Lifeline:

Social Online Spectrometry

Steve Stein Yuri Mirokhin Manor Askenazi

Lifeline-S.O.S: 'Crowd Curation' of Unidentified GC-(EI)MS spectra through Social Online Spectrometry

Manor Askenazi[1], Yuri Mirokhin[2], Stephen Stein[2]

[1]*Biomedical Hosting LLC*, Arlington, MA [2]*NIST*, Gaithersburg, MD

email: lifeline@biomedical-hosting.net **website: http://lifeline.biomedical-hosting.net** **blog: http://blogline.biomedical-hosting.net**

Introduction

The physicochemical relationship between molecular structures and their EI-MS spectra is famously both reproducible and opaque (difficult to predict): whereas reproducibility has underpinned the success of mass spectral libraries, opacity has stymied de-novo structure elucidation for decades. Consequently, unmatched spectra are simply ignored by most identification pipelines. In an effort to gather, prioritize and elucidate unidentified EI spectra we are introducing 'Lifeline' -- a feature of the NIST MS Search software enabling users to submit unidentified spectra to the Lifeline-SOS website where they will be scored for quality, classified into broad molecular category and assigned to appropriate spectral clusters. The submitter can then exchange hypotheses with other cluster members as well as annotations for the individual peaks comprising the cluster centroid.

Methods

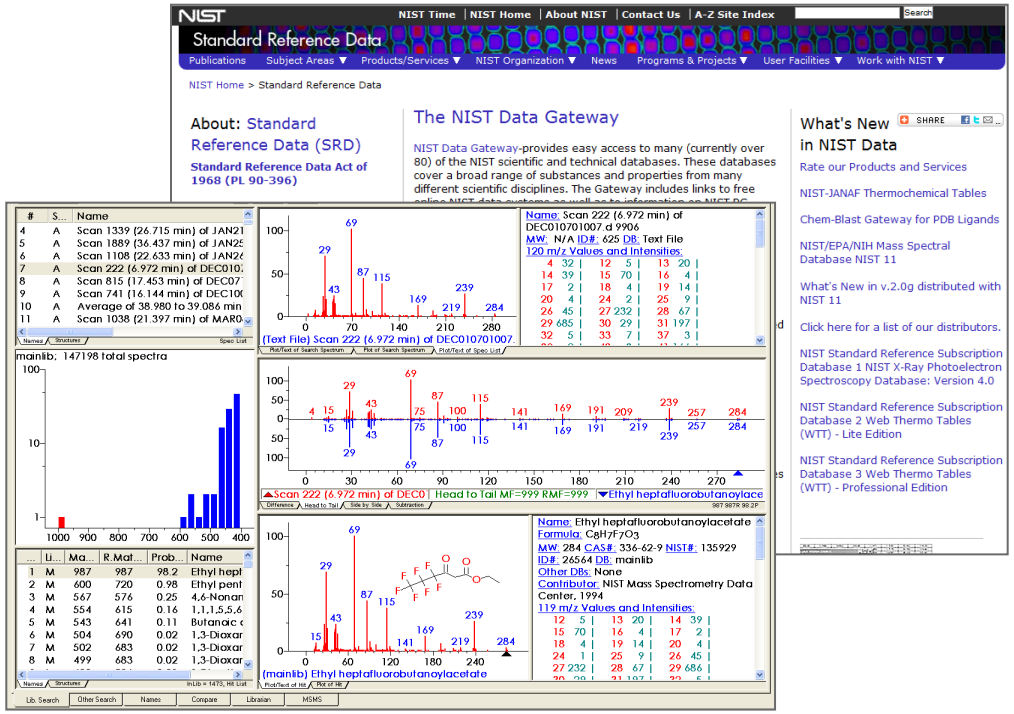
In contrast to other online spectral repositories, Lifeline-S.O.S is intended entirely for individual, unidentified spectra. Furthermore, the website emphasizes communication over algorithmic 'cleverness' by seeking to make submitters aware of other researchers which may be stymied by the same unknown spectra. Importantly, the submitters are not required to disclose their identity nor the identity of the lab in which the spectrum was acquired, neither are they required to provide an exhaustive characterization of the sample or experimental context of the originating acquisition. Users will be able to register for alerts regarding their submitted spectrum (such as a matching submission by another user or the posting of a hypothesis/annotation by another cluster-member) and are, of course, encouraged to participate in the cooperative elucidation process...

Preliminary Results/Abstract

The ability to submit unidentified spectra to the Lifeline-SOS website has been incorporated directly into the latest release (14) of NIST's MS Search software (this feature is currently in alpha and is disabled by default– if you would like to enable it, please email a request to the following address: **lifeline@biomedical-hosting.net**). We expect that a facile submission process with essentially no (compulsory) reporting requirements, will encourage the elicitation of challenging spectra from the user community. Submissions to the Lifeline-SOS site will, following an aggregation and curation process, form the basis for an 'open source' spectral library, similar to NIST's peptide spectral library. Furthermore, the library will serve as the default destination for recurring and unidentified spectra generated by NIST's internal spectral acquisition efforts, e.g. the site has been pre-populated with recurrent unidentified spectra generated from internal metabolic samples. Nevertheless, the project is in its infancy and one of our goals at this juncture is to validate the overall design of the system as well as its underlying principles: Does the system provide sufficient feedback to justify the submission of spectra? Is an emergent community, assembled by a common spectral cluster, more effective at spectral identification than the individual members on their own? Does the promise of cooperative prioritization and elucidation constitute sufficient incentive for the users to curate their submissions with e.g. information regarding the experimental context or the nature of the sample? More generally, will users be willing to communicate unidentified data when not mandated by e.g. the academic publication process?

Novel Aspect

A website accessible directly through NIST's MS Search software, enabling submitters to communicate, prioritize and ultimately elucidate unidentified GC-(EI)MS spectra.



NIST Spectral Library
GC-EI/MS, 1992

(a brief history…)

Public Law 90-396
90th Congress, H. R. 6279
July 11, 1968

An Act

To provide for the collection, compilation, critical evaluation, publication, and sale of standard reference data.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

DECLARATION OF POLICY

SECTION 1. The Congress hereby finds and declares that reliable standardized scientific and technical reference data are of vital importance to the progress of the Nation's science and technology. It is therefore the policy of the Congress to make critically evaluated reference data readily available to scientists, engineers, and the general public. It is the purpose of this Act to strengthen and enhance this policy

Standard Reference Data Act
H.R. 6279, 1968

NIST

(a brief history…)

United States Constitution
Article 1, Section 8, 1787

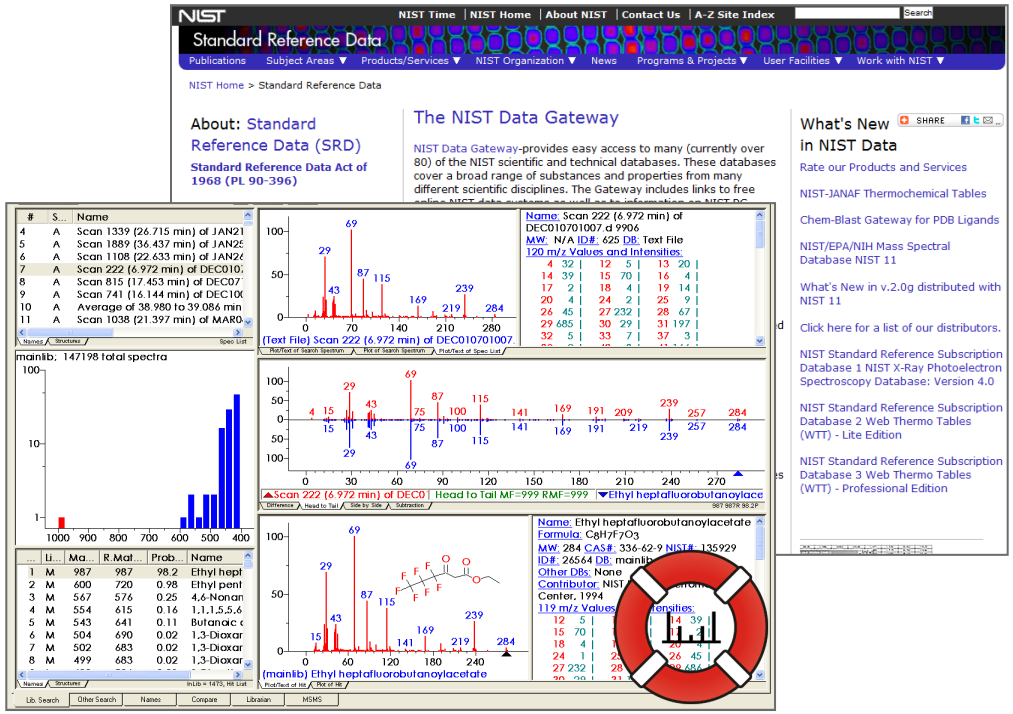
(a brief history…)

"מֹאזְנֵי צֶדֶק אַבְנֵי־צֶדֶק … יִהְיֶה לָכֶם"

"Just balances, just weights … shall ye have"

Leviticus, 19:35

(a brief history…)



NIST Spectral Library release 14
276,248 GC-MS Spectra, <u>today</u>

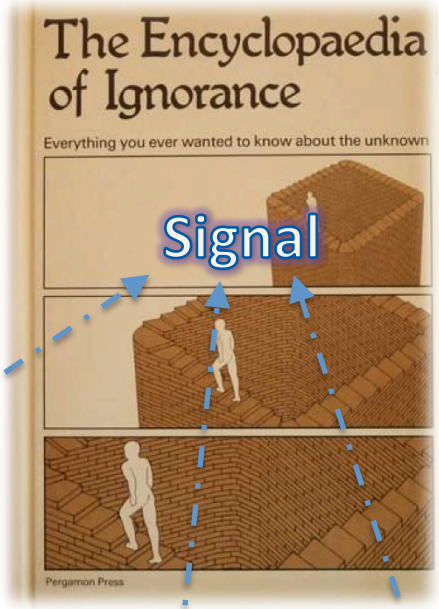
NIST

(back to the present time…)

identified

unidentified

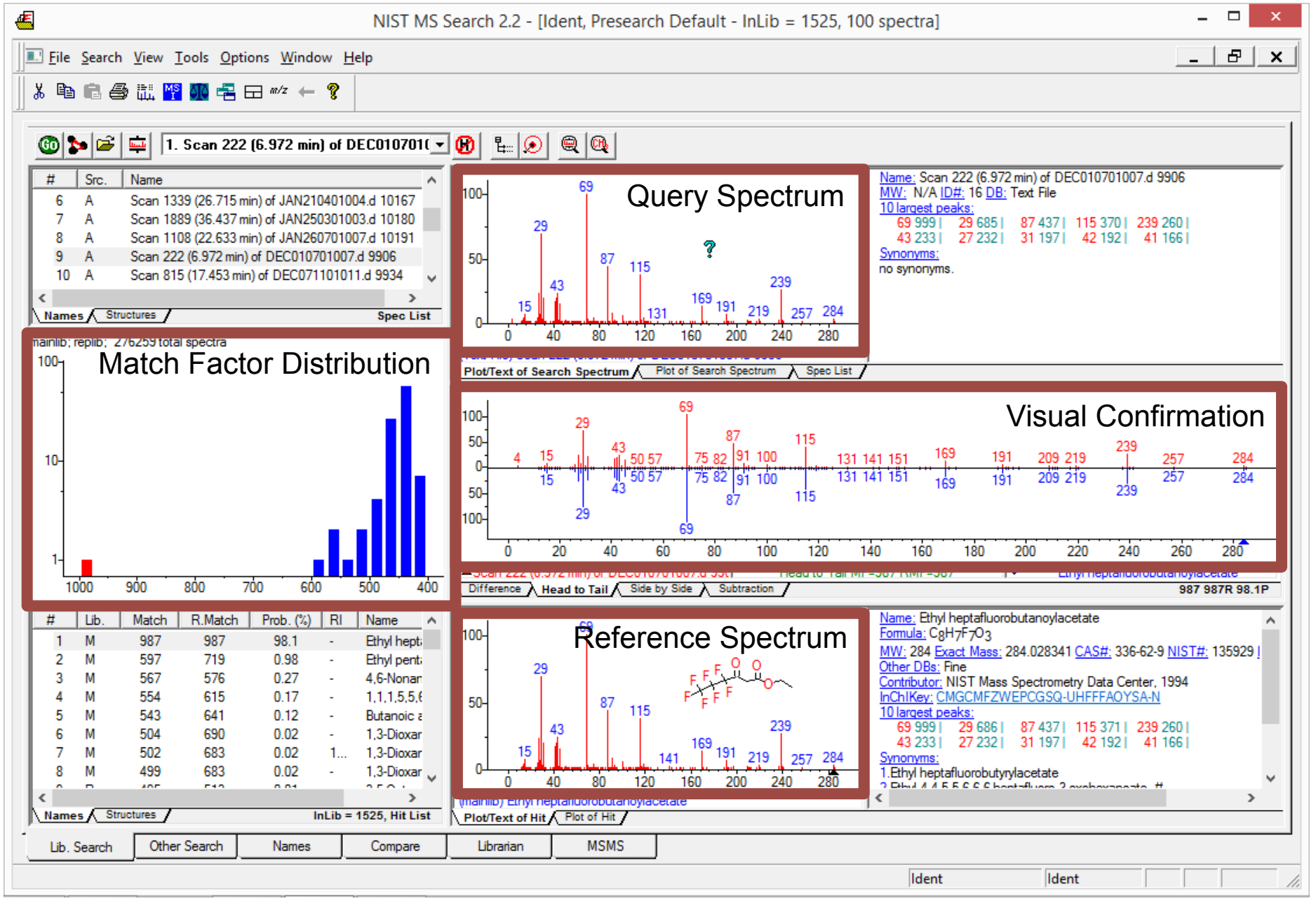
reproducible



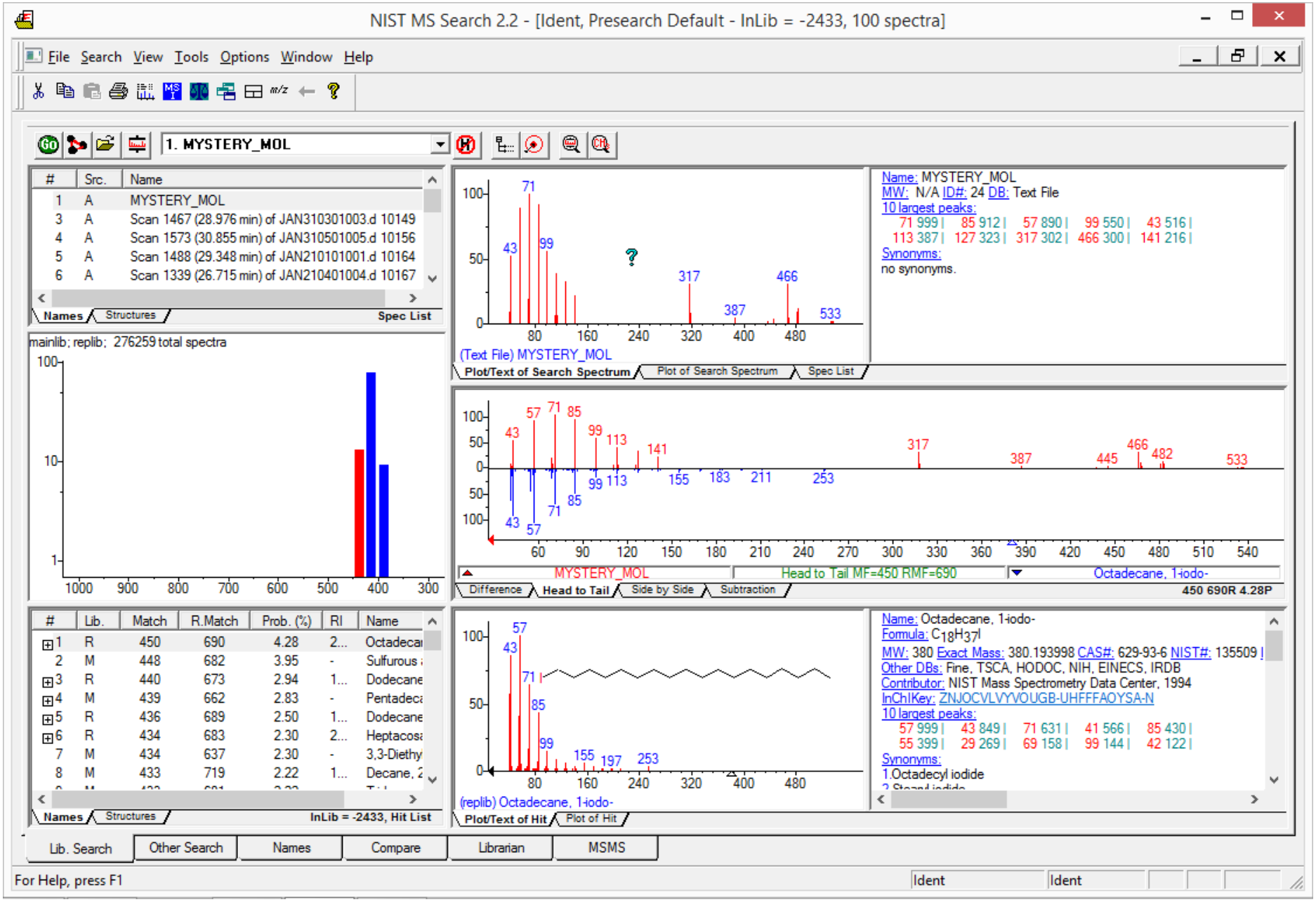
irreproducible

Noise Noise Noise Noise Noise Noise Noise Noise Noise Noise Noise Noise Noise Noise Noise Noise Noise Noise

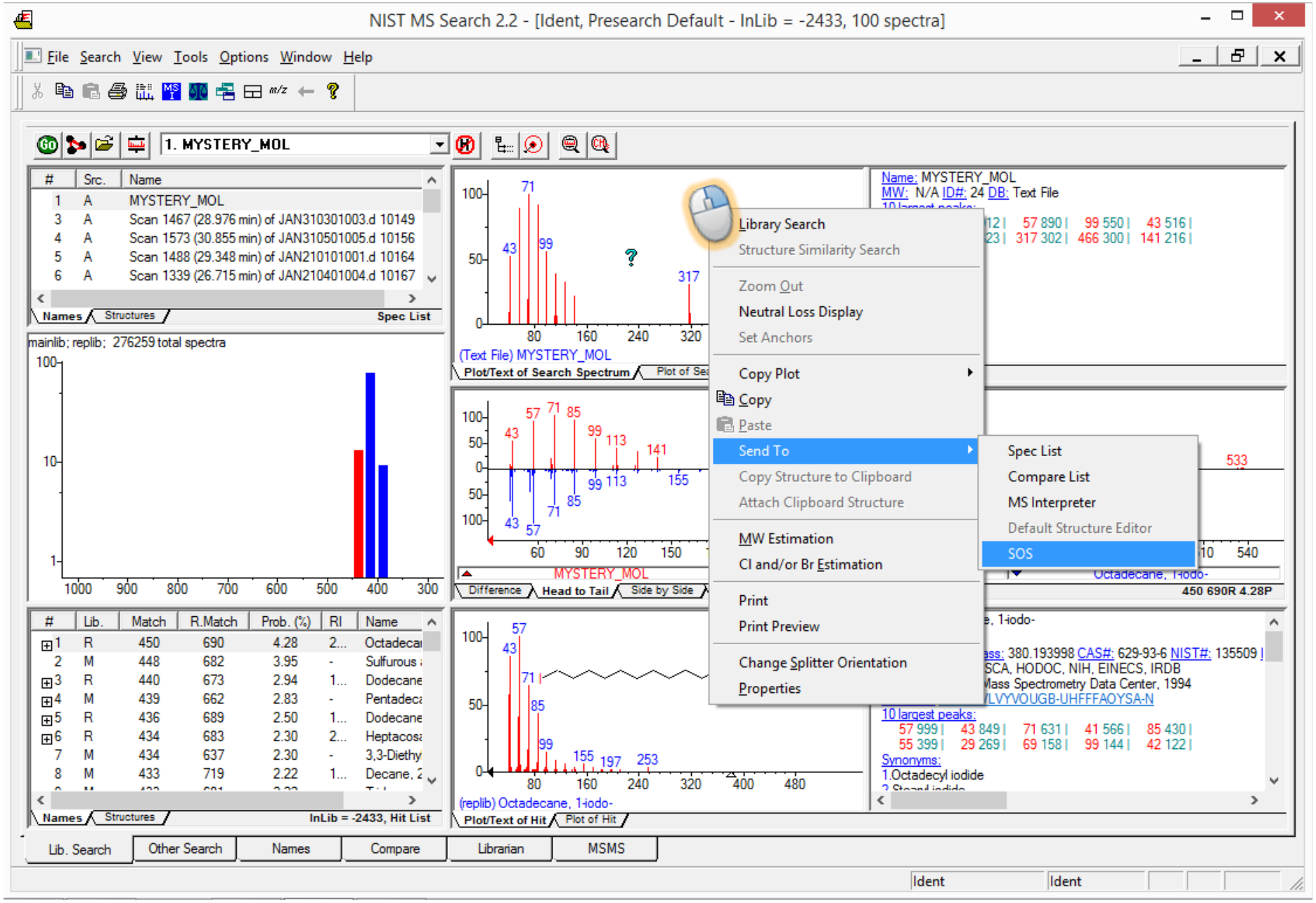
(user-submitted spectra will help to “rescue” unidentified signal from noise…)



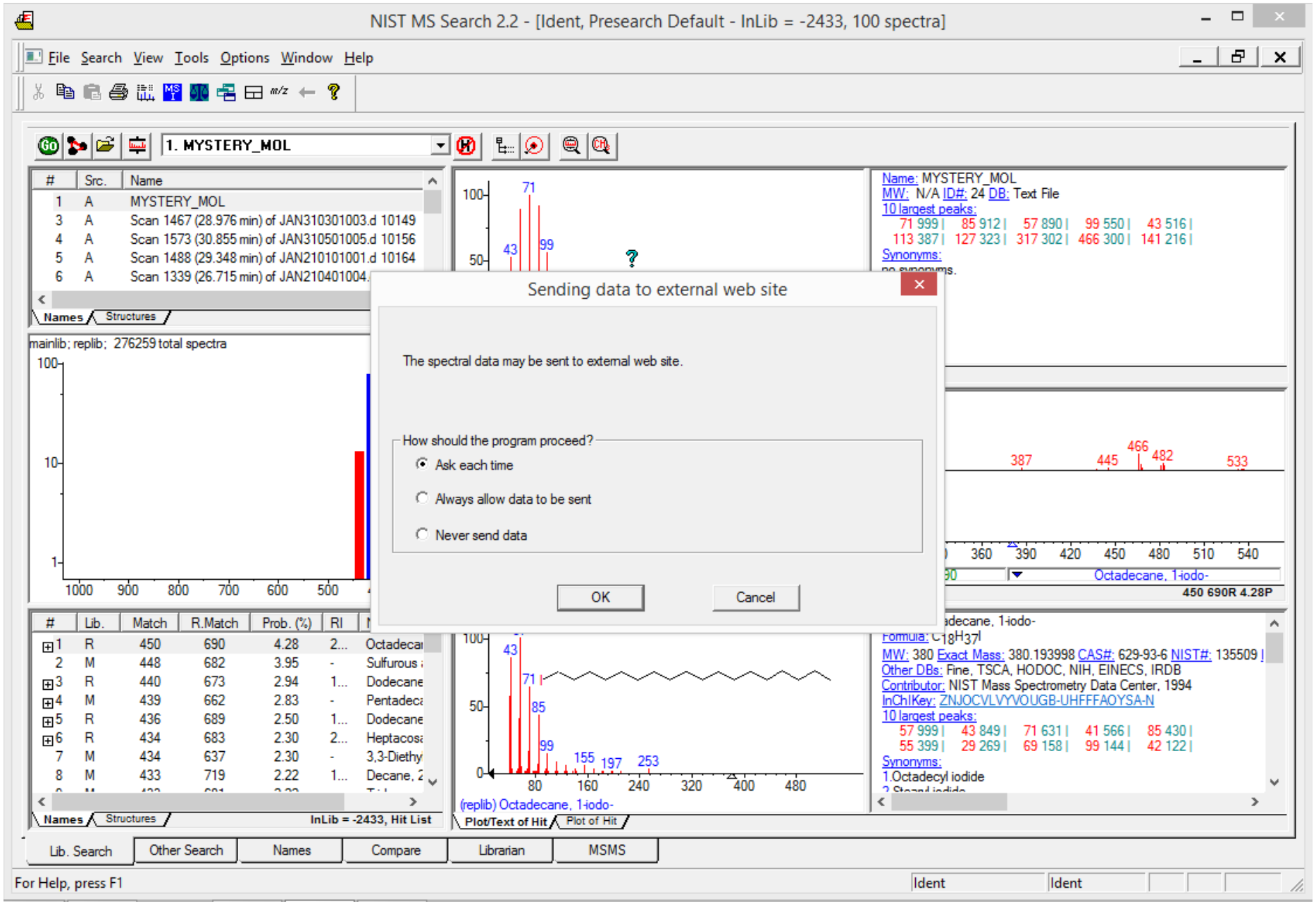
(NIST MS Search when things are going well…)



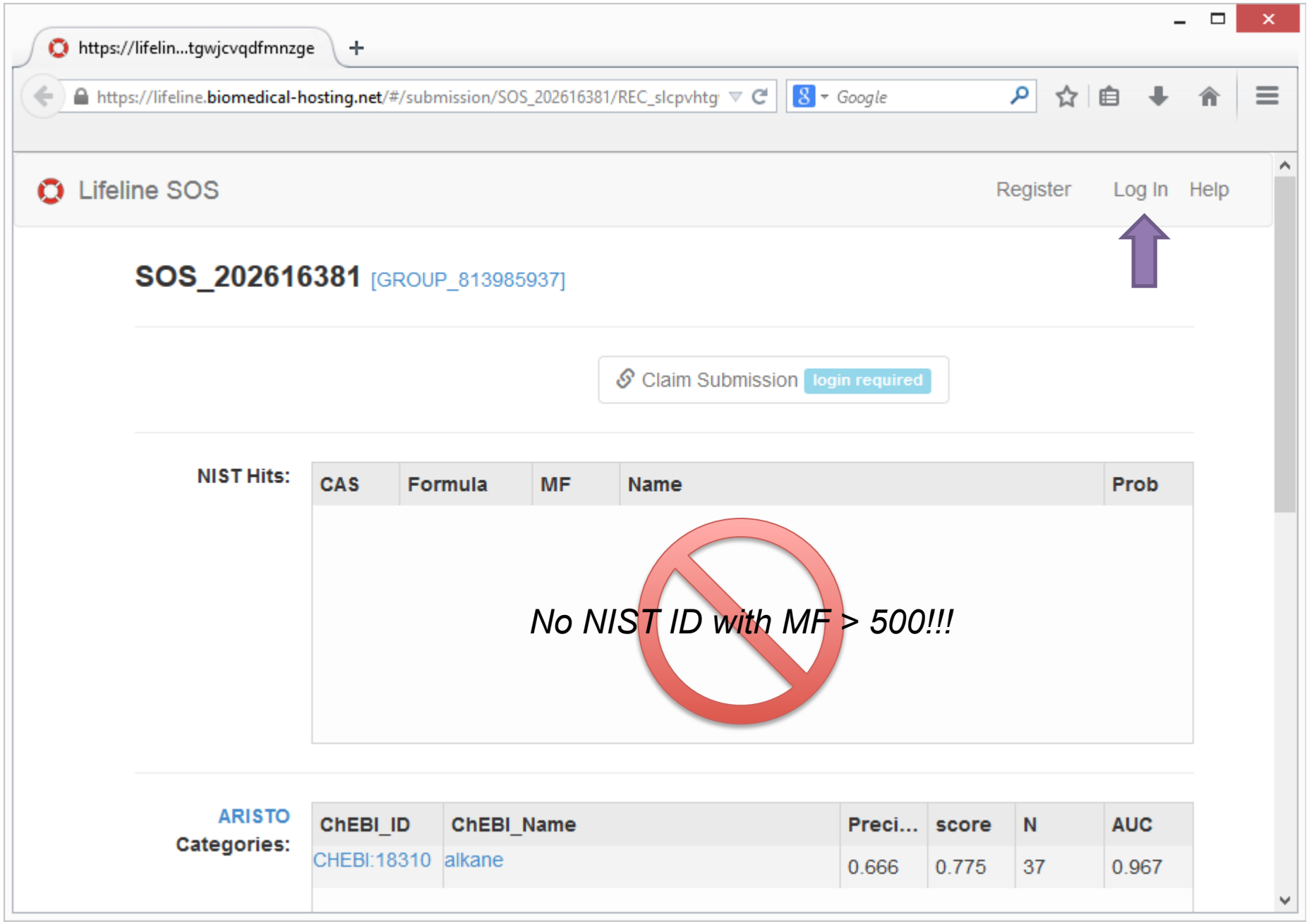
(NIST MS Search when no match is found…)



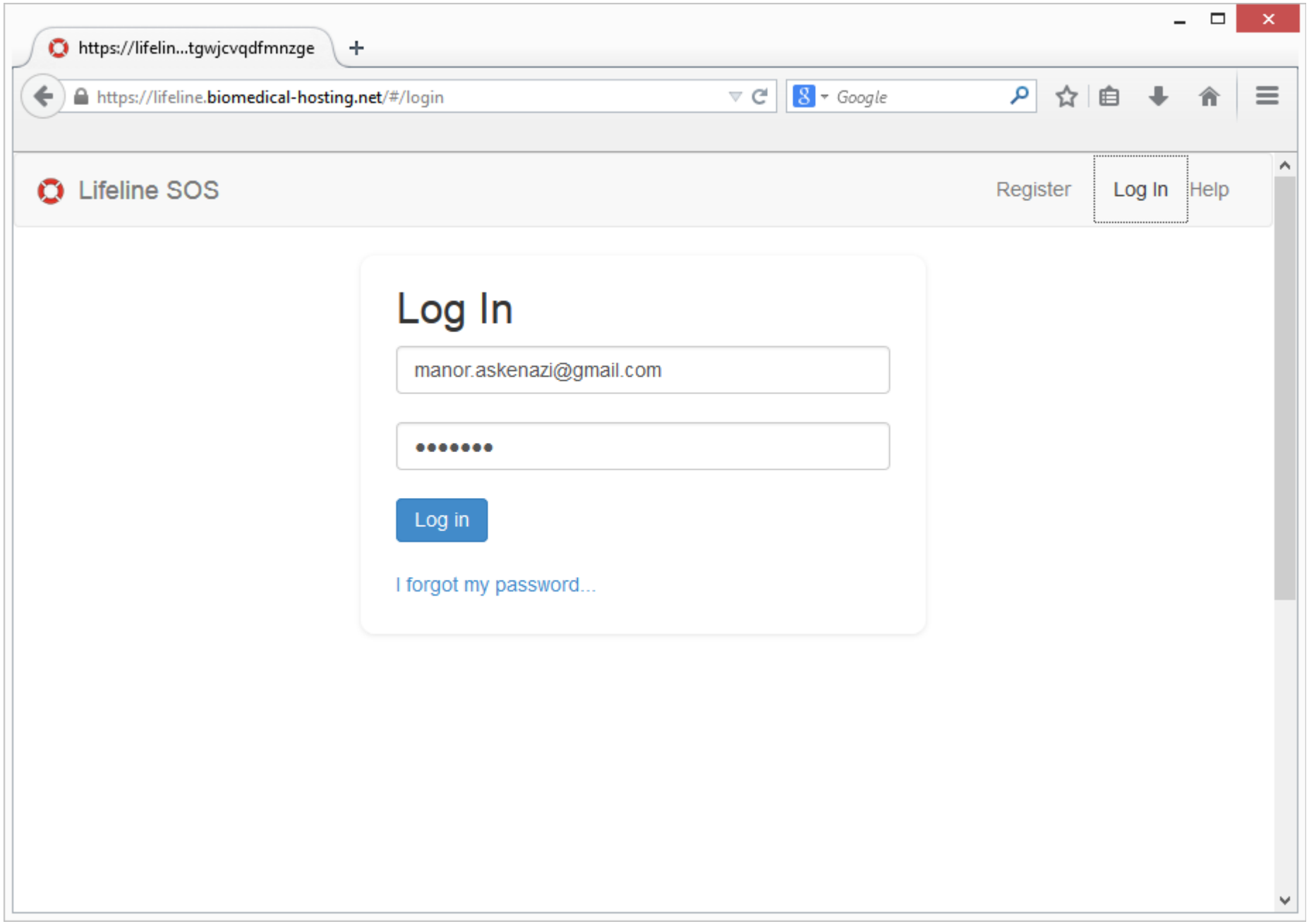
(The user can now provide feedback regarding the failure to identify…)



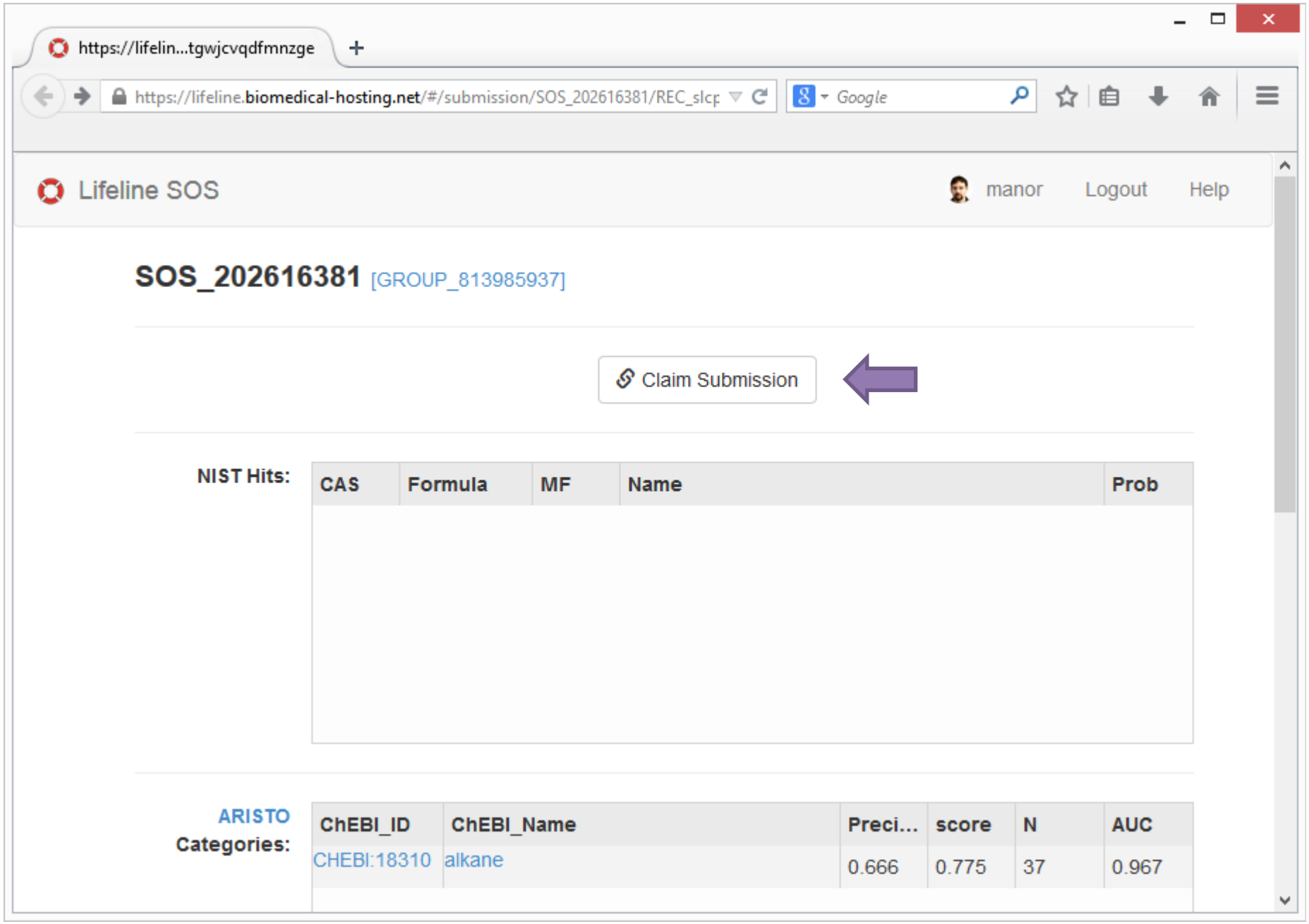
(Note: this will involve user-data leaving the lab…)



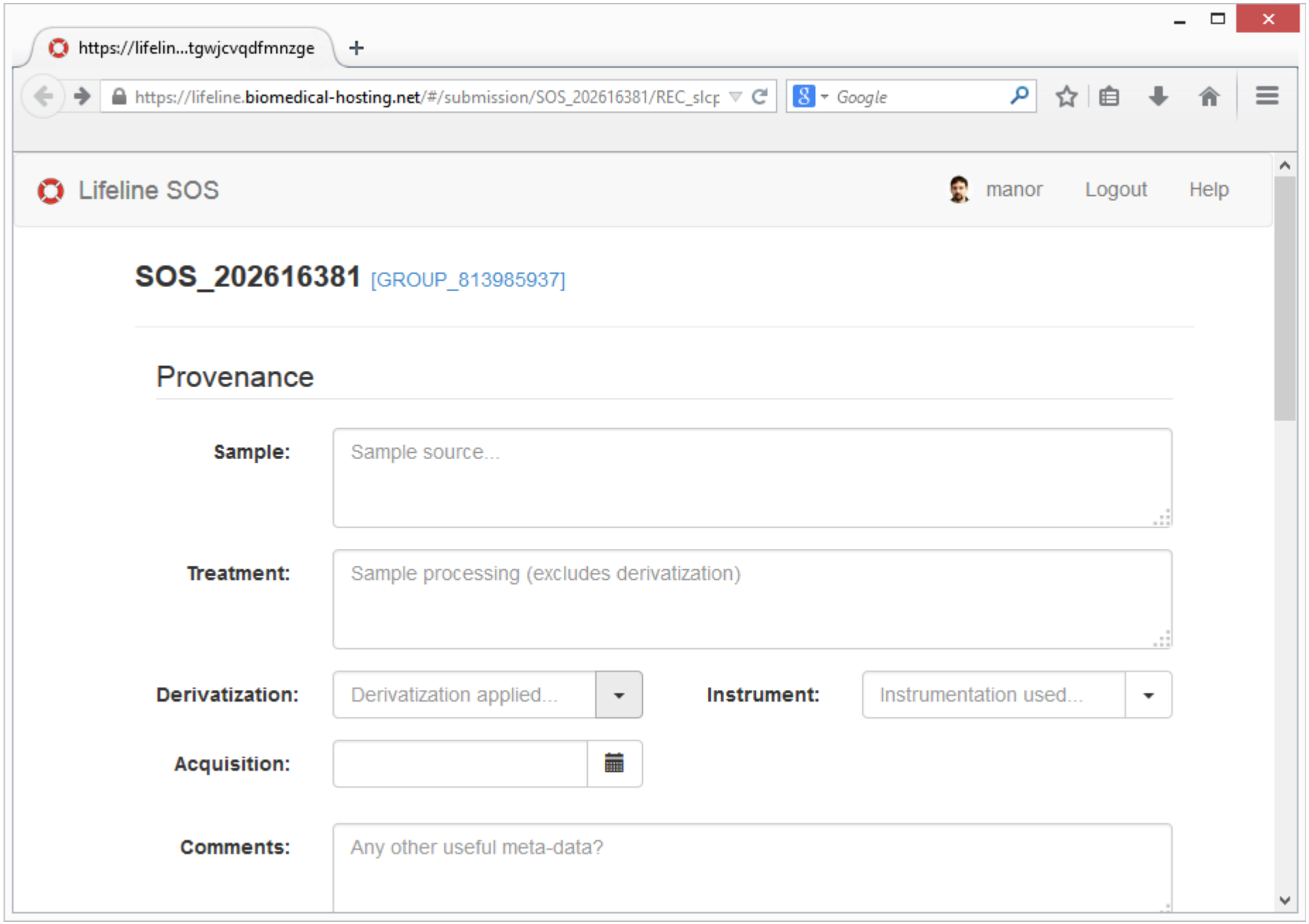
(By default an accession is provided, but the submission remains anonymous…)



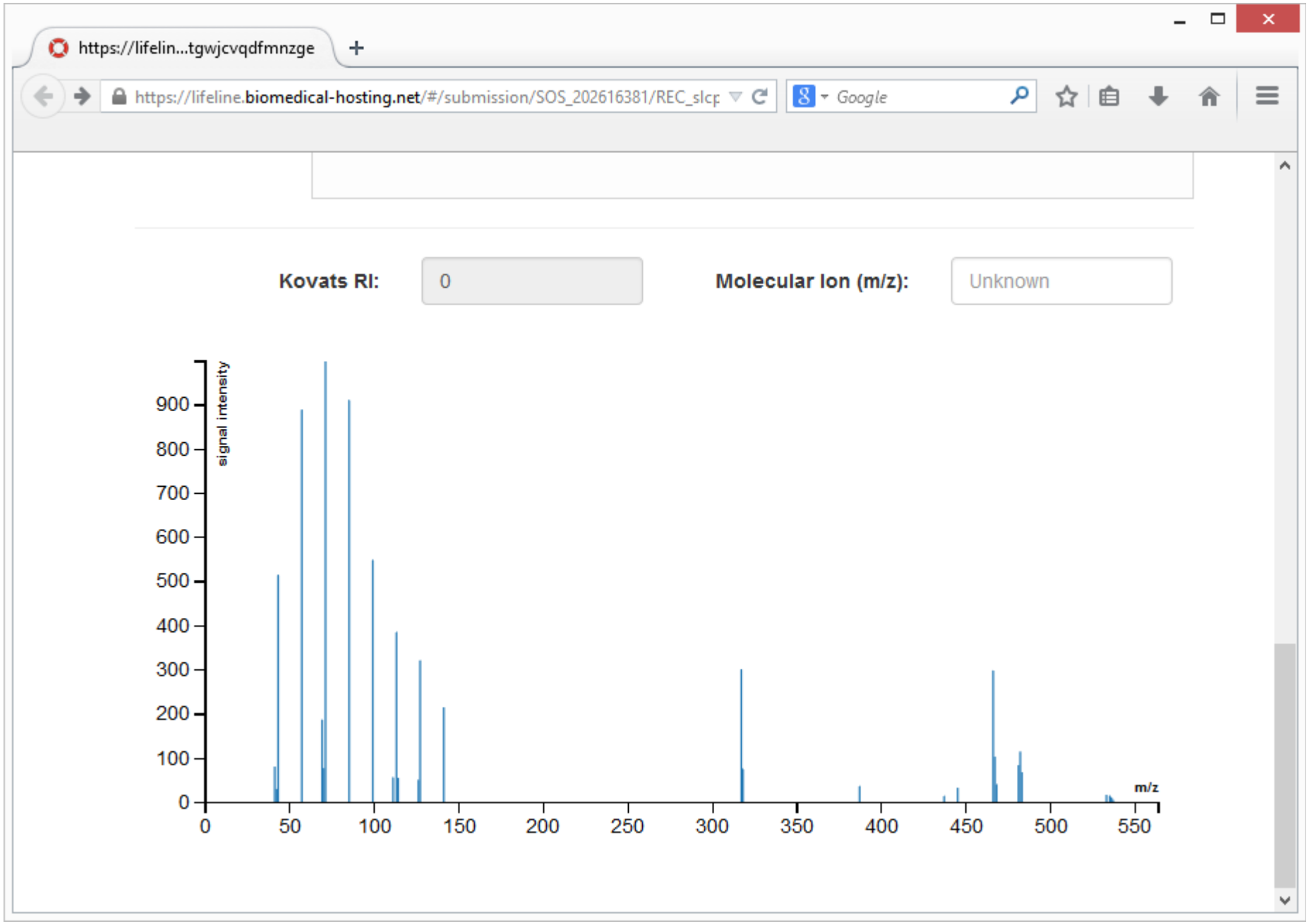
(By logging in, the user can “claim” the submitted spectrum…)



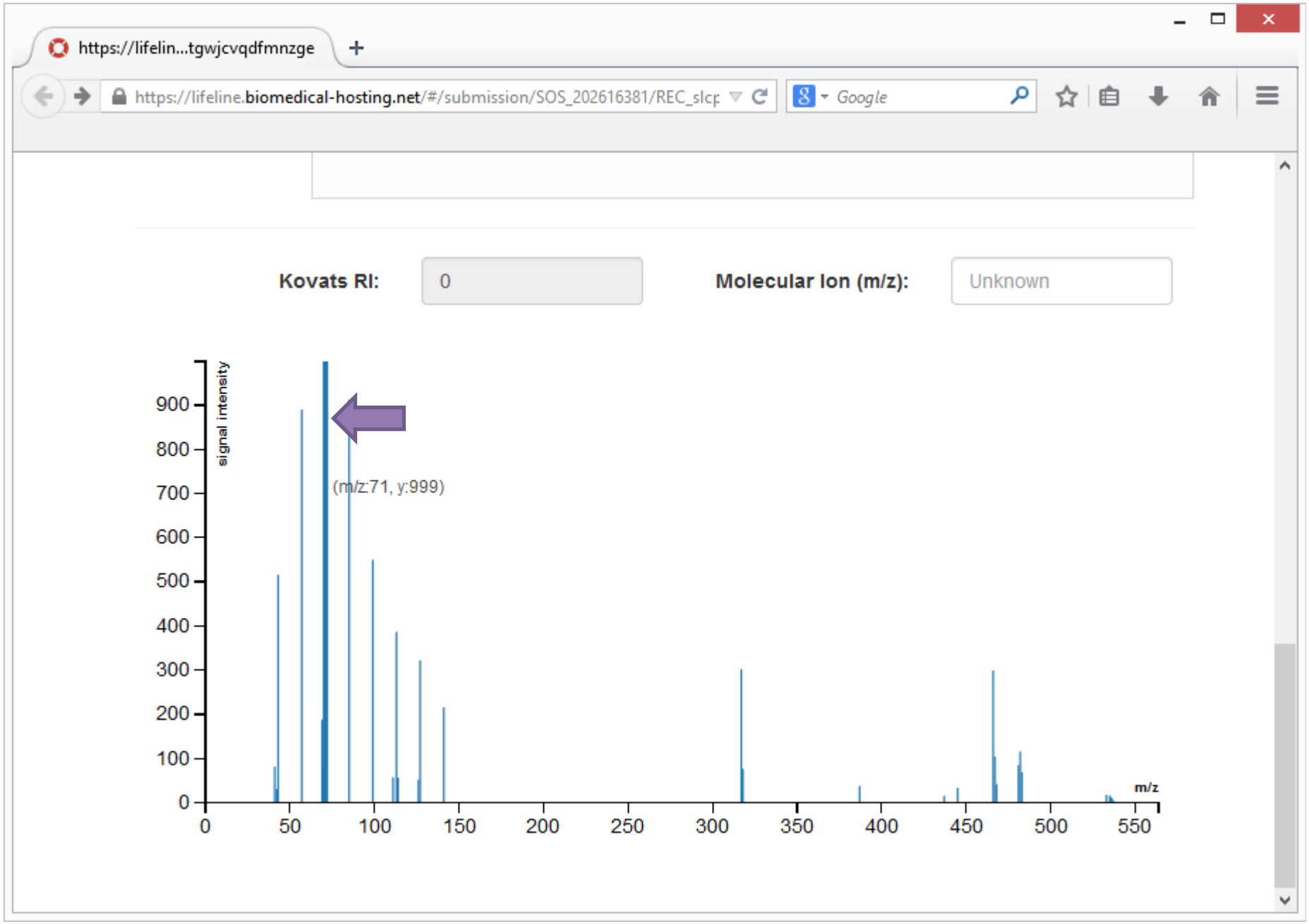
(By logging in, the user can “claim” the submitted spectrum…)



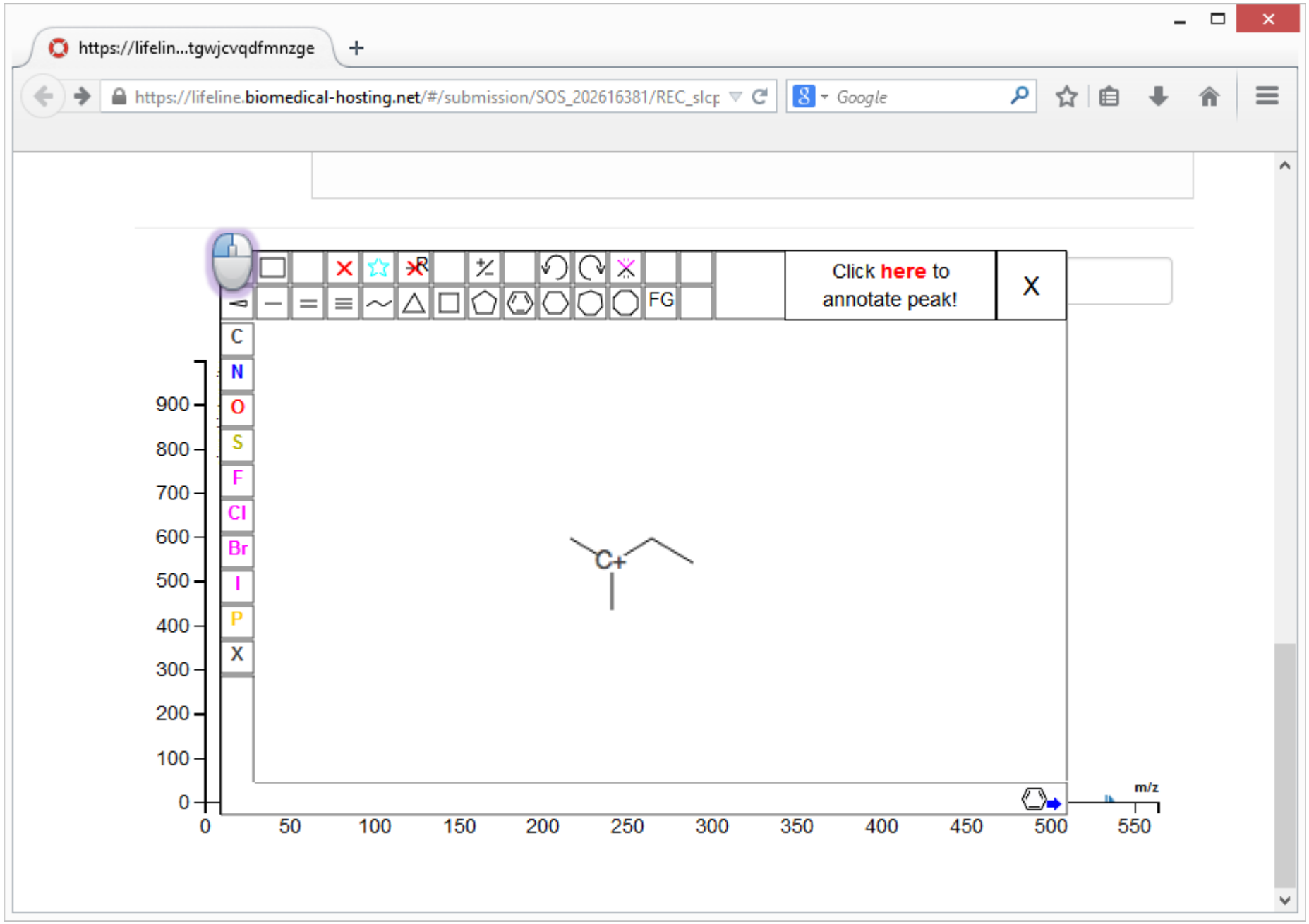
(S/he is then able to provide meta-data describing the submission…)



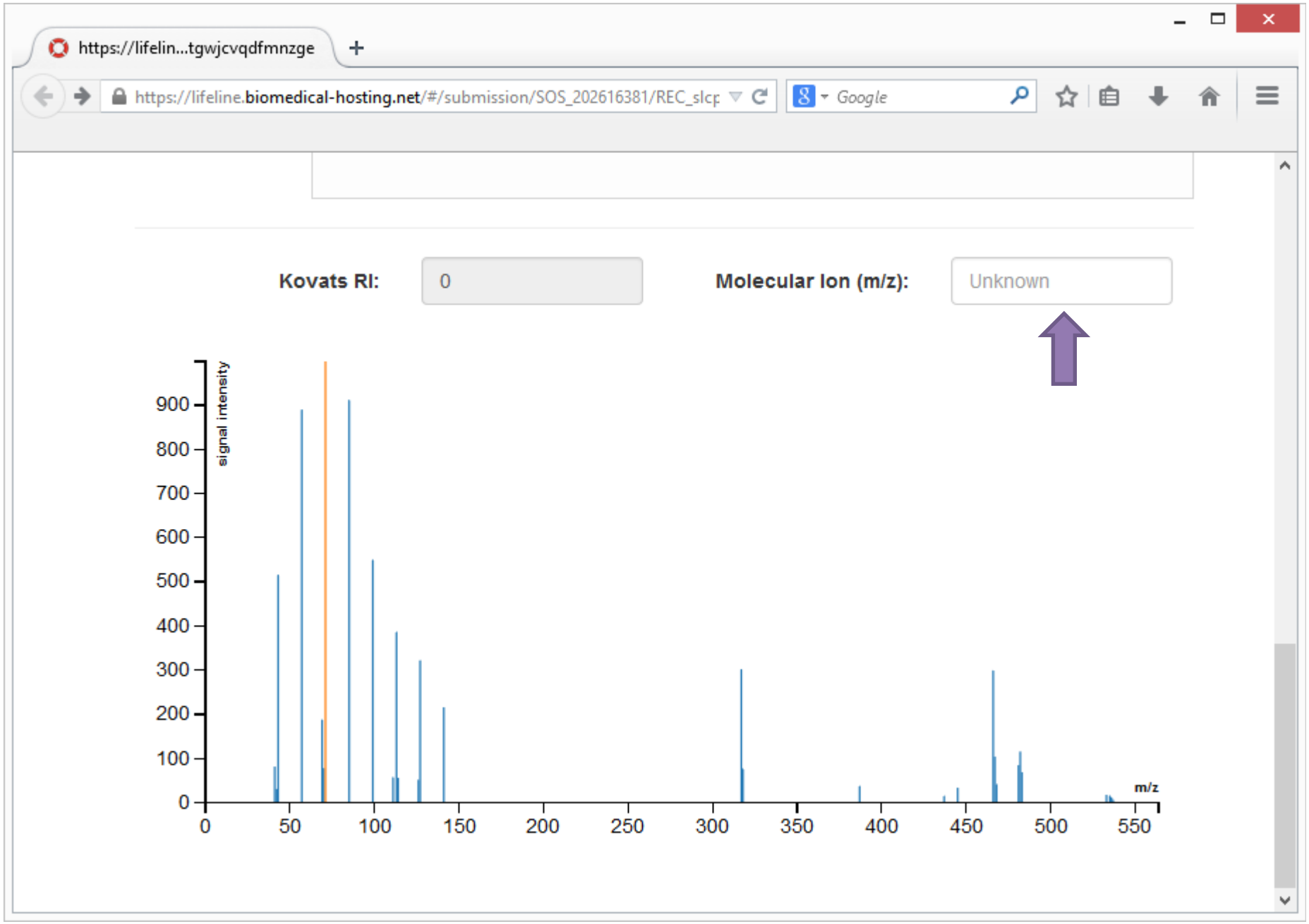
(Visually inspect the resulting spectral submission…)



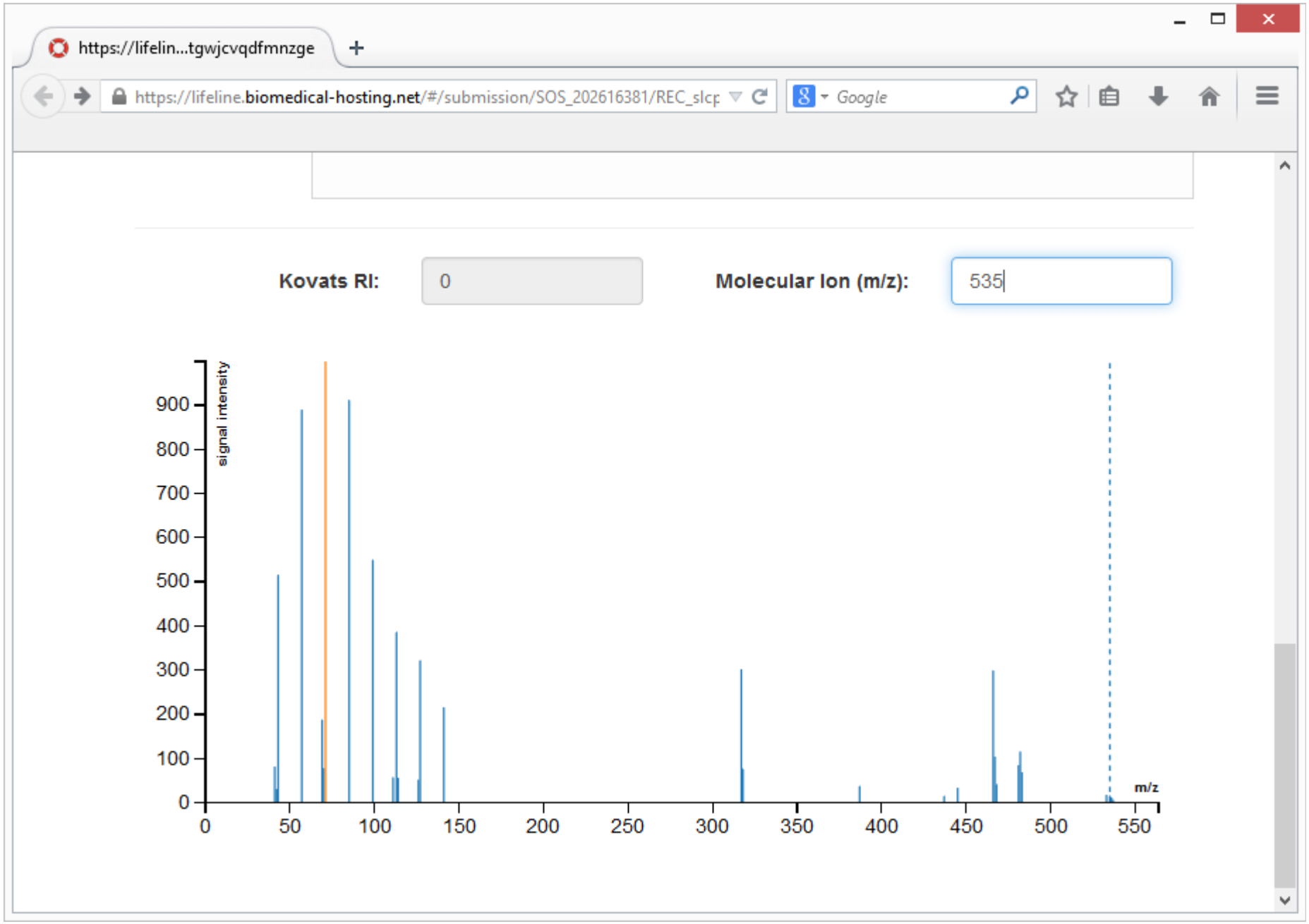
(Provide structural peak annotation…)



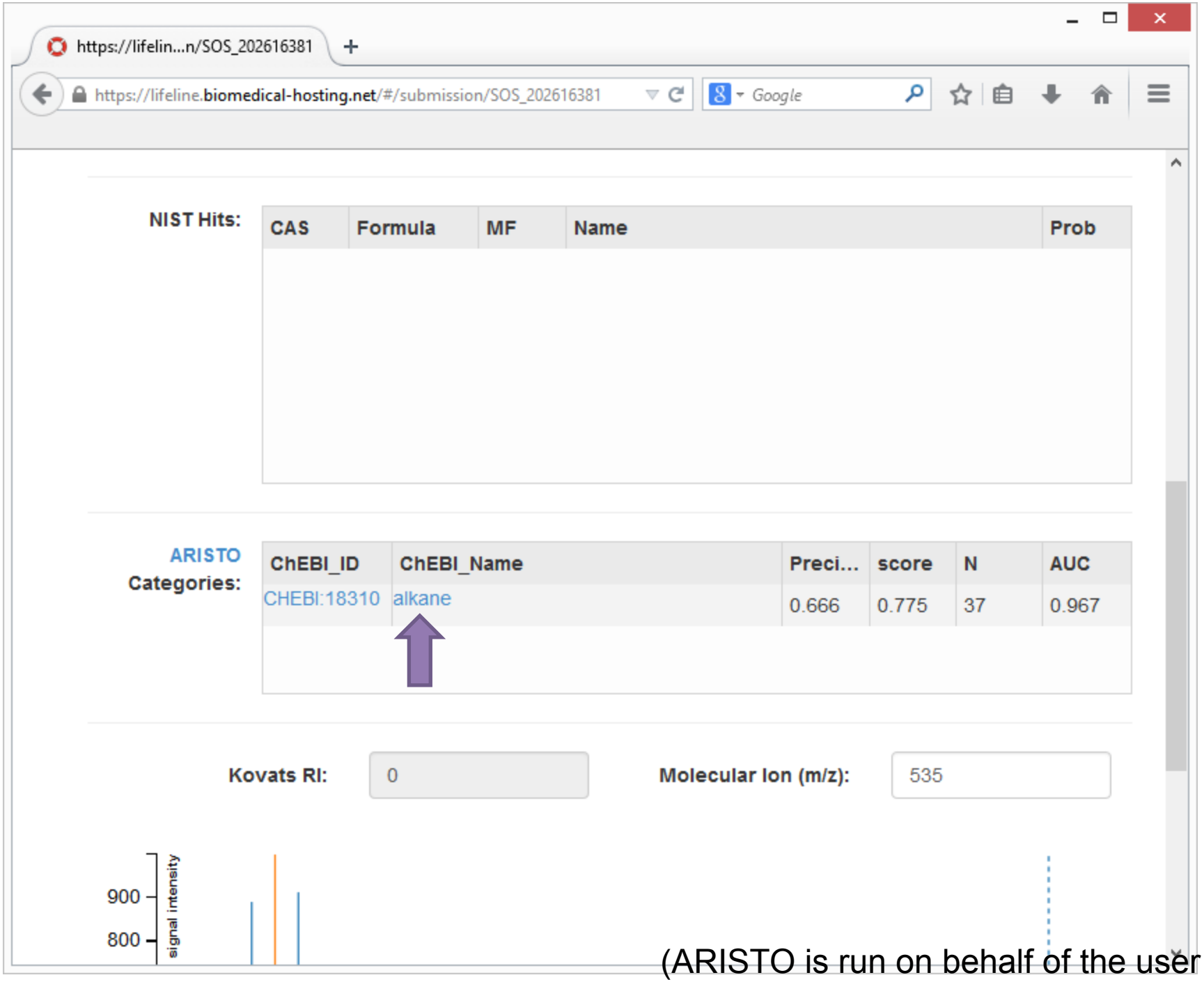
(Provide structural peak annotation…)



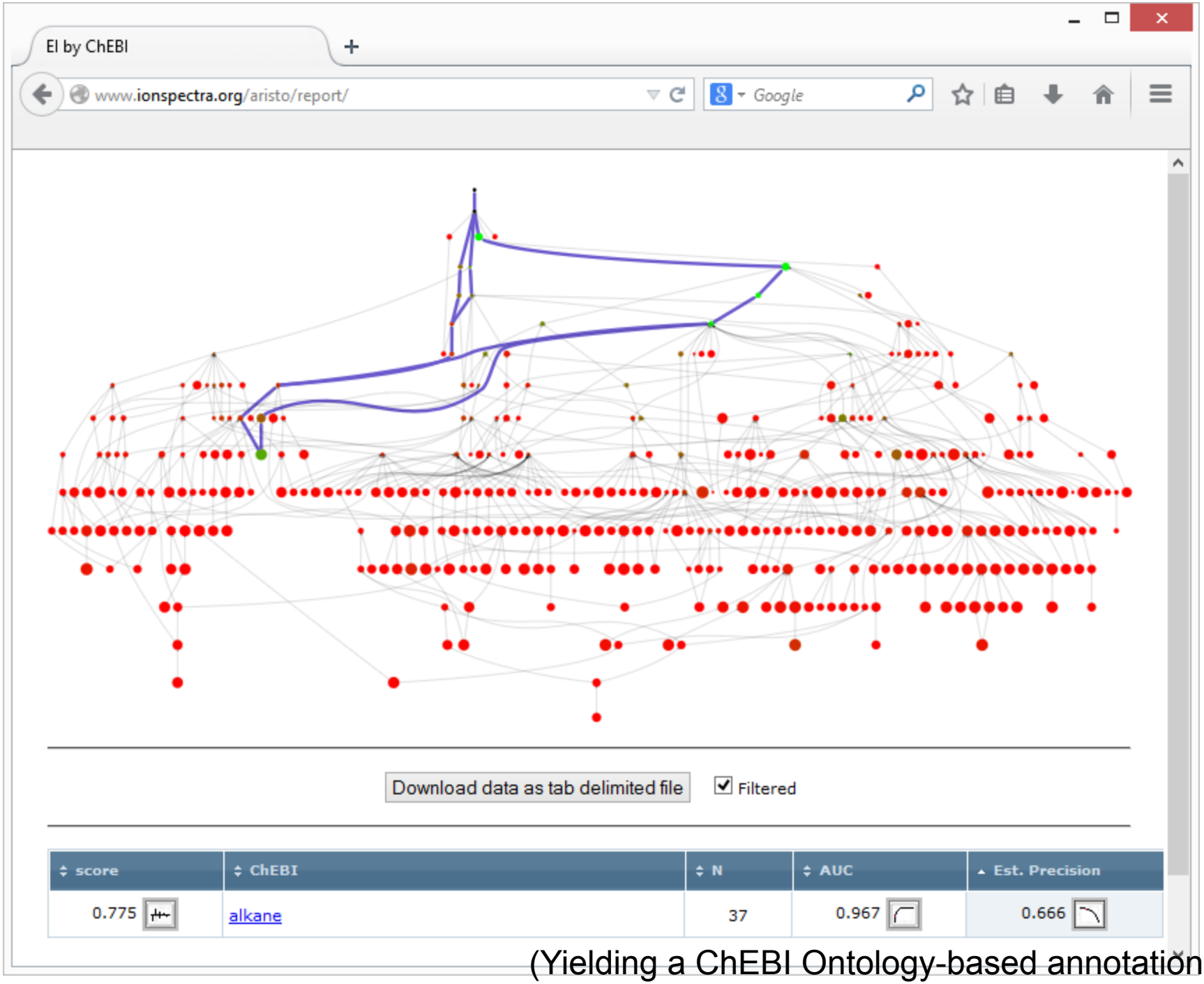
(And suggest the mass of the Molecular Ion…)



(And suggest the mass of the Molecular Ion…)



(ARISTO is run on behalf of the user…)

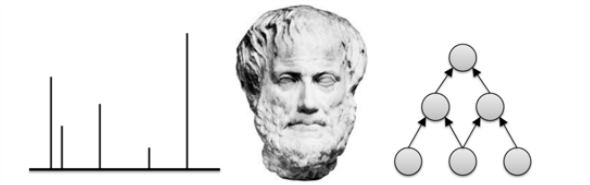


(Yielding a ChEBI Ontology-based annotation…)

(Yielding a ChEBI Ontology-based annotation…)

EI by ChEBI

www.ionspectra.org/aristo/report/



Automatic Reduction of Ion Spectra To Ontology

Nucleic Acids Research, 2011, Vol. 39, Web Server issue ***W505–W510***
doi:10.1093/nar/gkr403

ARISTO: ontological classification of small molecules by electron ionization-mass spectrometry

Manor Askenazi[1,2,*] and Michal Linial[1]

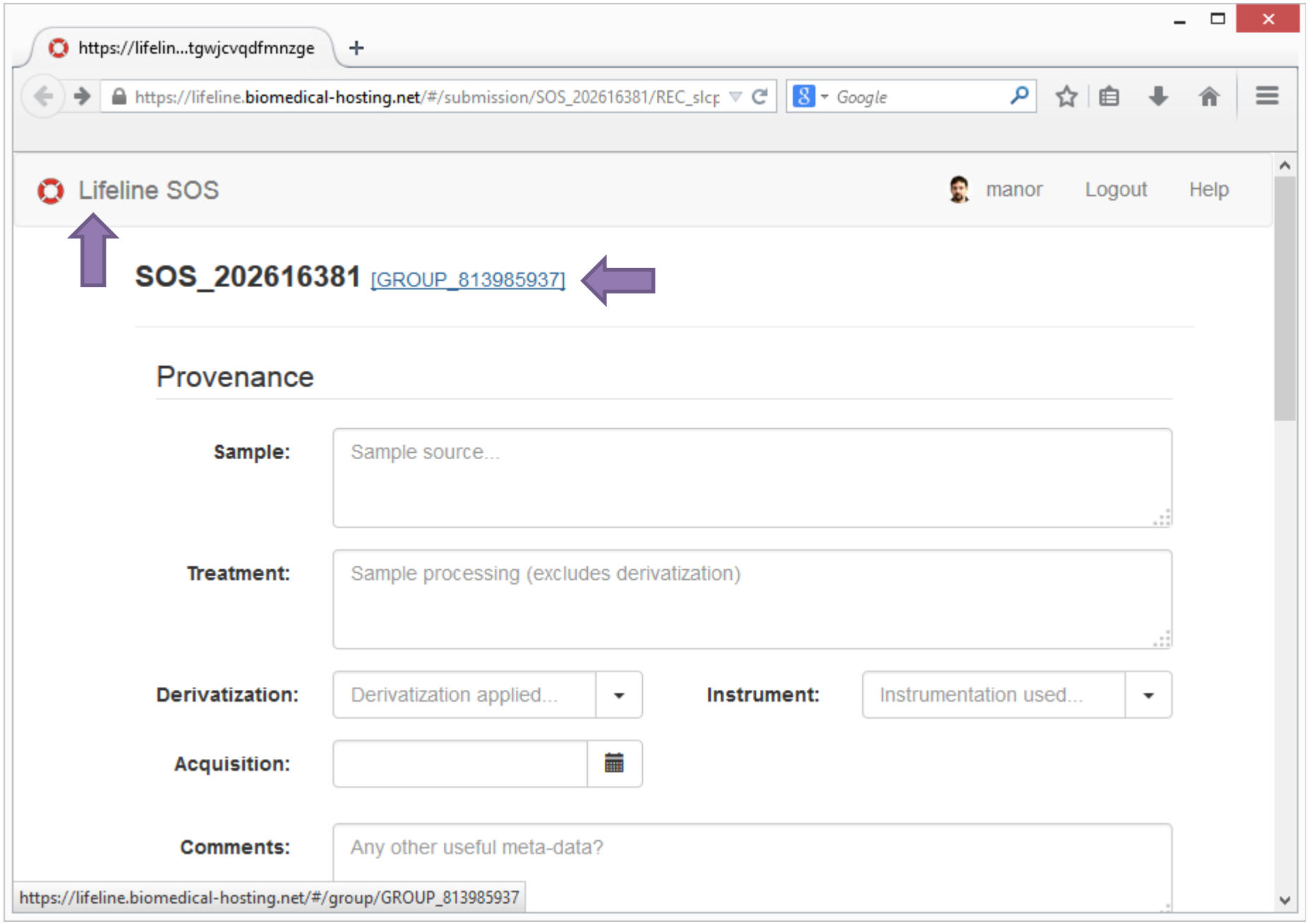
[1]Sudarsky Center for Computational Biology, Department of Biological Chemistry, Hebrew University of Jerusalem, Israel and [2]Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA, USA

Received February 15, 2011; Revised May 2, 2011; Accepted May 5, 2011

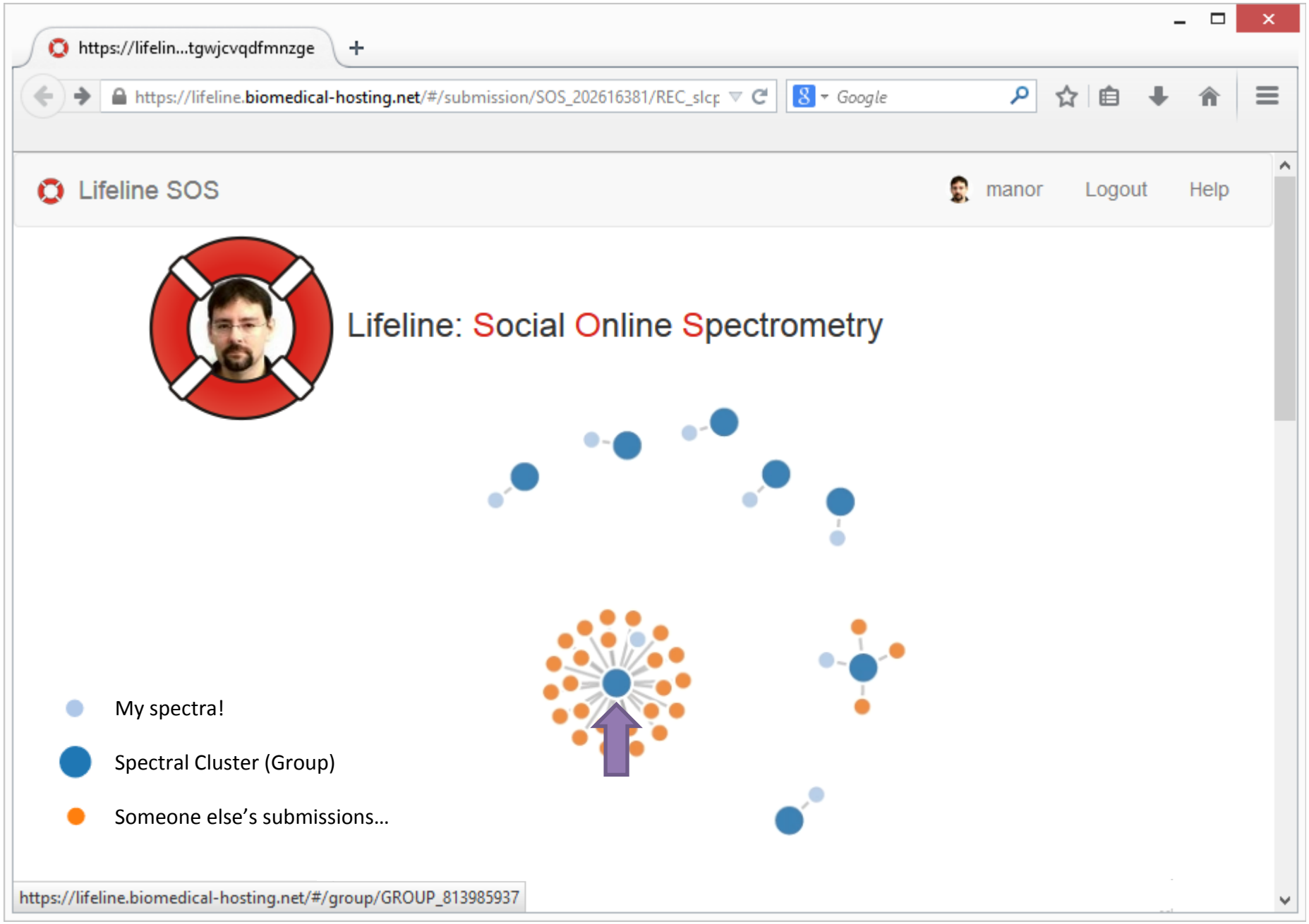
Download data as tab delimited file ☑ Filtered

score	ChEBI	N	AUC	Est. Precision
0.775	alkane	37	0.967	0.666

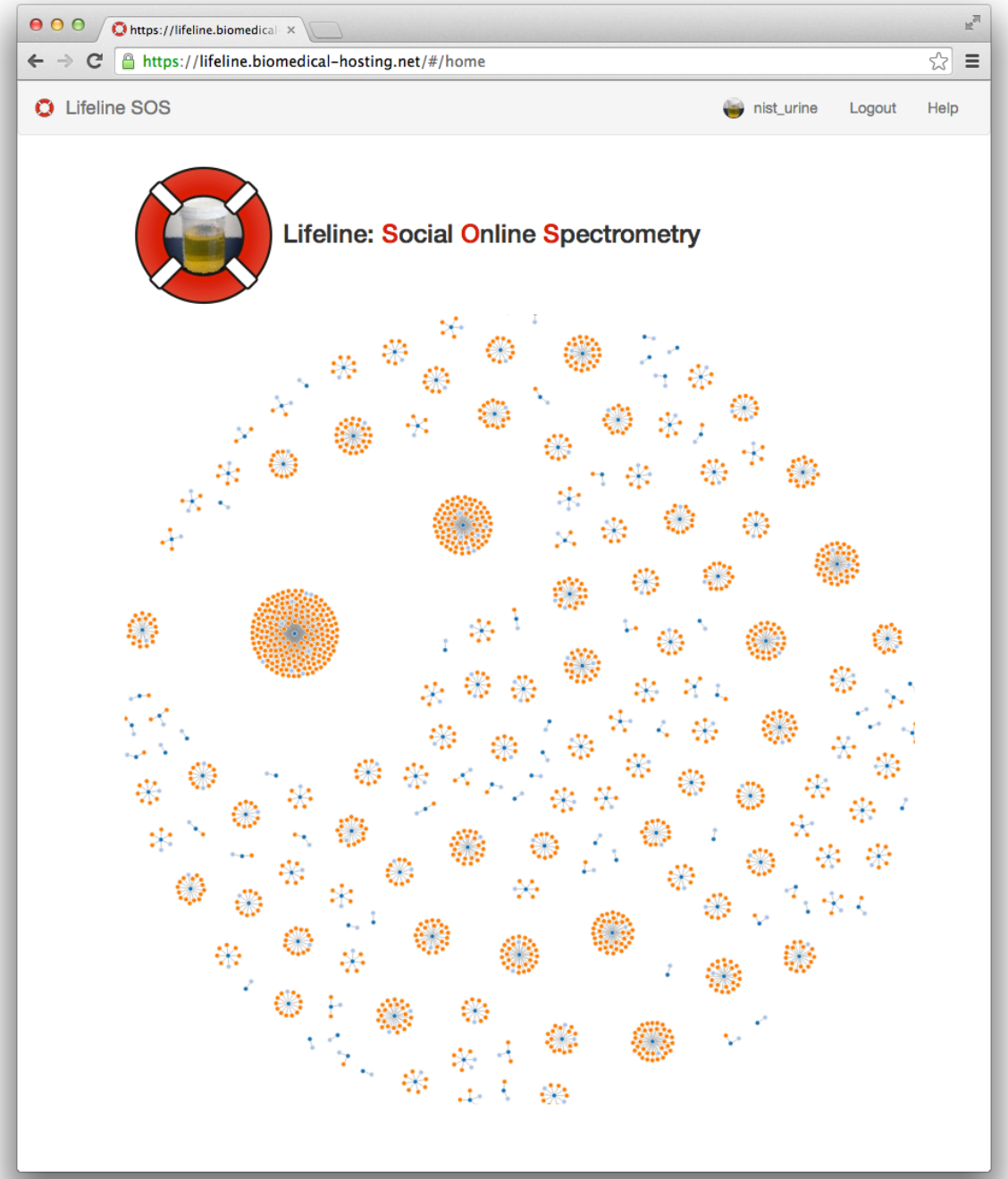
(Yielding a ChEBI Ontology-based annotation…)



(Every submission is associated with a Spectral Cluster…)



(The user's homepage shows his/her spectra in context…)



(Bulk loading of spectra is possible, but only through administration tools…)

https://lifelin...tgwjcvqdfmnzge

https://lifeline.biomedical-hosting.net/#/group/GROUP_813985937

Google

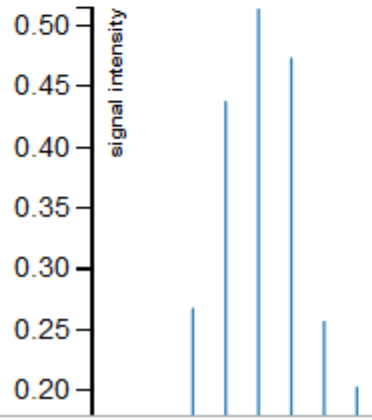
Lifeline SOS

manor Logout Help

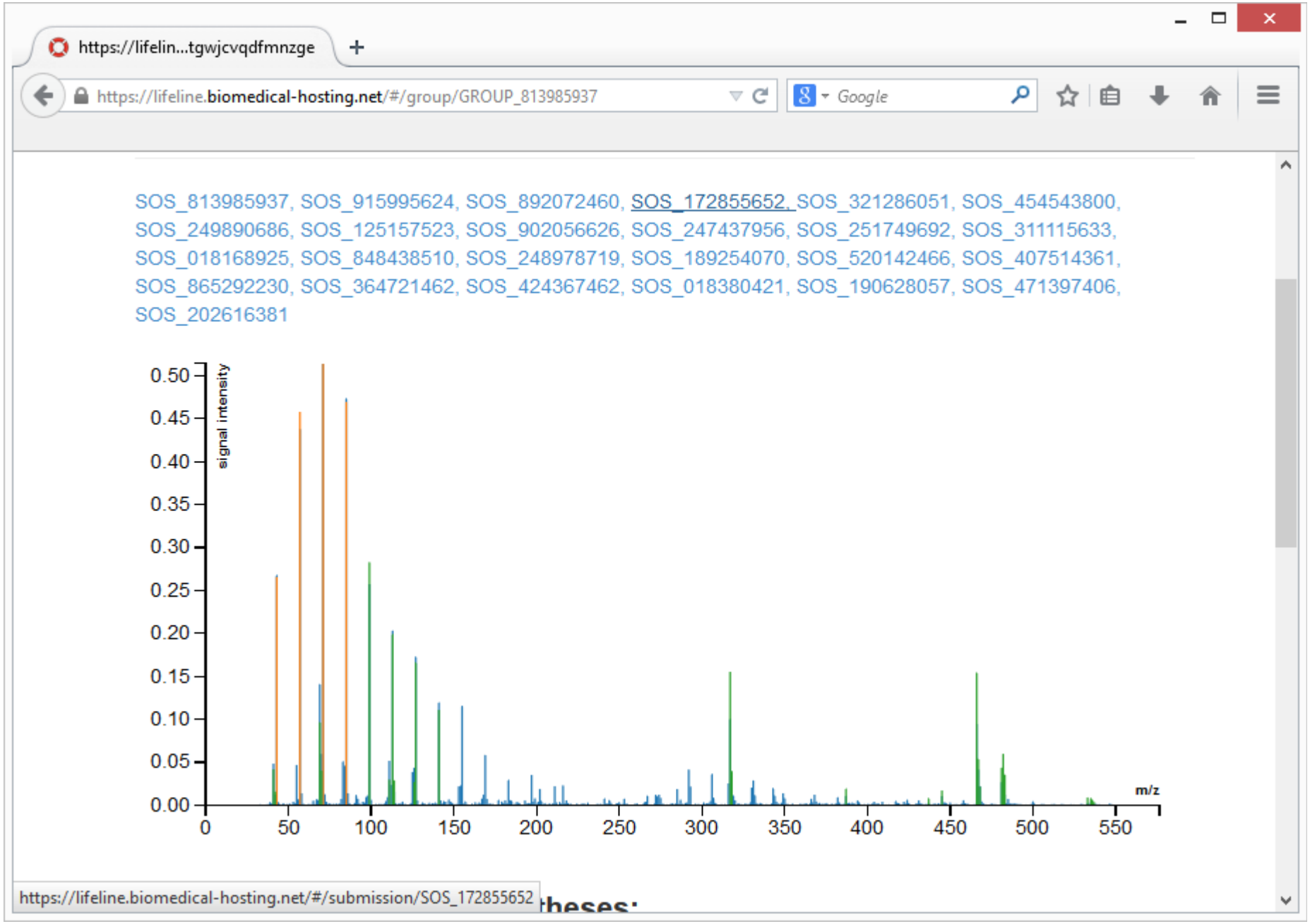
GROUP_813985937 (25 spectra)

Best NIST ID: *MF=615* Eicosane, 2-methyl- [C21H44, CAS# 1560845] **Source Spectrum:** SOS_892072460

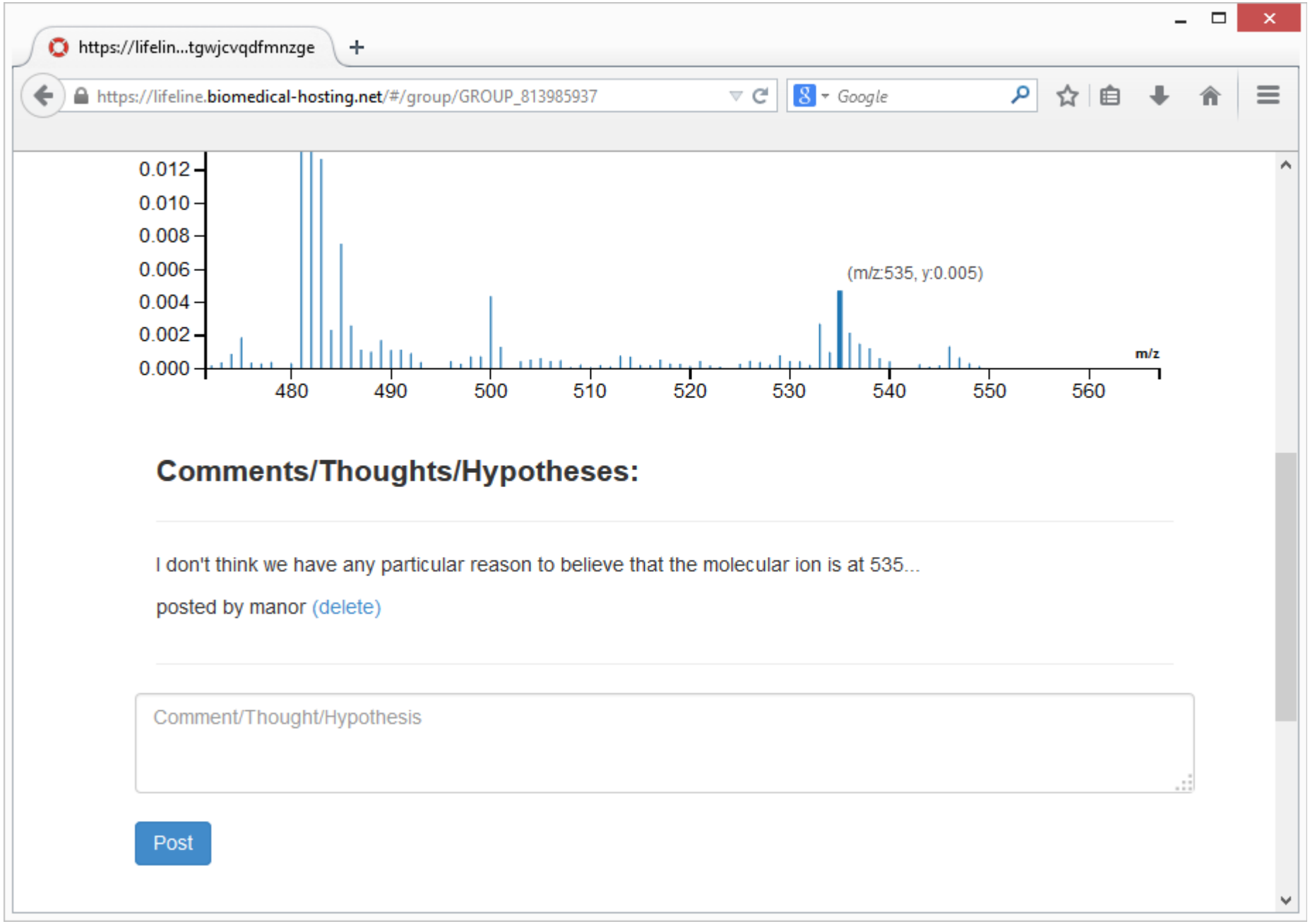
SOS_813985937, SOS_915995624, SOS_892072460, SOS_172855652, SOS_321286051, SOS_454543800, SOS_249890686, SOS_125157523, SOS_902056626, SOS_247437956, SOS_251749692, SOS_311115633, SOS_018168925, SOS_848438510, SOS_248978719, SOS_189254070, SOS_520142466, SOS_407514361, SOS_865292230, SOS_364721462, SOS_424367462, SOS_018380421, SOS_190628057, SOS_471397406, SOS_202616381



(The Group Page shows the average spectrum as well as all participating submissions)



(Individual spectra can be visualized as overlays on top of the consensus spectrum)



Users can communicate indirectly through the comments section of the Group Page)

lifeline@biomedical-hosting.net

http://lifeline.biomedical-hosting.net
(http://blogline.biomedical-hosting.net)

"לְכוּ־נָא וְנִוָּכְחָה"

"Come now, and let us reason together"

Isaiah, 1:18