Biochimica et Biophysica Acta xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Review

Antibody informatics for drug discovery ☆

Hiroki Shirai [a], Catherine Prades [b], Randi Vita [c], Paolo Marcatili [d], Bojana Popovic [e], Jianqing Xu [e], John P. Overington [f], Kazunori Hirayama [a], Shinji Soga [a], Kazuhisa Tsunoyama [a], Dominic Clark [f], Marie-Paule Lefranc [g], Kazuyoshi Ikeda [f,*]

[a] Molecular Medicine Research Laboratories, Drug Discovery Research, Astellas Pharma Inc., 21, Miyukigaoka, Tsukuba-shi, Ibaraki 305-8585, Japan
[b] Global Biotherapeutics, Bioinformatics, Sanofi-Aventis Recherche & Développement, Centre de recherche Vitry-sur-Seine, 13, quai Jules Guesde, BP 14, 94403 Vitry-sur-Seine Cedex, France
[c] Immune Epitope Database and Analysis Project, La Jolla Institute for Allergy & Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA
[d] Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Anker Engelunds Vej 1, 2800 Lyngby, Denmark
[e] MedImmune Ltd, Milstein Building, Granta Park, Cambridge CB21 6GH, UK
[f] The EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[g] IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire (LIGM), Université Montpellier 2, Institut de Génétique Humaine, UPR CNRS 1142, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France

ARTICLE INFO

Article history:
Received 27 March 2014
Received in revised form 4 July 2014
Accepted 11 July 2014
Available online xxxx

Keywords:
Antibody informatics
Antibody modeling
Antibody database
Antibody numbering
Drug discovery

ABSTRACT

More and more antibody therapeutics are being approved every year, mainly due to their high efficacy and antigen selectivity. However, it is still difficult to identify the antigen, and thereby the function, of an antibody if no other information is available. There are obstacles inherent to the antibody science in every project in antibody drug discovery. Recent experimental technologies allow for the rapid generation of large-scale data on antibody sequences, affinity, potency, structures, and biological functions; this should accelerate drug discovery research. Therefore, a robust bioinformatic infrastructure for these large data sets has become necessary. In this article, we first identify and discuss the typical obstacles faced during the antibody drug discovery process. We then summarize the current status of three sub-fields of antibody informatics as follows: (i) recent progress in technologies for antibody rational design using computational approaches to affinity and stability improvement, as well as *ab-initio* and homology-based antibody modeling; (ii) resources for antibody sequences, structures, and immune epitopes and open drug discovery resources for development of antibody drugs; and (iii) antibody numbering and IMGT. Here, we review "antibody informatics," which may integrate the above three fields so that bridging the gaps between industrial needs and academic solutions can be accelerated. This article is part of a Special Issue entitled: Recent advances in molecular engineering of antibody.

1. Introduction

Recent advances in experimental technologies allow researchers to rapidly generate an enormous amount of data using a variety of molecular biological methods. This data-driven science should be transformed into a model-based science. Pharmaceutical companies need to handle large biological data sets since molecular biology is significantly involved in drug discovery, development, and manufacturing. However, the expense involved in catching up with this rapid progress prevents any single company from adapting to these large biological data sets quickly and efficiently. Some efforts toward pre-competitive collaborations are underway. For instance, since sales of antibody therapeutics continue to rise, the EMBL European Bioinformatics Institute (EMBL-EBI) — Industry Programme [1] has focused on antibody or biologics informatics for both academia and industry. Four of the ten highest selling drugs from October 2012 to September 2013 were biologics, and the launch of biosimilars will make this situation even more interesting. There are many more informatics resources available for the analysis of small molecule therapeutics than for the antibody drug discovery process. In this paper, we review "antibody informatics" to create a synergetic resource of related efforts. We summarize some of the obstacles for antibody drug discovery and approaches to overcome these obstacles using antibody informatics. We first map the obstacles faced and their relevant informatics tools for the workflow of antibody drug discovery such as antibody modeling tools, antibody databases, and

Abbreviations: 2D, two-dimensional; 3D, three-dimensional; CDR, complementarity determining region; CMC, chemistry, manufacturing and control; CPCA, composite protein for clinical applications; CS, canonical structure; EMBL-EBI, EMBL European Bioinformatics Institute; FPIA, fusion protein for immune applications; FR, framework region; HTS, high throughput sequencing; IG, immunoglobulin; IgSF, immunoglobulin superfamily; IMGT, the international ImMunoGeneTics information system; MH, major histocompatibility; MhSF, MH superfamily; NGS, next generation sequencing; PD, pharmacodynamics; PK, pharmacokinetics; SPR, surface plasmon resonance; RMSD, root mean square deviation; TR, T cell receptor; VH, variable region of heavy chain; VL, variable region of light chain

☆ This article is part of a Special Issue entitled: Recent advances in molecular engineering of antibody.

* Corresponding author at: Level Five Co.,Ltd., Shiodome Shibarikyu Bldg., 1-2-3 Kaigan, Minato-ku, Tokyo 105-0022, Japan. Tel.: +81 3 5403 5917.
E-mail address: kazikeda@me.com (K. Ikeda).

http://dx.doi.org/10.1016/j.bbapap.2014.07.006
1570-9639/

accurate antibody numbering. We then discuss the current status of those three elements in detail.

2. Obstacles for drug discovery to be tackled using antibody informatics

Fig. 1 maps the obstacles faced during the workflow of antibody drug discovery and presents approaches used by antibody informatics tools. In the top box, a rough workflow is described: A host is immunized with a selected immunogen in order to obtain antigen specific antibodies, whose affinity and *in vitro* activity are measured. The researchers select a lead antibody among them mainly based on the *in vitro* activity, and then proceed to engineer (*e.g.*, through complementarity determining region (CDR) grafting) an optimized antibody. Next, pharmacokinetics (PK), pharmacodynamics (PD) and toxicological properties are measured for the selected antibody. Finally, mass production and chemistry, manufacturing, and control (CMC) are performed for clinical trials. The second and third boxes of Fig. 1 describe the obstacles and relevant informatics tools, respectively.

The design of therapeutic antibodies is a very difficult problem. First, obtaining an antibody specific to a target molecule can often prove difficult, and for some antigens, no specific antibody can be generated. Even when many antibodies with high affinity to their antigens are generated, they still may not possess enough functional activity due to their non-ideal binding sites. In order to solve this problem, careful design of the immunogen used for raising antibodies is required, for example, for generation of a stable active form or dissecting the functional domain. Unfortunately, some antibodies may have lower affinity or activity, and shorter PK than the expected values. Poor physicochemical properties, such as lower thermal stability and aggregation tendency, may cause further problems. To deal with these problems, engineering of the antibody, such as reducing the physicochemical problems [2,3], enhancing the affinity, or elongating half-life [4,5] is needed. Many antibody or antigen designs can be performed by computational ("*in silico*") approaches [6]. Knowledge, experience, and intuition can also be helpful in the antibody design process. In the latter case, the researchers, usually non-informatics researchers, need to be familiar with the antibody as well as its epitope. Antibody modeling and protein docking are often used to construct antibody–antigen tertiary structural models from amino acid sequences and play an important role both in the *in silico* design process and in understanding protein functions [6,7]. Although determination of the three-dimensional (3D) structure of a protein by X-ray crystallography has become easier, it still consumes a great deal of time and expense, and is not always successful. Now more than ever, as the number of antibody sequences available has been rapidly increasing, there is a demand for high quality of antibody modeling and protein docking, as a rapid suitable alternative for generating structural data is in demand.

Even after successful modeling and design, a functional antibody that met the PK/PD and toxicity criteria may have problems in later stages, such as mass production or CMC, because of its poor physicochemical properties. Therefore, new antibody selection criteria that predict difficulties at later stages are needed. Methodologies that prioritize therapeutic antibodies, based on evaluation of their druggability or developability by considering the features of antibody sequence, structure, and its physicochemical nature, would be ideal [8,9].

Administration of a therapeutic antibody is accompanied by the risks of developing an anti-antibody immune response. Methodological developments in CDR grafting and transgenic animals for generating humanized and human antibodies, respectively, have reduced the risk of immunogenicity in clinical trials. However, these advances are still not perfect. Prediction and elimination of T cell epitopes are two ways to tackle this problem, but the mechanisms of immunogenicity are complex and the causes are still unclear. Antibody aggregation can also be an

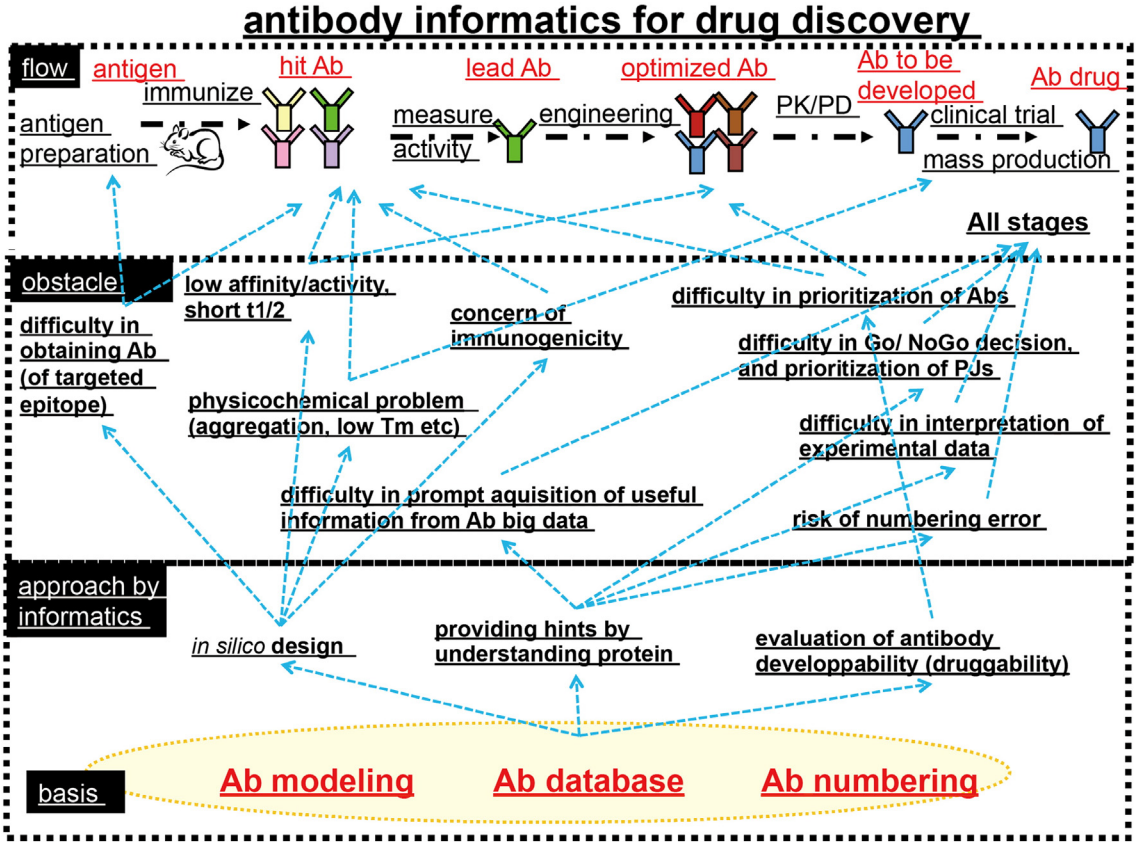

Fig. 1. Antibody informatics approaches to antibody drug discovery. $t_{1/2}$: half-life and Ab: antibody.

Please cite this article as: H. Shirai, et al., Antibody informatics for drug discovery, Biochim. Biophys. Acta (2014), http://dx.doi.org/10.1016/j.bbapap.2014.07.006