abYsis: Integrated Antibody Sequence and Structure—Management, Analysis, and Prediction

Mark B. Swindells[1], **Craig T. Porter**[2], **Matthew Couch**[2], **Jacob Hurst**[2], **K.R. Abhinandan**[2], **Jens H. Nielsen**[3], **Gary Macindoe**[3], **James Hetherington**[3] and **Andrew C.R. Martin**[2]

1 - *Chemogenomix,* 12 High Street, Easton on the Hill PE9 3LR, UK
2 - *Institute of Structural and Molecular Biology,* Division of Biosciences, University College London, Gower Street, London WC1E 6BT, UK
3 - *Research Software Development Group,* Research IT Services, University College London, Gower Street, London WC1E 6BT, UK

Correspondence to Andrew C.R. Martin: andrew@bioinf.org.uk
http://dx.doi.org/10.1016/j.jmb.2016.08.019
Edited by Michael Sternberg

Abstract

abYsis is a web-based antibody research system that includes an integrated database of antibody sequence and structure data. The system can be interrogated in numerous ways—from simple text and sequence searches to sophisticated queries that apply 3D structural constraints. The publicly available version includes pre-analyzed sequence data from the European Molecular Biology Laboratory European Nucleotide Archive (EMBL-ENA) and Kabat as well as structure data from the Protein Data Bank. A researcher's own sequences can also be analyzed through the web interface.

A defining characteristic of abYsis is that the sequences are automatically numbered with a series of popular schemes such as Kabat and Chothia and then annotated with key information such as complementarity-determining regions and potential post-translational modifications. A unique aspect of abYsis is a set of residue frequency tables for each position in an antibody, allowing "unusual residues" (those rarely seen at a particular position) to be highlighted and decisions to be made on which mutations may be acceptable. This is especially useful when comparing antibodies from different species.

abYsis is useful for any researcher specializing in antibody engineering, especially those developing antibodies as drugs.

abYsis is available at www.abysis.org.

Introduction

Antibodies are among the most important classes of proteins involved in the adaptive immune system. They act as adapter molecules between the virtually infinite range of possible antigens and the constant effector molecules such as Fc receptors and C1q of the complement system. Through the DNA-level splicing of V and J gene segments in the light chain and V, D, and J gene segments in the heavy chain, they create a repertoire of around 10^{11} possible sequences [1]. The fact that antibodies are capable of binding to a virtually infinite set of antigens, generally with high specificity and affinity, has led to a huge interest in using them as drugs in the treatment of human disease [2].

Therapeutic antibodies can act through a variety of mechanisms. First, they may act simply to bind to a target in the way that most small-molecule drugs work. For example, they may bind to receptors acting as agonists or antagonists or to enzymes acting as inhibitors or as allosteric activators. Since they have a much larger interacting surface than typical small-molecule drugs (which generally bind into pockets on a protein surface), they can also interact with much flatter protein surfaces and can therefore be used to bind to hormones or disrupt protein–protein interactions. Second, antibodies can be used to trigger the immune system in the normal way to kill cells such as tumor cells or virus-infected cells. For example, the antibody, Palivizumab (Synergis), is used to prevent respiratory syncytial virus infection in high-risk infants.

Third, antibodies can be used as a delivery mechanism, either through antibody–drug conjugates with cytotoxic moieties or through "antibody-directed enzyme prodrug therapy", which aims to restrict the action of a cytotoxic drug to cancer sites [3]. Fourth, more recent developments have led to bispecific antibodies capable of binding two antigens, either for convenience of having a single agent or for deliberately cross-linking different molecules [4–6].

It has been estimated that around a third of all drugs in development are monoclonal antibodies [7,8], and as of May 2016, according to the Antibody Society, 55 monoclonal antibody products had been approved in the US or Europe with around 4 new products being approved per year [9]. Therapeutic antibodies have been used for a wide variety of conditions such as cancer [10], transplant rejection [11], rheumatoid arthritis [12], antiviral prophylaxis [13], and Crohn's disease [14].

Despite the success of antibodies in the clinic, there are a number of problems in developing successful therapeutics. In addition to high affinity and specificity, antibodies must be non-immunogenic and have sufficiently high bioavailability. Moreover, stability is important to avoid denaturation and aggregation, not only for long shelf life and persistent bioavailability but also because denaturation and aggregation can lead to increased immunogenicity [15–17]. Consequently, there is still room for improvement in building better antibodies having higher affinity and lower immunogenicity [18].

Traditionally, producing fully human antibodies has been difficult, while antibodies from other species are likely to lead to an immune response (the "Human Anti-Mouse Antibody" response [19,20] or, more generally, the "Anti-Antibody Response" [21]). This prevents the repetitive administration of the antibody for treatment and may lead to anaphylactic shock. Traditional methods for producing monoclonal antibodies [22] do not work in humans [23,24]. Techniques such as the Selected Lymphocyte Antibody Method (SLAM) [25] allow the screening of B cells harvested from peripheral blood, spleen, bone marrow, or lymph nodes and, in addition to being applicable to nonhuman species, can be used to identify therapeutic antibody candidates from human peripheral blood. While this can be useful for creating anti-infective agents, the majority of targets are human proteins against which humans do not produce an immune response. While the production of antibodies from phage display libraries has been very popular [26,27], the success of these agents in the clinic has so far been limited. Another more recent development is the production of fully human antibodies from transgenic mice [23], where the antibody repertoire is fully human, but this route is extremely expensive. Consequently, murine (or other nonhuman) antibodies are frequently engineered to make them appear "more human". The process of humanization aims to reduce the immunogenicity of nonhuman antibodies by using human constant and variable regions as an "acceptor" into which nonhuman complementarity-determining regions (CDRs) from the "donor" antibody with the required specificity are inserted [28,29]. However, in general, naïvely grafted antibodies do not bind well, and some framework residues need to be "back-mutated" to the equivalent donor residue in order to restore the binding affinity [30].

Despite the interest in antibodies and the need to engineer them, the range of freely available web-based tools and resources to aid in this process is limited. Probably the best known resource is the Immunogenetics Database (IMGT) [31], a comprehensive collection of data and services available over the web. There are a number of web-based antibody modeling servers, the performance of which was recently compared in the second antibody modeling assessment (AMA-II) [32]. These include PiGS[†] [33], SAbPred[‡] [34] Rosetta Antibody [35,36], Kotai Antibody Builder[§] [37], and SmrtMolAntibody[¶] [38]. SAbPred also includes antibody numbering [39] and epitope and paratope prediction. Other resources include SAbDab[††] [40], a database of antibody structures; Tabhu[‡‡], a set of tools to aid with antibody humanization [41]; and Digit[§§] [42], a database of antibody variable domain sequences.

For many years, we have offered a number of web-based antibody tools including KabatMan [43], SACS [44], AbNum [7], expressed-sequence-scored humanness [45], germline-scored humanness [46], VH/VL packing-angle prediction [47], assignment of Chothia canonical classes [48], and testing a sequence for unusual residues [43]. These are all available via our Website[¶¶].

Here, we describe our abYsis system, a database of antibody sequence and structure that also integrates many of the analysis tools described above.

Results

abYsis contains antibody protein sequences from EMBLIG, which contains antibody information extracted from the EMBL-ENA databank (Couch, Porter, Swindells, and Martin, in preparation), the Kabat collection [49] (which has not been updated since July 2000), and the Protein Data Bank (PDB) [50]. Data from other sources such as IMGT can also be loaded, but the vast majority of sequence information in IMGT is contained in EMBLIG or the PDB, which we handle separately. In fact, EMBLIG contains many antibody sequences that are not present in IMGT. Comparing the two datasets as described in the Materials and Methods, we found that the two datasets overlap by 28,388 sequences. EMBLIG contains 94.8% of the sequences in IMGT, but 27.9% of sequences in EMBLIG were not present in IMGT; EMBLIG contains 1.3 times more sequences than IMGT (see Table 1).

Table 1. Comparison of the data in IMGT (release 16 Jul 2016) and EMBLIG (release 18 Apr 2016). Sequences in IMGT are filtered to include only antibody variable domain sequences derived from EMBL-ENA. Both datasets were filtered to ensure that translations were at least 50 residues in length.

Total sequences in EMBLIG	127,450
Total sequences in IMGT	96,970
Total sequences present in both datasets	91,898
Total missing from EMBLIG that are present in IMGT	5072
Total missing from IMGT that are present in EMBLIG	35,552

The ability to integrate data from other sources is particularly useful for people who decide to have abYsis in-house as they can include their own sequence data in the resource. Table 2 shows the processed content of the latest release of abYsis (V2.7.3) as of May 2015. Automated numbering of sequences using abnum [7] is key to the annotation of sequences—the current release successfully numbers 122,134 sequences. The system also provides automated pairing of light and heavy chains. Resources such as EMBL-ENA (and therefore EMBLIG) and Kabat do not contain a clear pairing field to link data for light and heavy chains, although a "DR/partner" field is being populated in IMGT, and we exploit this when reading IMGT data. Consequently, as described in the Materials and Methods, we attempt to take a cautious approach to pairing for other data sources. For EMBLIG data, we use the clone information to perform the pairing: 7110 chains are paired into 3555 antibodies, while antibodies containing a single chain occur 119,312 times because this clone information is only present for ~56% of chains in EMBLIG. The most highly represented species are *Homo sapiens* (88,662), *Mus musculus* (41,254), *Macaca mulatta* (3561), *Oryctolagus cuniculus* (2668), *Equus caballus* (1138), *Oncorhynchus mykiss* (979), *Lama glama* (955), *Camelus dromedarius* (930), and *Ginglymostoma cirratum* (893), with synthetic constructs coming in fourth place with 2968 entries.

The reason that numbering is key to the annotation is that it enables us to identify regions of the sequences (CDRs and frameworks). It provides an implicit sequence alignment (which can be displayed in abYsis using the Javascript Sequence Alignment Viewer [51]) and the mapping of sequence to structure, allowing us to look at the frequency of residue types at each position and to identify key residues that define the conformation of the CDRs [52] or which are important in antibody humanization [30,53].

abYsis can be used in three main ways: (a) by searching the database to find information about sequences (and structures) including detailed annotations, identification of unusual residues, post-translational modification sites, etc.; (b) by analyzing trends across sequences in the database; and (c) by entering your own sequences for analysis. These three approaches are described below.

Searching the database

The main search page allows searches based on identifiers, names, antigens (where described and parsable in the source data), references, authors, organisms, etc., and searches can be restricted to a given dataset (PDB, Kabat, EMBLIG, or other datasets that have been included) or a given class of chain. In addition, searches can be structure-based by requiring CDRs to match specific canonical classes or by specifying residues within a given distance of a particular position; for example, one could find all antibodies with a lysine within 4 Å of residue L23. This can be done either in known structures or by using averaged distances calculated across all known structures to extrapolate these constraints to sequences for which no structure is available. Finally, one can also search on the basis of sequence. Searches can specify motifs (either exact or using regular expressions) present throughout the sequence or within particular regions (CDRs or frameworks), region lengths, or residues present or not present at particular positions.

A simple example search is shown in Fig. 1, together with two views of the results—as a simple table or as a sequence alignment. The results come from searching for "hyhel" and therefore display the HyHEL set of antibodies binding to hen egg-white lysozyme described by Smith-Gill and colleagues [54].

Having selected a particular sequence of interest, we provide much more detailed information in what

Table 2. Processed content of the abYsis database as of May 2016. In addition, DNA data are available for sequences from EMBLIG and Kabat. The National Center for Biotechnology Information (NCBI) Germline data and germline information from V-BASE are also included.

Data	Protein Sequences			Classified			
Sources	Sequences	≥70 aa	Numbered	Total	Heavy	Light	Paired
EMBLIG	126,422	119,422	101,969	122,745	92,179	30,566	3555
Kabat	19,399	14,114	12,631	19,399	11,842	7557	2572
PDB	7712	7712	7534	7534	3820	3714	3397
Total	153,533	141,248	122,134	149,678	107,841	41,837	9524
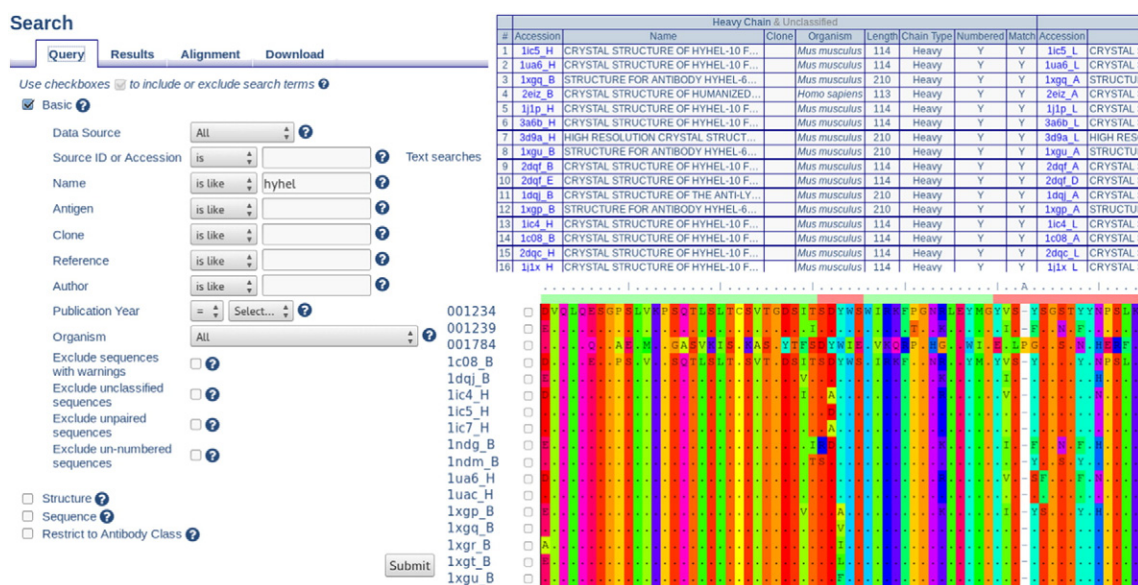
Fig. 1. Simple searches of abYsis. On the left is the main search page showing the default, simple options. Searching for "hyhel" in the Name field results in a table of hits shown on the upper right. For clarity, this has been truncated—the full view would show the complete information for the light chains to the right of the heavy chains, where paired sequences have been found. Bottom right shows the alignment view of the sequences.

we term the "Key Annotation" page, which provides a set of tabs with different information. The initial summary page provides the sequence with Chothia numbering applied and the Chothia structural loops highlighted. Unusual residues occurring in <1% of sequences across all species are highlighted, as are a range of predicted post-translational modification sites. Information about the sequence (accession code, data source, dates, name, organism, chain type, partner chain, and references) is also provided. A button at the bottom of the page allows the sequence to be searched against all the other sequences using BLAST.

The "Numbering & Regions" tab (part of which is shown in Fig. 2a) gives more control over numbering schemes, allowing Kabat, Chothia, or Martin (Enhanced Chothia) [7] schemes and displaying CDRs/structural loops according to definitions provided by Kabat, Chothia, AbM (useful for modeling purposes), and contacts [55]. On the right of the panel, a residue distribution graph for the currently selected residue—in this case, residue H1—is shown. As the graph shows, the most common residues (across all species) are glutamate (E) and glutamine (Q); the amino acid present in this particular sequence is highlighted with an arrow above the relevant bar.

The "Unusual Residues" tab allows the same information to be provided but allows the comparison to be restricted to a specific species, while the "Humanization" tab gives a much more detailed view of this information. Again, clicking on an amino acid

will show a graph of its distribution, which can be restricted to a particular species. The "heatmap", which is part of the humanization tab and shown in part in Fig. 2b, gives a summary of the graphical view of residue frequencies for each position in the sequence. Two thresholds are defined using sliders at the bottom of the page; for each position, amino acid types with a frequency below the lower threshold (i.e., unusual residues) are shown in red, those above the higher threshold (i.e., common) are shown in green, with the remaning residues shown in yellow, the CDRs being shown in darker shades. At the top of the heatmap, the sequence is shown with colors based on a first reference sequence and then with colors against a second reference sequence, which is also used for the rest of the heatmap. Clicking on a residue in the middle of the heatmap will update the sequence shown at the bottom, which can then be exported. Typically, one might set mouse as the first reference and human as the second reference, thus allowing the identification of unusual human residues.

Other tabs provided include (a) assignment of Chothia conformational canonical classes based on sequence; (b) a "Structure" tab, which is only displayed if the selected sequence comes from a PDB file, in which the structure can be viewed and rotated; (c) a "Germline" tab, which maps a sequence to the closest germline sequences; and (d) a "Humanness" tab, which implements the method of Abhinandan and Martin for assessing how human-like a sequence is [45].
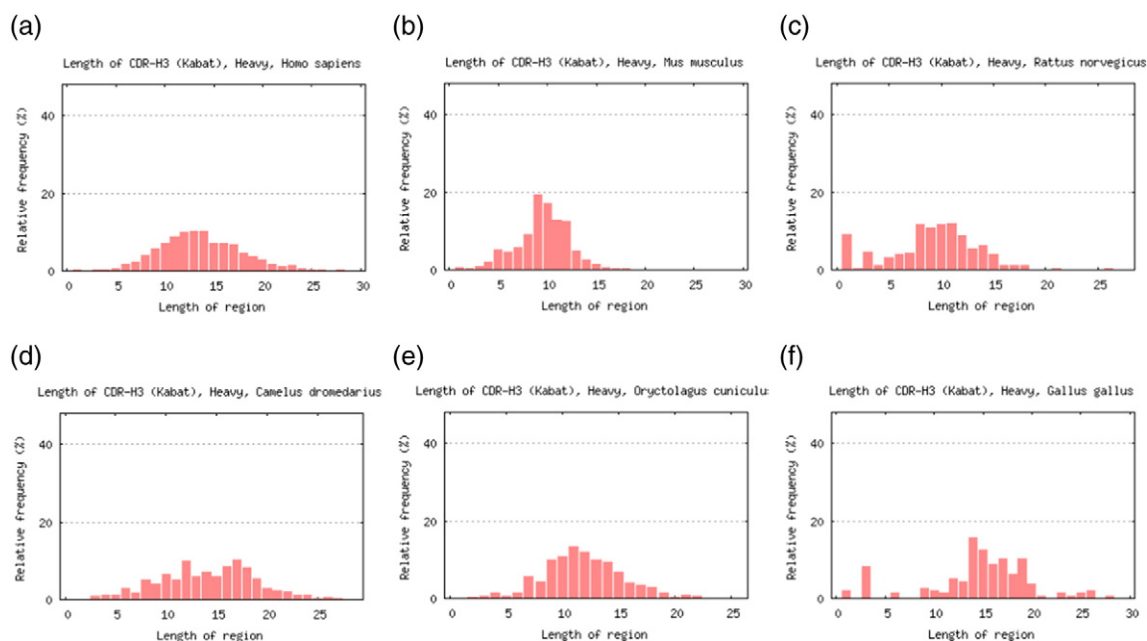
(a)

(b)
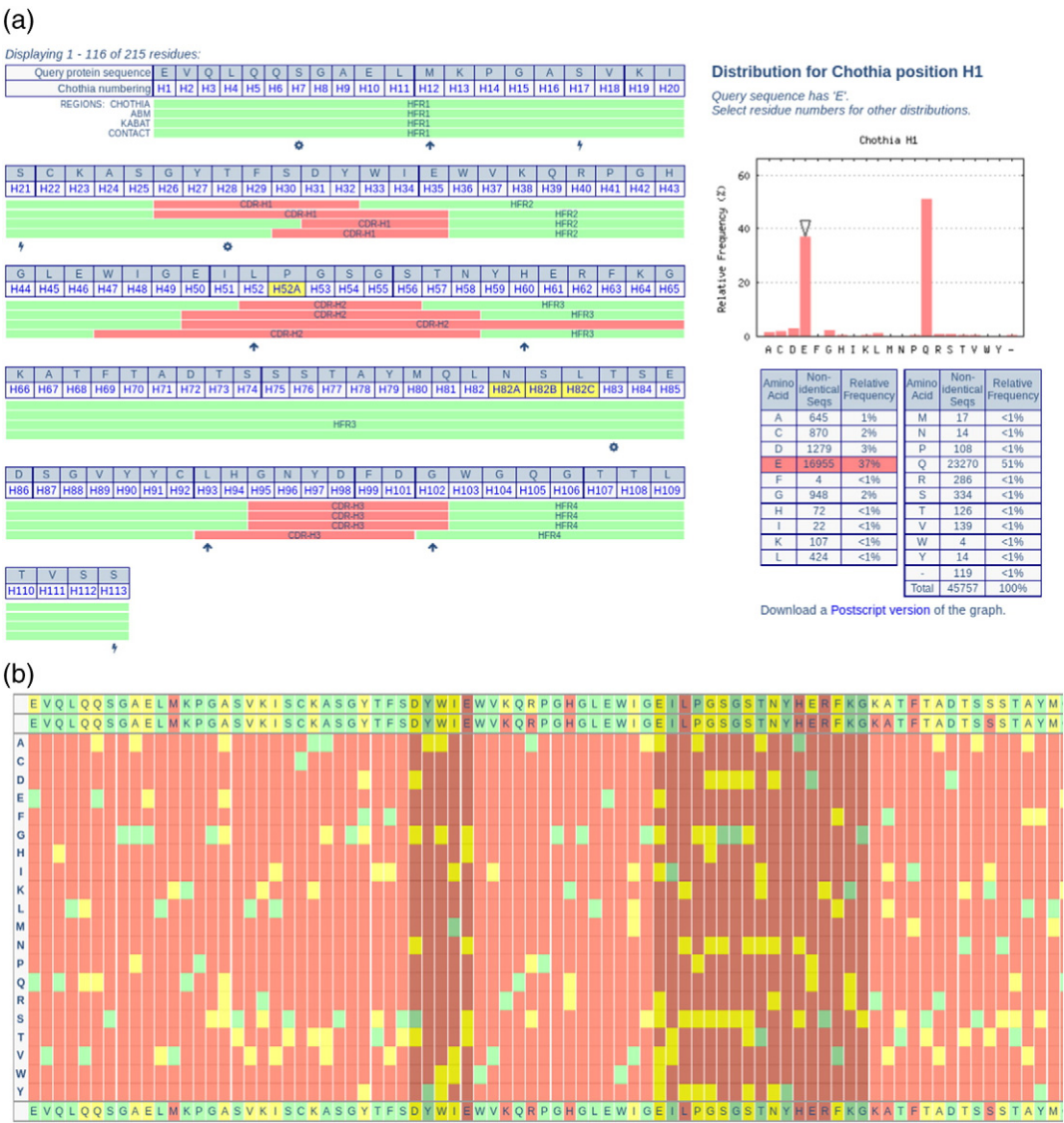
Fig. 2. Key annotation results for the heavy chain of antibody HyHEL-5. (a) the "Numbering & Regions" tab—unusual residues are highlighted with an upwards arrow; predicted CK2 phosphorylation sites with a gear wheel and predicted PKC phosphorylation sites with a lightning symbol. (b) The "Humanize" tab which shows the sequence horizontally and the summarized residue frequency information for a chosen species vertically below each residue.

Analyzing trends

The distributions search page allows the distributions of residues at particular positions to be analyzed outside the context of a particular sequence. It also allows the distributions of the lengths of regions (CDRs and frameworks) to be analyzed. This has been an area of some considerable previous interest. For example, Wu *et al.* [56] analyzed the distribution of the lengths of CDR-H3 loops in humans and other species and concluded that "human sequences vary from 2 to 26 amino acids residues, but less extensively in other species". They updated their analysis in 1993 [57],

refining the analysis in an attempt to look at defined specficities. Similar detailed analyses of length distributions were performed by Collis *et al.* [58] and Barrios *et al.* [59].

Such analyses become straightforward using abYsis, and in Fig. 3, we show the length distributions of CDR-H3 in humans, mice, rats, camel, rabbit, and chicken. The graph confirms earlier observations that mouse antibodies tend to have shorter CDR-H3 lengths than humans, but it also shows that this trend extends to rat antibodies. Despite the recent popularity of rabbit antibodies in generating potential clinical antibodies, the CDR-H3 lengths still tend to be shorter than human antibodies, but camels and

(a)

Length of CDR-H3 (Kabat), Heavy, Homo sapiens

(b)

Length of CDR-H3 (Kabat), Heavy, Mus musculus

(c)

Length of CDR-H3 (Kabat), Heavy, Rattus norvegicus

(d)

Length of CDR-H3 (Kabat), Heavy, Camelus dromedarius

(e)

Length of CDR-H3 (Kabat), Heavy, Oryctolagus cuniculu

(f)

Length of CDR-H3 (Kabat), Heavy, Gallus gallus

Fig. 3. Distributions of CDR-H3 lengths in (a) humans, (b) mice, (c) rats, (d) camel, (e) rabbit, and (f) chicken.

chickens tend to have CDR-H3s that are even longer than humans.

Entering your own sequences

The sequence input pages allow one or more protein and/or DNA sequences to be entered. These are numbered and analyzed in exactly the same way as sequences that are stored in the database. Consequently, all the "key annotation" information described above is also available for these. Sequences can also be searched against the database using BLAST to find the most similar antibodies.

The "Humanization" tab is of particular interest since it allows one to upload a naïvely humanized antibody (i.e., where the CDRs have been grafted from a donor species onto a human acceptor framework) and compare this with unusual residues identified in the mouse framework, which may be functionally important. One can also look for unusual residues in the human acceptor, which may have an effect on binding or lead to immunogenicity.

The locally installed version of abYsis also allows sequences to be uploaded and added to the database.

Materials and methods

abYsis is based around a PostgreSQL relational database populated with sequence and structure data. All sequences are numbered using Kabat, Chothia, and Martin numbering schemes using abnum [7],

and the database is populated using scripts written in Perl. The web interface is implemented using Perl and Mason, and the Javascript Sequence Alignment Viewer [51] is used to display sequence alignments. JSmol [60] is used for structure display.

Comparison of EMBLIG and IMGT

A Perl script was written to compare the content of EMBLIG and IMGT. The IMGT data file contains sequences that are not antibodies and, while mostly coming from EMBL-ENA, also comes from sources such as the PDB (that we treat separately). Consequently the IMGT data were filtered to ensure that the sequence (a) was present in EMBL-ENA, (b) was labeled as IG-Light or IG-Heavy, (c) contains a variable domain, (d) has a translation of at least 50 residues provided, and (e) is not a pseudogene. EMBLIG data were also filtered to include only sequences with translations of at least 50 residues.

Pairing antibody chains

For Kabat data, we pair antibody chains based on the provided "name" field and author information to pair light and heavy chains as described by Martin [43]. For pairing chains within PDB files, we examine the proximity of residues normally found in the interface between light and heavy chains. For EMBLIG data derived from EMBL-ENA, there is no clear name field so we currently use the clone name given in the "FT/source/clone=" field. In addition to matching the clone

name, pairs of light and heavy chains must also share publication data.

Conclusion

abYsis provides a unique, integrated web-based environment for exploring antibody sequence and structure, bringing together data from different public sources. In addition, users who install the system locally are able to integrate their own sequence data. A key advantage of abYsis is that we try to pair light and heavy chain sequences. However, we take a cautious approach to this, and in the case of EMBLIG data derived from EMBL-ENA, we use clone information that is only present for ~56% of chains. Consequently, future work will include exploiting the information in the "FT/CDS/product=" information, which is used to provide a "Name" in abYsis. However, this will require cleanup to remove uninformative text (such as "immunoglobulin", "gamma", "heavy", "light", or "chain") and, for safety, would need to be combined with author and/or publication information. The system provides annotation of sequences stored in the system, or which a user uploads, and allows individual sequences to be compared with those in the underlying database. Integration of sequence and structure data allows the analysis of sequence data in a structural context, and the system provides tools that can aid with antibody engineering.

Acknowledgments

Present address: J. Hurst, Immunocore Limited, 101 Park Drive, Milton Park, Abingdon, Oxon OX14 4RY, UK.
Present address: K.R. Abhinandan, Integrated Biologics Profiling Unit, Novartis Pharma AG, Basel 4057, Switzerland.

†www.biocomputing.it/pigs
‡opig.stats.ox.ac.uk/webapps/sabdab-sabpred/Welcome-SAbPred.php
§kotaiab.org
¶www.macromoltek.com
††opig.stats.ox.ac.uk/webapps/sabdab-sabpred/Welcome.php
‡‡circe.med.uniroma1.it/tabhu/
§§circe.med.uniroma1.it/digit/
¶¶www.bioinf.org.uk/abs/

Abbreviations used:
CDR, complementarity-determining region; PDB, Protein Data Bank; IMGT, Immunogenetics Database; EMBL-ENA, European Molecular Biology Laboratory European Nucleotide Archive.

References

[1] J. Glanville, W. Zhai, J. Berka, D. Telman, G. Huerta, G.R. Mehta, I. Ni, L. Mei, P.D. Sundar, G.M.R. Day, D. Cox, A. Rajpal, J. Pons, Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 20,216–20,221.

[2] T. Yamada, Therapeutic monoclonal antibodies, Keio J. Med. 60 (2011) 37–46.

[3] K.D. Bagshawe, Targeting: the ADEPT story so far, Curr. Drug Targets 10 (2009) 152–157.

[4] U. Reusch, K. Harrington, C. Gudgeon, I. Fucek, K. Ellwanger, M. Weichel, S. Knackmuss, E. Zhukovsky, J.A. Fox, L.A. Kunkel, J. Guenot, R.B. Walter, Characterization of CD33/CD3 tetravalent bispecific tandem diabodies (TandAbs) for the treatment of acute myeloid leukemia, Clin. Cancer Res. (2016), http://dx.doi.org/10.1158/1078-0432.CCR-16-0350 (Epub ahead of print]).

[5] X. Chen, N. Haddish-Berhane, P. Moore, T. Clark, Y. Yang, H. Li, D. Xuan, H. Barton, A. Betts, F. Barletta, Mechanistic projection of first in human dose for bispecific immuno-modulatory P-cadherin LP-DART—an integrated PK/PD modeling approach, Clin. Pharmacol. Ther. 100 (2016) 232–241.

[6] T. Torres, M. Romanelli, A. Chiricozzi, A revolutionary therapeutic approach for psoriasis: bispecific biological agents, Expert Opin. Investig. Drugs 25 (2016) 751–754.

[7] K.R. Abhinandan, A.C.R. Martin, Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains, Mol. Immunol. 45 (2008) 3832–3839.

[8] J.M. Reichert, V.E. Valge-Archer, Development trends for monoclonal antibody cancer therapeutics, Nat. Rev. Drug Discov. 6 (2007) 349–356.

[9] D.M. Ecker, S.D. Jones, H.L. Levine, The therapeutic monoclonal antibody market, MAbs 7 (2015) 9–14.

[10] R.A. Larson, M. Boogaerts, E. Estey, C. Karanes, E.A. Stadtmauer, E.L. Sievers, P. Mineur, J.M. Bennett, M.S. Berger, C.B. Eten, M. Munteanu, M.R. Loken, J.J.M. van Dongen, I.D. Bernstein, F.R. Appelbaum, Antibody-targeted chemotherapy of older patients with acute myeloid leukemia in first relapse using Mylotarg (gemtuzumab ozogamicin), Leukemia 16 (2002) 1627–1636.

[11] J.L. Berard, R.L. Velez, R.B. Freeman, S.M. Tsunoda, A review of interleukin-2 receptor antagonists in solid organ transplantation, Pharmacotherapy 19 (1999) 1127–1137.

[12] R. Maini, E.W.S. Clair, F. Breedveld, D. Furst, J. Kalden, M. Weisman, J. Smolen, P. Emery, G. Harriman, M. Feldmann, P. Lipsky, Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) *versus* placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial, Lancet 354 (1999) 1932–1939.

[13] X. Sáez-Llorens, E. Castaño, D. Null, J. Steichen, P.J. Sánchez, O. Ramilo, F.H. Top, E. Connor, Safety and pharmacokinetics of an intramuscular humanized monoclonal antibody to respiratory syncytial virus in premature infants and infants with bronchopulmonary dysplasia, Pediatr. Infect. Dis. J. 17 (1998) 787–791.

[14] W.J. Sandborn, S.B. Hanauer, Antitumor necrosis factor therapy for inflammatory bowel disease: a review of agents, pharmacology, clinical results, and safety, Inflamm. Bowel Dis. 5 (1999) 119–133.

[15] W. Wang, S. Singh, D.L. Zeng, K. King, S. Nema, Antibody structure, instability, and formulation, J. Pharm. Sci. 96 (2007) 1–26.

[16] I.R. Mackay, F.S. Rosen, P.J. Delves, I.M. Roitt, The immune system, N. Engl. J. Med. 343 (2000) 37–49.

[17] O.H. Brekke, I. Sandlie, Therapeutic antibodies for human diseases at the dawn of the twenty-first century, Nat. Rev. Drug Discov. 2 (2003) 52–62.

[18] G.J. Weiner, Building better monoclonal antibody-based therapeutics, Nat. Rev. Cancer 15 (2015) 361–370.

[19] R.W. Schroff, K.A. Foon, S.M. Beatty, R.K. Oldham, A.C. Morgan, Human anti-murine immunoglobulin responses in patients receiving monoclonal antibody therapy, Cancer Res. 45 (1985) 879–885.

[20] D.L. Shawler, R.M. Bartholomew, L.M. Smith, R.O. Dillman, Human immune response to multiple injections of murine monoclonal IgG, J. Immunol. 135 (1985) 1530–1535.

[21] W.Y.K. Hwang, J. Foote, Immunogenicity of engineered antibodies, Methods 36 (2005) 3–10.

[22] G. Kohler, C. Milstein, Continuous cultures of fused cells secreting antibody of predefined specificity, Nature (London) 256 (1975) 495–497.

[23] L.L. Green, Antibody engineering via genetic engineering of the mouse: XenoMouse strains are a vehicle for the facile generation of therapeutic human monoclonal antibodies, J. Immunol. Methods 231 (1999) 11–23.

[24] G. Winter, C. Milstein, Man-made antibodies, Nature (London) 349 (1991) 293.

[25] S. Tickle, R. Adams, D. Brown, M. Griffiths, D. Lightwood, A. Lawson, High-throughput screening for high affinity antibodies, J. Lab. Autom. 14 (2009) 303–307.

[26] J. McCafferty, A.D. Griffiths, G. Winter, D.J. Chiswell, Phage antibodies—filamentous phage displaying antibody variable domains, Nature (London) 348 (1990) 552–554.

[27] A.E. Nixon, D.J. Sexton, R.C. Ladner, Drugs derived from phage display: from candidate identification to clinical practice, MAbs 6 (2014) 73–85.

[28] P.T. Jones, P.H. Dear, J. Foote, M.S. Neuberger, G. Winter, Replacing the complementarity-determining regions in a human antibody with those from a mouse, Nature (London) 321 (1986) 522–525.

[29] M. Verhoeyen, C. Milstein, G. Winter, Reshaping human antibodies: grafting an antilysozyme activity, Science 239 (1988) 1534–1536.

[30] L. Riechmann, M. Clark, H. Waldmann, G. Winter, Reshaping human antibodies for therapy, Nature 332 (1988) 323–327.

[31] M.-P. Lefranc, V. Giudicelli, P. Duroux, J. Jabado-Michaloud, G. Folch, S. Aouinti, E. Carillon, H. Duvergey, A. Houles, T. Paysan-Lafosse, S. Hadi-Saljoqi, S. Sasorith, G. Lefranc, S. Kossida, IMGT, the international ImMunoGeneTics information system 25 years on, Nucleic Acids Res. 43 (2015) D413–D422.

[32] J.C. Almagro, A. Teplyakov, J. Luo, R.W. Sweet, S. Kodangattil, F. Hernandez-Guzman, G.L. Gilliland, Second antibody modeling assessment (AMA-II), Proteins Struct. Funct. Genet. 82 (2014) 1553–1562.

[33] P. Marcatili, A. Rosi, A. Tramontano, PIGS: automatic prediction of antibody structures, Bioinformatics 24 (2008) 1953–1954.

[34] J. Dunbar, K. Krawczyk, J. Leem, C. Marks, J. Nowak, C. Regep, G. Georges, S. Kelm, B. Popovic, C.M. Deane, SAbPred: a structure-based antibody prediction server, Nucleic Acids Res. 44 (W1) (2016) W474–W478.

[35] A. Sircar, E.T. Kim, J.J. Gray, RosettaAntibody: antibody variable region homology modeling server, Nucleic Acids Res. 37 (2009) W474–W479.

[36] B.D. Weitzner, D. Kuroda, N. Marze, J. Xu, J.J. Gray, Blind prediction performance of RosettaAntibody 3.0: grafting, relaxation, kinematic loop modeling, and full CDR optimization, Proteins 82 (2014) 1611–1623.

[37] H. Shirai, K. Ikeda, K. Yamashita, Y. Tsuchiya, J. Sarmiento, S. Liang, T. Morokata, K. Mizuguchi, J. Higo, D.M. Standley, H. Nakamura, High-resolution modeling of antibody structures by a combination of bioinformatics, expert knowledge, and molecular simulations, Proteins 82 (2014) 1624–1635.

[38] M. Berrondo, S. Kaufmann, M. Berrondo, Automated Aufbau of antibody structures from given sequences using Macromoltek's SmrtMolAntibody, Proteins 82 (2014) 1636–1645.

[39] J. Dunbar, C.M. Deane, ANARCI: antigen receptor numbering and receptor classification, Bioinformatics 32 (2016) 298–300.

[40] J. Dunbar, K. Krawczyk, J. Leem, T. Baker, A. Fuchs, G. Georges, J. Shi, C.M. Deane, SAbDab: the structural antibody database, Nucleic Acids Res. 42 (2014) D1140–D1146.

[41] P.P. Olimpieri, P. Marcatili, A. Tramontano, Tabhu: tools for antibody humanization, Bioinformatics 31 (2015) 434–435.

[42] A. Chailyan, A. Tramontano, P. Marcatili, A database of immunoglobulins with integrated tools: DIGIT, Nucleic Acids Res. 40 (2012) D1230–D1234.

[43] A.C. Martin, Accessing the Kabat antibody sequence database by computer, Proteins 25 (1996) 130–133.

[44] L.C. Allcorn, A.C.R. Martin, SACS—self-maintaining database of antibody crystal structure information, Bioinformatics 18 (2002) 175–181.

[45] K.R. Abhinandan, A.C.R. Martin, Analyzing the "degree of humanness" of antibody sequences, J. Mol. Biol. 369 (2007) 852–862.

[46] P. Thullier, O. Huish, T. Pelat, A.C.R. Martin, The humanness of macaque antibody sequences, J. Mol. Biol. 396 (2010) 1439–1450.

[47] K.R. Abhinandan, A.C.R. Martin, Analysis and prediction of VH/VL packing in antibodies, Protein Eng. Des. Sel. 23 (2010) 689–697.

[48] A.C. Martin, J.M. Thornton, Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies, J. Mol. Biol. 263 (1996) 800–815.

[49] G. Johnson, T.T. Wu, Kabat database and its applications: future directions, Nucleic Acids Res. 29 (2001) 205–206.

[50] H.M. Berman, T. Battistuz, T.N. Bhat, W.F. Bluhm, P.E. Bourne, K. Burkhardt, Z. Feng, G.L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J.D. Westbrook, C. Zardecki, The Protein Data Bank, Acta Crystallogr. D Biol. Crystallogr. 58 (2002) 899–907.

[51] A.C.R. Martin, Viewing multiple sequence alignments with the JavaScript sequence alignment viewer (JSAV), F1000Res. 3 (2014) 249.

[52] B. Al-Lazikani, A.M. Lesk, C. Chothia, Standard conformations for the canonical structures of immunoglobulins, J. Mol. Biol. 273 (1997) 927–948.

[53] J.R. Adair, Engineering antibodies for therapy, Immunol. Rev. 130 (1992) 5–40.

[54] S.J. Smith-Gill, C.R. Mainhart, T.B. Lavoie, S. Rudikoff, M. Potter, VL-VH expression by monoclonal antibodies recognizing avian lysozyme, J. Immunol. 132 (1984) 963–967.

[55] R.M. MacCallum, A.C. Martin, J.M. Thornton, Antibody–antigen interactions: contact analysis and binding site topography, J. Mol. Biol. 262 (1996) 732–745.

[56] T.T. Wu, G. Johnson, E.A. Kabat, Length distribution of CDRH3 in antibodies, Proteins 16 (1993) 1–7.

[57] G. Johnson, T.T. Wu, Preferred CDRH3 lengths for antibodies with defined specificities, Int. Immunol. 10 (1998) 1801–1805.

[58] A.V.J. Collis, A.P. Brouwer, A.C.R. Martin, Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen, J. Mol. Biol. 325 (2003) 337–354.

[59] Y. Barrios, P. Jirholt, M. Ohlin, Length of the antibody heavy chain complementarity determining region 3 as a specificity-determining factor, J. Mol. Recognit. 17 (2004) 332–338.

[60] R.M. Hanson, J. Prilusky, R. Zhou, T. Nakane, J.L. Sussman, JSmol and the next-generation web-based representation of 3D molecular structure as applied to Proteopedia, Isr. J. Chem. 53 (2013) 207–216.