Data Mining And Computational Analysis Of Human Growth Hormone Gene (gh1) Sequence In Normal Population to Identify Potential Variants With Disease Causing Effects.

Sonia Verma and Amit V Pandey

Pediatric Endocrinology, University Children's Hospital, Bern, and Department for BioMedical Research, University of Bern, Bern Switzerland.

Introduction

Mutations in GH1 gene cause isolated growth hormone deficiency. Several disease-causing mutations from patients with IGHD have been reported. These mutations have been shown to: (a) produce shorter isoforms of GH that does not bind to growth hormone receptor, (b) cause diminished secretion of GH or (c) result in misfolded or truncated GH protein. A large amount of genomics data from sequencing studies from non-clinical population is now available which shows several hundred genetic variations in GH1 gene. Role of common polymorphic variants in GH1 in relation to effects GH protein has not been systematically studied.

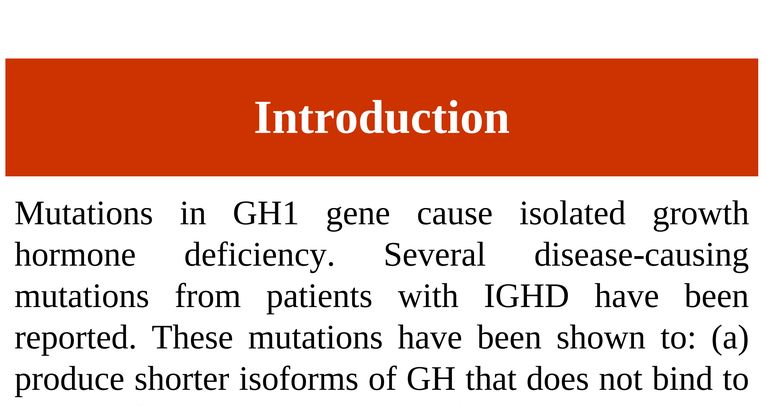

Figure 1. An overview of the secreted human GH1 (N-26) amino acid sequence. Location of some common mutants and polymorphisms and structural features in GH are shown. From: *Pandey AV. Endocrine Dev. 2012 23:71-85.*

Objective

Searching the genomics data to find and analyse the effects of potentially disease-causing variants in GH1 gene by using computational simulations.

Methods

The human GH protein sequence (NP_000506.2) was used as a query sequence to search the homologous sequences in the Uniref90 database by performing a Psi-blast search. The selected GH sequences were aligned. The MSA was used as input in the Consurf analysis, which calculates evolutionary (functional and structural) conserved residues in the protein based on the phylogenetic and structural similarity between sequences from different species. The conservation score for each residue position in MSA represents the conservation of the residue.

The human GH crystal structure solved with its receptor (PDB Id: 3HHR, dimeric receptor) has few missing residues in the loop regions (Chain A: 148-154; Chain B & C: 57-62, 73-78), those were modeled by using loop modeling. For each loop region ten loop conformations were generated and out of them an energy minimized conformation was selected for the loop replacement. The structure models were visualized with Pymol (www.pymol.org) and rendered as ray-traced images with POVRAY (www.povray.org)

Results

Clinical Significance	Protein Residue	AA Pos	Reference Protein Residue	ConSurf Conservation	Residue Contact Receptor 1	Residue Contact Receptor 2	UniProt Disease Variant
	Gln [Q]	28	Pro [P]	e		Yes (2)	
	Thr [T]	30	Ile [I]	b		Yes (4)	
	Lys [K]	34	Arg [R]	e		Yes (8)	
	His [H]	38	Asn [N]	e		Yes (12)	
	Phe [F]	41	Leu [L]	b		Yes (15)	
	His [H]	42	Arg [R]	e, f		Yes (16)	Yes (R>C)
	Thr [T]	43	Ala [A]	b, s			
	Arg [R]	44	His [H]	e	Yes (18)		
	Tyr [Y]	47	His [H]	e, f	Yes (21)		
	Thr [T]	50	Ala [A]	b, s			
B	Tyr [Y]	51	Phe [F]	e	Yes (25)		
	Pro [P]	71	Leu [L]	e	Yes (45)		
	Lys [K]	73	Asn [N]	e			Yes (N>D)
	Thr [T]	74	Pro [P]	e	Yes (48)		
P	Ser [S]	79	Cys [C]	b, s			
	Asp [D]	82	Glu [E]	e			
	Cys [C]	88	Ser [S]	e	Yes (62)		
	Lys [K]	89	Asn [N]	e	Yes (63)		
	His [H]	103	Arg [R]	e			Yes (R>C)
P	Cys [C]	103					
	Cys [C]	105	Ser [S]	b, s			
	Glu [E]	110	Gln [Q]	e, f			
	Arg [R]	117	Gln [Q]	e			Yes (Q>L)
P	Gly [G]	138	Asp [D]	e			Yes (D>G)
	Glu [E]	142	Asp [D]	e		Yes (116)	
	Asp [D]	145	Glu [E]	e		Yes (119)	
	Ser [S]	146	Gly [G]	b		Yes (120)	
	Met [M]	149	Thr [T]	b		Yes (123)	
	Pro [P]	188	Leu [L]	b, s			
	Glu [E]	195	Asp [D]	e, f			
	His [H]	190	Tyr [Y]	b	Yes (164)		
	Asn [N]	198	Lys [K]	e, f	Yes (172)		
	Lys [K]	200	Glu [E]	e	Yes (174)		
	Met [M]	205	Ile [I]	b	Yes (179)		
	Arg [R]	208	Cys [C]	b, s	Yes (182)		
P	His [H]	209	Arg [R]	e			Yes (R>H)
	Tyr [Y]	215	Cys [C]	e, f	Yes (189)		
	Ser [S]	216	Gly [G]	e	Yes (190)		

Table 1. Summary of potentially disease causing variants in GH1.

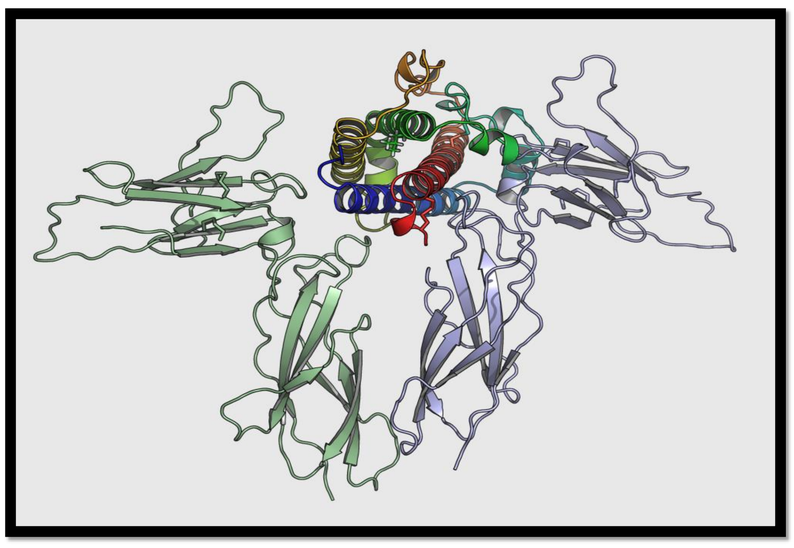

Figure 2. Complex of GH1 with its receptor.

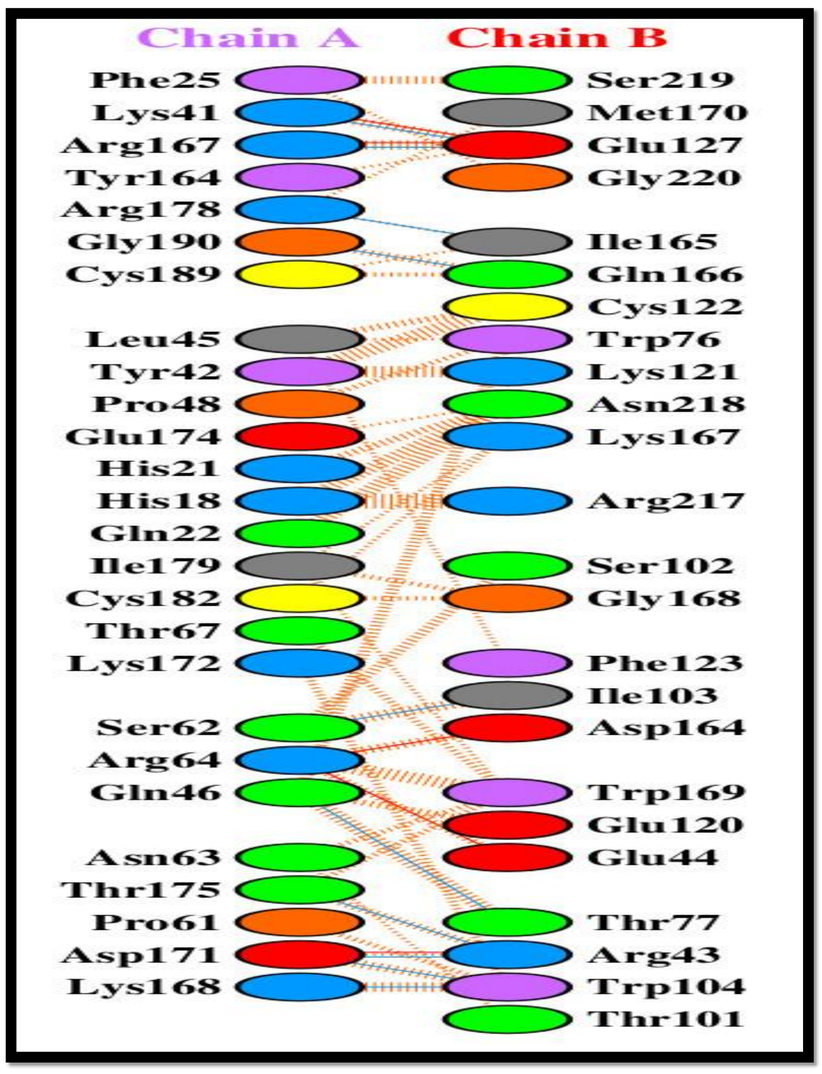

Figure 3. Contact points of GH1 with its receptor.

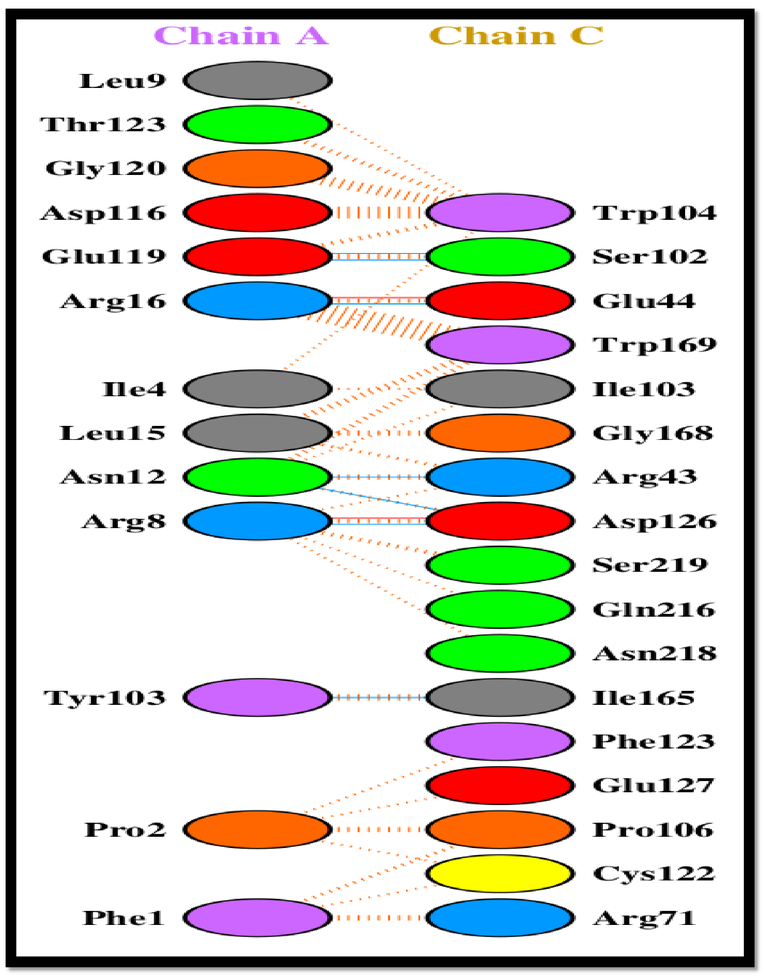

Figure 4. Contact of GH1 with its receptor.

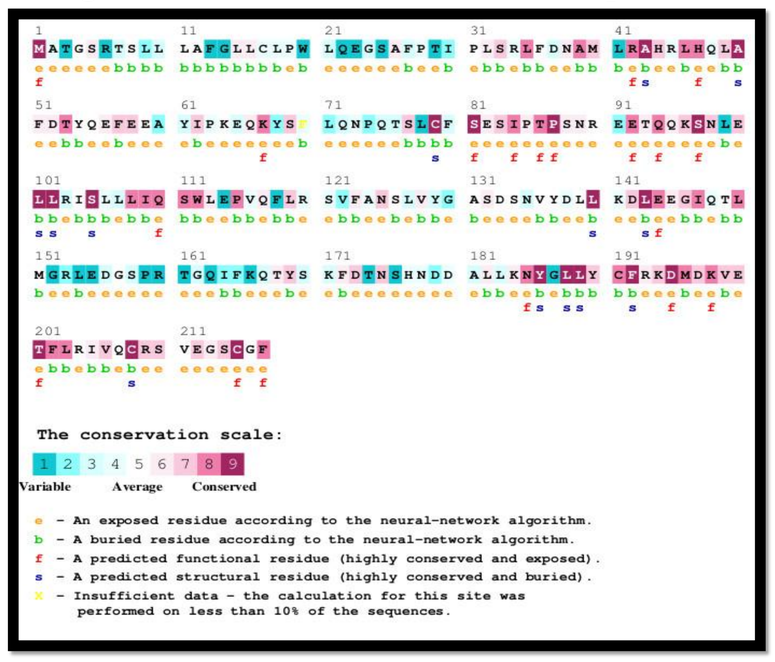

Figure 5. Evolutionary sequence conservation of amino acids in GH1 based on analysis of different GH proteins available in genome databases.

Conclusions

Identification of potentially negative effects from some variations in GH1 gene from non-clinical populations can be utilized to study links to growth variations and distribution of certain common variants in GH1 gene. Identification of potentially disease-causing variants in GH1 will help in further functional characterization of these variants when these are later found in patients and linked to growth hormone deficiency.

1. *Pandey AV. Bioinformatics tools and databases for the study of human growth hormone. Endocrine Dev. 2012 23:71-85.*

Email: amit.pandey@dbmr.unibe.ch

Acknowledgements

This research is supported by unrestricted research grants from Novo Nordisk AG, Switzerland and Sandoz Pharmaceuticals AG, Switzerland.

Disclosure Statement: The Authors have no potential conflict of interest.