3DM'S PATENT ANALYSIS SUITE BY BIO-PRODICT



1.3DM introduction

2. 3DM's patent analysis suite

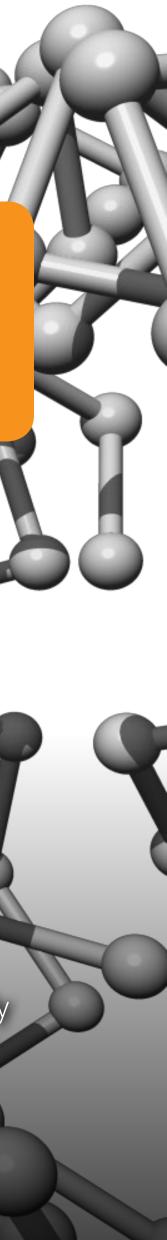
1.3DM INTRODUCTION





In the first step of the generation of a 3DM superfamily platform 3DM superimposes all available structures for the protein family and groups the structures into subfamilies. From these groups one template structure is selected that is used to align sequences for which no structures are available thereby forming subfamily alignments. Here 140 of such template structures are shown for the kinase protein family. The resulting 140 subfamilies are combined into one large superfamily alignment. In total the alignment for the kinases contains over 114.000 WT sequences. A 3D-numbering scheme, in which all structural equivalent residues get the same number (e.g. are in the same column in the alignment), is applied to all sequences, all structures, and to the alignments. Here, the labels of the purple residues located at 3D-position 27 are shown. The 3D-numbers synchronize all sequences and structures.

The 3D-numbers synchronize all sequences and allows for the transfer of mutation data from patents to any target protein.





Kinases (2012) ♡

Alignment Superfamily 3DM

Subset Default

Numbering scheme 3DM

Dashboard

Alignment (structures excluded)

3DM System

🔒 Start

			0			2		3			
	Alignment		Overall consensus 1BYGA	O Subfamily			K GSFO	BEVYKARDK DVMLGEYR	GEEVAV		REE
			3KFAA	Subfamily		кнк		EVYER VWK	SITVAV		
- du	Alignment statistics		2X2LA	Subfamily	EFPRKNLVL	GKT					
	5		3CLYA 1LUFA	O Subfamily O Subfamily	EYPRNNIEY	GKF					
			1RDOE	O Subfamily	K Y S L D D F E R	K	$ \rightarrow $	140 su	ihfam	ilida	
<,	Correlated mutations		3CBLA	O Subfamily	VLNHEDVVL	GEL		140 30		IIICS	
			106LA	 Subfamily Subfamily 	KVTPNDFDF			11/ C		iaual	Λ/T
5 y	Visualize		1P40A 2EVAA	O Subfamily		EEN	\rightarrow	>114.0		ique	
			2XIRA	O Subfamily	E F P R <mark>D</mark> R L K L	GKF					
.	Phylogeny	~	3F66B	 Subfamily Subfamily 	VIPPESLIV	GEN					
	, nyiogony	·	3GENA 1MP8A	Subfamily		GRC					
	D 11 ·		200LA	O Subfamily	E I Q R E R I E L E I D A S C I K I	EK	(31)-r	number	s are	in tac	nt a
	Panel design		2VGPB	Subfamily	QWTLDDFEI			IGIIIDCI	o arc,		
			2YFXA 3BRBA	 Subfamily Subfamily 		GKA	ccha	eme for	nnciti	nnc ir	h th
f	Patents	\sim	3C4ZA	O Subfamily	PVTKNTFRQ		SCIIC		positi	0115 11	I UI
			3EORA	O Subfamily	LIGEKDLTL			0011000		datru	ct
1	Proteins	~	3LXPA 3POZA	O Subfamily O Subfamily	VFEKRELKK	GK	All S	equenc	es and	i Stru	CLU
•**	1 locomo	·	3RCJA	O Subfamily	KKGVEDFKF	GK K EK					Ι.
~	o		2SRCA	O Subfamily	EIPRESLKL	EKK	renu	Imbered) acco	raina	ΟJ
ά	Search	~	3A8XB	 Subfamily Subfamily 	S L G L <mark>Q D F D</mark> L Q F E E R H L K F					J	
			3LXLA 3PVUA	Subfamily		HR					
3	System	\sim	3V50B	Subfamily	HIKRRDIVL	KRE					
-			2RKUA	O Subfamily O Subfamily	E T T I R R Y C R	GRF	Thic	connec	te tha	alianr	ma
			1J1BB 1XBBA	O Subfamily	FLDRSLLTL	GR EK QGE CE	IIIIJ	CONNEC		augin	IIC
			3A7IA	O Subfamily	EDPEELFTK	LEP	tha (3DM too			- + -
			3C0IA 3KMUA	O Subfamily O Subfamily			UIE .			ן מנו ט	larc
			3MIYA	O Subfamily	VIDPSELTF		inalı	Idina th			
			3KK8A	O Subfamily	T R F T <mark>D E</mark> Y Q L	F E E K E	INCIL	uding th	e udla		
			1PHKA 1XJDA	O Subfamily O Subfamily	D S F Y Q K Y E P R L G L D D F N F			0			
			2WEIA	O Subfamily	ESISDDYKL QNVEDFYEI - LDLDNLKL	GKK	pate	nts.			
			3BHYA	Subfamily		GEE					
			3G2FA	 Subfamily Subfamily 							
			3IECA 3KVWA	O Subfamily	DHLAYRYEV	LKVIG	KGSFO	QVVKAYDH	GQHVAL	KIIRNK	HRC
			2BUJA	O Subfamily	- L G G A R Y L F	IQKLG		S Y V D L V E G L		KRILCH	REE
			2J0IA	O Subfamily O Subfamily	E D P R K K Y T N E <mark>L K D D D L E K</mark>	FEKIG		STVYTAT <mark>D</mark> K SVVSKVRHK		K Q M N L R K L I H L E	K E I K K C
			3EOCA 3G51A	O Subfamily	KADESDEEL	IKVIG	GSEC	KVFLVRKV	GQLYAM	KVLKKD	QDC
			3MI9A	O Subfamily	CRCVDKFEK	LGKIG	EGTYO	Q V Y K A R D K		KKVRMD	
			1TKIA	O Subfamily O Subfamily	- ELSSKYAI DVLTERYEI	A E E L G V S T L G	RGQFG	S I V H R C V E I S K V V E C I D H	K K T F M A G E R V A L	K F V K V K K I I K N V	QAL REA
			2EU9A 2H6DA	O Subfamily	EIKIGHYIL	GKTLG	EGTFO	KVKLAEHQ		KILNRQ	VGK
			3HMMA	O Subfamily	RTIARQIQL	VECIG	KGRYO	BEVWRGKWR	GEKVAV	KIFSSR	ER
			3LM5A	O Subfamily O Subfamily	E P F N N F Y T L R L K R E D F E I			V V R K C I E K E V A V V R H K		K F L R K R K I L N K W	R A C
			3TKUA 4AAAA	O Subfamily	SNVMEKYEK	L K V I G L G K V G	R GAFO	V V M K C R N K	G K V Y A M G Q I V A I	KKFKES	KK
			1U5RB	O Subfamily	ENPTKLFSD	LREIG		BAVYFA RD V	NEVVAI	ккмзүз	W Q D
			2W5AA	O Subfamily O Subfamily	Q S R V E D Y E V Q D P L G I F E L			R C Q K I R R K Q V Y K G R H V	GKILVW	K E L D Y G K V M D V T	EEE
			2X7FA 3DTCA	O Subfamily		EEIIG		SKVYRAFWR	GEEVAV	KAARQD	
			3HMPA	O Subfamily	•••• <mark>K</mark> Y S	LKQIG	SGGSS	KVFQVLNE	KQIYAI	KYVNLE	I E S
	This is		3185E	O Subfamily	SSCINNYLSE	KKILG	C GAFG	EVHLVEDN	KLYKIV	KILKKS	

This is a zoomed out picture of the sequence alignment that results from the superposition of the 140 kinase template structures. Residues that are aligned are p structural equivalent and have the same 3D-number in the 3DM system. Because they are structural equivalent they likely have the same role in the different proteins. This means that data, such as mutation data, can be transferred between residues that have the same 3D-humber?

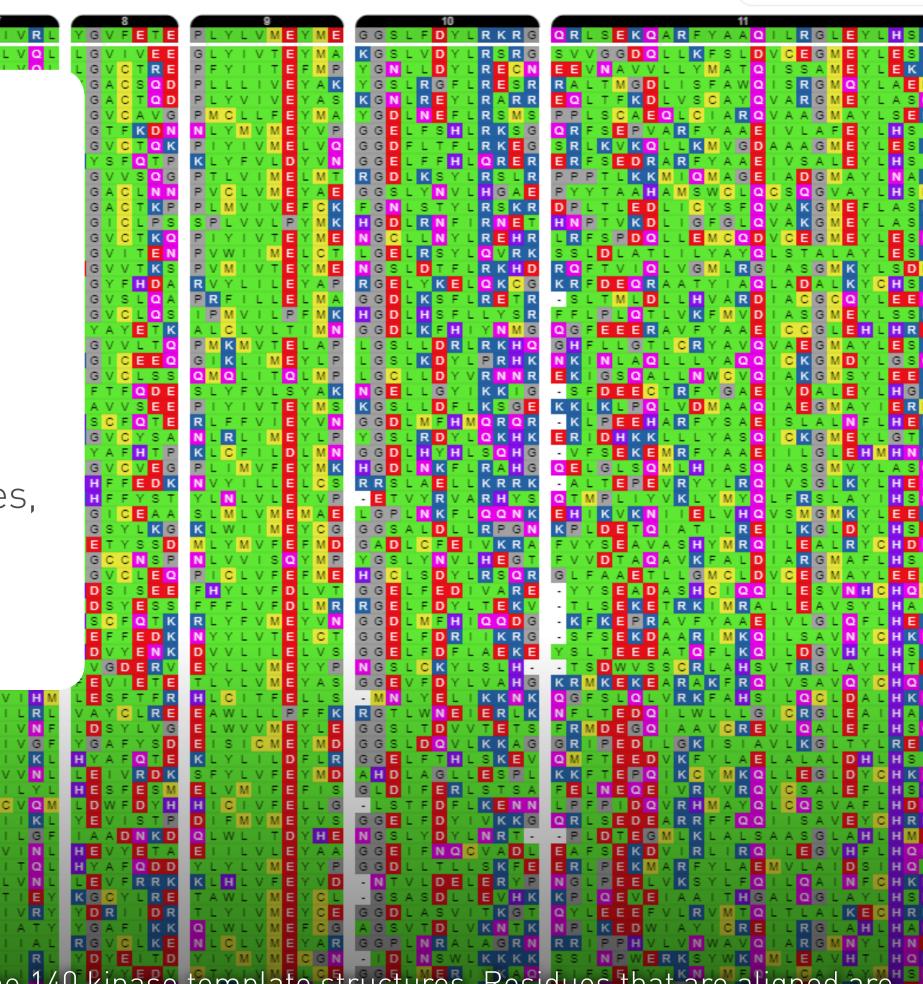
3ZYAA 3ZZWB

A F L A E A S V M T Q L RHKNL

sequences

a numbering he alignment. ures are the alignment.

ent, the structures, ta to each other, ollects for the



Structures exclude

 \sim

HOTS

G	T	1	12 H	R	D	L	ę	1
ĸ	F		н	R	D	L	K A	
N	F	V	н	R	D	Ľ	A A A A A K A K A K A A A A A K A A K A A K R K A K K A K K A K K N A K	1
κ	L	V	н	R	D	L	А	ł
ĸ	С	1	H	R	D	L.	A	1
K.	-	÷	H	R	D D	÷	ĸ	k
N	c	i.	H	R	D	Ľ	A	7
Ν	1	V	Υ	R	D	L	κ	F
к	F	v	H	R	D	Ļ.	A	/
ĸ	C	÷	н	R	Ы	÷	K A	1
ĸ	F	V I I V I I V I I V I I V I	$\mathbf{H} \mathbf{H} \mathbf{H} \mathbf{H} \mathbf{H} \mathbf{H} \mathbf{H} \mathbf{H} $	R	D		Â	1
Q	F		н	R	D	L	A	ł
R	F	V	H	R	D	÷	A	1
н	v	Ť	н	R	D	÷	ĸ	Í
н	F	Ť.	н	R	D	Ť.	A	1
Ν	F	Т	н	R	D	L	А	ł
R	-	4	Ŷ	R	D	L	ĸ	E
Q	Ŷ	÷	н	R	D	t	Â	Ź
R	L	 	н	R	D	L L L L	A	1
G	I		н	R	D	L	к	F
N	Ŷ	1	Ĥ	R	D	L	R	1
R	F C F C F F F Y V F F F Y I Y Y V	i	H	R	D	È	Â	
F	V	V	Υ	R	D	L	κ	F
H	F	V	H	R	D	L	A	1
R	F I V	-	н	R	D	÷	ĸ	l
N	F	v	Н	R	D	t	Ā	1
R	к	I V I P	н	RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR		T	к	1
Ν	1	1	H	R	D	Y	к	F
0 0 1 Z Z Z 0 Z H A 0 Z 0 Z 0 Z 0 Z H A 0 Z 0 Z X X X X Z Z 0 Z 0 Z 0 Z 0 Z X X X X	F I V	I	H Y H H H H H H H H H H H	R	P			F L F A F S A F
G	v	v	н	R	D	È	ĸ	Ē
Ν	T	V	н	R	D	L	κ	F
G	1	1	Y	R	D	÷	ĸ	l
R	1	V		R F R R	D D	L	к к	F
А	Ť.	A S V I A I	н	R	D D D D	L	N	\$
Ν	T	V	H	R	D	L	κ	1
K	T Y		Н	C	D	÷	ĸ	F
G	V	î	н	R	D D D	ĩ	ĸ	5
к	T	М	н	R	D	V	κ	F
G	1	M - L G T V A V	Y	R	D		ĸ	F
N	1	G	н	R	D	÷	R	í
к	Ĺ.	т	H H H H H H H	т	D	É.	ĸ	E
М	L	V	н	R	D		κ	F
A	1	A	Н	R	D	L	ĸ	0.0
G	Ŷ	v	н	R	D	L	ĸ	E E
N	I V Y I M	V I	H H	R R R F T R R L R R		1	к	F
Ν	М	L	Н	R	D D D		κ	1
ĸ	T V	÷	H H	R	D	L	ĸ	1
P	V I	j	н	R R R S	D	Ĺ	K	-
G	I	۷	н	S	D	L	к	I
2 R A Z K G G K G Z G K S A Z G Z Z F K P G Z G C Z P	1	L I V A V I I V	н	K R	D	Ļ	X X X X X X X X X X X X X X X X X X X	
0		V	H H	R R	D D		ĸ	
N			н	R	D	L	K	
Ρ		۷	н	R R R	D	L		1
N A S G	v	L A I V	Н	R	D D D D D	L		F
A		A	H H	R	D D	FL		-
G			н	R	D	L		I
Н					D	L	A	1
					- D			

MANY DIFFERENT DATA TYPES ARE COLLECTED FOR ALL SEQUENCES IN THESE LARGE ALIGNMENTS

ightarrow Mutation data

From literature: 264.605... **From patents:** 1.683 mutations have been extrated from claimes.

ightarrow Structure data

PPI, ligand contacts, bridges, solvent accessibility, flexibility/RMSD, ...

 \rightarrow Alignment data

Conservation, correlated mutations, family specific residues, ...

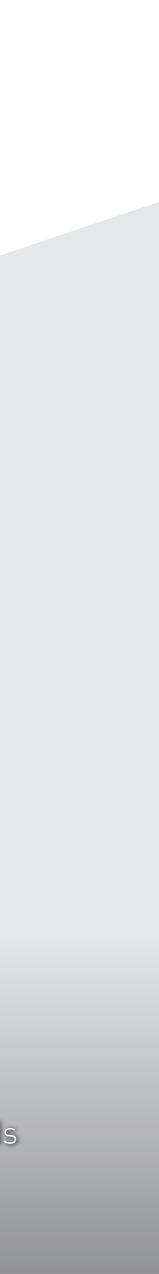
ightarrow SNP data

SNP databases, genome sequencing projects

ightarrow Data- and tool integration

All data are stored connected to the 3D numbering scheme and thus to the alignment, to the structures, all 3DM tools, and each other.

3DM collects many different data types from all kinds of sources for all the 114.000 proteins that are available in this kinase protein family. These are all stored connected to the 3D-numbers. This connects all data to each other making it possible to make very complex queries. For example, it now takes a couple of seconds to find out if mutations reported in the literature that cause effects on specificity are more often found at positions that show a correlated mutation behavior in the alignment. Manually such comparisons can easily be a year work.



2. 3DM'S PATENT ANALYSIS SUITE



PATENT DATA EXTRACTION

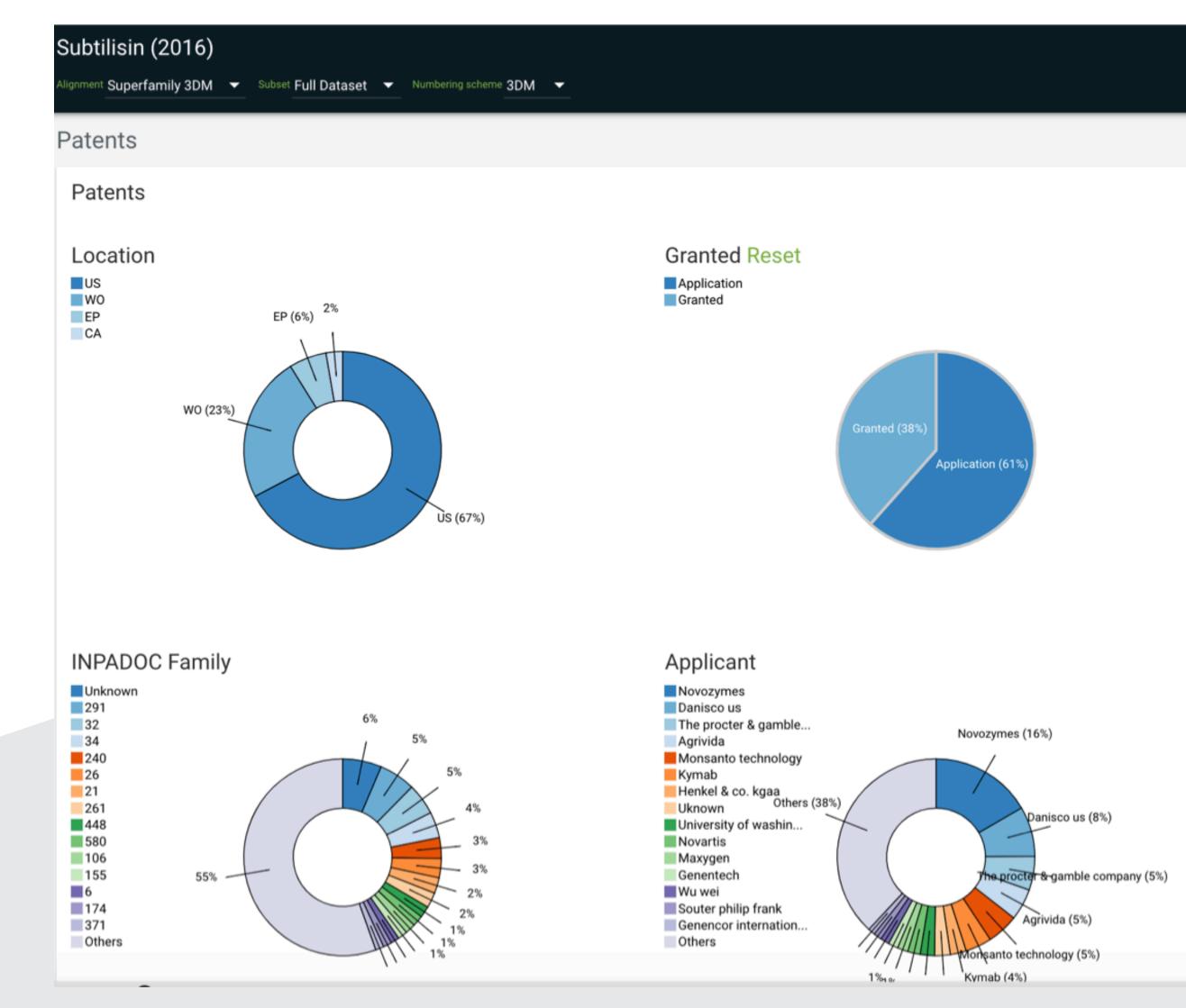
 \rightarrow

Collection of patents for complete protein families

- > Claimed and mentioned sequences
- > Claimed sequence spaces
- > Claimed mutations > Patent family information
- > Bibliographic data:
 - I Priority, filing and publication dates
 - II Applicants
 - III Titles & abstracts
 - IV Many others...
- > Inventors
- > Many other data types

3DM collects the patents for complete protein superfamilies by combining complex patent sequence searches with a smart patent family search. For important heavily patented protein families, such as the kinase protein family or the antibody protein family, 3DM collects many (7.944 and 44.859 respectively) patents. From these patents many different data types are extracted, such as claimed sequences and claimed mutations. For the kinase and antibody protein families 3DM extracted 131.578 and 703.356 sequences, and 1.683 and 5.180 mutations, respectively.

DIFFERENT ONLINE SELECTION TOOLS FOR ANALYSIS OF THE PATENTS

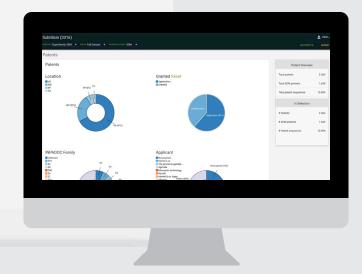




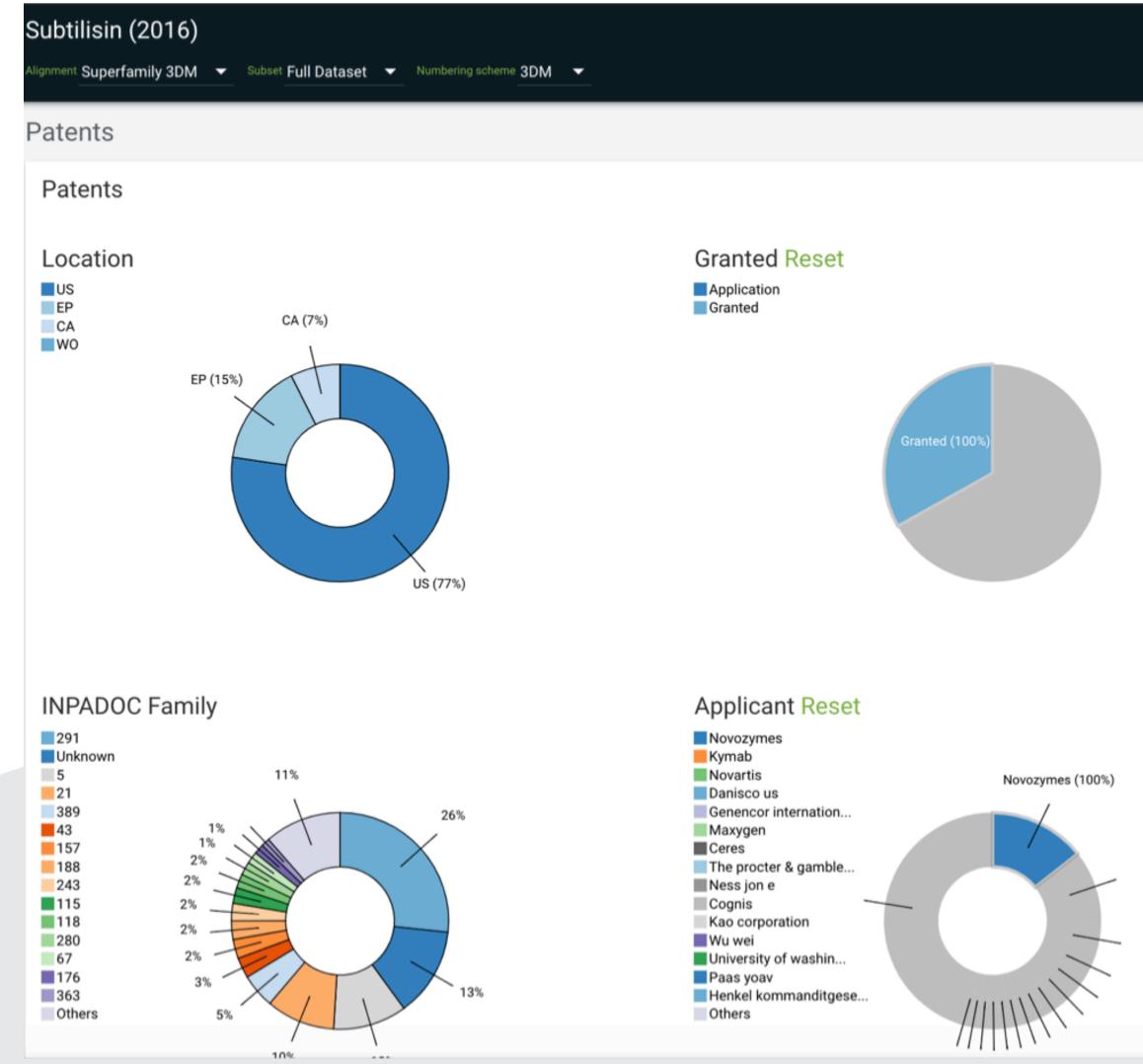
👱 Henk-

HOTSPOTS SUBSE

Patent Overview	
Total patents	2.464
Total 3DM proteins	1.635
Total patent sequences	13.894
In Selection	
# Patents	2.464
# 3DM proteins	1.635
# Patent sequences	13.894



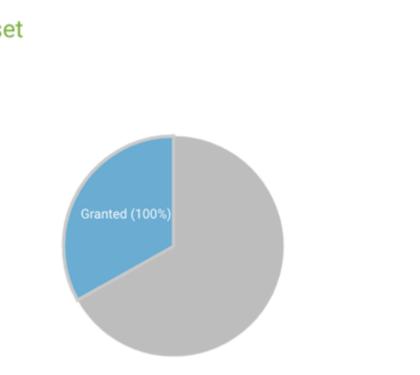
DIFFERENT ONLINE SELECTION TOOLS FOR ANALYSIS OF THE PATENTS

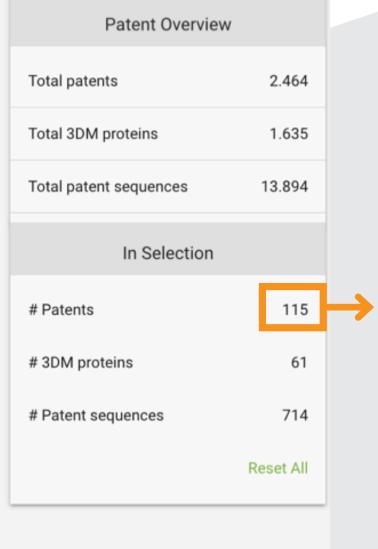




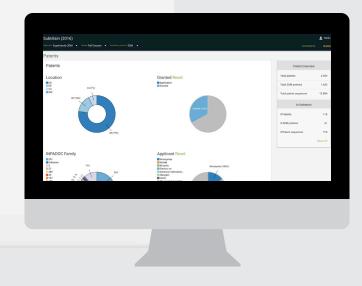
👱 Henł

HOTSPOTS SUBS

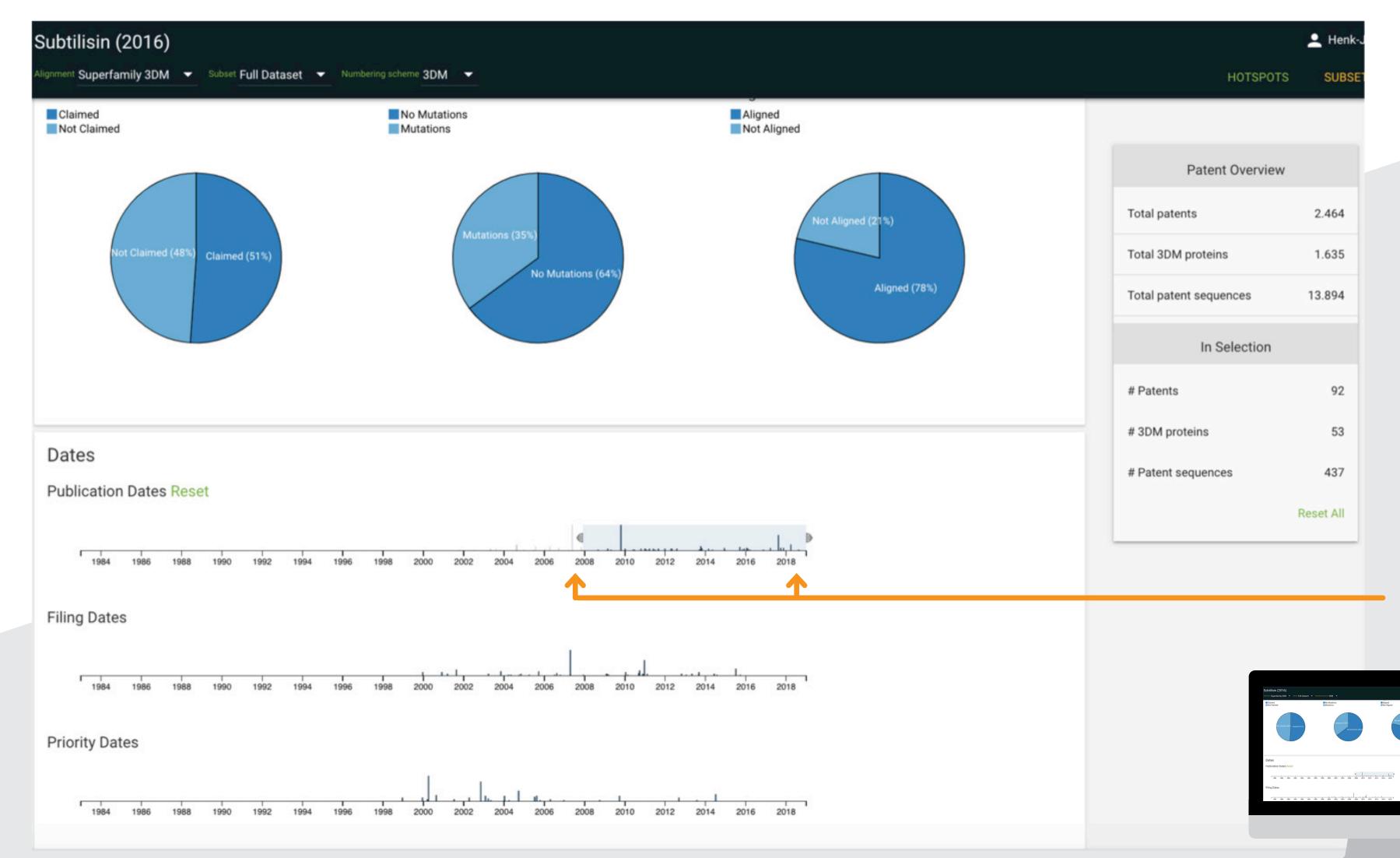




Here all granted patents from Novozymes are selected. Novozymes has 115 granted patents in this protein family.



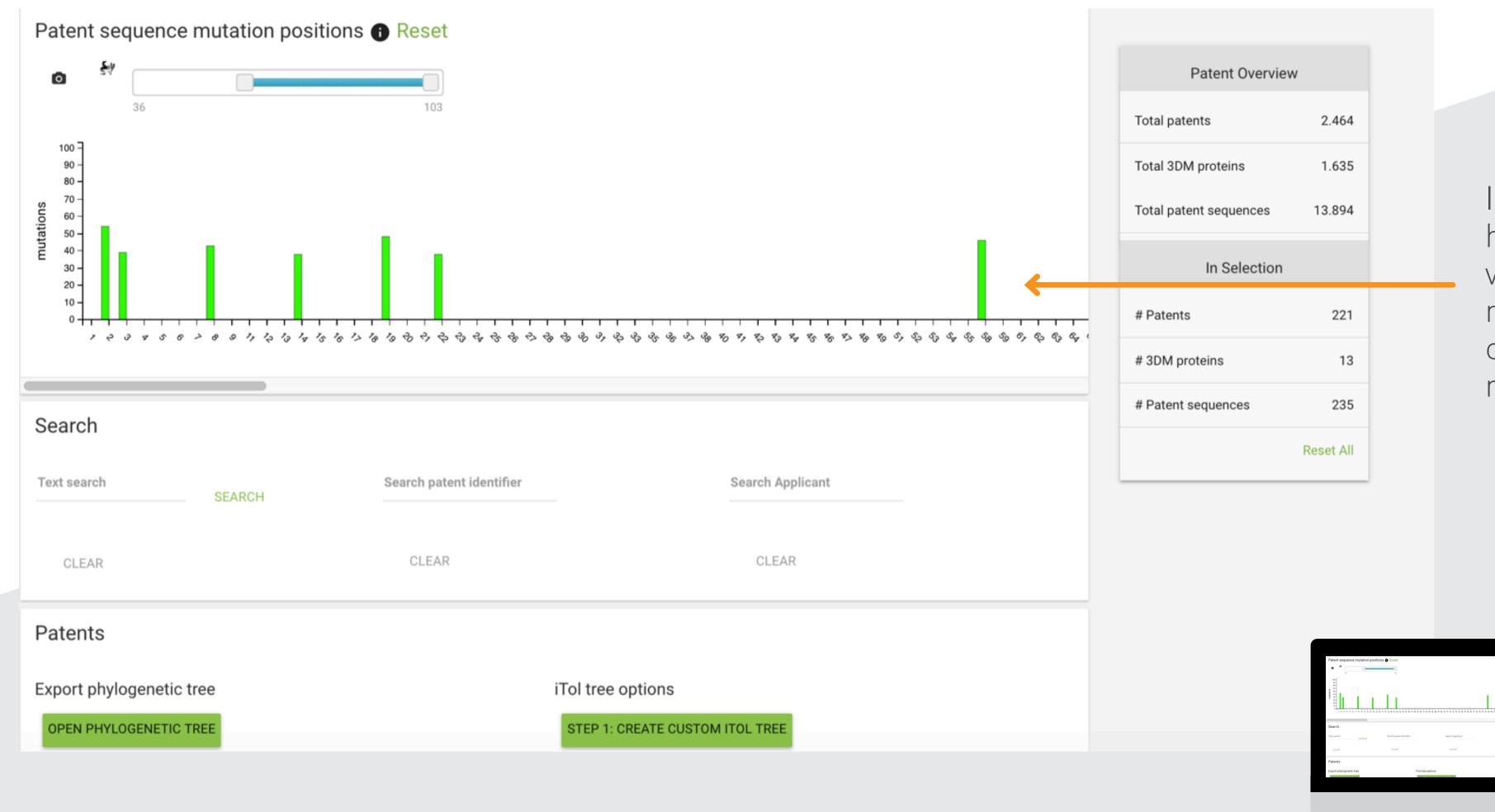
DIFFERENT ONLINE SELECTION TOOLS FOR ANALYSIS OF THE PATENTS





Of which 92 have been published since 2008

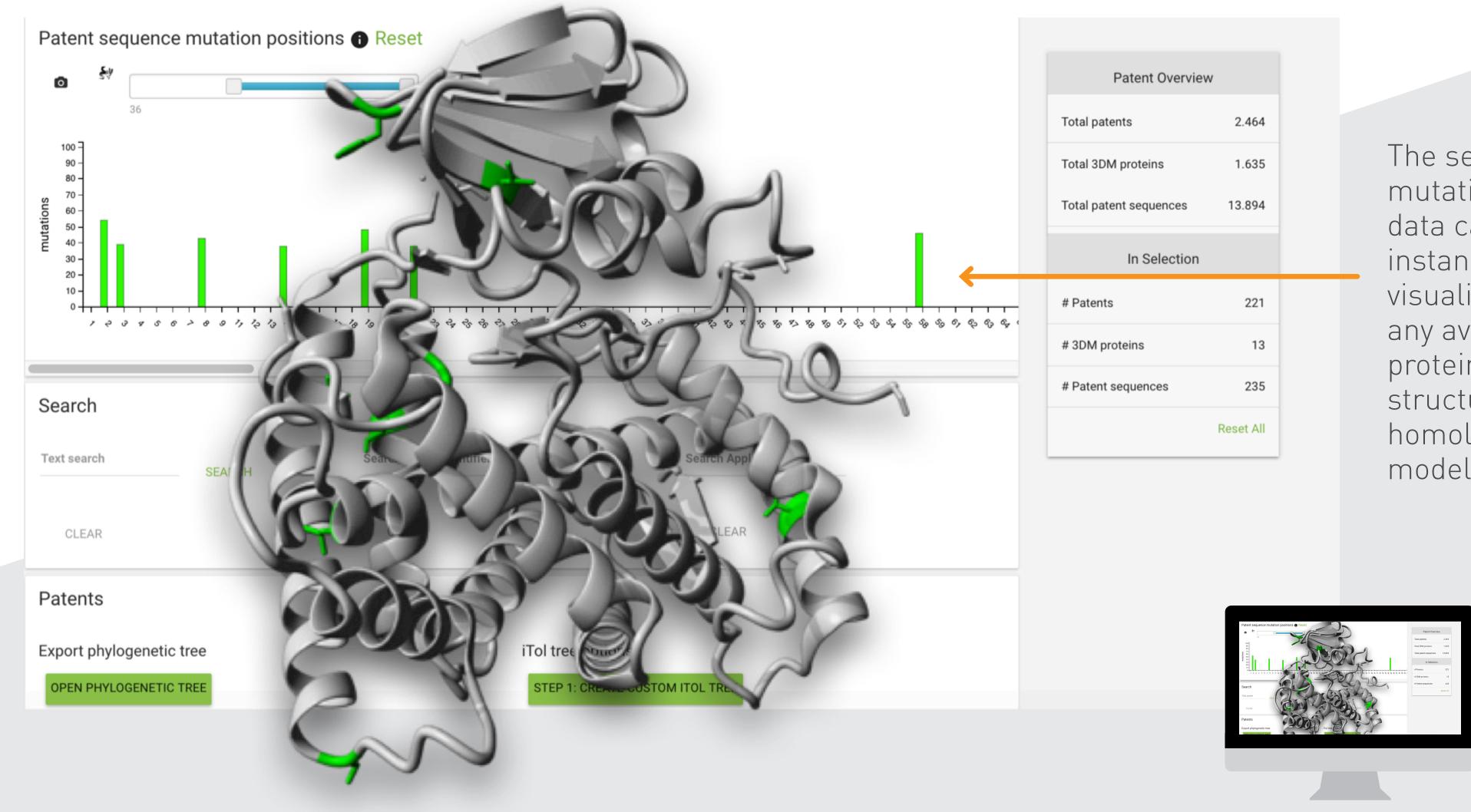
DIFFERENT VISUALISATION OPTIONS



Interactive histogram visualising mutations claimed by novozymes.

In Selection

MANY OPTIONS FOR VISUALIZING PATENT DATA



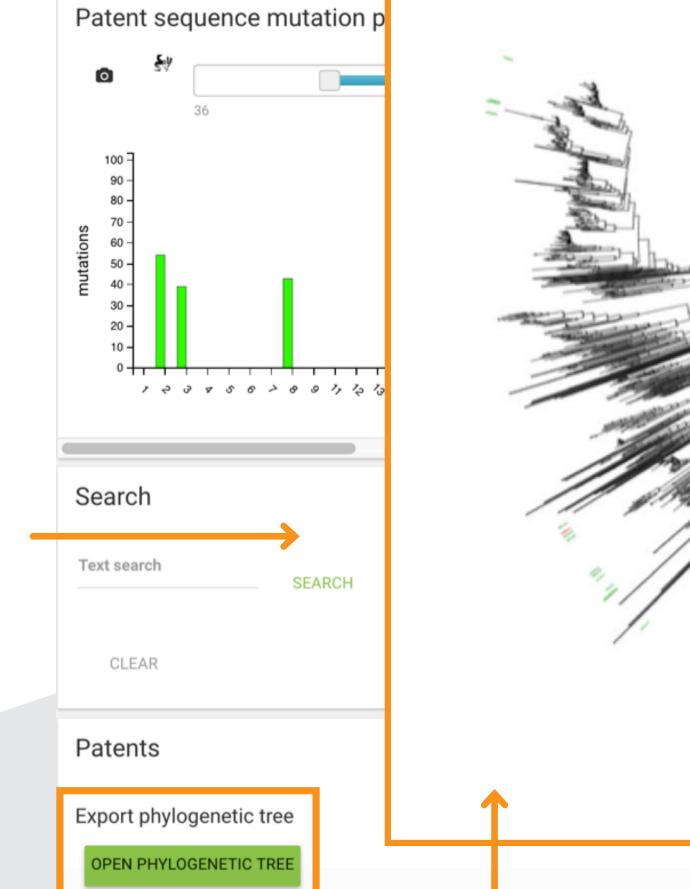


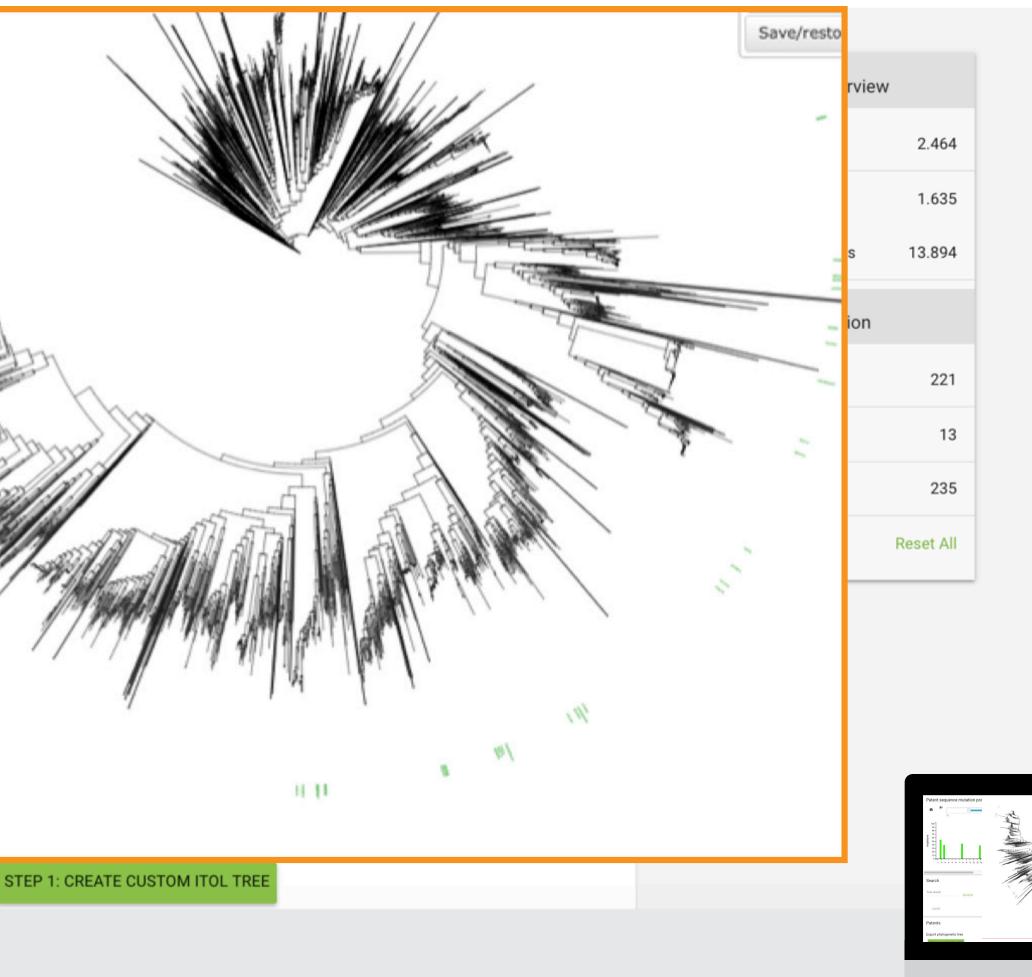
The selected mutation data can instantly be visualized in any available protein structure or homology model.

MANY OPTIONS FOR VISUALIZING PATENT DATA

The patented sequences can instantly be visualized in the phylogenetic tree of the superfamily for FTO analysis.



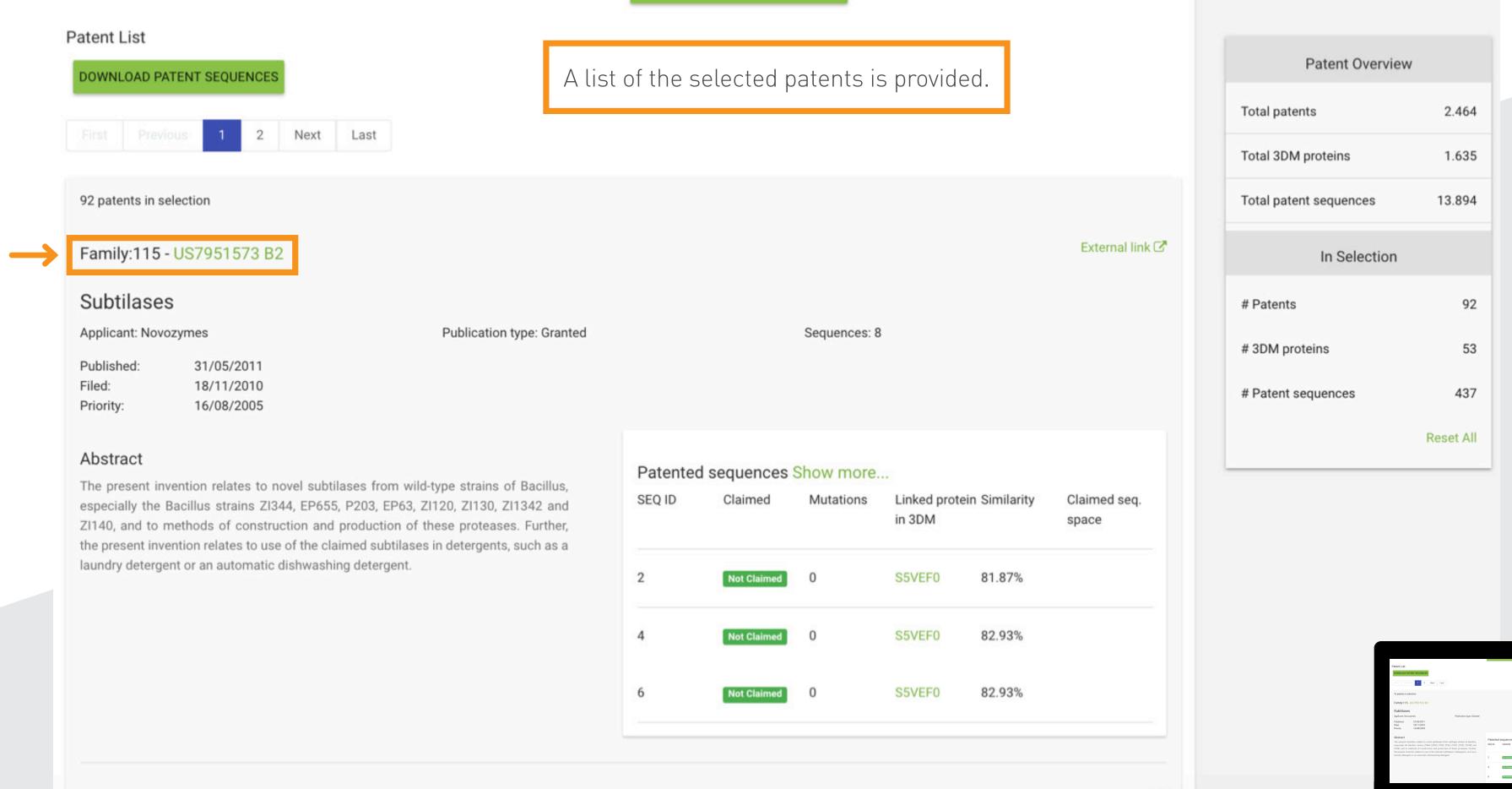






SUMMARY OF THE SELECTED PATENTS IS PROVIDED

Link to interactive analysis tool for each patent



External link CA



INTERACTIVE TOOL FOR ANALYSIS A PATENT SHOWING CLAIMED SEQUENCES AND MUTATIONS

tent US91	33423 B2							
SUBTILASE	VARIANTS						Simple family memb	ers
Applicant: NOVOZ	YMES A/S,					External link 🖉	US20100120091 A1	
subtilase in one catalytic activity.	or more properties including: w The variants of the invention ch as laundry detergent compos	variants exhibiting improvements vash performance, thermal stabilit n are suitable for use in e.g., o itions and dish wash composition:	y, storage stability or leaning or detergent		Type: Published: Filed: Priority:	Granted 15/09/2015 15/01/2010 08/07/2005	modifications: S9R, A15T, V68 wherein the variant has prote	VIEW MINED DATA comprising the following set of BA, S99G, A228V, Q245R, N261D; ase activity and each position he amino acid sequence of SEQ ID
_inked pate	ent sequences						NO: 1.	
EQ ID	Claimed	Linked protein in 3DM	Similarity	Claimed Seq. space	Linked muta	itions	the modifications K27R, *36D	h further comprises one or more of S56P, N62D, N76D, S87N, G97N, V104I, V104N, V104Y, S106A,
A Q S V P Y G V	S9R , A15T , K27R , N62D , V68A N252K	LDN21592904 7, G97, S101, N123, Y167, Q206, A, N76D, S87N, G97N, S101G, S10 S N V K V A V I D S G I D S S	1R, N123S, Y167A, Q20	5E , N218S , N218D , M222S , M22			N204D, V205I, Q206E, L217D, T224S, A232V, K235L, Q236H S101G+V104N, S87N+S101G K27R+V104Y+N123S+T274A,	+V104N,
	Not Claimed	P41362	100%		0		4. The composition of claim 3 cellulase, a lipase, an amylase hemicellulase, an esterase, a polygalacturonase, a beta-gal thereof.	lactase, a glycoam
A OSVPWGI	SRVQAPAAHNRGLTO	G S G V K V A V L D T G I S T H	PDLNIRGGASFV	P G E P S T Q D G N G H G T	H V A G T I A A	LNNSIGVL	5. An isolated DNA sequence 1.	encoding a subtila



MOVING MOUSE OVER DATA WILL HIGHLIGHT THE CLAIM IN WHICH 3DM FOUND THE SELECTED DATA POINT

Subtilisin (2016)

Alignment Superfamily 3DM 🔻 Subset Full Dataset 🔻 Numbering scheme 3DM 🔫

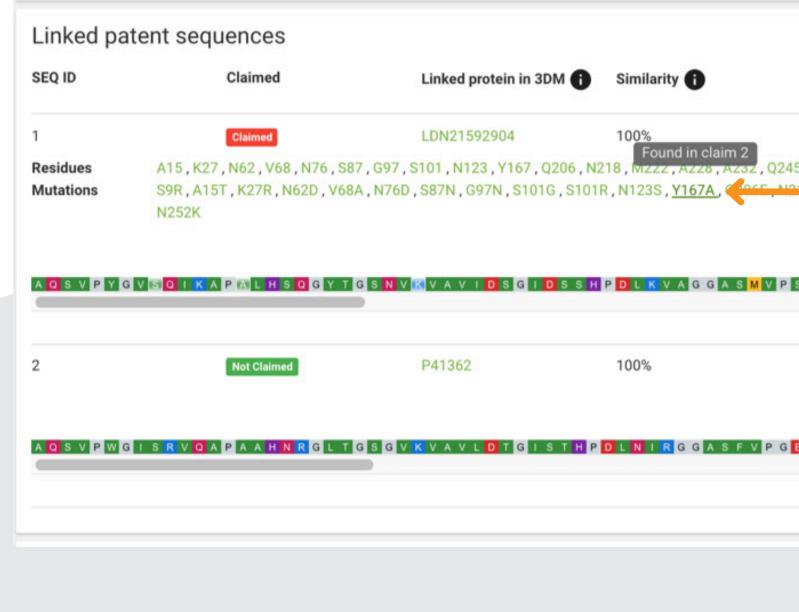
patent US9133423 B2

SUBTILASE VARIANTS

Applicant: NOVOZYMES A/S,

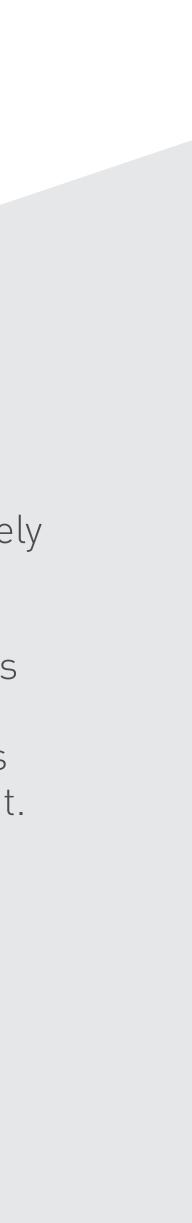
Abstract

The present invention relates to novel subtilase variants exhibiting improvements relative to the parent subtilase in one or more properties including: wash performance, thermal stability, storage stability or catalytic activity. The variants of the invention are suitable for use in e.g., cleaning or detergent compositions, such as laundry detergent compositions and dish wash compositions, including automatic dish wash compositions.

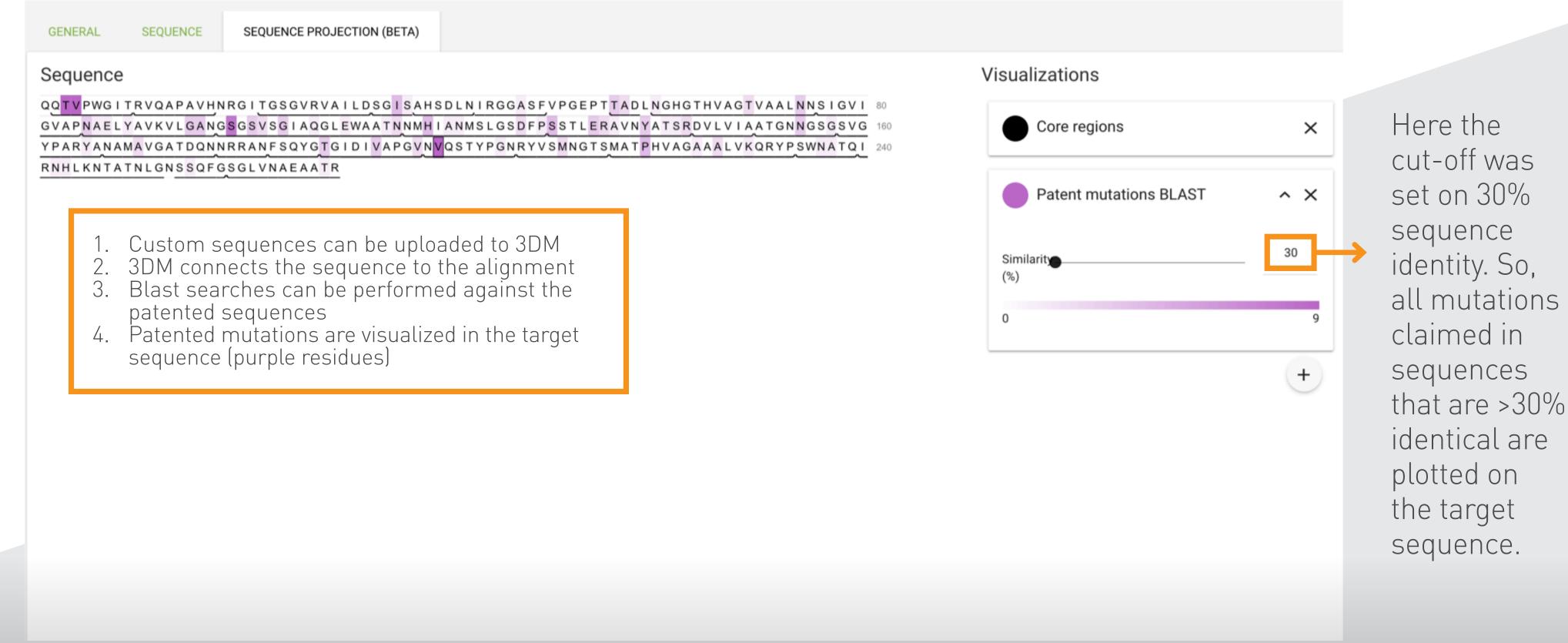




			👤 Henk-Jan Joosten	
		нотѕр	ots subsets Q	
	External link 🗗	Simple family members		
	Type: Granted Published: 15/09/2015 Filed: 15/01/2010 Priority: 08/07/2005	Claims 1. A variant of a subtilisin 309 comprisin modifications: S9R, A15T, V68A, S99G, A wherein the variant has protease activity corresponds to a position of the amino a NO: 1.	228V, Q245R, N261D; and each position	Interactive analyse
Claimed Seq. space	Linked mutations	2. The variant of claim 1, which further co the modifications K27R, *36D, S56P, N62 S101G, S101R, S103A, V104A, V104I, V1 H120D, H120N, N123S, G159D, Y167A, R N204D, V205I, Q206E, L217D, N218S, N2 T224S, A232V, K235L, Q236H, N248D, N2 S101G+V104N, S87N+S101G+V104N, K27R+V104Y+N123S+T274A, and N76D+	D, N76D, S87N, G97N, 04N, V104Y, S106A, 1705, R170L, A194P, 18D, M222S, M222A, 252K, T274A,	claimed sequences and mutations in a patent
SETNPFQDNWSHG	T H W A G T V A A L <mark>12 N</mark> S I G	3. A cleaning or detergent composition, o claim 1 and a surfactant.	comprising a variant of	in a paterit
E P S T Q D G N G H G T H	0 VAGTIAALNNSIGVL	4. The composition of claim 3, which add cellulase, a lipase, an amylase, a cutinas hemicellulase, an esterase, a lactase, a g polygalacturonase, a beta-galactosidase thereof.	e, a prot	Image: Second
		5. An isolated DNA sequence encoding a 1.	No Subtrained in the initial initinitinitial initial ini	Attenting variable Date Analysis 1



TOOL FOR PROJECTING PATENT DATA ON A CUSTOM TARGET SEQUENCE



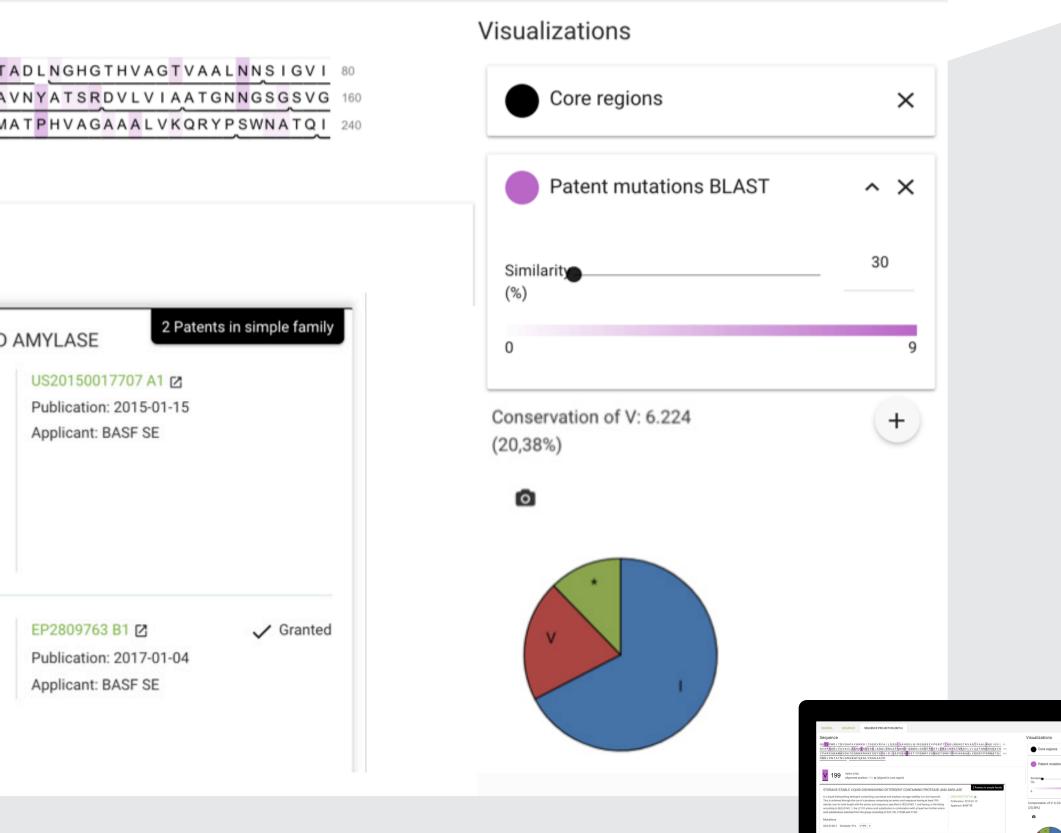
This is a screenshot of 3DM's sequence projection tool. This tool allows for the visualization of data from homologues sequences directly on your target protein. This feature is unique to 3DM and makes 3DM such a powerful tool. As different homologues proteins have different numbering schemes this is very difficult and time consuming to do manually. For instance, if you are interested in a certain amino acid in your target, it is very difficult to find all patents that claim a mutation at that specific position. The 3D-numbering scheme synchronizes the patented sequences to any protein target making such complex searches very easy.



CLICKING ON A PURPLE RESIDUE PROVIDES A LIST OF PATENTS THAT HAVE CLAIMED MUTATIONS AT THE SELECTED POSITION

V PWG I TRVQAPAVHNRG I TGSGVRVA I LDSG I SAHSDLNIRGGASFVPGEPT TA VAPNAELYAVKVLGANGSGSVSGIAQGLEWAATNNMH I ANMSLGSDFPSSTLERAV PARYANAMAVGATDQNNRRANFSQYGTGIDIVAPGVNVQSTYPGNRYVSMNGTSMA HLKNTATNLGNSSQFGSGLVNAEAATR V 199 Valine (Val) Alignment position: 191 ☑ (aligned in core region) STORAGE-STABLE LIQUID DISHWASHING DETERGENT CONTAINING PROTEASE AND In a liquid dishwashing detergent comprising a protease and amylase, storage stability is to be improved. This is achieved through the use of a protease comprising an amino acid sequence having at least 70% identity over its total length with the amino acid substitution in combination with at least two further amino		SEQUENCE	SEQUENCE PROJECTION (BETA)	
VAPNAELYAVKVLGANGSGSVSGIAQGLEWAATNNMHIANMSLGSDFPSSTLERAN PARYANAMAVGATDQNNRRANFSQYGTGIDIVAPGVNVQSTYPGNRYVSMNGTSMA IHLKNTATNLGNSSQFGSGLVNAEAATR Valine (Val) Alignment position: 191 ☑ (aligned in core region) STORAGE-STABLE LIQUID DISHWASHING DETERGENT CONTAINING PROTEASE AND A In a liquid dishwashing detergent comprising a protease and amylase, storage stability is to be improved. This is achieved through the use of a protease comprising an amino acid sequence having at least 70% identity over its total length with the amino acid sequence specified in SEQ ID NO. 1 and having, in the listing according to SEQ ID NO. 1, the L211D amino acid substitution in combination with at least two further amino	equence			
Varian Ama VGATDQNNRRANFSQYGTGIDIVAPGVNVQSTYPGNRYVSMNGTSMA IHLKNTATNLGNSSQFGSGLVNAEAATR Valine (Val) Alignment position: 191 [2] (aligned in core region) STORAGE-STABLE LIQUID DISHWASHING DETERGENT CONTAINING PROTEASE AND A In a liquid dishwashing detergent comprising a protease and amylase, storage stability is to be improved. This is achieved through the use of a protease comprising an amino acid sequence having at least 70% identity over its total length with the amino acid sequence specified in SEQ ID NO. 1 and having, in the listing according to SEQ ID NO. 1, the L211D amino acid substitution in combination with at least two further amino				_
V 199 Valine (Val) Alignment position: 191 ☑ (aligned in core region) STORAGE-STABLE LIQUID DISHWASHING DETERGENT CONTAINING PROTEASE AND A In a liquid dishwashing detergent comprising a protease and amylase, storage stability is to be improved. This is achieved through the use of a protease comprising an amino acid sequence having at least 70% identity over its total length with the amino acid sequence specified in SEQ ID NO. 1 and having, in the listing according to SEQ ID NO. 1, the L211D amino acid substitution in combination with at least two further amino		^		
Alignment position: 191 🖾 (aligned in core region)		^		
	STORAGE In a liquid dis This is achiev identity over	Alignment p -STABLE LIQUI shwashing detergent wed through the use	position: 191 ☑ (aligned in core region) ID DISHWASHING DETERGENT CONTAINING PROTEA nt comprising a protease and amylase, storage stability is to be impr e of a protease comprising an amino acid sequence having at least 7	oved.
	acid substitu Mutations		n the group consisting of S3T, V4I, V193M and V199I.	-
SEQ ID NO:1 Similarity: 91% V199I V	acid substitu Mutations SEQ ID NO:1	tions selected from Similarity: 91%	n the group consisting of S3T, V4I, V193M and V199I.	-





For more information visit our website: www.bio-prodict.nl

or request information via: info@bio-prodict.nl



