

DockQ v2: Improved automatic quality measure for protein multimers, nucleic acids, and small molecules

Claudio Mirabello^{1,2} and Björn Wallner¹✉

¹Division of Bioinformatics, Department of Physics, Chemistry and Biology, Linköping University, 581 83 Linköping, Sweden

²National Bioinformatics Infrastructure Sweden, Science for Life Laboratory

Abstract

It is important to assess the quality of modeled biomolecules to benchmark and assess the performance of different prediction methods. DockQ has emerged as the standard tool for assessing the quality of protein interfaces in model structures against given references. However, as predictions of large multimers with multiple chains become more common, there is a need to update DockQ with more functionality for robustness and speed. Moreover, as the field progresses and more methods are released to predict interactions between proteins and other types of molecules, such as nucleic acids and small molecules, it becomes necessary to have a tool that can assess all types of interactions. Here, we present a complete reimplementation of DockQ in pure Python. The updated version of DockQ is more portable, faster and introduces novel functionalities, such as automatic DockQ calculations for multiple interfaces and automatic chain mapping with multi-threading. These enhancements are designed to facilitate comparative analyses of protein complexes, particularly large multi-chain complexes. Furthermore, DockQ is now also able to score interfaces between proteins, nucleic acids, and small molecules.

Code: <https://wallnerlab.org/DockQ>

Correspondence: bjorn.wallner@liu.se

Introduction

Protein interactions are crucial for most biological functions, including metabolism, gene expression, cell signaling, and immune response. Structural characterization of protein interactions at the molecular level is important to understanding basic biology, as well as disease mechanisms. Thus, elucidating protein structures has been the subject of immense research in previous decades. Experimental methods such as X-ray crystallography, NMR, and cryo-EM have determined the three-dimensional structure of over 200,000 proteins. However, experimental determination of protein structure is difficult, time-consuming, and sometimes even impossible given experimental constraints. Also, 200,000 structures are a small number compared to the over 250,000,000 proteins known to date. Thus, there has been a massive drive to develop computational methods to close this gap by predicting the structural properties of proteins.

To be able to compare the performance of computational methods it is important to have robust measures that can assess the quality of predicted models. Many such meth-

ods have been developed over the years, mostly for assessing monomer structures: MaxSub (1), GDT_TS (2), TM-score (3), and LDDT (4). Lately, methods have been developed also for multimers: DockQ (5), oligoLDDT (6), or US-align (7). In contrast to oligoLDDT and US-align, which measure the overall structural similarity of multimers, DockQ assesses the quality of the interactions between chains in multimers. This is an advantage since the quality of these interactions might be obscured in the overall structural similarity.

Lately, next-generation prediction tools, such as RoseTTAFold-All Atom (8) and AlphaFold 3 (9), have been released that can predict interactions between proteins, nucleic acids, and small molecules alike. In order to benchmark these new methods, it is important for a tool like DockQ to be able to evaluate interactions between all these types of molecules rather than forcing users to rely on multiple specialized tools.

We now introduce a new version of DockQ (v2). DockQ v2 is faster than the previous version, easier to install, and more portable. In this version, we include new functionalities: DockQ v2 can score multiple interfaces simultaneously in complexes with more than two chains and automatically detects the model-to-native chain mapping that maximizes the global quality of the complex. We also include the possibility of scoring interactions between proteins, nucleic acids, and small molecules. To the best of our knowledge, DockQ is the only tool that can seamlessly handle all these types of molecules.

In addition, we have also added interface clash detection, multithreading, transparent handling of mmCIF and gzipped files, the possibility to install it with *pip*, and the possibility to import it as Python module.

Method description

DockQ - assessing the quality of multimers. DockQ assesses the quality of a multimeric model by comparing it to a reference structure (typically an experimentally determined structure). The comparison is performed by assessing the similarity of the interfaces between protein chains. For a given protein-protein or protein-nucleic acid interface, the DockQ score is the average of the fraction of native contacts (f_{nat}), scaled interface RMSD (iRMSD), and scaled ligand RMSD (LRMSD):

$$\text{DockQ} = \frac{1}{3} \left(f_{\text{nat}} + \frac{1}{1 + (\frac{i\text{RMSD}}{1.5})^2} + \frac{1}{1 + (\frac{\text{LRMSD}}{8.5})^2} \right) \quad (1)$$

Where 1.5 and 8.5 are scaling parameters that monitors how fast iRMSD and LRMSD, respectively, should go to zero. These parameters were determined by maximizing the ability of DockQ to classify protein interfaces according to the CAPRI metrics ("Incorrect", "Acceptable", "Medium" and "High quality") (10). The advantage of the DockQ score is that it is a continuous score in the [0, 1] range, from incorrect to high quality, which facilitates method comparisons since distributions, instead of classifications, of model quality scores can be compared. This also makes it better suited for machine learning (11).

To calculate the DockQ score using Eq. 1, the f_{nat} , iRMSD, and LRMSD need to be calculated, below we describe how these measures are defined.

Fraction of correctly predicted reference contacts. The f_{nat} measure is the fraction of correctly predicted reference or native contacts in the interface. An interface is a set of interacting amino acids or nucleotides sitting in separate chains with at least one pair of heavy atoms within 5 Å of each other (4 Å in the case of peptides). A contact is correct if a pair of interfacial amino acids in the model are also interacting in the reference. The total number of correct contacts is then divided by the total number of contacts in the model interfaces. A problem with f_{nat} is that it does not consider the number of false predictions. Thus, a model with a large number of false positives might still receive a fairly high f_{nat} . Also, models with clashes will obtain higher f_{nat} , taking it to the extreme, a model with all (x, y, z) coordinates at 0 will have $f_{\text{nat}}=1.0$. This is, in general, only a problem for models with many clashes. Therefore, the number of clashes is automatically detected and reported to the user in DockQ v2.

Interface RMSD. The interface RMSD, iRMSD, is the backbone RMSD of the interface residues. Interface residues are defined as pairs of residues between protein and/or nucleic acid chains where any two heavy atoms are within 10 Å of each other (8 Å between pairs of $C\beta$ atoms in case of peptides).

Ligand RMSD. The ligand RMSD, LRMSD, is the backbone RMSD of the ligand when superimposing on the receptor. The receptor is defined as the larger of the two chains by number of amino acids or nucleotides. If two chains are equal in size, the first will be considered the receptor. In the case of small molecule ligands, LRMSD is calculated on the position of all heavy atoms in the ligand when superimposing on the receptor interface (*pocket-aligned LRMSD*).

Workflow and new functionalities. DockQ used to be protein-only. We have now added the functionality to also score interactions involving nucleic acids, and small molecule ligands. In the case of small molecule ligands,

only the pocket-aligned ligand RMSD (LRMSD) is reported, as commonly done in other methods (12, 13) and the corresponding DockQ score is calculated only based on the LRMSD component in Eq. 1 ($f_{\text{nat}} = i\text{RMSD} = 0$). Since small molecules are not ordered like proteins and nucleic acids, we use NetworkX's (14) graph matching functionality and report the solution with lowest LRMSD when multiple matches are possible (e.g. symmetric molecules). Graph edges are calculated according to the atoms' covalent radii (15).

DockQ reads the coordinates of a model and reference structure in PDB or mmCIF format, either uncompressed or compressed with *gzip*, to calculate the DockQ score (Eq. 1) for each interface in the reference structure. It returns the DockQ score for each interface as well as an overall *GlobalDockQ*, defined as the average of DockQ scores over all interfaces. Equivalent residues in the model and reference are mapped using sequence alignment (default) or by following the residue numbering.

For multimeric complexes with subunit stoichiometry larger than one, there is a need to find the optimal chain mapping, i.e. the chain mapping that maximizes the overall *GlobalDockQ*. This is also needed when multiple small molecules of the same type are being evaluated at once. The number of possible chain mappings grows with the factorial of the number of equivalent molecules, i.e., a homodimer of stoichiometry A2 has $2! = 2$ possible mappings, while a homooctamer of stoichiometry A8 has $8! = 40,320$. The default behavior is that DockQ will identify equivalent subunits by sequence alignment (by molecule name in the case of small molecule ligands) and exhaustively permute the chain mappings to find the one with the highest *GlobalDockQ*. It is also possible to provide a given chain mapping with the *--mapping* option or to limit the search to a subset of native chains. The search is parallelized over multiple threads (default=8) using LRU caching and with computationally intensive functions, such as distance calculations, implemented in Cython for speed.

DockQ v2 is released as a pip-installable Python package. This allows you to import and use DockQ as a module in other software.

Results

To verify that DockQ v2 is able to reproduce the DockQ scores generated with the previous version, we compared DockQ scores calculated for 17,409 models from 37 CASP15 targets. The agreement is almost perfect ($R=1.000$) between the new and the previous version; see Figure 1a. Thus, the scores from both versions can be used interchangeably.

The updated version is generally faster, especially for larger targets, as shown in Figure 1b. The linear portion of the time complexity is decreased approximately four times (4.08) compared to the previous version. This leads to significantly reduced runtimes in practice.

An important new feature is automatic chain mapping, which simplifies the problem of finding the optimal chain mapping between symmetric chains. Previously, the optimal

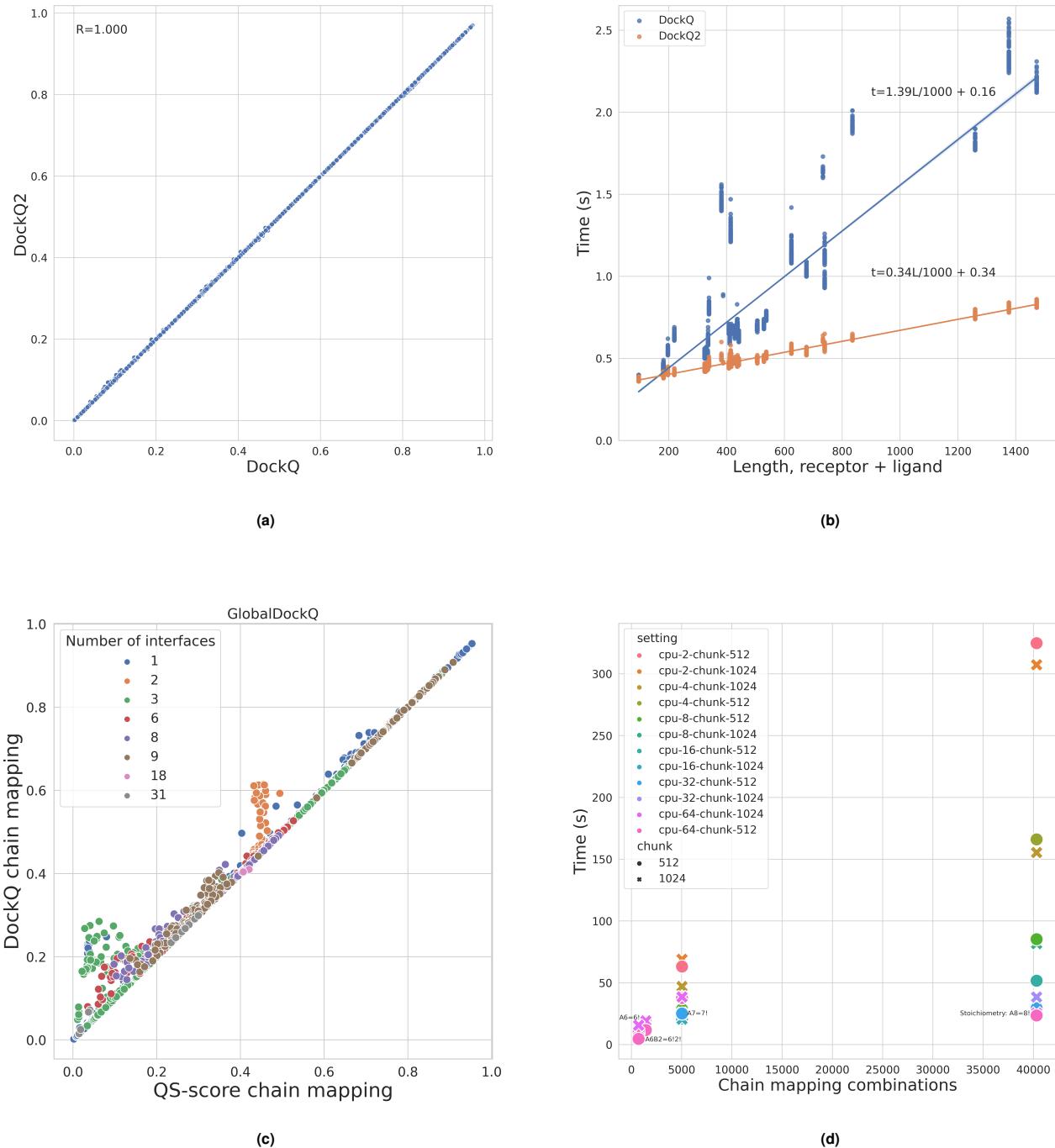


Fig. 1. (a) We evaluated DockQ v2 against the old DockQ (v1.0) on a large benchmark of 17,049 models from 37 CASP15 targets. The DockQ scores are equivalent between the two versions. (b) The new version of DockQ is faster than the previous due to optimised implementation of computationally expensive functions in Cython, especially for larger complexes. (c) Comparison of methods to perform automatic mappings between model and native chains: the automatic mapping done in DockQ yields the same or better GlobalDockQ than when using mappings optimised by QS-score. (d) Time to exhaustively optimise the mapping between sets of chains for complexes with different stoichiometry. The time complexity is linear with the number of chain mapping combinations, and can be proportionally cut down by increasing the number of parallel threads in DockQ.

chain mapping has either been ignored, assuming a one-to-one correspondence between the chains in model and reference, or had to be determined with some other software like the chain mapping routine in QS-score (16). The lat-

ter was used when calculating DockQ in the evaluation of multimer predictions in CASP15 (17). We compared the DockQ scores obtained using the new automatic chain mapping to the DockQ scores obtained by using the mapping

from the chain mapping routine in QS-score. In all cases, the automatic chain mapping in DockQ produces an identical or higher GlobalDockQ compared to using the mapping provided by QS-score, Figure 1c. This is expected since the mapping using QS-score is optimized using QS-score, while the automatic chain mapping in DockQ is optimized on GlobalDockQ. To make the automatic chain mapping faster, it is parallelized on multiple threads, making it possible to exhaustively map a homomer with stoichiometry A8 ($8! = 40,320$ chain combinations) in less than a minute on 8 threads or one with stoichiometry A7 ($7! = 5040$ combinations) in less than 30 seconds, see Figure 1d.

Taken together, we believe that this new, improved version of DockQ will become an essential tool for evaluating models of biomolecular complexes, regardless of the type of molecule they may include.

ACKNOWLEDGEMENTS

This work was supported by the Wallenberg AI, Autonomous System and Software Program (WASP) from Knut and Alice Wallenberg Foundation (KAW), Swedish Research Council grant, 2020-03352, The Swedish e-Science Research Center, and Carl Tryggers stiftelse för Vetenskaplig Forskning, 20:453. CM is financially supported by the Knut and Alice Wallenberg (KAW) Foundation and the "BeyondFold" Technological Development Project at SciLifeLab. The computations were performed on resources provided by KAW and NSC (Berzelius).

Bibliography

1. N Siew, A Elofsson, L Rychlewski, and D Fischer. MaxSub: An automated measure to assess the quality of protein structure predictions. *Bioinformatics*, 16(9):776–785, 2000.
2. Adam Zemla. LGA: a method for finding 3D similarities in protein structures. *Nucleic Acids Research*, 31(13):3370–3374, 2003. ISSN 0305-1048. doi: 10.1093/nar/gkg571.
3. Yang Zhang and Jeffrey Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710, 2004.
4. Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013. ISSN 1367-4803. doi: 10.1093/bioinformatics/btt473.
5. Sankar Basu and Björn Wallner. DockQ: a quality measure for protein-protein docking models. *PLoS one*, 11(8):e0161879, 2016.
6. Juergen Haas, Rafal Gumienny, Alessandro Barbato, Flavio Ackermann, Gerardo Tauriello, Martino Bertoni, Gabriel Studer, Anna Smolinski, and Torsten Schwede. Introducing “best single template” models as reference baseline for the Continuous Automated Model Evaluation (CAMEO). *Proteins: Structure, Function, and Bioinformatics*, 87(12):1378–1387, 2019. ISSN 0887-3585. doi: 10.1002/prot.25815.
7. Chengxin Zhang, Morgan Shine, Anna Marie Pyle, and Yang Zhang. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature Methods*, 19(9):1109–1115, 2022. ISSN 1548-7091. doi: 10.1038/s41592-022-01585-1.
8. Generalized biomolecular modeling and design with RoseTTAFold All-Atom.
9. Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, pages 1–3, 2024.
10. Keeley W. Collins, Matthew M. Copeland, Guillaume Brysbaert, Shoshana J. Wodak, Alexandre M.J.J. Bonvin, Petras J. Kundrotas, Ilya A. Vakser, and Marc F. Lensink. CAPRI-Q: The CAPRI resource evaluating the quality of predicted structures of protein complexes. *Journal of Molecular Biology*, page 168540, 2024. ISSN 0022-2836. doi: 10.1016/j.jmb.2024.168540.
11. Sankar Basu and Björn Wallner. Finding correct protein–protein docking models using Pro-QDock. *Bioinformatics*, 32(12):i262–i270, 2016.
12. Eric W Bell and Yang Zhang. Dockrmsd: an open-source tool for atom mapping and rmsd calculation of symmetric molecules through graph isomorphism. *Journal of cheminformatics*, 11:1–9, 2019.
13. Rocco Meli and Philip C Biggin. spryrmsd: symmetry-corrected RMSD calculations in Python. *Journal of Cheminformatics*, 12(1):49, 2020.
14. Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.
15. Beatriz Cordero, Verónica Gómez, Ana E Platero-Prats, Marc Revés, Jorge Echeverría, Eduard Cremades, Flavio Barragán, and Santiago Alvarez. Covalent radii revisited. *Dalton Transactions*, (21):2832–2838, 2008.
16. Martino Bertoni, Florian Kiefer, Marco Biasini, Lorenza Bordoli, and Torsten Schwede. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports*, 7(1):10480, 9 2017. doi: 10.1038/s41598-017-09654-8.
17. Gabriel Studer, Gerardo Tauriello, and Torsten Schwede. Assessment of the assessment—All about complexes. *Proteins: Structure, Function, and Bioinformatics*, 91(12):1850–1860, 2023. ISSN 0887-3585. doi: 10.1002/prot.26612.