Received: 20 January 2023 | Revised: 20 April 2023 | Accepted: 16 May 2023

DOI: 10.1002/minf.202300018

molecular informatics

APPLICATION NOTE

The VEGA web service: multipurpose online tools for molecular modelling and docking analyses

Alessandro Pedretti | **Serena Vittorio** | **Emanuela Sabato** | **Giulio Vistoli** | **Angelica Mazzolari**

Dipartimento di Scienze Farmaceutiche, Università degli Studi di Milano, Milano

Correspondence
Alessandro Pedretti, Dipartimento di Scienze Farmaceutiche, Università degli Studi di Milano, Via Mangiagalli, 25, I-20133 Milano.
Email: alessandro.pedretti@unimi.it

Abstract

The paper presents the VEGA Online web service, which includes a set of freely available tools deriving from the development of the VEGA suite of programs. In detail, the paper is focused on two tools: the VEGA Web Edition (WE) and the Score tool. The former is a versatile file format converter including relevant features for 2D/3D conversion, for surface mapping and for editing/preparing input files. The Score application allows rescoring docking poses and in particular includes the MLP Interactions Scores (MLPInS) for describing hydrophobic interactions. To the best of our knowledge, this web service is the only available resource by which one can calculate both the virtual log P of a given input molecule according to the MLP approach plus the corresponding MLP surface.

KEYWORDS
atom typing, MLPInS scores, molecular surface, rescore, VEGA program, virtual logP, web service

During the last years, a significant increase of the number of available online tools and web services was observed in all scientific fields [1]. Such a trend was further boosted by the recent pandemic crisis reasonably as the online resources allow reinforcing scientific collaborations while preserving physical distance and safety rules. Remarkably, these tools are usually free resources and are part of the open science movement, which has been gaining momentum in recent years.

Molecular modelling and more specifically *in silico* drug design are certainly no exception and a huge number of dedicated online resources are routinely utilized by researchers for a wide variety of applications [2]. The online resources can be employed for docking simulations and virtual screenings [3], for protein modelling and analysis [4, 5] as well as for cheminformatics applications often paired with AI-based analyses. [6] Not to mention the huge number of bioinformatics online platforms developed to assist omics analyses (e.g., refs. [7, 8]).

Here, we present the VEGA web service (https://www.ddl.unimi.it/vegaol), which comprises a set of resources deriving from the development of the VEGA suite of programs [9]. The VEGA software is a complete molecular modelling suite which includes several features for editing, converting, viewing and analysing molecular structures, molecular databases and trajectories supporting a wide variety of file formats. The suite also includes graphical user interfaces for various external programs such as AMMP, NAMD, MOPAC, Plants thus representing a versatile platform to assist different kinds of computational experiments. Finally, VEGA has a highly expandable architecture which allows additional features to be implemented by scripts and plug-in tools. The VEGA program is freely downloadable for non-profit users [10].

The VEGA web service includes several tools some of which have been already described in the previous studies, which originated them [11, 12]. All online tools are freely accessible to everyone and their utilization does

not require user's registration. Here, the attention is focused on two online tools: (1) the VEGA Web Edition (VEGA WE), which is a powerful file format converter also including relevant features for 2D/3D conversion, for surface mapping and for editing/preparing input files as well as (2) the Score application for rescoring docking poses [13] with a special focus on the MLP Interactions Scores (MLPInS) [14].

Notably, applications similar to the here proposed VEGA WE webserver are already available: they comprise the online version of well-known programs, such as OpenBabel [15] and ChemAxon JChem, [16] as well as web services for specific tasks, such as the management of SMILES strings [17] or the generation of input files for MM/MD programs [18]. Nevertheless, the VEGA WE tool includes a variety of features, which cannot be found in other resources and go far beyond the simple file format conversion and editing.

Figure 1 shows the graphical interface of the VEGA WE tool and evidences its key features. As input structure, the user can either upload a file or enter the text in a specific form. The tool automatically recognizes the format of the input molecule and supports the following file formats: Alchemy, AMMP, Accelrys Insight .car, Accelrys Quanta/CHARMm CRD, AutoDock 4 PDBQT, Cambridge Data File (CSSR), Cartesian coordinates (XYZ), Chem3D, ChemSol, CIF, CML, CML 2.0, CPMD XYZ, CRT, GAMESS Cartesian input and output, Gromos/Gromacs .gro, HyperChem .hin, InChI, InChI + aux data, MDL Molfile, MDL Molfile V3000, mmCIF, Mopac Cartesian, Mopac Gaussian Z-matrix, Mopac internal, Protein Data Bank (PDB), Protein Data Bank ATDL (PDBA), Protein Data Bank Fat (PDBF), Protein Data Bank Large (PDBL), PSF X-Plor, QMC, Tinker XYZ, and Tripos Sybyl (Mol2) [19].

FIGURE 1 Graphical interface of the VEGA WE tool accessible at https://www.ddl.unimi.it/vegaol/vegawe.htm.

On the input structure, a set of relevant tasks can be performed. Figure 1 shows the supported procedures which can be summarized as follows: (1) assigning the atomic charges according to the Gasteiger-Marsili method [20], (2) normalizing the coordinates, (3) removing water molecules (automatically recognized by the system), (4) adding the hydrogen atoms according to a predefined pH value; (5) removing the hydrogen atoms by selecting if removing all hydrogens or only the apolar ones, and (6) detecting the torsion angles (to be set if the output format requires them, e.g., PDBQT). If the input file corresponds to a protein structure, the tool can also add the missing side chains.

Among the implemented features, the capacity to assign atom types is of particular relevance to support the generation of input files for various applications. In detail, this function performs the atom typing based on a specified force field template written in ATDL format (Atom Type Description Language) [21]. In order to assign the correct atom types, VEGA applies a multi-step algorithm which (1) generates the connectivity table; (2) attributes the hybridization to each atom; (3) searches for the rings; (4) detects the aromatic systems using the Hückel rule; (5) parses the selected ATDL template and finally (6) assigns the resulting atom types. The detailed description of the ATDL templates available online can be found in [22].

The resulting output file can be generated by supporting the same formats already listed when describing the input files. Additionally, the VEGA online tool can generate the canonical SMILES string as well as the FASTA file if the input file includes a protein structure. The web service can also convert the input file in the Interchange File Format (IFF) format [23]. This is a binary format with a chunk structure in which all chunks are optional (apart from the first one) and the structure is totally expandable. This allows the storage of a huge number of information in a very compact and flexible way. Further details on the architecture of the IFF files can be found at [24].

The VEGA online resource can also generate four types of molecular surfaces storing them in CSV, VRML and SRF (binary) formats. Further details on the characteristics of these surface's formats can be found at [25]. The four types of calculated surface are: (1) the simple Van der Waals surface; (2) the Molecular Electrostatic Potential (MEP); (3) the Molecular Lipophilic Potential (MLP) [26] and (4) Molecular Lipophilic Index (ILM) [27]. For each surface, the user can define both the probe radius (by default equal to 1.4 Å) and the dot

density (i.e., number of dots per Å^2, by default equal to 10). The MEP surface requires the attribution of the atomic charges, while the ILM surface requires a solvated system since it is based on the distance between the solute's atoms and the solvent's molecules as described in [27]. To the best of our knowledge, the VEGA web service is the only online resource able to calculate the MLP surface according to [22]. Notably, the VEGA web service also includes an online tool to calculate the conformer-dependent log P value accordingly to the MLP approach (the so-called virtual log P) [28]. As previously described, the input form accepts both a file and a text and supports the file formats already listed for the VEGA WE tool. Overall, the VEGA web service offers an online integrated system to predict the log P value of a given input structure based on the MLP 3D method as well as to generate the corresponding MLP surface.

To support the modelling of chemical structures, the VEGA webserver also includes a tool for editing 2D molecular structure and for the 2D→3D conversion. The molecular editor is based on the Ketcher JavaScript applet [29] while the 2D/3D conversion involves the AMMP molecular mechanics software [30]. In detail, the 2D/3D conversion comprises the following steps: (1) adding the hydrogen atoms (when necessary); (2) assigning the atom charges (by the Gasteiger–Marsili method); (3) saving the molecule in the AMMP file format. Then, AMMP optimizes the structure in two steps: (1) distance geometry optimization by Gauss–Seidel algorithm (50 steps); (2) conjugate gradients minimization (500 steps, toler = 0.01). The minimized structure is finally converted into the requested file format. Notice that the implemented minimization procedure is intended to perform an initial roughing of the 3D structure. Each user will then optimize the generated structure based on his preferred protocols.

Figure 2 shows the graphical web interface of the Score tool which can be considered as a focused online version of the Rescore+ software [31] and is purposely designed for rescoring analyses of docking results. The tool accepts two types of input files since one may submit either the structure of the complex indicating the molecule number corresponding to the ligand or the two separate files (receptor plus ligand). The supported formats for the input files are the same listed above.

Table 1 lists the computed scoring functions, which can be subdivided into three main groups accordingly to the type of interaction they encode. Concerning the ionic interactions, the online tool utilizes the canonical

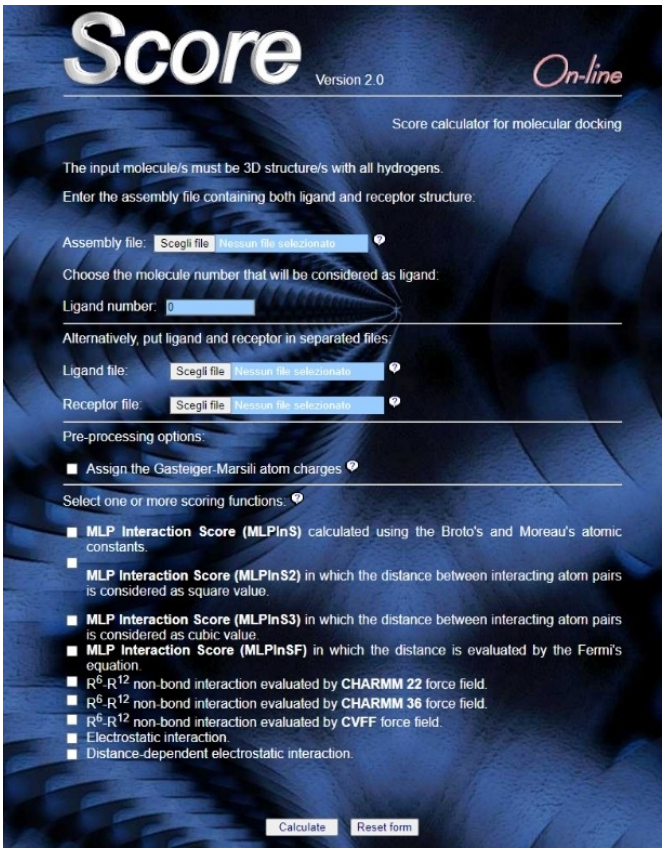

FIGURE 2 Graphical interface of the Score tool accessible at https://www.ddl.unimi.it/vegaol/score.htm.

TABLE 1 Scoring functions computed by the Score tool.

Scoring function	Description
MLP_{InS}	Hydrophobic interaction calculated using the Broto and Moreau atomic constants.
MLP_{InS2}	Hydrophobic interaction with a square distance between interacting atoms.
MLP_{InS3}	Hydrophobic interaction with a cubic distance between interacting atoms.
MLP_{InSF}	Hydrophobic interaction in which the distance between atoms is evaluated by the Fermi's equation.
CHARMM 22 R6-R12	R^6-R^{12} non-bond interaction evaluated by CHARMM 22 force field.
CHARMM 36 R6-R12	R^6-R^{12} non-bond interaction evaluated by CHARMM 36 force field.
CVFF R6-R12	R^6-R^{12} non-bond interaction evaluated by CVFF force field.
Electrostatic	Electrostatic interaction with dielectric constant equal to 1.
Electrostatic distance dependent	Electrostatic interaction with a distance-dependent dielectric function.

Coulomb's equation by applying either a dielectric constant equal to 1 or a distance dependent dielectric function.

Clearly, these scores require that the atomic charges have been assigned. If the input files lack these data, the tool can calculate them on the fly by applying the Gasteiger–Marsili method. With regard to Van der Waals contacts, the tool allows their calculation by applying the Lennard-Jones equation as parameterized by three different force fields, CHARMM22 [32], CHARMM36 [33] and CVFF [34]. Notice that a precise evaluation of these scoring functions would require that the analysed complexes are previously minimized by utilizing the same force field. The third group of scoring functions were proposed by us to encode for hydrophobic contacts and are based on the Broto and Moreau lipophilic increments [35].

In detail, the MLP_{InS} scores [12] are computed by employing a Coulomb-like equation in which the Broto and Moreau atomic increments replace the atomic charges and the distance function can be selected among four possible functions: linear (MLP_{InS}), square (MLP_{InS2}), cubic (MLP_{InS3}) and Fermi's type (MLP_{InSF}) distance.

The output file is written in XML format and includes the following main data for each selected score: (1) the computed score value; (2) the contribution of each atom to the total score; and (3) the contribution of each residue to the total score.

To conclude, the VEGA web service represents a freely available online resource for converting, preparing and editing files for molecular structures. It also includes features for 2D/3D conversion, surface mapping and atom typing, the last being of particular relevance for preparing input files for MM/MD runs. The Score tool allows the calculation of a set of scoring functions to parameterize polar, non-polar and hydrophobic interactions.

Special emphasis is given here on lipophilicity: to the best of our knowledge, this is the only free online resource by which one can calculate the virtual log P according to the MLP approach also generating the corresponding MLP surface. The MLPInS scores also move in this direction since they use the same lipophilic atomic increments to evaluate MLP-based hydrophobic interactions. These web services are intended to parallel the features constantly added to the VEGA suite of programs and could be expanded by supporting additional file formats (for the VEGA WE tool) or by implementing novel scoring function (for the Score tool).

COMPUTATIONAL METHODS

VEGA Online comprises two components: a PHP script and the VEGA command line version. When the form is completed and submitted by the user, the data are sent to the Web server via the standard POST command. The script saves the input molecule into a suitable temporary file and processes the user options transforming them in a set of commands for the VEGA command line version. Thus, the VEGA command line is launched and script waits for the end of the calculation and then processes the resulting output file to be compatible with the HTML standard. Finally, the so prepared output is sent to the user by the webserver.

A similar workflow is also utilized by the Score tool. Here the related script sends the input file to the VEGA command line which generates the complex keeping the original atom coordinates, if the user provides two separated files. If a single file is submitted, VEGA detects the molecules and recognizes the ligand by the molecule number indicated in the specific field. If required, VEGA assigns the atomic charges and then evaluates the scores selected by the user. The calculated output scores are encoded in XML format as described above. Notice that the Score tool does not perform docking calculations and the submitted structures should be completed with hydrogens and preferably optimized.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Alessandro Pedretti http://orcid.org/0000-0001-5916-2029

REFERENCES

1. https://www.statista.com/topics/3071/cloud-software-as-a-service-saas.
2. V. Potemkin, A. Potemkin, M. Grishina, *Curr. Top. Med. Chem.* **2018**, *18*, 1955–1975.
3. J. Krüger, P. Thiel, I. Merelli, R. Grunzke, S. Gesing, *Curr. Drug Targets.* **2016**, *17*, 1649–1660.
4. V. A. Paiva, I. S. Gomes, C. R. Monteiro, M. V. Mendonça, P. M. Martins, C. A. Santana, V. Gonçalves-Almeida, S. C. Izidoro, R. C. Melo-Minardi, S. A, Silveira, *Comput. Biol. Med.* **2022**, *147*, 105695.

5. L. Kurgan, *Methods.* **2022**, *204*, 132–141.
6. J. L. Medina-Franco, N. Sánchez-Cruz, E. López-López, B. I. Díaz-Eufracio, *J. Comput. Aided. Mol. Des.* **2022**, *36*, 341–354.
7. G. R. Bhat, I. Sethi, B. Rah, R. Kumar, D. Afroze, *Front. Genet.* **2022**, *13*, 865182.
8. M. Mahjoubin-Tehran, S. Rezaei, A. Jalili, A. Sahebkar, S. H. Aghaee-Bakhtiari, *Brief. Bioinform.* **2022**, *23*, bbab381.
9. A. Pedretti, A. Mazzolari, S. Gervasoni, L. Fumagalli, G. Vistoli, *Bioinformatics.* **2021**, *37*, 1174–1175.
10. https://www.ddl.unimi.it/cms/index.php?Software_projects:VEGA_ZZ:Download.
11. G. Russo, L. Grumetto, F. Barbato, G. Vistoli, A. Pedretti, *Eur. J. Pharm. Sci.* **2017**, *99*, 173–184.
12. A. Pedretti, A. Mazzolari, S. Gervasoni, G. Vistoli, *Appl. Sci.* **2020**, *10*(21), 7704.
13. A. Pedretti, A. Mazzolari, S. Gervasoni, G. Vistoli, *Int. J. Mol. Sci.* **2019**, *20*(9), 2060.
14. G. Vistoli, A. Pedretti, A. Mazzolari, B. Testa, *Bioorg. Med. Chem.* **2010**, *18*(1), 320–329.
15. N. M. O′Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J. Cheminform.* **2011**, *3*, 33.
16. F. Csizmadia, *J. Chem. Inf. Comput. Sci.* **2000**, *40*(2), 323–324.
17. https://cactus.nci.nih.gov/translate/.
18. S. Jo, T. Kim, V. G. Iyer, W. Im, *J. Comput. Chem.* **2008**, *29*, 1859–1865.
19. A. Pedretti, L. Villa, G. Vistoli, *J Mol Graph Model.* **2002**, *21*(1), 47–49.
20. J. Gasteiger, M. Marsili, *Tetrahedron Lett.* **1978**, *19*(34), 3181–3184.
21. A. Pedretti, L. Villa, G. Vistoli, *Theor. Chem. Acc.* **2003**, *109*, 229–232.
22. https://www.ddl.unimi.it/manual/pages/force_field.htm.
23. A. Pedretti, L. Villa, G. Vistoli, *J. Comput. Aided. Mol. Des.* **2004**, *18*(3), 167–73.
24. https://www.ddl.unimi.it/manual/pages/appx_d_iff.htm.
25. https://www.ddl.unimi.it/manual/pages/gl_srf.htm.
26. P. Gaillard, P. A. Carrupt, B. Testa, A. Boudon, *J. Comput. Aided. Mol. Des.* **1994**, *8*(2), 83–96.
27. A. Pedretti, A. M. Villa, L. Villa, G. Vistoli, *Internet J. Chem.* **2000**, *45*(7), Art. 13.
28. https://www.ddl.unimi.it/vegaol/vlogp.htm.
29. https://lifescience.opensource.epam.com/ketcher.
30. R. W. Harrison, *J. Comput. Chem.* **1993**, *14*, 1112–1122.
31. A. Pedretti, C. Granito, A. Mazzolari, G. Vistoli, *Mol. Inform.* **2016**, *35*(8-9), 424–433.
32. A. D. Jr. MacKerell, D. Bashford, M. Bellott, R. L. Jr. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. III Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B.* **1998**, *102*, 3586–3616.
33. R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. Jr. MacKerell, *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
34. P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, A. T. Hagler, *Proteins.* **1988**, *4*(1), 31–47.
35. P. Broto, G. Moreau, C. Vandycke, *Eur. J. Med. Chem.* **1984**, *19*, 71–78.

How to cite this article: A. Pedretti, S. Vittorio, E. Sabato, G. Vistoli, A. Mazzolari, *Molecular Informatics* **2023**, *42*, e202300018. https://doi.org/10.1002/minf.202300018