

A faint, light-colored network pattern of interconnected lines and nodes is visible in the background, resembling a molecular or structural model.

CCDC

advancing structural science

What's Up

Customer Update Webinar

19th November 2020



Today's presenters



Seth Wiggin

Senior Scientific
Editor



Carmen Nitsche

General Manager
CCDC Inc.



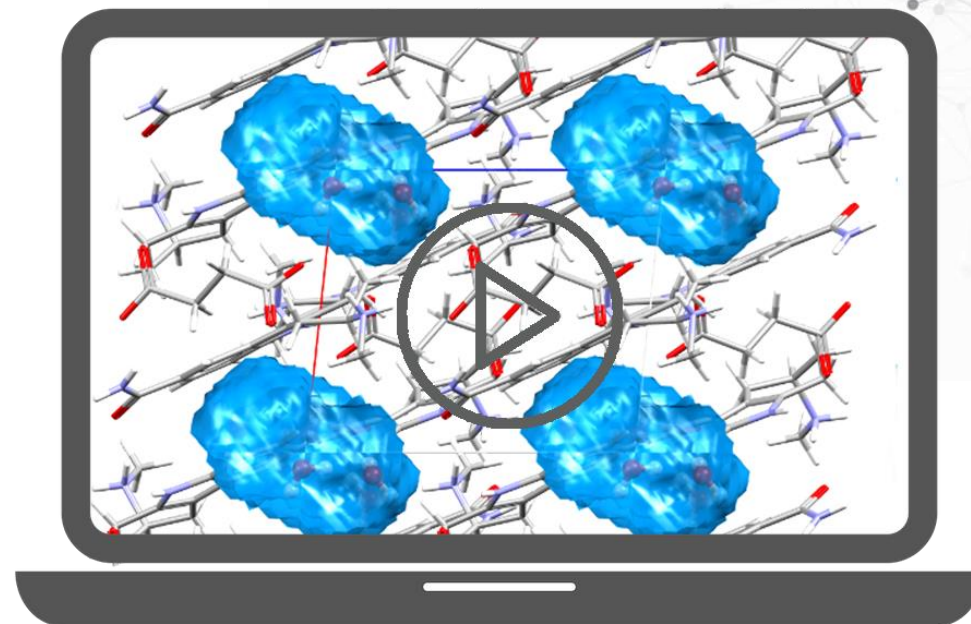
Ian Bruno

Head of Strategic
Partnerships

Overview

In this webinar we will discuss:

- Latest updates and news
- Using the Python API in Mercury for custom reporting and analyses
- CCDC software Partnerships
- Q&A: the floor is yours



Latest updates and news from CCDC

- 7th Crystal Structure Prediction (CSP) Blind Test just started.
<https://www.ccdc.cam.ac.uk/Community/initiatives/cspblindtests/csp-blind-test-7/>
- Latest CCDC Whitepaper: **Ultra-large GOLD docking on cloud resources**. Download from our website.
<https://info.ccdc.cam.ac.uk/whitepaper-ultra-large-gold-docking-on-cloud-resources>
- CrystEngComm celebrates the CSD 1 million structures in a special issue.

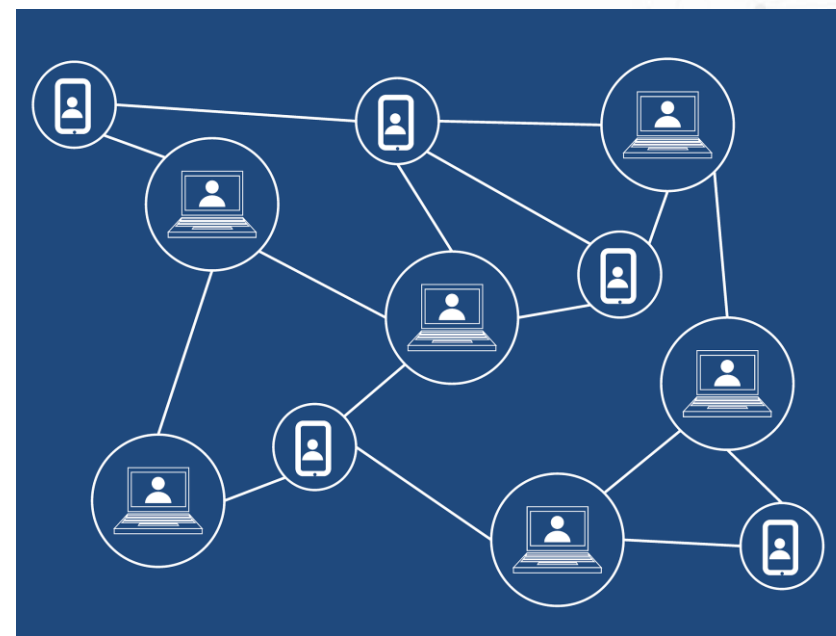


Latest updates and news from CCDC

>Events

- **CCDC Virtual Workshop: Aromatic Analyser**
Live session on 24th November
 - 3 pm – 4.30 pm (GMT)
- **CCDC and BACG Crystal Conversations**
3rd December
 - Virtual event
 - 2 pm to 4 pm (GMT)

→ Register for all CCDC events here
<https://www.ccdc.cam.ac.uk/News/Events>



Using the Python API in Mercury

For custom reporting and analyses.



Seth Wiggin

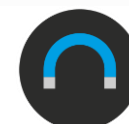
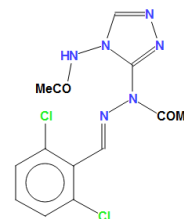
Senior Scientific Editor

What is the CSD Python API?

- The CSD Python API (Application Programming Interface) enables you to use many capabilities of the CSD-System without being bound by graphical (or command line) interfaces
- You can readily create CSD-driven analyses and workflows, tailor them to your needs and then publish them to your own menu in Mercury for specialist analysis and easier communication

CSD Python API - Example

```
In [10]: from codo import io, diagram
In [26]: import IPython.core.display
import StringIO
In [27]: # Set up CSD entry reader and find the first entry in the database
csd = io.EntryReader('csd')
csd_entry = csd[0]
csd_entry.identifier
Out[27]: u'AABHTZ'
In [30]: # Generate a diagram for that CSD entry
diagram_generator = diagram.DiagramGenerator()
diagram_generator.settings.font_size = 12
img = diagram_generator.image(csd_entry)
In [31]: # Display the 2d diagram
output = StringIO.StringIO()
img.save(output, "PNG")
contents = output.getvalue()
IPython.core.display.display_png(contents, raw=True)
```



Functions include:

- Full search capabilities
- Geometry analysis
- Interaction analysis
- Descriptor calculation
- 2D diagram generation

1.

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Application Note

Hotspots API: A Python Package for the Detection of Small Molecule Binding Hotspots and Application to Structure-Based Drug Design

Peter R. Curran,*¹ Chris J. Radoux,¹ Mihaela D. Smlouva,¹ Richard A. Sykes,¹ Alicia P. Higuera,¹ Anthony R. Bradley,¹ Brian D. Marsden,¹ David R. Spring,¹ Tom L. Blundell,¹ Andrew R. Leach,¹ William R. Pitt,¹ and Jason C. Cole²

On This: *J. Chem. Inf. Model.* 2020, 60, 1911–1916 | Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Methods that survey protein surfaces for binding hotspots can help to evaluate target tractability and guide exploration of potential ligand binding regions. Fragment Hotspot Maps builds upon interaction data mined from the CSD (Cambridge Structural Database) and exploits the idea of identifying hotspots using small chemical fragments, which is now widely used to design new drug leads. Prior to this publication, Fragment Hotspot Maps was only publicly available through a web application. To increase the accessibility of this algorithm we present the Hotspots API (application programming interface), a toolkit that offers programmatic access to the core Fragment Hotspot Maps algorithm, thereby facilitating the interpretation and application of the analysis. To demonstrate the package's utility, we present a workflow which automatically derives protein–hydrogen-bond constraints for molecular docking. The Hotspots API is available from <https://github.com/pcrcan/hotspot> under the MIT license and is dependent on the commercial CSD Python API.

INTRODUCTION

In the context of protein–ligand interactions, the term “hotspot” describes a region within a pocket that contributes a disproportionately large amount to the overall binding energy.¹ We previously described a hotspot as “the minimum binding site that will bind a fragment, maintaining the fragment binding position, once it has been elaborated.”² Fragment experiments can yield useful information about the tractability of a target,³ or be used to guide structure-based drug designs.⁴ It follows that the presence of a computationally determined hotspot can be used in the same way.⁵

There have been many computational approaches to map potential protein–ligand interactions within pockets.^{6–12} Fragment Hotspot Maps,¹³ and other more recent methods,^{14–17} go further by differentiating between the available interactions, highlighting the most preferential.

Prior to this communication, the Fragment Hotspot Maps method was only publicly available through a web application (<http://fragment-hotspot-maps.ccdc.cam.ac.uk/>). While this provided easy access to the method, it allowed only for visual inspection of the results, limiting how the information could be used, particularly for large scale applications or as part of existing structure-based drug design (SBD) workflows.

Herein, we present the Hotspots API. For the general user, this provides direct access to the calculation, enabling analysis of confidential structures and facilitating the integration of results with other SBD methods. For those users, we provide example workflows as “cookbook” examples in documentation which include tractability assessment, coframe searching, and docking. As an example, we show how to use Fragment Hotspot Maps to automatically GOLD¹⁸ docking constraints. For developers, the open code base offers a platform for collaboration and for researchers from several institutions to start project, new features and applications.

Fragment Hotspot Maps Background. In SuperStar, The Fragment Hotspot Maps approach the previous work of Hoström¹⁹ and SuperStar,²⁰ in 1999. In SuperStar, patterns of interactions are coded searching for all structures containing a given predefined functional groups, which are then searched for nonbonded contacts between the two central group and a contact group). Each 3D hit is treated such that the central groups are superimposed. This set of contact group atomic positions at supergroup (Figure 1a). As described in the original paper,²⁰ supergroups can be converted into grids by

Received: October 28, 2019
Published: March 26, 2020

ACS Publications | © 2020 American Chemical Society 1911 | <https://doi.org/10.1021/acs.jcim.2002009>
J. Chem. Inf. Model. 2020, 60, 1911–1916

2.

The Journal of Physical Chemistry Letters | Letter
pubs.acs.org/jpclett

Machine Learning in Chemistry | Hot Paper

Co-crystal Prediction by Artificial Neural Networks^{26,27}

Jan-Joris Devogeleer, Hugo Meekes, Paul Tinnemans, Elias Vlieg, and René de Gelder*

How to cite: International Edition: doi.org/10.1002/anie.202009467
German Edition: doi.org/10.1002/ange.202009467

ABSTRACT: A significant amount of attention has been given to the design and synthesis of co-crystals by both industry and academia because of its potential to change a molecule's physicochemical properties. Yet, difficulties arise when searching for adequate combinations of molecules (or coformers) to form co-crystals, hampering the efficient exploration of the target's solid-state landscape. This paper reports on the application of a data-driven co-crystal prediction method based on two types of artificial neural network models and co-crystal data present in the Cambridge Structural Database. The models accept pairs of coformers and predict whether a co-crystal is likely to form. By combining the output of multiple models of both types, our approach shows to have excellent performance on the proposed co-crystal training and validation sets and has an estimated accuracy of 80% for molecules for which previous co-crystallization data is unavailable.

INTRODUCTION

Molecular solids²⁸ appear in many different ways, and the solid-state landscape of a molecule may cover various crystalline forms, ranging from polymorphs and hydrates to more complex multicomponent crystals.²⁹ In the latter, the formation of new intermolecular interactions between the target and an auxiliary compound has proven to be an excellent tool to modify physico-chemical characteristics of a target compound, such as the (aqueous) solubility, bioavailability, density, and melting point.^{30–32} Multicomponent crystals therefore find their application in various fields (e.g., fertilizers,^{33,34} pigments^{35,36} and medicine^{37,38}), and play a pivotal role in the effective formulation of pharmaceuticals.

The design of multicomponent crystals is non-trivial and new forms are often identified via trial and error. Unlike salts, where proton transfer leads to strong ionic/covalent interactions, solvates and co-crystals are assembled through weaker, non-covalent interactions (e.g., hydrogen bonding, π-π interactions, ...). Such intermolecular interactions between functional groups are often used to rationalize the possibility of aggregation,³⁹ but with no guarantee that the potential interactions will emerge.

Whereas polymorphic salts and solvates are commonly screened using automated high throughput systems,^{40,41} the experimental screening of co-crystals remains labour-intensive and time-consuming. In order to shorten this process, a variety of computational tools, based on hydrogen-bond propensities^{42–44} statistical analysis and modeling of molecular descriptors,^{45,46} electrostatic potential maps,^{47,48} crystal structure prediction,^{49,50} COSMO-RS^{51,52} molecular dynamics,⁵³ or PIXEL calculations^{54,55} and Hirshfeld surface analysis⁵⁶ (as for instance implemented in the *Crystal Explorer* or *software package*⁵⁷), have been developed to aid in the discovery of adequate combinations of the constituents or coformers. Although these computer-aided methods have succeeded in enhancing co-crystal screening protocols, some of the shortcomings include their bias towards small or structurally related datasets, oversimplified assumptions regarding the mechanisms of interaction, and, in some cases, their computational costs.

Recently, we introduced a holistic approach to study co-crystallization using network science and link prediction.^{58,59} Analysis of a network of coformers extracted from the Cambridge Structural database (CSD)⁶⁰ shows that, rather than being a random assembly of coformers, it represents a rational source of co-crystal information that can form a basis for prediction. Therefore, it would be very appealing to develop a method that utilizes all this co-crystal information, and is able to predict co-crystals for coformers lacking any experimental data on co-crystal formation. Such a tool would for instance enable the evaluation of the co-crystal formation propensity for in silico determined drug candidates (prior to their actual synthesis), or aid in the (co-)crystallization of molecules that are amorphous in their pure form.

Artificial neural networks, and in particular deep learning,⁶¹ have emerged as promising tools for data-driven prediction. Given an adequate molecular representation, artificial neural networks can be used to, for example, predict physico-chemical properties (e.g., solubility) or classify molecules, hereby assigning the input to a certain class (e.g., toxic or non-toxic).

Driven by the recent advances in artificial neural networks and the promising source of co-crystal information present in the CSD, we introduce a new approach to predict co-crystal formation using neural networks. Two neural network model types are introduced that each accept a pair of coformers as input and classify the combination as a possible co-crystal or not. By optimizing the configuration

Research Articles | Angewandte Chemie International Edition | Wiley Online Library

Received: July 27, 2019
Accepted: September 23, 2019
Published: September 23, 2019

© 2020 The Authors. Published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial, and no modifications or adaptations are made.

Angew. Chem. Int. Ed. 2020, 59, 1–10

3.

The Journal of Physical Chemistry Letters | Letter
pubs.acs.org/jpclett

Virtual Screening for High Carrier Mobility in Organic Semiconductors

and Harald Oberhofer*

Research Center, Technische Universität München, Lichtenbergstr. 4, D-85747

ABSTRACT: To improve the use of organic semiconductors in screening approaches we compute the electronic energies as two main descriptors for charge states extracted from the Cambridge Structural of the calculated coupling values, we identify ion pathways. Thus, we readily find many of the most promising materials that have not yet been electronically. Together with the unique descriptor database allows us to extract further accelerate the theoretical design and discovery.

INTRODUCTION

Organic semiconductors have attracted much attention in the field of electronics and photonics. The former mostly reflects the quantum mechanical overlap of the frontier orbitals involved in the charge transport, while the latter quantity accounts for the response of all other charges in the system to the local change in charge state. While both quantities are intuitively understood in the context of hopping type transport, they are also meaningful descriptors beyond this regime.¹ The electronic coupling for example directly enters the tight-binding description of the carrier effective mass in the band-transport picture,^{2,30} and λ can, albeit more indirectly, be viewed as a measure of the electron/phonon coupling strength,³¹ which, for example, plays a role in the relaxation time approximation of the band transport mobility.³² Finally, another factor potentially influencing transport properties in organic semiconductors is the degree of disorder present in the system,³³ leading for example to a variation in site energies of localized charges on inequivalent sites. However, this effect is mostly of importance in amorphous systems,^{33,35} and can thus for the sake of brevity be omitted for the crystalline semiconductors considered here (see the Supporting Information for an analysis of site energy variations).

In a pioneering screening study, Sokolov et al.³⁴ employed λ to assess the suitability of seven modifications to the well-known anaphtho-biendioophene (DNTT) organic semiconductor.³⁵ The choice for the reorganization energy was thereby not arbitrary but motivated by the fact that λ can, with some limitations be reduced to a single-molecule property due to the comparatively low contribution of the external

Received: July 27, 2016
Accepted: September 23, 2016
Published: September 23, 2016

© 2016 The Authors. Published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial, and no modifications or adaptations are made.

Angew. Chem. Int. Ed. 2016, 54, 3971–3977

1. <https://doi.org/10.1021/acs.jcim.9b00996>
2. <https://doi.org/10.1002/anie.202009467>
3. <https://doi.org/10.1021/acs.jpcllett.6b01657>



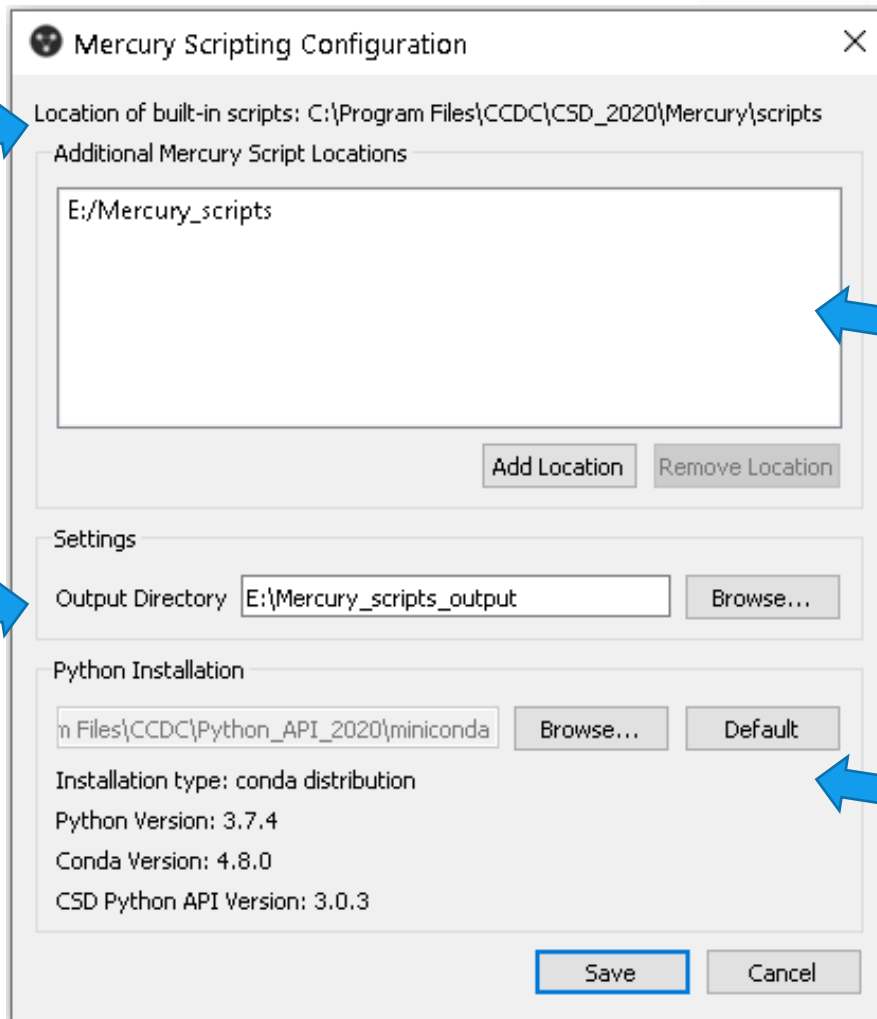
Benefits of scripting workflows



- Automating reporting and analyses saves time
- Results will be consistent and predictable
- Having a scripted workflow makes it easy to share knowledge with colleagues

CSD Python API Mercury interface

Mercury contains built-in scripts offering a range of search and analysis



The script output can be written to any specified location



Additional script locations can be added for custom scripts from a variety of sources



The Python installation used by Mercury is configurable



Mercury Demo

The screenshot displays the Mercury software interface. The main window shows a 3D ball-and-stick model of a complex organic molecule. A report window titled 'AABHTZ' is overlaid on the model. The report contains the following text:

Crystal Structure Report for AABHTZ

Crystal Structure Analysis

The report also includes the CCDC logo and a chemical structure diagram of a fragment with a MeCO group, an NH group, and a chlorine atom (Cl).

At the bottom of the Mercury window, there are 'Display Options' and 'Display' sections with checkboxes for Packing, Asymmetric Unit, Auto centre, Short Contact, and H-Bond. A 'Reset' button is also present.

Conclusions

- Using Mercury's CSD Python API interface allows easy access to the API's functionality for reports and analysis
- Automating workflows via Python scripting saves time and effort
- It is simple to customize scripts for your own specialist analysis and communication needs
- Thank you for your attention

CCDC Software Partnerships

Past, present and future plans.



Ian Bruno

Head of Strategic Partnerships

CCDC Partnerships

Structure Solution

Rigaku
oxford diffraction

BRUKER

Malvern Panalytical
a spectris company

Crystals

OlexSys

Phenix

GΦL
Global Phasing Limited

Data Publication

ACS
Chemistry for Life®

IUGr

ROYAL SOCIETY OF CHEMISTRY

WILEY

Taylor & Francis Group
an informa business

ELSEVIER

SPRINGER NATURE

DE GRUYTER

Hindawi

DataCite

Data Linking

ICSD

ICDD
International Centre for Diffraction Data

WORLDWIDE PDB
PROTEIN DATA BANK

PubChem

ChemSpider
The free chemical database

PPDB

DRUGBANK

ChEMBL

Clarivate Analytics

InChI TRUST

Software Integrations

BIOVIA

Chemical Computing Group

SCHRÖDINGER

WAVEFUNCTION

BioSolveIT

KNIME
Open for Innovation

optibrium

cresset

CCDC Software Partnerships

Structure Solution

Rigaku
oxford diffraction

BRUKER

Malvern Panalytical
a spectris company

Crystals

OlexSys

Phenix

GΦL
Global Phasing Limited

Data Publication

ACS
Chemistry for Life®

IUGT

ROYAL SOCIETY OF CHEMISTRY

WILEY

Taylor & Francis Group
an informa business

ELSEVIER

SPRINGER NATURE

DE GRUYTER

Hindawi

DataCite

Data Linking

ICSD

ICDD
International Centre for Diffraction Data

PDB
PROTEIN DATA BANK

PubChem

ChemSpider
The free chemical database

PPDB

DRUGBANK

ChEMBL

Clarivate Analytics

InChI TRUST

Software Integrations

BIOVIA

Chemical Computing Group

SCHRÖDINGER

WAVEFUNCTION

BioSolveIT

KNIME
Open for Innovation

optibrium

cresset

General Priorities for Software Partnerships

- Make CSD data and functionality more readily accessible to chemists across an organisation
- Make it possible to combine CSD data and functionality with other scientific methods
- Make it easier for CSD-based data functionality to be used in custom workflows
- Enable the use of CSD data to help with structure solution and refinement in chemistry and biology

Current Integrations

Protein-Ligand Docking



Discovery Studio



MOE



StarDrop



GOLD

CSD Searching

WAVEFUNCTION



Spartan



Materials Studio



ConQuest

Materials Studio also interfaces to Motif Searching in CSD-Materials

Ideas Generation



MOE CSD Linker Database



Spark CSD Fragment Database



ReCore CSD Index

BioSolveIT

CSD-based libraries for third party systems available to download at <https://www.ccdc.cam.ac.uk/support-and-resources/downloads/>

Workflow Environments









CCDC Pipeline Pilot Collection







CCDC KNIME Components

Current Integrations

Geometry Analysis

 SCHRÖDINGER	Maestro	<i>Validation of molecular geometries</i>
 Chemical Computing Group	MOE	<i>Validation of molecular geometries</i>
 GFL Global Phasing Limited	Grade	<i>Optimisation of ligand geometries in proteins</i>
 Phenix	Phenix	
 Crystals	Crystals	<i>Restrains for structure refinement based on experimental data</i>
 Mogul		

Structure Solution

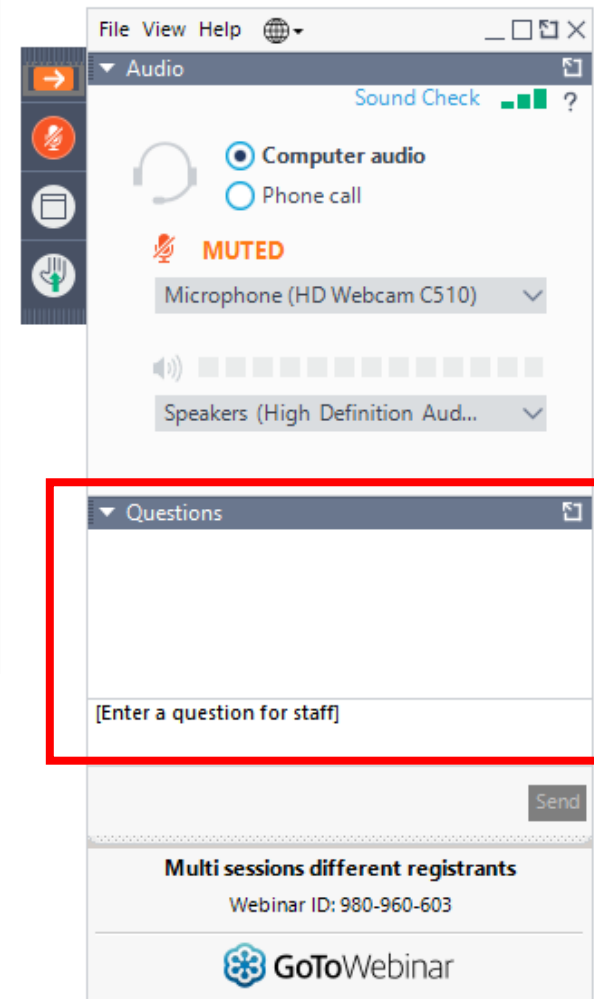
 Rigaku oxford diffraction	Rigaku / Oxford Diffraction	
 BRUKER	Bruker	<i>Reduced cell searching using CellCheckCSD</i>
 Malvern Panalytical	HighScore	<i>CSD-based Search/Match database available for download</i>
 OlexSys	Olex²	<i>Direct link to CCDC Deposition Services</i>

Current Priorities

- Move existing integrations to use the CSD Python API
 - Current integrations use command line options to CCDC software – the CSD Python API offers greater flexibility and extensibility
- Extend to other CCDC functionality
 - Interest has been expressed in the ability to access Full Interaction Maps from within other packages – we are discussing this with the relevant partners
- Improve existing integrations
 - Refreshing existing interfaces to GOLD to accommodate new functionality – we would be interested in your feedback on options you would like to see exposed

Q&A

- Type your questions in the box as shown

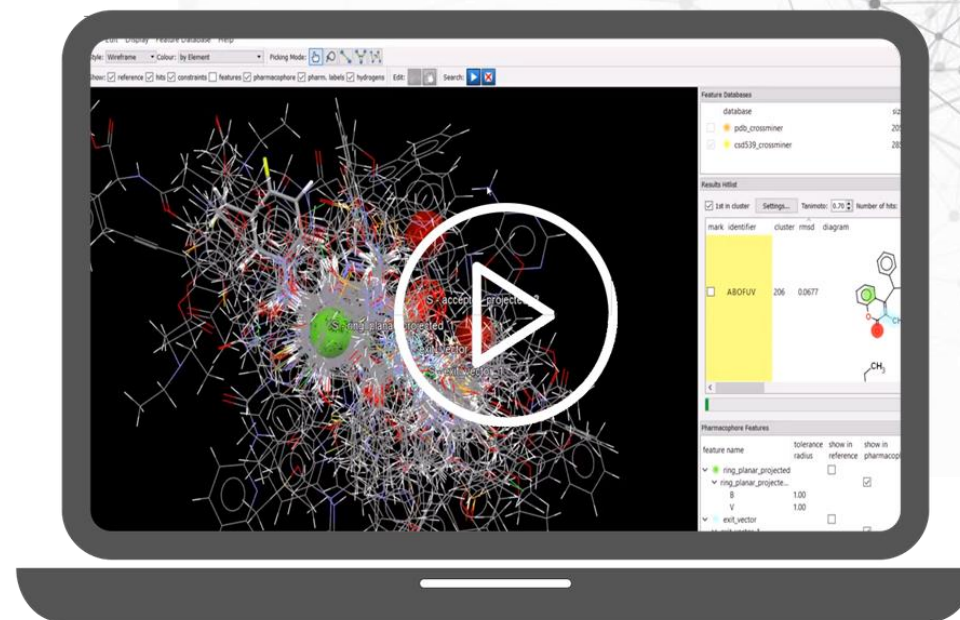


The screenshot displays the GoToWebinar interface. At the top, there is a menu bar with 'File', 'View', and 'Help'. Below this is the 'Audio' control panel, which includes a 'Sound Check' indicator and options for 'Computer audio' (selected) and 'Phone call'. A 'MUTED' status is shown with a microphone icon. The selected microphone is 'Microphone (HD Webcam C510)' and the speakers are 'Speakers (High Definition Aud...'. Below the audio settings is a 'Questions' section, which is highlighted with a red box. This section contains a text input field with the placeholder text '[Enter a question for staff]' and a 'Send' button. At the bottom of the interface, there is a section for 'Multi sessions different registrants' with the 'Webinar ID: 980-960-603' and the 'GoToWebinar' logo.

Next What's Up Webinar

- Next webinar: January 21st
- Follow us on social media
- Send us your ideas and news

hello@ccdc.cam.ac.uk



Thank you

hello@ccdc.cam.ac.uk

The Cambridge Crystallographic Data Centre
12 Union Road, Cambridge CB2 1EZ, United Kingdom
[Registered Charity No. 800579](#)

CCDC