

Structural Signature Variation

SSV is a method to calculate if a mutation could be beneficial or not for an enzyme.

Summary

Summary	2
SSV documentation	3
What is SSV?	
Case study: Second-Generation Biofuel problem	
How does SSV method work?	5
Inputs	5
SSV steps for the case study	6
Running online	7
Accessing the "Run online" panel	
Running an example (H228T)	8
Run by scripts	
Requirements	
Scripts	12

SSV documentation

What is SSV?

SSV (Structural Signature Variation) is a method to propose mutations for enzymes used in industrial applications. **SSV** uses structural signatures to detect patterns, which can improve the activity of enzymes. A real and important application for SSV is the second-generation biofuel production.



Case study: Second-Generation Biofuel problem

We used SSV as a method to propose mutations for β -glucosidase enzymes used in Second-Generation Biofuel production. SSV uses structural signatures to detect patterns, which can improve the activity of nontolerant β -glucosidases, based on a manually curated database of glucose-tolerant β -glucosidases.

Second-Generation Biofuel is produced by residues obtained from the first generation biofuel production. They can be obtained from biomass, such as corn, algae, sugarcane, and so on. The production process is based on the extraction of fermentable sugars from cellulose.

Cellulose is decomposed by the action of three enzymes: endoglucanases, exoglucanases, and β glucosidases. Endoglucanases act first, cleaving cellulose in oligosaccharides of several lengths. Then, exoglucanases cleave the oligosaccharides in disaccharides, such as cellobiose. In the end, betaglucosidases cleaves the cellobiose in two molecules of glucose, that will be used in the fermentation process to obtaining biofuel.



 β -glucosidases are strongly inhibited by high glucose concentrations (same used for biofuel production).



In addition, it increases cellobiose concentration that inhibits endoglucanases and exoglucanases.



Glucose-tolerant β -glucosidases: β -glucosidases of high resistance to glucose inhibition, also called glucose-tolerants, can help to improve the biofuel production. Also, some mutations can turn non-tolerant in glucose-tolerant β -glucosidases.



How does SSV method work?

SSV method constructs structural signatures for wild and mutant proteins and compares the signature's variation with a manually curated database. In the case study, we used a database of glucose-tolerant β -glucosidases, called **Betagdb** (available at <u>http://bioinfo.dcc.ufmg.br/betagdb</u>).



Inputs

SSV requires three proteins:

1. Wild protein

- 2. Mutant protein
- 3. Template (reference)

SSV steps for the case study



- 1. We constructed structural signatures using aCSM-ALL. We used cutoff ranging from oÅ until 10Å, and step distance of o.1Å. However, we did not use the whole protein, only the region near to the active site. First, we performed the extraction of the catalytic pocket residues. This is not necessary for all projects run in SSV, however, if you use the whole protein, this increase the computational costs. The mutant and wild protein are aligned with 3VIK, the only beta-glucosidase in complex with cellobiose detected in the literature. The residues of the catalytic pocket of 3VIK are known: Q45, H148, W149, N192, S193, L195, T196, D199, M207, N253, I254, N255, Y273, N335, F336, Y337, T338, L340, W374, E402, W444, E451, W452, and F460. We performed alignments using MultiProt (http://bioinfo3d.cs.tau.ac.il/MultiProt);
- We previously calculated the signature of 23 glucose-tolerant beta-glucosidases of betagdb. We calculated the distance among all glucose-tolerant beta-glucosidase and wild and mutant. The lowest values correspond to the templates;
- 3. We calculated the distance between wild and its template (Δ SSV_{Wt}), and mutant and template (Δ SSV_{Mt}). The difference between both distances is the variation of the signature variations ($\Delta\Delta$ SSV).

Running online

Running online is the best way to execute GTS. It is available online at: <<u>http://bioinfo.dcc.ufmq.br/qts</u>>.

Accessing the "Run online" panel

Click on "Run online" or in the "Run now!" Button.



The "Run online" panel requires some information to execute:

- **Project name**: define a name for your project. You can use any name (the system will create a unique ID for each project);
- E-mail: declare your e-mail (optional);
- Mutation evaluated: insert the point mutation or multiple mutations evaluated (optional);
- Wild PDB: input the wild PDB file;
- Mutant PDB: input the mutant PDB file;
- **Templates database:** input a set of proteins that it will be used as templates (send a zip file).

Run online

Project name: Define a name for your project (required)	Wild PDB (required): Escolher arquivo Nenhum arquivo selecionado
E-mail: Receive the results in your e-mail (optional)	Mutant PDB (required): Escolher arquivo Nenhum arquivo selecionado Limit: 2MB. Please, send only one chain.
Mutations evaluated: E.g. E96K (optional)	Templatea database (required; ZIP format): Escolher arquivo Nenhum arquivo selecionado
	Download sample dataset
	Submit

We made available a database with 27 mutations evaluated in the paper and the 23 proteins of betagdb (the templates for the case study). We can download the database and perform analysis using SSV online. We also made available a table with expected values and the values found.

SEV	Dov	wnload samp	ble dataset Down	nload			bad He X
	id	File (Wild)	File (Mutant)	Mutation	∆∆GTS expected	ΔΔGTS score	
Run on	1	w1.pdb	m1.pdb	H228T	ΔΔGTS < 0	-186.18	
	2	w2.pdb	m2.pdb	V174C/A404V/L441F	$\Delta\Delta GTS < 0$	-246.22	
Project name:	3	w3.pdb	m3.pdb	H184F	$\Delta\Delta GTS < 0$	100.37	
Define a name for	4	w4.pdb	m4.pdb	P172L	$\Delta\Delta GTS < 0$	-6.29	
	5	w5.pdb	m5.pdb	P172L/F250A	$\Delta\Delta GTS < 0$	-6.29	
E-mail:	6	w6.pdb	m6.pdb	L167W	$\Delta\Delta GTS < 0$	-602.80	
Receive the result	7	w7.pdb	m7.pdb	L167W/P172L	$\Delta\Delta GTS < 0$	-615.46	
Mutations evaluated:	8	w8.pdb	m8.pdb	L167W/P172L/P338F	$\Delta\Delta GTS < 0$	-615.46	
E.g. E96K (option	9	w9.pdb	m9.pdb	V168Y	$\Delta\Delta GTS > 0$	330.56	
	10	w10.pdb	m10.pdb	F225S	$\Delta\Delta GTS > 0$	-365.07	
	11	w11.pdb	m11.pdb	Y308F	$\Delta\Delta GTS > 0$	34.19	
	12	w12.pdb	m12.pdb	Y308A	$\Delta\Delta GTS > 0$	-108.62	
	13	w13.pdb	m13.pdb	1207V	$\Delta\Delta GTS < 0$	-71.56	
	14	w14.pdb	m14.pdb	N218H	ΔΔGTS < 0	-230.61	
	15	w15.pdb	m15.pdb	N273V	$\Delta\Delta GTS > 0$	-55.26	
	16	w16.pdb	m16.pdb	F252I	$\Delta\Delta GTS > 0$	86.70	

Running an example (H228T)

Now, we will run the first example of that database. The mutation H228T, where detected for a nontolerant beta-glucosidase and improved its glucose tolerance. For this reason, we expected a $\Delta\Delta$ SSV negative.

Download the dataset:

SSV	Dov	wnload samp	le dataset Dowr	nload		
	id	File (Wild)	File (Mutant)	Mutation	∆∆GTS expected	ΔΔGTS score
Run on	1	w1.pdb	m1.pdb	H228T	ΔΔGTS < 0	-186.18
	2	w2.pdb	m2.pdb	V174C/A404V/L441F	ΔΔGTS < 0	-246.22
Project name:	3	w3.pdb	m3.pdb	H184F	ΔΔGTS < 0	100.37
Bgl1B_H228T	4	w4.pdb	m4.pdb	P172L	ΔΔGTS < 0	-6.29
	5	w5.pdb	m5.pdb	P172L/F250A	ΔΔGTS < 0	-6.29
E-mail:	6	w6.pdb	m6.pdb	L167W	ΔΔGTS < 0	-602.80
diego@dcc.ufmg.	7	w7.pdb	m7.pdb	L167W/P172L	ΔΔGTS < 0	-615.46
Mutations evaluated:	8	w8.pdb	m8.pdb	L167W/P172L/P338F	ΔΔGTS < 0	-615.46
H228T	9	w9.pdb	m9.pdb	V168Y	$\Delta\Delta GTS > 0$	330.56
	10	w10.pdb	m10.pdb	F225S	$\Delta\Delta GTS > 0$	-365.07
	11	w11.pdb	m11.pdb	Y308F	$\Delta\Delta GTS > 0$	34.19
	12	w12.pdb	m12.pdb	Y308A	$\Delta\Delta GTS > 0$	-108.62
	13	w13.pdb	m13.pdb	1207V	ΔΔGTS < 0	-71.56
	14	w14.pdb	m14.pdb	N218H	ΔΔGTS < 0	-230.61
	15	w15.pdb	m15.pdb	N273V	$\Delta\Delta GTS > 0$	-55.26

Extract the files:

				📄 dataset				
Et	= • =		*~			0	Q Buscar	
Nova Pasta	Caminho	Visualizar	Ação	Organizar		Obter Informações	Buscar	
Nome					^	Data de Modificação	Tamanho	Tipo
🗋 m1.	pdb					8 de ago de 2016 16:01	16 KB	Documt 2.app
🗋 m2	.pdb					8 de ago de 2016 16:01	17 KB	Documt 2.app
📄 m3	.pdb					8 de ago de 2016 16:02	17 KB	Documt 2.app
🗋 m4	.pdb					8 de ago de 2016 16:01	17 KB	Documt 2.app
🗋 m5	.pdb					8 de ago de 2016 16:01	17 KB	Documt 2.app
m6	.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
🗋 m7.	pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
- 🗋 m8	.pdb					8 de ago de 2016 16:02	18 KB	Documt 2.app
′ 🗋 m9	.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
🗋 m1	0.pdb					8 de ago de 2016 16:01	17 KB	Documt 2.app
📄 m1:	1.pdb					8 de ago de 2016 16:01	17 KB	Documt 2.app
🗋 m1	2.pdb					8 de ago de 2016 16:01	17 KB	Documt 2.app
📄 m1	3.pdb					8 de ago de 2016 16:01	17 KB	Documt 2.app
📄 m1-	4.pdb					8 de ago de 2016 16:02	17 KB	Documt 2.app
📄 m1	5.pdb					8 de ago de 2016 16:01	17 KB	Documt 2.app
🗋 m1	6.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
📄 m1	7.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
🗋 m1	B.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
📄 m1	9.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
🗋 m2	0.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
📄 m2	1.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
🗋 m2	2.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
📄 m2	3.pdb					8 de ago de 2016 16:02	18 KB	Documt 2.app
🗋 m2	4.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
📄 m2	5.pdb					8 de ago de 2016 16:01	17 KB	Documt 2.app
🗋 m2	6.pdb					8 de ago de 2016 16:02	17 KB	Documt 2.app
📄 m2	7.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app
Then ten	nplates.zip					Ontem 16:00	93 KB	ZIP
📄 w1.	pdb					8 de ago de 2016 16:02	17 KB	Documt 2.app
) 🗋 w2	.pdb					8 de ago de 2016 16:02	17 KB	Documt 2.app
📄 w3	.pdb					8 de ago de 2016 16:02	17 KB	Documt 2.app
) 🗋 w4	.pdb					8 de ago de 2016 16:01	18 KB	Documt 2.app

Run SSV using w1.pdb (wild), m1.pdb (mutant), and templates.zip (23 templates):

Sev	About	Documentation	Run online	Download	Help
Run online	Nenhum	arquivo selecionado			
Project name: BgI1B_H228T	Wild PDB (required) Escolher arquivo	w1.pdb			
E-mail: diego@dcc.ufmg.br	Mutant PDB (require Escolher arquivo Limit: 2MB. Please,	d): m1.pdb send only one chain.			
Mutations evaluated: H228T	Templates database Escolher arquivo	(required; ZIP format): templates.zip			
	Download san	nple dataset			
Su	ıbmit				
©2018 SSV by LBS 0	created by Diego Marian	D.			

10

After submitting the data, SSV online will process your requisition. A unique ID will be created for your project. When the process to finish, you can click at the link.

Sev	About	Documentation	Run online	Download	Help
Your project was created. You can access it at the link: SSV48B6462					

While SSV runs your project, you will receive this message:



In the end, you will be redirected to the individual page of the project. This page is identified by the unique ID, described under the project name (in the green sector).

Below the green sector, there are three important sectors:

- 1. The results of the $\Delta\Delta$ SSV calculation;
- 2. Wild visualization;
- 3. Mutant visualization.



In the main panel, it is shown:

- ΔΔSSV score;
- Classification;
- ΔSSV_{Wt};
- ΔSSV_{Mt} .

Run by scripts

We made available the scripts to run SSV. However, we cannot have sure that you will be able to reproduce the software environment. If you have problems, run the online version.

We strongly recommend running SSV online.

Requirements

- O.S. Linux 64bit (recommended Ubuntu 16.04)
- Python (version 2)
- Library Numpy
- Perl
- aCSM
- MultiProt

Scripts

Script	Input	Output
extractSite.py	PDBs files inserted into the folder "pdb".	PDBs files with the catalytic pocket extracted.
aCSM.pl	List with the addresses of PDBs files with the catalytic pocket extracted.	CSV file with the total of atoms in the cutoff distances.
dgts3.py	CSV file obtained from aCSM. The first 23 lines must represent the beta-glucosidases from betagdb in alphabetical order.	ΔΔSSV and templates for wild and mutant save in the file "result.txt".