



Daylight Chemical Information Systems, Inc.
120 Vantis - Suite 550 - Aliso Viejo, CA 92656
tel +1 949-831-9990 - fax +1 949-831-9902 - www.daylight.com

Daylight Web Services Manual

Table of Contents

<u>Daylight Web Services Manual</u>	1
<u>1. Introduction</u>	1
<u>2. Prerequisites</u>	1
<u>2.1 Daylight Software Requirements</u>	1
<u>2.2 Third-Party Software Requirements</u>	1
<u>2.3 Installation</u>	2
<u>3. Web Services</u>	2
<u>3.1 canonicalizeSmiles</u>	2
<u>3.2 convertStructure</u>	2
<u>3.3 deriveScaffold</u>	5
<u>3.4 deriveSDClusters</u>	6
<u>3.5 getProperties</u>	7
<u>3.6 getDepiction</u>	8
<u>3.7 getTransform</u>	10
<u>3.8 getTautomers</u>	10
<u>3.9 desaltSmiles</u>	11
<u>3.10 normalizeSmiles</u>	12
<u>3.11 getClogP</u>	12
<u>3.12 generateRTable</u>	13
<u>3.13 executeProgram</u>	13

Daylight Web Services Manual

Daylight Version 4.9
Release Date 08/01/11

Copyright Notice

This document and the programs described herein are Copyright © 2007-2011, Daylight Chemical Information Systems, Inc., Aliso Viejo, CA. Daylight explicitly grants permission to reproduce this document under the condition that it is reproduced in its entirety including this notice, and without alteration. All other rights are reserved.

1. Introduction

Web services use standard, open protocols to provide access to a wide range of programs over a network. Passing parameter data along with a request for a particular service triggers an action and sends back a response. Thus the web services model provides a valuable mechanism for delivering complex chemistry-oriented functionality such as format conversion or property calculation within an organization.

Daylight offers a series of application components such as canonicalization and depiction as Java web services for use on their own or for inclusion in user-designed or workflow applications. The "Web Services" access Daylight application libraries and toolkits written in C through the Java Native Interface (JNI) framework. SOAP is used as the messaging format for all currently available Daylight Web Services.

2. Prerequisites

No particular programming skills are required for use of the Web Services. However, installation and set-up requires a general knowledge of UNIX and Daylight software.

2.1 Daylight Software Requirements

Web services are included with the standard Daylight distribution (versions 4.93 or later). The standard distribution is available for download from Daylight's web site (<http://www.daylight.com>). In order to use any of the web services, an appropriate Daylight license for each particular web service is required. Note: The server code will only run on supported Solaris and Linux platforms.

2.2 Third-Party Software Requirements

The following third-party software packages are required for the server:

Java JDK 5 - <http://www.sun.com/java/>

Tomcat 5.5 - <http://tomcat.apache.org/>

Axis 1.4 (testing only) - <http://tomcat.apache.org/>

Note: Only 32-bit versions of the web services are currently available.

2.3 Installation

See the [Daylight Installation Manual](#) for specific instructions on setting up the web services.

3. Web Services

The following sections describe the currently available Web Services. The response objects for all services contain the content of the response, processing or error messages, or both. When options are provided to web services in the form of a list of alternating strings of names and values, the general rule is that a repeated name will have its last value used. All names must have associated values.

Each web service publishes a Web Services Description Language (WSDL) file which represents the definitive specification of all the inputs and outputs (including exceptions) for each service. A copy of the WSDL in \$DY_ROOT/webservices.

All of the Web Services will optionally report errors generated during an action as part of returned message when an ERRORLEVEL input parameter is supplied.

Standard error levels are as follows:

- 0 = no messages returned
- 1 = warnings, notes and errors returned
- 2 = warnings and errors returned
- 3 = all errors returned
- 4 = serious errors only

3.1 canonicalizeSmiles

This Web Service parses a list of molecules or reactions and generates the corresponding canonical SMILES.

Input SOAP Message

List of SMILES strings
ISO option string
ERRORLEVEL

Output List:

List of objects with one object per input SMILES
[(SMILES string, error messages)]

Option:

ISO - Sets returned canonical SMILES to contain isomeric information
[TRUE|FALSE] default is FALSE

3.2 convertStructure

This Web Service interconverts data and structures between MDL chemical table-based file formats [molfile (MOL), SDfile (SDF), RGfile, RXNfile (RDF) and RDfile (RDF)] and Daylight SMILES-based formats [SMILES (SMI), isomeric SMILES (ISM), SMARTS (SMA), SMIRKS (SMRK), and Thor Data Tree (TDT)]. Detailed descriptions of conversion formats and options are available in the [Daylight Conversion Manual](#).

MDL format to SMILES conversions are based upon default ptable values unless specific ptable changes are provided.

Input SOAP Message:

List of input strings
 Input format string
 Output format string
 List of options as name-value pairs
 Optional ptable changes as a list of
 [atom number, atom symbol, atom mass, list of valence-charge pairs]
 ERRORLEVEL

Output SOAP Message:

List of objects with one object per input SMILES
 [(output string, error messages)]

Valid Input/Output Combinations:

SMI --> SDF or RDF
 TDT --> SDF or RDF
 SDF --> SMI, ISM, SMA, TDT or TDTSMA
 RDF --> SMI, ISM, SMA, SMRK, TDT, TDTSMA or TDTSMRK

Note: MOL is a valid input value that is interchangeable with SDF regardless of the actual input format. In addition, MOL is a valid output format value if the input value is SMI or TDT. However, the output will always be written in SDF format even if there is no associated data. In addition, either MOL or SDF can be used with RGfile input. Lastly, rxnfile format is not recognized as a separate format. RDF is used for both rxnfiles and RDfiles.

Valid Conversion/Option Combinations:

Conversion	Option
SMI --> SDF or RDF	SMI_COMMENT
TDT --> SDF or RDF	NAME_DATATAG SMI_COMMENT SPLIT_FIELDS SMI_WITH_TUPLES USE_3D
SDF or RDF --> SMI, ISM or TDT	ADD_2D (TDT output only) ADD_3D (TDT output only) CHI_EXPLICIT_H DB_EXPLICIT_H DB_RING_CISTRANS FIX_RADICAL_RINGS ID_FIELD IMPLICIT_CHIRALITY M_ISO_ARE_DEFECTS

Daylight Web Services Manual

	PREFIX (RDF input only) SMI_IS_ISM (TDT output only) SPLIT_FIELDS (SDF input only)
SDF or RDF --> SMA or TDTSMA	ADD_2D (TDTSMA output only) ADD_3D (TDTSMA output only) DAYLIGHT_LIKE DAYLIGHT_CHI_H DAYLIGHT_HCOUNT DAYLIGHT_STEREO DB_RING_CISTRANS FIX_RADICAL_RINGS ID_FIELD IMPLICIT_CHIRALITY M_ISO_ARE_DEFECTS PREFIX (RDF input only) SPLIT_FIELDS (SDF input only)
RDF --> SMRK or TDTSMRK	ADD_2D (TDTSMRK output only) ADD_3D (TDTSMRK output only) DAYLIGHT_CHI_H DB_RING_CISTRANS FIX_RADICAL_RINGS ID_FIELD IMPLICIT_CHIRALITY M_ISO_ARE_DEFECTS PREFIX

Options:

ADD_2D - Adds 2D coordinates to Daylight output
[TRUEIFALSE] default is true

ADD_3D - Adds 3D coordinates to Daylight output
[TRUEIFALSE] default is true

CHI_EXPLICIT_H - Determines whether chiral atoms must have explicit hydrogens
[TRUEIFALSE] default is false

DAYLIGHT_LIKE - Sets all three DAYLIGHT options
[TRUEIFALSE] default is true

DAYLIGHT_HCOUNT - Determines whether both the explicit H and H-count fields are used
[TRUEIFALSE] default is true

DAYLIGHT_STEREO - Determines whether only specified stereochemistry is used
[TRUEIFALSE] default is true

DAYLIGHT_CHI_H - Determines whether chiral atoms in the input file must have explicit hydrogens

- [TRUEIFALSE] default is true
- DB_EXPLICIT_H - Determines whether double bonds in the input file must have explicit hydrogens
[TRUEIFALSE] default is false
- DB_RINGS_CISTRANS - Determines whether stereochemistry for ring double bonds is indicated
[TRUEIFALSE] default is false
- FIX_RADICAL-RINGS - Determines if radical rings are converted to aromatic
[TRUEIFALSE] default is true
- ID_FIELD- Specifies the data field identifier to be used as the unique ID
[NAME] default is first line of header block for molecules and \$RIREG for reactions
- IMPLICIT_CHIRALITY - Specifies how chirality is determined in order to detect implicit chiral centers
[TRUEIFALSE] default is false
- M_ISO_ARE_DEFECTS - Indicates whether values in the M ISO line
are mass defects or actual masses
[TRUEIFALSE] default is false
- NAME_DATATAG - Designates the data tag to be used as the unique ID
[NAME] default is LINE1, if available otherwise \$NAM
- PREFIX - Parses the designated prefix from data field identifiers
[NAME] default is to use the full \$DTYPE name
- SMI_COMMENT - Determines whether the SMILES is placed in the comment line
of the connection table
[TRUEIFALSE] default is false
- SMI_IS_ISM - Replaces SMILES with isomeric SMILES in the output
[TRUEIFALSE] default is false
- SMI_WITH_TUPLES - Determines whether tuple information is associated with SMILES
or isomeric SMILES
[TRUEIFALSE] default is true
- SPLIT-FIELDS - Splits data that is spread across multiple lines in an input into separate entries
[TRUEIFALSE] default is false
- USE_3D - Designates whether 3D coordinates are included in the MDL output
[TRUEIFALSE] default is false

3.3 deriveScaffold

This Web Service generates a single scaffold that captures all common substructure elements including ring topology from a list of SMILES. Please note that this process can time intensive. Therefore, the server timeout may need to be adjusted to accommodate large processes.

Input SOAP Message:

List of SMILES strings
List of options as name-value pairs
ERRORLEVEL

Output SOAP Message:

SMARTS scaffold
Error messages

Option:

TOPO_OPTION - Sets topology to full (DX_CSS_DEFAULT, default), simple with no ring bond counts used (DX_CSS_SIMPLE_TOPOLOGY) and uses only atoms with TOORDER properties set, e.g. as a result of using transforms, (DX_CSS_USE_TORDER_ONLY)
If multiple TOPO_OPTION-value pairs are supplied, the individual values will be combined and all will be used.

MIN_FRAGMENT - Sets the minimum fingerprint path size
[INTEGER] default is 0, range is 0 to 19
Increasing the minimum fingerprint path size eliminates scaffolds that are smaller than the set size.

3.4 deriveSDClusters

This Web Service partitions small to moderate-sized sets of input SMILES into clusters with significant scaffolds. Like most clustering algorithms, scaffold-directed clustering limit on the web service permits clustering up to 10,000 structures which can take tens of minutes for drug-like molecules. The server timeout may need to be adjusted depending on the use. For more information see the [Clustering Manual](#).

Input SOAP Message:

List of SMILES strings
List of options as name-value pairs
ERRORLEVEL

Output SOAP Message:

List of objects with one object per output cluster
[(SMARTS scaffold, list of member SMILES, properties
(cluster id, minimum coverage, number of members), error messages)]

Options:

MIN_FP_PATH_SIZE - Sets the minimum fingerprint path size
[INTEGER] default is 0, range is 0 to 19
Increasing the minimum fingerprint path size eliminates scaffolds that are smaller than the set size.

MAX_FP_PATH_SIZE - Sets the maximum fingerprint path size
[INTEGER] default is 19, range is 0 to 19

MIN_COVERAGE - Sets the minimum scaffold coverage
[NUMBER] default is 0.3, range is 0.0 to 1.0

TOPO_OPTION - Sets topology to full or simple, i.e., no ring bond counts used
[DX_CSS_DEFAULT|DX_CSS_SIMPLE_TOPOLOGY] default is DX_CSS_DEFAULT

3.5 getProperties

This Web Service calculates values for a specified list of different physical properties for one or more input SMILES using Daylight algorithms. See the [Daylight Properties Manual](#) for additional information.

Input SOAP Message:

List of objects SMILES strings
List of properties
SINGLE_PART option string
RXNDIFF option string
Optional SMARTS string for MATCH_COUNT
ERRORLEVEL

Output SOAP Message:

List of objects with one object per input SMILES
[(list of computed property values, error messages)]

Options:

SINGLE_PART - Treats entire input SMILES as a single molecule
[TRUE|FALSE] default is FALSE
Computed property values can be a comma separated string if the input SMILES has multiple parts and flag is set to false

RXNDIFF - Returns the difference between the property values of the product and the reactant
[TRUE|FALSE] default is FALSE
If rxndiff is true, then single_part cannot be FALSE

Properties:

ACCURATE_MASS - Molecular weight in atomic mass units using the the most common isotope of each element

ATOM-COUNT - Count of heavy atoms in a molecule

AVERAGE_MOL-WEIGHT - Molecular weight based on average atomic weights for naturally occurring element

DEPICTION - Planar coordinates for explicit atoms

FINGERPRINT - Fingerprint using default parameters

FLEXIBILITY - Ratio of rotatable bonds to the total count of bonds

FRAGMENT_COUNT - Number of fragments formed by removal of the isolated carbons from the structure

Daylight Web Services Manual

HACCEPTOR_COUNT - Number of hydrogen-bond acceptor sites

HDONOR_COUNT - Number of hydrogen-bond donor sites

MATCH_COUNT - Number of unique matches using a user defined SMARTS

MOLAR_VOLUME - Average molar volume based on Schroedinger's method

MOL_FORM - Molecular formula in Hill order

PARACHOR - Molar surface tension in dynes per centimeter using McGowan's method

PART_COUNT - Number of components

POLAR_SURFACE_AREA - Topological polar surface area according to the method of Ertl, Rohde, and Selzer

RIGIDITY - Tanimoto similarity value between a molecule and version of itself with rotatable bonds removed

RING_COUNT - Number of smallest set of smallest rings

ROTBOND_COUNT - Number of rotatable bonds using defined SMARTS pattern

STEREOCENTER_COUNT - Number of stereocenters using a particular set of defined SMARTS patterns

3.6 getDepiction

This Web Service parses a list of name-value strings (alternate name and values), one of which pairs must be either "SMILES" and a valid SMILES string or "TDT" and a valid TDT string and returns a structural diagram in GIF format.

Input SOAP Message:

List of options as name-value pairs one of which must be SMILES or TDT

ERRORLEVEL

Output SOAP Message:

GIF (binary array)

Error messages

Options:

COLORMODE - Specifies output color scheme for foreground and background
[COB, COW, COP, BOW, BOP, WOB, or WOP] default is COB

FROMTO - Specifies output horizontal alignment by aligning depiction to -1 and -2 wildcard atoms
([*-1] and [*-2])
[TRUE|FALSE] default is false; overridden by the orient option)

HEIGHT- Specifies output height

- [PIXELS] default is 300
- HIDE_CHI_H - Specifies hide chiral hydrogens in output
[TRUEIFALSE] default is true
- HIGHLIGHT - Specifies a SMARTS query string to be used to highlight the matching portion of the input SMILES or TDT structure.
[SMARTS]
- HLEN_PCT - Specifies scaling length for bonds to hydrogen in output
[NUMBER] default is 1.00, range is 0.67 to 1.0
- HYDROGENS - Specifies that aliphatic hydrogen and carbons are to be shown in output
[TRUEIFALSE] default is false
- NONEXHAUSTIVE - Specifies whether exhaustive or nonexhaustive SMARTS matching is used for highlighting the depiction
[TRUEIFALSE] default is false
- NUMCOLORS - Specifies the number of output atom colors for input TDT with ALAB specified
[NUMBER] CPK color scheme is used by default
- OLD_STYLE - Specifies pre-v4.83 bond style rendition to be used in output
[TRUEIFALSE] default is false
- ORIENT - Specifies automatic orientation of 2D layout in output to the longest axis
[TRUEIFALSE] default is false; overrides the fromto option)
- OUTPUT - Specifies whether the out is in gif or png format
- REACTION - Specifies that the input SMILES is a reaction with atom-mapping
[TRUEIFALSE] default is false
- SCALE - Specifies the output number of pixels per angstrom
[NUMBER] default is 100; overrides width and height options
- SCHEMATIC - Specifies that the output be a skeleton frame with no hydrogen atoms or aromatic bonds
[TRUEIFALSE] default is false
- SMILES - Indicates that the input is a SMILES
[valid-SMILES-string]
- SMIRKS - Indicates that the input is a general reaction that may contain SMARTS expressions for atoms
[TRUEIFALSE] default is false
- TDT - Indicates that the input is a TDT
[valid-TDT-string]
- WIDTH - Specifies output width
[PIXELS] default is 400

XSMILES - Specifies that output be in XSMILES or Kekule format
[TRUE|FALSE] default is false

3.7 getTransform

This Web Service applies a specified transform to one or more input SMILES. See the [Daylight Theory Manual](#) for additional information on SMIRKS reaction transforms.

Input SOAP Message:

List of input SMILES
Single SMIRKS transform
ISO option string
List of options as name-value pairs
ERRORLEVEL

Output SOAP Message:

List of objects with one object per input SMILES [(transformed SMILES string, error messages)

Options:

ISO - Sets returned SMILES to contain isomeric information
[TRUE|FALSE] default is FALSE

EXHAUSTIVE_SEARCH - TRUE returns all molecules, FALSE returns a single molecule.
[TRUE|FALSE] default is FALSE.

FULL_RXN - Determines if a full reaction is returned or only the reaction product.
[TRUE|FALSE] default is TRUE.

DIRECTION - Determines the direction in which the transform is performed.
[DX_FORWARD|DX_REVERSE] default is DX_FORWARD.

3.8 getTautomers

The Web Service calculates tautomers for input SMILES. See the [Daylight Properties Manual](#) for additional information on tautomers.

Input SOAP Message:

List of input SMILES
ISO option string
List of options as name-value pairs
FIXED_SUBSTRUCTURE option list
ERRORLEVEL

Output SOAP Message:

List of objects with one object per input SMILES [(list of tautomeric SMILES string, error messages)

Options:

ISO - Sets returned SMILES to contain isomeric information
[TRUE|FALSE] default is FALSE

NO_ENOL - Restricts hydrogen donors and acceptors to only heteroatoms, i.e., suppresses keto-enol type tautomerism. [TRUE|FALSE] The default is FALSE.

KEKULE - Controls whether kekule structures are generated using dt_xsmiles() (TRUE) or canonical SMILES using dt_cansmiles() (FALSE).
[TRUE|FALSE] The default is FALSE.

UNIQUE - Determines if the output is generated by using the relative electronegativities of the atom types (O>S>Se>Te>N>C) as graph invariants to preferentially assign double bond and hydrogen positions in the tautomer. Although this canonical tautomer often corresponds to the lowest energy form, this is not guaranteed as extended electronic factors are not considered.
[TRUE|FALSE] The default is FALSE

ITERATION_LIMIT - Maximum number of donor or acceptor positions for iteration. If a structure has more donors or acceptors than the specified limit, then no tautomer enumeration is performed. The default limit allows the program to generate tautomers for every input structure until all possible tautomers have been generated. A reasonable value for limit to minimize long-running, pathological cases, is 10.
[INTEGER] The default is 0

FIXED_SUBSTRUCTURE - This option is useful for excluding specific functional groups from the calculation. If an input molecule matches one of the SMARTS in the supplied comma-separated list it is marked as non-tautomerizable.
[LIST OF SMARTS] The default is no matches.

3.9 desaltSmiles

This Web Service parses a list of SMILES and removes salts based upon a salt table or an actual list of salts that is provided as part of the input message. A copy of the default salt table (salts.dat) is located in \$DY_ROOT/data. The format is one salt with a class number per line, i.e., [Na+] 0.

In order to utilize a user-provided table instead of the default, the environment variable DY_SALT_DATA must be set to the location of the new table. If the user-provided table has more than one class listed, then the class number to be used can be specified in the input message. Lastly, if a list of salts is provided with the input message, then this list is used in place of either the default or user-provided table.

Input SOAP Message:

List of input SMILES
Comma-separated list of salts
ISO option string
Class number
ERRORLEVEL

Output SOAP Message:

List of objects with one object per input SMILES [(SMILES string, error messages)

Options:

ISO - Sets returned SMILES to contain isomeric information
[TRUE|FALSE] default is FALSE

3.10 normalizeSmiles

This Web Service parses a list of SMILES and normalizes the structure based upon a transform table or an actual list of SMIRKS transforms that is provided as part of the input message. A copy of the default transform table (transforms.dat) is located in \$DY_ROOT/data. The format is one SMIRKS, reaction direction, and class number per line.

In order to utilize a user-provided table instead of the default, the environment variable DY_TRANSFORM_DATA must be set to the location of the new table. If the user-provided table has more than one class listed, then the class number to be used can be specified in the input message. Lastly, if a list of transforms is provided with the input message, then this list is used in place of either the default or user-provided table. In this case, the direction (forward/reverse) to be used for the transformation is specified in input message.

Input SOAP Message:

List of input SMILES
Comma-separated list of SMIRKS transforms
ISO option string
Direction of SMIRKS reaction
Class number
ERRORLEVEL

Output SOAP Message:

List of objects with one object per input SMILES [(SMILES string, error messages)

Options:

ISO - Sets returned SMILES to contain isomeric information
[TRUE|FALSE] default is FALSE

3.11 getClogP

This Web Service parses a list of SMILES and calculates the both the logarithm of the computed octanol-water partition coefficient (clogp) and molar refractivity (cmr). See the [Daylight ClogP Manual](#) and the [Daylight CMR Manual](#) for more information.

Input SOAP Message:

List of input SMILES
ERRORLEVEL

Output SOAP Message:

List of objects with one object per input SMILES [(clogp result string, cmr result string, error messages)

3.12 generateRTable

This Web Service parses a single, scaffold-based cluster of molecules such as those generated by [deriveSDclusters](#) and determines the RGroups for that scaffold.

Please note that in order to generate an rtable, the input SMARTS scaffold cannot have more than four fragments. In fact, the most useful R-tables are those where the number of scaffold fragments is kept to one or two. Therefore if you are using deriveSDclusters/deriveScaffold to generate the input information, you may need to set the MIN_FP_PATH_SIZE/MIN_FRAGMENT option to a larger value in order to get a better scaffold. Also be aware that if an input scaffold is highly symmetric then the program will automatically switch to non-exhaustive matches.

Input SOAP Message:

Cluster object consisting of SMARTS scaffold, list of member SMILES, list of cluster properties as name-value pairs

ERRORLEVEL

The list of cluster properties is optional and may be the same as that generated by deriveSDCluster.

Output SOAP Message:

List of RTable row objects with one object per input SMILES

[RTable row (molecule ID and Rgroups consisting of an array of SMARTS strings)]
error messages

3.13 executeProgram

This Web Service enables the use of PipeTalk for two-way communication with an external process such as that described for ClogP. See the Program Object Toolkit section of the [Daylight Programmer's Guide](#) for additional information. Note: In order for executeProgram to function, the environment variable DY_WSPATH must be defined (see [Daylight Installation Manual](#)) and the program being called by must be installed in a path below DY_WSPATH.

Input SOAP Message:

List of lists of strings where each list is one input record

Path to program relative to DY_WSPATH (ascending is not permitted)

List of program arguments

Output SOAP Message:

List of lists where each list is one output record