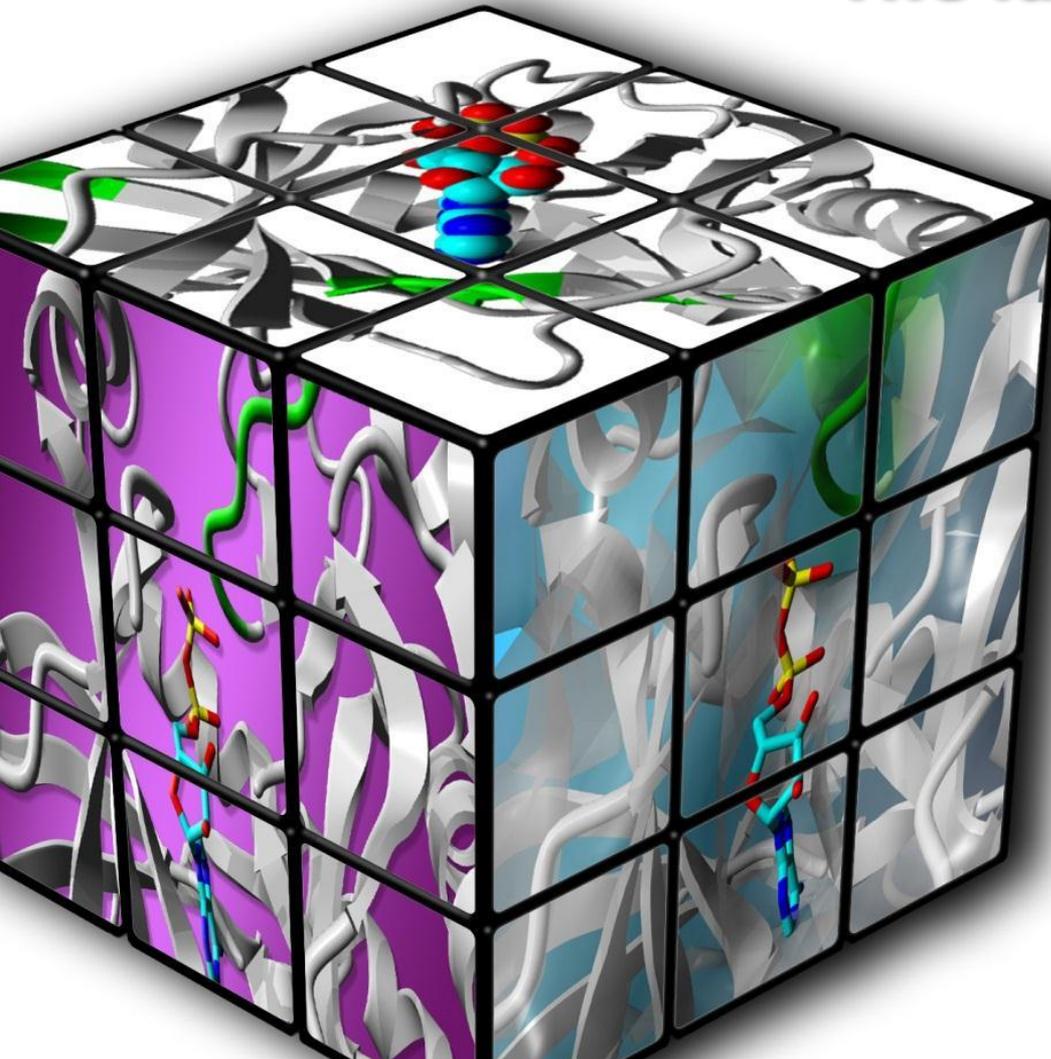


Variant Analysis with HOPE

The last piece of the puzzle



Hanka Venselaar

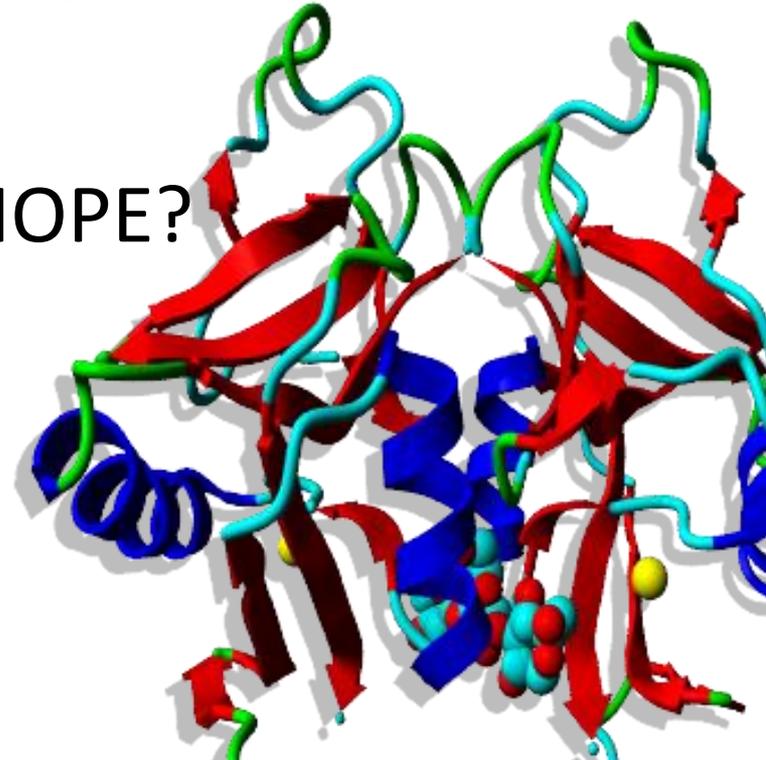
RadboudUMC Nijmegen

**Centre for Molecular and
Biomolecular Informatics**

(Hanka.Venselaar@radboudumc.nl)

Outline

- Structural bioinformatics at Radboud UMC
- Project HOPE, the technical details
- Practical work, how to use HOPE?

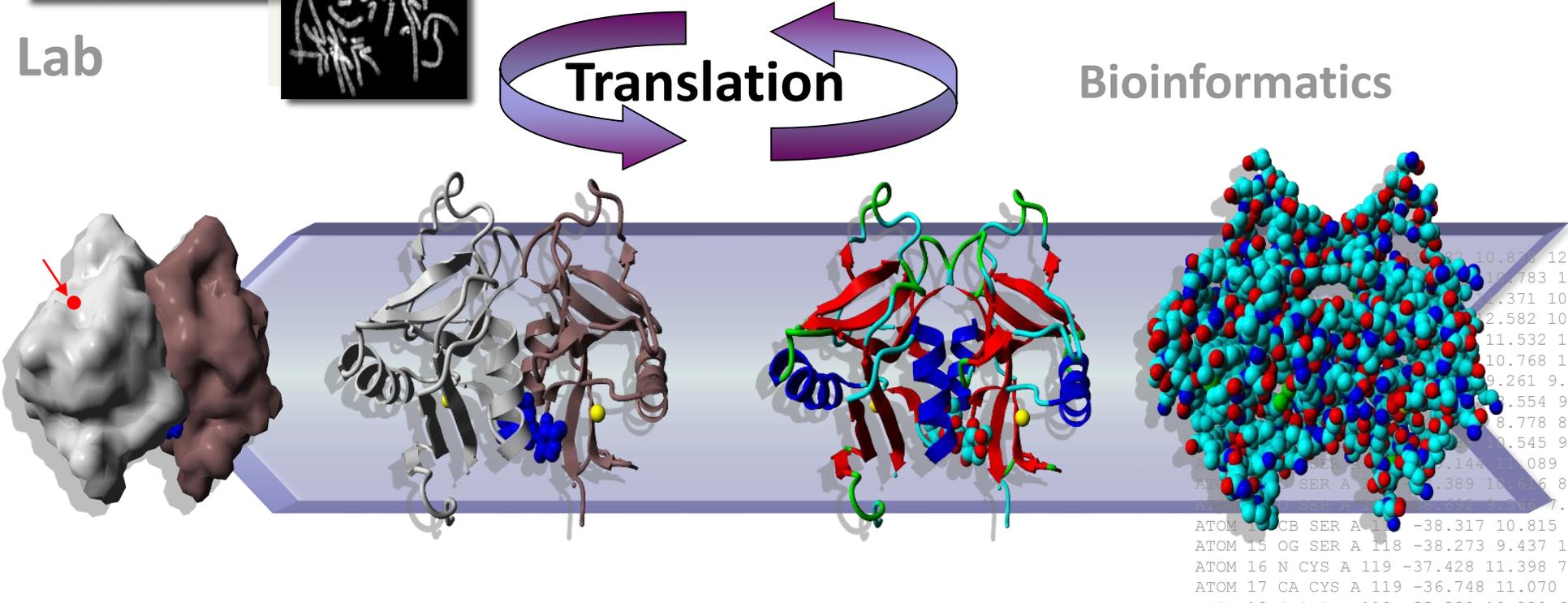
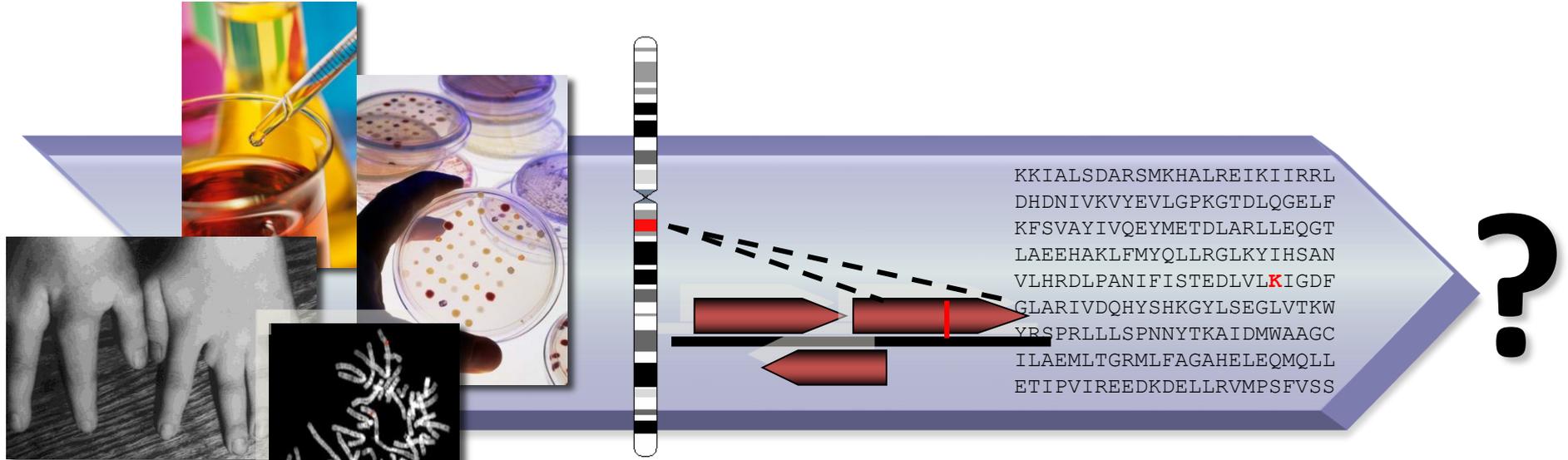




Academic hospital RadboudUMC
Focus on curing patients + understanding disease

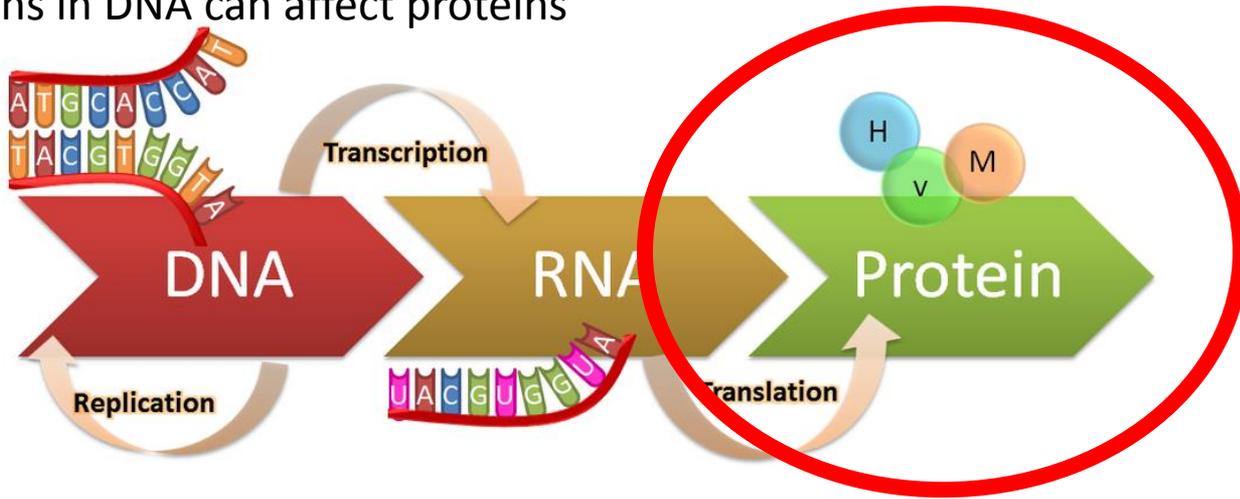


From lab to protein structures ...and back

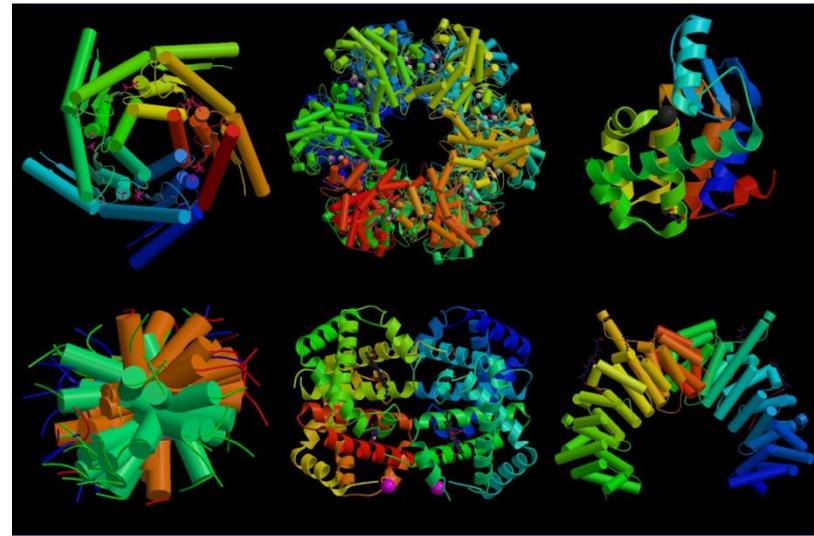
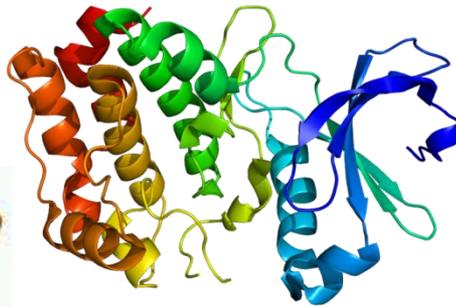
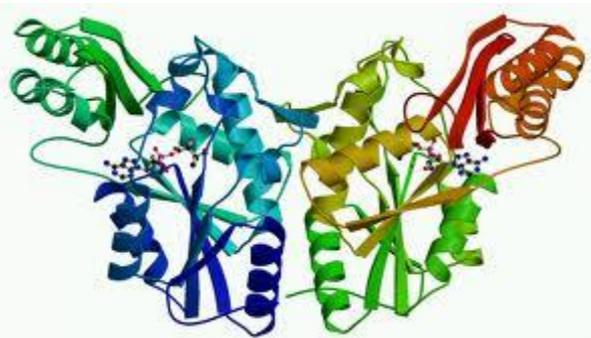


2 things to remember...

1: Mutations in DNA can affect proteins



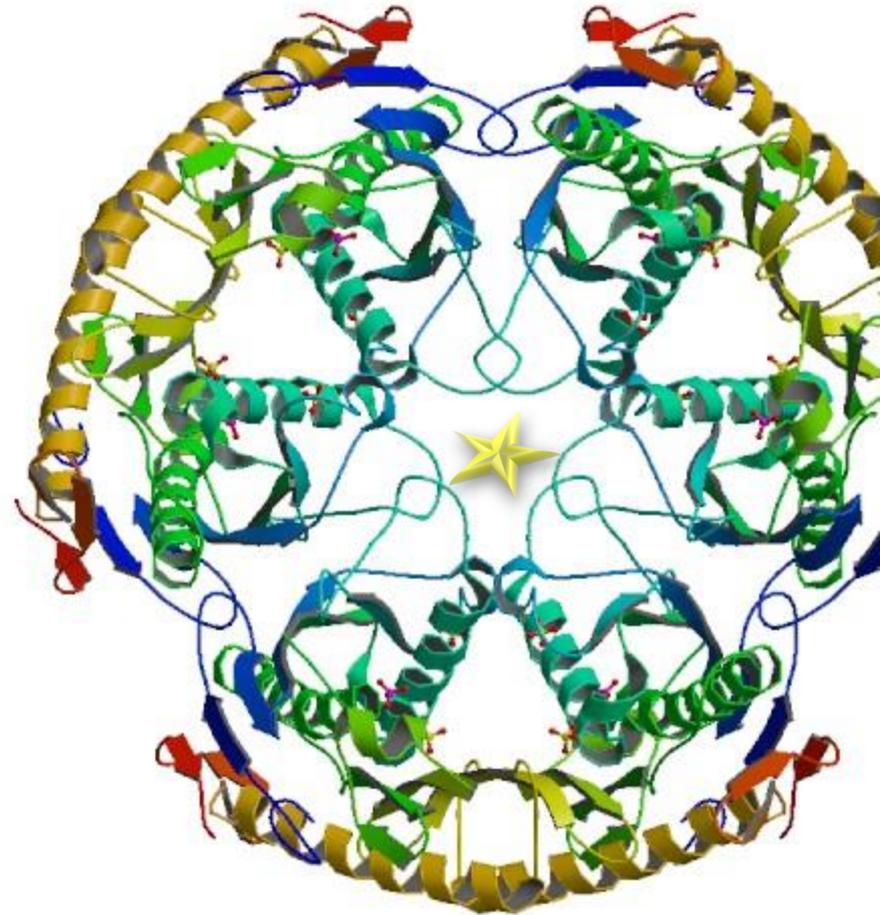
2: Each protein has its own structure that defines its function



Hidden secrets in protein structures

For example:

- Active sites
- Ligands / co-factors
- Surface interactions with other molecules
- Stabilizing interactions
- Transmembrane regions
- Flexible domains
- Etc...etc..etc...

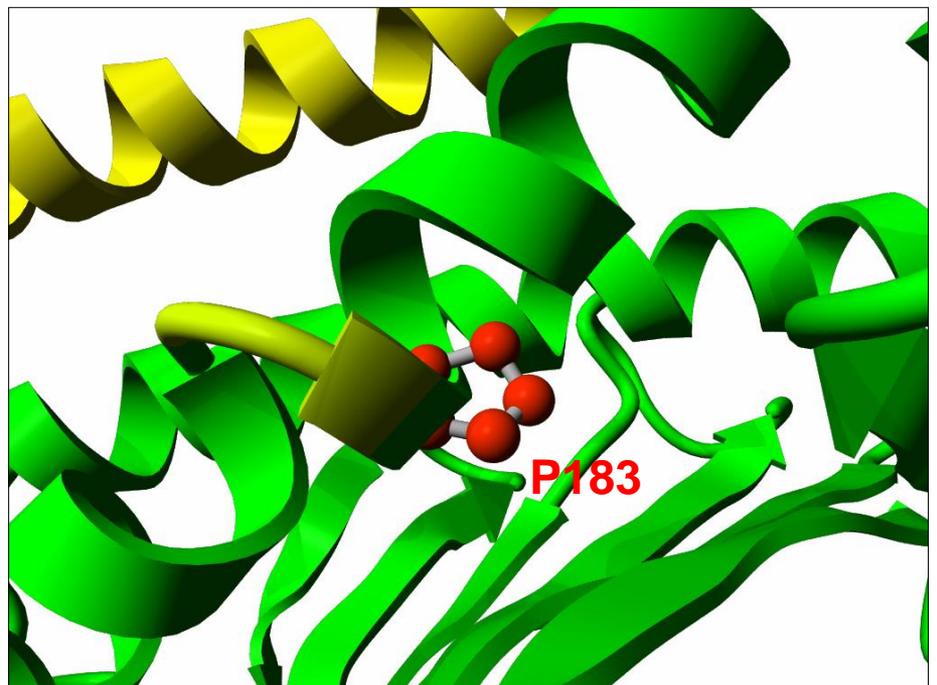
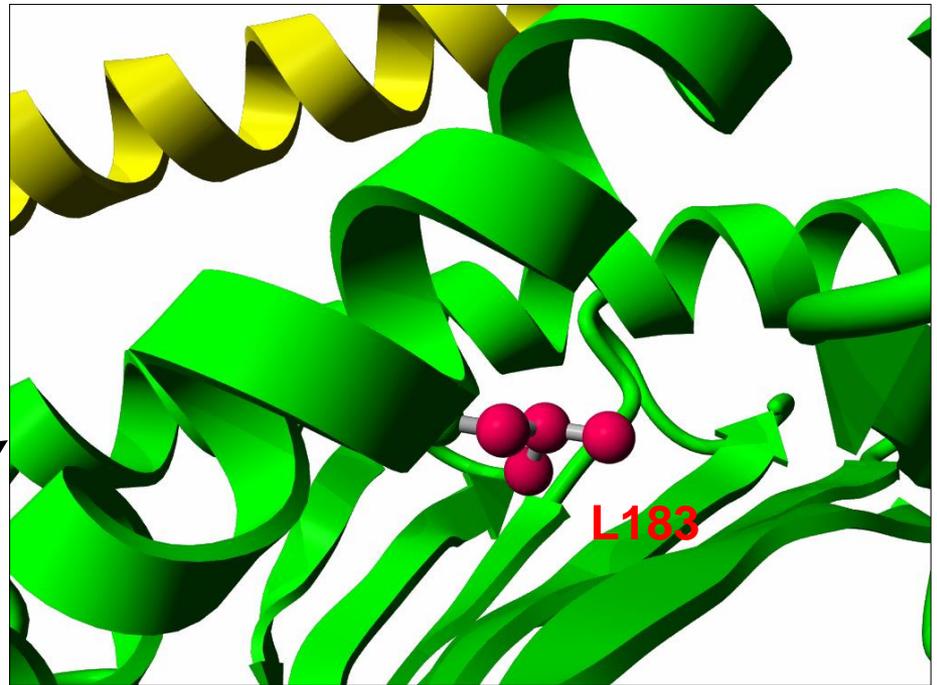
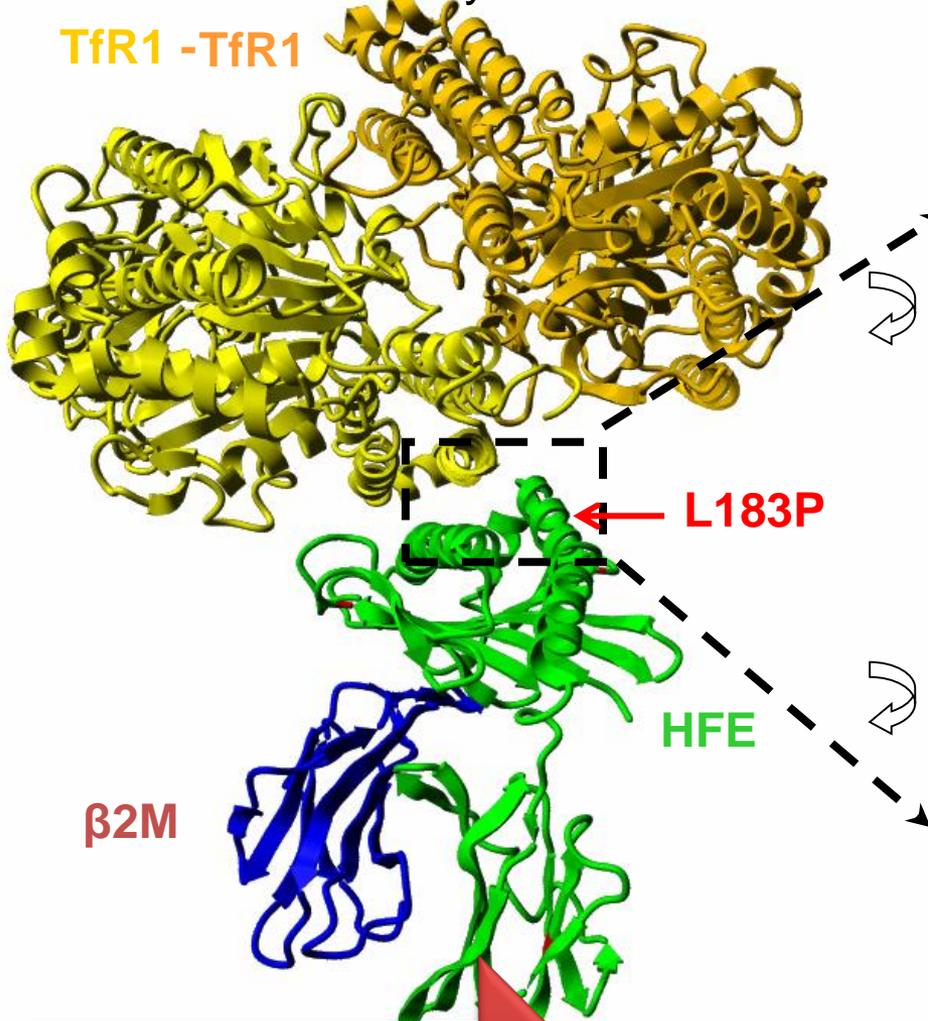


Example: HFE protein

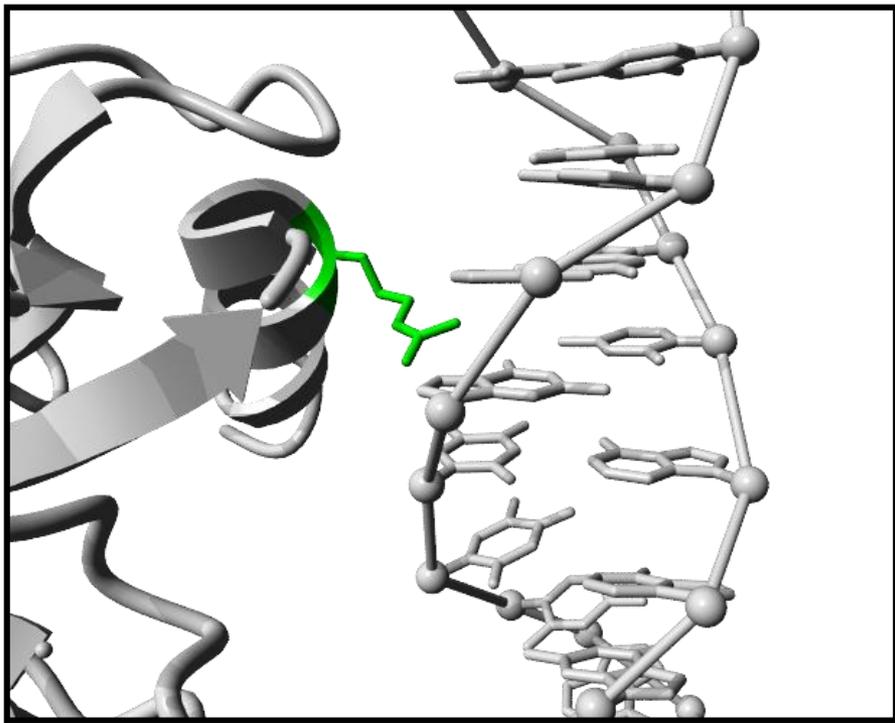
Department of Clinical Chemistry

Mutations -> Hereditary Hemochromatosis

TfR1 -TfR1



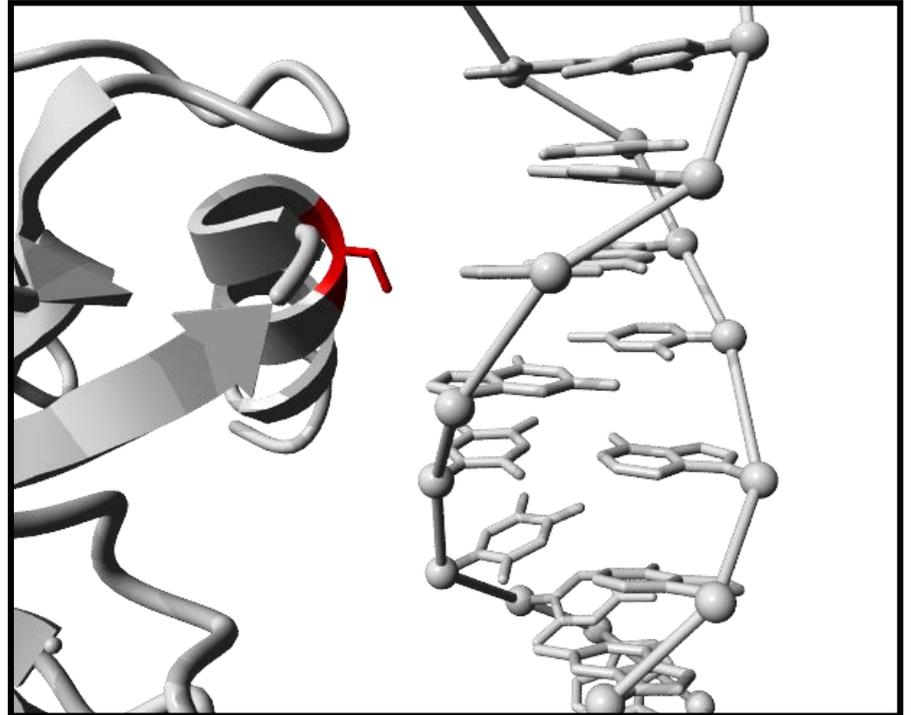
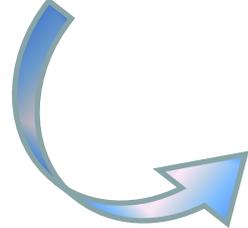
Now used for diagnostics



Arginine

- Loss of negative charge
- Loss of interaction with the DNA

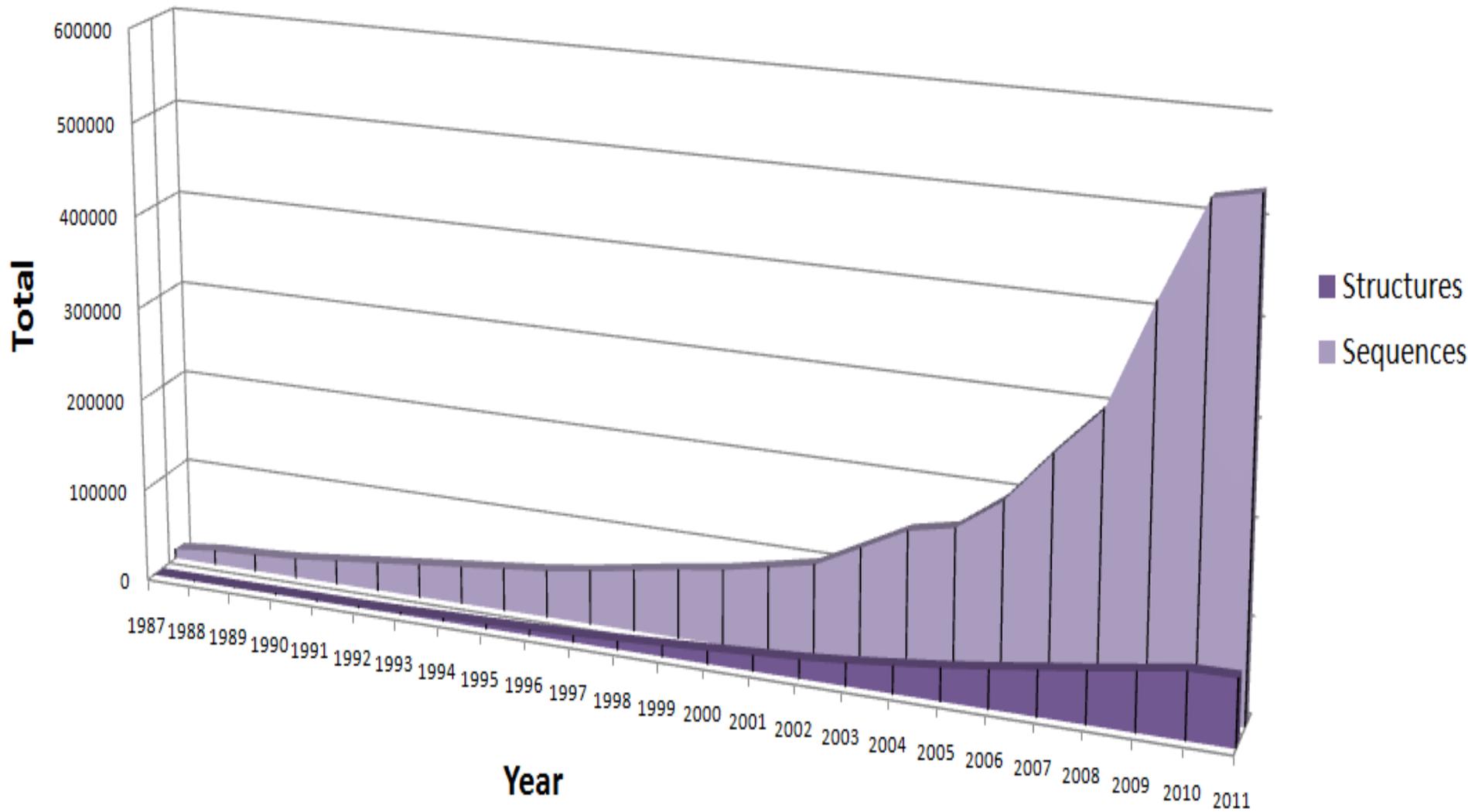
Mutation R→S
In p63



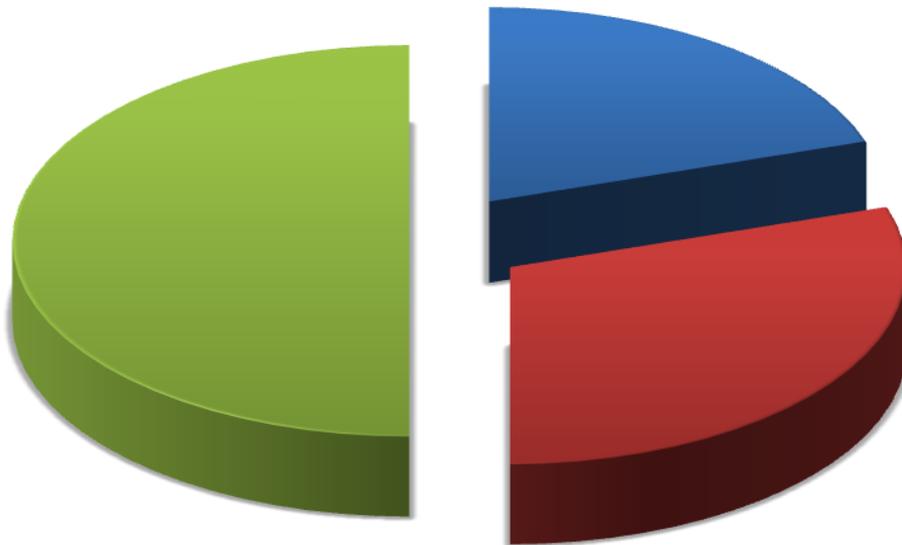
Serine

Protein structures contain a wealth of information...

...there is just 1 problem....



3D-structures for human proteins



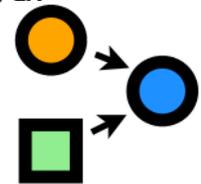
■ 20% Solved 3D-structures



■ 30% Modelling templates



■ 50% No structural information



Predictions/Annotations

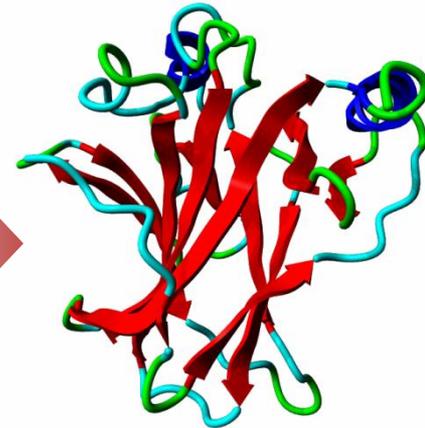
Homology modelling....

MSASTQTNEFLSPEVFQHIWDFLEQPICSVQP
 IDLNFVDEPSEDGATNKIEISMDCIRMQDSDL
 SDMWPQYTNLGLLNSMDQQIQNGSSSTSPYNT
 DHAQNSVTAPSPYAQPSSTFDALSPSPAIPSN
 TDYPGPHSFDVVSFQQSSTAKSATWTYSTELKK
 LYCQIAKTCPIQIKVMTPPPQGAVIRAMPVYK
 KAEHVTEVVKRCPNHELSEFNEGQIAPPSHL
 IRVEGNSHAQYVEDPITGRQSVLVPYEPQVG
 TEFTTVLYNFMCSNCVGGMNRRLIIVTLE
 TRDGQVLGRRCFEARICACPRDRKADEDSIR
 KQVSDSTKNGDGTKRPFQNTHTGIQMTSIIK
 RRSPPDELLYLPVRGRETYEMLLKIKESLELM
 QYLPQHTIETYRRVIDAVRFTLRQTISFPPRD
 EWNDFNFDMDARRNKQQRKEEGE

BLAST against the PDB

SPAIPSN TDYPGPHSFDVVSFQQSSTAKSATW
 SSSVPSQKTYQGSYGFRLGLHSGTAKSVTC

GAVIRAMPVYKKA~~EHVTEVVKRCPN~~HELSE
 GTRVRAMAIYKQSQHMTEVVRRCPHHERCSD



Known 3D structure

MSASTQTNEFLSPEVFQHIWDFLEQPICSVQP
 IDLNFVDEPSEDGATNKIEISMDCIRMQDSDL
 SDMWPQYTNLGLLNSMDQQIQNGSSSTSPYNT
 DHAQNSVTAPSPYAQPSSTFDALSPSPAIPSN
 TDYPGPHSFDVVSFQQSSTAKSATWTYSTELKK
 LYCQIAKTCPIQIKVMTPPPQGAVIRAMPVYK
 KAEHVTEVVKRCPNHELSEFNEGQIAPPSHL
 IRVEGNSHAQYVEDPITGRQSVLVPYEPQVG
 TEFTTVLYNFMCSNCVGGMNRRLIIVTLE
 TRDGQVLGRRCFEARICACPRDRKADEDSIR
 KQVSDSTKNGDGTKRPFQNTHTGIQMTSIIK
 RRSPPDELLYLPVRGRETYEMLLKIKESLELM
 QYLPQHTIETYRRVIDAVRFTLRQTISFPPRD
 EWNDFNFDMDARRNKQQRKEEGE

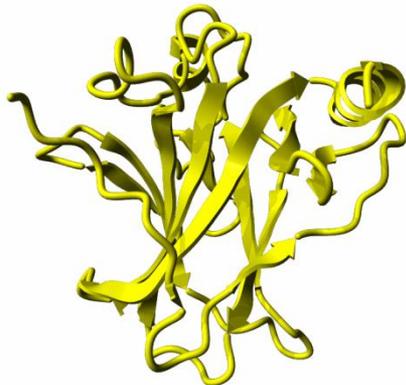
BLAST against the PDB

SPAIPSN TDYPGPHSFDVVSFQQSSTAKSATW
 -----EFLKSSRLTVDS---VDAKATPF

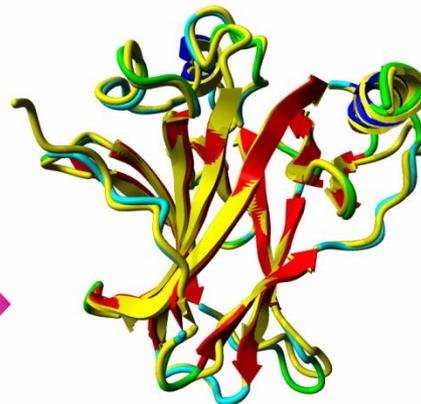
GAVIRAMPVYKKA~~EHVTEVVKRCPN~~HELSE
 ALKMRAMP-----EFLCMNWLNSDDMELS--



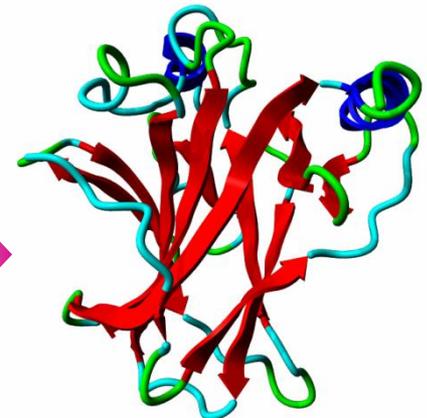
Unknown 3D structure



Homologous 3D structure

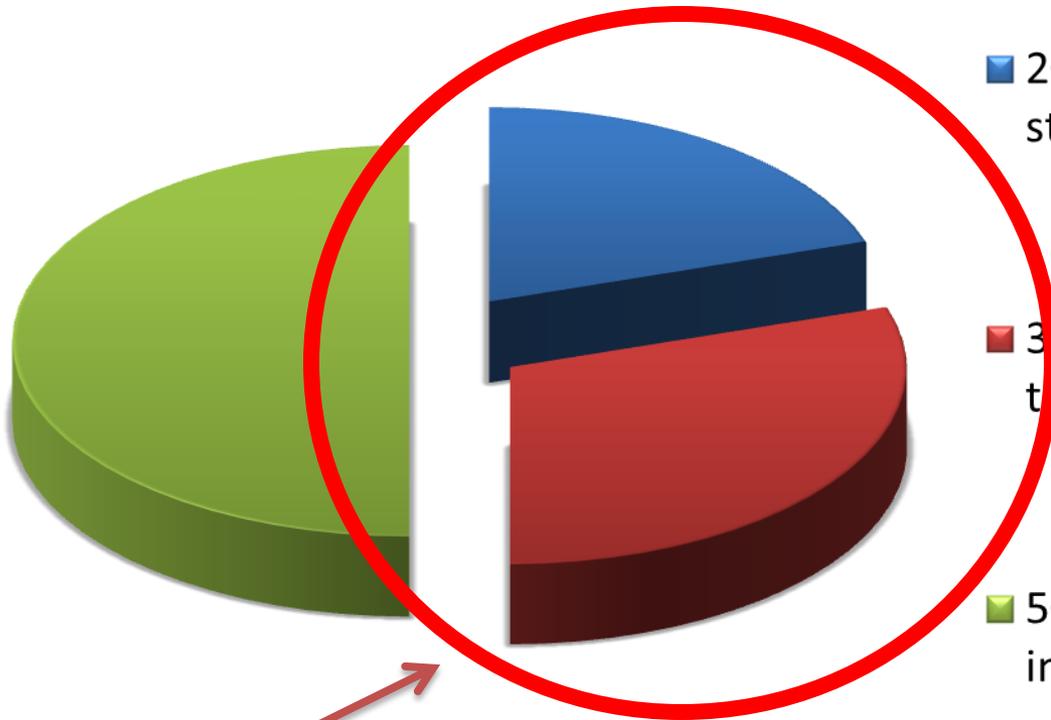


Use the homologous structure to predict the unknown 3D structure



Homology model

3D-structures for human proteins



This is what HOPE will use...

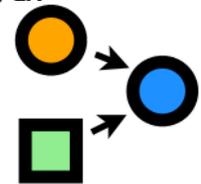
■ 20% Solved 3D-structures



■ 30% Modelling templates

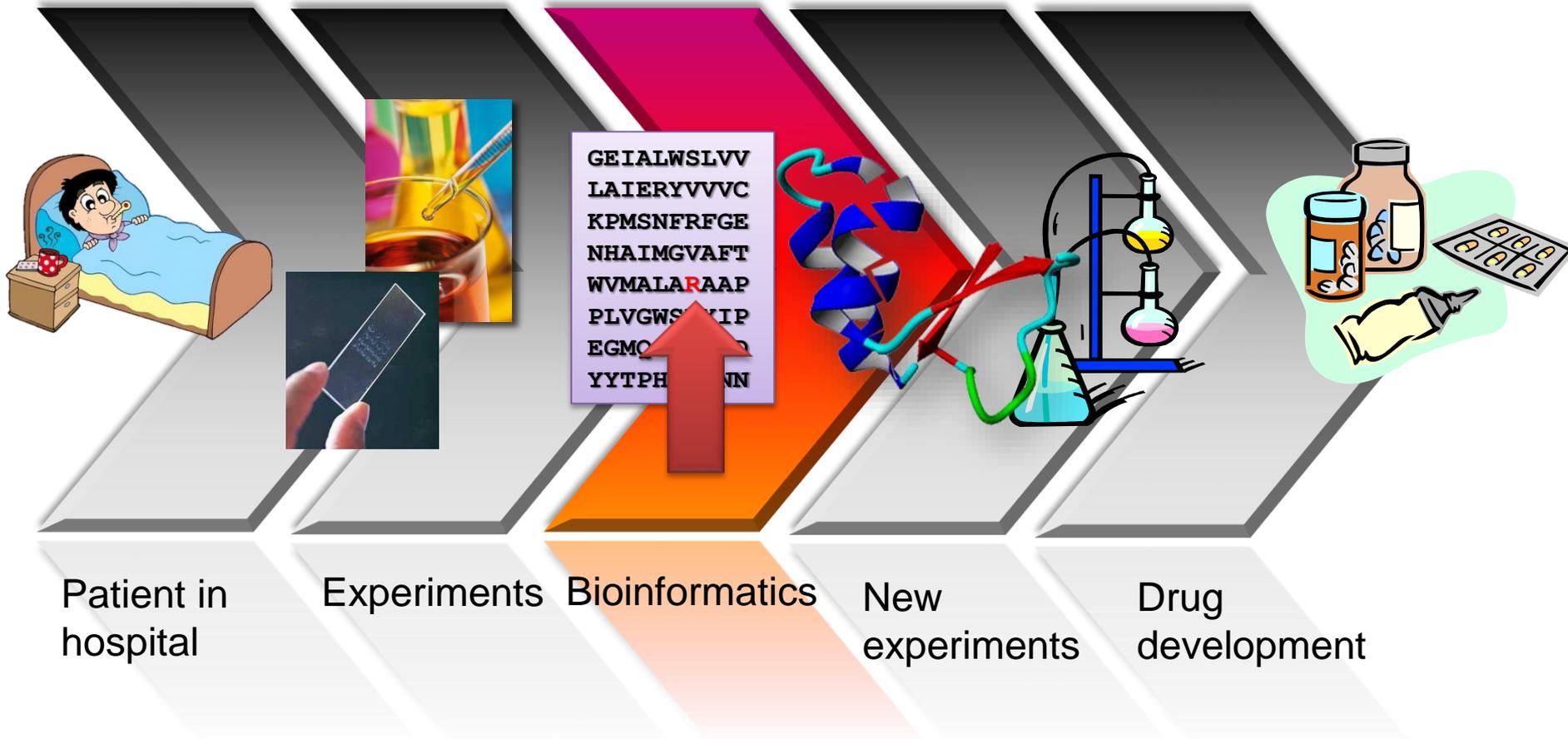


■ 50% No structural information

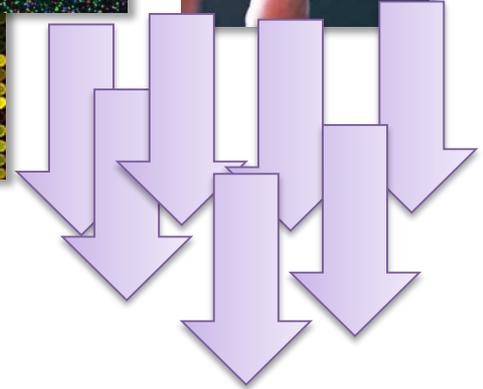
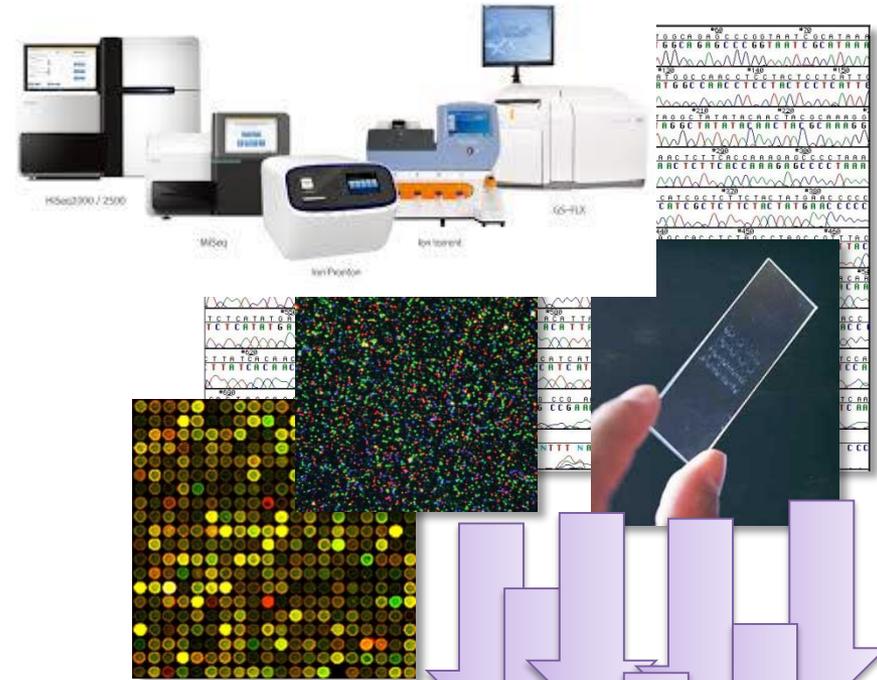
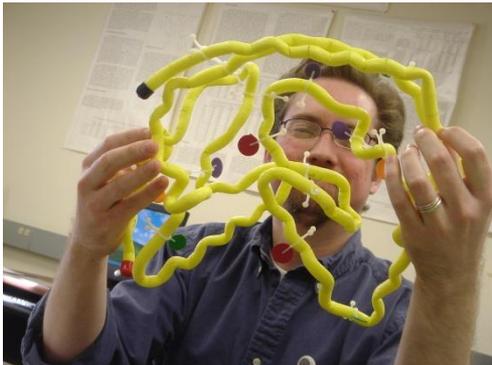
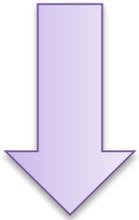


Predictions/Annotations

The Molecular Puzzle: from patient to medicine



Yesteryear vs Nowadays

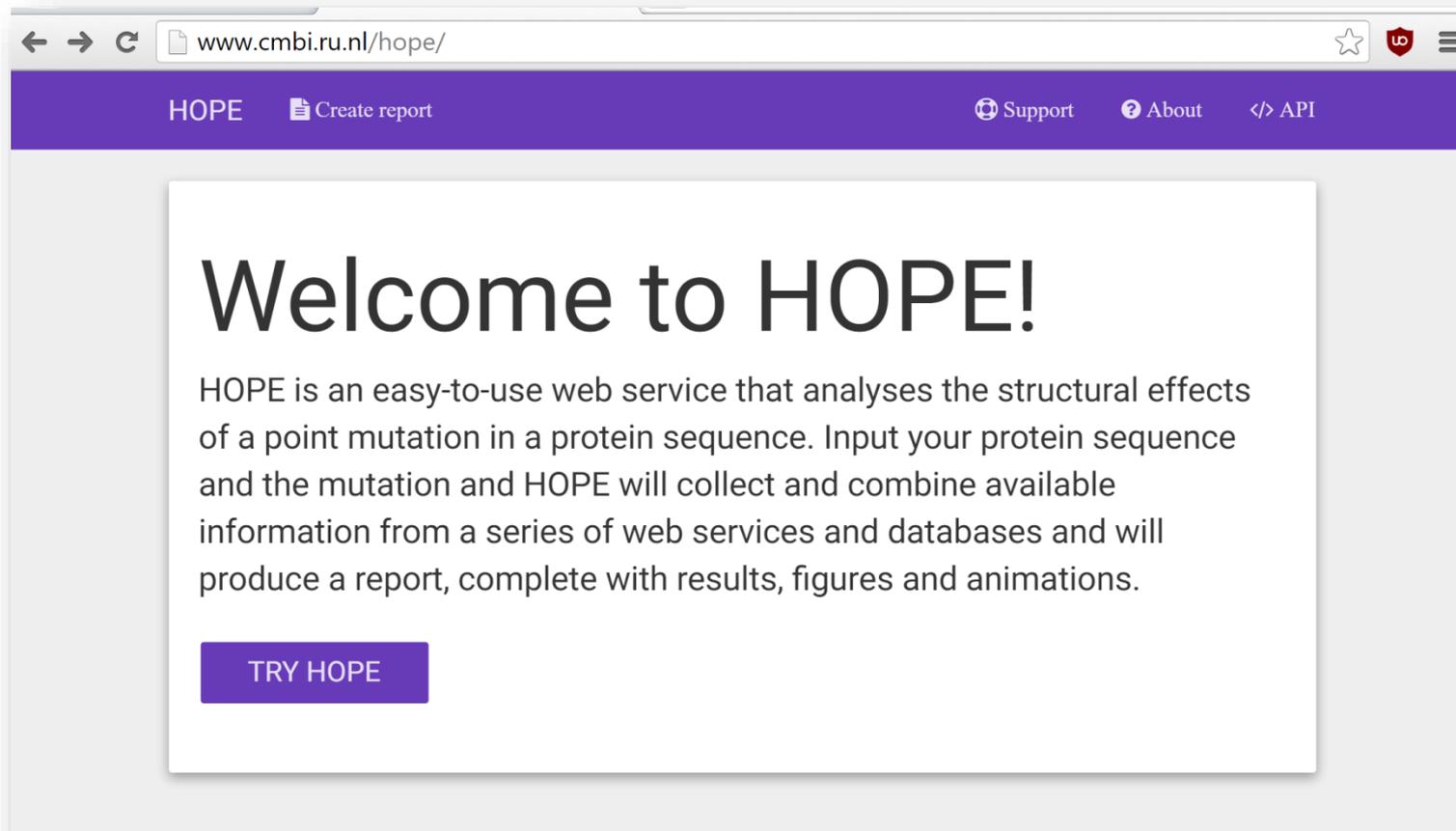


HOPE

So, we developed HOPE....

HOPE, an automatic web server that:

Is easy to use for (bio)medical scientists & Provides structural explanation of point mutations



URL: www.cmbi.ru.nl/hope

Input page: Sequence + Mutation

www.cmbi.ru.nl/hope/input

HOPE Create report Support About API

Input Sequence

Enter a protein sequence or PDB accession code

NEXT

Select Wildtype

Select Mutation

Data collection & combination is hidden for user

www.cmbi.ru.nl/hope/report/574bf68bb6d01b173ce7e36e

HOPE Create report Support About API

Method

The exact 3D-structure of your protein of interest is unknown. However, HOPE is able to build a model of your protein of interest based on a homologous structure. The model will be built using the Yasara & WHAT IF Twinset. Structural information was collected using information from WHAT IF Web services, the UniProt database, and a series of DAS-servers. As a possible modelling template, HOPE will use the structure of your protein of interest can be found in UniProt entry: UBP4...

...the mutation of a methionine into a valine at position 41: ...the schematic structures of the original (left) and the mutant (right) residue, which are the same for each amino acid, is colored red. The side chain of the mutant residue is colored black.

H₂N-CH(CH₂-S-CH₂-CH₂-CH₂-OH)-COOH Mutates into H-CH(CH₂-CH₂-CH₂-OH)-COOH

Each amino acid has its own specific size, charge, and hydrophobicity-value. 1 newly introduced mutant residue often differ in these properties.

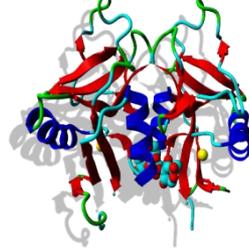
The mutant residue is smaller than the wild-type residue.

The report will evaluate the effect of the mutation on the following features: C residue, structural domains in which the residue is located, modifications on t...

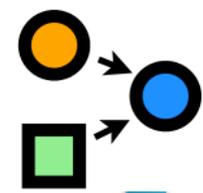
Overview of the protein in ribbon-presentation. The protein is coloured by element, α -helix=blue, β -strand = red, turn=green, 3/10 helix=yellow and random coil=cyan. Other molecules in the complex are coloured grey when present.

Movies

Report page: Extensive explanation of effects on molecular level



Data collection

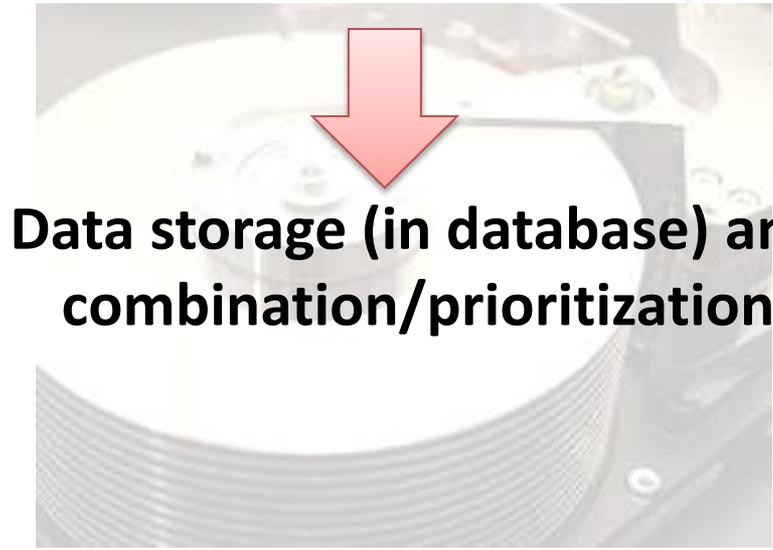
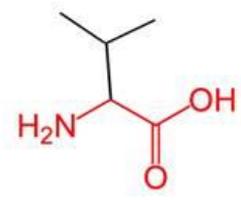
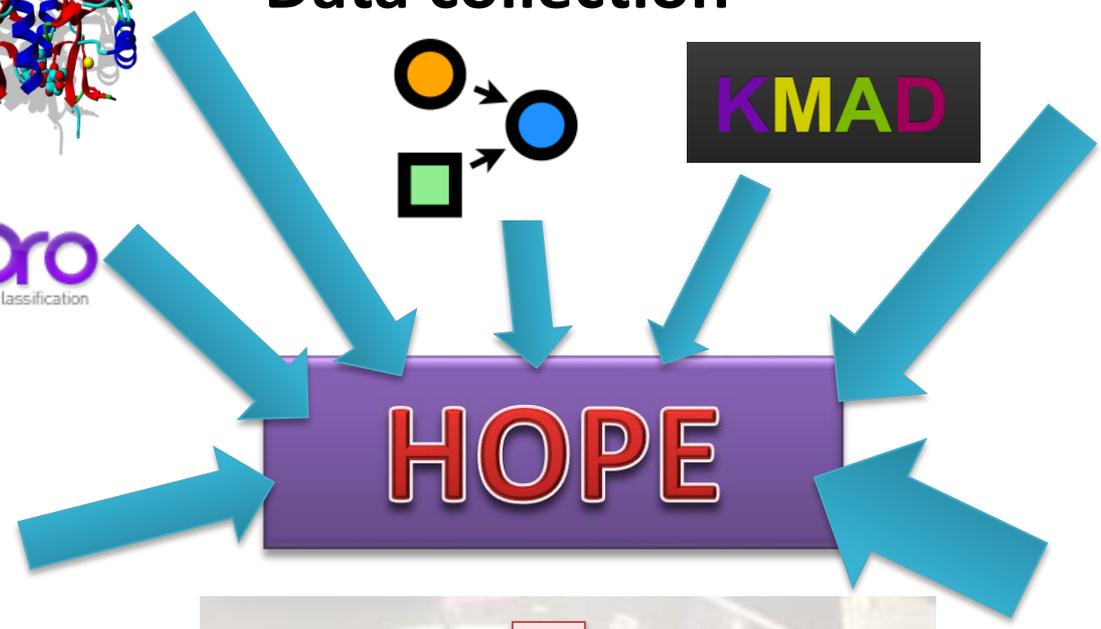


KMAD

HSSP

Interpro
Protein sequence analysis & classification

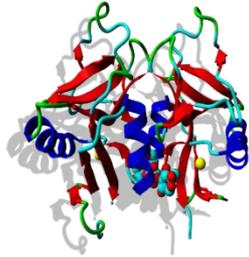
UniProt



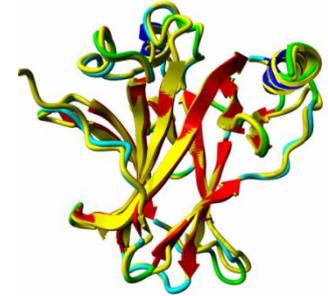
Data storage (in database) and combination/prioritization

Information sources for HOPE:

3D Structures



3D Structure Homolog



Annotated SwissProt/
UniProt features



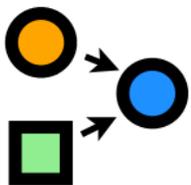
Conservation scores from
multiple sequence alignments



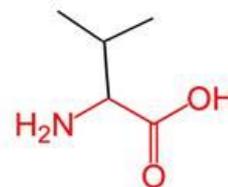
Domain GO-terms



Disorder prediction & motifs

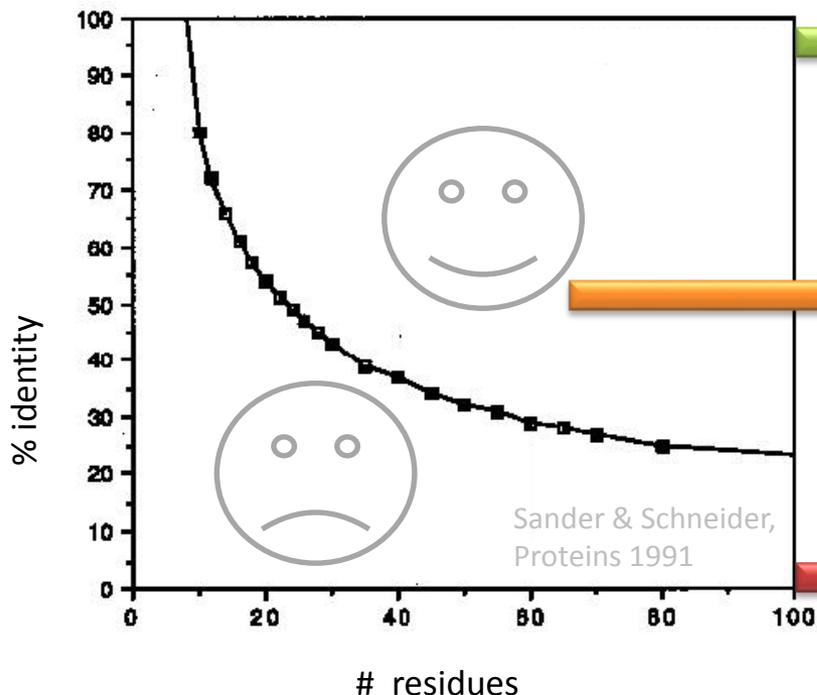


Sequence-based predictions



Amino Acid properties

Data Collection: Structural information



100% BLAST hit, PDB-file = structure
(Decision tree defines best structure based on NMR vs X-ray, resolution, length of the hit)

Enough sequence identity to build model
(Decision tree defines best template based on %identity, length)

No (homologous) structure, use sequence-based predictions



Structure analysis by



Web services.

Examples: metal-contacts, ligand-contacts, hydrogenbonds, ionic interactions, disulfide bridges, torsion angles, secondary structure, accessibility

Data collection: SwissProt annotations



- Collect the sequence features in the SwissProt record
- Information is annotated, checked and sometimes cross linked

Features			
metal	1232	1232	1 Zinc 4.
binding	1142	1142	1 S-adenosyl-L-methionine.
binding	1226	1226	1 S-adenosyl-L-methionine; via amide nitrogen.
mod_res	466	466	1 Phosphoserine.
var_seq	805	825	21 DAEGSTCLHLAAKKGHYEVVQ -> IQKTSKVYTESQETQR SQTIL (in isoform 2). /FTid=VSP_002222.
var_seq	826	1267	442 Missing (in isoform 2). /FTid=VSP_002223.
var_seq	1150	1153	4 DGEV -> ISSA (in isoform 3). /FTid=VSP_002224.

Sequence information
 Length: **1267 aa**, molecular weight **138254 Da**, CRC64 checksum **FF221BEAE9022010**

```

MAADEGSAEK QAGEAHMAAD GETNGSCENS DASSHANAANK HTQDSARVNP QDGTNTLTRI
AENGVSERDS EAAKQNHVTA DDFVQTSVIG SNGYILNKPA LQAQPLRTTS TLASSLPGHA
AKTLPGGAGK GRTPSAFPQT PAAPPATLGE GSADTEDRKL PAPGADVVKVH RARKTMPKSV
VGLHAASKDP REVREARDHK EPKEEINKNI SDFGRQQLLP PFPSLHQSLP QNQCVMATTK
SQTACLPLFVL AAASVSRKKKR RMGTYSVLPK KTKVLKQRT VIEMFKSITH STVGSKGEKD
LGASSLHVNG ESLEMSDED DSEEELEDDG HGAEQAAAFP TEDSRTSKES MSEADRAQKM
DGESEEEQES VDTGEEEEEGG DESDLSSESS IKKKFLKRKG KTDSPWIKPA RKRRRRSRKK
PSGALGSESY KSSAGSAEQT APGDSTGYME VSLDSLRLRV KGILSSQAEG LANGPDVLET
DGLQEVPLCS CRMETPKSRE ITTLANNQCM ATESVDHELG RCTNSVVKYE LMRPSNKAPL
LVLCEDHRGR MVKHQCCPGC GYFCTAGNFM ECQPESSISH RFHKDCASRV NNASYCPHCG
EESKAKEVT IAKADTTSTV TPVPGQEKGS ALEGRADTTT GSAAGPPLSE DDKLQGAASH
VPEGFDPTGP AGLGRPTPGL SQPGKETLE SALIALDSEK PKKLRFHFKQ LYFSARQGEL
  
```

Example Uniprot-features:
 Regions, domains, motifs,
 mutagenesis sites, variants,
 active site, signal,
 glycosylation sites, etc

Data collection: Conservation scores



- Calculated from the multiple sequence alignment in HSSP
- Profile gives % occurrence of residue type on that position
- Precalculated (for PDB-files/templates) or newly generated (sequence)

# SEQUENCE PROFILE AND ENTROPY																											
SeqNo	PDBNo	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D	NOCC	NDEL	NINS	ENTROPY	RELENT	WEIGHT
1	126 A	0	0	0	0	0	0	0	0	67	0	33	0	0	0	0	0	0	0	0	0	3	0	0	0.637	21	0.29
2	127 A	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	291	0	0	0.000	0	1.00
3	128 A	7	0	0	0	0	0	0	0	2	1	87	2	0	0	0	0	0	1	0	0	291	0	0	0.554	18	0.74
4	129 A	0	0	1	0	0	0	1	0	0	3	1	14	0	1	0	0	23	0	53	0	293	0	0	1.366	45	0.32
5	130 A	0	0	1	0	0	0	0	0	1	0	5	59	0	0	0	27	0	3	2	2	294	0	0	1.178	39	0.38
6	131 A	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	1	0	3	3	63	294	0	0	0.995	33	0.47
7	132 A	1	0	0	0	1	0	96	0	0	0	0	0	0	2	0	0	0	0	0	0	295	0	0	0.241	8	0.92
8	133 A	0	0	0	0	0	0	0	0	10	69	0	0	1	4	0	0	14	0	0	0	296	0	0	1.043	34	0.57
9	134 A	0	0	0	0	0	0	0	99	0	0	1	0	0	0	0	0	0	0	0	0	296	0	0	0.079	2	0.98
10	135 A	1	3	1	0	0	0	0	0	3	42	14	4	1	0	0	2	0	8	10	11	297	0	0	1.844	61	0.24
11	136 A	0	5	0	0	2	1	46	0	0	0	0	0	1	46	0	0	0	0	0	0	300	0	0	1.031	34	0.43
12	137 A	1	0	0	0	0	0	0	46	2	0	21	5	0	12	1	0	1	2	5	3	303	0	0	1.673	55	0.37
13	138 A	0	4	0	0	96	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	309	0	0	0.186	6	0.98
14	139 A	0	0	0	0	0	0	0	0	0	0	1	4	4	7	23	3	9	22	1	26	309	0	0	1.868	62	0.23
15	140 A	44	45	8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	309	0	0	1.024	34	0.67
16	141 A	0	1	0	0	0	0	0	31	0	0	36	15	0	1	9	0	0	3	2	1	309	0	0	1.623	54	0.32
17	142 A	0	1	0	0	97	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	311	0	0	0.197	6	0.96
18	143 A	0	33	1	0	0	0	0	3	4	2	4	0	0	0	0	0	50	1	1	1	311	3	296	1.347	44	0.18
19	144 A	0	0	0	0	0	0	0	0	0	7	0	0	0	21	0	7	58	4	2	1	308	0	0	1.276	42	0.50
20	145 A	0	0	0	0	0	0	0	0	1	93	1	0	0	0	0	0	0	0	3	1	309	0	0	0.376	12	0.86
21	146 A	0	0	0	0	0	0	0	42	2	1	44	0	0	0	0	8	0	0	2	0	309	2	308	1.182	39	0.43
22	147 A	1	0	0	0	0	0	0	0	86	2	2	9	0	0	0	0	0	0	0	0	309	0	0	0.562	18	0.73
23		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000	0	1.00
24	149 A	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	98	1	0	0	0	310	0	0	0.121	4	0.95
25	150 A	0	0	0	0	1	0	0	0	0	0	96	1	0	0	0	1	0	0	1	0	311	0	0	0.229	7	0.90
26	151 A	43	0	0	1	0	0	0	0	45	1	0	8	0	0	0	0	1	0	1	0	311	0	0	1.109	37	0.35
27	152 A	0	0	0	6	1	0	0	0	0	2	0	89	0	0	0	0	1	0	0	1	311	0	0	0.503	16	0.77
28	153 A	0	0	0	0	1	56	4	0	0	0	9	0	31	0	0	0	0	0	0	0	311	0	0	1.071	35	0.07
29	154 A	0	0	1	0	0	0	0	0	0	0	0	97	0	0	0	0	0	1	0	0	313	6	179	0.173	5	0.93
30	155 A	0	0	0	0	1	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0	315	0	0	0.111	3	0.99
31	156 A	0	0	0	0	0	0	0	0	0	99	0	0	0	1	0	0	0	0	0	0	321	0	0	0.074	2	0.96
32	157 A	1	0	1	0	0	0	1	0	1	53	1	27	0	0	0	0	0	13	0	2	322	0	0	1.269	42	0.37
33	158 A	2	20	4	2	0	0	0	1	7	4	15	2	1	0	0	4	1	26	0	11	322	0	0	2.136	71	0.14
34	159 A	0	97	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	325	0	0	0.183	6	0.90
35	160 A	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	50	1	0	46	2	325	0	0	0.905	30	0.49
36	161 A	1	0	0	2	0	0	0	0	0	0	0	0	0	0	1	96	0	0	0	0	325	0	0	0.226	7	0.90
37	162 A	2	88	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	327	0	0	0.430	14	0.95

Data collection: Gene Ontology-terms per domain

- InterPro Predicts functional domains
- Annotated with GO-terms, a controlled vocabulary that describes the protein
- Cellular component, molecular function and biological process
- Examples: GO: 0003677 DNA-binding, GO: 0005155 Protein binding,

GO: 003700 sequence-specific DNA binding transcription factor activity



Data collection: Disorder predictions & Motifs

- Knowledge Based **M**ultiple Sequence **A**lignment for Intrinsically **D**isordered Proteins
- Predicts % disorder of the protein, and whether the mutation of interest is located in a disordered region
- Uses Short Linear Motifs from ELM to make alignments and therefore it can also predict whether a mutation will disturb such a motif

The logo for KMAD, featuring the letters K, M, A, and D in a stylized, colorful font. The 'K' is purple, 'M' is yellow, 'A' is green, and 'D' is pink. The letters are set against a dark gray rectangular background.

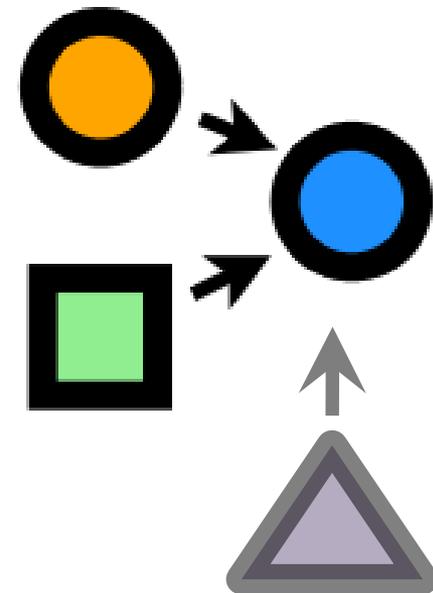
KMAD

Data Collection: Sequence based predictions

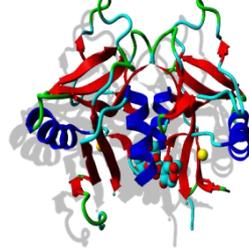
- Prediction servers, based on sequence only
- Accessible using the Uniprot accession code or sequence

Predictions:

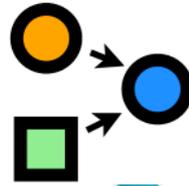
transmembrane domains, secondary structure, phosphorylation sites, accessibility



→Other servers/services can be added



Data collection



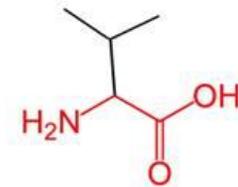
KMAD

HSSP

Interpro
Protein sequence analysis & classification

UniProt

HOPE



Data storage (in database) and
combination/prioritization

Report creation

URL for HOPE

www.cmbi.ru.nl/hope



The gateway page

The screenshot displays the HOPE gateway interface. At the top, a purple navigation bar contains the HOPE logo, a 'Create report' button, and links for 'Support', 'About', and 'API'. Below this, a main content area features a 'Progress' section with a table of tasks and their statuses. A red error message is visible at the bottom, and a 'RETRY' button is located in the bottom right corner.

Task	Status
Blast Pdb	REVOKED
Blast Swissprot Uniprot	FAILURE
Get Reprof Features	PENDING
Get Esp Features	PENDING
Get Interpro Features	REVOKED
Get Pisa Features	PENDING
Get Kmad Features	REVOKED
Get Structure Features	PENDING
Get Uniprot Features	PENDING
Get Hssp Features	PENDING
Create Stills	PENDING
Create Movies	PENDING

HOPE failed to create your report. The system administrator has been notified.

RETRY

Method

The exact 3D-structure of your protein of interest is unknown. However, HOPE is able to build a model of your protein of interest based on a homologous structure. The model will be built using the Yasara & WHAT IF Twinset. Structural information was collected using information from WHAT IF Web services, the UniProt database, and a series of DAS-servers. As a possible modelling template, HOPE identified PDB: [2F1Z](#). More information about your protein of interest can be found in UniProt entry: [UBP48_HUMAN](#). See the [method page](#) for more information.

Amino Acids

You are interested in the mutation of a methionine into a valine at position 415.

The figure below shows the schematic structures of the original (left) and the mutant (right) amino acid. The backbone, which is the same for each amino acid, is colored red. The side chain, unique for each amino acid, is colored black.



Each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties.

The mutant residue is smaller than the wild-type residue.

The report will evaluate the effect of the mutation on the following features: Contacts made by the mutated

The report

Contacts

In the 3D-structure can be seen that the wild-type was involved in a metal-ion contact.

The mutant residue is bigger than the wild-type residue.

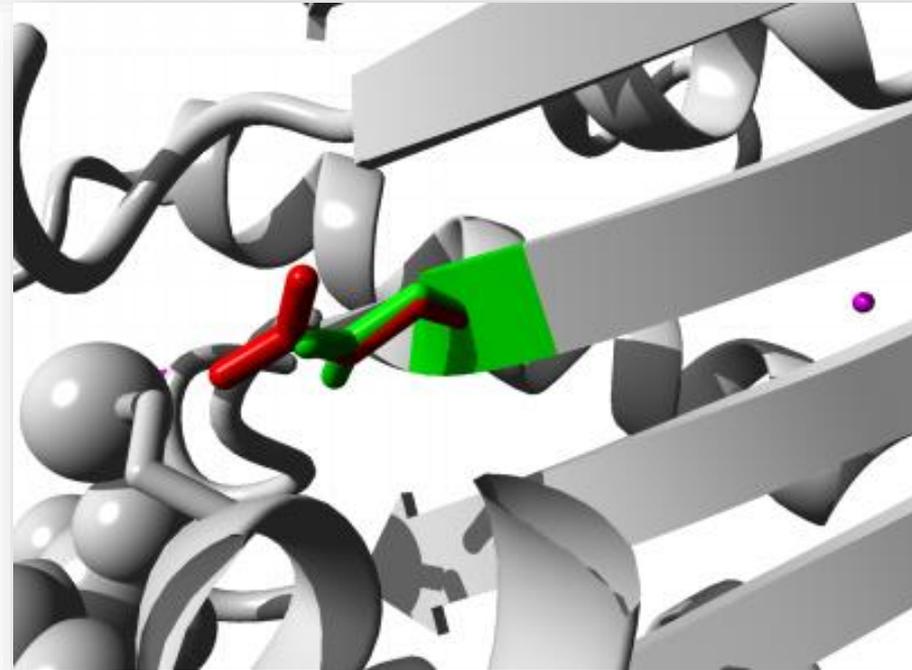
The size differences between the wild-type and mutant residue disturb the interaction with the metal-ion: "MG".

The wild-type residue forms a hydrogen bond with: serine at position 101, serine at position 101,

The size difference between wild-type and mutant residue makes that the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue did.

The wild-type residue forms a salt bridge with: lysine at position 100, arginine at position 200,

This paragraph mentions all contacts made by the wildtype residue and the changes caused by the mutation.



The report

Structure

The mutation is located within a domain, annotated in UniProt as:

- **Kinesin motor**

The mutation introduces an amino acid with different properties, which can disturb this domain and abolish its function.

Structure

In the 3D-structure can be seen that the wild-type residue is located in its preferred secondary structure, a turn. The mutant residue prefers to be in another secondary structure, therefore the local conformation will be slightly destabilized.

Structure

The wild-type residue is a glycine, the most flexible of all residues. This flexibility might be necessary for the protein's function. Mutation of this glycine can abolish this function.

This paragraph will mention special structural features such as transmembrane stretches, regions and secondary structure elements

The report

Predicts whether (and how much of) the protein is disordered and whether the mutation occurs in such a disordered region.



Disorder

The protein is predicted to contain disordered regions (disorder percentage: 52%), but the residue of interest is not predicted to be located in a disordered region.

Conservation

The wild-type residue is very conserved, but a few other residue types have been observed at this position too.

Neither your mutant residue nor another residue type with similar properties was observed at this position in other homologous sequences. Based on conservation scores this mutation is probably damaging to the protein.

Your mutant residue is located near a highly conserved position.



Indicates whether the residue is very conserved (=important) or more variable.

The report

Domains

This residue is part of an interpro domain named [Transforming Growth Factor Beta, Conserved Site](#) ([IPR017948](#)).

This domain is annotated with the following Gene-Ontology (GO) terms to indicate its function:

- [Growth Factor Activity](#) ([GO:0008083](#))

More broadly speaking, these GO annotations indicate the domain has a function in:

- [Protein Binding](#) ([GO:0005515](#))

This residue is part of an interpro domain named [Transforming Growth Factor-Beta-Related](#) ([IPR015615](#)).

This residue is part of an interpro domain named [Transforming Growth Factor-Beta, C-Terminal](#) ([IPR001839](#)).

This domain is annotated with the following Gene-Ontology (GO) terms to indicate its function:

- [Growth Factor Activity](#) ([GO:0008083](#))

More broadly speaking, these GO annotations indicate the domain has a function in:

- [Protein Binding](#) ([GO:0005515](#))

This residue is part of an interpro domain named [Cystine-Knot Cytokine](#) ([IPR029034](#)).

The mutated residue is located in a domain that is important for binding of other molecules. The mutated residue is in contact with residues in another domain. It is possible that the mutation disturbs these contacts.

The mutated residue is located in a domain that is important for binding of other molecules and in contact with residues in a domain that is also important for binding. The mutation might disturb the interaction between these two domains and as such affect the function of the protein.

The mutated residue is located in a domain that is important for binding of other molecules and in contact with residues in a domain that is important for the activity of the protein. The mutation might affect this interaction and thereby disturb signal transfer from binding domain to the activity domain.



Domain paragraph: mentions and links the InterPro domains, their respective GO-terms, and whether 2 domains with an annotated function are in contact.

The report

Predicts whether the residue was located in an annotated motif (taken from ELM database) and whether this mutation changes that motif

Motifs

The following motifs are predicted at this position:

- **Y-based sorting signal** (TRG_ENDOCYTIC_2). The motif is not damaged by the mutation.

Amino Acid Properties

There is a difference in charge between the wild-type and mutant amino acid.

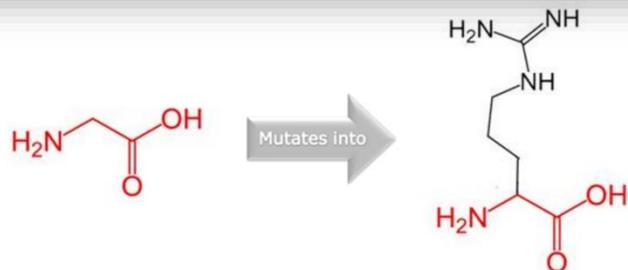
The mutation introduces a charge at this position, this can cause repulsion between the mutant residue and neighboring residues.

The wild-type and mutant amino acids differ in size.

The mutant residue is bigger than the wild-type residue.

The residue is located on the surface of the protein, mutation of this residue can disturb interactions with other molecules or other parts of the protein.

The torsion angles for this residue are unusual. Only glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure.



Last remark: the amino acid compared on their properties such as size, hydrophobicity and charge. Accessibility (if available) is used here.

The report

This will tell you whether the mutation is annotated in dbSNP and Swissprot (Disease vs Polymorphism)



Variants

This mutation matches a previously described variant, with the following description: `In dbSNP:rs11559290.`

See the ExPASy site about this variant: [VAR_062966](#).

The variant is annotated with severity: `POLYMORPHISM`.

Modifications

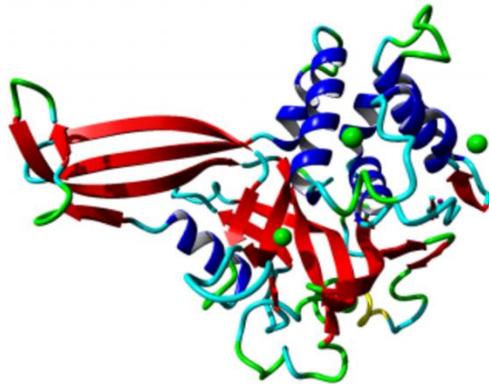
The wild-type residue is predicted (by KMAD) to be a phosphorylation site.

Only serine, threonine and tyrosine residues can be phosphorylated, mutation into another residue type will disturb this modification.



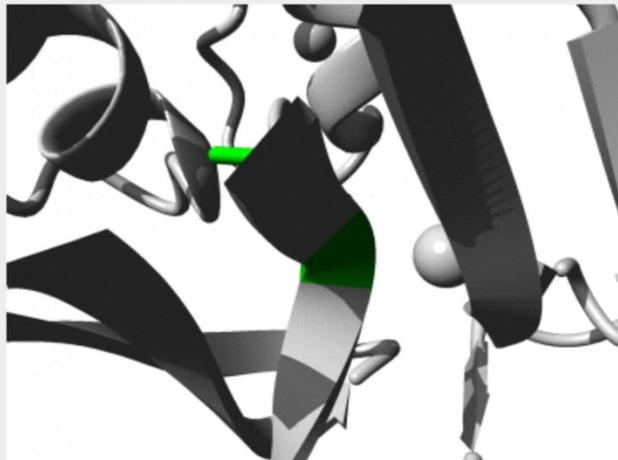
This paragraph will list the annotated/predicted modifications at this position.

The report



Overview of the protein in ribbon-presentation. The protein is coloured by element; α -helix=blue, β -strand = red, turn=green, 3/10 helix=yellow and random coil=cyan. Other molecules in the complex are coloured grey when present.

Movies



Pictures and animated gifs

HOPE → Understanding protein structures and mutations

www.cmbi.ru.nl/hope/report/574bf68bb6d01b173ce7e36e

HOPE Create report Support About API

Method

The exact 3D-structure of your protein of interest is unknown. However, HOPE is able to build a model of your protein of interest based on a homologous structure. The model will be built using the Yasara & WHAT IF Twinset. Structural information was collected using information from WHAT IF Web services, the UniProt database, and a series of DAS-servers. As a possible modelling template, HOPE identified PDB: 2F1Z. More information about your protein of interest can be found in UniProt entry: UBP48_HUMAN. See the method page for more information.

Amino Acids

You are interested in the mutation of a methionine into a valine at position 415.

The figure below shows the schematic structures of the original (left) and the mutant (right) amino acid. The backbone, which is the same for each amino acid, is colored red. The side chain, unique for each amino acid, is colored black.



Each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties.

The mutant residue is smaller than the wild-type residue.

The report will evaluate the effect of the mutation on the following features: Contacts made by the mutated residue, structural domains in which the residue is located, modifications on this residue and known variants

Citing HOPE

Please use the following citation when referencing the results in your report:

Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces.

BMC Bioinformatics. 2010 Nov 8;11(1):548. DOI: [10.1186/1471-2105-11-548](https://doi.org/10.1186/1471-2105-11-548). PubMed: [21059217](https://pubmed.ncbi.nlm.nih.gov/21059217/).

nature genetics

Mutations of *LRT* reading frames, ca

Zubair M Ahmed^{1,13}, Saber Masmo Rob W J Collin^{3,4}, Saima Riazuiddi Abdelaziz Thilif⁵, Bert van der Zwaag Andrew J Griffiths⁶, Ilhem Charfeddine Abdelmoum Ghorbel⁷, Shekhi Ri

Many proteins necessary for sound trans identified through positional cloning of g deafness¹⁻³. We report here that mutation associated with profound nonsyndromic I DFNB63 locus on human chromosome 1 *LRTOMT* has two alternative reading fr two different proteins, *LRTOMT1* and *LK* by protein blot analyses. *LRTOMT2* is a methyltransferase. During evolution, new through partial or complete coalescence evidence that in the primate lineage: *LRT* fusion of two neighboring ancestral genes separate genes (*Lrnc51* and *Tom3*) in rod

We mapped recessive deafness-associated loci segregating in eight families to a 2.64-Mb interval 11q13.3-q13.4 (refs. 5-7). This interval contains the gene for the form (MIM610706) characterized by microtia, mi agnesia^{8,9}. Three of the eight families we recessive mutations of *FGF3* with all of the 6 We used the meiotic recombinations from 1 negative families to refine the linkage interval (Supplementary Fig. 1 online). This interval predicted genes (NCBI build 36.1). Using gen members, we sequenced the protein-coding a approximately 100 bp flanking each exon of a four pathogenic mutations in an unchar

¹Laboratory of Molecular Genetics, National Institute Diagnostic et la Therapie, Centre de Biotechnologie Universitaire Nijmegen Medical Centre, Nijmegen 6500 Tabor 61000, Turkey, ³Center for Molecular and B Neuroscience and Pharmacology, Rudolf Magnus Inst Excellence in Molecular Biology, University of the Poi Dordrecht, Rockville, Maryland 20850, USA, ⁴Serv Medicine, Karadeniz Technical University, Trabzon 6 Behavou, Radoud University Nijmegen, Nijmegen, H.K. (H.Kremers@art.umcn.nl), H.A. (Gidre@acta.un Received 9 April; accepted 28 August; published on

NATURE GENETICS ADVANCE ONLINE PUBLICATION

LETTERS

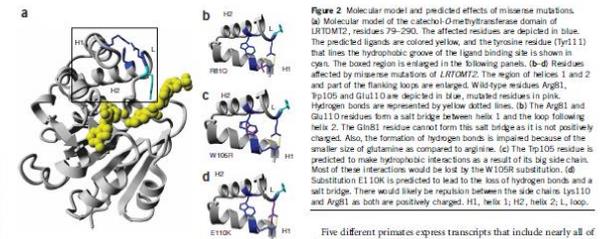


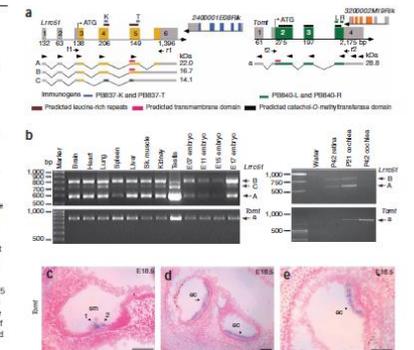
Figure 2 Molecular model and predicted effects of missense mutations. (a) Molecular model of the catalytic O-methyltransferase domain of LRTOMT2, residues 729-920. The affected residues are depicted in blue. The predicted ligands are colored yellow, and the tyrosine residue (Tyr1111) that lines the hydrophobic groove of the ligand binding site is shown in cyan. The boxed region is enlarged in the following panels. (b-d) Residues affected by missense mutations of LRTOMT2. The region of helices 1 and 2 and part of the flanking loops are enlarged. Wild-type residues Arg81, Trp105 and Glu110 are depicted in blue, mutated residues in pink. Hydrogen bonds are represented by yellow dotted lines. (e) The Arg81 and Glu110 residues form a salt bridge between helix 1 and the loop following helix 2. The Glu81 residue cannot form this salt bridge as it is not positively charged. Also, the formation of hydrogen bonds is impaired because of the smaller size of glutamine as compared to arginine. (f) The Trp105 residue is predicted to make hydrophobic interactions as a result of its side chain. Most of these interactions would be lost by the W105R substitution. (g) Substitution E110K is predicted to lead to the loss of hydrogen bonds and a salt bridge. These would likely see re-equilibration between the side chain, Lys110 and Arg81, as both are positively charged. H1, helix 1; H2, helix 2; L, loop.

Five different primates express transcripts that include nearly all of the exons of *LRTOMT* as well as a separate transcript equivalent to rodent *Lrnc51* (Supplementary Fig. 7 online). Inspection of the mouse genome reveals that, in a hypothetical fusion transcript between *Lrnc51* and *Tom3*, if the first translation start codon (ATG in exon 5) were to be used in rodents, an in-frame translation stop codon would be present four codons downstream (Fig. 1d). A fusion protein between *Lrnc51* and *Tom3* in rodents is also unlikely because the first exon of *Tom3* does not have an in-frame consensus splice acceptor site (Supplementary Fig. 2c).

Mouse *Lrnc51* has six exons and is predicted to encode Lrnc51, a 253 residue protein that has two leucine-rich repeats (Fig. 3a). The four

An animal model of *LRTOMT* would be valuable in evaluating the pathophysiology of these mutations. However, in rodents, there are two separate genes, designated *Lrnc51* and *Tom3* (Fig. 3a), which together are orthologous to primate *LRTOMT*. We were unable to detect fusion transcripts of *Lrnc51* with *Tom3* by either RACE or RT-PCR of cDNAs from brain, heart or liver of mouse and rat using all possible combinations of primers depicted in Figure 3a. *LRTOMT* fusion transcripts could be readily amplified from human liver and heart cDNAs (Fig. 1a,b).

Figure 3 Mouse *Lrnc51* and *Tom3* (a) Chromosomal region 7q43 shows conserved synteny with human chromosome 11q13.3. Unlike human, mouse has two separate genes, *Lrnc51* and *Tom3*, encoding *Lrnc51* and *Tom3*, respectively. Translation of *Tom3* mRNA starts in exon 2. This ATG of *LRTOMT* conserved in primates and located in human exon 5 (Supplementary Fig. 2d). Right-pointing arrowheads in exons 1-5 of *Lrnc51* indicate forward RT-PCR primers used in all possible combinations with reverse primers (left-pointing arrowheads) in *Tom3* and cDNAs from mouse brain, liver and heart. No mouse fusion transcripts were recovered (data not shown). Arrows (1), (2) and (3) indicate primer pairs for expression profiling. (b) PCR analyses of *Lrnc51* and *Tom3* transcripts show ubiquitous expression. Primers are as in (a). (c-d) *Tom3* sense and antisense cDNA probes were hybridized to sagittal sections of whole mouse embryos from embryonic day 12.5 to 18.5. No signal was detected using the control sense probe (data not shown). (e) At E18.5 specific staining is visible in the region of the sensory cells of the cochlea where outer hair cells (arrowhead 1) and inner hair cells (arrowhead 2) are located. (f) At E18.5 in the utricle (arrow and saccule (arrowhead)), a clear signal can be observed in the region of the sensory cells. (g) In E18.5 sensory epithelium of the cristae ampullares, *Tom3* mRNA was detected (arrow). No other tissues showed staining for *Tom3* at E18.5. Scale bars, 100 μm. sm, scala media; ec, endolymph compartment.

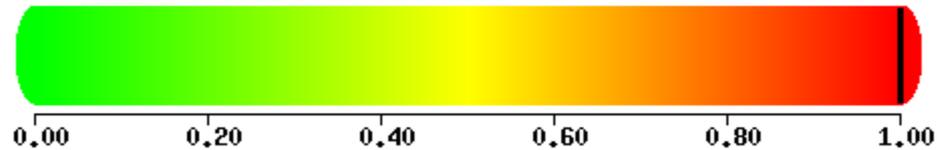


NATURE GENETICS ADVANCE ONLINE PUBLICATION

3

The difference between HOPE and other methods

This mutation is predicted to be **PROBABLY DAMAGING** with a score of **1.000** (sensitivity: **0.00**; specificity: **1.00**)



Method	Result
Grantham-score	89
PolyPhen-2	Probably damaging
SIFT	Not tolerated
SNP&GO	Disease 10
SNAP	Non-neutral 70%
SNPs3D	-1.43
MutPred	0.82
nsSNPanalyzer	Neutral
Panther	-2.93
PHD-SNP	Disease 2

Method	Result
Grantham-score	101
PolyPhen-2	Probably damaging
SIFT	Not tolerated
SNP&GO	Disease 10
SNAP	Non-neutral 93%
SNPs3D	-2,35
MutPred	0.918
nsSNPanalyzer	Disease
Panther	-6,95
PHD-SNP	Disease 9



HOPE report R287W

Contacts

The wildtype residue forms a hydrogen bond with the Leucine on position 271 and with the Tyrosine on position 237. The size difference between wild-type and mutant residue makes that the new residue is not in the correct position to make the same hydrogenbond as the original wild-type residue did. The difference in hydrophobicity will affect hydrogenbond formation. The difference in charge will disturb the ionic interaction made by the original, wild-type residue.

Conservation

Only this residue type was found at this position. Mutation of a 100% conserved residue is usually damaging for the protein.

Domains

The mutated residue is buried in a domain that is important for binding of other molecules. The differences between the wild-type and mutant residue might disturb the core structure of this domain and thereby affect the binding properties.

Amino acid properties

There is a difference in charge between the wild-type and mutant amino acid. The charge of the buried wild-type residue is lost by this mutation. The wild-type and mutant amino acids differ in size. The mutant residue is bigger than the wild-type residue. The wild-type residue was buried in the core of the protein. The mutant residue is bigger and probably will not fit.

Validation experiment: Grantham scores, SIFT, PolyPhen and HOPE

Dataset 1:
Mutations
selected from
publications
115

Dataset 2:
Mutations from
in-house
projects
66

Neutral SNPs
46

Analysis of these variants by 4 different methods. Results were compared to our own manual analyses.

Grantham scores

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	-2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	0	-2	-3	-2	1	0	-4	-2	-3	
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-1	0	-1	-4	-3	-3	
C	0	-3	-3	0	9	-3	-4	-3	-3	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1	
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	-1	-2	-3	-1	0	-1	-3	-2	
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-2	-3	-3	-2	0	2	-2	-3	-3	
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	-2	
I	-1	-3	-3	-3	-1	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	-3	
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	1	-1	-3	-2	
F	-2	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	0	6	-4	0	-1	-1	-2	
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	4	7	0	-1	-1	-2	
S	-1	-1	1	0	-1	0	0	-1	-2	0	-1	-2	0	-1	-2	1	-1	-1	-1	
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	0	-1	-1	-1	
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-3	-1	-1	-4	0	-1	-1	-1	-4	
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	0	-1	-1	-3	
V	0	3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	1	-2	0	-1	-1	-2	

SIFT

SIFT predicts whether an amino acid substitution affects protein function. SIFT predictions are based on the degree of conservation of amino acid residues to sequence alignments derived from closely related sequences collected through PSI-BLAST. SIFT can be applied to naturally occurring missense mutations, polymorphisms or laboratory induced sequence variations.

Submitted to: I. Craig Bester

Protein: SIFT Human SNPs

Method: SIFT Human SNPs

Database: UniProtKB/Swiss-Prot

Query: SIFT Human SNPs

Results: SIFT Human SNPs

Summary: SIFT Human SNPs

Details: SIFT Human SNPs

Legend: SIFT Human SNPs

Footer: SIFT Human SNPs

PolyPhen-2

PolyPhen-2 Prediction of functional effects of human nsSNPs

Home About Help Downloads Batch query Write's log

PolyPhen-2 (Polymorphism Phenotyping v2) is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. Please, use the form below to submit your query.

15-Feb-2012: PolyPhen-2 server has been updated to utilize version 2.2.2 of the software; protein sequences from UniProtKB/UniRef100 Release 2011_12 (14-Dec-2011), structures from PDB/OSDP Snapshot 03-Jan-2012 (78,304 entries) and UCSC Multiz multiple alignments of 45 vertebrate genomes with hg19/GACv37 human genome (08-Oct-2009)

Query Data

Protein or SNP Identifier:

Protein sequence in FASTA format:

Position:

Substitution: AA: ARNDCEGQHILKMFPSSTWYV
AA: ARNDCEGQHILKMFPSSTWYV

Query description:

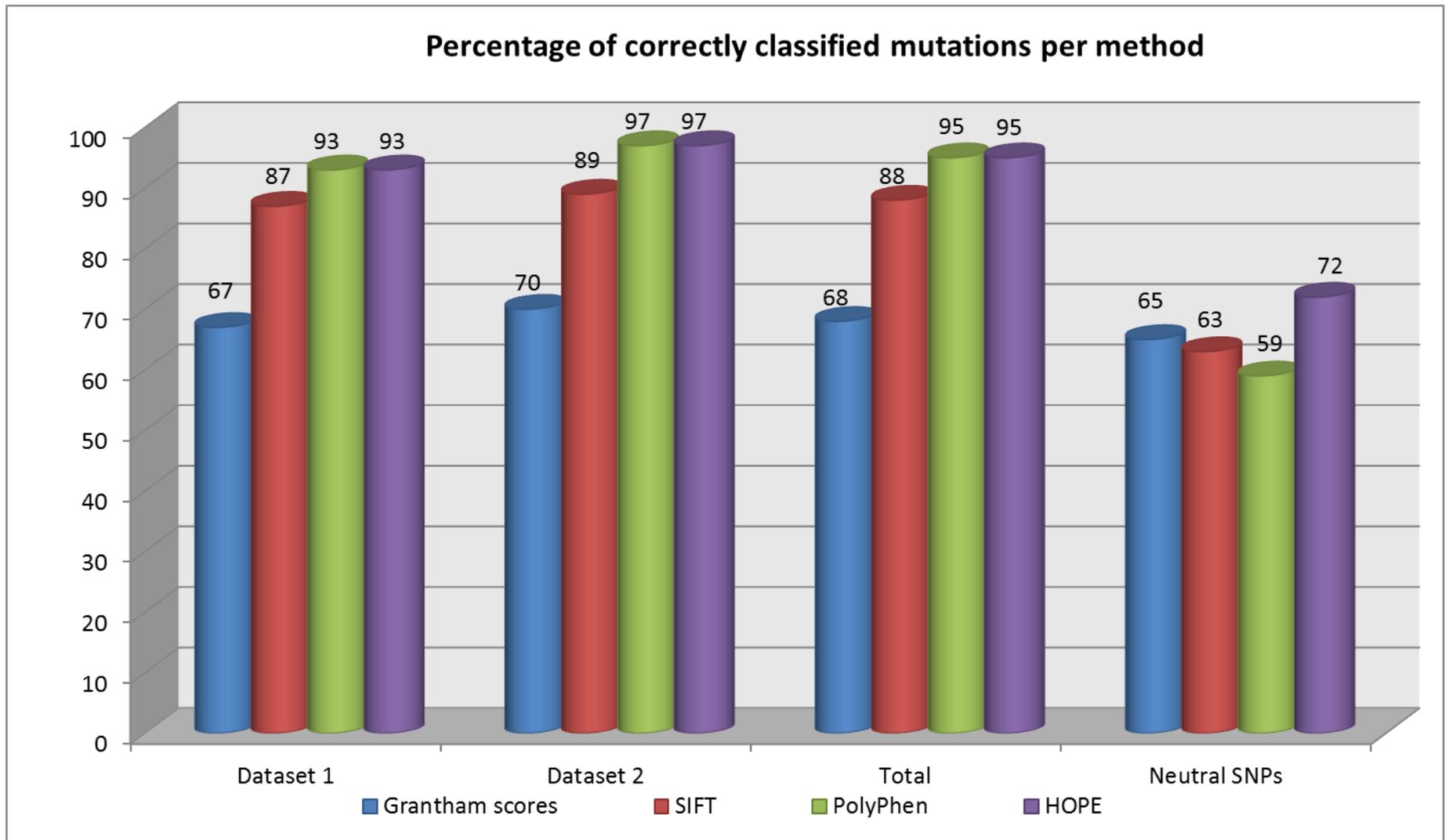
HOPE

Amino Acid / Sequence based



Structure based

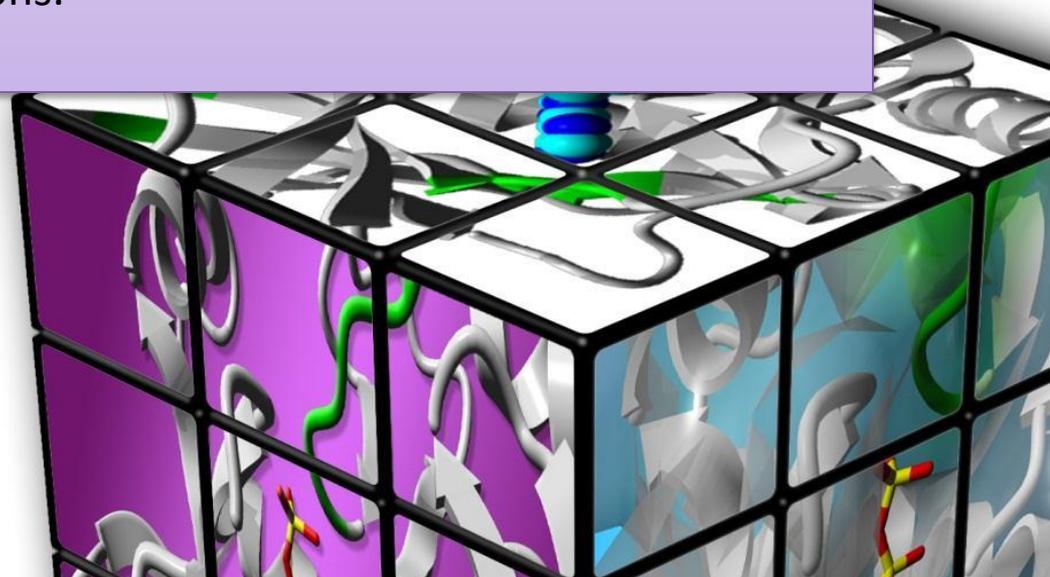
The results: HOPE's score equals PolyPhen2



So, using multiple sequences improves the predictions, using 3D-structure information improves the results even more.

Take home messages....

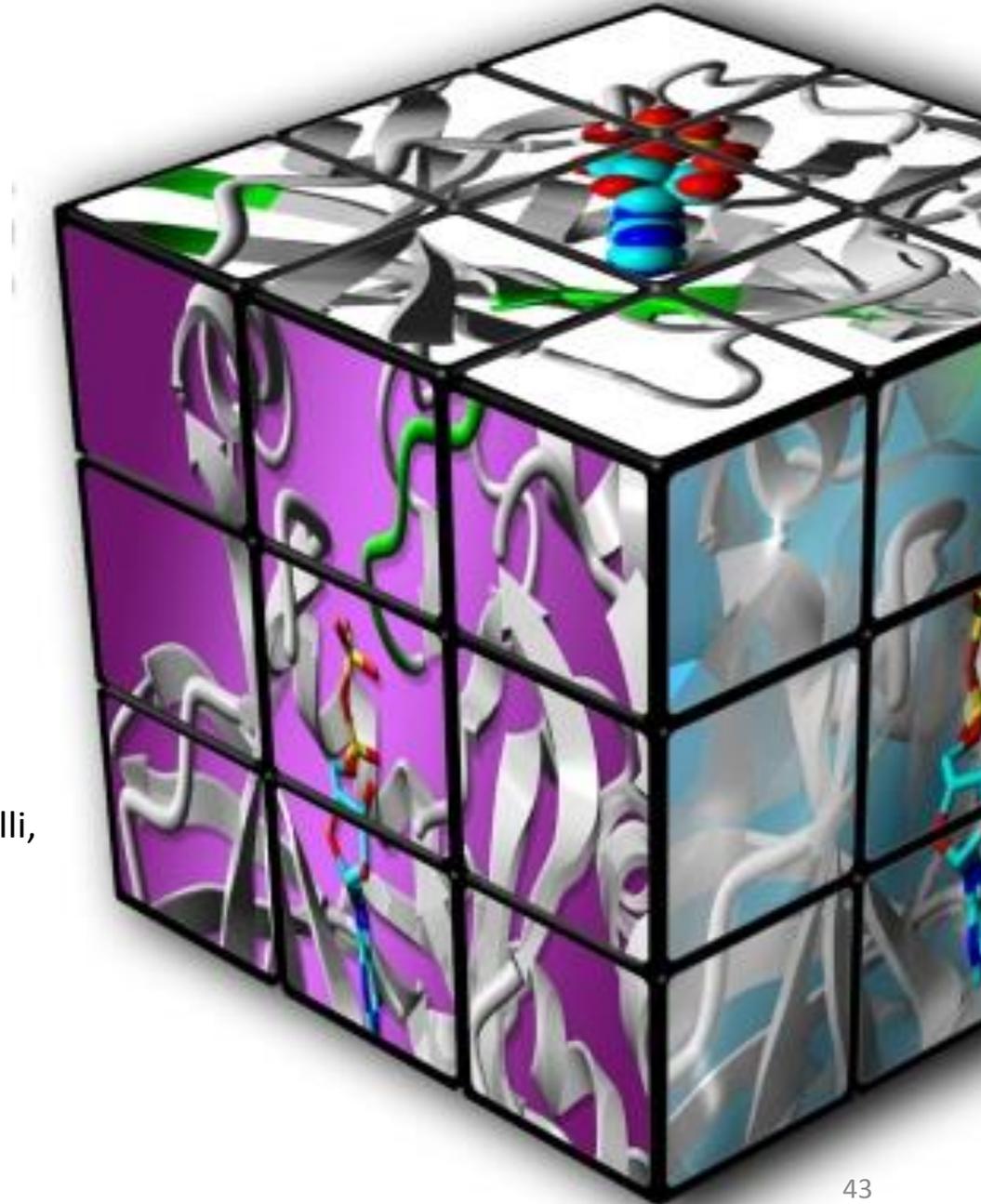
- HOPE can analyze the effect of point mutations on molecular level.
- Use HOPE for those mutations you have identified as possibly harmful, and for which you are missing the last piece of the puzzle to understand structural effects.
- HOPE's performance is comparable to PolyPhen (and probably to other tools that use structural information).
- HOPE is not a classifier! HOPE provides insight and understanding of the structure and effect of the mutations.

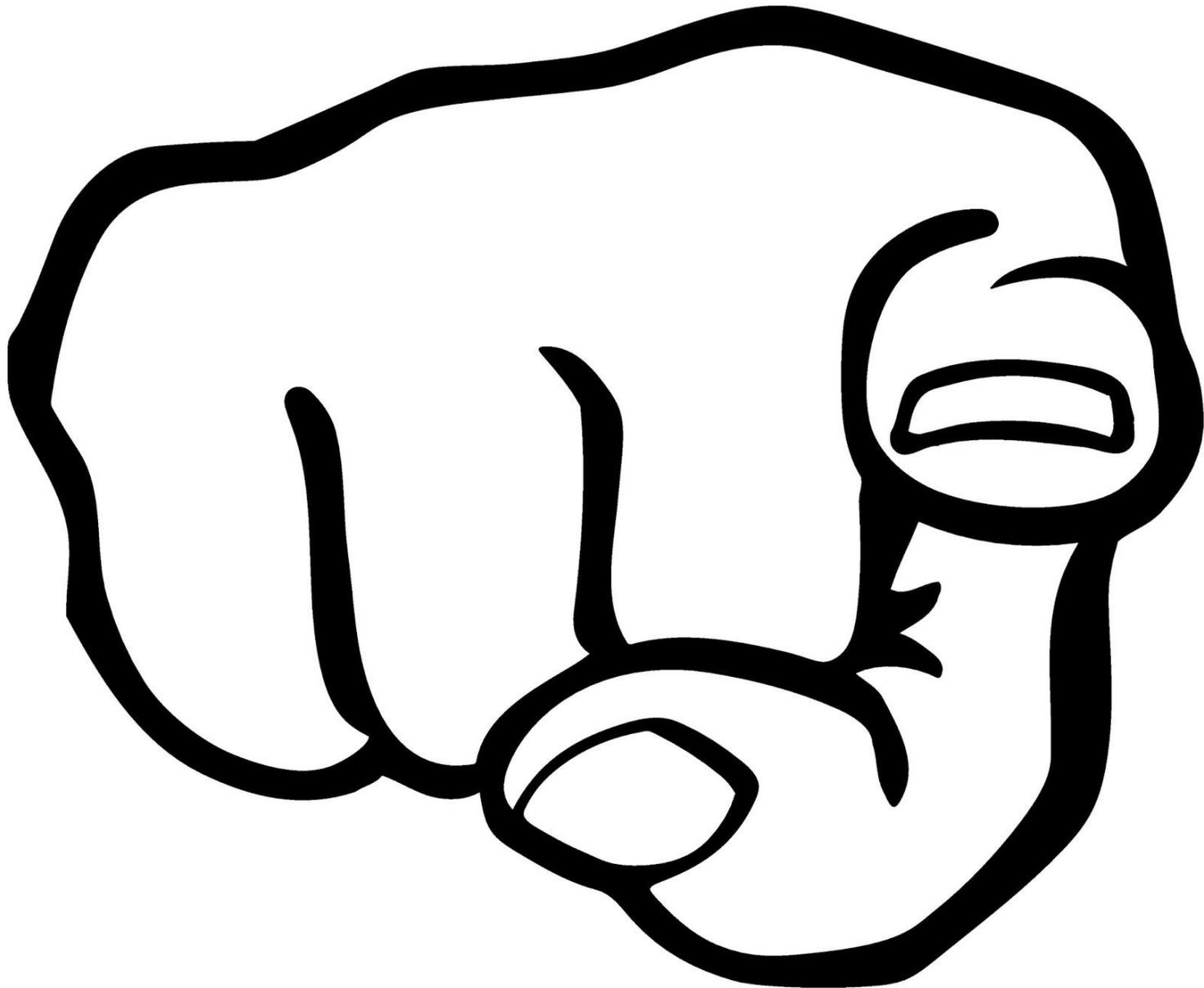


Thanks to...

- Tim te Beek
 - Remko Kuipers
 - Maarten Hekkelman
 - Coos Baakman
 - Elmar Krieger
 - Jules Kersenmakers
 - Jon Black
 - Joanna Lange
-
- Students: Annika Borman, Franscesca Camilli, Shimah Golizadeh, Marlou Snelleman
 - All the HOPE users, every collaborator, and every scientist in the world who added something to the numerous databases available on the web

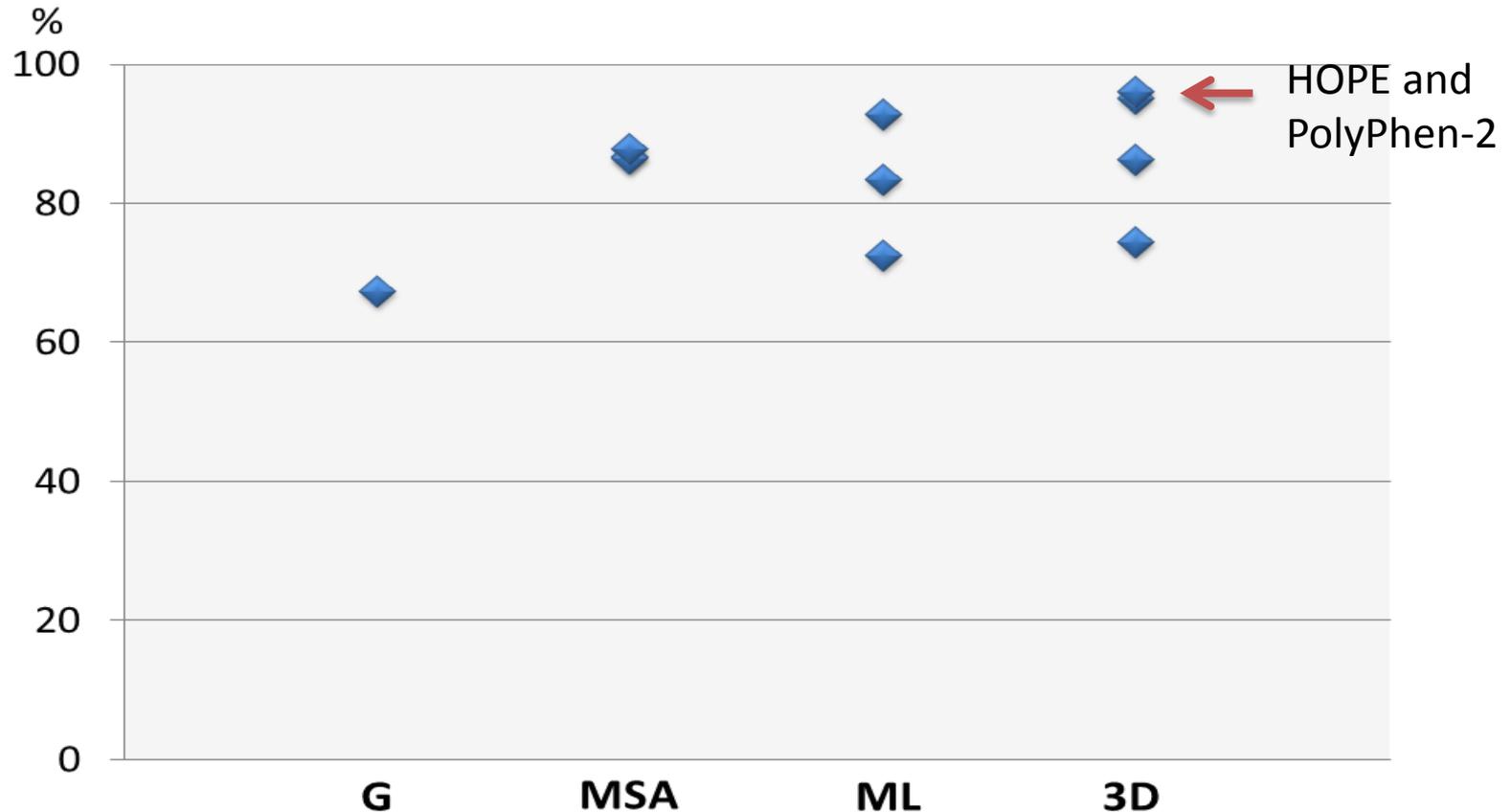
Hanka.Venselaar@radboudumc.nl





Extra slide:

Performance of mutation analysis servers



- Results improve when we use more datasources
- Multiple sequence alignments largely improve Grantham scores
- Machine learning techniques/more data sources can improve the scores
- HOPE/PolyPhen-2 make the best use of 3D-structures - >best scores

Validation of HOPE, an experiment

Extra slide:

- Dataset of >200 mutations (181 damaging/46 benign) collected from literature
- Mutations were analyzed using 10 online servers and HOPE

Method	Damaging mutations	Benign SNPs
Grantham score	67,4%	65,2%
PhD-SNP	85,6%	73,9%
Panther	86,5%	35,1%
SIFT	87,8%	64,4%
SNPs&GO	72,5%	77,8%
SNAP	83,4%	56,5%
MutPred	92,8%	85,7%
nsSNPanalyzer	74,5%	67,6%
SNPs3D	86,3%	62,8%
PolyPhen-2	95,0%	58,6%
HOPE	96,1%	76,1%

G = differences between amino acids

MSA = Conservation scores from multiple sequence alignments, and amino acid differences

ML = Conservation scores and GO-terms or predictions based on sequence, amino acid differences

3D = 3D-Structural features, Swissprot annotations, conservation scores, amino acid differences

HOPE demo: practical work

The following examples will guide you through HOPE's report pages.

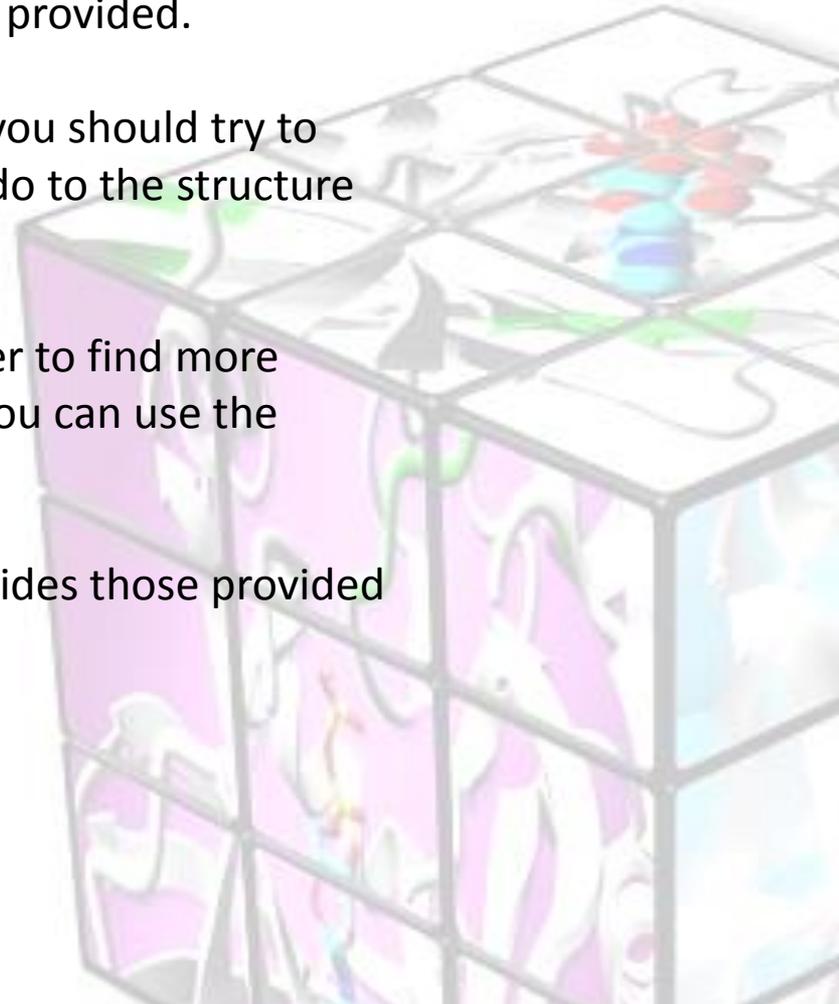
You can simply copy and paste the sequence we provided.

We have added some questions, but in general you should try to answer the question "What does this mutation do to the structure and function of the protein?"

Sometimes you might want to use another server to find more information for the protein of interest, for this you can use the accession codes.

You are free to try other random mutations besides those provided here....

→ If you have questions, let me know!

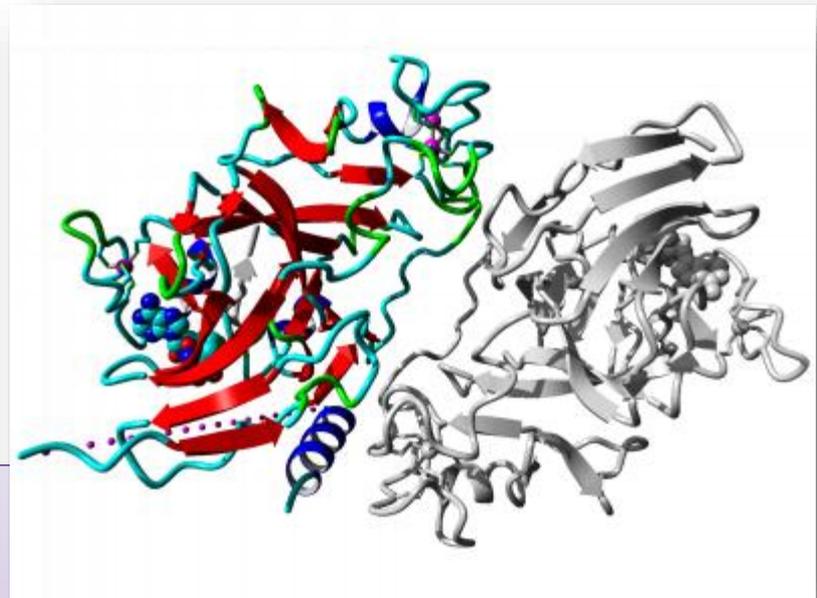


Example 1: Euchromatic Histone Methyl Transferase

Accession codes: EHMT1_HUMAN Q9H9B1

Sequence:

```
MAAADAEAVP ARGEQQDCC VKTELLGEET PMAADEGSAE KQAGEAHMAA DGETNGSCEN
SDASSHANAA KHTQDSARVN PQDGNTLTR IAENGVSERD SEAAKQNHVT ADDFVQTSVI
GSNGYIILNKP ALQAQLRRTT STLASSLPGH AAKTLPGGAG KGRTPSAFPQ TPAAPPATLG
EGSADTEDRK LPAPGADVKV HRARKTMPKS VVGLHAASKD PREVREARDH KEPKEEINKN
ISDFGRQQLL PFPFSLHQSL PQNQCYMATT KSQTAQLPFV LAAAVSRKKK RRMGTYSLVP
KKKTKVLKQR TVIEMFKSIT HSTVGSKGEK DLGASSLHVN GESLEMSDE DDSEEELEEDD
GHGAEQAAAF PTEDSRSTKE SMSEADRAQK MDGESEEEQE SVDTGEEEGE GDESDLSESSE
SIKKKFLKRRK GKTDSPWIKP ARKRRRRSRK KPSGALGSES YKSSAGSAEQ TAPGDSTGYM
EVSLDSLDR VKGILSSQAE GLANGPDVLE TDGLQEVPLC SCRNETPKSR EITTLANNQC
MATESVDHEL GRCTNSVVKY ELMRPSNKAP LLVLCEDHRG RMVKHQCCPG CGYFCTAGNF
MECQPESSIS HRFHKDCASR VNNASYCPHC GEESKAKEV TIAKADTTST VTPVPGQEK
SALEGRADTT TGSAAGPPLS EDDKLQGAAS HVPEGFDPGT PAGLGRPTPG LSQGPQKETL
ESALIALDSE KPKKLRFPK QLYFSARQGE LQKVLLMLVD GIDPNFKMEH QNKRSPLHAA
AEAGHVDICH MLVQAGANID TCSEDQRTPL MEAAENNHLE AVKYLKAGA LVDPKDAEGS
TCLHLAAKKG HYEYVQYLLS NGQMDVNCQD DGGWTPMIWA TEYKHVDLVK LLLSKGSDIN
IRDNEENICL HWAAFSGCVD IAELLAAKC DLHAVNIHGD SPLHIAAREN RYDCVVLFLS
RDSDVTLKNK EGETPLQCAS LNSQVWSALQ MSKALQDSAP DRPSPVERIV SRDIARGYER
IPIPCVNAV DSEPCPSNYKY VSQNCVTSMP NIDRNITHLQ YCVCIDDCSS SNCMCGQLSM
RCWYDKDGR LPEFNMAEPP LIFEENHACS CWRNCRNRV VQNGLRARLQL YRTRDMGWGV
RSLQDIIPGT FVCEYVGE LI SDSEADVREE DSYLFDLDNK DGEVYCIDAR FYGNVSRFIN
HHCEPNLVPV RVFMAHQDLR FPRIAFFSTR LIEAGEQLGF DYGERFWDIK GKLFSCRCGS
PKCRHSSAAL AQRQASAAQE AQEDGLPPTS SAAAADPL
```



Mutations to try:

N1200V (what kind of contacts does this residue make, how are they disturbed, how conserved is the residue?)

C1115Y (what kind of contact does the residue make, can you see this in the pictures?)

Y1018F (what kind of contact does this residue make, what kind of interactions are now missing?)

Mutation N1200V, important notes

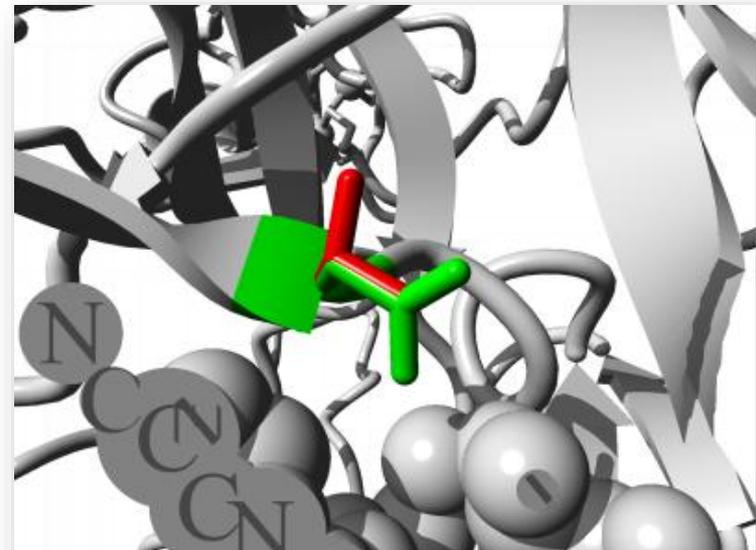
First of all: The mutant residue is smaller and more hydrophobic (cannot make hydrogenbonds)



The wildtype made contacts with a ligand (click on the link), and it made hydrogenbonds (which the mutant cannot make). PISA identifies contacts with another molecule, in this case that is the ligand again.

The wildtype and its neighbor are 100% conserved, so they are important! (for ligand binding, something you can also see in the pictures)

Everything tells you that the residue is located in the ligand binding domain and important for interactions with the molecule. Mutation will cause loss of these interactions and therefore cause loss of function.

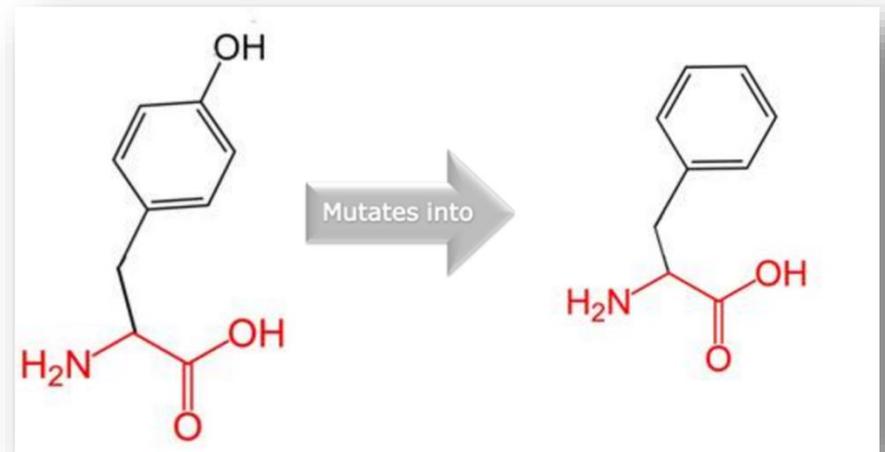


Mutation Y1018F, important notes

The wildtype and mutant residue differ slightly, the mutant is smaller and missing the -OH, so therefore it cannot make the same hydrogenbonds

The wildtype residue is located on the surface, where it makes a hydrogenbond. The pictures, and PISA, indicate that this is a hydrogenbond to the other monomer. Losing this bond might mean loss of dimerisation.

However! Conservation scores indicate that the mutant residue was seen at this position in other structures. This makes our conclusion less clear and you might need experimental backup to underline your conclusions.



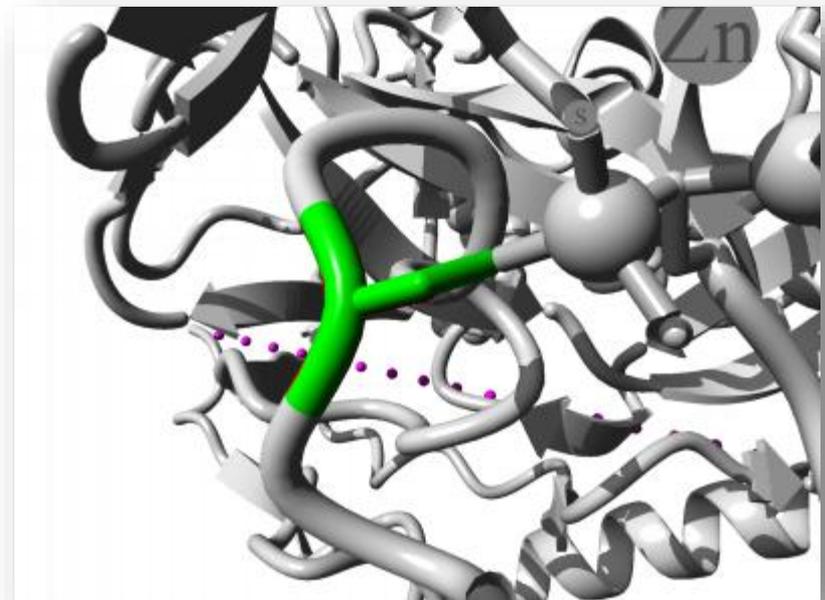
Mutation C1115A, important notes

The wildtype is a S, often involved in disulfide bonds. Mutant A is smaller and cannot make these bonds.



However, the contacts section doesn't mention a disulfid bond, but a metal contact instead. The pictures show the metal-interaction too.

Conservation scores indicate a damaging mutation too... The mutation will cause loss of stability of the SET-domain because A cannot bind metal similar to S.

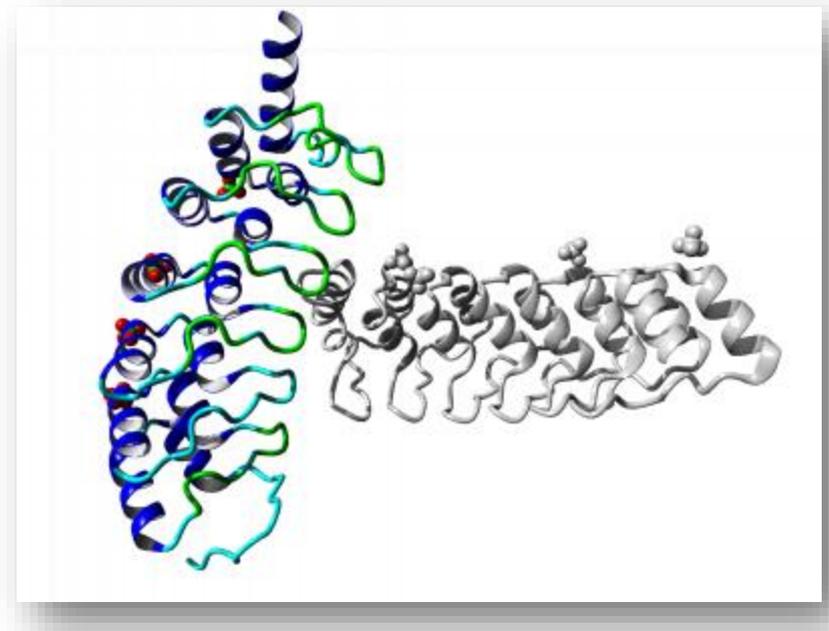


Continued Example 1: Euchromatic Histone Methyl Transferase

Accession codes: EHMT1_HUMAN Q9H9B1

Sequence :

```
MAAADAEAVP ARGEQQDCC VKTELLGEET PMAADEGSAE KQAGEAHMAA DGETNGSCEN
SDASSHANAA KHTQDSARVN PQDGTNTLTR IAENGVSERD SEAAKQNHVT ADDFVQTSVI
GSGYIILNKP ALQAQPLRRT STLASSLPGH AAKTLPGGAG KGRTPSAFPQ TPAAPPATLG
EGSADTEDRK LPAPGADVKV HRARKTMPKS VVGLHAASKD PREVREARDH KEPKEEINKN
ISDFGRQQLL PFPFSLHQSL PQNQCYMATT KSQTAQLPFV LAAAVSRKKK RRMGTYSLVP
KKKTKVLKQR TVIEMFKSIT HSTVGSKGEK DLGASSLHVN GESLEMSDE DDSEEELEDD
GHGAEQAAAF PTEDSRTSKE SMSEADRAQK MDGESEEEQE SVDTGEEEEE GDESDLSSSE
SIKKKFLKRRK GKTDSPWIKP ARKRRRRSRK KPSGALGSES YKSSAGSAEQ TAPGDSTGYM
EVSLDSLDRL VKGILSSQAE GLANGPDVLE TDGLQEVPLC SCRMETPKSR EITTLANNQC
MATESVDHEL GRCTNSVVKY ELMRPSNKAP LLVLCEDHRG RMVKHQCCPG CGYFCTAGNF
MECQPESIS HRFHKDCASR VNNASYCPHC GEESKAKEV TIAKADTTST VTPVPGQEK
SALEGRADTT TGSAAGPPLS EDDKLQGAAS HVPEGFDPTG PAGLGRPTPG LSQGPGETL
ESALIALDSE KPKKLRFPK QLYFSARQGE LQKVLLMLVD GIDPNFKMEH QNKRSPLHAA
AEAGHVDICH MLVQAGANID TCSEDQRTPL MEAAENNHLE AVKYLKAGA LVDPKDAEGS
TCLHLAAKKG HYEYVQYLLS NGQMDVNCQD DGGWTPMIWA TEYKHVDLVK LLLSKGSDIN
IRDNEENICL HWAAFSGCVD IAELLAAKC DLHAVNIHGD SPLHIAAREN RYDCVVLFLS
RDSDVTLKNK EGETPLQCAS LNSQVWSALQ MSKALQDSAP DRPSPVERIV SRDIARGYER
IPIPCVNAV DSEPCPSNYKY VSQNCVTPM NIDRNITHLQ YCVCIDDCSS SNCMCGQLSM
RCWYDKDGR LPEFNMAEPP LIFECHNACS CWRNCRNRV QNGLRARLQL YRTRDMGWGV
RSLQDIPPGT FVCEYVGE LI SDSEADVREE DSYLFDLNDK DGEVYCIDAR FYGNVSRFIN
HHCEPNLVPV RVFMAHQDLR FPRIAFFSTR LIEAGEQLGF DYGERFWDIK GKLFSCRCGS
PKCRHSSAAL AQRQASAAQE AQEDGLPPTS SAAAADPL
```



More Mutations to try:

H770A: For this analysis HOPE will use a different structure than for the previous mutations, why? Also check the conservation scores. What do you think of this?

A43V: Is there a structure for this mutation? Is this a known mutation?

Mutation H770AA, important notes

As you can see HOPE uses PDB file 3B7B for its structural analysis. This was PDB file 3HNA for the previous examples. Why?

Method

The 3D-structure of your protein of interest is known. Information from this 3D-structure will be obtained using WHAT IF Web services, the UniProt database and a series of DAS-servers. The structural information was obtained from the analysis of PDB: [3B7B](#). Annotations about this protein were obtained from UniProt entry [EHMT1_HUMAN](#). See the [method page](#) for more information.

To figure this out you can perform a BLAST search against the PDB database to see which parts of the EHMT1 protein are experimentally solved.

Go to: <http://mrs.cmbi.ru.nl/blast>

The screenshot shows the NCBI BLAST search interface. The search parameters are: Search for `gnl|sprot|EHMT1_HUMAN` in the `PDB` database. The search results table shows 22 hits, with the first hit being `gnl|sprot|EHMT1_HUMAN` in the `pdb` database, with 250 hits found. Red arrows point to the sequence input field, the PDB database dropdown, the Run Blast button, and the search results table.

Home Blast Align Status

Search All Databanks for

Enter one protein sequence in *FastA* format

```
>gnl|sprot|EHMT1_HUMAN
MAAADAEAVPARGEPQQDCCVKTELLGEEETPMAADEGSAEKQAGEAHMAADGETNGSCENSASSHANAANKH
TQDSARVNPQDGTNLTTRIAENGVSERDSEAAKQNHVTADDFVQTSVIGSNGYILNKPALQAQPLRTTSTLA
SSLPGHAAKTLPGGAGKGRTPSAFPQTPAAPATLGEGSADTEDRKLPAAGADVKVHRARKTMPKSVVGLHA
ASKDPREVREARDHKEPKEEINKNISDFGRQQLPPFPPLHQSLPQNCYMATTKSQTAQLPFVLAADVSRK
```

Databank to search: PDB

Filter low complexity:

E-value cutoff: 10.0

advanced options

Run Blast

Blast results

Nr	ID	Databank	Status
22	gnl sprot EHMT1_HUMAN	pdb	250 hits found

Paste the EHMT1 sequence here
Choose the PDB database here
Click here to run BLAST
Click to see your results

Your list with BLAST hits will appear. These are all PDB files of which the sequence shows a % identity with your sequence of interest. Somewhere around no 6, you will find a hit for 3HNA, click on the grey/coloured bar to see this:

mrs.cmbi.ru.nl/blast

5 3HNA.A  crystal structure of catalytic domain of human euchromatic histone methyltransferase 1 in complex with sah and mono-methylated h3k9 peptide (HYDROLASE/HYDROLASE REGULATOR 30-MA) 1 611

6 3HNA.B  crystal structure of catalytic domain of human euchromatic histone methyltransferase 1 in complex with sah and mono-methylated h3k9 peptide (HYDROLASE/HYDROLASE REGULATOR 30-MA) 1 611

Hsp nr	Alignment	Score	Bitscore	E-value	Length	Identity	Similarity
1		1576	611	5.59e-174	284	284 (100%)	284 (100%)
Q:	982 NSQVWSALQMSKALQDSAPDRPSPVERIVSRDIARGYERIPICVNAV DSEPCPSNYKYV		1041				
S:	3 NSQVWSALQMSKALQDSAPDRPSPVERIVSRDIARGYERIPICVNAV DSEPCPSNYKYV		62				
Q:	1042 SQNCV TSPMNIDRNITHLQYCV CIDD C S S S N C M C G Q L S M R C W Y D K D G R L L P E F N M A E P P L		1101				
S:	63 SQNCV TSPMNIDRNITHLQYCV CIDD C S S S N C M C G Q L S M R C W Y D K D G R L L P E F N M A E P P L		122				
Q:	1102 IFECNHACSCWRNCRNRVQNGLRARLQLYRTRDMGWVRS LQDIPPGTFVCEYV GELIS		1161				
S:	123 IFECNHACSCWRNCRNRVQNGLRARLQLYRTRDMGWVRS LQDIPPGTFVCEYV GELIS		182				
Q:	1162 DSEADVREEDSYLFDLDNKDGEVYCIDARFYGNVSRFINHHCEPNLVPVRVFM AHQDLRF		1221				
S:	183 DSEADVREEDSYLFDLDNKDGEVYCIDARFYGNVSRFINHHCEPNLVPVRVFM AHQDLRF		242				
Q:	1222 PRIAFFSTR LIEAGEQLGFDYGERFWDIKGLFSCRCGSPKCRH	1265					
S:	243 PRIAFFSTR LIEAGEQLGFDYGERFWDIKGLFSCRCGSPKCRH	286					

7 3MO0.A  human g9a-like (glp, also known as ehmt1) in complex with inhibitor e11 (TRANSFERASE 22-APR-10 3MO0); histone-lysine n-methyltransferase, h3 ly 1 611

← Q = Sequence of interest (EHMT1)
 ← S = Sequence of the structure found in PDB

You will see the alignment of the matching part of your sequence of interest with the structure found in the PDB database.

Do you understand the result? Can you see that the structure found contains residue 982-1265 of your sequence of interest?

Find BLAST hit 3B7B (somewhere around 25). Do you understand which part of the structure can be found in this PDB file?

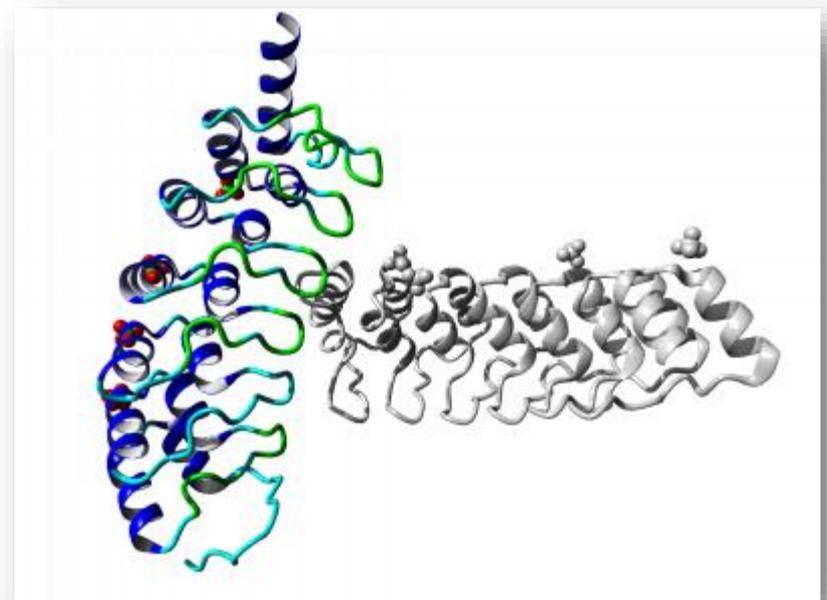
Mutation H770A, important notes

So, HOPE uses a different PDB file for structural analysis of mutation H770, simply because 3HNA does not contain this position, and 3B7B does.

The amino acid properties differ again between H and A. But have a look at the conservation score. What does the conservation tell you?

Does the wildtype have any special contacts? Or special modifications? Follow the domain link to read more about “Ankyrin repeat containing domains”. Can you find a special function for the domain?

It doesn't look like this mutation does much damage to the protein, you will need more experimental results.



Mutation A43V, important notes

As you can see HOPE cannot find a source of structural information. You can use the BLAST results from the previous example to see if there is indeed no structure known for this part of the protein.

It doesn't look like there is anything known about the function of the wildtype residue. And the two residue types only differ in size.



However, the mutation was found before and is known in the ExPASy database

Variants

This mutation matches a previously described variant, with the following description: **In a breast cancer sample; somatic mutation.**

See the ExPASy site about this variant: [VAR_036345](#).

The severity of this variant is unknown.

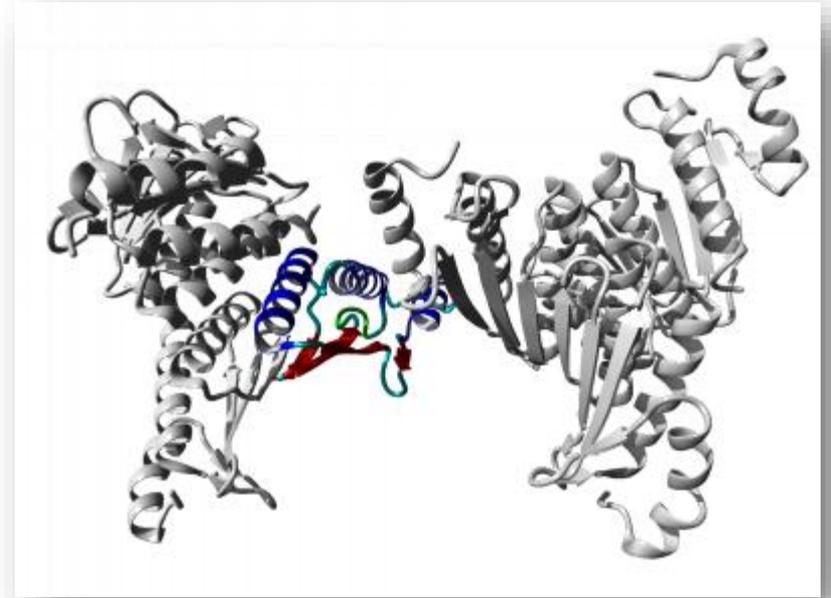
You can follow the link to read more about this variant, would you classify this as a damaging mutation?

Example 2: Phenylalanine hydroxylase-stimulating protein

Accession code: PHS_HUMAN P61457

Sequence :

```
MAGKAHRLSA EERDQLLPNL RAVGWNELEG  
RDAIFKQFHF KDFNRAFGFM TRVALQAEKL  
DHHPEWFNVY NKVHITLSTH ECAGLSERDI  
NLASFIEQVA VSMT
```



Mutations to try:

H63S: Look at the interactions with other molecules?

N44P: What happens when a proline is introduced in a helix?

Mutation H63S, important notes

H and S are very different residues and conservation is high.

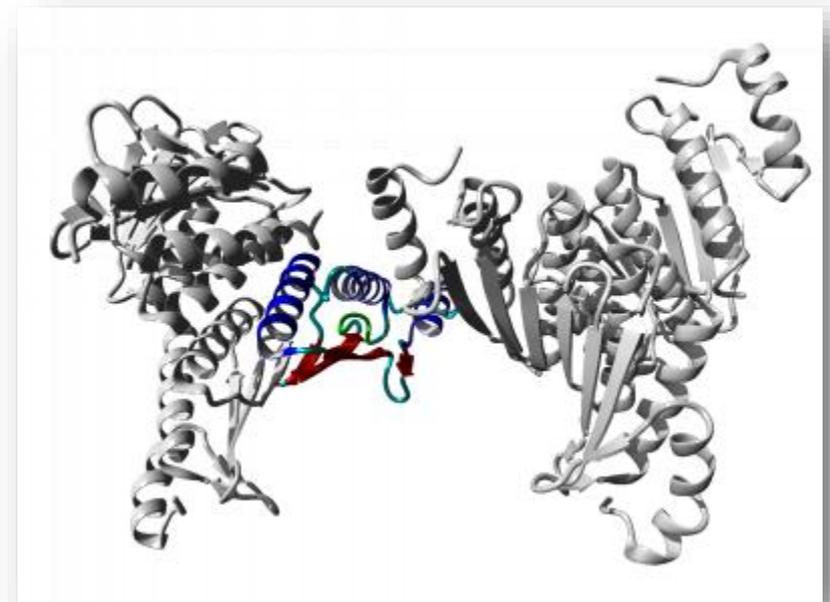
Read the information about contacts indicated by PISA and look at the pictures. Do you agree that this residue is in contact with other molecules?

Does the domain information agree with the fact that this protein needs to function in a complex?

Look at the moving animations, how many mutations can you see? Why? In what kind of complex does this protein function?

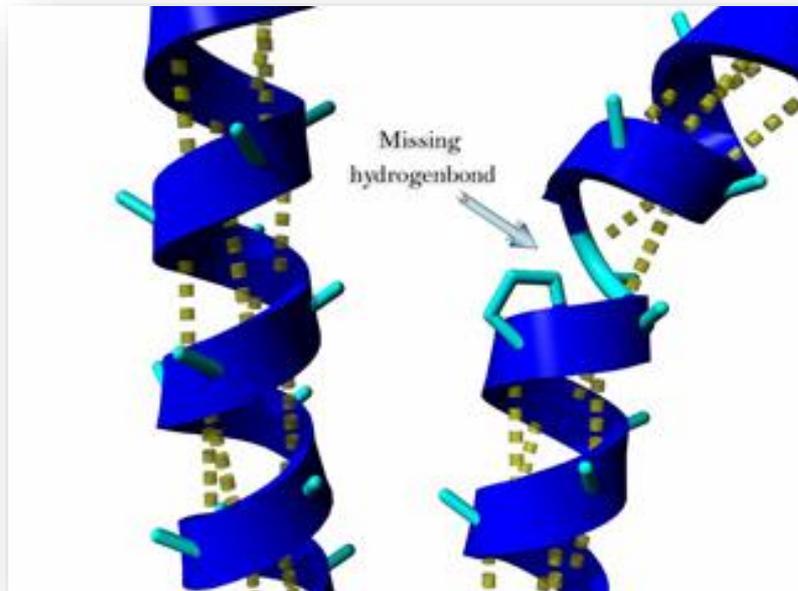
If you want you can read more information about the protein here:

http://mrs.cmbi.ru.nl/entry?db=sprot&nr=300243&q=phs_human



Mutation N44P, important notes

N and P have a few differences in amino acid properties. You can see that due to size and hydrophobicity changes a few contacts will be lost.



More important for this protein is the fact that a proline will be introduced in a helix structure, this will cause a kink and destabilization of the structure. The pictures show that the helix is important for interactions with the other subunits in the complex.

Example 3: Transcription factor 7-like 2

Accession code: TF7L2_HUMAN Q9NQB0

Sequence:

```
MPQLNGGGGDDLGANDELISFKDEGEQEEKSSENSAERDLADVKSSLVNESETNQNSSS  
DSEAERRPPRSESEFRDKSRESLEEAQRQDGGFLFKGPPYPGYPFIMIPDLTSPYLPNGS  
LSPTARTLHFQSGSTHYSAYKTIEHQIAVQYLQMKWPLLDVQAGSLQSRQALKDARSPSP  
AHIVSNKVPVVQHPHHVHPLTPLITYSNEHFTPGNPPPHLPADVDPKTGIPRPPHPPDIS  
PYYPLSPGTVGQIPHPLGWLVPQQGQPVYPITTTGGFRHPYPTALTVNASMSRFPPhMVPP  
HHTLHTTGIPHPAIVTPTVKQESSQSDVGSLSHSSKHQDSKKEEEKKKPHIKKPLNAFMLY  
MKEMRAKVVAECTLKESAAINQILGRRWHALSREEQAKYYELARKERQLHMQLYPGWSAR  
DNYGKKKKRKRDKQPGETNEHSECFLNPCLSLPPITDLSAPKKCRARFGLDQQNNWCGPC  
RRKKKCVRYIQGEGSCLSPSSDGSLLDSPPSPNLLGSPPRDAKSQTEQTQPLSLSLKP  
DPLAHL SMMPPPPALLLAEATHKASALCPNGALDLPPAALQPAAPSSSIAQPSTSSLHSH  
SSLAGTQPQPLSLVTKSLE
```

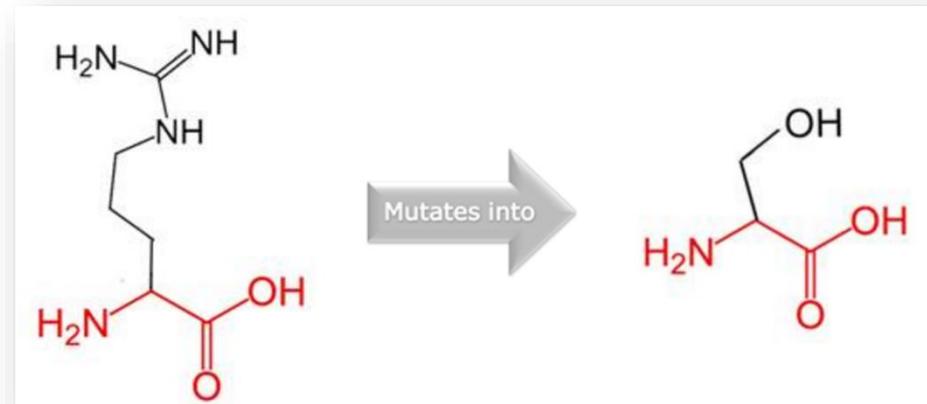
Mutations to try:

R387S: For what function is the wildtype residue required?

T212S: What function does the residue have and will it really be lost by this mutation?

Mutation R387S, important notes

The mutant residue S is much smaller and neutrally charged when compared to wildtype R. This causes a potential loss of many interactions.



The contact paragraph tells us that a residue close to the mutated one had an important function in DNA binding. HOPE will look at important functions of neighboring residues.

Contacts

The mutated residue did not interact with DNA, however, one of its neighbor residues was found to interact with DNA. This interaction might be affected by the mutation.

Many stabilizing interactions will be lost. If the structure destabilizes locally, the residues important for DNA binding will also be affected. Remember the function/name of the protein?

Mutation T212S, important notes

These two residues show very minor differences.



Look at the variant and modification paragraphs. The wildtype residue is phosphorylated. During mutagenesis experiments was found that mutation of this residue to a V (so removal of phosphorylation) has an enhancing effect. Can you understand why?

Variants

Mutagenesis experiments have been performed on this position. Mutation of the wild-type residue into **V** has the following effect: **Reduced phosphorylation by NLK and enhanced DNA-binding; when associated with V-201.**

This is an example of a mutation that causes the removal of regulation. Sometimes that results in a constitutively active protein.

However, since S can also be phosphorylated the effect of this mutation might be less severe.

Your own examples.....

Now it's time to try your own examples....

1: Find the amino acid sequence of your protein of interest. And submit the sequence in HOPE

Hint: use MRS or Uniprot to search for your protein:
<http://mrs.cmbi.ru.nl/> or <http://www.uniprot.org/>

2: Choose a mutation (if you don't know an interesting mutation, you can try to find an interesting residue in MRS or Uniprot and see whether HOPE finds the same info). Submit this mutation to HOPE.

3: Wait for HOPE to finish the report and read it. See if you understand the results. Write down any questions you have so we can discuss these results.

You can also send any questions to Hanka.Venselaar@radboudumc.nl