

Chapter 10

Network Analysis and Integration in a Virtual Cell Environment

Björn Sommer

Abstract Integrative Bioinformatics combines a number of different disciplines related to biology as well as informatics. One major target of this research area is the creation of a virtual cell. Naturally, this topic is accompanied by a vast amount of problems which arise due to the fact that a large number of highly specific disciplines have to be addressed. In this publication a subproblem will be discussed, functional cell modeling. Beginning with a virtual cell environment which provides cell components featuring different subcompartmental layers, protein-related networks will be localized and visualized. For the localization, a data warehouse will be accessed. Special interactive techniques will be applied to the semiautomatic analysis of localization entries found in databases. Finally, different visualization approaches will be shown, and 2D and 3D network visualization will be discussed, as well as quantitative illustrations using charts.

10.1 Introduction

One of the most demanding tasks of today's scientific community is revealing the secrets of the biological cell. Three aspects make it nearly impossible to accomplish this task: (1) the unimaginable complexity of the cell, (2) its extremely small scale, and (3) the amount of differing disciplines to unite for this task.

The modeling and simulation of a virtual cell is one important approach to understanding the functioning of the cell. In the past, approaches like VCell and E-Cell were developed which mathematically simulated cells by using differential

B. Sommer (✉)
Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany
e-mail: bjoern@CELLmicrocosmos.org

equations [15, 24]. Naturally, these mathematical simulations address only quantitative simulations ignoring molecular interactions. Atomistic simulations using, e.g., molecular dynamic techniques are able to simulate small fragments of a cell like a membrane patch or vesicles [4, 5, 10]. However, it is not possible to simulate an entire cell taking molecular interactions into account. Alternatively, many different cell visualization approaches were developed in recent years, such as The Interactorium, MetNetVR, or Meta!Blast [26–28]. They visualize cells at the mesoscopic scale, where cell components can be differentiated but not molecular structures.

10.1.1 CELLmicrocosmos PathwayIntegration

This book is devoted to Integrative Bioinformatics which – as was discussed in the preceding chapters – comprises a number of different disciplines uniting biology and informatics. In this chapter a set of these disciplines will be applied to correlate a virtual cell environment with metabolic networks. In this way the aforementioned cellular mesoscopic level is combined with the functional level. For this purpose, the CELLmicrocosmos 4.2 PathwayIntegration (CmPI) will be used, a cell modeling and visualization environment [20, 21].

The mesoscopic level is represented by a set of cell components, such as the cell membrane, the cytosol, and the extracellular matrix, the mitochondrion, and the nucleus. These cell components are represented by three-dimensional models as well as by color codes as shown in Fig. 10.5. These contrasting color codes are based on the color alphabet developed by Green-Armytage, enabling a good visual differentiation [9, 20].

The functional level is covered by two well-known metabolic pathways: the citrate cycle and the glycolysis. Both pathways are interrelated, because the glycolysis generates pyruvate which is needed to initiate the citrate cycle. An important fact for subsequent analysis is that the localization of both pathways is well known. The cytosol – the intracellular fluid surrounding all membrane-based organelles – is the reaction chamber of the glycolysis. After a number of reactions, the final product of the glycolysis, the pyruvate, is transported by a specific protein through the inner mitochondrial membrane to the mitochondrial matrix. There, the pyruvate is oxidatively decarboxylized by the pyruvate dehydrogenase complex, resulting in the product acetyl CoA. This compound enters the citrate cycle which is also located in the mitochondrial matrix. The final result of the citrate cycle is the citrate [2, 6].

The following sections will discuss a workflow combining the mesoscopic and the functional level by using Integrative Bioinformatics techniques.

10.2 Relevant Databases

10.2.1 Metabolic Pathways from KEGG

First of all, the metabolic pathways have to be acquired from an electronic source. One of the most acknowledged biochemical databases and one of the most frequently used sources in Bioinformatics is the Kyoto Encyclopedia of Genes and Genomes (KEGG) [14]. It includes genomic, chemical, as well as systemic functional information, and it is available at <http://www.kegg.jp/kegg/>.

KEGG is partially freely, partly commercially available and has been developed in the Kanehisa Laboratories of Kyoto University. It is well known for its two-dimensional interlinked pathway maps. Figure 10.1 shows the human citrate cycle. The KEGG identifier is “hsa00020,” where “hsa” is an abbreviation for the organism (Homo sapiens) and 00020 is the KEGG-internal number of the pathway map.

KEGG contains a large number of different eukaryotic and prokaryotic organisms which are all linked to different versions of the same pathway. The enzymes

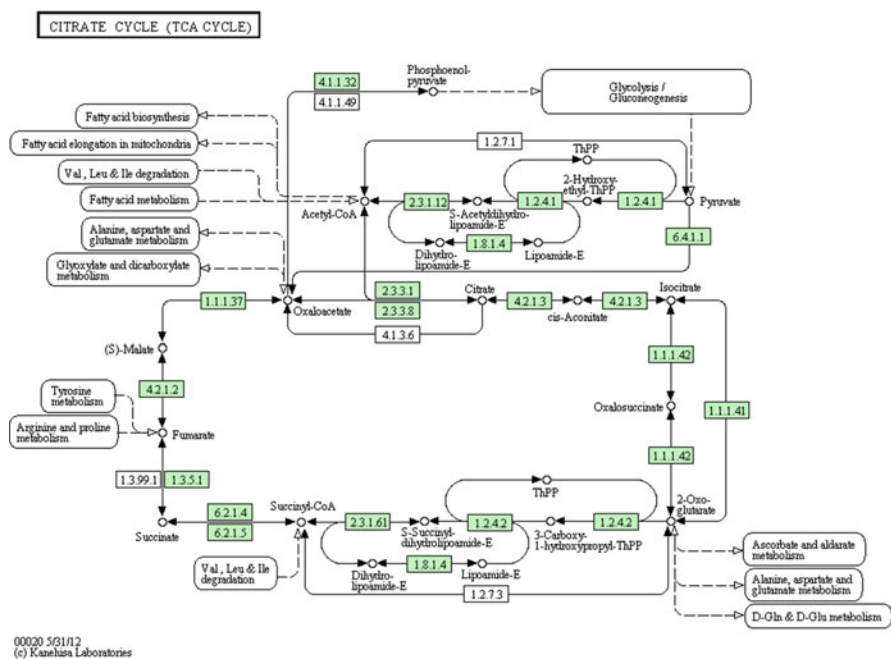


Fig. 10.1 KEGG: the citrate cycle pathway of Homo sapiens, hsa00020 (Courtesy of/Copyright 2013 by Kanehisa Laboratories, source: <http://www.kegg.jp>. Reprinted with permission from [14])

in Fig. 10.1 are coded following the EC standard discussed in Sect. 10.2.2.2. The enzymes are connected to their products and substrates which are named by their commonly known identifiers. The connections symbolize the reactions and the reactions' directions are indicated by arrows. All other elements are interrelated metabolic pathways. Here, the connection to the glycolysis can also be found.

Alternatively, it is possible to load externally created pathways into CmPI. For this purpose, a SBML import has been integrated [12]. Biological network reconstruction tools like VANESA (see Chap. 8: Biological Network Modeling and Analysis) can be used to model networks which are then imported to CmPI and localized in a virtual cell environment [13].

10.2.2 Protein Localization Databases

The digital data describing metabolic pathways is now available, and now the question arises, how can this structure – visualized as a two-dimensional image as was seen in Fig. 10.1 – be combined with the spatial structure of a cell? Because KEGG does not provide information about the localization of the networks, alternative sources have to be accessed. These sources will be discussed in this section. Four databases will be introduced which can be applied to this problem.

10.2.2.1 Reactome

First of all, the Reactome databases should be introduced. It is developed by the European Bioinformatics Institute (EBI) and different American institutes. It is a freely available, curated Open Source project. Similar to KEGG, it contains a large variety of different pathways and of course also a number of metabolic pathways. The major focus lies on the human organism. Expert users may integrate experimental data into Reactome. For each protein complex part, specific pathway localization information is available [7]. The database is found at <http://www.reactome.org>.

10.2.2.2 BRENDA and the Enzyme Classification

In contrast to the previous databases, the one following does not contain pathway maps. BRENDA (BRAunschweig ENzyme Database) is developed and curated at the TU Braunschweig. It can be freely accessed at: <http://www.brenda-enzymes.org>

A commercial version of BRENDA also exists which contains additional current information. In BRENDA, the user finds functional structural and property-related data which is mainly based on manually annotated references from primary literature [19].

The classification of the different enzyme types – which was already shown in Sect. 10.2.1 – follows the Enzyme Commission number (EC) classification, which basically consists of four numbers subdivided by a period [25]. The first number codes, e.g., (1) oxidoreductases, (2) transferases, (3) hydrolases, (4) lyases, (5) isomerases, and (6) ligases.

It is important to note that the EC numbers do not usually describe one particular protein. Instead, it applies to a number of different proteins which meet the criteria of the specific EC definition. Therefore, a single EC number may describe a large set of different proteins. The BRENDA databases link many EC numbers to specific databases based on manually curated literature. Of course, the aforementioned problem applies also to the localization, because it is not linked to a specific protein, but a protein family.

10.2.2.3 UniProt

In contrast to BRENDA, UniProt contains information linked to specific proteins. It is the freely accessible and regularly updated universal protein database for curated as well as automatic acquired data. UniProt is a collaboration between the European Bioinformatics Institute (EBI), the Protein Information Resource (PIR), and the Swiss Institute of Bioinformatics (SIB). It is linked to various external databases, such as BRENDA, Gene Ontology, and Reactome [8]. It is available at <http://www.UniProt.org>.

UniProt contains different sub-databases, but for the localization of proteins, this work will focus at the UniProt Knowledgebase (UniProtKB). The website and the database contain a number of categories holding localization information, for example:

- General annotation (comments)
 - Subcellular location
- Ontologies
 - Keywords
 - Cellular component
- Gene Ontology
 - Cellular component

The terms found in these categories may be a concrete cell component, an intra-compartmental location, or a sentence describing location-related facts.

10.2.2.4 The Gene Ontology and the Redundancy of Terms

A problem which arises when dealing with the large variety of different components of a cell is redundancy. Using the different databases previously introduced, a large

number of different terms may describe the same entity. Just to give an example, an excerpt of different terms provided by the databases BRENDA, UniProt, and Reactome will now be listed for the term “plasma membrane.”

- BRENDA
 - Cellular component
 - Cell membrane
 - Plasma membrane
 - Cytoplasmic membrane
 - Cell outer membrane
- UniProt
 - Associated with the synaptic plasma membrane (by similarity)
 - Integral to plasma membrane
 - Intrinsic to internal side of plasma membrane
 - Localized on the cell surface
- Reactome
 - Integrin cell surface interactions

Obviously, each of these terms is associated with the plasma membrane. But it also can be seen that each term contains additional information which might be relevant in different contexts.

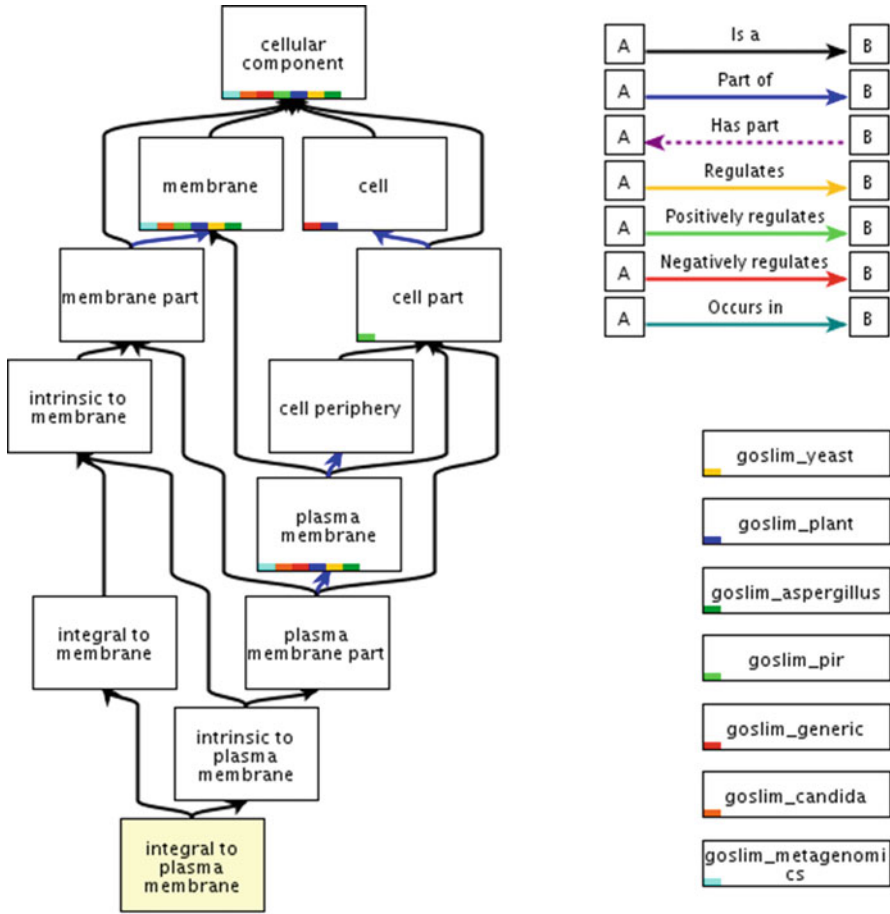
For a long time, the Gene Ontology (GO) has been addressing this problem [1, 3]. This database contains gene-related protein information in conjunction with structured, controlled vocabularies. The ontologies contain all these terms and link them to the so-called GO-terms which are a quasi standard in the Bioinformatics community. GO is located at <http://www.geneontology.org>.

Now, the GO vocabularies should be examined by looking again at the major term “plasma membrane” which has the GO identifier “GO:005886.” Directly correlated to this term is each of the three terms “cell membrane,” “plasma membrane,” and “cytoplasmic membrane.”

But the other previously listed terms are more specific and are linked to the following GO-terms:

- GO:0005887: integral to plasma membrane
- GO:0009279: cell outer membrane
- GO:0031235: intrinsic to internal side of plasma membrane

Now the question arises, how are these terms connected to the previously mentioned term GO:005886 representing the plasma membrane? These hierarchical dependencies are also addressed by GO, as can be seen in Fig. 10.2 showing the GO Graph View for GO:0005887. The terms GO:005886 and GO:0031235 are also contained and show spatial interdependencies: “integral to plasma membrane” → “intrinsic to plasma membrane” (→ “plasma membrane part”) → “plasma membrane.”



QuickGO - <http://www.ebi.ac.uk/QuickGO>

Fig. 10.2 Gene Ontology: graph view for the GO-term “integral to plasma membrane” (Courtesy of/Copyright 2013 by The Gene Ontology, AmiGO version 1.8, <http://amigo.geneontology.org>. Reprinted with permission from [1, 3])

But it also has to be mentioned that currently no GO identifiers were found for the following entries:

- Localized on the cell surface
- Associated with the synaptic plasma membrane (by similarity)
- Integrin cell surface interactions

The problem which applies to all databases discussed here is that they have to be continuously curated and extended.

10.2.3 DAWIS-M.D.

CmPI uses all previously discussed databases to solve the following application case. To enable the fast access of this large amount of diverse data, all databases were integrated in a data warehouse. The used data warehouse called DAWIS-M.D. was already extensively discussed in Chap. 4: Data Warehouses in Bioinformatics [11].

10.2.4 ANDCell

The previously mentioned databases contain a large amount of curated data. In many cases there is no data found for a specific localization in these databases. But if the user searches PubMed via its web interface, a publication might be found which was published in the most recent year. Of course, this information will usually find the way into the previously mentioned databases within the next new releases. But sometimes this will take some time.

Therefore, an alternative way is needed to acquire this information directly from PubMed. For this purpose, text mining is an appropriate approach. CmPI uses the ANDCell database for this purpose [18]. This database was already discussed in Chap. 6: Text Mining on PubMed.

10.3 Localizing Metabolic Pathways Using Integrative Bioinformatics Techniques

Now that the basics has been discussed, the correlation of the mesoscopic and the functional level has to be addressed by using CmPI. First of all, the metabolic pathways will be downloaded, and then these pathways will be correlated with a cellular environment by using the localization databases.

10.3.1 Downloading the Citrate Cycle and the Glycolysis

CmPI connects to DAWIS-M.D. to download the two pathways hsa00010, the glycolysis, and hsa00020, the citrate cycle. Figure 10.3 shows the resulting two-dimensional visualization of hsa00010 in the 2D viewer of CmPI and Fig. 10.4 hsa00020. Both layouts are based on the original KEGG layouts, the so-called KGML (KEGG Markup Language) pathway maps. These pathway maps are well known from the KEGG website which provides images for each pathway. If comparing the original KGML layouts (Fig. 10.1) from the websites with the 2D visualization of CmPI (Fig. 10.4), an important difference will be noted. The KGML layouts often contain multiple instances of a protein. For example, the original

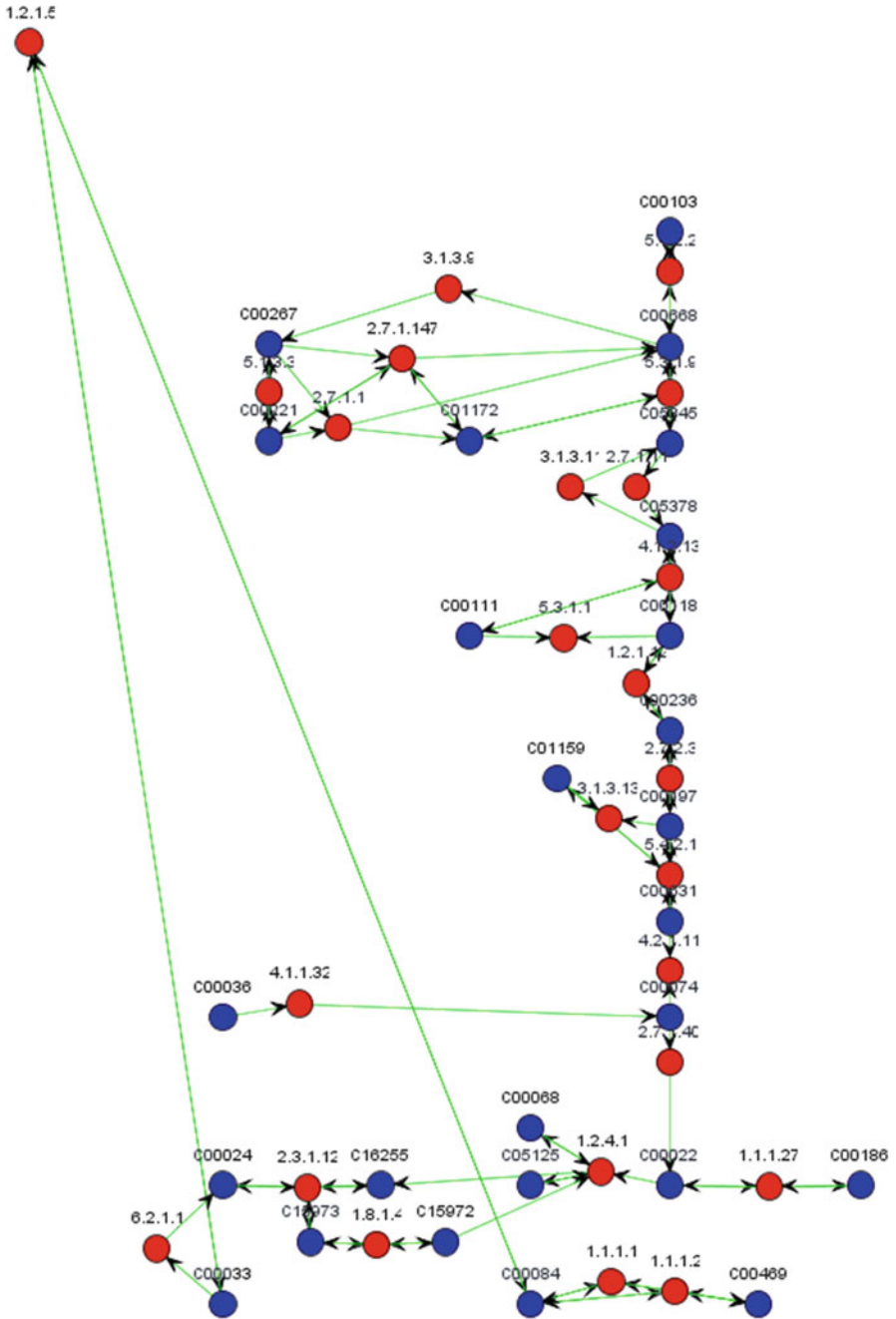


Fig. 10.3 2D visualization of the glycolysis (hsa00010) in CmPI

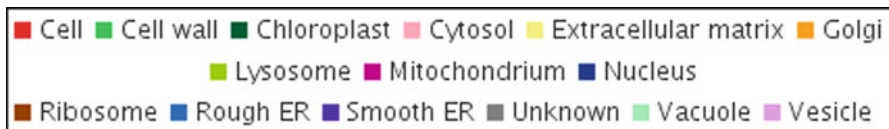


Fig. 10.5 Color codes for all cell components coded in the following figures

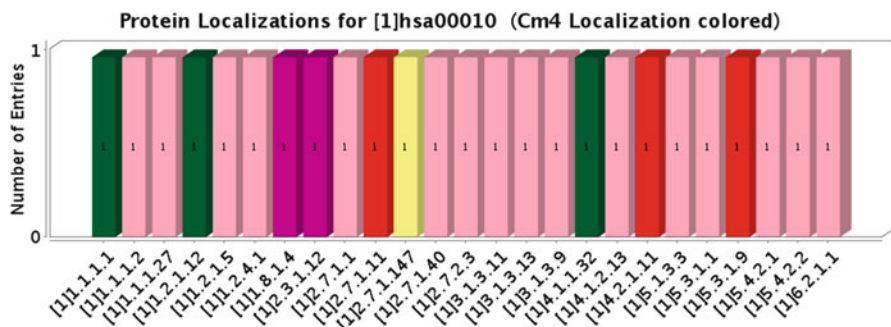


Fig. 10.6 Initial Localization Chart, category “Localizations/Cm4” for hsa00010

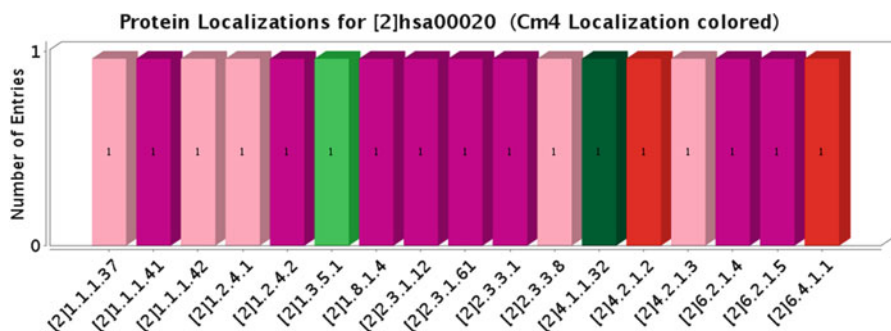


Fig. 10.7 Initial Localization Chart, category “Localizations/Cm4” for hsa00020

First, the downloaded metabolic pathways are localized by using the connection between CmPI and DAWIS-M.D. For this purpose, the EC identifier of each protein is sent to the localization databases combined with the information that the organism of interest is “Homo sapiens.” Otherwise, the localization results will contain a lot of information from other organisms. Now, a number of diverse localizations are retrieved, as shown in Fig. 10.5 which also lists the color codes used during the following analysis.

To analyze and filter this data, CmPI provides a special visualization, the Subcellular Localization Charts.

Figures 10.6 and 10.7 show the initially assigned localizations by CmPI which automatically selects the first entry in the alphabetically ordered list downloaded

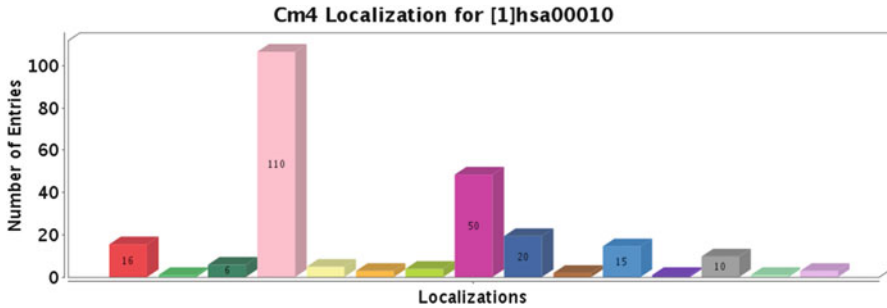


Fig. 10.8 Initial Localization Chart, category “Protein Localizations/Cm4” for hsa00010

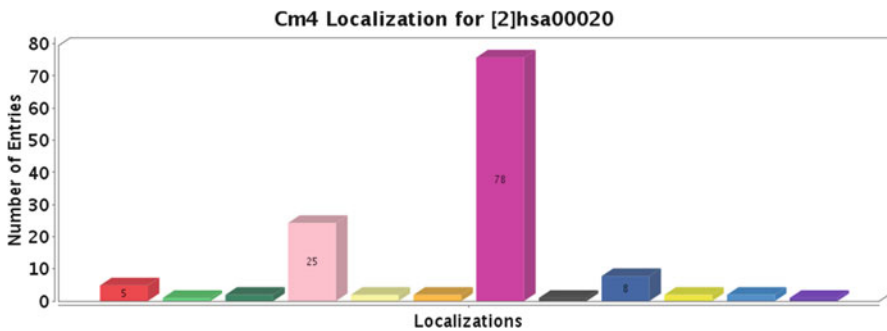


Fig. 10.9 Initial Localization Chart, category “Protein Localizations/Cm4” for hsa00020

from the databases as the potential localizations for each protein. Of course, the result does not meet the expectations given by the literature. The glycolysis and the citrate cycle each show five different localizations. For example, the result “chloroplast” does not meet the expectations of a Homo sapiens-related cell. It will be shown that only a few clicks are needed to assign the correct localizations.

10.3.3 Investigating the Preliminary Localizations

Now the Localization Charts will be used to investigate all localization entries found in the databases. Each entry is one single result from one of the five localization databases pointing to one single localization. Therefore, a single database like UniProt usually provides multiple localization entries for a single protein. All localization entries for glycolysis is shown in Fig. 10.8 and the citrate cycle in Fig. 10.9.

And obviously, recalling the initial expectations for both pathways seem to be confirmed by a first glance at the images. While the enzymes of hsa00010 are mainly localized at the cytosol, those of the hsa00020 are concentrated at the mitochondrion. Moreover, it is interesting to note that the localization with the

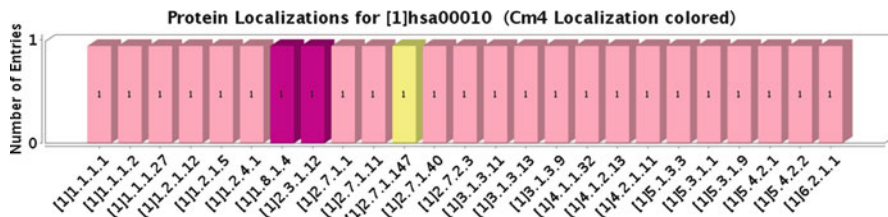


Fig. 10.10 Localization Chart, category “Protein Localizations/Cm4” after assigning the localization “cytosol” to hsa00010

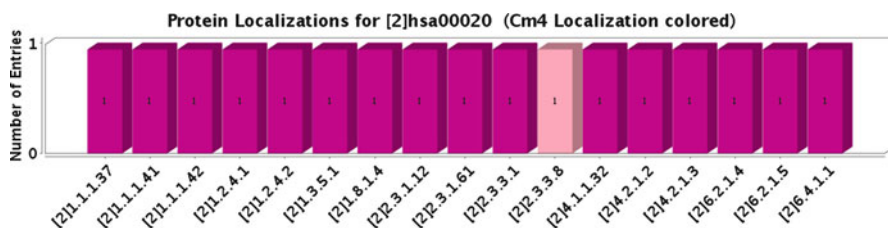


Fig. 10.11 Localization Chart, category “Protein Localizations/Cm4” after assigning the localization “mitochondria” to hsa00020

second highest incidence is – in both cases – the localization with the highest priority of the pathway vis-a-vis.

These images give a first idea of a localization, but it is very important to be aware of the fact that the shown localizations apply to the complete pathway. This visualization does not show any information about the single proteins. Therefore, it might be that 50 % of the proteins cannot be localized to the localization confirmed by the literature.

Using the Localization Charts, there are different ways to find a solution. Here, a method will be chosen which quickly leads to the final localization.

A double-click is performed on each of the bars representing the dominant localizations in Fig. 10.8, the cytosol, and Fig. 10.9, the mitochondrion. This action is already sufficient to assign most of the proteins to the correct localizations for both pathways, as can be seen by the resulting Localization Charts in Figs. 10.10 and 10.11.

Twenty two of twenty five enzymes were localized at the cytosol for the glycolysis and 16 of 17 enzymes at the mitochondrion for the citrate cycle. Therefore, the initial assumption was verified, but in addition to this, it is also possible to interpret more. Glycolysis and citrate cycle interact, but by checking the textbook, it is not known which concrete proteins are involved in the transition between cytosol and the mitochondrion. By looking at Fig. 10.10, it is possible to directly identify these enzymes; 1.8.1.4 and 2.3.1.12 are part of both pathways, and the only localization in the context of these two pathways is the mitochondrion. In addition, the enzyme 2.3.3.8, which here is only part of the citrate cycle, seems to be an enzyme not localized in the mitochondrion but in the cytosol.

2	1.8.1.4	mitochondrial matrix ...	mitochondrial matrix	1/5 Matrix	REACTOME: by G.O. (1/5)
2	1.2.4.2	mitochondrial inner me...	mitochondrial inner membrane	2/5 Inner Mem...	BRENDA: Reviewed (1/7)
2	2.3.1.61	mitochondrial matrix ...	mitochondrial matrix	1/5 Matrix	REACTOME: by G.O. (1/5)
2	6.2.1.5	mitochondrial chromos...	mitochondrial chromosome	1/5 Matrix	ANDCell: PubMed (1/5)
2	6.2.1.4	mitochondrial chromos...	mitochondrial chromosome	1/5 Matrix	ANDCell: PubMed (1/4)
2	1.1.1.42	mitochondrial matrix ...	mitochondrial matrix	1/5 Matrix	REACTOME: by G.O. (4/9)
2	1.1.1.41	mitochondrial matrix ...	mitochondrial matrix	1/5 Matrix	REACTOME: by G.O. (1/8)
2	2.3.3.8	citrate lyase complex H...	citrate lyase complex	1/1 Cytosol	GO: TAS:ProtInc (1/7)
2	1.3.5.1	mitochondrial respirator...	mitochondrial respiratory cha...	2/5 Inner Mem...	GO: ISS:UniProtKB (3/9)
2	4.2.1.2	mitochondrial matrix ...	mitochondrial matrix	1/5 Matrix	REACTOME: by G.O. (4/5)
2	4.2.1.3	mitochondrial chromos...	mitochondrial chromosome	1/5 Matrix	ANDCell: PubMed (10/20)
2	2.3.3.1	mitochondrial matrix ...	mitochondrial matrix	1/5 Matrix	GO: IDA:UniProtKB (1/3)
2	1.1.1.37	mitochondrial matrix ...	mitochondrial matrix	1/5 Matrix	REACTOME: by G.O. (4/7)
2	6.4.1.1	mitochondrial matrix ...	mitochondrial matrix	1/5 Matrix	REACTOME: by G.O. (3/6)
2	1.2.4.1	mitochondrial matrix ...	mitochondrial matrix	1/5 Matrix	GO: IEA:UniProtKB-SubCell (2/12)
2	2.3.1.12	mitochondrial pyruvate ...	mitochondrial pyruvate dehy...	1/5 Matrix	GO: NAS:UniProtKB (1/7)
2	4.1.1.32	mitochondrial matrix ...	mitochondrial matrix	1/5 Matrix	REACTOME: by G.O. (8/10)

Fig. 10.12 An excerpt of the Localization Table showing only hsa00020

Now, a more accurate examination of the different localizations should be done by looking at the Localization Terms. The Localization Terms describe the concrete entries downloaded from the databases which were used to map them to the localizations inside CmPI. For example, the terms listed in Sect. 10.2.2.4 – “associated with the synaptic plasma membrane (by similarity),” “integral to plasma membrane,” and “intrinsic to internal side of plasma membrane” – each are mapped to the Cm4 Localization, the cell membrane. The Localization Term used to map 2.3.3.8 to the cytosol is the “citrate lyase complex.” By examining Fig. 10.4, it can be seen that 2.3.3.8 is directly involved in the generation of the citrate (C00158) which is processed in the citrate lyase complex converting citrate to oxaloacetate [2, 6].

Next, the accuracy of the Membrane Localization of the citrate cycle should be verified by looking at the Localization Table part of CmPI in Fig. 10.12. This figure shows an excerpt of the Localization Table showing all citrate cycle-associated enzymes. At a first glance it can be seen that most of the enzymes were correctly localized to the mitochondrial matrix. But for the localization of the enzymes 1.2.4.2 and 1.3.5.1, the mitochondrial inner membrane was selected. The Localization Table shown in Fig. 10.12 is interactive. Therefore, all found localizations can also be found here by clicking the corresponding entry. For 1.3.5.1, an alternative option to the mitochondrial matrix is only the outer membrane of the mitochondrion. Therefore, the selection of the inner membrane seems to be correct. In contrast to this, the enzyme 1.2.4.2 shows also an entry for the matrix which can be directly selected in the Localization Table.

10.3.4 Examining an Outsider by Direct Access to External Sources

Because the glycolysis is localized at the cytosol, Membrane Localizations as those discussed for the citrate cycle (e.g., mitochondrial matrix or mitochondrial inner

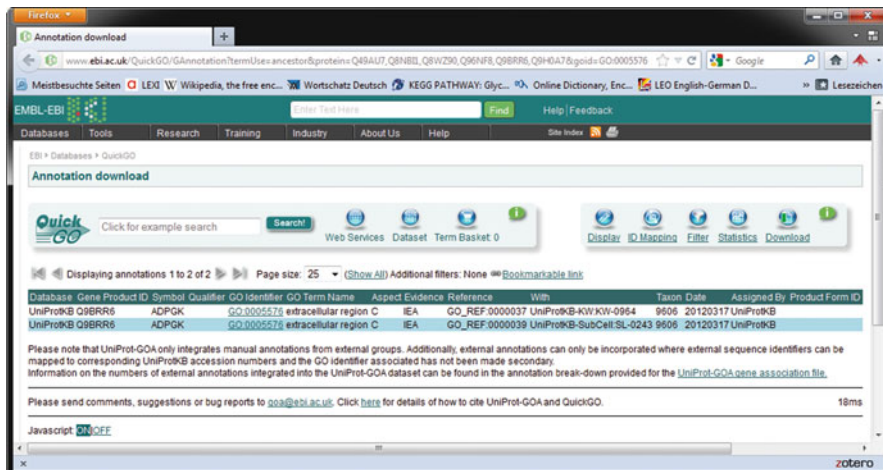


Fig. 10.13 The link to 2.7.1.147 from the Localization Table in CmPI; it shows additional localization information

membrane) are not of relevance. But still there is one outsider enzyme – 2.7.1.147 – because this enzyme was localized at the extracellular matrix by the Localization Term “extracellular region.” By using the Localization Table, the localization can be examined with respect to its source. The database which provided this result was GO based on the Localization Term “Inferred from electronic annotation. Source: UniProtKB-SubCell.” The Localization Table can also be used to click on the provided link which directly opens a web browser with the address “<http://www.ebi.ac.uk/QuickGO/>.” Here, the link to the UniProt entry Q9BRR6 is shown, based on two references (Fig. 10.13). Examining the UniProt entry shows that this enzyme is also involved in the glycolysis. Concluding these observations there are two potential reasons why the enzyme 2.7.1.147 was localized at extracellular matrix:

1. The enzyme 2.7.1.147 is in fact located in the extracellular matrix during the involvement in the glycolysis.
2. Due to the fact that the distance between the extracellular matrix and the cytosol is too large, the second option is more probable. There is currently no experimental proof available in the databases proving that this enzyme is located in the cytosol.

In both cases it can be predicted that 2.7.1.147 will most probably be located at the cytosol during the glycolysis, because all interacting enzymes are also found in this cell component.

10.3.5 *Localization Result*

Finally, it can be stated that the preliminary Localization Charts discussed in Sect. 10.3.2 are visually equal to the final result. But two changes according to the Membrane Localizations of two enzymes have to be done. The resulting localization priority list would look like this:

1. Mitochondrial matrix
2. Mitochondrial inner membrane
3. Cytosol
4. Extracellular space

Therefore, it was shown that nearly no foreknowledge would have been needed to predict the localization of these two pathways by using CmPI.

10.3.6 *3D Visualization*

Of course, the Subcellular Localization Charts of CmPI can be used to analyze protein-related data sets without the intention to generate a virtual cell environment. However, for a number of application cases, it will be relevant to visualize the networks in correlation with a cell model.

In the previous sections, it was discussed how the databases are queried to get (1) the metabolic pathways (the glycolysis and the citrate cycle) and (2) to localize these pathways. Now, the networks have to be correlated with the cell components inside the cell model. For this purpose, a geodesic layout is combined with an Inverted Self-Organizing Map (ISOM) layout [17] and then mapped onto the surface of the cell components [20, 21].

This process is subdivided into three steps. First, the nodes – representing the enzymes, substrates, and products – are distributed onto the surface of a unit sphere by using the geodesic layout [21]. By doing so, the layout tries to achieve a node distribution where the distance between a given node and its neighboring nodes is equal. In the second step, this initial layout is used to apply the ISOM layout [17]. This layout moves interconnected nodes – these are nodes connected by a reaction – closer to each other, whereas those nodes without interconnections try to move away from each other.

After the first two steps have been accomplished, the resulting layout is shown in Fig. 10.14. This image shows an abstract SphereCell containing only cell components relevant for the final localizations (from inside to outside): mitochondrion, cytosol, cell membrane, and extracellular matrix [20]. The applied contrast color-coding was also used for the enzyme localization and is described in Fig. 10.5.

The third step of the layout process is the mapping of the nodes onto the surface of the cell components. This process is not needed in the case of the SphereCell, because the layout is directly applied to its spherical cell components. But of course, this extremely simplified representation of a cell is usually not sufficient, because

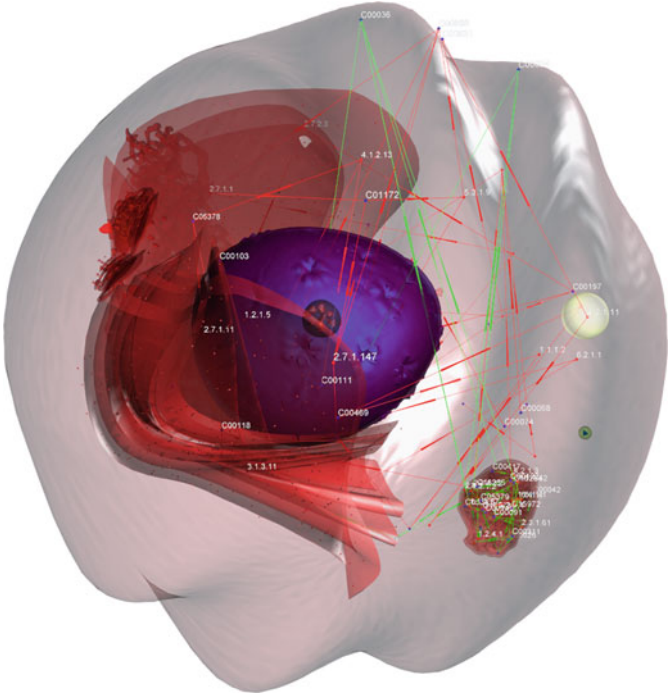


Fig. 10.15 An animal cell model correlated with the citrate cycle and glycolysis (color figure online)

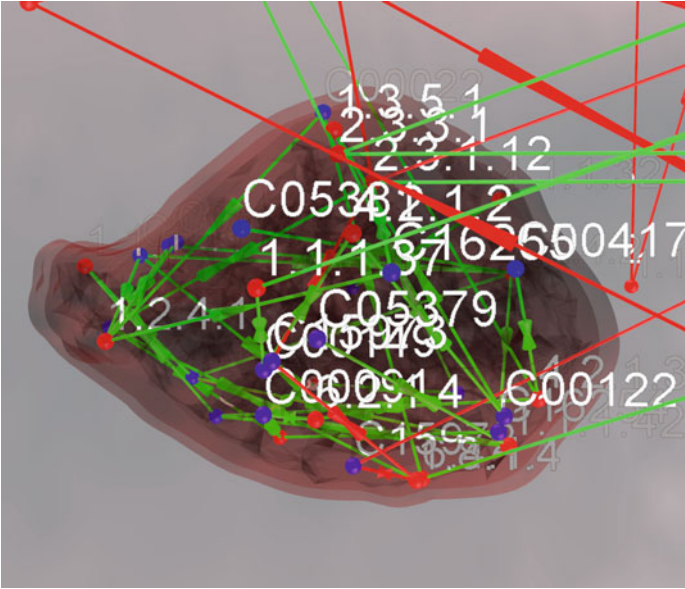


Fig. 10.16 A mitochondrion model based on microscopic data correlated with the citrate cycle and glycolysis (color figure online)

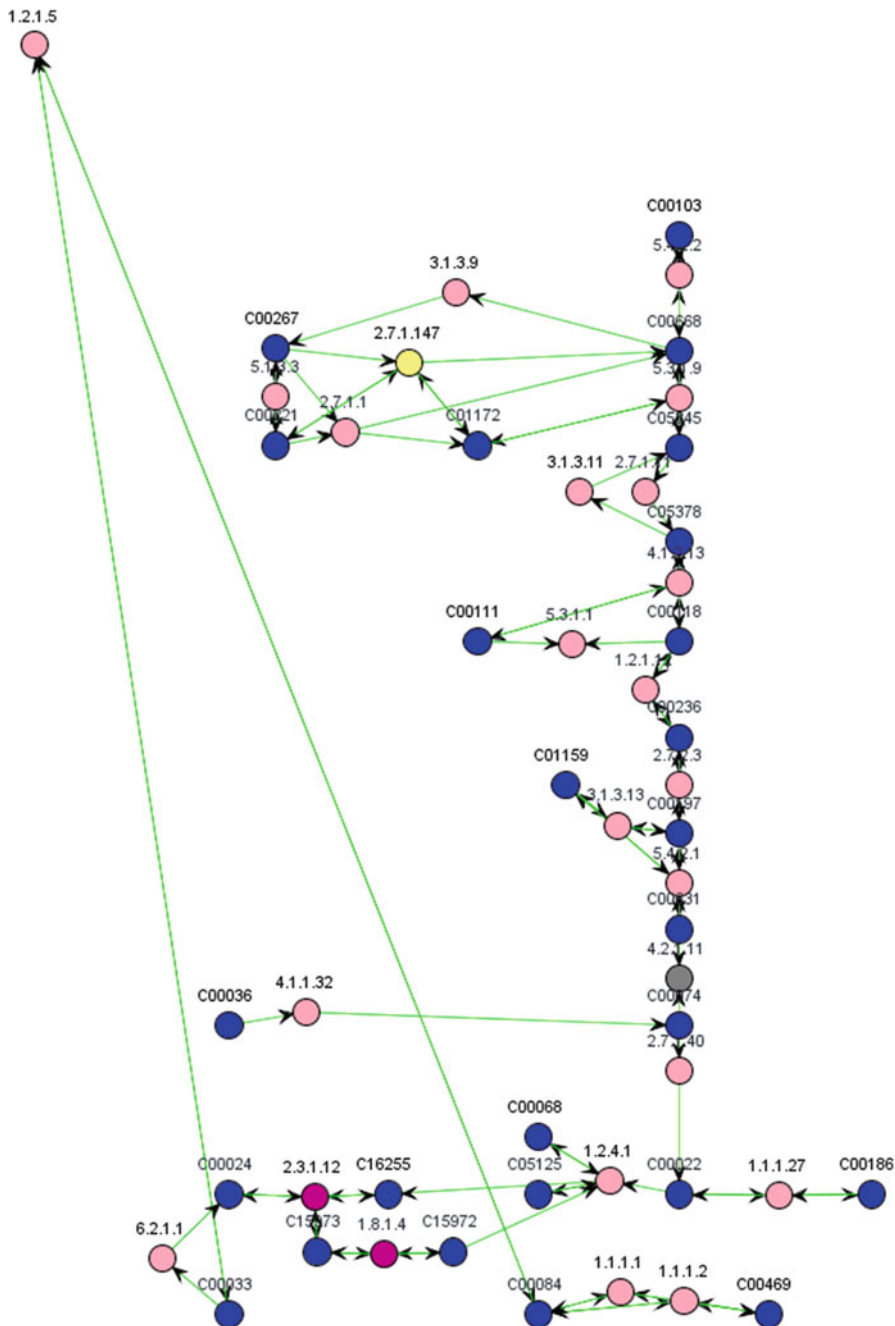


Fig. 10.17 2D visualization of the glycolysis (hsa00010) in CmPI using the localization colors for the enzymes

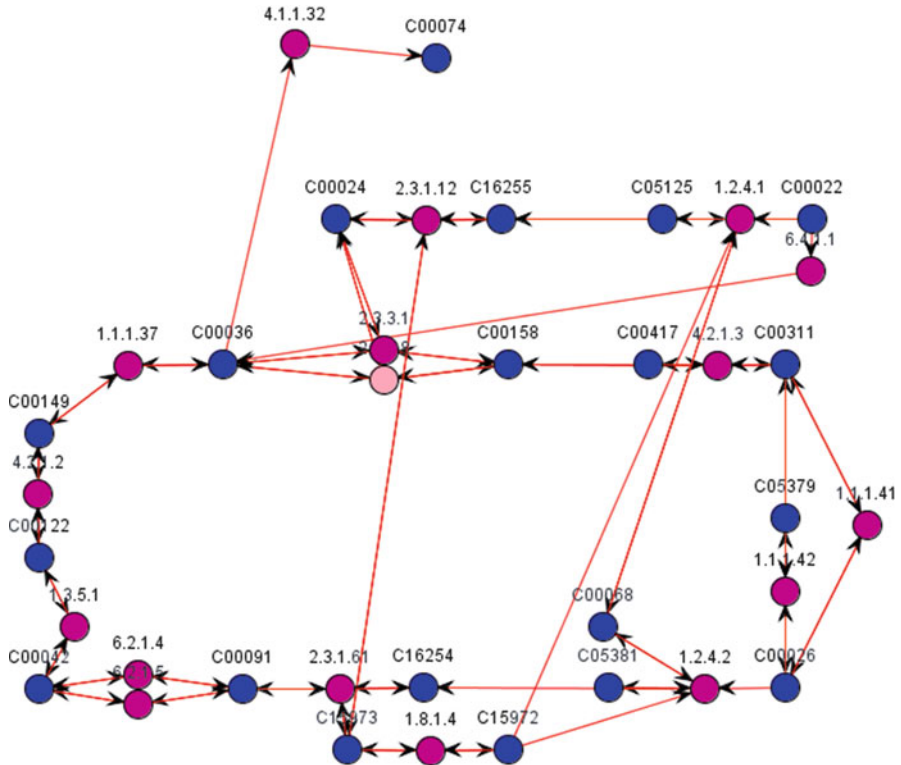


Fig. 10.18 2D visualization of the citrate cycle (hsa00020) in CmPI using the Localization Colors for the enzymes

by coloring the enzymes according to the color codes known from Fig. 10.5. And Fig. 10.19 shows all potential enzyme localizations found in the databases, providing a good overview of all available localization information.

10.4 Conclusions

In conclusion, it can be stated that Integrative Bioinformatics techniques can be used to generate and analyze biological networks, to predict the localization of its components, and to use different visualization techniques to enable the discussion of potential results. It should be mentioned again that the information provided by KEGG concerning the involved enzymes is quite vague, because the EC classification does not describe specific proteins, but protein families. Therefore, many different localizations for each enzyme were available. Of course, CmPI was

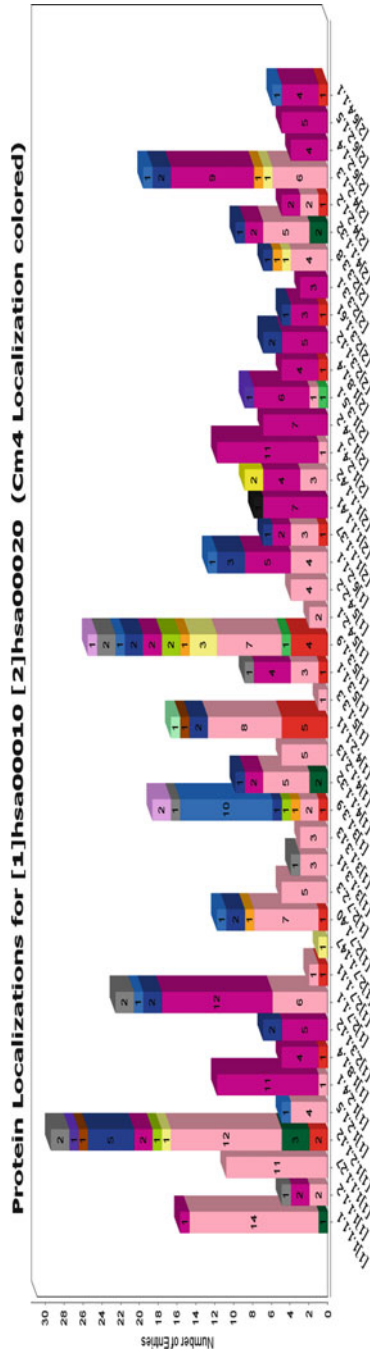


Fig. 10.19 Localization Chart, category “Protein Localizations/Cm4” showing all potential enzyme localizations found in the databases

already used to localize specific proteins in a cellular environment, for example, to examine the localization of a cardiovascular disease-related protein set [22, 23].

Despite this fact, it was shown that the CmPI can be used (1) for the prediction and analysis of the localization of protein-related networks or sets by using the Subcellular Localization Charts and (2) to visualize and explore the results in two and three dimensions. Therefore, the functional and mesoscopic level of cytology can be combined; microscopic data sets – as known, e.g., from the Cell Centered Database [16] – can be used as a base to combine spatial cellular structures with metabolic pathways.

References

1. Ashburner M, Ball CA, Blake JA et al (2000) Gene Ontology: tool for the unification of biology (The Gene Ontology Consortium). *Nat Gen* 25:25–29
2. Berg JM, Tymoczko JL, Stryer L (2006) *Biochemistry*, 6th edn. W.H. Freeman, New York
3. Blake JA, Chan J, Kishore R (2013) Gene Ontology annotations and resources. *Nucl Acids Res* 41:D530–D535
4. Brooks BR, Brooks CL, Mackerell Jr AD et al (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30:1545–1614
5. Case DA, Cheatham TE, Darden T et al (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688
6. Cooper GM, Hausman RE (2007) *The cell: a molecular approach*, 4th edn. ASM, Washington
7. Croft D, O’Kelly G, Wu G (2011) *Reactome: a database of reactions, pathways and biological processes*. *Nucl Acids Res* 39:D691–D697
8. Dimmer EC, Huntley RP, Alam-Faruque Y (2012) The UniProt-GO annotation database in 2011. *Nucl Acids Res* 40:D565–D570
9. Green-Armytage P (2010) A colour alphabet and the limits of colour coding. *Colour Des Creat* 5:1–23
10. Hess B, Kutzner C, van der Spoel D et al (2008) Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447
11. Hippe K, Kormeier B, Töpel T et al (2010) DAWIS-MD—a data warehouse system for metabolic data. *GI Jahrestag* 2:720–725
12. Hucka M, Finney A, Sauro HM et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19:524–531
13. Janowski S, Kormeier B, Töpel T et al (2010) Modeling of cell-to-cell communication processes with Petri nets using the example of quorum sensing. *In Silico Biol* 10:27–48
14. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucl Acids Res* 40:D109–D114
15. Loew LM, Schaff JC (2001) The Virtual Cell: a software environment for computational cell biology. *TRENDS Biotechnol* 19:401–406
16. Martone ME, Gupta A, Wong M et al (2002) A cell-centered database for electron tomographic data. *J Struct Biol* 138:145–155
17. Meyer B (1998) Self-organizing graphs—a neural network perspective of graph layout. *Lect Notes Comput Sci (Graph Drawing)* 1547:246–262
18. Podkolodnaya OA, Yarkova EE, Demenkov PS et al (2011) Application of the ANDCell computer system to reconstruction and analysis of associative networks describing potential relationships between myopia and glaucoma. *Russ J Genet Appl Res* 1:21–28

19. Scheer M, Grote A, Chang A (2011) BRENDA, the enzyme information system in 2011. *Nucl Acids Res* 39:D670–D676
20. Sommer B (2012) CELLmicrocosmos – integrative cell modeling at the molecular, mesoscopic and functional level. Bielefeld University, Bielefeld
21. Sommer B, Künsemöller J, Sand N et al (2010) CELLmicrocosmos 4.1: an interactive approach to integrating spatially localized metabolic networks into a virtual 3D cell environment. In: Fred A, Filipe J, Gamboa H (eds) BIOINFORMATICS 2010 – proceedings of the 1st international conference on bioinformatics, part of the 3rd international joint conference on biomedical engineering systems and technologies (BIOSTEC 2010), INSTICC, Valencia
22. Sommer B, Tiys E S, Kormeier B et al (2010) Visualization and analysis of a cardio vascular disease- and MUPP1-related biological network combining text mining and data warehouse approaches. *J Integr Bioinform* 7:148
23. Sommer B, Kormeier B, Demenkov P S et al (2013) Subcellular localization charts: a new visual methodology for the semi-automatic localization of protein-related data sets. *J Bioinform Comput Biol* 11:1340005
24. Takahashi K, Ishikawa N, Sadamoto Y (2003) E-Cell 2: multi-platform E-Cell simulation system. *Bioinformatics* 19:1727–1729
25. Webb EC (1992) Enzyme nomenclature – recommendations of the nomenclature committee of the international union of biochemistry and molecular biology on the nomenclature and classification of enzymes, NC-IUBMB. Academic, San Diego
26. Widjaja YY, Pang CNI, Li SS, Wilkins MR et al (2009) The interactorium: visualising proteins, complexes and interaction networks in a virtual 3D cell. *Proteomics* 9:5309–5315
27. Wurtele ES, Li J, Diao L et al (2003) MetNet: software to build and model the biogenetic lattice of Arabidopsis. *Comp Funct Genom* 4:239–245
28. Wurtele ES, Bassham DC, Dickerson J et al (2010) Meta!Blast: a serious game to explore the complexities of structural and metabolic cell biology. In: Proceedings of the ASME 2010 world conference on innovative virtual reality, ASME, Ames
29. Yamaguchi R, Lartigue L, Perkins G (2008) Opa1-mediated cristae opening is Bax/Bak and BH3 dependent, required for apoptosis, and independent of Bak oligomerization. *Mol Cell* 31:557–569