αCharges

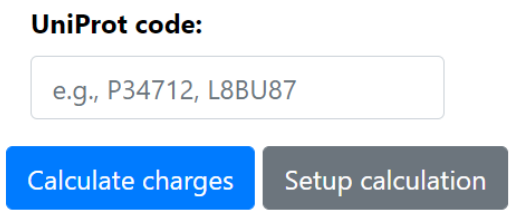
Contents

Overview

AlphaCharges (αCharges) is a web application providing a simple to use interface for calculation of partial atomic charges on protein structures predicted by the AlphaFold2 algorithm [Jumper2021] and deposited in the AlphaFoldDB database [Varadi2022]. The application consists of four pages: Main Page, the Calculation settings page, the Calculation progress page, and the Calculation results page.

Main Page

The main page allows uploading the chosen structure from AlphaFoldDB. The structure must be defined by its UniProt [UniProt2019] identifier (so-called UniProt code):



The UniProt codes of the proteins can be obtained directly from AlphaFoldDB. Here, you can search for your protein, gene, or organism.

Examples of UniProt codes, which are accepted by αCharges: A0A159JYF7, Q96247, F4HT52, or C0SV66.
Using extended UniProt codes (e.g., A0A159JYF7_9DIPT, AUX1_ARATH, F4HT52_ARATH) or UniProt codes of fragments (e.g., Q8WZ42-F1, Q8WZ42-F2) is not allowed by αCharges.

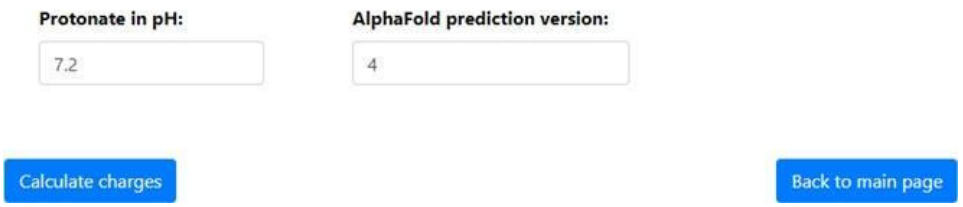
By clicking on the "Calculate charges" button, the computation of charges will start with default parameters (i.e., physiological pH 7.2 and AlphaFoldDB version 4), and the user is redirected to the Calculation progress page.

Moreover, it is possible to modify the computation setup using the "Setup calculation" button. This button redirects the user to the Calculation settings page.

The Main page also offers three use cases – proteins, which partial atomic charges provide interesting information for the research community.

Calculation settings page

This page allows a user to define the settings of the calculation. The „Protonate in pH" box makes it possible to decide which pH value to use to protonate the chosen protein. The pH values can be from 0 to 14. The "**AlphaFold prediction version"** box allows setting the desired database release. Currently, versions v1, v2, v3 and v4 are available; see the details on the AlphaFoldDB FTP page.



Clicking on "Calculate charges" will redirect the user to the Calculation progress page.

Calculation progress page

This page informs a user how the calculation proceeds. It shows the following steps:

Step 1/6: Structure downloaded.

Step 2/6: Structure protonated.

Step 3/6: Molecule loaded.

Step 4/6: Parameters assigned.

Step 5/6: Submolecules created.

Step 6/6: Calculation of partial atomic charges.

It provides also information about the duration of individual steps.

Details about the individual steps:

- Step 1: The structure is downloaded from AlphaFoldDB in the PDB format.
- Step 2: The structure is protonated by PROPKA3 [Olsson2011] at the default pH or

user-defined pH. A PDB file containing hydrogen atoms is created.

- Step 3: The structure is loaded using the python library RDKit (https://www.rdkit.org) and converted into an internal representation.
- Step 4: Relevant SQE+qp [Schindler2021] parameters are assigned to all the atoms.
- Step 5: The molecule is divided into parts (submolecules) according to the cover approach [Ionescu2015]. Partial atomic charges are calculated separately for each submolecule. This approach significantly speeds up the following calculation at the cost of minimal error.
- Step 6: Partial atomic charges are calculated using the SQE+qp empirical charge calculation method [Schindler2021], parameterized using B3LYP/6-31G*/NPA quantum mechanical charges. The parameterization was done specifically for AlphaFoldDB molecules.

Calculation results page

This page includes information about the charge calculation process, visualization of the charged structure and download data section.

Information about the charge calculation process

This information includes the UniProt structure code, its number of atoms, AlphaFold prediction version, pH and calculation time. An example of the information is the following:

UniProt code: A0A1P8BEE7

Number of atoms: 5137

AlphaFold2 prediction version: 4

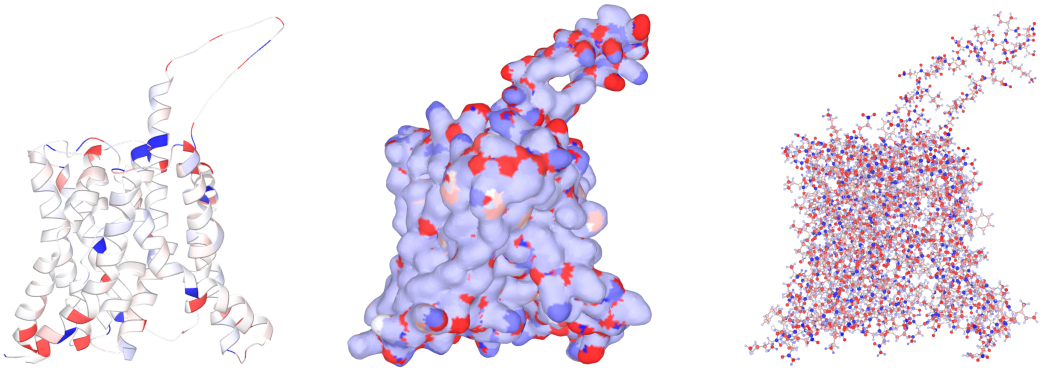
pH: 8.0

Calculation time: 4.63 seconds

Visualization of the structure

αCharges integrates LiteMol suite [Sehnal2017] to show the calculated charges. The user can select between three different charge visualization modes:

- Cartoon mode (Figure 1a), in which the colour of individual amino acids is determined by the sum of charges on its atoms.
- Surface mode (Figure 1b), in which the protein surface is coloured by partial atomic charges of the amino acids, forming the protein's surface. The charges of amino acids are again computed as the sum of charges on their atoms.
- Balls and sticks mode (Figure 1c), where the individual atoms are coloured by their partial atomic charge.



a b C

Figure 1: Visualization of PIN5 protein partial atomic charges (UniProt code Q9FFD0): a) Cartoon mode, b) Surface mode, c) Balls and sticks mode.

The redder the colour, the more negative the charge, and the bluer the colour, the more positive the charge.
The colour scale can be relative (from the lowest to the biggest charge) or absolute (from a user-defined minimal value to a user-defined maximal value).

Results download

The ”Download charges and protonated structure” button allows the user to download a ZIP file containing four files. Two of them contain coordinates of the protonated structure, specifically:

- PDB file: It contains coordinates of the protonated structure in PDB format.
- mmCIF file: It contains coordinates of the protonated protein in mmCIF format.

The remaining two files contain information about charges:

- TXT file: It contains charges of individual atoms present in the protonated structure. The charges are ordered by indexes of the atoms for which they were computed.
- PQR file: It contains information about the coordinates and charges of the protonated structure. The file looks the following way:

```
ATOM      1  N   MET     1      -23.494 -19.088  -2.588 -0.7600 2.0000
ATOM      2  CA  MET     1      -22.509 -19.864  -3.374 -0.1289 2.0000
ATOM      3  C   MET     1      -21.922 -18.915  -4.401  0.6910 1.7000
ATOM      4  CB  MET     1      -23.176 -21.067  -4.061 -0.4312 2.0000
ATOM      5  O   MET     1      -22.699 -18.157  -4.963 -0.6850 1.4000
```

Where values of partial atomic charges are marked **blue**.

Limitations

- **αCharges** cannot compute charges for AlphaFoldDB structures, which contain some error caused by its prediction via the AlphaFold2 algorithm. For example, in the structure with UniProt code Q55GB6, the bond between atoms 269 and 270 goes through an aromatic ring of tyrosin. In this case, **αCharges** reports the following error message:

```
There is an error with atom with index <index of the atom>! The structure is probably incorrectly predicted by AlphaFold, or incorrectly protonated by PROPKA3.
```

- αCharges cannot compute charges for structures which are wrongly protonated by PROPKA3. For example, when the protonation causes the carbon with five hydrogens or some other incorrect protonation. The solution to these errors is now in process (see https://github.com/Electrostatics/pdb2pqr/issues/304). In this case, αCharges reports the same error message as in the previous one.

Bug reporting

If you encounter an error or if you have an idea for an improvement, please send a report to Ondrej Schindler (ondrej.schindler@mail.muni.cz). Thank you!

References

- [Ionescu2015] Ionescu, C. M., Sehnal, D., Falginella, F. L., Pant, P., Pravda, L., Bouchal, T., ..., Koča, J. (2015). *AtomicChargeCalculator: interactive web-based calculation of atomic charges in large biomolecular complexes and drug-like molecules*. **Journal of cheminformatics**, 7(1), 1-13.
- [Jumper2021] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ..., Hassabis, D. (2021). *Highly accurate protein structure prediction with AlphaFold*. **Nature**, 596(7873), 583-589.
- [Olsson2011] Olsson, M. H., Søndergaard, C. R., Rostkowski, M., Jensen, J. H. (2011). *PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions*. **Journal of chemical theory and computation**, 7(2), 525-537.
- [Sehnal2017] Sehnal, D., Deshpande, M., Svobodová Vařeková, R., Mir, S., Berka, K., Midlik, A., ..., Koča, J. (2017). *LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data*. **Nature methods**, 14(12), 1121-1122.
- [Schindler2021] Schindler, O., Raček, T., Maršavelski, A., Koča, J., Berka, K., Svobodová, R. (2021). *Optimized SQE atomic charges for peptides accessible via a web application*. **Journal of cheminformatics**, 13(1), 1-11.
- [UniProt2019] UniProt Consortium. (2019). *UniProt: a worldwide hub of protein knowledge*. **Nucleic acids research**, 47(D1), D506-D515.
- [Varadi2022] Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., ..., Velankar, S. (2022). *AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models*. **Nucleic acids research**, 50(D1), D439-D444.