Beilstein-Institut

ESCEC, March 19th – 23rd, 2006, Rüdesheim/Rhein, Germany

The Estimation of Kinetic Parameters in Systems Biology by Comparing Molecular Interaction Fields of Enzymes

Matthias Stein, Razif R. Gabdoulline, Bruno Besson, Rebecca C. Wade

EML Research gGmbH, Molecular and Cellular Modeling Group, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany

E-Mail: matthias.stein@eml-r.villa-bosch.de

Received: 16th June 2006 / Published: 31st August 2007

Abstract

The kinetic modelling of biochemical pathways requires a consistent set of enzymatic kinetic parameters. We report results from software development to assist the user in systems biology, allowing the retrieval of heterogeneous protein sequence, structural and kinetic data. For the simulation of biological networks, missing enzymatic kinetic parameters can be calculated using a similarity analysis of the enzymes' molecular interaction fields. The quantitative PIPSA (qPIPSA) methodology relates changes in the molecular interaction fields of the enzymes with variations in the enzymatic rate constants or binding affinities. As an illustrative example, this approach is used to predict kinetic parameters for glucokinases from *Escherichia coli* based on experimental values for a test set of enzymes. The best correlation of the electrostatic potentials with kinetic parameters is found for the open form of the glucokinases. The similarity analysis was extended to a large set of glucokinases from various organisms.

INTRODUCTION

One of the aims of systems biology is to provide a mathematical description of metabolic or signalling protein networks. This can be achieved by constructing a set of differential equations describing changes in concentrations of compounds with time [1]. Enzyme-specific parameters, such as ligand binding affinity and catalytic turnover, are needed for solving these equations. These parameters need to be valid under the desired experimental conditions. Despite recent developments in enzymatic high-throughput assays, experimental values of many of the required parameters often are not available for the chosen organism or enzyme, or have not been determined at the desired temperature or pH [2].

For the construction of a kinetic model, it is essential to have a consistent and reliable set of enzymatic kinetic parameters. The importance of the uniformity of the measurement and reporting of enzymatic functional data has been emphasized in [3].

Molecular systems biology deals with the intrinsic molecular interactions and enzymatic reaction mechanisms of each enzyme involved in the systems biology network [4]. The generation of quantitative structure–function relationships which relate the enzyme's activity to molecular interactions between the substrate molecules and critical components of the enzyme represents one of the challenges of modern enzymology [5].

The SYCAMORE (**SY**stems biology's **C**omputational **A**nalysis and **MO**deling **R**esearch **E**nvironment) is being developed as part of the German systems biology initiative "HepatoSys" [6] (Platform Bioinformatics and Modelling, Groups of Dr Ursula Kummer and Dr Rebecca Wade, EML Research) and aims at providing guidance to the user in setting up a biochemical kinetic model, running and analysing the results (see legend of Fig. 1 for details). When kinetic parameters are absent or inconsistent, structure-based modelling of the missing kinetic parameters is started.

PIPSA (Protein Interaction Property Similarity Analysis) is used as a means of comparing the molecular interaction fields of a test set of proteins and relating differences in enzymatic rate constants to variations in the electrostatic potentials exerted by the protein. The PIPSA methodology has been used previously to cluster different proteins according to the similarity of their electrostatic potentials. Applications include PH domains [7], E2 domains [8], triose phosphate isomerases [9], and Cu,Zn-superoxide dismutases [10]. We have extended the use of PIPSA to a more quantitative approach (qPIPSA) to relate the variations of the protein electrostatic potential within a family of enzymes to kinetic parameters.

The aim of this paper is to present an example of the application of the structure-based modelling module of the SYCAMORE project. We demonstrate the retrieval of heterogeneous protein structural and sequence information from distributed sources. The information on a protein from related organisms is then used to estimate the kinetic parameters for a corresponding protein from a different organism using the PIPSA methodology. This approach enables the user to detect inconsistent experimental values of kinetic parameters.

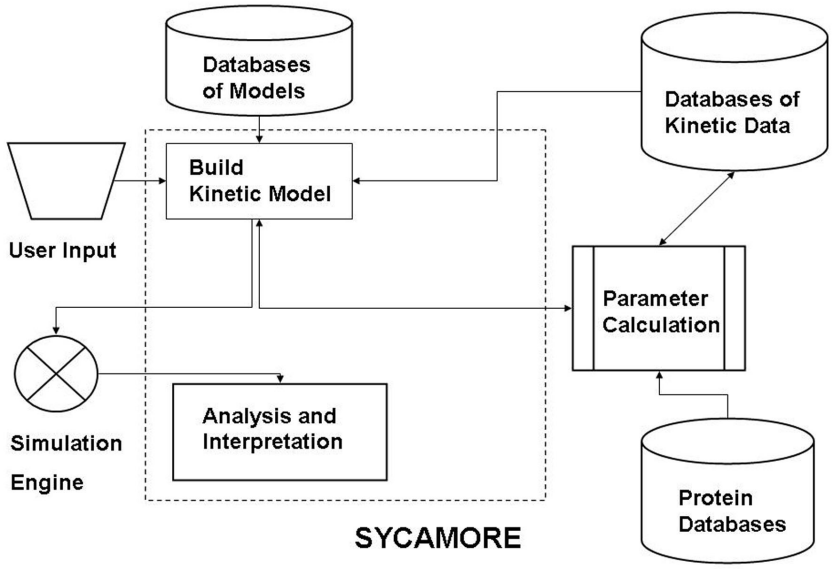


Figure 1. The SYCAMORE (SYstems biology's Computational Analysis and MOdeling Research Environment) assists the user in setting up and performing simulations in systems biology**.** The user can create a mathematical model by hand or use models from a depository such as Biomodels [43] or JWS online [44]**.** During the setup of the model**,** experimental kinetic parameters can be retrieved from BRENDA [14] or SABIO-RK[15]. When experimental parameters are not available for the desired organism but for a related organism or obtained under different environmental conditions**,** the modelling of these parameters from protein sequence and structural information can be initiated**.** The generated data then flow back into the kinetic model before the complete model is given to an external simulation engine (such as COPASI [45])**.** The final step is the analysis and interpretation of the results of the network modelling**.**

As a test case, we apply the method to the discrimination between mammalian and non-mammalian glucokinases and in particular to the assignment of a K_m value to the enzyme from *Escherichia coli*. The biochemistry and evolution of glucokinases has been reviewed in [11 – 13].

Methods

Retrieval of enzymatic structural and kinetic information

The structure-based modelling module within SYCAMORE is a link between the databases of experimental kinetic data, protein sequence and structure databases and the mathematical kinetic model (see Fig. 1). It is coded in Java as a server–client architecture and browser-based to allow for maximum portability and ease of accessibility.

This module is still under development. Currently the user can query the BRENDA [14] and SABIO-RK [15] databases for existing experimental kinetic parameters. Protein structural models can be retrieved from the Protein Data Bank (PDB) [16], theoretical models from ModBase [17] and from the Swiss-Model Repository [18]. Protein sequences are taken from the SwissProt/UniProt database [19].

The module uses servlets and core classes. The results pages are generated using Java Server Pages (JSP) which allow static HTML to be mixed with dynamically-generated HTML pages so that the generated web pages have a dynamic content. The result pages display in any web browser compliant with XHTML and ECMAscript (Javascript). The Systems Biology Standard Markup Language (SMBL) [20] was chosen as the file format standard to communicate between the various applications and modules.

The user has the opportunity to choose retrieved sequence, structural and kinetic data from the various sources and in the end to review his choice, modify parameters or insert user-generated alternative values.

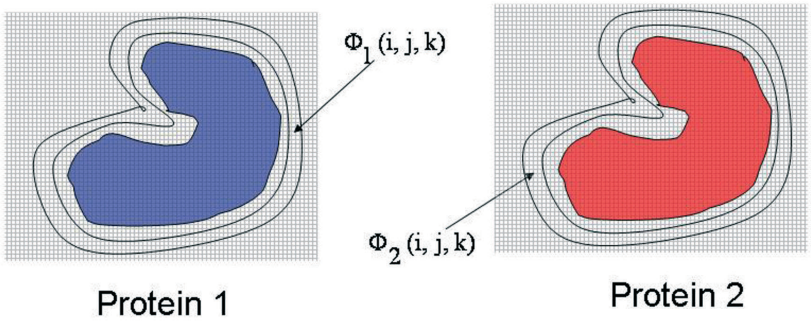
Protein interaction property similarity analysis

The structure-based systems biology calculations are performed by comparing molecular interaction fields such as the electrostatic potential or a hydrophobic field. The PIPSA method has been described elsewhere [7, 21].

The molecular interaction fields of proteins are compared on a three-dimensional grid over the superimposed proteins. The difference in the molecular interaction fields can be quantified by the calculation of similarity indices which were originally developed for the comparison of small molecules. The Hodgkin similarity index detects differences in sign, magnitude and spatial behaviour in the potential [22, 23].

Generation of protein models

Protein amino acid sequences were taken from the SwissProt database [19]. Multiple sequence alignment of amino acid sequences was performed using the program ClustalW [24]. Comparative protein structural modelling was done using Modeller 8v1 [25]. Polar hydrogens were added using the program WHATIF [26]. The OPLS non-bonded parameter set was used to assign partial atomic charges and radii. The electrostatic potentials were calculated with the program UHBD [27]. The linearized form of the Poisson–Boltzmann equation (LPBE) was solved using the Choleski preconditioned conjugate gradient method. An ionic strength of 50 mM, a grid dimension of $150 \times 150 \times 150$ Å^3 and a grid spacing of 1.0 Å was employed. The relative dielectric constant of the solvent was 78.0 and that of the solute was set to 4.0.



$$(\mathbf{p}_1, \mathbf{p}_2) = \sum_{i,j,k} \Phi_1(i,j,k)\Phi_2(i,j,k)$$

Hodgkin Similarity Index

$$\mathbf{SI}_{12} = \frac{2(\mathbf{p}_1, \mathbf{p}_2)}{(\mathbf{p}_1, \mathbf{p}_1) + (\mathbf{p}_2, \mathbf{p}_2)}$$

Figure 2. Calculation of molecular fields Φ_1 and Φ_2 on three dimensional cubic grids for two proteins and definition of the scalar product of the molecular interaction fields by summing over every grid point on a skin**.** The Hodgkin similarity index [22,23,46] is a measure of the pair-wise similarity of the molecular fields**.**

Results and Discussion

Here we give an illustrative example of the application of structure-based systems biology for the detection of inconsistent kinetic parameters and the generation of missing parameters for use in mathematical modelling of biochemical protein networks.

The conversion of chemical energy in the glycolytic (Emden–Meyerhof) pathway is one of the best investigated and understood metabolic pathways. The glucokinases (EC 2.7.1.2) catalyse the first chemical reaction in glycolysis. They phosphorylate glucose at the 6 position by abstracting a phosphate group from ATP. This yields glucose-6-phosphate and ADP. The virtually irreversible reaction is one of the control sites in glycolysis since the mammalian glucokinase is not product inhibited.

$$\text{Glucose} + \text{ATP} \rightarrow \text{Glucose-6-phosphate} + \text{ADP} \qquad (1)$$

We create here the scenario of a user wanting to model the glucokinase from *E. coli* by starting from knowledge about the enzyme in *Homo sapiens*.

Retrieval of protein information from distributed resources

In the structure-based estimation of kinetic parameters, the user is faced with the distribution of necessary data over various resources. The protein information retrieval module within SYCAMORE simplifies the accession to distributed protein sequence, structural and kinetic information.

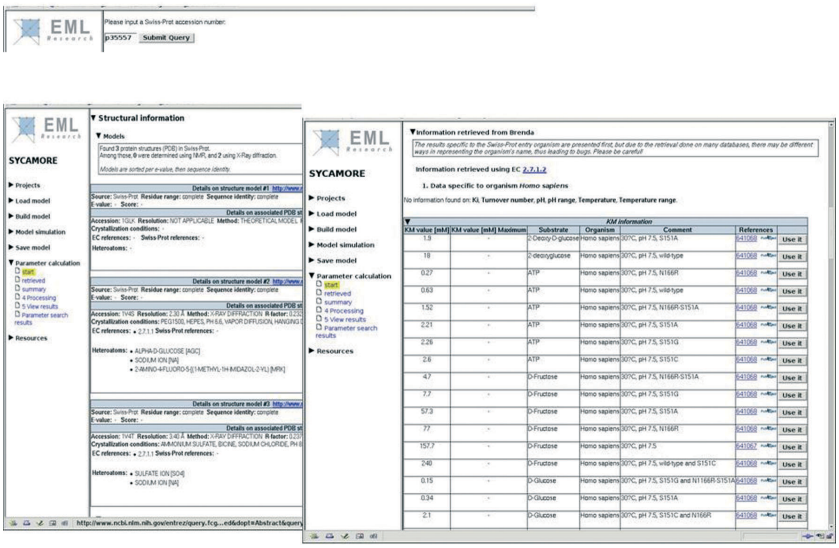


Figure 3. Snapshot of protein information retrieval module within SYCAMORE**.** It retrieves heterogeneous protein information such as protein structure, existing experimental kinetic data and sequence information (see text for details).

Figure 3 shows a snapshot of the protein information retrieval module within SYCAMORE. When querying for the glucokinase from *Homo sapiens* (SwissProt ID P35557) in SwissProt, three related protein structures are found: these are the X-ray crystal structures of the enzyme from *Homo sapiens* in its closed form (PDB entry 1V4S) and its open form (PDB entry 1V4T) [28] plus a theoretical model for the human glucokinase (PDB code 1GLK) based on its homology to the enzyme from yeast. The user may select one of the three models for subsequent structural modelling.

Below, relevant additional structural information for kinetic modelling from the IntAct [29] database at EBI are given, such as the interaction of human glucokinase with the glucokinase regulatory protein (GCKR) and the 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase I.

The next screen displays relevant kinetic information that was found in BRENDA [14] when searching for enzymes with the same EC number. First data for the glucokinase from *Homo sapiens* such as K_m values and specific activities for a range of substrates and the influence of single point mutations on K_m are reported. Then data specific to other organisms are also reported.

The user may choose any of the reported parameters for subsequent mathematical modelling by clicking on the "use it" button. The user then has the option to review his choice of parameters, correct or modify them or insert his own parameters manually for the mathematical modelling of the enzyme glucokinase.

PIPSA of the electrostatic potential of glucokinases

Here we present an illustrative case of the structure-based generation of kinetic parameters from a PIPSA of the electrostatic potential of glucokinases. We analyse the similarity of the electrostatic potentials of a test set of 8 different glucokinases for which experimental K_m constants for the substrate glucose could be found in the BRENDA database. We set our focus on the glucokinase from *E. coli* and demonstrate a procedure to assist the user in the choice of an appropriate K_m value when constructing a kinetic model.

Kinetic constants and comparative protein structural modelling

For the glucokinases from *Homo sapiens*, *Rattus norvegicus*, *Escherichia coli*, *Aspergillus niger*, *Hansenula polymorpha*, *Saccharomyces cervisiae*, *Streptococcus mutans* and *Zymomonas mobilis* K_m values for the substrate glucose could be found in the BRENDA database. They all catalyse an identical chemical reaction. However, they do so with very different substrate binding affinity, represented by the K_m value.

The experimental values found in BRENDA are 0.028 mM (*S. cerevisiae*) [30], 0.05 mM (*H. polymorpha*) [30], 0.063 mM (*Asp. niger*) [30], 0.095 mM (*Z. mobilis*) [31], 0.61 mM (*S. mutans*) [32] to 6 mM (*H. sapiens*) [33] and 7.7 mM *(R. norvegicus*) [34] and thus cover a range of more than 2 orders of magnitude.

For the glucokinase from *E. coli*, the available experimental K_m values range from 0.78 mM [35] to 0.15 mM [36]. Since no experimental error bars are given, we would like to check the completeness and consistency of these values. The user has to make a choice when setting up a kinetic model of glycolysis in *E. coli*. We apply the PIPSA method to compare the electrostatic potentials around the active site and correlate with experimental K_m values from other organisms to suggest a value for *E. coli*.

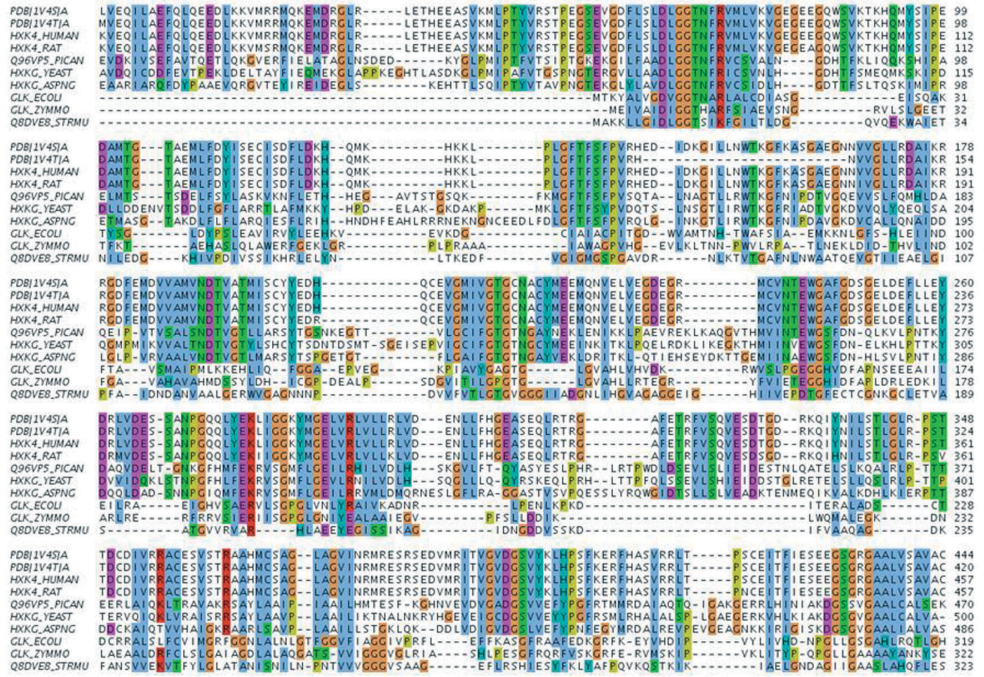


Figure 4. Multiple sequence alignment of glucokinases from *Homo sapiens, Rattus norvegicus, Hansenula polymorpha*, *Saccharomyces cerevisiae*, *Aspergillus niger*, *Escherichia coli*, *Zymmomonas mobilis* and *Streptococcus mutans*. The amino acid sequences of the template structures of the open (PDB code 1V4T) and closed (PDB code 1V4S) [28] forms of the human glucokinases are also given.

Figure 4 shows the ClustalW multiple sequence alignment of glucokinases with the sequences from *Homo sapiens* of the closed (1V4S) and open forms (1V4T) of the enzyme. The multiple sequence alignment was used to generate protein structural models by mapping the target sequences from *Homo sapiens*, *Rattus norvegicus*, *Escherichia coli*, *Aspergillus niger*, *Hansenula polymorpha*, *Saccharomyces cerevisiae*, *Streptococcus mutans* and *Zymomonas mobilis* to the template protein structure of the open (PDB code 1V4T) and closed forms (PDB code 1V4S) of human glucokinase. For each of the generated protein models, the electrostatic potential was calculated.

Calculation and comparison of the electrostatic potentials for glucokinases

The mammalian glucokinase undergoes a large conformational change upon substrate binding [28]. Two of the three layers of the small domain of glucokinase rotate at an angle of 99 ° around a hinge region [28]. The substrate glucose binds to the bottom of the deep cleft between the large domain and the small domain. In the closed form, glucose is coordinated by residues from the large and the small domains.

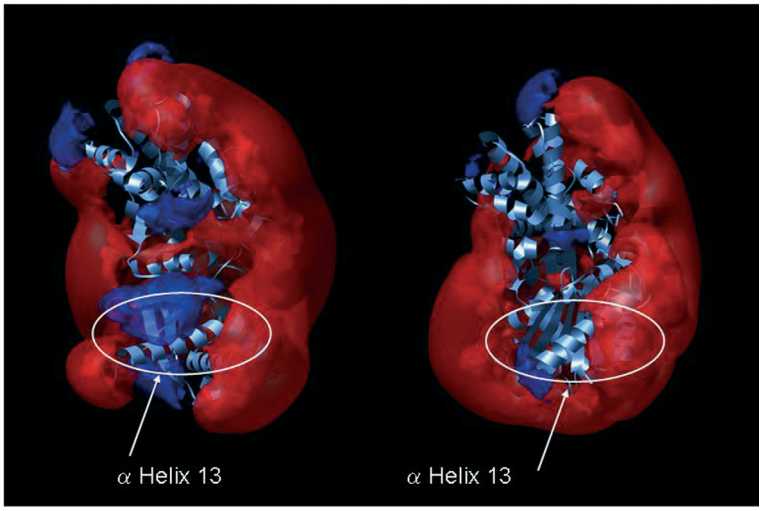


Figure 5. Calculated electrostatic isopotential isosurface at (0.6 kcal mol^{-1} e^{-1} of the open (left) and closed (right) form of the Hexokinase IV from *Homo sapiens* [28].

Figure 5 shows the calculated electrostatic potential for the open form (1V4T; left in Fig. 5) and the closed form (1V4S, right in Fig. 5). The two forms differ in electrostatic potential in particular around the α13 helix which moves in a different direction to the small domain upon conformational change [28].

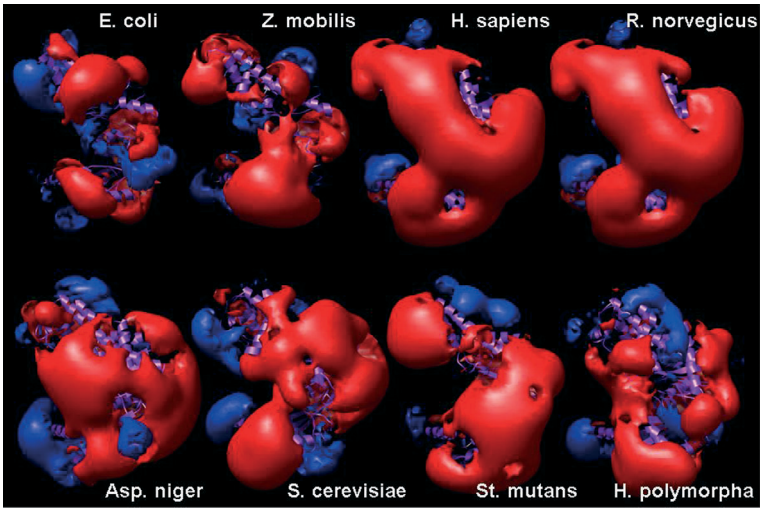


Figure 6. Calculated electrostatic potentials of glucokinases from eight organisms for which substrate K_m values were found in the BRENDA database**.** The isosurfaces are shown at 0.6 kcal mol^{-1} $\mathbf{e^{-1}}$ for the open form of the enzyme**.**

Figure 6 shows the computed electrostatic potential of the glucokinases in the other organisms. All have a large negative patch near the ATP binding region (right side) and a more positive patch on the left. Visual inspection shows that the electrostatic potential of the glucokinases from *Homo sapiens* and *Rattus norvegicus* appear indistinguishable. There is, however, a large variation in the distribution of the electrostatic potential across the organisms.

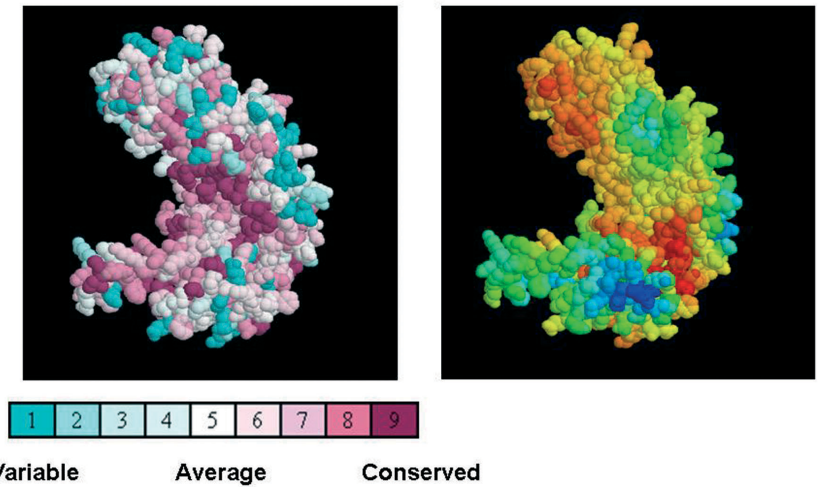


Figure 7:

Left: Conservation of the amino acid residues in the multiple sequence alignment displayed on the open form of human hexokinase IV (1V4T) using the Consurf algorithm (47).

Right: Conservation of the calculated electrostatic potential. Pairwise comparison of the calculated electrostatic potentials.

Figure 7 shows the conservation of the positions of amino acid residues of the eight glucokinases mapped onto the crystal structure of the human enzyme in its open form (left). The most conserved amino acid residues are found in the cleft between the large and small domains: this is the site where the ligand co-crystallizes in the closed form; and a patch of conserved amino acid residues in proximity to the ligand binding site, potentially the entry channel of the substrate. Figure 7 (right) shows the conservation of the electrostatic potential. The most conserved patches of the electrostatic potential of the set of glucokinases, ranging from blue (no conservation), yellow (intermediate) to patches of high conservation (coloured in red). The most conserved electrostatic region approximately overlaps with the region of most conserved amino acid sequences between the two protein domain and may refer to the entry channel of the substrate. The electrostatic potential near the ligand binding site, however, is not strictly conserved. The variations in the electrostatic potentials at this spot may explain the large range of K_m values between mammalian and non-mammalian enzymes.

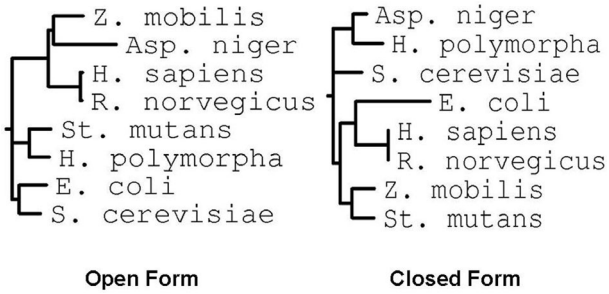


Figure 8: Tree diagram of the similarities of the electrostatic potentials of glucokinases in a region of radius 15 Å around the ligand binding site.

A more quantitative comparison of the electrostatic potentials is possible with the Hodgkin similarity indices. The pairwise similarities can be easily visualized in phylogenetic trees [8]. Figure 8 displays tree diagrams of the similarities of the electrostatic potentials in the test set of eight glucokinases of the open (left) and closed (right) forms. We used a radius of 15 Å around the ligand binding site for the comparison of the electrostatic potentials since the conservation of the active site was also observed in a phylogenetic analysis of the primary sequences of hexokinases [11].

For the closed form, the nearest neighbours of the glucokinase from *E. coli* are the mammalian glucokinases from *Homo sapiens* and *Rattus norvegicus*. This would suggest a K_m value of the *E. coli* glucokinase in the mM range. This assignment seems improbable since the sequence identity is very low between glucokinases from *E. coli* and *Homo sapiens* (14% overall sequence identity).

The mammalian glucokinases in liver (hexokinases IV) possess a high K_m value (6 – 7 mM) and act as a sensor of high glucose levels in the blood since the physiological role of glucokinases in vertebrates is significantly different from that of invertebrates. In mammalians, the glucokinase (hexokinase IV) is the liver-specific isozyme with a glucose sensor function in hepatocytes [11] and represents 95% of the total hexokinase activity of hexokinases. The liver enzymes phosphorylate glucose only when it has reached a high concentration in the blood. Thus, isozymes in brain and muscle, which have 50-fold lower K_m values, are activated first. Only when glucose is abundant, is the liver isozyme active and ensures that glucose is not wasted.

When the electrostatic potentials are computed for protein structural models of the open form (Fig. 8, left), the closest glucokinase to *E. coli* is from *S. cerevisiae* and suggests a K_m value around 0.03 mM for *E. coli*. This predicted K_m value is clearly outside the range of K_m values retrieved from BRENDA: 0.15 mM [35] to 0.78 mM [36]. This discrepancy was analysed further. The glucokinase from *E. coli* displays only weak similarity to the other glucokinases. This had been noticed already by Cardenas *et al*. [37]. The absence of homology with other hexokinases suggested an early divergent evolution of hexokinases

in plants, vertebrates, yeast and bacterial hexokinases. The current investigation suggests that the glucokinase from *E. coli* is a very specific hexokinase with a predicted very low K_m value of the same order of magnitude as yeast.

The recently solved X-ray structure of the ATP-dependent glucokinase from *E. coli* displayed a RNase H-like fold [38] which is also found for *Homo sapiens* [28] and yeast [39] glucokinases and justifies *a posteriori* the use of the template protein structure from *Homo sapiens* despite the low sequence identity.

When searching for additional investigations of the kinetics of the glucokinase from *E. coli* that are not yet included in BRENDA, we found a recent report by Millar and Raines of a K_m value of the glucokinase from *E. coli* of 0.076 mM [40]. This is significantly lower than the K_m values reported previously ranging from 0.15 mM to 0.78 mM.

This supports our assignment of the glucokinase from *E. coli* to the family of very specific bacterial glucokinases with a very low K_m value: 0.028 mM (*S. cerevisiae*) and 0.063 mM (*Asp. niger*).

In general, we found a better correlation of the kinetic parameters for the open form of the enzyme. This was also noticed by Xu *et al*. who correlated calculated interaction energies of various sugars with measured k_{cat}/K_m values [41]. They came to the conclusion that the substrate sugar molecules are recognized by binding to the open form of glucokinase.

PIPSA of a large set of glucokinases

The previous application of the PIPSA classification of glucokinases was limited to a small set of eight experimentally characterized organisms. In systems biology one aims at an understanding of enzymes in context and also across a larger number of organisms.

The investigation of the similarity of the electrostatic potentials of glucokinases was extended to a larger set of proteins. All protein sequences that were annotated as either glucokinases or classified with the EC number 2.7.1.2 were aligned according to their amino acid sequence identity. Sequences which were annotated as polyphosphate glucokinases, ROK (repressor, open reading frame, and kinase) or for which only fragments were available, were removed. This led to a set of 164 aligned protein sequences. Protein structural models were generated based on the template structure of the human hexokinase IV (HXK4_HUMAN) in its open form. Electrostatic potentials were calculated by solving the linearized Poisson–Boltzmann equation (as described above in detail).

The 164 proteins were classified according to their Hodgkin similarity indices of the electrostatic potential in a region of 15 Å radius around the ligand binding site (see Fig. 9). The inserts show magnifications of selected glucokinases from *E. coli*, Yeast and *Homo sapiens*.

The nearest neighbours to *E. coli* are the glucokinases from *E. coli O6*, *Shigella flexnen*, *Salmonella typhi* and *Salmonella typhimarium*. The enzyme from yeast is closest to various glucokinases from *Xylella fastidiosa* and *Yersinia pestis*. From PIPSA of the electrostatic potentials, one may expect glucokinases from *Sparus aurata* (Gilthead sea bream), *Cyprinus carpio* (Common carp), hexokinase IV from *mouse*, *Oncorhynchus mykiss* (Rainbow trout) to exhibit similar kinetic parameters to the enzymes from *Homo sapiens* and *R. norvegicus*. Also the glucokinase EMI2_Yeast (Early Meiotic Induction Protein 2 [42] is predicted to possess similar kinetic parameters. This glucokinase is involved in sporulation and is required for the full activation of the early meiotic inducer EMI1 [41]. This glucokinase performs a different physiological role from bacterial glucokinases and thus a high K_m value may be expected.

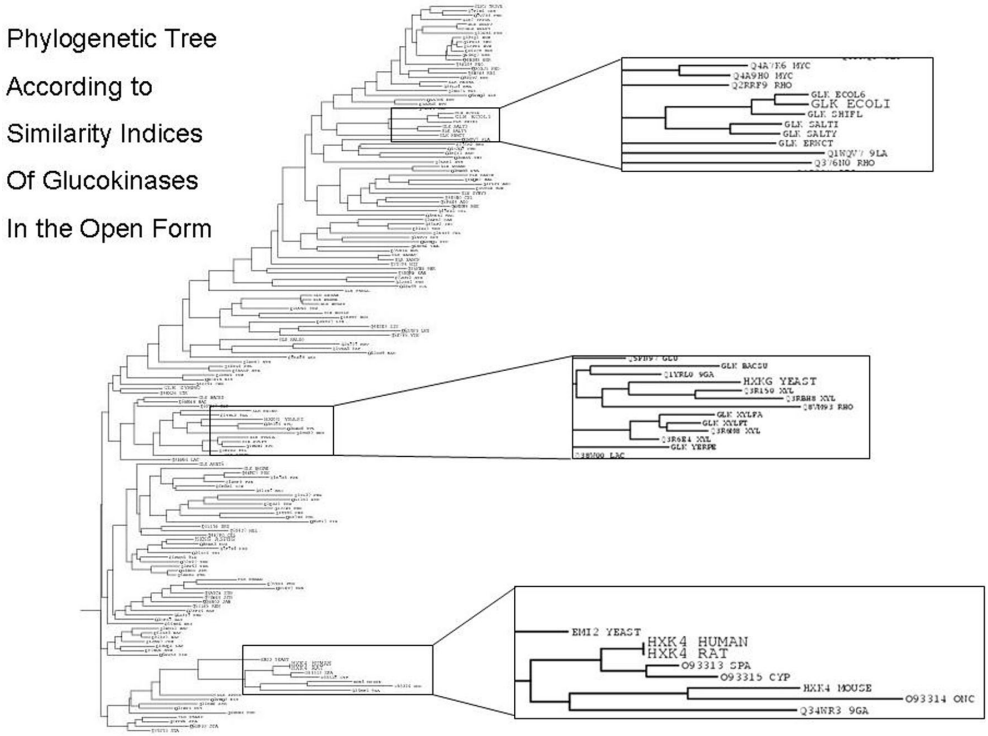


Figure 9: Tree diagram of 164 glucokinases EC 2.1.7.2 classified by their similarity in electrostatic potential of the open form in a region of radius 15 Å around the ligand binding site.

Conclusion and Outlook

Structure-based systems biology provides detailed insight into cellular processes at a molecular level. It is thus complementary to the abstract mathematical modelling of protein signaling or metabolic networks. The PIPSA method provides a quantitative structure to function relationship for enzymes. It quantifies the similarity of molecular interactions between the substrate molecule and the protein active site for the same enzyme from a large number of organisms. The large-scale application of PIPSA allows the classification of enzymes previously uncharacterized and the detection of relationships with other enzymes.

Furthermore, the PIPSA method can be used to detect outliers from a series of well-characterized enzymes. For this use it is critical to have:

i) an extensive annotation of experimental conditions
ii) a detailed and consistent set of experimental data.

Further application and extension of the qPIPSA method to predicting enzymatic K_m and k_{cat}/K_m values and the comparative modelling of the glycolytic pathway across multiple organisms is in progress.

Acknowledgements

We thank the Klaus-Tschira Foundation and the BMBF Systems Biology Initiative HepatoSys for financial support. We are grateful to Dr A. Weidemann for the design and layout of the SYCAMORE user interface and Dr U. Kummer and Dr. I. Rojas for the scientific collaboration.

References

[1] Kitano, H. (2002) *Nature* 420, 206–210.

[2] Gabdoulline, R.R., Kummer, U., Olsen, L.F. & Wade, R.C. (2003) *Biophys. J.* 85, 1421–1428.

[3] Apweiler, R., Cornish-Bowden, A., Hofmeyr, J.-H.S., Kettner, C., Leyh, T.S., Schomburg, D. & Tipton, K. (2005) *Trends in Biochemical Sciences* 30, 11–12.

[4] Aloy, P. & Russell, R.B. (2005) *FEBS Letters* 579, 1854–1858.

[5] Kettner, C. & Hicks, M.G. (2005) *Curr. Enz. Inhib.* 1, 171–181.

[6] http://www.systembiologie.de.

[7] Blomberg, N., Gabdoulline, R.R., Nilges, M. & Wade, R.C. (1999) *Proteins* 37, 379–387.

[8] Winn, P.J., Religa, T.L., Battey, J.N., Banerjee, A. & Wade, R.C. (2004) *Structure* 12, 1563–1574.

[9] Wade, R.C., Gabdoulline, R.R. & Luty, B. (1998) *Proteins* 31, 406–416.

[10] Wade, R.C., Gabdoulline, R.R., Luedemann, S. & Lounnas, V. (1998) *Proc. Natl. Acad. Sci. USA* 95, 5942–5949.

[11] Cardenas, M.L., Cornish-Bowden, A. & Ureta, T. (1998) *Biochim. Biophys. Acta* 1401, 242–264.

[12] Cardenas, M.L. (1997) *Biochemical Society Transactions* 25, 131–135.

[13] Cardenas, M.L. (2003) in *Glucokinase and Glycemic Diseases*, ed. Magnuson, M.A. (Karger, Basel).

[14] Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. & Schomburg, D. (2004) *Nucleic Acids Res.* 32, D431–3.

[15] Rojas, I., Kania, R., Wittig, U., Weidemann, A., Goblewski, M. & Krebs, O. (2005) in *Proceedings of the 4th Workshop on Computation of Biochemical Pathways and Genetic Networks*, eds. Kummer, U., Pahle, J., Surovtsova, I. & Zobeley, J. (Logos Verlag, Berlin), pp. 63–67.

[16] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000) *Nucl. Acids Res.* 28, 235–242.

[17] Pieper, U., Eswar, N., Davis, F.P., Braberg, H., Madhusudhan, M.S., Rossi, A., Marti-Renom, M., Karchin, R., Webb, B.M., Eramian, D., Shen, M.Y., Kelly, L., Melo, F. & Sali, A. (2006) *Nucleic Acids Research.* 34, D291–295.

[18] Kopp, J. & Schwede, T. (2004) *Nucl. Acids Res.* 32, D230-D234.

[19] Bairoch, A., Boeckmann, B., Ferro, S. & Gasteiger, E. (2004) *Brief. Bioinform.* 5, 39–55.

[20] Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I.I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.-H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Le Novere, N., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J. & Wang, J. (2003) *Bioinformatics* 19, 524–531.

[21] Wade, R.C., Gabdoulline, R.R. & Rienzo, F.D. (2001) *Intl. J. Quant. Chem.* 83, 122–127.

[22] Good, A.C., Hodgkin, E.E. & Richards, W.G. (1992) *Journal of Computer-Aided Molecular Design* 6, 513–520.

[23] C. Burt, W.G. Richards & Huxley, P. (1990) *Journal of Computational Chemistry* 11, 1139–1146.

[24] Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) *Nucl. Acids Res.* 22, 4673–4680.

[25] Sali, A. & Blundell, T.L. (1993) *J. Mol. Biol.* 234, 779–815.

[26] Vriend, G. (1990) *J. Mol. Graph.* 8, 52–56.

[27] Davis, M.E., Madura, J.D., Luty, B.A. & McCammon, J.A. (1991) *Computer Physics Communications* 62, 187–197.

[28] Kamata, K., Mitsuya, M., Nishimura, T., Eiki, J. & Nagata, Y. (2004) *Structure* 12, 429–438.

[29] Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S., Vingron, M., Roechert, B., Roepstorff, P., Valencia, A., Margalit, H., Armstrong, J., Bairoch, A., Cesareni, G., Sherman, D. & Apweiler, R. (2004) *Nucl. Acids Res.* 32, D452–455.

[30] Laht, S., Karp, H., Kotka, P., Jarviste, A. & Alamae, T. (2002) *Gene* 296, 195–203.

[31] Scopes, R.K. & Bannon, D.R. (1995) *Biochim. Biophys. Acta* 1249, 173–9.

[32] Porter, V. & Chassy, B.M. (1982) *Methods Enzymol.* 90, 25–30.

[33] Xu, L.Z., Harrison, R.W., Weber, I.T. & Pilkis, S.J. (1995) *J. Biol. Chem.* 270, 9939–46.

[34] Tu, J. & Tuch, B.E. (1996) *Diabetes* 45, 1068–75.

[35] Meyer, D., Schneider-Fresenius, C., Horlacher, R., Peist, R. & Boos, W. (1997) *J. Bacteriol.* 179, 1298–306.

[36] Arora, K.K. & Pedersen, P.L. (1995) *Arch. Biochem. Biophys.* 319, 574–579.

[37] Cardenas, M.L. (2004) in *Glucokinase and Glycemic Diseases*, eds. Matschinsky, F.M. & Magnuson, M.A. (Karger, Basel), Vol. 16, pp. 31–41.

[38] Lunin, V.V., Li, Y., Schrag, J.D., Iannuzzi, P., Cygler, M. & Matte, A. (2004) *Journal of Bacteriology* 186, 6915–6927.

[39] Kuser, P.R., Krauchenco, S., Antunes, O.A. & Polikarpov, I. (2000) *J. Biol. Chem.* 275, 20814–20821.

[40] Miller, B.G. & Raines, R.T. (2004) *Biochemistry* 43, 6387–6392.

[41] Xu, L.Z., Weber, I.T., Harrison, R.W., Gidh-Jain, M. & Pilkis, S.J. (1995) *Biochemistry* 34, 6083–6092.

[42] Enyenihi, A.H. & Saunders, W.S. (2003) *Genetics* 163, 47–54.

[43] Le Novere, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J.L. & Hucka, M. (2006) *Nucl. Acids Res.* 34, D689–691.

[44] Olivier, B.G. & Snoep, J.L. (2004) *Bioinformatics* 20, 2143–2144.

[45] http://www.copasi.org.

[46] Hodgkin, E.E. & Richards, W.G. (1987) *Int. J. Quant. Chem. Quant. Biol. Symp.*, 105–110.

[47] Glaser, F., Pupko, T., Paz, I., Bell, R.F., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003) *Bioinformatics* 19, 163–164.