Validator^{DB}

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 100000 structures of biomolecules and their complexes. Such large bodies of data, especially accumulated over a short period of time using high throughput techniques, will inherently be plagued by various 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. Ligand validation, as well as the validation of uncommon residues, are very challenging tasks, because of the high diversity and nontriviality of their structure, and the general lack of information about correct structures. Therefore, software tools focused on ligand validation were developed relatively recently[1,2], and the topic is still under active development[3]. These tools are able to validate one or more structures (even thousands of structures), but they are not able to provide the broad scientific community with a more complex image of the quality of structures in dedicated and well established structural databases. For example, a general overview and corresponding statistical evaluation of validation results for residues and ligands in the entire PDB is not yet available, despite the exponential growth of the PDB and the development of structural validation tools in recent years.

We had recently developed **MotiveValidator**[4], an interactive platform for the speedy validation of ligands, residues and fragments using a novel, straightforward approach based on the validation of residue annotation. **MotiveValidator** employs advanced algorithms for the detection and comparison of structural motifs[5,6], along with tools for chirality verification[7] and interactive visualization of 3D structures[8]. Using **MotiveValidator**, we further created **ValidatorDB**, a comprehensive resource of validation results for residues and ligands in the Protein Data Bank. Along with validation results for individual residues and ligands, **ValidatorDB** also provides a summary and statistical evaluation of the validation results at various levels of detail within the PDB. Thus, **ValidatorDB** offers a comprehensive overview of the quality of the ligand structures in the entire PDB.

1 Lütteke T, von der Lieth C-W. *BMC Bioinformatics* (2004) 5, 69.
2 Kleywegt GJ, Harris MR. *Acta Crystallographica D. Biological Crystallography* (2007) 63: 935–8.
3 ftp://ftp.ebi.ac.uk/pub/databases/pdb/validation_reports/
4 Svobodová Vařeková R, Jaiswal D, Sehnal D, Ionescu CM, Geidl S, Pravda L, Horský V, Wimmerová M, Koča J, *Nucleic Acids Research* (2014) doi:10.1093/nar/gku426
5 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.
6 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.
7 O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR, *Journal of Chemoinformatics* 3, (2011): 33.
8 http://www.chemdoodle.com

2. How to use this manual

In the following sections we offer an extensive tour through the **Validator^{DB}** features. The elements of the user interface are described as we go along, both textually and visually. In this manual we employ use cases in order to better illustrate efficient ways to extract useful information from **Validator^{DB}**.

Note that the **Validator^{DB}** interface provides lots of *tool tips* to aid you along the way. Moreover, if you prefer a fast walk-through rather than going through all the explanations in the present manual, please view the *tutorial* available on the **Validator^{DB}** web pages.

Enjoy exploring **Validator^{DB}**!

3. Availability and technical details

3.1. Where to find Validator^{DB}

Validator^{DB} is freely available via the internet since May 2014 at http://ncbr.muni.cz/MotiveValidatorDB. There is no login requirement for accessing **Validator^{DB}**.

3.2. What you need in order to access Validator^{DB}

Validator^{DB} is basically a database, or rather a collection of validation results for ligands and residues in Protein Data Bank. The database is maintained on the ncbr.chemi.muni.cz server at the National Centre for Biomolecular Research within Masaryk University, Czech Republic, and updated weekly. All you need in order to access **Validator^{DB}** is an internet browser that is up to date and has JavaScript enabled, and a working internet connection. 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/ValidatorDB, you will reach the **Validator^{DB} synopsis page**, which contains a brief, general description of **Validator^{DB}**, along with 3 tabs (Figure 1A). The different tabs on the **Validator^{DB} synopsis page** provide access to overviews and statistical evaluation of validation results for the entire PDB, for each residue across all PDB IDs containing that residue, and for all analyzed residues in each PDB ID, in graphical or tabular form. Click on each tab to discover what type of overview can be accessed.

Further, the **Validator^{DB} specifics page** (Figure 1B), which is accessible by looking up specific residues or PDB IDs in the **synopsis page**, allows to view the results for selected residues in more detail. The **specifics page** is also organized into tabs that allow different levels of analysis of the results.

Last but not least, remember to check the *tool tips* by hovering the mouse cursor over any graphical or textual element in the **Validator^{DB}** interface.

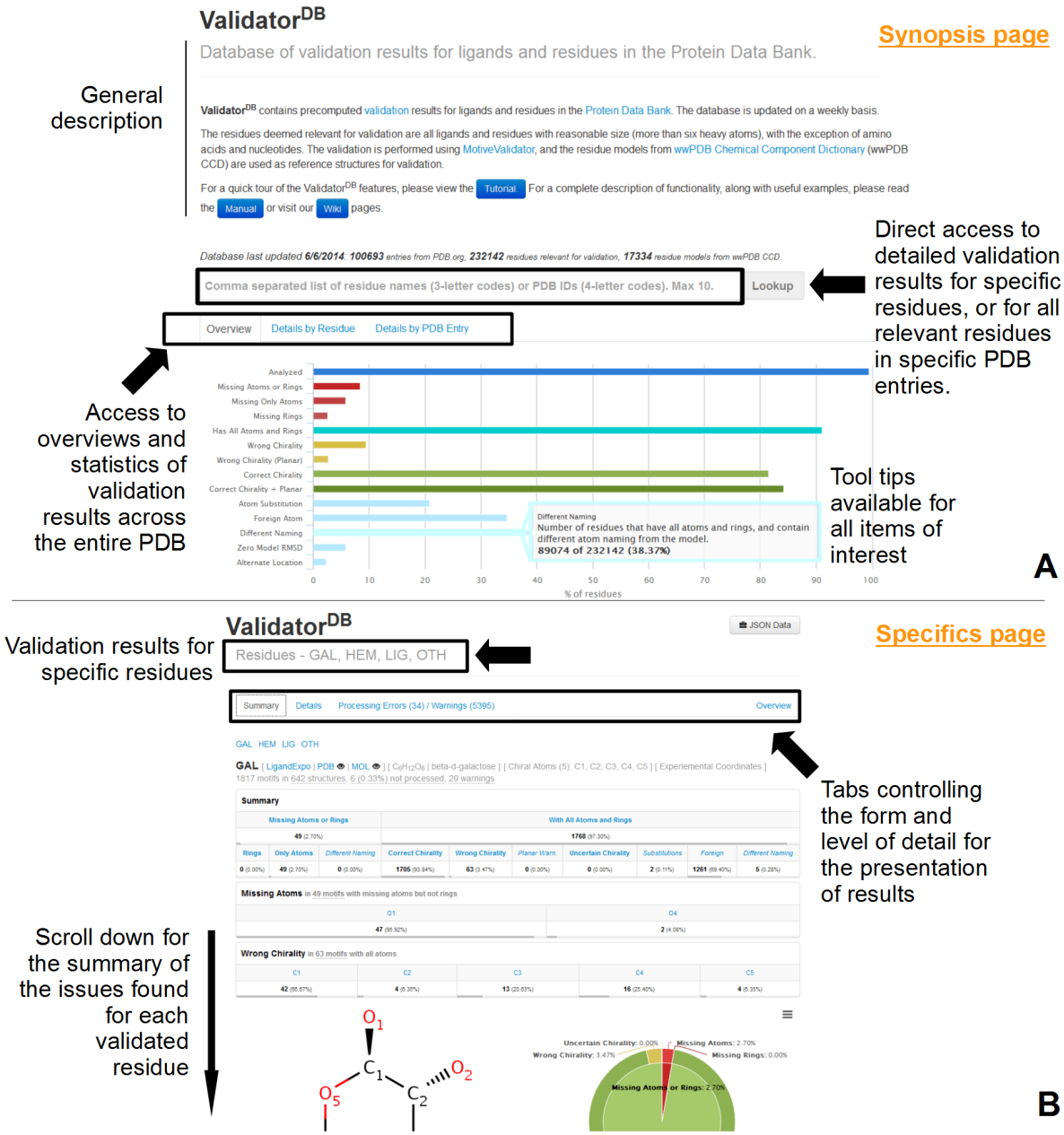

Figure 1: ***ValidatorDB*** *web pages*

A) The ***synopsis page*** *contains a general description, and is organized in 3 tabs, namely Overview, Details by Residue, and Details by PDB Entry. These tabs provide access to graphical and tabular overviews and statistical evaluation of validation results for the entire PDB, for each residue across all PDB IDs containing that residue, and for all analyzed residues in each PDB ID, respectively. Click on each tab to discover what type of overview can be accessed.*

B) The ***specifics page*** *contains validation results for specific residues of interest, which can be viewed in a few different tabs, which control the form and level of detail that the results will be presented in. The tabs Summary and Details, allow different levels of analysis, while the tab Overview presents a rough overview of the results in a graph similar to the Overview tab on the* ***Synopsis page****. An additional tab with Processing Errors/Warnings may appear if issues are detected in the input files.*

4. Basic terms and principles

Before moving on to more extensive descriptions of features, it is important to clearly establish the meaning of a few key terms and principles within the **Validator^{DB}** environment.

4.1. Residue

The term *residue* is used 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 **Validator^{DB}** 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 **Validator^{DB}** term *residue* (section 4.1) thus fully covers ligands, in addition to typical components like amino acids and nucleotides. Within the **Validator^{DB}** environment, any statements pertaining to *ligands* hold also for *residues*.

4.3. Residues and ligands relevant for validation

As mentioned in section 1, well studied residues like amino acids and nucleotides are routinely validated upon submission of new structures to the PDB. Furthermore, reports of the quality of their structure are already accessible[9]. The challenge addressed by **Validator^{DB}** lies in providing access to validation results for residues other than the well studied amino acids and nucleotides. This generally includes ligands and uncommon residues (e.g., substituted amino acids), which exhibit high diversity and nontriviality in their structure, and for which there is generally much less information regarding correct structures.

Thus, within the **Validator^{DB}** environment, we further refine the meaning of the terms *residue* and *ligand* to refer to *residues and ligands relevant for validation*. Specifically, these are all ligands and residues with reasonable size (more than six heavy atoms), with the exception of amino acids and nucleotides. All other features of the terms *residue* and *ligand* described in sections 4.1 and 4.2 are maintained. Henceforth, all references to *residues* and *ligands* in this manual will have the meaning of *residues and ligands relevant for validation*. Similarly, all references to *residues* and *ligands* in the **Validator^{DB}** web pages (including Wiki and tutorial) have the meaning of *residues and ligands relevant for validation*. The PDB currently holds over 17000 residues and ligands relevant for validation.

9 ftp://ftp.ebi.ac.uk/pub/databases/pdb/validation_reports/

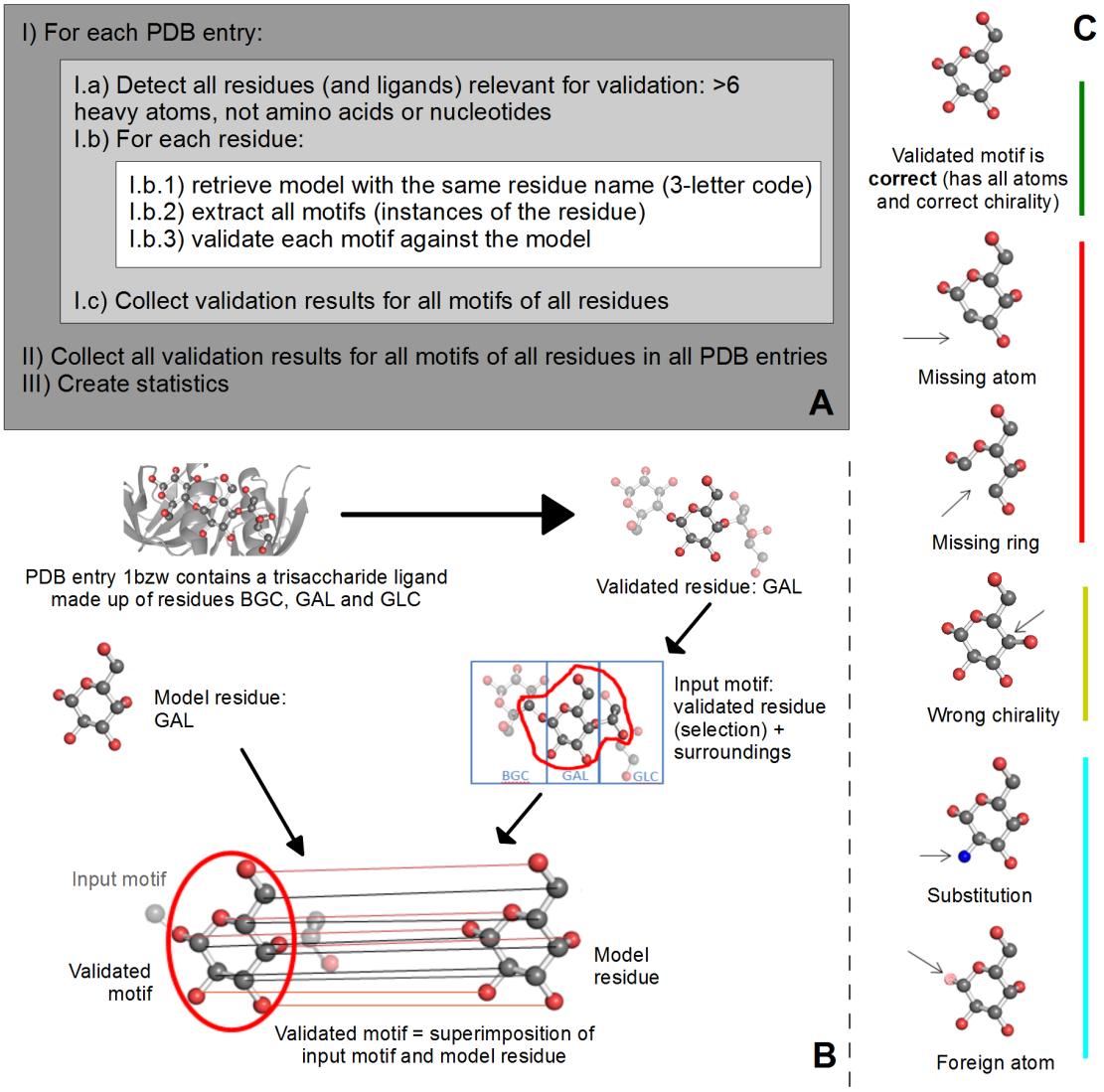

Figure 2: Validation within ***Validator**DB***

A) Overview of validation process. For each PDB entry, the relevant residues are detected based on their name (3-letter code) and number of atoms. Then, for each validated residue, the corresponding model is retrieved from wwPDB CCD, and each motif of this residue is validated against the model. ***Validator**DB** *is then built as the collection of validation results for all motifs of all residues in all PDB entries.*

B) Example of validation for a galactose (GAL) motif in the PDB entry 1bzw. The validated residue GAL is extracted from PDB entry 1bzw in the form of an input motif, which contains 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). By superimposing the input motif and model residue, the validated motif results as the subset of atoms in the input motif which correspond to atoms in the model residue. Evaluation of the validated motif (which atoms are present and where, compared to the model residue) produces the validation results.

C) Typical validation results for a GAL motif. If the validated motif contains all atoms at their expected positions, it is marked as correct (green). If any atoms are missing, the validated motif is marked as incomplete (red). If some atoms have wrong chirality, the validated motif is marked accordingly (dark yellow). If any unusual features are detected, the validated motif is marked with a warning (cyan).

4.4. Motif

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 **Validator^{DB}** 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, **Validator^{DB}** 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).

We can generally say that, within the **Validator^{DB}** environment, all *residues* can be thought of as *motifs*. Therefore, different *instances of the same residue* (such as copies of the same ligand in different monomers) can be considered and processed as different motifs, making their identification straightforward and unambiguous. The PDB currently holds over 230000 motifs of over 17000 different residues and ligands relevant for validation (section 4.3).

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. Within the **Validator^{DB}** environment, this structure was used as reference template in the validation process, whereby each instance of each residue was compared against the model residue with the same name (3-letter code). The origin of the *models* is be the wwPDB chemical component dictionary[10] (wwPDB CCD).

4.6. Validation

As stated in section 3, the validation results stored in **Validator^{DB}** are updated every week. Within the **Validator^{DB}** environment, the term *validation* refers to the process of determining whether a residue or ligand is structurally complete and correctly annotated. This means checking if the topology and chirality of each motif of a validated residue (section 4.4) correspond to those of the model residue (section 4.5) with the same name as the validated residue.

The validation of residues and ligands in the entire PDB takes place in a few distinct steps. First, for each PDB entry, the residues which are relevant for validation are detected based on their name (3-letter code) and number of atoms (more than 6 heavy atoms). Amino acid residues and nucleotides are excluded based on their residue name. Then, for each validated residue, the corresponding model (same 3-letter code as the validated residue) is retrieved from wwPDB CCD, and each motif of this residue is validated against the model. **Validator^{DB}** is then built as the collection of validation results for all motifs of all residues in all PDB entries (Figure 2A).

The validation of each motif against the model residue can be illustrated on a galactose (GAL) motif from the PDB entry 1bzw (Figure 2B). The *validated residue* GAL is extracted form PDB entry 1bzw in the form of an *input motif*, which contains 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). Then, by superimposing the *input motif* and *model residue*, the *validated motif* is obtained as the subset of atoms in the input motif which have a correspondent atom in the model residue. Comparing each atom and bond in the validated motif to those in the model residue produces the validation results.

10 http://www.wwpdb.org/ccd.html

4.7. Validation results

For each validated motif, **Validator^{DB}** contains several types of results. Since the evaluation of the validated motif relies on comparing all atoms and bonds in the validated motif to those in the model residue, the first results that can be encountered are *errors*. Namely:

- *Missing atoms*: an atom in the model residue has no corresponding atom in the validated motif
- *Missing rings*: at least one missing atom originates from cycles (rings)
- *Wrong chirality*: an atom from the validated motif has different chirality than the corresponding atom from the model residue
- *Wrong chirality (planar)*: the chirality error was found on a planar chiral center. Because of their spacial distribution, planar chiral centers are very sensitive even to small perturbations in the position of the substituents. Therefore, some of the errors reported here might not be significant.
- *Uncertain chirality*: the presence of unusual bonds may cause an improper evaluation of chirality

Chirality is only evaluated for those motifs which are complete. This is because the absence of some atoms can prevent the proper evaluation of chirality on the chiral centers present in the validated motif. Therefore, note that all motifs which are counted in the *Wrong chirality* category are in fact complete. At the same time, the motifs with no missing atoms and no chirality error are actually counted in a category called *Correct chirality*.

Suspicious discrepancies between the atoms and inter-atomic bonds in the validated motif and in the model residue are reported as *warnings*. Namely:

- *Substitution*: an atom from the validated motif is of a different chemical element than the corresponding atom in the model residue (e.g., O mapped to N). This happens often at linkage sites.
- *Different naming*: an atom from the validated motif has a different PDB atom name than the corresponding atom from the model residue (e.g., the C1 atom mapped to the C7 atom). This happens often when the original PDB files were produced by different software.
- *Foreign atom*: an atom from the model residue was mapped to an atom from outside the validated residue (i.e., from its surroundings).
- *Alternate locations*: in the original PDB file, the validated residue contains atoms which were given in alternate locations (i.e., most probably different rotamers). Only the first rotamer was considered during validation.
- *Zero model RMSD*: the superimposition between the model residue and the validated motif has a root mean square deviation of zero, i.e., the validated motif is identical to the model residue used as reference.

Disabling discrepancies between the atoms and inter-atomic bonds in the validated motif and in the model residue are reported as *processing errors*, and such motifs are not validated.

Typical validation results that can be found in **Validator^{DB}** are illustrated on the galactose motif mentioned in section 4.6 (Figure 2C). As a general rule, in the **Validator^{DB}** interface, errors are marked in red (missing atoms) or dark yellow (wrong chirality), correct structures in green, and warnings in cyan.

5. Database contents

Validator^{DB} contains precomputed validation results for ligands and residues in the Protein Data Bank. The database is updated on a weekly basis. The validation is performed using **Validator**, and the residue models from wwPDB CCD are used as reference templates for validation. All residues of significant size (a minimum of 6 heavy atoms) have been included in **Validator^{DB}**, with the exception of amino acids and nucleotides, which are checked thoroughly upon submission of the structure to the PDB, and thus do not require additional validation.

The validation results available in **Validator^{DB}** inform whether each motif (occurrence, instance) of a ligand or residue in the PDB exhibits the appropriate topology and stereochemistry expected from its annotation (3-letter code), or how it differs from this annotation. Additionally, all issues related to incorrect or suspicious topology and stereochemistry are explicitly described in a comprehensive and intuitive manner (e.g., location of missing atoms or chirality inversions).

Validator^{DB} is organized on two main levels, namely PDB-wide results (**synopsis page**), and results restricted to specific residues of interest (**specifics page**). The two levels present the same type of validation results (as described in section 4.7), although the available features differ to some extent (e.g., the **specifics page** allows 3D visualization of motifs). We shall describe each level of the database in detail below.

5.1. Synopsis page

The **Validator^{DB} synopsis page** (Figures 1A,3) contains a brief description of **Validator^{DB}**, along with information about the last database update (date and number of structures that have been processed during the validation). Specifically, in May 2014, over 100000 PDB entries had been processed, containing over 230000 motifs of 17000 residues relevant for validation.

Additionally, the **synopsis page** allows to access the validation results for specific residues of interest via the *LookUp bar* (Figure 1A). Simply type a comma separated list of residue names (3-letter codes) into the *LookUp bar*, and you will be redirected to the **specifics page** containing validation results for the residues you requested. If you specify a list of PDB IDs (4-letter codes) instead, then the corresponding **specifics page** will contain validation results for all relevant residues and ligands in the PDB entries you specified. See section 5.2 for a description of the contents of the **specifics page**, and how to interpret these contents.

The **Validator^{DB} synopsis page** further provides access to various data sets of PDB-wide validations via 3 different tabs, namely *Overview*, *Details by Residue*, and *Details by PDB entry*. A full description of each of these tab is given below (sections 5.1.1-5.1.3).

5.1.1. Overview

The *Overview* tab of the **synopsis page** provides a very general statistical evaluation of results across the entire PDB in graphical form (Figures 1A,3A). The elements of the graph represent percentages of the total number of motifs (over 200000) of residues relevant for validation. A graphic element will be displayed in the *Overview* graph only if it represents at least 0.5% of the total number of motifs. Each element of the graph is described in a tool tip, but note that here the term *residue* actually refers to *occurrence of residue* (motif).

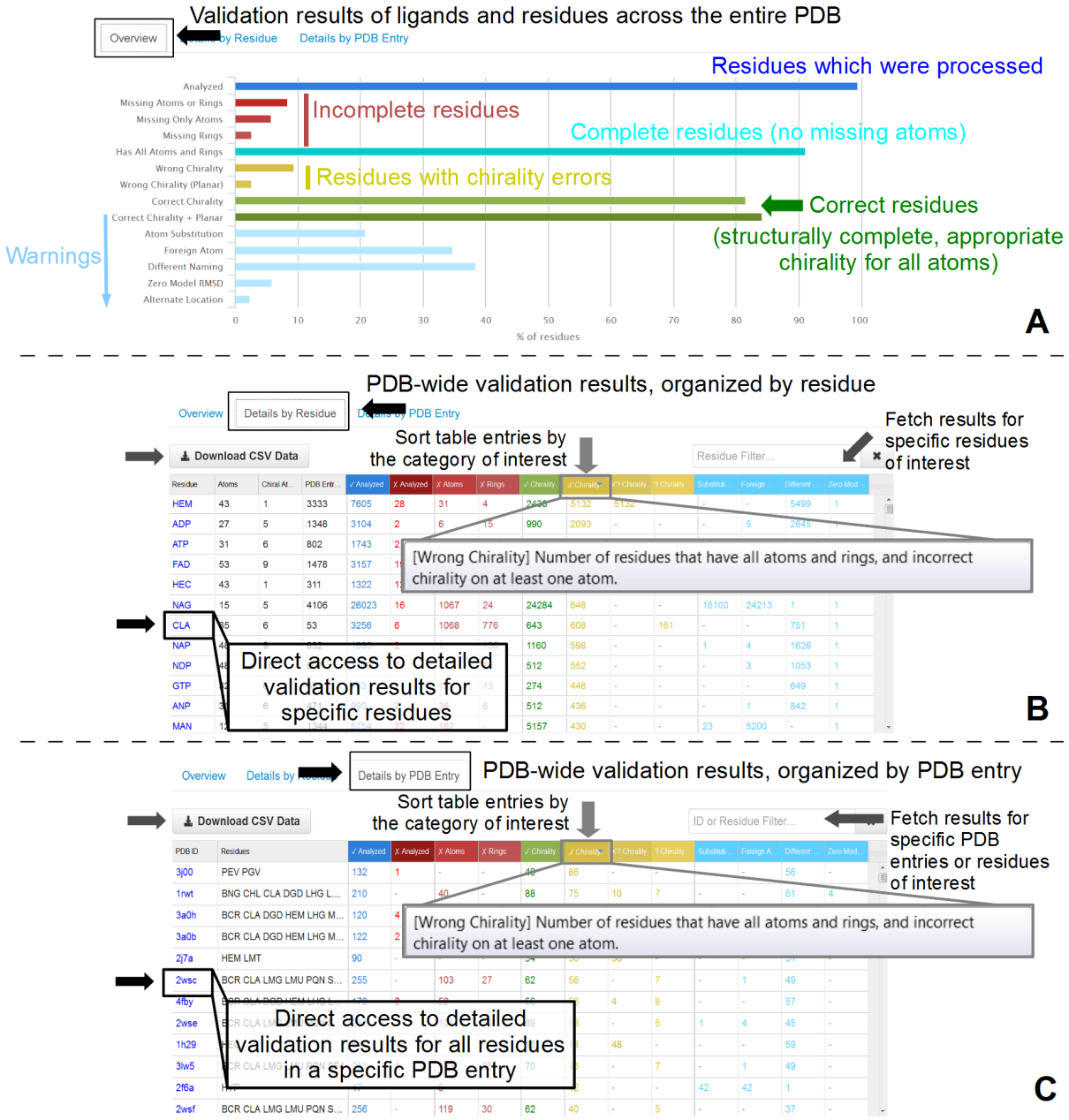

*Figure 3: The **Validator^{DB} synopsys page**.*

A) The Overview tab (see also Figure 1A): graph with the general statistical evaluation of results across the entire PDB. Each graphical element is described in a tool tip, and can be looked up in this manual in sections 4.7 and 5.1.1. Note that here the term residue actually refers to occurrence of residue (motif). Graphical elements can be assigned to roughly 6 categories: general information, incomplete residues, chirality issues, correct residues, warnings, etc.).

*B) The Details by Residue tab: interactive table summarizing the results for each validated residue across the entire PDB. Validation results (section 4.7) keep the color coding convention, and all results can be downloaded in .csv format. A filter allows to display results only for selected residues. Click on any residue name to access the **Validator^{DB} specifics page** with detailed validation results for that residue.*

*C) The Details by PDB Entry tab: interactive table summarizing the results for all validated residues in each PDB entry.Validation results (section 4.7) keep the color coding convention, and all results can be downloaded in .csv format. A filter allows to display results only for selected residues or selected PDB entries. Click on any PDB ID to access the **Validator^{DB} specifics page** with detailed validation results for all residues in that PDB entry.*

The elements of the graph can be assigned to roughly 6 categories, depending on which kind of information they contain (e.g., incomplete residue, chirality issues, warnings, etc.). The categories are marked by different colors (Figure 3A). Most of the graph elements have been explained in section 4.7 of this manual. The additional elements are *Analyzed*, which refers to the total number of motifs that could be processed, *Missing Atoms or Rings*, which is the sum of *Missing (Only) Atoms* and *Missing Rings*, and *Has All Atoms and Rings*, which is the total number of complete residues.

5.1.2. Details by Residue

The *Details by Residue* tab (Figure 3B) contains an interactive table summarizing the results for each residue validated across the entire PDB. Each row corresponds to one residue, identified by its residue name (3-letter code). The information in the table is organized according to the validation results as presented in section 4.7 of this manual. The color coding for the table header and the font inside the table is the same as in the categories defined in the *Overview* tab. Each element of the table header is described in a tool tip, but note that here the term *residue* actually refers to *occurrence of residue* (motif).

The table is interactive. Clicking on any element in the table header allows to sort the table entries according to that element. Click on any residue name in order to access the **Validator^{DB} specifics page** with detailed validation results for that residue (see section 5.2).

The filter at the top right corner allows to retrieve the table row with a specific residue. Simply type the residue name into the filter. All results can be downloaded in .csv format using the download button at the top left corner.

5.1.3. Details by PDB entry

The *Details by PDB Entry* tab (Figure 3C) contains an interactive table summarizing the results for all residues validated in each PDB entry. Each row corresponds to one PDB entry, identified by its PDB ID (4-letter code). The information in the table is organized according to the validation results as presented in section 4.7 of this manual. The color coding for the table header and the font inside the table is the same as in the categories defined in the *Overview* tab. Each element of the table header is described in a tool tip, but note that here the term *residue* actually refers to *occurrence of residue* (motif).

The table is interactive. Clicking on any element in the table header allows to sort the table entries according to that element. Click on any PDB in order to access the **Validator^{DB} specifics page** with detailed validation results for all residues in that PDB entry (see section 5.2).

The filter at the top right corner allows to retrieve the table rows with a specific residue, or the table rows with selected PDB IDs. Simply type the residue name or PDB ID into the filter. All results can be downloaded in .csv format using the download button at the top left corner.

5.2. Specifics page

The **Validator^{DB} specifics page** is accessible from the **synopsis page**, either via the *LookUp* bar on the *Overview* tab (Figure 1A), or via the residue names and PDB IDs in the interactive tables on the tabs *Details by Residue* (Figure 3B) and *Details by PDB Entry* (Figure 3C), respectively. Depending on how it was accessed, the **specifics page** might retrieve validation results for one or more residues, a fact mentioned at the very top of the page (Figure 1B).

The **ValidatorDB specifics page** (Figures 1B,4,5) 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 several tabs on the **specifics page**, namely *Summary*, *Details*, *Processing Errors/Warnings*, and *Overview*. These tabs will be described in detail in sections 5.2.1-5.2.4.

Inspecting the tabular and graphical validation reports accessible on the **specifics page** is the most comfortable and effective way to evaluate the results. Additionally, you may use the *JSON Data* download button at the top right corner of the **specifics page** in order to download the complete validation reports and perform any additional analyses on your own.

*Figure 4: The **ValidatorDB specifics page** defaults to the Summary tab. The list of residues for which results are available is given at the top of the page. 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, the **Summary** tab provides an overview of the validation results in tabular and graphical form. The issues reported are related to incomplete structure or incorrect chirality, as well as useful warnings (section 4.7). The color coding convention is maintained. Further, problematic atoms are highlighted, to better locate problems in the structures.*

5.2.1. Summary

On the **ValidatorDB specifics page**, the first view of the results is available in the *Summary* tab (Figures 1B,4A). For each validated residue, the *Summary* tab provides an overview of potential issues encountered, as described in section 4.7.

If more than one residue were validated in one run, a list of these residues will be at the top of the page. 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. Validation statistics are given as absolute numbers and percentages over all the motifs that were processed for each residue.

The table with the validation report 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 makes it difficult 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.

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.

5.2.2. Details

Whereas the *Summary* tab provides statistics of the issues over all validated motifs for each validated residue, the *Details* tab of the **Validator^{DB} specifics page** allows you to inspect the issues in select groups of motifs, and further in each individual motif (Figure 5A). 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* by its name (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, usually helpful in identifying the source of the error in the structure. The column *Id* refers to a unique identifier assigned to each motif in order to keep a transparent trace of the motif's origin, as it contains the PDB ID, as well as the serial index of the first atom in the motif, as it appears in the original PDB entry. 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 sort the motifs according to the property specified in the header.

5.2.3. 3D visualization

The *3D viewer* implemented in the **Validator^{DB}** interface offers one step further in the analysis of each individual validated motif, and is accessible via the *Details* tab on the **specifics page**

(Figure 5B). In the table, 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.

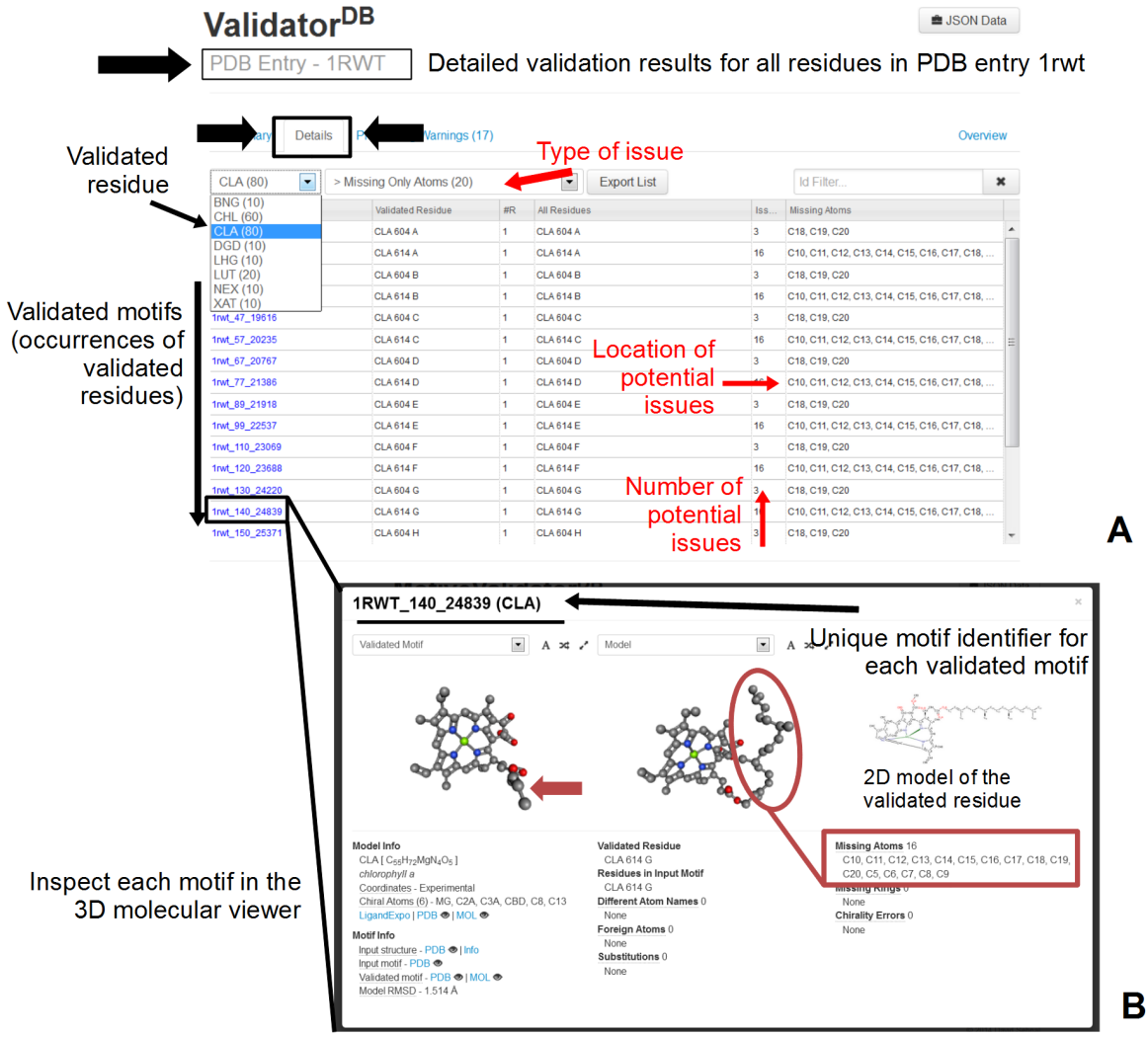

Figure 5: The Details tab of the ***ValidatorDB specifics page***

A) The Details tab allows to inspect the issues in selected groups of motifs, and further in each individual motif. A motif is a single instance (occurrence) of a validated residue, and is assigned a unique motif identifier based on the PDB ID of the PDB entry of origin, and the PDB atom identifier of the first atom in the motif. 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 motif.

B) 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. A 2D model is provided for clarity, especially helpful when working with larger and complex residues. Basic information about the validated residue in question is also given, along with a complete report of the validation results.*

5.2.4. Processing warnings and processing errors

The validation reports in **Validator^{DB}** also mention various unusual aspects encountered during validation. Sometimes the processed PDB entries contain information that is ambiguous, conflicting or which deviates strongly from the expected reference. **Validator^{DB}** reports such events as *processing warnings* or *processing errors*, depending on the severity of the deviations. Such information can be found in the *Processing Errors/Warnings* tab on the **specifics page**. The selection field at the top of the page 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.

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. **Validator^{DB}** typically reports several kinds of warnings: substitutions, foreign atoms, different naming, alternate locations and zero model RMSD, planar chiral center, unusual bond lengths, etc. 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.

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. 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 is 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 either the **synopsis** or **specifics page**. The number of motifs with processing errors can be easily calculated as the difference between total motifs of relevant residues in the PDB, and the number of analyzed motifs (currently around 2000 motifs, about 0.5% of relevant motifs in the PDB). Further, because **Validator^{DB}** automatically extracts all motifs of a relevant residue and assigns them a unique and informative motif Id, you will be able to easily find the motif in its original PDB entry, and explore it.

5.2.5. Overview

To keep consistency with the **synopsis page**, the **specifics page** also allows visualization of general validation statistics for a selected number of residues via the *Overview* tab. This representation is entirely compatible with that of the *Overview* tab on the **synopsis page** (Figure 5A), and in fact makes up a subset of that data set. All color coding conventions are kept, and tool tips provide descriptions of each graphical element.