Bioscience Reports (2019) **39** BSR20181940
https://doi.org/10.1042/BSR20181940

PORTLAND PRESS

Review Article

The molecular basis of transient heme-protein interactions: analysis, concept and implementation

Amelie Wißbrock, Ajay Abisheck Paul George, Hans Henning Brewitz, Toni Kühl and Diana Imhof

Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute, University of Bonn, An der Immenburg 4, Bonn, Germany

Correspondence: Diana Imhof (dimhof@uni-bonn.de)

Deviant levels of available heme and related molecules can result from pathological situations such as impaired heme biosynthesis or increased hemolysis as a consequence of vascular trauma or bacterial infections. Heme-related biological processes are affected by these situations, and it is essential to fully understand the underlying mechanisms. While heme has long been known as an important prosthetic group of various proteins, its function as a regulatory and signaling molecule is poorly understood. Diseases such as porphyria are caused by impaired heme metabolism, and heme itself might be used as a drug in order to downregulate its own biosynthesis. In addition, heme-driven side effects and symptoms emerging from heme-related pathological conditions are not fully comprehended and thus impede adequate medical treatment. Several heme-regulated proteins have been identified in the past decades, however, the molecular basis of transient heme-protein interactions remains to be explored. Herein, we summarize the results of an in-depth analysis of heme binding to proteins, which revealed specific binding modes and affinities depending on the amino acid sequence. Evaluating the binding behavior of a plethora of heme-peptide complexes resulted in the implementation of a prediction tool (SeqD-HBM) for heme-binding motifs, which eventually led and will perspectively lead to the identification and verification of so far unknown heme-regulated proteins. This systematic approach resulted in a broader picture of the alternative functions of heme as a regulator of proteins. However, knowledge on heme regulation of proteins is still a bottomless barrel that leaves much scope for future research and development.

Introduction

Heme is a valued, versatile, and vital molecule [1–3]. As a prosthetic group of hemoglobin, heme was initially described by Fritz Ludwig Hünefeld in the 1840s [4]. Two Nobel Prizes awarded to Hans Fischer for the synthesis of hemin in 1930, and to Max Perutz and John Kendrew in 1962, who explored the structure of hemoglobin and myoglobin, honored the eminent role of the molecule which was already recognized in the early 19th century (*cf.* www.Nobelprize.org). However, even though the nature of heme and related physiological processes had been investigated for many decades, another substantial function was only identified in the 1990s: Heme may act as an effector and signaling molecule [5]. Part of this function includes transient heme-protein interactions as found for the human Aminolevulinic acid synthase (ALAS) by Lathrop and Timko in 1993 [6]. Lathrop and Timko are the eponyms of the nowadays broadly used term 'heme-regulatory motif' (HRM) that originally described a distinct conserved motif involved in heme-mediated regulation of ALAS [6]. Over the years, the term HRM was refined and meanwhile describes a short amino acid sequence that includes a heme-coordination site and is located on the protein surface [7]. Heme binding to

Received: 17 October 2018
Revised: 18 December 2018
Accepted: 02 January 2019

Accepted Manuscript Online: 08 January 2019
Version of Record published: 30 January 2019

Bioscience Reports (2019) **39** BSR20181940
https://doi.org/10.1042/BSR20181940

such motifs may alter protein stability and/or function or it can result in the formation of a catalytically active heme-peptide/protein complex [5,8]. Motifs including a cysteine-proline (CP) dipeptide are specified by the term 'CP motif' [7]. The latter one is the most prominent representative amongst HRMs [7,9–14]. Today, after more than two decades of research, CP motifs are still the best explored HRMs, nevertheless, there is no doubt about HRMs occurring in much more versatile ways considering also other coordinating amino acids such as histidine- and tyrosine-based motifs. If no functional impact occurs upon heme binding to a protein, the term 'heme-binding motif' (HBM) is used to describe a protein sequence stretch that interacts with heme. An intriguing question that arises at this point is, what are the specific requirements for transient heme-protein interactions?

Given the chemical nature of heme, one can easily imagine that a heme-protein interaction occurs at various molecular levels. A transient interaction requires fast heme association and disassociation in order to allow for a situation-dependent response. For regulatory heme, a coordinative bond of the central iron ion to a heteroatom-containing amino acid side chain is observed in the first place. The most prominent heme-coordinating residues are cysteine, histidine, and tyrosine, while methionine and lysine are less frequently found [10]. In addition, hydrophobic interactions and π-π stacking conveyed by the porphyrin ring system as well as electrostatic interactions and hydrogen bonding via the propionate side chains contribute to heme binding and influence binding mode and affinity [10,15]. Therefore, not only the coordinative residue but also the surrounding amino acids are responsible for the decision if and how heme interacts with a specific protein [10,15]. Following the aforementioned initial studies, several heme-regulated proteins have been identified in the last 25 years [5,8]. These heme-regulated proteins take part in diverse biological processes including transcription and translation (e.g. DGCR8 [16], Rev-Erbβ [17]), ion channel modulation (e.g. hSlo1 [18]), circadian rhythm (Per2 [19]), and cell-cycle regulation (p53 [20]) [5,8]. Moreover, the function of various extracellular proteins such as complement factors (e.g. C1q [21], C3 [22]), and coagulation factors (e.g. FVIII [23]) is altered by transient heme binding. Heme-binding to eminent medical targets as, for instance, the cystathionine-β-synthase [24] or the amyloid β (Aβ) peptide [25], known for its crucial role in Alzheimer's disease, promoted further interest in potential (patho-)physiological implications of transient heme-protein interplays. Heme interaction with proteins may be of particular interest in the case of amplified hemolysis, for example, as a consequence of vascular injury or action of bacterial hemolysins, resulting in ominously augmented concentrations of biologically available heme. Extending the observations of a heme-mediated change of a protein's activity and/or stability, a peroxidase-like activity [25] was shown for the Aβ-heme complex raising the question of heme-mediated physiological responses depending on the cellular milieu in specific events, e.g. oxidative stress. On the other hand, heme-mediated protein regulation can be used for therapy as in the case of ALAS mentioned above [26]. In an acute attack, porphyria patients may receive heme preparations that are thought to downregulate ALAS activity in the liver and thereby decrease toxic heme intermediates [26]. Due to the fundamental role of heme, as an oxygen-binding molecule in e.g. hemoglobin and its omnipresence in the blood, it is inevitably necessary to understand basic heme-associated processes as well as heme-mediated protein regulation in order to take appropriate measures as required in the case of risk-bearing pathophysiological conditions.

Even though the awareness of alternative roles of heme increased over the years, a systematic approach to explore the molecular basis of transient heme-protein interactions was missing until 2010, when we started the search for specific interaction patterns, binding motifs, and structural insights concerning the transient binding of heme to peptides and proteins.

Sequence criteria for heme binding identified by a combinatorial peptide library screening

Since short protein-derived sequences (∼9 amino acids) were shown to be suitable to study heme-binding behavior [9,12,27,28], our initial studies included the construction of a combinatorial nonapeptide library based on histidine, tyrosine, and cysteine as heme axial ligand (at position P^0) as these are the most striking heme-coordinating residues (Figure 1A) [29]. Besides the coordinating amino acid all positions were randomized, and the library was constructed as $X_4(C/H/Y)^0X_4$ (X: all amino acids except Cys and Met, but including Nle). While methionine was required for technical reasons (special peptide elimination procedure) and thus could not be used at the randomized positions [30], additional cysteines were excluded to avoid intramolecular disulfide formation. After library synthesis, the peptide-bound resin beads were incubated with varying concentrations of heme (0.01 to 100 nM) [29]. Upon incubation a yellow-green color occurred in the case of heme binding and allowed to manually pick the respective beads using a stereomicroscope (Figure 1A).

Sequence elucidation by PED-MALDI-TOF mass spectrometry [30] as well as on-bead automated Edman degradation revealed distinct sequence characteristics [29]. Evaluation of the obtained peptides revealed a predominance

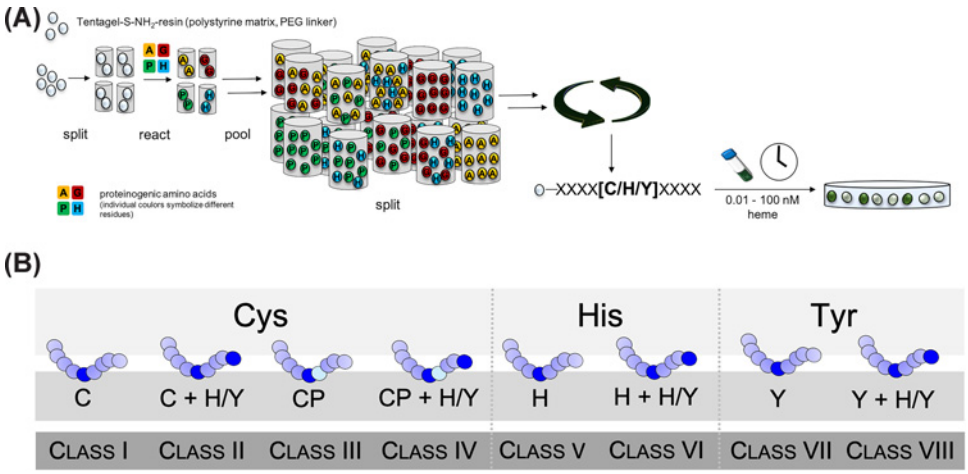

Figure 1. Investigation and classification of heme-binding sequences by means of a combinatorial peptide library

(**A**) A combinatorial peptide library $X_4(C/H/Y)^0X_4$ was constructed in order to investigate sequence specificities of heme-binding peptides. As a result of the screening, a classification system for HRMs/HBMs (**B**) was compiled for cysteine-, histidine-, and tyrosine-based motifs including extra classes for CP motifs [13,29,33,34].

of histidine and tyrosine residues (∼40% each) as heme axial ligands over cysteines (∼20%). In addition, corresponding sequence specificities at the termini were identified for all peptides: primarily polar residues as e.g. E, D, Q, N, R, K, H, Y emerged and also, to a lesser extent, hydrophobic amino acids like L, V, F, and Y. Such residues facilitate the interaction with the functional groups of the porphyrin ring. Interestingly, the appearance of additional coordination sites (His/Tyr) was observed in more than 50% of the peptide hits.

The identified hit sequences were analyzed to derive consensus sequences for the different classes of HRMs. To identify potential heme-regulated proteins, a database screening with the consensus sequences was subsequently performed by means of the ScanProsite tool [31] (ExPASy Proteomics server). Several search runs revealed potential heme-binding bacterial as well as human proteins suggesting that the underlying molecular concept is evolutionarily conserved [5,8,32]. Moreover, already published HRMs were evaluated for similarities and sequence characteristics, too. Based on these findings, further fine-tuned consensus sequences were derived and screened as described before. Among the proposed heme-binding sequences, the human dipeptidyl-peptidase 8 (DPP8) was identified as an interesting target protein that was later shown to be regulated by heme [13,29].

Classifying heme-binding motifs

The findings of the peptide library screening inspired us to develop a classification system for HBMs according to the axial ligand (H, Y, C) (Figure 1B) [33,34]. The observation that more than 50% of the library-derived heme-binding peptides exhibited ancillary coordination sites in close proximity [29], justified a further division of the three classes considering the presence or absence of additional potential coordination sites beyond position P^0. Thus, a total number of eight classes of HBMs was established (Figure 1B). To allow for profound analysis, suitable peptide representatives of each class as well as proteins carrying the corresponding motif were selected to study their heme-binding behavior with the help of sophisticated spectroscopic methods [13,29,33,34]. The subsequent investigations were based on peptide sequences as model compounds bearing in mind that the gained knowledge was to be transferred to the protein level in following studies.

In-depth analysis of heme-binding motifs by different spectroscopic methods

The spectroscopic methods applied within our studies complemented each other and allowed to draw a comprehensive picture of the heme-peptide interaction taking place (Figure 2). The fact that relatively large substance quantities are required is common to all of these methods, restricting the practical applicability for molecules with limited availability (e.g. proteins). In the case of nonapeptides, the amount available is usually not the limiting factor, but physicochemical properties affecting solubility, for example, may prevent the implementation of the individual method. Hereafter, a short, simplified insight into the individual spectroscopic methods is given (Figure 2).

Bioscience Reports (2019) **39** BSR20181940
https://doi.org/10.1042/BSR20181940

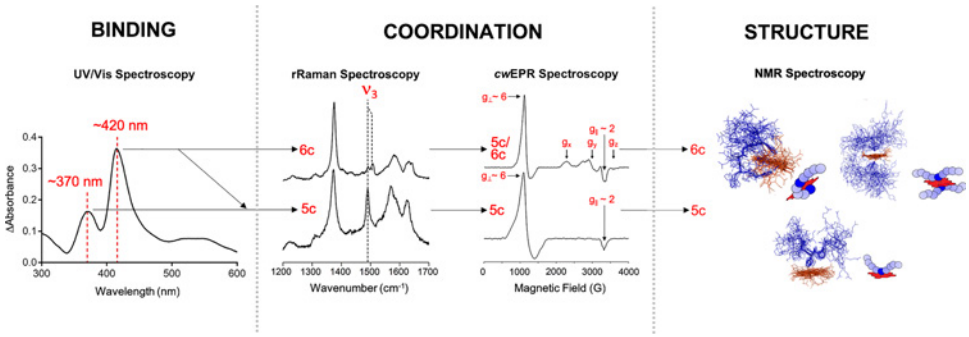

Figure 2. Investigating heme binding to peptides or proteins by various spectroscopic methods

Complex formation can be detected by UV/Vis spectroscopy, in particular by a shift of the heme-characteristic Soret-band. Upon binding the heme-iron coordination state, i.e. the occurrence of a penta- and/or hexa-coordinated complex, can be investigated by rRaman and *cw*EPR spectroscopy, focusing amongst others on the ν_3 band (rRaman) and the signals around $g \sim 6$ and $g \sim 2$ (*cw*EPR). The topology and structure of the formed complexes can be clarified by applying 2D- or 3D-NMR spectroscopy [32,33].

Due to the number of conjugated double bonds found in the porphyrin scaffold, heme shows a characteristic absorbance spectrum [35,36]. Especially worth mentioning is the B-band, called Soret-band after its discoverer Jacques-Louis Soret [37]. This band is visible at ∼400 nm. Besides the Soret-band, there are less pronounced Q-bands in the range of ∼450-700 nm [38]. The exact position of the bands depends on the oxidation and spin state of the heme-iron ion as well as the immediate surrounding of the heme molecule [39,40]. As described in the next section, a shift of the Soret-band to ∼370 nm seems to correlate primarily with a penta-coordinated complex, while a shift to ∼420 nm is found for penta- and hexa-coordinated complexes [33,34]. Therefore, the formation of heme-peptide/protein complexes is detectable in the UV/Vis spectrum, in particular with regard to a shift of the Soret-band as is commonly observed for heme binding to amino acid sequences. Moreover, titrations with different concentrations of either interaction partner enable the determination of binding constants such as the dissociation constant K_D.

Once a heme-peptide/protein interaction is confirmed by UV/Vis spectroscopy, the heme-iron coordination state (complex geometry), i.e. a penta- or hexa-coordinated iron ion, is of great interest. The iron ion is bound to four nitrogen atoms of the planar porphyrin ring system while there are two open positions remaining that allow for one or two additional coordinative bonds. These positions can be occupied by ligands possessing sulfur, oxygen or nitrogen atoms such as trifunctional amino acids. Coordination to these ligands will result in the formation of a penta- or hexa-coordinated heme-complex. It is worth mentioning that in hemoproteins the sixth ligand can be a solvent or gas molecule such as water, oxygen, carbon monoxide or nitrogen monoxide, which is usually classified as a penta-coordination with respect to the protein ligand [41]. Methods that allow to determine the coordination state of the heme iron are *resonance* Raman (rRaman) spectroscopy and *continuous wave* electron spin resonance (*cw*EPR) spectroscopy (Figure 2) [42,43]. rRaman spectroscopy is based on laser-induced vibrations (370-430 nm) characteristic for the heme molecule [42,44]. The so-called ν_3 band is of special interest regarding the iron coordination state, since it shifts according to the coordination occurring during complex formation [44]. The ν_3 band emerges around ∼1491 cm^{-1} in the case of hemin only (chloride acts as ligand) and a penta-coordinated heme-peptide/protein complex, while in case of a hexa-coordinated complex, the band appears around 1505 cm^{-1} (Figure 2) [44]. Moreover, mixtures of penta- and hexa-coordinated complexes can be detected as a double band [42]. Additional bands such as the ν_7 band can give further insight into the complex geometry present [45].

*cw*EPR spectroscopy is based on the spin state of the iron ion [43]. Simplified, free hemin as well as a penta-coordinated heme complex exhibit a high-spin state (S = 5/2), while a hexa-coordinated complex results in a low-spin state (S = 6/2) [43]. The respective signals appear at $g \sim 6$ and $g_{||} \sim 2$ in the case of penta-coordination, whereas a hexa-coordination leads to the occurrence of three signals with g-values (g_x, g_y, g_z) in the range of $g \sim 1.5$ to $g \sim 3$ (Figure 2) [34,43]. In-depth information on the structure and topology of heme-peptide/protein complexes can be obtained by applying 2D/3D-NMR spectroscopy (Figure 2) [13]. Changes upon heme complexation are identified by comparing the complex structure to the unbound peptide/protein structure, i.e. evaluation of the chemical shifts of residues before and after heme incubation is necessary [13,46]. NMR spectroscopy of longer sequences is extremely time consuming and challenging primarily due to high structural flexibility [47,48]. Depending on the sequence composition (^{1}H, ^{13}C) HSQC (heteronuclear single quantum coherence) spectra that are based on the natural

abundance of ^{13}C are used among other experiments. NMR structural analysis of large peptides and proteins is also possible, yet recombinant expression of ^{13}C and/or ^{15}N labeled molecules is required, which is usually achieved by adding e.g. $^{15}NH_4Cl$ and $^{13}C_6$-glucose to the respective growth media. The fact that the paramagnetic Fe^{3+} interferes with the surrounding amino acids leads to opposing effects on the intensities and to a broadening of the resonances. Therefore, in many studies other metal porphyrins, e.g. Ga^{III}-PPIX, were used instead, which are supposed to interact with peptides/proteins in a similar manner as heme [49–51].

Furthermore, it is possible to use the knowledge obtained regarding sequence requirements for transient heme-peptide/protein interactions to identify HBMs within known heme-binding/heme-regulated proteins. Therefore, we developed a computational tool that allows for the evaluation of potential motifs upon input of the protein sequence.

Sequence-dependent characteristics of heme-binding motifs

Since the most prominent motifs so far are the aforementioned CP-motifs, our initial studies focused on cysteine-based sequences representing HRM-classes I-IV (Figure 1) [13,33]. Li *et al.* showed that the sole occurrence of a heme-coordination site as well as the presence of a CP-motif does not necessarily result in an interaction with heme [10]. This observation was confirmed within all of our studies, since nonapeptides consisting of a CP-dipeptide surrounded by four respectively three alanine residues did not interact with heme [13,29,33,34]. These findings support the original idea that specific sequence features are required for heme binding. For the cysteine-based motifs (class I-IV) high to moderate binding affinities with K_D values ranging from 0.40 ± 0.19 μM to 6.36 ± 2.61 μM were determined by UV/Vis spectroscopy [13,29,33]. The cysteine-based sequences with additional coordinating residues (tyrosine, histidine) generally appeared to lead to higher binding affinities [13,33]. The affinities determined seem plausible because a transient interaction requires a fast and uncomplicated association and dissociation of the respective molecules and, in addition, heme is a rather small interaction partner [25,47,52]. The K_D values obtained are also supported by examples of regulatory heme binding described by other groups [52,53].

Furthermore, evaluation of the observed UV spectra revealed the occurrence of four different kinds of spectra which we categorized as UV-groups I-IV [54]. While a shift of the Soret-band to ∼370 nm primarily represented a penta-coordinated complex as was predominantly found for CP-peptides, a shift to ∼420-430 nm cannot be uniquely assigned to a distinct coordination state although spectra with both maxima frequently exhibited a mixture of penta- and hexa-coordinated complexes as revealed by rRaman and *cw*EPR spectroscopy (Figure 2) [13,33]. In contrast to the CP-peptides, for cysteine-based motifs (without proline) all forms of coordination states were observed [33].

Structural investigation of selected representatives of classes I-IV by NMR spectroscopy gave insight into the particular role of the proline residue within the CP-motif [13,33]. On the peptide level there was a clear difference between C- and CP-based peptides [33]. The proline residue seemed to reinforce a more defined backbone structure of the free peptide compared to the rather flexible structure of the cysteine-based peptide [33]. Application of heme did not lead to an increase of rigidity in the case of the cysteine-based motifs, however, CP-based peptides showed increased backbone rigidity upon heme binding, in particular in close proximity of the CP-motif. It was found that the proline residue confers a distinct conformation to the subsequent backbone, which - as a consequence - is directed away from the porphyrin ring [33]. NMR spectroscopy also revealed penta-coordinated heme complexes for CP-motifs (class III and IV) independently of the presence or absence of an additional possible heme coordination site [33]. Analysis of the cysteine-based motifs without a proline displayed hexa-coordination for both classes (I and II) (Figures 1 and 3). On the one hand, the cysteine-based peptide with no additional coordination site revealed binding of two peptide molecules to one heme molecule in a 'sandwich-like' structure (class I) (Figure 3). On the other hand, for class II (e.g. HXXXC) it was shown that a spacer length of three amino acids is required to obtain a 'loop-like' (respectively 'clamp-like') hexa-coordinated complex with one coordinating residue being the central cysteine and the other one being a distal histidine residue [33] (Figure 3).

To complete the picture of heme-binding sequences, histidine- and tyrosine-based motifs representing classes V-VIII (Figure 1) were investigated in the same manner as described above [34]. In contrast to the earlier findings for cysteine-based motifs, several sequences did not interact with heme or did not show saturation upon increasing heme concentrations, thereby hampering the determination of K_D values [34]. The K_D values determined ranged from 0.24 ± 0.17 μM to 6.25 ± 1.44 μM [34] again revealing high to moderate binding affinities as expected. In-depth analysis of selected sequences revealed that histidine-based motifs usually formed mixed or hexa-coordinated complexes with e.g. an HXXXH-motif exhibiting a loop-like structure, while tyrosine-based motifs predominantly occurred in a penta-coordinated fashion [34]. In the case of an additional coordination site in tyrosine-based peptides no loop

Bioscience Reports (2019) **39** BSR20181940
https://doi.org/10.1042/BSR20181940

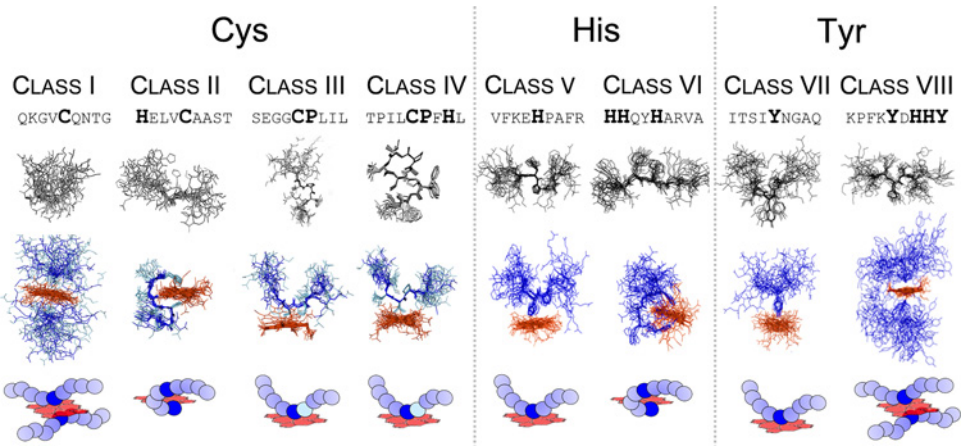

Figure 3. Structural elucidation of heme-binding peptides (classes I–VIII) using 2D-NMR spectroscopy

Different binding modes occurred depending on the peptide sequence composition and the formation of penta-coordinated (III, IV, V, and VII) and hexa-coordinated complexes (I, II, VI, and VIII). The latter ones emerged in different forms: a sandwich-like complex including two peptides interacting with one heme molecule (I and VIII) and a loop/clamp-like complex for peptides that exhibit additional coordination sites (II and VI) [32,33].

formation was found [34]. In general, no increased backbone rigidity as found for the CP-motifs was observed within these studies [34].

Comparing the amino acid composition of the heme-binding sequences revealed a crucial role of the net charge of the nonapeptides. While a negative net charge appeared to inhibit heme interaction, a positive net charge was usually accompanied by a comparably high binding affinity [34]. All the information gained from UV/Vis, rRaman, *cw*EPR, and 2D-NMR spectroscopy revealed insight into the specific characteristics of heme binding to peptides/proteins on the level of primary sequences and secondary structures. In the presented study, peptides served as model system to examine a broad range of primary sequence motifs of heme-binding peptides and proteins. It is worth noting that various other studies have applied peptide-based approaches to investigate heme binding and hemoproteins. These studies addressed functional, structural, stability, and specificity issues of the respective heme complexes [28]. The sequences and secondary structural elements of the examined peptides vary broadly. Whereas some studies use protein-derived sequences, others utilize specifically designed peptides which exhibit desired functional and structural properties. In contrast to the study using nonapeptides, specially designed peptides are often characterized by secondary structures that facilitate distinct functions, e.g. intended heme binding [28,55–59]. Among these are heme-binding multi-stranded β-sheet peptides [55], β-hairpin conformation [56], heme-Cage β-Sheet miniproteins [57], as well as helical sequence stretches [59]. Besides basic research these heme-binding peptides and miniproteins are intended for industrial and biomedical applications [58].

Subsequently the established consensus sequences mentioned above were screened against the protein data base ScanProsite tool [31] (ExPASy Proteomics server) aiming to identify so far unknown heme-regulated proteins. The hits obtained were further assessed taking into account the accessibility of the suggested HBM/HRM for heme binding, the protein structure if available, and the possibility to experimentally test the impact of heme on the protein activity. Several potential heme-regulated proteins were identified using this approach. Heme binding to these proteins was verified by spectroscopic methods and the functional impact of heme was shown *in vitro* [13,34,54]. This approach was successful for bacterial proteins such as FeoB [54], chloramphenicol-acetyltransferase (Cat) [32], and hemolysin C [60] as well as human proteins as the aforementioned dipeptidylpeptidase 8 [13]. To summarize the knowledge obtained regarding distinct sequence features of HBMs/HRMs, a procedure that enables evaluation of a protein sequence to comprise HBMs in a stepwise manner was generated (see below). A first assessment of the sequence can be drawn on the basis of the primary sequence. Additional experiments such as UV/Vis spectroscopy will then facilitate pre-evaluation of the heme-binding mode based on general structural features of the complex formed (pre-selection). In order to facilitate the HBM/HRM evaluation process for other users, we recently developed an algorithm termed *SeqD-HBM*. The basics of the *SeqD-HBM* are explained below.

Evaluating heme-binding capacity of protein sequences using *SeqD-HBM*

A handful of computational tools are available to predict ligand binding to proteins. Some of these have a broader scope of being able to predict the binding of multiple ligands to proteins e.g. *TargetS* [61], whereas other tools are specific for the prediction of heme binding (e.g. *HemeBIND+* [62] and *HemeBIND* [63]). While these tools use complex algorithms and base their predictions on extracting several novel features (e.g. Depth index, Protrusion index, Surface complementarity etc.), the nature of binding usually described is the strong irreversible binding of heme (mostly within a binding pocket) rather than a transient interaction. Moreover, most of these tools rely heavily on the availability of structural information as a basis for their predictions [63]. With this in mind, we introduce a novel tool named "*SeqD-HBM"* for sequence-based identification of HBM using the protein sequence as its primary input (Supplementary Material). The program processes the input sequence through a systematic stepwise validation extracting at each step relevant features from the sequence to produce a tabulated list of the possible heme-coordination sites available for the given sequence. Besides reporting the potential heme-coordination sites (Cys, His, Tyr), the program outputs the 9mer motifs associated with the coordination site and the net charge. Finally, in a column named "comment" useful hints regarding the predicted 9mer motif, such as identification of a CP motif, are provided to the user. This comprehensive output further guides the user to make informed judgments on the nature of heme interaction with the protein of interest.

SeqD-HBM in its current stand-alone form has two distinct modes of operation. The *default mode* assumes that the user has no information on the structure related to the input sequence. This consequently means that there is no possibility to determine if a predicted coordination site is "surface-exposed" or "buried". Prediction of a buried residue as a potential coordination site would be a false positive, defeating the purpose of this tool. To overcome this roadblock, we pass the input sequence through a sequence-based solvent accessibility meta-predictor namely *WESA* (Weighted Ensemble Solvent Accessibility) [64,65]. *WESA* determines the solvent accessibility of each residue in a sequence using an ensemble of five methods: Bayesian statistics, multiple linear regression, decision tree, neural network, and support vector machine. A weighted sum of individual predictions determines the final prediction. This deems *WESA* to be a robust and reliable tool to distinguish between the buried and exposed states of residues and has a published accuracy of 80%. *WESA* is invoked from within *SeqD-HBM* and only those coordination sites that are predicted to be "exposed" are considered for the final tabulated output of the *SeqD-HBM* program.

The second mode of operation called the *structure mode* assumes that the user is aware of the structure or is in possession of the structure data of a sequence that is passed as input to *SeqD-HBM*. In this case, *SeqD-HBM* does not invoke WESA and the HBM validation checks are done on every possible coordination site available in the sequence. The user is expected to use the known structural information to manually filter out false positives (i.e. ignore a coordination site prediction if it is known from the structure that the site is a buried residue) from the *SeqD-HBM* prediction. The operation of the application is presented as a flowchart in Figure 4. The implementation and testing details of *SeqD-HBM* are discussed in the Supplementary Material.

Conclusion

To gain a deeper insight into the molecular basis of transient heme binding, nonapeptides were used as models in order to allow for a global investigation of heme-binding characteristics. More than 200 heme-peptide complexes based on cysteine, histidine, or tyrosine as heme axial ligand have been examined so far using UV/Vis spectroscopy and, in part, methods such as rRaman, *cw*EPR, and 2D-NMR spectroscopy. Evaluation of the data obtained revealed specific sequence features such as a positive net charge of the heme-binding sequence or the existence of hydrophobic residues which have a positive effect on the heme-peptide/protein interplay. Depending on the sequence composition, different binding modes have been observed, e.g. penta- vs. hexa-coordination or mixtures thereof. The knowledge derived from the detailed analysis may first be used to predict heme binding to proteins based on consensus sequences and respective data base searches. Prediction and verification of unknown potentially heme-binding proteins based on such a consensus sequences search has been successful in several cases, i.e. bacterial FeoB, and HlyC. Second, it may be utilized to assess the heme-binding capacity of proteins which were shown to bind heme and to identify HBMs in such sequences. To make the evaluation of heme-binding sequences available to the public, our knowledge was incorporated into the program '*SeqD-HBM*' for the determination of HBMs in proteins. The software evaluates the motifs contained in a protein on the basis of the primary sequence and, if possible, takes structural features into account. We expect that such a tool will be useful to decipher molecular details on heme-binding/regulated proteins and in this way support basic research concerning the previously mentioned heme-related pathological scenarios.

Bioscience Reports (2019) **39** BSR20181940
https://doi.org/10.1042/BSR20181940

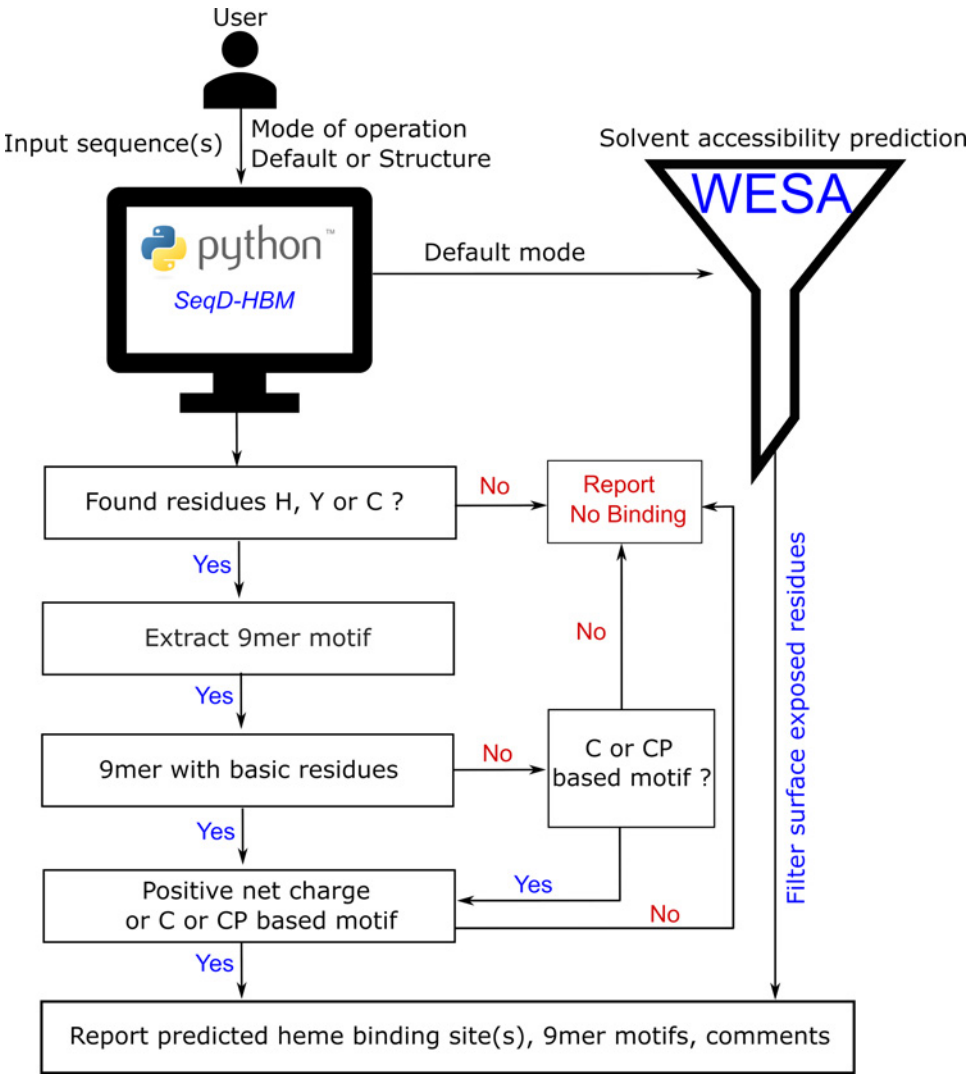

Figure 4. Computational prediction of heme-binding protein sequences

Computational evaluation of the heme-binding potential of various peptide/protein sequences based on the knowledge gained from in-depth spectroscopic studies on heme-peptide complexes.

Acknowledgements

We acknowledge Oliver Ohlenschläger (FLI Jena) for the performance of NMR structure analysis and his useful suggestions and scientific discussions. The authors also thank Jürgen Popp (IPHT Jena) and Ute Neugebauer (University Hospital Jena) for support with rRaman spectroscopy and Olaf Schiemann and Gregor Hagelüken (University of Bonn) for support with *cw*EPR spectroscopy.

Funding

This work was financially supported by the Deutsche Forschungsgemeinschaft (DFG) within [FOR1738 and SFB 813 (to D.I.)].

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Author Contribution

D.I. designed and directed the project. A.A.P.G. developed the algorithm of SeqD-HBM. All authors contributed to the manuscript.

Abbreviations

ALAS, aminolevulinic acid synthase; Aβ, amyloid β; *cw*EPR, continuous-wave electron spin resonance spectroscopy; CP, cysteine-proline; HBM, heme-binding motif; HBP, heme-binding protein; HRM, heme-regulatory motif; MALDI-TOF, matrix-assisted laser desorption/ionization time of flight; PED, partial Edman degradation; rRaman spectroscopy, resonance Raman spectroscopy; SeqD-HBM, sequence-based detection of heme binding motifs; WESA, weighted ensemble solvent accessibility.

References

1 Mense, S.M. and Zhang, L. (2006) Heme: a versatile signaling molecule controlling the activities of diverse regulators ranging from transcription factors to MAP kinases. *Cell Res.* **16**, 681–692, https://doi.org/10.1038/sj.cr.7310086
2 Severance, S. and Hamza, I. (2009) Trafficking of heme and porphyrins in metazoa. *Chem. Rev.* **109**, 4596–4616, https://doi.org/10.1021/cr9001116
3 Poulos, T.L. (2014) Heme enzyme structure and function. *Chem. Rev.* **114**, 3919–3962, https://doi.org/10.1021/cr400415k
4 Hünefeld, F.L. (1840) *Der Chemismus in der thierischen Organization*, F. A. Brockhaus, Leipzig
5 Zhang, L. (2011) *Heme Biology: The Secret Life of Heme in Regulating Diverse Biological Processes*, World Sci., Singapore; Hackensack, NJ
6 Lathrop, J.T. and Timko, M.P. (1993) Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. *Science* **259**, 522–525, https://doi.org/10.1126/science.8424176
7 Shimizu, T. (2012) Binding of cysteine thiolate to the Fe(III) heme complex is critical for the function of heme sensor proteins. *J. Inorg. Biochem.* **108**, 171–177, https://doi.org/10.1016/j.jinorgbio.2011.08.018
8 Kühl, T. and Imhof, D. (2014) Regulatory FeII/III heme: the reconstruction of a molecule's biography. *Chem. Bio. Chem.* **15**, 2024–2035, https://doi.org/10.1002/cbic.201402218
9 Zhang, L. and Guarente, L. (1995) Heme binds to a short sequence that serves a regulatory function in diverse proteins. *EMBO J.* **14**, 313–320, https://doi.org/10.1002/j.1460-2075.1995.tb07005.x
10 Li, T., Bonkovsky, H.L. and Guo, J. (2011) Structural analysis of heme proteins: implications for design and prediction. *BMC Struct. Biol.* **11**, 13, https://doi.org/10.1186/1472-6807-11-13
11 Hou, S., Reynolds, M.F., Horrigan, F.T., Heinemann, S.H. and Hoshi, T. (2006) Reversible binding of heme to proteins in cellular signal transduction. *Acc. Chem. Res.*, https://doi.org/10.1021/ar040020w
12 Igarashi, J., Murase, M., Iizuka, A., Pichierri, F., Martinkova, M. and Shimizu, T. (2008) 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, https://doi.org/10.1074/jbc.M801400200
13 Kühl, T., Wißbrock, A., Goradia, N., Sahoo, N., Galler, K., Neugebauer, U. et al. (2013) Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins. *ACS Chem. Biol.* **8**, 1785–1793, https://doi.org/10.1021/cb400317x
14 Westberg, J.A., Jiang, J. and Andersson, L.C. (2011) Stanniocalcin 1 binds hemin through a partially conserved heme regulatory motif. *Biochem. Biophys. Res. Commun.* **409**, 266–269, https://doi.org/10.1016/j.bbrc.2011.05.002
15 Schneider, S., Marles-Wright, J., Sharp, K.H. and Paoli, M. (2007) Diversity and conservation of interactions for binding heme in b-type heme proteins. *Nat. Prod. Rep.* **24**, 621–630, https://doi.org/10.1039/b604186h
16 Faller, M., Matsunaga, M., Yin, S., Loo, J.A. and Guo, F. (2007) Heme is involved in microRNA processing. *Nat. Struct. Mol. Biol.* **14**, 23–29, https://doi.org/10.1038/nsmb1182
17 Raghuram, S., Stayrook, K.R., Huang, P., Rogers, P.M., Nosie, A.K., McClure, D.B. et al. (2007) Identification of heme as the ligand for the orphan nuclear receptors REV-ERBα and REV-ERBβ. *Nat. Struct. Mol. Biol.* **14**, 1207–1213, https://doi.org/10.1038/nsmb1344
18 Tang, X.D., Xu, R., Reynolds, M.F., Garcia, M.L., Heinemann, S.H. and Hoshi, T. (2003) Haem can bind to and inhibit mammalian calcium-dependent Slo1 BK channels. *Nature* **425**, 531–535, https://doi.org/10.1038/nature02003
19 Yang, J., Kim, K.D., Lucas, A., Drahos, K.E., Santos, C.S., Mury, S.P. et al. (2008) A novel heme-regulatory motif mediates heme-dependent degradation of the circadian factor period 2. *Mol. Cell. Biol.* **28**, 4697–4711, https://doi.org/10.1128/MCB.00236-08
20 Shen, J., Sheng, X., Chang, Z., Wu, Q., Wang, S., Xuan, Z. et al. (2014) Iron metabolism regulates p53 signaling through direct Heme-p53 interaction and modulation of p53 localization, stability, and function. *Cell Rep.* **7**, 180–193, https://doi.org/10.1016/j.celrep.2014.02.042
21 Roumenina, L.T., Radanova, M., Atanasov, B.P., Popov, K.T., Kaveri, S.V., Lacroix-Desmazes, S. et al. (2011) Heme interacts with C1q and inhibits the classical complement pathway. *J. Biol. Chem.* **286**, 16459–16469, https://doi.org/10.1074/jbc.M110.206136
22 Frimat, M., Tabarin, F., Dimitrov, J.D., Poitou, C., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V. et al. (2013) Complement activation by heme as a secondary hit for atypical hemolytic uremic syndrome. *Blood* **122**, 282–292, https://doi.org/10.1182/blood-2013-03-489245
23 Repessé, Y., Dimitrov, J.D., Peyron, I., Moshai, E.F., Kiger, L., Dasgupta, S. et al. (2012) Heme binds to factor VIII and inhibits its interaction with activated factor IX. *J. Thromb. Haemost.* **10**, 1062–1071, https://doi.org/10.1111/j.1538-7836.2012.04724.x
24 Kumar, A., Wißbrock, A., Goradia, N., Bellstedt, P., Ramachandran, R., Imhof, D. et al. (2018) Heme interaction of the intrinsically disordered N-terminal peptide segment of human cystathionine-β-synthase. *Sci. Rep.* **8**, 2474
25 Atamna, H. and Boyle, K. (2006) Amyloid-beta peptide binds with heme to form a peroxidase: relationship to the cytopathologies of Alzheimer's disease. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3381–3386, https://doi.org/10.1073/pnas.0600134103
26 Balwani, M. and Desnick, R.J. (2012) The porphyrias: advances in diagnosis and treatment. *Blood.* **120** ((23)), 4496–4504, https://doi.org/10.1182/blood-2012-05-423186
27 Goodfellow, B.J., Dias, J.S., Ferreira, G.C., Henklein, P., Wray, V. and Macedo, A.L. (2001) The solution structure and heme binding of the presequence of murine 5-aminolevulinate synthase. *FEBS Lett.* **505**, 325–331, https://doi.org/10.1016/S0014-5793(01)02818-6

28 Lombardi, A., Nastri, F. and Pavone, V. (2001) Peptide-based heme–protein models. *Chem. Rev., Am Chem. Soc.* **101**, 3165–3190, https://doi.org/10.1021/cr000055j
29 Kühl, T., Sahoo, N., Nikolajski, M., Schlott, B., Heinemann, S.H. and Imhof, D. (2011) Determination of hemin-binding characteristics of proteins by a combinatorial peptide library approach. *Chem. Bio. Chem.* **12**, 2846–2855, https://doi.org/10.1002/cbic.201100556
30 Sweeney, M.C., Wavreille, A.S., Park, J., Butchar, J.P., Tridandapani, S. and Pei, D. (2005) Decoding protein-protein interactions through combinatorial chemistry: sequence specificity of SHP-1, SHP-2, and SHIP SH2 domains. *Biochemistry* **44**, 14932–14947, https://doi.org/10.1021/bi051408h
31 de Castro, E., Sigrist, C. J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E. et al. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–W365, https://doi.org/10.1093/nar/gkl124
32 Brewitz, H.H., Hagelueken, G. and Imhof, D. (2016) Structural and functional diversity of transient heme binding to bacterial proteins. *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 683–697, https://doi.org/10.1016/j.bbagen.2016.12.021
33 Brewitz, H.H., Kühl, T., Goradia, N., Galler, K., Popp, J., Neugebauer, U. et al. (2015) Role of the chemical environment beyond the coordination site: structural insight into Fe(III) protoporphyrin binding to cysteine-based heme-regulatory protein motifs. *Chem. Bio. Chem.* **16**, 2216–2224, https://doi.org/10.1002/cbic.201500331
34 Brewitz, H.H., Goradia, N., Schubert, E., Galler, K., Kühl, T., Syllwasschya, B. et al. (2016) Heme interacts with histidine- and tyrosine-based protein motifs and inhibits enzymatic activity of chloramphenicol acetyltransferase from E. coli. *Biochim. Biophys. Acta - Gen. Subj.* **1860**, 1343–1353, https://doi.org/10.1016/j.bbagen.2016.03.027
35 Marcelli, A., Jelovica Badovinac, I., Orlic, N., Salvi, P.R. and Gellini, C. (2013) Excited-state absorption and ultrafast relaxation dynamics of protoporphyrin IX and hemin. *Photochem. Photobiol. Sci.* **12**, 348–355, https://doi.org/10.1039/C2PP25247C
36 Nienhaus, K. and Nienhaus, G.U. (2005) Probing heme protein-ligand interactions by UV/visible absorption spectroscopy. *Methods Mol. Biol.* **305**, 215–242
37 Soret, J.-L. (1883) Analyse spectrale: Sur le spectre d'absorption du sang dans la partie violette et ultra-violette. *Comptes rendus l'Académie des Sci* **97**, 1269–1270
38 Uttamlal, M. and Sheila Holmes-Smith, A. (2008) The excitation wavelength dependent fluorescence of porphyrins. *Chem. Phys. Lett.* **454**, 223–228, https://doi.org/10.1016/j.cplett.2008.02.012
39 Papadopoulos, P.G., Walter, S.A., Li, J. and Baker, G.M. (1991) Proton interactions in the resting form of cytochrome oxidase. *Biochemistry* **30**, 840–850, https://doi.org/10.1021/bi00217a038
40 Luthra, A., Denisov, I.G. and Sligar, S.G. (2011) Spectroscopic features of cytochrome P450 reaction intermediates. *Arch. Biochem. Biophys.* **507** ((1)), 26–35, https://doi.org/10.1016/j.abb.2010.12.008
41 Sono, M., Roach, M.P., Coulter, E.D. and Dawson, J.H. (1996) Heme-containing oxygenases. *Chem. Rev.* **96**, 2841–2888, https://doi.org/10.1021/cr9500500
42 Spiro, T.G. (1985) Resonance Raman spectroscopy as a probe of heme protein structure and dynamics. *Adv. Protein Chem.* **37**, 111–159, https://doi.org/10.1016/S0065-3233(08)60064-9
43 Nakamura, M., Ikeue, T., Ohgo, Y., Takahashi, M. and Takeda, M.T. (2002) Highly saddle shaped (porphyrinato)iron(III) iodide with a pure intermediate spin state. *Chem. Commun.* 1198–1199, https://doi.org/10.1039/b202768b
44 Spiro, T.G. and Burke, J.M. (1976) Protein control of porphyrin conformation. comparison of resonance raman spectra of heme proteins with mesoporphyrin IX analogs. *J. Am. Chem. Soc.* **98**, 5482–5489, https://doi.org/10.1021/ja00434a013
45 Kitagawa, T., Abe, M., Kyogoku, Y., Ogoshi, H., Watanabe, E. and Yoshida, Z. (1976) Resonance Raman spectra of metallooctaethylporphyrins. Low frequency vibrations of porphyrin and iron-axial ligand stretching modes. *J. Phys. Chem.* **80**, 1181–1186, https://doi.org/10.1021/j100552a012
46 Bagai, I., Sarangi, R., Fleischhacker, A.S., Sharma, A., Hoffman, B.M., Zuiderweg, E.R.P. et al. (2015) Spectroscopic studies reveal that the heme regulatory motifs of heme oxygenase-2 are dynamically disordered and exhibit redox-dependent interaction with heme. *Biochemistry* **54**, 2693–2708, https://doi.org/10.1021/bi501489r
47 Yin, L., Dragnea, V., Feldman, G., Hammad, L.A., Karty, J.A., Dann, C.E. et al. (2013) Redox and light control the heme-sensing activity of AppA. *MBio* **4**, e00563–13, https://doi.org/10.1128/mBio.00563-13
48 Ishikawa, H., Nakagaki, M., Bamba, A., Uchida, T., Hori, H., O'Brian, M.R. et al. (2011) Unusual heme binding in the bacterial iron response regulator protein: spectral characterization of heme binding to the heme regulatory motif. *Biochemistry* **50**, 1016–1022, https://doi.org/10.1021/bi101895r
49 Caillet-Saguy, C., Piccioli, M., Turano, P., Lukat-Rodgers, G., Wolff, N., Rodgers, K.R. et al. (2012) Role of the iron axial ligands of heme carrier HasA in heme uptake and release. *J. Biol. Chem.* **287**, 26932–26943, https://doi.org/10.1074/jbc.M112.366385
50 Caillet-Saguy, C., Piccioli, M., Turano, P., Izadi-Piuneyre, N., Delepierre, M., Bertini, I. et al. (2009) Mapping the interaction between the hemophore HasA and its outer membrane receptor HasR using CRINEPT-TROSY NMR spectroscopy. *J. Am. Chem. Soc.* **131**, 1736–1744, https://doi.org/10.1021/ja804783x
51 Moriwaki, Y., Caaveiro, J.M.M., Tanaka, Y., Tsutsumi, H., Hamachi, I. and Tsumoto, K. (2011) Molecular basis of recognition of antibacterial porphyrins by heme-transporter IsdH-NEAT3 of Staphylococcus aureus. *Biochemistry* **50**, 7311–7320, https://doi.org/10.1021/bi200493h
52 Hu, R.-G., Wang, H., Xia, Z. and Varshavsky, A. (2008) The N-end rule pathway is a sensor of heme. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 76–81, https://doi.org/10.1073/pnas.0710568105
53 Lechardeur, D., Cesselin, B., Liebl, U., Vos, M.H., Fernandez, A., Brun, C. et al. (2012) Discovery of intracellular heme-binding protein HrtR, which controls heme efflux by the conserved HrtB-HrtA transporter in Lactococcus lactis. *J. Biol. Chem.* **287**, 4752–4758, https://doi.org/10.1074/jbc.M111.297531
54 Schubert, E., Florin, N., Duthie, F., Brewitz, H.H., Kühl, T., Imhof, D. et al. (2015) Spectroscopic studies on peptides and proteins with cysteine-containing heme regulatory motifs (HRM). *J. Inorg. Biochem.* **148**, 49–56, https://doi.org/10.1016/j.jinorgbio.2015.05.008

55 D'Souza, A., Mahajan, M. and Bhattacharjya, S. (2016) Designed multi-stranded heme binding β-sheet peptides in membrane. *Chem. Sci. R. Soc. Chem.* **7**, 2563–2571
56 Nagarajan, D., Sukumaran, S., Deka, G., Krishnamurthy, K., Atreya, H.S. and Chandra, N. (2018) Design of a heme-binding peptide motif adopting a β-hairpin conformation. *J. Biol. Chem.* **293**, 9412–9422, https://doi.org/10.1074/jbc.RA118.001768
57 D'Souza, A., Wu, X., Yeow, E.K.L. and Bhattacharjya, S. (2017) Designed heme-cage β-sheet miniproteins. *Angew. Chem.* **129**, 5998–6002
58 Rai, J. (2017) Mini Heme-Proteins: designability of structure and diversity of functions. *Curr. Protein Pept. Sci.* **18**, 1132–1140, https://doi.org/10.2174/1389203718666170515144037
59 Shifman, J.M., Gibney, B.R., Sharp, R.E. and Dutton, P.L. (2000) Heme redox potential control in de novo designed four-α-helix bundle proteins. *Biochem. Am. Chem. Soc.* **39**, 14813–14821
60 Peherstorfer, S., Brewitz, H.H.B., Paul George, A.A., Wißbrock, A., Adam, J.M., Schmitt, L. et al. (2018) Insights into mechanism and functional consequences of heme binding to hemolysin-activating lysine acyltransferase HlyC from Escherichia coli. *Biochim. Biophys. Acta* **1862**, 1964–1972, https://doi.org/10.1016/j.bbagen.2018.06.012
61 Yu, D.J., Hu, J., Yang, J., Shen, H.B., Tang, J. and Yang, J.Y. (2013) Designing template-free predictor for targeting protein-ligand binding sites with classifier ensemble and spatial clustering. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **10**, 994–1008, https://doi.org/10.1109/TCBB.2013.104
62 Liu, R. and Hu, J. (2011) Computational prediction of heme-binding residues by exploiting residue interaction network. *PLoS ONE* **6**, e25560, https://doi.org/10.1371/journal.pone.0025560
63 Liu, R. and Hu, J. (2011) HemeBIND: a novel method for heme binding residue prediction by combining structural and sequence information. *BMC Bioinformatics* **12**, 207–219, https://doi.org/10.1186/1471-2105-12-207
64 Chen, H. and Zhou, H.X. (2005) Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. *Nucleic Acids Res.* **33**, 3193–3199, https://doi.org/10.1093/nar/gki633
65 Shan, Y., Wang, G. and Zhou, H.X. (2001) Fold recognition and accurate query-template alignment by a combination of PSI-BLAST and threading. *Proteins Struct. Funct. Genet.* **42**, 23–37, https://doi.org/10.1002/1097-0134(20010101)42:1%3c23::AID-PROT40%3e3.0.CO;2-K

Supplementary Material

The Molecular Basis of Transient Heme-Protein Interactions: Analysis, Concept and Implementation

Amelie Wißbrock[1], Ajay Abisheck Paul George[1],
Hans Henning Brewitz[1], Toni Kühl[1], and Diana Imhof[1]*

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

SeqD-HBM development

The current stand-alone version of *SeqD-HBM,* designed to be used both on Linux and Microsoft Windows based operating systems was written and tested in *Python 3.6.4* and *Python 2.7.15* on a Linux workstation running *Ubuntu 18.4* and on workstations running Windows 7 and Windows 10 operating systems. Parallel installations of *Python 2.X* and *3.X* versions are required on the same system to automatically post of the sequences to the WESA server for solvent accessibility predictions. This part of the logic was developed using hints from the WESA documentation for running batch jobs (http://pipe.sc.fsu.edu/PostHandler/WESA-PostHandler.htm). Windows users also need the *wget* program present as an executable in the working directory.

SeqD-HBM usage

With all of the programs and scripts correctly organized and the requirements fulfilled, *SeqD-HBM* can be run on the command line by issuing a command such as the following.

python SeqD-HBM_Linux_Win.py <input_fasta_file> default OR *python SeqD-HBM_Linux_Win.py <input_fasta_file> structure*

In the line above *<input_fasta_file>* is the name of the FASTA file with the sequence(s) and the *default* and *structure* arguments indicate the mode of operation. It must be noted that the input file must be placed in the same directory as the *SeqD-HBM* python script. The program can take multiple sequences given one below the other in a file as long as they are specified in the FASTA format. Even in the case of a single sequence, the program expects the sequence to be given in a input file with a header as per the FASTA format. In the *default* mode, the program automatically posts the sequence to the WESA server and attempts to fetch the solvent accessibility prediction once every two minutes until a successful output is obtained. The *default* mode will not process sequences more than 2000 amino acid residues long since this is a limit set by WESA.While using the program for a large number of sequences, it is recommended to redirect the output into a text file.

SeqD-HBM output

The current version of *SeqD-HBM* first checks all of the sequences in the file for junk or non-standard characters. The program only accepts the 20 standard amino acids as input which is the same case with WESA and informs the user if there are errors in the input file. The main output for each sequence is the table containing the predicted coordination site, the associated 9mer motif and the net charge on the motif. An additional column named “comment” provides useful hints to the user regarding each predicted motif. This includes special instances such as when the motif has a net charge of 0 or less but if the coordinating residue

is a cysteine or if the motif is a CP motif. Another message that can occur in the "comment" column of the output is hinting to the user the possible occurrence of disulfide bonds in the predicted motif when the sequence contains multiple cysteine residues.

SeqD-HBM testing

SeqD-HBM was tested for its accuracy and performance on different datasets. The performance of the program was tested on a set of ~600 protein sequences with an average sequence length of 135 amino acids. In the *structure* mode, where the WESA computation is skipped, the program took only 1.62 seconds to process the output for these proteins. Of course, since WESA uses five different machine learning algorithms, it does prove to be a performance bottleneck with respect to execution time especially while operating on a large set of sequences. However, this is an acceptable tradeoff considering the accuracy of the prediction when absolutely no structural information is available.

Test sequence

```
>sp|P06736|HLYC_ECOLX Hemolysin-activating lysine-acyltransferase HlyC
OS=Escherichia coli OX=562 GN=hlyC PE=1 SV=1
MNINKPLEILGHVSWLWASSPLHRNWPVSLFAINVLPAIQANQYVLLTRDDYPVAYCSWA
NLSLENEIKYLNDVTSLVAEDWTSGDRKWFIDWIAPFGDNGALYKYMRKKFPDELFRAIR
VDPKTHVGKVSEFHGGKIDKQLANKIFKQYHHELITEVKRKSDFNFSLTG
```

We use a 170 residue long sequence of the protein HlyC (Hemolysin-activating lysine-acyltransferase Uniprot ID P06736) from *Escherichia coli* as a test sequence to demonstrate the usage of *SeqD-HBM* in both its *default* and *structure* modes of operation. This protein has no experimentally determined structure available. *SeqD-HBM* used in with the *structure* mode took 0.05 seconds to execute and predict H23, C57, Y104, Y106, H126, H134, Y150, H151 and H152 as the potential heme binding sites. Running the same sequence in the default mode took 7.64 minutes (458.39 seconds) for the overall execution with the WESA computation but the prediction eliminated the residues C57, Y104 and Y150 as buried hence improving the accuracy by removing 3 false positives.