Computational and Structural Biotechnology Journal 15 (2017) 222–231

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Antibody H3 Structure Prediction

C. Marks*, C.M. Deane

Department of Statistics, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, United Kingdom

ARTICLE INFO

Article history:
Received 16 December 2016
Received in revised form 24 January 2017
Accepted 27 January 2017
Available online 1 February 2017

Keywords:
Antibodies
H3
Loop modelling
Protein structure prediction

ABSTRACT

Antibodies are proteins of the immune system that are able to bind to a huge variety of different substances, making them attractive candidates for therapeutic applications. Antibody structures have the potential to be useful during drug development, allowing the implementation of rational design procedures. The most challenging part of the antibody structure to experimentally determine or model is the H3 loop, which in addition is often the most important region in an antibody's binding site. This review summarises the approaches used so far in the pursuit of accurate computational H3 structure prediction.

1. Introduction

Antibodies are proteins that bind to foreign objects that find their way into an organism, preventing them from causing harm and marking them for removal. A huge number of different antibodies can be produced – estimates vary, but it is thought that humans have the potential to of produce up to 10^{13} different antibodies [1] – making them capable of binding to a huge range of substances, ranging from proteins on the cell surface of bacteria to non-biological small molecules [2]. The substance that an antibody binds to is known as an antigen, and the specific region of the antigen to which the antibody binds is called the epitope. Mature antibodies bind with high affinity and are specific, meaning that they bind to other epitopes only very weakly, or not at all [3].

The ability of antibodies to bind with high affinity and specificity to their targets means that they are good candidates for therapeutic and diagnostic applications. Since the first antibody treatment, muromonab, was approved in 1986 for the prevention of transplant rejection, the market has grown rapidly [4]. By 2012, antibody therapies accounted for over a third of the total sales in the biopharmaceutical sector in the US, and they are currently the biggest-selling class of biopharmaceuticals [5].

Although molecules from biological sources tend to be larger, more complex and far more difficult to characterise than traditional small molecule drugs [6], they are promising as therapeutic agents [7]. Antibodies have been used for many disease areas: some currently on the market include infliximab (Remicade) and adalimumab (Humira) for the treatment of rheumatoid arthritis; trastuzumab (Herceptin) and bevacizumab (Avastin) for cancer; and alemtuzumab (Lemtrada) for multiple sclerosis [8].

Knowledge of an antibody's structure is extremely useful when developing a novel therapeutic, allowing it to be engineered more rationally. This knowledge can be used to increase binding affinity by guiding residues to be mutated, through the use of computational techniques such as binding affinity prediction [e.g. Ref. 9], epitope and paratope prediction [10,11], stability measurements [e.g. Ref. 12], and docking [e.g. Ref. 13]. Computational tools have already been used successfully to increase the binding affinity of antibodies [e.g. Refs. 14;15;16;17;18]. However, since experimental structure determination is time-consuming and expensive, the ability to computationally build accurate models of antibody structures (in particular their antigen-binding sites) from their sequences is highly desirable. This has become even more important as next-generation sequencing (NGS) data for antibodies has become available [1, 19].

2. Antibody Structure and the H3 Loop

Antibodies vary from large, multi-chain and multi-domain complexes, like those found in humans, to small, single domain molecules, such as nanobodies [20]. However, binding always occurs in a similar fashion, through interactions between the antigen and a number of loops on the antibody called complementarity determining regions (or CDRs). In standard mammalian antibodies, there are six of these loops; three on the heavy chain and three on the light

* Corresponding author.
E-mail addresses: claire.marks@stats.ox.ac.uk (C. Marks), deane@stats.ox.ac.uk (C. Deane).

http://dx.doi.org/10.1016/j.csbj.2017.01.010

chain (labelled L1, L2, L3 and H1, H2, H3 respectively). In contrast, for camelid antibodies, which lack a light chain, there are only three. The CDRs are the most variable parts of the whole antibody structure, and they govern the majority of the antigen-binding properties of an antibody.

The conformational diversity of five of the six CDRs (L1, L2, L3, H1 and H2) is thought to be limited. For these CDRs, only a small number of different shapes have been observed, forming a set of discrete conformational classes known as canonical structures [21]. Since its proposal in 1987 [21], the idea has been reinvestigated many times as the number of known antibody structures has increased [e.g. Refs. 22-24]. These studies have led to the identification of particular amino acids at certain positions that are thought to be structure-determining; the canonical class of a CDR of unknown structure can therefore be predicted from its sequence with high accuracy. The least diverse CDR is L2, with around 99% of known structures belonging to the same class [24].

Unlike the other five CDRs, the H3 loop has not been classified into canonical forms; a huge range of structures have been observed (Fig. 1). This is due to how antibody sequences are encoded in the genome. The complete nucleotide sequence coding for an antibody heavy chain is created by combining gene segments from different locations (this is known as V(D)J recombination, after the 'variable', 'diversity', and 'joining' segments). The DNA encoding the H3 loop is found at the join between the V, D and J gene segments, which, with the addition of a process called junctional diversification, leads to a huge range of possible sequences. H3 loops vary widely in length: most are between 3 and 20 residues but they are occasionally far longer (Fig. 1). Bovine antibodies, for example, have H3s that are 50 or even 60 residues in length [26]. For comparison, the canonical CDRs each have a most 8 different lengths, and are normally far shorter - the longest canonical form is 17 residues long, but there are few examples of these five loops with lengths over 15 [23].

The 'torso' of H3 loops (the residues nearest to the anchors) has been observed to adopt one of two conformations, labelled kinked (or bulged) or extended (or non-bulged — see Fig. 2). The majority of H3 loops are kinked [23,27]. Proposals have been made about why this is the case, such as the interaction of a basic residue in the C-anchor with an asparagine located within the loop, which have led to the development of rules that aim to predict which conformation

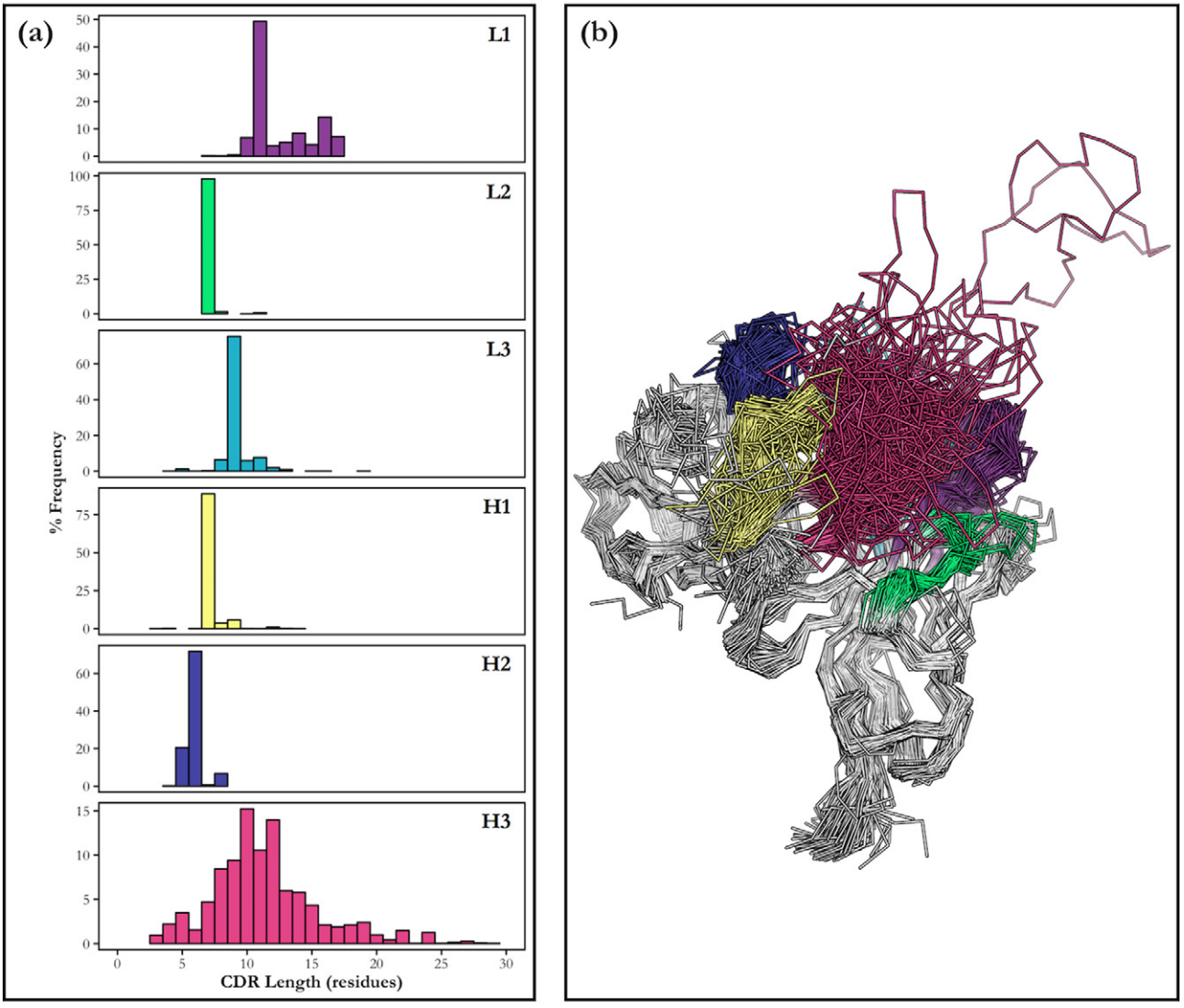

Fig. 1. (a) The frequency of observed loop lengths for the six CDRs. Data shown is calculated from all structures in SAbDab [25] . The H3 loop displays greater diversity in length than the canonical CDRs. (b) The structures of a set of antibodies with up to 80% sequence identity and a resolution of up to 3 Å, as downloaded from SAbDab [25] . Framework regions are shown in grey, while the CDRs are coloured (L1 — purple, L2 — green, L3 — blue, H1 — yellow, H2 — dark blue, H3 — pink). H3 loops display more conformational diversity than the other parts of the antibody. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)