

V

V Domain

- [Variable \(V\) Domain](#)

V Gene

- [Variable \(V\) Gene](#)

V Type Domain

- [Variable \(V\) Domain](#)

Vaccine

Ramachandran Srinivasan
G.N. Ramachandran Knowledge Centre for Genome Informatics, Institute of Genomics and Integrative Biology, Delhi, India

Definition

Vaccine is a formulation of killed or attenuated pathogens, or antigens derived from them which help prevent, ameliorate, or treat infectious disease by stimulating antibody production or cellular immunity against the pathogen.

Vaccine Antigen Databases

Hifzur Rahman Ansari¹, Darren R. Flower² and Gajendra Raghava¹

¹Bioinformatics Centre, Institute of Microbial Technology, Chandigarh, Chandigarh, India

²School of Life and Health Sciences, Aston University, Birmingham, UK

Definition

Vaccine is a substance that trains the host immune system to protect itself against a pathogen. Vaccine antigen databases are compilations of different types of resources directly or indirectly leading to the development of vaccines. In order to train, a variety of antigens are used to activate different arms of immune system such as B- and T-cell epitopes, and innate immune components. This entry describes databases of various components of the immune system used to design a vaccine.

Characteristics

Vaccine Antigen Databases

MHCBN

Updated version 4.0 of the MHCBN (Lata et al. 2009) contains more than 25,000 experimentally characterized T-cell and MHC binding peptides with affinity and activity data compiled from ~1,500 literatures and other public databases. Coverage of ~1,000 transport-associated ► [antigen](#) processing (TAP) and ~4,000 MHC non-binder peptides put MHCBN at a position

distinct from other similar immunological databases. Database covers all major hosts like human, mouse, rat, monkey, etc., and provide information of 20 major MHC-linked diseases. Database provides a user-friendly search interface which can be used for simple and customized queries. This also provides several tools for peptide searching, mapping, and BLAST search against MHCBN data.

AntiJen

AntiJen v2.0 (Toseland et al. 2005) while retaining the B- and T-cell ► [epitopes](#) archived quantitative data on a variety of immunological interactions which makes it both unique and useful. The database contains more than 31,000 entries with 11 sub-databases. These interactions include peptide binding to TAP, peptide binding to MHC, peptide-MHC (p-MHC) binding to TCR, and protein–protein interactions in terms of molecular thermodynamic, kinetic, and biophysical data such as diffusion coefficient and receptor copy number. All of these parameters are key components of effective vaccine design.

IEDB

The Immune Epitope Database (IEDB) (Vita et al. 2010) is the most comprehensive and progressive database to date in its field. IEDB contains more than 75,000 experimentally characterized B- and T-cell epitopes, MHC binding and elution experiment data compiled from ~10,500 literature sources, and data directly submitted by different labs. They have covered ~2,500 organisms based on the priority set by National Institute of Allergy and Infectious Diseases (NIAID) which includes allergens and autoantigens. Database provides useful complementary tools based on the latest algorithms for the analysis and prediction of epitopes and MHC binding. It provides improved user interface for general and advanced searching or browsing by source organism or MHC allele type.

IEDB has greatly strengthened the field of immunoinformatics by employing the concept of Ontology of Immune Epitopes (ONTIEs) which has enabled curators and users to establish relationships between published literature and database entries.

AntigenDB

Earlier, the principal means of vaccine development was empirical: a live virulent or attenuated pathogen was used which are now considered poor vaccine

candidates despite its potent ► [immunogenicity](#). These are more likely to cause disease in children or immune-compromised individuals. Due to safety and availability of powerful ► [adjuvants](#), the trend is shifting toward antigen, acellular, or ► [subunit vaccine](#) and ► [epitope](#)-based vaccines. Approaches for the identification of antigens include the cumbersome assays for epitopes, antibodies, cytokine levels, and testing for enhanced survival in animal models. Most of the databases mentioned above catalog those antigens for which B- and T-cell epitopes are known, otherwise despite being ► [immunogenic](#) as a whole they are lost in the literature. Such antigens or similar pathogenic proteins might prove useful in vaccine design. AntigenDB database project (Ansari et al. 2010) started with the objective of compiling all immunogenic antigens from all major pathogens for which epitope information is known or unknown. Application of AntigenDB is to help researchers working for the development of antigen prediction algorithms or vaccinologists who would like to use these antigens as ► [subunit vaccine](#) candidates. Currently AntigenDB contains more than 500 antigens coming from 44 important pathogenic species. A database entry contains information regarding sequence, structure, origin, etc., and additional information like B- and T-cell epitope, MHC binding, function, and gene expression information if available. It can be searched by using simple keywords or browsing using pathogen names, after which data can be easily exported to an Excel file. AntigenDB also provides useful tools like BLAST for analysis of the antigenic similarity, epitope search to check availability of that or similar epitopes in the database, and peptide mapping to see whether query protein contains any known epitopes or not. Data submission tool is also provided to submit data online which is included in the database after validation.

Allergen Databases

Allergy is a condition caused by allergenic proteins due to IgE antibody-mediated immune responses that affect more than 30% of the population in developed countries with major economic burden (Kamble and Bharmal 2009). Allergenic proteins, a class of antigens, induce the specific IgE antibody production upon binding to surface-bound IgE on mast cells or basophils leading to release of inflammatory mediators. A person sensitized to pollen allergens may also react to several plant foods due to similarity in

Vaccine Antigen Databases, Table 1 Vaccine antigen resources

Resource	Url	Remarks	Reference
IEDB	www.iedb.org	B- and T-cell epitopes, MHC binding/nonbinding peptides and elution assays	(Vita et al. 2010)
AntigenDB	www.imtech.res.in/raghava/antigendb	B- and T-cell epitopes, antigen's information	(Ansari et al. 2010)
MHCBN	www.imtech.res.in/raghava/mhcbn	MHC binding and nonbinding peptides	(Lata et al. 2009)
BCIPEP	www.imtech.res.in/raghava/bcipep	B-cell epitope	(Saha et al. 2005)
AntiJen	http://www.darrenflower.info/antijen/	Quantitative data on MHC, B- and T-cell epitopes	(Toseland et al. 2005)
varDB	http://www.vardb.org/vardb/	Data on variable gene families	(Diez et al. 2010)
<i>Organism-/disease-specific resources</i>			
HIV	http://hivdb.stanford.edu/index.html	135,271 sequences, 123,047 isolates, 2,050,366 mutations, 53,102 phenotypes	(Shafer, 2006)
Cancer	TANTIGEN http://cvc.dfci.harvard.edu/tadb/index.html	4,006 curated antigen entries representing 251 unique proteins	
TB	TBDB (http://www.tbdb.org/) Tuberculist (http://genolist.pasteur.fr/TubercuList/) BioHealthBase (http://patricbrc.vbi.vt.edu/portal/portal/patric/Taxon?cType=taxon&cId=1763)		(Reddy et al. 2009)
FLU (influenza)	IRD (http://www.fludb.org/brc/home.do?decorator=influenza)		(Squires et al. 2008)
Malaria	MalVac (http://malvac.igib.res.in/)		(Chaudhuri et al. 2008)
Allergy	SDAP (http://fermi.utmb.edu/SDAP/) ADFS (http://allergen.nihs.go.jp/ADFS/), allergome (http://www.allergome.org)	Allergenic proteins from thousands of sources and tissues	(Ivanciuc et al. 2003; Mari et al. 2006)
<i>Community efforts and compendia</i>			
The vaccine Page	http://www.vaccines.org/		
Violin	http://www.violinet.org/		
NIAID	http://www.niaid.nih.gov/		

sequence and structure of cross-reactive allergenic proteins. Owing to its importance, a number of databases and analysis tools have been created in the past. Fortunately we have now the standard for the naming of allergens maintained by Allergen Nomenclature Sub-committee of the World Health Organization and International Union of Immunological Societies (WHO/IUIS). Major active databases in this area are summarized below.

SDAP (Structural Database of Allergenic Proteins)

The database was released in 2001 with continuous updates (Ivanciuc et al. 2003). To date SDAP contains

1,396 allergens, 1,221 protein sequences and 28 IgE and IgG epitopes, 70 allergens with PDB structures, and 582 3D models. The database provides information on allergen name, IUIS nomenclature, type, source species, sequence, IgE epitopes, literature, family domain, and structure information when available. Database can be searched through a user-friendly interface. It also hosts several useful computational tools for the prediction of allergenicity and cross-reactivity.

ADFS (Allergen Database for Food Safety)

ADFS was an initiative of Novel Foods and Immunology of the Japanese National Institute of Health

Sciences with new updates in February 2010 mainly dedicated to food safety. It contains 1,285 sequences, 77 molecular structures and 91 epitopes, and 88 sugar attached allergens derived from air, pathogens, food, and venom/salivary allergens. The database provides domain information, links to other external databases, and tools for the allergen analysis and prediction.

Allergome

Allergome (Mari et al. 2006) is the most comprehensive resource for allergens to date, encompassing 173 biological functions, 1,759 allergen sources, 53 tissues, ~700 structures, 17,000 bibliographies, and 5,276 entries. Allergome is updated on a daily basis and latest statistics can be found at the dedicated <http://www.allergome.org/script/statistic.php> page. Entries contain information regarding allergen name, source tissue and organism, route of exposure, taxonomy, primary sequence, motifs, post-translational factors, and cross-reactivity. The concept of Allergenicity Score was introduced which is a real-time characterization status of a given allergen. The score can be visualized on the dedicated page of the allergen using a 3-color, 11-dot sequence. Dots represent status-like data generation, sequence availability, and different functional/non-functional tests. Users can search the database for the desired characteristics based on these scores. The database also includes a user-friendly reference archive and an “allergome aligner” tool for the analysis of allergens.

Antigen Variation and Pathogen-Specific Resources

varDB

Vertebrates have developed an efficient immune system and pathogens need to adapt to this very hostile environment; therefore they devise different counter strategies like antigenic drift and variation to evade a host immune system. Some pathogens mutate their antigens so that antibodies are no longer able to recognize them while some hide themselves from the immune cells like malaria causing pathogen plasmodium to reside in red blood cells. Genome sequencing efforts revealed the existence of a large number of multi-copy gene families with suspected involvement in antigenic variation and virulence. The genomics and proteomics projects release a huge amount of sequence and phenotypic data. The varDB project (Diez et al. 2010)

aims to produce a dedicated platform for the antigen variation, phenotypes, and clinical manifestations. So far the varDB database has covered 42 gene families, accommodating 67,000 sequences from 27 species of organisms causing ~19 different diseases. varDB has created a framework for automatic data download, processing, and annotation. Entries can be queried and sorted based on the UniProt ID, gene products, isolation source, host, locus tag, etc. varDB has implemented a user-friendly tool, a shopping cart where sequences can be added and analyzed, and a correlation with disease outcome can be revealed leading to the identification of functionally important residues and drug targets.

In addition to generalized antigen vaccine databases, there is a growing battery of databases fully dedicated to a particular pathogen/disease or communities working toward the data compilation and understanding of major diseases in terms of vaccines (see Table 1).

Remaining Challenges

- Like peptide there should be quantitative data on protein affinity measures while binding to B-cell receptors or antibodies.
- Databases should contain potential antigen class based on *in silico* predictions.
- All antigen databases should coordinate and define the universal format like other protein sequences and structure databases.
- Terminology should be properly defined like the vaccine ontology initiative taken by IEDB.
- There should be a wider reach of resources for vaccine information networks like Violin and vaccine Page.
- There should be regular updating of each database which is a challenging task and whether we need generalized databases or organism specific databases are more intuitive.

Cross-References

- [Adjuvants](#)
- [Antigen](#)
- [Epitope](#)
- [Immunogen](#)
- [Subunit Vaccine](#)

References

- Ansari HR, Flower DR, Raghava GPS (2010) AntigenDB: an immunoinformatics database of pathogen antigens. *Nucleic Acids Res* 38:D847–D853
- Chaudhuri R, Ahmed S, Ansari FA, Singh HV, Ramachandran S (2008) MalVac: database of malarial vaccine candidates. *Malar J* 7:184
- Diez D, Hayes N, Joannin N, Normark J, Kanehisa M, Wahlgren M, Wheelock CE, Goto S (2010) varDB: a database of antigenic variant sequences—current status and future prospects. *Acta Trop* 114:144–151
- Ivanciu O, Schein CH, Braun W (2003) SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res* 31:359–362
- Kamble S, Bharmal M (2009) Incremental direct expenditure of treating asthma in the United States. *J Asthma* 46:73–80
- Lata S, Bhasin M, Raghava GPS (2009) MHCBN 4.0: a database of MHC/TAP binding peptides and T-cell epitopes. *BMC Res Notes* 2:61
- Mari A, Scala E, Palazzo P, Ridolfi S, Zennaro D, Carabella G (2006) Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The allergome platform as a model. *Cell Immunol* 244:97–100
- Reddy TB, Riley R, Wymore F, Montgomery P, DeCaprio D, Engels R, Gellesch M, Hubble J, Jen D, Jin H (2009) TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res* 37:D499–D508
- Saha S, Bhasin M, Raghava GP (2005) Bcipep: a database of B-cell epitopes. *BMC Genomics* 6:79
- Shafer RW (2006) Rationale and uses of a public HIV drug-resistance database. *J Infect Dis* 194(Suppl 1): S51–S58
- Squires B, Macken C, Garcia-Sastre A, Godbole S, Noronha J, Hunt V, Chang R, Larsen CN, Klem E, Biersack K (2008) BioHealthBase: informatics support in the elucidation of influenza virus host pathogen interactions and virulence. *Nucleic Acids Res* 36:D497–D503
- Toseland CP, Clayton DJ, McSparron H, Hemsley SL, Blythe MJ, Paine K, Doytchinova IA, Guan P, Hattotuwigama CK, Flower DR (2005) AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res* 1:4
- Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B (2010) The immune epitope database 2.0. *Nucleic Acids Res* 38: D854–D862

Vaccine Design

- [Reverse Vaccinology](#)

Vaccine Development

- [Reverse Vaccinology](#)

Vaccine Development Modeling

- [Vaccinomics](#)

Vaccine Molecular Signatures

- [Systems Immunology, Novel Evaluation of Vaccine](#)

Vaccine-induced Immune Response

- [Immunomonitoring](#)

Vaccinomics

Bertrand Bellier, Adrien Six, Véronique Thomas-Vaslin and David Klatzmann
Immunology-Immunopathology-Immunotherapy,
UPMC University Paris 06, UMR7211, CNRS,
UMR7211, INSERM, U959, Paris, France

Synonyms

[Systems vaccinology](#); [Vaccine development modeling](#)

Definition

Vaccinomics is the branch of omics, which deals with vaccine analysis. Vaccinomics encompasses the fields of immunogenetics and immunogenomics as applied to understanding the mechanisms of heterogeneity in immune responses to vaccines (Poland and Oberg 2010; Poland et al. 2008).

Cross-References

► [Systems Immunology, Novel Evaluation of Vaccine](#)

References

- Poland GA, Oberg AL (2010) Vaccinomics and bioinformatics: accelerants for the next golden age of vaccinology. *Vaccine* 20:3509–3510
- Poland GA, Ovsyannikova IG, Jacobson RM (2008) Personalized vaccines: the emerging field of vaccinomics. *Expert Opin Biol Ther* 11:1659–1667

Validation

► [CellML Model Curation](#)

Variable (V) Domain

Marie-Paule Lefranc
Laboratoire d'ImmunoGénétique Moléculaire,
Institut de Génétique Humaine UPR 1142, Université
Montpellier 2, Montpellier, France

Synonyms

V domain; V type domain; Variable domain

Definition

The Variable (V) domain is a type of structural unit (domain) that, with the ► [Constant \(C\) domain](#), characterizes a protein chain belonging to the ► [immunoglobulin superfamily \(IgSF\)](#). The V domain comprises the V-DOMAIN of the immunoglobulins (IG) or antibodies and T cell receptors (TR), and the V-LIKE-DOMAIN of the IgSF proteins other than IG and TR. The V domain description per receptor type and chain type, based on the ► [IMGT-ONTOLOGY concepts](#), is shown in [Table 1](#). IMGT[®] labels are in capital letters.

A V-DOMAIN results from the rearrangements of two or three genes (V-J rearrangement between a ► [variable \(V\) gene](#) and a ► [joining \(J\) gene](#), or V-D-J rearrangement between a V gene, a ► [diversity \(D\) gene](#) and a J gene) (► [Immunoglobulin Synthesis](#)) (Lefranc and Lefranc 2001a, b), whereas a V-LIKE-DOMAIN is usually encoded by one exon of a gene. Despite this major difference in the biosynthesis, the structure is very similar. Indeed, in both cases, the V domain is made of nine antiparallel beta strands (A, B, C, C', C'', D, E, F, and G) linked by beta turns (AB, CC', C''D, DE, and EF) and loops (BC, C'C'', and FG), forming a sandwich of two sheets (four strands on one sheet and five strands on the other).

The beta strands form the framework regions (FR) (► [Framework region \(FR-IMGT\)](#)), whereas the three loops that bind to the antigen and confer the specificity to the IG or TR correspond to the complementarity determining regions (CDR) (► [Complementarity determining region \(CDR-IMGT\)](#)).

According to the ► [IMGT unique numbering for V domain](#) (Lefranc et al. 2003), the five conserved amino acids of a V-DOMAIN always have the same position, cysteine 23 (1st-CYS), tryptophan 41 (CONSERVED-TRP), conserved hydrophobic amino acid 89, cysteine 104 (2nd-CYS), and tryptophan or phenylalanine 118 (J-TRP or J-PHE) (Lefranc et al. 2003). Amino acids of a V-LIKE-DOMAIN may be less conserved, particularly the amino acid at position 118.

Analysis of V domain sequences can be performed by tools of IMGT[®], the international ImMunoGeneTics information system[®] (<http://www.imgt.org>) (► [IMGT[®] information system](#)):

- Nucleotide sequences of IG and TR V-DOMAIN by IMGT/V-QUEST and IMGT/JunctionAnalysis (Lefranc 2009; Ehrenmann et al. 2010a) and for sequences from deep sequencing by IMGT/HighV-QUEST
- Amino acid sequences of V-DOMAIN and V-LIKE-DOMAIN by IMGT/DomainGapAlign (Ehrenmann et al. 2010b)

These tools align the user sequences with the closest V and J regions (for V-DOMAIN) and the closest V domains (for V-LIKE-DOMAIN) of the IMGT reference directory, create gaps according to the IMGT unique numbering for V domain (Lefranc et al. 2003), delimit the FR-IMGT and

Variable (V) Domain, Table 1 V domain description per receptor type and chain type

V domain	Receptor type	V domain description per chain type	Sequence description
V-DOMAIN	IG	VH ^a	V-D-J-REGION
		V-KAPPA ^b	V-J-REGION
		V-LAMBDA ^b	V-J-REGION
	TR	V-ALPHA	V-J-REGION
		V-BETA	V-D-J-REGION
		V-GAMMA	V-J-REGION
		V-DELTA	V-D-J-REGION
V-LIKE-DOMAIN	IgSF other than IG and TR	V-LIKE-DOMAIN	EX (exon with its corresponding number)

^aVH: V domain of IG-Heavy chain

^bIf the IG-Light chain type is not specified, V-KAPPA and V-LAMBDA are described as VL (V domain of IG-Light chain)

CDR-IMGT, highlight differences with the closest reference(s), and generate the ► [IMGT Collier de Perles](#).

IMGT/DomainGapAlign is widely used for the standardized delimitation of the FR-IMGT and CDR-IMGT, particularly in antibody humanization by CDR grafting (Lefranc 2009; Ehrenmann et al. 2010a). IMGT/DomainGapAlign allows to estimate the potential immunogenicity of a V-DOMAIN by providing the percentage of identity of the V-REGION framework of an humanized antibody compared with the closest human V-REGION and by characterizing the amino acid changes in the 4 FR-IMGT using the 11 amino acid physicochemical classes (Pommié et al. 2004).

Analysis of V domain three-dimensional (3D) structures is available in IMGT/3Dstructure-DB (Ehrenmann et al. 2010b). For antigen receptors crystallized with their ligands, a detailed description of the amino acids in contact (► [Paratope](#) of the V-DOMAIN, on the antigen receptor side, and ► [Epitope](#) or antigenic determinant, on the antigen side) is provided for the IG/antigen (IG/Ag) complexes (IG Paratope/antigen or B cell epitope) and for the TR/peptide/major histocompatibility (TR/pMH) complexes (TR paratope/pMH or T cell epitope).

Cross-References

- [Complementarity Determining Region \(CDR-IMGT\)](#)
- [Constant \(C\) Domain](#)
- [Diversity \(D\) Gene](#)
- [Epitope](#)
- [Framework Region \(FR-IMGT\)](#)

- [IMGT Collier de Perles](#)
- [IMGT Unique Numbering](#)
- [IMGT[®] Information System](#)
- [IMGT-ONTOLOGY](#)
- [Immunoglobulin Superfamily \(IgSF\)](#)
- [Joining \(J\) Gene](#)
- [Paratope](#)
- [Variable \(V\) Gene](#)

References

- Ehrenmann F, Duroux P, Giudicelli V, Lefranc M-P (2010a) Standardized sequence and structure analysis of antibody using IMGT[®]. In Kontermann R, Dübel S (eds) *Antibody engineering*, vol 2, 2nd edn. Springer-Verlag Berlin, Heidelberg, pp 11–31
- Ehrenmann F, Kaas Q, Lefranc M-P (2010b) IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res* 38:D301–317
- Lefranc M-P (2009) Antibody databases and tools: The IMGT[®] experience. In Zhiqiang An (ed) *Therapeutic monoclonal antibodies: from bench to clinic*. Wiley, Hoboken, New Jersey, pp 91–114
- Lefranc M-P, Lefranc G (2001a) *The immunoglobulin FactsBook*. Academic Press, London, pp 1–458.
- Lefranc M-P, Lefranc G (2001b) *The T cell receptor FactsBook*. Academic Press, London, pp 1–398
- Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 27:55–77
- Pommié C, Levadoux S, Sabatier R, Lefranc M-P (2004) IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J Mol Recognit* 17:17–32

Variable (V) Gene

Marie-Paule Lefranc

Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine UPR 1142, Université Montpellier 2, Montpellier, France

Synonyms

V gene; Variable gene

Definition

The variable (V) gene, or “*variable*,” is a ► [leafconcept](#) of the “► [GeneType](#)” concept of identification (generated from the ► [IDENTIFICATION Axiom](#)) of ► [IMGT-ONTOLOGY](#), the global reference in ► [immunogenetics](#) and ► [immunoinformatics](#) (Giudicelli and Lefranc 1999; Lefranc et al. 2004, 2005, 2008; Duroux et al. 2008), built by IMGT[®], the international ImMunoGeneTics information system[®] (<http://www.imgt.org>) (► [IMGT[®] Information System](#)). “*Variable*” identifies a gene that rearranges at the DNA level and codes the variable region of the variable domain (► [Variable \(V\) Domain](#)) of an immunoglobulin (IG) or antibody or of a T cell receptor (TR) chain (► [Chain Type](#)).

It is one of the four leafconcepts that are characteristics of the IG and TR loci (Lefranc and Lefranc 2001a, b), the other three being “diversity” (D), “joining” (J), and “constant” (C) (► [Diversity \(D\) Gene](#), ► [Joining \(J\) Gene](#), ► [Constant \(C\) Gene](#)).

An IG or TR variable (V) gene has two possible and exclusive configurations (► [Configuration Type](#)): a germline configuration (before DNA rearrangement) and a rearranged configuration (after DNA rearrangement). In the germline configuration, a variable (V) gene possesses in 3' (downstream) a recombination signal (V-RS) (► [Recombination Signal \(RS\)](#)). This V-RS and the 5' recombination signal of the J gene (5'J-RS) or D gene (5'D-RS) are specifically recognized by the enzyme recombinase that allows the V gene to be rearranged to a J gene or to a D gene, depending on the locus (► [Immunoglobulin Synthesis](#)).

The sequence resulting from the V-J or V-D-J rearrangement encodes the V-DOMAIN (► [Variable \(V\) Domain](#)) of the IG and TR chains (► [Chain Type](#)) (Lefranc and Lefranc 2001a, b).

Cross-References

- [Chain Type](#)
- [Configuration Type](#)
- [Constant \(C\) Gene](#)
- [Conventional Gene](#)
- [Diversity \(D\) Gene](#)
- [Gene Type](#)
- [IMGT[®] Information System](#)
- [IMGT-ONTOLOGY](#)
- [IMGT-ONTOLOGY, IDENTIFICATION Axiom](#)
- [IMGT-ONTOLOGY, Leafconcept](#)
- [Immunogenetics](#)
- [Immunoinformatics](#)
- [Immunoglobulin Synthesis](#)
- [Joining \(J\) Gene](#)
- [Recombination Signal \(RS\)](#)
- [Variable \(V\) Domain](#)

References

- Duroux P, Kaas Q, Brochet X, Lane J, Ginestoux C, Lefranc M-P, Giudicelli V (2008) IMGT-Kaleidoscope, the Formal IMGT-ONTOLOGY paradigm. *Biochimie* 90:570–583
- Giudicelli V, Lefranc M-P (1999) Ontology for immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics* 12:1047–1054
- Lefranc M-P, Lefranc G (2001a) The immunoglobulin FactsBook. Academic, London, pp 1–458
- Lefranc M-P, Lefranc G (2001b) The T cell receptor FactsBook. Academic, London, pp 1–398
- Lefranc M-P, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudou D, Jabado-Michaloud J, Magris S, Scaviner D, Thouvenin V, Combres K, Girod D, Jeanjean S, Protat C, Yousfi Monod M, Duprat E, Kaas Q, Pommie C, Chaume D, Lefranc G (2004) IMGT-ONTOLOGY for immunogenetics and immunoinformatics. *In Silico Biol* 4:17–29
- Lefranc M-P, Clément O, Kaas Q, Duprat E, Chastellan P, Coelho I, Combres K, Ginestoux C, Giudicelli V, Chaume D, Lefranc G (2005) IMGT-Choreography for immunogenetics and immunoinformatics. *In Silico Biol* 5:45–60
- Lefranc M-P, Giudicelli V, Regnier L, Duroux P (2008) IMGT, a system and an ontology that bridge biological and computational spheres in bioinformatics. *Brief Bioinform* 9:263–275

Variable Domain

- ▶ [Variable \(V\) Domain](#)

Variable Gene

- ▶ [Variable \(V\) Gene](#)

Variable Selection

- ▶ [Feature Selection](#)

Variable Subset Selection

- ▶ [Feature Selection](#)

Vector Quantization

- ▶ [Cluster Analysis](#)

Version Control

- ▶ [Distributed Version Control System \(DVCS\)](#)

Viability Explanation

- ▶ [Explanation, Functional](#)

Viral Respiratory Tract Infections

Christian Schönbach
Department of Bioscience and Bioinformatics, Kyushu
Institute of Technology, Iizuka, Fukuoka, Japan

Synonyms

[Acute upper respiratory tract infections](#); [Cold and flu-like illness](#); [Common cold](#)

Definition

Viral infections of the respiratory tract (▶ [Respiratory Tract, Upper and Lower](#)) are one of the most common disease causes in humans. Typical symptoms of upper respiratory tract infections are cough, wheezing, and a sore throat. Lower respiratory tract infections may manifest themselves in bronchiolitis, pneumonia, and exacerbations of asthma and chronic obstructive pulmonary disease (Nichols et al. [2008](#)). Respiratory tract infection–associated viruses belong to four families of RNA viruses and two families of DNA viruses ([Table 1](#)). Human rhinoviruses (HRVs), the focus of this essay, are thought to be the cause of more than half of all acute upper respiratory tract infections (common cold), worldwide (Lee et al. [2007](#)). The cytopathology of HRV is low. However, symptoms of the inflammatory host response and exacerbations of asthmatic and allergic conditions contribute to the high direct and indirect economic cost of HRV infections. Non-influenza virus-associated cost of respiratory tract infections including infections caused by HRV were estimated to be at least 40 billion US\$ per year, in the USA alone (Fendrick et al. [2003](#)). Approved vaccines and specific antiviral drugs are available for influenza A but not for most of the other viruses causing respiratory tract illness. Various degrees of clinical significance, high serotype diversity, insufficient *in vitro* propagation cell culture, and small animal models and/or limited understanding of associations between infected host molecular pathways and virus strains contribute to the lack of specific treatment or preventive measures. For a number of newly emerged viruses

Viral Respiratory Tract Infections, Table 1 Respiratory tract infection–associated viruses

Virus group	Virus family	Human (H) representative species/strains
(–)ssRNA	<i>Paramyxoviridae</i>	Respiratory syncytial virus (HRSV)
		Parainfluenza virus (HPIV)
		Metapneumovirus (HMPV)
	<i>Orthomyxoviridae</i>	Influenza A, B and C (A/H1N1 2009 pandemic)
(+)ssRNA	<i>Picornaviridae</i>	Rhinovirus A, B, and C
	<i>Coronaviridae</i>	Severe acute respiratory syndrome (SARS) coronavirus (CoV); CoV NL63 and HKU1
ssDNA	<i>Parvoviridae</i>	Bocavirus (HBoV)
dsDNA	<i>Adenoviridae</i>	Adenovirus (HAV)

(–)ssRNA: negative sense single-stranded RNA; (+)ssRNA: positive sense single-stranded RNA; ssDNA: single-stranded DNA; dsDNA: double-stranded DNA

or strains, for example, human bocavirus (HBoV), human coronavirus (HCoV) NL63 (Sloots et al. 2008), and HRV C strains (Lee et al. 2007), the pathology and global epidemiology has yet to be established.

Characteristics

Human Rhinovirus (HRV)

HRV, originally named JH virus (Johns Hopkins virus) was discovered in 1956. Since then 101 serotypes of HRV A and B groups have been identified which are still circulating. More recently, PCR of genomic regions and their phylogenetic analyses revealed a new group of HRV viruses, HRV C and 26 new strains among known HRV groups. Globally, the extent of HRV diversity is probably far greater. Unlike influenza A, the HRV diversity does not seem to be the result of antigenic drift (► [Antigenic Drift and Shift](#)) (Lee et al. 2007).

HRV Infections and Disease

HRV infections frequently affect both upper and lower respiratory tract causing cold and flu-like illness with symptoms ranging from sore throat, cough, rhinitis, to sinusitis. Asthma symptoms were reported in up to 25% of children infected with RV, worldwide (Arden and Mackay 2009). On average HRV infects adults 2–4 times and children 4–10 times per year (Yoon et al. 1999). In immunocompetent individuals HRV infections proceed relatively mildly, requiring only physician consultation. If HRV infects immunocompromised and asthmatic adults, it affects the lower respiratory tract, necessitating hospitalization. The induction of ► [pro-inflammatory mediators](#) and mediators of lung epithelial pathways in vitro and in vivo

including IL6, IL4, IL13, IL1RN, IL8, CXCL5, CCL10, CCL5, CXCL10, CSF2, ICAM1, MUC5AC4, and VCAMs, among others, can severely exacerbate existing illness conditions (Edwards et al. 2006; Bartlett et al. 2008). It is the host response to HRV infection rather than the HRV cytopathicity that causes CLFIs and asthma exacerbations. Therefore therapeutic approaches target the suppression of inflammation in combination with antiviral compounds.

Therapeutic Approaches

Application of general antiviral mediators IFNB1 and/or INFA2 in experimental HRV infections including asthmatic conditions showed mixed results due to strong side effects. Currently available antiviral compounds target capsid proteins (pirodivir) or viral replication (pleconaril and ribavirin) and reduce symptoms (Nichols et al. 2008). However pleconaril is ineffective against HRV4, HRV5, HRV42, HRV84, HRV93, HRV97, and HRV99, whereas ribavirin does not possess antiviral activity against HRV15, HRV3, and HRV5 (Choi et al. 2010). Attempts to identify more effective compounds led to 7-O-d-glucosid of banaba leaves (*Lagerstroemia speciosa* L.), which are used in the Philippines as tea to treat diabetes. 7-O-d-glucosid had in vitro antiviral activity for HRV1B, HRV2, HRV40, HRV6, and HRV14, including pleconaril- and ribavirin-resistant strains HRV3 and HRV5 (Choi et al. 2010).

Nasal sprays or lozenges containing Zn²⁺, which binds to ICAM1, the entry receptor for 90% of RV viruses showed no or mixed antiviral activity (Nichols et al. 2008). On the other hand inhalation of soluble ICAM1 (Tremacamra) reduced CFLI symptoms and

HRV 39 replication in an experimental infection model (Edwards et al. 2006).

Glucosteroids, whose anti-inflammatory effect is mediated via the glucocorticoid receptor signaling pathway, did not show sufficient amelioration of HRV-induced asthma exacerbation. In the light of a recent study using mice that were first infected (challenged) with HRV (allergen) and then challenged (infected) with allergen (HRV), the mixed outcome is unsurprising. HRV-infected mice challenged with allergen showed an augmented helper T (Th) cell 1 response (► [Th1 Response](#)) (e.g., INFG) whereas the response of allergen-challenged mice infected with HRV was skewed toward ► [Th2 response](#) cytokines IL4 and IL13 (Bartlett et al. 2008). Development of efficient and specific anti-inflammatory therapeutics will require further investigations into the cytokine-driven Th2 and Th1 polarization process in lung epithelial cells.

NFKB1 (NF-kappaB-p50) and RELA (p65) are required to induce transcription of pro-inflammatory cytokines (► [Pro-inflammatory Mediators](#)) in lung epithelial cells. Thus, inhibition of NF-kappa B signaling appears to be a promising anti-inflammatory therapeutic approach for HRV infections. Yet, further studies are necessary to clarify the extent and role of alternative activation of NFKB1 and RELA via IKBKE, TBK1, and TANK instead of CHUCK/IKBKB/IKBKG, and NF-kappa B mediated activation of INFB1, which has antiviral activity in context of HRV infections (Edwards et al. 2006).

Perspectives

The high serotype diversity and extent of newly discovered strains render HRV an unlikely candidate to develop a preventive vaccine. Moreover, frequently occurring mild infections of immunocompetent hosts are considered important to develop a broad immunity against RV strains.

A global disease burden study predicted that in 2020, chronic obstructive pulmonary disease (COPD) and lower respiratory tract infections will rank third and fourth among death causes (Murray and Lopez 1997). Given the contribution of HRV in exacerbations of COPD and asthma, new strategies in the development of specific anti-inflammatory and antiviral strategies are needed. These include (1) genome sequencing of all strains, (2) establishing cell lines that allow propagation of all strains to fulfill ► [Koch's postulates](#), (3) determining associations

between strains, symptoms, and epidemiology, and (4) differences in molecular pathways of host responses.

Cross-References

- [Antigenic Drift and Shift](#)
- [Koch's Postulates](#)
- [Pro-inflammatory Mediators](#)
- [Respiratory Tract, Upper and Lower](#)
- [Th1 Response](#)
- [Th2 Response](#)

References

- Arden KE, Mackay IM (2009) Human rhinoviruses: coming in from the cold. *Genome Med* 1(4):44
- Bartlett NW, Walton RP, Edwards MR, Aniscenko J, Caramori G, Zhu J, Glanville N, Choy KJ, Jourdan P, Burnet J, Tuthill TJ, Pedrick MS, Hurle MJ, Plumpton C, Sharp NA, Bussell JN, Swallow DM, Schwarze J, Guy B, Almond JW, Jeffery PK, Lloyd CM, Papi A, Killington RA, Rowlands DJ, Blair ED, Clarke NJ, Johnston SL (2008) Mouse models of rhinovirus-induced disease and exacerbation of allergic airway inflammation. *Nat Med* 14(2):199–204
- Choi HJ, Bae EY, Song JH, Baek SH, Kwon DH (2010) Inhibitory effects of orobol 7-O-D-glucoside from banaba (*Lagerstroemia speciosa* L.) on human rhinoviruses replication. *Lett Appl Microbiol* 51(1):1–5
- Edwards MR, Kebabdz T, Johnson MW, Johnston SL (2006) New treatment regimes for virus-induced exacerbations of asthma. *Pulm Pharmacol Ther* 19(5):320–334
- Fendrick AM, Monto AS, Nightengale B, Sarnes M (2003) The economic burden of non-influenza-related viral respiratory tract infection in the United States. *Arch Intern Med* 163(4):487–494
- Lee WM, Kiesner C, Pappas T, Lee I, Grindle K, Jartti T, Jakiela B, Lemanske RF Jr, Shult PA, Gern JE (2007) A diverse group of previously unrecognized human rhinoviruses are common causes of respiratory illnesses in infants. *PLoS ONE* 2(10):e966
- Murray CJ, Lopez AD (1997) Alternative projections of mortality and disability by cause 1990–2020: global burden of disease study. *Lancet* 349(9064):1498–1504
- Nichols WG, Peck Campbell AJ, Boeckh M (2008) Respiratory viruses other than influenza virus: impact and therapeutic advances. *Clin Microbiol Rev* 21(2):274–290
- Sloots TP, Whitley DM, Lambert SB, Nissen MD (2008) Emerging respiratory agents: new viruses for old diseases? *J Clin Virol* 42(3):233–243
- Yoon HJ, Zhu Z, Gwaltney JM Jr, Elias JA (1999) Rhinovirus regulation of IL-1 receptor antagonist in vivo and in vitro: a potential mechanism of symptom resolution. *J Immunol* 162(12):7461–7469

Virtual Cell (VCell) Modeling and Analysis Platform

Ion Moraru

R.D. Berlin Center for Cell Analysis and Modeling,
University of Connecticut Health Center, Farmington,
CT, USA

Definition

The Virtual Cell (VCell) is a unique computational environment for modeling and simulation in cell biology. It has been specifically designed to be a tool for a wide range of scientists, from experimental biologists to theoretical biophysicists. The creation of models can range from the simple to evaluate hypotheses or to interpret experimental data, to the complex, where multilayered models can be used to probe the predicted behavior of large, highly nonlinear systems.

VCell is deployed as a distributed application that was designed to be primarily used over the Internet, with a rich-client Java interface and a cluster-based server-side hosted at the University of Connecticut Center for Cell Analysis and Modeling. However, comprehensive stand-alone capabilities have been introduced starting with VCell 5. Distinct biological and mathematical frameworks are encompassed within a single graphical interface. Users can build complex models which specify compartmental topology and geometry, molecular characteristics, and relevant interaction parameters. VCell automatically converts the biological description into a corresponding mathematical system of ordinary and/or partial differential equations. The mathematically savvy user may directly specify the complete equation-based description of the model, bypassing the schematic interface. Either way, VCell will then solve the equations by applying numerical solvers and generate appropriate software code to perform and analyze simulations. Results can be displayed and analyzed online or downloaded to the user's computer in a variety of formats.

VCell is freely accessible to all members of the scientific community; the latest version can be found at <http://vcell.org/>. All VCell proprietary software code is currently developed as Open Source and is

made available under the MIT license at <http://sourceforge.net/projects/vcell/>.

Characteristics

The VCell Database

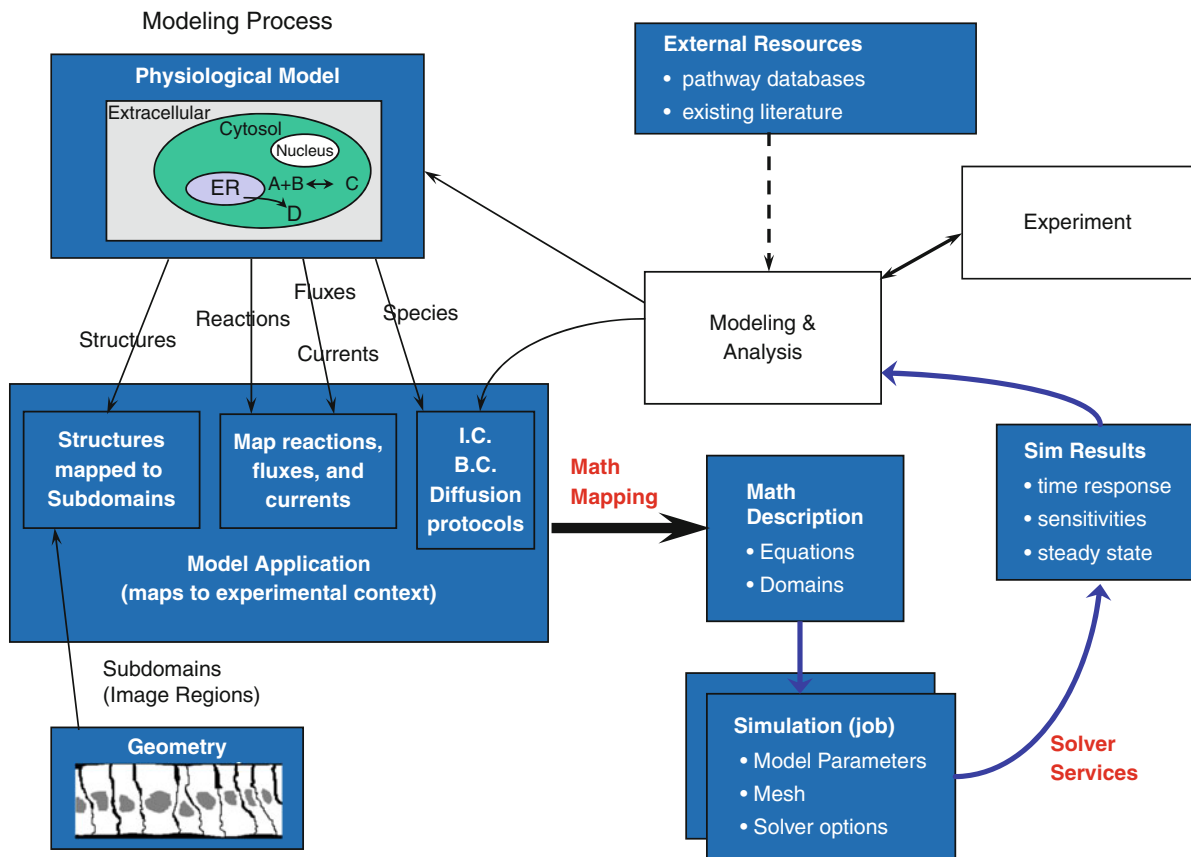
The distributed version of VCell maintains a database of models developed by users. A model can be made private (accessible exclusively to the owner; this is the default), public (available for viewing and copying to all VCell users), or can be shared with a customizable list of users (specified by the model owner). The model database serves several important features. First, it provides backward compatibility, by allowing automated machine testing whenever a new version of the VCell is released. Second, the possibility to flexibly share models provides a unique environment for collaborative research or for educational purposes (courses, workshops, etc.). Third, models can be made public such that everyone can read and create copies of them, and additionally, they can be “frozen” into an immutable version for persistent archival (e.g., to guarantee future reproducibility of published work).

VCell has achieved a large following among biologists doing computational modeling: As of November 2011, more than 3,000 worldwide users have run simulations on VCell's servers. The VCell database contained more than 30,000 models and more than 200,000 simulation results, of which more than 600 models were made publicly available.

The Modeling Paradigm in VCell

The typical modeling process carried out in VCell is illustrated in [Fig. 1](#). It begins with the physiological model being defined in terms of species, structures, reactions, fluxes, and currents. Note, not all models must contain all terms; it will depend on the complexity of the model.

The model is associated with one or more Application components, which allow users to separate context-specific assumptions from the rest of the model. For example, this can involve mapping topological structures to specific subdomains (image regions defined in analytical or experimental geometries), defining initial conditions, boundary conditions, fast reaction approximations, diffusion parameters, electrophysiological protocols, events, etc.



Virtual Cell (VCell) Modeling and Analysis Platform, Fig. 1 The VCell BioModel consists of one *Physiological Model* that provides the conceptual representation of the model elements: structures, molecules, connectivity map, kinetics. It includes one or more *Model Applications* – analogous to “virtual experiments” that can be simulated: initial conditions, actual morphologies, electrical protocols, etc. An Application can be compartmental or it can use a *Geometry*. Each Application has

a mathematical representation (*Math Description*) that is automatically generated and can be viewed in VCMDL format or through an equation browser. For each Application, a user can specify one or more *Simulations* (time length, resolution, solvers to use, parameter overrides, etc.) that will run and produce *Simulation Results*; these can be viewed in VCell, in a customized version of VisIt, or exported to the user’s computer in a variety of formats (raw data, spreadsheets, movies, images)

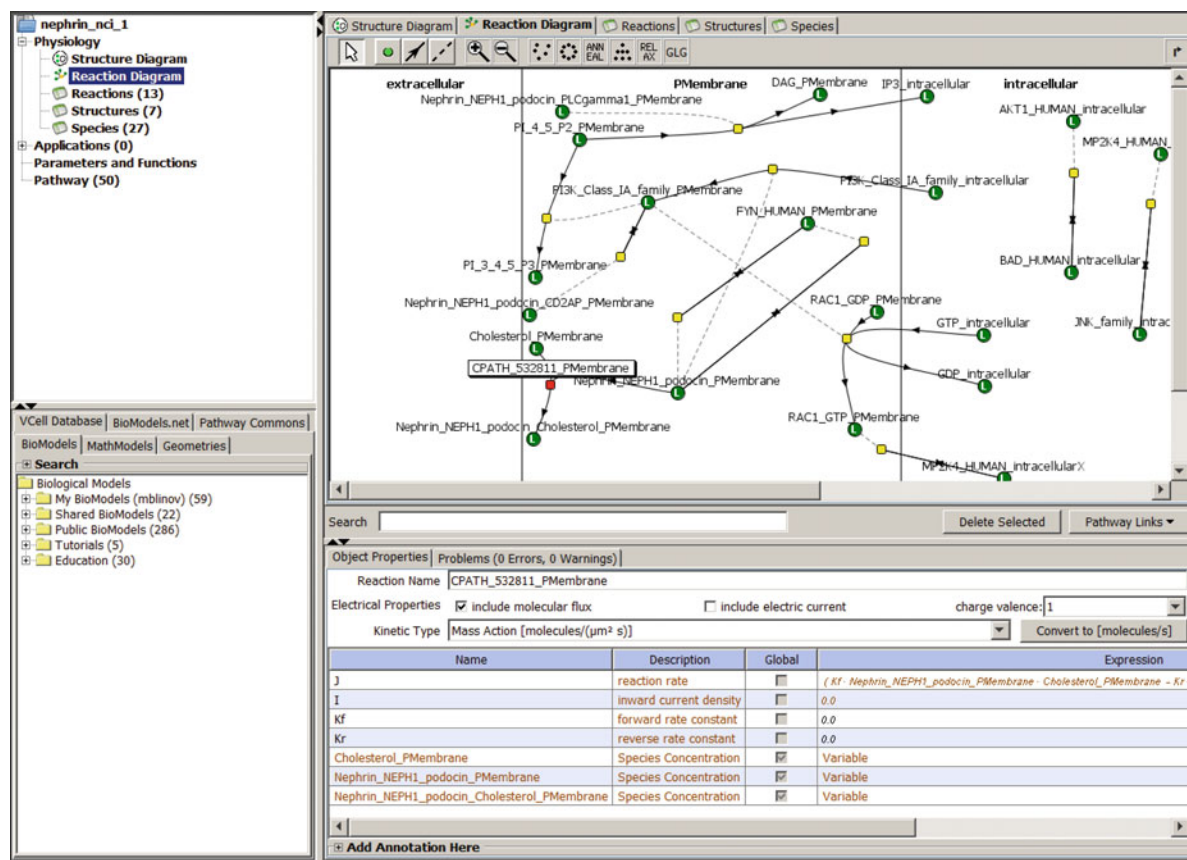
Once an application is created, a detailed math description of the model and application is automatically and transparently generated by the software. One can then run a simulation which will include additional model details such as model parameters, mesh size, and solver options. The simulation results can provide information regarding time response, sensitivities, and steady state. Such results can be used for analysis, further modeling, and for refining the experimental protocol.

VCell Modeling Features

A user can specify many features of biological systems necessary for efficient predictive modeling. Specifically, a user may specify location and nesting

of intracellular compartments, including compartmental volumes and membrane sizes. For species, a user can specify concentrations and/or molecule counts, as well as 3D (volumetric) diffusion, 2D (surface)diffusion, and advection (flow/velocity). Bimolecular or multimolecular interactions can be specified in each compartment, as well as on the boundary between compartments (membrane reactions and transmembrane transport). At compartment boundaries, either molecular flux or electrical current, or both, can be specified, as well as membrane potential.

The VCell user interface provides a possibility to specify reaction network graphically (through drag and



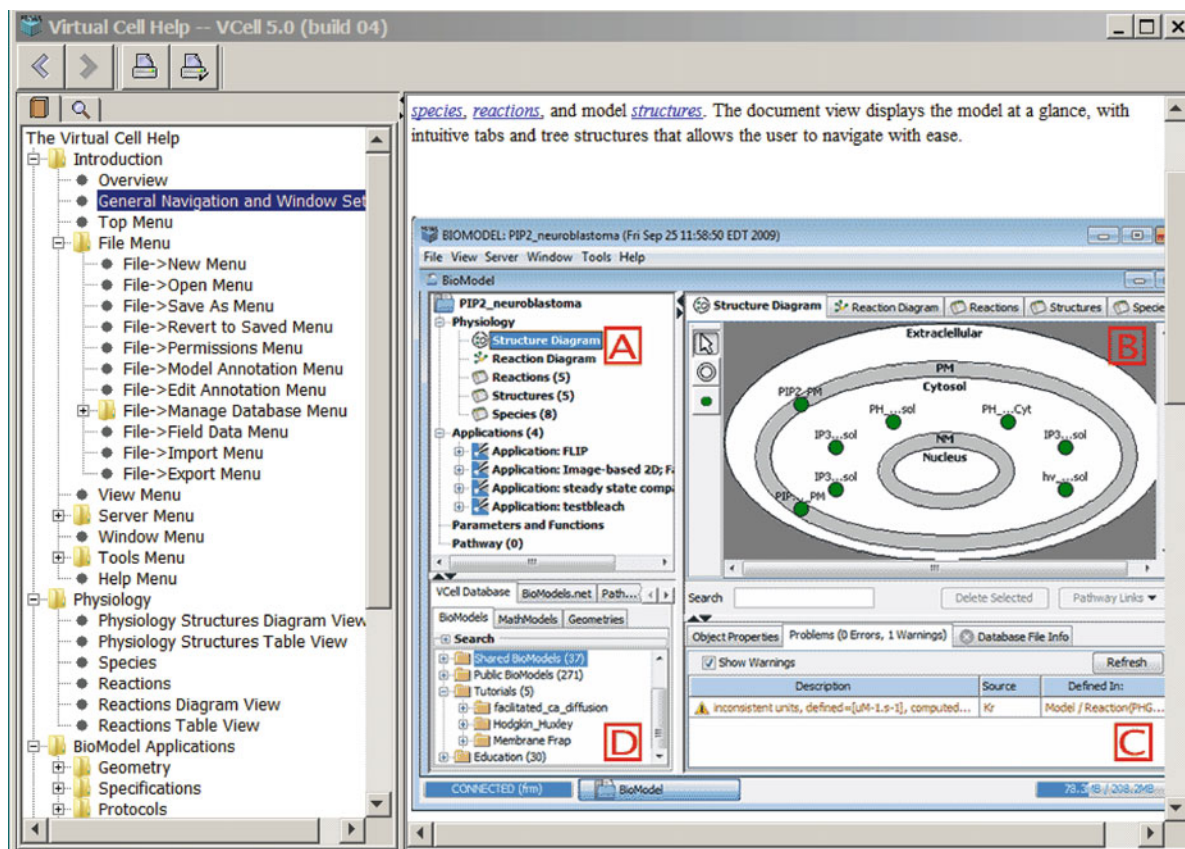
Virtual Cell (VCell) Modeling and Analysis Platform, Fig. 2 The Physiology Model includes reaction network representation as a diagram (shown) or as lists of reactions and species. Each reaction and species has a list of editable properties displayed below. A user can get access to every model

element and multiple model views from the *top left panel*, and to public modeling resources (VCell database, BioModels database, Pathway Commons databases) from the *bottom left panel*

drop graphical interface), as a list of species and reactions (through a tabular interface allowing copy-paste of multiple objects), as shown in Fig. 2. In addition, VCell has many convenience features throughout, such as word and mathematical expression autocomplete, extensive search options, sort/filter-capable tables, etc., as well as a searchable online Help window describing all details of the VCell interface and modeling process (Fig. 3).

VCell has been originally designed to work with spatially resolved models, and it has a wide range of capabilities to specify 2D or 3D geometries for compartments. A user is able to create geometries through reusing previously defined geometries from other VCell models (including shared models from other users).

New analytic shapes can be created by defining analytic equations that describe compartment distributions in space or by manually defining compartments on a regular grid of points in space (image painting). An important VCell feature is a possibility to use geometries from external resources. Rasterized 2D or 3D (Z-stacks) images can be imported, as well as CAD specifications such as stereolithography and constructive solid geometry. In a pre-segmented image, compartments can be defined as areas of different pixel values. VCell provides a set of image processing functions to segment a non-segmented image to define compartments. Additionally, surface mesh files can be imported and sampled within a spatial domain to define



Virtual Cell (VCell) Modeling and Analysis Platform, Fig. 3 The online help window of VCell provides description of all windows, commands, and messages of VCell

compartments. Geometrically defined compartments can be viewed as individual image planes in the x,y,z dimension or as rendered surface boundaries between compartment regions.

Rule-based modeling is also available in VCell via the BioNetGen@VCell service – an interface to the BioNetGen rule-based modeling software that provides a possibility to specify and run rule-based models, as well as create equivalent flattened VCell BioModels.

A user is able to import into VCell models (or parts thereof) that have been created in other software packages that have been specified (or converted) in SBML or CellML formats. Additionally, embedded links are provided to models in the BioModels database with an option to import them directly into VCell. Conversely, models created in the VCell can be used by many tools, as they can be exported in SBML or MATLAB formats.

Last but not least, external resources such as pathway databases are being made available to the VCell software as a *Pathway Commons @VCell* component. For example, physical entities and interactions from nine public pathway databases accessible through Pathway Commons (such as Reactome, NCI PID, MetaCyc, and more) can be searched and imported into the Virtual Cell; model species and reactions can be associated with the corresponding entities from pathway databases (Fig. 4).

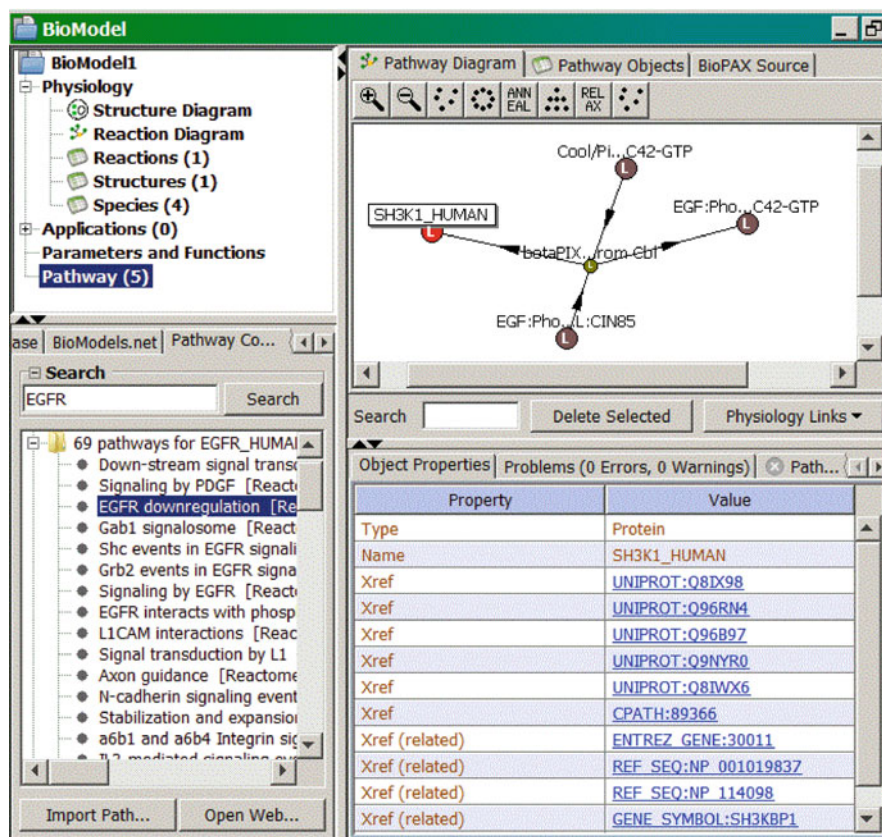
VCell Simulation Capabilities

VCell includes a wide range of simulation capabilities. For nonspatial models, a user has an option to use multiple deterministic (ordinary differential equations, ODE) or stochastic solvers. Available nonspatial deterministic solvers are the following:

- Forward Euler (First Order, Fixed Time Step)
- Runge-Kutta (Second Order, Fixed Time Step)
- Runge-Kutta (Fourth Order, Fixed Time Step)

Virtual Cell (VCell) Modeling and Analysis Platform, Fig. 4

After a search of Pathway Commons (bottom left panel), the entries of interest can be imported and stored alongside a VCell model as a Pathway Model. For each entry in a Pathway Model, the user has all the necessary information, including a list of identifiers from different databases with clickable web links that lead to relevant entries in any of these databases. The user is then able either to create a new VCell model based on the pathway data, or to link existing elements of a VCell model to entities in a Pathway Model. Linked species and reaction in the VCell model are marked with a letter L (as are the corresponding physical entities and interactions in the Pathway Model)



- Adams-Moulton (Fifth Order, Fixed Time Step)
- Runge-Kutta-Fehlberg (Fifth Order, Variable Time Step)
- IDA (Variable Order, Variable Time Step, ODE/DAE)
- CVODE (Variable Order, Variable Time Step)
- Combined stiff solver CVODE/IDA

Available nonspatial stochastic solvers are the following:

- Gibson (Next Reaction Stochastic Method)
- Hybrid (Gibson + Euler-Maruyama Method)
- Hybrid (Gibson + Milstein Method)
- Hybrid (Gibson + Milstein Method, Variable Time Step)

Also, for nonspatial simulations, VCell incorporates a special simulation option to perform parameter sensitivity analysis (for any variable and/or parameter). Additionally VCell integrates parameter estimation capabilities from COPASI (<http://www.copasi.org>), providing more than ten different algorithms to search for parameter values to fit externally provided data.

For spatial models, a user can use multiple deterministic (partial differential equations, PDE) or

stochastic solvers. Available spatial deterministic solvers are the following:

- Semi-Implicit Finite Volume, Regular Grid (Fixed Time Step)
- Semi-Implicit Finite Volume Compiled, Regular Grid (Fixed Time Step)
- Fully Implicit Finite Volume, Regular Grid (Variable Time Step)

Available spatial stochastic solvers are the following:

- Smoluchowski dynamics particle-based simulator (integrated Smoldyn solver – <http://www.smoldyn.org>)
- Semi-Implicit Finite Volume, Regular Grid (Fixed Time Step) – hybrid deterministic/stochastic

References

- Cowan AE, Moraru II, Schaff JC, Slepchenko BM, Loew LM (2012) Spatial modeling of cell signaling networks. *Methods Cell Biol* 110:195–221

Moraru II, Schaff JC, Loew LM (2006) Think simulation – think experiment: the virtual cell paradigm. In Perrone LF, Wieland FP, Liu J, Lawson BG, Nicol DM, Fujimoto R M (eds) Proceedings of the 2006 winter simulation conference, pp 1713–1719

Moraru II, Schaff JC, Slepchenko BM, Blinov M, Morgan F, Lakshminarayana A, Gao F, Li Y, Loew LM (2008) The virtual cell modeling and simulation software environment. IET Syst Biol 2:352–362

Slepchenko BM, Loew LM (2010) Use of virtual cell in studies of cellular dynamics. Int Rev Cell Mol Biol 283:1–5.6

Virtual Laboratories

- [Collaborative and Distributed Biomedical Applications](#)

Virtual Plant Modeling

- [Functional-Structural Plant Modeling](#)

Virus Reference Genomes

- [NCBI Viral Genomes Resources](#)

Visual Processing Unit (VPU)

- [GPU](#)

Visualization of Disease Networks

Sanjeev Kumar¹ and Shipra Agrawal²

¹BioCOS Life Sciences Private Limited, Bangalore, Karnataka, India

²BioCOS Life Sciences Private Limited, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India

Definition

Network Visualization

With huge amount of biological data, network analysis of the data has become almost inevitable. Complex scientific data can be conveniently represented in the

form of graphical models such as graphs having nodes and edges forming networks. The vital information can be easily visualized and hence be extracted, which otherwise may lay hidden in such large and complex data.

Visualization of Disease Networks

Network visualization has enabled scientists to hypothesize the causal mechanisms behind the diseases more easily and extract hidden information from the entities (nodes) and the relations between them (edges). It can facilitate analysis of disease-related molecular data in multiple dimensions as the data could be visualized simultaneously, which can further assist in decision making.

The disease-oriented networks could be described through the proper visualization and topologically significant network diagrams. The application of network visualization toward the understanding of disease biology from the disease networks could be exemplified from following two examples:

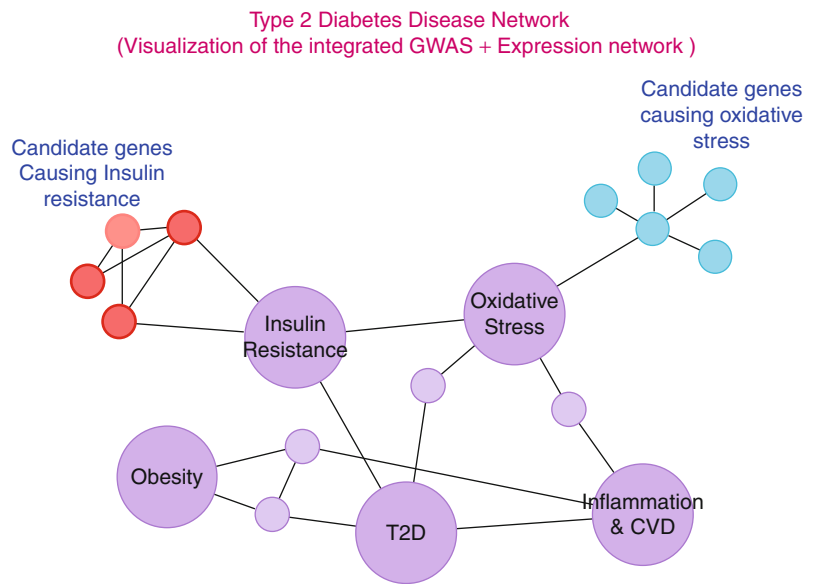
- ► [Co-expression networks](#) built from the differentially expressed genes of a diseased system can facilitate identification of the causal genes and their functional relations through a statistically significant network topology diagram. For example, the ► [hub](#) genes signify highly connected and centrally important genes and modules (► [Module Network](#)) infer the group of highly connected and functionally correlated genes (Sieberts and Schadt 2007).
- An integrated visualization of gene ► [co-expression](#) and genetic association networks facilitates an insight on epistatic mechanisms or measure allelic effects on expression molecular mechanisms through networks derived from case control data of various diseases (Azuaje and Dopaz 2005).

Tools for Visualizing Disease Networks

Several network visualizer tools are available in open source domain. Functionality of these tools is based on graph applications, which map the highthroughput biological data as 2D graphs to visualize biological interactions at molecular level. It facilitates analyses of protein-protein interactions, dynamic simulations on pathway data, gene networks, integrated molecular networks, and analysis of evolutionary relationships between proteins involved in molecular characterization of a disease. Among these, the Cytoscape is freely

Visualization of Disease Networks,

Fig. 1 Visualization of a conceptual Type 2 diabetes (T2D) disease network: The *orange* and *blue* nodes (genes) correspond to known causal genes for insulin resistance and oxidative stress, respectively. The *larger purple* metanodes could be used to describe the resultant pathological conditions and progressive complications of T2D



available standalone software; it performs visualization of the molecular interaction networks and integration of multiple types of molecular data (<http://www.cytoscape.org/>). It allows the user to identify functionally and statistically significant disease network modules. Another highly interactive application, Gene networks is suitable to analyze large-scale disease networks. This tool has the capability to integrate multiple types of molecular networks. BioLayout Express3D (<http://www.bioblayout.org/>) is a tool for visualization and clustering of large-scale networks in both 3D and 2D. It supports both unweighted and weighted graphs together with edge annotation of pairwise relationships. This tool is very useful for analyzing co-expression disease networks. **Figure 1**

describes a sample visualization of Type 2 diabetes (T2D) gene network and its pathophysiology through obesity risk, insulin resistance, and inflammation. The visualization describes progression of T2D into cardiovascular complications.

References

- Azuaje F, Dopaz J (2005) Data analysis and visualization in genomics and proteomics. Wiley, Hoboken, pp 1–257. ISBN 0470094397
- BioLayout Express3D (<http://www.bioblayout.org/>)
- Cytoscape (<http://www.cytoscape.org/>)
- Sieberts SK, Schadt EE (2007) Moving toward a system genetics view of disease. *Mamm Genome* 18:389–401