

# ChelomEx: Isotope-Assisted Discovery of Metal Chelates in Complex Media Using High-Resolution LC-MS

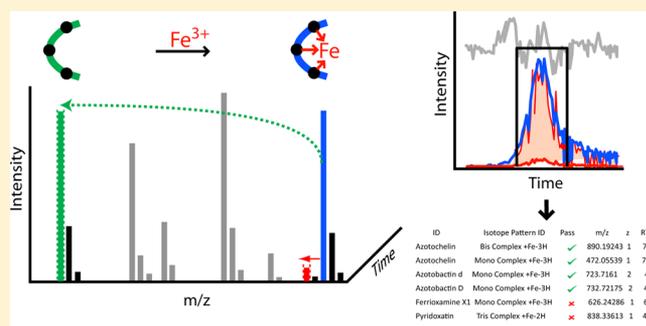
Oliver Baars,<sup>\*,†</sup> François M. M. Morel,<sup>†</sup> and David H. Perlman<sup>‡,§</sup>

<sup>†</sup>Department of Geosciences, Princeton University, Princeton, New Jersey 08544, United States

<sup>‡</sup>Princeton Collaborative Proteomics and Mass Spectrometry Core, Departments of Molecular Biology and Chemistry, and the Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, 08544 New Jersey, United States

## S Supporting Information

**ABSTRACT:** Chelating agents can control the speciation and reactivity of trace metals in biological, environmental, and laboratory-derived media. A large number of trace metals (including Fe, Cu, Zn, Hg, and others) show characteristic isotopic fingerprints that can be exploited for the discovery of known and unknown organic metal complexes and related chelating ligands in very complex sample matrices using high-resolution liquid chromatography mass spectrometry (LC-MS). However, there is currently no free open-source software available for this purpose. We present a novel software tool, ChelomEx, which identifies isotope pattern-matched chromatographic features associated with metal complexes along with free ligands and other related adducts in high-resolution LC-MS data. High sensitivity and exclusion of false positives are achieved by evaluation of the chromatographic coherence of the isotope pattern within chromatographic features, which we demonstrate through the analysis of bacterial culture media. A built-in graphical user interface and compound library aid in identification and efficient evaluation of results. ChelomEx is implemented in MatLab. The source code, binaries for MS Windows and MAC OS X as well as test LC-MS data are available for download at SourceForge (<http://sourceforge.net/projects/chelomex>).



Although the presence of strong, biologically derived ligands for Fe and other trace metals in environmental and biological samples is well-known, through indirect speciation measurements (e.g., electrochemistry or liquid chromatography inductively coupled-plasma mass spectrometry<sup>1–3</sup>), their chemical nature often remains unknown due to low concentrations in highly complex sample matrices.<sup>4–6</sup> Metal ions also play a central role as catalysts in chemical laboratories, and the ability to discover chelating agents in complex combinatorial-synthetic mixtures is highly desirable.<sup>7,8</sup> Developments in high-resolution liquid chromatography mass spectrometry (LC-MS) instrumentation make it possible to resolve and characterize a growing number of metabolites and other molecules in complex samples<sup>9</sup> and this new technology may be used as a direct approach in the discovery of metal chelating agents.

However, resolving the chemical composition of a measured species to reveal that it contains a metal ion is a challenge. The number of possible elemental compositions that can be assigned to a given  $m/z$  value within the measurement error increases exponentially with the molecular mass and number of different elements to consider in the compound. Therefore, despite the high mass accuracy of current MS instruments (at or below the ppm level), it is not possible in most cases to

assign a unique sum formula to the measured mass, and thereby reveal metal-containing compounds in a facile manner.<sup>10</sup>

One promising approach is to exploit characteristic natural stable-isotopic fingerprints that are associated with many metals in order to recognize their presence in organic complexes. Detection of the <sup>54</sup>Fe–<sup>56</sup>Fe or the <sup>69</sup>Ga–<sup>71</sup>Ga pair has been previously used as a screen for siderophores in LC-MS data from seawater<sup>11–13</sup> and culture extracts.<sup>14</sup> Yet, there is no freely available software to aid in the systematic detection of defined isotope patterns at low abundance in highly complex sample matrices.

Here, we present an open-source software package called ChelomEx (short for Chelomics Explorer) for targeted and untargeted recognition of organic metal complexes and metal chelating compounds from complex LC-MS data. ChelomEx builds on algorithms and procedures for isotope pattern filtering that have already been described in the literature,<sup>14,15</sup> but it includes additional features that address challenges specific to detecting isotope patterns associated with metal chelates. Poor chromatographic peak shapes are often observed for metal-adducted species, due to interaction of charged

Received: August 11, 2014

Accepted: October 21, 2014

Published: October 21, 2014

binding sites with standard chromatographic media, surfaces of glass capillaries, etc., which may prevent effective analysis by available peak picking algorithms. ChelomEx recognizes isotope pattern-matched chromatographic features independent of peak shape and at signal intensities close to the detection limit. In many samples, unbound ligand species may be more abundant than their metal complexes, especially at low metal concentrations or at the low pH values of acidic mobile phases that are often used in reversed-phase LC-MS. Identification of the unbound ligand species strengthens the assignment of metal complexes and is highly useful for compound identification and structural determination. ChelomEx includes an algorithm to relate metal complexes to their free ligands based on chromatographic coelution or related MS/MS spectra. A built-in compound library and graphical user interface aid in identification, evaluation, and prioritization of results. We demonstrate the accuracy and sensitivity of the software in the example of Fe complexation using siderophore standards and extracts from bacterial culture supernatants that contain a set of known catechol siderophores.

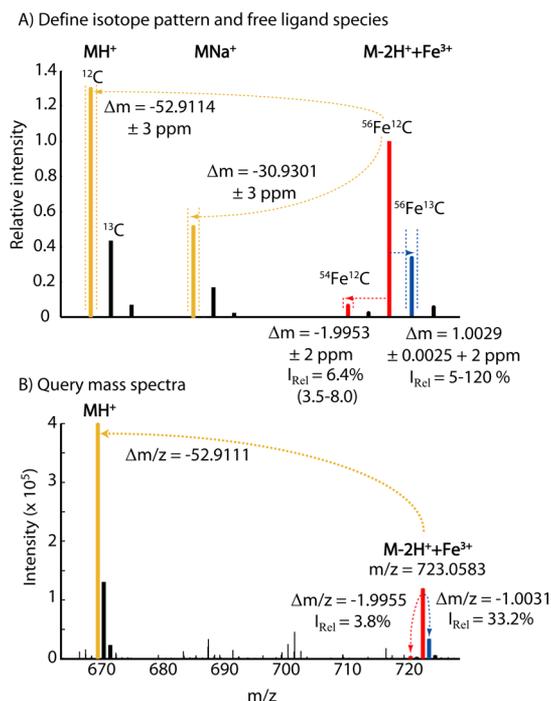
## ■ APPROACH

High-resolution LC-MS data is typically composed of time series of individual, highly complex mass spectra (full scan MS1), often interspersed by one or more tandem mass spectra (MS/MS, MS2). Intensity contours of individual species in the spectra can be stitched together over the time dimension (generating extracted ion chromatograms, EICs) to reveal chromatographic elution peaks of the individual ions. One of the central challenges for our software is to differentiate mass spectral patterns that result from a defined isotopic fingerprint of an organic metal complex (true positives) from those that may be caused by a combination of mass spectral peaks from two or more species with unrelated elemental compositions (false positives), which incidentally coelute. To address this, even at low signal intensities and with complex matrices, the algorithm consists of three main parts: (1) identification of pattern-matched isotope clusters, (2) deconvolution of these clusters into pattern-matched chromatographic features, and (3) evaluation of the chromatographic coherence of the isotope pattern in these features. Subsequent steps consist of the detection of related free ligands and other user-defined adduct species, which can give further confidence to the metal complex assignment and be highly useful in compound identification and structural determination.

### Identification of Pattern-Matched Isotope Clusters.

The isotope pattern for the metal complex and related free ligands or other adducts are defined by the user (Figure 1). The isotope pattern definition consists of the expected mass differences ( $\Delta m$ ) and intensity ratios ( $I_{\text{Rel}}$ ) between the isotopologues together with measurement error boundaries. A list of  $\Delta m$  and  $I_{\text{Rel}}$  values for a number of biologically relevant elements is given in Table S-1 (Supporting Information). For each charge state ( $z$ ) to be included in the analysis, the algorithm then tests each peak within the MS1 spectrum to determine if it falls within the defined  $\Delta m/z$  and  $I_{\text{Rel}}$  boundaries. If all required mass spectral peaks within an isotope cluster are found, the observed pattern-matched isotope cluster is added to a list of hits.

Additional functionality allows the user to define the occurrence of particular isotopic peaks as either optional or forbidden. If an optional isotopologue is detected, then its presence is simply noted, whereas if a forbidden isotope is

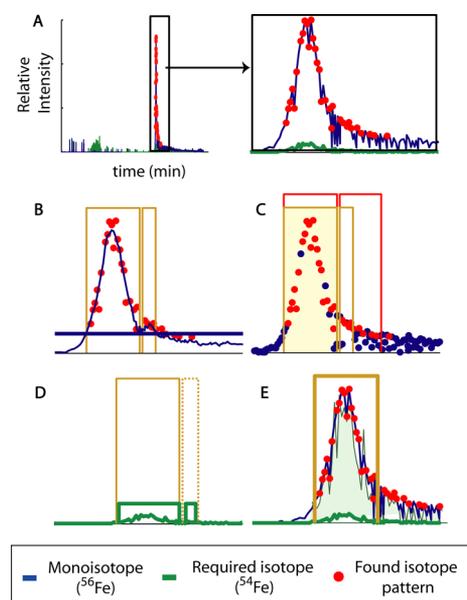


**Figure 1.** (A) Example of the definition of an isotope pattern for the detection of Fe complexes, consisting of mass difference ( $\Delta m$ ) and relative intensity ( $I_{\text{Rel}}$ ) windows between the  $^{56}\text{Fe}^{12}\text{C}$  monoisotopic species and a required  $^{54}\text{Fe}^{12}\text{C}$  isotopologue shown in red, as well as an optional  $^{56}\text{Fe}^{13}\text{C}$  isotopologue shown in blue. Mass differences to related free ligand species or other adducts, shown in yellow, can also be defined. Error margins for  $\Delta m$  and  $I_{\text{Rel}}$  are indicated by dotted lines around the position of the expected mass spectral signals. The algorithm then queries experimental mass spectra for the defined mass differences and relative intensities. (B) Mass spectrum of Fe-enterobactin ( $m/z = 723.0583$ ), including the algorithm's detection of the Fe isotope pattern and its recognition of the free ligand species.

detected, an otherwise matching isotope pattern is excluded from the list of pattern-matched hits. Using this feature, the user can make the isotope pattern search results much more specific to a given element (see also Table S-1, Supporting Information).

In a second step, the pattern-matched isotope clusters that span multiple MS1 spectra (i.e., different retention times) and possess the same charge states are grouped together if their monoisotopic species (i.e., the isotopologue that combines the principal isotope of each atom in the molecule, for small organic molecules, the monoisotopic species is the most abundant isotopologue) have the same  $m/z$  value within the defined measurement error. This grouping proceeds in order of decreasing intensity, in accordance with published chromatogram building algorithms.<sup>16</sup> Finally, using the  $m/z$  - intensity pairs for the signals in a given pattern-matched isotope cluster group, an intensity weighted  $m/z$  average is calculated for each isotopologue.

**Identification of Candidates for Pattern-Matched Chromatographic Features.** To identify isotope pattern-matched chromatographic features, extracted ion chromatograms (EICs) are calculated using the averaged  $m/z$  values of each isotopologue in a given pattern-matched isotope cluster. Chromatographic peaks that correspond to times of consecutive observed pattern-matched clusters are then identified, as described in Figure 2.



**Figure 2.** Detection of isotope pattern-matched chromatographic features. Shown is an example of  $^{54}\text{Fe}$  and  $^{56}\text{Fe}$  isotopologues associated with the Fe–prothochelin complex detected in *Azotobacter vinelandii* culture supernatants. (A) EICs are calculated with the averaged  $m/z$  values of each isotopologue in a given pattern-matched isotope cluster. (B) Peaks are then detected by baseline subtraction with moving-average smoothed EICs, whereby the baseline height is a user defined fraction  $F$  of the mean signal intensity for the isotopologue in the grouped pattern-matched isotope cluster. Based on samples analyzed in this study, we find that the default value of  $F = 1/3$  generally ensures the required sensitivity to detect the defined isotope pattern while excluding potential interferences from unrelated baseline species. The identified peaks are indicated by orange boxes. (C) In the next step, times of consecutive identified isotope clusters are connected when less than a given number of scans without a pattern match lie between two found isotope clusters (skip parameter, e.g.  $S = 4$ ) as shown by red lines. A peak is retained for further evaluation if its apex is part of a connected isotope cluster. If no peak meets this requirement, the isotope cluster group is marked by the program as having failed the chromatographic coherence test. (D) If a peak of any isotopologue is detected (shown in green boxes) that does not fall between the boundaries of an identified monoisotopic peak, and if this peak is detected within a time window of one peak width around the selected monoisotopic peak, then the peak of the monoisotopic species is extended to include the additional isotopologue peak. (E) The final chromatographic peak feature contains the peaks of all isotopologues in the cluster. The green area represents the EIC of the  $^{54}\text{Fe}$  isotopologue normalized to the  $^{56}\text{Fe}$  monoisotopic species to illustrate the chromatographic coherence of both isotopologues, which is evaluated in more detail in the next step of the algorithm.

**Evaluation of the Chromatographic Coherence of the Isotope Pattern.** Using the identified chromatographic feature boundaries, a set of parameters is calculated to further evaluate whether the isotope pattern is chromatographically consistent (Table 1). These parameters are then used as filters to eliminate false positives. The value of each parameter is compared to a predefined range that is based on evaluations of samples analyzed in this study. This range may be modified by the user, for example, to reduce the number of false positive features identified, at the expense of an increased likelihood of missing true positives. The program uses two ranges, a narrow range that accommodates the expected range of parameter

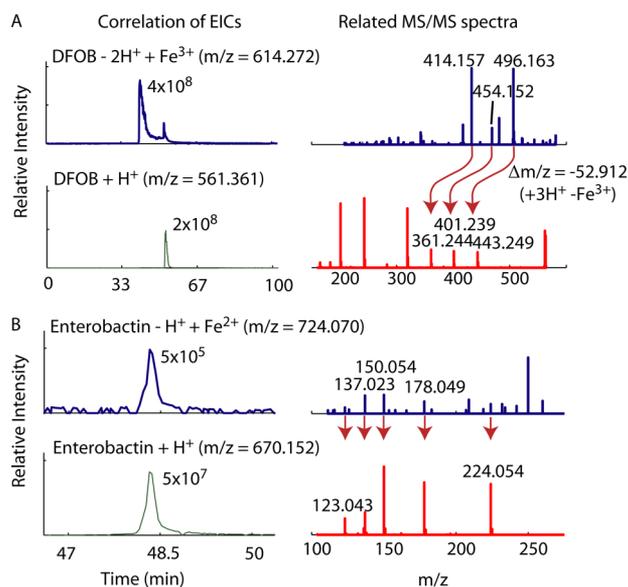
**Table 1. Parameters to Test the Chromatographic Consistency of the Isotope Pattern within Identified Chromatographic Features**

parameter	expected range <sup>a</sup>	wider range (outliers) <sup>a</sup>
number of pattern-matched isotope clusters per feature	$\geq 5$	$\geq 2$
fraction of scans with pattern-matched isotope clusters per feature	$> 0.5$	$> 0.25$
intensity fraction of monoisotopic signals that are part of a pattern-matched isotope cluster	$> 0.5$	$> 0.25$
intensity weighed average of observed $\Delta m/z$ for each required isotopologue	within user defined range	
intensity weighed average $I_{\text{Rel}}$ for each required isotopologue	within user defined range	below expected lowest intensity
Pearson correlation coefficient between signal intensities of each required isotopologue and the monoisotopic species	$\geq 0.7$	$\geq 0.4$

<sup>a</sup>The predefined range of parameter values is based on evaluations of samples analyzed in this study and may be modified by the user.

values and a wider range that is used to flag possible measurement outliers for later manual inspection. Thus, the chromatographic peak feature is annotated as having passed all tests if all parameter values fall within the narrow boundaries. If the wide range for any of the parameters is missed or if more than one parameter associated with an isotope misses the narrow range, then the peak group is flagged as having failed the chromatographic coherence test. In all other cases, if a parameter is within the wider range, the pattern-matched isotope cluster group is flagged for manual inspection by the user. The pattern-matched isotope cluster group is also flagged if a “forbidden” isotope is defined and found within the  $\Delta m/z$  and  $I_{\text{Rel}}$  boundaries and with a high Pearson correlation coefficient between the forbidden isotope and the monoisotope intensities ( $R \geq 0.7$ ).

**Detection of Related Free Ligands and Other User-Defined Adduct Species.** To find related free ligands or other user-defined adducts, the program uses the known charge state and  $m/z$  of the monoisotopic species and the  $\Delta m/z$  values associated with the user-defined related adducts to derive their EICs. A corresponding signal is tested for relatedness to the metal complex if its maximum signal intensity is above a user-defined minimum. Two complementary approaches are followed. One indication for relatedness is chromatographic coelution of the species with the metal complex (Figure 3, left panels), which may be the result of similar interactions of the species with the stationary LC phase, or arise due to in-source formation or dissociation of metal complexes. The program defines a free ligand or adduct to be related to the pattern-matched isotope cluster if it coelutes within the retention time window of the identified chromatographic feature with high intensity (summed intensity of adduct species  $> 2 \times$  summed intensity of the monoisotopic species in the pattern-matched isotope cluster) or if the intensities of the EICs of the adducts are strongly correlated to those of the monoisotopic species in the pattern-matched isotope cluster ( $R > 0.7$ ). A second, independent indicator of relatedness is the presence of common characteristic fragment ions or related fragment ions that differ by the presence of the metal in any provided MS/MS spectra associated with the species (Figure 3, right panels). To obtain related MS/MS features, the program creates a list of all fragment ions above a given noise level or above a given relative

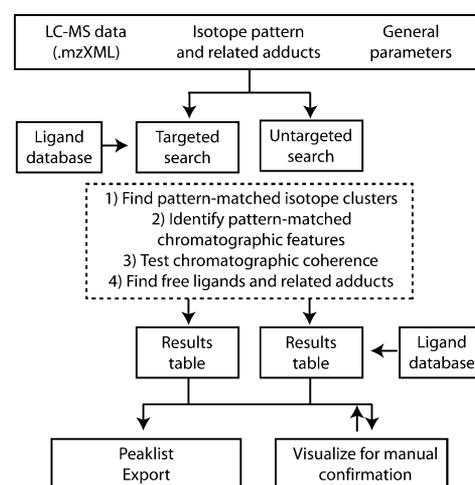


**Figure 3.** Identification of free ligands in the example of siderophore standards: (A) DFOB ([DFOB] = 50  $\mu$ M, [FeDFOB] = 50  $\mu$ M) and (B) enterobactin ([Fe-enterobactin] = 1  $\mu$ M) based on chromatographic correlation in EICs between the complex and the free ligand (left panels) and shared fragment ions or related fragment ions that differ by the presence of the metal in the MS/MS spectra (right panels).

intensity to the most abundant fragment that has a minimum  $m/z$  difference of 50 from the parent ion. This  $m/z$  difference has been chosen as large enough to exclude common neutral losses that are uninformative in terms of relatedness of the parent ions (e.g.,  $M-H_2O$ ,  $M-NH_3$ ,  $M-2H_2O$ , etc.), yet small enough so as not to exclude, for example, the loss of the smallest amino acid glycine from peptides by backbone fragmentation ( $\Delta m = 57$ ). The algorithm then tests whether the fragment ion lists of the two species have one or more signals at the same  $m/z$  value, or whether common  $m/z$  differences between the parent ion and the fragment ions are observed for both species. The MS/MS spectra of the metal complex and the free ligand may show common related ions that differ by the presence of the metal if the metal remains bound to the molecule during the fragmentation process, as seen in the example MS/MS spectra of Fe-complexed vs protonated deferoxamine B (DFOB) (Figure 3A). Conversely, common fragment ions may be observed when the metal is lost from the complex during fragmentation, as seen in the MS/MS spectra of Fe-complexed vs protonated enterobactin (Figure 3B). Filtering for coelution is particularly useful to identify free ligands of labile metal complexes, while high-quality MS/MS spectra can routinely be recorded for complexes that remain intact and may be measured with higher intensities.

## IMPLEMENTATION

**Input.** ChelomEx is implemented in MatLab (R2013b) and guides the user through the workflow shown in Figure 4, via a graphical user interface (Figures 5 and S-4–S-6, Supporting Information). The input consists of the LC-MS data files converted to the common, vendor-neutral mzXML format, the definition of the metal-associated isotope pattern and related adducts (Figure S-5, Supporting Information), and configuration of the above-mentioned general parameters, which are involved in the algorithm. Mass error windows for the isotope

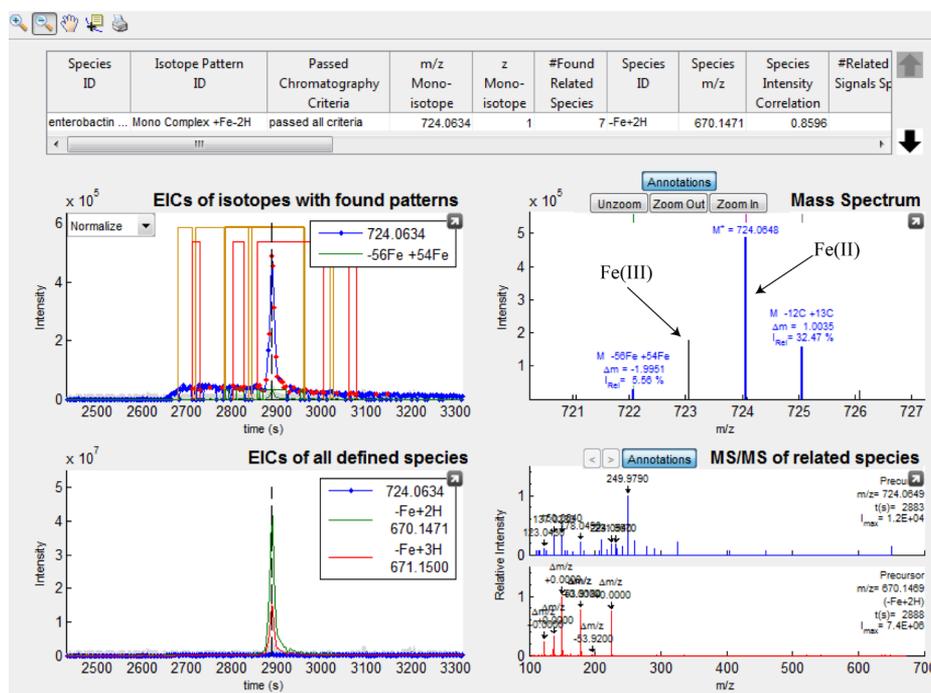


**Figure 4.** Schematic of the ChelomEx workflow.

pattern are given individually for each isotopologue as the sum of an absolute  $m/z$  offset (in amu) and a relative error in ppm of the  $m/z$  of the monoisotopic species. Associated with the relative error, a minimum absolute error may be defined (useful at low  $m/z$  values). The  $m/z$  offset allows for tolerance of metal isotopologues that may overlap with isotopologues of the main elements of the organic compound (e.g.,  $^{13}C$  in complexes with Mo, Figure S-3, Supporting Information) or with coeluting species that can have a mass difference of one proton due to different redox states of the metal center (e.g.,  $Fe^{2+}$  and  $Fe^{3+}$ ). The metal complex and its putative free ligand may be present as a family of different species. It is therefore possible to provide a list of mass differences between the monoisotopic species of the metal complexes and the free ligands that account for different redox states of the metal and various adducts to the ligand (e.g.,  $Na^+$  or  $NH_4^+$  adducts). The general parameter settings include a common mass error (relative ppm error and a minimum error at low  $m/z$  values) that is used to calculate all EICs.

**Processing in Targeted and Untargeted Modes.** After all inputs have been provided, the software can be run in targeted or untargeted modes. The targeted search uses a built-in compound library, which presently contains names, elemental compositions, and exact neutral masses of known siderophores. This database was assembled from an online siderophore database<sup>17</sup> and literature reviews of siderophores,<sup>18</sup> and may also be expanded by the user. The user then selects one or more of the defined adduct species that contain the mass difference between the free siderophore ( $M$ ) and the different possible redox states of the metal complex (e.g., in the case of iron:  $M-3H^+ + Fe^{3+}$  and  $M-2H^+ + Fe^{2+}$ ). For each entry in the database, the program calculates the  $m/z$  values of possible complexes, including 1:1, 2:1, and 3:1 ligand:metal complexes. In this case, for identification of pattern-matched isotope clusters, ChelomEx uses only the subset of all mass spectral peaks that have  $m/z$  values corresponding to the  $m/z$  values in the target list (Step 1 in Figure 5). In the results list that ChelomEx produces, the algorithm reports charge states, as well as mono-, bis-, or tris- complexes involving the same ligand, and all found complexes are annotated with the corresponding database entry.

In untargeted mode, all mass spectral peaks are included in the isotope pattern search, and all metal complexes that show the required isotope pattern are matched, irrespective of



**Figure 5.** Visualization of the results for manual inspection. Information about the selected pattern-matched chromatographic feature is at the top of the window; the four panels below the table consist of (clockwise from the top right) individual mass spectra displaying the pattern-matched isotope cluster, the aligned MS/MS spectra of the metal complex and any found related species, the EICs of the primary feature and all found related species, and the EICs of the isotopologues that make up the pattern-matched feature. The mass spectrum in the top right panel includes a manual annotation of the redox speciation of the metal center inferred from the mass difference between the monoisotopic species of the metal complex and the free ligand.

whether the mass of their ligand corresponds to a database entry. To facilitate the annotation of known compounds in the list of results, the compound library may be expanded with new entries and interactively searched. The database search may include various adducts and derivatives of the known ligands, as well as dimers, trimers, etc. (Figure S-6, Supporting Information)

**Manual Confirmation of Results and Export of Peaklists.** Because the program selectively returns only a small subset of all chromatographic features that have a defined characteristic isotopic signature, the manual inspection option is an important feature in the software and provides a means to confirm the results of the algorithm and inspect marginal or nonconforming features that were flagged for manual inspection by the algorithm (Figure 5). Confirmed pattern-matched chromatographic features can be annotated manually in the results table within the software. Finally, a peaklist with identified peaks can be exported as an ASCII formatted table or via copy-paste from the results table in the graphical user interface (Figure S-4, Supporting Information). The peaklists can then be used for further targeted processing (e.g., sample alignment and statistical analyses) with available software, such as with XCMS,<sup>19</sup> MZmine 2,<sup>20</sup> or Maven,<sup>21</sup> or they can be used as an input for additional targeted LC-MS/MS analyses.

## APPLICATION EXAMPLE

**Methods. Sample Preparation and High-Resolution Nano-UPLC-MS.** The samples included Fe complexes with the siderophore standards deferrioxamine B (DFOB) and enterobactin, as well as extracts of sterile growth media (control) and conditioned media from Fe-limited *A. vinelandii* strain OP cultures. The extraction was performed on reversed-

phase columns (Amberlite XAD-16 and Oasis HLB) and methanolic (50% and 100%) eluates were lyophilized before resuspension and injection on a high-resolution high-mass-accuracy reversed-phase nano-UPLC LTQ-Orbitrap XL or LTQ-Orbitrap Velos platform. Further details on sample preparation and measurement are given in the Supporting Information.

**Preprocessing of Data, Definition of Isotope Patterns and Input Parameters.** Raw LC-MS data collected in profile mode were centroided and converted from the Thermo .raw format to .mzXML using the open-source software tool MSConvert which is part of the ProteoWizard toolkit.<sup>22</sup> The data set was further processed to connect MS/MS spectra with their high-resolution precursor masses using Proteome Discoverer (v1.4). The output file with MS/MS data in the .mgf format was also converted to .mzXML using MSConvert and both .mzXML files were imported into ChelomEx.

The error ranges in the example of the Fe isotope pattern measurement were evaluated experimentally with the Fe-protochelin complex. For isotope intensities that were at least 10× above the noise level ( $I_{\text{noise}} \sim 3000$  counts), the observed average mass difference  $\Delta m/z$  between  $^{54}\text{Fe}$ -Protochelin and  $^{56}\text{Fe}$ -Protochelin agreed with the expected value within  $\pm 0.5$  ppm, and the deviation from  $I_{\text{Rel}}$  was <15%. Closer to the noise level, the errors increased significantly. Here we use a common relative mass error of 2 ppm (with a minimum absolute error of 0.000 75 amu at low  $m/z$  values) for all isotopologues and allow for a relatively large error in  $I_{\text{Rel}}$  (Table S-2, Supporting Information).

Higher errors are possible for metals with isotopologues that are unresolved from the isotopologues of the elements of the organic ligand (e.g., Mo, Figure S-3, Supporting Information)

and for metals that may be present at different redox states. Therefore, an isotopologue-specific constant mass error is added to the common relative error, based on simulated isotope patterns for the metal complexes with protochelin at the experimental resolution (see the Supporting Information for details). The specific error tolerances for the isotope patterns of Fe, Mo, and Zn complexes along with defined free ligands and other related adducts are given in Tables S-2 and S-3 (Supporting Information). As further input parameters, we define singly and doubly charged species to be included in the search and set the general  $m/z$  error to 3 ppm, with a minimum absolute  $m/z$  error of 0.002 amu. Other general parameters are given in the Supporting Information in Table S-4.

**Analysis of Siderophore Standards.** We demonstrate the approach of ChelomEx for screening LC-MS data for characteristic metal-isotope patterns using two different classes of siderophore standards: the hydroxamate siderophore DFOB and the catechol siderophore enterobactin (for structures, see Figure S-1, Supporting Information). The algorithm successfully identifies both Fe complexes, based on their isotope patterns, and links the complexes to free ligand species in the sample because of coelution and common MS/MS fragment ions (enterobactin) or related fragment ions differing by the presence of the metal (DFOB) (Figure 4). The observed Fe-complex-to-free-ligand peak height ratios suggest strong complexation of Fe to DFOB, while enterobactin appears to be present mainly as the free ligand, in agreement with the expected lability of the Fe–enterobactin complex at the low pH of the formic acid mobile phase buffer (pH  $\sim$  2.7) in this study.<sup>18</sup>

Using the mass differences between the metal complexes and the free ligands, the algorithm finds Fe in DFOB as Fe<sup>3+</sup> ( $\Delta m(-\text{Fe}^{3+} + 3\text{H}^+) = -52.9115$ ), while the Fe in the enterobactin complex is present as both Fe<sup>2+</sup> ( $\Delta m(-\text{Fe}^{2+} + 2\text{H}^+) = -53.9193$ ) and Fe<sup>3+</sup> (Figure 5). In the case of enterobactin, both redox species are linked to the same free ligand. The occurrence of both redox species may be explained by the low binding strength of enterobactin at low pH values and a corresponding positive shift in the redox potential.<sup>23</sup>

Interestingly, the ChelomEx search reveals the presence of several additional Fe complexes present at low abundances in the DFOB standard, with masses that correspond to other ferrioxamine species (e.g., ferrioxamine H, A1, A2, D1, acyl-ferrioxamine 1). The additional compounds could be enriched and partially isolated by high-pressure liquid chromatography, which confirmed their presence as contaminants in the DFOB standard and ruled out potential measurement artifacts. The MS/MS spectra of these species were in agreement with the structures of the putative ferrioxamines. The abundances of the additional ferrioxamines were more than 10 $\times$  below that of DFOB, and their Fe complex isotope patterns were close to the detection limit, suggesting that these species were minor contaminants in the DFOB standard, previously unreported presumably due to the difficulty of their detection without high-resolution LC-MS and the aid of a dedicated software tool. Since our commercial DFOB standard is purified from microbial cultures, it is possible that the additional ferrioxamines were also biologically produced and coisolated with DFOB.

**Untargeted Analysis of Metal Complexes in Bacterial Media Extracts.** The acid-lability of Fe complexes with catechol siderophores makes them challenging to analyze by positive mode reversed-phase electrospray LC-MS, due to the

acidic mobile phases that are typically used. To test the performance of ChelomEx in identifying these compounds in a complex matrix, we compared media extracts of cultures with the nitrogen fixing bacterium *A. vinelandii* to extracts of the sterile growth medium (control). *A. vinelandii* produces three catechol siderophores with known structures and molecular masses within the instrumental scan range (Azotochelin, Protochelin, and Azotobactin; for structures, see Figure S-2, Supporting Information). We mined the LC-MS data for the characteristic isotope patterns of Fe, Mo, and Zn complexes potentially present in the sample. All three known *A. vinelandii* siderophores were identified as Fe complexes by ChelomEx, along with additional complexes that are only detected in the conditioned media extracts, but not in the control samples (Table 2). Consistent with the high affinity of the known *A.*

**Table 2. Number of Identified Pattern Matched Isotope Cluster Groups and Peak Features for Fe, Mo, and Zn with Extracts of Conditioned Media from Fe-Limited *A. vinelandii* (A.v.) Cultures and Comparison to the Sterile Growth Medium (ctrl.)**

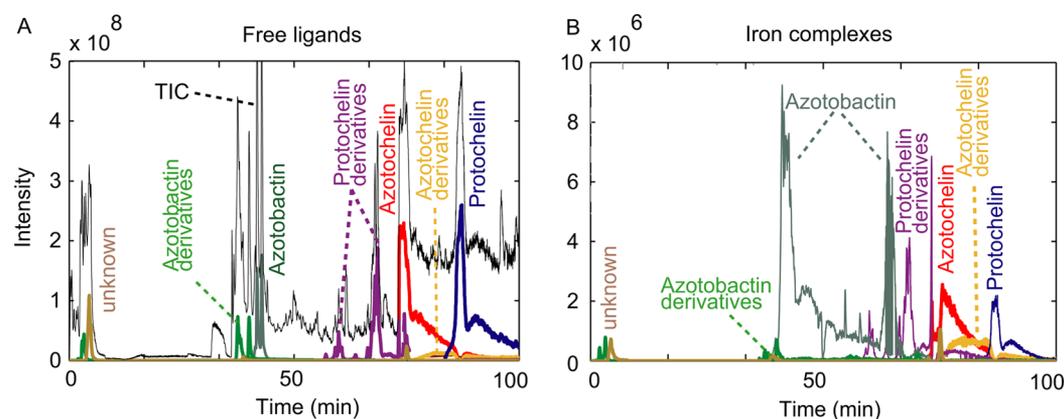
	Fe		Mo		Zn	
	A.v.	ctrl.	A.v.	ctrl.	A.v.	ctrl.
$\geq 1$ pattern-matched isotope clusters	1630	238	12	6	46	5
$\geq 5$ pattern-matched isotope clusters	745	48	8	0	21	4
pattern-matched chromatographic features	519	90	28	3	14	2
flagged for inspection	144	22	6	0	8	1
passed tests	99	5 <sup>a</sup>	1	0	3	0

<sup>a</sup>3 of the 5 species are cross contamination from A.v. present at low intensities (free ligand  $I_{\text{Control}}/I_{\text{A.v.}} < 1\%$ )

*vinelandii* siderophores for Fe, we found that by far the largest number of identified complexes were associated with Fe. Nevertheless, some compounds were also detected with isotope patterns that indicate Mo and Zn complexation. In this context, the *A. vinelandii* siderophores have been previously reported to have multiple roles and also bind Mo<sup>24</sup> and likely Zn,<sup>25</sup> yet with lower affinity than Fe.

**Discrimination of Isotope Patterns and Chromatographic Features.** In the case of the Fe isotope pattern searches, a large number of apparent isotope clusters were matched in the conditioned media and in the control sample that did not show chromatographic correlation, caused by the coincidental co-occurrence of mass spectral signals that are within the specified error windows of the isotope pattern (e.g., Figure S-7, Supporting Information). The number of false positive isotope patterns in the conditioned media and the control sample is larger for Fe compared to Mo and Zn due to the fact that the distinguishing isotope pattern for Fe is defined by the presence of only one isotopologue in addition to the monoisotopic species, while those for Mo and Zn include several isotopologues (Tables S-1 and S-2, Supporting Information). The false positive matches to Fe complexes in the spent media were sequentially eliminated by the peak feature detection and chromatographic coherence tests, and essentially all species that passed these tests were validated by manual inspection as correctly annotated.

A previously described procedure for the isotope-assisted discovery of Fe siderophore complexes<sup>14</sup> did not include the



**Figure 6.** Results for the analysis of Fe complexes in *A. vinelandii* spent media showing (A) EICs of the most abundant free ligands together with the total ion chromatogram (TIC) and (B) EICs of the most abundant Fe complexes. The Fe complexes represent minor species under the experimental conditions, exemplative of species only feasibly discoverable through the use of a systematic software tool. The distribution of Fe complexes and free ligands that is revealed by ChelomEx allows the user to prioritize them for further analysis. The found compounds include all known *A. vinelandii* siderophores within the mass spectral scan range and additional putative relatives of the known siderophores, as indicated by their masses and related MS/MS spectra.

identification of chromatographic peak features. Therefore, all acquired mass spectra were used to calculate a set of parameters to eliminate false positives. These parameters included  $\Delta m/z$ , and  $I_{Rel}$  values and the intensity correlation between the  $^{54}\text{Fe}$ – $^{56}\text{Fe}$  isotopologue pair. By toggling off the chromatographic peak detection and modifying the chromatographic consistency filters in Table 1, we can force ChelomEx to behave similarly for comparison. The number of authentic isotope patterns in the *A. vinelandii* sample identified in this way decreased by  $\sim 50\%$  due to the increased inclusion of signals that overlay or are isobaric with the metal complex isotopologues, and which are interfering in the  $\Delta m/z$ ,  $I_{Rel}$ , and isotope intensity correlation calculations. At the same time, additional false positives were also included in the identified patterns ( $\sim 40\%$ ), defined by correlated intensities of the putative isotopologues that are not linked to consecutive matched isotope pattern clusters. Still, there are some ( $\sim 10\%$ ) isotopologue intensity correlations that were stronger for authentic metal complexes without the chromatographic peak detection enabled, because the algorithm of ChelomEx excludes low intensity regions of the peaks, such as those caused by chromatographic tailing, to reduce potential interferences. These species, however, were still made visible to the user as part of the group of compounds flagged by the software for manual inspection in the graphical user interface. Thus, the identification of pattern-matched chromatographic features by ChelomEx significantly enhanced the sensitivity and the discrimination of signals associated with authentic metal complexes from false positive features.

**Free Ligands and Related Adducts.** As described above (see the Approach section), the software links the monoisotopic species in a pattern-matched chromatographic feature to mass spectral signals that match  $m/z$  values of theoretical free ligand species. For the analysis of the *A. vinelandii* conditioned media sample, 65 out of the total of 99 Fe–complex chromatographic features were associated with coeluting free ligand species. Five of these were matched based on their related high-resolution MS/MS spectra; this relatively small number was a consequence of the low Fe-complex abundances for the putative catechol siderophores under the experimental conditions (pH  $\sim 2.7$ ), which limited the collection of high-

quality MS/MS spectra for these species. The link between putative complexes and free ligands that ChelomEx helps to reveal is powerful, as it can provide further confidence in the assignment of metal complexes, which may be minor species by comparison to the free ligands, as illustrated in Figure 6. In that way, it becomes possible for the user to get an overview of the most prominent species in the chromatogram and, on a selective basis, confirm the software results or inspect features that are flagged for manual inspection.

**Targeted Analysis.** To prioritize the analysis of known chelating agents of interest, ChelomEx includes a targeted analysis feature that is based on a built-in library that contains exact masses and chemical compositions of known siderophores and any additional user defined compounds. The results (Figure 6) show all three known siderophores produced by *A. vinelandii* are among the species with the highest intensities. The results from the targeted analysis can also inform further analysis of unidentified compounds that may be related to the known species.

**Discovery of New Siderophores.** As can be seen in Table 2, the number of putative Fe complexes discovered in the untargeted analysis is much larger than the three siderophores previously known to be produced by *A. vinelandii* within our mass spectral scan range. A number of them likely are related to the known siderophores. One reason for this is that each siderophore complex may appear as several species (e.g., bis-complexes, mixed ligand complexes, different charge and redox states) or undergo in-source fragmentation reactions, leading to observed coeluting complexes. Additionally, reactions in the sample (e.g., hydrolysis, oxidation, etc.) may potentially result in derivatization of the known compounds. Some species may also be novel biologically produced siderophores. Here we focus on the species with the highest peak intensities of the free ligand and the highest peak intensities of the metal complexes. The additional found Fe complexes and free ligands possess mass differences with respect to the known siderophores that suggest that they are close relatives (e.g., M-2H, M-2H+O, M-CH<sub>2</sub>, etc.), a hypothesis that is further substantiated by related MS/MS spectra between the free ligands of the newly found species and the known siderophores (Figure 6 and Tables S-5 and S-6, Supporting Information). For example, we find an Fe isotope pattern ( $m/z = 664.147$ ) that corresponds to a putative

protochelin ( $m/z = 678.163$ ) relative with a mass difference of one  $\text{CH}_2$  group ( $\Delta m/z = -14.016$ ). The intensity of the Fe complex is close to the detection limit of the isotope pattern so that a systematic discovery of such species in our dataset is only feasible with the aid of a software tool, such as ChelomEx. The corresponding unbound ligand is present at roughly 20 $\times$  higher abundance. The MS/MS spectra of the new siderophore and of protochelin show common fragment ions and related fragments that differ by the mass of the  $\text{CH}_2$  group. This confirms that the new compound is most likely a protochelin relative and allows one to suggest the position of the missing  $\text{CH}_2$  group in this new species (Figure S-8, Supporting Information).

## CONCLUSIONS

ChelomEx is a useful new tool for the discovery and study of metal chelation in natural, biological, and laboratory-derived media. The results of these studies can give new insights into the biological availability and cycling of trace metals in the environment as well as strategies of microbial communities to cope with trace-metal limitation and toxicity. New insights may also be gained in other research fields with an interest in metal speciation, ranging from areas of food and health research to metal-catalyzed organic synthesis.

## ASSOCIATED CONTENT

### Supporting Information

Supporting Information is available as noted in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*O. Baars. E-mail: [obaars@princeton.edu](mailto:obaars@princeton.edu).

### Present Address

<sup>§</sup>Department of Chemistry, Princeton University, Princeton, New Jersey 08544

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Y. Xu (Stanford University) and X. Zhang (Princeton University) for assistance with bacterial cultures. LC-MS measurements were facilitated by S. Kyin and H. Shwe (Princeton University). Funding for this study was provided by the United States National Science Foundation [OCE 1315200 to FM3].

## REFERENCES

- (1) Boiteau, R. M.; Fitzsimmons, J. N.; Repeta, D. J.; Boyle, E. A. *Anal. Chem.* **2013**, *85*, 4357–4362.
- (2) Popp, M.; Hann, S.; Koellensperger, G. *Anal. Chim. Acta* **2010**, *668*, 114–129.
- (3) Xue, H.; Sigg, L. In *Environmental Electrochemistry*; Taillefert, M., Rozan, T. F., Eds.; ACS Symposium Series; American Chemical Society: Washington DC, 2002; Vol. 811, pp 336–370.
- (4) Essen, S. A.; Bylund, D.; Holmstrom, S. J. M.; Moberg, M.; Lundstrom, U. S. *Biometals* **2006**, *19*, 269–282.
- (5) Gledhill, M.; Buck, K. N. *Front. Microbiol.* **2012**, *3*, 69.
- (6) Vraspir, J. M.; Butler, A. *Annu. Rev. Mar. Sci.* **2009**, *1*, 43–63.
- (7) Francis, M. B.; Jacobsen, E. N. *Angew. Chem., Int. Ed.* **1999**, *38*, 937–941.

- (8) Marshall, G. R.; Reddy, P. A.; Schall, O. F.; Naik, A.; Beusen, D. D.; Ye, Y.; Slomczynska, U. In *Advances in Supramolecular Chemistry*; Gokel, G. W., Ed.; Cerberus Press: Miami, 2002.
- (9) Patti, G. J.; Yanes, O.; Siuzdak, G. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263–269.
- (10) Rojas-Cherto, M.; Kasper, P. T.; Willighagen, E. L.; Vreeken, R. J.; Hankemeier, T.; Reijmers, T. H. *Bioinformatics* **2011**, *27*, 2376–2383.
- (11) Gledhill, M. *Analyst* **2001**, *126*, 1359–1362.
- (12) Mawji, E.; Gledhill, M.; Milton, J. A.; Tarran, G. A.; Ussher, S.; Thompson, A.; Wolff, G. A.; Worsfold, P. J.; Achterberg, E. P. *Environ. Sci. Technol.* **2008**, *42*, 8675–8680.
- (13) Velasquez, I.; Nunn, B. L.; Ibsanmi, E.; Goodlett, D. R.; Hunter, K. A.; Sander, S. G. *Mar. Chem.* **2011**, *126*, 97–107.
- (14) Lehner, S. M.; Atanasova, L.; Neumann, N. K. N.; Krska, R.; Lemmens, M.; Druzhinina, I. S.; Schuhmacher, R. *Appl. Environ. Microbiol.* **2013**, *79*, 18–31.
- (15) Zhu, P.; Tong, W.; Alton, K.; Chowdhury, S. *Anal. Chem.* **2009**, *81*, 5910–5917.
- (16) Benton, H. P.; Wong, D. M.; Trauger, S. A.; Siuzdak, G. *Anal. Chem.* **2008**, *80*, 6382–6389.
- (17) Bertrand, S. [http://bertrandsamuel.free.fr/siderophore\\_base/index.php](http://bertrandsamuel.free.fr/siderophore_base/index.php), 2014.
- (18) Hider, R. C.; Kong, X. L. *Nat. Prod. Rep.* **2010**, *27*, 637–657.
- (19) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.
- (20) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. *BMC Bioinf.* **2010**, *11*, 395.
- (21) Melamud, E.; Vastag, L.; Rabinowitz, J. D. *Anal. Chem.* **2010**, *82*, 9818–9826.
- (22) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. *Bioinformatics* **2008**, *24*, 2534–2536.
- (23) Harrington, J. M.; Crumbliss, A. L. *Biometals* **2009**, *22*, 679–689.
- (24) Bellenger, J. P.; Arnaud-Neu, F.; Asfari, Z.; Myneni, S. C. B.; Stiefel, E. I.; Kraepiel, A. M. L. *JBIC, J. Biol. Inorg. Chem.* **2007**, *12*, 367–376.
- (25) Huyer, M.; Page, W. J. *Appl. Environ. Microbiol.* **1988**, *54*, 2625–2631.