Update of KDBI: Kinetic Data of Bio-molecular Interaction database

Pankaj Kumar1, B. C. Han1, Z. Shi1, J. Jia1, Y. P. Wang2, Y. T. Zhang2, L. Liang2, Q. F. Liu2, Z. L. Ji2 and Y. Z. Chen1,*

1Bioinformatics and Drug Design Group, Centre for Computational Science and Engineering, National University of Singapore, Blk S16, Level 8, 3 Science Drive 2, Singapore 117543 and 2Bioinformatics Research Group, School of Life Sciences, Xiamen University, Xiamen 361005, FuJian Province, P. R. China

Received September 15, 2008; Revised October 10, 2008; Accepted October 14, 2008

ABSTRACT

Knowledge of the kinetics of biomolecular interactions is important for facilitating the study of cellular processes and underlying molecular events, and is essential for quantitative study and simulation of biological systems. Kinetic Data of Bio-molecular Interaction database (KDBI) has been developed to provide information about experimentally determined kinetic data of protein–protein, protein–nucleic acid, protein–ligand, nucleic acid–ligand binding or reaction events described in the literature. To accommodate increasing demand for studying and simulating biological systems, numerous improvements and updates have been made to KDBI, including new ways to access data by pathway and molecule names, data file in System Biology Markup Language format, more efficient search engine, access to published parameter sets of simulation models of 63 pathways, and 2.3-fold increase of data (19,263 entries of 10,532 distinctive biomolecular binding and 11,954 interaction events, involving 2,635 proteins/protein complexes, 847 nucleic acids, 1,603 small molecules and 45 multi-step processes). KDBI is publicly available at http://bidd.nus.edu.sg/group/kdbi/kdbi.asp.

INTRODUCTION

Biomolecular interactions, via individual and network actions, play fundamental roles in biological, disease and therapeutic processes (1–4). Extensive experimental and computational studies have significantly advanced our understanding of the characteristics, organization, evolution and complexity of biomolecular interaction networks in biological systems (5–8), and enabled the generation of genome-scale protein–protein interactions and the development prediction tools (6,7,9–12).

Many databases have been developed for providing information about biomolecular interactions [e.g. MIPS (13), DIP (14), BIND (15), Biocyc (16), MINT (17), Bio-models (18), STRING (19) and IntAct (20)], and biological networks and pathways [KEGG (21), BioGRID (22), NetworKIN (23), STITCH (24), DOMINE (25), CellCircuits (26), Reactome (27) and enzyme reactions (28)].

In view that quantitative as well as mechanistic understanding of biomolecular interactions is important for exploration and engineering of biological networks and for the development of novel therapeutics to combat diseases (29,30), kinetic data of biomolecular interactions have been provided in some databases. For instance, BRENDA (31) and SABIO-RK (32) provide kinetic constants of enzymatic activities, DOQCS contains kinetic parameters of simulation models of cellular signaling derived from experimental and other sources (33). To complement these databases for providing the kinetic data not yet covered by other databases, we have developed the Kinetic Data of Bio-molecular Interactions database [KDBI; (34)] to provide experimentally measured kinetic data for protein–protein, protein–nucleic acid and protein–small molecule interactions aimed at facilitating mechanistic investigation, quantitative study and simulation of cellular processes and events (31–33,35–39). Kinetic data in KDBI have been manually collected from literatures, a substantial percentage of which are not yet available in other databases (e.g. some protein–protein interactions in thrombin, translation initiation, DNA repair, and ion transport pathways, and individual protein–nucleic acid interactions).

In the updated KDBI, apart from 2.3-fold increase of experimental kinetic data, we added four new features. The first is the access of KDBI entries via the list of nucleic acid and pathway names. The second is the inclusion of literature-reported kinetic parameter sets of 63 pathway
simulation models (35–44) for facilitating the applications, assessments and further development of these pathway models. The third is the facility for collectively accessing the available kinetic data of multi-step processes (e.g. metabolism and pathway segments) collected in KDBI. The fourth is the availability of SBML (45) files for all records of the kinetic parameter sets of pathway simulation models for facilitating the use of the relevant data in such software tools as Celldesigner (46), Copasi (47), cPath (48), PaVESy (49) and SBMLeditor (50).

**EXPERIMENTAL KINETIC DATA AND ACCESS**

Additional sets of the experimentally determined kinetic data of biomolecular interactions were collected from published literatures. Compared to the last version of KDBI, the number of entries in the updated KDBI is increased by 2.3-fold to 19,263, which include 2,635 protein–protein, 1,711 protein–nucleic acid, 11,873 protein–small molecule and 1,995 nucleic acid–small molecule interactions. Each entry provides detailed description about binding or reaction event, participating molecules, binding or reaction equation, kinetic data and related references. As shown in Figures 1–3, kinetic data for protein–protein, small molecule–nucleic acid and protein–small molecule interactions are provided in terms of one or a combination of kinetic quantities as given in the literature of a particular event. These quantities include association/dissociation rate constant, on/off rate constant, first/second/third/... order rate constant, catalytic rate constant, equilibrium association/dissociation constant, inhibition constant, and binding affinity constant, IC50, etc. and experimental conditions (pH value and temperature).

These data can be accessed via input of names of molecules and bio-events (association, dissociation, complex formation, electron transfer, inhibition, etc.), and via selection of pathway and protein name from the pathway list and protein list fields in KDBI webpage. The kinetic data of an event are searchable by several methods. One method is via the name of participating molecules (protein, nucleic acid, small peptide, ligand or ion) or pathway involved in an event. In some events described in the literature, a participating entity is an unidentified molecule located in the membrane of a cell or on the surface of a virus. In these entries, only the name of the cell or virus is given. An entry can also be searched through a Swiss-Prot AC number for a protein or the CAS number for a small molecule ligand. Moreover, keyword-based text search is also supported. To facilitate convenient access of relevant data, partial lists of proteins and nucleic acid are provided. Searches involving combination of these methods or selection fields are also supported.

**PARAMETER SETS OF PATHWAY SIMULATION MODELS**

As part of the efforts for facilitating the understanding and quantitative analysis of complex biological processes and network responses, mathematical simulation models of various pathways have been developed and extensively used for studying and quantitative understanding of signaling dynamics (35–39), signal-specific sensing (40) and discrimination (44), feedback regulations and cross-talks (42,43), and receptor cross-activation (41) and internalization (42). These mathematical models typically use ordinary differential equations (ODEs) to describe the temporal dynamic behavior of molecular species in the pathway. The kinetic rate constants of protein–protein,
protein–small molecule, protein–nucleic acid and other interactions (e.g. binding association rate $K_a$, binding dissociation rate $K_d$, reaction rate $K_r$, reaction turnover rate $K_{cat}$, Michaelis–Menten constant $K_m$) are needed to establish these ODEs, which have been primarily generated by combinations of experimental data, computed theoretical values and empirically fitted values computational (39–44). To facilitate further applications, developments and assessments of the published pathway models, we collected and included in KDBI the parameter sets of 63 published ODE-based models, which can be accessed from the pathway list in the ‘Pathway Simulation Parameters’ field in KDBI webpage. Moreover, we added kinetic data type to every entry to clearly distinguish its original source (experimental or simulation model). In particular, for the kinetic data of a simulation model that have been obtained from other publications, cross-reference to the original source is provided. A typical search result is shown in Figure 4.

**KINETIC DATA FOR MULTI-STEP PROCESSES**

Some published studies provide information about the experimental kinetic data for multiple components of
Examples of these processes include RNA binding activity to translation initiation factors eIF4G, 70-kDa heat shock protein polymerization, control of platelet function by cyclic AMP, GroEL interaction with conformational states of horse cytochrome c, intermolecular catalysis by hairpin ribozymes, antisense RNA interaction with its complimentary RNA and nucleotide binding to actin. To facilitate the development of pathway simulation models based on these building blocks, we provided direct access to the collection of the kinetic data for each of these processes, which can be accessed via a separate search field ‘Multi-step processes’ in KDBI webpage. A typical search result is shown in Figure 5.

**Figure 4.** Pathway parameter set page. This page provides kinetic data and reaction equation (while available) as well as the name of participating molecules and description of event.

**Figure 5.** Multi-process kinetic data page. This page provides kinetic data and reaction equation (while available) as well as the name of participating molecules and description of event.

**KINETIC DATA FILES IN THE SYSTEMS BIOLOGY MARKUP LANGUAGE FORMAT**

Systems Biology Markup Language (SBML) has been developed as a free, open and XML-based format for...
representing biochemical reaction networks, and it is a software-independent language for describing models common to computational biology research, including cell signaling pathways, metabolic pathways, gene regulation and others (54). Many pathway simulation and analysis software tools have built-in SBML compatibility features to allow the input, manipulation, simulation and analysis of different pathway models and parameters (45, 54–58). To facilitate the input of the pathway parameter sets into these software tools, we created the SBML file for the parameter sets of all 63 pathway simulation models included in KDBI, which can be downloaded via the link provided on the top of the page that displays the relevant kinetic data. SBML file for the user-selected kinetic data can be dynamically generated and exported by clicking the selected entries and then the SBML file export button.

**REMARKS**

The updated version of KDBI is intended to be a more useful resource for convenient access of available biomolecular kinetic data to complement other biomolecular interaction and pathway databases in facilitating quantitative studies of biomolecular interactions and networks. New technologies have been developed in employing surface plasmon resonance technology for deriving real-time dynamics and kinetic data (59), and in using protein microarrays (60) and solution NMR spectroscopy (61) for monitoring and characterizing biomolecular interactions. Moreover, new experimental designs of the well-establishe
d technologies such as isothermal titration calorimetry allow the measurement and estimate of previously inaccessible kinetic parameters (62). Resources for collecting and accessing the increasing amount of kinetic data can better serve the need for mechanistic investigation, quantitative study and simulation of biological processes and events.

**FUNDING**

Funding for open access charge: New Century Excellent Talents in University (NCET) of MOE, China; grant (2007/019) from the National Natural Science Foundation of China (NSFC).

**Conflict of interest statement.** None declared.

**REFERENCES**

1. Kitano, H. (2007) Innovation – a robustness-based approach to systems-oriented drug design. *Nat. Rev. Drug Discov.*, 6, 202–210.
2. Legrain, P., Wojcik, J. and Gauthier, J. M. (2001) Protein-protein interaction maps: a lead towards cellular functions. *Trends Genet.*, 17, 346–352.
3. Downward, J. (2001) The ins and outs of signalling. *Nature*, 411, 759–762.
4. Lengeler, J. W. (2000) Metabolic networks: a signal-oriented approach to cellular models. *Biol. Chem.*, 381, 911–920.
5. Beyer, A., Bandopadhyay, S. and Ideker, T. (2007) Integrating physical and genetic maps: from genomes to interaction networks. *Nat. Rev. Genet.*, 8, 699–710.
6. Drees, B. L., Sundin, B., Brazeau, E., Caviston, J. P., Chen, G. C., Guo, W., Kozminski, K. G., Lau, M. W., Moskov, J. J., Tong, A. et al. (2001) A protein interaction map for cell polarity development. *J. Cell Biol.*, 154, 549–571.
7. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415, 141–147.
8. Qian, J., Lin, J., Luscombe, N. M., Yu, H. and Gerstein, M. (2003) Prediction of regulatory networks: genome-wide identification of transcription factor targets from gene expression data. *Bioinformatics*, 19, 1917–1926.
9. Lo, S. L., Cai, C. Z., Chen, Y. Z. and Chung, M. C. M. (2005) Effect of training datasets on support vector machine prediction of protein–protein interactions. *Proteomics*, 5, 876–884.
10. Phizicky, E., Bastaens, P. I., Zhu, H., Snyder, M. and Fields, S. (2003) Protein analysis on a proteomic scale. *Nature*, 422, 208–215.
11. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. and Yeates, T. O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl Acad. Sci. USA*, 96, 4285–4288.
12. Dandekar, T., Snel, B., Huynen, M. and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.*, 23, 324–328.
13. Mewes, H. W., Frishman, D., Guldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Rosenberg, M., Sturmgrothe, M., Rudd, S. and Weil, B. (2002) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.*, 30, 31–34.
14. Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U. and Eisenberg, D. (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res.*, 32, D449–D451.
15. Alfarano, C., Andrade, C. E., Anthony, K., Bahroos, N., Bajec, M., Bantoft, K., Betel, D., Bobechko, B., Boutiller, K., Burgess, E. F. *et al.* (2005) The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res.*, 33, D418–D424.
16. Karp, P. P., Ouzounis, C. A., Moore-Kochlaes, C., Goldovsky, L., Koppa, P., Ahren, D., Tsoka, S., Darzentas, N., Kunn, V. and Lopez-Bigas, N. (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res.*, 33, 6083–6089.
17. Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M. and Cesareni, G. (2002) MINT: a molecular INteraction database. *FEBS Lett.*, 513, 135–140.
18. Le Novère, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dhahiri, H., Li, L., Sacco, H., Silhvestra, M., Shapiro, B. *et al.* (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.*, 34, D689–D691.
19. von Mering, C., Jensen, L. J., Kuhn, M., Chaffron, S., Doerkts, T., Kruger, B., Snel, B. and Bork, P. (2007) STRING 7 – recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.*, 35, D358–D362.
20. Kerrien, S., Alam-Farquey, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feurmann, M., Friedrichsen, A., Huntley, R. *et al.* (2007) IntAct – open source resource for molecular interaction data. *Nucleic Acids Res.*, 35, D651–D655.
21. Okuda, S., Yamada, T., Hamajima, T., Itoh, M., Katayama, T., Bork, P., Goto, S. and Kanehisa, M. (2008) KGCG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res.*, 36, W423–W426.
22. Breitkreutz, B. J., Stark, C., Reguly, T., Boucher, L., Breitkreutz, A., Livstone, M., Oughtred, R., Lackner, D. H., Bahler, J., Wood, V. *et al.* (2008) The BioGRID interaction database: 2008 update. *Nucleic Acids Res.*, 36, D637–D640.
23. Linding, R., Jensen, L. J., Pascaleeau, A., Olhovsky, M., Colwill, K., Bork, P., Yaffe, M. B. and Ponsaw, T. (2008) NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.*, 36, D695–D699.
24. Kuhn, M., von Mering, C., Campillos, M., Jensen, L. J. and Bork, P. (2008) STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res.*, 36, D684–D688.
25. Rughavachari, B., Tasnem, A., Przytycka, T. M. and Joithi, R. (2008) DOMINE: A database of protein domain interactions. *Nucleic Acids Res.*, 36, D656–D661.
26. Mak,H.C., Daly,M., Gruebel,B. and Ideker,T. (2007) CellCircuits: a database of protein network models. *Nucleic Acids Res.*, 35, D538–D545.

27. Joshi-Tope,G., Gillespie,M., Vastryk,I., D’Eustachio,P., Schmidt,E., de Bono,B., Jassal,B., Gopinath,G.R., Wu,G.R., Matthews,L. et al. (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, 33, D428–432.

28. Goto,S., Okuno,Y., Hattori,M., Nishioka,T. and Kanehisa,M. (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.*, 30, 402–404.

29. Fabrizi,F., Bunnapradist,S., Lunghi,G. and Martin,P. (2003) The systems biology markup language (SBML): a versatile modeling tool for biochemical networks. *Proc. IEEE*, 96, 1254–1265.

30. Zhou,S., Chan,E., Lim,L.Y., Boehler,P., Li,S.C., Wang,J., Zhang,Q., Huang,M. and Xu,A. (2004) Therapeutic drugs that behave as mechanism-based inhibitors of cytochrome P450 3A4. *Curr. Drug Metab.*, 5, 415–442.

31. Schomburg,I., Chang,A. and Schomburg,D. (2002) BRENDA, a database of kinetic data and metabolic information. *Nucleic Acids Res.*, 30, 47–49.

32. Rojas,I., Golebiewski,M., Kania,R., Krebs,O., Mir,S., Weidemann,A. and Wittig,U. (2007) Storing and annotating of kinetic data. In *Silico Biol.*, 7, S37–S44.

33. Sivakumaran,S., Harirhharampur,S., Mishra,J. and Bhalla,U.S. (2003) The database of quantitative cellular signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics*, 19, 408–415.

34. Ji,Z.L., Chen,X., Zhen,C.J., Yao,L.X., Han,L.Y., Yeo,W.K., Chung,P.C., Puy,H.S., Tay,Y.T., Muhammad,A. et al. (2003) KDBI: kinetic data of bio-molecular interactions database. *Nucleic Acids Res.*, 31, 255–257.

35. Fussenegger,M., Bailey,J.E. and Schomburg,D. (2002) BRENDASE, enzyme data and metabolic information. *Nucleic Acids Res.*, 30, 47–49.

36. Haugh,J.M., Wells,A. and Laufbender,D.A. (2000) Mathematical modeling of epidermal growth factor receptor signaling through the phospholipase C pathway: mechanistic insights and predictions for molecular interventions. *Biotechnol. Bioeng.*, 70, 225–238.

37. J. Biol. Chem., 280, 899–910.

38. van den Broek,B., Noorm,M.C. and Wuite,G.J. (2005) DNA-tension dependence of restriction enzyme activity reveals mecanochemical properties of the reaction pathway. *Nucleic Acids Res.*, 33, 2676–2684.

39. Schoeberl,B., Eichler-Jonsson,C., Gilles,E.D. and Muller,G. (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGFR receptors. *Nat. Biotechnol.*, 20, 370–375.

40. Sasagawa,S., Ozaki,Y., Fujita,K. and Kuroda,S. (2005) Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nat. Cell Biol.*, 7, 365–373.

41. Birtwistle,M.R., Hatakeyama,M., Yamoto,N., Oguma,B., Hoek,J.B. and Kholodenko,B.N. (2007) Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. *Mol. Syst. Biol.*, 3, 144.

42. Ung,C.Y., Li,H., Ma,X.H., Jia,J., Li,B.W., Low,C.B. and Chen,Y.Z. (2008) Simulation of the regulation of EGFR endocytosis and EGFR-ERK signaling by endophilin-mediated RhoA-EGFR crosstalk. *FEBS Lett.*, 582, 2283–2290.

43. Suresh,B.C.V., Babar,S.M.E., Song,E.J., Oh,E. and Yoo,Y.S. (2008) Kinetic analysis of the MAPK and PI3K/Akt signaling pathways. *Mol. Cells*, 25, 397–406.

44. Altan-Bonnet,G. and Germain,R.N. (2004) Modeling C1 cell antigen discrimination based on feedback control of dual ERK responses. *PLoS Biol.*, 3, 1925–1938.

45. Bornstein,B.J., Keating,S.M., Jouraku,A. and Hucka,M. (2008) LibSBML: an API library for SBML. *Bioinformatics*, 24, 880–881.

46. Funahashi,A., Matsuoka,Y. and Goto,Y. (1996) Interaction of GroEL and GroES: inhibition of GroEL-dependent ATPase activity reveals morphological changes inGroEL and GroES. *J. Mol. Biol.*, 267, 570–580.

47. Hoops,S., Sahle,S., Gauges,R., Lee,C., Pahle,J., Simus,N., Singhal,M., Xu,L., Mendes,P. and Kummer,U. (2006) COPASI: a complex pathway simulator. *Bioinformatics*, 22, 3067–3074.

48. Cerami,E.G., Bader,G.D., Gross,B.E. and Sander,C. (2006) cPath: pathway visualization and editing system. *Bioinformatics*, 20, 2841–2844.

49. Nicolas,R., Donizelli,M. and Le Novere,N. (2007) SBMLEdit: effective creation of models in the systems biology markup language (SBML). *BMC Bioinformatics*, 8, 79.

50. Franch,T., Jouraku,A., Magnusson,E.H., Jacobsen,J.P. and Gerdes,K. (1999) Antisense RNA regulation in prekaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *J. Mol. Biol.*, 294, 1115–1125.

51. Korneeva,N.L., Lamphbear,B., Hennig,F.L.C., Merrick,W.C. and Rhoads,R.E. (2001) Characterization of the two eIF4A-binding sites on human eIF4G-1. *J. Biol. Chem.*, 276, 2872–2879.

52. Hoshino,M., Kawaata,Y. and Goto,Y. (1996) Interaction of GroEL with conformational states of horse cytochrome c. *J. Mol. Biol.*, 262, 575–587.

53. 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. et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19, 524–531.

54. Alves,R., Antunes,F. and Salvador,A. (2006) Tools for kinetic modeling of biochemical networks. *Nat. Biotechnol.*, 24, 667–672.

55. Deckard,A., Bergmann,F.T. and Sauro,H.M. (2006) Supporting the SBML layout extension. *Bioinformatics*, 22, 2966–2967.

56. Schmidt,H., Drews,G., Vera,J. and Wolkenhauer,O. (2007) SBML export interface for the systems biology toolbox for MATLAB. *Bioinformatics*, 23, 1297–1298.

57. Hoxit,Z. and Klipp,E. (2006) SBML-PET: a systems biology markup language-based parameter estimation tool. *Bioinformatics*, 22, 2704–2705.

58. Huber,W. and Mueller,F. (2006) Biochemical interaction analysis in drug discovery using surface plasmon resonance technology. *Curr. Pharm. Des.*, 12, 3999–4021.

59. Yu,X.B., Xue,D.K. and Cheng,Q. (2006) Label-free detection methods for protein microarrays. *Proteomics*, 6, 5493–5503.

60. Pellecchia,M. (2005) Solution nuclear magnetic resonance spectroscopy techniques for probing intermolecular interactions. *Chem. Biol.*, 12, 961–971.

61. Surolia,N.J. and Haq,I. (2007) Advances in the analysis of isothermal titration calorimetry data for ligand-DNA interactions. *Methods*, 42, 162–172.