Int. J. Mol. Sci. **2011**, *12*, 226-251; doi:10.3390/ijms12010226

OPEN ACCESS

International Journal of
Molecular Sciences
ISSN 1422-0067
www.mdpi.com/journal/ijms

Article

Computational Docking of Antibody-Antigen Complexes, Opportunities and Pitfalls Illustrated by Influenza Hemagglutinin

Mattia Pedotti †, Luca Simonelli †, Elsa Livoti and Luca Varani *

Institute for Research in Biomedicine, via Vela 6, 6500 Bellinzona, Switzerland;
E-Mails: mattia.pedotti@irb.unisi.ch (M.P.); luca.simonelli@irb.unisi.ch (L.S.);
elsa.livoti@irb.unisi.ch (E.L.)

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: luca.varani@irb.unisi.ch;
Tel.: +41-91-820-0321; Fax: +41-91-820-0302.

Received: 2 November 2010; in revised form: 22 December 2010 / Accepted: 4 January 2011 / Published: 5 January 2011

Abstract: Antibodies play an increasingly important role in both basic research and the pharmaceutical industry. Since their efficiency depends, in ultimate analysis, on their atomic interactions with an antigen, studying such interactions is important to understand how they function and, in the long run, to design new molecules with desired properties. Computational docking, the process of predicting the conformation of a complex from its separated components, is emerging as a fast and affordable technique for the structural characterization of antibody-antigen complexes. In this manuscript, we first describe the different computational strategies for the modeling of antibodies and docking of their complexes, and then predict the binding of two antibodies to the stalk region of influenza hemagglutinin, an important pharmaceutical target. The purpose is two-fold: on a general note, we want to illustrate the advantages and pitfalls of computational docking with a practical example, using different approaches and comparing the results to known experimental structures. On a more specific note, we want to assess if docking can be successful in characterizing the binding to the same influenza epitope of other antibodies with unknown structure, which has practical relevance for pharmaceutical and biological research. The paper clearly shows that some of the computational docking predictions can be very accurate, but the algorithm often fails to discriminate them from inaccurate

solutions. It is of paramount importance, therefore, to use rapidly obtained experimental data to validate the computational results.

Keywords: antibody modeling; computational docking; influenza; hemagglutinin; antibody-antigen complexes

1. Introduction

Individuals that recover from the attack of a pathogen have antibodies (Abs) capable of detecting and neutralizing the same pathogen in a future encounter, usually conferring life-long protection from it. Detection and neutralization are initiated by the binding of these antibodies to antigens, often surface proteins, through specific atomic interactions between the antibody and the region of the antigen (Ag) that it recognizes (epitope). A better understanding of these interactions is expected to accelerate vaccine development, since most current vaccines are based on the generation of neutralizing antibody responses. If we understand the structural rules governing Ab-Ag interactions in a given virus, for instance, then we have the molecular basis to attempt to design and synthesize new epitopes to be used as vaccines, optimize the antibodies themselves for passive immunization or design new drugs mimicking the antibodies or their effect.

In addition to pharmaceutical development, antibodies play an increasingly relevant role in basic research and industrial processes, where they are starting to be used as recognition elements sensitive to the presence of a given antigen. Designing and synthesizing new antibodies with desired properties would, therefore, have a profound impact, but we are very far away from being able to do that. Despite antibodies having been known and characterized for several decades [1,2], in fact, we still know remarkably little about their interactions. Given an antibody structure, for instance, we cannot even predict whether it can bind a protein, nucleic acid or sugar, let alone the specific antigen or conformational epitope that it recognizes. The study of Ab-Ag complexes should further our understanding of the general principles of recognition and, in the long run, gives us the basis for the successful design of new molecules or the rational optimization of existing ones.

The best way to study atomic interaction is to obtain the three-dimensional structure of antibody-antigen complexes. Traditionally, this is achieved by experimental techniques like X-ray crystallography, an often long and laborious process with high failure rate. Thanks to advances in algorithms and processing power, however, we can now use computational techniques for the structural characterization of intermolecular complexes. Computational docking—the process of predicting the conformation of a complex starting from its separated components—provides a fast and inexpensive route to obtain structures, including those which are not suitable for experimental determination. Although computational docking is still in its infancy and marred by several limitations, there is no doubt that it will become more and more accurate, relevant and widespread in the coming years.

Here we first illustrate the application of computational docking to the study of antibody-antigen interactions, and then highlight the strengths and weaknesses of the approach by predicting the binding of two different antibodies to hemagglutinin, the surface protein of influenza virus and an important

pharmaceutical target. Being able to accurately predict those structures, for which X-ray information is available, would strengthen our belief that computational techniques can be used to characterize the binding of new antibodies against the same epitope.

1.1. Computational Docking

Computational docking, a relatively new and constantly evolving technique, is the process of predicting the structure of a multi-molecular complex from the structures of its separated components. Its progress has been monitored since 2002 by the "Critical Assessment of PRediction of Interactions" project (CAPRI) [3], a comparative evaluation of protein-protein docking algorithms on a set of known targets. Here we focus on docking of antibodies to protein antigens, which presents specific challenges but also has peculiar features exploitable to ease the calculations.

In a typical docking protocol, the structures of the antigen and antibody are separated by approximately 25 Å and subsequently brought together by the chosen algorithm. The first necessary step, therefore, is obtaining the structures of the isolated antigen and antibody. The starting structure may be defined as follows:

(i) "Bound", if it originates from an experimental structure of the complex that needs to be docked. This is interesting when developing docking procedures but it is generally not biologically attractive, because computational docking is unlikely to add relevant information if an experimental structure is already available.

(ii) "Unbound", if it originates from an experimental structure of the molecule not bound to the partner that needs to be docked, *i.e.*, either free or bound to a different partner. This is the most common scenario for antigens, especially since the number of available protein structures is increasing thanks to several structural genomics efforts. Structures of free antibodies, instead, are usually not available, nor they would be particularly useful since Abs are known to drastically change conformation upon binding [4].

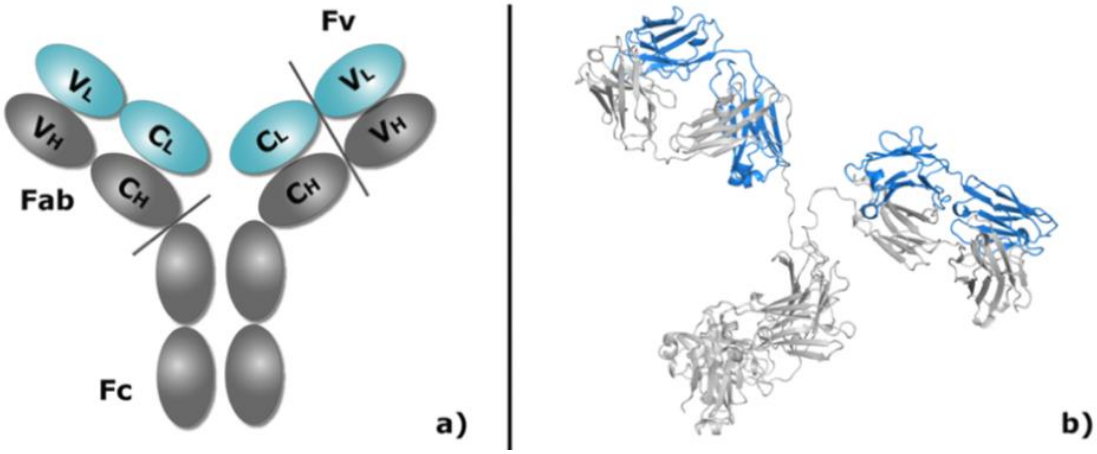
(iii) "Modeled", if it has been predicted by homology modeling and/or other computational techniques like *ab initio* predictions or molecular dynamics. A thorough description of homology modeling for protein antigens is beyond the scope of this manuscript. Suffice to say that the results are remarkably accurate if the target protein has sequence similarity to a protein with known structure and that even *ab initio* predictions are starting to produce accurate results, albeit much less than homology modeling [5–7]. Antibody structures can be predicted with remarkable accuracy and precision as well; the process is relatively different from standard protein modeling and is covered in the next sections.

1.2. Antibody Structure, Implications for Modeling

Antibodies are large (~150 kDa), y-shaped molecules containing a so-called Fc region (Fragment, Crystallizable, it binds to various cell receptors and mediates a response of the immune system) and two Fab regions (Fragment, Antigen Binding). The latter are composed by one heavy and one light chain, each with a constant and a variable domain called F_V (Figure 1). The F_V is the only domain responsible for antigen binding and, therefore, the only one that needs to be considered for docking. It

is further subdivided in a framework region, highly conserved in both sequence and conformation, and six highly variable CDR loops (Complementarity Determining Region), three from each chain and often referred to as L1, L2, L3, H1, H2, and H3.

Figure 1. Schematic (**a**) and cartoon (**b**) representation of a full antibody structure. Antigens bind to the tip of the V_H and V_L domains.



Despite their high sequence variability, five of the six loops (all except H3) can assume just a small repertoire of main-chain conformations, called "canonical structures" [5–7]. These conformations are determined by the length of the loops and by the presence of key residues at specific positions in the antibody sequence. The specific pattern of residues that determines each canonical structure forms a signature that can be recognized in the sequence of an antibody of unknown structure, allowing successful prediction of the canonical structure itself with high accuracy [8,9]. Uncertainties arise in the relatively rare cases when a loop is particularly long and/or does not follow canonical structures. The H3 loop does not appear to adopt canonical structures, instead, and predicting its conformation requires more sophisticated and less accurate approaches.

The framework regions can also be reliably predicted since known structures with high sequence identity are often available. Due to the presence of conserved residues at the interface between the light and heavy chain, the relative geometry of these domains is also well preserved [10]. Correct assembling of the heavy and light chain is nonetheless critical for the accurate orientation of the antigen binding interface and errors may arise in the modeling.

It is important to note that the rules and templates used for modeling are based on structures of antibodies bound to their antigen and are therefore accurate in the context of the bound conformation of an antibody.

1.3. Antibody Modeling Based on Canonical Structures, the PIGS Server

PIGS (Prediction of ImmunoGlobulin Structure [11]) is a web-based server for the automatic prediction of antibody structure [12] based on the canonical structure method [13]. The Web Antibody Modelling server, WAM [14], utilizes the same approach but offers less features and is generally less convenient to utilize.

In the canonical structure method, the sequence of each variable domain (V_L and V_H) of the antibody of unknown structure (target) is independently aligned with the corresponding variable domain sequences of all the immunoglobulins of known structure. For this step, standard database searching (e.g., BLAST) [15], and multiple sequence alignment (e.g., Clustalw) [16] programs can be used, but it is important to verify that residues at key structural positions are correctly aligned. The backbone structure of the framework is then modeled using the known structures with highest sequence identity as template. The rationale for this is that, in general, the higher the residue identity in the core of two proteins the more similar the conformation in this region [8] and, hence, the higher the quality of the model. Similarly, the conformation of the CDR loops is predicted using known templates with the same canonical loop conformation and high sequence identity. Different combinations of templates can be used as illustrated below.

(i) Best heavy and light chains. Use the chains with highest sequence identity as templates. Since they come from different antibodies, the two chains need to be packed together by a least-squares fit of the residues conserved at the interface. This may introduce errors in the relative orientation of the two chains, with adverse consequences for the accurate modeling of the antigen binding site.

(ii) Same canonical structures. Use a template whose CDR loops have the same canonical structures as the target even if a template with higher sequence identity exists for one or both chains. If framework and loops are taken from different templates, then the loops need to be grafted in, possibly introducing errors: the residues adjacent to the loop are superimposed to the framework by a weighted least-square fit of the main chain.

(iii) Same antibody. Use the same antibody as template for both heavy and light chain, even if templates with higher sequence identity exist. This does not require optimization of the relative orientation of the two chains and thus avoids the errors illustrated earlier.

(iv) Same antibody and canonical structures. The template is an antibody with the same canonical structures as the target and it is used to model both framework and the CDR loops. This option does not require optimization of framework orientation nor loop grafting and may offer more accurate results even if templates with higher sequence identity are available for one of the chains. The approach tends to fail, however, if the identity is too low.

The conformation of five of the six CDR loops can be modeled as described but no canonical structure is known for the H3 loop. However, the so-called “torso” region, *i.e.*, the H3 residues closer to the framework, can still be predicted by similarity to antibodies sharing the same torso conformation [17–19]. The “head” region of H3, instead, follows rules of standard protein hairpins and can be predicted by similarity to protein loops (not just antibodies) with high sequence identity, but the result is usually less accurate than for other CDR loops.

The subsequent step consists in the modeling of the side chains conformations. At sites where the parent structure and the model have the same amino acid the conformation of the parent structure is retained. Otherwise, the side chain conformation is copied from antibodies with high sequence similarity or imported from standard rotamer libraries [20]. Finally, the model is refined by a few cycles of energy minimization to improve the stereochemistry, especially in those regions where segments of structures coming from different immunoglobulins have been joined, but not to significantly refine the models.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.