Public API Documentation

POST /structure endpoint

Introduction

This documentation describes in detail how to use the POST /structure endpoint of 3decision Public API to automatically import a set of structures into 3decision.

You can import structures in different ways into 3decision. Publicly available structures from the PDB are imported automatically for example. Users can also upload their own structures via the graphical user interface of 3decision using the “User data” layout in the dashboard.

However, if you want to import for example all the historical data from your company, or a larger set of structures, then 3decision provides you with a more automated way of doing this. The POST /structure endpoint is described in detail in this document.

Requirements

In order to import a bulk of structures you will need :

1. Your input structures in PDB format and all associated files you would like to register with them, like PDF reports, electron density files etc…
2. A csv file indicating which structure to register with bits and pieces of information that we collect together with a structure.

1 - Structural data and associated files

You have the freedom to organize your structural data the way you want to. However, we would recommend to put at least all structural information etc in a folder a part from the CSV file. Ideally organize a folder per structure, if you have a lot of supplementary files to upload together with a structure. But in the end it is up to you.

2 – Input Archive

The POST /structure endpoint is using an archive zipped file named ‘post_structure.zip’ that must contain:

- single tabulated (or ; separated) file. This file should follow the general formatting guidelines outlined in the next section, and be named ‘post_structure_configuration.csv’
- The files (structure files, reports) that are described in the ‘post_structure_configuration.csv’ file

```
post_structure.zip
├── post_structure_configuration.csv
├── struct1
    ├── structure.pdb
    ├── report.pdf
```

Input CSV file format

3decision associates to each structure a unique structure identifier. Thus, to identify which information corresponds to which structure in your structure upload you have to specify an identifier relating the information to the correct structure(s) you want to register. For example, you want to register a set of PDB files with no further information? Then simply associate to each file a different structureId. Note that the structureId you choose will not be registered by default and used by 3decision.

If you want to register other information together with a structure, another line should be added to the csv file with the same structureId and a keyword will tell 3decision where to register that supplementary information.

You can specify several types of information (column infoValue1, infoValue2, etc…) that are associated to different keywords (column infoType) and keywords must be in uppercase. Some keywords are mandatory, but most are optional.

Again, all the information associated to the same structure to upload must have to the same structureId (column structureId).

Column headers structureId, infoType and infoValue are mandatory. Here a very **basic example** using all mandatory fields:

structureId	infoType	InfoValue1
Struct1	TITLE	The title of the structure
Struct1	STRUCTURE_FILE	projetX/structure.pdb
Struct1	METHOD	X-RAY DIFFRACTION
Struct1	SOURCE	User
Struct1	CREATED_BY	discngine
Struct2	TITLE	The title of the structure2
Struct2	STRUCTURE_FILE	projetA/mystructure.pdb
Struct2	METHOD	X-RAY DIFFRACTION
Struct2	SOURCE	User
Struct2	CREATED_BY	discngine

If you have downloaded this document from the API documentation endpoint, an example archive that can be sent to the POST /structure endpoint will be available at this document location.

Mandatory keywords

STRUCTURE_FILE

structureId	infoType	infoValue1
Struct1	STRUCTURE_FILE	projetX/structure.pdb

Please use unix based file paths ideally without spaces, special characters etc.

TITLE

structureId	infoType	infoValue1
Struct1	TITLE	The title of the structure

METHOD

structureId	infoType	infoValue1
Struct1	METHOD	X-RAY DIFFRACTION

Warning: For now, the method must already exist in 3decision DB. Refer to the end of the documentation for all available experimental methods that you can specify here.

SOURCE

structureId	infoType	infoValue1
Struct1	SOURCE	User

Warning: For now, the structure source must already exist in 3decision (allowed entries are PDB, Registry, Swiss-Model, Historical, User).

CREATED_BY

structureId	infoType	infoValue1
Struct1	CREATED_BY	discngine

Warning: The usename defined in InfoValue1 must be a valid 3decision username

Optional keywords:

LIGAND

This keyword is used to specify a reference 2D structure for ligands present in the structure. This help to avoid registration problems associate to bond order or bond definitions, usually not well represented in pdb structures.

You can specify up to one line per ligand present in the structure, so the same structureId can be associate to multiple LIGAND keywords.

The reference structure can be a SMILES string or an SD File. The same SD File can contain and refer to multiple ligands.

structureId	infoType	infoValue1
Struct1	LIGAND	COc1ccc(CC(Nc2cccc(c2)[Cl])=O)cc1
Struct2	LIGAND	postStructure/MyRefLigands.sdf

Warning: File path and SMILES are case sensitive and reference structure file must have an .sd or .sdf file extension

ANNOTATION

This keyword is used to specify annotations that will be added and associated to the structure in 3decision.

Here, the infoValue1 columns must corresponds to an existing annotation type.

The infoValue2 column corresponds to the value of the annotation.

You can specify multiple annotations for the same structure and so have multiple ANNOTATION keywords for the same structureId.

Annotation Link : You can append a link to an annotation by adding an optionnal infoValue3 column to structure annotation

structureId	infoType	infoValue1	infoValue2	infoValue3
Struct1	ANNOTATION	Internal ID	X-PDB1000	
Struct2	ANNOTATION	Comment	Test Structure	
Struct2	ANNOTATION	PFAM Annotation	PFAM ID	link

PROJECT_ID

This keyword is used to specify the projects that will be associated to the structure in 3decision. If the project does not exists, or you don't have writing permission on it, then a error will be returned.

You can specify multiple project identifiers for the same structure and so have multiple PROJECT_ID keywords for the same structureId.

structureId	infoType	infoValue1
Struct1	PROJECT_ID	1
Struct2	PROJECT_ID	2

NEW_PROJECT

This keyword is used to specify new project to be created and to associate the structures to.

Each project must have a project id (that will not be used to register projects in 3decision, but that will be used to mapped linked files), which must be in the format : 'Project_xxx'. The project name can be specified in the infoValue2 (100 characters maximum) column, and the project description can be specified in the infoValue3 (400 characters maximum) column. If several lines contain the NEW_PROJECT with the same project name, only one will be created and each structure that is mapped to it in the file will be added to the project. If several NEW_PROJECT lines exists and have different description, an error will be returned.

structureId	infoType	infoValue1	infoValue2	infoValue3
Struct1	NEW_PROJECT	Project_001	New Project	Project description
Struct2	NEW_PROJECT	Project_001	New Project	Project description

Warning:

The structures will always be associated to your personal project, even if you specify other projects to for the structures to be mapped to.

RESOLUTION

This keyword is used to specify the structure resolution if relevant.

structureId	infoType	infoValue1
Struct1	RESOLUTION	1.61

Warning: Resolution must use a dot (.) as decimal separator.

UNIPROT_MAPPING

This keyword is used to specify manually which chain corresponds to which protein in the structure. Here you can associate a chain to either a uniprot id or parts of the chain to a uniprot id. In infoValue1 column, you can specify the protein chain to apply the mapping to. In infoValue2, you can specify the uniprot code to map the chain to. Optionally, in infoValue3, you can specify the residue sequence as described above.

The sequence range is optional but must be specified with the format ‘xxx-yyy’ if used. Note, if no uniprot mapping is specified , 3decision will try to automatically guess the protein you have in your structure files. This works reasonably well for most of the cases. If you want to make sure that 3decision uses your protein assignments, then you must specify this section for each chain in the structure. Furthermore, if you have a structure of a chimera, we recommend to define your chain mappings carefully yourself in the input CSV file.

structureId	infoType	infoValue1	infoValue2	infoValue3
Struct1	UNIPROT_MAPPING	A	NUDT7_HUMAN	
Struct1	UNIPROT_MAPPING	B	NUDT7_HUMAN	100-200

Warning:

- You can’t have two lines associate to the same chain with and without a sequence range
- If you specified the mapping, you must do it for all the protein chains of the structure
- You can't map less than 90% of the residues resolves for a chain

LINKED_FILE

3decision allows you to register files together with a structure. This is particularly useful to keep a structure together with original reports (pdf files for instance) or electron density maps (ccp4 or mtz). The LINKED_FILE keyword allows you to add these files to a structure.

A file description can be added in the infoValue2 column. A structure can be linked to a structure AND a project, which ids can be specified in the infoValue3 columns (you can use existing project ids, or project ids described in the infoValue1 of NEW_PROJECT lines). If none are specified, the structure file will be linked to your personal project. You can also specify ‘All’ to associate the structure files to all mapped projects, and your personal project. **If you specify ‘None’ in infoValue3, the file will be directly linked to the structure (see warning below).**

You can link more than one file to a structure, so you can have more than one LINKED_FILE for a structureId.

structureId	infoType	infoValue1	infoValue2	infoValue3
Struct1	LINKED_FILE	Struct1/report1.pdf	A description	All
Struct1	LINKED_FILE	Struct1/report2.pdf		1, 2, Project_001

Warning:

- File paths are case sensitive.
- The file extension must be registered in 3decision DB. Refer to the end of the documentation for all supported file types.
- **Specifying 'None' in infoValue3 will link the file directly to the structure and everyone that has access to the structure will have access to the file, regardless of which project they are a part of.**

STRUCTURE_RELATION

You can also register now relations between structures in 3decision. This is particularly useful, if you refined a structure and want to register the refined version, while keeping track of the unrefined version. You can also register derived structures, like biomolecular assemblies or structures where crystallization stabilizers (protein chains) have been stripped off. To properly annotate from which structure such an upload originates you can specify a "relation".

A relation is defined between a current structureId and either another structureId in the same bulk import, or an existing structure code in 3decision (PDB code, or 6 alphanumerical code in 3decision). In infoValue1, specify if the structure specified in structureId is the parent or the child in the relation. (Accepted values are 'Parent' and 'Child'). In infoValue2, specify the type of the relation (see below for accepted values). In infoValue3, specify the structureId of the structure to put in relation, and in infoValue4 the description of the relation.

structureId	infoType	infoValue1	infoValue2	infoValue3	infoValue4
Struct1	STRUCTURE_RELATION	Parent	Derived	1uyd	description

Note that if the associated structure is an existing private structure, you must have the writing rights to create the relation.

Note that for now structure relations can only encompass the following types :

Refined: Child structure is obtained after modification of parent structure parameters to better match to experimental data

Derived: Child structure is obtained after modification of parent structure. For example, you reconstruct a biomolecule from an asymmetric unit or you drop a chain from an initial structure, then it can be defined as derived.

Prepared for docking: Child structure is a ready-to-dock version of the parent structure

CUSTOM_TRANSFORMATION

In 3decision, in the project dashboard, a user can select a reference structure for a given project. All other structures in the project will be superimposed onto the chosen reference, to ease up the loading of project of superposition of structures within 3decision. However, some structures can be challenging to superimpose on the fly. 3decision now provides a way for users to specify the transformation matrix to be used to superimpose a structure to the selected reference of the project. The format to specify the custom transformation matrix is as follows: the InfoType column should be set to the 'CUSTOM_TRANSFORMATION' keyword, and the infoValue1 must represent a 4x4 transformation matrix, with each number separated by an underscore '_' to avoid all ambiguities between commas, semicolons, points, in the numbers and csv format. In the example below, the infoValue1 has been splitted with new lines due to the document format, however the **value must be read as a single line**:

0.9396196538634071_-0.19129380982798871_-0.28376329118912447_-65.1830952000429_0.2583808525231708_0.940256245471734_0.2217148602698105_-89.94231339097247_0.22439755606405637_-0.28164673155705944_0.932909855426588_0.5420021367033776_0.0_0.0_0.0_1.0

A custom transformation must be associated to a structure AND a project: the project ids to map the custom transformation to must be referenced in the infoValue2 column. You can use existing project ids, or project ids described in the infoValue1 of NEW_PROJECT lines. Multiple projects can be specified, provided that the project ids are separated by commas.

structureId	infoType	infoValue1	infoValue2
Struct1	CUSTOM_TRANSFORMATION	0.9396196538634071_ -0.19129380982798871_ -0.28376329118912447_ -65.1830952000429_ 0.2583808525231708_ 0.940256245471734_ 0.2217148602698105_ -89.94231339097247_ 0.22439755606405637_ -0.28164673155705944_ 0.932909855426588_ 0.5420021367033776_ 0.0_ 0.0_ 0.0_ 1.0	1, Project_001

Please note that the workspace of a project that has at least one custom transformation cannot be opened without first setting a reference for the project. Using custom transformation for the structure reference of a project is also supported. Beware that we do not have ways to programmatically manage the custom transformations, but new endpoints to do this will be added in an upcoming release.

LIGAND_EXTERNAL_ID

3decision allows to add identifiers to small molecules in structures. These identifiers will be then searchable and displayed within 3decision. To add an identifier to a ligand, you must specify the residue name (3 letter code) to which the identifier refers to in the **infoValue1** column, and the identifier itself in the **infoValue2** column. A residue name for a given structure can only have a single identifier.

structureId	infoType	infoValue1	infoValue2
Struct1	LIGAND_EXTERNAL_ID	LIG	L12345

Allowed experimental methods

The following list resumes all experimental methods that are currently allowed in 3decision. This list is based on available methods in the RCSB PDB. If one is missing in this list, please contact us.

```
ELECTRON CRYSTALLOGRAPHY
ELECTRON MICROSCOPY
ELECTRON MICROSCOPY;
ELECTRON MICROSCOPY; SOLID-STATE NMR
ELECTRON MICROSCOPY; SOLUTION SCATTERING
FIBER DIFFRACTION
FIBER DIFFRACTION; SOLID-STATE NMR
FLUORESCENCE TRANSFER
INFRARED SPECTROSCOPY
Model
NEUTRON DIFFRACTION
NEUTRON DIFFRACTION; SOLUTION NMR
NEUTRON DIFFRACTION; X-RAY DIFFRACTION
POWDER DIFFRACTION
SOLID-STATE NMR
SOLID-STATE NMR; ELECTRON MICROSCOPY
SOLID-STATE NMR; SOLUTION SCATTERING; ELECTRON MICROSCOPY
SOLUTION NMR
SOLUTION NMR ; THEORETICAL MODEL
SOLUTION NMR; SOLID-STATE NMR
SOLUTION NMR; SOLUTION SCATTERING
SOLUTION NMR; THEORETICAL MODEL
SOLUTION SCATTERING
SOLUTION SCATTERING; SOLUTION NMR
Swiss-Model
X-RAY DIFFRACTION
X-RAY DIFFRACTION; EPR
```

```
X-RAY DIFFRACTION; NEUTRON DIFFRACTION
X-RAY DIFFRACTION; SOLUTION NMR
```

Allowed file types

The following list resumes all file extensions that are currently supported in 3decision. This list is going to evolve, so if one is missing in this list, please contact us.

```
brix
ccp4
cns
csv
cube
doc
docx
dsn6
dx
dxbin
ent
gro
map
mmcif
mmtf
mol
mol2
mrc
mtz
pdb
pdf
ppt
pptx
pqr
sd
sdf
txt
xls
xlsx
xplor
zip
```