MESPEUS

(MEtal Sites in Proteins at Edinburgh University)

http://mespeus.nchu.edu.tw/

Version 2023

MESPEUS is being developed by

National Chung Hsing University, Taiwan

Contact: kunyi.hsin@nchu.edu.tw

Acknowledgements

Developers would like to thank Prof. Malcolm D. Walkinshaw and Dr. Marjorie M. Harding for insightful comments, and Dr. Reuven Pnini for thoughtful discussion. This project was supported by grants of National Science and Technology Council and Council of Agriculture Executive Yuan, Republic of China (Grant 110-2313-B-005-005 and 112AS-9.2.1-AD-U1).

Contents

Introduction

MESPEUS is a freely accessible database which uses carefully selected metal coordination groups found in metalloprotein structures taken from the Protein Data Bank. The database contains geometrical information of metal sites within proteins. Metal types include Metals selected are the 10 elements common in biology, known as Na, Mg, K, Ca, Mn, Fe, Co, Ni, Cu, and Zn. Additional 30 metals are also covered, including Ag, Al, Au, Ba, Be, Cd, Cr, Cs, Ga, Gd, Hg, Ir, Li, Mo, Pb, Pd, Pr, Pt, Rb, Re, Rh, Ru, Sr, Tb, Tl, U, V, W, Y, and Yb. In order to completely determine the metal coordination, the symmetry-related units of a given protein structure are taken into and are generated using the appropriate space group symmetry operations. MESPEUS is automatically updated monthly by synchronizing with the Protein Data Bank database. MESPEUS is now available at http://mespeus.nchu.edu.tw/.

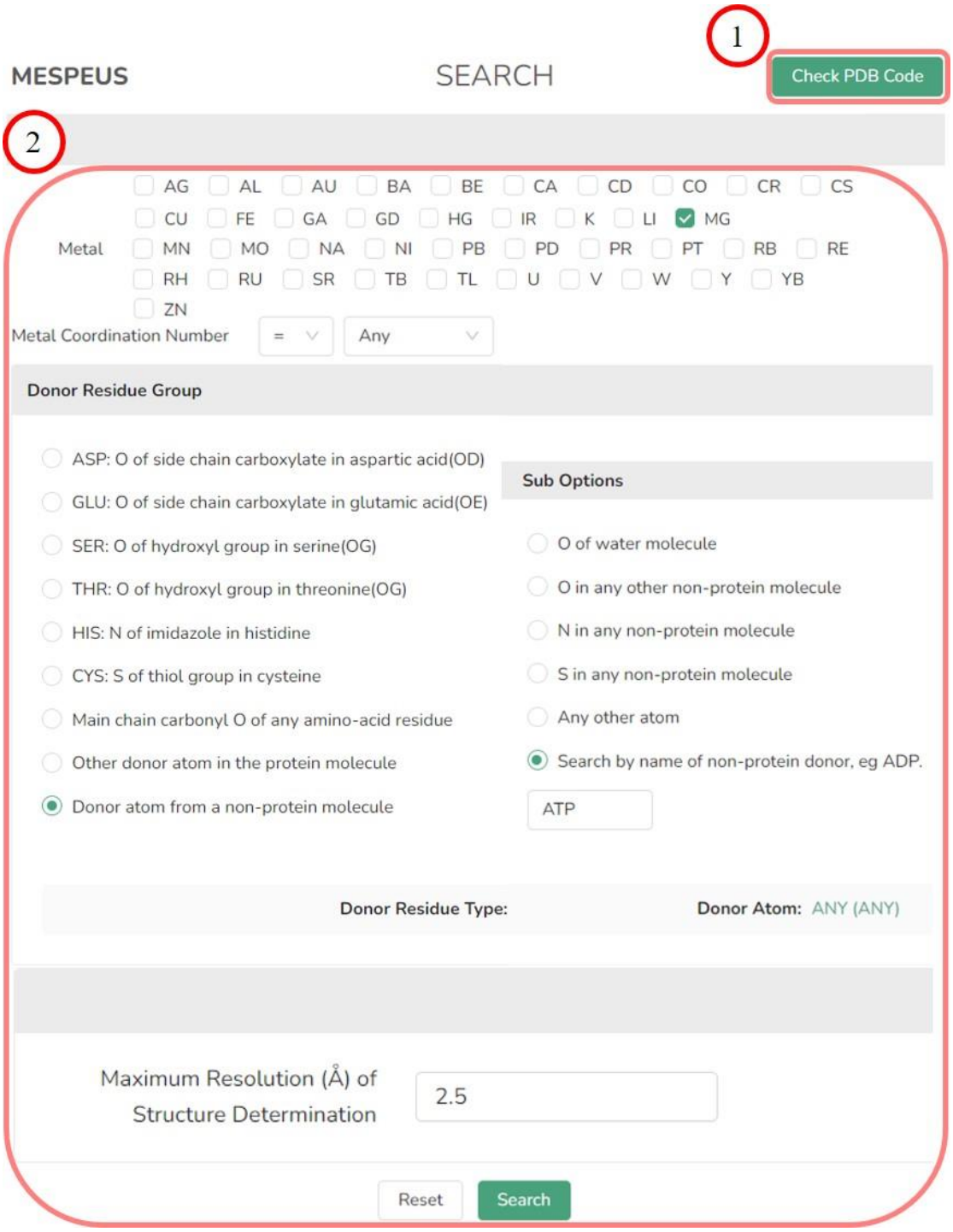
Major Features

1. Both **x-ray** and **cryoEM** models are included in the collection of MESPEUS.
2. Covering **40 metals**: selected are the 10 elements common in biology, known as Na, Mg, K, Ca, Mn, Fe, Co, Ni, Cu, and Zn. Additional 30 metals are also covered, including Ag, Al, Au, Ba, Be, Cd, Cr, Cs, Ga, Gd, Hg, Ir, Li, Mo, Pb, Pd, Pr, Pt, Rb, Re, Rh, Ru, Sr, Tb, Tl, U, V, W, Y, and Yb.
3. In order to completely determine the metal coordination, the **symmetry-related units** of a given protein structure are taken into and are generated using the appropriate space group symmetry operations. This allows a more complete description of the metal coordination geometry by including coordinating atoms that are not in the asymmetric unit.
4. The idea of a '**target distance**' has been applied to the recognition coordination donors which are the most likely distances found from previous studies between the metal of interest and various ligand atoms.
5. The MESPEUS system has been developed to be capable of **automatically updating** its database monthly by synchronizing its data with the storage of Protein Data Bank database and performing metal site analysis as well as content renewal.
6. The database is **freely accessible** and open to all users without login requirement.

Interface operation

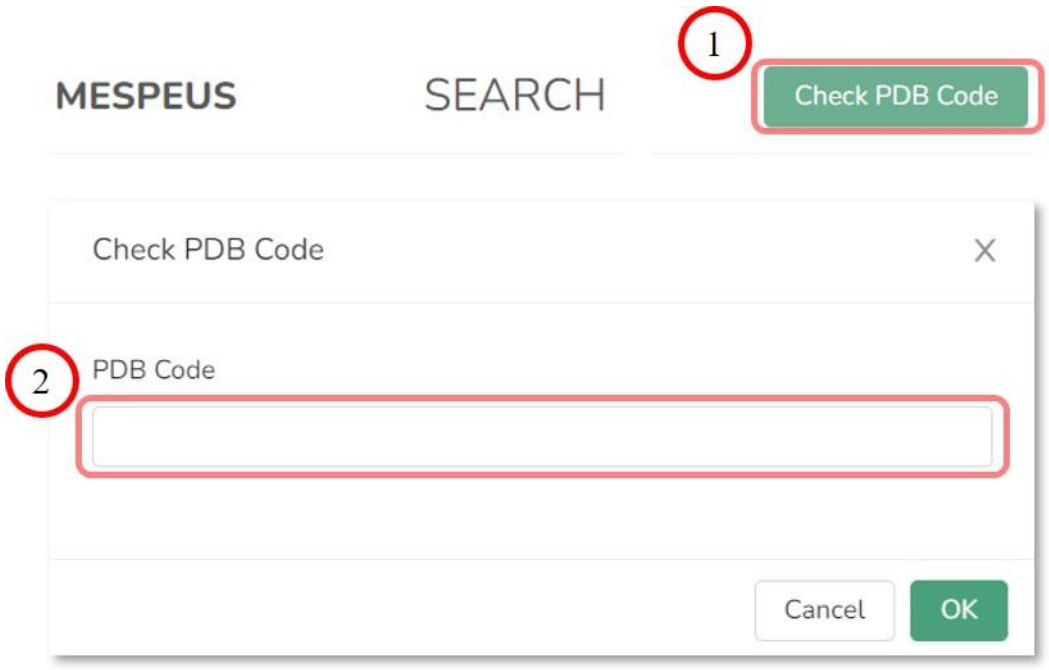
Main query page of the MESPEUS web interface. Two methods proceed query:

1. By directly inputting a PDB ID to investigate all of the metal sites in a protein structure.
2. By specifying a selection of query criteria to search for metal sites of interest.



1. Inputting a PDB ID

1.1 Users can investigate all of the metal sites in a proteins by simply inputting a PDB ID



1. Click on the Check PDB Code button.

2. Input a PDB ID to investigate all of the metal sites in that protein.

1.2 Qeury results of giving a PDB ID

Results showing the metal sites in a metalloprtoein (PDB ID: 6AAP).

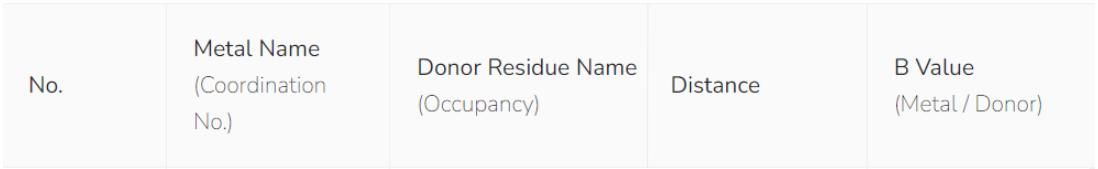
1. The structure model is displayed in 3D by JSmol with a series of interative buttons to operate the model with ease. If the metal site(s) contains donor atoms from a symmetry-related unit(s), the model of the symmetry-related unit(s) will be displayed in translucent mode. Funcitons of the buttons are described as below:

Cartoon on	Display the structure model in cartoon mode.
Gray	Change the color to gray.
Translucent	Display the strucutre model in translucent mode.
Cell	Display the lattice cell.
Reset	Reset the rendering.
Spin On	Spin the strucutre model.
Spin Off	Stop sping the strucutre model.

2. Information of the protein strucutre model derived from the PDB file.

PDB ID	Protein Data Bank Identification Number.
Title	The title of the protein structure.
Class	The classification or family of the prtoein.
Resolution	The structural resoluton of the protein model.
Space Group	The space group of the protein model.
Cell	The information of the lattice cell.
Refinement Program	The program applied to the refinement process.
R Value	The R value of the structural model.
Free R	The R free of the structural model.
No. Metal Sites	The number of the metal sites found in the protein structure.

3. Information of Metal Site(s)

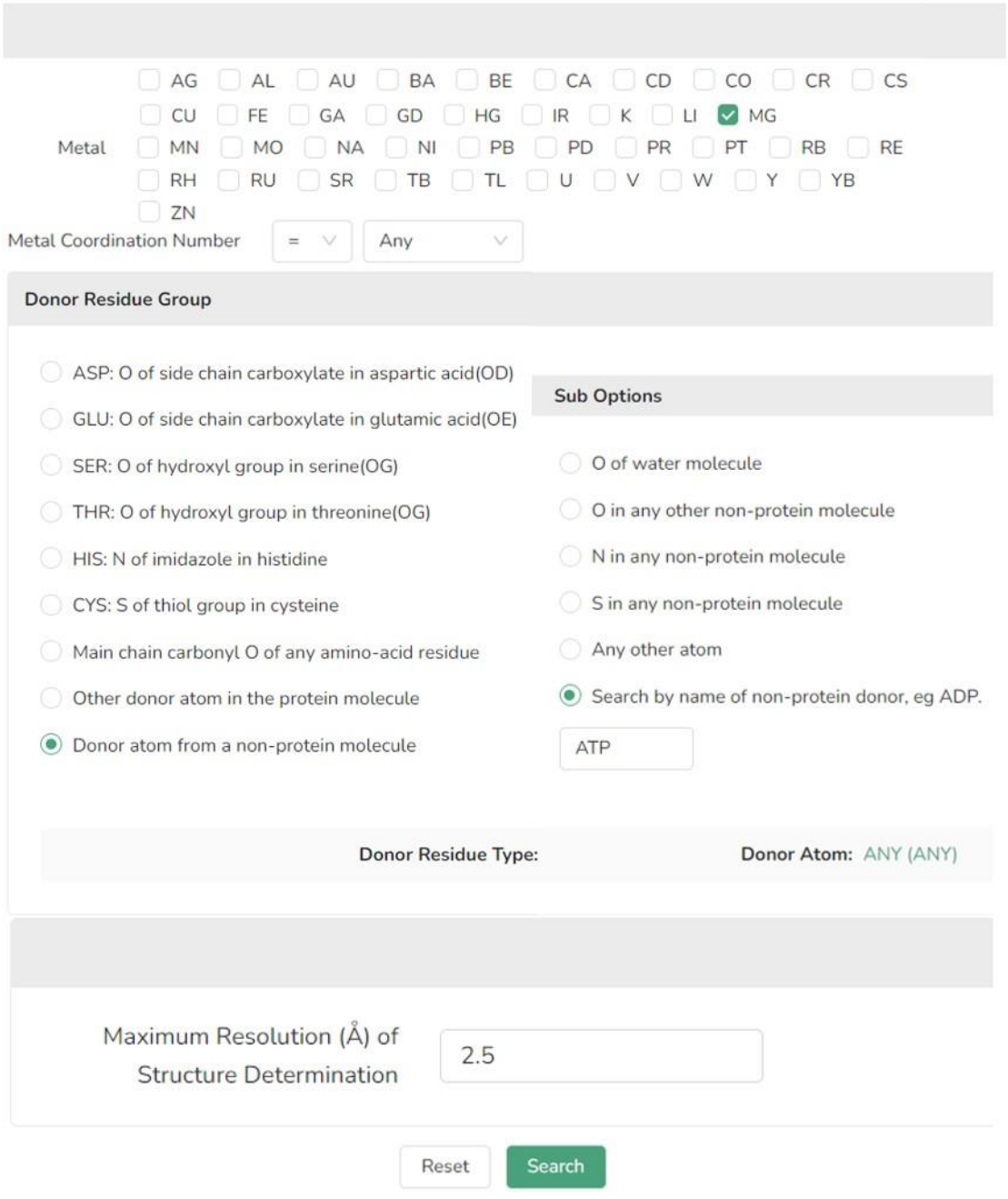


No.	The ID of the metal site found in the structure.
Metal Name (Coordination No.)	Information of the metal and its coordination number. The metal name, metal's residue ID and chain ID are displayed. It is clickable to interactively point out and centre the metal shown on the JSmol.
Donor Residue Name (Occupancy)	The information of the donor residue(s). The donor atom, residue name, chain ID, residue ID, and the occupancy value are displayed. It is clickable to open a page showing the details of the metal site, as described in Section 2.3 .
Distance	The observed metal-donor atom distance.
B Value (Metal/Donor)	The B value of the metal and donor.

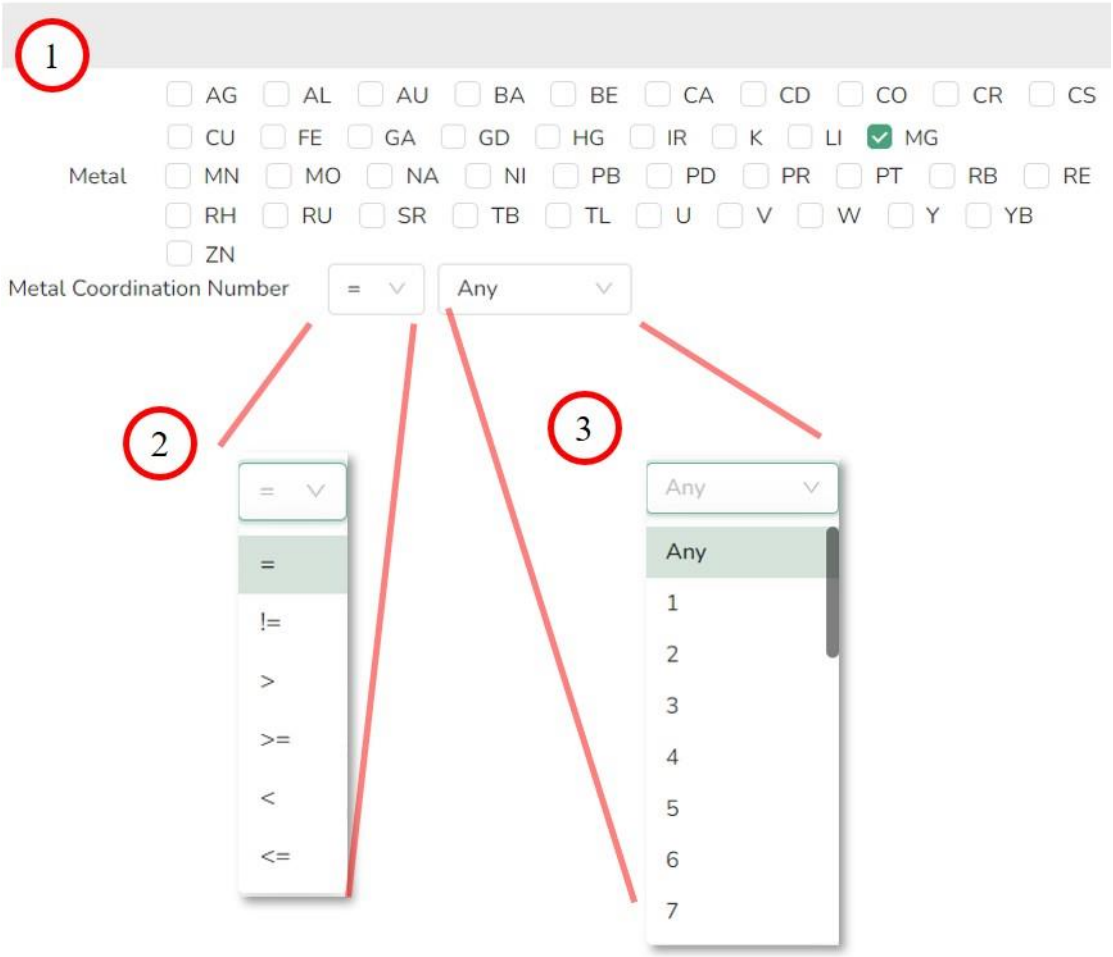
2. Specify query criteria.

2.1 Specifying the query criteria.

The MESPEUS web interface allows users to submit a query through various search options for types of interactions observed in a metal site with a specific resolution.



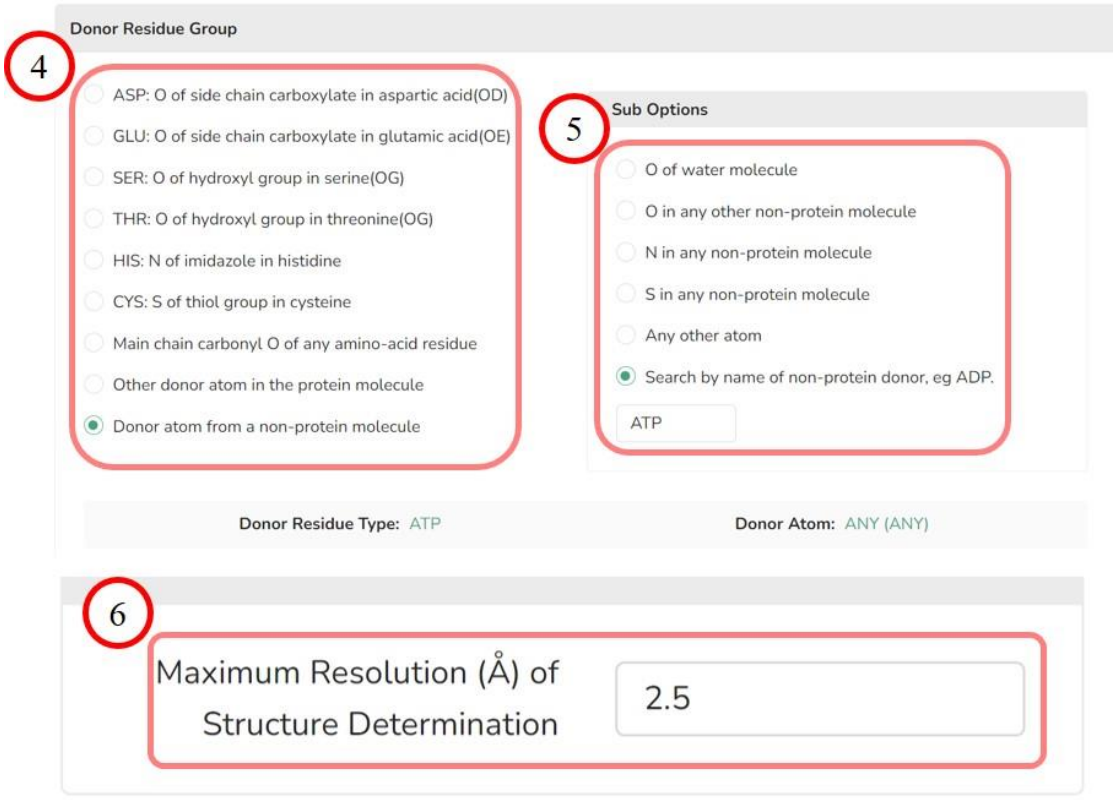
1. Selecting metal(s) of interest for a query.
2. Specifying the logical operator for determining the cooridination number.
3. Specifying the coordination number of the metal site.



4. Specifying the donor residue or molecule involved in the metal site. A list of options defining metal-donor interactions is as tabulated below the figure.

5. Specifying the donor atom or molecule interacting with the metal, which can be an atom type or a molecule identity as shown in the PDB file, such as "ADP" (i.e. the identity shown in the HETATM residue column).

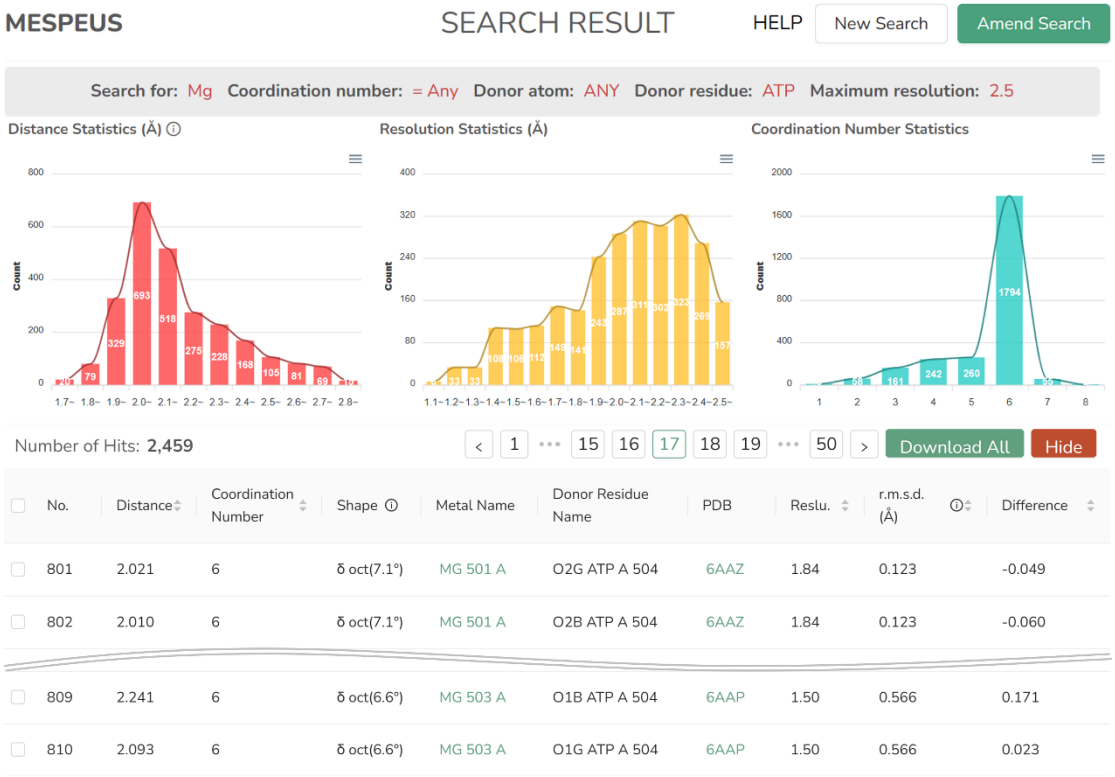
6. Setting up the maximum structural resolution (Å) of the hit proteins.



1	ASP	O of side chain carboxylate in aspartic acid (OD)	Any Monodentate Bidentate
2	GLU	O of side chain carboxylate in glutamic acid (OE)	Any Monodentate Bidentate
3	SER	O of hydroxyl group in serine (OG)	-
4	THR	O of hydroxyl group in threonine (OG)	-
5	HIS	N of imidazole in histidine	ND and NE N of imidazole in histidine (ND) N of imidazole in histidine (NE)
6	CYS	S of thiol group in cysteine	-
7	Main chain carbonyl O of any amino-acid residue		-
8	Other donor atom in the protein molecule		O of amide group of asparagine (OD) O of amide group of glutamine (OE) O of phenolate group of tyrosine (OH) S of methionine side chain (SD) Main Chain N, Any AA (Rare) N in LYS side chain (Very Rare) N in ARG side chain (Very Rare) Any other atom
9	Donor atom from a non-protein molecule		O of water molecule O in any other non-protein molecule N in any non-protein molecule S in any non-protein molecule Any other atom Search by name of non-protein donor

2.2 Search result

Result page of searching for Mg-ATP interaction with the query criteria of metal = Mg, coordination number = Any, donor atom = Any, donor residue = ATP and maximum resolution = 2.5. Users can click on the question marks shown on the page to obtain the relevant information. The details of the page content are described as following sections.

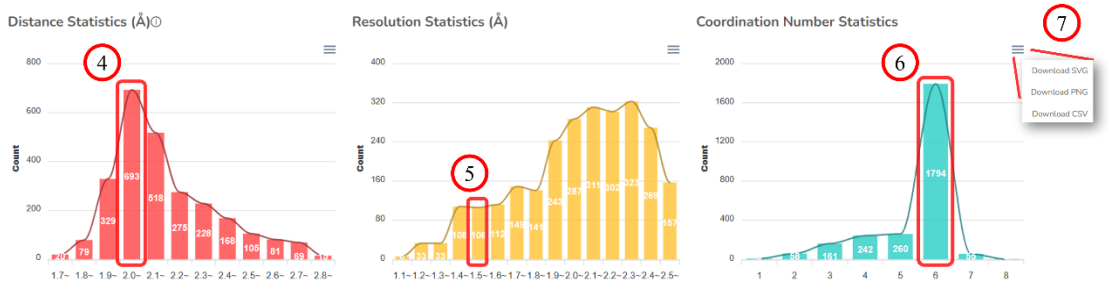


No.	Distance	Coordination Number	Shape	Metal Name	Donor Residue Name	PDB	Reslu.	r.m.s.d. (Å)	Difference
801	2.021	6	δ oct(7.1°)	MG 501 A	O2G ATP A 504	6AAZ	1.84	0.123	-0.049
802	2.010	6	δ oct(7.1°)	MG 501 A	O2B ATP A 504	6AAZ	1.84	0.123	-0.060
809	2.241	6	δ oct(6.6°)	MG 503 A	O1B ATP A 504	6AAP	1.50	0.566	0.171
810	2.093	6	δ oct(6.6°)	MG 503 A	O1G ATP A 504	6AAP	1.50	0.566	0.023

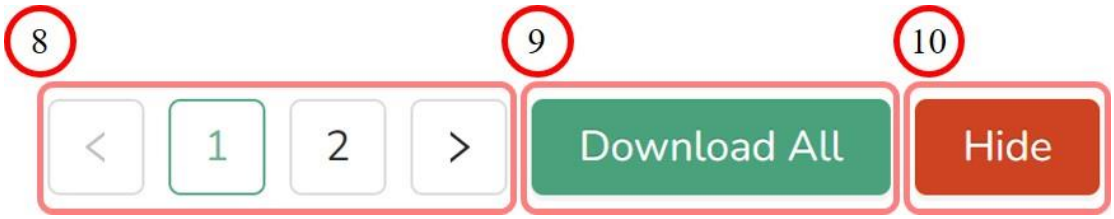
1. New Search button：click on to perform a new search.
2. Amend Search button：click on to perform a amend search.
3. Showing the query criteria of present search.

A set of histograms will be generated to statistically illustrate the search result after a query. Apart from data visualisation, the histograms also function as filters to retrieve a subset from the search result by clicking on the bins. In addition, the curves shown on each histogram display the dynamically updated information. The curve reflects the distribution of the subset when users clicked on the bin of the histogram. The bins of the histograms are separated in order for users to inspect and click on them with ease. The histograms are provided with drilldown enabled function allowing users to examine query results to the desired level of detail. Information of the histograms can be downloaded as svg, png and csv formats. Details as below:

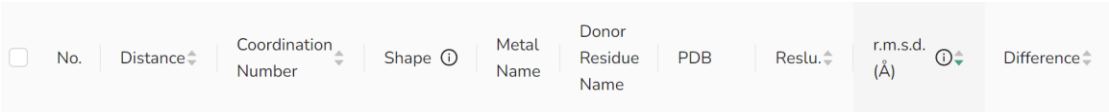
4. A histogram showing the distribution of metal-donor distances of the hit metal sites.
5. A histogram showing the distribution of structure resolution of the hit proteins.
6. A histogram showing the distribution of coordination number of the hit metal sites.
7. Downloading the historgram information in different file formats.



8. Pages showing the query results. Users can click on it to change the page.
9. Click on to download the query results in format of csv for further investigation locally.
10. An exclusion function is available for de-selecting the undesired entries before downloading or calculating distributions.



The columns of the query result table from the left side lists the metal-donor distances measured, coordination number, estimated shape with the RMSD value from the ideal geometry, metal name, donor group (atom name, residue name, chain ID and residue ID), PDB ID, resolution, the RMSD between the metal-donor atom distances and the target distances observed in the metal site and the difference between the two distances.

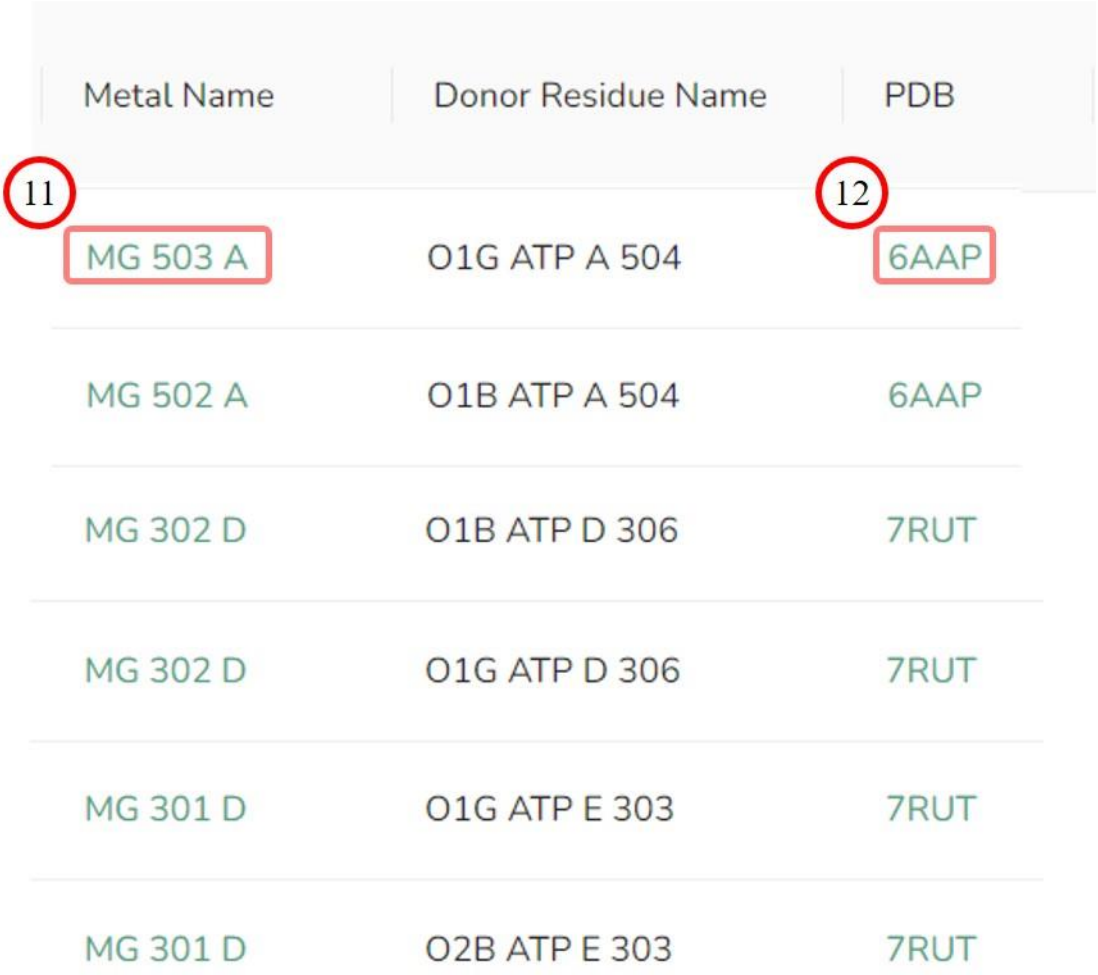


No.	The IDs of the hit entries where user can click on the checkbox to select desired entries before downloading or calculating distributions.
Distance	The observed metal-donor distances (Å). A sorting function is available by clicking on the head of the column.
Coordination Number	The coordination number of a metal site. A sorting function is available by clicking on the head of the column.
Shape	The determined shape of the coordination group. Regardless of bond type, coordination numbers 4, 5 and 6 have regular shapes and interbond angles of a coordination group can be 1) a tetrahedral with equivalent bond angles at 109.5°, or a square planar with angles at 90°, 2) trigonal bipyramid or tetragonal pyramid, and 3) a octahedral with angles at 90°, respectively. A question mark on the head of the column can be clicked on to show the details of the shape determination.
Metal Name	The metal name of the coordination group, presenting the metal identity, residue ID and chain ID as shown in the PDB file. The entries are clickable to open another page to show the detail of the metal site.
Donor Residue Name	The identies of the donor residue, presenting the atom type, residue name, chain ID and residue number as shown in the PDB file.
PDB	The PDB ID of the metalloprotein. The entries are clickable to open another page to show the details of the metalloprotein. A sorting function is available by clicking on the head of the column.
Reslu.	The structural resolution of the metalloprotein model (Å). A sorting function is available by clicking on the head of the

	column.
r.m.s.d. (Å)	Metal "r.m.s.d." is the r.m.s. difference between metal-donor atom distances calculated from the PDB file coordinates, and the target distances. This is quite useful as a quality indicator for the reported distances and the geometry of the site. A question mark on the head of the column can be clicked on to show the details of the shape determination. A sorting function is available by clicking on the head of the column.
Difference	The difference between the the metal-donor atom distances and the target distances observed in the metal site. A sorting function is available by clicking on the head of the column.

11. Click to pop up a webpage with information of the selected metal site, detailed in Section 2.3.

12. Click to pop up a webpage with information of the selected metalloprotein, detailed in Section 1.2.



2.3 Information of a metal site

Page displaying the specified metal site found in a protein (PDB ID: 6AAP). The buttons next to the graphs allow users to manipulate the 3D pictures with ease, such as to centre and rotate the view and to add other information. If the metal site(s) contains donor atoms from a symmetry-related unit(s), the model and the interacting donor(s) of the symmetry-related unit(s) will be displayed in translucent mode.

1. The coordination group of the metal site is shown in 3D by JSmol. Clicking on the "Donor Atom(s)" button can display the identities of the donor groups; "Angle(s)" to display the angles of donor-metal-donor combinations; "Distance(s)" for the distance between the metal and donor atom(s); "2D" for showing the coordination group in 2D generated by LigPlot.

2. The relation of the metal site to the whole metalloprotein. The structure of symmetry-related units displayed by translucent color.

3. The determined shape of the metal site is displayed with the RMSD indicating metal coordination geometry and distortions away from ideality.

4. List of donors. For instance, the metal, Mg (cyan colored), is coordinated with the O atoms from GLU, ATP phosphates and two water molecules, respectively. Among them, two donors from GLU and HOH are located within symmetry-related units. Distance(s) of the donor atom(s) to the metal and the referred B vlaue(s) are displayed.