ZPRED: Zeta Potential Prediction in a General Aqueous Electrolyte Solution

Version 1.0.0

User Manual

Contents

Introduction to ZPRED

ZPRED computes the zeta (or electrokinetic) potential of a macromolecule from its atomic-level structure, as specified by the atomic coordinates of a protein data bank (PDB) file(s) input by the user. The software has been experimentally validated on small, rigid globular proteins; however, it has been developed to handle any molecule as long as the structure is known. ZPRED's computation involves modeling an electric double layer around an input molecular structure and averaging the electric potentials at the solvation edge to define the zeta potential. Essential solution properties (relative dielectric, density, and viscosity) are defined by empirical relations fit to experimentally measured values from other works.

Installation

ZPRED's user interface and software is freely available on GitHub. Please download the repository linked below to get started with the installation:

https://github.com/molecularcollisions/ZPRED

Due to the multiple software components required to run ZPRED, it is essential to carefully execute the following procedure. First, begin with installing the user interface, which can also guide you through the rest of installation

- Open a terminal (Ctrl+Alt+T)
- Go to the downloaded ZPRED folder file location
- Give executable permission to the shell script, "Compile_and_Run.sh"

  ```
  chmod +x Compile_and_Run.sh
  ```
- Run the shell script in the terminal

  ```
  ./Compile_and_Run.sh
  ```

After running the shell script, the ZPRED user interface should appear. On the right hand side, navigate to the "Program File Paths" panel. For each software component, there is an "Installation" button that provides details for its installation. Additionally, a description is provided below for each component.

APBS (Adaptive Poisson-Boltzmann Solver [1]**)** v1.5

- Download APBS 1.5 from the link below:

https://sourceforge.net/projects/apbs/files/apbs/apbs-1.5/APBS-1.5-linux64.tar.gz

- After downloading the compressed APBS package, decompress it in the terminal:

```
tar xzf APBS-1.5-linux64.tar.gz --strip-components 1 -C /usr/local
```

- Install a necessary library:

```
sudo apt-get install libgfortran3
```

- Correct a possible linking error with libreadline.so.6:

```
cd /lib/x86_64-linux-gnu
sudo ln -s libreadline.so.7 libreadline.so.6
```

- Decompress the downloaded APBS package to a known location (such as the Software folder of the downloaded ZPRED package) in order to locate the following executable binaries:
  - APBS-1.5-linux64/bin/apbs
  - APBS-1.5-linux64/share/apbs/tools/bin/multivalue

MSMS (Micheal Sanners Molecular Surface [2]**)** v2.6.1

- Download MSMS 2.6.1 from the link below:

http://mgltools.scripps.edu/downloads/tars/releases/MSMSRELEASE/REL2.6.1/msms_i86Linux2_2.6.1.tar.gz

- After downloading the compressed MSMS package, decompress it, and the following files should be ready:
  - msms.i86Linux2.2.6.1 (MSMS executable)
  - atmtypenumbers
  - pdb_to_xyzr
  - pdb_to_xyzrn

PDB2PQR [3, 4] v1.8

- Download PDB2PQR 1.8 (includes PROPKA 3.0) from the link below:

https://sourceforge.net/projects/pdb2pqr/files/pdb2pqr/pdb2pqr-1.8/pdb2pqr-1.8.tar.gz

- Python and Numpy should be installed with the terminal commands below:

```
sudo apt-get install python-dev
sudo apt-get install python-pip
pip install --upgrade pip
sudo pip install numpy scipy
```

- When installing PDB2PQR, configure without pdb2pka using the following commands:

```
./configure --disable-pdb2pka
sudo make
sudo make install
```

- After installation, the following python file should be ready:
  - main.py

HullRad [5]

- Python code is already included in the ZPRED package and automatically unpacks itself at run-time (see Files/editedHullRad.py computation file)

- **Note**: Python2.7 is required to run this software

HYDROPRO [6] v10 (Optional)

- This hydrodynamics software is optional and was found to perform poorly relative to HullRad. However, it is included for those who wish to compare the two softwares in computing hydrodynamic properties (e.g. translational diffusivity)
- Download HYDROPRO 10 from the link below:

http://leonardo.inf.um.es/macromol/programs/hydropro/hydropro10.zip

- After downloading the compressed package, decompress it, and the HYDROPRO executable file (below) should be ready:
  - hydropro10-lnx.exe

ZPRED Driver Software

- ZPRED C++ source code is already included in the initial download containing the user interface. It can be located in Software/ZPRED/C++/ along with the shell script, "CompileZPRED.sh". To compile the code, the following dependencies must be installed first:
- Boost system and filesystem libraries should be obtained with the terminal command:

  sudo apt-get install libboost-all-dev
- The C++ compiler, g++, should be obtained with the terminal command:

  sudo apt-get install g++
- OpenSSL Crypto library should be obtained with the terminal command:

  sudo apt-get install libssl-dev
- POSIX Pthread library should be obtained with the terminal command:

  sudo apt-get install libpthread-stubs0-dev
- Compile the ZPRED executable (ZPRED.exe) using the included shell script, "CompileZPRED.sh"

Running ZPRED

ZPRED is a driver program that automates the use of the above mentioned software components to compute the electrokinetic (or zeta) potential of a macromolecule. Thus, successful implementation of ZPRED requires that all software components are installed correctly.

To use ZPRED, it is necessary to specify solution conditions (solvent, temperature, ion concentrations, etc.) along with providing a structure or ensemble of structures. Together this information is used to model the collection of solvation layers around the molecule (i.e. the electric double layer) and capture the average electric potential at the molecule's predicted solvation edge.

ZPRED's user interface (shown below) is designed to setup computations for a single molecule in multiple solution conditions. Multiple structure (e.g. PDB) files are treated as an ensemble representing a single, flexible molecule in solution (like folded proteins). It is recommended to always use multiple structure files sampled from a molecular dynamics simulation (e.g. Amber 2015 [7]) allowing the molecule to relax and jiggle between conformations as it would in solution. The number of files submitted for computation should reflect the size and how "jiggly" the input molecule is in solution. In other words, a bigger protein is expected to require more input structure files to represent itself in solution compared to a smaller protein.

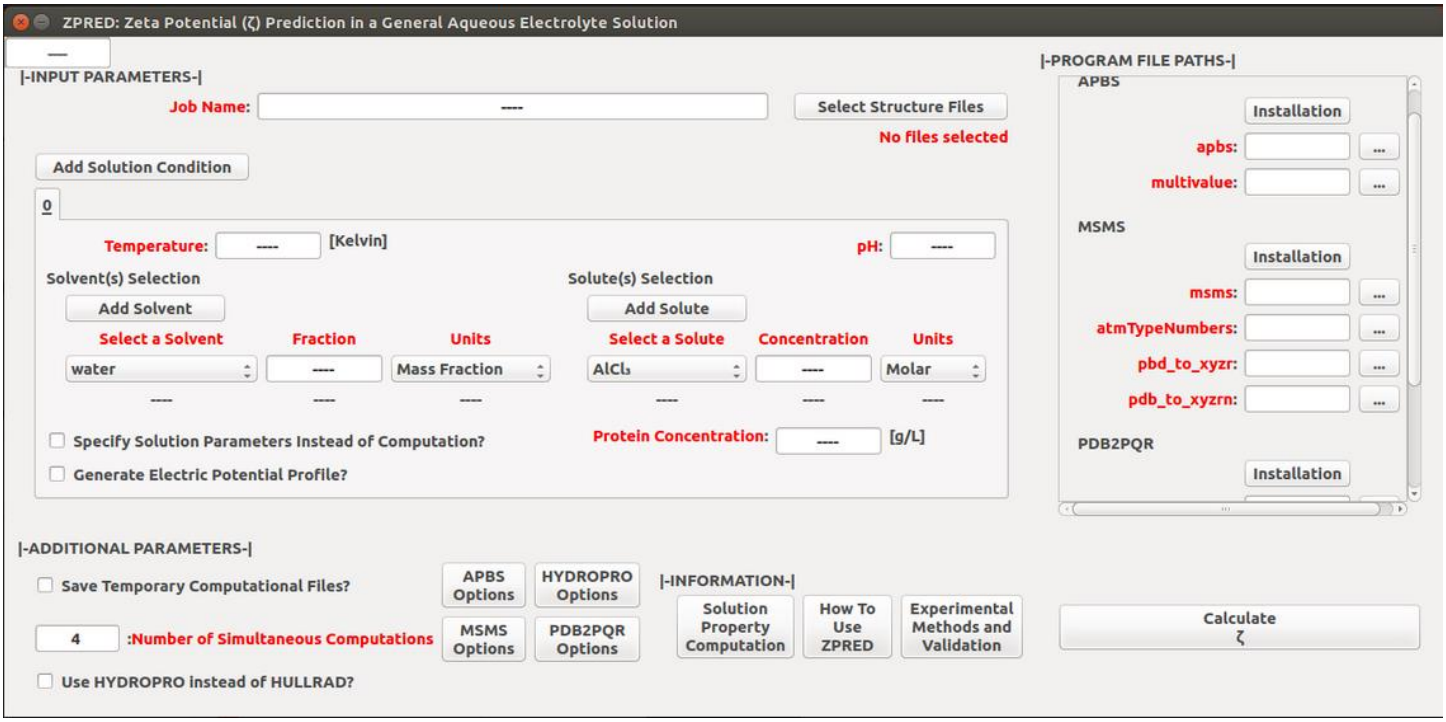

Figure 1. ZPRED User Interface

As shown in the user interface (**Fig. 1**), required input is initially red and turns black once a value is provided. Inputting a value requires pressing Enter on the keyboard each time. A new job name should be specified for each ZPRED computation. An output folder will be generated with this job name and will store all files used during the computation. With the exception of output and log files, the computational files are deleted once computation is complete to save memory. Optionally, these computational files can be saved as well using the check box (Save Temporary Computational Files?) under Addition Parameters. For a description of all these files, refer to the ZPRED Output section. Each ZPRED computation can contain multiple solution conditions, and as previously mentioned, multiple structure files (ZPRED treats multiple files as an ensemble of structures representing a single molecule in solution). More details on the user interface features are described below.

PROGRAM FILE PATHS

• **Auto-Fill** Button	appears after selecting software component file paths for the first time
APBS	
• **apbs** TextInput	APBS executable (located in bin/)
• **multivalue** TextInput	APBS Tool, multivalue, executable (located in share/apbs/tools/bin/)
HYDROPRO (Optional)	
• **hydropro-lnx.exe** TextInput	HYDROPRO executable (hydropro-lnx.exe)
MSMS	
• **msms** TextInput	MSMS executable (msms.i86Linux2.2.6.1)
• **atmtypenumbers** TextInput	MSMS atmtypenumbers file (atmtypenumbers)
• **pdb_to_xyzr** TextInput	MSMS pdb_to_xyzr file (pdb_to_xyzr)
• **pdb_to_xyzrn** TextInput	MSMS pdb_to_xyzrn file (pdb_to_xyzrn)
PDB2PQR	
• **main.py** TextInput	PDB2PQR Python file (main.py)

ZPRED

• **ZPRED.exe** TextInput	ZPRED executable (ZPRED.exe)

INPUT PARAMETERS

• **Job Name** TextInput	specifies the name of the computation and creates a folder for containing the computational files and final output
• **Select Structure Files** Button	specify the file path to one or more structures to use in the computation
• **Add Solution Condition** Button	adds a solution condition to include in the computation
• **Remove Solution Condition** Button	removes the last added solution condition from the computation
• **Temperature** TextInput	specifies the absolute temperature of the solution
• **pH** TextInput	specifies the pH of the solution (be sure to include acid and/or base in the solutes selection)

Solvent(s) Selection

• **Add Solvent** Button	adds another solvent to be included in the computation
• **Select a Solvent** DropdownBox	[default: water] specifies the solvent to use in the computation
• **Fraction** TextInput	specifies the fraction (mole, mass or volume) that the solvent occupies (e.g. a fraction of 1 means the solvent occupies the entire volume)
• **Units** DropdownBox	specifies the fraction type (mole, mass or volume)

Solute(s) Selection

• **Add Solute** Button	adds a solute to be included in the computation
• **Select a Solute** DropdownBox	specifies the solute dissolved in the solvent
• **Concentration** TextInput	specifies the concentration of the solute
• **Units** DropdownBox	specifies the concentration type (molar or molal)
• **Protein Concentration** TextInput	specifies the protein's mass concentration in solution in units of grams/Liter
• **Specify Solution Parameters Instead of Computation?**	allows user to specify solution properties rather than calculating them with an empirical relation derived from experimental data
• **Generate Electric Potential Profile?**	outputs the computed electric potential profile from the molecular surface for the specific solution condition

ADDITIONAL PARAMETERS

• **Save Temporary Computational Files?**	specifies whether or not to save all files used in the computation (can require large amounts of memory)
• **Number of Simultaneous Computations** TextInput	specifies the number of parallel ZPRED computations to execute simultaneously

• **Use HYDROPRO instead of HULLRAD?**	specifies the use of HYDROPRO hydrodynamic software rather than the superior HullRad (for comparison purposes)
• **APBS Options** Button	allows user to adjust critical APBS parameters
• **HYDROPRO Options** Button	allows user to adjust HYDROPRO calculation type
• **MSMS Options** Button	allows user to adjust mesh size of generated solvent-excluded surfaces
• **PDB2PQR Options** Button	allows user to adjust the force field used for determining charge

INFORMATION

• **Solution Property Computation** Button	provides information about the empirical relations used to model pertinent solvent properties (relative dielectric, density, viscosity)
• **How to Use ZPRED** Button	provides information about using ZPRED
• **Experimental Methods and Validation** Button	provides information on the experimental validation of the software

As ZPRED carries out its computation, progress can be seen in the terminal used to start the user interface. An example of the progress display is shown below.

Figure 2. ZPRED Computation in Progress

Although much debugging has been performed, the progress display can sometimes be faulty. This does not always indicate an error occurred, but rather the error output stream is picking up additional signals from one of the software components. To view an actual error and each software components' output, the log file (located in Files/LOG/) for the specific computation should be carefully viewed to trace the error. With no errors, completion of ZPRED computation looks like that shown below in **Fig. 3**.

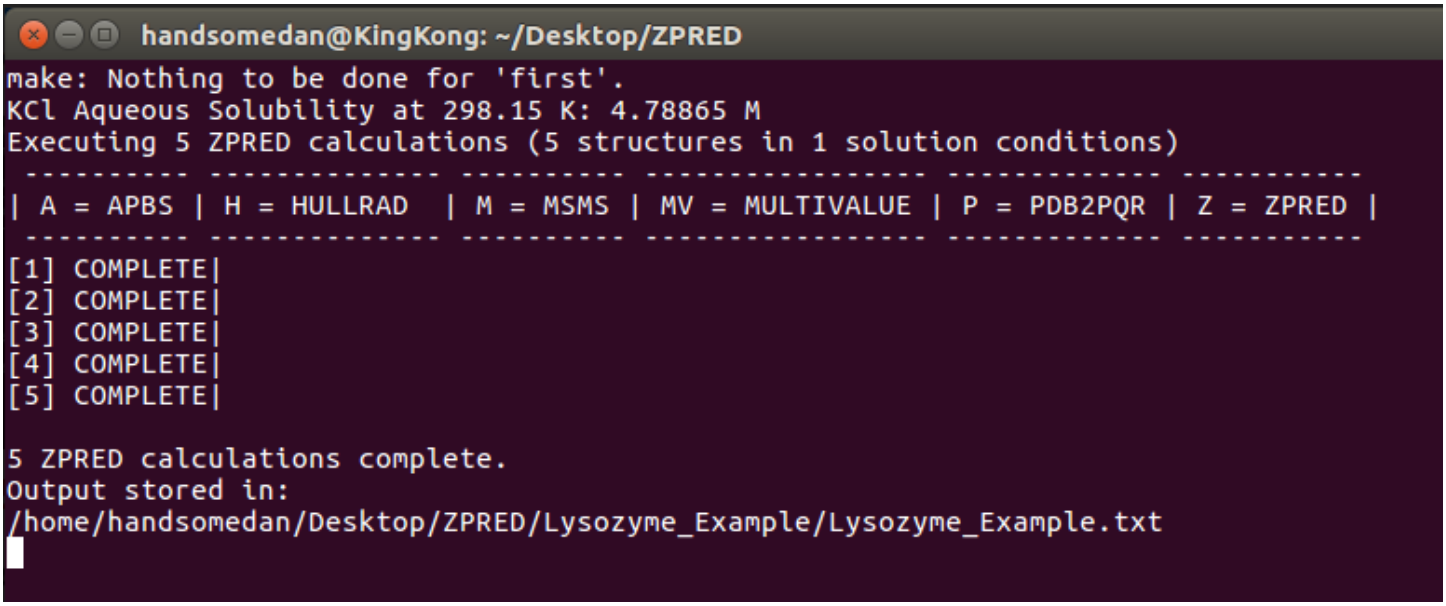

Figure 3. ZPRED Computation Complete

ZPRED Output

Upon successful completion of a computation, a new folder with the computation's job name will have been generated containing a variety of files. To store all files used during the computation, you must check the option (Save Temporary Computational Files?) before running ZPRED via the user interface. An overview of all computational files (including the output) is presented below in reverse order of their creation. File organization is also presented with folders in bold, important files in red and user provided variables in brackets.

ZPRED File Organization

<Job Name>
- **<Job Name>.txt** — concise output
- **Files**
  - **DX**
    - **<PDB File Name>_<Solute Concentration>**
      - **pot**
        - **pH<pH>T<Temperature>SR<Ion Radius>.dx** — APBS generated electric potentials file
  - **HP**
    - **<PDB File Name>_<Computation Number>_pH<pH>_T<Temperature>_<Salt Concentration>**
      - **hullrad_output_<Computation Number>.txt** — hydrodynamic properties output file
  - **IN**
    - **<PDB File Name>_<Solute Concentration>**
      - **pH<pH>_T<Temperature>_SR<Ion Radius>_SD<Solution Dielectric>.in** APBS input command file
  - **LOG**
    - **<PDB File Name>**
      - **<Computation Number>.log** — ZPRED log file
  - **msms**
    - **<PDB File Name>_<Computation Number>_pH<pH>_T<Temperature>_<Salt Concentration>**
  - **PQR**
    - **<PDB File Name>**
      - **pH<pH>_<Computation_Number>.pqr** — PDB2PQR generated PQR file
  - **PROPKA**
    - **<PDB File Name>**
      - **pH<pH>_<Computation_Number>.propka** — PROPKA generated propka file
  - **S**
    - **<PDB File Name>_<Solute Concentration>**
      - **CGO**
      - **colorCGO**
      - **CSV**
      - **fromMSMS**
      - **fromMV**
      - **SAS**
  - **Z**
    - **<PDB File Name>_<Solute Concentration>**
      - **pH<pH>T<Temperature>PC<Protein Concentration>_<Computation_Number>.txt**detailed output

Concise Output

```
Solution Condition #0
pH 7.00 298.15 K
100 mol% water
0.01 M KCl
Relative Dielectric: 78.18514375
Density: 0.997511251894 kg/L
Viscosity: 0.000890087892 Ns/m^2
Hydration Layer Thickness: 2.3 A
Protein Concentration: 3 g/L
-----------------------------------------------------------------------------------------------------------------------|
|||||||||-|-|-|-|SEM-----------|Henry---------|Kuwabara------|--------------|
|||||||||Anhydrous------|Solvated-------|Zeta----------|-|Electro.------|Electro.------|Electro.------|--------------|
|||||||||Radius---------|Radius---------|Potential-----|Q|Mobility------|Mobility------|Mobility------|Diffusivity---|
||||||||||[A]----------|[A]----------|[Volts]-------|-|[umcm/Vs]-----|[umcm/Vs]-----|[umcm/Vs]-----|[m^2/s]-------|
6LYZ_1   |16.354387666567|18.728915492229|0.039485522497|8|1.69979405799 |2.08723765704 |2.016198598557|0.000000000131|
6LYZ_5   |16.4471026131  |18.728915492229|0.038466817887|8|1.658963022083|2.033388081608|1.964461039458|0.000000000131|
6LYZ_4   |16.44337760032 |18.728915492229|0.038378579405|8|1.655415304861|2.028723721844|1.960712681864|0.000000000131|
6LYZ_3   |16.5970258389l3|18.872984072938|0.035383349163|8|1.53312090127 |1.870763872432|1.80831161914 |0.00000000013 |
6LYZ_2   |16.569960535741|18.872984072938|0.035894212825|8|1.553950418081|1.897773901284|1.834701917959|0.00000000013 |
-----------------------------------------------------------------------------------------------------------------------
Average  |16.482370850928|18.786542924513|0.037521696355|8|1.620248740857|1.983577446842|1.916877171396|0.000000000131|
Std.Dev. |0.09995215959  |0.078909611482 |0.001782228082|0|0.072542519095|0.094013131612|0.090263189874|0.000000000001|
-----------------------------------------------------------------------------------------------------------------------|
```

Figure 4. ZPRED Output File (Concise View)

- located in generated output folder and named <Job Name>.txt
- contains values from all structure files at all input solution conditions
- provides solution properties used during computation
- provides output table with pertinent values (solvation layer thickness, zeta potential, diffusivity)
- also, results from three different generalized electrokinetic models provide electrophoretic mobility values that can be compared with experimental values determined by free electrophoresis:
  (**1**) Standard Electrokinetic Model (SEM) [8]
  (**2**) Henry Model (recommended for dilute protein solutions) [9]
  (**3**) Kuwabara Model [10, 11]
- if selected (see below for details), contains the generated average electric potential profile

Detailed Output

```
////////////////////////////////////////////////////
// Physical Shape Descriptor
.shape: sphere
// Asphericity
.asphericity: 0.057930133299
// Shape Parameter
.shapeParameter: 0.024088462504
// Gyration Radius [A]
.gyrationRadius: 14.352406255665
// Solution Condition Number
.solCondNum: 0
// Name of Protein Conformation
.proteinName: 6LYZ_3
// Anhydrous Protein Radius [A]
.proteinRadius: 16.597025838913
// Solvate (Hydrated) Protein Radius [A]
.solvatedRadius: 18.572854072935
// Protein Diffusivity [m^2/s]
.diffusivity: 0.000000000013
// Solution Debye Length [A]
.debyeLength: 30.359499574551
// Dimensionless Electrokinetic Radius
.kn: 0.6216500321468
// Henry Correction Factor for Electrophoretic Retardation
.henryCorrection: 0.6797985122133
// Protein Net Valence [e]
.charge: 8
// Hydration Layer Thickness [A]
.Xsp: 2.25
// Electric Potential (2.25 A from surface) [V]
.zetaPotential: 0.0353833349163
// Absolute Temperature [K]
.temperature: 298.15
// Solution pH
.pH: 7.00
// Solution Viscosity [Ns/m^2]
.viscosity: 0.000890057892
// Solution Relative Dielectric
.dielectric: 78.15514375
// Solution Density [kg/L]
.density: 0.997511251894
// Solvent Components (Solvent Names)
.solvent: water;
// Solvent Components (Solvent Concentrations) [Mole Fraction]
.solventConc: 1.000000000000;
// Solvent Components (Solute Names)
.solute: KCl;
// Solvent Components (Solute Concentrations) [Molar]
.soluteConc: 0.010000000000;
// Specific Volume [L/g]
.specificVolume: 0.000725514457
// Volume Fraction
.volumeFraction: 0.002176543572
// Electrophoretic Mobility [umcm/(Vs)] Defined from ZPRED computed zeta potential and Henry Correction Factor
.henryMobility: 1.870763572432
// Electrophoretic Mobility [umcm/(Vs)] Defined from ZPRED computed zeta potential and Kuwabara Correction Factor
.kuwabaraMobility: 1.805311161914
// Kuwabara Correction Factor for Particle Concentration
.kuwabaraCorrection: 0.8571042224129
// Electrophoretic Mobility from Henry Equation [umcm/(Vs)]
.henryModelMobility: 2.545305241524
// Electrophoretic Mobility from O'Brien & White Algorithm (solves standard electrokinetic model (SEM)) [umcm/(Vs)]
.semMobility: 1.533120901127
// Electrophoretic Mobility from O'Brien & White Algorithm with Kuwabara Correction Factor [umcm/(Vs)]
.semMobility2: 1.481940251276
////////////////////////////////////////////////////
kn:     0.6216500321468
Km:     1.404305762234
Compare Km to:  1.2055501312332
Chahine Approximation Appropriate!!!
Chahine Correction: 0.8609785212205
Electrophoretic Mobility: 1.5159742778126 umcm/(Vs)
// Electric Potential Profile from Protein Surface
.generateProfile: false
```

Figure 5. ZPRED Output File (Detailed View)

- located in /Files/Z/ folder and named based on the structure file and solution condition used during computation
- contains value for single structure file in single solution condition
- contains everything in the concise view plus more
- shape assessment (is molecule cylindrical or spherical?)
- electrophoretic effects assessment

Notes

For optimal results, it is essential to use the exact same solution conditions used during experimental measurements and to use PDB files that truly represent the molecule in solution.

Literature References

[1] N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, "Electrostatics of nanosystems applications to microtubules and the ribosome," *Proceedings of the National Academy of Sciences,* vol. 98, no. 18, pp. 10037-10041, 2001.

[2] M. F. Sanner, A. J. Olson and J.-C. Spehner, "Reduced Surface: An Efficient Way to Compute Molecular Surfaces," *Biopolymers,* vol. 38, no. 3, pp. 305-320, 1996.

[3] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations," *Nucleic Acids Research,* vol. 32, pp. W665-W667, 2004.

[4] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe and N. A. Baker, "PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations," *Nucleic Acids Research,* vol. 35, 2007.

[5] P. J. Fleming and K. G. Fleming, "HullRad: Fast Calculations of Folded and Disordered Protein and Nucleic Acid Hydrodynamics Properties," *Biophysical Journal,* vol. 114, pp. 856-869, 2018.

[6] J. G. de la Torre, M. L. Huertas and B. Carrasco, "Calculation of hydrodynamic properties of globular proteins from their atomic-level structure," *Biophysical Journal,* vol. 78, pp. 719-730, 2000.

[7] D. A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. M. Merz, G. Monard, P. Needham, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, R. Salomon-Ferrer, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, D. M. York and P. A. Kollman, "AMBER 2015," San Francisco, 2015.

[8] R. W. O'Brien and L. R. White, "Electrophoretic Mobility of a Spherical Colloidal Particle," *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics,* vol. 74, pp. 1607-1626, 1978.

[9] D. C. Henry, "The Cataphoresis of Suspended Particles. Part I. The Equation of Cataphoresis," *Proceedings of the*

Royal Society A, vol. 133, pp. 106-129, 1931.

[10] S. Kuwabara, "The forced experienced by randomly distributed parallel circular cylinders or spheres in a viscous flow at small Reynolds numbers," *Journal of the Physical Society of Japan,* vol. 14, no. 4, pp. 527-532, 1959.

[11] S. Levine and G. H. Neale, "The Prediction of Electrokinetic Phenomena within Multiparticle Systems I. Electrophoresis and Electroosmosis," *Journal of Colloid and Interface Science,* vol. 47, no. 2, pp. 520-529, 1974.