Issues in Science and Technology Librarianship

Electronic Resources Reviews

Navigating the National Center for Biotechnology Information's Databases on the Medicinal Chemistry of Homocystinuria

Sarah H. Jeong, MLIS
Research & Instruction Librarian-Science
Wake Forest University
jeongsh@wfu.edu

Introduction

The National Center for Biotechnology Information's (NCBI) databases are authoritative, current information sources intended for researchers, faculty, graduate students, information professionals, and the public for finding the genetic, protein, and structural molecular biological data (NCBI Resource Coordinators 2016). NCBI Gene, Nucleotide, Protein, and Structure databases are considered the four core linked, annotated genetic and protein sequence information sources curated by NCBI scientists based on the scientists' raw data deposited into GenBank, which became publicly available in 1982 (Choudhuri 2014). MedGen is an authoritative information portal for inherited human diseases and was launched in 2012 by NCBI (Louden 2020). MedGen uses standardized terminology from "NLM's Unified Medical Language system (UMLS®), the NIH Genetic Testing Registry (GTR®), and ClinVar" (Halavi et al. 2018). OMIM (Online Mendelian Inheritance in Man) is a curated database for finding the genotype and phenotype of inherited human diseases, which was created in 1985 through collaboration between the National Library of Medicine and the William H. Welch Medical Library of Johns Hopkins University and developed by NCBI in 1995 (About OMIM 2021). PubChem, launched in 2004 by NCBI, is a linked data repository of standardized chemical compounds and substances with provenance of chemical structures (Hähnke et al. 2018). The Bioassay section of PubChem has been a legacy tool since November 1, 2018 (About PubChem 2021). In this article, the reader will understand how to find an allele associated with the phenotype of an inherited human disease, the biomolecular pathway causing the physical manifestation of this disease, and finally, the treatment in the management of the condition.

NCBI Database Search Strategy of the Biochemical Pathway and Medicinal Chemistry of Homocystinuria

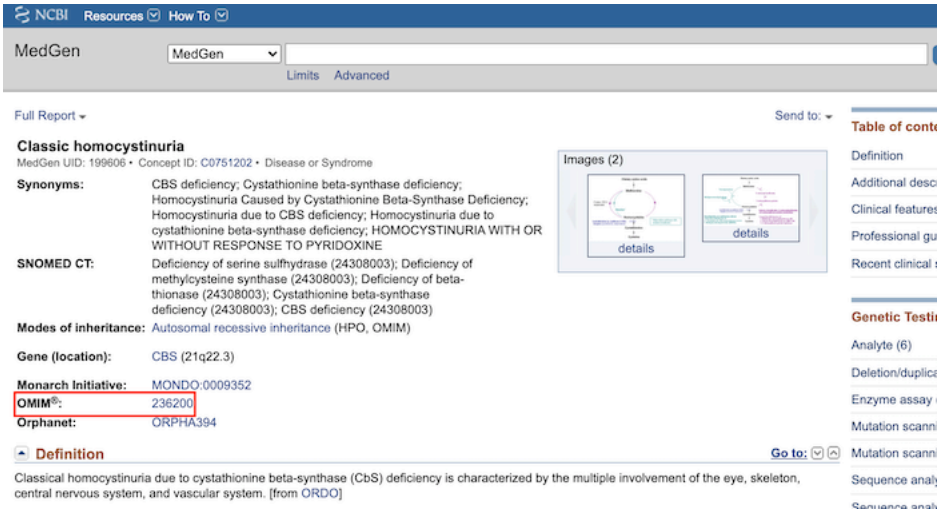


Figure 1. Entry for Homocystinuria in the NCBI MedGen database.

Start by searching for Homocystinura in the NCBI MedGen database, and follow the link to the OMIM database by clicking on 236200 https://www.ncbi.nlm.nih.gov/medgen/199606 (Figure 1).

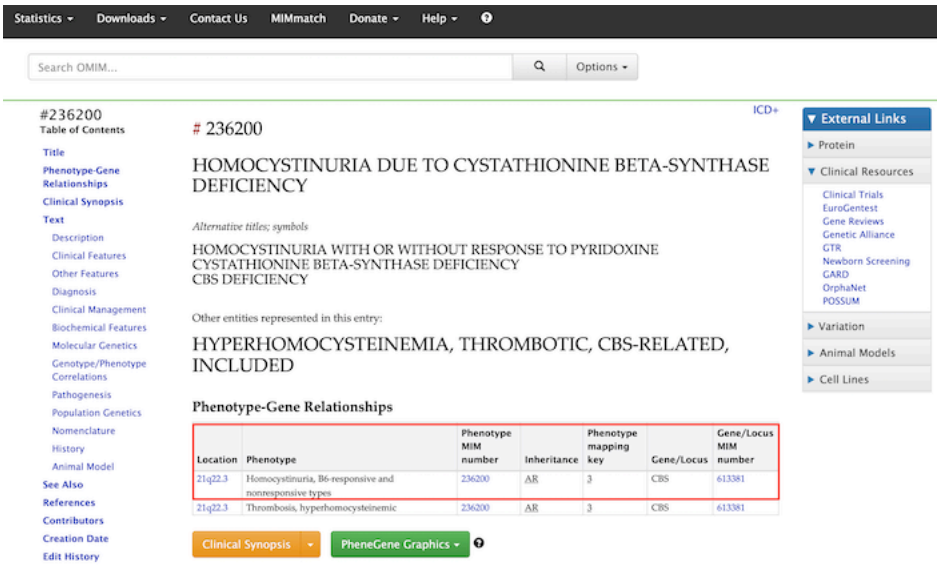


Figure 2. OMIM phenotype of Homocystinuria.

From OMIM #236200 Homocystinuria https://omim.org/entry/236200 (Figure 2), follow the link to the Gene/Locus MIM #613381.

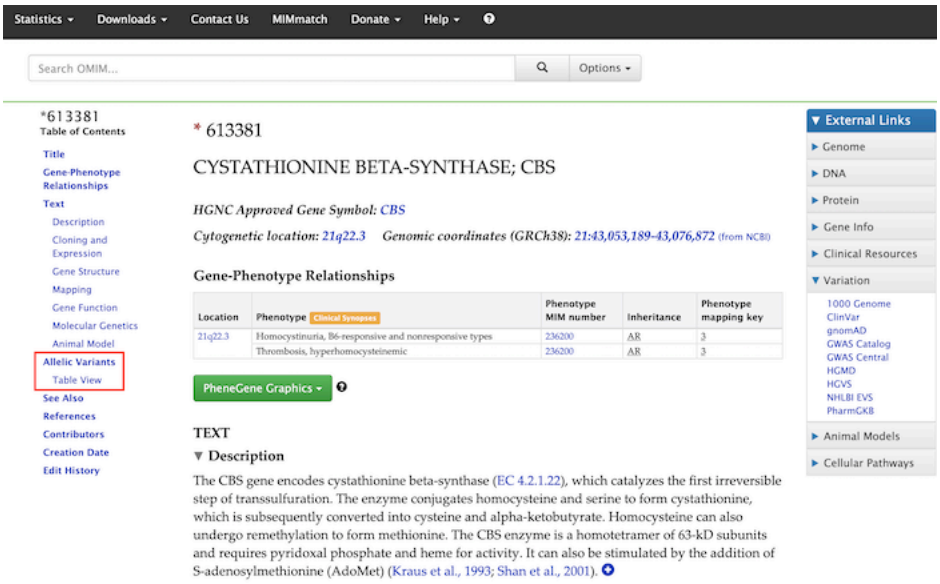


Figure 3. Cystathionine Beta-Synthase (CBS) gene.

The Cystathionine Beta-Synthase (CBS) gene https://omim.org/entry/613381 (Figure 3) encodes for a key enzyme in metabolism, and its deficiency causes Homocystinuria.

From OMIM database (https://omim.org/entry/613381) follow the link to see Allelic Variants. This takes you to: https://omim.org/entry/613381#allelicVariants

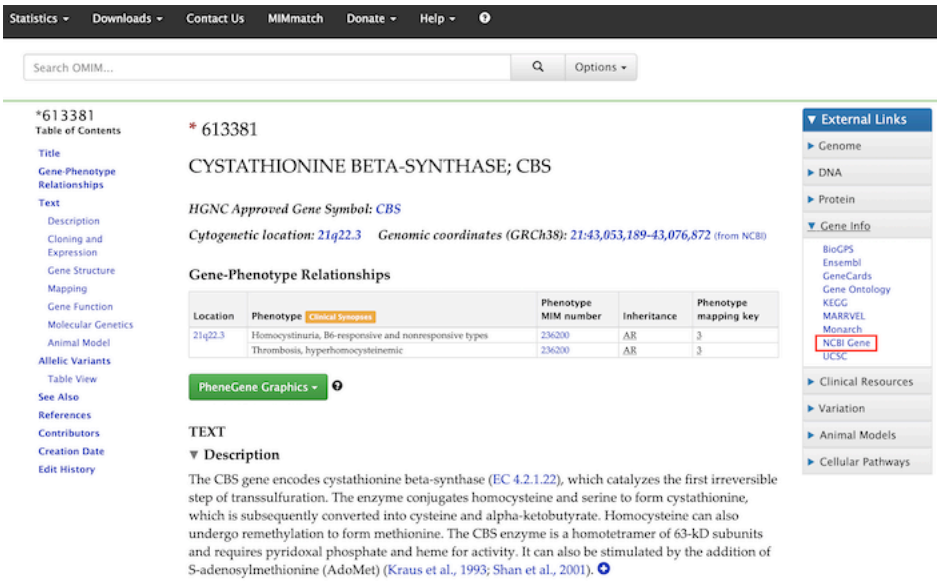


Figure 4. OMIM #613381 Cystathionine Beta-Synthase gene.

From the OMIM entry for the Cystathionine Beta-Synthase gene https://omim.org/entry/613381 (Figure 4), follow the link to the NCBI Gene database to get the canonical protein.

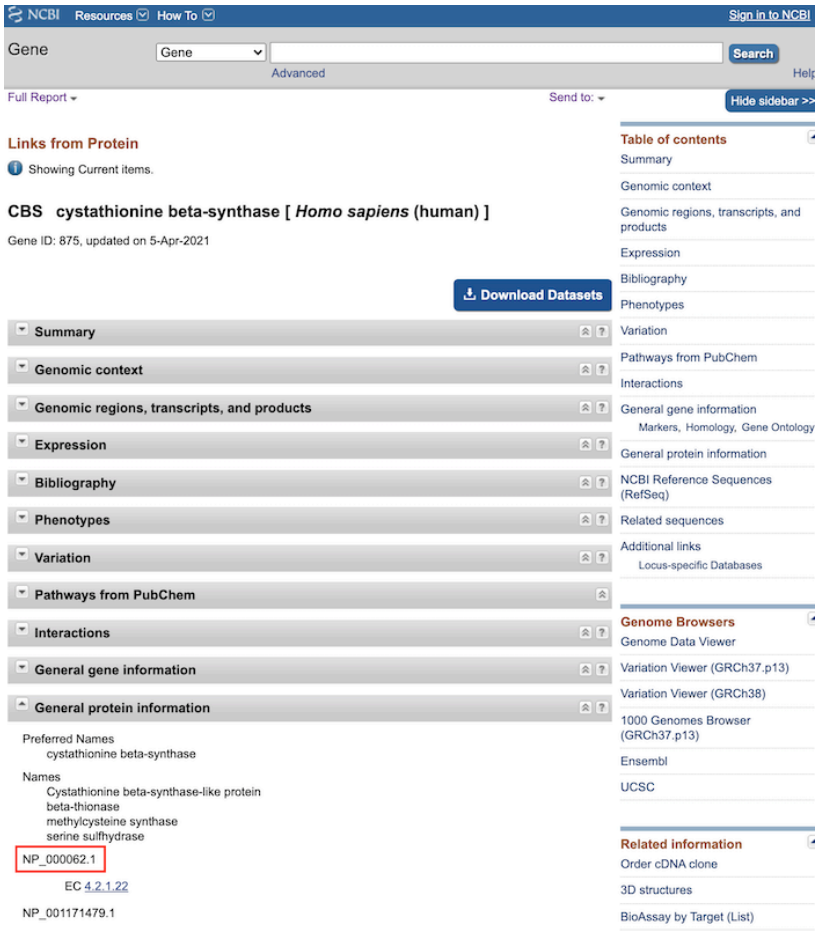


Figure 5. Entry for human cystathionine beta-synthase gene in the NCBI Gene database.

From the NCBI Gene record of the human cystathionine beta-synthase gene https://www.ncbi.nlm.nih.gov/gene/875 (Figure 5), look up the canonical protein (NP_000062.1) associated with Homocystinuria in the NCBI Protein database https://www.ncbi.nlm.nih.gov/protein/NP_000062.1.

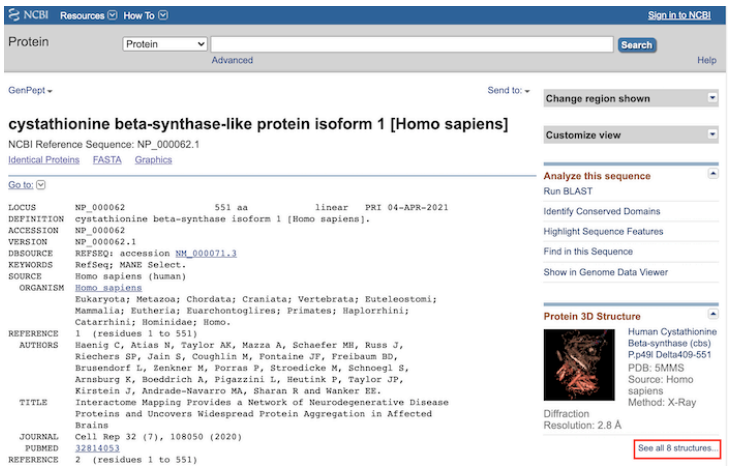


Figure 6. Entry for the canonical protein associated with Homocystinuria in the NCBI Protein database.

You can link to 8 protein 3-D structures of cystathionine beta-synthase-like protein isoform 1 from NCBI Protein database by following the link to "See all 8 structures…" https://www.ncbi.nlm.nih.gov/protein/NP_000062.1 (Figure 6).

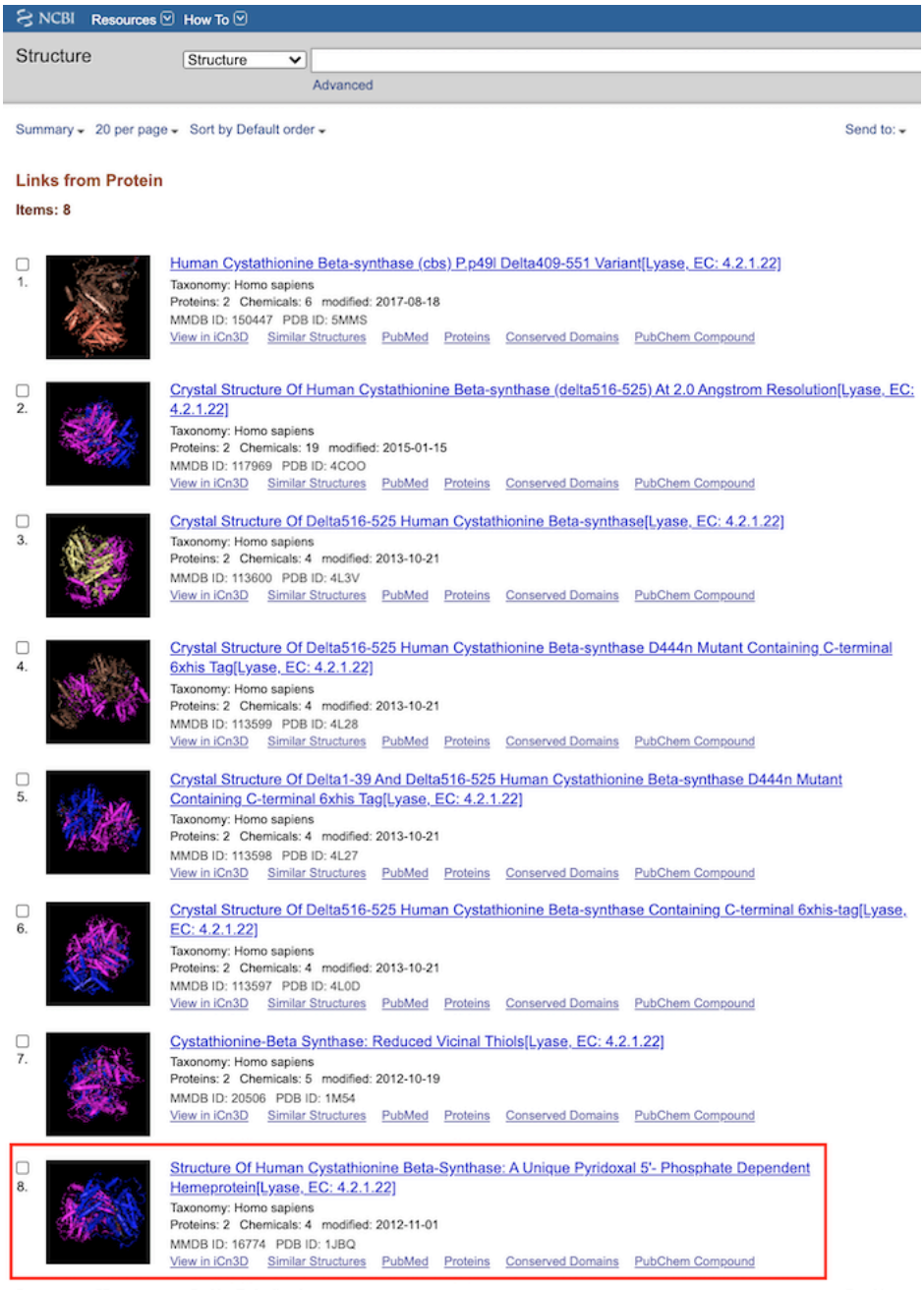


Figure 7. 3-D structures of cystathionine beta-synthase-like protein isoform 1 in the NCBI Structure database.

In NCBI Structure database, protein structure #8 https://www.ncbi.nlm.nih.gov/structure?Db=structure&DbFrom=protein&Cmd=Link&LinkName=protein_structure&LinkReadableName=Structure&IdsFromResult=4557415 (Figure 7) seems to be the wild type protein structure of human Cystathionine Beta-Synthase.

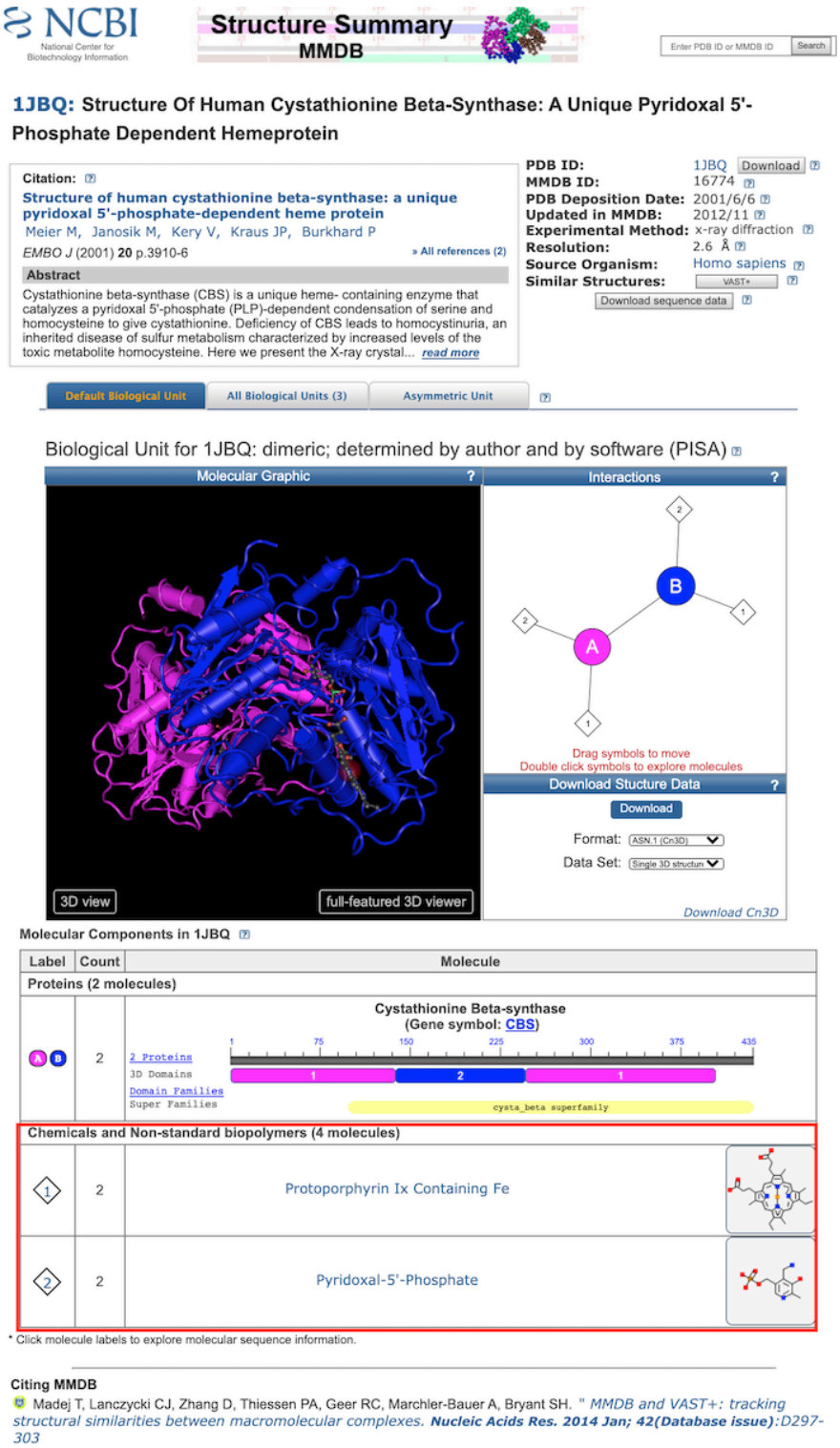


Figure 8. Entry for 3-D structure of the wild type, canonical protein PDB ID: 1JBQ Human Cystathionine Beta-Synthase in the NCBI Structure database.

The NCBI Structure entry for PDB ID: 1JBQ Human Cystathionine Beta-Synthase https://www.ncbi.nlm.nih.gov/Structure/pdb/1JBQ (Figure 8) has links to PubChem entries for the co-factor pyridoxal 5' phosphate https://pubchem.ncbi.nlm.nih.gov/substance/152137797 and the heme https://pubchem.ncbi.nlm.nih.gov/substance/823350, both of which are required for

Cystathionine Beta-Synthase enzyme activity. On the 3-D conformer part, under 'Interactive Chemical Structure Model' of the PubChem record, click the Space-filling radio button to get a rendering of the molecule.

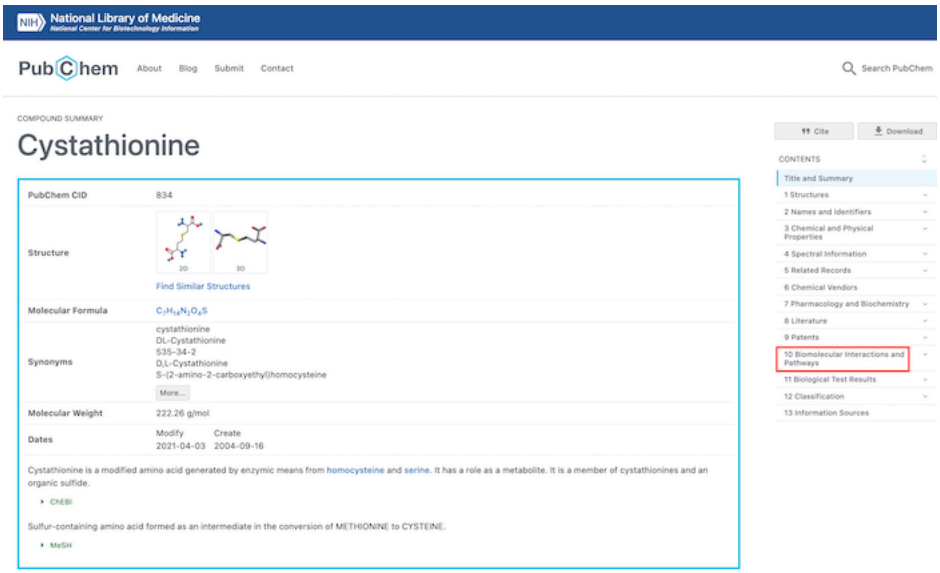


Figure 9. Entry for Cystathionine compound in the NCBI PubChem database.

You can also search for all the relevant substrates and products and follow the link to Biomolecular Interactions and Pathways of Cystathionine (Figure 9) in the NCBI PubChem database https://pubchem.ncbi.nlm.nih.gov/compound/834.

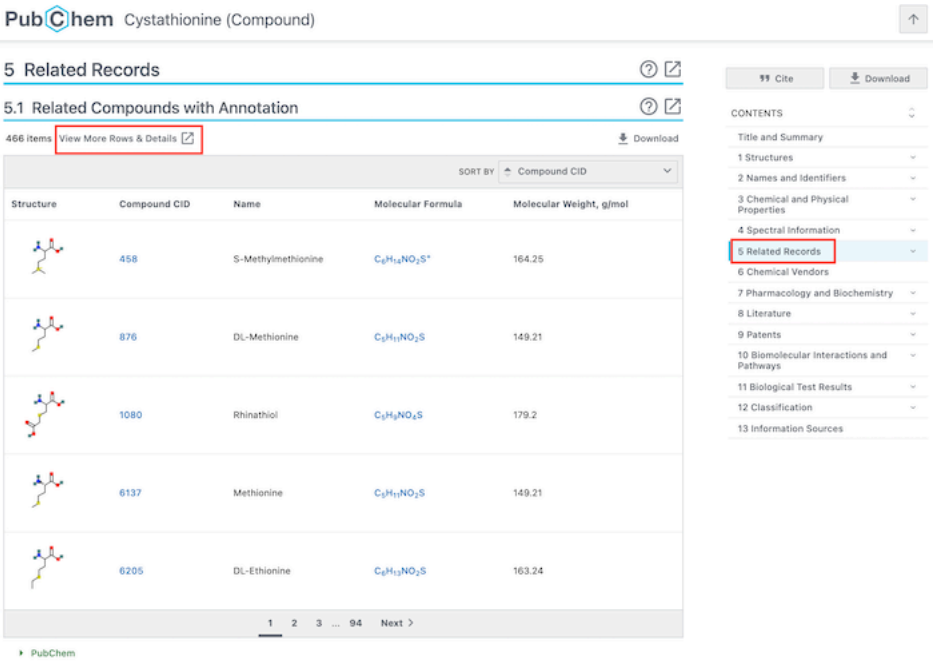
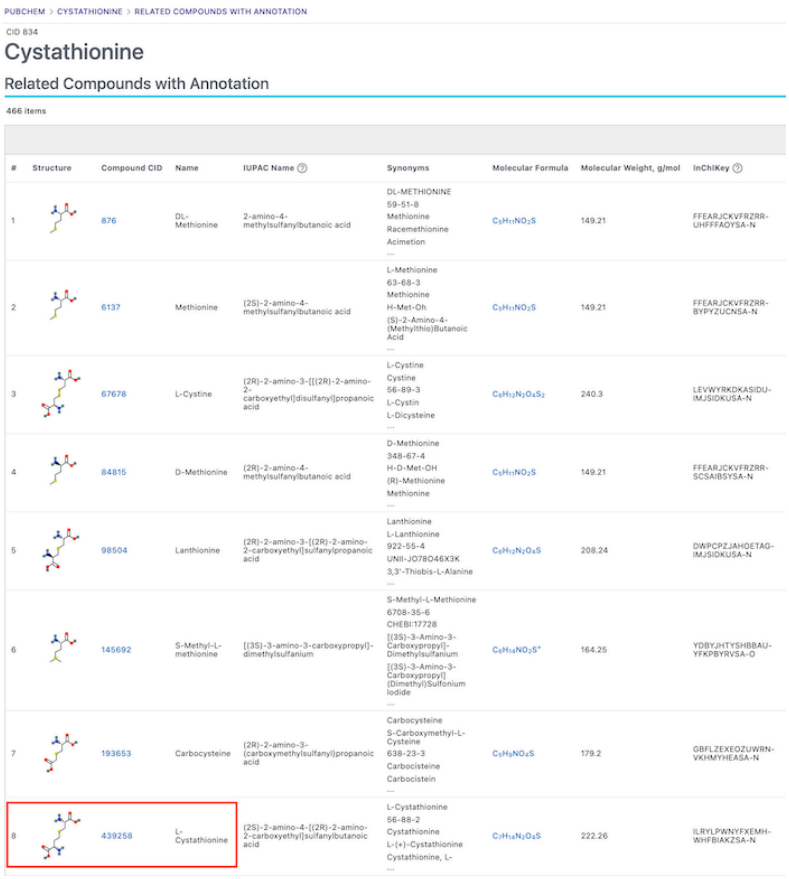


Figure 10. Related Compounds for Cystathionine in the NCBI PubChem database.

In PubChem Section 5.1 Related Compounds with Annotations for Cystathionine https://pubchem.ncbi.nlm.nih.gov/compound/834#section=Related-Compounds-with-

Annotation&fullscreen=true (Figure 10), click on “View More Rows & Details,” then sort by Create Date and follow the link to L-Cystathionine to find the biomolecular pathway that causes Homocystinuria.



PUBCHEM > CYSTATHIONINE > RELATED COMPOUNDS WITH ANNOTATION

CID 834

Cystathionine

Related Compounds with Annotation

466 items

#	Structure	Compound CID	Name	IUPAC Name	Synonyms	Molecular Formula	Molecular Weight, g/mol	InChIKey
1		876	DL-Methionine	2-amino-4-methylsulfanylbutanoic acid	DL-METHIONINE 59-51-8 Methionine Racemethionine Acimetion ...	$C_5H_{11}NO_2S$	149.21	FFEARJCKVFRZRR-UHFFFAOYSA-N
2		6137	Methionine	(2S)-2-amino-4-methylsulfanylbutanoic acid	L-Methionine 63-68-3 Methionine H-Met-Oh (S)-2-Amino-4-(Methylthio)Butanoic Acid ...	$C_5H_{11}NO_2S$	149.21	FFEARJCKVFRZRR-BYPYZUCNSA-N
3		67678	L-Cystine	(2R)-2-amino-3-[[(2R)-2-amino-2-carboxyethyl]disulfanyl]propanoic acid	L-Cystine Cystine 56-89-3 L-Cystin L-Dicysteine ...	$C_6H_{12}N_2O_4S_2$	240.3	LEVWYRKDKASIDU-IMJSIDKUSA-N
4		84815	D-Methionine	(2R)-2-amino-4-methylsulfanylbutanoic acid	D-Methionine 348-67-4 H-D-Met-OH (R)-Methionine Methionine ...	$C_5H_{11}NO_2S$	149.21	FFEARJCKVFRZRR-SCSAIBSYSA-N
5		98504	Lanthionine	(2R)-2-amino-3-[(2R)-2-amino-2-carboxyethyl]sulfanylpropanoic acid	Lanthionine L-Lanthionine 922-55-4 UNII-JO78O46X3K 3,3'-Thiobis-L-Alanine ...	$C_6H_{12}N_2O_4S$	208.24	DWPCPZJAHQETAG-IMJSIDKUSA-N
6		145692	S-Methyl-L-methionine	[(3S)-3-amino-3-carboxypropyl]-dimethylsulfanium	S-Methyl-L-Methionine 6708-35-6 CHEBI:17728 [(3S)-3-Amino-3-Carboxypropyl]-Dimethylsulfanium [(3S)-3-Amino-3-Carboxypropyl]-(Dimethyl)Sulfonium Iodide ...	$C_6H_{14}NO_2S^+$	164.25	YDBYJHTYSHBBAU-YFKPBYRVSA-O
7		193653	Carbocysteine	(2R)-2-amino-3-(carboxymethylsulfanyl)propanoic acid	Carbocysteine S-Carboxymethyl-L-Cysteine 638-23-3 Carbocisteine Carbocistein ...	$C_5H_9NO_4S$	179.2	GBFLZEXEOZUWRN-VKHMYHEASA-N
8		439258	L-Cystathionine	(2S)-2-amino-4-[(2R)-2-amino-2-carboxyethyl]sulfanylbutanoic acid	L-Cystathionine 56-88-2 Cystathionine L-(+)-Cystathionine Cystathionine, L- ...	$C_7H_{14}N_2O_4S$	222.26	ILRYLPWNYFXEMH-WHFBIAKZSA-N

Figure 11. Related compounds of Cystathionine in the NCBI PubChem database.

In NCBI PubChem database, sort the list of related compounds of Cystathionine by “Create Date” and follow the link to L-Cystathionine (Figure 11).

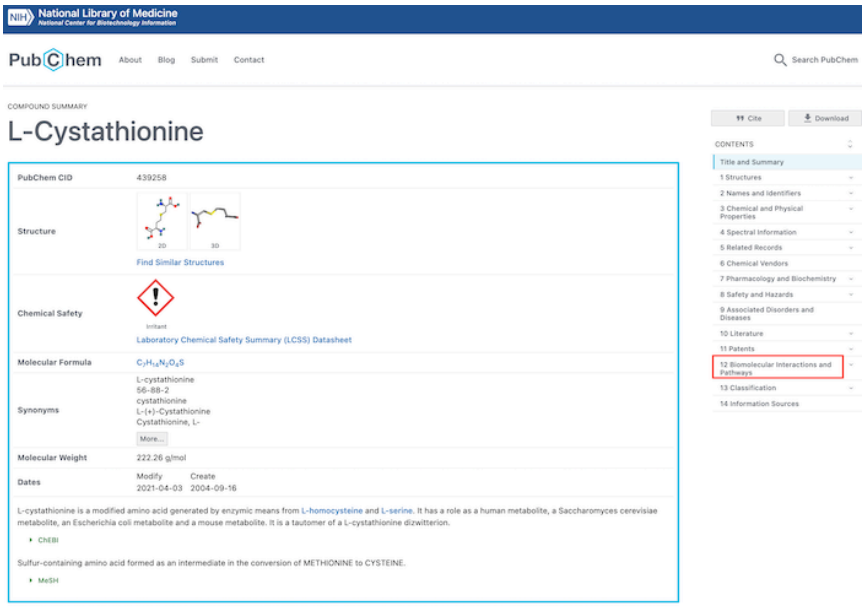


Figure 12. Compound Summary of L-Cystathionine in the NCBI PubChem database.

From the Compound Summary of L-Cystathionine from NCBI PubChem database https://pubchem.ncbi.nlm.nih.gov/compound/439258 (Figure 12), expand the section on "Biomolecular Interactions and Pathways," then click "Pathways."

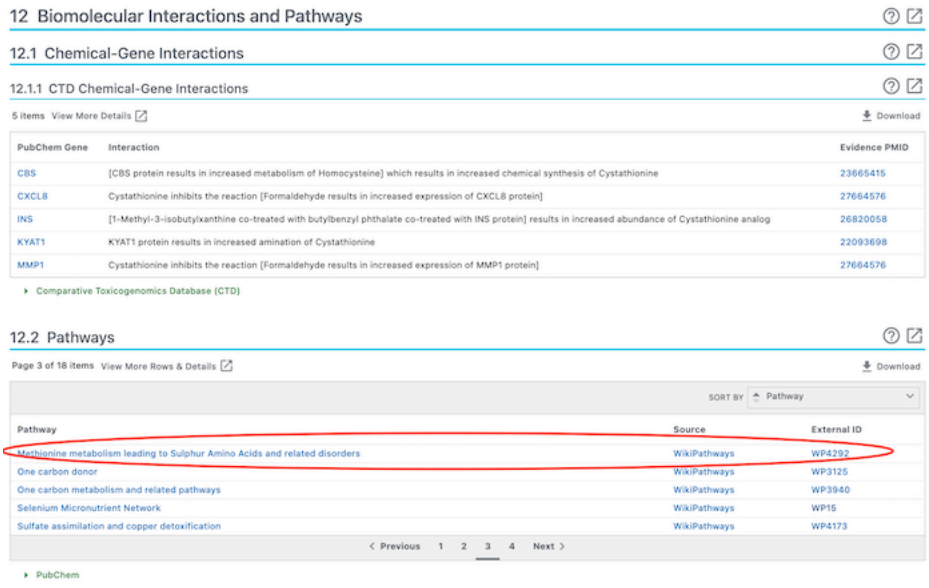


Figure 13. Biomolecular Pathways of L-Cystathionine in the NCBI PubChem database.

In PubChem Section 12.2 Pathways of L-Cystathionine https://pubchem.ncbi.nlm.nih.gov/compound/439258#section=Pathways (Figure 13), follow the link to WikiPathways database for "Methionine metabolism leading to Sulphur Amino Acids and related disorders."

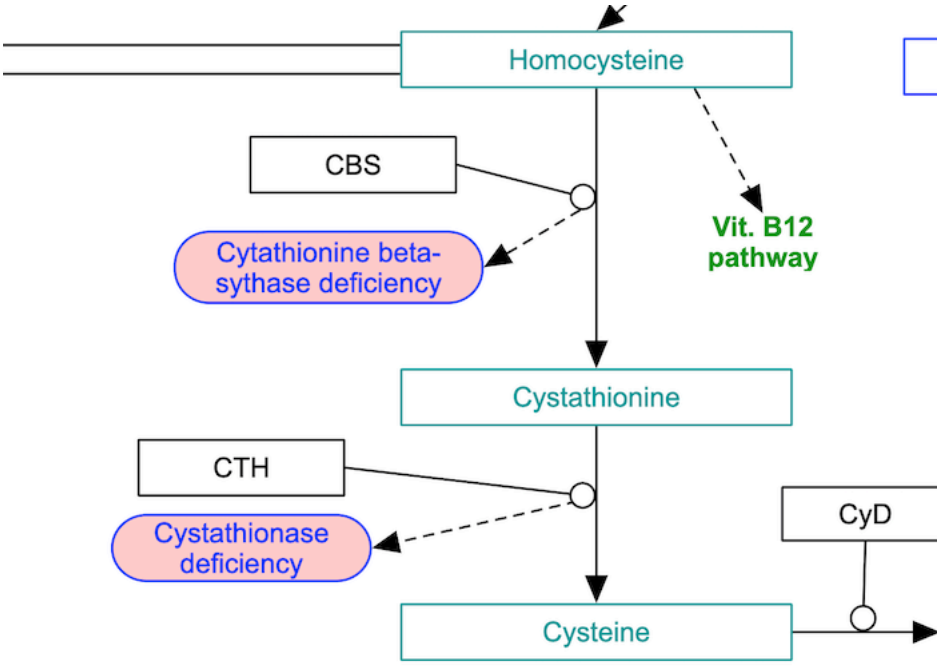


Figure 14. Section of Methionine metabolism leading to sulfur amino acids and related disorders (Homo sapiens) in the WikiPathways database.

In WikiPathways database, the reaction we are looking for is homocysteine to cystathionine in methionine metabolism leading to sulfur amino acids and related disorders (Homo sapiens) https://www.wikipathways.org/index.php/Pathway:WP4292 (Figure 14), where deficiency of the Cystathionine Beta-Synthase enzyme causes Homocystinuria.

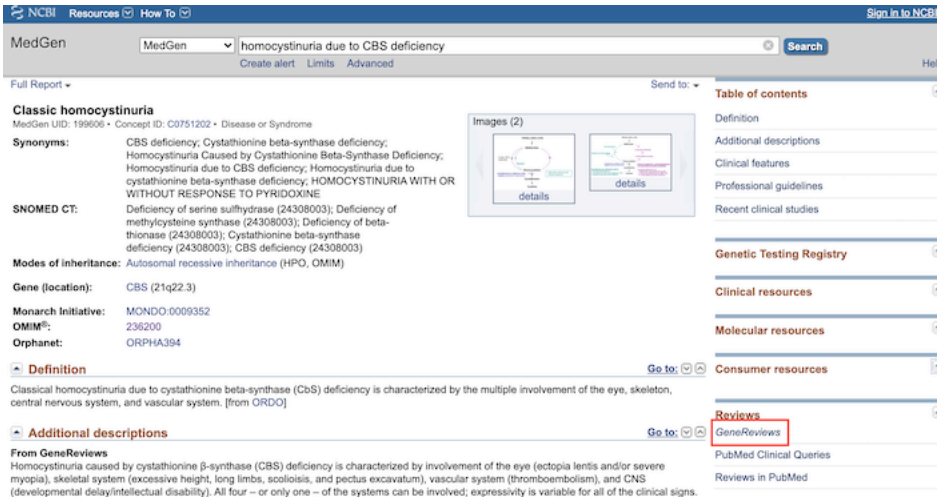


Figure 15. Entry for Classic homocystinuria in the NCBI MedGen database.

Go to the GeneReviews chapter that is linked from the entry for Classic homocystinuria in the NCBI MedGen database. https://www.ncbi.nlm.nih.gov/medgen/199606 (Figure 15).

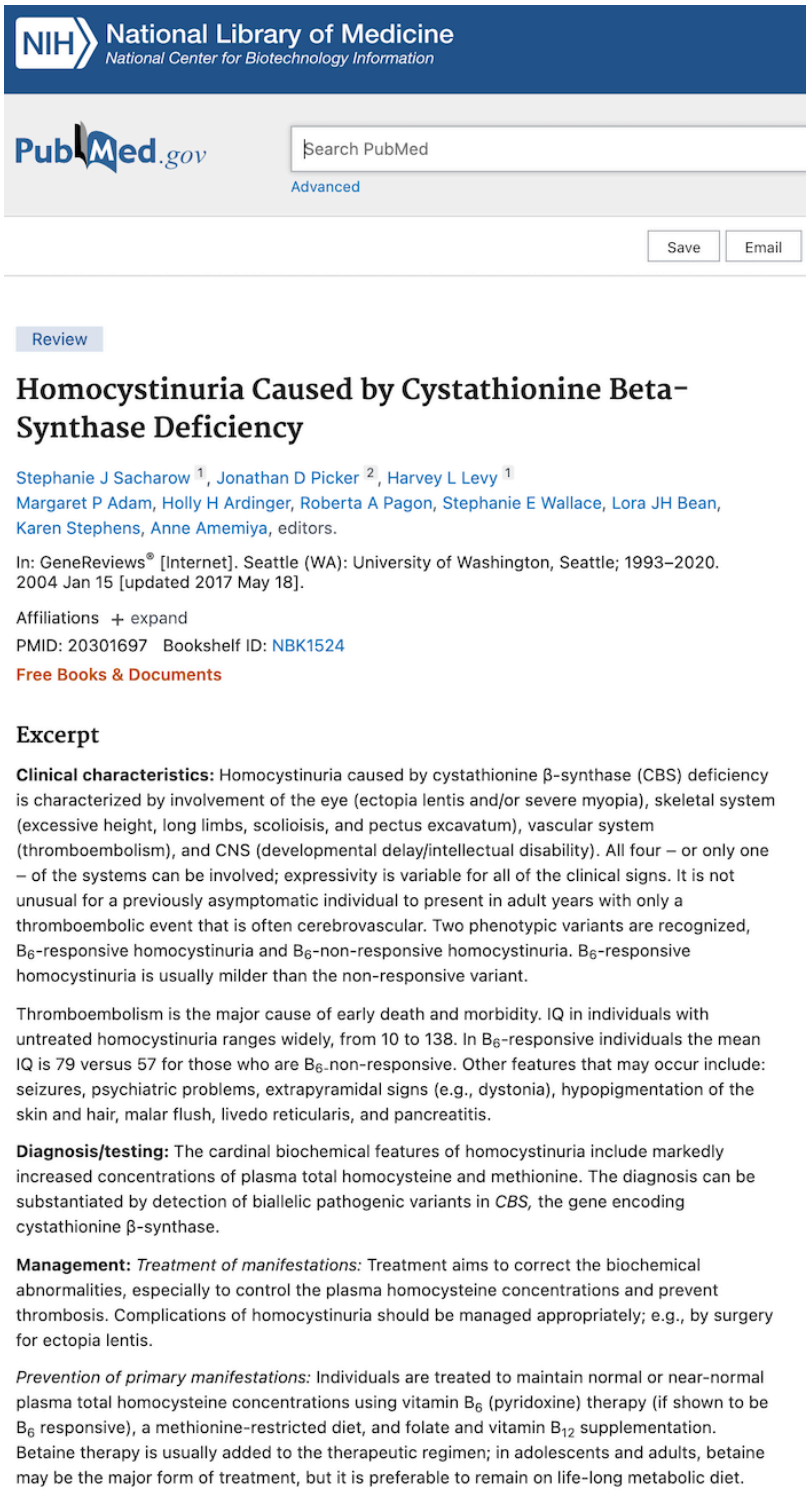


Figure 16. GeneReviews chapter on "Homocystinuria Caused by Cystathionine Beta-Synthase Deficiency" in the PubMed database.

The GeneReviews chapter covers management of the treatment for "Homocystinuria Caused by Cystathionine Beta-Synthase Deficiency" https://pubmed.ncbi.nlm.nih.gov/20301697/ (Figure 16).

Prevention of Primary Manifestations

The principles of treatment are to correct the biochemical abnormalities – especially to control the elevated plasma homocysteine concentrations as much as possible, to prevent or at least reduce the complications of homocystinuria [Yap & Naughten 1998], and to prevent further complications such as thrombosis [Morris et al 2017].

The best results have been reported in those individuals identified by newborn screening and treated shortly after birth in whom the plasma free homocystine concentration is maintained below 11 µmol/L (preferably, ≤5 µmol/L) [Yap et al 2001b]. This corresponds to a plasma total homocysteine concentration below 120 µmol/L or, preferably, below 100 µmol/L [Morris et al 2017]. For B_6-responsive individuals, the goal for plasma total homocysteine is below 50 µmol/L [Morris et al 2017].

These goals may need revision when very long-term data becomes available.

Measures used to control total plasma homocysteine concentration include vitamin B_6 (pyridoxine) therapy (if shown to be B_6 responsive), methionine-restricted diet, and folate and vitamin B_{12} supplementation. Betaine therapy is usually added to the therapeutic regimen; in adolescents and adults betaine may be the major form of treatment but it is preferable to remain on life-long metabolic diet. In those who have already had a vascular event, betaine therapy alone may prevent recurrent events [Lawson-Yuen & Levy 2010].

Details about each aspect of treatment follow.

Vitamin B_6 (pyridoxine) therapy. In those who are shown to be B_6 responsive, treatment with pyridoxine in a dose of approximately 200 mg/day or the lowest dose that produces the maximum biochemical benefit (i.e., lowest plasma homocysteine and methionine concentrations), as determined by measurement of total homocysteine and amino acid levels, should be given.

Pyridoxine may also be included in treatment despite evidence of B_6 non-responsiveness, typically in doses of 100-200 mg daily (although some adults receive 500-1000 mg daily).

Dietary treatment. B_6-non-responsive neonates or those only very poorly responsive to pyridoxine require a **methionine-restricted diet** with frequent metabolic monitoring. This diet should be continued indefinitely. Dietary treatment should be considered for clinically diagnosed individuals but often is not tolerated if begun in mid-childhood or later.

The majority of B_6-responsive individuals also require a methionine-restricted diet for metabolic control.

The diet for homocystinuria is very complex and the skills of an experienced metabolic dietician must be utilized. Dietary treatment reduces methionine intake by restricting natural protein intake. However, to prevent protein malnutrition, a methionine-free amino acid formula supplying the other amino acids (as well as cysteine, which may be an essential amino acid in CBS deficiency) is provided. Breast feeding may be continued in combination with the methionine-free amino acid infant formula [MacDonald et al 2006]. The amount of methionine required is calculated by a metabolic dietician and supplied in natural food and special low-protein foods and monitored on the basis of plasma concentrations of total homocysteine as well as methionine.

Folate and vitamin B_{12} supplementation. Folate and vitamin B_{12} optimize the conversion of homocysteine to methionine by methionine synthase, thus helping to decrease the plasma homocysteine concentration. When the red blood cell folate concentration and serum B_{12} concentration are reduced, folic acid is given orally at 5 mg per day; and vitamin B_{12} is given as hydroxycobalamin at 1 mg IM per month.

Betaine treatment. Treatment with betaine provides an alternate remethylation pathway to convert excess homocysteine to methionine (see Figure 1) and may help to prevent complications, particularly thrombosis [Yap et al 2001a, Lawson-Yuen & Levy 2010]. By converting homocysteine to methionine, betaine lowers plasma total homocysteine concentrations but raises the plasma concentration of methionine.

Figure 17. Prevention of Primary Manifestations of Homocystinuria from GeneReviews chapter, "Homocystinuria Caused by Cystathionine Beta-Synthase Deficiency."

According to the GeneReviews chapter, "Homocystinuria Caused by Cystathionine Beta-Synthase Deficiency," the homocystinuria condition responds to vitamins B6 (pyridoxine), B-12 (Cyanocobalamin) and folate and/or a methionine restricted diet and betaine https://www.ncbi.nlm.nih.gov/books/NBK1524/#homocystinuria.Management (Figure 17).

Though not drugs in the traditional sense these are all entities that have PubChem records. Search with the terms below in PubChem Compound at https://pubchem.ncbi.nlm.nih.gov/.

Pyridoxine https://pubchem.ncbi.nlm.nih.gov/compound/1054

Cyanocobalamin does not have a 3-D structure since it is a complex but it has a 2-D rendering https://pubchem.ncbi.nlm.nih.gov/compound/5311498s

Folic Acid https://pubchem.ncbi.nlm.nih.gov/compound/135398658

L-methionine https://pubchem.ncbi.nlm.nih.gov/compound/6137

Betaine https://pubchem.ncbi.nlm.nih.gov/compound/247

Multiple articles have been written about searching NCBI databases (NCBI Resource Coordinators 2016) including GenBank (Benson et al. 2013), Gene (Brown et al. 2015), MedGen (Louden 2020), OMIM (Amberger et al. 2015; Amberger & Hamosh 2017), and PubChem (Kim et al. 2021; Kim et al. 2016) for further reading. For additional guidance on NCBI databases, please refer to the National Center for Biotechnology Information's (NCBI) YouTube channel https://www.youtube.com/ncbinlm.

Acknowledgement

The author would like to thank Peter Cooper, Ph.D. of the National Center for Biotechnology Information for his consultation work on the search strategy.

References

About OMIM [Internet]. Baltimore (MD): Johns Hopkins University; c1966-2021 [cited 2021 Mar 23]. Available from https://www.omim.org/about.

About PubChem [Internet]. Legacy Bioassay Tools. Bethesda (MD): National Center for Biotechnology Information; [cited 2021 Mar 23]. Available from https://pubchemdocs.ncbi.nlm.nih.gov/legacy-bioassay-tools.

Amberger J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. & Hamosh, A. 2015. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research.* 43(D1):D789-D798. DOI: 10.1093/nar/gku1205.

Amberger, J.S. & Hamosh, A. 2017. Searching Online Mendelian Inheritance in Man (OMIM): A knowledgebase of human genes and genetic phenotypes. *Current Protocols in Bioinformatics.* 58:1.2.1-1.2.12. DOI: 10.1002/cpbi.27.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. 2013. GenBank. *Nucleic Acids Research.* 41(D1):D36-D42. DOI: 10.1093/nar/gks1195.

Brown, G.R., Hem, V., Katz, K.S., Ovetsky, M., Wallin, C., Ermolaeva, O., Tolstoy, I., Tatusova, T., Pruitt, K.D., Maglott, D.R., et al. 2015. Gene: A gene-centered information resource at NCBI. *Nucleic Acids Research.* 43(D1):D36-D42. DOI: 10.1093/nar/gku1055.

Choudhuri, S. 2014. Data, databases, data format, database search, data retrieval systems, and genome browsers. In: Choudhuri, S., editor. *Bioinformatics for Beginners.* Oxford (UK): Academic Press. p. 77–131.

Hähnke, V.D., Kim, S. & Bolton, E.E. 2018. PubChem chemical structure standardization. *Journal of Cheminformatics.* 10(1):1–40. DOI: 10.1186/s13321-018-0293-8.

Halavi, M., Maglott, D., Gorelenkov, V. & Rubinstein, W. 2018. MedGen. In: Beck, J. et al. editors. *The NCBI Handbook.* 2nd ed. Bethesda (MD): National Center for Biotechnology Information (US). Available from https://www.ncbi.nlm.nih.gov/books/NBK159970/.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B., et al. 2021. PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Research.* 49(D1):D1388–D1395. DOI: 10.1093/nar/gkaa971.

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., et al. 2016. PubChem Substance and Compound databases. *Nucleic Acids Research.* 44(D1):D1202–D1213. DOI: 10.1093/nar/gkv951.

Louden, D.N. 2020. MedGen: NCBI's portal to information on medical conditions with a genetic component. *Medical Reference Services Quarterly.* 39(2):183–191. DOI: 10.1080/02763869.2020.1726152.

NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research.* 44(D1):D7-19. DOI: 10.1093/nar/gkv1290.