

A community web-based resource and  
repository for pathway research:  
[pathwaydatabase.org](http://pathwaydatabase.org)

Semen A. Trygubenko\*, Tetyana V. Bogdan and David J. Wales

*University Chemical Laboratories, Lensfield Road,*

*Cambridge CB2 1EW, UK*

December 11, 2006

**Abstract**

We introduce a new platform for accelerating communication between researchers who are interested in locating pathways with special properties. Pathwaydatabase.org is a web repository of paths and associated metadata that supports a variety of pathway representations and enables efficient searching, classification and deposition of path-

---

\*E-mail: [sat39@cam.ac.uk](mailto:sat39@cam.ac.uk)

way data, thereby promoting pathway research and facilitating the dissemination of knowledge in this interdisciplinary field.

## 1 Introduction

The notion of a pathway (or path) is frequently used to describe dynamic and kinetic information of a process in fields such as chemical physics, molecular biology, genomics and topology. For example, in physiology a pathway could be defined as a fixed route formed by a chain of nerve cells along which impulses of a particular kind usually travel. In biochemistry it is a sequence of biochemical reactions undergone by a compound in a living organism as part of its metabolism. In molecular modelling a path on the potential energy surface can be defined as a sequence of transition states and intervening local minima that connect two given endpoints.<sup>1</sup>

Although such definitions may seem dissimilar, an appropriate coarse-graining of the pathway may allow us to use common language and employ common methodologies to problems arising in different fields. For example, metabolic pathways represented as directed graphs, with the substrates, products, and enzymes as nodes, and the chemical reactions catalysed by the enzymes as edges, can be analysed by a relabelling graph transformation rule.<sup>2</sup> Pathways on the potential energy surface can also be naturally

described in a graph representation to which another graph transformation method can be applied to extract kinetic properties of the reaction from the sets of stationary points that comprise the pathway.<sup>3</sup>

According to the Oxford English dictionary, the concept of a pathway was originally introduced in scientific context around 1905. Since then a vast amount of pathway data has been accumulated both experimentally and computationally. With rapid development of systems biology in recent years hundreds of kinetic, signalling, metabolic and biochemical pathway databases each specialising in a small class of molecules have been established. Examples of such databases include the Kinase Database,<sup>4</sup> the Systems Biology Database<sup>5</sup> and the Encyclopedia of Metabolic Pathways,<sup>6</sup> to name just a few.

The pathway resource list maintained by Badger, Cary and Sanders<sup>7,8</sup> predicts there are more than 190 web-accessible biological pathway and network databases. Each year the first issue of Oxford's Nucleic Acids Research journal is devoted to biological databases, and an issue in July is devoted to papers describing web-based software resources of value to the biological community.<sup>9</sup> Michael Galperin maintains the Molecular Biology Database Collection, a compendium of databases that includes all those databases described in Nucleic Acids Research, as well as selected other databases relevant to biologists.<sup>10</sup> 139 new databases were included into 2006 update, totalling to considerable 858.

In the fields of computational chemistry and physics similar pathway repositories are not yet available. However, other databases are emerging, with one of the first successful projects being BioSimGrid, which provides infrastructure for deposition and analysis of molecular dynamics trajectories in protein simulations.<sup>7</sup> It has been our experience that there is a general need for data integration via a facility that would provide access and searching capabilities to data and information types involved with pathway generation and research. We thus propose to establish [pathwaydatabase.org](http://pathwaydatabase.org) as a scientific content management system, where data and algorithms can be deposited and analysed by the pathway research community in a convenient online collaborative environment.

The main objective of the project is developing the facilities to share, unify and disseminate the knowledge about pathways relevant to computational chemistry and physics communities, and about methods that can locate paths with special properties, in a standardised machine-searchable form. The site also aims to provide access to an easy-to-use resource of tests and benchmarks to assist comparison and evaluation of computational techniques that are used to locate paths. We thus hope that the portal will accelerate communication between scientists by providing new forms of online communication, e.g. forum-like discussions, history of improvements on previously deposited pathways, multiviews and statistical information about

the pathway deposits.

In the following sections we describe the data content and technical implementation of the resource, and provide examples of data that has been posted so far.

## 2 Pathway database

The technical infrastructure of the web resource allows any registered member to deposit data related to pathway research. Examples of data that can be deposited include pathways themselves and the metadata that describes the pathways, such as the system for which the pathway was found, any special property this pathway possesses, or the technique used in locating this pathway, along with benchmark data. The history of depositions is preserved to produce statistics about the database and can be retrieved if necessary. The list of most recent changes is displayed in a dedicated portlet.

The resource supports a wiki-like information management suite and a forum for submission and organisation of text and graphical information to facilitate discussions. News and events sections, which can be formed into customisable RSS feeds, are available and can be edited by any member.

## 2.1 Implementation

Currently the portal is driven by the Zope<sup>8</sup> and Apache<sup>9</sup> web application servers with a PostgreSQL<sup>10</sup> database backend. GUI and content-management are powered by Plone,<sup>11</sup> with Ploneboard communication product providing forum-like capabilities. Incremental backups are performed daily with rdiff-backup.<sup>12</sup> Subversion is used for source version control.<sup>13</sup> The underlying OS platform is Debian/GNU Linux.<sup>14</sup>

## 2.2 Content

The resource can be immediately used for identifying the pathways that have been found before for a system of interest, and also if one wants to improve on existing techniques for locating the paths. Furthermore, pathways with particular properties, such as the shortest or the fastest which are found by different methods, can be compared.

Currently the database includes shortest and fastest paths between ‘important’ regions of potential energy surfaces of a number of clusters bound by the Lennard-Jones potential. Representative pathways for a variety of model proteins are in the process of being deposited.

## 2.3 Future development

Efforts are underway to integrate numerical tools for quick calculation of pathway properties such as cooperativity or symmetry. We also want to make extensive use of the plugin capabilities and standard graphics libraries to provide embedded pathway visualisation and animation.

Pathwaydatabase.org is also expected to incorporate semantic web technologies such as Resource Description Framework (RDF) and extensible markup language (XML) and make use of emerging standard libraries and parsers.<sup>?,15,16</sup>

## 3 Conclusions

The problem of finding and understanding pathways occupies the minds of researchers from various branches of science. The large number of possible applications and lack of communication between scientists can potentially lead to knowledge rediscovery and duplication of scientific effort in the future.

We attempt to unify pathway knowledge by introducing pathwaydatabase.org, a means for sharing pathway data in a form that allows efficient deposition, searching and manipulation, yet flexible enough to accommodate a variety of formats and representations.

We expect this web repository to evolve as new data is added and to ex-

pand its functionality to meet new demands. We hope pathwaydatabase.org will be of benefit for all scientists from physical and chemical community involved in pathway research.

## 4 Acknowledgements

S.A.T. is grateful to Darwin College for the financial support. S.A.T. and T.V.B. are thankful to EPSRC for the financial support.

## References

- [1] D. J. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, Cambridge University Press, Cambridge (2003).
- [2] A. Corradini, H. Ehrig, U. Montanari, L. Ribeiro and G. Rozenberg (eds.), *Graph Transformations, Third International Conference, ICGT 2006, Natal, Rio Grande do Norte, Brazil, September 17-23, 2006, Proceedings*, vol. 4178 of *Lecture Notes in Computer Science*. Springer (2006).
- [3] S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.* **124**, 234110 (2006).
- [4] *Kinase Pathway Database*, <http://kinasedb.ontology.ims.u-tokyo.ac.jp> (2004).

- [5] *Systems Biology Pathway Database*, <http://www.proteinlounge.com/> (2006).
- [6] *MetaCyc*, <http://metacyc.org/> (2005).
- [7] B. Wu, K. Tsai, S. Murdock, M. Ng, S. Johnston, H. Fangohr, P. Jeffreys, S. Cox, J. Essex and M. Sansom, Proceedings of UK e-Science All Hands Meeting (2003).
- [8] *Zope web application server*, <http://zope.org> (2006).
- [9] *The Apache Software Foundation*, <http://www.apache.org> (2006).
- [10] *PostgreSQL: The world's most advanced open source database*, <http://www.postgresql.org> (2006).
- [11] *Plone content management system*, <http://plone.org> (2006).
- [12] *rdiff-backup*, <http://www.nongnu.org/rdiff-backup/> (2006).
- [13] *SVN, a version control system*, [subversion.tigris.org](http://subversion.tigris.org) (2006).
- [14] *Debian — The Universal Operating System*, <http://www.debian.org/> (2006).
- [15] W. W. W. Consortium, *Semantic Web*, <http://www.w3.org/2001/sw/> (2006).

- [16] N. Shadbolt, T. Berners-Lee and W. Hall, *IEEE Intelligent Systems* **21**, 96 (2006).