The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes

G.C.P. van Zundert [1], M. Trellet [1], J. Schaarschmidt [1], Z. Kurkcuoglu [1], M. David [2], M. Verlato [3], A. Rosato [4] and A.M.J.J. Bonvin [1]

1 - Bijvoet Center for Biomolecular Research, *Faculty of Science – Chemistry, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands*
2 - LIP - Laboratório de Instrumentação e Física Experimental de Particulãs, *Avenida Elias Garcia 14, 1000 Lisbon, Portugal*
3 - Istituto Nazionale di Fisica Nucleare - Sezione di Padova, *Via Marzolo 8, 35131 Padova, Italy*
4 - Magnetic Resonance Center and Department of Chemistry "Ugo Schiff", *University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy*

Correspondence to A.M.J.J. Bonvin: *a.m.j.j.bonvin@uu.nl*
http://dx.doi.org/10.1016/j.jmb.2016.11.032
Edited by Michael Sternberg

Abstract

Structure determination of complex molecular machines requires a combination of an increasing number of experimental methods with highly specialized software geared toward each data source to properly handle the gathered data. Recently, we introduced the two software packages PowerFit and DisVis. These combine high-resolution structures of atomic subunits with density maps from cryo-electron microscopy or distance restraints, typically acquired by chemical cross-linking coupled with mass spectrometry, respectively. Here, we report on recent advances in both GPGPU-accelerated software packages: PowerFit is a tool for rigid body fitting of atomic structures in cryo-electron density maps and has been updated to also output reliability indicators for the success of fitting, through the use of the Fisher *z*-transformation and associated confidence intervals; DisVis aims at quantifying the information content of distance restraints and identifying false-positive restraints. We extended its analysis capabilities to include an analysis of putative interface residues and to output an average shape representing the putative location of the ligand. To facilitate their use by a broad community, they have been implemented as web portals harvesting both local CPU resources and GPGPU-accelerated EGI grid resources. They offer user-friendly interfaces, while minimizing computational requirements, and provide a first interactive view of the results. The portals can be accessed freely after registration via http://milou.science.uu.nl/services/DISVIS and http://milou.science.uu.nl/services/POWERFIT.

Introduction

As experimental methods for the structure determination of macromolecular complexes are getting ever more diverse, new computational tools are required to combine the resulting datasets and provide a consistent structural interpretation. Cryo-electron microscopy (cryo-EM) [1] and chemical cross-links coupled with mass spectrometry (CXMS) [2] are popular and powerful methods to elucidate the architecture of macromolecular complexes. Yet, the resulting data are of widely different nature: cryo-EM provides density information, where the information content depends on the resolution, while CXMS experiments provide upper bound distance restraints within and between subunits. Given the variety of experimental methods in the toolbox of structural biologists, highly specialized software is required to properly handle the resulting data and their information content.

Recently, we introduced PowerFit [3] and DisVis [4], two software packages that combine cryo-EM and CXMS data, respectively, with structural models. PowerFit is an integrative modeling tool, which aims to fit atomic structures as rigid bodies into the lower-resolution cryo-EM density maps using an objective six-dimensional cross-correlation search. PowerFit

Please cite this article as: G.C.P. van Zundert, et al., The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes, J. Mol. Biol. (2016), http://dx.doi.org/10.1016/j.jmb.2016.11.032

distinguishes itself from other fitting software by its speed and, more importantly, its sensitivity. PowerFit has been recently extended to include reliability measures and extensively benchmarked to determine fitting success rates as a function of cryo-EM map resolution and subunit sizes [5]. In contrast, DisVis aims to quantify the information content of intersubunit distance restraints, which are determined, for example, by using CXMS. It employs the concept of the accessible interaction space, the set of all possible complex conformations consistent with the provided data. In consequence, DisVis is rather an explorative than an integrative modeling tool. It samples millions to billions of conformations and counts all poses consistent with the data while providing additional global statistics to detect the presence of false-positive restraints through a bootstrapping technique.

Here, we describe the recent developments in PowerFit and DisVis, together with an overview of the new web server interfaces providing user-friendly access to the software. We first give a short description of both software packages and lay out the methodological advances in detail. After that, we describe the general workflow of the servers and how users can either run their jobs on our local CPU resources or benefit from GPGPU-enabled resources provided by the European Grid Infrastructure (EGI)[†] initiative.

The servers are freely available after registration via http://milou.science.uu.nl/services/DISVIS/ and http://milou.science.uu.nl/services/POWERFIT.

Overview and advances

PowerFit is an open source Python package and command line tool for automatic rigid body fitting of macromolecules in cryo-EM density maps[‡]. It performs a full, exhaustive six-dimensional cross-correlation search of the three translational and three rotational degrees of freedom to find the optimal placement of the provided atomic structure into the density. PowerFit uses fast Fourier transforms (FFTs) to accelerate the translational correlation scans while minimizing the required rotational sampling by its use of optimal rotation sets [6]. It further leverages available computational hardware by providing support for multiprocessor systems or even GPGPU resources through the OpenCL framework. In addition to its algorithmic optimization, PowerFit introduced the new sensitive core-weighted local cross-correlation score [3,7] in combination with the edge-enhancing Laplace pre-filter [8], further expanding the applicable resolution range to successfully fit a subunit in the density. Recently, we performed a comprehensive fitting analysis using PowerFit to ascertain reliability measures and success indicators [5]. We demonstrated that the Fisher *z*-transformation in combination with its associated confidence intervals allows to detect successful fittings [9,10]. The Fisher *z*-transformation is given by:

$$z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho},$$

where ρ is the cross-correlation value. The associated estimated error is defined as follows:

$$\sigma_z = \frac{1}{\sqrt{\frac{MV}{FC} - 3}},$$

where MV is the molecular volume of the fitted chain, and FC is the resolution of the map. In a previous investigation [10], we fitted 379 individual chains in 6 ribosome cryo-EM density maps starting at a resolution of 6 Å up to 30 Å. Successful fits were obtained in >99% of the cases for which the difference in error-normalized *z*-score ($\frac{z}{\sigma_z}$) between the two top-scoring solutions was larger than ~2.5. Conversely, in less than 3% of the failed cases did the difference exceed 1. We furthermore investigated the influence of the size and shape of the fitted chain on the success rate. As expected, fitting larger subunits is easier, consequently the required resolution for a successful fit is lower; surprisingly, however, the influence of the shape on the success rate was negligible. The latest version of PowerFit automatically includes the *z*-score analysis and provides for each solution the *z*-score, [Eq. (1), *z*], $\frac{z}{\sigma_z}$, and the difference in $\frac{z}{\sigma_z}$ to the best fit to easily estimate the reliability of the fits.

DisVis is another open source Python package and command line tool[§] using similar algorithmic techniques as PowerFit. While PowerFit is geared toward the modeling of high-resolution atomic structures into cryo-EM maps, DisVis was developed to explore the information content of distance restraints as provided. It quantifies and visualizes the accessible interaction space, that is, the set of all models of a binary complex consistent with the given data. To this end, DisVis also performs a six-dimensional exhaustive docking search of the ligand's translational and rotational degrees of freedom to sample millions to billions of conformations. It uses well-known FFT docking techniques by mapping the atomic structures onto grids [11], representing the receptor by its core and interaction shape, and the ligand solely by its core region. The volume of clashes (the overlap between the two core shapes) and interaction (overlap between the receptor core and ligand interaction shape—the outer shell of a molecule) can then be rapidly calculated using the cross-correlation theorem. A conformation is deemed a complex if its interaction volume is of sufficient size and its clashing volume is limited. The clashing and interaction parameters were determined using the Protein–Protein Docking Benchmark 4 [12], where, for each complex, the clashing and interaction volume was calculated using the superimposed unbound

Please cite this article as: G.C.P. van Zundert, et al., The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes, J. Mol. Biol. (2016), http://dx.doi.org/10.1016/j.jmb.2016.11.032

structures on the bound complex. For the bulk of the complexes, the clashing volume was lower than 200 Å^3 and its interaction volume larger than 300 Å^3, except for the most difficult cases where the conformational change associated with the binding event involved the displacement of whole domains. The number thus reflects a bare minimum requirement for a complex within the rigid body approximation. The regions of the interaction space that are consistent with the defined distance restraints can straightforwardly be determined using simple geometric arguments [4].

During the search, DisVis counts the number of complex conformations consistent with increasing numbers of distance restraints and also analyzes how often each restraint is violated. This allows to quantify the information content of the restraints while also providing insight into their consistency and reliability. In addition, DisVis outputs a discrete density map where each grid point represents the center of mass position of the ligand with respect to the receptor with an associated value corresponding to the maximum number of consistent restraints found at that particular position. This allows to visualize the space in which the ligand can be positioned with respect to the receptor based on a defined number of distance restraints. Although DisVis uses well-known techniques present in macromolecular docking, its aims and output deviate from typical docking software. While docking aims at predicting the structure of the complex, DisVis' main aim is to explore the information content and consistency of the distance restraints and to visualize the accessible interaction space defined by these. We have thus dubbed the different paradigm followed by DisVis as *explorative modeling*.

In the updated DisVis version, we have extended the analysis capabilities: DisVis now also provides an average occupancy grid and can infer potential interface residues from a statistical analysis of the consistent models. In the following, we first present the mathematical intricacies to obtain the average occupancy grids. Starting from the receptor core (R_{core}) and interaction (R_{inter}) shapes, and the ligand core (L_{core}) shape, the clashing (C) and interacting (I) volume for a particular rotation can be rapidly calculated via the cross-correlation theorem and FFT algorithm:

$$C = F^{-1}[F(L_{core})^* \times F(R_{\text{core}})]$$
$$I = F^{-1}[F(L_{core})^* \times F(R_{\text{inter}})]$$

where F is the FFT. The accessible interaction space, that is, the translations where a conformation is considered a valid complex, is defined as follows:

$$A(\vec{r}) = \begin{cases} \text{if } C(\vec{r}) \leq C_{\max} \wedge I(\vec{r}) \geq I_{\min} : 1 \\ else : 0 \end{cases},$$

where $C_{\max}$ and $I_{\min}$ are the maximum allowed clashing volume and minimum required interaction volume, respectively. The reduced accessible interaction space for the rotation is defined by the product of the accessible interaction space and the distance restraint space consistent with at least N restraints L_N [4]:

$$A_{,N} = A \times L_N$$

Next, the occupancy of the ligand molecule at each grid point considering all data-consistent translations for a particular rotation is then given by the convolution of the ligand core shape and the reduced accessible interaction space, that is:

$$O_N = F^{-1}[F(L_{\text{core}}) \times F(A_N)]$$

The average occupancy grid is the weighted summation of the translational occupancy grids over all rotations, normalized by the number of accessible complexes consistent with at least N restraints (N_A):

$$\overline{O}_N = \frac{\sum_P w_R O_{N,R}}{N_A},$$

where the summation is over all rotations P, and w_R is the weight factor of rotation R to properly average over the rotation space [6]. The values at each grid point can then be interpreted as the probability of that grid point being occupied by the ligand given the required distance restraints consistency. The occupancy grids are outputted in MRC format, a density map format used, for example, in cryo-EM. They can easily be inspected with molecular viewers, such as UCSF Chimera [13]. The average occupancy grids thus provide a proxy for the binding mode of the ligand and indicate the preferred occupied regions of space. It comes, however, at a cost, as all atomic information of the ligand is lost and now replaced by an average shape.

Next to the average occupancy grids, DisVis can now also infer interface residues from the reduced accessible interaction space through the concept of the average number of interactions per complex (AIC) for each residue, defined as follows:

$$\text{AIC} = \frac{\sum_P w_R \sum_{C_R} I_C}{N_A},$$

where the first summation is over all orientations P indexed by R, and w_R is the weight of each rotation to properly average over the orientation space; the second summation is over all complexes consistent with the distance restraints at rotation R indexed by C, I_C is the number of interactions a particular residue forms in complex C, and N_A is the total number of complexes consistent with at least N restraints. For this, DisVis requires, as an additional input, the residue numbers of all solvent-accessible residues for both

Please cite this article as: G.C.P. van Zundert, et al., The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes, J. Mol. Biol. (2016), http://dx.doi.org/10.1016/j.jmb.2016.11.032

receptor and ligand. During the rotational search, DisVis counts the number of interactions each selected receptor residue makes with each selected ligand residue for complexes consistent with at least N restraints through a brute force method, that is, it calculates all distances between the two residue sets. However, this approach can be computationally demanding, as the number of distances that need to be calculated scales with $N \times M$, where N and M are the number of atoms in the selected ligand and receptor residue set. To reduce the computational costs, DisVis only considers CA or O3′ atoms for proteins and nucleic acids, respectively. An interaction is then defined as a distance smaller or equal to 10 Å. This is done (by default) only for complexes consistent with all distance restraints, up to complexes violating, at most, three restraints. The total number of interactions each selected residue makes is ultimately normalized by the total number of accessible complexes consistent with at least N restraints. The output is the AIC of each selected residue, where highly accessed residues in the reduced accessible interaction space are more likely to be at the interface. This information is particularly valuable, for example, to select residues for mutagenesis studies and the identification of hot spots or to drive the modeling of a complex in our information-driven docking approach HADDOCK [14,15].

Web portals

PowerFit and DisVis both have significant requirements in terms of computational resources. Furthermore, a local installation, especially one harvesting the power of GPGPUs, is non-trivial and time-consuming. To lift these hurdles and provide user-friendly access to both software packages, we developed two new web portals that grant access to large-scale resources, either through local multithreading calculations on our CPU cluster or through the GPGPU capabilities of the EGI. The use of these web portals eliminates the need for any local software installation and guarantees a robust user experience through a reliable framework.

The portals have been built on the open source Flask framework[‖] where different modules handle the different components of the web servers:

- User/admin mailing system (Flask-Mail)
- Authentication through SQL database usage (Flask-SQLAlchemy)
- Forms creation and validation (Flask-WTF)
- HTML/CSS/JavaScript framework (Flask-Bootstrap)
- Admin management (Flask-Script)

The web portals are hosted within our lab facilities and communicate with either a grid User Interface (UI) using gLite [16] commands to submit to the GPGPU-enabled grid sites of EGI or a local CPU cluster. They are accessible from our main portal[¶].

The web portals' front-end consists of five main sections:

(1) A home page with basic information about the software
(2) A submission page with an input form
(3) A registration page
(4) An exemplary result page
(5) A help page with a detailed description of the expected input (parameters/files) and provided output

Workflow

The workflow behind the DisVis and PowerFit web portals is summarized in Fig. 1. The submission page holds the form with mandatory and optional parameters (the latter are prefilled with default values), which are sent to our local server. Input and output of DisVis and PowerFit will be described in the following two sections (see below). Once the submission form is correctly filled and submitted by the user, a validation process takes place to verify any provided input. At this step, only basic syntax and type errors are verified through Flask-WTF validators for common fields and home-made validators for the software-related files and parameters (e.g., PDB files or chain selection). Credentials are sent to an internal database to check if the user is already registered. Authentication will be extended in the future to make use of Single Sign-On possibilities, allowing users from recognized institutions (e.g., INSTRUCT[††]) to use the web portals. If any error is detected or if the user e-mail address is unknown, the form is sent back to the submission page and a dynamic message with an informative error is displayed next to the problematic field. In this way, the user is alerted within seconds of any wrongly filled field or problems with input files. The pre-processing/validation stage thus allows a smooth operation of the servers.

If no error is detected upon form submission, the files and parameters are copied and processed by the master node to create a job package. In the case of a local submission, input files and the job package are sent to a working node through a local batch system. The online progress page is constantly updated by parsing the job's status.

In the case of submission to the grid, input files are sent to a local grid UI at the Utrecht LSG-BCBR grid site managed by the Dutch NGI via SURFsara. The UI handles the submission, monitoring, and retrieval of the grid jobs using custom scripts making use of the gLite middleware. Output files are automatically sent to our server when a job is finished. The computations on the grid are executed via Docker containers customized with all the required software dependencies for DisVis and PowerFit and enabled for GPGPU access[‡‡].

Please cite this article as: G.C.P. van Zundert, et al., The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes, J. Mol. Biol. (2016), http://dx.doi.org/10.1016/j.jmb.2016.11.032

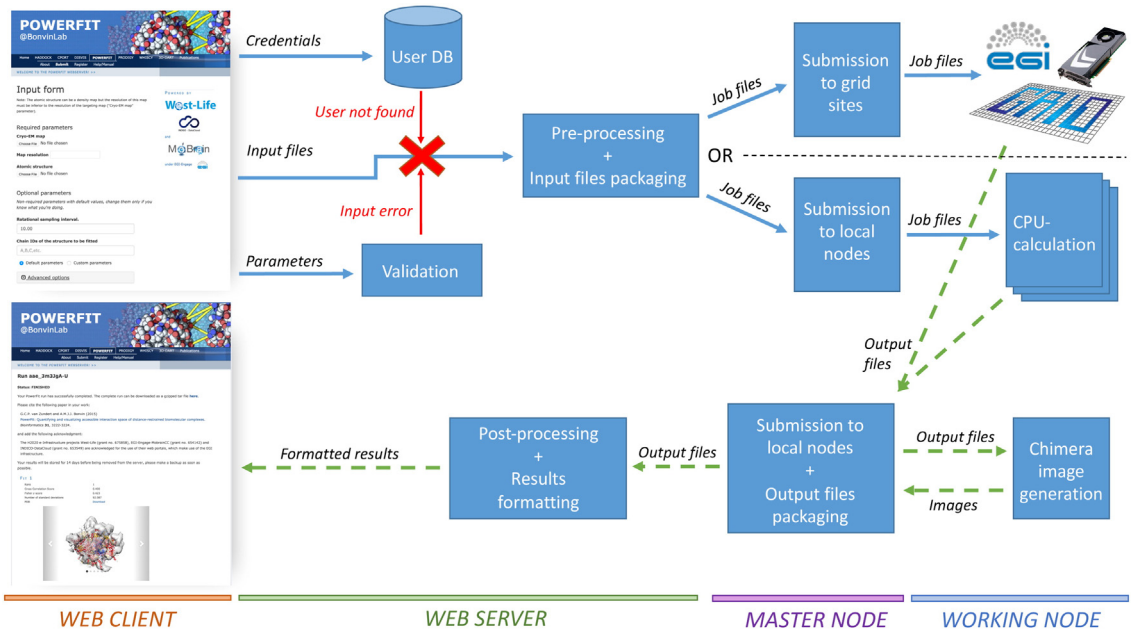

Fig. 1. Overview of the web portals' workflow connecting the web clients to the working nodes. The blue connectors report the input process from the form's parameter submission to the software execution. The green dashed connectors report the output process starting from the output files' retrieval up to the display of the results.

For both local and grid submission cases, a unique results directory is generated on the server upon completion, and the output files are then post-processed by the web server. Depending on the status of a finished job, different result pages will be presented to the user. In case of failure, the log files of the concerned job are made available. In case of success, the resulting output files are post-processed and the results presented in a user-friendly manner. Next to the text display of the generated tables, a graphical representation of the accessible interaction space allowing the direct online visualization by the end user is displayed as well (for details, see below).

The average run times of sample PowerFit and DisVis runs with various settings comparing the performance on local CPU and grid GPGPU resources are summarized in Table 1. Pre- and post-processing are handled within a few minutes; however, the wall time of specific jobs might be significantly affected by the load of the selected resource.

PowerFit input and output

The PowerFit web server exposes in its input form the same parameters as those in the command line application with the exception of the number of CPU processors, the GPGPU flag, and the results directory name. These are defined automatically by the web server to match the selected hardware resources and the local directory structure of the server. Among the 12 exposed parameters, 3 of them are mandatory and 9 are optional (set to default values). The three required parameters/files are: (1) the cryo-EM map in MRC or CCP4 format, (2) its associated resolution, and (3) a PDB or mmCIF file with the atomic coordinates of the structure to be fitted into the map. The optional advanced parameters group comprises (1) the rotational sampling interval (10.0 degrees by default, can be increased to reduce the rotational space sampling and speed up the calculation), (2) chain selection by ID (the whole structure is considered by default), (3) the possibility to disable the Laplace pre-filter, (4) the possibility to disable the core-weighted scoring, (5) the number of solutions to be written to PDB format (10 by default), (6) the possibility to disable the density map resampling, (7) the resampling rate (2.0 by default, not used if the density map is not resampled), (8) the possibility to disable the trimming of the density map, and (9) the trimming cutoff to be used for the density map trimming (by default, 10% of the maximum intensity value found in the density). These parameters can be simply left to their default values or customized by the user to suit his/her needs.

The default parameters are chosen for optimal scoring function sensitivity with both the Laplace pre-filter and core-weighted correlation function activated. Through resampling and trimming the map in combination with a modest rotational sampling density, the computational cost remains at an acceptable level. In cases where the PowerFit results are unsatisfactory, the user is advised to explore using different combinations of the scoring function, for example, without both the Laplace pre-filter and core-weighted correlation function or by increasing the rotational sampling to 5° (denser sampling has shown no substantial benefit). The user can furthermore adjust the trimming behavior

Please cite this article as: G.C.P. van Zundert, et al., The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes, J. Mol. Biol. (2016), http://dx.doi.org/10.1016/j.jmb.2016.11.032

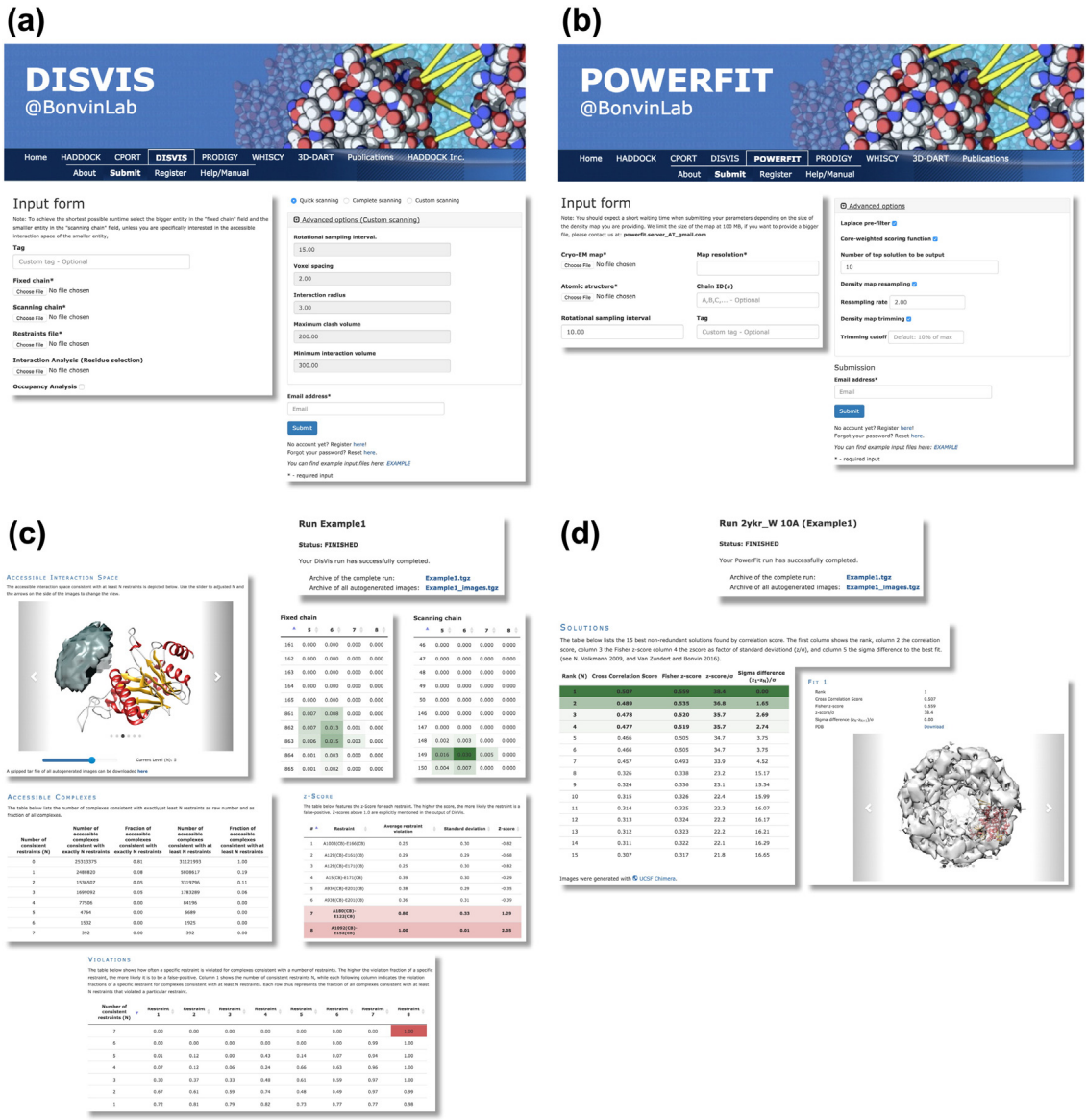

Fig. 2. Excerpts from the submission (top) and results (bottom) pages of (a and c) DisVis and (b and d) PowerFit.

Table 1. Average run times and standard deviation of sample PowerFit and DisVis runs in minutes

	PowerFit		DisVis			
	Rotational Sampling		"Quick scanning"	"Complete scanning"		
	10°	5°		–	Occupancy analysis	Occupancy + Interface analysis
GPGPU (grid)	5.2 ± 0.0	11.3 ± 0.4	4.3 ± 0.7	17.4 ± 1.9	45.7 ± 12.2	48.7 ± 14.6
CPU (local)	9.4 ± 3.5	68.7 ± 19.8	2.4 ± 0.6	105.0 ± 1.9	266.3 ± 28.0	288.5 ± 23.9

Average run times were calculated based on five runs for each group. For PowerFit, chain C of the GroEL-ES complex (PDB: 3zpz, 525 residues) was fitted into the corresponding map (EMD-2325; resolution: 8.9 Å, estimated volume: 936 nm^3) with a rotational sampling interval of 5° and 10°, respectively. For DisVis, the analysis was performed with the RNA polymerase II complex of *Saccharomyces cerevisiae* (1WCM; chain A—1416 residues, and chain E—215 residues) with the six experimental and two false-positive restraints previously described [4]. For the runs, four different settings were employed: (1) quick scanning, (2) complete scanning, (3) complete scanning with occupancy analysis, and (4) complete scanning with occupancy analysis, and interface analysis for 10 fixed chain and 10 scanning chain residues.

Please cite this article as: G.C.P. van Zundert, et al., The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes, J. Mol. Biol. (2016), http://dx.doi.org/10.1016/j.jmb.2016.11.032

of PowerFit in cases where the default trimming options might be too aggressive with important regions being trimmed from the provided map. An example archive containing a cryo-EM map and a PDB file is available for download.

Once a job has successfully completed, its output files are processed to be displayed in a user-friendly way. A tar archive of the job can be downloaded with the different input and output files. A post-processing script based on Chimera [13] is used to generate images of the top N (maximum of 10) best solutions generated by PowerFit. For each model, images of the structure fitted into the cryo-EM density map are generated with Chimera for six different views of the molecular scene: top, bottom, front, back, right, and left. The user can switch between images through the use of a Bootstrap carousel view. Due to the computing time of the Chimera script, image generation is decoupled from the main post-processing, and images are not displayed until this step has been completed. Therefore, the results page reloads every 30 s until the images are available and a short message about the image generation in progress is displayed. For each of the top N solutions, the following statistics are displayed: their rank, cross-correlation score, Fisher *z*-score (*z*), the *z*-score divided by the estimated error $(\frac{z}{\sigma_z})$, and the difference in $\frac{z}{\sigma_z}$ to the best fit. The latter is given to better appreciate the reliability of the fit: if the difference in $\frac{z}{\sigma_z}$ between the best and the second best fit is greater than ~1.5, it can be considered a reliable fit (note that in case of symmetry, multiple solutions might have similar *z*-scores); conversely, if the difference is below 1 for multiple solutions, all of these should be critically inspected and additional experimental information in conjunction with integrative approaches need to be applied to ascertain a reliable model. A link is also provided to download separately the PDB file of each solution. A summary table sorted according to the scores allows for a quick comparison of the 15 best solutions by cross-correlation score (see Fig. 2).

DisVis input and output

Similar to PowerFit, the DisVis input form takes the same parameters as the command line version with the exception of the number of CPU processors, the GPGPU flag, and the results directory name as well. All exposed parameters can be tuned by filling the different fields of the form. Yet, only three parameters are required and do not have default values: the fixed chain structure file (usually the larger of the two molecules), the scanning chain structure file, and the restraints file. The restraints must follow a strict syntax illustrated below with an example:

<chainid 1>	<resid 1>	<atomname 1>	<chainid 2>	<resid 2>	<atomname 2>	<mindis>	<maxdis>
A	18	CA	F	27	CB	10.0	20.0

The seven remaining parameters can be changed via three distinct modes: two modes have predefined values allowing either for a “complete scanning” of the fixed chain, which corresponds to the default values of DisVis, or a “quick scanning”, which substantially increases the speed of the search at the cost of its accuracy [4]. A summary of the parameters together with their respective values with respect to the mode is described in Table 2. The third mode, called “custom scanning”, allows the user to modify the parameters to his/her own convenience. The last two parameters in Table 2 are specific to DisVis-2.0. The Interaction analysis will report the frequency at which the provided putative interface residues are observed at the interface based on the reduced accessible interaction space (see above). The occupancy analysis will generate four volume files giving a normalized indication of how frequent a grid point is occupied by the ligand for at least a given number of consistent restraints N with N varying between *max* (the total number of restraints provided by the user) and *max*-3. Since the frequencies encoded in the volume file have a continuous scale and the discretization of such a scale without extra information is complicated, we chose to skip the visualization of these files to lighten the results' content. They are, however, present in the archive file available to the user.

Again, Chimera is used to visualize the results by generating the images of the available interaction space density map at all levels. The density represents the center of mass of the scanning chain consistent with N restraints at every position in space. Six views (front, back, top, bottom, left, and right) of the fixed chain represented as cartoon and the density map as volume are generated for each level. A slider allows the users to change the level of N, allowing again a fast initial visualization of the results online. Plain text output files are displayed as interactive tables of fixed size allowing for a quick navigation between the results. A short description of the respective output is located at the top of each table. Putative false-positive restraints are highlighted in the *z*-score and violations table if no complex consistent with all restraints exists (see Fig. 2).

A typical DisVis investigation should start with determining the consistency of the restraints dataset using the “quick scanning” option. After inspecting the results, the dataset can be filtered to discard false-positive restraints using the violation analysis. When the consistency of the dataset has been secured, additional runs can be performed with the “complete scanning” setting and optionally targeted interface analysis. With the resulting information, interface residues can be inferred and, through the occupancy

Please cite this article as: G.C.P. van Zundert, et al., The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes, J. Mol. Biol. (2016), http://dx.doi.org/10.1016/j.jmb.2016.11.032

Table 2. Summary of the DisVis parameters from the two pre-defined search modes

Parameter	Complete scanning	Quick scanning
Rotational sampling interval (degrees)	9.72	15.0
Voxel spacing (Å)	1.0	2.00
Interaction radius (Å)	3.0	3.0
Maximum clash volume (Å^3)	200.0	200.0
Minimum interaction volume (Å^3)	300.0	300.0
Residue selection counting for interactions	None	None
Occupancy analysis?	Yes	No

analysis, the putative binding pose of the ligand approximated. These insights are of use in determining future experiments to further characterize the interaction.

Conclusions

We have described in this article the recent advances in PowerFit and DisVis and the new possibilities offered by their web interface. In its latest version, PowerFit now provides an assessment of the reliability of the fit through a *z*-score and standard deviation analysis. DisVis has been extended to generate an average occupancy grid and to infer the interface residues from a statistical analysis of the models consistent with the restraints. Both software packages are now freely available through user-friendly web portals providing access to the GPGPU resources of EGI, facilitating their use by entry-level users who can easily set up a job without having to worry about any installation procedure. Furthermore, the servers, through their post-processing, allow for initial online visualization and analysis of the results. These developments should greatly facilitate the use of DisVis and PowerFit by end users and open the route for their implementation in workflows.

Acknowledgments

This work is supported by grants from the Netherlands Organization for Scientific Research (NWO; ECHO grant no. 711.011.009) and by European H2020 e-Infrastructure grants (EGI-Engage, grant no. 654142; INDIGO-DataCloud, grant no. 653549; West-Life grant no. 675858; and BioExcel grant no. 675728). Andrea Giachetti (Centro di Risonanze Magnetiche, University of Florence) is acknowledged for technical support on GPGPUs. The EGI infrastructure with the dedicated support of CESNET-MetaCloud, INFN-PADOVA, NCG-INGRID-PT, RAL-LCG2, TW-NCHC, SURFsara, and NIKHEF, and the additional support of the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, South Africa, Malaysia, Taiwan, and the US Open Science Grid are acknowledged.

Received 1 September 2016;
Received in revised form 29 November 2016;
Accepted 30 November 2016
Available online xxxx

Keywords:
cryo-electron microscopy;
chemical cross-links coupled with mass spectrometry;
fast Fourier transform;
GPGPU acceleration

†www.egi.eu
‡https://github.com/haddocking/powerfit
§https://github.com/haddocking/disvis
||http://flask.pocoo.org/
¶http://milou.science.uu.nl
††www.structuralbiology.eu
‡‡https://github.com/indigo-dc/docker-disvis/
https://github.com/indigo-dc/docker-powerfit

Abbreviations used:
cryo-EM, cryo-electron microscopy; CXMS, cross-links coupled with mass spectrometry; EGI, European Grid Infrastructure; FFT, fast Fourier transform; AIC, average number of interactions per complex; UI, User Interface.

References

[1] X.-C.B. Bai, G. McMullan, S.H.W. Scheres, How cryo-EM is revolutionizing structural biology, Trends Biochem. Sci. 40 (2015) 49–57.
[2] A. Leitner, M. Faini, F. Stengel, R. Aebersold, Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines, Trends Biochem. Sci. 41 (2016) 20–32.
[3] G.C.P. Van Zundert, A.M.J.J. Bonvin, Fast and sensitive rigid-body fitting into cryo-EM density maps with PowerFit, AIMS Biophys. 2 (2015) 73–87.
[4] G.C.P. Van Zundert, A.M.J.J. Bonvin, DisVis: quantifying and visualizing the accessible interaction space of distance-restrained biomolecular complexes, Bioinformatics 31 (2015) 3222–3224.
[5] G.C.P. Van Zundert, A.M.J.J. Bonvin, Defining the limits and reliability of rigid-body fitting in cryo-EM maps using multi-scale image pyramids, J. Struct. Biol. 195 (2016) 252–258.
[6] C.F. Karney, Quaternions in molecular modeling, J. Mol. Graph. Model. 25 (2007) 595–604.
[7] X. Wu, J.L. Milne, M.J. Borgnia, A.V. Rostapshov, S. Subramaniam, B.R. Brooks, A core-weighted fitting method

for docking atomic structures into low-resolution maps: application to cryo-electron microscopy, J. Struct. Biol. 141 (2003) 63–76.
[8] P. Chacón, W. Wriggers, Multi-resolution contour-based fitting of macromolecular structures, J. Mol. Biol. 317 (2002) 375–384.
[9] R.A. Fisher, On the "probable error" of a coefficient of correlation deduced from a small sample, Metron 1 (1921) 1–32.
[10] N. Volkmann, Confidence intervals for fitting of atomic models into low-resolution densities, Acta Crystallogr. D Biol. Crystallogr. 65 (2009) 679–689.
[11] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, I.A. Vakser, Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 2195–2199.
[12] H. Hwang, T. Vreven, J. Janin, Z. Weng, Protein–protein docking benchmark version 4.0, Proteins 78 (2010) 3111–3114.
[13] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[14] C. Dominguez, R. Boelens, A.M.J.J. Bonvin, HADDOCK: a protein–protein docking approach based on biochemical or biophysical information, J. Am. Chem. Soc. 125 (2003) 1731–1737.
[15] G.C.P. Van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries, A.M.J.J. Bonvin, The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes, J. Mol. Biol. 428 (2016) 720–725.
[16] T. Ferrari, L. Gaido, Resources and services of the EGEE production infrastructure, J. Grid Comput. 9 (2011) 119–133.

Please cite this article as: G.C.P. van Zundert, et al., The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes, J. Mol. Biol. (2016), http://dx.doi.org/10.1016/j.jmb.2016.11.032