Evolutionary alternatives examined with three examples: Amino acids, coiled coils and strategies of iron-cycling bacteria

**Dissertation
zur Erlangung des akademischen Grades "doctor
rerum naturalium" (Dr. rer. nat.)**

**vorgelegt dem Rat der Fakultät für Biowissenschaften der
Friedrich-Schiller-Universität Jena**

von M. Sc. Chemische Biologie, André Then
geboren am 27. Februar 1993 **in** Frankfurt a. M.

Gutachter

Prof. Dr. Stefan Schuster, Friedrich-Schiller-Universität Jena.
Prof. Dr. Bashar Ibrahim, Gulf University for Science and Technology Kuwait.
Prof. Dr. Dirk Walther, Max-Planck-Institut für Molekulare Pflanzenphysiologie Potsdam.

Verteidigung

Die Verteidigung erfolgte am 21.11.2022.

Abstract

The questions of why things are the way they are and if they could have been any different are frequently occupying the minds of us human beings. One way out is provided by the assumption of contingency, the view that long-term development is mainly dependent on the results of many random events. However, I argue that from a scientific point of view, fundamentally revolving around skepticism and the search for underlying patterns, contingency does not provide a comfortable answer and should always be perceived as a preliminary resort. This thesis revolves around the investigation of evolutionary alternatives related to case studies at three different levels of biological complexity. Key aspects of evolutionary alternatives are the pool of available elements to choose from, the pressures which lead to the preference for the selection of certain choices over others, and the conditions under which these selections comprise a viable or even optimal choice for the organism(s).

On the lowest level of biological complexity, we will deal with the selection and optimality of the 20 proteinogenic amino acids. Experiments on prebiotic chemistry show that the pool of amino acids available for incorporation into the system of genetic encoding contained many more alternative options. Although pressures for the preference of some amino acids from the pool over others are also discussed, the main focus in this part resides on the question of the optimality of the 20 proteinogenic amino acids. Previous studies try to answer this question by examining the occupation of chemical space by the 20 proteinogenic amino acids and comparing it to randomly chosen sets of putatively available amino acids at the emergence of life. Complementary, our approach focuses more on the unique functional roles of each amino acid. Therefore, we developed a median-based classification approach using binary feature patterns to classify the 20 proteinogenic amino acids. In comparison to the famous classification of Ramsay Taylor our approach provides an optimal solution as it requires only the minimum number of features for complete classification.

The second part considers the design space of coiled coils. Coiled coils are a prominent type of supersecondary structure. They form oligomers of α-helices and take part in a multitude of different functions, such as transcriptional regulation, catalytic activity, and provision of structural stability. Critical for the stability of coiled coils are non-covalent interactions between α-helix residues pointing toward the core of the coiled-coil. Geometrically, the distances with which these interactions reoccur along the sequence are restricted to certain discrete values. We developed a recursive algorithm to compute all possible decompositions of inter-contact distances for any period length. To further narrow down the design space of coiled coils, experimentally determined coiled-coil decompositions are extracted from publicly available databases. One key aspect is the forces driving the realization of demanding coiled-coil decompositions deviating from the canonical heptad repeat.

In the last section, we arrive at the level of microbial interactions. The system of bacteria employing iron ions as electron donors or electron acceptors, respectively, is characterized by an astonishing amount of interdependencies. At a basic level, Fe(III) reducing and Fe(II) oxidizing bacteria provide each other with the required substrate for their respiration and therefore allow for the cycling of iron between different ionization states. In detail, the efficiency of this process is dependent on a multitude of factors, such as spatial distribution of the bacteria, temporal oscillations of oxygen, pH, and the choice of microbial strategies pursued. We developed an agent-based model to derive preconditions for efficient cooperation between these bacteria, with a focus on the dynamics of iron nanoparticles.

Zusammenfassung

Die Fragen warum die Dinge sind, wie sie sind und ob sie anders hätten sein können, haben uns Menschen seit jeher beschäftigt. Ein Ausweg stellt die Annahme der Kontingenz dar, also die Ansicht, dass die langfristige Entwicklung hauptsächlich ein Produkt der Resultate vieler Zufallsereignisse darstellt. Aus wissenschaftlicher Perspektive, deren grundlegende Charakteristika Zweifel und die Suche nach zugrundeliegenden Mustern darstellen, kann Kontingenz nur als eine Zwischenlösung dienen. Diese Arbeit behandelt die Untersuchung evolutionärer Alternativen in Hinblick auf Fallstudien auf drei unterschiedlichen Ebenen biologischer Komplexität. Schlüsselaspekte evolutionärer Alternativen sind der Möglichkeitsraum auszuwählender Elemente, die Drücke, die die Bevorzugung gewisser Auswahlmöglichkeiten mit sich bringen und die Bedingungen unter denen diese gewählten Möglichkeiten eine funktionierende oder gar optimale Lösung aus Sicht der Organismen darstellen.

Auf der niedrigsten Ebene biologischer Komplexität widmen wir uns der Auswahl und Optimalität der 20 proteinogenen Aminosäuren. Experimente der präbiotischen Chemie haben gezeigt, dass der Pool für den Einbau in den genetischen Code verfügbarer Aminosäuren neben den 20 proteinogenen Aminosäuren viele andere Optionen beinhaltete. Auch wenn die Drücke, die die Bevorzugung gewisser Aminosäuren des Pools gegenüber anderer diskutiert werden, verbleibt der Fokus bei der Frage hinsichtlich der Optimalität der 20 proteinogenen Aminosäuren. Vorausgehende Studien versuchen diese Frage zu beantworten, indem sie die Besetzung des chemischen Raums durch die 20 proteinogenen Aminosäuren untersuchen und sie mit der Besetzung durch zufällige Sets mutmaßlich verfügbarer Aminosäuren zum Zeitpunkt der Entstehung des Lebens vergleichen. Komplementär dazu verfolgt unser Ansatz die einzigartigen funktionellen Rollen unterschiedlicher Aminosäuren. Dafür haben wir einen Ansatz zur Median-basierten Klassifikation anhand binärer Eigenschaftsmuster der 20 proteinogenen Aminosäuren entworfen. Im Vergleich mit der berühmten Klassifikation Ramsay Taylors liefert unser Ansatz eine optimale Lösung, da zur vollständig disjunkten Klassifikation nur die minimale Anzahl von Eigenschaften benötigt wird.

Der zweite Teil der Arbeit beschäftigt sich mit der Beschreibung des Raums möglicher Coiled Coil Strukturen. Coiled Coils stellen einen verbreiteten Typ von Supersekundärstrukturen dar. Sie setzen sich aus einzelnen α-Helices zusammen und partizipieren in einer Reihe unterschiedlicher Funktionen, z. B. Regulation der Transkription, katalytische Aktivität und die Bereitstellung struktureller Stabilität. Kritisch für die Stabilität von Coiled Coils sind non-kovalente Wechselwirkungen zwischen Resten der α-Helices die in Richtung des Coiled Coil Zentrums orientiert sind. Geometrisch sind die Abstände in denen diese Wechselwirkungen entlang der Sequenz auftreten auf wenige diskrete Werte limitiert. Wir entwickelten einen rekursiven Algorithmus um alle möglichen Zerlegungen dieser Abstände zwischen non-kovalenten Kontakten für jede beliebige Periodenlänge zu ermitteln. Um den Möglichkeitsraum für die Realisierung von Coiled Coils genauer zu beschreiben, wurden verfügbare Datenbanken auf experimentell bestimmte Coiled Coil Zerlegungen analysiert. Ein Schlüsselaspekt stellen die Kräfte dar, die die Realisierung anspruchsvoller Coiled Coil Zerlegungen antreiben, also solcher die vom kanonischen Heptad Motiv abweichen.

Im letzten Teil sind wir auf der Ebene mikrobieller Interaktionen angelangt. Das System von Bakterien, die Eisenionen als Elektronendonoren oder -akzeptoren nutzen, zeichnet sich durch ein erstaunliches Maß an gegenseitigen Abhängigkeiten aus. Grundlegend betrachtet stellen Fe(III) reduzierende und Fe(II) oxidierende Bakterien einander die benötigten Substrate für die jeweiligen Respirationsprozesse zur Verfügung und ermöglichen dadurch einen Zyklus zwischen Eisenionen unterschiedlicher Wertigkeit. Im Detail jedoch hängt die Effizienz dieses Prozesses von einer Vielzahl von Faktoren ab, so etwa der räumlichen Organisation der Bakterien, der zeitlichen Oszillation von Sauerstoff, dem pH und

der Realisierung mikrobieller Strategien. Wir entwickelten ein Agenten-basiertes Model um die Voraussetzungen für eine effiziente Kooperation zwischen diesen Bakterien zu ermitteln. Einen Fokus stellte dabei die Dynamik von Eisen-Nanopartikeln dar.

Contents

1.Introduction

1.1. Evolutionary alternatives for proteinogenic amino acids

1.1.1. Available amino acids at the origin of life

For the treatment of the question of, if and how the 20 proteinogenic amino acids of the genetic code represent an optimal choice, it is necessary to first summarize what is known about the pool of amino acids available at the origin of life, approximately 4.5 Ga to 3.9 Ga[1]. The presumed composition of amino acids depends on the processes which took part in the formation of the amino acids. Among the potential sources of amino acids discussed is the formation in hydrothermal vents, an introduction via external sources, such as comets or meteorites, or atmospheric and photolytic reactions, triggered for example by electric discharge[2]. Probably the most well-known experiment trying to mimic the conditions of prebiotic amino acid synthesis is the Miller-Urey experiment[3,4].

Miller and Urey performed sparked and silent discharges on a highly reducing atmosphere composed of CH_4, NH_3, H_2O, and H_2, while capturing the produced non-volatile organic compounds in an aqueous solution. Via paper chromatography they were able to detect milligram amounts of simple proteinogenic amino acids such as glycine, alanine, aspartic acid, and glutamic acid. In addition, they detected also other, non-proteinogenic α-amino acids, for example, α-amino butyric acid. β-amino acids, such as β-alanine, were less prevalent. This led to the hypothesis that the reaction mechanism for the most amino acid is a variation of the Strecker synthesis[5]. In the Strecker synthesis an aldehyde, ammonium, and cyanide react to acetonitrile which is subsequently hydrolyzed into an amino acid. The Strecker synthesis produces mainly α-amino acids. In respect to the Miller-Urey experiment, the assumption is that the educts of the reaction are formed in the gaseous phase, whereas the reaction to the amino acids then occurs in solution. β-amino acids are then assumed to be formed by alternative reaction mechanisms.

Since the original experiment by Miller and Urey analytical techniques made substantial progress, now allowing for the quantification and detection even of low abundant compounds. One study employed modern analytical techniques on original samples of one of Millers experiments where he included H_2S as a sulfur source[6]. The authors were able to detect and quantify 23 amino acids and 4 amines, including amino acids not detected by Miller himself, such as threonine, serine, leucine, isoleucine, and methionine. Although cysteine itself was not detectable, its known degradation products cysteamine and homocysteic acid were found. A recent review notes that except for tyrosine, tryptophan, and glutamine, all proteinogenic amino acids were abiotically synthesized in at least one experiment mimicking prebiotic chemistry[7].

Although most single experiments on prebiotic chemistry produce only a limited amount of amino acids[8], combined they provide an ample pool of possible candidates for usage as monomer building blocks at the origin of life. This can also be seen by analyzing the amino acid diversity of meteorites, which are considered an external input source of amino acids[9]. Meteorites have been shown to sometimes contain over 80 different aliphatic amino acids of a chain length up to eight carbon atoms[10]. Although it might be tempting to include the amino acids accessible via enzymatic synthesis into the pool of available amino acids as well, this reasoning appears questionable. The product range of the enzymatic machinery is dependent on the 20 proteinogenic amino acids living organisms actually incorporated in their genetic code. Without further evidence, the enzymatically accessible amino acids

can therefore not be safely regarded as available options at the onset of life. The range of compounds accessible via enzyme synthesis could be regarded as a criterion for the optimality of the set of 20 proteinogenic amino acids, though. The impressive range of accessible compounds is exemplified by the natural occurrence of all theoretically possible branched and unbranched amino acids solely composed of single bonded carbon up to a chain length of 5[11]. Overall it seems reasonable to assume that at least a few hundred amino acids were available for incorporation into the earliest version of a genetic code. Of course, it is safe to assume that they varied in abundance, as well as in terms of regional distribution. Abundance itself does not provide the most important criterion for the selection of amino acids in early life though[12]. In the following chapter we will explore how functional criteria might have driven the selection of a set of early amino acids.

1.1.2. Functional criteria and optimality of the 20 proteinogenic amino acids

In an early attempt to explain why these particular 20 amino acids form the basis for the protein synthesis of all organisms, Weber and Miller divided the amino acids into functional groups and reasoned about why they are preferable to other available options within the same group[12]. For example, longer linear alkyl chains in hydrophobic amino acids as they occur in norvaline and norleucine might be disadvantageous as they come with a higher degree of rotational flexibility, thereby impeding the formation of ordered tertiary protein structures. As another reason for the exclusion of norleucine, they mention the structural similarity of norleucine to methionine supported by the fact that methionine has been replaced by norleucine in some proteins without the loss of enzymatic activity. Such reasoning about the preferability of amino acids in terms of functional properties is especially relevant in cases where equal abundance would have made it possible to have incorporated alternative amino acids as well.

Aside from attempting to find plausible, but non-testable reasons for the occurrence of these particular 20 proteinogenic amino acids, other studies focused more on identifying quantitative measures which then form the basis for comparison with random sets of amino acids sampled from a pool of reasonable candidates. The research of Philip and Freeland is part of the latter category as they compare the coverage of the chemical space for certain physicochemical characteristics by the proteinogenic amino acids to those of random sets[13]. The coverage itself consists of range, *i. e.* the difference between the maximum and the minimum value for the physicochemical characteristic, and evenness, the variance of the intervals between consecutive members of amino acids sorted based on their value for the physicochemical characteristic. The authors mention coverage as an improvement to their previously employed measure on the adaptive value of amino acid alphabets, which is the statistical variance of the amino acid alphabet on the physicochemical characteristic[14]. This measurement is flawed in that it favors amino acid alphabets whose members cluster towards the ends of the range of the occupied chemical space. As the amino acids exhibit a high redundancy in the considered physicochemical parameters in question, those alphabets should be considered as alphabets with low adaptive value. In their study, Philip and Freeland choose size, charge, and hydrophobicity as relevant physicochemical characteristics. They show that only a low percentage of random sets outperforms the proteinogenic amino acids in coverage of the chemical space associated with these characteristics and combinations between them (between 0.0 and 3.4 %).

Whereas Philip and Freeland only considered an amino acid pool size of at most 76 amino acids, a latter study extended the pool to a library of 1913 amino acids within the molecular weight range of the proteinogenic amino acids[15]. They found that from the 10^8 possible sets, only six achieved a better coverage of the three-dimensional chemical space of size, charge, and hydrophobicity. Interestingly

they also showed that all six sets exhibit higher summed heats of formation ($\Delta H_f°$). The free energy of formation ΔG, a thermodynamic quantity related to the heat of formation, has been previously shown to correlate well with the abundancies of amino acids from prebiotic synthesis experiments and the chemical analysis of meteorites[16]. In combination, this strongly supports natural selection rather than chance being responsible for the selection of this particular set of 20 proteinogenic amino acids. Amino acids might actually have been chosen based on their coverage of chemical space. From the range of suitable candidate sets, the abundancies of the amino acid elements might then have been the decisive factor.

Another way to look at the adaptability of the 20 proteinogenic amino acids is to look at their historic development. Based on their abundancies and complexity It is deemed unrealistic that all the proteinogenic amino acids were part of the coding set from the start. Instead, there is support for a consensus set of nine early amino acids consisting of Ala, Asp, Glu, Gly, Leu, Pro, Ser, Thr, and Val with high prebiotic abundancies and sufficient capability of forming foldable proteins[17]. This set of an early set of amino acids could then have been complemented by stepwise integration of more complex amino acids via codon evolution[18]. This idea has been further supported by experimental and theoretical mutation studies which showed that a reduced set of amino acids still allows for the proper folding of extant proteins[19,20]. This raises the question of why this functional early set of amino acids had to be further extended in the first place.

As Higgs pointed out[21], translational errors increase when more amino acids are added to the genetic code. This disadvantage has to be justified by the newly incorporated amino acid adding to the existing chemical diversity of the set. The larger the set already is, the smaller the benefit upon the addition of new amino acids. Examining the consensus set of ten early amino acids it becomes apparent that they lack key catalytic groups. The three amino acids which exhibit high catalytic propensity, His, Cys, and Arg, are assumed to be late introductions into the genetic code[22]. This hypothesis is corroborated by a study that analyzed the gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) as a proxy for reactivity[23]. It has been shown that amino acids assumed to be late additions to the genetic code, have a lower HOMO-LUMO gap and are therefore more reactive. In conclusion, the early set of amino acids might have been sufficient in providing the structural basis for protein folding, for a diverse range of catalytic activity though the recruitment of further amino acids was necessary.

1.1.3. Classification approaches for amino acids

When talking about extending the genetic code, both in the initial phases of the evolution of life and nowadays in synthetic biology, it is useful to classify the amino acid for characterizing their functional roles and redundancies. The most famous classification of amino acids has been conducted by William Ramsay Taylor[24]. His aim was to condense the information of Dayhoffs' matrix depicting the odds that amino acid will mutate into another[25] in a more accessible format. Initial visualization attempts showed that there is a good correspondence between two physicochemical properties, notably size and hydrophobicity, on one side, and Dayhoffs' mutational odds on the other. The eight features Taylor employed - aliphatic, aromatic, charged, hydrophobic, polar, positive, small, and tiny – are therefore mostly derived from these two properties.

For visualization, Taylor uses an Euler diagram, although he erroneously terms it a Venn diagram. The difference is that in a Venn diagram every possible intersection between the features has to be displayed. In Taylors visualization, this is not the case, as for example there is no intersection displayed between the features positive and tiny. The classification does not provide complete segregation of

amino acids, as the pairs, Tyr-Trp, Ile-Leu, and Ala-Gly reside together in the same sub-sets. Furthermore, even though Taylor mentions that for the purpose of simplicity as few sets as possible have been introduced, in this sense his classification is not optimal as he uses more than the minimum number of five features to classify the 20 proteinogenic amino acids. The usefulness of his classification is demonstrated by applying logical operations on the features to construct the smallest set or sub-sets to describe all conserved amino acids in a specific sequence alignment.

The aforementioned sets resulting from logical operations on the features are often difficult to interpret, as combinations of Taylors' features, which center mainly around size and hydrophobicity, are not always well suited to describe the characteristics of amino acids responsible for their conservation in a sequence alignment. Later approaches detach themselves even further from interpretability and focus on identifying classifications optimally representing the amino acid conservation of the sequence alignment at hand. Kosiol *et al.* assume a Markov process for the description of protein evolution and aim at identifying almost invariant sets of amino acids[26]. Amino acids within such sets often mutate into each other in protein sequence alignment but rarely into members of other almost invariant sets. This strength of mixing within a Markov process can be quantified by calculating the conductance, a measure for the tendency of elements to mutate around states. Considering this measure, the classification approach has been shown to outperform Taylors' classification based on physicochemical properties of the amino acids, although this comes at the price of further reducing interpretability, as the algorithm does not consider any amino acid properties despite conservation in the sequence alignment at hand. A further disadvantage is that the algorithm requires the specification of the number of sets as input and does not provide a way to decide on the best number of sets.

Many more reduction schemes of amino acids have been suggested over the years, based on, chemistry, sequence alignments, structure alignments, contact potential, and protein blocks[27]. Their benefits show in many machine learning applications by simplifying feature extraction, boosting predictive performance for example in protein fold and antimicrobial peptide prediction[28,29]. Unfortunately, most classification approaches for improving performance in machine learning heavily rely on data only available for the 20 proteinogenic amino acids. Therefore, these classification schemes are not capable of easily extending to non-proteinogenic amino acids and do not provide any guidance on which amino acids could provide functional benefits upon addition to the genetic code. A discussion sparked by the rapid developments in synthetic biology.

1.1.4. Extending the genetic code with non-proteinogenic amino acids

Neither the assignment of the 64 genetic codons, nor the composition of proteinogenic amino acids are carved in stone. In fact, genetic code expansion towards the integration of selenocysteine and pyrrolysine is naturally occurring via the repurposing of stop codons[30]. These 21st and 22nd proteinogenic amino acids differ though in how they are synthesized. Selenocysteine requires serine bound to the selenocysteine tRNA[Sec] as a precursor. An enzyme called selenocysteine synthase then catalyzes the transfer of selenium from selenophosphate. A stem-loop secondary structure in the mRNA is then required to recognize the opal stop codon (UGA) for a selenocysteine codon. Pyrrolysine in contrast is directly attached to its tRNA[Pyl] and is inserted in amber stop codons (TAG). There are 25 proteins in humans that contain selenocysteine[31]. Selenocysteine has probably been introduced to the genetic code because of its suitability for catalyzing redox-reactions, as all functionally catalyzed proteins belong to this category[32,33]. This is in good agreement with selenocysteine having a smaller HOMO-LUMO gap than any of the other proteinogenic amino acids[23]. Pyrrolysine occurs only in a few enzymes with methyltransferase activity and has been shown to be important for the binding of

methylammonium[34]. Thus, both proteinogenic amino acid additions seem to have taken place primarily for reasons of catalytic activity.

Building on the insights of natural genetic code expansion, the process has also been made synthetically available for introducing a variety of non-proteinogenic amino acids. Young and Schultz summarize the requirements for the artificial expansion of the genetic code in their review[35]. A codon which does not code for any of the natural amino acids has to be reutilized. Possible options include the three stop codons, rarely used codons for an amino acid that has multiple codons, or engineered quadruplet codons. Because it has the lowest frequency of all the three stop codons, the most commonly used codon is the amber codon (TAG)[36]. Furthermore, an orthogonal pair of an aaRS (amino acyl-tRNA synthethase) and a tRNA for the non-proteinogenic amino acid is required. Orthogonality means, that the aaRS only amino-acylates tRNA for the non-proteinogenic amino acid and none of the endogenous tRNAs. Furthermore, the aaRS has to specifically amino-acylate its partner tRNA with the non-proteinogenic amino acid. Finally, the non-proteinogenic amino acid has to be available to the cell via transport across the membrane, nontoxic, and stable. Employing these principles, researchers have been able to incorporate more than 200 different non-proteinogenic amino acids into proteins[37].

The applications for genetic code expansion are extremely diverse[38]. Most studies focus on the investigation of specific proteins and are not yet attempting to answer questions about the putative advantages of globally enhancing the design space by supplying new building blocks to the protein expression machinery. A prominent field comprises the incorporation of amino acids with functional groups for bioorthogonal labeling of proteins[39]. The functional groups of the introduced amino acids rarely occur in nature and therefore provide the specificity for selective reactions under physiological conditions conjugating them to interacting proteins or ligands. In general, these studies focus mostly on finding one accessible and functional non-proteinogenic amino acid. Other studies on genetic code expansion applications already include the functional characterization of a variety of non-proteinogenic amino acids, albeit mostly for one specific protein. An early example is the investigation of replacements for a proline residue critical for the ion-channel activity of the 5-Hydroxytryptamine type 3 receptor[40]. Upon replacing the proline with different analogues of similar structure they were able to deduce that the *cis-trans* energy gap is critical for switching between the closed and open state of the ion-channel.

Recent developments which increase the ease of incorporation and number of available non-proteinogenic amino acids[41], complete control over the amber codon placement via genome engineering[42], and the introduction of a new base pair between unnatural nucleotides[43] brings the scientific community close to the genome-wide introduction of non-proteinogenic amino acids[37]. This prospect calls for novel computational methods which classify proteinogenic and non-proteinogenic amino acids into functional categories. They can help to decide which amino acids to prioritize for genome-wide introduction on the basis of how much they might contribute to expanding the design space of available protein functions and structures. These amino acids should not be part of a functional category already occupied by one of the proteinogenic amino acids. To determine the relevant features, we develop a classification scheme which allows for the unique assignment of each proteinogenic amino acid to one functional category defined by repeated median-based separation along the selected features. By choosing the minimal number of features to uniquely group each proteinogenic amino acid, we obtain $5^2 = 32$ functional categories. Twenty of them are occupied by proteinogenic amino acids, twelve are left unoccupied. As the classification approach is developed by separating the proteinogenic amino acids selected by evolution, we deem it likely that non-canonical amino acids which occupy the remaining twelve categories may greatly expand the design space of proteins upon genome-wide introduction.

1.2. Evolutionary alternatives for decompositions of coiled-coil periods

1.2.1. Structural description of protein helices and coiled coils

Protein helices are secondary structures which have in common that they are stabilized by hydrogen bonds between the backbone carbonyl and amine group of non-adjacent amino acid residues. The distance along the sequence between amino acids that participate in hydrogen bonds with each other varies with the type of the helix. Hydrogen bonding amino acids are two residues apart in the 2.2_7-helix, three residues in the 3_{10}-helix, four residues in the α-helix (3.6_{13}-helix), and five residues in the π-helix (4.4_{16}-helix), whose propensity takes part as a binary feature in the optimal classification achieved in the amino acid related part of this thesis. The x_y refers to x for the number of amino acids required to complete one turn and to y for the number of atoms within the backbone enclosed by one hydrogen bond. The most abundant helix in proteins is the α-helix. This helix and in particular the coiled-coil superhelix formed by oligomerization of at least two α-helices will be the focus of this work.

The structural parameters mentioned in the following for the description of the α-helix and coiled-coil geometry are retrieved from the textbook "Fibrous proteins: Structures and mechanisms" by D. A. Parry and J. M. Squire[44]. The axial translation describes the distance covered along the helical axis with each amino acid residue and amounts to 0.15 nm for the α-helix. In combination with the information, there are 3.6 amino acid residues per turn, which amounts to a pitch of 0.54 nm. The pitch describes the distance covered along the helical axis upon performing one full turn and is an important parameter for the structure and stability of the different types of helices. The π-helix for example has a comparatively short pitch of 0.47 nm. This favors the incorporation of aromatic amino acids, as the aromatic rings can come into close distance to each other and engage in strong stacking interactions to further stabilize the helix structure[45]. On the other hand, $C_β$-branched amino acids have been shown to destabilize the α-helix in general due to steric clashes with the backbone, although in some local contexts β-branched amino acids might also stabilize the helix due to reduced entropic costs for participating in Van der Waals interactions[46]. If the α-helix occurs in natural proteins consisting of L-amino acids, it is right-handed.

The α-helix produces a macrodipole nearly parallel to its helical axis with putative contributions to functions such as the binding/long-range attraction of charged residues and enzymatic catalysis[47]. However, the potential effect the electrical field of the macrodipole has on processes such as oligomerization and stability of α-helices seems to have been overestimated. Baker *et al.* showed that energetically favored salt bridge orientations between lysine and glutamate are oriented exactly opposite (namely $K_i \rightarrow E_{i+4}$ preferred over $E_i \rightarrow K_{i+4}$) to what would have been expected if the electrostatic interactions with the macrodipole are of major importance[48]. They, therefore, argue that macrodipole involving explanations of phenomena whose local scale goes beyond the helix termini, where the macrodipole is strongest, should be taken with caution. For example, Parry and Squire showed that the preference for anti-parallel α-helices in coiled coils, which was assumed to be caused by favorable macrodipole interactions, disappears if considering only intermolecular coiled-coils[44]. This hints at sterical demands in intramolecular coiled coils as a more plausible explanation for the preference of anti-parallel α-helix orientation. Also, the hypothesis that the macrodipole limits the length of α-helices to around 40 residues has been rejected by identifying helix lengths of up to 200 amino acids[49]. This is important, as macrodipoles do not preclude the realization of longer decompositions in coiled coils. Therefore, these decompositions remain in the pool of evolutionary alternatives.

The propensity of an amino acid is defined by its abundance in a particular secondary structure divided by its total abundance in proteins. Propensity values in connection with the subsequent structural analysis form the basis to decide which kind of structure stabilizes or destabilizes the secondary structure in hand. From the analysis of α-helix propensities important insights into the structural requirements of helix stability have been obtained[50,51]. First, proline and glycine destabilize the α-helix and are frequently called helix breakers. Proline destabilizes the α-helix because of its secondary amine, which is unable to participate in helix-stabilizing hydrogen bonds and causes strain in the helix backbone. The incorporation of glycine comes with high entropic costs, as the incorporation into the helix restricts the available ϕ- and ψ-angles, which can otherwise assume a wide range of different values in glycine as it lacks a side chain[52]. Important factors for the interpretation of the relative propensities of the other amino acids include the burial of non-polar surface area of the side chain close to the helix backbone (preference of Ile over Val), sufficient chain length for electrostatic interactions with other side chains (Glu over Asp), and again the entropic costs associated with a restriction of conformational space (Ala over Leu).

Coiled coils are supersecondary structures consisting of at least two α-helices wound around each other. They are either called parallel if all the α-helices are aligned with their N-termini on one end of the coiled coil and the C-termini on the other, or anti-parallel if N- and C-termini of different α-helices are located on opposite ends of the coiled coil. The canonical interaction responsible for the stabilization of coiled coils is the knobs-into-holes packing[53]. For this interaction, typically, hydrophobic side chains of one helix are located into a diamond-shaped cavity formed by four hydrophobic side chains of the partner helix. For the standard heptad decomposition those interactions take place alternating between every third and fourth residues. Therefore, on average there is an interaction every 3.5 residues. (Super-)helices are either left-handed (anti-clockwise rotation around the axis) or right-handed (clockwise rotation around the axis). As the period of the α-helix is 3.6, the helices in the coiled-coil have to be wound around each other in a left-handed manner to engage in interactions every 3.5 residues. The canonical heptad motif can be interrupted in sequences by insertion of another interhelical contact after 3 (stammer) or 4 (stutter) residues, which affects the structural parameters of the coiled coil, for example the pitch or the handedness[54]. Non-canonical motifs were already envisaged by Pauling and include the motifs 4/1, 10/3, 11/3, 15/4, and 18/5, where the first number states the period length and the second the number of interhelical contacts[55].

In the following chapter I am going to shed light on how the pattern and frequency of hydrophobic interactions influence the structural dynamics of coiled coils. The aforementioned stammers and stutters allow for different decompositions, *i. e.* sequences of distances between hydrophobic contacts, of coiled coil periodicities for higher period lengths. As an algorithm for the complete identification of all possible decompositions for any period length comprises the core of the publication associated with this section of the introduction, it is important to better understand the impact different decompositions can have on protein functionality associated with coiled coils.

1.2.2. Influential factors for coiled-coil dynamics and applications in protein design

Because of their suitability for experimental methods and their simple heptad motif, coiled coils were in the initial phase of research traditionally associated with fibrous proteins, such as α-keratin and kinesin[44]. In these cases, the opposite handedness of the coiled coil compared to the α-helix and the frequent hydrophobic interactions of the canonical heptad motif provide mechanical stability. Only later the dynamic behavior of coiled coils in connection with molecular recognition became apparent. Well-known examples include leucine zippers as transcription factors[56] and HAMP (histidine kinase and methyl-accepting proteins) domains in signal transduction[57]. Specificities of interaction are often

conveyed by packing geometry and interhelical salt bridges, whereas signal transduction often takes places via axial shifts, tilts, or rotations of the coiled coil[44].

Though the plasticity of coiled coils is certainly higher than initially anticipated, there are limits to the amount of strain coiled coils are able to tolerate. Hartmann *et al.* experimentally confirmed a prediction that 10/3 = 3.33 residues per turn comprises the lower limit of supercoiling a coiled coil can endure[58,59]. Upon introduction of a stammer, the resulting strain causes the formation of a short 3_{10}-helix fragment. Because this lower limit is 0.3 residues per turn lower than the periodicity of a strainless straight helix, they estimate the upper limit to be around 3.9 residues per turn[60]. This reasoning assumes symmetry in the resulting strains caused by over- or underwinding the coiled coil. The highest observed periodicity in a coiled coil so far is 3.8 residues per turn[61]. If the strain is locally confined, as it is the case for a single one or two residue insertion, the strain can be relieved by the formation of a π-turn[62] or β-layer[60], respectively. In general, the periodicity is not necessarily averaged and the resulting strain homogenously distributed over the coiled coil. Instead, coiled coils can consist of separate regions which assume different periodicities and sometimes even switch from left-handed to right-handed or *vice versa*.

The determination of the correct decomposition for a certain periodicity is crucial for differentiation between these cases. For example, the membrane-spanning coiled coil of a transmembrane receptor has been shown to realize the rather complicated decomposition (3 + 4 + 3 + 4) + (4 + 4) instead of the more balanced decomposition (3 + 4 + 4) + (3 + 4 + 4) for the periodicity of 22 / 6 = 3.67 residues per turn[63]. This causes the coiled coil to switch along the helical axis from a slightly left-handed "body" to a shorter, strongly right-handed "neck" segment. Transition regions can also contain characteristic motifs, as it is the case with the YxD motif for right-handed coiled-coil regions in trimeric autotransporter adhesins[61]. The preceding examples already show that a complete enumeration of all possible decompositions for a certain period length or periodicity is critical for the evaluation of the coiled-coil design space.

Because coiled-coils are understood very well by parametric equations and their dynamic behavior in response to different environmental cues can be fine-tuned by residue mutation and/or realization of alternative decompositions, they meet a large interest in protein design. Cysteine and histidine residues of α-helices pointing towards the core of the coiled coil are ideally suited for the complexation of metals and the promotion of metal-catalytic activity. It has been shown that by altering the pitch of the coiled coil, the introduction of a stammer in coiled coils can alter the associated catalytic activities by distorting the local complexation geometry[64]. Coiled coils with cysteine/histidine layers are often capable to bind multiple metals and can therefore in principle catalyze different reactions depending on the bound metal cofactor. Deletion/Insertion of the discontinuities has been used to fine-tune the metal-binding and catalytic profile[65]. Asymmetry opens up the design space of complexation geometry and has for example been employed to boost copper-catalyzed nitrate reductase activity[66]. When applying anti-parallel coiled coils with non-heptad motifs, precise knowledge about how different decompositions of coiled-coil periods affect the geometry is of utmost importance.

Not only the geometry of the coiled coil in its folded state is of interest to protein-design applications, but also its dynamic behavior, switching between the folded and unfolded state upon exposition to different environmental cues. One field of application is the drug delivery by liposomes containing membrane-spanning coiled coils. The idea is to transport drugs with high cytotoxicity in liposomes with relatively low cytotoxicity and release them at the target site by exposure to changing environmental conditions. Reja *et al.* proved the principle of this method by designing coiled coils with environment-sensitive amino acids as part of the hydrophobic core and incorporating them into liposomes[67]. Once these liposomes enter lysosomes and are exposed to a pH lowered from 7.4 to 5.5, the coiled coils

unfold, destabilizing the liposome membrane and thereby causing the leakage of the drug. In a similar study, an increase in temperature was used to trigger doxorubicin release liposomes with leucine zippers anchored to them[68]. Interestingly, the susceptibility of coiled-coil stability to changing pH can be affected by altering the period decomposition. This has been shown for the coiled coil of influenza hemagglutinin HA2[69]. Insertion of a stutter conveys coiled-coil stability over a pH-range of 4.5 to 7.1, whereas coiled coils with the canonical heptad motif depend on low pH for correct folding. The stutter causes a local unwinding which gives two glutamate residues the flexibility to align for proper hydrogen bonding. This hydrogen bond prevents the deprotonation of the glutamate residues upon neutral pH, which causes electrostatic repulsion and according destabilization in heptad coiled coils.

Fig. 1 provides a scheme of how the pool of theoretically possible decompositions provides evolutionary alternatives for the differentiation and fine-tuning of various functions. Notably, as non-heptad decompositions can be achieved by the insertion/deletion/mutation of a single amino acid, the transition between heptad and non-heptad or between alternate non-heptad decompositions can readily take place. On the structural level non-heptad decompositions affect the pitch and accordingly the handedness of the coiled coil, but can also introduce kinks or local unwinding. That many non-heptad decompositions can be found in nature[55] hints at the functional benefits they provide. Some benefits were introduced in the aforementioned sections, including the resilience of coiled-coil dynamics to changing environmental conditions or the realization of different catalytic functions. By developing an algorithm for the complete enumeration of all possible coiled-coil decompositions and comparing them with the experimentally confirmed coiled-coil decompositions, the article associated with this section provides the basis for a systematic assessment of different coiled-coil decompositions and the functional benefits they provide.
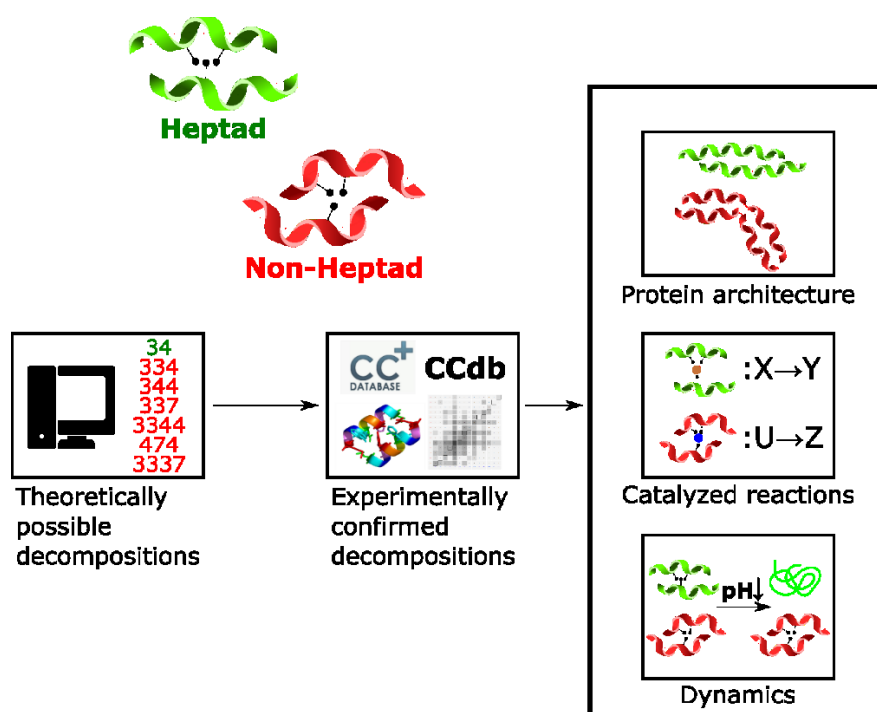

Figure 1: Scheme for the conducted research on alternative decompositions in coiled coils. From the theoretically possible decompositions enumerated by the algorithm developed in the second article, only a subset has been experimentally determined in coiled coils. Based on examples in the literature, alternate non-heptad decompositions can alter coiled coil structure, enable the catalysis of different reactions by altering cofactor binding geometry, or provide enhanced resilience to changes in environmental conditions, such as pH or temperature.

1.3. Evolutionary alternatives for cooperative strategies of iron-cycling bacteria

1.3.1. The study of microbial cooperation through agent-based modeling

Agent-based models (ABMs) are characterized by the rule-based simulation of individuals in a spatial 2D- or 3D-environment. Their implementation was driven by the insight that reductionist approaches like ordinary differential equation models (ODE-models) are either not suitable or excessively complicated to study phenomena emerging as a result of the local interactions of a population of agents[70]. Over the last 20 years, ABMs have been increasingly employed for the simulation of microbial communities[71,72]. They are most suitable for systems with high spatial heterogeneity, agent diversity, stochasticity, and adaptivity in behavior. In the following, I will exemplify the usefulness of ABMs for studying microbiological systems for three research areas which fulfill most if not all of the aforementioned criteria: Chemotaxis, biofilms, and microbial cross-feeding.

Chemotaxis is the non-random movement guided by a concentration gradient of a chemical mediator. As the distribution of these chemical mediators in the environment changes with time, the modeling of chemotaxis is often accompanied by the implementation of one or multiple random diffusion processes. The explicit implementation of diffusion can however quickly become expensive in computing power and might get in conflict with obtaining meaningful results for processes that unfold over larger time scales. In Brownian motion, the mean square displacement of the particle is proportional to time[73]. For small-scale environments associated with microbiological systems, diffusion subroutines have to be called frequently as otherwise there would always be complete equilibration between discrete time steps. The associated computation becomes intractable if the total runtime of the simulation is in the region of hours or days which is often required to assess processes like for example the influence of microbial metabolism on the habitat. ABMs on chemotaxis have to circumvent this problem by limiting the simulation time and thereby the scope of the study, using heuristic approaches instead of explicitly implementing diffusion and/or choosing an environment with high viscosity to lower the diffusion coefficient.

Tokarski *et al.* used an ABM to investigate the influence of different movement strategies of phagocytes on its clearing efficiency on conidia of the opportunistic pathogen *Aspergillus fumigatus*[74]. They were not only able to rank the different strategies in regard to clearing efficiency but also showed how the efficiency of the strategy depends on related parameters such as initial placement of the phagocytes or conidia, and communication thresholds (for chemotaxis along a gradient of chemokines released by other phagocytes). But ABMs can also be used to study the protein signaling network underlying chemotaxis[75]. In this case, the agents are membrane and cytosolic proteins which interact with themselves for oligomerization/ligation and change states upon ligand binding. The model has been developed for hypotheses testing and showed for example that receptor heterogeneity and dependence of coupling strength on receptor methylation increase the dynamic range of the stimulus, *i. e.* the concentration of the ligand, the system can appropriately respond to.

The dynamics of many biofilms is closely linked with chemotaxis. Sweeney *et al.* extended the iDynoMiCS platform, which has been developed for individual-based modeling of biofilms[76], to simulate the dependence of *Helicobacter pylori* biofilm structure on the production and sensing of the quorum-sensing molecule autoinducer-2 (AI2)[77]. By conducting simulations with agents representing wild-type *H. pylori* and mutants deficient in chemotaxis or AI-2 production, they confirmed that the heterogenous surface structure correlates with AI-2 mediated chemorepulsion. This study highlights many advantages of ABMs, as it includes individual heterogeneity, spatial complexity, and the spontaneous emergence of higher-order structures (biofilm structure) by rules guiding local

interactions (chemorepulsion). The last aspect is also illustrated by another study which test different hypotheses about local interactions to assess which are able to reproduce experimental observations on biofilm formation of *Pseudomonas aeruginosa* in microwell arrays[78].

Another interesting research area that benefited from the application of ABMs is the study of metabolic interactions within microbial communities. BacArena for example is a modeling framework which combines flux balance analysis[79] with ABMs to investigate the influence of cross-feeding on the spatial structure of heterogeneous communities[80]. One key connection the authors were able to observe when applying the framework is that the balance between species abundances in the gut microbiome critically depends on the mucus secretion of the host epithelium. Impaired mucus secretion can thereby contribute to inflammatory diseases of the gut. Germerodt *et al.* investigated the formation of local cooperation clusters of mixed microbial communities of wild-type bacteria and different auxotrophs (species deficient in the synthesis of at least one proteinogenic amino acid)[81]. The ABM simulations confirmed the evolutionary stability of local cooperation clusters for nutrient exchange over a broad range of different parameter values. This study is an example of how ABMs can be helpful to test the validity of a hypothesis without requiring parameter values to be exactly known.

All the aforementioned fields, where ABMs have already proven beneficial as a complementary modeling approach, are related to the system under study in the third publication of this thesis. Iron-cycling bacteria are challenged with sensing and migrating towards their non-soluble source of electron acceptors/donators (see subsequent subsection 1.3.2). They adhere to minerals and thereby form spatially dense communities. And finally, as we investigate a system of two species, one reducing iron and one oxidizing iron, they metabolically cooperate by consuming the product of the other species' respiration. In the following, the biological processes will be described which were eventually implemented in an ABM framework.

1.3.2. Principles of iron respiration

When laymen think about respiration they might usually associate it with mammals breathing for air to survive. From a scientific point of view, the process of aerobic respiration requires the transport of electrons along a chain of enzymes via redox reactions until they eventually reduce oxygen to molecular water. Part of the energy generated during this process is used to transport protons across the membrane to establish the gradient. Finally, the backflow of said protons across the membrane and down the proton gradient is coupled to the synthesis of adenosine triphosphate (ATP) via the enzyme ATP synthase. When we further abstract from aerobic respiration to general respiration, we see that the use of many electron acceptors other than oxygen is theoretically possible and to be found in nature.

The maximal energy to be acquired from a respiratory process depends on the difference between the midpoint potential (E_m) of the electron donor and electron acceptor. E_m increases with the tendency of a molecular species to accept electrons. For our system of interest, the relevant values at neutral pH are +0.77 V for Fe(II)/Fe(III) (the reduction of Fe(III) to Fe(II)) and +0.82 for H_2O/O_2 (the reduction of oxygen to molecular water)[82]. This means that upon the usage of the same electron donor, the organism can potentially draw a higher amount of energy from the reduction of O_2 compared to the reduction of Fe(III). The energetic disadvantage is also the reason why Fe(III)-reducing bacteria (FeRB) only fall back to using Fe(III) as an electron acceptor if oxygen is not available anymore, *i. e.* under anoxic conditions.

In contrast to the mostly insoluble Fe(III)-minerals, the product of microbial dissimilatory Fe(III) reduction, namely Fe(II), dissolves to a considerable degree in aqueous media. Fe(II) can be oxidized

back to Fe(III) when used as an electron donor in respiratory processes by Fe(II)-oxidizing bacteria (FeOB). Of course, the corresponding electron acceptor has to have a higher E_m than the Fe(II)/Fe(III) redox pair as the FeOB would otherwise not be able to conserve energy from the associated electron transport processes. The most common electron acceptors in the natural environment matching this requirement are O_2 and NO_3^-, where oxygen is energetically favored[82]. In environments where nitrate is absent, this already hints at how the oxygen requirements differ for FeRB and FeOB. FeRB require oxygen concentrations close to zero, because otherwise they would not conduct Fe(III) reduction, but aerobic respiration instead. In contrast, FeOB require at least some oxygen for Fe(II) oxidation, as oxygen is the terminal electron acceptor of their respiratory chain. This either leads to FeOB and FeRB coexisting in slightly different locations at the aerobic/anerobic interface[83] or at the same location with temporal oscillations of oxygen concentrations[84]. In the latter case the activity phases of Fe(II) oxidation by FeOB and Fe(III) reduction by FeRB switch between each other in accordance with the oxygen concentration of the environment.

Although FeOB benefit from the higher accessibility of Fe(II) in solution, they have to compete with the abiotic, homogeneous oxidation of Fe(II). Druschel *et al.* measured the percentage contribution of biotic Fe(II) oxidation to the total rate of Fe(II) oxidation for the neutrophilic FeOB *Sideroxydans lithotrophicus* under different oxygen concentrations[85]. They showed that the highest contribution (88 %) is achieved for 15 µM O_2 compared to 4 % at 275 µM O_2 (the oxygen concentration of water in equilibration with atmospheric oxygen). They did not observe growth bands of *S. lithotrophicus* above oxygen concentrations of 50 µM O_2. This seems to be the upper limit for competition with the rapid abiotic Fe(II) oxidation at circumneutral pH.

Instead of conducting experiments on different oxygen concentrations, one might wonder why the relative contribution of biotic/abiotic Fe(II) oxidation is not simply determined by a universal rate equation. In their review, Melton *et al.* concisely pointed out the difficulties in formulating such an equation[86]. First, the product of oxidation, Fe(III) oxyhydroxides, provides a catalytic surface for heterogeneous Fe(II) oxidation, thereby increasing the abiotic competition for Fe(II). Furthermore, both biotic and abiotic heterogeneous Fe(II) oxidation depend on the FeOB species. The latter depends on if the Fe(III) oxyhydroxides formed on the surface of the FeOB are embedded in exopolysaccharides as this limits the sorption capacity of catalytic sites for Fe(II). Finally, the composition of the aqueous medium influences the reaction rates by potentially providing impurities upon Fe(III)-oxyhydroxide crystallization which in turn affects the reaction rates. These interdependencies make the straightforward implementation of an ODE-model very cumbersome and support the implementation of a simplified ABM beforehand to identify the relevant processes associated with this system.

The FeRB do not face equally strong competition with abiotic Fe(III)-reduction processes as the FeOB. Nevertheless, the efficiency with which they are able to reduce Fe(III) strongly depends on a number of environmental parameters. One key influence of the biotic Fe(III) reduction rate is the crystallization state of the Fe(III) minerals[87]. Upon comparison of fresh and aged biogenic iron oxides, it has been shown that the main influential factor is not the mineral composition or thermodynamic properties but the morphology of the minerals, especially the specific surface area[88]. Interestingly, the crystallization state is influenced by the FeRB indirectly via Fe(II) produced upon mineral reduction[89]. As the Fe(II) concentration declines with the activity of FeOB, the two species mutually influence each other's reduction/oxidation rates. Another example of this it the production of iron nanoparticles by the FeOB[90]. These nanoparticles provide substrates for FeRB and have an increased surface-area-normalized reduction rate compared to macroparticles[91].

1.3.3. Microbial strategies to face the challenges associated with iron respiration

FeRB face the challenge to locate and move towards an insoluble substrate, namely the Fe(III) minerals. In contrast to regular chemotaxis the insoluble Fe(III) minerals do not provide any concentration gradient by themselves to guide the bacteria through chemotaxis. From microscopic observations, it became apparent that the model FeRB *Shewanella oneidensis* exhibits chemotaxis towards insoluble electron acceptors (IEA) characterized by an increase in speed and flagellar reversal frequency when in close proximity (approximately within 60 µm distance) of the IEA[92,93]. There exist some strong arguments that the chemotaxis is mediated by sensing the electronic state of redox-active compounds[94]. By correlating the applied oxidative potential which causes a switch in chemotactic behavior with the reduction potential of the compound, Oram and Jeuken deem flavins most likely to be the mediators of chemotaxis[95].

Instead of locating and moving towards the IEA, FeRB have also developed strategies for non-contact-dependent Fe(III) reduction. Marsili *et al.* proved the secretion and usage of flavins as electron shuttles by different species of *Shewanella*[96]. Electron shuttles are small organic redox-active compounds which are reduced upon contact with the surface of FeRB, diffuse through the medium, and in turn reduce the Fe(III) mineral upon contact. FeRB do not only benefit from the production of these compounds because they enable long-range electron transfer, but also because the flavins establish redox gradients towards the IEA important for chemotaxis, as mentioned before. The electron shuttles provide an important means for Fe(III) reduction. Experiments suggest that they account for up to 75 % of total Fe(III) reduction by FeRB[97]. Despite of electron shuttles, some FeRB also employ chelators[86]. Instead of transporting the electrons towards the Fe(III) mineral, upon binding they solubilize the Fe(III) and make it thereby accessible for FeRB swimming in the medium.

One of the two key processes for which we tested alternative hypotheses in our ABM is the adhesion of FeRB on the Fe(III) mineral. Adhesion is the prerequisite for contact-dependent Fe(III) reduction. The options discussed for maintaining the adhesion of FeRB are electrostatic interactions and the utilization of pili or polymers to form a biofilm for higher cell densities[98]. Electrostatic interactions have been shown to play a role, as the strength of the adhesion depends on the redox state of MtrC, the protein which mediates the reduction of Fe(III)[99]. Interestingly, with continuous contact-dependent reduction, the charge of the Fe(III) mineral gets more negative, which makes continued adhesion more difficult as the cell surface of the FeRB is also negatively charged[100]. In contrast, the strength of adhesion based on the formation of biofilms should increase with time, as more extracellular polymeric substances (EPS) are produced. This means that two adhesion mechanisms are proposed which differ from each other in fundamental ways, which makes it worthwhile to assess their influence on system variables via mathematical modeling.

The other key strategy concerns encrustation prevention of the FeOB. FeOB have to prevent the excessive precipitation of Fe(III) oxyhydroxides to maintain the exchange of metabolites with their environment. The two main competing hypotheses are shedding of the extracellular mineral sheet or prevention by locally lowering of the extracellular pH of the environment. Schmid *et al.* observed extracellular iron sheets on *Acidovorax* sp. strain BoFeN1[101]. Interestingly, although some of the cells were nearly completely surrounded by the iron sheets, others seemed not to be affected by encrustation at all. This suggests that the microscopic images capture cells at different points of time between right after the shedding of the iron sheets and the peak of cell encrustation. The other hypothesis includes the prevention of cell encrustation by lowering the local pH and thereby directing the precipitation of Fe(III) oxyhydroxides at a distance from the cell surface. This strategy has been observed for *Thiodyction* sp. strain F4[102]. As of now it is unclear which strategy is employed by the model FeOB of our study, *Sideroxydans* spp., we tested the influence of both strategies.

Two main research questions we investigate with the ABM developed as part of the third article are depicted in **Fig. 2**. First, we test the possible combinations of the aforementioned evolutionary strategies for adhesion (FeRB) or the prevention of encrustation (FeOB). Particular interest resides in the question whether the combinations are equally efficient for cooperation between FeRB and FeOB in terms of total iron-cycling and nanoparticle reduction. This is of interest, as for a specific FeRB/FeOB species the employed strategy might be often be unknown. For predicting the dynamics of metals in the natural environment it is therefore of importance to assess the sensitivity of biological redox-processes with respect to the underlying assumptions about microbial strategies. The second question considers the effect of changing pH on both, the macroaggregate iron-cycling and the cycling of iron nanoparticles. When discussing evolutionary alternatives for microbial strategies, it is important to determine within what boundaries they remain stable. FeRB can either adhere to macroaggregate iron minerals, allowing for a relatively constant reduction of Fe(III), or they swim in the liquid medium and reduce nanoparticles which have a much higher surface-normalized reduction rate. The second strategy depends strongly on the production of nanoparticles by FeOB, but also on the aggregation behavior of the nanoparticles themselves. By literature survey we determine the influence of pH on the key processes of our model and perform simulations in order to clarify under what pH efficient reduction of nanoparticles can take place.
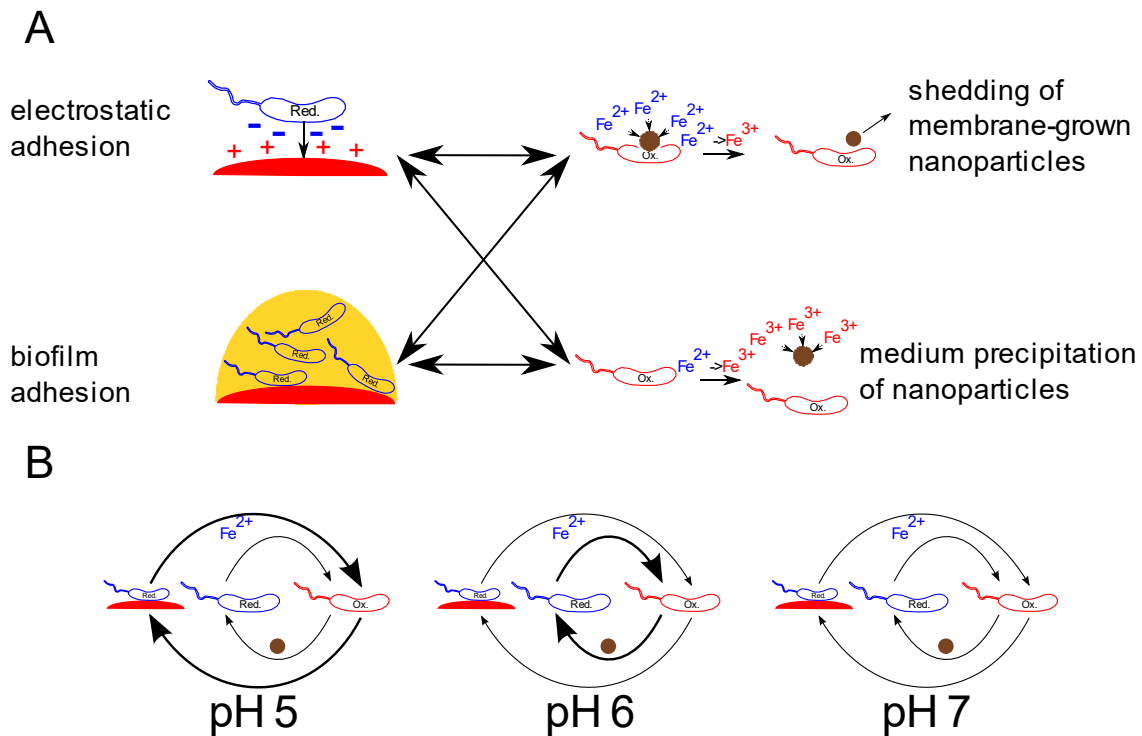

Figure 2: Two research questions about evolutionary alternatives in connection to iron-cycling bacteria. **A** Adhesion of FeRB can be mediated by electrostatic forces or biofilm formation, whereas encrustation of the FeOB cell membrane can be prevented by shedding of membrane-grown nanoparticles or by shifting the precipitation of Fe(III) to the surrounding medium by lowering the local pH around the cell membrane. We test all possible combinations of these strategies. **B** pH influences the viability of cooperative strategies between iron-cycling bacteria. We test the influence of pH on the reduction of macroaggregate Fe(III)-reduction and nanoparticle Fe(III)-reduction.

2. Article 1

A novel method for achieving an optimal classification of the proteinogenic amino acids

André Then, Karel Mácha, Bashar Ibrahim, Stefan Schuster

In this study we develop a novel median-based classification to group the 20 proteinogenic amino acids into distinct categories via separation with the minimum number of five features. We thereby improve Ramsay Taylors' classification, as he used eight features for amino acid classification and was not able to separate some amino acids from each other. Our optimal classification is tightly linked to the questions why exactly these 20 proteinogenic amino acids have been integrated into the genetic code and if and how they are optimal. Whereas previous studies tried to answer these questions be examining the coverage of chemical space of the proteinogenic amino acid set and compared it with the coverage of random sets of prebiotic amino acids, we argue that it might be sufficient that amino acids occupy different functional niches although these niches might be located in close proximity in chemical space. Leucine and isoleucine provide an excellent example. In four of the five features we found to enable optimal classification, they have very similar values. But in contrast to isoleucine, leucine frequently occurs in α-helices. Thereby the encoding of both amino acids can be justified, as they may contribute their similar physicochemical characteristics in different structural environments. Finally, we demonstrate the applicability of our classification approach to other evolutionary problems, as it also allows for ranking the amino acids in terms of how redundant their physicochemical contribution to the set is.

Manuskript Nr. 1

Titel des Manuskriptes: A novel method for achieving an optimal classification of the proteinogenic amino acids

Autoren: André Then, Karel Mácha, Bashar Ibrahim, Stefan Schuster

Bibliographische Informationen: Then, A., Mácha, K., Ibrahim, B., & Schuster, S. (2020). A novel method for achieving an optimal classification of the proteinogenic amino acids. *Scientific reports*, *10*(1), 1-11.

Der Kandidat / Die Kandidatin ist

☐ Erstautor/-in, [X] Ko-Erstautor/-in, ☐ Korresp. Autor/-in, ☐ Koautor/-in.

Status: Publiziert

Anteile (in %) der Autoren / der Autorinnen an den vorgegebenen Kategorien der Publikation

Autor/-in	Konzeptionell	Datenanalyse	Experimentell	Verfassen des Manuskriptes	Bereitstellung von Material
A. Then	30	50	75	75	-
K. Mácha	20	25	25	0	-
B. Ibrahim	20	15	0	10	-
S. Schuster	30	10	0	15	-
Summe:	100 %	100 %	100 %	100 %	100 %

3.Article 2

Bioinformatics analysis of the periodicity in proteins with coiled-coil structure - Enumerating all decompositions of sequence periods

André Then, Haotian Zhang, Bashar Ibrahim, Stefan Schuster

Under review

In the second article we investigate the design space of coiled coil structures in respect to the decomposition of core-contact residues. For this purpose, we develop a recursive algorithm to enumerate all decompositions for any period length. By comparing the output with the decompositions actually found in databases of experimentally determined coiled-coil structures, we discuss the limitations which narrow down the design space. Furthermore, coiled coils with decompositions deviating from the heptad repeat are examined and the putative functional advantages they provide are discussed. The study points to further interesting research questions, for example what determines the distribution of local periodicity along the coiled-coil structure. From our observations, the decompositions alone do not allow for sufficient prediction of these distributions.

Manuskript Nr. 2

Titel des Manuskriptes: Bioinformatics analysis of the periodicity in proteins with coiled-coil structure - Enumerating all decompositions of sequence periods

Autoren: André Then, Haotian Zhang, Bashar Ibrahim, Stefan Schuster

Bibliographische Informationen: -

Der Kandidat / Die Kandidatin ist

[X] Erstautor/-in, ☐ Ko-Erstautor/-in, ☐ Korresp. Autor/-in, ☐ Koautor/-in.

Status: Zur Publikation eingereicht

Anteile (in %) der Autoren / der Autorinnen an den vorgegebenen Kategorien der Publikation

Autor/-in	Konzeptionell	Datenanalyse	Experimentell	Verfassen des Manuskriptes	Bereitstellung von Material
A. Then	60	90	100	75	-
H. Zhang	0	10	0	0	-
B. Ibrahim	10	0	0	5	-
S. Schuster	30	0	0	20	-
Summe:	100 %	100 %	100 %	100 %	100 %

4. Article 3

Agent-based modeling of iron cycling bacteria provides a framework for testing alternative environmental conditions and modes of action

André Then, Jan Ewald, Natalie Söllner, Rebecca E. Cooper, Kirsten Küsel, Bashar Ibrahim, Stefan Schuster

The third article uses an agent-based modeling framework to investigate the conditions under which iron reducing/oxidizing bacteria interact to maximize the throughput of iron between its redox states Fe(II) and Fe(III). The key processes are carefully selected and parameterized from the literature. Most experimental studies deal with a single species under specific environmental conditions with emphasis on a single system variable. As several more recent studies hint at the importance of the interaction between FeRB and FeOB, we investigate how their interactions depend on environmental parameters or assumptions about their strategies. One important aspect are the nanoparticles produced by the FeOB, as they have shown to be reduced much faster by FeRB than macroaggregates.

Manuskript Nr. 3

Titel des Manuskriptes: Agent-based modelling of iron cycling bacteria provides a framework for testing alternative environmental conditions and modes of action

Autoren: André Then, Jan Ewald, Natalie Söllner, Rebecca E. Cooper, Kirsten Küsel, Bashar Ibrahim, Stefan Schuster

Bibliographische Informationen: Then, A., Ewald, J., Söllner, N., Cooper, R. E., Küsel, K., Ibrahim, B., & Schuster, S. (2022). Agent-based modelling of iron cycling bacteria provides a framework for testing alternative environmental conditions and modes of action. *Royal Society Open Science*, *9*(5), 211553.

Der Kandidat / Die Kandidatin ist

[X] Erstautor/-in, ☐ Ko-Erstautor/-in, ☐ Korresp. Autor/-in, ☐ Koautor/-in.

Status: Publiziert

Anteile (in %) der Autoren / der Autorinnen an den vorgegebenen Kategorien der Publikation

Autor/-in	Konzeptionell	Datenanalyse	Experimentell	Verfassen des Manuskriptes	Bereitstellung von Material
A. Then	70	85	85	75	-
J. Ewald	5	10	0	0	-
N. Söllner	0	0	10	0	-
R. E. Cooper	0	0	5	5	-
K. Küsel	5	0	0	10	-
B. Ibrahim	10	5	0	5	-
S. Schuster	10	0	0	5	-
Summe:	100 %	100 %	100 %	100 %	100 %

5.Discussion

5.1. Applying the optimal classification scheme to non-canonical amino acids

In the first article we developed a classification scheme to uniquely separate the 20 proteinogenic amino acids into functional categories by median-based separation along a minimal number of five features. Candidate features were retrieved from the AAindex database[103,104]. Just running our algorithm on the complete set of features would have led to a high number of solutions, with most of them consisting of features not easily accessible to interpretation and probably also of little biological significance. We therefore manually selected features for inclusion in our search based on the following criteria: Interpretability, relatedness to biological contexts, and determinability for non-proteinogenic amino acids.

The criterion of interpretability has been chosen because we assumed that concepts which are thoroughly discussed in science, and therefore are known to many scholars in biological sciences, are of higher relevance for the description of amino acid roles. The rule of thumb for this criterion is related to our experience in teaching, notably whether undergraduate students get acquainted with the feature during introductory courses. The criterion of relatedness to biological contexts excludes features such as solubility in organic solvents, which are easily interpretable but unlikely to have played a role in the selection during the course of evolution. The determinability for non-proteinogenic amino acids is a requirement for the application of our classification scheme to synthetic biology.

The identified eighteen different solutions for median-based, optimal classification of the 20 proteinogenic amino acids are highly redundant in terms of the features they employ. The features electron-ion interaction potential, α-helix propensity, and π-helix propensity are part of every solution, whereas the two remaining features volume and hydrophobicity can only be replaced by highly similar features. To give an example: Hydrophobicity can be replaced with the percentage of buried residues in proteins, but this feature is of course strongly correlated with hydrophobicity. In the following, I will shed some light on the definition of the five different features and how they are measured. As mentioned in the Introduction, one particular point of interest is the classification of non-proteinogenic amino acids for giving an estimate of the functional/structural benefits of genetic code expansion.

The feature 'volume' is here defined as the side chain volume minus the constant peptide volume in $Å^3$ [105,106]. The constant peptide volume is the glycine residue volume minus the estimated glycine side-chain volume of 3 $Å^3$. The volume can easily be calculated for every non-proteinogenic amino acid. Hydrophobicity has been determined by measuring the energy involved in the transfer of the free amino acid between water and oil[107,108]. Therefore, in contrast to the other four features, low values correspond to a high hydrophobicity. This led to a mistake in the Venn diagram **Fig. 4** of our publication. The yellow set correspond to high values for the free energy for the transfer of the amino acids from water to oil and hence represents the hydrophilic amino acids. Hydrophobic interactions reduce the non-polar surface area in contact with water. To calculate hydrophobicity by this means in a strict sense requires experimental measurements. For deciding whether to group a non-proteinogenic amino acid into the hydrophobic or hydrophilic category, computational octanol-water partition coefficients like the XLOGP3-AA should be sufficient though[109].

The electron-ion interaction potential is a combination of the two quantities Z* and W[110]. Z* is the "average quasi-valence number" and is completely determined by the chemical composition of the

compound. For some atoms, the quasi-valence number equals the ordinary valence number, whereas for others the values deviate without the authors giving a theoretical justification. The second part, W, is the calculated energy of a charge brought within 5 Å of the investigated compound. The method for the derivation of the final formula involves complex equations but has been criticized by Linus Pauling as being inaccurate and incompletely described[111]. At the time the method has been published, accurate calculations of electrostatic interaction energies would have been too difficult to compute, which may justify the usage of a less accurate pseudo-potential. However, nowadays these kind of computations are standard in molecular mechanics calculations and can be easily applied to non-proteinogenic amino acids[112]. In conclusion, it is possible to determine the electron-ion interaction potential also for non-proteinogenic amino acids.

The α-helix propensity gives the empirical preference for amino acids to occur in an α-helix sequence compared to the relative abundance of the amino acid in protein sequences in general[113]. As the α-helix propensity is calculated from large sequence/structural databases, strictly it can only be calculated for proteinogenic amino acids. However, it is possible to take advantage of the fact that for functional characterization within our classification scheme it is sufficient to assign the non-proteinogenic amino acid either to the left or right side of the median. Horovitz *et al.* discuss structural features which contribute to the helix-stabilizing or -destabilizing tendencies of amino acids[114]. Proline is a helix-breaker because its backbone nitrogen is part of the pyrrolidine ring and can therefore not participate in hydrogen bonds. Although glycine is a helix-breaker as well, the reason differs. Here the conformational restriction of the main chain in a helix compared to the random coil is unfavorable in entropic terms. Entropy also plays a role in the helix-destabilizing effect of C_β-branched amino acids, as the dihedral torsion involving the branch along the side chain is restricted to the *trans* conformation. For linear side chains or side chains that branch farther away from the backbone, the conformation is not restricted in the same manner. Although conformational preferences of non-proteinogenic amino acids have been used to identify helix-stabilizers in the Noncoded Amino acids Database in specific cases[115], there is no straight-forward connection between conformational preferences and helix propensity. Other effects, such as the local environment of the helix and the possibility of non-covalent interactions between residues of subsequent turns also play a role. Structural features of non-proteinogenic amino acids might therefore only be sufficient to group them left or right of the median for α-helices in clear-cut cases, as for example (a-methyl)-alanine (strong conformational preference for fitting into a helix backbone) and 1-aminocyclopropanecarboxylic acid (similar to proline)[116].

The assignment of π-helix propensity to non-proteinogenic amino acids faces similar problems. Further complications arise due to the small π-helix sample size in comparison to α-helices and the deviation of real-world π-helix parameters from the canonical case[117]. Both issues are related to the fact that hydrogen bonds are formed between residue *i* and *i*+5, instead of *i* and *i*+4 for the α-helix. This results in the formation of an energetically unfavorable central helical hole with a diameter of approximately 1 Å. This hole is too small to allow the occupation of water molecules. As a consequence, the canonical helix is not formed in real-world proteins. Instead, irregular backbone conformations enable the stabilization of this structure by cross-core van der Waals interactions. These irregular backbone conformations make it even more difficult to assign π-helix propensity values based on the conformational preferences of the amino acids than for α-helices. Fortunately, some structure-propensity connections were observed based on the evaluation of the available π-helices[45]. The authors found that large and aromatic amino acids are generally preferred over smaller ones[45]. As the side chains of hydrogen bonding residues (*i* and *i*+5) are located close to each other due to the small unit rise of the π-helix (only 1.2 Å), these kinds of residues enable energetically favorable Van der Waals or aromatic interactions. As most π-helices are rather short containing 7 to 13 residues, polar residues like Asn, Glu, Ser, and Thr are preferred in positions where they can stabilize backbone carbonyl groups

which have no partner to form main-chain hydrogen bonds. Again, as we only aim to group non-proteinogenic amino acids left or right of the median in terms of π-helix propensity, this information might already be sufficient for functional categorization in some cases.

With our method of classification, each of the 20 proteinogenic amino acids is functionally categorized into a unique role. In **Fig. 3** we visualize the introduction of new functional roles over the course of natural genetic code expansion. For this, we assume the consensus order of amino acid appearance as described by Trifonov[18,118].
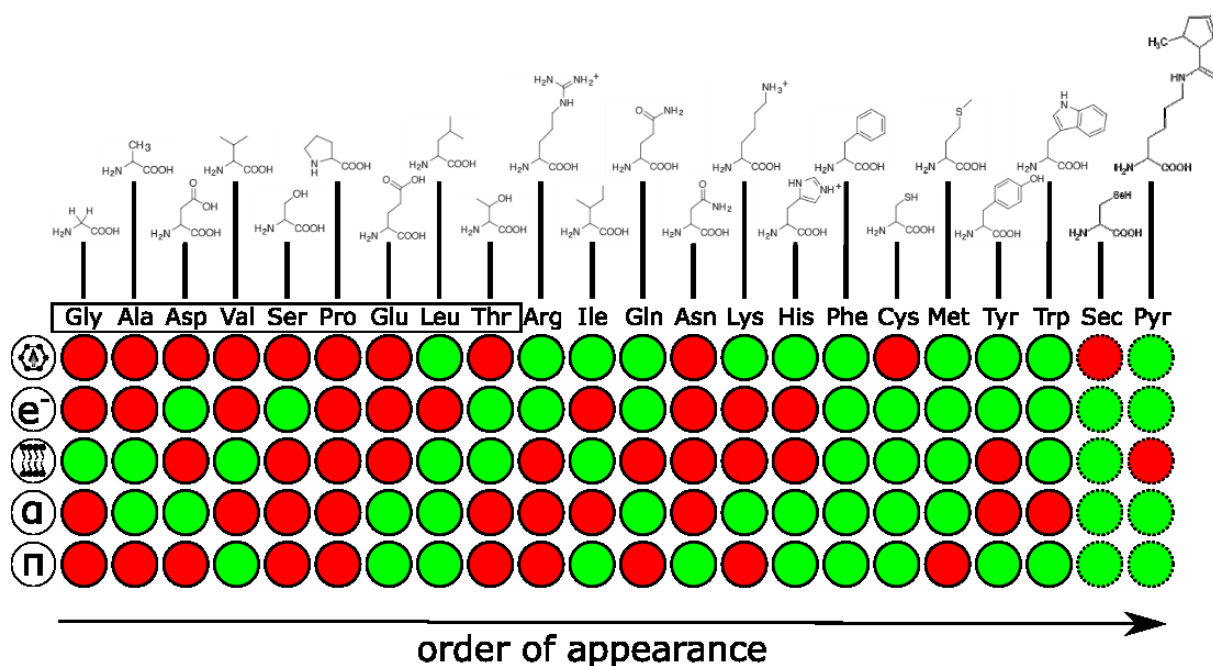
Figure 3: Functional classification of the 20 + 2 proteinogenic amino acids according to our classification scheme. The features are from top to bottom: Volume, electron-ion interaction potential, hydrophobicity, α-helix propensity, and π-helix propensity. A green circle signifies that the corresponding amino acid exceeds the median and possesses the corresponding characteristic, *e. g.* large for volume or hydrophobic for hydrophobicity. A red circle means that the amino acid lacks the characteristic. The boxed amino acids are the consensus early amino acids[17], available at the onset of life via prebiotic chemistry. For selenocysteine and pyrrolysine the feature values are not accessible in the AAindex database. See the main text for the assignment of these two amino acids into groups.

The characteristics of large, electrostatically interactive, and π-helix abundant are mostly present among amino acids which comprise late additions to the genetic code. This is in accordance with the ideas summarized in the Introduction: Small/simple amino acids which were available in high abundance at the onset of life form a core set, which is sufficient for the synthesis of foldable proteins. The more complex amino acids, which provide catalytic functionality by being more electrostatically interactive and flexible in terms of a longer side chain, were introduced later. As discussed earlier, there is a connection between the size and aromaticity (which in turn requires a certain minimum size) of the amino acid side chains and the π-helix propensity. It is therefore not surprising that also most π-helix abundant amino acids are late additions to the genetic code.

For applications in synthetic biology, it would be interesting to group also non-proteinogenic amino acids into functional categories. The non-proteinogenic amino acids categorized into hitherto unoccupied functional categories can be expected to open up new possibilities for the structural evolution and design of proteins. Although the AAindex database does not provide the feature values for amino acids beyond the 20 proteinogenic ones, taking selenocysteine and pyrrolysine as examples, I will show that an assignment to functional categories is feasible. The categorization on the features size and electron-ion interaction potential is straight-forward, as the exact values can either be found

in other databases (size) or calculated by using the provided equation (electron-ion interaction potential)[119]. For deciding whether selenocysteine and pyrrolysine are hydrophobic or not, we can either conduct experiments on the octanol/water partition coefficient or simply compute the coefficient via methods like the XLOGP3-AA, which correlate well with experimental data[109]. The difficult part is the decision of whether the two amino acids have a high α-/π-helix propensity or not. As the definition of propensity prevents the exact determination of this feature value for amino acids beyond the 20 proteinogenic ones, we have to rely on the connections observed between structure and propensity. As selenocysteine is nearly identical in structure to cysteine we can assume that they are located on the same side of the median (high helix propensities). However, as cysteine itself is close to the median in terms of the two propensities, this assignment has to be taken with caution. Serine is also very similar to selenocysteine and is grouped in both cases on the other side of the median (low helix propensities). As pyrrolysine is not C_β-branched and the bulky ring is located far away from the backbone, we might assume that it is α-helix abundant. As stated previously larger, aromatic groups tend towards higher π-helix propensities. The long side-chain of pyrrolysine might stabilize the π-helix via Van der Waals interactions and should therefore be grouped towards high π-helix propensities.

In conclusion, this shows that our classification scheme is in principle extendable towards non-proteinogenic amino acids. For the helix propensities, experiments which incorporate the candidates for genetic code expansion into helices and measure/model the changes in stability are recommended to increase the certainty of classification. Although pyrrolysine falls into a functional category unoccupied by any of the 20 proteinogenic amino acids, this might not necessarily be the main reason why it has been added to the genetic code in some organisms. In the Introduction, it was explicated that both, selenocysteine and pyrrolysine, are required for specific catalytic functions. Nevertheless, as the development in synthetic biology proceeds, scientists will put more emphasis on the question of which non-proteinogenic amino acids to choose for genetic code addition to expand the structural and functional possibilities of protein design.

5.2. Drawing a connection between coiled-coil decompositions and structure

In the publication "Bioinformatics analysis of the periodicity in proteins with coiled-coil structure – Enumerating all decompositions of sequence periods" I developed an algorithm to assess the theoretically possible design space for coiled-coil periods in terms of their possible decompositions. Two mistakes were discovered after publication. First, when discussing the mass fragmentation problem in the mathematical background of the methods, x_i on the left side of the equation should be replaced by $x[m]$. Second, when giving an example for possible decompositions in the methods section, the decompositions [3, 3, 4], [3, 4, 3] and [4, 3, 3] correspond to a period length of 10 instead of 11.

Upon comparison with experimentally confirmed coiled coils from the CC+ database[120], we were able to draw connections between the distributions of theoretically possible and experimentally confirmed coiled coils over increasing period lengths. We retrieved experimentally confirmed coiled coils in agreement with our assumptions on allowed hydrophobic distances whose helix lengths range from 7 up to 140. The minimum helix length is assumed to be 5, which is the lowest length allowing for the formation of an intrahelical hydrogen bond between residue i and i+4[121].

However, this helix length is not visible in our analyses as its period cannot be decomposed with any combination of the allowed interhelical contact residue distances of 3, 4, and 7. This simplification accurately represents around 88 % of the dimeric coiled coils in the database. We showed that the

most common of the interhelical contact residue distances different from 3, 4, and 7 are 1, 10, and 11. Localized one residue insertions can be accommodated by the formation of a wider π-turn[62]. Hydrophobic distances of 10 and 11 still exhibit tolerated periodicities of 10/3 = 3.33 and 11/3 = 3.67. They are expected to be accompanied by lower hydrophobic distances though, as they exhibit a low density of hydrophobic distances which are required for coiled-coil stabilization.

Another key observation was the oscillating pattern for the number of decompositions realized in experimentally confirmed coiled coils for increasing period lengths. This was not expected based on the number of theoretically possible decompositions alone, as they were found to be generally increasing with higher period lengths, although there are a few exceptions (for example there are fewer theoretically possible decompositions for a period length of 19 than for 18). Instead, if we canceled the theoretical decompositions which contain two or more insertions of stammers or stutters, we retrieved an oscillatory pattern that closely resembles the one for the naturally occurring decompositions. This is in accordance with the fact that we did not observe any natural decompositions containing such insertion of two or more stammers/stutters in the analyzed dimeric coiled coils.

By taking a closer look at the strain applied to the coiled coil, the symmetry of stammer/stutter repeats does not hold true. The periodicity of stammers (3) deviates stronger from the proposed lower boundary of possible coiled-coil periodicities (3.33) than the periodicity of stutters (4) from the proposed upper boundary of possible coiled-coil periodicities (3.9)[122]. Accordingly, Hartmann *et al.* showed that the insertion of two or more stammers leads to the formation of α/β-coiled coils as the strain breaks the helices into short β-strands[60]. In contrast, the same author identified a 3 + 4 + 4 + 4 decomposition in a coiled-coil tetramer without the disruption of coiled-coil structure[63]. This shows that rather than the number of stammer/stutter repeats, the average periodicity of decomposition should be regarded as an approximation to predict if it can be realized.

The average periodicity of a decomposition might be a good predictor of whether the decomposition can be realized, but without further information, it is not sufficient to predict the coiled-coil geometry. Experimental examples illustrate this. For a transmembrane coiled coil, it has been shown that the coiled coil is split into two separate regions, a longer left-handed region and a shorter, strongly right-handed region[63]. The total decomposition is 3 + 4 + 3 + 4 + 4 + 4 and has the same periodicity as the hendecad motif 3 + 4 + 4, namely 11/3 = 3.67. If one only takes into account the periodicity, both coiled coils should be nearly parallel, as has been shown for the surface layer protein tetrabrachion[123]. However, from this example and our analysis on non-heptad coiled coils it became apparent that the periodicity is not necessarily homogenously distributed along with the coiled coil. Rather the coiled coil can split into regions with different periodicities.

An interesting research topic to pin down the design space associated with different decompositions is to determine which separations in regions with distinct periodicity are possible for a given decomposition. It might be difficult to find hard criteria for the exclusion of certain separations though. The aforementioned example with the (3 + 4 + 3 + 4) + (4 + 4) separation showed that even regions which have periodicities above the proposed upper boundary, namely the short strongly right-handed segment, can exist. We were also able to identify (3 + 3) segments that are below to proposed lower boundary for allowed coiled-coil periodicities. Interestingly, upon geometric analysis, these segments showed local periodicities higher than both the theoretical periodicity of 3 and the lower boundary periodicity of 3.33. This is only possible for very short sections and probably leads to distorted knobs-into-holes packing, which might be on the edge of what SOCKET tolerates in the identification of hydrophobic interactions. An in-depth structural analysis to see what kind of interactions besides the hydrophobic contacts between helices stabilize these coiled-coil geometries would be insightful.

Another important research question emerging from our analysis of coiled-coil decompositions is what sequence patterns allow us to predict the realization of a particular segmentation from the possible alternatives. Alvarez *et al.* showed that in YadA, the adhesin of enteropathogenic *Yersiniae*, right-handed segments of the coiled coil are accompanied by the sequence motif YxD, whereas left-handed segments often show the motif RxD[61]. The further stabilization to the coiled-coil superhelix through additional hydrogen bonds between the tyrosine and aspartate residues of the YxD motif is only geometrically possible in right-handed coiled coils. These motifs are therefore likely to be part of a right-handed coiled-coil segment. Analogous reasoning applies to the RxD motif for left-handed segments of a coiled-coil.

I suggest a data-driven approach to identify additional sequence motifs correlated with specific segments or transition regions in-between segments. For this purpose, our algorithm can easily be extended to enumerate all possible segmentations for each decomposition. The distributions of local periodicities of different segmentations are then fitted to the corresponding experimental data available at CCdb[124]. The sequences are then segmented and further analyzed to identify characteristic patterns. This approach might be limited though by the influence of the protein structure environment surrounding the coiled coil.

The coronavirus spike glycoprotein (PDB-ID: 6B3O)[125] contains a coiled coil from residues 973 to 1062 and might be a reasonable example for studying such effects (**Fig. 4**). Interestingly the decomposition, as analyzed with SOCKET2 (cut-off distance of 7.6 Å)[126] is symmetric: (4 + 7 + 4 + 7 + 7 + 7 + 7 + 4 + 7 + 7 + 7 + 7) + (4 + 7 + 4). The brackets indicate the long left-handed segment of the coiled coil and the short right-handed segment with the pentadecad motif, respectively. Although in principal the asymmetric segmentation of a symmetric decomposition might be due to differences in amino-acid sequence, the difference in protein environment, in this case, is very distinctive. The C-terminus of the coiled coil is surrounded by a clamp and engages in a multitude of different interactions, whereas the only potential interaction partners at the N-terminus are three surrounding α-helices. The fact, that the right-handed segment begins as soon as the coiled coil enters the clamp environment hints at the potential role the interactions might play for enforcing the right-handed orientation.
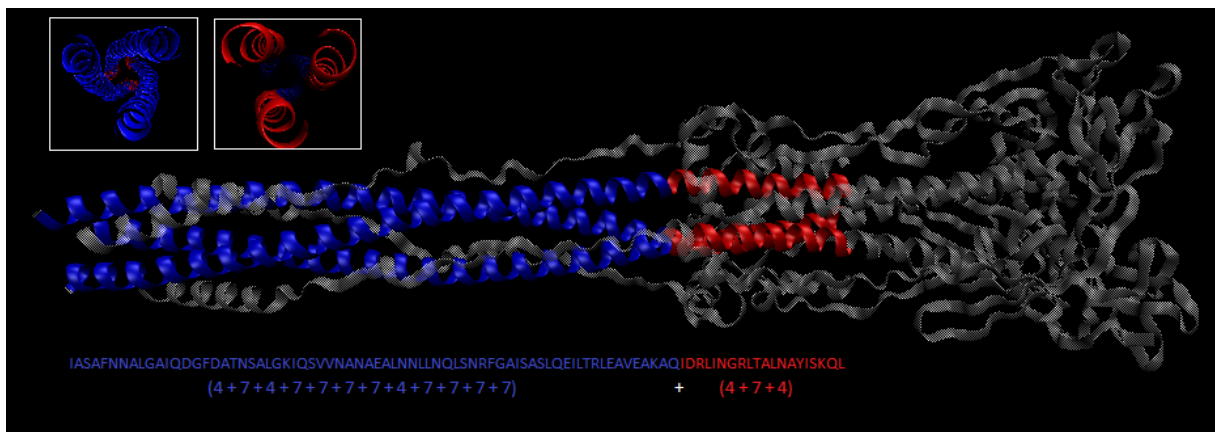

Figure 4: Coronavirus spike glycoprotein (PDB-ID: 6B3O). The left-handed segment of the coiled coil is colored blue, whereas the right-handed segment is colored red. The remainder of the protein structure is transparently white. The inlets highlight the different orientations of the two segments. Beneath the protein structure, the amino acid sequence (residue 972 to 1062) and decomposition of the coiled coil are depicted.

5.3. Assessment of the preconditions and implications for different strategies of iron-cycling bacteria

In our ABM we integrated and carefully parameterized the key processes for describing the interplay between FeRB and FeOB. As the Introduction already pointed out the astonishing complexity of this system, especially when studied in their natural environment, the scope of this pioneering model is not to give highly accurate, quantitative predictions for a broad range of different species and environmental conditions. Instead, we deliver an ABM for testing hypotheses about the underlying mechanisms for the emergence of characteristic patterns observed in experiments. Furthermore, the model provides the option to test the influence of the implemented processes on the outcome of any variable of interest via sensitivity analysis. This will be helpful in the implementation of minimal models, which describe parts of the system with higher accuracy for prediction. These minimal models can build on the information provided by the sensitivity analysis of our ABM to neglect the processes which have been shown to have little influence on the variables of interest.

In the manuscript, we provide an example for testing hypotheses for the mechanisms underlying an experimental pattern. The example consists of an observation made by Barcellos *et al.*, which states that high frequencies (60 h anoxic, 10 h oxic) in oscillations of the environment between oxic and anoxic states of the environment cause an increase in the rate of Fe(III) reduction[127]. For low frequencies (240 h anoxic, 40 h oxic) no such increase has been observed. The authors hypothesize that the frequent recurrence of anoxic phases with active Fe(III) reduction prevents the formation of secondary minerals with higher crystallinity. This is supported by the findings of Mejia *et al.*, who observed a higher abundance of magnetite (high crystallinity) compared to lepidocrocite and ferrihydrite (low crystallinity) when the Fe(III) reduction activity of FeRB is lowered by the availability of the alternative electron acceptor nitrate[128].

In our model, we tested if the linear decline in reduction susceptibility due to a prolonged break from microbial Fe(III) reduction, which prevents transition into a secondary mineral with higher crystallinity, is sufficient to explain the experimentally observed pattern. We indeed were able to show that the implementation of this mechanism causes a stronger increase in Fe(III) reduction rate for high frequencies compared to low frequencies. That we were able to reproduce this phenomenon is especially remarkable in the view that it was not required to integrate more detailed information. A specialized ODE-model focusing on this particular process integrated growth kinetics, physical protection against oxidation for a percentage of Fe(II), and initial distribution of Fe(III) between states of high and low reduction susceptibility to observe the same pattern[129]. This finding supports the case for the complementary of ABMs. ODE-models might be more successful for accurate prediction if fitted to experimental data, but ABMs can often provide a faster way for the reproduction of emergent system patterns and the testing of underlying mechanisms.

We then went from the testing of hypotheses to the comparison of the system behavior for different pH values. Computational simulations of this kind are of importance, as there is a large variety of environmental conditions FeRB/FeOB species can drive in. By carefully reviewing the literature, we identified the key processes affected by pH. First, the abiotic oxidation rate declines with decreasing pH. The E_m of the redox couple O_2/H_2O increases with decreasing pH, driving the oxidation reaction[82]. However, this affects both, abiotic and biotic oxidation. The reason why eventually biotic oxidation can profit more from this change is assumed to be the shift at low pH from ferrous hydroxides towards ferrous ions which are less susceptible to abiotic oxidation[130,131]. Second, via influencing the charge of Fe(III) macroaggregates and nanoparticles, pH affects the adhesion of bacteria to these substrates. A previous study on the adhesion of *Shewanella putrefaciens* on ferrihydrite has observed a decline of

bacteria attached to the mineral from 80 to 20 % upon increasing the pH from 5 to 6[98]. A continued increase in pH does not result in a further decline of the percentage of attached bacteria. Finally, for similar electrostatic reasons, pH also affects the aggregation of nanoparticles. Here the decline of aggregation due to electrostatic repulsion with decreasing pH is extreme, comprising two orders of magnitude upon a drop of pH from 7 to 5[100].

Again, our focus was less on making extremely accurate, quantitative predictions and more on discovering patterns of general system behavior. We found for example that the efficient reduction of nanoparticles seems to be constricted to a narrow range around a pH of 6. At lower pH, the adhesion between the FeRB and the macroaggregates becomes so strong, that basically none of the FeRB are left swimming in the medium. This contrasts the oscillations of the number of attached FeRB upon switching between the oxic and anoxic phase in simulations of higher pH. In their experimental study Roberts *et al.* observed that even under adjustment of pH to circumneutral values, the adhesive forces are too strong to be overcome by environmental conditions driving the bacteria towards detachment[98]. Using or ABM for testing alternative hypotheses on the underlying mechanisms of this observation might provide new insights in the future. The strong adhesion is probably the reason why nanoparticle reduction is not working efficiently at lower pH. The number of nanoparticles attached to the surface of the FeRB is reduced, which makes sense, as attached FeRB have less surface area available for nanoparticle adsorption.

Higher pH prevents the efficient reduction of nanoparticles for different reasons. Here the strong aggregation leads to a declining number in nanoparticles. The nanoparticles are on average larger, but the total surface area critical for Fe(III) reduction is declining in comparison to a higher number of small nanoparticles. Therefore, under conditions of high pH nanoparticle reduction is slowed down. Interestingly, we observed that the beneficial effect of nanoparticles on the efficiency of iron cycling is not only due to their higher surface-area-normalized reduction rate. We also observed that the buildup of a nanoparticle pool with their associated high surface areas functions as a sink for ferrous ion (Fe^{2+}). Upon contact-dependent reduction of the Fe(III) macroaggregates, FeRB passivate the surface of the macroaggregate they are adhered to with ferrous ion. This leads to a decline in the local Fe(III) reduction rate and might eventually lead to the detachment of the FeRB due to electrostatic repulsion. However, as ferrous ion is soluble it equilibrates with the solution and other available surfaces in the environment which allows for its adhesion. If nanoparticles provide a high surface area for ferrous ion, this ultimately leads to larger periods of uninterrupted attachment and thereby to a more efficient Fe(III) reduction of FeRB. Our findings also explain the catalytic effect of nanoparticles on FeRB macroaggregate reduction Bosch *et al.* identified[132]. In addition to the kinetic enhancement, also a three to four times increase in maximum reduction of macroaggregates was detected. This agrees with the function of nanoparticles as a sink for ferrous ions due to prolonged reduction phases upon attachment.

To lay the path for further improving the accuracy of predictions it is important to discuss some of the uncertainties associated with the behavior of nanoparticles in connection with FeRB and FeOB. First, although or model assumes a positive correlation between available surface area of nanoparticles and their rate of reduction by FeRB, experimental studies show that this might only be an accurate prediction for certain range of nanoparticle sizes[91,133]. Both studies show that for very small nanoparticles (< 10 nm in diameter) the surface-area-normalized reduction rate declines, whereas comparatively large nanoparticles (> 100 nm in diameter) have an unexpectedly high surface-area-normalized reduction rate. The putative reasons for this behavior are still discussed, though it seems likely that multiple factors contribute. From the thermodynamic point of view, it seems that narrower band gaps between the O 2p and the Fe 3d orbitals of larger nanoparticles lead to a higher electron

affinity[134]. In conclusion, larger nanoparticles are stronger electron acceptors which cause an increase in the FeRB Fe(III) reduction kinetics.

In our model, we assume for reasons of simplicity that nanoparticles are perfectly round spheres and upon aggregation combine into a new, proportionally larger round sphere. In reality, however, the shape of nanoparticles and their aggregates varies. Associated parameters like the density of packing and the size of cavities have been shown to be influential factors for the microbial Fe(III) reduction rate[91]. However, these parameters are not easily predicted and depend on a multitude of factors themselves, such as salinity, pH, and content of natural organic matter[100]. Natural organic matter for example influences not only the size and shape of the nanoparticles but often coprecipitates with the nanoparticles into a mixed aggregate. This hampers discerning whether natural organic matter exhibits a positive or negative effect on microbial Fe(III) reduction rate. One study about the reduction of alginate-hematite composites by *Shewanella oneidensis* MR-1 shows that the initial Fe(III) reduction rate is reduced compared to pure hematite[135]. It has been proposed that the reason for this is the impaired accessibility of hematite caused by the surrounding alginate matrix. In contrast, another study focusing on the reduction of oxyhydroxide nanoparticles by *Shewanella putrefaciens* CIP 80.40 T in the presence of humic acids observe an eight times increase in Fe(III) reduction rate compared to nano-lepidocrocite[136]. Therefore, the influence of natural organic matter on Fe(III) reduction rate seems to be complex and probably depends on a multitude of different factors.

Another layer of complexity is added if one considers that iron is not only used as an electron acceptor for the dissimilatory reduction but also comprises a micronutrient whose acquisition is critical for any species of bacteria. The acquisition can occur via a multitude of mechanisms, for example by complexation through siderophores or, interestingly, also the direct penetration of the outer membrane by nanoparticles. Dehner *et al.* have shown that this latter option requires a diameter of less than 10 nm for *Pseudomonas mendocina* and hematite nanoparticles[137]. From the view of the FeRB, the presence of *Pseudomonas* species could therefore be beneficial or detrimental, depending on the environmental conditions. The produced siderophores increase the bioavailability of Fe(III), whereas nanoparticles penetrating the outer membrane of other bacteria results in a loss of potential electron acceptor for the FeRB. From the modelers' perspective, this opens up the perspective of an interesting tri-partite system of interaction in the context of microbial iron cycling. However, for the accurate description of conditions supporting mutually beneficial interactions of the participating species, further research is required on how the size, shape, and state of Fe nanoparticles influence important characteristics like bioavailability and Fe(III) reduction rate. I am confident that in this regard as well, mathematical modeling can contribute to the identification of key patterns from the multitude of available experimental studies.

References

1   Pearce, B. K., Tupper, A. S., Pudritz, R. E. & Higgs, P. G. Constraining the time interval for the origin of life on Earth. *Astrobiology* **18**, 343-364 (2018).

2   McCollom, T. M. Miller-Urey and Beyond: What Have We Learned About Prebiotic Organic Synthesis Reactions in the Past 60 Years? *Annual Review of Earth and Planetary Sciences* **41**, 207-229, doi:10.1146/annurev-earth-040610-133457 (2013).

3   Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **117**, 528-529 (1953).

4   Miller, S. L. Production of some organic compounds under possible primitive earth conditions1. *Journal of the American Chemical Society* **77**, 2351-2361 (1955).

5   Miller, S. L. The mechanism of synthesis of amino acids by electric discharges. *Biochimica et Biophysica Acta* **23**, 480-489 (1957).

6   Parker, E. T. *et al.* Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment. *Proc Natl Acad Sci U S A* **108**, 5526-5531, doi:10.1073/pnas.1019191108 (2011).

7   Kitadai, N. & Maruyama, S. Origins of building blocks of life: A review. *Geoscience Frontiers* **9**, 1117-1153, doi:10.1016/j.gsf.2017.07.007 (2018).

8   Bada, J. L. & Lazcano, A. Prebiotic soup--revisiting the miller experiment. *Science* **300**, 745-746 (2003).

9   Bada, J. L. & Lazcano, A. Some like it hot, but not the first biomolecules. *Science* **296**, 1982-1983 (2002).

10   Pizzarello, S., Cooper, G. & Flynn, G. The nature and distribution of the organic material in carbonaceous chondrites and interplanetary dust particles. *Meteorites and the Early Solar System II* **1**, 625-651 (2006).

11   Fichtner, M., Voigt, K. & Schuster, S. The tip and hidden part of the iceberg: Proteinogenic and non-proteinogenic aliphatic amino acids. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1861**, 3258-3269 (2017).

12   Weber, A. L. & Miller, S. L. Reasons for the occurrence of the twenty coded protein amino acids. *Journal of Molecular Evolution* **17**, 273-284 (1981).

13   Philip, G. K. & Freeland, S. J. Did evolution select a nonrandom "alphabet" of amino acids? *Astrobiology* **11**, 235-240, doi:10.1089/ast.2010.0567 (2011).

14   Lu, Y. & Freeland, S. J. A quantitative investigation of the chemical space surrounding amino acid alphabet formation. *Journal of Theoretical Biology* **250**, 349-361 (2008).

15   Ilardo, M., Meringer, M., Freeland, S., Rasulev, B. & Cleaves, H. J., 2nd. Extraordinarily adaptive properties of the genetically encoded amino acids. *Sci Rep* **5**, 9414, doi:10.1038/srep09414 (2015).

16   Higgs, P. G. & Pudritz, R. E. A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology* **9**, 483-490 (2009).

17   Longo, L. M. & Blaber, M. Protein design at the interface of the pre-biotic and biotic worlds. *Arch Biochem Biophys* **526**, 16-21, doi:10.1016/j.abb.2012.06.009 (2012).

18   Trifonov, E. N. Consensus temporal order of amino acids and evolution of the triplet code. *Gene* **261**, 139-151 (2000).

19   Shibue, R. *et al.* Comprehensive reduction of amino acid set in a protein suggests the importance of prebiotic amino acids for stable proteins. *Sci Rep* **8**, 1227, doi:10.1038/s41598-018-19561-1 (2018).

20   Solis, A. D. Reduced alphabet of prebiotic amino acids optimally encodes the conformational space of diverse extant protein folds. *BMC Evol Biol* **19**, 158, doi:10.1186/s12862-019-1464-6 (2019).

21   Higgs, P. G. A four-column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code. *Biol Direct* **4**, 16, doi:10.1186/1745-6150-4-16 (2009).

22   Georgiou, C. D. Functional Properties of Amino Acid Side Chains as Biomarkers of Extraterrestrial Life. *Astrobiology* **18**, 1479-1496, doi:10.1089/ast.2018.1868 (2018).

23   Granold, M., Hajieva, P., Tosa, M. I., Irimie, F. D. & Moosmann, B. Modern diversification of the amino acid repertoire driven by oxygen. *Proc Natl Acad Sci U S A* **115**, 41-46, doi:10.1073/pnas.1717100115 (2018).

24   Taylor, W. R. The classification of amino acid conservation. *Journal of Theoretical Biology* **119**, 205-218 (1986).

25   Dayhoff, M. O. *Atlas of protein sequence and structure*. (National Biomedical Research Foundation., 1972).

26   Kosiol, C., Goldman, N. & Buttimore, N. H. A new criterion and method for amino acid classification. *J Theor Biol* **228**, 97-106, doi:10.1016/j.jtbi.2003.12.010 (2004).

27   Stephenson, J. D. & Freeland, S. J. Unearthing the root of amino acid similarity. *J Mol Evol* **77**, 159-169, doi:10.1007/s00239-013-9565-0 (2013).

28   Peterson, E. L., Kondev, J., Theriot, J. A. & Phillips, R. Reduced amino acid alphabets exhibit an improved sensitivity and selectivity in fold assignment. *Bioinformatics* **25**, 1356-1362 (2009).

29   Dong, G. F., Zheng, L., Huang, S. H., Gao, J. & Zuo, Y. C. Amino Acid Reduction Can Help to Improve the Identification of Antimicrobial Peptides and Their Functional Activities. *Front Genet* **12**, 669328, doi:10.3389/fgene.2021.669328 (2021).

30   Yuan, J. *et al.* Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. *FEBS Lett* **584**, 342-349, doi:10.1016/j.febslet.2009.11.005 (2010).

31   Kryukov, G. V. *et al.* Characterization of mammalian selenoproteomes. *Science* **300**, 1439-1443 (2003).

32   Labunskyy, V. M., Hatfield, D. L. & Gladyshev, V. N. Selenoproteins: molecular pathways and physiological roles. *Physiological Reviews* **94**, 739-777 (2014).

33   Papp, L. V., Lu, J., Holmgren, A. & Khanna, K. K. From selenium to selenoproteins: synthesis, identity, and their role in human health. *Antioxidants & Redox Signaling* **9**, 775-806 (2007).

34   Gaston, M. A., Jiang, R. & Krzycki, J. A. Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr Opin Microbiol* **14**, 342-349, doi:10.1016/j.mib.2011.04.001 (2011).

35   Young, T. S. & Schultz, P. G. Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J Biol Chem* **285**, 11039-11044, doi:10.1074/jbc.R109.091306 (2010).

36   Shandell, M. A., Tan, Z. & Cornish, V. W. Genetic Code Expansion: A Brief History and Perspective. *Biochemistry* **60**, 3455-3469, doi:10.1021/acs.biochem.1c00286 (2021).

37   Tang, H., Zhang, P. & Luo, X. Recent Technologies for Genetic Code Expansion and their Implications on Synthetic Biology Applications. *J Mol Biol*, 167382, doi:10.1016/j.jmb.2021.167382 (2021).

38   Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu Rev Biochem* **79**, 413-444, doi:10.1146/annurev.biochem.052308.105824 (2010).

39   Lang, K. & Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem Rev* **114**, 4764-4806, doi:10.1021/cr400355w (2014).

40   Lummis, S. C. *et al.* Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* **438**, 248-252, doi:10.1038/nature04130 (2005).

41   Wang, Y. *et al.* Expanding the Structural Diversity of Protein Building Blocks with Noncanonical Amino Acids Biosynthesized from Aromatic Thiols. *Angew Chem Int Ed Engl* **60**, 10040-10048, doi:10.1002/anie.202014540 (2021).

42   Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342**, 357-360 (2013).

43   Zhang, Y. *et al.* A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **551**, 644-647 (2017).

44   Parry, D. A. & Squire, J. M. *Fibrous proteins: structures and mechanisms*. (Springer, 2017).

45   Fodje, M. & Al-Karadaghi, S. Occurrence, conformational features and amino acid propensities for the π-helix. *Protein Engineering, Design and Selection* **15**, 353-358 (2002).

46   Cornish, V. W., Kaplan, M. I., Veenstra, D. L., Kollman, P. A. & Schultz, P. G. Stabilizing and destabilizing effects of placing. beta.-branched amino acids in protein. alpha.-Helixes. *Biochemistry* **33**, 12022-12031 (1994).

47   Hol, W., Van Duijnen, P. T. & Berendsen, H. The α-helix dipole and the properties of proteins. *Nature* **273**, 443-446 (1978).

48   Baker, E. G. *et al.* Local and macroscopic electrostatic interactions in single alpha-helices. *Nat Chem Biol* **11**, 221-228, doi:10.1038/nchembio.1739 (2015).

49   Wolny, M. *et al.* Characterization of long and stable de novo single alpha-helix domains provides novel insight into their stability. *Sci Rep* **7**, 44341, doi:10.1038/srep44341 (2017).

50   Pace, C. N. & Scholtz, J. M. A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical Journal* **75**, 422-427 (1998).

51   Blaber, M., Zhang, X.-j. & Matthews, B. W. Structural basis of amino acid α helix propensity. *Science* **260**, 1637-1640 (1993).

52   Ho, B. K. & Brasseur, R. The Ramachandran plots of glycine and pre-proline. *BMC Structural Biology* **5**, 1-11 (2005).

53   Crick, F. The packing of α-helices: simple coiled-coils. *Acta Crystallographica* **6**, 689-697 (1953).

54   Offer, G., Hicks, M. R. & Woolfson, D. N. Generalized Crick equations for modeling noncanonical coiled coils. *Journal of Structural Biology* **137**, 41-53 (2002).

55   Gruber, M. & Lupas, A. N. Historical review: another 50th anniversary--new periodicities in coiled coils. *Trends Biochem Sci* **28**, 679-685, doi:10.1016/j.tibs.2003.10.008 (2003).

56   Landschulz, W. H., Johnson, P. F. & McKnight, S. L. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759-1764 (1988).

57   Aravind, L. & Ponting, C. P. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiology Letters* **176**, 111-116 (1999).

58   Hartmann, M. D. *et al.* A coiled-coil motif that sequesters ions to the hydrophobic core. *Proceedings of the National Academy of Sciences* **106**, 16950-16955 (2009).

59   Brown, J. H., Cohen, C. & Parry, D. A. Heptad breaks in α-helical coiled coils: Stutters and stammers. *Proteins: Structure, Function, and Bioinformatics* **26**, 134-145 (1996).

60   Hartmann, M. D. *et al.* alpha/beta coiled coils. *Elife* **5**, doi:10.7554/eLife.11861 (2016).

61   Alvarez, B. H. *et al.* A transition from strong right-handed to canonical left-handed supercoiling in a conserved coiled-coil segment of trimeric autotransporter adhesins. *J Struct Biol* **170**, 236-245, doi:10.1016/j.jsb.2010.02.009 (2010).

62   Lupas, A. Coiled coils: new structures and new functions. *Trends in Biochemical Sciences* **21**, 375-382 (1996).

63   Hartmann, M. D. *et al.* A soluble mutant of the transmembrane receptor Af1503 features strong changes in coiled-coil periodicity. *J Struct Biol* **186**, 357-366, doi:10.1016/j.jsb.2014.02.008 (2014).

64   Pinter, T. B. J. *et al.* Making or Breaking Metal-Dependent Catalytic Activity: The Role of Stammers in Designed Three-Stranded Coiled Coils. *Angew Chem Int Ed Engl* **59**, 20445-20449, doi:10.1002/anie.202008356 (2020).

65   Pinter, T. B. J. *et al.* Open Reading Frame 1 Protein of the Human Long Interspersed Nuclear Element 1 Retrotransposon Binds Multiple Equivalents of Lead. *J Am Chem Soc* **143**, 15271-15278, doi:10.1021/jacs.1c06461 (2021).

66   Koebke, K. J. *et al.* Nitrite reductase activity within an antiparallel de novo scaffold. *J Biol Inorg Chem* **26**, 855-862, doi:10.1007/s00775-021-01889-1 (2021).

67   Reja, R. M. *et al.* pH sensitive coiled coils: a strategy for enhanced liposomal drug delivery. *Nanoscale* **8**, 5139-5145, doi:10.1039/c5nr07734f (2016).

68   Al-Ahmady, Z. S. *et al.* Lipid–peptide vesicle nanoscale hybrids for triggered drug release by mild hyperthermia in vitro and in vivo. *ACS Nano* **6**, 9335-9346 (2012).

69   Higgins, C. D., Malashkevich, V. N., Almo, S. C. & Lai, J. R. Influence of a heptad repeat stutter on the pH-dependent conformational behavior of the central coiled-coil from influenza hemagglutinin HA2. *Proteins* **82**, 2220-2228, doi:10.1002/prot.24585 (2014).

70   Capra, F. & Luisi, P. L. *The systems view of life: A unifying vision*. (Cambridge University Press, 2014).

71   Ferrer, J., Prats, C. & Lopez, D. Individual-based modelling: an essential tool for microbiology. *J Biol Phys* **34**, 19-37, doi:10.1007/s10867-008-9082-3 (2008).

72   Klimenko, A. I. *et al.* A review of simulation and modeling approaches in microbiology. *Russian Journal of Genetics: Applied Research* **6**, 845-853, doi:10.1134/s2079059716070066 (2016).

73   Azimi, M., Jamali, Y. & Mofrad, M. R. Accounting for diffusion in agent based models of reaction-diffusion systems with application to cytoskeletal diffusion. *PLoS One* **6**, e25306, doi:10.1371/journal.pone.0025306 (2011).

74   Tokarski, C. *et al.* Agent-based modeling approach of immune defense against spores of opportunistic human pathogenic fungi. *Front Microbiol* **3**, 129, doi:10.3389/fmicb.2012.00129 (2012).

75   Miller, J., Parker, M., Bourret, R. B. & Giddings, M. C. An agent-based model of signal transduction in bacterial chemotaxis. *PLoS One* **5**, e9454, doi:10.1371/journal.pone.0009454 (2010).

76   Lardon, L. A. *et al.* iDynoMiCS: next-generation individual-based modelling of biofilms. *Environ Microbiol* **13**, 2416-2434, doi:10.1111/j.1462-2920.2011.02414.x (2011).

77   Sweeney, E. G. *et al.* Agent-Based Modeling Demonstrates How Local Chemotactic Behavior Can Shape Biofilm Architecture. *mSphere* **4**, doi:10.1128/mSphere.00285-19 (2019).

78   Wilmoth, J. L. *et al.* A Microfluidics and Agent-Based Modeling Framework for Investigating Spatial Organization in Bacterial Colonies: The Case of Pseudomonas Aeruginosa and H1-Type VI Secretion Interactions. *Front Microbiol* **9**, 33, doi:10.3389/fmicb.2018.00033 (2018).

79   Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nature Biotechnology* **28**, 245-248 (2010).

80   Bauer, E., Zimmermann, J., Baldini, F., Thiele, I. & Kaleta, C. BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. *PLoS Comput Biol* **13**, e1005544, doi:10.1371/journal.pcbi.1005544 (2017).

81   Germerodt, S. *et al.* Pervasive Selection for Cooperative Cross-Feeding in Bacterial Communities. *PLoS Comput Biol* **12**, e1004986, doi:10.1371/journal.pcbi.1004986 (2016).

82   Bird, L. J., Bonnefoy, V. & Newman, D. K. Bioenergetic challenges of microbial iron metabolisms. *Trends Microbiol* **19**, 330-340, doi:10.1016/j.tim.2011.05.001 (2011).

83   Roden, E. E., Sobolev, D., Glazer, B. & Luther, G. W. Potential for microscale bacterial Fe redox cycling at the aerobic-anaerobic interface. *Geomicrobiology Journal* **21**, 379-391 (2004).

84   Hines, M. E. in *The Interactions Between Sediments and Water*   159-172 (Springer, 2006).

85   Druschel, G. K., Emerson, D., Sutka, R., Suchecki, P. & Luther III, G. W. Low-oxygen and chemical kinetic constraints on the geochemical niche of neutrophilic iron (II) oxidizing microorganisms. *Geochimica et Cosmochimica Acta* **72**, 3358-3370 (2008).

86   Melton, E. D., Swanner, E. D., Behrens, S., Schmidt, C. & Kappler, A. The interplay of microbially mediated and abiotic reactions in the biogeochemical Fe cycle. *Nat Rev Microbiol* **12**, 797-808, doi:10.1038/nrmicro3347 (2014).

87   Roden, E. E. Fe (III) oxide reactivity toward biological versus chemical reduction. *Environmental Science & Technology* **37**, 1319-1324 (2003).

88   Najem, T., Langley, S. & Fortin, D. A comparison of Fe(III) reduction rates between fresh and aged biogenic iron oxides (BIOS) by Shewanella putrefaciens CN32. *Chemical Geology* **439**, 1-12, doi:10.1016/j.chemgeo.2016.06.006 (2016).

89   Han, R. *et al.* Dependence of Secondary Mineral Formation on Fe(II) Production from Ferrihydrite Reduction by Shewanella oneidensis MR-1. *ACS Earth and Space Chemistry* **2**, 399-409, doi:10.1021/acsearthspacechem.7b00132 (2018).

90   Schädler, S. *et al.* Formation of Cell-Iron-Mineral Aggregates by Phototrophic and Nitrate-Reducing Anaerobic Fe(II)-Oxidizing Bacteria. *Geomicrobiology Journal* **26**, 93-103, doi:10.1080/01490450802660573 (2009).

91   Bose, S. *et al.* Bioreduction of hematite nanoparticles by the dissimilatory iron reducing bacterium Shewanella oneidensis MR-1. *Geochimica et Cosmochimica Acta* **73**, 962-976, doi:10.1016/j.gca.2008.11.031 (2009).

92   Harris, H. W., El-Naggar, M. Y. & Nealson, K. H. Shewanella oneidensis MR-1 chemotaxis proteins and electron-transport chain components essential for congregation near insoluble electron acceptors. *Biochem Soc Trans* **40**, 1167-1177, doi:10.1042/BST20120232 (2012).

93   Harris, H. W. *et al.* Redox Sensing within the Genus Shewanella. *Front Microbiol* **8**, 2568, doi:10.3389/fmicb.2017.02568 (2017).

94   Starwalt-Lee, R., El-Naggar, M. Y., Bond, D. R. & Gralnick, J. A. Electrolocation? The evidence for redox-mediated taxis in Shewanella oneidensis. *Mol Microbiol* **115**, 1069-1079, doi:10.1111/mmi.14647 (2021).

95   Oram, J. & Jeuken, L. J. Tactic response of Shewanella oneidensis MR-1 toward insoluble electron acceptors. *MBio* **10**, e02490-02418 (2019).

96   Marsili, E. *et al.* Shewanella secretes flavins that mediate extracellular electron transfer. *Proceedings of the National Academy of Sciences* **105**, 3968-3973 (2008).

97   Kotloski, N. J. & Gralnick, J. A. Flavin electron shuttles dominate extracellular electron transfer by Shewanella oneidensis. *MBio* **4**, doi:10.1128/mBio.00553-12 (2013).

98   Roberts, J. A., Fowle, D. A., Hughes, B. T. & Kulczycki, E. Attachment Behavior ofShewanella putrefaciensonto Magnetite under Aerobic and Anaerobic Conditions. *Geomicrobiology Journal* **23**, 631-640, doi:10.1080/01490450600964441 (2007).

99   Tokunou, Y. & Okamoto, A. Geometrical Changes in the Hemes of Bacterial Surface c-Type Cytochromes Reveal Flexibility in Their Binding Affinity with Minerals. *Langmuir* **35**, 7529-7537, doi:10.1021/acs.langmuir.8b02977 (2019).

100   Baalousha, M. Aggregation and disaggregation of iron oxide nanoparticles: influence of particle concentration, pH and natural organic matter. *Science of the Total Environment* **407**, 2093-2101 (2009).

101   Schmid, G. *et al.* 3-D analysis of bacterial cell-(iron)mineral aggregates formed during Fe(II) oxidation by the nitrate-reducing Acidovorax sp. strain BoFeN1 using complementary microscopy tomography approaches. *Geobiology* **12**, 340-361, doi:10.1111/gbi.12088 (2014).

102   Hegler, F., Schmidt, C., Schwarz, H. & Kappler, A. Does a low-pH microenvironment around phototrophic Fe(II) -oxidizing bacteria prevent cell encrustation by Fe(III) minerals? *FEMS Microbiol Ecol* **74**, 592-600, doi:10.1111/j.1574-6941.2010.00975.x (2010).

103   Kawashima, S. & Kanehisa, M. AAindex: amino acid index database. *Nucleic Acids Research* **28**, 374-374 (2000).

104   Kawashima, S. *et al.* AAindex: amino acid index database, progress report 2008. *Nucleic Acids Research* **36**, D202-D205 (2007).

105   Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862-864 (1974).

106   Goldsack, D. & Chalifoux, R. Contribution of the free energy of mixing of hydrophobic side chains to the stability of the tertiary structure of proteins. *Journal of Theoretical Biology* **39**, 645-651 (1973).

107   Prabhakaran, M. The distribution of physical, chemical and conformational properties in signal and nascent peptides. *Biochemical Journal* **269**, 691-696 (1990).

108   Engelman, D., Steitz, T. & Goldman, A. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annual Review of Biophysics and Biophysical Chemistry* **15**, 321-353 (1986).

109   Cheng, T. *et al.* Computation of octanol– water partition coefficients by guiding an additive model with knowledge. *Journal of Chemical Information and Modeling* **47**, 2140-2148 (2007).

110   Veljkovic, V. *A theoretical approach to the preselection of carcinogens and chemical carcinogenesis.* (Gordon & Breach Publishing Group, 1980).

111   Pauling, L. *The Quarterly Review of Biology* **57**, 228-229 (1982).

112   Im, W., Beglov, D. & Roux, B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Computer Physics Communications* **111**, 59-75 (1998).

113   Creighton, T. E. *Proteins: structures and molecular properties.* (Macmillan, 1993).

114   Horovitz, A., Matthews, J. M. & Fersht, A. R. α-Helix stability in proteins: II. Factors that influence stability at an internal position. *Journal of Molecular Biology* **227**, 560-568 (1992).

115   Revilla-Lopez, G. *et al.* Integrating the intrinsic conformational preferences of noncoded alpha-amino acids modified at the peptide bond into the noncoded amino acids database. *Proteins* **79**, 1841-1852, doi:10.1002/prot.23009 (2011).

116   Alias, M., Ayuso-Tejedor, S., Fernandez-Recio, J., Cativiela, C. & Sancho, J. Helix propensities of conformationally restricted amino acids. Non-natural substitutes for helix breaking proline and helix forming alanine. *Org Biomol Chem* **8**, 788-792, doi:10.1039/b919671d (2010).

117   Riek, R. P. & Graham, R. M. The elusive pi-helix. *J Struct Biol* **173**, 153-160, doi:10.1016/j.jsb.2010.09.001 (2011).

118   Trifonov, E. N. The origin of the genetic code and of the earliest oligopeptides. *Research in Microbiology* **160**, 481-486 (2009).

119   Veljkovic, V., Veljkovic, N., Este, J., Huther, A. & Dietrich, U. Application of the EIIP/ISM bioinformatics concept in development of new drugs. *Current Medicinal Chemistry* **14**, 441-453 (2007).

120   Testa, O. D., Moutevelis, E. & Woolfson, D. N. CC+: a relational database of coiled-coil structures. *Nucleic Acids Research* **37**, D315-D322 (2009).

121   Qin, Z., Fabre, A. & Buehler, M. J. Structure and mechanism of maximum stability of isolated alpha-helical protein domains at a critical length scale. *Eur Phys J E Soft Matter* **36**, 53, doi:10.1140/epje/i2013-13053-8 (2013).

122   Lupas, A. N. & Bassler, J. Coiled Coils - A Model System for the 21st Century. *Trends Biochem Sci* **42**, 130-140, doi:10.1016/j.tibs.2016.10.007 (2017).

123   Stetefeld, J. *et al.* Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer. *Nature Structural Biology* **7**, 772-776 (2000).

124   Szczepaniak, K., Bukala, A., da Silva Neto, A. M., Ludwiczak, J. & Dunin-Horkawicz, S. A library of coiled-coil domains: from regular bundles to peculiar twists. *Bioinformatics* **36**, 5368-5376 (2020).

125   Walls, A. C. *et al.* Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proceedings of the National Academy of Sciences* **114**, 11157-11162 (2017).

126   Kumar, P. & Woolfson, D. N. Socket2: A Program for Locating, Visualising, and Analysing Coiled-coil Interfaces in Protein Structures. *Bioinformatics*, doi:10.1093/bioinformatics/btab631 (2021).

127   Barcellos, D., Cyle, K. T. & Thompson, A. Faster redox fluctuations can lead to higher iron reduction rates in humid forest soils. *Biogeochemistry* **137**, 367-378, doi:10.1007/s10533-018-0427-0 (2018).

128   Mejia, J., Roden, E. E. & Ginder-Vogel, M. Influence of Oxygen and Nitrate on Fe (Hydr)oxide Mineral Transformation and Soil Microbial Communities during Redox Cycling. *Environ Sci Technol* **50**, 3580-3588, doi:10.1021/acs.est.5b05519 (2016).

129   Ginn, B., Meile, C., Wilmoth, J., Tang, Y. & Thompson, A. Rapid Iron Reduction Rates Are Stimulated by High-Amplitude Redox Fluctuations in a Tropical Forest Soil. *Environ Sci Technol* **51**, 3250-3259, doi:10.1021/acs.est.6b05709 (2017).

130   St Clair, B., Pottenger, J., Debes, R., Hanselmann, K. & Shock, E. Distinguishing biotic and abiotic iron oxidation at low temperatures. *ACS Earth and Space Chemistry* **3**, 905-921 (2019).

131   Singer, P. C. & Stumm, W. Acidic mine drainage: the rate-determining step. *Science* **167**, 1121-1123 (1970).

132   Bosch, J., Heister, K., Hofmann, T. & Meckenstock, R. U. Nanosized iron oxide colloids strongly enhance microbial iron reduction. *Appl Environ Microbiol* **76**, 184-189, doi:10.1128/AEM.00417-09 (2010).

133   Liu, J. *et al.* Particle size effect and the mechanism of hematite reduction by the outer membrane cytochrome OmcA of Shewanella oneidensis MR-1. *Geochimica et Cosmochimica Acta* **193**, 160-175, doi:10.1016/j.gca.2016.08.022 (2016).

134   Chernyshova, I. V., Ponnurangam, S. & Somasundaran, P. On the origin of an unusual dependence of (bio)chemical reactivity of ferric hydroxides on nanoparticle size. *Phys Chem Chem Phys* **12**, 14045-14056, doi:10.1039/c0cp00168f (2010).

135   Weihe, S. H. C., Mangayayam, M., Sand, K. K. & Tobler, D. J. Hematite Crystallization in the Presence of Organic Matter: Impact on Crystal Properties and Bacterial Dissolution. *ACS Earth and Space Chemistry* **3**, 510-518, doi:10.1021/acsearthspacechem.8b00166 (2019).

136   Pedrot, M., Le Boudec, A., Davranche, M., Dia, A. & Henin, O. How does organic matter constrain the nature, size and availability of Fe nanoparticles for biological reduction? *J Colloid Interface Sci* **359**, 75-85, doi:10.1016/j.jcis.2011.03.067 (2011).

137   Dehner, C. A., Barton, L., Maurice, P. A. & DuBois, J. L. Size-dependent bioavailability of hematite (α-Fe2O3) nanoparticles to a common aerobic bacterium. *Environmental Science & Technology* **45**, 977-983 (2011).

Acknowledgements

Angaben zum Eigenanteil

Manuskript Nr. 1

Kurzreferenz: Then *et al.* (2020), Sci. Rep.

Beitrag des Doktoranden / der Doktorandin

Beitrag des Doktoranden / der Doktorandin zu Abbildungen, die experimentelle Daten wiedergeben (nur für Originalartikel):

Abbildung(en) # 1-6	[X]	100 % (die in dieser Abbildung wiedergegebenen Daten entstammen vollständig experimentellen Arbeiten, die der Kandidat/die Kandidatin durchgeführt hat)
	☐	0 % (die in dieser Abbildung wiedergegebenen Daten basieren ausschließlich auf Arbeiten anderer Koautoren)
	☐	Etwaiger Beitrag des Doktoranden / der Doktorandin zur Abbildung: _____% Kurzbeschreibung des Beitrages: *(z. B. „Abbildungsteile a, d und f" oder „Auswertung der Daten" etc)*

Manuskript Nr. 2

Kurzreferenz -

Beitrag des Doktoranden / der Doktorandin

Beitrag des Doktoranden / der Doktorandin zu Abbildungen, die experimentelle Daten wiedergeben (nur für Originalartikel):

Abbildung(en) # 1-8	[X]	100 % (die in dieser Abbildung wiedergegebenen Daten entstammen vollständig experimentellen Arbeiten, die der Kandidat/die Kandidatin durchgeführt hat)
	☐	0 % (die in dieser Abbildung wiedergegebenen Daten basieren ausschließlich auf Arbeiten anderer Koautoren)
	☐	Etwaiger Beitrag des Doktoranden / der Doktorandin zur Abbildung: _____% Kurzbeschreibung des Beitrages: *(z. B. „Abbildungsteile a, d und f" oder „Auswertung der Daten" etc)*

Manuskript Nr. 3

Kurzreferenz: Then *et al.* (2022), R. Soc. Open Sci.

Beitrag des Doktoranden / der Doktorandin

Beitrag des Doktoranden / der Doktorandin zu Abbildungen, die experimentelle Daten wiedergeben (nur für Originalartikel):

Abbildung(en) # 1-7	[X]	100 % (die in dieser Abbildung wiedergegebenen Daten entstammen vollständig experimentellen Arbeiten, die der Kandidat/die Kandidatin durchgeführt hat)
	☐	0 % (die in dieser Abbildung wiedergegebenen Daten basieren ausschließlich auf Arbeiten anderer Koautoren)
	☐	Etwaiger Beitrag des Doktoranden / der Doktorandin zur Abbildung: _____% Kurzbeschreibung des Beitrages: *(z. B. „Abbildungsteile a, d und f" oder „Auswertung der Daten" etc)*

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe. Mir ist die geltende Promotionsordnung der Fakultät für Biowissenschaften bekannt. Bei der Auswahl und Auswertung des Materials, sowie auch bei der Herstellung des Manuskripts haben mich Prof. Dr. Stefan Schuster und Prof. Dr. Bashar Ibrahim unterstützt. Ich habe weder die Hilfe eines Promotionsberaters in Anspruch genommen, noch haben Dritte unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die vorgelegte Dissertation wurde zuvor nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht. Ferner habe ich weder die gleiche, noch eine abgewandelte Form der vorliegenden Arbeit, oder eine andere Abhandlung bei einer anderen Hochschule als Dissertation eingereicht.

Jena, den   _____

                                                                                        André Then