.org/10.1186/s12859-020-3420-2.

[24] L. Zhang, L. Guarente, Heme binds to a short sequence that serves a regulatory function in diverse proteins, EMBO J. 14 (1995) 313–320, https://doi.org/10.1002/j.1460-2075.1995.tb07005.x.

[25] A. Lombardi, F. Nastri, V. Pavone, Peptide-based heme-protein models, Chem. Rev. 101 (2001) 3165–3189, https://doi.org/10.1021/cr000055j.

[26] J. Igarashi, M. Murase, A. Iizuka, F. Pichierri, M. Martinkova, T. Shimizu, Elucidation of the heme binding site of heme-regulated eukaryotic initiation factor 2α kinase and the role of the regulatory motif in heme sensing by spectroscopic and catalytic studies of mutant proteins, J. Biol. Chem. 283 (2008) 18782–18791, https://doi.org/10.1074/jbc.M801400200.

[27] S. Hira, T. Tomita, T. Matsui, K. Igarashi, M. Ikeda-Saito, Bach1, a heme-dependent transcription factor, reveals presence of multiple heme binding sites with distinct coordination structure, IUBMB Life 59 (2007) 542–551, https://doi.org/10.1080/15216540701225941.

[28] J. Yang, K. Ishimori, M.R. O'Brian, Two heme binding sites are involved in the regulated degradation of the bacterial iron response regulator (Irr) protein, J. Biol. Chem. 280 (2005) 7671–7676, https://doi.org/10.1074/jbc.M411664200.

[29] J. Yang, K.D. Kim, A. Lucas, K.E. Drahos, C.S. Santos, S.P. Mury, D.G.S. Capelluto, C.V. Finkielstein, A novel heme-regulatory motif mediates heme-dependent degradation of the circadian factor period 2, Mol. Cell. Biol. 28 (2008) 4697–4711, https://doi.org/10.1128/mcb.00236-08.

[30] H. Ishikawa, M. Kato, H. Hori, K. Ishimori, T. Kirisako, F. Tokunaga, K. Iwai, Involvement of heme regulatory motif in heme-mediated ubiquitination and degradation of IRP2, Mol. Cell 19 (2005) 171–181, https://doi.org/10.1016/j.molcel.2005.05.027.

[31] L. Yin, N. Wu, M.A. Lazar, Nuclear receptor rev-Erbα: a heme receptor that coordinates circadian rhythm and metabolism, Nucl. Recept. Signal. 8 (2010), https://doi.org/10.1621/nrs.08001 nrs.08001.

[32] S. Raghuram, K.R. Stayrook, P. Huang, P.M. Rogers, A.K. Nosie, D.B. McClure, L.L. Burris, S. Khorasanizadeh, T.P. Burris, F. Rastinejad, Identification of heme as the ligand for the orphan nuclear receptors REV-ERBα and REV-ERBβ, Nat. Struct. Mol. Biol. 14 (2007) 1207–1213, https://doi.org/10.1038/nsmb1344.

[33] D. Lechardeur, B. Cesselin, U. Liebl, M.H. Vos, A. Fernandez, C. Brun, A. Gruss, P. Gaudu, Discovery of intracellular heme-binding protein HrtR, which controls heme efflux by the conserved HrtB-HrtA transporter in *Lactococcus lactis*, J. Biol. Chem. 287 (2012) 4752–4758, https://doi.org/10.1074/jbc.M111.297531.

[34] E.B. Draganova, N. Akbas, S.A. Adrian, G.S. Lukat-Rodgers, D.P. Collins, J.H. Dawson, C.E. Allen, M.P. Schmitt, K.R. Rodgers, D.W. Dixon, Heme binding by *Corynebacterium diphtheriae* HmuT: function and heme environment, Biochemistry. 54 (2015) 6598–6609, https://doi.org/10.1021/acs.biochem.5b00666.

[35] N. Gupta, S.W. Ragsdale, Thiol-disulfide redox dependence of heme binding and heme ligand switching in nuclear hormone receptor rev-erbβ, J. Biol. Chem. 286 (2011) 4392–4403, https://doi.org/10.1074/jbc.M110.193466.

[36] Y. Kabe, T. Nakane, I. Koike, T. Yamamoto, Y. Sugiura, E. Harada, K. Sugase, T. Shimamura, M. Ohmura, K. Muraoka, A. Yamamoto, T. Uchida, S. Iwata, Y. Yamaguchi, E. Krayukhina, M. Noda, H. Handa, K. Ishimori, S. Uchiyama, T. Kobayashi, M. Suematsu, Haem-dependent dimerization of PGRMC1/Sigma-2 receptor facilitates cancer proliferation and chemoresistance, Nat. Commun. 7 (2016) 1–13, https://doi.org/10.1038/ncomms11030.

[37] S.L. Freeman, H. Kwon, N. Portolano, G. Parkin, U. Venkatraman Girija, J. Basran, A.J. Fielding, L. Fairall, D.A. Svistunenko, P.C.E. Moody, J.W.R. Schwabe, C.P. Kyriacou, E.L. Raven, Heme binding to human CLOCK affects interactions with the E-box, Proc. Natl. Acad. Sci. U. S. A. 116 (2019) 19911–19916, https://doi.org/10.1073/pnas.1905216116.

[38] X.D. Tang, R. Xu, M.F. Reynolds, M.L. Garcia, S.H. Heinemann, T. Hoshi, Haem can bind to and inhibit mammalian calcium-dependent Slo1 BK channels, Nature 425 (2003) 531–535, https://doi.org/10.1038/nature02003.

[39] F.T. Horrigan, S.H. Heinemann, T. Hoshi, Heme regulates allosteric activation of the Slo1 BK channel, J. Gen. Physiol. 126 (2005) 7–21, https://doi.org/10.1085/jgp.200509262.

[40] M.J. Burton, S.M. Kapetanaki, T. Chernova, A.G. Jamieson, P. Dorlet, J. Santolini, P.C.E.E. Moody, J.S. Mitcheson, N.W. Davies, R. Schmid, E.L. Raven, N.M. Storey, A heme-binding domain controls regulation of ATP-dependent potassium channels, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 3785–3790, https://doi.org/10.1073/pnas.1600211113.

[41] S.M. Kapetanaki, M.J. Burton, J. Basran, C. Uragami, P.C.E. Moody, J.S. Mitcheson, R. Schmid, N.W. Davies, P. Dorlet, M.H. Vos, N.M. Storey, E. Raven, A mechanism for CO regulation of ion channels, Nat. Commun. 9 (2018) 907, https://doi.org/10.1038/s41467-018-03291-z.

[42] H. Atamna, W.H. Frey, A role for heme in Alzheimer's disease: heme binds amyloid β and has altered metabolism, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 11153–11158, https://doi.org/10.1073/pnas.0404349101.

[43] H. Atamna, K. Boyle, Amyloid-β peptide binds with heme to form a peroxidase: relationship to the cytopathologies of Alzheimer's disease, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 3381–3386, https://doi.org/10.1073/pnas.0600134103.

[44] J. Flemmig, M. Zámocký, A. Alia, Amyloid β and free heme: bloody new insights into the pathogenesis of Alzheimer's disease, Neural Regen. Res. 13 (2018) 1170–1174, https://doi.org/10.4103/1673-5374.235021.

[45] D. Pramanik, S.G. Dey, Active site environment of heme-bound amyloid β peptide associated with Alzheimers disease, J. Am. Chem. Soc. 133 (2011) 81–87, https://doi.org/10.1021/ja1084578.

[46] H. Atamna, W.H. Frey, N. Ko, Human and rodent amyloid-β peptides differentially bind heme: relevance to the human susceptibility to Alzheimer's disease, Arch. Biochem. Biophys. 487 (2009) 59–65, https://doi.org/10.1016/j.abb.2009.05.003.

[47] E. Chiziane, H. Telemann, M. Krueger, J. Adler, J. Arnhold, A. Alia, J. Flemmig, Free heme and amyloid-β: a fatal liaison in Alzheimer's disease, J. Alzheimers Dis. 61 (2018) 963–984, https://doi.org/10.3233/JAD-170711.

[48] H. Atamna, Heme binding to amyloid-β peptide: mechanistic role in Alzheimer's disease, J. Alzheimers Dis. 10 (2006) 255–266, https://doi.org/10.3233/JAD-2006-102-310.

[49] W.J. Geldenhuys, C.J. Van der Schyf, Role of serotonin in Alzheimer's disease, CNS Drugs. 25 (2011) 765–781, https://doi.org/10.2165/11590190-000000000-00000.

[50] S. Mukherjee, M. Seal, S.G. Dey, Kinetics of serotonin oxidation by heme-Aβ relevant to Alzheimer's disease, J. Biol. Inorg. Chem. 19 (2014) 1355–1365, https://doi.org/10.1007/s00775-014-1193-7.

[51] A. Wißbrock, D. Imhof, A tough nut to crack: intracellular detection and quantification of heme in malaria parasites by a genetically encoded protein sensor, ChemBioChem. 18 (2017) 1561–1564, https://doi.org/10.1002/cbic.201700274.

[52] I. Pal, A.K. Nath, M. Roy, M. Seal, C. Ghosh, A. Dey, S.G. Dey, Formation of compound I in heme bound Aβ-peptides relevant to Alzheimer's disease, Chem. Sci. 10 (2019) 8405–8410, https://doi.org/10.1039/c9sc01679a.

[53] C. Caillet-Saguy, M. Delepierre, A. Lecroisey, I. Bertini, M. Piccioli, P. Turano, Direct-detected 13C NMR to investigate the iron (III) hemophore HasA, J. Am. Chem. Soc. 128 (2006) 150–158, https://doi.org/10.1021/ja054902h.

[54] C. Caillet-Saguy, P. Turano, M. Piccioli, G.S. Lukat-Rodgers, M. Czjzek, B. Guigliarelli, N. Izadi-Pruneyre, K.R. Rodgers, M. Delepierre, A. Lecroisey, Deciphering the structural role of histidine 83 for heme binding in hemophore HasA, J. Biol. Chem. 283 (2008) 5960–5970, https://doi.org/10.1074/jbc.M703795200.

[55] S.W. Ragsdale, L. Yi, Evidence that the heme regulatory motifs in heme oxygenase-2 serve as a thiol/disulfide redox switch regulating heme binding, J. Biol. Chem. 282 (2007) 21056–21067, https://doi.org/10.1074/jbc.M700664200.

[56] A.N.P. Hiner, E.L. Raven, R.N.F. Thorneley, F. García-Cánovas, J.N. Rodríguez-López, Mechanisms of compound I formation in heme peroxidases, J. Inorg. Biochem. 91 (2002) 27–34, https://doi.org/10.1016/S0162-0134(02)00390-2.

[57] M. Paoli, J. Marles-Wright, A. Smith, Structure-function relationships in heme-proteins, DNA Cell Biol. 21 (2002) 271–280, https://doi.org/10.1089/104454902753759690.

[58] C.D. Putnam, A.S. Arvai, Y. Bourne, J.A. Tainer, Active and inhibited human catalase structures: ligand and NADPH binding and catalytic mechanism, J. Mol. Biol. 296 (2000) 295–309, https://doi.org/10.1006/jmbi.1999.3458.

[59] P.R. Ortiz de Montellano, Catalytic mechanisms of heme peroxidases, in: E. Torres, M. Ayala (Eds.), Biocatalysis Based on Heme Peroxidases, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 79–107, , https://doi.org/10.1007/978-3-642-12627-7.

[60] J. Rai, S. Raghothama, D. Sahal, Tyrosine-heme ligation in heme-peptide complex: design based on conserved motif of catalase, J. Pept. Sci. 13 (2007) 406–412, https://doi.org/10.1002/psc.862.

[61] A. Pîrnau, M. Bogdan, Investigation of the interaction between naproxen and human serum albumin, Rom. J. Biophys. 18 (2008) 49–55.

[62] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins Struct. Funct. Genet. 59 (2005) 687–696, https://doi.org/10.1002/prot.20449.

[63] E. Krieger, G. Vriend, YASARA view – molecular graphics for all devices – from smartphones to workstations, Bioinformatics 30 (2014) 2981–2982, https://doi.org/10.1093/bioinformatics/btu426.

[64] M.-T. Hopp, N. Alhanafi, A.A. Paul George, N.S. Hamedani, A. Biswas, J. Oldenburg, B. Pötzsch, D. Imhof, Molecular Insights and Functional Consequence of the Interaction of Heme with Activated Protein C, (2019) submitted.

[65] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38, https://doi.org/10.1016/0263-7855(96)00018-5.

35

High-affinity binding and catalytic activity of His/Tyr-based sequences: Extending heme-regulatory motifs beyond CP

Benjamin Franz Syllwasschy[a], Maximilian Steve Beck[a], Ivona Družeta[a], Marie-Thérèse Hopp[a], Anuradha Ramoji[b,c], Ute Neugebauer[b,c], Senada Nozinovic[d], Dirk Menche[d], Dieter Willbold[e,f], Oliver Ohlenschläger[g], Toni Kühl[a*], and Diana Imhof[a*]

[a]Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute, University of Bonn, 53121 Bonn, Germany
[b]Center for Sepsis Control and Care (CSCC), Jena University Hospital, 07747 Jena, Germany
[c]Leibniz Institute of Photonic Technology, Albert-Einstein-Str. 9, 07745 Jena, Germany
[d]Institute for Organic Chemistry and Biochemistry, University of Bonn, 53121 Bonn, Germany
[e]Jülich Research Centre, Institute of Complex Systems – Structural Biochemistry (ICS-6), 52425 Jülich, Germany
[f]Institute of Physical Biology, University of Düsseldorf, 40225 Düsseldorf, Germany
[g]Leibniz Institute on Aging – Fritz Lipmann Institute, 07745 Jena, Germany

Appendix A. Supplementary data
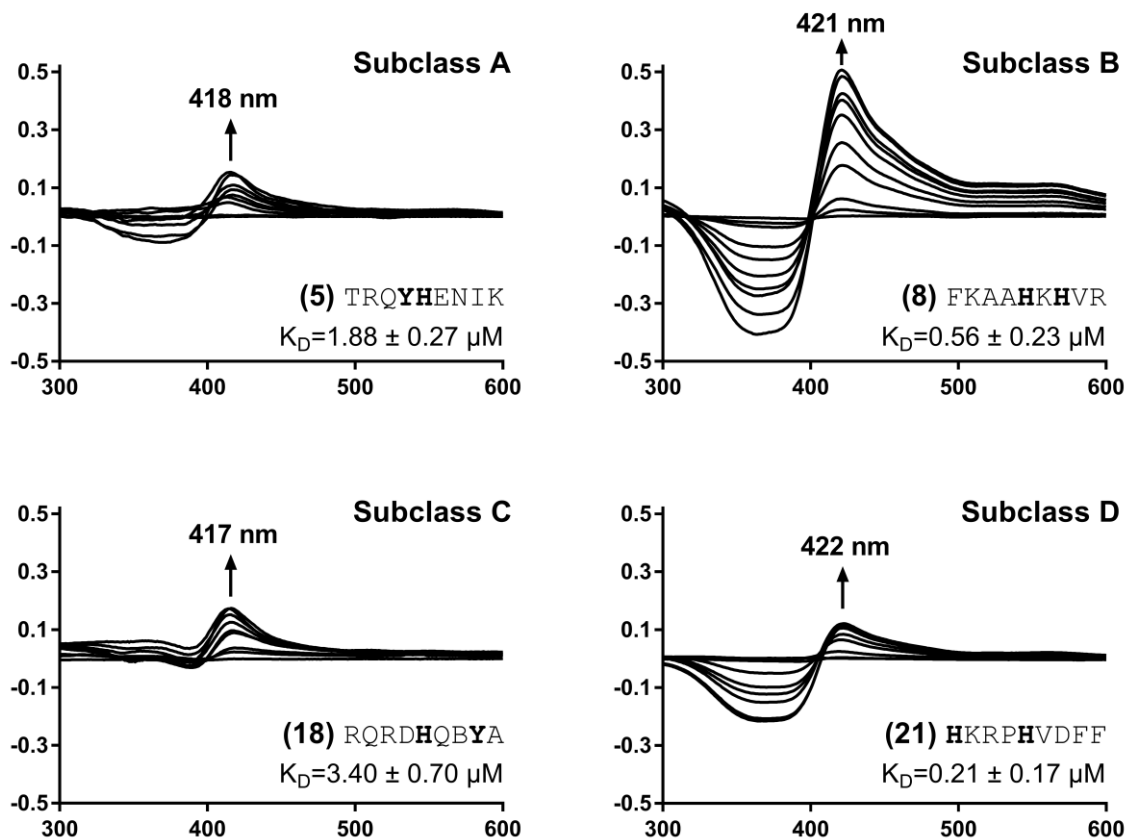
Fig. S1 Examples of UV/vis difference spectra from peptide representatives of subclasses A-D. Peptides with the strongest heme-binding affinity of each subclass are shown. Peptide concentrations are 20 µM and heme concentrations ranged from 0.4 µM – 40 µM. All subclasses primarily show shifts to ~420 nm and belong to UV class 1 [1–4].

Fig. S2 Resonance Raman spectra of the investigated peptides.
Peptide numbers are given in the spectra. The ν3 shift to higher wavenumbers is indicative of hexacoordination of the iron ion [1–4].

Tab. S1 Analytical data of all peptides used in this study.

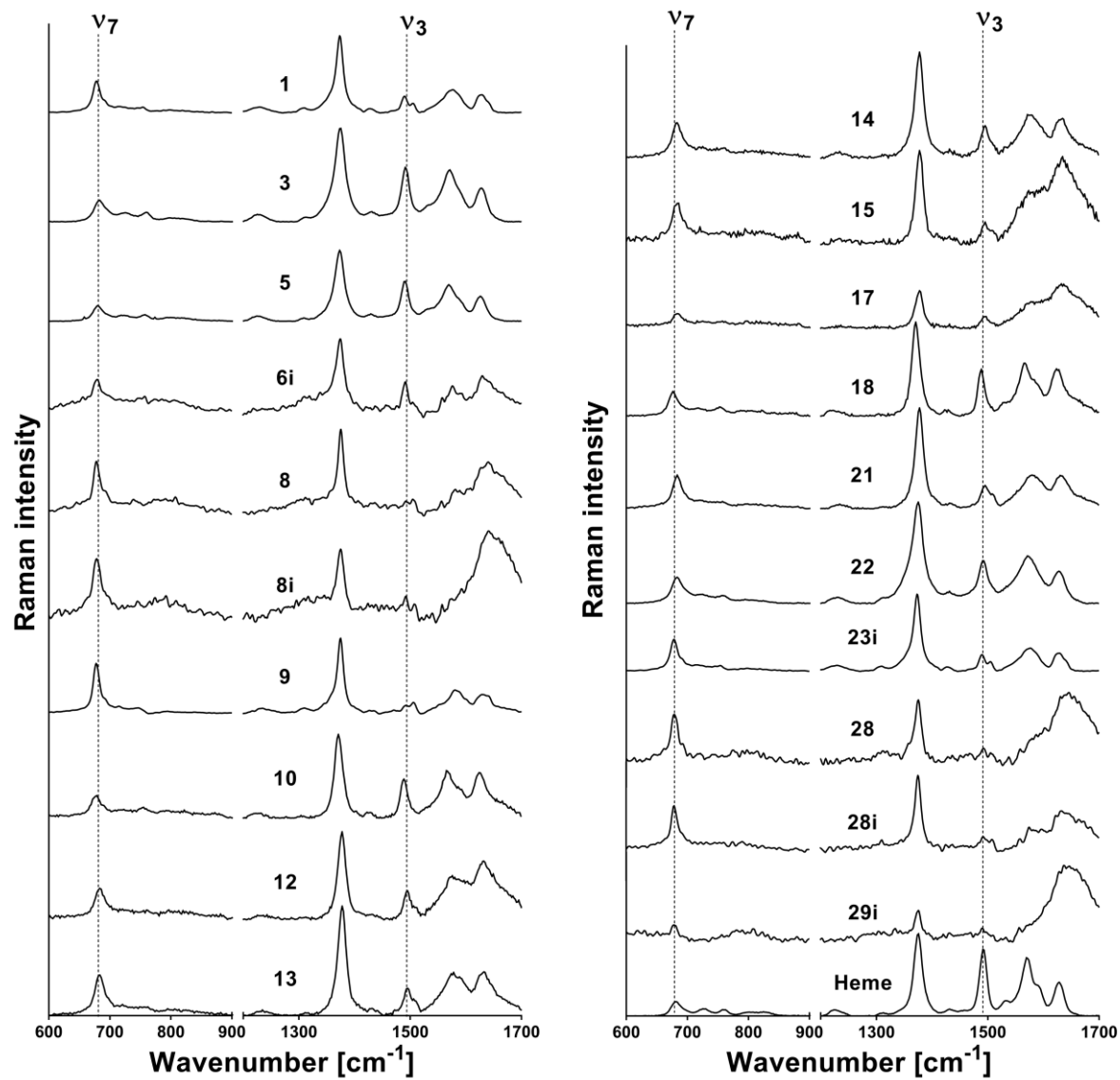| No.	Sequence[a] [Ref]	$M_{theor.}$[b]	$M_{exp.}$[c]	t_R HPLC[g]	R_F TLC[n]
control	AAA**HH**AAAA	788.39	789.42[d]	12.15[h]	0.47[o]
control	AAA**HY**AAAA	814.40	815.42[d]	17.19[h]	0.49[o]
control	AAA**YH**AAAA	814.40	815.41[d]	16.48[h]	0.49[o]
control	AAA**YY**AAAA	840.40	841.51[d]	21.80[h]	0.52[o]
1	RQR**HH**KEFK	1263.70	1264.72[d]	18.44[h]	18.12[p]
2	GGF**YH**LAAD[1]	948.45	949.47[d]	28.08[i]	0.46[q]
3	QKR**YH**EDIF	1233.61	1234.62[d]	28.77[h]	0.22[q]
4	FLF**YH**SQSG[5]	1083.51	542.77[f]	20.40[j]	0.57[q]
5	TRQ**YH**ENIK	1186.61	1187.62[d]	18.52[h]	0.15[q]
control	AA**H**A**H**AAAA	788.39	789.42[d]	12.62[h]	0.51[o]
control	AA**H**A**Y**AAAA	814.40	815.42[d]	17.33[h]	0.57[o]
control	AA**Y**A**H**AAAA	814.40	815.42[d]	16.86[h]	0.38[o]
control	AA**Y**A**Y**AAAA	840.40	863.38[e]	22.34[h]	21.43[p]
6	BVQL**H**K**H**SG[1]	1016.59	1017.57[d]	21.21[i]	0.45[q]
6i	GS**H**K**H**LQVB	1016.58	1017.59[d]	25.65[h]	0.29[o]
7	QG**H**G**H**QNFP[1]	1019.47	1020.48[d]	20.20[i]	0.33[q]
8	FKAA**H**K**H**VR[1]	1091.65	1092.58[d]	18.92[i]	0.66[r]
8i	RV**H**K**H**AAKF[1]	1091.64	1092.65[d]	20.07[h]	20.08[p]
9	WI**H**G**H**IRDK[6]	1159.67	580.83[f]	14.70[k]	0.23[s]
10	PRQA**H**V**Y**RA[1]	1095.60	1096.60[d]	19.02[i]	0.13[o]
11	TVBL**Y**N**H**RI[1]	1126.66	1127.64[d]	29.98[i]	0.39[q]
12	KR**Y**I**H**RDLA	1169.67	1170.69[d]	22.23[h]	0.12[q]
13	MN**Y**I**H**RDLR	1215.62	1216.63[d]	24.88[h]	0.67[q]
14	RN**Y**I**H**RDLR	1201.60	1202.57[d]	23.32[h]	0.67[q]
15	MN**Y**V**H**RDLR	1240.68	1241.67[d]	22.56[h]	0.25[q]
16	FLEN**Y**L**Y**ER[1]	1244.62	1249.40[d]	22.10[i]	0.52[s]
control	A**H**AA**H**AAAA	788.39	789.43[d]	12.67[h]	0.53[o]
control	A**H**AA**Y**AAAA	814.40	815.40[d]	17.78[h]	0.56[o]
control	A**Y**AA**H**AAAA	814.40	815.41[d]	16.35[h]	0.59[o]
control	A**Y**AA**Y**AAAA	840.40	863.39[e]	22.48[h]	0.59[o]
17	Q**H**LV**H**FDRR	1205.64	1206.62[d]	26.46[h]	0.25[q]
18	RQRD**H**QB**Y**A[1]	1184.62	1185.64[d]	22.10[i]	0.18[o]
19	GFGT**Y**SW**H**E[1]	1081.46	1082.36[d]	29.79[i]	0.42[q]
20	W**Y**AA**Y**TKPS[1]	1084.53	1085.54[d]	15.90[k]	0.43[s]
control	**H**AAA**H**AAAA	788.39	789.42[d]	11.69[h]	0.29[o]
control	**H**AAA**Y**AAAA	814.40	815.43[d]	16.66[h]	17.78[p]
control	**Y**AAA**H**AAAA	814.40	815.43[d]	14.67[h]	0.38[o]
control	**Y**AAA**Y**AAAA	840.40	863.44[e]	19.65[h]	20.96[p]
21	**H**KRP**H**VDFF	1180.61	1181.62[d]	31.09[h]	0.15[q]
22	**H**REL**H**GDFV	1107.55	1108.52[d]	25.15[h]	0.37[q]
23	**H**RKE**H**FEAF[1]	1198.60	1199.61[d]	26.64[i]	0.13[q]
23i	FAEF**H**EKR**H**	1198.59	1199.61[d]	23.44[h]	0.20[o]
24	**Y**NLL**H**BAFD[1]	1103.58	1104.58[d]	19.20[l]	0.51[s]
25	KSVI**H**NLV**Y**[1]	1070.62	1071.65[d]	30.32[i]	0.48[o]
26	VRRE**H**FEF**Y**[1]	1280.64	1281.66[d]	29.28[i]	0.33[o]
27	AEFR**H**DSG**Y**[7]	1079.48	540.75[f]	21.80[m]	0.47[o]
28	**H**PFP**Y**IWKA[1]	1156.62	1157.62[d]	14.20[l]	0.43[q]
28m	APFP**Y**IWKA	1090.60	1091.59[d]	16.30[l]	0.37[q]
28i	AKWI**Y**PFP**H**	1156.61	1157.62[d]	33.85[h]	0.22[s]
29	**H**ADT**Y**FGWR[1]	1150.53	1151.47[d]	30.83[i]	0.38[q]
29i	RWGF**Y**TDA**H**	1150.52	1151.53[d]	30.26[h]	0.33[s]
	TVEI**H**DLFF[1]	1118.58	1119.46[d]	38.41[i]	0.49[o]
	QVRL**H**WLSP[1]	1133.65	1134.61[d]	33.38[i]	0.59[s]
	VFKE**H**PAFR[4]	1128.62	1131.02[d]	15.03[k]	0.23[o]
	WELDY**F**QWK[1]	1312.62	1313.62[d]	38.91[i]	0.52[q]
	DNFR**Y**BIPN[1]	1149.59	1150.74[d]	32.61[i]	0.48[q]

IBFRYSSLK[1]	1124.67	1125.56[d]	32.04[i]	0.71[s]

i, inverted sequence of previously reported peptides; m, mutated sequence of previously reported peptides; [a]Peptides were synthesized as C-terminal amides. B, norleucin; [b]Theoretical monoisotopic mass [M]; [c]MALDI or ESI mass spectrometry was performed and monoisotopic mass peaks were detected as: [d][M+H]$^{+}$; [e][M+Na]$^{+}$; [f][M+2H]$^{2+}$. [g]Retention times are stated in minutes as observed by C18-RP-HPLC, eluent A: water with 0.1 % TFA, eluent B: acetonitrile with 0.1 % TFA; [h]0 – 60% eluent B in 60 minutes; [i]0 – 40 % eluent B in 40 minutes; [j]10 – 50 % eluent B in 40 minutes; [k]10 – 40 % eluent B in 30 minutes; [l]20 – 50 % eluent B in 30 minutes; [m]0 – 30 % eluent B in 30 minutes. [n]TLC, thin layer chromatography, was performed using silica gel 60 layer and elution system: [o]Pyridine:ethyl acetate:acetic acid:water (5:5:1:3, v/v); [p]Instead of TLC, HPLC with a C8 Column was performed using a gradient of 0 – 60% eluent B in 60 minutes the value given is the retention time in minutes; [q]Ammonia (25%):2-propanol (3:7, v/v); [r]TLC was performed using RP-8 (C$_8$, reversed phase) layer and elution system methanol:water:TFA (50:50:1, v/v); [s]n-Butanol/acetic acid/water (48:18:24, v/v).

Tab. S2: Chemical shifts of peptide 16 (FLENYLYER) without (black) and with Ga(III)protoporphyrin IX (red) and 9 (WIHGHIRDK).

No.	Res.	H	Hα		Hβ		Hγ			Hδ		other
Pep. 16												
1	Phe	-	4.13	-	2.86	2.63	-	-	-	-	-	-
		-	4.00	-	2.99	2.89	-	-	-	7.14	-	Hε 7.01
2	Leu	-	4.10	-	1.33	1.33	-	-	-	0.68	0.64	-
		-	4.10	-	1.33	1.33	1.27	-	-	0.68	0.64	-
3	Glu	8.41	3.94	-	1.70	1.67	2.03	2.03	-	-	-	-
		8.42	3.93	-	1.68	1.68	2.03	2.03	-	-	-	-
4	Asn	8.27	4.38	-	2.54	2.51	-	-	-	-	-	-
		8.28	4.39	-	2.53	2.51	-	-	-	7.43	6.68	-
5	Tyr	7.92	4.23	-	2.73	2.73	-	-	-	6.81	-	-
		7.92	4.23	-	2.73	2.73	-	-	-	6.80	-	Hε 6.54
6	Leu	7.76	3.95	-	1.23	1.11	1.23	-	-	0.63	0.56	-
		7.75	3.94	-	1.23	1.10	-	-	-	0.63	0.56	-
7	Tyr	7.68	4.26	-	2.84	2.72	-	-	-	6.89	-	-
		7.67	4.25	-	2.83	2.72	-	-	-	6.88	-	Hε 6.59
8	Glu	8.10	3.99	-	1.80	1.71	2.01	2.01	-	-	-	-
		8.08	4.00	-	1.80	1.71	2.01	2.01	-	-	-	-
9	Arg	8.06	4.01	-	1.65	1.53	1.42	1.41	-	2.95	2.87	Hε 7.01
		8.06	4.01	-	1.65	1.52	1.40	1.40	-	2.94	2.93	-
									-			
Pep. 9												
1	Trp	-	4	-	3.03	3.03	-	-	-	6.94	-	Hε 7.24, 7.46, Hζ 6.97, 7.30 Hη 6.83
2	Ile	-	3.82	-	1.37	-	1.02	1.02	0.50	0.75	-	-
3	His	8.11	4.23	-	2.83	2.76	-	-	-	6.78	-	-
4	Gly	8.3	3.69	3.56	-	-	-	-	-	-	-	-
5	His	8.12	4.35	-	2.85	2.82	-	-	-	6.73	-	-
6	Ile	8.03	3.81	-	1.54	-	1.16	0.86	0.60	0.57		-
7	Arg	8.36	3.99	-	1.52	1.44	1.31	1.27	-	2.83	2.83	-
8	Asp	8.15	4.30	-	2.46	2.37	-	-	-	-	-	-
9	Arg	8.17	3.96	-	1.51	1.63	1.22	1.15	-	1.42	1.42	Hε 2.73, 2.73

Tab. S3: Statistical data for the structure calculation of peptide 16 (FLENYLYER) and 9 (WIHGHIRDK) performed with YASARA.

Peptide	**16** free	**16** + GaPPIX	**9** free
Number of NOEs			
total	145	173	79
used in calculation	142	98	79
Avg. RMSD of backbone[Å]	1.76 ± 0.59	1.19 ± 0.36	1.47 ± 0.47
Avg. RMSD of backbone AS[2-8][Å]	1.21 ± 0.41	0.46 ± 0.15	0.83 ± 0.38
Best Z-score	-0.969	-2.342	-1.857
Lowest restraint violation energy [kJ/mol]	0.742	0.277	0.340

References

[1]   H.H. Brewitz, N. Goradia, E. Schubert, K. Galler, T. Kühl, B.F. Syllwasschy, J. Popp, U. Neugebauer, G. Hagelueken, O. Schiemann, O. Ohlenschläger, D. Imhof, Heme interacts with histidine-and tyrosine-based protein motifs and inhibits enzymatic activity of chloramphenicol acetyltransferase from Escherichia coli, Biochim. Biophys. Acta - General Subj. 1860 (2016) 1343–1353. doi:10.1016/j.bbagen.2016.03.027.

[2]   H.H. Brewitz, T. Kühl, N. Goradia, K. Galler, J. Popp, U. Neugebauer, O. Ohlenschläger, D. Imhof, Role of the Chemical Environment beyond the coordination Site: Structural insight into FeIII protoporphyrin binding to cysteine-based heme-regulatory protein motifs, ChemBioChem. 16 (2015) 2216–2224. doi:10.1002/cbic.201500331.

[3]   T. Kühl, A. Wißbrock, N. Goradia, N. Sahoo, K. Galler, U. Neugebauer, J. Popp, S.H. Heinemann, O. Ohlenschläger, D. Imhof, Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins, ACS Chem. Biol. 8 (2013) 1785–1793. doi:10.1021/cb400317x.

[4]   T. Kühl, N. Sahoo, M. Nikolajski, B. Schlott, S.H. Heinemann, D. Imhof, Determination of hemin-binding characteristics of proteins by a combinatorial peptide library approach, ChemBioChem. 12 (2011) 2846–2855. doi:10.1002/cbic.201100556.

[5]   A. Wißbrock, N.B. Goradia, A. Kumar, A.A. Paul George, T. Kühl, P. Bellstedt, R. Ramachandran, P. Hoffman, K. Galler, U. Neugebauer, K. Hampel, B. Zimmermann, S. Adams, M. Wiendl, G. Krönke, I. Hamza, S.H. Heinemann, S. Frey, A. Hueber, O. Ohlenschläger, D. Imhof, Structural insights into heme binding to IL-36α proinflammatory cytokine, Sci. Rep. 9 (2019) 16893. doi:10.1038/s41598-019-53231-0.

[6]   M.-T. Hopp, N. Alhanafi, A.A. Paul George, N.S. Hamedani, A. Biswas, J. Oldenburg, B. Pötzsch, D. Imhof, Molecular insights and functional consequence of the interaction of heme with activated protein C, (2019) submitted.

[7]   A. Wißbrock, T. Kühl, K. Silbermann, A.J. Becker, O. Ohlenschläger, D. Imhof, Synthesis and Evaluation of Amyloid β Derived and Amyloid β Independent Enhancers of the Peroxidase-like Activity of Heme, J. Med. Chem. 60 (2017) 373–385. doi:10.1021/acs.jmedchem.6b01432.

4.1.3 Summary

The established class system of heme-binding peptides consisting of classes I-VIII was successfully enriched by the creation of the four new subclasses A-D out of classes VI and VIII. 58 representatives of each of these subclasses, corresponding to more than 10 peptides for each subclass, were successfully synthesized and analyzed for heme binding. K_D values between 0.21 ± 0.17 μM and 8.86 ± 0.67 μM were obtained. Subclasses B and D, i.e. (H/Y)X(H/Y) and (H/Y)XXX(H/Y), contained the largest fraction of strong and intermediate binders. It was established that subclasses A, i.e. (H/Y)(H/Y), and C, i.e. (H/Y)XX(H/Y), are not favorable for heme binding and therefore do not contain suitable motifs. Structural analysis showed that subclasses A and C formed primarily pentacoordinated heme complexes, while subclasses B, i.e. (H/Y)X(H/Y), and D , i.e. (H/Y)XXX(H/Y), displayed mixtures of penta- and hexacoordination. NMR structures of two representatives of subclass B were solved and found to be relatively flexible in the free state, while in complex with heme pentacoordinate complexes emerged undergoing a slight structural stabilization. Interestingly, only three peptides of the set displayed a significant peroxidase activity. The motifs HXH, HXXXH, and HXXXY were identified as the most promising putative heme-binding motifs from the combined experimental data.

This manuscript lays the foundation for the following chapters, which expand on both the wealth of peptide data generated as well as the postulated motifs. In Chapter II, the peptides reported in this study can be utilized as training data for a machine learning algorithm, while in Chapter III and Chapter IV, the motifs and structural information are transferred to the protein level with the examples of JAK2 and TLR4.

4.2 Chapter II – HeMoQuest: a webserver for qualitative prediction of transient heme binding to protein motifs

Authors*

Ajay Abisheck Paul George, Mauricio Lacerda, Benjamin Franz Syllwasschy, Marie-Thérèse Hopp, Amelie Wißbrock, and Diana Imhof

This article was published in:

4.2.1 Introduction

The prediction of (transiently) heme-binding proteins is of great interest, since many heme-regulated proteins are presumably still unrecognized (see Section 2.4). Several attempts have been made to predict heme binding computationally, but none were based on data specifically generated for the subject of transient heme binding and many require protein structures as input[165–169]. However, such a distinction is necessary, since permanent and transient heme binding are vastly different in regard to their binding modes as well as structural and sequence requirements (see Section 0). In fact, most previously utilized training sets contain relatively homogenous data with almost two thirds of the hemoproteins being either hemoglobin or cytochrome variants[24]. Consequently, currently available heme-prediction algorithms are unable to correctly identify sites of transient heme interaction, achieving an accuracy of $< 60\%$ (see Section 2.4)[24]. A first algorithm predicting regulatory heme binding was published by the group of Prof. Imhof in 2019, termed SeqD-HBM[27]. This algorithm, however, was not available to a wide scientific audience, but rather used as an in-house tool[24].

It was therefore required, to generate a publicly available algorithm able to predict transient heme binding accurately. Prerequisite for this venture was the wealth of heme-binding peptides generated by previous studies[29–32,156] and in Chapter I. In this chapter the generation and validation of such an algorithm termed HeMoQuest will be described.

*Contributions:
AAPG and ML designed the HeMoQuest web application. MH, BFS and AW curated and prepared the peptide and heme-regulated protein datasets. AAPG, and DI conceived the presented idea. DI designed and planned the project. The manuscript was written through the contribution of all authors. The authors read and approved the final manuscript.

4.2.2 Article

On the following pages, the article is printed in its published form.

BMC Bioinformatics

RESEARCH ARTICLE

Open Access

HeMoQuest: a webserver for qualitative prediction of transient heme binding to protein motifs

Ajay Abisheck Paul George, Mauricio Lacerda, Benjamin Franz Syllwasschy, Marie-Thérèse Hopp, Amelie Wißbrock and Diana Imhof*

Abstract

Background: The notion of heme as a regulator of many physiological processes via transient binding to proteins is one that is recently being acknowledged. The broad spectrum of the effects of heme makes it important to identify further heme-regulated proteins to understand physiological and pathological processes. Moreover, several proteins were shown to be functionally regulated by interaction with heme, yet, for some of them the heme-binding site(s) remain unknown. The presented application HeMoQuest enables identification and qualitative evaluation of such heme-binding motifs from protein sequences.

Results: We present HeMoQuest, an online interface (http://bit.ly/hemoquest) to algorithms that provide the user with two distinct qualitative benefits. First, our implementation rapidly detects transient heme binding to nonapeptide motifs from protein sequences provided as input. Additionally, the potential of each predicted motif to bind heme is qualitatively gauged by assigning binding affinities predicted by an ensemble learning implementation, trained on experimentally determined binding affinity data. Extensive testing of our implementation on both existing and new manually curated datasets reveal that our method produces an unprecedented level of accuracy (92%) in identifying those residues assigned "heme binding" in all of the datasets used. Next, the machine learning implementation for the prediction and qualitative assignment of binding affinities to the predicted motifs achieved 71% accuracy on our data.

Conclusions: Heme plays a crucial role as a regulatory molecule exerting functional consequences via transient binding to surfaces of target proteins. HeMoQuest is designed to address this imperative need for a computational approach that enables rapid detection of heme-binding motifs from protein datasets. While most existing implementations attempt to predict sites of permanent heme binding, this application is to the best of our knowledge, the first of its kind to address the significance of predicting transient heme binding to proteins.

Keywords: Heme, Heme-regulated protein, Transient heme binding, Heme-binding site prediction, Web application, Machine learning

Background

Heme (iron protoporphyrin IX) is an astoundingly prevalent molecule found within humans, animals and plants, fulfilling a plethora of functions [1, 2]. It is the oxygen carrying moiety of hemoglobin, the gas-sensing molecule of NO-sensors and the redox active part of cytochromes [1, 2]. Besides its well-known binding to these hemoproteins as a prosthetic group, heme has been established as a biologically available molecule. Human targets of

* Correspondence: dimhof@uni-bonn.de
Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute, An der Immenburg 4, University of Bonn, 53121 Bonn, Germany

transient heme binding have been reviewed extensively [1–6] (Additional Table 1, supplementary data 3). Well-known representatives are δ-aminolevulinic acid synthase 1 (ALAS1) and transcription regulator protein Bach1, which bind heme via CP-containing motifs [1, 7, 8]. Recent reports have expanded the knowledge on transient heme binding. Some of the published heme-regulated proteins were shown to bind heme, but no information on the heme binding motif is available so far. A curated list of such proteins is available in additional Table 1. In this work we highlight the need for an exclusive computational method that is able to pinpoint heme-binding residues in protein sequences.

Despite the apparent abundance of interest in computational solutions to predict heme binding, all of the former approaches were focused on prediction of permanent heme binding as opposed to the prediction of transient heme binding and its associated regulatory function. In 2011 and 2012, the groups of Liu, Li and Xiong were the first to present publicly available heme-binding prediction algorithms [9–12]. These approaches had in common that they started from sets of published structures of hemoproteins. A large number of such structures is available, i.e. the gene ontology term "heme binding" (GO:0020037) currently yields over 4500 PDB structures. However, these structures are highly redundant and some are of low resolution, which was compensated by the authors by applying a cutoff at 25% or 30% sequence identity and at 3 Å resolution [9–12]. The resulting datasets were used to train machine learning algorithms based on structural features. In 2013, Yu et al. took the challenge to predict heme binding to proteins without available 3D structures, since all the so far known webservers were working with template-based methods. Therefore, the webserver "TargetS" (http://www.csbio.sjtu.edu.cn/ bioinf/TargetS/) was established to predict binding of ligands (i.a. heme) starting from primary sequences via a recursive spatial clustering algorithm. It included different aspects, such as evolutionary information, ligand-specific properties and secondary structure. A dataset of 233 structures of heme-binding proteins with a cutoff of 40% sequence identity was extracted from the BioLIP database [13], and used for training and testing the webserver. Derived from a scoring card method (SCM), the latest prediction method for heme binding to proteins "SCMHBP" benefits from an evaluation of heme-binding tendencies of 400 dipeptides and 20 amino acids, which is transferred onto protein sequences. Consequently, two non-redundant training datasets were designed with 747 heme-binding proteins and 747 non-heme-binding proteins, and two already existing datasets were taken into account for testing the SCMHBP, resulting in a mean accuracy of 85.9% [9, 10, 14]. In another approach, Zhang et al. clustered 4003 X-ray structures of heme-binding proteins via Blastclust [15] with a

sequence identity of less than 30% and selected 260 representatives for testing [16]. In addition, the training datasets from earlier studies were included [10, 17]. On this data, a novel predictor, i.e. "HEMEsPred" (http://www.inforstation.com/HEMEsPred/), was generated including sequence- and structure-based features, a fast-adaptive ensemble learning scheme and a more specific model for different heme ligands [16]. A summary of previous algorithms can be found in Fig. 1a.

However, all of these approaches employed relatively generic training data by merely querying available databases. Large parts of these training data consist of biologically redundant data such as different variants of hemoglobins (29.2%), and cytochromes (34.8%). Furthermore, previous studies were entirely focused on permanent heme binding and thus, disregarded transient, regulatory heme binding entirely. In contrast to hemoproteins, transient heme binding is believed to occur on protein surfaces and not in deep heme-binding pockets [1, 2, 6]. Strikingly, previous implementations have strongly focused on residues in the protein core, at times even specifically excluding surface exposed residues. As a consequence, all previous algorithms fail to predict transient heme-binding sites, and attain a maximum accuracy of less than 60% when challenged with the prediction of transient heme binding. The overarching aim of this work is to provide a computational tool exclusively developed for the prediction and qualitative evaluation of transient heme binding to protein surfaces based on our recently established SeqD-HBM algorithm [6].

Results

The first version of the algorithm was produced as a standalone Python script(s) wherein the SeqD-HBM algorithm for detecting heme-binding motifs was implemented only [6]. At this stage, the program was used as an in-house tool and was shared with users upon request. The full-fledged web application, in contrast, is a multi-fold improvement on the former version. It is built on the Django 2.3 framework (https://www.djangoproject.com/), which is known for its robustness, scalability and security. As shown in Fig. 1b, the usage of RabbitMQ and Celery for the management of user requests in the housekeeping module makes it evident that the application was built to scale well on heavy load. This is especially important since the application handles not only inbound and outbound user requests but also makes API calls to the (external) weighted ensemble solvent accessibility (WESA) [18, 19] application in the process of providing predictions. That being the case, additional measures had to be taken to constantly check all of the application end-points and update the status of a user request in the database. HeMoQuest was able to correctly predict residues (and the associated motifs)
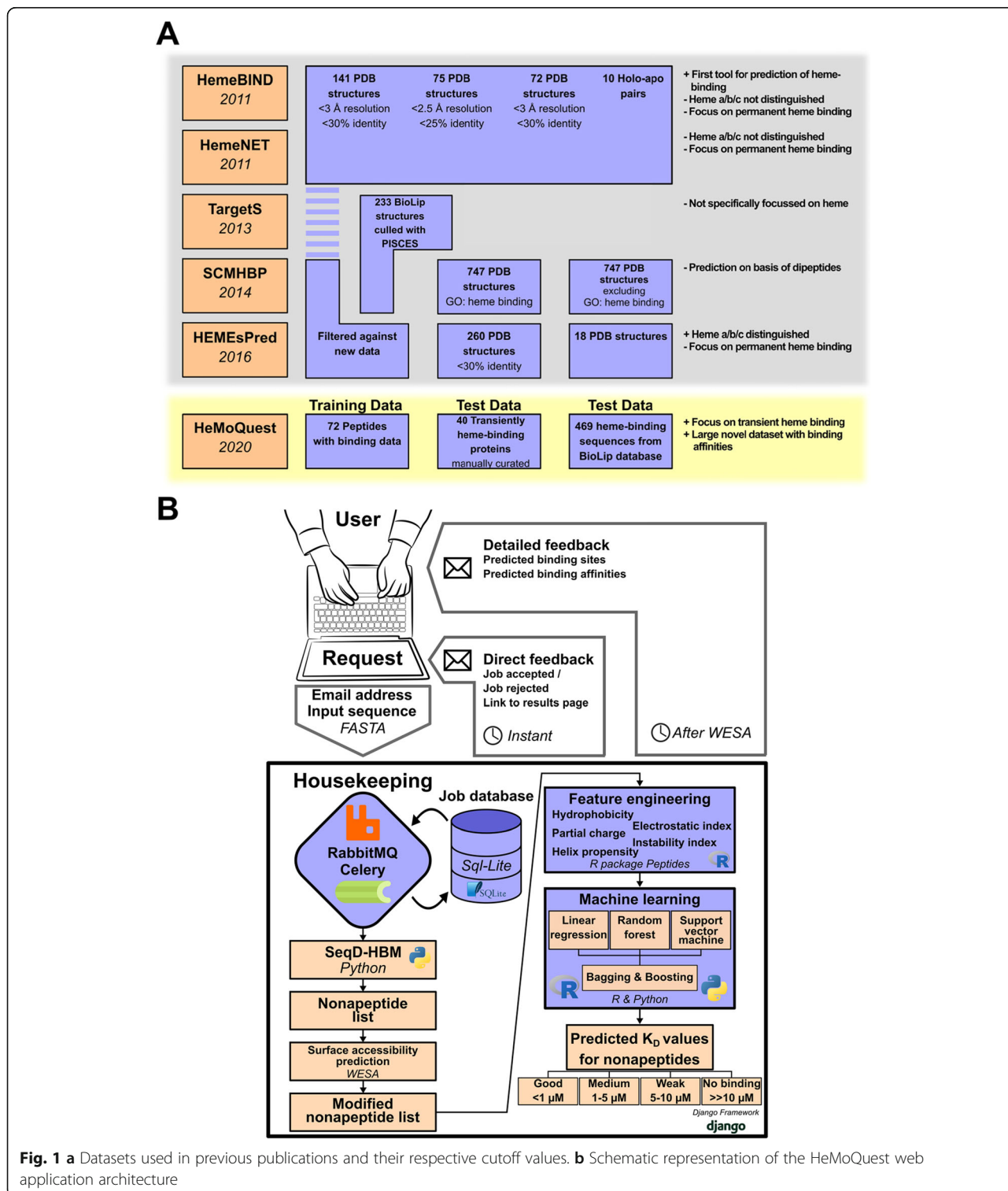
Fig. 1 a Datasets used in previous publications and their respective cutoff values. **b** Schematic representation of the HeMoQuest web application architecture

annotated as "heme-binding" with accuracies of 92.1% while using the WESA mode and 96.2% when WESA solvent accessibility was skipped from the 469 sequences gathered from BioLip (Supplementary data 2). In the WESA mode a small dip in accuracy was due to the fact

that most CP motifs were predicted to be buried by WESA. At the same time, though the non-WESA mode of operation an increase in accuracy was produced, the number of new potential motifs predicted increased by ~ 15%. However, it must be noted that these additionally

predicted motifs cannot be deemed as false positives with certainty since there exists no experimental evidence to prove their inability to transiently bind heme. Going beyond detecting heme-binding motifs from input sequences, the fact that a reasonable amount of in-house data with published binding affinities existed (Additional Table 2, supplementary data 1), provided the scope for a predictive machine-learning based solution to be implemented. In essence, the idea is to train the sequence data with known affinities on different predictive models so that a new input sequence could have the approximate binding affinities of their nonapeptide motifs predicted. We explicitly state here that the predicted affinities are not to be compared with the experimental values published at the protein level but are only to be compared among the different nonapeptides from the same sequence as an indicator of the best binding motif in the input sequence. Features from the input sequences were engineered by using the *Peptides* R package (https://github.com/dosorio/Peptides/). For any input sequence *Peptides* can produce over 100 different physiochemical values that define features such as hydrophobicity, instability, partial charge etc. (Supplementary data 4). It is infeasible to train a model on such a large number of features and hence the feature engineering consisted of choosing a reasonable number of diverse features that describe the input sequence. The Pearson correlation coefficient was calculated between all physiochemical features available in *Peptides* on the binding affinity data and the top 8 most diverse features that produced the best correlation were chosen as the final features for predictions. These features, which described hydrophobicity, instability, helix propensity, partial charge and electrostatics, were all sequence invariants, i.e. they are strongly dependent on the order of the amino acid residues in the sequence. All of the machine learning models were built using the Scikit-learn machine learning package [20].

Since the aim was to predict binding affinities, regression was used in all models built. First, a multivariate linear regression model was built using the Ordinary Least Squares method. Next, a random forest regressor, which is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting, was built. Finally, a regression model of a SVM was built using the Epsilon-Support Vector Regression method available in Scikit-learn. Predictions from the individual models were further subjected to an independent voting scheme to produce the best prediction. The AdaBoost and BaggingRegressor methods were used for the purpose, respectively.

The application was tested on the three independently collected datasets (see Methods). Overall, the heme-binding residue prediction module was able to successfully identify the predefined heme-coordinating residues and the associated motifs. Our method was able to successfully predict every single heme-coordinating residue and its associated motif for every single sequence tested. This is mainly due to the specific checks done on H, C, and Y residues explicitly on all the sequences. Consequently, the algorithm predicted more motifs than what is mentioned in the validation set. Though this results in an overall false positive rate of ~ 15%, one cannot be sure that these "additional" motifs are other potential heme-binding motifs, not accessible for prediction via experimental approaches as seen in earlier reports [21]. The prediction of affinities was again impressive with the support vector machine producing the best predictions of 71% accuracy on the training set. In terms of the qualitative classification, we were able to correctly classify with an overall accuracy of 68%, between the "good", "moderate", and "weak" binding motifs on the test set.

It was further observed that within the individual prediction algorithms (Fig. 1a), the linear regressor preformed the worst since it was clear that none of the physiochemical features used to describe the sequence data had strong correlation to the K_D values. This is a common pitfall for linearly uncorrelated features and is frequent in small datasets. However, the random forest and SVM predictions outweighed poor performance of the linear regression, thereby producing overall predictions of acceptable quality.

Discussion

Under hemolytic conditions, such as thalassemia, sickle cell disease, or distinct bacterial infections, red blood cells are destroyed and release both hemoglobin and heme [3]. Heme is initially bound and neutralized by heme scavengers such as hemopexin and albumin, but once their capacity is exceeded, vast amounts of free heme arise [22]. Heme can consequently bind and regulate a number of proteins, for most of which the interaction site is unknown. For example, heme has been suggested to inhibit the classical complement pathway by interaction with C1q [23] and to activate the alternative complement pathway by deposition of C3 [24]. Using Surface Plasmon Resonance spectroscopy, the heme dissociation constant of C1q was determined to be approximately 1–2 μM, but no binding site could be identified yet [25]. The latter applies to C3 as well, yet molecular docking suggested a binding site close to the functionally important thioester bond [24]. Furthermore, heme influences factor VIII and fibrinogen in seemingly contradictory fashion, but partially due to the lack of structural information this dissonance has not been unraveled yet [26–28].

The opposite approach, i.e. the prediction of unknown heme-regulated proteins from peptide sequences, has also been fruitful. Building on a combinatorial peptide library screening approach we predicted and validated transient

heme binding to proteins such as chloramphenicol acetyltransferase, hemolysin C, and interleukin-36α [6, 29–31].

Conclusion

As demonstrated with evidence thus far, heme plays a key regulatory role in a multitude of processes. Strikingly, for most of the proteins involved, little is known about the sites and affinities of heme binding. Experimental mapping of heme-protein interactions requires vast effort, e.g. the expression of protein mutants or the co-crystallization of heme with the protein of interest, at times with conflicting results [8, 32]. As an extension of existing experimental work on peptide-heme binding, HeMoQuest provides a shortcut to the identification of heme-regulated proteins. Due to the convincing accuracy of the presented algorithm, researchers may be able to bypass the necessity of producing peptides as models and may even be able to rationally design heme-binding peptides and proteins. Several efforts have been undertaken to predict heme binding to proteins [9–12, 14, 16, 17, 33], however, HeMoQuest is fundamentally different from previous tools because of its exclusive focus on transient heme binding. The tool is built on a dataset created solely for this purpose. As with any data-driven approach, HeMoQuest is poised to only get better with time as more data becomes available.

Methods

Web application architecture

Multiple programming languages and frameworks were effectively utilized in this work to construct a user-friendly and effective web application, called HeMoQuest (heme-binding motif quest). This webserver (Fig. 1b) was built on the Django framework version 2.3.1 (https://www.django-project.com/) running Python 3.6.5 under the hood. The user is given access to three pages: 1) the landing page that can also be used to submit sequences for analysis, 2) the results page containing the analysis of a single request in a tabular format, and 3) the analysis page hyperlinked from the results providing a stepwise analysis of how a prediction was produced for an input sequence. User input and the analysis results are saved in a SQlite database (https://www.sqlite.org/index.html). The schema consists of three tables namely: 'jobs', 'sequences' and 'results'. The job is saved whenever the user sends enough information to be processed, such as a file or an input sequence. The sequences table stores for each job all sequences successfully read (even if there are false inputs). The results table contains the predicted nonapeptide motifs and the predicted binding affinities.

Web application control flow

Django first checks user submissions for the following: There must be at least one sequence or one file in the FASTA format. File sizes are restricted to 2 MB. If solvent accessibility prediction via WESA, i.e. the Weighted Ensemble Solvent Accessibility predictor tool [18, 19], is requested, an email address is required to be entered by the user. With the basic checks done, the input is read and a status (either "failed", "queued" or "processed") is assigned. Each sequence is analyzed and the possible binding sites are saved in the database. If solvent accessibility prediction was not requested, the analysis for each sequence is generated and saved, the status is changed to "processed". If an email address was supplied, a message is sent with the link to the results page and finally the user is redirected to the results page. If solvent accessibility prediction was requested, the analysis for each sequence is also generated and saved, under the status "queued". The user will receive an email with uniform resource locator (URL) link to the results page that displays the status of how many sequences are being processed and, as the sequences are processed, their results. This process will continue as long as the celery job detects the existence of queued sequences with the WESA detection mode. The task scheduler takes care of the automatic execution of this process.

Prediction algorithms

The nonapeptide motif prediction algorithm SeqD-HBM published earlier [6] as a standalone python script was overhauled and rewritten for the web application. For the prediction of binding affinities, a set of 73 nonapeptide sequences (Additional Table 1) synthesized and validated in-house with experimentally determined and published binding affinities were used [29, 30, 34–36]. Three different predictors namely, linear regression, random forest and support vector machine (SVM) methods were used to perform regression, predicting the target variable, i.e. the binding affinity. An ensemble-based voting scheme was used to obtain the final prediction from the individual predictors (Fig. 1b). An 80/20 train-test split was employed in all cases.

Dataset preparation

Three independent datasets were used for this study. The first of which comprised of 469 sequences (supplementary data 2) from the BioLip database [13] (March 2019 release) extracted for the HEM ligand code which relates to "heme". The second dataset consisted of a cumulative set of all of the data used in previous studies (Fig. 1b). Finally, a set of 40 proteins (Additional Table 2, supplementary data 3) known to bind heme transiently, was chosen manually.

Supplementary information

> **Additional file 1:.** Additional Table 1. Human heme-regulated proteins and reported heme-binding sites. Additional Table 2. Peptides sequences and binding data used for the initial training of HeMoQuest

Additional file 2 Supplementary data Additionally, the datasets used to train and validate the application is also available for download from the "HeMoQuest Datasets" section of the webserver. A description of the files provided is given below. 1. Training data. 1a. Title: HeMoQuest K_D prediction training data. 1b. Description: This comma separated values file contains 72 sequences along with their K_D values used in the training of the ML algorithms of HeMoQuest. Column 1 (ID) contains a sequence identifier, column 2 (Seq) contains the sequence and column 3 (K_D) contains the experimentally determined K_D value of for the peptide sequence. 2. Test data. 2a. Title: HeMoQuest test data for heme binding residue and motif prediction. 2b. Description: This file contains 469 sequences in fasta format, obtained from the BioLip database, all of which are said to bind heme. This data was used to test HeMoQuest's ability to detect heme binding residues in comparison to existing algorithms. 3. Test data. 3a. Title: HemoQuest test data with manually curated transient heme binding protein sequences. 3b. Description: This file contains 45 sequences in fasta format from 40 manually curated proteins (from Additional Table 1) that are known from literature to be transient heme binding proteins. Few of the proteins have their origins in more than one species and hence we end up with 45 sequences for 40 proteins. 4. Training features. 4a. Title: Features used in training the HeMoQuest K_D prediction. 4b. Description: This comma separated values file contains 76 initial features that were generated for the K_D prediction training from the R package *Peptides*. The final set of features used are from the columns 'charge_vec', 'hydrof_vec_octanolScale_pH8', 'acidic', 'kideraFac3', 'vhseScale5_vec', 'vhseScale7_vec', 'protFP5_vec' and 'fasgaiVec4'.

Abbreviations
ALAS: Aminolevulinic acid synthase; NO: Nitric oxide; SeqD-HBM: Sequence based detection of heme-binding motifs; SVM: Support vector machine; URL: Uniform resource locator; WESA: Weighted ensemble solvent accessibility

Authors' contributions
AAPG and ML designed the HemoQuest web application. MH, BS and AW curated and prepared the HRP datasets. AAPG, and DI conceived the presented idea. DI designed and planned the project. The manuscript was written through the contribution of all authors. The authors read and approved the final manuscript.

Availability of data and materials
The HeMoQuest webserver is freely available at http://bit.ly/hemoquest All the data used in this study are available in additional files 1 and 2.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Zhang L. Heme biology: the secret life of heme in regulating diverse biological processes. World scientific; 2011. https://doi.org/10.1142/7484.
2. Kühl T, Imhof D. Regulatory FeII/III heme: the reconstruction of a molecule's biography. ChemBioChem. 2014;16:2024–35.
3. Roumenina LT, Rayes J, Lacroix-Desmazes S, Dimitrov JD. Heme: modulator of plasma systems in hemolytic diseases. Trends Mol Med. 2016;22:200–13. https://doi.org/10.1016/j.molmed.2016.01.004.
4. Gozzelino R. The pathophysiology of Heme in the brain. Curr Alzheimer Res. 2016;13:174–84 http://www.ncbi.nlm.nih.gov/pubmed/26391040.
5. Comer J, Zhang L. Experimental methods for studying cellular Heme signaling. Cells. 2018;7:47.
6. Wißbrock A, Paul George AA, Brewitz HH, Kühl T, Imhof D. The molecular basis of transient heme-protein interactions: analysis, concept and implementation. Biosci Rep. 2019;39:BSR20181940. https://doi.org/10.1042/bsr20181940.
7. Lathrop JT, Timko MP. Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. Science (80-). 1993;259: 522–5. https://doi.org/10.1126/science.8424176.
8. Satoh T, Satoh H, Iwahara SI, Hrkal ZZ, Peyton DH, Muller-Eberhard U. Roles of heme iron-coordinating histidine residues of human hemopexin expressed in baculovirus-infected insect cells. Proc Natl Acad Sci U S A. 1994;91:8423–7.
9. Liu R, Hu J. HemeBIND: a novel method for heme binding residue prediction by combining structural and sequence information. BMC Bioinformatics. 2011;12:207.
10. Liu R, Hu J. Computational prediction of heme-binding residues by exploiting residue interaction network. PLoS One. 2011;6:e25560.
11. Li T, Bonkovsky HL, Guo J. Structural analysis of heme proteins: implications for design and prediction. BMC Struct Biol. 2011;11:13.
12. Xiong Y, Liu J, Zhang W, Zeng T. Prediction of heme binding residues from protein sequences with integrative sequence profiles. Proteome Sci. 2012; 10(Suppl 1):1–8.
13. Yang J, Roy A, Zhang Y. BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. Nucleic Acids Res. 2013;41:D1096–103.
14. Liou YF, Charoenkwan P, Srinivasulu YS, Vasylenko T, Lai SC, Lee HC, et al. SCMHBP: prediction and analysis of heme binding proteins using propensity scores of dipeptides. BMC Bioinformatics. 2014;15:1–14.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
16. Zhang J, Chai H, Gao B, Yang G, Ma Z. HEMEsPred: structure-based ligand-specific Heme binding residues prediction by using fast-adaptive ensemble learning scheme. IEEE/ACM Trans Comput Biol Bioinforma. 2016;15:147–56.
17. Fufezan C, Zhang J, Gunner MR. Ligand preference and orientation in b- and c-type heme-binding proteins. Proteins Struct Funct Genet. 2008;73: 690–704. https://doi.org/10.1002/prot.22097.
18. Chen H, Zhou HX. Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. Nucleic Acids Res. 2005;33:3193–9.
19. Shan Y, Wang G, Zhou HX. Fold recognition and accurate query-template alignment by a combination of PSI-BLAST and threading. Proteins Struct Funct Genet. 2001;42:23–37.
20. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–30 http://arxiv.org/abs/1201.0490.
21. Peherstorfer S, Brewitz HH, Paul George AA, Wißbrock A, Adam JM, Schmitt L, et al. Insights into mechanism and functional consequences of heme binding to hemolysin-activating lysine acyltransferase HlyC from Escherichia coli. Biochim Biophys Acta - Gen Subj. 1862;2018:1964–72.
22. Chiabrando D, Vinchi F, Fiorito V, Mercurio S, Tolosano E. Heme in pathophysiology: a matter of scavenging, metabolism and trafficking across cell membranes. Front Pharmacol. 2014;5:1–24.
23. Roumenina LT, Radanova M, Atanasov BP, Popov KT, Kaveri SV, Lacroix-Desmazes S, et al. Heme interacts with C1q and inhibits the classical complement pathway. J Biol Chem. 2011;286:16459–69.
24. Frimat M, Tabarin F, Dimitrov JD, Poitou C, Halbwachs-Mecarelli L, Fremeaux-Bacchi V, et al. Complement activation by heme as a secondary hit for atypical hemolytic uremic syndrome. Blood. 2013;122:282–92.
25. Roumenina LT, Sène D, Radanova M, Blouin J, Halbwachs-Mecarelli L, Dragon-Durey M-A, et al. Functional complement C1q abnormality leads to impaired immune complexes and apoptotic cell clearance. J Immunol. 2011;187:4369–73.
26. Repessé Y, Dimitrov JD, Peyron I, Moshai EF, Kiger L, Dasgupta S, et al. Heme binds to factor VIII and inhibits its interaction with activated factor IX. J Thromb Haemost. 2012;10:1062–71.
27. Orino K. Functional binding analysis of human fibrinogen as an iron- and heme-binding protein. BioMetals. 2013;26:789–94.
28. Ke Z, Huang Q. Haem-assisted dityrosine-cross-linking of fibrinogen under non-thermal plasma exposure: one important mechanism of facilitated blood coagulation. Sci Rep. 2016;6:1–8. https://doi.org/10.1038/srep26982.

29. Brewitz HH, Goradia N, Schubert E, Galler K, Kühl T, Syllwasschy B, et al. Heme interacts with histidine- and tyrosine-based protein motifs and inhibits enzymatic activity of chloramphenicol acetyltransferase from Escherichia coli. Biochim Biophys Acta - Gen Subj. 1860;2016:1343–53.

30. Brewitz HH, Kühl T, Goradia N, Galler K, Popp J, Neugebauer U, et al. Role of the chemical environment beyond the coordination site: structural insight into FeIII Protoporphyrin binding to cysteine-based Heme-regulatory protein motifs. ChemBioChem. 2015;16:2216–24.

31. Wißbrock A, Goradia NB, Kumar A, Paul George AA, Kühl T, Bellstedt P, et al. Structural insights into heme binding to IL-36α proinflammatory cytokine. Sci Rep. 2019;9(1):16893.

32. Paoli M, Anderson BF, Baker HM, Morgan WT, Smith A, Baker EN. Crystal structure of hemopexin reveals a novel high-affinity heme site formed between two β-propeller domains. Nat Struct Biol. 1999;6:926–31.

33. Yu DJ, Hu J, Yang J, Bin SH, Tang J, Yang JY. Designing template-free predictor for targeting protein-ligand binding sites with classifier ensemble and spatial clustering. IEEE/ACM Trans Comput Biol Bioinforma. 2013;10:994–1008.

34. Kühl T, Sahoo N, Nikolajski M, Schlott B, Heinemann SH, Imhof D. Determination of hemin-binding characteristics of proteins by a combinatorial peptide library approach. ChemBioChem. 2011;12:2846–55.

35. Kühl T, Wißbrock A, Goradia N, Sahoo N, Galler K, Neugebauer U, et al. Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins. ACS Chem Biol. 2013;8:1785–93.

36. Wißbrock A, Kühl T, Silbermann K, Becker AJ, Ohlenschläger O, Imhof D. Synthesis and evaluation of amyloid β derived and amyloid β independent enhancers of the peroxidase-like activity of Heme. J Med Chem. 2017;60:373–85.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

HeMoQuest: A webserver for qualitative prediction of transient heme binding to protein motifs

Ajay Abisheck Paul George[1], Mauricio Lacerda[1], Benjamin Franz Syllwasschy[1], Marie-Thérèse Hopp[1], Amelie Wißbrock[1], and Diana Imhof[1,*]

[1]Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute, An der Immenburg 4, University of Bonn, 53121 Bonn, Germany

*To whom correspondence should be addressed: Diana Imhof (Email: dimhof@uni-bonn.de).

Additional information

Additional Table 1. Human heme-regulated proteins and reported heme-binding sites

Number	Protein	Origin	Published heme-binding sites	References	HeMoQuest Prediction*	Comment
1	δ-aminolevulinic acid synthase 1 (ALAS1)	*R. norvegicus/ H. sapiens*	RC^{8}PFLS NC^{33}PKMM KC^{108}PFLA	[1–3]	All published sites correctly predicted.	CP motifs missed in the WESA mode.
2	δ-aminolevulinic acid synthase 2 (ALAS2)	*M. musculus/ H. sapiens*	C^{11}PVLA C^{38}PILA	[1, 4]	All published sites correctly predicted.	CP motifs missed in the WESA mode
3	Amyloid β	*H. sapiens/ rodents*	Y^{10}EVH^{13}H^{14}	[5–9]	Published site correctly predicted.	Used as training data.
4	Arginyl-Transferase (ATE1)	*M. musculus/ S. cerevisiae*	HSC^{411}P	[10]	Published site correctly predicted.	CP motifs missed in the WESA mode.
5	Arginyl-tRNA-Synthetase (ArgRS)	*H. sapiens*	QC^{115}NSAM	[11]	All published sites correctly predicted.	-
6	Bach1	*M. musculus*	EC^{438}PWLG NC^{464}PFIS PCP^{495}YAC DC^{649}PLSF	[12–15]	All published sites correctly predicted.	CP motifs missed in the WESA mode.
7	Bach2	*H. sapiens*	C^{369}P C^{499}P C^{506}P	[16–18]	All published sites correctly predicted.	CP motifs C^{499}P C^{506}P missed and C^{369}P predicted correctly in WESA mode.
8	BK channel	*H. sapiens*	C^{612}XXCH^{616}	[19–23]	Published sites correctly predicted.	-
9	C1q	*H. sapiens*	not known	[24, 25]	Up to 6 potential	-

					motifs predicted	
10	C3(a/b)	*H. sapiens*	not known	[26]	20 potential motifs predicted	-
11	CLOCK	*M. musculus*	H^{144} C^{195}	[27]	All published sites correctly predicted.	-
12	Cystathionine-β-synthase (CBS)	*H. sapiens*	C^{15}PHRSGPH C^{52} H^{65}	[28–31]	All published sites correctly predicted.	-
13	DiGeorge critical region 8 (DGCR8)	*H. sapiens*	$IPC^{352}L$ (from two subunits)	[32–34]	All published sites correctly predicted.	-
14	Dipeptidyl peptidase 8 (DPP8)	*H. sapiens*	$SDFKC^{506}PIKE$	[35]	All published sites correctly predicted.	-
15	Eukaryotic translation initiation factor 2a kinase 1 (eIF2α, HRI)	*M. musculus*	$AC^{167}PYVM$ $RC^{309}PVQA$	[36–38]	All published sites correctly predicted.	-
16	Anti-hemophilic factor (Factor VIII)	*H. sapiens*	not known	[39, 40]	99 potential motifs predicted.	WESA could not be applied since the sequence is over 2000 residues.
17	Fibrinogen (Factor I)	*H. sapiens*	not known	[41, 42]	12 potential motifs predicted.	-
18	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	*H. sapiens*	not known	[43–47]	3 potential motifs predicted.	-
19	Heme oxygenase 2	*R. norvegicus/ H. sapiens*	$XH^{45}X$ $KC^{265}PFYA$ (+H^{256}) $SC^{282}PFRT$	[48–51]	All published sites correctly predicted.	-
20	Human period circadian protein homolog 2 (hPer2)	*H. sapiens*	$SC^{841}PA$ $AC^{962}PA$	[52]	All published sites correctly predicted.	-
21	Immunoglobin (IgG)	*H. sapiens*	not known	[53]	8 potential sites predicted	-
22	Interleukin-36α	*H. sapiens*	$SEGGC^{136}PLIL$ $FLFY^{108}HSQSG$	[54]	All published sites correctly predicted.	CP motifs missed by WESA mode.
23	Iron regulatory protein 1 (IRP1)	*H. sapiens*	$C^{118}P$ $C^{300}P$	[55, 56]	All published sites correctly predicted.	CP motifs missed by WESA mode.
24	Iron regulatory protein 2 (IRP2)	*H. sapiens*	$C^{201}PXH^{204}XXXXP$ $C^{375}P$	[55–58]	All published sites correctly predicted.	CP motifs missed by WESA mode.
25	Janus kinase 2 (JAK2)	*H. sapiens*	$RPDGC^{1094}PDEI$	[59]	All published sites	CP motifs missed by

					correctly predicted.	WESA mode.
26	K$_{ATP}$ channels	*R. norvegicus*	C^{628}XXH(X)$_{16}$H^{648}	[60]	All published sites correctly predicted.	-
27	Neudesin	*M. musculus*	not known	[61]	Up to 5 potential motifs predicted.	-
28	Neuferricin	*M. musculus*	not known	[62]	6 potential motifs predicted.	-
29	NMDA Receptor	*M. musculus*	not known	[63, 64]	33 potential motifs predicted.	-
30	p53	*H. sapiens*	C^{275}AC^{277}P	[65, 66]	All published sites correctly predicted.	CP motifs missed by WESA mode.
31	p63	*H. sapiens*	CXCP	[65]	All published sites correctly predicted.	CP motifs missed by WESA mode.
32	p73	*H. sapiens*	CXCP	[65]	All published sites correctly predicted.	CP motifs missed by WESA mode.
33	PGRMC1/Sigma-2 receptor	*H. sapiens*	YGPEGPY^{113}GVFA	[67, 68]	All published sites correctly predicted.	-
34	Rev-erbα	*H. sapiens*	H^{602}	[69–71]	All published sites correctly predicted.	-
35	Rev-erbβ	*H. sapiens*	C^{384} H^{568}	[70, 72–75]	All published sites correctly predicted.	-
36	Src	*H. sapiens*	not known (CP motif presumed)	[59]	Five potential motifs predicted.	Predicted CP motif C^{486}P.
37	Stanniocalcin-1	*H. sapiens*	RC^{114}STFQ	[76]	All published sites correctly predicted.	-
38	Stanniocalcin-2	*H. sapiens*	SRKC^{125}PAIREM	[77]	All published sites correctly predicted.	-
39	Toll like Receptor 4 (TLR4)	*H. sapiens*	not known	[78]	13 potential sites predicted.	-
40	Tryptophanyl-tRNA-Synthetase	*H. sapiens*	H^{130} (H^{129}, H^{336}, H^{445} presumed)	[79]	All published sites correctly predicted.	-

*Sometimes, more potential binding sites than the reported ones were found.

Additional Table 2. Peptides sequences and binding data used for the initial training of HeMoQuest

Number	Sequence*	K_D [µM]	Reference
1	RAFFYCKAC	0.28 ± 0.09	[80]
2	AFFYCKACH	0.39 ± 0.09	[80]
3	AESFCTNQD	2.23 ± 1.86†	[81]
4	KWEDCVFDS	1.78 ± 0.78	[81]
5	LWQFCAFSS	0.79 ± 0.22	[81]
6	WRSNCSHQH	12.03 ± 0.59†	[81]
7	VEKMCIETA	5.56 ± 1.07	[81]
8	AQEPHLEBK	100 ± 0.00	[82]
9	LINQHATSF	6.23 ± 1.35 (n.sat.)	[82]
10	ETIFHTVQQ	100 ± 0.00	[82]
11	YPGQHISNE	100 ± 0.00	[82]
12	BVQLHKHSG	2.41 ± 0.73	[82]
13	RQRDHQBYA	3.38 ± 0.68	[82]
14	PRQAHVYRA	4.78 ± 1.27	[82]
15	IBFRYSSLK	0.34 ± 0.26	[82]
16	EKBWYWAEA	100 ± 0.00	[82]
17	EBRPYETGD	100 ± 0.00	[82]
18	LLPYYVQED	100 ± 0.00	[82]
19	GFGTYSWHE	6.25 ± 1.44	[82]
20	PPKTYQGEG	100 ± 0.00	[82]
21	DTPDYDYSR	100 ± 0.00	[82]
22	HADTYFGWR	19.01 ± 1.76 (n.sat.)	[82]
23	ITSIYNGAQ	100 ± 0.00	[82]
24	LRAVYEKDA	2.48 ± 0.48 (n.sat.)	[82]
25	FKQYHHELI	2.31 ± 0.69†	[82]
26	EQREHANVI	100 ± 0.00	[82]
27	HDENYTPPE	100 ± 0.00	[82]
28	KPFKYDHHY	1.96 ± 0.72	[82]
29	TLDLHLEVS	100 ± 0.00	[82]
30	AAAAAAAAA	100 ± 0.00	[9]
31	RHDSGYEVHH	80.70 ± 4.00	[9]
32	DAEFRHDSGYEVHHQKLV	0.59 ± 0.05	[9]
33	NVNLTSNALLYHYWIAVSHKAPA	3.00 ± 1.19	[9]
34	NVNLTSNHLLYHYWIAVSAKAPA	1.89 ± 0.72	[9]
35	NVNLTSNALLYHYWIAVSAKAPA	0.88 ± 0.94	[9]
36	NVNLTSNHLLYAYWIAVSHKAP	19.84 ± 4.11	[9]
37	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV	0.10 ± 0.07	[9]
38	AEFRHDSGY	100 ± 0.00	[9]
39	HDSGYEVHH	100 ± 0.00	[9]
40	RHLPCDICV	0.13 ± 0.09	[83]
41	FYWDCNHYW	0.67 ± 0.12	[80]
42	RADICVHLN	0.40 ± 0.17	[80]
43	SGGLPAPSDFKCPIKEEIAITSG	1.42 ± 0.24	[80]
44	AAAACAAAA	100 ± 0.00	[80]
45	QFSQCRIBN	2.78 ± 0.78	[80]
46	YVSRCIBBA	3.44 ± 0.64	[80]
47	HHQYHARVA	0.87 ± 0.05	[80]
48	AAAAHAAAA	100 ± 0.00	[80]
49	AAAAYAAAA	100 ± 0.00	[80]
50	HPFPYIWKA	0.33 ± 0.25	[80]
51	NVNLTSNHLLYHYWIAVSHKAPA	1.40 ± 0.13	[80]
52	VRMDTLAHVLYYPQKPLVTTRSM	0.51 ± 0.27	[80]
53	APSRCTQWL	4.41 ± 0.80	[35]
54	SSIPCLFYK	0.28 ± 0.19	[35]
55	SQSSCPAVP	6.43 ± 0.53	[35]
56	DESACPVYM	13.26 ± 1.44	[35]
57	RPDGCPDEI	0.50 ± 0.23	[35]
58	EDKDCPIKE	48.40 ± 2.73	[35]
59	ALTGCPWHD	5.32 ± 0.55	[35]
60	QFSSCPHYW	1.94 ± 0.42	[35]
61	IGVVCPFVR	0.87 ± 0.41	[35]

62	ARLGCPVIP	1.37 ± 0.33	[35]
63	SEGGCPLIL	3.75 ± 0.77	[35, 54]
64	TPILCPFHL	0.60 ± 0.41	[35]
65	AAAACPAAA	4.73 ± 1.83	[35]
66	SSIPCLHYK	0.81 ± 0.51	[35]
67	DESACPYVM	3.27 ± 1.44	[35]
68	AIRRCSTFQ	0.50 ± 0.20	[81]
69	HELVCAAST	0.40 ± 0.19	[81]
70	QKGVCQNTG	2.25 ± 0.70	[81]
71	AAHYHTYER	0.83 ± 0.33	[82]
72	FKAAHKHVR	0.99 ± 0.21	[82]
73	FLFYHSQSG	4.48 ± 2.20	[54]

*B, norleucine as substituent for methionine; [†]these K_D values have been reevaluated

References

1. Lathrop JT, Timko MP. Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. Adv Sci. 1993;259:522–6.

2. Munakata H, Sun JY, Yoshida K, Nakatani T, Honda E, Hayakawa S, et al. Role of the heme regulatory motif in the heme-mediated inhibition of mitochondrial import of 5-aminolevulinate synthase. J Biochem. 2004;136:233–8.

3. Kubota Y, Nomura K, Katoh Y, Yamashita R, Kaneko K, Furuyama K. Novel mechanisms for heme-dependent degradation of ALAS1 protein as a component of negative feedback regulation of heme biosynthesis. J Biol Chem. 2016;291:20516–29.

4. Goodfellow BJ, Dias JS, Ferreira GC, Henklein P, Wray V, Macedo AL. The solution structure and heme binding of the presequence of murine 5-aminolevulinate synthase. FEBS Lett. 2001;505:325–31.

5. Atamna H, Frey WH. A role for heme in Alzheimer's disease: Heme binds amyloid β and has altered metabolism. Proc Natl Acad Sci U S A. 2004;101:11153–8.

6. Atamna H, Frey WH, Ko N. Human and rodent amyloid-β peptides differentially bind heme: Relevance to the human susceptibility to Alzheimer's disease. Arch Biochem Biophys. 2009;487:59–65.

7. Pramanik D, Dey SG. Active site environment of heme-bound amyloid β peptide associated with Alzheimers Disease. J Am Chem Soc. 2011;133:81–7.

8. Zhou Y, Wang J, Liu L, Wang R, Lai X, Xu M. Interaction between amyloid-β peptide and heme probed by electrochemistry and atomic force microscopy. ACS Chem Neurosci. 2013;4:535–9.

9. Wißbrock A, Kühl T, Silbermann K, Becker AJ, Ohlenschläger O, Imhof D. Synthesis and evaluation of amyloid β derived and amyloid β independent enhancers of the peroxidase-like activity of heme. J Med Chem. 2017;60:373–85.

10. Hu R-G, Wang H, Xia Z, Varshavsky A. The N-end rule pathway is a sensor of heme. Proc Natl Acad Sci U S A. 2008;105:76–81.

11. Yang F, Xia X, Lei HY, Wang ED. Hemin binds to human cytoplasmic arginyl-tRNA synthetase and inhibits its catalytic activity. J Biol Chem. 2010;285:39437–46.

12. Ogawa K, Sun J, Taketani S, Nakajima O, Nishitani C, Sassa S, et al. Heme mediates derepression of Maf recognition element through direct binding to transcription repressor Bach1. EMBO J. 2001;20:2835–43.

13. Hira S, Tomita T, Matsui T, Igarashi K, Ikeda-Saito M. Bach1, a heme-dependent transcription factor, reveals presence of multiple heme binding sites with distinct coordination structure. IUBMB Life. 2007;59:542–51.

14. Zenke-Kawasaki Y, Dohi Y, Katoh Y, Ikura T, Ikura M, Asahara T, et al. Heme induces ubiquitination and degradation of the transcription factor Bach1. Mol Cell Biol. 2007;27:6962–71.

15. Segawa K, Watanabe-Matsui M, Matsui T, Igarashi K, Murayama K. Functional heme binding to the intrinsically disordered C-terminal region of bach1, a transcriptional repressor. Tohoku J Exp Med. 2018;247:153–9.

16. Watanabe-Matsui M, Muto A, Matsui T, Itoh-Nakadai A, Nakajima O, Murayama K, et al. Heme regulates B-cell differentiation, antibody class switch, and heme oxygenase-1 expression in B cells as a ligand of Bach2. Blood. 2011;117:5438–48.

17. Watanabe-Matsui M, Matsumoto T, Matsui T, Ikeda-Saito M, Muto A, Murayama K, et al. Heme binds to an intrinsically disordered region of Bach2 and alters its conformation. Arch Biochem Biophys. 2015;565:25–31.

18. Suenaga T, Watanabe-Matsui M, Uejima T, Shima H, Matsui T, Ikeda-Saito M, et al. Charge-state-distribution analysis of Bach2 intrinsically disordered heme binding region. J Biochem. 2016;160:291–8.

19. Tang XD, Xu R, Reynolds MF, Garcia ML, Heinemann SH, Hoshi T. Haem can bind to and inhibit mammalian calcium-dependent Slo1 BK channels. Nature. 2003;425:531–5.

20. Williams SEJ, Wootton P, Mason HS, Bould J, Iles DE, Riccardi D, et al. Hemoxygenase-2 is an oxygen sensor for a calcium-sensitive potassium channel. Science (80-). 2004;306:2093–7.

21. Horrigan FT, Heinemann SH, Hoshi T. Heme regulates allosteric activation of the Slo1 BK channel. J Gen Physiol. 2005;126:7–21.

22. Jaggar JH, Li A, Parfenova H, Liu J, Umstot ES, Dopico AM, et al. Heme is a carbon monoxide receptor for large-conductance Ca 2+-activated K+ channels. Circ Res. 2005;97:805–12.

23. Yi L, Morgan JT, Ragsdale SW. Identification of a thiol/disulfide redox switch in the human BK channel that controls its affinity for heme and CO. J Biol Chem. 2010;285:20117–27.

24. Dimitrov JD, Roumenina LT, Doltchinkova VR, Vassilev TL. Iron ions and haeme modulate the binding properties of complement subcomponent C1q and of immunoglobulins. Scand J Immunol. 2007;65:230–9.

25. Roumenina LT, Radanova M, Atanasov BP, Popov KT, Kaveri S V., Lacroix-Desmazes S, et al. Heme interacts with C1q and inhibits the classical complement pathway. J Biol Chem. 2011;286:16459–69.

26. Frimat M, Tabarin F, Dimitrov JD, Poitou C, Halbwachs-Mecarelli L, Fremeaux-Bacchi V, et al. Complement activation by heme as a secondary hit for atypical hemolytic uremic syndrome. Blood. 2013;122:282–92.

27. Lukat-Rodgers GS, Correia C, Botuyan MV, Mer G, Rodgers KR. Heme-based Sensing by the Mammalian Circadian Protein, CLOCK. Inorg Chem. 2010;49:6349–65.

28. Meier M, Janosik M, Kery V, Kraus JP, Burkhard P. Structure of human cystathionine β-synthase: A unique pyridoxal 5′-phosphate-dependent heme protein. EMBO J. 2001;20:3910–6.

29. Taoka S, Lepore BW, Kabil Ö, Ojha S, Ringe D, Banerjee R. Human cystathionine β-synthase is a heme sensor protein. Evidence that the redox sensor is heme and not the vicinal cysteines in the CXXC motif seen in the crystal structure of the truncated enzyme. Biochemistry. 2002;41:10454–61.

30. Weeks CL, Singh S, Madzelan P, Banerjee R, Spiro TG. Heme regulation of human cystathionine β-synthase activity: Insights from fluorescence and Raman spectroscopy. J Am Chem Soc. 2009;131:12809–16.

31. Kumar A, Wißbrock A, Goradia N, Bellstedt P, Ramachandran R, Imhof D, et al. Heme interaction of the intrinsically disordered N-terminal peptide segment of human cystathionine-β-synthase. Sci Rep. 2018;8:1–9.

32. Faller M, Matsunaga M, Yin S, Loo JA, Guo F. Heme is involved in microRNA processing. Nat Struct Mol Biol. 2007;14:23–9.

33. Barr I, Smith AT, Senturia R, Chen Y, Scheidemantle BD, Burstyn JN, et al. DiGeorge Critical Region 8 (DGCR8) is a double-cysteine-ligated heme protein. J Biol Chem. 2011;286:16716–25.

34. Weitz SH, Gong M, Barr I, Weiss S, Guo F. Processing of microRNA primary transcripts requires heme in mammalian cells. Proc Natl Acad Sci. 2014;111:1861–6.

35. Kühl T, Wißbrock A, Goradia N, Sahoo N, Galler K, Neugebauer U, et al. Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins. ACS Chem Biol. 2013;8:1785–93.

36. Mense SM, Zhang L. Heme: a versatile signaling molecule controlling the activities of diverse regulators ranging from transcription factors to MAP kinases. Cell Res. 2006;16:681–92.

37. Miksanova M, Igarashi J, Minami M, Sagami I, Yamauchi S, Kurokawa H, et al. Characterization of heme-regulated eIF2α kinase: Roles of the N-terminal domain in the oligomeric state, heme binding, catalysis, and inhibition. Biochemistry. 2006;45:9894–905.

38. Igarashi K, Murase M, Iizuka A, Pichierri F, Martinkova M, Shimizu T, et al. Elucidation of the heme binding site of heme-regulated eukaryotic initiation factor 2α kinase and the role of the regulatory motif in heme sensing by spectroscopic and catalytic studies of mutant proteins. J Biol Chem. 2008;283:18782–91.

39. Green D, Furby FH, Berndt MC. The interaction of the VIII/von Willebrand factor complex with hematin. Thromb Haemost. 1986;56:277–82.

40. Repessé Y, Dimitrov JD, Peyron I, Moshai EF, Kiger L, Dasgupta S, et al. Heme binds to factor VIII and inhibits its interaction with activated factor IX. J Thromb Haemost. 2012;10:1062–71.

41. Orino K. Functional binding analysis of human fibrinogen as an iron- and heme-binding protein. BioMetals. 2013;26:789–94.

42. Ke Z, Huang Q. Haem-assisted dityrosine-cross-linking of fibrinogen under non-thermal plasma exposure: One important mechanism of facilitated blood coagulation. Sci Rep. 2016;6 May:1–8. doi:10.1038/srep26982.

43. Grdisa M, White MK. Expression of glyceraldehyde-3-phosphate dehydrogenase during differentiation of HD3 cells. Eur J Cell Biol. 1996;71:177–82.

44. Campanale N, Nickel C, Daubenberger CA, Wehlan DA, Gorman JJ, Klonis N, et al. Identification and characterization of heme-interacting proteins in the malaria parasite, Plasmodium falciparum. J Biol Chem. 2003;278:27354–61.

45. Famin O, Ginsburg H. The treatment of Plasmodium falciparum -infected erythrocytes with chloroquine leads to accumulation of ferriprotoporphyrin IX bound to particular parasite proteins and to the inhibition of the parasite's 6-phosphogluconate dehydrogenase . Parasite. 2003;10:39–50.

46. Chakravarti R, Aulak KS, Fox PL, Stuehr DJ. GAPDH regulates cellular heme insertion into inducible nitric oxide synthase. Proc Natl Acad Sci. 2010;107:18004–9.

47. Hannibal L, Collins D, Brassard J, Chakravarti R, Vempati R, Dorlet P, et al. Heme binding properties of glyceraldehyde-3-phosphate dehydrogenase. Biochemistry. 2012;51:8514–29.

48. McCoubrey WK, Huang TJ, Maines MD. Heme oxygenase-2 is a hemoprotein and binds heme through heme regulatory motifs that are not involved in heme catalysis. J Biol Chem. 1997;272:12568–74.

49. Yi L, Ragsdale SW. Evidence that the heme regulatory motifs in heme oxygenase-2 serve as a thiol/disulfide redox switch regulating heme binding. J Biol Chem. 2007;282:21056–67.

50. Yi L, Jenkins PM, Leichert LI, Jakob U, Martens JR, Ragsdale SW. Heme regulatory motifs in heme oxygenase-2 form a thiol/disulfide redox switch that responds to the cellular redox state. J Biol Chem. 2009;284:20556–61.

51. Fleischhacker AS, Sharma A, Choi M, Spencer AM, Bagai I, Hoffman BM, et al. The C-terminal heme regulatory motifs of heme oxygenase-2 are redox-regulated heme binding sites. Biochemistry. 2015;54:2709–18.

52. Yang J, Kim KD, Lucas A, Drahos KE, Santos CS, Mury SP, et al. A novel heme-regulatory motif mediates heme-dependent degradation of the circadian factor Period 2. Mol Cell Biol. 2008;28:4697–711.

53. Dimitrov JD, Roumenina LT, Doltchinkova VR, Mihaylova NM, Lacroix-Desmazes S, Kaveri S V., et al. Antibodies use heme as a cofactor to extend their pathogen elimination activity and to acquire new effector functions. J Biol Chem. 2007;282:26696–706.

54. Wißbrock A, Goradia NB, Kumar A, Paul George AA, Kühl T, Bellstedt P, et al. Structural insights into heme binding to IL-36α proinflammatory cytokine. Sci Rep. 2019;9:16893.

55. Ogura M, Endo R, Ishikawa H, Takeda Y, Uchida T, Iwai K, et al. Redox-dependent axial ligand replacement and its functional significance in heme-bound iron regulatory proteins. J Inorg Biochem. 2018;182:238–48. doi:10.1016/j.jinorgbio.2018.01.007.

56. Nishitani Y, Okutani H, Takeda Y, Uchida T, Iwai K, Ishimori K. Specific heme binding to heme regulatory motifs in iron regulatory proteins and its functional significance. J Inorg Biochem. 2019;198 June:110726. doi:10.1016/j.jinorgbio.2019.110726.

57. Yamanaka K, Ishikawa H, Megumi Y, Tokunaga F, Kanie M, Rouault TA, et al. Identification of the ubiquitin-protein ligase that recognizes oxidized IRP2. Nat Cell Biol. 2003;5:336–40.

58. Ishikawa H, Kato M, Hori H, Ishimori K, Kirisako T, Tokunaga F, et al. Involvement of heme regulatory motif in heme-mediated ubiquitination and degradation of IRP2. Mol Cell. 2005;19:171–81.

59. Yao X, Balamurugan P, Arvey A, Leslie C, Zhang L. Heme controls the regulation of protein tyrosine kinases Jak2 and Src. Biochem Biophys Res Commun. 2010.

60. Burton MJ, Kapetanaki SM, Chernova T, Jamieson AG, Dorlet P, Santolini J, et al. A heme-binding domain controls regulation of ATP-dependent potassium channels. Proc Natl Acad Sci. 2016;113:3785–90.

61. Kimura I, Nakayama Y, Yamauchi H, Konishi M, Miyake A, Mori M, et al. Neurotrophic activity of neudesin, a novel extracellular heme-binding protein, is dependent on the binding of heme to its cytochrome b 5-like heme/steroid-binding domain. J Biol Chem. 2008;283:4323–31.

62. Kimura I, Nakayama Y, Konishi M, Kobayashi T, Mori M, Ito M, et al. Neuferricin, a novel extracellular heme-binding protein, promotes neurogenesis. J Neurochem. 2010;112:1156–67.

63. Chernova T, Steinert JR, Guerin CJ, Nicotera P, Forsythe ID, Smith AG. Neurite Degeneration Induced by

Heme Deficiency Mediated via Inhibition of NMDA Receptor-Dependent Extracellular Signal-Regulated Kinase 1/2 Activation. J Neurosci. 2007;27:8475–85.

64. Kannan M, Steinert JR, Forsythe ID, Smith AG, Chernova T. Mevastatin accelerates loss of synaptic proteins and neurite degeneration in aging cortical neurons in a heme-independent manner. Neurobiol Aging. 2010;31:1543–53. doi:10.1016/j.neurobiolaging.2008.09.004.

65. Shen J, Sheng X, Chang ZN, Wu Q, Wang S, Xuan Z, et al. Iron metabolism regulates p53 signaling through direct Heme-p53 interaction and modulation of p53 localization, stability, and function. Cell Rep. 2014;7:180–93. doi:10.1016/j.celrep.2014.02.042.

66. Shen J, Sheng X, Chang ZN, Wu Q, Xie D, Wang F, et al. The heme–p53 interaction: Linking iron metabolism to p53 signaling and tumorigenesis. Mol Cell Oncol. 2016;3:5–7.

67. Min L, Strushkevich N V., Harnastai IN, Iwamoto H, Gilep AA, Takemori H, et al. Molecular identification of adrenal inner zone antigen as a heme-binding protein. FEBS J. 2005;272:5832–43.

68. Kabe Y, Nakane T, Koike I, Yamamoto T, Sugiura Y, Harada E, et al. Haem-dependent dimerization of PGRMC1/Sigma-2 receptor facilitates cancer proliferation and chemoresistance. Nat Commun. 2016;7 May 2015:1–13. doi:10.1038/ncomms11030.

69. Yin L, Wu N, Curtin JC, Qatanani M, Szwergold NR, Reid RA, et al. Rev-erbα, a heme sensor that coordinates metabolic and circadian pathways. Science (80-). 2007;318:1786–9.

70. Raghuram S, Stayrook KR, Huang P, Rogers PM, Nosie AK, McClure DB, et al. Identification of heme as the ligand for the orphan nuclear receptors REV-ERBα and REV-ERBβ. Nat Struct Mol Biol. 2007;14:1207–13.

71. Marvin KA, Reinking JL, Lee AJ, Pardee K, Krause HM, Burstyn JN. Nuclear receptors Homo sapiens rev-erbβ and Drosophila melanogaster E75 are thiolate-ligated heme proteins which undergo redox-mediated ligand switching and bind CO and NO. Biochemistry. 2009;48:7056–71.

72. Pardee KI, Xu X, Reinking J, Schuetz A, Dong A, Liu S, et al. The structural basis of gas-responsive transcription by the human nuclear hormone receptor REV-ERBβ. PLoS Biol. 2009;7:0384–98.

73. Gupta N, Ragsdale SW. Thiol-disulfide redox dependence of heme binding and heme ligand switching in nuclear hormone receptor rev-erbβ. J Biol Chem. 2011;286:4392–403.

74. Carter EL, Gupta N, Ragsdale SW. High affinity heme binding to a heme regulatory motif on the nuclear receptor rev-erbβ leads to its degradation and indirectly regulates its interaction with nuclear receptor corepressor. J Biol Chem. 2016;291:2196–222.

75. Carter EL, Ramirez Y, Ragsdale SW. The heme-regulatory motif of nuclear receptor Rev-erbβ is a key mediator of heme and redox signaling in circadian rhythm maintenance and metabolism. J Biol Chem. 2017;292:11280–99.

76. Westberg JA, Jiang J, Andersson LC. Stanniocalcin 1 binds hemin through a partially conserved heme regulatory motif. Biochem Biophys Res Commun. 2011;409:266–9. doi:10.1016/j.bbrc.2011.05.002.

77. Jiang J, Westberg JA, Andersson LC. Stanniocalcin 2, forms a complex with heme oxygenase 1, binds hemin and is a heat shock protein. Biochem Biophys Res Commun. 2012;421:274–9. doi:10.1016/j.bbrc.2012.03.151.

78. Figueiredo RT, Fernandez PL, Mourao-Sa DS, Porto BN, Dutra FF, Alves LS, et al. Characterization of heme as activator of toll-like receptor 4. J Biol Chem. 2007;282:20221–9.

79. Wakasugi K. Human tryptophanyl-tRNA synthetase binds with heme to enhance its aminoacylation activity. Biochemistry. 2007;46:11291–8.

80. Nikolajski M, Kühl T, Heinemann SH, Schlott B, Imhof D, Sahoo N. Determination of Hemin-Binding Characteristics of Proteins by a Combinatorial Peptide Library Approach. ChemBioChem. 2011;12:2846–55.

81. Brewitz HH, Kühl T, Goradia N, Galler K, Popp J, Neugebauer U, et al. Role of the Chemical Environment beyond the Coordination Site: Structural Insight into FeIII Protoporphyrin Binding to Cysteine-Based Heme-Regulatory Protein Motifs. ChemBioChem. 2015.

82. Brewitz HH, Goradia N, Schubert E, Galler K, Kühl T, Syllwasschy BF, et al. Heme interacts with histidine-and tyrosine-based protein motifs and inhibits enzymatic activity of chloramphenicol acetyltransferase from Escherichia coli. Biochim Biophys Acta (BBA)-General Subj. 2016;1860:1343–53.

83. Schubert E, Florin N, Duthie F, Brewitz HH, Kühl T, Imhof D, et al. Spectroscopic studies on peptides and proteins with cysteine-containing heme regulatory motifs (HRM). J Inorg Biochem. 2015;148:49–56.

4.2.3 Summary

A web application for the prediction of transient heme-binding sites from protein sequence, available to international scientists, was successfully developed under the name HeMoQuest[24]. The first step of the HeMoQuest web application is an implementation of the SeqD-HBM algorithm, which identifies a list of suitable HBMs. This list is then optionally filtered against the predictor weighted ensemble solvent accessibility (WESA) for solvent accessibility. Subsequently, three machine learning methods are employed to predict K_D values for the HBM list. These machine learning methods were trained on a large set of peptides with known K_D values. The parameters chosen for training were hydrophobicity, electrostatic index, partial charge, instability index, and helix propensity[24].

HeMoQuest represents the logical extension of the knowledge gathered with heme-binding peptides and thus the first algorithm to predict regulatory heme binding. The prediction accuracy, predicting 100% of known heme-regulatory motifs in human proteins, and the dissociation constants of published heme-binding peptides with 71% accuracy is unparalleled so far and will likely lead to the discovery of numerous yet unknown heme-binding proteins in the future. Interestingly, for many of the known transient heme-binding proteins, the exact interaction site is not known. HeMoQuest allows for the prediction of possible motifs, reducing the amount of model peptides that have to be synthesized, and putative interaction sites that have to be mutated on the protein level, thus reducing the overall experimental effort. In the future, HeMoQuest is prepared to further improve its prediction accuracy with each subsequent set of heme-binding peptides that are published with K_D values, since it can easily be retrained with an expanded set of peptides. HeMoQuest is available as a webserver and thus provides its service to a broad international audience and thus conveys the spirit of the information era, in which *in silico* methods facilitate research processes.

This chapter is fundamentally based on basic studies of heme-binding peptides as presented in Chapter I. All peptide binding data from Chapter I can be integrated to further improve the K_D prediction accuracy. Furthermore, the algorithm is the foundation for Chapter III and Chapter IV, where it is used to predict heme-binding sites in JAK2 and TLR4.

4.3 Chapter III - Structural insights into the interaction of heme with protein tyrosine kinase JAK2

Authors*

Benjamin Franz Schmalohr, Al-Hassan Mohamed Mustafa, Oliver Holger Krämer, and Diana Imhof

4.3.1 Introduction

Heme as regulator of protein function is especially relevant in erythrocytes and their precursors during the massive effort of hemoglobin synthesis. It is well-established that heme acts on the synthesis of globin chains and its own biosynthesis (see Section 2.4). However, the tyrosine kinase JAK2, which is the most important signal transducing kinase in erythrocyte precursors, has also been implicated in heme binding. Intriguingly, the heme-binding sites have not yet been elucidated and the effect has not been shown in a directly relevant cell model.

The following manuscript utilizes knowledge gained from previous studies, Chapter I, and Chapter II to investigate heme binding to JAK2. The effect of heme on JAK2 is investigated in the K562 cell line, which is a model of undifferentiated erythrocytes. Furthermore, putative heme-binding sites in JAK2 are predicted using both consensus sequences and HeMoQuest. Two promising HBMs are investigated on the peptide level and possible mechanistic effects are discussed.

*Contributions:
BFS and DI selected peptides. BFS synthesized and characterized the peptides. BFS and AMM planned the *in vitro* assay. AMM performed the *in vitro* assay and immunoblotting. BFS performed *in silico* analysis of heme-binding motifs. All authors discussed and interpreted the data. BFS wrote the manuscript draft. The manuscript was finalized through contributions of all authors.

4.3.2 Article

On the following pages, the article is printed in its published form.

ChemBioChem

Communications
doi.org/10.1002/cbic.202000730

Chem
Europ

European C
Societies P

Check for
updates

Structural Insights into the Interaction of Heme with Protein Tyrosine Kinase JAK2**

Benjamin Franz Schmalohr,[a] Al-Hassan M. Mustafa,[b] Oliver H. Krämer,[b] and Diana Imhof*[a]

Janus kinase 2 (JAK2) is the most important signal-transducing tyrosine kinase in erythropoietic precursor cells. Its malfunction drives several myeloproliferative disorders. Heme is a small metal-ion-carrying molecule that is incorporated into hemoglobin in erythroid precursor cells to transport oxygen. In addition, heme is a signaling molecule and regulator of various biochemical processes. Here, we show that heme exposure leads to hyperphosphorylation of JAK2 in a myeloid cancer cell line. Two peptides identified in JAK2 are heme-regulatory motifs and show low-micromolar affinities for heme. These peptides map to the kinase domain of JAK2, which is essential for downstream signaling. We suggest these motifs to be the interaction sites of heme with JAK2, which drive the heme-induced hyperphosphorylation. The results presented herein could facilitate the development of heme-related pharmacological tools to combat myeloproliferative disorders.

Janus kinase 2 (JAK2), a protein tyrosine kinase, is crucial for the transduction of cytokine and hormone signals in humans and other vertebrates.[1] Defective JAK2 signaling is involved in myeloproliferative disorders such as polycythemia vera, leukemia, and lymphoma.[2] The Janus homology 2 (JH2) pseudokinase domain inhibits the catalytic activity of JAK2.[3] Its action is further regulated by cellular inhibitors, such as lymphocyte adapter protein (LNK), casitas B-lineage lymphoma protein (CBL), suppressor of cytokine signaling 3 (SOCS3), and protein tyrosine phosphatases (PTPs).[4–6] Pharmacological inhibitors against aberrant JAK2 activity have recently been developed.[7]

In 2010, JAK2 was found to be regulated by heme (iron(II/III) protoporphyrin IX).[8] Heme is a known regulator of protein function and stability, controlling hemoglobin synthesis, its own biosynthesis, as well as inflammatory processes.[9–11] Regulatory heme binding occurs primarily through heme-regulatory motifs (HRMs)[9], short sequence stretches on the protein surface, which contain a central, iron-coordinating amino acid.[10,12] The best-known HRM is the Cys-Pro dipeptide (CP) motif.[13]

However, we have recently identified potential new HRMs based on histidine and tyrosine, for example, HXH, HXXXY, and HXXXH.[14,15] The concept of heme binding to JAK2 is conceivable as the cells in which it is expressed, such as erythroid precursor cells, exhibit high heme concentrations.[16,17] JAK2 activity is activated by erythropoietin, which drives hemoglobin synthesis.[18] At the same time, heme is accumulated in erythroid precursors and then increases globin production by inactivating the transcription inhibitor Bach1 (BTB domain and CNC homologue 1).[19] The aforementioned JAK2-heme interaction was initially identified by a database search for CP-containing peptides,[8] but no proof of heme binding to a CP motif in JAK2 exists as of yet. We used K562 cells to investigate the effect of heme on JAK2 and hypothesize that a histidine/tyrosine (H/Y)-based and a CP motif of the JH1 domain are involved in heme-induced JAK2 phosphorylation.

To investigate the effect of heme on JAK2 we utilized the erythroleukemia cell line K562. Under normal conditions, these cells have low intrinsic JAK2 signaling and therefore low amounts of Y^{1007}/Y^{1008} phosphorylated JAK2. However, heme is ubiquitously present, so that cells and the cell culture additive fetal bovine serum (FBS) contain a pool of labile heme.[20] This naturally occurring heme can mask heme-induced effects and therefore needs to be reduced to minimal amounts.[8] We blocked intrinsic heme synthesis with the inhibitor succinyl acetone (SA) and extrinsic heme by using heme-depleted FBS.[20] Western blot analysis of cell lysates from K562 cells treated in this way showed minimal JAK2 phosphorylation (Figure 1). Upon addition of 2 μM and 4 μM heme, respectively, JAK2 was strongly phosphorylated with a relative density increase from 1 to 1.7 and 1.8 (Figure 1). This effect did not increase further upon addition of 10 μM heme (data not shown). The activating phosphorylation of JAK2 was confirmed by STAT3, which was also increasingly phosphorylated upon heme addition (Figure 1). NaOH (negative control), which was used to dissolve heme, did not show any effect. We thus demonstrate that heme has a marked activating effect on JAK2 and STAT3 in a human erythroleukemia cell line.

We subsequently examined whether the heme effect was due to a direct heme-JAK2 interaction and if so, where the interaction site(s) could be located. Recent investigations revealed that His/Tyr-based motifs, such as (H/Y)X(H/Y), are interesting candidates for HRMs beside the CP motif.[14] In a consensus sequence-based search,[12] we identified the motif KRYIH^{974}RDLA (peptide 1) as potential HRM in the JH1 domain of JAK2.[14] This motif fulfills all criteria for regulatory heme binding.[12,15] It has a positive net charge, stemming from basic amino acids, and contains advantageous hydrophobic amino

[a] B. F. Schmalohr, Prof. D. Imhof
     Pharmaceutical Biochemistry and Bioanalytics
     Pharmaceutical Institute, University of Bonn
     An der Immenburg 4, 53121 Bonn (Germany)
     E-mail: dimhof@uni-bonn.de
[b] A.-H. M. Mustafa, Prof. O. H. Krämer
     University Medical Center Mainz, Institute of Toxicology
     Obere Zahlbacher Straße 67, 55131 Mainz (Germany)
[**] A previous version of this manuscript has been deposited on a preprint
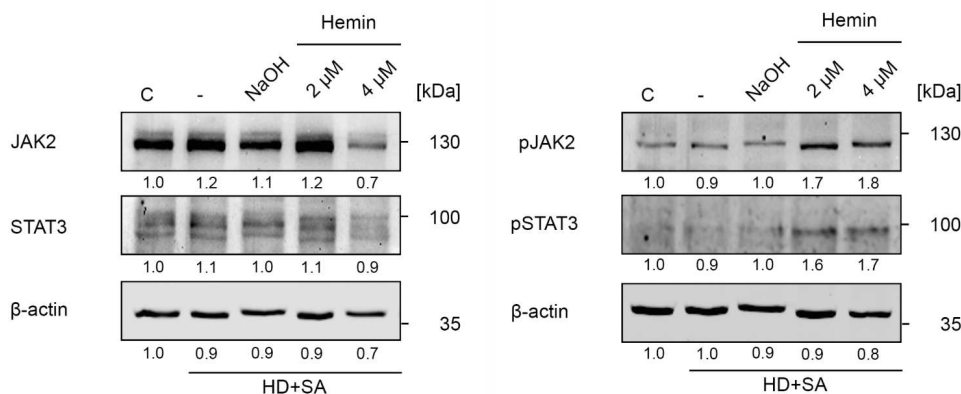   server (https://doi.org/10.1101/2020.08.13.246454)

65

ChemBioChem

Communications
doi.org/10.1002/cbic.202000730

Chemistry
Europe
European Chemical
Societies Publishing

Figure 1. JAK2 is hyperphosphorylated upon heme exposure. Detection of the non-phosphorylated proteins (left) and phosphorylated JAK2/STAT3 (right) was performed on lysates of K562 cells, depleted of heme by addition of succinyl acetone (SA) and heme-depleted medium (HD). Densiometric analysis is indicated below each line. Western blot analysis showed a clear increase in JAK2 phosphorylation at Y^{1007}/Y^{1008} and STAT3 phosphorylation at Y^{705}. NaOH did not increase JAK2 or STAT3 phosphorylation. The results are representative of three independent experiments.

acids. The central histidine, here H974, has been shown to effectively coordinate the heme iron ion, especially in the context of an H/Y motif such as YXH.[14] Previously, it was hypothesized that a CP motif might be responsible for heme binding to JAK2,[14] but no proof of this has been brought forward yet.[8] There is only a single CP motif in JAK2, namely RPDG**C**[1092]**P**DEI (peptide 2), which is also located within the JH1 domain (Figure 2).

Peptide 1 and the CP-containing peptide 2 were produced by solid-phase peptide synthesis and analyzed for heme binding using UV/Vis spectroscopy, as previously described.[14,21] Both peptides were able to bind heme with good to high affinity, as apparent from the dissociation constants (K_D) 2.49 ± 0.13 μM (peptide 1) and 0.50 ± 0.23 μM (peptide 2).[14,21] The peptides form a pentacoordinate 1:1 complex with heme as identified from a shift of the ν3-band to 1495 (peptide 1) and

1491 cm^{-1} (peptide 2) in resonance Raman spectroscopy and from UV/Vis spectroscopy.[14,21] In consequence, both peptides represent suitable candidates for the heme interaction sites in JAK2.

In order to visualize the heme-binding motifs, we used a JH1-JH2 domain model, which was generated from high-resolution crystal structures of JH1 and JH2 using extensive all-atom molecular dynamics (MD) simulations.[22] The sequence corresponding to peptide 1 spans over the majority of the catalytic loop of the JH1 domain, which is located in the domains core (Figure 2).[23] Parts of the motif are solvent-accessible, but heme binding might interfere with substrate binding at this site. Peptide 2, on the other hand, is located in a flexible loop on the surface of the JH1 domain, which has been hypothesized to be part of the JH1-Src-homology 2-like domain (SH2L) interface.[24] Its surface accessibility would be beneficial for unhindered heme binding, as has been shown for other heme-regulated proteins.[25] It is also conceivable that further motifs are involved in heme binding to JAK2. The algorithm HeMoQuest[26] suggested 14 further motifs. However, the majority of these were predicted to possess a lower binding affinity compared to peptides 1 and 2 or are not favorably positioned on the surface of the protein.

Here we suggest regulation of JAK2 by heme in K562 cells, which is mediated by two potential heme-regulatory motifs. We confirmed these motifs on the peptide level and mapped their localization within the JAK2 structure to the catalytic loop and the putative JH1/SH2 L domain interface. Our results are in good agreement with previous results by Zhang et al., which showed that JAK2 is heme-regulated in HeLa cells.[8] In contrast to this report, however, we had a closer look at the possible heme interaction sites. Peptide 1 might not be favored for heme binding, since heme binding to the catalytic loop could interfere with phosphorylation, for example by blocking access to the catalytic center. In contrast, heme binding to peptide 2 might disrupt the JH1-SH2 L interaction and thus release the JH1 domain into the putative elongated active state, which
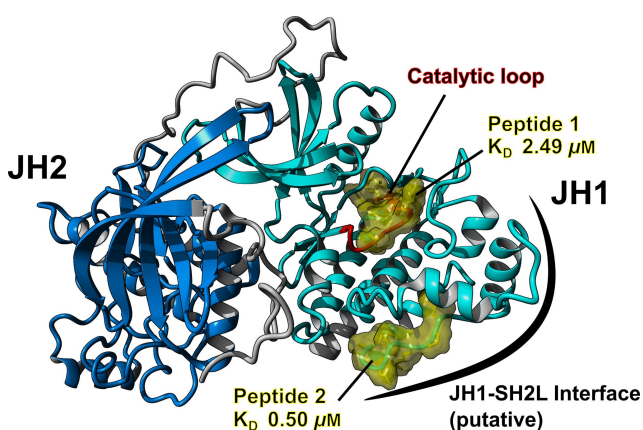

Figure 2. Visualization of the suggested motifs in the JH1 domain of JAK2. The motifs KR**YIH**[974]RDLA (peptide 1) and RPDG**C**[1092]**P**DEI (peptide 2) are shown in yellow in the context of JH1 (light blue ribbons) and JH2 (dark blue ribbons). Peptide 1 is located in the catalytic loop (red ribbons), whereas peptide 2 is located in the putative JH1-SH2 L interface.[24] Visualization was performed with YASARA[27] and a model of JH1-JH2.[22]

would explain the observed hyperphosphorylation.[24] Alternatively, one could speculate that the increased heme phosphorylation could stem from an indirect effect. Heme could either directly bind to proteins/peptides that are upstream of JAK2 (e.g., type I or type II receptors or their ligands),[28] or it could induce other pathways, which might eventually lead to JAK2 activation (e.g., the TLR4-NFκB axis by secretion of cytokines).[29] Future studies will shed light on the underlying mechanism, yet are dependent on the sufficient availability, quality, and integrity of the 120 kDa protein JAK2.

In summary, we confirm heme binding to JAK2 and pinpoint two potential binding sites, as supported by peptides representing the respective motifs. This study provides deeper insight into the regulatory effect of heme and might aid in unraveling the role of heme in JAK2-related diseases. Especially in myeloid leukemia, the heme-degrading enzyme heme oxygenase 1 has recently been identified as druggable target.[30] Consequently, localizing heme-binding sites may aid in the development of novel research tools and subsequently new targeted drugs.

Experimental Section

Reagents: Endotoxin-free heme (Fe^{III} protoporphyrin IX chloride) was purchased from Frontier Scientific, Logan, UT. Peptides were synthesized by Fmoc solid-phase peptide synthesis and characterized as described earlier.[14,23] Succinylacetone (4,6-dioxoheptanoic acid) was purchased from Sigma-Aldrich. FBS was obtained from Thermo Fisher Scientific. Heme-depleted medium was prepared from heme-depleted FBS, as described earlier.[20] K562 cells were cultured at 37 °C and a 5 % CO$_2$ humidified conditions in RPMI-1640 medium supplemented with 5–10 % heme-depleted FBS, and 1 % penicillin/streptomycin (Sigma–Aldrich). For immunoblot analysis, anti-JAK2 (#3230), anti-rabbit IgG HRP-linked (#7074), anti-mouse IgG HRP-linked (#7076) were purchased from Cell Signaling Technology, Frankfurt, Germany; anti-pJAK2 (Tyr1007/1008; sc-21870), anti-pSTAT3 (Tyr705; sc-7993), and anti-β-actin (sc-47778) were purchased from Santa Cruz, Heidelberg, Germany.

Heme addition to K562 cells and measurement of JAK2 phosphorylation: K562 cells were depleted of endogenous heme by incubation with HD medium and 0.5 mM SA for 24 h prior to heme addition. Heme was dissolved in 30 mM NaOH to a concentration of 500 μM for 30 min in the dark and sterilized using a 0.2 μm filter. After filtration, the concentration was normalized using an extinction coefficient of 32.482 mM^{-1} cm^{-1} at 398 nm[14] and further diluted to in DPBS to 10x final assay concentration. Cells were seeded in serum-free medium with 0.5 mM SA in 6-well plates and the freshly prepared heme solutions were added. After incubation for 24 h, the cells were harvested by centrifugation (300 g, 5 min). Cells were lysed in NET–N buffer (100 mM NaCl, 10 mM Tris·HCl pH 8, 1 mM EDTA, 10 % glycerine, 0.5 % NP-40; plus cOmplete^{TM} Protease Inhibitor Cocktail tablets (Roche) and phosphatase inhibitor cocktail 2 (Sigma)) for 30 min on ice, sonicated (10 s, 20 % amplitude) and centrifuged (18 800 g, 20 min, 4 °C). Protein content of lysates was estimated in a Bradford assay. Proteins were detected and analyzed by SDS-PAGE and western blotting using enhanced chemiluminescence on an iBright CL1000 imaging system (Invitrogen). Densiometric analysis was performed using Image J.[31]

[1] E. Parganas, D. Wang, D. Stravopodis, D. J. Topham, J. C. Marine, S. Teglund, E. F. Vanin, S. Bodner, O. R. Colamonici, J. M. Van Deursen, G. Grosveld, J. N. Ihle, *Cell* **1998**, *93*, 385–395.

[2] S. R. Hubbard, *Front. Endocrinol.* **2018**, *8*, 361.

[3] P. Saharinen, K. Takaluoma, O. Silvennoinen, *Mol. Cell. Biol.* **2000**, *20*, 3387–3395.

[4] A. Bersenev, C. Wu, J. Balcerek, W. Tong, *J. Clin. Invest.* **2008**, *118*, 2832–2844.

[5] K. Lv, J. Jiang, R. Donaghy, C. R. Riling, Y. Cheng, V. Chandra, K. Rozenova, W. An, B. C. Mohapatra, B. T. Goetz, V. Pillai, X. Han, E. A. Todd, G. R. Jeschke, W. Y. Langdon, S. Kumar, E. O. Hexner, H. Band, W. Tong, *Genes Dev.* **2017**, *31*, 1007–1023.

[6] R. Shuai, B. Liu, *Nat. Rev. Immunol.* **2003**, *3*, 900–911.

[7] P. Bose, S. Verstovsek, *Blood* **2017**, *130*, 115–125.

[8] X. Yao, P. Balamurugan, A. Arvey, C. Leslie, L. Zhang, *Biochem. Biophys. Res. Commun.* **2010**, *403*, 30–35.

[9] L. Zhang, L. Guarente, *EMBO J.* **1995**, *14*, 313–320.

[10] T. Kühl, D. Imhof, *ChemBioChem* **2015**, *16*, 2024–2035.

[11] A. W. Munro, H. M. Girvan, K. J. McLean, M. R. Cheesman, D. Leys, in *Tetrapyrroles Birth, Life Death* (Eds.: M. J. Warren, A. G. Smith), Springer, **2009**, pp. 160–183.

[12] A. Wißbrock, A. A. Paul George, H. H. Brewitz, T. Kühl, D. Imhof, *Biosci. Rep.* **2019**, *39*, BSR20181940.

[13] L. Zhang, *Heme Biology: The Secret Life of Heme in Regulating Diverse Biological Processes*, World Scientific, **2011**.

[14] B. F. Syllwasschy, M. S. Beck, I. Družeta, M.-T. Hopp, A. Ramoji, U. Neugebauer, S. Nozinovic, D. Menche, D. Willbold, O. Ohlenschläger, T. Kühl, D. Imhof, *Biochim. Biophys. Acta Gen. Subj.* **2020**, *1864*, 129603.

[15] H. H. Brewitz, N. Goradia, E. Schubert, K. Galler, T. Kühl, B. F. Syllwasschy, J. Popp, U. Neugebauer, G. Hagelueken, O. Schiemann, O. Ohlenschläger, D. Imhof, *Biochim. Biophys. Acta Gen. Subj.* **2016**, *1860*, 1343–1353.

[16] A. Oda, K. Sawada, B. J. Druker, K. Ozaki, H. Takano, K. Koizumi, Y. Fukada, M. Handa, T. Koike, Y. Ikeda, *Blood*, **1998**, *92*, 443–451.

[17] D. Chiabrando, S. Marro, S. Mercurio, C. Giorgi, S. Petrillo, F. Vinchi, V. Fiorito, S. Fagoonee, A. Camporeale, E. Turco, G. R. Merlo, L. Silengo, F. Altruda, P. Pinton, E. Tolosano, *J. Clin. Invest.* **2012**, *122*, 4569–4579.

[18] T. D. Richmond, M. Chohan, D. L. Barber, *Trends Cell Biol.* **2005**, *15*, 146–155.

[19] K. Ogawa, J. Sun, S. Taketani, O. Nakajima, C. Nishitani, S. Sassa, N. Hayashi, M. Yamamoto, S. Shibahara, H. Fujita, K. Igarashi, *EMBO J.* **2001**, *20*, 2835–2843.

[20] C. Chen, T. K. Samuel, M. Krause, H. A. Dailey, I. Hamza, *J. Biol. Chem.* **2012**, *287*, 9601–9612.

[21] T. Kühl, A. Wißbrock, N. Goradia, N. Sahoo, K. Galler, U. Neugebauer, J. Popp, S. H. Heinemann, O. Ohlenschläger, D. Imhof, *ACS Chem. Biol.* **2013**, *8*, 1785–1793.

[22] Y. Shan, K. Gnanasambandan, D. Ungureanu, E. T. Kim, H. Hammarén, K. Yamashita, O. Silvennoinen, D. E. Shaw, S. R. Hubbard, *Nat. Struct. Mol. Biol.* **2014**, *21*, 579–584.

[23] K. Lindauer, T. Loerting, K. R. Liedl, R. T. Kroemer, *Protein Eng.* **2001**, *14*, 27–37.

ChemBioChem

Communications
doi.org/10.1002/cbic.202000730

Chemistry
Europe
European Chemical
Societies Publishing

[24] P. Ayaz, H. M. Hammarén, J. Raivola, D. Sharon, S. R. Hubbard, O. Silvennoinen, Y. Shan, D. E. Shaw, *bioRxiv* **2019**, doi:10.1101/727727.

[25] A. Wißbrock, N. B. Goradia, A. Kumar, A. A. Paul George, T. Kühl, P. Bellstedt, R. Ramachandra, P. Hoffmann, K. Galler, J. Popp, U. Neugebauer, K. Hampel, B. Zimmermann, S. Adam, M. Wiendl, G. Schett, I. Hamza, S. H. Heinemann, S. Frey, et al., *Sci. Rep.* **2019**, *9*, 16893.

[26] A. A. Paul George, M. Lacerda, B. F. Syllwasschy, M.-T. Hopp, A. Wißbrock, D. Imhof, *BMC Bioinf.* **2020**, *21*, 124.

[27] E. Krieger, G. Vriend, *Bioinformatics* **2014**, *30*, 2981–2982.

[28] M. Sakatsume, K. Igarashi, K. D. Winestock, G. Garotta, A. C. Larner, D. S. Finbloom, *J. Biol. Chem.* **1995**, *270*, 17528–17534.

[29] S. Lin, Q. Yin, Q. Zhong, F.- L. Lv, Y. Zhou, J.- Q. Li, J.- Z. Wang, B. Su, Q. - W. Yang, *J. Neuroinflammation* **2012**, *9*, 1–14.

[30] L. Salerno, G. Romeo, M. N. Modica, E. Amata, V. Sorrenti, I. Barbagallo, V. Pittalà, *Eur. J. Med. Chem.* **2017**, *142*, 163–178.

[31] C. A. Schneider, W. S. Rasband, K. W. Eliceiri, *Nat. Methods* **2012**, *9*, 671–675.

COMMUNICATIONS
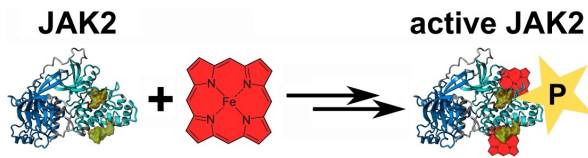
Where it's happening: JAK2 is of paramount importance in erythropoietic precursor cells and its dysfunction drives myeloid cancers. Heme is accumulated by these cells in order to produce hemoglobin. Here we show that heme can drive JAK2 phosphorylation in K562 myeloid tumor cells. We identify two possible interaction sites based on distinct heme-binding motifs, that is, cysteine-proline and histidine-X-tyrosine. Structural insights into the interaction of heme with protein tyrosine kinase JAK2 (Imhof @UniBonn)

*B. F. Schmalohr, A.-H. M. Mustafa,
Prof. O. H. Krämer, Prof. D. Imhof**

1 – 5

Structural Insights into the Interaction of Heme with Protein Tyrosine Kinase JAK2

4.3.3   Summary

In this chapter, it was shown that JAK2 represents a prime example of a heme-regulated protein confirming earlier studies, which suggested the protein to possess heme-regulatory motifs[201]. In K562 cells, which resemble early-stage erythrocytes, heme is able to markedly increase the phosphorylation of JAK2 at concentrations as low as 2 µM, shown by Western blot. This phosphorylation induces downstream signaling, i.e. through STAT3, which was also hyperphosphorylated upon heme exposure. Two potential HRMs were identified and produced as peptides. One peptide contained a CP motif, whereas the other contained a YXH motif belonging to subclass B, which has been suggested as favorable in Chapter I. Both exhibited strong heme binding, with the CP peptide displaying a slightly lower K_D value. The putative HRMs were both located in the highly relevant JH1 domain, which is responsible for the initiation of JAK2 downstream signaling. Interestingly, the CP motif was located on the interface of the JH1 and SH2L domain. It has been suggested that a destabilization of this interface will lead to an increased activity of JAK2.

In summary, this chapter shows the successful application of previous research (see Section 2.3 and Chapter I and Chapter II) to the protein level. The results allow for a deeper understanding of the regulatory role heme plays in the regulation of globin synthesis and the etiology of myeloproliferative disorders.

4.4 Chapter IV – In silico analysis of the interaction of heme with the TLR4/MD2 complex

4.4.1 Introduction

In previous studies, TLR4 emerged at the junction of heme-related pathologies, driving inflammation and thrombosis in conditions of hemolysis. As described in 2.5.3, much is known already about TLR4 in general and in part also on the interaction with heme. However, two aspects are still subject of dispute. Firstly, the exact nature of the interaction of heme with TLR4 remains elusive. Does heme bind to TLR4, to MD2 or only to a complex of both? What are the thermodynamic parameters of the interaction such as the stoichiometry, the association and dissociation constants, as well as the corresponding rate constants? Where are the interaction sites located? The clarification of these issues is paramount for the elucidation of the second aspect, the physiological relevance of the heme-TLR4 interaction. Arising questions in this regard are: Can (patho)physiological heme concentrations be sufficiently high for TLR4 activation to occur? Which downstream effectors of TLR4 are induced by heme? Is there a biased signaling in response to heme? Investigation of all of these aspects and the subsequent improved understanding of heme biology in regard to TLR4 may allow for the application of targeted medicines in hemolytic patients to prevent organ damage as a result of heme toxicity.

The first step on this path is the physicochemical characterization of the heme-TLR4 interaction. A first attempt has been made by the group of Belcher et al. in 2020. It was proposed that heme does not bind to TLR4 but to its essential co-receptor MD2[256]. The authors presented a UV/Vis spectrum and an agarose pull-down experiment. However, the UV/Vis spectrum exhibits a maximum absorbance at the Soret band of 0.025, which is arguably below the threshold of reliable measurement with this method and leads to an unacceptably low signal to noise ratio[50]. In addition, the utilized heme-biotin streptavidin-agarose pull-down assay has not been confirmed in regard to specificity. As further method, an MD2 mutant protein was generated and tested in a luciferase-based nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) assay system[256]. Interestingly, the control experiments using this assay did not reflect previous results by other groups, which reported an over-additive effect of heme and LPS on NF-κB[222,251]. Nonetheless, Belcher et al. conclude that W23, Y34, and Y36 in MD2 are involved in heme binding[256]. In the reported data, the mutations W23A and Y34A reduce heme binding, while Y36A increases heme binding, despite the close proximity to Y34 [256]. As a control for mutation-induced structural changes, the authors check for LPS sensitivity of the mutants, which remains largely unaltered[256]. On a side note, all three suggested amino acids are located in the direct vicinity of a cysteine bridge, which might impair heme binding.

As mentioned above, it would be very interesting to elucidate the heme-binding site of TLR4 or MD2 and to confirm the results presented by Belcher et al. The tools for this task have been presented in Chapter I and Chapter II. Consequently, this chapter will afford a second perspective on possible binding sites of heme on TLR4 and MD2.

4.4.2 Methods

IN SILICO BINDING PREDICTIONS

Heme-binding sites on TLR4 and MD2 were predicted from the canonical human sequences sourced from UniProt (IDs O00206 and Q9Y6Y9)[257]. For this purpose, the

published algorithm HeMoQuest was used[24]. Furthermore, several docking experiments were performed on a fully glycosylated crystal structure of the extracellular part of the TLR4/MD2 complex (PDB: 4G8A) using YASARA version 19.12.14[109]. From the crystal structure several different arrangements for heme interaction were extracted. Firstly, one half of the dimer was removed and docking was performed on a single TLR4/MD2/LPS complex. In a further approach, LPS was also removed and docking was performed only on the TLR4/MD2 complex. As a ligand, Fe(III)protoporphyrin IX was obtained from the ChemSpider database (ID 401223, Chloride removed)[258]. Each docking experiment was performed with the *dock_runensemble* macro in the AutoDock VINA mode with a fully flexible ligand. 10 different receptor side chain rotamers were sampled and for each member of the receptor ensemble, 50 docking runs were performed, resulting in a total of 500 docking runs per structure.

Furthermore, local docking runs were performed on the H68/Y72 area by creating a cubic simulation cell with 15 Å edge length around this motif. Subsequently, docking was performed within this simulation cell analogous to the whole receptor docking.

4.4.3   Results

SUITABLE MOTIFS OCCUR PRIMARILY ON TLR4 AND NOT ON MD2

In order to elucidate the heme-binding site of the TLR4/MD2 complex, HeMoQuest was used to predict putative HRMs. In TLR4, 13 suitable motifs were identified, 10 of which were located in the extracellular domain (see Table **1**). Motifs 1 and 2 are both centered around an interesting HXXXY motif, which was shown to be favorable in Chapter I, and additionally have a positive overall net charge, as well as ample aromatic amino acids. Consequently, these motifs represent the most promising candidates in TLR4. Motifs 3 and 5-7 possess only one coordinating amino acid, which has been shown to lead to low or intermediate affinity. Similarly, motifs 4 and 10, which are cysteine-based might not show high heme-binding affinity, since they do not contain a CP motif. Motifs 8 and 9 are also very interesting, since they contain an HXH motif and a further basic arginine as well as hydrophobic amino acids. Therefore, these motifs represent the second interesting binding site. Motifs 11-13 are most likely irrelevant for heme binding, since they are located intracellularly.

In MD2, only two motifs were identified, one of which was involved in the C25-C51 disulfide bridge, which might hinder heme binding. Interestingly, when compared to the motifs in MD2 suggested by Belcher et al., only the first motif overlapped with W23 (see Table **1**). This motif, however, is also involved in the C25-C51 disulfide bridge and contains one N-glycosylated asparagine. All three motifs suggested by Belcher et al. contained disulfide-bonded cysteines. Furthermore, only the motif around W23, which is almost identical to that predicted by HeMoQuest, possesses a positive net charge. The other motifs around Y34 and Y36 have zero or negative net charge, which has been shown to be detrimental to heme binding on the peptide level. Therefore, from the perspective of peptide models, none of the motifs in MD2 appear promising, which is in contrast to the findings in an earlier report by Belcher et al., which suggested MD2 to be the heme-binding part of the TLR4/MD2 complex[256].

Table 1: Overview of the putative HRMs in TLR4.

Motifs are shown as predicted with HeMoQuest and by Belcher et al. [24,256]. Disulfide bridges are notified by arrows, glycosylation by the respective sugar moieties.

Putative HRMs in TLR4 predicted with HeMoQuest				
No.	Coord. residue	9mer motif	Net charge	Comment
1	H68	NPLRHLGSY	+2	
2	Y72	HLGSYSFFS	+1	
3	H148	FPIGHLKTL	+2	
4	C192	QSIYCTDLR	0(CYS motif)	Cys is reduced
5	H199	LRVLHQMPL	+2	
6	H229	EIRLHKLTL	+2	
7	H431	LDFQHSNLK	+1	
8	H456	LDISHTHTR	+2	
9	H458	ISHTHTRVA	+3	
10	C542	FPYKCLNSL	+1	Cys is reduced
11	Y667	GCIKYGRGE	+1	Intracellular
12	H728	HEGFHKSRK	+4	Intracellular
13	H805	VLGRHIFWR	+3	Intracellular

Putative HRMs in MD2 predicted with HeMoQuest				
No.	Coord. residue	9mer motif	Net charge	Comment
1	Y22	AQKQYWVCN	+1	C25←→C51 N26-GalNAc
2	Y75	LKQLYFNLY	+1	

Putative HRMs in MD2 reported by Belcher et al.[256]				
No.	Coord. residue	9mer motif	Net charge	Comment
1	W23	QKQYWVCNS	+1	C25←→C51 N26-GalNAc
2	Y34	ASISYTYCD	-1	C37←→C148
3	Y36	ISYTYCDKM	0	C37←→C148

DOCKING CONFIRMS H68/Y72 AND THE LPS BINDING SITE AS SUITABLE MOTIFS

As a second route of *in silico* assessment of heme-binding motifs in TLR4 and MD2, docking experiments with heme were performed on the TLR4/MD2 complex both with and without LPS. Therein, the heme molecule was kept flexible, to allow for the sampling of a larger conformational space, while for the protein, 10 side chain rotamers were tested. For the TLR4/MD2/LPS complex, 59 docking clusters were identified with a highest binding energy of 10.9 kcal/mol and an *in silico* dissociation constant of 10.7 nM (see Figure **10**A). The structural ensemble without LPS yielded 51 docking clusters with a highest binding energy of 11.9 kcal/mol and an *in silico* dissociation constant of 1.9 nM (see Figure **10**B). The docking results showed two major clusters of suitable poses. On the one hand, the H68/Y72 motif was favored, and on the other hand the opening of the LPS-binding pocket was highly populated (see Figure **10**A). This result is in agreement with the predictions obtained by HeMoQuest, which found this motif to be the most suitable candidate. However, the second motif at H456/H458 was surprisingly not found in the docking experiments. A reason for this may be the lack of aromatic amino acids displayed by this motif. Accumulation of docking poses on the opening of the LPS-binding pocket might be explained by the hydrophobic properties of the binding pocket and LPS itself. It is possible that heme interacts with LPS directly under physiological conditions, but this hypothesis is in need of experimental validation.
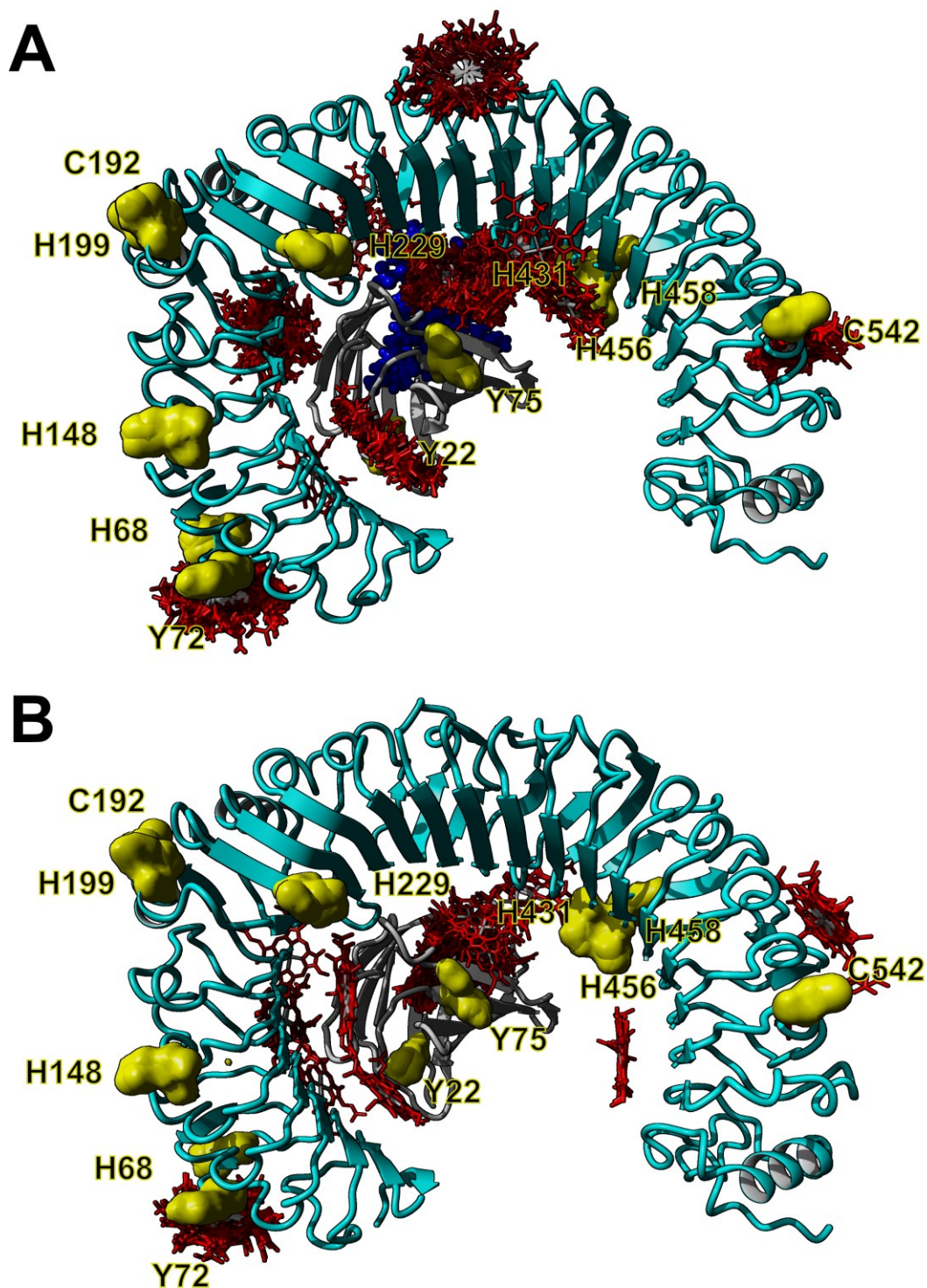
Figure 10: Docking experiments with heme and TLR4.
A: Docking of heme to the TLR4/MD2/LPS complex. The most suitable docking poses are close to H68/Y72 and to the opening of the LPS-binding pocket. **B:** Docking of heme to the TLR4/MD2 complex without LPS. The most suitable docking poses are close to H68/Y72 and within the LPS-binding pocket. TLR4 is shown in light blue ribbons, MD2 as grey ribbons. Heme is depicted as red sticks and LPS as dark blue spheres. The coordinating amino acids predicted by HeMoQuest are shown in yellow. Visualization was performed with YASARA 20.4.24[109].

When heme was docked to TLR4 without LPS, a large number of docking poses was found within the LPS-binding pocket of MD2 (see Figure **10**B). Heme might be able to bind into the LPS-binding pocket of MD2 in place of LPS due to its hydrophobic

nature. However, there have also been reports on a synergistic effect of LPS and heme on the activation of TLR4 and even on intracellular heme levels[251]. This could be realized, for example, by heme stabilizing LPS in its binding pocket via hydrophobic interaction. In order to resolve these questions, these results have to be corroborated by *in vitro* experiments. Surprisingly, none of the motifs suggested by Belcher and colleagues were identified in any of the docking experiments.
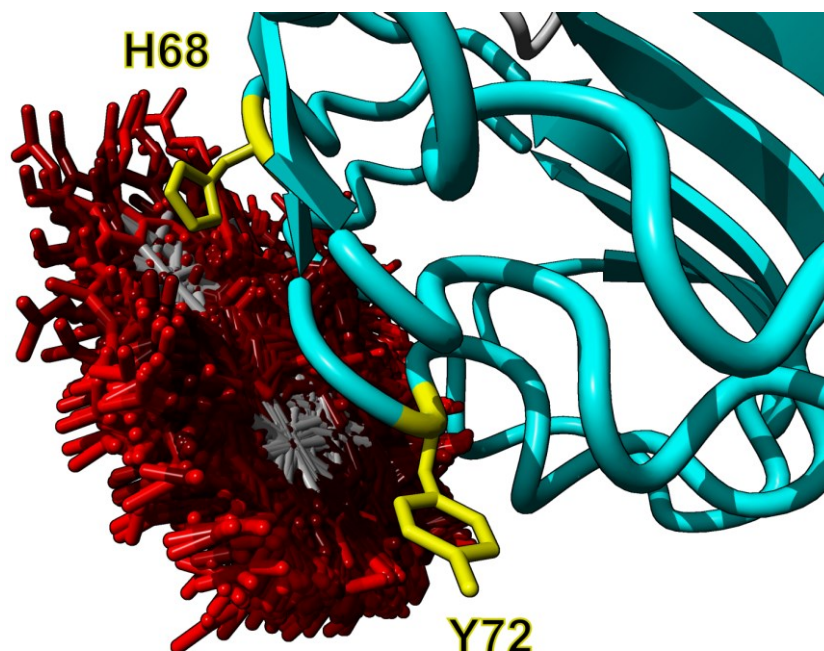
Figure 11: Local docking of heme to H68/Y72 in TLR4.
TLR4 is shown in light blue ribbons, Heme is depicted as red sticks. The coordinating amino acids predicted by HeMoQuest are shown in yellow. Visualization was performed with YASARA 20.4.24[109].

In order to examine the chemical space around the promising H68/Y72 motif in more detail, a focused docking was performed (see Figure **11**). 24 distinct docking clusters were identified with a maximum binding energy of 9.2 kcal/mol and a minimum *in silico* dissociation constant of 168.1 nM. H68 was found to be able to coordinate the central heme iron ion in the docking experiment, whereas Y72 was not found to be the coordinating amino acid in any docking position, but may stabilize the interaction via aromatic interactions with the porphyrin ring system. Overall, docking confirmed the motif NPLRH[68]LGSY[72] as a suitable site for heme interaction, but suggested H68 to be the primary heme-coordinating amino acid. According to the data shown here, this motif represents the most likely candidate for heme binding to TLR4.

NO SUITABLE MOTIFS ARE FOUND ON MD2 AND HEME DOES NOT DOCK TO MD2

Contrary to literature reports of heme binding to MD2[256], no significant docking poses were found on MD2 except for the LPS-binding pocket. Neither the motifs suggested by HeMoQuest, nor those suggested by Belcher and colleagues were identified in the docking clusters (see Figure **12**). This corroborates the assessment performed on the peptide level, which suggested the identified motifs on MD2 as not highly suitable.
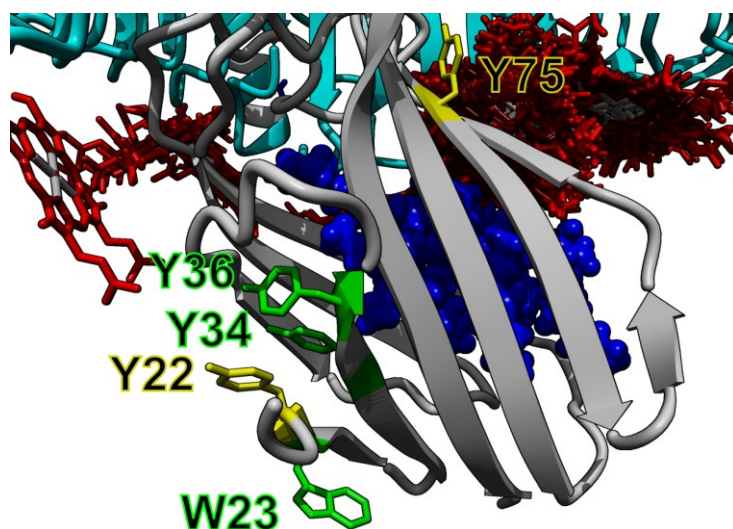
Figure 12: Close-up of MD2 in heme docking experiment to TLR4/MD2/LPS.
No docking pose with heme binding to any of the suggested motifs in MD2 was identified. TLR4 is shown in light blue ribbons, MD2 as grey ribbons. Heme is depicted as red sticks and LPS as dark blue spheres. The coordinating amino acids predicted by HeMoQuest are shown in yellow, while the amino acids suggested by Belcher and colleagues are depicted as green sticks[256]. Visualization was performed with YASARA 20.4.24[109].

4.4.4 Discussion

Using the heme-binding prediction algorithm HeMoQuest, a total of 10 extracellular motifs were suggested. By use of docking experiments, this selection was narrowed down to only one most suitable motif around H68. Interestingly, all other motifs, which were predicted, were not found within the heme docking poses. This might be explained by the fact that during a docking run, all motifs simultaneously compete for one heme molecule, and that thus only the most suitable motifs will be found as a docking result. Since most further motifs did not show optimal sequence requirements, they can be expected to bind heme with lower affinity.

The interaction of heme with MD2 through the suggested peptide motifs could not be confirmed by the *in silico* experiments performed in this chapter. This result was not expected, since several different experimental approaches performed by others led to conclude that heme does bind MD2[256]. Nonetheless, as explained previously (see Section 4.4.1), all of the approaches showed considerable weaknesses and thus the results presented herein might warrant an additional experimental examination of heme binding to MD2. One location that was not previously predicted to be heme-binding was the LPS-binding pocket. Interestingly, in the docking experiments it was identified both in the presence and absence of LPS in the structure. This behavior could explain the reported interaction of heme with MD2 that was seen in UV/Vis, even if the suggested motifs would not be involved[256]. It might also hint at a reason for the observed over-additive effect of simultaneous administration of both heme and LPS on TLR4[251].

It must be considered, however, that the results presented herein rely exclusively on *in silico* work, which has to be validated by experimental work. In the following paragraph, strategies for the evaluation of the arising hypotheses are discussed.

ROUTES FOR EXPERIMENTAL VALIDATION

Since TLR4 is present as a dimer in its active form, the docking studies could be repeated on a dimeric structure. This approach might afford motifs on the interface of two

TLR4 units, which have not been recognized in the monomer. The validity of the suggested motifs *per se* could be evaluated by synthesizing the corresponding nonapeptides and analyzing them for their ability to bind heme. This method has been shown to reliably identify suitable motifs, which could later be confirmed on the protein level[26,30,32,33,132].

A protein level approach could also be envisioned to rely on the direct measurement of the interaction of TLR4/MD2 with heme. Such an approach has been reported for MD2 via UV/Vis interaction measurement, but with relatively noisy results[256]. The difficulty with direct interaction measurement lies in the arduous expression and purification of sufficient quantities of TLR4 and MD2[259]. In any case, UV/Vis measurements, require relatively large quantities of either protein, which makes this method less attractive[50]. Due to the difficulties in the expression of large quantities of human TLR4 and MD2, crystal structure elucidation of the TLR4/MD2/heme complex is not feasible, although it would theoretically give excellent insight. In a crystal structure, the location of the interaction with heme, as well as the involved amino acids could be meticulously analyzed. As a second method, albeit more difficult with heme, 3D-NMR spectroscopy could be performed, giving insight into a solution structure of the complex. Due to limits in protein size, this technique would have to be performed on either MD2 only, or parts of TLR4[260]. In recent years, methods which require significantly less protein have been used with heme. These methods include surface plasmon resonance (SPR) spectroscopy, isothermal titration calorimetry (ITC), or biolayer interferometry (BLI)[50]. All of these methods have the advantage that e.g. K_D value and binding mode can be determined with high precision. This data could, in return, provide a relation of the importance of heme binding to TLR4, MD2 or their complex. However, in order to derive information on possible motifs, mutational studies have to be performed, which again, require large efforts in terms of protein expression and purification.

4.4.5 Summary

This chapter builds on previous work, which raised TLR4 as highly interesting node in a heme biology network[261]. As exploratory *in silico* work, this chapter utilized the knowledge and tools on peptide motifs obtained in Chapter I and Chapter II, to analyze possible heme-binding motifs on TLR4.

Two primary putative interaction sites of heme on TLR4, NPLR**H**[68]LGS**Y**[72] and IS**H**[456]T**H**[458]TRVA, were suggested by HeMoQuest. The first site around H68/Y72 was confirmed in docking experiments, while the second site was not confirmed therein.

No suitable heme interaction sites were identified by HeMoQuest on MD2, which is contrary to the results obtained by Belcher and colleagues[256]. However, heme was found to dock close to the LPS-binding site of MD2, or even into the LPS-binding site if LPS was not present. This binding behavior could explain the synergistic effects of LPS and heme on TLR4 that have been observed by other groups.

5 Conclusion

In order to shed light on H/Y-based motifs, 58 nonapeptides with combinations of exactly two coordination sites (histidine or tyrosine) with different spacer lengths were synthesized in Chapter I. The sequences were classified into a newly devised subclass scheme, consisting of subclasses A-D, corresponding to spacer lengths from 0-3. Extensive heme binding and structural studies were performed on these peptides using UV/Vis, resonance Raman, and NMR spectroscopy. In addition, the peroxidase activity in complex with heme was determined for select peptide candidates.

Extensive studies on peptide models for heme regulatory motifs have been performed in the past. These studies included cysteine-, histidine- and tyrosine-based peptides, and the established CP motif[29–33,156]. However, distinct combinations of histidine and tyrosine as basis for heme-binding motifs were previously not well-characterized. This study represents the first systematic analysis of these alternative H/Y-based motifs and demonstrated that they represent important class of heme-binding peptides.

The peptides displayed K_D values from 0.21 ± 0.17 µM to 8.86 ± 0.67 µM, with subclasses B (one spacer amino acid) and D (three spacer amino acids) exhibiting the largest number of high affinity peptides. These affinities are only slightly lower than what was previously reported for CP peptides[29–31]. Histidine-based peptides displayed stronger heme-binding affinity than tyrosine-based peptides. This is in good agreement with previous studies, which also found that heme-binding affinity diminishes with the coordinating amino acid in the order of C > H > Y[29–33,156]. Subclasses A and C displayed predominantly pentacoordinate heme complexes, while B and D showed mixtures of penta- and hexacoordination. In subclass B, this could be realized by a sandwich-like complex, while for subclass D a clamp-like formation is possible.

In previous studies, NMR structures of two representatives with multiple heme-coordinating sites had already been solved (see Figure **13**)[29,32]. However, no structures with isolated H/Y-based motifs were known. This thesis provided two experimental structures for representatives of subclass B to address this lack of knowledge (see Figure **13**). Indeed, as predicted by earlier studies[31], a clamp-like formation did not form and a stabilizing effect of heme on the peptide structure was observed.

The motifs HXH, HXXXY, and HXXXH emerged as possible heme-binding motifs, characterized by the highest heme-binding affinity and most frequent occurrence. Previously observed conductive characteristics for heme binding, such as further basic amino acids, aromatic, and lipophilic amino acids, were also observed in the investigated peptides. With these results, a comprehensive analysis of HRMs in peptide models has been concluded. When directly compared, the CP motif appears to have slightly higher affinity on the peptide level than the H/Y-based motifs and appears to occur more often in heme-regulated proteins (see Chapter II)[29–33,156]. Nonetheless, in certain examples both on the peptide as well as the protein level, H/Y-based motifs appear to play a considerable role.

With the study of peptide models, some limitations arise. Inherently, models represent simplifications of natural situations which are too complex to be studied directly. The peptides used in this study are merely models for the proteins that contain them. Thus, care has to be taken when extending insights from peptide models to proteins. For instance, the validity of nonapeptides as a model for protein structures can be questioned *per se*. The latter concern can be countered by the fact that nonapeptides have

been used successfully to predict heme binding to a previously unknown heme-binding protein, i.e. with DPP8, HlyC, IL-36, and APC[26,30,33,132].
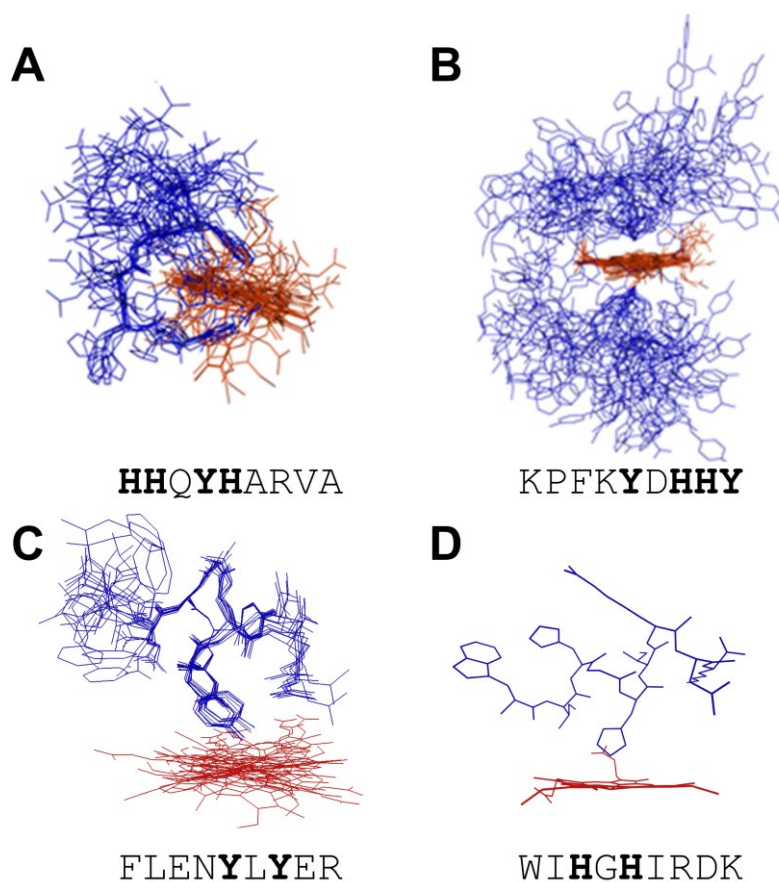

Figure 13: Structures of heme-peptide complexes with H/Y-based motifs.
This study contributed two structures of peptides with H/Y-based motifs, which were solved using 2D-NMR experiments. Prior to this study, only two structures with multiple histidines and tyrosines were known. The peptides are shown as blue sticks, while heme is shown as red sticks. **A:** The peptide HHQYHARVA from class VI shows a clamp-like structure with two histidines coordinating the central iron ion. **B**: The peptide KPFKYDHHY from class VIII shows sandwich-like hexacoordinative binding. **C, D**: This thesis provided two further structures with a YXY and HXH motif from subclass B. Visualization was adapted from [29,32] or performed with VMD[262].

There are multiple possible approaches to investigate models which are potentially closer to entire proteins. Possible extensions of the nonapeptide models studied herein could include longer peptides or combinations of more than two coordinating amino acids. In addition, however, it was shown, that only amino acids in the direct vicinity of the coordinating amino acid (up to four residues distance) have an influence on heme binding[263]. The systematic investigation of models with more than two coordinating amino acids would potentially be possible, but requires considerable effort. For example, three coordinating amino acids would result in at least 16 different classes. With sufficient representatives per class, this would correspond to at least 160 peptides, which would have to be synthesized and analyzed. Furthermore, such motifs were not found in the known heme-regulated proteins so far[24]. Since further studies on peptide motifs will be laborious and might not be promising, this thesis represents a significant milestone for the elucidation of regulatory heme binding. In combination with previously obtained knowledge, it represents the basis for the further investigation of H/Y-based motifs in proteins and the identification of suitable interaction sites.

The second goal of this thesis was the dissemination of the knowledge obtained from peptide models generated in this and previous studies. Thus, in Chapter II of this thesis, all previously published algorithms to predict heme binding to proteins were assessed, and it was found that all of them were focused on permanent heme binding. A specific algorithm was needed to predict transient, regulatory heme binding from protein sequence. This algorithm, HeMoQuest, was built on the collation of all previous data. Specifically, the motif selection scheme provided by SeqD-HBM was integrated with a machine learning algorithm, which is capable of predicting association constants. In order to validate HeMoQuest, a dataset of all known human transiently heme-binding protein was manually assembled and curated meticulously. This dataset in itself provides great value for researchers in the heme field, since an up-to-date collection of this data was not available beforehand. HeMoQuest correctly identified every single literature-known heme-binding site in this dataset, proving its indisputable value. However, in some cases, significantly more heme-binding sites than expected were predicted, which might indicate erroneous prediction of false positives. With the tested protein examples, one cannot be certain that the predicted motifs are not indeed true positives, which have not been identified yet, since several heme molecules can bind to one protein. The number of false positives could potentially be reduced in future iterations, by applying a cutoff on the predicted K_D values, for example at 20 µM. Predicted low-affinity HBMs could then be discarded. Prerequisite for such a cutoff would be high accuracy in the correct prediction of motif affinities. The current version of HeMoQuest has an accuracy of 71% in assigning correct K_D values on the training set, which is likely not sufficient for this procedure. However, it significantly outperformed all previously published algorithms in the correct identification of HBMs.

The machine learning functionality was designed to be trained with peptide sequences and their experimentally determined dissociation constants. This approach has the advantage that it can be extended indefinitely with peptide data that might be generated in the future. With the incorporation of more training data, the accuracy is likely to rise. Furthermore, HeMoQuest is capable of operating on sequence data only and does not require structural information. On the one hand, this drastically enhances its applicability, since for many proteins, no structural data is available. On the other hand, the performed prediction of solvent accessibility by WESA is not flawless and could potentially result in incorrect results. Future improvements of HeMoQuest could therefore encompass other methods of taking the surface exposure of HBMs into account. This is especially interesting, since recent advances in the field of *in silico* structure prediction have been staggering[264].

Chapter III investigated a protein containing both a CP and a H/Y-based motif: JAK2. JAK2 is highly expressed in erythrocyte precursors, which are essentially cellular hemoglobin factories[265]. Hemoglobin synthesis is the junction of several pathways of fine-tuned globin and heme synthesis. JAK2 is responsible for transducing growth signals through the EPO-R and has been reported to be hyperphosphorylated by heme in HeLa cells[201]. In this chapter, the published effect was translated to the highly relevant cell line K562, which represents an erythrocyte-like cancer cell line. In this cell line, the effect of heme on JAK2 could be confirmed and it was furthermore shown that the effect also extends to the downstream transcription factor STAT3. Through the motifs and characteristics analyzed and processed in Chapter I and Chapter II, it was possible to suggest two suitable interaction sites, in the catalytically active JH1 domain. One of them contained a CP motif and the other one a YXH motif. Both motifs exhibit advantageous

sequence characteristics and bind heme with high affinity on the peptide level. In the protein context, the CP motif might be favored, since it is located on the domain surface in a position where heme binding might disrupt inter-domain interactions and thus lead to activation of JAK2. The YXH motif, on the other hand, is part of the catalytic loop and heme binding to it might hinder catalytic activity.

However, the results presented herein do not include measurement of a direct interaction of heme with JAK2 or single domains thereof. It can therefore be questioned, whether JAK2 is directly bound by heme and thus activated, or whether alternative, unknown effects lead to the activation of JAK2 when heme is added. The unambiguous clarification of this concern can only be solved by assessing direct interaction via physicochemical methods, such as UV/Vis, ITC, SPR or BLI. Providing such measurements may be daunting since the heterologous expression of JAK2 is notoriously difficult and crystal structures exist only of single domains[266,267]. Fortunately, the latter three methods require only minuscule amounts of purified protein[84] and might therefore be feasible in the future.

A further issue that awaits answering is the physiological relevance of the heme-JAK2 interaction. Due to the well-established regulatory network of hemoglobin synthesis in erythrocyte precursors and the low concentrations of heme needed to activate JAK2, a physiological relevance is entirely feasible. Nevertheless, *in vivo* experiments are needed to unequivocally demonstrate such a relevance. Fortunately, several mouse models with mutated JAK2 alleles already exist, although complete knock-out of JAK2 is lethal[268–270]. These mouse models could be challenged with heme and possible differences dependent on JAK2 could be observed. In summary, JAK2 represents a highly interesting example of HRMs in a protein with physiological relevance in heme and hemoglobin biology.

As a consequence of the pivotal position of TLR4 in previous studies and a text mining-based network analysis, termed heme knowledge graph (HemeKG)[261], a more detailed *in silico* analysis of TLR4 was performed in Chapter IV. Therein, HeMoQuest was utilized to predict HBMs in TLR4 and MD2, while the results were compared with the putative heme-binding sites reported by Belcher and colleagues[256]. Two suitable motifs for heme binding were identified in TLR4, while no suitable motifs were found in MD2. Docking experiments corroborated a heme-binding site at H68, while the reported sites on MD2 were not confirmed. However, the approach pursued here is exclusively *in silico* and has to be confirmed by experimental studies. Most importantly, a direct interaction between heme and TLR4 or MD2 would have to be confirmed, but further possibilities are also discussed in Chapter IV, such as the use of 3D-NMR or advanced biophysical methods. In combination with the mechanistic studies of heme in the TLR4 pathway, a holistic picture can possibly be painted in future studies.

In conclusion, this thesis provides a comprehensive account of nonapeptide model studies focused around one and two possible coordinating amino acids. The CP motif remains the preferred motif in protein examples but in several cases, H/Y-based motifs, such as HXH, HXXXH, and HXXXY, are also highly relevant. This array of suitable motifs and characteristics that define them can now be used to enter the next phase of heme research, which is the systematic identification of heme-binding proteins and their relevance in disease. In order to aid this cause, the knowledge gained from this and numerous other studies was made publicly available through the tool HeMoQuest. Two proteins, which are already known to be disease-relevant, JAK2 and TLR4, have been studied here. The knowledge obtained here and in the future with the provided tools will

advance our understanding of the role heme plays in pathogenesis of diseases, such as cardiovascular disease, ischemia reperfusion injury, transfusion injury and many more. Equipped with improved understanding, researchers may be able to apply known heme scavengers in a targeted fashion, or develop new medicines to manage heme toxicity.

Abbreviations

Aβ	Amyloid beta
ALA	δ-Aminolevulinic acid
ALAS1	δ-Aminolevulinate synthase 1
ALAS2	δ-Aminolevulinate synthase 2
APC	Activated protein C
ATP	Adenosine triphosphate
Bach1	BTB domain and CNC homolog 1
BMAL1	Brain and muscle ARNT-like 1
C1q	Complement component 1q
CD14	Cluster of differentiation 14
CD163	Cluster of differentiation 163
CLOCK	Circadian locomotor output cycles kaput
COSY	Correlation spectroscopy
CP	Cysteine-proline
DAMP	Damage-associated molecular pattern
DFT	Density-functional theory
DGCR8	DiGeorge syndrome critical region 8
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EGFR	Epidermal growth factor receptor
eIF2	Eukaryotic initiation factor 2
EPO	Erythropoietin
EPO-R	Erythropoietin receptor
EPR	Electron paramagnetic resonance
FERM	Band 4.1, ezrin, radixin, moesin
FID	Free induction decay
FLVCR1a	Feline leukemia virus subgroup C receptor-related protein 1 isoform a
FLVCR1b	Feline leukemia virus subgroup C receptor-related protein 1 isoform b
FLVCR2	Feline leukemia virus subgroup C receptor-related protein 2
GATA1	Globin transcription factor 1
GM-CSF	Granulocyte-macrophage colony-stimulating factor
H/Y	Histidine/tyrosine
HBM	Heme-binding motif
HemeKG	Heme knowledge graph
HO-1	Heme oxygenase 1
HO-2	Heme oxygenase 2
Hp	Haptoglobin
HRG1	Heme-responsive gene 1 protein homolog
HRI	Heme-regulated inhibitor kinase
HRM	Heme-regulatory motif
HSA	Humans serum albumin
HSQC	Heteronuclear single quantum coherence
IgG	Immunoglobulin G
IL-3	Interleukin-3
IL-36α	Interleukin-36 alpha
IRP2	Iron-responsive element-binding protein 2
Irr	Iron response regulator
JAK2	Janus kinase 2
JH1	Janus homology 1
JH2	Janus homology 2

K_{ATP}	ATP-sensitive potassium channel
LPS	Lipopolysaccharide
Maf	Musculoaponeurotic fibrosarcoma
MAPK	Mitogen-activated protein kinase
MD2	Myeloid differentiation factor 2
MyD88	Myeloid differentiation primary response 88
NAD	Nicotinamide adenine dinucleotide
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NOESY	Nuclear Overhauser enhancement spectroscopy
PDB	Protein data bank
PGRMC1	Progesterone receptor membrane component 1
PI3K	Phosphoinositide 3-kinase
PPIX	Protoporphyrin IX
RBCK1	RanBP-type and C3HC4-type zinc finger-containing protein 1
Rev-ERB α	Nuclear receptor subfamily 1 group D member 1
Rev-ERB β	Nuclear receptor subfamily 1 group D member 2
ROS	Reactive oxygen species
SbnI	Staphyloferrin B cluster 9-gene product
SCD	Sickle cell disease
SCF	Stem cell factor
sGC	Soluble guanylyl cyclase
SH2	Src homology 2
SH2L	Src homology 2-like
Slo1	Large conductance calcium-activated potassium channel
STAT3	Signal transducer and activator of transcription 3
STAT5	Signal transducer and activator of transcription 5
TNFα	Tumor necrosis factor alpha
TLR4	Toll-like receptor 4
TOCSY	Total correlated spectroscopy
TrpRS	Tryptophanyl-tRNA-synthetase
UV/Vis	Ultraviolet-visible
WESA	Weighted ensemble solvent accessibility

List of Figures

List of Tables

References

1. Hamza, I. & Dailey, H. A. One ring to rule them all: Trafficking of heme and heme synthesis intermediates in the metazoans. *Biochim. Biophys. Acta - Mol. Cell Res.* **1823,** 1617–1632 (2012).

2. Lukin, J. A., Simplaceanu, V., Zou, M., Ho, N. T. & Ho, C. NMR reveals hydrogen bonds between oxygen and distal histidines in oxyhemoglobin. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 10354–10358 (2000).

3. Munro, A. W., Girvan, H. M., McLean, K. J., Cheesman, M. R. & Leys, D. Heme and hemoproteins. in *Tetrapyrroles: Birth, life and death* (eds. Warren, M. J. & Smith, A. G.) 160–183 (Springer Science & Business Media, 2009).

4. Poulos, T. L. Soluble guanylate cyclase. *Curr. Opin. Struct. Biol.* **16,** 736–743 (2006).

5. Guengerich, F. P. Cytochromes P450, drugs, and diseases. *Mol. Interv.* **3,** 194–204 (2003).

6. Dunford, H. B., Araiso, T., Job, D., Ricard, J., Rutter, R., Hager, L. P., Wever, R., Kast, W. M., Boelens, R. & Ellfolk, N. Peroxidases. in *The Biological Chemistry of Iron* (eds. Dunford, B. H., Dolphin, D., Raymond, K. N. & Sieker, L.) 337–355 (Springer Netherlands, 1982).

7. Sies, H., Bücher, T., Oshino, N. & Chance, B. Heme occupancy of catalase in hemoglobin-free perfused rat liver and of isolated rat liver catalase. *Arch. Biochem. Biophys.* **154,** 106–116 (1973).

8. Montecinos, L., Eskew, J. D. & Smith, A. What is next in this "age" of heme-driven pathology and protection by hemopexin? An update and links with iron. *Pharmaceuticals* **12,** 144 (2019).

9. Frimat, M., Boudhabhay, I. & Roumenina, L. T. Hemolysis derived products toxicity and endothelium: Model of the second hit. *Toxins (Basel).* **11,** 1–34 (2019).

10. Kumar, S. & Bandyopadhyay, U. Free heme toxicity and its detoxification systems in human. *Toxicol. Lett.* **157,** 175–188 (2005).

11. Ascenzi, P., Bocedi, A., Visca, P., Altruda, F., Tolosano, E., Beringhelli, T. & Fasano, M. Hemoglobin and heme scavenging. *IUBMB Life* **57,** 749–759 (2005).

12. Tolosano, E., Fagoonee, S., Morello, N., Vinchi, F. & Fiorito, V. Heme scavenging and the other facets of hemopexin. *Antioxidants Redox Signal.* **12,** 305–320 (2010).

13. Chiabrando, D., Vinchi, F., Fiorito, V. & Tolosano, E. Haptoglobin and hemopexin in heme detoxification and iron recycling. in *Acute phase proteins - regulation and functions of acute phase proteins* (ed. Veas, F.) 261–288 (InTech, 2011).

14. Zhang, L. *Heme biology: The secret life of heme in regulating diverse biological processes.* (World Scientific, 2011).

15. Kühl, T. & Imhof, D. Regulatory FeII/III heme: The reconstruction of a molecule's biography. *ChemBioChem* **15,** 2024–2035 (2014).

16. Ponka, P., Sheftel, A. D., English, A. M., Bohle, D. S. & Garcia-Santos, D. Do mammalian cells really need to export and import heme? *Trends Biochem. Sci.* **42,** 395–406 (2017).

17. Korolnek, T. & Hamza, I. Like iron in the blood of the people: The requirement for heme trafficking in iron metabolism. *Front. Pharmacol.* **5,** 126 (2014).

18. Lathrop, J. T. & Timko, M. P. Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. *Science* **259,** 522–525 (1993).

19. Tang, X. D., Xu, R., Reynolds, M. F., Garcia, M. L., Heinemann, S. H. & Hoshi, T. Haem can bind to and inhibit mammalian calcium-dependent Slo1 BK channels. *Nature* **425,** 531–535 (2003).

20. Munakata, H., Sun, J. Y., Yoshida, K., Nakatani, T., Honda, E., Hayakawa, S., Furuyama, K. & Hayashi, N. Role of the heme regulatory motif in the heme-mediated inhibition of mitochondrial import of 5-aminolevulinate synthase. *J. Biochem.* **136,** 233–238 (2004).

21. Faller, M., Matsunaga, M., Yin, S., Loo, J. A. & Guo, F. Heme is involved in microRNA processing. *Nat. Struct. Mol. Biol.* **14,** 23–29 (2007).

22. Yin, L., Wu, N., Curtin, J. C., Qatanani, M., Szwergold, N. R., Reid, R. A., Waitt, G. M., Parks, D. J., Pearce, K. H., Wisely, G. B. & Lazar, M. A. Rev-erbα, a heme sensor that coordinates metabolic and circadian pathways. *Science* **318,** 1786–1789 (2007).

23. Yin, L., Wu, N. & Lazar, M. A. Nuclear receptor rev-erbα: A heme receptor that coordinates circadian rhythm and metabolism. *Nucl. Recept. Signal.* **8,** nrs.08001 (2010).

24. Paul George, A. A., Lacerda, M., Syllwasschy, B. F., Hopp, M.-T., Wißbrock, A. & Imhof, D. HeMoQuest: A webserver for qualitative prediction of transient heme binding to protein motifs. *BMC Bioinformatics* **21,** 124 (2020).

25. Roumenina, L. T., Radanova, M., Atanasov, B. P., Popov, K. T., Kaveri, S. V., Lacroix-Desmazes, S., Frémeaux-Bacchi, V. & Dimitrov, J. D. Heme interacts with C1q and inhibits the classical complement pathway. *J. Biol. Chem.* **286,** 16459–16469 (2011).

26. Wißbrock, A., Goradia, N. B., Kumar, A., Paul George, A. A., Kühl, T., Bellstedt, P., Ramachandra, R., Hoffmann, P., Galler, K., Popp, J., Neugebauer, U., Hampel, K., Zimmermann, B., Adam, S., Wiendl, M., Schett, G., Hamza, I., Heinemann, S. H., Frey, S., *et al.* Heme regulates human proinflammatory IL-36 cytokines. *Sci. Rep.* **9,** 16893 (2019).

27. Wißbrock, A., Paul George, A. A., Brewitz, H. H., Kühl, T. & Imhof, D. The molecular basis of transient heme-protein interactions: Analysis, concept and implementation. *Biosci. Rep.* **39,** BSR20181940 (2019).

28. Zhang, L. & Guarente, L. Heme binds to a short sequence that serves a regulatory function in diverse proteins. *EMBO J.* **14,** 313–320 (1995).

29. Kühl, T., Sahoo, N., Nikolajski, M., Schlott, B., Heinemann, S. H. & Imhof, D. Determination of hemin-binding characteristics of proteins by a combinatorial peptide library approach. *ChemBioChem* **12,** 2846–2855 (2011).

30. Kühl, T., Wißbrock, A., Goradia, N., Sahoo, N., Galler, K., Neugebauer, U., Popp, J., Heinemann, S. H., Ohlenschläger, O. & Imhof, D. Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins. *ACS Chem. Biol.* **8,** 1785–1793 (2013).

31. Brewitz, H. H., Kühl, T., Goradia, N., Galler, K., Popp, J., Neugebauer, U., Ohlenschläger, O. & Imhof, D. Role of the chemical environment beyond the coordination site: Structural insight into FeIII protoporphyrin binding to cysteine-based heme-regulatory protein motifs. *ChemBioChem* **16,** 2216–2224 (2015).

32. Brewitz, H. H., Goradia, N., Schubert, E., Galler, K., Kühl, T., Syllwasschy, B. F., Popp, J., Neugebauer, U., Hagelueken, G., Schiemann, O., Ohlenschläger, O. & Imhof, D. Heme interacts with histidine- and tyrosine-based protein motifs and inhibits enzymatic activity of chloramphenicol acetyltransferase from Escherichia coli. *Biochim. Biophys. Acta - Gen. Subj.* **1860,** 1343–1353 (2016).

33. Peherstorfer, S., Brewitz, H. H., Paul George, A. A., Wißbrock, A., Adam, J. M., Schmitt, L. & Imhof, D. Insights into mechanism and functional consequences of heme binding to hemolysin-activating lysine acyltransferase HlyC from Escherichia coli. *Biochim. Biophys. Acta - Gen. Subj.* **1862,** 1964–1972 (2018).

34. Kumar, A., Wißbrock, A., Goradia, N., Bellstedt, P., Ramachandran, R., Imhof, D. & Ohlenschläger, O. Heme interaction of the intrinsically disordered N-terminal peptide segment of human cystathionine-β-synthase. *Sci. Rep.* **8,** 1–9 (2018).

35. Hünefeld, F. L. *Der Chemismus in der thierischen Organisation: Physiologisch-chemische Untersuchungen der materiellen Veränderungen, oder des Bildungslebens im thierischen Organismus, insbesondere des Blutbildungsprocesses, der Natur der Blutkörperchen und ihrer Kernchen.* (Friedrich Arnold Brockhaus, 1840).

36. Schwedt, G. *Die Chemie des Lebens.* (Wiley-VCH Verlag GmbH & Co. KGaA, 2011).

37. Küster, W. Beiträge zur Kenntnis des Bilirubins und Hämins. *Hoppe-Seyler´s Zeitschrift für Physiol. Chemie* **82,** 463–483 (1912).

38. Fischer, H. & Zeile, K. Synthese des Hämatoporphyrins, Protoporphyrins und Hämins. *Justus Liebig's Ann. der Chemie* **468,** 98–116 (1929).

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

40. D'Alessandro, A. & Zolla, L. Proteomic analysis of red blood cells and the potential for the clinic: What have we learned so far? *Expert Rev. Proteomics* **14,** 243–252 (2017).

41. Warren, M. & Smith, A. *Tetrapyrroles: Birth, life, and death.* (Springer Science & Business Media, 2009).

42. Frankenberg, N., Schobert, M., Moser, J., Raux, E., Graham, R., Warren, M. J. & Jahn, D. The biosynthesis of hemes, siroheme, vitamin B12 and linear tetrapyrroles in pseudomonads. in *Pseudomonas* (ed. Ramos, J.-L.) 111–146 (2004).

43. Hederstedt, L. Heme A biosynthesis. *Biochim. Biophys. Acta - Bioenerg.* **1817,** 920–927 (2012).

44. Caughey, W. S. & York, J. L. Isolation and some properties of the green heme of cytochrome oxidase from beef heart muscle. *J. Biol. Chem.* **237,** 2414–2416 (1962).

45. Ponka, P. Cell biology of heme. *Am. J. Med. Sci.* **318,** 241–256 (1999).

46. Bowman, S. E. J. & Bren, K. L. The chemistry and biochemistry of heme c: Functional bases for covalent attachment. *Nat. Prod. Rep.* **25,** 1118–1130 (2008).

47. Mathews, F. S. The structure, function and evolution of cytochromes. *Prog. Biophys. Mol. Biol.* **45,** 1–56 (1985).

48. Beale, S. I. Biosynthesis of chlorophylls and hemes. in *The chlamydomonas*

sourcebook (eds. Harris, E. H., Stern, D. B. & Witman, G. B.) 731–798 (Elsevier, 2009).

49. Moss, G. P. Nomenclature of tetrapyrroles. *Eur. J. Biochem.* **178,** 277–328 (1988).

50. Hopp, M.-T., Schmalohr, B. F., Kühl, T., Detzel, M. S., Wissbrock, A. & Imhof, D. Heme determination and quantification methods and their suitability for practical applications and every-day-use. *Anal. Chem.* **92,** 9429–9440 (2020).

51. Larsen, R., Gouveia, Z., Soares, M. P. & Gozzelino, R. Heme cytotoxicity and the pathogenesis of immune-mediated inflammatory diseases. *Front. Pharmacol.* **3,** 77 (2012).

52. Cannon, J. B., Yunker, M. H. & Luoma, N. The effect of aggregation inhibitors and antioxidants on the stability of hemin solutions. *PDA J. Pharm. Sci. Technol.* **49,** 77–82 (1995).

53. Fufezan, C., Zhang, J. & Gunner, M. R. Ligand preference and orientation in b- and c-type heme-binding proteins. *Proteins Struct. Funct. Bioinforma.* **73,** 690–704 (2008).

54. Siegert, S. W. K. & Holt, R. J. Physicochemical properties, pharmacokinetics, and pharmacodynamics of intravenous hematin: A literature review. *Adv. Ther.* **25,** 842–857 (2008).

55. Kostrzewska, E., Gregor, A. & Tarczyńska-Nosal, S. Heme arginate (Normosang) in the treatment of attacks of acute hepatic porphyrias. *Polish J. Med. Pharm.* **23,** 259–262 (1991).

56. Reddi, A. R. & Hamza, I. Heme mobilization in animals: A metallolipid's journey. *Acc. Chem. Res.* **49,** 1104–1110 (2016).

57. Gouveia, Z., Carlos, A. R., Yuan, X., Aires-da-Silva, F., Stocker, R., Maghzal, G. J., Leal, S. S., Gomes, C. M., Todorovic, S., Iranzo, O., Ramos, S., Santos, A. C., Hamza, I., Gonçalves, J. & Soares, M. P. Characterization of plasma labile heme in hemolytic conditions. *FEBS J.* **284,** 3278–3301 (2017).

58. Belcher, J. D., Beckman, J. D., Balla, G., Balla, J. & Vercellotti, G. Heme degradation and vascular injury. *Antioxidants Redox Signal.* **12,** 233–248 (2010).

59. Asher, C., De Villiers, K. A. & Egan, T. J. Speciation of ferriprotoporphyrin IX in aqueous and mixed aqueous solution is controlled by solvent identity, pH, and salt concentration. *Inorg. Chem.* **48,** 7994–8003 (2009).

60. Imada, Y., Nakamura, H. & Takano, Y. Density functional study of porphyrin distortion effects on redox potential of heme. *J. Comput. Chem.* **39,** 143–150 (2018).

61. Petrich, J. W., Poyart, C. & Martin, J. L. Photophysics and reactivity of heme proteins: A femtosecond absorption study of hemoglobin, myoglobin, and protoheme. *Biochemistry* **27,** 4049–4060 (1988).

62. Duprat, A. F., Traylor, T. G., Wu, G. Z., Coletta, M., Sharma, V. S., Walda, K. N. & Magde, D. Myoglobin-NO at low pH: Free four-coordinated heme in the protein pocket. *Biochemistry* **34,** 2634–2644 (1995).

63. Battistuzzi, G., Bellei, M., Bortolotti, C. A. & Sola, M. Redox properties of heme peroxidases. *Arch. Biochem. Biophys.* **500,** 21–36 (2010).

64. Granik, S. Evolution of heme and chlorophyll. in *Evolving genes and proteins* (eds.

Bryson, V. & Vogel, H. J.) 67–88 (Elsevier, 1965).

65. Smith, A. G. & Witty, M. *Heme, Chlorophyll, and Bilins*. *Heme, Chlorophyll, and Bilins* (Humana Press, 2001).

66. Atamna, H., Liu, J. & Ames, B. N. Heme deficiency selectively interrupts assembly of mitochondrial complex IV in human fibroblasts: Relevance to aging. *J. Biol. Chem.* **276,** 48410–48416 (2001).

67. De Montellano, P. R. O. *Cytochrome P450: structure, mechanism, and biochemistry*. (Springer Science & Business Media, 2005).

68. Winterbourn, C. C. Toxicity of iron and hydrogen peroxide: The Fenton reaction. *Toxicol. Lett.* **82–83,** 969–974 (1995).

69. Ishikawa, S., Tamaki, S., Ohata, M., Arihara, K. & Itoh, M. Heme induces DNA damage and hyperproliferation of colonic epithelial cells via hydrogen peroxide produced by heme oxygenase: A possible mechanism of heme-induced colon cancer. *Mol. Nutr. Food Res.* **54,** 1182–1191 (2010).

70. Gozzelino, R. The pathophysiology of heme in the brain. *Curr. Alzheimer Res.* **13,** 174–184 (2015).

71. Chiabrando, D., Vinchi, F., Fiorito, V., Mercurio, S. & Tolosano, E. Heme in pathophysiology: A matter of scavenging, metabolism and trafficking across cell membranes. *Front. Pharmacol.* **5,** 61 (2014).

72. Putnam, C. D., Arvai, A. S., Bourne, Y. & Tainer, J. A. Active and inhibited human catalase structures: Ligand and NADPH binding and catalytic mechanism. *J. Mol. Biol.* **296,** 295–309 (2000).

73. Das, D. K. & Medhi, O. K. The role of heme propionate in controlling the redox potential of heme: Square wave voltammetry of protoporphyrinato IX iron(III) in aqueous surfactant micelles. *J. Inorg. Biochem.* **70,** 83–90 (1998).

74. Luck, A. N., Yuan, X., Voronin, D., Slatko, B. E., Hamza, I. & Foster, J. M. Heme acquisition in the parasitic filarial nematode Brugia malayi. *FASEB J.* **30,** 3501–3514 (2016).

75. Kapishnikov, S., Grolimund, D., Schneider, G., Pereiro, E., McNally, J. G., Als-Nielsen, J. & Leiserowitz, L. Unraveling heme detoxification in the malaria parasite by in situ correlative X-ray fluorescence microscopy and soft X-ray tomography. *Sci. Rep.* **7,** 7610 (2017).

76. Muller-Eberhard, U., Javid, J., Liem, H. H., Hanstein, A. & Hanna, M. Plasma concentrations of hemopexin, haptoglobin and heme in patients with various hemolytic diseases. *Blood* **32,** 811–815 (1968).

77. Wißbrock, A. & Imhof, D. A tough nut to crack: Intracellular detection and quantification of heme in malaria parasites by a genetically encoded protein sensor. *ChemBioChem* **18,** 1561–1564 (2017).

78. Hanna, D. A., Hu, R., Kim, H., Martinez-Guzman, O., Torres, M. P. & Reddi, A. R. Heme bioavailability and signaling in response to stress in yeast cells. *J. Biol. Chem.* **293,** 12378–12393 (2018).

79. Syntichaki, P., Samara, C. & Tavernarakis, N. The vacuolar H+-ATPase mediates intracellular acidification required for neurodegeneration in C. elegans. *Curr. Biol.* **15,** 1249–1254 (2005).

80. Egan, T. J., Mavuso, W. W. & Ncokazi, K. K. The mechanism of β-hematin formation in acetate solution. Parallels between hemozoin formation and biomineralization processes. *Biochemistry* **40,** 204–213 (2001).

81. Pek, R. H., Yuan, X., Rietzschel, N., Zhang, J., Jackson, L. K., Nishibori, E., Ribeiro, A., Simmons, W. R., Jagadeesh, J., Sugimoto, H., Alam, M. Z., Garrett, L., Haldar, M., Ralle, M., Phillips, J. D., Bodine, D. M., Hamza, I., Alam, Z., Garrett, L., *et al.* Hemozoin produced by mammals confers heme tolerance. *Elife* **8,** e49503 (2019).

82. Kuter, D., Venter, G. A., Naidoo, K. J. & Egan, T. J. Experimental and time-dependent density functional theory characterization of the UV-visible spectra of monomeric and μ-oxo dimeric ferriprotoporphyrin IX. *Inorg. Chem.* **51,** 10233–10250 (2012).

83. Srinivas, V. & Mohan Rao, C. Time profile of hemin aggregation: An analysis. *Biochem. Int.* **21,** 849–855 (1990).

84. Comer, J. & Zhang, L. Experimental methods for studying cellular heme signaling. *Cells* **7,** 47 (2018).

85. De Villiers, K. A., Kaschula, C. H., Egan, T. J. & Marques, H. M. Speciation and structure of ferriprotoporphyrin IX in aqueous solution: Spectroscopic and diffusion measurements demonstrate dimerization, but not μ-oxo dimer formation. *J. Biol. Inorg. Chem.* **12,** 101–117 (2007).

86. Brown, S. B., Jones, P. & Lantzke, I. R. Infrared evidence for an oxo-bridged (Fe-O-Fe) haemin dimer. *Nature* **223,** 960–961 (1969).

87. Loup, C., Lelièvre, J., Benoit-Vical, F. & Meunier, B. Trioxaquines and heme-artemisinin adducts inhibit the in vitro formation of hemozoin better than chloroquine. *Antimicrob. Agents Chemother.* **51,** 3768–3770 (2007).

88. Gouterman, M. Spectra of porphyrins. *J. Mol. Spectrosc.* **6,** 138–163 (1961).

89. Soret, J.-L. Analyse spectrale. Sur le spectre d'absorption du sang dans la partie violette et ultra-violette. *C. R. Hebd. Seances Acad. Sci.* **97,** 1269–1270 (1883).

90. Uttamlal, M. & Holmes-Smith, A. S. The excitation wavelength dependent fluorescence of porphyrins. *Chem. Phys. Lett.* **454,** 223–228 (2008).

91. Thordarson, P. Determining association constants from titration experiments in supramolecular chemistry. *Chem. Soc. Rev.* **40,** 1305–1323 (2011).

92. Syllwasschy, B. F., Beck, M. S., Družeta, I., Hopp, M.-T., Ramoji, A., Neugebauer, U., Nozinovic, S., Menche, D., Willbold, D., Ohlenschläger, O., Kühl, T. & Imhof, D. High-affinity binding and catalytic activity of His/Tyr-based sequences: Extending heme-regulatory motifs beyond CP. *Biochim. Biophys. Acta - Gen. Subj.* **1864,** 129603 (2020).

93. Spiro, T. G. & Burke, J. M. Protein Control of Porphyrin Conformation. Comparison of Resonance Raman Spectra of Heme Proteins with Mesoporphyrin IX Analogs. *J. Am. Chem. Soc.* **98,** 5482–5489 (1976).

94. Neugebauer, U., März, A., Henkel, T., Schmitt, M. & Popp, J. Spectroscopic detection and quantification of heme and heme degradation products. *Anal. Bioanal. Chem.* **404,** 2819–2829 (2012).

95. Spiro, T. G., Smulevich, G. & Su, C. Probing protein structure and dynamics with resonance Raman spectroscopy: Cytochrome c peroxidase and hemoglobin.

Biochemistry **29,** 4497–4508 (1990).

96. Kitagawa, T., Abe, M., Kyogoku, Y., Ogoshi, H., Watanabe, E. & Yoshida, Z. Resonance Raman spectra of metallooctaethylporphyrins. Low frequency vibrations of porphyrin and iron-axial ligand stretching modes. *J. Phys. Chem.* **80,** 1181–1186 (1976).

97. Czernuszewicz, R. S., Macor, K. A., Li, X. Y., Kincaid, J. R. & Spiro, T. G. Resonance Raman spectroscopy reveals a1u vs. a2u character and pseudo-Jahn-Teller distortion in radical cations of nickel(II), copper(II), and chloroiron(III) octaethyl- and tetraphenylporphyrins. *J. Am. Chem. Soc.* **111,** 3860–3869 (1989).

98. Farrar, T. C. & Becker, E. D. *Pulse and Fourier transform NMR: Introduction to theory and methods*. (Academic Press Inc., 1971).

99. Roberts, J. D. *Nuclear magnetic resonance: Applications to organic chemistry*. (McGraw-Hill Book Company, 1959).

100. Köhler, F. H. Paramagnetic Complexes in Solution: The NMR Approach. in *Encyclopedia of Magnetic Resonance* (eds. Harris, R. K. & Wasylishen, R.) (John Wiley & Sons, Ltd, 2011).

101. Wolff, N., Izadi-Pruneyre, N., Couprie, J., Habeck, M., Linge, J., Rieping, W., Wandersman, C., Nilges, M., Delepierre, M. & Lecroisey, A. Comparative analysis of structural and dynamic properties of the loaded and unloaded hemophore HasA: Functional implications. *J. Mol. Biol.* **376,** 517–525 (2008).

102. Kumar, A., Wißbrock, A., Bellstedt, P., Lang, A., Ramachandran, R., Wiedemann, C., Imhof, D. & Ohlenschläger, O. 1H, 13C, and 15N resonance assignments of the cytokine interleukin-36β isoform-2. *Biomol. NMR Assign.* **13,** 155–161 (2019).

103. Wißbrock, A., Goradia, N. B., Kumar, A., Paul George, A. A., Kühl, T., Bellstedt, P., Ramachandran, R., Hoffmann, P., Galler, K., Popp, J., Neugebauer, U., Hampel, K., Zimmermann, B., Adam, S., Wiendl, M., Krönke, G., Hamza, I., Heinemann, S. H., Frey, S., *et al.* Structural insights into heme binding to IL-36α proinflammatory cytokine. *Sci. Rep.* **9,** 16893 (2019).

104. Sinclair, P. R., Gorman, N. & Jacobs, J. M. Measurement of heme concentration. *Curr. Protoc. Toxicol.* **00,** 8.3.1-8.3.7 (1999).

105. Zámocký, M. & Koller, F. Understanding the structure and function of catalases: Clues from molecular evolution and in vitro mutagenesis. *Prog. Biophys. Mol. Biol.* **72,** 19–66 (1999).

106. Bernroitner, M., Zamocky, M., Furtmüller, P. G., Peschek, G. A. & Obinger, C. Occurrence, phylogeny, structure, and function of catalases and peroxidases in cyanobacteria. *J. Exp. Bot.* **60,** 423–440 (2009).

107. Smith, L. J., Kahraman, A. & Thornton, J. M. Heme proteins - Diversity in structural characteristics, function, and folding. *Proteins Struct. Funct. Bioinforma.* **78,** 2349–2368 (2010).

108. Smith, M. L., Paul, J., Ohlsson, P. I., Hjortsberg, K. & Paul, K. G. Heme-protein fission under nondenaturing conditions. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 882–886 (1991).

109. Krieger, E. & Vriend, G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* **30,** 2981–2982 (2014).

110. Heinemann, I. U., Jahn, M. & Jahn, D. The biochemistry of heme biosynthesis.

Arch. Biochem. Biophys. **474,** 238–251 (2008).

111. Goodfellow, B. J., Dias, J. S., Ferreira, G. C., Henklein, P., Wray, V. & Macedo, A. L. The solution structure and heme binding of the presequence of murine 5-aminolevulinate synthase. *FEBS Lett.* **505,** 325–331 (2001).

112. Kubota, Y., Nomura, K., Katoh, Y., Yamashita, R., Kaneko, K. & Furuyama, K. Novel mechanisms for heme-dependent degradation of ALAS1 protein as a component of negative feedback regulation of heme biosynthesis. *J. Biol. Chem.* **291,** 20516–20529 (2016).

113. Eisenstein, R. S. Iron regulatory proteins and the molecular control of mammalian iron metabolism. *Annu. Rev. Nutr.* **20,** 627–662 (2000).

114. Ishikawa, H., Kato, M., Hori, H., Ishimori, K., Kirisako, T., Tokunaga, F. & Iwai, K. Involvement of heme regulatory motif in heme-mediated ubiquitination and degradation of IRP2. *Mol. Cell* **19,** 171–181 (2005).

115. Ogawa, K., Sun, J., Taketani, S., Nakajima, O., Nishitani, C., Sassa, S., Hayashi, N., Yamamoto, M., Shibahara, S., Fujita, H. & Igarashi, K. Heme mediates derepression of Maf recognition element through direct binding to transcription repressor Bach1. *EMBO J.* **20,** 2835–2843 (2001).

116. Tahara, T., Sun, J., Nakanishi, K., Yamamoto, M., Mori, H., Saito, T., Fujita, H., Igarashi, K. & Taketani, S. Heme positively regulates the expression of β-globin at the locus control region via the transcriptional factor Bach1 in erythroid cells. *J. Biol. Chem.* **279,** 5480–5487 (2004).

117. Sun, J., Brand, M., Zenke, Y., Tashiro, S., Groudine, M. & Igarashi, K. Heme regulates the dynamic exchange of Bach1 and NF-E2-related factors in the Maf transcription factor network. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 1461–1466 (2004).

118. Suzuki, H., Tashiro, S., Hira, S., Sun, J., Yamazaki, C., Zenke, Y., Ikeda-Saito, M., Yoshida, M. & Igarashi, K. Heme regulates gene expression by triggering Crm1-dependent nuclear export of Bach1. *EMBO J.* **23,** 2544–2553 (2004).

119. Zenke-Kawasaki, Y., Dohi, Y., Katoh, Y., Ikura, T., Ikura, M., Asahara, T., Tokunaga, F., Iwai, K. & Igarashi, K. Heme induces ubiquitination and degradation of the transcription factor Bach1. *Mol. Cell. Biol.* **27,** 6962–6971 (2007).

120. Miksanova, M., Igarashi, J., Minami, M., Sagami, I., Yamauchi, S., Kurokawa, H. & Shimizu, T. Characterization of heme-regulated eIF2α kinase: Roles of the N-terminal domain in the oligomeric state, heme binding, catalysis, and inhibition. *Biochemistry* **45,** 9894–9905 (2006).

121. Kaasik, K. & Lee, C. C. Reciprocal regulation of haem biosynthesis and the circadian clock in mammals. *Nature* **430,** 467–471 (2004).

122. Yang, J., Kim, K. D., Lucas, A., Drahos, K. E., Santos, C. S., Mury, S. P., Capelluto, D. G. S. & Finkielstein, C. V. A novel heme-regulatory motif mediates heme-dependent degradation of the circadian factor period 2. *Mol. Cell. Biol.* **28,** 4697–4711 (2008).

123. Freeman, S. L., Kwon, H., Portolano, N., Parkin, G., Girija, U. V., Basran, J., Fielding, A. J., Fairall, L., Svistunenko, D. A., Moody, P. C. E., Schwabe, J. W. R., Kyriacou, C. P. & Raven, E. L. Heme binding to human CLOCK affects interactions with the E-box. *Proc. Natl. Acad. Sci. U. S. A.* **116,** 19911–19916

(2019).

124. Raghuram, S., Stayrook, K. R., Huang, P., Rogers, P. M., Nosie, A. K., McClure, D. B., Burris, L. L., Khorasanizadeh, S., Burris, T. P. & Rastinejad, F. Identification of heme as the ligand for the orphan nuclear receptors REV-ERBα and REV-ERBβ. *Nat. Struct. Mol. Biol.* **14,** 1207–1213 (2007).

125. Carter, E. L., Ramirez, Y. & Ragsdale, S. W. The heme-regulatory motif of nuclear receptor Rev-erbβ is a key mediator of heme and redox signaling in circadian rhythm maintenance and metabolism. *J. Biol. Chem.* **292,** 11280–11299 (2017).

126. Horrigan, F. T., Heinemann, S. H. & Hoshi, T. Heme regulates allosteric activation of the Slo1 BK channel. *J. Gen. Physiol.* **126,** 7–21 (2005).

127. Burton, M. J., Kapetanaki, S. M., Chernova, T., Jamieson, A. G., Dorlet, P., Santolini, J., Moody, P. C. E., Mitcheson, J. S., Davies, N. W., Schmid, R., Raven, E. L. & Storey, N. M. A heme-binding domain controls regulation of ATP-dependent potassium channels. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 3785–3790 (2016).

128. Sahoo, N., Goradia, N., Ohlenschläger, O., Schönherr, R., Friedrich, M., Plass, W., Kappl, R., Hoshi, T. & Heinemann, S. H. Heme impairs the ball-and-chain inactivation of potassium channels. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E4036–E4044 (2013).

129. Dimitrov, J. D., Roumenina, L. T., Doltchinkova, V. R., Mihaylova, N. M., Lacroix-Desmazes, S., Kaveri, S. V. & Vassilev, T. L. Antibodies use heme as a cofactor to extend their pathogen elimination activity and to acquire new effector functions. *J. Biol. Chem.* **282,** 26696–26706 (2007).

130. Figueiredo, R. T., Fernandez, P. L., Mourao-Sa, D. S., Porto, B. N., Dutra, F. F., Alves, L. S., Oliveira, M. F., Oliveira, P. L., Graça-Souza, A. V. & Bozza, M. T. Characterization of heme as activator of toll-like receptor 4. *J. Biol. Chem.* **282,** 20221–20229 (2007).

131. Repessé, Y., Dimitrov, J. D., Peyron, I., Moshai, E. F., Kiger, L., Dasgupta, S., Delignat, S., Marden, M. C., Kaveri, S. V. & Lacroix-Desmazes, S. Heme binds to factor VIII and inhibits its interaction with activated factor IX. *J. Thromb. Haemost.* **10,** 1062–1071 (2012).

132. Hopp, M.-T., Alhanafi, N., Paul George, A. A., Hamedani, N. S., Biswas, A., Oldenburg, J., Pötzsch, B. & Imhof, D. Molecular insights and functional consequences of the interaction of heme with activated protein C. *Antioxid. Redox Signal.* **34,** 32–48 (2021).

133. Pardee, K. I., Xu, X., Reinking, J., Schuetz, A., Dong, A., Liu, S., Zhang, R., Tiefenbach, J., Lajoie, G., Plotnikov, A. N., Botchkarev, A., Krause, H. M. & Edwards, A. The structural basis of gas-responsive transcription by the human nuclear hormone receptor REV-ERBβ. *PLoS Biol.* **7,** e1000043 (2009).

134. Caillet-Saguy, C., Turano, P., Piccioli, M., Lukat-Rodgers, G. S., Czjzek, M., Guigliarelli, B., Izadi-Pruneyre, N., Rodgers, K. R., Delepierre, M. & Lecroisey, A. Deciphering the structural role of histidine 83 for heme binding in hemophore HasA. *J. Biol. Chem.* **283,** 5960–5970 (2008).

135. Atamna, H. & Frey, W. H. A role for heme in Alzheimer's disease: Heme binds amyloid β and has altered metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 11153–11158 (2004).

136. Dias, J. S., Macedo, A. L., Ferreira, G. C., Peterson, F. C., Volkman, B. F. & Goodfellow, B. J. The first structure from the SOUL/HBP family of heme-binding proteins, murine P22HBP. *J. Biol. Chem.* **281,** 31553–31561 (2006).

137. Mense, S. M. & Zhang, L. Heme: a versatile signaling molecule controlling the activities of diverse regulators ranging from transcription factors to MAP kinases. *Cell Res.* **16,** 681–692 (2006).

138. Yang, J., Ishimori, K. & O'Brian, M. R. Two heme binding sites are involved in the regulated degradation of the bacterial iron response regulator (Irr) protein. *J. Biol. Chem.* **280,** 7671–7676 (2005).

139. Hira, S., Tomita, T., Matsui, T., Igarashi, K. & Ikeda-Saito, M. Bach1, a heme-dependent transcription factor, reveals presence of multiple heme binding sites with distinct coordination structure. *IUBMB Life* **59,** 542–551 (2007).

140. Igarashi, J., Murase, M., Iizuka, A., Pichierri, F., Martinkova, M. & Shimizu, T. Elucidation of the heme binding site of heme-regulated eukaryotic initiation factor 2α kinase and the role of the regulatory motif in heme sensing by spectroscopic and catalytic studies of mutant proteins. *J. Biol. Chem.* **283,** 18782–18791 (2008).

141. Yamanaka, K., Ishikawa, H., Megumi, Y., Tokunaga, F., Kanie, M., Rouault, T. A., Morishima, I., Minato, N., Ishimori, K. & Iwai, K. Identification of the ubiquitin-protein ligase that recognizes oxidized IRP2. *Nat. Cell Biol.* **5,** 336–340 (2003).

142. Noé, R., Bozinovic, N., Lecerf, M., Lacroix-Desmazes, S. & Dimitrov, J. D. Use of cysteine as a spectroscopic probe for determination of heme-scavenging capacity of serum proteins and whole human serum. *J. Pharm. Biomed. Anal.* **172,** 311–319 (2019).

143. Aragão, D., Frazão, C., Sieker, L., Sheldrick, G. M., LeGall, J. & Carrondo, M. A. Structure of dimeric cytochrome c3 from Desulfovibrio gigas at 1.2 Å resolution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59,** 644–653 (2003).

144. Correia, I. J., Paquete, C. M., Coelho, A., Almeida, C. C., Catarino, T., Louro, R. O., Frazão, C., Saraiva, L. M., Carrondo, M. A., Turner, D. L. & Xavier, A. V. Proton-assisted two-electron transfer in natural variants of tetraheme cytochromes from Desulfomicrobium sp. *J. Biol. Chem.* **279,** 52227–52237 (2004).

145. Fülöp, V., Sam, K. A., Ferguson, S. J., Ginger, M. L. & Allen, J. W. A. Structure of a trypanosomatid mitochondrial cytochrome c with heme attached via only one thioether bond and implications for the substrate recognition requirements of heme lyase. *FEBS J.* **276,** 2822–2832 (2009).

146. Stevens, J. M., Daltrop, O., Allen, J. W. A. & Ferguson, S. J. C -type cytochrome formation: Chemical and biological enigmas. *Acc. Chem. Res.* **37,** 999–1007 (2004).

147. Yi, L., Morgan, J. T. & Ragsdale, S. W. Identification of a thiol/disulfide redox switch in the human BK channel that controls its affinity for heme and CO. *J. Biol. Chem.* **285,** 20117–20127 (2010).

148. Ogura, M., Endo, R., Ishikawa, H., Takeda, Y., Uchida, T., Iwai, K., Kobayashi, K. & Ishimori, K. Redox-dependent axial ligand replacement and its functional significance in heme-bound iron regulatory proteins. *J. Inorg. Biochem.* **182,** 238–248 (2018).

149. Brewitz, H. H., Hagelueken, G. & Imhof, D. Structural and functional diversity of

transient heme binding to bacterial proteins. *Biochim. Biophys. Acta - Gen. Subj.* **1861,** 683–697 (2017).

150. Paoli, M., Anderson, B. F., Baker, H. M., Morgan, W. T., Smith, A. & Baker, E. N. Crystal structure of hemopexin reveals a novel high-affinity heme site formed between two β-propeller domains. *Nat. Struct. Biol.* **6,** 926–931 (1999).

151. Hrkal, Z., Vodrážka, Z. & Kalousek, I. Transfer of heme from ferrihemoglobin and ferrihemoglobin isolated chains to hemopexin. *Eur. J. Biochem.* **43,** 73–78 (1974).

152. Detzel, M. S., Syllwasschy, B. F., Steinbock, F., Ramoji, A., Hopp, M.-T., Paul George, A. A., Neugebauer, U. & Imhof, D. Revisiting the interaction of heme with hemopexin: Recommendations for the responsible use of an emerging drug. *bioRxiv* (2020).

153. Wardell, M., Wang, Z., Ho, J. X., Robert, J., Ruker, F., Ruble, J. & Carter, D. C. The atomic structure of human methemalbumin at 1.9 Å. *Biochem. Biophys. Res. Commun.* **291,** 813–819 (2002).

154. Zunszain, P. A., Ghuman, J., Komatsu, T., Tsuchida, E. & Curry, S. Crystal structural analysis of human serum albumin complexed with hemin and fatty acid. *BMC Struct. Biol.* **3,** 1–9 (2003).

155. Adams, P. A. & Berman, M. C. Kinetics and mechanism of the interaction between human serum albumin and monomeric haemin. *Biochem. J.* **191,** 95–102 (1980).

156. Wißbrock, A., Kühl, T., Silbermann, K., Becker, A. J., Ohlenschläger, O. & Imhof, D. Synthesis and evaluation of amyloid β derived and amyloid β independent enhancers of the peroxidase-like activity of heme. *J. Med. Chem.* **60,** 373–385 (2017).

157. Wakasugi, K. Human tryptophanyl-tRNA synthetase binds with heme to enhance its aminoacylation activity. *Biochemistry* **46,** 11291–11298 (2007).

158. Kabe, Y., Nakane, T., Koike, I., Yamamoto, T., Sugiura, Y., Harada, E., Sugase, K., Shimamura, T., Ohmura, M., Muraoka, K., Yamamoto, A., Uchida, T., Iwata, S., Yamaguchi, Y., Krayukhina, E., Noda, M., Handa, H., Ishimori, K., Uchiyama, S., *et al.* Haem-dependent dimerization of PGRMC1/Sigma-2 receptor facilitates cancer proliferation and chemoresistance. *Nat. Commun.* **7,** 11030 (2016).

159. Min, L., Strushkevich, N. V., Harnastai, I. N., Iwamoto, H., Gilep, A. A., Takemori, H., Usanov, S. A., Nonaka, Y., Hori, H., Vinson, G. P. & Okamoto, M. Molecular identification of adrenal inner zone antigen as a heme-binding protein. *FEBS J.* **272,** 5832–5843 (2005).

160. Lukat-Rodgers, G. S., Correia, C., Botuyan, M. V., Mer, G. & Rodgers, K. R. Heme-based sensing by the mammalian circadian protein CLOCK. *Inorg. Chem.* **49,** 6349–6365 (2010).

161. Marvin, K. A., Reinking, J. L., Lee, A. J., Pardee, K., Krause, H. M. & Burstyn, J. N. Nuclear receptors Homo sapiens rev-erbβ and Drosophila melanogaster E75 are thiolate-ligated heme proteins which undergo redox-mediated ligand switching and bind CO and NO. *Biochemistry* **48,** 7056–7071 (2009).

162. Laakso, H. A., Marolda, C. L., Pinter, T. B., Stillman, M. J. & Heinrichs, D. E. A heme-responsive regulator controls synthesis of Staphyloferrin B in Staphylococcus aureus. *J. Biol. Chem.* **291,** 29–40 (2016).

163. Singleton, C., White, G. F., Todd, J. D., Marritt, S. J., Cheesman, M. R., Johnston,

A. W. B. & Le Brun, N. E. Heme-responsive DNA binding by the global iron regulator Irr from Rhizobium leguminosarum. *J. Biol. Chem.* **285,** 16023–16031 (2010).

164. Hopp, M.-T., Domingo-Fernández, D., Gadiya, Y., Detzel, M. S., Schmalohr, B. F., Steinbock, F., Imhof, D. & Hofmann-Apitius, M. Unravelling the debate on heme effects in COVID-19 infections. *bioRxiv* (2020).

165. Liu, R. & Hu, J. HemeBIND: A novel method for heme binding residue prediction by combining structural and sequence information. *BMC Bioinformatics* **12,** 207 (2011).

166. Liu, R. & Hu, J. Computational prediction of heme-binding residues by exploiting residue interaction network. *PLoS One* **6,** e25560 (2011).

167. Yu, D. J., Hu, J., Yang, J., Shen, H. Bin, Tang, J. & Yang, J. Y. Designing template-free predictor for targeting protein-ligand binding sites with classifier ensemble and spatial clustering. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **10,** 994–1008 (2013).

168. Liou, Y. F., Charoenkwan, P., Srinivasulu, Y. S., Vasylenko, T., Lai, S. C., Lee, H. C., Chen, Y. H., Huang, H. L. & Ho, S. Y. SCMHBP: Prediction and analysis of heme binding proteins using propensity scores of dipeptides. *BMC Bioinformatics* **15,** 1–14 (2014).

169. Zhang, J., Chai, H., Gao, B., Yang, G. & Ma, Z. HEMEsPred: Structure-based ligand-specific heme binding residues prediction by using fast-adaptive ensemble learning scheme. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **15,** 147–156 (2016).

170. Krieger, E. & Vriend, G. New ways to boost molecular dynamics simulations. *J. Comput. Chem.* **36,** 996–1007 (2015).

171. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular docking: A powerful approach for structure-based drug discovery. *Curr. Comput. Aided-Drug Des.* **7,** 146–157 (2011).

172. Novoa, E. M., De Pouplana, L. R., Barril, X. & Orozco, M. Ensemble docking from homology models. *J. Chem. Theory Comput.* **6,** 2547–2557 (2010).

173. Chen, Y.-C. Beware of docking! *Trends Pharmacol. Sci.* **36,** 78–95 (2015).

174. Hammes, G. G. Multiple conformational changes in enzyme catalysis. *Biochemistry* **41,** 8221–8228 (2002).

175. Ribeiro, J. M. L., Tsai, S. T., Pramanik, D., Wang, Y. & Tiwary, P. Kinetics of ligand-protein dissociation from all-atom simulations: Are we there yet? *Biochemistry* **58,** 156–165 (2019).

176. Hub, J. S., De Groot, B. L., Grubmüller, H. & Groenhof, G. Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.* **10,** 381–390 (2014).

177. Paul George, A. A., Heimer, P., Leipold, E., Schmitz, T., Kaufmann, D., Tietze, D., Heinemann, S. H. & Imhof, D. Effect of Conformational Diversity on the Bioactivity of μ-Conotoxin PIIIA Disulfide Isomers. *Mar. Drugs* **17,** 390 (2019).

178. Wouters, H. J. C. M., van der Klauw, M. M., de Witte, T., Stauder, R., Swinkels, D. W., Wolffenbuttel, B. H. R. & Huls, G. Association of anemia with health-related quality of life and survival: a large population-based cohort study.

Haematologica **104,** 468–476 (2019).

179. Alam, M. Z., Devalaraja, S. & Haldar, M. The heme connection: Linking erythrocytes and macrophage biology. *Front. Immunol.* **8,** 6–11 (2017).

180. Sassa, S. Why heme needs to be degraded to iron, biliverdin IXα, and carbon monoxide? *Antioxidants Redox Signal.* **6,** 819–824 (2004).

181. Wang, C. Y. & Babitt, J. L. Liver iron sensing and body iron homeostasis. *Blood* **133,** 18–29 (2019).

182. Vallelian, F., Schaer, C. A., Deuel, J. W., Ingoglia, G., Humar, R., Buehler, P. W. & Schaer, D. J. Revisiting the putative role of heme as a trigger of inflammation. *Pharmacol. Res. Perspect.* **6,** e00392 (2018).

183. Mateus, V., Rocha, J., Mota-Filipe, H., Sepodes, B. & Pinto, R. Hemin reduces inflammation associated with TNBS-induced colitis. *Clin. Exp. Gastroenterol.* **11,** 325–334 (2018).

184. Kayama, H., Kohyama, M., Okuzaki, D., Motooka, D., Barman, S., Okumura, R., Muneta, M., Hoshino, K., Sasaki, I., Ise, W., Matsuno, H., Nishimura, J., Kurosaki, T., Nakamura, S., Arase, H., Kaisho, T. & Takeda, K. Heme ameliorates dextran sodium sulfate-induced colitis through providing intestinal macrophages with noninflammatory profiles. *Proc. Natl. Acad. Sci. U. S. A.* **115,** 8418–8423 (2018).

185. Fortes, G. B., Alves, L. S., De Oliveira, R., Dutra, F. F., Rodrigues, D., Fernandez, P. L., Souto-Padron, T., De Rosa, M. J., Kelliher, M., Golenbock, D., Chan, F. K. M. & Bozza, M. T. Heme induces programmed necrosis on macrophages through autocrine TNF and ROS production. *Blood* **119,** 2368–2375 (2012).

186. Lin, T., Kwak, Y. H. H., Sammy, F., He, P., Thundivalappil, S., Sun, G., Chao, W. & Warren, H. S. S. Synergistic inflammation is induced by blood degradation products with microbial Toll-like receptor agonists and is blocked by hemopexin. *J. Infect. Dis.* **202,** 624–632 (2010).

187. Lin, S., Yin, Q., Zhong, Q., Lv, F.-L., Zhou, Y., Li, J.-Q., Wang, J.-Z., Su, B. & Yang, Q.-W. Heme activates TLR4-mediated inflammatory injury via MyD88/TRIF signaling pathway in intracerebral hemorrhage. *J. Neuroinflammation* **9,** 46 (2012).

188. Doty, R. T., Yan, X., Lausted, C., Munday, A. D., Yang, Z., Yi, D., Jabbari, N., Liu, L., Keel, S. B., Tian, Q. & Abkowitz, J. L. Single-cell analyses demonstrate that a heme-GATA1 feedback loop regulates red cell differentiation. *Blood* **133,** 457–469 (2019).

189. Chung, J., Wittig, J. G., Ghamari, A., Maeda, M., Lodish, H., Bauer, D. E., Orkin, S. H., Cantor, A. B., Maeda, T., Dailey, H. A. & Paw, B. H. Erythropoietin signaling regulates heme biosynthesis. *Blood* **128,** 543–543 (2016).

190. Mercurio, S., Petrillo, S., Chiabrando, D., Bassi, Z. I., Gays, D., Camporeale, A., Vacaru, A., Miniscalco, B., Valperga, G., Silengo, L., Altruda, F., Baron, M. H., Santoro, M. M. & Tolosano, E. The heme exporter Flvcr1 regulates expansion and differentiation of committed erythroid progenitors by controlling intracellular heme accumulation. *Haematologica* **100,** 720–729 (2015).

191. Socolovsky, M., Lodish, H. F. & Daley, G. Q. Control of hematopoietic differentiation: Lack of specificity in signaling by cytokine receptors. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 6573–6575 (1998).

192. Funakoshi-Tago, M., Pelletier, S., Moritake, H., Parganas, E. & Ihle, J. N. Jak2 FERM domain interaction with the erythropoietin receptor regulates Jak2 kinase activity. *Mol. Cell. Biol.* **28,** 1792–1801 (2008).

193. Feng, J., Witthuhn, B. A., Matsuda, T., Kohlhuber, F., Kerr, I. M. & Ihle, J. N. Activation of Jak2 catalytic activity requires phosphorylation of Y1007 in the kinase activation loop. *Mol. Cell. Biol.* **17,** 2497–2501 (1997).

194. Kerenyi, M. A., Grebien, F., Gehart, H., Schifrer, M., Artaker, M., Kovacic, B., Beug, H., Moriggl, R. & Müllner, E. W. Stat5 regulates cellular iron uptake of erythroid cells via IRP-2 and TfR-1. *Blood* **112,** 3878–3888 (2008).

195. Oda, A., Sawada, K., Druker, B. J., Ozaki, K., Takano, H., Koizumi, K., Fukada, Y., Handa, M., Koike, T. & Ikeda, Y. Erythropoietin induces tyrosine phosphorylation of Jak2, STAT5A, and STAT5B in primary cultured human erythroid precursors. *Blood* **92,** 443–451 (1998).

196. Yao, H., Ma, Y., Hong, Z., Zhao, L., Monaghan, S. A., Hu, M. C. & Huang, L. J. Activating JAK2 mutants reveal cytokine receptor coupling differences that impact outcomes in myeloproliferative neoplasm. *Leukemia* **31,** 2122–2131 (2017).

197. Gutiérrez, L., Caballero, N., Fernández-Calleja, L., Karkoulia, E. & Strouboulis, J. Regulation of GATA1 levels in erythropoiesis. *IUBMB Life* **72,** 89–105 (2020).

198. Rio, S., Gastou, M., Karboul, N., Derman, R., Suriyun, T., Manceau, H., Leblanc, T., El Benna, J., Schmitt, C., Azouzi, S., Larghéro, J., Karim, Z., MacIas-Garcia, A., Chen, J. J., Hermine, O., Courtois, G., Puy, H., Gouya, L., Mohandas, N., *et al.* Regulation of globin-heme balance in Diamond-Blackfan anemia by HSP70/GATA1. *Blood* **133,** 1358–1370 (2019).

199. Rose, M. Y. & Olson, J. The kinetic mechanism of heme biding to human apohemoglobin. *J. Biol. Chem.* **258,** 4298–4303 (1983).

200. Duffy, S. P., Shing, J., Saraon, P., Berger, L. C., Eiden, M. V, Wilde, A. & Tailor, C. S. The Fowler syndrome-associated protein FLVCR2 is an importer of heme. *Mol. Cell. Biol.* **30,** 5318–5324 (2010).

201. Yao, X., Balamurugan, P., Arvey, A., Leslie, C. & Zhang, L. Heme controls the regulation of protein tyrosine kinases Jak2 and Src. *Biochem. Biophys. Res. Commun.* **403,** 30–35 (2010).

202. Camaschella, C. New insights into iron deficiency and iron deficiency anemia. *Blood Rev.* **31,** 225–233 (2017).

203. Vinchi, F., Ingoglia, G., Chiabrando, D., Mercurio, S., Turco, E., Silengo, L., Altruda, F. & Tolosano, E. Heme exporter FLVCR1a regulates heme synthesis and degradation and controls activity of cytochromes P450. *Gastroenterology* **146,** 1325–1338 (2014).

204. Khan, A. A. & Quigley, J. G. Heme and FLVCR-related transporter families SLC48 and SLC49. *Mol. Aspects Med.* **34,** 669–682 (2013).

205. Andersson, L. C., Nilsson, K. & Gahmberg, C. G. K562—A human erythroleukemic cell line. *Int. J. Cancer* **23,** 143–147 (1979).

206. Hubbard, S. R. Mechanistic insights into regulation of JAK2 tyrosine kinase. *Front. Endocrinol.* **8,** 361 (2018).

207. Dale, M., Abraham, E., Editor, A., Wagner, P. D. & Spivak, J. L. Narrative review: Thrombocytosis, polycythemia vera, and JAK2 mutations: The phenotypic

mimicry of chronic myeloproliferation. *Ann. Intern. Med.* **152,** 300–306 (2010).

208. Saharinen, P., Takaluoma, K. & Silvennoinen, O. Regulation of the JAK2 tyrosine kinase by its pseudokinase domain. *Mol. Cell. Biol.* **20,** 3387–3395 (2000).

209. Ungureanu, D., Wu, J., Pekkala, T., Niranjan, Y., Young, C., Jensen, O. N., Xu, C. F., Neubert, T. A., Skoda, R. C., Hubbard, S. R. & Silvennoinen, O. The pseudokinase domain of JAK2 is a dual-specificity protein kinase that negatively regulates cytokine signaling. *Nat. Struct. Mol. Biol.* **18,** 971–976 (2011).

210. Ayaz, P., Hammarén, H. M., Raivola, J., Sharon, D., Hubbard, S. R., Silvennoinen, O., Shan, Y. & Shaw, D. E. Structural models of full-length JAK2 kinase. *bioRxiv* (2019).

211. Effenberger-Neidnicht, K. & Hartmann, M. Mechanisms of hemolysis during sepsis. *Inflammation* **41,** 1569–1581 (2018).

212. Vinchi, F., Da Silva, M. C., Ingoglia, G., Petrillo, S., Brinkman, N., Zuercher, A., Cerwenka, A., Tolosano, E. & Muckenthaler, M. U. Hemopexin therapy reverts heme-induced proinflammatory phenotypic switching of macrophages in a mouse model of sickle cell disease. *Blood* **127,** 473–486 (2016).

213. Ashley-Koch, A., Yang, Q. & Olney, R. S. Sickle hemoglobin (Hb S) allele and sickle cell disease: A HuGE review. *Am. J. Epidemiol.* **151,** 839–845 (2000).

214. Morris, C. R., Kuypers, F. A., Kato, G. J., Lavrisha, L., Larkin, S., Singer, T. & Vichinsky, E. P. Hemolysis-associated pulmonary hypertension in thalassemia. *Ann. N. Y. Acad. Sci.* **1054,** 481–485 (2005).

215. Orf, K. & Cunnington, A. J. Infection-related hemolysis and susceptibility to Gram-negative bacterial co-infection. *Front. Microbiol.* **6,** 666 (2015).

216. Famin, O. & Ginsburg, H. The treatment of Plasmodium falciparum -infected erythrocytes with chloroquine leads to accumulation of ferriprotoporphyrin IX bound to particular parasite proteins and to the inhibition of the parasite's 6-phosphogluconate dehydrogenase. *Parasite* **10,** 39–50 (2003).

217. Wiseman, G. M. The hemolysins of staphylococcus aureus. *Bacteriol. Rev.* **39,** 317–344 (1975).

218. Rafikova, O., Al Ghouleh, I. & Rafikov, R. Focus on early events: Pathogenesis of pulmonary arterial hypertension development. *Antioxidants Redox Signal.* **31,** 933–953 (2019).

219. Chiziane, E., Telemann, H., Krueger, M., Adler, J., Arnhold, J., Alia, A. & Flemmig, J. Free heme and amyloid-β: A fatal liaison in Alzheimer's disease. *J. Alzheimer's Dis.* **61,** 963–984 (2018).

220. Frimat, M., Tabarin, F., Dimitrov, J. D., Poitou, C., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V. & Roumenina, L. T. Complement activation by heme as a secondary hit for atypical hemolytic uremic syndrome. *Blood* **122,** 282–292 (2013).

221. Litvinov, R. I. & Weisel, J. W. Role of red blood cells in haemostasis and thrombosis. *ISBT Sci. Ser.* **12,** 176–183 (2017).

222. Wang, L., Vijayan, V., Jang, M. S., Thorenz, A., Greite, R., Rong, S., Chen, R., Shushakova, N., Tudorache, I., Derlin, K., Pradhan, P., Madyaningrana, K., Madrahimov, N., Bräsen, J. H., Lichtinghagen, R., van Kooten, C., Huber-Lang, M., Haller, H., Immenschuh, S., *et al.* Labile heme aggravates renal inflammation

and complement activation after ischemia reperfusion injury. *Front. Immunol.* **10,** 2975 (2019).

223. Kato, G. J., Steinberg, M. H. & Gladwin, M. T. Intravascular hemolysis and the pathophysiology of sickle cell disease. *J. Clin. Invest.* **127,** 750–760 (2017).

224. Ghosh, S., Adisa, O. A., Chappa, P., Tan, F., Jackson, K. A., Archer, D. R. & Ofori-Acquah, S. F. Extracellular hemin crisis triggers acute chest syndrome in sickle mice. *J. Clin. Invest.* **123,** 4809–4820 (2013).

225. Ofori-Acquah, S. F., Hazra, R., Orikogbo, O. O., Crosby, D., Flage, B., Ackah, E. B., Lenhart, D., Tan, R. J., Vitturi, D. A., Paintsil, V., Owusu-Dabo, E. & Ghosh, S. Hemopexin deficiency promotes acute kidney injury in sickle cell disease. *Blood* **135,** 1044–1048 (2020).

226. Chen, L., Zhang, X., Chen-Roetling, J. & Regan, R. F. Increased striatal injury and behavioral deficits after intracerebral hemorrhage in hemopexin knockout mice: Laboratory investigation. *J. Neurosurg.* **114,** 1159–1167 (2011).

227. Letarte, P. B., Lieberman, K., Nagatani, K., Haworth, R. A., Odell, G. B. & Duff, T. A. Hemin: levels in experimental subarachnoid hematoma and effects on dissociated vascular smooth-muscle cells. *J. Neurosurg.* **79,** 252–255 (1993).

228. Hanafy, K. A. The role of microglia and the TLR4 pathway in neuronal apoptosis and vasospasm after subarachnoid hemorrhage. *J. Neuroinflammation* **10,** 83 (2013).

229. Schipper, H. M., Liberman, A. & Stopa, E. G. Neural heme oxygenase-1 expression in idiopathic Parkinson's disease. *Exp. Neurol.* **150,** 60–68 (1998).

230. Buehler, P. W., Humar, R. & Schaer, D. J. Haptoglobin therapeutics and compartmentalization of cell-free hemoglobin toxicity. *Trends Mol. Med.* **26,** 683–697 (2020).

231. Reiter, C. D., Wang, X., Tanus-Santos, J. E., Hogg, N., Cannon, R. O., Schechter, A. N. & Gladwin, M. T. Cell-free hemoglobin limits nitric oxide bioavailability in sickle-cell disease. *Nat. Med.* **8,** 1383–1389 (2002).

232. Deuel, J. W., Vallelian, F., Schaer, C. A., Puglia, M., Buehler, P. W. & Schaer, D. J. Different target specificities of haptoglobin and hemopexin define a sequential protection system against vascular hemoglobin toxicity. *Free Radic. Biol. Med.* **89,** 931–943 (2015).

233. Graversen, J. H., Madsen, M. & Moestrup, S. K. CD163: A signal receptor scavenging haptoglobin-hemoglobin complexes from plasma. *Int. J. Biochem. Cell Biol.* **34,** 309–314 (2002).

234. Hvidberg, V., Maniecki, M. B., Jacobsen, C., Højrup, P., Møller, H. J. & Moestrup, S. K. Identification of the receptor scavenging hemopexin-heme complexes. *Blood* **106,** 2572–2579 (2005).

235. Pasternack, R. F., Gibbs, E. J., Hoeflin, E., Kosar, W. P., Kubera, G., Skowronek, C. A., Wong, N. M. & Muller-Eberhard, U. Hemin binding to serum proteins and the catalysis of interprotein transfer. *Biochemistry* **22,** 1753–1758 (1983).

236. Sen, N. F. A., Bjerrum, P. J. & Siggaard-Andersen, O. Ionic binding, net charge, and Donnan effect of human serum albumin as a function of pH. *Clin. Chem.* **39,** 48–52 (1993).

237. Jacobsen, J. & Brodersen, R. Albumin-bilirubin binding mechanism. *J. Biol.*

Chem. **258,** 6319–6326 (1983).

238. Morgan, W. T., Heng Liem, H., Sutor, R. P. & Muller-Eberhard, U. Transfer of heme from heme-albumin to hemopexin. *Biochem. Biophys. Acta - Gen. Subj.* **444,** 435–445 (1976).

239. Karnaukhova, E., Krupnikova, S. S., Rajabi, M. & Alayash, A. I. Heme binding to human alpha-1 proteinase inhibitor. *Biochim. Biophys. Acta - Gen. Subj.* **1820,** 2020–2029 (2012).

240. Kristiansson, A., Bergwik, J., Alattar, A. G., Flygare, J., Gram, M., Hansson, S. R., Olsson, M. L., Storry, J. R., Allhorn, M. & Åkerström, B. Human radical scavenger α1-microglobulin protects against hemolysis in vitro and α1-microglobulin knockout mice exhibit a macrocytic anemia phenotype. *Free Radic. Biol. Med.* in press (2020).

241. Fasano, M., Mattu, M., Coletta, M. & Ascenzi, P. The heme-iron geometry of ferrous nitrosylated heme-serum lipoproteins, hemopexin, and albumin: A comparative EPR study. *J. Inorg. Biochem.* **91,** 487–490 (2002).

242. Janciauskiene, S., Tumpara, S., Wiese, M., Wrenger, S., Vijayan, V., Gueler, F., Chen, R., Madyaningrana, K., Mahadeva, R., Welte, T., Immenschuh, S. & Chorostowska-Wynimko, J. Alpha1-antitrypsin binds hemin and prevents oxidative activation of human neutrophils: putative pathophysiological significance. *J. Leukoc. Biol.* **102,** 1127–1141 (2017).

243. Piazza, M., Damore, G., Costa, B., Gioannini, T. L., Weiss, J. P. & Peri, F. Hemin and a metabolic derivative coprohemin modulate the TLR4 pathway differently through different molecular targets. *Innate Immun.* **17,** 293–301 (2011).

244. Gay, N. J. & Keith, F. J. Drosophila Toll and IL-1 receptor. *Nature* **351,** 355–356 (1991).

245. Kawai, T. & Akira, S. TLR signaling. *Cell Death Differ.* **13,** 816–825 (2006).

246. Lu, Y.-C., Yeh, W.-C. & Ohashi, P. S. LPS/TLR4 signal transduction pathway. *Cytokine* **42,** 145–151 (2008).

247. Krishnan, J., Anwar, M. A. & Choi, S. TLR4 (Toll-Like Receptor 4). in *Encyclopedia of Signaling Molecules* (ed. Choi, S.) 1–13 (Springer New York, 2016).

248. Akira, S. & Takeda, K. Toll-like receptor signalling. *Nat. Rev. Immunol.* **4,** 499–511 (2004).

249. Nagai, Y., Akashi, S., Nagafuku, M., Ogata, M., Iwakura, Y., Akira, S., Kitamura, T., Kosugi, A., Kimoto, M. & Miyake, K. Essential role of MD-2 in LPS responsiveness and TLR4 distribution. *Nat. Immunol.* **3,** 667–672 (2002).

250. Patra, M. C., Kwon, H.-K., Batool, M. & Choi, S. Computational Insight Into the Structural Organization of Full-Length Toll-Like Receptor 4 Dimer in a Model Phospholipid Bilayer. *Front. Immunol.* **9,** 489 (2018).

251. Sudan, K., Vijayan, V., Madyaningrana, K., Gueler, F., Igarashi, K., Foresti, R., Motterlini, R. & Immenschuh, S. TLR4 activation alters labile heme levels to regulate BACH1 and heme oxygenase-1 expression in macrophages. *Free Radic. Biol. Med.* **137,** 131–142 (2019).

252. Shih, R. & Yang, C. Induction of heme oxygenase-1 attenuates expression in mouse brain endothelial cells. *J. Neuroinflammation* **7,** 86 (2010).

253. Belcher, J. D., Chen, C., Nguyen, J., Milbauer, L., Abdulla, F., Alayash, A. I., Smith, A., Nath, K. A., Hebbel, R. P. & Vercellotti, G. M. Heme triggers TLR4 signaling leading to endothelial cell activation and vaso-occlusion in murine sickle cell disease. *Blood* **123,** 377–390 (2014).

254. Nath, K. A., Belcher, J. D., Nath, M. C., Grande, J. P., Croatt, A. J., Ackerman, A. W., Katusic, Z. S. & Vercellotti, G. M. Role of TLR4 signaling in the nephrotoxicity of Heme and heme proteins. *Am. J. Physiol. - Ren. Physiol.* **314,** F906–F914 (2018).

255. Vinchi, F., De Franceschi, L., Ghigo, A., Townes, T., Cimino, J., Silengo, L., Hirsch, E., Altruda, F. & Tolosano, E. Hemopexin therapy improves cardiovascular function by preventing heme-induced endothelial toxicity in mouse models of hemolytic diseases. *Circulation* **127,** 1317–1329 (2013).

256. Belcher, J. D., Zhang, P., Nguyen, J., Kiser, Z. M., Nath, K. A., Hu, J., Trent, J. O. & Vercellotti, G. M. Identification of a heme activation site on the MD-2/TLR4 complex. *Front. Immunol.* **11,** 1370 (2020).

257. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47,** D506–D515 (2019).

258. Pence, H. E. & Williams, A. ChemSpider: An online chemical information resource. *J. Chem. Educ.* **87,** 1123–1124 (2010).

259. Kim, H. M., Park, B. S., Kim, J.-I., Kim, S. E., Lee, J., Oh, S. C., Enkhbayar, P., Matsushima, N., Lee, H., Yoo, O. J. & Lee, J.-O. Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist eritoran. *Cell* **130,** 906–917 (2007).

260. Gauto, D. F., Estrozi, L. F., Schwieters, C. D., Effantin, G., Macek, P., Sounier, R., Sivertsen, A. C., Schmidt, E., Kerfah, R., Mas, G., Colletier, J.-P., Güntert, P., Favier, A., Schoehn, G., Schanda, P. & Boisbouvier, J. Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. *Nat. Commun.* **10,** 2697 (2019).

261. Humayun, F., Domingo-Fernández, D., Paul George, A. A., Hopp, M.-T., Syllwasschy, B. F., Detzel, M. S., Hoyt, C. T., Hofmann-Apitius, M. & Imhof, D. A computational approach for mapping heme biology in the context of hemolytic disorders. *Front. Bioeng. Biotechnol.* **8,** 74 (2020).

262. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).

263. Li, T., Bonkovsky, H. L. & Guo, J. T. Structural analysis of heme proteins: Implication for design and prediction. *BMC Struct. Biol.* **11,** 13 (2011).

264. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K. & Hassabis, D. Improved protein structure prediction using potentials from deep learning. *Nature* **577,** 706–710 (2020).

265. Parganas, E., Wang, D., Stravopodis, D., Topham, D. J., Marine, J. C., Teglund, S., Vanin, E. F., Bodner, S., Colamonici, O. R., Van Deursen, J. M., Grosveld, G. & Ihle, J. N. Jak2 is essential for signaling through a variety of cytokine receptors. *Cell* **93,** 385–395 (1998).

266. McNally, R., Toms, A. V. & Eck, M. J. Crystal structure of the FERM-SH2 module

of human Jak2. *PLoS One* **11,** e0156218 (2016).

267. Lucet, I. S., Fantino, E., Styles, M., Bamert, R., Patel, O., Broughton, S. E., Walter, M., Burns, C. J., Treutlein, H., Wilks, A. F. & Rossjohn, J. The structural basis of Janus kinase 2 inhibition by a potent and specific pan-Janus kinase inhibitor. *Blood* **107,** 176–183 (2006).

268. Grisouard, J., Hao-Shen, H., Dirnhofer, S., Wagner, K.-U. & Skoda, R. C. Selective deletion of Jak2 in adult mouse hematopoietic cells leads to lethal anemia and thrombocytopenia. *Haematologica* **99,** e52–e54 (2014).

269. Grisouard, J., Li, S., Kubovcakova, L., Rao, T. N., Meyer, S. C., Lundberg, P., Hao-Shen, H., Romanet, V., Murakami, M., Radimerski, T., Dirnhofer, S. & Skoda, R. C. JAK2 exon 12 mutant mice display isolated erythrocytosis and changes in iron metabolism favoring increased erythropoiesis. *Blood* **128,** 839–851 (2016).

270. Wagner, K.-U., Krempler, A., Triplett, A. A., Qi, Y., George, N. M., Zhu, J. & Rui, H. Impaired alveologenesis and maintenance of secretory mammary epithelial cells in JAK2 conditional knockout mice. *Mol. Cell. Biol.* **24,** 5510–5520 (2004).

Acknowledgements

Publikationen

Artikel

1. Detzel, M. S.*, **Schmalohr, B. F.***, Hopp, M.-T., Paul George, A. A., Imhof, D. **(2021)** Revisiting the interaction of heme with hemopexin: Recommendations for the responsible use of an emerging drug. Biological Chemistry, in Revision

2. **Schmalohr, B. F.**, Mustafa, A-H. M., Krämer, O., Imhof, D. **(2021)** Structural insights into the interaction of heme with protein tyrosine kinase JAK2. *ChemBioChem,* im Druck

3. Hopp, M.-T., **Schmalohr, B. F.**, Kühl, T., Detzel, M., Wissbrock, A., Imhof, D. **(2020)** Heme Determination and Quantification Methods and Their Suitability for Practical Applications and Every-Day-Use. Anal. Chem., 92, 9429-9440 (**Front cover, grafisches Design**)

4. **Syllwasschy, B. F.,** Beck, M. S., Družeta, I., Hopp, M.-T., Ramoji, A., Neugebauer, U., Nozinovic, S., Menche, D., Willbold, D., Ohlenschläger, O., Kühl, T., Imhof, D. **(2020)** High-affinity binding and catalytic activity of His/Tyr-based sequences: Extending heme-regulatory motifs beyond CP. Biochim. Biophys. Acta - General Subjects, 1864, 129603

5. Paul George, A. A., Lacerda, M., **Syllwasschy, B. F.,** Hopp, M.-T., Wißbrock, A., Imhof, D. **(2020)** HeMoQuest: A webserver for qualitative prediction of transient heme binding to protein motifs. *BMC Bioinformatics*, 21, 124

6. Humayun, F., Domingo-Fernandez, D., Paul George, A. A., Hopp, M.-T., **Syllwasschy, B. F.,** Detzel, M. S., Hoyt, T. C., Hofmann-Apitius, M., Imhof, D. **(2020)** A computational approach for mapping heme biology in the context of hemolytic disorders. *Front. Bioeng. Biotechnol.*, 8, 74

7. Brewitz, H. H., Goradia, N., Schubert, E., Galler, K., Kühl, T., **Syllwasschy, B. F.,** Popp, J., Neugebauer, U., Hagelueken, G., Schiemann, O., Ohlenschläger, O., Imhof, D. **(2016)** Heme interacts with histidine- and tyrosine-based protein motifs and inhibits enzymatic activity of chloramphenicol acetyltransferase from Escherichia coli. Biochim. Biophys. Acta - General Subjects, 1860, 1343-1353

Patente

1. Patent: U.S. 63/056351 (vorläufig): Imhof, D., Hamza, I., Paul George, A. A., **Syllwasschy, B. F.**, Wißbrock, A. (**2020**) Heme-Binding Peptides and Methods of Using the Same, Patent angemeldet

Vorträge

1. **Syllwasschy, B. F.,** Hopp, M.-T., Paul George, A. A., Wißbrock, A., Imhof, D. (**März 2019**) His/Tyr-based peptide motifs to study regulatory heme binding. 14. Deutsches Peptidsymposium (GPS), Köln, Deutschland

Poster

1. Detzel, M. S., **Schmalohr, B. F.,** Steinbock, F., Hopp, M.-T., Paul George, A. A., Imhof, D. (**Januar 2021**) Revisiting the interaction of heme with hemopexin. Advances in Chemical Biology, Virtuelle Konferenz

2. Kühl, T., **Syllwasschy, B. F.,** Beck, M. S., Družeta, I., Hopp, M.-T, Kühl, T., Imhof, D. (**September 2019**) His-/Tyr-based heme-binding peptides: From heme binding to catalytic function. Jahrestagung der Deutschen Pharmazeutischen Gesellschaft (DPhG), Heidelberg, Deutschland

2. **Syllwasschy, B. F.,** Hopp, M.-T., Paul George, A. A., Wißbrock, A., Imhof, D. (**März 2019**) His/Tyr-based peptide motifs to study regulatory heme binding. 14. Deutsches Peptidsymposium (GPS), Köln, Deutschland

3. **Syllwasschy, B. F.,** Hopp, M.-T., Paul George, A. A., Wißbrock, A., Imhof, D. (**August 2018**) Heme binding beyond Cys/Pro motifs? A Systematic Analysis of His/Tyr-Based Motifs in Heme Binding Peptides and Proteins. 35. Europäisches Peptidsymposium (EPS), Dublin, Irland

4. **Syllwasschy, B. F.,** Wißbrock, A., Brewitz, H. H., Schubert, E., Kühl, T., Ohlenschläger, O., Schiemann, O., Hagelüken, G., Imhof, D. (**Oktober 2017**) Investigation heme peptide complexes as models to explore heme regulated proteins. SFB 813 - Internationales Symposium 2017, Bonn, Deutschland

5. Brewitz, H. H., Wißbrock, A., Goradia, N., **Syllwasschy, B. F.,** Kühl, T., Ohlenschläger, O., Imhof, D. (**März 2017**) Binding studies of heme with peptides reveal sequence requirements for heme association and enable prediction of new heme-regulated proteins. 13. Deutsches Peptidsymposium (GPS), Erlangen, Deutschland