UNIVERSITÀ DI SIENA 1240

A bioinformatic tool for the investigation of HGD mutations responsible for Alkaptonuria

Vittoria Cicaloni[1,2] Andrea Bernini[1], Annalisa Santucci[1], Ottavia Spiga[1]

[1]Dept. of Biotechnology, Chemistry and Pharmacy, University of Siena

[2]Toscana Life Sciences, Siena

Introduction

Alkaptonuria (AKU) [OMIM 203500] is the first described inborn error of metabolism (Garrod 1908) characterised by deficiency of homogentisate dioxygenase (HGD) involved in metabolism of tyrosine (La Du et al. 1958). In AKU, the metabolic block causes accumulation of homogentisic acid (HGA) that polymerizes, forming a dark brown ochronotic pigment that is deposited in connective tissue, mainly in joints cartilage, causing a severe form of arthropathy. For now, no cure exists for the disorder. So far, mutation analysis has been performed in about 400 AKU patients showing that the missense mutations are the most numerous (68%) (Zatkova et al. 2011). The crystal structure of the HGD protein is organised as a highly complex and dynamic hexamer comprising two disc-like trimers (Titus et al. 2000). An intricate network of non-covalent interactions is required to maintain the spatial structure of the protomer, the trimer and finally, the hexamer. This delicate structure can be easily disrupted by mutations leading to aberrant effects on enzyme function, which is confirmed by high proportion of missense mutations identified in AKU patients.

We showed that the missense mutations are predicted to affect the activity of the enzyme by three mechanisms: decrease of stability of individual protomers, disruption of protomer–protomer interactions or modification of residues in the region of the active site (Nemethova et al. 2016). In present work we focus on the structural positioning of the mutations in the dynamic model of HGD in its environment.

Results

We have developed a tool for the evaluation of all the missense mutations in HGD protein, showing that the loss of enzymatic activity is due to modifications of structurally critical and evolutionary conserved residues. With the analysis of the structure, missense mutations have been divided into 4 groups according to their position: active site, core, surface and interface between monomers (Fig 1).

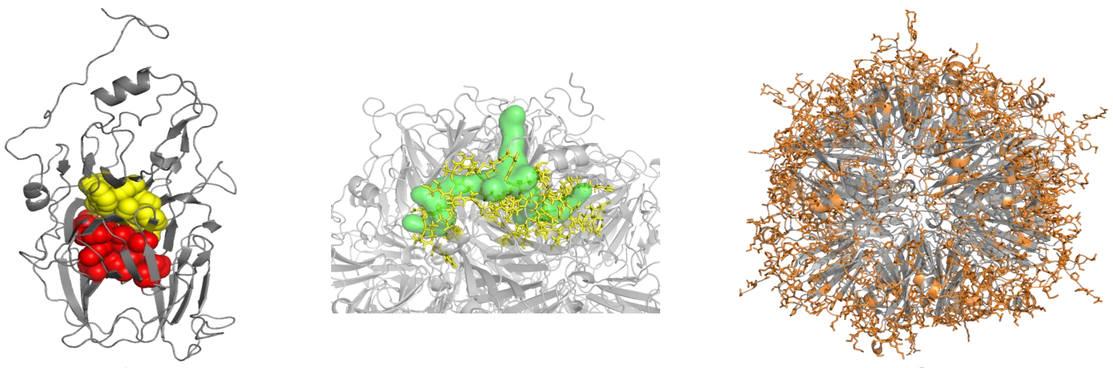

Figure 1: (A) Representation of the 'core1' (yellow) located in the C-terminal domain and the 'core2' (red) located in the N-terminal domain (Chain A) of a HGD monomer; (B) Graphic representation of the pore defined by the pore volume of the cavity (green) and residue (yellow) delimiting the cavity of the active site. (C) Residues (orange) of the HGD hexamer located at the surface.

With the analysis of the sequence, Shannon entropy has been calculated on a multiple sequence alignment in order to highlight conserved positions, where missense mutations have a high probability to be deleterious. The evaluation of variants effects on the HGD structure-function was extended to a 'dynamic' simulation of HGD in water which provides a detailed assessment of dynamic and structural properties of the HGD protein. Thanks to the study of occupancy, we have found that the large part of the residues placed on monomer interface, are affected by missense mutations. These findings highlight the relevance of non-covalent interactions in the modulation of protein stability and in the quaternary structure folding.

Seq no	Seq res	Mutazioni missense	H+	Interfacce res	Superficie res	Core protomero	Sito attivo
292	H	R	•	•	•	·	•
321	R	P	•	•	•	·	•
33	Q	R	•	•	•	·	·
225	R	H;L;P	•	•	•	·	·
322	W	R	•	•	•	·	·
330	R	S	•	·	•	·	•
374	D	H	•	·	•	·	•
13	E	K	•	•	·	·	·
25	L	P	•	•	•	·	·
40	Y	S	•	•	·	·	·
42	E	A	•	•	·	·	·
44	L	F	•	•	·	·	·
47	S	L	•	•	·	·	·
53	R	W;Q	•	•	·	·	·
60	W	G	•	•	·	·	·
73	F	L	•	•	·	·	·
92	P	T	•	•	·	·	·
136	F	Y	•	·	·	•	·
137	L	P	•	·	•	·	·
147	F	S	•	·	·	•	·
159	Q	H	•	·	•	·	·
167	T	I	•	·	·	•	·
168	E	K;D	•	•	·	·	·
171	K	N	•	·	•	·	·
181	V	F	•	·	·	•	·
197	R	G	•	·	•	·	·
216	I	T	•	•	·	·	·
217	G	W	•	•	·	·	·
219	N	S	•	•	·	·	·
269	H	R	•	·	·	·	·
274	P	L	•	·	·	•	·
276	K	N	•	·	•	·	·
291	K	D	•	•	·	·	·
305	S	F	•	•	·	·	·
309	G	V	•	•	·	·	·
329	F	C	•	•	·	·	·
337	N	D	•	•	·	·	·
353	K	Q	•	·	·	·	•
359	P	L	•	·	•	·	·
368	M	V	•	•	·	·	·
371	H	R	•	·	·	·	•
373	P	L	•	·	·	·	•
401	E	Q	•	·	·	·	•
18	D	N	•	·	·	·	·
57	K	N	•	·	·	·	·
61	L	P	•	·	·	·	·
62	Y	C	•	·	·	·	·
97	W	R;G	•	·	·	·	·
115	G	R	•	·	·	·	·
116	L	P	•	·	·	·	·
120	C	F;W	•	·	·	·	·
122	A	D;V	•	·	·	·	·
123	G	R;A	•	·	·	·	·
149	N	K	•	·	·	·	·
152	G	A	•	·	·	·	·
153	D	G	•	·	·	·	·
158	P	R;L	•	·	·	·	·
161	G	R	•	·	·	·	·
172	M	T	•	·	·	·	·
178	E	D;G	•	·	·	·	·
183	Q	R	•	·	·	·	·
186	M	K	•	·	·	·	·
187	R	G	•	·	·	·	·
189	S	I	•	·	·	·	·
198	G	D	•	·	·	·	·
205	G	D	•	·	·	·	·
227	F	S	•	·	·	·	·

Figure 2: Seq no: number of residue in the HGD sequence. Seq res: residue corresponding to the sequence number. A large part of missense mutations: they are listed in descending order. H+: in this group are all the mutations are justified by the analysis of entropy.

Methods

Homo sapiens HGD (HGDHs) sequence in FASTA format was obtained from UniProt and was used to search homologous proteins using BLAST (Altschul et al. 1990). CLUSTALXv2.1 (Larkin et al. 2007) was used for a multiple alignment of obtained sequences. For each position, Shannon Entropy values were computed by Bio3D (Grant et al. 2006), a package of R software v3.0.3 (Team 2008). Shannon entropy values were normalized between 0 (corresponding to high conservation) and 1 (high variability). The missing segments in protein structure were reconstructed by homology model through Swiss-Pdb Viewer v4.1.0 (Guex and Peitsch 1997) using Pseudomonas putida HGD (HGDPp) structure. The protein model was energetically minimized with GROMACS 5.0.2 (Berendsen et al. 1995) which is also used for molecular dynamic simulation, for the calculation of the Root mean square deviation (RMSD), the Root mean square fluctuation (RMSF) and the occupancy of hydrogen bonds (fraction of time that the hydrogen bond is formed, between 0 and 1).

Conclusion

Using a bioinformatics approach, we obtained an in-depth comprehension of the HGD mutation-function and mutation-structure relationships: featuring every missense mutation on the basis of structural and entropic information we collected as many data as possible from both structural and functional point of view in order to develop a suitable tool for the prediction of the variants effects responsible for AKU.