iCn3D, a Web-based 3D Viewer for the Visualization of Biomolecular Structure and Sequence Annotation

Jiyao Wang[1*], Philippe Youkharibache[1,2], Dachuan Zhang[1], Christopher J. Lanczycki[1], Renata C. Geer[1], Tom Madej[1,2], Lon Phan[1], Minghong Ward[1], Shennan Lu[1], Gabriele H. Marchler[1], Yanli Wang[1], Stephen H. Bryant[1], Lewis Y. Geer[1], Aron Marchler-Bauer[1]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, MD, 20894, USA
[2]National Cancer Institute, National Institutes of Health, 8600 Rockville Pike, Bethesda, MD, 20894, USA
* wangjiy@ncbi.nlm.nih.gov

Keywords: molecular visualization, web-based, 3D viewer, biological annotation, sequence annotation, functional annotation

Abstract

Summary: iCn3D (I-see-in-3D) is a web-based 3D molecular structure viewer implemented in JavaScript. It invites users to explore sequence-structure relationships by displaying 3D structures, protein and nucleotide sequences, and biological annotations such as protein domains, Single Nucleotide Variations, functional sites, molecular contacts, as well as user defined features. Customized and user-annotated displays can be shared with colleagues or publishers via a simple URL, or embedded in web pages as live 3D displays. iCn3D is unique in its ability to simultaneously show 3D structure, 2D molecular networks and interactions, and 1D protein and nucleotide sequences through an integrated sequence/annotation browser. Arbitrary molecular features can be selected in any of the 1D/2D/3D windows as sets of molecules or residues, utilizing flexible user interfaces. These selections are synchronized dynamically in all displays. They can, for example, be shown as tracks in the 1D sequence/annotation browser with simultaneous 3D representations. They can be saved and reused throughout a single user session, or exported for subsequent sessions and further external analysis. iCn3D is rich in structural analysis and visualization features and has a built-in user interface that supports both command line- and menu-driven operation. Session commands can be saved, replayed and shared. iCn3D can display superimposed structures obtained from NCBI's VAST+ structural alignment resource, for comparative analysis of tertiary or quaternary structure. iCn3D can also display electron density maps for crystal

structures and export files for 3D printing. The following example URL exemplifies some of the 1D/2D/3D representations:
https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?mmdbid=1TUP&showanno=1&show2d=1&showsets=1.

Availability and implementation: iCn3D is freely available to the public. Its source code is available at https://github.com/ncbi/icn3d.

To whom correspondence should be addressed: wangjiy@mail.nih.gov

1 Introduction

With the widespread availability of powerful personal computers and mobile devices, the visualization of biomolecular 3D structures is no longer restricted to stand-alone 3D viewer applications such as PyMol, Chimera, Jmol, Swiss-Pdb, or Cn3D, to name a few (Craig, et al., 2013), which all require the installation of an application; interactive visualization of large molecular structures can now be achieved via web-based technologies such as WebGL. One of the earliest web-based 3D viewers was GLmol, developed by Takanori Nakane in 2011 (http://webglmol.osdn.jp/index-en.html). It used the Three.js library and could display both proteins and nucleic acids. Subsequently, iview (Li, et al., 2014), itself based on GLmol, introduced a wider range of user interface features. Recently NGL (Rose, et al., 2018; Rose and Hildebrand, 2015) and LiteMol (Sehnal, et al., 2017) have been developed at RCSB (PDB) and PDBe, to visualize large structures by parsing compressed binary files MMTF (Bradley, et al., 2017) and BinaryCIF (Sehnal, et al., 2017), respectively. Other WebGL-based 3D viewers include JSmol (Hanson, et al., 2013), 3Dmol (Rego and Koes, 2015), Molmil (Bekker, et al., 2016), Web3DMol (Shi, et al., 2017), among others.

NCBI developed and has maintained "Cn3D" (Wang, et al., 2000), a stand-alone application written in C++. As such, Cn3D users need to install the application and configure their web browsers to view and manipulate molecular structures in 3D. One of Cn3D's strengths is its ability to visualize multiple 3D structures and 1D sequences as well as sequence alignments simultaneously. The downsides of a stand-alone application, such as the installation overhead and the need to release updates to roll out novel features, can now be eliminated by taking advantage of advanced web technology. To this end we have developed iCn3D, which covers many of the features provided by Cn3D, such as sequence displays in sync with the 3D structure displays. iCn3D also introduces novel features as suggested in "Twelve Elements of Visualization and Analysis" (Youkharibache, 2017). As described below, these features include:

1) selection behavior synchronized between displays of 3D structure, 2D interaction schematic, 1D sequence, and a menu of (pre-)defined sets, many of which are defined automatically during an interactive session to track user selections;
2) annotations such as the location of non-synonymous SNPs (automatically extracted and mapped to 3D structures from NCBI dbSNP database (Sherry, et al., 2001)), protein domains, functional sites, molecular contacts, and user annotations that can be defined interactively or through file/text based mechanisms, and can be displayed as "custom tracks" in the sequence/annotation viewer;
3) custom sequence/structure (sub)set selections;
4) "Share Link" URLs to share visual displays of interest over the internet, with all the underlying annotations and user-provided labels; and
5) stereolithography (STL) or Virtual Reality Modeling Language (VRML) files to support 3D printing.
6) electron density map for crystal structures.

2 Implementation

2.1 Third-party Libraries and Tools

iCn3D Structure Viewer is a WebGL-based 3D viewer using Three.js and jQuery. Its source code is available at GitHub (https://github.com/ncbi/icn3d). To view 3D structures, no installation is required. iCn3D displays the 3D structures directly in a web browser. One example is shown in Figure 1, which can also be opened in a live browser window with the shortened URL https://d55qc.app.goo.gl/YnHHnGrakpwcc4gq6. iCn3D has been developed based on the 3D styles of iview (Li, et al., 2014) and GLmol, the surface display of 3Dmol (Rego and Koes, 2015), and the sphere and cylinder display of NGL using the impostor technology (Rose, et al., 2018; Rose and Hildebrand, 2015).

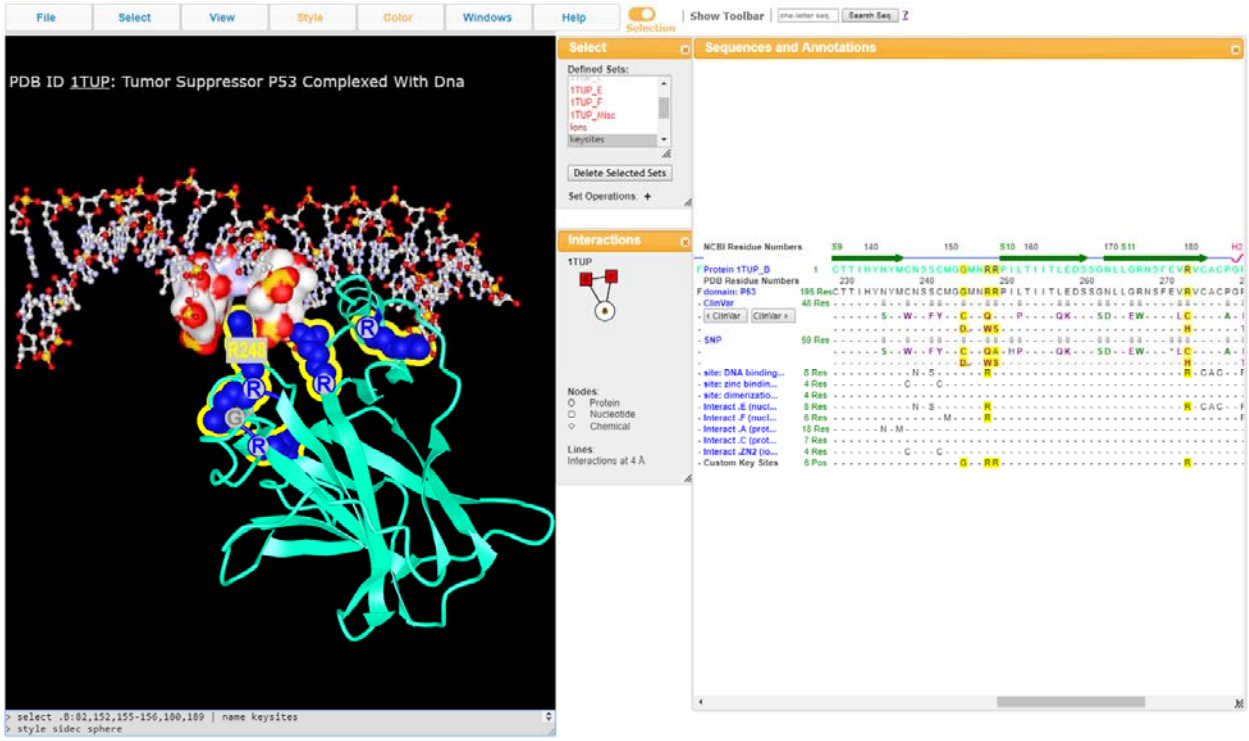

Figure 1: iCn3D visualizing the structure of tumor suppressor P53 complexed with DNA (PDB code 1TUP). The interactive view includes 3D coordinates (left), a menu of defined sets (top-middle), a 2D schematic of the molecules involved in the complex structure (bottom-middle), and biomolecular sequences and annotations, such as domains, SNPs, ClinVar (Landrum, et al., 2016), functional sites, interactions (molecular contacts), and custom tracks (right). This display can be reproduced in a live browser window with the shortened share link: https://d55qc.app.goo.gl/YnHHnGrakpwcc4gq6.

2.2 Data Types

iCn3D reads 3D structure data from a variety of sources as shown in Table 1. The type of input structure identifier determines the file format that will be loaded into iCn3D for a given structure, and may slightly affect details of the display. The MMTF ID is preferable to PDB ID and mmCIF ID because MMTF data cover all 3D structures, are highly compressed, include DSSP-calculated secondary structures if the secondary structures are not provided originally, and include bond types such as double bond or triple bonds in the chemicals. The MMDB ID loads a structure in NCBI's MMDB file format, which provides unique views not available in the other formats, such as the 2D interaction schematic display. iCn3D also reads local PDB files, mmCIF files, Mol2 files, SDF files, XYZ files, or files from a server via URLs.

Data Type	URL Parameter
PDB ID	pdbid=1tup
mmCIF ID	mmcifid=1tup
MMTF ID (Bradley, et al., 2017)	mmtfid=1tup
NCBI MMDB ID	mmdbid=1tup
NCBI gi	gi=1310961
NCBI VAST+ alignment	align=1HHO,4N7N
NCBI PubChem (Kim, et al., 2016) CID	cid=2244

Table 1: iCn3D reads 3D structure data from a variety of sources. For example, the URL to display the PDB ID 1TUP is formatted as the base URL "https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?" plus the parameter "pdbid=1tup".

Different views in iCn3D (such as the 2D interaction schematic) can be shown or hidden using URL parameters shown in Table 2. Users can also add commands in the URL directly using a parameter such as "&command=color+secondary+structure".

View	Parameter to Show	Parameter to Hide
2D Interaction Schematic	&show2d=1	&show2d=0 (default)
Biological Annotations	&showanno=1	&showanno=0 (default)
Defined Sets	&showsets=1	&showsets=0 (default)
Menu	&showmenu=1 (default)	&showmenu=0
Command Window	&showcommand=1 (default)	&showcommand=0

Table 2: Different views in iCn3D can be shown or hidden using URL parameters. For example, the URL to show the 2D interaction schematic is formatted as the base URL "https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?mmdbid=1tup" plus the parameter "&show2d=1".

2.3 Selection Mode

All operations in iCn3D apply to either all atoms, or to the current selection. By default, iCn3D uses the "All atoms" selection mode and applies the "style" and "color" options to all atoms. When users select part of the structure, the mode becomes "Selection", and the mode switch button (next to the "Help" menu, Figure 1), as well as the "Style" and "Color" menus, become orange. The next operation applies only on the selected atoms. If users want to change style or color for all atoms, they can toggle the mode button back to "All atoms".

2.4 Hydrogen Bonds in Chemical Binding

Users can select "Chemical Binding" in the View menu to view hydrogen bonds between chemicals and proteins. A pair of atoms form a hydrogen bond if their elements are among nitrogen, oxygen, or fluorine and the distance between them is less or equal to 3.5 angstrom. The side chains of hydrogen donors or acceptors are shown in the Stick style. Hydrogen bonds are shown as dashed green lines. The view zooms in to all chemicals and their hydrogen-bonded residues. If users are interested in only one chemical, they can select the chemical by holding the "Alt" key and clicking on the chemical in the 3D structure, and then click "H-Bonds to Selection" in the View menu to show the hydrogen bonds with this chemical.

2.5 Surface

iCn3D can compute and display multiple surfaces for multiple selections. Three types of surfaces are available: Van der Waals Surface, Molecular Surface, and Solvent

Accessible Surface (Xu and Zhang, 2009). By default, the surfaces are computed for selected atoms, without considering surrounding atoms. Users can select surface types containing "with Context" to show the surfaces considering the surrounding atoms. The color of a surface segment is determined by the color of the proximal atoms. Surfaces are generated with a maximum grid size of 180 angstroms. If the size of the structure is larger than the maximum grid size, a scale factor is used to fit the structure into this maximum grid size.

2.6 Electron Density Map

Figure 2 shows both electron density map and B-factor (temperature factor) values for the PDB structure 3GVU. The 2Fo-Fc electron density map (contours in cyan) is contoured at 1.5 σ by clicking "Electron Density > 2Fo-Fc Map" in the Style menu. The structure is colored by B-factor values in the range of 0 (blue) -100 (red). The thickness of the B-factor tube is proportional to the B-factor value. The electron density map is consistent with the B-factor values: residues with large B-factor (white thick tubes) have poor electron density map.

iCn3D can show both 2Fo-Fc and Fo-Fc electron density maps for any subset of a crystal structure using the DSN6 file retrieved from RCSB. Users can choose a sigma value between 0 and 10 σ for the electron density maps. The default value is 1.5 σ for 2Fo-Fc map and 3.0 σ for Fo-Fc map. In the Fo-Fc map, the positive sigma values are colored in green and the negative sigma values are colored in red.

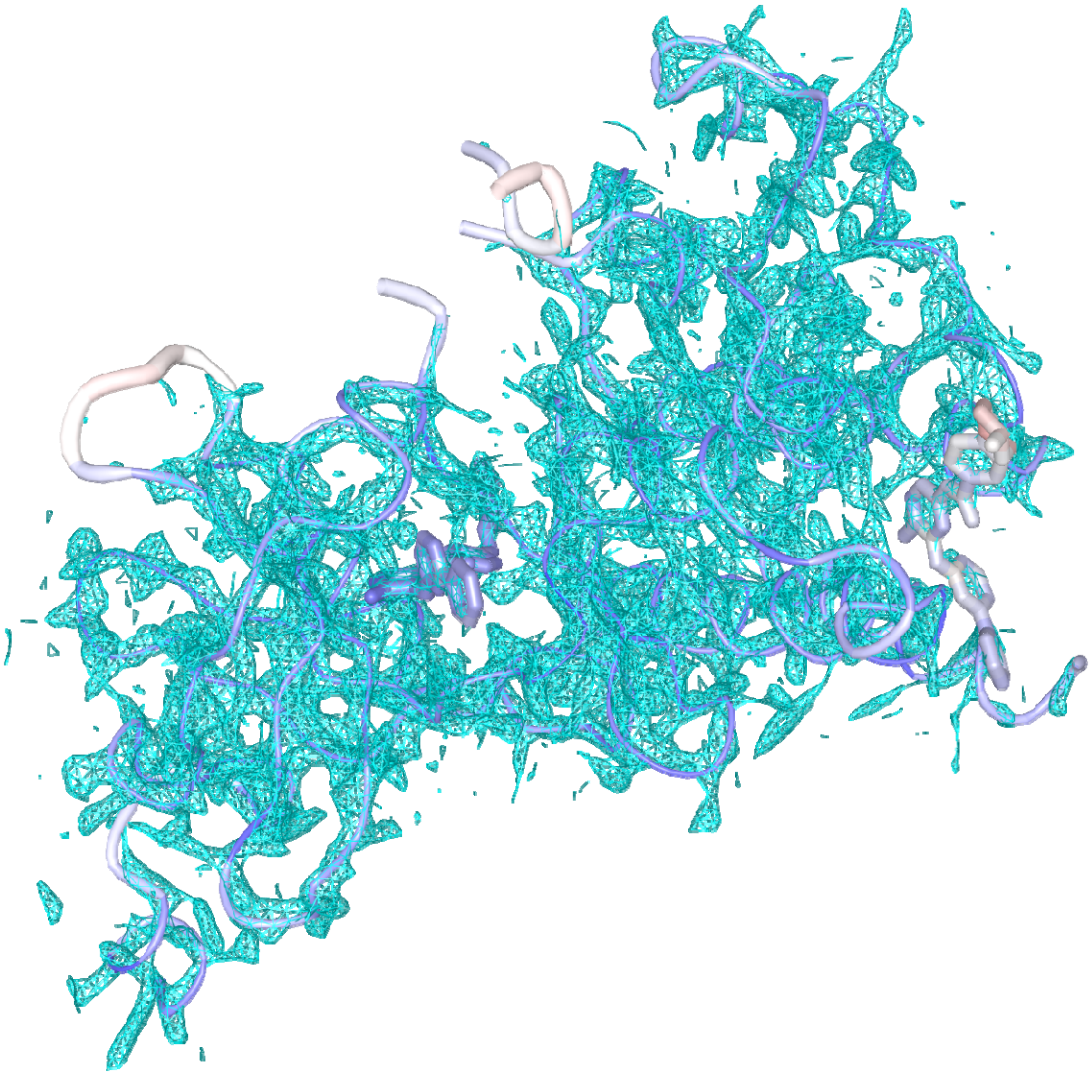

Figure 2. Electron density map and B-factor values of the PDB structure 3GVU. The 2Fo-Fc electron density map is contoured at 1.5 σ. The tube represents the asymmetric unit of the crystal structure. Its color changes from blue to red as the B-factor increases from 0 to 100. The tube is thicker when the B-factor value is larger. Share link: https://d55qc.app.goo.gl/MtW8mBGsvYCNEMJh8.

2.7 Style

Several commonly used representations such as Ribbon, Cylinder and Plate, C Alpha Trace, B-factor Tube, Stick, Sphere, etc. are available to display molecules. In addition, iCn3D has a novel display style called "Schematic", which is available for proteins, nucleotides, and chemicals. The schematic style shows one-letter residue abbreviations in the positions of the C-alpha atoms for proteins or the O3' atoms for nucleotides. For chemicals, the schematic style renders bonds as sticks, and labels all non-carbon and non-hydrogen atoms with their element names.

The shiny surface display for 3D structures is achieved by using the material MeshPhongMaterial in Three.js. The arrows in the ribbon style of beta sheets are achieved by adding a triangle at the end of each strand. The triangle follows the curve of the strand, which are smoothed using the cubic spline algorithm.

If the structure is specified via a PDB ID or a PDB file, and the secondary structure assignments are missing, the DSSP method (Sicheri, et al.) is used to calculate the secondary structures.

2.8 Color

iCn3D offers many coloring schemes: "Spectrum", "Secondary", "Charge", "Hydrophobic", "Chain", "Residue", "Atom", "B-factor", "Unicolor", and "Color Picker". The "Charge" coloring scheme renders Arg and Lys residues in blue, and all nucleotides, Asp and Glu residues in red. The "Hydrophobic" coloring scheme renders Gly, Pro, Ala, Val, Leu, Ile, and Phe residues in green, and the remaining residues in grey. iCn3D also allows users to pick a color using "Color Picker" (https://github.com/tovic/color-picker).

2.9 Commands

iCn3D has a defined command associated with each operation. All commands are sequentially saved in memory. These commands enable users to share URL links (e.g., https://d55qc.app.goo.gl/YnHHnGrakpwcc4gq6), save the visualization's current

state in a state file, undo and redo display changes, and directly modify the display using commands in the grey command window that appears beneath the 3D display.

The command corresponding to each operation is shown in that command window. Users can also find both the commands and the corresponding JavaScript functions by clicking “Web APIs > Commands” in the Help menu. On the Web APIs page, the “Menu”, “Command”, and “Method” columns list the operation name in the menus, the command name, and the JavaScript function, respectively.

2.10 Performance

iCn3D displays large structures very quickly. For example, it can display the complex structures of the large ribosomal subunit of Haloarcula marismortui (https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?mmtfid=1ffk) and of a virus capsid (https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?mmtfid=2bbv) in a few seconds. These structures contain 64,281 and 469,020 atoms, respectively.

Several technologies are used to speed up the display. First, iCn3D uses the impostor technology (Rose, et al., 2018; Rose and Hildebrand, 2015) to render spheres and cylinders directly with shaders when the WebGL extension on fragment depth value is available in the browser. Since spheres and cylinders are used a lot, the impostor technology speeds up the rendering significantly. Second, iCn3D uses Instanced Buffer Geometry in Three.js to render biological assemblies. Each biological assembly consists of many copies of an asymmetric unit. The Instanced Buffer Geometry technology speeds up the display by creating the geometry of the asymmetric unit only once, transforming it according to the symmetry matrices to generate translated/rotated copies, and rendering the transformed copies. Third, iCn3D accepts the highly compressed binary MMTF data format. This greatly speeds up the loading of 3D coordinates.

2.11 Embed iCn3D with iframe or JavaScript libraries

iCn3D can be embedded in a web page by including the URL in HTML iframe, e.g. <iframe src="https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?mmdbid=1tup&showanno=1&show2d=1&showsets=1&width=800&height=450&showmenu=1&showcommand=1&rotate=right" width="900" height="600" style="border:none"></iframe>. This method always shows the most recent version of iCn3D.

To embed iCn3D with JavaScript libraries, the following libraries need to be included: jQuery, jQuery UI, Three.js, and iCn3D library. An html div tag to hold the 3D viewer is added: “ ”. The iCn3D widget is initialized with the custom defined parameter “cfg”: “var icn3dui = new iCn3DUI(cfg); icn3dui.show3DStructure();”. Multiple iCn3D widgets can be embedded in a single page. Users can choose to show the most recent version of iCn3D, or a locked version of iCn3D. To show the most recent version, use the library files without the version postfix as shown in the iCn3D Web API page (https://www.ncbi.nlm.nih.gov/Structure/icn3d/icn3d.html#HowToUse). To show a locked version, use the library files with the version postfix as shown in the source code of iCn3D page (https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?mmdbid=1tup). If the input is provided as an MMDB ID, both library files and backend cgis are versioned so that the 3D display will be stable.

3 Features

3.1 Share link

One important feature of iCn3D is the facility to share a display, which includes both selections and rendering choices, with other users through a single URL. One example URL is https://d55qc.app.goo.gl/YnHHnGrakpwcc4gq6 (Figure 1), which is shortened using google Firebase Dynamic Links from the original URL: https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?mmdbid=1tup&showanno=1&show2d=1&showsets=1&command=view+annotations;+set+annotation+all;+set+view+detailed+view;+select+saved+atoms+1TUP_E+or+1TUP_F;+color+atom;+style+nucleotides+ball+and+stick;+select+.B:155;+select+zone+cutoff+4;+select+saved+atoms+1TUP_E+or+1TUP_F+and+sphere.B:R155-4A;+set+surface+Van+der+Waals+surface;+select+saved+atoms+1TUP_B+or+1TUP_E+or+1TUP_F;+show+selection;+add+track+|+chainid+1TUP_B+|+title+Custom+Key+Sites+|+text+82+R,+152+G,+155-156+RR,+180+R,+189+R;+select+.B:82,156,180,189;+color+blue;+add+residue+labels;+select+.B:152;+color+grey;+add+residue+labels;+select+.B:155;+color+blue;+add+label+R248+|+x+51.76+y+21.37+z+83.13+|+size+40+|+color+ffff00+|+background+cccccc+|+type+custom;+select+.B:82,152,155-156,180,189+|+name+keysites;+style+sidec+sphere|||{"factor":"1.1127","mouseChange":{"x":"0.065352","y":"0.032319"},"quaternion":{"_x":"0.10007","_y":"-0.22373","_z":"0.21026","_w":"0.94643"}}. The original URL includes the following commands, where "+" symbols are replaced by spaces. First, show sequences and annotations, 2D interaction schematic, and defined sets for the PDB structure 1TUP using the parameters "showanno=1", "show2d=1", and "showsets=1". Second, show all annotations and view the sequence details using the commands "view annotations; set annotation all; set view detailed view;". Third, color the nucleotide chains by atoms and display them with the style of ball and stick using the commands "select saved atoms 1TUP_E or 1TUP_F; color atom; style nucleotides ball and stick;". Fourth, show the surface of nucleotides within 4 angstroms from Arg 248 in the chain B. When the input is MMDB ID, iCn3D uses NCBI residue numbers instead of PDB residue numbers. These two numbers are the same for most structures. But they are different by 93 for the PDB structure 1TUP. Thus, the corresponding NCBI residue number for Arg 248 is 155. We use the following commands to show the surface: "select .B:155; select zone cutoff 4; select saved atoms 1TUP_E or 1TUP_F and sphere.B:R155-4A; set surface Van der Waals surface;". Fifth, display only the chains 1TUP_B, 1TUP_E, and 1TUP_F using the commands

“select saved atoms 1TUP_B or 1TUP_E or 1TUP_F; show selection;”. Sixth, add a custom track with the key residues arginine 175, 248-249, 273 and 282, and glycine 245 using the commands “add track | chainid 1TUP_B | title Custom Key Sites | text 82 R, 152 G, 155-156 RR, 180 R, 189 R;”. Note the residue numbers used in the commands are 93 smaller than the “PDB Residue Numbers”. Seventh, color the key residues Arg in blue and Gly in grey and add residue labels using the commands “select .B:82,156,180,189; color blue; add residue labels; select .B:152; color grey; add residue labels; select .B:155; color blue; add label R248 | x 51.76 y 21.37 z 83.13 | size 40 | color ffff00 | background cccccc | type custom;”. Eighth, display the side chains of the key residues with the style of sphere using the commands “select .B:82,152,155-156,180,189 | name keysites; style sidec sphere;”. Finally, the orientation of the 3D display is set using the parameters after “|||”.

The Share Link above conveniently reproduces the custom display in Figure 1. This link can be shared with colleagues, inserted into presentations, or included in a publication, etc. It can be generated by clicking “Share Link” in the File menu. It is also available in the last line of each saved iCn3D PNG image, which is generated by clicking “Save Files > iCn3D PNG Image” in the File menu. Images generated by clicking the right mouse button over the 3D display, or by printing the screen do not have the embedded link. Furthermore, this saved iCn3D PNG image can be used to reproduce the custom display by clicking “Open File > iCn3D PNG Image”. Thus, users can use these iCn3D PNG images to save their custom displays or reproduce these custom displays later.

3.2 Synchronized Selection among 3D structure, 2D Interactions, Defined Sets, and 1D Sequences and Annotations

As shown in Figure 1, iCn3D highlights (in yellow) key residues involved in the interaction of the tumor suppressor P53 with DNAs (PDB ID 1TUP): arginine 175, 248-249, 273 and 282, and glycine 245 (Cho, et al., 1994). The residue Arg 248 at chain B inserts into the major groove of the DNA double-helix. Other highlighted residues are also involved in protein-DNA interactions or protein-protein interactions. Mutations of these key residues may affect the interaction of P53 with DNAs and cause cancer (Vogelstein, et al., 2000). In the "Sequences and Annotations" window, users can click on the buttons "< ClinVar" and "ClinVar >" to browse through ClinVar (clinically relevant variants) annotation, which will be highlighted and labeled on the 3D display. Users can also mouse over the residues in the ClinVar track of the "Sequences and Annotations" window to see details about each variant. The residues colored in purple are pathogenic and those colored in green have no known clinical significance. The "SNP" track shows the locations of variants from NCBI's dbSNP database.

iCn3D simultaneously shows 3D coordinates, 2D interaction schematic, a list of defined sets, and 1D sequences and annotations. (The 2D interaction schematic is displayed only when the input data type is MMDB ID.) A selection in any of these four displays triggers the highlight in the other three displays. To select the residue Arg with NCBI residue number 155, users can put the mouse over the residue in the groove to see the mouseover text "Arg155", then hold the "Alt" key and click on the residue. The Ray Casting technology (https://threejs.org/docs/#api/core/Raycaster) is used to pick an atom in the structure. Once an atom is picked, the residue containing the atom is selected by default. Other options under "Select on 3D" in the Select menu include "Chain", "Strand/Helix", "Atom". Once an atom is selected, users can use the up and down arrows to switch among different levels: atom, residue, strand/helix, chain, and the whole structure. Users can hold the "Ctrl" key to combine selections and hold the "Shift" key to select a range of residues. To toggle the highlight or remove the labels, users can click on "Show Toolbar" at the top-right corner and click on the button "Toggle Highlight" or "Remove Labels".

To select chain B on the 2D interaction schematic, users can click on the node with the label B in the interaction schematic. Users can also click on the line that connects chain B with chain E. This operation selects the residues in chain B that interact with chain E. Two residues are considered as interacting residues if any of their non-hydrogen atoms are found within 4 angstroms of each other. The selected residue set is saved with a name “inter_B_E” in the list of “Defined Sets”. If only part of the chain is selected, a smaller node is shown inside the original node in the 2D interaction schematic. Selections on 2D interactions can be combined by holding the “Ctrl” key. Only the selected chains on 2D interactions show up in the “Sequences and Annotations” window.

Users can also click on “1TUP_E” in the list of “Defined Sets” to select the DNA chain E. Only the selected chain E will show up in the “Sequences and Annotations” window. Users can hold the “Ctrl” key to select multiple sets in “Defined Sets” and choose to display only these sets by clicking the button “View Only Selection” in the View menu. Users can also change the “union/or” of multiple sets to “intersection/and” or “exclusion/not” by clicking the “+” sign after “Set Operations” and clicking the button “Save Selection” in the window of “Defined Sets”. For example, if users click four sets ":1-10", ":11-20", ":5-15", and ":7-8" in the "Defined Sets", the command "saved atoms :1-10 or :11-20 or :5-15 or :7-8" will show up after the text “Select”. Users can modify the command to "saved atoms :1-10 or :11-20 and :5-15 not :7-8", which combines all residues 1-10 and 11-20 to get the residues 1-20, then intersects with the residues 5-15 to get the residues 5-15, then exclude the residues 7-8 to get the final residues 5-6 and 9-15.

3.3 Sequences & Annotations

The “Sequences and Annotations” window has two views, “Summary” and “Details”. The “Summary” view shows a graphical representation of each sequence as a colored bar, and the “Details” view shows the detailed sequence, including individual residues as well as any biological annotations users have chosen to display. iCn3D provides

several annotations originating from various NCBI databases: SNPs (from dbSNP), ClinVar (annotated SNPs from ClinVar), Conserved Domains (from CDD (Marchler-Bauer, et al., 2017)), 3D Domains (from MMDB), Functional Sites (from CDD), and Interactions (from MMDB). Each annotation contains the title followed by the sequence. Users can click on the title to select all sequences in this track. To select multiple tracks, click the titles while holding the "Ctrl" key. To select a few residues in the sequence, users can simply use mouse to drag on the sequence. Each additional drag adds more residues to the selection. To deselect a residue, drag on the residue again. To show the sequences and annotations of all chains, click the button "Show All Chains" in this window.

SNPs and ClinVar tracks show the pathogenic variants in purple, other ClinVar clinical significance variants in green, and variants, which are not in ClinVar but are in dbSNP, in black (Figure 1). (dbSNP annotates variants to RefSeq protein sequences as part of its regular build release process. The mapping of dbSNP variants to PDB structures was done by using the annotated RefSeq protein sequences to BLAST search against PDB database. The BLAST alignments were filtered to include only the top PDB protein hits with E-value less than 0.0001 and were used to project the variants from RefSeq proteins to the corresponding positions on the PDB proteins.) Conserved Domains show the location of conserved domain footprints as pre-computed by CDD annotation services. 3D Domains are domain models based on 3D structures and reflect compact structural units within a protein that have been identified using purely geometric criteria. Functional Sites are obtained from pre-computed CDD annotation and are defined as sets of residues that are associated with molecular function and presumed conserved in a domain family, such as binding sites, or active/catalytic sites. Interactions shows the residues in one molecule/chain in contact with another molecule, and are calculated on-the-fly, or pre-calculated at NCBI by identifying non-hydrogen atoms that are found within 4 angstroms of each other. They can also be selected from the 2D interaction schematic.

3.4 Custom Tracks

Users can also add custom tracks in the Sequences and Annotations window to complement or supplement existing annotations. To add custom tracks, click the button “Add Track” in each chain section. Currently there are five ways to add custom tracks. First, the current selection can be shown as a track. Users can make any selection and show the selection as a track. Second, users can input a protein sequence in the FASTA format. The sequence is aligned with the sequence of the current chain using the BLAST algorithm. If a match is confirmed, the matched part of the input sequence will be shown as a custom track. Otherwise, an error message “cannot be aligned” will be displayed. Third, users can input a BED file to show multiple tracks. The BED file can include color information. Fourth, users can add any custom text as a track (e.g., “Custom Key Sites” in Figure 1). To save space, the input is a list of comma-separated position and text pairs. For example, "152 G, 155-156 RR" defines a character "G" at the position 152 and two continuous characters "RR" at positions from 155 to 156. The starting position is 1. Fifth, users can input NCBI protein accessions to add a track. The sequence of the protein will be aligned with the sequence of the current chain using BLAST, and the matched sequence will be shown as a custom track.

3.5 Search Sequences and Advanced Selection

iCn3D features a search button at the top-right corner. Users can search a segment of sequence in the one-letter format (e.g., “GXXRR”). Matched residues will be highlighted in 3D structure, 2D interactions, the list of Defined Sets, and 1D Sequences and Annotations. The matched residues are also saved as a defined set with the name such as “:GXXRR”. The symbol “:” means residues and “X” is a wildcard character to represent any single residue. The search shares the same specification as the “Advanced” option in the Select menu, except that it is not necessary to include the “:” symbol when using the “Search Seq.” box to find residues within sequences.

The “Advanced” selection is a powerful tool to build custom selections. Users can input a selection command with a simple specification and save the selection with a name. For example, in the selection “$1TUP.A,B:5-10,R,ions@CA,C”, the term “$1TUP” uses the prefix “$” to indicate a list of comma-separated structures; “.A,B” uses the prefix “.” to indicate macromolecules/chains; “:5-10,R,ions” uses the prefix “:” to indicate individual residues. Residues can be specified via residue numbers or ranges of residue numbers (e.g., “5-10”), one-letter residue codes (e.g., “R”), or via predefined names: “proteins”, “nucleotides”, “chemicals”, “ions”, and “water”. The prefix “@” is used to specify individual atoms. For example, “@CA,C” instructs iCn3D to select the alpha carbon and carbon atoms in the specified regions of the structure. Partial specifications are allowed, e.g., the selection “:5-10” selects residues with the residue numbers in the range of 5-10 in all structures and chains.

3.6 Visualization of Structure Alignment

iCn3D can visualize 3D structure alignments calculated by NCBI’s VAST+ service (Madej, et al., 2014). As shown in Figure 3, the 3D structure of oxy-hemoglobin (1HHO) is aligned with that of deoxy-hemoglobin (4N7N). Users can alternate these two structures by pressing the key “a”. To view both structures again, click “View Only Selection” in the View menu.

The aligned residues are colored in red. Oxy-hemoglobin (1HHO) and deoxy-hemoglobin (4N7N) are displayed with the style “C Alpha Trace”. In the 2D interactions window, pairs of aligned protein molecules have the same number in red, e.g., both nodes 1HHO_A_1 and 4N7N_C are labeled with the number 3 colored in red.

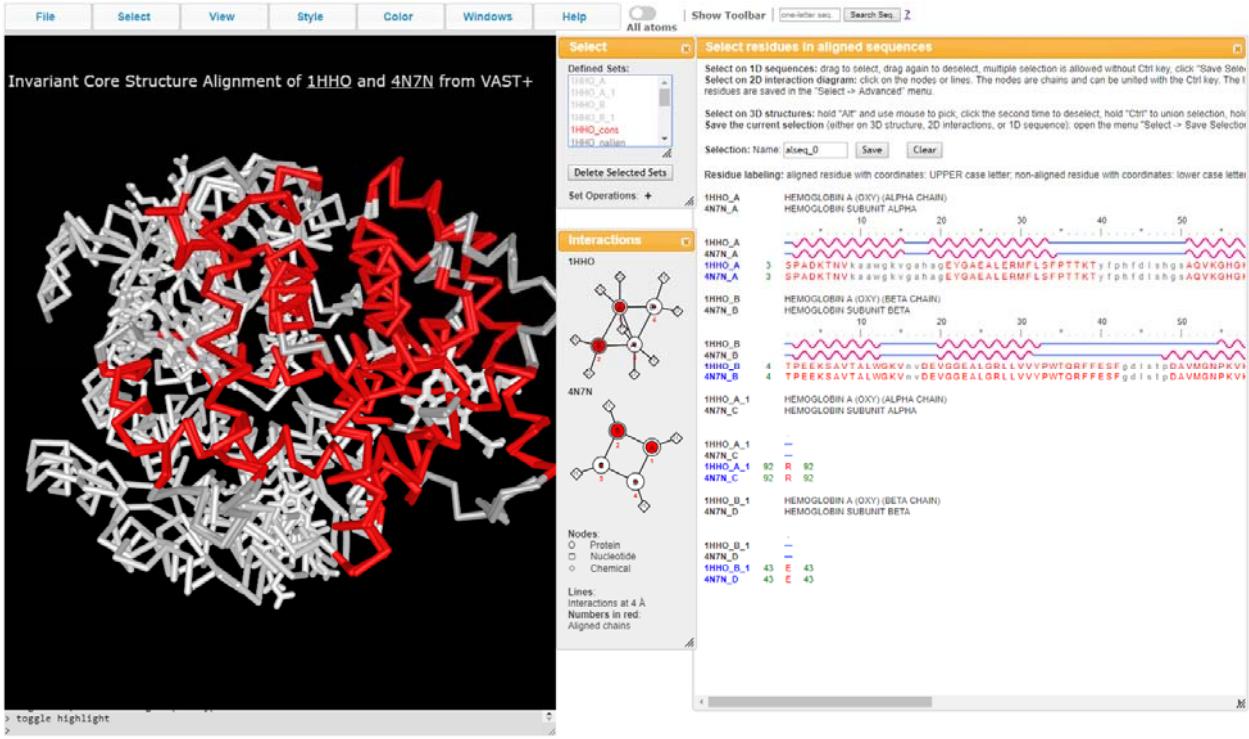

Figure 3: iCn3D visualizing the structure superposition and corresponding alignments between human oxy and deoxyhemoglobin (PDB ID 1HHO and 4N7N), as computed by VAST+. (Share Link: https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?align=1HHO,4N7N&showalignseq=1&atype=1&show2d=1&showsets=1)

3.7 3D Printing

iCn3D enables users to export STL or VRML files for 3D printing. An STL file can be printed in most 3D printers, but it has only one color. A VRML file can be used for 3D color printing. Some common 3D display styles such as “Ribbon” are suitable for 3D visualization but may require extra supporting structures/material in 3D printing. iCn3D adds default stabilizers if the export method is “STL with Stabilizers” or “VRML (Color, with Stabilizers)” in the File menu. The stabilizers for nucleotides are the hydrogen bonds. The stabilizers for proteins are connections made from every third residue to another residue that is within 4 angstroms. The stabilizers for chemicals and ions are connections from every 10th atoms in the chemicals or ions to another residue that is within 4 angstroms. Users can customize the size of the 3D objects by clicking “3D Printing > Set Thickness” in the File menu.

To build a printable 3D model, users can click “3D Printing > Add All Stabilizers” in the File menu to evaluate whether the 3D model appears stable. If not, they may click “Add One Stabilizer” to add extra stabilizers. Then users can click “STL” or “VRML (Color)” to export files. When users print a large biological assembly (e.g., a virus with 60 asymmetric units), each asymmetric unit is exported as a separate file. Users can then use external software to merge these files into a single file for 3D printing.

4 Web Applications

(a) (b)

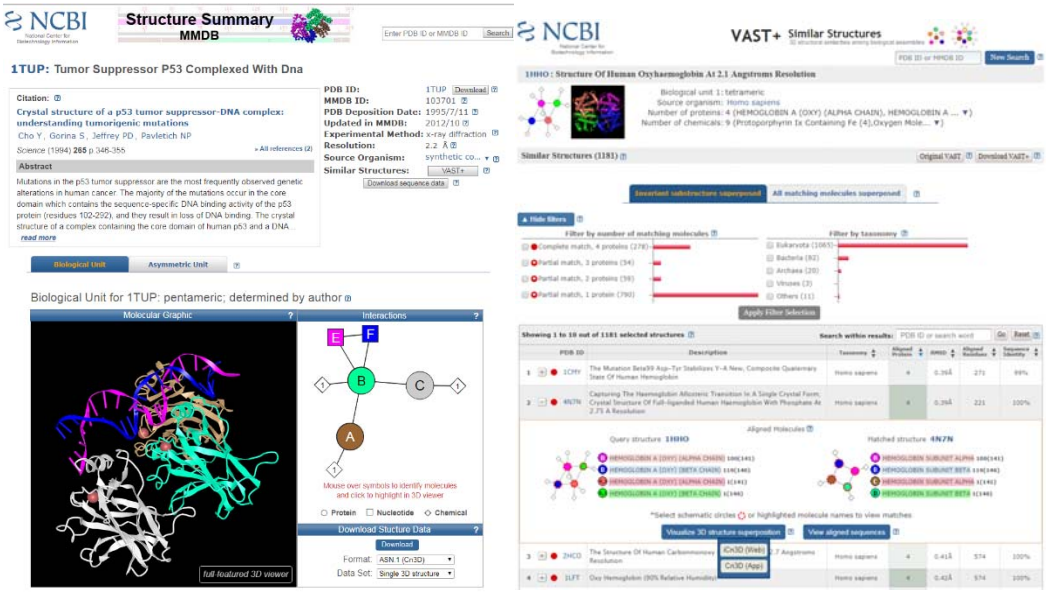

Figure 4: (a) MMDB page showing the PDB structure 1TUP. (Link: https://www.ncbi.nlm.nih.gov/Structure/pdb/1TUP) (b) VAST+ page showing the similar structures of the PDB structure 1HHO. (Link: https://www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi?uid=1HHO)

As shown in Figure 4a, iCn3D has been implemented as the default viewing option in the structure summary pages of NCBI’s structure database MMDB (e.g., https://www.ncbi.nlm.nih.gov/Structure/pdb/1TUP). The interactive 3D displays can be rendered without requiring users to install a stand-alone application, such as the

Cn3D viewer (Wang, et al., 2000). The structure summary page shows a static molecular graphic image by default. Users can click the button "3D view" at the bottom-left corner of the image to view the structure using iCn3D or click the button "full-featured 3D viewer" at the bottom-right corner to launch iCn3D in a separate window.

As shown in Figure 4b, iCn3D has also been implemented as a viewer in the VAST+ resource pages, which summarize the results of 3D comparisons between similar macromolecular complexes, including superpositions and corresponding alignments, e.g., https://www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi?uid=1HHO. To view structure superpositions, users need to select one of the aligned structures and expand the corresponding table row to show details about the comparison. In the resulting panel, click on the button "Visualize 3D structure superposition", and choose "iCn3D (Web)" in the popup to launch iCn3D.

5 Discussion

iCn3D has been developed as a replacement for Cn3D with respect to visualization tasks involving 3D structure data provided by NCBI. As with its predecessor Cn3D, the combination of macromolecular structure (3D) and sequence (1D) information, and the additional visualization of macromolecular interactions (2D schematic), facilitate the study, interpretation, and understanding of 3D structures and structural complexes. A powerful collection of methods for generating and refining selections and the ability to import and export user-defined selections enable users to communicate complex functional aspects of molecular systems. Additionally, the "Share Link" feature not only enables users to show a snapshot to others, but also makes it possible to replace static 3D structure images in web pages with interactive 3D viewers.

The annotations, such as mutations from ClinVar or dbSNP, functional sites, interaction interfaces, structural or conserved domains, and their simultaneous

visualization in 1D/2D/3D can provide useful and compelling evidence relating sequence, structure and function. A cross comparison among all these and any additional user-defined annotations, represented as tracks in the (1D) sequence/annotation window (e.g., Custom Key Sites in Figure 1), may facilitate analyses and make it easier to spot cross correlations between sequence and structure relating to function. More detailed genome level annotations could be added to the tracks in the future, merging genome/protein sequence browser related capabilities with corresponding molecular structures and structural complexes.

Acknowledgments

We would like to thank Dr. Terry S. Yoo and Dr. David T. Chen for their helpful discussions on generating models for 3D printing. Dr. Alexander S. Rose provided valuable suggestions on reading MMTF files and implementing impostor methods. We also thank Qiangling Li, Spencer Bliven, Eli Draizen, Jose Duarte, Keiichiro Oto, and Tim Schaefer for their hackathon on the 2D interaction schematic during the 2016 ISMB meeting.

Funding

This work was supported by the Intramural Research Program of the National Library of Medicine at National Institutes of Health/DHHS. Comments, suggestions and questions are welcome and should be directed to: info@ncbi.nlm.nih.gov. Funding to pay the Open Access publication charges for this article was provided by the Intramural Research Program of the National Library of Medicine at National Institutes of Health/DHHS.

Conflict of Interest: none declared.

References

Bekker, G.J., Nakamura, H. and Kinjo, A.R. Molmil: a molecular viewer for the PDB and beyond. *J Cheminform* 2016;8(1):42.
Bradley, A.R., *et al.* MMTF-An efficient file format for the transmission, visualization, and analysis of macromolecular structures. *PLoS Comput Biol* 2017;13(6):e1005575.
Cho, Y., *et al.* Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* 1994;265(5170):346-355.
Craig, P.A., Michel, L.V. and Bateman, R.C. A survey of educational uses of molecular visualization freeware. *Biochem Mol Biol Educ* 2013;41(3):193-205.
Hanson, R.M., *et al.* JSmol and the Next-Generation Web-Based Representation of 3D Molecular Structure as Applied to Proteopedia. *Israel Journal of Chemistry* 2013;53(3-4):207-216.
Kim, S., *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res* 2016;44(D1):D1202-1213.
Landrum, M.J., *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016;44(D1):D862-868.
Li, H., *et al.* iview: an interactive WebGL visualizer for protein-ligand complex. *BMC Bioinformatics* 2014;15:56.
Madej, T., *et al.* MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Res* 2014;42(Database issue):D297-303.
Marchler-Bauer, A., *et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 2017;45(D1):D200-D203.
Rego, N. and Koes, D. 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* 2015;31(8):1322-1324.
Rose, A.S., *et al.* NGL Viewer: Web-based molecular graphics for large complexes. *Bioinformatics* 2018;doi: 10.1093/bioinformatics/bty419.
Rose, A.S. and Hildebrand, P.W. NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res* 2015;43(W1):W576-579.
Sehnal, D., *et al.* LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nat Methods* 2017;14(12):1121-1122.
Sherry, S.T., *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29(1):308-311.
Shi, M., Gao, J. and Zhang, M.Q. Web3DMol: interactive protein structure visualization based on WebGL. *Nucleic Acids Res* 2017;45(W1):W523-W527.
Sicheri, F., Moarefi, I. and Kuriyan, J. Crystal structure of the Src family tyrosine kinase Hck. *Nature* 1997;385(6617):602-609.
Vogelstein, B., Lane, D. and Levine, A.J. Surfing the p53 network. *Nature* 2000;408(6810):307-310.
Wang, Y., *et al.* Cn3D: sequence and structure views for Entrez. *Trends Biochem Sci* 2000;25(6):300-302.
Xu, D. and Zhang, Y. Generating triangulated macromolecular surfaces by Euclidean Distance Transform. *PLoS One* 2009;4(12):e8140.
Youkharibache, P. Twelve Elements of Visualization and Analysis for Tertiary and Quaternary Structure of Biological Molecules. *bioRxiv 153528* 2017;doi:10.1101/153528.