MotiveValidator

User Manual

Contents

1. Statement of purpose

The advancement of research in structural biology has provided a large body of structural data deposited in various databases. One great example is the Protein Data Bank (PDB), which has been growing exponentially, and which currently consists of more than 90000 structures of biomolecules and their complexes. Such large bodies of data, especially accumulated over a short period of time and using high throughput techniques, will inherently be plagued by a variety of problems.

Validation arose as a major issue in the structural biology community when it became apparent that some published structures contained serious errors, either documented (e.g., due to insufficient electron density in a certain area), or not. Structural databases generally require that the new submissions be checked prior to acceptance. The tools employed for presubmission validations work fairly well for well studied residues like amino acids or nucleotides. However, an essential step in the validation process is checking the ligand structure, because ligands play a key role in protein function, and also because they are the main source of errors in structures. A notable case of ligand validation is the analysis of carbohydrate structures, because they have complex topology and many chiral atoms. Yet carbohydrates are involved in a variety of fundamental biological processes and they have large pharmaceutical and diagnostic potential. Additionally, more than 60% of nontrivial-sized ligands (> 10 atoms) from the Protein Data Bank contain a carbohydrate. In recent years, many algorithms for validation, ligand validation and carbohydrate validation have been developed. Nonetheless, significant limitations persist, such as insufficient coverage of ligands and time inefficiency (i.e., calculations are time demanding and only one entry can be validated in each run).

We have developed **MotiveValidator**, a user-friendly, interactive and platform independent environment for the speedy validation of ligands, residues and fragments (denoted as *structural motifs*). **MotiveValidator** covers all standard and custom residues and ligands, and was successfully tested in six research labs on more than 50.000 input samples.

2. How to use this manual

In the following sections we offer an extensive tour through the **MotiveValidator** functionality. The elements of the user interface are described as we go along. The explanations are both visual and textual.

Note that the web page also provides a quick tutorial in addition to this manual. The tutorial is meant for a quick start for first time users, who would like to try out **MotiveValidator** without going through all the explanations in the manual. The tutorial offers a brief, graphical walk through of job submission and result analysis using **MotiveValidator**. Additionally, to illustrate the way results can be analyzed, a few sample calculations are available on the web page for execution and download.

Enjoy working with MotiveValidator!

3. Availability and technical details

3.1. Where to find MotiveValidator

MotiveValidator is freely available via the internet since September 2013 at http://ncbr.muni.cz/MotiveValidator. There is no login requirement for using **MotiveValidator**.

MotiveValidator employs advanced algorithms for the comparison of structural motifs developed in our labs (SiteBinder[1]), and takes advantage of our know-how from the development of web applications for the detection of protein structural motifs (MOLE[2]). Additionally, **MotiveValidator** employs Open Babel[3] for chirality verification and ChemDoodle[4] for interactive visualization of 3D structures. The complete theoretical and methodological background is described in the respective papers or web resources. Finally, **MotiveValidator** employs an in-house chemical language for the detection and extraction of residues (MotiveQuery, D. Sehnal et al., unpublished work), and in-house algorithms for the statistical evaluation of results.

3.2. What you need in order to run MotiveValidator

MotiveValidator is basically a collection of several web applications, therefore you do not need to install it on your computer. It runs on the ncbr.chemi.muni.cz server at the National Centre for Biomolecular Research within Masaryk University, Czech Republic. All you need in order to use **MotiveValidator** is an internet browser that is up to date and has JavaScript enabled, and a working internet connection. Since all calculations are run on the server, the only functionality that relies on your system is the display of 3D models, for which your browser will need to support WebGL. If you experience trouble displaying the 3D models, please check http://get.webgl.org in order to find out how to enable WebGL on your system.

3.3. How to get around the web page

As soon as you type in the address http://ncbr.muni.cz/MotiveValidator, you will reach the **MotiveValidator submission page**, which contains a brief, general description of **MotiveValidator**, along with 5 tabs (Figure 1 A). These tabs allow you to access the various parts of the **MotiveValidator** functionality. To see what a tab is meant for, just click on it. Once you have submitted your calculation and the results are ready, you will be redirected to the **MotiveValidator results page**, which allows you to analyze the results and download the data (Figure 1 B). The **results page** is also organized into tabs that allow different levels of analysis of the results.

Basic orientation in the web page is always available in the **Quick Help** tab of the **submission page** (Figure 1 A), which is the default tab open when you first access **MotiveValidator**. Last but not least, note that all sections of the **MotiveValidator** web pages contain tool tips to aide you along the way. When in doubt about what a button does or what something means, simply hover the mouse over it.

---

1 Sehnal D, Svobodová Vařeková R, Huber HJ, Geidl S, Ionescu CM, Wimmerová M, Koča J, *Journal of Chemical Information and Modeling* 52(2), (2012): 343-359.
2 Berka K, Hanák O, Sehnal D, Banáš P, Navrátilová V, Jaiswal D, Ionescu C, Svobodová Vařeková R, Koča J, Otyepka M, *Nucleic Acids Research* 40, W1 (2012): W222-W227.
3 O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR, *Journal of Chemoinformatics* 3, (2011): 33.
4 http://www.chemdoodle.com

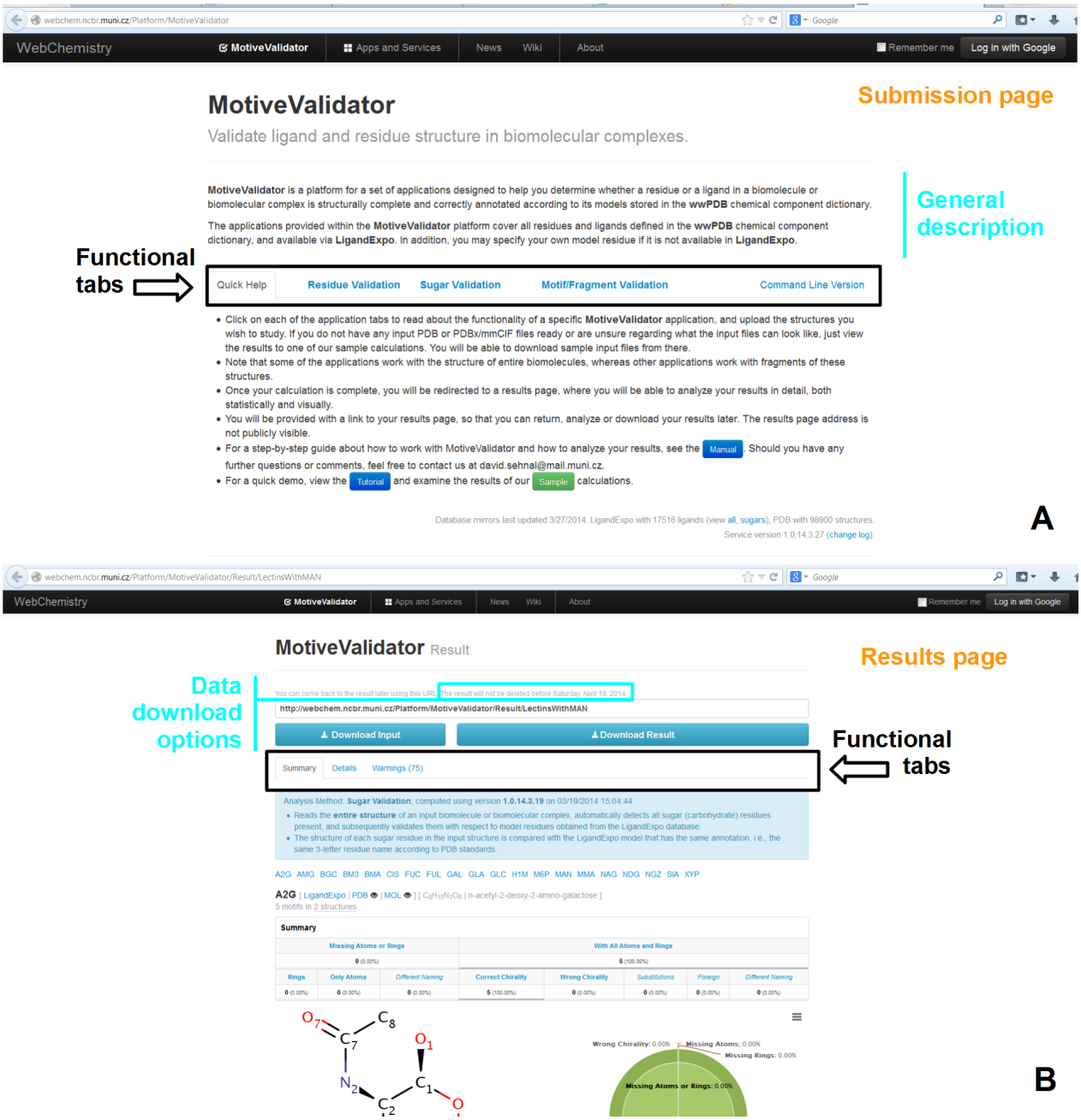

Figure 1: ***MotiveValidator*** *web pages: A) The* ***submission page*** *contains a general description, and is organized in 5 functional tabs, namely Quick Help, Residue Validation, Sugar Validation, Motif/Fragment Validation, and Command line version. Click on a tab to see the functionality it provides. B) The* ***results page*** *contains a few options to access your data, and is organized in a few functional tabs, namely Summary and Details, which allow different levels of analysis. An additional tab with Warnings or Processing errors may appear if issues are detected in the input files. Your results will be available on the server for a month. You may download the input or output data using the Download buttons.*

4. Basic terms

Before moving on to more extensive descriptions of functionality, it is important to clearly establish the meaning of a few key terms within the **MotiveValidator** environment.

4.1. Residue

We use the term *residue* to refer to any component of a biomacromolecule or a biomacromolecular complex. This includes amino acid residues and nucleotides, which are commonly referred to as residues as they form proteins and nucleic acids. Within the **MotiveValidator** environment, any collection of atoms bound by chemical bonds (covalent, coordinative or ionic) can be considered a residue as long as this fact is appropriately indicated in the input PDB file. Specifically, all the atoms that make up a residue should have the same *residue name* (3-letter code) and *residue identifier* (index internal to the input PDB file).

4.2. Ligand

We use the term *ligand* to refer to a chemical compound which forms a complex with a biomacromolecule (e.g., sugar, drug, heme). Ions can also function as self standing ligands, or they can be part of a residue (such as Fe in heme). In the PDB format, a ligand has its own residue identifier and 3-letter code, and is composed from HETATM records. The **MotiveValidator** term *residue* (section 4.1) thus fully covers ligands, in addition to typical components like amino acids and nucleotides.

4.3. Sugar

We use the term *sugar* to refer to the special case when a residue belongs to a carbohydrate (saccharide).

4.4. Motif / Fragment

With respect to the chemistry of biomolecules, the term *motif* is used to refer to a well defined distribution of structural elements in a biomolecule or biomolecular complex, with characteristics generally associated with a specific function. Within the **MotiveValidator** environment, a *motif* is generally a fragment of a biomacromolecule, biomacromolecular complex or ligand, made up of one or more residues or parts of residues. A *motif* can in principle be any fragment of a biomolecule. Nonetheless, **MotiveValidator** is focused on the validation of residues, thus here motif generally refers to a fragment made up from the residue under study, together with its surroundings (i.e., atoms from neighboring residues). Note that the terms *fragment* and *motif* are used as synonyms in this manual.

We can generally say that, within the **MotiveValidator** environment, all *residues* can be thought of as *motifs*. Therefore, different *instances of the same residue* (such as multiple arginine residues throughout the sequence of a protein, or copies of the same ligand in different monomers) can be considered and processed as different motifs, making their identification straightforward and unambiguous.

4.5. Model residue

We use the term *model residue* (or simply *model*) to refer to a particular structure that is known to be correct. This structure will then be used as reference template in the validation process, whereby a query residue with the same name (3-letter code) as the model will be compared to the model. Within the **MotiveValidator** environment, a model contains one residue. The origin of the *model* can be the wwPDB chemical component dictionary accessible via LigandExpo[5], or a custom model provided by the user.

5. Input file requirements

MotiveValidator is meant for residue based validations against LigandExpo models, and thus accepts input files in PDB format. The PDB format is necessary due to the fact that it contains residue information (the 3-letter code residue name and residue identifier). However, especially in the case of more unusual residues or user defined motifs, it is many times useful to also submit your structures in SD/SDF/MOL format along with the PDB format. This is to insure that **MotiveValidator** identifies inter-atomic bonds correctly. Additionally, since the representation of large biomacromolecules and their complexes is moving towards a more general format, **MotiveValidator** also allows to upload the structure to be validated in PDBx/mmCIF format.

The PDB file format[6] is well established. The following fields must appear correctly in your input PDB files: atom index, atom name, element symbol, residue name, residue index, 3D coordinates. Alternate locations of atoms are ignored. If **MotiveValidator** finds any issues in the input files, it will report them as warnings or processing errors (details in section 6.2.4).

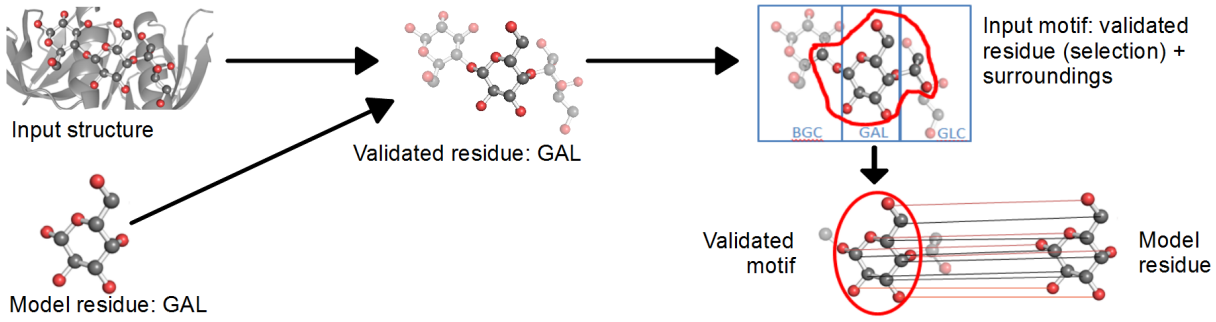

*Figure 2: **MotiveValidator** identifies the validated residue in the input structure based on the name (3-letter code) of the model residue. Then it extracts the validated residue in the form of an input motif, containing all the atoms of the validated residue, together with all atoms found within one or two bonds of any atom from the validated residue (surroundings). The validated motif results as the subset of atoms in the input motif which correspond to atoms in the model residue.*

6. Functionality

MotiveValidator is a platform for a set of applications designed to help you determine whether a residue, ligand, or fragment in a biomolecule or biomolecular complex is structurally complete and correctly annotated. Specifically, **MotiveValidator** checks if the topology and chirality are correct. Validation is performed against model residues from the wwPDB chemical component dictionary. **MotiveValidator** covers all residues and ligands accessible via LigandExpo.

5 Ligand Expo, available at http://ligand-expo.rcsb.org/, provides chemical and structural information about small molecules within the structure entries of the Protein Data Bank.

6 Read more about the PDB format at http://wwpdb.org/documentation/format33/v3.3.html

Additionally, you may upload your own custom model in case the residue, ligand or fragment of your interest does not have any model available in LigandExpo. However, please note that the user bears all responsibility for the correctness of uploaded models.

In its most basic form, the function of **MotiveValidator** is to take a residue and validate it by comparing its structure against the structure of a model residue that shares the same name (3-letter code) with the residue of interest (Figure 2). Therefore, **MotiveValidator** generally requires two input files, namely a file containing the residue to be validated (*input structure*), and a file containing the *model* to be used as reference. Note that, most commonly, the input structure containing the *validated residue* actually includes entire biomolecules (or their fragments) and other ligands. For this reason, the interface of **MotiveValidator** in fact uses the term *biomolecule* to refer to the input file containing *validated residues*. This simply means that you need not extract the *validated residue* yourself before you can validate it. **MotiveValidator** will do that for you.

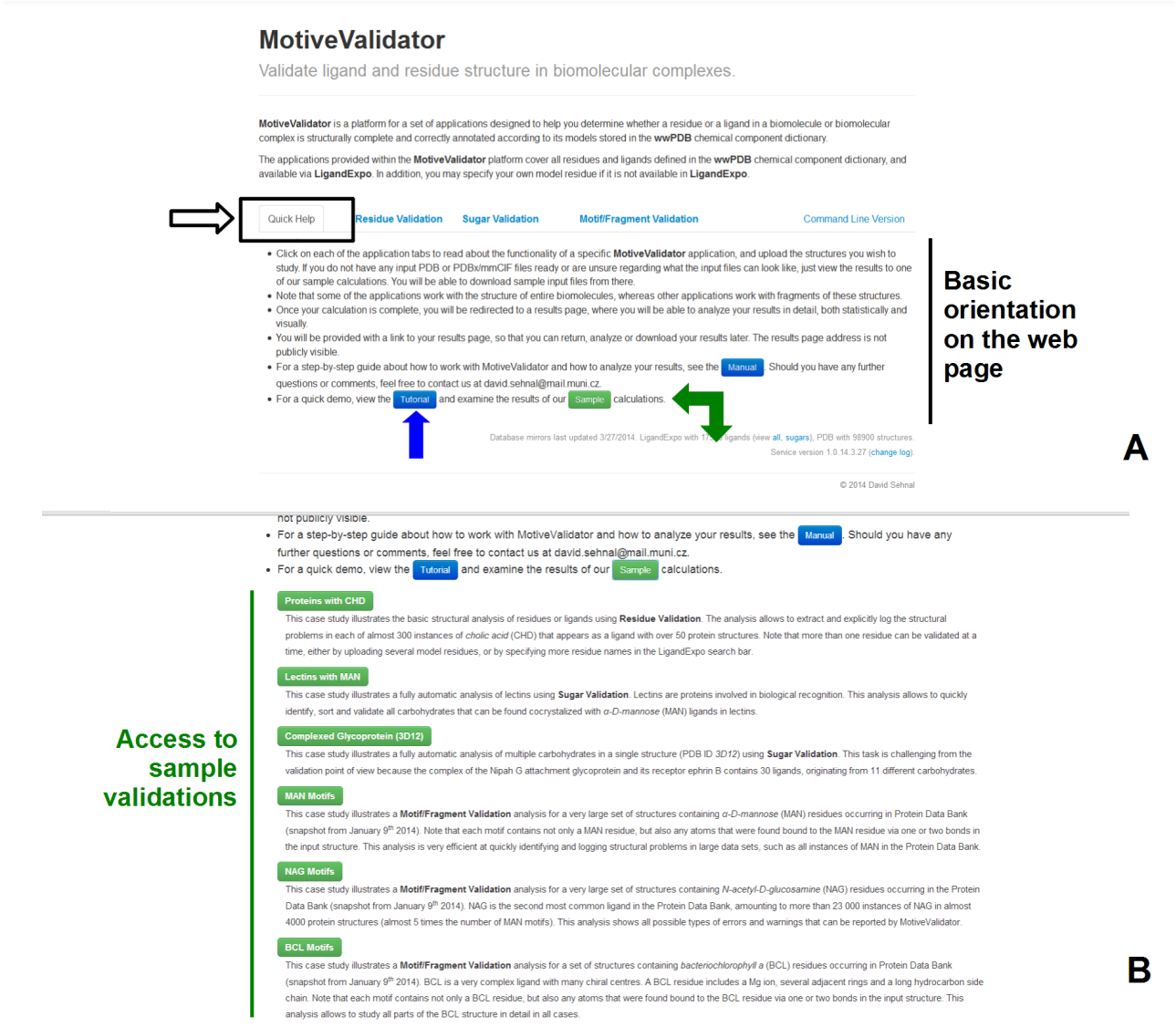

*Figure 3: A) The **Quick Help** tab offers basic tips on how to get oriented on the web page. Additionally, it provides access to the tutorial and sample calculations via two buttons at the bottom of the page. B) Once you click on the Sample button, scroll down to see a list of 6 sample validations that you can access by their respective buttons. Brief descriptions of the sample validations are also given here.*

Once you have initiated a validation, **MotiveValidator** redirects you to the **results page**. The **MotiveValidator** output provides a straightforward report of the validation results, including a summary and detailed information in both tabular and graphical form, along with a 3D structure visualizer for closer inspection of the problematic structures.

Depending on which type of validation you wish to perform, the procedure to load input files and initiate a validation is slightly different, and thus discussed separately in this manual (sections 6.1.1-6.1.3). Nonetheless, regardless of the type of validation you chose on the **submission page**, the presentation of results in **MotiveValidator** is unified, and will therefore be treated as a single unit in this manual (section 6.2).

6.1. File submission

All submissions of input files are made on the **MotiveValidator submission page**, via the appropriate functional tab (Figure 1 A). Which tab is appropriate depends on the type of validation you would like to perform. **MotiveValidator** currently allows three kinds of validations, each in its respective tab: **Residue Validation**, **Sugar Validation**, and **Motif/Fragment Validation**. Two additional tabs appear on the submission page, namely **Quick Help** and **Command line version**. Simply click on any functional tab in order to access it.

The **Quick Help** tab (Figure 3 A) is meant for basic orientation on the **MotiveValidator** web page, and also provides access to the tutorial and 6 sample calculations, accessible via their respective buttons (Figure 3 B). A brief description is provided for each sample calculation on the **Quick Help** tab.

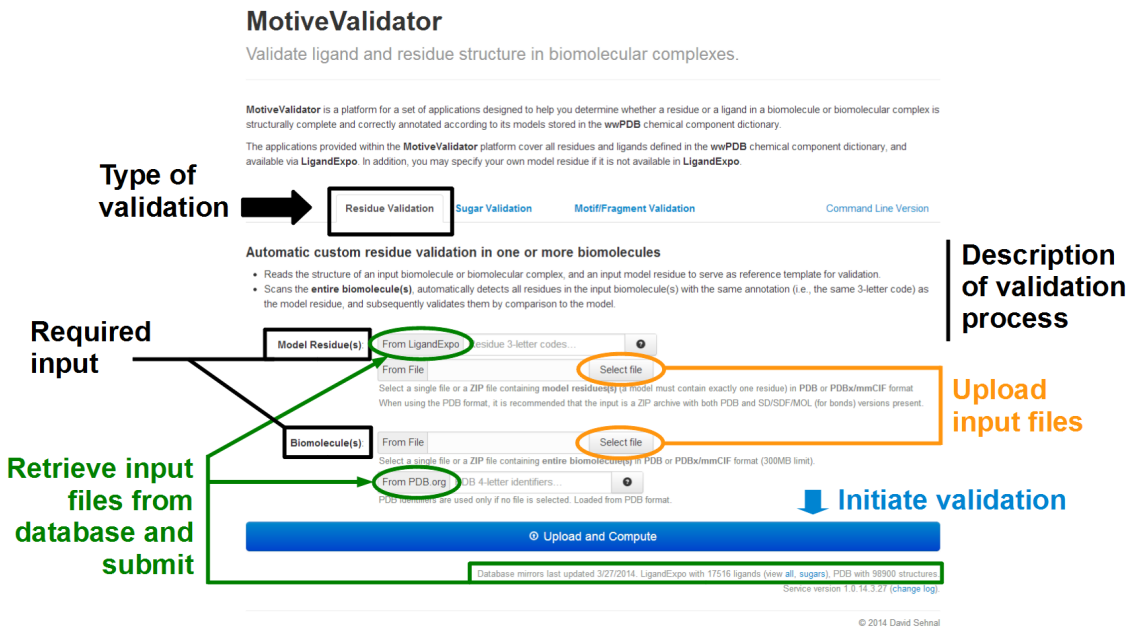

*Figure 4: The **Residue Validation** tab allows to validate residues or ligands in one or more biomolecules. Two input files are necessary: the first with the structure of the model that will be used as reference, and the second containing the residue to be validated. Input files can be retrieved from mirrors of LigandExpo and Protein Data Bank maintained on our server, or uploaded from your computer in PDB or PDBx/mmCIF format. More residues can be validated in one run if you retrieve a list of comma separated residue names (3-letter codes) from LigandExpo, or upload a .zip archive with several model residues. Similarly, thousands of biomolecules can be processed in one run if you upload a .zip archive with biomolecule files, or retrieve a list of comma separated PDB IDs (4-letter identifiers). Once you have picked your input files, click the Upload and Compute button at the bottom of the page to initiate the validation.*

We further describe in detail the functional tabs **Residue Validation**, **Sugar Validation**, and **Motif/Fragment Validation** on the **submission page** (Figures 4-6), as they allow you to upload input files and initiate validations.

6.1.1. Residue Validation

The first type of validation that can be performed using **MotiveValidator** can be initiated via the **Residue Validation** tab on the **submission page** (Figure 4). The **Residue Validation** tab provides a brief description of the validation process, and several options for input file submission.

Remember that the term *residue* refers here to any collection of atoms bound by chemical bonds (covalent, coordinative or ionic), as long as all the atoms have the same *residue name* (3-letter code) and *residue identifier* (index internal to the input PDB file). Therefore any set of atoms that can be defined as a residue (or ligand) within the **MotiveValidator** environment can be validated using **Residue Validation**. **MotiveValidator** will first read in the *model* file and establish the residue name and structural information for all its atoms. **MotiveValidator** will then scan the entire input biomolecule and extract all instances of residues with the same name (3-letter code) as the model. Each of these instances will be considered a separate motif, receive a unique motif identifier and subsequently be validated against the structure of the model residue.

The input files may be uploaded from your computer, or retrieved from the corresponding databases. The model may be retrieved from LigandExpo by specifying its 3-letter identifier according to the PDB residue naming convention. Similarly, the biomolecule may be retrieved from the Protein Data Bank by specifying its 4-letter PDB ID. Capitalization is not important, so 1tqn, 1TQN and 1Tqn will retrieve the same structure. Notice that we keep mirrors of LigandExpo and the Protein Data Bank on our server, so that file retrieval is as speedy as possible. The date the database mirrors were last updated is always displayed at the bottom of the **submission page**.

One feature that makes **MotiveValidator** very efficient is that it allows the validation of multiple residues in multiple files in a single run. If you need to validate more than one residue at a time, simply request to retrieve several models from LigandExpo by specifying a list of 3-letter codes separated by commas. Alternatively, you may upload more files with model residues in a .zip archive. You may also wish to validate a residue in more than one biomolecule, in which case simply upload a .zip archive of all your input structures, or request to retrieve several structures from the Protein Data Bank by specifying a list of 4-letter PDB IDs separated by commas. Thousands of biomolecules can be scanned at a time.

Once you have specified your desired input for the model and for the biomolecules to be validated, click the Upload and Compute button at the bottom of the page. Check out section 6.2 to find out how to analyze your validation results.

6.1.2. Sugar Validation

The second type of validation that can be performed using **MotiveValidator** can be initiated via the **Sugar Validation** tab on the **submission page** (Figure 5). As in the case of the **Residue Validation** tab, the **Sugar Validation** tab provides a brief description of the validation process, and several options for input file submission.

Sugar validation is a special case of **residue validation** where the residues to be validated are carbohydrates. This special analysis mode allows **MotiveValidator** to perform fully automated validations. Specifically, **MotiveValidator** keeps a pre-established list of sugar residues and their models from LigandExpo, containing the residue name and structural information for all atoms in

each sugar residue. The consequence of this fact is that you need not provide the sugar models yourself, and need only specify the biomolecule file. **MotiveValidator** will automatically scan all residues in the biomolecule file, and extract all instances of all sugars based on residue names and residue identifiers. Each instance of each sugar will be considered a separate motif, receive a unique motif identifier and subsequently be validated against its corresponding sugar model. All sugar validations take place in a single run, making **sugar validation** very efficient.

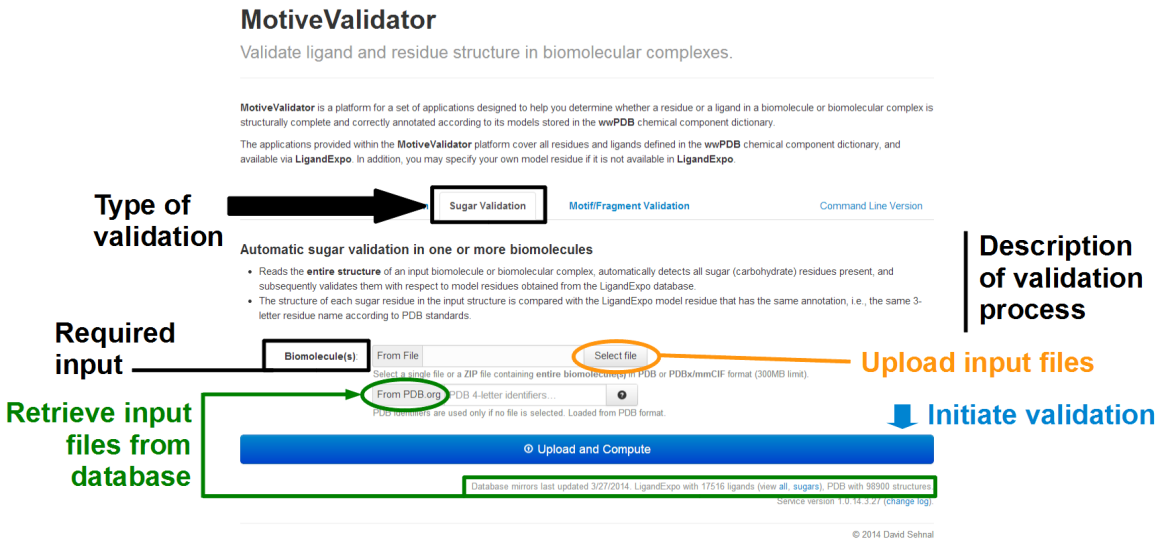

*Figure 5: The **Sugar Validation** tab allows to automatically validate all sugar residues in one or more biomolecules. **MotiveValidator** keeps a pre-established list of all sugar residues and their structural models from LigandExpo. Only the biomolecule input file(s) containing sugar residues to be validated must be specified. Input files can be retrieved from the mirror of Protein Data Bank maintained on our server, or can be uploaded from your computer in PDB or PDBx/mmCIF format. All sugar residues will be automatically validated in one run. Thousands of biomolecules can be processed in one run if you upload a .zip archive with biomolecule files, or retrieve a list of comma separated PDB IDs (4-letter identifiers). Once you have picked your input file(s), click the Upload and Compute button at the bottom of the page in order to initiate the validation.*

The biomolecule input files may be uploaded from your computer. Alternatively, the biomolecule may be retrieved from the Protein Data Bank by specifying its 4-letter identifier PDB ID. Capitalization is not important, so 1tqn, 1TQN and 1Tqn will retrieve the same structure. Notice that we keep mirrors of these databases on our server, so that file retrieval is as speedy as possible. The date the database mirrors were last updated is always displayed at the bottom of the **submission page**. The efficiency of **MotiveValidator** is further enhanced by the possibility to perform the validation of sugars in multiple biomolecules at a time, in which case simply upload a .zip archive of all your input structures, or retrieve several structures from the Protein Data Bank by specifying a list of 4-letter PDB IDs separated by commas. Thousands of biomolecules can be scanned at a time.

Once you have specified your desired input for the model and for the biomolecules to be validated, click the Upload and Compute button at the bottom of the page. Check out section 6.2 to find out how to analyze your validation results.

6.1.3. Motif/Fragment Validation

The third type of validation that can be performed using **MotiveValidator** can be initiated via the **Motif/Fragment Validation** tab on the **submission page** (Figure 6). The **Motif/Fragment Validation** tab provides a brief description of the validation process, and several options for input file submission.

Remember that the term *motif* refers here to a fragment of a biomolecule or ligand, made up of one or more residues or parts of residues. Therefore, any set of atoms that can be defined as a motif within the **MotiveValidator** environment can be validated using **Motif/Fragment Validation**. **MotiveValidator** will first read in the *model* file and collect the structural information for all its atoms. **MotiveValidator** will then read in the motif file and assign it a unique motif identifier. Then, the entire structure of the motif as a whole is validated against the model, regardless of the residue names and identifiers in the motif file.

The model may be retrieved from LigandExpo by specifying its 3-letter identifier according to the PDB residue naming convention. Capitalization is not important, so man, MAN and mAn will retrieve the same structure from LigandExpo. Notice that we keep a mirror of LigandExpo on our server, so that file retrieval is as speedy as possible. The date the database mirrors were last updated is always displayed at the bottom of the **submission page**. Alternatively, the model files may be uploaded from your computer in PDB or PDBx/mmCIF format, along with the motif input files. If you would like to extract your own fragments from biomolecule files and then analyze them as motifs using **MotiveValidator**, you may use our tool **MotifExtractor** (details in section 8.2)

Thousands of motifs can be validated in a single run if you upload a .zip archive with motif input files. Also, each motif can be validated against several reference models at a time. In order to do so, simply request to retrieve several models from LigandExpo by specifying a list of 3-letter codes separated by commas, or upload more model files in a .zip archive.

Once you have specified your desired input for the model and for the motifs to be validated, click the Upload and Compute button at the bottom of the page. Check out section 6.2 to find out how to analyze your validation results.

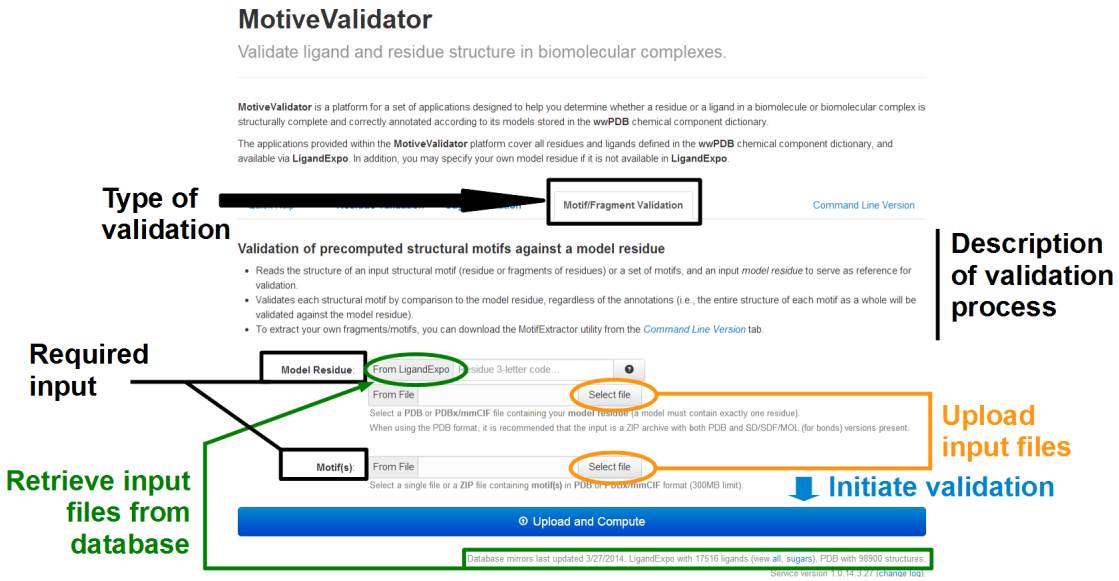

*Figure 6: The **Motif/Fragment Validation** tab allows to validate fragments of biomolecules or ligands that we refer to as motifs. A motif may contain one or more residues or fragments of residues. Two input files are necessary: the first file with the structure of the model, and the second file containing the motif to be validated. Unlike the other types of validation that **MotiveValidator** allows, **Motif/Fragment Validation** takes the entire structure of the input motif as a whole, regardless of residue names and identifiers, and compare it to the model residue. Models can be retrieved from the mirror of LigandExpo maintained on our server, or uploaded in PDB or PDBx/mmCIF format along with the motif input files. You can validate a motif against several model residues if you retrieve a list of comma separated residue names (3-letter codes) from LigandExpo, or upload a .zip archive with model files. Thousands of motifs can be processed in one run if you upload a .zip archive instead of a single motif file. Once you picked your input file(s), click the Upload and Compute button at the bottom of the page to initiate the validation.*

6.2. Analysis of results

As soon as you initiate a validation by providing the necessary input files and clicking the Upload and Compute button at the bottom of any functional tabs on the **MotiveValidator submission page**, you will be redirected to the **results page**. If you are working with large volumes of data, **MotiveValidator** will display the progress of the calculation. However, even in such cases, where thousands of structures are processed, the total duration of any validation is on the order of seconds.

Regardless of the type of validation you chose on the **submission page**, the presentation of results in **MotiveValidator** is unified. The organization of the **results page** is in principle similar to that of the **submission page**. Namely, general options and information are given first, whereas specific reports of the validation results are available via the functional tabs (Figure 1 B).

You may access the **results page** of each specific calculation via its own URL (web address), given at the top of the **results page**. Your results are stored on our server for a month. The exact date till which you may return and view or download your results from our server is also specified at the top of the **results page**.

MotiveValidator provides a straightforward report of the validation results, including a summary and detailed information in both tabular and graphical form, along with a 3D structure visualizer for closer inspection of the problematic structures. These reports are accessible via the functional tabs on the **results page**, which we shall describe in detail in sections 6.2.1-6.2.4.

Inspecting the tabular and graphical validation reports accessible on the **results page** is the most comfortable and effective way to evaluate your results. Additionally, you may use the blue Download buttons in order to download any part of the data involved in the validation, and perform additional analyses on your own. The structures of the model residue, input motifs and validated motifs are also available for download in PDB format. See section 7 for a description of how the downloadable validation output is organized.

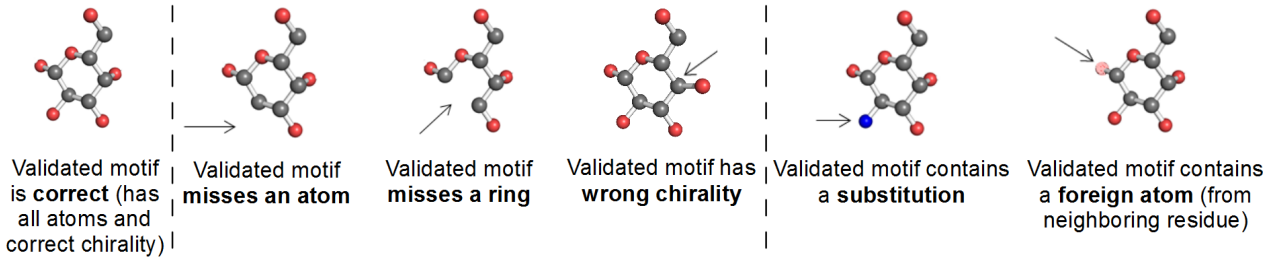

*Figure 7: **MotiveValidator** reports correct structures, as well as all potential issues found during validation, namely structures that are wrong either because they are incomplete (missing atoms or rings), or because the chirality of some atoms is incorrect. Additionally, **MotiveValidator** reports substitutions of native atoms due to chemical modifications of the residue, or atoms of neighboring residues found to replace native atoms at linkage sites, as well as unexpected PDB atom identifiers (not shown here).*

6.2.1. Summary analysis

The first view of the results is available in the **Summary** tab of the **results page** (Figure 8). First, a description of the validation process is given. Then, for each validated residue, you are provided with an overview of potential issues encountered. **MotiveValidator** reports issue related to incomplete structure or incorrect chirality, as well as other useful notes.

If more than one residue were validated in one run, a list of these residues will be available

right under the description. In order to examine the validation summary for each residue, you will need to either click on that specific residue in the list, or just scroll down the page till you reach it. Each validated residue is identified by its 3-letter code, as well as its chemical formula and common name. Each input structure with *validated residues* may contain one or more instances of each *validated residue*, which are denoted as motifs (see definitions in section 4). Each motif was validated independently, and thus statistics of these results are given, as absolute numbers and percentages. **MotiveValidator** reports several kinds of issues found in validated motifs (Figure 7).

The table with issues is organized into two main sections, referring to incomplete (Missing Atoms or Rings) and complete structures (With All Atoms and Rings) respectively. The formal distinction between *ring* atoms and non-ring atoms (simply denoted as *atoms*) is meant to allow a quick localization of potential issues in residues containing rings, especially where atom identifiers are not useful. Chirality is evaluated only for the complete structures, since the absence of some atoms would make it impossible to check the chirality of some of the remaining atoms. Further, the problematic atoms are highlighted, in order to better localize the problems in the structures.

Aside from issues related to incomplete structure and wrong chirality, **MotiveValidator** also reports any unusual annotation for each atom. *Different Naming* refers to the cases where atoms in the validated motif have a different PDB atom identifier compared to the model. This happens if different atom naming conventions were used when creating the input files, but it can also be an indicator that the mapping of the validated motif and model residue might be flawed. *Substitutions* refers to the cases where, mainly due to chemical modifications of the native residue (e.g., glycosylation of amino acid residues), at least one atom in the validated motif is replaced with an atom of a different chemical symbol. *Foreign* (or *foreign atoms*) refers to the cases where, mainly due to binding of multiple residues into chains, the validated motif contains at least one atom which formally belongs to a neighboring residue. This generally happens when residues are extracted from larger structures based on bonding information, but may also be an indication that the validated residue is a derivative of the model.

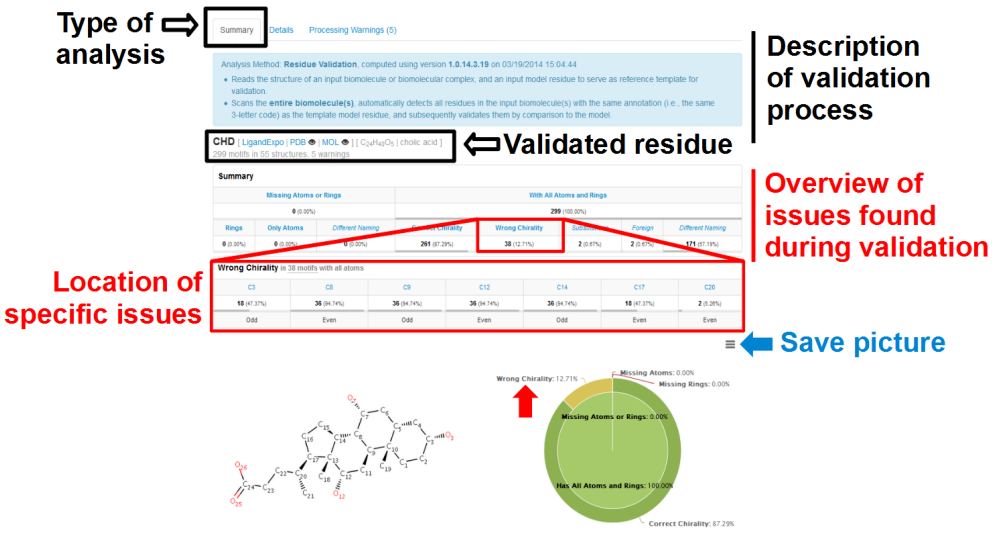

*Figure 8: The **Summary** tab first provides a description of the validation process, and then a summary of the results in tabular and graphical form. Each input structure with validated residues may contain one or more instances of each validated residue, denoted as motifs. If you validated more than one residue in a single **MotiveValidator** run, the list of residues appears right under the description. In the list, click on the residue of interest to get the summary of its validation, or simply scroll down the page till you reach it. For each residue, **MotiveValidator** provides an overview of issues related to incomplete structure or incorrect chirality, as well as other useful notes (different atom naming, substitutions, foreign atoms). Then, problematic atoms are highlighted, to better localize problems in the structures.*

Additionally, **MotiveValidator** reports various issues encountered when processing the input files. *Processing warnings* are issues that may cause incorrect validation, such as atoms that are too close in the 3D space, or unusual bond lengths given by the CONECT records. *Processing errors* are major issues preventing the finalization of the validation, such as parts of the residue which are completely disconnected from the rest of the structure, probably due to missing atoms at multiple locations throughout the structure.

Last, a 2D representation of the model residue, and a pie chart with the validation results are provided for visual representation purposes. You can download them via the small icon at the top right corner of the chart, and later use them in your presentations.

6.2.2. Detailed analysis

Whereas the **Summary** tab provides statistics of the issues over all validated motifs for each validated residue, the **Details** tab allows you to inspect the issues in select groups of motifs, and further in each individual motif (Figure 9). Note that you may also access the details of any particular group of motifs also by clicking on a specific issue in any **Summary** tab table.

The **Details** tab is organized into a table where each row contains information regarding a single validated motif. The content of the table (i.e., which motifs are included, and what information is displayed) is dictated by the values of three selection fields at the top of the table. Click on the first field, and select the *validated residue* (3-letter code) from the drop down menu. Only the motifs that were matched to that residue name will be displayed in the table. Click on the second field and select the *type of issue* (e.g., wrong chirality) from the drop down menu. Only the motifs which exhibit that type of issue will be displayed in the table. The number of motifs that fit each selection is given in brackets. If you want to make your selection even more specific, use the selection filed *Id filter.*

Which table columns are filled depend mostly on the type of issue selected in the filter. The most important columns are *Id*, *Issues/Warnings*, *Missing atoms/rings*, *Atoms*, *Processing warnings*. The other columns give additional information, not essential for the validation results, but many times useful in identifying the source of the error in the structure. The column *Id* refers to a unique

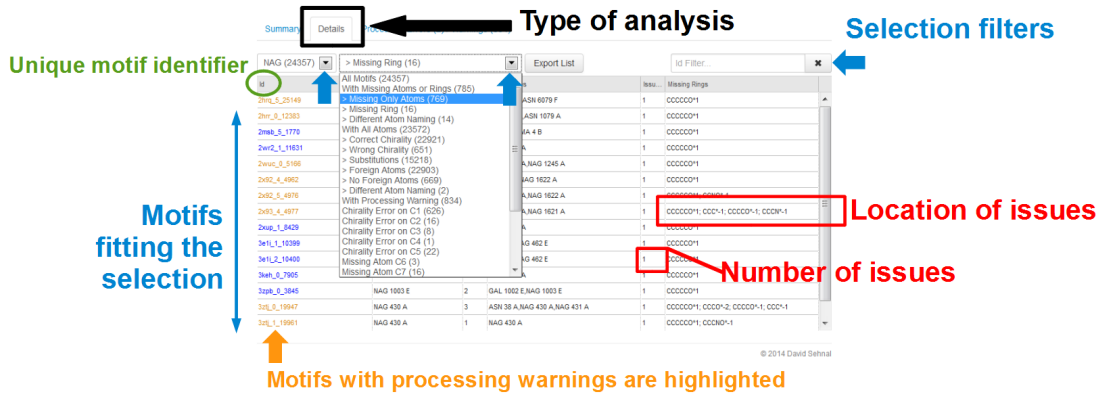

*Figure 9: The **Details** tab allows to inspect the issues in select groups of motifs, and further in each individual motif. A motif is a single instance of a query residue identified in the input file, and is assigned a unique motif identifier. Each row in the table contains information regarding a single motif. In order to select which motifs will be displayed in the table, you may specify the residue name and type of issue in the drop down menus of the selection fields at the top of the table. Further, you may refine your selection using the Id Filter. The number and location of any issues identified in the structure of each motif is given, along with additional notes regarding which residues appear in the vicinity of the validated residue. Motifs where there were processing warnings are highlighted in orange font.*

identifier that **MotiveValidator** assigns to each motif. The Id is useful not just for handling large numbers of motifs, but also to keep a transparent trace of the motif's origin, as it contains the PDB ID of the original input structure, as well as the serial index of the first atom in the motif, as it appears in the original input file. The column *Issues/Warnings* reports the number of issues or warnings found for each particular motif. The column *Missing atoms/rings* explains which atoms are missing in each validated motif, whereas *Atoms* shows the position of incorrect chirality. Missing atoms are listed by their atom identifier in the model, whereas atoms with wrong chirality are listed by their identifier in the validated motif. Clicking on a column header allows to order the motifs according to the property specified in the header.

6.2.3. 3D visualization

The **3D viewer** implemented in **MotiveValidator** offers one step further in the analysis of each individual validated motif accessible via the **Details** tab (Figure 10). Simply click on the Id of a motif of interest, in order to open the 3D viewer, where you can inspect the structural inaccuracies more closely.

Here you will be able to view and manipulate with the 3D representations of the validated motif and model residue, to help you better assess the position and relevance of the structural issues found during validation. Additionally, a 2D representation of the model is provided for clarity, which is especially helpful for larger motifs. Basic information about the validated motif is also given, along with a complete report of the validation results, where all the potential issues are listed.

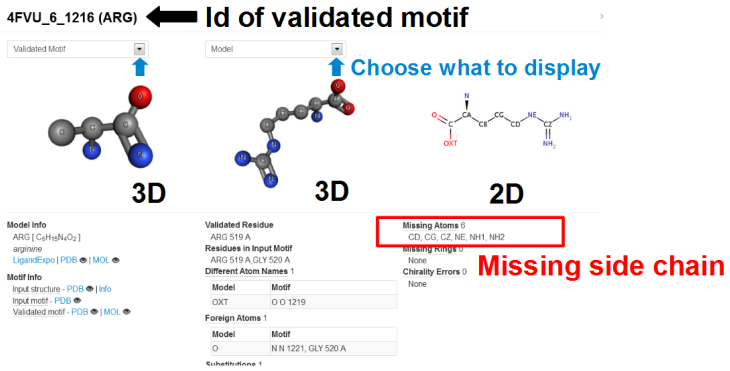

*Figure 10: You can open the **3D visualizer** by clicking on a motif's Id in the **Details** tab. You may manipulate (rotate in the 3D space) with the 3D representations of the validated motif and model residue, so as to better assess the position and relevance of the structural issues found during validation. Additionally, a 2D model is provided for clarity, which is especially helpful when working with larger motifs. Basic information about the residue in question is also given, along with a complete report of the validation results. In this particular example, motif 4FVU_25_1216, which was matched as an arginine residue, is missing most of the side chain atoms. It may be that the side chain was very flexible, and thus not well defined in the electron density of the crystal structure of the parent biomolecule. The flexibility of the side chain may be related to a biologically relevant conformational switch of the arginine at this particular location.*

6.2.4. Processing warnings and processing errors

Sometimes the input files contain information that is ambiguous, conflicting or which deviates strongly from the expected reference. **MotiveValidator** reports such occurrences as processing warnings or processing errors, depending on the severity of the deviations (Figure 11). The selection field helps filter the warnings and errors. Simply click on the drop down menu and select the category of warnings or errors that you would like to explore.

Cases where two atoms are too close to each other in the 3D space, or where the bond lengths were found to be unusual are reported as *processing warnings*. It is always good to check and make sure that negative validation results (e.g., missing atoms) are not in fact caused by ignoring some atoms in an ill-formed structure. Any major errors in the input file, such as atoms that are completely disconnected from the rest of the structure, are reported as *processing errors*, and these structures are not processed at all.

It is important to note the difference between processing warnings and processing errors. A warning may simply lead to ignoring a faulty atom, but the motif will be validated. On the other hand, a processing error prevents entire motifs from being validated, so you will not find these motifs in the statistics available on the **Summary** tab, or in the tables available on the **Details** tab. However, because **MotiveValidator** automatically extracts all instances of a validated residue and assigns them a unique and informative motif Id, you will be able to easily find the motif in its original PDB input file. You may be able to fix the reported errors, or at least investigate these deviations using different means.

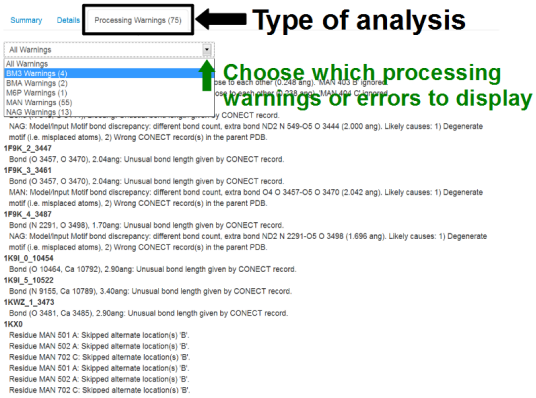

*Figure 11: **MotiveValidator** first reads in all input files. If it encounters information that is ambiguous, conflicting or which deviates strongly from the expected reference, **MotiveValidator** will report these occurrences as processing warnings or processing errors. Processing warnings are typically atoms that lie too close to other atoms in the 3D space, or where bond information in CONECT records is unusual. Processing errors are mainly caused by major flaws in the input file, such as atoms that are completely disconnected from the rest of the structure. The problematic structures are not processed at all, and the validation statistics will not include them. Click on an item in the drop down menu to select the category of processing warnings or errors that you would like to explore.*

7. Additional output

As previously mentioned, the most comfortable and efficient way to analyze the validation results is to directly use the statistics and graphics provided on the web page, as this information is very easy to sort, filter and synthesize into a presentation. However, should you want to further process the validation output data yourself, you may obtain it via the blue Download Result button at the top of the **results page**. All structures (model residue, input motifs and validated motifs, but not motifs with processing errors) are given in PDB and SDF/MDL/MOL format. All data that appears in the **Summary** and **Details** tabs in tabular form is given in .csv format. Additionally, the atom pairing between the model residue and the validated motif is provided.

8. Additional resources

Along with all the functionality available via the web interface, we provide additional tools for your convenience. These tools are command line based, and thus available in the **Command line version** tab (Figure 12) of the **MotiveValidatior submission page**.

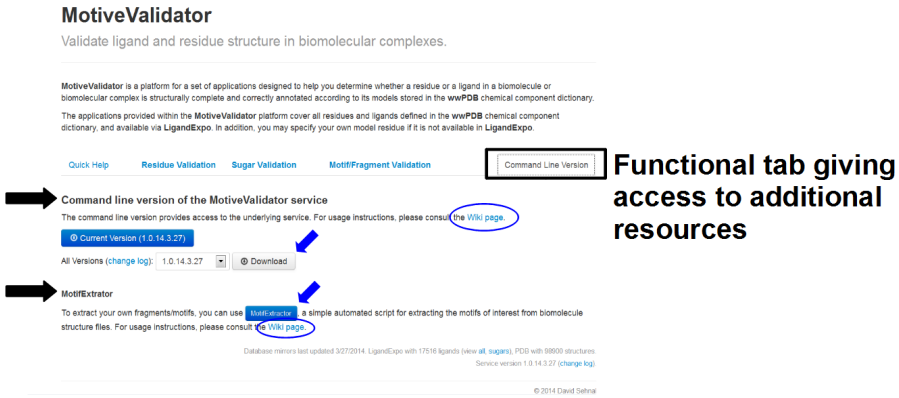

*Figure 12: The most straightforward, comfortable and effective way to take advantage of **MotiveValidator** is to run it from our server. However, if you would like to run **MotiveValidator** from your system, the **Command line version** tab provides access to the command line version of **MotiveValidator**. You may download the latest version of **MotiveValidator** using the blue download button, whereas older versions can be chosen from the drop down list and downloaded via the gray download button. Furthermore, here you may download **MotifExtractor**, a simple automated script for extracting the motifs of interest from biomolecule structure files. Instructions on how to run **MotiveValidator** and **MotfiExtractor** on your system are available on our online Wiki pages page.*

8.1. Command line version of MotiveValidator

In this manual we described in detail how to use **MotiveValidator** directly from our server, since that is the most straightforward, comfortable and effective way to take advantage of **MotiveValidator**. Nonetheless, should you wish to run the validations on your system instead of on our server, you may do so. All you need to do is click on the **Command line version** tab, and download the command line version of **MotiveValidator** (Figure 12). Instructions on how to run **MotiveValidator** on your system can be found on our web page at http://webchem.ncbr.muni.cz/Wiki/MotiveValidator_Command_Line_Help.

8.2. MotifExtractor

For **Residue Validation** and **Sugar Validation**, input files may contain any additional structural elements in addition to the validated residue. **MotiveValidator** will identify and extract the validated residues based on the annotations in the model file. Thus you need not process the biomolecule input files before running **MotiveValidator** in either of these two validation modes.

However, in the case of **Motif/Fragment Validation**, the entire structure of each input motif will be compared to the model without checking annotations (residue name and identifier) in the model file. Therefore, you will need to provide the input motifs yourself. To save you time in this task, we provide **MotifExtractor**, a simple automated script for extracting the motifs of interest from biomolecule structure files. Instructions on how to run **MotifExtractor** on your system can be found on our web page at http://webchem.ncbr.muni.cz/Wiki/MotifExtractor_Help.