


DATABASE

Open Access



Institute collection and analysis of Nanobodies (iCAN): a comprehensive database and analysis platform for nanobodies

Jing Zuo[†], Jian Li^{*†} , Rongxin Zhang, Longsheng Xu, Hanhan Chen, Xiaohuan Jia, Zhipeng Su, Linhong Zhao, Xing Huang and Wei Xie^{*}

Abstract

Background: Nanobodies are single-domain antibodies that contain the unique structural and functional properties of naturally-occurring heavy chain in camelidae. As a novel class of antibody, they show many advantages compared with traditional antibodies such as smaller size, higher stability, improved specificity, more easily expressed in microorganisms. These unusual hallmarks make them as promising tools in basic research and clinical practice. Although thousands of nanobodies are known to be published, no single database provides searchable, unified annotation and integrative analysis tools for these various nanobodies.

Results: Here, we present the database of Institute Collection and Analysis of Nanobodies (iCAN). It is built for the aim that addressing the above gap to expand and accelerate the nanobody research. iCAN, as the first database of nanobody, contains the most comprehensive information to date on nanobodies and related antigens. So far, iCAN incorporates 2391 entries which include 2131 from patents and 260 from publications and provides a simple user interface for researchers to retrieve and view the detailed information of nanobodies. In addition to the data collection, iCAN also provides online bioinformatic tools for sequence analysis and characteristic feature extraction.

Conclusions: In summary, iCAN enables researchers to analyze nanobody features and explore the applications of nanobodies more efficiently. iCAN is freely available at <http://ican.ils.seu.edu.cn>.

Keywords: Nanobody, Nanobody sequence analysis, Database

Background

Camelidae (*Camelus ferus*, *Camelus bactrianus*, *Camelus dromedarius*, *Vicugna vicugna*, *Vicugna pacos*, *Lama guanicoe*, *Lama glama*) and some cartilaginous fish (nurse shark, wobbegong and dogfish sharks) were discovered to produce functional antibodies which naturally lacked light chains [1, 2]. The single domain antigen-binding fragments of camelid heavy chain antibodies are referred to VHs or nanobodies. Nanobody entity contains four framework regions (FR1–4), forming the core structure of the domain, which are alternated

with three complementary determining regions (CDR1–3). Compared with traditional antibodies, nanobodies show obvious advantages: higher affinity, higher solubility, higher domain stability, smaller size (~15 kda) and recognition of hidden antigenic sites, reduced aggregation tendencies, more easily expressed in microorganisms [3–8].

In the latest years, there has been an enormous growth of interest in studying nanobody for its unique characteristic structure and being the ideal candidate for the development of sophisticated nanobiotechnologies in diverse fields. In basic research, nanobodies are developed into various research tools such as affinity purification [9], gene activation or inactivation [10], immunoprecipitation [11]. Due to the nanoscopic size and high affinity against

* Correspondence: jianli2014@seu.edu.cn; wei.xie@seu.edu.cn

[†]Equal contributors

The Key Laboratory of Developmental Genes and Human Disease, Ministry of Education, Institute of Life Sciences, Southeast University, Nanjing, China



intracellular signaling molecules, nanobodies and their derivative formats are used as nanotracer in intracellular bioimaging [12]. In addition, nanobodies are applied in disease diagnosis, for example, molecular diagnosis for breast cancer [13]. The application of nanobodies in targeting therapeutics is in an early stage but shows good prospects in the therapy of acute thrombotic thrombocytopenic purpura [14], infectious disease [15], rheumatoid arthritis [16], central nervous system disease [17], breast and ovarian cancers [18] and so on. Besides, nanobodies are used to be enzyme inhibitors to increase starch content [19] and also to do food testing in agriculture.

Currently, the application research of nanobody is still in exploration. Although thousands of nanobodies have been published, there exists no database of nanobodies integrating nanobody resources. iCAN is built as the first comprehensive nanobody database, which is presented for the aim that providing searchable, unified annotation and integrative analysis tools for both academic and industrial researchers to expand and accelerate the nanobody research. In order to accomplish this, the annotation data was collected and integrated from public sources. The first version of iCAN has 2391 entries, comprising 74 nanobodies from RCSB PDB, 156 nanobodies from EMBL and 2131 nanobodies from patents. Information related to nanobody sequence, structure, target antigens, function, taxonomy of the source organism is shown. In addition, iCAN also provides online bioinformatic tools which are designed for sequence analysis such as Blast, Clustal Omega, Motif extraction and CDR prediction tool. These tools can be utilized to fast discover characteristic feature of nanobodies against one special antigen in sequence. Altogether,

iCAN will aid the scientific/industrial community in analyzing the nanobody features and exploring the applications of nanobodies.

Construction and content

Data collection and organization

Sequences and structural information of nanobodies were retrieved from databases of RCSB PDB (http://www.rcsb.org/pdb/home/home.do), EMBL (http://www.embl.org/), PubMed (https://www.ncbi.nlm.nih.gov/pubmed) and public patents using combination of keywords like ‘nanobody’, ‘single domain antibody’, and ‘VHH’. Information on nanobody sequence, structure, targeted antigen, source organism and publication details was manually curated. Currently, iCAN includes curated and unified annotations of 2391 unique nanobody entries. The structure information of 74 nanobodies is shown in Structure interface separately. Because there is no a nomenclature for nanobodies until now, we renamed all curated nanobodies following a format ‘CAN_serial number’ as their temporary names and unique identifiers in our database.

Database architecture

iCAN is a user-friendly website with concise interfaces which enable easy and fast navigation. Figure 1 illustrates the architecture of this database site. The detailed introduction to website interfaces can be viewed in the Help page of iCAN or the web manual in Additional file 1.

iCAN is architected to provide a unified resource for the scientific/industrial community allowing users to locate items in this database by any of these queries: (1) list the related information such as targeted antigen,

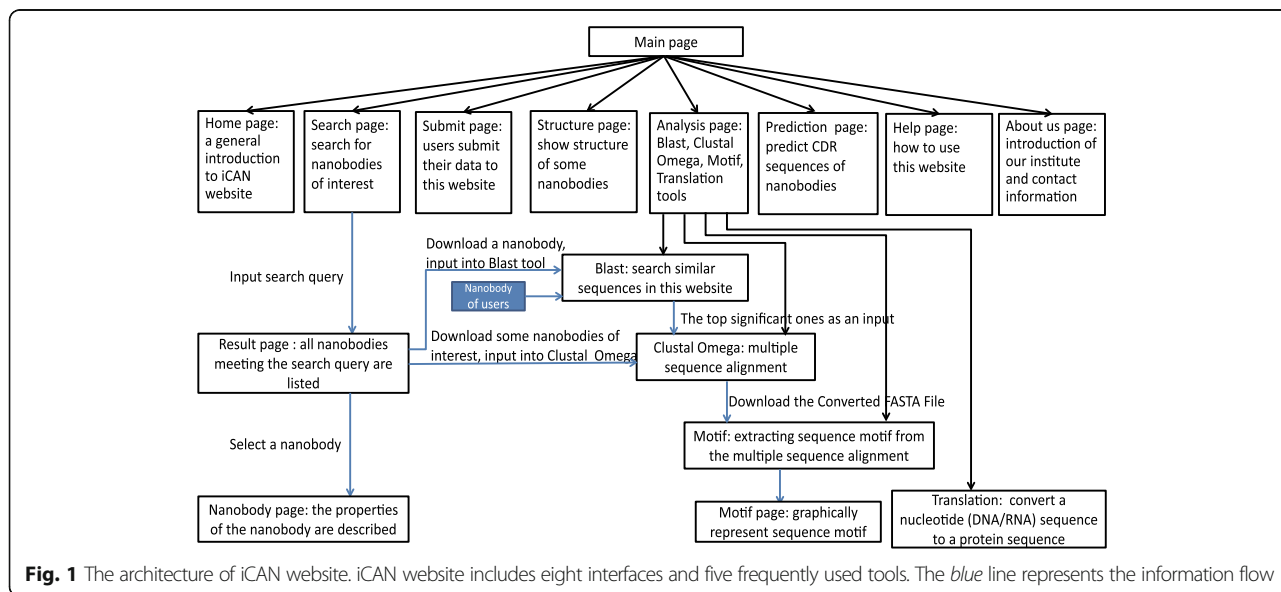


Fig. 1 The architecture of iCAN website. iCAN website includes eight interfaces and five frequently used tools. The blue line represents the information flow

'CAN_serial number' of nanobodies, function and source organism; (2) query by raw sequence such as nanobody full length sequence or CDR1, 2, 3 domain sequence; (3) list the identifier of published origin like PDB ID, PubMed ID. Besides, users can view structure information through PDB link after screening nanobodies of interest in Structure interface (Additional file 2: Figure S1a). iCAN is also architected to aid researchers to do sequence analysis with four frequently-used tools (Blast, Clustal Omega, Translation and Motif) in Analysis interface and CDR prediction tool in Prediction interface. In order to collect as many nanobodies as possible to achieve the aim of nanobody prediction in the future, iCAN provides Submit interface (Additional file 2: Figure S1b) for researchers to upload their own nanobodies.

Database implementation

iCAN is built on Apache HTTP Server (V2.4.43) with PHP (V5.6.24). The back-end functions using MySQL Server (V5.5.53) and the front-end works using Cascading Style Sheets (CSS), HyperText Markup Language (HTML). The required HTML pages are returned in response to user query. After the user's Web browser sends a HTTP query to Web server, the required script including database query is executed. Then the PHP script dynamically generates results in the form of an HTML page and sends to the user's computer (Additional file 3: Figure S2).

Statistical results

The iCAN database currently contains 2391 nanobodies (Additional file 4). The source organism of nanobody is mainly *Lama glama* (1395), *Camelus bactrianus* (48) and *Camelus dromedarius* (27). Most nanobodies are obtained by immunizing animals, whereas, some nanobodies are yielded through phage screening natural

antibody repertoires from non-immunized camels or from synthesized library (Fig. 2a). Recently, nanobodies are mostly applied in clinical practice (1863), basic research (130). Besides, they are also used in fields of crystallization aids, food testing and structure biology (Fig. 2b).

Utility

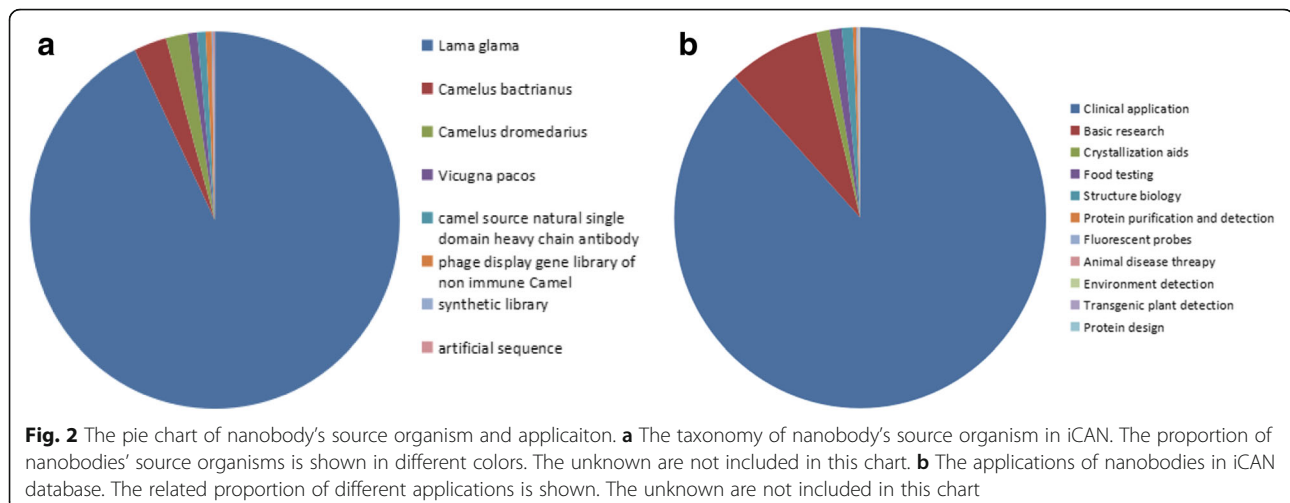
Obtaining detailed information and sequence feature for nanobody of interest is necessary for nanobody research and application. iCAN provides users with such information and convenience. Here we present an example that one can get the sequence feature of nanobodies for a given antigen (Human Epidermal Growth Factor Receptor, EGFR) by using iCAN data and analytic tools. To achieve the above goal, one can follow the below pipeline: obtaining nanobody sequence by searching antigen name or by blasting a given similar nanobody sequence, followed by downloading the data, multiple sequence alignment and motif extraction.

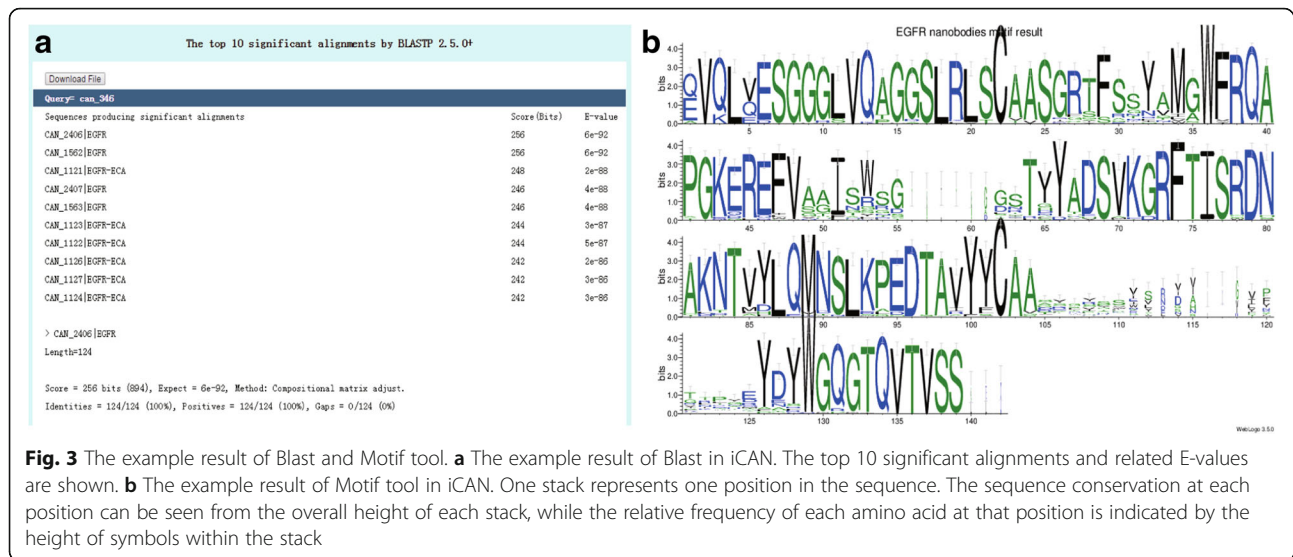
Obtaining nanobody information by searching

Basic search and advanced search are very useful functions, which were developed to identify nanobody based on name, antigen, PDB ID, function, source organism, etc. The detailed information about nanobodies can be viewed by clicking on search result that links to general information page. For example, we get 68 entries of the resulting nanobodies using Basic Search by typing EGFR as antigen.

Obtaining nanobody sequence by BLAST

BLAST is one of the most important tools for searching nanobodies with sequence as the keyword. It can be used to search nanobody from all nanobodies or, specially, from the patented nanobodies recorded in iCAN





database. In the page of Blast search result, the top 10 significant alignments are shown (Fig. 3a).

Data download

Users can download the search results in FASTA format in a batch which will be used in multiple sequence alignments.

Multiple sequence alignments browsing

The nanobody sequence alignments can be viewed by Clustal Omega. We input the downloaded results to Clustal Omega module and then download the converted FASTA file.

Sequence motif extraction

The downloaded converted FASTA file from the last step is used as input data for extracting sequence motif. A graphical result of sequence motif feature from the multiple sequence alignment is shown in Fig. 3b. We can get the primary information about each position of nanobody sequence and corresponding proportion of EGFR nanobodies.

Discussion

Besides our current work, there are still many aspects of the database that can be further improved in the future. First, a large fraction of nanobody sequences are likely not to be published for some reasons such as commercial interest, which are not included in this database. It is very necessary to expand the number of nanobodies of iCAN in the future with the help of all researchers. Second, structure of nanobody is key feature for specially recognizing and binding with antigen, thus exploration of nanobody structure is important for designing novel functional nanobody for target antigen. In the current version of the

database, the structure data of many nanobodies is vacant. More structure data are required. As fast increased nanobody structure data will be uploaded to PDB database, more nanobody 3D structure data will be collected and help us get an idea about the relationship between nanobody sequence and its 3D structure.

Additionally, as nanobody has family-specific sequence composition, with fast increasing sequence data of nanobodies, there is a possibility that one can mine the sequence to discover nanobody sequence feature and try to de novo design novel nanobody sequence. We will further build-up and embed protein homology modeling tools in our database to describe nanobody structure feature. The above work will be completed with more accumulation of nanobody sequence and novel tools involvement. At that time, iCAN will provide more powerful functions for users.

Conclusions

In this study, we present the first database of nanobody, iCAN. That is a comprehensive and free database of nanobody which not only collects the nanobody information, but also provides tools for sequence alignment, pattern feature creation and keywords-based search function. The information and tools in iCAN will enable researchers to work on nanobody more efficiently.

Availability and requirements

The datasets generated and/or analyzed during the current study are available in the iCAN repository, <http://ican.ils.seu.edu.cn>

Browser requirement: we recommend the use of the IE, Chrome or Firefox web browsers for an optional experience.

Datasets in iCAN are freely available for academic users. For all other uses, please contact with us by email: ican@seu.edu.cn

Additional files

Additional file 1: iCAN web manual.doc: the manual of iCAN website. (DOCX 91 kb)

Additional file 2: Structure and Submit pages of iCAN website. (a) Structure page of iCAN (b) Submit page of iCAN. (PDF 639 kb)

Additional file 3: Figure S1. The workflow of iCAN website. After the user's Web browser sends a HTTP query, the required HTML pages are returned in response to user's query. (PDF 294 kb)

Additional file 4: Figure S2. all data.xls: the source data of all curated nanobody. (XLS 2544 kb)

Abbreviations

CDR: Complementarity determining region; CSS: Cascading style sheets; EGFR: Human epidermal growth factor receptor; FR: Framework region; HTML: Hyper text markup language; iCAN: Institute collection and analysis of nanobodies

Acknowledgements

We are thankful to the anonymous reviewers for their valuable suggestions.

Funding

This study is supported by Natural Science Foundation of Jiangsu Province, Prospective Joint Research Project of Jiangsu, China (BK20151403, BY2015070-24) and the Fundamental Research Funds for the Central Universities (2242015K10011, 3231006205, 2242016K40094, 2242017K3DN23, 2242017K41041). We also thank for the support from Natural Science Foundation of China (81603019 and 81502975).

Authors' contributions

JL and WX conceived and supervised this work. JL, RZ and JZ constructed the database and tools. JZ, LX, HC, XJ, ZS, LZ, XH and JL collected data. JZ mainly curated the data. JZ drafted manuscript. JL, LZ, XH and WX modified the manuscript. All authors read and approved the manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 21 February 2017 Accepted: 9 October 2017

Published online: 17 October 2017

References

- Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, Bendahman N, Hamers R. Naturally occurring antibodies devoid of light chains. *Nature*. 1993;363(6428):446–8.
- Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature*. 1995;374(6518):168–73.
- Dumoulin M, Conrath K, Van Meirhaeghe A, Meersman F, Heremans K, Frenken LG, Muyldermans S, Wyns L, Matagne A. Single-domain antibody fragments with high conformational stability. *Protein Sci*. 2002;11(3):500–15.
- Lauwereys M, Arbabi Ghahroudi M, Desmyter A, Kinne J, Holzer W, De Genst E, Wyns L, Muyldermans S. Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. *EMBO J*. 1998;17(13):3512–20.
- Muyldermans S, Lauwereys M. Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies. *J Mol Recognit*. 1999;12(2):131–40.
- Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R. Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng*. 1994;7(9):1129–35.
- van der Linden RH, Frenken LG, de Geus B, Harmsen MM, Ruuls RC, Stok W, de Ron L, Wilson S, Davis P, Verrips CT. Comparison of physical chemical properties of llama VHH antibody fragments and mouse monoclonal antibodies. *Biochim Biophys Acta*. 1999;1431(1):37–46.
- Arbabi Ghahroudi M, Desmyter A, Wyns L, Hamers R, Muyldermans S. Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Lett*. 1997;414(3):521–6.
- De Genst EJ, Guillems T, Wellens J, O'Day EM, Waudby CA, Meehan S, Dumoulin M, Hsu ST, Cremades N, Verschueren KH, et al. Structure and properties of a complex of alpha-synuclein and a single-domain camelid antibody. *J Mol Biol*. 2010;402(2):326–43.
- Tang JCY, Szikra T, Kozorovitskiy Y, Teixeira M, Sabatini BL, Roska B, Cepko CL. A Nanobody-based system using fluorescent proteins as scaffolds for cell-specific gene manipulation. *Cell*. 2013;154(4):928–39.
- Trong ND, Peeters E, Muyldermans S, Charlier D, Hassanzadeh-Ghassabeh G. Nanobody (R)-based chromatin immunoprecipitation/micro-array analysis for genome-wide identification of transcription factor DNA binding sites. *Nucleic Acids Res*. 2013;41(5):e59.
- Leduc C, Si S, Gautier JJ, Gao ZH, Shibu ES, Gautreau A, Giannone G, Cognet L, Lounis B. Single-molecule imaging in live cell using gold nanoparticles. *Method Cell Biol*. 2015;125:13–27.
- Vaneycken I, Devoogdt N, Van Gassen N, Vincke C, Xavier C, Wernery U, Muyldermans S, Lahoutte T, Cavelliers V. Preclinical screening of anti-HER2 nanobodies for molecular imaging of breast cancer. *FASEB J*. 2011;25(7):2433–46.
- Holz JB. The TITAN trial—assessing the efficacy and safety of an anti-von Willebrand factor Nanobody in patients with acquired thrombotic thrombocytopenic purpura. *Transfus Apher Sci*. 2012;46(3):343–6.
- Strokappe N, Szyndol A, Aasa-Chapman M, Gorlani A, Quigley AF, Hulsik DL, Chen L, Weiss R, de Haard H, Verrips T. Llama antibody fragments recognizing various epitopes of the CD4bs neutralize a broad range of HIV-1 subtypes a, B and C. *PLoS One*. 2012;7(3):e33298.
- Van Roy M, Ververken C, Beirnaert E, Hoefman S, Kolkman J, Vierboom M, Breedveld E, t Hart BA, Poelmans S, Bontinck L, et al. The preclinical pharmacology of the high affinity anti-IL-6R Nanobody (R) ALX-0061 supports its clinical development in rheumatoid arthritis. *Arthritis Res Ther*. 2015;17:135.
- Li T, Bourgeois JP, Celli S, Glacial F, Le Sourd AM, Mecheri S, Weksler B, Romero I, Couraud PO, Rougeon F, et al. Cell-penetrating anti-GFAP VHH and corresponding fluorescent fusion protein VHH-GFP spontaneously cross the blood-brain barrier and specifically recognize astrocytes: application to brain imaging. *FASEB J*. 2012;26(10):3969–79.
- Van de Broek B, Devoogdt N, D'Hollander A, Gijls HL, Jans K, Lagae L, Muyldermans S, Maes G, Borghs G. Specific cell targeting with nanobody conjugated branched gold nanoparticles for photothermal therapy. *ACS Nano*. 2011;5(6):4319–28.
- Jobling SA, Jarman C, Teh MM, Holmberg N, Blake C, Verhoeyen ME. Immunomodulation of enzyme function in plants by single-domain antibody fragments. *Nat Biotechnol*. 2003;21(1):77–80.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

