

Scipion web tools: Easy to use cryo-EM image processing over the web

Pablo Conesa Mingo, José Gutierrez, Adrián Quintana, José Miguel de la Rosa Trevín, Airén Zaldívar-Peraza, Jesús Cuenca Alba, Mohsen Kazemi, Javier Vargas, Laura del Cano, Joan Segura, Carlos Oscar S. Sorzano, and Jose María Carazo*

Centro Nacional de Biotecnología (CNB-CSIC), Cantoblanco, Madrid, 28049, Spain

Received 31 July 2017; Accepted 28 September 2017

DOI: 10.1002/pro.3315

Published online 3 October 2017 proteinscience.org

Abstract: Macromolecular structural determination by Electron Microscopy under cryogenic conditions is revolutionizing the field of structural biology, interesting a large community of potential users. Still, the path from raw images to density maps is complex, and sophisticated image processing suites are required in this process, often demanding the installation and understanding of different software packages. Here, we present Scipion Web Tools, a web-based set of tools/workflows derived from the Scipion image processing framework, specially tailored to nonexpert users in need of very precise answers at several key stages of the structural elucidation process.

Keywords: cryo-electron microscopy; web tools; validation; visualization; workflow

Introduction

The field of cryo-Electron Microscopy (cryo-EM) is going through a technological “revolution” that is making possible the direct observation of noncrystalline, flexible and short-lived macromolecular complexes to quasi atomic resolution.¹ The interest that these new structural capabilities has raised is enormous, attracting to the field a whole new range of researchers demanding simple to use yet very efficient image processing software. Nowadays, there

are many good cryo-EM image processing packages, but all of them need to be installed, either in the researcher’s computer or in the local HPC facility. HPC is required in many cases, since some workflows require far more computing power than it is available on a laptop or desktop computer.

On the other hand, the availability of remote web tools is starting to be a reality in cryo-EM image processing, and it is certainly already very popular in Bioinformatics. In fact, there is a complete yearly issue of Nucleic Acid Research devoted to bioinformatics web applications, clearly conveying the importance for that field of the existence of easily accessible web tools. These web applications, referred to in a more general way as web tools, are usually focused on a specific workflow that is considered to be of general interest (for example, a simplified version of a more complex processing approach). Input and output parameters are simple and the real processing takes place in a remote server. This

Grant sponsor: Spanish Ministry of Economy and Competitiveness; Grant numbers: AIC-A-2011-0638 and BIO2016-76400-R; Grant sponsor: Fundación General CSIC (Programa ComFuturo) to Javier Vargas; Grant sponsor: European Union (EU) and Horizon 2020; Grant numbers: EINFRA-2014-2: EGI Engage (Proposal: 654142) and EINFRA-2015-1: West-Life (Proposal: 675858).

*Correspondence to: Jose María Carazo, Darwin 3, Universidad Autónoma de Madrid 28049 Madrid Spain. E-mail: carazo@cnb.csic.es

latter fact means researchers do not need to rely on their own local computing resources. Web tools in Structural Biology are very well known in certain scientific domains, such as CCP4² (on line) for X-ray crystallography (www.ccp4.ac.uk/ccp4online/) or the very successful weNMR³ (www.wenmr.eu). However, in the cryo-EM field the instances of web tools are scarcer, although certainly not absent. Along this line we may refer to the tilt pair validation server <http://www.ebi.ac.uk/pdbe/emdb/validation/tiltpair/> and the Fourier shell correlation server <http://www.ebi.ac.uk/pdbe/emdb/validation/fsc/>, hosted at the European Bioinformatics Institute, as well as Maskiton⁴ for image classification (<http://maskiton.nysbc.org>). It is in this context, and building on the web interface of the workflow software integrator Scipion⁵ and its capability to transparently bridge across different software packages, that we present Scipion Web Tools (SWT). In this work, we have singularized some common steps used in most cryo-EM image processing workflows and, specifically for them, we have created new predefined workflows paying special attention to the clarity of both the user interface and all input/output. In this way, SWT can be accessed from any computer platform just using a web browser to perform powerful image analysis workflows over multiple image processing packages running remotely. These workflows include some common tasks in single particle analysis (SPA), making special emphasis on maps and models reliability.

SWT starts with a relatively reduced set of tools, aiming at attracting users so as to drive future developments based on user demand. The criteria for choosing which webtools to publish are the following: The first one is “usefulness,” which represents our judgment about the capacity of delivering an important piece of new information. The second one is “complementarity and independency,” meaning that SWT is expected to be used as additional tools within an image processing workflow, with clear and defined input and output easily extracted from the user’s current workflow. The third criterion addresses the need to reach a balance between network and computational demands, which should not be very high, while the fourth and final one is that the software should not be commercial. Each of the protocols used in any of the webtools displays all the references to papers that ought to be cited in order to credit the software developers authors; we also provide the required links according to the software license. Note that we do not have an agreement in place with the authors of third party software, we just use the terms of the licenses. All tools are provided with detailed online help, guiding users through the different steps of the particular workflow they are interested, as shown in Figure 1.

Results

Scipion web tools server

Scipion Web Tools (SWT) is a web application built over Scipion framework software, offering access to

predefined image processing workflows over the web at <http://scipion.cnb.csic.es/m/services/>. It is an open access service and no login is required. SWT is technically a “web wrapper” for Scipion software, therefore it offers most of its features over the web. One of Scipion features is the ability to easily switch from one Cryo-EM software package to another, allowing users to smoothly make the output data from one package, the input data for a different package, thus trying out different software methods in a very simple context and with complete project traceability. All map visualization over the Web are done using NGL viewer.⁶

As in many analysis websites, data must be moved from the client computer to the SWT server. Clearly, this is the main disadvantage compared to a local software installation, where data is usually located on site. Data size varies depending on which stage of the Cryo-EM workflow users are in. Roughly speaking, a typical SPA workflow starts with thousands of Direct Electron Detector movies (in the range of TBs), whereas input for later stages (micrographs, particles and volumes) are in the range of GBs. Indeed, many tasks in cryo-EM handle relatively large data sets, explaining the scarcity of web tools in this scientific domain. Consequently, data transfer is a key element to be analyzed when designing and developing web tools. Since uploading big files could be, not only time consuming, but also sensitive to network reliability, SWT has implemented a resumable file upload. It is based on the standard HTTP file transfer, so it will not be faster than a regular upload, but it provides a robust way of transferring big files, even TB’s, in poor network conditions. Resumable upload make use of a django module (<https://github.com/jeanphix/django-resumable>) which basically splits a file into small parts which are then uploaded. The server side keeps track of those parts that have been already uploaded, maintaining the communication with the javascript browser to skip them in case of later upload.

All tools start with the creation of a project, which is basically a private storage space where users can start uploading their files and run some Cryo-EM processes. Since we do not require any registration, privacy of data relies on the privacy of the project URL. This address is not searchable or browsable, and it is the responsibility of the user to track this URL, for example bookmarking it. All projects expire after 2 weeks, which is deemed to be enough time for the analysis of the results of any of the provided tools. Additionally, only our system administrators and developers have been authorized to access the server that runs the presented system for maintenance and upgrade purposes.

For teaching and illustrating purposes, SWT offers 2 learning entry points: “Explore” and “Try it!” “Explore” entry point forwards users to a read

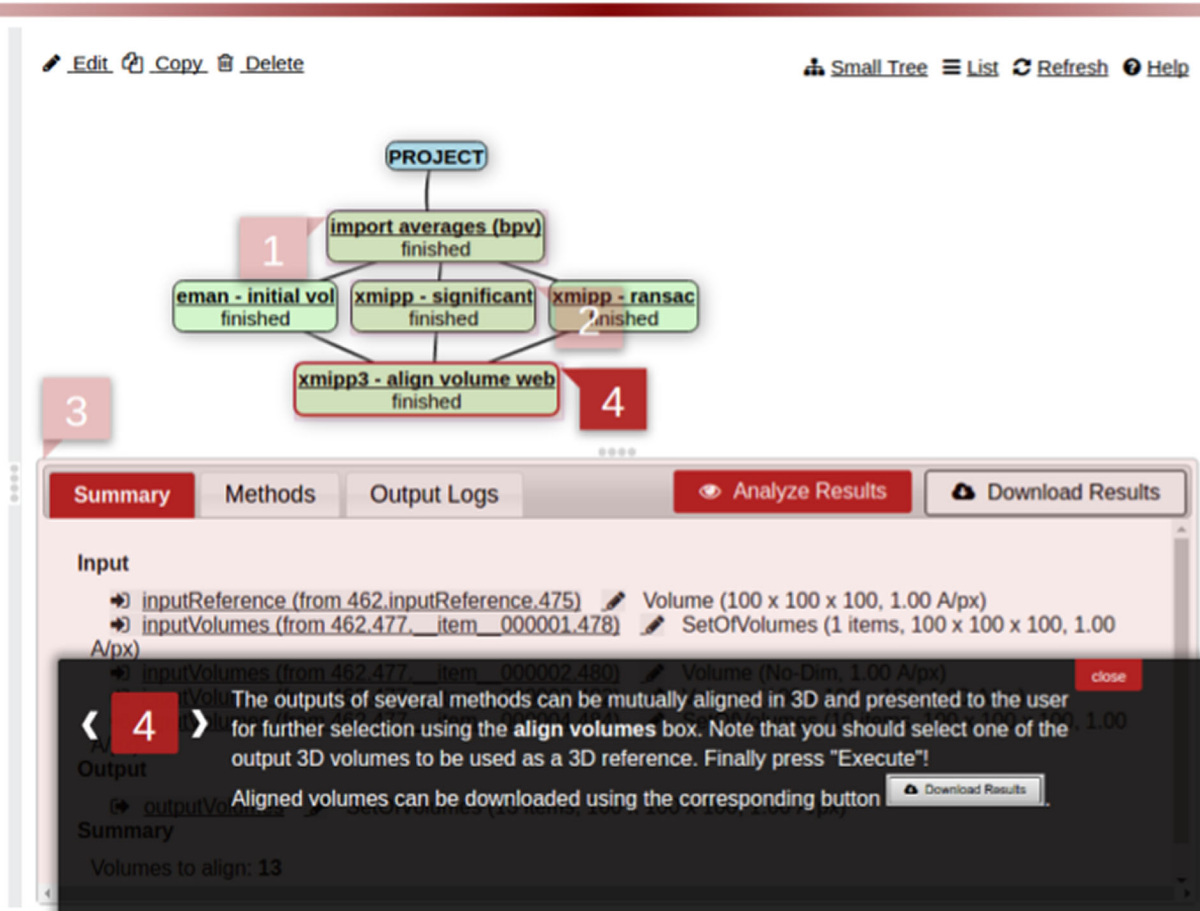


Figure 1. Sample project presenting a typical SWT workflow. Note that help is activated, highlighting the different steps of the process, and that the description corresponding to step 4 is shown

only, already executed project, where they can explore how data, parameters and output files are handled by that workflow. In turn, the second learning entry point (“Try it!”) offers users several data sets that are automatically loaded into the SWT workflow. In this latter case, data and parameters are provided by us, so that users can focus on the details of the graphic interface and on the processing workflow, running the processes with already tested data and parameters. The third entry point is “Use it!”, and it creates an empty project, without any data, where users should upload their own data and define and run everything from the beginning.

The actual hardware supporting SWT is located at CNB-CSIC, but Scipion is fully Cloud deployable,⁷ so it can be immediately moved to private (European Federated Cloud) or public (Amazon) clouds, according to user demands in the future.

Current web tools list

Currently we offer two sections of services, that we refer to as “Single particle analysis tools” and “Reliability tools.” The first section offers some basic

and relatively small workflows that most users will probably need to perform at some point during their image processing analysis workflow. Currently, we are offering 3 tools: “My movie alignment,” “My first map,” and “My resolution map.” The second section focuses on ways to analyze some quality parameters of maps (gathered under “My reliability tool”), and a protein interaction explorer tool accessing 3DIANA⁸ to study proposed atomic models.

Section 1: Single particle analysis tools. We have designed three basic tools to cover the initial step of processing (Beam Induced Movement correction), an intermediate step (de novo calculation of an Initial Map) and one of the final tasks (local resolution estimation). They are presented in more detail in the following:

My movie alignment (MMA). This tool addresses the beam induced motion (BIM) correction step, which is required for high resolution data collection with Direct Electron Detectors (DEDs). My Movie Alignment (MMA) is not a typical web tool, since it

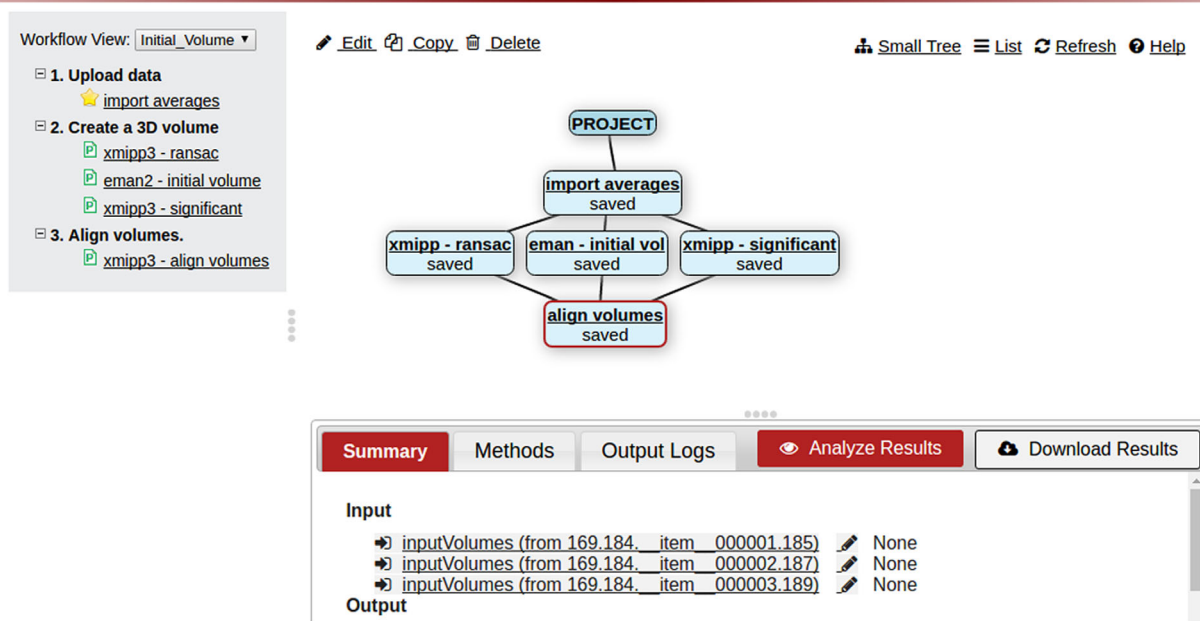


Figure 2. Workflow of My first map service

is intended mostly as a testbed and showcase of the different types of processing that can be done locally or remotely considering computational cost and data transfer speed. Indeed, BIM correction is the step involving the largest amount of data, making it a good candidate for a process to be performed locally.

Currently MMA workflow offers the Optical Flow algorithm,⁹ which delivers as output a corrected averaged micrograph. Additionally, MMA provides access to all graphical capabilities of Optical Flow 2D, allowing to easily analyze the precise pattern of BIM characteristics of the sample being imaged. Efficient uploading of very large data sets was a must in the design of MMA. Typically, the size of a movie is in the order of a few GB's, with each project requiring thousands of them. Consequently, the standard Scipion web upload function using a web browser was modified into a resumable, more robust but not faster upload. A second option was also enabled, allowing users to run the rsync command, which gets average transfer rates over Europe in the order of several tens of MB/s. At this speed, a typical project takes several hours to upload (for a deeper analysis of data transfer rates and opportunities for Cloud computing in cryo-EM, see Ref. 7). Still, the simplicity of use of rsync, together with its availability in most operating systems, makes it an interesting option to keep.

My first map (MFM). The macromolecular structural determination process in cryo-EM as applied to purified complexes in vitro is normally referred to as

Single Particle Analysis (SPA). Typical SPA image processing workflows require the user to supply a low resolution estimation of the structure, which we refer to as an initial model. The efficient determination of this initial model is still an open and challenging problem, especially when structural flexibility is important. MFM offers three initial map estimation algorithms, such as e2initialmodel,¹⁰ RANSAC¹¹ and Significant,¹² generating a gallery of pooled initial model candidates. Naturally, many more algorithms to deal with the initial model problem have been proposed in the literature (reviewed in Ref. 12), and the three selected ones are just an initial choice that can be enriched in subsequent versions of SWT. Figure 2 shows the standard MFM workflow. It starts by importing images, which most commonly refers to sets of 2D averages, and it follows with the different algorithms receiving the same averages as input data. The workflow ends with the listing of all suggested and mutually aligned initial models. Each of the volumes can be displayed either by slices or by surface rendering. Finally, user can download the ranked volumes in a very simple way from the lower SWT panel, just by clicking on "Download results."

My resolution map (MRM). The image end result of a cryo-EM study is a three-dimensional density map in which the different parts of the volume may present different degrees of quality (resolution) as a result of, most typically, flexibility or compositional differences of the macromolecule under consideration. Consequently, maps are better described using

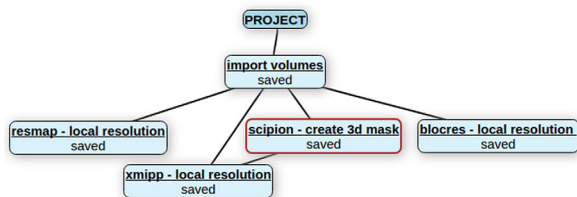


Figure 3. Workflow of My resolution map service

the concept of “local (per voxel) resolution,” rather than by a single global value. However, the concept of what “resolution” truly means is not simple, and several approaches have been proposed to measure resolution (for a review on resolution measures in cryo-EM, see Ref. 13). In fact, the existence of different methods, included in different software packages, provides a very good case for Scipion Web Tools. In this way, we show in Figure 3 the current MRM workflow, that includes three approaches to local resolution that are offered to the user in one single interface, namely Resmap,¹⁴ the currently most popular approach, Bloccres,¹⁵ the first one to be introduced in the field, and MonoRes (Vilas et al., under review), a new and fully automated approach. Note that Bloccres requires the user to supply two “half maps” and not just the final map; for simplicity, test data sets provided by SWT only refer to the case of calculating the local resolution of a final map and, therefore, only ResMap and MonoRes are shown in test mode.

We make special stress in that the availability of a web-based approach like SWT makes very easy to offer this workflow in a variety of new contexts, like a third-party pointing to a specific SWT to enhance a certain service being offered. Indeed, this is currently the case of use of SWT done by the Electron Microscopy Data Base (EMDB) at the European Bioinformatics Institute <https://www.ebi.ac.uk/pdbe/>

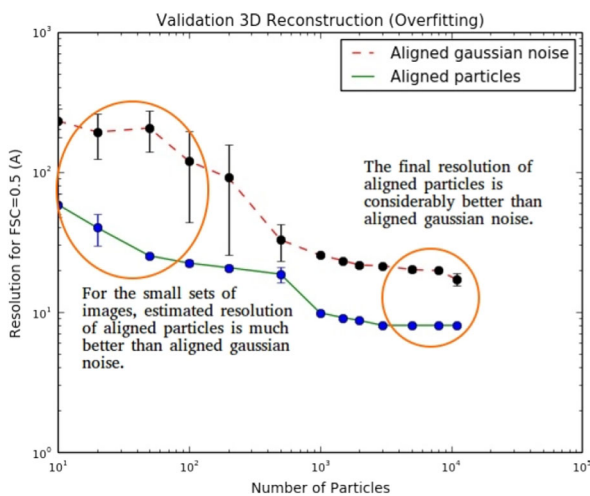


Figure 4. Phantom in the noise output chart as explained in SWT help

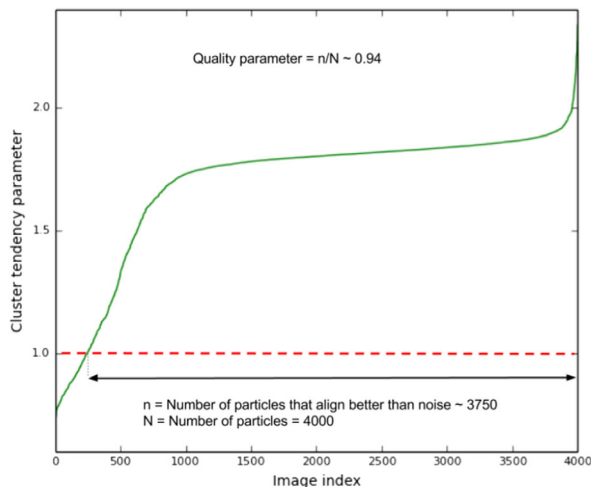


Figure 5. Angular reliability method output chart as explained in SWT help

emdb.validators.html/. In this way, EMDB at EBI, under EM Resources - Validation, provides a link to MRM, so that users can easily calculate the local resolution of any cryo-EM map, either already submitted or still in the analysis process.

Section 2: Reliability tools

My reliability tool (MRT). Cryo-EM methods can in many cases render a density map that seems reasonable but, is it correct?. Can users trust the outcome of a given processing workflow? This is a key question, specially in those many cases in which quasi atomic information is not reached. Consequently, validation and methods to assess the quality and reliability of maps have received much attention in recent years.¹⁶ “My reliability tool” (MRT) aims to provide several map reliability methods (two in this first version) that may help users to assess the quality of their reconstruction without any software installation requirement and without acquiring any additional experimental images (for other approaches, such as the use of tilt pairs, the interested reader is referred to other servers, like <https://www.ebi.ac.uk/pdbe/emdb/validators.html/>). One of the methods provided by SWT is referred to as “Phantom in the noise,”¹⁷ that compares the effect of signal and noise images in reconstructing a map (Fig. 4). The second method included in MRT assesses the angular assignment reliability between a set of particle projections and the map calculated from them,¹⁸ rendering the corresponding cluster tendency chart (Fig. 5).

Domain-domain interaction analysis. SWT includes the possibility to easily add Web-based tools from other portals, so that they incorporate a specific functionality into a new context. This is the case of 3DIANA protein domain-domain interaction tool,⁸ that we include under “My reliability tools” as

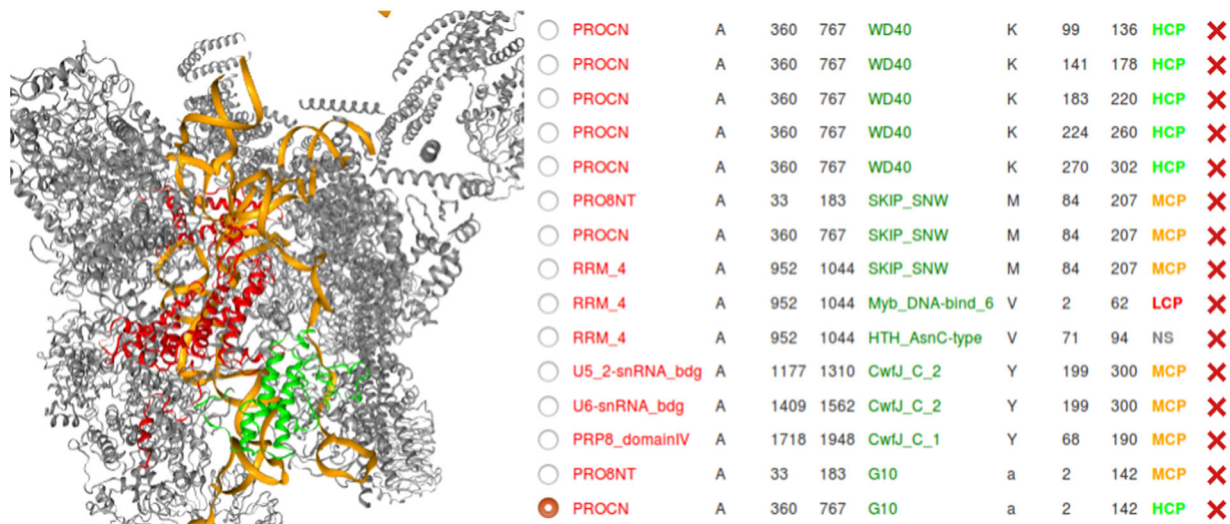


Figure 6. Protein Domain-Domain interaction analysis of the yeast spliceosome (EMDB entry: EMD-6413 and pdb structure: 3JB9). On the right side table, we display DIMERO prediction scores of domain interacting pairs. Location of selected domains within the structure is shown using the 3D viewer presented on the left-hand side panel. Note that on the right hand panel we have selected the interaction between PROCN and G10 domains of the splicing factors cf5 and spp42. Then, these domains are marked on the left hand-side panel in red (PROCN) and in green (G10), respectively (the rest of the complex is in gray, while RNA is in orange)

an approach to analyze interactions between domains of a proposed hybrid model. Protein domain-domain interactions are computed measuring the atomic distances between the different subunits in a complex, so that those pairs of domains that are at any residue closer than 6 Å are considered to interact. Then, the domain pairs are evaluated in terms of DIMERO scores.¹⁹ This method computes the likelihood of interaction between two domains based on different topological metrics measured on multiple protein-protein interaction networks. When this tool is used, the web server displays a table where all interacting domain pairs from the different subunits are displayed with their evaluation scores (Fig. 6). Moreover, a link to 3DID database²⁰ is provided for those cases in which the structure of a particular domain pair has been already experimentally solved in other complexes. Finally, the structure of the complex is displayed using the NGL viewer (Rose et al, 2015) and the different interacting domain pairs can be mapped in the structure interactively.

Discussion

Computing is being transformed in a way in which transparent access to remote computational facilities for the performance of multiple everyday tasks is very common. In the scientific domain, the availability of web based remote services is becoming more and more accepted as one of the key component of the provision of computational infrastructures. Certainly, some areas have a stronger tradition of web tools than others, but Structural Biology is not

orphan of them, as reviewed in the Introduction. It is in this context that we present here the first version of Scipion Web Tools (SWT), designed to be simple and intuitive to use and, at the same time, easily expandible. SWT targets users that are already in the process of image processing of electron micrographs for the analysis of macromolecular complexes following the so called Single Particle Analysis approach. These users will probably be following the proposed workflows provided by any of the packages in the field but, at certain key points of the process, may want to enhance/enrich these workflows with additional information coming from other packages but without the need to perform any local installation. SWT is designed for them, providing very easy and intuitive interaction, as well as very simple manners to import and export information. SWT are currently hosted at <http://scipion.cnb.csic.es/m/services/>. Additionally, as part of the support to e-Infrastructures for research in Structural Biology provided through the WestLife project, it is deployed at the EGI Federated Cloud.²¹ SWT is free and accessible to all users upon publication.

Acknowledgments

The authors want to thanks Alp Kucukelbir and Hemant Tagare for their collaboration in ResMap re-factoring for web usage. Also, we want to thanks Roberto Melero and Carmen San Martin for testing the services and reviewing the data. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro GP100 GPU used for parts of this development.

References

1. Vinothkumar KR, Henderson R (2016) Single particle electron microscopy: trends, issue and future perspective. *Q Rev Biophys* 49:e13.
2. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D* 67:235–242.
3. TA, Wassenaar M, van Dijk N, Loureiro-Ferreira G, van der Schot SJ, de Vries C, Schmitz J, van der Zwan R, Boelens A, Giachetti L, Ferella A, Rosato I, Bertini T, Herrmann HRA, Jonker A, Bagaria V, Jaravine P, Güntert H, Schwalbe WF, Vranken JF, Doreleijers G, Vriend GW, Vuister D, Franke A, Kikhney DI, Svergun RH, Fogh J, Ionides ED, Laue C, Spronk S, Jurkša M, Verlato S, Badoer S, Dal Pra M, Mazzucato E, Frizziero Alexandre MJJ Bonvin (Dec 2012) Wenmr: Structural biology on the grid. *J Grid Comput* 10:743–767.
4. Yoshioka C, Lyumkis D, Carragher B, Potter CS (2013) Maskiton: Interactive, web-based classification of single-particle electron microscopy images. *J Struct Biol* 182:156–163.
5. de la Rosa-Trevín JM, Quintana A, del Cano L, Zaldívar A, Foche I, Gutiérrez J, Gómez-Blanco J, Burguet-Castell J, Cuenca-Alba J, Abrishami V, Vargas J, Otón J, Sharov G, Vilas JL, Navas J, Conesa P, Kazemi M, Marabini R, Sorzano COS, Carazo JM (2016) Scipion: A software framework toward integration, reproducibility and validation in 3d electron microscopy. *J Struct Biol* 195:93–99.
6. Rose AS, Hildebrand PW (2015) NGL Viewer: a web application for molecular visualization. *Nucl. Acids Res* 43:576–579.
7. Cuenca-Alba J, del Cano L, Gómez Blanco J, de la Rosa Trevín JM, Conesa P, Marabini R, Sorzano COS, Carazo JM (2017) Scipioncloud: An integrative and interactive gateway for large scale cryo electron microscopy image processing on commercial and academic clouds. *J Struct Biol* 1047–8477.
8. Segura J, Sanchez-Garcia R, Tabas-Madrid D, Cuenca-Alba J, Sorzano COS, Carazo JM (2016) 3DIANA: 3D domain interaction analysis: A toolbox for quaternary structure modeling. *Biophys J* 110:766–755.
9. Abrishami V, Vargas J, Li X, Cheng Y, Marabini R, Sorzano COS, Carazo JM (2015) Alignment of direct detection device micrographs using a robust optical flow approach. *J Struct Biol* 189:163–176.
10. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ (2007) Eman2: An extensible image processing suite for electron microscopy. *J Struct Biol* 157: 38–46.
11. Vargas J, Alvarez-Cabrera A-L, Marabini R, Carazo JM, Sorzano COS (2014) Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics* 30:2891–2898.
12. Sorzano C, Vargas J, Rosa-Trevín JM, Otón J, Álvarez Cabrera A, Abrishami V, Sesmero E, Marabini R, Carazo J (2015) A statistical approach to the initial volume problem in single particle analysis by electron microscopy. *J Struct Biol* 189:213–219.
13. Sorzano C, Vargas J, Otón J, Abrishami V, Rosa-Trevín JM, Gómez-Blanco J, Vilas JL, Marabini R, Carazo J (2017) A review of resolution measures and related aspects in 3d electron microscopy. *Progr Biophys Mol Biol* 124:1–30.
14. Kucukelbir A, Sigworth FJ, Tagare HD (2014) Quantifying the local resolution of cryo-em density maps. *Nat Methods* 11:63–65.
15. Cardone G, Heymann JB, Steven AC (2013) One number does not fit all: Mapping local variations in resolution in cryo-em reconstructions. *J Struct Biol* 184:226–236.
16. Rosenthal PB, Rubinstein JL (2015) Validating maps from single particle electron cryomicroscopy. *Curr Opin Struct Biol* 34:135–144.
17. Heymann JB (2014) The Phantom in the Noise and Validation of 3D EM Reconstructions. *Microsc Microanal* 20:792–793.
18. Vargas J, Otón J, Marabini R, Carazo JM, Sorzano COS (2016) Particle alignment reliability in single particle electron cryomicroscopy: a general approach. *Sci Rep* 6.
19. Segura J, Sorzano COS, Cuenca-Alba J, Aloy P, Carazo JM (2015) Using neighborhood cohesiveness to infer interactions between protein domains. *Bioinformatics* 31:2545–2552.
20. Mosca R, Céol A, Stein A, Olivella R, Aloy P (2013) 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucl Acids Res* 42:374–379.
21. Fernández-del Castillo E, Scardaci D, Lopez García A (2015) The egi federated cloud e-infrastructure. *Procedia Comput Sci* 68:196–205.