BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

NuProPlot: nucleic acid and protein interaction analysis and plotting program

Lagnajeet Pradhan and Hyun-Joo Nam*

Department of Bioengineering, University of Texas at Dallas, 800 W. Campbell Road RL10, Richardson, TX 75080, USA.
*Correspondence e-mail: hnam@utdallas.edu

Received 2 October 2014
Accepted 6 January 2015

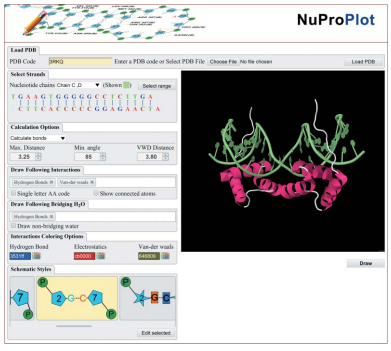
Keywords: *NuProPlot*; protein–nucleic acid complexes.

Growing numbers of protein and nucleic acid complex structures are being determined and deposited in the Protein Data Bank and the Nucleic Acid Database. With the increasing complexity of these structures, it is challenging to analyse and visualize the three-dimensional interactions. The currently available programs for such analysis and visualization are limited in their applications. They can only analyse a subset of protein–nucleic acid complexes and require multiple iterations before obtaining plots that are suitable for presentation. An interactive web-based program, *NuProPlot* (http://www.nuproplot.com), has been developed which can automatically identify hydrogen, electrostatic and van der Waals interactions between proteins and nucleic acids and generate a plot showing all of the interactions. Protein–DNA and protein–RNA interactions can be visualized in simple two-dimensional schematics. Interactive schematic drawing options allow selection of the plotted area and repositioning of the individual interactions for better legibility. *NuProPlot* is a fully automated and user-friendly program providing various custom options. *NuProPlot* represents a greatly improved option for analysis and presentation of protein–nucleic acid interactions.

1. Introduction

Protein and nucleic acid interaction is an essential part of many cellular activities such as transcription, translation, chromatin organization and regulation of gene expression. To help to understand the details of these vital processes, many studies have provided three-dimensional structures of molecules in action as complexes of protein and nucleic acids. Rapidly increasing numbers of protein and nucleic acid complex structures have been deposited in the wwPDB (Berman *et al.*, 2003), with the current count exceeding 4500. The three-dimensional structures contain a plethora of information regarding their functions and regulations. Through analysis and comparison of these structures, we can obtain an insight into their functions, which can be exploited for the development of treatment strategies for various diseases.

All DNA/RNA-binding proteins recognize or bind to nucleic acids using various interactions such as hydrogen bonding, hydrophobic, van der Waals and electrostatic interactions. Detailed analyses of such interactions are not always easy to perform owing to the large number of atoms and their complex arrangements. Specifically, individual amino acids responsible for nucleic acid sequence recognition and enzymatic manipulation can only be identified through the use of sophisticated three-dimensional graphics software. After analysis of the interactions, it is also important to present the results in a format that can be easily understood by scientists in various disciplines.



Previously, *NUCPLOT* has accomplished this through automated analyses of protein–DNA interactions and two-dimensional schematization of the interactions (Luscombe *et al.*, 1997). *NUCPLOT* uses *HBPLUS* to analyse protein–DNA interactions and generate a list of all of the hydrogen bonds and van der Waals interactions in a given protein–nucleic acid complex. *HBPLUS* performs H-atom positioning and hydrogen-bond calculation using an algorithm developed based on X-ray and neutron diffraction studies (McDonald & Thornton, 1994). Currently, many programs such as *PyMOL* (Schrödinger, New York, USA), *Accelrys Discovery Studio* (Accelrys, San Diego, California, USA) and *Jmol* (Willighagen & Howard, 2007) can also detect potential hydrogen bonds to generate such lists. *NUCPLOT* identifies atoms that belong to nucleic acids from a coordinate file and displays amino acids and solvent residues interacting with the nucleic acid atoms in two-dimensional schematics. Although bonding information can be easily obtained using various programs, *NUCPLOT* is the only program which has the ability to automatically generate two-dimensional schematics. It depicts interactions in protein–DNA complexes and, for simple cases, protein–RNA interactions as well. It is a command-line-based program, and to customize the output schematics the input file needs to be edited multiple times through cumbersome correction processes.

NUCPLOT cannot handle nucleic acids with complex secondary structures such as RNA cloverleafs and four-stranded DNA, and currently no software is available that can show protein–nucleic acid interactions in the context of complex RNA or DNA structures. Programs such as *RNAview* (Yang *et al.*, 2003), *VRNA* (Darty *et al.*, 2009), *loopDloop* (Gilbert, 1997) and *PseudoViewer* (Byun & Han, 2009) can show complex RNA secondary structures in two dimensions, but they lack the ability to plot interactions with proteins.

Here, we describe *NuProPlot* (http://www.nuproplot.com), a web-based automatic analysis and two-dimensional schematization program for protein–nucleic acid complexes. It can handle any type of nucleic acid ranging from a few base pairs of simple DNA to thousands of RNA bases in complex protein–RNA structures. It also provides all the advantages of a user-friendly web-based program with interactive options. *NuProPlot* is a versatile protein–nucleic acid interaction analysis and presentation tool.

2. Materials and methods

2.1. Program design

NuProPlot is written using the latest web technologies combined with several existing algorithms. It allocates the computational loads to both the server and the client's computer to speed up the process. Fig. 1 shows the components and structure of the program. It uses atomic coordinates of protein–nucleic acid complexes in PDB format as input. From the selected PDB file, the program first identifies nucleic acid residues and determines their base-pairing information using the server's computing machinery. The complex is then visualized using *Jmol* and the nucleic acid base-pairing information is displayed in the user interface (UI) window on the client's side (Willighagen & Howard, 2007; Fig. 1, Load). Alternatively, the program provides an option to use *JSMol* instead of *JMol* since Java is not enabled in many devices owing to security concerns (Hanson, 2013).

From the UI, the user selects ranges of nucleic acids for interaction analyses and the atomic coordinate subset is generated at the server based on the input. Protein atoms within a 4 Å radius of the nucleic acid atom subset are then selected and protein–nucleotide interactions are calculated. This greatly decreases the computation time, as unnecessary protein–protein and nucleotide–nucleotide interactions are excluded from the calculation. This calculation is carried out on the users' browser, which keeps the server from becoming overloaded and allows the users to run their calculation in real time from their own devices.

For the two-dimensional schematization, the server produces two-dimensional coordinates for DNA and RNA schematics. The program adopts the *ViennaRNA* package for the generation of two-dimensional schematics of RNA secondary structures (Lorenz *et al.*, 2011). These algorithms were written in C and loaded as PHP extensions on the server side to speed up the computation. *NuProPlot* generates the two-dimensional schematics of nucleic acids and adds the calculated protein–nucleic acid interactions to the plot. (Fig. 1,

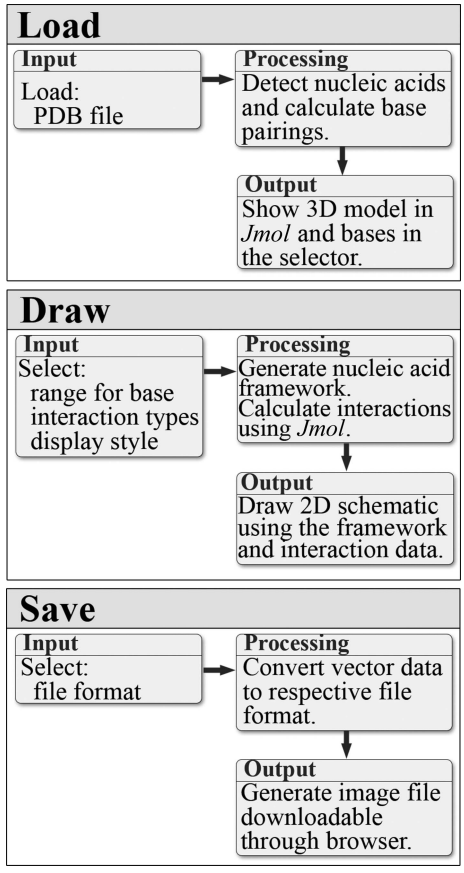


Figure 1
Program-design flowchart.

Draw). A user can customize the plot interactively at this stage, and the vector coordinate of the final schematic is sent to the server and converted to the image format selected by the user. The image file is then downloaded to the user's computing device (Fig. 1, Save). We used the latest HTML5 features along with the Ajax and jQuery Javascript framework for the user interface and for the generation of the schematic diagrams (Powell, 2008; Chaffer & Swedberf, 2010).

2.2. Nucleic acid–nucleic acid interactions

In *NuProPlot*, the base-pairing information of the nucleic acids is first determined before calculating the protein and nucleic acid interactions of a given PDB file. It is not always easy to obtain the base-pairing information of nucleic acids from PDB files. Relying solely on the chain label of the atoms to determine base pairing can result in errors, especially since RNAs frequently form base pairs within a single strand which can develop into complex conformations. For example, tRNAs and riboswitch RNAs fold into multi-lobed structures with complex base-pair patterns (Westhof & Fritsch, 2000). In addition, RNAs with pseudoknot structures must be identified from the coordinate files and handled accordingly (Wyatt *et al.*, 1989).

Among several excellent programs dealing with nucleic acid geometry, we have incorporated *RNAView* into *NuProPlot* for the detection of base-pairing information (Yang *et al.*, 2003). *RNAView* identifies all the base-pair and multi-strand interactions in RNA and DNA structures and sorts them into canonical and noncanonical base pairs (Yang *et al.*, 2003; Lu & Olson, 2008). The base-pairing information is depicted using dot–bracket notation (DBN) in the user interface of *NuProPlot* (Staple & Butcher, 2005). DBN can represent complex secondary-structure information of RNA as well as canonical DNA pairs. Figs. 2(*a*) and 2(*b*) show examples of DNA double-strand and RNA strand secondary structures represented in DBN in the *NuProPlot* user interface. The same nucleic acids are later plotted in two-dimensional schematic

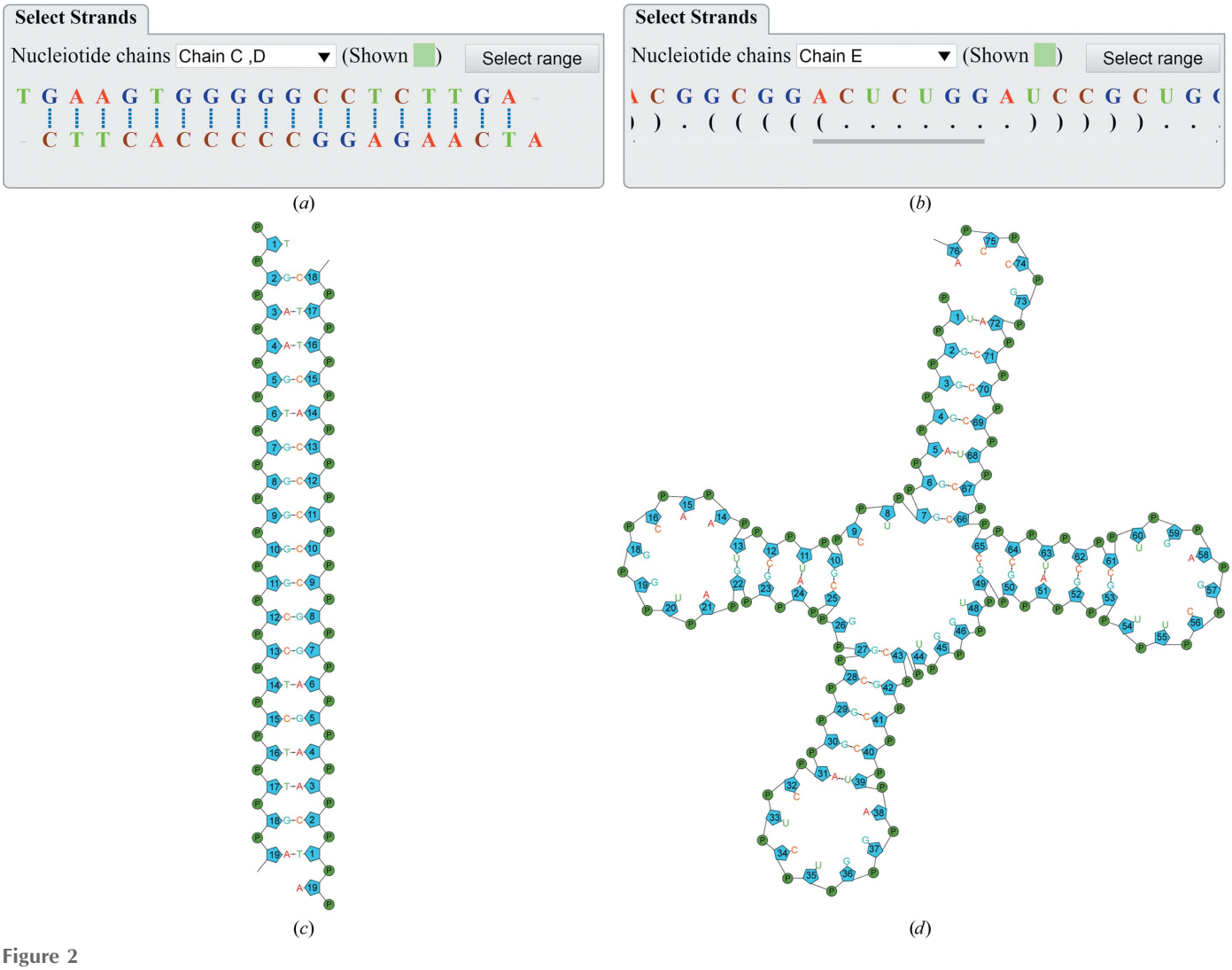


Figure 2
Schematic display of nucleic acids. Examples of DNA and RNA (in DBN notations) are shown in (*a*) and (*b*), respectively. The corresponding DNA and RNA are also shown as two-dimensional schematic diagrams in (*c*) and (*d*), respectively. The two-dimensional diagrams make it easier to visualize base-to-base interactions and secondary structures.

diagrams which make it much easier to visualize base-to-base interactions and the secondary structures (Figs. 2*c* and 2*d*).

We incorporated the RNA plotting algorithm described in the *ViennaRNA* package to generate the two-dimensional schematic for RNA secondary structure (Lorenz *et al.*, 2011). Some RNA structures are so complex that they clutter the two-dimensional schematics if not drawn properly. The current version of *NuProPlot* disregards RNA structures such as pseudoknots and kissing hairpins during the schematization process to make the final plot more legible. A subroutine was implemented to exclude the base pairing in these structures using the previously described 'nesting' convention for non-pseudoknotted structures (Rivas & Eddy, 1999).

Although DNA can also form complex secondary structures, identification of DNA base pairs is relatively easier than RNA. We used the same set of algorithms to determine the base-pairing information for DNA strands. Once the base-pairing information becomes available, *Jmol*/*JSMol* scripting is used to determine whether the strand is a DNA or an RNA, after which the base-pairing information is displayed accordingly in the selector panel (Figs. 2*a* and 2*b*).

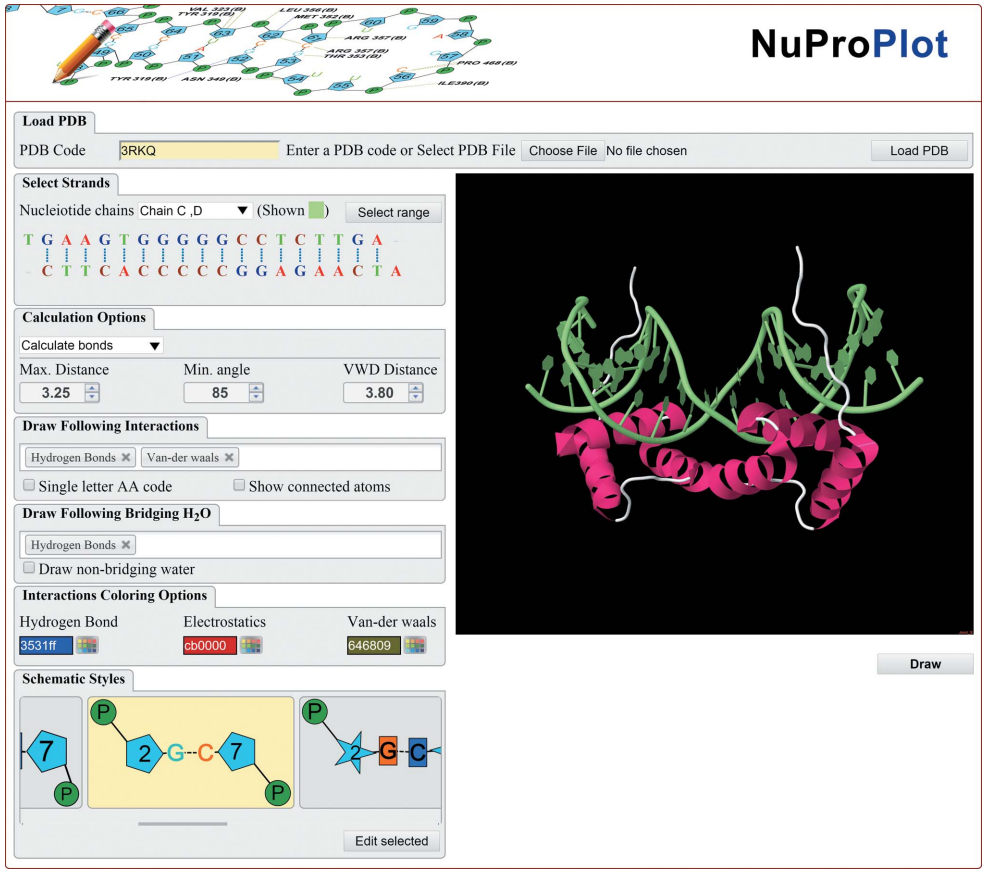


Figure 3
User-interface and display windows. The user-interface window is designed for custom option selections (left panel). The three-dimensional display of an example, the structure of the complex of NKX2.5 HD and DNA, is shown on the right.

2.3. Protein–nucleic acid interactions

To calculate the interactions between nucleic acids and proteins, *NuProPlot* first identifies protein atoms within a 4 Å radius and only calculates interactions between the nucleic acid and these protein atoms. This speeds up the calculation tremendously. The hydrogen-bond prediction algorithm used in *Jmol* is identical to *HBPLUS* but the H-atom placement algorithm is slightly different (Willighagen & Howard, 2007; McDonald & Thornton, 1994). While *HBPLUS* predicts the position of an H atom based on hybridization of electron orbitals, *JMol* uses the bond-angle and bond-length information (Momany *et al.*, 1975). This results in slightly different predictions between the two programs. Electrostatic interaction prediction is a novel addition in *NuProPlot*. It is determined by taking into account the distance between the two atoms and their polarity (Xu *et al.*, 1997). The nonpolar interactions among neighbouring atoms are categorized as van der Waals interactions and calculated based on the distance between two atoms. The cutoff distance for van der Waals radius can be defined by users.

In *NuProPlot*, the calculation of interactions runs separately from plot generation. Thus, users can take advantage of the automated bond-prediction software built into the program or use lists generated by other programs such as *HBPLUS* (McDonald & Thornton, 1994) or *PyMOL* (Schrödinger). Calculated interactions can then be used for plotting. The predicted interactions are shown as lines in the two-dimensional schematics, and various interactions are represented as distinct colored lines connecting the two respective groups.

2.4. User interface and custom options

The highly modularized nature and the separation of calculation and plotting allow the flexibility necessary to provide a variety of custom options in *NuProPlot*. Fig. 3 shows the user-friendly UI window conveniently designed for the selection of custom options. PDB-format coordinates can be uploaded from a file directly or accessed by simply typing a PDB code into the 'Load PDB' tab of the main window. When a PDB code is used, *NuProPlot* uploads the coordinate file as presented in the database, which is the content of an

asymmetric unit in most cases. In cases where the biologically functional complex spans multiple asymmetric units or unit cells, the coordinates of the whole biological unit need to first be generated by the user before it is uploaded into *NuProPlot*. Such functional assemblies can be identified using programs such as *PISA* (Krissinel & Henrick, 2007). Once the coordinate file has been uploaded, the structure is displayed using *Jmol*/*JSMol* on the right side of the browser and five additional tabs appear under the 'Load PDB' panel on the left side. These are the 'Select Strands', 'Calculation Options', 'Draw Following Interactions', 'Draw Following Bridging Interactions' and 'Schematic Styles' tabs, respectively.

The 'Select Strands' tab shows the nucleic acid chains and their base-pairing information. It allows the user to select only the nucleic acid region of interest from the protein–nucleic acid complex. Depending on whether the nucleic acid is DNA or RNA, the tab shows the base pairing as double-stranded DNA pairing or in dot–bracket notation, respectively. For DNA strands with overlapping pairing, *NuProPlot* calculates the transitivity relationship between the pairs. This allows the detection of chains in many complicated structures. For example, the model of catabolic gene-activator protein and DNA (PDB entry 1j59) has four DNA strands (*C*, *D*, *F* and *E*) with some overlaps, as shown below (Parkinson *et al.*, 1996).

```
CCCCCCCCCCCCCDDDDDDDDDDDDDDDD
FFFFFFFFFFFFFFEEEEEEEEEEEEEE
```

NuProPlot identifies the pairs and calculates transitivity, and the 'Chain Selector' lists multiple options for DNA chain pairs within the complex. They are listed as 'Chain C, F', 'Chain D, F', 'Chain D, E' and 'Chain C, D, E, F', and the user can simply select specific chains and base pairs of the nucleic acid for interaction analyses. After deciding on the chains, the user can select the range of base pairs for schematization. The 'Select range' tab shows a two-way range slider that can be used to set the start and end of the range.

'Calculation Options' allows the user to set parameters for hydrogen-bond calculation and the van der Waals distance (Fig. 3). The user can also upload interaction information calculated by *HBPLUS* or any other program of their choice using this tab (McDonald & Thornton, 1994). Types of interactions for presentation can be selected by the user from the drop-down menu in the 'Draw Following Interactions' section. Water molecules involved in bridging interactions between nucleic acid and protein can also be visualized through the 'Draw Following Bridging H_2O' tab.

The drawing options consist of two sections: 'Interactions Coloring Options' and 'Schematic Styles' (Fig. 3). The 'Interactions Coloring Options' helps the user to set the colors of different bonds such as hydrogen, electrostatic and van der Waals bonds. The schematic style section provides templates for different elements such as the phosphate backbone, sugar moiety and bases. It also features a schematic style editor, which can be invoked by clicking on the 'Edit Selected' button. The style editor can be used to customize the shape, size, color, font, bond length and bond angle of nucleic acid schematics. This is a unique feature of *NuProPlot* that grants greater freedom in schematic styles instead of forcing users to use only predefined styles. Several template shapes are provided for the various groups, but the user can use an SVG string to define any custom shapes. The editor panel allows the dragging and repositioning of the groups and bonds in a single nucleotide monomer, while the preview panel shows how a DNA or RNA molecule would be presented with the applied style. Once a style is defined, clicking on the apply button will add the newly created style to the already existing schematic styles list. The user-generated schematic styles can be saved as files on the user's computer for future use. This will ensure that all the schematics generated by a user will have an identical appearance.

2.5. Schematization

Once the calculation and drawing options have been selected, the schematization process is initiated by clicking on the 'Draw' button (Fig. 3). This opens up a new window with the two-dimensional schematic of the nucleic acid, DNA or RNA drawn according to the selected styling options. The program then starts plotting the protein–nucleic acid interactions based on the selected calculation options. In the plot, the individual protein residues and interactions can be dragged and repositioned to improve the legibility of the plot. Once the user is satisfied with the plot, it can then be saved as an image file in JPG, PNG or SVG format. The generated image files are of high resolution suitable for publication. Additionally, the SVG file has no resolution limit and can be opened in any graphics editor for additional editing.

3. Results

We present here two examples of *NuProPlot* schematics. The first example is the NKX2.5 homeodomain and DNA complex structure (PDB entry 3rkq). NKX2.5 belongs to the NK2 family of transcription factors and binds to DNA through a homeodomain (HD). The NKX2.5 HD is composed of three helices and an extended N-terminal arm. The interactions between the NKX2.5 HD and DNA are mediated mostly by residues from the N-terminal extension and the third helix. The NK2 family HDs recognize a TAAG motif, distinct from other HDs, which bind to TAAT. The Tyr54 conserved in NK2 family HDs interacts with the C residue complementary to the G in the motif, thus determining the specificity (Pradhan *et al.*, 2012). Fig. 3 shows the structure of the NKX2.5 HD in complex with its target DNA shown in a *Jmol* window. Using *NuProPlot*, the detailed interactions between the protein and DNA are plotted. The pull-down menu options shown in Fig. 3 are used to generated the draft plot (Fig. 4*a*). In this plot, labels may overlap to each other or present themselves on top of the DNA schematics, making it difficult to comprehend the schematics. After a brief session of interactive adjustment, the same plot can be modified for better viewing as presented in Fig. 4(*b*). The plot shows multiple van der Waals interactions between Tyr54 and Cys15.

The second example is a 'transamidosome'. In archaea and bacteria, glutamyl synthetase (GluRS) transfers Glu to both $tRNA^{Glu}$ and $tRNA^{Gln}$. Glu–$tRNA^{Gln}$ is then modified to Gln–$tRNA^{Gln}$ by the action of a tRNA-dependent amidotransferase. During the two-step process, a tight complex of GluRS and the tRNA-dependent amidotransferase with $tRNA^{Gln}$, referred to as a 'transamidosome', is maintained (Wilcox & Nirenberg, 1968; Curnow *et al.*, 1997; Bailly *et al.*, 2007; Huot *et al.*, 2011). In the bacterium *Thermotoga maritima*, heterotimeric GatCAB makes up the tRNA-dependent amidotransferase and structural studies of a 'transamidosome' show GluRS, $tRNA^{Gln}$ and GatCAB in a tight complex conformation (Ito & Yokoyama, 2010). When the coordinates of the 'transamindosome' (PDB entry 3al0) were analysed using the *PISA* program, two sets of GluRS, $tRNA^{Gln}$ and GatCAB complexes from two asymmetric units were identified as the probable biological unit (Ito & Yokoyama, 2010; Krissinel & Henrick, 2007). The two sets of complexes are related by twofold rotational symmetry, but only one amino-acid residue from the second set of complexes is within interacting distance of the $tRNA^{Gln}$. Further discussions on protein–nucleic interactions, therefore, are limited to one set of the GluRS, $tRNA^{Gln}$ and GatCAB complex.

Fig. 5(*a*) shows a three-dimensional schematic of a GluRS, $tRNA^{Gln}$ and GatCAB complex. The GluRS interacts with sections of the $tRNA^{Gln}$ which contain similar features in $tRNA^{Gln}$ and $tRNA^{Glu}$. L-Glutamyl-sulfamoyl adenosine (Glu-SA), a nonhydrolyzable analogue of glutamyl-AMP, is bound to the GluRS and is located near the acceptor arm of the $tRNA^{Gln}$. Among the GatCAB complex, only chain *B* responsible for the catalytic activity makes direct contacts with the $tRNA^{Gln}$ (Fig. 5*a*).

All of the contacts between the proteins and the tRNA can be analysed and visualized by *NuProPlot* almost instantly. The two-dimensional diagram of the 'transamidosome' in Fig. 5(*b*) shows an overall view of all the amino-acid residues contacting the tRNA. It shows that the tRNA makes extensive contacts with GluRS and to a lesser degree with the GatCAB *B* chain. This 'transamidosome' structure captures one stage of a multistep process. With the presence of Glu-SA, it shows that the GluRS molecule is in the productive form before the GatCAB molecule acts on the misacylated residue. The *NuProPlot* two-dimensional plot shows that GluRS (chain *G*) makes most contacts with the acceptor stem, anticodon loop and the stem section of the tRNA D-arm. The GatCAB *B* chain, on the other hand, interacts with the D loop and the stem section of the TψC arm. Sequence and conformational variation between $tRNA^{Gln}$ and $tRNA^{Glu}$ is pronounced on the D loop, specifically near U20 (Ito & Yokoyama, 2010). Contacts between the tRNA and GatCAB are concentrated

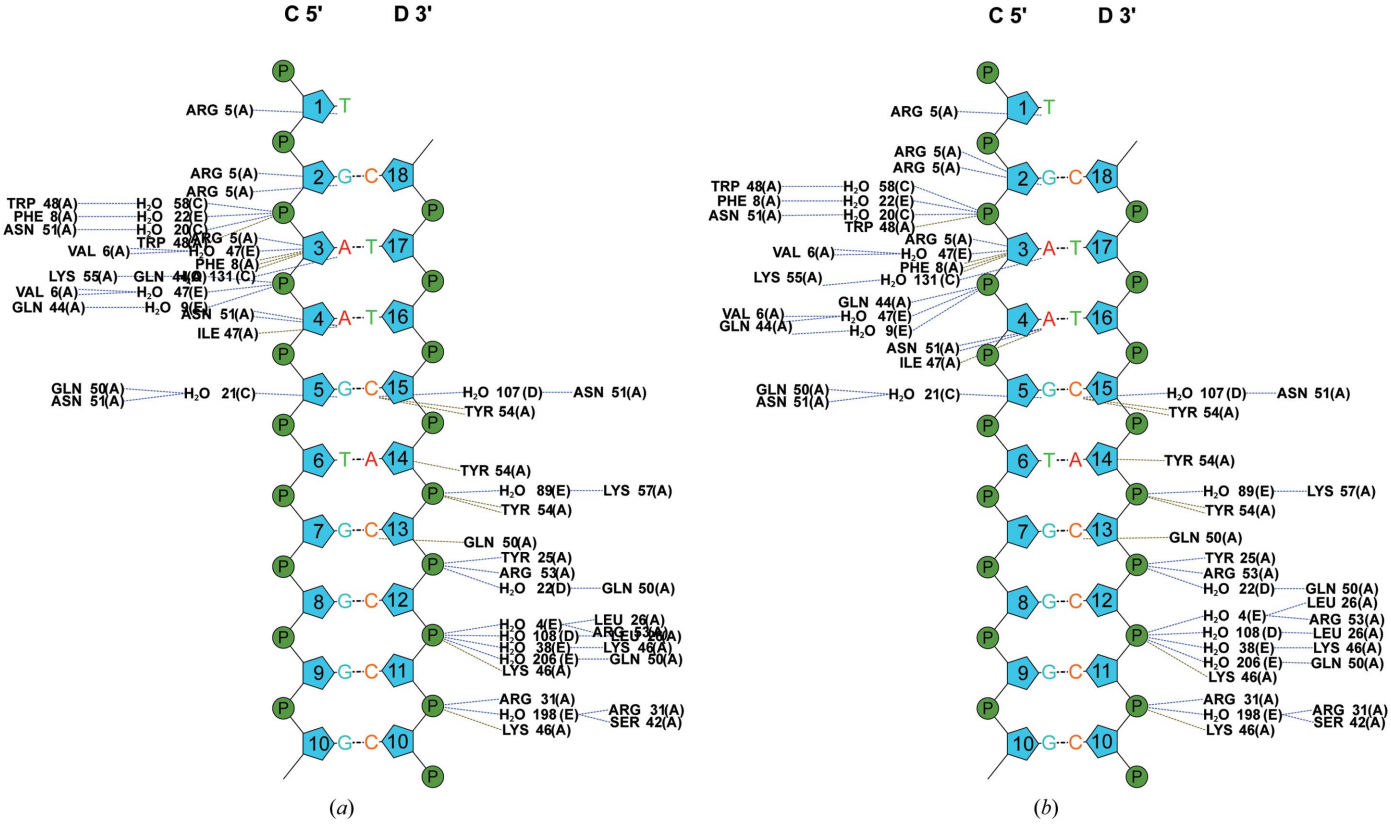


Figure 4
Examples of two-dimensional schematic diagrams generated using *NuProPlot*. (*a*) shows a plot generated using the NKX2.5 HD–DNA complex model (PDB entry 3rkq) before interactive adjustments. By adjusting the positions of interacting amino acids the legibility of the diagram is enhanced, as shown in (*b*).

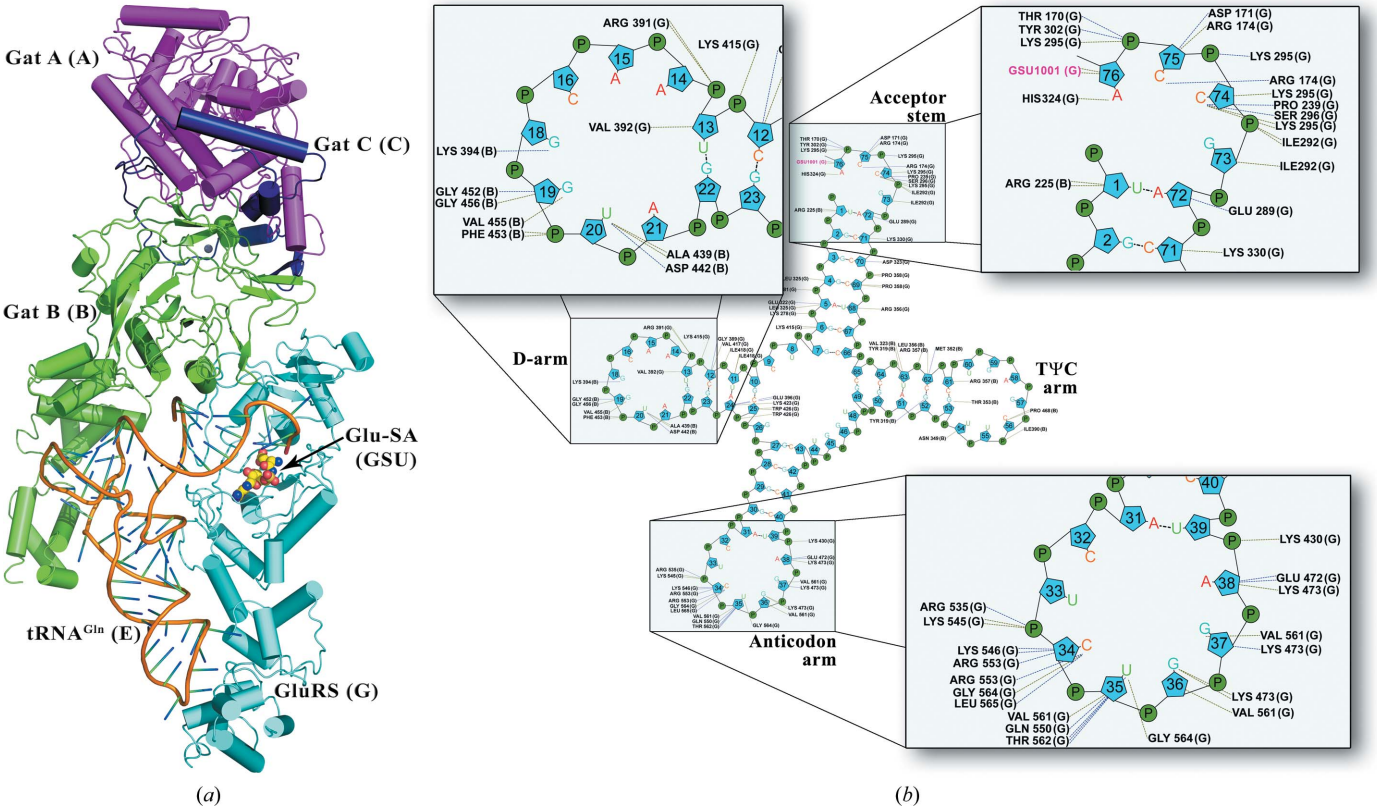


Figure 5
NuProPlot representation of a protein–RNA complex. (*a*) A three-dimensional model of the 'transamidosome' (PDB entry 3al0). Three protein chains of GatCAB and GluRS are shown with tRNA^{Gln}. The ID of each polypeptide chain is shown in parentheses. The figure was generated using *PyMOL* (Schrödinger). (*b*) *NuProPlot* diagrams of the protein–RNA complex. The two-dimensional diagram of tRNA^{Gln} is shown with interacting amino-acid residues. The boxed areas show enlarged views of the selected tRNA^{Gln} arms. Glu-SA is labelled in magenta.

around this region, which is likely to serve as a specificity-determining factor for GatCAB. The two-dimensional diagram also shows a contact site of Glu-SA near the acceptor arm (Fig. 5*b*).

4. Discussion

NuProPlot provides a powerful yet user-friendly tool to investigate and schematize protein–nucleic acid interactions. It can be used for protein–nucleic acid models generated by any experimental or computational methods. *NuProPlot* facilitates detailed analyses of interactions at the protein and nucleic acid interface and provides a two-dimensional schematic which can be saved as a high-resolution image file in several formats. Being an online program makes it more versatile and truly platform-independent. *NuProPlot* provides a convenient and advanced tool for researchers seeking to better understand protein–nucleic acid macromolecular complexes. Future improvement of the program will include the ability to specify hydrophobic interactions, the ability to draw the schematics horizontally and the ability to use pan and zoom functions to visualize larger schematics.

Acknowledgements

We thank Richard. T. Moore and Jankikeerthi Dharmarpandi for their critical reading of the manuscript.

References

Bailly, M., Blaise, M., Lorber, B., Becker, H. D. & Kern, D. (2007). *Mol. Cell*, **28**, 228–239.
Berman, H., Henrick, K. & Nakamura, H. (2003). *Nature Struct. Biol.* **10**, 980.
Byun, Y. & Han, K. (2009). *Bioinformatics*, **25**, 1435–1437.
Chaffer, J. & Swedberf, K. (2010). *jQuery 1.4 Reference Guide.* Birmingham: Packt Publishing.
Curnow, A. W., Hong, K., Yuan, R., Kim, S., Martins, O., Winkler, W., Henkin, T. M. & Söll, D. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 11819–11826.
Darty, K., Denise, A. & Ponty, Y. (2009). *Bioinformatics*, **25**, 1974–1975.
Gilbert, D. G. (1997). *loopDloop* v.2.02: *A Tool for Drawing RNA Secondary Structures in Molecular Biology*. Indiana University, Bloomington, USA.
Hanson, R. (2013). *JSMol*. http://wiki.jmol.org/index.php/JSmol#JSmol.
Huot, J. L., Fischer, F., Corbeil, J., Madore, É., Lorber, B., Diss, G., Hendrickson, T. L., Kern, D. & Lapointe, J. (2011). *Nucleic Acids Res.* **39**, 9306–9315.
Ito, T. & Yokoyama, S. (2010). *Nature (London)*, **467**, 612–616.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F. & Hofacker, I. L. (2011). *Algorithms Mol. Biol.* **6**, 26–40.
Lu, X.-J. & Olson, W. K. (2008). *Nature Protoc.* **3**, 1213–1227.
Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. (1997). *Nucleic Acids Res.* **25**, 4940–4945.

McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
Momany, F. A., McGuire, R. F., Burgess, A. W. & Scheraga, H. A. (1975). *J. Phys. Chem.* **79**, 2361–2381.
Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. H., Ebright, R. E. & Berman, H. M. (1996). *J. Mol. Biol.* **260**, 395–408.
Powell, T. A. (2008). *Ajax: The Complete Reference*. New York: McGraw–Hill.
Pradhan, L., Genis, C., Scone, P., Weinberg, E. O., Kasahara, H. & Nam, H.-J. (2012). *Biochemistry*, **51**, 6312–6319.
Rivas, E. & Eddy, S. R. (1999). *J. Mol. Biol.* **285**, 2053–2068.
Staple, D. W. & Butcher, S. E. (2005). *PLoS Biol.* **3**, e213.
Westhof, E. & Fritsch, V. (2000). *Structure*, **8**, R55–R65.
Wilcox, M. & Nirenberg, M. (1968). *Proc. Natl Acad. Sci. USA*, **61**, 229–236.
Willighagen, E. & Howard, M. (2007). *Jmol: An Open-Source Java Viewer for Chemical Structures in 3D. http://www.jmol.org/.*
Wyatt, J. R., Puglisi, J. D. & Tinoco, I. Jr (1989). *Bioessays*, **11**, 100–106.
Xu, D., Tsai, C.-J. & Nussinov, R. (1997). *Protein Eng.* **10**, 999–1012.
Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H. & Westhof, E. (2003). *Nucleic Acids Res.* **31**, 3450–3460.