HYDE Commandline Documentation Version 2.0

Sascha Jung & Marcus Gastreich

May 12, 2023

Contents

1 Introduction

All links, references, table of contents lines etc. in this pdf are clickable.

HYDE is a tool for scoring protein-ligand complexes at the commandline. HYDE is our award-winning, desolvation-aware scoring function.[1] It is neither trained on particular data nor calibrated to specific targets. It is fully based on physical principles, namely the two major driving forces of binding in aqueous solution, desolvation and interactions, and combines both in a scientifically rigid manner. As a result, it calculates realistic total binding affinities (free energies of binding) of a ligand to its target. Additionally, the affinity contributions can also be broken down to individual atoms (per-atom contributions), revealing those molecule parts with favorable and unfavorable binding contributions.

HYDE is also a "component" within our flagship 3D modeling platform SeeSAR that has been conceived for drug researchers of any discipline and educational level (`https://www.biosolveit.de/SeeSAR`). Within SeeSAR, you can easily score your results with HYDE after every step of your design process throughout the various Modes. Additionally, the per-atom contributions are visualized as green or red coronas which makes it straightforward to identify molecule parts that boost the affinity of your ligand. This enables you to easily rationalize the results and select compounds for further investigation, all under full visual control.

HYDE at the commandline features:

- scoring and optimization of protein-bound ligands
- calculation of estimated binding affinities
- calculation and optimization of H-bond networks
- consideration of pharmacophore constraints as post-filter
- processing of unlimited ligand numbers

Please note that this package is a commandline package.

2 Technical Prerequisites

2.1 Required Software

HYDE is a commandline application. Control through a graphical user interface (GUI) is available in SeeSAR, BioSolveIT's 3D flagship modeling suite. We highly recommend you also install SeeSAR to swiftly carry out all sorts of preparational steps for the receptor; the preparation can then be written to a `*.hydescorer` file and used at the commandline with HYDE (this package).

Technically, you will need:

- The HYDE **application package** (from `https://biosolveit.de/download`). Depending on your operating system, some libraries may have to be installed. Get in touch with us if that is the case: `mailto:support@biosolveit.com`. Please mention any errors/warnings that you see in your mail.
- A **shell** (Linux/Unix) or a terminal (macOS), or a commandline environment (Windows; e.g.: cmd.exe or PowerShell)
- A valid **license** (from `mailto:license@biosolveit.de`), see below.

2.2 Licensing

HYDE needs a license to operate which is available from us. There are various sorts of licenses, but in most of the cases your early testing will employ a license file that simply needs to be put next to the executable, see below.

The license setup instructions will come with the license that we will send out — or that has already been sent out to you. In case you do not have a license yet, please get in touch with us at `mailto:contact@biosolveit.de`, and provide us with the necessary information. Please note: a SeeSAR license will be read as "valid" by HYDE.

License File Locations A "test license" that you can request online and that is sent to you instantaneously can simply be placed next to the executable (`hydescorer.exe`, `Hydescorer` or `hydescorer` — depending on your operating system). For macOS please read on...

macOS Specialties On MacOS, the executable will typically reside inside the *.app package:

```
/Applications/Hydescorer.app/Contents/MacOS/Hydescorer
```

To place the short term test license there, you will have to go into the *.app package using a right mouse click (or CTRL-click) on Hydescorer.app in the Finder, and click on "Show package contents". In there, you will see the Contents/ subfolder, in there the MacOS subfolder, and in there, the Hydescorer executable. If you are about to use the **test license**, place is right there, next to the executable. A longer term license will be handled separately, we will tell you how when we send that very license.

When you call Hydescorer for the first time, go to the Finder, and navigate to the Applications folder. Do a right(!) click on Hydescorer.app, and — if applicable — confirm that you want to open the program. It will flash up once, and you are good to go at the terminal prompt from there on.

Obtaining a License File Using `--license-info` you can obtain information about the specification of your license server machine, the searched directories, and the validity of the currently used license files. This may also be useful when Hydescorer is not starting up as you would expect it to.

Call Hydescorer with the `--license-info` option, to see an output like this:

```
./hydescorer --license-info
================================================================================
 License Information:
================================================================================
Host-ID: "28f0763fa408 38c98636b9cb"
BIOSOLVE_LICENSE_FILE: /Applications/BioSolveIT/License
LM_LICENSE_FILE:

Currently used key/path after environment variables:
/software/hydescorer-2.0.0/hydescorer

Hydescorer:
 >> Valid key, expires on Thursday, 30 June 2023

Request a license:
***
*** https://www.biosolveit.de/license/?product=hyde&operating_system=linux...
***
```

Request an evaluation or longer-term license using the link that is provided at the very bottom of the output. Also, this output may help us to find out if there are any problems with your license or its setup.

3 Help

An overview of all commmandline options is available by calling Hydescorer with `--help`; alternatively get help for a command with a tailing `-h` or `--help`. Default vaues are bracketed:

```
./hydescorer  -h

 Program options:
-i [ --input ] arg            Library input molecule file. Supported file types are *.mol, *.mol2 and
                              *.sdf.
                              Note: 3D coordinates must be provided, otherwise molecule is skipped.
-o [ --output ] arg           Output file (suffix is required). Only '.sdf' is supported.
-p [ --protein ] arg          Protein file. Only '.pdb' is supported.
                              Note: Can't be used together with '--binding-site-definition'.
-r [ --refligand ] arg        Reference ligand file. Supported file types are *.mol, *.mol2 and *.sdf.
                              Note: The reference ligand must have 3D coordinates and lie in the pocket in
                              the protein as it is used to determine the binding site for scoring.
                              Note: Can't be used together with '--binding-site-definition'.
--binding-site-definition arg Runs hyde scoring on the basis of this binding site definition file (a
                              .hydescorer file that can be exported from SeeSAR's Docking or Binding Site
                              mode) - it contains the protein with the predefined binding site and
                              conserved waters.
                              Note: Can't be used together with '--protein' or '--refligand'.
--hbond-only [=arg(=1)] (=0)  Runs H-bond network optimization only.
                              Note: No hyde scores are available.
--write-protein arg           Output base file name for writing protein with solution. Only '.pdb' is
                              supported.
                              Note: For each solution, a separate file is created.

General options:
-h [ --help ]                 Print this help message
--license-info                Print license info
--thread-count arg            Maximum number of threads used for calculations. The default is to use all
                              available cores.
--version                     Print version info
-v [ --verbosity ] arg (=2)   Set verbosity level
                                   0 [silent]
                                   1 [error]
                                   2 [warning]
                                   3 [workflow]
                                   4 [steps]
```

Please note that the abbreviated, one-letter options are preceded with one dash – whereas the longer, named options are preceded with two dashes: `--`.

In addition, we are available to support your endeavors with HYDE by email (`mailto:support@biosolveit.de`). We try to answer within a day, during business hours.

4 Jump Start: A Swift HYDE Scoring

To run a HYDE scoring at the commandline with all defaults, you will need at least:

- An input file with compounds that have 3D coordinates in the binding pocket of the protein (typically results from prior docking with FlexX or a native ligand from the crystal structure).

(a) Protein and reference ligand

- A protein input file (`.pdb`) **and**
- An input ligand file (i.e., the so-called reference ligand) with 3D coordinates in a binding cavity of the input protein (`.sdf`, `.mol` or `.mol2` format). The reference ligand is needed to define the binding site.

(b) A binding site definition file

- A binding site definition file (`.hydescorer`) that you have exported from within SeeSAR's Docking, Binding Site Mode or Analyzer Mode.

Now run this — with the file names replaced by your own names of course:
In case (a), that is, if you provided a protein file and a reference ligand file:

```
./hydescorer -i MyDockingResults.sdf -o MyHYDEoutput.sdf
             -p MyProtein.pdb -r My3DReferenceLigand.sdf
```

In case (b), that is, if you have prepared and exported a binding site definition file from SeeSAR:

```
./hydescorer -i MyDockingResults.sdf -o MyHYDEoutput.sdf
             --binding-site-definition MyDefinitionFile.hydescorer
```

Depending on the operating system you use, please certainly also adapt the commandline usage, e.g., use a slash or a backslash etc.

The above calls run the HYDE scoring and create the output file `MyHYDEoutput.sdf` that contains the HYDE-optimized pose for every input ligand. The SD output

file will also contain the respective HYDE estimated affinity range (SD property fields named `<BIOSOLVEIT.HYDE_ESTIMATED_AFFINITY_LOWER_BOUNDARY>` and `<BIOSOLVEIT.HYDE_ESTIMATED_AFFINITY_UPPER_BOUNDARY>`). Additionally, there are also information on clashes and torsion quality available (see Section 7 for more information).

5 Program Options

`-i` / `--input (arg)`
Specify the ligand input file in `.sdf`, `.mol` and `.mol2` format. The input file should contain all the compounds which you want to score. Please make sure that all molecules have 3D coordinates and lie in the very binding pocket of the protein that you either specified with the `-p` option and a suitable reference ligand via the `-r` option *or* with the `--binding-site-definition` option, see below. Otherwise, the molecules cannot be processed.

`-o` / `--output (arg)`
Specify a name for the output file (`.sdf` format, suffix is required) containing the HYDE-optimized poses. The output file will contain all successfully optimized and scored molecules from the input file along with detailed information on the affinity range and quality of the respective pose (see Section 7).

`-p` / `--protein (arg)`
Specify a file containing the protein in `.pdb` format. If you do so, please make sure you also specify a reference ligand with the `-r` option (necessary to define the binding pocket, see below). The `-p` option is obsolete if you have prepared a binding-site definition file (see `--binding-site-definition` option below).

`-r` / `--refligand (arg)`
Provide a file containing the reference ligand in `.sdf`, `.mol` and `.mol2` format. The reference ligand will be used to determine the binding site for scoring. Therefore, it is important you make sure the reference ligand has 3D coordinates and occupies a pocket of the protein which you specified via the `-p` option. To define the binding pocket, the Hydescorer automatically detects all protein residues which have at least one atom lying in a 6.5 Å sphere around the reference ligand atoms (see Figure 1 for an example). The `-r` option is obsolete if you have prepared a binding-site definition file (see below).

`--binding-site-definition (arg)`
As an alternative way to using a separate protein and reference ligand file (see above) you can run a HYDE scoring on the basis of a binding site definition file (`.hydescorer` file). The `.hydescorer` file must have been exported from SeeSAR's Docking, Binding Site or Analyzer Mode and contains the protein with

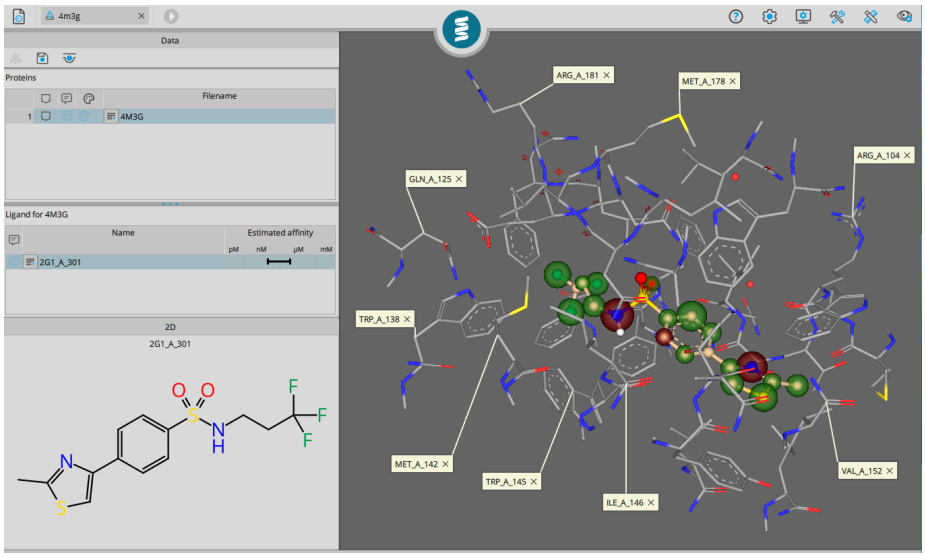

Figure 1: *Example for the automated binding site definition on the basis of a reference ligand (PDB 4M3G). All protein residues which have at least one atom lying in the 6.5 Å sphere defined around every ligand heavy atom are included.*

a predefined binding site. Additionally, it may also contain pharmacophore constraints which serve as post-filter to eliminate irrelevant results. The output file will only contain those results which fulfill the constraints. Pharmacophore constraints can be defined and exported either in SeeSAR's Docking or Analyzer Mode. Please note that this option cannot be used together with the `-p` and `-r` option.

`--hbond-only`
If this option is specified, only the calculation and optimization of the hydrogen bond network is performed for every input ligand. No HYDE scoring and pose optimization is executed, e.g. no HYDE scores are available in the output file.

`--write-protein (arg)`
Specify an output base file name for writing the HYDE-optimized poses in complex with the protein (`.pdb` and `.ent` are supported). Suffixes are not required, a `.pdb` file is written by default. Please note: A separate file is written for every molecule in the input ligand file (if the optimization was successful). The optimized protein-ligand complexes are written *in addition* to the ligand-only output file (`-o` option).

6 General Options

`-h` / `--help`
Displays the commandline help with short descriptions for every argument option. For more information see Section 3.

`--license-info`
Displays detailed information about the license setup you currently use. For more information see Section 2.2.

`--thread-count (arg)`
Specifies the maximum number of threads used for the HYDE scoring. By default, all available logical cores of your computer are used. You may want to reduce the number of threads used if you want to run other computations on your computer at the same time, or if you share the compute resource. Takes an integer as argument.

`--version`
Displays information on the version of HYDE on the commandline. In quoting HYDE, please mention this version number.

`-v` / `--verbosity (arg)`
You can set the verbosity level, e.g. the level of console output, by giving an integer as argument. The default value is 2. The following options are available:

- `0` Silent. No messages will be displayed in the console during HYDE run. Errors will be ignored whenever possible.
- `1` Error. Only error messages leading to failures of the HYDE scoring will be displayed.
- `2` Warning. The default setting. Warnings and error messages will be displayed.
- `3` Workflow. In addition to errors and warnings, the different steps of the HYDE optimization are displayed.
- `4` Steps. In addition to the 'Workflow' option, the progress of HYDE scoring is displayed in detail.

7 HYDE Output

The HYDE output file (`.sdf`) contains the optimized ligand poses along with detailed information on estimated affinity ranges, pose quality and druglikeness:

- `BIOSOLVEIT.HYDE_ATOM_SCORES [kJ/mol]`:
  Detailed information on the per-atom HYDE scores. For every atom, the total atom contribution (Total), the receptor desolvation contribution (Rec-Desolv), the receptor interaction contribution (Rec-IA), the ligand desolvation contribution (Lig-Desolv) and the ligand interaction contribution (Lig-IA) are listed.
- `BIOSOLVEIT.HYDE_ESTIMATED_AFFINITY_LOWER_BOUNDARY [nM]`:
  Lower boundary for the estimated binding affinity in nanomolar concentration.
- `BIOSOLVEIT.HYDE_ESTIMATED_AFFINITY_UPPER_BOUNDARY [nM]`:
  Upper boundary for the estimated binding affinity in nanomolar concentration. The "real" affinity is predicted to be in the range between lower and upper boundary.
- `BIOSOLVEIT.HYDE_LIGAND_EFFICIENCY range: ++,+,0,-,--`:
  Estimated ligand efficiency range calculated based on the predicted affinity range. Range is gradually from very high ligand efficiency (++) to very low efficiency (- -).
- `BIOSOLVEIT.HYDE_LIGAND_LIPOPHILIC_EFFICIENCY range: ++,+,0,-,--`:
  Estimated lipophilic ligand efficiency range calculated based on the predicted affinity range and the calculated logP. Range is gradually from very high efficiency (++) to very low efficiency (- -).
- `BIOSOLVEIT.INTER_CLASH range: red, yellow, green`:
  Rating of clashes between ligand and protein atoms. Red = one and more major/intolerable clashes between ligand and protein atoms *or* more than two medium/tolerable clashes; yellow = maximum of two medium/tolerable clashes; green = no clashes between ligand and protein atoms.
- `BIOSOLVEIT.INTRA_CLASH range: red, yellow, green`:
  Rating of clashes between ligand atoms (intra clashes). Red = one and more major/intolerable intra clashes *or* more than two medium/tolerable intra clashes; yellow = maximum of two medium/tolerable intra clashes; green = no clashes between ligand atoms.
- `BIOSOLVEIT.LOGP`:
  Calculated logP of the molecule.

- `BIOSOLVEIT.MOLECULAR_WEIGHT`:
  Calculated molecular weight of the molecule in g/mol.
- `BIOSOLVEIT.MOLECULE_CHECKSUM`:
  Checksum for internal validation.
- `BIOSOLVEIT.TORSION_QUALITY range: red, yellow, green, not rotatable`:
  Rating of the torsion quality of the pose. Red = more than one unfavorable/intolerable torsion; yellow = maximum of one unfavorable torsion and maximum of two unfavorable but still tolerable torsions *or* more than two unfavorable but still tolerable torsions; green = maximum of two unfavorable but still tolerable torsions; not rotatable = no rotatable bonds in the molecule, no torsion quality assessment possible.
- `BIOSOLVEIT.TPSA`:
  Calculated topological polar surface area of the molecule in $Å^2$.

8 Further Reading, References, How to Cite

The original ideas behind the HYDE scoring function have been cited many times; they are covered in the original publication by Reulecke and Rarey.[1] HYDE is constantly improved in a collaboration with BAYER, Hamburg University, and BioSolveIT.[2]

Additional information on the tool is available at

`https://biosolveit.de/products/#HYDE`.

Cite HYDE as:

```
HYDE Version 2.0.0, BioSolveIT GmbH,
St. Augustin, Germany, 2023, biosolveit.de/HYDE
```

Complementary tools, especially also the graphical platforms SeeSAR and infiniSee, can be obtained from the BioSolveIT website (`https://biosolveit.com`).

We wish you great success and much joy with HYDE!

References

[1] Ingo Reulecke, Gudrun Lange, Jürgen Albrecht, Robert Klein, and Matthias Rarey. Towards an integrated description of hydrogen bonding and dehydration: Decreasing false positives in virtual screening with the hyde scoring function. *ChemMedChem*, 3:885–897, 6 2008.

[2] Nadine Schneider, Gudrun Lange, Sally Hindle, Robert Klein, and Matthias Rarey. A consistent description of hydrogen bond and dehydration energies in protein-ligand complexes: Methods behind the hyde scoring function. *Journal of Computer-Aided Molecular Design*, 27:15–29, 2013.